The Annals of Statistics
2022, Vol. 50, No. 4, 1960–1991
https://doi.org/10.1214/21-AOS2165

LIMIT THEOREMS FOR DISTRIBUTIONS INVARIANT UNDER GROUPS OF TRANSFORMATIONS

BY MORGANE AUSTERN[1,a] AND PETER ORBANZ[2,b]

[1]*Department of Statistics, Harvard University,* [a]*maustern@fas.harvard.edu*
[2]*Gatsby Computational Neuroscience Unit, University College London,* [b]*porbanz@gatsby.ucl.ac.uk*

A distributional symmetry is invariance of a distribution under a group of transformations. Exchangeability and stationarity are examples. We explain that a result of ergodic theory implies a law of large numbers for such invariant distributions: If the group satisfies suitable conditions, expectations can be estimated by averaging over subsets of transformations, and these estimators are strongly consistent. We show that, if a mixing condition holds, the averages also satisfy a central limit theorem, a Berry–Esseen bound, and concentration. These are extended further to apply to triangular arrays, to randomly subsampled averages, and to a generalization of U-statistics. As applications, we obtain a general limit theorem for exchangeable random structures, and new results on stationary random fields, network models, and a class of marked point processes. We also establish asymptotic normality of the empirical entropy for a large class of processes. Some known results are recovered as special cases, and can hence be interpreted as an outcome of symmetry. The proofs adapt Stein's method.

1. Introduction. Statistical models that can be characterized by symmetry, or transformation invariance, include stationary processes [42], graphon and graphex models of networks [2, 5, 9, 14, 30, 45], the exchangeable random partitions that underpin much of Bayesian nonparametrics [24, 39], and rotation- and shift-invariant random fields [7, 26]. Examples from related fields are various models for relational data and preference prediction used in machine learning [35], point process representations of nearest neighbor methods and Voronoi tesselations [22, 23, 38], or self-similar stochastic processes [28]. Recent advances in spin glass theory rely crucially on exchangeable arrays [36].

We consider estimation under such invariant models. For each example above, a canonical estimator for expectations is known. We explain that these estimators are special cases of a general class of averages. For such averages, the ergodic theorem of Lindenstrauss [31] implies what a statistician would call a (strong) law of large numbers. Starting from this result, we establish central limit theorems, Wasserstein Berry–Esseen bounds—that is, Berry–Esseen type bounds that measure distance to a limiting variable in terms of the Wasserstein distance—and a concentration inequality. We then develop several applications in detail.

1.1. *Overview.* The remainder of this section is an informal summary of our approach, and of the main results. For the purposes of this introduction, we sidestep technicalities: A key quantity throughout is an infinite group $\mathbb{G}$. We assume for now that $\mathbb{G}$ is countable, and postpone general definitions to Section 2.

Consider a random element X of a space $\mathbf{X}$, and a real-valued function f. Suppose the group $\mathbb{G}$ consists of measurable bijections $\phi : \mathbf{X} \to \mathbf{X}$. We can then transform X by ϕ, where

Received May 2021; revised November 2021.
MSC2020 subject classifications. Primary 62G20; secondary 37A30, 62M99, 60F05, 60G09.
Key words and phrases. Asymptotic normality, central limit theorems, Berry–Esseen theorems, pointwise ergodic theorems, Stein's method, symmetry, exchangeability, ergodicity.

we use the notation $\phi(X)$ and ϕX interchangeably. The purpose of this work is to understand under what conditions the expectation $\mathbb{E}[f(X)]$ can be estimated by

$$\mathbb{F}_n(f, X) := \frac{1}{|\mathbf{A}_n|} \sum_{\phi \in \mathbf{A}_n} f(\phi X), \tag{1}$$

where $\mathbf{A}_1, \mathbf{A}_2, \ldots$ are finite subsets of $\mathbb{G}$, and $|\cdot|$ denotes cardinality. (For uncountable groups, $\mathbb{F}_n$ integrates over a compact set $\mathbf{A}_n$.) Such averages occur in dynamical systems [20] and statistical mechanics [37]. Various examples are used in statistics:

EXAMPLE.

(i) The window estimator for a random field on a grid [7, 26]. In this case, $X = (X_{ij})_{i,j \in \mathbb{Z}}$ is a collection of real-valued random variables. Let f be a function that depends only on the value at the origin, so $f(X) = g(X_{00})$ for some function g. A transformation that shifts the grid is of the form $\phi = (k, l)$ for some $k, l \in \mathbb{Z}$. If we choose $\mathbf{A}_n := \{-n, \ldots, n\}^2$, then

$$\mathbb{F}_n(f, X) = \frac{1}{|\mathbf{A}_n|} \sum_{(k,l) \in \mathbf{A}_n} f\big((X_{i+k, j+l})_{i,j \in \mathbb{Z}}\big) = \frac{1}{(2n+1)^2} \sum_{|i|,|j| \leq n} g(X_{ij}) \tag{2}$$

averages g over all locations on the subgrid of radius n around the origin. The group $\mathbb{G}$ is the group $\mathbb{Z}^2$ of all shifts, with addition as group operation.

More generally, X is a random object—such as a random sequence, matrix, field, or graph—and f is a function that typically depends only on "a small part" of X. The group $\mathbb{G}$ is a set of transformations that "move the domain" of f over X, and $\mathbf{A}_n$ contains those elements of $\mathbb{G}$ that cover a suitably defined sample, whose size is a function of n. The next two examples choose $\mathbf{A}_n$ as $\mathbb{S}_n$, the set of all permutations of the set $\{1, \ldots, n\}$.

EXAMPLES.

(ii) The sample average over a random sequence $X = (X_1, X_2, \ldots)$. Consider a function $f(X) = g(X_1)$ of the first entry, and let each permutation $\phi \in \mathbb{S}_n$ transform X by permuting entries, $\phi X := (X_{\phi(1)}, \ldots, X_{\phi(n)}, X_{n+1}, X_{n+2}, \ldots)$. Then

$$\mathbb{F}_n(f, X) = \frac{1}{|\mathbb{S}_n|} \sum_{\phi \in \mathbb{S}_n} f(\phi X) = \frac{1}{n!} \sum_{\phi \in \mathbf{A}_n} g(X_{\phi(1)}) = \frac{1}{n} \sum_{i \leq n} g(X_i). \tag{3}$$

In this case, the group is $\mathbb{G} = \bigcup_n \mathbb{S}_n$, the set of all finite permutations of $\mathbb{N}$.

(iii) The triangle density in network analysis [2, 5]. Here, X is a random undirected, simple graph with vertex set $\mathbb{N}$. Denote by $X[i_1, \ldots, i_k]$ the induced subgraph on the vertices $i_1, \ldots, i_k \in \mathbb{N}$. Let g be a function defined on graphs with three vertices, and set $f(X) := g(X[1, 2, 3])$. Suppose each $\phi \in \mathbb{S}_n$ transforms the graph by permuting the first n vertices, so $(\phi X)[1, 2, \ldots] = X[\phi(1), \ldots, \phi(n), n+1, n+2, \ldots]$. Then $\mathbb{F}_n$ averages g over all subgraphs of size 3 in the finite graph $X[1, \ldots, n]$:

$$\mathbb{F}_n(f, X) = \frac{1}{|\mathbb{S}_n|} \sum_{\phi \in \mathbb{S}_n} g\big(X\big[\phi(1), \phi(2), \phi(3)\big]\big) = \frac{1}{n(n-1)(n-2)} \sum g\big(X[i, j, k]\big),$$

where the sum on the right runs over all distinct triples $i, j, k \leq n$.

TABLE 1

Examples of invariant random structures and their ergodic special cases

$\mathbb{G}$-invariant objects X	$\mathbb{G}$-ergodic objects	ξ explained by	Eq. (7) specializes to
exchangeable sequences	i.i.d. sequences	de Finetti's theorem [28]	law of large numbers
stationary sequence	ergodic stationary sequ. [42]	see [42], Theorem I.4.10	Birkhoff's theorem [28]
exchangeable graphs	graphon models [10, 17]	Aldous–Hoover thm. [29]	graph limit convergence [10]
graphs generated by inv. point processes	graphex models [14]	Kallenberg's representation theorem [29]	empirical graphex [9, 45]
exchangeable arrays	dissociated arrays [29]	Aldous–Hoover theorem	Kallenberg's LLN [27]

Tools from ergodic theory. To characterize the behavior of $\mathbb{F}_n$, we borrow from ergodic theory: Two key conditions are

$$\text{(4)} \qquad \text{(i)} \quad \phi X \stackrel{\mathrm{d}}{=} X \quad \text{and} \quad \text{(ii)} \quad |\phi \mathbf{A}_n \cap \mathbf{A}_n|/|\mathbf{A}_n| \xrightarrow{n\to\infty} 1 \quad \text{for all } \phi \in \mathbb{G},$$

where $\stackrel{\mathrm{d}}{=}$ is equality in distribution. If (4)(i) holds, X is called $\mathbb{G}$-invariant. If it also satisfies

$$\text{(5)} \qquad P(X \in A) \in \{0, 1\} \quad \text{for every Borel set } A \text{ with } \phi A = A \text{ for all } \phi \in \mathbb{G},$$

it is called $\mathbb{G}$-ergodic. (Uncountable groups require more general formulations of (4)(ii) and (5), see Section 2.) The same terminology is applied to the distribution of X, so a $\mathbb{G}$-ergodic probability measure is the law of $\mathbb{G}$-ergodic random element, etc. Table 1 lists examples.

To motivate the conditions informally, first observe that $\mathbb{F}_n$ attempts to estimate $\mathbb{E}[f(X)]$ from surrogate values $f(\phi X)$. That should require $\mathbb{E}[f(X)] = \mathbb{E}[f(\phi X)]$, which is in turn implied by (4)(i). Any valid estimator $\mathbb{F}_n$ of $\mathbb{E}[f(X)]$ must satisfy $\mathbb{F}_n \approx \mathbb{E}[f(X)]$ in some suitable sense for large enough n, so it must also satisfy

$$\mathbb{F}_n(f, X) \approx \mathbb{E}\big[f(X)\big] = \mathbb{E}\big[f(\phi X)\big] \approx \mathbb{F}_n(f, \phi X).$$

That is true if $\phi \mathbf{A}_n \approx \mathbf{A}_n$, which is guaranteed by (4)(ii). In statistics, this condition was first used by Charles Stein, to characterize groups for which the Hunt–Stein theorem establishes minimaxity of invariant tests [8]. Ergodicity can be motivated as follows: We hope to establish strong consistency of estimates, that is, $\mathbb{F}_n(f, X) \to \mathbb{E}[f(X)]$ almost surely as $n \to \infty$. That means the event $\{\mathbb{F}_n(f, X) \to a\}$ must have probability 1 for $a = \mathbb{E}[f(X)]$, and 0 otherwise. Since invariance implies $\mathbb{E}[f(\phi X)] = \mathbb{E}[f(X)]$, these events are invariant sets for all $a \in \mathbb{R}$ (and indeed any invariant measurable set can be characterized in this way for some f and some invariant X). In this sense, $\mathbb{G}$-ergodic distributions form a class for which strong consistency might hold, provided one can establish a suitable strong law of large numbers.

This law of large numbers is due to Lindenstrauss [31]: If (4)(ii) holds, and X is $\mathbb{G}$-ergodic,

$$\text{(6)} \qquad \mathbb{F}_n(f, X) \xrightarrow{n\to\infty} \mathbb{E}\big[f(X)\big] \quad \text{almost surely}$$

for any function f with $\mathbb{E}[|f(X)|] < \infty$. The sets $\mathbf{A}_n$ must satisfy certain additional fine print, but they can always be modified to do so if they satisfy (4)(ii). Theorem 1 in Section 2 gives a proper statement.

The theorem can be extended to the $\mathbb{G}$-invariant case. The two cases are related by a property known as ergodic decomposition: $\mathbb{G}$-invariant distributions are mixtures of $\mathbb{G}$-ergodic ones. More formally, if X is $\mathbb{G}$-invariant, there is a random element ξ of the set of $\mathbb{G}$-ergodic distributions such that $X|\xi \sim \xi$ (see Theorem 2 for details). If X is $\mathbb{G}$-invariant, (6) becomes

$$\text{(7)} \qquad \mathbb{F}_n(f, X) \xrightarrow{n\to\infty} \mathbb{E}\big[f(X)|\xi\big] = \int f(x)\xi(dx) \quad \text{almost surely.}$$

For example, a random sequence $(X_i)_{i\in\mathbb{Z}}$ is stationary if it is $\mathbb{Z}$-invariant (adding elements of $\mathbb{Z}$ shifts the index set). In this case, ergodic decomposition is a classic result of time series theory [42], and (6) specializes to Birkhoff's ergodic theorem. An exchangeable (i.e., permutation-invariant) sequence is ergodic if it is i.i.d.—see Example (vii) for details. Thus, $X|\xi \sim \xi$ means X is "conditionally i.i.d.", which is de Finetti's theorem, and (6) is the strong law of large numbers.

Sketch of the main results. Our results provide rates of convergence for $\mathbb{F}_n$. Like certain convergence results for stationary processes, they use a mixing condition to control dependence within X: Recall that a typical mixing condition for a discrete-time process $(X_1, X_2, \dots)$ would be that any pair (X_j, X_k), for $j < k$, is approximately independent of the tail $(X_{k+n}, X_{n+k+1}, \dots)$ for large n [6]. Informally, we replace the tail by $(f(\psi X))_{\psi\in G}$, for a set $G \subset \mathbb{G}$, and require

$$\big(f(\phi_1 X), f(\phi_2 X)\big) \perp\!\!\!\perp \big(f(\psi X)\big)_{\psi\in G}|\xi \quad \text{approximately} \tag{8}$$

whenever $\phi_1, \phi_2 \in \mathbb{G}$ are far from G. The condition is tailored to second-order results, hence the pair on the left. Since $X|\xi \sim \xi$, conditional independence given ξ suffices. Section 3 gives a precise definition.

Our first result is a central limit theorem: If $\mathbb{E}[|f(X)|^{2+\varepsilon}] < \infty$ for some $\varepsilon > 0$, and the conditional mixing property above holds, then

$$\sqrt{|\mathbf{A}_n|}\big(\mathbb{F}_n(f, X) - \mathbb{E}[f(X)|\xi]\big) \xrightarrow{\mathrm{d}} \eta Z \quad \text{for } Z \sim N(0, 1).$$

The asymptotic variance η^2 is a random variable, independent of Z, and constant if X is $\mathbb{G}$-ergodic. That is Theorem 4. If $\mathbb{E}[|f(X)|^{4+2\varepsilon}] < \infty$, Theorem 6 bounds the approximation error as

$$d_{\mathrm{W}}\bigg(\frac{\sqrt{|\mathbf{A}_n|}}{\eta}\big(\mathbb{F}_n(f, X) - \mathbb{E}[f(X)|\xi]\big), Z\bigg) \le u(\mathbf{A}_n, \eta)$$

for a suitable function u and the Wasserstein distance d_{W}. This result generalizes the Berry–Esseen theorem (using d_{W} instead of total variation). In either case, the moment condition can be relaxed to $\varepsilon = 0$, at the price of stronger mixing.

In statistics, asymptotic normality results are often applied to quantify uncertainty. Theorem 5 shows that, if $z_{1-\frac{\alpha}{2}}$ is the $(1-\alpha)$-quantile of the standard normal distribution,

$$\limsup_{n\to\infty} P\bigg(\mathbb{E}[f(X)|\xi] \in \bigg[\mathbb{F}_n(f, X) - z_{1-\frac{\alpha}{2}}\frac{\hat{\eta}}{\sqrt{|\mathbf{A}_n|}}, \mathbb{F}_n(f, X) + z_{1-\frac{\alpha}{2}}\frac{\hat{\eta}}{\sqrt{|\mathbf{A}_n|}}\bigg]\bigg) \le \alpha$$

holds under the distribution P of X, where $\hat{\eta}$ is an empirical variance that can be computed from a sample of size n. In other words, the interval estimate above is a consistent confidence interval.

In Section 5, we generalize $\mathbb{F}_n$ along three lines: (i) f and X may change with n. (ii) Averages may be subsampled or randomized. In the simplest case, that means replacing $\mathbf{A}_n^{k_n}$ by a random subset $\widehat{\mathbf{A}}_n$, and generalizing $\mathbb{F}_n$ to

$$\widehat{\mathbb{F}}_n(f_n, X_n) = \frac{1}{|\widehat{\mathbf{A}}_n|}\sum_{\phi\in\widehat{\mathbf{A}}_n} f_n(\phi X_n) - \mathbb{E}[f_n(X_n)|\xi_n].$$

More generally, $\widehat{\mathbb{F}}_n$ is defined by a random measure μ_n on $\mathbf{A}_n$, so that random subsets are the special case where μ_n is uniform on $\widehat{\mathbf{A}}_n$. (iii) Each ϕ may be substituted by a vector of transformations, with some number k_n of elements, replacing $\mathbb{G}$ by $\mathbb{G}^{k_n}$ and $\mathbf{A}_n$ by $\mathbf{A}_n^{k_n}$. Our main results are a central limit theorem (Theorem 10) and a Wasserstein Berry–Esseen bound (Theorem 11) for $\widehat{\mathbb{F}}_n$. We use the result for k_n-tuples to formulate a class of generalized U-statistics (Corollary 12).

Since certain asymptotic properties of i.i.d. sequences generalize to $\mathbb{G}$-invariant objects, it is natural to ask whether finite-sample properties do so, too. Section 6 gives a concentration inequality of the form

$$\mathbb{P}\big(\widehat{\mathbb{F}}_n(f, X) \geq t\big) \leq 2e^{-\omega_n t^2} \quad \text{for all } t > 0,$$

for certain constants ω_n, where $\mathbb{P}$ denotes probability under the joint distribution of X and the (possibly randomized) average $\widehat{\mathbb{F}}_n$.

1.2. *Applications*. The remaining sections apply the theorems summarized above to obtain new results for a number of specific models, and also highlight how certain known results can be phrased as instances of invariance. We consider two specific types of invariance—stationarity and exchangeability—in some detail because of their importance to statistics, but also discuss other applications, to point processes and entropy.

For stationary random fields, the group $\mathbb{G}$ plays a dual role, as index set and a set of shifts. Substituting a field indexed by the grid $\mathbb{Z}^d$ into Theorem 4 recovers Bolthausen's central limit theorem [7]. Substituting other groups generalizes this result, as Corollary 7 illustrates for a continuous field on $\mathbb{G} = \mathbb{R}^d$. Corollary 8 is a Berry–Esseen bound for Bolthausen's theorem. If $\mathbb{G}$ is uncountable, the estimator $\mathbb{F}_n$ becomes an integral. In applications where computing the integral is not be feasible, it can be discretized to a sum, by applying Theorems 10 and 11, with μ_n chosen as almost surely discrete. Making μ_n nonrandom makes the discretization deterministic. Corollary 16 illustrates both cases, again for a continuous random field.

Modeling assumptions made in statistics often imply some form of invariance under the permutation group $\mathbb{S}_\infty$. Examples are i.i.d. sequences, Bayesian models appealing to de Finetti's theorem (which involve exchangeable sequences), stochastic block models and graphon models (which generate exchangeable graphs), and finite and Dirichlet process mixtures (which generate exchangeable partitions). Theorem 17 is a general central limit theorem for such models: Any $\mathbb{S}_\infty$-invariant random object X satisfies

$$\sqrt{n}\bigg(\frac{1}{n!}\sum_{\phi \in \mathbb{S}_n} f(\phi X) - \mathbb{E}\big[f(X)|\mathbb{S}_\infty\big]\bigg) \xrightarrow{d} \eta Z \quad \text{where } \eta \perp\!\!\!\perp Z \sim N(0,1).$$

The result does not require a mixing condition.

Some models do not make an invariance assumption on the data source, but rather use invariant random objects as approximations or latent variables. One example are nonparametric stochastic block models that increase the number of "communities" in the model with sample size [15]. An observed graph with n vertices is explained by an exchangeable graph X_n, whose distribution changes with n. Corollary 18 shows how to estimate a statistic, where we choose the triangle density for illustration. Informally,

$$\frac{\sqrt{n}}{\eta_n}\big(\text{empirical triangle density}(n) - \text{population triangle density under } X_n\big) \xrightarrow{d} Z,$$

where η_n is determined by the law of X_n. Another example are graphex random graphs [9, 14, 45], which are not themselves exchangeable, but generated by a latent point process with an exchangeability property. Section 8.5 explains how to apply Theorem 17 to such models by extracting an exchangeable surrogate object. Corollary 20 is an example: The standard estimator for the graphex equivalent of the edge density satisfies

$$\sqrt{s}\big(\text{empirical graphex subgraph density}(s) - \text{graphex subgraph density}\big) \xrightarrow{d} \eta Z,$$

where s is the relevant notion of sample size.

In Section 9, we obtain a central limit theorem and the speed of convergence for so-called random geometric measures. These are point processes with dependence between neighboring points, and have been applied to nearest-neighbor methods and tesselations. Whether points are neighbors is defined by shifting an "observation window" over the sample space. Representing these shifts as elements of a group makes Theorems 4 and 6 applicable.

Section 10 concerns entropy: The entropy of a stochastic process is defined as a limit of so-called empirical entropies, computed from the first n values of the process. This definition can be extended to certain invariant random objects, by defining the nth empirical entropy using the transformation set $\mathbf{A}_n$. The fact that the limit exists is, in the classical case, known as the Shannon–McMillan–Breiman theorem. Lindenstrauss [31] has generalized it to the invariant case. Theorem 23 establishes asymptotic normality: Under suitable conditions,

$$\sqrt{|\mathbf{A}_n|}\big(\text{empirical entropy}(n) - \text{entropy}\big) \xrightarrow{d} \eta Z \quad \text{as } n \to \infty.$$

2. Background and definitions. Throughout, $\mathbb{G}$ is a group, with identity element e. By $\mathbf{X}$, we always mean a standard Borel space, with Borel σ-algebra $\mathcal{B}(\mathbf{X})$, and by $\mathcal{P}(\mathbf{X})$ the space of probability measures on $\mathbf{X}$, topologized by weak convergence. For a random element X of $\mathbf{X}$, and $p > 0$, define the norm $\|f\|_p := \mathbb{E}[|f(X)|^p]^{1/p}$ for measurable functions $f : \mathbf{X} \to \mathbb{R}$. The set of functions with $\|f\|_p < \infty$ is denoted $\mathbf{L}_p(X)$. By $f \in \mathbf{L}_p(X)$, we refer to a function f, rather than an equivalence class.

2.1. *Conditions on the group*. To explain the estimator (1) for an uncountable group $\mathbb{G}$, we must define a topology and a measure on $\mathbb{G}$. Finite sets then generalize to compact ones, and sums over group elements to integrals. To cohere with group structure, the topology must make the group operation continuous. If that is the case, and the topology is locally compact, second-countable, and Hausdorff, or lcscH, then $\mathbb{G}$ is a *lcscH group*. If $\mathbb{G}$ is countable, the discrete topology is lcscH, and $\mathbb{G}$ is a *discrete group*. We always equip $\mathbb{G}$ with its Borel σ-algebra $\mathcal{B}(\mathbb{G})$. On every lcscH group, there is a σ-finite measure $|\cdot|$ that satisfies

$$|\phi^{-1}A| = |A| \quad \text{for all } \phi \in \mathbb{G} \text{ and } A \in \mathcal{B}(\mathbb{G}), \tag{9}$$

called a *Haar measure*. It is unique up to positive scaling, so $c|\cdot|$ is again a Haar measure for $c > 0$ [28]. If a set $A \subset \mathbb{G}$ is compact, then $|A| < \infty$. Informally, Haar measures generalize volume, and (9) shows that a set can be shifted without changing its volume. Examples of Haar measures are Lebesgue measure on the groups $(\mathbb{R}^r, +)$, for $r \in \mathbb{N}$, or counting measure (cardinality) on a discrete group. Our results do not assume a specific scaling c, but in examples we always choose $|\cdot|$ as cardinality if $\mathbb{G}$ is discrete.

Like volume, distance can be defined in a shift-invariant way: If $\mathbb{G}$ is lcscH, there exists a metric d on $\mathbb{G}$ that is *left-invariant*,

$$d(\phi^{-1}\cdot, \phi^{-1}\cdot) = d(\cdot, \cdot) \quad \text{for all } \phi \in \mathbb{G}. \tag{10}$$

We write $\mathbf{B}_t(\phi) := \{\psi \in \mathbb{G} | d(\psi, \phi) \le t\}$ for a metric ball centered at ϕ, and abbreviate by $\mathbf{B}_t := \mathbf{B}_t(e)$ a metric ball around the identity. One can always choose a left-invariant metric on $\mathbb{G}$ such that $\mathbf{B}_n$ "grows evenly" with n,

$$\frac{|\mathbf{B}_{n+1} \setminus \mathbf{B}_n|}{|\mathbf{B}_n \setminus \mathbf{B}_{n-1}|} = O(1), \tag{11}$$

see [32]. If G and A are sets in $\mathbb{G}$, we write $GA := \{\phi\psi | \phi \in G, \psi \in A\}$. A *Følner sequence* is a sequence of compact sets $\mathbf{A}_1, \mathbf{A}_2, \ldots \subset \mathbb{G}$ such that

$$\frac{|G\mathbf{A}_n \cap \mathbf{A}_n|}{|\mathbf{A}_n|} \xrightarrow{n \to \infty} 1 \quad \text{for every compact } G \subset \mathbb{G}. \tag{12}$$

If $\mathbb{G}$ is discrete, its compact sets are the finite sets, and (12) is equivalent to (4)(ii). A lcscH group that contains a Følner sequence is called *amenable* [20]. A Følner sequence is *tempered* if

(13) $$\Big|\bigcup_{k<n}\mathbf{A}_k^{-1}\mathbf{A}_n\Big| \leq c|\mathbf{A}_n| \quad \text{for some } c>0 \text{ and all } n\in\mathbb{N}.$$

Not every Følner sequence is tempered, but every lcscH group containing a Følner sequence also contains a tempered Følner sequence [31], Proposition 1.4.

CONVENTION. We use the shorthand *nice group* for an amenable lcscH group $\mathbb{G}$ equipped with a metric d satisfying (10) and (11).

EXAMPLES.

(iv) The group $\mathbb{S}_\infty$ of all permutations of $\mathbb{N}$ with finite support: Define $\mathbb{S}_n$ as the group of permutations of $\{1,\dots,n\}$, and $\mathbb{S}_\infty := \bigcup_{n\in\mathbb{N}}\mathbb{S}_n$. The canonical metric on $\mathbb{S}_\infty$ is

(14) $$d(\phi,\phi') := \min\{n\in\mathbb{N}|\phi(n,n+1,\dots)=\phi'(n,n+1,\dots)\}.$$

The sequence $(\mathbb{S}_n)$ is a tempered Følner sequence: Each $\phi\in\mathbb{G}$ is in $\mathbb{S}_n$ for n sufficiently large, so $\phi\mathbb{S}_n\cap\mathbb{S}_n=\mathbb{S}_n$ eventually, and (4)(ii) holds. Since $\mathbb{S}_k^{-1}\mathbb{S}_n=\mathbb{S}_n$ whenever $k\leq n$, the sequence is tempered.

(v) The shifts of the r-dimensional grid $\mathbb{Z}^r$ form the group $(\mathbb{Z}^r,+)$: An element $\mathbf{j}$ of the group shifts a grid point $\mathbf{i}$ to $\mathbf{i}+\mathbf{j}$. Its canonical metric

(15) $$d(\mathbf{i},\mathbf{j}) = \min_{k\leq r}|i_k - j_k|$$

is left-invariant and satisfies (11). The balls $\mathbf{B}_n=\{-n,\dots,n\}^r$, for $n\in\mathbb{N}$, form a tempered Følner sequence, and so do the sets $\{1,\dots,n\}^r$.

(vi) Similarly, $(\mathbb{R}^r,+)$ is the shift group of $\mathbb{R}^r$. Lebesgue measure is a Haar measure, Euclidean distance is a left-invariant metric satisfying (11), and the balls $\mathbf{B}_n$ and the sets $[0,n]^r$ both form tempered Følner sequences.

Recall from the Introduction that $|\mathbf{A}_n|$ can be interpreted as sample size. If $\mathbb{G}$ is compact, $|\mathbf{A}_n|\leq|\mathbb{G}|<\infty$. It is hence essential for asymptotics that $\mathbb{G}$ is not compact. Examples of nice, noncompact groups include the groups above, the group $(\mathbb{R}_{>0},\cdot)$ (which characterizes self-similarity of stochastic processes), the group of translations and rotations of a Euclidean space, and discrete and continuous Heisenberg groups [18]. See [20, 32] for more.

2.2. *Invariance and ergodicity.* We now let elements of $\mathbb{G}$ transform elements of a space $\mathbf{X}$. We must specify what that means: Permuting a matrix, say, could mean permuting rows, or columns, or entries. Such a specification is called an action: A *measurable action* of $\mathbb{G}$ on $\mathbf{X}$ is a jointly measurable map $(\phi,x)\mapsto T_\phi(x)$ that satisfies

(16) $$T_e(x)=x \quad\text{and}\quad T_{\phi\phi'}(x)=T_\phi(T_{\phi'}(x)) \quad \text{for } x\in\mathbf{X} \text{ and } \phi,\phi'\in\mathbb{G}.$$

The conditions ensure that the set of transformations T_ϕ defined by $\mathbb{G}$ on $\mathbf{X}$ is itself a group. We usually simplify notation and write $\phi(x):=T_\phi(x)$. A random element X of $\mathbf{X}$ with distribution P is $\mathbb{G}$-*invariant* if

$$\phi(X)\stackrel{\mathrm{d}}{=}X \quad\text{or equivalently}\quad P=P\circ\phi^{-1} \quad\text{for all } \phi\in\mathbb{G}.$$

We then call P a $\mathbb{G}$-invariant measure. A Borel set $A\in\mathcal{B}(\mathbf{X})$ is *almost invariant* if $P(\phi A\,\triangle\,A)=0$ for all $\phi\in\mathbb{G}$ and all $\mathbb{G}$-invariant P, where $\triangle$ denotes symmetric difference.

The almost invariant sets form a σ-algebra $\sigma(\mathbb{G})$, and we abbreviate conditioning on $\sigma(\mathbb{G})$ as

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

A probability measure is $\mathbb{G}$-*ergodic* if it is $\mathbb{G}$-invariant and $P(A) \in \{0, 1\}$ for all $A \in \sigma(\mathbb{G})$. This condition is equivalent to (5) if $\mathbb{G}$ is countable [20]. A random element is $\mathbb{G}$-ergodic if its distribution is.

2.3. *Estimation*. We now come to the general form of the estimator (1). For a group $\mathbb{G}$ acting measurably on $\mathbf{X}$, a Følner sequence $(\mathbf{A}_n)$ on $\mathbb{G}$, and a Borel function f on $\mathbf{X}$, define

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

If $\mathbb{G}$ is discrete, $\mathbb{F}_n$ simplifies to the sum (1). The cornerstone of our work is a result of Lindenstrauss, which concluded a long line of work by Ornstein, Weiss, and others (e.g., [46]).

THEOREM 1 (E. Lindenstrauss [31]). *If a random element X of a standard Borel space is invariant under a measurable action of a nice group, and if $(\mathbf{A}_n)$ is a tempered Følner sequence, then*

$$\mathbb{F}_n(f, X) \xrightarrow{n\to\infty} \mathbb{E}\big[f(X)|\mathbb{G}\big] \quad \textit{almost surely for all } f \in \mathbf{L}_1(X), \tag{17}$$

where $\mathbb{E}[f(X)|\mathbb{G}] = \mathbb{E}[f(X)]$ almost surely if X is ergodic.

Where convenient, we center $\mathbb{F}_n$ around the limit as

$$\overline{\mathbb{F}}_n(f, X) := \mathbb{F}_n(f, X) - \mathbb{E}\big[f(X)|\mathbb{G}\big]. \tag{18}$$

The next result gives an interpretation of the limit: If X is invariant, it can be generated by selecting an ergodic measure ξ at random, and then drawing X from ξ. The limit $\mathbb{E}[f(X)|\mathbb{G}]$ is the expectation of f under the instance of the latent measure ξ that has generated X.

THEOREM 2 (Ergodic decomposition, Varadarajan [44]). *If a lcscH group $\mathbb{G}$ acts measurably on a standard Borel space $\mathbf{X}$, the set of $\mathbb{G}$-invariant probability measures is convex. Its set of extreme points is the set $\mathbf{E}$ of $\mathbb{G}$-ergodic measures, and is measurable in $\mathcal{P}(\mathbf{X})$. A random element X of $\mathbf{X}$ is $\mathbb{G}$-invariant if and only if*

$$P[X \in \cdot|\mathbb{G}] = \xi(\cdot) \quad \textit{almost surely} \tag{19}$$

for a random element ξ of $\mathbf{E}$. The law of ξ is uniquely determined by that of X.

Thus, conditioning on $\sigma(\mathbb{G})$ means conditioning on ξ. Another implication is that $\xi = P$ almost surely if P is itself ergodic, and therefore

$$\mathbb{E}\big[f(X)|\mathbb{G}\big] = \int f(x)\,d\xi(x) = \mathbb{E}\big[f(X)\big] \quad \text{if } X \text{ is ergodic.}$$

Taking expectations on both sides of (19) shows that P is $\mathbb{G}$-invariant if and only if

$$P(X \in \cdot) = \int_{\mathbf{E}} m(\cdot)\mathbb{P}(\xi \in dm).$$

That provides a more geometric interpretation: The integral represents P as a generalized convex combination of extreme points, also known as a barycenter. Recall that every element of a polytope in Euclidean space is a convex combination of extreme points. The theorems of Krein–Milman and Choquet generalize this property from polytopes to certain compact convex sets [1]. Theorem 2 shows that, if the elements of the convex set are specifically $\mathbb{G}$-invariant measures, compactness is not required.

EXAMPLES.

(vii) Let $\mathbf{X}$ be the space $\mathbb{R}^{\mathbb{N}}$ of real-valued sequences. Define an action of the permutation group $\mathbb{S}_\infty$ as $\phi(x) := (x_{\phi(1)}, x_{\phi(2)}, \dots)$, for $x \in \mathbf{X}$ and $\phi \in \mathbb{S}_\infty$. An *exchangeable sequence* is a $\mathbb{S}_\infty$-invariant random sequence $X = (X_i)_{i\in\mathbb{N}}$. It is ergodic if and only if it is i.i.d., a fact known as the Hewitt–Savage 0–1 law [28]. It follows that ξ factorizes as $\xi = \xi_0^{\otimes\mathbb{N}}$, for some random probability measure ξ_0 on $\mathbb{R}$. Theorem 2 then takes the form

$$P(X \in \cdot) = \int_{\mathcal{P}(\mathbb{R}^{\mathbb{N}})} m(\cdot)\mathbb{P}(\xi \in dm) = \int_{\mathcal{P}(\mathbb{R})} m_0^{\otimes\mathbb{N}}(\cdot)\mathbb{P}(\xi_0 \in dm_0),$$

which is de Finetti's theorem [28]. Let $f(x) = g(x_1)$ be a function of the first sequence entry, as in (3). Theorem 1 becomes

$$\frac{1}{n!}\sum_{\phi\in\mathbb{S}_n} f(\phi X) = \frac{1}{n}\sum_{i\le n} g(X_i) \xrightarrow{n\to\infty} \int_{\mathbb{R}} g(x_1)\xi_0(dx_1) \quad \text{a.s.}$$

For ergodic X, this is the strong law of large numbers for i.i.d. sequences.

(viii) Fix $r \in \mathbb{N}$, and let $\mathbf{X} = \mathbb{R}^{\mathbb{Z}^r}$ be the set of scalar fields $x = (x_i)_{i\in\mathbb{Z}^r}$ on an r-dimensional grid. Define an action of $\mathbb{G} = \mathbb{Z}^r$ on $\mathbf{X}$ as $\phi(x) := (x_{i+\phi})_{i\in\mathbb{Z}^r}$ for $\phi \in \mathbb{Z}^r$. A *stationary random field* is a $\mathbb{Z}^r$-invariant random element X of $\mathbf{X}$. Recall from Example (v) that $\mathbf{A}_n = \{-n, \dots, n\}^r$ defines a Følner sequence. Write $\Omega_n := \{-n, \dots, n\}^r$ to distinguish the subset Ω_n of the *index* set $\mathbb{Z}^r$ from the subset $\mathbf{A}_n$ of the *group* $\mathbb{Z}^r$. In this case, (12) can be rephrased in terms of the index set: Since $\Omega_n := \mathbf{A}_n(0, \dots, 0)$,

$$|\partial\Omega_n|/|\Omega_n| \xrightarrow{n\to\infty} 0 \quad \text{where } \partial\Omega_n = \Omega_n \setminus \Omega_{n-1}.$$

In this form, the condition is well known in statistics [7, 26]. For a function $f(x) = g(x_{0,\dots,0})$ at the origin, $\mathbb{F}_n$ is given by (2). We also noted already that $\mathbf{A}_n$ can alternatively be chosen as $\{1, \dots, n\}^r$. For the case $r = 1$ of stationary sequences, Theorem 1 then takes the form $n^{-1}\sum_{i=1}^n g(X_i) \to \mathbb{E}[g(X_1)|\mathbb{G}]$, which is Birkhoff's ergodic theorem [42].

3. Conditional mixing. This section formalizes the mixing condition sketched in (8). The label "mixing" is used for a range of conditions, whose common denominator is typically that they quantify dependence using terms of the form $|P(A)P(B) - P(A \cap B)|$. Their strengths and purposes vary—an extensive list of mixing conditions for stationary processes, for example, is surveyed by Bradley [12]. Ergodic theory defines mixing conditions to verify ergodicity, which are typically much weaker [20]. Our notion of mixing more closely resembles that used in random field asymptotics [7, 21].

Fix $f \in \mathbf{L}_1(X)$. Given a set $G \subset \mathbb{G}$, the events in $\mathbf{X}$ that can be formulated in terms of $(f(\phi X))_{\phi\in G}$ form the σ-algebra

$$\sigma_f(G) := \sigma(f \circ \phi, \phi \in G) = \sigma\Bigl(\bigcup_{\phi\in G} (f \circ \phi)^{-1}\mathcal{B}(\mathbb{R})\Bigr),$$

where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra of $\mathbb{R}$. Write $\mathbf{B}_t(G) := \bigcup_{\phi\in G} \mathbf{B}_t(\phi)$. The set of group elements whose distance from G exceeds t is $\mathbb{G} \setminus \mathbf{B}_t(G)$. The set of events we consider is

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

The *mixing coefficient* for f and P is the function

$$\alpha(t) := \sup_{(A,B)\in\mathcal{C}(t)} \bigl|P(A)P(B) - P(A \cap B)\bigr| \quad \text{for } t > 0,$$

and P is *mixing* with respect to f if $\alpha(t) \to 0$ as $t \to \infty$. Similarly,

$$\alpha(t|\mathbb{G}) := \sup_{(A,B)\in\mathcal{C}(t)} \mathbb{E}\big[\big|P(A|\mathbb{G})P(B|\mathbb{G}) - P(A\cap B|\mathbb{G})\big|\big] \quad \text{for } t > 0$$

is the *conditional mixing coefficient*, and P is *conditionally mixing* if $\alpha(t|\mathbb{G}) \to 0$ as $t \to \infty$. Both coefficients are decreasing in t, since $\mathcal{C}(t_1) \supset \mathcal{C}(t_2)$ if $t_1 \leq t_2$.

LEMMA 3. *The mixing coefficients satisfy* $\alpha(k|\mathbb{G}) \leq 4\alpha(k)$ *for all* $k \in \mathbb{N}$.

Thus, mixing implies conditional mixing. The first example below shows that the converse need not be true. The second example describes a case where both properties hold.

EXAMPLES.

(ix) Any exchangeable sequence $X = (X_1, X_2, \ldots)$ is conditionally mixing with respect to $f : (x_1, x_2, \ldots) \mapsto x_1$: By de Finetti's theorem, its entries are conditionally independent. For subsets $F, G \subset \mathbb{N}$, that implies

$$(X_i)_{i\in F} \perp\!\!\!\perp (X_j)_{j\in G}|\mathbb{G} \quad \text{if} \min_{i\in F, j\in G} |i-j| \geq 1,$$

and hence $\alpha(k|\mathbb{G}) = 0$ for all $k \in \mathbb{N}$. It need not be mixing: Draw once from a random variable Y, and set $X_i := Y$ for all $i \in \mathbb{N}$. Then X is exchangeable, but dependence of X_1 and X_i does not diminish as i grows.

(x) Let $X = (X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^r}$ be a stationary random field with the Markov property: For each $\mathbf{i} \in \mathbb{Z}^r$, $X_{\mathbf{i}} \perp\!\!\!\perp (X_{\mathbf{j}})_{\mathbf{j}\in\mathbb{Z}^d\setminus\{\mathbf{i}\}}|(X_{\mathbf{j}})_{\mathbf{j}\in\mathbf{B}_1(\mathbf{i})}$. Suppose X satisfies the so-called Dobrushin condition,

$$\vartheta := \sup_{\mathbf{i}|d(\mathbf{i},0)=1} \sup_{A,B\in\mathcal{B}(\mathbf{X})} \big|P(X_0 \in A|X_{\mathbf{i}} \in B) - P(X_0 \in A)\big| \leq \frac{1}{2r}.$$

If so, it is mixing with respect to all coordinate functions: There are positive constants c_1 and c_2 such that $\alpha(k) \leq c_1 e^{-c_2 k}$ for all $k \in \mathbb{N}$ (e.g., [21], 8.28). By Lemma 3, that also implies conditional mixing. In general, if (X_i) is a stationary sequence, $\alpha(\cdot|\mathbb{G})$ can be bounded by the classical α-mixing coefficients (e.g., [12]).

(xi) If X is conditionally mixing for f, it is conditionally mixing for $g \circ f$, for any measurable function g.

4. Basic limit theorems. The central limit theorem requires conditional mixing and a second-moment condition. The strength of each can be traded off against the other: The theorems in this section assume either

$$\text{(20)} \qquad \text{(i)} \quad \mathbb{E}\big[f(X)^2\big] < \infty, \qquad \text{(ii)} \quad \alpha(K|\mathbb{G}) = 0 \quad \text{for some } K \in \mathbb{N},$$

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

$$\text{(21)} \qquad \text{(i)} \quad \mathbb{E}\big[f(X)^{2+\varepsilon}\big] < \infty, \qquad \text{(ii)} \quad \int_{\mathbb{G}} \alpha\big(d(e,\phi)|\mathbb{G}\big)^{\frac{\varepsilon}{2+\varepsilon}} |d\phi| < \infty,$$

where e is the identity element of $\mathbb{G}$. If $\mathbb{G}$ is discrete, (21)(ii) simplifies to

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

We note only en passant that the quantity $|\mathbf{B}_{n+1} \setminus \mathbf{B}_n|$ plays a crucial role in group theory, where it is known as the growth rate of $\mathbb{G}$ [32].

THEOREM 4. *Let $\mathbb{G}$ be a nice group with tempered Følner sequence $(\mathbf{A}_n)$, acting measurably on a standard Borel space $\mathbf{X}$. If a $\mathbb{G}$-invariant random element X of $\mathbf{X}$ and a function $f : \mathbf{X} \to \mathbb{R}$ satisfy either (20) or (21), then*

$$\sqrt{|\mathbf{A}_n|}\big(\mathbb{F}_n(f, X) - \mathbb{E}[f(X)|\mathbb{G}]\big) \xrightarrow{d} \eta Z \quad \text{for } Z \sim N(0,1). \tag{22}$$

The asymptotic variance η^2 is a random variable distributed as

$$\eta^2 \stackrel{\mathrm{d}}{=} \int_{\mathbb{G}} \eta^2(\phi)|d\phi| \quad \text{for } \eta^2(\phi) := \mathbb{E}[f(X)f(\phi X)|\mathbb{G}], \tag{23}$$

and satisfies $\eta^2 < \infty$ almost surely. It is independent of Z, and constant almost surely if X is $\mathbb{G}$-ergodic.

The rate of convergence in Lindenstrauss' theorem is thus $|\mathbf{A}_n|^{-\frac{1}{2}}$, and depends only on the Følner sequence. The choice of action does not affect the rate, but does affect the mixing coefficient and constants. The ergodic decomposition property is visible in the independence of η and Z: Theorem 2 shows $\mathbb{E}[\cdot|\mathbb{G}] = \mathbb{E}[\cdot|\xi]$, so η is a function of ξ, and constant if X is $\mathbb{G}$-ergodic. Informally, the randomness of Z is due to $X|\xi$, that of η is due to ξ.

In statistical terms, $\mathbb{F}_n(X, f)$ is an estimate of $\mathbb{E}[f(X)|\mathbb{G}]$ computed from a sample of size $|\mathbf{A}_n|$. Theorem 1 shows this estimator is (strongly) consistent, and Theorem 4 provides the rate of convergence and shows the estimation error is asymptotically normal. That can be used to obtain a consistent confidence interval, as the next result shows. The additional condition on η ensures—in the nonergodic case, where η is not almost surely constant—that its law does not place too much mass very close to 0, which could lead to effectively degenerate behavior even if $\eta > 0$ almost surely.

THEOREM 5. *Assume the conditions of Theorem 4, and additionally that $P(\eta < t) \to 0$ if $t \searrow 0$. Let (b_n) be an increasing sequence of positive integers satisfying*

$$\text{(i)} \quad b_n \to \infty, \qquad \text{(ii)} \quad |\mathbf{B}_{b_n}| = o\big(\sqrt{|\mathbf{A}_n|}\big), \qquad \text{(iii)} \quad |\mathbf{A}_n \setminus \mathbf{B}_{b_n}\mathbf{A}_n| = o\big(|\mathbf{A}_n|\big),$$

and define the empirical variance

$$\hat{\eta}_n^2 := \frac{1}{|\mathbf{A}_n|}\int_{\mathbf{A}_n}\int_{\mathbf{B}_{b_n}(\phi)} \big(f(\phi X) - \mathbb{F}_n(f, X)\big)\big(f(\phi' X) - \mathbb{F}_n(f, X)\big)|d\phi'||d\phi|.$$

For any $\alpha \in (0,1)$, let $z_{1-\frac{\alpha}{2}}$ be the positive scalar satisfying $P(|Z| > z_{1-\frac{\alpha}{2}}) = \alpha$. Then

$$\limsup_{n\to\infty} P\left(\mathbb{E}\big(f(X)|\mathbb{G}\big) \in \left[\mathbb{F}_n(f, X) \pm z_{1-\frac{\alpha}{2}}\frac{\hat{\eta}_n}{\sqrt{|\mathbf{A}_n|}}\right]\right) \le \alpha.$$

The left- and right-hand side in (22) can be compared in terms of the Wasserstein distance d_{W}. For two random elements Y and Y' of $\mathbb{R}$, this is

$$d_{\mathrm{W}}(Y, Y') := \sup_{h\in\mathcal{L}}\big|\mathbb{E}[h(Y)] - \mathbb{E}[h(Y')]\big|,$$

where $\mathcal{L}$ are the Lipschitz functions on $\mathbb{R}$ with Lipschitz constant 1 (e.g., [40]). We denote normalized moments of f by

$$s_p := \mathbb{E}\left[\left|\frac{f(X)}{\eta}\right|^p\right]^{\frac{1}{p}} = \left\|\frac{f(X)}{\eta}\right\|_p \quad \text{for } p > 0.$$

The bound on d_{W} depends both on the value of the integral in (21)(ii), and on the decay of its tail, and we define

$$\tau(b) := \int_{\mathbb{G}\setminus\mathbf{B}_b} \alpha\big(d(e,\phi)|\mathbb{G}\big)^{\frac{\varepsilon}{2+\varepsilon}}|d\phi| \quad \text{for } b \geq 0. \tag{24}$$

Condition (21)(ii) then amounts to $\tau(0) < \infty$. The next result can be read as a generalization of the Berry–Esseen theorem (phrased in terms of d_{W} rather than total variation). It quantifies the speed of convergence in Theorem 4, and the coverage of the confidence interval.

THEOREM 6. *Assume the conditions of Theorem 4, with η defined as in (23), and let Z be a standard normal variable. If (20) holds, and $K \in \mathbb{N}$ is the smallest number for which $\alpha(K|\mathbb{G}) = 0$, then*

$$d_{\mathrm{W}}\bigg(\frac{\sqrt{|\mathbf{A}_n|}}{\eta}\overline{\mathbb{F}}_n(f,X), Z\bigg) \leq \kappa s_2^2 \frac{|\mathbf{A}_n \vartriangle \mathbf{B}_K\mathbf{A}_n|}{|\mathbf{A}_n|} + \kappa\frac{\max(s_4^3,1)|\mathbf{B}_K|^2}{\sqrt{|\mathbf{A}_n|}}$$

for a positive constant κ. If f satisfies (21) for some $\varepsilon > 0$,

$$d_{\mathrm{W}}\bigg(\frac{\sqrt{|\mathbf{A}_n|}}{\eta}\overline{\mathbb{F}}(f,X), Z\bigg) \leq \kappa s_{2+\varepsilon}^2 \frac{|\mathbf{A}_n| - |\mathbf{A}_n \cap \mathbf{B}_{b_n}\mathbf{A}_n|}{|\mathbf{A}_n|} + \kappa\max\big(s_{4+2\varepsilon}^3,1\big)\tau(0)\bigg(\tau(b_n) + \frac{|\mathbf{B}_{b_n}|}{\sqrt{|\mathbf{A}_n|}}\bigg)$$

for a positive constant κ, and any sequence $b_1 < b_2 < \cdots$ of positive scalars.

The choice of (b_n) trades off $|\mathbf{B}_b|$, which increases with b, against $\tau(b)$, which decreases.

EXAMPLE.

(xii) Let X be an i.i.d. sequence, and hence exchangeable and ergodic. For $f \in \mathbf{L}_2(X_1)$, we have $\alpha(1|\mathbb{G}) = 0$, and Theorem 4 is the elementary central limit theorem. Theorem 6 is a Wasserstein metric counterpart to the Berry–Esseen bound (which uses the total variation metric, e.g., [40]): Hypothesis (20) holds, the first term of the bound satisfies $\mathbf{A}_n \vartriangle \mathbf{B}_1\mathbf{A}_n = O(1/n)$, and the second term collapses to $1/\sqrt{n}$.

A less elementary application is a real-valued random field $(X_\phi)_{\phi\in\mathbb{G}}$ that is stationary, that is, invariant under the group $\mathbb{G}$ acting on the index set $\mathbb{G}$. For the groups $\mathbb{Z}^r$ and $\mathbb{R}^r$, for instance, substituting into Theorem 4 yields the following corollary.

COROLLARY 7. *Let $X = (X_\phi)_{\phi\in\mathbb{G}}$ be a stationary random field, and f a real-valued function that satisfies (21). If $\mathbb{G} = (\mathbb{Z}^r, +)$ for some $r \in \mathbb{N}$,*

$$\sqrt{n^r}\bigg(\frac{1}{n^r}\sum_{\mathbf{i}\in\{0,\ldots,n\}^r} f(X_{\mathbf{i}}) - \mathbb{E}\big[f(X)|\mathbb{Z}^r\big]\bigg) \xrightarrow{\mathrm{d}} \eta Z \quad \text{as } n \to \infty$$

for $\eta^2 := \sum_{\mathbf{i}\in\mathbb{Z}^r} \mathbb{E}[f(X_0)f(X_{\mathbf{i}})|\mathbb{Z}^r]$. If $\mathbb{G} = (\mathbb{R}^r, +)$ instead, then

$$\sqrt{n^r}\bigg(\frac{1}{n^r}\int_{[0,n]^r} f(X_t)|dt| - \mathbb{E}\big[f(X)|\mathbb{R}^r\big]\bigg) \xrightarrow{\mathrm{d}} \eta Z \quad \text{as } n \to \infty,$$

where $\eta^2 := \int_{\mathbb{R}^r} \mathbb{E}[f(X_0)f(X_t)|\mathbb{R}^r]|dt|$. In either case, $\eta \perp\!\!\!\perp Z$.

The case $\mathbb{G} = \mathbb{Z}^r$ is Bolthausen's central limit theorem [7]. Thus, Theorem 4 implies a generalization of Bolthausen's theorem to random fields indexed by nice groups, as the second case illustrates. If X satisfies the condition $\vartheta < 1/(2r)$ in Example (x), it is conditionally mixing with respect to each coordinate function, and the corollary holds for all functions $f(X) = g(X_0)$ with $g \in \mathbf{L}_{2+\varepsilon}(X_0)$.

If we quantify the approximation error using Theorem 6, additional properties of the group play a role, and we hence consider a specific class: $(\mathbb{Z}^r, +)$ is a so-called finitely generated nilpotent group of rank r. Such groups are nice, and each contains a finite set called a generator. The minimal number of elements of this set required to transform one group element into another is a metric, the word metric, whose metric balls $\mathbf{B}_n$ satisfy (12) and $1/|\mathbf{B}_n| = O(n^{-r})$. We refer to [32] for details. Substituting into Theorem 6 yields the following.

COROLLARY 8. *Let $\mathbb{G}$ be a finitely generated, nilpotent group of rank $r \in \mathbb{N}$, and set $\mathbf{A}_n := \mathbf{B}_n$ for the word metric of a finite generator. If there exist $\varepsilon, \delta > 0$ such that $\alpha(k|\mathbb{G}) = O(k^{-(r+\delta)})$ and $f(X)/\eta \in \mathbf{L}_{4+2\varepsilon}(X)$, then*

$$d_{\mathrm{W}}\biggl(\frac{\sqrt{|\mathbf{A}_n|}}{\eta}\bigl(\mathbb{F}_n(f, X) - \mathbb{E}[f(X)|\mathbb{G}]\bigr), Z\biggr) = O\bigl(n^{-r\delta/(2(r+\delta))}\bigr) \quad \textit{for } Z \sim N(0,1),$$

where η is defined as in Corollary 7 and independent of Z.

For the case $\mathbb{G} = \mathbb{Z}^r$, the unit coordinate vectors in $\mathbb{Z}^r$ are a finite generator, and the word metric it defines is the metric (15).

5. Generalized limit theorems. This section extends our main theorems to a generalized version of the estimator $\mathbb{F}_n(f, X)$. We begin with an informal overview; proper definitions follow in Section 5.1. The generalized estimator we will define is of the form

$$\frac{1}{\mu_n(\mathbf{A}_n^{k_n})} \int_{\mathbf{A}_n^{k_n}} f_n\bigl(T_n(\boldsymbol{\phi}, X_n)\bigr)\mu_n(d\boldsymbol{\phi}),$$

and again involves a random quantity, now denoted X_n, a real-valued function f_n, and a group action T_n. Additionally, μ_n is a random measure, and $k_n \in \mathbb{N}$. The estimator combines three separate extensions of $\mathbb{F}_n$:

- *Triangular arrays*. We permit the function f_n and the law of X_n to depend on n. That generalizes an invariant random object X in a similar way as triangular arrays generalize i.i.d. sequences (e.g., [28]). Changing X_n with n may involve changing the sample space $\mathbf{X}_n$ and the action T_n. An application example is a nonparametric network model in Section 8.3, which uses $m(n)$ parameters to explain an observed graph of size n. In this case, f_n and $\mathbf{X}_n$ are fixed, but $m(n)$, and hence the distribution of X_n, depends on n.
- *Randomization*. The set $\mathbf{A}_n$ may be randomized, which we formalize as a random measure μ_n on $\mathbf{A}_n$. For example, if $\phi_{n1}, \dots, \phi_{nj_n}$ are sampled with replacement from $\mathbf{A}_n$,

$$\mu_n := j_n^{-1} \sum_{i \le j_n} \delta_{\phi_{ni}} \quad \text{yields the average } j_n^{-1} \sum_{i \le j_n} f(\phi_{ni} X).$$

  More generally, if $\mathbb{G}$ is countable, μ_n may generate subsets (sampling without replacement), multisets (sampling with replacement), or sets of weighted points. In the uncountable case, μ_n may be discrete (which discretizes the integral in $\mathbb{F}_n$ to a sum), or generate uncountable subsets. An illustration is Corollary 16, which subsamples a rotation group.

- *U-statistics*. Consider a function $g : \mathbb{R}^k \to \mathbb{R}$ and a random sequence $(Y_1, Y_2, \dots)$ in $\mathbb{R}$. A U-statistic can be defined in several equivalent ways (see Section 5.5), one of which is

  $$n^{-k} \sum_{i_1,\dots,i_k \le n} g(Y_{i_1}, \dots, Y_{i_k}).$$

  It can be expressed in terms of shifts: If $f : (y_i) \mapsto y_1$ is the first coordinate function and shifts in the set $\mathbf{A}_n := \{0, 1, \dots, n-1\}$ act on the index set of Y by addition, we have

  $$n^{-k} \sum_{i_1,\dots,i_k \le n} g(Y_{i_1}, \dots, Y_{i_k}) = |\mathbf{A}_n|^{-k} \sum_{\phi_1,\dots,\phi_k \in \mathbf{A}_n} g\big(f(\phi_1 Y), \dots, f(\phi_k Y)\big).$$

  If we instead choose g as a function $g : \mathbf{X}_n^{k_n} \to \mathbb{R}$ and replace the set of shifts by a subset $\mathbf{A}_n$ of a general nice group $\mathbb{G}$, we obtain a generalized U-statistic

  $$\frac{1}{|\mathbf{A}_n|^{k_n}} \int_{\mathbf{A}_n^{k_n}} g(\phi_1 X_n, \dots, \phi_k X_n) |d\boldsymbol{\phi}|^{\otimes k}$$

  for tuples $\boldsymbol{\phi} = (\phi_1, \dots, \phi_{k_n})$. To average over tuples, we must choose T_n as an action of $\mathbb{G}^{k_n}$, and we permit the dimension k_n to grow with n. Similarly as elementary U-statistics, these generalized U-statistics are asymptotically normal under suitable conditions (Corollary 12). A variant of this idea is used in the proof of Theorem 17, to approximate permutations by tuples of shifts.

These generalizations can be used in combination with each other, and we hence formulate results simultaneously for all three. Theorem 10 is a central limit theorem, and Theorem 11 gives the speed of convergence. The conditions of Theorems 4 and 6—invariance, a moment condition, and conditional mixing—are still applicable in principle, but since they become rather restrictive in the general case, we introduce the following relaxations:

- Suppose T_n is an action of $\mathbb{G}^{k_n}$. Applying Theorem 4 would require invariance under all tuples $\boldsymbol{\phi} = (\phi_1, \dots, \phi_{k_n})$. The U-statistic above illustrates how strong this assumption is: Even if $(Y_i)_{i \in \mathbb{Z}}$ is stationary, the random field $(g(Y_{i_1}, \dots, Y_{i_{k_n}}))_{i_1,\dots,i_{k_n} \in \mathbb{Z}}$ is not invariant under shifts in $\mathbb{Z}^k$. To obtain a more suitable condition, we observe that the field is invariant under "diagonal" shifts

  $$\big(g(Y_{i_1}, \dots, Y_{i_{k_n}})\big) \quad \mapsto \quad \big(g(Y_{i_1+j}, \dots, Y_{i_{k_n}+j})\big) \quad \text{for } j \in \mathbb{Z}.$$

  More generally, if $(j_1, \dots, j_n)$ is any fixed tuple, applying this tuple as a shift may change the distribution, but the shifted field $(g(Y_{i_1+j_1}, \dots, Y_{i_k+j_k}))$ is again invariant under diagonal shifts. The notion of invariance assumed in this section, defined in (27), generalizes this property from $\mathbb{Z}$ to a general group $\mathbb{G}$.
- Recall that conditional mixing formulates conditions on pairs (ϕ, ϕ') in $\mathbb{G}$ that are far away from a set G. For tuples, this condition becomes stronger as k_n grows—loosely speaking because distances are larger in high dimensions. The marginal mixing condition defined in Section 5.2 measures entry-wise distances, which tend to be smaller.
- The bound on moments is relaxed to uniform integrability, similar to conditions assumed by central limit theorems for triangular arrays.

Randomization requires an additional condition: To guarantee convergence, μ_n must not concentrate on an "unrepresentatively small" part of $\mathbf{A}_n$. Section 5.3 makes that precise.

5.1. *Definitions*. Let $0 < k_1 \leq k_2 \leq \cdots$ be integers. For each $n \in \mathbb{N}$, let X_n be a random element of a standard Borel space $\mathbf{X}_n$, and $f_n : \mathbf{X}_n \to \mathbb{R}$ a measurable function. If $\mathbb{G}$ is a nice group with Haar measure $|\cdot|$, the product space $\mathbb{G}^{k_n}$ is a nice group with Haar measure $|\cdot|^{\otimes k_n}$. Similarly, if $(\mathbf{A}_n)_n$ is a tempered Følner sequence in $\mathbb{G}$, so is $(\mathbf{A}_i^{k_n})_{i\in\mathbb{N}}$ in $\mathbb{G}^{k_n}$. To randomize averages, let μ_n be a random measure on $\mathbb{G}^{k_n}$ that satisfies

$$\text{(i)} \quad \mu_n \text{ is } \sigma\text{-finite,} \qquad \text{(ii)} \quad \mu_n(\mathbf{A}_n^{k_n}) > 0 \quad \text{almost surely.} \tag{25}$$

(Formally, we equip the set of σ-finite measures on $\mathbb{G}^{k_n}$ with the σ-algebra generated by the maps $\mu \mapsto \mu(A)$, for all Borel sets $A \subset \mathbb{G}^{k_n}$. By a random measure, we mean a random element of this space (e.g., [28]).) Let $T_n : \mathbb{G}^{k_n} \times \mathbf{X}_n \to \mathbf{X}_n$ be a measurable action of $\mathbb{G}^{k_n}$, and write

$$\boldsymbol{\phi} x := T_n(\phi_1, \ldots, \phi_{k_n}, x) \quad \text{for } x \in \mathbf{X}_n \text{ and } \boldsymbol{\phi} = (\phi_1, \ldots, \phi_{k_n}) \in \mathbb{G}^{k_n}.$$

The *diagonal action* associated with T_n consists of all transformations

$$(\phi, \ldots, \phi)x = T_n(\phi, \ldots, \phi, x) \quad \text{for } \phi \in \mathbb{G}. \tag{26}$$

The notion of invariance assumed in this section is

$$T_n\big((\phi, \ldots, \phi), T_n(\boldsymbol{\psi}, X_n)\big) \stackrel{\text{d}}{=} T_n(\boldsymbol{\psi}, X_n) \quad \text{for every } \phi \in \mathbb{G} \text{ and } \boldsymbol{\psi} \in \mathbb{G}^{k_n}$$

or equivalently, in more concise notation,

$$(\phi, \ldots, \phi)\boldsymbol{\psi} X_n \stackrel{\text{d}}{=} \boldsymbol{\psi} X_n \quad \text{for every } \phi \in \mathbb{G} \text{ and } \boldsymbol{\psi} \in \mathbb{G}^{k_n}. \tag{27}$$

That is a stronger requirement than diagonal invariance, but weaker than T_n-invariance. To define conditioning, we denote by $\sigma_n(\mathbb{G})$ the σ-algebra

$$\sigma_n(\mathbb{G}) := \big\{A \subset \mathbf{X}_n \text{ Borel} | (\phi, \ldots, \phi)A = A \text{ for all } \phi \in \mathbb{G}\big\},$$

and abbreviate $\mathbb{E}[\cdot|\mathbb{G}] := \mathbb{E}[\cdot|\sigma_n(\mathbb{G})]$ and $P(\cdot|\mathbb{G}) = P(\cdot|\sigma_n(\mathbb{G}))$. We then consider the random, conditionally centered average

$$\widehat{\mathbb{F}}_n(f_n, X_n) := \frac{1}{\mu_n(\mathbf{A}_n^{k_n})} \int_{\mathbf{A}_n^{k_n}} f_n(\boldsymbol{\phi} X_n) - \mathbb{E}\big[f_n(\boldsymbol{\phi} X_n)|\mathbb{G}\big]\, \mu_n(d\boldsymbol{\phi}). \tag{28}$$

If $k_n = 1$, and $\mu_n(\cdot) = |\cdot|$ for all n, and if all X_n and all $\mathbf{X}_n$ are identical, we recover $\sigma_n(\mathbb{G}) = \sigma(\mathbb{G})$ and $\widehat{\mathbb{F}}_n = \overline{\mathbb{F}}_n$.

5.2. *Marginal mixing*. To formulate a suitable mixing condition, we modify the definitions in Section 3: Again consider two elements $\boldsymbol{\phi}$ and $\boldsymbol{\phi}'$ and a subset G, now all in $\mathbb{G}^{k_n}$. We measure how close the entries ϕ_i and ϕ'_k are to the remaining entries of $\boldsymbol{\phi}$ or $\boldsymbol{\phi}'$, or to any entry of vectors in G. To do so, we define the set of "all other" entries,

$$\mathcal{E}_{i,k}(\boldsymbol{\phi}, \boldsymbol{\phi}', G) := \{\phi_j | j \neq i\} \cup \big\{\phi'_j | j \neq k\big\} \cup \{\pi_j | \pi \in G, j \leq k_n\}.$$

In terms of the metric d on $\mathbb{G}$, the shortest distance from ϕ_i or ϕ'_k to any of these is

$$\delta_{i,k}(\boldsymbol{\phi}, \boldsymbol{\phi}', G) := \inf\big\{d\big(\{\phi_i, \phi'_k\}, \psi\big) | \psi \in \mathcal{E}_{i,k}\big\}. \tag{29}$$

For the given function f_n, we then define the set of events

$$\mathcal{C}_{i,k}(t) := \bigcup \sigma_{f_n}(\boldsymbol{\phi}) \otimes \sigma_{f_n}(\boldsymbol{\phi}') \otimes \sigma_{f_n}(G),$$

where the union runs over all pairs $(\boldsymbol{\phi}, \boldsymbol{\phi}')$ and all measurable sets G in $\mathbb{G}^{k_n}$ with $\delta_{i,k}(\boldsymbol{\phi}, \boldsymbol{\phi}', G) \geq t$. Recall that the conditional mixing coefficient was defined in terms of the conditional $P(\cdot|\mathbb{G})$. Using Lindenstrauss' theorem, the latter can be written as

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

To measure the effect of transforming only by coordinates i and k, we substitute this by

$$P_{i,k}(A, A') := \lim_{m\to\infty} \frac{1}{|\mathbf{A}_m|} \int_{\mathbf{A}_m} \mathbb{I}\{e_{i,\psi} X_n \in A, e_{k,\psi} X_n \in A'\} |d\psi|,$$

where $e_{i,\psi} := (e, \dots, e, \psi, e, \dots, e)$ has k_n dimensions and ψ is the ith coordinate. We then define the *marginal mixing coefficient*

$$\alpha_n(t|\mathbb{G}) := \sup_{i\le k_n} \sup_{(A,A',B)\in\mathcal{C}_{i,k}(t)} \bigl|P(A, A', B|\mathbb{G}) - \mathbb{E}[P_{i,k}(A, A')\mathbb{I}\{X_n \in B\}|\mathbb{G}]\bigr|.$$

Choosing (k_n, f_n, X_n) as $(1, f, X)$ for all n recovers $\alpha_n(\cdot|\mathbb{G}) = \alpha(\cdot|\mathbb{G})$.

REMARK. Applying conditional mixing to tuples would measure distance between $(\boldsymbol{\phi}, \boldsymbol{\phi}')$ and G in the product space metric on $\mathbb{G}^{k_n}$. Marginal mixing weakens the condition by replacing this metric by (29). Loosely speaking, since (29) tends to be smaller, the condition $\delta_{i,k}(\boldsymbol{\phi}, \boldsymbol{\phi}', G) \ge t$ then tends to exclude more triples $(\boldsymbol{\phi}, \boldsymbol{\phi}', G)$ in the definition of $\mathcal{C}_{i,k}$ than the metric would, which results in a smaller supremum α_n.

If we specifically consider processes of the form $X_n = f_n(f(\phi_1 X), \dots, f(\phi_{k_n} X))$, the definitions of conditional and marginal mixing can be compared directly. In this case, the intuition in the previous remark can be made precise.

PROPOSITION 9. *Let X be $\mathbb{G}$-invariant, $f \in \mathbf{L}_1(X)$, and set $\mathbf{X}_n = \mathbb{R}^{k_n}$. Then the conditional mixing coefficient of $(f(\phi X))_{\phi\in\mathbb{G}}$ and the marginal mixing coefficient of $(f_n(f(\phi_1 X), \dots, f(\phi_{k_n} X)))_{\boldsymbol{\phi}\in\mathbb{G}^{k_n}}$ satisfy $\alpha_n(\cdot|\mathbb{G}) \le \alpha(\cdot|\mathbb{G})$.*

5.3. *Spreading conditions for randomization.* The random measure μ_n should not concentrate on a subset of $\mathbf{A}_n^{k_n}$ that is "too small". That is formalized as follows: For $A \in \mathcal{B}(\mathbb{G}^{2k_n})$ and any measure ν on $\mathbb{G}^{k_n}$, define

$$\mathbb{T}_n(A, \nu) := \frac{1}{\nu(\mathbf{A}_n^{k_n})^2} \int_{\mathbf{A}_n^{2k_n}} \mathbb{I}\bigl((\boldsymbol{\phi}, \boldsymbol{\psi}) \in A\bigr)\nu(d\boldsymbol{\phi})\nu(d\boldsymbol{\psi}).$$

Consider the random variable

$$\Gamma_n^2(A, \boldsymbol{\phi}) := \frac{1}{\mathbb{T}_n(A, |\cdot|^{\otimes k_n})\mu_n(\mathbf{A}_n^{k_n})} \int_{\mathbf{A}_n^{k_n}} \mathbb{I}\bigl((\boldsymbol{\phi}, \boldsymbol{\psi}) \in A)\bigr)\mu_n(d\boldsymbol{\psi}).$$

Informally, one would expect the integrals

$$\frac{1}{\mu_n(\mathbf{A}_n^{k_n})} \int_{\mathbf{A}_n^{k_n}} \Gamma_n^2(A, \boldsymbol{\phi})\mu_n(d\boldsymbol{\phi}) = \frac{\mathbb{T}_n(A, \mu_n)}{\mathbb{T}_n(A, |\cdot|^{\otimes k_n})}$$

to be bounded if μ_n spreads out its mass sufficiently. As μ_n might be discrete even if the Haar measure is not, bounds should be formulated only in terms of "sufficiently large" sets A. We define the family of such sets as

$$\Sigma_n := \bigl\{A \in \mathcal{B}(\mathbb{G}^{2k_n}) | A \text{ is connected and } |\mathrm{pr}_k(A)| \ge 1 \text{ for all } k \le 2k_n\bigr\},$$

where pr_k denotes projection onto the kth coordinate. A weak notion of boundedness suffices for asymptotic normality: We call the sequence (μ_n) *well-spread* if the variables Γ_n^2 are uniformly integrable for large sets,

$$\sup_n \sup_{A\in\Sigma_n} \biggl\| \frac{1}{\mu_n(\mathbf{A}_n^{k_n})} \int_{\mathbf{A}_n^{k_n}} \Gamma_n^2(A, \phi)\mathbb{I}\bigl(|\Gamma_n^2(A, \phi)| \ge \beta\bigr)\, d\mu_n(\phi) \biggr\|_1 \xrightarrow{\beta\to\infty} 0.$$

A Berry–Esseen bound requires a stricter bound and a fourth-order condition: We similarly define

$$\mathbb{T}_n^*(A,\nu) := \frac{1}{\nu(\mathbf{A}_n^{k_n})^4}\int_{\mathbf{A}_n^{4k_n}} \mathbb{I}\big((\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\boldsymbol{\phi}_3,\boldsymbol{\phi}_4)\in A\big)\nu^{\otimes 4}(d\boldsymbol{\phi}_1,d\boldsymbol{\phi}_2,d\boldsymbol{\phi}_3,d\boldsymbol{\phi}_4),$$

now for subset A of and a measure ν on $\mathbb{G}^{4k_n}$, and

$$\Sigma_n^* := \{A\in\mathcal{B}(\mathbb{G}^{4k_n}) \mid A \text{ is connected and } |\mathrm{pr}_k(A)|\geq 1 \text{ for all } k\leq 4k_n\}.$$

We call (μ_n) *strongly well-spread* if

$$\mathcal{S} := \sup_n \mathcal{S}^n < \infty \quad \text{where } \mathcal{S}^n := \sup_{A\in\Sigma_n^*}\left\|\frac{\mathbb{T}_n^*(A,\mu_n)}{\mathbb{T}_n^*(A,|\cdot|^{\otimes k_n})}\right\|_1,$$

with *spreading coefficient* $\mathcal{S}$. Since the existence of higher moments implies uniform integrability, strongly well-spread implies well-spread. Either condition can be applied to a single random measure μ, by setting $\mu_n := \mu$ for all n.

EXAMPLES.

(xiii) Let Π be a Poisson point process on $\mathbb{G}^k$, for some $k\in\mathbb{N}$. Then the random measure $\mu(\cdot) := |\Pi\cap\cdot|^{\otimes k}$ is strongly well-spread if

$$\sup_{A\in\mathcal{B}(\mathbb{G}^k),|A|^{\otimes k}<\infty} \frac{\mathbb{E}[|\Pi\cap A|^{\otimes k}]}{|A|^{\otimes k}} < \infty.$$

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

$$\Pi_n\cap\mathbf{A}_n^{k_n} \mid \big(|\Pi_n\cap\mathbf{A}_n^{k_n}| = m\big) \stackrel{\mathrm{d}}{=} (\Phi_1,\ldots,\Phi_m) \quad \text{for all } m\in\mathbb{N},$$

where the Φ_i are drawn uniformly with or without replacement from $\mathbf{A}_n^{k_n}$. The sequence defined by $\mu_n(\cdot) := |\Pi_n\cap\cdot|^{\otimes k_n}$ is strongly well-spread.

5.4. *Results*. If the dimension k_n grows with n, we must quantify how much f_n changes with n: For $p>0$ and $i\leq k_n$, define

$$c_{i,p}(f_n) := \sup_{\psi\in\mathbb{G},\boldsymbol{\phi}\in\mathbb{G}^{k_n}} \frac{1}{2}\big\|f_n\circ\boldsymbol{\phi} - f_n\circ(e,\ldots,e,\psi,e,\ldots,e)\boldsymbol{\phi}\big\|_p,$$

where ψ is the ith coordinate. Hypotheses (20) and (21) are then replaced by one of the following conditions: Either

$$\text{(i)}\quad \sup_n \alpha_n(K|\mathbb{G}) = 0, \qquad \text{(ii)}\quad \sup_n \sum_{i\leq k_n} c_{i,2}(f_n) < \infty,$$
$$\text{(iii)}\quad \big(f_n(\boldsymbol{\phi}X_n)^2\big)_{n\in\mathbb{N},\phi\in\mathbb{G}^{k_n}} \quad \text{is uniformly integrable} \tag{30}$$

holds for some $K\in\mathbb{N}$, or

$$\text{(i)}\quad \sup_n \int_{\mathbb{G}} \alpha_n\big(d(e,\phi)|\mathbb{G}\big)^{\frac{\epsilon}{2+\epsilon}}|d\phi| < \infty, \qquad \text{(ii)}\quad \sup_n \sum_{i\leq k_n} c_{i,2+\varepsilon}(f_n) < \infty,$$
$$\text{(iii)}\quad \big(f_n(\boldsymbol{\phi}X_n)^{2+\varepsilon}\big)_{n\in\mathbb{N},\phi\in\mathbb{G}^{k_n}} \quad \text{is uniformly integrable} \tag{31}$$

holds for some $\varepsilon>0$. In either case, (iii) implies (ii) if the sequence (k_n) is bounded. To assemble the asymptotic variance, set

$$\widehat{\mathbb{F}}_{\infty,i}(\psi) := \lim_{m\to\infty}\frac{1}{|\mathbf{A}_m|^{k_n}}\int_{\boldsymbol{\phi}\in\mathbf{A}_m^{k_n}} f_n\big((\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_{i-1},\psi,\boldsymbol{\phi}_{i+1},\ldots,\boldsymbol{\phi}_{k_n})X_n\big)|d\boldsymbol{\phi}|^{\otimes k_n-1}.$$

Let μ_n^i be the ith coordinate marginal of μ_n, scaled to $\mu_n^i(\mathbf{A}_n) = \sqrt{|\mathbf{A}_n|}$,

$$\mu_n^i(\cdot) := \frac{\sqrt{|\mathbf{A}_n|}}{\mu_n(\mathbf{A}_n^{k_n})}\mu_n(\mathbf{A}_n, \dots, \mathbf{A}_n, \cdot, \mathbf{A}_n, \dots, \mathbf{A}_n),$$

and set

$$\widehat{\eta}_{nm} := \sum_{i,j \le k_n} \iint_{\phi\mathbf{A}_n, \psi \in \mathbf{B}_m(\phi)} \mathbb{E}[\widehat{\mathbb{F}}_{\infty,i}(e)\widehat{\mathbb{F}}_{\infty,j}(\phi^{-1}\psi)|\mathbb{G}]\mu_n^i(d\phi)\mu_n^j(d\psi).$$

The central limit theorem then takes the following form.

THEOREM 10. *Let (X_n) be invariant in the sense of (27) for each n, and let (μ_n) be well-spread and independent of (X_n). Assume either condition (30) or (31) holds. If $k_n = o(|\mathbf{A}_n|^{\frac{1}{4}})$, and if the limits*

$$\widehat{\eta}_{nm} \xrightarrow{p} \eta_m \quad \text{as } n \to \infty \quad \text{and} \quad \eta_m \xrightarrow{\mathbf{L}_2} \eta \quad \text{as } m \to \infty \tag{32}$$

exist, then

$$\sqrt{|\mathbf{A}_n|}\widehat{\mathbb{F}}_n(f_n, X_n) \xrightarrow{d} \eta Z \quad \text{as } n \to \infty,$$

for a standard normal variable Z that is independent of η.

The Wasserstein Berry–Esseen bound in Theorem 6 generalizes similarly the following.

THEOREM 11. *Assume the conditions of Theorem 10 hold, require that (μ_n) is strongly well-spread, and define η as in (32). If condition (30) holds for some $K \in \mathbb{N}$,*

$$d_{\mathrm{W}}\left(\frac{\sqrt{|\mathbf{A}_n|}}{\eta}\widehat{\mathbb{F}}_n(f_n, X_n), Z\right) \le \kappa \frac{k_n^2(\mathcal{S}^n \wedge 1)((\sum_i c_{i,4})^3 \wedge 1)|\mathbf{B}_K|^2}{\sqrt{|\mathbf{A}_n|}} + \left\|\frac{\widehat{\eta}_{n,K}^2 - \eta^2}{\eta^2}\right\|,$$

for a positive constant κ. If (31) holds instead, set

$$\mathcal{R}_n(b) := \sum_{t \ge b} |\mathbf{B}_{t+1} \setminus \mathbf{B}_t|\alpha_n(t|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}} \quad \text{for } b \in \mathbb{N},$$

and fix any sequence $0 < b_1 < b_2 < \cdots$ of integers. Then

$$\begin{aligned} d_{\mathrm{W}}\left(\frac{\sqrt{|\mathbf{A}_n|}}{\eta}\widehat{\mathbb{F}}_n(f_n, X_n), Z\right) &\le \kappa \mathcal{R}_n(b_n)\left(\sum_i c_{i,2+\epsilon}\right)^2 (\mathcal{S}^n \wedge 1) + \left\|\frac{\widehat{\eta}_{n,b_n}^2 - \eta^2}{\eta^2}\right\| \\ &\quad + \kappa\left(\left(\sum_i c_{i,4+2\epsilon}\right)^3 \wedge 1\right)(\mathcal{S}^n \wedge 1)\mathcal{R}_n(0)\frac{k_n^2|\mathbf{B}_{b_n}|}{\sqrt{|\mathbf{A}_n|}} \end{aligned}$$

for a positive constant κ.

If we choose $(k_n, X_n, \widehat{\mathbb{F}}_n)$ as $(1, X, \mathbb{F}_n)$ for all n, the conditions specialize to (20) and (21), and the results to Theorems 4 and 6.

5.5. *Generalized U-statistics.* The generalized notion of invariance defined in (27) allows us to formulate a useful generalization of U-statistics, denoted $X_{\boldsymbol{\psi}}$ in the next result. Substituting these into Theorem 10 shows they are asymptotically normal.

COROLLARY 12. *Consider a $\mathbb{G}$-invariant random element Y of $\mathbf{X}$, a function $h : \mathbf{X}^k \to \mathbb{R}$, and define $X_{\boldsymbol{\psi}} := h(\psi_1 Y, \ldots, \psi_k Y)$ for $\boldsymbol{\psi} \in \mathbb{G}^k$. Suppose there is an $\varepsilon > 0$ for which the conditional mixing coefficient of Y satisfies $\int_{\mathbb{G}} \alpha^{\frac{\varepsilon}{2+\varepsilon}}(d(e,\phi)|\mathbb{G})|d\phi| < \infty$, and $(X_{\boldsymbol{\psi}}^{2+\varepsilon})_{\boldsymbol{\psi}\in\mathbb{G}^k}$ is uniformly integrable. Then*

$$|\mathbf{A}_n|^{\frac{1}{2}-k} \int_{\mathbf{A}_n^k} (X_{\boldsymbol{\psi}} - \mathbb{E}[X_{\boldsymbol{\psi}}|\mathbb{G}^k])|d\boldsymbol{\psi}|^{\otimes k} \xrightarrow{\mathrm{d}} \eta Z$$

for $\eta \perp\!\!\!\perp Z$ and $Z \sim N(0,1)$. If we denote

$$H_i(\phi) := \lim_{m\to\infty} \frac{1}{|\mathbf{A}_m|^{k-1}} \int_{\mathbf{A}_m^{k-1}} X_{\psi_1,\ldots,\psi_{i-1},\phi,\psi_{i+1},\ldots,\psi_{k_n}} |d\psi_1|\cdots|d\psi_{i-1}||d\psi_{i+1}|\cdots|d\psi_k|,$$

the asymptotic variance is $\eta^2 = \sum_{i,j\le k}\int_{\mathbb{G}} \operatorname{Cov}[H_i(e), H_j(\phi)|\mathbb{G}]|d\phi|$.

To clarify the relationship to U-statistics, recall that a U-statistic for an i.i.d. sequence $(Y_i)_{i\in\mathbb{Z}}$ is usually defined in one of two ways, namely

$$U_n := \binom{n}{k}^{-1} \sum_{\phi\in\mathbb{S}_n} h(Y_{\phi(1)},\ldots,Y_{\phi(k)}) \quad \text{or} \quad V_n := \frac{1}{n^k}\sum_{i_1,\ldots,i_k\le n} h(Y_{i_1},\ldots,Y_{i_k}).$$

The definitions are equivalent, in the sense that $\sqrt{n}(U_n - V_n) \to 0$ in probability [41]. The corollary shows $n^{-1/2}(V_n - \mathbb{E}[V_n]) \xrightarrow{\mathrm{d}} \eta Z$, if we choose $\mathbb{G}$ as $\mathbb{Z}$ and $\mathbf{A}_n$ as $\{1,\ldots,n\}$. Although $h(Y_{i_1},\ldots,Y_{i_k})$ satisfies the relaxed invariance (27), it is not $\mathbb{Z}^k$-invariant, since arbitrary shifts may break independence of (Y_i) by duplicating indices.

6. Concentration. The theorems above show that certain asymptotic properties of i.i.d. processes generalize to symmetric random objects. We show next that certain finite-sample properties generalize similarly. We use the definitions of Section 5, but somewhat restrict the spaces and functions involved: Fix two Borel spaces $\mathbf{X}$ and $\mathbf{Y}$, two sequences (f_n) and (g_n) of measurable functions $f_n : \mathbf{X} \to \mathbf{Y}$ and $g_n : \mathbf{Y}^{k_n} \to \mathbf{X}$, and let (X_n) be a sequence of $\mathbb{G}$-invariant random elements of $\mathbf{X}$. We consider concentration for quantities of the form $g_n(f_n(\phi_1 X_n),\ldots,f_n(\phi_{k_n}X_n))$. To this end, define

$$Y^n := (Y_\phi^n)_{\phi\in\mathbb{G}} \quad \text{where } Y_\phi^n := f_n(\phi X_n).$$

That implies $(Y_\phi^n) \overset{\mathrm{d}}{=} (Y_{\psi\phi}^n)$ for $\psi \in \mathbb{G}$. We again work with (conditionally) centered averages: For $\boldsymbol{\phi} = (\phi_1,\ldots,\phi_{k_n})$, set

$$h_n(\boldsymbol{\phi}X_n) := g_n(Y_{\phi_1}^n,\ldots,Y_{\phi_{k_n}}^n) - \mathbb{E}[g_n(Y_{\phi_1}^n,\ldots,Y_{\phi_{k_n}}^n)|\mathbb{G}] \quad \text{for } \boldsymbol{\phi}\in\mathbb{G}^{k_n}.$$

The average $\widehat{\mathbb{F}}_n$, as defined in the previous section, is then

$$\widehat{\mathbb{F}}_n(h_n, X_n) = \frac{1}{\mu_n(\mathbf{A}_n^{k_n})}\int_{\mathbf{A}_n^{k_n}} h_n(\boldsymbol{\phi}X_n)\mu_n(d\boldsymbol{\phi}).$$

EXAMPLE.

(xv) Let $X := (X_i)_{i\in\mathbb{Z}}$ be a stationary, real-valued process, so $\mathbf{X} = \mathbb{R}^{\mathbb{Z}}$. Choose all f_n as the coordinate function $(x_i) \mapsto x_0$ at index 0. If (g_n) is any sequence of measurable functions $g_n : \mathbb{R}^k \to \mathbb{R}$, we obtain random fields $Y^n = (g_n(X_{i_1},\ldots,X_{i_k}))_{i_1,\ldots,i_k\in\mathbb{Z}}$, and

$$\widehat{\mathbb{F}}_n(h_n, X_n) = \frac{1}{n^k}\sum_{i_1,\ldots,i_k\le n}\big(g_n(X_{i_1},\ldots,X_{i_k}) - \mathbb{E}(g_n(X_{i_1},\ldots,X_{i_k}))\big).$$

As we had already observed in the introduction of Section 5, each Y^n is invariant under the diagonal action $Y^n \mapsto (g_n(X_{i_1+\phi},\ldots,X_{i_k+\phi}))$, for $\phi\in\mathbb{Z}$, since X is stationary.

A function $f : \mathbf{Y}^k \to \mathbb{R}$ is *self-bounded* if there are constants $\delta_1, \ldots, \delta_k$, the *self-bounding coefficients*, such that

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

see, for example, [11]. We call f *uniformly* $\mathbf{L}_1$*-continuous* in $\mathbb{G}$ if

$$\sup_{\substack{\boldsymbol{\phi},\boldsymbol{\psi}\in\mathbb{G}^k \\ d(\phi_i,\psi_i)\leq\epsilon \text{ for } i\leq k}} \big\|f(\boldsymbol{\phi}X_n) - f(\boldsymbol{\psi}X_n)\big\|_1 \longrightarrow 0 \quad \text{as } \epsilon \to 0.$$

We measure interactions within a process $Y = (Y_\phi)_{\phi\in\mathbb{G}}$ as follows: Write $\mathcal{L}$ for the law of a random variable, $\|\cdot\|_{\mathrm{TV}}$ for the total variation norm, and abbreviate $Y_{\neq\phi} := (Y_\psi)_{\psi\neq\phi}$. If $\mathbb{G}$ is countable, define

$$\Lambda[Y] := \sum_{\phi\in\mathbb{G}\setminus\{e\}} \sup_{\substack{\mathbf{x},\mathbf{y}\in\mathbf{Y}^{\mathbb{G}} \\ \mathbf{x}_{\neq\phi}=\mathbf{y}_{\neq\phi}}} \big\|\mathcal{L}(Y_e|Y_{\neq e} = \mathbf{x}_{\neq e}) - \mathcal{L}(Y_e|Y_{\neq e} = \mathbf{y}_{\neq e})\big\|_{\mathrm{TV}}.$$

If $\mathbb{G}$ is uncountable, we discretize: For $\epsilon > 0$, a set $C \subset \mathbb{G}$ is an ϵ*-net* if

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

A decreasing sequence of nets is a sequence $(C_i)_{i\in\mathbb{N}}$, where C_i is an ϵ_i-net and $\epsilon_i \to 0$. Define

$$\rho[Y] := \sup\bigg(1 - \lim_{i\to\infty} \frac{1 - \Lambda[(Y_\phi)_{\phi\in C_i}]}{|\mathbf{B}_{\epsilon_i}|}\bigg),$$

where the supremum is taken over all decreasing sequences of nets for which the limit on the right exists. Discretizing continuous processes on nets is a standard tool in the context of concentration inequalities (see [11], Chapter 13). Note that $\rho = \Lambda$ if the group is discrete. For $\mathbb{G} = \mathbb{Z}$, it is known as the Dobrushin interdependence coefficient [43]. A continuous example is a Markov process $Y = (Y_t)_{t\in\mathbb{R}}$ on $\mathbb{G} = \mathbb{R}$, where

$$\rho[Y] = \lim_{t\to\infty} \frac{1}{t} \sup_{x,y\in\mathbb{R}} \big\|\mathcal{L}(Y_0|Y_t = x) - \mathcal{L}(Y_0|Y_t = y)\big\|_{\mathrm{TV}}.$$

THEOREM 13. *Let $(\mathbf{A}_n)$ be a tempered Følner sequence in $\mathbb{G}$, let (c_i) be the self-bounding coefficients of h_n, and require that (h_n) is uniformly $\mathbf{L}_1$-continuous in $\mathbb{G}$. Define*

$$\tau_n := \sup_{j\leq k_n} \sup_{B\in\mathcal{B}(\mathbb{G})} \frac{|\mathbf{A}_n|\mu_n(\mathbf{A}_n^{j-1} \times (B \cap \mathbf{A}_n) \times \mathbf{A}_n^{k_n-j}|\mathbf{A}_n^{k_n})}{|B \cap \mathbf{A}_n|}.$$

Then

$$\mathbb{P}\big(\widehat{\mathbb{F}}_n(h_n, X_n) \geq t\big) \leq 2\mathbb{E}\bigg(\exp\bigg(-\frac{(1-\rho[Y^n])|\mathbf{A}_n|}{(\sum_{i\leq k_n} c_i)^2\tau_n^2}t^2\bigg)\bigg) \quad \textit{for all } t > 0.$$

The coefficients τ_n are only required if averages are randomized. If (μ_n) is nonrandom, the statement can simplify considerably. For example, we have the following corollary.

COROLLARY 14. *If $\mu_n = |\cdot|^{\otimes k_n}$ almost surely for each n, then*

$$\mathbb{P}\big(\widehat{\mathbb{F}}_n(h_n, X_n) \geq t\big) \leq 2\exp\bigg(-\frac{(1-\rho_n)|\mathbf{A}_n|}{(\sum_{i\leq k_n} c_i)^2}t^2\bigg) \quad \textit{for } t > 0 \textit{ and } n \in \mathbb{N}.$$

EXAMPLE.

(xvi) For illustration, compare to the i.i.d. case: Choose X, f and g as in Example (xv) at the beginning of this section, and assume additionally that the sequence X is i.i.d. If (c_i) are the self-bounding coefficients of g, the corollary shows

$$\mathbb{P}(\widehat{\mathbb{F}}_n(h, X) \geq t) \leq 2\exp\left(-\frac{nt^2}{(\sum_{i \leq k} c_i)^2}\right) \quad \text{for } t > 0 \text{ and } n \in \mathbb{N},$$

which is a version of McDiarmid's inequality.

Theorem 13 implicitly assumes fairly strong mixing: If $\mathbb{G}$ is discrete, for example, then $\alpha(n|\mathbb{G}) \leq c_1\alpha(n) \leq c_2\Lambda[X]$ for some positive constants c_1 and c_2 and all $n \in \mathbb{N}$. The mixing condition is hence no weaker than that required in the asymptotic case, and conditioning on $X_{\neq e}$ in the definition of $\Lambda[X]$ means it is typically stronger.

7. Approximation by subsets of transformations. According to Theorem 10, $\mathbb{F}_n$ may be computed using only a subset of $\mathbf{A}_n$. We briefly discuss a few cases in more detail. First suppose we "factor out" a compact subgroup $\mathbb{K}$ of $\mathbb{G}$ to obtain a subgroup $\mathbb{H}$, and then compute $\mathbb{F}_n$ using a Følner sequence of $\mathbb{H}$. For exchangeable sequences, factoring $\mathbb{S}_k$ out of $\mathbb{S}_\infty$ amounts to including only every kth observation in the sample average, so rates slow by a constant. The general behavior is similar.

PROPOSITION 15. *Let $\mathbb{G}$ be generated by the union of a noncompact group $\mathbb{H}$ and a compact group $\mathbb{K}$, and let $(\mathbf{A}_n^{\mathbb{H}})$ be a Følner sequence in $\mathbb{H}$. Then $\mathbf{A}_n := \mathbf{A}_n^{\mathbb{H}}\mathbb{K}$ is a Følner sequence in $\mathbb{G}$. If X is $\mathbb{G}$-invariant, and $f \in \mathbf{L}_2(X)$ satisfies (21) with respect to $\mathbb{G}$, there exist random variables $\eta, \eta_{\mathbb{H}} \in \mathbf{L}_2(X)$ and an independent variable $Z \sim N(0, 1)$ such that*

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

and

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

where η is defined by (32). The ratio $\beta := \sqrt{|\mathbb{K}|}\frac{\eta_{\mathbb{H}}}{\eta}$, and hence $\eta_{\mathbb{H}}$, 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, let $X = (X_t)_{t \in \mathbb{R}^r}$ be a continuous random field that is both shift- and rotation invariant. Thus, $\mathbb{G} = \mathbb{R}^r \times \mathbb{O}_r$, where $\mathbb{O}_r$ is the (compact) orthogonal group of order r. Factoring out $\mathbb{O}_r$ means we average only over shifts. Convergence then slows by a factor

$$\beta^2 - 1 = \frac{1}{\eta^2}\mathbb{E}\left[f(X) \int_{\mathbb{R}^r} \int_{\mathbb{O}_r} (f(X + \phi) - f(\theta X + \phi))|d\theta||d\phi|\right]. \tag{33}$$

One might also discretize $\mathbf{A}_n$ (e.g., to avoid integration), or subsample it. For example: A tempered Følner sequence in $\mathbb{R}^r \times \mathbb{O}_r$ is given by $([-n, n]^r \times \mathbb{O}_r)_n$ [32]. If we discretize $[-n, n]^r$ deterministically, and $\mathbb{O}_r$ at random, we obtain the following corollary.

COROLLARY 16. *Let $X = (X_t)_{t\in\mathbb{R}^r}$ be a random field invariant under rotations and translations of $\mathbb{R}^r$, and require (21). Fix $m \in \mathbb{N}$. For $z \in \mathbb{Z}^r$, let $\Theta_1^z, \ldots, \Theta_m^z$ be independent, uniform random elements of $\mathbb{O}_r$. Then*

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

as $n\to\infty$, for an almost surely finite random variable $\eta_m \perp\!\!\!\perp Z$. Relative to $\mathbb{F}_n$ defined by integration over the entire set $[-n,n]^r \times \mathbb{O}_r$, convergence slows by a coefficient $\beta_m^2 - 1 = (\beta^2-1)/(2m^2\eta_m^2)$, where β is given by (33).

If the random rotations Θ_j^z are not independent—for example, if one generates m rotations once and uses them repeatedly—the rate may slow.

8. Applications I: Exchangeable structures.

One of the most common distributional symmetries is permutation invariance, often referred to as exchangeability. It can broadly be categorized into three types: *Finite exchangeability* is invariance under $\mathbb{S}_n$, for some fixed $n \in \mathbb{N}$ [29]. This is an example of invariance under a compact group, and has no asymptotic theory. Countably infinite exchangeability, or henceforth simply *exchangeability*, is invariance under $\mathbb{S}_\infty$. This type is common in statistics and probability. By *uncountable exchangeability*, we refer to invariance under permutation groups of uncountable sets. Such groups are not nice, and Lindenstrauss' theorem is not applicable, but Section 8.5 gives an example where reduction to our results is possible.

8.1. *Exchangeability.*

The next theorem adapts our results to exchangeable structures, including the examples in Table 2. In this case, the mixing condition can be eliminated.

THEOREM 17. *Let X be a random element of a standard Borel space $\mathbf{X}$, and invariant under a measurable action of $\mathbb{S}_\infty$. Let f be a function satisfying $\mathbb{E}[f(X)^2] < \infty$ and*

$$\sum_{i\in\mathbb{N}} \limsup_j \|f(X) - f(\tau_{ij}X)\|_2 < \infty, \tag{34}$$

where τ_{ij} denotes the transposition of i and j. As $n\to\infty$,

$$\sqrt{n}\mathbb{F}_n(f,X) = \sqrt{n}\bigg(\frac{1}{n!}\sum_{\phi\in\mathbb{S}_n} f(\phi X) - \mathbb{E}\big[f(X)|\mathbb{S}_\infty\big]\bigg) \xrightarrow{d} \eta Z, \tag{35}$$

where $Z \sim N(0,1)$ is independent of η. Define

$$\mathbb{F}^i(\phi) := \lim_{n\to\infty} \frac{1}{|\mathbb{S}_n^i|} \sum_{\phi'\in\mathbb{S}_n^i} f\big(\phi'\phi X\big) \quad \textit{where } \mathbb{S}_n^i := \big\{\phi \in \mathbb{S}_n | \phi(i) = i\big\}.$$

The asymptotic variance satisfies

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

If in addition $\mathbb{E}[f(X)^4/\eta^4] < \infty$ and $\sum_{i\in\mathbb{N}} \limsup_j \|\frac{f(X)-f(\tau_{ij}X)}{\eta}\|_4 < \infty$, the Wasserstein distance to the limit is

$$d_{\mathrm{W}}\bigg(\frac{\sqrt{n}}{\eta}\mathbb{F}_n(f,X), Z\bigg) = O\bigg(\min_{k\in\mathbb{N}}\bigg[\frac{k^2}{\sqrt{n}} + \sum_{i>k} \max\bigg(\limsup_j \bigg\|\frac{f(X)-f(\tau_{ij}X)}{\eta}\bigg\|_4, 1\bigg)\bigg]\bigg).$$

TABLE 2
Examples of exchangeable random structures

Random structure X	Ergodic structures	CLT (35) due to
exchangeable sequence [29]	i.i.d. sequences	H. Bühlmann [13]
exchangeable partition [39]	"paint-box" distributions	
exchangeable graph [17]	graphon distributions	Bickel et al. [5], Ambroise and Matias [2]
jointly exch. array [29]	dissociated arrays	Eagleson and Weber [19], Davezies et al. [16]
separately exch. array [29]	dissociated arrays	

Typically, X is of the form $(X_t)_{t\in T}$ for some countable set T, and permutations act on X by acting on T. If f depends only on a finite number of these indices—for example, if X is a random matrix and f a function of a finite number of entries—(34) always holds, although this condition is far from necessary. If X is conditionally mixing for f, the result can be deduced from Theorem 4. The proof of the general case defines surrogate variables $X_n := (f(\tau_{1,i_1}\circ\cdots\circ\tau_{k_n n,i_{k_n}}X))_{i_1,\ldots,i_{k_n}}$ for a suitable sequence (k_n), and applies an idea similar to the generalized U-statistics of Corollary 12.

REMARK. (a) Our definition of exchangeability as an arbitrary action of $\mathbb{S}_\infty$ permits trivial cases, for example: Mapping each $\phi\in\mathbb{S}_\infty$ to the identity map of $\mathbf{X}$ is a valid action. It makes all distributions exchangeable, point masses are ergodic, and $\mathbb{F}_n(f,X)=\mathbb{E}[f(X)|\mathbb{G}]=f(X)$ for all n. (b) Exchangeability can also be defined as invariance under the group $\mathbb{S}(\mathbb{N})$ of all bijections of $\mathbb{N}$, as is often done in Bayesian statistics. This definition is equivalent to ours, in the sense that any measurable action of $\mathbb{S}(\mathbb{N})$ and its restriction to $\mathbb{S}_\infty\subset\mathbb{S}(\mathbb{N})$ have the same invariant and ergodic measures [34], but the group $\mathbb{S}(\mathbb{N})$ is not nice.

8.2. *Jointly exchangeable arrays.* We discuss one class of examples in Table 2, the jointly exchangeable arrays, in more detail. These are defined as follows: A collection $x=(x_{i_1,\ldots,i_r})_{i_1,\ldots,i_r\in N}$ of scalars is called an r-array indexed by $N\subseteq\mathbb{N}$. The subarray indexed by $M\subset N$ is denoted $x[M]$. We let permutations ϕ of N act on x by permuting each index dimension separately, $\phi(x):=(x_{\phi(i_1),\ldots,\phi(i_r)})$. A *jointly exchangeable array* is a random array X that is indexed by $N=\mathbb{N}$ and satisfies $\phi(X)\stackrel{\mathrm{d}}{=}X$ for all $\phi\in\mathbb{S}_\infty$.

The ergodic exchangeable arrays are characterized explicitly, by the Aldous–Hoover theorem [29]: To keep notation simple, assume $r=2$. Then X is $\mathbb{S}_\infty$-ergodic if and only if there is a measurable function $h:[0,1]^3\to\mathbb{R}$ such that

$$X\stackrel{\mathrm{d}}{=}\big(h(U_i,U_j,U_{ij})\big)_{i,j\in\mathbb{N}}\quad\text{where } (U_i,U_{ij})_{i,j\in\mathbb{N}}\sim_{\text{iid}}\text{Uniform}[0,1]. \tag{36}$$

Thus, X is $\mathbb{S}_\infty$-ergodic if h is fixed, and $\mathbb{S}_\infty$-invariant if h is random. For $r>2$, the function h has additional arguments [29]. Kallenberg [27] first proved the relevant case of Lindenstrauss' theorem: If X is $\mathbb{S}_\infty$-ergodic,

$$\frac{1}{n!}\sum_{\phi\in\mathbb{S}_n} f\big((X_{\phi(i_1),\ldots,\phi(i_r)})_{i_1,\ldots,i_r}\big)\xrightarrow{n\to\infty}\mathbb{E}\big[f(X)\big]\quad\text{a.s. for } f\in\mathbf{L}_1(X).$$

Eagleson and Weber [19] proved an early version of (35) for such averages (under stronger conditions than Theorem 17). Under suitable additional assumptions, one can obtain a uniform result [16].

An *exchangeable graph* is an exchangeable 2-array with binary entries and almost surely zero diagonal [17]. We interpret the array as the adjacency matrix of a random graph with vertex set $\mathbb{N}$. Since that makes the range of h binary, one can eliminate one degree of freedom in the representation above: An exchangeable graph is ergodic if and only if (36) holds for a measurable function $w:[0,1]^2 \to [0,1]$ and $h(u,v,z) := \mathbb{I}\{z \leq w(u,v)\}$. For undirected graphs, w can be chosen to satisfy $w(u,v) = w(v,u)$, and is called a *graphon* [10].

For a finite graph y with vertex set $\{1,\dots,k\}$ and the subgraph $X[1,\dots,k]$ of X on the same set, consider the subgraph probability $t(y) := P(X[1,\dots,k] = y)$. Some authors interpret $t(y)$ as a moment statistic [5]. For $n \geq k$ and a graph x with vertex set $\{1,\dots,n\}$, the *homomorphism density* $t(x,y) := 1/n! \sum_{\phi \in \mathbb{S}_n} \mathbb{I}\{x[\phi(1),\dots,\phi(k)] = y\}$ is the (normalized) number of times y occurs as a subgraph of x [10, 33]. If X is ergodic, and a finite subgraph $X[1,\dots,n]$ is observed as data, substituting into Kallenberg's result above shows

$$t\big(X[1,\dots,n],\cdot\big) \xrightarrow{n\to\infty} P\big(X[1,\dots,k] = \cdot\big) = t(\cdot) \quad \text{almost surely.} \tag{37}$$

In other words, the sample homomorphism density $t(X[1,\dots,n],y)$ is a strongly consistent estimator of $t(y)$. Borgs et al. [10] and Lovász and Szegedy [33] have also obtained (37), using different arguments. For these estimators, (35) is due to Bickel et al. [5] and Ambroise and Matias [2].

8.3. *Stochastic block models with a growing number of classes.* Suppose we choose h in (36) as follows: Fix some $m \in \mathbb{N}$. Choose a measurable function $\pi:[0,1] \to \{1,\dots,m\}$ and a symmetric function $v:\{1,\dots,m\}^2 \to [0,1]$. For each $i \leq m$, set $\pi_i := \mathbb{P}(\pi(U) = i)$, where U is uniform in $[0,1]$. We can read $(\pi_i)_{i\leq m}$ as a distribution on m categories, and v as a matrix $(v(i,j))_{i,j\leq m}$. Define a random undirected graph with vertex set $\mathbb{N}$ as

$$X(\pi,v) := \big(\mathbb{I}\big\{U_{ij} < v\big(\pi(U_i),\pi(U_j)\big)\big\}\big)_{i<j\in\mathbb{N}}.$$

Since this is a special case of (36), $X(\pi,v)$ is an ergodic exchangeable graph, represented by the piece-wise constant graphon $w = v \circ (\pi \otimes \pi)$. A family of such distributions, indexed by some range of pairs (π,v), is a *stochastic block model* with m classes (e.g., [2]). Since each law is specified by a finite vector (π_i) and matrix v, the model is parametric. Nonparametric extensions let m grow with sample size (e.g., [15]): Choose an increasing function $m:\mathbb{N}\to\mathbb{N}$ and a parameter sequence $(\pi^n,v^n)_{n\in\mathbb{N}}$ such that $X_n := X(\pi^n,v^n)$ has $m(n)$ classes. An observed graph on n vertices is then explained as the finite subgraph $X_n[1,\dots,n]$.

In the nonparametric case, no asymptotic normality results seem to be known, but can easily be obtained from our results. Since X_n changes with sample size, Theorem 17 is not applicable, but Theorems 10 and 11 can be used instead. As a concrete example, let y be the complete graph on three vertices. In this case, $P(X_n[1,2,3] = y)$ is often called the triangle density. If vertex 1 is in class i, but the classes of 2 and 3 are unknown, the probability that $X_n[1,2,3]$ is a triangle is

$$E_i(n) := \mathbb{E}\big[f(X_n)|\pi^n\big(U_1^n\big) = i\big] = \sum_{j\leq m(n)} \pi_j^n\bigg(v^n(i,j)\sum_{k\leq m(n)} \pi_k^n v^n(i,k)v^n(j,k)\bigg).$$

Applying Theorems 10 and 11 to $f(x) := \mathbb{I}\{x[1,2,3] = y\}$ yields the following.

COROLLARY 18. *Let Z be a standard normal variable. As $n \to \infty$,*

$$\frac{\sqrt{n}}{\eta_n}\bigg(\frac{1}{n(n-1)(n-2)}\sum \mathbb{I}\big\{X_n[i_1,i_2,i_3] = y\big\} - P\big(X_n[1,2,3] = y\big)\bigg) \xrightarrow{d} Z,$$

where the sum runs over all distinct triples $i_1,i_2,i_3 \leq n$, and

$$\eta_n^2 = \sum_{i\leq m(n)} \pi_i^n E_i(n)\bigg(E_i(n) - \sum_{j\leq m(n)} \pi_j^n E_j(n)\bigg) \quad \textit{almost surely.}$$

The Wasserstein distance to the limit is $O(\eta_n^{-3} n^{-\frac{1}{2}} \|f(X_n)\|_4^{\frac{3}{4}})$.

The simplest SBM is an *Erdős–Rényi* (ER) graph, where each edge is an independent Bernoulli variable with success probability p, that is, $v := p$ is constant. This model has been thoroughly studied, and we can relate the corollary to some known results.

EXAMPLES.

(xvii) If X is an ER graph, $t(X[1,\dots,k],\cdot)$ satisfies a degenerate central limit theorem, with $\eta = 0$, see [2]. To see this in the corollary, set $X_n := X$ for all n. We can then consider the limit $\eta_n Z$. Since $E_i(n)$ does not depend on i nor n, we obtain $\eta_n = 0$.

(xviii) Let each X_n be an ER graph, with edge probability $p(n)$, and let $p(n) \to 0$. In principle, Corollary 18 holds: The limiting triangle density is 0, and $\eta_n = 0$. However, more bespoke results rescale by $1/\sqrt{p(n)}$ to make small-scale behavior visible [25]. These do not follow from Theorem 10, since the variables $\mathbb{I}\{X_n[1,2,3] = y\}/p(n)$ are not uniformly integrable.

8.4. *Separate exchangeability.* A random r-array X is *separately exchangeable* if it is invariant under the action

$$\boldsymbol{\phi} x := x_{\phi_1(i_1),\dots,\phi_r(i_r)} \quad \text{for all } x \in \mathbf{X} \text{ and } \boldsymbol{\phi} = (\phi_1,\dots,\phi_r) \in \mathbb{S}_\infty^r.$$

Comparing to (26) shows that joint exchangeability is the diagonal invariance corresponding to separate exchangeability. Some models for relational data in machine learning assume separate exchangeability for matrices whose rows and columns are indexed by distinct sets (e.g., consumers and products), and joint exchangeability if the sets are identical (e.g., vertices of a graph) [35]. Separate exchangeability is the stronger property, and results in a faster rate and simpler asymptotic variance.

COROLLARY 19. *Let X be a separately exchangeable r-array, and let $f \in \mathbf{L}_2(X)$ be a function that satisfies* (34). *As $n \to \infty$,*

$$\sqrt{n^r}\,\overline{\mathbb{F}}_n(f,X) = \sqrt{n^r}\left(\frac{1}{(n!)^r}\sum_{\phi\in\mathbb{S}_n^r} f(\phi X) - \mathbb{E}\left[f(X)|\mathbb{S}_\infty^r\right]\right) \xrightarrow{d} \eta Z,$$

where Z is standard normal and independent of η. The asymptotic variance satisfies $\eta^2 = \mathrm{Var}[f(X)|\mathbb{S}_\infty^r] < \infty$ almost surely.

EXAMPLE.

(xix) The convergence rate for homomorphism densities is in general $n^{-1/2}$ if a graph is exchangeable, but n^{-1} if it is Erdős–Rényi (e.g., [2]). Corollary 19 shows that is a consequence of additional symmetries in ER graphs, since they are not only jointly but even separately exchangeable.

8.5. *Graphex models*. Caron and Fox [14] have proposed a class of random graphs that, with extensions and refinements by other authors [9, 45], are referred to as *graphex models*. Recall from (36) how an ergodic exchangeable graph is generated by a graphon $w : [0,1]^2 \to [0,1]$ and independent uniform variables. A graphex model is defined similarly, by a symmetric measurable function $\omega : \mathbb{R}^2_{\geq 0} \to [0,1]$ and a unit-rate Poisson process $\Pi = \{(U_1,V_1),(U_2,V_2),\dots\}$ on $\mathbb{R}^2_{\geq 0}$. Let U_{ij}, for $i \leq j \in \mathbb{N}$, again be i.i.d. uniform elements of $[0,1]$. Define a random countable subset X_ω of $\mathbb{R}^2_{\geq 0}$ as

$$(V_i,V_j) \in X_\omega \quad \Longleftrightarrow \quad U_{ij} < \omega(U_i,U_j).$$

This set is interpreted as a graph, in which vertices V_i and V_j are connected if the pair (V_i, V_j) is in X_ω. The set X_ω thus functions as a form of adjacency matrix, but each vertex is identified by the value V_i, rather than the index i. A subgraph is not selected as an $n \times n$ submatrix, but by placing a rectangle $[0, s)^2$ in the plane: The subgraph $g_s(X_\omega)$ for $s \in (0, \infty]$ is

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

Suppose an instance $g_s(X_\omega)$ with N vertices is observed. Veitch and Roy [45] have shown that one can estimate the restriction $\omega|_{[0,s]^2}$ of ω, provided s is known: Subdivide $[0, s)^2$ into quadratic patches I_{ij}, and define a piece-wise constant function $\hat{\omega}_s$ on $[0, s)^2$ by specifying its value on each patch as

$$\hat{\omega}_s|_{I_{ij}} := \mathbb{I}\{(i, j) \in G\} \quad \text{where } I_{ij} := \left[\frac{i-1}{N}s, \frac{i}{N}s\right) \times \left[\frac{j-1}{N}s, \frac{j}{N}s\right).$$

This estimator is consistent on bounded domains $[0, t)^2$, in the following sense: Regard $\hat{\omega}_s$ as a function $\mathbb{R}^2_{\geq 0} \to [0, 1]$, with constant value 0 outside $[0, s)^2$. Generate $X_{\hat{\omega}_s}$ according to (8.5), using a Poisson process and uniform variables that are independent of X_ω. Then

$$g_t(X_{\hat{\omega}_s}) \xrightarrow{\text{d}} g_t(X_\omega) \quad \text{as } s \to \infty, \tag{38}$$

for every fixed $t \in (0, \infty)$ [45]. If f is a measurable function of finite graphs, the Veitch–Roy estimator of $\mathbb{E}[f(g_t(X_\omega))]$ is therefore

$$\hat{f}_s := \mathbb{E}[f(g_t(X_{\hat{\omega}_s}))|g_s(X_\omega)].$$

The distributional convergence in (38) implies $\hat{f}_s \to \mathbb{E}[f(g_t(X_\omega))]$ almost surely as $s \to \infty$.

We illustrate how to obtain rates for a simple example: Fix $t > 0$. For a finite graph g, choose f as

$$f(g) := \frac{1}{t^2}|\text{edge set of } g| \quad \text{hence } f(g_t(X_\omega)) = \frac{1}{t^2}|X_\omega \cap [0, t)^2|. \tag{39}$$

The function $(\omega, t) \mapsto \mathbb{E}[f(g_t(X_\omega))]$ is then similar to the edge density in a graphon model. Consider the random sets

$$\mathcal{V}_{mn} := X_\omega \cap [m, m+1) \times [n, n+1) \quad \text{for } m, n \in \mathbb{N}.$$

If we choose $s \in \mathbb{N}$, we have

$$\hat{f}_s = \frac{1}{t^2} \sum_{(i,j) \in g_s(X_\omega)} P((i, j) \in g_t(X_{\hat{\omega}_s})|g_s(X_\omega)) = \frac{1}{s^2} \sum_{m,n<s} |\mathcal{V}_{mn}|.$$

It follows from the construction of X_ω that the random array $(|\mathcal{V}_{mn}|)_{m,n}$ is jointly exchangeable and ergodic. We can hence apply Theorem 17, and obtain the following corollary.

COROLLARY 20. *Let* $\omega : \mathbb{R}^2_{\geq 0} \to [0, 1]$ *be a measurable and symmetric function, and fix* $t > 0$. *Define* f *as in* (39). *Then, for* $Z \sim N(0, 1)$,

$$\sqrt{s}(\hat{f}_s - \mathbb{E}[f(g_t(X_\omega))]) \xrightarrow{d} \eta Z \quad \text{as } s \to \infty,$$

where $\eta^2 = 4\,\mathrm{Cov}[|X_\omega \cap [0, 1]^2|, |X_\omega \cap [0, 1] \times [0, 2]||\mathbb{G}]$ *is a finite constant.*

The random set X_ω is invariant under an uncountable permutation group that transforms each axis $\mathbb{R}_{\geq 0}$ [14], and is in fact ergodic [29]. That is an example of uncountable exchangeability, as described at the beginning of this section. The local counts $|\mathcal{V}_{mn}|$ are a device to reduce uncountable to countable exchangeability, and hence to invariance under a nice group.

9. Applications II: Marked point processes. Random geometric measures are point processes whose behavior at a given point may depend on points nearby. They originate from so-called germ-grain models in physics [22], and are used to study for example, nearest neighbor methods and Voronoi tesselations [23, 38]. Theorems 4 and 6 are directly applicable.

9.1. *Setup.* The following definitions are adapted from those of Penrose [38], with some simplifications: Consider two Polish spaces, $\mathbf{X}$ (which we think of as a set of points) and $\mathbf{Y}$ (a set of marks or covariates), both equipped with their Borel σ-algebras. Denote by $\mathbf{M}$ the space of σ-finite measures on $\mathbf{X} \times \mathbf{Y}$, equipped with the σ-algebra generated by the evaluation maps, and by $\mathcal{F}$ the set of finite subsets of $\mathbf{X} \times \mathbf{Y}$. Let $\mu : \mathbf{X} \times \mathbf{Y} \times \mathcal{F} \to \mathbf{M}$ be a measurable map, and $W \subset \mathbf{X} \times \mathbf{Y}$ a compact set. Loosely speaking, μ assigns to each marked point (x, y) a measure $\mu(x, y, F)$ that depends on a set F of points near x, and on their marks. These nearby points are collected by using W as an observation window, which is moved over $\mathbf{X} \times \mathbf{Y}$ by elements of a group: Let $\mathbb{G}$ be a nice group that acts measurably on $\mathbf{X}$. We extend the action to one on $\mathbf{X} \times \mathbf{Y}$ by defining

$$\phi(x, y) := \big(\phi(x), y\big) \quad \text{for all } \phi \in \mathbb{G}, (x, y) \in \mathbf{X} \times \mathbf{Y}. \tag{40}$$

For compact $\mathbf{A}_1, \mathbf{A}_2, \ldots \subset \mathbb{G}$, write $\mathbf{A}_n W = \{(\phi(x), y) | \phi \in \mathbf{A}_n, (x, y) \in W\}$. If Π is a point process on $\mathbf{X} \times \mathbf{Y}$, then

$$\nu_n(\cdot) := \frac{1}{|\mathbf{A}_n|} \sum_{(x,y)\in\Pi_n} \mu(x, y, \Pi_n)(\cdot) \quad \text{for } \Pi_n := \Pi \cap \mathbf{A}_n W \tag{41}$$

is a random measure on $\mathbf{X} \times \mathbf{Y}$. The sequence (ν_n) is called a *random geometric measure* if Π is invariant under the action (40), and if the sets Π_n are almost surely finite. See [4, 38] for similar definitions.

9.2. *Asymptotic normality.* A central theme in the literature on random geometric measures is the limiting behavior of statistics of the form

$$\nu_n(h) := \int_{\mathbf{X}\times\mathbf{Y}} h(x, y)\nu_n(dx, dy) \quad \text{for } h : \mathbf{X} \times \mathbf{Y} \to \mathbb{R}.$$

Such results typically require that the window does not collect any point more than once. A simple condition that excludes such repetitions is as follows: Require (i) that $\mathbb{G}$ contains a subgroup $\mathbb{H}$ such that $\phi(W) \cap \psi(W) = \varnothing$ for distinct $\phi, \psi \in \mathbb{H}$, and (ii) that $\mathbb{H}W = \mathbb{G}W$. Informally, $\mathbb{H}W$ "tiles" the set $\mathbb{G}W \subset \mathbf{X}$ of points reached by the window. We also require that (iii) the set $\{\phi \in \mathbb{G} | \phi(W) \cap W \neq \varnothing\}$ is compact. If $\mathbb{G} = \mathbf{X} = \mathbb{R}^2$, for example, one might choose $W = [-1, 1]^2$ and $\mathbb{H} = \{(2i, 2j) | i, j \in \mathbb{Z}\}$. The relationship to our results becomes clear if we define

$$f_n(F) := \int_{\mathbf{X}\times\mathbf{Y}} h\big(x', y'\big) \sum_{(x,y)\in F\cap W} \mu(x, y, \Pi_n)\big(dx', dy'\big)$$

and

$$f(F) := \int_{\mathbf{X}\times\mathbf{Y}} h\big(x', y'\big) \sum_{(x,y)\in F\cap W} \mu(x, y, \Pi)\big(dx', dy'\big)$$

for $F \in \mathcal{F}$, and observe that $\nu_n(h) \approx \frac{1}{|\mathbf{A}_n \cap \mathbb{H}|} \int_{\mathbf{A}_n\cap\mathbb{H}} f_n(\phi(\Pi)) |d\phi| = \mathbb{F}_n(f_n, \Pi)$. We apply Theorem 4 and 6, and obtain the following proposition.

PROPOSITION 21. *Require that* (*i*)–(*iii*) *above hold, and that the sets* $\mathbf{A}_n$ *in* (41) *form a tempered Følner sequence. For each* n, *let* $\alpha^{(n)}(\cdot|\mathbb{G})$ *be the conditional mixing coefficient of* Π *and* f_n. *If*

$$\sup_n \int_{\mathbb{G}} \alpha^{(n)}\big(d(e,\phi)|\mathbb{G}\big)^{\frac{\varepsilon}{2+\varepsilon}} |d\phi| < \infty \quad \textit{and} \quad \|f_n(\Pi)^{2+\varepsilon}\|_1 < \infty$$

holds for some $\varepsilon \geq 0$, *then as* $n \to \infty$,

$$\sqrt{|\mathbf{A}_n \cap \mathbb{H}|}\big(\nu_n(h) - \mathbb{E}[\nu_n(h)|\mathbb{G}]\big) \xrightarrow{d} \eta Z \quad \textit{for } Z \sim N(0,1),$$

where $\eta^2 = \int_{\mathbb{H}} \mathrm{Cov}[f(\Pi), f(\phi\Pi)|\mathbb{G}]|d\phi|$ *and* $\eta \perp\!\!\!\perp Z$. *Moreover*,

$$d_{\mathrm{W}}\bigg(\frac{\sqrt{|\mathbf{A}_n \cap \mathbb{H}|}}{\eta}\big(\nu_n(h) - \mathbb{E}[\nu_n(h)|\mathbb{G}]\big), Z\bigg) = O\bigg(\frac{1}{\sqrt{|\mathbf{A}_n \cap \mathbb{H}|}} \max\big\{1, \|f_n(\Pi)\|_4^3\big\}\bigg).$$

9.3. *Relationship to existing results.* Versions of the result above are known in the case where $\mathbf{X}$ is $\mathbb{R}^r$, $\mathbb{G} = \mathbb{R}^r$ consists of shifts, and $\mathbf{A}_n$ is the Euclidean ball $\mathbf{B}_n$ [22, 23, 38]. These are not phrased in terms of conditional mixing, but instead use a "stabilization condition" (e.g. [4]). The next result translates stabilization to mixing conditions. There is no standardized definition of stabilization; the one we state below is similar to Definition 2.4 of Penrose [38], which he calls *power-law stabilizing* of order q. For $(x, y) \in \mathbf{X} \times \mathbf{Y}$ and $F \in \mathcal{F}$, let F_t be the truncated set $\{(\tilde{x}, \tilde{y}) \in F | d(x, \tilde{x}) \leq t\}$. The *stabilization radius* of μ is

$$R(x, y, F) := \inf\big\{t > 0 | \mu(x, y, F) = \mu(x, y, F_t)\big\},$$

where we use the convention $\inf \varnothing = \infty$. If

$$\sup_{s>0} \sup_{(x,y)\in W} s^q P\big(R(x, y, \Pi) > s\big) < \infty \quad \text{for some } q > 1, \tag{42}$$

μ is *polynomially stable* of order q. The condition implies conditional mixing if the metric balls in $\mathbb{G}$ do not expand too quickly.

PROPOSITION 22. *Let* Π *be a Poisson process, and let* μ *be polynomially stable of order* q. *If the metric balls* $\mathbf{B}_n$ *in* $\mathbb{G}$ *satisfy* $\sup_{n\in\mathbb{N}} n^{-r}|\mathbf{B}_n| < \infty$ *for some* $r > 0$, *then*

$$\sup_n \int_{\mathbb{G}} \alpha^{(n)}\big(d(e,\phi)|\mathbb{G}\big)^{\frac{\varepsilon}{2+\varepsilon}} |d\phi| < \infty \quad \textit{whenever } q > \frac{2+\varepsilon}{\varepsilon} r.$$

That holds in particular for the groups $\mathbb{R}^r$, since an r-dimensional Euclidean ball has volume $|\mathbf{B}_n| = (\sqrt{\pi}n)^r / \Gamma(\frac{r}{2}+1)$. Geometric group theory provides further examples: A group that satisfies $\sup_{n\in\mathbb{N}} n^{-r}|\mathbf{B}_n| < \infty$ and is also finitely generated is said to be of *polynomial growth* [32]. Nice groups of polynomial growth include $\mathbb{Z}^d$, the groups in Corollary 8, or the discrete Heisenberg groups (e.g., [18]).

10. Applications III: Entropy. The entropy of a stationary process is defined as a limit. This limit exists almost surely, by the Shannon–McMillan–Breiman (SMB) theorem [42]. It has a natural generalization to invariant processes (e.g., [20]), which again converges almost surely [31]. An adaptation of Theorem 4 gives conditions under which it is asymptotically normal. In this section, we assume $\mathbb{G}$ is discrete, and *finitely generated*, which means there is a finite subset $G \subset \mathbb{G}$ such that $\mathbb{G}$ is the smallest group containing G. That is, for example, true for $\mathbb{Z}^r$ (choose G as the set of unit coordinate vectors), but not for $\mathbb{S}_\infty$.

10.1. *Entropy.* Let Y be a discrete random variable with mass function $p(k) := P(Y = k)$ for $k \in \mathbb{N}$. If $Y_1, Y_2, \ldots$ are i.i.d. copies of Y, the law of large numbers guarantees almost sure convergence

$$-\frac{1}{n}\log\bigl(p(Y_1) \times \cdots \times p(Y_n)\bigr) \xrightarrow{n\to\infty} -\mathbb{E}\bigl[\log p(Y)\bigr] =: H[Y].$$

The constant $H[Y]$ is the *entropy* of Y [28]. If $X = (X_i)_{i\in\mathbb{Z}}$ is a stochastic process with values in the finite set $[K]$, the entropy can be defined similarly: If p_n is the joint mass function of $(X_1, \ldots, X_n)$, and X is stationary and ergodic, there is a constant $h[X] \geq 0$ such that

$$-\frac{1}{n}\log p_n(X_1, \ldots, X_n) \xrightarrow{n\to\infty} h[X] \quad \text{almost surely.} \tag{43}$$

This is the SMB theorem, and $h[X]$ is again called the entropy, or the entropy rate [42]. The term $-\frac{1}{n}\log p_n(X_1, \ldots, X_n)$ is the *empirical entropy*.

10.2. *Entropy of invariant distributions.* Let $\mathbb{G}$ be countable, and $(\mathbf{A}_n)$ a tempered Følner sequence with $|\mathbf{A}_n|/\log(n) \to \infty$. Let X be a $\mathbb{G}$-ergodic random element of $\mathbf{X}$. To define entropy, regard $(\phi X)_{\phi\in\mathbb{G}}$ as a stochastic process on the group, and discretize its state space: Choose a partition $\lambda := (\lambda_1, \ldots, \lambda_K)$ of $\mathbf{X}$ into a finite number of Borel sets, and write $\lambda(x) = k$ if $x \in \lambda_k$. Let p_n be the joint mass function of $(\lambda(\phi X))_{\phi\in\mathbf{A}_n}$. Then there is a constant $h_\lambda[X] \geq 0$ such that

$$h_n(\lambda, X) := -\frac{1}{|\mathbf{A}_n|}\log p_n\bigl(\bigl(\lambda(\phi X)\bigr)_{\phi\in\mathbf{A}_n}\bigr) \xrightarrow{n\to\infty} h_\lambda[X] \quad \text{almost surely.}$$

This result is again due to Lindenstrauss [31]. To recover (43), choose X as a stationary process $(X_i)_{i\in\mathbb{Z}}$, and $\lambda_k := \{x = (x_i)_{i\in\mathbb{Z}} | x_0 = k\}$.

10.3. *Asymptotic normality.* Suppose $\mathbb{G}$ admits a total order $\preceq$ that is left-invariant (i.e., $\phi \preceq \psi$ if and only if $\pi\phi \preceq \pi\psi$ for $\phi, \psi, \pi \in \mathbb{G}$). The process values indexed by a set $G \subset \mathbb{G}$ are predictive of the value at ϕ if

$$L_\phi(G) := \log P\bigl[\lambda(\phi X) | \lambda(\psi X), \psi \in G\bigr]$$

is large, where P denotes probability under the law of X. The scalar

$$\rho_m := \sup_{A\subset\mathbb{G}} \bigl\| L_e(A) - L_e(A \cap \mathbf{B}_m)\bigr\|_2$$

measures how well the value at the identity is predicted by values within a radius m. Recall that the definition of mixing in Section 3 uses pairs ϕ_1, ϕ_2 in $\mathbb{G}$. We extend it to k-tuples: For $k \in \mathbb{N}$, define

$$\mathcal{C}(t, k) := \bigl\{(A, B) \in \sigma_f(\phi_1, \ldots, \phi_k) \otimes \sigma_f(G) | G \subset \mathbb{G}, \phi_1, \ldots, \phi_k \in \mathbb{G} \setminus \mathbf{B}_t(G)\bigr\},$$

and $\alpha(t, k) := \sup_{(A,B)\in\mathcal{C}(t,k)} |P(A, B) - P(A)P(B)|$. The mixing coefficient in Section 3 is hence $\alpha(t) = \alpha(t, 2)$.

THEOREM 23. *Let $\mathbb{G}$ be a finitely generated, nice group with left-invariant total order, and let X be $\mathbb{G}$-ergodic with* $\sup_{A\subset\mathbb{G}} \|L_e(A)\|_{2+\varepsilon} < \infty$ *for some $\varepsilon > 0$. Choose a tempered Følner sequence satisfying $|\mathbf{A}_n \triangle \mathbf{B}_{b_n}\mathbf{A}_n|/|\mathbf{A}_n| \to 0$ and $\sqrt{|\mathbf{A}_n|}\rho_{b_n} \to 0$, for some sequence (b_n) of positive scalars. If*

$$\sum_{i\in\mathbb{N}} |\mathbf{B}_i| \min_{m\leq i}\bigl(\rho_m + \alpha\bigl(i - m, |\mathbf{B}_m|\bigr)^{\frac{\varepsilon}{2+\varepsilon}}\bigr) < \infty \tag{44}$$

holds for the mixing coefficient of the function $f := \lambda$, *then*

$$\sqrt{|\mathbf{A}_n|}\big(h_n(\lambda, X) - h_\lambda[X]\big) \xrightarrow{d} \eta Z \quad \text{as } n \to \infty,$$

where the asymptotic variance is independent of Z *and satisfies*

$$\eta^2 = \sum_{\phi \in \mathbb{G}} \mathrm{Cov}\big[L_e\big(\{\psi \preceq e\}\big), L_\phi\big(\{\psi \preceq \phi\}\big)\big] < \infty \quad \text{almost surely.}$$

Condition (44) can be interpreted as follows: The proof represents h_n as

$$\log p_n\big((\lambda(\phi X))_{\phi \in \mathbf{A}_n}\big) = \sum_{\phi \in \mathbf{A}_n} L_\phi\big(\{\psi \in \mathbf{A}_n | \psi \preceq \phi\}\big),$$

and approximates it by the average $\mathbb{F}_n$ of $f'(X) = L_\phi(\{\psi | \psi \preceq \phi\} \cap \mathbf{B}(\phi, m))$. The approximation error is a function of ρ_m, and decreasing in m. Mixing, on the other hand, involves tuples in $\mathbf{B}_m$, and since $\alpha(\cdot, |\mathbf{B}_m|)$ is nondecreasing in $|\mathbf{B}_m|$, a smaller m means better mixing. Informally, dependence within the process is both beneficial (it makes predicting one value from others easier) and detrimental (it reduces mixing).

REMARK. (a) Left-invariance of the order is not required for asymptotic normality, but simplifies η. Provided it holds, η does not depend on the choice of $\preceq$. (b) Examples of groups satisfying Theorem 23 are $(\mathbb{Z}^r, +)$ and the groups in Corollary 8, or discrete Heisenberg groups [32]. (c) Existence of a total order implies $\phi^m \neq e$ for all $m \in \mathbb{N}$, unless $\phi = e$. In algebraic terms, $\mathbb{G}$ is torsion-free [32].

Acknowledgments. Wenda Zhou suggested randomized averages, Ryan P. Adams the setting of Section 7, and Ismaël Castillo the stochastic block model application in Section 8.3. Victor Veitch provided clarifications regarding graphex models. We are indebted to the associate editor and to three referees for very thorough and helpful reports that have improved the paper substantially. This work was partly funded by the Gatsby Charitable Foundation (PO).

SUPPLEMENTARY MATERIAL

Supplementary Material to 'Limit theorems for distributions invariant under groups of transformations' (DOI: 10.1214/21-AOS2165SUPP; .pdf). The supplement [3] contains proofs of all results in this article, and further auxiliary results that are used in the proofs.

REFERENCES

[1] ALFSEN, E. M. (1971). *Compact Convex Sets and Boundary Integrals*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete*, *Band* 57. Springer, New York. MR0445271

[2] AMBROISE, C. and MATIAS, C. (2012). New consistent and asymptotically normal parameter estimates for random-graph mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 3–35. MR2885838 https://doi.org/10.1111/j.1467-9868.2011.01009.x

[3] AUSTERN, M. and ORBANZ, P. (2022). Supplement to "Limit theorems for distributions invariant under groups of transformations." https://doi.org/10.1214/21-AOS2165SUPP

[4] BARYSHNIKOV, YU. and YUKICH, J. E. (2005). Gaussian limits for random measures in geometric probability. *Ann. Appl. Probab.* **15** 213–253. MR2115042 https://doi.org/10.1214/105051604000000594

[5] BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. MR2906868 https://doi.org/10.1214/11-AOS904

[6] BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR1324786

[7] BOLTHAUSEN, E. (1982). On the central limit theorem for stationary mixing random fields. *Ann. Probab.* **10** 1047–1050. MR0672305

[8] BONDAR, J. V. and MILNES, P. (1981). Amenability: A survey for statistical applications of Hunt–Stein and related conditions on groups. *Z. Wahrsch. Verw. Gebiete* **57** 103–128. MR0623457 https://doi.org/10.1007/BF00533716

[9] BORGS, C., CHAYES, J. T., COHN, H. and HOLDEN, N. (2017). Sparse exchangeable graphs and their limits via graphon processes. *J. Mach. Learn. Res.* **18** Paper No. 210, 71. MR3827098

[10] BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2008). Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851. MR2455626 https://doi.org/10.1016/j.aim.2008.07.008

[11] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[12] BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. MR2178042 https://doi.org/10.1214/154957805100000104

[13] BÜHLMANN, H. (1958). Le problème "limite central" pour les variables aléatoires échangeables. *C. R. Acad. Sci. Paris Sér. I Math.* **246** 534–536. MR0092265

[14] CARON, F. and FOX, E. B. (2017). Sparse graphs using exchangeable random measures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1295–1366. MR3731666 https://doi.org/10.1111/rssb.12233

[15] CHOI, D. S., WOLFE, P. J. and AIROLDI, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99** 273–284. MR2931253 https://doi.org/10.1093/biomet/asr053

[16] DAVEZIES, L., D'HAULTFŒUILLE, X. and GUYONVARCH, Y. (2021). Empirical process results for exchangeable arrays. *Ann. Statist.* **49** 845–862. MR4255110 https://doi.org/10.1214/20-aos1981

[17] DIACONIS, P. and JANSON, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* (7) **28** 33–61. MR2463439

[18] DIACONIS, P. and MALLIARIS, M. (2021). Complexity and randomness in the Heisenberg groups (and beyond). *New Zealand J. Math.* **52** 403–426. MR4378158 https://doi.org/10.53733/134

[19] EAGLESON, G. K. and WEBER, N. C. (1978). Limit theorems for weakly exchangeable arrays. *Math. Proc. Cambridge Philos. Soc.* **84** 123–130. MR0501275 https://doi.org/10.1017/S0305004100054967

[20] EINSIEDLER, M. and WARD, T. (2011). *Ergodic Theory*. Springer, Berlin.

[21] GEORGII, H.-O. (2011). *Gibbs Measures and Phase Transitions*, 2nd ed. *De Gruyter Studies in Mathematics* **9**. de Gruyter, Berlin. MR2807681 https://doi.org/10.1515/9783110250329

[22] HEINRICH, L. and MOLCHANOV, I. S. (1999). Central limit theorem for a class of random measures associated with germ-grain models. *Adv. in Appl. Probab.* **31** 283–314. MR1724553 https://doi.org/10.1239/aap/1029955136

[23] HUG, D., LAST, G. and SCHULTE, M. (2016). Second-order properties and central limit theorems for geometric functionals of Boolean models. *Ann. Appl. Probab.* **26** 73–135. MR3449314 https://doi.org/10.1214/14-AAP1086

[24] JAMES, L. F., LIJOI, A. and PRÜNSTER, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scand. J. Stat.* **36** 76–97. MR2508332 https://doi.org/10.1111/j.1467-9469.2008.00609.x

[25] JANSON, S., ŁUCZAK, T. and RUCINSKI, A. (2000). *Random Graphs*. *Wiley-Interscience Series in Discrete Mathematics and Optimization*. Wiley Interscience, New York. MR1782847 https://doi.org/10.1002/9781118032718

[26] JENSEN, J. L. and KÜNSCH, H. R. (1994). On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Ann. Inst. Statist. Math.* **46** 475–486. MR1309718

[27] KALLENBERG, O. (1999). Multivariate sampling and the estimation problem for exchangeable arrays. *J. Theoret. Probab.* **12** 859–883. MR1702867 https://doi.org/10.1023/A:1021692202530

[28] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR1876169 https://doi.org/10.1007/978-1-4757-4015-8

[29] KALLENBERG, O. (2005). *Probabilistic Symmetries and Invariance Principles*. *Probability and Its Applications* (*New York*). Springer, New York. MR2161313

[30] KLOPP, O., TSYBAKOV, A. B. and VERZELEN, N. (2017). Oracle inequalities for network models and sparse graphon estimation. *Ann. Statist.* **45** 316–354. MR3611494 https://doi.org/10.1214/16-AOS1454

[31] LINDENSTRAUSS, E. (2001). Pointwise theorems for amenable groups. *Invent. Math.* **146** 259–295. MR1865397 https://doi.org/10.1007/s002220100162

[32] LÖH, C. (2017). *Geometric Group Theory: An Introduction*. *Universitext*. Springer, Cham. MR3729310 https://doi.org/10.1007/978-3-319-72254-2

[33] LOVÁSZ, L. and SZEGEDY, B. (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** 933–957. MR2274085 https://doi.org/10.1016/j.jctb.2006.05.002

[34] MAITRA, A. (1977). Integral representations of invariant measures. *Trans. Amer. Math. Soc.* **229** 209–225. MR0442197 https://doi.org/10.2307/1998506

[35] ORBANZ, P. and ROY, D. M. (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 437–461.

[36] PANCHENKO, D. (2013). Spin glass models from the point of view of spin distributions. *Ann. Probab.* **41** 1315–1361. MR3098679 https://doi.org/10.1214/11-AOP696

[37] PARISI, G. (1988). *Statistical Field Theory*. Addison-Wesley, Reading.

[38] PENROSE, M. D. (2007). Gaussian limits for random geometric measures. *Electron. J. Probab.* **12** 989–1035. MR2336596 https://doi.org/10.1214/EJP.v12-429

[39] PITMAN, J. (2006). *Combinatorial Stochastic Processes*. *Lecture Notes in Math.* **1875**. Springer, Berlin. MR2245368

[40] ROSS, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8** 210–293. MR2861132 https://doi.org/10.1214/11-PS182

[41] SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0595165

[42] SHIELDS, P. C. (1996). *The Ergodic Theory of Discrete Sample Paths*. *Graduate Studies in Mathematics* **13**. Amer. Math. Soc., Providence, RI. MR1400225 https://doi.org/10.1090/gsm/013

[43] STROOCK, D. W. and ZEGARLIŃSKI, B. (1992). The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Comm. Math. Phys.* **149** 175–193. MR1182416

[44] VARADARAJAN, V. S. (1963). Groups of automorphisms of Borel spaces. *Trans. Amer. Math. Soc.* **109** 191–220. MR0159923 https://doi.org/10.2307/1993903

[45] VEITCH, V. and ROY, D. M. (2019). Sampling and estimation for (sparse) exchangeable graphs. *Ann. Statist.* **47** 3274–3299. MR4025742 https://doi.org/10.1214/18-AOS1778

[46] WEISS, B. (2003). Actions of amenable groups. In *Topics in Dynamics and Ergodic Theory*. *London Mathematical Society Lecture Note Series* **310** 226–262. Cambridge Univ. Press, Cambridge. MR2052281 https://doi.org/10.1017/CBO9780511546716.012

APPENDIX A: PROOF OVERVIEW AND AUXILIARY RESULTS

The proofs are presented in three parts, for the basic limit theorems in Appendix B, for the general ones in Appendix C, and for all other results in Appendix D. The basic results (Theorems 4 and 6) are special cases of the general ones (Theorems 10 and 11), but we prove them first to clarify the approach. The general proofs require changes, but follow the same layout.

A.1. Proof overview. The proofs of the main results, Theorems 4, 6, 10 and 11, use Stein's method [e.g. 7]: For the function class

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

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

$$d_{\mathrm{w}}(W, Z) \leq \sup_{t \in \mathcal{F}} \left| \mathbb{E}[W t(W) - t'(W)] \right| \qquad \text{for } Z \sim N(0,1)\,. \tag{2}$$

The distance d_{w} metrizes convergence in distribution for variables with a first moment [e.g. 7]. One can therefore establish a central limit theorem for a sequence (W_n) of such variables by showing $d_{\mathrm{w}}(W_n, Z) \to 0$, and hence by showing that the right-hand side of (2) vanishes as $n \to \infty$.

Basic case. In broad strokes, Theorems 4 and 6 are proven as follows:

- Choose $W = W_n$ as a suitably scaled version of $\eta(n)^{-1}\mathbb{F}_n$, where $\eta(n)$ is a (for now unspecified) positive random variable.
- To upper-bound (2), split W at a cut-off distance b_n in $\mathbb{G}$, into a short-range and a long-range term. Adapting (2) to these modifications yields a refined bound, in Lemma 7. The leg work of the proof is then to control each term in this bound.
- Stein's method involves the notion of "dependency neighborhoods" [7]: A set, say $\mathcal{N}(i)$, of indices for a random variable X_i such that $X_i \perp\!\!\!\perp X_j$ if $j \notin \mathcal{N}(i)$. In our proofs, the neighborhood is the area within the cut-off b_n, but terms inside and outside the neighborhood are not completely independent. We hence bound long-range terms using conditional mixing.
- Split f into small and large values at a threshold γ_n. Since no fourth moment is assumed, large values must be controlled explicitly.
- The resulting bound is a function of $\eta(n)$. Choose $\eta(n)$ as an approximation to the quantity η defined in the statement of Theorem 4.

The central limit theorem then follows by showing that the bound vanishes as $n \to \infty$, and the Berry-Esseen bound by additionally requiring a third and fourth moment, and substituting these into the bound.

REMARK. Limit theorems for random structures often require a condition called stable convergence [e.g. 4]. That is not required here; instead, the proof shows that $\eta(n)^{-1}\mathbb{F}_n$ converges to Z conditionally on $\sigma(\mathbb{G})$, which is then used to obtain convergence of $\mathbb{F}_n$ to ηZ. That is possible because η is constant given $\sigma(\mathbb{G})$. In terms of Theorem 2, η is a function of ξ, and hence $\sigma(\mathbb{G})$-measurable.

General case. Proving Theorems 10 and 11 requires a number of modifications:

- Since the dimension k_n of the group may grow with n, we work with surrogate functions that depend only on the first few entries of $\boldsymbol{\phi} \in \mathbb{G}^{k_n}$.
- Working in $\mathbb{G}^{k_n}$ complicates the dependency neighborhoods.
- Since $\widehat{\mathbb{F}}_n$ is now random, we must also control the probability of selecting elements of the dependency neighborhood, using the spreading conditions.

A.2. Comments on other proof techniques. Central limit theorems can be proven with a range of tools, including Fourier techniques, Lindeberg's replacement trick, or martingale methods. Unlike Stein's method, these do not seem adaptable to our problems. In the case of concentration, the Efron-Stein inequality and other standard techniques similarly fail. There are several obstacles: *(i) Topology of the group.* Many martingale proofs, and the Efron-Stein approach to concentration, combine observations into blocks, and control dependence between blocks via an isoperimetric argument (i.e. block boundaries are of negligible size). That applies to some groups, such as $\mathbb{G}=\mathbb{Z}$, but fails even for $\mathbb{G}=\mathbb{Z}^2$. Bolthausen [1] used Stein's method to address an instance of this problem. *(ii) Lack of a total order.* Replacement arguments (e.g. Lindeberg's method and 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 permutation groups. *(iii) Group size*, since replacement arguments require countability.

REMARK. Martingales are applicable if $\mathbb{G}$ contains compact subgroups $\mathbb{G}_1 \subset \mathbb{G}_2 \subset \ldots$ such that $\mathbb{G}=\cup_n \mathbb{G}_n$. That is the case for $\mathbb{S}_\infty$, with $\mathbb{G}_n=\mathbb{S}_n$. If so, $(\mathbb{G}_n)$ is a Følner sequence, and $(\mathbb{F}_n)$ is a reverse martingale adapted to the filtration $\sigma(\mathbb{G}_1)\supset\sigma(\mathbb{G}_2)\supset\ldots$. That implies (17). The corresponding case of Theorem 4 (with more restrictive moment and mixing conditions) follows from the reverse martingale central limit theorem. Such arguments are used in [6] for convergence, and in [3] for asymptotic normality. However, the method has limitations even for $\mathbb{G}=\mathbb{S}_\infty$. For example: If (X_i) is an exchangeable sequence and h a function of two arguments, $(h(X_i,X_j))_{ij}$ is an exchangeable array, but even with proper normalization, $\sum_{i<j} h(X_i,X_j)$ is not a reverse martingale unless $h(x,y)=h(y,x)$.

A.3. Auxiliary results. We begin with a result that allows us to bound the Wasserstein distance d_{w}. Recall that $\mathcal{L}$ denotes the set of Lipschitz functions with constant 1. It is a standard result that

$$d_{\mathrm{w}}(X,Y) \;=\; \sup_{h\in\mathcal{L}} |\mathbb{E}[h(X)]-\mathbb{E}[h(Y)]| \;=\; \inf \mathbb{E}[|X'-Y'|]\,, \tag{3}$$

where the infimum is taken over all couplings (X',Y') of X and Y. This identity is sometimes known as the Kantorovich-Rubinstein formula. In analogy to d_{w}, we define the conditional (and hence random) distance

$$d_{\mathrm{w}}(X,Y|\mathbb{G}) \;:=\; \sup_{h\in\mathcal{L}} |\mathbb{E}[h(X)|\mathbb{G}]-\mathbb{E}[h(Y)|\mathbb{G}]|\,.$$

The next lemma shows how it relates to d_{w}.

LEMMA 1. *Let X and Y be random variables in $\mathbf{L}_1(\mathbb{R})$, defined on an abstract probability space $(\Omega,\mathcal{A},\mathbb{P})$. Then*

$$d_{\mathrm{w}}(X,Y|\mathbb{G}) = \inf\; \mathbb{E}\big[|X'-Y'|\,\big|\mathbb{G}\big] \qquad \mathbb{P}\text{-a.s.},$$

where the infimum runs over all couplings (X',Y') of the conditional variables $X|\sigma(\mathbb{G})$ and $Y|\sigma(\mathbb{G})$, and

$$d_{\mathrm{w}}(X,Y) \le \mathbb{E}[d_{\mathrm{w}}(X,Y|\mathbb{G})] \;=\; \|d_{\mathrm{w}}(X,Y|\mathbb{G})\|_1\,.$$

PROOF. Since both random variables are real-valued, we can choose regular conditional distributions p for X and q for Y. That is, $p_\omega = P(\,\bullet\,|\mathbb{G})(\omega)$ holds for $\mathbb{P}$-almost all $\omega\in\Omega$, the map $\omega\mapsto p_\omega$ is measurable, and the same holds for q and Y. We can then apply (3) pointwise in ω, which shows that $\mathbb{P}$-almost surely,

$$\sup_{h\in\mathcal{L}} |\mathbb{E}[h(X)|\mathbb{G}](\omega)-\mathbb{E}[h(Y)|\mathbb{G}](\omega)| \;=\; d_{\mathrm{w}}(p_\omega,q_\omega) \;=\; \inf \mathbb{E}|X'-Y'|\,,$$

where the infimum is taken over all (X',Y') with marginal distributions p_ω and q_ω. That shows the first identity. The second claim holds since

$$\begin{aligned} d_{\mathrm{w}}(X,Y) &= \sup_{h\in\mathcal{L}}\bigl|\mathbb{E}[h(X)]-\mathbb{E}[h(Y)]\bigr| = \sup_{h\in\mathcal{L}}\bigl|\mathbb{E}[\mathbb{E}[h(X)|\mathbb{G}]-\mathbb{E}[h(Y)|\mathbb{G}]]\bigr| \\ &\le \mathbb{E}[\sup_{h\in\mathcal{L}}\bigl|\mathbb{E}[h(X)|\mathbb{G}]-\mathbb{E}[h(Y)|\mathbb{G}]\bigr|] = \mathbb{E}[d_{\mathrm{w}}(X,Y|\mathbb{G})]\,, \end{aligned}$$

where we have used the tower property and the relation $\sup\mathbb{E}\le\mathbb{E}\sup$. □

Conditioning in d_{w} lets us swap a random variable Y (which in the proofs will be the asymptotic variance) between arguments:

LEMMA 2 (Random scaling). *Let X, Y, and Z be random variables in $\mathbf{L}_2(\mathbb{R})$, such that Y is $\sigma(\mathbb{G})$-measurable. If $Y\ge c$ almost surely for some $c>0$,*

$$d_{\mathrm{w}}(X,Z/Y)\le\|d_{\mathrm{w}}(XY,Z|\mathbb{G})\|_1/c\,.$$

PROOF. The second part of Lemma 1 shows that $d_{\mathrm{w}}(X,Z/Y)\le\|d_{\mathrm{w}}(X,Z/Y|\mathbb{G})\|_1$. Fix any $\epsilon>0$. Since Y is $\sigma(\mathbb{G})$-measurable, there is a coupling (X',Z') of the conditional variables $X|\sigma(\mathbb{G})$ and $Z|\sigma(\mathbb{G})$ such that $\mathbb{E}[|X'Y-Z'||\mathbb{G}]\le d_{\mathrm{w}}(XY,Z|\mathbb{G})+\epsilon$, now by the first part of Lemma 1. This coupling satisfies

$$\mathbb{E}[|X'Y/Y-Z'/Y||\mathbb{G}]\;\le\;\mathbb{E}[|X'Y-Z'||\mathbb{G}]/c\;\le\;\bigl(d_{\mathrm{w}}(XY,Z|\mathbb{G})+\epsilon\bigr)/c\,.$$

Since ϵ is arbitrary, it follows that $d_{\mathrm{w}}(X,Z/Y)\le\|d_{\mathrm{w}}(XY,Z|\mathbb{G})\|_1/c$. □

We must repeatedly use bounds of the form $\|\mathbb{E}[\,\bullet\,|\mathbb{G}]\|_1\lesssim\|\bullet\|_{\frac{2+\varepsilon}{2}}\alpha(k|\mathbb{G})^{\frac{2}{2+\varepsilon}}$ to "separate off" conditioning. The next two lemmas capture all cases needed in the proofs, for both $\varepsilon=0$ and $\varepsilon>0$. The first version applies to conditional mixing. Recall this involves a pair ϕ_1,ϕ_2 of distance at least k from a set $G\subset\mathbb{G}$, which here is of finite size m. The lemma shows that, if a transformation π does not move the pair too close to G, the desired inequality holds.

LEMMA 3 (Conditional mixing bound). *Let X be $\mathbb{G}$-invariant, Y a real-valued random variable, and $h\colon\mathbf{X}^{k+2}\times\mathbb{R}\to\mathbb{R}$ a measurable function with $\mathbb{E}[|h(X,\dots,X,Y)|]<\infty$. Fix $\phi_1,\phi_2,\psi_1,\dots,\psi_m\in\mathbb{G}$, and set*

$$H_\tau:=h(\psi_1X,\dots,\psi_mX,\tau^{-1}\phi_1X,\tau^{-1}\phi_2X,Y)\quad\textit{for } \tau\in\mathbb{G}\,.$$

Let π be an element of $\mathbb{G}$. If

$$Y\perp\!\!\!\perp X\,|\,\sigma(\mathbb{G})\quad\textit{and}\quad k\le\min_{i\le 2,j\le m}d(\tau^{-1}\phi_i,\psi_j)$$

for both $\tau=\pi$ and the identity $\tau=e$, then

$$\bigl\|\mathbb{E}[H_\pi|\mathbb{G},Y]-\mathbb{E}[H_e|\mathbb{G},Y]\bigr\|_1\;\le\;4\|H_\pi-H_e\|_{\frac{2+\varepsilon}{2}}\alpha(k|\mathbb{G})^{\frac{2}{2+\varepsilon}}$$

for any $\varepsilon\ge 0$.

PROOF. *Case 1: $\|H_\pi-H_e\|_\infty$ finite*. We approximate h by a step function

$$h^*(\,\bullet\,,\bullet\,,\bullet\,)=\textstyle\sum_{i=1}^N c_i\mathbb{I}(\,\bullet\in A_i,\,\bullet\in B_i,\,\bullet\in C_i)\,, \tag{4}$$

for some $N\in\mathbb{N}$, measurable sets A_i in $\mathbf{X}^m$, B_i in $\mathbf{X}^2$ and C_i in $\mathbb{R}$, and scalars $|c_i|\le\|h\|_\infty$. Define H^*_τ analogously to H_τ, by substituting h^* for h. Fix any $\delta>0$. Since h is integrable, h^* can be chosen to make $\|h-h^*\|_1$ arbitrarily small, and hence such that $\|(H_\pi - H_e)-(H^*_\pi - H^*_e)\|_1 \leq \delta$. If we abbreviate

$$I_i := \mathbb{I}_{A_i}(\psi_1 X,\ldots,\psi_k X)\,\mathbb{I}_{C_i}(Y)\left(\mathbb{I}_{B_i}(\phi_1 X,\phi_2 X) - \mathbb{I}_{B_i}(\pi^{-1}(\phi_1 X,\phi_2 X))\right)$$

and $E_i := \mathbb{E}[I_i|\mathbb{G},Y]$, we have $\|\mathbb{E}[H^*_\pi - H^*_e|\mathbb{G},Y]\|_1 \leq \sum_{i=1}^{N_\delta}|c_i|\|E_i\|_1$ for some $N_\delta \in \mathbb{N}$. Using the definition of conditional mixing, we have

$$\begin{aligned}\textstyle\sum_{i\leq N_\delta}|c_i|\|E_i\|_1 \;&\leq\; \mathbb{E}\big[\textstyle\sum_{i|E_i>0}|c_i||E_i| + \sum_{i|E_i\leq 0}|c_i||E_i|\big]\\ &\leq\; \textstyle\sum_{i|E_i>0}|c_i|\mathbb{E}[E_i] - \sum_{i|E_i\leq 0}|c_i|\mathbb{E}[E_i]\\ &\leq\; \max_i |c_i|\big(\|\textstyle\sum_{i|E_i>0}E_i\|_1 + \|\sum_{i|E_i\leq 0}E_i\|_1\big)\\ &\leq\; 2\|H_\pi - H_e\|_\infty\,\alpha(k|\mathbb{G})\;.\end{aligned}\tag{5}$$

Since the right-hand side does not depend on δ or h^*, that implies

$$\|\mathbb{E}[H_\pi - H_e|\mathbb{G},Y]\|_1 \leq 2\|H_\pi - H_e\|_\infty\,\alpha(k|\mathbb{G})\;.$$

Case 2: $\|H_\pi - H_e\|_\infty$ *infinite*. For $r>0$, define

$$\Delta H := H_\pi - H_e \qquad \Delta H_r := \Delta H\cdot\mathbb{I}\{\Delta H\leq r\} \qquad \overline{\Delta H_r} := \Delta H - \Delta H_r\;.$$

The triangle inequality gives $\|\mathbb{E}[H_\pi - H_e|\mathbb{G},Y]\|_1 \leq \|\Delta H_r\|_1 + \|\overline{\Delta H_r}\|_1$, and case 1 above implies $\|\Delta H_r\|_1 \leq 2r\alpha(k|\mathbb{G})$. Since $\|h\|_{\frac{2+\varepsilon}{2}}$ is finite, we can assume $\|\Delta H\|_{\frac{2+\varepsilon}{2}} \leq 1$ without loss of generality. By Hölder's inequality,

$$\|\overline{\Delta H_r}\|_1 \;\leq\; \|\Delta H\|_{\frac{2+\varepsilon}{2}}\cdot\|\mathbb{I}\{\Delta H > r\}\|_{\frac{2+\varepsilon}{\varepsilon}} \;\leq\; 2r^{-\frac{\varepsilon}{2}}\;.$$

We hence obtain $\|\mathbb{E}[H_\pi - H_e|\mathbb{G},Y]\|_1 \leq 2r\alpha(k|\mathbb{G}) + 2r^{-\frac{\varepsilon}{2}} = 4\alpha(k|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}}$ by choosing $r = \alpha(k|\mathbb{G})^{\frac{-2}{2+\varepsilon}}$. □

The second version is the analogous result for marginal mixing coefficients. As in Section 5, $e_{i,\tau} = (e,\ldots,e,\tau,e,\ldots,e)$ denotes a vector with k_n entries and τ as the ith entry, and $\delta_{i,j}$ is defined as in (29).

LEMMA 4 (Marginal mixing bound). *Let X be a random element of $\mathbf{X}_n$, invariant under the diagonal action of $\mathbb{G}^{k_n}$, and Y a real-valued random variable. Let $h\colon \mathbf{X}_n^{k+2}\times\mathbb{R}\to\mathbb{R}$ be measurable, with $\mathbb{E}[|h(X,\ldots,X,Y)|]<\infty$. Fix $\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\boldsymbol{\psi}_1,\ldots,\boldsymbol{\psi}_m \in \mathbb{G}^{k_n}$. For any $i,j\leq k_n$, set*

$$H^{ij}_\tau := h(\boldsymbol{\psi}_1 X,\ldots,\boldsymbol{\psi}_k X, e_{i,\tau}\boldsymbol{\phi}_1 X, e_{j,\tau}\boldsymbol{\phi}_2 X, Y) \quad \textit{for } \tau\in\mathbb{G}\;,$$

where Let π be an element of $\mathbb{G}$. If

$$Y \perp\!\!\!\perp X \,|\, \sigma(\mathbb{G}) \quad \textit{and} \quad k \leq \delta_{ij}(e_{i,\tau}\boldsymbol{\phi}_1, e_{j,\tau}\boldsymbol{\phi}_2, \{\boldsymbol{\psi}_1,\ldots,\boldsymbol{\psi}_m\})$$

for both $\tau=\pi$ and the identity $\tau=e$, then

$$\big\|\mathbb{E}[H^{ij}_\pi|\mathbb{G},Y] - \mathbb{E}[H^{ij}_e|\mathbb{G},Y]\big\|_1 \;\leq\; 4\|H^{ij}_\pi - H^{ij}_e\|_{\frac{2+\varepsilon}{2}}\alpha_n(k|\mathbb{G})^{\frac{2}{2+\varepsilon}}$$

for any $\varepsilon\geq 0$.

Since the proof is almost identical to that of Lemma 3, we only highlight the requisite changes.

PROOF. If $\|H^{ij}_\pi - H^{ij}_e\|_\infty$ is finite, again use (4), now with measurable sets A_i in $\mathbf{X}^m_n$ and B_i in $\mathbf{X}^2_n$, and define H^{ij*}_τ by substituting h^* for h. For $\delta>0$ given, choose h^* such that $\|(H^{ij}_\pi - H^{ij}_e) - (H^{ij*}_\pi - H^{ij*}_e)\|_1 \leq \delta$. If we change the definition of I_i to

$$I_i := \mathbb{I}_{A_i}(\boldsymbol{\psi}_1 X,\ldots,\boldsymbol{\psi}_k X)\,\mathbb{I}_{C_i}(Y)\,\big(\mathbb{I}_{B_i}(\boldsymbol{\phi}_1 X,\boldsymbol{\phi}_2 X) - \mathbb{I}_{B_i}(e_{i,\pi}\boldsymbol{\phi}_1 X, e_{j,\pi}\boldsymbol{\phi}_2 X)\big)\;,$$

repeating (5) shows $\sum_{i\leq N_\delta}|c_i|\|E_i\|_1 \leq 2\|H^{ij}_\pi - H^{ij}_e\|_\infty\,\alpha_n(k|\mathbb{G})$, and hence

$$\|\mathbb{E}[H^{ij}_\pi - H^{ij}_e|\mathbb{G},Y]\|_1 \leq 2\|H^{ij}_\pi - H^{ij}_e\|_\infty\,\alpha_n(k|\mathbb{G})\;.$$

If $\|H^{ij}_\pi - H^{ij}_e\|_\infty$ is infinite, we set $\Delta H := H^{ij}_\pi - H^{ij}_e$. Repeating the argument in the previous proof shows $\|\Delta H_r\|_1 \leq 2r\alpha_n(k|\mathbb{G})$ and $\|\overline{\Delta H_r}\|_1 \leq 2r^{-\frac{\varepsilon}{2}}$ for any $r>0$, and hence $\|\mathbb{E}[H^{ij}_\pi - H^{ij}_e|\mathbb{G},Y]\|_1 \leq 4\alpha_n(k|\mathbb{G})^{\frac{\epsilon}{2+\epsilon}}$. $\square$

The next result is used to relate mixing to the growth of volume under the metric d. We phrase this in terms of a generic function g, which is later chosen as $t\mapsto\alpha(t|\mathbb{G})^{\frac{\epsilon}{2+\epsilon}}$ in the basic case, and $t\mapsto\alpha_n(t|\mathbb{G})^{\frac{\epsilon}{2+\epsilon}}$ in the general case.

LEMMA 5. *Let $g:[0,\infty)\to[0,\infty)$ be a measurable function. Then*

$$\frac{\sum_{i\geq m}|\mathbf{B}_{i+1}\setminus\mathbf{B}_i|g(i)}{\int_{\mathbb{G}\setminus\mathbf{B}_{m-1}} g(d(e,\phi))|d\phi|} < \infty \qquad \textit{for all } m\in\mathbb{N}\;.$$

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

$$\textstyle\sum_{i\geq m}|\mathbf{B}_{i+1}\setminus\mathbf{B}_i|g(i) \leq r\sum_{i\geq m}|\mathbf{B}_i\setminus\mathbf{B}_{i-1}|g(i) \leq r\displaystyle\int_{\mathbb{G}\setminus\mathbf{B}_{m-1}} g(d(e,\phi))|d\phi|\;,$$

where we have used (11). $\square$

Finally, we note that assuming $\mathbb{E}[f(X)|\mathbb{G}]=0$ incurs no loss of generality:

LEMMA 6 (Conditional centering). *Let X be $\mathbb{G}$-invariant, and $p\geq 1$. For any $g\in\mathbf{L}_p(X)$, the random function $f(\bullet):=g(\bullet)-\mathbb{E}[g(X)|\mathbb{G}]$ is $\sigma(\mathbb{G})$-measurable random element of $\mathbf{L}_p(X)$. For all $n\in\mathbb{N}$,*

$$\mathbb{F}_n(f,X) = \overline{\mathbb{F}}_n(g,X) \quad \textit{and} \quad \alpha_f(n|\mathbb{G}) = \alpha_g(n|\mathbb{G}) \quad \textit{almost surely,}$$

where $\alpha_\bullet(n|\mathbb{G})$ is the conditional mixing coefficient defined by $(\bullet, X)$.

PROOF. For $p\geq 1$, $\mathbf{L}_p$-norms contract under conditioning [5]. That makes f a $\sigma(\mathbb{G})$-measurable random element of $\mathbf{L}_p(X)$. Since $f(\phi X) = g(\phi X) - \mathbb{E}[g(X)|\mathbb{G}]$ for any $\phi\in\mathbb{G}$, we have $\mathbb{F}_n(f,X) = \overline{\mathbb{F}}_n(g,X)$. To prove the second claim, consider events $A\in\sigma_f(\{\phi_1,\phi_2\})$ and $B\in\sigma_f(G)$, for any $G\subset\mathbb{G}$ and $\phi_1,\phi_2\in\mathbb{G}\setminus\mathbf{B}_t(G)$. Fix any $\delta>0$. By definition of σ_f, we can choose sets $S_i\in\sigma_g(\phi_1,\phi_2)$, sets $T_i\in\sigma(\mathbb{G})$, and constants $c_i\in[0,1]$ such that $\|\sum_i c_i\mathbb{I}(S_i,T_i) - \mathbb{I}(A)\|_1 \leq \delta$. As the sets T_i are in $\sigma(\mathbb{G})$, we have

$$\begin{aligned}&\|\textstyle\sum_i c_i\big(\mathbb{P}(S_i,T_i,B|\mathbb{G}) - P(S_i,T_i|\mathbb{G})P(B|\mathbb{G})\big)\|_1\\ &= \|\textstyle\sum_i c_i\mathbb{I}(T_i)\big(\mathbb{P}(S_i,B|\mathbb{G}) - P(S_i|\mathbb{G})P(B|\mathbb{G})\big]\|_1 \leq \alpha(t|\mathbb{G})\;,\end{aligned}$$

where the final inequality uses the definition α and $c_i\in[0,1]$. As δ is arbitrary, this implies $\|P(A,B|\mathbb{G}) - P(A|\mathbb{G})P(B|\mathbb{G})\|_1 \leq \alpha(t|\mathbb{G})$. $\square$

APPENDIX B: PROOFS OF THE BASIC LIMIT THEOREMS

We first adapt the upper bound on d_{w} given by Stein's inequality to our problem in B.1, and then apply it to prove the limit theorems in B.2.

B.1. Bounds on the Wasserstein distance. By Lemma 6, it suffices to establish Theorems 4 and 6 for elements f of

$$\overline{\mathbf{L}}_p(X,\mathbb{G}) := \{ f(\bullet) = g(\bullet) - \mathbb{E}[f(X)|\mathbb{G}] \,|\, g \in \mathbf{L}_p(X) \} .$$

Given $f \in \overline{\mathbf{L}}_p(X,\mathbb{G})$, we choose the variable W in Stein's inequality as

$$W \;:=\; \frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{F}_n(f,X) \;=\; \frac{1}{\eta_n}\int_{\mathbf{A}_n} f(\phi X)|d\phi| \qquad \text{where } \eta_n := \eta(n)\sqrt{|\mathbf{A}_n|}\,.$$

Here $\eta(n)$ is for now any positive, $\sigma(\mathbb{G})$-measurable random variable, but will be chosen in the next section as a specific approximation to the asymptotic variance. For a fixed element $\phi \in \mathbb{G}$, conditional mixing allows us to control dependence for elements ϕ' far away from ϕ. To treat terms close to ϕ separately, we choose $b > 0$, and decompose W into long-range and short-range contributions,

$$W_b^\phi := \frac{1}{\eta_n}\int_{\mathbf{A}_n} \mathbb{I}\{d(\phi,\phi') \geq b\} f(\phi' X)|d\phi'| \qquad \text{and} \qquad \Delta_b^\phi := W - W_b^\phi\,.$$

For our purposes, Stein's inequality then takes the following form.

LEMMA 7. *Assume the conditions of Theorem 4, and define W as above, for a $\sigma(\mathbb{G})$-measurable random element $\eta(n)$ of $(0,\infty)$. Then*

$$\begin{aligned}
\mathbb{E}[d_{\mathrm{w}}(W,Z|\mathbb{G})] \;\leq\; & \sup_{t\in\mathcal{F}} \Big\| \mathbb{E}\big[\frac{1}{\eta_n}\int_{\mathbf{A}_n} f(\phi X) t(W_b^\phi)|d\phi|\big|\mathbb{G}\big] \|_1 \\
& + \sup_{t\in\mathcal{F}} \Big\| \mathbb{E}\big[\frac{1}{\eta_n}\int_{\mathbf{A}_n} f(\phi X)(t(W) - t(W_b^\phi) - \Delta_b^\phi t'(W))|d\phi|\big|\mathbb{G}\big] \Big\|_1 \qquad (6)\\
& + \sqrt{\frac{2}{\pi}} \Big\| 1 - \frac{1}{\eta_n}\mathbb{E}\big[\int_{\mathbf{A}_n} f(\phi X)\Delta_b^\phi |d\phi| \,\big|\mathbb{G}\big] \Big\|_1 \\
& + \sqrt{\frac{2}{\pi}} \Big\| \frac{1}{\eta_n}\int_{\mathbf{A}_n} f(\phi X)\Delta_b^\phi - \mathbb{E}[f(\phi X)\Delta_b^\phi|\mathbb{G}]|d\phi| \Big\|_1 \\
& =: \text{(a)} + \text{(b)} + \text{(c)} + \text{(d)}
\end{aligned}$$

where Z is a standard normal variable and $b > 0$.

PROOF. The triangle inequality yields

$$\begin{aligned}
& \|\mathbb{E}[Wt(W) - t'(W)|\mathbb{G}]\|_1 \\
& = \big\| \mathbb{E}\big[\int_{\mathbf{A}_n} \tfrac{f(\phi X)}{\eta_n}\big(t(W) - t(W_b^\phi) + t(W_b^\phi)\big) - t'(W)|d\phi|\big|\mathbb{G}\big] \big\|_1 \\
& \leq \big\| \mathbb{E}\big[\int_{\mathbf{A}_n} \frac{f(\phi X)}{\eta_n}(t(W) - t(W_b^\phi)) - t'(W)|d\phi|\big|\mathbb{G}\big] \big\|_1 \\
& + \big\| \mathbb{E}\big[\int_{\mathbf{A}_n} \frac{f(\phi X)}{\eta_n} t(W_b^\phi)|d\phi|\big|\mathbb{G}\big] \big\|_1 .
\end{aligned}$$

Using $t \in \mathcal{F}$, the first term can be bounded further as

$$
\begin{aligned}
&\big\|\mathbb{E}\big[\int_{\mathbf{A}_n} \frac{f(\phi X)(t(W)-t(W_b^\phi))}{\eta n} - t'(W)|d\phi|\big|\mathbb{G}\big]\big\|_1 \\
&\quad\leq \Big\|\mathbb{E}\big[\int_{\mathbf{A}_n} \tfrac{f(\phi X)(t(W)-t(W_b^\phi))-\Delta_b^\phi t'(W)}{\eta_n}|d\phi|\big|\mathbb{G}\big]\Big\|_1 \\
&\quad+ \Big\|\mathbb{E}\big[t'(W)\big(1-\int_{\mathbf{A}_n}\frac{f(\phi X)}{\eta n}\Delta_b^\phi|d\phi|\big)\big|\mathbb{G}\big]\Big\|_1 \\
&\quad\leq \Big\|\mathbb{E}\big[\int_{\mathbf{A}_n} \frac{f(\phi X)(t(W)-t(W_b^\phi))-\Delta_b^\phi t'(W)}{\eta n}|d\phi|\big|\mathbb{G}\big]\Big\|_1 \\
&\quad+ \sqrt{\frac{2}{\pi}}\big\|1-\frac{\mathbb{E}[\int_{\mathbf{A}_n} f(\phi X)\Delta_b^\phi|d\phi||\mathbb{G}]}{\eta n}\big\|_1 \\
&\quad+ \sqrt{\frac{2}{\pi}}\big\|\frac{1}{\eta n}\int_{\mathbf{A}_n} f(\phi X)\Delta_b^\phi - \mathbb{E}[f(\phi X)\Delta_b^\phi|\mathbb{G}]|d\phi|\big\|_1 .
\end{aligned}
$$

Substituting into the right-hand side of (2) yields the result. □

The main work of the proof is to control the terms (a)–(d) in Lemma 7. To handle large values of f, we split the function in its range, into

(7) $$f^{<\gamma}(x) := f(x)\mathbb{I}\{|f(x)|<\gamma\} \text{ and } f^{\geq\gamma}(x) := f(x)\mathbb{I}\{|f(x)|\geq\gamma\} .$$

The next result refines the terms (a)–(d) using Lemma 3, and by handling $f^{<\gamma}$ and $f^{\geq\gamma}$ separately.

LEMMA 8. *Require the assumptions of Lemma 7. Fix $b>0$ and $\gamma>0$, and let τ be defined as in (24). Choose $p,q>0$ to be such that $\frac{1}{p}+\frac{1}{q}=1$. Then*

$$
\begin{aligned}
\|d(W,Z|\mathbb{G})\|_1 \leq\ & 4\big\|\frac{f(X)}{\eta(n)}\big\|_{2+\varepsilon}^2\tau(b) + 4|\mathbf{B}_b|\big\|\frac{f^{\geq\gamma}(X)}{\eta(n)}\big\|_{2+\varepsilon}\big\|\frac{f(X)}{\eta(n)}\big\|_{2+\varepsilon} \\
&+ \frac{8|\mathbf{B}_b|}{\sqrt{|\mathbf{A}_n|}}\big\|\frac{f(X)}{\eta(n)}\big\|_{2q(1+\varepsilon/2)}^2\big\|\frac{f^{<\gamma}(X)}{\eta(n)}\big\|_{p(1+\varepsilon/2)}\int_{\mathbb{G}}\alpha(d(e,\phi)|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}}d|\phi| \\
&+ \sqrt{2/\pi}\Big(\mathbb{E}\big[\big|\frac{\eta(n)^2-\eta_b^2}{\eta(n)^2}\big|\big] + \big\|\frac{f(X)}{\eta(n)}\big\|_2^2\frac{|\mathbf{A}_n \vartriangle \mathbf{B}_b\mathbf{A}_n|}{|\mathbf{A}_n|}\Big) \\
&+ 4\frac{|\mathbf{B}_b|}{\sqrt{|\mathbf{A}_n|}}\big\|\frac{f^{<\gamma}(X)}{\eta(n)}\big\|_{4+2\varepsilon}^2\big(\int_{\mathbb{G}}\alpha(d(e,\phi)|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}}|d\phi|\big)^{\frac{1}{2}} .
\end{aligned}
$$

PROOF. To bound (a), fix any $\delta>0$. Then

(8) $$\big\|\mathbb{E}\big[\frac{1}{\eta n}f(\phi X)t(W_b^\phi)\big|\mathbb{G}\big]\big\|_1 \leq \sum_{j\geq\lfloor|\mathbf{B}_b|/\delta\rfloor}\Big\|\mathbb{E}\Big[f(\phi X)\frac{t(W_{j\delta}^\phi)-t(W_{(j+1)\delta}^\phi)}{\eta n}\Big|\mathbb{G}\Big]\Big\|_1$$

An application of Lemma 3 to the summand gives

$$\Big\|\mathbb{E}\Big[\frac{f(\phi X)(t(W_{j\delta}^\phi)-t(W_{(j+1)\delta}^\phi))}{\eta n}\Big|\mathbb{G}\Big]\Big\|_1 \leq 4\Big\|\frac{f(\phi X)(t(W_{j\delta}^\phi)-t(W_{(j+1)\delta}^\phi))}{\eta n}\Big\|_{\frac{2+\varepsilon}{2}}\alpha(j\delta|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}} .$$

By Hölder's inequality,

$$\Big\|\frac{f(\phi X)(t(W_{j\delta}^\phi)-t(W_{(j+1)\delta}^\phi))}{\eta(n)}\Big\|_{\frac{2+\varepsilon}{2}} \leq \Big\|\frac{f(X)}{\eta(n)}\Big\|_{2+\varepsilon}\big\|t(W_{j\delta}^\phi)-t(W_{(j+1)\delta}^\phi)\big\|_{2+\varepsilon}$$

and since t is Lipschitz, $\|t(W^{\phi}_{j\delta}) - t(W^{\phi}_{(j+1)\delta})\|_{2+\varepsilon} \leq \|W^{\phi}_{j\delta} - W^{\phi}_{(j+1)\delta}\|_{2+\varepsilon}$. In summary, the right-hand side of (8) is bounded by

$$\text{rhs (8)} \leq 4\sqrt{\frac{2}{\pi|\mathbf{A}_n|}}\sum_{j\geq\lfloor|\mathbf{B}_b|/\delta\rfloor}\Big\|\frac{f(X)}{\eta(n)}\Big\|_{2+\varepsilon}\big\|W^{\phi}_{j\delta} - W^{\phi}_{(j+1)\delta}\big\|_{2+\varepsilon}\alpha(j\delta|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}}\,.$$

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

$$\text{(a)} \;\leq\; 4\Big\|\frac{f(X)}{\eta(n)}\Big\|^2_{2+\varepsilon}\int_{\mathbb{G}\setminus\mathbf{B}_b}\alpha(d(e,\phi)|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}}|d\phi| \;=\; 4\Big\|\frac{f(X)}{\eta(n)}\Big\|^2_{2+\varepsilon}\tau(b)\,,$$

For (b), we decompose $f = f^{<\gamma} + f^{\geq\gamma}$. The triangle inequality gives

$$\begin{aligned}
&\Big\|\mathbb{E}\big[\int_{\mathbf{A}_n} f(\phi X)\frac{t(W) - t(W^{\phi}_b) - \Delta^{\phi}_b t'(W)}{\eta n}|d\phi|\big|\mathbb{G}\big]\Big\|_1\\
\leq\;&\Big\|\mathbb{E}\big[\int_{\mathbf{A}_n} f^{\geq\gamma}(\phi X)\frac{t(W) - t(W^{\phi}_b) - \Delta^{\phi}_b t'(W)}{\eta n}|d\phi|\big|\mathbb{G}\big]\Big\|_1\\
+\;&\Big\|\mathbb{E}\big[\int_{\mathbf{A}_n} f^{<\gamma}(\phi X)\frac{t(W) - t(W^{\phi}_b) - \Delta^{\phi}_b t'(W)}{\eta n}|d\phi|\big|\mathbb{G}\big]\Big\|_1 =: \text{(b1)} + \text{(b2)}\,.
\end{aligned}$$

Since t is an element of $\mathcal{F}$, it satisfies

$$|t(x+y) - t(x) - yt'(x)| \leq 2|y|\sup_{z\in[x,x+y]}|t'(z)| \qquad \text{for } x, y \in \mathbb{R} \tag{9}$$

and $\sup|t'(z)| \leq \sqrt{2/\pi} \leq 1$. Choosing $y = \Delta^{\phi}_b$ yields

$$\begin{aligned}
\text{(b1)} &\leq 2\Big\|\mathbb{E}\big[\frac{1}{\eta n}\int_{\mathbf{A}_n}|f^{\geq\gamma}(\phi X)||\Delta^{\phi}_b|\;|d\phi|\big|\mathbb{G}\big]\Big\|_1\\
&\leq 2\Big\|\frac{f^{\geq\gamma}(X)}{\eta(n)}\Big\|_{2+\varepsilon}\Big\|\frac{f(X)}{\eta(n)}\Big\|_{2+\varepsilon}\frac{\int_{\mathbf{A}_n^2}\mathbb{I}\{d(\phi,\phi')\leq b\}|d\phi||d\phi'|}{|\mathbf{A}_n|}\\
&\leq 2|\mathbf{B}_b|\Big\|\frac{f^{\geq\gamma}(X)}{\eta(n)}\Big\|_{2+\varepsilon}\Big\|\frac{f(X)}{\eta(n)}\Big\|_{2+\varepsilon}
\end{aligned}$$

For (b2), fix $p, q > 0$ with $1/p + 1/q = 1$. A Taylor expansion gives

$$|t(W) - t(W^{\phi}_b) - \Delta^{\phi}_b t'(W)| \;\leq\; \frac{1}{2}\sup_w|t''(w)|(\Delta^{\phi}_b)^2 \;\leq\; (\Delta^{\phi}_b)^2\,.$$

Substituting $(\Delta^{\phi}_b)^2 = (\frac{1}{\eta_n}\int_{\mathbf{A}_n}\mathbb{I}\{d(\phi,\phi')\leq b\}f(\phi' X)|d\phi'|)^2$ into (b2) yields

$$\begin{aligned}
\text{(b2)} &\leq \Big\|\int_{\mathbf{A}_n^3}\frac{\mathbb{E}\big[f^{<\gamma}(\phi X)\mathbb{I}\{d(\phi,\psi), d(\phi,\pi)\leq b\}f(\psi X)f(\pi X)|\mathbb{G}\big]}{\eta_n^3}|d\phi||d\psi||d\pi|\Big\|_1\\
&\leq \frac{8|\mathbf{B}_b|}{\sqrt{|\mathbf{A}_n|}}\big\|\tfrac{f(X)}{\eta(n)}\big\|^2_{2q(1+\frac{\varepsilon}{2})}\big\|\frac{f^{<\gamma}(X)}{\eta(n)}\big\|_{p(1+\frac{\varepsilon}{2})}\int_{\mathbb{G}}\alpha^{\frac{\varepsilon}{2+\varepsilon}}(d(e,\phi)|\mathbb{G})d|\phi|\,.
\end{aligned}$$

To bound (c), write $\eta_b^2 := \int_{\phi\in\mathbf{B}_b}\eta^2(\phi)|d\phi|$ again apply the triangle inequality, which yields

$$\begin{aligned}
\text{(c)}\cdot\sqrt{\frac{\pi}{2}} &= \Big\|\frac{\eta(n)^2 - \int_{\mathbf{A}_n^2}\frac{1}{|\mathbf{A}_n|}\mathbb{E}[\mathbb{I}\{d(\phi,\phi')\leq b\}f(\phi X)f(\phi' X)|\mathbb{G}]|d\phi||d\phi'|}{\eta(n)^2}\Big\|_1\\
&\leq \mathbb{E}\Big[\Big|\frac{\eta(n)^2 - \eta_b^2}{\eta(n)^2}\Big|\Big] + \Big\|\frac{\eta_b^2 - \int_{\mathbf{A}_n^2}|\mathbf{A}_n|^{-1}\mathbb{E}[\mathbb{I}\{d(\phi,\phi')\leq b\}f(\phi X)f(\phi' X)|\mathbb{G}]|d\phi||d\phi'|}{\eta(n)^2}\Big\|_1\\
&\leq \mathbb{E}\Big[\Big|\frac{\eta(n)^2 - \eta_b^2}{\eta(n)^2}\Big|\Big] + \Big\|\frac{f(X)}{\eta(n)}\Big\|_2^2\frac{|\mathbf{A}_n\,\triangle\,\mathbf{B}_b\mathbf{A}_n|}{|\mathbf{A}_n|}.
\end{aligned}$$

For (d), we again use $f = f^{<\gamma} + f^{\geq\gamma}$ and the triangle inequality. For a pair (ϕ_1, ϕ_2) of group elements, abbreviate

$$F^{<\gamma}_{\phi_1\phi_2} := \frac{1}{\eta(n)^2}\big(f^{<\gamma}(\phi_1 X)f^{<\gamma}(\phi_2 X) - \mathbb{E}[f^{<\gamma}(\phi_1 X)f^{<\gamma}(\phi_2 X)|\mathbb{G}]\big)\,,$$

and define $F^{\geq\gamma}_{\phi_1\phi_2}$ as $F^{\leq\infty}_{\phi_1\phi_2} - F^{\leq\gamma}_{\phi_1\phi_2}$. For any quadruple $\phi_1, \ldots, \phi_4 \in \mathbb{G}$,

$$\big\|\text{Cov}[F^{<\gamma}_{\phi_1\phi_2}, F^{<\gamma}_{\phi_3\phi_4}|\mathbb{G}]\big\|_1 \leq 4\Big\|\frac{f^{\leq\gamma}(X)}{\eta(n)}\Big\|^4_{4+2\varepsilon}\alpha\big(d((\phi_1,\phi_2),(\phi_3,\phi_4))|\mathbb{G}\big)^{\frac{\varepsilon}{2+\varepsilon}}$$

holds by Lemma 3, which implies

$$\Big\|\int_{\mathbf{A}_n\times\mathbf{A}_n\mathbf{B}_b}F^{<\gamma}_{\phi_1\phi_2}\frac{|d\phi_1||d\phi_2|}{|\mathbf{A}_n|}\Big\|_1 \leq \frac{4|\mathbf{B}_b|}{\sqrt{|\mathbf{A}_n|}}\Big\|\frac{f^{<\gamma}(X)}{\eta(n)}\Big\|^2_{4+2\varepsilon}\big(\int_{\mathbb{G}}\alpha(d(e,\phi)|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}}|d\phi|\big)^{\frac{1}{2}}\,.$$

For $f^{\geq\gamma}$, we obtain

$$\Big\|\int_{\mathbf{A}_n\times\mathbf{A}_n\mathbf{B}_b}F^{\geq\gamma}_{\phi_1,\phi_2}\frac{|d\phi_1||d\phi_2|}{|\mathbf{A}_n|}\Big\|_1 \leq 2|\mathbf{B}_b|\Big\|\frac{f^{\geq\gamma}(X)}{\eta(n)}\Big\|_2\Big\|\frac{f(X)}{\eta(n)}\Big\|_2 =: \text{(d')}\,,$$

and hence

$$\text{(d)}\cdot\frac{\sqrt{\pi}}{\sqrt{2}} \leq 4\frac{|\mathbf{B}_b|}{\sqrt{|\mathbf{A}_n|}}\Big\|\frac{f^{<\gamma}(X)}{\eta(n)}\Big\|^2_{4+2\varepsilon}\big(\int_{\mathbb{G}}\alpha(d(e,\phi)|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}}|d\phi|\big)^{\frac{1}{2}} + \text{(d')}\,.$$

Rearranging terms within (a)+(b)+(c)+(d) yields the statement. □

B.2. Proof of the limit theorems. We first prove the central limit theorem under hypothesis (21). The result under hypothesis (20), and the Berry-Esseen bound, then follow with minimal adjustments.

PROOF OF THEOREM 4 ASSUMING (21). Set $S_n := \sqrt{|\mathbf{A}_n|}\mathbb{F}_n(X)$, and let $Z \sim N(0,1)$ be independent of (X,η). We must show $S_n \xrightarrow{\text{d}} \eta Z$. By Lemma 3,

$$\begin{aligned}\|\eta^2\|_1 &\leq \int_{\mathbb{G}}\|\mathbb{E}[f(X)f(\phi X)|\mathbb{G}]\|_1|d\phi| \\ &\leq \|f(X)\|_{2+\epsilon}\textstyle\sum_{b\in\mathbb{N}}|\mathbf{B}_{b+1}\setminus\mathbf{B}_b|\alpha(b|\mathbb{G})^{\frac{\epsilon}{2+\epsilon}} < \infty\,,\end{aligned}$$

which shows $\eta < \infty$ almost surely. Since ηZ and $S_n := \sqrt{|\mathbf{A}_n|}\mathbb{F}_n(X)$ have first moments, $S_n \xrightarrow{\text{d}} \eta Z$ holds if $d_{\text{w}}(S_n, \eta Z) \to 0$, as $n \to \infty$.

To show that is the case, we may assume $f \in \overline{\mathbf{L}}_1(X)$, by Lemma 6. We first choose suitable sequences (γ_n) and (b_n). By definition, $|\mathbf{A}_n| \to \infty$. Set $\gamma_n := |\mathbf{A}_n|^{1/6}$. That implies $\gamma_n \to \infty$, and hence $\|f^{\geq\gamma_n}(X)\|_{2+\epsilon} \to 0$. Since $|\mathbf{A}_n|$ diverges, we can choose a divergent sequence (b_n) such that

$$|\mathbf{B}_{b_n}| \leq |\mathbf{A}_n|^{1/12}, \quad |\mathbf{B}_{b_n}|\|f^{\geq\gamma_n}(X)\|_2 \qquad \text{and} \qquad \frac{|\mathbf{A}_n \vartriangle \mathbf{B}_{b_n}\mathbf{A}_n|}{|\mathbf{A}_n|} \to 0\,.$$

Collecting terms in Lemma 8, we then have

$$r_n := \frac{|\mathbf{B}_{b_n}|\gamma_n^2}{\sqrt{|\mathbf{A}_n|}} + |\mathbf{B}_{b_n}|\|f^{\geq\gamma_n}(X)\|_2 \to 0 \quad \text{and} \quad \tilde{r}_n := \frac{|\mathbf{A}_n \vartriangle \mathbf{B}_{b_n}\mathbf{A}_n|}{|\mathbf{A}_n|} \to 0\,.$$

The next step is to construct $\eta(n)$ in Lemma 8 as an approximation to η. To this end, we set $u_n := \max\{r_n, \tilde{r}_n, \tau(b_n)\}^{1/8}$ and $v_n := \max\{r_n, \tilde{r}_n, \tau(b_n)\}^{-1/2}$. As $n \to \infty$, we hence have $u_n \to 0$ and $v_n \to \infty$, and observe that

(10) (i) $u_n < v_n$ eventually (ii) $\dfrac{v_n}{u_n^3}\big(r_n + \tilde{r}_n + \tau(b_n)\big) \to 0$ (iii) $v_n P(\eta < u_n) \to 0\,.$

Set $\eta(n) := \eta\mathbb{I}\{\eta \in [u_n, v_n]\} + u_n\mathbb{I}\{\eta \notin [u_n, v_n]\}$, and note that $\eta(n) \perp\!\!\!\perp Z$. Then

$$\begin{aligned} d_{\mathrm{w}}(S_n, \eta Z) \;&\leq\; d_{\mathrm{w}}(S_n, \eta(n)Z) + d_{\mathrm{w}}(\eta(n)Z, \eta Z) \\ &\leq\; d_{\mathrm{w}}(S_n, \eta(n)Z) + \|Z\|_1 \|(\eta - u_n)\mathbb{I}\{\eta \notin [u_n, v_n]\}\|_1 \,. \end{aligned}$$

Since we have already shown $\|\eta^2\|_1 < \infty$, the last term satisfies

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

It thus suffices to show $d_{\mathrm{w}}(S_n, \eta(n)Z) \to 0$. Using the Markov inequality we note that

$$P\big(\eta \notin [u_n, v_n]\big) \leq P\big(\eta < u_n\big) + \frac{\|\eta^2\|_1}{v_n^2}. \tag{11}$$

Using Lemma 2 with $Y = \frac{1}{\eta(n)}$,

$$d_{\mathrm{w}}(S_n, \eta(n)Z) \;\leq\; v_n \mathbb{E}\big[d_{\mathrm{w}}\big(\tfrac{S_n}{\eta(n)}, Z \big| \mathbb{G}\big)\big] \,,$$

since $1/\eta(n) \geq 1/v_n$. Substituting $W = \frac{S_n}{\eta(n)}$ into Lemma 8 gives

$$\begin{aligned} v_n \mathbb{E}\big[d_{\mathrm{w}}\big(\tfrac{S_n}{\eta(n)}, Z\big|\mathbb{G}\big)\big] \leq{}& \frac{v_n}{u_n^2}\Big(5\big\|f(X)\big\|_{2+\varepsilon}^2 \tau(b_n) \\ &+ 4|\mathbf{B}_{b_n}| \|f^{\geq\gamma_n}(X)\|_{2+\varepsilon} \big\|f(X)\big\|_{2+\varepsilon} + \frac{8|\mathbf{B}_{b_n}| \big\|f(X)\big\|_{2+\varepsilon}^2 \gamma_n \tau(0)}{u_n\sqrt{|\mathbf{A}_n|}} \\ &+ \sqrt{2/\pi}\big(u_n^2 P(\eta \notin [u_n, v_n]) + \big\|f(X)\big\|_2^2 \tilde{r}_n\big) + 4\frac{|\mathbf{B}_{b_n}|\gamma_n^2\sqrt{\tau(0)}}{\sqrt{|\mathbf{A}_n|}}\Big) \\ \leq{}& \frac{8v_n}{\min(u_n^3, 1)}\Big(\big\|f(X)\big\|_{2+\varepsilon}^2 \tau(b_n) + \max(\big\|f(X)\big\|_{2+\varepsilon}^2 \tau(0), 1)[r_n + \tilde{r}_n]\Big) \\ &+ v_n P(\eta < u_n) + \frac{\|\eta^2\|_1}{v_n} \,. \end{aligned}$$

This final bound vanishes as $n \to \infty$: The first term by (10), the second since $u_n \to 0$ and $v_n \to \infty$. That shows $d_{\mathrm{w}}(S_n, \eta(n)Z) \to 0$, which implies $d_{\mathrm{w}}(S_n, \eta Z) \to 0$ and completes the proof. □

Since the Berry-Esseen bound assumes a third and fourth moment, it can be proven by applying Lemma 8 directly:

PROOF OF THEOREM 6. The sequence (b_n) is given by hypothesis. Fix any divergent sequence (γ_n) in $(0, \infty)$. For each γ_n,

$$\|f(X)\mathbb{I}\{|f(X) \leq \gamma_n|\}\|_{3(1+\frac{\varepsilon}{2})} \;\leq\; \|f(X)\|_{3(1+\frac{\varepsilon}{2})} \,.$$

We can hence apply Lemma 8 with $p = \frac{3}{2}$ and $q = 3$, and Theorem 6 follows for $n \to \infty$. □

PROOF OF THEOREM 4 ASSUMING (20). There is a finite $k \in \mathbb{N}$ such that $\alpha(k|\mathbb{G}) = 0$. We can hence repeat the argument in the above for $b_1 = b_2 = \ldots := k$ and $\varepsilon = 0$, which again yields $d_w(S_n, \eta(n)Z) \to 0$ for $n \to \infty$.

□

B.3. Derivation of the confidence interval. We now prove Theorem 5. Using the centered average $\overline{\mathbb{F}}_n$, the statement of the theorem can be phrased as: Under the conditions of Theorem 4, and assuming $\hat{\eta}_n$ is defined using a suitable sequence (b_n),

$$\limsup_{n\to\infty} P\Big(\big|\overline{\mathbb{F}}_n(f,X)\big| > \frac{\hat{\eta}_n}{\sqrt{|\mathbf{A}_n|}} z_{1-\frac{\alpha}{2}}\Big) \leq \alpha\,.$$

Since we assume the hypothesis of Theorem 4, we can reuse part of its proof: The scaled average S_n in that proof assumed $\mathbb{E}[f(X)|\mathbb{G}]=0$. Since we now use $\overline{\mathbb{F}}_n$, we have $S_n=\sqrt{\mathbf{A}_n}\,\overline{\mathbb{F}}_n(f,X)$. We hence already know that $d_{\mathrm{w}}(S_n,\eta Z)\to 0$ for the asymptotic variance η, and the key to obtaining a confidence interval is to show that this also implies

$$d_{\mathrm{w}}(S_n,\hat{\eta}_n Z)\to 0 \tag{12}$$

for the empirical variance $\hat{\eta}_n$. The proof has three steps:

1. We first show that (12) holds if $\|\eta^2-\hat{\eta}_n^2\|_1\to 0$.
2. The main technical work is then to show $\|\eta^2-\hat{\eta}_n^2\|_1\to 0$, which we do using similar arguments as the proof of the central limit theorem.
3. Given (12), we deduce the result.

PROOF OF THEOREM 5. *Step 1*. Since $\hat{\eta}_n$ is, by its definition, independent of Z, the triangle inequality shows

$$d_{\mathrm{w}}(S_n,\hat{\eta}_n Z)\leq d_{\mathrm{w}}(S_n,\eta Z)+d_{\mathrm{w}}(\eta Z,\hat{\eta}_n Z)\leq d_{\mathrm{w}}(S_n,\eta Z)+\|\eta-\hat{\eta}_n\|_1\|Z\|_1\,.$$

Since $d_{\mathrm{w}}(S_n,\eta Z)\to 0$, (12) holds if $\|\eta-\hat{\eta}_n\|_1\to 0$. For any $\epsilon>0$, we have

$$\begin{aligned}\|\eta-\hat{\eta}_n\|_1 &\leq \|(\eta-\hat{\eta}_n)\mathbb{I}(\max(\eta,\hat{\eta}_n\leq\epsilon)\|_1+\|(\eta-\hat{\eta}_n)\mathbb{I}(\max(\eta,\hat{\eta}_n)>\epsilon)\|_1\\ &\leq 2\epsilon+\frac{\|\eta^2-\hat{\eta}_n^2\|_1}{\epsilon}\,,\end{aligned}$$

so it suffices to establish $\|\eta^2-\hat{\eta}_n^2\|_1\to 0$.

Step 2. We first observe that, similarly to the bound of term (c) in the proof of Theorem 4,

$$\begin{aligned}&\Big\|\hat{\eta}_n^2-\frac{1}{|\mathbf{A}_n|}\int_{\mathbf{A}_n^2}\mathbb{I}(d(\phi,\phi')\leq b_n)f(\phi X)f(\phi'(X)|d\phi'||d\phi|+\Big(\frac{1}{|\mathbf{A}_n|}\int_{\mathbf{A}_n}f(\phi X)|d\phi|\Big)^2\Big\|_1\\ &\leq\|f(X)\|_2^2\frac{|\mathbf{A}_n\,\triangle\,\mathbf{A}_n\mathbf{B}_{b_n}|}{|\mathbf{A}_n|}\,.\end{aligned}$$

Applying the triangle inequality to $\|\eta^2-\hat{\eta}_n^2\|_1$ hence gives

$$\begin{aligned}\|\eta^2-\hat{\eta}_n^2\|_1\leq&\Big\|\eta^2-\int_{\mathbf{A}_n^2}\frac{\mathbb{I}\{d(\phi,\phi')\leq b_n\}}{|\mathbf{A}_n|}\mathrm{Cov}[\mathrm{f}(\phi\mathrm{X}),\mathrm{f}(\phi'\mathrm{X})|\mathbb{G}]|\mathrm{d}\phi||\mathrm{d}\phi'|\Big\|_1\\ &+\Big\|\frac{1}{|\mathbf{A}_n|}\int_{\mathbf{A}_n^2}\mathbb{I}(d(\phi,\phi')\leq b_n)\big[f(\phi X)f(\phi'(X)-\mathbb{E}(f(\phi X)f(\phi'X)|\mathbb{G})\big]|d\phi'||d\phi|\Big\|_1\\ &+|\mathbf{B}_{b_n}|\Big\|\Big(\frac{1}{|\mathbf{A}_n|}\int_{\mathbf{A}_n}f(\phi X)|d\phi|\Big)^2-\mathbb{E}(f(X)|\mathbb{G})^2\Big\|_{L_2}\\ &+\|f(X)\|_2^2\frac{|\mathbf{A}_n\,\triangle\,\mathbf{A}_n\mathbf{B}_{b_n}|}{|\mathbf{A}_n|}\\ &=:\text{(a)}+\text{(b)}+\text{(c)}+\|f(X)\|_2^2\frac{|\mathbf{A}_n\,\triangle\,\mathbf{A}_n\mathbf{B}_{b_n}|}{|\mathbf{A}_n|}\end{aligned}$$

Since $|\mathbf{A}_n \triangle \mathbf{A}_n\mathbf{B}_{b_n}|/|\mathbf{A}_n| \to 0$ as $n \to \infty$ by hypothesis (iii), the last term vanishes asymptotically. We bound each of the remaining terms individually. For term (a), abbreviate $\eta_b^2 := \int_{\phi\in\mathbf{B}_b} \eta^2(\phi)|d\phi|$. Then

$$\text{(a)} \leq \mathbb{E}\big[\big|\eta^2 - \eta_{b_n}^2\big|\big] + \Big\|\eta_{b_n}^2 - \int_{\mathbf{A}_n^2} |\mathbf{A}_n|^{-1}\mathbb{I}\{d(\phi,\phi') \leq b_n\}\mathrm{Cov}[\mathrm{f}(\phi\mathrm{X}),\mathrm{f}(\phi'\mathrm{X})|\mathbb{G}]|\mathrm{d}\phi||\mathrm{d}\phi'|\Big\|_1$$

$$\leq 4\|f(X)\|_{2+\epsilon}^2\tau(b_n) + 4\big\|f(X)\big\|_2^2\frac{|\mathbf{A}_n \triangle \mathbf{A}_n\mathbf{B}_{b_n}|}{|\mathbf{A}_n|}.$$

Since $\tau(b_n) \to 0$ by hypothesis (i), that implies $(a) \to 0$ as $n \to \infty$. Term (c) satisfies

$$\begin{aligned}\text{(c)} &\leq 4\|f(X)\|_2|\mathbf{B}_{b_n}|\Big\|\frac{1}{|\mathbf{A}_n|}\int_{\mathbf{A}_n} f(\phi X) - \mathbb{E}(f(X)|\mathbb{G})|d\phi|\Big\|_2 \\ &\leq \frac{4\|f(X)\|_2\|f(X)\|_{2+\epsilon}\tau(0)|\mathbf{B}_{b_n}|}{\sqrt{|\mathbf{A}_n|}}\end{aligned}$$

and hence (c) $\to 0$ as $n \to \infty$, by hypothesis (ii).

For term (b), we have to argue similarly as in the proof of Theorem 4: Since we do not assume a fourth moment exists, we must split f. To this end, define $f^{<\gamma}$ and $f^{\geq\gamma}$ as in (7), and choose a sequence (γ_n) of positive scalars satisfying

$$\gamma_n \to \infty \qquad \text{and} \qquad \gamma_n^2\frac{|\mathbf{B}_{b_n}|}{\sqrt{|\mathbf{A}_n|}} \to 0 \qquad \text{as } n \to \infty\,.$$

For a pair $\phi_1, \phi_2 \in \mathbb{G}$, abbreviate

$$F_{\phi_1\phi_2}^{<\gamma_n} := \big(f^{<\gamma_n}(\phi_1 X)f^{<\gamma_n}(\phi_2 X) - \mathbb{E}[f^{<\gamma_n}(\phi_1 X)f^{<\gamma_n}(\phi_2 X)|\mathbb{G}]\big)\,.$$

For any two pairs (ϕ_1, ϕ_2) and (ϕ_3, ϕ_4), we then have

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

by the conditional mixing bound in Lemma 3. That implies

$$\begin{aligned}\Big\|\int_{\mathbf{A}_n\times\mathbf{A}_n\mathbf{B}_{b_n}} F_{\phi_1\phi_2}^{<\gamma_n}\frac{|d\phi_1||d\phi_2|}{|\mathbf{A}_n|}\Big\|_1 &\leq \frac{4|\mathbf{B}_{b_n}|}{\sqrt{|\mathbf{A}_n|}}\big\|f^{<\gamma_n}(X)\big\|_{4+2\varepsilon}^2\Big(\int_{\mathbb{G}}\alpha(d(e,\phi)|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}}|d\phi|\Big)^{\frac{1}{2}} \\ &\leq \frac{4|\mathbf{B}_{b_n}|\gamma_n^2}{\sqrt{|\mathbf{A}_n|}}\Big(\int_{\mathbb{G}}\alpha(d(e,\phi)|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}}|d\phi|\Big)^{\frac{1}{2}}\,.\end{aligned}$$

The residual term $F_{\phi_1\phi_2}^{\leq\infty} - F_{\phi_1\phi_2}^{\leq\gamma_n}$ satisfies

$$\Big\|\int_{\mathbf{A}_n\times\mathbf{A}_n\mathbf{B}_{b_n}}(F_{\phi_1\phi_2}^{\leq\infty} - F_{\phi_1\phi_2}^{\leq\gamma_n})\tfrac{|d\phi_1||d\phi_2|}{|\mathbf{A}_n|}\Big\|_1 \leq 2|\mathbf{B}_{b_n}|\big\|f^{\geq\gamma_n}(X)\big\|_2\big\|f(X)\big\|_2\,,$$

and combining the two shows

$$\text{(b)} \leq 4\frac{|\mathbf{B}_{b_n}|}{\sqrt{|\mathbf{A}_n|}}\gamma_n^2\Big(\int_{\mathbb{G}}\alpha(d(e,\phi)|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}}|d\phi|\Big)^{\frac{1}{2}} + 2|\mathbf{B}_{b_n}|\big\|f^{\geq\gamma_n}(X)\big\|_2\big\|f(X)\big\|_2 \to 0.$$

In summary, (a) + (b) + (c) $\to 0$ holds as $n \to \infty$. That implies $\|\eta^2 - \hat{\eta}_n^2\|_1 \to 0$, and we have established (12).

Step 3. By definition of the Wasserstein metric d_{w}, we can find a sequence of couplings $(S_n, \hat{\eta}_n Z)$ that satisfy $\|S_n - \hat{\eta}_n Z\|_1 \leq 2d_{\mathrm{w}}(S_n, \hat{\eta}_n Z)$. The hypothesis $P(\eta < t) \to 0$ for $t \searrow 0$ implies there is a sequence (t_n) of positive reals that satisfies

$$t_n \xrightarrow{n\to\infty} 0 \quad \text{and} \quad \|\eta - \hat{\eta}_n\|_1 = o(t_n) \quad \text{and} \quad d_{\mathrm{w}}\big(S_n,\ \hat{\eta}_n Z\big) = o\big(t_n\sqrt{P(\hat{\eta}_n < t_n)}\big)\,.$$

The truncated empirical variance $\tilde{\eta}_n := \min(\hat{\eta}_n, t_n)$ then satisfies

$$\|\tilde{\eta}_n - \hat{\eta}_n\|_1 \;\leq\; t_n P(\hat{\eta}_n < t_n) \;\xrightarrow{n\to\infty}\; 0\,.$$

Since also

$$P(\hat{\eta}_n < t_n) \;\leq\; P(\eta < \frac{t_n}{2}) + P(|\eta - \hat{\eta}_n| \geq \frac{t_n}{2}) \;\leq\; P(\eta < \frac{t_n}{2}) + 2\frac{\|\eta - \hat{\eta}_n\|_1}{t_n} \;\xrightarrow{n\to\infty}\; 0\,,$$

it follows that $d_{\mathrm{w}}(S_n, \tilde{\eta}_n Z) \leq d_{\mathrm{w}}(S_n, \hat{\eta}_n Z) + t_n P(\hat{\eta}_n < t_n) \to 0$. To express probabilities in terms of $\tilde{\eta}_n$, we define a sequence (ϵ_n) as $\epsilon_n = t_n\sqrt{P(\hat{\eta}_n < t_n)}$. Then $\epsilon_n = o(t_n)$, and the triangle inequality shows

$$\begin{aligned}
&P\Big(\big|\overline{\mathbb{F}}_n(f,X)\big| > \frac{\hat{\eta}_n}{\sqrt{|\mathbf{A}_n|}} z_{1-\frac{\alpha}{2}}\Big)\\
&\leq P\Big(\big|\sqrt{|\mathbf{A}_n|}\overline{\mathbb{F}}_n(f,X)\big| > \tilde{\eta}_n z_{1-\frac{\alpha}{2}} - \epsilon_n\Big) + P\Big(\big|\hat{\eta}_n - \tilde{\eta}_n\big| z_{1-\frac{\alpha}{2}} > \epsilon_n\Big)\\
&\leq P\Big(\big|\tilde{\eta}_n Z\big| > \tilde{\eta}_n z_{1-\frac{\alpha}{2}} - 2\epsilon_n\Big) + P(\big|\sqrt{|\mathbf{A}_n|}\overline{\mathbb{F}}_n - \tilde{\eta}_n Z\big| > \epsilon_n) + P\Big(\big|\hat{\eta}_n - \tilde{\eta}_n\big| z_{1-\frac{\alpha}{2}} > \epsilon_n\Big)\\
&\leq \text{(a')} + \text{(b')} + \text{(c')}\,.
\end{aligned}$$

We again bound each term successively. By Markov's inequality,

$$\begin{aligned}
\text{(c')} \;=\; P\Big(\big|\hat{\eta}_n - \tilde{\eta}_n\big| z_{1-\frac{\alpha}{2}} > \epsilon_n\Big) \;\leq\; \frac{z_{1-\frac{\alpha}{2}}\|\hat{\eta}_n - \tilde{\eta}_n\|_1}{\epsilon_n} \;&\leq\; \frac{t_n P(\hat{\eta}_n < t_n) z_{1-\frac{\alpha}{2}}}{\epsilon_n}\\
&=\; z_{1-\frac{\alpha}{2}}\sqrt{P(\hat{\eta}_n < t_n)} \;\to\; 0\,.
\end{aligned}$$

Since the coupling $(S_n, \hat{\eta}_n Z)$ is chosen to satisfy $\|\sqrt{|\mathbf{A}_n|}\,\overline{\mathbb{F}}_n - \hat{\eta}_n Z\|_1 \leq 2d_{\mathrm{w}}(S_n, \hat{\eta}_n)$,

$$\begin{aligned}
\text{(b')} \;&=\; P(\big|\sqrt{|\mathbf{A}_n|}\overline{\mathbb{F}}_n(f,X) - \tilde{\eta}_n Z\big| > \epsilon_n) \;\leq\; \frac{\big\|\sqrt{|\mathbf{A}_n|}\overline{\mathbb{F}}_n(f,X) - \tilde{\eta}_n Z\big\|_1}{\epsilon_n}\\
&\leq\; \frac{\big\|\sqrt{|\mathbf{A}_n|}\overline{\mathbb{F}}_n(f,X) - \hat{\eta}_n Z\big\|_1 + \|\tilde{\eta}_n - \hat{\eta}_n\|_1}{\epsilon_n} \;\leq\; \frac{2d_{\mathrm{w}}(S_n, \hat{\eta}_n) + t_n P(\hat{\eta}_n < t_n)}{\epsilon_n} \to 0.
\end{aligned}$$

Finally,

$$\begin{aligned}
\text{(a')} \;=\; P\Big(\big|\tilde{\eta}_n Z\big| > \tilde{\eta}_n z_{1-\frac{\alpha}{2}} - 2\epsilon_n\Big) \;&=\; P\Big(\big|Z\big| > z_{1-\frac{\alpha}{2}} - \frac{2\epsilon_n}{\tilde{\eta}_n}\Big) \;\leq\; P\Big(\big|Z\big| > z_{1-\frac{\alpha}{2}} - \frac{2\epsilon_n}{t_n}\Big)\\
&\leq\; P\Big(\big|Z\big| > z_{1-\frac{\alpha}{2}} - 2\sqrt{P(\hat{\eta}_n < t_n)}\Big) \;=\; \alpha + o_n(1)\,.
\end{aligned}$$

Substituting the upper bounds on (a'), (b') and (c') into the bound above, we obtain

$$\limsup_n P\Big(\big|\overline{\mathbb{F}}_n(f,X)\big| > \frac{\hat{\eta}_n}{\sqrt{|\mathbf{A}_n|}} z_{1-\frac{\alpha}{2}}\Big) \leq \alpha\,,$$

which is the statement of the theorem. □

APPENDIX C: PROOFS OF THE GENERAL LIMIT THEOREMS

We next prove Theorems 10 and 11. Recall that the proof in the basic case adapts Stein's inequality in Lemma 7, bounds the constituent terms individually, and then deduces both limit theorems from this bound. The structure in the general case is similar: Lemma 10 below substitutes for Lemma 7, and the main work is again to upper-bound each term on its right-hand side, which we do in Sections C.3–C.6. The theorems are then deduced in Sections C.7 and C.8. Although the steps remain similar, the terms in the bounds change:

- The generalization of invariance to (27) makes the dependency neighborhoods (which above were balls of radius b_n around group elements) more complicated.
- The fact that k_n may grow with n complicates terms involving f_n. Their moments are handled using telescopic sums $\bar{h}_n^i$, defined below.
- Large values of f were previously controlled using $f(x)\mathbb{I}\{|f(x)|<\gamma\}$ and its remainder. Similar quantities now have to be phrased in terms of $\bar{h}_n^i$ and the coefficients $c_{i,p}$.
- Randomized averages have to be phrased in terms of μ_n, see the definitions of P_{μ_n} and $\mathbb{E}_{\mu_n}$ below.
- Since we have to control the influence of randomization by μ_n, spreading coefficients $\mathcal{S}^n$ or $\mathcal{S}_w^n$ appear in the bounds.
- Since we make no specific restrictions on how a group action may apply the entries of a vector $\boldsymbol{\phi}\in\mathbb{G}^{k_n}$, arguments that compare pairs of such vectors often have to compare all possible combinations of coordinates.

As a result, the bounds become lengthy, and we first introduce some additional notation to summarize quantities that occur frequently.

C.1. Notation. Recall that sequences (k_n) and (b_n) are given by hypothesis. In addition, we will use a non-decreasing integer sequence (k_n') with $k_n'\leq k_n$. In the proofs, the functions f_n always appear in a centered form, which is the (random) function

$$h_n(X_n):=f_n(X_n)-\mathbb{E}[f_n(X_n)|\mathbb{G}]\,.$$

We frequently have to restrict random measures to subsets. If μ is a random measure on $\mathbb{G}^{k_n}$ and A a measurable subset, write

$$P_\mu(\,\bullet\,|A):=\frac{\mu(\,\bullet\cap A)}{\mu(A)}$$

provided $\mu(A)>0$ almost surely. Since $P_\mu(\,\bullet\,|A)$ is almost surely a probability measure even if μ is not, the usual rules of conditioning apply and explain expressions such as $P_\mu(\,\bullet\,|A,Y)$ for a random quantity Y. If f is a measurable function on $\mathbb{G}^{k_n}$, set

$$\mathbb{E}_\mu[f(\boldsymbol{\phi})|A]:=\int f(\boldsymbol{\phi})P_\mu(d\boldsymbol{\phi}|A)=\frac{1}{\mu(A)}\int_A f(\boldsymbol{\phi})\mu(d\boldsymbol{\phi})\,.$$

The distance $d_{\mathrm{w}}(W_n,Z)$ in Stein's inequality is then applied to

$$W_n:=\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n}[h_n(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n}]\,.$$

Recall from the proof overview that Stein's method considers dependency neighborhoods around an index i. We generalize these to sets of coordinates of a vector $\boldsymbol{\phi}$ that are similar to $\psi\in\mathbb{G}$,

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

Two types of averages of h_n appear in the upper bounds on d_{w}. One holds entries outside a neighborhood $\mathcal{I}_{b,k}(\psi,\boldsymbol{\phi})$, of size $I:=|\mathcal{I}_{b,k}(\psi,\boldsymbol{\phi})|$, fixed,

$$\bar{h}_n^{\psi,b,k}(\boldsymbol{\phi}X_n):=\lim_{m\to\infty}\frac{1}{|\mathbf{A}_m|^I}\int_{\{\boldsymbol{\theta}\in\mathbf{A}_m^{k_n}|\theta_i=\phi_i\text{ for }i\not\in\mathcal{I}_{b,k}(\psi,\boldsymbol{\phi})\}}h_n(\boldsymbol{\theta}X_n)|d\boldsymbol{\theta}|^{\otimes I}\;.$$

The other appears in particular in the context of moments. It fixes the first k_n-i entries, and can be written as a telescopic sum

$$\bar{h}_n^i(\boldsymbol{\phi}X_n):=g_n^i(\boldsymbol{\phi}X_n)-g_n^{i+1}(\boldsymbol{\phi}X_n)$$

with summands

$$g_n^i(\boldsymbol{\phi}X_n):=\lim_{m\to\infty}\frac{1}{|\mathbf{A}_m|^i}\int_{\mathbf{A}_m^i}h((\phi_1,\dots,\phi_{k_n-i},\theta_1,\dots,\theta_i)X_n)|d\theta_1|\cdots|d\theta_i|\;.$$

Higher moments of $h_n/\eta(n)$ are controlled using a sequence (γ_n) with $\gamma_n\to\infty$. That leads to bounds involving the terms

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

More generally, for any function f_n on $\mathbf{X}_n$ and the coefficients $c_{i,p}$ defined in Section 5, we write

$$M_p(f_n):=\sup_{\boldsymbol{\phi}\in\mathbb{G}^{k_n}}\big\|\tfrac{f_n(\boldsymbol{\phi}X_n)}{\eta(n)}\big\|_p\quad\text{ and }\quad C_p(f_n):=\textstyle\sum_{i=1}^{\infty}c_{i,p}(f_n)\;.$$

Terms in the bounds that quantify the behavior of μ_n involve vectors $\boldsymbol{\phi}\in\mathbb{G}^{k_n}$ whose entries are “not too close” to each other. To this end, we write

$$\partial(\boldsymbol{\phi}):=\min_{i\neq j}d(\phi_i,\phi_j)\;.$$

In particular, we must consider $\mu_n^*(\bullet):=\mu_n(\bullet\cap\{\boldsymbol{\phi}|\partial(\boldsymbol{\phi})\geq b_n\})$. This is again a random measure on $\mathbb{G}^{k_n}$, with

$$P_{\mu_n^*}(\boldsymbol{\phi}\in\bullet|\mathbf{A}_n^{k_n})=\mathbb{E}_{\mu_n}\big[\mathbb{I}\{\boldsymbol{\phi}\in\bullet,\partial(\boldsymbol{\phi})\geq b_n\}\big|\mathbf{A}_n^{k_n}\big]\;.\tag{13}$$

Moments of μ_n are controlled using a sequence (β_n) with $\beta_n\to\infty$. They lead to rather complicated terms, which we encapsulate using the sets

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

In words, a random vector $\boldsymbol{\phi}$ is generated by P_{μ_n}, conditionally on its entries not being too similar (hence $P_{\mu_n^*}$), and the set contains those vectors $\boldsymbol{\psi}$ unlikely to have an entry similar to ϕ_i. Finally, for a strongly well-spread sequence, the spreading coefficient $\mathcal{S}^n$ was defined in Section 5. A similar coefficients in the well-spread case is

$$\mathcal{S}_w^n:=\sup_{A\in\Sigma_n,n\in\mathbb{N}}\mathbb{E}\Big[\frac{1}{\mathbb{T}_n(A,|\bullet|^{\otimes k_n})}P_{\mu_n\otimes\mu_n}\big((\boldsymbol{\phi},\boldsymbol{\phi}')\in A\big|\mathbf{A}_n^{2k_n}\big)\Big]\;.$$

C.2. Main lemmas. We first bound the error incurred by excluding vectors whose entries are close to each other, i.e. of substituting μ_n^* for μ_n:

LEMMA 9. *For a positive random variable $\eta(n)$ with $\eta(n) \perp\!\!\!\perp_{\mathbb{G}} X_n$ and a standard normal variable Z^*, write*

$$E(\mu_n) := \tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n}[h_n(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n}]\,.$$

Then

$$\|d_{\mathrm{w}}(E(\mu_n),Z^*|\mathbb{G}) - d_{\mathrm{w}}(E(\mu_n^*),Z^*|\mathbb{G})\|_1 \leq \frac{k_n^2 C_1(\tfrac{h_n}{\eta(n)})|\mathbf{B}_{b_n}|\mathcal{S}_w^n}{\sqrt{|\mathbf{A}_n|}}\,.$$

PROOF. By definition of the Wasserstein distance,

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

We bound the final term: Since μ_n and X_n are independent, we can apply the definition of the spreading coefficient $\mathcal{S}_w^n$ to obtain

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

which yields the desired result. □

The main bound on the Wasserstein distance is formulated in terms of μ_n^*:

LEMMA 10. *Let $\eta(n)$ be a positive random variable with $\eta(n)\perp\!\!\!\perp_{\mathbb{G}}X_n$, and $\mathcal{F}$ the function class (1). Let*

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

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

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

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

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

PROOF. By Stein's inequality, $d_{\mathrm{w}}(W^*,Z^*)\leq \sup|\mathbb{E}[W^*t(W^*)-t'(W^*)]|$. We decompose the right-hand side: Since $h_n=\sum_i \bar{h}_n^i$,

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

The final term can be bounded further using the triangle inequality as

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

where $(*)$ uses the fact that $\sup_{x\in\mathbb{R}}|t'(x)|\leq\sqrt{2/\pi}$. □

C.3. Bounding the first term in Lemma 10. We proceed to bound each term on the right-hand side of Lemma 10. For the first term, we observe:

LEMMA 11. *Assume the conditions of Theorem 11, and define a random measure $\mu_n^{i-j}(\bullet):=|\mathbf{A}_n|P_{\mu_n\otimes\mu_n}(\boldsymbol{\phi}_j^{-1}\boldsymbol{\phi}_i'\in\bullet|\mathbf{A}_n^{2k_n})$ on $\mathbb{G}$. Then*

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

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

PROOF. The first statement follows from the definition of $\widehat{\mathbb{F}}$, as

$$
\begin{aligned}
&\|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)\| = \big\|\lim_m \tfrac{1}{|\mathbf{A}_m|^{k_n}}\textstyle\int_{\mathbf{A}_m^{k_n}} h_n(\boldsymbol{\phi}_{1:i-1}e\boldsymbol{\phi}_{i+1:k_n}X_n)-h_n(\boldsymbol{\phi}X_n)|d\boldsymbol{\phi}|\big\|_p\\
&\leq \lim_m\frac{1}{|\mathbf{A}_m|^{k_n}}\int_{\mathbf{A}_m^{k_n}}\|h_n(\boldsymbol{\phi}_{1:i-1}e\boldsymbol{\phi}_{i+1:k_n}X_n)-h_n(\boldsymbol{\phi}X_n)\|_p|d\boldsymbol{\phi}|\;\leq\; c_{i,p}(h_n)\,.
\end{aligned}
$$

Since $\mathbb{E}[\mathbb{E}_{\mu_n^{i-j}}[\mathbb{I}_{\mathbf{B}_b}]] = |\mathbf{A}_n|\mathbb{E}[\mathbb{E}_{\mu_n\otimes\mu_n}[\mathbb{I}_{\phi_j^{-1}\phi_i'\in\mathbf{B}_b}|\mathbf{A}_n^{2k_n}]] \leq \mathcal{S}_w^n|\mathbf{B}_b|$, the second statement also holds. $\square$

LEMMA 12. *Assume hypothesis* (30). *Then*

$$\sup_{t\in\mathcal{F}}\Big\|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}[\bar{h}_n^i(\boldsymbol{\phi}X_n)t(W_{in}^{\boldsymbol{\phi}})|\mathbf{A}_n^{k_n}]\big|\mathbb{G}\big]\Big\|_1 \leq K_1C_2\big(\tfrac{h_n}{\eta(n)}\big)\textstyle\sum_{k_n'<i}c_{i,2}\big(\tfrac{h_n}{\eta(n)}\big),$$

where $K_1 = O(|\mathbf{B}_K|\mathcal{S}_w^n)$. *If hypothesis* (31) *holds instead,*

$$\begin{aligned}\sup_{t\in\mathcal{F}}\Big\|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}[\bar{h}_n^i(\boldsymbol{\phi}X_n)t(W_{in}^{\boldsymbol{\phi}})|\mathbf{A}_n^{k_n}]\big|\mathbb{G}\big]\Big\|_1&\\ \leq K_2C_{2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\big[\tfrac{k_n}{\sqrt{|\mathbf{A}_n|}}+C_{2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\big]\mathcal{R}_n(b_n)&\\ +K_3|\mathbf{B}_{b_n}|C_2\big(\tfrac{h_n}{\eta(n)}\big)\textstyle\sum_{k_n'<i}c_{i,2}\big(\tfrac{h_n}{\eta(n)}\big)&\,,\end{aligned}$$

where $K_2 = O(\mathcal{S}_w^n)$ *and* $K_3 = O(\mathcal{S}_w^n)$.

PROOF. We prove the (harder) case of hypothesis (31) first, and then modify it for (30). Similar to $W_{in}^{\boldsymbol{\phi}}$, we abbreviate

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

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

$$\begin{aligned}&\textstyle\sum_i\Big\|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)t(W_{in}^{\boldsymbol{\phi}})\big|\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\Big\|_1\\ &\overset{(*)}{\leq}\textstyle\sum_i\mathbb{E}\Big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[|\bar{h}_n^i(\boldsymbol{\phi}X_n)||W_{in}^{\boldsymbol{\phi}}-W_{ib_nk_n}^{\boldsymbol{\phi}}|\big|\mathbf{A}_n^{k_n}\big]\Big]\\ &\quad+\textstyle\sum_i\Big\|\mathbb{E}\Big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)t(W_{ib_nk_n}^{\boldsymbol{\phi}})\big|\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\Big]\Big\|_1\end{aligned}\tag{14}$$

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

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

where $\mathcal{J}_n(\phi_i,\boldsymbol{\phi}') = \mathcal{I}_{b_n,k_n}(\phi_i,\boldsymbol{\phi}')\setminus\mathcal{I}_{b_n,k_n'}(\phi_i,\boldsymbol{\phi}')$.

To bound the second term, consider the vector $\boldsymbol{\phi}\in\mathbb{G}^{k_n}$ in (14). We define a sequence $(\boldsymbol{\phi}^{i,j})_{j\in\mathbb{N}}$ in $\mathbb{G}^{k_n}$ whose coordinates differ increasingly from the ith coordinate of $\boldsymbol{\phi}$ as j increases: Set $\boldsymbol{\phi}^{i,0} = \boldsymbol{\phi}$. For $j\geq 1$, choose

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

for each $k\leq k_n$. By definition of μ_n^*, we have $\boldsymbol{\phi}^{i,j} = \boldsymbol{\phi}$ for $j\leq b_n$. Then

$$\textstyle\sum_i\Big\|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\bar{h}_n^i(\boldsymbol{\phi}X_n)t(W_{b_n,k_n,i}^{\boldsymbol{\phi}})\big|\mathbb{G}\big]\Big\|_1$$

$$\leq \sum_i \sum_{j\geq b_n} \Big\|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\bar{h}_n^i(\boldsymbol{\phi}^{i,j+1}X_n)[t(W^{\boldsymbol{\phi}}_{ijk_n})-t(W^{\boldsymbol{\phi}}_{i(j+1)k_n})]\big|\mathbb{G}\big]\Big\|_1$$

$$+ \sum_i \sum_{j\geq b_n} \Big\|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}[\bar{h}_n^i(\boldsymbol{\phi}^{i,j+1}X_n)-\bar{h}_n^i(\boldsymbol{\phi}^{i,j}X_n)]t(W^{\boldsymbol{\phi}}_{ijk_n})\big|\mathbb{G}\big]\Big\|_1$$

$$\overset{(*)}{\leq} 4\sqrt{\tfrac{2}{\pi}}\sum_{j,\,j\geq b_n}\sum_i c_{i,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)|\mathbf{A}_n|\big\|W^{\boldsymbol{\phi}}_{ijk_n}-W^{\boldsymbol{\phi}}_{i(j+1)k_n}\big\|_{2+\varepsilon}\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(j|\mathbb{G})$$

$$+ 4\sum_i c_{i,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\sum_{j\geq b_n}\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(j|\mathbb{G})\sqrt{|\mathbf{A}_n|}\mathbb{I}\{d(\boldsymbol{\phi}_i,\boldsymbol{\phi}_{\setminus i})\in[j,j+1]\}\,.$$

Here, $(*)$ is obtained using Lemma 3, and the fact that

$$\sup_{x\in\mathbb{R}}|t'(x)|\leq\sqrt{2/\pi}\quad\text{and}\quad\sup_{x\in\mathbb{R}}|t(x)|\leq 1$$

Since that is true for any $\boldsymbol{\phi}\in\mathbb{G}^{k_n}$, using the definition of $\mathcal{S}_w^n$ we conclude

$$\sum_i \Big\|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big(\bar{h}_n^i(\boldsymbol{\phi}X_n)t(W^{\boldsymbol{\phi}}_{b_n,k_n,i})\big|\mathbf{A}_n^{k_n}\big)\big|\mathbb{G}\big]\Big\|_1$$

$$\leq 4\sqrt{\tfrac{2}{\pi}}\sum_i c_{i,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\sum_j c_{j,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)$$

$$\times\,\mathbb{E}\big[\mathbb{E}_{\mu_n^{\otimes 2}}\big(\mathbb{I}\{j\notin\mathcal{I}_{b_n,k_n}(\boldsymbol{\phi}_i,\boldsymbol{\phi}')\}|\mathbf{A}_n|\alpha^{\frac{\varepsilon}{2+\varepsilon}}(d(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j)|\mathbb{G})|\mathbf{A}_n^{2k_n}\big)\big]$$

$$+4\sum_i c_{i,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\sum_{j\neq i}\mathbb{E}\big[\mathbb{E}_{\mu_n^*}\big(\sqrt{|\mathbf{A}_n|}\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(d(\boldsymbol{\phi}_i,\boldsymbol{\phi}_j)|\mathbb{G})|\mathbf{A}_n^{k_n}\big)\big]$$

$$\leq 4\sum_i c_{i,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\mathcal{S}_w^n\Big(\tfrac{k_n}{\sqrt{|\mathbf{A}_n|}}+\sqrt{\tfrac{2}{\pi}}\sum_i c_{i,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\Big)$$

$$\sum_{i\geq b_n}\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(i|\mathbb{G})|\mathbf{B}_{i+1}\setminus\mathbf{B}_i|\,.$$

That establishes the result under (31). If (30) is assumed instead, the second term of Eq. (14) vanishes by Lemma 3. We hence have

$$\sup_{t\in\mathcal{F}}\Big\|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\sum_i\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)t(W^{\boldsymbol{\phi}}_{in})|\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\Big\|_1$$

$$\leq\;|\mathbf{B}_K|\mathcal{S}_w^n\sum_i\sum_{k'_n<j\leq k_n}c_{i,2}\big(\tfrac{h_n}{\eta(n)}\big)c_{j,2}\big(\tfrac{h_n}{\eta(n)}\big)\,,$$

which shows result also holds under (30). □

C.4. The second term in Lemma 10. The strategy is to use a Taylor expansion, and to bound

$$\big|\bar{h}_n^i(\boldsymbol{\phi}X_n)\big(t(W^*)-t(W^{\boldsymbol{\phi}}_{in})-\Delta^{\boldsymbol{\phi}}_{in}t'(W^*)\big)\big|\leq\frac{\sup_{x\in\mathbb{R}}|t''(x)|}{2}\big|\bar{h}_n^i(\boldsymbol{\phi}X_n)\big|(\Delta^{\boldsymbol{\phi}}_{in})^2$$

As $\bar{h}_n^i(X_n)$ might not admit a third moment we first upper-bound it using the sequence (γ_n). To bound $|\bar{h}_n^i(\boldsymbol{\phi}X_n)\mathbb{I}(\bar{h}_n^i(\boldsymbol{\phi}X_n)\leq\gamma_n c_{i,2}(h_n))|(\Delta^{\boldsymbol{\phi}}_{in})^2$, we must control the probability that random triples $\boldsymbol{\phi},\boldsymbol{\phi}',\boldsymbol{\phi}''\in\mathbb{G}^{k_n}$ satisfy

$$d(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j),d(\boldsymbol{\phi}_i,\boldsymbol{\phi}''_l)\leq b_n\quad\text{and}\quad\boldsymbol{\phi}'\in V_{i,\beta_n}(n)\tag{15}$$

for some $i\leq k_n$ and $j,l\leq k'_n$, and either

$$\text{(i)}\ \min_{l\leq k_n}d(\boldsymbol{\phi}'_j,\boldsymbol{\phi}''_l)\in[k,k+1]\quad\text{or}\quad\text{(ii)}\ \min_{\substack{l\leq k_n\\ l\neq i}}d(\boldsymbol{\phi}'_j,\boldsymbol{\phi}_l)\in[k,k+1]\,.\tag{16}$$

We quantify these as follows: The upper bound on the term in Lemma 10 must be established for fixed values of n and β_n. Given such values, we choose a constant $S^*_2(k_n)$ that satisfies

$$\frac{|\mathbf{A}_n|^2 \left\|\mathbb{E}_{\mu_n^{\otimes 3}}\left[\mathbb{I}\{\phi,\phi'',\phi'' \text{ satisfies (15) and (16}i)\}\middle|\mathbf{A}_n^{3k_n}\right]\right\|_1}{|\mathbf{B}_{k+1}\setminus\mathbf{B}_k||\mathbf{B}_{b_n}|k_n} \leq S^*_2(k_n)$$

$$\text{and}\quad \frac{|\mathbf{A}_n|^2 \left\|\mathbb{E}_{\mu_n^{\otimes 3}}\left[\mathbb{I}\{\phi,\phi'',\phi'' \text{ satisfies (15) and (16}ii)\}\middle|\mathbf{A}_n^{3k_n}\right]\right\|_1}{|\mathbf{B}_{k+1}\setminus\mathbf{B}_k||\mathbf{B}_{b_n}|k_n} \leq S^*_2(k_n)\,.$$

Similarly, we choose a constant S^*_0 such that

$$\frac{|\mathbf{A}_n|}{|\mathbf{B}_m|}\left\|\mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{I}\{d(\phi_i,\boldsymbol{\phi'}_j)\leq m \text{ and } \phi'\notin V_{i,\beta_n}(n)\}|\mathbf{A}_n^{2k_n}]\right\|_1 \leq S^*_0$$

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

LEMMA 13. *Assume (30) holds. Then for $t\in\mathcal{F}$, and any $p,q>0$ satisfying $\frac{1}{p}+\frac{1}{q}=1$,*

$$\sup_{H\in\mathcal{F}}\left\|\mathbb{E}\Big(\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big(\bar{h}_n^i(\phi X_n)(t(W^*)-t(W_{in}^{\phi})-\Delta_{in}^{\phi}t'(W^*))|\mathbf{A}_n^{k_n}\big)\Big|\mathbb{G}\Big)\right\|_1$$
$$\leq K_1\tfrac{k_nk'_n}{\sqrt{|\mathbf{A}_n|}}C_{2q}\big(\tfrac{h_n}{\eta(n)}\big)^2\; S^*_2(k'_n)\textstyle\sum_i\Gamma_{i,p}(\gamma_n)$$
$$+K_2S^*_0C_2(\tfrac{h_n}{\eta(n)})^2+K_3C_2(\tfrac{h_n}{\eta(n)})\textstyle\sum_i c_{i,2}\big(\tfrac{\bar{h}_n^i(\phi X_n)}{\eta(n)}\mathbb{I}\{\tfrac{\bar{h}_n^i(\phi X_n)}{\eta(n)}\geq\gamma_n\}\big)\,,$$

where $K_1=O(|\mathbf{B}_k|^2)$ and $K_2=O(|\mathbf{B}_K|)$ and $K_3=O(\mathcal{S}_w^n|\mathbf{B}_K|)$. If (31) holds instead,

$$\sup_{H\in\mathcal{F}}\left\|\mathbb{E}\Big(\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big(\bar{h}_n^i(\phi X_n)(t(W^*)-t(W_{in}^{\phi})-\Delta_{in}^{\phi}t'(W^*))|\mathbf{A}_n^{k_n}\big)\Big|\mathbb{G}\Big)\right\|_1$$
$$\leq K_1\frac{k_nk'_n|\mathbf{B}_{b_n}|S^*_2(k'_n)}{\sqrt{|\mathbf{A}_n|}}C_{(2+\varepsilon)q}\big(\tfrac{h_n}{\eta(n)}\big)^2\;\sum_i\Gamma_{i,p(1+\frac{\varepsilon}{2})}(\gamma_n)$$
$$+K_2|\mathbf{B}_{b_n}|S^*_0C_2\big(\tfrac{h_n}{\eta(n)}\big)^2+K_3|\mathbf{B}_{b_n}|C_2\big(\tfrac{h_n}{\eta(n)}\big)\textstyle\sum_i c_{i,2}\big(\tfrac{\bar{h}_n^i(\phi X_n)}{\eta(n)}\mathbb{I}\{\tfrac{\bar{h}_n^i(\phi X_n)}{\eta(n)}\geq\gamma_n\}\big)$$

where $K_1=O(\mathcal{R}_n(0))$ and $K_2=O(1)$ and $K_3=O(\mathcal{S}_w^n)$.

PROOF. Suppose first (30) holds. Since $h_n(X_n)$ may not have a third moment, we upper-bound it using the sequence (γ_n). By the triangle inequality,

$$\left\|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[\underbrace{\bar{h}_n^i(\phi X_n)(t(W^*)-t(W_{in}^{\phi})-\Delta_{in}^{\phi}t'(W^*))}_{:=T}|\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\right\|_1$$
$$\leq\left\|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[T\,\mathbb{I}\{|\tfrac{\bar{h}_n^i(\phi X_n)}{c_{i,2}(h_n)}|>\gamma_n\}\,|\,\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\right\|_1$$
$$+\left\|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[T\,\mathbb{I}\{|\tfrac{\bar{h}_n^i(\phi X_n)}{c_{i,2}(h_n)}|\leq\gamma_n\}\,|\,\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\right\|_1.$$

We again bound each term separately. Since $t\in\mathcal{F}$, it satisfies (9), hence

$$\left\|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[T\,\mathbb{I}\{|\tfrac{\bar{h}_n^i(\phi X_n)}{c_{i,2}(h_n)}|>\gamma_n\}\,|\,\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\right\|_1$$
$$\leq 2\textstyle\sum_i\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[|\bar{h}_n^i(\phi X_n)|\mathbb{I}\{\tfrac{|\bar{h}_n^i(\phi X_n)|}{c_{i,2}(h_n)}>\gamma_n\}|\Delta_{in}^{\phi}||\mathbf{A}_n^{k_n}\big]\big];.$$

For all $\boldsymbol{\phi} \in \mathbb{G}^{k_n}$ and $i \in \mathbb{N}$ we have, by definition of $\Delta_{in}^{\boldsymbol{\phi}}$,

$$\textstyle\sum_i \mathbb{E}\big[\frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[|\bar{h}_n^i(\boldsymbol{\phi}X_n)|\mathbb{I}\{\frac{|\bar{h}_n^i(\boldsymbol{\phi}X_n)|}{c_{i,2}(h_n)} > \gamma_n\}|\Delta_{in}^{\boldsymbol{\phi}}||\mathbf{A}_n^{k_n}\big]\big]$$
$$\leq |\mathbf{A}_n|\Big(c_{j,2}\big(\frac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{\eta(n)}\mathbb{I}\{\frac{|\bar{h}_n^i(\boldsymbol{\phi}X_n)|}{c_{i,2}(h_n)} > \gamma_n\}\big)\mathbb{E}\big[\mathbb{E}_{\mu_n^*}[\mathbb{I}\{d(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j') \leq b_n\}|\mathbf{A}_n^{k_n}, \boldsymbol{\phi}]\big]\Big)$$
$$\textstyle\sum_{j \leq k_n'} c_{i,2}\big(\frac{\bar{h}_n^i}{\eta(n)}\big)\ .$$

Using the definition of $\mathcal{S}_w^n$, this implies

$$\big\|\mathbb{E}\big[\frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i \mathbb{E}_{\mu_n^*}\big[T\,\mathbb{I}\{|\frac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{c_{i,2}(h_n)}| > \gamma_n\}\,|\,\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\big\|_1$$
$$\leq \mathcal{S}_w^n|\mathbf{B}_{b_n}|C_2\big(\frac{h_n}{\eta(n)}\big)\textstyle\sum_i c_{i,2}\big(\frac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{\eta(n)}\mathbb{I}\{\frac{|\bar{h}_n^i(\boldsymbol{\phi}X_n)|}{c_{i,2}(h_n)} > \gamma_n\}\big)\ .$$

To bound the second term, we abbreviate

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

Again using the triangle inequality, we have

$$\big\|\mathbb{E}\big[\frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i \mathbb{E}_{\mu_n^*}\big[T\,\mathbb{I}\{|\frac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{c_{i,2}(h_n)}| \leq \gamma_n\}\,|\,\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\big\|_1$$
$$\leq\big\|\mathbb{E}\big[\frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)\mathbb{I}\{|\frac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{c_{i,2}(h_n)}|\leq\gamma_n\}\big(t(\tilde{W}_{in}^{\boldsymbol{\phi}}) - t(W_{in}^{\boldsymbol{\phi}})\big)\big|\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\big\|_1$$
$$+\big\|\mathbb{E}\big[\frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)\mathbb{I}\{|\frac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{c_{i,2}(h_n)}|\leq\gamma_n\}(\Delta_{in}^{\boldsymbol{\phi}} - \tilde{\Delta}_{in}^{\boldsymbol{\phi}})t'(W^*)|\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\big\|_1$$
$$+\big\|\mathbb{E}\big[\frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)$$
$$\mathbb{I}\{|\frac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{c_{i,2}(h_n)}|\leq\gamma_n\}\big(t(\tilde{W}_{in}^{\boldsymbol{\phi}}) - t(\tilde{W}^*) - \tilde{\Delta}_{in}^{\boldsymbol{\phi}}t'(W^*)\big)\big|\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\big\|_1$$
$$=: \text{(a)} + \text{(b)} + \text{(c)}\ ,$$

and must bound (a)—(c) further. Since t is 1-Lipschitz,

$$\text{(a)} \leq \mathbb{E}\big[\frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[\big|\bar{h}_n^i(\boldsymbol{\phi}X_n)\big|\mathbb{I}\{|\frac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{c_{i,2}(h_n)}|\leq\gamma_n\}\big|\tilde{W}_{in}^{\boldsymbol{\phi}} - W_{in}^{\boldsymbol{\phi}}\big|\big|\mathbf{A}_n^{k_n}\big]\big]$$
$$\leq 2\textstyle\sum_i c_{i,2}\big(\frac{h_n}{\eta(n)}\big)\sum_{j\leq k_n'} c_{j,2}\big(\frac{h_n}{\eta(n)}\big)$$
$$\mathbb{E}[|\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{I}\{d(\boldsymbol{\phi}_i, \boldsymbol{\phi'}_j)\leq b_n, \boldsymbol{\phi'}\notin V_{i\beta_n}\}|\mathbf{A}_n^{2k_n}]]$$
$$\leq 2C_2\big(\frac{h_n}{\eta(n)}\big)^2\sup_{i,j}\mathbb{E}[|\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}\big[\mathbb{I}\{d(\boldsymbol{\phi}_i, \boldsymbol{\phi'}_j)\leq b_n, \boldsymbol{\phi'}\notin V_{i\beta_n}\}|\mathbf{A}_n^{2k_n}\big]\big].$$

Analogously, we have

$$\text{(b)} \leq 2|\mathbf{B}_{b_n}|\textstyle\sum_i c_{i,2}\big(\frac{h_n}{\eta(n)}\big)\sum_{j\leq k_n} c_{j,2}\big(\frac{h_n}{\eta(n)}\big)$$
$$\sup_{i,j}\frac{1}{|\mathbf{B}_{b_n}|}\mathbb{E}[|\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{I}\{d(\boldsymbol{\phi}_i, \boldsymbol{\phi'}_j)\leq b_n, \phi'\notin V_{i\beta_n}\}|\mathbf{A}_n^{2k_n}]]\ .$$

To bound (c), we first observe

$$\text{(c)} \leq \frac{1}{2}\sup_{x\in\mathbb{R}}|h''(x)|\mathbb{E}\big[\frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[|\bar{h}_n^i(\boldsymbol{\phi}X_n)|\mathbb{I}\{|\frac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{c_{i,2}(h_n)}|\leq\gamma_n\}(\tilde{\Delta}_{in}^{\boldsymbol{\phi}})^2\,|\,\mathbf{A}_n^{k_n}\big]\big]\ .$$

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

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

Note such a sequence always exists. Abbreviate
$$G(\boldsymbol{\phi}'):=\big[h_n(\boldsymbol{\phi}'X_n)-\bar{h}_n^{\boldsymbol{\phi}'_i,b_n,k'_n}(\boldsymbol{\phi}'X_n)\big]\mathbb{I}\{\boldsymbol{\phi}'\in V_i(\beta_n)\}.$$
An application of the triangle inequality and of Lemma 3 yields
$$\begin{aligned}
&\big\|\mathbb{E}[\tfrac{|\bar{h}_n^i(\boldsymbol{\phi}X_n)|}{\eta(n)^3}\mathbb{I}\{\tfrac{|\bar{h}_n^i(\boldsymbol{\phi}X_n)|}{c_{i,2}(h_n)}\leq\gamma_n\}G(\boldsymbol{\phi}')G(\boldsymbol{\phi}'')|\mathbb{G}]\big\|_1\\
&\leq\textstyle\sum_l\Big\|\mathbb{E}[\tfrac{|\bar{h}_n^i(\boldsymbol{\phi}X_n)|}{\eta(n)^3}\mathbb{I}\{\tfrac{|\bar{h}_n^i(\boldsymbol{\phi}X_n)|}{c_{i,2}(h_n)}\leq\gamma_n\}G(\boldsymbol{\phi}')G(\boldsymbol{\psi}^l)\big|\mathbb{G}]\\
&\qquad\qquad-\mathbb{E}[\tfrac{|\bar{h}_n^i(\boldsymbol{\phi}X_n)|}{\eta(n)^3}\mathbb{I}\{\tfrac{|\bar{h}_n^i(\boldsymbol{\phi}X_n)|}{c_{i,2}(h_n)}\leq\gamma_n\}G(\boldsymbol{\phi}')G(\boldsymbol{\psi}^{l-1})\big|\mathbb{G}]\Big\|_1\\
&\leq 4\Gamma_{i,q(1+\frac{\varepsilon}{2})}(\gamma_n)\textstyle\sum_{j,k}c_{k,2p(1+\frac{\varepsilon}{2})}\big(\tfrac{h_n}{\eta(n)}\big)c_{j,2p(1+\frac{\varepsilon}{2})}\big(\tfrac{h_n}{\eta(n)}\big)\\
&\qquad\mathbb{I}\{\boldsymbol{\phi}''\in V_i(\beta_n)\}\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(\min\{\bar{d}(\boldsymbol{\phi}''_k,\boldsymbol{\phi}),\bar{d}(\boldsymbol{\phi}''_k,\boldsymbol{\phi}')\}|\mathbb{G})\,,
\end{aligned}$$
where the final term sums over $j\in\mathcal{I}_{b_n,k'_n}(\boldsymbol{\phi},\boldsymbol{\phi}')$ and $l\in\mathcal{I}_{b_n,k'_n}(\boldsymbol{\phi},\boldsymbol{\phi}'')$. By Taylor expansion, we hence obtain
$$\begin{aligned}
\text{(c)}&\leq\frac{1}{2}\sup_{x\in\mathbb{R}}|h''(x)|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[|\bar{h}_n^i(\boldsymbol{\phi}X_n)|\mathbb{I}\{|\tfrac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{c_{i,2}(h_n)}|\leq\gamma_n\}(\tilde{\Delta}_{in}^{\boldsymbol{\phi}})^2\,\big|\,\mathbf{A}_n^{k_n}\big]\big]\\
&\leq 4\textstyle\sum_{i\leq k_n,j,k\leq k'_n}\Gamma_{i,q(1+\frac{\varepsilon}{2})}(\gamma_n)c_{k,2p(1+\frac{\varepsilon}{2})}\big(\tfrac{h_n}{\eta(n)}\big)c_{j,2p(1+\frac{\varepsilon}{2})}\big(\tfrac{h_n}{\eta(n)}\big)\sum_b\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(b|\mathbb{G})\\
&\qquad\big\|\mathbb{E}_{\mu_n^{\otimes 3}}\big[\bar{d}(\boldsymbol{\phi}''_k,\boldsymbol{\phi})\in[b,b+1],\ \boldsymbol{\phi}'',\boldsymbol{\phi}'\in\mathbf{B}_{b_n}(\boldsymbol{\phi}),\ \mathbb{I}\{\boldsymbol{\phi}''\in V_i(\beta_n)\}\big|A_n^{3k_n}\big]\big\|_1\\
&\quad+4\textstyle\sum_{i\leq k_n,j,k\leq k'_n}\Gamma_{i,q(1+\frac{\varepsilon}{2})}(\gamma_n)c_{k,2p(1+\frac{\varepsilon}{2})}\big(\tfrac{h_n}{\eta(n)}\big)c_{j,2p(1+\frac{\varepsilon}{2})}\big(\tfrac{h_n}{\eta(n)}\big)\sum_b\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(b|\mathbb{G})\\
&\qquad\big\|\mathbb{E}_{\mu_n^{\otimes 3}}\big[\bar{d}(\boldsymbol{\phi}''_k,\boldsymbol{\phi}')\in[b,b+1],\boldsymbol{\phi}'',\boldsymbol{\phi}'\in\mathbf{B}_{b_n}(\boldsymbol{\phi}),\mathbb{I}\{\boldsymbol{\phi}''\in V_i(\beta_n)\}\big|A_n^{3k_n}\big]\big\|_1\\
&\leq\tfrac{1}{\sqrt{|\mathbf{A}_n|}}\Big(8k'_nk_n|\mathbf{B}_{b_n}|\big(\textstyle\sum_i c_{i,q(1+\frac{\varepsilon}{2})}(\tfrac{h_n}{\eta(n)})\big)^2\big(\sum_i\Gamma_{i,p(1+\frac{\varepsilon}{2})}(\gamma_n)\big)S_2^*(k'_n)\mathcal{R}_n(0)\Big)\,.
\end{aligned}$$
This establishes the result under hypothesis (31). If (30) holds instead, we modify the proof above as follows: There is now 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,
$$\begin{aligned}
&\big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)(t(W^*)-t(W_{in}^{\boldsymbol{\phi}})-\Delta_{in}^{\boldsymbol{\phi}}t'(W^*))|\mathbf{A}_n^{k_n}\big]\big]\big|\\
&\quad\leq\ \frac{4k'_nk_n|\mathbf{B}_K|^2}{\sqrt{|\mathbf{A}_n|}}\big(\textstyle\sum_i c_{i,2q}\big(\tfrac{h_n}{\eta(n)}\big)\big)^2\big(\sum_i\Gamma_{i,p(1+\frac{\varepsilon}{2})}(\gamma_n)\big)S_2^*(k'_n)\\
&\quad+2\textstyle\sum_i c_{i,2}\big(\tfrac{h_n}{\eta(n)}\big)\sum_j c_{j,2}\big(\tfrac{h_n}{\eta(n)}\big)\\
&\qquad\qquad\qquad\qquad\mathbb{E}[|\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{I}\{\bar{d}(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_{1:j})\leq K,\boldsymbol{\phi}\notin V_{i\beta_n}\}|\mathbf{A}_n^{2k_n}]]\\
&\quad+2\mathcal{S}_w^n|\mathbf{B}_K|k'_n\textstyle\sum_i c_{i,2}\big(\dfrac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{\eta(n)}\mathbb{I}\{|\dfrac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{c_{i,2}(h_n)}|>\gamma_n\}\big)M_2\big(\tfrac{h_n}{\eta(n)}\big)\,,
\end{aligned}$$

and the result holds under (30). □

C.5. The third term in Lemma 10.

LEMMA 14. *Fix $p, q > 0$ such that $\frac{1}{p} + \frac{1}{q} = 1$. If (30) holds,*

$$\big\|1 - \tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \mathbb{E}\big[\mathbb{E}_{\mu_n^*}\big[h_n(\boldsymbol{\phi} X_n)\Delta_{in}^{\boldsymbol{\phi}} \big|\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\big\|_1$$
$$\leq \mathbb{E}\big[\big|\tfrac{\eta(n)^2 - \hat{\eta}_{n,K}^2}{\eta(n)^2}\big|\big] + K_1 C_2\big(\tfrac{h_n}{\eta(n)}\big) \textstyle\sum_{j > k_n'} c_{j,2}\big(\tfrac{h_n}{\eta(n)}\big) + \tfrac{K_2 k_n^4}{|\mathbf{A}_n|} C_2\big(\tfrac{h_n}{\eta(n)}\big)^2 ,$$

where $K_1 = O(\mathcal{S}_w^n |\mathbf{B}_K|)$ and $K_2 = O(\mathcal{S}_w^n |\mathbf{B}_K|^2)$. If (31) holds instead,

$$\big\|1 - \tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \mathbb{E}\big[\mathbb{E}_{\mu_n^*}\big[h_n(\boldsymbol{\phi} X_n)\Delta_{in}^{\boldsymbol{\phi}} \big|\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\big\|_1$$
$$\leq K_2 |\mathbf{B}_{b_n}| C_2\big(\tfrac{h_n}{\eta(n)}\big) \textstyle\sum_{j > k_n'} c_{j,2}\big(\tfrac{h_n}{\eta(n)}\big) + K_1 \mathcal{S}_w^n \dfrac{k_n^4 |\mathbf{B}_{b_n}|^2}{|\mathbf{A}_n|} C_2(\dfrac{h_n}{\eta(n)})^2$$
$$+ K_3 C_{2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)^2 \tfrac{k_n^2 |\mathbf{B}_{b_n}|}{|\mathbf{A}_n|} \mathcal{S}_w^n \mathcal{R}_n(b_n) + \mathbb{E}\big[\big|\tfrac{\eta(n)^2 - \hat{\eta}_{n,b_n}^2}{\eta(n)^2}\big|\big] ,$$

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

PROOF. Assume first that (31) holds. We use the abbreviation $G^k(\boldsymbol{\phi}') := h_n(\boldsymbol{\phi}' X_n) - \bar{h}_n^{\phi_i, b_n, k}(\boldsymbol{\phi}' X_n)$. By the triangle inequality,

$$\Big\|\frac{\eta(n)^2 - |\mathbf{A}_n| \sum_i \mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{E}[\bar{h}_n^i(\boldsymbol{\phi} X_n) G^{k_n'}(\boldsymbol{\phi}')|\mathbb{G}]|\mathbf{A}_n^{2k_n}]}{\eta(n)^2}\Big\|_1$$
$$\leq \Big\|\frac{|\mathbf{A}_n| \sum_i \mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{E}[\bar{h}_n^i(\boldsymbol{\phi} X_n)(\bar{h}_n^{\phi_i, b_n, k_n'}(\boldsymbol{\phi}' X_n) - \bar{h}_n^{\phi_i, b_n, k_n}(\boldsymbol{\phi}' X_n))|\mathbb{G}]\big|\mathbf{A}_n^{2k_n}]}{\eta(n)^2}\Big\|_1 \tag{17}$$
$$+ \Big\|\frac{\hat{\eta}_{n,b_n}^2 - |\mathbf{A}_n| \sum_i \mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{E}[\bar{h}_n^i(\boldsymbol{\phi} X_n) G^{k_n}(\boldsymbol{\phi}')|\mathbb{G}]|\mathbf{A}_n^{2k_n}]}{\eta(n)^2}\Big\|_1 + \mathbb{E}\big[\big|\tfrac{\eta(n)^2 - \hat{\eta}_{n,b_n}^2}{\eta(n)^2}\big|\big]$$
$$=: \text{(a)} + \text{(b)} + \text{(c)} .$$

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

$$\text{(a)} \leq \sum_i \Big\|\frac{|\mathbf{A}_n| \mathbb{E}_{\mu_n^{\otimes 2}}[\sum_{j \in \mathcal{I}_{b_n, k_n}(\phi_i, \boldsymbol{\phi}') \setminus \mathcal{I}_{b_n, k_n'}(\phi_i, \boldsymbol{\phi}')} c_{i,2}(\frac{h_n}{\eta(n)}) c_{j,2}(\frac{h_n}{\eta(n)})|\mathbf{A}_n^{2k_n}]}{\eta(n)^2}\Big\|_1$$
$$\leq \mathcal{S}_w^n |\mathbf{B}_{b_n}| \textstyle\sum_i \sum_{j > k_n'} c_{i,2}\big(\tfrac{h_n}{\eta(n)}\big) c_{j,2}\big(\tfrac{h_n}{\eta(n)}\big) .$$

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

$$\mathcal{J}(\boldsymbol{\phi}, \boldsymbol{\phi}') := \{i, j | d(\phi_i, \phi_j') \leq b_n\} . \tag{18}$$

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

$$\psi_i = \phi_i \quad \text{and} \quad \psi_j' = \phi_j', \qquad \mathcal{J}(\boldsymbol{\psi}, \boldsymbol{\psi}') = \{i, j\} . \tag{19}$$

For the remainder of the proof, denote the concatenation of two vectors as

$$[\boldsymbol{\phi}, \boldsymbol{\psi}] := (\phi_1, \ldots, \phi_m, \psi_1, \ldots, \psi_n) \quad \text{ for } \boldsymbol{\phi} = (\phi_1, \ldots, \phi_m), \boldsymbol{\psi} = (\psi_1, \ldots, \psi_n) .$$

Using a telescopic sum, we have

$$\begin{aligned}
&\big\|\mathbb{E}\big[\tfrac{1}{\eta(n)^2}H(\boldsymbol{\phi},\boldsymbol{\phi}')\big|\mathbb{G}\big]-\mathbb{E}\big[\tfrac{1}{\eta(n)^2}H(\boldsymbol{\psi},\boldsymbol{\psi}')\big|\mathbb{G}\big]\big\|_1\\
&\leq \textstyle\sum_{l=0}^{k_n-1}\big\|\mathbb{E}\big[\tfrac{1}{\eta(n)^2}\big(H([\boldsymbol{\psi}_{1:l},\boldsymbol{\phi}_{l+1:k_n}],\boldsymbol{\phi}')-H([\boldsymbol{\psi}_{1:l+1},\boldsymbol{\phi}_{l+2:k_n}],\boldsymbol{\phi}')\big)\big|\mathbb{G}\big]\big\|_1\\
&+\textstyle\sum_{l=0}^{k_n-1}\big\|\mathbb{E}\big[\tfrac{1}{\eta(n)^2}\big(H(\boldsymbol{\psi},[\boldsymbol{\psi}'_{1:l},\boldsymbol{\phi}'_{l+1:k_n}])-H(\boldsymbol{\psi},[\boldsymbol{\psi}'_{1:l+1},\boldsymbol{\phi}'_{l+2:k_n}])\big)\big|\mathbb{G}\big]\big\|_1\\
&\overset{(*)}{\leq} 8\textstyle\sum_{l\neq i}c_{l,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)c_{j,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}\big(\bar{d}([\boldsymbol{\psi}_l,\boldsymbol{\phi}_l],\ [\boldsymbol{\phi}',\boldsymbol{\phi}_{l+1:k_n},\boldsymbol{\psi}_{1:l-1}])\big|\mathbb{G}\big)\\
&+8\textstyle\sum_{l\neq j}c_{l,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)c_{i,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}\big(\bar{d}([\boldsymbol{\psi}'_l,\boldsymbol{\phi}'_l],\ [\boldsymbol{\phi},\boldsymbol{\phi}'_{l+1:k_n},\boldsymbol{\psi}'_{1:l-1}])|\mathbb{G}\big)\,.
\end{aligned}$$

where (*) is follows from Lemma 3 and inequality

$$\begin{aligned}
&\Big\|\tfrac{1}{\eta(n)^2}\big(H([\boldsymbol{\psi}_{1:l},\boldsymbol{\phi}_{l+1:k_n}],\boldsymbol{\phi}')-H([\boldsymbol{\psi}_{1:l+1},\boldsymbol{\phi}_{l+2:k_n}],\boldsymbol{\phi}')\big)\Big\|_{1+\frac{\varepsilon}{2}}\\
&\leq 2c_{l,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)c_{j,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\,.
\end{aligned}$$

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

$$\begin{aligned}
&\Big\|\mathbb{E}\big[\tfrac{1}{\eta(n)^2}H(\boldsymbol{\phi},\boldsymbol{\phi}')\big|\mathbb{G}\big]-\mathbb{E}\big[\tfrac{1}{\eta(n)^2}\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,\boldsymbol{\phi}_i)\widehat{\mathbb{F}}_{\infty,j}(h_n,X_n,\boldsymbol{\phi}'_j)|\mathbb{G}\big]\Big\|_1\\
&\leq 8\textstyle\sum_{l\neq i}c_{l,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)c_{j,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}\big(\bar{d}(\boldsymbol{\phi}_l,[\boldsymbol{\phi}',\boldsymbol{\phi}_{l+1:k_n}])\big|\mathbb{G}\big)\\
&+8\textstyle\sum_{l\neq j}c_{l,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)c_{i,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}\big(\bar{d}(\boldsymbol{\phi}'_l,[\boldsymbol{\phi}'_{l+1:k_n},\boldsymbol{\phi}_i])|\mathbb{G}\big)\,.
\end{aligned}$$

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

$$\begin{aligned}
&\Big\|\mathbb{E}_{\mu_n^{\otimes 2}}\big[\tfrac{\mathbb{I}\{\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')=\{i,j\}\}(H(\boldsymbol{\phi},\boldsymbol{\phi}')-\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,\boldsymbol{\phi}_i)\widehat{\mathbb{F}}_{\infty,j}(h_n,X_n,\boldsymbol{\phi}'_j))}{\eta(n)^2}\big|\mathbf{A}_n^{2k_n}\big]\Big\|_1\\
&\leq 32\big(\textstyle\sum_l c_{l,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\big)^2\tfrac{k_n^2|\mathbf{B}_{b_n}|}{|\mathbf{A}_n|}\textstyle\sum_{m\geq b_n}|\mathbf{B}_{m+1}\setminus\mathbf{B}_m|\mathcal{S}_w^n\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(m|\mathbb{G})\,.
\end{aligned}$$

We can then upper-bound (b) as

$$\begin{aligned}
&\big\||\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}\big[\tfrac{\mathbb{I}\{\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')=\{i,j\}\}(H(\boldsymbol{\phi},\boldsymbol{\phi}')-\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,\boldsymbol{\phi}_i)\widehat{\mathbb{F}}_{\infty,j}(h_n,X_n,\boldsymbol{\phi}'_j))}{\eta(n)^2}\big|\mathbf{A}_n^{2k_n}\big]\big\|_1\\
&+\big\||\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}\big[\tfrac{\mathbb{I}\{\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')\subsetneq\{i,j\}\}(H(\boldsymbol{\phi},\boldsymbol{\phi}')-\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,\boldsymbol{\phi}_i)\widehat{\mathbb{F}}_{\infty,j}(h_n,X_n,\boldsymbol{\phi}'_j))}{\eta(n)^2}\big|\mathbf{A}_n^{2k_n}\big]\big\|_1\\
&=:b_{ij}^1+b_{ij}^2\ \geq \text{(b)}\,.
\end{aligned}$$

We have already obtained a bound for b_{ij}^1 above. For b_{ij}^2, the Cauchy-Schwartz inequality yields

$$\begin{aligned}
\textstyle\sum_{ij}b_{ij}^2 &\leq 4|\mathbf{A}_n|M_2\big(\tfrac{h_n}{\eta(n)}\big)^2\textstyle\sum_{ij}\mathbb{E}\Big[\mathbb{E}_{\mu_n^{\otimes 2}}\big[\mathbb{I}\{\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')\subsetneq\{i,j\}\}\big|A_n^{2k_n}\big]\Big]\\
&\leq 4\tfrac{\mathcal{S}_w^n|\mathbf{B}_{b_n}|^2k_n^4}{|\mathbf{A}_n|}M_2\big(\tfrac{h_n}{\eta(n)}\big)^2\,.
\end{aligned}$$

Substituting the bounds for (a) and (b) so obtained back into (17) then completes the proof under hypothesis (31). If (30) holds instead, correlations between elements separated by a distance exceeding some constant K have no effect. In this case,

$$\big\|t'(W^*)\big(1-\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}\big[\mathbb{E}_{\mu_n^*}\big[h_n(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}\big|\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\big\|_1$$
$$\leq\sqrt{\tfrac{2}{\pi}}\Big(\mathbb{E}\big[\big|\tfrac{\eta(n)^2-\hat{\eta}_{n,K}^2}{\eta(n)^2}\big|\big]+\mathcal{S}_w^n|\mathbf{B}_K|\textstyle\sum_i\sum_{k_n'<j\leq k_n}c_{i,2}\big(\tfrac{h_n}{\eta(n)}\big)c_{j,2}\big(\tfrac{h_n}{\eta(n)}\big)$$
$$+4\tfrac{\mathcal{S}_w^n|\mathbf{B}_K|^2k_n^4}{|\mathbf{A}_n|}M_2\big(\tfrac{h_n}{\eta(n)}\big)^2\Big)\,,$$

which completes the proof. □

C.6. The fourth term in Lemma 10. The final term in Lemma 10 represents variation of $\eta(n)$, and we upper-bound it in terms of its variance. As in the proof of the basic case, $\eta(n)$ can be thought of as an empirical variance, and its variance is a fourth-order quantity. Since the fourth moment of $h_n(X_n)$ may not exist, we control it using the sequence (γ_n). To bound the standard deviation, 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 consider the events

$$d(\boldsymbol{\phi}_{1,i},\boldsymbol{\phi}_{2,j})\leq b_n\qquad d(\boldsymbol{\phi}_{3,l},\boldsymbol{\phi}_{4,m})\leq b_n \tag{20}$$
$$\text{and}\quad\boldsymbol{\phi}_1\in V_{i,\beta_n}\quad\boldsymbol{\phi}_2\in V_{j,\beta_n}\quad\boldsymbol{\phi}_3\in V_{l,\beta_n}\quad\boldsymbol{\phi}_4\in V_{m,\beta_n}\,. \tag{21}$$

Since n is fixed, we can then choose a constant S_4^* such that

$$\tfrac{|\mathbf{A}_n|^3}{|A||\mathbf{B}_{b_n}|^2}\big\|\mathbb{E}_{\mu_n^{\otimes 4}}\big[\mathbb{I}\{\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_4\text{ satisfy (20), (21) and }\boldsymbol{\phi}_{2,j}^{-1}\boldsymbol{\phi}_{3,m}\in A\}\big|\mathbf{A}_n^{4k_n}\big]\big\|\leq S_4^*$$

holds for every Borel set $A\subset\mathbb{G}^{k_n}$ with $|pr_j(A)|\geq 1$ for all $j\leq k_n$.

LEMMA 15. *Fix $p,q>0$ with $\frac{1}{p}+\frac{1}{q}=1$. Assume (30) holds. Then*

$$\textstyle\sum_i\big\|\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}-\mathbb{E}[\bar{h}_n^i(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}|\mathbb{G}]\big|\mathbf{A}_n^{k_n}\big]\big\|_1$$
$$\leq K_1\tfrac{{k_n'}^2}{\sqrt{|\mathbf{A}_n|}}\Gamma^2_{4(1+\frac{\varepsilon}{2})}(\gamma_n)\sqrt{S_4^*}+\tfrac{K_2k_n^4}{|\mathbf{A}_n|}C_2^2\big(\tfrac{h_n}{\eta(n)}\big)+K_3\big[C_2\big(\tfrac{h_n}{\eta(n)}\big)^2|\mathbf{B}_k|S_0^*$$
$$+C_2\big(\tfrac{h_n}{\eta(n)}\big)\textstyle\sum_i\Big(\mathbb{E}\Big[\dfrac{|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|^2}{\eta(n)^2}\mathbb{I}\{|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|>\gamma_n c_{i,2}(h_n)\}\Big]\Big)^{\frac{1}{2}}\Big]\,,$$

where $K_1=O(|\mathbf{B}_K|^{\frac{3}{2}})$ and $K_2=O(\mathcal{S}_w^n|\mathbf{B}_K|^2)$ and $K_3=O(\mathcal{S}_w^n|B_K|)$. If (31) holds instead, then

$$\textstyle\sum_i\big\|\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}-\mathbb{E}[\bar{h}_n^i(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}|\mathbb{G}]\big|\mathbf{A}_n^{k_n}\big]\big\|_1$$
$$\leq K_1\Big(|\mathbf{B}_{b_n}|S_0^*C_{2+\varepsilon}^2\big(\tfrac{h_n}{\eta(n)}\big)+\tfrac{|\mathbf{B}_{b_n}|^2k_n^4}{|\mathbf{A}_n|}C_{2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)^2\Big)$$
$$+K_2\mathcal{S}_w^n\frac{k_n^2|\mathbf{B}_{b_n}|\mathcal{R}_n(b_n)C_{2+\varepsilon}^2\big(\frac{h_n}{\eta(n)}\big)}{|\mathbf{A}_n|}$$
$$+K_3|\mathbf{B}_{b_n}|C_{2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\textstyle\sum_i\Big(\mathbb{E}\big[\tfrac{|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|^2\mathbb{I}\{|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|>\gamma_n c_{i,2}(h_n)\}}{\eta(n)^2}\big]\Big)^{\frac{1}{2}}$$
$$+K_4\frac{|\mathbf{B}_{b_n}|{k_n'}^2\Gamma^2_{4(1+\frac{\varepsilon}{2})}(\gamma_n)}{\sqrt{|\mathbf{A}_n|}}\sqrt{S_4^*}\,,$$

for $K_1=O(\mathcal{S}_w^n)$ and $K_2=O(1)$ and $K_3=O(\mathcal{S}_w^n)$ and $K_4=O(\mathcal{R}_n^{\frac{1}{2}}(0))$.

PROOF. First suppose (31) holds. As in Lemma 14, we abbreviate

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

where we now keep track of the index i. We conditionally center H,

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

Interactions between $\hat{\mathbb{F}}_{\infty,i}$ for different values of i involve terms of the form

$$\begin{aligned}F_{ij}(\phi,\phi',\tau)=\;&\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,\phi)\mathbb{I}\{\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,\phi)\leq\tau c_{i,2}(h_n(X_n))\}\\&\times\widehat{\mathbb{F}}_{\infty,j}(h_n,X_n,\phi')\mathbb{I}\{\widehat{\mathbb{F}}_{\infty,j}(h_n,X_n,\phi')\leq\tau c_{j,2}(h_n(X_n))\}\end{aligned}$$

for $\phi,\phi'\in\mathbb{G}$ and some threshold $\tau\in(0,\infty]$. We again center conditionally,

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

Abbreviate $J_{ij}=\mathbb{I}\{j\in\mathcal{I}_{b_n,k_n'}(\boldsymbol{\phi}_i,\boldsymbol{\phi}'),(\boldsymbol{\phi},\boldsymbol{\phi}')\in V_{i,\beta_n}\times V_{j,\beta_n}\}$. Using the triangle inequality, we obtain:

$$\begin{aligned}&\textstyle\sum_i\big\|\frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}-\mathbb{E}[\bar{h}_n^i(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}|\mathbb{G}]\big|\mathbf{A}_n^{k_n}\big]\big\|_1\\&\leq\textstyle\sum_{i,j}\big\|\frac{|\mathbf{A}_n|}{\eta(n)^2}\mathbb{E}_{\mu_n^{\otimes 2}}\big[J_{ij}\overline{H}(\boldsymbol{\phi},\boldsymbol{\phi}',i)\big|\mathbf{A}_n^{2k_n}\big]\big\|_1\\&+\textstyle\sum_{i,j}\big\|\frac{|\mathbf{A}_n|}{\eta(n)^2}\mathbb{E}_{\mu_n^{\otimes 2}}\big[(1-J_{ij})\overline{H}(\boldsymbol{\phi},\boldsymbol{\phi}',i)\big|\mathbf{A}_n^{2k_n}\big]\big\|_1\\&=:\textstyle\sum_{i,j}a_{ij}+\sum_{i,j}b_{ij}\;.\end{aligned}$$

Consider a_{ij} first. By the triangle inequality

$$\begin{aligned}a_{ij}\leq&\Big|\mathbb{E}\Big[\tfrac{|\mathbf{A}_n|}{\eta(n)^2}\mathbb{E}_{\mu_n^{\otimes 2}}\Big[\mathbb{I}\{j\in\mathcal{I}_{b_n,k_n'}(\boldsymbol{\phi}_i,\boldsymbol{\phi}')\}\big|\mathbb{E}[|\overline{H}(\boldsymbol{\phi},\boldsymbol{\phi}',i)|\,|\mathbb{G}]-\mathbb{E}[|\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}_j',\infty)|\,|\mathbb{G}]\big|\Big|\mathbf{A}_n^{2k_n}\Big]\Big]\Big|\\&+\Big\|\tfrac{|\mathbf{A}_n|}{\eta(n)^2}\mathbb{E}_{\mu_n^{\otimes 2}}\big[J_{ij}\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}_j',\infty)\big]\Big\|_1\\=:&\;a_{ij}'+a_{ij}''\;.\end{aligned}$$

To bound a_{ij}', we proceed similarly as in the proof of Lemma 14: Recall the index set $\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')$ in (18). If $\boldsymbol{\phi},\boldsymbol{\phi}'\in\mathbb{G}^{k_n}$ satisfy $\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')=\{i,j\}$, we have

$$\begin{aligned}&\Big\|\tfrac{1}{\eta(n)^2}\big(\mathbb{E}\big[|\overline{H}(\boldsymbol{\phi},\boldsymbol{\phi}',i)|\big|\mathbb{G}\big]-\mathbb{E}\big[|\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}_j',\infty)|\big|\mathbb{G}\big]\big)\Big\|_1\\&\leq 8\textstyle\sum_{l\neq i}c_{l,2+\varepsilon}\big(\frac{h_n}{\eta(n)}\big)c_{j,2+\varepsilon}\big(\frac{h_n}{\eta(n)}\big)\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}\big(d(\boldsymbol{\phi}_l,[\boldsymbol{\phi}',\boldsymbol{\phi}_{l+1:k_n}])\big|\mathbb{G}\big)\\&+8\textstyle\sum_{l\neq j}c_{l,2+\varepsilon}\big(\frac{h_n}{\eta(n)}\big)c_{i,2+\varepsilon}\big(\frac{h_n}{\eta(n)}\big)\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}\big(d(\boldsymbol{\phi}_l',[\boldsymbol{\phi}_{l+1:k_n}',\boldsymbol{\phi}_i])|\mathbb{G}\big)\;.\end{aligned}$$

Applying Lemma 3 and the definition of the random measure μ_n^* gives

$$\begin{aligned}&\textstyle\sum_{i,j}\mathbb{E}\Big[\frac{|\mathbf{A}_n|}{\eta(n)^2}\mathbb{E}_{\mu_n^{\otimes 2}}\Big[\mathbb{I}\{\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')=\{i,j\}\}\\&\qquad\qquad\big|\mathbb{E}[|\overline{H}(\boldsymbol{\phi},\boldsymbol{\phi}',i)|\big|\mathbb{G}]-\mathbb{E}[|\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}_j',\infty)|\big|\mathbb{G}]\big|\;\Big|\mathbf{A}_n^{2k_n}\Big]\Big]\\&\leq 32\,\mathcal{S}_w^n\,\tfrac{k_n^2|\mathbf{B}_{b_n}|}{|\mathbf{A}_n|}\big(\textstyle\sum_l c_{l,2+\varepsilon}\big(\frac{h_n}{\eta(n)}\big)\big)^2\sum_{i\geq b_n}|\mathbf{B}_{i+1}\backslash\mathbf{B}_i|\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(i|\mathbb{G})\;.\end{aligned}$$

Again similarly to the proof of Lemma 14, we obtain

$$\sum_{i,j}\mathbb{E}\Big[\tfrac{|\mathbf{A}_n|}{\eta(n)^2}\mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{I}\{\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')\not\subset\{i,j\}\}$$
$$\times\big|\mathbb{E}[|\overline{H}(\boldsymbol{\phi},\boldsymbol{\phi}',i)|\big|\mathbb{G}]-\mathbb{E}[|\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j,\infty)|\big|\mathbb{G}]\big||\mathbf{A}_n^{2k_n}]\Big]$$
$$\leq 4|\mathbf{B}_{b_n}|^2 M_2\big(\tfrac{h_n}{\eta(n)}\big)^2|\mathbf{A}_n|\mathbb{E}\Big[\mathbb{E}_{\mu_n^{\otimes 2}}\big[\mathbb{I}\{\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')\not\subset\{i,j\}\}|A_n^{2k_n}\big]\Big]$$
$$\leq \tfrac{4\mathcal{S}_w^n|\mathbf{B}_{b_n}|^2k_n^4}{|\mathbf{A}_n|}M_2\big(\tfrac{h_n}{\eta(n)}\big)^2\,.$$

Hence

$$\sum_{i,j}a'_{ij}\leq\;32\,\mathcal{S}_w^n\tfrac{k_n^2|\mathbf{B}_{b_n}|}{|\mathbf{A}_n|}\big(\textstyle\sum_l c_{l,2+\varepsilon}\big(\tfrac{h_n}{\eta(n)}\big)\big)^2\sum_{i\geq b_n}|\mathbf{B}_{i+1}\setminus\mathbf{B}_i|\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(i|\mathbb{G})$$
$$+\tfrac{4\mathcal{S}_w^n|\mathbf{B}_{b_n}|^2k_n^4}{|\mathbf{A}_n|}M_2\big(\tfrac{h_n}{\eta(n)}\big)^2\,.$$

To bound a''_{ij}, abbreviate $J'_{ij}:=\mathbb{I}\{\boldsymbol{\phi}\in V_{i,\beta_n},\boldsymbol{\phi}'\in V_{j,\beta_n},d(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j)\leq b_n\}$. Then

$$a''_{ij}\leq\big\||\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}\big[J'_{ij}\tfrac{\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j,\infty)-\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j,\gamma_n)}{\eta(n)^2}\big|\mathbf{A}_n^{2k_n}\big]\big\|_1$$
$$+\big\||\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}\big[J'_{ij}\tfrac{\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j,\gamma_n)}{\eta(n)^2}\big|\mathbf{A}_n^{2k_n}\big]\big\|_2$$

The first term can be bounded using Cauchy-Schwartz, as

$$\big\|\textstyle\sum_{i,j}|\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}\big[J'_{ij}\tfrac{\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j,\infty)-\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j,\gamma_n)}{\eta(n)^2}\big|\mathbf{A}_n^{2k_n}\big]\big\|_1$$
$$\leq 4\sum_{\min\{i,j\}\leq k'_n}\Big(\mathbb{E}\big[\frac{|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|^2\mathbb{I}\{|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|>\gamma_n c_{i,2}(h_n)\}}{\eta(n)^2}\big]\Big)^{\frac{1}{2}}$$
$$c_{j,2}\big(\tfrac{h_n}{\eta(n)}\big)]\mathbb{E}[\mathbb{E}_{\mu_n^{\otimes 2}}\mathbb{I}\{\boldsymbol{\phi}_i^{-1}\boldsymbol{\phi}'_j\in\mathbf{B}_{b_n}\}|\mathbf{A}_n^{2k_n}]]$$
$$\leq 8\mathcal{S}_w^n|\mathbf{B}_{b_n}|\big(\textstyle\sum_j c_{j,2}\big(\tfrac{h_n}{\eta(n)}\big)\big)$$
$$\sum_i\big(\mathbb{E}[|\frac{1}{\eta(n)^2}\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|^2\mathbb{I}\{|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|>\gamma_n c_{i,2}(h_n)\}]\big)^{\frac{1}{2}}\,.$$

The second term involves four-way interactions, so some abbreviations are helpful: Set $\zeta_i:=\|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)\mathbb{I}\{|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|\leq\gamma_n c_{i,2}(h_n)\}\|_{4+2\varepsilon}$ and $\widehat{\mathbb{F}}^{\gamma_n}_{\infty,i}:=\min\{\widehat{\mathbb{F}}_{\infty,i},\gamma_n\}$. For $\phi,\phi',\psi,\psi'\in\mathbb{G}$ and indices i,j,l,m, we have

$$\big\|\mathrm{Cov}[\widehat{\mathbb{F}}^{\gamma_n}_{\infty,i}(h_n,X_n,\phi)\widehat{\mathbb{F}}^{\gamma_n}_{\infty,l}(h_n,X_n,\phi'),\,\widehat{\mathbb{F}}^{\gamma_n}_{\infty,j}(h_n,X_n,\psi)\widehat{\mathbb{F}}^{\gamma_n}_{\infty,m}(h_n,X_n,\psi')]\big\|_1$$
$$\leq 4\,\zeta_i\,\zeta_j\,\zeta_l\,\zeta_m\,\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}\big(\bar{d}((\phi,\phi'),(\psi,\psi'))\big|\mathbb{G}\big)$$

Therefore, by definition of S_4^*, we have

$$\sum_{i\leq k_n,j\leq k'_n}\big\||\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}\big[\mathbb{I}\{\boldsymbol{\phi}'\in V_{j,\beta_n},d(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j)\leq b_n\}\tfrac{\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j,\gamma_n)}{\eta(n)^2}\big|\mathbf{A}_n^{2k_n}\big]\big\|_2$$
$$\leq 8\frac{|\mathbf{B}_{b_n}|{k'_n}^2}{\sqrt{|\mathbf{A}_n|}}\Big(S_4^*\textstyle\sum_i|\mathbf{B}_{i+1}\setminus\mathbf{B}_i|\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(i|\mathbb{G})\Big)^{\frac{1}{2}}\sum_{i\leq k_n,j\leq k'_n}\zeta_i\,\zeta_j\,.$$

In summary, we can upper-bound a''_{ij} as

$$\sum_{i\leq k_n, j\leq k'_n} a''_{ij} \leq 8\frac{|\mathbf{B}_{b_n}|k_n'^{\,2}}{\sqrt{|\mathbf{A}_n|}}\Big(S_4^*\sum_i |\mathbf{B}_{i+1}\setminus\mathbf{B}_i|\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(i|\mathbb{G})\Big)^{\frac{1}{2}}\sum_{i\leq k_n, j\leq k'_n}\zeta_i\zeta_j$$
$$+\,8\,\mathcal{S}_w^n\,|\mathbf{B}_{b_n}|\big(\sum_j c_{j,2}\big(\tfrac{h_n}{\eta(n)}\big)\big)$$
$$\sum_i\big(\mathbb{E}[|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|^2\mathbb{I}\{|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|>\gamma_n c_{i,2}(h_n)\}]\big)^{\frac{1}{2}}\,.$$

An upper bound on the final term $\sum_{i,j}(b_{ij})$ is, by Cauchy-Schwartz,

$$2\big(\sum_i c_{i,2}\big(\tfrac{h_n}{\eta(n)}\big)\big)^2\sup_{i,j}\mathbb{E}[\mathbb{E}_{\mu_n^{\otimes 2}}[|\mathbf{A}_n|\mathbb{I}\{\boldsymbol{\phi}'\notin V_i(\beta_n), d(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j)\leq b_n\}|\mathbf{A}_n^{2k_n}]]\,,$$

and we have hence obtained all terms in the bound under hypothesis (31). If (30) holds instead, there is again a constant distance K beyond which correlations vanish, and

$$\text{(a)} \leq \frac{8\mathcal{S}_w^n|\mathbf{B}_K|^2k_n^4}{|\mathbf{A}_n|}M_2^2\big(\tfrac{h_n}{\eta(n)}\big)+|\mathbf{B}_{b_n}|S_0^*\mathcal{S}_w^n\big(\sum_i c_{i,2}\big(\tfrac{h_n}{\eta(n)}\big)\big)^2$$
$$\text{(b)} \leq 2\frac{|\mathbf{B}_K|^{\frac{3}{2}}k_n'^{\,2}}{\sqrt{|\mathbf{A}_n|}}\sqrt{S_4^*}\sum_{i\leq k_n,j\leq k'_n}\zeta_i\zeta_j\,,$$

which completes the proof of the lemma. □

C.7. Proof of the central limit theorem. We complete the proof of Theorem 10 by showing $d_{\mathrm{w}}(\sqrt{|\mathbf{A}_n|}\,\widehat{\mathbb{F}}_n(h_n,X_n),\eta Z)\to 0$. We first note that

$$\|\widehat{\eta}^2_{m,n}-\eta^2_m\|_1 \xrightarrow{n\to\infty} 0 \qquad \text{for all } m\in\mathbb{N}\,. \tag{22}$$

That is the case since, for every $\varepsilon>0$, we have

$$\mathbb{E}[|\widehat{\eta}^2_{m,n}-\eta^2_m|]\leq\varepsilon+\mathbb{E}[\eta^2_m\mathbb{I}\{|\widehat{\eta}^2_{m,n}-\eta^2_m|>\varepsilon\})+\mathbb{E}\big[\widehat{\eta}^2_{m,n}\mathbb{I}\{|\widehat{\eta}^2_{m,n}-\eta^2_m|>\varepsilon\}\big]$$
$$\leq\varepsilon+\mathbb{E}[\eta^2_m\mathbb{I}\{|\widehat{\eta}^2_{m,n}-\eta^2_m|>\varepsilon\}]$$
$$+\,|\mathbf{B}_m|\mathcal{S}_w^n\big(\sum_i c_{i,2}(h_n\mathbb{I}\{|\widehat{\eta}^2_{m,n}-\eta^2_m|>\varepsilon\})\big)^2\,,$$

and (22) follows by uniform integrability of $(h_n(\boldsymbol{\phi}X_n)^2)_{\boldsymbol{\phi},n}$.

We next must specify suitable sequences of coefficients γ_n, β_n, k_n, k'_n, and b_n for which the relevant terms in the bounds in Lemma 9 and 10 converge to 0 as $n\to\infty$. We first choose (γ_n) and (β_n) to satisfy

$$r_n^1 := \beta_n\gamma_n^2k_n^2/\sqrt{|\mathbf{A}_n|}\;\longrightarrow\;0\,.$$

Such sequences exist, since $k_n^2/\sqrt{|\mathbf{A}_n|}\to 0$. Because of (22), (b_n) can be chosen to additionally satisfy

$$\|\widehat{\eta}^2_{b_n,n}-\eta^2_{b_n}\|_1\xrightarrow{n\to\infty}0\,.$$

In addition we ask that (k'_n) and (b_n) satisfy

$$r_n^2 := |\mathbf{B}_{b_n}|k'_nS_0^*\;\longrightarrow\;0$$
$$r_n^3 := |\mathbf{B}_{b_n}|k'_n\sum_i c_{i,2}\big(\bar{h}_n^i(\boldsymbol{\phi}X_n)\mathbb{I}\{\tfrac{|\bar{h}_n^i(\boldsymbol{\phi}X_n)|}{c_{i,2}(h_n)}>\gamma_n\}\big)\;\longrightarrow\;0$$
$$r_n^4 := |\mathbf{B}_{b_n}|\Big(\sum_{k'_n<i}c_{i,2+\varepsilon}(h_n)\Big)\;\longrightarrow\;0$$
$$r_n^5 := |\mathbf{B}_{b_n}|\frac{k_n^2\gamma_n^2}{\sqrt{|\mathbf{A}_n|}}+\mathcal{R}_n(b_n)+k'_nr_n^1\;\longrightarrow\;0$$

as $n \to \infty$, which is possible since $S_0^* \to 0$ as $\beta_n \to \infty$. Consequently, we can choose sequences (δ_n) and (ε_n), with $\delta_n \to \infty$ and $\varepsilon_n \to \infty$ such that

$$\delta_n/\varepsilon_n^3 \to 0 \quad \text{ and } \quad \delta_n r_n^j/\varepsilon_n^3 \xrightarrow{n\to\infty} 0 \quad \text{ for } j = 1, \ldots, 5\,.$$

Because of (22), these sequences can be chosen to additionally satisfy

$$\frac{\delta_n}{\varepsilon_n^3}\|\widehat{\eta}^2_{b_n,n} - \eta^2_{b_n}\|_1 \xrightarrow{n\to\infty} 0\,.$$

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

$$\eta(n) := \eta\mathbb{I}\{\eta \in [u_n, v_n]\} + \varepsilon_n\mathbb{I}\{\eta \notin [u_n, v_n]\}\,.$$

Then using Lemma 2 we obtain

$$d_{\mathrm{w}}(S_n, \eta(n)Z) \leq \delta_n\mathbb{E}\big[d_{\mathrm{w}}\big(\tfrac{S_n}{\eta(n)}, Z\big|\mathbb{G}\big)\big] \quad \text{ for } \quad S_n := \sqrt{|\mathbf{A}_n|}\,\widehat{\mathbb{F}}_n(h_n, X_n)\,.$$

To apply Lemma 9 and Lemma 10, we note that

$$\sup_n \textstyle\sum_i c_{i,2}\big(\bar{h}_n^i(\boldsymbol{\phi}X_n)\mathbb{I}\{|\frac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{c_{i,2}(h_n)}| > \gamma_n\}\big) \to 0 \qquad \text{as } \gamma_n \to \infty\,.$$

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 the sequences satisfying:

$$S_2^* \leq k'_n\beta_n\mathcal{S}_w^n \qquad S_4^* \leq {k'_n}^2\beta_n^2\mathcal{S}_w^n \qquad S_0^* \to 0\,.$$

Moreover, we have $\sum_{i\leq k_n, j\leq k'_n}\zeta_i\,\zeta_j \leq \frac{\gamma_n^2}{\varepsilon_n^2}\big[\sum_i c_{i,2+\varepsilon}(h_n)\big]^2$ and

$$\textstyle\sum_i \big\|\bar{h}_n^i(\boldsymbol{\phi}X_n)\mathbb{I}\{|\bar{h}_n^i(\boldsymbol{\phi}X_n)| \leq \gamma_n c_{i,2}\big(\frac{h_n}{\eta(n)}\big)\}\big\|_{L_\infty} \leq \gamma_n\sum_i c_{2,i}(h_n)\,.$$

Substituting into Lemma 9 and 10, we then obtain an upper bound on $\mathbb{E}\big[d_{\mathrm{w}}\big(\frac{S_n}{\eta(n)}, Z\big|\mathbb{G}\big)\big]$ and hence, as shown above, on $d_{\mathrm{w}}(S_n, Z)$ as claimed.

C.8. Proof of the Berry-Esseen theorem. To prove Theorem 11, let μ_n^* be the random measure defined in Eq. (13). We consider the variable

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

and similarly define W^* by substituting μ_n^* for μ_n, as in Lemma 10. If (b_n) is the increasing sequence chosen in the theorem, Lemma 9 shows

$$\big|d_{\mathrm{w}}(W, Z) - d_{\mathrm{w}}(W^*, Z)\big| \;\leq\; \frac{k_n^2 C_1(\frac{h_n}{\eta(n)})|\mathbf{B}_{b_n}|\mathcal{S}_w^n}{\sqrt{|\mathbf{A}_n|}}\,.$$

(If hypothesis Eq. (30) is assumed, we can in particular choose $b_n = K$ for all n and some K.) We can apply Lemma 10, where we choose $\eta(n) := \eta$ and $k'_n := k_n$ for all n. In Lemma 12–15, we can set $p = \frac{3}{2}$ and $q = \frac{1}{3}$. The constants S_2^*, S_4^* and the weak spreading coefficient $\mathcal{S}_w^n$ can then be bounded in terms of the (strong) spreading coefficients as

$$S_2^* \leq \mathcal{S}^n \qquad S_4^* \leq \mathcal{S}^n \qquad \mathcal{S}_w^n \leq \mathcal{S}^n\,,$$

and substitute these into the bounds in Lemma 12–15. The sequences (β_n) and (γ_n), which respectively controls moments of (μ_n) and $\frac{h_n}{\eta(n)}$, are relevant in the proof of the central limit theorem; here, we can set $\beta_n = \gamma_n = \infty$ for all n, and note that

$$\|\bar{h}_n^i(\boldsymbol{\phi}X_n)\mathbb{I}\{|\bar{h}_n^i(\boldsymbol{\phi}X_n)|\leq\gamma_n c_{i,2}\big(\tfrac{h_n}{\eta(n)}\big)\}\|_{3(1+\frac{\varepsilon}{2})} \;=\; \|\bar{h}_n^i(\boldsymbol{\phi}X_n)\|_{3(1+\frac{\varepsilon}{2})} \leq\; c_{i,3(1+\frac{\varepsilon}{2})}\big(\tfrac{h_n}{\eta}\big)$$

and $\zeta_i \leq c_{4+2\epsilon,i}\big(\frac{h_n}{\eta}\big)$. Substituting all terms into Lemma 10 completes the proof.

APPENDIX D: OTHER PROOFS

This appendix collects the proofs of all results aside from the main limit theorems—on mixing coefficients, concentration, and applications—in the order they appear in the text.

D.1. Properties of mixing coefficients.

PROOF OF LEMMA 3. Fix $n \in \mathbb{N}$ and $(A, B) \in \mathcal{C}(n)$. Using the triangle inequality,

$$\begin{aligned}&\mathbb{E}\big[|P(A|\mathbb{G})P(B|\mathbb{G}) - P(A\cap B|\mathbb{G})|\big]\\&\leq 2 \sup_{C\in\sigma(\mathbb{G})} \mathbb{E}\big[\mathbb{I}(C)\big(P(A|\mathbb{G})P(B|\mathbb{G}) - P(A\cap B|\mathbb{G})\big)\big] \;\leq\; 2 \sup_{C\in\sigma(\mathbb{G})} \big(a+b\big)\end{aligned}$$

where we have abbreviated

$$\begin{aligned}a &:= \mathbb{E}\big[\mathbb{I}(C)P(A|\mathbb{G})P(B|\mathbb{G}) - P(A)P(B\cap C)\big]\\ \text{and}\;\; b &:= \mathbb{E}\big[P(A)P(B\cap C) - \mathbb{I}(C)P(A\cap B|\mathbb{G})\big]\;.\end{aligned}$$

It follows from the tower property that

$$b \;\leq\; \big|P(A\cap B\cap C) - P(A)P(B\cap C)\big| \leq \alpha(n)\;,$$

and therefore $b \leq \alpha(n)$. Similarly,

$$\begin{aligned}a \;&\leq\; \big|\mathbb{E}\big[P(A)P(B\cap C) - \mathbb{I}(A)P(B\cap C|\mathbb{G})\big]\big|\\&\leq \big|P(A)P(B\cap C) - \mathbb{E}\big[\mathbb{I}(A)P(B\cap C|\mathbb{G})\big]\big| \leq \lim_{k\to\infty} \alpha(k) = 0\;.\end{aligned}$$

In summary, $\mathbb{E}[|P(A|\mathbb{G})P(B|\mathbb{G}) - P(A\cap B|\mathbb{G})|] \leq 4\alpha(n)$. Since that is the case for all $n \in \mathbb{N}$ and $(A, B) \in \mathcal{C}(n)$, we conclude $\alpha(n|\mathbb{G}) \leq 4\alpha(n)$ □

To relate marginal and conditional mixing coefficients, we use Lemma 3:

PROOF OF PROPOSITION 9. Fix $i, j \leq k$. We can choose a subset $G \subset \mathbb{G}$ and vectors $\boldsymbol{\phi}, \boldsymbol{\phi}', \boldsymbol{\psi}, \boldsymbol{\psi}' \in \mathbb{G}^k$ such that $\delta_{i,j}(\boldsymbol{\phi}, \boldsymbol{\phi}', G) \geq t$ and $\delta_{i,j}(\boldsymbol{\psi}, \boldsymbol{\psi}', G) \geq t$ and

$$\psi_l = \begin{cases}\pi\phi_i & \text{if } l = i\\ \phi_l & \text{otherwise}\end{cases} \qquad \psi'_l = \begin{cases}\pi\phi'_j & \text{if } l = j\\ \phi'_l & \text{otherwise}\end{cases} \qquad \text{for some } \pi \in \mathbb{G}\;.$$

For Borel sets $A \subset \mathbb{R}^2$ and $B \subset \mathbb{R}^G$, Lemma 3 shows

$$\begin{aligned}&\big\|\mathbb{E}\big[\mathbb{I}[(X_{\boldsymbol{\phi}}, X_{\boldsymbol{\phi}'})\in A]\mathbb{I}[X_G\in B]|\mathbb{G}\big] - \mathbb{E}\big[\mathbb{I}[(X_{\boldsymbol{\psi}}, X_{\boldsymbol{\psi}'})\in A]\mathbb{I}[X_G\in B]|\mathbb{G}\big]\big\|_1\\&\leq \alpha(t|\mathbb{G}).\end{aligned}$$

Substituting into the definition of $P_{i,j}(\cdot)$ gives

$$|P(A, B|\mathbb{G}) - \mathbb{E}[P_{i,j}(A)\mathbb{I}\{X_n \in B\}|\mathbb{G}_n]| \leq \alpha(t|\mathbb{G})$$

for all $i, j \leq k$, and hence $\alpha_n(t|\mathbb{G}) \leq \alpha(t|\mathbb{G})$ as claimed. □

D.2. Concentration. To prove concentration, we use the "exchangeable pairs" variant of Stein's method, in this form due to Chatterjee [2].

PROOF OF THEOREM 13. The proof strategy is to regard $\mathbb{E}_{\mu_n}[h_n(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n}]$ as an integral, approximate this integral by sums, and establish concentration of each sum. These sums are constructed as follows: For each $m \in \mathbb{N}$, let C_m be an ϵ_m-net with $\epsilon_m = 1/m$. Let λ_m be a partition of $\mathbb{G}$ into a countable number of measurable sets; we write $\lambda_m(\phi)$ for the set containing a given $\phi \in \mathbb{G}$. Clearly, this partition can be chosen such that

$$\text{each } \phi \in C_m \text{ is in a separate set of } \lambda_m \quad \text{ and } \quad \lambda_m(\phi) \subset \mathbf{B}_{1/m}(\phi)\,.$$

Since λ_m partitions $\mathbb{G}$, the product $\lambda_m^{k_n} := \lambda_m \times \ldots \times \lambda_m$ partitions $\mathbb{G}^{k_n}$, and we discretize the integral as

$$\Sigma_{nm} := \textstyle\sum_{\boldsymbol{\phi} \in C_m^{k_n}} \mathbb{E}_{\mu_n}\big[\lambda_m^{k_n}(\boldsymbol{\phi})|\mathbf{A}_n^{k_n}\big] h_n(\boldsymbol{\phi}X_n)\,.$$

For each fixed $n \in \mathbb{N}$, the approximation error satisfies

$$\big\|\Sigma_{nm} - \mathbb{E}_{\mu_n}[h_n(\boldsymbol{\phi}X_n)\big|\mathbf{A}_n^{k_n}]\big\|_1 \leq \sup_{\substack{\boldsymbol{\phi},\boldsymbol{\phi}' \in \mathbb{G}^{k_n} \\ d(\boldsymbol{\phi}_i',\boldsymbol{\phi}_i) \leq \epsilon_m,\; i \leq k_n}} \|h_n(\boldsymbol{\phi}X_n) - h_n(\boldsymbol{\phi}'X_n)\|_1 \xrightarrow{m} 0.$$

Thus, $\|\Sigma_{nm} - \mathbb{E}_{\mu_n}[h_n(\boldsymbol{\phi}X_n)\big|\mathbf{A}_n^{k_n}]\| \to 0$ as $m \to \infty$. Since h_n is $\mathbf{L}_1$-uniformly continuous,

$$\mathbb{P}(|\mathbb{E}_{\mu_n}[h_n(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n}]| > t \,|\, \mu_n) \;\leq\; \limsup_m \mathbb{P}(|\Sigma_{nm}| \geq t \,\big|\, \mu_n) \qquad \text{for } t > 0\,.$$

Now apply the method of exchangeable pairs: Consider the sets of vectors

$$\lambda_m^{-i}(\phi) := \{(\psi_1, \ldots, \psi_{k_n}) \in \mathbf{A}_n^{k_n} | \psi_i \in \lambda_m(\phi)\}.$$

Recall that Σ_{nm} is self-bounded by hypothesis. For each $i \leq k_n$, the self-bounding coefficient is $\sum_{\boldsymbol{\phi} \in C_m^{k_n}} c_i \mathbb{E}_{\mu_n}[\lambda_m^{-i}(\boldsymbol{\phi}_i)|\mathbf{A}_n^{k_n}]$. Using [2, Theorem 4.3], we obtain

$$\begin{aligned}
\mathbb{P}(|\Sigma_{mn}| \geq t|\mu_n) \quad &\leq \quad 2\,\mathbb{E}\Big[\exp\Big(-\frac{(1-\Lambda[(X_\phi)_{\phi \in C_m}])t^2}{\sum_{\phi \in C_m}(\sum_i c_i \mathbb{E}_{\mu_n}[\lambda_m^{-i}(\boldsymbol{\phi}_i)|\mathbf{A}_n^{k_n}])^2}\Big)\Big] \\
&\leq \quad 2\,\mathbb{E}\Big[\exp\Big(-|\mathbf{A}_n|\frac{(1-\Lambda[(X_\phi)_{\phi \in C_m}])t^2}{\tau_n^2 |\mathbf{B}_{1/m}| \big(\sum_i c_i\big)^2}\Big)\Big]\,,
\end{aligned}$$

where the second inequality uses the definition of τ_n. That holds for any m, and any decreasing sequence (C_m) of nets. For $m \to \infty$, we hence obtain

$$\mathbb{P}(|\mathbb{E}_{\mu_n}(h_n(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n})| \geq t \,|\, \mu_n) \leq 2\,\mathbb{E}\big[\exp\big(-|\mathbf{A}_n|\tfrac{(1-\rho_n)t^2}{[\sum_i c_i]^2 \tau_n^2}\big)\big]$$

as claimed, where we have substituted in the definition of ρ_n. □

D.3. Approximation by subsets of transformations. Recall that we may assume $\mathbb{E}[f(X)|\mathbb{G}] = 0$ without loss of generality, by Lemma 6.

PROOF OF PROPOSITION 15. Set $f' := f - \mathbb{E}[f(X)|\mathbb{G}]$. By Theorem 10,

$$\int_{\mathbf{A}_n} \frac{f'(\phi X)}{\sqrt{|\mathbf{A}_n|}} |d\phi| \xrightarrow{\text{d}} \eta Z\,.$$

For the measures (μ_n) chosen as $\mu_n(A) := |A \bigcap \mathbb{H}|$, the theorem shows

$$\int_{\mathbf{A}_n \cap \mathbb{H}} \frac{f'(\phi X)}{\sqrt{|\mathbf{A}_n \cap \mathbb{H}|}} |d\phi| \xrightarrow{\text{d}} \eta_H Z \quad \text{ and } \quad \int_{\mathbf{A}_n} \frac{f'(\phi X)}{\sqrt{|\mathbf{A}_n|}} |d\phi| \xrightarrow{\text{d}} \eta Z\,.$$

Since the random variables η and η_H satisfy

$$
\begin{aligned}
|\mathbb{K}|\eta_H^2 - \eta^2 &= |\mathbb{K}|\int_{\mathbb{H}} \mathbb{E}[f(X)f(\phi X)|\mathbb{G}]|d\phi| - \eta^2 \\
&= \int_{\mathbb{H}}\int_{\mathbb{K}} \mathbb{E}[f(X)[f(\phi X) - f(\phi\theta X)]|\mathbb{G}]|d\theta||d\phi|
\end{aligned}
$$

almost surely, the result follows. □

D.4. Applications. We first establish Theorem 17, on exchangeable structures. The idea of the proof is to represent $(f(\phi X))_{\phi\in\mathbb{S}_n}$ approximately, by a certain random field X_n on $\mathbb{Z}^{k_n}$ that is invariant under diagonal action of shifts. That allows us to apply Theorems 10 and 11. The proof can be read as an example of the generalized U-statistics in Corollary 12.

PROOF OF THEOREM 17. For $i \in \mathbb{N}$, we denote

$$
d_i := \limsup_j \|f(X) - f(\tau_{ij}X)\|_2 \quad \text{and} \quad d_i(\eta) := \limsup_j \big\|\tfrac{f(X)-f(\tau_{ij}X)}{\eta}\big\|_2 .
$$

Consider the segment $[i] = \{1,\dots,i\}$, and write $\mathbb{S}_m^{[i]} = \{\phi\in\mathbb{S}_m | \phi[i] = [i]\}$ for the set of permutations that leave it invariant.

Step 1: Approximation. We define

$$
\bar{f}^i(x) := \lim_{m\to\infty} \tfrac{1}{|\mathbb{S}_m^{[i]}|}\textstyle\sum_{\psi\in\mathbb{S}_m^{[i]}} f(\psi x) ,
$$

and use $\bar{f}^i(\phi X)$ as a surrogate of $f(\phi X)$ that depends only on the image $\phi[i]$. Averaging out the kth coordinate gives

$$
\bar{f}^{i,k}(x) := \lim_{m\to\infty} \tfrac{1}{m}\textstyle\sum_{l\le m} \bar{f}^i(\tau_{l,k}x) .
$$

We will show that for any increasing, divergent sequence (k_n),

$$
\tfrac{\sqrt{n}}{|\mathbb{S}_n|}\textstyle\sum_{\phi\in\mathbb{S}_n}\big(f(\phi X) - \bar{f}^{k_n}(\phi X)\big) \xrightarrow{\mathbf{L}_1} 0 \qquad \text{as } n\to\infty .
$$

Indeed, since $(f - \bar{f}^{k_n}) = \sum_{k\ge k_n}(\bar{f}^{k+1} - \bar{f}^k)$, we have

$$
\begin{aligned}
&\big\|\tfrac{\sqrt{n}}{|\mathbb{S}_n|}\textstyle\sum_{\phi\in\mathbb{S}_n}\big(f(\phi X) - \bar{f}^{k_n}(\phi X)\big)\big\|_1^2 \le \big\|\tfrac{\sqrt{n}}{|\mathbb{S}_n|}\textstyle\sum_{\phi\in\mathbb{S}_n}\big(f(\phi X) - \bar{f}^{k_n}(\phi X)\big)\big\|_2^2 \\
&\le \tfrac{n}{|\mathbb{S}_n|^2}\textstyle\sum_{\phi,\psi\in\mathbb{S}_n}\mathbb{E}\big[\big(f(\phi X) - \bar{f}^{k_n}(\phi X)\big)\big(f(\psi X) - \bar{f}^{k_n}(\psi X)\big)\big] \\
&\le \tfrac{n}{|\mathbb{S}_n|^2}\textstyle\sum_{k\ge k_n}\sum_{\phi,\psi\in\mathbb{S}_n}\mathbb{E}\big[\big(\bar{f}^{k+1}(\phi X) - \bar{f}^{k}(\phi X)\big)\big(f(\psi X) - \bar{f}^{k_n}(\psi X)\big)\big]
\end{aligned}
$$

Consider the summands on the right-hand side. Observe that

$$
\begin{aligned}
&\mathbb{E}\big[(\bar{f}^k(\phi X) - \bar{f}^{k-1}(\phi X))\bar{f}^{\infty,m}(\psi X)\big] = 0 \\
\text{and}\quad &\mathbb{E}\big[(\bar{f}^k(\phi X) - \bar{f}^{k-1}(\phi X))\bar{f}^{k_n,m}(\psi X)\big] = 0
\end{aligned}
$$

whenever $\psi(m) = \phi(k)$ for $k \le m$. Each summand is hence bounded as

$$
\begin{aligned}
&\big|\mathbb{E}\big[(\bar{f}^{k+1}(\phi X) - \bar{f}^k(\phi X))(f(\psi X) - \bar{f}^{k_n}(\psi X))\big]\big| \\
&= \big|\mathbb{E}\big[(\bar{f}^{k+1}(\phi X) - \bar{f}^k(\phi X))(f(\psi X) - \bar{f}^{\infty,m}(\psi X) - \bar{f}^{k_n}(\psi X) + \bar{f}^{k_n,m}(\psi X))\big]\big| \\
&\le \big\|\bar{f}^{k+1}(\phi X) - \bar{f}^k(\phi X)\big\|_2\big\|f(\psi X) - \bar{f}^{\infty,m}(\psi X) - \bar{f}^{k_n}(\psi X) + \bar{f}^{k_n,m}(\psi X)\big\|_2 \\
&\le 2d_k d_m .
\end{aligned}
$$

Substituting into the bound yields

$$
\begin{aligned}
&\Big\|\tfrac{\sqrt{n}}{|\mathbb{S}_n|}\textstyle\sum_{\phi\in\mathbb{S}_n}(f(\phi X)-\bar{f}^{k_n}(\phi X))\Big\|_1^2\\
&\leq\ \tfrac{n}{|\mathbb{S}_n|^2}\textstyle\sum_{k\geq k_n}\sum_{m\in\mathbb{N}}\sum_{\phi,\psi\in\mathbb{S}_n}\mathbb{I}\{\phi(k)=\psi(m)\}d_kd_m\\
&\leq\ 2\big(\textstyle\sum_{k\geq k_n}d_k\big)\big(\sum_{m\in\mathbb{N}}d_m\big)\ \longrightarrow\ 0.
\end{aligned}
$$

It hence suffices to show $\frac{\sqrt{n}}{|\mathbb{S}_n|}\sum_{\phi\in\mathbb{S}_n}\bar{f}^{k_n}(\phi X)$ is asymptotically normal if $k_n=o(n^{1/4})$.

Step 2: Representation by random fields. For each $n\in\mathbb{N}$, we construct a scalar random field X_n on $\mathbb{Z}^{k_n}$ as follows: For $\mathbf{j}=(j_1,\ldots,j_k)\in\mathbb{Z}^k$, define the permutation $\phi_{\mathbf{j}}:=\tau_{1,j_1}\circ\cdots\circ\tau_{k,j_k}$. Note that $\phi_{\mathbf{j}}[k]=\mathbf{j}$. Then

$$
X_n:=\big(Y_{\mathbf{j}}\big)_{\mathbf{j}\in\mathbb{Z}^{k_n}}\qquad\text{where}\qquad Y_{\mathbf{j}}:=\begin{cases}\bar{f}^{k_n}(\phi_{\mathbf{j}}X) & \text{if } \mathbf{j}_l\neq\mathbf{j}_k \text{ for all } l\neq k\\ 0 & \text{otherwise}\end{cases}
$$

is a random element of $\mathbf{X}_n:=\mathbb{R}^{\mathbb{Z}^{k_n}}$. The shift $(\mathbf{i},(x_{\mathbf{j}})_{\mathbf{j}\in\mathbb{Z}^{k_n}})\mapsto(x_{\mathbf{j}+\mathbf{i}})_{\mathbf{j}\in\mathbb{Z}^{k_n}}$ is an action of the group $\mathbb{Z}^{k_n}$ on $\mathbf{X}_n$. Since X is exchangeable, X_n is by construction invariant under the diagonal action of $\mathbb{Z}^{k_n}$, and its marginal mixing coefficients satisfy $\alpha_n(t|\mathbb{G})=0$ for all $t>0$. Theorem 10 then shows convergence as in (35) holds, for $\eta\perp\!\!\!\perp Z$.

Step 3: Berry-Esseen bound. The reasoning is similar: For $k\in\mathbb{N}$, we have

$$
d_{\mathrm{W}}\big(\tfrac{\sqrt{n}}{\eta|\mathbb{S}_n|}\textstyle\sum_{\phi\in\mathbb{S}_n}f(\phi X),\tfrac{\sqrt{n}}{\eta|\mathbb{S}_n|}\sum_{\phi\in\mathbb{S}_n}\bar{f}^k(\phi X)\big)\leq 2\big(\sum_{l\geq k}d_l(\eta)\big)\big(\sum_{m\in\mathbb{N}}d_m(\eta)\big).
$$

We denote $\eta^2(n):=\sum_{i,j\leq k}\mathrm{Cov}[\mathbb{F}^i(X,e)\mathbb{F}^j(X,\phi)|\mathbb{G}]$, and observe that

$$
\begin{aligned}
\Big\|\frac{\eta^2(n)-\eta^2}{\eta^2}\Big\|\ &\leq\ \Big\|\frac{\sum_{l=k}^{\infty}\sum_{m\in\mathbb{N}}\mathrm{Cov}[\mathbb{F}^l(X,e)\mathbb{F}^m(X,\phi)|\mathbb{G}]}{\eta^2}\Big\|\\
&\leq\ 2\big(\textstyle\sum_{m\in\mathbb{N}}d_m(\eta)\big)\sum_{l\geq k}d_l(\eta).
\end{aligned}
$$

Substituting into Theorem 11 gives

$$
d_{\mathrm{W}}\big(\frac{\sqrt{n}}{\eta|\mathbb{S}_n|}\textstyle\sum_{\phi\in\mathbb{S}_n}f(\phi X),Z\big)\ \leq\ C\big(\frac{k^2}{\sqrt{n}}+\sum_{l\geq k}d_l(\eta)\big)\,,
$$

for some $C<\infty$. □

PROOF OF PROPOSITION 21. Write $\mathbb{L}:=\{\phi\in\mathbb{G}|\phi(W)\cap W\neq\varnothing\}$. Observe that, if we choose ϕ to be an element of $\mathbb{H}\setminus(\mathbf{A}_n\cap\mathbb{H})$ that is such that $\phi(W)\cap\mathbf{A}_nW\neq\varnothing$, then we have $\phi\in\mathbf{A}_n\mathbb{L}\cap\mathbb{H}$. This implies that

$$
\begin{aligned}
&\big\|\sqrt{|\mathbf{A}_n\cap\mathbb{H}|}\big(\nu_n(h)-\frac{1}{|\mathbf{A}_n\cap\mathbb{H}|}\int_{\mathbf{A}_n\cap\mathbb{H}}f(\phi(\Pi))d|\phi|\big)\big\|_2^2\\
&\leq\frac{|(\mathbf{A}_n\vartriangle\mathbf{A}_n\mathbb{L})\cap\mathbb{H}|}{|\mathbf{A}_n\cap\mathbb{H}|}\|f(\Pi)\|_{2+\epsilon}^2\sum_{i\in\mathbb{N}}|\mathbf{B}_{i+1}\setminus\mathbf{B}_i|\alpha_i(i|\mathbb{G})^{\frac{\epsilon}{2+\epsilon}}\to 0\,,
\end{aligned}
$$

and Theorem 4 shows $\frac{1}{\sqrt{|\mathbf{A}_n\cap\mathbb{H}|}}\int_{\mathbf{A}_n\cap\mathbb{H}}f(\phi(\Pi))-\mathbb{E}(f(\Pi)|\mathbb{G})d|\phi|\xrightarrow{d}\eta Z$. □

PROOF OF PROPOSITION 22. By hypothesis, $\sup_{i>0}i^{-r}|\mathbf{B}_i|<\infty$, polynomial stability holds with index $q>\frac{(2+2\epsilon)r}{\epsilon}$, and Π is a Poisson process. We have to show that

$$
\int_{\mathbb{G}}\alpha^{(n)}(d(e,\phi)|\mathbb{G})^{\frac{\epsilon}{2+\epsilon}}|d\phi|<\infty\,.
$$

For $b \in \mathbb{N}$, a subset $G_1 \subset \mathbb{G}$, and $F \in \mathcal{F}$, define

$$f_{n,b}(F) = f_n(F \cap \mathbf{B}_b) \quad Y(G_1) := (f_n(\phi(\Pi)))_{\phi \in G_1} \quad Y_b(G_1) := (f_{n,b}(\phi(\Pi)))_{\phi \in G_1} \,.$$

Write $\mathcal{L}(\bullet)$ for the law of a random variable. Then

$$\big\|\mathcal{L}(Y(G_1)) - \mathcal{L}(Y_b(G_1))\big\|_{\text{TV}} \leq P(Y(G_1) \neq Y_b(G_1)) \leq \mathbb{E}\Big[\sum_{(x,y)\in G_1 W \cap \Pi} \mathbb{I}(R(x,y,\Pi_n) > b)\Big]$$

An application of Campbell's theorem for Poisson processes [5] shows there are constants $C_1, C_2 > 0$ and $\gamma := |\{\phi \in \mathbb{H} | \phi(W) \cap G_1 W \neq \varnothing\}|$ such that

$$\big\|\mathcal{L}(Y(G_1)) - \mathcal{L}(Y_b(G_1))\big\|_{\text{TV}} \,\leq\, C_1 \gamma \sup_{(x,m)\in W} P(R(x,m,\Pi_n) > b) \,\leq\, C_2 \gamma b^{-q} \,. \tag{23}$$

Let $\bar{d}$ be the Hausdorff metric induced by d, and denote $\bar{d}$-balls by $\bar{\mathbf{B}}$. Take $G_1 := \{\phi, \phi'\}$ with elements $\phi, \phi' \in \mathbb{G}$ and let G_2 be another subset of $\mathbb{G}$ with $\bar{d}(G_1, G_2) \geq b$. Then there is $C_3 < \infty$ such that

$$\begin{aligned} \big\|\mathcal{L}(Y(G_2)) - \mathcal{L}(Y_{\bar{d}(G_1,G_2)-\frac{b}{2}}(G_2))\big\|_{\text{TV}} &\leq P(Y(G_2) \neq Y_{\bar{d}(G_1,G_2)-\frac{b}{2}}(G_2)) \\ &\leq \textstyle\sum_{j\geq b} P\big(Y(\bar{\mathbf{B}}_{j+1}(G_1) \setminus \bar{\mathbf{B}}_j(G_1)) \neq Y_{\frac{2j-b}{2}}(\bar{\mathbf{B}}_{j+1}(G_1) \setminus \bar{\mathbf{B}}_j(G_1)\big) \\ &\leq C_3 \textstyle\sum_{j\geq 0} (j + \frac{b}{2})^{-q} (j+b)^{r-1} \end{aligned} \tag{24}$$

where the second inequality applies the union bound, and the third follows by substituting the growth rate and the definition of stability into Eq. (23). Whenever G_1 and G_2 satisfy $|G_1| \leq 2$ and $\bar{d}(G_1, G_2) \geq b$, and A, B are measurable sets, there is hence a constant C_4 such that

$$\begin{aligned} &\big|P(Y(G_1) \in A, Y(G_2) \in B) - P(Y(G_1) \in A) P(Y(G_2) \in B)\big| \\ &\quad \leq \|\mathcal{L}(Y_{b/2}(G_1)) - \mathcal{L}(Y(G_1))\|_{\text{TV}} + \|\mathcal{L}(Y_{\bar{d}(G_1,G_2)-b/2}(G_2)) - \mathcal{L}(Y(G_2))\|_{\text{TV}} \\ &\quad \leq C_4 \big(\tfrac{b}{2}\big)^{r-q} \end{aligned}$$

The first inequality holds by independence of $Y_{b/2}(G_1)$ and $Y_{\bar{d}(G_1,G_2)-b/2}(G_2)$, the second follows from Eq. (23) and (24). That implies $\alpha^{(n)}(b|\mathbb{G}_2) \leq C_4 (b/2)^{r-q}$, and hence the desired result since $q > 2\frac{1+\epsilon}{\epsilon} r$. □

PROOF OF THEOREM 23. 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 declaring $\phi_{i-1}^n \prec \phi_i^n$ for all $i \in \mathbb{N}$. 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

$$\tfrac{1}{\sqrt{|\mathbf{A}_n|}} \big(\log P(S_{\mathbf{A}_n}) - \mathbb{E}[\log P(S_{\mathbf{A}_n})]\big) = \tfrac{1}{\sqrt{|\mathbf{A}_n|}} \textstyle\sum_{\phi \in \mathbf{A}_n} f_n(S,\phi) \,.$$

Now consider a ϕ such that $\mathcal{T}_n(\phi) \cap \mathbf{B}_m = \mathcal{T}(\phi) \cap \mathbf{B}_m$. Then

$$\|f_n(S,\phi) - g_m(S,\phi)\|_2 \leq \rho_m \,.$$

The number of $\phi \in \mathbf{A}_n$ for which that is *not* the case is

$$|\{\phi \in \mathbf{A}_n \,|\, \mathcal{T}_n(\phi) \cap \mathbf{B}_m \neq \mathcal{T}(\phi) \cap \mathbf{B}_m\}| \;\leq\; |\mathbf{A}_n \vartriangle \mathbf{B}_m \mathbf{A}_n|$$

Denote $M_p := \sup_{\phi \in \mathbb{G}, A \subset \mathbb{G}} \| \log P(X_\phi | X_A) \|_p$. For any $\phi, \phi' \in \mathbb{G}$ that satisfy $d(\phi, \phi') \geq i$ and any $k \in \mathbb{N}$, we have

$$\begin{aligned} &\mathrm{Cov}\big[f_n(S,\phi) - g_m(S,\phi), f_n(S,\phi') - g_m(S,\phi')\big] \\ &\leq 4\min(\rho_m,\rho_k)^2 + 8\min(\rho_m,\rho_k) M_2 + 4M_{2+\varepsilon}^2 \alpha^{\frac{\varepsilon}{2+\varepsilon}}(i-k, |\mathbf{B}_m|)\,. \end{aligned}$$

Therefore for any sequence (b_n) satisfying $\frac{|\mathbf{A}_n \vartriangle \mathbf{B}_{b_n} \mathbf{A}_n|}{|\mathbf{A}_n|} \to 0$ and $b_n \to \infty$ we have

$$\tfrac{1}{\sqrt{|\mathbf{A}_n|}} \textstyle\sum_{\phi \in \mathbf{A}_n} f_n(S,\phi) - g_{b_n}(S,\phi) \xrightarrow{L_2} 0\,.$$

Let α^m be the mixing coefficient of g_m. Then $\alpha^m(i) \leq \alpha(i - 2m, |\mathbf{B}_m|)$. Theorem 10 hence implies

$$\tfrac{1}{\sqrt{|\mathbf{A}_n|}} \textstyle\sum_{\phi \in \mathbf{A}_n} g_m(\phi X) \xrightarrow{d} \eta_m Z \quad \text{for} \quad \eta_m^2 := \sum_\phi \mathrm{Cov}[g_m(X), g_m(\phi X)]\,.$$

Since $\eta_m \xrightarrow{m \to \infty} \eta$, the result follows. □

REFERENCES

[1] BOLTHAUSEN, E. (1982). On the central limit theorem for stationary mixing random fields. *Ann. Probab.* **10** 1047–1050.

[2] CHATTERJEE, S. (2005). Concentration inequalities with exchangeable pairs, PhD thesis.

[3] EAGLESON, G. K. and WEBER, N. C. (1978). Limit theorems for weakly exchangeable arrays. *Math. Proc. Cambridge Philos. Soc.* **84** 123–130.

[4] HÄUSLER, E. and LUSCHGY, H. (2015). *Stable convergence and stable limit theorems*. Springer.

[5] KALLENBERG, O. (2001). *Foundations of Modern Probability*, 2nd ed. Springer.

[6] LOVÁSZ, L. and SZEGEDY, B. (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** 933–957.

[7] ROSS, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8** 210–293.