Uniform Inference for Kernel Density Estimators with Dyadic Data

Supplemental Appendix

Matias D. Cattaneo[1] Yingjie Feng[2] William G. Underwood[1*]

October 13, 2023

Contents

[1]Department of Operations Research and Financial Engineering, Princeton University
[2]School of Economics and Management, Tsinghua University
[*]Corresponding author: wgu2@princeton.edu

SA1 Introduction

We describe the setup of dyadic density estimation, define the dyadic kernel density estimator, give some notation and state our assumptions.

SA1.1 Setup and estimator

Fix $n \geq 2$ and suppose there is a probability space carrying the latent random variables $\mathbf{A}_n = (A_i : 1 \leq i \leq n)$ and $\mathbf{V}_n = (V_{ij} : 1 \leq i < j \leq n)$. Suppose that the A_i are i.i.d., the V_{ij} are i.i.d., and that $\mathbf{A}_n$ is independent of $\mathbf{V}_n$. Define the observable dyadic random variables $W_{ij} = W(A_i, A_j, V_{ij})$ where W is some unknown real-valued function. There are $\frac{1}{2}n(n-1)$ such variables; one for each unordered pair of distinct indices $i < j$. Note that if $i < j$ and $i' < j'$ are all distinct, then W_{ij} is independent of $W_{i'j'}$. However, W_{ij} is not in general independent of $W_{ij'}$, as they may both depend on the latent variable A_i. This data generating process is justified by the Aldous–Hoover representation theorem for exchangeable arrays (Aldous, 1981; Hoover, 1979).

Denote the kernel weight of the data point W_{ij} at the evaluation point w with bandwidth h (see Section SA1.3 for details) by $k_h(W_{ij}, w)$. Then the dyadic kernel density estimator is defined as

$$\widehat{f}_W(w) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_h(W_{ij}, w).$$

SA1.2 Notation

SA1.2.1 Norms

For real vectors, $\|\cdot\|_p$ is the standard L^p norm defined for $p \in [1, \infty]$. For real square matrices, $\|\cdot\|_p$ is the operator norm induced by the corresponding vector norm. In particular, $\|\cdot\|_1$ is the maximum absolute column sum, $\|\cdot\|_\infty$ is the maximum absolute row sum, and $\|\cdot\|_2$ is the maximum singular value. For real symmetric matrices, $\|\cdot\|_2$ coincides with the maximum absolute eigenvalue. We use $\|\cdot\|_{\max}$ to denote the largest absolute entry of a real matrix. For real-valued functions, $\|\cdot\|_\infty$ denotes the (essential) supremum norm. The total variation norm of a real-valued function of a single real variable is $\|g\|_{\mathrm{TV}} = \sup_{n \geq 1} \sup_{x_1 \leq \cdots \leq x_n} \sum_{i=1}^{n-1} |g(x_{i+1}) - g(x_i)|$.

SA1.2.2 Inequalities

For deterministic non-negative sequences a_n and b_n, write $a_n \lesssim b_n$ or $a_n = O(b_n)$ to indicate that there exists a positive constant C which does not depend on n (although might depend on other quantities, depending on context) satisfying $a_n \leq C b_n$ for all sufficiently large n. Write $a_n \ll b_n$ or $a_n = o(b_n)$ to indicate that $a_n / b_n \to 0$. If $a_n \lesssim b_n \lesssim a_n$, write $a_n \asymp b_n$. For random non-negative sequences A_n and B_n, write $A_n \lesssim_{\mathbb{P}} B_n$ or $A_n = O_{\mathbb{P}}(B_n)$ to indicate that for any $\varepsilon > 0$ there exists a deterministic positive constant C_ε satisfying $\mathbb{P}(A_n \leq C_\varepsilon B_n) \geq 1 - \varepsilon$ for all sufficiently large n. Write $A_n = o_{\mathbb{P}}(B_n)$ if $A_n / B_n \to 0$ in probability.

SA1.2.3 Sets

For $x \in \mathbb{R}$ and $a \geq 0$, we use $[x \pm a]$ to denote the compact interval $[x - a, x + a]$. For a bounded set $\mathcal{X} \subseteq \mathbb{R}$ and $a \geq 0$ we use $[\mathcal{X} \pm a]$ to denote the compact interval $[\inf \mathcal{X} - a,\ \sup \mathcal{X} + a]$. For measurable subsets of $\mathbb{R}^d$ we use Leb to denote the Lebesgue measure, and for finite sets we use $|\cdot|$ for the cardinality.

SA1.2.4 Sums

We use $\sum_i$ to indicate $\sum_{i=1}^{n}$ when clear from context. Similarly we use $\sum_{i<j}$ for $\sum_{i=1}^{n-1} \sum_{j=i+1}^{n}$ and $\sum_{i<j<r}$ for $\sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{r=j+1}^{n}$.

SA1.2.5 Function classes

Let $\mathcal{X} \subseteq \mathbb{R}$ be an interval and $\beta > 0$. Define $\underline{\beta}$ as the largest integer which is *strictly* smaller than β. Let $\mathcal{C}^{\underline{\beta}}(\mathcal{X})$ be the class of functions from $\mathbb{R}$ to $\mathbb{R}$ which are $\underline{\beta}$ times continuously differentiable on $\mathcal{X}$. Note that $\mathcal{C}^0(\mathcal{X})$ is the class of functions which are continuous on $\mathcal{X}$. For $C > 0$, define the Hölder class with smoothness $\beta > 0$ by

$$\mathcal{H}^{\beta}_{C}(\mathcal{X}) = \Big\{ g \in \mathcal{C}^{\underline{\beta}}(\mathcal{X}) : \max_{1 \le r \le \underline{\beta}} \big|g^{(r)}(x)\big| \le C \text{ and } \big|g^{(\underline{\beta})}(x) - g^{(\underline{\beta})}(x')\big| \le C|x - x'|^{\beta - \underline{\beta}}, \quad \text{for all } x, x' \in \mathcal{X} \Big\}.$$

Note that $\mathcal{H}^{1}_{C}(\mathcal{X})$ is the class of functions which are C-Lipschitz on $\mathcal{X}$, and observe that the functions in $\mathcal{H}^{\beta}_{C}(\mathcal{X})$ are not uniformly bounded on $\mathcal{X}$.

SA1.3 Assumptions

Assumption SA1 (Data generation)
Fix $n \ge 2$ and let $\mathbf{A}_n = (A_i : 1 \le i \le n)$ be i.i.d. real-valued random variables supported on $\mathcal{A} \subseteq \mathbb{R}$. Let $\mathbf{V}_n = (V_{ij} : 1 \le i < j \le n)$ be i.i.d. real-valued random variables with a Lebesgue density f_V on $\mathbb{R}$. Suppose that $\mathbf{A}_n$ is independent of $\mathbf{V}_n$. Let $W_{ij} = W(A_i, A_j, V_{ij})$ and $\mathbf{W}_n = (W_{ij} : 1 \le i < j \le n)$, where W is some unknown real-valued function which is symmetric in its first two arguments. Let $\mathcal{W} \subseteq \mathbb{R}$ be a compact interval with positive Lebesgue measure $\mathrm{Leb}(\mathcal{W})$. *Assume that the conditional distribution of W_{ij} given A_i and A_j admits a Lebesgue density denoted $f_{W|AA}(w \mid A_i, A_j)$, and define $f_{W|A}(w \mid a) = \mathbb{E}\big[f_{W|AA}(w \mid A_i, a)\big]$ and $f_W(w) = \mathbb{E}\big[f_{W|AA}(w \mid A_i, A_j)\big]$. Take $C_{\mathrm{H}} > 0$ and $\beta \ge 1$, and suppose that $f_W \in \mathcal{H}^{\beta}_{C_{\mathrm{H}}}(\mathcal{W})$ and that $f_{W|AA}(\cdot \mid a, a') \in \mathcal{H}^{1}_{C_{\mathrm{H}}}(\mathcal{W})$ for all $a, a' \in \mathcal{A}$. Assume that $\sup_{w \in \mathcal{W}} \|f_{W|A}(w \mid \cdot)\|_{\mathrm{TV}} < \infty$.*

Remark. If $W(a_1, a_2, v)$ is strictly monotonic and continuously differentiable in its third argument, we can give the conditional density of W_{ij} explicitly using the usual change-of-variables formula: with $w = W(a_1, a_2, v)$, we have $f_{W|AA}(w \mid a_1, a_2) = f_V(v)\left|\frac{\partial W(a_1, a_2, v)}{\partial v}\right|^{-1}$.

Remark. By Lemma SA43, Assumption SA1 implies that the densities f_W, $f_{W|A}$ and $f_{W|AA}$ are all uniformly bounded by $C_{\mathrm{d}} := 2\sqrt{C_{\mathrm{H}}} + 1/\mathrm{Leb}(\mathcal{W})$.

Assumption SA2 (Kernels and bandwidth)
Let $h = h(n) > 0$ be a sequence of bandwidths satisfying $h \log n \to 0$ and $\frac{\log n}{n^2 h} \to 0$. For each $w \in \mathcal{W}$ let $k_h(\cdot, w)$ be a real-valued function supported on $[w \pm h] \cap \mathcal{W}$. Let $p \ge 1$ be an integer and suppose that k_h belongs to a family of boundary bias-corrected kernels of order p, which is to say that

$$\int_{\mathcal{W}} (s - w)^r k_h(s, w)\, \mathrm{d}s \quad \begin{cases} = 1 & \text{for all } w \in \mathcal{W} \text{ if } r = 0, \\ = 0 & \text{for all } w \in \mathcal{W} \text{ if } 1 \le r \le p - 1, \\ \neq 0 & \text{for some } w \in \mathcal{W} \text{ if } r = p. \end{cases}$$

Suppose also that for some $C_{\mathrm{L}} > 0$, the kernels satisfy $k_h(s, \cdot) \in \mathcal{H}^{1}_{C_{\mathrm{L}}/h^2}(\mathcal{W})$ for all $s \in \mathcal{W}$.

Remark. The kernel functions required by Assumption SA2 can be constructed using polynomials on $[w \pm h] \cap \mathcal{W}$, solving a family of linear systems to find the coefficients.

Remark. By Lemma SA43, Assumption SA2 implies that if $h \le 1$ then k_h is uniformly bounded by C_{k}/h where $C_{\mathrm{k}} := 2C_{\mathrm{L}} + 1 + 1/\mathrm{Leb}(\mathcal{W})$.

SA2 Main results

SA2.1 Bias

Lemma SA1 is a standard result in kernel density estimation with boundary bias correction, and does not rely on the dyadic structure of the data.

Lemma SA1 (Bias of $\widehat{f}_W$)
Suppose that Assumptions SA1 and SA2 hold. For $w \in \mathcal{W}$ define the leading bias term as

$$B_p(w) = \frac{f_W^{(p)}(w)}{p!} \int_{\mathcal{W}} k_h(s,w) \left(\frac{s-w}{h}\right)^p \, \mathrm{d}s.$$

for $1 \le p \le \underline{\beta}$. Then we have the following bias bounds.

(i) If $p \le \underline{\beta} - 1$,

$$\sup_{w \in \mathcal{W}} \left| \mathbb{E}\big[\widehat{f}_W(w)\big] - f_W(w) - h^p B_p(w) \right| \le \frac{2C_{\mathrm{k}} C_{\mathrm{H}}}{(p+1)!} h^{p+1}.$$

(ii) If $p = \underline{\beta}$,

$$\sup_{w \in \mathcal{W}} \left| \mathbb{E}\big[\widehat{f}_W(w)\big] - f_W(w) - h^p B_p(w) \right| \le \frac{2C_{\mathrm{k}} C_{\mathrm{H}}}{\underline{\beta}!} h^{\beta}.$$

(iii) If $p \ge \underline{\beta} + 1$,

$$\sup_{w \in \mathcal{W}} \left| \mathbb{E}\big[\widehat{f}_W(w)\big] - f_W(w) \right| \le \frac{2C_{\mathrm{k}} C_{\mathrm{H}}}{\underline{\beta}!} h^{\beta}.$$

Noting that $\sup_{\mathcal{W}} |B_p(w)| \le 2C_{\mathrm{k}} C_{\mathrm{H}}/p!$, we deduce that for $h \le 1$,

$$\sup_{w \in \mathcal{W}} \left| \mathbb{E}\big[\widehat{f}_W(w)\big] - f_W(w) \right| \le \frac{4C_{\mathrm{k}} C_{\mathrm{H}}}{(p \wedge \underline{\beta})!} h^{p \wedge \beta} \lesssim h^{p \wedge \beta}.$$

SA2.2 Uniform consistency

In this section we demonstrate uniform consistency of the dyadic kernel density estimator. Lemma SA2 provides a U-statistic decomposition of the estimator and Lemma SA4 employs this decomposition to establish uniform concentration. Theorem SA1 then combines this with the bias result from Lemma SA1 to show uniform consistency. Lemma SA3 provides a useful trichotomy for interpreting our results in various classes of data distributions.

Lemma SA2 (Hoeffding-type decomposition for $\widehat{f}_W$)
Suppose that Assumptions SA1 and SA2 hold. Define the linear term (Hájek projection), quadratic term and error term of $\widehat{f}_W(w)$ as

$$L_n(w) = \frac{2}{n} \sum_{i=1}^{n} l_i(w), \qquad Q_n(w) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} q_{ij}(w), \qquad E_n(w) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} e_{ij}(w)$$

respectively, where

$$\begin{aligned}
l_i(w) &= \mathbb{E}\left[k_h(W_{ij}, w) \mid A_i\right] - \mathbb{E}\left[k_h(W_{ij}, w)\right], \\
q_{ij}(w) &= \mathbb{E}\left[k_h(W_{ij}, w) \mid A_i, A_j\right] - \mathbb{E}\left[k_h(W_{ij}, w) \mid A_i\right] - \mathbb{E}\left[k_h(W_{ij}, w) \mid A_j\right] + \mathbb{E}\left[k_h(W_{ij}, w)\right], \\
e_{ij}(w) &= k_h(W_{ij}, w) - \mathbb{E}\left[k_h(W_{ij}, w) \mid A_i, A_j\right].
\end{aligned}$$

Then the following Hoeffding-type decomposition holds:

$$\widehat{f}_W(w) = \mathbb{E}\big[\widehat{f}_W(w)\big] + L_n(w) + Q_n(w) + E_n(w).$$

Further, the stochastic processes L_n, Q_n and E_n are all mean-zero, since

$$\mathbb{E}\big[L_n(w)\big] = \mathbb{E}\big[Q_n(w)\big] = \mathbb{E}\big[E_n(w)\big] = 0$$

for all $w \in \mathcal{W}$. Also they are mutually orthogonal in $L^2(\mathbb{P})$ as

$$\mathbb{E}\big[L_n(w) Q_n(w')\big] = \mathbb{E}\big[L_n(w) E_n(w')\big] = \mathbb{E}\big[Q_n(w) E_n(w')\big] = 0$$

for all $w, w' \in \mathcal{W}$.

Lemma SA3 (Trichotomy of degeneracy)
Suppose that Assumptions SA1 and SA2 hold, and define the non-negative upper and lower degeneracy constants

$$D_{\text{up}}^2 = \sup_{w \in \mathcal{W}} \text{Var}\left[f_{W|A}(w \mid A_i)\right], \qquad D_{\text{lo}}^2 = \inf_{w \in \mathcal{W}} \text{Var}\left[f_{W|A}(w \mid A_i)\right]$$

respectively. Then precisely one of the following three statements must hold.

(i) Total degeneracy: $D_{\text{up}} = D_{\text{lo}} = 0$. *Then* $L_n(w) = 0$ *for all* $w \in \mathcal{W}$ *almost surely.*

(ii) No degeneracy: $D_{\text{lo}} > 0$. *Then* $\inf_{w \in \mathcal{W}} \text{Var}[L_n(w)] \geq \frac{2D_{\text{lo}}}{n}$ *for all large enough* n.

(iii) Partial degeneracy: $D_{\text{up}} > D_{\text{lo}} = 0$. *There exists* $w \in \mathcal{W}$ *with* $\text{Var}\left[f_{W|A}(w \mid A_i)\right] = 0$*; such a point is labelled* degenerate *and satisfies* $\text{Var}[L_n(w)] \leq 64 C_{\text{k}} C_{\text{H}} C_{\text{d}} \frac{h}{n}$. *There also exists a point* $w' \in \mathcal{W}$ *with* $\text{Var}\left[f_{W|A}(w' \mid A_i)\right] > 0$*; such a point is labelled* non-degenerate *and satisfies* $\text{Var}[L_n(w')] \geq \frac{2}{n} \text{Var}\left[f_{W|A}(w' \mid A_i)\right]$ *for all large enough* n.

Remark. The trichotomy of total/partial/no degeneracy given in Lemma SA3 is useful for understanding the asymptotic behavior of the dyadic kernel density estimator. Note that our need for uniformity in w complicates the simpler degeneracy/no degeneracy dichotomy observed for pointwise results by Graham et al. (2022).

Lemma SA4 (Uniform concentration of $\widehat{f}_W$)
Suppose Assumptions SA1 and SA2 hold. Then with L_n, Q_n *and* E_n *defined as in Lemma SA2, we have*

$$\mathbb{E}\left[\sup_{w \in \mathcal{W}} |L_n(w)|\right] \lesssim \frac{D_{\text{up}}}{\sqrt{n}}, \qquad \mathbb{E}\left[\sup_{w \in \mathcal{W}} |Q_n(w)|\right] \lesssim \frac{1}{n}, \qquad \mathbb{E}\left[\sup_{w \in \mathcal{W}} |E_n(w)|\right] \lesssim \sqrt{\frac{\log n}{n^2 h}},$$

where $\lesssim$ *is up to constants which depend on the underlying data distribution and the choice of kernel. Note that the* Q_n *term is dominated by the* L_n *term uniformly in the bandwidth* h. *Therefore by Lemma SA2*

$$\mathbb{E}\left[\sup_{w \in \mathcal{W}} \left|\widehat{f}_W(w) - \mathbb{E}[\widehat{f}_W(w)]\right|\right] \lesssim \frac{D_{\text{up}}}{\sqrt{n}} + \sqrt{\frac{\log n}{n^2 h}}.$$

Theorem SA1 (Uniform consistency of $\widehat{f}_W$)
Suppose Assumptions SA1 and SA2 hold. Then

$$\mathbb{E}\left[\sup_{w \in \mathcal{W}} \left|\widehat{f}_W(w) - f_W(w)\right|\right] \lesssim h^{p \wedge \beta} + \frac{D_{\text{up}}}{\sqrt{n}} + \sqrt{\frac{\log n}{n^2 h}},$$

where $\lesssim$ *is up to constants which depend on the underlying data distribution and the choice of kernel.*

Remark. In light of the degeneracy trichotomy in Lemma SA3, we interpret Theorem SA1.

(i) Partial or no degeneracy: when $D_{\text{up}} > 0$, any bandwidth sequence satisfying $\frac{\log n}{n} \lesssim h \lesssim n^{-\frac{1}{2(p \wedge \beta)}}$ gives the bandwidth-independent "parametric" rate noted by Graham et al. (2022):

$$\mathbb{E}\left[\sup_{w \in \mathcal{W}} \left|\widehat{f}_W(w) - \mathbb{E}[\widehat{f}_W(w)]\right|\right] \lesssim \frac{1}{\sqrt{n}},$$

(ii) Total degeneracy: when $D_{\text{up}} = 0$, minimizing the upper bound by setting $h \asymp \left(\frac{\log n}{n^2}\right)^{\frac{1}{2(p \wedge \beta)+1}}$ yields

$$\mathbb{E}\left[\sup_{w \in \mathcal{W}} \left|\widehat{f}_W(w) - f_W(w)\right|\right] \lesssim h^{p \wedge \beta} + \sqrt{\frac{\log n}{n^2 h}} \lesssim \left(\frac{\log n}{n^2}\right)^{\frac{p \wedge \beta}{2(p \wedge \beta)+1}},$$

SA2.3 Minimax optimality

In this section we demonstrate minimax optimality of our estimator under uniform convergence by providing upper and lower bounds in expectation uniformly over some classes of dyadic distributions.

Theorem SA2 (Minimax optimality of $\widehat{f}_W$)
Fix $\beta \geq 1$ and $C_{\mathrm{H}} > 0$, and take $\mathcal{W}$ a compact interval with positive Lebesgue measure. Define $\mathcal{P} = \mathcal{P}(\mathcal{W}, \beta, C_{\mathrm{H}})$ as the class of dyadic distributions satisfying Assumption SA1. Define $\mathcal{P}_{\mathrm{d}}$ as the subclass of $\mathcal{P}$ containing only those distributions which are totally degenerate on $\mathcal{W}$ in the sense that $\sup_{w\in\mathcal{W}} \operatorname{Var}\left[f_{W|A}(w \mid A_i)\right] = 0$. Then

$$\inf_{\widetilde{f}_W} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\sup_{w\in\mathcal{W}} \left|\widetilde{f}_W(w) - f_W(w)\right|\right] \asymp \frac{1}{\sqrt{n}},$$

$$\inf_{\widetilde{f}_W} \sup_{\mathbb{P}\in\mathcal{P}_{\mathrm{d}}} \mathbb{E}_{\mathbb{P}}\left[\sup_{w\in\mathcal{W}} \left|\widetilde{f}_W(w) - f_W(w)\right|\right] \asymp \left(\frac{\log n}{n^2}\right)^{\frac{\beta}{2\beta+1}},$$

where $\widetilde{f}_W$ is any estimator depending only on the data $\mathbf{W}_n = (W_{ij} : 1 \leq i < j \leq n)$ distributed according to the dyadic law $\mathbb{P}$. The constants in $\asymp$ depend only on $\mathcal{W}$, β and C_{H}.

Remark. Theorem SA2 verifies that the rates of uniform consistency derived in Theorem SA1 are minimax-optimal when using a kernel of sufficiently high order ($p \geq \beta$). It also shows that both the L_n and E_n terms are important in the consistency of $\widehat{f}_W$, and that their relative magnitude in the supremum norm is determined by the degeneracy type of the underlying distribution.

SA2.4 Covariance structure

Lemma SA5 (Covariance structure)
Suppose Assumptions SA1 and SA2 hold. Define the covariance function of the dyadic kernel density estimator by

$$\Sigma_n(w, w') = \operatorname{Cov}\left[\widehat{f}_W(w), \widehat{f}_W(w')\right]$$

for $w, w' \in \mathcal{W}$. Then Σ_n admits the following representations.

$$\begin{aligned}\Sigma_n(w, w') &= \frac{2}{n(n-1)} \operatorname{Cov}\big[k_h(W_{ij}, w), k_h(W_{ij}, w')\big] + \frac{4(n-2)}{n(n-1)} \operatorname{Cov}\big[k_h(W_{ij}, w), k_h(W_{ir}, w')\big] \\ &= \frac{2}{n(n-1)} \operatorname{Cov}\big[k_h(W_{ij}, w), k_h(W_{ij}, w')\big] + \frac{4(n-2)}{n(n-1)} \operatorname{Cov}\big[\mathbb{E}[k_h(W_{ij}, w) \mid A_i], \mathbb{E}[k_h(W_{ij}, w') \mid A_i]\big],\end{aligned}$$

where $1 \leq i < j < r \leq n$.

Lemma SA6 (Variance bounds)
Suppose that Assumptions SA1 and SA2 hold. Then for all large enough n,

$$\frac{D_{\mathrm{lo}}^2}{n} + \frac{1}{n^2 h} \inf_{w\in\mathcal{W}} f_W(w) \lesssim \inf_{w\in\mathcal{W}} \Sigma_n(w, w) \leq \sup_{w\in\mathcal{W}} \Sigma_n(w, w) \lesssim \frac{D_{\mathrm{up}}^2}{n} + \frac{1}{n^2 h}.$$

SA2.5 Strong approximation

In this section we give a strong approximation for the empirical process $\widehat{f}_W$. We begin by using the Kómlos–Major–Tusnády (KMT) approximation to obtain a strong approximation for L_n in Lemma SA7. Since E_n is an empirical process of i.n.i.d. variables, the KMT approximation is not valid. Instead we apply a conditional version of Yurinskii's coupling to obtain a conditional strong approximation for E_n in Lemma SA8, and then construct an unconditional strong approximation for E_n in Lemma SA9. These approximations are combined to give a strong approximation for $\widehat{f}_W$ in Theorem SA3. We do not need to construct a strong approximation for the negligible Q_n.

This section is largely concerned with distributional properties, and as such will frequently involve *copies* of processes. We say that X' is a copy of a random variable X if they have the same distribution, though they may be defined on different probability spaces. To ensure that all of the joint distributional properties of such processes are preserved, we also carry over a copy of the latent variables $(\mathbf{A}_n, \mathbf{V}_n)$ to the new space.

Many of the technical details regarding the copying and embedding of stochastic processes are covered by the Vorob'ev–Berkes–Philipp Theorem, which is stated and discussed in Lemma SA28. In particular, this theorem can be used for random vectors or for stochastic processes indexed by a compact rectangle in $\mathbb{R}^d$ with a.s. continuous sample paths.

Lemma SA7 (Strong approximation of L_n)
Suppose that Assumptions SA1 and SA2 hold. For each $n \geq 2$ there exists on some probability space a copy of $\left(\mathbf{A}_n, \mathbf{V}_n, L_n\right)$, denoted $\left(\mathbf{A}'_n, \mathbf{V}'_n, L'_n\right)$, and a mean-zero Gaussian process $Z_n^{L'}$ indexed on $\mathcal{W}$ satisfying

$$\mathbb{P}\left(\sup_{w\in\mathcal{W}} \left|\sqrt{n}L'_n(w) - Z_n^{L'}(w)\right| > D_{\text{up}}\frac{t + C_1 \log n}{\sqrt{n}}\right) \leq C_2 e^{-C_3 t},$$

for some positive constants C_1, C_2, C_3 and for all $t > 0$. By integration of tail probabilities,

$$\mathbb{E}\left[\sup_{w\in\mathcal{W}} \left|\sqrt{n}L'_n(w) - Z_n^{L'}(w)\right|\right] \lesssim \frac{D_{\text{up}} \log n}{\sqrt{n}}.$$

Further, $Z_n^{L'}$ has the same covariance structure as $\sqrt{n}L'_n$ in the sense that for all $w, w' \in \mathcal{W}$,

$$\mathbb{E}\left[Z_n^{L'}(w) Z_n^{L'}(w')\right] = n\mathbb{E}\left[L'_n(w) L'_n(w')\right].$$

It also satisfies the following trajectory regularity property for any $\delta_n \in (0, 1/2]$:

$$\mathbb{E}\left[\sup_{|w-w'|\leq\delta_n} \left|Z_n^{L'}(w) - Z_n^{L'}(w')\right|\right] \lesssim D_{\text{up}}\delta_n\sqrt{\log 1/\delta_n},$$

and has continuous trajectories. The process $Z_n^{L'}$ is a function only of $\mathbf{A}'_n$ and some random noise which is independent of $(\mathbf{A}'_n, \mathbf{V}'_n)$.

Lemma SA8 (Conditional strong approximation of E_n)
Suppose that Assumptions SA1 and SA2 hold. For each $n \geq 2$ and $t_n > 0$ with $|\log t_n| \lesssim \log n$, there exists on some probability space a copy of $\left(\mathbf{A}_n, \mathbf{V}_n, E_n\right)$, denoted $\left(\mathbf{A}'_n, \mathbf{V}'_n, E'_n\right)$, and a process $\widetilde{Z}_n^{E'}$ which is Gaussian conditional on $\mathbf{A}'_n$ and mean-zero conditional on $\mathbf{A}'_n$, satisfying

$$\mathbb{P}\left(\sup_{w\in\mathcal{W}} \left|\sqrt{n^2h}E'_n(w) - \widetilde{Z}_n^{E'}(w)\right| > t_n \;\middle|\; \mathbf{A}'_n\right) \leq C_1 t_n^{-2} n^{-1/2} h^{-3/4} (\log n)^{3/4},$$

$\mathbf{A}'_n$-almost surely for some constant $C_1 > 0$. Setting $t_n = n^{-1/4}h^{-3/8}(\log n)^{3/8}R_n$ for any sequence $R_n \to \infty$ and taking an expectation gives

$$\sup_{w\in\mathcal{W}} \left|\sqrt{n^2h}E'_n(w) - \widetilde{Z}_n^{E'}(w)\right| \lesssim_{\mathbb{P}} n^{-1/4}h^{-3/8}(\log n)^{3/8}R_n.$$

Further, $\widetilde{Z}_n^{E'}$ has the same conditional covariance structure as $\sqrt{n^2h}E'_n$ in the sense that for all $w, w' \in \mathcal{W}$,

$$\mathbb{E}\left[\widetilde{Z}_n^{E'}(w)\widetilde{Z}_n^{E'}(w') \mid \mathbf{A}'_n\right] = n^2h\mathbb{E}\left[E'_n(w)E'_n(w') \mid \mathbf{A}'_n\right].$$

It also satisfies the following trajectory regularity property for any $\delta_n \in (0, 1/(2h)]$:

$$\mathbb{E}\left[\sup_{|w-w'|\leq\delta_n} \left|\widetilde{Z}_n^{E'}(w) - \widetilde{Z}_n^{E'}(w')\right|\right] \lesssim \frac{\delta_n}{h}\sqrt{\log\frac{1}{h\delta_n}},$$

and has continuous trajectories.

Lemma SA9 (Unconditional strong approximation of E_n)
Suppose that Assumptions SA1 and SA2 hold. Let $(\mathbf{A}'_n, \mathbf{V}'_n, \widetilde{Z}_n^{E\prime})$ *be defined as in Lemma SA8. For each* $n \geq 2$ *there exists (on some probability space) a copy of* $(\mathbf{A}'_n, \mathbf{V}'_n, \widetilde{Z}_n^{E\prime})$*, denoted* $(\mathbf{A}''_n, \mathbf{V}''_n, \widetilde{Z}_n^{E\prime\prime})$*, and a centered Gaussian process* $Z_n^{E\prime\prime}$ *satisfying*

$$\mathbb{E}\left[\sup_{w\in\mathcal{W}}\left|\widetilde{Z}_n^{E\prime\prime}(w) - Z_n^{E\prime\prime}(w)\right|\right] \lesssim n^{-1/6}(\log n)^{2/3}.$$

Further, $Z_n^{E\prime\prime}$ *has the same (unconditional) covariance structure as* $\widetilde{Z}_n^{E\prime\prime}$ *and* $\sqrt{n^2h}E_n$ *in the sense that for all* $w, w' \in \mathcal{W}$*,*

$$\mathbb{E}\left[Z_n^{E\prime\prime}(w)Z_n^{E\prime\prime}(w')\right] = \mathbb{E}\left[\widetilde{Z}_n^{E\prime\prime}(w)\widetilde{Z}_n^{E\prime\prime}(w')\right] = n^2h\,\mathbb{E}\left[E_n(w)E_n(w')\right].$$

It also satisfies the following trajectory regularity property for any $\delta_n \in (0, 1/(2h)]$*:*

$$\mathbb{E}\left[\sup_{|w-w'|\leq\delta_n}\left|Z_n^{E\prime\prime}(w) - Z_n^{E\prime\prime}(w')\right|\right] \lesssim \frac{\delta_n}{h}\sqrt{\log\frac{1}{h\delta_n}}.$$

Finally, $Z_n^{E\prime\prime}$ *is independent of* $\mathbf{A}''_n$ *and has continuous trajectories.*

Remark. Note that the process $\widetilde{Z}_n^{E\prime}$, constructed in Lemma SA8, is a conditionally Gaussian process but is not in general a Gaussian process. The process $Z_n^{E\prime\prime}$, constructed in Lemma SA9, is a true Gaussian process.

Theorem SA3 (Strong approximation of $\widehat{f}_W$)
Suppose that Assumptions SA1 and SA2 hold. For each $n \geq 2$ *and any sequence* $R_n \to \infty$ *there exists on some probability space a centered Gaussian process* $Z_n^{f\prime}$ *and a copy of* $\widehat{f}_W$*, denoted* $\widehat{f}'_W$*, satisfying*

$$\sup_{w\in\mathcal{W}}\left|\widehat{f}'_W(w) - \mathbb{E}[\widehat{f}'_W(w)] - Z_n^{f\prime}(w)\right| \lesssim_{\mathbb{P}} n^{-1}\log n + n^{-5/4}h^{-7/8}(\log n)^{3/8}R_n + n^{-7/6}h^{-1/2}(\log n)^{2/3}.$$

Further, $Z_n^{f\prime}$ *has the same covariance structure as* $\widehat{f}'_W(w)$ *in the sense that for all* $w, w' \in \mathcal{W}$*,*

$$\mathbb{E}\left[Z_n^{f\prime}(w)Z_n^{f\prime}(w')\right] = \mathrm{Cov}\left[\widehat{f}'_W(w), \widehat{f}'_W(w')\right] = \Sigma_n(w, w').$$

It also has continuous trajectories satisfying the following trajectory regularity property for any $\delta_n \in (0, 1/2]$*:*

$$\mathbb{E}\left[\sup_{|w-w'|\leq\delta_n}\left|Z_n^{f\prime}(w) - Z_n^{f\prime}(w')\right|\right] \lesssim \frac{D_{\mathrm{up}}}{\sqrt{n}}\delta_n\sqrt{\log\frac{1}{\delta_n}} + \frac{1}{\sqrt{n^2h}}\frac{\delta_n}{h}\sqrt{\log\frac{1}{h\delta_n}}.$$

Remark. The interpretation of Theorem SA3 is deferred to Section SA2.6, in which we scale the processes by their pointwise variance in order to better understand the role of degeneracy on strong approximation rates.

SA2.6 Infeasible uniform confidence bands

We use the strong approximation and bias results from Theorem SA3 and Lemma SA1 respectively to construct uniform confidence bands for the true density function f_W. From now on we will drop the prime notation for copies of processes in the interest of clarity. In this section we will assume oracle knowledge of the true covariance function Σ_n, which is not typically available in practice. For feasible versions of these results which use a covariance estimator, see Section SA2.9. We also assume that the true density f_W is bounded away from zero on the domain of inference, which is a standard assumption when constructing confidence bands.

Lemma SA10 (Infeasible Gaussian approximation of the standardized t-statistic)
Let Assumptions SA1 and SA2 hold and suppose that $f_W(w) > 0$ *on* $\mathcal{W}$*. Define for* $w \in \mathcal{W}$

$$T_n(w) = \frac{\widehat{f}_W(w) - f_W(w)}{\sqrt{\Sigma_n(w,w)}} \quad \text{and} \quad Z_n^T(w) = \frac{Z_n^f(w)}{\sqrt{\Sigma_n(w,w)}}.$$

Then with $R_n \to \infty$ as in Theorem SA3,

$$\sup_{w\in\mathcal{W}} \left|T_n(w) - Z_n^T(w)\right| \lesssim_{\mathbb{P}} \frac{n^{-1/2}\log n + n^{-3/4}h^{-7/8}(\log n)^{3/8}R_n + n^{-2/3}h^{-1/2}(\log n)^{2/3} + n^{1/2}h^{p\wedge\beta}}{D_{\text{lo}} + 1/\sqrt{nh}}.$$

For the coverage rate error in the upcoming Theorem SA4 to converge to zero in large samples, we require further restrictions on the bandwidth sequence. These restrictions depend on the degeneracy type of the dyadic distribution, and are given as Assumption SA3.

Assumption SA3 (Rate restriction for uniform confidence bands)
Suppose that one of the following holds.

(i) No degeneracy: $D_{\text{lo}} > 0$ and $n^{-6/7}\log n \ll h \ll (n\log n)^{-\frac{1}{2(p\wedge\beta)}}$.

(ii) Partial or total degeneracy: $D_{\text{lo}} = 0$ and $n^{-2/3}(\log n)^{7/3} \ll h \ll (n^2\log n)^{-\frac{1}{2(p\wedge\beta)+1}}$.

Theorem SA4 (Infeasible uniform confidence bands)
Let Assumptions SA1, SA2 and SA3 hold and suppose that $f_W(w) > 0$ on $\mathcal{W}$. Let $\alpha \in (0,1)$ be a confidence level and define $q_{1-\alpha}$ as the quantile satisfying

$$\mathbb{P}\left(\sup_{w\in\mathcal{W}} \left|Z_n^T(w)\right| \le q_{1-\alpha}\right) = 1-\alpha.$$

Then

$$\mathbb{P}\left(f_W(w) \in \left[\widehat{f}_W(w) \pm q_{1-\alpha}\sqrt{\Sigma_n(w,w)}\right] \text{ for all } w \in \mathcal{W}\right) \to 1-\alpha.$$

Remark. By Theorem SA1, the asymptotically optimal bandwidth choice under uniform convergence is given by $h \asymp (n^{-2}\log n)^{\frac{1}{2(p\wedge\beta)+1}}$. This bandwidth satisfies Assumption SA3 only in the case of no degeneracy. Thus in degenerate cases, one must undersmooth by choosing a bandwidth which is smaller than the optimal bandwidth. This robust bias correction can be achieved in practice by selecting an approximately optimal bandwidth for a kernel of order p, but then using a kernel of higher order $p' > p$ for constructing the confidence bands.

SA2.7 Covariance estimation

In this section we provide a consistent estimator for the covariance function Σ_n. In Lemma SA11 we define the estimator and demonstrate that it converges in probability in a suitable sense In Lemma SA12 we give an alternative representation which is more amenable to computation.

Lemma SA11 (Covariance estimation)
Let Assumptions SA1 and SA2 hold and suppose that $nh \gtrsim \log n$ and $f_W(w) > 0$ on $\mathcal{W}$. For $w, w' \in \mathcal{W}$ define

$$\widehat{\Sigma}_n(w,w') = \frac{4}{n^2(n-1)^2}\sum_{i<j} k_h(W_{ij},w)k_h(W_{ij},w') + \frac{24}{n^2(n-1)^2}\sum_{i<j<r} S_{ijr}(w,w') - \frac{4n-6}{n(n-1)}\widehat{f}_W(w)\widehat{f}_W(w'),$$

where

$$\begin{aligned} S_{ijr}(w,w') = \frac{1}{6}\Big(&k_h(W_{ij},w)k_h(W_{ir},w') + k_h(W_{ij},w)k_h(W_{jr},w') + k_h(W_{ir},w)k_h(W_{ij},w') \\ &+ k_h(W_{ir},w)k_h(W_{jr},w') + k_h(W_{jr},w)k_h(W_{ij},w') + k_h(W_{jr},w)k_h(W_{ir},w')\Big). \end{aligned}$$

Then $\widehat{\Sigma}_n$ is uniformly entrywise-consistent in the sense that

$$\sup_{w,w'\in\mathcal{W}} \left|\frac{\widehat{\Sigma}_n(w,w') - \Sigma_n(w,w')}{\sqrt{\Sigma_n(w,w) + \Sigma_n(w',w')}}\right| \lesssim_{\mathbb{P}} \frac{\sqrt{\log n}}{n}.$$

Lemma SA12 (Alternative covariance estimator representation)
Suppose that Assumptions SA1 and SA2 hold and let $\widehat{\Sigma}_n$ be the covariance estimator defined in Lemma SA11. Then the following alternative representation for $\widehat{\Sigma}_n$ holds, which may be easier to compute as it does not involve any triple summations over the data.

$$\widehat{\Sigma}_n(w,w') = \frac{4}{n^2}\sum_{i=1}^{n} S_i(w)S_i(w') - \frac{4}{n^2(n-1)^2}\sum_{i<j} k_h(W_{ij},w)k_h(W_{ij},w') - \frac{4n-6}{n(n-1)}\widehat{f}_W(w)\widehat{f}_W(w'),$$

where

$$S_i(w) = \frac{1}{n-1}\left(\sum_{j=1}^{i-1} k_h(W_{ji},w) + \sum_{j=i+1}^{n} k_h(W_{ij},w)\right)$$

is an estimator of $\mathbb{E}[k_h(W_{ij},w) \mid A_i]$.

Remark. The covariance estimator $\widehat{\Sigma}_n$ is not necessarily almost surely positive semi-definite.

SA2.8 Positive semi-definite covariance estimation

In this section we provide a positive semi-definite estimator $\widehat{\Sigma}_n^+$ which is uniformly entrywise-consistent for Σ_n. Define $\widehat{\Sigma}_n$ as in Lemma SA11 and consider the following optimization problem over bivariate functions.

$$\begin{aligned} \text{minimize:} \quad & \sup_{w,w'\in\mathcal{W}} \left| \frac{M(w,w') - \widehat{\Sigma}_n(w,w')}{\sqrt{\widehat{\Sigma}_n(w,w) + \widehat{\Sigma}_n(w',w')}} \right| \quad \text{over } M : \mathcal{W}\times\mathcal{W} \to \mathbb{R} \\ \text{subject to:} \quad & M \text{ is symmetric and positive semi-definite,} \\ & \left|M(w,w') - M(w,w'')\right| \le \frac{4}{nh^3} C_{\mathrm{k}} C_{\mathrm{L}} |w' - w''| \text{ for all } w,w',w'' \in \mathcal{W}. \end{aligned} \tag{1}$$

Lemma SA13 (Consistency of $\widehat{\Sigma}_n^+$)
Suppose that Assumptions SA1 and SA2 hold and that $nh \gtrsim \log n$ and $f_W(w) > 0$ on $\mathcal{W}$. Then the optimization problem (1) has an approximately optimal solution $\widehat{\Sigma}_n^+$ which is uniformly entrywise-consistent for Σ_n in the sense that

$$\sup_{w,w'\in\mathcal{W}} \left| \frac{\widehat{\Sigma}_n^+(w,w') - \Sigma_n(w,w')}{\sqrt{\Sigma_n(w,w) + \Sigma_n(w',w')}} \right| \lesssim_{\mathbb{P}} \frac{\sqrt{\log n}}{n}.$$

Remark. The optimization problem (1) is stated for functions rather than for matrices so is infinite-dimensional. However, when restricting to finite-size matrices, Lemma SA13 still holds and does not depend on the size of the matrices. Furthermore, the problem then becomes a semi-definite program and so can be solved to arbitrary precision in polynomial time in the size of the matrices (Laurent and Rendl, 2005).

The Lipschitz-type constraint in the optimization problem (1) ensures that $\widehat{\Sigma}_n^+$ is sufficiently smooth and is a technicality required by some of the later proofs. In practice this constraint is readily verified.

Lemma SA14 (Positive semi-definite variance estimator bounds)
Suppose that Assumptions SA1 and SA2 hold and that $nh \gtrsim \log n$ and $f_W(w) > 0$ on $\mathcal{W}$. Then $\widehat{\Sigma}_n^+(w,w) \ge 0$ almost surely for all $w \in \mathcal{W}$ and

$$\frac{D_{\mathrm{lo}}^2}{n} + \frac{1}{n^2h} \lesssim_{\mathbb{P}} \inf_{w\in\mathcal{W}} \widehat{\Sigma}_n^+(w,w) \le \sup_{w\in\mathcal{W}} \widehat{\Sigma}_n^+(w,w) \lesssim_{\mathbb{P}} \frac{D_{\mathrm{up}}^2}{n} + \frac{1}{n^2h}.$$

SA2.9 Feasible uniform confidence bands

Now we use the strong approximation derived in Section SA2.5 and the positive semi-definite covariance estimator introduced in Section SA2.8 to construct feasible uniform confidence bands.

Lemma SA15 (Proximity of the standardized and studentized t-statistics)
Let Assumptions SA1 and SA2 hold and suppose that $nh \gtrsim \log n$ *and* $f_W(w) > 0$ *on* $\mathcal{W}$. *Define for* $w \in \mathcal{W}$

$$\widehat{T}_n(w) = \frac{\widehat{f}_W(w) - f_W(w)}{\sqrt{\widehat{\Sigma}_n^+(w,w)}}.$$

Then

$$\sup_{w\in\mathcal{W}} \left|\widehat{T}_n(w) - T_n(w)\right| \lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{n}} \left(\sqrt{\log n} + \frac{\sqrt{n}h^{p\wedge\beta}}{D_{\mathrm{lo}} + 1/\sqrt{nh}}\right) \frac{1}{D_{\mathrm{lo}} + 1/\sqrt{nh}}.$$

Lemma SA16 (Feasible Gaussian approximation of the infeasible Gaussian process)
Let Assumptions SA1 and SA2 hold and suppose that $nh \gtrsim \log n$ *and* $f_W(w) > 0$ *on* $\mathcal{W}$. *Define a process* $\widehat{Z}_n^T(w)$ *which, conditional on the data* $\mathbf{W}_n$, *is conditionally mean-zero and conditionally Gaussian and whose conditional covariance structure is*

$$\mathbb{E}\left[\widehat{Z}_n^T(w)\widehat{Z}_n^T(w') \mid \mathbf{W}_n\right] = \frac{\widehat{\Sigma}_n^+(w,w')}{\sqrt{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')}}$$

Then the following conditional Kolmogorov–Smirnov result holds.

$$\sup_{t\in\mathbb{R}} \left|\mathbb{P}\left(\sup_{w\in\mathcal{W}} \left|Z_n^T(w)\right| \le t\right) - \mathbb{P}\left(\sup_{w\in\mathcal{W}} \left|\widehat{Z}_n^T(w)\right| \le t \;\middle|\; \mathbf{W}_n\right)\right| \lesssim_{\mathbb{P}} \frac{n^{-1/6}(\log n)^{5/6}}{D_{\mathrm{lo}}^{1/3} + (nh)^{-1/6}}.$$

Lemma SA17 (Feasible Gaussian approximation of the studentized t-statistic)
Let Assumptions SA1, SA2 and SA3 hold and suppose that $f_W(w) > 0$ *on* $\mathcal{W}$. *Then*

$$\sup_{t\in\mathbb{R}} \left|\mathbb{P}\left(\sup_{w\in\mathcal{W}} \left|\widehat{T}_n(w)\right| \le t\right) - \mathbb{P}\left(\sup_{w\in\mathcal{W}} \left|\widehat{Z}_n^T(w)\right| \le t \;\middle|\; \mathbf{W}_n\right)\right| \ll_{\mathbb{P}} 1.$$

Theorem SA5 (Feasible uniform confidence bands)
Let Assumptions SA1, SA2 and SA3 hold and suppose that $f_W(w) > 0$ *on* $\mathcal{W}$. *Let* $\alpha \in (0,1)$ *be a confidence level and define* $\widehat{q}_{1-\alpha}$ *as the conditional quantile satisfying*

$$\mathbb{P}\left(\sup_{w\in\mathcal{W}} \left|\widehat{Z}_n^T(w)\right| \le \widehat{q}_{1-\alpha} \;\middle|\; \mathbf{W}_n\right) = 1 - \alpha.$$

Then

$$\mathbb{P}\left(f_W(w) \in \left[\widehat{f}_W(w) \pm \widehat{q}_{1-\alpha}\sqrt{\widehat{\Sigma}_n^+(w,w)}\right] \text{ for all } w \in \mathcal{W}\right) \to 1 - \alpha.$$

Remark. In practice, suprema over $\mathcal{W}$ can be replaced by maxima over a sufficiently fine finite partition of $\mathcal{W}$. The conditional quantile $\widehat{q}_{1-\alpha}$ can be estimated by Monte Carlo simulation, resampling from the Gaussian process defined by the law of $\widehat{Z}_n^T \mid \mathbf{W}_n$.

SA2.10 Counterfactual dyadic density estimation

As an application we provide methodology for estimation and inference on counterfactual dyadic density functions, following the reweighting approach of DiNardo et al. (1996). We give the counterfactual data generating process as Assumption SA4.

Assumption SA4 (Counterfactual data generation)
For each $r \in \{0,1\}$, let $\mathbf{W}_n^r$, $\mathbf{A}_n^r$ and $\mathbf{V}_n^r$ be as in Assumption SA1. Let X_i^r be finitely-supported variables, setting $\mathbf{X}_n^r = (X_1^r, \dots, X_n^r)$. Suppose that (A_i^r, X_i^r) are independent over $1 \le i \le n$ and that $\mathbf{X}_n^r$ is independent of $\mathbf{V}_n^r$. Assume that $W_{ij}^r \mid X_i^r, X_j^r$ has a Lebesgue density $f_{W|XX}^r(\cdot \mid x_1, x_2) \in \mathcal{H}_{C_H}^{\beta}(\mathcal{W})$ and that X_i^r has positive probability mass function $p_X^r(x)$ on a common support $\mathcal{X}$. Suppose that $(\mathbf{A}_n^0, \mathbf{V}_n^0, \mathbf{X}_n^0)$ and $(\mathbf{A}_n^1, \mathbf{V}_n^1, \mathbf{X}_n^1)$ are independent.

The counterfactual density of W_{ij} in population 1 had X_i, X_j followed the distribution in population 0 by

$$f_W^{1\triangleright 0}(w) = \mathbb{E}\left[f_{W|XX}^1\big(w \mid X_1^0, X_2^0\big)\right] = \sum_{x_1 \in \mathcal{X}} \sum_{x_2 \in \mathcal{X}} f_{W|XX}^1(w \mid x_1, x_2)\psi(x_1)\psi(x_2)p_X^1(x_1)p_X^1(x_2)$$

where $\psi(x) = p_X^0(x)/p_X^1(x)$ for $x \in \mathcal{X}$. This counterfactual density can be estimated by the counterfactual dyadic kernel density estimator

$$\widehat{f}_W^{1\triangleright 0}(w) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \widehat{\psi}(X_i^1)\widehat{\psi}(X_j^1)k_h(W_{ij}^1, w)$$

with $\widehat{\psi}(x) = \mathbb{I}\{\widehat{p}_X^1(x) > 0\}\widehat{p}_X^0(x)/\widehat{p}_X^1(x)$ and $\widehat{p}_X^r(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{X_i^r = x\}$. Note that since $p_X^r(x) > 0$,

$$\begin{aligned}
\widehat{\psi}(x) - \psi(x) &= \frac{\widehat{p}_X^0(x) - p_X^0(x)}{p_X^1(x)} - \frac{p_X^0(x)}{p_X^1(x)}\frac{\widehat{p}_X^1(x) - p_X^1(x)}{p_X^1(x)} + \frac{\widehat{p}_X^1(x) - p_X^1(x)}{p_X^1(x)}\frac{\widehat{p}_X^1(x)p_X^0(x) - \widehat{p}_X^0(x)p_X^1(x)}{\widehat{p}_X^1(x)p_X^1(x)} \\
&= \frac{1}{n}\sum_{r=1}^{n} \kappa(X_r^0, X_r^1, x) + O_{\mathbb{P}}\left(\frac{1}{n}\right)
\end{aligned}$$

is an asymptotic linear representation where

$$\kappa(X_i^0, X_i^1, x)] = \frac{\mathbb{I}\{X_i^0 = x\} - p_X^0(x)}{p_X^1(x)} - \frac{p_X^0(x)}{p_X^1(x)}\frac{\mathbb{I}\{X_i^1 = x\} - p_X^1(x)}{p_X^1(x)}$$

satisfies $\mathbb{E}[\kappa(X_i^0, X_i^1, x)] = 0$. We now establish uniform consistency and feasible strong approximation results for the counterfactual density estimator.

Lemma SA18 (Bias of $\widehat{f}_W^{1\triangleright 0}$)
Suppose that Assumptions SA1, SA2 and SA4 hold. Then

$$\sup_{w \in \mathcal{W}} \left|\mathbb{E}\big[\widehat{f}_W^{1\triangleright 0}(w)\big] - f_W^{1\triangleright 0}(w)\right| \lesssim h^{p \wedge \beta} + \frac{1}{n}.$$

Lemma SA19 (Hoeffding-type decomposition for $\widehat{f}_W^{1\triangleright 0}$)
Suppose that Assumptions SA1, SA2 and SA4 hold. Writing $k_{ij} = k_h(W_{ij}^1, w)$, $\kappa_{ri} = \kappa(X_r^0, X_r^1, X_i^1)$ and $\psi_i = \psi(X_i^1)$, define the projections

$$\begin{aligned}
u &= \mathbb{E}\left[k_{ij}\psi_i\psi_j\right], \\
u_i &= \frac{2}{3}\psi_i\mathbb{E}\left[k_{ij}\psi_j \mid A_i^1\right] + \frac{2}{3}\mathbb{E}\left[k_{jr}\psi_j\kappa_{ir} \mid X_i^0, X_i^1\right] - \frac{2}{3}u, \\
u_{ij} &= \frac{1}{3}\psi_i\psi_j\mathbb{E}\left[k_{ij} \mid A_i^1, A_j^1\right] + \frac{1}{3}\psi_i\mathbb{E}\left[k_{ir}\psi_r \mid A_i^1\right] + \frac{1}{3}\psi_i\mathbb{E}\left[k_{ir}\kappa_{jr} \mid A_i^1, X_j^0, X_j^1\right] + \frac{1}{3}\kappa_{ji}\mathbb{E}\left[k_{ir}\psi_r \mid A_i^1\right] \\
&\quad + \frac{1}{3}\psi_j\mathbb{E}\left[k_{jr}\psi_r \mid A_j^1\right] + \frac{1}{3}\psi_j\mathbb{E}\left[k_{jr}\kappa_{ir} \mid X_i^0, X_i^1, A_j^1\right] + \frac{1}{3}\kappa_{ij}\mathbb{E}\left[k_{jr}\psi_r \mid A_j^1\right] - u_i - u_j + u, \\
u_{ijr} &= \frac{1}{3}\psi_i\psi_j\mathbb{E}\left[k_{ij} \mid A_i^1, A_j^1\right] + \frac{1}{3}\psi_i\kappa_{rj}\mathbb{E}\left[k_{ij} \mid A_i^1, A_j^1\right] + \frac{1}{3}\psi_j\kappa_{ri}\mathbb{E}\left[k_{ij} \mid A_i^1, A_j^1\right] + \frac{1}{3}\psi_i\psi_r\mathbb{E}\left[k_{ir} \mid A_i^1, A_r^1\right] \\
&\quad + \frac{1}{3}\psi_i\kappa_{jr}\mathbb{E}\left[k_{ir} \mid A_i^1, A_r^1\right] + \frac{1}{3}\psi_r\kappa_{ji}\mathbb{E}\left[k_{ir} \mid A_i^1, A_r^1\right] + \frac{1}{3}\psi_j\psi_r\mathbb{E}\left[k_{jr} \mid A_j^1, A_r^1\right] \\
&\quad + \frac{1}{3}\psi_j\kappa_{ir}\mathbb{E}\left[k_{jr} \mid A_j^1, A_r^1\right] + \frac{1}{3}\psi_r\kappa_{ij}\mathbb{E}\left[k_{jr} \mid A_j^1, A_r^1\right] - u_{ij} - u_{ir} - u_{jr} + u_i + u_j + u_r - u, \\
v_{ijr} &= \frac{1}{3}k_{ij}\big(\psi_i\psi_j + \psi_i\kappa_{rj} + \psi_j\kappa_{ri}\big) + \frac{1}{3}k_{ir}\big(\psi_i\psi_r + \psi_i\kappa_{jr} + \psi_r\kappa_{ji}\big) + \frac{1}{3}k_{jr}\big(\psi_j\psi_r + \psi_j\kappa_{ir} + \psi_r\kappa_{ij}\big).
\end{aligned}$$

With $l_i^{1\triangleright 0}(w) = u_i$ *and* $e_{ijr}^{1\triangleright 0}(w) = v_{ijr} - u_{ijr}$, *set*

$$L_n^{1\triangleright 0}(w) = \frac{3}{n}\sum_{i=1}^{n} l_i^{1\triangleright 0}(w) \qquad \text{and} \qquad E_n^{1\triangleright 0}(w) = \frac{6}{n(n-1)(n-2)}\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{r=i+1}^{n} e_{ijr}^{1\triangleright 0}(w).$$

Then the following Hoeffding-type decomposition holds, where $O_{\mathbb{P}}(1/n)$ *is uniform in* $w \in \mathcal{W}$.

$$\widehat{f}_W^{1\triangleright 0}(w) = \mathbb{E}\big[\widehat{f}_W^{1\triangleright 0}(w)\big] + L_n^{1\triangleright 0}(w) + E_n^{1\triangleright 0}(w) + O_{\mathbb{P}}\left(\frac{1}{n}\right).$$

Further, the stochastic processes $L_n^{1\triangleright 0}$ *and* $E_n^{1\triangleright 0}$ *are mean-zero and mutually orthogonal in* $L^2(\mathbb{P})$. *Define the upper and lower degeneracy constants as*

$$\left(D_{\text{up}}^{1\triangleright 0}\right)^2 = \limsup_{n\to\infty}\sup_{w\in\mathcal{W}} \operatorname{Var}\left[l_i^{1\triangleright 0}(w)\right] \qquad \text{and} \qquad \left(D_{\text{lo}}^{1\triangleright 0}\right)^2 = \liminf_{n\to\infty}\inf_{w\in\mathcal{W}} \operatorname{Var}\left[l_i^{1\triangleright 0}(w)\right].$$

Lemma SA20 (Uniform consistency of $\widehat{f}_W^{1\triangleright 0}$)
Suppose that Assumptions SA1, SA2 and SA4 hold. Then

$$\mathbb{E}\left[\sup_{w\in\mathcal{W}}\left|\widehat{f}_W^{1\triangleright 0}(w) - f_W^{1\triangleright 0}(w)\right|\right] \lesssim h^{p\wedge\beta} + \frac{D_{\text{up}}^{1\triangleright 0}}{\sqrt{n}} + \sqrt{\frac{\log n}{n^2 h}}.$$

Lemma SA21 (Strong approximation of $\widehat{f}_W^{1\triangleright 0}$)
On an appropriately enlarged probability space and for any sequence $R_n \to \infty$, *there exists a mean-zero Gaussian process* $Z_n^{f,1\triangleright 0}$ *with the same covariance structure as* $\widehat{f}_W^{1\triangleright 0}(w)$ *satisfying*

$$\sup_{w\in\mathcal{W}}\left|\widehat{f}_W^{1\triangleright 0}(w) - \mathbb{E}\big[\widehat{f}_W^{1\triangleright 0}(w)\big] - Z_n^{f,1\triangleright 0}(w)\right| \lesssim_{\mathbb{P}} n^{-1}\log n + n^{-5/4}h^{-7/8}(\log n)^{3/8}R_n + n^{-7/6}h^{-1/2}(\log n)^{2/3}.$$

Lemma SA22 (Counterfactual covariance structure)
Writing k'_{ij} *for* $k_h(W_{ij}^1, w')$ *etc., the counterfactual covariance function is*

$$\begin{aligned}\Sigma_n^{1\triangleright 0}(w,w') &= \operatorname{Cov}\left[\widehat{f}_W^{1\triangleright 0}(w), \widehat{f}_W^{1\triangleright 0}(w')\right] \\ &= \frac{4}{n}\mathbb{E}\left[\Big(\psi_i\mathbb{E}\big[k_{ij}\psi_j \mid A_i^1\big] + \mathbb{E}\left[k_{rj}\psi_r\kappa_{ij} \mid X_i^0, X_i^1\right]\Big)\Big(\psi_i\mathbb{E}\big[k'_{ij}\psi_j \mid A_i^1\big] + \mathbb{E}\left[k'_{rj}\psi_r\kappa_{ij} \mid X_i^0, X_i^1\right]\Big)\right] \\ &\quad + \frac{2}{n^2}\mathbb{E}\left[k_{ij}k'_{ij}\psi_i^2\psi_j^2\right] - \frac{4}{n}\mathbb{E}\left[k_{ij}\psi_i\psi_j\right]\mathbb{E}\left[k'_{ij}\psi_i\psi_j\right] + O\left(\frac{1}{n^{3/2}} + \frac{1}{\sqrt{n^4 h}}\right).\end{aligned}$$

Lemma SA23 (Infeasible Gaussian approximation of the standardized counterfactual t-statistic)
Let Assumptions SA1, SA2 and SA4 hold and suppose that $f_W^{1\triangleright 0}(w) > 0$ *on* $\mathcal{W}$. *Define for* $w \in \mathcal{W}$

$$T_n^{1\triangleright 0}(w) = \frac{\widehat{f}_W^{1\triangleright 0}(w) - f_W^{1\triangleright 0}(w)}{\sqrt{\Sigma_n^{1\triangleright 0}(w,w)}} \quad \text{and} \quad Z_n^{T,1\triangleright 0}(w) = \frac{Z_n^{f,1\triangleright 0}(w)}{\sqrt{\Sigma_n^{1\triangleright 0}(w,w)}}.$$

Then with $R_n \to \infty$ *as in Lemma SA21,*

$$\sup_{w\in\mathcal{W}}\left|T_n^{1\triangleright 0}(w) - Z_n^{T,1\triangleright 0}(w)\right| \lesssim_{\mathbb{P}} \frac{n^{-1/2}\log n + n^{-3/4}h^{-7/8}(\log n)^{3/8}R_n + n^{-2/3}h^{-1/2}(\log n)^{2/3} + n^{1/2}h^{p\wedge\beta}}{D_{\text{lo}}^{1\triangleright 0} + 1/\sqrt{nh}}.$$

Theorem SA6 (Infeasible counterfactual uniform confidence bands)
Let Assumptions SA1, SA2, SA3 and SA4 hold and suppose that $f_W^{1\triangleright 0}(w) > 0$ *on* $\mathcal{W}$. *Let* $\alpha \in (0,1)$ *be a confidence level and define* $q_{1-\alpha}^{1\triangleright 0}$ *as the quantile satisfying*

$$\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|Z_n^{T,1\triangleright 0}(w)\right| \le q_{1-\alpha}^{1\triangleright 0}\right) = 1-\alpha.$$

Then

$$\mathbb{P}\left(f_W^{1\triangleright 0}(w) \in \left[\widehat{f}_W^{1\triangleright 0}(w) \pm q_{1-\alpha}^{1\triangleright 0}\sqrt{\Sigma_n^{1\triangleright 0}(w,w)}\right] \text{ for all } w \in \mathcal{W}\right) \to 1-\alpha.$$

To conclude this section we propose an estimator for the counterfactual covariance function $\Sigma_n^{1\triangleright 0}$. First let

$$\widehat{\kappa}(X_i^0, X_i^1, x)] = \frac{\mathbb{I}\{X_i^0 = x\} - \widehat{p}_X^0(x)}{\widehat{p}_X^1(x)} - \frac{\widehat{p}_X^0(x)}{\widehat{p}_X^1(x)} \frac{\mathbb{I}\{X_i^1 = x\} - \widehat{p}_X^1(x)}{\widehat{p}_X^1(x)}$$

and define the leave-out conditional expectation estimators

$$S_i^{1\triangleright 0}(w) = \widehat{\mathbb{E}}\left[k_h(W_{ij}^1, w)\psi(X_j^1) \mid A_i^1\right] = \frac{1}{n-1}\left(\sum_{j=1}^{i-1} k_h(W_{ji}^1, w)\widehat{\psi}(X_j^1) + \sum_{j=i+1}^{n} k_h(W_{ij}^1, w)\widehat{\psi}(X_j^1)\right),$$

$$\widetilde{S}_i^{1\triangleright 0}(w) = \widehat{\mathbb{E}}\left[k_h(W_{rj}^1, w)\psi(X_r^1)\kappa(X_i^0, X_i^1, X_j^1) \mid X_i^0, X_i^1\right] = \frac{1}{n-1}\sum_{j=1}^{n} \mathbb{I}\{j \neq i\}\widehat{\kappa}(X_i^0, X_i^1, X_j^1) S_j^{1\triangleright 0}(w).$$

Then set

$$\begin{aligned}\widehat{\Sigma}_n^{1\triangleright 0}(w, w') = \frac{4}{n^2}\sum_{i=1}^{n}\left(\widehat{\psi}(X_i^1)S_i^{1\triangleright 0}(w) + \widetilde{S}_i^{1\triangleright 0}(w)\right)\left(\widehat{\psi}(X_i^1)S_i^{1\triangleright 0}(w') + \widetilde{S}_i^{1\triangleright 0}(w')\right) \\ - \frac{4}{n^3(n-1)}\sum_{i<j} k_h(W_{ij}^1, w)k_h(W_{ij}^1, w')\widehat{\psi}(X_i^1)^2\widehat{\psi}(X_j^1)^2 - \frac{4}{n}\widehat{f}_W^{1\triangleright 0}(w)\widehat{f}_W^{1\triangleright 0}(w').\end{aligned}$$

We then use a positive semi-definite approximation to $\widehat{\Sigma}_n^{1\triangleright 0}$, denoted by $\widehat{\Sigma}_n^{+,1\triangleright 0}$, following the methodology of Section SA2.8. We omit the proof of consistency of these covariance estimators in the interest of brevity. To construct feasible uniform confidence bands, define a process $\widehat{Z}_n^{T,1\triangleright 0}(w)$ which, conditional on the data $\mathbf{W}_n^1$, $\mathbf{X}_n^0$ and $\mathbf{X}_n^1$ is conditionally mean-zero and conditionally Gaussian and whose conditional covariance structure is

$$\mathbb{E}\left[\widehat{Z}_n^{T,1\triangleright 0}(w)\widehat{Z}_n^{T,1\triangleright 0}(w') \mid \mathbf{W}_n^1, \mathbf{X}_n^0, \mathbf{X}_n^1\right] = \frac{\widehat{\Sigma}_n^{+,1\triangleright 0}(w, w')}{\sqrt{\widehat{\Sigma}_n^{+,1\triangleright 0}(w, w)\widehat{\Sigma}_n^{+,1\triangleright 0}(w', w')}}.$$

Let $\alpha \in (0, 1)$ be a confidence level and define $\widehat{q}_{1-\alpha}^{1\triangleright 0}$ as the conditional quantile satisfying

$$\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{Z}_n^{T,1\triangleright 0}(w)\right| \leq \widehat{q}_{1-\alpha}^{1\triangleright 0} \;\middle|\; \mathbf{W}_n^1, \mathbf{X}_n^0, \mathbf{X}_n^1\right) = 1 - \alpha.$$

Then assuming that the covariance estimator is appropriately consistent, we have that

$$\mathbb{P}\left(f_W^{1\triangleright 0}(w) \in \left[\widehat{f}_W^{1\triangleright 0}(w) \pm \widehat{q}_{1-\alpha}^{1\triangleright 0}\sqrt{\widehat{\Sigma}_n^{+,1\triangleright 0}(w, w)}\right] \text{ for all } w \in \mathcal{W}\right) \to 1 - \alpha.$$

SA3 Technical lemmas

In this section we present some lemmas which provide the technical foundations for several of our main results. These lemmas are stated in as much generality as is reasonably possible, and we believe that they may be of some independent interest.

SA3.1 Maximal inequalities for i.n.i.d. empirical processes

Firstly we provide a maximal inequality for empirical processes of independent but not necessarily identically distributed (i.n.i.d.) random variables, indexed by a class of functions. This result is an extension of Theorem 5.2 from Chernozhukov et al. (2014b), which only covers i.i.d. random variables, and is proven in the same manner. Such a result is useful in the study of dyadic data because when conditioning on latent variables, we may encounter random variables which are conditionally independent but which do not necessarily follow the same conditional distribution.

Lemma SA24 (A maximal inequality for i.n.i.d. empirical processes)
Let $X_1, \dots, X_n$ be independent but not necessarily identically distributed (i.n.i.d.) random variables taking values in a measurable space $(S, \mathcal{S})$. Denote the joint distribution of $X_1, \dots, X_n$ by $\mathbb{P}$ and the marginal distribution of X_i by $\mathbb{P}_i$, and let $\bar{\mathbb{P}} = n^{-1} \sum_i \mathbb{P}_i$. Let $\mathcal{F}$ be a class of Borel measurable functions from S to $\mathbb{R}$ which is pointwise measurable (i.e. it contains a countable subclass which is dense under pointwise convergence). Let F be a strictly positive measurable envelope function for $\mathcal{F}$ (i.e. $|f(s)| \leq |F(s)|$ for all $f \in \mathcal{F}$ and $s \in S$). For a distribution $\mathbb{Q}$ and some $q \geq 1$, define the $(\mathbb{Q}, q)$-norm of $f \in \mathcal{F}$ as $\|f\|_{\mathbb{Q},q}^q = \mathbb{E}_{X \sim \mathbb{Q}}[f(X)^q]$ and suppose that $\|F\|_{\bar{\mathbb{P}},2} < \infty$. For $f \in \mathcal{F}$ define the empirical process

$$G_n(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(f(X_i) - \mathbb{E}[f(X_i)] \Big).$$

Let $\sigma > 0$ satisfy $\sup_{f \in \mathcal{F}} \|f\|_{\bar{\mathbb{P}},2} \leq \sigma \leq \|F\|_{\bar{\mathbb{P}},2}$ and $M = \max_{1 \leq i \leq n} F(X_i)$. Then with $\delta = \sigma / \|F\|_{\bar{\mathbb{P}},2} \in (0, 1]$,

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} \big| G_n(f) \big| \right] \lesssim \|F\|_{\bar{\mathbb{P}},2} \, J\big(\delta, \mathcal{F}, F\big) + \frac{\|M\|_{\mathbb{P},2} \, J(\delta, \mathcal{F}, F)^2}{\delta^2 \sqrt{n}},$$

where $\lesssim$ is up to a universal constant, and $J(\delta, \mathcal{F}, F)$ is the covering integral

$$J\big(\delta, \mathcal{F}, F\big) = \int_0^\delta \sqrt{1 + \sup_{\mathbb{Q}} \log N(\mathcal{F}, \rho_{\mathbb{Q}}, \varepsilon \|F\|_{\mathbb{Q},2})} \, \mathrm{d}\varepsilon,$$

with the supremum taken over finite discrete probability measures $\mathbb{Q}$ on $(S, \mathcal{S})$.

Lemma SA25 (A VC class maximal inequality for i.n.i.d. empirical processes)
Assume the same setup as in Lemma SA24, and suppose further that $\mathcal{F}$ forms a VC-type class in that

$$\sup_{\mathbb{Q}} N(\mathcal{F}, \rho_{\mathbb{Q}}, \varepsilon \|F\|_{\mathbb{Q},2}) \leq (C_1 / \varepsilon)^{C_2}$$

for all $\varepsilon \in (0, 1]$, for some constants $C_1 \geq e$ (where e is the standard exponential constant) and $C_2 \geq 1$. Then for $\delta \in (0, 1]$ we have the covering integral bound

$$J\big(\delta, \mathcal{F}, F\big) \leq 3\delta \sqrt{C_2 \log(C_1 / \delta)},$$

and so by Lemma SA24,

$$\begin{aligned} \mathbb{E}\left[\sup_{f \in \mathcal{F}} \big| G_n(f) \big| \right] &\lesssim \sigma \sqrt{C_2 \log(C_1/\delta)} + \frac{\|M\|_{\mathbb{P},2} C_2 \log(C_1/\delta)}{\sqrt{n}} \\ &\lesssim \sigma \sqrt{C_2 \log\big(C_1 \|F\|_{\bar{\mathbb{P}},2} / \sigma\big)} + \frac{\|M\|_{\mathbb{P},2} C_2 \log\big(C_1 \|F\|_{\bar{\mathbb{P}},2} / \sigma\big)}{\sqrt{n}}, \end{aligned}$$

where $\lesssim$ is up to a universal constant.

SA3.2 Strong approximation results

Next we provide two strong approximation results. The first is a corollary of the KMT approximation (Komlós et al., 1975) which applies to bounded-variation functions of i.i.d. variables. The second is an extension of the Yurinskii coupling (Belloni et al., 2019) which applies to Lipschitz functions of i.n.i.d. variables.

Lemma SA26 (A KMT approximation corollary)
For $n \geq 1$ let $X_1, \dots, X_n$ be i.i.d. real-valued random variables and $g_n : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a function satisfying the total variation bound $\sup_{x \in \mathbb{R}} \|g_n(\cdot, x)\|_{\mathrm{TV}} < \infty$. Then on some probability space there exist independent copies of $X_1, \dots, X_n$, denoted $X_1', \dots, X_n'$, and a mean-zero Gaussian process $Z_n(x)$ such that if we define the empirical process

$$G_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(g_n(X_i', x) - \mathbb{E}\big[g_n(X_i', x)\big] \Big),$$

then for some universal positive constants C_1, C_2 and C_3,

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}}\left|G_n(x)-Z_n(x)\right| > \sup_{x\in\mathbb{R}}\|g_n(\cdot,x)\|_{\mathrm{TV}}\,\frac{t+C_1\log n}{\sqrt{n}}\right) \le C_2 e^{-C_3 t}.$$

Further, Z_n has the same covariance structure as G_n in the sense that for all x, $x' \in \mathbb{R}$,

$$\mathbb{E}\big[Z_n(x)Z_n(x')\big] = \mathbb{E}\big[G_n(x)G_n(x')\big].$$

By independently sampling from the law of Z_n conditional on $X_1', \dots, X_n'$, we can assume that Z_n is a function only of $X_1', \dots, X_n'$ and some independent random noise.

Lemma SA27 (Yurinskii coupling for Lipschitz i.n.i.d. empirical processes)
For $n \ge 2$ let $X_1, \dots, X_n$ be independent but not necessarily identically distributed (i.n.i.d.) random variables taking values in a measurable space $(S, \mathcal{S})$ and let $\mathcal{X}_n \subseteq \mathbb{R}$ be a compact interval with $|\log \mathrm{Leb}(\mathcal{X}_n)| \le C_1 \log n$ where $C_1 > 0$ is a constant. Let g_n be a measurable function on $S \times \mathcal{X}_n$ satisfying $\sup_{\xi\in S}\sup_{x\in\mathcal{X}_n}|g_n(\xi,x)| \le M_n$ and $\sup_{x\in\mathcal{X}_n}\max_{1\le i\le n}\mathrm{Var}[g_n(X_i,x)] \le \sigma_n^2$, with $|\log M_n| \le C_1\log n$ and $\left|\log \sigma_n^2\right| \le C_1 \log n$. Suppose that g_n satisfies the following uniform Lipschitz condition:

$$\sup_{\xi\in S}\sup_{x,x'\in\mathcal{X}_n}\left|\frac{g_n(\xi,x)-g_n(\xi,x')}{x-x'}\right| \le l_{n,\infty},$$

and also the following L^2 Lipschitz condition:

$$\sup_{x,x'\in\mathcal{X}_n}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left|\frac{g_n(X_i,x)-g_n(X_i,x')}{x-x'}\right|^2\right]^{1/2} \le l_{n,2},$$

where $0 < l_{n,2} \le l_{n,\infty}$, $|\log l_{n,2}| \le C_1 \log n$ and $|\log l_{n,\infty}| \le C_1 \log n$. Then for any $t_n > 0$ with $|\log t_n| \le C_1 \log n$, there is a probability space carrying independent copies of $X_1, \dots, X_n$ denoted $X_1', \dots, X_n'$ and a mean-zero Gaussian process $Z_n(x)$ such that if we define the empirical process

$$G_n(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(g_n(X_i',x) - \mathbb{E}[g_n(X_i',x)]\Big),$$

then

$$\mathbb{P}\left(\sup_{x\in\mathcal{X}_n}\left|G_n(x)-Z_n(x)\right| > t_n\right) \le \frac{C_2\sigma_n\sqrt{\mathrm{Leb}(\mathcal{X}_n)}\sqrt{\log n}\sqrt{M_n+\sigma_n\sqrt{\log n}}}{n^{1/4}t_n^2}\sqrt{l_{n,2}\sqrt{\log n}+\frac{l_{n,\infty}}{\sqrt{n}}\log n}$$

where $C_2 > 0$ is a constant depending only on C_1. Further, Z_n has the same covariance structure as G_n in the sense that for all $x, x' \in \mathcal{X}_n$,

$$\mathbb{E}\big[Z_n(x)Z_n(x')\big] = \mathbb{E}\big[G_n(x)G_n(x')\big].$$

SA3.3 The Vorob'ev–Berkes–Philipp theorem

We present a generalization of the Vorob'ev–Berkes–Philipp theorem (Dudley, 1999) which allows one to "glue" multiple random variables or stochastic processes onto the same probability space, while preserving some pairwise distributions. We begin with some definitions.

Definition SA1 (Tree)
A tree *is an undirected graph with finitely many vertices which is connected and contains no cycles or self-loops.*

Definition SA2 (Polish Borel probability space)
A Polish Borel probability space *is a triple $(\mathcal{X}, \mathcal{F}, \mathbb{P})$, where $\mathcal{X}$ is a Polish space (a topological space metrizable by a complete separable metric), $\mathcal{F}$ is the Borel σ-algebra induced on $\mathcal{X}$ by its topology, and $\mathbb{P}$ is a probability measure on $(\mathcal{X}, \mathcal{F})$. Important examples of Polish spaces include $\mathbb{R}^d$ and the Skorokhod space $\mathcal{D}[0,1]^d$ for some $d \ge 1$. In particular, one can consider vectors of real-valued random variables or stochastic processes indexed by compact subsets of $\mathbb{R}^d$ which have almost surely continuous trajectories.*

Definition SA3 (Projection of a law)
Let $(\mathcal{X}_1, \mathcal{F}_1)$ and $(\mathcal{X}_2, \mathcal{F}_2)$ be measurable spaces, and let $\mathbb{P}_{12}$ be a law on the product space $(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$. The projection *of $\mathbb{P}_{12}$ onto $\mathcal{X}_1$ is the law $\mathbb{P}_1$ defined on $(\mathcal{X}_1, \mathcal{F}_1)$ by $\mathbb{P}_1 = \mathbb{P}_{12} \circ \pi_1^{-1}$ where $\pi_1(x_1, x_2) = x_1$ is the first-coordinate projection.*

Lemma SA28 (Vorob'ev–Berkes–Philipp theorem, tree form)
Let $\mathcal{T}$ be a tree with vertex set $\mathcal{V} = \{1, \dots, n\}$ and edge set $\mathcal{E}$. Suppose that attached to each vertex i is a Polish Borel probability space $(\mathcal{X}_i, \mathcal{F}_i, \mathbb{P}_i)$. Suppose that attached to each edge $(i, j) \in \mathcal{E}$ (where $i < j$ without loss of generality) is a law $\mathbb{P}_{ij}$ on $(\mathcal{X}_i \times \mathcal{X}_j, \mathcal{F}_i \otimes \mathcal{F}_j)$. Assume that these laws are pairwise-consistent in the sense that the projection of $\mathbb{P}_{ij}$ onto $\mathcal{X}_i$ (resp. $\mathcal{X}_j$) is $\mathbb{P}_i$ (resp. $\mathbb{P}_j$) for each $(i, j) \in \mathcal{E}$. Then there exists a law $\mathbb{P}$ on

$$\left(\prod_{i=1}^{n} \mathcal{X}_i, \bigotimes_{i=1}^{n} \mathcal{F}_i \right)$$

such that the projection of $\mathbb{P}$ onto $\mathcal{X}_i \times \mathcal{X}_j$ is $\mathbb{P}_{ij}$ for each $(i, j) \in \mathcal{E}$, and therefore also the projection of $\mathbb{P}$ onto $\mathcal{X}_i$ is $\mathbb{P}_i$ for each $i \in \mathcal{V}$.

Remark. The requirement that $\mathcal{T}$ must contain no cycles is necessary in general. To see this, consider the Polish Borel probability spaces given by $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}_3 = \{0, 1\}$, their respective Borel σ-algebras, and the pairwise-consistent probability measures:

$$1/2 = \mathbb{P}_1(0) = \mathbb{P}_2(0) = \mathbb{P}_3(0)$$
$$1/2 = \mathbb{P}_{12}(0,1) = \mathbb{P}_{12}(1,0) = \mathbb{P}_{13}(0,1) = \mathbb{P}_{13}(1,0) = \mathbb{P}_{23}(0,1) = \mathbb{P}_{23}(1,0).$$

That is, each measure $\mathbb{P}_i$ places equal mass on 0 and 1, while $\mathbb{P}_{ij}$ asserts that each pair of realizations is a.s. not equal. The graph of these laws forms a triangle, which is not a tree. Suppose that (X_1, X_2, X_3) has distribution given by $\mathbb{P}$, where $X_i \sim \mathbb{P}_i$ and $(X_i, X_j) \sim \mathbb{P}_{ij}$ for each i, j. But then by definition of $\mathbb{P}_{ij}$ we have $X_1 = 1 - X_2 = X_3 = 1 - X_1$ a.s., which is a contradiction.

Remark. Two important applications of Lemma SA28 include the embedding of a random vector into a stochastic process and the coupling of stochastic processes onto the same probability space:

(i) Let X_1 and X_2 be stochastic processes with trajectories in $\mathcal{D}[0,1]$. For $x_1, \dots, x_n \in [0,1]$ let $\widetilde{X}_1 = (X_1(x_1), \dots, X_1(x_n))$ be a random vector and suppose that $\widetilde{X}_1'$ is a copy of $\widetilde{X}_1$. Then there is a law $\mathbb{P}$ on $\mathcal{D}[0,1] \times \mathbb{R}^n \times \mathcal{D}[0,1]$ such that restriction of $\mathbb{P}$ to $\mathcal{D}[0,1] \times \mathbb{R}^n$ is the law of $(X_1, \widetilde{X}_1)$, while the restriction of $\mathbb{P}$ to $\mathbb{R}^n \times \mathcal{D}[0,1]$ is the law of $(\widetilde{X}_1', X_2)$. In other words, we can embed the vector $\widetilde{X}_1'$ into a stochastic process X_1 while maintaining the joint distribution of $\widetilde{X}_1'$ and X_2.

(ii) Let $X_1, X_1', \dots, X_n, X_n'$ be stochastic processes with trajectories in $\mathcal{D}[0,1]$, where X_i' is a copy of X_i for each $1 \leq i \leq n-1$. Suppose that $\mathbb{P}\big(\|X_{i+1} - X_i'\| > t\big) \leq r_i$ for each $1 \leq i \leq n-1$, where $\|\cdot\|$ is a norm on $\mathcal{D}[0,1]$. Then there exist copies of $X_1, \dots, X_n$ denoted $X_1'', \dots, X_n''$ satisfying $\mathbb{P}\big(\|X_{i+1}'' - X_i''\| > t\big) \leq r_i$ for each $1 \leq i \leq n$. That is, all of the approximation inequalities can be satisfied simultaneously on the same probability space.

Remark. Note that while we discuss trees in Lemma SA28, these refer to the dependency graph of the relevant Polish space-valued random variables and do not have any direct relation to the networks studied throughout the main paper.

SA4 Additional empirical results

We present some additional empirical results using the International Monetary Fund's Direction of Trade Statistics (DOTS) data set, to complement those given in the main paper. This data set contains information about the yearly trade flows among $n = 207$ economies ($N = 21\,321$ pairs), and we focus on the years 1995, 2000 and 2005.

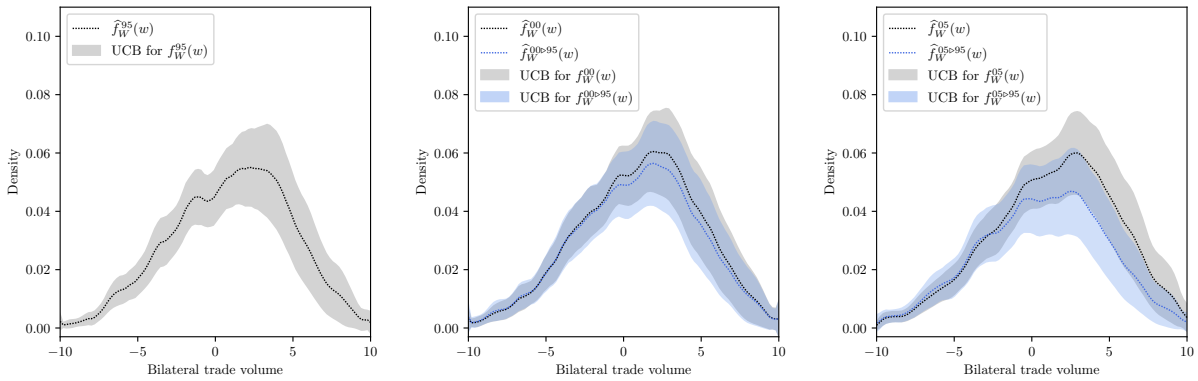
We define the *trade volume* between countries i and j as the logarithm of the sum of the trade flow (in billions of US dollars) from i to j and the trade flow from j to i. In each year several pairs of countries did not trade directly, yielding trade flows of zero and hence a trade volume of $-\infty$. We therefore assume that the distribution of trade volumes is a mixture of a point mass at $-\infty$ and a Lebesgue density on $\mathbb{R}$. The local nature of our estimator means that observations taking the value of $-\infty$ can simply be removed from the data set.

For counterfactual analysis we use the gross domestic product (GDP) of each country as a covariate, using 10%-percentiles to group the values into 10 different levels for ease of estimation. This allows for a comparison of the observed distribution of trade at each year with, for example, the counterfactual distribution of trade had the GDP distribution remained as it was in 1995. As such we can measure how much of the change in trade distribution is attributable to a shift in the GDP distribution.

To estimate the trade volume density function we use the counterfactual dyadic kernel density estimator from Section SA2.10 with $d = 100$ equally-spaced evaluation points in $[-10, 10]$, using the rule-of-thumb bandwidth selector $\widehat{h}_{\mathrm{ROT}}$ described in the main paper with $p = 2$ and $C(K) = 2.435$. For inference we use an Epanechnikov kernel of order $p = 4$ and resample the Gaussian process $B = 10\,000$ times. We also estimate the counterfactual trade distributions in 2000 and 2005 respectively, replacing the GDP distribution with that from 1995. For each year, Figure 1 plots the real and counterfactual density estimates along with their respective uniform confidence bands (UCB) at the nominal coverage rate of 95%. Our empirical results show that the counterfactual distribution drifts further from the truth in 2005 compared with 2000, indicating a more significant shift in the GDP distribution.

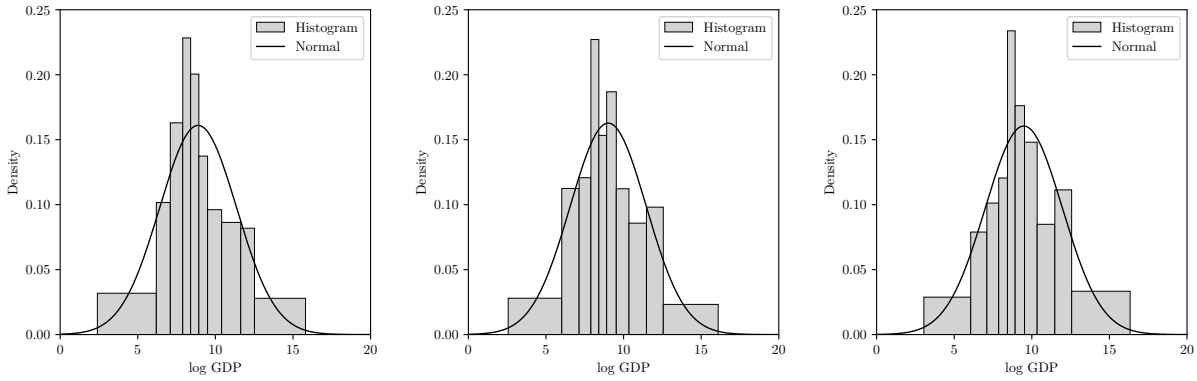
In Figure 2 we illustrate how, in the preliminary step of the counterfactual analysis, the distribution of log GDP is approximated using the histogram estimators $\widehat{p}_X^0$ and $\widehat{p}_X^1$ defined in Section SA2.10. We also plot the density function of a normal distribution, fitted using maximum likelihood estimation, and this seems to capture the distribution of log GDP reasonably well. Such a parametric approach to the preliminary step may be favored in cases where a choice of model is clear or where the histogram estimators perform poorly.

To demonstrate the relative robustness of our counterfactual analysis to the choice of preliminary estimation step, we also provide results using a parametric estimator of the distribution of GDP. Figure 3 repeats the procedure used for Figure 1, but this time replacing the histogram estimators by parametric estimators of the log GDP based on normal likelihood maximization. The point estimates are qualitatively similar, with the counterfactual distribution drifting in the same direction over time. The confidence bands are also similar, with the band based on the parametric fit being slightly narrower in general. This could be due to the more stringent model specification leading to less estimated variance in the fitted values.



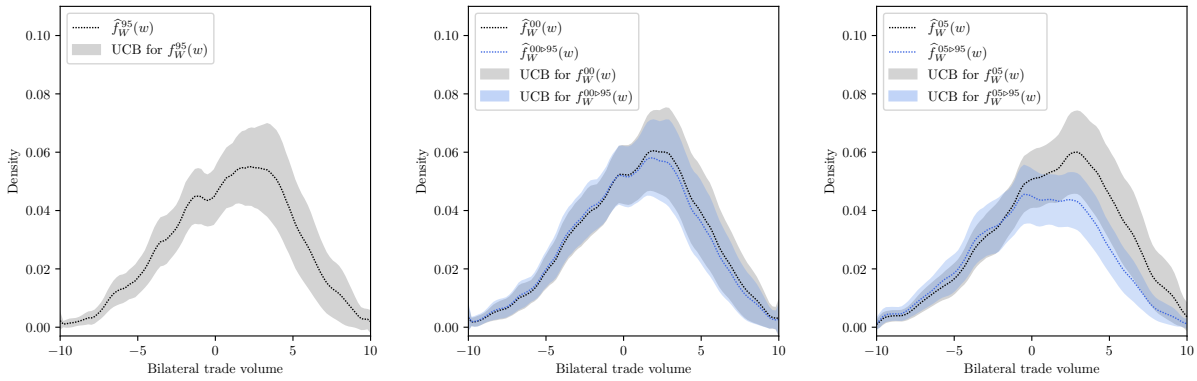
(a) Year 1995, $\widehat{h}_{\mathrm{ROT}} = 1.27$ (b) Year 2000, $\widehat{h}_{\mathrm{ROT}} = 1.31$ (c) Year 2005, $\widehat{h}_{\mathrm{ROT}} = 1.37$

Figure 1: Real and counterfactual density estimates and confidence bands for the DOTS data with histogram-based covariate estimation.



(a) Year 1995 (b) Year 2000 (c) Year 2005

Figure 2: Estimated GDP distributions for the DOTS data.



(a) Year 1995, $\widehat{h}_{\text{ROT}} = 1.27$ (b) Year 2000, $\widehat{h}_{\text{ROT}} = 1.31$ (c) Year 2005, $\widehat{h}_{\text{ROT}} = 1.37$

Figure 3: Real and counterfactual density estimates and confidence bands for the DOTS data with parametric covariate estimation.

SA5 Proofs

SA5.1 Preliminary lemmas

In this section we list some results in probability and U-statistic theory which are used in proofs of this paper's main results. Other auxiliary lemmas will be introduced when they are needed.

SA5.1.1 Standard probabilistic results

Lemma SA29 (Bernstein's inequality for independent random variables)
Let $X_1, \ldots, X_n$ be independent real-valued random variables with $\mathbb{E}[X_i] = 0$ and $|X_i| \leq M$ and $\mathbb{E}[X_i^2] \leq \sigma^2$, where M and σ are non-random. Then for all $t > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq t\right) \leq 2\exp\left(-\frac{t^2 n}{2\sigma^2 + \frac{2}{3}Mt}\right).$$

Proof (Lemma SA29)
See for example Lemma 2.2.9 in van der Vaart and Wellner (1996). □

Lemma SA30 (The matrix Bernstein inequality)
For $1 \le i \le n$ let X_i be independent symmetric $d \times d$ real random matrices with expected values $\mu_i = \mathbb{E}[X_i]$. Suppose that $\|X_i - \mu_i\|_2 \le M$ almost surely for all $1 \le i \le n$ where M is non-random, and define $\sigma^2 = \left\| \sum_i \mathbb{E}[(X_i - \mu_i)^2] \right\|_2$. Then there exists a universal constant $C > 0$ such that for any $t > 0$ and $q \ge 1$,

$$\mathbb{P}\left(\left\| \sum_{i=1}^n (X_i - \mu_i) \right\|_2 \ge 2\sigma\sqrt{t} + \frac{4}{3}Mt \right) \le 2de^{-t},$$
$$\mathbb{E}\left[\left\| \sum_{i=1}^n (X_i - \mu_i) \right\|_2^q \right]^{1/q} \le C\sigma\sqrt{q + \log 2d} + CM(q + \log 2d).$$

Another simplified version of this is as follows: suppose that $\|X_i\|_2 \le M$ almost surely, so that $\|X_i - \mu_i\|_2 \le 2M$. Then since $\sigma^2 \le nM^2$, we have

$$\mathbb{P}\left(\left\| \sum_{i=1}^n (X_i - \mu_i) \right\|_2 \ge 4M\big(t + \sqrt{nt}\big) \right) \le 2de^{-t},$$
$$\mathbb{E}\left[\left\| \sum_{i=1}^n (X_i - \mu_i) \right\|_2^q \right]^{1/q} \le CM\big(q + \log 2d + \sqrt{n(q + \log 2d)}\big).$$

Proof (Lemma SA30)
See Lemma 3.2 in Minsker and Wei (2019). □

Lemma SA31 (A maximal inequality for Gaussian vectors)
Take $n \ge 2$. Let $X_i \sim \mathcal{N}(0, \sigma_i^2)$ for $1 \le i \le n$ (not necessarily independent), with $\sigma_i^2 \le \sigma^2$. Then

$$\mathbb{E}\left[\max_{1 \le i \le n} X_i \right] \le \sigma\sqrt{2 \log n}, \tag{2}$$
$$\mathbb{E}\left[\max_{1 \le i \le n} |X_i| \right] \le 2\sigma\sqrt{\log n}. \tag{3}$$

If Σ_1 and Σ_2 are constant positive semi-definite $n \times n$ matrices and $N \sim \mathcal{N}(0, I_n)$, then

$$\mathbb{E}\Big[\big\| \Sigma_1^{1/2} N - \Sigma_2^{1/2} N \big\|_\infty \Big] \le 2\sqrt{\log n}\, \big\| \Sigma_1 - \Sigma_2 \big\|_2^{1/2}. \tag{4}$$

If further Σ_1 is positive definite, then

$$\mathbb{E}\Big[\big\| \Sigma_1^{1/2} N - \Sigma_2^{1/2} N \big\|_\infty \Big] \le \sqrt{\log n}\, \lambda_{\min}(\Sigma_1)^{-1/2}\, \big\| \Sigma_1 - \Sigma_2 \big\|_2. \tag{5}$$

Proof (Lemma SA31)
For $t > 0$, Jensen's inequality on the concave logarithm function gives

$$\mathbb{E}\left[\max_{1 \le i \le n} X_i \right] = \frac{1}{t}\mathbb{E}\left[\log \exp \max_{1 \le i \le n} tX_i \right] \le \frac{1}{t} \log \mathbb{E}\left[\exp \max_{1 \le i \le n} tX_i \right] \le \frac{1}{t} \log \sum_{i=1}^n \mathbb{E}\left[\exp tX_i \right]$$
$$= \frac{1}{t} \log \sum_{i=1}^n \exp\left(\frac{t^2 \sigma_i^2}{2} \right) \le \frac{1}{t} \log n + \frac{t\sigma^2}{2},$$

where we use the Gaussian moment generating function. Minimizing this upper bound over t by setting $t = \sqrt{2 \log n}/\sigma$ yields Equation 2:

$$\mathbb{E}\left[\max_{1 \le i \le n} X_i \right] \le \sigma\sqrt{2 \log n}.$$

For Equation 3, we use the symmetry of the Gaussian distribution:

$$\mathbb{E}\left[\max_{1\le i\le n}|X_i|\right] = \mathbb{E}\left[\max_{1\le i\le n}\{X_i, -X_i\}\right] \le \sigma\sqrt{2\log 2n} \le 2\sigma\sqrt{\log n}.$$

For Equations 4 and 5, note that $\Sigma_1^{1/2}N - \Sigma_2^{1/2}N$ is a Gaussian vector with covariance matrix $\left(\Sigma_1^{1/2} - \Sigma_2^{1/2}\right)^2$. The variances of of its components are the diagonal elements of this matrix, namely

$$\sigma_i^2 = \mathrm{Var}\left[\left(\Sigma_1^{1/2}N - \Sigma_2^{1/2}N\right)_i\right] = \left(\left(\Sigma_1^{1/2} - \Sigma_2^{1/2}\right)^2\right)_{ii}.$$

Note that if e_i is the ith standard unit basis vector, then for any real symmetric matrix A, we have $e_i^{\mathsf{T}}A^2e_i = (A^2)_{ii}$, so in particular $(A^2)_{ii} \le \|A\|_2^2$. Therefore

$$\sigma_i^2 \le \left\|\Sigma_1^{1/2} - \Sigma_2^{1/2}\right\|_2^2 =: \sigma^2.$$

Applying Equation 3 then gives

$$\mathbb{E}\left[\left\|\Sigma_1^{1/2}N - \Sigma_2^{1/2}N\right\|_\infty\right] \le 2\sqrt{\log n}\left\|\Sigma_1^{1/2} - \Sigma_2^{1/2}\right\|_2.$$

By Theorem X.1.1 in Bhatia (1997), we can deduce

$$\left\|\Sigma_1^{1/2} - \Sigma_2^{1/2}\right\|_2 \le \left\|\Sigma_1 - \Sigma_2\right\|_2^{1/2},$$

giving Equation 4. If further Σ_1 is positive definite, then by Theorem X.3.8 in Bhatia (1997),

$$\left\|\Sigma_1^{1/2} - \Sigma_2^{1/2}\right\|_2 \le \frac{1}{2}\lambda_{\min}(\Sigma_1)^{-1/2}\left\|\Sigma_1 - \Sigma_2\right\|_2,$$

giving Equation 5. □

Lemma SA32 (Maximal inequalities for Gaussian processes)
Let Z be a separable mean-zero Gaussian process indexed by $x \in \mathcal{X}$. Recall that Z is separable for example if $\mathcal{X}$ is Polish and Z has continuous trajectories. Define its covariance structure on $\mathcal{X}\times\mathcal{X}$ by $\Sigma(x,x') = \mathbb{E}[Z(x)Z(x')]$, and the corresponding semimetric on $\mathcal{X}$ by

$$\rho(x,x') = \mathbb{E}\left[\left(Z(x) - Z(x')\right)^2\right]^{1/2} = \left(\Sigma(x,x) - 2\Sigma(x,x') + \Sigma(x',x')\right)^{1/2}.$$

Let $N(\varepsilon,\mathcal{X},\rho)$ denote the ε-covering number of $\mathcal{X}$ with respect to the semimetric ρ. Define $\sigma = \sup_x \Sigma(x,x)^{1/2}$. Then there exists a universal constant $C > 0$ such that for any $\delta > 0$,

$$\mathbb{E}\left[\sup_{x\in\mathcal{X}}|Z(x)|\right] \le C\sigma + C\int_0^{2\sigma}\sqrt{\log N(\varepsilon,\mathcal{X},\rho)}\,\mathrm{d}\varepsilon,$$
$$\mathbb{E}\left[\sup_{\rho(x,x')\le\delta}|Z(x) - Z(x')|\right] \le C\int_0^{\delta}\sqrt{\log N(\varepsilon,\mathcal{X},\rho)}\,\mathrm{d}\varepsilon.$$

Proof (Lemma SA32)
See Corollary 2.2.8 in van der Vaart and Wellner (1996), noting that for any $x, x' \in \mathcal{X}$, we have $\mathbb{E}[|Z(x)|] \lesssim \sigma$ and $\rho(x,x') \le 2\sigma$, implying that $\log N(\varepsilon,\mathcal{X},\rho) = 0$ for all $\varepsilon > 2\sigma$. □

Lemma SA33 (Anti-concentration for Gaussian process absolute suprema)
Let Z be a separable mean-zero Gaussian process indexed by a semimetric space $\mathcal{X}$ satisfying $\mathbb{E}[Z(x)^2] = 1$ for all $x \in \mathcal{X}$. Then for any $\varepsilon > 0$,

$$\sup_{t\in\mathbb{R}}\mathbb{P}\left(\left|\sup_{x\in\mathcal{X}}|Z(x)| - t\right| \le \varepsilon\right) \le 4\varepsilon\left(1 + \mathbb{E}\left[\sup_{x\in\mathcal{X}}|Z(x)|\right]\right).$$

Proof (Lemma SA33)
See Corollary 2.1 in Chernozhukov et al. (2014a). □

Lemma SA34 (No slowest rate of convergence in probability)
Let X_n be a sequence of real-valued random variables with $X_n = o_{\mathbb{P}}(1)$. Then there exists a deterministic sequence $\varepsilon_n \to 0$ such that $\mathbb{P}\big(|X_n| > \varepsilon_n\big) \leq \varepsilon_n$ for all $n \geq 1$.

Proof (Lemma SA34)
Define the following deterministic sequence for $k \geq 1$.

$$\tau_k = \sup\big\{n \geq 1 : \mathbb{P}\big(|X_n| > 1/k\big) > 1/k\big\} \vee (\tau_{k-1} + 1)$$

with $\tau_0 = 0$. Since $X_n = o_{\mathbb{P}}(1)$, each τ_k is finite and so we can define $\varepsilon_n = \frac{1}{k}$ where $\tau_k < n \leq \tau_{k+1}$. Then, noting that $\varepsilon_n \to 0$, we have $\mathbb{P}\big(|X_n| > \varepsilon_n\big) = \mathbb{P}\big(|X_n| > 1/k\big) \leq 1/k = \varepsilon_n$. □

SA5.1.2 U-statistics

Lemma SA35 (General Hoeffding-type decomposition)
Let $\mathcal{U}$ be a vector space. Let $u_{ij} \in \mathcal{U}$ be defined for $1 \leq i, j \leq n$ and $i \neq j$. Suppose that $u_{ij} = u_{ji}$ for all i, j. Then for any $u_i \in \mathcal{U}$ (for $1 \leq i \leq n$) and any $u \in \mathcal{U}$, the following decomposition holds:

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\big(u_{ij} - u\big) = 2(n-1)\sum_{i=1}^{n}\big(u_i - u\big) + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\big(u_{ij} - u_i - u_j + u\big).$$

Proof (Lemma SA35)
We compute the left hand side minus the right hand side, beginning by observing that all of the u_{ij} and u terms clearly cancel.

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\big(u_{ij} - u\big) - 2(n-1)\sum_{i=1}^{n}\big(u_i - u\big) - \sum_{i=1}^{n}\sum_{j\neq i}\big(u_{ij} - u_i - u_j + u\big) \\
&\quad = -2(n-1)\sum_{i=1}^{n} u_i - \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\big(-u_i - u_j\big) = -2(n-1)\sum_{i=1}^{n} u_i + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} u_i + \sum_{j=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n} u_j \\
&\quad = -2(n-1)\sum_{i=1}^{n} u_i + (n-1)\sum_{i=1}^{n} u_i + (n-1)\sum_{j=1}^{n} u_j = 0.
\end{aligned}$$

□

Lemma SA36 (A U-statistic concentration inequality)
Let $(S, \mathcal{S})$ be a measurable space and $X_1, \ldots, X_n$ be i.i.d. S-valued random variables. Let $H : S^m \to \mathbb{R}$ be a function of m variables satisfying the symmetry property $H(x_1, \ldots, x_m) = H(x_{\tau(1)}, \ldots, x_{\tau(m)})$ for any m-permutation τ. Suppose also that $\mathbb{E}[H(X_1, \ldots, X_m)] = 0$. Let $M = \|H\|_\infty$ and $\sigma^2 = \mathbb{E}\big[\mathbb{E}[H(X_1, \ldots, X_m) \mid X_1]^2\big]$. Define the (not necessarily degenerate) U-statistic

$$U_n = \frac{m!(n-m)!}{n!}\sum_{1 \leq i_1 < \cdots < i_m \leq n} H(X_1, \ldots, X_n).$$

Then for any $t > 0$,

$$\mathbb{P}\left(|U_n| > t\right) \leq 4\exp\left(-\frac{nt^2}{C_1(m)\sigma^2 + C_2(m)Mt}\right),$$

where $C_1(m)$, $C_2(m)$ are positive constants depending only on m.

Proof (Lemma SA36)
See Theorem 2 in Arcones (1995). □

Lemma SA37 (A second-order U-process maximal inequality)
Let $X_1 \ldots, X_n$ be i.i.d. random variables taking values in a measurable space $(S, \mathcal{S})$ with distribution $\mathbb{P}$. Let $\mathcal{F}$ be a class of measurable functions from $S \times S$ to $\mathbb{R}$ which is also pointwise measurable. Define the degenerate second-order U-process

$$U_n(f) = \frac{2}{n(n-1)} \sum_{i<j} \Big(f(X_i, X_j) - \mathbb{E}\big[f(X_i, X_j) \mid X_i\big] - \mathbb{E}\big[f(X_i, X_j) \mid X_j\big] + \mathbb{E}\big[f(X_i, X_j)\big] \Big)$$

for $f \in \mathcal{F}$. Suppose that each $f \in \mathcal{F}$ is symmetric in the sense that $f(s_1, s_2) = f(s_2, s_1)$ for all $s_1, s_2 \in S$. Let F be a measurable envelope function for $\mathcal{F}$ satisfying $|f(s_1, s_2)| \leq F(s_1, s_2)$ for all $s_1, s_2 \in S$. For a law $\mathbb{Q}$ on $(S \times S, \mathcal{S} \otimes \mathcal{S})$, define the $(\mathbb{Q}, q)$-norm of $f \in \mathcal{F}$ by $\|f\|_{\mathbb{Q},q}^q = \mathbb{E}_{\mathbb{Q}}[|f|^q]$. Assume that $\mathcal{F}$ is VC-type in the following manner.

$$\sup_{\mathbb{Q}} N\big(\mathcal{F}, \|\cdot\|_{\mathbb{Q},2}, \varepsilon \|F\|_{\mathbb{Q},2}\big) \leq (C_1/\varepsilon)^{C_2}$$

for some constants $C_1 \geq e$ and $C_2 \geq 1$, and for all $\varepsilon \in (0, 1]$, where $\mathbb{Q}$ ranges over all finite discrete laws on $S \times S$. Let $\sigma > 0$ be any deterministic value satisfying $\sup_{f \in \mathcal{F}} \|f\|_{\mathbb{P},2} \leq \sigma \leq \|F\|_{\mathbb{P},2}$, and define the random variable $M = \max_{i,j} |F(X_i, X_j)|$. Then there exists a universal constant $C_3 > 0$ satisfying

$$n\mathbb{E}\left[\sup_{f \in \mathcal{F}} \big|U_n(f)\big|\right] \leq C_3 \sigma \Big(C_2 \log\big(C_1 \|F\|_{\mathbb{P},2}/\sigma\big)\Big) + \frac{C_3 \|M\|_{\mathbb{P},2}}{\sqrt{n}} \Big(C_2 \log\big(C_1 \|F\|_{\mathbb{P},2}/\sigma\big)\Big)^2.$$

Proof (Lemma SA37)
Apply Corollary 5.3 from Chen and Kato (2020) with the order of the U-statistic fixed at $r = 2$, and with $k = 2$. □

Lemma SA38 (A U-statistic matrix concentration inequality)
Let $X_1, \ldots, X_n$ be i.i.d. random variables taking values in a measurable space $(S, \mathcal{S})$. Suppose $H : S^2 \to \mathbb{R}^{d \times d}$ is a measurable matrix-valued function of two variables satisfying the following assumptions:

(i) $H(X_1, X_2)$ is an almost surely symmetric matrix.

(ii) $\|H(X_1, X_2)\|_2 \leq M$ almost surely.

(iii) H is a symmetric function in its arguments in that $H(X_1, X_2) = H(X_2, X_1)$ almost surely.

(iv) H is degenerate in the sense that $\mathbb{E}[H(X_1, x_2)] = 0$ for all $x_2 \in S$.

Let $U_n = \sum_i \sum_{j \neq i} H(X_i, X_j)$ be a U-statistic, and define the variance-type constant

$$\sigma^2 = \mathbb{E}\left[\big\|\mathbb{E}\left[H(X_i, X_j)^2 \mid X_j\right]\big\|_2\right].$$

Then for a universal constant $C > 0$ and for all $t > 0$,

$$\mathbb{P}\left(\|U_n\|_2 \geq C\sigma n(t + \log d) + CM\sqrt{n}(t + \log d)^{3/2}\right) \leq Ce^{-t}.$$

We remark here that clearly by Jensen's inequality, $\sigma^2 \leq \mathbb{E}[\|H(X_i, X_j)^2\|_2] = \mathbb{E}[\|H(X_i, X_j)\|_2^2] \leq M^2$, giving the weaker but simpler concentration inequality

$$\mathbb{P}\left(\|U_n\|_2 \geq 2CMn(t + \log d)^{3/2}\right) \leq Ce^{-t}.$$

From this last inequality we can deduce the following moment bound by integration of tail probabilities.

$$\mathbb{E}\left[\|U_n\|_2\right] \lesssim Mn(\log d)^{3/2}.$$

Proof (Lemma SA38)
We apply results from Minsker and Wei (2019).

Part 1: decoupling
Let $\bar{U}_n = \sum_{i=1}^n \sum_{j=1}^n H(X_i^{(1)}, X_j^{(2)})$ be a decoupled matrix U-statistic, where $X^{(1)}$ and $X^{(2)}$ are i.i.d. copies of the sequence $X_1, \ldots, X_n$. By Lemma 5.2 in Minsker and Wei (2019), since we are only stating this result for degenerate U-statistics of order 2, there exists a universal constant D_2 such that for any $t > 0$, we have

$$\mathbb{P}\left(\|U_n\|_2 \geq t\right) \leq D_2 \mathbb{P}\left(\|\bar{U}_n\|_2 \geq t/D_2\right).$$

Part 2: concentration of the decoupled U-statistic
By Equation 11 in Minsker and Wei (2019), we have the following concentration inequality for decoupled degenerate U-statistics. For some universal constant C_1 and for any $t > 0$,

$$\mathbb{P}\left(\|\bar{U}_n\|_2 \geq C_1 \sigma n(t + \log d) + C_1 M\sqrt{n}(t + \log d)^{3/2}\right) \leq e^{-t}.$$

Part 3: concentration of the original U-statistic
Hence we have

$$\begin{aligned}
&\mathbb{P}\left(\|U_n\|_2 \geq C_1 D_2 \sigma n(t + \log d) + C_1 D_2 M\sqrt{n}(t + \log d)^{3/2}\right) \\
&\quad \leq D_2 \mathbb{P}\left(\|\bar{U}_n\|_2 \geq C_1 \sigma n(t + \log d) + C_1 M\sqrt{n}(t + \log d)^{3/2}\right) \\
&\quad \leq D_2 e^{-t}.
\end{aligned}$$

The main result follows by setting $C = C_1 + C_1 D_2$.

Part 4: moment bound
The final equation, giving a moment bound for the simplified version, can be deduced as follows. We already have that

$$\mathbb{P}\left(\|U_n\|_2 \geq 2CMn(t + \log d)^{3/2}\right) \leq Ce^{-t}.$$

This implies that for any $t \geq \log d$, we have

$$\mathbb{P}\left(\|U_n\|_2 \geq 8CMnt^{3/2}\right) \leq Ce^{-t}.$$

Defining $s = 8CMnt^{3/2}$, or equivalently $t = \left(\frac{s}{8CMn}\right)^{2/3}$, shows that for any $s \geq 8CMn(\log d)^{3/2}$,

$$\mathbb{P}\left(\|U_n\|_2 \geq s\right) \leq Ce^{-\left(\frac{s}{8CMn}\right)^{2/3}}.$$

Hence the moment bound is obtained:

$$\begin{aligned}
\mathbb{E}\left[\|U_n\|_2\right] &= \int_0^\infty \mathbb{P}\left(\|U_n\|_2 \geq s\right) \mathrm{d}s = \int_0^{8CMn(\log d)^{3/2}} \mathbb{P}\left(\|U_n\|_2 \geq s\right) \mathrm{d}s + \int_{8CMn(\log d)^{3/2}}^\infty \mathbb{P}\left(\|U_n\|_2 \geq s\right) \mathrm{d}s \\
&\leq 8CMn(\log d)^{3/2} + \int_0^\infty Ce^{-\left(\frac{s}{8CMn}\right)^{2/3}} \mathrm{d}s = 8CMn(\log d)^{3/2} + 8CMn \int_0^\infty e^{s^{-2/3}} \mathrm{d}s \\
&\lesssim Mn(\log d)^{3/2}.
\end{aligned}$$

□

SA5.2 Technical results

SA5.2.1 Maximal inequalities for i.n.i.d. empirical processes

Before presenting the proof of Lemma SA24, we give some auxiliary lemmas; namely a symmetrization inequality (Lemma SA39), a Rademacher contraction principle (Lemma SA40) and a Hoffman–Jørgensen inequality (Lemma SA41). Recall that the Rademacher distribution places probability mass of $1/2$ on each of the points -1 and 1.

Lemma SA39 (A symmetrization inequality for i.n.i.d. variables)
Let $(S, \mathcal{S})$ be a measurable space and $\mathcal{F}$ a class of Borel-measurable functions from S to $\mathbb{R}$ which is pointwise measurable (i.e. it contains a countable dense subset under pointwise convergence). Let $X_1, \ldots, X_n$ be independent but not necessarily identically distributed S-valued random variables. Let $a_1, \ldots, a_n$ be arbitrary points in S and ϕ a non-negative non-decreasing convex function from $\mathbb{R}$ to $\mathbb{R}$. Define $\varepsilon_1, \ldots, \varepsilon_n$ as independent Rademacher random variables, independent of $X_1, \ldots, X_n$. Then

$$\mathbb{E}\left[\phi\left(\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \Big(f(X_i) - \mathbb{E}[f(X_i)]\Big)\right|\right)\right] \le \mathbb{E}\left[\phi\left(2\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i\Big(f(X_i) - a_i\Big)\right|\right)\right].$$

Note that in particular this holds with $a_i = 0$ and also holds with $\phi(t) = t \vee 0$.

Proof (Lemma SA39)
See Lemma 2.3.6 in van der Vaart and Wellner (1996). □

Lemma SA40 (A Rademacher contraction principle)
Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent Rademacher random variables and $\mathcal{T}$ be a bounded subset of $\mathbb{R}^n$. Define $M = \sup_{t\in\mathcal{T}} \max_{1\le i\le n} |t_i|$. Then, noting that the supremum is measurable because $\mathcal{T}$ is a subset of a separable metric space and is therefore itself separable,

$$\mathbb{E}\left[\sup_{t\in\mathcal{T}}\left|\sum_{i=1}^n \varepsilon_i t_i^2\right|\right] \le 4M\,\mathbb{E}\left[\sup_{t\in\mathcal{T}}\left|\sum_{i=1}^n \varepsilon_i t_i\right|\right].$$

This gives the following corollary. Let $X_1, \ldots, X_n$ be mutually independent and also independent of $\varepsilon_1, \ldots, \varepsilon_n$. Let $\mathcal{F}$ be a pointwise measurable class of functions from a measurable space $(S, \mathcal{S})$ to $\mathbb{R}$, with measurable envelope F. Define $M = \max_i F(X_i)$. Then we obtain that

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)^2\right|\right] \le 4\mathbb{E}\left[M\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)\right|\right].$$

Proof (Lemma SA40)
We apply Theorem 4.12 from Ledoux and Talagrand (1991) with F as the identity function and

$$\psi_i(s) = \psi(s) = \min\left(\frac{s^2}{2M}, \frac{M}{2}\right).$$

This is a weak contraction (i.e. 1-Lipschitz) because it is continuous, differentiable on $(-M, M)$ with derivative bounded by $|\psi'(s)| \le |s|/M \le 1$, and constant outside $(-M, M)$. Note that since $|t_i| \le M$ by definition, we have $\psi_i(t_i) = t_i^2/(2M)$. Hence by Theorem 4.12 from Ledoux and Talagrand (1991),

$$\mathbb{E}\left[F\left(\frac{1}{2}\sup_{t\in\mathcal{T}}\left|\sum_{i=1}^n \varepsilon_i\psi_i(t_i)\right|\right)\right] \le \mathbb{E}\left[F\left(\sup_{t\in\mathcal{T}}\left|\sum_{i=1}^n \varepsilon_i t_i\right|\right)\right], \quad \mathbb{E}\left[\frac{1}{2}\sup_{t\in\mathcal{T}}\left|\sum_{i=1}^n \varepsilon_i \frac{t_i^2}{2M}\right|\right] \le \mathbb{E}\left[\sup_{t\in\mathcal{T}}\left|\sum_{i=1}^n \varepsilon_i t_i\right|\right],$$
$$\mathbb{E}\left[\sup_{t\in\mathcal{T}}\left|\sum_{i=1}^n \varepsilon_i t_i^2\right|\right] \le 4M\,\mathbb{E}\left[\sup_{t\in\mathcal{T}}\left|\sum_{i=1}^n \varepsilon_i t_i\right|\right].$$

To see the corollary, set $\mathcal{T} = \big\{\big(f(X_1), \ldots, f(X_n)\big) : f \in \mathcal{F}\big\}$ and note that for fixed realization $X_1, \ldots, X_n$,

$$\mathbb{E}_\varepsilon\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)^2\right|\right] = \mathbb{E}_\varepsilon\left[\sup_{t\in\mathcal{T}}\left|\sum_{i=1}^n \varepsilon_i t_i^2\right|\right] \le 4\mathbb{E}_\varepsilon\left[M\sup_{t\in\mathcal{T}}\left|\sum_{i=1}^n \varepsilon_i t_i\right|\right] = 4\mathbb{E}_\varepsilon\left[M\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)\right|\right].$$

Taking an expectation over $X_1, \ldots, X_n$ and applying Fubini's theorem yields the result. □

Lemma SA41 (A Hoffmann–Jørgensen inequality)
Let $(S, \mathcal{S})$ be a measurable space and $X_1, \ldots, X_n$ be S-valued random variables. Suppose that $\mathcal{F}$ is a pointwise

measurable class of functions from S to $\mathbb{R}$ with finite envelope F. Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent Rademacher random variables which are independent of $X_1, \ldots, X_n$. Then for any $q \in (1, \infty)$,

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)\right|^q\right]^{1/q} \leq C_q\left(\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)\right|\right] + \mathbb{E}\left[\max_{1\leq i\leq n}\sup_{f\in\mathcal{F}}\left|f(X_i)\right|^q\right]^{1/q}\right),$$

where C_q is a positive constant depending only on q.

Proof (Lemma SA41)
We use Talagrand's formulation of a Hoffmann–Jørgensen inequality. Consider the independent $\ell^\infty(\mathcal{F})$-valued random functionals u_i defined by $u_i(f) = \varepsilon_i f(X_i)$, where $\ell^\infty(\mathcal{F})$ is the Banach space of bounded functions from $\mathcal{F}$ to $\mathbb{R}$, equipped with the norm $\|u\|_\mathcal{F} = \sup_{f\in\mathcal{F}}|u(f)|$. Then Remark 3.4 in Kwapien et al. (1991) gives

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n u_i(f)\right|^q\right]^{1/q} \leq C_q\left(\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n u_i(f)\right|\right] + \mathbb{E}\left[\max_{1\leq i\leq n}\sup_{f\in\mathcal{F}}|u_i(f)|^q\right]^{1/q}\right)$$

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)\right|^q\right]^{1/q} \leq C_q\left(\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)\right|\right] + \mathbb{E}\left[\max_{1\leq i\leq n}\sup_{f\in\mathcal{F}}\left|f(X_i)\right|^q\right]^{1/q}\right).$$

□

Proof (Lemma SA24)
We follow the proof of Theorem 5.2 from Chernozhukov et al. (2014b), using our i.n.i.d. versions of the symmetrization inequality (Lemma SA39), Rademacher contraction principle (Lemma SA40) and Hoffmann–Jørgensen inequality (Lemma SA41).

Without loss of generality, we may assume that $J(1, \mathcal{F}, F) < \infty$ as otherwise there is nothing to prove, and that $F > 0$ everywhere on S. Let $\mathbb{P}_n = n^{-1}\sum_i \delta_{X_i}$ be the empirical distribution of X_i, and define the empirical variance bound $\sigma_n^2 = \sup_\mathcal{F} n^{-1}\sum_i f(X_i)^2$. By the i.n.i.d. symmetrization inequality (Lemma SA39),

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|G_n(f)\right|\right] = \frac{1}{\sqrt{n}}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n\Big(f(X_i) - \mathbb{E}[f(X_i)]\Big)\right|\right] \leq \frac{2}{\sqrt{n}}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)\right|\right],$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent Rademacher random variables which are independent of $X_1, \ldots, X_n$. Then the standard entropy integral inequality from the proof of Theorem 5.2 in the supplemental materials for Chernozhukov et al. (2014b) gives for a universal constant $C_1 > 0$,

$$\frac{1}{\sqrt{n}}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)\right| \;\middle|\; X_1, \ldots, X_n\right] \leq C_1\|F\|_{\mathbb{P}_n,2}J(\sigma_n/\|F\|_{\mathbb{P}_n,2}, \mathcal{F}, F).$$

Taking marginal expectations and applying Jensen's inequality along with a convexity result for the covering integral, as in Lemma A.2 in Chernozhukov et al. (2014b), gives

$$Z := \frac{1}{\sqrt{n}}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)\right|\right] \leq C_1\|F\|_{\mathbb{P},2}J(\mathbb{E}[\sigma_n^2]^{1/2}/\|F\|_{\mathbb{P},2}, \mathcal{F}, F).$$

Next we use the symmetrization inequality (Lemma SA39), the contraction principle (Lemma SA40), the

Cauchy–Schwarz inequality and the Hoffmann–Jørgensen inequality (Lemma SA41) to deduce that

$$\begin{aligned}
\mathbb{E}[\sigma_n^2] &= \mathbb{E}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n f(X_i)^2\right] \leq \sup_{f\in\mathcal{F}} \mathbb{E}_{\bar{\mathbb{P}}}\left[f(X_i)^2\right] + \frac{1}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n f(X_i)^2 - \mathbb{E}\left[f(X_i)^2\right]\right|\right] \\
&\leq \sigma^2 + \frac{2}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)^2\right|\right] \leq \sigma^2 + \frac{8}{n}\mathbb{E}\left[M\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)\right|\right] \\
&\leq \sigma^2 + \frac{8}{n}\mathbb{E}\left[M^2\right]^{1/2}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)\right|^2\right]^{1/2} \\
&\leq \sigma^2 + \frac{8}{n}\|M\|_{\mathbb{P},2}\, C_2\left(\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(X_i)\right|\right] + \mathbb{E}\left[\max_{1\leq i\leq n}\sup_{f\in\mathcal{F}}\left|f(X_i)\right|^2\right]^{1/2}\right) \\
&= \sigma^2 + \frac{8C_2}{n}\|M\|_{\mathbb{P},2}\left(\sqrt{n}Z + \|M\|_{\mathbb{P},2}\right) \lesssim \sigma^2 + \frac{\|M\|_{\mathbb{P},2}Z}{\sqrt{n}} + \frac{\|M\|_{\mathbb{P},2}^2}{n},
\end{aligned}$$

where $\lesssim$ indicates a bound up to a universal constant. Hence taking a square root we see that, following the notation from the proof of Theorem 5.2 in the supplemental materials to Chernozhukov et al. (2014b),

$$\sqrt{\mathbb{E}[\sigma_n^2]} \lesssim \sigma + \|M\|_{\mathbb{P},2}^{1/2} Z^{1/2} n^{-1/4} + \|M\|_{\mathbb{P},2} n^{-1/2} \lesssim \|F\|_{\bar{\mathbb{P}},2}\left(\Delta \vee \sqrt{DZ}\right),$$

where $\Delta^2 = \|F\|_{\bar{\mathbb{P}},2}^{-2}\big(\sigma^2 \vee (\|M\|_{\mathbb{P},2}^2/n)\big) \geq \delta^2$ and $D = \|M\|_{\mathbb{P},2} n^{-1/2}\|F\|_{\bar{\mathbb{P}},2}^{-2}$. Thus returning to our bound on Z, we now have

$$Z \lesssim \|F\|_{\bar{\mathbb{P}},2} J(\Delta \vee \sqrt{DZ}, \mathcal{F}, F).$$

The final steps proceed as in the proof of Theorem 5.2 from Chernozhukov et al. (2014b), considering cases separately for $\Delta \geq \sqrt{DZ}$ and $\Delta < \sqrt{DZ}$, and applying convexity properties of the entropy integral J. □

Proof (Lemma SA25)
We are assuming the VC-type condition that

$$\sup_{\mathbb{Q}} N(\mathcal{F}, \rho_{\mathbb{Q}}, \varepsilon\|F\|_{\mathbb{Q},2}) \leq (C_1/\varepsilon)^{C_2}$$

for all $\varepsilon \in (0,1]$, for some constants $C_1 \geq e$ and $C_2 \geq 1$. Hence for $\delta \in (0,1]$, the entropy integral can be bounded as follows.

$$\begin{aligned}
J\big(\delta, \mathcal{F}, F\big) &= \int_0^\delta \sqrt{1 + \sup_{\mathbb{Q}} \log N(\mathcal{F}, \rho_{\mathbb{Q}}, \varepsilon\|F\|_{\mathbb{Q},2})}\,\mathrm{d}\varepsilon \leq \int_0^\delta \sqrt{1 + C_2\log(C_1/\varepsilon)}\,\mathrm{d}\varepsilon \\
&\leq \int_0^\delta \left(1 + \sqrt{C_2\log(C_1/\varepsilon)}\right)\mathrm{d}\varepsilon = \delta + \sqrt{C_2}\int_0^\delta \sqrt{\log(C_1/\varepsilon)}\,\mathrm{d}\varepsilon \\
&\leq \delta + \sqrt{\frac{C_2}{\log(C_1/\delta)}}\int_0^\delta \log(C_1/\varepsilon)\,\mathrm{d}\varepsilon = \delta + \sqrt{\frac{C_2}{\log(C_1/\delta)}}\big(\delta + \delta\log(C_1/\delta)\big) \leq 3\delta\sqrt{C_2\log(C_1/\delta)}.
\end{aligned}$$

The remaining equations now follow by Lemma SA24. □

SA5.2.2 Strong approximation results

Before proving Lemma SA26, we require the elementary characterization of bounded-variation functions given in Lemma SA42.

Lemma SA42 (A characterization of bounded-variation functions)
Let $\mathcal{V}_1$ be the class of real-valued functions on $[0,1]$ which are 0 at 1 and have total variation bounded by 1.

Also define the class of half-interval indicator functions $\mathcal{I} = \{\mathbb{I}[0,t] : t \in [0,1]\}$. *For any topological vector space* $\mathcal{X}$, *define the symmetric convex hull of a subset* $\mathcal{Y} \subseteq \mathcal{X}$ *as*

$$\operatorname{symconv} \mathcal{Y} = \left\{ \sum_{i=1}^{n} \lambda_i y_i : \sum_{i=1}^{n} \lambda_i = 1,\ \lambda_i \geq 0,\ y_i \in \mathcal{Y} \cup -\mathcal{Y},\ n \in \mathbb{N} \right\}.$$

Denote its topological closure by $\overline{\operatorname{symconv}}\ \mathcal{Y}$. *Then under the topology induced by pointwise convergence,*

$$\mathcal{V}_1 \subseteq \overline{\operatorname{symconv}}\ \mathcal{I}.$$

Proof (Lemma SA42)
Firstly, let $\mathcal{D} \subseteq \mathcal{V}_1$ be the class of real-valued functions on $[0,1]$ which are 0 at 1, have total variation exactly 1, and are weakly monotone decreasing. Therefore for $g \in \mathcal{D}$, we have $\|g\|_{\mathrm{TV}} = g(0) = 1$. Let $S = \{s_1, s_2, \dots\} \subseteq [0,1]$ be the countable set of discontinuity points of g. We want to find a sequence of convex combinations of elements of $\mathcal{I}$ which converges pointwise to g. To do this, first define the sequence of meshes

$$A_n = \{s_k : 1 \leq k \leq n\} \cup \{k/n : 0 \leq k \leq n\}$$

which satisfies $\bigcup_n A_n = S \cup ([0,1] \cap \mathbb{Q})$. Endow A_n with the ordering induced by the canonical order on $\mathbb{R}$, giving $A_n = \{a_1, a_2, \dots\}$, and define the sequence of functions

$$g_n(x) = \sum_{k=1}^{|A_n|-1} \mathbb{I}[0, a_k]\big(g(a_k) - g(a_{k+1})\big),$$

where clearly $\mathbb{I}[0,a_k] \in \mathcal{I}$ and $g(a_k) - g(a_{k+1}) \geq 0$ and $\sum_{k=1}^{|A_n|-1} \big(g(a_k) - g(a_{k+1})\big) = g(0) - g(1) = 1$. Therefore g_n is a convex combination of elements of $\mathcal{I}$. Further, note that for $a_k \in A_n$, we have

$$g_n(a_k) = \sum_{j=k}^{|A_n|-1} \big(g(a_j) - g(a_{j+1})\big) = g(a_k) - g(a_{|A_n|}) = g(a_k) - g(1) = g(a_k).$$

Hence if $x \in S$, then eventually $x \in A_n$ so $g_n(x) \to g(x)$. Alternatively in $x \notin S$, then g is continuous at x. But $g_n \to g$ on the dense set $\bigcup_n A_n$, so also $g_n(x) \to g(x)$. Hence $g_n \to g$ pointwise on $[0,1]$.

Now take $f \in \mathcal{V}_1$. By the Jordan decomposition for total variation functions (Royden and Fitzpatrick, 1988), we can write $f = f^+ - f^-$, with f^+ and f^- weakly decreasing, $f^+(1) = f^-(1) = 0$, and $\|f^+\|_{\mathrm{TV}} + \|f^-\|_{\mathrm{TV}} = \|f\|_{\mathrm{TV}}$. Supposing that both $\|f^+\|_{\mathrm{TV}}$ and $\|f^-\|_{\mathrm{TV}}$ are strictly positive, let g_n^+ approximate the unit-variation function $f^+/\|f^+\|_{\mathrm{TV}}$ and g_n^- approximate $f^-/\|f^-\|_{\mathrm{TV}}$ as above. Then since trivially

$$f = \|f^+\|_{\mathrm{TV}} f^+/\|f^+\|_{\mathrm{TV}} - \|f^-\|_{\mathrm{TV}} f^-/\|f^-\|_{\mathrm{TV}} + \big(1 - \|f^+\|_{\mathrm{TV}} - \|f^-\|_{\mathrm{TV}}\big) \cdot 0,$$

we have that the convex combination

$$g_n^+ \|f^+\|_{\mathrm{TV}} - g_n^- \|f^-\|_{\mathrm{TV}} + \big(1 - \|f^+\|_{\mathrm{TV}} - \|f^-\|_{\mathrm{TV}}\big) \cdot 0$$

converges pointwise to f. This also holds if either of the total variations $\|f^\pm\|_{\mathrm{TV}}$ are zero, since then the corresponding sequence $g_n^\pm$ need not be defined. Now note that each of g_n^+, $-g_n^-$ and 0 are in $\operatorname{symconv} \mathcal{I}$, so $f \in \overline{\operatorname{symconv}}\ \mathcal{I}$ under pointwise convergence. □

Proof (Lemma SA26)
We follow the Gaussian approximation method given in Section 2 of Giné et al. (2004). The KMT approximation theorem (Komlós et al., 1975) asserts the existence of a probability space carrying n i.i.d. uniform random variables $\xi_1, \dots, \xi_n \sim \mathcal{U}[0,1]$ and a standard Brownian motion $B_n(s) : s \in [0,1]$ such that if

$$\alpha_n(s) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\mathbb{I}\{\xi_i \leq s\} - s\big), \qquad \beta_n(s) := B_n(s) - sB_n(1),$$

then for some universal positive constants C_1, C_2, C_3 and for all $t > 0$,

$$\mathbb{P}\left(\sup_{s\in[0,1]} \big|\alpha_n(s) - \beta_n(s)\big| > \frac{t + C_1 \log n}{\sqrt{n}}\right) \leq C_2 e^{-C_3 t}.$$

We can view α_n and β_n as random functionals defined on the class of half-interval indicator functions $\mathcal{I} = \big\{\mathbb{I}[0,s] : s \in [0,1]\big\}$ in the following way.

$$\alpha_n(\mathbb{I}[0,s]) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\mathbb{I}[0,s](\xi_i) - \mathbb{E}[\mathbb{I}[0,s](\xi_i)]\big),$$
$$\beta_n(\mathbb{I}[0,s]) = \int_0^1 \mathbb{I}[0,s](u)\,\mathrm{d}B_n(u) - B_n(1)\int_0^1 \mathbb{I}[0,s](u)\,\mathrm{d}u,$$

where the integrals are defined as Itô and Riemann–Stieltjes integrals in the usual way for stochastic integration against semimartingales (Le Gall, 2016, Chapter 5). Now we extend their definitions to the class $\mathcal{V}_1$ of functions on $[0,1]$ which are 0 at 1 and have total variation bounded by 1. This is achieved by noting that by Lemma SA42, we have $\mathcal{V}_1 \subseteq \overline{\text{symconv}}\,\mathcal{I}$ where $\overline{\text{symconv}}\,\mathcal{I}$ is the smallest symmetric convex class containing $\mathcal{I}$ which is closed under pointwise convergence. Thus by the dominated convergence theorem, every function in $\mathcal{V}_1$ is approximated in L^2 by finite convex combinations of functions in $\pm\mathcal{I}$, and the extension to $g \in \mathcal{V}_1$ follows by linearity and L^2 convergence of (stochastic) integrals:

$$\alpha_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(g(\xi_i) - \mathbb{E}[g(\xi_i)]\big), \qquad \beta_n(g) = \int_0^1 g(s)\,\mathrm{d}B_n(s) - B_n(1)\int_0^1 g(s)\,\mathrm{d}s.$$

Now we show that the norm induced on $(\alpha_n - \beta_n)$ by the function class $\mathcal{V}_1$ is a.s. identical to the supremum norm. Writing the sums as integrals and using integration by parts for finite-variation Lebesgue–Stieltjes and Itô integrals, and recalling that $g(1) = \alpha_n(0) = B_n(0) = 0$, we see

$$\begin{aligned}
\sup_{g\in\mathcal{V}_1} \big|\alpha_n(g) - \beta_n(g)\big| &= \sup_{g\in\mathcal{V}_1} \left|\int_0^1 g(s)\,\mathrm{d}\alpha_n(s) - \int_0^1 g(s)\,\mathrm{d}B_n(s) + B_n(1)\int_0^1 g(s)\,\mathrm{d}s\right| \\
&= \sup_{g\in\mathcal{V}_1} \left|\int_0^1 \alpha_n(s)\,\mathrm{d}g(s) - \int_0^1 B_n(s)\,\mathrm{d}g(s) + B_n(1)\int_0^1 s\,\mathrm{d}g(s)\right| \\
&= \sup_{g\in\mathcal{V}_1} \left|\int_0^1 \big(\alpha_n(s) - \beta_n(s)\big)\,\mathrm{d}g(s)\right| = \sup_{s\in[0,1]} \big|\alpha_n(s) - \beta_n(s)\big|,
\end{aligned}$$

where in the last line the upper bound is because $\|g\|_{\text{TV}} \leq 1$ and the lower bound is by taking $g_\varepsilon = \pm\mathbb{I}[0, s_\varepsilon]$ where $|\alpha_n(s_\varepsilon) - \beta_n(s_\varepsilon)| \geq \sup_s |\alpha_n(s) - \beta_n(s)| - \varepsilon$. Hence we obtain

$$\mathbb{P}\left(\sup_{g\in\mathcal{V}_1} \big|\alpha_n(g) - \beta_n(g)\big| > \frac{t + C_1 \log n}{\sqrt{n}}\right) \leq C_2 e^{-C_3 t}. \tag{6}$$

Now define $V_n = \sup_{x\in\mathbb{R}} \|g_n(\cdot, x)\|_{\text{TV}}$, noting that if $V_n = 0$ then the result is trivially true by setting $Z_n = 0$. Let F_X be the common c.d.f. of X_i, and define the quantile function $F_X^{-1}(s) = \inf\{u : F_X(u) \geq s\}$ for $s \in [0,1]$, writing $\inf\emptyset = \infty$ and $\inf\mathbb{R} = -\infty$. Consider the function class

$$\mathcal{G}_n = \big\{V_n^{-1} g_n\big(F_X^{-1}(\cdot), x\big) - V_n^{-1} g_n\big(F_X^{-1}(1), x\big) : x \in \mathbb{R}\big\},$$

noting that $g_n(\cdot, x)$ is finite-variation so $g_n(\pm\infty, x)$ can be interpreted as the relevant limit. By monotonicity of F_X and the definition of V_n, the members of $\mathcal{G}_n$ have total variation of at most 1 and are 0 at 1, implying that $\mathcal{G}_n \subseteq \mathcal{V}_1$. Noting that α_n and β_n are random linear operators which a.s. annihilate constant functions, define

$$Z_n(x) = \beta_n\Big(g_n\big(F_X^{-1}(\cdot), x\big)\Big) = V_n \beta_n\Big(V_n^{-1} g_n\big(F_X^{-1}(\cdot), x\big) - V_n^{-1} g_n\big(F_X^{-1}(1), x\big)\Big),$$

which is a mean-zero Gaussian process with continuous trajectories. Its covariance structure is

$$
\begin{aligned}
&\mathbb{E}[Z_n(x)Z_n(x')] \\
&= \mathbb{E}\Bigg[\left(\int_0^1 g_n\big(F_X^{-1}(s),x\big)\,\mathrm{d}B_n(s) - B_n(1)\int_0^1 g_n\big(F_X^{-1}(s),x\big)\,\mathrm{d}s \right) \\
&\quad \times \left(\int_0^1 g_n\big(F_X^{-1}(s),x'\big)\,\mathrm{d}B_n(s) - B_n(1)\int_0^1 g_n\big(F_X^{-1}(s),x'\big)\,\mathrm{d}s \right) \Bigg] \\
&= \mathbb{E}\left[\int_0^1 g_n\big(F_X^{-1}(s),x\big)\,\mathrm{d}B_n(s) \int_0^1 g_n\big(F_X^{-1}(s),x'\big)\,\mathrm{d}B_n(s) \right] \\
&\quad - \int_0^1 g_n\big(F_X^{-1}(s),x\big)\,\mathrm{d}s\ \mathbb{E}\left[B_n(1)\int_0^1 g_n\big(F_X^{-1}(s),x'\big)\,\mathrm{d}B_n(s) \right] \\
&\quad - \int_0^1 g_n\big(F_X^{-1}(s),x'\big)\,\mathrm{d}s\ \mathbb{E}\left[B_n(1)\int_0^1 g_n\big(F_X^{-1}(s),x\big)\,\mathrm{d}B_n(s) \right] \\
&\quad + \int_0^1 g_n\big(F_X^{-1}(s),x\big)\,\mathrm{d}s \int_0^1 g_n\big(F_X^{-1}(s),x'\big)\,\mathrm{d}s\ \mathbb{E}\left[B_n(1)^2\right] \\
&= \int_0^1 g_n\big(F_X^{-1}(s),x\big)g_n\big(F_X^{-1}(s),x'\big)\,\mathrm{d}s - \int_0^1 g_n\big(F_X^{-1}(s),x\big)\,\mathrm{d}s \int_0^1 g_n\big(F_X^{-1}(s),x'\big)\,\mathrm{d}s \\
&= \mathbb{E}\Big[g_n\big(F_X^{-1}(\xi_i),x\big)g_n\big(F_X^{-1}(\xi_i),x'\big)\Big] - \mathbb{E}\Big[g_n\big(F_X^{-1}(\xi_i),x\big)\Big]\mathbb{E}\Big[g_n\big(F_X^{-1}(\xi_i),x'\big)\Big] \\
&= \mathbb{E}\Big[g_n\big(X_i,x\big)g_n\big(X_i,x'\big)\Big] - \mathbb{E}\Big[g_n\big(X_i,x\big)\Big]\mathbb{E}\Big[g_n\big(X_i,x'\big)\Big] = \mathbb{E}\big[G_n(x)G_n(x')\big]
\end{aligned}
$$

as desired, where we used the Itô isometry for stochastic integrals, writing $B_n(1) = \int_0^1 \mathrm{d}B_n(s)$; and noting that $F_X^{-1}(\xi_i)$ has the same distribution as X_i. Finally, note that

$$
G_n(x) = \alpha_n\Big(g_n\big(F_X^{-1}(\cdot),x\big)\Big) = V_n\alpha_n\Big(V_n^{-1}g_n\big(F_X^{-1}(\cdot),x\big) - V_n^{-1}g_n\big(F_X^{-1}(1),x\big)\Big),
$$

and so by Equation 6

$$
\mathbb{P}\left(\sup_{x\in\mathbb{R}}\Big|G_n(x) - Z_n(x)\Big| > V_n\frac{t + C_1\log n}{\sqrt{n}}\right) \leq \mathbb{P}\left(\sup_{g\in\mathcal{V}_1}\big|\alpha_n(g) - \beta_n(g)\big| > \frac{t + C_1\log n}{\sqrt{n}}\right) \leq C_2 e^{-C_3 t}.
$$

□

Proof (Lemma SA27)
Take $0 < \delta_n \leq \mathrm{Leb}(\mathcal{X}_n)$ and let $\mathcal{X}_n^\delta = \{x_1,\ldots,x_{|\mathcal{X}_n^\delta|}\}$ be a δ_n-covering of $\mathcal{X}_n$ with cardinality $|\mathcal{X}_n^\delta| \leq \mathrm{Leb}(\mathcal{X}_n)/\delta_n$. Suppose that $|\log\delta_n| \lesssim C_1\log n$ up to a universal constant. We first use the Yurinskii coupling to construct a Gaussian process Z_n which is close to G_n on this finite cover. Then we bound the fluctuations in G_n and in Z_n using entropy methods.

Part 1: Yurinskii coupling
Define the i.n.i.d. and mean-zero variables

$$
h_i(x) = \frac{1}{\sqrt{n}}\Big(g_n(X_i',x) - \mathbb{E}[g_n(X_i',x)]\Big),
$$

where $X_1',\ldots,X_n'$ are independent copies of $X_1,\ldots,X_n$ on some new probability space, so that we have $G_n(x) = \sum_{i=1}^n h_i(x)$ in distribution. Also define the length-$|\mathcal{X}_n^\delta|$ random vector

$$
h_i^\delta = \big(h_i(x) : x\in\mathcal{X}_n^\delta\big).
$$

By an extension of Yurinskii's coupling to general norms (Belloni et al., 2019, supplemental materials, Lemma 38), there exists on the new probability space a Gaussian length-$|\mathcal{X}_n^\delta|$ vector Z_n^δ which is mean-zero and with the same covariance structure as $\sum_{i=1}^n h_i^\delta$ satisfying

$$
\mathbb{P}\left(\left\|\sum_{i=1}^n h_i^\delta - Z_n^\delta\right\|_\infty > 3t_n\right) \leq \min_{s>0}\left(2\mathbb{P}\big(\|N\|_\infty > s\big) + \frac{\beta s^2}{t_n^3}\right),
$$

where

$$\beta = \sum_{i=1}^{n} \Big(\mathbb{E}\big[\|h_i^\delta\|_2^2 \, \|h_i^\delta\|_\infty \big] + \mathbb{E}\big[\|z_i\|_2^2 \, \|z_i\|_\infty \big] \Big),$$

with $z_i \sim \mathcal{N}(0, \mathrm{Var}[h_i^\delta])$ independent and $N \sim \mathcal{N}(0, I_{|\mathcal{X}_n^\delta|})$. By the almost sure and variance bounds on g_n,

$$\mathbb{E}\big[\|h_i^\delta\|_2^2 \, \|h_i^\delta\|_\infty \big] \le \frac{M_n}{\sqrt{n}} \mathbb{E}\big[\|h_i^\delta\|_2^2 \big] = \frac{M_n}{\sqrt{n}} \sum_{x \in \mathcal{X}_n^\delta} \mathbb{E}\big[h_i(x)^2 \big] \le \frac{M_n}{\sqrt{n}} \frac{|\mathcal{X}_n^\delta| \sigma_n^2}{n} \le \frac{M_n \sigma_n^2 \, \mathrm{Leb}(\mathcal{X}_n)}{n^{3/2} \delta_n}.$$

By the fourth moment bound for Gaussian variables,

$$\begin{aligned} \mathbb{E}\big[\|z_i\|_2^4 \big] &\le |\mathcal{X}_n^\delta| \, \mathbb{E}\big[\|z_i\|_4^4 \big] \le |\mathcal{X}_n^\delta|^2 \max_j \mathbb{E}\big[(z_i^{(j)})^4 \big] \le 3 |\mathcal{X}_n^\delta|^2 \max_j \mathbb{E}\big[(z_i^{(j)})^2 \big]^2 \\ &= 3 |\mathcal{X}_n^\delta|^2 \max_{x \in \mathcal{X}_n^\delta} \mathbb{E}\big[h_i(x)^2 \big]^2 \le \frac{3 \sigma_n^4 \, \mathrm{Leb}(\mathcal{X}_n)^2}{n^2 \delta_n^2}. \end{aligned}$$

Also by Jensen's inequality and for $|\mathcal{X}_n^\delta| \ge 2$, assuming $C_1 > 1$ without loss of generality,

$$\begin{aligned} \mathbb{E}\big[\|z_i\|_\infty^2 \big] &\le \frac{4\sigma_n^2}{n} \log \mathbb{E}\big[e^{\|z_i\|_\infty^2 / (4\sigma_n^2/n)} \big] \le \frac{4\sigma_n^2}{n} \log \mathbb{E}\left[\sum_{j=1}^{|\mathcal{X}_n^\delta|} e^{(z_i^{(j)})^2 / (4\sigma_n^2/n)} \right] \le \frac{4\sigma_n^2}{n} \log \big(2 |\mathcal{X}_n^\delta| \big) \\ &\le \frac{4\sigma_n^2}{n} \left(\log 2 + \log \mathrm{Leb}(\mathcal{X}_n) - \log \delta_n \right) \le \frac{12 C_1 \sigma_n^2 \log n}{n}, \end{aligned}$$

where we used the moment generating function of a χ_1^2 random variable. Therefore we can apply the Cauchy–Schwarz inequality to obtain

$$\mathbb{E}\big[\|z_i\|_2^2 \, \|z_i\|_\infty \big] \le \sqrt{\mathbb{E}\big[\|z_i\|_2^4 \big]} \sqrt{\mathbb{E}\big[\|z_i\|_\infty^2 \big]} \le \sqrt{\frac{3\sigma_n^4 \, \mathrm{Leb}(\mathcal{X}_n)^2}{n^2 \delta_n^2}} \sqrt{\frac{12 C_1 \sigma_n^2 \log n}{n}} \le \frac{6 \sigma_n^3 \, \mathrm{Leb}(\mathcal{X}_n) \sqrt{C_1 \log n}}{n^{3/2} \delta_n}.$$

Now summing over the n samples gives

$$\beta \le \frac{M_n \sigma_n^2 \, \mathrm{Leb}(\mathcal{X}_n)}{\sqrt{n} \delta_n} + \frac{6 \sigma_n^3 \, \mathrm{Leb}(\mathcal{X}_n) \sqrt{C_1 \log n}}{\sqrt{n} \delta_n} = \frac{\sigma_n^2 \, \mathrm{Leb}(\mathcal{X}_n)}{\sqrt{n} \delta_n} \Big(M_n + 6 \sigma_n \sqrt{C_1 \log n} \Big).$$

By a union bound and Gaussian tail probabilities, we have that $\mathbb{P}\big(\|N\|_\infty > s \big) \le 2 |\mathcal{X}_n^\delta| e^{-s^2/2}$. Thus we get the following Yurinskii coupling inequality for all $s > 0$:

$$\mathbb{P}\left(\left\| \sum_{i=1}^{n} h_i^\delta - Z_n^\delta \right\|_\infty > t_n \right) \le \frac{4 \, \mathrm{Leb}(\mathcal{X}_n)}{\delta_n} e^{-s^2/2} + \frac{\sigma_n^2 \, \mathrm{Leb}(\mathcal{X}_n) s^2}{\sqrt{n} \delta_n t_n^3} \Big(M_n + 6 \sigma_n \sqrt{C_1 \log n} \Big).$$

Note that Z_n^δ now extends by the Vorob'ev–Berkes–Philipp Theorem (Lemma SA28) to a mean-zero Gaussian process Z_n on the compact interval $\mathcal{X}_n$ with covariance structure given by

$$\mathbb{E}\big[Z_n(x) Z_n(x') \big] = \mathbb{E}\big[G_n(x) G_n(x') \big],$$

satisfying for any $s' > 0$

$$\mathbb{P}\left(\sup_{x \in \mathcal{X}_n^\delta} \big| G_n(x) - Z_n(x) \big| > t_n \right) \le \frac{4 \, \mathrm{Leb}(\mathcal{X}_n)}{\delta_n} e^{-s^2/2} + \frac{\sigma_n^2 \, \mathrm{Leb}(\mathcal{X}_n) s^2}{\sqrt{n} \delta_n t_n^3} \Big(M_n + 6 \sigma_n \sqrt{C_1 \log n} \Big).$$

Part 2: Regularity of G_n

Next we bound the fluctuations in the empirical process G_n. Consider the following classes of functions on S and their associated (constant) envelope functions. By continuity of g_n, each class is pointwise measurable (to see this, restrict the index sets to rationals).

$$\begin{aligned}\mathcal{G}_n &= \{g_n(\cdot, x) : x \in \mathcal{X}_n\},\\ \mathrm{Env}(\mathcal{G}_n) &= M_n,\\ \mathcal{G}_n^\delta &= \{g_n(\cdot, x) - g_n(\cdot, x') : x, x' \in \mathcal{X}_n, |x - x'| \le \delta_n\},\\ \mathrm{Env}(\mathcal{G}_n^\delta) &= l_{n,\infty}\delta_n.\end{aligned}$$

We first show that for each n these are VC-type classes. To see this, note that by the uniform Lipschitz assumption we have that

$$\big\|g_n(\cdot, x) - g_n(\cdot, x')\big\|_\infty \le l_{n,\infty}|x - x'|$$

for all $x, x' \in \mathcal{X}_n$. Therefore with $\mathbb{Q}$ ranging over the finitely-supported distributions on $(S, \mathcal{S})$, noting that any $\|\cdot\|_\infty$-cover is a $\rho_\mathbb{Q}$-cover,

$$\sup_\mathbb{Q} N\big(\mathcal{G}_n, \rho_\mathbb{Q}, \varepsilon l_{n,\infty}\operatorname{Leb}(\mathcal{X}_n)\big) \le N\big(\mathcal{G}_n, \|\cdot\|_\infty, \varepsilon l_{n,\infty}\operatorname{Leb}(\mathcal{X}_n)\big) \le N\big(\mathcal{X}_n, |\cdot|, \varepsilon \operatorname{Leb}(\mathcal{X}_n)\big) \le 1/\varepsilon.$$

Replacing ε by $\varepsilon M_n/(l_{n,\infty}\operatorname{Leb}(\mathcal{X}_n))$ gives

$$\sup_\mathbb{Q} N\big(\mathcal{G}_n, \rho_\mathbb{Q}, \varepsilon M_n\big) \le \frac{l_{n,\infty}\operatorname{Leb}(\mathcal{X}_n)}{\varepsilon M_n},$$

and so $\mathcal{G}_n$ is a VC-type class. To see that $\mathcal{G}_n^\delta$ is also a VC-type class, we construct a cover in the following way. Let $\mathcal{F}_n$ be an ε-cover for $(\mathcal{G}_n, \|\cdot\|_\infty)$. Then by the triangle inequality, $\mathcal{F}_n - \mathcal{F}_n$ is a 2ε-cover for $(\mathcal{G}_n - \mathcal{G}_n, \|\cdot\|_\infty)$ of cardinality at most $|\mathcal{F}_n|^2$, where the subtractions are defined as set subtractions. Since $\mathcal{G}_n^\delta \subseteq \mathcal{G}_n - \mathcal{G}_n$, we see that $\mathcal{F}_n - \mathcal{F}_n$ is a 2ε-external cover for $\mathcal{G}_n^\delta$. Thus

$$\sup_\mathbb{Q} N\big(\mathcal{G}_n^\delta, \rho_\mathbb{Q}, \varepsilon l_{n,\infty}\operatorname{Leb}(\mathcal{X}_n)\big) \le N\big(\mathcal{G}_n^\delta, \|\cdot\|_\infty, \varepsilon l_{n,\infty}\operatorname{Leb}(\mathcal{X}_n)\big) \le N\big(\mathcal{G}_n, \|\cdot\|_\infty, \varepsilon l_{n,\infty}\operatorname{Leb}(\mathcal{X}_n)\big)^2 \le 1/\varepsilon^2.$$

Replacing ε by $\varepsilon\delta_n/\operatorname{Leb}(\mathcal{X}_n)$ gives

$$\sup_\mathbb{Q} N\big(\mathcal{G}_n^\delta, \rho_\mathbb{Q}, \varepsilon l_{n,\infty}\delta_n\big) \le \frac{\operatorname{Leb}(\mathcal{X}_n)^2}{\varepsilon^2\delta_n^2} \le (C_{1,n}/\varepsilon)^2$$

with $C_{1,n} = \operatorname{Leb}(\mathcal{X}_n)/\delta_n$, demonstrating that $\mathcal{G}_n^\delta$ forms a VC-type class. We now apply the maximal inequality for i.n.i.d. data given in Lemma SA25. To do this, note that $\sup_{\mathcal{G}_n^\delta}\|g\|_{\bar{\mathbb{P}},2} \le l_{n,2}\delta_n$ by the L^2 Lipschitz condition, and recall $\mathrm{Env}(\mathcal{G}_n^\delta) = l_{n,\infty}\delta_n$. Therefore Lemma SA25 with $\|F\|_{\bar{\mathbb{P}},2} = l_{n,\infty}\delta_n$, $\|M\|_{\mathbb{P},2} = l_{n,\infty}\delta_n$ and $\sigma = l_{n,2}\delta_n$ gives, up to universal constants

$$\begin{aligned}\mathbb{E}\left[\sup_{g\in\mathcal{G}_n^\delta}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(g(X_i) - \mathbb{E}[g(X_i)]\Big)\right|\right] &\lesssim \sigma\sqrt{2\log\big(C_{1,n}\|F\|_{\bar{\mathbb{P}},2}/\sigma\big)} + \frac{\|M\|_{\mathbb{P},2}2\log\big(C_{1,n}\|F\|_{\bar{\mathbb{P}},2}/\sigma\big)}{\sqrt{n}}\\ &\lesssim l_{n,2}\delta_n\sqrt{C_1\log n} + \frac{l_{n,\infty}\delta_n}{\sqrt{n}}C_1\log n,\end{aligned}$$

and hence by Markov's inequality,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{|x-x'|\le\delta_n}\left|G_n(x)-G_n(x')\right|>t_n\right)\\
&=\mathbb{P}\left(\sup_{|x-x'|\le\delta_n}\frac{1}{\sqrt{n}}\left|\sum_{i=1}^n\Big(g_n(X_i,x)-\mathbb{E}[g_n(X_i,x)]-g_n(X_i,x')+\mathbb{E}[g_n(X_i,x')]\Big)\right|>t_n\right)\\
&=\mathbb{P}\left(\sup_{g\in\mathcal{G}_n^\delta}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(g(X_i)-\mathbb{E}[g(X_i)]\Big)\right|>t_n\right)\le\frac{1}{t}\mathbb{E}\left[\sup_{g\in\mathcal{G}_n^\delta}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(g(X_i)-\mathbb{E}[g(X_i)]\Big)\right|\right]\\
&\lesssim\frac{l_{n,2}\delta_n}{t_n}\sqrt{C_1\log n}+\frac{l_{n,\infty}\delta_n}{t_n\sqrt{n}}C_1\log n.
\end{aligned}$$

Part 3: regularity of Z_n

Next we bound the fluctuations in the Gaussian process Z_n. Let ρ be the following semimetric:

$$\begin{aligned}
\rho(x,x')^2&=\mathbb{E}\big[\big(Z_n(x)-Z_n(x')\big)^2\big]=\mathbb{E}\big[\big(G_n(x)-G_n(x')\big)^2\big]\\
&=\frac{1}{n}\sum_{i=1}^n\mathbb{E}\big[\big(h_i(x)-h_i(x')\big)^2\big]\le l_{n,2}^2\,|x-x'|^2.
\end{aligned}$$

Hence $\rho(x,x')\le l_{n,2}\,|x-x'|$. By the Gaussian process maximal inequality from Lemma SA32, we obtain that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{|x-x'|\le\delta_n}\left|Z_n(x)-Z_n(x')\right|\right]\lesssim\mathbb{E}\left[\sup_{\rho(x,x')\le l_{n,2}\delta_n}\left|Z_n(x)-Z_n(x')\right|\right]\le\int_0^{l_{n,2}\delta_n}\sqrt{\log N(\varepsilon,\mathcal{X}_n,\rho)}\,\mathrm{d}\varepsilon\\
&\le\int_0^{l_{n,2}\delta_n}\sqrt{\log N(\varepsilon/l_{n,2},\mathcal{X}_n,|\cdot|)}\,\mathrm{d}\varepsilon\le\int_0^{l_{n,2}\delta_n}\sqrt{\log\left(1+\frac{\mathrm{Leb}(\mathcal{X}_n)l_{n,2}}{\varepsilon}\right)}\,\mathrm{d}\varepsilon\\
&\le\int_0^{l_{n,2}\delta_n}\sqrt{\log\left(\frac{2\,\mathrm{Leb}(\mathcal{X}_n)l_{n,2}}{\varepsilon}\right)}\,\mathrm{d}\varepsilon\le\log\left(\frac{2\,\mathrm{Leb}(\mathcal{X}_n)}{\delta_n}\right)^{-1/2}\int_0^{l_{n,2}\delta_n}\log\left(\frac{2\,\mathrm{Leb}(\mathcal{X}_n)l_{n,2}}{\varepsilon}\right)\,\mathrm{d}\varepsilon\\
&=\log\left(\frac{2\,\mathrm{Leb}(\mathcal{X}_n)}{\delta_n}\right)^{-1/2}\left(l_{n,2}\delta_n\log\left(2\,\mathrm{Leb}(\mathcal{X}_n)l_{n,2}\right)+l_{n,2}\delta_n+l_{n,2}\delta_n\log\left(\frac{1}{l_{n,2}\delta_n}\right)\right)\\
&=\log\left(\frac{2\,\mathrm{Leb}(\mathcal{X}_n)}{\delta_n}\right)^{-1/2}l_{n,2}\delta_n\left(1+\log\left(\frac{2\,\mathrm{Leb}(\mathcal{X}_n)}{\delta_n}\right)\right)\lesssim l_{n,2}\delta_n\sqrt{\log\left(\frac{\mathrm{Leb}(\mathcal{X}_n)}{\delta_n}\right)}\lesssim l_{n,2}\delta_n\sqrt{C_1\log n},
\end{aligned}$$

where we used that $\delta_n\le\mathrm{Leb}(\mathcal{X}_n)$. So by Markov's inequality,

$$\mathbb{P}\left(\sup_{|x-x'|\le\delta_n}\left|Z_n(x)-Z_n(x')\right|>t_n\right)\lesssim t_n^{-1}l_{n,2}\delta_n\sqrt{C_1\log n}.$$

Part 4: conclusion

By the results of the previous parts, we have up to universal constants that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{x\in\mathcal{X}_n}\left|G_n(x)-Z_n(x)\right|>t_n\right)\\
&\le\mathbb{P}\left(\sup_{x\in\mathcal{X}_n^\delta}\left|G_n(x)-Z_n(x)\right|>t_n/3\right)+\mathbb{P}\left(\sup_{|x-x'|\le\delta_n}\left|G_n(x)-G_n(x')\right|>t_n/3\right)\\
&\quad+\mathbb{P}\left(\sup_{|x-x'|\le\delta_n}\left|Z_n(x)-Z_n(x')\right|>t_n/3\right)\\
&\lesssim\frac{4\,\mathrm{Leb}(\mathcal{X}_n)}{\delta_n}e^{-s^2/2}+\frac{\sigma_n^2\,\mathrm{Leb}(\mathcal{X}_n)s^2}{\sqrt{n}\delta_n t_n^3}\Big(M_n+6\sigma_n\sqrt{C_1\log n}\Big)+\frac{l_{n,2}\delta_n}{t_n}\sqrt{C_1\log n}+\frac{l_{n,\infty}\delta_n}{t_n\sqrt{n}}C_1\log n.
\end{aligned}$$

Choosing an approximately optimal mesh size of

$$\delta_n = \sqrt{\frac{\sigma_n^2 \operatorname{Leb}(\mathcal{X}_n) \log n}{\sqrt{n} t_n^3} \left(M_n + \sigma_n \sqrt{\log n} \right)} \Bigg/ \sqrt{t_n^{-1} l_{n,2} \sqrt{\log n} \left(1 + \frac{l_{n,\infty} \sqrt{\log n}}{l_{n,2} \sqrt{n}} \right)}$$

satisfies $\log |\delta_n| \lesssim C_1 \log n$ up to a universal constant, so taking s a large enough multiple of $\sqrt{\log n}$ gives

$$\begin{aligned}
&\mathbb{P}\left(\sup_{x \in \mathcal{X}_n} \left| G_n(x) - Z_n(x) \right| > t_n \right) \\
&\lesssim \frac{4 \operatorname{Leb}(\mathcal{X}_n)}{\delta_n} e^{-s^2/2} + \frac{\sigma_n^2 \operatorname{Leb}(\mathcal{X}_n) s^2}{\sqrt{n} \delta_n t_n^3} \left(M_n + 6 \sigma_n \sqrt{C_1 \log n} \right) + \frac{l_{n,2} \delta_n}{t_n} \sqrt{C_1 \log n} + \frac{l_{n,\infty} \delta_n}{t_n \sqrt{n}} C_1 \log n \\
&\lesssim \delta_n \frac{l_{n,2} \sqrt{\log n}}{t_n} \left(1 + \frac{l_{n,\infty} \sqrt{\log n}}{l_{n,2} \sqrt{n}} \right) \\
&\lesssim \frac{\sigma_n \sqrt{\operatorname{Leb}(\mathcal{X}_n)} \sqrt{\log n} \sqrt{M_n + \sigma_n \sqrt{\log n}}}{n^{1/4} t_n^2} \sqrt{l_{n,2} \sqrt{\log n} + \frac{l_{n,\infty}}{\sqrt{n}} \log n}
\end{aligned}$$

up to constants depending only on C_1. □

SA5.2.3 The Vorob'ev–Berkes–Philipp theorem

Proof (Lemma SA28)
The proof is by induction on the number of vertices in the tree. Let $\mathcal{T}$ have n vertices, and suppose that vertex n is a leaf connected to vertex $n-1$ by an edge, relabelling the vertices if necessary. By the induction hypothesis we assume that there is a probability measure $\mathbb{P}^{(n-1)}$ on $\prod_{i=1}^{n-1} \mathcal{X}_i$ whose projections onto $\mathcal{X}_i$ are $\mathbb{P}_i$ and whose projections onto $\mathcal{X}_i \times \mathcal{X}_j$ are $\mathbb{P}_{ij}$, for $i, j \leq n-1$. Now apply the original Vorob'ev–Berkes–Philipp theorem, which can be found as Theorem 1.1.10 in Dudley (1999), to the spaces $\prod_{i=1}^{n-2} \mathcal{X}_i$, $\mathcal{X}_{n-1}$ and $\mathcal{X}_n$; and to the laws $\mathbb{P}^{(n-1)}$ and $\mathbb{P}_{n-1,n}$. This gives a law $\mathbb{P}^{(n)}$ which agrees with $\mathbb{P}_i$ at every vertex by definition, and agrees with $\mathbb{P}_{ij}$ for all $i, j \leq n-1$. It also agrees with $\mathbb{P}_{n-1,n}$, and this is the only edge touching vertex n. Hence $\mathbb{P}^{(n)}$ satisfies the desired properties. □

SA5.3 Main results

We give our main results on consistency, minimax optimality, strong approximation, covariance estimation, feasible inference and counterfactual estimation.

SA5.3.1 Bias

We begin with a basic fact about Lipschitz functions.

Lemma SA43 (Lipschitz kernels are bounded)
Let $\mathcal{X} \subseteq \mathbb{R}$ be a connected set. Let $f : \mathcal{X} \to \mathbb{R}$ satisfy the Lipschitz condition $|f(x) - f(x')| \leq C|x - x'|$ for some $C > 0$ and all $x, x' \in \mathcal{X}$. Suppose also that f is a kernel in the sense that $\int_{\mathcal{X}} f(x) \, \mathrm{d}x = 1$. Then we have

$$\sup_{x \in \mathcal{X}} |f(x)| \leq C \operatorname{Leb}(\mathcal{X}) + \frac{1}{\operatorname{Leb}(\mathcal{X})}.$$

Now let $g : \mathcal{X} \to [0, \infty)$ satisfy $|g(x) - g(x')| \leq C|x - x'|$ for some $C > 0$ and all $x, x' \in \mathcal{X}$. Suppose also that g is a sub-kernel in the sense that $\int_{\mathcal{X}} g(x) \, \mathrm{d}x \leq 1$. Then for any $M \in \big(0, \operatorname{Leb}(\mathcal{X})\big]$, we have

$$\sup_{x \in \mathcal{X}} f(x) \leq CM + \frac{1}{M}.$$

Remark. Applying Lemma SA43 to the density and kernel functions defined in Assumptions SA1 and SA2 yields the following. Firstly, since $k_h(\cdot, w)$ is C_{L}/h^2-Lipschitz on $[w \pm h] \cap \mathcal{W}$ and integrates to one, we have

by the first inequality in Lemma SA43 that

$$|k_h(s,w)| \le \frac{2C_{\mathrm{L}}+1}{h} + \frac{1}{\mathrm{Leb}(\mathcal{W})}.$$

Since each of $f_{W|AA}(\cdot \mid a,a')$, $f_{W|A}(\cdot \mid a)$ and f_W is non-negative and C_{H}-Lipschitz on $\mathcal{W}$ and integrates to at most one over $\mathcal{W}$, taking $M = \frac{1}{\sqrt{C_{\mathrm{H}}}} \wedge \mathrm{Leb}(\mathcal{W})$ in the second inequality in Lemma SA43 gives

$$f_{W|AA}(w \mid a,a') \le 2\sqrt{C_{\mathrm{H}}} + \frac{1}{\mathrm{Leb}(\mathcal{W})}, \quad f_{W|A}(w \mid a) \le 2\sqrt{C_{\mathrm{H}}} + \frac{1}{\mathrm{Leb}(\mathcal{W})}, \quad f_W(w) \le 2\sqrt{C_{\mathrm{H}}} + \frac{1}{\mathrm{Leb}(\mathcal{W})}.$$

Proof (Lemma SA43)
We begin with the first inequality. Note that if $\mathrm{Leb}(\mathcal{X}) = \infty$ there is nothing to prove. Suppose for contradiction that $|f(x)| > C\,\mathrm{Leb}(\mathcal{X}) + \frac{1}{\mathrm{Leb}(\mathcal{X})}$ for some $x \in \mathcal{X}$. If $f(x) \ge 0$ then by the Lipschitz property, for any $y \in \mathcal{X}$,

$$f(y) \ge f(x) - C|y-x| > C\,\mathrm{Leb}(\mathcal{X}) + \frac{1}{\mathrm{Leb}(\mathcal{X})} - C\,\mathrm{Leb}(\mathcal{X}) = \frac{1}{\mathrm{Leb}(\mathcal{X})}.$$

Similarly if $f(x) \le 0$ then

$$f(y) \le f(x) + C|y-x| < -C\,\mathrm{Leb}(\mathcal{X}) - \frac{1}{\mathrm{Leb}(\mathcal{X})} + C\,\mathrm{Leb}(\mathcal{X}) = -\frac{1}{\mathrm{Leb}(\mathcal{X})}.$$

But then either $\int_{\mathcal{X}} f(x)\,\mathrm{d}x > \int_{\mathcal{X}} 1/\mathrm{Leb}(\mathcal{X})\,\mathrm{d}x = 1$ or $\int_{\mathcal{X}} f(x)\,\mathrm{d}x < \int_{\mathcal{X}} -1/\mathrm{Leb}(\mathcal{X})\,\mathrm{d}x = -1 < 1$, giving a contradiction.

For the second inequality, assume that f is non-negative on $\mathcal{X}$ and take $M \in \big(0, \mathrm{Leb}(\mathcal{X})\big]$. Suppose for contradiction that $f(x) > CM + \frac{1}{M}$ for some $x \in \mathcal{X}$. Then again by the Lipschitz property, we have $f(y) \ge 1/M$ for all y such that $|y-x| \le M$. Since $\mathcal{X}$ is connected, we have $\mathrm{Leb}(\mathcal{X} \cap [x \pm M]) \ge M$ and so we deduce that $\int_{\mathcal{X}} f(x)\,\mathrm{d}x > M/M = 1$ which is a contradiction. □

Proof (Lemma SA1)
Begin by defining

$$P_p(s,w) = \sum_{r=0}^{p} \frac{f_W^{(r)}(w)}{r!}(s-w)^r$$

for $s, w \in \mathcal{W}$ as the degree-p Taylor polynomial of f_W, centered at w and evaluated at s. Note that for $p \le \underline{\beta} - 1$, by Taylor's theorem with Lagrange remainder,

$$f_W(s) - P_p(s,w) = \frac{f_W^{(p+1)}(w')}{(p+1)!}(s-w)^{p+1}$$

for some w' between w and s. Also note that for any p,

$$\int_{\mathcal{W}} k_h(s,w)\big(P_p(s,w) - P_{p-1}(s,w)\big)\,\mathrm{d}s = \int_{\mathcal{W}} k_h(s,w)\frac{f_W^{(p)}(w)}{p!}(s-w)^p\,\mathrm{d}s = h^p B_p(w).$$

Further, by the order of the kernel,

$$\begin{aligned}\mathbb{E}\big[\widehat{f}_W(w)\big] - f_W(w) &= \int_{\mathcal{W}} k_h(s,w) f_W(s)\,\mathrm{d}s - f_W(w) = \int_{\mathcal{W}} k_h(s,w)\big(f_W(s) - f_W(w)\big)\,\mathrm{d}s \\ &= \int_{\mathcal{W}} k_h(s,w)\big(f_W(s) - P_{p-1}(s,w)\big)\,\mathrm{d}s.\end{aligned}$$

Part 1: low-order kernel
Suppose that $p \leq \underline{\beta} - 1$. Then

$$\begin{aligned}
\sup_{w\in\mathcal{W}} \left|\mathbb{E}[\widehat{f}_W(w)] - f_W(w) - h^p B_p(w)\right| &= \sup_{w\in\mathcal{W}} \left|\int_{\mathcal{W}} k_h(s,w)\big(f_W(s) - P_{p-1}(s,w)\big)\,\mathrm{d}s - h^p B_p(w)\right| \\
&= \sup_{w\in\mathcal{W}} \left|\int_{\mathcal{W}} k_h(s,w)\big(f_W(s) - P_p(s,w) + P_p(s,w) - P_{p-1}(s,w)\big)\,\mathrm{d}s - h^p B_p(w)\right| \\
&= \sup_{w\in\mathcal{W}} \left|\int_{\mathcal{W}} k_h(s,w)\big(f_W(s) - P_p(s,w)\big)\,\mathrm{d}s\right| = \sup_{w\in\mathcal{W}} \left|\int_{\mathcal{W}} k_h(s,w)\frac{f_W^{(p+1)}(w')}{(p+1)!}(s-w)^{p+1}\,\mathrm{d}s\right| \\
&\leq \sup_{w\in\mathcal{W}} \left|\int_{[w\pm h]} \frac{C_{\mathrm{k}}}{h}\frac{C_{\mathrm{H}}}{(p+1)!}h^{p+1}\,\mathrm{d}s\right| \leq \frac{2C_{\mathrm{k}}C_{\mathrm{H}}}{(p+1)!}h^{p+1}.
\end{aligned}$$

Part 2: order of kernel matches smoothness
Suppose that $p = \underline{\beta}$. Then

$$\begin{aligned}
\sup_{w\in\mathcal{W}} \left|\mathbb{E}[\widehat{f}_W(w)] - f_W(w) - h^p B_p(w)\right| &= \sup_{w\in\mathcal{W}} \left|\int_{\mathcal{W}} k_h(s,w)\big(f_W(s) - P_{\underline{\beta}-1}(s,w)\big)\,\mathrm{d}s - h^p B_p(w)\right| \\
&= \sup_{w\in\mathcal{W}} \left|\int_{\mathcal{W}} k_h(s,w)\big(f_W(s) - P_{\underline{\beta}}(s,w) + P_{\underline{\beta}}(s,w) - P_{\underline{\beta}-1}(s,w)\big)\,\mathrm{d}s - h^{\underline{\beta}} B_{\underline{\beta}}(w)\right| \\
&= \sup_{w\in\mathcal{W}} \left|\int_{\mathcal{W}} k_h(s,w)\big(f_W(s) - P_{\underline{\beta}}(s,w)\big)\,\mathrm{d}s\right| = \sup_{w\in\mathcal{W}} \left|\int_{\mathcal{W}} k_h(s,w)\frac{f_W^{(\underline{\beta})}(w') - f_W^{(\underline{\beta})}(w)}{\underline{\beta}!}(s-w)^{\underline{\beta}}\,\mathrm{d}s\right| \\
&\leq \sup_{w\in\mathcal{W}} \left|\int_{[w\pm h]} \frac{C_{\mathrm{k}}}{h}\frac{C_{\mathrm{H}}h^{\beta-\underline{\beta}}}{\underline{\beta}!}h^{\underline{\beta}}\,\mathrm{d}s\right| \leq \frac{2C_{\mathrm{k}}C_{\mathrm{H}}}{\underline{\beta}!}h^{\beta}.
\end{aligned}$$

Part 3: high-order kernel
Suppose that $p \geq \underline{\beta} + 1$. Then as in the previous part

$$\sup_{w\in\mathcal{W}} \left|\mathbb{E}[\widehat{f}_W(w)] - f_W(w)\right| = \sup_{w\in\mathcal{W}} \left|\int_{[w\pm h]\cap\mathcal{W}} k_h(s,w)\big(f_W(s) - P_{\underline{\beta}}(s,w)\big)\,\mathrm{d}s\right| \leq \frac{2C_{\mathrm{k}}C_{\mathrm{H}}}{\underline{\beta}!}h^{\beta}.$$

□

SA5.3.2 Uniform consistency

Proof (Lemma SA2)
Part 1: Hoeffding-type decomposition
Note that

$$\begin{aligned}
\widehat{f}_W(w) - E_n(w) - \mathbb{E}[\widehat{f}_W(w)] &= \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\Big(\mathbb{E}[k_h(W_{ij},w)\mid A_i,A_j] - \mathbb{E}[k_h(W_{ij},w)]\Big) \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j\neq i}\Big(\mathbb{E}[k_h(W_{ij},w)\mid A_i,A_j] - \mathbb{E}[k_h(W_{ij},w)]\Big)
\end{aligned}$$

and apply Lemma SA35 with

$$u_{ij} = \frac{1}{n(n-1)}\mathbb{E}\big[k_h(W_{ij},w)\mid A_i,A_j\big], \quad u_i = \frac{1}{n(n-1)}\mathbb{E}\big[k_h(W_{ij},w)\mid A_i\big], \quad u = \frac{1}{n(n-1)}\mathbb{E}\big[k_h(W_{ij},w)\big],$$

to see

$$\widehat{f}_W(w) - E_n(w) - \mathbb{E}[\widehat{f}_W(w)] = \frac{2}{n}\sum_{i=1}^{n}\big(u_i - u\big) + \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\big(u_{ij} - u_i - u_j + u\big)$$
$$= \frac{2}{n}\sum_{i=1}^{n} l_i(w) + \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n} q_{ij}(w) = L_n + Q_n.$$

Part 2: Expectation and covariance of L_n, Q_n and E_n
Observe that L_n, Q_n and E_n are clearly mean-zero. For orthogonality, note that their summands have the following properties, for any $1 \leq i < j \leq n$ and $1 \leq r < s \leq n$, and for any $w, w' \in \mathcal{W}$:

$$\mathbb{E}\big[l_i(w)q_{rs}(w')\big] = \mathbb{E}\big[l_i(w)\mathbb{E}\big[q_{rs}(w') \mid A_i\big]\big] = 0,$$
$$\mathbb{E}\big[l_i(w)e_{rs}(w')\big] = \begin{cases}\mathbb{E}\big[l_i(w)\big]\mathbb{E}\big[e_{rs}(w')\big], & \text{if } i \notin \{r,s\}\\ \mathbb{E}\big[l_i(w)\mathbb{E}\big[e_{rs}(w') \mid A_r, A_s\big]\big], & \text{if } i \in \{r,s\}\end{cases}$$
$$= 0,$$
$$\mathbb{E}\big[q_{ij}(w)e_{rs}(w')\big] = \begin{cases}\mathbb{E}\big[q_{ij}(w)\big]\mathbb{E}\big[e_{rs}(w')\big], & \text{if } \{i,j\}\cap\{r,s\} = \emptyset\\ \mathbb{E}\big[\mathbb{E}\big[q_{ij}(w) \mid A_i\big]\mathbb{E}\big[e_{rs}(w') \mid A_i\big]\big], & \text{if } \{i,j\}\cap\{r,s\} = \{i\}\\ \mathbb{E}\big[\mathbb{E}\big[q_{ij}(w) \mid A_j\big]\mathbb{E}\big[e_{rs}(w') \mid A_j\big]\big], & \text{if } \{i,j\}\cap\{r,s\} = \{j\}\\ \mathbb{E}\big[q_{ij}(w)\mathbb{E}\big[e_{rs}(w') \mid A_r, A_s\big]\big], & \text{if } \{i,j\} = \{r,s\}\end{cases}$$
$$= 0,$$

where we used mutual independence of $\mathbf{A}_n$ and $\mathbf{V}_n$ and also $\mathbb{E}[q_{rs}(w) \mid A_i] = 0$ and $\mathbb{E}[e_{ij}(w) \mid A_i, A_j] = 0$. □

Proof (Lemma SA3)
Part 1: total degeneracy
Suppose $D_{\text{lo}} = 0$, so $\operatorname{Var}[f_{W|A}(w \mid A_i)] = 0$ for all $w \in \mathcal{W}$. Therefore for all $w \in \mathcal{W}$, we have $f_{W|A}(w) = f_W(w)$ almost surely. By taking a union over $\mathcal{W} \cap \mathbb{Q}$ and by continuity of $f_{W|A}$ and f_W, this implies that $f_{W|A}(w) = f_W(w)$ for all $w \in \mathcal{W}$ almost surely. Thus

$$\mathbb{E}\left[k_h(W_{ij}, w) \mid A_i\right] = \int_{\mathcal{W}} k_h(s,w) f_{W|A}(s \mid A_i)\,\mathrm{d}s = \int_{\mathcal{W}} k_h(s,w) f_W(s)\,\mathrm{d}s = \mathbb{E}\left[k_h(W_{ij}, w)\right]$$

for all $w \in \mathcal{W}$ almost surely. Hence $l_i(w) = 0$ and therefore $L_n(w) = 0$ for all $w \in \mathcal{W}$ almost surely.

Part 2: no degeneracy
Suppose that $D_{\text{lo}} > 0$. Now since $f_{W|A}(\cdot \mid a)$ is C_{H}-Lipschitz for all $a \in \mathcal{A}$ and since $|k_h| \leq C_{\text{k}}/h$,

$$\sup_{w\in\mathcal{W}}\left|\mathbb{E}[k_h(W_{ij}, w) \mid A_i] - f_{W|A}(w \mid A_i)\right| = \sup_{w\in\mathcal{W}}\left|\int_{\mathcal{W}} k_h(s,w) f_{W|A}(s \mid A_i)\,\mathrm{d}s - f_{W|A}(w \mid A_i)\right|$$
$$= \sup_{w\in\mathcal{W}}\left|\int_{\mathcal{W}\cap[w\pm h]} k_h(s,w)\left(f_{W|A}(s \mid A_i) - f_{W|A}(w \mid A_i)\right)\,\mathrm{d}s\right|$$
$$\leq 2h\frac{C_{\text{k}}}{h}C_{\text{H}}h \leq 2C_{\text{k}}C_{\text{H}}h$$

almost surely. Therefore since $f_{W|A}(w \mid a) \leq C_{\text{d}}$, we have

$$\sup_{w\in\mathcal{W}}\left|\operatorname{Var}\left[\mathbb{E}[k_h(W_{ij}, w) \mid A_i]\right] - \operatorname{Var}\left[f_{W|A}(w \mid A_i)\right]\right| \leq 16C_{\text{k}}C_{\text{H}}C_{\text{d}}h$$

whenever h is small enough that $2C_{\text{k}}C_{\text{H}}h \leq C_{\text{d}}$. Thus

$$\inf_{w\in\mathcal{W}}\operatorname{Var}\left[\mathbb{E}[k_h(W_{ij}, w) \mid A_i]\right] \geq \inf_{w\in\mathcal{W}}\operatorname{Var}[f_{W|A}(w \mid A_i)] - 16C_{\text{k}}C_{\text{H}}C_{\text{d}}h.$$

Therefore if $D_{\text{lo}} > 0$ then eventually $\inf_{w\in\mathcal{W}}\operatorname{Var}\left[\mathbb{E}[k_h(W_{ij}, w) \mid A_i]\right] \geq D_{\text{lo}}/2$. Finally note that

$$\inf_{w\in\mathcal{W}}\operatorname{Var}[L_n(w)] = \frac{4}{n}\inf_{w\in\mathcal{W}}\operatorname{Var}\left[\mathbb{E}[k_h(W_{ij}, w) \mid A_i]\right] \geq \frac{2D_{\text{lo}}}{n}.$$

Part 3: partial degeneracy
Since $f_{W|A}(w \mid A_i)$ is bounded by C_d and C_H-Lipschitz in w, we must have that $\mathrm{Var}[f_{W|A}(w \mid A_i)]$ is a continuous function on $\mathcal{W}$. Thus if $D_\mathrm{lo} = 0$, there must be at least one point $w \in \mathcal{W}$ for which $\mathrm{Var}[f_{W|A}(w \mid A_i)] = 0$ by compactness. Let w be any such degenerate point. Then by the previous part,

$$\mathrm{Var}[L_n(w)] = \frac{4}{n}\,\mathrm{Var}\left[\mathbb{E}[k_h(W_{ij}, w) \mid A_i]\right] \leq 64 C_\mathrm{k} C_\mathrm{H} C_\mathrm{d} \frac{h}{n}.$$

If conversely w is not a degenerate point then $\mathrm{Var}[f_{W|A}(w \mid A_i)] > 0$ so eventually

$$\mathrm{Var}[L_n(w)] = \frac{4}{n}\,\mathrm{Var}\left[\mathbb{E}[k_h(W_{ij}, w) \mid A_i]\right] \geq \frac{2}{n}\,\mathrm{Var}[f_{W|A}(w \mid A_i)].$$

□

Proof (Lemma SA4)
We establish VC-type properties of some function classes and apply results from empirical process theory.

Part 1: establishing VC-type classes
Consider the following function classes:

$$\mathcal{F}_1 = \Big\{ W_{ij} \mapsto k_h(W_{ij}, w) : w \in \mathcal{W} \Big\},$$
$$\mathcal{F}_2 = \Big\{ (A_i, A_j) \mapsto \mathbb{E}\big[k_h(W_{ij}, w) \mid A_i, A_j\big] : w \in \mathcal{W} \Big\},$$
$$\mathcal{F}_3 = \Big\{ A_i \mapsto \mathbb{E}\big[k_h(W_{ij}, w) \mid A_i\big] : w \in \mathcal{W} \Big\}.$$

For $\mathcal{F}_1$, take $0 < \varepsilon \leq \mathrm{Leb}(\mathcal{W})$ and let $\mathcal{W}_\varepsilon$ be an ε-cover of $\mathcal{W}$ of cardinality at most $\mathrm{Leb}(\mathcal{W})/\varepsilon$. Since

$$\sup_{s,w,w' \in \mathcal{W}} \left| \frac{k_h(s,w) - k_h(s,w')}{w - w'} \right| \leq \frac{C_\mathrm{L}}{h^2}$$

almost surely, we see that

$$\sup_{\mathbb{Q}} N\left(\mathcal{F}_1, \rho_\mathbb{Q}, \frac{C_\mathrm{L}}{h^2}\varepsilon\right) \leq N\left(\mathcal{F}_1, \|\cdot\|_\infty, \frac{C_\mathrm{L}}{h^2}\varepsilon\right) \leq \frac{\mathrm{Leb}(\mathcal{W})}{\varepsilon},$$

where $\mathbb{Q}$ ranges over Borel probability measures on $\mathcal{W}$. Since $\frac{C_\mathrm{k}}{h}$ is an envelope function for $\mathcal{F}_1$,

$$\sup_{\mathbb{Q}} N\left(\mathcal{F}_1, \rho_\mathbb{Q}, \frac{C_\mathrm{k}}{h}\varepsilon\right) \leq \frac{C_\mathrm{L}}{C_\mathrm{k}} \frac{\mathrm{Leb}(\mathcal{W})}{h\varepsilon}.$$

Thus for all $\varepsilon \in (0, 1]$,

$$\sup_{\mathbb{Q}} N\left(\mathcal{F}_1, \rho_\mathbb{Q}, \frac{C_\mathrm{k}}{h}\varepsilon\right) \leq \frac{C_\mathrm{L}}{C_\mathrm{k}} \frac{\mathrm{Leb}(\mathcal{W}) \vee 1}{h\varepsilon} \leq (C_1/(h\varepsilon))^{C_2},$$

where $C_1 = \frac{C_\mathrm{L}}{C_\mathrm{k}}(\mathrm{Leb}(\mathcal{W}) \vee 1)$ and $C_2 = 1$. Next, $\mathcal{F}_2$ forms a smoothly parametrized class of functions since for $w, w' \in \mathcal{W}$ we have by the uniform Lipschitz properties of $f_{W|AA}(\cdot \mid A_i, A_j)$ and $k_h(s, \cdot)$, with $|w - w'| \leq h$,

$$\begin{aligned}
&\big|\mathbb{E}\big[k_h(W_{ij}, w) \mid A_i, A_j\big] - \mathbb{E}\big[k_h(W_{ij}, w') \mid A_i, A_j\big]\big| \\
&= \left| \int_{[w \pm h] \cap \mathcal{W}} k_h(s, w) f_{W|AA}(s \mid A_i, A_j)\,\mathrm{d}s - \int_{[w' \pm h] \cap \mathcal{W}} k_h(s, w') f_{W|AA}(s \mid A_i, A_j)\,\mathrm{d}s \right| \\
&= \left| \int_{[w \pm 2h] \cap \mathcal{W}} \big(k_h(s, w) - k_h(s, w')\big) f_{W|AA}(s \mid A_i, A_j)\,\mathrm{d}s \right| \\
&= \left| \int_{[w \pm 2h] \cap \mathcal{W}} \big(k_h(s, w) - k_h(s, w')\big) \big(f_{W|AA}(s \mid A_i, A_j) - f_{W|AA}(w \mid A_i, A_j)\big)\,\mathrm{d}s \right| \\
&\leq 4h \frac{C_\mathrm{L}}{h^2} |w - w'| 2 C_\mathrm{H} h \leq 8 C_\mathrm{L} C_\mathrm{H} |w - w'| \leq C_3 |w - w'|,
\end{aligned}$$

where $C_3 = 8C_{\mathrm{L}}C_{\mathrm{H}}$. The same holds for $|w - w'| > h$ because Lipschitzness is a local property. By taking $\mathbb{E}[\cdot \mid A_i]$, it can be seen by the contraction property of conditional expectation that the same holds for the singly-conditioned terms:

$$\big|\mathbb{E}\big[k_h(W_{ij}, w) \mid A_i\big] - \mathbb{E}\big[k_h(W_{ij}, w') \mid A_i\big]\big| \leq C_3|w - w'|.$$

Therefore $\mathcal{F}_3$ is also smoothly parametrized in exactly the same manner. Let

$$\begin{aligned} C_4 &= \sup_{w\in\mathcal{W}} \operatorname*{ess\,sup}_{A_i,A_j} \big|\mathbb{E}\big[k_h(W_{ij}, w) \mid A_i, A_j\big]\big| = \sup_{w\in\mathcal{W}} \operatorname*{ess\,sup}_{A_i,A_j} \left|\int_{[w\pm h]\cap\mathcal{W}} k_h(s, w) f_{W|AA}(s \mid A_i, A_j)\,\mathrm{d}s\right| \\ &\leq 2h\frac{C_{\mathrm{k}}}{h}C_{\mathrm{d}} \leq 2C_{\mathrm{k}}C_{\mathrm{d}}. \end{aligned}$$

For any $\varepsilon \in (0, 1]$, take an $(\varepsilon C_4/C_3)$-cover of $\mathcal{W}$ of cardinality at most $C_3\,\mathrm{Leb}(\mathcal{W})/(\varepsilon C_4)$. By the above smooth parametrization properties, this cover induces an εC_4-cover for both $\mathcal{F}_2$ and $\mathcal{F}_3$:

$$\begin{aligned} \sup_{\mathbb{Q}} N\big(\mathcal{F}_2, \rho_{\mathbb{Q}}, \varepsilon C_4\big) &\leq N\big(\mathcal{F}_2, \|\cdot\|_\infty, \varepsilon C_4\big) \leq C_3\,\mathrm{Leb}(\mathcal{W})/(\varepsilon C_4), \\ \sup_{\mathbb{Q}} N\big(\mathcal{F}_3, \rho_{\mathbb{Q}}, \varepsilon C_4\big) &\leq N\big(\mathcal{F}_3, \|\cdot\|_\infty, \varepsilon C_4\big) \leq C_3\,\mathrm{Leb}(\mathcal{W})/(\varepsilon C_4). \end{aligned}$$

Hence $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ form VC-type classes with envelopes $F_1 = C_{\mathrm{k}}/h$ and $F_2 = F_3 = C_4$ respectively:

$$\begin{aligned} &\sup_{\mathbb{Q}} N\left(\mathcal{F}_1, \rho_{\mathbb{Q}}, \varepsilon C_{\mathrm{k}}/h\right) \leq (C_1/(h\varepsilon))^{C_2}, && \sup_{\mathbb{Q}} N\big(\mathcal{F}_2, \rho_{\mathbb{Q}}, \varepsilon C_4\big) \leq (C_1/\varepsilon)^{C_2}, \\ &\sup_{\mathbb{Q}} N\big(\mathcal{F}_3, \rho_{\mathbb{Q}}, \varepsilon C_4\big) \leq (C_1/\varepsilon)^{C_2}, \end{aligned}$$

for some constants $C_1 \geq e$ and $C_2 \geq 1$, where we augment the constants if necessary.

Part 2: controlling L_n
Observe that $\sqrt{n}L_n$ is the empirical process of the i.i.d. variables A_i indexed by $\mathcal{F}_3$. We apply Lemma SA25 with $\sigma = C_4$:

$$\mathbb{E}\left[\sup_{w\in\mathcal{W}} \left|\sqrt{n}L_n(w)\right|\right] \lesssim C_4\sqrt{C_2 \log C_1} + \frac{C_4 C_2 \log C_1}{\sqrt{n}} \lesssim 1.$$

By Lemma SA3, the left hand side is zero whenever $D_{\mathrm{up}} = 0$, so we can also write

$$\mathbb{E}\left[\sup_{w\in\mathcal{W}} \left|\sqrt{n}L_n(w)\right|\right] \lesssim D_{\mathrm{up}}.$$

Part 3: controlling Q_n
Observe that nQ_n is the completely degenerate second-order U-process of the i.i.d. variables A_i indexed by $\mathcal{F}_2$. This function class is again uniformly bounded and VC-type, so applying the U-process maximal inequality from Lemma SA37 yields with $\sigma = C_4$

$$\mathbb{E}\left[\sup_{w\in\mathcal{W}} \left|nQ_n(w)\right|\right] \lesssim C_4 C_2 \log C_1 + \frac{C_4(C_2 \log C_1)^2}{\sqrt{n}} \lesssim 1.$$

Part 4: controlling E_n
Conditional on $\mathbf{A}_n$, note that nE_n is the empirical process of the conditionally i.n.i.d. variables W_{ij} indexed by $\mathcal{F}_1$. We apply Lemma SA25 conditionally with

$$\begin{aligned} \sigma^2 &= \sup_{w\in\mathcal{W}} \mathbb{E}\Big[\big(k_h(W_{ij}, w) - \mathbb{E}[k_h(W_{ij}, w) \mid A_i, A_j]\big)^2 \mid A_i, A_j\Big] \leq \sup_{w\in\mathcal{W}} \mathbb{E}\Big[k_h(W_{ij}, w)^2 \mid A_i, A_j\Big] \\ &\leq \sup_{w\in\mathcal{W}} \int_{[w\pm h]\cap\mathcal{W}} k_h(s, w)^2 f_{W|AA}(s \mid A_i, A_j)\,\mathrm{d}s \leq 2h\frac{C_{\mathrm{k}}^2}{h^2} \lesssim 1/h \end{aligned}$$

and noting that we have a sample size of $\frac{1}{2}n(n-1)$, giving

$$\begin{aligned}\mathbb{E}\left[\sup_{w\in\mathcal{W}}\left|nE_n(w)\right|\right] &\lesssim \sigma\sqrt{C_2\log\left((C_1/h)F_1/\sigma\right)} + \frac{F_1C_2\log\left((C_1/h)F_1/\sigma\right)}{n} \\ &\lesssim \frac{1}{\sqrt{h}}\sqrt{C_2\log\left((C_1/h)(C_\mathrm{k}/h)\sqrt{h}\right)} + \frac{(C_\mathrm{k}/h)C_2\log\left((C_1/h)(C_\mathrm{k}/h)\sqrt{h}\right)}{n} \\ &\lesssim \sqrt{\frac{\log 1/h}{h}} + \frac{\log\left(1/h\right)}{nh} \lesssim \sqrt{\frac{\log n}{h}},\end{aligned}$$

where the last line follows by the bandwidth assumption of $\frac{\log n}{n^2h} \to 0$. □

Proof (Theorem SA1)
This follows from Lemma SA1 and Lemma SA4. □

SA5.3.3 Minimax optimality

Before proving Theorem SA2 we first give a lower bound result for parametric point estimation in Lemma SA44.

Lemma SA44 (A Neyman–Pearson result for Bernoulli random variables)
Recall that the Bernoulli distribution $\mathrm{Ber}(\theta)$ *places mass* θ *at* 1 *and mass* $1-\theta$ *at* 0. *Define* $\mathbb{P}_\theta^n$ *as the law of* $(A_1, A_2, \ldots, A_n, V)$, *where* $A_1, \ldots, A_n$ *are i.i.d.* $\mathrm{Ber}(\theta)$, *and* V *is an* $\mathbb{R}^d$*-valued random variable for some* $d \geq 1$ *which is independent of the* A *variables and with a fixed distribution that does not depend on* θ. *Let* $\theta_0 = \frac{1}{2}$ *and* $\theta_{1,n} = \frac{1}{2} + \frac{1}{\sqrt{8n}}$. *Then for any estimator* $\widetilde{\theta}_n$ *which is a function of* $(A_1, A_2, \ldots, A_n, V)$ *only,*

$$\mathbb{P}_{\theta_0}^n\left(\left|\widetilde{\theta}_n - \theta_0\right| \geq \frac{1}{\sqrt{32n}}\right) + \mathbb{P}_{\theta_{1,n}}^n\left(\left|\widetilde{\theta}_n - \theta_{1,n}\right| \geq \frac{1}{\sqrt{32n}}\right) \geq \frac{1}{2}.$$

Proof (Lemma SA44)
Let $f : \{0,1\}^n \to \{0,1\}$ be any function. Considering this function as a statistical test, the Neyman–Pearson lemma and Pinsker's inequality (Giné and Nickl, 2021) give

$$\begin{aligned}\mathbb{P}_{\theta_0}^n\left(f=1\right) + \mathbb{P}_{\theta_{1,n}}^n\left(f=0\right) &\geq 1 - \mathrm{TV}\left(\mathbb{P}_{\theta_0}^n, \mathbb{P}_{\theta_{1,n}}^n\right) \geq 1 - \sqrt{\frac{1}{2}\mathrm{KL}\left(\mathbb{P}_{\theta_0}^n \,\middle\|\, \mathbb{P}_{\theta_{1,n}}^n\right)} \\ &= 1 - \sqrt{\frac{n}{2}\mathrm{KL}\left(\mathrm{Ber}(\theta_0) \,\middle\|\, \mathrm{Ber}(\theta_{1,n})\right) + \frac{n}{2}\mathrm{KL}\left(V \,\middle\|\, V\right)} \\ &= 1 - \sqrt{\frac{n}{2}\mathrm{KL}\left(\mathrm{Ber}(\theta_0) \,\middle\|\, \mathrm{Ber}(\theta_{1,n})\right)},\end{aligned}$$

where TV is the total variation distance and KL is the Kullback–Leibler divergence. In the penultimate line we used the tensorization of Kullback–Leibler divergence (Giné and Nickl, 2021), noting that the law of V is fixed and hence does not contribute. We now evaluate this Kullback–Leibler divergence at the specified parameter values.

$$\begin{aligned}\mathbb{P}_{\theta_0}^n\left(f=1\right) + \mathbb{P}_{\theta_{1,n}}^n\left(f=0\right) &\geq 1 - \sqrt{\frac{n}{2}\mathrm{KL}\left(\mathrm{Ber}(\theta_0) \,\middle\|\, \mathrm{Ber}(\theta_{1,n})\right)} \\ &= 1 - \sqrt{\frac{n}{2}}\sqrt{\theta_0\log\frac{\theta_0}{\theta_{1,n}} + (1-\theta_0)\log\frac{1-\theta_0}{1-\theta_{1,n}}} \\ &= 1 - \sqrt{\frac{n}{2}}\sqrt{\frac{1}{2}\log\frac{1/2}{1/2+1/\sqrt{8n}} + \frac{1}{2}\log\frac{1/2}{1/2-1/\sqrt{8n}}} \\ &= 1 - \frac{\sqrt{n}}{2}\sqrt{\log\frac{1}{1-1/(2n)}} \geq 1 - \frac{\sqrt{n}}{2}\sqrt{\frac{1}{n}} = \frac{1}{2},\end{aligned}$$

where in the penultimate line we used that $\log \frac{1}{1-x} \leq 2x$ for $x \in [0, 1/2]$. Now define a test f by $f = 1$ if $\widetilde{\theta}_n > \frac{1}{2} + \frac{1}{\sqrt{32n}}$ and $f = 0$ otherwise, to see

$$\mathbb{P}^n_{\theta_0}\left(\widetilde{\theta}_n > \frac{1}{2} + \frac{1}{\sqrt{32n}}\right) + \mathbb{P}^n_{\theta_{1,n}}\left(\widetilde{\theta}_n \leq \frac{1}{2} + \frac{1}{\sqrt{32n}}\right) \geq \frac{1}{2}.$$

By the triangle inequality, recalling that $\theta_0 = \frac{1}{2}$ and $\theta_{1,n} = \frac{1}{2} + \frac{1}{\sqrt{8n}}$, we have the event inclusions

$$\left\{\widetilde{\theta}_n > \frac{1}{2} + \frac{1}{\sqrt{32n}}\right\} \subseteq \left\{\left|\widetilde{\theta}_n - \theta_0\right| \geq \frac{1}{\sqrt{32n}}\right\}$$
$$\left\{\widetilde{\theta}_n \leq \frac{1}{2} + \frac{1}{\sqrt{32n}}\right\} \subseteq \left\{\left|\widetilde{\theta}_n - \theta_{1,n}\right| \geq \frac{1}{\sqrt{32n}}\right\}.$$

Thus by the monotonicity of measures,

$$\mathbb{P}^n_{\theta_0}\left(\left|\widetilde{\theta}_n - \theta_0\right| \geq \frac{1}{\sqrt{32n}}\right) + \mathbb{P}^n_{\theta_{1,n}}\left(\left|\widetilde{\theta}_n - \theta_{1,n}\right| \geq \frac{1}{\sqrt{32n}}\right) \geq \frac{1}{2}.$$

□

Proof (Theorem SA2)

Part 1: lower bound for $\mathcal{P}$

By translation and scaling of the data, we may assume without loss of generality that $\mathcal{W} = [-1, 1]$. We may also assume that $C_{\mathrm{H}} \leq 1/2$, since reducing C_{H} can only shrink the class of distributions. Define the dyadic distribution $\mathbb{P}_\theta$ with parameter $\theta \in [1/2, 1]$ as follows: $A_1, \ldots, A_n$ are i.i.d. $\mathrm{Ber}(\theta)$, while V_{ij} for $1 \leq i < j \leq n$ are i.i.d. and independent of $\mathbf{A}_n$. The distribution of V_{ij} is given by its density function $f_V(v) = \frac{1}{2} + C_{\mathrm{H}} v$ on $[-1, 1]$. Finally generate $W_{ij} = W(A_i, A_j, V_{ij}) := (2A_iA_j - 1)V_{ij}$. Note that the function W does not depend on θ. The conditional and marginal densities of W_{ij} are for $w \in [-1, 1]$

$$f_{W|AA}(w \mid A_i, A_j) = \begin{cases} \frac{1}{2} + C_{\mathrm{H}} w & \text{if } A_i = A_j = 1 \\ \frac{1}{2} - C_{\mathrm{H}} w & \text{if } A_i = 0 \text{ or } A_j = 0, \end{cases}$$
$$f_{W|A}(w \mid A_i) = \begin{cases} \frac{1}{2} + (2\theta - 1) C_{\mathrm{H}} w & \text{if } A_i = 1 \\ \frac{1}{2} - C_{\mathrm{H}} w & \text{if } A_i = 0, \end{cases}$$
$$f_W(w) = \frac{1}{2} + (2\theta^2 - 1) C_{\mathrm{H}} w.$$

Clearly $f_W \in \mathcal{H}^\beta_{C_{\mathrm{H}}}(\mathcal{W})$ and $f_{W|AA}(\cdot \mid a, a') \in \mathcal{H}^1_{C_{\mathrm{H}}}(\mathcal{W})$. Also $\sup_{w \in \mathcal{W}} \|f_{W|A}(w \mid \cdot)\|_{\mathrm{TV}} \leq 1$. Therefore $\mathbb{P}_\theta$ satisfies Assumption SA1 and so $\left\{\mathbb{P}_\theta : \theta \in [1/2, 1]\right\} \subseteq \mathcal{P}$.

Note that $f_W(1) = \frac{1}{2} + (2\theta^2 - 1)C_{\mathrm{H}}$, so $\theta^2 = \frac{1}{2C_{\mathrm{H}}}(f_W(1) - 1/2 + C_{\mathrm{H}})$. Thus if $\widetilde{f}_W$ is some density estimator depending only on the data $\mathbf{W}_n$, we can naturally define the non-negative parameter point estimator

$$\widetilde{\theta}^2_n := \frac{1}{2C_{\mathrm{H}}}\left(\widetilde{f}_W(1) - \frac{1}{2} + C_{\mathrm{H}}\right) \vee 0.$$

This gives the inequality

$$\left|\widetilde{\theta}^2_n - \theta^2\right| = \left|\frac{1}{2C_{\mathrm{H}}}\left(\widetilde{f}_W(1) - \frac{1}{2} + C_{\mathrm{H}}\right) \vee 0 - \frac{1}{2C_{\mathrm{H}}}\left(f_W(1) - \frac{1}{2} + C_{\mathrm{H}}\right)\right| \leq \frac{1}{2C_{\mathrm{H}}} \sup_{w \in \mathcal{W}} \left|\widetilde{f}_W(w) - f_W(w)\right|.$$

Therefore since also $\widetilde{\theta} \geq 0$ and $\theta \geq \frac{1}{2}$,

$$\left|\widetilde{\theta}_n - \theta\right| = \frac{\left|\widetilde{\theta}^2_n - \theta^2\right|}{\widetilde{\theta}_n + \theta} \leq \frac{1}{C_{\mathrm{H}}} \sup_{w \in \mathcal{W}} \left|\widetilde{f}_W(w) - f_W(w)\right|.$$

Now we apply the point estimation lower bound from Lemma SA44, setting $\theta_0 = \frac{1}{2}$ and $\theta_{1,n} = \frac{1}{2} + \frac{1}{\sqrt{8n}}$, noting that the estimator $\widetilde{\theta}_n$ is a function of $\mathbf{W}_n$ only, thus is a function of $\mathbf{A}_n$ and $\mathbf{V}_n$ only and so satisfies the conditions.

$$\begin{aligned}
&\mathbb{P}_{\theta_0}\left(\sup_{w\in\mathcal{W}}\left|\widetilde{f}_W(w) - f_W^{(0)}(w)\right| \geq \frac{1}{C\sqrt{n}}\right) + \mathbb{P}_{\theta_{1,n}}\left(\sup_{w\in\mathcal{W}}\left|\widetilde{f}_W(w) - f_W^{(1)}(w)\right| \geq \frac{1}{C\sqrt{n}}\right) \\
&\quad \geq \mathbb{P}_{\theta_0}\left(\left|\widetilde{\theta}_n - \theta_0\right| \geq \frac{1}{CC_{\mathrm{H}}\sqrt{n}}\right) + \mathbb{P}_{\theta_{1,n}}\left(\left|\widetilde{\theta}_n - \theta_{1,n}\right| \geq \frac{1}{CC_{\mathrm{H}}\sqrt{n}}\right) \\
&\quad \geq \mathbb{P}_{\theta_0}\left(\left|\widetilde{\theta}_n - \theta_0\right| \geq \frac{1}{\sqrt{32n}}\right) + \mathbb{P}_{\theta_{1,n}}\left(\left|\widetilde{\theta}_n - \theta_{1,n}\right| \geq \frac{1}{\sqrt{32n}}\right) \geq \frac{1}{2},
\end{aligned}$$

where we set $C \geq \frac{\sqrt{32}}{C_{\mathrm{H}}}$. Therefore we deduce that

$$\inf_{\widetilde{f}_W}\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widetilde{f}_W(w) - f_W(w)\right| \geq \frac{1}{C\sqrt{n}}\right) \geq \frac{1}{4}$$

and so

$$\inf_{\widetilde{f}_W}\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\left[\sup_{w\in\mathcal{W}}\left|\widetilde{f}_W(w) - f_W(w)\right|\right] \geq \frac{1}{4C\sqrt{n}}.$$

Part 2: lower bound for $\mathcal{P}_{\mathrm{d}}$

For the subclass of totally degenerate distributions, we rely on the main theorem from Khasminskii (1978). Let $\mathcal{P}_0$ be the subclass of $\mathcal{P}_{\mathrm{d}}$ consisting of the distributions which satisfy $A_1 = \cdots = A_n = 0$ and $W_{ij} := A_i + A_j + V_{ij} = V_{ij}$, so that W_{ij} are i.i.d. with common density $f_W = f_V$. Define the class

$$\mathcal{F} = \left\{f \text{ density function on } \mathbb{R},\ f \in \mathcal{H}^{\beta}_{C_{\mathrm{H}}}(\mathcal{W})\right\}.$$

Write $\mathbb{E}_f$ for the expectation under W_{ij} having density f. Then by the main theorem in Khasminskii (1978),

$$\liminf_{n\to\infty}\inf_{\widetilde{f}_W}\sup_{f\in\mathcal{F}}\mathbb{E}_f\left[\left(\frac{n^2}{\log n}\right)^{\frac{\beta}{2\beta+1}}\sup_{w\in\mathcal{W}}\left|\widetilde{f}_W(w) - f_W(w)\right|\right] > 0,$$

where $\widetilde{f}_W$ is any density estimator depending only on the $\frac{1}{2}n(n-1)$ i.i.d. data samples $\mathbf{W}_n$. Now every density function in $\mathcal{H}^{\beta}_{C_{\mathrm{H}}}(\mathcal{W})$ corresponds to a distribution in $\mathcal{P}_0$ and therefore to a distribution in $\mathcal{P}_{\mathrm{d}}$. Thus for large enough n and some positive constant C,

$$\inf_{\widetilde{f}_W}\sup_{\mathbb{P}\in\mathcal{P}_{\mathrm{d}}}\mathbb{E}_{\mathbb{P}}\left[\sup_{w\in\mathcal{W}}\left|\widetilde{f}_W(w) - f_W(w)\right|\right] \geq \frac{1}{C}\left(\frac{\log n}{n^2}\right)^{\frac{\beta}{2\beta+1}}.$$

Part 3: upper bounds

The corresponding upper bounds follow by using a dyadic kernel density estimator $\widehat{f}_W$ with a boundary bias-corrected Lipschitz kernel of order $p \geq \beta$ and using a bandwidth of h. Firstly Lemma SA1 gives

$$\sup_{\mathbb{P}\in\mathcal{P}}\sup_{w\in\mathcal{W}}\left|\mathbb{E}_{\mathbb{P}}\left[\widehat{f}_W(w)\right] - f_W(w)\right| \leq \frac{4C_{\mathrm{k}}C_{\mathrm{H}}}{\beta!}h^{\beta}.$$

Then, treating the degenerate and non-degenerate cases separately and noting that all inequalities hold uniformly over $\mathcal{P}$ and $\mathcal{P}_{\mathrm{d}}$, the proof of Lemma SA4 shows that

$$\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\left[\sup_{w\in\mathcal{W}}\left|\widehat{f}_W(w) - \mathbb{E}_{\mathbb{P}}[\widehat{f}_W(w)]\right|\right] &\lesssim \frac{1}{\sqrt{n}} + \sqrt{\frac{\log n}{n^2 h}}, \\
\sup_{\mathbb{P}\in\mathcal{P}_{\mathrm{d}}}\mathbb{E}_{\mathbb{P}}\left[\sup_{w\in\mathcal{W}}\left|\widehat{f}_W(w) - \mathbb{E}_{\mathbb{P}}[\widehat{f}_W(w)]\right|\right] &\lesssim \sqrt{\frac{\log n}{n^2 h}}.
\end{aligned}$$

Thus combining these yields that

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\sup_{w\in\mathcal{W}} \left|\widehat{f}_W(w) - f_W(w)\right|\right] \lesssim h^\beta + \frac{1}{\sqrt{n}} + \sqrt{\frac{\log n}{n^2 h}},$$
$$\sup_{\mathbb{P}\in\mathcal{P}_{\mathrm{d}}} \mathbb{E}_{\mathbb{P}}\left[\sup_{w\in\mathcal{W}} \left|\widehat{f}_W(w) - f_W(w)\right|\right] \lesssim h^\beta + \sqrt{\frac{\log n}{n^2 h}}.$$

Set $h = \left(\frac{\log n}{n^2}\right)^{\frac{1}{2\beta+1}}$ and note that $\beta \geq 1$ implies that $\left(\frac{\log n}{n^2}\right)^{\frac{\beta}{2\beta+1}} \ll \frac{1}{\sqrt{n}}$. Therefore for some constant $C > 0$,

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\sup_{w\in\mathcal{W}} \left|\widehat{f}_W(w) - f_W(w)\right|\right] \lesssim \frac{1}{\sqrt{n}} + \left(\frac{\log n}{n^2}\right)^{\frac{\beta}{2\beta+1}} \leq \frac{C}{\sqrt{n}},$$
$$\sup_{\mathbb{P}\in\mathcal{P}_{\mathrm{d}}} \mathbb{E}_{\mathbb{P}}\left[\sup_{w\in\mathcal{W}} \left|\widehat{f}_W(w) - f_W(w)\right|\right] \leq C\left(\frac{\log n}{n^2}\right)^{\frac{\beta}{2\beta+1}}.$$

□

SA5.3.4 Covariance structure

Proof (Lemma SA5)
Throughout this proof we will write k_{ij} for $k_h(W_{ij}, w)$ and k'_{ij} for $k_h(W_{ij}, w')$, in the interest of brevity.

$$\begin{aligned}
\Sigma_n(w,w') &= \mathbb{E}\Big[\big(\widehat{f}_W(w) - \mathbb{E}[\widehat{f}_W(w)]\big)\big(\widehat{f}_W(w') - \mathbb{E}[\widehat{f}_W(w')]\big)\Big] \\
&= \mathbb{E}\left[\left(\frac{2}{n(n-1)}\sum_{i<j}\big(k_{ij} - \mathbb{E}k_{ij}\big)\right)\left(\frac{2}{n(n-1)}\sum_{r<s}\big(k'_{rs} - \mathbb{E}k'_{rs}\big)\right)\right] \\
&= \frac{4}{n^2(n-1)^2}\sum_{i<j}\sum_{r<s}\mathbb{E}\left[\big(k_{ij} - \mathbb{E}k_{ij}\big)\big(k'_{rs} - \mathbb{E}k'_{rs}\big)\right] = \frac{4}{n^2(n-1)^2}\sum_{i<j}\sum_{r<s}\mathrm{Cov}\left[k_{ij}, k'_{rs}\right].
\end{aligned}$$

Note first that for i, j, r, s all distinct, k_{ij} is independent of k'_{rs} and so the covariance is zero. By a counting argument, it can be seen that there are $n(n-1)/2$ summands where $|\{i,j,r,s\}| = 2$, and $n(n-1)(n-2)$ summands where $|\{i,j,r,s\}| = 3$. Therefore since the samples are identically distributed, the value of the summands depends only on the number of distinct indices and we have the decomposition

$$\begin{aligned}
\Sigma_n(w,w') &= \frac{4}{n^2(n-1)^2}\left(\frac{n(n-1)}{2}\mathrm{Cov}[k_{ij}, k'_{ij}] + n(n-1)(n-2)\,\mathrm{Cov}[k_{ij}, k'_{ir}]\right) \\
&= \frac{2}{n(n-1)}\mathrm{Cov}[k_{ij}, k'_{ij}] + \frac{4(n-2)}{n(n-1)}\mathrm{Cov}[k_{ij}, k'_{ir}],
\end{aligned}$$

giving the first representation. To obtain the second representation, note that since W_{ij} and W_{ir} are independent conditional on A_i,

$$\begin{aligned}
\mathrm{Cov}\left[k_{ij}k'_{ir}\right] &= \mathbb{E}\left[k_{ij}k'_{ir}\right] - \mathbb{E}[k_{ij}]\mathbb{E}[k'_{ir}] = \mathbb{E}\left[\mathbb{E}\left[k_{ij}k'_{ir} \mid A_i\right]\right] - \mathbb{E}[k_{ij}]\mathbb{E}[k'_{ir}] \\
&= \mathbb{E}\left[\mathbb{E}[k_{ij} \mid A_i]\mathbb{E}[k'_{ir} \mid A_i]\right] - \mathbb{E}[k_{ij}]\mathbb{E}[k'_{ir}] = \mathrm{Cov}\left[\mathbb{E}[k_{ij} \mid A_i], \mathbb{E}[k'_{ir} \mid A_i]\right].
\end{aligned}$$

□

Proof (Lemma SA6)
By Lemma SA5, the diagonal elements of Σ_n are

$$\Sigma_n(w,w) = \frac{2}{n(n-1)}\mathrm{Var}\left[k_h(W_{ij}, w)\right] + \frac{4(n-2)}{n(n-1)}\mathrm{Var}\left[\mathbb{E}[k_h(W_{ij}, w) \mid A_i]\right].$$

We bound each of the two terms separately. Firstly, note that since k_h is bounded by C_{k}/h,

$$\operatorname{Var}\big[k_h(W_{ij},w)\big] \le \mathbb{E}\big[k_h(W_{ij},w)^2\big] = \int_{\mathcal{W}\cap[w\pm h]} k_h(s,w)^2 f_W(s)\,\mathrm{d}s \le 2C_{\mathrm{k}}^2/h.$$

Conversely, since $\big|\mathbb{E}[k_h(W_{ij},w)]\big| = \big|\int_{[w\pm h]\cap\mathcal{W}} k_h(s,w)f_W(s)\,\mathrm{d}s\big| \le 2C_{\mathrm{k}}C_{\mathrm{d}}$, Jensen's integral inequality shows

$$\begin{aligned}
\operatorname{Var}\big[k_h(W_{ij},w)\big] &\ge \int_{\mathcal{W}\cap[w\pm h]} k_h(s,w)^2 f_W(s)\,\mathrm{d}s - 4C_{\mathrm{k}}^2C_{\mathrm{d}}^2 \\
&\ge \inf_{w\in\mathcal{W}} f_W(w)\frac{1}{2h}\left(\int_{\mathcal{W}\cap[w\pm h]} k_h(s,w)\,\mathrm{d}s\right)^2 - 4C_{\mathrm{k}}^2C_{\mathrm{d}}^2 \\
&\ge \frac{1}{2h}\inf_{w\in\mathcal{W}} f_W(w) - 4C_{\mathrm{k}}^2C_{\mathrm{d}}^2 \ge \frac{1}{4h}\inf_{w\in\mathcal{W}} f_W(w)
\end{aligned}$$

for small enough h, noting that this is trivially true if the infimum is zero. For the other term, we have

$$\operatorname{Var}\big[\mathbb{E}[k_h(W_{ij},w)\mid A_i]\big] \le \operatorname{Var}\big[f_{W|A}(w\mid A_i)\big] + 16C_{\mathrm{H}}C_{\mathrm{k}}C_{\mathrm{d}}h \le 2D_{\mathrm{up}}^2$$

for small enough h, by a result from the proof of Lemma SA3. Also

$$\operatorname{Var}\big[\mathbb{E}[k_h(W_{ij},w)\mid A_i]\big] \ge \operatorname{Var}\big[f_{W|A}(w\mid A_i)\big] - 16C_{\mathrm{H}}C_{\mathrm{k}}C_{\mathrm{d}}h \ge \frac{D_{\mathrm{lo}}^2}{2}$$

for small enough h. Combining these four inequalities yields that for all large enough n,

$$\begin{aligned}
\frac{2}{n(n-1)}\frac{1}{4h}\inf_{w\in\mathcal{W}} f_W(w) + \frac{4(n-2)}{n(n-1)}\frac{D_{\mathrm{lo}}^2}{2} &\le \inf_{w\in\mathcal{W}}\Sigma_n(w,w) \\
\le \sup_{w\in\mathcal{W}}\Sigma_n(w,w) &\le \frac{2}{n(n-1)}\frac{2C_{\mathrm{k}}^2}{h} + \frac{4(n-2)}{n(n-1)}2D_{\mathrm{up}}^2,
\end{aligned}$$

so that

$$\frac{D_{\mathrm{lo}}^2}{n} + \frac{1}{n^2h}\inf_{w\in\mathcal{W}} f_W(w) \lesssim \inf_{w\in\mathcal{W}}\Sigma_n(w,w) \le \sup_{w\in\mathcal{W}}\Sigma_n(w,w) \lesssim \frac{D_{\mathrm{up}}^2}{n} + \frac{1}{n^2h}.$$

□

SA5.3.5 Strong approximation

Proof (Lemma SA7)
To obtain the strong approximation, we apply the KMT corollary from Lemma SA26. Define the functions

$$k_h^A(a,w) = 2\mathbb{E}[k_h(W_{ij},w)\mid A_i = a],$$

which are of bounded variation in a uniformly over w since

$$\begin{aligned}
\sup_{w\in\mathcal{W}}\|k_h^A(\cdot,w)\|_{\mathrm{TV}} &= 2\sup_{w\in\mathcal{W}}\sup_{m\in\mathbb{N}}\sup_{a_0\le\cdots\le a_m}\sum_{i=1}^m \big|k_h^A(a_i,w) - k_h^A(a_{i-1},w)\big| \\
&= 2\sup_{w\in\mathcal{W}}\sup_{m\in\mathbb{N}}\sup_{a_0\le\cdots\le a_m}\sum_{i=1}^m \left|\int_{[w\pm h]\cap\mathcal{W}} k_h(s,w)\big(f_{W|A}(s\mid a_i) - f_{W|A}(s\mid a_{i-1})\big)\,\mathrm{d}s\right| \\
&\le 2\sup_{w\in\mathcal{W}}\int_{[w\pm h]\cap\mathcal{W}} |k_h(s,w)|\sup_{m\in\mathbb{N}}\sup_{a_0\le\cdots\le a_m}\sum_{i=1}^m \big|f_{W|A}(s\mid a_i) - f_{W|A}(s\mid a_{i-1})\big|\,\mathrm{d}s \\
&\le 2\sup_{w\in\mathcal{W}}\int_{[w\pm h]\cap\mathcal{W}} |k_h(s,w)|\,\big\|f_{W|A}(w\mid\cdot)\big\|_{\mathrm{TV}}\,\mathrm{d}s \\
&\le 4C_{\mathrm{k}}\sup_{w\in\mathcal{W}}\big\|f_{W|A}(w\mid\cdot)\big\|_{\mathrm{TV}} \lesssim D_{\mathrm{up}},
\end{aligned}$$

where the last line is by observing that the total variation is zero whenever $D_{\text{up}} = 0$. Hence by Lemma SA26 there exist (on some probability space) n independent copies of A_i, denoted A_i' and a centered Gaussian process $Z_n^{L'}$ such that if we define

$$L_n'(w) = \frac{1}{n}\sum_{i=1}^{n}\big(k_h^A(A_i', w) - \mathbb{E}[k_h^A(A_i', w)]\big),$$

then for some positive constants C_1, C_2, C_3, by defining the processes as zero outside $\mathcal{W}$ we have

$$\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\sqrt{n}L_n'(w) - Z_n^{L'}(w)\right| > D_{\text{up}}\frac{t + C_1\log n}{\sqrt{n}}\right) \le C_2 e^{-C_3 t}.$$

Integrating tail probabilities shows that

$$\mathbb{E}\left[\sup_{w\in\mathcal{W}}\left|\sqrt{n}L_n'(w) - Z_n^{L'}(w)\right|\right] \le D_{\text{up}}\frac{C_1\log n}{\sqrt{n}} + \int_0^\infty \frac{D_{\text{up}}}{\sqrt{n}} C_2 e^{-C_3 t}\,\mathrm{d}t \lesssim \frac{D_{\text{up}}\log n}{\sqrt{n}}.$$

Further, $Z_n^{L'}$ has the same covariance structure as $G_n^{L'}$ in the sense that for all $w, w' \in \mathcal{W}$,

$$\mathbb{E}\big[Z_n^{L'}(w) Z_n^{L'}(w')\big] = \mathbb{E}\big[G_n^{L'}(w) G_n^{L'}(w')\big],$$

and clearly L_n' is equal in distribution to L_n. To obtain the trajectory regularity property of $Z_n^{L'}$, note that it was shown in the proof of Lemma SA4 that for all $w, w' \in \mathcal{W}$,

$$\big|k_h^A(A_i, w) - k_h^A(A_i, w')\big| \le C|w - w'|$$

for some constant $C > 0$. Therefore since the A_i are i.i.d.,

$$\begin{aligned}
\mathbb{E}\left[\left|Z_n^{L'}(w) - Z_n^{L'}(w')\right|^2\right]^{1/2} &= \sqrt{n}\,\mathbb{E}\left[\left|L_n(w) - L_n(w')\right|^2\right]^{1/2} \\
&= \sqrt{n}\,\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\Big(k_h^A(A_i, w) - k_h^A(A_i, w') - \mathbb{E}\big[k_h^A(A_i, w)\big] + \mathbb{E}\big[k_h^A(A_i, w')\big]\Big)\right|^2\right]^{1/2} \\
&= \mathbb{E}\left[\left|k_h^A(A_i, w) - k_h^A(A_i, w') - \mathbb{E}\big[k_h^A(A_i, w)\big] + \mathbb{E}\big[k_h^A(A_i, w')\big]\right|^2\right]^{1/2} \lesssim |w - w'|.
\end{aligned}$$

Therefore by the regularity result for Gaussian processes in Lemma SA32, with $\delta_n \in (0, 1/2]$:

$$\mathbb{E}\left[\sup_{|w-w'|\le\delta_n}\left|Z_n^{L'}(w) - Z_n^{L'}(w')\right|\right] \lesssim \int_0^{\delta_n}\sqrt{\log 1/\varepsilon}\,\mathrm{d}\varepsilon \lesssim \delta_n\sqrt{\log 1/\delta_n} \lesssim D_{\text{up}}\delta_n\sqrt{\log 1/\delta_n},$$

where the last inequality is because $Z_n^{L'} \equiv 0$ whenever $D_{\text{up}} = 0$. There is a modification of $Z_n^{L'}$ with continuous trajectories by Kolmogorov's continuity criterion (Le Gall, 2016, Theorem 2.9). Note that L_n' is $\mathbf{A}_n'$-measurable and so by Lemma SA26 we can assume that $Z_n^{L'}$ depends only on $\mathbf{A}_n'$ and some random noise which is independent of $(\mathbf{A}_n', \mathbf{V}_n')$. Finally in order to have $\mathbf{A}_n', \mathbf{V}_n', L_n'$ and $Z_n^{L'}$ all defined on the same probability space, we note that $\mathbf{A}_n$ and $\mathbf{V}_n$ are random vectors while L_n' and $Z_n^{L'}$ are stochastic processes with continuous sample paths indexed on the compact interval $\mathcal{W}$. Hence the Vorob'ev–Berkes–Philipp theorem (Lemma SA28) allows us to "glue" them together in the desired way on another new probability space, giving $\big(\mathbf{A}_n', \mathbf{V}_n', L_n', Z_n^{L'}\big)$, where we retain the single prime notation for clarity. □

Proof (Lemma SA8)
We apply Lemma SA27 conditional on $\mathbf{A}_n$. While this lemma is not in its current form stated for conditional distributions, the Yurinskii coupling on which it depends can be readily extended by following the proof of Belloni et al. (2019, Lemma 38), using a conditional version of Strassen's theorem (Chen and Kato, 2020, Theorem B.2). Care must similarly be taken in embedding the conditionally Gaussian vectors into a conditionally Gaussian process, using the Vorob'ev–Berkes–Philipp theorem (Lemma SA28).

By the mutual independence of A_i and V_{ij}, we have that the observations W_{ij} are independent (but not necessarily identically distributed) conditionally on $\mathbf{A}_n$. Note that $\sup_{s,w\in\mathcal{W}}|k_h(s,w)| \lesssim M_n = h^{-1}$ and $\mathbb{E}[k_h(W_{ij},w)^2 \mid \mathbf{A}_n] \lesssim \sigma_n^2 = h^{-1}$. The following uniform Lipschitz condition holds with $l_{n,\infty} = C_{\mathrm{L}}h^{-2}$, by the Lipschitz property of the kernels:

$$\sup_{s,w,w'\in\mathcal{W}}\left|\frac{k_h(s,w)-k_h(s,w')}{w-w'}\right| \le l_{n,\infty}.$$

Also, the following L^2 Lipschitz condition holds uniformly with $l_{n,2} = 2C_{\mathrm{L}}\sqrt{C_{\mathrm{d}}}h^{-3/2}$:

$$\begin{aligned}
&\mathbb{E}\big[\big|k_h(W_{ij},w)-k_h(W_{ij},w')\big|^2 \mid \mathbf{A}_n\big]^{1/2}\\
&\quad\le \frac{C_{\mathrm{L}}}{h^2}|w-w'|\left(\int_{([w\pm h]\cup[w'\pm h])\cap\mathcal{W}} f_{W|AA}(s\mid\mathbf{A}_n)\,\mathrm{d}s\right)^{1/2}\\
&\quad\le \frac{C_{\mathrm{L}}}{h^2}|w-w'|\sqrt{4hC_{\mathrm{d}}} \le l_{n,2}|w-w'|.
\end{aligned}$$

So we can apply Lemma SA27 conditionally on $\mathbf{A}_n$ to the $\frac{1}{2}n(n-1)$ observations, noting that

$$\sqrt{n^2h}E_n(w) = \sqrt{\frac{2nh}{n-1}}\sqrt{\frac{2}{n(n-1)}}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\Big(k_h(W_{ij},w)-\mathbb{E}[k_h(W_{ij},w)\mid A_i,A_j]\Big),$$

to deduce that for $t_n>0$ there exist (an enlarged probability space) conditionally mean-zero and conditionally Gaussian processes $\widetilde{Z}_n^{E\prime}(w)$ with the same conditional covariance structure as $\sqrt{n^2h}E_n(w)$ and satisfying

$$\begin{aligned}
&\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\sqrt{n^2h}E_n(w)-\widetilde{Z}_n^{E\prime}(w)\right|>t_n \;\middle|\; \mathbf{A}_n'\right)\\
&\quad=\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\sqrt{\frac{n(n-1)}{2}}E_n(w)-\sqrt{\frac{n-1}{2nh}}\widetilde{Z}_n^{E\prime}(w)\right|>\sqrt{\frac{n-1}{2nh}}t_n \;\middle|\; \mathbf{A}_n'\right)\\
&\quad\lesssim \frac{\sigma_n\sqrt{\mathrm{Leb}(\mathcal{W})}\sqrt{\log n}\sqrt{M_n+\sigma_n\sqrt{\log n}}}{n^{1/2}t_n^2/h}\sqrt{l_{n,2}\sqrt{\log n}+\frac{l_{n,\infty}}{n}\log n}\\
&\quad\lesssim \frac{h^{-1/2}\sqrt{\log n}\sqrt{h^{-1}+h^{-1/2}\sqrt{\log n}}}{n^{1/2}t_n^2/h}\sqrt{h^{-3/2}\sqrt{\log n}+\frac{h^{-2}}{n}\log n}\\
&\quad\lesssim \sqrt{\frac{\log n}{n}}\frac{\sqrt{1+\sqrt{h\log n}}}{t_n^2}\sqrt{\sqrt{\frac{\log n}{h^3}}\left(1+\sqrt{\frac{\log n}{n^2h}}\right)}\\
&\quad\lesssim \sqrt{\frac{\log n}{n}}\frac{1}{t_n^2}\left(\frac{\log n}{h^3}\right)^{1/4} \lesssim t_n^{-2}n^{-1/2}h^{-3/4}(\log n)^{3/4},
\end{aligned}$$

where we used $h\lesssim 1/\log n$ and $\frac{\log n}{n^2h}\lesssim 1$. To obtain the trajectory regularity property of $\widetilde{Z}_n^{E\prime}$, note that for $w,w'\in\mathcal{W}$, by conditional independence,

$$\begin{aligned}
&\mathbb{E}\left[\left|\widetilde{Z}_n^{E\prime}(w)-\widetilde{Z}_n^{E\prime}(w')\right|^2 \mid \mathbf{A}_n'\right]^{1/2}\\
&\quad=\sqrt{n^2h}\,\mathbb{E}\left[\left|E_n(w)-E_n(w')\right|^2 \mid \mathbf{A}_n\right]^{1/2}\\
&\quad\lesssim \sqrt{n^2h}\,\mathbb{E}\left[\left|\frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\Big(k_h(W_{ij},w)-k_h(W_{ij},w')\Big)\right|^2 \;\middle|\; \mathbf{A}_n\right]^{1/2}\\
&\quad\lesssim \sqrt{h}\,\mathbb{E}\left[\left|k_h(W_{ij},w)-k_h(W_{ij},w')\right|^2 \mid \mathbf{A}_n\right]^{1/2} \lesssim h^{-1}|w-w'|.
\end{aligned}$$

Therefore by the regularity result for Gaussian processes in Lemma SA32, with $\delta_n \in (0, 1/(2h)]$:

$$\mathbb{E}\left[\sup_{|w-w'|\le\delta_n} \left|\widetilde{Z}_n^{E\prime}(w) - \widetilde{Z}_n^{E\prime}(w')\right| \mid \mathbf{A}_n'\right] \lesssim \int_0^{\delta_n/h} \sqrt{\log(\varepsilon^{-1}h^{-1})}\,\mathrm{d}\varepsilon \lesssim \frac{\delta_n}{h}\sqrt{\log\frac{1}{h\delta_n}},$$

and there exists a modification with continuous trajectories. Finally in order to have $\mathbf{A}_n', \mathbf{V}_n', E_n'$ and $\widetilde{Z}_n^{E\prime}$ all defined on the same probability space, we note that $\mathbf{A}_n$ and $\mathbf{V}_n$ are random vectors while E_n' and $\widetilde{Z}_n^{E\prime}$ are stochastic processes with continuous sample paths indexed on the compact interval $\mathcal{W}$. Hence the Vorob'ev–Berkes–Philipp theorem (Lemma SA28) allows us to "glue together" $(\mathbf{A}_n, \mathbf{V}_n, E_n)$ and $(E_n', \widetilde{Z}_n^{E\prime})$ in the desired way on another new probability space, giving $(\mathbf{A}_n', \mathbf{V}_n', E_n', \widetilde{Z}_n^{E\prime})$, where we retain the single prime notation for clarity.

Note that the trajectories of the conditionally Gaussian processes $\widetilde{Z}_n^{E\prime}$ depend on the choice of t_n, necessitating the use of a divergent sequence R_n to establish bounds in probability on the coupling error. □

Proof (Lemma SA9)

Part 1: defining $Z_n^{E\prime\prime}$

Pick $\delta_n \to 0$ with $\log 1/\delta_n \lesssim \log n$. Let $\mathcal{W}_\delta$ be a δ_n-covering of $\mathcal{W}$ with cardinality $\mathrm{Leb}(\mathcal{W})/\delta_n$ which is also a δ_n-packing. Let $\widetilde{Z}_{n,\delta}^{E\prime}$ be the restriction of $\widetilde{Z}_n^{E\prime}$ to $\mathcal{W}_\delta$. Let $\widetilde{\Sigma}_n^E(w,w') = \mathbb{E}\big[\widetilde{Z}_n^{E\prime}(w)\widetilde{Z}_n^{E\prime}(w') \mid \mathbf{A}_n'\big]$ be the conditional covariance function of $\widetilde{Z}_n^{E\prime}$, and define $\Sigma_n^E(w,w') = \mathbb{E}\big[\widetilde{\Sigma}_n^E(w,w')\big]$. Let $\widetilde{\Sigma}_{n,\delta}^E$ and $\Sigma_{n,\delta}^E$ be the restriction matrices of $\widetilde{\Sigma}_n^E$ and Σ_n^E respectively to $\mathcal{W}_\delta \times \mathcal{W}_\delta$, noting that, as (conditional) covariance matrices, these are (almost surely) positive semi-definite.

Let $N \sim \mathcal{N}(0, I_{|\mathcal{W}_\delta|})$ be independent of $\mathbf{A}_n'$, and define using the matrix square root $\widetilde{Z}_{n,\delta}^{E\prime\prime} = \big(\widetilde{\Sigma}_{n,\delta}^E\big)^{1/2}N$, which has the same distribution as $\widetilde{Z}_{n,\delta}^{E\prime}$, conditional on $\mathbf{A}_n'$. Extend it using the Vorob'ev–Berkes–Philipp theorem (Lemma SA28) to the compact interval $\mathcal{W}$, giving a conditionally Gaussian process $\widetilde{Z}_n^{E\prime\prime}$ which has the same distribution as $\widetilde{Z}_n^{E\prime}$, conditional on $\mathbf{A}_n'$. Define $Z_{n,\delta}^{E\prime\prime} = \big(\Sigma_{n,\delta}^E\big)^{1/2}N$, noting that this is independent of $\mathbf{A}_n'$, and extend it using the Vorob'ev–Berkes–Philipp theorem (Lemma SA28) to a Gaussian process $Z_n^{E\prime\prime}$ on the compact interval $\mathcal{W}$, which is independent of $\mathbf{A}_n'$ and has covariance structure given by Σ_n^E.

Part 2: closeness of $Z_n^{E\prime\prime}$ and $\widetilde{Z}_n^{E\prime\prime}$ on the mesh

Note that conditionally on $\mathbf{A}_n'$, $\widetilde{Z}_{n,\delta}^{E\prime\prime} - Z_{n,\delta}^{E\prime\prime}$ is a length-$|\mathcal{W}_\delta|$ Gaussian random vector with covariance matrix $\big(\big(\widetilde{\Sigma}_{n,\delta}^E\big)^{1/2} - \big(\Sigma_{n,\delta}^E\big)^{1/2}\big)^2$. So by the Gaussian maximal inequality in Lemma SA31 applied conditionally on $\mathbf{A}_n'$, we have

$$\mathbb{E}\left[\max_{w\in\mathcal{W}_\delta}\left|\widetilde{Z}_n^{E\prime\prime}(w) - Z_n^{E\prime\prime}(w)\right| \,\middle|\, \mathbf{A}_n'\right] \lesssim \sqrt{\log n}\left\|\widetilde{\Sigma}_{n,\delta}^E - \Sigma_{n,\delta}^E\right\|_2^{1/2},$$

since $\log|\mathcal{W}_\delta| \lesssim \log n$. Next, we apply some U-statistic theory to $\widetilde{\Sigma}_{n,\delta}^E - \Sigma_{n,\delta}^E$, with the aim of applying the matrix concentration result for second-order U-statistics presented in Lemma SA38. Firstly we note that since the conditional covariance structures of $\widetilde{Z}_n^{E\prime}$ and $\sqrt{n^2h}E_n$ are equal in distribution, we have, writing $E_n(\mathcal{W}_\delta)$ for the vector $\big(E_n(w) : w \in \mathcal{W}_\delta\big)$ and similarly for $k_h(W_{ij}, \mathcal{W}_\delta)$,

$$\begin{aligned}
\widetilde{\Sigma}_{n,\delta}^E &= n^2h\mathbb{E}[E_n(\mathcal{W}_\delta)E_n(\mathcal{W}_\delta)^\mathsf{T} \mid \mathbf{A}_n] \\
&= n^2h\frac{4}{n^2(n-1)^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \\
&\qquad \mathbb{E}\left[\Big(k_h(W_{ij},\mathcal{W}_\delta) - \mathbb{E}\left[k_h(W_{ij},\mathcal{W}_\delta) \mid \mathbf{A}_n\right]\Big)\Big(k_h(W_{ij},\mathcal{W}_\delta) - \mathbb{E}\left[k_h(W_{ij},\mathcal{W}_\delta) \mid \mathbf{A}_n\right]\Big)^\mathsf{T} \mid \mathbf{A}_n\right] \\
&= \frac{4h}{(n-1)^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} u(A_i, A_j),
\end{aligned}$$

where we define the random $|\mathcal{W}_\delta| \times |\mathcal{W}_\delta|$ matrices

$$u(A_i, A_j) = \mathbb{E}\left[k_h(W_{ij},\mathcal{W}_\delta)k_h(W_{ij},\mathcal{W}_\delta)^\mathsf{T} \mid \mathbf{A}_n\right] - \mathbb{E}\left[k_h(W_{ij},\mathcal{W}_\delta) \mid \mathbf{A}_n\right]\mathbb{E}\left[k_h(W_{ij},\mathcal{W}_\delta) \mid \mathbf{A}_n\right]^\mathsf{T}.$$

Let $u(A_i) = \mathbb{E}[u(A_i, A_j) \mid A_i]$ and $u = \mathbb{E}[u(A_i, A_j)]$. The following Hoeffding decomposition holds, by Lemma SA35:

$$\widetilde{\Sigma}^E_{n,\delta} - \Sigma^E_{n,\delta} = \widetilde{L} + \widetilde{Q},$$

where

$$\widetilde{L} = \frac{4h}{n-1} \sum_{i=1}^{n} \big(u(A_i) - u\big),$$
$$\widetilde{Q} = \frac{4h}{(n-1)^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \big(u(A_i, A_j) - u(A_i) - u(A_j) + u\big).$$

Next, we seek an almost sure upper bound on $\|u(A_i, A_j)\|_2$. Since this is a symmetric matrix, we have by Hölder's inequality

$$\|u(A_i, A_j)\|_2 \le \|u(A_i, A_j)\|_1^{1/2} \|u(A_i, A_j)\|_\infty^{1/2} = \max_{1 \le k \le |\mathcal{W}_\delta|} \sum_{l=1}^{|\mathcal{W}_\delta|} |u(A_i, A_j)_{kl}|.$$

The terms on the right hand side can be bounded as follows, writing w, w' for the kth and lth points in $\mathcal{W}_\delta$ respectively:

$$\begin{aligned}
|u(A_i, A_j)_{kl}| &= \big|\mathbb{E}\left[k_h(W_{ij}, w) k_h(W_{ij}, w') \mid \mathbf{A}_n\right] - \mathbb{E}\left[k_h(W_{ij}, w) \mid \mathbf{A}_n\right] \mathbb{E}\left[k_h(W_{ij}, w') \mid \mathbf{A}_n\right] \big| \\
&\lesssim \mathbb{E}\left[|k_h(W_{ij}, w) k_h(W_{ij}, w')| \mid \mathbf{A}_n\right] + \mathbb{E}\left[|k_h(W_{ij}, w)| \mid \mathbf{A}_n\right] \mathbb{E}\left[|k_h(W_{ij}, w')| \mid \mathbf{A}_n\right] \\
&\lesssim h^{-1} \mathbb{I}\big\{|w - w'| \le 2h\big\} + 1 \lesssim h^{-1} \mathbb{I}\big\{|k - l| \le 2h/\delta_n\big\} + 1,
\end{aligned}$$

where we used that $|w - w'| \ge |k - l| \delta_n$ because $\mathcal{W}_\delta$ is a δ_n-packing. Hence

$$\begin{aligned}
\|u(A_i, A_j)\|_2 &\le \max_{1 \le k \le |\mathcal{W}_\delta|} \sum_{l=1}^{|\mathcal{W}_\delta|} |u(A_i, A_j)_{kl}| \lesssim \max_{1 \le k \le |\mathcal{W}_\delta|} \sum_{l=1}^{|\mathcal{W}_\delta|} \Big(h^{-1} \mathbb{I}\big\{|k - l| \le 2h/\delta_n\big\} + 1\Big) \\
&\lesssim 1/\delta_n + 1/h + |\mathcal{W}_\delta| \lesssim 1/\delta_n + 1/h.
\end{aligned}$$

Clearly the same bound holds for $\|u(A_i)\|_2$ and $\|u\|_2$, by Jensen's inequality. Therefore applying the matrix Bernstein inequality (Lemma SA30) to the zero-mean matrix $\widetilde{L}$ gives

$$\mathbb{E}\left[\left\|\widetilde{L}\right\|_2\right] \lesssim \frac{h}{n} \left(\frac{1}{\delta_n} + \frac{1}{h}\right) \left(\log |\mathcal{W}_\delta| + \sqrt{n \log |\mathcal{W}_\delta|}\right) \lesssim \left(\frac{h}{\delta_n} + 1\right) \sqrt{\frac{\log n}{n}}.$$

Applying the matrix U-statistic concentration inequality (Lemma SA38) to the zero-mean matrix $\widetilde{Q}$ gives

$$\mathbb{E}\left[\left\|\widetilde{Q}\right\|_2\right] \lesssim \frac{h}{n^2} n \left(\frac{1}{\delta_n} + \frac{1}{h}\right) (\log |\mathcal{W}_\delta|)^{3/2} \lesssim \left(\frac{h}{\delta_n} + 1\right) \frac{(\log n)^{3/2}}{n}.$$

Hence putting everything together, taking a marginal expectation and applying Jensen's inequality,

$$\begin{aligned}
\mathbb{E}\left[\max_{w \in \mathcal{W}_\delta} \left|\widetilde{Z}_n^{E\prime\prime}(w) - Z_n^{E\prime\prime}(w)\right|\right] &\lesssim \sqrt{\log n}\, \mathbb{E}\left[\left\|\widetilde{\Sigma}^E_{n,\delta} - \Sigma^E_{n,\delta}\right\|_2^{1/2}\right] \lesssim \sqrt{\log n}\, \mathbb{E}\left[\left\|\widetilde{\Sigma}^E_{n,\delta} - \Sigma^E_{n,\delta}\right\|_2\right]^{1/2} \\
&\lesssim \sqrt{\log n}\, \mathbb{E}\left[\left\|\widetilde{L} + \widetilde{Q}\right\|_2\right]^{1/2} \lesssim \sqrt{\log n}\, \mathbb{E}\left[\left\|\widetilde{L}\right\|_2 + \left\|\widetilde{Q}\right\|_2\right]^{1/2} \\
&\lesssim \sqrt{\log n} \left(\left(\frac{h}{\delta_n} + 1\right) \sqrt{\frac{\log n}{n}} + \left(\frac{h}{\delta_n} + 1\right) \frac{(\log n)^{3/2}}{n}\right)^{1/2} \\
&\lesssim \sqrt{\frac{h}{\delta_n} + 1} \frac{(\log n)^{3/4}}{n^{1/4}}.
\end{aligned}$$

Part 3: regularity of Z_n^E and $\widetilde{Z}_n^{E\prime}$
Define the semimetrics

$$\rho(w,w')^2 = \mathbb{E}\left[\left|Z_n^{E\prime\prime}(w) - Z_n^{E\prime\prime}(w')\right|^2\right], \qquad \widetilde{\rho}(w,w')^2 = \mathbb{E}\left[\left|\widetilde{Z}_n^{E\prime\prime}(w) - \widetilde{Z}_n^{E\prime\prime}(w')\right|^2 \mid \mathbf{A}_n\right].$$

We can bound $\widetilde{\rho}$ as follows, since $\widetilde{Z}_n^{E\prime\prime}$ and $\sqrt{n^2h}E_n$ have the same conditional covariance structure:

$$\widetilde{\rho}(w,w') = \mathbb{E}\left[\left|\widetilde{Z}_n^{E\prime\prime}(w) - \widetilde{Z}_n^{E\prime\prime}(w')\right|^2 \mid \mathbf{A}_n'\right]^{1/2} = \sqrt{n^2h}\,\mathbb{E}\left[\left|E_n(w) - E_n(w')\right|^2 \mid \mathbf{A}_n'\right]^{1/2} \lesssim h^{-1}|w-w'|,$$

uniformly in $\mathbf{A}_n'$, where the last line was shown in the proof of Lemma SA8. Note that also

$$\rho(w,w') = \sqrt{\mathbb{E}[\widetilde{\rho}(w,w')^2]} \lesssim h^{-1}|w-w'|.$$

Thus Lemma SA32 applies directly to Z_n^E and conditionally to $\widetilde{Z}_n^{E\prime}$, with $\delta_n \in (0, 1/(2h)]$, yielding

$$\mathbb{E}\left[\sup_{|w-w'|\le\delta_n}\left|\widetilde{Z}_n^{E\prime\prime}(w) - \widetilde{Z}_n^{E\prime\prime}(w')\right| \mid \mathbf{A}_n'\right] \lesssim \int_0^{\delta_n/h}\sqrt{\log(1/(\varepsilon h))}\,\mathrm{d}\varepsilon \lesssim \frac{\delta_n}{h}\sqrt{\log\frac{1}{h\delta_n}},$$
$$\mathbb{E}\left[\sup_{|w-w'|\le\delta_n}\left|Z_n^{E\prime\prime}(w) - Z_n^{E\prime\prime}(w')\right|\right] \lesssim \int_0^{\delta_n/h}\sqrt{\log(1/(\varepsilon h))}\,\mathrm{d}\varepsilon \lesssim \frac{\delta_n}{h}\sqrt{\log\frac{1}{h\delta_n}}.$$

Continuity of trajectories follows from this.

Part 4: conclusion
We use the previous parts to deduce that

$$\begin{aligned}\mathbb{E}\left[\sup_{w\in\mathcal{W}}\left|\widetilde{Z}_n^{E\prime\prime}(w) - Z_n^{E\prime\prime}(w)\right|\right] &\lesssim \mathbb{E}\left[\max_{w\in\mathcal{W}_\delta}\left|\widetilde{Z}_n^{E\prime\prime}(w) - Z_n^{E\prime\prime}(w)\right|\right] + \mathbb{E}\left[\sup_{|w-w'|\le\delta_n}\left|\widetilde{Z}_n^{E\prime\prime}(w) - \widetilde{Z}_n^{E\prime\prime}(w')\right|\right] \\ &\quad + \mathbb{E}\left[\sup_{|w-w'|\le\delta_n}\left|Z_n^{E\prime\prime}(w) - Z_n^{E\prime\prime}(w')\right|\right] \\ &\lesssim \sqrt{\frac{h}{\delta_n}+1}\frac{(\log n)^{3/4}}{n^{1/4}} + \frac{\delta_n\sqrt{\log n}}{h}.\end{aligned}$$

Setting $\delta_n = h\left(\frac{\log n}{n}\right)^{1/6}$ gives

$$\mathbb{E}\left[\sup_{w\in\mathcal{W}}\left|\widetilde{Z}_n^{E\prime\prime}(w) - Z_n^{E\prime\prime}(w)\right|\right] \lesssim n^{-1/6}(\log n)^{2/3}.$$

Finally, independence of $Z_n^{E\prime\prime}$ and $\mathbf{A}_n''$ follows by another application of the Vorob'ev–Berkes–Philipp theorem from Lemma SA28, this time conditionally on $\mathbf{A}_n'$, to the random variables $\left(\mathbf{A}_n', \widetilde{Z}_n^{E\prime}\right)$ and $\left(\widetilde{Z}_n^{E\prime\prime}, Z_n^{E\prime\prime}\right)$. □

Proof (Theorem SA3)
We add together the strong approximations for the L_n and E_n terms, and then add an independent Gaussian process to account for the variance of Q_n.

Part 1: gluing together the strong approximations
Let $\left(\mathbf{A}_n', \mathbf{V}_n', L_n', Z_n^{L\prime}\right)$ be the strong approximation for L_n from Lemma SA7. Likewise let $\left(\mathbf{A}_n'', \mathbf{V}_n'', E_n'', \widetilde{Z}_n^{E\prime\prime}\right)$ and $\left(\mathbf{A}_n''', \mathbf{V}_n''', \widetilde{Z}_n^{E\prime\prime\prime}, Z_n^{E\prime\prime\prime}\right)$ be the conditional and unconditional strong approximations for E_n given in Lemmas SA8 and SA9 respectively. The first step is to define copies of all of these variables and processes on the same probability space. This is achieved by applying the Vorob'ev–Berkes–Philipp Theorem (Lemma SA28). In particular, dropping the prime notation for clarity, we construct $\left(\mathbf{A}_n, \mathbf{V}_n, L_n, Z_n^L, E_n, \widetilde{Z}_n^E, Z_n^E\right)$ with the following properties:

(i) $\sup_{w\in\mathcal{W}}\left|\sqrt{n}L_n(w) - Z_n^L(w)\right| \lesssim_{\mathbb{P}} n^{-1/2}\log n$,

(ii) $\sup_{w\in\mathcal{W}}\left|\sqrt{n^2h}E_n(w)-\widetilde{Z}_n^E(w)\right|\lesssim_{\mathbb{P}} n^{-1/4}h^{-3/8}(\log n)^{3/8}R_n$,

(iii) $\sup_{w\in\mathcal{W}}\left|\widetilde{Z}_n^E(w)-Z_n^E(w)\right|\lesssim_{\mathbb{P}} n^{-1/6}(\log n)^{2/3}$,

(iv) Z_n^L is independent of Z_n^E.

Note that the independence of Z_n^L and Z_n^E follows since Z_n^L depends only on $\mathbf{A}_n$ and some independent random noise, while Z_n^E is independent of $\mathbf{A}_n$. Therefore (Z_n^L, Z_n^E) are jointly Gaussian. To obtain the strong approximation result for $\widehat{f}_W$, define the mean-zero Gaussian process

$$Z_n^f(w)=\frac{1}{\sqrt{n}}Z_n^L(w)+\frac{1}{n}Z_n^Q(w)+\frac{1}{\sqrt{n^2h}}Z_n^E(w),$$

where $Z_n^Q(w)$ is a mean-zero Gaussian process independent of everything else and with covariance

$$\mathbb{E}\big[Z_n^Q(w)Z_n^Q(w')\big]=n^2\mathbb{E}\big[Q_n(w)Q_n(w')\big].$$

As shown in the proof of Lemma SA4, the process $Q_n(w)$ is uniformly Lipschitz and uniformly bounded in w. Thus by Lemma SA32, we have $\mathbb{E}\big[\sup_{w\in\mathcal{W}}|Z_n^Q(w)|\big]\lesssim 1$. Therefore the uniform approximation error is given by

$$\begin{aligned}
&\sup_{w\in\mathcal{W}}\left|\widehat{f}_W(w)-\mathbb{E}[\widehat{f}_W(w)]-Z_n^f(w)\right| \\
&\quad=\sup_{w\in\mathcal{W}}\left|\frac{1}{\sqrt{n}}Z_n^L(w)+\frac{1}{n}Z_n^Q(w)+\frac{1}{\sqrt{n^2h}}Z_n^E(w)-\Big(L_n(w)+Q_n(w)+E_n(w)\Big)\right| \\
&\quad\le\sup_{w\in\mathcal{W}}\left(\frac{1}{\sqrt{n}}\left|Z_n^L(w)-\sqrt{n}L_n(w)\right|+\frac{1}{\sqrt{n^2h}}\left|\widetilde{Z}_n^E(w)-\sqrt{n^2h}E_n(w)\right|+\frac{1}{\sqrt{n^2h}}\left|Z_n^E(w)-\widetilde{Z}_n^E(w)\right|\right. \\
&\qquad\left.+\left|Q_n(w)\right|+\frac{1}{n}\left|Z_n^Q(w)\right|\right) \\
&\quad\lesssim_{\mathbb{P}} n^{-1}\log n+n^{-5/4}h^{-7/8}(\log n)^{3/8}R_n+n^{-7/6}h^{-1/2}(\log n)^{2/3}.
\end{aligned}$$

Part 2: covariance structure
Since L_n, Q_n and E_n are mutually orthogonal in L^2 (as shown in Lemma SA2), we have the following covariance structure:

$$\begin{aligned}
\mathbb{E}\big[Z_n^f(w)Z_n^f(w')\big]&=\frac{1}{n}\mathbb{E}\big[Z_n^L(w)Z_n^L(w')\big]+\frac{1}{n^2}\mathbb{E}\big[Z_n^Q(w)Z_n^Q(w')\big]+\frac{1}{n^2h}\mathbb{E}\big[Z_n^E(w)Z_n^E(w')\big] \\
&=\mathbb{E}\big[L_n(w)L_n(w')\big]+\mathbb{E}\big[Q_n(w)Q_n(w')\big]+\mathbb{E}\big[E_n(w)E_n(w')\big] \\
&=\mathbb{E}\big[\big(\widehat{f}_W(w)-\mathbb{E}[\widehat{f}_W(w)]\big)\big(\widehat{f}_W(w')-\mathbb{E}[\widehat{f}_W(w')]\big)\big].
\end{aligned}$$

Part 3: trajectory regularity
The trajectory regularity of the process Z_n^f follows directly by adding the regularities of the processes $\frac{1}{\sqrt{n}}Z_n^L$ and $\frac{1}{n}Z_n^Q$ and $\frac{1}{\sqrt{n^2h}}Z_n^E$. Similarly, Z_n^f has continuous trajectories. □

SA5.3.6 Infeasible uniform confidence bands

Proof (Lemma SA10)
Note that

$$\left|T_n(w)-Z_n^T(w)\right|=\frac{\left|\widehat{f}_W(w)-f_W(w)-Z_n^f(w)\right|}{\sqrt{\Sigma_n(w,w)}}.$$

By Theorem SA3 and Lemma SA1, the numerator can be bounded above by

$$\begin{aligned}
&\sup_{w\in\mathcal{W}}\left|\widehat{f}_W(w)-f_W(w)-Z_n^f(w)\right|\le\sup_{w\in\mathcal{W}}\left|\widehat{f}_W(w)-\mathbb{E}\big[\widehat{f}_W(w)\big]-Z_n^f(w)\right|+\sup_{w\in\mathcal{W}}\left|\mathbb{E}\big[\widehat{f}_W(w)\big]-f_W(w)\right| \\
&\quad\lesssim_{\mathbb{P}} n^{-1}\log n+n^{-5/4}h^{-7/8}(\log n)^{3/8}R_n+n^{-7/6}h^{-1/2}(\log n)^{2/3}+h^{p\wedge\beta}.
\end{aligned}$$

By Lemma SA6 with $\inf_{\mathcal{W}} f_W(w) > 0$, the denominator is bounded below by

$$\inf_{w\in\mathcal{W}} \sqrt{\Sigma_n(w,w)} \gtrsim \frac{D_{\text{lo}}}{\sqrt{n}} + \frac{1}{\sqrt{n^2 h}},$$

and the result follows. □

Proof (Theorem SA4)
Note that the covariance structure of Z_n^T is given by

$$\text{Cov}\left[Z_n^T(w), Z_n^T(w')\right] = \frac{\Sigma_n(w,w')}{\sqrt{\Sigma_n(w,w)\Sigma_n(w',w')}}.$$

We apply an anti-concentration result to establish that all quantiles of the random variable $\sup_{w\in\mathcal{W}} \left|Z_n^T(w)\right|$ exist. To do this, we must first establish regularity properties of Z_n^T.

Part 1: L^2 regularity of Z_n^T
Writing k'_{ij} for $k_h(W_{ij}, w')$ etc., note that by Lemma SA5,

$$\begin{aligned}
&\left|\Sigma_n(w,w') - \Sigma_n(w,w'')\right| \\
&= \left|\frac{2}{n(n-1)}\text{Cov}\left[k_{ij}, k'_{ij}\right] + \frac{4(n-2)}{n(n-1)}\text{Cov}\left[k_{ij}, k'_{ir}\right] - \frac{2}{n(n-1)}\text{Cov}\left[k_{ij}, k''_{ij}\right] - \frac{4(n-2)}{n(n-1)}\text{Cov}\left[k_{ij}, k''_{ir}\right]\right| \\
&\leq \frac{2}{n(n-1)}\left|\text{Cov}\left[k_{ij}, k'_{ij} - k''_{ij}\right]\right| + \frac{4(n-2)}{n(n-1)}\left|\text{Cov}\left[k_{ij}, k'_{ir} - k''_{ir}\right]\right| \\
&\leq \frac{2}{n(n-1)}\|k_{ij}\|_\infty \|k'_{ij} - k''_{ij}\|_\infty + \frac{4(n-2)}{n(n-1)}\|k_{ij}\|_\infty \|k'_{ir} - k''_{ir}\|_\infty \\
&\leq \frac{4}{nh^3} C_{\text{k}} C_{\text{L}} |w' - w''| \lesssim n^{-1}h^{-3}|w' - w''|
\end{aligned}$$

uniformly in $w, w', w'' \in \mathcal{W}$. Therefore by Lemma SA6, with $\delta_n \leq n^{-2}h^2$, we have

$$\begin{aligned}
\inf_{|w-w'|\leq\delta_n} \Sigma_n(w,w') &\gtrsim \frac{D_{\text{lo}}^2}{n} + \frac{1}{n^2 h} - n^{-1}h^{-3}\delta_n \gtrsim \frac{D_{\text{lo}}^2}{n} + \frac{1}{n^2 h} - \frac{1}{n^3 h} \gtrsim \frac{D_{\text{lo}}^2}{n} + \frac{1}{n^2 h}, \\
\sup_{|w-w'|\leq\delta_n} \Sigma_n(w,w') &\lesssim \frac{D_{\text{up}}^2}{n} + \frac{1}{n^2 h} + n^{-1}h^{-3}\delta_n \lesssim \frac{D_{\text{up}}^2}{n} + \frac{1}{n^2 h} + \frac{1}{n^3 h} \lesssim \frac{D_{\text{up}}^2}{n} + \frac{1}{n^2 h}.
\end{aligned}$$

The L^2 regularity of Z_n^T is

$$\mathbb{E}\left[\left(Z_n^T(w) - Z_n^T(w')\right)^2\right] = 2 - 2\frac{\Sigma_n(w,w')}{\sqrt{\Sigma_n(w,w)\Sigma_n(w',w')}}.$$

Applying the elementary result that for $a, b, c > 0$,

$$1 - \frac{a}{\sqrt{bc}} = \frac{b(c-a) + a(b-a)}{\sqrt{bc}\left(\sqrt{bc} + a\right)},$$

with

$$a = \Sigma_n(w,w'), \qquad b = \Sigma_n(w,w), \qquad c = \Sigma_n(w',w')$$

and noting that $|c-a| \lesssim n^{-1}h^{-3}|w-w'|$ and $|b-a| \lesssim n^{-1}h^{-3}|w-w'|$ and $\frac{D_{\text{lo}}^2}{n} + \frac{1}{n^2 h} \lesssim a, b, c \lesssim \frac{D_{\text{up}}^2}{n} + \frac{1}{n^2 h}$ yields

$$\mathbb{E}\left[\left(Z_n^T(w) - Z_n^T(w')\right)^2\right] \lesssim \frac{(D_{\text{up}}^2/n + 1/(n^2 h))n^{-1}h^{-3}|w-w'|}{(D_{\text{lo}}^2/n + 1/(n^2 h))^2} \lesssim \frac{n^2 h^{-4}|w-w'|}{n^{-4}h^{-2}} \lesssim n^2 h^{-2}|w-w'|.$$

Thus the semimetric induced by Z_n^T on $\mathcal{W}$ is

$$\rho(w,w') := \mathbb{E}\left[\left(Z_n^T(w) - Z_n^T(w')\right)^2\right]^{1/2} \lesssim nh^{-1}\sqrt{|w-w'|}.$$

Part 2: trajectory regularity of Z_n^T
By the bound on ρ established in the previous part, we can deduce the following covering number bound:

$$N(\varepsilon, \mathcal{W}, \rho) \lesssim N\big(\varepsilon, \mathcal{W}, nh^{-1}\sqrt{|\cdot|}\big) \lesssim N\big(n^{-1}h\varepsilon, \mathcal{W}, \sqrt{|\cdot|}\big) \lesssim N\big(n^{-2}h^2\varepsilon^2, \mathcal{W}, |\cdot|\big) \lesssim n^2h^{-2}\varepsilon^{-2}.$$

Now apply the Gaussian process regularity result from Lemma SA32.

$$\begin{aligned}\mathbb{E}\left[\sup_{\rho(w,w')\le\delta}\left|Z_n^T(w) - Z_n^T(w')\right|\right] &\lesssim \int_0^\delta \sqrt{\log N(\varepsilon, \mathcal{W}, \rho)}\,\mathrm{d}\varepsilon \lesssim \int_0^\delta \sqrt{\log(n^2h^{-2}\varepsilon^{-2})}\,\mathrm{d}\varepsilon \\ &\lesssim \int_0^\delta \left(\sqrt{\log n} + \sqrt{\log 1/\varepsilon}\right)\,\mathrm{d}\varepsilon \lesssim \delta\left(\sqrt{\log n} + \sqrt{\log 1/\delta}\right),\end{aligned}$$

and so

$$\mathbb{E}\left[\sup_{|w-w'|\le\delta_n}\left|Z_n^T(w) - Z_n^T(w')\right|\right] \lesssim \mathbb{E}\left[\sup_{\rho(w,w')\le nh^{-1}\delta_n^{1/2}}\left|Z_n^T(w) - Z_n^T(w')\right|\right] \lesssim nh^{-1}\sqrt{\delta_n \log n},$$

whenever $1/\delta_n$ is at most polynomial in n.

Part 3: existence of the quantile
Apply the Gaussian anti-concentration result from Lemma SA33, noting that Z_n^T is separable, mean-zero and has unit variance:

$$\sup_{t\in\mathbb{R}}\mathbb{P}\left(\left|\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right| - t\right| \le 2\varepsilon_n\right) \le 8\varepsilon_n\left(1 + \mathbb{E}\left[\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right|\right]\right).$$

To bound the supremum on the right hand side, apply the Gaussian process maximal inequality from Lemma SA32 with $\sigma \le 1$ and $N(\varepsilon, \mathcal{W}, \rho) \lesssim n^2h^{-2}\varepsilon^{-2}$:

$$\mathbb{E}\left[\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right|\right] \lesssim 1 + \int_0^2 \sqrt{\log(n^2h^{-2}\varepsilon^{-2})}\,\mathrm{d}\varepsilon \lesssim \sqrt{\log n}.$$

Therefore

$$\sup_{t\in\mathbb{R}}\mathbb{P}\left(\left|\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right| - t\right| \le \varepsilon\right) \lesssim \varepsilon\sqrt{\log n}.$$

Letting $\varepsilon \to 0$ shows that the distribution function of $\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right|$ is continuous, and therefore all of its quantiles exist.

Part 4: validity of the infeasible uniform confidence band
Under Assumption SA3 and with a sufficiently slowly diverging sequence R_n, the strong approximation rate established in Lemma SA10 is

$$\begin{aligned}\sup_{w\in\mathcal{W}}\left|T_n(w) - Z_n^T(w)\right| &\lesssim_{\mathbb{P}} \frac{n^{-1/2}\log n + n^{-3/4}h^{-7/8}(\log n)^{3/8}R_n + n^{-2/3}h^{-1/2}(\log n)^{2/3} + n^{1/2}h^{p\wedge\beta}}{D_{\mathrm{lo}} + 1/\sqrt{nh}} \\ &\ll \frac{1}{\sqrt{\log n}}.\end{aligned}$$

So by Lemma SA34, take ε_n such that

$$\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|T_n(w) - Z_n^T(w)\right| > \varepsilon_n\right) \le \varepsilon_n\sqrt{\log n}$$

and $\varepsilon_n\sqrt{\log n} \to 0$. So by the previously established anti-concentration result,

$$\begin{aligned}
&\mathbb{P}\left(\left|\widehat{f}_W(w) - f_W(w)\right| \le q_{1-\alpha}\sqrt{\Sigma_n(w,w)} \text{ for all } w \in \mathcal{W}\right) \\
&= \mathbb{P}\left(\sup_{w\in\mathcal{W}} |T_n(w)| \le q_{1-\alpha}\right) \\
&\le \mathbb{P}\left(\sup_{w\in\mathcal{W}} \left|Z_n^T(w)\right| \le q_{1-\alpha} + \varepsilon_n\right) + \mathbb{P}\left(\sup_{w\in\mathcal{W}} \left|T_n(w) - Z_n^T(w)\right| > \varepsilon_n\right) \\
&\le \mathbb{P}\left(\sup_{w\in\mathcal{W}} \left|Z_n^T(w)\right| \le q_{1-\alpha}\right) + \mathbb{P}\left(\left|\sup_{w\in\mathcal{W}} \left|Z_n^T(w)\right| - q_{1-\alpha}\right| \le \varepsilon_n\right) + \varepsilon_n\sqrt{\log n} \\
&\le 1 - \alpha + 2\varepsilon_n\sqrt{\log n}.
\end{aligned}$$

The lower bound follows analogously:

$$\begin{aligned}
&\mathbb{P}\left(\left|\widehat{f}_W(w) - f_W(w)\right| \le q_{1-\alpha}\sqrt{\Sigma_n(w,w)} \text{ for all } w \in \mathcal{W}\right) \\
&\ge \mathbb{P}\left(\sup_{w\in\mathcal{W}} \left|Z_n^T(w)\right| \le q_{1-\alpha} - \varepsilon_n\right) - \varepsilon_n\sqrt{\log n} \\
&\ge \mathbb{P}\left(\sup_{w\in\mathcal{W}} \left|Z_n^T(w)\right| \le q_{1-\alpha}\right) - \mathbb{P}\left(\left|\sup_{w\in\mathcal{W}} \left|Z_n^T(w)\right| - q_{1-\alpha}\right| \le \varepsilon_n\right) - \varepsilon_n\sqrt{\log n} \\
&\le 1 - \alpha - 2\varepsilon_n\sqrt{\log n}.
\end{aligned}$$

Finally we apply $\varepsilon_n\sqrt{\log n} \to 0$ to see

$$\left|\mathbb{P}\left(\left|\widehat{f}_W(w) - f_W(w)\right| \le q_{1-\alpha}\sqrt{\Sigma_n(w,w)} \text{ for all } w \in \mathcal{W}\right) - (1-\alpha)\right| \to 0.$$

□

SA5.3.7 Covariance estimation

Before proving Lemma SA11, we provide the following useful concentration inequality. This is essentially a corollary of the U-statistic concentration inequality given in Theorem 3.3 in Giné et al. (2000).

Lemma SA45 (A concentration inequality)
Let X_{ij} be mutually independent random variables for $1 \le i < j \le n$ taking values in a measurable space $\mathcal{X}$. Let h_1, h_2 be measurable functions from $\mathcal{X}$ to $\mathbb{R}$ satisfying the following for all i and j.

$$\begin{aligned}
\mathbb{E}\big[h_1(X_{ij})\big] &= 0, & \mathbb{E}\big[h_2(X_{ij})\big] &= 0, \\
\mathbb{E}\big[h_1(X_{ij})^2\big] &\le \sigma^2, & \mathbb{E}\big[h_2(X_{ij})^2\big] &\le \sigma^2, \\
\big|h_1(X_{ij})\big| &\le M, & \big|h_2(X_{ij})\big| &\le M.
\end{aligned}$$

Consider the sum

$$S_n = \sum_{1\le i<j<r\le n} h_1(X_{ij})h_2(X_{ir}).$$

Then S_n satisfies the concentration inequality

$$\mathbb{P}\big(|S_n| \ge t\big) \le C\exp\left(-\frac{1}{C}\min\left\{\frac{t^2}{n^3\sigma^4}, \frac{t}{\sqrt{n^3\sigma^4}}, \frac{t^{2/3}}{(nM\sigma)^{2/3}}, \frac{t^{1/2}}{M}\right\}\right)$$

for some universal constant $C > 0$ and for all $t > 0$.

Proof (Lemma SA45)
We proceed in three main steps. Firstly we write S_n as a second-order U-statistic where we use double indices instead of single indices. Then we use a decoupling result to introduce extra independence. Finally a concentration result is applied to the decoupled U-statistic.

Part 1: writing S_n as a second-order U-statistic
Note that we can write S_n as the second-order U-statistic

$$S_n = \sum_{1\le i<j\le n} \sum_{1\le q<r\le n} h_{ijqr}(X_{ij}, X_{qr}),$$

where

$$h_{ijqr}(a,b) = h_1(a)h_2(b)\,\mathbb{I}\{j<r,\, q=i\}.$$

Although this may look like a fourth-order U-statistic, it is in fact second-order. This is due to independence of the variables X_{ij}, and by treating (i,j) as a single index.

Part 2: decoupling
By the decoupling result of Theorem 1 from de la Peña and Montgomery-Smith (1995), there exists a universal constant $C_1 > 0$ satisfying

$$\mathbb{P}\big(|S_n| \ge t\big) \le C_1 \mathbb{P}\left(C_1 |\widetilde{S}_n| \ge t\right),$$

where

$$\widetilde{S}_n = \sum_{1\le i<j\le n} \sum_{1\le q<r\le n} h_{ijqr}(X_{ij}, X'_{qr}),$$

with (X'_{ij}) an independent copy of (X_{ij}).

Part 3: U-statistic concentration
The U-statistic kernel $h_{ijqr}(X_{ij}, X'_{qr})$ is totally degenerate in the sense that

$$\mathbb{E}[h_{ijqr}(X_{ij}, X'_{qr}) \mid X_{ij}] = \mathbb{E}[h_{ijqr}(X_{ij}, X'_{qr}) \mid X'_{qr}] = 0.$$

Define and bound the following quantities:

$$\begin{aligned}
A &= \max_{ijqr} \|h_{ijqr}(X_{ij}, X'_{qr})\|_\infty \le M^2,\\
B &= \max\left\{ \left\| \sum_{1\le i<j\le n} \mathbb{E}\Big[h_{ijqr}(X_{ij}, X'_{qr})^2 \mid X_{ij}\Big] \right\|_\infty, \left\| \sum_{1\le q<r\le n} \mathbb{E}\Big[h_{ijqr}(X_{ij}, X'_{qr})^2 \mid X'_{qr}\Big] \right\|_\infty \right\}^{1/2}\\
&= \max\left\{ \left\| \sum_{1\le i<j\le n} h_1(X_{ij})^2 \mathbb{E}\big[h_2(X'_{qr})^2\big] \mathbb{I}\{j<r, q=i\} \right\|_\infty, \right.\\
&\qquad \left. \left\| \sum_{1\le q<r\le n} \mathbb{E}\big[h_1(X_{ij})^2\big] h_2(X'_{qr})^2 \mathbb{I}\{j<r, q=i\} \right\|_\infty \right\}^{1/2}\\
&\le \max\left\{n^2M^2\sigma^2, nM^2\sigma^2\right\}^{1/2} = nM\sigma,\\
C &= \left(\sum_{1\le i<j\le n} \sum_{1\le q<r\le n} \mathbb{E}\big[h_{ijqr}(X_{ij}, X'_{qr})^2\big] \right)^{1/2}\\
&= \left(\sum_{1\le i<j<r\le n} \mathbb{E}\big[h_1(X_{ij})^2 h_2(X'_{ir})^2\big] \right)^{1/2} \le \sqrt{n^3\sigma^4},\\
D &= \sup_{f,g} \left\{ \sum_{1\le i<j\le n} \sum_{1\le q<r\le n} \mathbb{E}\big[h_{ijqr}(X_{ij}, X'_{qr}) f_{ij}(X_{ij}) g_{qr}(X'_{qr})\big] \;:\right.
\end{aligned}$$

$$\sum_{1\le i<j\le n}\mathbb{E}\big[f_{ij}(X_{ij})^2\big]\le 1,\ \sum_{1\le q<r\le n}\mathbb{E}\big[g_{qr}(X'_{qr})^2\big]\le 1\Bigg\}$$

$$= \sup_{f,g}\Bigg\{\sum_{1\le i<j<r\le n}\mathbb{E}\big[h_1(X_{ij})f_{ij}(X_{ij})\big]\mathbb{E}\big[h_2(X'_{ir})g_{ir}(X'_{ir})\big]\ :$$

$$\sum_{1\le i<j\le n}\mathbb{E}\big[f_{ij}(X_{ij})^2\big]\le 1,\ \sum_{1\le q<r\le n}\mathbb{E}\big[g_{qr}(X'_{qr})^2\big]\le 1\Bigg\}$$

$$\le \sup_{f,g}\Bigg\{\sum_{1\le i<j<r\le n}\mathbb{E}\big[h_1(X_{ij})^2\big]^{1/2}\mathbb{E}\big[f_{ij}(X_{ij})^2\big]^{1/2}\mathbb{E}\big[h_2(X'_{ir})^2\big]^{1/2}\mathbb{E}\big[g_{ir}(X'_{ir})^2\big]^{1/2}\ :$$

$$\sum_{1\le i<j\le n}\mathbb{E}\big[f_{ij}(X_{ij})^2\big]\le 1,\ \sum_{1\le q<r\le n}\mathbb{E}\big[g_{qr}(X'_{qr})^2\big]\le 1\Bigg\}$$

$$\le \sigma^2\sup_{f,g}\Bigg\{\sum_{1\le i<j\le n}\mathbb{E}\big[f_{ij}(X_{ij})^2\big]^{1/2}\sum_{1\le r\le n}\mathbb{E}\big[g_{ir}(X'_{ir})^2\big]^{1/2}\ :$$

$$\sum_{1\le i<j\le n}\mathbb{E}\big[f_{ij}(X_{ij})^2\big]\le 1,\ \sum_{1\le q<r\le n}\mathbb{E}\big[g_{qr}(X'_{qr})^2\big]\le 1\Bigg\}$$

$$\le \sigma^2\sup_{f,g}\Bigg\{\left(n^2\sum_{1\le i<j\le n}\mathbb{E}\big[f_{ij}(X_{ij})^2\big]\right)^{1/2}\left(n\sum_{1\le r\le n}\mathbb{E}\big[g_{ir}(X'_{ir})^2\big]\right)^{1/2}\ :$$

$$\sum_{1\le i<j\le n}\mathbb{E}\big[f_{ij}(X_{ij})^2\big]\le 1,\ \sum_{1\le q<r\le n}\mathbb{E}\big[g_{qr}(X'_{qr})^2\big]\le 1\Bigg\}\le\sqrt{n^3\sigma^4}.$$

By Theorem 3.3 in Giné et al. (2000), we have that for some universal constant $C_2>0$ and for all $t>0$,

$$\begin{aligned}\mathbb{P}\left(|\widetilde{S}_n|\ge t\right)&\le C_2\exp\left(-\frac{1}{C_2}\min\left\{\frac{t^2}{C^2},\frac{t}{D},\frac{t^{2/3}}{B^{2/3}},\frac{t^{1/2}}{A^{1/2}}\right\}\right)\\&\le C_2\exp\left(-\frac{1}{C_2}\min\left\{\frac{t^2}{n^3\sigma^4},\frac{t}{\sqrt{n^3\sigma^4}},\frac{t^{2/3}}{(nM\sigma)^{2/3}},\frac{t^{1/2}}{M}\right\}\right)\end{aligned}$$

Part 4: Conclusion
By the previous parts and absorbing constants into a new constant $C>0$, we therefore have

$$\begin{aligned}\mathbb{P}\left(|S_n|\ge t\right)&\le C_1\mathbb{P}\left(C_1|\widetilde{S}_n|\ge t\right)\\&\le C_1C_2\exp\left(-\frac{1}{C_2}\min\left\{\frac{t^2}{n^3\sigma^4C_1^2},\frac{t}{\sqrt{n^3\sigma^4C_1}},\frac{t^{2/3}}{(nM\sigma C_1)^{2/3}},\frac{t^{1/2}}{MC_1^{1/2}}\right\}\right)\\&\le C\exp\left(-\frac{1}{C}\min\left\{\frac{t^2}{n^3\sigma^4},\frac{t}{\sqrt{n^3\sigma^4}},\frac{t^{2/3}}{(nM\sigma)^{2/3}},\frac{t^{1/2}}{M}\right\}\right).\end{aligned}$$

□

Proof (Lemma SA11)
Throughout this proof we will write k_{ij} for $k_h(W_{ij},w)$ and k'_{ij} for $k_h(W_{ij},w')$, in the interest of brevity. Similarly we write S_{ijr} to denote $S_{ijr}(w,w')$. The estimand and estimator are reproduced below for clarity.

$$\Sigma_n(w,w')=\frac{2}{n(n-1)}\mathbb{E}[k_{ij}k'_{ij}]+\frac{4(n-2)}{n(n-1)}\mathbb{E}[k_{ij}k'_{ir}]-\frac{4n-6}{n(n-1)}\mathbb{E}[k_{ij}]\mathbb{E}[k'_{ij}]$$

$$\widehat{\Sigma}_n(w,w')=\frac{2}{n(n-1)}\frac{2}{n(n-1)}\sum_{i<j}k_{ij}k'_{ij}+\frac{4(n-2)}{n(n-1)}\frac{6}{n(n-1)(n-2)}\sum_{i<j<r}S_{ijr}-\frac{4n-6}{n(n-1)}\widehat{f}_W(w)\widehat{f}_W(w'),$$

where

$$S_{ijr} = \frac{1}{6}\Big(k_{ij}k'_{ir} + k_{ij}k'_{jr} + k_{ir}k'_{ij} + k_{ir}k'_{jr} + k_{jr}k'_{ij} + k_{jr}k'_{ir}\Big).$$

We will prove uniform consistency of each of the three terms separately.

Part 1: uniform consistency of the $\widehat{f}_W(w)\widehat{f}_W(w')$ term

By boundedness of f_W and Theorem SA1, $\widehat{f}_W$ is uniformly bounded in probability. Noting that $\mathbb{E}[\widehat{f}_W(w)] = \mathbb{E}[k_{ij}]$ and by Lemma SA6,

$$\begin{aligned}
&\sup_{w,w'\in\mathcal{W}} \left| \frac{\widehat{f}_W(w)\widehat{f}_W(w') - \mathbb{E}\big[k_{ij}\big]\mathbb{E}\big[k_{ij'}\big]}{\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}} \right| \\
&= \sup_{w,w'\in\mathcal{W}} \left| \frac{\widehat{f}_W(w)\widehat{f}_W(w') - \mathbb{E}\big[\widehat{f}_W(w)\big]\mathbb{E}\big[\widehat{f}_W(w')\big]}{\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}} \right| \\
&\leq \sup_{w,w'\in\mathcal{W}} \left| \frac{\widehat{f}_W(w) - \mathbb{E}\big[\widehat{f}_W(w)\big]}{\sqrt{\Sigma_n(w,w)}}\widehat{f}_W(w') + \frac{\widehat{f}_W(w') - \mathbb{E}\big[\widehat{f}_W(w')\big]}{\sqrt{\Sigma_n(w',w')}}\mathbb{E}\big[\widehat{f}_W(w)\big] \right| \\
&\lesssim_{\mathbb{P}} \sup_{w\in\mathcal{W}} \left| \frac{\widehat{f}_W(w) - \mathbb{E}\big[\widehat{f}_W(w)\big]}{\sqrt{\Sigma_n(w,w)}} \right| \lesssim_{\mathbb{P}} \sup_{w\in\mathcal{W}} \left| \frac{L_n(w)}{\sqrt{\Sigma_n(w,w)}} \right| + \sqrt{n^2h}\sup_{w\in\mathcal{W}}|Q_n(w)| + \sqrt{n^2h}\sup_{w\in\mathcal{W}}|E_n(w)| \\
&\lesssim_{\mathbb{P}} \sup_{w\in\mathcal{W}} \left| \frac{L_n(w)}{\sqrt{\Sigma_n(w,w)}} \right| + \sqrt{n^2h}\frac{1}{n} + \sqrt{n^2h}\sqrt{\frac{\log n}{n^2h}} \lesssim_{\mathbb{P}} \sup_{w\in\mathcal{W}} \left| \frac{L_n(w)}{\sqrt{\Sigma_n(w,w)}} \right| + \sqrt{\log n}.
\end{aligned}$$

Now consider the function class

$$\mathcal{F} = \left\{ a \mapsto \frac{\mathbb{E}\big[k_h(W_{ij},w) \mid A_i = a\big] - \mathbb{E}\big[k_h(W_{ij},w)\big]}{\sqrt{n\Sigma_n(w,w)}} : w \in \mathcal{W} \right\},$$

noting that

$$\frac{L_n(w)}{\Sigma_n(w,w)^{1/2}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_w(A_i)$$

is an empirical process evaluated at $g_w \in \mathcal{F}$. By the lower bound on $\Sigma_n(w,w)$ from Lemma SA6 with $\inf_{\mathcal{W}} f_W(w) > 0$ and since $nh \gtrsim \log n$, the class $\mathcal{F}$ has a constant envelope function given by $F(a) \lesssim \sqrt{nh}$. Clearly $M = \sup_a F(a) \lesssim \sqrt{nh}$. Also by definition of Σ_n and orthogonality of L_n, Q_n and E_n, we have $\sup_{f\in\mathcal{F}} \mathbb{E}[f(A_i)^2] \leq \sigma^2 = 1$. To verify a VC-type condition on $\mathcal{F}$ we need to establish the regularity of the process. By Lipschitz properties of L_n and Σ_n derived in the proofs of Lemma SA4 and Theorem SA4 respectively, we have

$$\begin{aligned}
\left| \frac{L_n(w)}{\sqrt{\Sigma_n(w,w)}} - \frac{L_n(w')}{\sqrt{\Sigma_n(w',w')}} \right| &\lesssim \frac{\big|L_n(w) - L_n(w')\big|}{\sqrt{\Sigma_n(w,w)}} + |L_n(w')| \left| \frac{1}{\sqrt{\Sigma_n(w,w)}} - \frac{1}{\sqrt{\Sigma_n(w',w')}} \right| \\
&\lesssim \sqrt{n^2h}|w-w'| + \left| \frac{\Sigma_n(w,w) - \Sigma_n(w',w')}{\Sigma_n(w,w)\sqrt{\Sigma_n(w',w')}} \right| \\
&\lesssim \sqrt{n^2h}|w-w'| + (n^2h)^{3/2}\,|\Sigma_n(w,w) - \Sigma_n(w',w')| \\
&\lesssim \sqrt{n^2h}|w-w'| + (n^2h)^{3/2}n^{-1}h^{-3}|w-w'| \lesssim n^4|w-w'|
\end{aligned}$$

uniformly over $w, w' \in \mathcal{W}$. Therefore by compactness of $\mathcal{W}$ we have the covering number bound

$$N(\mathcal{F}, \|\cdot\|_\infty, \varepsilon) \lesssim N(\mathcal{W}, |\cdot|, n^{-4}\varepsilon) \lesssim n^4\varepsilon^{-1}.$$

Thus by Lemma SA25,

$$\mathbb{E}\left[\sup_{w\in\mathcal{W}}\left|\frac{L_n(w)}{\sqrt{\Sigma_n(w,w)}}\right|\right] \lesssim \sqrt{\log n} + \frac{\sqrt{nh}\log n}{\sqrt{n}} \lesssim \sqrt{\log n}.$$

Thus

$$\sup_{w,w'\in\mathcal{W}}\left|\frac{\widehat{f}_W(w)\widehat{f}_W(w') - \mathbb{E}\big[k_{ij}\big]\mathbb{E}\big[k_{ij'}\big]}{\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}}\right| \lesssim_{\mathbb{P}} \sqrt{\log n}.$$

Part 2: decomposition of the S_{ijr} term

We first decompose the S_{ijr} term into two parts, and obtain a pointwise concentration result for each. This is extended to a uniform concentration result by considering the regularity of the covariance estimator process. Note that $\mathbb{E}[S_{ijr}] = \mathbb{E}[k_{ij}k'_{ir}]$, and hence

$$\frac{6}{n(n-1)(n-2)}\sum_{i<j<r}\big(S_{ijr} - \mathbb{E}[k_{ij}k'_{ir}]\big) = \frac{6}{n(n-1)(n-2)}\sum_{i<j<r}S^{(1)}_{ijr} + \frac{6}{n(n-1)(n-2)}\sum_{i<j<r}S^{(2)}_{ijr}$$

where

$$S^{(1)}_{ijr} = S_{ijr} - \mathbb{E}[S_{ijr}\mid \mathbf{A}_n], \qquad S^{(2)}_{ijr} = \mathbb{E}[S_{ijr}\mid \mathbf{A}_n] - \mathbb{E}[S_{ijr}].$$

Part 3: pointwise concentration of the $S^{(1)}_{ijr}$ term

By symmetry in i,j and r it is sufficient to consider only the first summand in the definition of S_{ijr}. By conditional independence properties, we have the decomposition

$$\begin{aligned}
&\frac{6}{n(n-1)(n-2)}\sum_{i<j<r}\Big(k_{ij}k'_{ir} - \mathbb{E}[k_{ij}k'_{ir}\mid\mathbf{A}_n]\Big)\\
&\quad= \frac{6}{n(n-1)(n-2)}\sum_{i<j<r}\Big(k_{ij}k'_{ir} - \mathbb{E}[k_{ij}\mid\mathbf{A}_n]\mathbb{E}[k'_{ir}\mid\mathbf{A}_n]\Big)\\
&\quad= \frac{6}{n(n-1)(n-2)}\sum_{i<j<r}\Big(\big(k_{ij}-\mathbb{E}[k_{ij}\mid\mathbf{A}_n]\big)\big(k'_{ir}-\mathbb{E}[k'_{ir}\mid\mathbf{A}_n]\big)\\
&\qquad+\big(k_{ij}-\mathbb{E}[k_{ij}\mid\mathbf{A}_n]\big)\mathbb{E}[k'_{ir}\mid\mathbf{A}_n] + \big(k'_{ir}-\mathbb{E}[k'_{ir}\mid\mathbf{A}_n]\big)\mathbb{E}[k_{ij}\mid\mathbf{A}_n]\Big)\\
&\quad= \frac{6}{n(n-1)(n-2)}\sum_{i<j<r}\Big(k_{ij}-\mathbb{E}[k_{ij}\mid\mathbf{A}_n]\Big)\Big(k'_{ir}-\mathbb{E}[k'_{ir}\mid\mathbf{A}_n]\Big) \qquad (7)\\
&\qquad+\frac{2}{(n-1)(n-2)}\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\Big(k_{ij}-\mathbb{E}[k_{ij}\mid\mathbf{A}_n]\Big)\cdot\frac{3}{n}\sum_{r=j+1}^{n}\mathbb{E}[k'_{ir}\mid\mathbf{A}_n] \qquad (8)\\
&\qquad+\frac{2}{(n-1)(n-2)}\sum_{i=1}^{n-2}\sum_{r=i+2}^{n}\Big(k'_{ir}-\mathbb{E}[k'_{ir}\mid\mathbf{A}_n]\Big)\cdot\frac{3}{n}\sum_{j=i+1}^{r-1}\mathbb{E}[k_{ij}\mid\mathbf{A}_n]. \qquad (9)
\end{aligned}$$

For the term in (7), note that conditional on $\mathbf{A}_n$, we have that $k_{ij}-\mathbb{E}[k_{ij}\mid\mathbf{A}_n]$ are conditionally mean-zero and conditionally independent, as the only randomness is from $\mathbf{V}_n$. Also $\mathrm{Var}[k_{ij}\mid\mathbf{A}_n]\lesssim\sigma^2 := 1/h$ and $|k_{ij}|\lesssim M := 1/h$ uniformly. The same is true for k'_{ij}. Thus by Lemma SA45 for some universal constant $C_1>0$:

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i<j<r}\Big(k_{ij}-\mathbb{E}[k_{ij}\mid\mathbf{A}_n]\Big)\Big(k'_{ir}-\mathbb{E}[k'_{ir}\mid\mathbf{A}_n]\Big)\right| > t \;\middle|\; \mathbf{A}_n\right)\\
&\quad\le C_1\exp\left(-\frac{1}{C_1}\min\left\{\frac{t^2}{n^3\sigma^4}, \frac{t}{\sqrt{n^3\sigma^4}}, \frac{t^{2/3}}{(nM\sigma)^{2/3}}, \frac{t^{1/2}}{M}\right\}\right)\\
&\quad\le C_1\exp\left(-\frac{1}{C_1}\min\left\{\frac{t^2h^2}{n^3}, \frac{th}{\sqrt{n^3}}, \frac{t^{2/3}h}{n^{2/3}}, t^{1/2}h\right\}\right),
\end{aligned}$$

and therefore with $t \geq 1$ and since $nh \gtrsim \log n$, introducing and adjusting a new constant C_2 where necessary,

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{6}{n(n-1)(n-2)}\sum_{i<j<r}\Big(k_{ij}-\mathbb{E}[k_{ij}\mid \mathbf{A}_n]\Big)\Big(k'_{ir}-\mathbb{E}[k'_{ir}\mid \mathbf{A}_n]\Big)\right| > t\frac{\log n}{\sqrt{n^3h^2}} \;\middle|\; \mathbf{A}_n\right)\\
&\quad\leq \mathbb{P}\left(\left|\sum_{i<j<r}\Big(k_{ij}-\mathbb{E}[k_{ij}\mid \mathbf{A}_n]\Big)\Big(k'_{ir}-\mathbb{E}[k'_{ir}\mid \mathbf{A}_n]\Big)\right| > tn^{3/2}h^{-1}\log n/24 \;\middle|\; \mathbf{A}_n\right)\\
&\quad\leq C_2\exp\left(-\frac{1}{C_2}\min\Big\{(t\log n)^2, t\log n, (t\log n)^{2/3}(nh)^{1/3}, (tnh\log n)^{1/2}n^{1/4}\Big\}\right)\\
&\quad\leq C_2\exp\left(-\frac{1}{C_2}\min\Big\{t\log n, t\log n, t^{2/3}\log n, t^{1/2}n^{1/4}\log n\Big\}\right)\\
&\quad= C_2\exp\left(-\frac{t^{2/3}\log n}{C_2}\right) = C_2 n^{-t^{2/3}/C_2}.
\end{aligned}$$

Now for the term in (8), note that $\frac{3}{n}\sum_{r=j+1}^{n}\mathbb{E}[k'_{ir}\mid \mathbf{A}_n]$ is $\mathbf{A}_n$-measurable and bounded uniformly in i,j. Also, using the previously established conditional variance and almost sure bounds on k_{ij}, Bernstein's inequality (Lemma SA29) applied conditionally gives for some constant $C_3>0$

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{2}{(n-1)(n-2)}\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\Big(k_{ij}-\mathbb{E}[k_{ij}\mid \mathbf{A}_n]\Big)\cdot\frac{3}{n}\sum_{r=j+1}^{n}\mathbb{E}[k'_{ir}\mid \mathbf{A}_n]\right| > t\sqrt{\frac{\log n}{n^2h}} \;\middle|\; \mathbf{A}_n\right)\\
&\quad\leq 2\exp\left(-\frac{t^2n^2\log n/(n^2h)}{C_3/(2h)+C_3t\sqrt{\log n/(n^2h)}/(2h)}\right)\\
&\quad= 2\exp\left(-\frac{t^2\log n}{C_3/2+C_3t\sqrt{\log n/(n^2h)}/2}\right) \leq 2\exp\left(-\frac{t^2\log n}{C_3}\right) = 2n^{-t^2/C_3}.
\end{aligned}$$

The term in (9) is controlled in exactly the same way. Putting these together, noting the symmetry in i,j,r and taking a marginal expectation, we obtain the unconditional pointwise concentration inequality

$$\mathbb{P}\left(\left|\frac{6}{n(n-1)(n-2)}\sum_{i<j<r}S^{(1)}_{ijr}\right| > t\frac{\log n}{\sqrt{n^3h^2}}+t\sqrt{\frac{\log n}{n^2h}}\right) \leq C_2n^{-t^{2/3}/C_2}+4n^{-t^2/(4C_3)}.$$

Multiplying by $\big(\Sigma_n(w,w)+\Sigma_n(w',w')\big)^{-1/2} \lesssim \sqrt{n^2h}$ gives (adjusting constants if necessary)

$$\mathbb{P}\left(\left|\frac{6}{n(n-1)(n-2)}\sum_{i<j<r}\frac{S^{(1)}_{ijr}}{\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}}\right| > t\frac{\log n}{\sqrt{nh}}+t\sqrt{\log n}\right) \leq C_2n^{-t^{2/3}/C_2}+4n^{-t^2/(4C_3)}.$$

Part 4: pointwise concentration of the $S^{(2)}_{ijr}$ term

We apply the U-statistic concentration inequality from Lemma SA36. Note that the terms $\mathbb{E}[S_{ijr}\mid \mathbf{A}_n]$ are permutation-symmetric functions of the random variables A_i, A_j and A_r only, making $S^{(2)}_{ijr}$ the summands of a (non-degenerate) mean-zero third-order U-statistic. While we could apply a third-order Hoeffding decomposition here to achieve degeneracy, it is unnecessary as Lemma SA36 is general enough to deal with the non-degenerate case directly. The quantity of interest here is

$$\frac{6}{n(n-1)(n-2)}\sum_{i<j<r}S^{(2)}_{ijr} = \frac{6}{n(n-1)(n-2)}\sum_{i<j<r}\Big(\mathbb{E}[S_{ijr}\mid \mathbf{A}_n]-\mathbb{E}[S_{ijr}]\Big).$$

Note that by conditional independence,

$$\big|\mathbb{E}\big[k_{ij}k_{ir}\mid \mathbf{A}_n\big]\big| = \big|\mathbb{E}\big[k_{ij}\mid \mathbf{A}_n\big]\mathbb{E}\big[k_{ir}\mid \mathbf{A}_n\big]\big| \lesssim 1,$$

and similarly for the other summands in S_{ijr}, giving the almost-sure bound $|S^{(2)}_{ijr}| \lesssim 1$. We also have

$$\begin{aligned}\operatorname{Var}\left[\mathbb{E}[k_{ij} \mid A_i]\mathbb{E}[k'_{ir} \mid A_i]\right] &\lesssim \operatorname{Var}\left[\mathbb{E}[k_{ij} \mid A_i]\right] + \operatorname{Var}\left[\mathbb{E}[k'_{ir} \mid A_i]\right] \lesssim n \operatorname{Var}[L_n(w)] + n \operatorname{Var}[L_n(w')] \\ &\lesssim n\Sigma_n(w,w) + n\Sigma_n(w',w')\end{aligned}$$

and similarly for the other summands in S_{ijr}, giving the conditional variance bound

$$\mathbb{E}[\mathbb{E}[S^{(2)}_{ijr} \mid A_i]^2] \lesssim n\Sigma_n(w,w) + n\Sigma_n(w',w').$$

So Lemma SA36 and Lemma SA6 give the pointwise concentration inequality

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{6}{n(n-1)(n-2)}\sum_{i<j<r} S^{(2)}_{ijr}\right| > t\sqrt{\log n}\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}\right) \\ &\quad\le 4\exp\left(-\frac{nt^2(\Sigma_n(w,w)+\Sigma_n(w',w'))\log n}{C_4(n\Sigma_n(w,w)+n\Sigma_n(w',w')) + C_4 t\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}\sqrt{\log n}}\right) \\ &\quad\le 4\exp\left(-\frac{t^2\log n}{C_4 + C_4 t(\Sigma_n(w,w)+\Sigma_n(w',w'))^{-1/2}\sqrt{\log n}/n}\right) \le 4\exp\left(-\frac{t^2\log n}{C_4 + C_4 t\sqrt{h}}\right) \le 4n^{-t^2/C_4}\end{aligned}$$

for some universal constant $C_4 > 0$ (which may change from line to line), since the order of this U-statistic is fixed at three.

Part 5: concentration of the S_{ijr} term on a mesh

Pick $\delta_n \to 0$ with $\log 1/\delta_n \lesssim \log n$. Let $\mathcal{W}_\delta$ be a δ_n-covering of $\mathcal{W}$ with cardinality $O(1/\delta_n)$. Then $\mathcal{W}_\delta \times \mathcal{W}_\delta$ is a $2\delta_n$-covering of $\mathcal{W}\times\mathcal{W}$ with cardinality $O(1/\delta_n^2)$, under the Manhattan metric $d\big((w_1,w_1'),(w_2,w_2')\big) = |w_1-w_2| + |w_1'-w_2'|$. By the previous parts, we have that for fixed w and w':

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{6}{n(n-1)(n-2)}\sum_{i<j<r}\frac{S_{ijr}(w,w') - \mathbb{E}[S_{ijr}(w,w')]}{\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}}\right| > t\frac{\log n}{\sqrt{nh}} + 2t\sqrt{\log n}\right) \\ &\quad\le C_2 n^{-t^{2/3}/C_2} + 4n^{-t^2/(4C_3)} + 4n^{-t^2/C_4}.\end{aligned}$$

Taking a union bound over $\mathcal{W}_\delta\times\mathcal{W}_\delta$, noting that $nh \gtrsim \log n$ and adjusting constants gives

$$\begin{aligned}&\mathbb{P}\left(\sup_{w,w'\in\mathcal{W}_\delta}\left|\frac{6}{n(n-1)(n-2)}\sum_{i<j<r}\frac{S_{ijr}(w,w') - \mathbb{E}[S_{ijr}(w,w')]}{\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}}\right| > t\sqrt{\log n}\right) \\ &\quad\lesssim \delta_n^{-2}\Big(C_2 n^{-t^{2/3}/C_2} + 4n^{-t^2/(4C_3)} + 4n^{-t^2/C_4}\Big) \lesssim \delta_n^{-2} n^{-t^{2/3}/C_5},\end{aligned}$$

for some constant $C_5 > 0$.

Part 6: regularity of the S_{ijr} term

Next we bound the fluctuations in $S_{ijr}(w,w')$. Writing $k_{ij}(w)$ for $k_h(W_{ij},w)$, note that

$$\begin{aligned}\big|k_{ij}(w_1)k_{ir}(w_1') - k_{ij}(w_2)k_{ir}(w_2')\big| &\lesssim \frac{1}{h}\big|k_{ij}(w_1) - k_{ij}(w_2)\big| + \frac{1}{h}\big|k_{ir}(w_1') - k_{ir}(w_2')\big| \\ &\lesssim \frac{1}{h^3}\Big(|w_1-w_2| + |w_1'-w_2'|\Big),\end{aligned}$$

using the Lipschitz property of the kernel. Similarly for the other summands in S_{ijr}. Therefore

$$\sup_{|w_1-w_2|\le\delta_n}\sup_{|w_1'-w_2'|\le\delta_n}\big|S_{ijr}(w_1,w_1') - S_{ijr}(w_2,w_2')\big| \lesssim \delta_n h^{-3}.$$

Also as noted in the proof of Theorem SA4,

$$\sup_{|w_1-w_2|\le\delta_n}\sup_{|w_1'-w_2'|\le\delta_n}\big|\Sigma_n(w_1,w_1') - \Sigma_n(w_2,w_2')\big| \lesssim \delta_n n^{-1}h^{-3}.$$

Therefore since $\sqrt{\Sigma_n(w,w)} \gtrsim \sqrt{n^2 h}$ and $|S_{ijr}| \lesssim h^{-2}$, using the elementary fact $\frac{a}{\sqrt{b}} - \frac{c}{\sqrt{d}} = \frac{a-c}{\sqrt{b}} + c\frac{d-b}{\sqrt{bd}\sqrt{b+d}}$,

$$\sup_{|w_1-w_2|\le\delta_n} \sup_{|w_1'-w_2'|\le\delta_n} \left| \frac{S_{ijr}(w_1,w_1')}{\sqrt{\Sigma_n(w_1,w_1)+\Sigma_n(w_1',w_1')}} - \frac{S_{ijr}(w_2,w_2')}{\sqrt{\Sigma_n(w_2,w_2)+\Sigma_n(w_2',w_2')}} \right|$$
$$\lesssim \delta_n h^{-3}\sqrt{n^2 h} + h^{-2}\delta_n n^{-1} h^{-3}(n^2 h)^{3/2} \lesssim \delta_n n h^{-5/2} + \delta_n n^2 h^{-7/2} \lesssim \delta_n n^6,$$

where in the last line we use that $1/h \lesssim n$.

Part 7: uniform concentration of the S_{ijr} term
By setting $\delta_n = n^{-6}\sqrt{\log n}$, the fluctuations can be at most $\sqrt{\log n}$, so we have for $t \ge 1$

$$\mathbb{P}\left(\sup_{w,w'\in\mathcal{W}} \left| \frac{6}{n(n-1)(n-2)} \sum_{i<j<r} \frac{S_{ijr}(w,w') - \mathbb{E}[S_{ijr}(w,w')]}{\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}} \right| > 2t\sqrt{\log n} \right) \lesssim \delta_n^{-2} n^{-t^{2/3}/C_5} \lesssim n^{12-t^{2/3}/C_5}.$$

This converges to zero for any sufficiently large t, so

$$\sup_{w,w'\in\mathcal{W}} \left| \frac{6}{n(n-1)(n-2)} \sum_{i<j<r} \frac{S_{ijr}(w,w') - \mathbb{E}[S_{ijr}(w,w')]}{\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}} \right| \lesssim_{\mathbb{P}} \sqrt{\log n}.$$

Part 8: decomposition of the $k_{ij}k_{ij}'$ term
We move on to the final term in the covariance estimator. We have the decomposition

$$\frac{2}{n(n-1)} \sum_{i<j} \left(k_{ij}k_{ij}' - \mathbb{E}\big[k_{ij}k_{ij}'\big] \right) = \frac{2}{n(n-1)} \sum_{i<j} S_{ij}^{(1)} + \frac{2}{n(n-1)} \sum_{i<j} S_{ij}^{(2)},$$

where

$$S_{ij}^{(1)} = k_{ij}k_{ij}' - \mathbb{E}\big[k_{ij}k_{ij}' \mid \mathbf{A}_n\big], \qquad S_{ij}^{(2)} = \mathbb{E}\big[k_{ij}k_{ij}' \mid \mathbf{A}_n\big] - \mathbb{E}\big[k_{ij}k_{ij}'\big].$$

Part 9: pointwise concentration of the $S_{ij}^{(1)}$ term
Conditioning on $\mathbf{A}_n$, the variables $S_{ij}^{(1)}$ are conditionally independent and conditionally mean-zero. They further satisfy $|S_{ij}^{(1)}| \lesssim h^{-2}$ and the conditional variance bound $\mathbb{E}\big[\big(S_{ij}^{(1)}\big)^2 \mid \mathbf{A}_n\big] \lesssim h^{-3}$. Therefore applying Bernstein's inequality (Lemma SA29) conditional on $\mathbf{A}_n$, we obtain the pointwise in w, w' concentration inequality

$$\begin{aligned}
\mathbb{P}\left(\left| \frac{2}{n(n-1)} \sum_{i<j} S_{ij}^{(1)} \right| > t\sqrt{\frac{\log n}{n^2 h^3}} \;\middle|\; \mathbf{A}_n \right) &\le 2\exp\left(-\frac{t^2 n^2 \log n/(n^2 h^3)}{C_6 h^{-3}/2 + C_6 t h^{-2}\sqrt{\log n/(n^2 h^3)}/2} \right) \\
&\le 2\exp\left(-\frac{t^2 \log n}{C_6/2 + C_6 t\sqrt{\log n/(n^2 h)}/2} \right) \\
&\le 2\exp\left(-\frac{t^2 \log n}{C_6} \right) = 2n^{-t^2/C_6},
\end{aligned}$$

where C_6 is a universal positive constant.

Part 10: pointwise concentration of the $S_{ij}^{(2)}$ term
We apply the U-statistic concentration inequality from Lemma SA36. Note that $S_{ij}^{(2)}$ are permutation-symmetric functions of the random variables A_i and A_j only, making them the summands of a (non-degenerate) mean-zero second-order U-statistic. Note that $\big|S_{ij}^{(2)}\big| \lesssim h^{-1}$ and so trivially $\mathbb{E}\big[\mathbb{E}[S_{ij}^{(2)} \mid A_i]^2\big] \lesssim h^{-2}$. Thus by Lemma SA36, since the order of this U-statistic is fixed at two, for some universal positive constant C_7 we

have

$$\mathbb{P}\left(\left|\frac{2}{n(n-1)}\sum_{i<j} S_{ij}^{(2)}\right| > t\sqrt{\frac{\log n}{nh^2}}\right) \le 2\exp\left(-\frac{t^2 n\log n/(nh^2)}{C_7 h^{-2}/2 + C_7 t h^{-1}\sqrt{\log n/(nh^2)}/2}\right)$$
$$\le 2\exp\left(-\frac{t^2\log n}{C_7/2 + C_7 t\sqrt{\log n/n}/2}\right)$$
$$\le 2\exp\left(-\frac{t^2\log n}{C_7}\right) = 2n^{-t^2/C_7}.$$

Part 11: concentration of the $k_{ij}k'_{ij}$ term on a mesh
As before, use a union bound on the mesh $\mathcal{W}_\delta \times \mathcal{W}_\delta$.

$$\mathbb{P}\left(\sup_{w,w'\in\mathcal{W}_\delta}\left|\frac{2}{n(n-1)}\sum_{i<j}\Big(k_{ij}k'_{ij} - \mathbb{E}\big[k_{ij}k'_{ij}\big]\Big)\right| > t\sqrt{\frac{\log n}{n^2h^3}} + t\sqrt{\frac{\log n}{nh^2}}\right)$$
$$\le \mathbb{P}\left(\sup_{w,w'\in\mathcal{W}_\delta}\left|\frac{2}{n(n-1)}\sum_{i<j} S_{ij}^{(1)}\right| > t\sqrt{\frac{\log n}{n^2h^3}}\right) + \mathbb{P}\left(\sup_{w,w'\in\mathcal{W}_\delta}\left|\frac{2}{n(n-1)}\sum_{i<j} S_{ij}^{(2)}\right| > t\sqrt{\frac{\log n}{nh^2}}\right)$$
$$\lesssim \delta_n^{-2}n^{-t^2/C_6} + \delta_n^{-2}n^{-t^2/C_7}.$$

Part 12: regularity of the $k_{ij}k'_{ij}$ term
Just as for the S_{ijr} term, we have

$$\big|k_{ij}(w_1)k_{ij}(w'_1) - k_{ij}(w_2)k_{ij}(w'_2)\big| \lesssim \frac{1}{h^3}\Big(|w_1 - w_2| + |w'_1 - w'_2|\Big).$$

Part 13: uniform concentration of the $k_{ij}k'_{ij}$ term
By setting $\delta_n = h^3\sqrt{\log n/(nh^2)}$, the fluctuations can be at most $\sqrt{\log n/(nh^2)}$, so we have for $t \ge 1$

$$\mathbb{P}\left(\sup_{w,w'\in\mathcal{W}}\left|\frac{2}{n(n-1)}\sum_{i<j}\Big(k_{ij}k'_{ij} - \mathbb{E}\big[k_{ij}k'_{ij}\big]\Big)\right| > t\sqrt{\frac{\log n}{n^2h^3}} + 2t\sqrt{\frac{\log n}{nh^2}}\right)$$
$$\le \mathbb{P}\left(\sup_{w,w'\in\mathcal{W}_\delta}\left|\frac{2}{n(n-1)}\sum_{i<j}\Big(k_{ij}k'_{ij} - \mathbb{E}\big[k_{ij}k'_{ij}\big]\Big)\right| > t\sqrt{\frac{\log n}{n^2h^3}} + t\sqrt{\frac{\log n}{nh^2}}\right)$$
$$+ \mathbb{P}\left(\sup_{|w_1-w_2|\le\delta_n}\sup_{|w'_1-w'_2|\le\delta_n}\big|k_{ij}(w_1)k_{ij}(w'_1) - k_{ij}(w_2)k_{ij}(w'_2)\big| > t\sqrt{\frac{\log n}{nh^2}}\right)$$
$$\lesssim \delta_n^{-2}n^{-t^2/C_6} + \delta_n^{-2}n^{-t^2/C_7} \lesssim n^{1-t^2/C_6}h^{-4} + n^{1-t^2/C_7}h^{-4} \lesssim n^{5-t^2/C_8},$$

where $C_8 > 0$ is a constant and in the last line we use $1/h \lesssim n$. This converges to zero for any sufficiently large t, so by Lemma SA6 we have

$$\sup_{w,w'\in\mathcal{W}}\left|\frac{2}{n(n-1)}\sum_{i<j}\frac{k_{ij}k'_{ij} - \mathbb{E}\big[k_{ij}k'_{ij}\big]}{\sqrt{\Sigma_n(w,w) + \Sigma_n(w',w')}}\right| \lesssim_{\mathbb{P}} \left(\sqrt{\frac{\log n}{n^2h^3}} + \sqrt{\frac{\log n}{nh^2}}\right)\sqrt{n^2h} \lesssim_{\mathbb{P}} \sqrt{\frac{n\log n}{h}}.$$

Part 14: conclusion
By the uniform bounds in probability derived in the previous parts, and with $nh \gtrsim \log n$, we conclude that

$$\begin{aligned}\sup_{w,w'\in\mathcal{W}}\left|\frac{\widehat{\Sigma}_n(w,w')-\Sigma_n(w,w')}{\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}}\right| &\leq \frac{2}{n(n-1)}\sup_{w,w'\in\mathcal{W}}\left|\frac{2}{n(n-1)}\sum_{i<j}\frac{k_{ij}k'_{ij}-\mathbb{E}\big[k_{ij}k'_{ij}\big]}{\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}}\right| \\ &\quad+\frac{4(n-2)}{n(n-1)}\sup_{w,w'\in\mathcal{W}}\left|\frac{6}{n(n-1)(n-2)}\sum_{i<j<r}\frac{S_{ijr}-\mathbb{E}\big[k_{ij}k'_{ir}\big]}{\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}}\right| \\ &\quad+\frac{4n-6}{n(n-1)}\sup_{w,w'\in\mathcal{W}}\left|\frac{\widehat{f}_W(w)\widehat{f}_W(w')-\mathbb{E}[k_{ij}]\mathbb{E}[k'_{ij}]}{\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}}\right| \\ &\lesssim_{\mathbb{P}}\sqrt{\frac{\log n}{n^3h}}+\frac{\sqrt{\log n}}{n}+\frac{\sqrt{\log n}}{n}\lesssim_{\mathbb{P}}\frac{\sqrt{\log n}}{n}.\end{aligned}$$

□

Proof (Lemma SA12)
Since there is no ambiguity, we may understand k_{ij} to mean $k_h(W_{ij},w)$ when $i<j$ and to mean $k_h(W_{ji},w)$ when $j<i$. We use a prime to denote evaluation at w' rather than w. In this notation we may write

$$S_i(w)=\frac{1}{n-1}\sum_{j\neq i}k_{ij}$$

Let $\sum_{i\neq j\neq r}$ indicate that all the indices are distinct. Then

$$\begin{aligned}\frac{4}{n^2}\sum_{i=1}^n S_i(w)S_i(w') &= \frac{4}{n^2}\sum_{i=1}^n\frac{1}{n-1}\sum_{j\neq i}k_{ij}\frac{1}{n-1}\sum_{r\neq i}k'_{ir}=\frac{4}{n^2(n-1)^2}\sum_{i=1}^n\sum_{j\neq i}\sum_{r\neq i}k_{ij}k'_{ir} \\ &= \frac{4}{n^2(n-1)^2}\sum_{i=1}^n\sum_{j\neq i}\left(\sum_{r\neq i,r\neq j}k_{ij}k'_{ir}+k_{ij}k'_{ij}\right) \\ &= \frac{4}{n^2(n-1)^2}\sum_{i\neq j\neq r}k_{ij}k'_{ir}+\frac{4}{n^2(n-1)^2}\sum_{i\neq j}k_{ij}k'_{ij} \\ &= \frac{24}{n^2(n-1)^2}\sum_{i<j<r}S_{ijr}(w,w')+\frac{8}{n^2(n-1)^2}\sum_{i<j}k_{ij}k'_{ij} \\ &= \widehat{\Sigma}_n(w,w')+\frac{4}{n^2(n-1)^2}\sum_{i<j}k_{ij}k'_{ir}+\frac{4n-6}{n(n-1)}\widehat{f}\widehat{f}',\end{aligned}$$

and the result follows. □

SA5.3.8 Positive semi-definite covariance estimation

Proof (Lemma SA13)
Firstly we prove that the true covariance function Σ_n is feasible for the optimization problem (1) in the sense that it satisfies the constraints. Clearly as a covariance function, Σ_n is symmetric and positive semi-definite. The Lipschitz constraint is satisfied because as established in the proof of Theorem SA4,

$$\big|\Sigma_n(w,w')-\Sigma_n(w,w'')\big|\leq\frac{4}{nh^3}C_{\mathrm{k}}C_{\mathrm{L}}|w'-w''|$$

for all $w,w',w''\in\mathcal{W}$. Denote the (random) objective function in the optimization problem (1) by

$$\mathrm{obj}(M)=\sup_{w,w'\in\mathcal{W}}\left|\frac{M(w,w')-\widehat{\Sigma}_n(w,w')}{\sqrt{\widehat{\Sigma}_n(w,w)+\widehat{\Sigma}_n(w',w')}}\right|.$$

By Lemma SA11 with $w = w'$ we deduce that $\sup_{w\in\mathcal{W}} \left|\frac{\widehat{\Sigma}_n(w,w)}{\Sigma_n(w,w)} - 1\right| \lesssim_{\mathbb{P}} \sqrt{h\log n}$ and so

$$\begin{aligned}
\mathrm{obj}(\Sigma_n) &= \sup_{w,w'\in\mathcal{W}} \left|\frac{\widehat{\Sigma}_n(w,w') - \Sigma_n}{\sqrt{\Sigma_n(w,w) + \Sigma_n(w',w')}}\right| \sqrt{\frac{\Sigma_n(w,w) + \Sigma_n(w',w')}{\widehat{\Sigma}_n(w,w) + \widehat{\Sigma}_n(w',w')}} \\
&\lesssim_{\mathbb{P}} \frac{\sqrt{\log n}}{n} \left(1 - \frac{\left|\widehat{\Sigma}_n(w,w) - \Sigma_n(w,w)\right|}{\Sigma_n(w,w)} - \frac{\left|\widehat{\Sigma}_n(w',w') - \Sigma_n(w',w')\right|}{\Sigma_n(w',w')}\right)^{-1/2} \\
&\lesssim_{\mathbb{P}} \frac{\sqrt{\log n}}{n} \left(1 - \sqrt{h\log n}\right)^{-1/2} \lesssim_{\mathbb{P}} \frac{\sqrt{\log n}}{n}.
\end{aligned}$$

Since the objective function is non-negative and because we have established at least one feasible function M with an almost surely finite objective value, we can conclude the following. Let $\mathrm{obj}^* = \inf_M \mathrm{obj}(M)$, where the infimum is over feasible functions M. Then for all $\varepsilon > 0$ there exists a feasible function M_ε with $\mathrm{obj}(M_\varepsilon) \le \mathrm{obj}^* + \varepsilon$, and we call such a solution ε-optimal. Let $\widehat{\Sigma}_n^+$ be an n^{-1}-optimal solution. Then

$$\mathrm{obj}(\widehat{\Sigma}_n^+) \le \mathrm{obj}^* + n^{-1} \le \mathrm{obj}(\Sigma_n) + n^{-1}.$$

Thus by the triangle inequality,

$$\sup_{w,w'\in\mathcal{W}} \left|\frac{\widehat{\Sigma}_n^+(w,w') - \Sigma_n(w,w')}{\sqrt{\Sigma_n(w,w) + \Sigma_n(w',w')}}\right| \le \mathrm{obj}(\widehat{\Sigma}_n^+) + \mathrm{obj}(\Sigma_n) \le 2\,\mathrm{obj}(\Sigma_n) + n^{-1} \lesssim_{\mathbb{P}} \frac{\sqrt{\log n}}{n}.$$

□

Proof (Lemma SA14)
Since $\widehat{\Sigma}_n^+$ is positive semi-definite, we must have $\widehat{\Sigma}_n^+(w,w) \ge 0$. Now Lemma SA13 implies that for all $\varepsilon \in (0,1)$ there exists a C_ε such that

$$\begin{aligned}
&\mathbb{P}\left(\Sigma_n(w,w) - C_\varepsilon \frac{\sqrt{\log n}}{n}\sqrt{\Sigma_n(w,w)} \le \widehat{\Sigma}_n^+(w,w) \le \Sigma_n(w,w) + C_\varepsilon \frac{\sqrt{\log n}}{n}\sqrt{\Sigma_n(w,w)} \text{ for all } w \in \mathcal{W}\right) \\
&\quad \ge 1 - \varepsilon.
\end{aligned}$$

Consider the function $g_a(t) = t - a\sqrt{t}$ and note that it is increasing on $\{t \ge a^2/4\}$. Applying this with $t = \Sigma_n(w,w)$ and $a = \frac{\sqrt{\log n}}{n}$, noting that by Lemma SA6 we have $t = \Sigma_n(w,w) \gtrsim \frac{1}{n^2 h} \gg \frac{\log n}{4n^2} = a^2/4$, shows that for n large enough,

$$\begin{aligned}
\inf_{w\in\mathcal{W}} \Sigma_n(w,w) - \frac{\sqrt{\log n}}{n}\sqrt{\inf_{w\in\mathcal{W}} \Sigma_n(w,w)} &\lesssim_{\mathbb{P}} \inf_{w\in\mathcal{W}} \widehat{\Sigma}_n^+(w,w), \\
\sup_{w\in\mathcal{W}} \widehat{\Sigma}_n^+(w,w) &\lesssim_{\mathbb{P}} \sup_{w\in\mathcal{W}} \Sigma_n(w,w) + \frac{\sqrt{\log n}}{n}\sqrt{\sup_{w\in\mathcal{W}} \Sigma_n(w,w)}.
\end{aligned}$$

Applying the bounds from Lemma SA6 yields

$$\begin{aligned}
\frac{D_{\mathrm{lo}}^2}{n} + \frac{1}{n^2 h} - \frac{\sqrt{\log n}}{n}\left(\frac{D_{\mathrm{lo}}}{\sqrt{n}} + \frac{1}{\sqrt{n^2 h}}\right) &\lesssim_{\mathbb{P}} \inf_{w\in\mathcal{W}} \widehat{\Sigma}_n^+(w,w), \\
\sup_{w\in\mathcal{W}} \widehat{\Sigma}_n^+(w,w) &\lesssim_{\mathbb{P}} \frac{D_{\mathrm{up}}^2}{n} + \frac{1}{n^2 h} + \frac{\sqrt{\log n}}{n}\left(\frac{D_{\mathrm{up}}}{\sqrt{n}} + \frac{1}{\sqrt{n^2 h}}\right)
\end{aligned}$$

and so

$$\frac{D_{\mathrm{lo}}^2}{n} + \frac{1}{n^2 h} \lesssim_{\mathbb{P}} \inf_{w\in\mathcal{W}} \widehat{\Sigma}_n^+(w,w) \le \sup_{w\in\mathcal{W}} \widehat{\Sigma}_n^+(w,w) \lesssim_{\mathbb{P}} \frac{D_{\mathrm{up}}^2}{n} + \frac{1}{n^2 h}.$$

□

SA5.3.9 Feasible uniform confidence bands

Proof (Lemma SA15)

$$\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w)-T_n(w)\right| = \sup_{w\in\mathcal{W}}\left\{\left|\widehat{f}_W(w)-f_W(w)\right|\cdot\left|\frac{1}{\sqrt{\widehat{\Sigma}_n^+(w,w)}}-\frac{1}{\sqrt{\Sigma_n(w,w)}}\right|\right\}$$
$$\leq \sup_{w\in\mathcal{W}}\left|\frac{\widehat{f}_W(w)-\mathbb{E}\big[\widehat{f}_W(w)\big]}{\sqrt{\Sigma_n(w,w)}}+\frac{\mathbb{E}\big[\widehat{f}_W(w)\big]-f_W(w)}{\sqrt{\Sigma_n(w,w)}}\right|\cdot\sup_{w\in\mathcal{W}}\left|\frac{\widehat{\Sigma}_n^+(w,w)-\Sigma_n(w,w)}{\sqrt{\Sigma_n(w,w)\widehat{\Sigma}_n^+(w,w)}}\right|$$

Now from the proof of Lemma SA11 we have that $\sup_{w\in\mathcal{W}}\left|\frac{\widehat{f}_W(w)-\mathbb{E}\left[\widehat{f}_W(w)\right]}{\sqrt{\Sigma_n(w,w)}}\right|\lesssim_{\mathbb{P}}\sqrt{\log n}$, while Lemma SA1 gives $\sup_{w\in\mathcal{W}}\left|\mathbb{E}\big[\widehat{f}_W(w)\big]-f_W(w)\right|\lesssim h^{p\wedge\beta}$. By Lemma SA6 we have $\sup_{w\in\mathcal{W}}\Sigma_n(w,w)^{-1/2}\lesssim\frac{1}{D_{\text{lo}}/\sqrt{n}+1/\sqrt{n^2h}}$, and similarly Lemma SA14 gives $\sup_{w\in\mathcal{W}}\widehat{\Sigma}_n^+(w,w)^{-1/2}\lesssim_{\mathbb{P}}\frac{1}{D_{\text{lo}}/\sqrt{n}+1/\sqrt{n^2h}}$, Thus, applying Lemma SA13 to control the covariance estimation error,

$$\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w)-T_n(w)\right|\lesssim_{\mathbb{P}}\left(\sqrt{\log n}+\frac{h^{p\wedge\beta}}{D_{\text{lo}}/\sqrt{n}+1/\sqrt{n^2h}}\right)\frac{\sqrt{\log n}}{n}\frac{1}{D_{\text{lo}}/\sqrt{n}+1/\sqrt{n^2h}}$$
$$\lesssim_{\mathbb{P}}\sqrt{\frac{\log n}{n}}\left(\sqrt{\log n}+\frac{\sqrt{n}h^{p\wedge\beta}}{D_{\text{lo}}+1/\sqrt{nh}}\right)\frac{1}{D_{\text{lo}}+1/\sqrt{nh}}.$$

□

Proof (Lemma SA16)

Firstly note that the process $\widehat{Z}_n^T$ exists by noting that $\widehat{\Sigma}_n^+(w,w')$ and therefore also $\frac{\widehat{\Sigma}_n^+(w,w')}{\sqrt{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')}}$ are positive semi-definite functions and appealing to the Kolmogorov consistency theorem (Giné and Nickl, 2021). To obtain the desired Kolmogorov–Smirnov result we discretize and use the Gaussian–Gaussian comparison result found in Lemma 3.1 in Chernozhukov et al. (2013).

Part 1: bounding the covariance discrepancy

Define the maximum discrepancy in the (conditional) covariances of $\widehat{Z}_n^T$ and Z_n^T by

$$\Delta := \sup_{w,w'\in\mathcal{W}}\left|\frac{\widehat{\Sigma}_n^+(w,w')}{\sqrt{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')}}-\frac{\Sigma_n(w,w')}{\sqrt{\Sigma_n(w,w)\Sigma_n(w',w')}}\right|.$$

This random variable can be bounded in probability in the following manner. First note that by the Cauchy–Schwarz inequality for covariances, $|\Sigma_n(w,w')|\leq\sqrt{\Sigma_n(w,w)\Sigma_n(w',w')}$. Hence

$$\Delta\leq\sup_{w,w'\in\mathcal{W}}\left|\frac{\widehat{\Sigma}_n^+(w,w')-\Sigma_n(w,w')}{\sqrt{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')}}\right|+\sup_{w,w'\in\mathcal{W}}\left|\frac{\sqrt{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')}-\sqrt{\Sigma_n(w,w)\Sigma_n(w',w')}}{\sqrt{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')}}\right|$$
$$\leq\sup_{w,w'\in\mathcal{W}}\left\{\sqrt{\frac{\Sigma_n(w,w)+\Sigma_n(w',w')}{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')}}\left|\frac{\widehat{\Sigma}_n^+(w,w')-\Sigma_n(w,w')}{\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}}\right|\right\}$$
$$+\sup_{w,w'\in\mathcal{W}}\left|\frac{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')-\Sigma_n(w,w)\Sigma_n(w',w')}{\sqrt{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')\Sigma_n(w,w)\Sigma_n(w',w')}}\right|.$$

For the first term note that $\inf_{w\in\mathcal{W}} \widehat{\Sigma}_n^+(w,w) \gtrsim \frac{D_{\text{lo}}^2}{n} + \frac{1}{n^2 h}$ by Lemma SA14 and $\sup_{w\in\mathcal{W}} \left|\frac{\widehat{\Sigma}_n(w,w)}{\Sigma_n(w,w)} - 1\right| \lesssim_{\mathbb{P}} \sqrt{h\log n}$ by the proof of Lemma SA13. Thus by Lemma SA13,

$$\sup_{w,w'\in\mathcal{W}} \left\{ \sqrt{\frac{\Sigma_n(w,w)+\Sigma_n(w',w')}{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')}} \left| \frac{\widehat{\Sigma}_n^+(w,w') - \Sigma_n(w,w')}{\sqrt{\Sigma_n(w,w)+\Sigma_n(w',w')}} \right| \right\} \lesssim_{\mathbb{P}} \frac{\sqrt{\log n}}{n} \frac{1}{D_{\text{lo}}/\sqrt{n} + 1/\sqrt{n^2 h}}$$
$$\lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{n}} \frac{1}{D_{\text{lo}} + 1/\sqrt{nh}}.$$

For the second term, we have by the same bounds

$$\sup_{w,w'\in\mathcal{W}} \left| \frac{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w') - \Sigma_n(w,w)\Sigma_n(w',w')}{\sqrt{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')\Sigma_n(w,w)\Sigma_n(w',w')}} \right|$$
$$\leq \sup_{w,w'\in\mathcal{W}} \left\{ \frac{\left|\widehat{\Sigma}_n^+(w,w) - \Sigma_n(w,w)\right|\widehat{\Sigma}_n^+(w',w')}{\sqrt{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')\Sigma_n(w,w)\Sigma_n(w',w')}} + \frac{\left|\widehat{\Sigma}_n^+(w',w') - \Sigma_n(w',w')\right|\Sigma_n(w,w)}{\sqrt{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')\Sigma_n(w,w)\Sigma_n(w',w')}} \right\}$$
$$\leq \sup_{w,w'\in\mathcal{W}} \left\{ \frac{\left|\widehat{\Sigma}_n^+(w,w) - \Sigma_n(w,w)\right|}{\sqrt{\Sigma_n(w,w)}} \frac{\sqrt{\widehat{\Sigma}_n^+(w',w')}}{\sqrt{\widehat{\Sigma}_n^+(w,w)\Sigma_n(w',w')}} \right\}$$
$$+ \sup_{w,w'\in\mathcal{W}} \left\{ \frac{\left|\widehat{\Sigma}_n^+(w',w') - \Sigma_n(w',w')\right|}{\sqrt{\Sigma_n(w',w')}} \frac{\sqrt{\Sigma_n(w,w)}}{\sqrt{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')}} \right\}$$
$$\lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{n}} \frac{1}{D_{\text{lo}} + 1/\sqrt{nh}}.$$

Therefore

$$\Delta \lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{n}} \frac{1}{D_{\text{lo}} + 1/\sqrt{nh}}.$$

Part 2: Gaussian comparison on a mesh

Let $\mathcal{W}_\delta$ be a δ_n-covering of $\mathcal{W}$ with cardinality $O(1/\delta_n)$, where $1/\delta_n$ is at most polynomial in n. The scaled (conditionally) Gaussian processes Z_n^T and $\widehat{Z}_n^T$ both have pointwise (conditional) variances of 1. Therefore by Lemma 3.1 in Chernozhukov et al. (2013),

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\sup_{w\in\mathcal{W}_\delta} Z_n^T(w) \leq t \right) - \mathbb{P}\left(\sup_{w\in\mathcal{W}_\delta} \widehat{Z}_n^T(w) \leq t \;\middle|\; \mathbf{W}_n \right) \right| \lesssim \Delta^{1/3} \left(1 \vee \log \frac{1}{\Delta\delta_n}\right)^{2/3}$$

uniformly in the data. By the previous part and since $x(\log 1/x)^2$ is increasing on $\left(0, e^{-2}\right)$,

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\sup_{w\in\mathcal{W}_\delta} Z_n^T(w) \leq t \right) - \mathbb{P}\left(\sup_{w\in\mathcal{W}_\delta} \widehat{Z}_n^T(w) \leq t \;\middle|\; \mathbf{W}_n \right) \right| \lesssim_{\mathbb{P}} \left(\sqrt{\frac{\log n}{n}} \frac{1}{D_{\text{lo}} + 1/\sqrt{nh}} \right)^{1/3} (\log n)^{2/3}$$
$$\lesssim_{\mathbb{P}} \frac{n^{-1/6}(\log n)^{5/6}}{D_{\text{lo}}^{1/3} + (nh)^{-1/6}}.$$

Part 3: trajectory regularity of Z_n^T

During the proof of Theorem SA4 we established that Z_n^T satisfies the following trajectory regularity property:

$$\mathbb{E}\left[\sup_{|w-w'|\leq\delta_n} \left| Z_n^T(w) - Z_n^T(w') \right| \right] \lesssim nh^{-1}\sqrt{\delta_n \log n},$$

whenever $1/\delta_n$ is at most polynomial in n.

Part 4: conditional L^2 regularity of $\widehat{Z}_n^T$
By Lemma SA13, with $nh \gtrsim \log n$, we have uniformly in w, w',

$$\left|\widehat{\Sigma}_n^+(w,w') - \widehat{\Sigma}_n^+(w,w)\right| \lesssim n^{-1}h^{-3}|w-w'|.$$

Taking $\delta_n \le n^{-2}h^2$, Lemma SA14 gives

$$\inf_{|w-w'|\le\delta_n} \widehat{\Sigma}_n^+(w,w') \gtrsim \frac{D_{\mathrm{lo}}^2}{n} + \frac{1}{n^2h} - n^{-1}h^{-3}\delta_n \gtrsim \frac{D_{\mathrm{lo}}^2}{n} + \frac{1}{n^2h} - \frac{1}{n^3h} \gtrsim \frac{D_{\mathrm{lo}}^2}{n} + \frac{1}{n^2h}.$$

The conditional L^2 regularity of $\widehat{Z}_n^T$ is

$$\mathbb{E}\left[\left(\widehat{Z}_n^T(w) - \widehat{Z}_n^T(w')\right)^2 \mid \mathbf{W}_n\right] = 2 - 2\frac{\widehat{\Sigma}_n^+(w,w')}{\sqrt{\widehat{\Sigma}_n^+(w,w)\widehat{\Sigma}_n^+(w',w')}}.$$

Applying the same elementary result as for Z_n^T in the proof of Theorem SA4 yields

$$\mathbb{E}\left[\left(\widehat{Z}_n^T(w) - \widehat{Z}_n^T(w')\right)^2 \mid \mathbf{W}_n\right] \lesssim_{\mathbb{P}} n^2h^{-2}|w-w'|.$$

Thus the conditional semimetric induced by $\widehat{Z}_n^T$ on $\mathcal{W}$ is

$$\widehat{\rho}(w,w') := \mathbb{E}\left[\left(\widehat{Z}_n^T(w) - \widehat{Z}_n^T(w')\right)^2 \mid \mathbf{W}_n\right]^{1/2} \lesssim_{\mathbb{P}} nh^{-1}\sqrt{|w-w'|}.$$

Part 5: conditional trajectory regularity of $\widehat{Z}_n^T$
Just as for Z_n^T in the proof of Theorem SA4 we apply Lemma SA32, this time conditionally, to obtain that also

$$\mathbb{E}\left[\sup_{|w-w'|\le\delta_n}\left|\widehat{Z}_n^T(w) - \widehat{Z}_n^T(w')\right| \,\middle|\, \mathbf{W}_n\right] \lesssim_{\mathbb{P}} nh^{-1}\sqrt{\delta_n \log n},$$

whenever $1/\delta_n$ is at most polynomial in n.

Part 6: uniform Gaussian comparison
Now we use the trajectory regularity properties to extend the Gaussian–Gaussian comparison result from a finite mesh to all of $\mathcal{W}$. Write the previously established approximation rate as

$$r_n = \frac{n^{-1/6}(\log n)^{5/6}}{D_{\mathrm{lo}}^{1/3} + (nh)^{-1/6}}.$$

Take $\varepsilon_n > 0$ and observe that uniformly in $t \in \mathbb{R}$,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{Z}_n^T(w)\right| \le t \,\middle|\, \mathbf{W}_n\right) \\
&\quad\le \mathbb{P}\left(\sup_{w\in\mathcal{W}_\delta}\left|\widehat{Z}_n^T(w)\right| \le t+\varepsilon_n \,\middle|\, \mathbf{W}_n\right) + \mathbb{P}\left(\sup_{|w-w'|\le\delta_n}\left|\widehat{Z}_n^T(w) - \widehat{Z}_n^T(w')\right| \ge \varepsilon_n \,\middle|\, \mathbf{W}_n\right) \\
&\quad\le \mathbb{P}\left(\sup_{w\in\mathcal{W}_\delta}\left|Z_n^T(w)\right| \le t+\varepsilon_n\right) + O_{\mathbb{P}}(r_n) + \mathbb{P}\left(\sup_{|w-w'|\le\delta_n}\left|\widehat{Z}_n^T(w) - \widehat{Z}_n^T(w')\right| \ge \varepsilon_n \,\middle|\, \mathbf{W}_n\right) \\
&\quad\le \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right| \le t+2\varepsilon_n\right) + O_{\mathbb{P}}(r_n) \\
&\qquad + \mathbb{P}\left(\sup_{|w-w'|\le\delta_n}\left|Z_n^T(w) - Z_n^T(w')\right| \ge \varepsilon_n\right) + \mathbb{P}\left(\sup_{|w-w'|\le\delta_n}\left|\widehat{Z}_n^T(w) - \widehat{Z}_n^T(w')\right| \ge \varepsilon_n \,\middle|\, \mathbf{W}_n\right) \\
&\quad\le \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right| \le t+2\varepsilon_n\right) + O_{\mathbb{P}}(r_n) + O_{\mathbb{P}}(\varepsilon_n^{-1}nh^{-1}\sqrt{\delta_n\log n}) \\
&\quad\le \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right| \le t\right) + \mathbb{P}\left(\left|\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right| - t\right| \le 2\varepsilon_n\right) + O_{\mathbb{P}}(r_n) + O_{\mathbb{P}}(\varepsilon_n^{-1}nh^{-1}\sqrt{\delta_n\log n}).
\end{aligned}$$

The converse inequality is obtained analogously as follows:

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{Z}_n^T(w)\right|\le t \;\middle|\; \mathbf{W}_n\right)\\
&\quad\ge \mathbb{P}\left(\sup_{w\in\mathcal{W}_\delta}\left|\widehat{Z}_n^T(w)\right|\le t-\varepsilon_n \;\middle|\; \mathbf{W}_n\right)-\mathbb{P}\left(\sup_{|w-w'|\le\delta_n}\left|\widehat{Z}_n^T(w)-\widehat{Z}_n^T(w')\right|\ge\varepsilon_n \;\middle|\; \mathbf{W}_n\right)\\
&\quad\ge \mathbb{P}\left(\sup_{w\in\mathcal{W}_\delta}\left|Z_n^T(w)\right|\le t-\varepsilon_n\right)-O_{\mathbb{P}}(r_n)-\mathbb{P}\left(\sup_{|w-w'|\le\delta_n}\left|\widehat{Z}_n^T(w)-\widehat{Z}_n^T(w')\right|\ge\varepsilon_n \;\middle|\; \mathbf{W}_n\right)\\
&\quad\ge \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right|\le t-2\varepsilon_n\right)-O_{\mathbb{P}}(r_n)\\
&\qquad-\mathbb{P}\left(\sup_{|w-w'|\le\delta_n}\left|Z_n^T(w)-Z_n^T(w')\right|\ge\varepsilon_n\right)-\mathbb{P}\left(\sup_{|w-w'|\le\delta_n}\left|\widehat{Z}_n^T(w)-\widehat{Z}_n^T(w')\right|\ge\varepsilon_n \;\middle|\; \mathbf{W}_n\right)\\
&\quad\ge \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right|\le t-2\varepsilon_n\right)-O_{\mathbb{P}}(r_n)-O_{\mathbb{P}}(\varepsilon_n^{-1}nh^{-1}\sqrt{\delta_n\log n})\\
&\quad\ge \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right|\le t\right)-\mathbb{P}\left(\left|\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right|-t\right|\le 2\varepsilon_n\right)-O_{\mathbb{P}}(r_n)-O_{\mathbb{P}}(\varepsilon_n^{-1}nh^{-1}\sqrt{\delta_n\log n}).
\end{aligned}
$$

Combining these uniform upper and lower bounds gives

$$
\begin{aligned}
&\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{Z}_n^T(w)\right|\le t \;\middle|\; \mathbf{W}_n\right)-\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right|\le t\right)\right|\\
&\qquad\lesssim_{\mathbb{P}}\sup_{t\in\mathbb{R}}\mathbb{P}\left(\left|\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right|-t\right|\le 2\varepsilon_n\right)+r_n+\varepsilon_n^{-1}nh^{-1/2}\delta_n^{1/2}\sqrt{\log n}.
\end{aligned}
$$

To bound the remaining term, we apply the anti-concentration result for Z_n^T from the proof of Theorem SA4:

$$
\sup_{t\in\mathbb{R}}\mathbb{P}\left(\left|\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right|-t\right|\le\varepsilon\right)\lesssim\varepsilon\sqrt{\log n}.
$$

Therefore

$$
\begin{aligned}
&\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{Z}_n^T(w)\right|\le t \;\middle|\; \mathbf{W}_n\right)-\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right|\le t\right)\right|\\
&\qquad\lesssim_{\mathbb{P}}\varepsilon_n\sqrt{\log n}+r_n+\varepsilon_n^{-1}nh^{-1/2}\delta_n^{1/2}\sqrt{\log n}.
\end{aligned}
$$

Taking $\varepsilon=r_n/\sqrt{\log n}$ and then $\delta_n=n^{-2}hr_n^2\varepsilon_n^2/\log n$ yields

$$
\left|\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{Z}_n^T(w)\right|\le t \;\middle|\; \mathbf{W}_n\right)-\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right|\le t\right)\right|\lesssim_{\mathbb{P}} r_n=\frac{n^{-1/6}(\log n)^{5/6}}{D_{\text{lo}}^{1/3}+(nh)^{-1/6}}.
$$

□

Proof (Lemma SA17)
Part 1: Kolmogorov–Smirnov approximation
Let Z_n^T and $\widehat{Z}_n^T$ be defined as in the proof of Lemma SA16. Write

$$
r_n=\frac{n^{-1/6}(\log n)^{5/6}}{D_{\text{lo}}^{1/3}+(nh)^{-1/6}}
$$

for the rate of approximation from Lemma SA16. Then for any $\varepsilon_n > 0$ and uniformly in $t \in \mathbb{R}$:

$$\begin{aligned}
\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{Z}_n^T(w)\right| \le t \;\middle|\; \mathbf{W}_n\right) &\le \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right| \le t\right) + O_{\mathbb{P}}(r_n) \\
&\le \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right| \le t-\varepsilon_n\right) + \mathbb{P}\left(\left|\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right| - t\right| \le \varepsilon_n\right) + O_{\mathbb{P}}(r_n) \\
&\le \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w)\right| \le t\right) + \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w) - Z_n^T(w)\right| \ge \varepsilon_n\right) + \mathbb{P}\left(\left|\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right| - t\right| \le \varepsilon_n\right) + O_{\mathbb{P}}(r_n) \\
&\le \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w)\right| \le t\right) + \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w) - Z_n^T(w)\right| \ge \varepsilon_n\right) + \varepsilon_n\sqrt{\log n} + O_{\mathbb{P}}(r_n),
\end{aligned}$$

where in the last line we used the anti-concentration result from Lemma SA33 applied to Z_n^T, as in the proof of Lemma SA16. The corresponding lower bound is as follows:

$$\begin{aligned}
\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{Z}_n^T(w)\right| \le t \;\middle|\; \mathbf{W}_n\right) &\ge \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right| \le t\right) - O_{\mathbb{P}}(r_n) \\
&\ge \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right| \le t+\varepsilon_n\right) - \mathbb{P}\left(\left|\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right| - t\right| \le \varepsilon_n\right) - O_{\mathbb{P}}(r_n) \\
&\ge \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w)\right| \le t\right) - \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w) - Z_n^T(w)\right| \ge \varepsilon_n\right) - \mathbb{P}\left(\left|\sup_{w\in\mathcal{W}}\left|Z_n^T(w)\right| - t\right| \le \varepsilon_n\right) - O_{\mathbb{P}}(r_n) \\
&\ge \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w)\right| \le t\right) - \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w) - Z_n^T(w)\right| \ge \varepsilon_n\right) - \varepsilon_n\sqrt{\log n} - O_{\mathbb{P}}(r_n).
\end{aligned}$$

Part 2: t-statistic approximation
To control the remaining term, note that by Lemma SA10 and Lemma SA15,

$$\begin{aligned}
\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w) - Z_n^T(w)\right| &\le \sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w) - T_n(w)\right| + \sup_{w\in\mathcal{W}}\left|T_n(w) - Z_n^T(w)\right| \\
&\lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{n}}\left(\sqrt{\log n} + \frac{\sqrt{n}h^{p\wedge\beta}}{D_{\text{lo}} + 1/\sqrt{nh}}\right)\frac{1}{D_{\text{lo}} + 1/\sqrt{nh}} \\
&\quad + \frac{n^{-1/2}\log n + n^{-3/4}h^{-7/8}(\log n)^{3/8}R_n + n^{-2/3}h^{-1/2}(\log n)^{2/3} + n^{1/2}h^{p\wedge\beta}}{D_{\text{lo}} + 1/\sqrt{nh}}
\end{aligned}$$

and denote this last quantity by r_n'. Then for any $\varepsilon_n \gg r_n'$, we have

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w)\right| \le t\right) - \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{Z}_n^T(w)\right| \le t \;\middle|\; \mathbf{W}_n\right)\right| \lesssim_{\mathbb{P}} \varepsilon_n\sqrt{\log n} + r_n + o(1).$$

Part 3: rate analysis
This rate can be made $o_{\mathbb{P}}(1)$ by some appropriate choice of ε_n whenever $r_n \to 0$ and $r_n'\sqrt{\log n} \to 0$, by Lemma SA34, along with a sufficiently slowly diverging sequence R_n. Explicitly, we require the following.

$$\begin{aligned}
\frac{n^{-1/2}(\log n)^{3/2}}{D_{\text{lo}} + 1/\sqrt{nh}} &\to 0, & \frac{h^{p\wedge\beta}\log n}{D_{\text{lo}}^2 + (nh)^{-1}} &\to 0, \\
\frac{n^{-1/2}(\log n)^{3/2}}{D_{\text{lo}} + 1/\sqrt{nh}} &\to 0, & \frac{n^{-3/4}h^{-7/8}(\log n)^{7/8}}{D_{\text{lo}} + 1/\sqrt{nh}} &\to 0, \\
\frac{n^{-2/3}h^{-1/2}(\log n)^{7/6}}{D_{\text{lo}} + 1/\sqrt{nh}} &\to 0, & \frac{n^{1/2}h^{p\wedge\beta}(\log n)^{1/2}}{D_{\text{lo}} + 1/\sqrt{nh}} &\to 0, \\
\frac{n^{-1/6}(\log n)^{5/6}}{D_{\text{lo}}^{1/3} + (nh)^{-1/6}} &\to 0.
\end{aligned}$$

Using the fact that $h \lesssim n^{-\varepsilon}$ for some $\varepsilon > 0$ and removing trivial statements leaves us with

$$\frac{n^{-3/4}h^{-7/8}(\log n)^{7/8}}{D_{\text{lo}} + 1/\sqrt{nh}} \to 0, \qquad \frac{n^{1/2}h^{p\wedge\beta}(\log n)^{1/2}}{D_{\text{lo}} + 1/\sqrt{nh}} \to 0.$$

Now we analyze these based on the degeneracy type and verify that they hold under Assumption SA3.

(i) No degeneracy: if $D_{\text{lo}} > 0$ then we need

$$n^{-3/4}h^{-7/8}(\log n)^{7/8} \to 0, \qquad n^{1/2}h^{p\wedge\beta}(\log n)^{1/2} \to 0.$$

These reduce to $n^{-6/7}\log n \ll h \ll (n\log n)^{-\frac{1}{2(p\wedge\beta)}}$.

(ii) Partial or total degeneracy: if $D_{\text{lo}} = 0$ then we need

$$n^{-1/4}h^{-3/8}(\log n)^{7/8} \to 0, \qquad nh^{(p\wedge\beta)+1/2}(\log n)^{1/2} \to 0.$$

These reduce to $n^{-2/3}(\log n)^{7/3} \ll h \ll (n^2\log n)^{-\frac{1}{2(p\wedge\beta)+1}}$.

□

Proof (Theorem SA5)

Part 1: existence of the conditional quantile

We argue as in the proof of Lemma SA16, now also conditioning on the data. In particular, using the anti-concentration result from Lemma SA33, the regularity property of $\widehat{Z}_n^T$ and the Gaussian process maximal inequality from Lemma SA32, we see that for any $\varepsilon > 0$,

$$\begin{aligned}\sup_{t\in\mathbb{R}} \mathbb{P}\left(\left|\sup_{w\in\mathcal{W}}\left|\widehat{Z}_n^T(w)\right| - t\right| \le 2\varepsilon \;\middle|\; \mathbf{W}_n\right) &\le 8\varepsilon\left(1 + \mathbb{E}\left[\sup_{w\in\mathcal{W}}\left|\widehat{Z}_n^T(w)\right| \;\middle|\; \mathbf{W}_n\right]\right)\\ &\lesssim \varepsilon\sqrt{\log n}.\end{aligned}$$

Thus letting $\varepsilon \to 0$ shows that the conditional distribution function of $\sup_{w\in\mathcal{W}}\left|\widehat{Z}_n^T(w)\right|$ is continuous, and therefore all of its conditional quantiles exist.

Part 2: validity of the confidence band

Define the following (conditional) distribution functions.

$$F_Z(t \mid \mathbf{W}_n) = \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{Z}_n^T(w)\right| \le t \;\middle|\; \mathbf{W}_n\right), \qquad F_T(t) = \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w)\right| \le t\right),$$

along with their well-defined right-quantile functions,

$$F_Z^{-1}(p \mid \mathbf{W}_n) = \sup\left\{t \in \mathbb{R} \,:\, F_Z(t \mid \mathbf{W}_n) = p\right\}, \qquad F_T^{-1}(p) = \sup\left\{t \in \mathbb{R} \,:\, F_T(t) = p\right\}.$$

Note that $t \le F_Z^{-1}(p \mid \mathbf{W}_n)$ if and only if $F_Z(t \mid \mathbf{W}_n) \le p$. Take $\alpha \in (0,1)$ and define the quantile $\widehat{q}_{1-\alpha} = F_Z^{-1}(1-\alpha \mid \mathbf{W}_n)$, so that $F_Z(\widehat{q}_{1-\alpha} \mid \mathbf{W}_n) = 1-\alpha$. By Lemma SA17, we have that

$$\sup_{t\in\mathbb{R}}\left|F_Z(t \mid \mathbf{W}_n) - F_T(t)\right| = o_{\mathbb{P}}(1)$$

Thus by Lemma SA34, this can be replaced by

$$\mathbb{P}\left(\sup_{t\in\mathbb{R}}\left|F_Z(t \mid \mathbf{W}_n) - F_T(t)\right| > \varepsilon_n\right) \le \varepsilon_n$$

for some $\varepsilon_n \to 0$. Therefore

$$\begin{aligned}
\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w)\right| \le \widehat{q}_{1-\alpha}\right) &= \mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w)\right| \le F_Z^{-1}(1-\alpha \mid \mathbf{W}_n)\right) \\
&= \mathbb{P}\left(F_Z\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w)\right| \;\middle|\; \mathbf{W}_n\right) \le 1-\alpha\right) \\
&\le \mathbb{P}\left(F_T\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w)\right|\right) \le 1-\alpha+\varepsilon_n\right) + \varepsilon_n \le 1-\alpha+3\varepsilon_n,
\end{aligned}$$

where in the last line we used the fact that for any real-valued random variable X with distribution function F, we have $\left|\mathbb{P}\big(F(X)\le t\big) - t\right| \le \Delta$, where Δ is the size of the largest jump discontinuity in F. By taking an expectation and uniform integrability, $\sup_{t\in\mathbb{R}}\left|F_Z(t) - F_T(t)\right| = o(\varepsilon_n)$. Since F_Z has no jumps, we must have $\Delta \le \varepsilon_n$ for F_T. Finally a lower bound is constructed in an analogous manner, giving

$$\mathbb{P}\left(\sup_{w\in\mathcal{W}}\left|\widehat{T}_n(w)\right| \le \widehat{q}_{1-\alpha}\right) \ge 1-\alpha-3\varepsilon_n.$$

Here ends the proof of the theorem. □

SA5.3.10 Counterfactual dyadic density estimation

Proof (Lemma SA18)
Writing $k_{ij} = k_h(W_{ij}^1, w)$, $\psi_i = \psi(X_i^1)$, $\widehat{\psi}_i = \widehat{\psi}(X_i^1)$ and $\kappa_{ij} = \kappa(X_i^0, X_i^1, X_j^1)$,

$$\begin{aligned}
\mathbb{E}\big[\widehat{f}_W^{1\triangleright 0}(w)\big] &= \mathbb{E}\left[\frac{2}{n(n-1)}\sum_{i<j}\widehat{\psi}_i\widehat{\psi}_j k_{ij}\right] \\
&= \frac{2}{n(n-1)(n-2)}\sum_{i<j}\sum_{r\notin\{i,j\}}\mathbb{E}\left[k_{ij}\Big(\psi_i\psi_j + \psi_i\kappa_{rj} + \psi_j\kappa_{ri}\Big)\right] + O\left(\frac{1}{n}\right) \\
&= \mathbb{E}\left[k_{ij}\psi_i\psi_j\right] + O\left(\frac{1}{n}\right) = \mathbb{E}\big[\psi_i\psi_j\mathbb{E}\left[k_h(W_{ij}^1, w) \mid X_i^1, X_j^1\right]\big] + O\left(\frac{1}{n}\right) \\
&= \mathbb{E}\big[\psi_i\psi_j f_{W|XX}^1(w \mid X_i^1, X_j^1) + O_{\mathbb{P}}\left(h^{p\wedge\beta}\right)\big] + O\left(\frac{1}{n}\right) = f_W^{1\triangleright 0}(w) + O\left(h^{p\wedge\beta} + \frac{1}{n}\right)
\end{aligned}$$

uniformly in $w \in \mathcal{W}$, following the proof of Lemma SA1 by Hölder-continuity of $f_{W|XX}^1(\cdot \mid x_1, x_2)$. □

Proof (Lemma SA19)

$$\begin{aligned}
\widehat{f}_W^{1\triangleright 0}(w) &= \frac{2}{n(n-1)}\sum_{i<j}\widehat{\psi}_i\widehat{\psi}_j k_{ij} \\
&= \frac{2}{n(n-1)}\sum_{i<j}\left(\psi_i + \frac{1}{n}\sum_{r=1}^n\kappa_{ri}\right)\left(\psi_j + \frac{1}{n}\sum_{r=1}^n\kappa_{rj}\right)k_{ij} + O_{\mathbb{P}}\left(\frac{1}{n}\right) \\
&= \frac{2}{n(n-1)}\sum_{i<j}\psi_i\psi_j k_{ij} + \frac{2}{n(n-1)}\sum_{i<j}\psi_i\frac{1}{n}\sum_{r\notin\{i,j\}}^n\kappa_{rj}k_{ij} + \frac{2}{n(n-1)}\sum_{i<j}\psi_j\frac{1}{n}\sum_{r\notin\{i,j\}}^n\kappa_{ri}k_{ij} \\
&\quad + O_{\mathbb{P}}\left(\frac{1}{n}\right) \\
&= \frac{2}{n(n-1)(n-2)}\sum_{i<j}\sum_{r\notin\{i,j\}}k_{ij}\Big(\psi_i\psi_j + \psi_i\kappa_{rj} + \psi_j\kappa_{ri}\Big) + O_{\mathbb{P}}\left(\frac{1}{n}\right) \\
&= \frac{6}{n(n-1)(n-2)}\sum_{i<j<r}v_{ijr} + O_{\mathbb{P}}\left(\frac{1}{n}\right)
\end{aligned}$$

where

$$v_{ijr} = \frac{1}{3}k_{ij}\Big(\psi_i\psi_j + \psi_i\kappa_{rj} + \psi_j\kappa_{ri}\Big) + \frac{1}{3}k_{ir}\Big(\psi_i\psi_r + \psi_i\kappa_{jr} + \psi_r\kappa_{ji}\Big) + \frac{1}{3}k_{jr}\Big(\psi_j\psi_r + \psi_j\kappa_{ir} + \psi_r\kappa_{ij}\Big)$$

So by the Hoeffding decomposition for third-order U-statistics,

$$\begin{aligned}\widehat{f}_W^{1\triangleright 0}(w) &= u + \frac{3}{n}\sum_{i=1}^{n} u_i + \frac{6}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} u_{ij} + \frac{6}{n(n-1)(n-2)}\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{r=j+1}^{n} u_{ijr} \\ &\quad + \frac{6}{n(n-1)(n-2)}\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{r=j+1}^{n}\left(v_{ijr} - u_{ijr}\right) + O_{\mathbb{P}}\left(\frac{1}{n}\right) \\ &= \mathbb{E}\big[\widehat{f}_W^{1\triangleright 0}(w)\big] + L_n^{1\triangleright 0}(w) + Q_n^{1\triangleright 0}(w) + T_n^{1\triangleright 0}(w) + E_n^{1\triangleright 0}(w) + O_{\mathbb{P}}\left(\frac{1}{n}\right).\end{aligned}$$

Noting that ψ_i, κ_{ij} and $\mathbb{E}[k_{ij} \mid A_i^1, A_j^1]$ are all bounded and that $\mathbb{E}[k_{ij} \mid A_i^1, A_j^1]$ is Lipschitz in w, we deduce by Lemma SA37 and Proposition 2.3 of Arcones and Giné (1993) that $\sup_{w\in\mathcal{W}} |Q_n^{1\triangleright 0}(w) + T_n^{1\triangleright 0}(w)| \lesssim_{\mathbb{P}} \frac{1}{n}$. □

Proof (Lemma SA20)
By Lemma SA25 we have $\sup_{w\in\mathcal{W}} \left|L_n^{1\triangleright 0}(w)\right| \lesssim_{\mathbb{P}} \frac{1}{\sqrt{n}}$. Note that in the proof of Lemma SA19 the terms $v_{ijr} - u_{ijr}$ depend only on V_{ij}, V_{ir} and V_{jr} after conditioning on $\mathbf{A}_n^1$, $\mathbf{X}_n^0$ and $\mathbf{X}_n^1$. Thus $E_n^{1\triangleright 0}(w)$ is a degenerate second-order U-statistic and so $\sup_{w\in\mathcal{W}} \left|E_n^{1\triangleright 0}(w)\right| \lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{n^2 h}}$ by Lemma SA37. □

Proof (Lemma SA21)
Note that $L_n^{1\triangleright 0}(w) = \frac{3}{n}\sum_{i=1}^{n} l_i^{1\triangleright 0}(w)$ where $l_i^{1\triangleright 0}(w)$ depends only on A_i^1, X_i^0 and X_i^1. Let $\gamma : \mathcal{X}\times\mathcal{X} \to \{1, \ldots, |\mathcal{X}|^2\}$ be a bijection and define $\text{logistic}(x) = \frac{1}{1+e^{-x}}$. Let $\widetilde{A}_i = \text{logistic}(A_i^1) + \gamma(X_i^0, X_i^1)$ so that $A_i^1 = \text{logistic}^{-1}\left(\widetilde{A}_i - \lfloor \widetilde{A}_i \rfloor\right)$ and $(X_i^0, X_i^1) = \gamma^{-1}(\lfloor \widetilde{A}_i \rfloor)$. Thus $l_i^{1\triangleright 0}(w)$ is a bounded-variation function of $\widetilde{A}_i$, uniformly in w, and so as in Lemma SA7 we have that on an appropriately enlarged probability space,

$$\mathbb{E}\left[\sup_{w\in\mathcal{W}} \left|\sqrt{n}L_n^{1\triangleright 0}(w) - Z_n^{L,1\triangleright 0}(w)\right|\right] \lesssim \frac{\log n}{\sqrt{n}}$$

where $Z_n^{L,1\triangleright 0}$ is a mean-zero Gaussian process with the same covariance structure as $\sqrt{n}L_n^{1\triangleright 0}$. For $E_n^{1\triangleright 0}(w)$, we first construct a strong approximation conditional on $\mathbf{A}_n$ and $\mathbf{X}_n$ as in Lemma SA8 and deduce an unconditional strong approximation as in Lemma SA9 to see

$$\sup_{w\in\mathcal{W}} \left|\sqrt{n^2 h}E_n^{1\triangleright 0}(w) - Z_n^{E,1\triangleright 0}(w)\right| \lesssim_{\mathbb{P}} n^{-1/4}h^{-3/8}(\log n)^{3/8}R_n + n^{-1/6}(\log n)^{2/3}$$

where $Z_n^{E,1\triangleright 0}$ is a mean-zero Gaussian process with the same covariance structure as $\sqrt{n^2 h}E_n^{1\triangleright 0}$. Arguing as in the proof of Theorem SA3 shows that the Gaussian processes are independent and can be summed to yield a single strong approximation. □

Proof (Lemma SA22)

Arguing by mean-zero properties and conditional independence,

$$\begin{aligned}
\Sigma_n^{1\triangleright 0}(w,w') &= \operatorname{Cov}\left[\widehat{f}_W^{1\triangleright 0}(w), \widehat{f}_W^{1\triangleright 0}(w')\right] \\
&= \frac{1}{n^2(n-1)^2(n-2)^2}\sum_{i\neq j}\sum_{r\notin\{i,j\}}\sum_{i'\neq j'}\sum_{r'\notin\{i',j'\}} \mathbb{E}\Big[\Big(k_{ij}\psi_i\psi_j - \mathbb{E}[k_{ij}\psi_i\psi_j] + k_{ij}\psi_i\kappa_{rj} + k_{ij}\psi_j\kappa_{ri}\Big) \\
&\quad \times \Big(k'_{i'j'}\psi_{i'}\psi_{j'} - \mathbb{E}[k'_{ij}\psi_i\psi_j] + k'_{i'j'}\psi_{i'}\kappa_{r'j'} + k'_{i'j'}\psi_{j'}\kappa_{r'i'}\Big)\Big] + O\left(\frac{1}{n^{3/2}} + \frac{1}{\sqrt{n^4 h}}\right) \\
&= \frac{2}{n^2}\mathbb{E}\left[k_{ij}\psi_i\psi_j k'_{ij}\psi_i\psi_j\right] + \frac{4}{n}\mathbb{E}\left[k_{ij}\psi_i\psi_j k'_{ir}\psi_i\psi_r\right] - \frac{4}{n}\mathbb{E}\left[k_{ij}\psi_i\psi_j\right]\mathbb{E}\left[k'_{ij}\psi_i\psi_j\right] \\
&\quad + \frac{4}{n}\mathbb{E}\left[k_{ij}\psi_i\kappa_{i'j}k'_{i'j'}\psi_{i'}\psi_{j'}\right] + \frac{4}{n}\mathbb{E}\left[k_{ij}\psi_i\psi_j k'_{i'j'}\psi_{i'}\kappa_{ij}\right] + \frac{4}{n}\mathbb{E}\left[k_{ij}k'_{i'j'}\psi_i\psi_{i'}\kappa_{rj}\kappa_{rj'}\right] \\
&\quad + O\left(\frac{1}{n^{3/2}} + \frac{1}{\sqrt{n^4 h}}\right) \\
&= \frac{4}{n}\mathbb{E}\left[\Big(\psi_i\mathbb{E}\big[k_{ij}\psi_j \mid i\big] + \mathbb{E}\left[k_{rj}\psi_r\kappa_{ij} \mid i\right]\Big)\Big(\psi_i\mathbb{E}\big[k'_{ij}\psi_j \mid i\big] + \mathbb{E}\left[k'_{rj}\psi_r\kappa_{ij} \mid i\right]\Big)\right] \\
&\quad + \frac{2}{n^2}\mathbb{E}\left[k_{ij}k'_{ij}\psi_i^2\psi_j^2\right] - \frac{4}{n}\mathbb{E}\left[k_{ij}\psi_i\psi_j\right]\mathbb{E}\left[k'_{ij}\psi_i\psi_j\right] + O\left(\frac{1}{n^{3/2}} + \frac{1}{\sqrt{n^4 h}}\right),
\end{aligned}$$

where all indices are distinct. □

Proof (Lemma SA23)
The proof is exactly the same as the proof of Lemma SA10. □

Proof (Theorem SA6)
This proof proceeds in the same manner as the proof of Theorem SA4. □

References

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598.

Arcones, M. A. (1995). A Bernstein-type inequality for U-statistics and U-processes. *Statistics & Probability Letters*, 22(3):239–247.

Arcones, M. A. and Giné, E. (1993). Limit theorems for U-processes. *The Annals of Probability*, pages 1494–1542.

Belloni, A., Chernozhukov, V., Chetverikov, D., and Fernández-Val, I. (2019). Conditional quantile processes based on series or many regressors. *Journal of Econometrics*, 213(1):4–29.

Bhatia, R. (1997). *Matrix Analysis*, volume 169. Springer, New York, NY.

Chen, X. and Kato, K. (2020). Jackknife multiplier bootstrap: finite sample approximations to the U-process supremum with applications. *Probability Theory and Related Fields*, 176(3):1097–1163.

Chernozhukov, V., Chetverikov, D., Kato, K., et al. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819.

Chernozhukov, V., Chetverikov, D., Kato, K., et al. (2014a). Anti-concentration and honest, adaptive confidence bands. *The Annals of Statistics*, 42(5):1787–1818.

Chernozhukov, V., Chetverikov, D., Kato, K., et al. (2014b). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics*, 42(4):1564–1597.

de la Peña, V. H. and Montgomery-Smith, S. J. (1995). Decoupling inequalities for the tail probabilities of multivariate U-statistics. *The Annals of Probability*, 23(2):806–816.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). The distribution of wages, 1973–1992: A semiparametric approach. *Econometrica*, 64(5):1001–1004.

Dudley, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge Studies in Advanced Mathematics. Cambridge University Press.

Giné, E., Koltchinskii, V., and Sakhanenko, L. (2004). Kernel density estimators: convergence in distribution for weighted sup-norms. *Probability Theory and Related Fields*, 130(2):167–198.

Giné, E., Latała, R., and Zinn, J. (2000). Exponential and moment inequalities for U-statistics. In *High Dimensional Probability II*, pages 13–38. Birkhäuser, Boston, MA.

Giné, E. and Nickl, R. (2021). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Graham, B. S., Niu, F., and Powell, J. L. (2022). Kernel density estimation for undirected dyadic data. *Journal of Econometrics*. forthcoming.

Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ*.

Khasminskii, R. Z. (1978). A lower bound on the risks of nonparametric estimates of densities in the uniform metric. *Theory of Probability and its Applications*, 23(4):794–798.

Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent RVs, and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32(1-2):111–131.

Kwapien, S., Szulga, J., et al. (1991). Hypercontraction methods in moment inequalities for series of independent random variables in normed spaces. *The Annals of Probability*, 19(1):369–379.

Laurent, M. and Rendl, F. (2005). Semidefinite programming and integer programming. In *Discrete Optimization*, volume 12 of *Handbooks in Operations Research and Management Science*, pages 393–514. Elsevier.

Le Gall, J.-F. (2016). *Brownian Motion, Martingales, and Stochastic Calculus*, volume 274. Springer, Berlin, Heidelberg.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces*. Springer, Berlin, Heidelberg.

Minsker, S. and Wei, X. (2019). Moment inequalities for matrix-valued U-statistics of order 2. *Electronic Journal of Probability*, 24(133):1–32.

Royden, H. L. and Fitzpatrick, P. (1988). *Real Analysis*, volume 32. Macmillan, New York.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer, New York, NY.