Estimation and Inference in Boundary Discontinuity Designs: Distance-Based Methods
Supplemental Appendix

Matias D. Cattaneo* Rocio Titiunik† Ruiqi (Rae) Yu‡

October 29, 2025

Abstract

This supplemental appendix presents more general theoretical results encompassing those reported in the paper, their theoretical proofs, and other technical results. In particular, it presents a new strong approximation result for multiplicative-separable empirical processes leveraging and extending ideas from Cattaneo and Yu [2025].

*Department of Operations Research and Financial Engineering, Princeton University.
†Department of Politics, Princeton University.
‡Department of Operations Research and Financial Engineering, Princeton University.

Contents

SA-1 Setup

This supplemental appendix considers a generalized version of the problems studied in the main paper. Specifically, the underlying bivariate location variable $\mathbf{X}_i$ is d-dimensional ($d \geq 1$) with support $\mathscr{X} \subseteq \mathbb{R}^d$, and the boundary region $\mathscr{B}$ is a low dimensional manifold with "effective dimension" $d-1$. The results in the paper correspond to $d = 2$, that is, $\mathbf{X}_i$ is bivariate and $\mathscr{B}$ is a one-dimensional (boundary assignment) curve.

Assumption 1 in the paper generalizes as follows.

Assumption SA–1 (Data Generating Process)**.** *Let* $t \in \{0, 1\}$.

(i) $(Y_1(t), \mathbf{X}_1^\top)^\top, \ldots, (Y_n(t), \mathbf{X}_n^\top)^\top$ *are independent and identically distributed random vectors with* $\mathscr{X} = \prod_{l=1}^d [a_l, b_l]$ *for* $-\infty < a_l < b_l < \infty$ *for* $l = 1, \cdots, d$.

(ii) *The distribution of* $\mathbf{X}_i$ *has a Lebesgue density* $f_X(\mathbf{x})$ *that is continuous and bounded away from zero on* $\mathscr{X}$.

(iii) $\mu_t(\mathbf{x}) = \mathbb{E}[Y_i(t)|\mathbf{X}_i = \mathbf{x}]$ *is* $(p+1)$*-times continuously differentiable on* $\mathscr{X}$.

(iv) $\sigma_t^2(\mathbf{x}) = \mathbb{V}[Y_i(t)|\mathbf{X}_i = \mathbf{x}]$ *is bounded away from zero and continuous on* $\mathscr{X}$.

(v) $\sup_{\mathbf{x} \in \mathscr{X}} \mathbb{E}[|Y_i(t)|^{2+v}|\mathbf{X}_i = \mathbf{x}] < \infty$ *for some* $v \geq 2$.

The support $\mathscr{X}$ is partitioned into two (assignment) areas, $\mathscr{A}_0 \subset \mathbb{R}^d$ and $\mathscr{A}_1 \subset \mathbb{R}^d$, representing the control and treatment regions, respectively. Thus, $\mathscr{X} = \mathscr{A}_0 \cup \mathscr{A}_1$ with $\mathscr{A}_0$ and $\mathscr{A}_1$ disjoint regions in $\mathbb{R}^d$. The observed outcome is $Y_i = \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_0) Y_i(0) + \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1) Y_i(1)$, and $\mathscr{B} = \mathtt{bd}(\mathscr{A}_0) \cap \mathtt{bd}(\mathscr{A}_1)$ is the boundary determined by the assignment regions, where $\mathtt{bd}(\mathscr{A}_t)$ denotes the topological boundary of $\mathscr{A}_t$.

The *conditional treatment effect curve at the boundary* is

$$\tau(\mathbf{x}) = \mathbb{E}[Y_i(1) - Y_i(0)|\mathbf{X}_i = \mathbf{x}], \qquad \mathbf{x} \in \mathscr{B}.$$

The univariate distance score induced by the bivariate location variable is

$$D_i(\mathbf{x}) = [\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1) - \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_0)] \mathscr{d}(\mathbf{X}_i, \mathbf{x}), \qquad \mathbf{x} \in \mathscr{B},$$

where $\mathscr{d}(\cdot, \cdot)$ denotes a distance function. The distance-based treatment effect estimator process along the boundary based is $(\tau(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$ is

$$\Big(\widehat{\vartheta}(\mathbf{x}) = \widehat{\theta}_{1,\mathbf{x}}(0) - \widehat{\theta}_{0,\mathbf{x}}(0) : \mathbf{x} \in \mathscr{B}\Big),$$

where, for $t \in \{0, 1\}$,

$$\widehat{\theta}_{t,\mathbf{x}}(0) = \mathbf{e}_1^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x}), \qquad \widehat{\boldsymbol{\gamma}}_t(\mathbf{x}) = \underset{\boldsymbol{\gamma} \in \mathbb{R}^{p+1}}{\arg\min}\, \mathbb{E}_n\Big[\big(Y_i - \mathbf{r}_p(D_i(\mathbf{x}))^\top \boldsymbol{\gamma}\big)^2 K_h(D_i(\mathbf{x})) \mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x}))\Big],$$

$\mathbf{r}_p(u) = (1, u, \cdots, u^p)^\top$ and $K_h(u) = K(u/h)/h^2$ with $K(\cdot)$ a univariate kernel and h a bandwidth parameter, and $\mathscr{I}_0 = (-\infty, 0)$ and $\mathscr{I}_1 = [0, \infty)$. More generally, the least squares projection is

$$\widehat{\theta}_{t,\mathbf{x}}(D_i(\mathbf{x})) = \mathbf{r}_p(D_i(\mathbf{x}))^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x}), \qquad t \in \{0, 1\}, \quad \mathbf{x} \in \mathscr{B}.$$

We impose the following assumptions on the kernel function, distance function, and assignment boundary manifold. Let

$$\mathbf{\Psi}_{t,\mathbf{x}} = \mathbb{E}\left[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)^\top K_h(D_i(\mathbf{x}))\mathbb{1}(D_i(\mathbf{x}) \in \mathscr{I}_t)\right],$$

for $t \in \{0,1\}$.

Assumption SA–2 (Kernel, Distance, and Boundary)**.** *Let $t \in \{0,1\}$.*

(i) *$\mathscr{B}$ is compact $(d-1)$-rectifiable, with $\mathfrak{H}^{d-1}(\mathscr{B})$ positive and finite.*

(ii) *$\mathscr{d} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ is a metric on $\mathbb{R}^d$ equivalent to the Euclidean distance, that is, there exists positive constants C_u and C_l such that $C_l \|\mathbf{x}-\mathbf{x}'\| \leq \mathscr{d}(\mathbf{x},\mathbf{x}') \leq C_u \|\mathbf{x}-\mathbf{x}'\|$ for all $\mathbf{x},\mathbf{x}' \in \mathscr{X}$.*

(iii) *$K : \mathbb{R} \to [0,\infty)$ is compact supported and Lipschitz continuous, or $K(u) = \mathbb{1}(u \in [-1,1])$.*

(iv) $\liminf_{h\downarrow 0} \inf_{\mathbf{x}\in\mathscr{B}} \lambda_{\min}(\mathbf{\Psi}_{t,\mathbf{x}}) \gtrsim 1$.

For each $t \in \{0,1\}$, the induced conditional expectation based on univariate distance is

$$\theta_{t,\mathbf{x}}(r) = \mathbb{E}[Y_i|D_i(\mathbf{x}) = r] = \mathbb{E}[Y_i|\mathscr{d}(\mathbf{X}_i,\mathbf{x}) = |r|, \mathbf{X}_i \in \mathscr{A}_t], \qquad r \in \mathscr{I}_t, \quad \mathbf{x} \in \mathscr{B}.$$

More rigorously, for each $t \in \{0,1\}$, and letting $S_{t,\mathbf{x}}(r) = \{\mathbf{v} \in \mathscr{X} : \mathscr{d}(\mathbf{v},\mathbf{x}) = r, \mathbf{v} \in \mathscr{A}_t\}$ for $r \geq 0$ and $\mathbf{x} \in \mathscr{B}$,

$$\theta_{t,\mathbf{x}}(r) = \frac{\int_{S_{t,\mathbf{x}}(|r|)} \mu_t(\mathbf{v}) f_X(\mathbf{v}) \mathfrak{H}^{d-1}(d\mathbf{v})}{\int_{S_{t,\mathbf{x}}(|r|)} f_X(\mathbf{v}) \mathfrak{H}^{d-1}(d\mathbf{v})},$$

for $|r| > 0, \mathbf{x} \in \mathscr{B}, t \in \{0,1\}$, and therefore (under our assumptions)

$$\theta_{t,\mathbf{x}}(0) = \lim_{r\to 0} \mathbb{E}[Y_i|\mathscr{d}(\mathbf{X}_i,\mathbf{x}) = |r|, \mathbf{X}_i \in \mathscr{A}_t] = \lim_{r\to 0} \frac{\int_{S_{t,\mathbf{x}}(|r|)} \mu_t(\mathbf{v}) f_X(\mathbf{v}) \mathfrak{H}^{d-1}(d\mathbf{v})}{\int_{S_{t,\mathbf{x}}(|r|)} f_X(\mathbf{v}) \mathfrak{H}^{d-1}(d\mathbf{v})}.$$

Thus, the population limit based on the induced conditional expectations is $\theta_{\mathbf{x}}(0) = \theta_{1,\mathbf{x}}(0) - \theta_{0,\mathbf{x}}(0)$. Theorem SA-1 shows that $\theta_{\mathbf{x}}(0) = \tau(\mathbf{x})$ under Assumptions SA–1 and SA–2.

The best mean square approximation is

$$\theta^*_{t,\mathbf{x}}(D_i(\mathbf{x})) = \mathbf{r}_p(D_i(\mathbf{x}))^\top \boldsymbol{\gamma}^*_t(\mathbf{x}),$$

where

$$\boldsymbol{\gamma}^*_t(\mathbf{x}) = \operatorname*{arg\,min}_{\boldsymbol{\gamma}\in\mathbb{R}^{p+1}} \mathbb{E}\Big[\big(Y_i - \mathbf{r}_p(D_i(\mathbf{x}))^\top \boldsymbol{\gamma}\big)^2 K_h(D_i(\mathbf{x}))\mathbb{1}(D_i(\mathbf{x}) \in \mathscr{I}_t)\Big],$$

and uniqueness will follow from the results below. The estimation error decomposes into *linear error*, *approximation error*, and *non-linear error*: for all $t \in \{0,1\}$ and $\mathbf{x} \in \mathscr{B}$,

$$\begin{aligned}
\widehat{\theta}_{t,\mathbf{x}}(0) - \theta_{t,\mathbf{x}}(0) &= \mathbf{e}_1^\top \widehat{\mathbf{\Psi}}_{t,\mathbf{x}}^{-1} \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big) K_h(D_i(\mathbf{x})) Y_i\Big] - \theta_{t,\mathbf{x}}(0) \\
&= \mathbf{e}_1^\top \widehat{\mathbf{\Psi}}_{t,\mathbf{x}}^{-1} \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big) K_h(D_i(\mathbf{x}))(Y_i - \theta^*_{t,\mathbf{x}}(D_i(\mathbf{x})))\Big] + \theta^*_{t,\mathbf{x}}(0) - \theta_{t,\mathbf{x}}(0) \\
&= \underbrace{\theta^*_{t,\mathbf{x}}(0) - \theta_{t,\mathbf{x}}(0)}_{\text{approximation error}} + \underbrace{\mathbf{e}_1^\top \mathbf{\Psi}_{t,\mathbf{x}}^{-1} \mathbf{O}_{t,\mathbf{x}}}_{\text{linear error}} + \underbrace{\mathbf{e}_1^\top (\widehat{\mathbf{\Psi}}_{t,\mathbf{x}}^{-1} - \mathbf{\Psi}_{t,\mathbf{x}}^{-1}) \mathbf{O}_{t,\mathbf{x}}}_{\text{non-linear error}},
\end{aligned} \tag{SA-1}$$

where

$$\mathbf{O}_{t,\mathbf{x}} = \mathbb{E}_n\left[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)K_h(D_i(\mathbf{x}))(Y_i - \theta^*_{t,\mathbf{x}}(D_i(\mathbf{x})))\mathbb{1}(D_i(\mathbf{x})\in\mathscr{I}_t)\right],$$

$$\widehat{\mathbf{\Psi}}_{t,\mathbf{x}} = \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)^\top K_h(D_i(\mathbf{x}))\mathbb{1}(D_i(\mathbf{x})\in\mathscr{I}_t)\Big],$$

and the misspecification bias is

$$\mathfrak{B}_t(\mathbf{x}) = \theta^*_{t,\mathbf{x}}(0) - \theta_{t,\mathbf{x}}(0).$$

Finally, we define the following for quantities for future analysis: for $t\in\{0,1\}$, $\mathbf{x}_1,\mathbf{x}_2\in\mathscr{B}$,

$$\begin{aligned}\widehat{\mathbf{\Upsilon}}_{t,\mathbf{x}_1,\mathbf{x}_2} &= h^d\mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x}_1)}{h}\Big)\mathbf{r}_p\Big(\frac{D_i(\mathbf{x}_2)}{h}\Big)^\top K_h(D_i(\mathbf{x}_1))K_h(D_i(\mathbf{x}_2))\\ &\qquad (Y_i-\widehat{\theta}_{t,\mathbf{x}_1}(D_i(\mathbf{x}_1)))(Y_i-\widehat{\theta}_{t,\mathbf{x}_2}(D_i(\mathbf{x}_2)))\mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x}_1))\Big],\\ \mathbf{\Upsilon}_{t,\mathbf{x}_1,\mathbf{x}_2} &= h^d\mathbb{E}\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x}_1)}{h}\Big)\mathbf{r}_p\Big(\frac{D_i(\mathbf{x}_2)}{h}\Big)^\top K_h(D_i(\mathbf{x}_1))K_h(D_i(\mathbf{x}_2))\\ &\qquad (Y_i-\theta^*_{t,\mathbf{x}_1}(D_i(\mathbf{x}_1)))(Y_i-\theta^*_{t,\mathbf{x}_2}(D_i(\mathbf{x}_2))),\end{aligned}$$

$$\widehat{\Xi}_{\mathbf{x}_1,\mathbf{x}_2} = \widehat{\Xi}_{0,\mathbf{x}_1,\mathbf{x}_2} + \widehat{\Xi}_{1,\mathbf{x}_1,\mathbf{x}_2}, \qquad \widehat{\Xi}_{t,\mathbf{x}_1,\mathbf{x}_2} = \frac{1}{nh^d}\mathbf{e}_1^\top\widehat{\mathbf{\Psi}}^{-1}_{t,\mathbf{x}_1}\widehat{\mathbf{\Upsilon}}_{t,\mathbf{x}_1,\mathbf{x}_2}\widehat{\mathbf{\Psi}}^{-1}_{t,\mathbf{x}_2}\mathbf{e}_1$$

and

$$\Xi_{\mathbf{x}_1,\mathbf{x}_2} = \Xi_{0,\mathbf{x}_1,\mathbf{x}_2} + \Xi_{1,\mathbf{x}_1,\mathbf{x}_2}. \qquad \Xi_{t,\mathbf{x}_1,\mathbf{x}_2} = \frac{1}{nh^d}\mathbf{e}_1^\top\mathbf{\Psi}^{-1}_{t,\mathbf{x}_1}\mathbf{\Upsilon}_{t,\mathbf{x}_1,\mathbf{x}_2}\mathbf{\Psi}^{-1}_{t,\mathbf{x}_2}\mathbf{e}_1.$$

In particular, $\widehat{\Xi}_{\mathbf{x}} = \widehat{\Xi}_{\mathbf{x},\mathbf{x}}$, $\Xi_{\mathbf{x}} = \Xi_{\mathbf{x},\mathbf{x}}$, $\mathfrak{B}(\mathbf{x}) = \mathfrak{B}_1(\mathbf{x}) - \mathfrak{B}_0(\mathbf{x})$, etc.

SA-1.1 Notation and Definitions

For textbook references on empirical process, see van der Vaart and Wellner [1996], Dudley [2014], and Giné and Nickl [2016]. For textbook reference on geometric measure theory, see Simon et al. [1984], Federer [2014], and Folland [2002].

(i) *Multi-index Notations.* For a multi-index $\mathbf{u} = (u_1,\dots,u_d)\in\mathbb{N}^d$, denote $|\mathbf{u}| = \sum_{i=1}^d u_d$, $\mathbf{u}! = \Pi_{i=1}^d u_d$. Denote $\mathbf{r}_p(\mathbf{u}) = (1, u_1,\dots,u_d,u_1^2,\dots,u_d^2,\dots,u_1^p,\dots,u_d^p)$, that is, all monomials $u_1^{\alpha_1}\cdots u_d^{\alpha_d}$ such that $\alpha_i\in\mathbb{N}$ and $\sum_{i=1}^d\alpha_i\le p$. Define $\mathbf{e}_{1+\boldsymbol{\nu}}$ to be the $p_d = \frac{(d+p)!}{d!p!}$-dimensional vector such that $\mathbf{e}_{1+\boldsymbol{\nu}}^\top\mathbf{r}_p(\mathbf{u}) = \mathbf{u}^{\boldsymbol{\nu}}$ for all $\mathbf{u}\in\mathbb{R}^d$.

(ii) *Norms.* For a vector $\mathbf{v}\in\mathbb{R}^k$, $\|\mathbf{v}\| = (\sum_{i=1}^k\mathbf{v}_i^2)^{1/2}$, $\|\mathbf{v}\|_\infty = \max_{1\le i\le k}|\mathbf{v}_i|$. For a matrix $A\in\mathbb{R}^{m\times n}$, $\|A\|_p = \sup_{\|\mathbf{x}\|_p=1}\|A\mathbf{x}\|_p$, $p\in\mathbb{N}\cup\{\infty\}$, and $\lambda_{\min}(A)$ denotes its minimum eigenvalue. For a function f on a metric space (S,d), $\|f\|_\infty = \sup_{\mathbf{x}\in\mathcal{X}}|f(\mathbf{x})|$. For a probability measure Q on $(\mathcal{S},\mathscr{S})$ and $p\ge 1$, define $\|f\|_{Q,p} = (\int_{\mathcal{S}}|f|^p dQ)^{1/p}$. For a set $E\subseteq\mathbb{R}^d$, denote by $\mathfrak{m}(E)$ the Lebesgue measure of E.

(iii) *Empirical Process.* We use standard empirical process notations: $\mathbb{E}_n[g(\mathbf{v}_i)] = \frac{1}{n}\sum_{i=1}^n g(\mathbf{v}_i)$ and

$\mathbb{G}_n[g(\mathbf{v}_i)] = \frac{1}{\sqrt{n}}\sum_{i=1}^n (g(\mathbf{v}_i) - \mathbb{E}[g(\mathbf{v}_i)])$. Let $(\mathscr{S}, d)$ be a semi-metric space. The covering number $N(\mathscr{S}, d, \varepsilon)$ is the minimal number of balls $B_s(\varepsilon) = \{t : d(t,s) < \varepsilon\}$ needed to cover $\mathscr{S}$. A $\mathbb{P}$-*Brownian bridge* is a mean-zero Gaussian random function $W_n(f), f \in L_2(\mathscr{X}, \mathbb{P})$ with the covariance $\mathbb{E}[W_{\mathbb{P}}(f)W_{\mathbb{P}}(g)] = \mathbb{P}(fg) - \mathbb{P}(f)\mathbb{P}(g)$, for $f, g \in L_2(\mathscr{X}, \mathbb{P})$. A class $\mathscr{F} \subseteq L_2(\mathscr{X}, \mathbb{P})$ is $\mathbb{P}$-*pregaussian* if there is a version of $\mathbb{P}$-Brownian bridge $W_{\mathbb{P}}$ such that $W_{\mathbb{P}} \in C(\mathscr{F}; \rho_{\mathbb{P}})$ almost surely, where $\rho_{\mathbb{P}}$ is the semi-metric on $L_2(\mathscr{X}, \mathbb{P})$ is defined by $\rho_{\mathbb{P}}(f,g) = (\|f-g\|_{\mathbb{P},2}^2 - (\int f \, d\mathbb{P} - \int g \, d\mathbb{P})^2)^{1/2}$, for $f, g \in L_2(\mathscr{X}, \mathbb{P})$.

(iv) *Geometric Measure Theory.* For a set $E \subseteq \mathscr{X}$, the *De Giorgi perimeter of E related to $\mathscr{X}$* is $\mathscr{L}(E) = \mathtt{TV}_{\{\mathbb{1}_E\},\mathscr{X}}$. For $d \in \mathbb{N}$ and $0 \le m \le d$, the m-dimensional Hausdorff (outer) measure is given by $\mathfrak{H}^m(A) = \lim_{\delta \downarrow 0} \mathfrak{H}_\delta^m(A)$, $A \subseteq \mathbb{R}^d$, where for each $\delta > 0$, $\mathfrak{H}_\delta^m(A)$ is defined by taking $\mathfrak{H}_\delta^m(\emptyset) = 0$, and for any non-empty $A \subseteq \mathbb{R}^d$, $\mathfrak{H}_\delta^m(A) = \frac{\pi^{m/2}}{\Gamma(m/2+1)} \inf \sum_{j=1}^\infty (\mathrm{diam}(C_j)/2)^m$, and the infimum is taken over all countable collections $C_1, C_2, \cdots$ of subsets of $\mathbb{R}^d$ such that $\mathrm{diam}(C_j) < \delta$ and $A \subseteq \cup_{j=1}^\infty C_j$. Integration against $\mathfrak{H}^m$ is defined via Carathéodory's Theorem following the classical measure-theoretic literature. The Hausdorff dimension $\dim_{\mathfrak{H}}(A)$ of A is defined by $\dim_{\mathfrak{H}}(A) = \inf\{t \ge 0 : \mathfrak{H}^t(A) = 0\}$. A set $A \subseteq \mathbb{R}^d$ is said to be k-rectifiable if A is of Hausdorff dimension k, and there exist a countable collection $\{f_i\}$ of continuously differentiable maps $f_i : \mathbb{R}^k \to \mathbb{R}^d$ such that $\mathfrak{H}^k(E \setminus \cup_{i=0}^\infty f_i(\mathbb{R}^k)) = 0$. B is a *rectifiable curve* if there exists a Lipschitz continuous function $\gamma : [0,1] \to \mathbb{R}$ such that $B = \gamma([0,1])$. We define the curve length function of B to be $\mathfrak{L}(B) = \sup_{\pi \in \Pi} s(\pi, \gamma)$, where $\Pi = \{(t_0, t_1, \ldots, t_N) : N \in \mathbb{N}, 0 \le t_0 < t_1 < \ldots \le t_N \le 1\}$ and $s(\pi, \gamma) = \sum_{i=0}^N \|\gamma(t_i) - \gamma(t_{i+1})\|_2$ for $\pi = (t_0, t_1, \ldots, t_N)$.

(v) *Bounds and Asymptotics.* For reals sequences $|a_n| = o(|b_n|)$ if $\limsup \frac{a_n}{b_n} = 0$, $|a_n| \lesssim |b_n|$ if there exists some constant C and $N > 0$ such that $n > N$ implies $|a_n| \le C|b_n|$. For sequences of random variables $a_n = o_{\mathbf{b}P}(b_n)$ if $\mathrm{plim}_{n\to\infty} \frac{a_n}{b_n} = 0$, $|a_n| \lesssim_{\mathbb{P}} |b_n|$ if $\limsup_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}[|\frac{a_n}{b_n}| \ge M] = 0$.

(vi) *Distributions and Statistical Distances.* For $\boldsymbol{\mu} \in \mathbb{R}^k$ and $\boldsymbol{\Sigma}$ a $k \times k$ positive definite matrix, $\mathsf{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. For $-\infty < a < b < \infty$, $\mathsf{Uniform}([a,b])$ denotes the uniform distribution on $[a,b]$. $\mathsf{Bernoulli}(p)$ denotes the Bernoulli distribution with success probability p. $\Phi(\cdot)$ denotes the standard Gaussian cumulative distribution function. For two distributions P and Q, $d_{\mathrm{KL}}(P,Q)$ denotes the KL-distance between P and Q, and $d_{\chi^2}(P,Q)$ denotes the χ^2 distance between P and Q.

SA-1.2 Mapping between Main Paper and Supplement

The results in the main paper are special cases of the results in this supplemental appendix as follows.

- Theorem 1 in the paper corresponds to Theorem SA-1 with $d = 2$.
- Theorem 2 in the paper is proven in Section SA-6.14.
- Theorem 3 in the paper is proven in Section SA-6.15.
- Theorem 4(i) in the paper corresponds in Theorem SA-2 with $d = 2$.
- Theorem 4(ii) in the paper corresponds in Theorem SA-3 with $d = 2$.
- Theorem 5(i) in the paper corresponds in Theorem SA-4 with $d = 2$.
- Theorem 5(ii) in the paper corresponds in Theorem SA-7 with $d = 2$.

- Theorem 6 in the paper is proven in Section SA-6.16.

SA-2 Preliminary Lemmas

Recall that $t \in \{0,1\}$.

The following lemma gives a sufficient condition for Assumption SA–2.

Lemma SA-1 (Gram Invertibility). *Suppose the following conditions hold:*

1. *Assumptions SA–1(i)(ii) and Assumption SA–2 (iii) hold.*
2. *$\mathscr{d}(\cdot,\cdot)$ is the Euclidean distance.*
3. *There exists a set $U \subseteq \mathbb{R}^d$, such that $K(\|\mathbf{u}\|) \geq \kappa > 0$ for all $\mathbf{u} \in U$, $\lambda_{\min}(\int_U \mathbf{r}_p(\|\mathbf{z}\|)\mathbf{r}_p(\|\mathbf{z}\|)^\top d\mathbf{z}) > 0$, and $\liminf_{h\downarrow 0} \inf_{\mathbf{x}\in\mathscr{B}} \int_U K(\|\mathbf{u}\|)\mathbb{1}(\mathbf{x}+h\mathbf{u} \in \mathscr{A}_t)d\mathbf{u} \gtrsim 1$.*

Then Assumption SA–2 (iv) holds.

Lemma SA-2 (Gram). *Suppose Assumptions SA–1(i)(ii) and SA–2 hold. If $\frac{nh^d}{\log(1/h)} \to \infty$, then*

$$\sup_{\mathbf{x}\in\mathscr{B}} \left\|\widehat{\mathbf{\Psi}}_{t,\mathbf{x}} - \mathbf{\Psi}_{t,\mathbf{x}}\right\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}, \qquad 1 \lesssim_{\mathbb{P}} \inf_{\mathbf{x}\in\mathscr{B}} \left\|\widehat{\mathbf{\Psi}}_{t,\mathbf{x}}\right\| \leq \sup_{\mathbf{x}\in\mathscr{B}} \left\|\widehat{\mathbf{\Psi}}_{t,\mathbf{x}}\right\| \lesssim_{\mathbb{P}} 1,$$
$$\sup_{\mathbf{x}\in\mathscr{B}} \left\|\widehat{\mathbf{\Psi}}_{t,\mathbf{x}}^{-1} - \mathbf{\Psi}_{t,\mathbf{x}}^{-1}\right\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}.$$

Lemma SA-3 (Stochastic Linear Approximation). *Suppose Assumptions SA–1(i)(ii)(iii)(v) and SA–2 hold. If $\frac{nh^d}{\log(1/h)} \to \infty$, then*

$$\sup_{\mathbf{x}\in\mathscr{B}} \left\|\mathbf{O}_{t,\mathbf{x}}\right\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d},$$
$$\sup_{\mathbf{x}\in\mathscr{B}} \left|\mathbf{e}_1^\top \mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}}\right| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d},$$
$$\sup_{\mathbf{x}\in\mathscr{B}} \left|\mathbf{e}_1^\top (\widehat{\mathbf{\Psi}}_{t,\mathbf{x}}^{-1} - \mathbf{\Psi}_{t,\mathbf{x}}^{-1})\mathbf{O}_{t,\mathbf{x}}\right| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} \left(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}\right).$$

Lemma SA-4 (Covariance). *Suppose Assumptions SA–1 and SA–2 hold. If $\frac{nh^d}{\log(1/h)} \to \infty$, then*

$$\sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathscr{B}} \left\|\widehat{\mathbf{\Upsilon}}_{t,\mathbf{x}_1,\mathbf{x}_2} - \mathbf{\Upsilon}_{t,\mathbf{x}_1,\mathbf{x}_2}\right\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d},$$
$$\sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathscr{B}} nh^d\left|\widehat{\Xi}_{t,\mathbf{x}_1,\mathbf{x}_2} - \Xi_{t,\mathbf{x}_1,\mathbf{x}_2}\right| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}.$$

If, in addition, $\frac{n^{\frac{v}{2+v}}h^d}{\log(1/h)} \to \infty$, then

$$\inf_{\mathbf{x}\in\mathscr{B}} \lambda_{\min}(\widehat{\mathbf{\Upsilon}}_{t,\mathbf{x},\mathbf{x}}) \gtrsim_{\mathbb{P}} 1, \qquad \inf_{\mathbf{x}\in\mathscr{B}} \widehat{\Xi}_{t,\mathbf{x},\mathbf{x}} \gtrsim_{\mathbb{P}} (nh^d)^{-1},$$

and

$$\sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathscr{B}}\left|\frac{\widehat{\Xi}_{t,\mathbf{x}_1,\mathbf{x}_2}}{\sqrt{\widehat{\Xi}_{t,\mathbf{x}_1,\mathbf{x}_2}\widehat{\Xi}_{t,\mathbf{x}_2,\mathbf{x}_2}}}-\frac{\Xi_{t,\mathbf{x}_1,\mathbf{x}_2}}{\sqrt{\Xi_{t,\mathbf{x}_2,\mathbf{x}_2}\Xi_{t,\mathbf{x}_2,\mathbf{x}_2}}}\right|\lesssim_{\mathbb{P}}\sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}.$$

Lemma SA-5 (Uniform Bias: Minimal Guarantee)**.** *Suppose Assumptions SA–1 (i)(ii)(iii) and SA–2 hold. If $h\to 0$, then*

$$\sup_{\mathbf{x}\in\mathscr{B}}|\mathfrak{B}(\mathbf{x})|\lesssim h.$$

SA-3 Identification and Point Estimation

Theorem SA-1 (Distance-Based Identification)**.** *Suppose Assumptions SA–1(i)-(iii) and SA–2 hold. Then, $\tau(\mathbf{x})=\lim_{r\downarrow 0}\theta_{1,\mathbf{x}}(r)-\lim_{r\uparrow 0}\theta_{0,\mathbf{x}}(r)$ for all $\mathbf{x}\in\mathscr{B}$.*

Theorem SA-2 (Pointwise Convergence Rate)**.** *Suppose Assumptions SA–1 and SA–2 hold. If $nh^d\to\infty$, then*

$$\big|\widehat{\vartheta}(\mathbf{x})-\tau(\mathbf{x})\big|\lesssim_{\mathbb{P}}\frac{1}{\sqrt{nh^d}}+\frac{1}{n^{\frac{1+v}{2+v}}h^d}+\big|\mathfrak{B}(\mathbf{x})\big|.$$

Theorem SA-3 (Uniform Convergence Rate)**.** *Suppose Assumptions SA–1 and SA–2 hold. If $\frac{nh^d}{\log(1/h)}\to\infty$, then*

$$\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{\vartheta}(\mathbf{x})-\tau(\mathbf{x})\big|\lesssim_{\mathbb{P}}\sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}+\sup_{\mathbf{x}\in\mathscr{B}}\big|\mathfrak{B}(\mathbf{x})\big|.$$

SA-4 Distributional Approximation and Inference

Let $\mathbf{W}=((\mathbf{X}_1^\top,Y_1),\cdots,(\mathbf{X}_n^\top,Y_n))$, and recall that $t\in\{0,1\}$. The feasible t-statistics is

$$\widehat{\mathrm{T}}(\mathbf{x})=\frac{\widehat{\vartheta}(\mathbf{x})-\tau(\mathbf{x})}{\sqrt{\widehat{\Xi}_{\mathbf{x},\mathbf{x}}}},\qquad \mathbf{x}\in\mathscr{B}.$$

The associated $100(1-\alpha)\%$ confidence interval estimator is

$$\widehat{\mathrm{I}}_\alpha(\mathbf{x})=\left[\;\widehat{\vartheta}(\mathbf{x})-\mathfrak{q}_\alpha\sqrt{\widehat{\Xi}_{\mathbf{x},\mathbf{x}}}\;,\;\widehat{\vartheta}(\mathbf{x})+\mathfrak{q}_\alpha\sqrt{\widehat{\Xi}_{\mathbf{x},\mathbf{x}}}\;\right],$$

where $\mathfrak{q}_\alpha$ denotes an appropriate quantile depending on the desired confidence level $\alpha\in(0,1)$, and coverage objective (pointwise vs. uniform over $\mathscr{B}$). The following theorem establishes pointwise asymptotic normality and validity of confidence intervals. Let $\Phi(\cdot)$ be the cumulative distribution function of a standard univariate Gaussian random variable.

Theorem SA-4 (Confidence Intervals)**.** *Suppose Assumptions SA–1 and SA–2 hold. If $n^{\frac{v}{2+v}}h^d\to\infty$ and*

$\sqrt{nh^d}|\mathfrak{B}(\mathbf{x})| \to 0$, *then*

$$\sup_{u\in\mathbb{R}}\Big|\mathbb{P}\big(\widehat{\mathrm{T}}(\mathbf{x}) \le u\big) - \Phi(u)\Big| = o(1), \qquad \mathbf{x}\in\mathscr{B},$$

and

$$\mathbb{P}\big(\tau(\mathbf{x}) \in \widehat{\mathrm{I}}_\alpha(\mathbf{x})\big) = 1-\alpha+o(1), \qquad \mathbf{x}\in\mathscr{B},$$

provided that $\mathfrak{q}_\alpha = \inf\{c>0 : \mathbb{P}(|\widehat{Z}| \ge c|\mathbf{W}) \le \alpha\}$ *with* $\widehat{Z}|\mathbf{W} \sim \mathsf{Normal}(0, \widehat{\Xi}_{\mathbf{x},\mathbf{x}})$.

To conduct uniform inference, and in particular construct confidence bands, we rely on a new strong approximation result established in Section SA-5. First, we approximate (uniformly over $\mathbf{x}\in\mathscr{B}$) the feasible statistic $\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}$ by the following linear statistic (which is a sum of independent random variables):

$$\overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) = \Xi_{\mathbf{x},\mathbf{x}}^{-1/2}\Big(\mathbf{e}_1^\top \boldsymbol{\Psi}_{1,\mathbf{x}}^{-1}\mathbf{O}_{1,\mathbf{x}} - \mathbf{e}_1^\top \boldsymbol{\Psi}_{0,\mathbf{x}}^{-1}\mathbf{O}_{0,\mathbf{x}}\Big), \qquad \mathbf{x}\in\mathscr{B}.$$

Theorem SA-5 (Stochastic Linearization)**.** *Suppose Assumptions SA–1 and SA–2 hold. If* $\frac{nh^d}{\log(1/h)} \to \infty$*, then*

$$\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{\mathrm{T}}(\mathbf{x}) - \overline{\mathrm{T}}(\mathbf{x})\big| \lesssim_{\mathbb{P}} \sqrt{\log(1/h)}\bigg(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}\bigg) + \sqrt{nh^d}\sup_{\mathbf{x}\in\mathscr{B}}|\mathfrak{B}(\mathbf{x})|.$$

The pointwise (in $\mathscr{B}$) analogue of this result removes the $\log(1/h)$ penalty. See the proof of Theorem SA-4 for more details. To establish a Gaussian strong approximation for $\overline{\mathrm{T}}(\mathbf{x})$, define the class of functions $\mathscr{G} = \{g_{\mathbf{x}} : \mathbf{x}\in\mathscr{B}\}$ and $\mathscr{M} = \{m_{\mathbf{x}} : \mathbf{x}\in\mathscr{B}\}$, where

$$\begin{aligned} g_{\mathbf{x}}(\mathbf{u}) &= \mathbb{1}(\mathbf{u}\in\mathscr{A}_1)\mathfrak{K}_1(\mathbf{u};\mathbf{x}) - \mathbb{1}(\mathbf{u}\in\mathscr{A}_0)\mathfrak{K}_0(\mathbf{u};\mathbf{x}),\\ m_{\mathbf{x}}(\mathbf{u}) &= -\mathbb{1}(\mathbf{u}\in\mathscr{A}_1)\mathfrak{K}_1(\mathbf{u};\mathbf{x})\theta^*_{1,\mathbf{x}}(\mathscr{d}(\mathbf{u},\mathbf{x})) + \mathbb{1}(\mathbf{u}\in\mathscr{A}_0)\mathfrak{K}_0(\mathbf{u};\mathbf{x})\theta^*_{0,\mathbf{x}}(\mathscr{d}(\mathbf{u},\mathbf{x})), \end{aligned} \tag{SA-2}$$

with

$$\mathfrak{K}_t(\mathbf{u};\mathbf{x}) = \frac{1}{\sqrt{n\Xi_{\mathbf{x},\mathbf{x}}}}\mathbf{e}_1^\top\boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{r}_p\Big(\frac{\mathscr{d}(\mathbf{u},\mathbf{x})}{h}\Big)K_h(\mathscr{d}(\mathbf{u},\mathbf{x})),$$

for all $\mathbf{u}\in\mathscr{X}$, $\mathbf{x}\in\mathscr{B}$, and $t\in\{0,1\}$. In addition, let $\mathscr{R}$ be the class of functions containing the singleton identity function $\mathrm{Id}:\mathbb{R}\mapsto\mathbb{R}$, $\mathrm{Id}(x)=x$. Then, $\overline{\mathrm{T}}(\mathbf{x})$ can be represented as

$$\overline{\mathrm{T}}(\mathbf{x}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big[g_{\mathbf{x}}(\mathbf{X}_i)\,\mathrm{Id}(y_i) + m_{\mathbf{x}}(\mathbf{X}_i) - \mathbb{E}\big[g_{\mathbf{x}}(\mathbf{X}_i)\,\mathrm{Id}(y_i) + m_{\mathbf{x}}(\mathbf{X}_i)\big]\Big].$$

Following Cattaneo and Yu [2025], we define the multiplicative separable empirical processes by

$$M_n(g,r) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\big[g(\mathbf{x}_i)r(y_i) - \mathbb{E}[g(\mathbf{x}_i)r(y_i)]\big], \qquad g\in\mathscr{G}, r\in\mathscr{R},$$

which implies that

$$\overline{\mathrm{T}}(\mathbf{x}) = M_n(g_{\mathbf{x}},\mathrm{Id}) + M_n(m_{\mathbf{x}},1), \qquad \mathbf{x}\in\mathscr{B}.$$

Leveraging ideas in Cattaneo and Yu [2025], Theorem SA-8 gives a new Gaussian strong approximation that can be applied to $\overline{\mathrm{T}}(\mathbf{x})$. This new theorem allows for polynomial moment bound on the conditional distribution of $Y_i|\mathbf{X}_i$.

Theorem SA-6 (Gaussian Strong Approximation: $\overline{\mathrm{T}}$)**.** *Suppose Assumptions SA–1 and SA–2 hold, and that there exists a constant $C > 0$ such that for $t \in \{0,1\}$ and for any $\mathbf{x} \in \mathscr{B}$, the De Giorgi perimeter of the set $E_{t,\mathbf{x}} = \{\mathbf{y} \in \mathscr{A}_t : (\mathbf{y} - \mathbf{x})/h \in \operatorname{Supp}(K)\}$ satisfies $\mathscr{L}(E_{t,\mathbf{x}}) \leq Ch^{d-1}$. If $\liminf_{n\to\infty} \frac{\log h}{\log n} > -\infty$ and $nh^d \to \infty$ as $n \to \infty$, then (on a possibly enlarged probability space) there exists a mean-zero Gaussian process Z indexed by $\mathscr{B}$ with almost surely continuous sample path such that*

$$\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}} \big|\overline{\mathrm{T}}(\mathbf{x}) - z(\mathbf{x})\big|\Big] \lesssim (\log(n))^{\frac{3}{2}} \Big(\frac{1}{nh^d}\Big)^{\frac{1}{2d+2}\frac{v}{v+2}} + \log(n)\Big(\frac{1}{n^{\frac{v}{2+v}}h^d}\Big)^{\frac{1}{2}},$$

where $\lesssim$ is up to a universal constant, and $Z^{(\boldsymbol{\nu})}$ has the same covariance structure as $\overline{\mathrm{T}}$; i.e., $\mathbb{C}\mathrm{ov}[\overline{\mathrm{T}}(\mathbf{x}_1), \overline{\mathrm{T}}(\mathbf{x}_2)] = \mathbb{C}\mathrm{ov}[Z(\mathbf{x}_1), Z(\mathbf{x}_2)]$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}$.

Theorem SA-6 can be used to construct confidence bands for $(\tau(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$. Let $(\widehat{Z}(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$ be a (conditionally on $\mathbf{W}$) mean-zero Gaussian process with feasible (conditional) covariance function

$$\mathbb{C}\mathrm{ov}\Big[\widehat{Z}(\mathbf{x}_1), \widehat{Z}(\mathbf{x}_2)\Big|\mathbf{W}\Big] = \frac{\sqrt{\widehat{\Xi}_{\mathbf{x}_1,\mathbf{x}_2}}}{\sqrt{\widehat{\Xi}_{\mathbf{x}_1,\mathbf{x}_1}}\sqrt{\widehat{\Xi}_{\mathbf{x}_2,\mathbf{x}_2}}}, \qquad \mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}.$$

Theorem SA-7 (Confidence Bands)**.** *Suppose the assumptions and conditions in Theorem SA-6 hold. If $\liminf_{n\to\infty} \frac{\log h}{\log n} > -\infty$, $\frac{n^{\frac{v}{2+v}}h^d}{(\log n)^3} \to \infty$ and $\sqrt{nh^d}\sup_{\mathbf{x}\in\mathscr{B}} |\mathfrak{B}(\mathbf{x})| \to 0$, then*

$$\sup_{u\in\mathbb{R}} \Big|\mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|\widehat{\mathrm{T}}(\mathbf{x})\big| \leq u\Big) - \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|\widehat{Z}(\mathbf{x})\big| \leq u\Big|\mathbf{W}\Big)\Big| = o_{\mathbb{P}}(1)$$

and

$$\mathbb{P}\Big[\tau^{(\boldsymbol{\nu})}(\mathbf{x}) \in \widehat{\mathrm{I}}^{(\boldsymbol{\nu})}_{\alpha}(\mathbf{x}),\ \textit{for all } \mathbf{x} \in \mathscr{B}\Big] = 1 - \alpha + o(1),$$

provided that $\mathfrak{q}_\alpha = \inf\big\{c > 0 : \mathbb{P}\big(\sup_{\mathbf{x}\in\mathscr{B}} \big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\big| \geq c\big|\mathbf{W}\big) \leq \alpha\big\}$.

SA-5 Gaussian Strong Approximation

We present a Gaussian strong approximation theorem, which is the key technical tool behind Theorem SA-6. The theorem builds on and generalizes the results in Cattaneo and Yu [2025]. Consider the *residual-based empirical process* given by

$$M_n[g,r] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[g(\mathbf{x}_i)r(y_i) - \mathbb{E}[g(\mathbf{x}_i)r(y_i)]\Big], \qquad g \in \mathscr{G}, r \in \mathscr{R}.$$

where $\mathscr{G}$ and $\mathscr{R}$ are classes of functions satisfying certain regularity conditions.

SA-5.1 Definitions for Function Spaces

Let $\mathscr{F}$ be a class of measurable functions from a probability space $(\mathbb{R}^q, \mathscr{B}(\mathbb{R}^q), \mathbb{P})$ to $\mathbb{R}$. We introduce several definitions that capture properties of $\mathscr{F}$.

(i) $\mathscr{F}$ is pointwise measurable if it contains a countable subset $\mathscr{G}$ such that for any $f \in \mathscr{F}$, there exists a sequence $(g_m : m \geq 1) \subseteq \mathscr{G}$ such that $\lim_{m\to\infty} g_m(\mathbf{u}) = f(\mathbf{u})$ for all $\mathbf{u} \in \mathbb{R}^q$.

(ii) Let $\operatorname{Supp}(\mathscr{F}) = \cup_{f\in\mathscr{F}} \operatorname{Supp}(f)$. A probability measure $\mathbb{Q}_{\mathscr{F}}$ on $(\mathbb{R}^q, \mathscr{B}(\mathbb{R}^q))$ is a surrogate measure for $\mathbb{P}$ with respect to $\mathscr{F}$ if

   (i) $\mathbb{Q}_{\mathscr{F}}$ agrees with $\mathbb{P}$ on $\operatorname{Supp}(\mathbb{P}) \cap \operatorname{Supp}(\mathscr{F})$.

   (ii) $\mathbb{Q}_{\mathscr{F}}(\operatorname{Supp}(\mathscr{F}) \setminus \operatorname{Supp}(\mathbb{P})) = 0$.

   Let $\mathscr{Q}_{\mathscr{F}} = \operatorname{Supp}(\mathbb{Q}_{\mathscr{F}})$.

(iii) For $q = 1$ and an interval $\mathscr{I} \subseteq \mathbb{R}$, the pointwise total variation of $\mathscr{F}$ over $\mathscr{I}$ is

$$\mathtt{pTV}_{\mathscr{F},\mathscr{I}} = \sup_{f\in\mathscr{F}} \sup_{P\geq 1} \sup_{\mathscr{P}_P\in\mathscr{I}} \sum_{i=1}^{P-1} |f(a_{i+1}) - f(a_i)|,$$

where $\mathscr{P}_P = \{(a_1, \ldots, a_P) : a_1 \leq \cdots \leq a_P\}$ denotes the collection of all partitions of $\mathscr{I}$.

(iv) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, the total variation of $\mathscr{F}$ over $\mathscr{C}$ is

$$\mathtt{TV}_{\mathscr{F},\mathscr{C}} = \inf_{\mathscr{U}\in\mathscr{O}(\mathscr{C})} \sup_{f\in\mathscr{F}} \sup_{\phi\in\mathscr{D}_q(\mathscr{U})} \int_{\mathbb{R}^q} f(\mathbf{u}) \operatorname{div}(\phi)(\mathbf{u}) d\mathbf{u} / \|\|\phi\|_2\|_\infty,$$

where $\mathscr{O}(\mathscr{C})$ denotes the collection of all open sets that contains $\mathscr{C}$, and $\mathscr{D}_q(\mathscr{U})$ denotes the space of infinitely differentiable functions from $\mathbb{R}^q$ to $\mathbb{R}^q$ with compact support contained in $\mathscr{U}$.

(v) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, the local total variation constant of $\mathscr{F}$ over $\mathscr{C}$, is a positive number $\mathtt{K}_{\mathscr{F},\mathscr{C}}$ such that for any cube $\mathscr{D} \subseteq \mathbb{R}^q$ with edges of length ℓ parallel to the coordinate axises,

$$\mathtt{TV}_{\mathscr{F},\mathscr{D}\cap\mathscr{C}} \leq \mathtt{K}_{\mathscr{F},\mathscr{C}} \ell^{d-1}.$$

(vi) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, the envelopes of $\mathscr{F}$ over $\mathscr{C}$ are

$$\mathtt{M}_{\mathscr{F},\mathscr{C}} = \sup_{\mathbf{u}\in\mathscr{C}} M_{\mathscr{F},\mathscr{C}}(\mathbf{u}), \qquad M_{\mathscr{F},\mathscr{C}}(\mathbf{u}) = \sup_{f\in\mathscr{F}} |f(\mathbf{u})|, \qquad \mathbf{u} \in \mathscr{C}.$$

(vii) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, the Lipschitz constant of $\mathscr{F}$ over $\mathscr{C}$ is

$$\mathtt{L}_{\mathscr{F},\mathscr{C}} = \sup_{f\in\mathscr{F}} \sup_{\mathbf{u}_1,\mathbf{u}_2\in\mathscr{C}} \frac{|f(\mathbf{u}_1) - f(\mathbf{u}_2)|}{\|\mathbf{u}_1 - \mathbf{u}_2\|_\infty}.$$

(viii) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, the L_1 bound of $\mathscr{F}$ over $\mathscr{C}$ is

$$\mathtt{E}_{\mathscr{F},\mathscr{C}} = \sup_{f\in\mathscr{F}} \int_{\mathscr{C}} |f| d\mathbb{P}.$$

(ix) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, the uniform covering number of $\mathscr{F}$ with envelope $M_{\mathscr{F},\mathscr{C}}$ over $\mathscr{C}$ is

$$\mathtt{N}_{\mathscr{F},\mathscr{C}}(\delta, M_{\mathscr{F},\mathscr{C}}) = \sup_{\mu} N(\mathscr{F}, \|\cdot\|_{\mu,2}, \delta \|M_{\mathscr{F},\mathscr{C}}\|_{\mu,2}), \qquad \delta \in (0,\infty),$$

where the supremum is taken over all finite discrete measures on $(\mathscr{C}, \mathscr{B}(\mathscr{C}))$. We assume that $M_{\mathscr{F},\mathscr{C}}(\mathbf{u})$ is finite for every $\mathbf{u} \in \mathscr{C}$.

(x) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, the uniform entropy integral of $\mathscr{F}$ with envelope $M_{\mathscr{F},\mathscr{C}}$ over $\mathscr{C}$ is

$$J_{\mathscr{C}}(\delta, \mathscr{F}, M_{\mathscr{F},\mathscr{C}}) = \int_0^{\delta} \sqrt{1 + \log \mathtt{N}_{\mathscr{F},\mathscr{C}}(\varepsilon, M_{\mathscr{F},\mathscr{C}})} d\varepsilon,$$

where it is assumed that $M_{\mathscr{F},\mathscr{C}}(\mathbf{u})$ is finite for every $\mathbf{u} \in \mathscr{C}$.

(xi) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, $\mathscr{F}$ is a VC-type class with envelope $M_{\mathscr{F},\mathscr{C}}$ over $\mathscr{C}$ if (i) $M_{\mathscr{F},\mathscr{C}}$ is measurable and $M_{\mathscr{F},\mathscr{C}}(\mathbf{u})$ is finite for every $\mathbf{u} \in \mathscr{C}$, and (ii) there exist $\mathtt{c}_{\mathscr{F},\mathscr{C}} > 0$ and $\mathtt{d}_{\mathscr{F},\mathscr{C}} > 0$ such that

$$\mathtt{N}_{\mathscr{F},\mathscr{C}}(\varepsilon, M_{\mathscr{F},\mathscr{C}}) \leq \mathtt{c}_{\mathscr{F},\mathscr{C}} \varepsilon^{-\mathtt{d}_{\mathscr{F},\mathscr{C}}}, \qquad \varepsilon \in (0,1).$$

If a surrogate measure $\mathbb{Q}_{\mathscr{F}}$ for $\mathbb{P}$ with respect to $\mathscr{F}$ has been assumed, and it is clear from the context, we drop the dependence on $\mathscr{C} = \mathscr{Q}_{\mathscr{F}}$ for all quantities in the previous definitions. That is, to save notation, we set $\mathtt{TV}_{\mathscr{F}} = \mathtt{TV}_{\mathscr{F},\mathscr{Q}_{\mathscr{F}}}$, $\mathtt{K}_{\mathscr{F}} = \mathtt{K}_{\mathscr{F},\mathscr{Q}_{\mathscr{F}}}$, $\mathtt{M}_{\mathscr{F}} = \mathtt{M}_{\mathscr{F},\mathscr{Q}_{\mathscr{F}}}$, $M_{\mathscr{F}}(\mathbf{u}) = M_{\mathscr{F},\mathscr{Q}_{\mathscr{F}}}(\mathbf{u})$, $\mathtt{L}_{\mathscr{F}} = \mathtt{L}_{\mathscr{F},\mathscr{Q}_{\mathscr{F}}}$, and so on, whenever there is no confusion.

SA-5.2 Multiplicative-Separable Empirical Process

The following theorem generalizes Cattaneo and Yu [2025, Theorem SA.1] by requiring only bounded polynomial moments for y_i conditional on $\mathbf{x}_i$.

Theorem SA-8 (Strong Approximation for $(M_n(g,r) + M_n(h,s) : g \in \mathscr{G}, r \in \mathscr{R}, h \in \mathscr{H}, s \in \mathscr{S})$)**.** *Suppose $(\mathbf{z}_i = (\mathbf{x}_i, y_i) : 1 \leq i \leq n)$ are i.i.d. random vectors taking values in $(\mathbb{R}^{d+1}, \mathscr{B}(\mathbb{R}^{d+1}))$ with common law $\mathbb{P}_Z$, where $\mathbf{x}_i$ has distribution $\mathbb{P}_X$ supported on $\mathscr{X} \subseteq \mathbb{R}^d$, y_i has distribution $\mathbb{P}_Y$ supported on $\mathscr{Y} \subseteq \mathbb{R}$, $\sup_{\mathbf{x}\in\mathscr{X}} \mathbb{E}[|y_i|^{2+v}|\mathbf{x}_i = \mathbf{x}] \leq 2$ for some $v > 0$, and the following conditions hold.*

(i) *$\mathscr{G}$ and $\mathscr{H}$ are real-valued pointwise measurable classes of functions on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d), \mathbb{P}_X)$.*

(ii) *There exists a surrogate measure $\mathbb{Q}_{\mathscr{G}\cup\mathscr{H}}$ for $\mathbb{P}_X$ with respect to $\mathscr{G} \cup \mathscr{H}$ such that $\mathbb{Q}_{\mathscr{G}\cup\mathscr{H}} = \mathfrak{m} \circ \phi_{\mathscr{G}\cup\mathscr{H}}$, where the normalizing transformation $\phi_{\mathscr{G}\cup\mathscr{H}} : \mathscr{Q}_{\mathscr{G}\cup\mathscr{H}} \mapsto [0,1]^d$ is a diffeomorphism.*

(iii) *$\mathscr{G}$ is a VC-type class with envelope $\mathtt{M}_{\mathscr{G},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}}$ over $\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}$ with $\mathtt{c}_{\mathscr{G},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}} \geq e$ and $\mathtt{d}_{\mathscr{G},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}} \geq 1$. $\mathscr{H}$ is a VC-type class with envelope $\mathtt{M}_{\mathscr{H},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}}$ over $\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}$ with $\mathtt{c}_{\mathscr{H},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}} \geq e$ and $\mathtt{d}_{\mathscr{H},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}} \geq 1$.*

(iv) *$\mathscr{R}$ and $\mathscr{S}$ are real-valued pointwise measurable classes of functions on $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathbb{P}_Y)$.*

(v) *$\mathscr{R}$ is a VC-type class with envelope $M_{\mathscr{R},\mathscr{Y}}$ over $\mathscr{Y}$ with $\mathtt{c}_{\mathscr{R},\mathscr{Y}} \geq e$ and $\mathtt{d}_{\mathscr{R},\mathscr{Y}} \geq 1$, where $M_{\mathscr{R},\mathscr{Y}}(y) + \mathtt{pTV}_{\mathscr{R},(-|y|,|y|)} \leq \mathtt{v}(1 + |y|)$ for all $y \in \mathscr{Y}$, for some $\mathtt{v} > 0$. $\mathscr{S}$ is a VC-type class with envelope $M_{\mathscr{S},\mathscr{Y}}$ over $\mathscr{Y}$ with $\mathtt{c}_{\mathscr{S},\mathscr{Y}} \geq e$ and $\mathtt{d}_{\mathscr{S},\mathscr{Y}} \geq 1$, where $M_{\mathscr{S},\mathscr{Y}}(y) + \mathtt{pTV}_{\mathscr{S},(-|y|,|y|)} \leq \mathtt{v}(1 + |y|)$ for all $y \in \mathscr{Y}$, for some $\mathtt{v} > 0$.*

(vi) *There exists a constant* $\mathtt{k}$ *such that* $|\log_2 \mathtt{E}| + |\log_2 \mathtt{TV}| + |\log_2 \mathtt{M}| \leq \mathtt{k} \log_2(n)$, *where* $\mathtt{E} = \max\{\mathtt{E}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}, \mathtt{E}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\}$, $\mathtt{TV} = \max\{\mathtt{TV}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}, \mathtt{TV}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\}$ *and* $\mathtt{M} = \max\{\mathtt{M}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}, \mathtt{M}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\}$.

Consider the empirical process

$$A_n(g,h,r,s) = M_n(g,r) + M_n(h,s), \qquad g \in \mathscr{G}, r \in \mathscr{R}, h \in \mathscr{H}, s \in \mathcal{S}.$$

Then, on a possibly enlarged probability space, there exists a sequence of mean-zero Gaussian processes $(Z_n^A(g,h,r,s) : g \in \mathscr{G}, h \in \mathscr{H}, r \in \mathscr{R}, s \in \mathcal{S})$ *with almost sure continuous trajectories such that:*

- $\mathbb{E}[A_n(g_1,h_1,r_1,s_1)A_n(g_2,h_2,r_2,s_2)] = \mathbb{E}[Z_n^A(g_1,h_1,r_1,s_1)Z_n^A(g_2,h_2,r_2,s_2)]$ *holds for all* (g_1,h_1,r_1,s_1), $(g_2,h_2,r_2,s_2) \in \mathscr{G} \times \mathscr{H} \times \mathscr{R} \times \mathcal{S}$, *and*
- $\mathbb{E}\big[\,\|A_n - Z_n^A\|_{\mathscr{G}\times\mathscr{H}\times\mathscr{R}\times\mathcal{S}}\big] \leq C\mathtt{v}((\mathtt{d}\log(\mathtt{c}n))^{\frac{3}{2}}\mathtt{r}_n^{\frac{v}{v+2}}(\sqrt{\mathtt{ME}})^{\frac{2}{v+2}} + \mathtt{d}\log(\mathtt{c}n)\mathtt{M}n^{-\frac{v/2}{2+v}} + \mathtt{d}\log(\mathtt{c}n)\mathtt{M}n^{-\frac{1}{2}}\Big(\frac{\sqrt{\mathtt{ME}}}{\mathtt{r}_n}\Big)^{\frac{2}{v+2}})$,

where C *is a universal constant,* $\mathtt{c} = \mathtt{c}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}} + \mathtt{c}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}} + \mathtt{c}_{\mathscr{R},\mathscr{Y}} + \mathtt{c}_{\mathcal{S},\mathscr{Y}} + \mathtt{k}$, $\mathtt{d} = \mathtt{d}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\mathtt{d}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\mathtt{d}_{\mathscr{R},\mathscr{Y}}\mathtt{d}_{\mathcal{S},\mathscr{Y}}\mathtt{k}$,

$$\mathtt{r}_n = \min\Big\{\frac{(\mathtt{c}_1^d\mathtt{M}^{d+1}\mathtt{TV}^d\mathtt{E})^{1/(2d+2)}}{n^{1/(2d+2)}}, \frac{(\mathtt{c}_1^{\frac{d}{2}}\mathtt{c}_2^{\frac{d}{2}}\mathtt{MTV}^{\frac{d}{2}}\mathtt{EL}^{\frac{d}{2}})^{1/(d+2)}}{n^{1/(d+2)}}\Big\},$$
$$\mathtt{c}_1 = d\sup_{\mathbf{x}\in\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\prod_{j=1}^{d-1}\sigma_j(\nabla\phi_{\mathscr{G}\cup\mathscr{H}}(\mathbf{x})), \qquad \mathtt{c}_2 = \sup_{\mathbf{x}\in\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\frac{1}{\sigma_d(\nabla\phi_{\mathscr{G}\cup\mathscr{H}}(\mathbf{x}))}.$$

SA-6 Proofs

SA-6.1 Proof of Lemma SA-1

Assumption SA–1 (ii) implies

$$\begin{aligned}
\mathbf{\Psi}_{t,\mathbf{x}} &= \mathbb{E}\Big[\mathbf{r}_p\Big(\frac{\|\mathbf{X}_i-\mathbf{x}\|}{h}\Big)\mathbf{r}_p\Big(\frac{\|\mathbf{X}_i-\mathbf{x}\|}{h}\Big)^\top K_h(\|\mathbf{X}_i-\mathbf{x}\|)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)\Big]\\
&= \int_{\mathscr{A}_t}\mathbf{r}_p\Big(\frac{\|\mathbf{u}-\mathbf{x}\|}{h}\Big)\mathbf{r}_p\Big(\frac{\|\mathbf{u}-\mathbf{x}\|}{h}\Big)^\top K_h(\|\mathbf{u}-\mathbf{x}\|)f(\mathbf{u})d\mathbf{u}\\
&= f(\mathbf{x})\int_{\mathscr{A}_t}\mathbf{r}_p\Big(\frac{\|\mathbf{u}-\mathbf{x}\|}{h}\Big)\mathbf{r}_p\Big(\frac{\|\mathbf{u}-\mathbf{x}\|}{h}\Big)^\top K_h(\mathbf{u}-\mathbf{x})d\mathbf{u} + o(1),
\end{aligned}$$

where in the last line we have used $\int_{\mathscr{A}_t}(\frac{\|\mathbf{u}-\mathbf{x}\|}{h})^{\mathbf{v}}K_h(\|\mathbf{u}-\mathbf{x}\|)d\mathbf{u} = O(1)$ for any multi-index $\mathbf{v}$ from standard change of variable argument.

I. Polynomial Representation of Minimum Eigenvalue

For simplicity, call

$$\mathbf{S}_{t,\mathbf{x}} = \lim_{h\to 0}\mathbf{S}_{t,\mathbf{x}}(h), \qquad \mathbf{S}_{t,\mathbf{x}}(h) = \int_{\mathscr{A}_t}\mathbf{r}_p\Big(\frac{\|\mathbf{u}-\mathbf{x}\|}{h}\Big)\mathbf{r}_p\Big(\frac{\|\mathbf{u}-\mathbf{x}\|}{h}\Big)^\top K_h(\|\mathbf{u}-\mathbf{x}\|)d\mathbf{u}.$$

A change of variable gives

$$\mathbf{S}_{t,\mathbf{x}}(h) = \int\mathbf{r}_p(\|\mathbf{z}\|)\mathbf{r}_p(\|\mathbf{z}\|)^\top K(\|\mathbf{z}\|)\mathbb{1}(\mathbf{x}+h\mathbf{z}\in\mathscr{A}_t)d\mathbf{z}.$$

Let $\mathbf{a} \in \mathbb{R}^{\mathfrak{p}_p}$, where $\mathfrak{p}_p = \frac{(d+p)!}{d!p!}$. Then the equivalent representation of minimum eigenvalue gives

$$\begin{aligned}\lambda_{\min}(\mathbf{S}_{t,\mathbf{x}}(h)) &= \min_{\|\mathbf{a}\|=1} \int (\mathbf{a}^\top \mathbf{r}_p(\|\mathbf{z}\|))^2 K(\|\mathbf{z}\|) \mathbb{1}(\mathbf{x} + h\mathbf{z} \in \mathscr{A}_t) d\mathbf{z} \\ &\geq \kappa \min_{\|\mathbf{a}\|=1} \int_U (\mathbf{a}^\top \mathbf{r}_p(\|\mathbf{z}\|))^2 \mathbb{1}(\mathbf{x} + h\mathbf{z} \in \mathscr{A}_t) d\mathbf{z}, \end{aligned} \tag{SA-3}$$

where in the last line we have used $K(\mathbf{u}) \geq \kappa$ for all $u \in U$.

II. Mass Retaining Ratio in Treatment/Control Region

Denote $E_h(\mathbf{x}, t) = \{\mathbf{z} \in U : \mathbf{x} + h\mathbf{z} \in \mathscr{A}_t\}$. Assumption SA–2 (iii) implies there is some upper bound $\Lambda > 0$ of $K(\cdot)$. Hence for $c_0 = 1/2 \ \liminf_{h \downarrow 0} \inf_{\mathbf{x} \in \mathscr{B}} \int_U K(\|\mathbf{u}\|) \mathbb{1}(\mathbf{x} + h\mathbf{u} \in \mathscr{A}_t) d\mathbf{u}$, we have

$$\Lambda \mathfrak{m}(E_h(\mathbf{x}, t)) \geq \int_U K(\|\mathbf{u}\|) \mathbb{1}(\mathbf{x} + h\mathbf{u} \in \mathscr{A}_t) \geq c_0$$

for small enough h, which implies

$$\mathfrak{m}(E_h(\mathbf{x}, t)) \geq \alpha \mathfrak{m}(U), \qquad \alpha = \frac{c_0}{\Lambda \mathfrak{m}(U)}. \tag{SA-4}$$

III. L_2 Integral of Polynomials in Full v.s. Treatment/Control Regions

Consider $S = \{f \in \mathscr{P}_{p+1} : \int_U f(\|\mathbf{u}\|)^2 d\mathbf{u} = 1\}$, where $\mathscr{P}_{p+1}$ is the collection of all $(p+1)$-order polynomials. Let $(\phi_j, 1 \leq j \leq p+1)$ be a set of orthonormal basis of $(\mathscr{P}_{p+1}, \|\cdot\|_{L_2})$. Then $T(\mathbf{a}) = \sum_{j=1}^{p+1} a_j \phi_j$ is an isometry. Since $T(S) = \{\mathbf{a} \in \mathbb{R}^{p+1} : \|\mathbf{a}\| = 1\}$ is compact, S is also compact in $(\mathscr{P}_{p+1}, \|\cdot\|_{L_2})$. Since $\mathscr{P}_{p+1}$ is $(p+1)$-dimensional, equivalent of norms implies that S is also compact in $(\mathscr{P}_{p+1}, \|\cdot\|_{L_\infty})$. Now consider

$$\Phi_q(\varepsilon) = \mathfrak{m}(\{\mathbf{u} \in U : |q(u)| < \varepsilon\}), \qquad q \in S, \varepsilon > 0,$$

and

$$\psi(q) = \sup\Big\{\varepsilon > 0 : \Phi_q(\varepsilon) \leq \frac{\alpha}{2} \mathfrak{m}(U)\Big\}.$$

Since $\int_U q^2 = 1$ and q is polynomial on norm, $\lim_{\varepsilon \downarrow 0} \Phi_q(\varepsilon) = 0$ and $\Phi_q(\|q\|_\infty) = \mathfrak{m}(U)$. Continuity and Lipchitzness of $q \in S$ imply $\psi(q) > 0$ for all $q \in S$.

Next, we want to show ψ is lower-semicontinous function on $(\mathscr{P}_{p+1}, \|\cdot\|_{L_\infty})$. Suppose $q_n \to q$ uniformly on U. For every $\varepsilon_0 \in (0, \psi(q))$, there exists $\eta > 0$ such that $\Phi_q(\varepsilon_0) \leq \frac{\alpha}{2} \mathfrak{m}(U) - \eta$. Continuity of polynomials and the fact that level sets of polynomials have zero Lebesgue measure imply $\mathbb{1}_{\{|q_n| < \varepsilon_0\}}(\cdot) \to \mathbb{1}_{\{|q| < \varepsilon_0\}}(\cdot)$ almost surely. By Dominated Convergence Theorem, $\Phi_{q_n}(\varepsilon_0) \to \Phi_q(\varepsilon_0)$. Hence for large enough n, $\Phi_{q_n}(\varepsilon_0) \leq \frac{\alpha}{2} \mathfrak{m}(U)$, which implies $\varepsilon_0 \leq \psi(q_n)$. This implies $\liminf_{n \to \infty} \psi(q_n) \geq \varepsilon_0$. Since ε_0 is arbitrary in $(0, \psi(q))$, we have $\liminf_{n \to \infty} \psi(q_n) \geq \psi(q)$.

Compactness of S and lower-semicontinuity of ψ implies ψ attains its minimum on S. Since $\psi(q) > 0$ for all $q \in S$, we know $\varepsilon_* = \inf_{q \in S} \psi(q) > 0$. Then for every $q \in S$,

$$\begin{aligned}\int_{E_h(\mathbf{x},t)} q^2 &\geq \varepsilon_*^2 \, \mathfrak{m}\Big(E_h(\mathbf{x},t) \setminus \{|q| \leq \varepsilon_*\}\Big) \\ &\geq \varepsilon_*^2 \, \Big(\mathfrak{m}(E_h(\mathbf{x},t)) - \mathfrak{m}(\{|q| \leq \varepsilon_*\})\Big) \\ &\geq \varepsilon_*^2 \, \frac{\alpha}{2} \mathfrak{m}(U).\end{aligned}$$

Scaling q from S gives

$$\int_{E_h(\mathbf{x},t)} q^2 \geq \varepsilon_*^2 \, \frac{\alpha}{2} \int_U q^2, \qquad q \in \mathscr{P}_{p+1}. \tag{SA-5}$$

IV. Lower Bound of Minimum Eigenvalue

Equations (SA-3), (SA-4) and (SA-5) together give for small enough h,

$$\begin{aligned}\inf_{\mathbf{x} \in \mathscr{B}} \lambda_{\min}(\mathbf{S}_{t,\mathbf{x}}(h)) &\geq \kappa \inf_{\mathbf{x} \in \mathscr{B}} \min_{\|\mathbf{a}\|=1} \int_{E_h(\mathbf{x},t)} (\mathbf{a}^\top \mathbf{r}_p(\|\mathbf{z}\|))^2 d\mathbf{z}, \\ &\geq \kappa \varepsilon_*^2 \, \frac{\alpha}{2} \min_{\|\mathbf{a}\|=1} \int_U (\mathbf{a}^\top \mathbf{r}_p(\|\mathbf{z}\|))^2 d\mathbf{z} \\ &\geq \kappa \varepsilon_*^2 \, \frac{\alpha}{2} \lambda_{\min}\Big(\int_U \mathbf{r}_p(\|\mathbf{z}\|) \mathbf{r}_p(\|\mathbf{z}\|)^\top d\mathbf{z}\Big),\end{aligned}$$

which implies $\liminf_{h \to 0} \inf_{\mathbf{x} \in \mathscr{B}} \lambda_{\min}(\mathbf{S}_{t,\mathbf{x}}(h)) > 0$.

SA-6.2 Proof of Lemma SA-2

Since $\widehat{\mathbf{\Psi}}_{t,\mathbf{x}}$ is a finite dimensional matrix, it suffices to show the stated rate of convergence for each entry. For $0 \leq v \leq p$, define $\mathscr{G} = \{g_n(\cdot, \mathbf{x}) \mathbb{1}(\cdot \in \mathscr{A}_t) : \mathbf{x} \in \mathscr{X}\}$ with

$$g_n(\xi, \mathbf{x}) = \Big(\frac{\mathscr{d}(\xi, \mathbf{x})}{h}\Big)^v \frac{1}{h^d} K\Big(\frac{\mathscr{d}(\xi, \mathbf{x})}{h}\Big), \qquad \xi, \mathbf{x} \in \mathscr{X}.$$

We will show $\mathscr{G}$ is a VC-type of class.

Constant Envelope Function. We assume K is continuous and has compact support, and hence there exists a constant C_1 such that $\sup_{\mathbf{x} \in \mathscr{X}} \|g_n(\cdot, \mathbf{x})\|_\infty \leq C_1 h^{-d} = G$.

Diameter of $\mathscr{G}$ in L_2. For each $\mathbf{x} \in \mathscr{X}$, $g_n(\cdot, \mathbf{x})$ is supported on $\{\xi : \mathscr{d}(\xi, \mathbf{x}) \leq h\}$. By Assumption SA–1(ii) and Assumption SA–2(i), $\sup_{\mathbf{x} \in \mathscr{X}} \mathbb{P}\left(\mathscr{d}(\mathbf{X}_i, \mathbf{x}) \leq h\right) \lesssim h^d$. It follows that $\sup_{\mathbf{x} \in \mathscr{X}} \|g_n(\cdot, \mathbf{x})\|_{\mathbb{P},2} \leq C_2 h^{-d/2}$ for some constant C_2. We can take C_1 large enough so that $\sigma = C_2 h^{-d/2} \leq G = C_1 h^{-d}$.

Ratio. For some constant C_3, $\delta = \frac{\sigma}{F} = C_3 \sqrt{h^d}$.

Covering Numbers. Case 1: K is Lipschitz. Let $\mathbf{x}, \mathbf{x}' \in \mathscr{X}$. By Assumption SA–2,

$$\begin{aligned}&\sup_{\xi \in \mathscr{X}} \big|g_n(\xi, \mathbf{x}) - g_n(\xi, \mathbf{x}')\big| \\ &\leq \sup_{\xi \in \mathscr{X}} \Big[\Big(\frac{\mathscr{d}(\xi, \mathbf{x})}{h}\Big)^v - \Big(\frac{\mathscr{d}(\xi, \mathbf{x}')}{h}\Big)^v\Big] K_h(\mathscr{d}(\xi, \mathbf{x})) + \Big(\frac{\mathscr{d}(\xi, \mathbf{x}')}{h}\Big)^v \Big[K_h(\mathscr{d}(\xi, \mathbf{x})) - K_h(\mathscr{d}(\xi, \mathbf{x}'))\Big] \\ &\lesssim h^{-d-1} \|\mathbf{x} - \mathbf{x}'\|_\infty.\end{aligned}$$

By Lipschitz continuity property of $\mathscr{G}$, for any $\varepsilon \in (0,1]$ and for any finitely supported measure Q and metric $\|\cdot\|_{Q,2}$ based on $L_2(Q)$,

$$N(\{g_n(\cdot,\mathbf{x}) : \mathbf{x} \in \mathscr{X}\}, \|\cdot\|_{Q,2}, \varepsilon\|G\|_{Q,2}) \leq N(\mathscr{X}, \|\cdot\|_\infty, \varepsilon\|G\|_{Q,2}h^{d+1}) \overset{(i)}{\lesssim} \Big(\frac{\operatorname{diam}(\mathscr{X})}{\varepsilon\|G\|_{Q,2}h^{d+1}}\Big)^d \lesssim \Big(\frac{\operatorname{diam}(\mathscr{X})}{\varepsilon h}\Big)^d,$$

where inequality (i) uses the fact that $\varepsilon\|G\|_{Q,2}h^{d+1} \lesssim \varepsilon h \lesssim 1$. Thus, $\mathscr{G}$ forms a VC-type class in that $\sup_Q N(\mathscr{G}, \|\cdot\|_{Q,2}, \varepsilon\|G\|_{Q,2}) \lesssim (C_1/\epsilon)^{C_2}$ for all $\epsilon \in (0,1]$ with $C_1 = \frac{\operatorname{diam}(\mathscr{X})}{h}$ and $C_2 = d$. Moreover, for any discrete measure Q, and for any $\mathbf{x},\mathbf{x}' \in \mathscr{X}$, $\|g_n(\cdot,\mathbf{x})\mathbb{1}(\cdot \in \mathscr{A}_t) - g_n(\cdot,\mathbf{x}')\mathbb{1}(\cdot \in \mathscr{A}_t)\|_{Q,2} \leq \|g_n(\cdot,\mathbf{x}) - g_n(\cdot,\mathbf{x}')\|_{Q,2}$. Therefore,

$$\sup_Q N(\mathscr{G}, \|\cdot\|_{Q,2}, \varepsilon\|G\|_{Q,2}) \leq N(\mathscr{G}, \|\cdot\|_{Q,2}, \varepsilon\|G\|_{Q,2}) \leq (C_1/\varepsilon)^{C_2}, \qquad \varepsilon \in (0,1],$$

where the supremum is taken over all finite discrete measures on $\mathscr{X}$.

Case 2: $k = \mathbb{1}(\cdot \in [-1,1])$. Consider

$$m_n(\xi,\mathbf{x}) = \Big(\frac{\mathscr{d}(\xi,\mathbf{x})}{h}\Big)^v \frac{1}{h}\mathbb{1}(\xi \in \mathscr{A}_t), \qquad \xi,\mathbf{x} \in \mathscr{X},$$

$\mathscr{M} = \{m_n(\mathscr{d}(\cdot,\mathbf{x}) : \mathbf{x} \in \mathscr{B}\}$ and the constant envelope function $M = C_4 h^{-v-1}$, for some constant C_4 only depending on diameter of $\mathscr{X}$. The same argument as before shows that for any discrete measure Q, we have

$$N(\mathscr{M}, \|\cdot\|_{Q,2}, \varepsilon\|M\|_{Q,2}) \leq N(\mathscr{X}, \|\cdot\|_\infty, \varepsilon\|M\|_{Q,2}h^{1+v+1}) \lesssim \Big(\frac{\operatorname{diam}(\mathscr{X})}{\varepsilon\|M\|_{Q,2}h^{1+v+1}}\Big)^d \lesssim \Big(\frac{\operatorname{diam}(\mathscr{X})}{\varepsilon h}\Big)^d.$$

The class $\mathscr{L} = \{\mathbb{1}((\cdot - \mathbf{x})/h \in [-1,1]^d) : \mathbf{x} \in \mathscr{B}\}$ has VC dimension no greater than $2d$ [van der Vaart and Wellner, 1996, Example 2.6.1], and by van der Vaart and Wellner [1996, Theorem 2.6.4],

$$\sup_Q N(\mathscr{G}, \|\cdot\|_{Q,2}, \varepsilon\|G\|_{Q,2}) \leq N(\mathscr{G}, \|\cdot\|_{Q,2}, \varepsilon\|G\|_{Q,2}) \leq (C_1/\varepsilon)^{C_2}, \qquad \varepsilon \in (0,1],$$

where the supremum is taken over all finite discrete measures on $\mathscr{X}$.

Maximal Inequality. By Chernozhukov et al. [2014b, Corollary 5.1] for the empirical process on class $\mathscr{G}$,

$$\begin{aligned}\mathbb{E}\Big[\sup_{l\in\mathscr{G}}\big|\mathbb{E}_n[l(\mathbf{X}_i)] - \mathbb{E}[l(\mathbf{X}_i)]\big|\Big] &\lesssim \frac{\sigma}{\sqrt{n}}\sqrt{C_2\log(C_1/\delta)} + \frac{\|G\|_{\mathbb{P},2}C_2\log(C_1/\delta)}{n} \\ &\lesssim \frac{1}{\sqrt{nh^d}}\sqrt{d\log\Big(\frac{\operatorname{diam}(\mathscr{X})}{h^{1+d/2}}\Big)} + \frac{1}{nh^d}d\log\Big(\frac{\operatorname{diam}(\mathscr{X})}{h^{1+d/2}}\Big) \\ &\lesssim \sqrt{\frac{\log n}{nh^d}}.\end{aligned}$$

Thus, $\sup_{\mathbf{x}\in\mathscr{X}}\big\|\widehat{\mathbf{\Psi}}_{t,\mathbf{x}} - \mathbf{\Psi}_{t,\mathbf{x}}\big\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{nh^d}}$.

By Weyl's Theorem, $\sup_{\mathbf{x}\in\mathscr{X}}|\lambda_{\min}(\widehat{\mathbf{\Psi}}_{t,\mathbf{x}}) - \lambda_{\min}(\mathbf{\Psi}_{t,\mathbf{x}})| \leq \sup_{\mathbf{x}\in\mathscr{X}}\|\widehat{\mathbf{\Psi}}_{t,\mathbf{x}} - \mathbf{\Psi}_{t,\mathbf{x}}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{nh^d}}$. Therefore, we can lower bound the minimum eigenvalue by $\inf_{\mathbf{x}\in\mathscr{X}}\lambda_{\min}(\widehat{\mathbf{\Psi}}_{t,\mathbf{x}}) \geq \inf_{\mathbf{x}\in\mathscr{X}}\lambda_{\min}(\mathbf{\Psi}_{t,\mathbf{x}}) - \sup_{\mathbf{x}\in\mathscr{X}}|\lambda_{\min}(\widehat{\mathbf{\Psi}}_{t,\mathbf{x}}) - \lambda_{\min}(\mathbf{\Psi}_{t,\mathbf{x}})| \gtrsim_{\mathbb{P}} 1$.

Finally, it follows that $\sup_{\mathbf{x}\in\mathcal{X}} \|\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1}\| \lesssim_{\mathbb{P}} 1$ and hence

$$\sup_{\mathbf{x}\in\mathcal{X}} \left\|\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1} - \boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\right\| \le \sup_{\mathbf{x}\in\mathcal{X}} \left\|\boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\right\| \left\|\boldsymbol{\Psi}_{t,\mathbf{x}} - \widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}\right\| \left\|\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1}\right\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{nh^d}},$$

which completes the proof. □

SA-6.3 Proof of Lemma SA-3

Consider the class $\mathscr{F} = \{(\mathbf{z}, u) \mapsto \mathbf{e}_\nu^\top g_{\mathbf{x}}(\mathbf{z})(u - h_{\mathbf{x}}(\mathbf{z})) : \mathbf{x} \in \mathscr{B}\}$, $0 \le \boldsymbol{\nu} \le p$, where for $\mathbf{z} \in \mathcal{X}$,

$$g_{\mathbf{x}}(\mathbf{z}) = \mathbf{r}_p\Big(\frac{\mathscr{d}(\mathbf{z},\mathbf{x})}{h}\Big)K_h(\mathscr{d}(\mathbf{z},\mathbf{x})), \qquad h_{\mathbf{x}}(\mathbf{z}) = \boldsymbol{\gamma}_t^*(\mathbf{x})^\top \mathbf{r}_p\left(\mathscr{d}(\mathbf{z},\mathbf{x})\right).$$

By definition of $\boldsymbol{\gamma}_t^*(\mathbf{x})$,

$$\boldsymbol{\gamma}_t^*(\mathbf{x}) = \mathbf{H}^{-1}\boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{S}_{t,\mathbf{x}}, \qquad \mathbf{S}_{t,\mathbf{x}} = \mathbb{E}\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)K_h(D_i(\mathbf{x}))Y_i\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big]. \tag{SA-6}$$

Assumption SA–1 implies $\mathbf{S}_{t,\mathbf{x}}$ is continuous in $\mathbf{x}$, hence $\sup_{\mathbf{x}\in\mathcal{X}} \|\mathbf{S}_{t,\mathbf{x}}\| \lesssim 1$. And by Assumption SA–2(ii), $\inf_{\mathbf{x}\in\mathcal{X}} \lambda_{\min}(\boldsymbol{\Psi}_{t,\mathbf{x}}) \gtrsim 1$. Hence

$$\sup_{\mathbf{x}\in\mathscr{B}} \left\|\boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{S}_{t,\mathbf{x}}\right\| \lesssim 1. \tag{SA-7}$$

Now, consider properties of $\mathscr{F}$. Definition of $\boldsymbol{\gamma}_t^*(\mathbf{x})$ implies $\mathbb{E}[f(\mathbf{X}_i, Y_i)] = 0$ for all $f \in \mathscr{F}$. Since K is compactly supported, there exists $C_1, C_2 > 0$ such that $F(\mathbf{z}, u) = C_1 h^{-d}(|u| + C_2)$ is an envelope function for $\mathscr{F}$. Denote $M = \max_{1\le i\le n} F(\mathbf{X}_i, Y_i)$, then

$$\begin{aligned}
\mathbb{E}[M^2]^{1/2} &\lesssim h^{-d}\mathbb{E}\left[\max_{1\le i\le n} |Y_i|^2 + 1\right]^{1/2} \lesssim h^{-d}\mathbb{E}\left[\max_{1\le i\le n} |Y_i|^{2+v}\right]^{1/(2+v)} \\
&\lesssim h^{-d}\left[\sum_{i=1}^n \mathbb{E}[|\varepsilon_i + \sum_{t\in\{0,1\}} \mathbb{1}(\mathbf{x} \in \mathscr{A}_t)\mu_t(\mathbf{x})|^{2+v}]\right]^{1/(2+v)} \lesssim h^{-d}n^{1/(2+v)},
\end{aligned}$$

where we have used $\mathbf{X}$ is compact and μ_t is continuous, hence $\sup_{\mathbf{x}\in\mathcal{X}} |\sum_{t\in\{0,1\}} \mathbb{1}(\mathbf{x} \in \mathscr{A}_t)\mu_t(\mathbf{x})| \lesssim 1$. Denote $\sigma = \sup_{f\in\mathscr{F}} \mathbb{E}[f(\mathbf{X}_i, Y_i)^2]^{1/2}$. Then,

$$\sigma^2 \lesssim \sup_{\mathbf{x}\in\mathscr{B}} \mathbb{E}[\|\mathbf{e}_\nu^\top g_{\mathbf{x}}\|_\infty^2(|Y_i| + \|\mathbf{e}_\nu^\top h_{\mathbf{x}}\|_\infty)^2\mathbb{1}(K_h(D_i(\mathbf{x})) \neq 0)] \lesssim h^{-d}.$$

To check for the covering number of $\mathscr{F}$, notice that compare to the proof of Lemma SA-2, we have one more term $\mathbf{e}_{\boldsymbol{\nu}}^\top g_{\mathbf{x}}h_{\mathbf{x}} = \mathbf{r}_p\left(\frac{\mathscr{d}(\mathbf{z},\mathbf{x})}{h}\right)K_h(\mathscr{d}(\mathbf{z},\mathbf{x}))\boldsymbol{\gamma}_t^*(\mathbf{x})^\top\mathbf{r}_p\left(\mathscr{d}(\mathbf{z},\mathbf{x})\right)$. All terms except for $\boldsymbol{\gamma}_t^*(\mathbf{x})$ can be handled as in the proof of Lemma SA-2. Recall Equation (SA-6), and consider $l_{t,\mathbf{x}} = \mathbf{e}_{\mathbf{v}}^\top[\mathbf{R}(\mathscr{d}(\cdot,\mathbf{x})/h)K_h(\mathscr{d}(\cdot,\mathbf{x}))\mu_t\mathbb{1}(\cdot \in \mathscr{A}_t)$ and $\mathscr{L}_t = \{l_{t,\mathbf{x}} : \mathbf{x} \in \mathscr{B}\}$, $\mathbf{v}$ is a any multi-index. Then, for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}$,

$$|\mathbf{S}_{t,\mathbf{x}_1} - \mathbf{S}_{t,\mathbf{x}_2}| \le \|l_{t,\mathbf{x}_1} - l_{t,\mathbf{x}_2}\|_{\mathbb{P}_X,2},$$

and hence

$$N(\{\mathbf{e}_{\mathbf{v}}^\top \mathbf{S}_{t,\mathbf{x}} : \mathbf{x} \in \mathscr{B}\}, |\cdot|, \varepsilon h^{-d}) \le N(\mathscr{L}_t, \|\cdot\|_{\mathbb{P}_X,2}, \varepsilon h^{-d}) \le \sup_Q N(\mathscr{L}_t, \|\cdot\|_{Q,2}, \varepsilon h^{-d}),$$

Same argument as paragraph **Covering Numbers** in the proof of Lemma SA-2 then shows

$$\sup_Q N(\{g_{\mathbf{x}} : \mathbf{x} \in \mathscr{B}\}, \|\cdot\|_{Q,2}, \varepsilon C_1 h^{-d}) \le \left(\frac{\operatorname{diam}(\mathscr{X})}{h\varepsilon}\right)^d, \qquad 0 < \varepsilon \le 1,$$
$$\sup_Q N(\{g_{\mathbf{x}} h_{\mathbf{x}} : \mathbf{x} \in \mathscr{B}\}, \|\cdot\|_{Q,2}, \varepsilon C_1 h^{-d}) \le \left(\frac{\operatorname{diam}(\mathscr{X})}{h\varepsilon}\right)^d, \qquad 0 < \varepsilon \le 1,$$

where sup is taken over all discrete measures on $\mathscr{X}$. Product $\{g_{\mathbf{x}} : \mathbf{x} \in \mathscr{B}\}$ with the singleton of identity function $\{u \mapsto u, u \in \mathbb{R}\}$, and adding $\{g_{\mathbf{x}} h_{\mathbf{x}} : \mathbf{x} \in \mathscr{B}\}$,

$$\sup_Q N(\mathscr{F}, \|\cdot\|_{Q,2}, \varepsilon \|F\|_{Q,2}) \le 2\left(\frac{2\operatorname{diam}(\mathscr{X})}{h\varepsilon}\right)^d, \qquad 0 < \varepsilon \le 1,$$

where sup is taken over all discrete measures on $\mathscr{X} \times \mathbb{R}$. Denote $\mathtt{C}_1 = d$, $\mathtt{C}_2 = \frac{2(2\operatorname{diam}(\mathscr{X}))^d}{h^d}$. Hence, by Chernozhukov et al. [2014b, Corollary 5.1]

$$\begin{aligned}
\mathbb{E}\left[\sup_{\mathbf{x} \in \mathscr{B}} |\mathbf{e}_{\boldsymbol{\nu}}^\top \mathbf{O}_{t,\mathbf{x}}|\right] &= \mathbb{E}\left[\sup_{f \in \mathscr{F}} |\mathbb{E}_n[f(\mathbf{X}_i, Y_i)] - \mathbb{E}[f(\mathbf{X}_i, Y_i)]|\right] \\
&\lesssim \frac{\sigma}{\sqrt{n}}\sqrt{\mathtt{C}_2 \log(\mathtt{C}_1 \|M\|_{\mathbb{P},2}/\sigma)} + \frac{\|M\|_{\mathbb{P},2}\mathtt{C}_2 \log(\mathtt{C}_1 \|M\|_{\mathbb{P},2}/\sigma)}{n} \\
&\lesssim \frac{1}{\sqrt{nh^d}}\sqrt{d \log\left(\frac{\operatorname{diam}(\mathscr{X})}{h^{1+d/2}}\right)} + \frac{1}{n^{\frac{1+v}{2+v}} h^d} d \log\left(\frac{\operatorname{diam}(\mathscr{X})}{h^{1+d/2}}\right) \\
&\lesssim \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}} h^d}.
\end{aligned}$$

The rest follows from finite dimensionality of $\mathbf{O}_{t,\mathbf{x}}$, and Lemma SA-2. □

SA-6.4 Proof of Lemma SA-4

Denote $\eta_{i,t,\mathbf{x}} = Y_i - \theta^*_{t,\mathbf{x}}(D_i(\mathbf{x}))$ and $\xi_{i,t,\mathbf{x}} = \theta^*_{t,\mathbf{x}}(D_i(\mathbf{x})) - \widehat{\theta}_{t,\mathbf{x}}(D_i(\mathbf{x}))$. Then

$$\widehat{\boldsymbol{\Upsilon}}_{t,\mathbf{x},\mathbf{y}} = \mathbb{E}_n\left[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)\mathbf{r}_p\left(\frac{D_i(\mathbf{y})}{h}\right)^\top h^d K_h(D_i(\mathbf{x})) K_h(D_i(\mathbf{y})) (\eta_{i,t,\mathbf{x}} + \xi_{i,t,\mathbf{x}})^2 \mathbb{1}(D_i(\mathbf{x}) \in \mathscr{I}_t)\right],$$

and we decompose the error into

$$\begin{aligned}
&\widehat{\boldsymbol{\Upsilon}}_{t,\mathbf{x},\mathbf{y}} - \boldsymbol{\Upsilon}_{t,\mathbf{x},\mathbf{y}} = \Delta_{1,t,\mathbf{x},\mathbf{y}} + \Delta_{2,t,\mathbf{x},\mathbf{y}} + \Delta_{3,t,\mathbf{x},\mathbf{y}}, \\
&\Delta_{1,t,\mathbf{x},\mathbf{y}} = \mathbb{E}_n\left[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)\mathbf{r}_p\left(\frac{D_i(\mathbf{y})}{h}\right)^\top h^d K_h(D_i(\mathbf{x})) K_h(D_i(\mathbf{y})) \xi_{i,t,\mathbf{x}}^2 \mathbb{1}(D_i(\mathbf{x}) \in \mathscr{I}_t)\right], \\
&\Delta_{2,t,\mathbf{x},\mathbf{y}} = 2\mathbb{E}_n\left[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)\mathbf{r}_p\left(\frac{D_i(\mathbf{y})}{h}\right)^\top h^d K_h(D_i(\mathbf{x})) K_h(D_i(\mathbf{y})) \eta_{i,t,\mathbf{x}}\xi_{i,t,\mathbf{x}} \mathbb{1}(D_i(\mathbf{x}) \in \mathscr{I}_t)\right],
\end{aligned}$$

$$
\begin{aligned}
\Delta_{3,t,\mathbf{x},\mathbf{y}} = \mathbb{E}_n\Bigg[&\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)\mathbf{r}_p\left(\frac{D_i(\mathbf{y})}{h}\right)^\top h^d K_h\left(D_i(\mathbf{x})\right)K_h\left(D_i(\mathbf{y})\right)\eta_{i,t,\mathbf{x}}^2 \mathbb{1}(D_i(\mathbf{x})\in\mathscr{I}_t)\Bigg]\\
&- \mathbb{E}\Bigg[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)\mathbf{r}_p\left(\frac{D_i(\mathbf{y})}{h}\right)^\top h^d K_h\left(D_i(\mathbf{x})\right)K_h\left(D_i(\mathbf{y})\right)\eta_{i,t,\mathbf{x}}^2 \mathbb{1}(D_i(\mathbf{x})\in\mathscr{I}_t)\Bigg].
\end{aligned}
$$

By Assumption SA–2, $K_h(D_i(\mathbf{x}))\neq 0$ implies $\|\mathbf{r}_p(D_i(\mathbf{x})/h)\|_2 \lesssim 1$. Hence by Lemma SA-2 and SA-3,

$$
\begin{aligned}
&\max_{t\in\{0,1\}}\max_{1\le i\le n}\sup_{\mathbf{x}\in\mathscr{B}}|\xi_{i,t,\mathbf{x}}|\\
&= \max_{t\in\{0,1\}}\max_{1\le i\le n}\sup_{\mathbf{x}\in\mathscr{B}}|\mathbf{r}_p(D_i(\mathbf{x}))^\top(\widehat{\boldsymbol{\gamma}}_{t,\mathbf{x}}-\boldsymbol{\gamma}^*_{t,\mathbf{x}})|\mathbb{1}(K_h(D_i(\mathbf{x}))\ge 0)\\
&= \max_{t\in\{0,1\}}\max_{1\le i\le n}\sup_{\mathbf{x}\in\mathscr{B}}|\mathbf{r}_p(D_i(\mathbf{x}))^\top\mathbf{H}^{-1}(\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}}+(\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1}-\boldsymbol{\Psi}_{t,\mathbf{x}}^{-1})\mathbf{U}_{t,\mathbf{x}})|\mathbb{1}(K_h(D_i(\mathbf{x}))\ge 0)\\
&\le \max_{t\in\{0,1\}}\sup_{\mathbf{x}\in\mathscr{B}}\left\|\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}}\right\|_2 + \max_{t\in\{0,1\}}\sup_{\mathbf{x}\in\mathscr{B}}\left\|(\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1}-\boldsymbol{\Psi}_{t,\mathbf{x}}^{-1})\mathbf{U}_{t,\mathbf{x}}\right\|_2\\
&\lesssim \sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d},
\end{aligned}
$$

where

$$
\mathbf{U}_{t,\mathbf{x}} = \mathbb{E}_n\left[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)K_h(D_i(\mathbf{x}))\theta^*_{t,\mathbf{x}}(\mathbf{X}_i)\mathbb{1}(D_i(\mathbf{x})\in\mathscr{I}_t)\right].
$$

Assuming $\frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}\to\infty$, similar maximal inequality as in the proof of Lemma SA-2 shows

$$
\begin{aligned}
\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}}\|\Delta_{1,t,\mathbf{x},\mathbf{y}}\| &\lesssim_{\mathbb{P}} \max_{t\in\{0,1\}}\max_{1\le i\le n}\sup_{\mathbf{x}\in\mathscr{B}}|\xi_{i,t,\mathbf{x}}|^2 \lesssim \left(\sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}\right)^2,\\
\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}}\|\Delta_{2,t,\mathbf{x},\mathbf{y}}\| &\lesssim_{\mathbb{P}} \max_{t\in\{0,1\}}\max_{1\le i\le n}\sup_{\mathbf{x}\in\mathscr{B}}|\xi_{i,t,\mathbf{x}}| \lesssim \sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}.
\end{aligned}
\tag{SA-8}
$$

Consider the (μ,ν) entry of $\Delta_{3,t,\mathbf{x},\mathbf{y}}$. Consider the class

$$
\mathscr{F} = \left\{(\mathbf{z},u)\mapsto\left(\frac{d(\mathbf{z},\mathbf{x})}{h}\right)^{\mu+\nu}h^dK_h(d(\mathbf{z},\mathbf{x}))K_h(d(\mathbf{z},\mathbf{y}))(u-\mathbf{r}_p(d(\mathbf{z},\mathbf{x}))^\top\boldsymbol{\gamma}^*_{t,\mathbf{x}})^2:\mathbf{x},\mathbf{y}\in\mathscr{X}\right\}.
$$

By Assumption SA–2 and SA–1(v), we have $\sup_{f\in\mathscr{F}}\mathbb{E}[f(\mathbf{X}_i,Y_i)^2]^{1/2}\lesssim h^{-d/2}$. Moreover, Assumption SA–2 and Equation (SA-7) imply there exists $C_1,C_2>0$ such that $F(\mathbf{z},u)=C_1h^{-d}(u^2+C_2)$ is an envelope function for $\mathscr{F}$, with

$$
\mathbb{E}\big[\max_{1\le i\le n}F(\mathbf{X}_i,Y_i)^2\big]^{\frac{1}{2}}\lesssim C_1h^{-d}(\mathbb{E}\big[\max_{1\le i\le n}Y_i^4\big]^{\frac{1}{2}}+C_2)\lesssim C_1h^{-d}(\mathbb{E}\big[\max_{1\le i\le n}Y_i^{2+v}\big]^{\frac{2}{2+v}}+C_2)\lesssim h^{-d}n^{\frac{2}{2+v}}.
$$

Apply Chernozhukov et al. [2014b, Corollary 5.1] similarly as in Lemma SA-3 gives

$$
\mathbb{E}\left[\sup_{f\in\mathscr{F}}|\mathbb{E}_n[f(\mathbf{X}_i,Y_i)]-\mathbb{E}[f(\mathbf{X}_i,Y_i)]|\right]\lesssim\sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}.
$$

Finite dimensionality of $\Delta_{3,t,\mathbf{x},\mathbf{y}}$ then implies

$$\mathbb{E}\left[\sup_{\mathbf{x},\mathbf{y}\in\mathcal{X}} \|\Delta_{3,t,\mathbf{x},\mathbf{y}}\|\right] \lesssim \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}} h^d}. \tag{SA-9}$$

Putting together Equations (SA-8), (SA-9) and Lemma SA-2 gives the result. □

SA-6.5 Proof of Lemma SA-5

By Theorem SA-1 and Equation (SA-6), we have

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathscr{B}} |\mathfrak{B}_{n,t}(\mathbf{x})| &= \sup_{\mathbf{x}\in\mathscr{B}} \left|\mathbf{e}_1^\top \mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{S}_{t,\mathbf{x}} - \mu_t(\mathbf{x})\right| \\
&= \sup_{\mathbf{x}\in\mathscr{B}} \left|\mathbf{e}_1^\top \mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbb{E}\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)K_h(D_i(\mathbf{x}))\mathbf{R}_p(D_i(\mathbf{x}))^\top(\mu_t(\mathbf{X}_i) - \mu_t(\mathbf{x}), 0, \cdots, 0))\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big]\right| \\
&\lesssim \sup_{\mathbf{x}\in\mathscr{B}}\sup_{\mathbf{z}\in\mathcal{X}} |\mu_t(\mathbf{x}) - \mu_t(\mathbf{z})|\mathbb{1}(K_h(\mathfrak{d}(\mathbf{z},\mathbf{x})) > 0) \\
&\lesssim h.
\end{aligned}$$

□

SA-6.6 Proof of Theorem SA-1

Since $\theta_{\mathbf{x}}(0) = \theta_{1,\mathbf{x}}(0) - \theta_{0,\mathbf{x}}(0)$ and $\tau(\mathbf{x}) = \mu_1(\mathbf{x}) - \mu_0(\mathbf{x})$, it is enough to prove the result for one treatment assignment group $t \in \{0, 1\}$. By Assumption SA–1(iii) and Assumption SA–2(ii), for any $r \neq 0$, for any $\mathbf{x} \in \mathscr{B}$ and $\mathbf{y} \in S_{t,\mathbf{x}}(r)$, $|\mu_t(\mathbf{y}) - \mu_t(\mathbf{x})| \lesssim |r|$. Hence, for any $r \neq 0$, for any $\mathbf{x} \in \mathscr{B}$, $t \in \{0, 1\}$,

$$|\theta_{t,\mathbf{x}}(r) - \mu_t(\mathbf{x})| \leq \frac{\int_{S_{t,\mathbf{x}}(|r|)} |\mu_t(\mathbf{y}) - \mu_t(\mathbf{x})| f_X(\mathbf{y})\mathfrak{H}^{d-1}(d\mathbf{y})}{\int_{S_{t,\mathbf{x}}(|r|)} f_X(\mathbf{y})\mathfrak{H}^{d-1}(d\mathbf{y})} \lesssim r.$$

implying

$$|\theta_{t,\mathbf{x}}(0) - \mu_t(\mathbf{x})| \leq \lim_{r\to 0} |\theta_{t,\mathbf{x}}(r) - \mu_t(\mathbf{x})| = 0,$$

which establishes the result. □

SA-6.7 Proof of Theorem SA-2

The proofs of Lemma SA-2 and Lemma SA-3 can be done when the index set is the singleton $\{\mathbf{x}\}$ instead of $\mathscr{B}$, replacing Chernozhukov et al. [2014b, Corollary 5.1] by Bernstein inequality, and thus obtaining

$$\begin{aligned}
\left|\mathbf{e}_1^\top \mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}}\right| &\lesssim_{\mathbb{P}} \sqrt{\frac{1}{nh^d}} + \frac{1}{n^{\frac{1+v}{2+v}} h^d}, \\
\left|\mathbf{e}_1^\top (\widehat{\mathbf{\Psi}}_{t,\mathbf{x}}^{-1} - \mathbf{\Psi}_{t,\mathbf{x}}^{-1})\mathbf{O}_{t,\mathbf{x}}\right| &\lesssim_{\mathbb{P}} \sqrt{\frac{1}{nh^d}}\Bigg(\sqrt{\frac{1}{nh^d}} + \frac{1}{n^{\frac{1+v}{2+v}} h^d}\Bigg).
\end{aligned}$$

for all $\mathbf{x} \in \mathscr{B}$. In words, uniformity only adds a $\log(1/h)$ penalty. Therefore, using decomposition (SA-1), the pointwise convergence rate follows. □

SA-6.8 Proof of Theorem SA-3

Follows from Lemma SA-2, Lemma SA-3 and decomposition (SA-1). $\square$

SA-6.9 Proof of Theorem SA-4

Define $\overline{\mathrm{T}}(\mathbf{x}) = \sum_{i=1}^n Z_i$, with $Z_i = Z_{1,i} - Z_{0,i}$ independent random variables ($i = 1, 2, \ldots, n$),

$$Z_{t,i} = \frac{1}{n}\Xi_{\mathbf{x},\mathbf{x}}^{-1/2}\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)K_h(D_i(\mathbf{x}))(Y_i - \theta_{t,\mathbf{x}}^*(D_i(\mathbf{x})))\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t),$$

$\mathbb{E}[Z_i] = 0$ and $\mathbb{V}[Z_i] = n^{-1}$. By the Berry-Essen Theorem,

$$\sup_{u\in\mathbb{R}}\Big|\mathbb{P}\big(\overline{\mathrm{T}}(\mathbf{x}) \leq u\big) - \Phi(u)\Big| \lesssim \sum_{i=1}^n \mathbb{E}\big[|Z_i|^3\big] \lesssim \sum_{i=1}^n \mathbb{E}\big[|Z_{1,i}|^3\big] + \sum_{i=1}^n \mathbb{E}\big[|Z_{0,i}|^3\big]$$

where

$$\begin{aligned}
\mathbb{E}\big[|Z_{t,i}|^3\big] &= \sum_{i=1}^n n^{-3}\Xi_{\mathbf{x},\mathbf{x}}^{-3/2}\mathbb{E}\Big[\Big|\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)K_h(D_i(\mathbf{x}))\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)(Y_i - \theta_{t,\mathbf{x}}^*(D_i(\mathbf{x}))\Big|^3\Big] \\
&\lesssim n^{-2}\Xi_{\mathbf{x},\mathbf{x}}^{-3/2}\mathbb{E}[|K_h(D_i(\mathbf{x}))(Y_i - \theta_{t,\mathbf{x}}^*(D_i(\mathbf{x})))|^3] \\
&\lesssim n^{-2}\Xi_{\mathbf{x},\mathbf{x}}^{-3/2}\mathbb{E}[|K_h(D_i(\mathbf{x}))(\mathbb{E}[|Y_i|^3|\mathbf{X}_i] + |\theta_{t,\mathbf{x}}^*(D_i(\mathbf{x}))|^3)|] \\
&\lesssim (nh^d)^{-1/2},
\end{aligned}$$

noting that $\sup_{\mathbf{x}\in\mathscr{B}}\big\|\mathbf{r}_p\big(\frac{D_i(\mathbf{x})}{h}\big)K_h(D_i(\mathbf{x}))\big\| \lesssim 1$ holds almost surely in $\mathbf{X}_i$, $\Xi_{\mathbf{x},\mathbf{x}} \gtrsim (nh^d)^{-1/2}$ by Lemma SA-4, $\mathbb{E}[|Y_i|^3|\mathbf{X}_i] \lesssim 1$ by Assumption SA–1(v), and $\max_{1\leq i\leq n}\sup_{\mathbf{x}\in\mathscr{B}}|\theta_{t,\mathbf{x}}^*(D_i(\mathbf{x}))| \lesssim 1$ because

$$\theta_{t,\mathbf{x}}^*(D_i(\mathbf{x})) = \gamma_t^*(\mathbf{x})^\top\mathbf{r}_p(D_i(\mathbf{x})) = (\mathbf{\Psi}_{t,\mathbf{x}}\mathbf{S}_{t,\mathbf{x}})^{-1}\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big).$$

Since

$$\big|\widehat{\mathrm{T}}(\mathbf{x}) - \overline{\mathrm{T}}(\mathbf{x})\big| \lesssim_{\mathbb{P}} \frac{1}{\sqrt{nh^d}} + \frac{1}{n^{\frac{v}{2+v}}h^d} + \sqrt{nh^d}|\mathfrak{B}(\mathbf{x})|,$$

the pointwise asymptotic normality follows, under the conditions imposed. Finally, validity of the confidence interval estimator is immediate. $\square$

SA-6.10 Proof of Theorem SA-5

We make the decomposition based on Equation (SA-1) and convergence of $\widehat{\Xi}_{\mathbf{x},\mathbf{x}}$,

$$\begin{aligned}
\widehat{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) - \overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) &= \widehat{\Xi}_{\mathbf{x},\mathbf{x}}^{-1/2}\Big(\sum_{t\in\{0,1\}}(-1)^{\frac{t+1}{2}}(\widehat{\theta}_{t,\mathbf{x}}(0) - \theta_{t,\mathbf{x}}(0))\Big) - \Xi_{\mathbf{x},\mathbf{x}}^{-1/2}\Big(\sum_{t\in\{0,1\}}(-1)^{\frac{t+1}{2}}\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}}\Big) \\
&= \widehat{\Xi}_{\mathbf{x},\mathbf{x}}^{-1/2}\Big(\sum_{t\in\{0,1\}}(-1)^{\frac{t+1}{2}}(\widehat{\theta}_{t,\mathbf{x}}(0) - \theta_{t,\mathbf{x}}(0)) - \sum_{t\in\{0,1\}}(-1)^{\frac{t+1}{2}}\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}}\Big) \qquad (= \Delta_{1,\mathbf{x}}) \\
&\quad + (\widehat{\Xi}_{\mathbf{x},\mathbf{x}}^{-1/2} - \Xi_{\mathbf{x},\mathbf{x}}^{-1/2})\sum_{t\in\{0,1\}}(-1)^{\frac{t+1}{2}}\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}} \qquad (= \Delta_{2,\mathbf{x}})
\end{aligned}$$

By Lemma SA-2 and SA-3, and the decomposition Equation (SA-1),

$$\begin{aligned}
&\sup_{\mathbf{x}\in\mathscr{X}}\left|\sum_{t\in\{0,1\}}(-1)^{\frac{t+1}{2}}(\widehat{\theta}_{t,\mathbf{x}}(0)-\theta_{t,\mathbf{x}}(0))-\sum_{t\in\{0,1\}}(-1)^{\frac{t+1}{2}}\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}}\right|\\
&\lesssim_{\mathbb{P}}\sqrt{\frac{\log(1/h)}{nh^d}}\left(\sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}\right)+\sup_{\mathbf{x}\in\mathscr{B}}\sum_{t\in\{0,1\}}|\theta^*_{t,\mathbf{x}}(0)-\theta_{t,\mathbf{x}}(0)|.
\end{aligned}$$

Together with Lemma SA-4,

$$\sup_{\mathbf{x}\in\mathscr{B}}|\Delta_{1,\mathbf{x}}|\lesssim_{\mathbb{P}}\frac{\log(1/h)}{\sqrt{nh^d}}+\frac{(\log(1/h))^{\frac{3}{2}}}{n^{\frac{1+v}{2+v}}h^d}+\sqrt{nh^d}\sup_{\mathbf{x}\in\mathscr{B}}\sum_{t\in\{0,1\}}|\theta^*_{t,\mathbf{x}}(0)-\theta_{t,\mathbf{x}}(0)|. \tag{SA-10}$$

By Lemma SA-2, Lemma SA-3 and Lemma SA-4, and assume $\frac{n^{\frac{v}{2+v}}h^d}{\log(1/h)}\to\infty$, then

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathscr{X}}\left|\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}}\left(\Xi_{\mathbf{x},\mathbf{x}}^{-1/2}-\widehat{\Xi}_{\mathbf{x},\mathbf{x}}^{-1/2}\right)\right|&\lesssim_{\mathbb{P}}\sqrt{nh^d}\left(\sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}\right)\left(\sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}\right)\\
&=\sqrt{\log(1/h)}\left(1+\sqrt{\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}}\right)\left(\sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}\right)\\
&\lesssim\sqrt{\log(1/h)}\left(\sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}\right).
\end{aligned}$$

Hence

$$\sup_{\mathbf{x}\in\mathscr{B}}|\Delta_{2,\mathbf{x}}|\lesssim_{\mathbb{P}}\frac{\log(1/h)}{\sqrt{nh^d}}+\frac{(\log(1/h))^{\frac{3}{2}}}{n^{\frac{1+v}{2+v}}h^d}. \tag{SA-11}$$

Putting together Equations (SA-10), (SA-11) give the result. □

SA-6.11 Proof of Theorem SA-6

We will verify the high level conditions stated in Theorem SA-8.

Without loss of generality, we can assume $\mathscr{X}=[0,1]^d$, and $\mathscr{Q}_{\mathscr{F}_t}=\mathbb{P}_X$ is a valid surrogate measure for $\mathbb{P}_X$ with respect to $\mathscr{G}$, and $\phi_{\mathscr{G}}=\mathrm{Id}$ is a valid normalizing transformation (as in Theorem SA-8). This implies the constants $\mathtt{c}_1$ and $\mathtt{c}_2$ from Theorem SA-8 are all 1.

Recall $\mathscr{G}=\{g_{\mathbf{x}}:\mathbf{x}\in\mathscr{B}\}$ where

$$g_{\mathbf{x}}(\mathbf{u})=\mathbb{1}(\mathbf{u}\in\mathscr{A}_1)\mathfrak{K}_1(\mathbf{u};\mathbf{x})-\mathbb{1}(\mathbf{u}\in\mathscr{A}_0)\mathfrak{K}_0(\mathbf{u};\mathbf{x}).$$

By standard arguments and [Cattaneo et al., 2024, Lemma 7], we get properties of $\mathscr{G}$ as follows:

$$\mathtt{M}_{\mathscr{G}}\lesssim h^{-d/2},\qquad \mathtt{E}_{\mathscr{G}}\lesssim h^{d/2},\qquad \mathtt{TV}_{\mathscr{G}}\lesssim h^{d/2-1},\qquad \sup_Q N(\mathscr{G},\|\cdot\|_{Q,2},\varepsilon(2c+1)^{d+1}\mathtt{M}_{\mathscr{G}})\leq 2\mathbf{c}'\varepsilon^{-d-1}+2.$$

By definition of $\theta^*_{t,\mathbf{x}}(\cdot)$, for each $\mathbf{x}\in\mathscr{B}$, $t\in\{0,1\}$,

$$\theta^*_{t,\mathbf{x}}(\mathscr{d}(\mathbf{u},\mathbf{x}))=\boldsymbol{\gamma}^*_t(\mathbf{x})^\top\mathbf{r}_p(\mathscr{d}(\mathbf{u},\mathbf{x}))=(\mathbf{H}^{-1}\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{S}_{t,\mathbf{x}})^\top\mathbf{r}_p(\mathscr{d}(\mathbf{u},\mathbf{x}))=(\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{S}_{t,\mathbf{x}})^\top\mathbf{r}_p\Big(\frac{\mathscr{d}(\mathbf{u},\mathbf{x})}{h}\Big),$$

recalling

$$\boldsymbol{\Psi}_{t,\mathbf{x}} = \mathbb{E}\left[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)^\top K_h(D_i(\mathbf{x}))\mathbb{1}(D_i(\mathbf{x}) \in \mathscr{I}_t)\right], \quad \mathbf{S}_{t,\mathbf{x}} = \mathbb{E}\left[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)K_h(D_i(\mathbf{x}))Y_i\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\right].$$

We can check that $\left\|\boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\right\| \lesssim 1$, $\|\mathbf{S}_{t,\mathbf{x}}\| \lesssim 1$ and

$$\mathtt{M}_{\mathscr{M}_t} \lesssim h^{-d/2}, \qquad \mathtt{E}_{\mathscr{M}_t} \lesssim h^{-d/2}, \qquad t \in \{0,1\}.$$

In what follows, we verify the entropy and total variation properties of $\mathscr{M}$. Using product rule we can verify

$$\sup_{\mathbf{u}\in\mathscr{X}} \sup_{\mathbf{x},\mathbf{x}'\in\mathscr{B}} \frac{|\theta^*_{t,\mathbf{x}}(\mathscr{d}(\mathbf{u},\mathbf{x})) - \theta^*_{t,\mathbf{x}}(\mathscr{d}(\mathbf{u},\mathbf{x}'))|}{\|\mathbf{x}-\mathbf{x}'\|} \lesssim h^{-1}.$$

Define $f_{t,\mathbf{x}}(\cdot) = \frac{h^{-d/2}}{\sqrt{n\Xi_{\mathbf{x},\mathbf{x}}}}\mathbf{e}_1^\top \boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{r}_p(\cdot)\, K(\cdot)(\boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{S}_{t,\mathbf{x}})^\top \mathbf{r}_p(\cdot)$. Then,

$$\mathfrak{K}_t(\mathbf{u};\mathbf{x})\theta^*_{t,\mathbf{x}}(\mathscr{d}(\mathbf{u},\mathbf{x})) = h^{-d/2} f_{t,\mathbf{x}}\Big(\frac{d(\mathbf{u},\mathbf{x})}{h}\Big), \qquad \mathbf{u}\in\mathscr{X}, \mathbf{x}\in\mathscr{B}, t\in\{0,1\}.$$

Take $\mathscr{M}_t = \{\mathfrak{K}_t(\cdot;\mathbf{x})\theta^*_{t,\mathbf{x}}(\mathscr{d}(\cdot,\mathbf{x})) : \mathbf{x}\in\mathscr{B}\}$, $t\in\{0,1\}$. For $t \in\{0,1\}$, $f_{t,\mathbf{x}}$ satisfies:

(i) boundedness $\quad \sup_{\mathbf{x}\in\mathscr{B}}\sup_{\mathbf{u}\in\mathscr{X}} |f_{t,\mathbf{x}}(\mathbf{u})| \leq \mathbf{c}$,

(ii) compact support $\quad \operatorname{supp}(f_{t,\mathbf{x}}(\cdot)) \subseteq [-\mathbf{c},\mathbf{c}]^d, \forall \mathbf{x}\in\mathscr{B}$,

(iii) Lipschitz continuity $\quad \sup_{\mathbf{x}\in\mathscr{B}}\sup_{\mathbf{u},\mathbf{u}'\in\mathscr{X}} \frac{|f_{\mathbf{x}}(\mathbf{u}) - f_{\mathbf{x}}(\mathbf{u}')|}{\|\mathbf{u}-\mathbf{u}'\|} \leq \mathbf{c}$

$$\sup_{\mathbf{u}\in\mathbf{X}}\sup_{\mathbf{x},\mathbf{x}'\in\mathscr{B}} \frac{|f_{\mathbf{x}}(\mathbf{u}) - f_{\mathbf{x}'}(\mathbf{u})|}{\|\mathbf{x}-\mathbf{x}'\|} \leq \mathbf{c}h^{-1},$$

for some constant $\mathbf{c}$ not depending on n. Then, by an argument similar to Cattaneo et al. [2024, Lemma 7], there exists a constant $\mathbf{c}'$ only depending on $\mathbf{c}$ and d that for any $0 \leq \varepsilon \leq 1$,

$$\sup_Q N\left(h^{d/2}\mathscr{H}_t, \|\cdot\|_{Q,1}, (2c+1)^{d+1}\varepsilon\right) \leq \mathbf{c}'\varepsilon^{-d-1} + 1,$$

where supremum is taken over all finite discrete measures. Taking a constant envelope function $\mathtt{M}_{\mathscr{M}_t} = (2c+1)^{d+1}h^{-d/2}$, we have for any $0 < \varepsilon \leq 1$,

$$\sup_Q N\left(\mathscr{H}_t, \|\cdot\|_{Q,1}, \varepsilon\mathtt{M}_{\mathscr{F}_t}\right) \leq \mathbf{c}'\varepsilon^{-d-1} + 1.$$

By Lemma SA-6, above implies the uniform covering number for $\mathscr{H}_t$ satisfies

$$\mathtt{N}_{\mathscr{M}_t}(\varepsilon) \leq 4\mathbf{c}'(\varepsilon/2)^{-d-1}, \qquad 0 < \varepsilon \leq 1.$$

Since $\mathscr{M} \subseteq \mathscr{M}_0 + \mathscr{M}_1$, here $+$ denotes the Minkowski sum, with $\mathtt{M}_{\mathscr{M}}$ taken to be $\mathtt{M}_{\mathscr{M}_0} + \mathtt{M}_{\mathscr{M}_1}$, a bound on the uniform covering number of $\mathscr{M}$ can be given by

$$\mathtt{N}_{\mathscr{M}}(\varepsilon) \leq 16(\mathbf{c}')^2(\varepsilon/2)^{-2d-2}, \qquad 0 < \varepsilon \leq 1.$$

With the assumption that $\mathscr{L}(E_{t,\mathbf{x}}) \leq Ch^{d-1}$ for $E_{t,\mathbf{x}} = \{\mathbf{y} \in \mathscr{A}_t : (\mathbf{y}-\mathbf{x})/h \in \mathrm{Supp}(K)\}$ for all $t \in \{0,1\}$, $\mathbf{x} \in \mathscr{B}$, and the fact that $\mathtt{TV}_{\mathscr{M}_t} \lesssim h^{d/2-1}$ for $t \in \{0,1\}$, the same argument as in the paragraph **Total Variation** in the proof of Theorem SA-8 shows

$$\mathtt{TV}_{\mathscr{M}} \lesssim h^{d/2-1}.$$

Now apply Theorem SA-8 with $\mathscr{G}$, $\mathscr{M}$ defined in Equation (SA-2), $\mathscr{R} = \{\mathrm{Id}\}$, $\mathscr{S} = \{1\}$, noticing that

$$(\overline{\mathrm{T}}_{\mathrm{dis}} : \mathbf{x} \in \mathscr{B}) = (A_n(g,m,r,s) : (g,m,r,s) \in \mathscr{F} \times \mathscr{R} \times \mathscr{S}), \qquad \mathscr{F} = \{(g_{\mathbf{x}}, m_{\mathbf{x}}) : \mathbf{x} \in \mathscr{B}\} \subseteq \mathscr{G} \times \mathscr{M},$$

the result then follows. □

Lemma SA-6 (VC Class to VC2 Class)**.** *Assume $\mathscr{F}$ is a VC class on a measure space $(\mathcal{X}, \mathscr{B})$: there exists an envelope function F and positive constants $c(\mathscr{F}), d(\mathscr{F})$ such that for all $\varepsilon \in (0,1)$,*

$$\sup_Q N(\mathscr{F}, \|\cdot\|_{Q,1}, \varepsilon \|F\|_{Q,1}) \leq c(\mathscr{F})\varepsilon^{-d(\mathscr{F})},$$

where the supremum is taken over all finite discrete measures. Then, $\mathscr{F}$ is also VC2 class: for all $\varepsilon \in (0,1)$,

$$\sup_Q N(\mathscr{F}, \|\cdot\|_{Q,2}, \varepsilon \|F\|_{Q,2}) \leq c(\mathscr{F})(\varepsilon^2/2)^{-d(\mathscr{F})},$$

where the supremum is taken over all finite discrete measures.

Proof of Lemma SA-6. Let Q be a finite discrete probability measure. Let $f, g \in \mathscr{F}$. Then, $\int |f-g|^2 dQ \leq 2\int |f-g||F| dQ$. Define another probability measure $\tilde{Q}(c_k) = F(c_k)Q(c_k)/\|F\|_{Q,1}$ on the support of Q, denoted by $\{c_1, \ldots, c_k, \ldots\}$. Then,

$$\int |f-g|^2 dQ \leq 2\|F\|_{Q,1} \int |f-g| d\tilde{Q} \leq 2\|F\|_{Q,1} \|f-g\|_{\tilde{Q},1}.$$

Hence, if we take an $\varepsilon^2/2$-net in $(\mathscr{F}, \|\cdot\|_{\tilde{Q},1})$ with cardinality no greater than $c(\mathscr{F})\varepsilon^{-d(\mathscr{F})}$, then for any $f \in \mathscr{F}$, there exists a $g \in \mathscr{F}$ such that $\|f-g\|_{\tilde{Q},1} \leq \varepsilon^2/2 \|F\|_{\tilde{Q},1}$, and hence

$$\|f-g\|_{Q,2}^2 \leq 2\varepsilon^2/2 \|F\|_{Q,1} \|F\|_{\tilde{Q},1} \leq \varepsilon^2 \|F\|_{Q,2}^2,$$

which gives the result. □

SA-6.12 Proof of Theorem SA-7

The result follows from Theorems SA-5 and SA-6, Chernozhukov et al. [2014a], and Chernozhuokov et al. [2022]. □

SA-6.13 Proof of Theorem SA-8

Since A_n is the addition of two M_n processes, indexed by $\mathscr{G} \times \mathscr{R}$ and $\mathscr{H} \times \mathscr{S}$ respectively, the Gaussian strong approximation error essentially depends on the *worst case scenario* between $\mathscr{G}$ and $\mathscr{H}$, and between $\mathscr{R}$ and $\mathscr{S}$. Hence (1) taking maximums $\mathtt{E} = \max\{\mathtt{E}_{\mathscr{G}}, \mathtt{E}_{\mathscr{H}}\}$, $\mathtt{M} = \max\{\mathtt{M}_{\mathscr{G}}, \mathtt{M}_{\mathscr{H}}\}$ and $\mathtt{TV} = \max\{\mathtt{TV}_{\mathscr{G}}, \mathtt{TV}_{\mathscr{H}}\}$; (2) noticing that A_n is still indexed by a VC-type class of functions, we can get the claimed result.

For a more rigor proof, we can not apply Cattaneo and Yu [2025, Theorem SA.1] on $(M_n(g,r) : g \in \mathscr{G}, r \in \mathscr{R})$ and $(M_n(h,s) : h \in \mathscr{H}, s \in \mathscr{S})$ directly, since this ignores the dependence structure between the two empirical processes. However, we can still project the functions onto a Haar basis, and control the *strong approximation error for projected process* and the *projection error* as in the proof of Cattaneo and Yu [2025, Theorem SA.1] and show both errors can be controlled via *worst case scenario* between $\mathscr{G}$ and $\mathscr{H}$, and between $\mathscr{R}$ and $\mathscr{S}$.

Reductions: Here we present some reductions to our problem. By the same argument as in Section SA-II.3 (Proofs of Theorem 1) in the supplemental appendix of Cattaneo and Yu [2025], we can show there exists $\mathbf{u}_i, 1 \leq i \leq n$ i.i.d $\mathsf{Uniform}([0,1]^d)$ on a possibly enlarged probability space, such that

$$f(\mathbf{x}_i) = f(\phi_{\mathscr{G}\cup\mathscr{H}}^{-1}(\mathbf{u}_i)), \qquad \forall f \in \mathscr{G} \cup \mathscr{H}, \forall 1 \leq i \leq n.$$

With the help of Cattaneo and Yu [2025, Lemma SA.10], we can assume w.l.o.g. that $\mathbf{x}_i$'s are i.i.d $\mathsf{Uniform}(\mathscr{X})$ with $\mathscr{X} = [0,1]^d$, and $\phi_{\mathscr{G}\cup\mathscr{H}} : [0,1]^d \to [0,1]^d$ is the identity function. Although we assume $\sup_{\mathbf{x}\in\mathscr{X}} \mathbb{E}[|Y_i|^{2+v}|\mathbf{X}_i = \mathbf{x}] < \infty$, we first present the result under the assumption $\sup_{\mathbf{x}\in\mathscr{X}} \mathbb{E}[\exp(|y_i|)|\mathbf{x}_i = \mathbf{x}] \leq 2$, which is the same as in Cattaneo and Yu [2025, Theorem 2]. Also in correspondence to the notations in Cattaneo and Yu [2025, Theorem 2], we set $\alpha = 1$ throughout this proof.

Cell Constructions and Projections: The constructions here are the same as those in Cattaneo and Yu [2025], and we present them here for completeness. Let $\mathscr{A}_{M,N}(\mathbb{P},1) = \{\mathscr{C}_{j,k} : 0 \leq k < 2^{M+N-j}, 0 \leq j \leq M+N\}$ be an axis-aligned cylindered quasi-dyadic expansion of $\mathbb{R}^{d+1}$, with depth M for the main subspace $\mathbb{R}^d$ and depth N for the multiplier subspace $\mathbb{R}$, with respect to $\mathbb{P}$, the joint distribution of $(\mathbf{x}_i, y_i)$ taking values in $\mathbb{R}^d \times \mathbb{R}$, as in Cattaneo and Yu [2025, Definition SA.4]. To see what $\mathscr{A}_{M,N}(\mathbb{P},1)$ is, it can be given by the following iterative partition procedure:

1. *Initialization ($q = 0$):* Take $\mathscr{C}_{M+N-q,0} = \mathscr{X} \times \mathbb{R}$ where $\mathscr{X} = [0,1]^d$.

2. *Iteration ($q = 1, \ldots, M$):* Given $\mathscr{C}_{K-l,k}$ for $0 \leq l \leq q-1, 0 \leq k < 2^l$, take $s = (q \mod d) + 1$, and construct $\mathscr{C}_{K-q,2k} = \mathscr{C}_{K-q+1,k} \cap \{(\mathbf{x}, y) \in [0,1]^d \times \mathbb{R} : \mathbf{e}_s^\top \mathbf{x} \leq c_{K-q+1,k}\}$ and $\mathscr{C}_{K-q,2k+1} = \mathscr{C}_{K-q+1,k} \cap \{(\mathbf{x}, y) \in [0,1]^d \times \mathbb{R} : \mathbf{e}_s^\top \mathbf{x} > c_{K-j+1,k}\}$ such that $\mathbb{P}(\mathscr{C}_{K-q,2k})/\mathbb{P}(\mathscr{C}_{K-q+1,k}) \in [\frac{1}{1+\rho}, \frac{\rho}{1+\rho}]$ for all $0 \leq k < 2^{q-1}$. Continue until $(\mathscr{C}_{N,k} : 0 \leq k < 2^M)$ has been constructed. By construction, for each $0 \leq l < M$, $\mathscr{C}_{N,l} = \mathscr{X}_{0,l} \times \mathscr{Y}_{0,N,0}$, with $\mathscr{Y}_{0,N,0} = \mathbb{R}$.

3. *Iteration ($q = M+1, \cdots, M+N$):* Given $\mathscr{C}_{K-l,k}$ for $0 \leq l \leq q-1, 0 \leq k < 2^l$, each $\mathscr{C}_{M+N-q,k}$ can be written as $\mathscr{X}_{0,l} \times \mathscr{Y}_{l,M+N-q,m}$ with $k = 2^{q-M}l + m$. Construct $\mathscr{C}_{M+N-q-1,2k} = \mathscr{X}_{0,l} \times \mathscr{Y}_{l,M+N-q-1,2m}$ and $\mathscr{C}_{M+N-q-1,2k+1} = \mathscr{X}_{0,l} \times \mathscr{Y}_{l,M+N-q-1,2m+1}$, such that there exists some $\mathfrak{q}_{M+N-q,k} \in \mathbb{R}$ with $\mathscr{Y}_{l,M+N-q-1,2m} = \mathscr{Y}_{l,M+N-q,m} \cap (-\infty, \mathfrak{q}_{M+N-q,k})$ and $\mathscr{Y}_{l,M+N-q-1,2m+1} = \mathscr{Y}_{l,M+N-q,m} \cap (\mathfrak{q}_{M+N-q,k}, \infty)$, $\mathbb{P}(y_i \in \mathscr{Y}_{l,M+N-q-1,2m}|\mathbf{x}_i \in \mathscr{X}_{0,l}) = \mathbb{P}(y_i \in \mathscr{Y}_{l,M+N-q-1,2m+1}|\mathbf{x}_i \in \mathscr{X}_{0,l}) = \frac{1}{2}\mathbb{P}(y_i \in \mathscr{Y}_{l,M+N-q-1,m}|\mathbf{x}_i \in \mathscr{X}_{0,l})$.

Consider the projection $\Pi_1(\mathscr{A}_{M,n}(\mathbb{P},1))$ given in Equation (SA-7) in Cattaneo and Yu [2025], noticing that $\mathscr{A}_{M,N}(\mathbb{P},1)$ is one special instance of $\mathscr{C}_{M,N}(\mathbb{P},\rho)$. That is, define $e_{j,k} = \mathbb{1}_{\mathscr{C}_{j,k}}$ and $\widetilde{e}_{j,k} = e_{j-1,2k} - e_{j-1,2k+1}$,

$$\Pi_1(\mathscr{C}_{M,N}(\mathbb{P},\rho))[g,r] = \gamma_{M+N,0}(g,r)e_{M+N,0} + \sum_{1\leq j\leq M+N} \sum_{0\leq k<2^{M+N-j}} \widetilde{\gamma}_{j,k}(g,r)\widetilde{e}_{j,k}, \tag{SA-12}$$

where $e_{j,k} = \mathbb{1}(\mathscr{C}_{j,k})$ and $\widetilde{e}_{j,k} = \mathbb{1}(\mathscr{C}_{j-1,2k}) - \mathbb{1}(\mathscr{C}_{j-1,2k+1})$, and

$$\gamma_{j,k}(g,r) = \begin{cases} \mathbb{E}[g(X)r(Y)|X \in \mathscr{X}_{j-N,k}], & \text{if } N \leq j \leq M+N, \\ \mathbb{E}[g(X)|X \in \mathscr{X}_{0,l}] \cdot \mathbb{E}[r(Y)|X \in \mathscr{X}_{0,l}, Y \in \mathscr{Y}_{l,0,m}], & \text{if } j < N, k = 2^{N-j}l + m, \end{cases}$$

and $\widetilde{\gamma}_{j,k}(g,r) = \gamma_{j-1,2k}(g,r) - \gamma_{j-1,2k+1}(g,r)$. We will use Π_1 as a shorthand for $\Pi_1(\mathscr{C}_{M,N}(\mathbb{P},\rho))$.

For simplicity, we denote $\Pi_1(\mathscr{A}_{M,n}(\mathbb{P},1))$ by Π_1 instead. Now define the projected empirical process

$$\Pi_1 A_n(g,h,r,s) = \Pi_1 M_n(g,r) + \Pi_1 M_n(h,s), \qquad g \in \mathscr{G}, h \in \mathscr{H}, r \in \mathscr{R}, s \in \mathscr{S},$$

where $\Pi_1 M_n(g,r)$ and $\Pi_1 M_n(h,s)$ are given in Equation (SA-10) in Cattaneo and Yu [2025], that is,

$$\Pi_1 M_n(g,r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\Pi_1[g,r](\mathbf{x}_i, y_i) - \mathbb{E}[\Pi_1[g,r](\mathbf{x}_i, y_i)]\big),$$
$$\Pi_1 M_n(h,s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\Pi_1[h,s](\mathbf{x}_i, y_i) - \mathbb{E}[\Pi_1[h,s](\mathbf{x}_i, y_i)]\big).$$

Construction of Gaussian Process Suppose $(\widetilde{\xi}_{j,k} : 0 \leq k < 2^{M+N-j}, 1 \leq j \leq M+N)$ are i.i.d. standard Gaussian random variables. Take $F_{(j,k),m}$ to be the cumulative distribution function of $(S_{j,k} - mp_{j,k})/\sqrt{mp_{j,k}(1-p_{j,k})}$, where $p_{j,k} = \mathbb{P}(\mathscr{C}_{j-1,2k})/\mathbb{P}(\mathscr{C}_{j,k})$ and $S_{j,k}$ is a $\mathrm{Bin}(m, p_{j,k})$ random variable, and $G_{(j,k),m}(t) = \sup\{x : F_{(j,k),m}(x) \leq t\}$. We define $U_{j,k}, \widetilde{U}_{j,k}$'s via the following iterative scheme:

1. *Initialization:* Take $U_{M+N,0} = n$.
2. *Iteration:* Suppose we've defined $U_{l,k}$ for $j < l \leq M+N, 0 \leq k < 2^{M+N-l}$, then solve for $U_{j,k}$'s s.t.

$$\widetilde{U}_{j,k} = \sqrt{U_{j,k} p_{j,k}(1-p_{j,k})} G_{(j,k),U_{j,k}} \circ \Phi(\widetilde{\xi}_{j,k}),$$
$$\widetilde{U}_{j,k} = (1-p_{j,k})U_{j-1,2k} - p_{j,k}U_{j-1,2k+1} = U_{j-1,2k} - p_{j,k}U_{j,k},$$
$$U_{j-1,2k} + U_{j-1,2k+1} = U_{j,k}, \quad 0 \leq k < 2^{M+N-j}.$$

Continue till we have defined $U_{0,k}$ for $0 \leq k < 2^{M+N}$.

Then, $\{U_{j,k} : 0 \leq j \leq K, 0 \leq k < 2^{M+N-j}\}$ have the same joint distribution as $\{\sum_{i=1}^n e_{j,k}(\mathbf{x}_i, y_i) : 0 \leq j \leq K, 0 \leq k < 2^{M+N-j}\}$. By Vorob'ev–Berkes–Philipp theorem [Dudley, 2014, Theorem 1.31], $\{\widetilde{\xi}_{j,k} : 0 \leq k < 2^{M+N-j}, 1 \leq j \leq M+N\}$ can be constructed on a possibly enlarged probability space such that the previously constructed $U_{j,k}$ satisfies $U_{j,k} = \sum_{i=1}^n e_{j,k}(\mathbf{x}_i)$ almost surely for all $0 \leq j \leq M+N, 0 \leq k < 2^{M+N-j}$. We will show $\widetilde{\xi}_{j,k}$'s can be given as a Brownian bridge indexed by $\widetilde{e}_{j,k}$'s.

Since all of $\mathscr{G}$, $\mathscr{H}$, $\mathscr{R}$ and $\mathscr{S}$ are VC-type, we can show $\mathscr{G} \times \mathscr{H} + \mathscr{R} \times \mathscr{S}$ is also VC-type, here $+$ is the Minkowski sum. Hence $\mathscr{F} = \mathscr{G} \times \mathscr{H} + \mathscr{R} \times \mathscr{S} \cup \Pi_1[G \times \mathscr{H} + \mathscr{R} \times \mathscr{S}]$ is pre-Gaussian.

Then, by Skorohod Embedding lemma [Dudley, 2014, Lemma 3.35], on a possibly enlarged probability space, we can construct a Brownian bridge $(Z_n(f) : f \in \mathscr{F})$ that satisfies

$$\widetilde{\xi}_{j,k} = \frac{\mathbb{P}(\mathscr{C}_{j,k})}{\sqrt{\mathbb{P}(\mathscr{C}_{j-1,2k})\mathbb{P}(\mathscr{C}_{j-1,2k+1})}} Z_n(\widetilde{e}_{j,k}),$$

for $0 \leq k < 2^{M+N-j}, 1 \leq j \leq M+N$. Moreover, call

$$V_{j,k} = \sqrt{n} Z_n(e_{j,k}), \qquad \widetilde{V}_{j,k} = \sqrt{n} Z_n(\widetilde{e}_{j,k}), \qquad \widetilde{\xi}_{j,k} = \frac{\mathbb{P}(\mathscr{C}_{j,k})}{\sqrt{n \mathbb{P}(\mathscr{C}_{j-1,2k}) \mathbb{P}(\mathscr{C}_{j-1,2k+1})}} \widetilde{V}_{j,k}.$$

for $0 \leq k < 2^{K-j}, 1 \leq j \leq K$. We have for $g \in \mathscr{G}, h \in \mathscr{H}, r \in \mathscr{R}, s \in \mathscr{S}$,

$$\begin{aligned}
\sqrt{n} \Pi_1 A_n(g,h,r,s) &= \sum_{j=1}^{M+N} \sum_{0 \leq k < 2^{M+N-j}} (\widetilde{\gamma}_{j,k}[g,r] + \widetilde{\gamma}_{j,k}[h,s]) \widetilde{U}_{j,k}, \\
\sqrt{n} \Pi_1 Z_n(g,h,r,s) &= \sum_{j=1}^{M+N} \sum_{0 \leq k < 2^{M+N-j}} (\widetilde{\gamma}_{j,k}[g,r] + \widetilde{\gamma}_{j,k}[h,s]) \widetilde{V}_{j,k}.
\end{aligned}$$

Decomposition Fix one $(g,h,r,s) \in \mathscr{G} \times \mathscr{H} \times \mathscr{R} \times \mathscr{S}$, we decompose by

$$\begin{aligned}
& A_n(g,h,r,s) - Z_n(g,h,r,s) \\
= \; & \underbrace{\Pi_1 A_n(g,h,r,s) - \Pi_1 Z_n(g,h,r,s)}_{\text{strong approximation (SA) error for projected}} + \underbrace{A_n(g,h,r,s) - \Pi_1 A_n(g,h,r,s) + \Pi_1 Z_n(g,h,r,s) - Z_n(g,h,r,s)}_{\text{projection error}}.
\end{aligned}$$

SA error for Projected Process The strong approximation error essentially depends on the Hilbertian pseudo norm

$$\sum_{j=1}^{M+N} \sum_{0 \leq k < 2^{M+N-j}} (\widetilde{\gamma}_{j,k}[g,r] + \widetilde{\gamma}_{j,k}[h,s])^2 \leq 2 \sum_{j=1}^{M+N} \sum_{0 \leq k < 2^{M+N-j}} (\widetilde{\gamma}_{j,k}[g,r])^2 + 2 \sum_{j=1}^{M+N} \sum_{0 \leq k < 2^{M+N-j}} (\widetilde{\gamma}_{j,k}[h,s])^2.$$

Hence, Cattaneo and Yu [2025, Lemma SA.19] gives with probability at least $1 - 2e^{-t}$,

$$|\Pi_1 A_n(g,h,r,s) - \Pi_1 Z_n(g,h,r,s)| \leq C_1 C_\alpha \sqrt{\frac{N^{2\alpha+1} 2^M \mathtt{EM}}{n} t} + C_1 C_\alpha \sqrt{\frac{(\|\Pi_1[g,r]\|_\infty + \|\Pi_1[h,s]\|_\infty)^2 (M+N)}{n} t},$$

where $C_1 > 0$ is a universal constant and $C_\alpha = 1 + (2\alpha)^{\alpha/2}$.

Projection Error For the projection error, we use the simple observation that

$$|A_n(g,h,r,s) - \Pi_1 A_n(g,h,r,s)| \leq |M_n(g,r) - \Pi_1 M_n(g,r)| + |M_n(h,s) - \Pi_1 M_n(h,s)|,$$

and Cattaneo and Yu [2025, Lemma SA.23] to get for all $t > N$,

$$\begin{aligned}
\mathbb{P}\Big[|A_n(g,h,r,s) - \Pi_1 A_n(g,h,r,s)| > C_2 \sqrt{C_{2\alpha}} \sqrt{N^2 \mathtt{V} + 2^{-N} \mathtt{M}^2} t^{\alpha + \frac{1}{2}} + C_2 C_\alpha \frac{\mathtt{M}}{\sqrt{n}} t^{\alpha+1}\Big] &\leq 4ne^{-t}, \\
\mathbb{P}\Big[|Z_n(g,h,r,s) - \Pi_1 Z_n(g,h,r,s)| > C_2 \sqrt{C_{2\alpha}} \sqrt{N^2 \mathtt{V} + C_2 C_\alpha 2^{-N} \mathtt{M}^2} t^{\frac{1}{2}} + C_2 C_\alpha \frac{\mathtt{M}}{\sqrt{n}} t\Big] &\leq 4ne^{-t},
\end{aligned}$$

where $C_\alpha = 1 + (2\alpha)^{\frac{\alpha}{2}}$ and $C_{2\alpha} = 1 + (4\alpha)^\alpha$ and C_2 is a constant that only depends on the distribution of $(\mathbf{x}_1, y_1)$, with

$$\mathtt{V} = \min\{2\mathtt{M}, \sqrt{d} \mathtt{L} 2^{-M/d}\} 2^{-M/d} \mathtt{TV}_{\mathscr{H}}.$$

Uniform SA Error: Since all of $\mathscr{G}$, $\mathscr{H}$, $\mathscr{R}$ and $\mathscr{S}$ are VC-type class, from a union bound argument and the same control over fluctuation error as in Cattaneo and Yu [2025, Lemma SA.18], denoting $\mathscr{F} = \mathscr{G}\times\mathscr{H}\times\mathscr{R}\times\mathscr{S}$, we get for all $t > 0$ and $0 < \delta < 1$,

$$\mathbb{P}\big[\,\|A_n - A_n \circ \pi_{\mathscr{F}_\delta}\|_{\mathscr{F}} + \|Z_n - Z_n \circ \pi_{\mathscr{F}_\delta}\|_{\mathscr{F}} > C_1 C_\alpha \mathsf{F}_n(t,\delta)\big] \leq \exp(-t),$$

where $C_\alpha = 1 + (2\alpha)^{\frac{\alpha}{2}}$ and

$$\mathsf{F}_n(t,\delta) = J(\delta)\mathtt{M} + \frac{(\log n)^{\alpha/2}\mathtt{M}J^2(\delta)}{\delta^2\sqrt{n}} + \frac{\mathtt{M}}{\sqrt{n}}t + (\log n)^\alpha \frac{\mathtt{M}}{\sqrt{n}}t^\alpha.$$

where

$$\begin{aligned} J(\delta) &= 3\delta\Big(\sqrt{\mathtt{d}_{\mathscr{G}}\log(\frac{2\mathtt{c}_{\mathscr{G}}}{\delta})} + \sqrt{\mathtt{d}_{\mathscr{H}}\log(\frac{2\mathtt{c}_{\mathscr{H}}}{\delta})} + \sqrt{\mathtt{d}_{\mathscr{R}}\log(\frac{2\mathtt{c}_{\mathscr{R}}}{\delta})} + \sqrt{\mathtt{d}_{\mathscr{S}}\log(\frac{2\mathtt{c}_{\mathscr{S}}}{\delta})}\Big) \\ &\lesssim \sqrt{\mathtt{d}\log(\mathtt{c}/\delta)}, \end{aligned}$$

recalling $\mathtt{c} = \mathtt{c}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}} + \mathtt{c}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}} + \mathtt{c}_{\mathscr{R},\mathscr{Y}} + \mathtt{c}_{\mathscr{S},\mathscr{Y}} + \mathtt{k}$, $\mathtt{d} = \mathtt{d}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\mathtt{d}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\mathtt{d}_{\mathscr{R},\mathscr{Y}}\mathtt{d}_{\mathscr{S},\mathscr{Y}}\mathtt{k}$. Choosing the optimal M^*, N^* gives $\mathbb{P}\big[\,\big\|A_n - Z_n^A\big\|_{\mathscr{F}} > C_1\mathtt{v}\mathsf{T}_n(t)\big] \leq C_2 e^{-t}$ for all $t > 0$, where

$$\mathsf{T}_n(t) = \min_{\delta\in(0,1)}\{\mathsf{A}_n(t,\delta) + \mathsf{F}_n(t,\delta)\},$$

with

$$\begin{aligned} \mathsf{A}_n(t,\delta) &= \sqrt{d}\min\Big\{\Big(\frac{\mathtt{c}_1^d\mathtt{E}\mathtt{TV}^d\mathtt{M}^{d+1}}{n}\Big)^{\frac{1}{2(d+1)}}, \Big(\frac{\mathtt{c}_1^d\mathtt{c}_2^d\mathtt{E}^2\mathtt{M}^2\mathtt{TV}^d\mathtt{L}^d}{n^2}\Big)^{\frac{1}{2(d+2)}}\Big\}(t + \log(n\mathtt{N}(\delta)N^*))^{\alpha+1} \\ &\quad + \sqrt{\frac{\mathtt{M}^2(M^* + N^*)}{n}}(\log n)^\alpha(t + \log(n\mathtt{N}(\delta)N^*))^{\alpha+1}, \\ \mathsf{F}_n(t,\delta) &= J(\delta)\mathtt{M} + \frac{(\log n)^{\alpha/2}\mathtt{M}J^2(\delta)}{\delta^2\sqrt{n}} + \frac{\mathtt{M}}{\sqrt{n}}\sqrt{t} + (\log n)^\alpha\frac{\mathtt{M}}{\sqrt{n}}t^\alpha, \end{aligned}$$

where

$$\begin{aligned} \mathscr{V}_{\mathscr{R}} &= \{\theta(\cdot, r) : r \in \mathscr{R}\}, \\ \mathtt{N}(\delta) &= \mathtt{N}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}(\delta/2, \mathtt{M}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}})\mathtt{N}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}(\delta/2, \mathtt{M}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}})\mathtt{N}_{\mathscr{R},\mathscr{Y}}(\delta/2, M_{\mathscr{R}})\mathtt{N}_{\mathscr{S},\mathscr{Y}}(\delta/2, M_{\mathscr{S},\mathscr{Y}}), \\ J(\delta) &= 2J_{\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}(\mathscr{G}, \mathtt{M}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}, \delta/2) + 2J_{\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}(\mathscr{H}, \mathtt{M}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}, \delta/2) + 2J_{\mathscr{Y}}(\mathscr{R}, M_{\mathscr{R},\mathscr{Y}}, \delta/2) + 2J_{\mathscr{Y}}(\mathscr{S}, M_{\mathscr{S},\mathscr{Y}}, \delta/2), \\ M^* &= \Big\lfloor \log_2\min\Big\{\Big(\frac{\mathtt{c}_1 n\mathtt{TV}}{\mathtt{E}}\Big)^{\frac{d}{d+1}}, \Big(\frac{\mathtt{c}_1\mathtt{c}_2 n\mathtt{LTV}}{\mathtt{EM}}\Big)^{\frac{d}{d+2}}\Big\}\Big\rfloor, \\ N^* &= \Big\lceil \log_2\max\Big\{\Big(\frac{n\mathtt{M}^{d+1}}{\mathtt{c}_1^d\mathtt{E}\mathtt{TV}^d}\Big)^{\frac{1}{d+1}}, \Big(\frac{n^2\mathtt{M}^{2d+2}}{\mathtt{c}_1^d\mathtt{c}_2^d\mathtt{TV}^d\mathtt{L}^d\mathtt{E}^2}\Big)^{\frac{1}{d+2}}\Big\}\Big\rceil. \end{aligned}$$

Truncation Argument for y_i's with Finite Moments The above result is derived under the assumption that $\sup_{\mathbf{x}\in\mathscr{X}}\mathbb{E}[\exp(|y_i|)|\mathbf{x}_i = \mathbf{x}] < \infty$. For the result under the condition $\sup_{\mathbf{x}\in\mathscr{X}}\mathbb{E}[|y_i|^{2+v}|\mathbf{x}_i = \mathbf{x}] < \infty$, we can use the same truncation argument as in [Cattaneo et al., 2025, Theorem SA-11 in the supplemental material] and the VC-type conditions for $\mathscr{G}, \mathscr{H}, \mathscr{R}, \mathscr{S}$ to get the stated conclusions. $\square$

SA-6.14 Proof of Theorem 2

Part I: Upper Bound.

The proof is essentially the proof for Lemma SA-5 with the data generating process ranging over $\mathscr{P}$. By Theorem SA-1 and Equation (SA-6), we have

$$\begin{aligned}
&\sup_{\mathbb{P}\in\mathscr{P}}\sup_{\mathbf{x}\in\mathscr{B}}|\mathfrak{B}_{n,t}(\mathbf{x})| \\
&= \sup_{\mathbb{P}\in\mathscr{P}}\sup_{\mathbf{x}\in\mathscr{B}}\left|\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{S}_{t,\mathbf{x}}-\mu_t(\mathbf{x})\right| \\
&= \sup_{\mathbb{P}\in\mathscr{P}}\sup_{\mathbf{x}\in\mathscr{B}}\left|\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbb{E}\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)K_h(D_i(\mathbf{x}))\mathbf{r}_p(D_i(\mathbf{x}))^\top(\mu_t(\mathbf{X}_i)-\mu_t(\mathbf{x}),0,\cdots,0))\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)\Big]\right| \\
&\lesssim \sup_{\mathbb{P}\in\mathscr{P}}\sup_{\mathbf{x}\in\mathscr{B}}\sup_{\mathbf{z}\in\mathscr{X}}\left|\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbb{E}\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)K_h(D_i(\mathbf{x}))\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)^\top\right| \\
&\qquad\cdot\sup_{\mathbb{P}\in\mathscr{P}}\sup_{\mathbf{x}\in\mathscr{B}}\sup_{\mathbf{z}\in\mathscr{X}}|\mu_t(\mathbf{x})-\mu_t(\mathbf{z})|\mathbb{1}(K_h(\mathscr{d}(\mathbf{z},\mathbf{x}))>0) \\
&\lesssim h.
\end{aligned}$$

Part II: Lower Bound.

The lower bound is proved by considering the following data generating process. Suppose $\mathbf{X}_i\sim\mathsf{Uniform}([-2,2]^2)$, and $\mu_0(x_1,x_2)=0$ and $\mu_1(x_1,x_2)=x_2$ for all $(x_1,x_2)\in\mathscr{X}=[-2,2]^2$. Suppose $Y_i(0)\sim\mathsf{Normal}(\mu_0(\mathbf{X}_i),1)$ and $Y_i(1)\sim\mathsf{Normal}(\mu_1(\mathbf{X}_i),1)$. Define the treatment and control region by $\mathscr{A}_1=\{(x,y)\in\mathscr{X}:x\geq 0,y\geq 0\}$, $\mathscr{A}_0=\mathscr{X}/\mathscr{A}_1$, $\mathscr{B}=\{(x,y)\in\mathbb{R}:0\leq x\leq 2,y=0\text{ or }x=0,0\leq y\leq 2\}$. Suppose $Y_i=\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_0)Y_i(0)+\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)Y_i(1)$. Suppose we choose $\mathscr{d}$ to be the Euclidean distance and $D_i(\mathbf{x})=\|\mathbf{X}_i-\mathbf{x}\|$. In this case, although the underlying conditional mean functions μ_t, $t\in\{0,1\}$ are smooth, the conditional mean given distance $\theta_{t,\mathbf{x}}$ may not even be differentiable. In this example,

$$\theta_{1,(s,0)}(r)=\begin{cases}\frac{2}{\pi r}, & \text{if } 0\leq r\leq s,\\ \frac{r+s}{\pi-\arccos(s/r)}, & \text{if } r>s.\end{cases}$$

Figure SA-1 plots $r\mapsto\theta_{1,(3/4,0)}(r)$ with the notation $\mathbf{x}_s=(s,0)$.

Under this data generating process, we can show

$$\inf_{0<h<1}\sup_{\mathbf{x}\in\mathscr{B}}\frac{|\mathfrak{B}_{n,1}(\mathbf{x})-\mathfrak{B}_{n,0}(\mathbf{x})|}{h}>0.$$

The proof proceeds in two steps. First, we show a scaling property of the asymptotic bias under our example, which gives a reduction to fixed-h bias calculation. Second, we prove the lower bound via the reduction from previous step.

Step 1: A Scaling Property

Let $0<h<1,0<s<1,0<C<1$. Define $h'=Ch$ and $s'=Cs$. Here C is the scaling factor and denote $\mathbf{x}_s=(s,0)$ and $\mathbf{x}_{s'}=(s',0)$. Denote bias for $\mathbf{x}_{s'}$ under bandwidth h' to be

$$\mathrm{bias}_{n,1}(h',s')=\mathbf{e}_1^\top\mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i((s',0))}{h'}\right)\mathbf{r}_p\left(\frac{D_i((s',0))}{h'}\right)^\top K_{h'}(D_i((s',0)))\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right]^{-1}$$

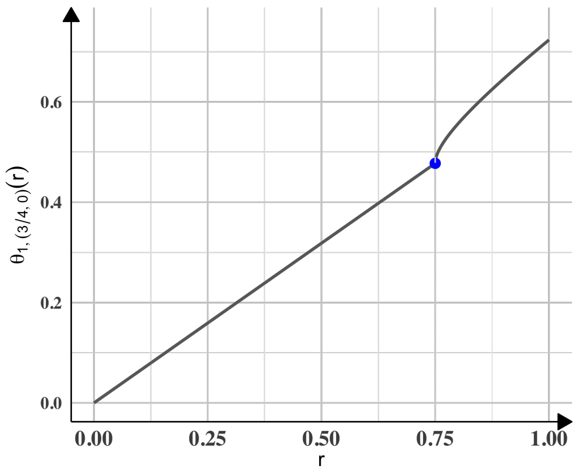

Figure SA-1: Conditional Mean Given Distance with One Kink

$$\mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i((s',0))}{h'}\right)K_{h'}\left(D_i((s',0))\right)\left(\mu_1(\mathbf{X}_i-(s',0))\right)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right], \tag{SA-13}$$

where we have used the fact that μ_1 is linear in our example, hence $\mu_1(\mathbf{X}_i)-\mu_1((s',0))=\mu_1(\mathbf{X}_i-(s',0))$. We reserve the notation $\mathfrak{B}_{n,t}$, $t=0,1$, to the bias when bandwidth is h, that is,

$$\mathfrak{B}_{n,t}(\mathbf{x}_s)\equiv \mathrm{bias}_{n,t}(h,s), \qquad h\in(0,1), s\in(0,1), t=0,1.$$

Inspecting each element of the last vector, for all $l\in\mathbb{N}$,

$$\begin{aligned}
&\mathbb{E}\bigg[\left(\frac{\|\mathbf{X}_i-(s',0)\|}{h'}\right)^l K_{h'}\left(\|\mathbf{X}_i-(s',0)\|\right)\left(\mu_1(\mathbf{X}_i-(s',0))\right)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\bigg]\\
&=\int_0^2\int_0^2\left(\frac{1}{h'}\right)^2\left(\frac{\|(u'-s',v')\|}{h'}\right)^l k\left(\frac{\|(u'-s',v')\|}{h'}\right)\mu_1\left((u',v')-(s',0)\right)\frac{1}{4}du'dv'\\
&\overset{(1)}{=}\int_0^{2/C}\int_0^{2/C}\left(\frac{1}{Ch}\right)^2\left(\frac{\|(Cu-Cs,Cv)\|}{Ch}\right)^l k\left(\frac{\|(Cu-Cs,Cv)\|}{Ch}\right)\mu_1\left(C(u-s,v)\right)\frac{C^2}{4}dudv\\
&=\int_0^{2/C}\int_0^{2/C}\left(\frac{1}{h}\right)^2\left(\frac{\|(u-s,v)\|}{h}\right)^l k\left(\frac{\|(u-s,v)\|}{h}\right)C\mu_1\left((u-s,v)\right)\frac{1}{4}dudv\\
&\overset{(2)}{=}\int_0^2\int_0^2\left(\frac{1}{h}\right)^2\left(\frac{\|(u-s,v)\|}{h}\right)^l k\left(\frac{\|(u,v)-(s,0)\|}{h}\right)C\mu_1\left((u,v)-(s,0)\right)\frac{1}{4}dudv\\
&=C\mathbb{E}\bigg[\left(\frac{\|\mathbf{X}_i-(s,0)\|}{h}\right)^l K_h\left(\|\mathbf{X}_i-(s,0)\|\right)\mu_1(\mathbf{X}_i-(s,0))\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\bigg],
\end{aligned}$$

where in (1) we have used a change of variable $(u,v)=\frac{1}{C}(u',v')$, and (2) holds since $k\left(\frac{\|\cdot-(s,0)\|}{h}\right)$ is supported in $(s,0)+hB(0,1)$, which is contained in $[0,2]\times[0,2]\subseteq[0,2/C]\times[0,2/C]$ for all $0<h<1$, $0<s<1$, $0<C<1$. This means

$$\mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i((s',0))}{h'}\right)K_{h'}\left(D_i((s',0))\right)\left(\mu_1(\mathbf{X}_i-(s',0))\right)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right]$$

$$= C\mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i((s,0))}{h}\right)K_h\left(D_i((s,0))\right)\left(\mu_1(\mathbf{X}_i-(s,0))\right)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right].$$

Similarly, for all $l \in \mathbb{N}$ and $0 < h < 1$, $0 < s < 1$, $0 < C < 1$,

$$\mathbb{E}\left[\left(\frac{D_i((s',0)))}{h'}\right)^l K_{h'}\left(D_i((s',0))\right)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right] = \mathbb{E}\left[\left(\frac{D_i((s,0))}{h}\right)^l K_h\left(D_i((s,0))\right)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right],$$

implying

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i((s',0))}{h'}\right)\mathbf{r}_p\left(\frac{D_i((s',0))}{h'}\right)^\top K_{h'}(D_i((s',0)))\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right]\\
&=\mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i((s,0))}{h}\right)\mathbf{r}_p\left(\frac{D_i((s,0))}{h}\right)^\top K_h(D_i((s,0)))\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right].
\end{aligned}$$

It then follows that for all $0 < h < 1$, $0 < s < 1$, $0 < C < 1$,

$$\mathrm{bias}_{n,1}(h',s') = C\,\mathrm{bias}_{n,1}(h,s).$$

Moreover, for all $0 < h < 1$, $0 < s < h$,

$$\mathfrak{B}_{n,1}(\mathbf{x}_s) = \mathrm{bias}_{n,1}(h,s) = h\,\mathrm{bias}_{n,1}\left(1,\frac{s}{h}\right). \tag{SA-14}$$

Since $\mu_0 \equiv 0$, it is easy to check that

$$\mathfrak{B}_{n,0}(\mathbf{x}_s) = \mathrm{bias}_{n,0}(h,s) \equiv 0, \qquad 0 < h < 1, 0 < s < h.$$

Step 2: Lower Bound on Bias

Now we want to show $\sup_{0\le s\le 1}|\,\mathrm{bias}_{n,1}(1,s)-\mathrm{bias}_{n,0}(1,s)| > 0$. By Equation (SA-13),

$$\begin{aligned}
\mathrm{bias}_{n,1}(1,s)-\mathrm{bias}_{n,0}(1,s) &= \mathbf{e}_1^\top\boldsymbol{\Psi}_s^{-1}\mathbf{S}_s-\mu_1(\mathbf{x}_s)-0=\mathbf{e}_1^\top\boldsymbol{\Psi}_s^{-1}\mathbf{S}_s,\\
\boldsymbol{\Psi}_s &= \mathbb{E}\left[\mathbf{r}_p\left(D_i(\mathbf{x}_s)\right)\mathbf{r}_p\left(D_i(\mathbf{x}_s)\right)^\top K(D_i(\mathbf{x}_s))\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right],\\
\mathbf{S}_s &= \mathbb{E}\left[\mathbf{r}_p\left(D_i(\mathbf{x}_s)\right)K(D_i(\mathbf{x}_s))\mu_1(\mathbf{X}_i)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right].
\end{aligned}$$

Changing to polar coordinates, we have

$$\begin{aligned}
\boldsymbol{\Psi}_s &= \int_0^\infty\int_{\Theta_s(r)}^\pi \mathbf{r}_p(r)\mathbf{r}_p(r)^\top K(r)r\,d\theta\,dr,\\
\mathbf{S}_s &= \int_0^\infty\int_{\Theta_s(r)}^\pi \mathbf{r}_p(r)K(r)r\sin(\theta)r\,d\theta\,dr,
\end{aligned}$$

with

$$\Theta_s(r) = \begin{cases} 0, & \text{if } 0\le r\le s,\\ \arccos(s/r), & \text{if } r > s.\end{cases}$$

For notation simplicity, denote

$$\mathbf{A}(s) = \int_0^\infty \int_{\Theta_s(u)}^\pi \mathbf{r}_p(u)\mathbf{r}_p(u)^\top K(u) u d\theta du = \mathbf{A}_1(s) + \mathbf{A}_2(s),$$
$$\mathbf{B}(s) = \int_0^\infty \int_{\Theta_s(u)}^\pi \mathbf{r}_p(u) K(u) u \sin(\theta) u d\theta du = \mathbf{B}_1(s) + \mathbf{B}_2(s),$$

where

$$\begin{aligned}
\mathbf{A}_1(s) &= \int_0^s \int_0^\pi \mathbf{r}_p(u)\mathbf{r}_p(u)^\top K(u) u d\theta du = \pi \int_0^s \mathbf{r}_p(u)\mathbf{r}_p(u)^\top K(u) u du,\\
\mathbf{A}_2(s) &= \int_s^\infty \int_{\arccos(s/u)}^\pi \mathbf{r}_p(u)\mathbf{r}_p(u)^\top K(u) u d\theta du = \int_s^\infty (\pi - \arccos(s/u))\mathbf{r}_p(u)\mathbf{r}_p(u)^\top K(u) u du,\\
\mathbf{B}_1(s) &= \int_0^s \int_0^\pi \mathbf{r}_p(u) K(u) u \sin(\theta) u d\theta du = 2\int_0^s \mathbf{r}_p(u) K(u) u^2 du,\\
\mathbf{B}_2(s) &= \int_s^\infty \int_{\arccos(s/u)}^\pi \mathbf{r}_p(u) K(u) u \sin(\theta) u d\theta du = \int_s^\infty (1 + \frac{s}{u})\mathbf{r}_p(u) K(u) u^2 du.
\end{aligned}$$

Evaluating the above at zero gives

$$\mathbf{A}(0) = \frac{\pi}{2}\int_0^\infty u\mathbf{r}_p(u)\mathbf{r}_p(u)^\top K(u) du, \quad \mathbf{B}(0) = \int_0^\infty u^2 \mathbf{r}_p(u) K(u) du.$$

Hence

$$\operatorname{bias}_{n,1}(1,0) - \operatorname{bias}_{n,0}(1,0) = \mathbf{e}_1^\top \mathbf{A}(0)^{-1}\mathbf{B}(0) = \mathbf{e}_1^\top \mathbf{A}(0)^{-1}\left[\frac{2}{\pi}\mathbf{A}(0)\mathbf{e}_2\right] = 0. \tag{SA-15}$$

Taking derivatives with respect to s, we have

$$\begin{aligned}
\dot{\mathbf{A}}_1(s) &= \pi \mathbf{r}_p(s)\mathbf{r}_p(s)^\top K(s) s,\\
\dot{\mathbf{A}}_2(s) &= -\pi \mathbf{r}_p(s)\mathbf{r}_p(s)^\top K(s) s + \int_s^\infty \frac{1}{\sqrt{u^2 - s^2}} u\mathbf{r}_p(u)\mathbf{r}_p(u)^\top K(u) du,\\
\dot{\mathbf{B}}_1(s) &= 2\mathbf{r}_p(s) K(s) s^2,\\
\dot{\mathbf{B}}_2(s) &= -2\mathbf{r}_p(s) K(s) s^2 + \int_s^\infty u\mathbf{r}_p(u) K(u) du.
\end{aligned}$$

Evaluating the above at zero gives

$$\dot{\mathbf{A}}(0) = \int_0^\infty \mathbf{r}_p(u)\mathbf{r}_p(u)^\top K(u) du, \quad \dot{\mathbf{B}}(0) = \int_0^\infty u\mathbf{r}_p(u) K(u) du.$$

Using matrix calculus, we know

$$\begin{aligned}
&\frac{d}{ds} \operatorname{bias}_{n,1}(1,s) - \operatorname{bias}_{n,0}(1,s)\Big|_{s=0}\\
&= \frac{d}{ds}\mathbf{e}_1^\top \mathbf{A}(s)^{-1}\mathbf{B}(s)\Big|_{s=0} &\text{(SA-16)}\\
&= -\mathbf{e}_1^\top \mathbf{A}(0)^{-1}\dot{\mathbf{A}}(0)\big[\mathbf{A}(0)^{-1}\mathbf{B}(0)\big] + \mathbf{e}_1^\top \mathbf{A}(0)^{-1}\dot{\mathbf{B}}(0) &\text{(SA-17)}
\end{aligned}$$

$$
\begin{aligned}
&= -\mathbf{e}_1^\top \mathbf{A}(0)^{-1}\dot{\mathbf{A}}(0)\left[\frac{2}{\pi}\mathbf{e}_2\right] + \mathbf{e}_1^\top\left[\frac{2}{\pi}\mathbf{e}_1\right] \\
&= -\frac{2}{\pi}\mathbf{e}_1^\top \mathbf{A}(0)^{-1}\int_0^\infty \begin{bmatrix} u \\ u^2 \\ \cdots \\ u^{p+1}\end{bmatrix} K(u)du + \mathbf{e}_1^\top\left[\frac{2}{\pi}\mathbf{e}_1\right] \\
&= -\frac{4}{\pi^2} + \frac{2}{\pi}. \qquad \text{(SA-18)}
\end{aligned}
$$

Combining Equations (SA-15) and (SA-16), and the fact that $\frac{d}{ds}\operatorname{bias}_{n,1}(1,s) - \operatorname{bias}_{n,0}(1,s)$ is continuous in s, we can show $\sup_{0\le s\le 1}|\operatorname{bias}_{n,1}(1,s) - \operatorname{bias}_{n,0}(1,s)| > 0$. Combining with Equation (SA-14), we have

$$
\begin{aligned}
\inf_{0<h<1}\sup_{\mathbf{x}\in\mathscr{B}} \frac{|\mathfrak{B}_{n,1}(\mathbf{x}) - \mathfrak{B}_{n,0}(\mathbf{x})|}{h} &\ge \inf_{0<h<1}\sup_{0<s<h}\frac{|\operatorname{bias}_{n,1}(s,h) - \operatorname{bias}_{n,0}(s,h)|}{h} \\
&= \inf_{0<h<1}\sup_{0<s<h}\left|\operatorname{bias}_{n,1}\left(1,\frac{s}{h}\right)\right| \\
&> 0.
\end{aligned}
$$

□

SA-6.15 Proof of Theorem 3

The proof of part (i) follows from part (ii) with $\mathscr{B}\cap B(\mathbf{x},\varepsilon)$ as the boundary. To prove part (ii), without loss of generality, we assume that $\iota = p+1$, and want to show $\sup_{\mathbf{x}\in\mathscr{B}^o}|\mathfrak{B}_{n,t}(\mathbf{x})| \lesssim h^{p+1}$. This means we have assumed that $\mathscr{B}$ has a one-to-one curve length parametrization γ that is C^{p+3} with curve length L, there exists $\varepsilon,\delta>0$ such that for all $\mathbf{x}\in\gamma([\delta, L-\delta])$ and $0<r<\varepsilon$, $S(\mathbf{x},r)$ intersects $\mathscr{B}$ with two points, $s(\mathbf{x},r)$ and $t(\mathbf{x},r)$. Define $a(\mathbf{x},r)$ and $b(\mathbf{x},r)$ to be the number in $[0,2\pi]$ such that

$$
[a(\mathbf{x},r), b(\mathbf{x},r)] = \{\theta : \mathbf{x} + r(\cos\theta,\sin\theta)\in\mathscr{A}_1\}.
$$

Then, for $\mathbf{x}\in\mathscr{B}$ and $0<r<\varepsilon$, $\theta_{1,\mathbf{x}}(r)$ has the following explicit representation:

$$
\theta_{1,\mathbf{x}}(r) = \frac{\int_{a(\mathbf{x},r)}^{b(\mathbf{x},r)}\mu_1(\mathbf{x}+r(\cos\theta,\sin\theta))f_X(\mathbf{x}+r(\cos\theta,\sin\theta))d\theta}{\int_{a(\mathbf{x},r)}^{b(\mathbf{x},r)} f_X(\mathbf{x}+r(\cos\theta,\sin\theta))d\theta}.
$$

Step 1: Curve length v.s. Distance to $\gamma(0)$

W.l.o.g., assume $\gamma(0) = \mathbf{x}$ and $\gamma'(0) = (1,0)$. Let $T:[0,\infty)\to[0,\infty)$ to be a continuous increasing function that satisfies

$$
\|\gamma\circ T(r)\|^2 = r^2, \qquad \forall r\in[0,h].
$$

Initial Case: $l = 1,2,3$.

We will show that T is C^l on $(0,h)$. For notational simplicity, define another function $\phi:[0,\infty)\to[0,\infty)$ by $\phi(t) = \|\gamma(t)\|^2$. Using implicit derivations iteratively,

$$
\phi\circ T(r) = r^2,
$$

$$
\begin{aligned}
&\phi'(T(r))T'(r) = 2r, \\
&\phi''(T(r))(T'(r))^2 + \phi'(T(r))T''(r) = 2, \\
&\phi'''(T(r))(T'(r))^3 + 3\phi''(T(r))T'(r)T''(r) + \phi'(T(r))T'''(r) = 0.
\end{aligned} \tag{1}
$$

From the above equalities, we get

$$
\begin{aligned}
T'(r) &= \frac{2r}{\phi'(T(r))}, \\
T''(r) &= \frac{2 - \phi''(T(r))\,(T'(r))^2}{\phi'(T(r))}, \\
T'''(r) &= -\frac{\phi'''(T(r))(T'(r))^3 + 3\phi''(T(r))T'(r)T''(r)}{\phi'(T(r))}.
\end{aligned}
$$

Since we have assumed γ is C^{p+3} on $(0,h)$, ϕ is also C^{p+1} on $(0,h)$. It follows from the above calculation that T is C^{p+3} on $(0,h)$. In order to find the limit of derivatives of T at 0, we need

$$
\begin{aligned}
\phi(t) &= \gamma_1(t)^2 + \gamma_2(t)^2, && \phi(0) = 0, \\
\phi'(t) &= 2\gamma_1(t)\gamma_1'(t) + 2\gamma_2(t)\gamma_2'(t), && \phi'(0) = 0, \\
\phi''(t) &= 2\gamma_1'(t)\gamma_1'(t) + 2\gamma_1(t)\gamma_1''(t) + 2\gamma_2'(t)\gamma_2'(t) + 2\gamma_2(t)\gamma_2''(t), && \phi''(0) = 2, \\
\phi'''(t) &= 6\gamma_1'(t)\gamma_1''(t) + 2\gamma_1(t)\gamma_1'''(t) + 6\gamma_2'(t)\gamma_2''(t) + 2\gamma_2(t)\gamma_2'''(t).
\end{aligned}
$$

Using L'Hôpital's rule

$$
\begin{aligned}
\lim_{r\downarrow 0} T'(r) &= \lim_{r\downarrow 0} \frac{2}{\phi''(T(r))T'(r)} = \frac{2}{2\lim_{r\downarrow 0} T'(r)} \implies \lim_{r\downarrow 0} T'(r) = 1, \\
\lim_{r\downarrow 0} T''(r) &= \lim_{r\downarrow 0} \frac{-\phi'''(T(r))(T'(r))^3 - \phi''(T(r))2T'(r)T''(r)}{\phi''(T(r))T'(r)} \\
&= \frac{-\phi^{(3)}(0) - 4\lim_{r\downarrow 0} T''(r)}{2} \\
&= \frac{-\phi^{(3)}(0)}{6}
\end{aligned}
$$

$$
\begin{aligned}
\lim_{r\downarrow 0} T^{(3)}(r) = &-\lim_{r\downarrow 0} \frac{\phi^{(4)}(T(r))(T'(r))^4 + \phi^{(3)}(T(r))3(T'(r))^2T''(r) + 3\phi^{(3)}(T(r))(T'(r))^2T''(r)}{\phi''(T(r))T'(r)} \\
&+ \lim_{r\downarrow 0} \frac{3\phi''(T(r))T'(r)T^{(3)}(r)}{\phi''(T(r))T'(r)} \\
&= -\frac{\phi^{(4)}(0) - (\phi^{(3)}(0))^2 + 6\lim_{r\downarrow 0} T^{(3)}(r)}{2} \\
&= -\frac{\phi^{(4)}(0) - (\phi^{(3)}(0))^2}{8}.
\end{aligned}
$$

Induction Step: $l \geq 4$.

Assume $\lim_{r\downarrow 0} T^{(i)}(r)$ exists and is finite for $0 \leq i \leq l-2$ and there exists a function $q(r)$ such that (i) $q(r)$ is a polynomial of $\phi^{(j)}(T(r)), 1 \leq j \leq l-1$ and $T^{(k)}(r), 1 \leq k \leq l-2$, (ii) $\lim_{r\downarrow 0} q(r) = 0$ and (iii)

$$q(r) + \phi'(T(r))T^{(l-1)}(r) = 0. \tag{2}$$

For $l = 4$, this assumption can be verified from Equation (1). Using L'hopital's rule,

$$\begin{aligned}\lim_{r\downarrow 0} T^{(l-1)}(r) &= \lim_{r\downarrow 0} -\frac{q(r)}{\phi'(T(r)}\\ &\overset{L'h}{=} \lim_{r\downarrow 0} -\frac{q'(r)}{\phi''(T(r))T'(r)}.\end{aligned}$$

From the previous paragraph, $\lim_{r\downarrow 0} \phi''(T(r))T'(r)$ exists and is finite. And $q'(r)$ is a polynomial of $\phi^{(j)}(T(r)), 1 \leq j \leq l$ and $T^{(k)}(r), 1 \leq k \leq l-1$. Hence $\lim_{r\downarrow 0} T^{(l-1)}(r)$ can be solved from the following equation and is finite:

$$\lim_{r\downarrow 0} q'(r) + \lim_{r\downarrow 0} \phi''(T(r))T'(r) \cdot \lim_{r\downarrow 0} T^{(l-1)}(r) = 0. \tag{3}$$

Taking derivatives on both sides of Equation (2),

$$q'(r) + \phi''(T(r))T'(r)T^{(l-1)}(r) + \phi'(T(r))T^{(l)}(r) = 0.$$

Take $q_2(r) = q'(r) + \phi''(T(r))T'(r)T^{(l-1)}(r)$. Then, (i) $q_2(r)$ is a polynomial of $\phi^{(j)}(T(r)), 1 \leq j \leq l$ and $T^{(k)}(r), 1 \leq k \leq l-1$, (ii) $\lim_{r\downarrow 0} q_2(r) = 0$, and (iii)

$$q_2(r) + \phi'(T(r))T^{(l)}(r) = 0.$$

Continue this argument till $l = p+3$, $\lim_{r\downarrow 0} T^{(j)}(r)$ exists and is a polynomial of $\phi^{(0)}(0), \ldots, \phi^{(j+1)}(0)$, which implies that it is bounded by a constant only depending on γ.

Step 2: $(p+1)$-times continuously differentiable S_r

We use the notation $\gamma(t) = (\gamma_1(t), \gamma_2(t))$. Define

$$A(t) = \angle\gamma(t) - \gamma(0), \gamma'(0) = \arcsin\left(\frac{\gamma_2(t)}{\|\gamma(t)\|}\right).$$

Since γ is C^{p+3}, we can Taylor expand γ at 0 to get

$$\gamma(t) = \begin{pmatrix}0\\0\end{pmatrix} + \begin{pmatrix}1\\0\end{pmatrix} t + \begin{pmatrix}u_2\\v_2\end{pmatrix} t^2 + \cdots \begin{pmatrix}u_{p+2}\\v_{p+2}\end{pmatrix} t^{p+2} + \begin{pmatrix}R_1(t)\\R_2(t)\end{pmatrix},$$

where we have used the fact that $\gamma_2'(0) = 0$ and $\|\gamma'(0)\| = 1$ and

$$R_1(t) = \int_0^t \frac{\gamma_1^{(p+3)}(s)(t-s)^{p+2}}{(p+2)!} ds, \qquad R_2(t) = \int_0^t \frac{\gamma_2^{(p+3)}(s)(t-s)^{p+2}}{(p+2)!} ds.$$

Since γ is C^{p+3}, $R_1(t)/t$ and $R_2(t)/t$ are C^{p+3} on $(0,\infty)$. We *claim* that $\lim_{t\downarrow 0}\frac{d^v}{dt^v}(R_1(t)/t)$ exists and is uniformly bounded for all $\mathbf{x}\in\mathscr{B}$, for all $0\le v\le p+1$. Define $\varphi(t)=R_1(t)/t$. Then

$$\begin{aligned}
\varphi'(t) &= -\frac{R_1(t)}{t^2}+\frac{R_1'(t)}{t},\\
\varphi''(t) &= \frac{2R_1(t)}{t^3}-\frac{2R_1'(t)}{t^2}+\frac{R_1''(t)}{t},\\
\varphi^{(3)}(t) &= -\frac{6R_1(t)}{t^4}+\frac{6R_1'(t)}{t^3}-\frac{3R_1^{(2)}(t)}{t^2}+\frac{R_1^{(3)}(t)}{t} \qquad \cdots
\end{aligned}$$

where

$$R_1'(t)=\int_0^t\frac{\gamma_1^{(p+1)}(s)(t-s)^{p-1}}{(p-1)!}ds,\qquad R_1''(t)=\int_0^t\frac{\gamma_1^{(p+1)}(s)(t-s)^{p-2}}{(p-2)!}ds,\qquad\cdots$$

Since γ_1 is C^{p+3}, there exists $C_1>0$ only depending on γ such that for all $0\le v\le p+3$,$\left|\frac{d^v}{dt^v}R_1(t)\right|\le C_1t^{p+1-v}$. Hence

$$\lim_{r\downarrow 0}\varphi^{(j)}(r)=0,\qquad\forall 0\le j\le p+1.$$

Similarly, $\lim_{r\downarrow 0}\frac{d^v}{dt^v}(R_2(t)/t)$ exists and is uniformly bounded for all $0\le v\le p+1$. Then

$$\frac{\gamma_2(t)}{\|\gamma(t)\|}=\frac{v_2t+\cdots+v_{p+2}t^{p+2}+R_2(t)/t}{\sqrt{(1+u_2t+\cdots u_{p+2}t^{p+2}+R_1(t)/t)^2+(v_2t+\cdots+v_{p+2}t^{p+2}+R_2(t)/t)^2}},t>0.$$

Notice that $\gamma_2(t)/\left\|\gamma(t)\right\|$ is of the form

$$p(t)(1+q(t))^\alpha,$$

where $\alpha<0$ and $p(t),q(t)$ are C^{p+1} on $(0,\infty)$ with $\lim_{r\downarrow 0}d^v/dt^vp(t)$ and $\lim_{r\downarrow 0}d^v/dt^vq(t)$ finite. Since the derivative of $p(t)(1+q(t))^\alpha$ is

$$p'(t)(1+q(t))^\alpha+p(t)\alpha(1+q(t))^{\alpha-1}q'(t),$$

which is the sum of two terms of the form $p_2(t)(1+q_2(t))^\alpha$ with p_2 and q_2 functions that are C^p with finite limits at 0. Continue this argument, we see that $\frac{\gamma_2(\cdot)}{\|\gamma(\cdot)\|}$ is C^{p+1} on $(0,\infty)$ and $\lim_{r\downarrow 0}\frac{d^v}{dt^v}\left(\gamma_2(t)/\left\|\gamma(t)\right\|\right)$ exist and are uniformly bounded for all $\mathbf{x}\in\mathscr{B}$ and for all $0\le v\le p+1$.

Since arcsin is C^{p+1} with bounded (higher order derivatives) on $[-1/2,1/2]$, A is C^{p+1} on $(0,\delta)$ and for all $0\le v\le p+1$, $\lim_{r\downarrow 0}A^{(v)}(t)$ exist and are uniformly bounded for all $\mathbf{x}\in\mathscr{B}$.

Step 3: $(p+1)$**-times continuously differentiable conditional density**

By the previous two steps, $a(\mathbf{x},r)=A\circ T(r)$ is C^{p+1} on $(0,\infty)$ with $|\lim_{r\downarrow 0}\frac{d^v}{dr^v}a(\mathbf{x},r)|<\infty$. Similarly, we can show that $b(\mathbf{x},r)$ is C^{p+1} in r with finite limits at $r=0$. By the assumption that f_X is C^{p+1} and bounded below by $\underline{f}$, $\theta_{1,\mathbf{x}}$ is C^{p+1} with $\lim_{r\downarrow 0}\frac{d^v}{dr^v}\theta_{1,\mathbf{x}}(r)$ uniformly bounded for all $\mathbf{x}\in\mathscr{B}$ and for all $0\le v\le p+1$.

This completes the proof. □

SA-6.16 Proof of Theorem 6

Let $s > 0$ be a parameter that is chosen later. Consider the following two data generating processes.

Data Generating Process $\mathbb{P}_0$.

Let $\mathcal{X} = \{r(\cos\theta, \sin\theta) : 0 \leq r \leq 1, 0 \leq \theta \leq \Theta(r)\}$, where

$$\Theta(r) = \begin{cases} \pi, & 0 \leq r < s, \\ \theta_k, & s + ks^2 \leq r < s + (k+1)s^2, 0 \leq k < K, \\ \theta_K, & s + Ks^2 \leq r < 1, \end{cases}$$

with $K = \lfloor \frac{1-s}{s^2} \rfloor$ and θ_k is the unique zero of

$$\frac{\sin(\theta)}{\theta} = \frac{(k + \frac{1}{2})s^2}{s + (k + \frac{1}{2})s^2}$$

over $\theta \in [0, \pi]$, and θ_K is the unique zero of

$$\frac{\sin(\theta)}{\theta} = \frac{Ks^2 + 1 - s}{s + Ks^2 + 1}$$

over $\theta \in [0, \pi]$. Suppose $\mathbf{X}_i$ has density f_X given by

$$f_X(r(\cos\theta, \sin\theta)) = \frac{1}{2\Theta(r)}, \qquad 0 \leq r \leq 1, 0 \leq \theta \leq \Theta(r).$$

Suppose

$$\mu_0(x_1, x_2) = \frac{1}{2} + \frac{1}{100}x_1, \qquad (x_1, x_2) \in \mathbb{R}^2.$$

Suppose $Y_i = \mathbb{1}(\eta_i \leq \mu(\mathbf{X}_i))$ where $(\eta_i : i : 1, \cdots, n)$ are i.i.d. random variables independent of $(\mathbf{X}_i : 1, \cdots, n)$. Let $\eta_0(r) = \mathbb{E}_{\mathbb{P}_0}[Y_i | \, \|\mathbf{X}_i - (0,0)\| = r]$, for $r \geq 0$. In particular, $\texttt{bd}(\mathcal{X})$ has length $\pi + 2$. Hence, $\texttt{bd}(\mathcal{X})$ is a rectifiable curve.

Data Generating Process $\mathbb{P}_1$.

Let $\mathcal{X} = \{r(\cos\theta, \sin\theta) : 0 \leq r \leq 1, 0 \leq \theta \leq \pi/2\}$, $\mathbf{X}_i$ is uniformly distributed on $\mathcal{X}$, and

$$\mu_1(x_1, x_2) = \frac{1}{2} + \frac{1}{100}(x_1 - s), \qquad (x_1, x_2) \in \mathbb{R}^2.$$

Suppose $Y_i = \mathbb{1}(\eta_i \leq \mu(\mathbf{X}_i))$ where $(\eta_i : 1, \cdots, n)$ are i.i.d random variables independent to $(\mathbf{X}_i : 1, \cdots, n)$. Let $\eta_1(r) = \mathbb{E}_{\mathbb{P}_1}[Y_i | \, \|\mathbf{X}_i - (0,0)\| = r]$, for $r \geq 0$. In particular, $\texttt{bd}(\mathcal{X})$ has length $\pi/2 + 2$. Hence, $\texttt{bd}(\mathcal{X})$ is a rectifiable curve.

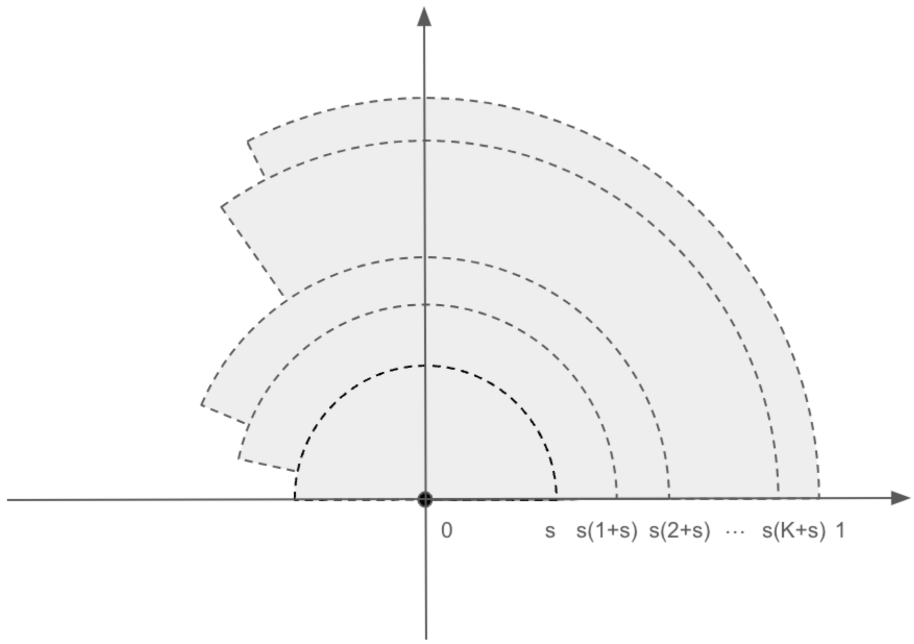

Figure SA-2: $\mathscr{X}$ from DGP $\mathbb{P}_0$

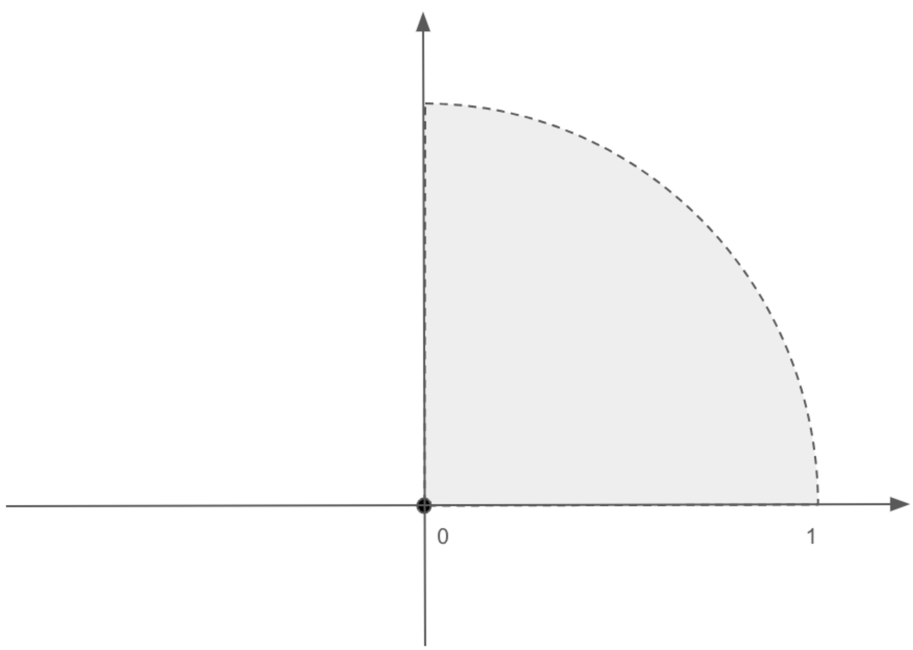

Figure SA-3: $\mathscr{X}$ from DGP $\mathbb{P}_1$

Minimax Lower Bound.

First, we show under the previous two models, $\mathbb{P}_0(\|\mathbf{X}_i\| \le r) = \mathbb{P}_1(\|\mathbf{X}_i\| \le r)$ for all $r \ge 0$. Since in $\mathbb{P}_1$, $\mathbf{X}_i$ is uniform distributed on $\mathbb{R}$, we know $\mathbb{P}_1(\|\mathbf{X}_i\| \le r) = r^2$, $0 \le r \le 1$.

$$\mathbb{P}_0(\|\mathbf{X}_i\| \le r) = \int_0^r \int_0^{\Theta(s)} \frac{1}{2\Theta(s)} s d\theta ds = r^2, \quad 0 \le r \le 1.$$

Hence, choosing $(0,0)$ as the point of evaluation in both $\mathbb{P}_0$ and $\mathbb{P}_1$, we have

$$\begin{aligned}
& d_{\mathrm{KL}}(\mathbb{P}_0(\|\mathbf{X}_i - (0,0)\|, Y_i), \mathbb{P}_1(\|\mathbf{X}_i - (0,0)\|, Y_i)) \\
&= \int_0^\infty \int_{-\infty}^\infty d\mathbb{P}_0(r,y) \log \frac{d\mathbb{P}_0(r,y)}{d\mathbb{P}_1(r,y)} \\
&= \int_0^\infty \int_{-\infty}^\infty d\mathbb{P}_0(r) d\mathbb{P}_0(y|r) \log \frac{d\mathbb{P}_0(r) d\mathbb{P}_0(y|r)}{d\mathbb{P}_1(r) d\mathbb{P}_1(y|r)} \\
&= \int_0^\infty d\mathbb{P}_0(r) \int_{-\infty}^\infty d\mathbb{P}_0(y|r) \log \frac{d\mathbb{P}_0(y|r)}{d\mathbb{P}_1(y|r)} \\
&= 2\int_0^1 d_{\mathrm{KL}}(\mathsf{Bernoulli}(\eta_0(r)), \mathsf{Bernoulli}(\eta_1(r))) r dr.
\end{aligned}$$

Under $\mathbb{P}_0$, $\mathbf{X}_i$ is uniformly distributed on $\{r(\cos\theta, \sin\theta) : 0 \le \theta \le \Theta(r)\}$ for each $0 < r \le 1$. Hence

$$\eta_0(r) = \frac{1}{2} + \frac{1}{100}\frac{1}{\Theta(r)}\int_0^{\Theta(r)} r\cos(u)du - \frac{s}{100} = \frac{1}{2} + \frac{1}{100} r \frac{\sin(\Theta(r))}{\Theta(r)}.$$

Thus, for $0 \le k < K$,

$$\begin{aligned}
\eta_0\Big(s + (k+\frac{1}{2})s^2\Big) &= \frac{1}{2} + \frac{1}{100}\Big((s + (k+\frac{1}{2})s^2)\frac{\sin(\Theta_k)}{\Theta_k}\Big) \\
&= \frac{1}{2} + \frac{1}{100}\Big((s + (k+\frac{1}{2})s^2)\frac{(k+\frac{1}{2})s^2}{s + (k+\frac{1}{2})s^2}\Big) \\
&= \eta_1\Big(s + (k+\frac{1}{2})s^2\Big).
\end{aligned}$$

Since both η_0 and η_1 are 1-Lipschitz on all intervals $[s + ks^2, s + (k+1)s^2]$ for all $0 \le k < K$, we know $|\eta_0(r) - \eta_1(r)| \le 2s^2$ for all $r \in [s, 1]$. Moreover, $\eta_0(r) = \frac{1}{2}$ for all $0 \le r \le s$ and $\eta_1(r) = \frac{1}{2} + \frac{1}{100}(r\frac{2}{\pi} - s)$. Hence $|\eta_0(r) - \eta_1(r)| \le s$ for all $0 \le r \le s$. Hence,

$$\begin{aligned}
\int_0^1 d_{\mathrm{KL}}(\mathsf{Bernoulli}(\eta_0(r)), \mathsf{Bernoulli}(\eta_1(r))) r dr &\le \int_0^1 d_{\chi^2}(\mathsf{Bernoulli}(\eta_0(r)), \mathsf{Bernoulli}(\eta_1(r))) r dr \\
&= \int_0^1 (\eta_1(r)\Big(\frac{\eta_0(r) - \eta_1(r)}{\eta_1(r)}\Big)^2 + (1 - \eta_1(r))\Big(\frac{\eta_0(r) - \eta_1(r)}{1 - \eta_1(r)}\Big)^2) r dr \\
&\le \frac{1}{\frac{1}{2} - \frac{3}{100}} \int_0^1 (\eta_0(r) - \eta_1(r))^2 r dr \\
&\le \frac{1}{\frac{1}{2} - \frac{3}{100}} \int_0^s s^2 r dr + \frac{1}{\frac{1}{2} - \frac{3}{100}} \int_s^1 (2s^2)^2 r dr \\
&\le \frac{5}{\frac{1}{2} - \frac{3}{100}} s^4.
\end{aligned}$$

Moreover, $|\mu_0(0,0) - \mu_1(0,0)| = \frac{1}{100}s$. Hence, by Tsybakov [2008, Theorem 2.2 (iii)], take $\frac{5}{\frac{1}{2}-\frac{3}{100}}s_*^4 = \frac{\log 2}{n}$, and conclude that

$$\inf_{T_n \in \mathscr{T}} \sup_{\mathbb{P} \in \mathscr{P}} \sup_{\mathbf{x} \in \mathscr{B}(\mathbb{P})} \mathbb{E}_{\mathbb{P}}[|T_n(\mathbf{U}_n(\mathbf{x})) - \mu(\mathbf{x})|] \geq \frac{1}{1600} s_* \gtrsim n^{-\frac{1}{4}}.$$

This concludes the proof. □

References

Matias D. Cattaneo and Ruiqi (Rae) Yu. Strong approximations for empirical processes indexed by lipschitz functions. *Annals of Statistics*, 53(3):1203–1229, 2025.

Matias D. Cattaneo, Rajita Chandak, Michael Jansson, and Xinwei Ma. Boundary adaptive local polynomial conditional density estimators. *Bernoulli*, 30(4):3193–3223, 2024.

Matias D. Cattaneo, Rocio Titiunik, and Ruiqi (Rae) Yu. Estimation and inference in boundary discontinuity designs: Location-based methods. *arXiv:2505.05670*, 2025.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Anti-concentration and honest, adaptive confidence bands. *Annals of Statistics*, 42(5):1787–1818, 2014a.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximation of suprema of empirical processes. *Annals of Statistics*, 42(4):1564–1597, 2014b.

Victor Chernozhuokov, Denis Chetverikov, Kengo Kato, and Yuta Koike. Improved central limit theorem and bootstrap approximations in high dimensions. *Annals of Statistics*, 50(5):2562–2586, 2022.

Richard M Dudley. *Uniform central limit theorems*, volume 142. Cambridge university press, 2014.

Herbert Federer. *Geometric measure theory*. Springer, 2014.

G.B. Folland. *Advanced Calculus*. Featured Titles for Advanced Calculus Series. Prentice Hall, 2002.

Evarist Giné and Richard Nickl. *Mathematical Foundations of Infinite-dimensional Statistical Models*. Cambridge University Press, New York, 2016.

Leon Simon et al. *Lectures on geometric measure theory*. Centre for Mathematical Analysis, Australian National University Canberra, 1984.

A.B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2008.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996.