Supplemental Appendix to "Boundary Adaptive Local Polynomial Conditional Density Estimators"[*]

Matias D. Cattaneo[†]   Rajita Chandak[*]   Michael Jansson[‡]   Xinwei Ma[§]

April 21, 2022

Abstract

This Supplemental Appendix contains general theoretical results encompassing those discussed in the main paper, includes proofs of those general results, and discusses additional methodological and technical results. A companion `R` package is available at https://nppackages.github.io/lpcde/.

Contents

SA-1 Setup

Let $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$ be continuously distributed random variables supported on $\mathcal{X} = [0,1]^d$ and $\mathcal{Y} = [0,1]$, respectively. We are interested in estimating the conditional distribution function and its derivatives:

$$\theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{\partial^\mu}{\partial y^\mu} \frac{\partial^{\boldsymbol{\nu}}}{\partial \mathbf{x}^{\boldsymbol{\nu}}} F(y|\mathbf{x}),$$

where $\mu \in \mathbb{N}$, and $\boldsymbol{\nu} \in \mathbb{N}^d$ representing multi-indices. (In the main paper we only consider derivatives along the y-direction, that is, we set $\boldsymbol{\nu} = 0$ and $\mu = \vartheta + 1$.)

To present our estimation strategy, we start from $\theta_{0,\boldsymbol{\nu}}$, the conditional distribution function and its derivatives with respect to the conditioning variable, and apply the local polynomial method:

$$\widehat{\frac{\partial^{\boldsymbol{\nu}}}{\partial \mathbf{x}^{\boldsymbol{\nu}}}} F(y|\mathbf{x}) = \mathbf{e}_{\boldsymbol{\nu}}^{\mathrm{T}} \hat{\boldsymbol{\gamma}}(y,\mathbf{x}), \qquad \hat{\boldsymbol{\gamma}}(y,\mathbf{x}) = \underset{\boldsymbol{\gamma} \in \mathbb{R}^{\mathfrak{q}_d+1}}{\operatorname{argmin}} \sum_{i=1}^n \Big(\mathbb{1}(y_i \leq y) - \mathbf{q}(\mathbf{x}_i - \mathbf{x})^{\mathrm{T}} \boldsymbol{\gamma}\Big)^2 L_h(\mathbf{x}_i;\mathbf{x}),$$

where, using standard multi-index notation, $\mathbf{q}(\mathbf{u})$ denotes the $(\mathfrak{q}_d + 1)$-dimensional vector collecting the ordered elements $\mathbf{u}^{\boldsymbol{\nu}}/\boldsymbol{\nu}!$ for $0 \leq |\boldsymbol{\nu}| \leq \mathfrak{q}$, where $\mathbf{u}^{\boldsymbol{\nu}} = u_1^{\nu_1} u_2^{\nu_2} \cdots u_d^{\nu_d}$, $|\boldsymbol{\nu}| = \nu_1 + \nu_2 + \cdots + \nu_d$ for $\mathbf{u} = (u_1, u_2, \cdots, u_d)^{\mathrm{T}}$, $\boldsymbol{\nu} = (\nu_1, \nu_2, \cdots, \nu_d)^{\mathrm{T}}$, and $\mathfrak{q}_d = (d+\mathfrak{q})!/(\mathfrak{q}!d!) - 1$. $L_h(\mathbf{u};\mathbf{x}) = L((\mathbf{u}-\mathbf{x})/h)/h^d$ for some kernel function L, and $\mathbf{e}_{\boldsymbol{\nu}}^{\mathrm{T}}$ is a basis vector extracting the corresponding estimate.

We can write the solution in closed form as

$$\widehat{\frac{\partial^{\boldsymbol{\nu}}}{\partial \mathbf{x}^{\boldsymbol{\nu}}}} F(y|\mathbf{x}) = \mathbf{e}_{\boldsymbol{\nu}}^{\mathrm{T}} \hat{\mathbf{S}}_{\mathbf{x}}^{-1} \left(\frac{1}{n} \sum_{i=1}^n \mathbb{1}(y_i \leq y) \mathbf{q}\Big(\frac{\mathbf{x}_i - \mathbf{x}}{h}\Big) L_h(\mathbf{x}_i;\mathbf{x}) \right),$$

where

$$\hat{\mathbf{S}}_{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{q}\Big(\frac{\mathbf{x}_i - \mathbf{x}}{h}\Big) \mathbf{q}\Big(\frac{\mathbf{x}_i - \mathbf{x}}{h}\Big)^{\mathrm{T}} L_h(\mathbf{x}_i;\mathbf{x}).$$

To estimate $\theta_{\mu,\boldsymbol{\nu}}$, we further smooth via local polynomials along the y-direction:

$$\hat{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \mathbf{e}_\mu^{\mathrm{T}} \hat{\boldsymbol{\beta}}(y,\mathbf{x}), \qquad \hat{\boldsymbol{\beta}}(y,\mathbf{x}) = \underset{\mathbf{b} \in \mathbb{R}^{\mathfrak{p}+1}}{\operatorname{argmin}} \sum_{i=1}^n \Big(\hat{F}(y_i|\mathbf{x}) - \mathbf{p}(y_i - y)^{\mathrm{T}} \mathbf{b}\Big)^2 K_h(y_i;y).$$

In the above $\mathbf{p}(u) = (1, u, u^2/2, \cdots, u^{\mathfrak{p}}/\mathfrak{p}!)^{\mathrm{T}}$ is the $\mathfrak{p}$-th order polynomial expansion, and $K_h(u;y) = K((u-y)/h)/h$ for some kernel function K. We can write the solution in closed-form as

$$\hat{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \mathbf{e}_\mu^{\mathrm{T}} \hat{\mathbf{S}}_y^{-1} \hat{\mathbf{R}}_{y,\mathbf{x}} \hat{\mathbf{S}}_{\mathbf{x}}^{-1} \mathbf{e}_{\boldsymbol{\nu}},$$

where

$$\hat{\mathbf{S}}_y = \frac{1}{n} \sum_{i=1}^{n} \mathbf{p}\left(\frac{y_i - y}{h}\right) \mathbf{p}\left(\frac{y_i - y}{h}\right)^{\mathrm{T}} K_h(y_i; y), \qquad \text{and}$$

$$\hat{\mathbf{R}}_{y,\mathbf{x}} = \frac{1}{n^2 h^{\mu + |\boldsymbol{\nu}|}} \sum_{j=1}^{n} \sum_{i=1}^{n} \mathbb{1}(y_i \le y_j) \mathbf{p}\left(\frac{y_j - y}{h}\right) K_h(y_j; y) \mathbf{q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}} L_h(\mathbf{x}_i; \mathbf{x}).$$

While in the above we considered local polynomial regressions along both the $\mathbf{x}$- and y-directions, it is also possible to employ a local smoothing technique. To be precise, let G be some function such that the following Lebesgue-Stieltjes integration is well-defined, then an alternative estimator can be constructed as

$$\check{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) = \mathbf{e}_\mu^{\mathrm{T}} \check{\boldsymbol{\beta}}(y, \mathbf{x}), \qquad \check{\boldsymbol{\beta}}(y, \mathbf{x}) = \operatorname*{argmin}_{\mathbf{b} \in \mathbb{R}^{\mathfrak{p}+1}} \int \left(\hat{F}(u|\mathbf{x}) - \mathbf{p}(u - y)^{\mathrm{T}} \mathbf{b} \right)^2 K_h(u; y) \mathrm{d}G(u),$$

which has the solution

$$\check{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) = \mathbf{e}_\mu^{\mathrm{T}} \mathbf{S}_y^{-1} \bar{\mathbf{R}}_{y,\mathbf{x}} \hat{\mathbf{S}}_{\mathbf{x}}^{-1} \mathbf{e}_{\boldsymbol{\nu}},$$

where

$$\mathbf{S}_y = \int_{\mathcal{Y}} \mathbf{p}\left(\frac{u - y}{h}\right) \mathbf{p}\left(\frac{u - y}{h}\right)^{\mathrm{T}} K_h(u; y) \mathrm{d}G(u), \quad \text{and}$$

$$\bar{\mathbf{R}}_{y,\mathbf{x}} = \frac{1}{n h^{\mu + |\boldsymbol{\nu}|}} \sum_{i=1}^{n} \left(\int_{\mathcal{Y}} \mathbb{1}(y_i \le u) \mathbf{p}\left(\frac{u - y}{h}\right) K_h(u; y) \mathrm{d}G(u) \right) \mathbf{q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}} L_h(\mathbf{x}_i; \mathbf{x}).$$

SA-1.1 List of Notations

Limits are taken with respect to the sample size tending to infinity and the bandwidth shrinking to zero (i.e., $n \to \infty$ and $h \to 0$). For two nonnegative sequences, $a_n \precsim b_n$ implies that $\limsup_{n \to \infty} |a_n/b_n| < \infty$. Similarly, $a_n \precsim_{\mathbb{P}} b_n$ means $|a_n/b_n|$ is asymptotically bounded in probability. We also adopt the small- and big-oh notation: $a_n = O_{\mathbb{P}}(b_n)$ is just $a_n \precsim_{\mathbb{P}} b_n$, and $a_n = o_{\mathbb{P}}(b_n)$ means a_n/b_n converges to zero in probability. Constants that do not depend on the sample size or the bandwidth will be denoted by $\mathfrak{c}$, $\mathfrak{c}_1$, $\mathfrak{c}_2$, etc.

We introduce another notation, $O_{\mathbb{TC}}$, which not only provides an asymptotic order but also controls the tail probability. To be specific, $a_n = O_{\mathbb{TC}}(b_n)$ if for any $\mathfrak{c}_1 > 0$, there exists some $\mathfrak{c}_2$ such that

$$\limsup_{n \to \infty} n^{\mathfrak{c}_1} \mathbb{P}\left[a_n \ge \mathfrak{c}_2 b_n\right] < \infty.$$

Here the subscript, $\mathbb{TC}$, stands for "tail control."

Finally, let $\mathbf{X} = (\mathbf{x}_1^{\mathrm{T}}, \cdots, \mathbf{x}_n^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{Y} = (y_1, \cdots, y_n)^{\mathrm{T}}$ be the data matrices.

- $F(y|\mathbf{x})$ and $f(y|\mathbf{x})$: the conditional distribution and density functions of y_i (at y) given $\mathbf{x}_i = \mathbf{x}$. The marginal distributions and densities are denoted by F_y, $F_{\mathbf{x}}$, f_y, and $f_{\mathbf{x}}$, respectively.

- y and $\mathbf{x}$: the evaluation points.

- $\mathcal{X} = [0,1]^d$ and $\mathcal{Y} = [0,1]$, the support of $\mathbf{x}_i$ and y_i, respectively.

- h: the bandwidth sequence.

- K: the kernel function, and L is the product kernel: $L(\mathbf{u}) = K(u_1)K(u_2)\cdots K(u_d)$.

- $\mathbf{p}$, $\mathbf{q}$: polynomial expansions. For a univariate argument y, $\mathbf{p}(u) = (1, u, u^2/2, \cdots, u^{\mathfrak{p}}/\mathfrak{p}!)^{\mathrm{T}}$, and for a multivariate argument $\mathbf{u}$, $\mathbf{q}(\mathbf{u})$ contains polynomials and interactions up to order $\mathfrak{q}$ in increasing order.

- $\mathbf{P}$ and $\mathbf{Q}$: defined as $\mathbf{p}(\cdot)K(\cdot)$ and $\mathbf{q}(\cdot)L(\cdot)$, respectively.

- $\mathbf{e}_\mu$ and $\mathbf{e}_{\boldsymbol{\nu}}$: standard basis vectors extracting the μ-th and $\boldsymbol{\nu}$-th element in the expansion of $\mathbf{p}$ and $\mathbf{q}$ for univariate and multivariate arguments, respectively.

- $G(\cdot)$ the weighting function used in $\check{\theta}_{\mu,\boldsymbol{\nu}}$, with its Lebesgue density denoted by $g(\cdot)$.

- Some matrices

$$\mathbf{S}_y = \int_{\frac{\mathcal{Y}-y}{h}} \mathbf{p}(u)\,\mathbf{P}(u)^{\mathrm{T}}\,g(y+hu)\mathrm{d}u, \qquad \hat{\mathbf{S}}_y = \frac{1}{nh}\sum_{i=1}^{n}\mathbf{p}\Big(\frac{y_i-y}{h}\Big)\mathbf{P}\Big(\frac{y_i-y}{h}\Big)^{\mathrm{T}},$$

$$\mathbf{c}_{y,\ell} = \int_{\frac{\mathcal{Y}-y}{h}} \frac{u^\ell}{\ell!}\mathbf{P}(u)\,g(y+hu)\mathrm{d}u, \qquad \hat{\mathbf{c}}_{y,\ell} = \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{\ell!}\Big(\frac{y_i-y}{h}\Big)^\ell \mathbf{P}\Big(\frac{y_i-y}{h}\Big)^{\mathrm{T}},$$

$$\mathbf{S}_{\mathbf{x}} = \int_{\frac{\mathcal{X}-\mathbf{x}}{h}} \mathbf{q}(\mathbf{v})\,\mathbf{Q}(\mathbf{v})^{\mathrm{T}} f_{\mathbf{x}}(\mathbf{x}+h\mathbf{v})\mathrm{d}\mathbf{v}, \qquad \hat{\mathbf{S}}_{\mathbf{x}} = \frac{1}{nh^d}\sum_{i=1}^{n}\mathbf{q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^{\mathrm{T}},$$

$$\mathbf{c}_{\mathbf{x},\mathbf{m}} = \int_{\frac{\mathcal{X}-\mathbf{x}}{h}} \frac{\mathbf{v}^{\mathbf{m}}}{\mathbf{m}!}\mathbf{Q}(\mathbf{v})\,f_{\mathbf{x}}(\mathbf{x}+h\mathbf{v})\mathrm{d}\mathbf{v}, \qquad \hat{\mathbf{c}}_{\mathbf{x},\mathbf{m}} = \frac{1}{nh^d}\sum_{i=1}^{n}\frac{1}{\mathbf{m}!}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^{\mathbf{m}}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big),$$

$$\mathbf{T}_y = \iint_{\frac{\mathcal{Y}-y}{h}} (u_1 \wedge u_2)\mathbf{P}(u_1)\,\mathbf{P}(u_2)^{\mathrm{T}}\,g(y+hu_1)g(y+hu_2)\mathrm{d}u_1\mathrm{d}u_2,$$

$$\mathbf{T}_{\mathbf{x}} = \int_{\frac{\mathcal{X}-\mathbf{x}}{h}} \mathbf{Q}(\mathbf{v})\,\mathbf{Q}(\mathbf{v})^{\mathrm{T}} f_{\mathbf{x}}(\mathbf{x}+h\mathbf{v})\mathrm{d}\mathbf{v}, \qquad \hat{\mathbf{T}}_{\mathbf{x}} = \frac{1}{nh^d}\sum_{i=1}^{n}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^{\mathrm{T}},$$

$$\hat{\mathbf{R}}_{y,\mathbf{x}} = \frac{1}{n^2 h^{1+d+\mu+|\boldsymbol{\nu}|}}\sum_{j=1}^{n}\sum_{i=1}^{n}\mathbb{1}(y_i \le y_j)\mathbf{P}\Big(\frac{y_j-y}{h}\Big)\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^{\mathrm{T}},$$

$$\bar{\mathbf{R}}_{y,\mathbf{x}} = \frac{1}{nh^{1+d+\mu+|\boldsymbol{\nu}|}}\sum_{i=1}^{n}\Big(\int_{\mathcal{Y}}\mathbb{1}(y_i \le u)\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\Big)\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^{\mathrm{T}}.$$

- Equivalent kernels:

$$\check{\mathscr{K}}^{\circ}_{\mu,\boldsymbol{\nu},h}(a,\mathbf{b};y,\mathbf{x}) = \frac{1}{h^{\mu+|\boldsymbol{\nu}|}}\mathbf{e}_{\mu}^{\mathrm{T}}\mathbf{S}_y^{-1}\left[\int_{\mathcal{Y}}\left(\mathbb{1}(a\le u)-\hat{F}(u|\mathbf{b})\right)\frac{1}{h}\mathbf{P}\left(\frac{u-y}{h}\right)dG(u)\right]\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{b}-\mathbf{x}}{h}\right)^{\mathrm{T}}\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}},$$

$$\hat{\mathscr{K}}^{\circ}_{\mu,\boldsymbol{\nu},h}(a,\mathbf{b};y,\mathbf{x}) = \frac{1}{h^{\mu+|\boldsymbol{\nu}|}}\mathbf{e}_{\mu}^{\mathrm{T}}\hat{\mathbf{S}}_y^{-1}\left[\frac{1}{n}\sum_{j=1}^{n}\left(\mathbb{1}(a\le y_j)-\hat{F}(y_j|\mathbf{b})\right)\frac{1}{h}\mathbf{P}\left(\frac{y_j-y}{h}\right)\right]\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{b}-\mathbf{x}}{h}\right)^{\mathrm{T}}\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}},$$

$$\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}(a,\mathbf{b};y,\mathbf{x}) = \frac{1}{h^{\mu+|\boldsymbol{\nu}|}}\mathbf{e}_{\mu}^{\mathrm{T}}\mathbf{S}_y^{-1}\left[\int_{\mathcal{Y}}\left(\mathbb{1}(a\le u)-F(u|\mathbf{b})\right)\frac{1}{h}\mathbf{P}\left(\frac{u-y}{h}\right)dG(u)\right]\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{b}-\mathbf{x}}{h}\right)^{\mathrm{T}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}},$$

$$\mathscr{K}_{\mu,\boldsymbol{\nu},h}(a,\mathbf{b};y,\mathbf{x}) = \frac{1}{h^{\mu+|\boldsymbol{\nu}|}}\mathbf{e}_{\mu}^{\mathrm{T}}\mathbf{S}_y^{-1}\left[\int_{\mathcal{Y}}\mathbb{1}(a\le u)\frac{1}{h}\mathbf{P}\left(\frac{u-y}{h}\right)dG(u)\right]\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{b}-\mathbf{x}}{h}\right)^{\mathrm{T}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}.$$

- Some rates:

$$\mathbf{r}_{\mathtt{B}} = h^{\mathfrak{q}+1-|\boldsymbol{\nu}|}+h^{\mathfrak{p}+1-\mu}, \qquad \mathbf{r}_{\mathtt{V}} = \sqrt{\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}}}, \qquad \mathbf{r}_{\mathtt{BE}} = \begin{cases} \frac{1}{\sqrt{nh^d}} & \text{if } \mu=0, \text{ and } \theta_{0,\mathbf{0}}\ne 0 \text{ or } 1 \\ \frac{1}{\sqrt{nh^{d+1}}} & \text{if } \mu>0, \text{ or } \theta_{0,\mathbf{0}}=0 \text{ or } 1 \end{cases},$$

$$\mathbf{r}_{\mathtt{VE}} = h^{\mathfrak{q}+\frac{1}{2}}+\sqrt{\frac{\log n}{nh^{d+1}}}, \qquad \mathbf{r}_{\mathtt{SE}} = \sqrt{\log n}\,\mathbf{r}_{\mathtt{VE}}, \qquad \mathbf{r}_{\mathtt{SA}} = \left(\frac{\log^{d+1} n}{nh^{d+1}}\right)^{\frac{1}{2d+2}}.$$

SA-1.2 Overview

In this subsection we provide an overview of the main results. Underlying assumptions and precise statements of the lemmas and theorems will be given in later sections. First consider $\check{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$, and we employ a conditional expectation decomposition:

$$\check{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = h^{-\mu-|\boldsymbol{\nu}|}\mathbf{e}_{\mu}^{\mathrm{T}}\mathbf{S}_y^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\int_{\mathcal{Y}}F(u|\mathbf{x}_i)\frac{1}{h}\mathbf{P}\left(\frac{u-y}{h}\right)dG(u)\right)\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^{\mathrm{T}}\right]\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}$$

$$+ h^{-\mu-|\boldsymbol{\nu}|}\mathbf{e}_{\mu}^{\mathrm{T}}\mathbf{S}_y^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\int_{\mathcal{Y}}\left(\mathbb{1}(y_i\le u)-F(u|\mathbf{x}_i)\right)\frac{1}{h}\mathbf{P}\left(\frac{u-y}{h}\right)dG(u)\right)\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^{\mathrm{T}}\right]\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}.$$

As we will show in Section SA-2, the first term above consists of the centering of the estimator (i.e., the parameter of interest $\theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$) and the smoothing bias. The second term, on the other hand, gives the asymptotic representation of the estimator. To be precise, we have

$$\check{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}(y_i,\mathbf{x}_i;y,\mathbf{x})$$

$$+ O_{\mathbb{P}}\left(h^{\mathfrak{q}+1-|\boldsymbol{\nu}|}+h^{\mathfrak{p}+1-\mu}+\sqrt{V_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\frac{\log n}{\sqrt{nh^d}}\right).$$

As a result, we can focus on establishing properties of the the first term, which provides an equivalent kernel expression. Denote its variance by $\mathsf{V}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})$. Then we show that the standardized process,

$$\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) = \frac{1}{n\sqrt{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}} \sum_{i=1}^{n} \mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(y_i, \mathbf{x}_i; y, \mathbf{x}),$$

is approximately normally distributed both pointwise and uniformly for $y \in \mathcal{Y}$ and $\mathbf{x} \in \mathcal{X}$. To be even more precise, we establish a strong approximation result, meaning that there exists a copy $\bar{\mathbb{S}}'_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})$, and a Gaussian process $\mathbb{G}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})$ with the same covariance structure, such that

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left| \bar{\mathbb{S}}'_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) - \mathbb{G}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) \right| = O_{\mathbb{P}}\left(\left(\frac{\log^{d+1} n}{nh^{d+1}} \right)^{\frac{1}{2d+2}} \right).$$

Together with a feasible variance-covariance estimator, the strong approximation result not only allows us to construct confidence bands for the target parameter and test shape restrictions, but also provides an explicit characterization of the coverage error probability for those procedures. On a related note, the (leading) variance of our estimator has the order $(nh^{d+2|\boldsymbol{\nu}|})^{-1}$ for $\mu = 0$, and $(nh^{d+2\mu+2|\boldsymbol{\nu}|-1})^{-1}$ for $\mu \geq 1$. For example, setting $\mu = 1$ and $\boldsymbol{\nu} = 0$, we have the leading variance to be $(nh^{d+1})^{-1}$ for conditional density estimation.

Inside the remainder term, $h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} + h^{\mathfrak{p}+1-\mu}$ is the order of the leading smoothing bias, and $(\log n)\sqrt{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})/(nh^d)}$ arises from the linearization step which replaces the random matrix $\hat{\mathbf{S}}_{\mathbf{x}}$ by its large-sample analogue $\mathbf{S}_{\mathbf{x}}$. It is worth mentioning that the order of the remainder term is uniformly valid for $y \in \mathcal{Y}$ and $\mathbf{x} \in \mathcal{X}$, which is why an extra logarithmic factor is present.

Now consider the other estimator, $\hat{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})$. While it is not possible to take a conditional expectation, we can still "center" the estimator with the conditional distribution function. That is,

$$\hat{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) = h^{-\mu-|\boldsymbol{\nu}|}\mathbf{e}_{\mu}^{\mathrm{T}}\hat{\mathbf{S}}_y^{-1}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} F(y_j|\mathbf{x}_i)\frac{1}{h}\mathbf{P}\left(\frac{y_j - y}{h}\right)\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}} \right]\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}$$

$$+ h^{-\mu-|\boldsymbol{\nu}|}\mathbf{e}_{\mu}^{\mathrm{T}}\hat{\mathbf{S}}_y^{-1}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \left(\mathbb{1}(y_i \leq y_j) - F(y_j|\mathbf{x}_i) \right)\frac{1}{h}\mathbf{P}\left(\frac{y_j - y}{h}\right)\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}} \right]\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}.$$

As before, the first term captures the target parameter and the smoothing bias. The analysis of the second term is more involved. Besides the asymptotic linear representation term, it also consists of a leave-in bias term (since the same observation is used twice) and a second order U-statistic. We show that the following expansion holds uniformly for $y \in \mathcal{Y}$ and $\mathbf{x} \in \mathcal{X}$:

$$\hat{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) - \theta_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} \mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(y_i, \mathbf{x}_i; y, \mathbf{x})$$

$$+ O_{\mathbb{P}}\left(h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} + h^{\mathfrak{p}+1-\mu} + \sqrt{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}\frac{\log n}{\sqrt{nh^d}} + \frac{\log n}{\sqrt{n^2 h^{d+2\mu+2|\boldsymbol{\nu}|+1}}} \right).$$

Here, the contribution of the U-statistic is represented by the order $\log n/(\sqrt{n^2 h^{d+2\mu+2|\boldsymbol{\nu}|+1}})$ in the remainder term. Interestingly, this term is negligible compared to the standard error, $\sqrt{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}$, provided that $\log n/(nh^2) \to \infty$.

The above demonstrates that important large-sample properties of the local regression based estimator, $\hat{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ – such as pointwise and uniform normal approximation – stem from the equivalent kernel representation. Here we note that the representation holds by setting $G = F_y$. In other words, $\hat{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ is first-order asymptotically equivalent to $\check{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ with the (infeasible) local smoothing using the marginal distribution F_y.

SA-1.3 Assumptions

We make the following assumptions on the joint distribution, the kernel function, and the weighting G.

Assumption SA-DGP (Data generating process)
(i) $\{y_i, \mathbf{x}_i\}_{1 \leq i \leq n}$ is a random sample from the joint distribution F supported on $\mathcal{Y} \times \mathcal{X} = [0,1]^{1+d}$.
(ii) The joint density, f, is continuous and is bounded away from zero.
(iii) $\theta_{2,\mathbf{0}}$ exists and is continuous.

Assumption SA-K (Kernel)
The kernel function K is nonnegative, symmetric, supported on $[-1,1]$, Lipschitz continuous, and integrates to one.

Assumption SA-W (Weighting function)
The weighting function G is continuously differentiable with a Lebesgue density denoted by g.

SA-2 Pointwise Large-sample Properties

We first present several uniform convergence results which will be used later to establish pointwise and uniform properties of our estimators.

Lemma SA-2.1 (Matrix convergence)
Let Assumptions SA-DGP, SA-K, and SA-W hold with $h \to 0$, $nh^d/\log n \to \infty$, and $G = F_y$. Then

$$\sup_{y \in \mathcal{Y}} \left| \hat{\mathbf{S}}_y - \mathbf{S}_y \right| = O_{\mathbb{TC}}\left(\sqrt{\frac{\log n}{nh}} \right), \qquad \sup_{y \in \mathcal{Y}} |\hat{\mathbf{c}}_{y,\ell} - \mathbf{c}_{y,\ell}| = O_{\mathbb{TC}}\left(\sqrt{\frac{\log n}{nh}} \right),$$

$$\sup_{\mathbf{x} \in \mathcal{X}} \left| \hat{\mathbf{S}}_{\mathbf{x}} - \mathbf{S}_{\mathbf{x}} \right| = O_{\mathbb{TC}}\left(\sqrt{\frac{\log n}{nh^d}} \right), \qquad \sup_{\mathbf{x} \in \mathcal{X}} |\hat{\mathbf{c}}_{\mathbf{x},\mathbf{m}} - \mathbf{c}_{\mathbf{x},\mathbf{m}}| = O_{\mathbb{TC}}\left(\sqrt{\frac{\log n}{nh^d}} \right),$$

$$\sup_{\mathbf{x} \in \mathcal{X}} \left| \hat{\mathbf{T}}_{\mathbf{x}} - \mathbf{T}_{\mathbf{x}} \right| = O_{\mathbb{TC}}\left(\sqrt{\frac{\log n}{nh^d}} \right).$$

If in addition that $nh^{d+1}/\log n \to \infty$, then

$$\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} \left| \mathbf{e}_\mu^{\mathrm{T}} \mathbf{S}_y^{-1} \left(\bar{\mathbf{R}}_{y,\mathbf{x}} - \mathbb{E}\left[\bar{\mathbf{R}}_{y,\mathbf{x}} | \mathbf{X} \right] \right) \right| = O_{\mathbb{TC}}(\mathbf{r}_1), \quad \text{where } \mathbf{r}_1 = \begin{cases} \sqrt{\frac{\log n}{nh^{d+2\mu+2|\nu|}}} & \text{if } \mu = 0 \\ \sqrt{\frac{\log n}{nh^{d+2\mu+2|\nu|-1}}} & \text{if } \mu > 0 \end{cases}.$$

We now follow the decomposition in Section SA-1.2 and study the leading bias of our estimators.

Lemma SA-2.2 (Bias)

Let Assumptions SA-DGP, SA-K and SA-W hold with $h \to 0$ and $nh^d/\log n \to \infty$. In addition, $\theta_{\mu',\nu'}$ exists and is continuous for all $\mu' + |\nu'| = \max\{\mathfrak{q} + 1 + \mu, \ \mathfrak{p} + 1 + |\nu|\}$. Then

$$\mathbf{e}_\mu^{\mathrm{T}} \mathbf{S}_y^{-1} \left[\frac{1}{nh^{\mu+|\nu|}} \sum_{i=1}^n \left(\int_{\mathcal{Y}} F(u|\mathbf{x}_i) \frac{1}{h} \mathbf{P}\left(\frac{u-y}{h}\right) \mathrm{d}G(u) \right) \frac{1}{h^d} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}} \right] \hat{\mathbf{S}}_{\mathbf{x}}^{-1} \mathbf{e}_\nu$$

$$= \theta_{\mu,\nu}(y,\mathbf{x}) + \mathsf{B}_{\mu,\nu}(y,\mathbf{x}) + o_{\mathbb{P}}\left(h^{\mathfrak{q}+1-|\nu|} + h^{\mathfrak{p}+1-\mu} \right),$$

where

$$\mathsf{B}_{\mu,\nu}(y,\mathbf{x}) = h^{\mathfrak{q}+1-|\nu|} \underbrace{\sum_{|\mathbf{m}|=\mathfrak{q}+1} \theta_{\mu,\mathbf{m}}(y,\mathbf{x}) \mathbf{c}_{\mathbf{x},\mathbf{m}}^{\mathrm{T}} \mathbf{S}_{\mathbf{x}}^{-1} \mathbf{e}_\nu}_{B_{(i),\mathfrak{q}+1}(y,\mathbf{x})} + h^{\mathfrak{p}+1-\mu} \underbrace{\theta_{\mathfrak{p}+1,\nu}(y,\mathbf{x}) \mathbf{c}_{y,\mathfrak{p}+1}^{\mathrm{T}} \mathbf{S}_y^{-1} \mathbf{e}_\mu}_{B_{(ii),\mathfrak{p}+1}(y,\mathbf{x})}.$$

Similarly,

$$\mathbf{e}_\mu^{\mathrm{T}} \hat{\mathbf{S}}_y^{-1} \left[\frac{1}{n^2 h^{\mu+|\nu|}} \sum_{i=1}^n \sum_{j=1}^n F(y_j|\mathbf{x}_i) \frac{1}{h} \mathbf{P}\left(\frac{y_j-y}{h}\right) \frac{1}{h^d} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}} \right] \hat{\mathbf{S}}_{\mathbf{x}}^{-1} \mathbf{e}_\nu$$

$$= \theta_{\mu,\nu}(y,\mathbf{x}) + \mathsf{B}_{\mu,\nu}(y,\mathbf{x}) + o_{\mathbb{P}}\left(h^{\mathfrak{q}+1-|\nu|} + h^{\mathfrak{p}+1-\mu} \right).$$

For future reference, we define the order of the leading bias as

$$\mathbf{r}_{\mathsf{B}} = h^{\mathfrak{q}+1-|\nu|} + h^{\mathfrak{p}+1-\mu}.$$

Remark SA-2.1 (Higher-order bias) Because the leading bias established in the lemma can be exactly zero, one may need to extract higher-order terms for bandwidth selection:

$$\mathsf{B}_{\mu,\nu}(y,\mathbf{x}) = h^{\mathfrak{q}+1-|\nu|} B_{(i),\mathfrak{q}+1}(y,\mathbf{x}) + h^{\mathfrak{p}+1-\mu} B_{(ii),\mathfrak{p}+1}(y,\mathbf{x})$$

$$+ h^{\mathfrak{q}+2-|\nu|} B_{(i),\mathfrak{q}+2}(y,\mathbf{x}) + h^{\mathfrak{p}+2-\mu} B_{(ii),\mathfrak{p}+2}(y,\mathbf{x}) + h^{\mathfrak{p}+\mathfrak{q}+2-\mu-|\nu|} B_{(iii),\mathfrak{p}+1,\mathfrak{q}+1}(y,\mathbf{x}),$$

where

$$B_{(i),\mathfrak{q}+2}(y,\mathbf{x}) = \sum_{|\mathbf{m}|=\mathfrak{q}+2} \theta_{\mu,\mathbf{m}}(y,\mathbf{x})\mathbf{c}_{\mathbf{x},\mathbf{m}}^{\mathrm{T}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}, \qquad B_{(ii),\mathfrak{p}+2}(y,\mathbf{x}) = \theta_{\mathfrak{p}+2,\boldsymbol{\nu}}(y,\mathbf{x})\mathbf{c}_{y,\mathfrak{p}+2}^{\mathrm{T}}\mathbf{S}_y^{-1}\mathbf{e}_{\mu},$$

$$B_{(iii),\mathfrak{p}+1,\mathfrak{q}+1}(y,\mathbf{x}) = \mathbf{e}_{\mu}^{\mathrm{T}}\mathbf{S}_y^{-1}\mathbf{c}_{y,\mathfrak{p}+1}\left(\sum_{|\mathbf{m}|=\mathfrak{q}+1}\theta_{\mathfrak{p}+1,\mathbf{m}}(y,\mathbf{x})\mathbf{c}_{\mathbf{x},\mathbf{m}}^{\mathrm{T}}\right)\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}.$$

Note that the last term, $h^{\mathfrak{p}+\mathfrak{q}+2-\mu-|\boldsymbol{\nu}|}B_{(iii),\mathfrak{p}+1,\mathfrak{q}+1}(y,\mathbf{x})$, is present only if $\mu = \mathfrak{p}$ and $|\boldsymbol{\nu}| = \mathfrak{q}$. ∎

Next we study the leading variance of our estimator, defined as

$$\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n}\mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(y_i,\mathbf{x}_i;y,\mathbf{x})\right].$$

Lemma SA-2.3 (Variance)
Let Assumptions SA-DGP, SA-K and SA-W hold with $h \to 0$ and $nh^d/\log n \to \infty$. Then
(i) $\mu = 0$ and $\theta_{0,\mathbf{0}} \neq 0$ or 1:

$$\mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{nh^{d+2|\boldsymbol{\nu}|}}\theta_{0,\mathbf{0}}(y,\mathbf{x})(1-\theta_{0,\mathbf{0}}(y,\mathbf{x}))\left(\mathbf{e}_{\boldsymbol{\nu}}^{\mathrm{T}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{T}_{\mathbf{x}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}\right) + O\left(\frac{1}{nh^{d+2|\boldsymbol{\nu}|-1}}\right).$$

(ii) $\mu = 0$ and $\theta_{0,\mathbf{0}} = 0$ or 1: $\mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x})$ has the order $\frac{1}{nh^{d+2|\boldsymbol{\nu}|-1}}$.
(iii) $\mu > 0$:

$$\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}}\theta_{1,\mathbf{0}}(y,\mathbf{x})\left(\mathbf{e}_{\mu}^{\mathrm{T}}\mathbf{S}_y^{-1}\mathbf{T}_y\mathbf{S}_y^{-1}\mathbf{e}_{\mu}\right)\left(\mathbf{e}_{\boldsymbol{\nu}}^{\mathrm{T}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{T}_{\mathbf{x}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}\right) + O\left(\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|-2}}\right).$$

For future reference, we will define

$$\mathsf{r}_{\mathsf{V}} = \sqrt{\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}}}.$$

Remark SA-2.2 (Vanishing boundary variance when $\mu = 0$) In case (ii), the true conditional distribution function is 0 or 1, which is why the leading variance shrinks faster. We do not provide a formula as the leading variance in this case takes a complicated form. ∎

Now, we propose two estimators for the variance that are valid for all three cases of Lemma SA-2.3, and hence will be useful for establishing a self-normalized distributional approximation later. Define

$$\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{n^2}\sum_{i=1}^{n}\check{\mathscr{K}}_{\mu,\boldsymbol{\nu},h}^{\circ}(y_i,\mathbf{x}_i;y,\mathbf{x})^2,$$

$$\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{n^2}\sum_{i=1}^{n}\hat{\mathscr{K}}_{\mu,\boldsymbol{\nu},h}^{\circ}(y_i,\mathbf{x}_i;y,\mathbf{x})^2.$$

Note that $\hat{V}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})$ is simply the plug-in variance estimator for $\hat{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})$ and $\check{V}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})$ is the plug-in variance estimator for $\check{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})$. The next lemma provides pointwise convergence results for the two variance estimators.

Lemma SA-2.4 (Variance estimation)
Let Assumptions SA-DGP, SA-K and SA-W hold with $h \to 0$ and $nh^{d+1}/\log n \to \infty$. In addition, $\theta_{0,\boldsymbol{\nu}}$ exists and is continuous for all $|\boldsymbol{\nu}| \leq \mathfrak{q} + 1$. Then
(i) $\mu = 0$ and $\theta_{0,\mathbf{0}} \neq 0$ or 1:

$$\left| \frac{\check{V}_{0,\boldsymbol{\nu}}(y, \mathbf{x}) - V_{0,\boldsymbol{\nu}}(y, \mathbf{x})}{V_{0,\boldsymbol{\nu}}(y, \mathbf{x})} \right| = O_{\mathbb{P}}\left(h^{\mathfrak{q}+1} + \sqrt{\frac{\log n}{nh^d}} \right).$$

(ii) $\mu > 0$, or $\theta_{0,\mathbf{0}} = 0$ or 1:

$$\left| \frac{\check{V}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) - V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})} \right| = O_{\mathbb{P}}\left(h^{\mathfrak{q}+\frac{1}{2}} + \sqrt{\frac{\log n}{nh^{d+1}}} \right).$$

Let $G = F_y$, then the same conclusions hold for $\hat{V}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})$.

Next, we study the large-sample distributional properties of the infeasible, standardized statistic

$$\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) = \frac{1}{n\sqrt{V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}} \sum_{i=1}^{n} \mathcal{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(y_i, \mathbf{x}_i; y, \mathbf{x}).$$

Note that this is equivalent to the scaled asymptotic linear representation of the estimator.

Theorem SA-2.1 (Asymptotic normality)
Let Assumptions SA-DGP, SA-K and SA-W hold with $h \to 0$. Then

$$\sup_{u \in \mathbb{R}} \left| \mathbb{P}\left[\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) \leq u \right] - \Phi(u) \right| = O\left(\mathbf{r}_{\mathtt{BE}} \right), \quad \text{where } \mathbf{r}_{\mathtt{BE}} = \begin{cases} \frac{1}{\sqrt{nh^d}} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{0}} \neq 0 \text{ or } 1 \\ \frac{1}{\sqrt{nh^{d+1}}} & \text{if } \mu > 0, \text{ or if } \theta_{0,\mathbf{0}} = 0 \text{ or } 1 \end{cases}.$$

While the theorem focuses on asymptotic normality of the infeasible t-statistic, $\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}^{\circ}(y, \mathbf{x})$, we show in the following remark that similar conclusions can be made for the t-statistics constructed with the estimators, $\hat{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})$ and $\check{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})$.

Remark SA-2.3 (Asymptotic normality of standardized statistics) We first introduce the statistic

$$\check{\mathbb{S}}_{\mu,\boldsymbol{\nu}}^{\circ}(y, \mathbf{x}) = \frac{\check{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) - \mathbb{E}\left[\check{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) | \mathbf{X} \right]}{\sqrt{V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}},$$

which is based on $\check{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})$. (In the main paper we directly center all statistics at the target parameter $\theta_{\mu,\boldsymbol{\nu}}$. For clarity, however, we will separate the discussion on distributional convergence

from the smoothing bias in this supplemental appendix. This is reflected by the superscript "circle.")
By combining the results of Lemmas SA-2.1 and SA-2.3, we have

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left| \check{\$}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \bar{\$}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \right| = O_{\mathbb{T}\mathbb{C}}\left(\frac{\log n}{\sqrt{nh^d}} \right).$$

As a result,

$$\sup_{u\in\mathbb{R}} \left| \mathbb{P}\left[\check{\$}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \le u \right] - \Phi(u) \right| = O\left(\frac{\log n}{\sqrt{nh^d}} + \mathtt{r}_{\mathrm{BE}} \right).$$

To present the pointwise distributional approximation result for the estimator $\hat{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$, we define the following statistic

$$\hat{\$}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{nh^{d+\mu+|\boldsymbol{\nu}|}\sqrt{V_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}} \sum_{i=1}^{n} \mathbf{e}_{\mu}^{\mathrm{T}} \hat{\mathbf{S}}_y^{-1} \left[\frac{1}{n}\sum_{j=1}^{n} \left[\mathbb{1}(y_i \le y_j) - F(y_j|\mathbf{x}_i) \right] \frac{1}{h}\mathbf{P}\left(\frac{y_j - y}{h} \right) \right]$$
$$\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h} \right)^{\mathrm{T}} \hat{\mathbf{S}}_{\mathbf{x}}^{-1} \mathbf{e}_{\boldsymbol{\nu}}.$$

It is worth mentioning that $\hat{\$}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ is not exactly centered, that is, it is not mean zero. Nevertheless, by the results of Lemmas SA-2.1 and SA-2.3, and the concentration inequality for second order U-statistics in Lemma SA-7.3, we have

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left| \hat{\$}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \check{\$}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \right| = O_{\mathbb{T}\mathbb{C}}\left(\frac{\log n}{\sqrt{nh^2}} \right),$$

for some constants $\mathfrak{c}_1$, $\mathfrak{c}_2$, and $\mathfrak{c}_3$. Then we can conclude that the coverage error satisfies

$$\sup_{u\in\mathbb{R}} \left| \mathbb{P}\left[\hat{\$}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \le u \right] - \Phi(u) \right| = O\left(\frac{\log n}{\sqrt{nh^{d\vee 2}}} + \mathtt{r}_{\mathrm{BE}} \right). \qquad \blacksquare$$

SA-3 Uniform Large-sample Properties

To conduct statistical inference on the entire function $\theta_{\mu,\boldsymbol{\nu}}$, such as constructing confidence bands or testing shape restrictions, we need uniform distributional approximations to our estimators. In this section, we will consider large-sample properties of our estimator which hold uniformly on $\mathcal{Y}\times\mathcal{X} = [0,1]^{d+1}$. We first establish the uniform convergence rate of our estimator.

Lemma SA-3.1 (Uniform rate of convergence)
Let Assumptions SA-DGP, SA-K and SA-W hold with $h \to 0$ and $nh^{d+1}/\log n \to \infty$. In addition, $\theta_{\mu',\boldsymbol{\nu}'}$ exists and is continuous for all $\mu' + |\boldsymbol{\nu}'| = \max\{\mathfrak{q}+1+\mu,\ \mathfrak{p}+1+|\boldsymbol{\nu}|\}$. Then

(i) $\mu = 0$:

$$\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} \left| \check{\theta}_{0,\boldsymbol{\nu}}(y,\mathbf{x}) - \theta_{0,\boldsymbol{\nu}}(y,\mathbf{x}) \right| = O_{\mathbb{TC}} \left(h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} + h^{\mathfrak{p}+1} + \sqrt{\frac{\log n}{nh^{d+2|\boldsymbol{\nu}|}}} \right);$$

(ii) $\mu > 0$:

$$\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} \left| \check{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \right| = O_{\mathbb{TC}} \left(h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} + h^{\mathfrak{p}+1-\mu} + \sqrt{\frac{\log n}{nh^{d+2\mu+2|\boldsymbol{\nu}|-1}}} \right).$$

In the next lemma, we characterize the uniform convergence rate of the variance estimators introduced in the previous section.

Lemma SA-3.2 (Uniform variance estimation)

Let Assumptions SA-DGP, SA-K and SA-W hold with $h \to 0$ and $nh^{d+1}/\log n \to \infty$. In addition, $\theta_{0,\boldsymbol{\nu}}$ exists and is continuous for all $|\boldsymbol{\nu}| \le \mathfrak{q}+1$. Then

$$\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} \left| \frac{\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})} \right| = O_{\mathbb{TC}}\left(\mathtt{r}_{\mathtt{VE}} \right), \quad \text{where } \mathtt{r}_{\mathtt{VE}} = h^{\mathfrak{q}+\frac{1}{2}} + \sqrt{\frac{\log n}{nh^{d+1}}}.$$

Let $G = F_y$, then the same conclusions hold for $\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$.

Now, we introduce the Studentized processes for each of the estimators, $\hat{\theta}_{\mu,\boldsymbol{\nu}}$ and $\check{\theta}_{\mu,\boldsymbol{\nu}}$:

$$\check{\mathbb{T}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \sqrt{\frac{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}{\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}} \check{\mathsf{S}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}), \qquad \hat{\mathbb{T}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \sqrt{\frac{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}{\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}} \hat{\mathsf{S}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}).$$

In the following lemma we study the error that arises from the Studentization of our estimators.

Lemma SA-3.3 (Studentization error)

Let Assumptions SA-DGP, SA-K and SA-W hold with $h \to 0$ and $nh^{d+1}/\log n \to \infty$. In addition, $\theta_{0,\boldsymbol{\nu}}$ exists and is continuous for all $|\boldsymbol{\nu}| \le \mathfrak{q}+1$. Then

$$\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} \left| \check{\mathbb{T}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \check{\mathsf{S}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \right| = O_{\mathbb{TC}}\left(\mathtt{r}_{\mathtt{SE}} \right), \qquad \text{where } \mathtt{r}_{\mathtt{SE}} = \sqrt{\log n} \, \mathtt{r}_{\mathtt{VE}}.$$

The same holds for $\hat{\mathbb{T}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \hat{\mathsf{S}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$.

Our next goal is to establish a uniform normal approximation to the process $\bar{\mathsf{S}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$. We first provide a few important properties of the equivalent kernel $\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}(a,\mathbf{b};y,\mathbf{x})$.

Lemma SA-3.4 (Boundedness and compact support)

Let Assumptions SA-K and SA-W hold. Then
(i) Both $\mathscr{K}_{\mu,\boldsymbol{\nu},h}(a,\mathbf{b};y,\mathbf{x})$ and $\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}(a,\mathbf{b};y,\mathbf{x})$ are bounded:

$$\sup_{a,\mathbf{b},y,\mathbf{x}} |\mathscr{K}_{\mu,\boldsymbol{\nu},h}(a,\mathbf{b};y,\mathbf{x})| + |\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}(a,\mathbf{b};y,\mathbf{x})| = O\left(h^{-d-\mu-|\boldsymbol{\nu}|} \right).$$

(ii) For any y and $\mathbf{x}$, $\mathscr{K}_{\mu,\boldsymbol{\nu},h}\left(\cdot,\cdot\ ;y,\mathbf{x}\right)$ is supported within an h-neighborhood of $(y,\mathbf{x}^{\mathrm{T}})^{\mathrm{T}}$ for all $\boldsymbol{\nu}$ and all $\mu \geq 1$.

Lemma SA-3.5 (Lipschitz continuity)

Let Assumptions SA-K and SA-W hold. Then

(i) Both $\mathscr{K}_{\mu,\boldsymbol{\nu},h}\left(a,\mathbf{b};y,\mathbf{x}\right)$ and $\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}\left(a,\mathbf{b};y,\mathbf{x}\right)$ are Lipschitz continuous with respect to a and $\mathbf{b}$:

$$\sup_{|a-a'|+|\mathbf{b}-\mathbf{b}'|>0,y,\mathbf{x}} \frac{\left|\mathscr{K}_{\mu,\boldsymbol{\nu},h}\left(a,\mathbf{b};y,\mathbf{x}\right)-\mathscr{K}_{\mu,\boldsymbol{\nu},h}\left(a',\mathbf{b}';y,\mathbf{x}\right)\right|+\left|\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}\left(a,\mathbf{b};y,\mathbf{x}\right)-\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}\left(a',\mathbf{b}';y,\mathbf{x}\right)\right|}{|a-a'|+|\mathbf{b}-\mathbf{b}'|}$$
$$= O\left(h^{-1-d-\mu-|\boldsymbol{\nu}|}\right).$$

(ii) Both $\mathscr{K}_{\mu,\boldsymbol{\nu},h}\left(a,\mathbf{b};y,\mathbf{x}\right)$ and $\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}\left(a,\mathbf{b};y,\mathbf{x}\right)$ are Lipschitz continuous with respect to y and $\mathbf{x}$:

$$\sup_{a,\mathbf{b},|y-y'|+|\mathbf{x}-\mathbf{x}'|>0} \frac{\left|\mathscr{K}_{\mu,\boldsymbol{\nu},h}\left(a,\mathbf{b};y,\mathbf{x}\right)-\mathscr{K}_{\mu,\boldsymbol{\nu}}\left(a,\mathbf{b};y',\mathbf{x}'\right)\right|+\left|\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}\left(a,\mathbf{b};y,\mathbf{x}\right)-\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu}}\left(a,\mathbf{b};y',\mathbf{x}'\right)\right|}{|y-y'|+|\mathbf{x}-\mathbf{x}'|}$$
$$= O\left(h^{-1-d-\mu-|\boldsymbol{\nu}|}\right).$$

Next, we prove a general result on the uniform covering number for function classes consisting of kernels, which may be of independent interest. Importantly, we allow the kernels in the function class to take different shapes. This is crucial for our purpose as the implied kernel in our estimator is boundary adaptive, and hence will take different forms for interior and boundary evaluation points.

Lemma SA-3.6 (Covering number)

Let $\mathcal{G} = \left\{g_{\mathbf{z}}\left(\frac{\cdot-\mathbf{z}}{h}\right), \mathbf{z} \in [0,1]^{d}\right\}$ be a class of functions, and $h > 0$. Let $\mathfrak{c}$ be a generic constant that does not depend on h. Assume

(i) boundedness $\sup_{\mathbf{z}} \sup_{\mathbf{z}'} |g_{\mathbf{z}}(\mathbf{z}')| \leq \mathfrak{c}$

(ii) compact support $\operatorname{supp}(g_{\mathbf{z}}(\cdot)) \subseteq [-\mathfrak{c},\mathfrak{c}]^{d},\ \forall\ \mathbf{z}$

(iii) Lipschitz continuity $\sup_{\mathbf{z}} |g_{\mathbf{z}}(\mathbf{z}') - g_{\mathbf{z}}(\mathbf{z}'')| \leq \mathfrak{c}|\mathbf{z}'-\mathbf{z}''|$

                             $\sup_{\mathbf{z}} |g_{\mathbf{z}'}(\mathbf{z}) - g_{\mathbf{z}''}(\mathbf{z})| \leq \mathfrak{c}h^{-1}|\mathbf{z}'-\mathbf{z}''|.$

Then, for any probability measure P, the $L^{1}(P)$-covering number of the class $\mathcal{G}$ satisfies

$$N\left((2\mathfrak{c}+1)^{d+1}\varepsilon,\ \mathcal{G},\ L^{1}(P)\right) \leq \mathfrak{c}'\frac{1}{\varepsilon^{d+1}} + 1,$$

where $\mathfrak{c}'$ is some constant that depends only on $\mathfrak{c}$ and d.

Remark SA-3.1 (On the covering number) This rate, ε^{-d-1}, is clearly suboptimal for very small ε. The reason is that when we fix h and consider how the covering number changes as $\varepsilon \downarrow 0$, the optimal rate is ε^{-d}, as in this case the class of functions is fixed (c.f. Theorem 2.7.11 in van der

Vaart and Wellner 1996). Such suboptimality is introduced because we prefer a covering number that depends only on ε (but not h). The result we derived performs better for moderate and large ε (relative to h).

Now consider how the above (a sharper result for moderate and large ε) manifests itself in our proof. Take a fixed ε. As the bandwidth shrinks to 0, we will be employing finer partitions of $[0,1]^d$. However, not all of the sets in the partition matter for bounding the covering number, because there are at most ε^{-1} sets carrying a probability mass larger than ε. Given that the functions we consider have compact support, most of them become irrelevant in our calculation of the covering number. Indeed, a function only makes a nontrivial contribution if its support intersects with some set in the (very fine) partition whose P-measure exceeds ε. Therefore, instead of considering all h^{-d} partitions, we only need to focus on ε^{-1} of them, which is why an extra ε^{-1} term is introduced. ∎

Corollary SA-3.1

Let Assumptions SA-K and SA-W hold, and $\mu \geq 1$. Then the function class,

$$\mathcal{K} = \left\{ h^{d+\mu+|\boldsymbol{\nu}|} \mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(\cdot,\cdot;y,\mathbf{x}): \ y \in \mathcal{Y}, \ \mathbf{x} \in \mathcal{X} \right\},$$

satisfies

$$\sup_{P} N\left(\varepsilon, \ \mathcal{K}, \ L^1(\mathbb{P})\right) \leq \mathfrak{c} \frac{1}{\varepsilon^{d+2}} + 1,$$

where the supremum is taken over all probability measures on $[0,1]^{d+1}$, and the constant $\mathfrak{c}$ does not depend on the bandwidth h.

Building on the properties of the equivalent kernel that we just established, we provide a uniform normal approximation to $\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ in the following lemma.

Theorem SA-3.1 (Strong approximation)

Let Assumptions SA-DGP, SA-K and SA-W hold with $h \to 0$ and $nh^{d+1}/\log n \to \infty$. Also assume $\mu \geq 1$. Define

$$\mathbf{r}_{\mathtt{SA}} = \left(\frac{\log^{d+1} n}{nh^{d+1}} \right)^{\frac{1}{2d+2}}.$$

Then there exist two centered processes, $\bar{\mathbb{S}}'_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ and $\mathbb{G}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$, such that (i) $\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ and $\bar{\mathbb{S}}'_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ have the same distribution, (ii) $\mathbb{G}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ is a Gaussian process and has the same covariance kernel as $\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$, and (iii) for some constants $\mathfrak{c}_1$, $\mathfrak{c}_2$, and $\mathfrak{c}_3$,

$$\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} \left| \bar{\mathbb{S}}'_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \mathbb{G}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \right| = O_{\mathbb{TC}}\left(\mathbf{r}_{\mathtt{SA}} \right).$$

and $\mathfrak{c}_3$ can be made arbitrarily large by appropriate choices of $\mathfrak{c}_1$.

The Gaussian approximation in the above lemma is not feasible, as its covariance kernel depends on unknowns. To be more precise, the covariance kernel takes the form

$$\mathsf{C}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') = \mathbb{C}\text{ov}\left[\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}),\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y',\mathbf{x}')\right] = \frac{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}')}{\sqrt{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\mathsf{V}_{\mu,\boldsymbol{\nu}}(y',\mathbf{x}')}},$$

where

$$\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') = \frac{1}{n}\mathbb{C}\text{ov}\left[\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}(y_i,\mathbf{x}_i;y,\mathbf{x}),\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}(y_i,\mathbf{x}_i;y',\mathbf{x}')\right].$$

We consider two estimators of the covariance kernel

$$\check{\mathsf{C}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') = \frac{\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}')}{\sqrt{\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y',\mathbf{x}')}}, \quad \hat{\mathsf{C}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') = \frac{\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}')}{\sqrt{\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y',\mathbf{x}')}},$$

and

$$\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') = \frac{1}{n^2}\sum_{i=1}^{n}\check{\mathscr{K}}^{\circ}_{\mu,\boldsymbol{\nu},h}(y_i,\mathbf{x}_i;y,\mathbf{x})\check{\mathscr{K}}^{\circ}_{\mu,\boldsymbol{\nu},h}(y_i,\mathbf{x}_i;y',\mathbf{x}')$$

$$\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') = \frac{1}{n^2}\sum_{i=1}^{n}\hat{\mathscr{K}}^{\circ}_{\mu,\boldsymbol{\nu},h}(y_i,\mathbf{x}_i;y,\mathbf{x})\hat{\mathscr{K}}^{\circ}_{\mu,\boldsymbol{\nu},h}(y_i,\mathbf{x}_i;y',\mathbf{x}').$$

Lemma SA-3.7 (Uniform consistency of the covariance estimator)
Let Assumptions SA-DGP, SA-K and SA-W hold with $h \to 0$ and $nh^{d+1}/\log n \to \infty$. In addition, $\theta_{0,\boldsymbol{\nu}}$ exists and is continuous for all $|\boldsymbol{\nu}| \le \mathfrak{q}+1$. Then

$$\sup_{y,y'\in\mathcal{Y},\mathbf{x},\mathbf{x}'\in\mathcal{X}}\left|\check{\mathsf{C}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') - \mathsf{C}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}')\right| = O_{\mathbb{TC}}\left(\mathtt{r}_{\mathtt{VE}}\right),$$

where $\mathtt{r}_{\mathtt{VE}}$ is defined in Lemma SA-3.2. Let $G = F_y$, then the same conclusion holds for $\hat{\mathsf{C}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}')$.

Lemma SA-3.8 (Gaussian comparison)
[Omitted long matching line]

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \le u \middle| \mathbf{Y},\mathbf{X}\right] - \mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\mathbb{G}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \le u\right]\right| = O_{\mathbb{P}}\left((\log n)\sqrt{\mathtt{r}_{\mathtt{VE}}}\right),$$

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\hat{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \le u \middle| \mathbf{Y},\mathbf{X}\right] - \mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\mathbb{G}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \le u\right]\right| = O_{\mathbb{P}}\left((\log n)\sqrt{\mathtt{r}_{\mathtt{VE}}}\right).$$

Theorem SA-3.2 (Feasible normal approximation)
Let Assumptions SA-DGP, SA-K and SA-W hold with $h \to 0$ and $nh^{d+1}/\log n \to \infty$. In addition,

16

$\theta_{0,\boldsymbol{\nu}}$ *exists and is continuous for all* $|\boldsymbol{\nu}| \leq \mathfrak{q}+1$. *Also assume* $\mu \geq 1$. *Then*

$$\sup_{u\in\mathbb{R}} \left| \mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\check{\mathbb{T}}^\circ_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u \right] - \mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\check{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u \,\bigg|\, \mathbf{Y},\mathbf{X} \right] \right| = O_\mathbb{P}\left(\sqrt{\log n}\mathbf{r}_{\mathtt{SA}} + (\log n)\sqrt{\mathbf{r}_{\mathtt{VE}}} \right),$$

$$\sup_{u\in\mathbb{R}} \left| \mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\hat{\mathbb{T}}^\circ_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u \right] - \mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\hat{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u \,\bigg|\, \mathbf{Y},\mathbf{X} \right] \right| = O_\mathbb{P}\left(\sqrt{\log n}\mathbf{r}_{\mathtt{SA}} + (\log n)\sqrt{\mathbf{r}_{\mathtt{VE}}} \right).$$

SA-4   Applications

SA-4.1   Confidence Bands

A natural corollary of Theorem SA-3.2 is that one can employ critical values computed from $\check{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ and $\hat{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ to construct confidence bands. To be very precise, define

$$\check{\mathbf{cv}}_{\mu,\boldsymbol{\nu}}(\alpha) = \inf\left\{ u: \ \mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\check{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u \,\bigg|\, \mathbf{Y},\mathbf{X} \right] \geq 1-\alpha \right\},$$

$$\hat{\mathbf{cv}}_{\mu,\boldsymbol{\nu}}(\alpha) = \inf\left\{ u: \ \mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\hat{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u \,\bigg|\, \mathbf{Y},\mathbf{X} \right] \geq 1-\alpha \right\}.$$

Then level $(1-\alpha)$ confidence bands can be constructed as

$$\check{\mathcal{C}}_{\mu,\boldsymbol{\nu}}(1-\alpha) = \left\{ \check{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \pm \check{\mathbf{cv}}_{\mu,\boldsymbol{\nu}}(\alpha)\sqrt{\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}: \ (y,\mathbf{x}) \in \mathcal{Y}\times\mathcal{X} \right\},$$

$$\hat{\mathcal{C}}_{\mu,\boldsymbol{\nu}}(1-\alpha) = \left\{ \hat{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \pm \hat{\mathbf{cv}}_{\mu,\boldsymbol{\nu}}(\alpha)\sqrt{\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}: \ (y,\mathbf{x}) \in \mathcal{Y}\times\mathcal{X} \right\},$$

whose coverage error is given in the following theorem.

Theorem SA-4.1 (Confidence band)
Consider the setting of Theorem SA-3.2. In addition, $\theta_{\mu',\boldsymbol{\nu}'}$ exists and is continuous for all $\mu'+|\boldsymbol{\nu}'| = \max\{\mathfrak{q}+1+\mu, \ \mathfrak{p}+1+|\boldsymbol{\nu}|\}$. Then

$$\mathbb{P}\left[\theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \in \check{\mathcal{C}}_{\mu,\boldsymbol{\nu}}(1-\alpha), \ \forall(y,\mathbf{x}) \in \mathcal{Y}\times\mathcal{X} \right] \geq 1-\alpha - O\left(\sqrt{\log n}\left(\mathbf{r}_{\mathtt{SA}} + \frac{\mathbf{r}_{\mathtt{B}}}{\mathbf{r}_{\mathtt{V}}} \right) + (\log n)\sqrt{\mathbf{r}_{\mathtt{VE}}} \right),$$

$$\mathbb{P}\left[\theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \in \hat{\mathcal{C}}_{\mu,\boldsymbol{\nu}}(1-\alpha), \ \forall(y,\mathbf{x}) \in \mathcal{Y}\times\mathcal{X} \right] \geq 1-\alpha - O\left(\sqrt{\log n}\left(\mathbf{r}_{\mathtt{SA}} + \frac{\mathbf{r}_{\mathtt{B}}}{\mathbf{r}_{\mathtt{V}}} \right) + (\log n)\sqrt{\mathbf{r}_{\mathtt{VE}}} \right).$$

SA-4.2   Parametric Specification Testing

In applications, it is not uncommon to estimate conditional densities or higher-order derivatives by specifying a parametric family of distributions. While such parametric restrictions may provide reasonable approximations, it is still worthwhile to conduct specification testing. To be specific,

17

assume the researcher postulates the following class

$$\left\{ \theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x};\boldsymbol{\gamma}) : \ \boldsymbol{\gamma} \in \Gamma_{\mu,\boldsymbol{\nu}} \right\},$$

where $\Gamma_{\mu,\boldsymbol{\nu}}$ is some compact parameter space. We abstract away from the specifics of the estimation technique, and assume that the researcher also picks some estimator (maximum likelihood, minimum distance, etc.) $\hat{\boldsymbol{\gamma}}$. Under fairly mild conditions, the estimator will converge in probability to some (possibly pseudo-true) parameter $\bar{\boldsymbol{\gamma}}$ in the parameter space $\Gamma_{\mu,\boldsymbol{\nu}}$. As before, we will denote the true parameter as $\theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$, and consider the following competing hypotheses:

$$\mathsf{H}_0: \ \theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x};\bar{\boldsymbol{\gamma}}) = \theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \quad \text{vs.} \quad \mathsf{H}_1: \ \theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x};\bar{\boldsymbol{\gamma}}) \neq \theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}).$$

The test statistics we employ takes the following form

$$\check{\mathbb{T}}_{\mathsf{PS}}(y,\mathbf{x}) = \frac{\check{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x};\hat{\boldsymbol{\gamma}})}{\sqrt{\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}}, \quad \hat{\mathbb{T}}_{\mathsf{PS}}(y,\mathbf{x}) = \frac{\hat{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x};\hat{\boldsymbol{\gamma}})}{\sqrt{\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}}.$$

Theorem SA-4.2 (Parametric specification testing)

Consider the setting of Theorem SA-3.2. In addition, $\theta_{\mu',\boldsymbol{\nu}'}$ exists and is continuous for all $\mu' + |\boldsymbol{\nu}'| = \max\{\mathfrak{q} + 1 + \mu, \ \mathfrak{p} + 1 + |\boldsymbol{\nu}|\}$. Assume the parametric estimate satisfies

$$\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} |\theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x};\hat{\boldsymbol{\gamma}}) - \theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x};\bar{\boldsymbol{\gamma}})| = O_{\mathbb{TC}}\left(\mathtt{r}_{\mathsf{PS}}\right),$$

for some $\mathtt{r}_{\mathsf{PS}}$. Then under the null hypothesis,

$$\mathbb{P}\left[\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} |\check{\mathbb{T}}_{\mathsf{PS}}(y,\mathbf{x})| > \check{\mathtt{cv}}_{\mu,\boldsymbol{\nu}}(\alpha) \right] \leq \alpha + O\left(\sqrt{\log n}\left(\mathtt{r}_{\mathsf{SA}} + \frac{\mathtt{r}_{\mathsf{B}} + \mathtt{r}_{\mathsf{PS}}}{\mathtt{r}_{\mathsf{V}}} \right) + (\log n)\sqrt{\mathtt{r}_{\mathsf{VE}}} \right),$$

$$\mathbb{P}\left[\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} |\hat{\mathbb{T}}_{\mathsf{PS}}(y,\mathbf{x})| > \hat{\mathtt{cv}}_{\mu,\boldsymbol{\nu}}(\alpha) \right] \leq \alpha + O\left(\sqrt{\log n}\left(\mathtt{r}_{\mathsf{SA}} + \frac{\mathtt{r}_{\mathsf{B}} + \mathtt{r}_{\mathsf{PS}}}{\mathtt{r}_{\mathsf{V}}} \right) + (\log n)\sqrt{\mathtt{r}_{\mathsf{VE}}} \right).$$

SA-4.3 Testing Shape Restrictions

Now consider shape restrictions on the conditional density or its derivatives. Let $c(y,\mathbf{x})$ be a pre-specified function, and we study the following one-sided competing hypotheses.

$$\mathsf{H}_0: \theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \leq c(y,\mathbf{x}) \quad \text{vs.} \quad \mathsf{H}_1: \theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) > c(y,\mathbf{x}).$$

The statistic we employ takes the form

$$\check{\mathbb{T}}_{\mathsf{SR}}(y,\mathbf{x}) = \frac{\check{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - c(y,\mathbf{x})}{\sqrt{\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}}, \quad \hat{\mathbb{T}}_{\mathsf{SR}}(y,\mathbf{x}) = \frac{\hat{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - c(y,\mathbf{x})}{\sqrt{\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}}.$$

18

and we will reject the null hypothesis if the test statistic exceeds a critical value.

Theorem SA-4.3 (Shape restriction testing)

Consider the setting of Theorem SA-3.2. In addition, $\theta_{\mu',\nu'}$ exists and is continuous for all $\mu' + |\nu'| = \max\{\mathfrak{q} + 1 + \mu, \ \mathfrak{p} + 1 + |\nu|\}$. Then under the null hypothesis,

$$
\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \check{\mathbb{T}}_{\mathsf{SR}}(y,\mathbf{x}) > \check{\mathtt{cv}}_{\mu,\nu}(\alpha)\right] \leq \alpha + O\left(\sqrt{\log n}\left(\mathtt{r}_{\mathsf{SA}} + \frac{\mathtt{r}_\mathsf{B} + \mathtt{r}_{\mathsf{PS}}}{\mathtt{r}_\mathsf{V}}\right) + (\log n)\sqrt{\mathtt{r}_{\mathsf{VE}}}\right),
$$

$$
\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \hat{\mathbb{T}}_{\mathsf{SR}}(y,\mathbf{x}) > \hat{\mathtt{cv}}_{\mu,\nu}(\alpha)\right] \leq \alpha + O\left(\sqrt{\log n}\left(\mathtt{r}_{\mathsf{SA}} + \frac{\mathtt{r}_\mathsf{B} + \mathtt{r}_{\mathsf{PS}}}{\mathtt{r}_\mathsf{V}}\right) + (\log n)\sqrt{\mathtt{r}_{\mathsf{VE}}}\right).
$$

SA-5 Bandwidth Selection

We assume throughout this section that $\mu > 0$. Using the bias expression derived in Lemma SA-2.2, and the leading variance is as characterized in Lemma SA-2.3, we can derive precise expressions for bandwidth selection.

SA-5.1 Pointwise Asymptotic MSE Minimization

Following from Fan and Gijbels (1996), the pointwise MSE-optimal bandwidth is defined as a minimizer of the following optimization problem

$$
h^*_{\mathfrak{p},\mathfrak{q},\mu,\nu}(y,\mathbf{x}) = \operatorname*{argmin}_{h>0}\left[\mathsf{V}_{\mu,\nu}(y,\mathbf{x}) + \mathsf{B}^2_{\mu,\nu}(y,\mathbf{x})\right]
$$

The solution to this equation gives an MSE-optimal bandwidth that depends on (i) the order of the polynomials, (ii) the order of the derivative to be estimated, and (iii) the position of the evaluation point.

Case 1: $\mathfrak{q} - |\nu| = \mathfrak{p} - \mu$, odd

In this case, both the leading bias constants, $B_{(i),\mathfrak{q}+1}(y,\mathbf{x})$ and $B_{(ii),\mathfrak{p}+1}(y,\mathbf{x})$, are nonzero. Therefore, the MSE-optimal bandwidth is

$$
\begin{aligned}
h^*_{\mathfrak{p},\mathfrak{q},\mu,\nu}(y,\mathbf{x}) &= \operatorname*{argmin}_{h>0}\left[\frac{1}{nh^{d+2|\nu|+2\mu-1}}V_{\mu,\nu}(y,\mathbf{x}) + h^{\mathfrak{p}+\mathfrak{q}+2-\mu-|\nu|}\left(B_{(i),\mathfrak{q}+1}(y,\mathbf{x}) + B_{(ii),\mathfrak{p}+1}(y,\mathbf{x})\right)^2\right] \\
&= \left[\frac{(d+2|\nu|+2\mu-1)V_{\mu,\nu}(y,\mathbf{x})}{(\mathfrak{p}+\mathfrak{q}+2-\mu-|\nu|)\left(B_{(i),\mathfrak{q}+1}(y,\mathbf{x}) + B_{(ii),\mathfrak{p}+1}(y,\mathbf{x})\right)^2}\frac{1}{n}\right]^{\frac{1}{d+\mathfrak{p}+\mathfrak{q}+|\nu|+\mu+1}}.
\end{aligned}
$$

Case 2: $\mathfrak{q} - |\boldsymbol{\nu}| = \mathfrak{p} - \mu$, even; either x or y is at or near the boundary

In this case, at least one of the leading bias constants, $B_{(i),\mathfrak{q}+1}(y, \mathbf{x})$ and $B_{(ii),\mathfrak{p}+1}(y, \mathbf{x})$, is nonzero. Therefore, the MSE-optimal bandwidth is the same as in **Case 1**:

$$
\begin{aligned}
h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y, \mathbf{x}) &= \operatorname*{argmin}_{h>0} \left[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) + h^{\mathfrak{p}+\mathfrak{q}+2-\mu-|\boldsymbol{\nu}|} \left(B_{(i),\mathfrak{q}+1}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+1}(y, \mathbf{x}) \right)^2 \right] \\
&= \left[\frac{(d + 2|\boldsymbol{\nu}| + 2\mu - 1) V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{(\mathfrak{p} + \mathfrak{q} + 2 - \mu - |\boldsymbol{\nu}|) \left(B_{(i),\mathfrak{q}+1}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+1}(y, \mathbf{x}) \right)^2} \frac{1}{n} \right]^{\frac{1}{d+\mathfrak{p}+\mathfrak{q}+|\boldsymbol{\nu}|+\mu+1}} .
\end{aligned}
$$

Case 3: $\mathfrak{q} - |\boldsymbol{\nu}| = \mathfrak{p} - \mu \neq 0$, even; both x and y are interior

In this case, both leading bias constants are zero. Therefore, the MSE-optimal bandwidth will depend on higher-order bias terms:

$$
\begin{aligned}
h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y, \mathbf{x}) &= \operatorname*{argmin}_{h>0} \left[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) + h^{\mathfrak{p}+\mathfrak{q}+4-\mu-|\boldsymbol{\nu}|} \left(B_{(i),\mathfrak{q}+2}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+2}(y, \mathbf{x}) \right)^2 \right] \\
&= \left[\frac{(d + 2|\boldsymbol{\nu}| + 2\mu - 1) V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{(\mathfrak{p} + \mathfrak{q} + 4 - \mu - |\boldsymbol{\nu}|) \left(B_{(i),\mathfrak{q}+2}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+2}(y, \mathbf{x}) \right)^2} \frac{1}{n} \right]^{\frac{1}{d+\mathfrak{p}+\mathfrak{q}+|\boldsymbol{\nu}|+\mu+3}} .
\end{aligned}
$$

Case 4: $\mathfrak{q} - |\boldsymbol{\nu}| = \mathfrak{p} - \mu = 0$, even; both x and y are interior

As in **Case 3**, both leading bias constants are zero. The difference, however, is that the leading bias will involve an extra term:

$$
\begin{aligned}
h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y, \mathbf{x}) &= \operatorname*{argmin}_{h>0} \left[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) + h^4 \left(B_{(i),\mathfrak{q}+2}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+2}(y, \mathbf{x}) + B_{(iii),\mathfrak{p}+1,\mathfrak{q}+1}(y, \mathbf{x}) \right)^2 \right] \\
&= \left[\frac{(d + 2|\boldsymbol{\nu}| + 2\mu - 1) V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{4 \left(B_{(i),\mathfrak{q}+2}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+2}(y, \mathbf{x}) + B_{(iii),\mathfrak{p}+1,\mathfrak{q}+1}(y, \mathbf{x}) \right)^2} \frac{1}{n} \right]^{\frac{1}{d+2|\boldsymbol{\nu}|+2\mu+3}} .
\end{aligned}
$$

Case 5: $\mathfrak{q} - |\boldsymbol{\nu}| < \mathfrak{p} - \mu$, $\mathfrak{q} - |\boldsymbol{\nu}|$ odd

In this case, the leading bias will involve only one term:

$$
\begin{aligned}
h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y, \mathbf{x}) &= \operatorname*{argmin}_{h>0} \left[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) + h^{2\mathfrak{q}+2-2|\boldsymbol{\nu}|} B_{(i),\mathfrak{q}+1}(y, \mathbf{x})^2 \right] \\
&= \left[\frac{(d + 2|\boldsymbol{\nu}| + 2\mu - 1) V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{(2\mathfrak{q} + 2 - 2|\boldsymbol{\nu}|) B_{(i),\mathfrak{q}+1}(y, \mathbf{x})^2} \frac{1}{n} \right]^{\frac{1}{d+2\mathfrak{q}+2\mu+1}} .
\end{aligned}
$$

Case 6: $\mathfrak{q} - |\boldsymbol{\nu}| = \mathfrak{p} - \mu - 1$, $\mathfrak{q} - |\boldsymbol{\nu}|$ even; x is interior

In this case, the leading bias will involve two terms:

$$
h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \underset{h>0}{\operatorname{argmin}} \left[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) + h^{\mathfrak{p}+\mathfrak{q}+3-\mu-|\boldsymbol{\nu}|} \left(B_{(i),\mathfrak{q}+2}(y,\mathbf{x}) + B_{(ii),\mathfrak{p}+1}(y,\mathbf{x}) \right)^2 \right]
$$

$$
= \left[\frac{(d+2|\boldsymbol{\nu}|+2\mu-1)V_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}{(\mathfrak{p}+\mathfrak{q}+3-\mu-|\boldsymbol{\nu}|)\left(B_{(i),\mathfrak{q}+2}(y,\mathbf{x}) + B_{(ii),\mathfrak{p}+1}(y,\mathbf{x}) \right)^2} \frac{1}{n} \right]^{\frac{1}{d+\mathfrak{p}+\mathfrak{q}+\mu+|\boldsymbol{\nu}|+2}}.
$$

Case 7: $\mathfrak{q} - |\boldsymbol{\nu}| < \mathfrak{p} - \mu - 1$, $\mathfrak{q} - |\boldsymbol{\nu}|$ even; x is interior

In this case, the leading bias will involve only one term:

$$
h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \underset{h>0}{\operatorname{argmin}} \left[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) + h^{2\mathfrak{q}+4-2|\boldsymbol{\nu}|} B_{(i),\mathfrak{q}+2}(y,\mathbf{x})^2 \right]
$$

$$
= \left[\frac{(d+2|\boldsymbol{\nu}|+2\mu-1)V_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}{(2\mathfrak{q}+4-2|\boldsymbol{\nu}|)B_{(i),\mathfrak{q}+2}(y,\mathbf{x})^2} \frac{1}{n} \right]^{\frac{1}{d+2\mathfrak{q}+2\mu+3}}.
$$

Case 8: $\mathfrak{q} - |\boldsymbol{\nu}| > \mathfrak{p} - \mu$, $\mathfrak{p} - \mu$ odd

In this case, the leading bias will involve only one term:

$$
h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \underset{h>0}{\operatorname{argmin}} \left[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) + h^{2\mathfrak{p}+2-2\mu} B_{(ii),\mathfrak{p}+1}(y,\mathbf{x})^2 \right]
$$

$$
= \left[\frac{(d+2|\boldsymbol{\nu}|+2\mu-1)V_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}{(2\mathfrak{p}+2-2\mu)B_{(ii),\mathfrak{p}+1}(y,\mathbf{x})^2} \frac{1}{n} \right]^{\frac{1}{d+2\mathfrak{p}+2|\boldsymbol{\nu}|+1}}.
$$

Case 9: $\mathfrak{q} - |\boldsymbol{\nu}| - 1 = \mathfrak{p} - \mu$, $\mathfrak{q} - |\boldsymbol{\nu}|$ even; y is interior

In this case, the leading bias will involve two terms:

$$
h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \underset{h>0}{\operatorname{argmin}} \left[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) + h^{\mathfrak{p}+\mathfrak{q}+3-\mu-|\boldsymbol{\nu}|} \left(B_{(i),\mathfrak{q}+1}(y,\mathbf{x}) + B_{(ii),\mathfrak{p}+2}(y,\mathbf{x}) \right)^2 \right]
$$

$$
= \left[\frac{(d+2|\boldsymbol{\nu}|+2\mu-1)V_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}{(\mathfrak{p}+\mathfrak{q}+3-\mu-|\boldsymbol{\nu}|)\left(B_{(i),\mathfrak{q}+1}(y,\mathbf{x}) + B_{(ii),\mathfrak{p}+2}(y,\mathbf{x}) \right)^2} \frac{1}{n} \right]^{\frac{1}{d+\mathfrak{p}+\mathfrak{q}+\mu+|\boldsymbol{\nu}|+2}}.
$$

Case 10: $\mathfrak{q} - |\boldsymbol{\nu}| - 1 > \mathfrak{p} - \mu$, $\mathfrak{p} - \mu$ even; y is interior

In this case, the leading bias will involve only one term:

$$
h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \underset{h>0}{\operatorname{argmin}} \left[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) + h^{2\mathfrak{p}+4-2\mu} B_{(ii),\mathfrak{p}+2}(y,\mathbf{x})^2 \right]
$$

$$
= \left[\frac{(d+2|\boldsymbol{\nu}|+2\mu-1)V_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}{(2\mathfrak{p}+4-2\mu)B_{(ii),\mathfrak{p}+2}(y,\mathbf{x})^2} \frac{1}{n} \right]^{\frac{1}{d+2\mathfrak{p}+2|\boldsymbol{\nu}|+3}}.
$$

SA-5.2 Rule-of-thumb Bandwidth Selection

This section outlines the methodology that the companion `R` package, `lpcde`, uses to construct the rule-of-thumb bandwidth selection.

The rule-of-thumb estimation uses the following assumptions in order to compute the optimal bandwidth:

- the data is jointly normal,

- $\mathbf{X}$ and $\mathbf{Y}$ are independent, and,

- $p - \mu = q - |\nu|$ are odd.

Using these assumptions, each of the terms in the formula given in Case 1 of Section SA-5.1 are computed as follows:

1. The densities and relevant derivatives are evaluated based on the joint normal distribution assumption.

2. $\mathbf{S}_y$, $\mathbf{T}_y$, $\mathbf{T}_\mathbf{x}$ and $\mathbf{S}_\mathbf{x}$ matrices are computed by plugging in for the range of the data, the evaluation point, the respective marginal densities, and the kernel used.

3. Similarly, the $\mathbf{c}_y$ and $\mathbf{c}_\mathbf{x}$ vectors are computed by using the range of the data, the evaluation point, kernel function, and the respective marginal densities.

4. Bias and variance estimates are constructed using the relevant entries of the vectors and matrices.

SA-6 Alternative Variance Estimators

SA-6.1 V-statistic variance estimator

We propose here an alternative variance estimator that is quick to implement in practice. We start by first observing that the estimator $\widehat{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})$ is a V-statistic. That is,

$$
\begin{aligned}
\widehat{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) &= \frac{1}{n^2} \sum_{i,j} \mathbb{1}(y_i \leq y_j) \mathbf{e}_\mu^{\mathrm{T}} \widehat{\mathbf{S}}_y^{-1} \mathbf{P}\left(\frac{y_j - y}{h}\right) \mathbf{Q}^{\mathrm{T}}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \widehat{\mathbf{S}}_\mathbf{x}^{-1} \mathbf{e}_{\boldsymbol{\nu}} \\
&= \frac{1}{n^2} \sum_{i=1}^{n} a(y_i, y) b(\mathbf{x}_i, \mathbf{x}) + \frac{1}{n^2} \sum_{i \neq j} \mathbb{1}(y_i \leq y_j) a(y_j, y) b(\mathbf{x}_i, \mathbf{x}), \quad \text{(SA-6.1)}
\end{aligned}
$$

where,

$$
a(y_j, y) = h^{1+\mu} \mathbf{e}_\mu^{\mathrm{T}} \widehat{\mathbf{S}}_y^{-1} \mathbf{P}\left(\frac{y_j - y}{h}\right), \qquad b(\mathbf{x}_i, \mathbf{x}) = h^{d+|\boldsymbol{\nu}|} \mathbf{e}_{\boldsymbol{\nu}}^{\mathrm{T}} \widehat{\mathbf{S}}_\mathbf{x}^{-1} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right).
$$

Note that $a(\cdot)$ and $b(\cdot)$ are scalar functions that are non-zero only for data points that are within h distance of the evaluation point. The second term in Equation SA-6.1 can now be symmetrized and treated as a U-statistic. Applying the Hoeffding decomposition to the symmetrized version of the second term and plugging it back into Equation SA-6.1, we get

$$\widehat{\theta}_{\mu,\nu}(y,\mathbf{x}) = \frac{1}{n}\mathbb{E}\left[a(y_i,y)b(\mathbf{x}_i,\mathbf{x})\right] + \frac{n-1}{n}\mathbb{E}[u_{i,j}]$$
$$+ \frac{1}{n^2}\sum_{i=1}^{n}\left(a(y_i,y)b(\mathbf{x}_i,\mathbf{x}) - \mathbb{E}\left[a(y_i,y)b(\mathbf{x}_i,\mathbf{x})\right]\right) + \frac{n-1}{n}L_{\mu,\nu}(y,\mathbf{x})$$
$$+ \frac{n-1}{n}W_{\mu,\nu}(y,\mathbf{x}) \tag{SA-6.2}$$

where

$$L_{\mu,\nu}(y,\mathbf{x}) = \frac{1}{n}\sum_{i} 2\left(\mathbb{E}\left[u_{i,j}|y_i,\mathbf{x}_i\right] - \mathbb{E}\left[u_{i,j}\right]\right)$$

$$W_{\mu,\nu}(y,\mathbf{x}) = \binom{n}{2}^{-1}\sum_{i<j}\left(u_{i,j} - \mathbb{E}\left[u_{i,j}|y_i,\mathbf{x}_i\right] - \mathbb{E}\left[u_{i,j}|y_j,\mathbf{x}_j\right] + \mathbb{E}\left[u_{i,j}\right]\right)$$

and

$$u_{i,j} = \frac{1}{2}\left(\mathbb{1}(y_i \le y_j)a(y_j,y)b(\mathbf{x}_i,\mathbf{x}) + \mathbb{1}(y_j \le y_i)a(y_i,y)b(\mathbf{x}_i,\mathbf{x})\right)$$

Dependence on polynomial orders is suppressed for notational simplicity. Since each of the terms in Equation SA-6.1 are orthogonal, the variance of the estimator can be expressed as the sum of the variance of each of the terms on the right hand side. Furthermore, we note that the first three terms and $W_{\mu,\nu}(y,\mathbf{x})$ have higher-order variance. Thus, we only need to look at the variance of $L_{\mu,\nu}(y,\mathbf{x})$.

$$\mathbb{V}\left(L_{\mu,\nu}(y,\mathbf{x})\right) = \mathbb{V}\left(\frac{2}{n}\sum_{i=1}^{n}\left(\mathbb{E}\left[u_{i,j} \mid y_i,\mathbf{x}_i\right] - \mathbb{E}\left[u_{i,j}\right]\right)\right)$$
$$= \frac{1}{n}\mathbb{V}\left(2\mathbb{E}\left[u_{i,j} \mid y_i,\mathbf{x}_i\right] - 2\mathbb{E}\left[u_{i,j}\right]\right)$$

where we know

$$2\mathbb{E}\left[u_{i,j} \mid y_i,\mathbf{x}_i\right] = \int \mathbb{1}(y_i \le u)a(u,y)\mathrm{d}F_{y_j|\mathbf{x}_i}(u)b(\mathbf{x}_i,\mathbf{x}) + F(y_i|\mathbf{x}_i)a(y_i,y)b(\mathbf{x}_i,\mathbf{x}).$$

We can expand and simplify this to get

$$\mathbb{V}\left(L_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\right) = \mathbb{E}\left[\left(\int \mathbb{1}(y_i \leq u)a(u,y)\mathrm{d}F_{y_j|\mathbf{x}_i}(u)b(\mathbf{x}_i,\mathbf{x}) + F(y_i|\mathbf{x}_i)a(y_i,y)b(\mathbf{x}_i,\mathbf{x})\right)^2\right]$$

$$= \mathbb{E}\left[\iint \mathbb{1}(y_i \leq \min\{u,v\})a(u,y)a(v,y)\mathrm{d}F_{y_j|\mathbf{x}_i}(u)\mathrm{d}F_{y_j|\mathbf{x}_i}(v)b^2(\mathbf{x}_i,\mathbf{x})\right.$$

$$+ \int \mathbb{1}(y_i \leq u)a(u,y)\mathrm{d}F_{y_j|\mathbf{x}_i}(u)F(y_i|\mathbf{x}_i)a(y_i,y)b^2(\mathbf{x}_i,\mathbf{x})$$

$$\left. + (F(y_i|\mathbf{x}_i)a(y_i,y)b(\mathbf{x}_i,\mathbf{x}))^2\right].$$

Note that this expression is identical to the variance expression derived in the proof of SA-2.3. This leads to a natural alternative jackknife covariance estimator,

$$\widehat{\mathsf{C}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') = \frac{1}{n}\sum_{i=1}^{n}\widehat{L}_{(i),\mu,\boldsymbol{\nu}}(y,\mathbf{x})\widehat{L}_{(i),\mu,\boldsymbol{\nu}}(y',\mathbf{x}').$$

where

$$\widehat{L}_{(i),\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{2}{n-1}\sum_{j\neq i}\left(u_{i,j} - \widehat{\theta}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\right).$$

In particular, note that if the two evaluation points are equivalent, we return the variance estimator,

$$\widehat{\mathsf{C}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y,\mathbf{x}) \equiv \widehat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{n-1}\sum_{i=1}^{n}\widehat{L}_{(i),\mu,\boldsymbol{\nu}}^2(y,\mathbf{x})$$

We can similarly construct a jackknife covariance estimator that takes on the form This estimator is implemented in the companion R package as the default variance-covariance matrix estimator.

SA-6.2   Asymptotic variance estimator

An alternative variance estimator is a sample version of the asymptotic variance derived in Lemma SA-2.3. That is, each of the matrices in the formula are replaced with sample-analogs. That is,
(i) $\mu = 0$:

$$\widehat{\mathsf{V}}_{0,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{nh^{d+2|\boldsymbol{\nu}|}}\widehat{\theta}_{0,\mathbf{0}}(y,\mathbf{x})(1 - \widehat{\theta}_{0,\mathbf{0}}(y,\mathbf{x}))\left(\mathbf{e}_{\boldsymbol{\nu}}^{\mathrm{T}}\widehat{\mathbf{S}}_{\mathbf{x}}^{-1}\widehat{\mathbf{T}}_{\mathbf{x}}\widehat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}\right)$$

(ii) $\mu > 0$:

$$\widehat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}}\widehat{\theta}_{1,\mathbf{0}}(y,\mathbf{x})\left(\mathbf{e}_{\mu}^{\mathrm{T}}\widehat{\mathbf{S}}_{y}^{-1}\widehat{\mathbf{T}}_{y}\widehat{\mathbf{S}}_{y}^{-1}\mathbf{e}_{\mu}\right)\left(\mathbf{e}_{\boldsymbol{\nu}}^{\mathrm{T}}\widehat{\mathbf{S}}_{\mathbf{x}}^{-1}\widehat{\mathbf{T}}_{\mathbf{x}}\widehat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}\right).$$

Similarly the covariance can be estimated with the following expression,

$$\widehat{\mathsf{C}}_{\mu,\boldsymbol{\nu}}(y, y'; \mathbf{x}) = \frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} \widehat{\theta}_{1,\mathbf{0}}(y, \mathbf{x}) \left(\mathbf{e}_\mu^{\mathrm{T}} \widehat{\mathbf{S}}_y^{-1} \widehat{\mathbf{T}}_{y,y'} \widehat{\mathbf{S}}_{y'}^{-1} \mathbf{e}_\mu \right) \left(\mathbf{e}_{\boldsymbol{\nu}}^{\mathrm{T}} \widehat{\mathbf{S}}_{\mathbf{x}}^{-1} \widehat{\mathbf{T}}_{\mathbf{x}} \widehat{\mathbf{S}}_{\mathbf{x}}^{-1} \mathbf{e}_{\boldsymbol{\nu}} \right).$$

where

$$\widehat{\mathbf{T}}_{y,y'} = \frac{1}{n^2 h^2} \sum_{i=1}^{n} \sum_{j \neq i} \left(\frac{y_i - y}{h} \wedge \frac{y_j - y'}{h} \right) \mathbf{P}\left(\frac{y_i - y}{h} \right) \mathbf{P}\left(\frac{y_j - y'}{h} \right)^{\mathrm{T}}$$

SA-7   Technical Lemmas and Proofs

SA-7.1   Technical Lemmas

Lemma SA-7.1 (Theorem 1.1 in Rio 1994)

Let $\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_n$ be iid random variables with continuous and strictly positive density on $[0,1]^d$, and $d \geq 2$. Let $\mathcal{G}$ be a class of functions from $[0,1]^d$ to $[-1,1]$, satisfying

$$\sup_P N(\varepsilon, \ \mathcal{G}, \ L^1(P)) \leq \mathfrak{c}_1 \varepsilon^{-\mathfrak{c}_2},$$

where the supremum is taken over all probability measures on $[0,1]^d$, and $\mathfrak{c}_1$ and $\mathfrak{c}_2$ are constants that can depend on $\mathcal{G}$. In addition, assume the following measurability condition holds: there exists a Suslin space $\mathcal{S}$ and a mapping $\mathbb{F} : \mathcal{S} \to \mathcal{G}$, such that $(s, \mathbf{z}) \mapsto \mathbb{F}(s, \mathbf{z})$ is measurable. Let

$$\mathrm{TV}_{\mathcal{G}} = \sup_{g \in \mathcal{G}} \sup_{\phi \in \mathcal{C}_1^\infty([0,1]^d)} \int_{[0,1]^d} g(\mathbf{z}) \mathrm{div}\phi(\mathbf{z}) \mathrm{d}\mathbf{z},$$

where div is the divergence operator, and $\mathcal{C}_1^\infty([0,1]^d)$ is the collection of infinitely differentiable functions with values in $\mathbb{R}^d$, support included in $[0,1]^d$, and supremum norm bounded by 1. Then on a possibly enlarged probability space, there exists a centered Gaussian process, $\mathbb{G}$, indexed by $\mathcal{G}$, such that (i) $\mathbb{C}\mathrm{ov}[\mathbb{G}(g), \mathbb{G}(g')] = \mathbb{C}\mathrm{ov}[g(\mathbf{z}_i), g'(\mathbf{z}_i)]$, and (ii) for any $t \geq \mathfrak{c}_3 \log n$,

$$\mathbb{P}\left[\sqrt{n} \sup_{g \in \mathcal{G}} |\mathbb{B}(g) - \mathbb{G}(g)| \geq \mathfrak{c}_3 \sqrt{n^{\frac{d-1}{d}} \mathrm{TV}_{\mathcal{G}} t} + \mathfrak{c}_3 t \sqrt{\log n} \right] \leq e^{-t}.$$

In the above, $\mathbb{B}$ is the empirical process indexed by $\mathcal{G}$:

$$\mathbb{B}(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(g(\mathbf{z}_i) - \mathbb{E}[g(\mathbf{z}_i)] \right),$$

and $\mathfrak{c}_3$ is some constant that only depends on d, $\mathfrak{c}_1$, and $\mathfrak{c}_2$.

Lemma SA-7.2 (Corollary 5.1 in Chernozhukov, Chetverikov, Kato and Koike 2019)

Let $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^{\ell_n}$ be two mean-zero Gaussian random vectors with covariance matrices $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$,

respectively. Further assume that the diagonal elements in $\boldsymbol{\Omega}_1$ are all one. Then

$$\sup_{\substack{A \subseteq \mathbb{R}^{\ell_n} \\ A \text{ rectangular}}} |\mathbb{P}\left[\mathbf{z}_1 \in A\right] - \mathbb{P}\left[\mathbf{z}_2 \in A\right]| \leq \mathfrak{c}\sqrt{\|\boldsymbol{\Omega}_1 - \boldsymbol{\Omega}_2\|_\infty} \log \ell_n,$$

where $\|\cdot\|_\infty$ denotes the supremum norm, and $\mathfrak{c}$ is an absolute constant.

Lemma SA-7.3 (Equation (3.5) in Giné, Latała and Zinn 2000)

For a degenerate and decoupled second order U-statistic, $\sum_{i,j=1,i\neq j}^n h_{ij}(x_i, \tilde{x}_j)$, the following holds:

$$\mathbb{P}\left[\left|\sum_{i,j,i\neq j}^n u_{ij}(x_i, \tilde{x}_j)\right| > t\right] \leq \mathfrak{c}\exp\left\{-\frac{1}{\mathfrak{c}}\min\left[\frac{t}{D}, \left(\frac{t}{B}\right)^{\frac{2}{3}}, \left(\frac{t}{A}\right)^{\frac{1}{2}}\right]\right\},$$

where $\mathfrak{c}$ is some absolute constant, and A, B and D are any constants satisfying

$$A \geq \max_{1\leq i,j\leq n}\sup_{u,v}|u_{ij}(u,v)|$$

$$B^2 \geq \max_{1\leq i,j\leq n}\left[\sup_v\left|\sum_{i=1}^n \mathbb{E}u_{ij}(x_i,v)^2\right|, \ \sup_u\left|\sum_{j=1}^n \mathbb{E}u_{ij}(u,\tilde{x}_j)^2\right|\right]$$

$$D^2 \geq \sum_{i,j=1,i\neq j}^n \mathbb{E}u_{ij}(x_i,\tilde{x}_j)^2.$$

where $\{x_i, 1 \leq i \leq n\}$ are independent random variables, and $\{\tilde{x}_i, 1 \leq i \leq n\}$ is an independent copy of $\{x_i, 1 \leq i \leq n\}$.

To apply the above lemma, an additional decoupling step is usually needed. Fortunately, the decoupling step only introduces an extra constant, but will not affect the order of the tail probability bound. Formally,

Lemma SA-7.4 (de la Peña and Montgomery-Smith 1995)

Consider the setting of Lemma SA-7.3. Then

$$\mathbb{P}\left[\left|\sum_{i,j,i\neq j}^n u_{ij}(x_i, x_j)\right| > t\right] \leq \mathfrak{c}\cdot\mathbb{P}\left[\mathfrak{c}\left|\sum_{i,j,i\neq j}^n u_{ij}(x_i, \tilde{x}_j)\right| > t\right],$$

where $\mathfrak{c}$ is an absolute constant.

As a result, we will apply Lemma SA-7.3 without explicitly mentioning the decoupling step or the extra constant it introduces.

Lemma SA-7.5 (Theorem 2.1 in Chernozhukov, Chetverikov and Kato 2014)

Let $\mathbb{G}$ be a centered and separable Gaussian process indexed by $g \in \mathcal{G}$ such that $\mathbb{V}[\mathbb{G}(g)] = 1$ for all

$g \in \mathcal{G}$. Assume $\sup_{g \in \mathcal{G}} \mathfrak{G}(g) < \infty$ almost surely. Define $C_{\mathcal{G}} = \mathbb{E}[\sup_{g \in \mathcal{G}} \mathfrak{G}(g)]$. Then for all $\varepsilon > 0$,

$$\sup_{u \in \mathbb{R}} \mathbb{P}\left[\left| \sup_{g \in \mathcal{G}} \mathfrak{G}(g) - u \right| \leq \varepsilon \right] \leq 4\varepsilon(C_{\mathcal{G}} + 1).$$

SA-7.2 Proof of Lemma SA-2.1

Part (i). We will prove the result for $\hat{\mathbf{S}}_{\mathbf{x}}$. The same proof strategy applies to $\hat{\mathbf{c}}_{\mathbf{x},\mathbf{m}}$, $\hat{\mathbf{S}}_y$, $\hat{\mathbf{c}}_{y,\ell}$ and $\hat{\mathbf{T}}_{\mathbf{x}}$.

To start, note that $\mathcal{X}$ is compact, then for any $\eta_n > 0$, one can find $\{\mathbf{x}_\ell : 1 \leq \ell \leq M_n\}$, such that $\mathcal{X} \subseteq \cup_{1 \leq \ell \leq M_n} B_\ell$, where $B_\ell := B(\mathbf{x}_\ell, \eta_n)$ is the Euclidean ball centered at $\mathbf{x}_\ell$ with radius η_n. Define $\mathbf{r} = \sqrt{\log n / (nh^d)}$. Then,

$$\sup_{\mathbf{x} \in \mathcal{X}} \left| \hat{\mathbf{S}}_{\mathbf{x}} - \mathbf{S}_{\mathbf{x}} \right| \leq \max_{1 \leq \ell \leq M_n} \sup_{\mathbf{x} \in B_\ell} \left| \hat{\mathbf{S}}_{\mathbf{x}} - \mathbf{S}_{\mathbf{x}} \right|$$

$$\leq \underbrace{\max_{1 \leq \ell \leq M_n} \left| \hat{\mathbf{S}}_{\mathbf{x}_\ell} - \mathbf{S}_{\mathbf{x}_\ell} \right|}_{(I)} + \underbrace{\sup_{1 \leq \ell \leq M_n} \sup_{\mathbf{x} \in B_\ell} \left| \hat{\mathbf{S}}_{\mathbf{x}} - \hat{\mathbf{S}}_{\mathbf{x}_\ell} \right|}_{(II)} + \underbrace{\sup_{1 \leq \ell \leq M_n} \sup_{\mathbf{x} \in B_\ell} \left| \mathbf{S}_{\mathbf{x}} - \mathbf{S}_{\mathbf{x}_\ell} \right|}_{(III)}.$$

Consider (III) first. We will take $\eta_n \precsim h$, then by the Lipschitz-h^{-1} continuity of $\mathbf{S}_{\mathbf{x}}$, the third term satisfies

$$\text{(III)} \precsim \frac{\eta_n}{h}.$$

Similarly, the random matrix, $\hat{\mathbf{S}}_{\mathbf{x}}$, is the average of Lipschitz-h^{-1-d} continuous functions, which means

$$\text{(II)} \precsim \frac{\eta_n}{h^{1+d}}.$$

Note that the above order is non-probabilistic.

Now consider the first term. By employing the union bound, we have that, for any constant $\mathfrak{c}_1 > 0$,

$$\mathbb{P}\left[(I) > \mathfrak{c}_1 \mathbf{r} \right] \leq M_n \max_{1 \leq \ell \leq M_n} \mathbb{P}\left[\left| \hat{\mathbf{S}}_{\mathbf{x}_\ell} - \mathbf{S}_{\mathbf{x}_\ell} \right| > \mathfrak{c}_1 \mathbf{r} \right].$$

To proceed, we recall the formula of $\hat{\mathbf{S}}_{\mathbf{x}}$:

$$\hat{\mathbf{S}}_{\mathbf{x}} = \frac{1}{nh^d} \sum_{i=1}^{n} \mathfrak{q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h} \right) \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h} \right)^{\mathrm{T}}.$$

It is easy to show that the summands in the above satisfies

$$\mathbb{V}\left[\frac{1}{h^d} \mathfrak{q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h} \right) \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h} \right)^{\mathrm{T}} \right] \leq C' h^{-d},$$

and

$$\left| \frac{1}{h^d} \mathfrak{q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h} \right) \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h} \right)^{\mathrm{T}} \right| \leq C' h^{-d},$$

where C' is a constant that does not depend on n, h or the evaluation point $\mathbf{x}$. Therefore, we apply Bernstein's inequality, which gives

$$M_n \max_{1 \leq \ell \leq M_n} \mathbb{P}\left[\left| \hat{\mathbf{S}}_{\mathbf{x}_\ell} - \mathbf{S}_{\mathbf{x}_\ell} \right| > \mathfrak{c}_1 \mathbf{r} \right] \leq 2 \exp\left\{ -\frac{1}{2} \frac{\mathfrak{c}_1^2 n^2 \mathbf{r}^2}{nC'h^{-d} + \frac{1}{3}\mathfrak{c}_1 C'h^{-d} n \mathbf{r}} + \log M_n \right\}$$

$$= 2 \exp\left\{ -\frac{1}{2} \frac{\mathfrak{c}_1^2 \log n}{C' + \frac{1}{3}\mathfrak{c}_1 C' \mathbf{r}} + \log M_n \right\}.$$

To complete the proof, we note that M_n is at most polynomial in n as long as η_n is. Therefore, one can choose η_n sufficiently small so that both (II) and (III) become negligible relative to (I), and hence for some constants $\mathfrak{c}_1$, $\mathfrak{c}_2$, and

27

$\mathfrak{c}_3$,

$$\mathbb{P}\left[\sup_{\mathbf{x}\in\mathcal{X}}\left|\hat{\mathbf{S}}_{\mathbf{x}} - \mathbf{S}_{\mathbf{x}}\right| > \mathfrak{c}_1\mathbf{r}\right] \leq \mathfrak{c}_2 n^{-\mathfrak{c}_3},$$

and $\mathfrak{c}_3$ can be made arbitrarily large with appropriate choices of $\mathfrak{c}_1$.

Part (ii). Next consider $\mathbf{e}_\mu^{\mathrm{T}}\mathbf{S}_y^{-1}(\bar{\mathbf{R}}_{y,\mathbf{x}} - \mathbb{E}[\bar{\mathbf{R}}_{y,\mathbf{x}}|\mathbf{X}])$, which takes the form

$$\mathbf{e}_\mu^{\mathrm{T}}\mathbf{S}_y^{-1}(\bar{\mathbf{R}}_{y,\mathbf{x}} - \mathbb{E}[\bar{\mathbf{R}}_{y,\mathbf{x}}|\mathbf{X}]) = \frac{1}{nh^{\mu+|\nu|}}\sum_{i=1}^{n}\mathbf{e}_\mu^{\mathrm{T}}\mathbf{S}_y^{-1}\int_{\frac{\mathcal{Y}-y}{h}}\left(\mathbb{1}(y_i \leq y + hu) - F(y + hu|\mathbf{x}_i)\right)\mathbf{P}(u)g(y + hu)\mathrm{d}u\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}}.$$

It is straightforward to see that

$$\left|\mathbf{e}_\mu^{\mathrm{T}}\mathbf{S}_y^{-1}\int_{\frac{\mathcal{Y}-y}{h}}\left(\mathbb{1}(y_i \leq y + hu) - F(y + hu|\mathbf{x}_i)\right)\mathbf{P}(u)g(y + hu)\mathrm{d}u\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}}\right| \leq C'h^{-d}$$

for some C' that holds uniformly for $y \in \mathcal{Y}$ and $\mathbf{x} \in \mathcal{X}$. We also have the following bound on the variance

$$\mathbb{V}\left[\mathbf{e}_\mu^{\mathrm{T}}\mathbf{S}_y^{-1}\int_{\frac{\mathcal{Y}-y}{h}}\left(\mathbb{1}(y_i \leq y + hu) - F(y + hu|\mathbf{x}_i)\right)\mathbf{P}(u)g(y + hu)\mathrm{d}u\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}}\right] \leq C'\begin{cases} h^{-d} & \text{if } \mu = 0 \\ h^{-d+1} & \text{if } \mu > 0. \end{cases}$$

Consider the first case above ($\mu = 0$). By a discretization $\{(y_\ell, \mathbf{x}_\ell) : 1 \leq \ell \leq M_n\}$ of $\mathcal{Y} \times \mathcal{X}$, we have the probabilistic bound due to Bernstein's inequality

$$\mathbb{P}\left[h^{\mu+|\nu|}\max_{1\leq\ell\leq M_n}\left|\mathbf{e}_\mu^{\mathrm{T}}\mathbf{S}_y^{-1}(\bar{\mathbf{R}}_{y,\mathbf{x}} - \mathbb{E}[\bar{\mathbf{R}}_{y,\mathbf{x}}|\mathbf{X}])\right| > \mathfrak{c}_1\mathbf{r}\right] \leq 2\exp\left\{-\frac{1}{2}\frac{\mathfrak{c}_1^2 n^2\mathbf{r}^2}{nC'h^{-d} + \frac{1}{3}\mathfrak{c}_1 C'h^{-d}n\mathbf{r}} + \log M_n\right\}$$

$$= 2\exp\left\{-\frac{1}{2}\frac{\mathfrak{c}_1^2\log n}{C' + \frac{1}{3}\mathfrak{c}_1 C'\mathbf{r}} + \log M_n\right\},$$

provided that we set $\mathbf{r} = \sqrt{\log n/(nh^d)}$. Using arguments similar to those in part (i), it is straightforward to show that M_n is at most polynomial in n, and the error from discretization can be ignored. This concludes the proof for the $\mu = 0$ case.

For $\mu > 0$, we set $\mathbf{r} = \sqrt{\log n/(nh^{d-1})}$, and the probabilistic bound takes the form

$$\mathbb{P}\left[h^{\mu+|\nu|}\max_{1\leq\ell\leq M_n}\left|\mathbf{e}_\mu^{\mathrm{T}}\mathbf{S}_y^{-1}(\bar{\mathbf{R}}_{y,\mathbf{x}} - \mathbb{E}[\bar{\mathbf{R}}_{y,\mathbf{x}}|\mathbf{X}])\right| > \mathfrak{c}_1\mathbf{r}\right] \leq 2\exp\left\{-\frac{1}{2}\frac{\mathfrak{c}_1^2 n^2\mathbf{r}^2}{nC'h^{-d+1} + \frac{1}{3}\mathfrak{c}_1 C'h^{-d}n\mathbf{r}} + \log M_n\right\}$$

$$= 2\exp\left\{-\frac{1}{2}\frac{\mathfrak{c}_1^2\frac{\log n}{h^{d-1}}}{\frac{C'}{h^{d-1}} + \frac{\mathfrak{c}_1 C'}{3h^d}\sqrt{\frac{\log n}{h^{d-1}}}} + \log M_n\right\}$$

$$= 2\exp\left\{-\frac{1}{2}\frac{\mathfrak{c}_1^2\log n}{C' + \frac{\mathfrak{c}_1 C'}{3}\sqrt{\frac{\log n}{nh^{d+1}}}} + \log M_n\right\}.$$

This concludes the proof for the second case, where $\mu > 0$.

SA-7.3 Proof of Lemma SA-2.2

The conditional expectation of $\bar{\mathbf{R}}_{y,\mathbf{x}}$ in $\check{\theta}_{\mu,\nu}$ is

$$\mathbb{E}\left[\frac{1}{nh^{\mu+|\nu|}}\sum_{i=1}^{n}\left[\int_{\mathcal{Y}}\mathbb{1}(y_i \leq u)\frac{1}{h}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\right]\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}}\bigg|\mathbf{X}\right]$$

$$= \frac{1}{nh^{\mu+|\nu|}}\sum_{i=1}^{n}\left[\int_{\mathcal{Y}}F(u|\mathbf{x}_i)\frac{1}{h}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\right]\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}}.$$

28

To proceed, we employ a Taylor expansion of the conditional distribution function to order s:

$$F(u|\mathbf{x}_i) = \sum_{\ell+|\mathbf{m}|\leq s} \theta_{\ell,\mathbf{m}}(y,\mathbf{x})\frac{1}{\ell!\mathbf{m}!}(u-y)^\ell(\mathbf{x}_i-\mathbf{x})^\mathbf{m} + o\left(\sum_{\ell+|\mathbf{m}|=s}|u-y|^\ell|\mathbf{x}_i-\mathbf{x}|^\mathbf{m}\right).$$

Then, the conditional expectation can be simplified as

$$\frac{1}{n}\sum_{i=1}^n\left[\int_{\mathcal{Y}} F(u|\mathbf{x}_i)\frac{1}{h}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\right]\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^\mathrm{T}$$

$$= \sum_{\ell+|\mathbf{m}|\leq s}\theta_{\ell,\mathbf{m}}(y,\mathbf{x})\left[\int_{\mathcal{Y}}\frac{1}{\ell!}(u-y)^\ell\frac{1}{h}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\right]\left[\frac{1}{n}\sum_{i=1}^n\frac{1}{h^d}\frac{1}{\mathbf{m}!}(\mathbf{x}_i-\mathbf{x})^\mathbf{m}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^\mathrm{T}\right]$$

$$+ o\left(\sum_{\ell+|\mathbf{m}|=s}\left[\int_{\mathcal{Y}}\frac{1}{\ell!}|u-y|^\ell\frac{1}{h}\left|\mathbf{P}\left(\frac{u-y}{h}\right)\right|\mathrm{d}G(u)\right]\left[\frac{1}{n}\sum_{i=1}^n\frac{1}{h^d}\frac{1}{\mathbf{m}!}|\mathbf{x}_i-\mathbf{x}|^\mathbf{m}\left|\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\right|\right]\right)$$

$$= \sum_{\ell+|\mathbf{m}|\leq s}h^{\ell+|\mathbf{m}|}\theta_{\ell,\mathbf{m}}(y,\mathbf{x})\mathbf{c}_{y,\ell}\hat{\mathbf{c}}_{\mathbf{x},\mathbf{m}}^\mathrm{T} + o_\mathbb{P}(h^s).$$

We note that

$$\mathbf{S}_y^{-1}\mathbf{c}_{y,\ell} = \mathbf{e}_\ell \text{ for all } 0\leq\ell\leq\mathfrak{p},$$

and

$$\hat{\mathbf{S}}_\mathbf{x}^{-1}\hat{\mathbf{c}}_{\mathbf{x},\mathbf{m}} = \mathbf{e}_\mathbf{m} \text{ for all } 0\leq|\mathbf{m}|\leq\mathfrak{q}.$$

Therefore,

$$\mathbb{E}[\breve{\theta}_{\mu,\boldsymbol{\nu}}|\mathbf{x}_1,\cdots,\mathbf{x}_n] = \theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) + h^{\mathfrak{q}+1-|\boldsymbol{\nu}|}\sum_{|\mathbf{m}|=\mathfrak{q}+1}\theta_{\mu,\mathbf{m}}(y,\mathbf{x})\hat{\mathbf{c}}_{\mathbf{x},\mathbf{m}}^\mathrm{T}\hat{\mathbf{S}}_\mathbf{x}^{-1}\mathbf{e}_{\boldsymbol{\nu}} + h^{\mathfrak{p}+1-\mu}\theta_{\mathfrak{p}+1,\boldsymbol{\nu}}(y,\mathbf{x})\mathbf{c}_{y,\mathfrak{p}+1}^\mathrm{T}\mathbf{S}_y^{-1}\mathbf{e}_\mu$$

$$+ o_\mathbb{P}\left(h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} + h^{\mathfrak{p}+1-\mu}\right).$$

By Lemma SA-2.1, the second term on the right-hand side satisfies

$$h^{\mathfrak{q}+1-|\boldsymbol{\nu}|}\sum_{|\mathbf{m}|=\mathfrak{q}+1}\theta_{\mu,\mathbf{m}}(y,\mathbf{x})\hat{\mathbf{c}}_{\mathbf{x},\mathbf{m}}^\mathrm{T}\hat{\mathbf{S}}_\mathbf{x}^{-1}\mathbf{e}_{\boldsymbol{\nu}} = h^{\mathfrak{q}+1-|\boldsymbol{\nu}|}\sum_{|\mathbf{m}|=\mathfrak{q}+1}\theta_{\mu,\mathbf{m}}(y,\mathbf{x})\mathbf{c}_{\mathbf{x},\mathbf{m}}^\mathrm{T}\mathbf{S}_\mathbf{x}^{-1}\mathbf{e}_{\boldsymbol{\nu}} + O_\mathbb{P}\left(h^{\mathfrak{q}+1-|\boldsymbol{\nu}|}\sqrt{\frac{\log n}{nh^d}}\right),$$

which means we can denote the leading bias as

$$\mathsf{B}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = h^{\mathfrak{q}+1-|\boldsymbol{\nu}|}\sum_{|\mathbf{m}|=\mathfrak{q}+1}\theta_{\mu,\mathbf{m}}(y,\mathbf{x})\mathbf{c}_{\mathbf{x},\mathbf{m}}^\mathrm{T}\mathbf{S}_\mathbf{x}^{-1}\mathbf{e}_{\boldsymbol{\nu}} + h^{\mathfrak{p}+1-\mu}\theta_{\mathfrak{p}+1,\boldsymbol{\nu}}(y,\mathbf{x})\mathbf{c}_{y,\mathfrak{p}+1}^\mathrm{T}\mathbf{S}_y^{-1}\mathbf{e}_\mu.$$

For the second claim of this lemma, we again consider a Taylor expansion

$$F(y_j|\mathbf{x}_i) = \sum_{\ell+|\mathbf{m}|\leq s}\theta_{\ell,\mathbf{m}}(y,\mathbf{x})\frac{1}{\ell!\mathbf{m}!}(y_j-y)^\ell(\mathbf{x}_i-\mathbf{x})^\mathbf{m} + o\left(\sum_{\ell+|\mathbf{m}|=s}|y_j-y|^\ell|\mathbf{x}_i-\mathbf{x}|^\mathbf{m}\right).$$

29

Then

$$\frac{1}{n^2 h^{d+1+\mu+|\boldsymbol{\nu}|}} \sum_{i,j=1}^{n} \mathbf{e}_\mu^{\mathrm{T}} \hat{\mathbf{S}}_y^{-1} \left[F(y_j|\mathbf{x}_i)\mathbf{P}\left(\frac{y_j - y}{h}\right)\right] \mathbf{P}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}} \hat{\mathbf{S}}_\mathbf{x}^{-1} \mathbf{e}_{\boldsymbol{\nu}}$$

$$= \frac{1}{n^2 h^{d+1+\mu+|\boldsymbol{\nu}|}} \sum_{i,j=1}^{n} \mathbf{e}_\mu^{\mathrm{T}} \hat{\mathbf{S}}_y^{-1} \left[\sum_{\ell+|\mathbf{m}|\leq s} \theta_{\ell,\mathbf{m}}(y,\mathbf{x}) \frac{1}{\ell!\mathbf{m}!} (y_j - y)^\ell (\mathbf{x}_i - \mathbf{x})^{\mathbf{m}} \mathbf{P}\left(\frac{y_j - y}{h}\right) \right] \mathbf{P}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}} \hat{\mathbf{S}}_\mathbf{x}^{-1} \mathbf{e}_{\boldsymbol{\nu}}$$

$$+ o\left(\frac{1}{n^2 h^{d+1+\mu+|\boldsymbol{\nu}|}} \mathbf{e}_\mu^{\mathrm{T}} \hat{\mathbf{S}}_y^{-1} \sum_{i,j=1}^{n} \left[\sum_{\ell+|\mathbf{m}|=s} |y_j - y|^\ell |\mathbf{x}_i - \mathbf{x}|^{\mathbf{m}} \left|\mathbf{P}\left(\frac{y_j - y}{h}\right)\right| \left|\mathbf{P}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right| \right] \hat{\mathbf{S}}_\mathbf{x}^{-1} \mathbf{e}_{\boldsymbol{\nu}} \right)$$

$$= \theta_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) + h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} \sum_{|\mathbf{m}|=\mathfrak{q}+1} \theta_{\mu,\mathbf{m}}(y,\mathbf{x}) \mathbf{c}_{\mathbf{x},\mathbf{m}}^{\mathrm{T}} \mathbf{S}_\mathbf{x}^{-1} \mathbf{e}_{\boldsymbol{\nu}} + h^{\mathfrak{p}+1-\mu}\theta_{\mathfrak{p}+1,\boldsymbol{\nu}}(y,\mathbf{x}) \mathbf{c}_{y,\mathfrak{p}+1}^{\mathrm{T}} \mathbf{S}_y^{-1} \mathbf{e}_\mu + o_{\mathbb{P}}(h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} + h^{\mathfrak{p}+1-\mu}).$$

SA-7.4   Proof of Lemma SA-2.3

Let $\mathbf{c}_1 = \mathbf{S}_y^{-1}\mathbf{e}_\mu$ and $\mathbf{c}_2 = \mathbf{S}_\mathbf{x}^{-1}\mathbf{e}_{\boldsymbol{\nu}}$.

$$\mathbb{V}\left[\int_{\mathcal{Y}} \left[\mathbb{1}(y_i \leq u) - F(u|\mathbf{x}_i) \right] \mathbf{c}_1^{\mathrm{T}} \frac{1}{h}\mathbf{P}\left(\frac{u - y}{h}\right) \mathrm{d}G(u) \frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}} \mathbf{c}_2 \right]$$

$$= \mathbb{E}\left[\mathbb{V}\left[\int_{\mathcal{Y}} \left[\mathbb{1}(y_i \leq u) - F(u|\mathbf{x}_i) \right] \mathbf{c}_1^{\mathrm{T}} \frac{1}{h}\mathbf{P}\left(\frac{u - y}{h}\right) \mathrm{d}G(u) \frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}} \mathbf{c}_2 \,\middle|\, \mathbf{X} \right] \right]$$

$$= \mathbb{E}\left[\iint_{\frac{\mathcal{Y}-y}{h}} \left(F(y + h(u_1 \wedge u_2)|\mathbf{x}_i) - F(y + hu_1|\mathbf{x}_i)F(y + hu_2|\mathbf{x}_i) \right) \mathbf{c}_1^{\mathrm{T}}\mathbf{P}(u_1)\,\mathbf{c}_1^{\mathrm{T}}\mathbf{P}(u_2)\,g(y + hu_1)g(y + hu_2)\mathrm{d}u_1\mathrm{d}u_2 \right.$$

$$\left. \left(\mathbf{c}_2^{\mathrm{T}} \frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right)^2 \right] \tag{*}$$

We make a further expansion:

$$F(y + h(u_1 \wedge u_2)|\mathbf{x}_i) - F(y + hu_1|\mathbf{x}_i)F(y + hu_2|\mathbf{x}_i)$$
$$= F(y|\mathbf{x}_i)(1 - F(y|\mathbf{x}_i)) + h(u_1 \wedge u_2)f(y|\mathbf{x}_i) - h(u_1 + u_2)f(y|\mathbf{x}_i)F(y|\mathbf{x}_i) + O(h^2).$$

Note that the remainder term, $O(h^2)$, holds uniformly for $y \in \mathcal{Y}$ and $\mathbf{x}_i \in \mathcal{X}$ since the conditional distribution function is assumed to have bounded second derivative.

Therefore,

$$(*) = \left(\mathbf{c}_1^{\mathrm{T}}\mathbf{c}_{y,0}\mathbf{c}_{y,0}^{\mathrm{T}}\mathbf{c}_1 \right) \mathbb{E}\left[F(y|\mathbf{x}_i)(1 - F(y|\mathbf{x}_i)) \left(\mathbf{c}_2^{\mathrm{T}} \frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right)^2 \right] + h\left(\mathbf{c}_1^{\mathrm{T}}\mathbf{T}_y\mathbf{c}_1 \right) \mathbb{E}\left[f(y|\mathbf{x}_i) \left(\mathbf{c}_2^{\mathrm{T}} \frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right)^2 \right]$$

$$- h\left[\mathbf{c}_1^{\mathrm{T}}\left(\mathbf{c}_{y,1}\mathbf{c}_{y,0}^{\mathrm{T}} + \mathbf{c}_{y,0}\mathbf{c}_{y,1}^{\mathrm{T}} \right)\mathbf{c}_1 \right] \mathbb{E}\left(f(y|\mathbf{x}_i)F(y|\mathbf{x}_i) \left[\mathbf{c}_2^{\mathrm{T}} \frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2 \right) + O\left(\frac{1}{h^{d-2}} \right)$$

$$= \mathbf{e}_\mu^{\mathrm{T}}\mathbf{e}_0 \mathbb{E}\left[F(y|\mathbf{x}_i)(1 - F(y|\mathbf{x}_i)) \left(\mathbf{c}_2^{\mathrm{T}} \frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right)^2 \right] + h\left(\mathbf{e}_\mu^{\mathrm{T}}\mathbf{S}_y^{-1}\mathbf{T}_y\mathbf{S}_y^{-1}\mathbf{e}_\mu \right) \mathbb{E}\left[f(y|\mathbf{x}_i) \left(\mathbf{c}_2^{\mathrm{T}} \frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right)^2 \right] + O\left(\frac{1}{h^{d-2}} \right).$$

To conclude the proof, we note that two scenarios can arise: $\mu = 0$ and $\mu > 0$. In the second case,

$$(*) = \frac{1}{h^{d-1}}\theta_{1,\mathbf{0}}\left(\mathbf{e}_\mu^{\mathrm{T}}\mathbf{S}_y^{-1}\mathbf{T}_y\mathbf{S}_y^{-1}\mathbf{e}_\mu \right)\left(\mathbf{e}_{\boldsymbol{\nu}}^{\mathrm{T}}\mathbf{S}_\mathbf{x}^{-1}\mathbf{T}_\mathbf{x}\mathbf{S}_\mathbf{x}^{-1}\mathbf{e}_{\boldsymbol{\nu}} \right) + O\left(\frac{1}{h^{d-2}} \right).$$

The first case is more involved. If $\theta_{0,\mathbf{0}} \neq 0, 1$, then

$$(*) = \frac{1}{h^d}\theta_{0,\mathbf{0}}(1 - \theta_{0,\mathbf{0}})\left(\mathbf{e}_{\boldsymbol{\nu}}^{\mathrm{T}}\mathbf{S}_\mathbf{x}^{-1}\mathbf{T}_\mathbf{x}\mathbf{S}_\mathbf{x}^{-1}\mathbf{e}_{\boldsymbol{\nu}} \right) + O_{\mathbb{P}}\left(\frac{1}{h^{d-1}} \right).$$

If $\theta_{0,\mathbf{0}} = 0$ or 1, then a further expansion is needed, which is why an extra h will be present in the leading variance.

30

SA-7.5 Proof of Lemma SA-2.4

SA-7.5.1 Consistency of $\check{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$

For the purposes of this proof, let $\mathbf{c}_1 = \mathbf{S}_y^{-1}\mathbf{e}_\mu$, $\hat{\mathbf{c}}_2 = \hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}$, and $\mathbf{c}_2 = \mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}$. To start, consider

$$\frac{1}{n^2 h^{2d+2\mu+2|\boldsymbol{\nu}|}} \sum_{i=1}^{n} \left[\int_{\mathcal{Y}} \left(\mathbb{1}\left(y_i \leq u\right) - \hat{F}(u|\mathbf{x}_i)\right) \frac{1}{h}\mathbf{c}_1^{\mathrm{T}}\mathbf{P}\left(\frac{u-y}{h}\right) \mathrm{d}G(u) \mathbf{c}_2^{\mathrm{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right]^2$$

$$= \frac{1}{n^2 h^{2d+2\mu+2|\boldsymbol{\nu}|}} \sum_{i=1}^{n} \left[\int_{\mathcal{Y}} \left(\mathbb{1}\left(y_i \leq u\right) - F(u|\mathbf{x}_i)\right) \frac{1}{h}\mathbf{c}_1^{\mathrm{T}}\mathbf{P}\left(\frac{u-y}{h}\right) \mathrm{d}G(u) \mathbf{c}_2^{\mathrm{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right]^2 \quad \text{(I)}$$

$$- \frac{2}{n^2 h^{2d+2\mu+2|\boldsymbol{\nu}|}} \sum_{i=1}^{n} \iint_{\frac{\mathcal{Y}-y}{h}} \left(\mathbb{1}\left(y_i \leq u_1\right) - F(u_1|\mathbf{x}_i)\right)\left(\hat{F}(u_2|\mathbf{x}_i) - F(u_2|\mathbf{x}_i)\right)\mathbf{c}_1^{\mathrm{T}}\mathbf{P}\left(u_1\right)\mathbf{c}_1^{\mathrm{T}}\mathbf{P}\left(u_2\right)$$

$$g(y+hu_1)g(y+hu_2)\mathrm{d}u_1\mathrm{d}u_2 \left[\mathbf{c}_2^{\mathrm{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right]^2 \quad \text{(II)}$$

$$+ \frac{1}{n^2 h^{2d+2\mu+2|\boldsymbol{\nu}|}} \sum_{i=1}^{n} \left[\int_{\mathcal{Y}} \left(\hat{F}(u|\mathbf{x}_i) - F(u|\mathbf{x}_i)\right) \frac{1}{h}\mathbf{c}_1^{\mathrm{T}}\mathbf{P}\left(\frac{u-y}{h}\right) \mathrm{d}G(u) \mathbf{c}_2^{\mathrm{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right]^2. \quad \text{(III)}$$

First consider term (III). With the uniform convergence result for the estimated conditional distribution function, it is clear that

$$|\text{(III)}| \lesssim_{\mathbb{P}} \frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}} \left(h^{2\mathfrak{q}+2} + \frac{\log n}{nh^d}\right) \lesssim_{\mathbb{P}} \begin{cases} \mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x})\left(h^{2\mathfrak{q}+2} + \frac{\log n}{nh^d}\right) & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{o}} \neq 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\left(h^{2\mathfrak{q}+1} + \frac{\log n}{nh^{d+1}}\right) & \text{if } \mu > 0, \text{ or } \theta_{0,\mathbf{o}} = 0 \text{ or } 1 \end{cases}.$$

Now we study term (I), which is clearly unbiased for $\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$. Therefore, we compute its variance.

$$\mathbb{V}\left[\text{(I)}\right] = \frac{1}{n^3 h^{4d+4\mu+4|\boldsymbol{\nu}|}}\mathbb{V}\left[\left(\int_{\mathcal{Y}} \left(\mathbb{1}\left(y_i \leq u\right) - F(u|\mathbf{x}_i)\right) \frac{1}{h}\mathbf{c}_1^{\mathrm{T}}\mathbf{P}\left(\frac{u-y}{h}\right) \mathrm{d}G(u) \mathbf{c}_2^{\mathrm{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right)^2\right]$$

$$\leq \frac{1}{n^3 h^{4d+4\mu+4|\boldsymbol{\nu}|}}\mathbb{E}\left[\left(\int_{\mathcal{Y}} \left(\mathbb{1}\left(y_i \leq u\right) - F(u|\mathbf{x}_i)\right) \frac{1}{h}\mathbf{c}_1^{\mathrm{T}}\mathbf{P}\left(\frac{u-y}{h}\right) \mathrm{d}G(u) \mathbf{c}_2^{\mathrm{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right)^4\right]$$

$$= \frac{1}{n^3 h^{4d+4\mu+4|\boldsymbol{\nu}|}}\mathbb{E}\left[\prod_{j=1}^{4}\left[\int_{\mathcal{Y}} \left(\mathbb{1}\left(y_i \leq u_j\right) - F(u_j|\mathbf{x}_i)\right) \frac{1}{h}\mathbf{c}_1^{\mathrm{T}}\mathbf{P}\left(\frac{u_j-y}{h}\right) \mathrm{d}G(u_j)\right]\left[\mathbf{c}_2^{\mathrm{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right]^4\right].$$

With iterative expectation (by conditioning on $\mathbf{x}_i$), the above further reduces to

$$\mathbb{V}\left[\text{(I)}\right] = \frac{1}{n^3 h^{3d+4\mu+4|\boldsymbol{\nu}|}}\theta_{0,\mathbf{o}}(1-\theta_{0,\mathbf{o}})(1-3\theta_{0,\mathbf{o}}(1-\theta_{0,\mathbf{o}}))\left[\mathbf{c}_1^{\mathrm{T}}\mathbf{c}_{y,0}\right]^4\mathbb{E}\left[\frac{1}{h^d}\left[\mathbf{c}_2^{\mathrm{T}}\mathbf{R}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right]^4\right] + O\left(\frac{h}{n^3 h^{4d+4\mu+4|\boldsymbol{\nu}|}}\right).$$

In other words,

$$\left|\text{(I)} - \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\right| \lesssim_{\mathbb{P}} \begin{cases} \sqrt{\frac{1}{n^3 h^{3d+4|\boldsymbol{\nu}|}}} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{o}} \neq 0 \text{ or } 1 \\ \sqrt{\frac{h}{n^3 h^{4d+4\mu+4|\boldsymbol{\nu}|}}} & \text{if } \mu > 0, \text{ or } \theta_{0,\mathbf{o}} = 0 \text{ or } 1 \end{cases} \lesssim \begin{cases} \mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x})\sqrt{\frac{1}{nh^d}} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{o}} \neq 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\sqrt{\frac{1}{nh^{d+1}}} & \text{if } \mu > 0, \text{ or } \theta_{0,\mathbf{o}} = 0 \text{ or } 1 \end{cases}.$$

Finally, we consider (II). Using the Cauchy-Schwartz inequality, we have

$$|\text{(II)}|^2 \leq |\text{(I)}| \cdot |\text{(III)}|.$$

As a result,

$$|\text{(II)}| \lesssim_{\mathbb{P}} \begin{cases} \mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x})\sqrt{h^{2\mathfrak{q}+2} + \frac{\log n}{nh^d}} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{o}} \neq 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\sqrt{h^{2\mathfrak{q}+1} + \frac{\log n}{nh^{d+1}}} & \text{if } \mu > 0, \text{ or } \theta_{0,\mathbf{o}} = 0 \text{ or } 1 \end{cases}.$$

To conclude the proof for $\check{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$, we note that replacing $\hat{\mathbf{c}}_2$ by $\mathbf{c}_2$ only leads to an additional multiplicative factor

$1 + O_{\mathbb{P}}(1/\sqrt{nh^d})$. See Lemma SA-2.1.

SA-7.5.2 Consistency of $\hat{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})$

For the purposes of this proof, let $\hat{\mathbf{c}}_1 = \hat{\mathbf{S}}_y^{-1}\mathbf{e}_\mu$, $\mathbf{c}_1 = \mathbf{S}_y^{-1}\mathbf{e}_\mu$, $\hat{\mathbf{c}}_2 = \hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_\nu$, and $\mathbf{c}_2 = \mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_\nu$. We first consider the following

$$\frac{1}{n^2 h^{2d+2\mu+2|\nu|}} \sum_{i=1}^n \left[\int_{\mathcal{Y}} \left(\mathbb{1}\,(y_i \le u) - \hat{F}(u|\mathbf{x}_i) \right) \frac{1}{h} \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{u-y}{h}\right) \mathrm{d}\hat{F}_y(u) \mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2$$

$$= \frac{1}{n^2 h^{2d+2\mu+2|\nu|}} \sum_{i=1}^n \left[\frac{1}{n^2} \sum_{j,k=1}^n \left(\mathbb{1}\,(y_i \le y_j) - F(y_j|\mathbf{x}_i)\right)\left(\mathbb{1}\,(y_i \le y_k) - F(y_k|\mathbf{x}_i)\right) \frac{1}{h^2} \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j - y}{h}\right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_k - y}{h}\right) \right] \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2 \tag{I}$$

$$- \frac{2}{n^2 h^{2d+2\mu+2|\nu|}} \sum_{i=1}^n \left[\frac{1}{n^2} \sum_{j,k=1}^n \left(\hat{F}(y_j|\mathbf{x}_i) - F(y_j|\mathbf{x}_i)\right)\left(\mathbb{1}\,(y_i \le y_k) - F(y_k|\mathbf{x}_i)\right) \frac{1}{h^2} \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j - y}{h}\right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_k - y}{h}\right) \right] \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2 \tag{II}$$

$$+ \frac{1}{n^2 h^{2d+2\mu+2|\nu|}} \sum_{i=1}^n \left[\frac{1}{n^2} \sum_{j,k=1}^n \left(\hat{F}(y_j|\mathbf{x}_i) - F(y_j|\mathbf{x}_i)\right)\left(\hat{F}(y_k|\mathbf{x}_i) - F(y_k|\mathbf{x}_i)\right) \frac{1}{h^2} \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j - y}{h}\right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_k - y}{h}\right) \right] \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2. \tag{III}$$

By the uniform convergence rate of the estimated conditional distribution function, we have

$$|(\mathrm{III})| \precsim_{\mathbb{P}} \frac{1}{nh^{d+2\mu+2|\nu|}} \left(h^{2\mathfrak{q}+2} + \frac{\log n}{nh^d} \right) \precsim_{\mathbb{P}} \begin{cases} \mathsf{V}_{0,\nu}(y,\mathbf{x}) \left(h^{2\mathfrak{q}+2} + \frac{\log n}{nh^d} \right) & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{0}} \ne 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\nu}(y,\mathbf{x}) \left(h^{2\mathfrak{q}+1} + \frac{\log n}{nh^{d+1}} \right) & \text{if } \mu > 0, \text{ or } \theta_{0,\mathbf{0}} = 0 \text{ or } 1 \end{cases}.$$

Next we consider (II). Using the Cauchy-Schwartz inequality, we have

$$|(\mathrm{II})|^2 \le |(\mathrm{I})| \cdot |(\mathrm{III})|.$$

Finally, consider term (I), which has the expansion

$$(\mathrm{I}) = \frac{1}{n^4 h^{2d+2\mu+2|\nu|+2}} \sum_{\substack{i,j,k=1 \\ \text{distinct}}}^n \left[\left(\mathbb{1}\,(y_i \le y_j) - F(y_j|\mathbf{x}_i)\right)\left(\mathbb{1}\,(y_i \le y_k) - F(y_k|\mathbf{x}_i)\right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j - y}{h}\right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_k - y}{h}\right) \right] \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2 \tag{I.1}$$

$$+ \frac{2}{n^4 h^{2d+2\mu+2|\nu|+2}} \sum_{\substack{i,j=1 \\ \text{distinct}}}^n \left[\left(\mathbb{1}\,(y_i \le y_j) - F(y_j|\mathbf{x}_i)\right)\left(1 - F(y_i|\mathbf{x}_i)\right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j - y}{h}\right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_i - y}{h}\right) \right] \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2 \tag{I.2}$$

$$+ \frac{1}{n^4 h^{2d+2\mu+2|\nu|+2}} \sum_{\substack{i,j=1 \\ \text{distinct}}}^n \left(\mathbb{1}\,(y_i \le y_j) - F(y_j|\mathbf{x}_i)\right)^2 \left[\mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j - y}{h}\right) \right]^2 \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2 \tag{I.3}$$

$$+ \frac{1}{n^4 h^{2d+2\mu+2|\nu|+2}} \sum_{i=1}^n \left(1 - F(y_i|\mathbf{x}_i)\right)^2 \left[\mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_i - y}{h}\right) \right]^2 \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2. \tag{I.4}$$

32

Then,

$$|(\mathrm{I.2})| \leq \frac{2}{n^4 h^{2d+2\mu+2|\boldsymbol{\nu}|+2}} \sum_{\substack{i,j=1 \\ \text{distinct}}}^{n} \left| \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j - y}{h}\right) \right| \cdot \left| \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_i - y}{h}\right) \right| \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2$$

$$\leq \frac{2}{n^2 h^{d+2\mu+2|\boldsymbol{\nu}|}} \left[\frac{1}{nh} \sum_{i=1}^{n} \left| \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_i - y}{h}\right) \right| \right] \left[\frac{1}{nh^{d+1}} \sum_{i=1}^{n} \left| \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_i - y}{h}\right) \right| \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2 \right]$$

$$\precsim_{\mathbb{P}} \frac{1}{n^2 h^{d+2\mu+2|\boldsymbol{\nu}|}} \left(1 + \sqrt{\frac{1}{nh}}\right)\left(1 + \sqrt{\frac{1}{nh^{d+1}}}\right) \precsim \frac{1}{n^2 h^{d+2\mu+2|\boldsymbol{\nu}|}} \precsim \begin{cases} \mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x})\frac{1}{n} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{o}} \neq 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\frac{1}{nh} & \text{if } \mu > 0, \text{ or } \theta_{0,\mathbf{o}} = 0 \text{ or } 1 \end{cases}.$$

Using similar techniques, one can show that

$$|(\mathrm{I.3})| \precsim_{\mathbb{P}} \begin{cases} \mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x})\frac{1}{nh} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{o}} \neq 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\frac{1}{nh^2} & \text{if } \mu > 0, \text{ or } \theta_{0,\mathbf{o}} = 0 \text{ or } 1 \end{cases} \qquad \text{and} \qquad |(\mathrm{I.4})| \precsim_{\mathbb{P}} \begin{cases} \mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x})\frac{1}{n^2 h} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{o}} \neq 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\frac{1}{n^2 h^2} & \text{if } \mu > 0, \text{ or } \theta_{0,\mathbf{o}} = 0 \text{ or } 1 \end{cases}.$$

To streamline the remaining derivation, define

$$\phi_{j,i} = \frac{1}{h}\left(\mathbb{1}\left(y_i \leq y_j\right) - F(y_j | \mathbf{x}_i) \right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j - y}{h}\right), \quad \phi_i = \mathbb{E}[\phi_{j,i} | y_i, \mathbf{x}_i], \quad \psi_i = \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2.$$

Then

$$(\mathrm{I.1}) = \frac{1}{n^4 h^{2d+2\mu+2|\boldsymbol{\nu}|}} \sum_{\substack{i,j,k=1 \\ \text{distinct}}}^{n} \phi_{j,i}\phi_{k,i}\psi_i$$

$$= \underbrace{\frac{1}{n^4 h^{2d+2\mu+2|\boldsymbol{\nu}|}} \sum_{\substack{i,j,k=1 \\ \text{distinct}}}^{n} (\phi_{j,i} - \phi_i)(\phi_{k,i} - \phi_i)\psi_i}_{(\mathrm{I.1.1})} + \left(2 + O\left(\frac{1}{n}\right)\right) \underbrace{\frac{1}{n^3 h^{2d+2\mu+2|\boldsymbol{\nu}|}} \sum_{\substack{i,j=1 \\ \text{distinct}}}^{n} (\phi_{j,i} - \phi_i)\phi_i \psi_i}_{(\mathrm{I.1.2})}$$

$$+ \left(1 + O\left(\frac{1}{n}\right)\right) \underbrace{\frac{1}{n^2 h^{2d+2\mu+2|\boldsymbol{\nu}|}} \sum_{i=1}^{n} \phi_i^2 \psi_i}_{(\mathrm{I.1.3})}.$$

We have studied the term (I.1.3) in the proof for $\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$. In particular,

$$\left| (\mathrm{I.1.3}) - \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \right| \precsim_{\mathbb{P}} \begin{cases} \mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x})\sqrt{\frac{1}{nh^d}} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{o}} \neq 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\sqrt{\frac{1}{nh^{d+1}}} & \text{if } \mu > 0, \text{ or } \theta_{0,\mathbf{o}} = 0 \text{ or } 1 \end{cases}.$$

Term (I.1.1) is a mean zero third order U-statistic. Consider its variance

$$\mathbb{V}\left[(\mathrm{I.1.1}) \right] = \left(\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}} \right)^2 \mathbb{E}\left[\frac{1}{n^6 h^{2d}} \sum_{\substack{i,j,k=1 \\ \text{distinct}}}^{n} \sum_{\substack{i',j',k'=1 \\ \text{distinct}}}^{n} (\phi_{j,i} - \phi_i)(\phi_{k,i} - \phi_i)(\phi_{j',i'} - \phi_{i'})(\phi_{k',i'} - \phi_{i'})\psi_i \psi_{i'} \right].$$

The above expectation is non-zero only in three scenarios: $(j = j', k = k', i \neq i')$, $(j = j', k = k', i = i')$ or $(j = i', k = k', i = j')$. Therefore,

$$|(\mathrm{I.1.1})| \precsim_{\mathbb{P}} \left(\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}} \right)\left(\frac{1}{nh} + \sqrt{\frac{1}{n^3 h^{d+2}}} \right) \precsim \left(\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}} \right)\frac{1}{nh}$$

$$\precsim \begin{cases} \mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x})\frac{1}{nh} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{o}} \neq 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\frac{1}{nh^2} & \text{if } \mu > 0, \text{ or } \theta_{0,\mathbf{o}} = 0 \text{ or } 1 \end{cases}.$$

33

Finally consider (I.1.2), which has a mean of zero. Its variance is

$$
\mathbb{V}\left[(\text{I.1.2})\right] = \left(\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}}\right)^2 \mathbb{E}\left[\frac{1}{n^4 h^{2d}} \sum_{\substack{i,j=1 \\ \text{distinct}}}^{n} \sum_{\substack{i',j'=1 \\ \text{distinct}}}^{n} (\phi_{j,i} - \phi_i)(\phi_{j',i'} - \phi_{i'})\phi_i\psi_i\phi_{i'}\psi_{i'}\right]
$$

$$
= \left(\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}}\right)^2 \mathbb{E}\left[\frac{1}{n^4 h^{2d}} \sum_{\substack{i,i',j=1 \\ \text{distinct}}}^{n} (\phi_{j,i} - \phi_i)(\phi_{j,i'} - \phi_{i'})\phi_i\psi_i\phi_{i'}\psi_{i'}\right]
$$

$$
+ \left(\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}}\right)^2 \mathbb{E}\left[\frac{1}{n^4 h^{2d}} \sum_{\substack{i,j=1 \\ \text{distinct}}}^{n} (\phi_{j,i} - \phi_i)^2 \phi_i^2 \psi_i^2\right] + \mathbb{E}\left[\frac{1}{n^4 h^{2d}} \sum_{\substack{i,j=1 \\ \text{distinct}}}^{n} (\phi_{j,i} - \phi_i)(\phi_{i,j} - \phi_j)\phi_i\psi_i\phi_j\psi_j\right]
$$

$$
\precsim \left(\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}}\right)^2 \left(\frac{1}{nh} + \frac{1}{n^2 h^{d+1}}\right).
$$

In addition, an extra h factor emerges if $\mu > 0$, or if $\theta_{0,\mathbf{o}} = 0$ or 1. As a result,

$$
|(\text{I.1.2})| \precsim_{\mathbb{P}} \begin{cases} \mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x})\sqrt{\frac{1}{nh}} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{o}} \neq 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\sqrt{\frac{1}{nh^2}} & \text{if } \mu > 0, \quad \text{or } \theta_{0,\mathbf{o}} = 0 \text{ or } 1 \end{cases}.
$$

To conclude the proof for $\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$, we note that replacing $\hat{\mathbf{c}}_1$ by $\mathbf{c}_1$ and $\hat{\mathbf{c}}_2$ by $\mathbf{c}_2$ only leads to an additional multiplicative factor $1 + O_{\mathbb{P}}(1/\sqrt{nh^d})$. See Lemma SA-2.1.

SA-7.6 Proof of Theorem SA-2.1

We will write

$$
\bar{\$}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{h^{d+\mu+|\boldsymbol{\nu}|}\mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(y_i,\mathbf{x}_i;y,\mathbf{x})}{\sqrt{\mathbb{V}\left[h^{d+\mu+|\boldsymbol{\nu}|}\mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(y_i,\mathbf{x}_i;y,\mathbf{x})\right]}}.
$$

Define $\mathbf{c}_1 = \mathbf{S}_y^{-1}\mathbf{e}_\mu$ and $\mathbf{c}_2 = \mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}$.

To apply the Berry-Esseen theorem, we first compute the third moment

$$
\mathbb{E}\left[\left|h^{d+\mu+|\boldsymbol{\nu}|}\mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(y_i,\mathbf{x}_i;y,\mathbf{x})\right|^3\right]
$$

$$
= \mathbb{E}\left[\left(\prod_{j=1}^{3}\left|\int_{\frac{\mathcal{Y}-y}{h}}\left(\mathbb{1}(y_i \leq y + hu_j) - F(y + hu_j|\mathbf{x}_i)\right)\mathbf{c}_1^{\mathrm{T}}\mathbf{P}(u_j)\,\mathrm{d}G(u_j)\right|\right)\left|\mathbf{c}_2^{\mathrm{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right|^3\right].
$$

The leading term in the above is simply

$$
\mathbb{E}\left[\left(\prod_{j=1}^{3}\left|\int_{\frac{\mathcal{Y}-y}{h}}\left(\mathbb{1}(y_i \leq y) - F(y|\mathbf{x}_i)\right)\mathbf{c}_1^{\mathrm{T}}\mathbf{P}(u_j)\,\mathrm{d}G(u_j)\right|\right)\left|\mathbf{c}_2^{\mathrm{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right|^3\right]
$$

$$
= \left|\mathbf{c}_1^{\mathrm{T}}\mathbf{c}_{y,0}\right|^3 \mathbb{E}\left[\left|\left(\mathbb{1}(y_i \leq y) - F(y|\mathbf{x}_i)\right)\right|^3\left|\mathbf{c}_2^{\mathrm{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right|^3\right]
$$

$$
= \left|\mathbf{c}_1^{\mathrm{T}}\mathbf{c}_{y,0}\right|^3 \mathbb{E}\left[\left(\theta_{0,\mathbf{o}}(y,\mathbf{x})(1 - \theta_{0,\mathbf{o}}(y,\mathbf{x}))(2\theta_{0,\mathbf{o}}(y,\mathbf{x})^2 - 2\theta_{0,\mathbf{o}}(y,\mathbf{x}) + 1)\right)\left|\mathbf{c}_2^{\mathrm{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right|^3\right] = O(h^d).
$$

Note that the above will be exactly zero in cases (ii) and (iii) of Lemma SA-2.3.

SA-7.7 Omitted Details of Remark SA-2.3

SA-7.7.1 Approximation and coverage error of $\check{\$}_{\mu,\nu}^{\circ}(y,\mathbf{x})$

To start,

$$
\check{\$}_{\mu,\nu}^{\circ}(y,\mathbf{x}) - \bar{\$}_{\mu,\nu}(y,\mathbf{x})
$$

$$
= \frac{1}{nh^{\mu+|\nu|}\sqrt{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}} \sum_{i=1}^{n} \mathbf{e}_{\mu}^{\mathrm{T}} \mathbf{S}_{y}^{-1} \left[\int_{\mathcal{Y}} \left[\mathbb{1}(y_i \le u) - F(u|\mathbf{x}_i) \right] \frac{1}{h} \mathbf{P}\Big(\frac{u-y}{h}\Big) \mathrm{d}G(u) \right] \frac{1}{h^d} \mathbf{Q}\Big(\frac{\mathbf{x}_i - \mathbf{x}}{h}\Big)^{\mathrm{T}} \left(\hat{\mathbf{S}}_{\mathbf{x}}^{-1} - \mathbf{S}_{\mathbf{x}}^{-1} \right) \mathbf{e}_{\nu}.
$$

By allowing the constant $\mathfrak{c}_1$ to take possibly different values in each term, we have

$$
\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left| \check{\$}_{\mu,\nu}^{\circ}(y,\mathbf{x}) - \bar{\$}_{\mu,\nu}(y,\mathbf{x}) \right| \ge \mathfrak{c}_1 \frac{\log n}{\sqrt{nh^d}} \right]
$$

$$
\le \mathbb{P}\left[\sup_{\mathbf{x}\in\mathcal{X}} \left| \hat{\mathbf{S}}_{\mathbf{x}} - \mathbf{S}_{\mathbf{x}} \right| > \mathfrak{c}_1 \sqrt{\frac{\log n}{nh^d}} \right] + \mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left| \frac{\mathbf{e}_{\mu}^{\mathrm{T}} \mathbf{S}_{y}^{-1} \left(\bar{\mathbf{R}}_{y,\mathbf{x}} - \mathbb{E}\left[\bar{\mathbf{R}}_{y,\mathbf{x}}|\mathbf{X} \right] \right)}{\sqrt{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}} \right| > \mathfrak{c}_1 \sqrt{\log n} \right] \le \mathfrak{c}_2 n^{-\mathfrak{c}_3},
$$

where the conclusions follow from the uniform rates established in Lemma SA-2.1 and the variance calculations in Lemma SA-2.3. Next, we consider the normal approximation error. Note that

$$
\mathbb{P}\left[\bar{\$}_{\mu,\nu}(y,\mathbf{x}) \le u - \mathfrak{c}_1 \frac{\log n}{\sqrt{nh^d}} \right] - \mathfrak{c}_2 n^{-\mathfrak{c}_3} \le \mathbb{P}\left[\check{\$}_{\mu,\nu}^{\circ}(y,\mathbf{x}) \le u \right] \le \mathbb{P}\left[\bar{\$}_{\mu,\nu}(y,\mathbf{x}) \le u + \mathfrak{c}_1 \frac{\log n}{\sqrt{nh^d}} \right] + \mathfrak{c}_2 n^{-\mathfrak{c}_3},
$$

which means

$$
\sup_{u\in\mathbb{R}} \left| \mathbb{P}\left[\check{\$}_{\mu,\nu}^{\circ}(y,\mathbf{x}) \le u \right] - \Phi(u) \right| \precsim \frac{\log n}{\sqrt{nh^d}} + \mathbf{r}_{\mathrm{BE}},
$$

where $\mathbf{r}_{\mathrm{BE}}$ is defined in Theorem SA-2.1.

SA-7.7.2 Approximation and coverage error of $\hat{\$}_{\mu,\nu}^{\circ}(y,\mathbf{x})$

To begin with, we decompose the double sum into

$$
\frac{1}{n^2 h^{d+1}} \sum_{i,j=1}^{n} \left[\mathbb{1}(y_i \le y_j) - F(y_j|\mathbf{x}_i) \right] \mathbf{P}\Big(\frac{y_j - y}{h}\Big) \mathbf{Q}\Big(\frac{\mathbf{x}_i - \mathbf{x}}{h}\Big)^{\mathrm{T}}
$$

$$
= \frac{1}{n^2 h^{d+1}} \sum_{i=1}^{n} \left(\left[1 - F(y_i|\mathbf{x}_i) \right] \mathbf{P}\Big(\frac{y_i - y}{h}\Big) - \int_{\mathcal{Y}} \left[\mathbb{1}(y_i \le u) - F(u|\mathbf{x}_i) \right] \mathbf{P}\Big(\frac{u-y}{h}\Big) \mathrm{d}G(u) \right) \mathbf{Q}\Big(\frac{\mathbf{x}_i - \mathbf{x}}{h}\Big)^{\mathrm{T}} \tag{I}
$$

$$
+ \frac{1}{n^2 h^{d+1}} \sum_{\substack{i,j=1 \\ \text{distinct}}}^{n} \left(\left[\mathbb{1}(y_i \le y_j) - F(y_j|\mathbf{x}_i) \right] \mathbf{P}\Big(\frac{y_j - y}{h}\Big) - \int_{\mathcal{Y}} \left[\mathbb{1}(y_i \le u) - F(u|\mathbf{x}_i) \right] \mathbf{P}\Big(\frac{u-y}{h}\Big) \mathrm{d}G(u) \right) \mathbf{Q}\Big(\frac{\mathbf{x}_i - \mathbf{x}}{h}\Big)^{\mathrm{T}}
$$
$$
\tag{II}
$$

$$
+ \frac{1}{nh^{d+1}} \sum_{i=1}^{n} \int_{\mathcal{Y}} \left[\mathbb{1}(y_i \le u) - F(u|\mathbf{x}_i) \right] \mathbf{P}\Big(\frac{u-y}{h}\Big) \mathrm{d}G(u) \mathbf{Q}\Big(\frac{\mathbf{x}_i - \mathbf{x}}{h}\Big)^{\mathrm{T}}, \tag{III}
$$

where we set $G = F_y$. Term (I) represents the leave-in bias, and it is straightforward to show that

$$
\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left| (\mathrm{I}) \right| > \mathfrak{c}_1 \frac{1}{n} \left(1 + \sqrt{\frac{\log n}{nh^{d+1}}} \right) \right] \le \mathfrak{c}_2 n^{-\mathfrak{c}_3},
$$

for some constants $\mathfrak{c}_1$, $\mathfrak{c}_2$, and $\mathfrak{c}_3$. See Lemma SA-2.1 for the proof strategy.

Term (II) is a degenerate U-statistic. Define

$$u_{i,j} = \mathbf{c}_1^{\mathrm{T}}\left(\left[\mathbb{1}(y_i \le y_j) - F(y_j|\mathbf{x}_i)\right]\mathbf{P}\left(\frac{y_j - y}{h}\right) - \int_{\mathcal{Y}}\left[\mathbb{1}(y_i \le u) - F(u|\mathbf{x}_i)\right]\mathbf{P}\left(\frac{u - y}{h}\right)\mathrm{d}G(u)\right)\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}}\mathbf{c}_2,$$

where $\mathbf{c}_1$ and $\mathbf{c}_2$ are arbitrary (fixed) vectors of conformable dimensions. Then with

$$A = C', \qquad B^2 = C'nh, \qquad D^2 = C'n^2h^{d+1},$$

for some constant C', and

$$t = C(\log n)\sqrt{n^2h^{d+1}}$$

for some large constant C, we apply Lemma SA-7.3, which gives (the value of C' may change for each line)

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\sum_{\substack{i,j=1\\ \text{distinct}}}^n u_{i,j}\right| > t\right] \le C'\exp\left\{-\frac{1}{C'}\min\left[\frac{t}{\sqrt{n^2h^{d+1}}}, \frac{t^{2/3}}{(nh)^{1/3}}, t^{1/2}\right] + \log n\right\}$$

$$= C'\exp\left\{-\frac{\sqrt{C}}{C'}\min\left[\log n, \left((\log n)^2 nh^d\right)^{\frac{1}{3}}, \left((\log n)^2 n^2 h^{d+1}\right)^{\frac{1}{4}}\right] + \log n\right\}.$$

As a result,

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|(\text{II})\right| > \mathfrak{c}_1\frac{\log n}{\sqrt{n^2h^{d+1}}}\right] \le \mathfrak{c}_2 n^{-\mathfrak{c}_3},$$

for some constants $\mathfrak{c}_1$, $\mathfrak{c}_2$, and $\mathfrak{c}_3$.

We now collect the pieces. The difference between $\hat{\boldsymbol{\$}}_{\mu,\nu}^{\circ}(y,\mathbf{x})$ and $\check{\boldsymbol{\$}}_{\mu,\nu}^{\circ}(y,\mathbf{x})$ is

$$\hat{\boldsymbol{\$}}_{\mu,\nu}^{\circ}(y,\mathbf{x}) - \check{\boldsymbol{\$}}_{\mu,\nu}^{\circ}(y,\mathbf{x})$$
$$= \frac{1}{h^{\mu+|\nu|}\sqrt{V_{\mu,\nu}(y,\mathbf{x})}}\mathbf{e}_\mu^{\mathrm{T}}\hat{\mathbf{S}}_y^{-1}\left[(\text{I}) + (\text{II})\right]\hat{\mathbf{S}}_\mathbf{x}^{-1}\mathbf{e}_\nu + \frac{1}{\sqrt{V_{\mu,\nu}(y,\mathbf{x})}}\mathbf{e}_\mu^{\mathrm{T}}\left(\hat{\mathbf{S}}_y^{-1} - \mathbf{S}_y^{-1}\right)\left[\bar{\mathbf{R}}_{y,\mathbf{x}} - \mathbb{E}\left[\bar{\mathbf{R}}_{y,\mathbf{x}}|\mathbf{X}\right]\right]\hat{\mathbf{S}}_\mathbf{x}^{-1}\mathbf{e}_\nu,$$

and the conclusion follows from Lemmas SA-2.1 and SA-2.3.

SA-7.8 Proof of Lemma SA-3.1

Part (i) Convergence of $\check{\theta}_{\mu,\nu} - \theta_{\mu,\nu}$. Recall that we have the following decomposition of our estimator

$$\check{\theta}_{\mu,\nu} - \theta_{\mu,\nu} = \frac{1}{nh^{1+d+\mu+|\nu|}}\sum_{i=1}^n\left[\int_{\mathcal{Y}}F(u|\mathbf{x}_i)\mathbf{e}_\mu^{\mathrm{T}}\mathbf{S}_y^{-1}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\right]\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}}\hat{\mathbf{S}}_\mathbf{x}^{-1}\mathbf{e}_\nu - \theta_{\mu,\nu} \qquad (\text{I})$$

$$+ \frac{1}{nh^{1+d+\mu+|\nu|}}\sum_{i=1}^n\left[\int_{\mathcal{Y}}\left(\mathbb{1}(y_i \le u) - F(u|\mathbf{x}_i)\right)\mathbf{e}_\mu^{\mathrm{T}}\mathbf{S}_y^{-1}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\right]\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}}\mathbf{S}_\mathbf{x}^{-1}\mathbf{e}_\nu \qquad (\text{II})$$

$$+ \frac{1}{nh^{1+d+\mu+|\nu|}}\sum_{i=1}^n\left[\int_{\mathcal{Y}}\left(\mathbb{1}(y_i \le u) - F(u|\mathbf{x}_i)\right)\mathbf{e}_\mu^{\mathrm{T}}\mathbf{S}_y^{-1}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\right]\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathrm{T}}\left(\hat{\mathbf{S}}_\mathbf{x}^{-1} - \mathbf{S}_\mathbf{x}^{-1}\right)\mathbf{e}_\nu.$$
$$(\text{III})$$

(I) is simply the conditional bias, whose order is given in Lemma SA-2.2. The convergence rate of (II) can be easily deduced from that of $\mathbf{e}_\mu^{\mathrm{T}}\mathbf{S}_y^{-1}\left(\bar{\mathbf{R}}_{y,\mathbf{x}} - \mathbb{E}\left[\bar{\mathbf{R}}_{y,\mathbf{x}}|\mathbf{X}\right]\right)$ in Lemma SA-2.1. Finally, it should be clear that (III) is negligible relative to (II).

Part (ii) Convergence of $\hat{\theta}_{\mu,\nu} - \theta_{\mu,\nu}$. This part follows from Remark SA-2.3.

SA-7.9   Proof of Lemma SA-3.2

SA-7.9.1   Uniform consistency of $\check{V}_{\mu,\nu}(y,\mathbf{x})$

For the purposes of this proof, let $\mathbf{c}_1 = \mathbf{S}_y^{-1}\mathbf{e}_\mu$, $\hat{\mathbf{c}}_2 = \hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_\nu$, and $\mathbf{c}_2 = \mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_\nu$. To start, consider

$$
\frac{1}{n^2 h^{2d+2\mu+2|\nu|}} \sum_{i=1}^n \left[\int_{\mathcal{Y}} \left(\mathbb{1}\left(y_i \le u\right) - \hat{F}(u|\mathbf{x}_i) \right) \frac{1}{h} \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{u-y}{h}\right) \mathrm{d}G(u) \mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2
$$

$$
= \frac{1}{n^2 h^{2d+2\mu+2|\nu|}} \sum_{i=1}^n \left[\int_{\mathcal{Y}} \left(\mathbb{1}\left(y_i \le u\right) - F(u|\mathbf{x}_i) \right) \frac{1}{h} \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{u-y}{h}\right) \mathrm{d}G(u) \mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2 \tag{I}
$$

$$
- \frac{2}{n^2 h^{2d+2\mu+2|\nu|}} \sum_{i=1}^n \iint_{\frac{\mathcal{Y}-y}{h}} \left(\mathbb{1}\left(y_i \le u_1\right) - F(u_1|\mathbf{x}_i) \right)\left(\hat{F}(u_2|\mathbf{x}_i) - F(u_2|\mathbf{x}_i) \right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(u_1\right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(u_2\right)
$$

$$
g(y+hu_1)g(y+hu_2)\mathrm{d}u_1\mathrm{d}u_2 \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2 \tag{II}
$$

$$
+ \frac{1}{n^2 h^{2d+2\mu+2|\nu|}} \sum_{i=1}^n \left[\int_{\mathcal{Y}} \left(\hat{F}(u|\mathbf{x}_i) - F(u|\mathbf{x}_i) \right) \frac{1}{h} \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{u-y}{h}\right) \mathrm{d}G(u) \mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2. \tag{III}
$$

First consider term (I). Clearly this term is unbiased for $\mathsf{V}_{\mu,\nu}(y,\mathbf{x})$. In the proof of Lemma SA-2.4, we showed that

$$
\mathbb{V}\left[\left(\int_{\mathcal{Y}} \left(\mathbb{1}\left(y_i \le u\right) - F(u|\mathbf{x}_i) \right) \frac{1}{h} \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{u-y}{h}\right) \mathrm{d}G(u) \mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right)^2 \right] \le C_1 \begin{cases} h^d & \text{if } \mu = 0 \\ h^{d+1} & \text{if } \mu > 0 \end{cases}.
$$

Also note that

$$
\left(\int_{\mathcal{Y}} \left(\mathbb{1}\left(y_i \le u\right) - F(u|\mathbf{x}_i) \right) \frac{1}{h} \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{u-y}{h}\right) \mathrm{d}G(u) \mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right)^2 \le C_2.
$$

In the above, the constants, C_1 and C_2, can be chosen to be independent of the evaluation point, the sample size, and the bandwidth. Then by a proper discretization of $\mathcal{Y} \times \mathcal{X}$, and applying the union bound and Bernstein's inequality, one has

$$
\mathbb{P}\left[\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} \left| \frac{(\mathrm{I}) - \mathsf{V}_{\mu,\nu}(y,\mathbf{x})}{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})} \right| > \mathfrak{c}_1 \mathfrak{r}_1 \right] \le \mathfrak{c}_2 n^{-\mathfrak{c}_3}, \qquad \mathfrak{r}_1 = \begin{cases} \sqrt{\frac{\log n}{nh^d}} & \text{if } \mu = 0 \\ \sqrt{\frac{\log n}{nh^{d+1}}} & \text{if } \mu > 0 \end{cases},
$$

for some constants $\mathfrak{c}_1$, $\mathfrak{c}_2$, and $\mathfrak{c}_3$. In addition, $\mathfrak{c}_3$ can be made arbitrarily large by appropriate choices of $\mathfrak{c}_1$. See the proof of Lemma SA-2.1 for an example of this proof strategy.

Next consider term (III). With the uniform convergence result for the estimated conditional distribution function, it is clear that

$$
\mathbb{P}\left[\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} \left| \frac{(\mathrm{III})}{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})} \right| > \mathfrak{c}_1 \mathfrak{r}_3 \right] \le \mathfrak{c}_2 n^{-\mathfrak{c}_3}, \qquad \mathfrak{r}_3 = h^{2\mathfrak{q}+1} + \frac{\log n}{nh^{d+1}}.
$$

Finally, we consider (II). Using the Cauchy-Schwartz inequality, we have

$$
|(\mathrm{II})|^2 \le |(\mathrm{I})| \cdot |(\mathrm{III})| .
$$

As a result,

$$
\mathbb{P}\left[\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} \left| \frac{(\mathrm{II})}{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})} \right| > \mathfrak{c}_1 \mathfrak{r}_2 \right] \le \mathfrak{c}_2 n^{-\mathfrak{c}_3}, \qquad \mathfrak{r}_2 = \sqrt{1 + \mathfrak{r}_1} \sqrt{\mathfrak{r}_3}.
$$

To conclude the proof for $\check{V}_{\mu,\nu}(y,\mathbf{x})$, we note that replacing $\hat{\mathbf{c}}_2$ by $\mathbf{c}_2$ only leads to an additional multiplicative factor $1 + O_{\mathbb{P}}(\sqrt{\log n/(nh^d)})$. See Lemma SA-2.1.

SA-7.9.2 Uniform consistency of $\hat{V}_{\mu,\nu}(y, \mathbf{x})$

For the purposes of this proof, let $\hat{\mathbf{c}}_1 = \hat{\mathbf{S}}_y^{-1} \mathbf{e}_\mu$, $\mathbf{c}_1 = \mathbf{S}_y^{-1} \mathbf{e}_\mu$, $\hat{\mathbf{c}}_2 = \hat{\mathbf{S}}_\mathbf{x}^{-1} \mathbf{e}_\nu$, and $\mathbf{c}_2 = \mathbf{S}_\mathbf{x}^{-1} \mathbf{e}_\nu$.

We consider the same decomposition used in the proof of Lemma SA-2.4:

$$\frac{1}{n^2 h^{2d+2\mu+2|\nu|}} \sum_{i=1}^n \left[\int_{\mathcal{Y}} \left(\mathbb{1}\left(y_i \le u\right) - \hat{F}(u|\mathbf{x}_i) \right) \frac{1}{h} \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{u-y}{h}\right) \mathrm{d}\hat{F}_y(u) \mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \right]^2$$

$$= \frac{1}{n^2 h^{2d+2\mu+2|\nu|}} \sum_{i=1}^n \left[\frac{1}{n^2} \sum_{j,k=1}^n \left(\mathbb{1}\left(y_i \le y_j\right) - F(y_j|\mathbf{x}_i) \right) \left(\mathbb{1}\left(y_i \le y_k\right) - F(y_k|\mathbf{x}_i) \right) \frac{1}{h^2} \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j-y}{h}\right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_k-y}{h}\right) \right] \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right) \right]^2$$

(I)

$$- \frac{2}{n^2 h^{2d+2\mu+2|\nu|}} \sum_{i=1}^n \left[\frac{1}{n^2} \sum_{j,k=1}^n \left(\hat{F}(y_j|\mathbf{x}_i) - F(y_j|\mathbf{x}_i) \right) \left(\mathbb{1}\left(y_i \le y_k\right) - F(y_k|\mathbf{x}_i) \right) \frac{1}{h^2} \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j-y}{h}\right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_k-y}{h}\right) \right] \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right) \right]^2$$

(II)

$$+ \frac{1}{n^2 h^{2d+2\mu+2|\nu|}} \sum_{i=1}^n \left[\frac{1}{n^2} \sum_{j,k=1}^n \left(\hat{F}(y_j|\mathbf{x}_i) - F(y_j|\mathbf{x}_i) \right) \left(\hat{F}(y_k|\mathbf{x}_i) - F(y_k|\mathbf{x}_i) \right) \frac{1}{h^2} \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j-y}{h}\right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_k-y}{h}\right) \right] \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right) \right]^2.$$

(III)

By the uniform convergence rate for the estimated conditional distribution function, we have

$$\mathbb{P}\left[\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} \left| \frac{(\text{III})}{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})} \right| > \mathfrak{c}_1 \mathfrak{r}_3 \right] \le \mathfrak{c}_2 n^{-\mathfrak{c}_3}, \qquad \mathfrak{r}_3 = h^{2\mathfrak{q}+1} + \frac{\log n}{n h^{d+1}}.$$

Employing the Cauchy-Schwartz inequality gives

$$|(\text{II})|^2 \le |(\text{I})| \cdot |(\text{III})|.$$

As a result, a probabilistic order for term (II) follows that of terms (I) and (III).

Finally, consider term (I), which has the expansion

$$(\text{I}) = \frac{1}{n^4 h^{2d+2\mu+2|\nu|+2}} \sum_{\substack{i,j,k=1 \\ \text{distinct}}}^n \left[\left(\mathbb{1}\left(y_i \le y_j\right) - F(y_j|\mathbf{x}_i) \right) \left(\mathbb{1}\left(y_i \le y_k\right) - F(y_k|\mathbf{x}_i) \right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j-y}{h}\right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_k-y}{h}\right) \right] \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right) \right]^2$$

(I.1)

$$+ \frac{2}{n^4 h^{2d+2\mu+2|\nu|+2}} \sum_{\substack{i,j=1 \\ \text{distinct}}}^n \left[\left(\mathbb{1}\left(y_i \le y_j\right) - F(y_j|\mathbf{x}_i) \right) \left(1 - F(y_i|\mathbf{x}_i) \right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j-y}{h}\right) \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_i-y}{h}\right) \right] \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right) \right]^2$$

(I.2)

$$+ \frac{1}{n^4 h^{2d+2\mu+2|\nu|+2}} \sum_{\substack{i,j=1 \\ \text{distinct}}}^n \left(\mathbb{1}\left(y_i \le y_j\right) - F(y_j|\mathbf{x}_i) \right)^2 \left[\mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j-y}{h}\right) \right]^2 \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right) \right]^2$$

(I.3)

$$+ \frac{1}{n^4 h^{2d+2\mu+2|\nu|+2}} \sum_{i=1}^n \left(1 - F(y_i|\mathbf{x}_i) \right)^2 \left[\mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_i-y}{h}\right) \right]^2 \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right) \right]^2.$$

(I.4)

Then,

$$|(\text{I.2})| \le \frac{1}{n^4 h^{2d+2\mu+2|\nu|+2}} \sum_{\substack{i,j=1 \\ \text{distinct}}}^n \left| \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_j-y}{h}\right) \right| \cdot \left| \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_i-y}{h}\right) \right| \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right) \right]^2$$

$$\le \left(\frac{1}{n h^{d+2\mu+2|\nu|}} \right) \frac{2}{n} \left[\frac{1}{nh} \sum_{i=1}^n \left| \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_i-y}{h}\right) \right| \right] \left[\frac{1}{n h^{d+1}} \sum_{i=1}^n \left| \mathbf{c}_1^{\mathrm{T}} \mathbf{P}\left(\frac{y_i-y}{h}\right) \right| \left[\mathbf{c}_2^{\mathrm{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right) \right]^2 \right],$$

38

which means

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{(\text{I.2})}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\right|\geq\mathfrak{c}_1\frac{1}{nh}\right]\leq\mathfrak{c}_2n^{-\mathfrak{c}_3}.$$

Using similar techniques, one can show that

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{(\text{I.3})}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\right|\geq\mathfrak{c}_1\frac{1}{nh^2}\right]\leq\mathfrak{c}_2n^{-\mathfrak{c}_3},\qquad\text{and}\qquad\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{(\text{I.4})}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\right|\geq\mathfrak{c}_1\frac{1}{n^2h^2}\right]\leq\mathfrak{c}_2n^{-\mathfrak{c}_3}.$$

To streamline the remaining derivation, define

$$\phi_{j,i}=\frac{1}{h}\left(\mathbb{1}\left(y_i\leq y_j\right)-F(y_j|\mathbf{x}_i)\right)\mathbf{c}_1^{\mathrm{T}}\mathbf{P}\left(\frac{y_j-y}{h}\right),\quad\phi_i=\mathbb{E}[\phi_{j,i}|y_i,\mathbf{x}_i],\quad\psi_i=\left[\mathbf{c}_2^{\mathrm{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\right]^2.$$

Then

$$(\text{I.1})=\frac{1}{n^4h^{2d+2\mu+2|\boldsymbol{\nu}|}}\sum_{\substack{i,j,k=1\\\text{distinct}}}^n\phi_{j,i}\phi_{k,i}\psi_i$$

$$=\underbrace{\frac{1}{n^4h^{2d+2\mu+2|\boldsymbol{\nu}|}}\sum_{\substack{i,j,k=1\\\text{distinct}}}^n(\phi_{j,i}-\phi_i)(\phi_{k,i}-\phi_i)\psi_i}_{(\text{I.1.1})}+\left(2+O\left(\frac{1}{n}\right)\right)\underbrace{\frac{1}{n^3h^{2d+2\mu+2|\boldsymbol{\nu}|}}\sum_{\substack{i,j=1\\\text{distinct}}}^n(\phi_{j,i}-\phi_i)\phi_i\psi_i}_{(\text{I.1.2})}$$

$$+\left(1+O\left(\frac{1}{n}\right)\right)\underbrace{\frac{1}{n^2h^{2d+2\mu+2|\boldsymbol{\nu}|}}\sum_{i=1}^n\phi_i^2\psi_i}_{(\text{I.1.3})}.$$

By employing the same techniques in the proof for $\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$, we have that

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{(\text{I.1.3})-\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\right|>\mathfrak{c}_1\mathbf{r}_1\right]\leq\mathfrak{c}_2n^{-\mathfrak{c}_3},\qquad\mathbf{r}_1=\begin{cases}\sqrt{\frac{\log n}{nh^d}}&\text{if }\mu=0\\\sqrt{\frac{\log n}{nh^{d+1}}}&\text{if }\mu>0\end{cases}.$$

Term (I.1.2) admits the following decomposition:

$$(\text{I.1.2})=\underbrace{\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}}\frac{n-1}{n^2}\sum_{j=1}^n\mathbb{E}\left[\frac{1}{h^d}(\phi_{j,i}-\phi_i)\phi_i\psi_i\Big|y_j,\mathbf{x}_j\right]}_{(\text{I.1.2.1})}$$

$$+\underbrace{\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}}\frac{1}{n^2h^d}\sum_{\substack{i,j=1\\\text{distinct}}}^n\left((\phi_{j,i}-\phi_i)\phi_i\psi_i-\mathbb{E}\left[\frac{1}{h^d}(\phi_{j,i}-\phi_i)\phi_i\psi_i\Big|y_j,\mathbf{x}_j\right]\right)}_{(\text{I.1.2.2})}.$$

Using the same techniques of Lemmas SA-2.1 and SA-2.4, we have

$$\mu=0\qquad\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{(\text{I.1.2.1})}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\right|>\mathfrak{c}_1\sqrt{\frac{\log n}{nh}}\right]\leq\mathfrak{c}_2n^{-\mathfrak{c}_3},$$

$$\mu>0\qquad\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{(\text{I.1.2.1})}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\right|>\mathfrak{c}_1\sqrt{\frac{\log n}{nh^2}}\right]\leq\mathfrak{c}_2n^{-\mathfrak{c}_3}.$$

Term (I.1.2.2) is a degenerate second order U-statistic. We adopt Lemma SA-7.3, which implies (see Remark SA-2.3

and its proof for an example)

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{(\text{I.1.2.2})}{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}\right|>\mathfrak{c}_1\sqrt{\frac{\log n}{n^2h^{d+3}}}\right]\le\mathfrak{c}_2n^{-\mathfrak{c}_3}.$$

To handle term (I.1.1), first consider the quantity $\phi_{j,i}-\phi_i$, which takes the form

$$\max_i\left|\frac{1}{n}\sum_{j=1}^n(\phi_{j,i}-\phi_i)\right|$$

$$\le\sup_{y'\in\mathcal{Y},\mathbf{x}'\in\mathcal{X}}\left|\frac{1}{nh}\sum_{j=1}^n\left[\left(\mathbb{1}\left(y'\le y_j\right)-F(y_j|\mathbf{x}')\right)\mathbf{c}_1^{\mathrm{T}}\mathbf{P}\left(\frac{y_j-y}{h}\right)-\int\left(\mathbb{1}\left(y'\le u\right)-F(u|\mathbf{x}')\right)\mathbf{c}_1^{\mathrm{T}}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\right]\right|.$$

Then it is straightforward to show that

$$\mathbb{P}\left[\max_i\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{1}{n}\sum_{j=1}^n(\phi_{j,i}-\phi_i)\right|\ge\mathfrak{c}_1\sqrt{\frac{\log n}{nh}}\right]\le\mathfrak{c}_2n^{-\mathfrak{c}_3}.$$

As a result,

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{(\text{I.1.1})}{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}\right|>\mathfrak{c}_1\frac{\log n}{nh^2}\right]\le\mathfrak{c}_2n^{-\mathfrak{c}_3}.$$

To conclude the proof for $\hat{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})$, we note that replacing $\hat{\mathbf{c}}_1$ by $\mathbf{c}_1$ and $\hat{\mathbf{c}}_2$ by $\mathbf{c}_2$ only leads to an additional multiplicative factor $1+O_{\mathbb{P}}(\sqrt{\log n/nh^d})$. See Lemma SA-2.1.

SA-7.10 Proof of Lemma SA-3.3

First consider $\check{\mathbb{T}}_{\mu,\nu}^{\circ}(y,\mathbf{x})$. The difference between $\check{\mathbb{T}}_{\mu,\nu}^{\circ}(y,\mathbf{x})$ and $\check{\$}_{\mu,\nu}^{\circ}(y,\mathbf{x})$ is

$$\check{\mathbb{T}}_{\mu,\nu}^{\circ}(y,\mathbf{x})-\check{\$}_{\mu,\nu}^{\circ}(y,\mathbf{x})=\left(\sqrt{\frac{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})}}-1\right)\check{\$}_{\mu,\nu}^{\circ}(y,\mathbf{x}).$$

From Lemma SA-3.2, we have

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\sqrt{\frac{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})}}-1\right|>\mathfrak{c}_1\mathtt{r}_{\mathtt{VE}}\right]\le\mathfrak{c}_2n^{-\mathfrak{c}_3}.$$

To close the proof, it is straightforward to verify that

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\check{\$}_{\mu,\nu}^{\circ}(y,\mathbf{x})\right|>\mathfrak{c}_1\sqrt{\log n}\right]\le\mathfrak{c}_2n^{-\mathfrak{c}_3},$$

which follows from the uniform convergence rate in Lemma SA-3.1. The same technique applies to the analysis of $\hat{\mathbb{T}}_{\mu,\nu}^{\circ}(y,\mathbf{x})-\hat{\$}_{\mu,\nu}^{\circ}(y,\mathbf{x})$.

SA-7.11 Proof of Lemma SA-3.4

We first rewrite the kernel using change-of-variable

$$h^{d+\mu+|\boldsymbol{\nu}|}\mathscr{K}_{\mu,\nu,h}(a,\mathbf{b};y,\mathbf{x})=\mathbf{e}_{\mu}^{\mathrm{T}}\mathbf{S}_y^{-1}\left[\int_{\frac{\mathcal{Y}-y}{h}}\mathbb{1}(a\le y+hv)\mathbf{P}(v)\,g(y+hv)\mathrm{d}v\right]\mathbf{Q}\left(\frac{\mathbf{b}-\mathbf{x}}{h}\right)^{\mathrm{T}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\nu}.$$

Then it should be clear that the kernel is bounded. The same holds for $h^{d+\mu+|\boldsymbol{\nu}|}\mathscr{K}_{\mu,\nu,h}^{\circ}(a,\mathbf{b};y,\mathbf{x})$.

Next consider two cases. If $(a-y)/h$ is larger than 1, then the integrand $\mathbb{1}(a\le y+hv)\mathbf{P}(v)$ will be zero because

$\mathbf{P}(v)$ is zero for $v \geq 1$. Therefore, the kernel defined above will be zero as well. For the case that $(a - y)/h \leq -1$, we can simply drop the indicator, as again $\mathbf{P}(v)$ will be zero for $v \leq -1$. Then the kernel becomes

$$h^{d+\mu+|\boldsymbol{\nu}|} \mathscr{K}_{\mu,\boldsymbol{\nu},h}(a, \mathbf{b}; y, \mathbf{x}) = \mathbf{e}_\mu^{\mathrm{T}} \mathbf{S}_y^{-1} \left[\int_{\frac{\mathcal{Y}-y}{h}} \mathbf{P}(v) g(y + hv) \mathrm{d}v \right] \mathbf{Q} \left(\frac{\mathbf{b} - \mathbf{x}}{h} \right)^{\mathrm{T}} \mathbf{S}_{\mathbf{x}}^{-1} \mathbf{e}_{\boldsymbol{\nu}}, \qquad a \leq -1.$$

Note that the matrix, $\mathbf{S}_y$, can be written as

$$\mathbf{S}_y = \int_{\frac{\mathcal{Y}-y}{h}} \mathbf{P}(v) \mathbf{p}(v)^{\mathrm{T}} g(y + hv) \mathrm{d}v,$$

which means $\int_{\frac{\mathcal{Y}-y}{h}} \mathbf{P}(v) g(y + hv) \mathrm{d}v$ is simply the first column of $\mathbf{S}_y$. As a result, $\mathscr{K}_{\mu,\boldsymbol{\nu},h}(a, \mathbf{b}; y, \mathbf{x})$ is zero provided that $\mu \geq 1$ and $|a - y|/h \geq 1$.

As for the second argument, $\mathbf{b}$, we note that $\mathbf{Q}((\mathbf{b} - \mathbf{x})/h)$ is zero if $\mathbf{b}$ lies outside of a h-cube around $\mathbf{x}$. This concludes our proof.

SA-7.12 Proof of Lemma SA-3.5

We will consider $h^{d+\mu+|\boldsymbol{\nu}|} \mathscr{K}_{\mu,\boldsymbol{\nu},h}(a, \mathbf{b}; y, \mathbf{x})$, which allows us to ignore the extra scaling factor. We first rewrite the kernel using change-of-variable

$$h^{d+\mu+|\boldsymbol{\nu}|} \mathscr{K}_{\mu,\boldsymbol{\nu},h}(a, \mathbf{b}; y, \mathbf{x}) = \mathbf{e}_\mu^{\mathrm{T}} \mathbf{S}_y^{-1} \left[\int_{\frac{\mathcal{Y}-y}{h}} \mathbb{1}(a \leq y + hv) \mathbf{P}(v) g(y + hv) \mathrm{d}v \right] \mathbf{Q} \left(\frac{\mathbf{b} - \mathbf{x}}{h} \right)^{\mathrm{T}} \mathbf{S}_{\mathbf{x}}^{-1} \mathbf{e}_{\boldsymbol{\nu}},$$

then it should be obvious that it is Lipschitz-h^{-1} continuous with respect to $\mathbf{b}$, as

$$\sup_{a,y,\mathbf{x}} h^{d+\mu+|\boldsymbol{\nu}|} |\mathscr{K}_{\mu,\boldsymbol{\nu},h}(a, \mathbf{b}; y, \mathbf{x}) - \mathscr{K}_{\mu,\boldsymbol{\nu},h}(a, \mathbf{b}'; y, \mathbf{x})|$$

$$\leq \left| \mathbf{Q} \left(\frac{\mathbf{b} - \mathbf{x}}{h} \right)^{\mathrm{T}} - \mathbf{Q} \left(\frac{\mathbf{b}' - \mathbf{x}}{h} \right)^{\mathrm{T}} \right| \sup_{a,y} \left| \mathbf{e}_\mu^{\mathrm{T}} \mathbf{S}_y^{-1} \left[\int_{\frac{\mathcal{Y}-y}{h}} \mathbb{1}(a \leq y + hv) \mathbf{P}(v) g(y + hv) \mathrm{d}v \right] \right| \sup_{\mathbf{x}} \left| \mathbf{S}_{\mathbf{x}}^{-1} \mathbf{e}_{\boldsymbol{\nu}} \right| \precsim h^{-1},$$

because $\mathbf{Q}(\cdot)$ is Lipschitz continuous.

Next consider the direction a. Again, we have

$$\sup_{\mathbf{b},y,\mathbf{x}} h^{d+\mu+|\boldsymbol{\nu}|} |\mathscr{K}_{\mu,\boldsymbol{\nu},h}(a, \mathbf{b}; y, \mathbf{x}) - \mathscr{K}_{\mu,\boldsymbol{\nu},h}(a', \mathbf{b}; y, \mathbf{x})|$$

$$\leq \sup_y \left| \int_{\frac{\mathcal{Y}-y}{h}} \left(\mathbb{1}(a \leq y + hv) - \mathbb{1}(a' \leq y + hv) \right) \mathbf{P}(v) g(y + hv) \mathrm{d}v \right| \sup_y \left| \mathbf{e}_\mu^{\mathrm{T}} \mathbf{S}_y^{-1} \right| \sup_{\mathbf{b},\mathbf{x}} \left| \mathbf{Q} \left(\frac{\mathbf{b} - \mathbf{x}}{h} \right)^{\mathrm{T}} \mathbf{S}_{\mathbf{x}}^{-1} \mathbf{e}_{\boldsymbol{\nu}} \right|$$

$$\precsim \sup_y \left| \int_{\frac{\mathcal{Y}-y}{h} \cap [-1,1] \cap \left[\frac{a-y}{h}, \frac{a'-y}{h} \right]} \mathbf{P}(v) g(y + hv) \mathrm{d}v \right|.$$

Therefore, the kernel is also Lipschitz-h^{-1} continuous with respect to a. The analysis of $h^{d+\mu+|\boldsymbol{\nu}|} \mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(a, \mathbf{b}; y, \mathbf{x})$ is similar.

Now we prove the second claim. First, it is not difficult to show that $\mathbf{S}_y$ is Lipschitz continuous with respect to y, with the Lipschitz constant having the order $1/h$. The same holds for its inverse, $\mathbf{S}_y^{-1}$, as $\mathbf{S}_y$ is uniformly bounded away from being singular. As a result,

$$\left| \mathbf{S}_y^{-1} - \mathbf{S}_{y'}^{-1} \right| \precsim \frac{1}{h} |y - y'|.$$

Similarly, one can show that

$$\left| \mathbf{S}_{\mathbf{x}}^{-1} - \mathbf{S}_{\mathbf{x}'}^{-1} \right| \precsim \frac{1}{h} |\mathbf{x} - \mathbf{x}'|.$$

41

Now consider the following difference

$$\sup_a \left| \int_{\mathcal{Y}} \mathbb{1}(a \leq u) \frac{1}{h} \left(\mathbf{P}\left(\frac{u-y}{h}\right) - \mathbf{P}\left(\frac{u-y'}{h}\right) \right) g(u) \mathrm{d}u \right| \sup_y \left| \mathbf{e}_\mu^\mathsf{T} \mathbf{S}_y^{-1} \right| \sup_{\mathbf{b},\mathbf{x}} \left| \mathbf{Q}\left(\frac{\mathbf{b}-\mathbf{x}}{h}\right)^\mathsf{T} \mathbf{S}_\mathbf{x}^{-1} \mathbf{e}_\nu \right|.$$

It is obvious that the kernel is Lipschitz-h^{-1} continuous in y and $\mathbf{x}$. The analysis of $h^{d+\mu+|\nu|} \mathscr{K}_{\mu,\nu,h}^\circ (a,\mathbf{b};y,\mathbf{x})$ is similar.

SA-7.13 Proof of Lemma SA-3.6

This proof is motivated by Lemma 4.1 in Rio (1994). Take $\ell = \lfloor 1/h \rfloor$, and partition each coordinate $[0,1]$ into ℓ intervals of equal length. This will lead to a partition $\mathcal{A} = \{A_j : 1 \leq j \leq \ell^d\}$ of $[0,1]^d$. Next, consider sets whose P-measure exceeds ε,

$$\mathcal{A}_{P,\varepsilon} = \{A \in \mathcal{A} : P[A] > \varepsilon\},$$

and their $\mathfrak{c}h$-enlargements

$$\mathcal{A}_{P,\varepsilon}^{\mathfrak{c}h} = \{A + [-\mathfrak{c}h, \mathfrak{c}h]^d : A \in \mathcal{A}_{P,\varepsilon}\}.$$

Importantly, if $\mathbf{z}$ does not belong to any set in $\mathcal{A}_{P,\varepsilon}^{\mathfrak{c}h}$, it means the support of the function $g_\mathbf{z}\left(\frac{\cdot-\mathbf{z}}{h}\right)$ will not intersect with any set in $\mathcal{A}_{P,\varepsilon}$. In this case,

$$\int \left| g_\mathbf{z}\left(\frac{\cdot-\mathbf{z}}{h}\right) \right| \mathrm{d}P \leq \mathfrak{c}P\left[h \cdot \mathrm{supp}(g_\mathbf{z}(\cdot)) + \mathbf{z}\right].$$

Define the complement of $\mathcal{A}_{P,\varepsilon}$ as

$$\mathcal{A}_{P,\varepsilon}^\perp = \{A \in \mathcal{A} : P[A] \leq \varepsilon\},$$

Then the set $h \cdot \mathrm{supp}(g_\mathbf{z}(\cdot)) + \mathbf{z}$ will be completely covered by sets in $\mathcal{A}_{P,\varepsilon}^\perp$. To determine the maximum number of intersections between $h \cdot \mathrm{supp}(g_\mathbf{z}(\cdot)) + \mathbf{z}$ and sets in $\mathcal{A}_{P,\varepsilon}^\perp$, it suffices to consider the Euclidean volume of the enlarged set $h \cdot \mathrm{supp}(g_\mathbf{z}(\cdot)) + \mathbf{z} + [-\ell^{-1}, \ell^{-1}]^d$, which is $(2\mathfrak{c}h + \ell^{-1})^d$. The Euclidean volume of each set in $\mathcal{A}_{P,\varepsilon}^\perp$ is ℓ^{-d}. Therefore, the set $h \cdot \mathrm{supp}(g_\mathbf{z}(\cdot)) + \mathbf{z}$ can intersect with at most

$$\frac{(2\mathfrak{c}h + \ell^{-1})^d}{\ell^{-d}} = (2\mathfrak{c}h\ell + 1)^d \leq (2\mathfrak{c} + 1)^d$$

sets in $\mathcal{A}_{P,\varepsilon}^\perp$. As a result, we conclude that

$$\int \left| g_\mathbf{z}\left(\frac{\cdot-\mathbf{z}}{h}\right) \right| \mathrm{d}P \leq \mathfrak{c}(2\mathfrak{c}+1)^d \varepsilon.$$

This leads to our first result. Let $A_{P,\varepsilon}^{\mathfrak{c}h} = \cup \mathcal{A}_{P,\varepsilon}^{\mathfrak{c}h}$ be the union of sets in $\mathcal{A}_{P,\varepsilon}^{\mathfrak{c}h}$, and

$$\mathcal{G}_1 = \left\{ g_\mathbf{z}\left(\frac{\cdot-\mathbf{z}}{h}\right) : \mathbf{z} \notin A_{P,\varepsilon}^{\mathfrak{c}h} \right\},$$

then

$$N\left((2\mathfrak{c}+1)^{d+1}\varepsilon, \ \mathcal{G}_1, \ L^1(P) \right) = 1, \ \forall \varepsilon \in (0,1].$$

As remark, we note that the function class $\mathcal{G}_1$ changes with respect to h, ε, as well as the probability measure P.

Next, we consider some $\mathbf{z}$ which belongs to some set in $\mathcal{A}_{P,\varepsilon}^{\mathfrak{c}h}$. Each set in $\mathcal{A}_{P,\varepsilon}^{\mathfrak{c}h}$ is a cube with edge length $\ell^{-1} + 2\mathfrak{c}h \leq 2(\mathfrak{c}+1)h$, because $h\ell \geq 0.5$. As a result,

$$N\left(h\varepsilon, \ A_{P,\varepsilon}^{\mathfrak{c}h}, \ |\cdot| \right) \leq \sum_{A \in \mathcal{A}_{P,\varepsilon}^{\mathfrak{c}h}} N(h\varepsilon, \ A, \ |\cdot|) \leq \mathrm{card}(\mathcal{A}_{P,\varepsilon}^{\mathfrak{c}h}) \cdot \mathfrak{c}' \frac{1}{\varepsilon^d} \leq \mathfrak{c}' \frac{1}{\varepsilon^{d+1}}.$$

Here, $\mathfrak{c}'$ is some fixed number that only depends on $\mathfrak{c}$ and d. Using the Lipschitz property, we have

$$\int \left| g_{\mathbf{z}}\left(\frac{\cdot - \mathbf{z}}{h}\right) - g_{\mathbf{z}'}\left(\frac{\cdot - \mathbf{z}'}{h}\right)\right| \mathrm{d}P \leq \int \left| g_{\mathbf{z}}\left(\frac{\cdot - \mathbf{z}}{h}\right) - g_{\mathbf{z}}\left(\frac{\cdot - \mathbf{z}'}{h}\right)\right| + \left| g_{\mathbf{z}}\left(\frac{\cdot - \mathbf{z}'}{h}\right) - g_{\mathbf{z}'}\left(\frac{\cdot - \mathbf{z}'}{h}\right)\right| \mathrm{d}P$$

$$\leq \mathfrak{c}h^{-1}|\mathbf{z} - \mathbf{z}'| + \int \left| g_{\mathbf{z}}\left(\frac{\cdot - \mathbf{z}'}{h}\right) - g_{\mathbf{z}'}\left(\frac{\cdot - \mathbf{z}'}{h}\right)\right| \mathrm{d}P$$

$$\leq \mathfrak{c}h^{-1}|\mathbf{z} - \mathbf{z}'| + \mathfrak{c}h^{-1}|\mathbf{z} - \mathbf{z}'| \leq 2\mathfrak{c}h^{-1}|\mathbf{z} - \mathbf{z}'|.$$

Now define

$$\mathcal{G}_2 = \mathcal{G}\backslash\mathcal{G}_1 = \left\{ g_{\mathbf{z}}\left(\frac{\cdot - \mathbf{z}}{h}\right) : \ \mathbf{z} \in A_{P,\varepsilon}^{\mathfrak{c}h} \right\},$$

then

$$N\left((2\mathfrak{c}+1)^{d+1}\varepsilon, \ \mathcal{G}_2, \ L^1(P)\right) \leq N\left(\frac{(2\mathfrak{c}+1)^{d+1}}{2\mathfrak{c}}h\varepsilon, \ A_{P,\varepsilon}^{\mathfrak{c}h}, \ |\cdot|\right) \leq \mathfrak{c}'\frac{1}{\varepsilon^{d+1}}.$$

Combining previous results, we have

$$N\left((2\mathfrak{c}+1)^{d+1}\varepsilon, \ \mathcal{G}, \ L^1(P)\right) \leq \mathfrak{c}'\frac{1}{\varepsilon^{d+1}} + 1.$$

SA-7.14   Proof of Corollary SA-3.1

This corollary follows directly from Lemma SA-3.6 and the properties of $\mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}$ given in Lemmas SA-3.4 and SA-3.5.

SA-7.15   Proof of Theorem SA-3.1

We will apply Lemma SA-7.1. To start, consider the process (i.e., without the additional scaling in $\bar{\mathbb{\$}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$)

$$\tilde{\mathbb{\$}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h^{d+\mu+|\boldsymbol{\nu}|}\mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(y_i, \mathbf{x}_i; y, \mathbf{x}),$$

which is the empirical process indexed by the function class

$$\mathcal{K} = \left\{ h^{d+\mu+|\boldsymbol{\nu}|}\mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(\cdot, \cdot; y, \mathbf{x}) : \ y \in \mathcal{Y}, \ \mathbf{x} \in \mathcal{X} \right\}.$$

From Lemma SA-3.4, the functions in the above class are uniformly bounded. Corollary SA-3.1 shows that the function class above is of VC type, and the covering number does not depend on the bandwidth. The measurability condition required in Lemma SA-7.1 also holds, as our function class is indexed by $(y,\mathbf{x}) \in [0,1]^{d+1}$, and the functions in $\mathcal{K}$ are continuous in y and $\mathbf{x}$.

Now the only missing ingredient is the total variation of the functions in $\mathcal{K}$. First, note that the function $h^{d+\mu+|\boldsymbol{\nu}|}\mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(\cdot, \cdot; y, \mathbf{x})$ is Lipschitz continuous with respect to the arguments, and the Lipschitz constant is of order h^{-1}. Therefore, its total variation is bounded by

$$\mathrm{TV}_{(y,\mathbf{x})} = \mathrm{TV}\left(h^{d+\mu+|\boldsymbol{\nu}|}\mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(\cdot, \cdot; y, \mathbf{x})\right) \precsim \frac{1}{h}\mathrm{vol}\left(\mathrm{supp}\left(\mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(\cdot, \cdot; y, \mathbf{x})\right)\right),$$

where $\mathrm{vol}\left(\mathrm{supp}\left(\cdot\right)\right)$ denotes the Euclidean volume of the support of $\mathscr{K}_{\mu,\boldsymbol{\nu},h}^{\circ}(\cdot, \cdot; y, \mathbf{x})$. Thanks to Lemma SA-3.4, the above total variation is further bounded by

$$\mathrm{TV}_{(y,\mathbf{x})} \precsim h^d,$$

which holds for all functions in $\mathcal{K}$. That is,

$$\mathrm{TV}_{\mathcal{K}} = \sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} \mathrm{TV}_{(y,\mathbf{x})} \precsim h^d.$$

43

Putting all pieces together, we conclude that there exists a centered Gaussian process, $\tilde{\mathfrak{G}}_{\mu,\nu}$ which has the same covariance kernel as $\tilde{\mathbb{S}}_{\mu,\nu}$, such that

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\tilde{\mathbb{S}}'_{\mu,\nu}(y,\mathbf{x})-\tilde{\mathfrak{G}}_{\mu,\nu}(y,\mathbf{x})\right|\geq\mathfrak{c}_1\left(\sqrt{\frac{h^d\log n}{n^{\frac{1}{d+1}}}}+\sqrt{\frac{\log^3 n}{n}}\right)\right]\leq\mathfrak{c}_2 n^{-\mathfrak{c}_3},$$

where $\tilde{\mathbb{S}}'_{\mu,\nu}(y,\mathbf{x})$ is a copy of $\tilde{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})$. This concludes our proof.

SA-7.16 Proof of Lemma SA-3.7

Consider $\check{\mathsf{C}}_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}')$. Note that we can decompose the difference into

$$\check{\mathsf{C}}_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}')-\mathsf{C}_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}')=\frac{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}')}{\sqrt{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})\check{\mathsf{V}}_{\mu,\nu}(y',\mathbf{x}')}}-\frac{\mathsf{V}_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}')}{\sqrt{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})\mathsf{V}_{\mu,\nu}(y',\mathbf{x}')}}$$

$$=\underbrace{\frac{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}')-\mathsf{V}_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}')}{\sqrt{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})\check{\mathsf{V}}_{\mu,\nu}(y',\mathbf{x}')}}}_{\text{(I)}}+\underbrace{\mathsf{C}_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}')\left(\sqrt{\frac{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})\mathsf{V}_{\mu,\nu}(y',\mathbf{x}')}{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})\check{\mathsf{V}}_{\mu,\nu}(y',\mathbf{x}')}}-1\right)}_{\text{(II)}}.$$

The probabilistic order of the second term is given in Lemma SA-3.2.

Using similar techniques as in the proof of Lemma SA-2.1 or SA-3.2, it is also straightforward to verify that term (I) has the same order. That is,

$$\mathbb{P}\left[\sup_{y,y'\in\mathcal{Y},\mathbf{x},\mathbf{x}'\in\mathcal{X}}|\text{(I)}|>\mathfrak{c}_1\mathtt{r}_{\mathtt{VE}}\right]\leq\mathfrak{c}_2 n^{-\mathfrak{c}_3}.$$

SA-7.17 Proof of Lemma SA-3.8

Consider an ε discretization of $\mathcal{Y}\times\mathcal{X}$, which is denoted by $\mathcal{A}_\varepsilon=\{(y_\ell,\mathbf{x}_\ell^{\mathrm{T}})^{\mathrm{T}}:\ 1\leq\ell\leq L\}$. Then one can define two Gaussian vectors, $\mathbf{z},\check{\mathbf{z}}\in\mathbb{R}^L$, such that

$$\mathbb{Cov}[z_\ell,z_{\ell'}]=\mathsf{C}(y_\ell,\mathbf{x}_\ell,y_{\ell'},\mathbf{x}_{\ell'}),\quad\mathbb{Cov}[\check{z}_\ell,\check{z}_{\ell'}|\text{Data}]=\check{\mathsf{C}}(y_\ell,\mathbf{x}_\ell,y_{\ell'},\mathbf{x}_{\ell'}).$$

Then we apply the Gaussian comparison result in Lemma SA-7.2 and the error rate in Lemma SA-3.7, which lead to

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}\left[\sup_{1\leq\ell\leq L}|\check{z}_\ell|\leq u\Big|\mathbf{Y},\mathbf{X}\right]-\mathbb{P}\left[\sup_{1\leq\ell\leq L}|z_\ell|\leq u\right]\right|$$

$$=\sup_{u\in\mathbb{R}}\left|\mathbb{P}\left[\sup_{1\leq\ell\leq L}|\check{\mathfrak{G}}_{\mu,\nu}(y_\ell,\mathbf{x}_\ell)|\leq u\Big|\mathbf{Y},\mathbf{X}\right]-\mathbb{P}\left[\sup_{1\leq\ell\leq L}|\mathfrak{G}_{\mu,\nu}(y_\ell,\mathbf{x}_\ell)|\leq u\right]\right|\precsim_{\mathbb{P}}\left[h^{\frac{\mathfrak{q}}{2}}+\left(\frac{\log n}{nh^{d+2}}\right)^{\frac{1}{4}}\right]\log\frac{1}{\varepsilon}.$$

Since ε only enters the above error bound logarithmically, one can choose $\varepsilon=n^{-c}$ for some c large enough, so that the error that arises from discretization becomes negligible. The same applies to $\hat{\mathfrak{G}}_{\mu,\nu}(y_\ell,\mathbf{x}_\ell)$.

SA-7.18 Proof of Theorem SA-3.2

First consider $\check{\mathbb{T}}^\circ_{\mu,\nu}(y,\mathbf{x})$. Since

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})|-\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{T}}^\circ_{\mu,\nu}(y,\mathbf{x})-\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})|\leq\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{T}}^\circ_{\mu,\nu}(y,\mathbf{x})|$$

$$\leq\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})|+\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{T}}^\circ_{\mu,\nu}(y,\mathbf{x})-\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})|,$$

then with Lemma SA-3.3,

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})|\le u-\mathfrak{c}_1\mathbf{r}_{\mathrm{SE}}\right]-\mathfrak{c}_2 n^{-\mathfrak{c}_3}\le\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{T}}^{\circ}_{\mu,\nu}(y,\mathbf{x})|\le u\right]$$
$$\le\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})|\le u+\mathfrak{c}_1\mathbf{r}_{\mathrm{SE}}\right]+\mathfrak{c}_2 n^{-\mathfrak{c}_3}.$$

In the above, we also used the fact that the difference $\check{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x})-\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})$ is negligible compared to $\mathbf{r}_{\mathrm{SE}}$ (see Remark SA-2.3).

By applying Lemma SA-3.1,

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\mathbb{G}_{\mu,\nu}(y,\mathbf{x})|\le u-\mathfrak{c}_1(\mathbf{r}_{\mathrm{SE}}+\mathbf{r}_{\mathrm{SA}})\right]-\mathfrak{c}_2 n^{-\mathfrak{c}_3}\le\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{T}}^{\circ}_{\mu,\nu}(y,\mathbf{x})|\le u\right]$$
$$\le\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\mathbb{G}_{\mu,\nu}(y,\mathbf{x})|\le u+\mathfrak{c}_1(\mathbf{r}_{\mathrm{SE}}+\mathbf{r}_{\mathrm{SA}})\right]+\mathfrak{c}_2 n^{-\mathfrak{c}_3}.$$

Finally, we apply the Gaussian comparison result in Lemma SA-3.8, which implies that

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\mathbb{G}_{\mu,\nu}(y,\mathbf{x})|\le u-\mathfrak{c}_1(\mathbf{r}_{\mathrm{SE}}+\mathbf{r}_{\mathrm{SA}})\right]-\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\mathbb{G}_{\mu,\nu}(y,\mathbf{x})|\le u\right]-\mathfrak{c}_2 n^{-\mathfrak{c}_3}-(\log n)\sqrt{\mathbf{r}_{\mathrm{VE}}}$$
$$\precsim_{\mathbb{P}}\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{T}}^{\circ}_{\mu,\nu}(y,\mathbf{x})|\le u\right]-\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{G}}_{\mu,\nu}(y,\mathbf{x})|\le u\bigg|\mathbf{Y},\mathbf{X}\right]$$
$$\precsim_{\mathbb{P}}\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\mathbb{G}_{\mu,\nu}(y,\mathbf{x})|\le u+\mathfrak{c}_1(\mathbf{r}_{\mathrm{SE}}+\mathbf{r}_{\mathrm{SA}})\right]-\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\mathbb{G}_{\mu,\nu}(y,\mathbf{x})|\le u\right]+\mathfrak{c}_2 n^{-\mathfrak{c}_3}+(\log n)\sqrt{\mathbf{r}_{\mathrm{VE}}}.$$

As a result,

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{T}}^{\circ}_{\mu,\nu}(y,\mathbf{x})|\le u\right]-\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{G}}_{\mu,\nu}y,\mathbf{x}|\le u\bigg|\mathbf{Y},\mathbf{X}\right]\right|$$
$$\precsim_{\mathbb{P}}\mathfrak{c}_2 n^{-\mathfrak{c}_3}+(\log n)\sqrt{\mathbf{r}_{\mathrm{VE}}}+\sup_{u\in\mathbb{R}}\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\mathbb{G}_{\mu,\nu}(y,\mathbf{x})|\in[u,u+\mathfrak{c}_1(\mathbf{r}_{\mathrm{SE}}+\mathbf{r}_{\mathrm{SA}})]\right].$$

Finally, due to Lemma SA-7.5, we have

$$\sup_{u\in\mathbb{R}}\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\mathbb{G}_{\mu,\nu}(y,\mathbf{x})|\in[u,u+\mathfrak{c}_1(\mathbf{r}_{\mathrm{SE}}+\mathbf{r}_{\mathrm{SA}})]\right]\precsim\sqrt{\log n}(\mathbf{r}_{\mathrm{SE}}+\mathbf{r}_{\mathrm{SA}}).$$

SA-7.19   Proof of Theorem SA-4.1

Note that $\theta_{\mu,\nu}(y,\mathbf{x})$ falls into the confidence band $\check{\mathcal{C}}_{\mu,\nu}(1-\alpha)$ if and only if

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{\check{\theta}_{\mu,\nu}(y,\mathbf{x})-\theta_{\mu,\nu}(y,\mathbf{x})}{\sqrt{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})}}\right|\le\check{\mathsf{cv}}_{\mu,\nu}(\alpha).$$

A sufficient condition would then be

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\check{\mathbb{T}}^{\circ}_{\mu,\nu}(y,\mathbf{x})\right|+\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{\mathbb{E}[\check{\theta}_{\mu,\nu}(y,\mathbf{x})|\mathbf{X}]-\theta_{\mu,\nu}(y,\mathbf{x})}{\sqrt{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})}}\right|\le\check{\mathsf{cv}}_{\mu,\nu}(\alpha).$$

The conclusion then follows from Theorem SA-3.2 and the bias calculation in Lemma SA-2.2. The same analysis applies to $\hat{\mathcal{C}}_{\mu,\nu}(1-\alpha)$.

SA-7.20   Proof of Theorem SA-4.2

To start, we decompose the test statistic into

$$\check{\mathbb{T}}_{\mathrm{PS}}(y, \mathbf{x}) = \check{\mathbb{T}}_{\mu,\boldsymbol{\nu}}^{\circ}(y, \mathbf{x}) + \frac{\mathbb{E}[\check{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})] - \theta_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{\sqrt{\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}} + \frac{\theta_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) - \theta_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}; \hat{\boldsymbol{\gamma}})}{\sqrt{\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}}.$$

Then by the leading bias order in Lemma SA-2.2 and the leading variance order in Lemma SA-2.3, we have that

$$\mathbb{P}\left[\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} \left| \frac{\mathbb{E}[\check{\theta}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})] - \theta_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{\sqrt{\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}} \right| > \mathfrak{c}_1 \frac{\mathbf{r}_{\mathrm{B}}}{\mathbf{r}_{\mathrm{V}}} \left(1 + \mathbf{r}_{\mathrm{VE}}\right) \right] \leq \mathfrak{c}_2 n^{-\mathfrak{c}_3}.$$

Similarly, under the null hypothesis,

$$\mathbb{P}\left[\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} \left| \frac{\theta_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) - \theta_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}; \hat{\boldsymbol{\gamma}})}{\sqrt{\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}} \right| > \mathfrak{c}_1 \frac{\mathbf{r}_{\mathrm{PS}}}{\mathbf{r}_{\mathrm{V}}} \left(1 + \mathbf{r}_{\mathrm{VE}}\right) \right] \leq \mathfrak{c}_2 n^{-\mathfrak{c}_3}.$$

Then we have the following error bound

$$\sup_{u \in \mathbb{R}} \left| \mathbb{P}\left[\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} |\check{\mathbb{T}}_{\mathrm{PS}}(y, \mathbf{x})| \leq u \right] - \mathbb{P}\left[\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} |\check{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})| \leq u \,\middle|\, \mathbf{Y}, \mathbf{X} \right] \right|$$

$$\precsim_{\mathbb{P}} \sqrt{\log n} \left(\mathbf{r}_{\mathrm{SE}} + \mathbf{r}_{\mathrm{SA}} + \frac{\mathbf{r}_{\mathrm{B}} + \mathbf{r}_{\mathrm{PS}}}{\mathbf{r}_{\mathrm{V}}} \right) + (\log n)\sqrt{\mathbf{r}_{\mathrm{VE}}}.$$

As a result,

$$\mathbb{P}\left[\sup_{y \in \mathcal{Y}, \mathbf{x} \in \mathcal{X}} |\check{\mathbb{T}}_{\mathrm{PS}}(y, \mathbf{x})| > \mathbf{cv}_{\mu,\boldsymbol{\nu}}(\alpha) \right] \leq \alpha + \mathfrak{c}\left(\sqrt{\log n} \left(\mathbf{r}_{\mathrm{SE}} + \mathbf{r}_{\mathrm{SA}} + \frac{\mathbf{r}_{\mathrm{B}} + \mathbf{r}_{\mathrm{PS}}}{\mathbf{r}_{\mathrm{V}}} \right) + (\log n)\sqrt{\mathbf{r}_{\mathrm{VE}}} \right).$$

The same strategy can be employed to establish results for $\hat{\mathbb{T}}_{\mathrm{PS}}(y, \mathbf{x})$. ∎

SA-7.21   Proof of Theorem SA-4.3

The conclusion follows directly from Theorem SA-4.1. ∎

References

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014). "Gaussian Approximation of Suprema of Empirical Processes," *Annals of Statistics*, *42*(4), 1564–1597.

Chernozhukov, V., Chetverikov, D., Kato, K., and Koike, Y. (2019). "Improved Central Limit Theorem and Bootstrap Approximations in High Dimensions," arXiv:1912.10529.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, New York: Chapman & Hall/CRC.

Giné, E., Latała, R., and Zinn, J. (2000). "Exponential and Moment Inequalities for U-statistics," In *High Dimensional Probability II*: Springer.

de la Peña, V. H. and Montgomery-Smith, S. J. (1995). "Decoupling Inequalities for the Tail Probabilities of Multivariate U-statistics," *Annals of Probability*, *23*(2), 806–816.

Rio, E. (1994). "Local Invariance Principles and Their Application to Density Estimation," *Probability Theory and Related Fields*, *98*(1), 21–45.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*: Springer.