Bootstrap-Assisted Inference for Generalized Grenander-type Estimators*
Supplemental Appendix

Matias D. Cattaneo† Michael Jansson‡ Kenichi Nagasawa§

March 23, 2023

*Cattaneo gratefully acknowledges financial support from the National Science Foundation through grant SES-1947805, and Jansson gratefully acknowledges financial support from the National Science Foundation through grant SES-1947662.

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

‡Department of Economics, University of California at Berkeley.

§Department of Economics, University of Warwick.

Contents

SA.1 Generalized Grenander-type estimators **2**
- SA.1.1 Asymptotic distribution . . . 2
- SA.1.2 Continuity of the limiting distribution . . . 3
- SA.1.3 Bootstrap approximation . . . 3

SA.2 Implementation **4**
- SA.2.1 Mean function estimation . . . 4
- SA.2.2 Bootstrap . . . 5

SA.3 Bootstrap inconsistency **8**

SA.4 Examples **8**
- SA.4.1 Monotone density function . . . 8
- SA.4.2 Monotone regression function . . . 10
- SA.4.3 Monotone hazard function . . . 12
- SA.4.4 Distribution function estimation with current status data . . . 13

SA.5 Proofs **14**
- SA.5.1 Switch relations . . . 14
- SA.5.2 Proof of Lemma SA-1 . . . 16
- SA.5.3 Proof of Theorem SA-1 . . . 17
- SA.5.4 Proof of Theorem SA-2 . . . 20
- SA.5.5 Proof of Lemma SA-2 . . . 21
- SA.5.6 Proof of Lemma SA-3 . . . 23
- SA.5.7 Proof of Theorem SA-3 . . . 27

SA.6 Verifying conditions in examples **28**
- SA.6.1 Monotone density function . . . 29
- SA.6.2 Monotone regression function . . . 33
- SA.6.3 Monotone hazard function . . . 36
- SA.6.4 Distribution function estimation with current status data . . . 37

SA.7 Rule-of-Thumb Step Size Selection **37**

SA.1 Generalized Grenander-type estimators

As discussed in the main paper, the parameter of interest $\theta_0(\mathsf{x})$ is characterized by $\theta_0(\mathsf{x}) = \partial_{-}\mathsf{GCM}_{[0,u_0]}(\Gamma_0 \circ \Phi_0^{-}) \circ \Phi_0(\mathsf{x})$ where ∂_{-} denotes left-differentiation operator, $\mathsf{GCM}_J(\cdot)$ is the greatest convex minorant operator over an interval J, and Γ_0, Φ_0 are some real-valued functions. Let $\widehat{\Phi}_n, \widehat{\Gamma}_n$ be estimators of Γ_0, Φ_0, and $\widehat{u}_n$ be a sample counterpart of u_0. Then, a generalized Grenander-type estimator takes the form

$$\widehat{\theta}_n(\mathsf{x}) = \partial_{-}\mathsf{GCM}_{[0,\widehat{u}_n]}\big(\widehat{\Gamma}_n \circ \widehat{\Phi}_n^{-}\big) \circ \widehat{\Phi}_n(\mathsf{x}).$$

Below we analyze the limiting distribution of $\widehat{\theta}_n(\mathsf{x})$ and its bootstrap-based distributional approximation.

SA.1.1 Asymptotic distribution

Let $\mathfrak{q} = \min\{j \in \mathbb{N} : \partial^j \theta_0(\mathsf{x}) \neq 0\}$ be the characteristic exponent of $\theta_0(\mathsf{x})$ and $\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) = \frac{\partial^{\mathfrak{q}}\theta_0(\mathsf{x})\partial\Phi_0(\mathsf{x})}{(\mathfrak{q}+1)!} v^{\mathfrak{q}+1}$.

Assumption SA-1. *For some $\delta > 0$, $\mathfrak{s} \geq 1$, and $\mathfrak{q} \in \mathbb{N}$,*

(1) *$I \subseteq \mathbb{R}$ is an interval and $I_{\mathsf{x}}^{\delta} := \{x : |x - \mathsf{x}| \leq \delta\} \subseteq I$.*

(2) *$\theta_0 : I \to \mathbb{R}$ is non-decreasing. In addition, $\theta_0(x)$ is $\lfloor \mathfrak{s} \rfloor$-times continuously differentiable on I_{x}^{δ} with*

$$\sup_{x \neq x' \in I_{\mathsf{x}}^{\delta}} \frac{|\partial^{\lfloor \mathfrak{s} \rfloor}\theta_0(x) - \partial^{\lfloor \mathfrak{s} \rfloor}\theta_0(x')|}{|x - x'|^{\mathfrak{s} - \lfloor \mathfrak{s} \rfloor}} < \infty.$$

Also, $\mathfrak{q} \leq \lfloor \mathfrak{s} \rfloor$.

(3) *$\Phi_0 : I \to [0, u_0]$ is non-decreasing and right-continuous. In addition, $\Phi_0(x)$ is $(\lfloor \mathfrak{s} \rfloor - \mathfrak{q} + 1)$-times continuously differentiable on I_{x}^{δ} with $\partial\Phi_0(\mathsf{x}) \neq 0$ and*

$$\sup_{x \neq x' \in I_{\mathsf{x}}^{\delta}} \frac{|\partial^{\lfloor \mathfrak{s} \rfloor - \mathfrak{q} + 1}\theta_0(x) - \partial^{\lfloor \mathfrak{s} \rfloor - \mathfrak{q} + 1}\theta_0(x')|}{|x - x'|^{\mathfrak{s} - \lfloor \mathfrak{s} \rfloor}} < \infty.$$

Define $a_n = n^{1/(2\mathfrak{q}+1)}$ and

$$\begin{aligned}\widehat{G}_{\mathsf{x},n}^{\mathfrak{q}}(v) = {} & \sqrt{na_n}[\widehat{\Gamma}_n(\mathsf{x} + va_n^{-1}) - \widehat{\Gamma}_n(\mathsf{x}) - \Gamma_0(\mathsf{x} + va_n^{-1}) + \Gamma_0(\mathsf{x})] \\ & - \theta_0(\mathsf{x})\sqrt{na_n}[\widehat{\Phi}_n(\mathsf{x} + va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x}) - \Phi_0(\mathsf{x} + va_n^{-1}) + \Phi_0(\mathsf{x})].\end{aligned}$$

Assumption SA-2.

(1) *$\widehat{G}_{\mathsf{x},n}^{\mathfrak{q}} \rightsquigarrow \mathcal{G}_{\mathsf{x}}$ where $\mathcal{G}_{\mathsf{x}}$ is a mean-zero Gaussian process with covariance kernel $\mathcal{C}_{\mathsf{x}}$.*

(2) $\sup_{x \in I} |\widehat{\Gamma}_n(x) - \Gamma_0(x)| = o_{\mathbb{P}}(1)$.

(3) *$\widehat{\Phi}_n$ is non-decreasing and right-continuous. For any $K > 0$,*

$$a_n \sup_{|v| \leq K} \big|\widehat{\Phi}_n(\mathsf{x} + va_n^{-1}) - \Phi_0(\mathsf{x} + va_n^{-1})\big| = o_{\mathbb{P}}(1).$$

Also, $\sup_{x \in I} |\widehat{\Phi}_n(x) - \Phi_0(x)| = o_{\mathbb{P}}(1)$.

(4) *For every* $s, t \in \mathbb{R}$,

$$\mathcal{C}_{\mathsf{x}}(s+t, s+t) - \mathcal{C}_{\mathsf{x}}(s+t, s) - \mathcal{C}_{\mathsf{x}}(s, s+t) + \mathcal{C}_{\mathsf{x}}(s, s) = \mathcal{C}_{\mathsf{x}}(t, t)$$

and

$$\mathcal{C}_{\mathsf{x}}(s\tau, t\tau) = \mathcal{C}_{\mathsf{x}}(s, t)\tau \qquad \textit{for every } \tau \geq 0.$$

In addition, $\mathcal{C}_{\mathsf{x}}(1,1) > 0$ *and* $\lim_{\delta \downarrow 0} \mathcal{C}_{\mathsf{x}}(1, \delta)/\sqrt{\delta} = 0$.

(5) *For some* $u_0 > \Phi_0(\mathsf{x})$, $\widehat{u}_n \geq u_0 + o_{\mathbb{P}}(1)$.
Also, $\{0, \widehat{u}_n\} \subseteq \widehat{\Phi}_n(I)$, *and* $\widehat{\Phi}_n^-([0, \widehat{u}_n])$ *and* $\widehat{\Phi}_n(I) \cap [0, \widehat{u}_n]$ *are closed.*

Theorem SA-1. *Under Assumptions SA-1 and SA-2,*

$$r_n(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})) \rightsquigarrow (\partial\Phi_0(\mathsf{x}))^{-1} \partial_- \mathsf{GCM}_{\mathbb{R}}\{\mathcal{G}_{\mathsf{x}} + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}\}(0).$$

where $r_n = n^{\frac{\mathfrak{q}}{2\mathfrak{q}+1}}$.

SA.1.2 Continuity of the limiting distribution

The proof of Thereom SA-1 relies on the continuity of the distribution function

$$x \mapsto \mathbb{P}\Big[\operatorname*{arg\,max}_{v \in \mathbb{R}} \Big\{\mathcal{G}_{\mathsf{x}}(v) - \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) + zv\Big\} \leq x\Big]$$

at $x = 0$ for each $z \in \mathbb{R}$. To show the continuity of the distribution function, we present a general lemma. Let $\{\mathbb{G}(s) : s \in \mathbb{R}\}$ be a Gaussian process with mean function μ and covariance kernel $\mathcal{K}$.

Lemma SA-1. *The Gaussian process* $\{\mathbb{G}(s) : s \in \mathbb{R}\}$ *has continuous sample paths, for every* $\tau > 0$, $s, t \in \mathbb{R}$, $\mathcal{K}(s\tau, t\tau) = \mathcal{K}(s, t)\tau$ *and* $\mathcal{K}(s+t, s+t) - \mathcal{K}(s+t, s) - \mathcal{K}(s, s+t) + \mathcal{K}(s, s) = \mathcal{K}(t, t)$, $\mathcal{K}(1,1) > 0$, *and* $\lim_{\delta \downarrow 0} \mathcal{K}(1, \delta)/\sqrt{\delta} = 0$. *Also,* $\limsup_{|s| \to \infty} \mu(s)/|s|^c = -\infty$ *for some* $c > 1$ *and* $\lim_{\tau \downarrow} |\mu(x + \tau) - \mu(x)|/\sqrt{\tau} = 0$ *for each* $x \in \mathbb{R}$. *Then, a unique maximizer of* $\mathbb{G}(s)$ *exists with probability one, and the distribution function*

$$x \mapsto \mathbb{P}\Big[\operatorname*{arg\,max}_{s \in \mathbb{R}} \mathbb{G}(s) \leq x\Big]$$

is continuous.

SA.1.3 Bootstrap approximation

Let $(\widehat{\Gamma}_n^*(x), \widehat{\Phi}_n^*(x), \widehat{u}_n^*)$ be the bootstrap version of $(\widehat{\Gamma}_n(x), \widehat{\Phi}_n(x), \widehat{u}_n)$, and

$$\widetilde{\Gamma}_n^*(x) = \widehat{\Gamma}_n^*(x) - \widehat{\Gamma}_n(x) + \widehat{\theta}_n(\mathsf{x})\widehat{\Phi}_n(x) + \widetilde{M}_{\mathsf{x},n}(x - \mathsf{x})$$

where $\widetilde{M}_{\mathsf{x},n}(v)$ is an estimator of $\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}$ as discussed in the main paper. Our bootstrap distributional approximation is based on the estimator

$$\widetilde{\theta}_n(\mathsf{x}) = \partial_- \mathsf{GCM}_{[0, \widehat{u}_n^*]}\big(\widetilde{\Gamma}_n^* \circ (\widehat{\Gamma}_n^*)^-\big) \circ \widehat{\Gamma}_n^*(\mathsf{x}).$$

Define

$$\widehat{G}^{\mathfrak{q},*}_{\mathsf{x},n}(v) = \sqrt{na_n}[\widehat{\Gamma}^*_n(\mathsf{x}+va_n^{-1}) - \widehat{\Gamma}^*_n(\mathsf{x}) - \widehat{\Gamma}_n(\mathsf{x}+va_n^{-1}) + \widehat{\Gamma}_n(\mathsf{x})]$$
$$- \widehat{\theta}_n(\mathsf{x})\sqrt{na_n}[\widehat{\Phi}^*_n(\mathsf{x}+va_n^{-1}) - \widehat{\Phi}^*_n(\mathsf{x}) - \widehat{\Phi}_n(\mathsf{x}+va_n^{-1}) + \widehat{\Phi}_n(\mathsf{x})].$$

We impose the following conditions to analyze the bootstrap procedure.

Assumption SA-3.

(1) $\widehat{G}^{\mathfrak{q},*}_{n,\mathsf{x}} \rightsquigarrow_{\mathbb{P}} \mathcal{G}_{\mathsf{x}}$.

(2) $\sup_{x\in I}|\widehat{\Gamma}^*_n(x) - \widehat{\Gamma}_n(x)| = o_{\mathbb{P}}(1)$.

(3) $\widehat{\Phi}^*_n$ *is non-decreasing and right-continuous. For any* $K>0$,

$$a_n \sup_{|v|\leq K}\left|\widehat{\Phi}^*_n(\mathsf{x}+va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x}+va_n^{-1})\right| = o_{\mathbb{P}}(1).$$

Also, $\sup_{x\in I}|\widehat{\Phi}^*_n(x) - \widehat{\Phi}_n(x)| = o_{\mathbb{P}}(1)$.

(4) For the same $\widehat{u}_n$ *in Assumption SA-2,* $\widehat{u}^*_n \geq \widehat{u}_n + o_{\mathbb{P}}(1)$.

Also, $\{0,\widehat{u}^*_n\} \subseteq \widehat{\Phi}^*_n(I)$, *and* $(\widehat{\Phi}^*_n)^-([0,\widehat{u}^*_n])$ *and* $\widehat{\Phi}^*_n(I)\cap[0,\widehat{u}^*_n]$ *are closed.*

Define

$$\widetilde{M}^{\mathfrak{q}}_{\mathsf{x},n}(v) = \sqrt{na_n}\widetilde{M}_{\mathsf{x},n}(va_n^{-1}).$$

Assumption SA-4. $\widetilde{M}^{\mathfrak{q}}_{\mathsf{x},n} \rightsquigarrow_{\mathbb{P}} \mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}$ *and for every* $K>0$,

$$\liminf_{\delta\downarrow 0}\liminf_{n\to\infty}\mathbb{P}\Big[\inf_{|v|>K^{-1}}\widetilde{M}_{\mathsf{x},n}(v)\geq\delta\Big] = 1 \qquad \textit{for every } K>0.$$

Theorem SA-2. *Suppose Assumptions SA-1, SA-2, SA-3, and SA-4 hold. Then,*

$$r_n(\widetilde{\theta}^*_n(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x})) \rightsquigarrow_{\mathbb{P}} (\partial\Phi_0(\mathsf{x}))^{-1}\partial_-\mathsf{GCM}_{\mathbb{R}}\{\mathcal{G}_{\mathsf{x}} + \mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}\}(0).$$

where $r_n = n^{\frac{\mathfrak{q}}{2\mathfrak{q}+1}}$.

By Lemma SA-1, the limit distribution is continuous, and thus, we have

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}^*_n\Big[\widetilde{\theta}^*_n(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x})\leq t\Big] - \mathbb{P}\Big[\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})\leq t\Big]\right| = o_{\mathbb{P}}(1)$$

where $\mathbb{P}^*_n$ is the bootstrap probability measure conditional on the data.

SA.2 Implementation

SA.2.1 Mean function estimation

Here we consider a construction of $\widetilde{M}^{\mathfrak{q}}_{\mathsf{x},n}$ and proivde a set of sufficient conditions implying Assumption SA-4.

For $j = 1,\ldots,\lfloor\mathfrak{s}\rfloor$, let

$$\mathcal{D}_j(\mathsf{x}) = \frac{\partial^{j+1}\Upsilon_0(\mathsf{x})}{(j+1)!}, \qquad \Upsilon_0(x) = \Gamma_0(x) - \theta_0(\mathsf{x})\Phi_0(x).$$

Under Assumption SA-1, $\mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}(v)=\mathcal{D}_{\mathfrak{q}}(\mathsf{x})v^{\mathfrak{q}+1}$.

In the main paper, we considered the following estimators of $\mathcal{D}_j(\mathsf{x})$:

$$\widetilde{\mathcal{D}}^{\mathtt{MA}}_{j,n}(\mathsf{x})=\epsilon_n^{-(j+1)}[\widehat{\Upsilon}_n(\mathsf{x}+\epsilon_n)-\widehat{\Upsilon}_n(\mathsf{x})],$$

$$\widetilde{\mathcal{D}}^{\mathtt{FD}}_{j,n}(\mathsf{x})=\epsilon_n^{-(j+1)}\sum_{k=1}^{j+1}(-1)^{k+j+1}\binom{j+1}{k}[\widehat{\Upsilon}_n(\mathsf{x}+k\epsilon_n)-\widehat{\Upsilon}_n(\mathsf{x})],$$

$$\widetilde{\mathcal{D}}^{\mathtt{BR}}_{j,n}(\mathsf{x})=\epsilon_n^{-(j+1)}\sum_{k=1}^{\underline{\mathfrak{s}}+1}\lambda^{\mathtt{BR}}_j(k)[\widehat{\Upsilon}_n(\mathsf{x}+c_k\epsilon_n)-\widehat{\Upsilon}_n(\mathsf{x})],$$

where $\epsilon_n>0$ is a vanishing sequence of tuning parameters, $\widehat{\Upsilon}_n(x)=\widehat{\Gamma}_n(x)-\widehat{\theta}_n(\mathsf{x})\widehat{\Phi}_n(x)$, the integer $\underline{\mathfrak{s}}$ is chosen by a researcher and assumed to satisfy $\underline{\mathfrak{s}}\leq\mathfrak{s}$, $\{c_k:1\leq k\leq\underline{\mathfrak{s}}+1\}$ are user chosen constants, and the scalars $\{\lambda^{\mathtt{BR}}_j(k):1\leq k\leq\underline{\mathfrak{s}}+1\}$ are defined by the property

$$\sum_{k=1}^{\underline{\mathfrak{s}}+1}\lambda^{\mathtt{BR}}_j(k)c_k^p=\mathbb{1}\{p=j+1\},\qquad p=1,\ldots,\underline{\mathfrak{s}}+1.$$

To analyze properties of the above numerical derivative estimators, for $\delta>0$, define

$$\begin{aligned}\widehat{G}_{\mathsf{x},n}(v;\delta)=&\sqrt{n\delta^{-1}}\big[\widehat{\Gamma}_n(\mathsf{x}+v\delta)-\widehat{\Gamma}_n(\mathsf{x})-\Gamma_0(\mathsf{x}+v\delta)+\Gamma_0(\mathsf{x})\big]\\&-\theta_0(\mathsf{x})\sqrt{n\delta^{-1}}\big[\widehat{\Phi}_n(\mathsf{x}+v\delta)-\widehat{\Phi}_n(\mathsf{x})-\Phi_0(\mathsf{x}+v\delta)+\Phi_0(\mathsf{x})\big],\end{aligned}$$

and

$$\widehat{R}_{\mathsf{x},n}(v;\delta)=\delta^{-1}\big[\widehat{\Phi}_n(\mathsf{x}+v\delta)-\widehat{\Phi}_n(\mathsf{x})-\Phi_0(\mathsf{x}+v\delta)+\Phi_0(\mathsf{x})\big].$$

Assumption SA-5. *For the same $\mathfrak{q}$ as in Assumption SA-1 and for every $\delta_n>0$ with $\delta_n=o(1)$ and $a_n^{-1}\delta_n^{-1}=O(1)$,*

$$\widehat{G}_{\mathsf{x},n}(1;\delta_n)=O_{\mathbb{P}}(1)\qquad\text{and}\qquad\widehat{R}_{\mathsf{x},n}(1;\delta_n)=o_{\mathbb{P}}(1).$$

Lemma SA-2. *Suppose Assumptions SA-1 and SA-5 are satisfied and that $r_n(\widehat{\theta}_n(\mathsf{x})-\theta_0(\mathsf{x}))=O_{\mathbb{P}}(1)$. If $\epsilon_n=o(1)$ and $a_n^{-1}\epsilon_n^{-1}=o(1)$, then*

$$\widetilde{\mathcal{D}}_{\mathfrak{q},n}(\mathsf{x})\to_{\mathbb{P}}\mathcal{D}_{\mathfrak{q}}(\mathsf{x}),\qquad\widetilde{\mathcal{D}}_{\mathfrak{q},n}\in\{\widetilde{\mathcal{D}}^{\mathtt{MA}}_{\mathfrak{q},n},\widetilde{\mathcal{D}}^{\mathtt{FD}}_{\mathfrak{q},n},\widetilde{\mathcal{D}}^{\mathtt{BR}}_{\mathfrak{q},n}\}$$

and

$$a_n^{\mathfrak{q}-j}\big(\widetilde{\mathcal{D}}^{\mathtt{BR}}_{j,n}-\mathcal{D}_j(\mathsf{x})\big)=O\big(a_n^{\mathfrak{q}-j}\epsilon_n^{\min(\underline{\mathfrak{s}}+1,\mathfrak{s})-j}\big)+o_{\mathbb{P}}(1),\qquad j=1,\ldots,\mathfrak{s}.$$

In particular, if $3\leq\bar{\mathfrak{q}}<\mathfrak{s}$, then

$$a_n^{\mathfrak{q}-(2\ell-1)}\big(\widetilde{\mathcal{D}}^{\mathtt{BR}}_{j,n}-\mathcal{D}_j(\mathsf{x})\big)=o_{\mathbb{P}}(1),\qquad\ell=1,\ldots,\lfloor(\bar{\mathfrak{q}}+1)/2\rfloor,\tag{SA.1}$$

holds provided that $n\epsilon_n^{1+2\bar{\mathfrak{q}}\min(\underline{\mathfrak{s}},\mathfrak{s}-1)/(\bar{\mathfrak{q}}-1)}\to0$ and $n\epsilon_n^{1+2\bar{\mathfrak{q}}}\to\infty$.

SA.2.2 Bootstrap

In the main paper, we considered how to construct bootstrap estimators $(\widehat{\Gamma}^*_n,\widehat{\Phi}^*_n)$ and provided primitive conditions that can be used to verify Assumption SA-3. Specifically, we assumed that the non-bootstrap estimators admit large-sample approximations

$$\widehat{\Gamma}_n(x) \approx \bar{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\gamma_0(x;\mathbf{Z}_i) \qquad \text{and} \qquad \widehat{\Phi}_n(x) \approx \bar{\Phi}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\phi_0(x;\mathbf{Z}_i).$$

Denoting estimators of γ_0 and ϕ_0 by $\widehat{\gamma}_n$ and $\widehat{\phi}_n$, we consider the bootstrap estimators

$$\widehat{\Gamma}^*_n(x) = \frac{1}{n}\sum_{i=1}^{n}W_{i,n}\widehat{\gamma}_n(x;\mathbf{Z}_i) \qquad \text{and} \qquad \widehat{\Phi}^*_n(x) = \frac{1}{n}\sum_{i=1}^{n}W_{i,n}\widehat{\phi}_n(x;\mathbf{Z}_i),$$

where $W_{1,n},\ldots,W_{n,n}$ denote exchangeable random variables, independent of $(\mathbf{Z}_1,\ldots,\mathbf{Z}_n,\widehat{\gamma}_n,\widehat{\phi}_n)$.

Let

$$\psi_{\mathsf{x}}(v;\mathbf{z}) = \gamma_0(\mathsf{x}+v;\mathbf{z}) - \gamma_0(\mathsf{x};\mathbf{z}) - \theta_0(\mathsf{x})[\phi_0(\mathsf{x}+v;\mathbf{z}) - \phi_0(\mathsf{x};\mathbf{z})],$$

and

$$\bar{\Gamma}^*_n(x) = \frac{1}{n}\sum_{i=1}^{n}W_{i,n}\gamma_0(x;\mathbf{Z}_i) \qquad \text{and} \qquad \bar{\Phi}^*_n(x) = \frac{1}{n}\sum_{i=1}^{n}W_{i,n}\phi_0(x;\mathbf{Z}_i).$$

For any class of functions $\mathfrak{F}$, let $N_U(\varepsilon,\mathfrak{F})$ denote the associated uniform covering number relative to L_2; that is, for any $\varepsilon > 0$, let

$$N_U(\varepsilon,\mathfrak{F}) = \sup_Q N(\varepsilon\|\bar{F}\|_{Q,2},\mathfrak{F},L_2(Q)),$$

where $\bar{F}$ is the minimal envelope function of $\mathfrak{F}$, $\|\cdot\|_{Q,2}$ is the $L_2(Q)$ norm, $N(\cdot)$ is the covering number, and the supermum is taken over every discrete probability measure Q with $\|\bar{F}\|_{Q,2} > 0$.

Assumption SA-6. *For the same $\mathfrak{q}$ as in Assumption SA-1, the following are satisfied:*

(1) $\mathbf{Z}_1,\mathbf{Z}_2,\ldots$ *are independent and identifcally distributed.*

(2) *For each $n\in\mathbb{N}$, $W_{1,n},\ldots,W_{n,n}$ are exchangeable random variables independent of $\mathbf{Z}_1,\ldots,\mathbf{Z}_n,\widehat{\gamma}_n,\widehat{\phi}_n$. In addition, for some $\mathfrak{r} > (4\mathfrak{q}+2)/(2\mathfrak{q}-1)$,*

$$\frac{1}{n}\sum_{i=1}^{n}W_{i,n} = 1, \qquad \frac{1}{n}\sum_{i=1}^{n}(W_{i,n}-1)^2 \to_{\mathbb{P}} 1, \qquad \textit{and} \qquad \mathbb{E}[|W_{1,n}|^{\mathfrak{r}}] = O(1).$$

(3) $\sup_{x\in I}|\widehat{\Gamma}_n(x) - \bar{\Gamma}_n(x)| = o_{\mathbb{P}}(1)$ *and* $\frac{1}{n}\sum_{i=1}^{n}\sup_{x\in I}|\widehat{\gamma}_n(x;\mathbf{Z}_i) - \gamma_0(x;\mathbf{Z}_i)|^2 = o_{\mathbb{P}}(1)$. *For every $K > 0$,*

$$\sqrt{na_n}\sup_{|v|\le K}\left|\widehat{\Gamma}_n(\mathsf{x}+va_n^{-1}) - \widehat{\Gamma}_n(\mathsf{x}) - \bar{\Gamma}_n(\mathsf{x}+va_n^{-1}) + \bar{\Gamma}_n(\mathsf{x})\right| = o_{\mathbb{P}}(1),$$

and

$$a_n\frac{1}{n}\sum_{i=1}^{n}\sup_{|v|\le K}\left|\widehat{\gamma}_n(\mathsf{x}+va_n^{-1};\mathbf{Z}_i) - \widehat{\gamma}_n(\mathsf{x};\mathbf{Z}_i) - \gamma_0(\mathsf{x}+va_n^{-1};\mathbf{Z}_i) + \gamma_0(\mathsf{x};\mathbf{Z}_i)\right|^2 = o_{\mathbb{P}}(1).$$

In addition, for some $V\in(0,2)$,

$$\limsup_{\varepsilon\downarrow 0}\frac{\log N_U(\varepsilon,\mathfrak{F}_\gamma)}{\varepsilon^{-V}} < \infty, \qquad \mathbb{E}[\bar{F}_\gamma(\mathbf{Z})^2] < \infty, \qquad \limsup_{\varepsilon\downarrow 0}\frac{\log N_U(\varepsilon,\hat{\mathfrak{F}}_{\gamma,n})}{\varepsilon^{-V}} = O_{\mathbb{P}}(1),$$

where $\mathfrak{F}_\gamma = \{\gamma_0(x;\cdot) : x \in I\}$, $\bar{F}_\gamma$ *is its minimal envelope, and* $\hat{\mathfrak{F}}_{\gamma,n} = \{\widehat{\gamma}_n(x;\cdot) : x \in I\}$*. Also,*

$$\limsup_{\delta\downarrow 0} \frac{\mathbb{E}[\bar{D}^\delta_\gamma(\mathbf{Z})^2 + \bar{D}^\delta_\gamma(\mathbf{Z})^4]}{\delta} < \infty,$$

where $\bar{D}^\delta_\gamma$ *is the minimal envelope of* $\{\gamma_0(x;\cdot) - \gamma_0(\mathsf{x};\cdot) : x \in I^\delta_{\mathsf{x}}\}$*.*

(4) $\widehat{\Phi}_n, \widehat{\Phi}^*_n$ *are non-decreasing and right-continuous on* I*.* $\sup_{x\in I} |\widehat{\Phi}_n(x) - \bar{\Phi}_n(x)| = o_{\mathbb{P}}(1)$, $\frac{1}{n}\sum_{i=1}^n \sup_{x\in I} |\widehat{\phi}_n(x;\mathbf{Z}_i) - \phi_0(x;\mathbf{Z}_i)|^2 = o_{\mathbb{P}}(1)$*, and* $a_n|\widehat{\Phi}_n(\mathsf{x}) - \bar{\Phi}_n(\mathsf{x})| = o_{\mathbb{P}}(1)$*. For every* $K > 0$*,*

$$\sqrt{na_n} \sup_{|v|\le K} \left|\widehat{\Phi}_n(\mathsf{x} + va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x}) - \bar{\Phi}_n(\mathsf{x} + va_n^{-1}) + \bar{\Phi}_n(\mathsf{x})\right| = o_{\mathbb{P}}(1),$$

and

$$a_n \frac{1}{n}\sum_{i=1}^n \sup_{|v|\le K} \left|\widehat{\phi}_n(\mathsf{x} + va_n^{-1};\mathbf{Z}_i) - \widehat{\phi}_n(\mathsf{x};\mathbf{Z}_i) - \phi_0(\mathsf{x} + va_n^{-1};\mathbf{Z}_i) + \phi_0(\mathsf{x};\mathbf{Z}_i)\right|^2 = o_{\mathbb{P}}(1).$$

In addition, for some $V \in (0,2)$*,*

$$\limsup_{\varepsilon\downarrow 0} \frac{\log N_U(\varepsilon, \mathfrak{F}_\phi)}{\varepsilon^{-V}} < \infty, \qquad \mathbb{E}[\bar{F}_\phi(\mathbf{Z})^2] < \infty, \qquad \limsup_{\varepsilon\downarrow 0} \frac{\log N_U(\varepsilon, \hat{\mathfrak{F}}_{\phi,n})}{\varepsilon^{-V}} = O_{\mathbb{P}}(1),$$

where $\mathfrak{F}_\phi = \{\phi_0(x;\cdot) : x \in I\}$, $\bar{F}_\phi$ *is its minimal envelope, and* $\hat{\mathfrak{F}}_{\phi,n} = \{\widehat{\phi}_n(x;\cdot) : x \in I\}$*. Also,*

$$\limsup_{\delta\downarrow 0} \frac{\mathbb{E}[\bar{D}^\delta_\phi(\mathbf{Z})^2 + \bar{D}^\delta_\phi(\mathbf{Z})^4]}{\delta} < \infty,$$

where $\bar{D}^\delta_\phi$ *is the minimal envelope of* $\{\phi_0(x;\cdot) - \phi_0(\mathsf{x};\cdot) : x \in I^\delta_{\mathsf{x}}\}$*.*

(5) *For every* $\delta_n > 0$ *with* $a_n\delta_n = O(1)$*,*

$$\sup_{v\neq v'\in[-\delta_n,\delta_n]} \frac{\mathbb{E}[|\psi_{\mathsf{x}}(v;\mathbf{Z}) - \psi_{\mathsf{x}}(v';\mathbf{Z})|]}{|v - v'|} = O(1)$$

and for all $s,t \in \mathbb{R}$*, and for some* $\mathcal{C}_{\mathsf{x}}$*,*

$$\frac{\mathbb{E}[\psi_{\mathsf{x}}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}}(t\delta_n;\mathbf{Z})]}{\delta_n} \to \mathcal{C}_{\mathsf{x}}(s,t).$$

Lemma SA-3. *Suppose Assumptions SA-1 and SA-6 are satisfied. Then, Assumption SA-2 (1)-(3) and Assumption SA-3 (1)-(3) are satisfied. If also*

$$\sqrt{n\delta_n^{-1}}\left[\widehat{\Gamma}_n(\mathsf{x} + \delta_n) - \widehat{\Gamma}_n(\mathsf{x}) - \bar{\Gamma}_n(\mathsf{x} + \delta_n) + \bar{\Gamma}_n(\mathsf{x})\right] = O_{\mathbb{P}}(1)$$

and

$$\sqrt{n\delta_n^{-1}}\left[\widehat{\Phi}_n(\mathsf{x} + \delta_n) - \widehat{\Phi}_n(\mathsf{x}) - \bar{\Phi}_n(\mathsf{x} + \delta_n) + \bar{\Phi}_n(\mathsf{x})\right] = O_{\mathbb{P}}(1)$$

for every $\delta_n > 0$ *with* $\delta_n = o(1)$ *and* $a_n^{-1}\delta_n^{-1} = O(1)$*, then Assumption SA-5 is satisfied.*

SA.3 Bootstrap inconsistency

In this section, we formally show the inconsistency of bootstrap distribution approximations. Consider the "naïve" bootstrap estimator

$$\widehat{\theta}_n^*(\mathsf{x}) = \big(\partial_-\mathsf{GCM}_{[0,\widehat{u}_n^*]}\big(\widehat{\Gamma}_n^* \circ (\widehat{\Phi}_n^*)^-\big)\big) \circ \widehat{\Phi}_n^*(\mathsf{x})$$

where

$$\widehat{\Gamma}_n^*(x) = \frac{1}{n}\sum_{i=1}^n W_{i,n}\widehat{\gamma}_n(x;\mathbf{Z}_i), \qquad \widehat{\Phi}_n^*(x) = \frac{1}{n}\sum_{i=1}^n W_{i,n}\widehat{\phi}_n(x;\mathbf{Z}_i).$$

Theorem SA-3. *Suppose Assumptions SA-1, SA-2, and SA-3 hold. Then,* $n^{\frac{\mathfrak{q}}{2\mathfrak{q}+1}}(\widehat{\theta}_n^*(\mathsf{x})-\widehat{\theta}_n(\mathsf{x})) \not\rightsquigarrow_{\mathbb{P}} (\partial\Phi_0(\mathsf{x}))^{-1}\partial_-\mathsf{GCM}_{\mathbb{R}}(\mathcal{G}_{\mathsf{x}}+\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}})(0)$ *i.e., the bootstrap approximation fails.*

This theorem implies the well-known result of the bootstrap inconsistency for Grenander estimator (e.g., Kosorok, 2008; Sen et al., 2010). Our result accommodates exchangeable bootstrap schemes and a wide class of generalized Grenander-type estimators.

SA.4 Examples

Let $\mathbf{Z}_i = (Y_i, \check{X}_i, \Delta_i, \mathbf{A}_i)'$, $i = 1, 2, \ldots, n$, be an observed random sample with $\check{X}_i = \min\{X_i, C_i\}$, $\Delta_i = \mathbb{1}(X_i \le C_i)$, and $\mathbf{A}_i$ denoting additional covariates. Assuming that $F_0(x) = \mathbb{P}[X_i \le x]$ is absolutely continuous on I, a Lebesgue density function is denoted by $f_0(x) = \partial F_0(x)$. If $\mathbb{P}[C_i \ge X_i] = 1$, then there is no (right) censoring and $\check{X}_i = X_i$. Other basic quantities of interest are the survival function $S_0(x) = \mathbb{P}[X_i > x]$ and the mean function $\mu_0(X_i) = \mathbb{E}[Y_i|X_i]$, as well as their conditional on $\mathbf{A}_i$ analogues $S_0(x|\mathbf{A}_i) = \mathbb{P}[X_i > x|\mathbf{A}_i]$ and $\mu_0(X_i, \mathbf{A}_i) = \mathbb{E}[Y_i|X_i, \mathbf{A}_i]$. $\mathfrak{q}$ is as defined in Assumption SA-1. Recall that we write $\partial^\ell g(x)$ for the ℓ-th derivative of a smooth function g and we use the convention $\partial^0 g(x) = g(x)$.

The examples below consider monotone estimation of $f_0(\mathsf{x})$, $\mu_0(\mathsf{x})$, $f_0(\mathsf{x})/S_0(\mathsf{x})$ and $F_0(\mathsf{x})$, under various assumptions related to censoring and covariate-adjustment. For each example, we provide a set of primitive conditions that imply Assumptions SA-1, SA-2, SA-3, and SA-5. For brevity, we only describe the covariance kernel and the mean function of the limiting Gaussian process, and do not repeat the conclusions of Theorems SA-1 and SA-2.

SA.4.1 Monotone density function

In this example, the parameter of interest is the density of X at a point x i.e., $\theta_0(\mathsf{x}) = f_0(\mathsf{x})$. We have $\Gamma_0(x) = F_0(x)$ and $\Phi_0(x) = x$. We take $\widehat{\Phi}_n(x) = \widehat{\Phi}_n^*(x) = x$. Also, for simplicity, we take u_0 as given.

SA.4.1.1 No censoring

First, consider the canonical case of no censoring: $\mathbb{P}[C_i \ge u_0] = 1$. The classical Grenander (1956) estimator sets $\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i \le x)$. The exchangeable bootstrap analogue is $\widehat{\Gamma}_n^*(x) =$

$\frac{1}{n}\sum_{i=1}^{n} W_{i,n}\mathbb{1}(X_i \leq x)$.

To analyze this example, we impose the following conditions.

Assumption SA.4.1.1

(1) The Lebesgue density f_0 of X is non-decreasing on $I = [0, u_0]$ and x is in the interior of I.

(2) The density f_0 satisfies Assumption SA-1 (2).

Under this assumption, the limit distribution of the Grenander estimator is characterized by

$$\mathcal{C}_{\mathsf{x}}(s,t) = f_0(\mathsf{x})\min\{|s|,|t|\}\mathbb{1}\{\text{sign}(s) = \text{sign}(t)\}, \quad \mathcal{D}_{\mathfrak{q}}(\mathsf{x}) = \frac{\partial^{\mathfrak{q}} f_0(\mathsf{x})}{(1+\mathfrak{q})!}.$$

SA.4.1.2 Independent right-censoring

Next, suppose that censoring occurs completely at random: $X_i \perp\!\!\!\perp C_i$. Huang and Wellner (1995) analyzed this example and related problems. In this case, we take $\widehat{\Gamma}_n(x) = 1 - \widehat{S}_n(x)$, where $\widehat{S}_n$ denotes an estimator of the survival function $S_0(x)$. For concreteness, let $\widehat{S}_n$ be the Kaplan-Meier estimator. For bootstrap, one possibility is to use the non-parametric bootstrap to resample the Kaplan-Meier estimator. Another approach is to use our framework in Section SA.2.2. For this purpose, let

$$\widehat{\gamma}_n(x;\mathbf{Z}) = \widehat{F}_n(x) + \widehat{S}_n(x)\left[\frac{\Delta\mathbb{1}(\check{X} \leq x)}{\widehat{S}_n(\check{X})\widehat{G}_n(\check{X})} - \int_0^{\check{X}\wedge x}\frac{d\widehat{\Lambda}_n(u)}{\widehat{S}_n(u)\widehat{G}_n(u)}\right]$$

where $\widehat{F}_n = 1 - \widehat{S}_n$, $\widehat{G}_n$ is the Kaplan Meier estimator for G_0, and $\widehat{\Lambda}_n$ is the cumulative hazard function associated with $\widehat{S}_n$ i.e., $\widehat{\Lambda}_n(x) = \int_0^x \widehat{S}_n(u)^{-1}d\widehat{F}_n(u)$. Then, the bootstrap objective function is $\widehat{\Gamma}_n^* = \sum_{i=1}^n W_{i,n}\widehat{\gamma}_n(x;\mathbf{Z}_i)$.

To analyze this example, we impose the following conditions.

Assumption SA.4.1.2

(1) The Lebesgue density f_0 of X is non-decreasing on $I = [0, u_0]$ and x is in the interior of I.

(2) $X \perp\!\!\!\perp C$.

(3) The density f_0 satisfies Assumption SA-1 (2). $G_0(c) = \mathbb{P}[C_i > c]$ is continuous on I, and $S_0(u_0)G_0(u_0) > 0$.

The last condition imposes that we set the interval I to be a strict subset of the support of X. The covariance kernel and the mean function in this setting have the form

$$\mathcal{C}_{\mathsf{x}}(s,t) = \frac{f_0(\mathsf{x})}{G_0(\mathsf{x})}\min\{|s|,|t|\}\mathbb{1}\{\text{sign}(x) = \text{sign}(t)\}, \quad \mathcal{D}_{\mathfrak{q}}(\mathsf{x}) = \frac{\partial^{\mathfrak{q}} f_0(\mathsf{x})}{(1+\mathfrak{q})!}.$$

SA.4.1.3 Conditionally independent right-censoring

We consider the case of censoring at random: $X_i \perp\!\!\!\perp C_i|\mathbf{A}_i$. See van der Laan and Robins (2003); Zeng (2004) and references therein for existing analysis of this problem. We set

$$\widehat{\gamma}_n(x;\mathbf{Z}) = \widehat{F}_n(x|\mathbf{A}) + \widehat{S}_n(x|\mathbf{A})\Big[\frac{\Delta\mathbb{1}(\check{X}\le x)}{\widehat{S}_n(\check{X}|\mathbf{A})\widehat{G}_n(\check{X}|\mathbf{A})} - \int_0^{\check{X}\wedge x}\frac{d\widehat{\Lambda}_n(u|\mathbf{A})}{\widehat{S}_n(u|\mathbf{A})\widehat{G}_n(u|\mathbf{A})}\Big],$$

where $\widehat{F}_n = 1-\widehat{S}_n$, $\widehat{S}_n(x|\mathbf{A}), \widehat{G}_n(c|\mathbf{A})$ denote preliminary estimates of the conditional survival functions $S_0(x|\mathbf{A}) = \mathbb{P}[X > x|\mathbf{A}]$, $G_0(c|\mathbf{A}) = \mathbb{P}[C > c|\mathbf{A}]$, respectively, and $\widehat{\Lambda}_n(x|\mathbf{A}) = \int_0^x \frac{\widehat{F}_n(du|\mathbf{A})}{\widehat{S}_n(u|\mathbf{A})}$. Using this function, we define $\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^n \widehat{\gamma}_n(x;\mathbf{Z}_i)$ and $\widehat{\Gamma}_n^*(x) = \frac{1}{n}\sum_{i=1}^n W_{i,n}\widehat{\gamma}_n(x;\mathbf{Z}_i)$. Notice that we employ the original first-step estimates $\widehat{S}_n, \widehat{G}_n, \widehat{\Lambda}_n$ in the bootstrap objective function, so it is not necessary to bootstrap these preliminary estimators.

Assumption SA.4.1.3 Let $\mathfrak{S}_n$, $\mathfrak{G}_n$ be sequences of function classes that contain $S_0(\cdot|\cdot)$, $G_0(\cdot|\cdot)$, respectively.

(1) The Lebesgue density f_0 of X is non-decreasing on $I = [0, u_0]$ and x is in the interior of I.
(2) $X \perp\!\!\!\perp C|A$ and the density f_0 satisfies Assumption SA-1 (2).
(3) For each $S \in \mathfrak{S}_n$, $x \mapsto S(x|\mathbf{A})$ is non-increasing almost surely, and $\{S(x|\cdot) : x \in I\}$ is a VC-subgraph class with the VC index bounded by a fixed constant. For all $S \in \mathfrak{S}_n$, $G \in \mathfrak{G}_n$, $0 < c \le S, G \le C < \infty$.
(4) For $K > 0$, $\sqrt{na_n}\sup_{|v|\le K}|\widehat{\Gamma}_n(\mathsf{x}+va_n^{-1}) - \widehat{\Gamma}_n(\mathsf{x}) - \Gamma_0(\mathsf{x}+va_n^{-1}) + \Gamma_0(\mathsf{x})| = o_{\mathbb{P}}(1)$.
(5) With probability approaching one, $\widehat{S}_n \in \mathfrak{S}_n$ and $\widehat{G}_n \in \mathfrak{G}_n$. $a_n\frac{1}{n}\sum_{i=1}^n \sup_{x\in I}|\widehat{S}_n(x|\mathbf{A}_i) - S_0(x|\mathbf{A}_i)|^2 = o_{\mathbb{P}}(1)$, and $a_n\frac{1}{n}\sum_{i=1}^n \sup_{x\in I}|\widehat{G}_n(x|\mathbf{A}_i) - G_0(x|\mathbf{A}_i)|^2 = o_{\mathbb{P}}(1)$.
(6) The conditional distribution of X given $\mathbf{A}$ has bounded Lebesgue density $f_{X|A}$, $\mathbb{E}[\frac{f_{X|A}(\mathsf{x}|\mathbf{A})}{G_0(\mathsf{x}|\mathbf{A})}] > 0$, and there are real-valued functions B, ω such that $\mathbb{E}[B(\mathbf{A})] < \infty$, $\lim_{\delta\downarrow 0}\omega(\delta) = 0$, and for $|x-\mathsf{x}|$ sufficiently small, $|\frac{f_{X|A}(x|\mathbf{A})}{S_0(x|\mathbf{A})G_0(x|\mathbf{A})} - \frac{f_{X|A}(\mathsf{x}|\mathbf{A})}{S_0(\mathsf{x}|\mathbf{A})G_0(\mathsf{x}|\mathbf{A})}| \le \omega(|x-\mathsf{x}|)B(\mathbf{A})$.

The condition (4) is high-level, and there are a few different approaches to verify them. See Westling and Carone (2020) for details. The covariance kernel and the mean function are

$$\mathcal{C}_{\mathsf{x}}(s,t) = \mathbb{E}\Big[\frac{f_{X|A}(\mathsf{x}|\mathbf{A})}{G_0(\mathsf{x}|\mathbf{A})}\Big]\min\{|s|,|t|\}\mathbb{1}\{\operatorname{sign}(s) = \operatorname{sign}(t)\}, \quad \mathcal{D}_{\mathfrak{q}}(\mathsf{x}) = \frac{\partial^{\mathfrak{q}} f_0(\mathsf{x})}{(1+\mathfrak{q})!}.$$

SA.4.2 Monotone regression function

The parameter of interest is the conditional mean function $\theta_0(\mathsf{x}) = \mu_0(\mathsf{x})$ for the classical case and $\theta_0(\mathsf{x}) = \mathbb{E}[\mu_0(\mathsf{x},\mathbf{A})]$ if covariates $\mathbf{A}$ are available. There is no censoring in this example. If $X \perp\!\!\!\perp \mathbf{A}$, the two objects coincide, but they are not the same in general. We have $\Gamma_0(x) = \mathbb{E}[Y\mathbb{1}\{X \le x\}]$, $\Phi_0(x) = F_0(x)$, I is the support of X, and $u_0 = 1$.

SA.4.2.1 Classical case

First consider the case without covariates. The classical isotonic regression estimator of Ayer et al. (1955) sets $\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^n Y_i \mathbb{1}\{X_i \le x\}$ and $\widehat{\Phi}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_i \le x\}$. The bootstrap analogue is $\widehat{\Gamma}_n^*(x) = \frac{1}{n}\sum_{i=1}^n W_{i,n} Y_i \mathbb{1}\{X_i \le x\}$ and $\widehat{\Phi}_n(x) = \frac{1}{n}\sum_{i=1}^n W_{i,n}\mathbb{1}\{X_i \le x\}$.

Assumption SA.4.2.1 Let $\varepsilon = Y - \mathbb{E}[Y|X]$.

(1) The regression function μ_0 is non-decreasing on I and x is in the interior of I.

(2) μ_0 satisfies Assumption SA-1 (2), and the cdf F_0 satisfies Assumption SA-1 (3).

(3) $\mathbb{E}[Y^2] < \infty$, $\sup_{|x-\mathsf{x}|\le\eta}\mathbb{E}[\varepsilon^4|X=x] < \infty$ for some $\eta > 0$, and $\sigma_0^2(x) = \mathbb{E}[\varepsilon^2|X=x]$ is continuous and positive at x.

The covariance kernel and the mean function are

$$\mathcal{C}_{\mathsf{x}}(s,t) = f_0(\mathsf{x})\sigma_0^2(\mathsf{x})\min\{|s|,|t|\}\mathbb{1}\{\operatorname{sign}(s) = \operatorname{sign}(t)\}, \qquad \mathcal{D}_{\mathfrak{q}}(\mathsf{x}) = \frac{f_0(\mathsf{x})\partial^{\mathfrak{q}}\mu_0(\mathsf{x})}{(1+\mathfrak{q})!}.$$

SA.4.2.2 With covariates

Now we consider a setting with covariates $\mathbf{A}$. A leading application of this framework in causal inference is discussed in Westling et al. (2020). The parameter of interest is $\theta_0(\mathsf{x}) = \mathbb{E}[\mathbb{E}[Y|X = \mathsf{x}, \mathbf{A}]]$. For $\widehat{\Phi}_n$, we set $\phi_0(x;\mathbf{Z}) = \mathbb{1}\{X \le x\}$, and for $\widehat{\Gamma}_n$, given a random sample $\{Y_i, X_i, \mathbf{A}_i\}_{i=1}^n$, we use

$$\widehat{\gamma}_n(x;\mathbf{Z}) = \mathbb{1}\{X \le x\}\left[\frac{Y - \widehat{\mu}_n(X,\mathbf{A})}{\widehat{g}_n(X,\mathbf{A})} + \frac{1}{n}\sum_{j=1}^n \widehat{\mu}_n(X,\mathbf{A}_j)\right]$$

where $\widehat{\mu}_n(X,\mathbf{A}), \widehat{g}_n(X,\mathbf{A})$ are preliminary estimates of $\mu_0(X,\mathbf{A})$ and $g_0(X,\mathbf{A}) = \frac{f_{X|A}(X,\mathbf{A})}{f_0(X)}$, respectively, and $f_{X|A}$ is the conditional Lebesgue density of X given $\mathbf{A}$.

Assumption SA.4.2.2 Define $\varepsilon = Y - \mathbb{E}[Y|X,\mathbf{A}]$ and $\sigma_0^2(X,\mathbf{A}) = \mathbb{E}[\varepsilon^2|X,\mathbf{A}]$. Let $\eta > 0$ be some fixed number.

(1) I is compact, the mapping $x \mapsto \mathbb{E}[\mu_0(x,\mathbf{A})]$ is non-decreasing on I, and x is in the interior of I.

(2) θ_0 satisfies Assumption SA-1 (2). The conditional distribution of X given $\mathbf{A}$ has a bounded Lebesgue density $f_{X|A}$, and there is $c > 0$ such that $g_0(X,\mathbf{A}) = \frac{f_{X|A}(X,\mathbf{A})}{f_X(X)} \ge c$ with probability one. The cdf F_0 satisfies Assumption SA-1 (3).

(3) For $K > 0$, $\sqrt{na_n}\sup_{|v|\le K}|\widehat{\Gamma}_n(\mathsf{x}+va_n^{-1}) - \widehat{\Gamma}_n(\mathsf{x}) - \Gamma_0(\mathsf{x}+va_n^{-1}) + \Gamma_0(\mathsf{x})| = o_{\mathbb{P}}(1)$.

(4) $a_n\frac{1}{n}\sum_{i=1}^n|\widehat{\mu}_n(X_i,\mathbf{A}_i) - \mu_0(X_i,\mathbf{A}_i)|^2 = o_{\mathbb{P}}(1)$, $a_n\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n|\widehat{\mu}_n(X_i,\mathbf{A}_j) - \mu_0(X_i,\mathbf{A}_j)|^2 = o_{\mathbb{P}}(1)$, $a_n\frac{1}{n}\sum_{i=1}^n\varepsilon_i^2|\widehat{g}_n(X_i,\mathbf{A}_i) - g_0(X_i,\mathbf{A}_i)|^2 = o_{\mathbb{P}}(1)$.

(5) There exists a real-valued function $\bar{\mu}$ such that $|\mu_0(x_1,\mathbf{A}) - \mu_0(x_2,\mathbf{A})| \le |x_1 - x_2|\bar{\mu}(\mathbf{A})$ for $|x_1 - x_2| \le \eta$ and $\mathbb{E}[\bar{\mu}(\mathbf{A})^2] < \infty$.

(6) $\mathbb{E}[\varepsilon^2] < \infty$, $\sup_{|x-\mathsf{x}|\le\eta}\mathbb{E}[\varepsilon^4|X=x] < \infty$, $\mathbb{E}[\frac{\sigma_0^2(\mathsf{x},\mathbf{A})}{g_0(\mathsf{x},\mathbf{A})}] > 0$, and there are real-valued functions B,ω such that $\mathbb{E}[B(\mathbf{A})] < \infty$, $\lim_{\delta\downarrow 0}\omega(\delta) = 0$, and for $|x-\mathsf{x}| \le \eta$, $|\frac{\sigma_0^2(x,\mathbf{A})f_{X|A}(x|\mathbf{A})}{g_0(x,\mathbf{A})^2} - \frac{\sigma_0^2(\mathsf{x},\mathbf{A})f_{X|A}(\mathsf{x}|\mathbf{A})}{g_0(\mathsf{x},\mathbf{A})^2}| \le \omega(|x-\mathsf{x}|)B(\mathbf{A})$.

The condition (3) is high-level, and there are different possibilities to verify them. See Westling and Carone (2020); Westling et al. (2020) for details. The covariance kernel and the mean function are

$$\mathcal{C}_{\mathsf{x}}(s,t) = f_0(\mathsf{x})\mathbb{E}\left[\frac{\sigma_0^2(\mathsf{x},\mathbf{A})}{g_0(\mathsf{x},\mathbf{A})}\right]\min\{|s|,|t|\}\mathbb{1}\{\text{sign}(s)=\text{sign}(t)\}, \qquad \mathcal{D}_{\mathfrak{q}}(\mathsf{x}) = \frac{f_0(\mathsf{x})\partial^{\mathfrak{q}}\theta_0(\mathsf{x})}{(1+\mathfrak{q})!}.$$

SA.4.3 Monotone hazard function

The parameter of interest is the hazard function of X, $\theta_0(\mathsf{x}) = f_0(\mathsf{x})/S_0(\mathsf{x})$. As pointed out by Westling and Carone (2020), the function Γ_0 takes the form $\Gamma_0(x) = \int_0^x \frac{f_0(u)}{S_0(u)}\Phi_0(du)$, and by taking $\Phi_0(x) = \int_0^x S_0(u)du$, $\Gamma_0(x) = F_0(x)$. Since Γ_0 is identical to the monotone density case, we can leverage the analysis done for the monotone density in Section SA.4.1. The interval I equals $[0, u_0^{\mathtt{MD}}]$ where $u_0^{\mathtt{MD}}$ is u_0 in the monotone density example. The u_0 for the monotone hazard function estimation is $u_0 = \Phi_0(u_0^{\mathtt{MD}})$, and we can take $\widehat{u}_n = \widehat{\Phi}_n(u_0^{\mathtt{MD}})$.

SA.4.3.1 Independent right-censoring

Consider the case of completely random censoring i.e., $X \perp\!\!\!\perp C$. We take $\widehat{\Gamma}_n(x) = 1 - \widehat{S}_n(x)$, where $\widehat{S}_n$ is the Kaplan-Meier estimator, and $\widehat{\Phi}_n(x) = \int_0^x \widehat{S}_n(u)du$. Using the same $\widehat{\gamma}_n(x)$ function as in Section SA.4.1.2, the bootstrap analogues are defined by $\widehat{\Gamma}_n^*(x) = \frac{1}{n}\sum_{i=1}^n W_{i,n}\widehat{\gamma}_n(x;\mathbf{Z}_i)$ and $\widehat{\Phi}_n^*(x) = \int_0^x[1-\widehat{\Gamma}_n^*(u)]du = \frac{1}{n}\sum_{i=1}^n W_{i,n}\widehat{\phi}_n(x;\mathbf{Z}_i)$ where $\widehat{\phi}_n(x;\mathbf{Z}) = x - \int_0^x \widehat{\gamma}_n(u;\mathbf{Z})du$.

Assumption SA.4.3.1

(1) Assumption SA.4.1.2 holds.

(2) $\sqrt{na_n}\sup_{|v|\le K}|\widehat{\Phi}_n(\mathsf{x}+va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x}) - \bar{\Phi}_n(\mathsf{x}+va_n^{-1}) + \bar{\Phi}_n(\mathsf{x})| = o_{\mathbb{P}}(1)$.

The second condition is high-level, and similar to verifying the condition $\sqrt{na_n}\sup_{|v|\le K}|\widehat{\Gamma}_n(\mathsf{x}+va_n^{-1}) - \widehat{\Gamma}_n(\mathsf{x}) - \bar{\Gamma}_n(\mathsf{x}+va_n^{-1}) + \bar{\Gamma}_n(\mathsf{x})| = o_{\mathbb{P}}(1)$ in the monotone density case, specific structures of the estimators facilitates the analysis. Alternatively, one may assume $n^{\frac{\mathfrak{q}}{2\mathfrak{q}+1}}[\widehat{\Gamma}_n(\mathsf{x}) - \bar{\Gamma}_n(\mathsf{x})] = o_{\mathbb{P}}(1)$, which is similar to the condition assumed in Theorem 7 of Westling and Carone (2020).

The covariance kernel and the mean function in this example have the form

$$\mathcal{C}_{\mathsf{x}}(s,t) = \frac{f_0(\mathsf{x})}{G_0(\mathsf{x})}\min\{|s|,|t|\}\mathbb{1}\{\text{sign}(x)=\text{sign}(t)\}, \quad \mathcal{D}_{\mathfrak{q}}(\mathsf{x}) = \frac{S_0(\mathsf{x})\partial^{\mathfrak{q}}f_0(\mathsf{x})}{(1+\mathfrak{q})!}.$$

SA.4.3.2 Conditionally independent right-censoring

Now suppose $X \perp\!\!\!\perp C|\mathbf{A}$ i.e., conditionally independent censoring. Using the same $\widehat{\gamma}_n$ function as in Section SA.4.1.3, we set $\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^n \widehat{\gamma}_n(x;\mathbf{Z}_i)$ and $\widehat{\Phi}_n(x) = \int_0^x[1-\widehat{\Gamma}_n(u)]du = \frac{1}{n}\sum_{i=1}^n \widehat{\phi}_n(x;\mathbf{Z}_i)$ where $\widehat{\phi}_n(x;\mathbf{Z}) = x - \int_0^x \widehat{\gamma}_n(u;\mathbf{Z})du$. The bootstrap analogues are $\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^n W_{i,n}\widehat{\gamma}_n(x;\mathbf{Z}_i)$ and $\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^n W_{i,n}\widehat{\phi}_n(x;\mathbf{Z}_i)$.

Assumption SA.4.3.2

(1) Assumption SA.4.1.3 holds.

(2) $\sqrt{na_n}\sup_{|v|\le K}|\widehat{\Phi}_n(\mathsf{x}+va_n^{-1})-\widehat{\Phi}_n(\mathsf{x})-\bar{\Phi}_n(\mathsf{x}+va_n^{-1})+\bar{\Phi}_n(\mathsf{x})|=o_{\mathbb{P}}(1)$.

The covariance kernel and the mean function take the form

$$\mathcal{C}_{\mathsf{x}}(s,t)=\mathbb{E}\Big[\frac{f_{X|A}(\mathsf{x}|\mathbf{A})}{G_0(\mathsf{x}|\mathbf{A})}\Big]\min\{|s|,|t|\}\mathbb{1}\{\mathrm{sign}(s)=\mathrm{sign}(t)\},\quad \mathcal{D}_{\mathfrak{q}}(\mathsf{x})=\frac{S_0(\mathsf{x})\partial^{\mathfrak{q}}f_0(\mathsf{x})}{(1+\mathfrak{q})!}.$$

SA.4.4 Distribution function estimation with current status data

The parameter of interest is the cdf of X at x, $\theta_0(\mathsf{x})=F_0(\mathsf{x})$. In this example, we do not observe $\tilde{X}=X\wedge C$. Instead, the observation consists of $\mathbf{Z}_i=(\Delta_i,C_i,\mathbf{A}_i)$. This setup is often referred to as current status data. Let $H_0(x)=\mathbb{P}[C\le x]$ be the cdf of the censoring time C. We can use $\Gamma_0(x)=\int_0^x F_0(u)H_0(du)$ and $\Phi_0(x)=H_0(x)$. The interval I is the support of X and $u_0=1$. The structure of the estimation problem turns out to be identical to the one for the monotone regression example, and we can leverage the common structure.

SA.4.4.1 Independent right-censoring

First we consider the case of completely at random censoring $X\perp\!\!\!\perp C$. See Groeneboom and Wellner (1992) for existing analysis. We set $\gamma_0(x;\mathbf{Z})=\Delta\mathbb{1}\{C\le x\}$ and $\phi_0(x;\mathbf{Z})=\mathbb{1}\{C\le x\}$. Thus, $\widehat{\Gamma}_n,\widehat{\Phi}_n,\widehat{\Gamma}_n^*,\widehat{\Phi}_n^*$ are defined using γ_0 and ϕ_0. Note that if the notation is mapped by $(\Delta,C)\leftrightarrow(Y,X)$, then these functions are identical to those of the classical monotone regression problem (Section SA.4.2.1). Thus, the following assumptions are identical to Assumption SA.4.2.1 up to notation and some changes due to boundedness of Δ.

Assumption SA.4.4.1 Let $\varepsilon=\Delta-\mathbb{E}[\Delta|C]$.

(1) The distribution function F_0 is non-decreasing on I and x is in the interior of I.

(2) F_0 satisfies Assumption SA-1 (2), and the cdf H_0 satisfies Assumption SA-1 (3).

(3) $\sigma_0^2(x)=\mathbb{E}[\varepsilon^2|C=x]$ is continuous and positive at x.

By the second assumption, H_0 is $\lfloor\mathfrak{s}\rfloor-\mathfrak{q}+1$ times continuously differentiable on I_{x}^{δ}, and we write $h_0(x)=\partial H_0(x)$. The covariance kernel and the mean function are

$$\mathcal{C}_{\mathsf{x}}(s,t)=h_0(\mathsf{x})\sigma_0^2(\mathsf{x})\min\{|s|,|t|\}\mathbb{1}\{\mathrm{sign}(s)=\mathrm{sign}(t)\},\qquad \mathcal{D}_{\mathfrak{q}}(\mathsf{x})=\frac{h_0(\mathsf{x})\partial^{\mathfrak{q}}F_0(\mathsf{x})}{(1+\mathfrak{q})!}.$$

SA.4.4.2 Conditionally independent right-censoring

We consider the case where right-censoring is conditionally independent i.e., $X\perp\!\!\!\perp C|\mathbf{A}$. van der Vaart and van der Laan (2006) analyzed this example as well as settings with time-varying covariates. We are focusing on time-invariant covariates. Define $F_0(C,\mathbf{A})=\mathbb{E}[\Delta|C,\mathbf{A}]$ and $g_0(C,\mathbf{A})=$

$\frac{h_{C|A}(C|\mathbf{A})}{h_0(C)}$ where $h_{C|A}$ is the conditional density of C given $\mathbf{A}$ and h_0 is the marginal density of C. We set $\phi_0(x;\mathbf{Z}) = \mathbb{1}\{C \le x\}$ and

$$\widehat{\gamma}_n(x;\mathbf{Z}) = \mathbb{1}\{C \le x\}\left[\frac{\Delta - \widehat{F}_n(C,\mathbf{A})}{\widehat{g}_n(C,\mathbf{A})} + \frac{1}{n}\sum_{j=1}^{n}\widehat{F}_n(C,\mathbf{A}_j)\right]$$

where $\widehat{F}_n(c,\mathbf{a})$ and $\widehat{g}_n(c,\mathbf{a})$ are preliminary estimators for $F_0(c,\mathbf{a})$ and $g_0(c,\mathbf{a})$, respectively. Similarly to the censoring completely at random case, with the change in the notation (i.e., $(\Delta, C) \leftrightarrow (Y, X)$), the setup is identical to that of the monotone regression with covariates (Section SA.4.2.2). The following assumption is identical to Assumption SA.4.2.2 up to notation and some changes due to boundedness of Δ.

Assumption SA.4.4.2 Define $\varepsilon = \Delta - \mathbb{E}[\Delta|C,\mathbf{A}]$ and $\sigma_0^2(C,\mathbf{A}) = \mathbb{E}[\varepsilon^2|C,\mathbf{A}]$. Let $\eta > 0$ be some fixed number.

(1) I is compact, F_0 is non-decreasing on I, and x is in the interior of I.

(2) $\theta_0 = F_0$ satisfies Assumption SA-1 (2). The conditional distribution of C given $\mathbf{A}$ has a bounded Lebesgue density $h_{C|A}$, and there is $c > 0$ such that $g_0(C,\mathbf{A}) \ge c$ with probability one. $\Phi_0 = H_0$ satisfies Assumption SA-1 (3).

(3) For $K > 0$, $\sqrt{na_n}\sup_{|v|\le K}|\widehat{\Gamma}_n(\mathsf{x}+va_n^{-1}) - \widehat{\Gamma}_n(\mathsf{x}) - \Gamma_0(\mathsf{x}+va_n^{-1}) + \Gamma_0(\mathsf{x})| = o_{\mathbb{P}}(1)$.

(4) $a_n\frac{1}{n}\sum_{i=1}^{n}|\widehat{F}_n(C_i,\mathbf{A}_i) - F_0(C_i,\mathbf{A}_i)|^2 = o_{\mathbb{P}}(1)$, $a_n\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|\widehat{F}_n(C_i,\mathbf{A}_j) - F_0(C_i,\mathbf{A}_j)|^2 = o_{\mathbb{P}}(1)$, $a_n\frac{1}{n}\sum_{i=1}^{n}|\widehat{g}_n(C_i,\mathbf{A}_i) - g_0(C_i,\mathbf{A}_i)|^2 = o_{\mathbb{P}}(1)$.

(5) There exists a real-valued function $\bar{F}$ such that $|F_0(c_1,\mathbf{A}) - F_0(c_2,\mathbf{A})| \le |c_1 - c_2|\bar{F}(\mathbf{A})$ for $|c_1 - c_2| \le \eta$ and $\mathbb{E}[\bar{F}(\mathbf{A})^2] < \infty$.

(6) $\mathbb{E}[\frac{\sigma_0^2(\mathsf{x},\mathbf{A})}{g_0(\mathsf{x},\mathbf{A})}] > 0$, and there are real-valued functions B, ω such that $\mathbb{E}[B(\mathbf{A})] < \infty$, $\lim_{\delta\downarrow 0}\omega(\delta) = 0$, and for $|x - \mathsf{x}| \le \eta$, $|\frac{\sigma_0^2(x,\mathbf{A})h_{C|A}(x|\mathbf{A})}{g_0(x,\mathbf{A})^2} - \frac{\sigma_0^2(\mathsf{x},\mathbf{A})h_{C|A}(\mathsf{x}|\mathbf{A})}{g_0(\mathsf{x},\mathbf{A})^2}| \le \omega(|x - \mathsf{x}|)B(\mathbf{A})$.

The covariance kernel and the mean function are

$$\mathcal{C}_{\mathsf{x}}(s,t) = h_0(\mathsf{x})\mathbb{E}\left[\frac{\sigma_0^2(\mathsf{x},\mathbf{A})}{g_0(\mathsf{x},\mathbf{A})}\right]\min\{|s|,|t|\}\mathbb{1}\{\mathrm{sign}(s) = \mathrm{sign}(t)\}, \qquad \mathcal{D}_{\mathfrak{q}}(\mathsf{x}) = \frac{h_0(\mathsf{x})\partial^{\mathfrak{q}}F_0(\mathsf{x})}{(1+\mathfrak{q})!}.$$

SA.5 Proofs

SA.5.1 Switch relations

For a real-valued function f, let $\mathsf{lsc}_J(f)$ be the lower semi-continuous minorant on a set J. For a real-valued function defined on a metric space, Lemma 4.3 of van der Vaart and van der Laan (2006) states that $\mathsf{lsc}_J(f)(s) = \liminf_{t\to s: t\in J} f(t)$.

A key technical tool to analyze generalized Grenander-type estimators is the following switching relationship.

Lemma SA-4. *Let Γ and Φ be real-valued functions defined on an interval $I \subseteq \mathbb{R}$, where Φ is nondecreasing and right continuous. Fix $l < u$ with $l, u \in \Phi(I)$. Let $\psi = \partial_-\mathrm{GCM}_{[l,u]}(\Gamma \circ \Phi^-)$ and*

$\theta = \psi \circ \Phi$. *If $\Phi(I) \cap [l, u]$ and $\Phi^-([l, u])$ are closed, then, for any $t \in \mathbb{R}$ and $\mathsf{x} \in I$ with $\Phi(\mathsf{x}) \in (l, u)$,*

$$\theta(\mathsf{x}) > t \quad \iff \quad \sup \operatorname*{arg\,max}_{x \in I^\star} \big\{ t\Phi(x) - \mathsf{lsc}_{I^\star}(\Gamma)(x) \big\} < \Phi^-\big(\Phi(\mathsf{x})\big)$$

where $I^\star := \Phi^-([l, u]) = \{x \in I : \Phi(x) \in [l, u]\}$.

Proof. In the sequel, $\mathsf{lsc}(\Gamma \circ \Phi^-) \equiv \mathsf{lsc}_{[l,u]}(\Gamma \circ \Phi^-)$ and $\mathsf{lsc}(\Gamma) \equiv \mathsf{lsc}_{I^\star}(\Gamma)$.

Lemma 4.2 of van der Vaart and van der Laan (2006) states that $\mathsf{GCM}_{[l,u]}(\mathsf{lsc}(\Gamma \circ \Phi^-)) = \mathsf{GCM}_{[l,u]}(\Gamma \circ \Phi^-)$ on the interior of $[l, u]$. Since $\Phi(\mathsf{x})$ is in the interior of $[l, u]$, we have $\theta(\mathsf{x}) = \partial_- \mathsf{GCM}_{[l,u]}(\mathsf{lsc}(\Gamma \circ \Phi^-))(\Phi(\mathsf{x}))$.

Now, by upper semi-continuity of $-(\mathsf{lsc}(\Gamma \circ \Phi^-))$, Lemma 4.1 of van der Vaart and van der Laan (2006) implies

$$\theta(\mathsf{x}) > t \quad \iff \quad \sup \operatorname*{arg\,max}_{y \in [l,u]} \big\{ ty - \mathsf{lsc}(\Gamma \circ \Phi^-)(y) \big\} < \Phi(\mathsf{x}).$$

Thus, it suffices to show

$$\sup \operatorname*{arg\,max}_{x \in I^\star} \big\{ t\Phi(x) - \mathsf{lsc}(\Gamma)(x) \big\} < \Phi^-\big(\Phi(\mathsf{x})\big) \iff \sup \operatorname*{arg\,max}_{y \in [l,u]} \big\{ ty - \mathsf{lsc}(\Gamma \circ \Phi^-)(y) \big\} < \Phi(\mathsf{x}). \tag{SA.2}$$

For $x \in I^\star$, $\Gamma(x) = \Gamma(\Phi^-(y))$ with $y = \Phi(x)$, and $\liminf_{s \to x : s \in I^\star} \Gamma(s) \geq \liminf_{v \to y : v \in [l,u]} \Gamma(\Phi^-(v))$. Also, by closedness of $\Phi^-([l, u])$, either $\Phi^-(y+) := \lim_{\eta \downarrow 0} \Phi^-(y + \eta) = \Phi^-(y)$ or $\Phi(\Phi^-(y+)) > y$. Therefore, $\liminf_{v \to y : v \in [l,u]} \Gamma(\Phi^-(v)) \geq \liminf_{s \to x : s \in I^\star} \Gamma(s)$. Then, for $x \in I^\star$ and $y = \Phi(x)$,

$$t\Phi(x) - \mathsf{lsc}(\Gamma)(x) = t\Phi(x) - \liminf_{s \to x : s \in I^\star} \Gamma(s) = ty - \liminf_{v \to y : v \in [l,u]} \Gamma(\Phi^-(v)) = ty - \mathsf{lsc}(\Gamma \circ \Phi^-)(y). \tag{SA.3}$$

Let $y^\star = \sup \operatorname{arg\,max}_{y \in [l,u]} \{ty - \mathsf{lsc}(\Gamma \circ \Phi^-)(y)\}$. First we show $y^\star \in \Phi(I)$. The case $\Phi(I) \supseteq [l, u]$ is obvious, so assume $\Phi(I) \not\supset [l, u]$. For contradiction, suppose $y^\star \notin \Phi(I)$. Since $\Phi(I)$ is closed and $y^\star \notin \{l, u\}$, there exists $\eta > 0$ such that $[y^\star - \eta, y^\star + \eta] \subset [l, u]$ and $[y^\star - \eta, y^\star + \eta] \cap \Phi(I) = \emptyset$. On the interval $[y^\star - \eta, y^\star + \eta]$, Φ^- is constant, and thus, $\mathsf{lsc}(\Gamma \circ \Phi^-) = \Gamma \circ \Phi^-$ is constant on $[y^\star - \eta/2, y^\star + \eta/2]$, which contradicts the definition of $y^\star = \sup \operatorname{arg\,max}_{y \in [l,u]} \{ty - \mathsf{lsc}(\Gamma \circ \Phi^-)(y)\}$. Therefore, $y^\star \in \Phi(I)$.

Since $y^\star \in \Phi(I)$, $\Phi(\Phi^-(y^\star)) = y^\star$. By (SA.3), $t\Phi(\Phi^-(y^\star)) - \mathsf{lsc}(\Gamma)(\Phi^-(y^\star)) = ty^\star - \mathsf{lsc}(\Gamma \circ \Phi^-)(y^\star)$, and thus, $\Phi^-(y^\star) \in \operatorname{arg\,max}_{x \in I^\star} \{t\Phi(x) - \mathsf{lsc}(\Gamma)(x)\}$. Also, for $\hat{x} \in \operatorname{arg\,max}_{x \in I^\star} \{t\Phi(x) - \mathsf{lsc}(\Gamma)(x)\}$, $\hat{x} \leq \Phi^-(y^\star)$ as shown below. Then, $\Phi^-(y^\star) = \sup \operatorname{arg\,max}_{x \in I^\star} \{t\Phi(x) - \mathsf{lsc}(\Gamma)(x)\}$.

Now, the LHS of (SA.2) implying the RHS follows from the non-decreasing property of Φ^-. If $y^\star < u$, $y^\star \in \Phi(I)$ implies $\Phi^-(y^\star) < \Phi^-(y^\star + \epsilon)$ for any $\epsilon > 0$, and thus, the RHS of (SA.2) implies the LHS. If $y^\star = u$, (SA.2) follows from $u \in \Phi(I)$ and $\Phi(\mathsf{x}) < u$.

It remains to show $\hat{x} \leq \Phi^-(y^\star)$ for $\hat{x} \in \operatorname{arg\,max}_{x \in I^\star} \{t\Phi(x) - \mathsf{lsc}(\Gamma)(x)\}$. By (SA.3) and $\Phi^-(\Phi(\hat{x})) = \hat{x}$,

$$t\Phi(\hat{x}) - \mathsf{lsc}(\Gamma \circ \Phi^-)(\Phi(\hat{x})) = t\Phi(\hat{x}) - \mathsf{lsc}(\Gamma)(\hat{x}) = t\Phi(\Phi^-(y^\star)) - \mathsf{lsc}(\Gamma)(\Phi^-(y^\star))$$

where the second equalty uses $\Phi^-(y^\star)$ being a maximizer of $x \in I^\star \mapsto t\Phi(x) - \mathsf{lsc}(\Gamma)(x)$. Thus, $\Phi(\hat{x}) \in \operatorname{arg\,max}_{y \in [l,u]} \{ty - \mathsf{lsc}(\Gamma \circ \Phi^-)(y)\}$. Then, $\Phi(\hat{x}) \leq y^\star$ because $y^\star$ is the largest element. Now, if $\Phi(\hat{x}) < y^\star$, then $\hat{x} < \Phi^-(y^\star)$. If $\Phi(\hat{x}) = y^\star$, then $\Phi^-(y^\star) = \Phi^-(\Phi(\hat{x})) = \hat{x}$, and thus, $\hat{x} \leq \Phi^-(y^\star)$ holds. $\square$

For the sake of completeness, we state the following version of a switching lemma.

Lemma SA-5. *Let $\Gamma : \mathbb{R} \to \mathbb{R}$ be a bounded, lower semi-continuous function such that $vt - \Gamma(t)$ has a unique maximum and satisfies $\lim_{|v|\to\infty} \Gamma(v)/|v| = \infty$. Then, for any $x, t \in \mathbb{R}$,*

$$\partial_- \mathsf{GCM}_{\mathbb{R}}(\Gamma)(x) > t \qquad \Longleftrightarrow \qquad \operatorname*{arg\,max}_{v\in\mathbb{R}} \{vt - \Gamma(v)\} < x.$$

Proof. Lemma A.1. of Sen et al. (2010) implies that there exists $c > |x|$ such that $\mathsf{GCM}_{\mathbb{R}}(\Gamma) = \mathsf{GCM}_{[-M,M]}(\Gamma)$ for $M \geq c$. For fixed t, there exists some $K > c$ such that $-\Gamma(0) > vt - \Gamma(v)$ for $|v| > K$. Thus, the above display is equivalent to

$$\partial_- \mathsf{GCM}_{[-K,K]}(\Gamma)(x) > t \qquad \Longleftrightarrow \qquad \operatorname*{arg\,max}_{v\in[-K,K]} \{vt - \Gamma(v)\} < x$$

and this statement follows from Lemma 4.1 of van der Vaart and van der Laan (2006). $\square$

SA.5.2 Proof of Lemma SA-1

First we show that a maximizer of $\mathbb{G}(s)$ over $s \in \mathbb{R}$ exists and is unique with probability one. We follow the proof strategy of Kim and Pollard (1990). Let $\widetilde{\mathbb{G}}(s) = \mathbb{G}(s) - \mu(s)$ be the centered process. If we can show that for $c > 1$ in the hypothesis,

$$\mathbb{P}\Big[\limsup_{|s|\to\infty} \frac{\widetilde{\mathbb{G}}(s)}{|s|^c} > \eta\Big] = 0 \qquad \text{for any } \eta > 0, \tag{SA.4}$$

then $\mathbb{G}(s) \to -\infty$ as $|s| \to \infty$ with probability one by $\limsup_{|s|\to\infty} \mu(s)/|s|^c \to -\infty$. Then, by continuous sample paths, a maximizer $\mathbb{G}(s)$ exists. Since $\mathcal{K}(s,s)+\mathcal{K}(t,t)-2\mathcal{K}(s,t) = \mathcal{K}(s-t, s-t) > 0$ for $s \neq t$, Lemma 2.6 of Kim and Pollard (1990) implies that this maximizer is unique with probability one. It remains to show (SA.4). Using the property $\mathcal{K}(s\tau, t\tau) = \tau\mathcal{K}(s,t)$,

$$\begin{aligned}
\sum_{k=2}^{\infty} \mathbb{P}\Big[\sup_{k-1\leq|s|\leq k} \frac{\widetilde{\mathbb{G}}(s)}{|s|^c} > \eta\Big] &\leq \sum_{k=2}^{\infty} \mathbb{P}\Big[\sup_{|s|\leq k} \widetilde{\mathbb{G}}(s) > |k-1|^c \eta\Big] \\
&\leq \sum_{k=2}^{\infty} \mathbb{P}\Big[\sup_{|s|\leq 1} \widetilde{\mathbb{G}}(s) > \frac{|k-1|^c}{\sqrt{k}} \eta\Big] \\
&\leq \mathbb{E}\Big[\sup_{|s|\leq 1} \widetilde{\mathbb{G}}(s)^2\Big] \eta^{-2} \sum_{k=2}^{\infty} k^{1-2c} < \infty.
\end{aligned}$$

Then, the Borel-Cantelli lemma implies the desired result.

To show the continuity of the distribution function $x \mapsto \mathbb{P}[\arg\max_{s\in\mathbb{R}} \mathbb{G}(s) \leq x]$, it suffices to show $\mathbb{P}[\arg\max_{s\in\mathbb{R}} \mathbb{G}(s) = x] = 0$ for $x \in \mathbb{R}$. Fix $x \in \mathbb{R}$ and define

$$\tilde{Z}(s) = \frac{\mathbb{G}(s) - \mathbb{G}(x)}{\sqrt{\mathcal{K}(s,s) + \mathcal{K}(x,x) - 2\mathcal{K}(s,x)}}, \qquad s \neq x$$

and $\tilde{Z}(x) = 0$. Note $\mathcal{K}(s,s) + \mathcal{K}(x,x) - 2\mathcal{K}(s,x) = \mathcal{K}(s-x, s-x)$. Then, $\max_{s\in\mathbb{R}} \tilde{Z}(s) \geq 0$, and

$$\mathbb{P}\Big[\operatorname*{arg\,max}_{s\in\mathbb{R}} \mathbb{G}(s) = x\Big] = \mathbb{P}\Big[\max_{s\in\mathbb{R}} \tilde{Z}(s) \leq 0\Big]$$

and the last probability is bounded by $\mathbb{P}[\max_{s\in\mathcal{S}} \tilde{Z}(s) \leq 0]$ with any subset $\mathcal{S} \subset \mathbb{R}$. In the sequel, we construct a suitable subset $\mathcal{S}$ to show that the probability can be made arbitrarily small. In particular, for given $\varepsilon \in (0,1)$ and $N \in \mathbb{N}$, we pick N points $\{s^{\varepsilon}_{1,N}, \dots, s^{\varepsilon}_{N,N}\} =: \mathcal{S}^{\varepsilon}_N$ such that

$$\mathbb{E}\tilde{Z}(s^{\varepsilon}_{i,N}) \geq -\varepsilon \qquad \text{for every } 1 \leq i \leq N \tag{SA.5}$$

and

$$\big|\mathrm{Cov}\big(\tilde{Z}\big(s^{\varepsilon}_{i,N}\big), \tilde{Z}\big(s^{\varepsilon}_{j,N}\big)\big)\big| \leq \varepsilon \qquad \text{for every } 1 \leq i \leq N. \tag{SA.6}$$

Then, using the fact that normal random vectors converge in distribution when their means and variances converge, we have

$$\mathbb{P}\Big[\max_{s\in\mathbb{R}} \tilde{Z}(s) \leq 0\Big] \leq \liminf_{\varepsilon\downarrow 0} \mathbb{P}\Big[\max_{s\in\mathcal{S}^{\varepsilon}_N} \tilde{Z}(s) \leq 0\Big] \leq \left|\int_{-\infty}^{0} \frac{\exp(-x^2/2)}{\sqrt{2\pi}} dx\right|^N = 2^{-N}.$$

By taking N large, we can make the left-hand side probability arbitrarily small.

Using the properties of $\mathcal{K}$ and of $\lim_{t\downarrow 0}[\mu(x+t) - \mu(x)]/\sqrt{t} = 0$, there exists $\bar{\tau}_{\varepsilon,1} \in (0,1)$ such that

$$\mathbb{E}[\tilde{Z}(x+\tau)] = \frac{\mu(x+\tau) - \mu(x)}{\sqrt{\mathcal{K}(\tau,\tau)}} = \frac{[\mu(x+\tau) - \mu(x)]/\sqrt{\tau}}{\sqrt{\mathcal{K}(1,1)}} > -\varepsilon, \qquad \forall\, \tau \in (0, \bar{\tau}_{\varepsilon,1}).$$

Also, for $0 < \tau_j < \tau_i < 1$, using condition $\mathcal{K}(s\tau, t\tau) = \mathcal{K}(s,t)\tau$,

$$\mathrm{Cov}\big(\tilde{Z}(x+\tau_i), \tilde{Z}(x+\tau_j)\big) = \frac{\mathcal{K}(\tau_i,\tau_j)}{\sqrt{\mathcal{K}(\tau_i,\tau_i)\mathcal{K}(\tau_j,\tau_j)}} = \frac{\mathcal{K}(1,\tau_j/\tau_i)/\sqrt{\tau_j/\tau_i}}{\mathcal{K}(1,1)}$$

and $\lim_{\delta\downarrow 0} \mathcal{K}(s, s\delta)/\sqrt{\delta} = 0$ implies that there exists $\bar{\tau}_{\varepsilon,2} \in (0,1)$ such that for all $\tau_j/\tau_i \in (0, \bar{\tau}_{\varepsilon,2})$, $|\mathrm{Cov}(\tilde{Z}(x+\tau_i), \tilde{Z}(x+\tau_j))| \leq \varepsilon$. Now, let $s^{\varepsilon}_{i,N} = x + \bar{\tau}^i_{\varepsilon}/2$ where $\bar{\tau}_{\varepsilon} = \min\{\bar{\tau}_{\varepsilon,1}, \bar{\tau}_{\varepsilon,2}\}$. This choice of $\{s^{\varepsilon}_{i,N}\}_{i=1}^N$ indeed satisfies (SA.5) and (SA.6). □

SA.5.3 Proof of Theorem SA-1

Let $r_n = n^{\frac{\mathfrak{q}}{2\mathfrak{q}+1}}$ and $\widehat{I}_n = \widehat{\Phi}^-_n([0, \widehat{u}_n])$. By Lemma SA-4 and change of variables,

$$\begin{aligned}\mathbb{P}[r_n(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})) > t] &= \mathbb{P}\Big[\sup \operatorname*{arg\,max}_{v\in\widehat{I}_n} \big\{[\theta_0(\mathsf{x}) + t r_n^{-1}]\widehat{\Phi}_n(v) - \mathsf{lsc}_{\widehat{I}_n}(\widehat{\Gamma}_n)(v)\big\} < \widehat{\Phi}^-_n\big(\widehat{\Phi}_n(\mathsf{x})\big)\Big] \\ &= \mathbb{P}\Big[\sup \operatorname*{arg\,min}_{v\in\widehat{V}^{\mathfrak{q}}_{\mathsf{x},n}} \big\{\mathsf{lsc}_{\widehat{V}^{\mathfrak{q}}_{\mathsf{x},n}}(\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n} + M^{\mathfrak{q}}_{\mathsf{x},n})(v) - t\widehat{L}^{\mathfrak{q}}_{\mathsf{x},n}(v)\big\} < \widehat{Z}^{\mathfrak{q}}_{\mathsf{x},n}\Big]\end{aligned}$$

where

$$\begin{aligned}
\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n}(v) &= \sqrt{na_n}[\widehat{\Gamma}_n(\mathsf{x}+va_n^{-1}) - \widehat{\Gamma}_n(\mathsf{x}) - \Gamma_0(\mathsf{x}+va_n^{-1}) + \Gamma_0(\mathsf{x})] \\
&\qquad - \theta_0(\mathsf{x})\sqrt{na_n}[\widehat{\Phi}_n(\mathsf{x}+va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x}) - \Phi_0(\mathsf{x}+va_n^{-1}) + \Phi_0(\mathsf{x})] \\
M^{\mathfrak{q}}_{\mathsf{x},n}(v) &= \sqrt{na_n}[\Gamma_0(\mathsf{x}+va_n^{-1}) - \Gamma_0(\mathsf{x})] - \theta_0(\mathsf{x})\sqrt{na_n}[\Phi_0(\mathsf{x}+va_n^{-1}) - \Phi_0(\mathsf{x})] \\
\widehat{L}^{\mathfrak{q}}_{\mathsf{x},n}(v) &= a_n[\widehat{\Phi}_n(\mathsf{x}+va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x})] \\
\widehat{Z}^{\mathfrak{q}}_{\mathsf{x},n} &= a_n[\widehat{\Phi}_n^{-}(\widehat{\Phi}_n(\mathsf{x})) - \mathsf{x}]
\end{aligned}$$

and $\widehat{V}^{\mathfrak{q}}_{\mathsf{x},n} = \{a_n(x-\mathsf{x}) : x \in \widehat{\Phi}_n^{-}([0,\widehat{u}_n])\}$ as defined in the main text. Note that $\widehat{\Phi}_n$ is continuous on $\widehat{I}_n$ because $\widehat{\Phi}_n(I) \cap [0,\widehat{u}_n]$ is closed.

We let $\mathsf{lsc}_{\widehat{V}^{\mathfrak{q}}_{\mathsf{x},n}}(\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n} + M^{\mathfrak{q}}_{\mathsf{x},n})(v) = \widehat{G}^{\mathfrak{q}}_{\mathsf{x},n}(v) + M^{\mathfrak{q}}_{\mathsf{x},n}(v)$ if $v \notin \widehat{V}^{\mathfrak{q}}_{\mathsf{x},n}$. With $V_n = \{a_n(x-\mathsf{x}) : x \in I\}$, we have

$$\mathsf{lsc}_{V_n}(\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n} + M^{\mathfrak{q}}_{\mathsf{x},n})(v) \le \mathsf{lsc}_{\widehat{V}^{\mathfrak{q}}_{\mathsf{x},n}}(\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n} + M^{\mathfrak{q}}_{\mathsf{x},n})(v) \le \widehat{G}^{\mathfrak{q}}_{\mathsf{x},n}(v) + M^{\mathfrak{q}}_{\mathsf{x},n}(v) \quad \forall\, v \in V_n, \quad \text{almost surely.}$$

For any compact interval K, Lemma 4.4 of van der Vaart and van der Laan (2006) implies that if $\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n} + M^{\mathfrak{q}}_{\mathsf{x},n} \rightsquigarrow \mathcal{G}_{\mathsf{x}} + \mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}$, then $\mathsf{lsc}_K(\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n} + M^{\mathfrak{q}}_{\mathsf{x},n}) \rightsquigarrow \mathcal{G}_{\mathsf{x}} + \mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}$ as $\mathcal{G}_{\mathsf{x}}$ has continuous sample paths with probability one and $\mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}$ is continuous. Thus, for our purpose, it is without loss of generality to look at $\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n} + M^{\mathfrak{q}}_{\mathsf{x},n}$ in place of $\mathsf{lsc}_{\widehat{V}^{\mathfrak{q}}_{\mathsf{x},n}}(\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n} + M^{\mathfrak{q}}_{\mathsf{x},n})$.

By Assumption SA-2 (1) and (3), $\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n} \rightsquigarrow \mathcal{G}_{\mathsf{x}}$, $\sup_{|v|\le C}|\widehat{L}^{\mathfrak{q}}_{\mathsf{x},n}(v) - L_{\mathsf{x}}(v)| = o_{\mathbb{P}}(1)$ with $L_{\mathsf{x}}(v) = \partial\Phi_0(\mathsf{x})v$, and $\widehat{Z}^{\mathfrak{q}}_{\mathsf{x},n} = o_{\mathbb{P}}(1)$.

For $M^{\mathfrak{q}}_{\mathsf{x},n}$ term, the function $\Gamma_0(\mathsf{x}+u) - \theta_0(\mathsf{x})\Phi_0(\mathsf{x}+u) - \Gamma_0(\mathsf{x}) + \theta_0(\mathsf{x})\Phi_0(\mathsf{x})$ converges to 0 as $u \to 0$. The derivative equals $[\theta_0(\mathsf{x}+u) - \theta_0(\mathsf{x})]\partial\Phi_0(\mathsf{x}+u)$, and $|\theta_0(\mathsf{x}+u) - \theta_0(\mathsf{x})|\partial\Phi_0(\mathsf{x}+u)/|u|^{\mathfrak{q}} \to \partial\Phi_0(\mathsf{x})\partial^{\mathfrak{q}}\theta_0(\mathsf{x})/\mathfrak{q}!$. Now by L'Hôpital's rule,

$$\begin{aligned}
\lim_{u\to 0}\frac{\Gamma_0(\mathsf{x}+u) - \theta_0(\mathsf{x})\Phi_0(\mathsf{x}+u) - \Gamma_0(\mathsf{x}) + \theta_0(\mathsf{x})\Phi_0(\mathsf{x})}{|u|^{\mathfrak{q}+1}} &= \lim_{u\to 0}\frac{|\theta_0(\mathsf{x}+u) - \theta_0(\mathsf{x})|\partial\Phi_0(\mathsf{x}+u)}{(\mathfrak{q}+1)|u|^{\mathfrak{q}}} \\
&= \frac{\partial^{\mathfrak{q}}\theta_0(\mathsf{x})\partial\Phi_0(\mathsf{x})}{(\mathfrak{q}+1)!}.
\end{aligned}$$

Thus, $M^{\mathfrak{q}}_{\mathsf{x},n} \to \mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}$ uniformly on compacta.

We apply an argmax continuous mapping theorem (e.g., Theorem 3.2.2 van der Vaart and Wellner, 1996) to prove

$$\sup \operatorname*{arg\,min}_{v\in\widehat{V}^{\mathfrak{q}}_{\mathsf{x},n}}\{\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n}(v) + \widehat{M}^{\mathfrak{q}}_{\mathsf{x},n}(v) - t\widehat{L}^{\mathfrak{q}}_{\mathsf{x},n}(v)\} \rightsquigarrow \operatorname*{arg\,min}_{v\in\mathbb{R}}\Big\{\mathcal{G}_{\mathsf{x}}(v) + \mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}(v) - t\partial\Phi_0(\mathsf{x})v\Big\},$$

which together with Lemma SA-1 implies

$$\mathbb{P}[r_n(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})) > t] \to \mathbb{P}\Big[\operatorname*{arg\,min}_{v\in\mathbb{R}}\Big\{\mathcal{G}_{\mathsf{x}}(v) + \mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}(v) - t\partial\Phi_0(\mathsf{x})v\Big\} < 0\Big].$$

That the limit distribution equals $\mathbb{P}[(\partial\Phi_0(\mathsf{x}))^{-1}\partial_{-}\mathsf{GCM}_{\mathbb{R}}\{\mathcal{G}_{\mathsf{x}} + \mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}\}(0) > t]$ follows from Lemma SA-5. To handle the n-varying domains, we use a modified version of argmax continuous mapping theorem by Cox (2022). To apply the continuous mapping theorem, let $\hat{v}_n = \sup\arg\min_{v\in\widehat{V}^{\mathfrak{q}}_{\mathsf{x},n}}[\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n}(v) + M^{\mathfrak{q}}_{\mathsf{x},n}(v) - t\widehat{L}^{\mathfrak{q}}_{\mathsf{x},n}(v)]$, and we need to verify (i) $\hat{v}_n = O_{\mathbb{P}}(1)$, (ii) $\widehat{V}^{\mathfrak{q}}_{\mathsf{x},n}$ converges to $\mathbb{R}$ in the Painlevé-Kuratowski sense (see Cox (2022) for the defintion), and (iii) the limit process is continuous and has a unique maximum with probability one. The limit process is continuous by the property of the covariance kernel, and the uniqueness of a maximizer follows as in the proof of Lemma SA-1.

To show $\hat{v}_n = O_{\mathbb{P}}(1)$, let $\widehat{\mathsf{x}}_n = \hat{v}_n a_n^{-1} + \mathsf{x}$, that is, $\widehat{\mathsf{x}}_n = \sup\arg\max_{v\in\widehat{I}_n}\{[\theta_0(\mathsf{x}) + tr_n^{-1}]\widehat{\Phi}_n(v) - \widehat{\Gamma}_n(v)\}$. Also, define $\widetilde{\mathsf{x}}_n = \widehat{\Phi}_n^{-}(\widehat{\Phi}_n(\mathsf{x}))$. Since $a_n(\widetilde{\mathsf{x}}_n - \mathsf{x}) = o_{\mathbb{P}}(1)$, it suffices to show $a_n(\widehat{\mathsf{x}}_n - \widetilde{\mathsf{x}}_n) = O_{\mathbb{P}}(1)$.

First we show $|\widehat{\mathsf{x}}_n - \widetilde{\mathsf{x}}_n| = o_{\mathbb{P}}(1)$. For any $\eta > 0$,

$$\begin{aligned}\mathbb{P}[|\widehat{\mathsf{x}}_n - \widetilde{\mathsf{x}}_n| > \eta] &\leq \mathbb{P}\Big[\sup_{v\in\widehat{I}_n:|v-\widetilde{\mathsf{x}}_n|>\eta}\big\{[\theta_0(\mathsf{x}) + tr_n^{-1}][\widehat{\Phi}_n(v) - \widehat{\Phi}_n(\widetilde{\mathsf{x}}_n)] - [\widehat{\Gamma}_n(v) - \widehat{\Gamma}_n(\widetilde{\mathsf{x}}_n)]\big\} \geq 0\Big] \\ &\leq \mathbb{P}\Big[\sup_{v\in I:|v-\mathsf{x}|>\frac{\eta}{2}}\big\{[\theta_0(\mathsf{x}) + tr_n^{-1}][\widehat{\Phi}_n(v) - \widehat{\Phi}_n(\widetilde{\mathsf{x}}_n)] - [\widehat{\Gamma}_n(v) - \widehat{\Gamma}_n(\widetilde{\mathsf{x}}_n)]\big\} \geq 0\Big] + o(1).\end{aligned}$$

The random function inside the bracket converges in probability (uniformly over $x \in I$) to $\theta_0(\mathsf{x})[\Phi_0(v) - \Phi_0(\mathsf{x})] - [\Gamma_0(v) - \Gamma_0(\mathsf{x})]$, which is uniquely maximized at $v = \mathsf{x}$, and $\sup_{v\in I:|v-\mathsf{x}|>\frac{\eta}{2}}\{\theta_0(\mathsf{x})[\Phi_0(v) - \Phi_0(\mathsf{x})] - [\Gamma_0(v) - \Gamma_0(\mathsf{x})]\} < 0$. Thus, $|\widehat{\mathsf{x}}_n - \widetilde{\mathsf{x}}_n| = o_{\mathbb{P}}(1)$ holds.

Following the argument similar to Theorem 3.2.5 of van der Vaart and Wellner (1996) and using $[\theta_0(\mathsf{x}) + tr_n^{-1}][\widehat{\Phi}_n(\widetilde{\mathsf{x}}_n) - \widehat{\Phi}_n(\mathsf{x})] - [\widehat{\Gamma}_n(\widetilde{\mathsf{x}}_n) - \widehat{\Gamma}_n(\mathsf{x})] = o_{\mathbb{P}}(a_n^{-(1+\mathfrak{q})})$, it suffices to bound for any small $\eta > 0$ and sufficiently large $M > 0$

$$\begin{aligned}&\sum_{j\geq M, 2^j\leq \eta a_n}\mathbb{P}\Big[\sup_{2^{j-1}<a_n|v|\leq 2^j} -\widehat{\Gamma}_n(\mathsf{x}+v) + \widehat{\Gamma}_n(\mathsf{x}) + [\theta_0(\mathsf{x}) + tr_n^{-1}][\widehat{\Phi}_n(\mathsf{x}+v) - \widehat{\Phi}_n(\mathsf{x})] \geq o_{\mathbb{P}}(a_n^{-(1+\mathfrak{q})})\Big] \\ &\leq \sum_{j\geq M, 2^j\leq \eta a_n}\mathbb{P}\Big[\sup_{a_n|v|\leq 2^j} -\widehat{G}^{\mathfrak{q}}_{n,\mathsf{x}}(va_n) + ta_n[\Phi_0(\mathsf{x}+v) - \Phi_0(v)] \geq c2^{(j-1)(\mathfrak{q}+1)} + o_{\mathbb{P}}(1)\Big]\end{aligned}$$

where we use $-\Gamma_0(\mathsf{x}+v) + \Gamma_0(\mathsf{x}) + \theta_0(\mathsf{x})[\Phi_0(\mathsf{x}+v) - \Phi_0(\mathsf{x})] \leq -c|v|^{\mathfrak{q}+1}$ for some $c > 0$ and $|v|$ close to zero and $a_n\sup_{|v-\mathsf{x}|\leq\eta}|\widehat{\Phi}_n(v) - \Phi_0(v)| = o_{\mathbb{P}}(1)$.

With some abuse of notation, we analyze the above probability by replacing $\widehat{G}^{\mathfrak{q}}_{n,\mathsf{x}}$ with $\mathcal{G}_{\mathsf{x}} + o_{\mathbb{P}}(1)$, which is possible by Dudley's representation theorem (e.g., Theorem 2.2 of Kim and Pollard, 1990). Then,

$$\begin{aligned}&\sum_{j\geq M, 2^j\leq \eta a_n}\mathbb{P}\Big[\sup_{a_n|v|\leq 2^j} -\mathcal{G}_{\mathsf{x}}(va_n) + ta_n[\Phi_0(\mathsf{x}+v) - \Phi_0(v)] \geq c2^{(j-1)(\mathfrak{q}+1)} - o_{\mathbb{P}}(1)\Big] \\ &\leq C\sum_{j\geq M, 2^j\leq \eta a_n}\frac{\mathbb{E}\Big[\sup_{|v|\leq 2^j}|\mathcal{G}_{\mathsf{x}}(v)|\Big] + a_n\sup_{|v|\leq 2^j a_n^{-1}}|\Phi_0(\mathsf{x}+v) - \Phi_0(\mathsf{x})|}{2^{j(\mathfrak{q}+1)}} \\ &\leq C\sum_{j\geq M, 2^j\leq \eta a_n}2^{-j(\mathfrak{q}+1)}\big[2^{j/2} + 2^j\big]\end{aligned}$$

where the last inequality uses a maximal inequality for Gaussian processes (e.g., Corollary 2.2.8 of van der Vaart and Wellner, 1996) and the property of the covariance kernel $|\mathcal{C}_{\mathsf{x}}(s,s) + \mathcal{C}_{\mathsf{x}}(t,t) - 2\mathcal{C}_{\mathsf{x}}(s,t)| = |s-t|\mathcal{C}_{\mathsf{x}}(1,1)$. Thus, taking M large enough makes the sum of the probabilities arbitrary small.

It remains to show $\widehat{V}^{\mathfrak{q}}_{\mathsf{x},n}$ converges to $\mathbb{R}$ in the Painlevé-Kuratowski sense. By Assumption SA-1, Φ_0 is strictly increasing and continuously differentiable on a neighborhood of x. Also,

$a_n \sup_{|v|\leq K} |\widehat{\Phi}_n(\mathsf{x}+va_n^{-1}) - \Phi_0(\mathsf{x}+va_n^{-1})| = o_{\mathbb{P}}(1)$ implies that

$$\sup_{|v|\leq K} \Big| a_n \big[\widehat{\Phi}_n^-(\Phi_0(\mathsf{x}+va_n^{-1})) - \Phi_0^-(\Phi_0(\mathsf{x}+va_n^{-1}))\big]\Big| = \sup_{|v|\leq K} \Big| a_n[\widehat{\Phi}_n^-(\Phi_0(\mathsf{x}+va_n^{-1})) - \mathsf{x}] - v\Big|$$

is $o_{\mathbb{P}}(1)$. Then, every $b \in \mathbb{R}$ belongs to the Painlevé-Kuratowski limit of $\widehat{V}_{\mathsf{x},n}^{\mathfrak{q}}$. □

SA.5.4 Proof of Theorem SA-2

The argument is analogous to the proof of Theorem SA-1 with appropriate changes for bootstrap.

Let $\widehat{G}_{\mathsf{x},n}^{\mathfrak{q},*}(v) = \sqrt{na_n}[\widehat{\Gamma}_n^*(\mathsf{x}+va_n^{-1}) - \widehat{\Gamma}_n^*(\mathsf{x}) - \widehat{\Gamma}_n(\mathsf{x}+va_n^{-1}) + \widehat{\Gamma}_n(\mathsf{x})] - \widehat{\theta}_n(\mathsf{x})\sqrt{na_n}[\widehat{\Phi}_n^*(\mathsf{x}+va_n^{-1}) - \widehat{\Phi}_n^*(\mathsf{x}) - \widehat{\Phi}_n(\mathsf{x}+va_n^{-1}) + \widehat{\Phi}_n(\mathsf{x})]$, $\widehat{L}_{\mathsf{x},n}^{\mathfrak{q},*}(v) = a_n[\widehat{\Phi}_n^*(\mathsf{x}+va_n^{-1}) - \widehat{\Phi}_n^*(\mathsf{x})]$, $\widehat{Z}_{\mathsf{x},n\mathsf{x}}^{\mathfrak{q},*} = a_n[(\widehat{\Phi}_n^*)^-(\widehat{\Phi}_n^*(\mathsf{x})) - \mathsf{x}]$, $\widehat{I}_n^* = (\widehat{\Phi}_n^*)^-([0,\widehat{u}_n^*])$, and $\widehat{V}_{\mathsf{x},n}^{\mathfrak{q},*} = \{a_n(x-\mathsf{x}) : x \in \widehat{I}_n^*\}$. Then,

$$\begin{aligned}
\mathbb{P}[r_n(\widetilde{\theta}_n^*(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x})) > t] &= \mathbb{P}\Big[\sup \operatorname*{arg\,max}_{v\in\widehat{I}_n^*} \big\{[\widehat{\theta}(\mathsf{x}) + tr_n^{-1}]\widehat{\Phi}_n^*(v) - \mathsf{lsc}_{\widehat{I}_n^*}(\widetilde{\Gamma}_n^*)(v)\big\} < (\widehat{\Phi}_n^*)^-(\widehat{\Phi}_n^*(\mathsf{x}))\Big] \\
&= \mathbb{P}\Big[\sup \operatorname*{arg\,min}_{v\in\widehat{V}_{\mathsf{x},n}^{\mathfrak{q},*}} \big\{\mathsf{lsc}_{\widehat{V}_{\mathsf{x},n}^{\mathfrak{q},*}}(\widehat{G}_{n,\mathsf{x}}^{\mathfrak{q},*} + \widetilde{M}_{\mathsf{x},n}^{\mathfrak{q}})(v) - t\widehat{L}_{\mathsf{x},n}^{\mathfrak{q},*}(v)\big\} < \widehat{Z}_{\mathsf{x},n\mathsf{x}}^{\mathfrak{q},*}\Big]
\end{aligned}$$

where the first equality uses Lemma SA-4 and the second equality uses changes of variables and continuity of $\widehat{\Phi}_n^*$ on $\widehat{I}_n^*$. With the same argument used for the proof of Theorem SA-1, it is without loss of generality to look at $\widehat{G}_{n,\mathsf{x}}^{\mathfrak{q},*} + \widetilde{M}_{\mathsf{x},n}^{\mathfrak{q}}$ in place of $\mathsf{lsc}_{\widehat{V}_{\mathsf{x},n}^{\mathfrak{q},*}}(\widehat{G}_{n,\mathsf{x}}^{\mathfrak{q},*} + \widetilde{M}_{\mathsf{x},n}^{\mathfrak{q}})$. By the hypothesis,

$$\widehat{G}_{n,\mathsf{x}}^{\mathfrak{q},*} + \widetilde{M}_{\mathsf{x},n}^{\mathfrak{q}} \rightsquigarrow_{\mathbb{P}} \mathcal{G}_{\mathsf{x}} + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}.$$

For the other term, $a_n \sup_{|v|\leq K} |\widehat{\Phi}_n(\mathsf{x}+va_n^{-1}) - \Phi_0(\mathsf{x}+va_n^{-1})| = o_{\mathbb{P}}(1)$ and $a_n \sup_{|v|\leq \eta} |\widehat{\Phi}_n^*(\mathsf{x}+va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x}+va_n^{-1})| = o_{\mathbb{P}}(1)$ imply $\sup_{|v|\leq K} |\widehat{L}_{\mathsf{x},n}^{\mathfrak{q},*}(v) - a_n[\Phi_0(\mathsf{x}+va_n^{-1}) - \Phi_0(\mathsf{x})]| = o_{\mathbb{P}}(1)$ and $\widehat{L}_{\mathsf{x},n}^{\mathfrak{q},*}(v)$ converges to $\partial\Phi_0(\mathsf{x})v$ uniformly on compacta. Also, $a_n[(\widehat{\Phi}_n^*)^-(\widehat{\Phi}_n^*(\mathsf{x})) - \mathsf{x}] = o_{\mathbb{P}}(1)$.

It remains to verify the hypothesis of the argmax continuous mapping theorem. Let $\widehat{x}_n^* = \operatorname*{arg\,max}_{v\in\widehat{I}_n^*}\{[\widehat{\theta}_n(\mathsf{x})+tr_n^{-1}]\widehat{\Phi}_n^*(v) - \widetilde{\Gamma}_n^*(v)\}$ and we check $a_n(\widehat{\mathsf{x}}_n^* - \widetilde{\mathsf{x}}_n^*) = O_{\mathbb{P}}(1)$ with $\widetilde{\mathsf{x}}_n^* = (\widehat{\Phi}_n^*)^-(\widehat{\Phi}_n^*(\mathsf{x}))$. With $\widehat{\Gamma}_{n,0}^*(v) = \widehat{\Gamma}_n^*(v) - \widehat{\Gamma}_n(v)$ and $\widehat{\Phi}_{n,0}^*(v) = \widehat{\Phi}_n^*(v) - \widehat{\Phi}_n(v)$,

$$\begin{aligned}
\mathbb{P}\big[|\widehat{\mathsf{x}}_n^* - \widetilde{\mathsf{x}}_n^*| > \eta\big] &\leq \mathbb{P}\Big[\sup_{|v-\widetilde{\mathsf{x}}_n^*|>\eta} \big\{\widehat{\theta}_n(\mathsf{x})[\widehat{\Phi}_{n,0}^*(v) - \widehat{\Phi}_{n,0}^*(\widetilde{\mathsf{x}}_n^*)] - [\widehat{\Gamma}_{n,0}^*(v) - \widehat{\Gamma}_{n,0}^*(\widetilde{\mathsf{x}}_n^*)] \\
&\qquad + tr_n^{-1}[\widehat{\Phi}_n^*(v) - \widehat{\Phi}_n^*(\widetilde{\mathsf{x}}_n^*)] - \widetilde{M}_{\mathsf{x},n}(v-\mathsf{x}) + \widetilde{M}_{\mathsf{x},n}(\widetilde{x}_n^* - \mathsf{x})\big\} \geq 0\Big] \\
&\leq \mathbb{P}\Big[\sup_{|v-\mathsf{x}|>\eta/2} \big\{-\widetilde{M}_{\mathsf{x},n}(v-\mathsf{x}) + \widetilde{M}_{\mathsf{x},n}(\widetilde{x}_n^* - \mathsf{x})\big\} \geq -o_{\mathbb{P}}(1)\Big] + o(1)
\end{aligned}$$

where for some $\delta > 0$, $-\widetilde{M}_{\mathsf{x},n}(v-\mathsf{x}) \leq -\delta$ for all $|v-\mathsf{x}| > \eta/2$ with probability approaching one by Assumption SA-4 and $\widetilde{M}_{\mathsf{x},n}(\widetilde{x}_n^* - \mathsf{x}) = (na_n)^{-1/2}[\widetilde{M}_{\mathsf{x},n}^{\mathfrak{q}}(a_n(\widetilde{x}_n^* - \mathsf{x})) \mp \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(a_n(\widetilde{x}_n^* - \mathsf{x}))] = o_{\mathbb{P}}(1)$ by $a_n(\widetilde{x}_n^* - \mathsf{x}) = o_{\mathbb{P}}(1)$ and $\widetilde{M}_{\mathsf{x},n}^{\mathfrak{q}} \rightsquigarrow \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}$. Thus, the majorant probability goes to zero.

Now, to show $a_n(\widehat{\mathsf{x}}_n^* - \widetilde{\mathsf{x}}_n^*) = O_{\mathbb{P}}(1)$, arguing as in the proof of Theorem SA-1, it suffices to bound for any small $\eta > 0$ and sufficiently large $M > 0$

$$\sum_{j \geq M, 2^j \leq \eta a_n} \mathbb{P}^* \left[\sup_{2^{j-1} < a_n |v| \leq 2^j} \widehat{G}^{\mathfrak{q},*}_{n,\mathsf{x}}(v a_n) + t a_n [\widehat{\Phi}^*_n(\mathsf{x}+v) - \widehat{\Phi}^*_n(\mathsf{x})] + \widetilde{M}^{\mathfrak{q}}_{\mathsf{x},n}(v a_n) \geq o_{\mathbb{P}}(1) \right]$$
$$\leq \sum_{j \geq M, 2^j \leq \eta a_n} \mathbb{P}^* \left[\sup_{a_n |v| \leq 2^j} \widehat{G}^{\mathfrak{q},*}_{n,\mathsf{x}}(v a_n) + t a_n [\Phi_0(\mathsf{x}+v) - \Phi_0(\mathsf{x})] \geq c 2^{j-1} + o_{\mathbb{P}}(1) \right]$$

where we used $\sup_{|v| \leq \eta} |\widetilde{M}^{\mathfrak{q}}_{\mathsf{x},n}(v) - \mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}(v)| = o_{\mathbb{P}}(1)$ and $\sup_{2^{j-1} < a_n |v| \leq 2^j} \mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}(a_n v) = -2^{j-1} \mathcal{D}_{\mathfrak{q}}(\mathsf{x})$. Using the weak convergence in probability of $\widehat{G}^{\mathfrak{q},*}_{n,\mathsf{x}}(v)$ to invoke the representation theorem, we can show $a_n(\widehat{x}^\star_n - \mathsf{x}_0) = O_{\mathbb{P}}(1)$. The set convergence of the domain follows from the argument similar to the one used in the proof of Theorem SA-1. □

SA.5.5 Proof of Lemma SA-2

Monomial approximation estimator

$$\widetilde{\mathcal{D}}^{\mathtt{MA}}_{\mathfrak{q},n}(\mathsf{x}) = \epsilon_n^{-(\mathfrak{q}+1)} [\Gamma_0(\mathsf{x}+\epsilon_n) - \Gamma_0(\mathsf{x}) - \theta_0(\mathsf{x})\{\Phi_0(\mathsf{x}+\epsilon_n) - \Phi_0(\mathsf{x})\}] \tag{SA.7}$$
$$+ \epsilon_n^{-(\mathfrak{q}+1/2)} n^{-1/2} \widehat{G}_{\mathsf{x},n}(1;\epsilon_n) \tag{SA.8}$$
$$- \epsilon_n^{-\mathfrak{q}} [\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})] \widehat{R}_{\mathsf{x},n}(1;\epsilon_n) \tag{SA.9}$$
$$- \epsilon_n^{-(\mathfrak{q}+1)} [\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})][\Phi_0(\mathsf{x}+\epsilon_n) - \Phi_0(\mathsf{x})]. \tag{SA.10}$$

The term (SA.7) converges to $\frac{\partial\Phi_0(\mathsf{x})\partial^{\mathfrak{q}}\theta_0(\mathsf{x})}{(1+\mathfrak{q})!}$ as argued in the proof of Theorem SA-1. The term (SA.8) is $O_{\mathbb{P}}([n\epsilon_n^{1+2\mathfrak{q}}]^{-1/2}) = o_{\mathbb{P}}(1)$, and the term (SA.9) and (SA.10) are $O_{\mathbb{P}}(n^{-\frac{\mathfrak{q}}{2\mathfrak{q}+1}} \epsilon_n^{-\mathfrak{q}}) = o_{\mathbb{P}}(1)$ by $\epsilon_n^{2\mathfrak{q}+1} n \to \infty$.

Forward difference estimator

$$\widetilde{\mathcal{D}}^{\mathtt{FD}}_{\mathfrak{q},n}(\mathsf{x}) = \epsilon_n^{-(\mathfrak{q}+1)} \sum_{k=1}^{\mathfrak{q}+1} (-1)^{k+\mathfrak{q}+1} \binom{\mathfrak{q}+1}{k} [\Upsilon_0(\mathsf{x}+k\epsilon_n) - \Upsilon_0(\mathsf{x})] \tag{SA.11}$$
$$+ \epsilon_n^{-(\mathfrak{q}+1/2)} n^{-1/2} \sum_{k=1}^{\mathfrak{q}+1} (-1)^{k+\mathfrak{q}+1} \binom{\mathfrak{q}+1}{k} \widehat{G}_{\mathsf{x},n}(k;\epsilon_n) \tag{SA.12}$$
$$- \epsilon_n^{-\mathfrak{q}} [\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})] \sum_{k=1}^{\mathfrak{q}+1} (-1)^{k+\mathfrak{q}+1} \binom{\mathfrak{q}+1}{k} \widehat{R}_{\mathsf{x},n}(k;\epsilon_n) \tag{SA.13}$$
$$- \epsilon_n^{-(\mathfrak{q}+1)} [\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})] \sum_{k=1}^{\mathfrak{q}+1} (-1)^{k+\mathfrak{q}+1} \binom{\mathfrak{q}+1}{k} [\Phi_0(\mathsf{x}+k\epsilon_n) - \Phi_0(\mathsf{x})]. \tag{SA.14}$$

(SA.11) converges to $\partial^{\mathfrak{q}+1}\Upsilon_0(\mathsf{x}) = \partial\Phi_0(\mathsf{x})\partial^{\mathfrak{q}}\theta_0(\mathsf{x})/(\mathfrak{q}+1)!$ by the standard forward difference formula. For (SA.12)-(SA.14), they are $o_{\mathbb{P}}(1)$ by the same argument as above for each k.

Bias-reduced estimator

$$\widetilde{\mathcal{D}}^{\mathtt{BR}}_{j,n}(\mathsf{x}) = \epsilon_n^{-(j+1)} \sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda_j^{\mathtt{BR}}(k)[\Upsilon_0(\mathsf{x}+c_k\epsilon_n) - \Upsilon_0(\mathsf{x})] \tag{SA.15}$$

$$+ \epsilon_n^{-(j+1/2)} n^{-1/2} \sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda_j^{\mathtt{BR}}(k)\widehat{G}_{\mathsf{x},n}(c_k;\epsilon_n) \tag{SA.16}$$

$$- \epsilon_n^{-j}[\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})] \sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda_j^{\mathtt{BR}}(k)\widehat{R}_{\mathsf{x},n}(c_k;\epsilon_n) \tag{SA.17}$$

$$- \epsilon_n^{-(j+1)}[\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})] \sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda_j^{\mathtt{BR}}(k)[\Phi_0(\mathsf{x}+c_k\epsilon_n) - \Phi_0(\mathsf{x})]. \tag{SA.18}$$

(SA.16) is $O_{\mathbb{P}}(\epsilon_n^{-(j+1/2)}n^{-1/2})$, (SA.17) is $o_{\mathbb{P}}(\epsilon_n^{-j}a_n^{-\mathfrak{q}})$, and (SA.18) is $O_{\mathbb{P}}(\epsilon_n^{-j}a_n^{-\mathfrak{q}})$. Note $\epsilon_n^{-(j+1/2)}n^{-1/2} = \epsilon_n^{-(j+1/2)}a_n^{-1/2}a_n^{-\mathfrak{q}}$. For (SA.15), by the definition of $\{\lambda_j^{\mathtt{BR}}(k) : k = 1, \ldots, \underline{\mathfrak{s}}\}$,

$$\epsilon_n^{-(j+1)} \sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda_j^{\mathtt{BR}}(k)[\Upsilon_0(\mathsf{x}+c_k\epsilon_n) - \Upsilon_0(\mathsf{x})] = \frac{\partial^{j+1}\Upsilon_0(\mathsf{x})}{(j+1)!} + O\big(\epsilon_n^{\min\{\mathfrak{s},\underline{\mathfrak{s}}+1\}-j}\big)$$

and the first part of the lemma follows. For the second part, note

$$a_n^{\mathfrak{q}-j}\big(\widetilde{\mathcal{D}}^{\mathtt{BR}}_{j,n}(\mathsf{x}) - \mathcal{D}_j(\mathsf{x})\big) = O\big(a_n^{\mathfrak{q}-j}\epsilon_n^{\min\{\mathfrak{s},\underline{\mathfrak{s}}+1\}-j}\big) + O_{\mathbb{P}}\Big([n\epsilon_n^{2\mathfrak{q}+1}]^{-\frac{j}{2\mathfrak{q}+1}}\Big)$$

and it is $o_{\mathbb{P}}(1)$ for every $(j,\mathfrak{q}) \in \{1,\ldots,\bar{q}\}^2$ if $n\epsilon_n^{(1+2\bar{\mathfrak{q}})\min\{\mathfrak{s}-1,\underline{\mathfrak{s}}\}/(\bar{\mathfrak{q}}-1)} \to 0$ and $n\epsilon_n^{2\bar{\mathfrak{q}}+1} \to \infty$. $\square$

SA.5.5.1 Higher-order expansion of the bias-reduced estimator

Here, we additionally assume that θ_0 is $(\underline{\mathfrak{s}}+1)$-times continusly differentiable and Φ_0 is $(\underline{\mathfrak{s}}+2)$-times continuously differentiable on I_{x}^{δ} for some $\delta > 0$. Then,

$$\begin{aligned}\epsilon_n^{-(j+1)} \sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda_j^{\mathtt{BR}}(k)[\Upsilon_0(\mathsf{x}+c_k\epsilon_n) - \Upsilon_0(\mathsf{x})] &- \frac{\partial^{j+1}\Upsilon_0(\mathsf{x})}{(j+1)!} \\ &= \epsilon_n^{\underline{\mathfrak{s}}+1-j}\frac{\partial^{\underline{\mathfrak{s}}+2}\Upsilon_0(\mathsf{x})}{(\underline{\mathfrak{s}}+2)!} \sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda_j^{\mathtt{BR}}(k)c_k^{\underline{\mathfrak{s}}+2} + o(\epsilon_n^{\underline{\mathfrak{s}}+1-j}).\end{aligned}$$

Also,

$$\epsilon_n^{-(j+1)} \sum_{k=1}^{\underline{\mathfrak{s}}+1} \lambda_j^{\mathtt{BR}}(k)[\Phi_0(\mathsf{x}+c_k\epsilon_n) - \Phi_0(\mathsf{x})] = \frac{\partial^{j+1}\Phi_0(\mathsf{x})}{(j+1)!} + O(\epsilon_n^{\underline{\mathfrak{s}}+1-j}),$$

and

$$(\text{SA.18}) = -[\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})]\frac{\partial^{j+1}\Phi_0(\mathsf{x})}{(j+1)!} + O\big(\epsilon_n^{\underline{\mathfrak{s}}+1-j}n^{-\frac{\mathfrak{q}}{2\mathfrak{q}+1}}\big).$$

The first term is independent of ϵ_n, so we can treat (SA.18) as a higher-order term (of smaller magnitude than (SA.15)).

SA.5.6 Proof of Lemma SA-3

We verify that Assumptions SA-1 and SA-6 imply Assumptions SA-2 SA-3-(3) and Assumption SA-3 (1)-(3). Verifying Assumption SA-5 is straightforward.

SA.5.6.1 Assumption SA-2 (1)

By Assumption SA-6 (3)-(4),

$$\widehat{G}^{\mathfrak{q}}_{\mathsf{x},n}(v) = \sqrt{\frac{a_n}{n}} \sum_{i=1}^{n} \left\{ \psi_{\mathsf{x}}(va_n^{-1}; \mathbf{Z}_i) - \mathbb{E}[\psi_{\mathsf{x}}(va_n^{-1}; \mathbf{Z})] \right\} + o_{\mathbb{P}}(1)$$

where the remainder term is uniformly small over $\{v : |v| \leq K\}$ for any fixed $K > 0$. Letting $\bar{\psi}_{\mathsf{x},n}(v; \mathbf{Z}_i) = \sqrt{a_n}\psi_{\mathsf{x}}(va_n^{-1}; \mathbf{Z}_i)$, we want to prove that the empirical process of $\{\bar{\psi}_{\mathsf{x},n}(v; \cdot) : |v| \leq K\}$ weakly converges to $\mathcal{G}_{\mathsf{x}}$. We verify finite-dimensional weak convergence and stochastic equicontinuity.

Letting $\delta_n = Ka_n^{-1}$,

$$n^{-1}\mathbb{E}[|\bar{\psi}_{\mathsf{x},n}(v; \mathbf{Z})|^4] \leq Cn^{-1}a_n^2 \mathbb{E}[\bar{D}^{\delta_n}_{\gamma}(\mathbf{Z})^4 + \bar{D}^{\delta_n}_{\phi}(\mathbf{Z})^4] = o(1)$$

for all $|v| \leq K$. Also, convergence of the covariance kernel is imposed in Assumption SA-6 (5). Thus, the Lyapunov central limit theorem implies the finite-dimensional convergence.

For stochastic equicontinuity, following the argument of Kim and Pollard (1990, Lemma 4.6) and using $a_n\mathbb{E}[\bar{D}^{\delta_n}_{\gamma}(\mathbf{Z})^2 + \bar{D}^{\delta_n}_{\phi}(\mathbf{Z})^2] = O(1)$, it suffices to show $\sup_{|v-s|\leq \eta_n, |v|\vee|s|\leq C} \frac{1}{n}\sum_{i=1}^{n} |\bar{\psi}_{\mathsf{x},n}(v; \mathbf{Z}_i) - \bar{\psi}_{\mathsf{x},n}(s; \mathbf{Z}_i)|^2 = o_{\mathbb{P}}(1)$ for any $\eta_n = o(1)$. For a constant $C > 0$,

$$\begin{aligned}
&\sup_{|v-s|\leq \eta_n, |v|\vee|s|\leq K} \frac{1}{n}\sum_{i=1}^{n} |\bar{\psi}_{\mathsf{x},n}(v; \mathbf{Z}_i) - \bar{\psi}_{\mathsf{x},n}(s; \mathbf{Z}_i)|^2 \\
&\leq \frac{1}{n}\sum_{i=1}^{n} a_n[\bar{D}^{\delta_n}_{\gamma}(\mathbf{Z}_i)^2 + \bar{D}^{\delta_n}_{\phi}(\mathbf{Z}_i)^2]\mathbb{1}\{\bar{D}^{\delta_n}_{\gamma}(\mathbf{Z}_i) + \bar{D}^{\delta_n}_{\phi}(\mathbf{Z}_i) > C\} \\
&\quad + a_n C \sup_{|v-s|\leq \eta_n, |v|\vee|s|\leq K} \mathbb{E}\big[|\psi_{\mathsf{x}}(va_n^{-1}; \mathbf{Z}) - \psi_{\mathsf{x}}(sa_n^{-1}; \mathbf{Z})|\big] \\
&\quad + Ca_n \sup_{|v-s|\leq \eta_n, |v|\vee|s|\leq K} \frac{1}{n}\sum_{i=1}^{n} \left\{ |\psi_{\mathsf{x}}(va_n^{-1}; \mathbf{Z}_i) - \psi_{\mathsf{x}}(sa_n^{-1}; \mathbf{Z}_i)| - \mathbb{E}\big[|\psi_{\mathsf{x}}(va_n^{-1}; \mathbf{Z}) - \psi_{\mathsf{x}}(sa_n^{-1}; \mathbf{Z})|\big] \right\}
\end{aligned}$$

where the first term after the inequality can be made arbitrarily small by making C large using $a_n\mathbb{E}[\bar{D}^{\delta_n}_{\gamma}(\mathbf{Z})^4 + \bar{D}^{\delta_n}_{\phi}(\mathbf{Z})^4] = O(1)$. The second term is $o_{\mathbb{P}}(1)$ by Assumption SA-6 (5). Finally, the third term is $O_{\mathbb{P}}(\sqrt{a_n/n})$ using Theorem 4.2 of Pollard (1989).

SA.5.6.2 Assumptions SA-2 (2)-(3)

For Assumption SA-2 (2),

$$\sup_{x\in I} \left|\widehat{\Gamma}_n(x) - \Gamma_0(x)\right| \leq \sup_{x\in I} \left|\widehat{\Gamma}_n(x) - \bar{\Gamma}_n(x)\right| + \sup_{x\in I} \left|\bar{\Gamma}_n(x) - \Gamma_0(x)\right|$$

where the first term after the inequality is assumed to be $o_{\mathbb{P}}(1)$ and the second term is $o_{\mathbb{P}}(1)$ by standard empirical process arguments. The identical argument implies $\sup_{x\in I} |\widehat{\Phi}_n(x) - \Phi_0(x)| = o_{\mathbb{P}}(1)$. For $a_n \sup_{|v|\leq K} |\widehat{\Phi}_n(\mathsf{x}+va_n^{-1}) - \Phi_0(\mathsf{x}+va_n^{-1})| = o_{\mathbb{P}}(1)$,

$$\begin{aligned}\sup_{|v|\leq K}\left|\widehat{\Phi}_n(\mathsf{x}+va_n^{-1}) - \Phi_0(\mathsf{x}+va_n^{-1})\right| \leq & \sup_{|v|\leq K}\left|\widehat{\Phi}_n(\mathsf{x}+va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x}) - \bar{\Phi}_n(\mathsf{x}+va_n^{-1}) + \bar{\Phi}_n(\mathsf{x})\right| \\ &+ \sup_{|v|\leq K}\left|\bar{\Phi}_n(\mathsf{x}+va_n^{-1}) - \bar{\Phi}_n(\mathsf{x}) - \Phi_0(\mathsf{x}+va_n^{-1}) + \Phi_0(\mathsf{x})\right| \\ &+ \left|\widehat{\Phi}_n(\mathsf{x}) - \bar{\Phi}_n(\mathsf{x})\right| + \left|\bar{\Phi}_n(\mathsf{x}) - \Phi_0(\mathsf{x})\right|\end{aligned}$$

where the first term after the inequality is $o_{\mathbb{P}}((na_n)^{-1/2}) = o_{\mathbb{P}}(a_n^{-1})$ and $|\widehat{\Phi}_n(\mathsf{x}) - \bar{\Phi}_n(\mathsf{x})| = o_{\mathbb{P}}(a_n^{-1})$. The remaining terms are also $o_{\mathbb{P}}(a_n^{-1})$ by standard arguments.

SA.5.6.3 Assumption SA-3 (1)

First we posit

$$\begin{aligned}\sqrt{na_n}\sup_{|v|\leq K}|\widehat{\Gamma}^*_n(\mathsf{x}+va_n^{-1}) - \widehat{\Gamma}^*_n(\mathsf{x}) - \bar{\Gamma}^*_n(\mathsf{x}+va_n^{-1}) + \bar{\Gamma}^*_n(\mathsf{x})| = o_{\mathbb{P}}(1),\\ \sqrt{na_n}\sup_{|v|\leq K}|\widehat{\Phi}^*_n(\mathsf{x}+va_n^{-1}) - \widehat{\Phi}^*_n(\mathsf{x}) - \bar{\Phi}^*_n(\mathsf{x}+va_n^{-1}) + \bar{\Phi}^*_n(\mathsf{x})| = o_{\mathbb{P}}(1),\end{aligned} \tag{SA.19}$$

which follows from the hypothesis of the lemma as shown below. By the above display,

$$\widehat{G}^{\mathsf{q},*}_{\mathsf{x},n}(v) = \sqrt{\frac{a_n}{n}}\sum_{i=1}^n W_{i,n}\Big\{\psi_{\mathsf{x}}(va_n^{-1};\mathbf{Z}_i) - \frac{1}{n}\sum_{j=1}^n \psi_{\mathsf{x}}(va_n^{-1};\mathbf{Z}_j)\Big\} + o_{\mathbb{P}}(1)$$

where we use $\sqrt{\frac{a_n}{n}}\sum_{i=1}^n W_{i,n}\{\phi_0(\mathsf{x}+va_n^{-1};\mathbf{Z}_i) - \phi_0(\mathsf{x};\mathbf{Z}_i) - \frac{1}{n}\sum_{j=1}^n[\phi_0(\mathsf{x}+va_n^{-1};\mathbf{Z}_j) - \phi_0(\mathsf{x};\mathbf{Z}_j)]\} = O_{\mathbb{P}}(1)$, and $\widehat{\theta}_n(\mathsf{x}) \to_{\mathbb{P}} \theta_0(\mathsf{x})$. Let $\widehat{\psi}_{\mathsf{x},n}(va_n^{-1};\mathbf{Z}) = \sqrt{a_n}[\psi_{\mathsf{x}}(va_n^{-1};\mathbf{Z}) - \frac{1}{n}\sum_{j=1}^n \psi_{\mathsf{x}}(va_n^{-1};\mathbf{Z}_j)]$ and to prove the finite-dimensional convergence, we apply Lemma 3.6.15 of van der Vaart and Wellner (1996). Assumption SA-6 (2) implies $\frac{1}{n}\sum_{i=1}^n (W_{i,n}-1)^2 \to_{\mathbb{P}} 1$ and $n^{-1}\max_{1\leq i\leq n} W_{i,n}^2 = o_{\mathbb{P}}(1)$. Since

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \widehat{\psi}_{\mathsf{x},n}(va_n^{-1};\mathbf{Z}_i)\widehat{\psi}_{\mathsf{x},n}(ua_n^{-1};\mathbf{Z}_i) = a_n\Bigg[&\frac{1}{n}\sum_{i=1}^n \psi_{\mathsf{x}}(va_n^{-1};\mathbf{Z}_i)\psi_{\mathsf{x}}(ua_n^{-1};\mathbf{Z}_i) \\ &- \frac{1}{n}\sum_{i=1}^n \psi_{\mathsf{x}}(va_n^{-1};\mathbf{Z}_i)\frac{1}{n}\sum_{i=1}^n \psi_{\mathsf{x}}(ua_n^{-1};\mathbf{Z}_i)\Bigg]\end{aligned}$$

and $\sup_{|v|\leq\delta}\psi_{\mathsf{x}}(v;\mathbf{Z}) \leq \bar{D}^{\delta}_{\gamma}(\mathbf{Z}) + |\theta_0(\mathsf{x})|\bar{D}^{\delta}_{\phi}(\mathbf{Z})$, for any $v,u \in \mathbb{R}$,

$$\frac{1}{n}\sum_{i=1}^n \widehat{\psi}_{\mathsf{x},n}(va_n^{-1};\mathbf{Z}_i)\widehat{\psi}_{\mathsf{x},n}(ua_n^{-1};\mathbf{Z}_i) - a_n\mathbb{E}[\psi_{\mathsf{x}}(va_n^{-1};\mathbf{Z})\psi_{\mathsf{x}}(ua_n^{-1};\mathbf{Z})] = o_{\mathbb{P}}(1).$$

Also, $\frac{1}{n}\sum_{i=1}^n \widehat{\psi}^4_{\mathsf{x},n}(va_n^{-1};\mathbf{Z}_i) = O_{\mathbb{P}}(1)$ and we verified the hypothesis of the lemma.

For stochastic equicontinuity, Lemma 3.6.7 of van der Vaart and Wellner (1996) implies that for any $n_0 \in \{1, \dots, n\}$,

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{|v-u|\le \eta_n, |v|\vee|u|\le K}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_{i,n}\big[\widehat{\psi}_{\mathsf{x},n}(va_n^{-1};\mathbf{Z}_i) - \widehat{\psi}_{\mathsf{x},n}(ua_n^{-1};\mathbf{Z}_i)\big]\Big|\Big|\{\mathbf{Z}_i\}_{i=1}^n\Big]\\
&\le C\frac{\sqrt{a_n}}{n}\sum_{i=1}^n\big[\bar{D}_\gamma^{\delta_n}(\mathbf{Z}_i) + \bar{D}_\phi^{\delta_n}(\mathbf{Z}_i)\big](n_0-1)\mathbb{E}\max_{1\le i\le n}|W_{i,n}|n^{-1/2}\\
&\quad + C\max_{n_0\le k\le n}\mathbb{E}\Big[\sup_{|v-u|\le \eta_n, |v|\vee|u|\le K}\Big|\frac{1}{\sqrt{k}}\sum_{i=n_0}^k\big[\widehat{\psi}_{\mathsf{x},n}(va_n^{-1};\mathbf{Z}_{R_i}) - \widehat{\psi}_{\mathsf{x},n}(ua_n^{-1};\mathbf{Z}_{R_i})\big]\Big|\Big|\{\mathbf{Z}_i\}_{i=1}^n\Big]
\end{aligned}$$

where $(R_1, \dots, R_n)$ is uniformly distributed on the set of all permutations of $\{1, \dots, n\}$, independent of $\{\mathbf{Z}_i\}_{i=1}^n$. Choose n_0 such that $n^{1/2-1/\mathfrak{r}}/n_0 \to \infty$ and $n_0/a_n \to \infty$, which is possible by $\mathfrak{r} > (4\mathfrak{q}+2)/(2\mathfrak{q}-1)$. Following the argument of van der Vaart and Wellner (1996, Theorem 3.6.13), it suffices to bound

$$\max_{n_0\le k\le n}\mathbb{E}^*\sup_{|v-u|\le \eta_n, |v|\vee|u|\le K}\Big|\frac{1}{\sqrt{k}}\sum_{i=1}^k\big[\widehat{\psi}_{\mathsf{x},n}(va_n^{-1};\mathbf{Z}_i^*) - \widehat{\psi}_{\mathsf{x},n}(ua_n^{-1};\mathbf{Z}_i^*)\big]\Big|$$

where $\{\mathbf{Z}_i^*\}_{i=1}^k$ denotes a random sample from the empirical CDF and $\mathbb{E}^*$ is the expectation under this empirical bootstrap law. Following the argument of Kim and Pollard (1990, Lemma 4.6), it suffices to show

$$\max_{n_0\le k\le n}\mathbb{E}^*\sup_{|v-u|\le \eta_n, |v|\vee|u|\le K}\frac{1}{k}\sum_{i=1}^k\big|\widehat{\psi}_{\mathsf{x},n}(va_n^{-1};\mathbf{Z}_i^*) - \widehat{\psi}_{\mathsf{x},n}(ua_n^{-1};\mathbf{Z}_i^*)\big|^2 = o_{\mathbb{P}}(1).$$

$$\begin{aligned}
&\mathbb{E}^*\sup_{|v-u|\le \eta_n, |v|\vee|u|\le K}\frac{1}{k}\sum_{i=1}^k\big|\widehat{\psi}_{\mathsf{x},n}(va_n^{-1};\mathbf{Z}_i^*) - \widehat{\psi}_{\mathsf{x},n}(ua_n^{-1};\mathbf{Z}_i^*)\big|^2\\
&\le a_n\frac{1}{n}\sum_{i=1}^n\big[\bar{D}_\gamma^{\delta_n}(\mathbf{Z}_i)^2 + \bar{D}_\phi^{\delta_n}(\mathbf{Z}_i)^2\big]\mathbb{1}\{\bar{D}_\gamma^{\delta_n}(\mathbf{Z}_i) + \bar{D}_\phi^{\delta_n}(\mathbf{Z}_i) > C\}\\
&\quad + Ca_n\sup_{|v-u|\le \eta_n, |v|\vee|u|\le K}\frac{1}{n}\sum_{i=1}^n\big|\psi_{\mathsf{x}}(va_n^{-1};\mathbf{Z}_i) - \psi_{\mathsf{x}}(ua_n^{-1};\mathbf{Z}_i)\big|\\
&\quad + Ca_n\mathbb{E}^*\sup_{|v-u|\le \eta_n, |v|\vee|u|\le K}\frac{1}{k}\sum_{i=1}^k\big|\psi_{\mathsf{x}}(va_n^{-1};\mathbf{Z}_i^*) - \psi_{\mathsf{x}}(ua_n^{-1};\mathbf{Z}_i^*)\big|\\
&\qquad\qquad - \mathbb{E}^*\big[\big|\psi_{\mathsf{x}}(va_n^{-1};\mathbf{Z}^*) - \psi_{\mathsf{x}}(ua_n^{-1};\mathbf{Z}^*)\big|\big].
\end{aligned}$$

The first term after the inequality does not depend on k and its expectation can be made arbitrarily small by taking C sufficiently large. The second term is independent of k and we can handle this term by adding and subtracting the expectation inside the summation. For the third term, applying Theorem 4.2 of Pollard (1989) again, it is bounded by a constant multiple of

$$a_nk^{-1/2}\Big(\frac{1}{n}\sum_{i=1}^n\big[\bar{D}_\gamma^{\delta_n}(\mathbf{Z}_i)^2 + \bar{D}_\phi^{\delta_n}(\mathbf{Z}_i)^2\big]\Big)^{1/2} = O_{\mathbb{P}}\big(\sqrt{a_n/k}\big),$$

which is $o_{\mathbb{P}}(1)$ by the choice of n_0.

Verifying (SA.19) We focus on the first display. By adding and subtracting the bootstrap means,

$$\begin{aligned}
&\widehat{\Gamma}_n^*(\mathsf{x}+va_n^{-1})-\widehat{\Gamma}_n^*(\mathsf{x})-\bar{\Gamma}_n^*(\mathsf{x}+va_n^{-1})+\bar{\Gamma}_n^*(\mathsf{x})\\
&=\frac{1}{n}\sum_{i=1}^{n}(W_{i,n}-1)[\widehat{\gamma}(\mathsf{x}+va_n^{-1};\mathbf{Z}_i)-\widehat{\gamma}(\mathsf{x};\mathbf{Z}_i)-\gamma_0(\mathsf{x}+va_n^{-1};\mathbf{Z}_i)+\gamma_0(\mathsf{x};\mathbf{Z}_i)]\\
&\quad+\check{\Gamma}_n(\mathsf{x}+va_n^{-1})-\check{\Gamma}_n(\mathsf{x})-\bar{\Gamma}_n(\mathsf{x}+va_n^{-1})+\bar{\Gamma}_n(\mathsf{x})
\end{aligned}$$

where $\check{\Gamma}_n(x)=\frac{1}{n}\sum_{i=1}^{n}\widehat{\gamma}(x;\mathbf{Z}_i)$. Let

$$\begin{aligned}
\widetilde{\gamma}_n(v;\mathbf{Z})=&\,\widehat{\gamma}(\mathsf{x}+va_n^{-1};\mathbf{Z}_i)-\widehat{\gamma}(\mathsf{x};\mathbf{Z}_i)-\gamma_0(\mathsf{x}+va_n^{-1};\mathbf{Z}_i)+\gamma_0(\mathsf{x};\mathbf{Z}_i)\\
&-\check{\Gamma}_n(\mathsf{x}+va_n^{-1})+\check{\Gamma}_n(\mathsf{x})-\bar{\Gamma}_n(\mathsf{x}+va_n^{-1})+\bar{\Gamma}_n(\mathsf{x}).
\end{aligned}$$

Lemma 3.6.7 of van der Vaart and Wellner (1996) implies

$$\begin{aligned}
&\sqrt{na_n}\mathbb{E}\Bigg[\sup_{|v|\le K}\left|\frac{1}{n}\sum_{i=1}^{n}W_{i,n}\widetilde{\gamma}_n(v;\mathbf{Z}_i)\right|\Bigg|\{\mathbf{Z}_i\}_{i=1}^n\Bigg]\\
&\le C\frac{\sqrt{a_n}}{n}\sum_{i=1}^{n}\sup_{|v|\le K}|\widetilde{\gamma}_n(v;\mathbf{Z}_i)|(n_0-1)\mathbb{E}[\max_{1\le i\le n}|W_{i,n}|]n^{-1/2}\\
&\quad+C\sqrt{a_n}\max_{n_0\le k\le n}\mathbb{E}\Bigg[\sup_{|v|\le K}\left|\frac{1}{\sqrt{k}}\sum_{i=n_0}^{k}\widetilde{\gamma}_n(v;\mathbf{Z}_{R_i})\right|\Bigg|\{\mathbf{Z}_i\}_{i=1}^n\Bigg]\\
&\le C\frac{\sqrt{a_n}}{n}\sum_{i=1}^{n}\sup_{|v|\le K}|\widetilde{\gamma}_n(v;\mathbf{Z}_i)|\frac{n_0n^{\mathfrak{r}}}{\sqrt{n}} &\text{(SA.20)}\\
&\quad+C\sqrt{a_n}\max_{n_0\le k\le n}\mathbb{E}^*\sup_{|v|\le K}\left|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\widetilde{\gamma}_n(v;\mathbf{Z}_i^*)\right|. &\text{(SA.21)}
\end{aligned}$$

For (SA.20),

$$\begin{aligned}
&\frac{\sqrt{a_n}}{n}\sum_{i=1}^{n}\sup_{|v|\le K}|\widetilde{\gamma}_n(v;\mathbf{Z}_i)|\\
&\le\frac{\sqrt{a_n}}{n}\sum_{i=1}^{n}\sup_{|v|\le K}|\widehat{\gamma}(\mathsf{x}+va_n^{-1};\mathbf{Z}_i)-\widehat{\gamma}(\mathsf{x};\mathbf{Z}_i)-\gamma_0(\mathsf{x}+va_n^{-1};\mathbf{Z}_i)+\gamma_0(\mathsf{x};\mathbf{Z}_i)|\\
&\quad+\sqrt{a_n}\sup_{|v|\le K}\left|\check{\Gamma}_n(\mathsf{x}+va_n^{-1})-\check{\Gamma}_n(\mathsf{x})-\bar{\Gamma}_n(\mathsf{x}+va_n^{-1})+\bar{\Gamma}_n(\mathsf{x})\right|
\end{aligned}$$

and both terms are $o_{\mathbb{P}}(1)$ by $a_n\frac{1}{n}\sum_{i=1}^{n}\sup_{|v|\le K}|\widehat{\gamma}(\mathsf{x}+va_n^{-1};\mathbf{Z}_i)-\widehat{\gamma}(\mathsf{x};\mathbf{Z}_i)-\gamma_0(\mathsf{x}+va_n^{-1};\mathbf{Z}_i)+\gamma_0(\mathsf{x};\mathbf{Z}_i)|^2=o_{\mathbb{P}}(1)$. For (SA.21), Corollary 4.3 of Pollard (1989) implies

$$\mathbb{E}^*\sup_{|v|\le K}\left|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\widetilde{\gamma}_n(v;\mathbf{Z}_i^*)\right|\le\frac{1}{n}\sum_{i=1}^{n}\sup_{|v|\le K}|\widehat{\gamma}(\mathsf{x}+va_n^{-1};\mathbf{Z}_i)-\widehat{\gamma}(\mathsf{x};\mathbf{Z}_i)-\gamma_0(\mathsf{x}+va_n^{-1};\mathbf{Z}_i)+\gamma_0(\mathsf{x};\mathbf{Z}_i)|^2$$

and this term is $o_{\mathbb{P}}(a_n^{-1})$ by Assumption SA-6 (3).

SA.5.6.4 Assumption SA-3 (2)

$$\sup_{x\in I}\left|\widehat{\Gamma}_n^*(x)-\widehat{\Gamma}_n(x)\right|\le\sup_{x\in I}\left|\widehat{\Gamma}_n^*(x)-\bar{\Gamma}_n^*(x)\right|+\sup_{x\in I}\left|\bar{\Gamma}_n^*(x)-\bar{\Gamma}_n(x)\right|+\sup_{x\in I}\left|\bar{\Gamma}_n(x)-\widehat{\Gamma}_n(x)\right|$$

where the last term is $o_{\mathbb{P}}(1)$ by the hypothesis. For $\widehat{\Gamma}_n^*(x) - \bar{\Gamma}_n^*(x)$,

$$\begin{aligned}\sup_{x\in I}\left|\widehat{\Gamma}_n^*(x) - \bar{\Gamma}_n^*(x)\right| &\le \frac{1}{n}\sum_{i=1}^n |W_{i,n}| \sup_{x\in I}\left|\widehat{\gamma}_n(x;\mathbf{Z}_i) - \gamma_0(x;\mathbf{Z}_i)\right| \\ &\le \sqrt{\frac{1}{n}\sum_{i=1}^n |W_{i,n}|^2 \frac{1}{n}\sum_{i=1}^n \sup_{x\in I}\left|\widehat{\gamma}_n(x;\mathbf{Z}_i) - \gamma_0(x;\mathbf{Z}_i)\right|^2}\end{aligned}$$

and the last term is $o_{\mathbb{P}}(1)$ by the hypothesis. For $\bar{\Gamma}_n^*(x) - \bar{\Gamma}_n(x)$, Lemma 3.6.7 of van der Vaart and Wellner (1996) implies

$$\begin{aligned}\mathbb{E}\Big[\sup_{x\in I}\left|\bar{\Gamma}_n^*(x) - \bar{\Gamma}_n(x)\right|\Big|\{\mathbf{Z}_i\}_{i=1}^n\Big] &\le C\frac{1}{n}\sum_{i=1}^n \sup_{x\in I}|\gamma_0(x;\mathbf{Z}_i)|\mathbb{E}[\max_{1\le i\le n}|W_{i,n}|]n^{-1/2} \\ &\quad + Cn^{-1/2}\max_{\lfloor\sqrt{n}\rfloor\le k\le n}\mathbb{E}\Big[\sup_{x\in I}\Big|\frac{1}{\sqrt{k}}\sum_{i=\lfloor\sqrt{n}\rfloor}^{k}\bar{\gamma}_n(x;\mathbf{Z}_{R_i})\Big|\Big|\{\mathbf{Z}_i\}_{i=1}^n\Big]\end{aligned}$$

where $\bar{\gamma}_n(x;\mathbf{z}) = \gamma_0(x;\mathbf{z}) - \frac{1}{n}\sum_{i=1}^n \gamma_0(x;\mathbf{Z}_i)$ and $(R_1,\ldots,R_n)$ is uniformly distributed on the set of all permutations of $\{1,\ldots,n\}$, independent of $\{\mathbf{Z}_i\}_{i=1}^n$. By the same argument as for verifying Assumption SA-3 (1), it suffices to show

$$n^{-1/2}\max_{\lfloor\sqrt{n}\rfloor\le k\le n}\mathbb{E}^*\sup_{x\in I}\left|\frac{1}{\sqrt{k}}\sum_{i=1}^k \bar{\gamma}_n(x;\mathbf{Z}_i^*)\right| = o_{\mathbb{P}}(1)$$

where $\{\mathbf{Z}_i^*\}_{i=1}^k$ denotes a random sample from the empirical CDF and $\mathbb{E}^*$ is the expectation under this empirical bootstrap law. Corollary 4.3 of Pollard (1989) implies the desired result.

SA.5.6.5 Assumption SA-3 (3)

$\sup_{x\in I}|\widehat{\Phi}_n^*(x) - \widehat{\Phi}_n(x)| = o_{\mathbb{P}}(1)$ follows from the same argument as for $\widehat{\Gamma}_n^*$.

$$\begin{aligned}a_n\sup_{|v|\le K}\left|\widehat{\Phi}_n^*(\mathsf{x}+va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x}+va_n^{-1})\right| &\le a_n\left|\widehat{\Phi}_n^*(\mathsf{x}) - \widehat{\Phi}_n(\mathsf{x})\right| \\ &\quad + a_n\sup_{|v|\le K}\left|\widehat{\Phi}_n^*(\mathsf{x}+va_n^{-1}) - \widehat{\Phi}_n^*(\mathsf{x}) - \widehat{\Phi}_n(\mathsf{x}+va_n^{-1}) + \widehat{\Phi}_n(\mathsf{x})\right|\end{aligned}$$

The first term after the inequality is $o_{\mathbb{P}}(1)$ by the hypothesis. The second term is $o_{\mathbb{P}}(1)$ by the stochastic equicontinuity argument for Assumption SA-3 (1).

SA.5.7 Proof of Theorem SA-3

The proof closely follows Kosorok (2008). Let $r_n = n^{\frac{\mathfrak{q}}{2\mathfrak{q}+1}}$. Suppose for contradiction that the bootstrap approximation is consistent i.e.,

$$r_n\big(\widehat{\theta}_n^*(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x})\big) \rightsquigarrow_{\mathbb{P}} (\partial\Phi_0(\mathsf{x}))^{-1}\partial_{-}\mathsf{GCM}_{\mathbb{R}}\{\mathcal{G}_{\mathsf{x}} + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}\}(0).$$

Then, by Theorem 2.2 of Kosorok (2008), we have

$$r_n\big(\widehat{\theta}_n^*(\mathsf{x}) - \theta_0(\mathsf{x})\big) \rightsquigarrow \sqrt{2}(\partial\Phi_0(\mathsf{x}))^{-1}\partial_{-}\mathsf{GCM}_{\mathbb{R}}\{\mathcal{G}_{\mathsf{x}} + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}\}(0) \qquad \text{(SA.22)}$$

where the convergence in distribution is unconditional.

Now, using the switching lemma, $\mathbb{P}[r_n(\widehat{\theta}_n^*(\mathsf{x}) - \theta_0(\mathsf{x})) > t]$ equals

$$\mathbb{P}\Big[\sup \operatorname*{arg\,max}_{v \in \widehat{I}_n} \big\{ -\widehat{\Gamma}_n^*(v) + [\theta_0(\mathsf{x}) + r_n^{-1}t]\widehat{\Phi}_n^*(v)\big\} < (\widehat{\Phi}_n^*)^{-}\big(\widehat{\Phi}_n^*(\mathsf{x})\big)\Big]$$

and to characterize the limiting distribution of $r_n(\widehat{\theta}_n^*(\mathsf{x}) - \theta_0(\mathsf{x}))$, it suffices to look at

$$-\frac{a_n^{\mathfrak{q}+1}}{n}\sum_{i=1}^n \bar{W}_{i,n}\{\widehat{\gamma}_n(\mathsf{x} + va_n^{-1}; \mathbf{Z}_i) - \widehat{\gamma}_n(\mathsf{x}; \mathbf{Z}_i) - \theta_0(\mathsf{x})\{\widehat{\phi}_n(\mathsf{x} + va_n^{-1}; \mathbf{Z}_i) - \widehat{\phi}_n(\mathsf{x}; \mathbf{Z}_i)\}\} \tag{SA.23}$$

$$- a_n^{\mathfrak{q}+1}\big[\widehat{\Gamma}_n(\mathsf{x} + va_n^{-1}) - \widehat{\Gamma}_n(\mathsf{x}) - \theta_0(\mathsf{x})\{\widehat{\Phi}_n(\mathsf{x} + va_n^{-1}) - \widehat{\Phi}_n(\mathsf{x})\}\big] \tag{SA.24}$$

$$+ a_n t\big[\widehat{\Phi}_n^*(\mathsf{x} + va_n^{-1}) - \widehat{\Phi}_n^*(\mathsf{x})\big] \tag{SA.25}$$

where $\bar{W}_{i,n} = W_{i,n} - 1$. The term (SA.23) conditionally weakly converge to $-\mathcal{G}_x$ on compacta and the term (SA.25) converges in probability to $t\partial\Phi_0(\mathsf{x})$. The term (SA.24) weakly converges to $-\mathcal{G}_x(v) - \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v)$ unconditionally. Thus,

$$\mathbb{P}\big[r_n\big(\widehat{\theta}_n^*(\mathsf{x}) - \theta_0(\mathsf{x})\big) > t\big] \to \mathbb{P}\Big[\operatorname*{arg\,max}_{v \in \mathbb{R}} \Big\{ -\sqrt{2}\mathcal{G}_{\mathsf{x}}(v) - \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) + t\partial\Phi_0(\mathsf{x})v\Big\} < 0\Big].$$

Note $\mathcal{G}_{\mathsf{x}}(av) =_d \sqrt{|a|}\mathcal{G}_{\mathsf{x}}(v)$, and using the change of variable $v = u2^{\frac{1}{2\mathfrak{q}+1}}$, the limit distribution equals

$$\begin{aligned}
&\mathbb{P}\Big[2^{\frac{1}{2\mathfrak{q}+1}} \operatorname*{arg\,max}_{u \in \mathbb{R}} \Big\{ -\mathcal{G}_{\mathsf{x}}(u) - \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) + 2^{-\frac{\mathfrak{q}}{2\mathfrak{q}+1}} t\partial\Phi_0(\mathsf{x})u\Big\} < 0\Big] \\
&= \mathbb{P}\Big[2^{\frac{\mathfrak{q}}{2\mathfrak{q}+1}}(\partial\Phi_0(\mathsf{x}))^{-1}\partial_{-}\mathsf{GCM}_{\mathbb{R}}\{\mathcal{G}_{\mathsf{x}} + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}\}(0) > t\Big].
\end{aligned}$$

Thus,

$$r_n\big(\widehat{\theta}_n^*(\mathsf{x}) - \theta_0(\mathsf{x})\big) \rightsquigarrow 2^{\frac{\mathfrak{q}}{2\mathfrak{q}+1}}(\partial\Phi_0(\mathsf{x}))^{-1}\partial_{-}\mathsf{GCM}_{\mathbb{R}}\{\mathcal{G}_{\mathsf{x}} + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}\}(0)$$

and this limit contradicts with (SA.22) as $2^{\frac{\mathfrak{q}}{2\mathfrak{q}+1}} \neq \sqrt{2}$, proving that the bootstrap estimator $\widehat{\theta}_n^*(\mathsf{x})$ fails to approximate the limit distribution. □

SA.6 Verifying conditions in examples

In this section, we demonstrate that our general theory is easily applicable to the examples considered in Section SA.4. For this purpose, one should verify Assumptions SA-1, SA-2 (4)-(5), and SA-6 for each example. Then, Lemma SA-3 implies that our general results (Theorems SA-1 and SA-2) apply. Since it is straightforward to check Assumptions SA-1, SA-2 (5), and SA-6 (1)-(2), we focus on SA-2 (4) and SA-6 (3)-(5).

When γ_0 is known (i.e., no preliminary estimations are needed), then Assumption SA-6 (3) reduces to: for some $V \in (0, 2)$,

$$\limsup_{\varepsilon \downarrow 0} \frac{\log N_U(\varepsilon, \mathfrak{F}_\gamma)}{\varepsilon^{-V}} < \infty, \qquad \mathbb{E}[\bar{F}_\gamma(\mathbf{Z})^2] < \infty, \qquad \limsup_{\delta \downarrow 0} \frac{\mathbb{E}[\bar{D}_\gamma^\delta(\mathbf{Z})^2 + \bar{D}_\gamma^\delta(\mathbf{Z})^4]}{\delta} < \infty. \tag{SA.26}$$

An identical remark applies to ϕ_0 and Assumption SA-6 (4).

In addition, as remarked in the main paper after Lemma 2, the second display of SA-2 (4) follows from the second display of SA-6 (5), and the first display of SA-2 (4) follows from

$$\lim_{n\to\infty} \delta_n^{-1}\mathbb{E}[\psi_{\mathsf{x}_n}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}_n}(t\delta_n;\mathbf{Z})] = \mathcal{C}_{\mathsf{x}}(s,t) \tag{SA.27}$$

for $a_n\delta_n = O(1)$ and any $\mathsf{x}_n \to \mathsf{x}$. To see the second claim,

$$\begin{aligned}
&\delta_n^{-1}\big\{\mathbb{E}[\psi_{\mathsf{x}}((s+t)\delta_n;\mathbf{Z})\psi_{\mathsf{x}}((s+t)\delta_n;\mathbf{Z})] - 2\mathbb{E}[\psi_{\mathsf{x}}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}}((s+t)\delta_n;\mathbf{Z})] \\
&\qquad + \mathbb{E}[\psi_{\mathsf{x}}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}}(s\delta_n;\mathbf{Z})]\big\} \to \mathcal{C}_{\mathsf{x}}(s+t,s+t) - \mathcal{C}_{\mathsf{x}}(s+t,s) - \mathcal{C}_{\mathsf{x}}(s,s+t) + \mathcal{C}_{\mathsf{x}}(s,s)
\end{aligned}$$

and at the same time, setting $\mathsf{x}_n = \mathsf{x} + s\delta_n$,

$$\begin{aligned}
&\delta_n^{-1}\big\{\mathbb{E}[\psi_{\mathsf{x}}((s+t)\delta_n;\mathbf{Z})\psi_{\mathsf{x}}((s+t)\delta_n;\mathbf{Z})] - 2\mathbb{E}[\psi_{\mathsf{x}}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}}((s+t)\delta_n;\mathbf{Z})] \\
&\qquad + \mathbb{E}[\psi_{\mathsf{x}}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}}(s\delta_n;\mathbf{Z})]\big\} \\
&= \delta_n^{-1}\big\{\mathbb{E}[\{\psi_{\mathsf{x}_n}(t\delta_n;\mathbf{Z}) - \psi_{\mathsf{x}_n}(-s\delta_n;\mathbf{Z})\}\{\psi_{\mathsf{x}_n}(t\delta_n;\mathbf{Z}) - \psi_{\mathsf{x}_n}(-s\delta_n;\mathbf{Z})\}] \\
&\qquad + 2\mathbb{E}[\psi_{\mathsf{x}_n}(-s\delta_n;\mathbf{Z})\{\psi_{\mathsf{x}_n}(t\delta_n;\mathbf{Z}) - \psi_{\mathsf{x}_n}(-s\delta_n;\mathbf{Z})\}] + \mathbb{E}[\psi_{\mathsf{x}_n}(-s\delta_n;\mathbf{Z})\psi_{\mathsf{x}_n}(-s\delta_n;\mathbf{Z})]\big\} \\
&= \delta_n^{-1}\mathbb{E}[\psi_{\mathsf{x}_n}(t\delta_n;\mathbf{Z})\psi_{\mathsf{x}_n}(t\delta_n;\mathbf{Z})] \to \mathcal{C}_{\mathsf{x}}(t,t)
\end{aligned}$$

and thus, $\mathcal{C}_{\mathsf{x}}(s+t,s+t) - \mathcal{C}_{\mathsf{x}}(s+t,s) - \mathcal{C}_{\mathsf{x}}(s,s+t) + \mathcal{C}_{\mathsf{x}}(s,s) = \mathcal{C}_{\mathsf{x}}(t,t)$ holds. Thus, for the two displays in SA-2 (4), it suffices to check SA-6 (5) and (SA.27).

SA.6.1 Monotone density function

For monotone density estimation, Φ_0 is the identity map, so Assumption SA-6 (4) holds trivially.

SA.6.1.1 No censoring

The covariance kernel is

$$\mathcal{C}_{\mathsf{x}}(s,t) = f_0(\mathsf{x})\min\{|s|,|t|\}\mathbb{1}\{\mathrm{sign}(s) = \mathrm{sign}(t)\}.$$

SA-2 (4) It is clear that $\mathcal{C}_{\mathsf{x}}(1,1) > 0$ from $f_0(\mathsf{x}) > 0$. Also, $\mathcal{C}_{\mathsf{x}}(1,\delta)/\sqrt{\delta} = f_0(\mathsf{x})\sqrt{\delta}\mathbb{1}\{\delta > 0\}$ for $|\delta| < 1$ and $\limsup_{\delta\downarrow 0}\mathcal{C}_{\mathsf{x}}(1,\delta)/\sqrt{\delta} = 0$ holds. The remaining conditions follow from verifying SA-6 (5) below.

SA-6 (3) In this example, $\gamma_0(x;\mathbf{Z}) = \mathbb{1}\{X \le x\}$ is known, so it suffices to verify (SA.26). The uniform covering number of $\{\mathbb{1}\{\cdot \le x\} : x \in \mathbb{R}\}$ grows linearly, and an envelope function can be taken to be 1. For an envelope function of $\{\mathbb{1}\{\cdot \le x\} - \mathbb{1}\{\cdot \le \mathsf{x}\} : |x - \mathsf{x}| \le \delta\}$, we can take $\mathbb{1}\{-\delta + \mathsf{x} \le \cdot \le \mathsf{x} + \delta\}$ and the moment bound is satisfied as $\mathbb{E}[\mathbb{1}\{-\delta + \mathsf{x} \le X \le \mathsf{x} + \delta\}] \le C\delta$.

SA-6 (5) Here $\psi_{\mathsf{x}}(v;\mathbf{Z}) = \mathbb{1}\{X \le \mathsf{x} + v\} - \mathbb{1}\{X \le \mathsf{x}\} - f_0(\mathsf{x})v$. Then,

$$\frac{\mathbb{E}[|\psi_{\mathsf{x}}(v;\mathbf{Z}) - \psi_{\mathsf{x}}(v';\mathbf{Z})|]}{|v - v'|} \le \frac{\mathbb{E}[\mathbb{1}\{\mathsf{x} + \min\{v,v'\} < X \le \mathsf{x} + \max\{v,v'\}\}]}{|v - v'|} + f_0(\mathsf{x}) \le C.$$

Also, $\psi_{\mathsf{x}_n}(s\delta_n;\mathbf{Z}) = \mathbb{1}\{\min\{\mathsf{x}_n,\mathsf{x}_n+s\delta_n\} < X \leq \max\{\mathsf{x}_n,\mathsf{x}_n+s\delta_n\}\} - f_0(\mathsf{x})s\delta_n$ and

$$\begin{aligned}\psi_{\mathsf{x}_n}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}_n}(t\delta_n;\mathbf{Z}) &= \mathbb{1}\{\mathsf{x}_n < X \leq \mathsf{x}_n + \delta_n \min\{s,t\}\}\mathbb{1}\{s>0,t>0\} \\ &\quad + \mathbb{1}\{\mathsf{x}_n + \max\{s,t\} < X \leq \mathsf{x}_n\}\mathbb{1}\{s<0,t<0\} \\ &\quad - \mathbb{1}\{\min\{\mathsf{x}_n,\mathsf{x}_n+s\delta_n\} < X \leq \max\{\mathsf{x}_n,\mathsf{x}_n+s\delta_n\}\}f_0(\mathsf{x})t\delta_n \\ &\quad - \mathbb{1}\{\min\{\mathsf{x}_n,\mathsf{x}_n+t\delta_n\} < X \leq \max\{\mathsf{x}_n,\mathsf{x}_n+t\delta_n\}\}f_0(\mathsf{x})s\delta_n \\ &\quad + f_0(\mathsf{x})^2 st\delta_n^2.\end{aligned}$$

Then, for any $s,t\in\mathbb{R}$ and $\mathsf{x}_n\to\mathsf{x}$, using continuity of f_0 at x,

$$\begin{aligned}\delta_n^{-1}\mathbb{E}[\psi_{\mathsf{x}_n}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}_n}(t\delta_n;\mathbf{Z})] &= \delta_n^{-1}\int_{\mathsf{x}_n}^{\mathsf{x}_n+\delta_n\min\{s,t\}} f_0(u)du\mathbb{1}\{s>0,t>0\} \\ &\quad + \delta_n^{-1}\int_{\mathsf{x}_n+\delta_n\max\{s,t\}}^{\mathsf{x}_n} f_0(u)du\mathbb{1}\{s<0,t<0\} + o(1) \\ &= f_0(\mathsf{x})[\min\{s,t\}\mathbb{1}\{s>0,t>0\} - \max\{s,t\}\mathbb{1}\{s<0,t<0\}] + o(1)\end{aligned}$$

and by $\min\{s,t\}\mathbb{1}\{s>0,t>0\} - \max\{s,t\}\mathbb{1}\{s<0,t<0\} = \min\{|s|,|t|\}\mathbb{1}\{\mathrm{sign}(s)=\mathrm{sign}(t)\}$, the desired result holds.

SA.6.1.2 Independent censoring

Let

$$\gamma_0(x;\mathbf{Z}) = \Gamma_0(x) + S_0(x)\left[\frac{\mathbb{1}\{\tilde{X}\leq x\}\Delta}{S_0(\tilde{X})G_0(\tilde{X})} - \int_0^{\tilde{X}\wedge x}\frac{\Lambda_0(du)}{S_0(u)G_0(u)}\right].$$

The covariance kernel is

$$\mathcal{C}_{\mathsf{x}}(s,t) = \frac{f_0(\mathsf{x})}{G_0(\mathsf{x})}\min\{|s|,|t|\}\mathbb{1}\{\mathrm{sign}(s)=\mathrm{sign}(t)\}.$$

SA-2 (4) $\mathcal{C}_{\mathsf{x}}(1,1) > 0$ follows from $f_0(\mathsf{x}) > 0$. $\lim_{\delta\downarrow 0}\mathcal{C}_{\mathsf{x}}(1,\delta)/\sqrt{\delta} = 0$ follows from the same computation as in the no censoring case. The remaining conditions follow from verifying SA-6 (5) below.

SA-6 (3) We have $\widehat{\Gamma}_n = 1 - \widehat{S}_n$ with $\widehat{S}_n$ the Kaplan-Meier estimator. By Theorem 1 of Lo and Singh (1986),

$$\sup_{x\in I}\left|\widehat{\Gamma}_n(x) - \frac{1}{n}\sum_{i=1}^n\gamma_0(x;\mathbf{Z})\right| = O_{\mathbb{P}}\Big(\Big|\frac{\log n}{n}\Big|^{3/4}\Big).$$

Since $\sqrt{na_n} = n^{\frac{\mathfrak{q}+1}{2\mathfrak{q}+1}} \leq n^{2/3}$ for $\mathfrak{q}\geq 1$, $\sup_x|\widehat{\Gamma}_n(x) - \Gamma_0(x)| = o_{\mathbb{P}}(1)$ and $\sqrt{na}\sup_{|v|\leq K}|\widehat{\Gamma}_n(\mathsf{x}+va_n^{-1}) - \widehat{\Gamma}_n(\mathsf{x}) - \bar{\Gamma}_n(\mathsf{x}+va_n^{-1}) + \bar{\Gamma}_n(\mathsf{x})| = o_{\mathbb{P}}(1)$ hold.

We have

$$\widehat{\gamma}_n(x;\mathbf{Z})-\gamma_0(x;\mathbf{Z})=\widehat{F}_n(x)-F_0(x)+\big[\widehat{S}_n(x)-S_0(x)\big]\left[\frac{\mathbb{1}\{\tilde{X}\le x\}\Delta}{\widehat{S}_n(\tilde{X})\widehat{G}_n(\tilde{X})}-\int_0^{\tilde{X}\wedge x}\frac{\widehat{\Lambda}_n(du)}{\widehat{S}_n(u)\widehat{G}_n(u)}\right]$$
$$+S_0(x)\mathbb{1}\{\tilde{X}\le x\}\Delta\frac{S_0(\tilde{X})G_0(\tilde{X})-\widehat{S}_n(\tilde{X})\widehat{G}_n(\tilde{X})}{S_0(\tilde{X})G_0(\tilde{X})\widehat{S}_n(\tilde{X})\widehat{G}_n(\tilde{X})}$$
$$-S_0(x)\int_0^{\tilde{X}\wedge x}\frac{S_0(u)G_0(u)-\widehat{S}_n(u)\widehat{G}_n(u)}{S_0(u)G_0(u)\widehat{S}_n(u)\widehat{G}_n(u)}\widehat{\Lambda}_n(du)$$
$$-S_0(x)\int_0^{\tilde{X}\wedge x}\frac{[\widehat{\Lambda}_n-\Lambda_0](du)}{S_0(u)G_0(u)}.$$

Using $S_0(u_0)G_0(u_0)>0$, $\sqrt{n}\sup_{x\in I}|\widehat{S}_n(x)-S_0(x)|=O_{\mathbb{P}}(1)$, $\sqrt{n}\sup_{x\in I}|\widehat{G}_n(x)-G_0(x)|=O_{\mathbb{P}}(1)$, and $\sqrt{n}\sup_{x\in I}|\widehat{\Lambda}_n(x)-\Lambda_0(x)|=O_{\mathbb{P}}(1)$, we have $\sqrt{n}\max_{1\le i\le n}\sup_{x\in I}|\widehat{\gamma}_n(x;\mathbf{Z}_i)-\gamma_0(x;\mathbf{Z}_i)|=o_{\mathbb{P}}(1)$, which in turn implies

$$a_n\frac{1}{n}\sum_{i=1}^n\sup_{|v|\le K}\left|\widehat{\gamma}_n(\mathsf{x}+va_n^{-1};\mathbf{Z}_i)-\widehat{\gamma}_n(\mathsf{x};\mathbf{Z}_i)-\gamma_0(\mathsf{x}+va_n^{-1};\mathbf{Z}_i)+\gamma_0(\mathsf{x};\mathbf{Z}_i)\right|^2=o_{\mathbb{P}}(1).$$

For the function class $\mathfrak{F}_\gamma$, we can take $\bar{F}_\gamma(\mathbf{Z})=1+[S_0(u_0)G_0(u_0)]^{-1}[1+\Lambda_0(u_0)]$ as a constant envelope. For the function class $\{S_0(x):x\in I\}$, given $m\in\mathbb{N}$, there exists $\{x_1,\dots,x_{m+1}\}\subset I$ such that $\sup_{x\in I}\min_{l=1,\dots,m+1}|S_0(x_l)-S_0(x)|\le 1/m$, which implies the uniform covering number is bounded by a linear function. The covering numbers of $\{\mathbb{1}\{\cdot\le s\}:s\in I\}$ and $\{\int_0^{\cdot\wedge s}[S_0(u)G_0(u)]^{-1}\Lambda_0(du):s\in I\}$ are also bounded by a linear function. By Lemma 5.1 of van der Vaart and van der Laan (2006), $\limsup_{\varepsilon\downarrow 0}\log N_U(\varepsilon,\mathfrak{F}_\gamma)\varepsilon^V<\infty$ holds for $V\in(0,2)$.

Now consider the uniform covering number of $\hat{\mathfrak{F}}_\gamma$. Given a realization of $(\widehat{S}_n,\widehat{G}_n)$, the mapping $x\mapsto\int_0^{x\wedge s}[\widehat{S}_n(u)\widehat{G}_n(u)]^{-1}\widehat{\Lambda}_n(du)$ is a composition of $x\mapsto x\wedge s$ and $x\mapsto\int_0^x[\widehat{S}_n(u)\widehat{G}_n(u)]^{-1}\widehat{\Lambda}_n(du)$. The latter mapping is monotone, and the first mapping is a VC-subgraph class, and Lemma 2.6.18 of van der Vaart and Wellner (1996) implies $\{\int_0^{\cdot\wedge s}[\widehat{S}_n(u)\widehat{G}_n(u)]^{-1}\widehat{\Lambda}_n(du):s\in I\}$ is a VC-subgraph class. Note that since S_0,G_0 are bounded away from zero, $\widehat{S}_n,\widehat{G}_n$ are bounded away from zero with probability approaching one. Thus, $\limsup_{\varepsilon\downarrow 0}\log N_U(\varepsilon,\hat{\mathfrak{F}}_\gamma)\varepsilon^V=O_{\mathbb{P}}(1)$ holds.

For $s\le t\in I$,

$$|\gamma_0(s;\mathbf{Z})-\gamma_0(t;\mathbf{Z})|\le C|F_0(s)-F_0(t)|+C|\mathbb{1}\{\tilde{X}\le s\}-\mathbb{1}\{\tilde{X}\le t\}|\Delta+\int_{\tilde{X}\wedge s}^{\tilde{X}\wedge t}\frac{\Lambda_0(du)}{S_0(du)G_0(du)}$$

and we can take $D_\gamma^\delta(\mathbf{Z})$ to be a constant multiple of $\sup_{|s|\le\delta}|F_0(\mathsf{x}+s)-F_0(\mathsf{x})|+\Delta\mathbb{1}\{|\tilde{X}-\mathsf{x}|\le\delta\}+\int_{x-\delta}^{x+\delta}\Lambda_0(du)/S_0(u)G_0(u)$. For $\delta>0$ small enough,

$$\mathbb{E}[D_\gamma^\delta(\mathbf{Z})^2+D_\gamma^\delta(\mathbf{Z})^4]\le Cf_0(\mathsf{x}+\delta)\delta.$$

SA-6 (5) We have

$$\psi_{\mathsf{x}}(v;\mathbf{Z})=S_0(\mathsf{x})\frac{(\mathbb{1}\{\tilde{X}\le\mathsf{x}+v\}-\mathbb{1}\{\tilde{X}\le\mathsf{x}\})\Delta}{S_0(\tilde{X})G_0(\tilde{X})}+O(|v|)$$

where $O(|v|)$ is uniformly over small enough $|v|$. Since

$$\mathbb{E}[|\mathbb{1}\{\tilde{X} \leq \mathsf{x}+v\} - \mathbb{1}\{\tilde{X} \leq \mathsf{x}+v'\}|\Delta] = \int_{\mathsf{x}+v\wedge v'}^{\mathsf{x}+v\vee v'} G_0(u)f_0(u)du \leq C|v-v'|,$$

the first display in (5) is satisfied. For the covariance kernel,

$$\begin{aligned}
&\mathbb{E}[\psi_{\mathsf{x}_n}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}_n}(t\delta_n;\mathbf{Z})] \\
&= S_0(\mathsf{x}_n)^2\Big(\int_{\mathsf{x}_n}^{\mathsf{x}_n+\delta_n\min\{s,t\}} \frac{f_0(u)}{S_0(u)^2G_0(u)}du\mathbb{1}\{s>0,t>0\} \\
&\qquad + \int_{\mathsf{x}_n+\delta_n\max\{s,t\}}^{\mathsf{x}_n} \frac{f_0(u)}{S_0(u)^2G_0(u)}du\mathbb{1}\{s<0,t<0\}\Big) + O(\delta_n^2) \\
&= \frac{S_0(\mathsf{x}_n)^2 f_0(\mathsf{x})}{S_0(\mathsf{x})^2G_0(\mathsf{x})}\min\{s,t\}\delta_n\mathbb{1}\{s>0,t>0\} \\
&\quad - \frac{S_0(\mathsf{x}_n)^2 f_0(\mathsf{x})}{S_0(\mathsf{x})^2G_0(\mathsf{x})}\max\{s,t\}\delta_n\mathbb{1}\{s<0,t<0\} + o(\delta_n)
\end{aligned}$$

where the last equality uses continuity of (S_0, G_0, f_0) at x i.e., $\int_{\mathsf{x}_n}^{\mathsf{x}_n+\delta_n}[\frac{f_0(u)}{S_0(u)^2G_0(u)} - \frac{f_0(\mathsf{x})}{S_0(\mathsf{x})^2G_0(\mathsf{x})}]du = o(1)\delta_n$. Thus, (SA.27) holds.

SA.6.1.3 Conditionally independent case

Let

$$\gamma_0(x;\mathbf{Z}) = F_0(x|\mathbf{A}) + S_0(x|\mathbf{A})\Big[\frac{\Delta\mathbb{1}\{\tilde{X}\leq x\}}{S_0(\tilde{X}|\mathbf{A})G_0(\tilde{X}|\mathbf{A})} - \int_0^{\tilde{X}\wedge x}\frac{\Lambda_0(du|\mathbf{A})}{S_0(u|\mathbf{A})G_0(u|\mathbf{A})}\Big]$$

where $F_0(x|A) = 1 - S_0(x|A)$. The covariance kernel is

$$\mathcal{C}_{\mathsf{x}}(s,t) = \mathbb{E}\Big[\frac{f_{X|A}(\mathsf{x}|A)}{G_0(\mathsf{x}|A)}\Big]\min\{|s|,|t|\}\mathbb{1}\{\mathrm{sign}(s) = \mathrm{sign}(t)\}.$$

SA-2 (4) $\mathcal{C}_{\mathsf{x}}(1,1) > 0$ follows from $\mathbb{E}[f_{X|A}(\mathsf{x}|A)/G_0(\mathsf{x}|A)] > 0$. $\lim_{\delta\downarrow 0}\mathcal{C}_{\mathsf{x}}(1,\delta)/\sqrt{\delta} = 0$ follows from the same computation as in the no censoring case. The remaining conditions follow from verifying SA-6 (5) below.

SA-6 (3) Since $\widehat{\Lambda}_n(x|\mathbf{A}) = \int_0^x \frac{\widehat{F}_n(du|\mathbf{A})}{\widehat{S}_n(u|\mathbf{A})}$,

$$|\widehat{\Lambda}_n(x|\mathbf{A}) - \Lambda_0(x|\mathbf{A})| \leq \sup_{u\in I}|\widehat{S}_n(u|\mathbf{A})^{-1} - S_0(u|\mathbf{A})^{-1}|\widehat{F}_n(x|\mathbf{A})$$

and $a_n\frac{1}{n}\sum_{i=1}^n \sup_{x\in I}|\widehat{\Lambda}_n(x|\mathbf{A}_i) - \Lambda_0(x|\mathbf{A}_i)|^2 = o_{\mathbb{P}}(1)$ holds. Using

$$\begin{aligned}
|\widehat{\gamma}_n(x;\mathbf{Z}) - \gamma_0(x;\mathbf{Z})| &\leq C\sup_{x\in I}|\widehat{S}_n(x|\mathbf{A}) - S_0(x|\mathbf{A})| + C\sup_{x\in I}|\widehat{G}_n(x|\mathbf{A}) - G_0(x|\mathbf{A})| \\
&\quad + C\sup_{x\in I}|\widehat{\Lambda}_n(x|\mathbf{A}) - \Lambda_0(x|\mathbf{A})|,
\end{aligned}$$

$\frac{1}{n}\sum_{i=1}^n \sup_{x\in I}|\widehat{\gamma}_n(x;\mathbf{Z}_i) - \gamma_0(x;\mathbf{Z}_i)|^2 = o_{\mathbb{P}}(1)$ and $a_n\frac{1}{n}\sum_{i=1}^n\sup_{|v|\leq K}\big|\widehat{\gamma}_n(\mathsf{x}+va_n^{-1};\mathbf{Z}_i) - \widehat{\gamma}_n(\mathsf{x};\mathbf{Z}_i) - \gamma_0(\mathsf{x}+va_n^{-1};\mathbf{Z}_i) + \gamma_0(\mathsf{x};\mathbf{Z}_i)\big|^2 = o_{\mathbb{P}}(1)$ hold.

For uniform covering numbers, the class $\{S(x|\cdot) : x \in I\}$ is assumed to be a VC-subgraph class. For $\{\int_0^{\cdot\wedge x} \frac{\Lambda(du|\cdot)}{S(u|\cdot)G(u|\cdot)} : x \in I\}$ with $(S,G) \in \mathfrak{S}_n \times \mathfrak{G}_n$ and $\Lambda(x|\mathbf{A}) = \int_0^x S(du|\mathbf{A})/S(u|\mathbf{A})$,

$$\left|\int_0^{\tilde{X}\wedge x_1} \frac{\Lambda(du|\mathbf{A})}{S(u|\mathbf{A})G(u|\mathbf{A})} - \int_0^{\tilde{X}\wedge x_2} \frac{\Lambda(du|\mathbf{A})}{S(u|\mathbf{A})G(u|\mathbf{A})}\right| \leq \frac{|S(x_1|\mathbf{A}) - S(x_2|\mathbf{A})|}{\inf_{u\in I} S(u|\mathbf{A})^2 G(u|\mathbf{A})}$$

and the class $\{\int_0^{\cdot\wedge x} \frac{\Lambda(du|\cdot)}{S(u|\cdot)G(u|\cdot)} : x \in I\}$ has a desired uniform coverig number bound by Lemma 5.1 of van der Vaart and van der Laan (2006).

For $\mathsf{x} + v \in I$,

$$|\gamma_0(\mathsf{x}+v;\mathbf{Z}) - \gamma_0(\mathsf{x};\mathbf{Z})| \leq C|S_0(\mathsf{x}+v|\mathbf{A}) - S_0(\mathsf{x}|\mathbf{A})| + C\Delta|\mathbb{1}\{\tilde{X} \leq \mathsf{x}+v\} - \mathbb{1}\{\tilde{X} \leq \mathsf{x}\}|$$

and using $1 - S_0(x|\cdot) = \int_0^x f_{X|A}(u|\cdot)du$ with $f_{X|A}$ being bounded, we can take

$$\bar{D}_\gamma^\delta(\mathbf{Z}) = C\Delta\mathbb{1}\{\mathsf{x} - \delta \leq \tilde{X} \leq \mathsf{x} + \delta\} + C\delta.$$

SA-6 (5) We have

$$\psi_{\mathsf{x}}(v;\mathbf{Z}) = S_0(\mathsf{x}|\mathbf{A})\frac{(\mathbb{1}\{\tilde{X} \leq \mathsf{x}+v\} - \mathbb{1}\{\tilde{X} \leq \mathsf{x}\})\Delta}{S_0(\tilde{X}|\mathbf{A})G_0(\tilde{X}|\mathbf{A})} + O(|v|)$$

and the first display follows as in the independent censoring case. For the covariance kernel,

$$\begin{aligned}\mathbb{E}[\psi_{\mathsf{x}_n}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}_n}(t\delta_n;\mathbf{Z})] &= \mathbb{E}\left[S_0(\mathsf{x}_n|\mathbf{A})^2 \int_{\mathsf{x}_n}^{\mathsf{x}_n+\delta_n s\wedge t} \frac{f_{X|A}(u|\mathbf{A})}{S_0(u|\mathbf{A})^2 G_0(u|\mathbf{A})} du\mathbb{1}\{s>0,t>0\}\right] \\ &\quad + \mathbb{E}\left[S_0(\mathsf{x}_n|\mathbf{A})^2 \int_{\mathsf{x}_n+\delta_n s\vee t}^{\mathsf{x}_n} \frac{f_{X|A}(u|\mathbf{A})}{S_0(u|\mathbf{A})^2 G_0(u|\mathbf{A})} du\mathbb{1}\{s<0,t<0\}\right] + O(\delta_n^2)\end{aligned}$$

and $\delta_n^{-1}\mathbb{E}[\psi_{\mathsf{x}_n}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}_n}(t\delta_n;\mathbf{Z})]$ converges to $\mathbb{E}[\frac{f_{X|A}(\mathsf{x}|\mathbf{A})}{G_0(\mathsf{x}|\mathbf{A})}]|s| \wedge |t|\mathbb{1}\{\text{sign}(s) = \text{sign}(t)\}$.

SA.6.2 Monotone regression function

SA.6.2.1 Classical case

The covariance kernel is

$$\mathcal{C}_{\mathsf{x}}(s,t) = f_0(\mathsf{x})\sigma_0^2(\mathsf{x})\min\{|s|,|t|\}\mathbb{1}\{\text{sign}(s) = \text{sign}(t)\}.$$

SA-2 (4) $\mathcal{C}_{\mathsf{x}}(1,1) > 0$ follows from $f_0(\mathsf{x})\sigma_0^2(\mathsf{x}) > 0$. $\lim_{\delta\downarrow 0}\mathcal{C}_{\mathsf{x}}(1,\delta)/\sqrt{\delta} = 0$ follows from the same computation as in the monotone density estimation. The remaining conditions follow from verifying SA-6 (5) below.

SA-6 (3) In this example, $\gamma_0(x;\mathbf{Z}) = Y\mathbb{1}\{X \leq x\}$ is known, so it suffices to verify (SA.26). The uniform covering number bound is straightforward as $\{\mathbb{1}\{\cdot \leq x\} : x \in \mathbb{R}\}$ is a VC-subgraph class. An envelope function is $|Y|$, whose second moment is finite. For $x \in I_{\mathsf{x}}^\delta$, $|\gamma_0(x;\mathbf{Z}) - \gamma_0(\mathsf{x};\mathbf{Z})| \leq |Y|\mathbb{1}\{\mathsf{x} - \delta \leq X \leq \mathsf{x} + \delta\}$, which we can take as $\bar{D}_\gamma^\delta(\mathbf{Z})$. Then, for $j = 2, 4$,

$$\mathbb{E}[\bar{D}_\gamma^\delta(\mathbf{Z})^j] \leq 2^{j-1}\int_{\mathsf{x}-\delta}^{\mathsf{x}+\delta} \left(|\mu_0(x)|^j + \mathbb{E}[\varepsilon^j|X = x]\right)f_0(x)dx \leq C\delta$$

and the desired bound holds.

SA-6 (4) $\widehat{\Phi}_n(x), \widehat{\Phi}_n^*(x)$ are step functions, and the sets $\widehat{\Phi}_n(I), \widehat{\Phi}_n^*(I), \widehat{\Phi}_n^-([0,1]), (\widehat{\Phi}_n^*)^-([0,1])$ are finite and thus closed. $\phi_0(x;\mathbf{Z}) = \mathbb{1}\{X \leq x\}$ is known, so it suffices to verify the analogue of (SA.26). The argument is the same as for checking SA-6 (3) in monotone density estimation with no censoring.

SA-6 (5) We have

$$\psi_{\mathsf{x}}(v;\mathbf{Z}) = \varepsilon(\mathbb{1}\{X \leq \mathsf{x}+v\} - \mathbb{1}\{X \leq \mathsf{x}\}) + (\mu_0(X) - \mu_0(\mathsf{x}))(\mathbb{1}\{X \leq \mathsf{x}+v\} - \mathbb{1}\{X \leq \mathsf{x}\}).$$

Then,

$$\mathbb{E}[|\psi_{\mathsf{x}}(v;\mathbf{Z}) - \psi_{\mathsf{x}}(v';\mathbf{Z})|] \leq \int_{\mathsf{x}+v\wedge v'}^{\mathsf{x}+v\vee v'} [\sigma_0(x) + |\mu_0(x) - \mu_0(\mathsf{x})|] f_0(x) dx \leq C|v - v'|$$

and the first display holds. For the covariance kernel, note $|(\mu_0(X)-\mu_0(\mathsf{x}_n))(\mathbb{1}\{X \leq \mathsf{x}_n+v\}-\mathbb{1}\{X \leq \mathsf{x}_n\}| \leq |v| \sup_{|x-\mathsf{x}|\leq 2\delta} |\partial\mu_0(x)|$ for $|x_n - \mathsf{x}| \vee |v| \leq \delta$ for $\delta < 0$ small enough. Then,

$$\begin{aligned}
&\mathbb{E}[\psi_{\mathsf{x}_n}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}_n}(t\delta_n;\mathbf{Z})] \\
&= \mathbb{E}[\varepsilon^2(\mathbb{1}\{X \leq \mathsf{x}_n + s\delta_n\} - \mathbb{1}\{X \leq \mathsf{x}_n\})(\mathbb{1}\{X \leq \mathsf{x}_n + t\delta_n\} - \mathbb{1}\{X \leq \mathsf{x}_n\})] + O(\delta_n^2) \\
&= \int_{\mathsf{x}_n}^{\mathsf{x}_n+\delta_n s\wedge t} \sigma_0^2(x) f_0(x) dx \mathbb{1}\{s > 0, t > 0\} + \int_{\mathsf{x}_n+\delta_n s\vee t}^{\mathsf{x}_n} \sigma_0^2(x) f_0(x) dx \mathbb{1}\{s < 0, t < 0\} + O(\delta_n^2)
\end{aligned}$$

and

$$\begin{aligned}
\delta_n^{-1}\mathbb{E}[\psi_{\mathsf{x}_n}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}_n}(t\delta_n;\mathbf{Z})] &\to \sigma_0^2(\mathsf{x}) f_0(\mathsf{x})\big(s \wedge t\mathbb{1}\{s > 0, t > 0\} - s \vee t\mathbb{1}\{s < 0, t < 0\}\big) \\
&= \sigma_0^2(\mathsf{x}) f_0(\mathsf{x})|s| \wedge |t|\mathbb{1}\{\mathrm{sign}(s) = \mathrm{sign}(t)\}
\end{aligned}$$

as desired.

SA.6.2.2 With covariates

Let

$$\gamma_0(x;\mathbf{Z}) = \mathbb{1}\{X \leq x\}\left[\frac{\varepsilon}{g_0(X,\mathbf{A})} + \theta_0(X)\right].$$

The covariance kernel is

$$\mathcal{C}_{\mathsf{x}}(s,t) = f_0(\mathsf{x})\mathbb{E}\left[\frac{\sigma_0^2(\mathsf{x},\mathbf{A})}{g_0(\mathsf{x},\mathbf{A})}\right] \min\{|s|,|t|\}\mathbb{1}\{\mathrm{sign}(s) = \mathrm{sign}(t)\}.$$

SA-2 (4) $\mathcal{C}_{\mathsf{x}}(1,1) > 0$ follows from $f_0(\mathsf{x})\mathbb{E}[\frac{\sigma_0^2(\mathsf{x},\mathbf{A})}{g_0(\mathsf{x},\mathbf{A})}] > 0$. $\lim_{\delta\downarrow 0} \mathcal{C}_{\mathsf{x}}(1,\delta)/\sqrt{\delta} = 0$ follows from the same computation as in the monotone density estimation. The remaining conditions follow from verifying SA-6 (5) below.

SA-6 (3)

$$|\widehat{\gamma}_n(x;\mathbf{Z})-\gamma_0(x;\mathbf{Z})|\leq \mathbb{1}\{X\leq x\}\Bigg[|\varepsilon|\Big|\widehat{g}_n(X,\mathbf{A})^{-1}-g_0(X,\mathbf{A})^{-1}\Big|+\frac{|\widehat{\mu}_n(X,\mathbf{A})-\mu_0(X,\mathbf{A})|}{\widehat{g}_n(X,\mathbf{A})}$$
$$+\frac{1}{n}\sum_{j=1}^{n}|\widehat{\mu}_n(X,\mathbf{A}_j)-\mu_0(X,\mathbf{A}_j)|+\left|\frac{1}{n}\sum_{j=1}^{n}\mu_0(X,\mathbf{A}_j)-\theta_0(X)\right|\Bigg].$$

The last sum is bounded by $\sup_{x\in I}|\frac{1}{n}\sum_{j=1}^{n}\mu_0(x,\mathbf{A}_j)-\theta_0(x)|$, and this object is $O_{\mathbb{P}}(n^{-1/2})$: to see this claim, first note that Assumption SA.4.2.2 (5) and Theorem 2.7.11 of van der Vaart and Wellner (1996) imply $\limsup_{\epsilon\downarrow 0}\log N_U(\epsilon,\{\mu(x,\cdot):x\in I\})\epsilon^V<\infty$ for some $V\in(0,2)$ and Theorem 4.2 of Pollard (1989) implies $\sup_{x\in I}|\frac{1}{n}\sum_{j=1}^{n}\mu_0(x,\mathbf{A}_j)-\theta_0(x)|=O_{\mathbb{P}}(n^{-1/2})$. Then, $a_n\frac{1}{n}\sum_{i=1}^{n}\sup_{x\in I}|\widehat{\gamma}_n(x;\mathbf{Z})-\gamma_0(x;\mathbf{Z})|^2=o_{\mathbb{P}}(1)$ holds.

The uniform covering numbers of $\mathfrak{F}_\gamma,\hat{\mathfrak{F}}_{\gamma,n}$ are the same order as for $\{\mathbb{1}\{\cdot\leq x\}:x\in I\}$. For $x\in I_{\mathsf{x}}^{\delta}$, $|\gamma_0(x;\mathbf{Z})-\gamma_0(\mathsf{x};\mathbf{Z})|\leq\mathbb{1}\{\mathsf{x}-\delta\leq X\leq\mathsf{x}+\delta\}(|\varepsilon|c^{-1}+\theta_0(\mathsf{x}+\delta))$. Then, $\limsup_{\delta\downarrow 0}\mathbb{E}[\bar{D}_{\gamma}^{\delta}(\mathbf{Z})^j]\delta^{-1}<\infty$ holds for $j=2,4$.

SA-6 (4) $\phi_0(x;\mathbf{Z})=\mathbb{1}\{X\leq x\}$ is known and the same as in the classical case, so the same argument applies.

SA-6 (5) We have

$$\psi_{\mathsf{x}}(v;\mathbf{Z})=(\mathbb{1}\{X\leq\mathsf{x}+v\}-\mathbb{1}\{X\leq\mathsf{x}\})\left[\frac{\varepsilon}{g_0(X,\mathbf{A})}+\theta_0(X)-\theta_0(\mathsf{x})\right].$$

Then, for $v,v'\in[-\delta,\delta]$ with sufficiently small $\delta>0$,

$$|\psi_{\mathsf{x}}(v;\mathbf{Z})-\psi_{\mathsf{x}}(v';\mathbf{Z})|\leq|\mathbb{1}\{X\leq\mathsf{x}+v\}-\mathbb{1}\{X\leq\mathsf{x}+v'\}|\big(c^{-1}|\varepsilon|+|X-\mathsf{x}|\sup_{x\in I_{\mathsf{x}}^{\delta}}|\partial\theta_0(x)|\big)$$

and $\sup_{v\neq v'\in[-\delta_n,\delta_n]}\mathbb{E}[|\psi_{\mathsf{x}}(v;\mathbf{Z})-\psi_{\mathsf{x}}(v';\mathbf{Z})|]/|v-v'|=O(1)$ holds.

For $s\delta_n$ small enough, $\psi_{\mathsf{x}}(s\delta_n;\mathbf{Z})=(\mathbb{1}\{X\leq\mathsf{x}+s\delta_n\}-\mathbb{1}\{X\leq\mathsf{x}\})\varepsilon g_0(X,\mathbf{A})^{-1}+O(\delta_n)$ and

$$\begin{aligned}&\mathbb{E}[\psi_{\mathsf{x}}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}}(t\delta_n;\mathbf{Z})]\\&=\mathbb{E}\left[\frac{\sigma_0^2(X,\mathbf{A})}{g_0(X,\mathbf{A})^2}(\mathbb{1}\{X\leq\mathsf{x}+s\delta_n\}-\mathbb{1}\{X\leq\mathsf{x}\})(\mathbb{1}\{X\leq\mathsf{x}+t\delta_n\}-\mathbb{1}\{X\leq\mathsf{x}\})\right]+O(\delta_n^2)\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}\left[\frac{\sigma_0^2(X,\mathbf{A})}{g_0(X,\mathbf{A})^2}(\mathbb{1}\{X\leq\mathsf{x}+s\delta_n\}-\mathbb{1}\{X\leq\mathsf{x}\})(\mathbb{1}\{X\leq\mathsf{x}+t\delta_n\}-\mathbb{1}\{X\leq\mathsf{x}\})\right]\\&=\mathbb{E}\left[\int_{\mathsf{x}}^{\mathsf{x}+\delta_n s\wedge t}\frac{\sigma_0^2(x,\mathbf{A})}{g_0(x,\mathbf{A})^2}f_{X|A}(x|\mathbf{A})dx\right]\mathbb{1}\{s>0,t>0\}\\&\quad+\mathbb{E}\left[\int_{\mathsf{x}+\delta_n s\vee t}^{\mathsf{x}}\frac{\sigma_0^2(x,\mathbf{A})}{g_0(x,\mathbf{A})^2}f_{X|A}(x|\mathbf{A})dx\right]\mathbb{1}\{s<0,t<0\}\\&=\mathbb{E}\left[\frac{\sigma_0^2(\mathsf{x},\mathbf{A})}{g_0(\mathsf{x},\mathbf{A})^2}f_{X|A}(\mathsf{x}|\mathbf{A})\right]\Big(\delta_n s\wedge t\mathbb{1}\{s>0,t>0\}-\delta_n s\vee t\mathbb{1}\{s<0,t<0\}\Big)+o(\delta_n).\end{aligned}$$

Since $\frac{f_{X|A}(x|\mathbf{A})}{g_0(x,\mathbf{A})^2} = \frac{f_0(x)}{g_0(x,\mathbf{A})}$, we have

$$\delta_n^{-1}\mathbb{E}[\psi_{\mathsf{x}}(s\delta_n;\mathbf{Z})\psi_{\mathsf{x}}(t\delta_n;\mathbf{Z})] \to f_0(\mathsf{x})\mathbb{E}\left[\frac{\sigma_0^2(\mathsf{x},\mathbf{A})}{g_0(\mathsf{x},\mathbf{A})}\right]\Big(s\wedge t\mathbb{1}\{s>0,t>0\} - s\vee t\mathbb{1}\{s<0,t<0\}\Big)$$

as desired.

SA.6.3 Monotone hazard function

For both cases, we use the same $\widehat{\gamma}_n$ function and assumptions as in the corresponding monotone density setting. Also, the covariance kernels are the same as in the monotone density case. Thus, SA-6 (3) and part of SA-2 (4) follow from the same argument. We focus on SA-6 (4) and (5).

In the sequel, γ_0 denotes the function defined for the corresponding monotone density example. That is, for the independent right-censoring case,

$$\gamma_0(x;\mathbf{Z}) = F_0(x) + S_0(x)\left[\frac{\mathbb{1}\{\tilde{X}\le x\}\Delta}{S_0(\tilde{X})G_0(\tilde{X})} - \int_0^{\tilde{X}\wedge x}\frac{\Lambda_0(du)}{S_0(u)G_0(u)}\right]$$

and for the conditionally independent case,

$$\gamma_0(x;\mathbf{Z}) = F_0(x|\mathbf{A}) + S_0(x|\mathbf{A})\left[\frac{\mathbb{1}\{\tilde{X}\le x\}\Delta}{S_0(\tilde{X}|\mathbf{A})G_0(\tilde{X}|\mathbf{A})} - \int_0^{\tilde{X}\wedge x}\frac{\Lambda_0(du|\mathbf{A})}{S_0(u|\mathbf{A})G_0(u|\mathbf{A})}\right].$$

The same remark applies to $\widehat{\gamma}_n$.

SA-6 (4) Since $\widehat{\Phi}_n, \widehat{\Phi}_n^*$ are integrals with non-negative integrands, they are non-decreasing and continuous. The closedness of range follows from continuity and I being a compact interval. For closedness of $\widehat{\Phi}_n^-([0,\widehat{u}_n])$ and $(\widehat{\Phi}_n^*)^-([0,\widehat{u}_n])$, the integrands of $\widehat{\Phi}_n, \widehat{\Phi}_n^*$ are non-increasing and non-negative, and thus, $\widehat{\Phi}_n^-([0,\widehat{u}_n])$ and $(\widehat{\Phi}_n^*)^-([0,\widehat{u}_n])$ are closed intervals. Note that

$$\widehat{\phi}_n(x;\mathbf{Z}) - \phi_0(x;\mathbf{Z}) = -\int_0^x[\widehat{\gamma}_n(u;\mathbf{Z}) - \gamma_0(u;\mathbf{Z})]du$$

and thus, $\frac{1}{n}\sum_{i=1}^n \sup_{x\in I}|\widehat{\phi}_n(x;\mathbf{Z}) - \phi_0(x;\mathbf{Z})|^2 = o_{\mathbb{P}}(1)$ and $a_n\frac{1}{n}\sum_{i=1}^n\sup_{|v|\le K}|\widehat{\phi}_n(\mathsf{x}+va_n^{-1};\mathbf{Z}_i) - \widehat{\phi}_n(\mathsf{x};\mathbf{Z}_i) - \phi_0(\mathsf{x}+va_n^{-1};\mathbf{Z}_i) + \phi_0(\mathsf{x};\mathbf{Z}_i)| = o_{\mathbb{P}}(1)$ follow from the analogous conditions on $\widehat{\gamma}_n$. To check $\sup_{x\in I}|\widehat{\Phi}_n(x) - \Phi_0(x)| = o_{\mathbb{P}}(1)$, note $\sup_{x\in I}|\frac{1}{n}\sum_{i=1}^n\phi_0(x;\mathbf{Z}_i) - \Phi_0(x)| = o_{\mathbb{P}}(1)$ follows from Glivenko-Cantelli, where $\phi_0(x;\mathbf{Z}) = x - \int_0^x\gamma_0(u;\mathbf{Z})du$ and

$$\sup_{x\in I}\left|\frac{1}{n}\sum_{i=1}^n\left[\widehat{\phi}_n(x;\mathbf{Z}_i) - \phi_0(x;\mathbf{Z}_i)\right]\right| \le \frac{1}{n}\sum_{i=1}^n\sup_{x\in I}|\widehat{\gamma}_n(x;\mathbf{Z}_i) - \gamma_0(x;\mathbf{Z}_i)| = o_{\mathbb{P}}(1),$$

where the last equality follows from $\frac{1}{n}\sum_{i=1}^n\sup_{x\in I}|\widehat{\gamma}_n(x;\mathbf{Z}_i) - \gamma_0(x;\mathbf{Z}_i)|^2 = o_{\mathbb{P}}(1)$. Now $\sup_{x\in I}|\widehat{\Phi}_n(x) - \Phi_0(x)| = o_{\mathbb{P}}(1)$ follows by the triangle inequality.

The conditions on the uniform covering number hold because γ_0 and $\widehat{\gamma}_n$ are bounded (for $\widehat{\gamma}_n$, with probability approaching one) and thus $|\phi_0(x_1;\mathbf{Z}) - \phi_0(x_2;\mathbf{Z})| \le C|x_1 - x_2|$ and $|\widehat{\phi}_n(x_1;\mathbf{Z}) - \widehat{\phi}_n(x_2;\mathbf{Z})| \le C|x_1 - x_2|$ with probability approaching one. By this Lipschitz property, the condition on $\bar{D}_\phi^\delta(\mathbf{Z})$ also holds.

SA-6 (5) Let $\psi_{\mathsf{x}}^{\mathtt{MD}}(v;\mathbf{Z}) = \gamma_0(\mathsf{x}+v;\mathbf{Z}) - \gamma_0(\mathsf{x};\mathbf{Z}) - \theta_0(\mathsf{x})v$ be the ψ_{x} function for the monotone density. Then, for x sufficiently close to x and $|v|$ small enough,

$$\begin{aligned}\psi_x(v;\mathbf{z}) &\overset{def}{=} \gamma_0(x+v;\mathbf{z}) - \gamma_0(x;\mathbf{z}) - \theta_0(x)[\phi_0(x+v;\mathbf{z}) - \phi_0(x;\mathbf{z})] \\ &= \psi_x^{\mathtt{MD}}(v;\mathbf{Z}) + \theta_0(x)\int_x^{x+v}\gamma_0(u;\mathbf{Z})du = \psi_x^{\mathtt{MD}}(v;\mathbf{Z}) + O(|v|).\end{aligned}$$

Then, the same argument as in the monotone density case implies the desired result.

SA.6.4 Distribution function estimation with current status data

As noted in Section SA.4.4, by mapping the notation $(\Delta, C) \leftrightarrow (Y, X)$, the arguments in Section SA.6.2 directly apply to the generalized Grenander-type estimators considered in Section SA.4.4.

SA.7 Rule-of-Thumb Step Size Selection

Here we develop a rule-of-thumb procedure to choose a step size for the bias-reduced numerical derivative estimator in the context of isotonic regression without covariates. Specifically, we consider the numerical derivative estimator

$$\widetilde{\mathcal{D}}_{j,n}^{\mathtt{BR}}(\mathsf{x}) = \epsilon_n^{-(j+1)}\sum_{k=1}^{\underline{\mathfrak{s}}+1}\lambda_j^{\mathtt{BR}}(k)[\widehat{\Upsilon}_n(\mathsf{x}+c_k\epsilon_n) - \widehat{\Upsilon}_n(\mathsf{x})]$$

with $\underline{\mathfrak{s}} = 3$, $c_1 = 1, c_2 = -1, c_3 = 2, c_4 = -2$. Then,

$$\lambda_1^{\mathtt{BR}}(1) = \frac{2}{3} = \lambda_1^{\mathtt{BR}}(2), \quad \lambda_1^{\mathtt{BR}}(3) = -\frac{1}{24} = \lambda_1^{\mathtt{BR}}(4),$$
$$\lambda_3^{\mathtt{BR}}(1) = -\frac{1}{6} = \lambda_3^{\mathtt{BR}}(2), \quad \lambda_3^{\mathtt{BR}}(3) = \frac{1}{24} = \lambda_3^{\mathtt{BR}}(4).$$

We use the (asymptotic) MSE-optimal step size discussed in the main paper. See also SA.5.5.1. Yet, with the choice of c_k's, part of the bias constant $\sum_{k=1}^{\underline{\mathfrak{s}}+1}\lambda_j^{\mathtt{BR}}(k)c_k^{\underline{s}+2}$ equals zero, and we need to turn to the next leading term of the bias, which is

$$\epsilon_n^{\underline{\mathfrak{s}}+2-j}\frac{\partial^{\underline{\mathfrak{s}}+3}\Upsilon_0(\mathsf{x})}{(\underline{\mathfrak{s}}+3)!}\sum_{k=1}^{\underline{\mathfrak{s}}+1}\lambda_j^{\mathtt{BR}}(k)c_k^{\underline{s}+3}.$$

Then, letting $\mathsf{B}_j^{\mathtt{BR}}(\mathsf{x}) = \frac{\partial^6\Upsilon_0(\mathsf{x})}{6!}\sum_{k=1}^4\lambda_j^{\mathtt{BR}}(k)c_k^6$, the MSE-optmal step size is

$$\epsilon_{j,n}^{\mathtt{BR}} = \left(\frac{(2j+1)\mathsf{V}_j^{\mathtt{BR}}(\mathsf{x})}{2(5-j)\mathsf{B}_j^{\mathtt{BR}}(\mathsf{x})^2}\right)^{1/11} n^{-1/11}.$$

The bias and variance constants depend on unknown features of the data generating process. Specifically, $\mathsf{B}_j^{\mathtt{BR}}(\mathsf{x})$ depends on the regression function θ_0, the Lebesgue density of X, and their derivatives at $X = \mathsf{x}$ while $\mathsf{V}_j^{\mathtt{BR}}(\mathsf{x})$ is determined by the density of X and the conditional variance of the regression error $\varepsilon = Y - \theta_0(X)$ at $X = \mathsf{x}$. To operationalize the construction of the step size, we posit a simple parametric model:

$$\mathbb{E}[Y|X] = \gamma_0 + \sum_{k=1}^{5} \gamma_k (X - \mathsf{x}_0)^k, \quad X \sim \text{Normal}(\mu, \sigma^2)$$

where $\{\gamma_0, \gamma_1, \gamma_2, \gamma_4, \gamma_4, \gamma_5, \mu, \sigma\}$ are parameters to be estimated. Once we estimate the parameters of this reference model, we can construct a rule-of-thumb step size $\epsilon_{j,n}^{\mathtt{ROT}}$ by replacing $\mathsf{B}_j^{\mathtt{BR}}(\mathsf{x})$ and $\mathsf{V}_j^{\mathtt{BR}}(\mathsf{x})$ with their estimates. Note that although the bias and variance constant estimators may not be consistent for the true $\mathsf{B}_j^{\mathtt{BR}}(\mathsf{x})$ and $\mathsf{V}_j^{\mathtt{BR}}(\mathsf{x})$, the rate of $\epsilon_{j,n}^{\mathtt{ROT}}$ is MSE-optimal, and the numerical derivative estimator converges to $\mathcal{D}_j(\mathsf{x})$ sufficiently fast to satisfy (SA.1).

References

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E. (1955), "An Empirical Distribution Function for Sampling with Incomplete Information," *Annals of Mathematical Statistics*, 26, 641–647.

Cox, G. (2022), "A Generalized Argmax Theorem with Applications," Working paper.

Grenander, U. (1956), "On the Theory of Mortality Measurement: Part II," *Scandinavian Actuarial Journal*, 39, 125–153.

Groeneboom, P., and Wellner, J. A. (1992), *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkhäuser Basel.

Huang, J., and Wellner, J. A. (1995), "Estimation of a Monotone Density or Monotone hazard under Random Censoring," *Scandinavian Journal of Statistics*, 22, 3–33.

Kim, J., and Pollard, D. (1990), "Cube Root Asymptotics," *Annals of Statistics*, 18, 191–219.

Kosorok, M. R. (2008), "Bootstrapping the Grenander Estimator," in *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, Institute of Mathematical Statistics, pp. 282–292.

Lo, S.-H., and Singh, K. (1986), "The Product-Limit Estimator and the Bootstrap: Some Asymptotic Representations," *Probability Theory and Related Fields*, 71, 455–465.

Pollard, D. (1989), "Asymptotics via Empirical Processes," *Statistical Science*, 4, 341–354.

Sen, B., Banerjee, M., and Woodroofe, M. (2010), "Inconsistency of Bootstrap: The Grenander Estimator," *Annals of Statistics*, 38, 1953–1977.

van der Laan, M. J., and Robins, J. M. (2003), *Unified Methods for Censored longitudinal Data and Causality*, Springer Science & Business Media.

van der Vaart, A. W., and van der Laan, M. J. (2006), “Estimating a Survival Distribution with Current Status Data and High-Dimensional Covariates,” *International Journal of Biostatistics*, 2.

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

Westling, T., and Carone, M. (2020), “A Unified Study of Nonparametric Inference for Monotone Functions,” *Annals of Statistics*, 48, 1001–1024.

Westling, T., Gilbert, P., and Carone, M. (2020), “Causal isotonic regression,” *Journal of the Royal Statistical Society, Series B*, 82, 719–747.

Zeng, D. (2004), “Estimating Marginal Survival Function by Adjusting for Dependent Censoring using Many Covariates,” *Annals of Statistics*, 32, 1533–1555.