Bootstrap-Assisted Inference for Generalized Grenander-type Estimators

Matias D. Cattaneo[1], Michael Jansson[2], and Kenichi Nagasawa[3]

April 2023

[1]Princeton University.
[2]UC-Berkeley.
[3]University of Warwick.

Setup: Grenander-Type Estimators [Westling & Carone, 2020]

Monotone function estimators $\widehat{\theta}_n(\mathsf{x})$ at interior point x exhibit:

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

$$\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) = \frac{\partial^{\mathfrak{q}}\theta_0(\mathsf{x})\partial\Phi_0(x)}{(1+\mathfrak{q})!}v^{1+\mathfrak{q}}, \qquad \mathfrak{q} = \min\left\{j \in \mathbb{N} : \partial^j\theta_0(\mathsf{x}) \neq 0\right\}.$$

as $n \to \infty$, where:

- $\theta_0(\cdot)$ is density, regression or hazard function, among other possibilities.
- $\mathcal{G}_{\mathsf{x}}$ is zero-mean Gaussian process (nonstationary, rough cov kernel).
- $\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}$ is non-random drift function (usually quadratic, but not always).
- Φ_0 unknown non-decreasing and càdlàg function.
- $\mathrm{GCM}_I(f)$ is greatest convex minorant of function f on interval I.

Motivation: Conducting inference can be challenging.

Setup: Grenander-Type Estimators [Westling & Carone, 2020]

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

$$\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v)=\frac{\partial^{\mathfrak{q}}\theta_0(\mathsf{x})\partial\Phi_0(x)}{(1+\mathfrak{q})!}v^{1+\mathfrak{q}}, \qquad \mathfrak{q}=\min\left\{j\in\mathbb{N}:\partial^{j}\theta_0(\mathsf{x})\neq 0\right\}.$$

Examples:

- Isotonic density [Grenander, 1956] and extensions (censoring, covariates).
- Isotonic regression [Brunk, 1970] and extensions (censoring, covariates).
- Monotone Hazard [Rao, 1970] and extensions (censoring, covariates).
- Current Status [Ayer et al. 1955] and extensions (censoring, covariates).

Problem: bootstrap inconsistent [Kosorok, 2008; Sen, Banerjee & Woodroofe, 2010].

- Restore bootstrap validity: modifying the distribution used when resampling (subsampling, m-out-of-n bootstrap, smooth bootstrap).
- <u>This paper</u>: Modifying ("reshaping") functional form estimator [CJN, 2020].

Leading Example (Today): Isotonic Density Estimation

Model:

- $X_1, \ldots, X_n$ i.i.d. with support $[0, 1]$.
- $F(x) = \mathbb{P}[X_i \leq x]$ absolutely continuous.
- $\partial F(x) = f(x)$ monotone (e.g., non-decreasing).

Estimand: $\theta_0(\mathsf{x}) = f(\mathsf{x})$ for interior point x.

Estimator: for $\mathcal{F}$ the class of non-decreasing densities supported on $[0, 1]$,

$$\widehat{\theta}_n(\cdot) = \underset{f \in \mathcal{F}}{\arg\max} \sum_{i=1}^{n} \log f(X_i)$$

$$\implies \qquad \widehat{\theta}_n(\mathsf{x}) = \partial_- \mathrm{GCM}_{[0,1]}(\widehat{\Gamma}_n)(\mathsf{x}), \qquad \widehat{\Gamma}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$$

Asy Dist: under regularity, $\mathfrak{q} = 1$ and $\Phi_0(x) = x$,

$$\sqrt[3]{n}\big(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})\big) \rightsquigarrow \partial_- \mathrm{GCM}_{\mathbb{R}}(\mathcal{G}_{\mathsf{x}} + \mathcal{M}_{\mathsf{x},1})(0) \sim \sqrt[3]{4 f(\mathsf{x}) \partial f(\mathsf{x})} \underset{v \in \mathbb{R}}{\operatorname{argmin}} \left\{ \mathcal{G}(v) + v^2 \right\}$$

Second Example: Isotonic Regression Estimation

Model:

- $(Y_1, X_1), \ldots, (Y_n, X_n)$ i.i.d. with X_i on support I, and $F(x) = \mathbb{P}[X_i \leq x]$.
- $\mu(x) = \mathbb{E}[Y_i | X_i = x]$ differentiable and monotone (e.g., non-decreasing).

Estimand: $\theta_0(\mathsf{x}) = \mu(\mathsf{x})$ for interior point x.

Estimator: for $\mathcal{F}$ the class of non-decreasing functions supported on I,

$$\widehat{\theta}_n(x) = \operatorname*{argmin}_{f \in \mathcal{F}} \sum_{i=1}^{n} (Y_{(i)} - f(X_{(i)}))^2$$

$$\Longrightarrow \qquad \widehat{\theta}_n(\mathsf{x}) = \Big(\partial_- \mathrm{GCM}_{[0,1]}(\widehat{\Gamma}_n \circ \widehat{\Phi}_n^-)\Big) \circ \widehat{\Phi}_n^-(\mathsf{x}),$$

$$\widehat{\Gamma}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x) Y_i, \qquad \widehat{\Phi}_n^- = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x).$$

Asy Dist: under regularity, $\mathfrak{q} = 1$ and $\Phi_0(x) = F(x)$,

$$\sqrt[3]{n}\big(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})\big) \rightsquigarrow \frac{1}{\partial \Phi_0(\mathsf{x})} \partial_- \mathrm{GCM}_{\mathbb{R}}(\mathcal{G}_{\mathsf{x}} + \mathcal{M}_{\mathsf{x},1})(0) \sim \sqrt[3]{\frac{4\sigma^2(\mathsf{x}) \partial \mu(\mathsf{x})}{f(\mathsf{x})}} \operatorname*{argmin}_{v \in \mathbb{R}} \{\mathcal{G}(v) + v^2\}$$

Third Example: Covariate-Adjusted Isotonic Regression

Model:

- $(Y_1, X_1, \mathbf{Z}_1'), \ldots, (Y_n, X_n, \mathbf{Z}_n')$ i.i.d. with X_i on support I, and $F(x) = \mathbb{P}[X_i \leq x]$.
- $\mu(x) = \mathbb{E}[\mathbb{E}[Y_i|X_i = x, \mathbf{Z}_i]]$ differentiable and monotone (e.g., non-decreasing).

Estimand: $\theta_0(\mathsf{x}) = \mathbb{E}\left[\mathbb{E}[Y_i|X_i = \mathsf{x}, \mathbf{Z}_i]\right]$.

Estimator:

$$\widehat{\theta}_n(\mathsf{x}) = \left(\partial_- \mathrm{GCM}_{[0,1]}(\widehat{\Gamma}_n \circ \widehat{\Phi}_n^-)\right) \circ \widehat{\Phi}_n^-(\mathsf{x})$$

$$\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i \leq x)\left[\frac{Y_i - \widehat{\mu}(X_i, \mathbf{Z}_i)}{\widehat{f}(\mathbf{Z}_i|X_i)} + \frac{1}{n}\sum_{j=1}^{n} \widehat{\mu}(X_i, \mathbf{Z}_j)\right], \quad \widehat{\Phi}_n^-(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i \leq x).$$

Asy Dist: under regularity, $\mathfrak{q} = 1$ and $\Phi_0(x) = F(x)$,

$$\sqrt[3]{n}\left(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})\right) \rightsquigarrow \frac{1}{\partial \Phi_0(\mathsf{x})} \partial_- \mathrm{GCM}_{\mathbb{R}}(\mathcal{G}_{\mathsf{x}} + \mathcal{M}_{\mathsf{x},1})(0) \sim c_0(\mathsf{x}) \operatorname*{argmin}_{v \in \mathbb{R}} \left\{\mathcal{G}(v) + v^2\right\}$$

Generalized Grenander-Type Estimators: Framework

Estimand: $\theta_0(\mathsf{x})$ at interior point x, a monotone function.

Estimator:

$$\widehat{\theta}_n(\mathsf{x}) = [\partial_- \mathrm{GCM}_{J_n}(\widehat{\Gamma}_n \circ \widehat{\Phi}_n^-)] \circ \widehat{\Phi}_n(\mathsf{x}),$$

for some (application specific) $\widehat{\Gamma}_n \to_{\mathbb{P}} \Gamma_0$ and $\widehat{\Phi}_n \to_{\mathbb{P}} \Phi_0$.

Asy Dist:

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

$$\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) = \frac{\partial^{\mathfrak{q}}\theta_0(\mathsf{x})\partial\Phi_0(x)}{(1+\mathfrak{q})!}v^{1+\mathfrak{q}}, \qquad \mathfrak{q} = \min\left\{j \in \mathbb{N} : \partial^j \theta_0(\mathsf{x}) \neq 0\right\}.$$

Our Goals:

- Develop valid bootstrap-assisted & automatic distributional approximations.
- Develop valid inference procedures: e.g., confidence intervals for $\theta_0(\mathsf{x})$.

Isotonic Density Estimation: Distribution Theory

- $\theta_0(\mathsf{x}) = f(\mathsf{x})$, a monotone density with $\Gamma_0(x) = \mathbb{P}(X_i \leq x)$.
- $\widehat{\theta}_n(\mathsf{x}) = \partial_- \mathrm{GCM}_{J_n}(\widehat{\Gamma}_n)(\mathsf{x})$ with $\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i \leq x)$.

Switching Relation:

$$\mathbb{P}\left[\sqrt[3]{n}\big(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})\big) > t\right] = \mathbb{P}\left[\operatorname*{arg\,max}_{v \in S_n}\left\{\widehat{G}_{\mathsf{x},n}(v) + t\widehat{L}_{\mathsf{x},n}(v) + M_{\mathsf{x},n}(v)\right\} < 0\right],$$

where

$$\widehat{G}_{\mathsf{x},n}(v) = -n^{2/3}\Big[\widehat{\Gamma}_n(\mathsf{x} + vn^{-1/3}) - \Gamma_0(\mathsf{x} + vn^{-1/3}) - \widehat{\Gamma}_n(\mathsf{x}) - \Gamma_0(\mathsf{x})\Big]$$

$$\widehat{L}_{\mathsf{x},n}(v) = v$$

$$M_{\mathsf{x},n}(v) = -n^{2/3}\Big[\Gamma_0(\mathsf{x} + vn^{-1/3}) - \Gamma_0(\mathsf{x}) - \theta_0(\mathsf{x})vn^{-1/3}\Big]$$

and

$$S_n = [-\mathsf{x}n^{1/3}, \infty) \qquad \text{with} \qquad \mathbb{1}(v \in S_n) \to \mathbb{1}(v \in I)$$

Isotonic Density Estimation: Distribution Theory

- $\theta_0(\mathsf{x}) = f(\mathsf{x})$, a monotone density with $\Gamma_0(x) = \mathbb{P}(X_i \leq x)$.
- $\widehat{\theta}_n(\mathsf{x}) = \partial_- \mathrm{GCM}_{J_n}(\widehat{\Gamma}_n)(\mathsf{x})$ with $\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i \leq x)$.

Switching Relation:

$$\mathbb{P}\left[\sqrt[3]{n}\big(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})\big) > t\right] = \mathbb{P}\left[\underset{v \in S_n}{\arg\max}\left\{\widehat{G}_{\mathsf{x},n}(v) + t\widehat{L}_{\mathsf{x},n}(v) + M_{\mathsf{x},n}(v)\right\} < 0\right],$$

where

$$\widehat{G}_{\mathsf{x},n}(v) = -n^{2/3}\left[\widehat{\Gamma}_n(\mathsf{x} + vn^{-1/3}) - \Gamma_0(\mathsf{x} + vn^{-1/3}) - \widehat{\Gamma}_n(\mathsf{x}) - \Gamma_0(\mathsf{x})\right] \rightsquigarrow \mathcal{G}_{\mathsf{x}}(v)$$

$$\widehat{L}_{\mathsf{x},n}(v) = v$$

$$M_{\mathsf{x},n}(v) = -n^{2/3}\left[\Gamma_0(\mathsf{x} + vn^{-1/3}) - \Gamma_0(\mathsf{x}) - \theta_0(\mathsf{x})vn^{-1/3}\right] \to -\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) = -\frac{\partial f(\mathsf{x})}{2}v^2$$

which implies

$$\underset{v \in S_n}{\arg\max}\left\{\widehat{G}_{\mathsf{x},n}(v) + t\widehat{L}_{\mathsf{x},n}(v) + M_{\mathsf{x},n}(v)\right\} \rightsquigarrow \underset{v \in I}{\arg\max}\left\{\mathcal{G}_{\mathsf{x}}(v) + tv - \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v)\right\}$$

Isotonic Density Estimation: Distribution Theory

$$\begin{aligned}
\mathbb{P}\left[\sqrt[3]{n}\big(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})\big) > t\right] &= \mathbb{P}\left[\operatorname*{arg\,max}_{v \in S_n}\left\{\widehat{G}_{\mathsf{x},n}(v) + t\widehat{L}_{\mathsf{x},n}(v) + M_{\mathsf{x},n}(v)\right\} < 0\right] \\
&\to \mathbb{P}\left[\operatorname*{arg\,max}_{v \in I}\{\mathcal{G}_{\mathsf{x}}(v) + tv - \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v)\} < 0\right] \\
&= \mathbb{P}\left[\partial_{-}\mathrm{GCM}_{\mathbb{R}}(\mathcal{G}_{\mathsf{x}} + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}})(0) > t\right]
\end{aligned}$$

where

$$\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}} = \frac{\partial f(\mathsf{x})}{2} v^2$$

- In this stylized example,

$$\mathfrak{q} = 1, \qquad Phi_0(x) = x, \qquad \mathcal{G}_{\mathsf{x}}(v) = \sqrt{f(x)}\mathcal{W}(v),$$

  with $\mathcal{W}(\cdot)$ two-sided Wiener process.

Next: let's investigate what happens when we apply the nonparametric bootstrap...

Isotonic Density Estimation: Bootstrapping

- $\theta_0(\mathsf{x}) = f(\mathsf{x})$ a monotone density with $\Gamma_0(x) = \mathbb{P}(X_i \le x)$.
- $\widehat{\theta}_n(\mathsf{x}) = \partial_- \mathrm{GCM}_{J_n}(\widehat{\Gamma}_n)(\mathsf{x})$ with $\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i \le x)$.
- $\widehat{\theta}_n^*(\mathsf{x}) = \partial_- \mathrm{GCM}_{J_n}(\widehat{\Gamma}_n^*)(\mathsf{x})$ with $\widehat{\Gamma}_n^*(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i^* \le x)$.

Switching Relation:

$$\mathbb{P}^*\left[\sqrt[3]{n}(\widehat{\theta}_n^*(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x})) > t\right] = \mathbb{P}^*\left[\operatorname*{arg\,max}_{v \in S_n^*}\left\{\widehat{G}_{\mathsf{x},n}^*(v) + t\widehat{L}_{\mathsf{x},n}^*(v) + M_{\mathsf{x},n}^*(v)\right\} < 0\right],$$

where

$$\widehat{G}_{\mathsf{x},n}^*(v) = -n^{2/3}\left[\widehat{\Gamma}_n^*(\mathsf{x} + vn^{-1/3}) - \widehat{\Gamma}_n(\mathsf{x} + vn^{-1/3}) - \widehat{\Gamma}_n^*(\mathsf{x}) - \widehat{\Gamma}_n(\mathsf{x})\right]$$

$$\widehat{L}_{\mathsf{x},n}^*(v) = v$$

$$M_{\mathsf{x},n}^*(v) = -n^{2/3}\left[\widehat{\Gamma}_n(\mathsf{x} + vn^{-1/3}) - \widehat{\Gamma}_n(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x})vn^{-1/3}\right]$$

and

$$S_n^* = [-xn^{1/3}, \infty) \qquad \text{with} \qquad \mathbb{1}(v \in S_n^*) \to \mathbb{1}(v \in I)$$

Isotonic Density Estimation: Bootstrapping

- $\theta_0(\mathsf{x}) = f(\mathsf{x})$ a monotone density with $\Gamma_0(x) = \mathbb{P}(X_i \leq x)$.
- $\widehat{\theta}_n(\mathsf{x}) = \partial_- \mathrm{GCM}_{J_n}(\widehat{\Gamma}_n)(\mathsf{x})$ with $\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i \leq x)$.
- $\widehat{\theta}^*_n(\mathsf{x}) = \partial_- \mathrm{GCM}_{J_n}(\widehat{\Gamma}^*_n)(\mathsf{x})$ with $\widehat{\Gamma}^*_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X^*_i \leq x)$.

NP Bootstrap Failure:

$$\operatorname*{arg\,max}_{v \in S^*_n} \left\{ \widehat{G}^*_{\mathsf{x},n}(v) + t\widehat{L}^*_{\mathsf{x},n}(v) + M^*_{\mathsf{x},n}(v) \right\} \not\rightsquigarrow_{\mathbb{P}} \operatorname*{arg\,max}_{v \in I} \{ \mathcal{G}_{\mathsf{x},\alpha}(v) + tv + \mathcal{M}_{\mathsf{x},\alpha}(v) \}$$

because

$$\widehat{G}^*_{\mathsf{x},n}(v) = -n^{2/3}[\widehat{\Gamma}^*_n(\mathsf{x} + vn^{-1/3}) - \widehat{\Gamma}_n(\mathsf{x} + vn^{-1/3}) - \widehat{\Gamma}^*_n(\mathsf{x}) - \widehat{\Gamma}_n(\mathsf{x})] \rightsquigarrow_{\mathbb{P}} \mathcal{G}_{\mathsf{x}}(v)$$

$$\widehat{L}^*_{\mathsf{x},n}(v) = v$$

$$M^*_{\mathsf{x},n}(v) = -n^{2/3}[\widehat{\Gamma}_n(\mathsf{x} + vn^{-1/3}) - \widehat{\Gamma}_n(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x})vn^{-1/3}] \not\rightarrow_{\mathbb{P}} -\mathcal{M}_{\mathsf{x},n}(v) = -\frac{\partial f(\mathsf{x})}{2}v^2$$

- **Recall**: for Asy Dist, we instead had

$$M_{\mathsf{x},n}(v) = -n^{2/3}[\Gamma_0(\mathsf{x} + vn^{-1/3}) - \Gamma_0(\mathsf{x}) - \theta_0(\mathsf{x})vn^{-1/3}] \rightarrow -\mathcal{M}_{\mathsf{x},n}(v) = -\frac{\partial f(\mathsf{x})}{2}v^2$$

Isotonic Density Estimation: Recap and Intuition

- $\theta_0(x) = f(x)$ a monotone density with $\Gamma_0(x) = \mathbb{P}(X_i \leq x)$.
- $\widehat{\theta}_n(x) = \partial_- \mathrm{GCM}_{J_n}(\widehat{\Gamma}_n)(\mathsf{x})$ with $\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i \leq x)$.

Asymptotic Distribution:

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}[\sqrt[3]{n}(\widehat{\theta}_n(x) - \theta_0(x)) \leq t] - \mathbb{P}\left[\partial_- \mathrm{GCM}_{\mathbb{R}}(\mathcal{G}_{\mathsf{x}} + \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}})(0) \leq t\right] \right| \to 0$$

NP bootstrap: $\widehat{\theta}_n^*(x) = \partial_- \mathrm{GCM}_{J_n}(\widehat{\Gamma}_n^*)(\mathsf{x})$ with $\widehat{\Gamma}_n^*(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i^* \leq x)$ is invalid,

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}^*[\sqrt[3]{n}(\widehat{\theta}_n^*(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x})) \leq t] - \mathbb{P}[\sqrt[3]{n}\big(\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x})\big) \leq t] \right| \not\to_{\mathbb{P}} 0$$

This paper: consistency can be achieved by reshaping $\widehat{\Gamma}_n^*$.

- Intuition:
  - around x, $\widehat{\Gamma}_n(x)$ has mean $\Gamma_0(x) \approx \Gamma_0(\mathsf{x}) + f(\mathsf{x})(x-\mathsf{x}) + \frac{1}{2}\partial f(\mathsf{x})(x-\mathsf{x})^2$
  - whereas the mean of $\widehat{\Gamma}_n^*(x)$ under the bootstrap distribution is given by $\widehat{\Gamma}_n(x)$.
- Reshaping**:** Let $\partial\tilde{f}_n(\mathsf{x})$ denote a consistent estimator of $\partial f(\mathsf{x})$, then

$$\tilde{\Gamma}_n^*(x) = \widehat{\Gamma}_n^*(x) - \widehat{\Gamma}_n(x) + \widehat{\theta}_n(\mathsf{x})(x-\mathsf{x}) + \frac{1}{2}\partial\tilde{f}_n(\mathsf{x})(x-\mathsf{x})^2$$

Isotonic Density Estimation: Bootstrap Consistency

- $\theta_0(\mathsf{x}) = f(\mathsf{x})$ a monotone density with $\Gamma_0(x) = \mathbb{P}(X_i \leq x)$.
- $\widehat{\theta}_n(\mathsf{x}) = \partial_-\mathrm{GCM}_{J_n}(\widehat{\Gamma}_n)(\mathsf{x})$ with $\widehat{\Gamma}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i \leq x)$.
- $\widehat{\theta}^*_n(\mathsf{x}) = \partial_-\mathrm{GCM}_{J_n}(\widehat{\Gamma}^*_n)(\mathsf{x})$ with $\widehat{\Gamma}^*_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X^*_i \leq x)$.
- $\tilde{\theta}^*_n(\mathsf{x}) = \partial_-\mathrm{GCM}_{J_n}(\tilde{\Gamma}^*_n)(\mathsf{x})$ with

$$\tilde{\Gamma}^*_n(x) = \widehat{\Gamma}^*_n(x) - \widehat{\Gamma}_n(x) + \widehat{\theta}_n(\mathsf{x})(x - \mathsf{x}) + \frac{1}{2}\partial\tilde{f}_n(\mathsf{x})(x - \mathsf{x})^2.$$

Bootstrap-Assisted Validity:

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}^*[\tilde{\theta}^*_n(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x}) \leq t] - \mathbb{P}[\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x}) \leq t] \right| \to_{\mathbb{P}} 0$$

provided that

$$\partial\tilde{f}_n(\mathsf{x}) \to_{\mathbb{P}} \partial f(\mathsf{x}).$$

- The consistency requirement is mild and easy to achieve automatically.
- Confidence intervals based on kernel estimator perform well in simulations.

Generalized Grenander-Type Estimators: General Case

- $\theta_0(\mathsf{x})$ at interior point x, a monotone function.
- $\widehat{\theta}_n(\mathsf{x}) = [\partial_- \mathrm{GCM}_{J_n}(\widehat{\Gamma}_n \circ \widehat{\Phi}_n^-)] \circ \widehat{\Phi}_n(\mathsf{x})$, with $(\widehat{\Gamma}_n, \widehat{\Phi}_n) \to_{\mathbb{P}} (\Gamma_0, \Phi_0)$.

Asymptotic Distribution: using their Generalized Switching Relation,

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

$$\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) = \frac{\partial^{\mathfrak{q}}\theta_0(\mathsf{x})\partial\Phi_0(x)}{(1+\mathfrak{q})!}v^{1+\mathfrak{q}}, \qquad \mathfrak{q} = \min\left\{j \in \mathbb{N} : \partial^j\theta_0(\mathsf{x}) \neq 0\right\}.$$

Bootstrap-assisted Inference using reshaped estimator:

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}^*[\tilde{\theta}_n^*(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x}) \leq t] - \mathbb{P}[\widehat{\theta}_n(\mathsf{x}) - \theta_0(\mathsf{x}) \leq t]\right| \to_{\mathbb{P}} 0$$

where

$$\tilde{\theta}_n^*(\mathsf{x}) = [\partial_- \mathrm{GCM}_{J_n}(\tilde{\Gamma}_n^* \circ (\widehat{\Phi}_n^*)^-)] \circ \widehat{\Phi}_n^*(\mathsf{x})$$

with

$$\tilde{\Gamma}_n^*(x) = \widehat{\Gamma}_n^*(x) - \widehat{\Gamma}_n(x) + \widehat{\theta}_n(\mathsf{x})\widehat{\Phi}_n(x - \mathsf{x}) + \tilde{\mathcal{M}}_{x,\mathfrak{q}}(x - \mathsf{x}),$$

and provided that

$$\tilde{\mathcal{M}}_{\mathsf{x},\mathfrak{q}}(v) \to_{\mathbb{P}} \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v).$$

Grenander-Type Estimators: Bootstrap-assisted Inference

Valid percentile CI for monotone function $\theta_0(\mathsf{x})$ at interior point x:

$$\mathrm{I}^*_{1-a}(\mathsf{x}) = \left[\widehat{\theta}_n(\mathsf{x}) - q^*_{1-a/2} \quad , \quad \widehat{\theta}_n(\mathsf{x}) - q^*_{a/2}\right]$$

where

- $\widehat{\theta}_n(\mathsf{x}) = [\partial_- \mathrm{GCM}_{J_n}(\widehat{\Gamma}_n \circ \widehat{\Phi}_n^-)] \circ \widehat{\Phi}_n(\mathsf{x}), \qquad (\widehat{\Gamma}_n, \widehat{\Phi}_n) \to_{\mathbb{P}} (\Gamma_0, \Phi_0).$

- $q^*_a = \inf\{t \in \mathbb{R} : \mathbb{P}^*[\tilde{\theta}^*_n(\mathsf{x}) - \widehat{\theta}_n(\mathsf{x}) \leq t] \geq a\},$

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

$$\tilde{\Gamma}^*_n(x) = \widehat{\Gamma}^*_n(x) - \widehat{\Gamma}_n(x) + \widehat{\theta}_n(\mathsf{x})\widehat{\Phi}_n(x - \mathsf{x}) + \tilde{\mathcal{M}}_{x,\mathfrak{q}}(x - \mathsf{x}),$$

$$\tilde{\mathcal{M}}_{\mathsf{x},\mathfrak{q}}(v) \to_{\mathbb{P}} \mathcal{M}^{\mathfrak{q}}_{\mathsf{x}}(v).$$

Key outstanding issue: How to construct $\tilde{\mathcal{M}}_{\mathsf{x},\mathfrak{q}}(v)$?

Bootstrap-assisted Inference: Drift Estimation

Recall: $\widehat{\theta}_n(\mathsf{x}) = [\partial_- \mathrm{GCM}_{J_n}(\widehat{\Gamma}_n \circ \widehat{\Phi}_n^-)] \circ \widehat{\Phi}_n(\mathsf{x})$, $\qquad (\widehat{\Gamma}_n, \widehat{\Phi}_n) \to_{\mathbb{P}} (\Gamma_0, \Phi_0)$,

$$\mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v) = \frac{\partial^{\mathfrak{q}}\theta_0(\mathsf{x})\partial\Phi_0(x)}{(1+\mathfrak{q})!} v^{1+\mathfrak{q}}, \qquad \mathfrak{q} = \min\left\{j \in \mathbb{N} : \partial^j \theta_0(\mathsf{x}) \neq 0\right\}.$$

- Sometimes, $\partial^{\mathfrak{q}}\theta_0(\mathsf{x})$ and $\partial\Phi_0(\mathsf{x})$ easy to characterize (if $\mathfrak{q}$ is known!) and estimate.
  - For example, $\pi_0(\mathsf{x}) = \partial f(\mathsf{x})$ and $\partial\Phi_0(\mathsf{x}) = 1$ in Isotonic Density Estimation.

- In general, if $\mathfrak{q}$ is known, $\tilde{\mathcal{M}}_{x,\mathfrak{q}}(v)$ based on numerical derivative estimator:

$$\frac{\widehat{\partial^{\mathfrak{q}}\theta_n(\mathsf{x})\partial\Phi}_n(\mathsf{x})}{(\mathfrak{q}+1)!} = \epsilon_n^{-(\mathfrak{q}+1)} \sum_{k=1}^{\mathfrak{q}+1} (-1)^{k+\mathfrak{q}+1} \binom{\mathfrak{q}+1}{k} [\widehat{\Upsilon}_n(\mathsf{x}+k\epsilon_n) - \widehat{\Upsilon}_n(\mathsf{x})]$$

  where

  - $\widehat{\Upsilon}_n(v) = \widehat{\Gamma}_n(v) - \widehat{\theta}_n(\mathsf{x})\widehat{\Phi}_n(v)$ $\qquad$ and $\qquad$ $\epsilon_n > 0$ is a (small) tuning parameter.
  - $\tilde{\mathcal{M}}_{\mathsf{x},\mathfrak{q}}(v) \to_{\mathbb{P}} \mathcal{M}_{\mathsf{x}}^{\mathfrak{q}}(v)$ requires $\epsilon_n \to 0$ and $n\epsilon_n^{1+2\mathfrak{q}} \to \infty$.
  - Under additional conditions, MSE-optimal ϵ_n can be obtained.

- Possible to develop estimator $\tilde{\mathcal{M}}_{\mathsf{x},\mathfrak{q}}(v)$ adaptive to unknown $\mathfrak{q} \leq \bar{\mathfrak{q}}$.

	DGP 1				DGP 2				DGP 3			
	$\tilde{\mathcal{D}}_{1,n}$	$\tilde{\mathcal{D}}_{3,n}$	Coverage	Length	$\tilde{\mathcal{D}}_{1,n}$	$\tilde{\mathcal{D}}_{3,n}$	Coverage	Length	$\tilde{\mathcal{D}}_{1,n}$	$\tilde{\mathcal{D}}_{3,n}$	Coverage	Length
Standard												
			0.832	0.370			0.835	0.516			0.904	0.028
m-out-of-n												
$m = \lceil n^{1/2} \rceil$			0.900	0.413			0.909	0.583			0.940	0.031
$m = \lceil n^{2/3} \rceil$			0.872	0.399			0.879	0.556			0.921	0.029
$m = \lceil n^{4/5} \rceil$			0.856	0.391			0.862	0.544			0.913	0.029
Reshaped												
Oracle	1.000	0.000	0.942	0.393	1.00	0.000	0.943	0.549	0.000	1.000	0.943	0.029
ND known q	1.045	0.000	0.950	0.396	1.04	0.000	0.944	0.543	0.000	1.012	0.935	0.028
ND robust	1.045	0.633	0.951	0.398	1.04	0.981	0.953	0.556	0.014	1.012	0.959	0.030

Discussion and Conclusion

- Nonparametric bootstrap fails for Grenander-Type Estimators.
- Other valid resampling methods available change the bootstrap distribution.
- This paper:
  - Employs standard nonparametric bootstrap.
  - Reshapes estimator to deal with bootstrap inconsistency.
- Our method applies to many problems in econ, stats, biostats, ML and beyond:
  - Isotonic density [Grenander, 1956] and extensions (censoring, covariates).
  - Isotonic regression [Brunk, 1970] and extensions (censoring, covariates).
  - Monotone Hazard [Rao, 1970] and extensions (censoring, covariates).
  - Current Status [Ayer et al. 1955] and extensions (censoring, covariates).
- **Coming soon**: Smoothed pairwise maximum rank correlation and related problems.
  - $\sqrt{n}$-consistent U-process optimizer but with Chernoff-type distribution