Inference on Directionally Differentiable Functions

Zheng Fang — Kansas State
Andres Santos — UC San Diego

November 9, 2016

General Question

For unknown $\theta_0 \in \mathbb{D}$ and known map $\phi : \mathbb{D} \to \mathbb{E}$, we consider the parameter

$$\phi(\theta_0)$$

Given estimator $\hat{\theta}_n$ for θ_0, what are the properties of the "plug-in" estimator

$$\phi(\hat{\theta}_n)$$

Under Differentiability

- Asymptotic Distribution by Delta Method.
- Bootstrap Validity for $\hat{\theta}_n \Rightarrow$ Bootstrap Validity of $\phi(\hat{\theta}_n)$.
- **Together:** Framework for conducting inference on $\phi(\theta_0)$ through $\phi(\hat{\theta}_n)$.

Question: Is there a similar conceptual framework for nondifferentiable ϕ?

Example 1

Andrews & Soares (2010): Let $X = (X^{(1)}, X^{(2)})' \in \mathbf{R}^2$, and consider

$$\max\{E[X^{(1)}], E[X^{(2)}]\}$$

Example 1

Andrews & Soares (2010): Let $X = (X^{(1)}, X^{(2)})' \in \mathbf{R}^2$, and consider

$$\max\{E[X^{(1)}], E[X^{(2)}]\}$$

Here $\theta_0 = E[X]$ and for any $\theta = (\theta^{(1)}, \theta^{(2)})' \in \mathbf{R}^2$ set $\phi : \mathbf{R}^2 \to \mathbf{R}$ to equal

$$\phi(\theta) = \max\{\theta^{(1)}, \theta^{(2)}\}$$

Example 1

Andrews & Soares (2010): Let $X = (X^{(1)}, X^{(2)})' \in \mathbf{R}^2$, and consider

$$\max\{E[X^{(1)}], E[X^{(2)}]\}$$

Here $\theta_0 = E[X]$ and for any $\theta = (\theta^{(1)}, \theta^{(2)})' \in \mathbf{R}^2$ set $\phi : \mathbf{R}^2 \to \mathbf{R}$ to equal

$$\phi(\theta) = \max\{\theta^{(1)}, \theta^{(2)}\}$$

Given a sample $\{X_i\}_{i=1}^n$ let $\hat{\theta}_n \equiv \bar{X} \equiv \frac{1}{n}\sum_{i=1}^n X_i$, in which case we have

$$\phi(\hat{\theta}_n) = \max\{\bar{X}^{(1)}, \bar{X}^{(2)}\}$$

Example 2

Andrews & Shi (2013): For $Y \in \mathbf{R}$ and $Z \in \mathbf{R}^{d_z}$ consider testing the null

$$E[Y|Z] \leq 0$$

For appropriate $\mathcal{F} \subseteq \ell^\infty(\mathbf{R}^{d_z})$ (space of bounded functions), equivalent to

$$\sup_{f \in \mathcal{F}} E[Y f(Z)] \leq 0$$

Example 2

Andrews & Shi (2013): For $Y \in \mathbf{R}$ and $Z \in \mathbf{R}^{d_z}$ consider testing the null

$$E[Y|Z] \leq 0$$

For appropriate $\mathcal{F} \subseteq \ell^\infty(\mathbf{R}^{d_z})$ (space of bounded functions), equivalent to

$$\sup_{f \in \mathcal{F}} E[Yf(Z)] \leq 0$$

Here $\theta_0 \in \ell^\infty(\mathcal{F})$ satisfies $\theta_0(f) = E[Yf(Z)]$ and $\phi : \ell^\infty(\mathcal{F}) \to \mathbf{R}$ is given by

$$\phi(\theta) = \sup_{f \in \mathcal{F}} \theta(f)$$

Given a sample $\{Y_i, Z_i\}_{i=1}^n$ let $\hat{\theta}_n(f) \equiv \frac{1}{n} \sum_{i=1}^n Y_i f(Z_i)$, in which case

$$\phi(\hat{\theta}_n) = \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n Y_i f(Z_i)$$

Other Examples

- **Study of Convex Identified Sets**

  Beresteanu & Molinari (2008), Bontemps et al. (2012).

- **Tests of Stochastic Dominance**

  Barret & Donald (2003), Linton et al. (2010).

- **Tests of Likelihood Ratio Ordering**

  Carolan and Tebbs (2005), Beare and Moon (2013).

Other Examples

- **Study of Convex Identified Sets**

  Beresteanu & Molinari (2008), Bontemps et al. (2012).

- **Tests of Stochastic Dominance**

  Barret & Donald (2003), Linton et al. (2010).

- **Tests of Likelihood Ratio Ordering**

  Carolan and Tebbs (2005), Beare and Moon (2013).

Key Observation

In all examples ϕ is directionally differentiable whenever it is not fully differentiable

General Outline

Question: How much structure does directional differentiability provide?

General Outline

Question: How much structure does directional differentiability provide?

General Reults

- Delta Method: Mild extension to Shapiro (1991) and Dumbgen (1993).
- Bootstrap Validity: If and only if characterization under Gaussianity.
- Bootstrap Alternative: Underlying logic behind existing approaches.

General Outline

Question: How much structure does directional differentiability provide?

General Reults

- Delta Method: Mild extension to Shapiro (1991) and Dumbgen (1993).
- Bootstrap Validity: If and only if characterization under Gaussianity.
- Bootstrap Alternative: Underlying logic behind existing approaches.

Inference

- Local Size Control: Guaranteed by subadditivity of derivative.
- Theoretical Illustration: Test of whether θ_0 belongs to convex set.

Related Literature

Partial Identification

Manski (2003), Imbens & Manski (2004), Pakes et al. (2006), Chernozhukov et al. (2007), Romano & Shaikh (2008, 2010), Bugni (2010), Canay (2010), Chernozhukov et al. (2013).

Bootstrap Validity

Hall (1992), Dumbgen (1993), Andrews (2000), Horowitz (2001).

Directional Differentiability

Hirano & Porter (2012), Song (2012), Kaido (2013), Kaido & Santos (2014).

New Applications

- Stochastic Monotonicity: Seo (2015).
- Density Ratio Ordering: Beare and Shi (2015).
- Regression Kink Design: Hansen (2015).
- Transaction Cost Estimation: Jha and Wolak (2015).
- Partially Identified Welfare Changes: Lee and Bhattacharya (2015).
- Derivative Estimation: Hong and Li (2015).

Directional Differentiability

Let $\phi : \mathbb{D} \to \mathbb{E}$ with $\mathbb{D}$ and $\mathbb{E}$ Banach Spaces with norms $\| \cdot \|_{\mathbb{D}}$ and $\| \cdot \|_{\mathbb{E}}$.

Then ϕ is directionally differentiable

$$\lim_{n\to\infty} \| \frac{\phi(\theta + t_n h \quad) - \phi(\theta)}{t_n} - \phi'_\theta(h) \|_{\mathbb{E}} = 0$$

for every sequence $t_n \downarrow 0$

Directional Differentiability

Let $\phi : \mathbb{D} \to \mathbb{E}$ with $\mathbb{D}$ and $\mathbb{E}$ Banach Spaces with norms $\| \cdot \|_{\mathbb{D}}$ and $\| \cdot \|_{\mathbb{E}}$.

Then ϕ is directionally differentiable in the Hadamard sense

$$\lim_{n\to\infty} \| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h) \|_{\mathbb{E}} = 0$$

for every sequence $t_n \downarrow 0$ and $h_n \to h$

Directional Differentiability

Let $\phi : \mathbb{D} \to \mathbb{E}$ with $\mathbb{D}$ and $\mathbb{E}$ Banach Spaces with norms $\|\cdot\|_{\mathbb{D}}$ and $\|\cdot\|_{\mathbb{E}}$.

Then ϕ is directionally differentiable in the Hadamard sense tangential to $\mathbb{D}_0$

$$\lim_{n\to\infty} \| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h)\|_{\mathbb{E}} = 0$$

for every sequence $t_n \downarrow 0$ and $h_n \to h$ with $h \in \mathbb{D}_0 \subseteq \mathbb{D}$.

Directional Differentiability

Let $\phi : \mathbb{D} \to \mathbb{E}$ with $\mathbb{D}$ and $\mathbb{E}$ Banach Spaces with norms $\|\cdot\|_{\mathbb{D}}$ and $\|\cdot\|_{\mathbb{E}}$.

Then ϕ is directionally differentiable in the Hadamard sense tangential to $\mathbb{D}_0$

$$\lim_{n\to\infty} \|\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h)\|_{\mathbb{E}} = 0$$

for every sequence $t_n \downarrow 0$ and $h_n \to h$ with $h \in \mathbb{D}_0 \subseteq \mathbb{D}$.

Comments

- $\phi'_\theta : \mathbb{D}_0 \to \mathbb{D}$ is necessarily continuous and homogenous of degree one.
- But ϕ'_θ does not need to be linear as required in full differentiability.
- In fact, ϕ is Hadamard differentiable at θ if and only if ϕ'_θ is linear.

Illustration

$$\phi(\theta) = |\theta|$$

Fully Differentiable at $\theta_0 \neq 0$

- For $\theta_0 > 0$: $t_n^{-1}\{\phi(\theta_0 + t_n h) - \phi(\theta_0)\} = h \Rightarrow \phi'_{\theta_0}(h) = h$
- For $\theta_0 < 0$: $t_n^{-1}\{\phi(\theta_0 + t_n h) - \phi(\theta_0)\} = -h \Rightarrow \phi'_{\theta_0}(h) = -h$

Illustration

$$\phi(\theta) = |\theta|$$

Fully Differentiable at $\theta_0 \neq 0$

- For $\theta_0 > 0$: $t_n^{-1}\{\phi(\theta_0 + t_n h) - \phi(\theta_0)\} = h \Rightarrow \phi'_{\theta_0}(h) = h$
- For $\theta_0 < 0$: $t_n^{-1}\{\phi(\theta_0 + t_n h) - \phi(\theta_0)\} = -h \Rightarrow \phi'_{\theta_0}(h) = -h$

Directionally Differentiable at $\theta_0 = 0$

- For $h > 0$: $t_n^{-1}\{\phi(\theta_0 + t_n h) - \phi(\theta_0)\} = t_n^{-1}\{0 + t_n h - 0\} \Rightarrow \phi'_{\theta_0}(h) = h$
- For $h < 0$: $t_n^{-1}\{\phi(\theta_0 + t_n h) - \phi(\theta_0)\} = t_n^{-1}\{0 - h - 0\} \Rightarrow \phi'_{\theta_0}(h) = -h$

Putting them together: At $\theta_0 = 0$, $\phi'_{\theta_0}(h) = |h|$ for all $h \in \mathbf{R}$

Example 1 (cont)

Recall $\theta = (\theta^{(1)}, \theta^{(2)})' \in \mathbf{R}^2$, and $\phi(\theta) = \max\{\theta^{(1)}, \theta^{(2)}\}$. Then we have:

$$\phi'_\theta(h) = \begin{cases} h^{(j^*)} & \text{if } \theta^{(1)} \neq \theta^{(2)} \\ \max\{h^{(1)}, h^{(2)}\} & \text{if } \theta^{(1)} = \theta^{(2)} \end{cases} .$$

for every $h = (h^{(1)}, h^{(2)})' \in \mathbf{R}^2$ and where $j^* = \arg\max_{j \in \{1,2\}} \theta^{(j)}$.

Comments

- ϕ'_θ is always continuous and homogeneous of degree one.
- ϕ is fully differentiable except when θ is such that $\theta^{(1)} = \theta^{(2)}$.
- ϕ'_θ is linear except when θ is such that $\theta^{(1)} = \theta^{(2)}$.
- Here $\mathbb{D} = \mathbf{R}^2$, $\mathbb{E} = \mathbf{R}$ and $\mathbb{D}_0 = \mathbf{R}^2$.

Example 2 (cont)

Recall $\theta \in \ell^\infty(\mathcal{F})$, $(\theta(f) = E[Yf(Z)])$ and $\phi(\theta) = \sup_{f \in \mathcal{F}} \theta(f)$. Then:

$$\phi'_\theta(h) = \sup_{f \in \Psi_{\mathcal{F}}(\theta)} h(f)$$

for every continuous $h : \mathcal{F} \to \mathbf{R}$ and where $\Psi_{\mathcal{F}}(\theta) \equiv \arg\max_{f \in \mathcal{F}} \theta(f)$.

Comments

- ϕ is fully differentiable except when $\Psi_{\mathcal{F}}(\theta)$ is not a singleton.
- ϕ'_θ is linear except when $\Psi_{\mathcal{F}}(\theta)$ is not a singleton.
- Here $\mathbb{D} = \ell^\infty(\mathcal{F})$, $\mathbb{E} = \mathbf{R}$, and $\mathbb{D}_0 = \mathcal{C}(\mathcal{F})$.

  $\Rightarrow$ Concept of Tangential Directional Hadamard differentiability needed.

Delta Method

Assumptions (D)

(i) $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \mathbb{D}$ and for some $r_n \uparrow \infty$, $r_n\{\hat{\theta}_n - \theta_0\} \xrightarrow{L} \mathbb{G}_0$.

(ii) $\phi : \mathbb{D} \to \mathbb{E}$ is Hadamard directionally differentiable at θ_0 tangential to $\mathbb{D}_0$.

(iii) $\mathbb{G}_0$ is tight and $P(\mathbb{G}_0 \in \mathbb{D}_0) = 1$.

Discussion

- D(i) The underlying data $\{X_i\}_{i=1}^n$ need not be i.i.d.
- D(ii) As in Example 2, it can be useful to allow $\mathbb{D}_0 \neq \mathbb{D}$.
- D(iii) Limiting law must concentrate on tangential set $\mathbb{D}_0$.

Note: Requirements completely analogous to standard Delta method.

Delta Method

Theorem (Shapiro, Dumbgen) If Assumption (D) holds, then it follows that

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \xrightarrow{L} \phi'_{\theta_0}(\mathbb{G}_0)$$

Addendum If in addition ϕ'_{θ_0} can be continuously extended to $\mathbb{D}$, then

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} = \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) + o_p(1)$$

Comments

- Directional differentiability of ϕ only assumed at θ_0.
- Conditions of addendum required for $\phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\})$ to make sense.
- Automatically satisfied if $\mathbb{D}_0$ is closed under $\|\cdot\|_{\mathbb{D}}$.
- Can be used to recover asymptotic distribution in all examples.

Proof Intuition

Step 1: Let $t_n = 1/r_n$ which satisfies $t_n \downarrow 0$ since $r_n \uparrow \infty$. Then we have

$$
\begin{aligned}
r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} &= \frac{1}{t_n}\{\phi(\theta_0 + t_n \times r_n\{\hat{\theta}_n - \theta_0\}) - \phi(\theta_0)\} \\
&\approx \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\})
\end{aligned}
$$

Step 2: Since $\phi'_{\theta_0} : \mathbb{D} \to \mathbb{E}$ is continuous, use continuous mapping theorem

$$
\phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) \xrightarrow{L} \phi'_{\theta_0}(\mathbb{G}_0)
$$

Proof Intuition

Step 1: Let $t_n = 1/r_n$ which satisfies $t_n \downarrow 0$ since $r_n \uparrow \infty$. Then we have

$$\begin{aligned} r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} &= \frac{1}{t_n}\{\phi(\theta_0 + t_n \times r_n\{\hat{\theta}_n - \theta_0\}) - \phi(\theta_0)\} \\ &\approx \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) \end{aligned}$$

Step 2: Since $\phi'_{\theta_0} : \mathbb{D} \to \mathbb{E}$ is continuous, use continuous mapping theorem

$$\phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) \xrightarrow{L} \phi'_{\theta_0}(\mathbb{G}_0)$$

Key Observation

Linearity of ϕ'_{θ_0} is irrelevant in the original proof of the Delta method.

Bootstrap Setup

Problem: How can we estimate the limiting distribution for inference?

What we know

- If bootstrap "works" for $\hat{\theta}_n$ and ϕ is differentiable $\Rightarrow$ it "works" for $\phi(\hat{\theta}_n)$.
- Examples where it fails when ϕ is not differentiable.
- Takeaway: Delta method generalizes, but not bootstrap consistency.

Bootstrap Setup

Problem: How can we estimate the limiting distribution for inference?

What we know

- If bootstrap "works" for $\hat{\theta}_n$ and ϕ is differentiable $\Rightarrow$ it "works" for $\phi(\hat{\theta}_n)$.
- Examples where it fails when ϕ is not differentiable.
- Takeaway: Delta method generalizes, but not bootstrap consistency.

Questions

- Does the bootstrap always fail when ϕ is not differentiable?
- When is bootstrap consistency for $\hat{\theta}_n$ inherited by $\phi(\hat{\theta}_n)$?

Next: Formalize the general setup in order to answer these questions.

Bootstrap Setup

For Banach space $\mathbf{A}$ with norm $\|\cdot\|_{\mathbf{A}}$, denote bounded Lipschitz functions

$$BL_1(\mathbf{A}) \equiv \{f : \mathbf{A} \to \mathbf{R} : \sup_{a \in \mathbf{A}} |f(a)| \leq 1 \text{ and } |f(a_1) - f(a_2)| \leq \|a_1 - a_2\|_{\mathbf{A}}\}$$

For laws L_1 and L_2 we measure distance by the bounded Lipschitz metric

$$d_{BL}(L_1, L_2) \equiv \sup_{f \in BL_1(\mathbf{A})} |\int f(a) dL_1(a) - \int f(a) dL_2(a)|$$

Comments

- Largest discrepancy in expectations assigned to functions in $BL_1(\mathbf{A})$.
- Metrizes weak convergence. Key in showing validity of critical values.
- Bootstrap consistency $\Leftrightarrow$ distance measured by d_{BL} is $o_p(1)$.

Bootstrap Setup

Informally: Assume the "bootstrapped" version $\hat{\theta}_n^*$ "works" for original $\hat{\theta}_n$.

Assumptions (B)

(i) $\hat{\theta}_n^* : \{X_i, W_i\}_{i=1}^n \to \mathbb{D}$ with $\{W_i\}_{i=1}^n$ independent of $\{X_i\}_{i=1}^n$.

(ii) $\sup_{f \in BL_1(\mathbb{D})} |E[f(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]| = o_p(1)$.

Discussion

- B(i) Includes nonparametric, Bayesian, block, and weighted bootstrap.
- B(ii) Law of $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ conditional on data is consistent for $\mathbb{G}_0$.
- Also need mild (asymptotic) measurability requirements.

Necessary and Sufficient

Theorem Suppose $\mathbb{G}_0$ is a Gaussian measure and Assumptions (D), (B), and regularity conditions hold. Then, $\phi : \mathbb{D}_\phi \to \mathbb{E}$ is (fully) Hadamard differentiable at $\theta_0 \in \mathbb{D}_\phi$ tangential to the support of $\mathbb{G}_0$ if and only if

$$\sup_{f \in BL_1(\mathbb{E})} |E[f(r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\})|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1)$$

Necessary and Sufficient

Theorem Suppose $\mathbb{G}_0$ is a Gaussian measure and Assumptions (D), (B), and regularity conditions hold. Then, $\phi : \mathbb{D}_\phi \to \mathbb{E}$ is (fully) Hadamard differentiable at $\theta_0 \in \mathbb{D}_\phi$ tangential to the support of $\mathbb{G}_0$ if and only if

$$\sup_{f \in BL_1(\mathbb{E})} |E[f(r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\})|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1)$$

Key Implication: If the bootstrap works for $\phi(\hat{\theta}_n)$ and $\mathbb{G}_0$ is Gaussian

$\Rightarrow \phi'_{\theta_0}$ must be linear $\Rightarrow \phi'_{\theta_0}(\mathbb{G}_0)$ must be Gaussian

Corollary Suppose $\mathbb{G}_0$ is Gaussian and previous assumptions hold. Then:

If the limiting distribution of $\phi(\hat{\theta}_n)$ is not Gaussian, then the bootstrap fails

Proof Intuition

Step 1: Use the Delta method to conclude that unconditionally on $\{X_i\}_{i=1}^n$:

$$r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\}$$

Proof Intuition

Step 1: Use the Delta method to conclude that unconditionally on $\{X_i\}_{i=1}^n$:

$$r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\}$$
$$= r_n\{\phi(\hat{\theta}_n^*) - \phi(\theta_0)\} - r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}$$

Proof Intuition

Step 1: Use the Delta method to conclude that unconditionally on $\{X_i\}_{i=1}^n$:

$$\begin{aligned} r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} \\ &= r_n\{\phi(\hat{\theta}_n^*) - \phi(\theta_0)\} - r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \\ &= \phi'_{\theta_0}(r_n\{\hat{\theta}_n^* - \theta_0\}) - \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) + o_p(1) \end{aligned}$$

Proof Intuition

Step 1: Use the Delta method to conclude that unconditionally on $\{X_i\}_{i=1}^n$:

$$
\begin{aligned}
& r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} \\
& \quad = r_n\{\phi(\hat{\theta}_n^*) - \phi(\theta_0)\} - r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \\
& \quad = \phi'_{\theta_0}(r_n\{\hat{\theta}_n^* - \theta_0\}) - \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) + o_p(1) \\
& \quad = \phi'_{\theta_0}(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\} + r_n\{\hat{\theta}_n - \theta_0\}) - \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) + o_p(1)
\end{aligned}
$$

Proof Intuition

Step 1: Use the Delta method to conclude that unconditionally on $\{X_i\}_{i=1}^n$:

$$
\begin{aligned}
r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} & \\
&= r_n\{\phi(\hat{\theta}_n^*) - \phi(\theta_0)\} - r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \\
&= \phi'_{\theta_0}(r_n\{\hat{\theta}_n^* - \theta_0\}) - \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) + o_p(1) \\
&= \phi'_{\theta_0}(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\} + r_n\{\hat{\theta}_n - \theta_0\}) - \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) + o_p(1)
\end{aligned}
$$

Step 2: Study the last expression conditional on $\{X_i\}_{i=1}^n$ to conclude that

$$
\phi'_{\theta_0}(\underbrace{r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}} + \underbrace{r_n\{\hat{\theta}_n - \theta_0\}}) - \phi'_{\theta_0}(\underbrace{r_n\{\hat{\theta}_n - \theta_0\}})
$$

Proof Intuition

Step 1: Use the Delta method to conclude that unconditionally on $\{X_i\}_{i=1}^n$:

$$\begin{aligned} r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} \\ &= r_n\{\phi(\hat{\theta}_n^*) - \phi(\theta_0)\} - r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \\ &= \phi'_{\theta_0}(r_n\{\hat{\theta}_n^* - \theta_0\}) - \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) + o_p(1) \\ &= \phi'_{\theta_0}(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\} + r_n\{\hat{\theta}_n - \theta_0\}) - \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) + o_p(1) \end{aligned}$$

Step 2: Study the last expression conditional on $\{X_i\}_{i=1}^n$ to conclude that

$$\phi'_{\theta_0}(\underbrace{r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}}_{\xrightarrow{L} \mathbb{G}_0} + \underbrace{r_n\{\hat{\theta}_n - \theta_0\}}_{\to h}) - \phi'_{\theta_0}(\underbrace{r_n\{\hat{\theta}_n - \theta_0\}}_{\to h})$$

Proof Intuition

Step 1: Use the Delta method to conclude that unconditionally on $\{X_i\}_{i=1}^n$:

$$
\begin{aligned}
& r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} \\
&\qquad = r_n\{\phi(\hat{\theta}_n^*) - \phi(\theta_0)\} - r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \\
&\qquad = \phi'_{\theta_0}(r_n\{\hat{\theta}_n^* - \theta_0\}) - \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) + o_p(1) \\
&\qquad = \phi'_{\theta_0}(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\} + r_n\{\hat{\theta}_n - \theta_0\}) - \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) + o_p(1)
\end{aligned}
$$

Step 2: Study the last expression conditional on $\{X_i\}_{i=1}^n$ to conclude that

$$
\phi'_{\theta_0}(\underbrace{r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}}_{\xrightarrow{L} \mathbb{G}_0} + \underbrace{r_n\{\hat{\theta}_n - \theta_0\}}_{\to h}) - \phi'_{\theta_0}(\underbrace{r_n\{\hat{\theta}_n - \theta_0\}}_{\to h})
$$

$\Rightarrow$ Bootstrap works iff $\phi'_{\theta_0}(\mathbb{G}_0 + h) - \phi'_{\theta_0}(h)$ is equal in distribution to $\phi'_{\theta_0}(\mathbb{G}_0)$.

Proof Intuition

So Far: Bootstrap consistency is equivalent to (for any h in support of $\mathbb{G}_0$)

$$\phi'_{\theta_0}(\mathbb{G}_0 + h) - \phi'_{\theta_0}(h) \stackrel{d}{=} \phi'_{\theta_0}(\mathbb{G}_0)$$

Note: Have not used Gaussianity. Similar implication to Dumbgen (1993).

Proof Intuition

So Far: Bootstrap consistency is equivalent to (for any h in support of $\mathbb{G}_0$)

$$\phi'_{\theta_0}(\mathbb{G}_0 + h) - \phi'_{\theta_0}(h) \stackrel{d}{=} \phi'_{\theta_0}(\mathbb{G}_0)$$

Note: Have not used Gaussianity. Similar implication to Dumbgen (1993).

Step 3: (Scalar Case) Suppose $\mathbb{G}_0 \sim N(0,1)$, then for any $r > 0$ and $t \in \mathbf{R}$

$$E[\exp\{it(\phi'_{\theta_0}(\mathbb{G}_0 + rh) - \phi'_{\theta_0}(rh))\}] = C(t)$$

Proof Intuition

So Far: Bootstrap consistency is equivalent to (for any h in support of $\mathbb{G}_0$)

$$\phi'_{\theta_0}(\mathbb{G}_0 + h) - \phi'_{\theta_0}(h) \stackrel{d}{=} \phi'_{\theta_0}(\mathbb{G}_0)$$

Note: Have not used Gaussianity. Similar implication to Dumbgen (1993).

Step 3: (Scalar Case) Suppose $\mathbb{G}_0 \sim N(0,1)$, then for any $r > 0$ and $t \in \mathbf{R}$

$$E[\exp\{it(\phi'_{\theta_0}(\mathbb{G}_0 + rh) - \phi'_{\theta_0}(rh))\}] = C(t)$$
$$\Rightarrow E[\exp\{it\phi'_{\theta_0}(\mathbb{G}_0 + rh)\}] = \exp\{itr\phi'_{\theta_0}(h)\}C(t)$$

Proof Intuition

So Far: Bootstrap consistency is equivalent to (for any h in support of $\mathbb{G}_0$)

$$\phi'_{\theta_0}(\mathbb{G}_0 + h) - \phi'_{\theta_0}(h) \stackrel{d}{=} \phi'_{\theta_0}(\mathbb{G}_0)$$

Note: Have not used Gaussianity. Similar implication to Dumbgen (1993).

Step 3: (Scalar Case) Suppose $\mathbb{G}_0 \sim N(0,1)$, then for any $r > 0$ and $t \in \mathbf{R}$

$$\begin{aligned}
&E[\exp\{it(\phi'_{\theta_0}(\mathbb{G}_0 + rh) - \phi'_{\theta_0}(rh))\}] = C(t) \\
&\qquad \Rightarrow E[\exp\{it\phi'_{\theta_0}(\mathbb{G}_0 + rh)\}] = \exp\{itr\phi'_{\theta_0}(h)\}C(t) \\
&\qquad \Rightarrow \frac{1}{\sqrt{2\pi}} \int \exp\{it\phi'_{\theta_0}(u)\} \exp\{-\frac{1}{2}(u - rh)^2\}du = \exp\{itr\phi'_{\theta_0}(h)\}C(t)
\end{aligned}$$

Proof Intuition

So Far: Bootstrap consistency is equivalent to (for any h in support of $\mathbb{G}_0$)

$$\phi'_{\theta_0}(\mathbb{G}_0 + h) - \phi'_{\theta_0}(h) \stackrel{d}{=} \phi'_{\theta_0}(\mathbb{G}_0)$$

Note: Have not used Gaussianity. Similar implication to Dumbgen (1993).

Step 3: (Scalar Case) Suppose $\mathbb{G}_0 \sim N(0,1)$, then for any $r > 0$ and $t \in \mathbf{R}$

$$\begin{aligned} &E[\exp\{it(\phi'_{\theta_0}(\mathbb{G}_0 + rh) - \phi'_{\theta_0}(rh))\}] = C(t) \\ &\quad \Rightarrow E[\exp\{it\phi'_{\theta_0}(\mathbb{G}_0 + rh)\}] = \exp\{itr\phi'_{\theta_0}(h)\}C(t) \\ &\quad \Rightarrow \frac{1}{\sqrt{2\pi}} \int \exp\{it\phi'_{\theta_0}(u)\} \exp\{-\frac{1}{2}(u - rh)^2\}du = \exp\{itr\phi'_{\theta_0}(h)\}C(t) \end{aligned}$$

$\Rightarrow$ Differentiate both sides w.r.t r to conclude $\phi'_{\theta_0}(h)$ is linear in h.

Proof Intuition

So Far: Bootstrap consistency is equivalent to (for any h in support of $\mathbb{G}_0$)

$$\phi'_{\theta_0}(\mathbb{G}_0 + h) - \phi'_{\theta_0}(h) \stackrel{d}{=} \phi'_{\theta_0}(\mathbb{G}_0)$$

Note: Have not used Gaussianity. Similar implication to Dumbgen (1993).

Step 3: (Scalar Case) Suppose $\mathbb{G}_0 \sim N(0,1)$, then for any $r > 0$ and $t \in \mathbf{R}$

$$\begin{aligned} &E[\exp\{it(\phi'_{\theta_0}(\mathbb{G}_0 + rh) - \phi'_{\theta_0}(rh))\}] = C(t) \\ &\qquad \Rightarrow E[\exp\{it\phi'_{\theta_0}(\mathbb{G}_0 + rh)\}] = \exp\{itr\phi'_{\theta_0}(h)\}C(t) \\ &\qquad \Rightarrow \frac{1}{\sqrt{2\pi}}\int \exp\{it\phi'_{\theta_0}(u)\}\exp\{-\frac{1}{2}(u - rh)^2\}du = \exp\{itr\phi'_{\theta_0}(h)\}C(t) \end{aligned}$$

$\Rightarrow$ Differentiate both sides w.r.t r to conclude $\phi'_{\theta_0}(h)$ is linear in h.

Step 4: Generalize scalar case by arguing as above through dual space $\mathbb{E}^*$.

Bootstrap Alternative

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \xrightarrow{L} \phi'_{\theta_0}(\mathbb{G}_0)$$

Intuition

- We need to estimate ϕ'_{θ_0} and the law of $\mathbb{G}_0$.
- By assumption $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ provides an estimate of the law of $\mathbb{G}_0$.
- Bootstrap fails for $\phi(\hat{\theta}_n)$ because it does not estimate ϕ'_{θ_0} appropriately.

Fix: For an estimator $\hat{\phi}'_n$ of ϕ'_{θ_0}, use the law conditional on $\{X_i\}_{i=1}^n$ of

$$\hat{\phi}'_n(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})$$

Bootstrap Alternative

Assumption (E) For every compact $K \subseteq \mathbb{D}_0$ and $\epsilon > 0$, $\hat{\phi}'_n : \mathbb{D} \to \mathbb{E}$ satisfies

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P\Big(\sup_{h \in K^\delta} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > \epsilon \Big) = 0$$

Discussion

- δ-enlargement needed because $r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}$ may not belong to $\mathbb{D}_0$.
- δ can sometimes be dropped – i.e. $\sup_{h \in K} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} = o_p(1)$.
- If $\hat{\phi}'_n$ is "smooth", then $\hat{\phi}'_n(h) = \phi'_{\theta_0}(h) + o_p(1)$ for all $h \in \mathbb{D}_0$ suffices.

Takeaway: In all examples additional structure makes (E) easy to verify.

Bootstrap Alternative

Theorem If Assumptions (B), (D), (E), and regularity conditions hold, then

$$\sup_{f \in BL_1(\mathbb{E})} |E[f(\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1)$$

Comments

- The law of $\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\})$ conditional $\{X_i\}_{i=1}^n$ consistent for $\phi'_{\theta_0}(\mathbb{G}_0)$.
- Implies consistency of critical values under standard conditions.
- The fact that ϕ'_{θ_0} is a directional derivative is never exploited ...

  $\Rightarrow$ More generally, a method for estimating distributions of the form

$$\tau(\mathbb{G}_0)$$

  where $\mathbb{G}_0$ is tight and $\tau : \mathbb{D} \to \mathbb{E}$ is an unknown continuous map.

Example 1 (cont)

Recall $\theta_0 = (E[X^{(1)}], E[X^{(2)}])'$ and for $j^* = \arg\max_{j \in \{1,2\}} E[X^{(j)}]$ we had:

$$\phi'_{\theta_0}(h) = \begin{cases} h^{(j^*)} & \text{if } E[X^{(1)}] \neq E[X^{(2)}] \\ \max\{h^{(1)}, h^{(2)}\} & \text{if } E[X^{(1)}] = E[X^{(2)}] \end{cases} .$$

Let $\hat{j}^* = \arg\max_{j \in \{1,2\}} \bar{X}^{(j)}$ and letting $\kappa_n \uparrow \infty$ satisfy $\kappa_n/\sqrt{n} \downarrow 0$ define

$$\hat{\phi}'_n(h) = \begin{cases} h^{(\hat{j}^*)} & \text{if } |\bar{X}^{(1)} - \bar{X}^{(2)}| > \kappa_n \\ \max\{h^{(1)}, h^{(2)}\} & \text{if } |\bar{X}^{(1)} - \bar{X}^{(2)}| \leq \kappa_n \end{cases} .$$

Comments

- $\hat{\phi}'_n$ trivially satisfies Assumption (E).
- $\hat{\phi}'_n(\sqrt{n}\{\bar{X}^* - \bar{X}\})$ reduces to generalized moment selection.

Example 2 (cont)

Recall $\theta_0(f) = E[Yf(Z)]$ and for $\Psi_{\mathcal{F}}(\theta) \equiv \arg\max_{f\in\mathcal{F}} \theta(f)$ we had that:

$$\phi'_\theta(h) = \sup_{f\in\Psi_{\mathcal{F}}(\theta)} h(f)$$

Suppose $\hat{\Psi}_{\mathcal{F}}(\theta_0)$ satisfies $d_H(\Psi_{\mathcal{F}}(\theta_0), \hat{\Psi}_{\mathcal{F}}(\theta_0), \|\cdot\|_{L^2(Z)}) = o_p(1)$, and let

$$\hat{\phi}'_n(h) = \sup_{f\in\hat{\Psi}_{\mathcal{F}}(\theta_0)} h(f)$$

Comments

- Easy to show $\hat{\phi}'_n$ satisfies Assumption (E).
- $\hat{\phi}'_n(\sqrt{n}\{\hat{\theta}^*_n - \hat{\theta}_n\})$ becomes special case of Andrews & Shi (2013).
- Also: Linton et al. (2010), Kaido (2013), Beare & Shi (2013).

Testing Implications

$$H_0 : \phi(\theta_0) \leq 0 \qquad\qquad H_1 : \phi(\theta_0) > 0$$

Proposed Test

- Employ $\sqrt{n}\phi(\hat{\theta}_n)$ as a test statistic.
- Unfeasible: $c_{1-\alpha}$ the $1-\alpha$ quantile of $\phi'_{\theta_0}(\mathbb{G}_0)$ (pointwise in P).
- Use $\hat{c}_{1-\alpha}$: the $1-\alpha$ quantile of $\hat{\phi}'_n(\sqrt{n}\{\hat{\theta}^*_n - \hat{\theta}_n\})$ conditional $\{X_i\}_{i=1}^n$.

Testing Implications

$$H_0 : \phi(\theta_0) \leq 0 \qquad\qquad H_1 : \phi(\theta_0) > 0$$

Proposed Test

- Employ $\sqrt{n}\phi(\hat{\theta}_n)$ as a test statistic.
- Unfeasible: $c_{1-\alpha}$ the $1-\alpha$ quantile of $\phi'_{\theta_0}(\mathbb{G}_0)$ (pointwise in P).
- Use $\hat{c}_{1-\alpha}$: the $1-\alpha$ quantile of $\hat{\phi}'_n(\sqrt{n}\{\hat{\theta}^*_n - \hat{\theta}_n\})$ conditional $\{X_i\}_{i=1}^n$.

Problem: So far analysis is pointwise in underlying distribution of $\{X_i\}_{i=1}^n$.

Goal: Examine when pointwise in P justified test can control size locally.

Local Setup

Assumption (L)

(i) $\{X_i\}_{i=1}^n$ is i.i.d. and $X_i \sim P \in \mathbf{P}$.

(ii) $\theta_0 \equiv \theta(P)$ for some known function $\theta : \mathbf{P} \to \mathbb{D}$.

(iii) $\hat{\theta}_n$ is a regular estimator for $\theta(P)$.

(iv) $P_{n,\lambda} \in \mathbf{P}$ and $\bigotimes_{i=1}^n P_{n,\lambda}$ is contiguous to $\bigotimes_{i=1}^n P$.

(v) For $\theta' : \Lambda \to \mathbb{D}$ linear, $\|r_n\{\theta(P_{n,\lambda}) - \theta(P)\} - \theta'(\lambda)\|_{\mathbb{D}} = o(1)$.

Discussion

- L(i) Imposed for notational simplicity.
- L(iii) Allows us to focus on irregularity generated by ϕ.
- L(v) Closely related to L(iii) - van der Vaart (1991).

Example 1 Intuition

$$H_0 : \max\{E[X^{(1)}], E[X^{(2)}]\} \leq 0 \qquad H_1 : \max\{E[X^{(1)}], E[X^{(2)}]\} > 0$$

But local to P with $\theta^{(1)}(P) = \theta^{(2)}(P) = 0$, set $\theta(P_{n,\lambda}) = \theta(P) + \lambda/\sqrt{n}$ to get

$$\sqrt{n}\phi(\hat{\theta}_n) \stackrel{L_n}{\to} \max\{\mathbb{G}_0^{(1)} + \lambda^{(1)}, \mathbb{G}_0^{(2)} + \lambda^{(2)}\}$$

Example 1 Intuition

$$H_0 : \max\{E[X^{(1)}], E[X^{(2)}]\} \leq 0 \qquad H_1 : \max\{E[X^{(1)}], E[X^{(2)}]\} > 0$$

But local to P with $\theta^{(1)}(P) = \theta^{(2)}(P) = 0$, set $\theta(P_{n,\lambda}) = \theta(P) + \lambda/\sqrt{n}$ to get

$$\begin{aligned} \sqrt{n}\phi(\hat{\theta}_n) &\stackrel{L_n}{\to} \max\{\mathbb{G}_0^{(1)} + \lambda^{(1)}, \mathbb{G}_0^{(2)} + \lambda^{(2)}\} \\ &\leq \max\{\mathbb{G}_0^{(1)}, \mathbb{G}_0^{(2)}\} \quad \text{(whenever } \lambda \leq 0) \end{aligned}$$

Example 1 Intuition

$$H_0 : \max\{E[X^{(1)}], E[X^{(2)}]\} \leq 0 \qquad H_1 : \max\{E[X^{(1)}], E[X^{(2)}]\} > 0$$

But local to P with $\theta^{(1)}(P) = \theta^{(2)}(P) = 0$, set $\theta(P_{n,\lambda}) = \theta(P) + \lambda/\sqrt{n}$ to get

$$\begin{aligned}\sqrt{n}\phi(\hat{\theta}_n) &\stackrel{L_n}{\to} \max\{\mathbb{G}_0^{(1)} + \lambda^{(1)}, \mathbb{G}_0^{(2)} + \lambda^{(2)}\} \\ &\leq \max\{\mathbb{G}_0^{(1)}, \mathbb{G}_0^{(2)}\} \quad \text{(whenever } \lambda \leq 0)\end{aligned}$$

Key Properties

- Local paths in null first order stochastically dominated by pointwise limit.
- $\theta(P)$ is regular at $P \Rightarrow$ no need to worry about it.

Subadditivity

$$\underbrace{\max\{\mathbb{G}_0^{(1)} + \lambda^{(1)}, \mathbb{G}_0^{(2)} + \lambda^{(2)}\}} \leq \underbrace{\max\{\mathbb{G}_0^{(1)}, \mathbb{G}_0^{(2)}\}} + \underbrace{\max\{\lambda^{(1)}, \lambda^{(2)}\}}$$

Subadditivity

$$\underbrace{\max\{\mathbb{G}_0^{(1)}+\lambda^{(1)},\mathbb{G}_0^{(2)}+\lambda^{(2)}\}}_{\phi'_{\theta_0}(\mathbb{G}_0+\lambda)} \leq \underbrace{\max\{\mathbb{G}_0^{(1)},\mathbb{G}_0^{(2)}\}}_{\phi'_{\theta_0}(\mathbb{G}_0)} + \underbrace{\max\{\lambda^{(1)},\lambda^{(2)}\}}_{\phi'_{\theta_0}(\lambda)}$$

Subadditivity

$$\underbrace{\max\{\mathbb{G}_0^{(1)}+\lambda^{(1)},\mathbb{G}_0^{(2)}+\lambda^{(2)}\}}_{\phi'_{\theta_0}(\mathbb{G}_0+\lambda)} \le \underbrace{\max\{\mathbb{G}_0^{(1)},\mathbb{G}_0^{(2)}\}}_{\phi'_{\theta_0}(\mathbb{G}_0)} + \underbrace{\max\{\lambda^{(1)},\lambda^{(2)}\}}_{\phi'_{\theta_0}(\lambda)}$$

$$\le \phi'_{\theta_0}(\mathbb{G}_0)$$

Subadditivity

$$\underbrace{\max\{\mathbb{G}_0^{(1)} + \lambda^{(1)}, \mathbb{G}_0^{(2)} + \lambda^{(2)}\}}_{\phi'_{\theta_0}(\mathbb{G}_0 + \lambda)} \leq \underbrace{\max\{\mathbb{G}_0^{(1)}, \mathbb{G}_0^{(2)}\}}_{\phi'_{\theta_0}(\mathbb{G}_0)} + \underbrace{\max\{\lambda^{(1)}, \lambda^{(2)}\}}_{\phi'_{\theta_0}(\lambda)}$$

$$\leq \phi'_{\theta_0}(\mathbb{G}_0)$$

Comments

- Key Condition: $\phi'_{\theta_0}(h_1 + h_2) \leq \phi'_{\theta_0}(h_1) + \phi'_{\theta_0}(h_2)$ (subadditivity).
- Andrews $\&$ Soares (2010), Andrews $\&$ Shi (2013), Linton et. al (2010).
- Equivalent to ϕ'_{θ_0} being convex due to homogeneity of degree one.

Size Control

Theorem If Assumptions (D), (B), (E), (L) hold, and $P_n \equiv \bigotimes_{i=1}^n P_{n,\lambda}$, then

$$\lim_{n\to\infty} P_n(\sqrt{n}\phi(\hat{\theta}_n) > \hat{c}_{1-\alpha}) = P(\phi'_{\theta_0}(\mathbb{G}_0 + \theta'(\lambda)) > c_{1-\alpha})$$

If in addition $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbf{R}$ is subadditive and $\phi(\theta(P_{n,\lambda})) \leq 0$ for all n, then

$$\limsup_{n\to\infty} P_n(\sqrt{n}\phi(\hat{\theta}_n) > \hat{c}_{1-\alpha}) \leq \alpha$$

Comments

- Key condition: subadditivity (of ϕ'_{θ_0}) and regularity (of $\hat{\theta}_n$)
- However, size control result is only local to $P \in \mathbf{P}$.
- But reassuring if subadditivity and regularity satisfied at all $P \in \mathbf{P}$.

Hypothesis

Let $\mathbb{H}$ be Hilbert space with norm $\|\cdot\|_{\mathbb{H}}$. For $\Lambda \subseteq \mathbb{H}$ closed and convex, test

$$H_0 : \theta_0 \in \Lambda \qquad\qquad H_1 : \theta_0 \notin \Lambda$$

Hypothesis

Let $\mathbb{H}$ be Hilbert space with norm $\|\cdot\|_{\mathbb{H}}$. For $\Lambda \subseteq \mathbb{H}$ closed and convex, test

$$H_0 : \theta_0 \in \Lambda \qquad\qquad H_1 : \theta_0 \notin \Lambda$$

Define the projection operator $\Pi_\Lambda : \mathbb{H} \to \Lambda$ which for each $\theta \in \mathbb{H}$ satisfies

$$\|\theta - \Pi_\Lambda \theta\|_{\mathbb{H}} = \inf_{h \in \Lambda} \|\theta - h\|_{\mathbb{H}}$$

$\Rightarrow$ Express original hypothesis in terms of the distance between θ_0 and Λ

$$H_0 : \|\theta_0 - \Pi_\Lambda \theta_0\|_{\mathbb{H}} = 0 \qquad\qquad H_1 : \|\theta_0 - \Pi_\Lambda \theta_0\|_{\mathbb{H}} > 0$$

Test Statistic

As a test statistic employ the (scaled) distance between $\hat{\theta}_n$ and the set Λ

$$r_n \|\hat{\theta}_n - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}}$$

Map To Our Framework

- Let $\phi : \mathbb{H} \to \mathbf{R}$ be given by $\phi(\theta) = \|\theta - \Pi_\Lambda \theta\|_{\mathbb{H}}$.
- Hypotheses are then $H_0 : \phi(\theta_0) = 0$ and $H_1 : \phi(\theta_0) > 0$.
- Test statistic is $r_n \phi(\hat{\theta}_n)$.

Key Steps

- Directional Differentiability of ϕ – Zaranotello (1971)
- Geometry enables easy construction of $\hat{\phi}'_n$

Takeaway: Very different problems can easily be handled in a unified way.

Examples

Suppose $X \in \mathbf{R}^d$ and consider the moment inequalities testing problem

$$H_0 : E[X] \leq 0 \qquad\qquad H_1 : E[X] \not\leq 0$$

Here $\mathbb{H} = \mathbf{R}^d$, Λ is the negative orthant ($\Lambda \equiv \{h \in \mathbf{R}^d : h \leq 0\}$), and

$$\phi(\theta) \equiv \|\Pi_\Lambda \theta - \theta\|_{\mathbb{H}} = \Big\{ \sum_{i=1}^{d} (E[X^{(i)}])_+^2 \Big\}^{\frac{1}{2}}$$

Comments

- Trivial extension to include weighting in projection.
- Applies to other θ and Λ– Wolak (1988), Kitamura $\&$ Stoye (2013).
- Also: First order stochastic dominance, conditional moment inequalites.

Examples

Let $(Y, D, X) \in \mathbf{R} \times \mathbf{R} \times \mathbf{R}^{d_z}$ and consider the quantile regression model

$$(\theta_0(\tau), \beta(\tau)) \equiv \arg \min_{\theta \in \mathbf{R}, \beta \in \mathbf{R}^{d_z}} E[\rho_\tau(Y - D\theta - Z'\beta)]$$

Standard result to get convergence of $\sqrt{n}\{\hat{\theta}_n - \theta_0\}$ for any $\epsilon > 0$ in space

$$\mathbb{H} \equiv \{\theta : [\epsilon, 1-\epsilon] \to \mathbf{R} : \langle \theta, \theta \rangle_{\mathbb{H}} < \infty\} \qquad \langle \theta_1, \theta_2 \rangle_{\mathbb{H}} \equiv \int_{\epsilon}^{1-\epsilon} \theta_1(\tau)\theta_2(\tau) d\tau$$

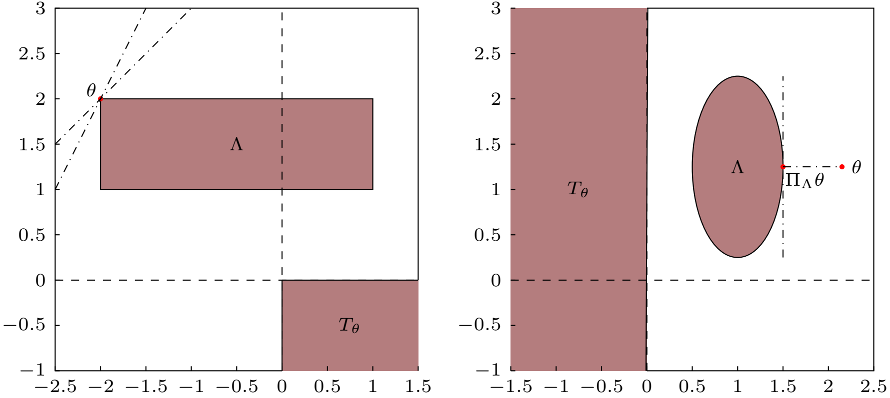
Comments

- Test for monotonicity of quantile treatment effects, correct specification.
- Other shape restrictions: concavity, convexity, homogeneity ...
- Similar: pricing kernel puzzle finds lack of predicted monotonicity.
- Also Related: Arellano et. al (2012), Escanciano & Zhu (2013).

Directional Differentiability

Definition For any $\theta \in \mathbb{H}$, the tangent cone of Λ at θ is given by:

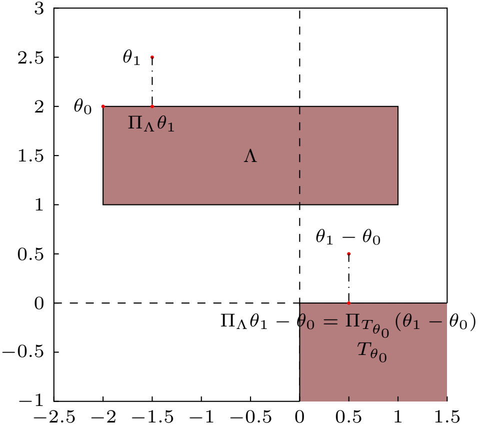
$$T_\theta \equiv \overline{\bigcup_{\alpha \geq 0} \alpha\{\Lambda - \Pi_\Lambda \theta\}}$$



Directional Differentiability

Zaranotello (1971) The directional derivative of Π_Λ at any $\theta \in \Lambda$ equals Π_{T_θ}

$$\Pi_\Lambda \theta_1 - \Pi_\Lambda \theta_0 \approx \Pi_{T_{\theta_0}}(\theta_1 - \theta_0)$$



Asymptotic Distribution

Proposition Let $\Lambda \subseteq \mathbb{H}$ be convex, and $r_n\{\hat{\theta}_n - \theta_0\} \xrightarrow{L} \mathbb{G}_0$. If $\theta_0 \in \Lambda$, then

$$\underbrace{r_n\|\hat{\theta}_n - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}}}_{r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}} \xrightarrow{L} \underbrace{\|\mathbb{G}_0 - \Pi_{T_{\theta_0}}\mathbb{G}_0\|_{\mathbb{H}}}_{\phi'_{\theta_0}(\mathbb{G}_0)}$$

Comments

- Quantiles of $\|\mathbb{G}_0\|_{\mathbb{H}}$ always provide valid (conservative?) critical values.
- If Λ is a cone, then quantiles of $\|\mathbb{G}_0 - \Pi_\Lambda \mathbb{G}_0\|_{\mathbb{H}}$ also valid.
- Possible to study projection $\Pi_\Lambda \theta_0$, and allow nonconvex Λ and $\theta_0 \notin \Lambda$.

Next: For inference, still need to construct suitable estimator $\hat{\phi}'_n$ for ϕ'_{θ_0}.

Bootstrap Alternative

$$\hat{\phi}_n'(h) \equiv \sup_{\theta \in \Lambda : \|\theta - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \leq \epsilon_n} \|h - \Pi_{T_\theta} h\|_{\mathbb{H}}$$

Intuition: Use the approximation by local “least favorable” tangent cone.

Bootstrap Alternative

$$\hat{\phi}_n'(h) \equiv \sup_{\theta\in\Lambda:\|\theta-\Pi_\Lambda\hat{\theta}_n\|_{\mathbb{H}}\leq\epsilon_n} \|h-\Pi_{T_\theta}h\|_{\mathbb{H}}$$

Intuition: Use the approximation by local “least favorable” tangent cone.

Proposition Let Λ be convex, $r_n\{\hat{\theta}_n-\theta_0\} \xrightarrow{L} \mathbb{G}_0$, $\phi_{\theta_0}'(h) \equiv \|h-\Pi_{T_{\theta_0}}h\|_{\mathbb{H}}$.

(A) If $\epsilon_n \downarrow 0$ an $\epsilon_n r_n \uparrow \infty$, then $\hat{\phi}_n'$ satisfies Assumption (E).

(B) $\phi_{\theta_0}' : \mathbb{H} \to \mathbf{R}$ satisfies $\phi_{\theta_0}'(h_1+h_2) \leq \phi_{\theta_0}'(h_1)+\phi_{\theta_0}'(h_2)$ (subadditive)

Comments

- Part (A) allows us to employ $\hat{\phi}_n'(r_n\{\hat{\theta}_n^*-\hat{\theta}_n\})$ if bootstrap works for $\hat{\theta}_n$.
- Part (B) implies local size control whenever $\hat{\theta}_n$ is regular at P.

Simulation Evidence

$$Y = \frac{\Delta}{\sqrt{n}} D \times U + Z'\beta + U$$

Where

- $D \in \{0, 1\}$ with $P(D = 1) = \frac{1}{2}$.
- $Z = (1, Z^{(1)}, Z^{(2)})'$ with $(Z^{(1)}, Z^{(2)}) \sim N(0, I_2)$.
- $U \sim U[0, 1]$ is unobserved, and $\beta = (0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})'$.
- D, Z, and U are all mutually independent.

Simulation Evidence

$$Y = \frac{\Delta}{\sqrt{n}} D \times U + Z'\beta + U$$

Where

- $D \in \{0, 1\}$ with $P(D = 1) = \frac{1}{2}$.
- $Z = (1, Z^{(1)}, Z^{(2)})'$ with $(Z^{(1)}, Z^{(2)}) \sim N(0, I_2)$.
- $U \sim U[0, 1]$ is unobserved, and $\beta = (0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})'$.
- D, Z, and U are all mutually independent.

It is then immediate that for $\theta_0(\tau) = \tau \frac{\Delta}{\sqrt{n}}$ and $\beta(\tau) \equiv (\tau, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})'$ we have

$$P(Y \leq D\theta_0(\tau) + Z'\beta(\tau) | D, Z) = \tau$$

Simulation Evidence

Goal Test whether $\theta_0(\tau)$ ($\approx$ QTE) is weakly increasing in τ.

Steps Using five thousand replications

- Compute quantile regression coefficient $\hat{\theta}_n$ on grid $\{0.2, 0.225, \ldots, 0.8\}$.
- Obtain $\Pi_\Lambda \hat{\theta}_n$ – projection onto monotone functions Λ.
- Compute two hundred bootstrap estimators of $\hat{\theta}_n^*$ on same grid.
- For each $\hat{\theta}_n^*$ obtain $\hat{\phi}_n'(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})$ (needs $\Pi_\Lambda \hat{\theta}_n$ and ϵ_n).

Evaluate

- Sensitivity to choice of $\epsilon_n = Cn^\kappa$ with $C \in \{0.01, 1\}$ and $\kappa \in \{\frac{1}{3}, \frac{1}{4}\}$.
- Accuracy of local approximation for different Δ ($\theta_0(\tau) = \Delta \frac{\tau}{\sqrt{n}}$).

Table: Empirical Size

Bandwidth		$n = 200$								
		$\alpha = 0.1$			$\alpha = 0.05$			$\alpha = 0.01$		
C	κ	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$
1	1/4	0.042	0.017	0.006	0.020	0.008	0.002	0.005	0.001	0.000
1	1/3	0.042	0.017	0.006	0.020	0.008	0.002	0.005	0.001	0.000
0.01	1/4	0.082	0.053	0.035	0.035	0.023	0.013	0.007	0.002	0.001
0.01	1/3	0.087	0.059	0.042	0.038	0.025	0.015	0.007	0.002	0.001
Theoretical		0.100	0.042	0.015	0.050	0.018	0.006	0.010	0.003	0.001

Bandwidth		$n = 500$								
		$\alpha = 0.1$			$\alpha = 0.05$			$\alpha = 0.01$		
C	κ	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$
1	1/4	0.051	0.020	0.007	0.026	0.011	0.002	0.005	0.001	0.000
1	1/3	0.051	0.020	0.007	0.026	0.011	0.002	0.005	0.001	0.000
0.01	1/4	0.096	0.058	0.038	0.047	0.025	0.015	0.009	0.005	0.001
0.01	1/3	0.103	0.065	0.045	0.049	0.030	0.017	0.009	0.005	0.001
Theoretical		0.100	0.042	0.015	0.050	0.018	0.006	0.010	0.003	0.001

Table: Local Power of 0.05 Level Test

Bandwidth		$n = 200$							
C	κ	$\Delta = -1$	$\Delta = -2$	$\Delta = -3$	$\Delta = -4$	$\Delta = -5$	$\Delta = -6$	$\Delta = -7$	$\Delta = -8$
1	1/4	0.061	0.155	0.321	0.555	0.782	0.934	0.989	1.000
1	1/3	0.061	0.155	0.321	0.555	0.782	0.934	0.989	1.000
0.01	1/4	0.078	0.172	0.330	0.558	0.783	0.934	0.989	1.000
0.01	1/3	0.081	0.174	0.331	0.559	0.783	0.934	0.989	1.000
Theoretical		0.120	0.245	0.423	0.623	0.796	0.911	0.970	0.992
Bandwidth		$n = 500$							
C	κ	$\Delta = -1$	$\Delta = -2$	$\Delta = -3$	$\Delta = -4$	$\Delta = -5$	$\Delta = -6$	$\Delta = -7$	$\Delta = -8$
1	1/4	0.071	0.181	0.355	0.576	0.789	0.925	0.981	0.997
1	1/3	0.071	0.181	0.355	0.576	0.789	0.925	0.981	0.997
0.01	1/4	0.094	0.201	0.370	0.583	0.791	0.925	0.981	0.997
0.01	1/3	0.098	0.204	0.371	0.585	0.791	0.925	0.981	0.997
Theoretical		0.120	0.245	0.423	0.623	0.796	0.911	0.970	0.992

Conclusion

Delta method

- Preserved under directional derivative (Shapiro 1991, Dumbgen 1993).
- Small extension to show it holds in probability.

Bootstrap

- Differentiability necessary and sufficient when $\mathbb{G}_0$ is Gaussian.
- Argued popular approaches implicitly estimate ϕ'_{θ_0}.

Inference

- Local size control guaranteed by subadditivity and regularity.
- Application to testing if θ_0 belongs to convex set.

$\Rightarrow$ Problems can be analyzed by studying $\mathbb{G}_0$ and ϕ'_{θ_0} (as in Delta method).