Inference on Directionally Differentiable Functions

Zheng Fang          Andres Santos
Kansas State        UC San Diego

November 9, 2016
General Question

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

$$\phi(\theta_0)$$

Given estimator $\hat{\theta}_n$ for $\theta_0$, 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 $\phi$?
Andrews & Soares (2010): Let $X = (X^{(1)}, X^{(2)})' \in \mathbb{R}^2$, and consider

$$\max\{E[X^{(1)}], E[X^{(2)}]\}$$
Andrews & Soares (2010): Let $X = (X^{(1)}, X^{(2)})' \in \mathbb{R}^2$, and consider
\[
\max\{E[X^{(1)}], E[X^{(2)}]\}
\]
Here $\theta_0 = E[X]$ and for any $\theta = (\theta^{(1)}, \theta^{(2)})' \in \mathbb{R}^2$ set $\phi : \mathbb{R}^2 \to \mathbb{R}$ to equal
\[
\phi(\theta) = \max\{\theta^{(1)}, \theta^{(2)}\}
\]
Example 1

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

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

Here $\theta_0 = E[X]$ and for any $\theta = (\theta^{(1)}, \theta^{(2)})' \in \mathbb{R}^2$ set $\phi: \mathbb{R}^2 \to \mathbb{R}$ 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 $Y \in \mathbb{R}$ and $Z \in \mathbb{R}^{d_z}$ consider testing the null

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

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

$$\sup_{f \in \mathcal{F}} E[Y f(Z)] \leq 0$$
Example 2

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

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

For appropriate $\mathcal{F} \subseteq \ell\infty(\mathbb{R}^{dz})$ (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 $\phi : \ell\infty(\mathcal{F}) \rightarrow \mathbb{R}$ is given by

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

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

$$\phi(\hat{\theta}_n) = \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n Y_if(Z_i)$$
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 $\phi$ is directionally differentiable whenever it is not fully differentiable
General Outline

Question: How much structure does directional differentiability provide?
Question: How much structure does directional differentiability provide?

General Results

• 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 Results**
- **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 $\theta_0$ belongs to convex set.
Related Literature

Partial Identification

Bootstrap Validity

Directional Differentiability
New Applications

- Derivative Estimation: Hong and Li (2015).
The Delta Method

The Bootstrap

Bootstrap Alternative

Inference Implications

Convex Set Membership
Directional Differentiability

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

Then $\phi$ is directionally differentiable

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

for every sequence $t_n \downarrow 0$.
Directional Differentiability

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

Then $\phi$ is directionally differentiable \textit{in the Hadamard sense} if

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

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

Comments

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

Let \( \phi : D \to E \) with \( D \) and \( E \) Banach Spaces with norms \( \| \cdot \|_D \) and \( \| \cdot \|_E \).

Then \( \phi \) is directionally differentiable in the Hadamard sense tangential to \( D_0 \)

\[
\lim_{n \to \infty} \left\| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'(\theta)(h) \right\|_E = 0
\]

for every sequence \( t_n \downarrow 0 \) and \( h_n \to h \) with \( h \in D_0 \subseteq D \).
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 $\phi$ is directionally differentiable in the Hadamard sense tangential to $\mathbb{D}_0$

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

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 $\phi'_\theta$ does not need to be linear as required in full differentiability.
- In fact, $\phi$ is Hadamard differentiable at $\theta$ if and only if $\phi'_\theta$ is linear.
\[ \phi(\theta) = |\theta| \]

**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 \( h \in \mathbb{R} \)
Example 1 (cont)

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

Comments

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

Santos. November 9, 2016. UCSD
Example 2 (cont)

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

$$\phi'_\theta(h) = \sup_{f \in \Psi_F(\theta)} h(f)$$

for every continuous $h : F \rightarrow \mathbb{R}$ and where $\Psi_F(\theta) \equiv \arg \max_{f \in F} \theta(f)$.

Comments

- $\phi$ is fully differentiable except when $\Psi_F(\theta)$ is not a singleton.
- $\phi'_\theta$ is linear except when $\Psi_F(\theta)$ is not a singleton.
- Here $\mathbb{D} = \ell^\infty(F)$, $\mathbb{E} = \mathbb{R}$, and $\mathbb{D}_0 = C(F)$.
- $\Rightarrow$ 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 \( \theta_0 \) 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

\[ r_n \{ \phi(\hat{\theta}_n) - \phi(\theta_0) \} \xrightarrow{L} \phi'_{\theta_0}(G_0) \]

Addendum If in addition \( \phi'_{\theta_0} \) 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 \( \phi \) only assumed at \( \theta_0 \).
- 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

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

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

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

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

\[
r_n \{ \phi(\hat{\theta}_n) - \phi(\theta_0) \} = \frac{1}{t_n} \{ \phi(\theta_0 + t_n \times r_n \{ \hat{\theta}_n - \theta_0 \}) - \phi(\theta_0) \} \\
\approx \phi'_{\theta_0} (r_n \{ \hat{\theta}_n - \theta_0 \})
\]

**Step 2:** Since \( \phi'_{\theta_0} : \mathbb{D} \rightarrow \mathbb{E} \) is continuous, use continuous mapping theorem

\[
\phi'_{\theta_0} (r_n \{ \hat{\theta}_n - \theta_0 \}) \xrightarrow{L} \phi'_{\theta_0} (G_0)
\]

**Key Observation**

Linearity of \( \phi'_{\theta_0} \) is irrelevant in the original proof of the Delta method.
1 The Delta Method

2 The Bootstrap

3 Bootstrap Alternative

4 Inference Implications

5 Convex Set Membership
Bootstrap Setup

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

What we know

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

**Questions**
- Does the bootstrap always fail when \( \phi \) 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 $A$ with norm $\| \cdot \|_A$, denote bounded Lipschitz functions

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

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

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

Comments

- Largest discrepancy in expectations assigned to functions in $BL_1(A)$.
- Metrizes weak convergence. Key in showing validity of critical values.
- Bootstrap consistency $\Leftrightarrow$ distance measured by $d_{BL}$ is $o_p(1)$.
Informally: Assume the “bootstrapped” version $\hat{\theta}_n^*$ “works” for original $\hat{\theta}_n$.

Assumptions (B)

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

(ii) $\sup_{f \in BL_1(D)} |E[f(r_n \{\hat{\theta}_n^* - \hat{\theta}_n\})|\{X_i\}_{i=1}^n] - E[f(G_0)]| = o_p(1)$.

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 $G_0$.
- Also need mild (asymptotic) measurability requirements.
Necessary and Sufficient

**Theorem** Suppose $G_0$ is a Gaussian measure and Assumptions (D), (B), and regularity conditions hold. Then, $\phi : \mathbb{D}_\phi \rightarrow \mathbb{E}$ is (fully) Hadamard differentiable at $\theta_0 \in \mathbb{D}_\phi$ tangential to the support of $G_0$ if and only if

$$\sup_{f \in BL_1(\mathbb{E})} |E[f(r_n \{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\})\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(G_0))]| = o_P(1)$$
Theorem Suppose $G_0$ 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 $G_0$ if and only if

$$\sup_{f \in BL_1(\mathbb{E})} |E[f(r_n \{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\}) | \{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(G_0))] | = o_P(1)$$

Key Implication: If the bootstrap works for $\phi(\hat{\theta}_n)$ and $G_0$ is Gaussian

$\Rightarrow \phi'_{\theta_0}$ must be linear $\Rightarrow \phi'_{\theta_0}(G_0)$ must be Gaussian

Corollary Suppose $G_0$ 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 \( \{X_i\}_{i=1}^n \):

\[
\begin{align*}
r_n \{ \phi(\hat{\theta}^*_n) - \phi(\hat{\theta}_n) \} \\
= r_n \{ \phi(\hat{\theta}^*_n) - \phi(\theta_0) \} - r_n \{ \phi(\hat{\theta}_n) - \phi(\theta_0) \}
\end{align*}
\]
Step 1: Use the Delta method to conclude that unconditionally on $\{X_i\}_{i=1}^n$:

$$
r_n \{ \phi(\hat{\theta}^*_n) - \phi(\hat{\theta}_n) \}
= r_n \{ \phi(\hat{\theta}^*_n) - \phi(\theta_0) \} - r_n \{ \phi(\hat{\theta}_n) - \phi(\theta_0) \}
= \phi'_{\theta_0} (r_n \{ \hat{\theta}^*_n - \theta_0 \}) - \phi'_{\theta_0} (r_n \{ \hat{\theta}_n - \theta_0 \}) + o_p(1)
$$
Step 1: Use the Delta method to conclude that unconditionally on $\{X_i\}_{i=1}^n$:

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

Santos. November 9, 2016.
Proof Intuition

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

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

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

\[
= \phi'_{\theta_0} (r_n \{ \hat{\theta}_n^* - \theta_0 \}) - \phi'_{\theta_0} (r_n \{ \hat{\theta}_n - \theta_0 \}) + o_p(1)
\]

\[
= \phi'_{\theta_0} (r_n \{ \hat{\theta}_n^* - \hat{\theta}_n \} + r_n \{ \hat{\theta}_n - \theta_0 \}) - \phi'_{\theta_0} (r_n \{ \hat{\theta}_n - \theta_0 \}) + o_p(1)
\]

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

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

Bootstrap works iff \( \phi'_{\theta_0} (G_0 + h) - \phi'_{\theta_0} (h) \) is equal in distribution to \( \phi'_{\theta_0} (G_0) \).

Santos. November 9, 2016. UCSD
Proof Intuition

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

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

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

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

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

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

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

$$\overset{L}{\rightarrow} G_0 \quad \rightarrow h \quad \rightarrow h$$
Proof Intuition

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

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

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

\[
\phi'_{\theta_0} (r_n \{ \hat{\theta}_n^* - \hat{\theta}_n \} + r_n \{ \hat{\theta}_n - \theta_0 \}) - \phi'_{\theta_0} (r_n \{ \hat{\theta}_n - \theta_0 \}) \xrightarrow{L} \mathbb{G}_0 \xrightarrow{\rightarrow} h \xrightarrow{\rightarrow} h
\]

\[\Rightarrow\] 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 $G_0$)

$$\phi'_{\theta_0}(G_0 + h) - \phi'_{\theta_0}(h) \overset{d}{=} \phi'_{\theta_0}(G_0)$$

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 $G_0$)

$$\phi'_{\theta_0} (G_0 + h) - \phi'_{\theta_0} (h) \overset{d}{=} \phi'_{\theta_0} (G_0)$$

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

Step 3: (Scalar Case) Suppose $G_0 \sim N(0, 1)$, then for any $r > 0$ and $t \in \mathbb{R}$

$$E[\exp\{it(\phi'_{\theta_0} (G_0 + rh) - \phi'_{\theta_0} (rh))\}] = C(t)$$
Proof Intuition

So Far: Bootstrap consistency is equivalent to (for any $h$ in support of $G_0$)

$$\phi'_{\theta_0}(G_0 + h) - \phi'_{\theta_0}(h) \overset{d}{=} \phi'_{\theta_0}(G_0)$$

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

Step 3: (Scalar Case) Suppose $G_0 \sim N(0, 1)$, then for any $r > 0$ and $t \in \mathbb{R}$

$$E[\exp\{it(\phi'_{\theta_0}(G_0 + rh) - \phi'_{\theta_0}(rh))\}] = C(t)$$

$$\Rightarrow E[\exp\{it\phi'_{\theta_0}(G_0 + rh)\}] = \exp\{itr\phi'_{\theta_0}(h)\}C(t)$$
Proof Intuition

So Far: Bootstrap consistency is equivalent to (for any \( h \) in support of \( G_0 \))

\[
\phi'_{\theta_0}(G_0 + h) - \phi'_{\theta_0}(h) \overset{d}{=} \phi'_{\theta_0}(G_0)
\]

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

Step 3: (Scalar Case) Suppose \( G_0 \sim N(0, 1) \), then for any \( r > 0 \) and \( t \in \mathbb{R} \)

\[
E[\exp\{it(\phi'_{\theta_0}(G_0 + rh) - \phi'_{\theta_0}(rh))\}] = C(t)
\]

\[
\Rightarrow E[\exp\{it\phi'_{\theta_0}(G_0 + rh)\}] = \exp\{itr\phi'_{\theta_0}(h)\}C(t)
\]

\[
\Rightarrow \frac{1}{\sqrt{2\pi}} \int \exp\{it\phi'_{\theta_0}(u)\} \exp\{-\frac{1}{2}(u - rh)^2\}du = \exp\{itr\phi'_{\theta_0}(h)\}C(t)
\]
Proof Intuition

So Far: Bootstrap consistency is equivalent to (for any $h$ in support of $G_0$)

$$
\phi'_{\theta_0}(G_0 + h) - \phi'_{\theta_0}(h) \overset{d}{=} \phi'_{\theta_0}(G_0)
$$

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

Step 3: (Scalar Case) Suppose $G_0 \sim N(0, 1)$, then for any $r > 0$ and $t \in \mathbb{R}$

$$
E[\exp\{it(\phi'_{\theta_0}(G_0 + rh) - \phi'_{\theta_0}(rh))\}] = C(t)
\Rightarrow E[\exp\{it\phi'_{\theta_0}(G_0 + rh)\}] = \exp\{itr\phi'_{\theta_0}(h)\}C(t)
\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)
\Rightarrow \text{Differentiate both sides w.r.t } r \text{ to conclude } \phi'_{\theta_0}(h) \text{ is linear in } h.
Proof Intuition

**So Far:** Bootstrap consistency is equivalent to (for any $h$ in support of $G_0$)

$$\phi'_{\theta_0}(G_0 + h) - \phi'_{\theta_0}(h) \overset{d}{=} \phi'_{\theta_0}(G_0)$$

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

**Step 3:** (Scalar Case) Suppose $G_0 \sim N(0, 1)$, then for any $r > 0$ and $t \in \mathbb{R}$

$$E[\exp\{it(\phi'_{\theta_0}(G_0 + rh) - \phi'_{\theta_0}(rh))\}] = C(t)$$

$$\Rightarrow E[\exp\{it\phi'_{\theta_0}(G_0 + rh)\}] = \exp\{itr\phi'_{\theta_0}(h)\}C(t)$$

$$\Rightarrow \frac{1}{\sqrt{2\pi}} \int \exp\{it\phi'_{\theta_0}(u)\} \exp\{-\frac{1}{2}(u - rh)^2\}du = \exp\{itr\phi'_{\theta_0}(h)\}C(t)$$

$$\Rightarrow$$ 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 $\mathbb{F}^*$. 
1. The Delta Method

2. The Bootstrap

3. Bootstrap Alternative

4. Inference Implications

5. Convex Set Membership
Bootstrap Alternative

\[ r_n \{ \phi(\hat{\theta}_n) - \phi(\theta_0) \} \xrightarrow{L} \phi'_{\theta_0}(G_0) \]

**Intuition**

- We need to estimate \( \phi'_{\theta_0} \) and the law of \( G_0 \).
- By assumption \( r_n \{ \hat{\theta}^*_n - \hat{\theta}_n \} \) provides an estimate of the law of \( G_0 \).
- Bootstrap fails for \( \phi(\hat{\theta}_n) \) because it does not estimate \( \phi'_{\theta_0} \) appropriately.

**Fix:** For an estimator \( \hat{\phi}'_n \) of \( \phi'_{\theta_0} \), use the law conditional on \( \{ X_i \}_{i=1}^n \) of

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

Santos. November 9, 2016. UCSD
Assumption (E) For every compact $K \subseteq \mathbb{D}_0$ and $\epsilon > 0$, $\hat{\phi}_n : \mathbb{D} \to \mathbb{E}$ satisfies

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

Discussion

- $\delta$-enlargement needed because $r_n \{ \hat{\theta}_n^* - \hat{\theta}_n \}$ may not belong to $\mathbb{D}_0$.
- $\delta$ 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

$$\sup_{f \in BL_1(E)} |E[f(\hat{\phi}'_n(r_n(\hat{\theta}^*_n - \hat{\theta}_n))) | \{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(G_0))]| = o_p(1)$$

**Comments**

- The law of $\hat{\phi}'_n(r_n(\hat{\theta}^*_n - \hat{\theta}_n))$ conditional $\{X_i\}_{i=1}^n$ consistent for $\phi'_{\theta_0}(G_0)$.
- Implies consistency of critical values under standard conditions.
- The fact that $\phi'_{\theta_0}$ is a directional derivative is never exploited ...

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

$$\tau(G_0)$$

where $G_0$ is tight and $\tau : D \rightarrow E$ 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 $\Psi_F(\theta) \equiv \arg \max_{f \in F} \theta(f)$ we had that:

$$\phi'_\theta(h) = \sup_{f \in \Psi_F(\theta)} h(f)$$

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

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

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).
1 The Delta Method
2 The Bootstrap
3 Bootstrap Alternative
4 Inference Implications
5 Convex Set Membership
Testing Implications

\[ H_0 : \phi(\theta_0) \leq 0 \quad \quad H_1 : \phi(\theta_0) > 0 \]

Proposed Test

- Employ \( \sqrt{n}\phi(\hat{\theta}_n) \) as a test statistic.
- Unfeasible: \( c_{1-\alpha} \) the \( 1 - \alpha \) quantile of \( \phi'_{\theta_0}(G_0) \) (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

\[ H_0 : \phi(\theta_0) \leq 0 \quad \text{\textit{\textup{\textbf{H}}}1 : \phi(\theta_0) > 0} \]

Proposed Test

- Employ \( \sqrt{n} \phi(\hat{\theta}_n) \) as a test statistic.
- Unfeasible: \( c_{1-\alpha} \) the \( 1 - \alpha \) quantile of \( \phi'_{\theta_0}(G_0) \) (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 \( X_i \sim P \in \mathcal{P} \).

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

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

(iv) \( P_{n,\lambda} \in \mathcal{P} \) 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 \( \phi \).
Example 1 Intuition

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

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

\[ \sqrt{n} \phi(\hat{\theta}_n) \overset{L_n}{\to} \max\{G_0^{(1)} + \lambda^{(1)}, G_0^{(2)} + \lambda^{(2)}\} \]
Example 1 Intuition

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

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

\[
\sqrt{n} \phi(\hat{\theta}_n) \xrightarrow{L_n} \max \{G_0^{(1)} + \lambda^{(1)}, G_0^{(2)} + \lambda^{(2)}\}
\]

\[ \leq \max \{G_0^{(1)}, G_0^{(2)}\} \quad (\text{whenever } \lambda \leq 0) \]
Example 1 Intuition

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

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

\[ \sqrt{n}\phi(\hat{\theta}_n) \xrightarrow{L_n} \max\{G^{(1)}_0 + \lambda^{(1)}, G^{(2)}_0 + \lambda^{(2)}\} \leq \max\{G^{(1)}_0, G^{(2)}_0\} \quad \text{(whenever } \lambda \leq 0) \]

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

$$\max\{G_0^{(1)} + \lambda^{(1)}, G_0^{(2)} + \lambda^{(2)}\} \leq \max\{G_0^{(1)}, G_0^{(2)}\} + \max\{\lambda^{(1)}, \lambda^{(2)}\}$$

Comments
- Key Condition: $$\phi'_{\theta_0}(h_1 + h_2) \leq \phi'_{\theta_0}(h_1) + \phi'_{\theta_0}(h_2)$$ (subadditivity).
- Equivalent to $$\phi'_{\theta_0}$$ being convex due to homogeneity of degree one.
Subadditivity

\[ \max \{ G_0^{(1)} + \lambda^{(1)}, G_0^{(2)} + \lambda^{(2)} \} \leq \max \{ G_0^{(1)}, G_0^{(2)} \} + \max \{ \lambda^{(1)}, \lambda^{(2)} \} \]

\[ \phi'_{\theta_0} (G_0 + \lambda) \leq \phi'_{\theta_0} (G_0) + \phi'_{\theta_0} (\lambda) \]
Subadditivity

\[
\max\{G_0^{(1)} + \lambda^{(1)}, G_0^{(2)} + \lambda^{(2)}\} \leq \max\{G_0^{(1)}, G_0^{(2)}\} + \max\{\lambda^{(1)}, \lambda^{(2)}\}
\]

\[
\phi'_{\theta_0}(G_0 + \lambda) \leq \phi'_{\theta_0}(G_0) + \phi'_{\theta_0}(\lambda)
\]

Comments

• Key Condition: \(\phi'_{\theta_0}(h_1 + h_2) \leq \phi'_{\theta_0}(h_1) + \phi'_{\theta_0}(h_2)\) (subadditivity).


• Equivalent to \(\phi'_{\theta_0}\) being convex due to homogeneity of degree one.
Subadditivity

\[
\max\{G_0^{(1)} + \lambda^{(1)}, G_0^{(2)} + \lambda^{(2)}\} \leq \max\{G_0^{(1)}, G_0^{(2)}\} + \max\{\lambda^{(1)}, \lambda^{(2)}\}
\]

\[
\phi'_{\theta_0}(G_0 + \lambda) \leq \phi'_{\theta_0}(G_0) + \phi'_{\theta_0}(\lambda)
\]

\[
\leq \phi'_{\theta_0}(G_0)
\]

Comments

• Key Condition: \(\phi'_{\theta_0}(h_1 + h_2) \leq \phi'_{\theta_0}(h_1) + \phi'_{\theta_0}(h_2)\) (subadditivity).


• Equivalent to \(\phi'_{\theta_0}\) 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 \( \phi'_{\theta_0} : \mathbb{D}_0 \to \mathbb{R} \) 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 \( \phi'_{\theta_0} \)) and regularity (of \( \hat{\theta}_n \))
- However, size control result is only local to \( P \in \mathbb{P} \).
- But reassuring if subadditivity and regularity satisfied at all \( P \in \mathbb{P} \).
The Delta Method

The Bootstrap

Bootstrap Alternative

Inference Implications

Convex Set Membership
Hypothesis

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

\[
H_0 : \theta_0 \in \Lambda \quad \quad H_1 : \theta_0 \notin \Lambda
\]
Let $H$ be Hilbert space with norm $\| \cdot \|_H$. For $\Lambda \subseteq H$ closed and convex, test

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

Define the projection operator $\Pi_\Lambda : H \rightarrow \Lambda$ which for each $\theta \in H$ satisfies

$$\| \theta - \Pi_\Lambda \theta \|_H = \inf_{h \in \Lambda} \| \theta - h \|_H$$

$\Rightarrow$ Express original hypothesis in terms of the distance between $\theta_0$ and $\Lambda$

$$H_0 : \| \theta_0 - \Pi_\Lambda \theta_0 \|_H = 0 \quad H_1 : \| \theta_0 - \Pi_\Lambda \theta_0 \|_H > 0$$
As a test statistic employ the (scaled) distance between $\hat{\theta}_n$ and the set $\Lambda$

$$r_n \| \hat{\theta}_n - \Pi_\Lambda \hat{\theta}_n \|_H$$

**Map To Our Framework**
- Let $\phi : \mathbb{H} \rightarrow \mathbb{R}$ be given by $\phi(\theta) = \| \theta - \Pi_\Lambda \theta \|_H$.
- 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 $\phi$ – Zaranotello (1971)
- Geometry enables easy construction of $\hat{\phi}'_n$

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

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

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

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

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

Comments

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

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

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

Standard result to get convergence of \(\sqrt{n}\{\hat{\theta}_n - \theta_0\}\) for any \(\epsilon > 0\) in space \(H\equiv\{\theta : [\epsilon, 1-\epsilon] \rightarrow \mathbb{R} : \langle \theta, \theta \rangle_H < \infty\}\)

\[
\langle \theta_1, \theta_2 \rangle_H \equiv \int_{\epsilon}^{1-\epsilon} \theta_1(\tau)\theta_2(\tau)d\tau
\]

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.
**Definition** For any \( \theta \in \mathbb{H} \), the tangent cone of \( \Lambda \) at \( \theta \) is given by:

\[
T_{\theta} \equiv \bigcup_{\alpha \geq 0} \alpha \{ \Lambda - \Pi_{\Lambda} \theta \}
\]
**Zaranotello (1971)** The directional derivative of $\Pi_\Lambda$ at any $\theta \in \Lambda$ equals $\Pi_{T\theta}$

$$\Pi_\Lambda \theta_1 - \Pi_\Lambda \theta_0 \approx \Pi_{T\theta_0} (\theta_1 - \theta_0)$$
Proposition Let $\Lambda \subseteq \mathbb{H}$ be convex, and $r_n\{\hat{\theta}_n - \theta_0\} \xrightarrow{L} G_0$. If $\theta_0 \in \Lambda$, then

$$r_n\|\hat{\theta}_n - \Pi_\Lambda \hat{\theta}_n\|_\mathbb{H} \xrightarrow{L} \|G_0 - \Pi_{T\theta_0} G_0\|_\mathbb{H}$$

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \xrightarrow{L} \phi'_{\theta_0}(G_0)$$

Comments

- Quantiles of $\|G_0\|_\mathbb{H}$ always provide valid (conservative?) critical values.
- If $\Lambda$ is a cone, then quantiles of $\|G_0 - \Pi_\Lambda G_0\|_\mathbb{H}$ also valid.
- Possible to study projection $\Pi_\Lambda \theta_0$, and allow nonconvex $\Lambda$ and $\theta_0 \notin \Lambda$.

Next: For inference, still need to construct suitable estimator $\hat{\phi}'_n$ for $\phi'_{\theta_0}$. 
\[ \hat{\phi}'_n(h) \equiv \sup_{\theta \in \Lambda: \|\theta - \Pi_{\Lambda} \hat{\theta}_n\|_H \leq \epsilon_n} \|h - \Pi_T \theta h\|_H \]

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

\[ \hat{\phi}_n'(h) \equiv \sup_{\theta \in \Lambda: \|\theta - \Pi_{\Lambda} \hat{\theta}_n\|_H \leq \epsilon_n} \| h - \Pi_{T_\theta} h \|_H \]

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

**Proposition** Let \( \Lambda \) be convex, \( r_n \{ \hat{\theta}_n - \theta_0 \} \overset{L}{\to} G_0 \), \( \phi'_{\theta_0}(h) \equiv \| h - \Pi_{T_{\theta_0}} h \|_H \).

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

(B) \( \phi'_{\theta_0} : \mathbb{H} \to \mathbb{R} \) 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

\[ Y = \Delta \sqrt{\frac{D}{n}} U + Z' \beta + U \]

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 \( \theta_0(\tau) = \tau \Delta \sqrt{\frac{1}{n}} \) 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 $\theta_0(\tau) \approx \text{QTE}$ is weakly increasing in $\tau$.

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 $\Lambda$.
- 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 $\epsilon_n$).

Evaluate

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

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>$\kappa$</th>
<th>$\Delta = 0$</th>
<th>$\Delta = 1$</th>
<th>$\Delta = 2$</th>
<th>$\Delta = 0$</th>
<th>$\Delta = 1$</th>
<th>$\Delta = 2$</th>
<th>$\Delta = 0$</th>
<th>$\Delta = 1$</th>
<th>$\Delta = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td></td>
<td>$\alpha = 0.1$</td>
<td></td>
<td></td>
<td>$\alpha = 0.05$</td>
<td></td>
<td></td>
<td>$\alpha = 0.01$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1/4</td>
<td>0.042</td>
<td>0.017</td>
<td>0.006</td>
<td>0.020</td>
<td>0.008</td>
<td>0.002</td>
<td>0.005</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>1/3</td>
<td>0.042</td>
<td>0.017</td>
<td>0.006</td>
<td>0.020</td>
<td>0.008</td>
<td>0.002</td>
<td>0.005</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>0.01</td>
<td>1/4</td>
<td>0.082</td>
<td>0.053</td>
<td>0.035</td>
<td>0.035</td>
<td>0.023</td>
<td>0.013</td>
<td>0.007</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>0.01</td>
<td>1/3</td>
<td>0.087</td>
<td>0.059</td>
<td>0.042</td>
<td>0.038</td>
<td>0.025</td>
<td>0.015</td>
<td>0.007</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>Theoretical</td>
<td></td>
<td>0.100</td>
<td>0.042</td>
<td>0.015</td>
<td>0.050</td>
<td>0.018</td>
<td>0.006</td>
<td>0.010</td>
<td>0.003</td>
<td>0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>$\kappa$</th>
<th>$\Delta = 0$</th>
<th>$\Delta = 1$</th>
<th>$\Delta = 2$</th>
<th>$\Delta = 0$</th>
<th>$\Delta = 1$</th>
<th>$\Delta = 2$</th>
<th>$\Delta = 0$</th>
<th>$\Delta = 1$</th>
<th>$\Delta = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td></td>
<td>$\alpha = 0.1$</td>
<td></td>
<td></td>
<td>$\alpha = 0.05$</td>
<td></td>
<td></td>
<td>$\alpha = 0.01$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1/4</td>
<td>0.051</td>
<td>0.020</td>
<td>0.007</td>
<td>0.026</td>
<td>0.011</td>
<td>0.002</td>
<td>0.005</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>1/3</td>
<td>0.051</td>
<td>0.020</td>
<td>0.007</td>
<td>0.026</td>
<td>0.011</td>
<td>0.002</td>
<td>0.005</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>0.01</td>
<td>1/4</td>
<td>0.096</td>
<td>0.058</td>
<td>0.038</td>
<td>0.047</td>
<td>0.025</td>
<td>0.015</td>
<td>0.009</td>
<td>0.005</td>
<td>0.001</td>
</tr>
<tr>
<td>0.01</td>
<td>1/3</td>
<td>0.103</td>
<td>0.065</td>
<td>0.045</td>
<td>0.049</td>
<td>0.030</td>
<td>0.017</td>
<td>0.009</td>
<td>0.005</td>
<td>0.001</td>
</tr>
<tr>
<td>Theoretical</td>
<td></td>
<td>0.100</td>
<td>0.042</td>
<td>0.015</td>
<td>0.050</td>
<td>0.018</td>
<td>0.006</td>
<td>0.010</td>
<td>0.003</td>
<td>0.001</td>
</tr>
</tbody>
</table>
### Table: Local Power of 0.05 Level Test

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>$C$</th>
<th>$\kappa$</th>
<th>$\Delta = -1$</th>
<th>$\Delta = -2$</th>
<th>$\Delta = -3$</th>
<th>$\Delta = -4$</th>
<th>$\Delta = -5$</th>
<th>$\Delta = -6$</th>
<th>$\Delta = -7$</th>
<th>$\Delta = -8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>$C$</td>
<td>$\kappa$</td>
<td>$\Delta = -1$</td>
<td>$\Delta = -2$</td>
<td>$\Delta = -3$</td>
<td>$\Delta = -4$</td>
<td>$\Delta = -5$</td>
<td>$\Delta = -6$</td>
<td>$\Delta = -7$</td>
<td>$\Delta = -8$</td>
</tr>
<tr>
<td>$n = 200$</td>
<td></td>
<td></td>
<td>0.061</td>
<td>0.155</td>
<td>0.321</td>
<td>0.555</td>
<td>0.782</td>
<td>0.934</td>
<td>0.989</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>1/4</td>
<td></td>
<td>0.061</td>
<td>0.155</td>
<td>0.321</td>
<td>0.555</td>
<td>0.782</td>
<td>0.934</td>
<td>0.989</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>1/3</td>
<td></td>
<td>0.078</td>
<td>0.172</td>
<td>0.330</td>
<td>0.558</td>
<td>0.783</td>
<td>0.934</td>
<td>0.989</td>
<td>1.000</td>
</tr>
<tr>
<td>0.01</td>
<td>1/4</td>
<td></td>
<td>0.071</td>
<td>0.181</td>
<td>0.355</td>
<td>0.576</td>
<td>0.789</td>
<td>0.925</td>
<td>0.981</td>
<td>0.997</td>
</tr>
<tr>
<td>0.01</td>
<td>1/3</td>
<td></td>
<td>0.071</td>
<td>0.181</td>
<td>0.355</td>
<td>0.576</td>
<td>0.789</td>
<td>0.925</td>
<td>0.981</td>
<td>0.997</td>
</tr>
<tr>
<td>Theoretical</td>
<td></td>
<td></td>
<td>0.120</td>
<td>0.245</td>
<td>0.423</td>
<td>0.623</td>
<td>0.796</td>
<td>0.911</td>
<td>0.970</td>
<td>0.992</td>
</tr>
<tr>
<td>$n = 500$</td>
<td></td>
<td></td>
<td>0.061</td>
<td>0.155</td>
<td>0.321</td>
<td>0.555</td>
<td>0.782</td>
<td>0.934</td>
<td>0.989</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>1/4</td>
<td></td>
<td>0.071</td>
<td>0.181</td>
<td>0.355</td>
<td>0.576</td>
<td>0.789</td>
<td>0.925</td>
<td>0.981</td>
<td>0.997</td>
</tr>
<tr>
<td>1</td>
<td>1/3</td>
<td></td>
<td>0.071</td>
<td>0.181</td>
<td>0.355</td>
<td>0.576</td>
<td>0.789</td>
<td>0.925</td>
<td>0.981</td>
<td>0.997</td>
</tr>
<tr>
<td>0.01</td>
<td>1/4</td>
<td></td>
<td>0.094</td>
<td>0.201</td>
<td>0.370</td>
<td>0.583</td>
<td>0.791</td>
<td>0.925</td>
<td>0.981</td>
<td>0.997</td>
</tr>
<tr>
<td>0.01</td>
<td>1/3</td>
<td></td>
<td>0.098</td>
<td>0.204</td>
<td>0.371</td>
<td>0.585</td>
<td>0.791</td>
<td>0.925</td>
<td>0.981</td>
<td>0.997</td>
</tr>
<tr>
<td>Theoretical</td>
<td></td>
<td></td>
<td>0.120</td>
<td>0.245</td>
<td>0.423</td>
<td>0.623</td>
<td>0.796</td>
<td>0.911</td>
<td>0.970</td>
<td>0.992</td>
</tr>
</tbody>
</table>
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 $G_0$ is Gaussian.
- Argued popular approaches implicitly estimate $\phi'_{\theta_0}$.

Inference
- Local size control guaranteed by subadditivity and regularity.
- Application to testing if $\theta_0$ belongs to convex set.

$\Rightarrow$ Problems can be analyzed by studying $G_0$ and $\phi'_{\theta_0}$ (as in Delta method).