On Bootstrap Inconsistency and Its Repair via the Biased Bootstrap

Mihai Giurcanu\textsuperscript{1} Brett Presnell\textsuperscript{2}

\textsuperscript{1}Department of Mathematics, University of Louisiana at Lafayette
\textsuperscript{2}Department of Statistics, University of Florida

November 19, 2009
Outline

Weak Convergence of Random Distributions

Some Examples of Bootstrap Inconsistency

Repairing Inconsistency via the Biased Bootstrap

Examples Revisited

Simulations.
Some Old Friends . . .

Theorem (Slutsky)

\[
\begin{aligned}
X_n \overset{d}{\to} X \\
Y_n \overset{Pr}{\to} c
\end{aligned}
\implies
\begin{aligned}
(X_n, Y_n) \overset{d}{\to} (X, c)
\end{aligned}
\]

Theorem (H. Rubin)

\[
\begin{aligned}
X_n \overset{d}{\to} X \\
g_n(x_n) \to g(x)
\end{aligned}
\implies
\begin{aligned}
g_n(X_n) \overset{d}{\to} g(X)
\end{aligned}
\]

\[\text{For all } x \in A \text{ and all sequences } x_n \to x, \text{ where } Pr(X \in A) = 1.\]
... Applied to Random Distributions ...  

**Setup:**

\[
\begin{align*}
S_1 & \quad S_2 & \quad S_3 & \quad \text{separable metric spaces} \\
\mathcal{P}_1 & \quad \mathcal{P}_2 & \quad \mathcal{P}_3 & \quad \text{spaces of probability measures on their Borel sets} \\
& & & \\
\rho_i & = \text{metric on } \mathcal{P}_i \text{ metrizing weak convergence} \\
& & & \\
X_n & = \text{random elements of } S_1 \\
Q_n & = \text{random elements of } \mathcal{P}_2 \\
\psi_n & = \text{measurable mappings of } (S_1 \times \mathcal{P}_2) \text{ into } \mathcal{P}_3
\end{align*}
\]

**Assume:**

\[
\begin{align*}
X_n & \xrightarrow{d} X \\
Q_n & \xrightarrow{Pr} Q \text{ (nonrandom)} \\
\psi_n(x_n, Q_n) & \rightarrow \psi(x, Q) \quad \text{for all}^\ast (x_n, Q_n) \rightarrow (x, Q)
\end{align*}
\]
Setup:

$S_1$ $S_2$ $S_3$ separable metric spaces
$P_1$ $P_2$ $P_3$ spaces of probability measures on their Borel sets

$\rho_i =$ metric on $P_i$ metrizing weak convergence

$X_n =$ random elements of $S_1$
$Q_n =$ random elements of $P_2$
$\psi_n =$ measurable mappings of $(S_1 \times P_2)$ into $P_3$

Assume:

$X_n \overset{d}{\to} X$
$Q_n \overset{\text{Pr}}{\to} Q$ (nonrandom)

$\psi_n(x_n, Q_n) \to \psi(x, Q)$ for all* $(x_n, Q_n) \to (x, Q)$
\[ X_n \xrightarrow{d} X \]
\[ Q_n \xrightarrow{Pr} Q \]
\[ \psi_n(x_n, Q_n) \rightarrow \psi(x, Q) \]
\[ \text{for all}^* (x_n, Q_n) \rightarrow (x, Q) \]

\[ \psi_n(X_n, Q_n) \xrightarrow{d} \psi(X, Q) \]
$S_1$, $S_2$, $S_3$ separable metric spaces

$X_n =$ random elements of $S_1$

$Q_n =$ random elements of $P_2$

$\psi_n =$ measurable mappings of $(S_1 \times P_2)$ into $P_3$

**Proposition**

**Assume:**

$X_n \xrightarrow{d} X$

$Q_n \xrightarrow{Pr} Q$ (*nonrandom*)

$\psi_n(x_n, Q_n) \rightarrow \psi(x, Q)$ for all* $(x_n, Q_n) \rightarrow (x, Q)$

**Then:**

$\psi_n(X_n, Q_n) \xrightarrow{d} \psi(X, Q)$
\(X_n, Y_n\) random elements of \(S_1\) and \(S_2\), resp.
\(g_n(x, y)\) \(S_3\)-valued function on \(S_1 \times S_2\)
\(\mathcal{A}_n\) sub-\(\sigma\)-algebra of underlying probability space

**Corollary**

Assume:

\[X_n \in \mathcal{A}_n \quad \text{and} \quad X_n \overset{d}{\to} X \sim P\]
\[\mathcal{L}(Y_n|\mathcal{A}_n) \overset{\text{Pr}}{\to} Q \quad \text{(nonrandom)}\]
\[g_n(x_n, y_n) \to g(x, y) \quad \text{for all}^* (x_n, y_n) \to (x, y)\]

Then:

\[\mathcal{L}(g_n(X_n, Y_n)|\mathcal{A}_n) \overset{d}{\to} \mathcal{L}(g(X, Y)|X), \quad X \sim P, \quad Y \sim Q, \quad X \perp \perp Y.\]
Examples: Generalities

Each of our examples follows the same pattern

- \( X_{1:n} := X_1, \ldots, X_n \sim \text{i.i.d.} \)
- \( X^*_1, \ldots, X^*_n \) a with-replacement random sample from \( X_{1:n} \)
- \( S_n \in \sigma(X_{1:n}) \) and \( S_n \xrightarrow{d} Z \sim N(0, \Sigma) \)
- \( \mathcal{L}(S^*_n | X_{1:n}) \xrightarrow{Pr} N(0, \Sigma) \)
- \( T_n = g_n(0, S_n; \theta) \)
- \( T^*_n = g_n(S_n, S^*_n; \theta) \)
- \( g_n(x_n, y_n; \theta) \to g(x, y; \theta) \) for all \( (x_n, y_n) \to (x, y) \)

Thus

- \( T_n \xrightarrow{d} g(0, Z; \theta) \quad Z \sim N(0, \Sigma) \)
- \( \mathcal{L}(T^*_n | X_{1:n}) \xrightarrow{d} \mathcal{L}(g(Z, W; \theta) | Z) \quad Z, W \sim \text{i.i.d. } N(0, \Sigma). \)
Examples: Generalities

Each of our examples follows the same pattern

1. $X_{1:n} := X_1, \ldots, X_n \sim \text{i.i.d.}$
2. $X_{1}^{*}, \ldots, X_{n}^{*}$ a with-replacement random sample from $X_{1:n}$
3. $S_n \in \sigma(X_{1:n})$ and $S_n \xrightarrow{d} Z \sim N(0, \Sigma)$
4. $\mathcal{L}(S_n^* | X_{1:n}) \xrightarrow{Pr} N(0, \Sigma)$
5. $T_n = g_n(0, S_n; \theta)$
6. $T_n^* = g_n(S_n, S_n^*; \theta)$
7. $g_n(x_n, y_n; \theta) \rightarrow g(x, y; \theta)$ for all* $(x_n, y_n) \rightarrow (x, y)$

Thus

1. $T_n \xrightarrow{d} g(0, Z; \theta)$ and $Z \sim N(0, \Sigma)$
2. $\mathcal{L}(T_n^* | X_{1:n}) \xrightarrow{d} \mathcal{L}(g(Z, W; \theta) | Z)$ and $Z, W \sim \text{i.i.d.} \ N(0, \Sigma)$.
Examples: Generalities

Each of our examples follows the same pattern

- $X_{1:n} := X_1, \ldots, X_n \sim \text{i.i.d.}$
- $X_1^*, \ldots, X_n^*$ a with-replacement random sample from $X_{1:n}$
- $S_n \in \sigma(X_{1:n})$ and $S_n \xrightarrow{d} Z \sim N(0, \Sigma)$
- $\mathcal{L}(S_n^*|X_{1:n}) \xrightarrow{\text{Pr}} N(0, \Sigma)$
- $T_n = g_n(0, S_n; \theta)$
- $T_n^* = g_n(S_n, S_n^*; \theta)$
- $g_n(x_n, y_n; \theta) \rightarrow g(x, y; \theta)$ for all $^* (x_n, y_n) \rightarrow (x, y)$

Thus

- $T_n \xrightarrow{d} g(0, Z; \theta)$ and $Z \sim N(0, \Sigma)$
- $\mathcal{L}(T_n^*|X_{1:n}) \xrightarrow{d} \mathcal{L}(g(Z, W; \theta)|Z)$ and $Z, W \sim \text{i.i.d. } N(0, \Sigma)$. 
Examples: Generalities

Each of our examples follows the same pattern

- $X_{1:n} := X_1, \ldots, X_n \sim \text{i.i.d.}$
- $X^*_1, \ldots, X^*_n$ a with-replacement random sample from $X_{1:n}$
- $S_n \in \sigma(X_{1:n})$ and $S_n \overset{d}{\rightarrow} Z \sim N(0, \Sigma)$
- $\mathcal{L}(S^*_n | X_{1:n}) \overset{\text{Pr}}{\rightarrow} N(0, \Sigma)$
- $T_n = g_n(0, S_n; \theta)$
- $T^*_n = g_n(S_n, S^*_n; \theta)$
- $g_n(x_n, y_n; \theta) \rightarrow g(x, y; \theta)$ for all $(x_n, y_n) \rightarrow (x, y)$

Thus

- $T_n \overset{d}{\rightarrow} g(0, Z; \theta)$ and $Z \sim N(0, \Sigma)$
- $\mathcal{L}(T^*_n | X_{1:n}) \overset{d}{\rightarrow} \mathcal{L}(g(Z, W; \theta) | Z)$, with $Z, W \sim \text{i.i.d. } N(0, \Sigma)$. 
Each of our examples follows the same pattern

- $X_{1:n} := X_1, \ldots, X_n \sim \text{i.i.d.}$
- $X_1^* , \ldots, X_n^*$ a with-replacement random sample from $X_{1:n}$
- $S_n \in \sigma(X_{1:n})$ and $S_n \xrightarrow{d} Z \sim N(0, \Sigma)$
- $\mathcal{L}(S_n^* | X_{1:n}) \xrightarrow{\Pr} N(0, \Sigma)$
- $T_n = g_n(0, S_n; \theta)$
- $T_n^* = g_n(S_n, S_n^*; \theta)$
- $g_n(x_n, y_n; \theta) \to g(x, y; \theta)$ for all $^* (x_n, y_n) \to (x, y)$

Thus

- $T_n \xrightarrow{d} g(0, Z; \theta)$ and $Z \sim N(0, \Sigma)$
- $\mathcal{L}(T_n^* | X_{1:n}) \xrightarrow{d} \mathcal{L}(g(Z, W; \theta) | Z)$ and $Z, W \sim \text{i.i.d. } N(0, \Sigma)$. 
Each of our examples follows the same pattern

- $X_{1:n} := X_1, \ldots, X_n \sim \text{i.i.d.}$
- $X_{1}^*, \ldots, X_{n}^*$ a with-replacement random sample from $X_{1:n}$
- $S_n \in \sigma(X_{1:n})$ and $S_n \overset{d}{\rightarrow} Z \sim \mathcal{N}(0, \Sigma)$
- $\mathcal{L}(S_n^*|X_{1:n}) \overset{\text{Pr}}{\rightarrow} \mathcal{N}(0, \Sigma)$
- $T_n = g_n(0, S_n; \theta)$
- $T_n^* = g_n(S_n, S_n^*; \theta)$
- $g_n(x_n, y_n; \theta) \rightarrow g(x, y; \theta)$ for all $(x_n, y_n) \rightarrow (x, y)$

Thus

- $T_n \overset{d}{\rightarrow} g(0, Z; \theta)$ and $Z \sim \mathcal{N}(0, \Sigma)$
- $\mathcal{L}(T_n^*|X_{1:n}) \overset{d}{\rightarrow} \mathcal{L}(g(Z, W; \theta)|Z)$ and $Z, W \sim \text{i.i.d. } \mathcal{N}(0, \Sigma)$. 
Each of our examples follows the same pattern

- $X_{1:n} := X_1, \ldots, X_n \sim \text{i.i.d.}$
- $X_1^*, \ldots, X_n^*$ a with-replacement random sample from $X_{1:n}$
- $S_n \in \sigma(X_{1:n})$ and $S_n \xrightarrow{d} Z \sim \mathcal{N}(0, \Sigma)$
- $\mathcal{L}(S_n^*|X_{1:n}) \xrightarrow{\text{Pr}} \mathcal{N}(0, \Sigma)$
- $T_n = g_n(0, S_n; \theta)$
- $T_n^* = g_n(S_n, S_n^*; \theta)$
- $g_n(x_n, y_n; \theta) \rightarrow g(x, y; \theta)$ for all* $(x_n, y_n) \rightarrow (x, y)$

Thus

- $T_n \xrightarrow{d} g(0, Z; \theta)$ and $Z \sim \mathcal{N}(0, \Sigma)$
- $\mathcal{L}(T_n^*|X_{1:n}) \xrightarrow{d} \mathcal{L}(g(Z, W; \theta)|Z)$ and $Z, W \sim \text{i.i.d.} \mathcal{N}(0, \Sigma)$. 
Each of our examples follows the same pattern

- $X_{1:n} := X_1, \ldots, X_n \sim \text{i.i.d.}$
- $X_1^*, \ldots, X_n^*$ a with-replacement random sample from $X_{1:n}$
- $S_n \in \sigma(X_{1:n})$ and $S_n \xrightarrow{d} Z \sim N(0, \Sigma)$
- $\mathcal{L}(S_n^*|X_{1:n}) \xrightarrow{\text{Pr}} N(0, \Sigma)$
- $T_n = g_n(0, S_n; \theta)$
- $T_n^* = g_n(S_n, S_n^*; \theta)$
- $g_n(x_n, y_n; \theta) \xrightarrow{\text{all}} g(x, y; \theta)$ for all* $(x_n, y_n) \rightarrow (x, y)$

Thus

- $T_n \xrightarrow{d} g(0, Z; \theta)$ $Z \sim N(0, \Sigma)$
- $\mathcal{L}(T_n^*|X_{1:n}) \xrightarrow{d} \mathcal{L}(g(Z, W; \theta)|Z)$ $Z, W \sim \text{i.i.d. } N(0, \Sigma)$. 

*Note: All* refers to the convergence of all random variables in the sequence to their respective limits.
Each of our examples follows the same pattern

- $X_{1:n} := X_1, \ldots, X_n \sim \text{i.i.d.}$
- $X_1^*, \ldots, X_n^*$ a with-replacement random sample from $X_{1:n}$
- $S_n \in \sigma(X_{1:n})$ and $S_n \xrightarrow{d} Z \sim N(0, \Sigma)$
- $\mathcal{L}(S_n^* | X_{1:n}) \xrightarrow{Pr} N(0, \Sigma)$
- $T_n = g_n(0, S_n; \theta)$
- $T_n^* = g_n(S_n, S_n^*; \theta)$
- $g_n(x_n, y_n; \theta) \rightarrow g(x, y; \theta)$ for all $(x_n, y_n) \rightarrow (x, y)$

Thus

- $T_n \xrightarrow{d} g(0, Z; \theta) \quad Z \sim N(0, \Sigma)$
- $\mathcal{L}(T_n^* | X_{1:n}) \xrightarrow{d} \mathcal{L}(g(Z, W; \theta) | Z) \quad Z, W \sim \text{i.i.d. } N(0, \Sigma)$.
\( T_n \overset{d}{\rightarrow} g(0, Z; \theta) \quad Z \sim N(0, \Sigma) \)

\( \mathcal{L}(T_n^*|X_{1:n}) \overset{d}{\rightarrow} \mathcal{L}(g(Z, W; \theta)|Z) \quad Z, W \sim \text{i.i.d. } N(0, \Sigma). \)

For most values of \( \theta \):

\[ g(x, y; \theta) = g(0, y; \theta) \]

so that

\[ \mathcal{L}(g(Z, W; \theta)|Z) = \mathcal{L}(g(0, W; \theta)|Z) = \mathcal{L}(g(0, W; \theta)) = \mathcal{L}(g(0, Z; \theta)) \checkmark \]

For other values of \( \theta \):

\( g(x, y; \theta) \) depends on \( x \) and

\[ \mathcal{L}(g(Z, W; \theta)|Z) \] is random (varies with \( Z \)) \xmark
Example: Hodges’ Estimator

\[ X_1, \ldots, X_n \sim \text{i.i.d. } (\mu, 1) \]

\[ 0 \leq b < 1 \]

\[ \hat{\mu}_n = \begin{cases} 
\overline{X}_n & \text{if } |\overline{X}_n| > n^{-1/4} \\
 b \overline{X}_n & \text{if } |\overline{X}_n| \leq n^{-1/4}
\end{cases} \]

\[ S_n = n^{1/2}(\overline{X}_n - \mu) \quad S^*_n = n^{1/2}(\overline{X}^*_n - \overline{X}_n) \]

\[ S_n \xrightarrow{d} Z \sim N(0, 1) \quad \mathcal{L}(S^*_n | X_1: n) \xrightarrow{\text{Pr}} N(0, 1) \]

\[ T_n = n^{1/2}(\hat{\mu}_n - \mu) \quad T^*_n = n^{1/2}(\hat{\mu}^*_n - \overline{X}_n) \]

\[ g_n(x, y; \mu) = y - (1 - b)(x + y + n^{1/2}\mu) \mathbb{I}(|x + y + n^{1/2}\mu| \leq n^{1/4}) \]

\[ g(x, y; \mu) = \begin{cases} 
y, & \text{if } \mu \neq 0, 
by + (1 - b)x, & \text{if } \mu = 0
\end{cases} \]
Example: Hodges’ Estimator (p2)

\[ g(x, y; \mu) = \begin{cases} 
  y, & \text{if } \mu \neq 0, \\
  by + (1 - b)x, & \text{if } \mu = 0
\end{cases} \]

\[ T_n \xrightarrow{d} g(0, Z; \mu) = \begin{cases} 
  Z & \text{if } \mu \neq 0 \\
  bZ & \text{if } \mu = 0
\end{cases} \quad Z \sim N(0, 1) \]

\[ \mathcal{L}(T_n^*|X_{1:n}) \xrightarrow{d} \mathcal{L}(g(Z, W; \mu)|Z) \quad Z, W \sim \text{i.i.d. } N(0, 1) \]

\[ \mu \neq 0 : \quad g(x, y; \mu) = g(0, y; \mu) = y \]

\[ \mathcal{L}(g(Z, W; \mu)|Z) = \mathcal{L}(W|Z) = \mathcal{L}(W) = N(0, 1) \checkmark \]

\[ \mu = 0 : \quad g(x, y; \mu) = by + (1 - b)x \]

\[ \mathcal{L}(g(Z, W; 0)|Z) = \mathcal{L}(bW + (1 - b)Z|Z) = N((1 - b)Z, b^2) \times \]
Example: Lindley’s Estimator

\[ X_1, \ldots, X_n \sim \text{i.i.d. } (\mu, I_p) \quad (p \geq 4) \]

\[ \hat{\mu}_n = m(\bar{X}_n)1_p + \left\{ 1 - \frac{p - 3}{n\|\bar{X}_n - m(\bar{X}_n)1_p\|^2} \right\} \{ \bar{X}_n - m(\bar{X}_n)1_p \} \]

where

\[ m(x) := \frac{1}{p} \sum_{i=1}^{p} x^{(i)}, \quad x \in \mathbb{R}^p \]
Example: Lindley’s Estimator (p2)

\[ S_n = n^{1/2}(\bar{X}_n - \mu) \quad S^*_n = n^{1/2}(\bar{X}^*_n - \bar{X}_n) \]

\[ S_n \xrightarrow{d} Z \sim N(0, I_p) \quad \mathcal{L}(S^*_n|X_{1:n}) \xrightarrow{Pr} N(0, I_p) \]

\[ T_n = n^{1/2}(\hat{\mu}_n - \mu) \quad T^*_n = n^{1/2}(\hat{\mu}^*_n - \bar{X}_n) \]

\[ g_n(x, y; \mu) = y - \frac{(p - 3)[(x + y) - m(x + y)1_p + n^{1/2}\{\mu - m(\mu)1_p\}]}{\| (x + y) - m(x + y)1_p + n^{1/2}\{\mu - m(\mu)1_p\} \|^2} \]

\[ g(x, y; \mu) = \begin{cases} y & \text{if } \mu \neq m(\mu)1_p \\ y - \frac{(p - 3)[(x + y) - m(x + y)1_p]}{\| (x + y) - m(x + y)1_p \|^2} & \text{if } \mu = m(\mu)1_p \end{cases} \]
Example: Lindley’s Estimator (p3)

\[ T_n \xrightarrow{d} g(0, Z; \mu) \quad Z \sim N(0, I_p) \]

\[ \mathcal{L}(T_n^*|X_{1:n}) \xrightarrow{d} \mathcal{L}(g(Z, W; \mu)|Z) \quad Z, W \sim \text{i.i.d. } N(0, I_p) \]

\[ \mu \neq m(\mu)1_p : \quad g(x, y; \mu) = g(0, y; \mu) = y \]

\[ T_n \xrightarrow{d} Z \sim N(0, I_p) \]

\[ \mathcal{L}(g(Z, W; \mu)|Z) = \mathcal{L}(W|Z) = \mathcal{L}(W) = N(0, I_p) \checkmark \]

\[ \mu = m(\mu)1_p : \]

\[ T_n \xrightarrow{d} Z - \frac{(p - 3)\{Z - m(Z)1_p\}}{\|Z - m(Z)1_p\|^2}, \quad Z \sim N(0, I_p) \]

\[ \mathcal{L}(g(Z, W; \mu)|Z) = \mathcal{L}\left(W - \frac{(p - 3)[(Z + W) - m(Z + W)1_p]}{\|(Z + W) - m(Z + W)1_p\|^2} \mid Z\right) \times \]
Example: The LASSO (Correlation Model)

\[(X_1, Y_1), \ldots, (X_n, Y_n) \text{ i.i.d.} \quad E[(Y_i - X_i^T \beta)X_i] = 0\]

\[\hat{\beta}_n = \arg\min_b \left\{ \sum_{i=1}^{n} (Y_i - X_i^T b)^2 + \lambda_n \sum_{j=1}^{p} |b^{(j)}| \right\} \]

Assume: \(E(\|(X_i, Y_i)\|_4^4) < \infty\)

\[V := \text{Var}[(Y_i - X_i^T \beta)X_i] \text{ positive definite}\]

\[C := E(X_iX_i^T) \text{ positive definite}\]

\[n^{-1/2} \lambda_n \to \lambda_0\]

Let: \(\Sigma = C^{-1}VC^{-1}\)

Bootstrapping Cases

\[(X_1^*, Y_1^*), \ldots, (X_n^*, Y_n^*) \text{ is a with-replacement random sample from } (X, Y)_{1:n} = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}.\]
Example: LASSO (Correlation Model) (p2)

\[ C_n = n^{-1} \sum_{i=1}^{n} X_i X_i^T \]

\[ \bar{\beta}_n = C_n^{-1} n^{-1} \sum_{i=1}^{n} X_i Y_i \quad (LSE) \]

\[ S_n = n^{1/2}(\bar{\beta}_n - \beta) \]

\[ T_n = n^{1/2}(\hat{\beta}_n - \beta) \]

\[ C_n \xrightarrow{Pr} C \]

\[ S_n \xrightarrow{d} Z \sim N(0, \Sigma) \]

\[ C_n^* = n^{-1} \sum_{i=1}^{n} X_i^* X_i^{*T} \]

\[ \bar{\beta}_n^* = C_n^{*-1} n^{-1} \sum_{i=1}^{n} X_i^* Y_i^* \]

\[ S_n^* = n^{1/2}(\bar{\beta}_n^* - \bar{\beta}_n) \]

\[ T_n^* = n^{1/2}(\hat{\beta}_n^* - \bar{\beta}_n) \]

\[ C_n^* \xrightarrow{Pr} C \]

\[ \mathcal{L}(S_n^*|X_{1:n}) \xrightarrow{Pr} N(0, \Sigma) \]
Example: LASSO (Correlation Model) (p3)

\[ g_n(x, y, C; \beta) = \]
\[ y + \arg\min_u \left\{ u^T Cu + n^{-1/2} \lambda_n \sum_{j=1}^{p} \left( |n^{1/2}\beta^{(j)} + (x + y + u)^{(j)}| - |n^{1/2}\beta^{(j)}| \right) \right\} \]

\[ g(x, y, C; \beta) = \]
\[ y + \arg\min_u \left\{ u^T Cu + \lambda_0 \sum_{j=1}^{p} \left( u^{(j)} \text{sign}(\beta^{(j)}) + |(x + y + u)^{(j)}| I(\beta^{(j)} = 0) \right) \right\} \]

Note: \( \beta^{(j)} \neq 0 \) for all \( j = 1, \ldots, p \) \( \implies \) \( g(x, y, C; \beta) = g(0, y, C; \beta) \)
Example: LASSO (Correlation Model) (p4)

\[ T_n \xrightarrow{d} g(0, Z, C; \beta), \quad Z \sim N(0, \Sigma) \]

\[ \mathcal{L}(T_n^*|X_{1:n}) \xrightarrow{d} \mathcal{L}(g(Z, W, C; \beta)|Z) \quad Z, W \sim \text{i.i.d. } N(0, \Sigma) \]

\( \beta(j) \neq 0 \) for all \( j = 1, \ldots, p \) : 

\[ g(x, y, C; \beta) = g(0, y, C; \beta) \]

\[ \mathcal{L}(g(Z, W, C; \beta)|Z) = \mathcal{L}(g(0, W, C; \beta)|Z) = \mathcal{L}(g(0, W, C; \beta)) \checkmark \]

\( \beta(j) = 0 \) for at least one \( j = 1, \ldots, p \) :

\[ \mathcal{L}(g(Z, W, C; \beta)|Z) \text{ varies with } Z \text{ and is random} \times \]
A Way to Repair the Bootstrap

Why is the bootstrap inconsistent in these examples?

▶ Limiting distribution of $T_n$ is discontinuous in $\mu$.
▶ Resampling from empirical distribution with mean $\overline{X}_n \neq \mu$.

Idea

▶ Resample from a distribution with mean $\tilde{\mu}_n$ chosen to fix bootstrap inconsistency.

Candidate Family of Distributions

▶ Weighted empirical distribution with weight $w_i$ on $X_i$, $i = 1, \ldots, n$.
▶ For a given value of $\mu$, choose the weights to
  ▶ minimize Kullback-Leibler divergence from empirical distribution
  ▶ subject to constraint $\sum_{i=1}^{n} w_i X_i = \mu$.
▶ Defines a pseudo-parametric family of distributions $\{\hat{P}_\mu\}$. 
Theorem

Suppose:

1. \( \Theta = \bigcup_{i \in \mathcal{I}} \Theta_i \subseteq \mathbb{R}^p. \)
2. \( L(S^*_n|X_{1:n}) \xrightarrow{P^n_\theta} Q_\theta \text{ for all } \theta \in \Theta. \)
3. \( Q_\theta(A_\theta) = 1 \text{ for all } \theta \in \Theta. \)
4. \( \tilde{\theta}_n \xrightarrow{P^n_\theta} \theta \text{ and } P^n_\theta(\tilde{\theta}_n \in \Theta_i) \rightarrow 1 \text{ for all } \theta \in \Theta_i \text{ and } i \in \mathcal{I}. \)
5. \( \text{For all } i \in \mathcal{I} \text{ and all } \theta \in \Theta_i, \ g_n(x_n; \theta_n) \rightarrow g(x; \theta) \text{ for all sequences } x_n \rightarrow x \in A_\theta \text{ and } \theta_n \rightarrow \theta \text{ with } \{\theta_n, n \geq 1\} \subseteq \Theta_i. \)

Then for all \( \theta \in \Theta: \ L\left( g_n(S^*_n, \tilde{\theta}_n) | X_{1:n} \right) \xrightarrow{P^n_\theta} L(g(Z, \theta)), Z \sim Q_\theta. \)

Corollary

If \( L(S_n) \xrightarrow{P^n_\theta} Q_\theta \text{ for all } \theta \in \Theta, \) then the conditional distribution of \( T^*_n = g_n(S^*_n, \tilde{\theta}_n) \) given \( X_{1:n} \) consistently estimates the distribution of \( T_n = g_n(S_n, \theta) \text{ under all values of } \theta \in \Theta. \)
Theorem

Suppose:

- $\Theta = \biguplus_{i \in I} \Theta_i \subset \mathbb{R}^p$.
- $L(S_n^* | X_{1:n}) \xrightarrow{P^n} Q_\theta$ for all $\theta \in \Theta$.
- $Q_\theta(A_\theta) = 1$ for all $\theta \in \Theta$.
- $\tilde{\theta}_n \xrightarrow{P^n} \theta$ and $P^n_\theta(\tilde{\theta}_n \in \Theta_i) \to 1$ for all $\theta \in \Theta_i$ and $i \in I$.
- For all $i \in I$ and all $\theta \in \Theta_i$, $g_n(x_n; \theta_n) \to g(x; \theta)$ for all sequences $x_n \to x \in A_\theta$ and $\theta_n \to \theta$ with $\{\theta_n, n \geq 1\} \subset \Theta_i$.

Then for all $\theta \in \Theta$: $L(g_n(S_n^*, \tilde{\theta}_n) | X_{1:n}) \xrightarrow{P^n} L(g(Z, \theta))$, $Z \sim Q_\theta$.

Corollary

If $L(S_n) \xrightarrow{P^n} Q_\theta$ for all $\theta \in \Theta$, then the conditional distribution of $T_n^* = g_n(S_n^*, \tilde{\theta}_n)$ given $X_{1:n}$ consistently estimates the distribution of $T_n = g_n(S_n, \theta)$ under all values of $\theta \in \Theta$. 
Theorem

Suppose:

- ▶ \( \Theta = \bigcup_{i \in I} \Theta_i \subset \mathbb{R}^p \).
- ▶ \( \mathcal{L}(S^*_n | X_{1:n}) \xrightarrow{P^n_{\theta}} Q_{\theta} \) for all \( \theta \in \Theta \).
- ▶ \( Q_{\theta}(A_{\theta}) = 1 \) for all \( \theta \in \Theta \).
- ▶ \( \tilde{\theta}_n \xrightarrow{P^n_{\theta}} \theta \) and \( P^n_{\theta}(\tilde{\theta}_n \in \Theta_i) \rightarrow 1 \) for all \( \theta \in \Theta_i \) and \( i \in I \).
- ▶ For all \( i \in I \) and all \( \theta \in \Theta_i \), \( g_n(x_n; \theta_n) \rightarrow g(x; \theta) \) for all sequences \( x_n \rightarrow x \in A_{\theta} \) and \( \theta_n \rightarrow \theta \) with \( \{\theta_n, n \geq 1\} \subset \Theta_i \).

Then for all \( \theta \in \Theta \): \( \mathcal{L}(g_n(S^*_n, \tilde{\theta}_n) | X_{1:n}) \xrightarrow{P^n_{\theta}} \mathcal{L}(g(Z, \theta)), Z \sim Q_{\theta} \).

Corollary

If \( \mathcal{L}(S_n) \xrightarrow{P^n_{\theta}} Q_{\theta} \) for all \( \theta \in \Theta \), then the conditional distribution of \( T^*_n = g_n(S^*_n, \tilde{\theta}_n) \) given \( X_{1:n} \) consistently estimates the distribution of \( T_n = g_n(S_n, \theta) \) under all values of \( \theta \in \Theta \).
Theorem

Assume

\[ X_1, \ldots, X_n \sim \text{i.i.d.} \ (\mu, \Sigma) \ (\text{finite}) \]

\( \tilde{\mu}_n \) is \( \sqrt{n} \)-consistent for \( \mu \)

Let

\[ X^*_1, \ldots, X^*_n \sim \text{i.i.d.} \ \hat{P}_{\tilde{\mu}_n} \ \text{conditional on} \ X_{1:n} = \{X_1, \ldots, X_n\} \]

Then

\[ \mathcal{L} \left( n^{1/2}(\bar{X}^*_n - \tilde{\mu}_n) | X_{1:n} \right) \xrightarrow{\text{Pr}} N(0, \Sigma) \]
Theorem

Assume

\[ X_1, \ldots, X_n \sim \text{i.i.d.} \ (\mu, \Sigma) \ (\text{finite}) \]

\[ \tilde{\mu}_n \text{ is } \sqrt{n}\text{-consistent for } \mu \]

Let

\[ X_1^*, \ldots, X_n^* \sim \text{i.i.d. } \hat{P}_{\tilde{\mu}_n} \quad \text{conditional on } X_{1:n} = \{X_1, \ldots, X_n\} \]

Then

\[ \mathcal{L} \left( n^{1/2}(\bar{X}_n^* - \tilde{\mu}_n) \mid X_{1:n} \right) \xrightarrow{\Pr} N(0, \Sigma) \]
Example (Hodges Estimator)

\( \Theta_0 = \{0\}, \Theta_1 = (-\infty, 0) \cup (0, \infty) \).

\( g_n(x; \mu) = x - (1 - b)(x + n^{1/2} \mu) I(|x + n^{1/2} \mu| \leq n^{1/4}) \)

\( g(x; \mu) = x - (1 - b)x I(\mu = 0) = \begin{cases} x & \text{if } \mu \neq 0 \\ bx & \text{if } \mu = 0 \end{cases} \)

\( \hat{\mu}_n = \bar{X}_n I(|\bar{X}_n| \geq a_n) \) with \( a_n \to 0 \) and \( n^{1/2} a_n \to \infty \).
Examples Revisited

Example (Lindley’s Estimator)

\( \Theta_0 = \{ \mu \in \mathbb{R}^p : \mu = m(\mu)1_p \} \), \( \Theta_1 = \{ \mu \in \mathbb{R}^p : \mu \neq m(\mu)1_p \} \).

\( g_n(x; \mu) = x - \frac{(p - 3)[(x) - m(x)1_p + n^{1/2}\{\mu - m(\mu)1_p\}]}{\| (x) - m(x)1_p + n^{1/2}\{\mu - m(\mu)1_p\} \|^2} \)

\( g(x; \mu) = \begin{cases} x, & \text{if } \mu \neq m(\mu)1_p, \\ x - \frac{(p - 3)[(x) - m(x)1_p]}{\| (x) - m(x)1_p \|^2}, & \text{if } \mu = m(\mu)1_p. \end{cases} \)

\( \tilde{\mu}_n = \begin{cases} \bar{X}_n, & \text{if } \| \bar{X}_n - m(\bar{X}_n)1_p \| > a_n \\ m(\bar{X}_n)1_p, & \text{if } \| \bar{X}_n - m(\bar{X}_n)1_p \| \leq a_n \end{cases} \)

with \( a_n \to 0 \) and \( n^{1/2}a_n \to \infty \).
Examples Revisited

Example (LASSO)

\[ \Theta = \mathbb{R}^p = \bigcup_{J \in \mathcal{I}} \Theta_J \text{ with } \mathcal{I} = \{ J : J \subset \{1, \ldots, p\} \} \quad \text{and} \]

\[ \Theta_J = \{ \beta \in \mathbb{R}^p : \beta^{(j)} = 0 \text{ for all } j \in J, \beta^{(j)} \neq 0 \text{ for all } j \notin J \} \]

\[ g_n(x, C; \beta) = x + \arg\min_u \left\{ u^T C u + n^{-1/2} \lambda_n \sum_{j=1}^p \left( |n^{1/2} \beta^{(j)} + (x + u)^{(j)}| - |n^{1/2} \beta^{(j)}| \right) \right\} \]

\[ g(x, C; \beta) = x + \arg\min_u \left\{ u^T C u + \lambda_0 \sum_{j=1}^p \left( u^{(j)} \text{ sign}(\beta^{(j)}) + |(x + u)^{(j)}| I(\beta^{(j)} = 0) \right) \right\} \]

\[ \tilde{\beta}_n = \text{an estimator with the "oracle" property, e.g., SCAD or adaptive-LASSO (but not the LASSO).} \]
Simulations: Hodges’ Estimator \((b = 0.5)\), \(a_n = n^{-1/2} \log_{10} n\)
Simulations: Hodges Estimator ($b = 0.5$)

$n = 25$

Mean KS Distance

- $n^{-1/4}$
- $n^{-1/3}$
- $n^{-1/2 \log_{10} n}$
- $n^{-2/3}$
- uniform

$\mu$
Simulations: Hodges Estimator ($b = 0.5$)
The End

No one will ever see this slide.
Proof of Corollary 1

Suppose \( x_n \to x \) and \( Y'_n \overset{d}{\to} Y \sim Q \).

Let \( h_n(y) := g_n(x_n, y) \) and \( h(y) := g(x, y) \).

\[
\begin{align*}
Y'_n \overset{d}{\to} Y \\
h_n(y_n) \to h(y) \\
\text{for all}^* \ y_n \to y
\end{align*}
\]

\[
\overset{\text{H.Rubin}}{\implies} h_n(Y'_n) \overset{d}{\to} h(Y) \\
g_n(x_n, Y'_n) \overset{d}{\to} g(x, Y)
\]

Translation: suppose \( x_n \to x \) and \( Q_n \to Q \). Then

\[
\psi_n(x_n, Q_n) \to \psi(x, Q)
\]

where

\[
\begin{align*}
\psi_n(x, Q') &:= \mathcal{L}(g_n(x, Y')) \\
\psi(x, Q') &:= \mathcal{L}(g(x, Y')) \quad \text{when} \ Y' \sim Q'
\end{align*}
\]
Proof of Corollary (p2)

\[ \psi_n(x, Q') := \mathcal{L}(g_n(x, Y')) \]
\[ \psi(x, Q') := \mathcal{L}(g(x, Y')) \]

\[ \psi_n(x_n, Q_n) \xrightarrow{\text{for all}^*} \psi(x, Q) \quad \text{for all}^* \ x_n \rightarrow x \quad \text{and} \ Q_n \rightarrow Q \]

Let \( Q_n = \mathcal{L}(Y_n|A_n) \). Then

\[ X_n \xrightarrow{d} X \]
\[ Q_n \xrightarrow{\text{Pr}} Q \]

\[ \psi_n(x_n, Q_n) \xrightarrow{\text{for all}^*} \psi(x, Q) \]

\[ \xrightarrow{\text{Proposition}} \]

\[ \psi_n(X_n, Q_n) \xrightarrow{d} \psi(X, Q) \]

Return to Corollary 1
Proof of Corollary (p3)

\[ \psi_n(x, Q') := \mathcal{L}(g_n(x, Y')) \quad Y' \sim Q' \]
\[ \psi(x, Q') := \mathcal{L}(g(x, Y')) \quad Y' \sim Q' \]
\[ Q_n := \mathcal{L}(Y_n|\mathcal{A}_n) \]
\[ \psi_n(X_n, Q_n) \xrightarrow{d} \psi(X, Q) \quad X \sim P \]

But

\[ X_n \in \mathcal{A}_n \]
\[ \mathcal{Q}_n := \mathcal{L}(Y_n|\mathcal{A}_n) \] \implies \[ \mathcal{L}(g_n(X_n, Y_n)|\mathcal{A}_n) = \psi_n(X_n, Q_n) \]

and

\[ \psi(X, Q) = \mathcal{L}(g(X, Y)|X), \quad X \sim P, \quad Y \sim Q, \quad X \perp \perp Y \]

so

\[ \mathcal{L}(g_n(X_n, Y_n)|\mathcal{A}_n) \xrightarrow{d} \mathcal{L}(g(X, Y)|X), \quad X \sim P, \quad Y \sim Q, \quad X \perp \perp Y \]