On Bootstrap Inconsistency and Its Repair via the Biased Bootstrap

Mihai Giurcanu[1] Brett Presnell[2]

[1]Department of Mathematics, University of Louisiana at Lafayette

[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)

$$\left.\begin{array}{l} X_n \xrightarrow{d} X \\ Y_n \xrightarrow{\text{Pr}} c \end{array}\right\} \implies (X_n, Y_n) \xrightarrow{d} (X, c)$$

Theorem (H. Rubin)

$$\left.\begin{array}{l} X_n \xrightarrow{d} X \\ \\ g_n(x_n) \to g(x) \quad \textit{for all}^* \ x_n \to x \end{array}\right\} \implies g_n(X_n) \xrightarrow{d} g(X)$$

* *For all* $x \in A$ *and all sequences* $x_n \to x$, *where* $\Pr(X \in A) = 1$.

. . . Applied to Random Distributions . . .

Setup:

$S_1 \quad S_2 \quad S_3$ separable metric spaces

$\mathscr{P}_1 \quad \mathscr{P}_2 \quad \mathscr{P}_3$ spaces of probability measures on their Borel sets

$\rho_i =$ metric on $\mathscr{P}_i$ metrizing weak convergence

$X_n =$ random elements of S_1

$\mathbb{Q}_n =$ random elements of $\mathscr{P}_2$

$\psi_n =$ measurable mappings of $(S_1 \times \mathscr{P}_2)$ into $\mathscr{P}_3$

Assume:

$$X_n \xrightarrow{\text{d}} X$$

$$\mathbb{Q}_n \xrightarrow{\text{Pr}} Q \text{ (nonrandom)}$$

$$\psi_n(x_n, Q_n) \to \psi(x, Q) \quad \text{for all}^* \ (x_n, Q_n) \to (x, Q)$$

... Applied to Random Distributions ...

Setup:

S_1 S_2 S_3 separable metric spaces

$\mathscr{P}_1$ $\mathscr{P}_2$ $\mathscr{P}_3$ spaces of probability measures on their Borel sets

$\rho_i =$ metric on $\mathscr{P}_i$ metrizing weak convergence

$X_n =$ random elements of S_1

$\mathbb{Q}_n =$ random elements of $\mathscr{P}_2$

$\psi_n =$ measurable mappings of $(S_1 \times \mathscr{P}_2)$ into $\mathscr{P}_3$

Assume:

$X_n \xrightarrow{\text{d}} X$

$\mathbb{Q}_n \xrightarrow{\text{Pr}} Q$ (nonrandom)

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

...

$$\left.\begin{array}{r} X_n \xrightarrow{\text{d}} X \\ \mathbb{Q}_n \xrightarrow{\text{Pr}} Q \end{array}\right\} \xRightarrow{\text{Slutsky}} (X_n, \mathbb{Q}_n) \xrightarrow{\text{d}} (X, Q)$$

$$\psi_n(x_n, Q_n) \to \psi(x, Q)$$

$$\text{for all}^* \ (x_n, Q_n) \to (x, Q)$$

$$\xRightarrow{\text{H.Rubin}} \psi_n(X_n, \mathbb{Q}_n) \xrightarrow{\text{d}} \psi(X, Q)$$

$S_1 \quad S_2 \quad S_3$ separable metric spaces

$X_n =$ random elements of S_1

$\mathbb{Q}_n =$ random elements of $\mathscr{P}_2$

$\psi_n =$ measurable mappings of $(S_1 \times \mathscr{P}_2)$ into $\mathscr{P}_3$

Proposition

Assume:

$$X_n \xrightarrow{d} X$$

$$\mathbb{Q}_n \xrightarrow{\Pr} Q \text{ (nonrandom)}$$

$$\psi_n(x_n, Q_n) \to \psi(x, Q) \quad \textit{for all}^* \ (x_n, Q_n) \to (x, Q)$$

Then:

$$\psi_n(X_n, \mathbb{Q}_n) \xrightarrow{d} \psi(X, Q)$$

X_n, Y_n	random elements of S_1 and S_2, resp.
$g_n(x, y)$	S_3-valued function on $S_1 \times S_2$
$\mathscr{A}_n$	sub-σ-algebra of underlying probability space

Corollary

Assume:

$$X_n \in \mathscr{A}_n \quad \textit{and} \quad X_n \xrightarrow{d} X \sim P$$

$$\mathscr{L}(Y_n | \mathscr{A}_n) \xrightarrow{\Pr} Q \text{ (nonrandom)}$$

$$g_n(x_n, y_n) \to g(x, y) \quad \textit{for all}^* \ (x_n, y_n) \to (x, y)$$

Then:

$$\mathscr{L}\big(g_n(X_n, Y_n)\big|\mathscr{A}_n\big) \xrightarrow{d} \mathscr{L}\big(g(X, Y)\big|X\big), \quad X \sim P, \quad Y \sim Q, \quad X \perp\!\!\!\perp Y.$$

▸ Proof

Examples: Generalities

Each of our examples follows the same pattern

- $X_{1:n} := X_1, \ldots, X_n \sim$ 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 $S_n \xrightarrow{\text{d}} Z \sim N(0, \Sigma)$
- $\mathscr{L}(S_n^* | X_{1:n}) \xrightarrow{\text{Pr}} N(0, \Sigma)$
- $T_n = g_n(0, S_n; \theta)$
- $T_n^* = g_n(S_n, S_n^*; \theta)$
- $g_n(x_n, y_n; \theta) \to g(x, y; \theta)$ for all* $(x_n, y_n) \to (x, y)$

Thus

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

Examples: Generalities

Each of our examples follows the same pattern

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

Thus

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

Examples: Generalities

Each of our examples follows the same pattern

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

Thus

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

Examples: Generalities

Each of our examples follows the same pattern

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

Thus

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

Examples: Generalities

Each of our examples follows the same pattern

- $X_{1:n} := X_1, \ldots, X_n \sim$ 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 $S_n \xrightarrow{\text{d}} Z \sim N(0, \Sigma)$
- $\mathscr{L}(S_n^* | X_{1:n}) \xrightarrow{\text{Pr}} N(0, \Sigma)$
- $T_n = g_n(0, S_n; \theta)$
- $T_n^* = g_n(S_n, S_n^*; \theta)$
- $g_n(x_n, y_n; \theta) \to g(x, y; \theta)$ for all* $(x_n, y_n) \to (x, y)$

Thus

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

Examples: Generalities

Each of our examples follows the same pattern

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

Thus

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

Examples: Generalities

Each of our examples follows the same pattern

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

Thus

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

Examples: Generalities

Each of our examples follows the same pattern

- $X_{1:n} := X_1, \ldots, X_n \sim$ 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 $S_n \xrightarrow{d} Z \sim N(0, \Sigma)$
- $\mathscr{L}(S_n^* | X_{1:n}) \xrightarrow{\Pr} N(0, \Sigma)$
- $T_n = g_n(0, S_n; \theta)$
- $T_n^* = g_n(S_n, S_n^*; \theta)$
- $g_n(x_n, y_n; \theta) \to g(x, y; \theta)$ for all* $(x_n, y_n) \to (x, y)$

Thus

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

Examples: Generalities

Each of our examples follows the same pattern

- $X_{1:n} := X_1, \ldots, X_n \sim$ 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 $S_n \xrightarrow{\text{d}} Z \sim N(0, \Sigma)$
- $\mathscr{L}(S_n^* | X_{1:n}) \xrightarrow{\text{Pr}} N(0, \Sigma)$
- $T_n = g_n(0, S_n; \theta)$
- $T_n^* = g_n(S_n, S_n^*; \theta)$
- $g_n(x_n, y_n; \theta) \to g(x, y; \theta)$ for all* $(x_n, y_n) \to (x, y)$

Thus

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

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

For most values of θ:

$$g(x, y; \theta) = g(0, y; \theta)$$

so that

$$\begin{aligned}\mathscr{L}\big(g(Z, W; \theta) \big| Z\big) &= \mathscr{L}\big(g(0, W; \theta) \big| Z\big) \\ &= \mathscr{L}\big(g(0, W; \theta)\big) = \mathscr{L}\big(g(0, Z; \theta)\big) ✔\end{aligned}$$

For other values of θ:

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

$$\mathscr{L}\big(g(Z, W; \theta) \big| Z\big) \text{ is random (varies with } Z) ✘$$

Example: Hodges' Estimator

$X_1, \ldots, X_n \sim$ i.i.d. $(\mu, 1)$ $\qquad 0 \le b < 1$

$$\hat{\mu}_n = \begin{cases} \bar{X}_n & \text{if } |\bar{X}_n| > n^{-1/4} \\ b\,\bar{X}_n & \text{if } |\bar{X}_n| \le n^{-1/4} \end{cases}$$

$$S_n = n^{1/2}(\bar{X}_n - \mu) \qquad S_n^* = n^{1/2}(\bar{X}_n^* - \bar{X}_n)$$

$$S_n \xrightarrow{\text{d}} Z \sim N(0,1) \qquad \mathscr{L}(S_n^* | X_{1:n}) \xrightarrow{\text{Pr}} N(0,1)$$

$$T_n = n^{1/2}(\hat{\mu}_n - \mu) \qquad T_n^* = n^{1/2}(\hat{\mu}_n^* - \bar{X}_n)$$

$$g_n(x, y; \mu) = y - (1-b)(x + y + n^{1/2}\mu)\,\mathrm{I}(|x + y + n^{1/2}\mu| \le n^{1/4})$$

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

Example: Hodges' Estimator (p2)

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

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

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

$\mu \neq 0 :\ g(x, y; \mu) = g(0, y; \mu) = y$

$$\mathscr{L}\big(g(Z, W; \mu) \big| Z\big) = \mathscr{L}(W|Z) = \mathscr{L}(W) = N(0,1) \ ✔$$

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

$$\mathscr{L}\big(g(Z, W; 0) \big| Z\big) = \mathscr{L}\big(bW + (1-b)Z \big| Z\big) = N\big((1-b)Z, b^2\big) \ ✘$$

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)

$$S_n = n^{1/2}(\bar{X}_n - \mu) \qquad S_n^* = n^{1/2}(\bar{X}_n^* - \bar{X}_n)$$

$$S_n \xrightarrow{\text{d}} Z \sim N(0, I_p) \qquad \mathscr{L}(S_n^* | X_{1:n}) \xrightarrow{\text{Pr}} N(0, I_p)$$

$$T_n = n^{1/2}(\hat{\mu}_n - \mu) \qquad T_n^* = n^{1/2}(\hat{\mu}_n^* - \bar{X}_n)$$

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

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

Example: Lindley's Estimator (p3)

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

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

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

$$T_n \xrightarrow{\text{d}} Z \sim N(0, I_p)$$

$$\mathscr{L}\big(g(Z, W; \mu)\big| Z\big) = \mathscr{L}(W|Z) = \mathscr{L}(W) = N(0, I_p) \checkmark$$

$\mu = m(\mu)1_p$:

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

$$\mathscr{L}\big(g(Z, W; \mu)\big| Z\big) = \mathscr{L}\left(W - \frac{(p-3)\big[(Z+W) - m(Z+W)1_p\big]}{\|(Z+W) - m(Z+W)1_p\|^2} \middle| Z\right) \times$$

Example: The LASSO (Correlation Model)

$(X_1, Y_1), \ldots, (X_n, Y_n)$ i.i.d. $\qquad E\left[(Y_i - X_i^T\beta)X_i\right] = 0$

$$\hat{\beta}_n = \underset{b}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} (Y_i - X_i^T b)^2 + \lambda_n \sum_{j=1}^{p} \left| b^{(j)} \right| \right\}$$

Assume: $E\left(\|(X_i, Y_i)\|^4\right) < \infty$

$V := \operatorname{Var}\left[(Y_i - X_i^T\beta)X_i\right]$ positive definite

$C := E\left(X_i X_i^T\right)$ positive definite

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

Let: $\Sigma = C^{-1} V C^{-1}$

Bootstrapping Cases

$(X_1^*, Y_1^*), \ldots, (X_n^*, Y_n^*)$ 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)

$$C_n = n^{-1} \sum_{i=1}^{n} X_i X_i^T \qquad C_n^* = n^{-1} \sum_{i=1}^{n} X_i^* X_i^{*T}$$

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

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

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

$$C_n \xrightarrow{\text{Pr}} C \qquad C_n^* \xrightarrow{\text{Pr}} C$$

$$S_n \xrightarrow{\text{d}} Z \sim N(0, \Sigma) \qquad \mathscr{L}(S_n^* | X_{1:n}) \xrightarrow{\text{Pr}} N(0, \Sigma)$$

Example: LASSO (Correlation Model) (p3)

$$g_n(x, y, C; \beta) =$$

$$y + \underset{u}{\operatorname{argmin}} \Big\{ u^T C u + n^{-1/2} \lambda_n \sum_{j=1}^{p} \Big(|n^{1/2} \beta^{(j)} + (x + y + u)^{(j)}| - |n^{1/2} \beta^{(j)}| \Big) \Big\}$$

$$g(x, y, C; \beta) =$$

$$y + \underset{u}{\operatorname{argmin}} \Big\{ u^T C u + \lambda_0 \sum_{j=1}^{p} \Big(u^{(j)} \operatorname{sign}(\beta^{(j)}) + \big|(x + y + u)^{(j)}\big| \, \mathrm{I}\big(\beta^{(j)} = 0\big) \Big) \Big\}$$

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)

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

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

$\beta^{(j)} \neq 0$ for all $j = 1, \ldots, p$: $g(x, y, C; \beta) = g(0, y, C; \beta)$

$$\mathscr{L}\big(g(Z, W, C; \beta) \big| Z\big) = \mathscr{L}\big(g(0, W, C; \beta) \big| Z\big) = \mathscr{L}\big(g(0, W, C; \beta)\big) \checkmark$$

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

$\mathscr{L}\big(g(Z, W, C; \beta) \big| Z\big)$ varies with Z and is random ✘

A Way to Repair the Bootstrap

Why is the bootstrap inconsistent in these examples?

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

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, $i = 1, \dots, n$.
- For a given value of μ, 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:

- $\Theta = \uplus_{i \in \mathcal{I}} \Theta_i \subset \mathbb{R}^p$.
- $\mathscr{L}(S_n^* | X_{1:n}) \xrightarrow{P_\theta^n} Q_\theta$ *for all* $\theta \in \Theta$.
- $Q_\theta(A_\theta) = 1$ *for all* $\theta \in \Theta$.
- $\tilde{\theta}_n \xrightarrow{P_\theta^n} \theta$ *and* $P_\theta^n(\tilde{\theta}_n \in \Theta_i) \to 1$ *for all* $\theta \in \Theta_i$ *and* $i \in \mathcal{I}$.
- *For all* $i \in \mathcal{I}$ *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 $\theta \in \Theta$: $\mathscr{L}\big(g_n(S_n^*, \tilde{\theta}_n) \big| X_{1:n}\big) \xrightarrow{P_\theta^n} \mathscr{L}\big(g(Z, \theta)\big)$, $Z \sim Q_\theta$.

Corollary

If $\mathscr{L}(S_n) \xrightarrow{P_\theta^n} Q_\theta$ *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:

- $\Theta = \uplus_{i \in \mathcal{I}} \Theta_i \subset \mathbb{R}^p$.
- $\mathscr{L}(S_n^* | X_{1:n}) \xrightarrow{P_\theta^n} Q_\theta$ for all $\theta \in \Theta$.
- $Q_\theta(A_\theta) = 1$ for all $\theta \in \Theta$.
- $\tilde{\theta}_n \xrightarrow{P_\theta^n} \theta$ and $P_\theta^n(\tilde{\theta}_n \in \Theta_i) \to 1$ for all $\theta \in \Theta_i$ and $i \in \mathcal{I}$.
- For all $i \in \mathcal{I}$ 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 $\theta \in \Theta$: $\mathscr{L}\left(g_n(S_n^*, \tilde{\theta}_n) \middle| X_{1:n}\right) \xrightarrow{P_\theta^n} \mathscr{L}\left(g(Z, \theta)\right)$, $Z \sim Q_\theta$.

Corollary

If $\mathscr{L}(S_n) \xrightarrow{P_\theta^n} Q_\theta$ *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:

- $\Theta = \uplus_{i \in \mathcal{I}} \Theta_i \subset \mathbb{R}^p$.
- $\mathscr{L}(S_n^* | X_{1:n}) \xrightarrow{P_\theta^n} Q_\theta$ *for all* $\theta \in \Theta$.
- $Q_\theta(A_\theta) = 1$ *for all* $\theta \in \Theta$.
- $\tilde{\theta}_n \xrightarrow{P_\theta^n} \theta$ *and* $P_\theta^n(\tilde{\theta}_n \in \Theta_i) \to 1$ *for all* $\theta \in \Theta_i$ *and* $i \in \mathcal{I}$.
- *For all* $i \in \mathcal{I}$ *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 $\theta \in \Theta$: $\mathscr{L}\big(g_n(S_n^*, \tilde{\theta}_n) \big| X_{1:n}\big) \xrightarrow{P_\theta^n} \mathscr{L}\big(g(Z, \theta)\big)$, $Z \sim Q_\theta$.

Corollary

If $\mathscr{L}(S_n) \xrightarrow{P_\theta^n} Q_\theta$ *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

Assume

$$X_1, \dots, X_n \sim \textit{i.i.d.}\ (\mu, \Sigma)\ \textit{(finite)}$$

$$\tilde{\mu}_n \textit{ is } \sqrt{n}\textit{-consistent for } \mu$$

Let

$$X_1^*, \dots, X_n^* \sim \textit{i.i.d.}\ \hat{P}_{\tilde{\mu}_n} \quad \textit{conditional on } X_{1:n} = \{X_1, \dots, X_n\}$$

Then

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

Theorem

Assume

$$X_1, \ldots, X_n \sim \textit{i.i.d. } (\mu, \Sigma) \textit{ (finite)}$$

$$\tilde{\mu}_n \textit{ is } \sqrt{n}\textit{-consistent for } \mu$$

Let

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

Then

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

Examples Revisited

Example (Hodges Estimator)

- $\Theta_0 = \{0\}$, $\Theta_1 = (-\infty, 0) \cup (0, \infty)$.
- $g_n(x;\mu) = x - (1-b)(x + n^{1/2}\mu)\,\mathrm{I}(|x + n^{1/2}\mu| \le n^{1/4})$
- $g(x;\mu) = x - (1-b)x\,\mathrm{I}(\mu = 0) = \begin{cases} x & \text{if } \mu \neq 0 \\ bx & \text{if } \mu = 0 \end{cases}$
- $\tilde{\mu}_n = \bar{X}_n I(|\bar{X}_n| \ge a_n)$ with $a_n \to 0$ and $n^{1/2} a_n \to \infty$.

Examples Revisited

Example (Lindley's Estimator)

- $\Theta_0 = \{\mu \in \mathbb{R}^p : \mu = m(\mu)1_p\}$, $\Theta_1 = \{\mu \in \mathbb{R}^p : \mu \neq m(\mu)1_p\}$.
- $$g_n(x;\mu) = x - \frac{(p-3)\left[(x) - m(x)1_p + n^{1/2}\{\mu - m(\mu)1_p\}\right]}{\left\|(x) - m(x)1_p + n^{1/2}\{\mu - m(\mu)1_p\}\right\|^2}$$
- $$g(x;\mu) = \begin{cases} x, & \text{if } \mu \neq m(\mu)1_p, \\ x - \dfrac{(p-3)[(x) - m(x)1_p]}{\|(x) - m(x)1_p\|^2}, & \text{if } \mu = m(\mu)1_p. \end{cases}$$
- $$\tilde{\mu}_n = \begin{cases} \bar{X}_n & \text{if } \|\bar{X}_n - m(\bar{X}_n)1_p\| > a_n \\ m(\bar{X}_n)1_p & \text{if } \|\bar{X}_n - m(\bar{X}_n)1_p\| \leq a_n \end{cases}$$

  with $a_n \to 0$ and $n^{1/2}a_n \to \infty$.

Examples Revisited

Example (LASSO)

- $\Theta = \mathbb{R}^p = \biguplus_{J \in \mathcal{I}} \Theta_J$ with $\mathcal{I} = \big\{J : J \subset \{1, \dots, p\}\big\}$ and

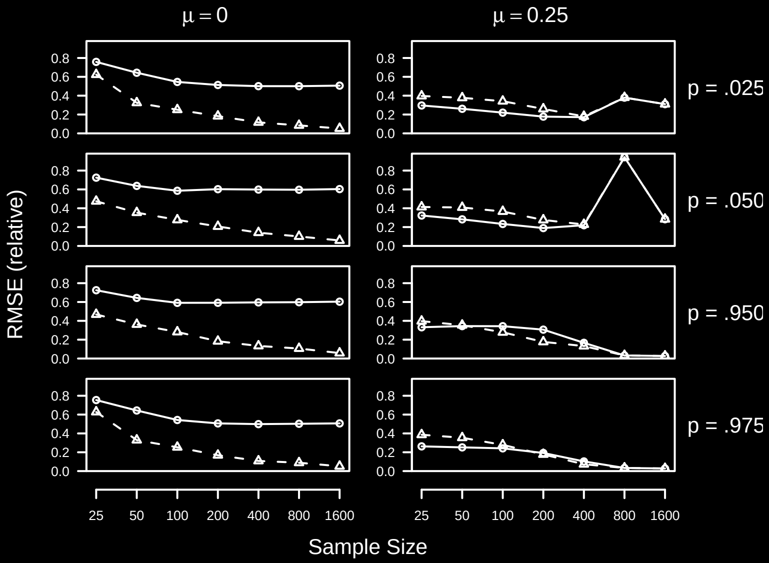
$$\Theta_J = \{\beta \in \mathbb{R}^p : \beta^{(j)} = 0 \text{ for all } j \in J,\ \beta^{(j)} \neq 0 \text{ for all } j \notin J\}$$

- $g_n(x, C; \beta) =$
$x + \operatorname{argmin}_u \Big\{ u^T C u + n^{-1/2} \lambda_n \sum_{j=1}^p \Big(|n^{1/2}\beta^{(j)} + (x+u)^{(j)}| - |n^{1/2}\beta^{(j)}| \Big) \Big\}$

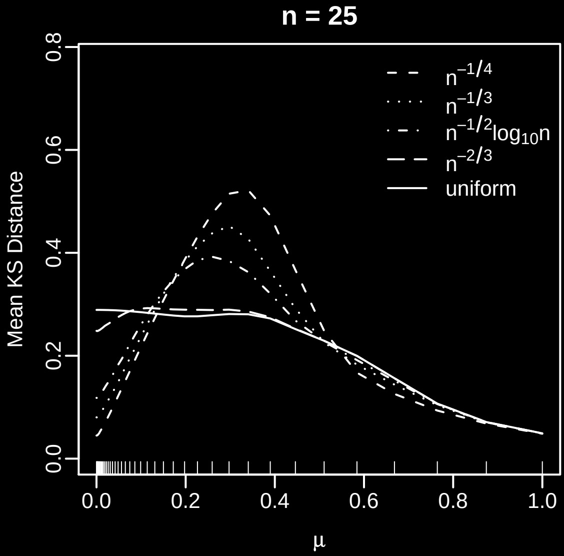
- $g(x, C; \beta) =$
$x + \operatorname{argmin}_u \Big\{ u^T C u + \lambda_0 \sum_{j=1}^p \Big(u^{(j)} \operatorname{sign}(\beta^{(j)}) + \big|(x+u)^{(j)}\big| \, \mathrm{I}(\beta^{(j)} = 0) \Big) \Big\}$

- $\tilde{\beta}_n =$ an estimator with the "oracle" property, e.g., SCAD or adaptive-LASSO (but not the LASSO).

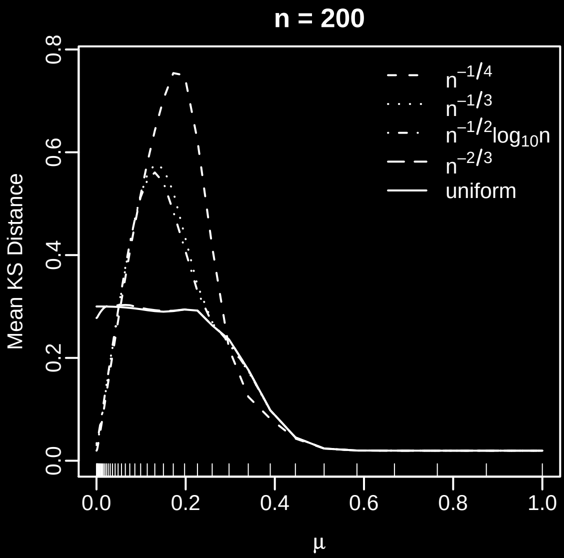
Simulations: Hodges' Estimator ($b = 0.5$), $a_n = n^{-1/2} \log_{10} n$



Simulations: Hodges Estimator ($b = 0.5$)



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 $Y'_n \xrightarrow{d} Y \sim Q$.
Let $h_n(y) := g_n(x_n, y)$ and $h(y) := g(x, y)$.

$$\left.\begin{array}{l} Y'_n \xrightarrow{d} Y \\ h_n(y_n) \to h(y) \\ \qquad \text{for all}^* \ y_n \to y \end{array}\right\} \xRightarrow{\text{H.Rubin}} h_n(Y'_n) \xrightarrow{d} h(Y) \quad \Updownarrow$$

$$g_n(x_n, Y'_n) \xrightarrow{d} g(x, Y)$$

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

$$\psi_n(x_n, Q_n) \to \psi(x, Q)$$

where

$$\begin{aligned} \psi_n(x, Q') &:= \mathscr{L}\big(g_n(x, Y')\big) \\ \psi(x, Q') &:= \mathscr{L}\big(g(x, Y')\big) \end{aligned} \quad \text{when } Y' \sim Q'$$

Return to Corollary 1

Proof of Corollary (p2)

$$\psi_n(x, Q') := \mathscr{L}\big(g_n(x, Y')\big) \qquad Y' \sim Q'$$
$$\psi(x, Q') := \mathscr{L}\big(g(x, Y')\big)$$

$$\psi_n(x_n, Q_n) \to \psi(x, Q) \qquad \text{for all}^* \ x_n \to x \text{ and } Q_n \to Q$$

Let $\mathbb{Q}_n = \mathscr{L}(Y_n | \mathscr{A}_n)$. Then

$$\left.\begin{array}{l} X_n \xrightarrow{\text{d}} X \\ \mathbb{Q}_n \xrightarrow{\text{Pr}} Q \\ \psi_n(x_n, Q_n) \to \psi(x, Q) \\ \qquad \text{for all}^* \ (x_n, Q_n) \to (x, Q) \end{array}\right\} \xRightarrow{\text{Proposition}} \psi_n(X_n, \mathbb{Q}_n) \xrightarrow{\text{d}} \psi(X, Q)$$

Return to Corollary 1

Proof of Corollary (p3)

$$\psi_n(x, Q') := \mathscr{L}\big(g_n(x, Y')\big) \qquad Y' \sim Q'$$
$$\psi(x, Q') := \mathscr{L}\big(g(x, Y')\big) \qquad Y' \sim Q'$$
$$\mathbb{Q}_n := \mathscr{L}(Y_n | \mathscr{A}_n)$$
$$\psi_n(X_n, \mathbb{Q}_n) \xrightarrow{\text{d}} \psi(X, Q) \qquad X \sim P$$

But

$$\left.\begin{aligned} X_n &\in \mathscr{A}_n \\ \mathbb{Q}_n &:= \mathscr{L}(Y_n | \mathscr{A}_n) \end{aligned}\right\} \implies \mathscr{L}\big(g_n(X_n, Y_n)\big|\mathscr{A}_n\big) = \psi_n(X_n, \mathbb{Q}_n)$$

and

$$\psi(X, Q) = \mathscr{L}\big(g(X, Y)\big|X\big), \quad X \sim P, \quad Y \sim Q, \quad X \perp\!\!\!\perp Y$$

so

$$\mathscr{L}\big(g_n(X_n, Y_n)\big|\mathscr{A}_n\big) \xrightarrow{\text{d}} \mathscr{L}\big(g(X, Y)\big|X\big), \quad X \sim P, \quad Y \sim Q, \quad X \perp\!\!\!\perp Y$$

□

Return to Corollary 1