Outline

1) Nonparametric Estimation
2) Plugin estimator
3) Bootstrap standard errors
4) Bootstrap bias estimator/correction
5) Bootstrap confidence intervals
6) Double bootstrap
Nonparametric Estimation

Setting Nonparametric iid sampling model

\[ X_1, \ldots, X_n \overset{iid}{\sim} P, \quad P \text{ unknown} \]

Want to do inference on some "parameter" \( \Theta(P) \)

\[ \begin{align*}
\text{Ex} & \quad \Theta(P) = \text{median}(P) \quad (X \in \mathbb{R}) \\
& \quad \Theta(P) = \lambda_{\max}(\text{Var}_P(X_i)) \quad (X \in \mathbb{R}^d) \\
& \quad \Theta(P) = \arg\min_{\theta \in \mathbb{R}^d} \mathbb{E}_P \left[ (Y_i - \theta'X_i)^2 \right] \quad (X_i, Y_i) \overset{iid}{\sim} P \\
& \quad \Theta(P) = \arg\min_{\theta \in \Omega} D_{KL}(P \parallel P_0) \quad \text{(best-fitting model)} \\
& \quad \quad \quad = \arg\max_{\theta} \mathbb{E}_P \left[ \ell(\theta; X_i) \right] \quad \text{(if misspec)}
\end{align*} \]

Recall the empirical dist. of \( X_1, \ldots, X_n \) is

\[ \hat{P}_n = \frac{1}{n} \sum \delta_{X_i} \quad (\hat{P}_n(A) = \#\{i : X_i \in A\}) \]

The plug-in estimator of \( \Theta(P) \) is \( \hat{\Theta} = \Theta(\hat{P}_n) \)

\[ \begin{align*}
& \text{a) Sample median} \\
& \text{b) } \lambda_{\max} \text{ (sample var)} \\
& \text{c) OLS estimator} \\
& \text{d) MLE for } \mathbb{P}_0 : \Theta \in \Theta^0
\end{align*} \]
Does plug-in estimator work? Depends
\[ \hat{P}_n \overset{p}{\to} P \] Dep. on what sense of convergence
\[ \hat{P}_n(A) \overset{p}{\to} P(A) \quad \text{for all } A \quad \checkmark \]

(IV)
\[ \sup_A |\hat{P}_n(A) - P(A)| \overset{p}{\to} 0 \quad \text{if } P \text{ cts } X \]

(\text{use } A_n = \{x_1, \ldots, x_n \})

\[ \sup_x |\hat{P}_n((-\infty, x]) - P((-\infty, x])| \overset{p}{\to} 0 \quad \text{for } X \in \mathbb{R} \quad \checkmark \]

Want \( \Theta(P) \) to be cts wrt some topology in which \( \hat{P}_n \overset{p}{\to} P \), then \( \Theta(\hat{P}_n) \overset{p}{\to} \Theta(P) \)

\[ \Theta(P) = \begin{cases} 
1 & \text{if } P_{x_1, x_2, \ldots} (x_1 = x_2) > 0 \\
0 & \text{otherwise} 
\end{cases} \]

If \( P \) cts then \( \Theta(P) = 0 \), but \( \Theta(\hat{P}_n) = 1 \ \forall n \)
Bootstrap standard errors

Suppose $\hat{\Theta}_n(X)$ is an estimator for $\Theta(P)$ (maybe plug-in, maybe not)

What is its standard error? Use plug-in:

$$\text{s.e.}(\hat{\Theta}_n) = \sqrt{\text{Var}_{\hat{P}_n}(\hat{\Theta}_n^*)} \quad \text{[use } \hat{\Theta}_n^* \text{ to indicate new sample } X_n^*, \text{ not } X]$$

$$\text{Var}_{\hat{P}_n}(\hat{\Theta}_n^*) = \text{Var}_{X_1^*, \ldots, X_n^* \sim \hat{P}_n}(\hat{\Theta}_n(X_1^*, \ldots, X_n^*))$$

How to compute? Monte Carlo:

For $b = 1, \ldots, B$:

- Sample $X_1^{*b}, \ldots, X_n^{*b} \sim \hat{P}_n$
- $\hat{\Theta}_n^{*b} = \hat{\Theta}(X_1^{*b}, \ldots, X_n^{*b})$
- $\overline{\Theta}^* = \frac{1}{B} \sum_{b=1}^B \hat{\Theta}_n^{*b}$
- $\text{s.e.}(\hat{\Theta}_n) = \sqrt{\frac{1}{B} \sum_{b=1}^B (\hat{\Theta}_n^{*b} - \overline{\Theta}^*)^2}$

Note this is a Monte Carlo numerical approx. to the idealized Bootstrap estimator, which we could compute by iterating through all $n^\pi$ possible $X_n^*(X_1^*, \ldots, X_n^*)$ vectors.
Bootstrap Bias Correction

\( \hat{\theta}_n \) some estimator. What is its bias?

\[ \text{Bias}_p(\hat{\theta}_n) = \mathbb{E}_p \left[ \hat{\theta}_n - \Theta(P) \right] \]

Idea: plug in \( \hat{P}_n \) for \( P \):

\[ \text{Bias}_{\hat{P}_n}(\hat{\theta}_n) = \mathbb{E}_{\hat{P}_n} \left[ \hat{\theta}_n - \Theta(\hat{P}_n) \right] \]

Monte Carlo:

For \( b = 1, \ldots, B \):

Sample \( X_1^b, \ldots, X_n^b \) iid \( \hat{P}_n \)

\( \hat{\theta}_n^b = \hat{\theta}(X_n^b) \)

\[ \bar{\theta}^* = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_n^b \]

\( \text{Bias}(\hat{\theta}_n) = \bar{\theta}^* - \Theta(\hat{P}_n) \)

We can use this to correct bias:

\( \hat{\theta}^{bc} = \hat{\theta}_n - \text{Bias}(\hat{\theta}_n) \)

Note: while \( \hat{\theta}_n - \text{Bias}(\hat{\theta}_n) \) is always better than \( \hat{\theta}_n \), \( \hat{\theta}_n - \text{Bias}(\hat{\theta}_n) \) may not be! Might be adding var.
Bispo

Real

World

Bootstrap World

Parameter

Data set

Estimator

Sampling dist

Sampling dist of estimator

\[ \Theta, E_p \hat{\Theta}, \Theta(\hat{\Theta}_n), E_p \hat{\Theta}^* \]

\[ \text{Bias}_p(\hat{\Theta}), \text{Bias}_{\hat{\Theta}_n}(\hat{\Theta}^*) \]

"Real World"

\[ P = \bigwedge \]

\[ \Theta(P) \]

\[ X_1, \ldots, X_n \overset{iid}{\sim} P \]

\[ \hat{\Theta}(X) \]

(hidden)

\[ \hat{P}_n(X) = \bigwedge \]

\[ \Theta(\hat{P}_n(X)) \]

\[ X^*_1, \ldots, X^* \overset{iid}{\sim} \hat{P}_n(X) \]

\[ \hat{\Theta}^* = \hat{\Theta}(X^*) \]

(generated at will)

\[ \hat{\Theta}^* \]

\[ \Theta(\hat{P}_n(X)) \]
Bootstrap Confidence Interval

How do we get a CI for $\theta(p)$?

Idea: What if we knew the distribution of $R_n(x, p) = \hat{\theta}_n(x) - \theta(p)$?

Define cdf $G_{n, p}(r) = P_p(\hat{\theta}(x) - \theta(p) \leq r)$

Lower $\alpha/2$ quantile $r_1 = G_{n, p}^{-1}(\alpha/2)$

Upper $\ldots$ $r_2 = G_{n, p}^{-1}(1 - \alpha/2)$

$$1 - \alpha = P_p( r_1 \leq \hat{\theta}_n - \theta \leq r_2 )$$

$$= P_p( \theta \in [\hat{\theta}_n - r_2, \hat{\theta}_n - r_1] )$$

Usually we don't know $G_{n, p}$ -- so bootstrap!

$$G_{n, \hat{p}_n}(r) = P_{\hat{p}_n}(\hat{\theta}(x^*) - \theta(\hat{p}_n) \leq r)$$

$G_{n, \hat{p}_n}(r)$ is a function only of $X$ (not of $P$)

Can use $C_{n, \alpha} = [\hat{\theta}_n - \hat{r}_2, \hat{\theta}_n - \hat{r}_1]$

with $\hat{r}_1 = G_{n, \hat{p}_n}^{-1}(\alpha/2)$, $\hat{r}_2 = G_{n, \hat{p}_n}^{-1}(1 - \alpha/2)$
**Bootstrap algo:**

For \( b = 1, \ldots, B \):

\[
X_{i}^{*b}, \ldots, X_{n}^{*b} \sim \text{iid } \tilde{P}_{n}^{a}
\]

\[R_n^{*b} = \hat{\Theta}(x^{*b}) - \Theta(\tilde{P}_n)\]

Return ecdf of \( R_n^{*b} \)

The quantity \( R_n(X, P) = \hat{\Theta}_n(X) - \Theta(P) \) is called a root (function of data + dist., used to make C.I.s)

Other examples:

\[
R_n(X, P) = \frac{\hat{\Theta}_n(X) - \Theta(P)}{\hat{\sigma}(X)}
\]

\[
R_n(X, P) = \frac{\hat{\Theta}_n(X)}{\Theta(P)}
\]

Want to choose \( R_n \) so its sampling dist. \( G_{n,p} \) changes slowly with \( P \) (so \( G_{n,\hat{P}_n} \approx G_{n,p} \))

Studentized root \( \frac{\hat{\Theta}_n - \Theta}{\hat{\sigma}} \) usually works better than \( \hat{\Theta}_n - \Theta \), then we get

\[
C_{n,\alpha} = \left[ \hat{\Theta}_n - \hat{\gamma}_2 \hat{\sigma}, \hat{\Theta}_n - \hat{\gamma}_1 \hat{\sigma} \right]
\]
Double Bootstrap

We might have theory that tells us, e.g.
\[ \sup_{a < b} \left| \hat{G}_{n, \hat{p}}([a, b]) - G_{n, p}([a, b]) \right| \overset{p}{\to} 0 \]
but still be worried about finite-sample coverage.

Let \( \gamma_{n, p}(\alpha) = P_p(C_{n, \alpha} \in \Theta(p)) \)
\[ \to 1 - \alpha \quad \text{if } C_{n, \alpha} \text{ has asy. coverage} \]
But in finite samples, might have
\[ \gamma_{n, p}(\alpha) < 1 - \alpha \]
e.g., "90% interval" has 87% coverage
\[ \gamma_{n, p}(0.1) = 0.87 < 0.9 \]
Solution? Double Bootstrap!

1. Estimate \( \gamma_{n, p}(\cdot) \) via plug-in \( \gamma_{n, \hat{p}}(\cdot) \)
2. Use \( C_{n, \hat{\alpha}}(X) \) where \( \hat{\gamma}(\hat{\alpha}) = 1 - \alpha \)
e.g., estimate "92% interval" has 90% coverage, \( \hat{\alpha} = .08 \)
Step 1 algo.

For $a = 1, \ldots, A$:

\[
X_{1}^{a}, \ldots, X_{n}^{a} \overset{iid}{\sim} \hat{P}_{n}^{a}
\]

\[
\hat{P}_{n}^{a} = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_{i}^{a}}
\]

For $b = 1, \ldots, B$:

\[
X_{1}^{a,a,b}, \ldots, X_{n}^{a,a,b} \overset{iid}{\sim} \hat{P}_{n}^{a}
\]

\[
R_{n}^{a,a,b} = \left( \hat{\theta}_{n}(X^{a,a,b}) - \theta(\hat{P}_{n}^{a}) \right) / \hat{\sigma}(X^{a,a,b})
\]

\[
\hat{G}_{n}^{a} = \text{ecdf}(R_{n}^{a,a,b})_{a,b}
\]

For $\alpha \in \text{grid}$:

\[
C_{n,\alpha}^{a} = \left[ \hat{\theta}_{n}^{a} - \hat{\sigma}_{n}^{a}, \hat{\theta}_{n}^{a} - \hat{\sigma}_{n}^{a} \right]
\]

For $\alpha \in \text{grid}$:

\[
\hat{\gamma}(\alpha) = \frac{1}{A} \sum_{a} \mathbb{1}\{C_{n,\alpha}^{a} \in \Theta(\hat{P}_{n})\}
\]

\[
\hat{\alpha} = \hat{\gamma}^{-1}(1-\alpha)
\]