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

$X_1, \ldots, X_n \overset{iid}{\sim} P, \quad P$ unknown

Want to do inference on some functional "parameter" $\theta(P)$

Ex a) $\theta(P) = \text{median}(P) \qquad (\mathcal{X} \subseteq \mathbb{R})$

b) $\theta(P) = \lambda_{max}(\text{Var}_P(X_i)) \qquad (\mathcal{X} \subseteq \mathbb{R}^d)$

c) $\theta(P) = \underset{\theta \in \mathbb{R}^d}{\text{argmin}} \; \mathbb{E}_P\left[(Y_i - \theta' X_i)^2\right]$

$(X_i, Y_i) \overset{iid}{\sim} P$

d) $\theta(P) = \underset{\theta \in \Theta}{\text{argmin}} \; D_{KL}(P \,\|\, P_\theta)$ (best-fitting model even if misspec.)

$= \underset{\theta}{\text{argmax}} \; \mathbb{E}_P\left[\ell_1(\theta; X_i)\right]$

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

$$\hat{P}_n = \frac{1}{n} \sum \delta_{X_i} \qquad \left(\hat{P}_n(A) = \frac{\#\{i : X_i \in A\}}{n}\right)$$

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

a) Sample median

b) λ_{max}(sample var)

c) OLS estimator

d) MLE for $\{P_\theta : \theta \in \Theta\}$

Does plug-in estimator work? Depends

$\hat{P}_n \xrightarrow{p} P$? Dep. on what sense of convergence

$\hat{P}_n(A) \xrightarrow{p} P(A)$ for all A ✓

(TV) $\sup_A |\hat{P}_n(A) - P(A)| \not\xrightarrow{p} 0$ if P cts ✗

(use $A_n = \{X_1, \dots, X_n\}$)

(Glivenko-Cantelli)

$\sup_x |\hat{P}_n((-\infty, x]) - P((-\infty, x])| \xrightarrow{p} 0$ for $\mathcal{X} \subseteq \mathbb{R}$ ✓

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

<u>Counterexamples</u>

$\theta(P) = 1\{P \text{ is absolutely cts}\}$ ($P \ll$ Lebesgue)

$\theta(P) = 1\{P \text{ is integrable}\}$ ($\mathbb{E}_P|X| < \infty$)

$\hat{P}_n$ always integrable, never abs. cts., for all 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:

$$\widehat{s.e.}(\hat{\theta}_n) = \sqrt{\mathrm{Var}_{\hat{P}_n}(\hat{\theta}_n^*)}$$

[use $\hat{\theta}_n^*$ to indicate new sample X^*, not X]

$$\mathrm{Var}_{\hat{P}_n}(\hat{\theta}_n^*) = \mathrm{Var}_{X_1^*, \ldots, X_n^* \overset{iid}{\sim} \hat{P}_n}\left(\hat{\theta}_n(X_1^*, \ldots, X_n^*)\right)$$

How to compute? Monte Carlo:

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

  Sample $X_1^{*b}, \ldots, X_n^{*b} \overset{iid}{\sim} \hat{P}_n$ ← Sample n points with replacement from original sample

  $\hat{\theta}^{*b} = \hat{\theta}(X_1^{*b}, \ldots, X_n^{*b})$

$$\bar{\theta}^* = \frac{1}{B}\sum_{b=1}^{B} \hat{\theta}^{*b}$$

$$\widehat{s.e.}(\hat{\theta}_n) = \sqrt{\frac{1}{B}\sum_b (\hat{\theta}^{*b} - \bar{\theta}^*)^2}$$

Note this is a Monte Carlo numerical approx. to the idealized Bootstrap estimator, which we could compute by iterating over all n^n possible $X^* = (X_1^*, \ldots, X_n^*)$ 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:

$$\text{Bias}_{\hat{P}_n}(\hat{\theta}_n^*) = \mathbb{E}_{\hat{P}_n}\left[\hat{\theta}_n^* - \underbrace{\theta(\hat{P}_n)}_{\text{NB}}\right]$$

Monte Carlo:

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

Sample $X_1^{*b}, \dots, X_n^{*b} \overset{\text{iid}}{\sim} \hat{P}_n$

$\hat{\theta}^{*b} = \hat{\theta}(X^{*b})$

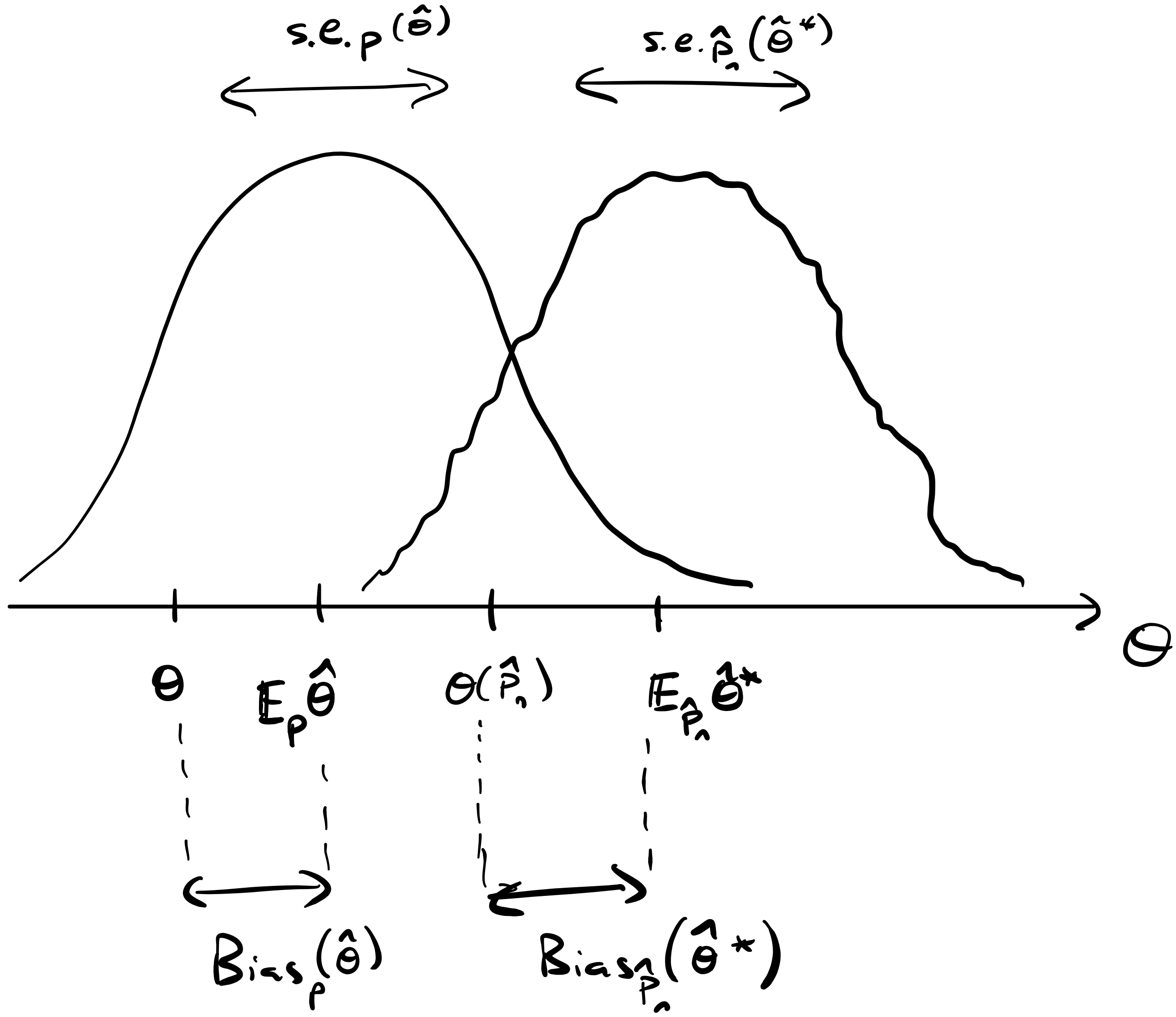
$$\overline{\theta}^* = \frac{1}{B}\sum_{b=1}^{B} \hat{\theta}^{*b}$$

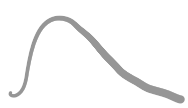
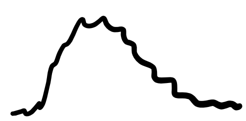
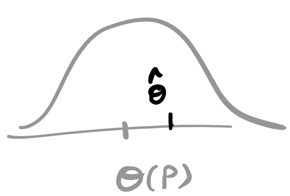
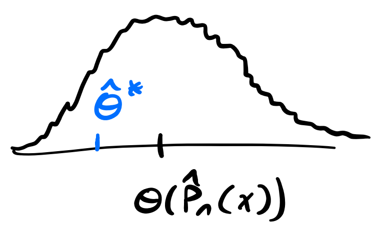
$$\widehat{\text{Bias}}(\hat{\theta}_n) = \overline{\theta}^* - \theta(\hat{P}_n)$$

We can use this to correct bias:

$$\hat{\theta}_n^{BC} = \hat{\theta}_n - \widehat{\text{Bias}}(\hat{\theta}_n)$$

<u>Note</u>: while $\hat{\theta}_n - \text{Bias}(\hat{\theta}_n)$ is always better than $\hat{\theta}_n$, $\hat{\theta}_n - \widehat{\text{Bias}}(\hat{\theta}_n)$ may not be! Might be adding var.



	"Real World"	"Bootstrap World"
Sampling dist.	P =  (hidden)	$\hat P_n(X)$ = 
Parameter	$\theta(P)$	$\theta(\hat P_n(X))$
Data set	$X_1,\dots,X_n \overset{iid}{\sim} P$	$X_1^*,\dots,X_n^* \overset{iid}{\sim} \hat P_n(X)$
Estimator	$\hat\theta(X)$ (observed once)	$\hat\theta^* = \hat\theta(X^*)$ (generated at will)
Sampling dist of estimator		

Bootstrap Confidence Interval

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

<u>Idea</u>: What if we knew the distribution of $R_n(X,P) = \hat{\theta}_n(X) - \theta(P)$?

Define cdf $G_{n,P}(r) = \mathbb{P}_P(\hat{\theta}(X) - \theta(P) \le r)$

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

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

$$1-\alpha = \mathbb{P}_P(r_1 \le \hat{\theta}_n - \theta \le r_2)$$

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

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

$$G_{n,\hat{P}_n}(r) = \mathbb{P}_{\hat{P}_n}(\hat{\theta}(X^*) - \theta(\hat{P}_n) \le r)$$

$G_{n,\hat{P}_n}(r)$ 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 $\hat{r}_1 = G_{n,\hat{P}_n}^{-1}(\alpha/2)$, $\hat{r}_2 = G_{n,\hat{P}_n}(1-\alpha/2)$

Bootstrap algo:

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

$X_1^{*b}, \ldots, X_n^{*b} \overset{iid}{\sim} \hat{P}_n$

$R_n^{*b} = \hat{\theta}(X^{*b}) - \theta(\hat{P}_n)$

Return ecdf of R_n^{*b}

The quantity $R_n(X, P) = \hat{\theta}_n(X) - \theta(P)$ is called a <u>root</u>

(function of data + dist., used to make CIs)

Other examples:

$$R_n(X, P) = \frac{\hat{\theta}_n(X) - \theta(P)}{\hat{\sigma}(X)}$$

[where $\hat{\sigma}(X)$ is some estimate of s.e.$(\hat{\theta}_n)$]

$$R_n(X, P) = \hat{\theta}_n(X) / \theta(P)$$

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

<u>Studentized</u> root $\frac{\hat{\theta}_n - \theta}{\hat{\sigma}}$ usually works better than $\hat{\theta}_n - \theta$, then we get

$$C_{n,\alpha} = \left[\hat{\theta}_n - \hat{r}_2 \hat{\sigma},\ \hat{\theta}_n - \hat{r}_1 \hat{\sigma}\right]$$

Double Bootstrap

We might have theory that tells us, e.g.

$$\sup_{a<b} \left| G_{n,\hat{P}_n}([a,b]) - G_{n,P}([a,b]) \right| \xrightarrow{P} 0$$

but still be worried about finite-sample coverage.

Let $\gamma_{n,P}(\alpha) = \mathbb{P}_P(C_{n,\alpha} \ni \theta(P))$

$\to 1-\alpha$ if $C_{n,\alpha}$ 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? <u>Double</u> Bootstrap!

1. Estimate $\gamma_{n,P}(\cdot)$ via plug-in $\gamma_{n,\hat{P}_n}(\cdot)$
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 $a = 1, \dots, A$:

$$X_1^{*a}, \dots, X_n^{*a} \overset{iid}{\sim} \hat{P}_n$$

$$\hat{P}_n^{*a} = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i^{*a}}$$

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

$$X_1^{**a,b}, \dots, X_n^{**a,b} \overset{iid}{\sim} \hat{P}_n^{*a}$$

$$R_n^{**a,b} = \left(\hat{\theta}_n(X^{**a,b}) - \theta(\hat{P}_n^{*a})\right) / \hat{\sigma}(X^{**a,b})$$

$$\hat{G}_n^{*a} = \mathrm{ecdf}\left(R_n^{**a,1}, \dots, R_n^{**a,B}\right)$$

For $\alpha \in$ grid:

$$C_{n,\alpha}^{*a} = \left[\hat{\theta}_n^{*a} - \hat{\sigma}^{*a} \cdot r_2(\hat{G}_n^{*a}),\ \hat{\theta}_n^{*a} - \hat{\sigma}^{*a} \cdot r_1(\hat{G}_n^{*a})\right]$$

For $\alpha \in$ grid:

$$\hat{\gamma}(\alpha) = \frac{1}{A} \sum_a 1\{C_{n,\alpha}^{*a} \ni \theta(\hat{P}_n)\}$$

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