Feynman Liang
STAT 210A
Lecture #25
11/17/2016

1 Plug-in estimator

Suppose we observe

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} P \quad X_i \in \mathbb{R}^d \tag{1.1}$$

Example 1.1. Say X_{ij} = score of student i on test in subject j.

Combine into $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and do PCA to estimate

$$\theta(p) = \underset{\|v\|=1}{\arg\max} \operatorname{Var}_p(v' X_i) \tag{1.2}$$

(assume unique).

One natural estimator is

$$\hat{\theta}(X_1, \ldots, X_n) = \underset{\|v\|=1}{\arg\max} \frac{1}{n} \sum_{i=1}^{n} \left[v'(X_i - \bar{X}) \right]^2 \tag{1.3}$$

where $\bar{X} = \frac{1}{n} \sum_i X_i$ is the sample average.

$$\hat{\theta} = \underset{\|v\|=1}{\arg\max} \operatorname{Var}_{\hat{p}_n}(v' X) \tag{1.4}$$

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \tag{1.5}$$

$$\text{For } A \subset \mathbb{R}^d : \hat{p}_n(A) = \frac{\#\{i \leq n : X_i \in A\}}{n} \overset{\text{a.s.}}{\to} P(A) \tag{1.6}$$

$\hat{p}$ is called the *empirical distribution*.

The *plug-in estimator* is

$$\hat{\theta} = \theta(\hat{p}_n) \tag{1.7}$$

Questions:

(a) What are bias & variance?

(b) Can we get a CI?

2 Bootstrap

We want $\text{Var}_p(\hat{\theta}_n)$.

Bootstrap estimates via plug-in

$$\hat{\text{Var}}(\hat{\theta}_n) = \underbrace{\text{Var}_{\hat{p}_n}(\hat{\theta}_n)}_{\text{can compute directly}} \tag{2.1}$$

$\text{Var}_{\hat{p}_n}(\hat{\theta}_n)$ integrates over possible samples

$$X_1^*, \ldots, X_n^* \overset{\text{iid}}{\sim} \hat{p}_n \tag{2.2}$$

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

1) Sample $X_1^{*b}, \ldots, X_n^{*b} \overset{\text{iid}}{\sim} \hat{p}_n$ with replacement
2) Compute $\hat{\theta}^{*b} = \hat{\theta}(x_1^{*b}, \ldots, x_n^{*b})$

$$\bar{\theta}^* = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^{*b} \tag{2.3}$$

$$\hat{\text{Var}} = \frac{1}{B} \sum_{b=1}^{B} (\hat{\theta}^{*b} - \bar{\theta}^*)^2 \underset{B \to \infty}{\overset{\text{a.s.}}{\to}} \text{Var}_{\hat{p}_n}(\hat{\theta}_n) \tag{2.4}$$

Notes:

- Not necc for $\hat{\theta}\theta(\hat{p}_n)$
- $\hat{p_n}$ need not be empirical distribution

2.1 Bias correction

$$\text{Bias}_p(\hat{\theta}) = \mathbb{E}_p[\hat{\theta}] - \theta(p) \tag{2.5}$$

$$\hat{\text{Bias}}(\hat{\theta}_n) = \text{Bias}_{\hat{p}_n}(\hat{\theta}) \tag{2.6}$$

$$= \mathbb{E}_{\hat{p}_n}(\hat{\theta}_n(X_1^*, \ldots, X_n^*)) - \theta(\hat{p}_n) \tag{2.7}$$

If we knew the true bias

$$\bar{\theta}_n = \hat{\theta}_n - \text{Bias}_p(\hat{\theta}_n) \tag{2.8}$$

Unfortunately, we don't know p so we don't know the bias, but we can estimate it with Boostrap

$$\check{\theta}_n = \hat{\theta}_n - \text{Bias}_{\hat{p}_n}(\hat{\theta}_n) \tag{2.9}$$

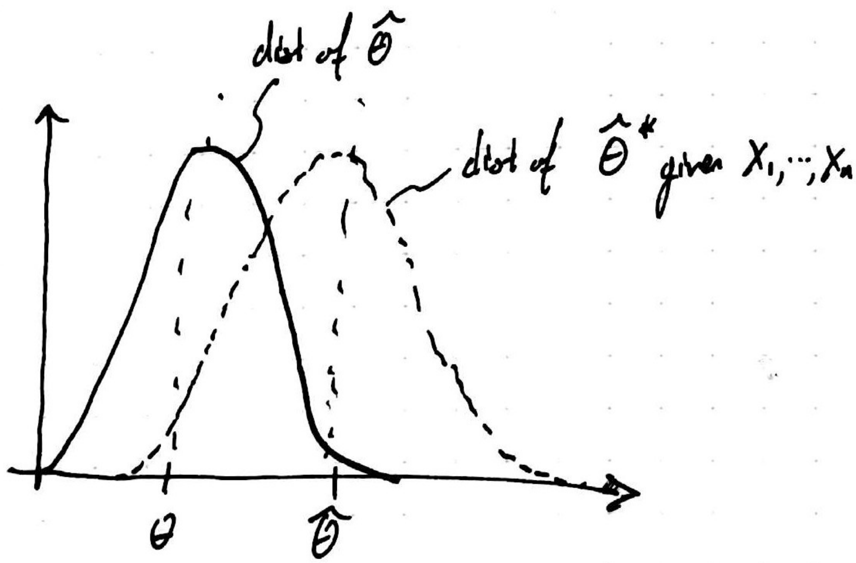

Figure 1: Typical picture showing bias in distribution of estimators

3 Bootstrap Confidence Regions

Definition 3.1. A *root* $R_n(X, \theta(p))$ depends on data X and parameter θ.

If we know the root's distribution

$$L_n(r; P) = \mathbb{P}_p[R_n(X, \theta(p)) \le r] \tag{3.1}$$

then we can make it into a confidence region $C(X; P)$

$$\begin{aligned} C_\alpha(X) &= \{\theta : L_n(R_n(X; \theta)) \le 1 - \alpha\} & (3.2) \\ &= \{\theta : R_n(X; \theta) \le L_n^{-1}(1 - \alpha)\} & (3.3) \end{aligned}$$

Then

$$\mathbb{P}_p(\theta(p) \in C_\alpha(X)) = \mathbb{P}_p\left(R_n(X, \theta(p)) \le L_n^{-1}(1 - \alpha; P)\right) \ge 1 - \alpha \tag{3.4}$$

Proposal: Use $L_n(\cdot, \hat{p}_n)$

Example 3.2. $R_n = |\hat{\theta}(X) - \theta(p)|$.
$\implies C(X) = \hat{\theta} = L_n^{-1}(1 - \alpha, \hat{p}_n)$

Example 3.3. $R_n = \frac{|\hat{\theta} - \theta|}{\hat{\sigma}(X)}$ ("studentized root")
$\implies C(X) = \hat{\theta} \pm \hat{\sigma} L_n^{-1}(1 - \alpha, \hat{p}_n)$

Example 3.4. $R_n = \|\hat{\theta} - \theta(p)\|_\infty$
$C = [\hat{\theta}_1 = L_n^{-1}(1 - \alpha)] \times \cdots \times [\hat{\theta}_d = L_n^{-1}(1 - \alpha)]$

Example 3.5. $\ell(\hat{\theta}(X)) - \ell(\theta(p))$ [ℓ any loss function]

Operationally (computing $L_n(r, \hat{p}_n))$), for $b = 1, \ldots, B$:
$X_1^{*b}, \ldots, X_n^{*b} \overset{\text{iid}}{\sim} \hat{p}_n$
$R_n^{*b} = R_n(X_1^{*b}, \ldots, X_n^{*b}; \theta(\hat{p}_n))$

Example 3.6. Wald

$$\hat{\theta} \pm \frac{1}{\sqrt{nJ_1(\hat{\theta})}} z_{\alpha/2} \tag{3.5}$$

R_n: $|\hat{\theta} - \theta| - \sqrt{nJ_1(\hat{\theta})} \Rightarrow |N(0,1)|$
[R_n is an *asymptotic pivot*] $L_n(\cdot; \hat{p}_n) \to L_n(\cdot; p)$