Lecture 23: Jackknife & Nonparametric Bootstrap

- Say we have iid sample $\underline{x} = (x_1, \dots, x_n)$ from an unknown prob. distribution F on some space $\mathcal{X}$

$$x_i \overset{iid}{\sim} F \quad \text{for } i = 1, 2, \dots, n.$$

- We can compute $\hat{\theta}$ applying some algorithm $s(\cdot)$ to $\underline{x}$ (i.e. $\hat{\theta} = s(\underline{x})$)

→ Our objective is to assign a standard error to $\hat{\theta}$.

Let $\underline{x}_{-i} = (x_1, x_2, \dots, \underbrace{x_{i-1}, x_{i+1}}_{x_i \text{ is removed}}, \dots x_n)^T$

and $\hat{\theta}_{-i} = s(\underline{x}_{-i})$

Definition: The Jackknife estimate of standard error is:

$$\hat{se}_{jack} = \left[\frac{n-1}{n} \sum_{i=1}^{n} (\hat{\theta}_{-i} - \hat{\theta}_{(\cdot)})^2 \right]^{1/2}$$

where $\hat{\theta}_{(\cdot)} = \sum_{i=1}^{n} \hat{\theta}_{-i} / n$

Example 1. Let $s(\cdot) = \bar{x}$. Then

$$\bar{x} = \sum_{i=1}^{n} x_i / n$$

$$\hat{\theta}_{-i} = \frac{n \cdot \bar{x} - x_i}{(n-1)}$$

$$\hat{\theta}_{(\cdot)} = \frac{n^2 \cdot \bar{x} - \sum_{i=1}^{n} x_i}{n(n-1)} = \frac{n^2 \cdot \bar{x} - n \cdot \bar{x}}{n(n-1)}$$

$$= \frac{\cancel{n} \bar{x} \cancel{(n-1)}}{\cancel{n(n-1)}} = \bar{x}.$$

$$\hat{\theta}_{-i} - \hat{\theta}_{(\cdot)} = \frac{n \cdot \bar{x} - x_i}{(n-1)} - \bar{x} = \frac{n \cdot \bar{x} - x_i - \bar{x} \cdot (n-1)}{(n-1)}$$

$$= \frac{\bar{x} - x_i}{(n-1)}.$$

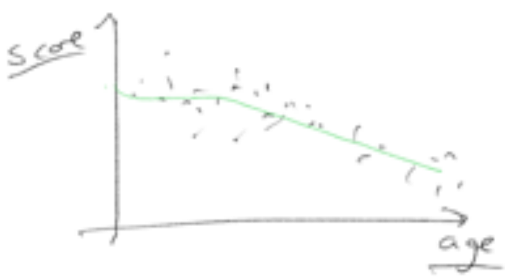
$$\hat{se}_{jack} = \left[\frac{\cancel{n-1}}{n} \cdot \sum_{i=1}^{n} \frac{(\bar{x} - x_i)^2}{(n-1)^{\cancel{2}}} \right]^{1/2}$$

$$= \left[\frac{1}{n(n-1)} \cdot \sum_{i=1}^{n} (\bar{x} - x_i)^2 \right]^{1/2}$$

$\Rightarrow$ This matches the classical formula for the standard error of the sample mean.

Example 2:

- we will use the kidney (or liver or whatever) data from chapter 1.



- we'll compute a Jackknife estimate of se. for the lowess regression estimator.

(i.e, the se of $\hat{y} = E(y \mid x)$)

Features of Jackknife estimator:

- Works for any distribution F.
- Need to be careful when definition of $s(\underline{x})$ changes with

n (think about the median)

- Mean weakness is its dependence on local derivatives. Unsmooth (choppy) statistics can result in erratic behavior of $\widehat{se}_{jack}$.

Nonparametric Bootstrap

- Bootstrap replaces unknown distribution F with an estimate $\hat{F} = \underline{x}$.

- Sample repeatedly to estimate $\widehat{se}_{boot}$ by computing the sample standard deviation of $\hat{\theta}$.

A bootstrap sample

$$\underline{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$$

where each x_i^* is randomly drawn with equal prob. & with replacement from $\{x_1, \ldots x_n\} = \underline{x}$.

Each bootstrap sample provides a bootstrap replication of $s(\cdot)$

$$\hat{\theta}^* = s(\underline{x}^*)$$

We repeat this process B times

$$\hat{\theta}^{*b} = s(\underline{x}^{*b}) \text{ for } b = 1, \ldots, B$$

Then, $$\widehat{se}_{boot} = \left[\frac{\sum_{b=1}^{B}(\hat{\theta}^{*b} - \hat{\theta}^{*\cdot})^2}{B-1}\right]^{1/2}$$

where

$$\hat{\theta}^{*\cdot} = \sum^{B} \hat{\theta}^{*b}$$

$$\overline{B}.$$

Note that $\hat{\theta}^*$ is obtained in two steps:

① $\underline{x}^{*b}$ is generated by iid sampling from $\hat{F}$

Empirical Probability Distribution (each point has prob $1/n$).

② $\hat{\theta}^{*b}$ is calculated from $\underline{x}^{*b}$.

Btw, it can be shown that
$\hat{F} = \{x_i \text{ w/prob } 1/n \text{ for all } i\}$
is the MLE estimate of F!

The true standard error of $\hat{\theta}$ is a function of F. Call this function $se(F)$.

$$se(F) = \left[\frac{\sum_{j=1}^{N} \left(\hat{\theta}^{(j)} - \hat{\theta}^{(\cdot)}\right)^2}{N-1} \right]^{1/2}$$

$$\text{where } \hat{\theta}^{(\cdot)} = \sum_{j=1}^{N} \hat{\theta}^{(j)} / N$$

$\hat{se}_{boot}$ is nothing but the plug-in estimate

$$\hat{se}_{boot} = se(\hat{F}).$$

- This is the one-sample nonparametric bootstrap. We'll cover other versions in the remaining lectures.

Properties of $\hat{se}_{boot}$

- "Shakes" the data much more than the Jackknife (doesn't depend on local derivatives).
- We can use bootstrap to estimate

- We ca...

  any other measure of variability (i.e. $E|\hat{\theta}-\theta|$).

- Much more computation than classical methods.

Example:

- 22 students took 5 tests
  - → we compute the sample correlation matrix & its eigenvalues λ_i's.

The eigenratio statistic $\theta = \dfrac{\max(\lambda_i)}{\sum \lambda_i}$

measures how closely the 5 scores can be described w/ a single #