Johns Hopkins Engineering

625.464 Computational Statistics

Basic Bootstrapping Methods

Module 9 Lecture 9B

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Nonparametric Bootstrap

$n \sim n^n$ possible pseudodatasets and so complete enumeration is not poss.

B ind, random bootstrap pseudodata (b.p.d.) from the empirical dist. $\hat{F}$ of the observed data.

ie. draw $\mathcal{X}_i^* = \{X_{i1}^*, X_{i2}^*, \dots, X_{in}^*\}$ for $i = 1, \dots, B$

and then use $R(\mathcal{X}_i^*, \hat{F})$ to approx the dist of $R(\mathcal{X}, F)$.

Nonparametric Bootstrap Example

$$\hat{F} = \begin{cases} 1 & 1/3 \\ 2 & 1/3 \\ 6 & 1/3 \end{cases}$$

Generate $\mathcal{X}_i^*$ by sampling $X_{i1}^*, X_{i2}^*, X_{i3}^*$ w/ replacement from $\{1, 2, 6\}$. Each $\mathcal{X}_i^*$ will yield an $\hat{\theta}^*$. $B = 1000$ to approx $P^*[\hat{\theta}^*]$

$\mathcal{X}^*$	$\hat{\theta}^*$	$P^*\left[\hat{\theta}^*\right]$	Observed Frequency
1 1 1	3/3	1/27	36/1000
1 1 2	4/3	3/27	101/1000
1 2 2	5/3	3/27	123/1000
2 2 2	6/3	1/27	25/1000
1 1 6	8/3	3/27	104/1000
1 2 6	9/3	6/27	227/1000
2 2 6	10/3	3/27	131/1000
1 6 6	13/3	3/27	111/1000
2 6 6	14/3	3/27	102/1000
6 6 6	18/3	1/27	40/1000

Parametric Bootstrap

Assumption

Believe that F is a parametric dist $F(X, \theta)$ with known form, but unknown parameter θ

Parametric Bootstrap

- Draw $X_1, \dots, X_n \sim$ iid $F(x, \theta)$
- use X to est $\theta \Rightarrow \hat{\theta}$
- Each parametric b.p.d. X^* is generated by drawing $x_1^*, \dots, x_n^* \sim$ iid $F(x, \hat{\theta})$

Parametric Bootstrap

Ex/ Assume $x_1, \ldots, x_n$ are Normal $(\theta, 1)$

- compute $\hat{\theta} = \sum_{i=1}^{n} x_i / n$
- draw $x_1^*, \ldots, x_n^*$ Normal $(\hat{\theta}, 1)$

Bootstrap Bias Correction

We want to know the bias of $T(F) = \theta$

$\therefore$ we are interested in using bootstrap analysis on

$$R(X, F) = T(\hat{F}) - T(F) = \hat{\theta} - \theta$$

We want to know

$$E[\text{Bias}] = E[\hat{\theta} - \theta] = E[\hat{\theta}] - \theta \quad \text{w.r.t. } F$$

using the bootstrap principle we calculate

$$E^*[\hat{\theta}^* - \hat{\theta}] = E^*[\hat{\theta}^*] - \hat{\theta} = \bar{\theta}^* - \hat{\theta}$$

w.r.t. $\hat{F}$

$$\boxed{\theta_{Bias} = \hat{\theta} - [\bar{\theta}^* - \hat{\theta}] = 2\hat{\theta} - \bar{\theta}^*}$$

where $\bar{\theta}^* = \sum_{j=1}^{B} \frac{\hat{\theta}_j^*}{B}$

drawn from $\hat{F}$

Bootstrap Estimation of Variance

Suppose we wish to estimate the variance of $T(\hat{F}) = \hat{\theta}$

- generate B $\hat{\theta}_j^*$ from $\hat{F}$

$$\hat{V}(\hat{\theta}) \approx \tilde{V}(\hat{\theta}^*) = \frac{1}{B-1} \sum_{j=1}^{B} \left(\hat{\theta}_j^* - \bar{\theta}^*\right)^2$$