Probability and Statistic (DeGroot 4th)

Chapter 12.6 The Bootstrap

Nonparametric bootstrap

$\eta(\mathbb{X}, \mathbb{F}) \to \eta(\mathbb{X}^*, \hat{\mathbb{F}})$, with mostly $\hat{F}$ is F_n, sample c.d.f

Parametric bootstrap

$\eta(\mathbb{X}, \mathbb{F}) \to \eta(\mathbb{X}^*, \hat{\mathbb{F}})$, with mostly using parameters of $\hat{F}$ as M.L.E

Exercises 1.

$X_1, X_2, ..., X_n \sim \theta e^{-\theta x} \text{ for } x \geq 0$

Use parametric bootstrap to estimate the variance of the sample average $Var(\bar{X})$ (No simulation is required)

We have $Var(\bar{X}) = \sigma^2/n$.

In exponential distribution, $\sigma^2 = \frac{1}{\lambda^2}$, we use M.L.E $\hat{\lambda}$ to replace λ.

Furthermore, $\hat{\lambda} = 1/\bar{X}$, So $Var(\bar{X}) = \bar{X}^2/n$.

Exercise 2.

$\mathbb{X}^* = (X_1^*, ..., X_n^*)$ is an i.i.d. sample from distribution F_n because we sample with replacement so

$X_i^* \text{ and } X_j^* \text{ will be independent and have same distribution } F_n$

Exercise 3

3. Let n be odd, and let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a sample of size n from some distribution. Suppose that we wish to use the nonparametric bootstrap to estimate some feature of the sample median. Compute the probability that the sample median of a nonparametric bootstrap sample will be the smallest observation from the original data $\boldsymbol{X}$.

Assume that in $\mathbb{X}$, we have l number equal the smallest one. So, the above event happens if and only if we sample at least $\frac{n+1}{2}$ smallest number, and we have probability to collect the smallest one is $\frac{l}{n}$.

Probability is :

$\sum_{i=(n+1)/2}^{n} C_n^i (l/n)^n (1-l/n)^{n-i}$

Exercise 4

Table 11.5 Boiling point of water in degrees Fahrenheit and atmospheric pressure in inches of mercury from Forbes' experiments. These data are taken from Weisberg (1985, p. 3).

Boiling Point	Pressure
194.5	20.79
194.3	20.79
197.9	22.40
198.4	22.67
199.4	23.15
199.9	23.35
200.9	23.89
201.1	23.99
201.4	24.02
201.3	24.01
203.6	25.14
204.6	26.57
209.5	28.49
208.6	27.76
210.7	29.04
211.9	29.88
212.2	30.06

4. Use the data in the first column of Table 11.5 on page 699. These data give the boiling points of water at 17 different locations from Forbes' experiment. Let F be the distribution from which these boiling points were drawn. We might not be willing to make many assumptions about F. Suppose that we are interested in the bias of the sample median as an estimator of the median of the distribution F. Use the nonparametric bootstrap to estimate this bias. First, do a pilot run to compute the simulation standard error of the simulation approximation, and then see how many bootstrap samples you need in order for your bias estimate (for distribution $\hat{F}$) to be within 0.02 of the true bias (for distribution $\hat{F}$) with probability at least 0.9.

$\eta(\mathbb{X}, \mathbb{F}) = E[\hat{\theta} - \theta]$ (1) , with $\hat{\theta}, \theta$ is sample median and median of distribution F

- It is hard to compute $\eta(\mathbb{X}, \mathbb{F})$, so we use $\eta(\mathbb{X}^*, \hat{\mathbb{F}})$. And we replace θ with $\hat{\theta} = 201.4$, from data in the first column of table.
- Then, we perform simulating a large number v simulation having size of n from $\mathbb{F}^*$, p.f uniform with 17 value $212.2, 211.9, 210.7, 208.6, ...$
- In each bootstrap sample i for $i = 1, ..., v$, we compute sample medium θ^{i*}, then we can have bias for this bootstrap sample $T^i = \theta^{i*} - \hat{\theta}$
- Then, $\frac{1}{v}\sum T^i$ is bootstrap estimator for bias
- To compute how many bootstrap samples in order for our bias estimate to be within 0.02 of true bias with probability at least 0.9. we use this equation:

$$v = \left[\Phi^{-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma}{\epsilon}\right]^2. \qquad (12.2.5)$$

- First, we based on v bootstrap samples to compute sample variance $\hat{\sigma}^2 = \frac{1}{n}\sum(T_i - \bar{T})^2$ ~~**no that is wrong, we should compute simulation standard error instead of sample stardard error**~~, $\epsilon = 0.2, \gamma = 0.9$. Finally, we replace $\mathbb{F}^*$ for c.d.f Φ , then we can easily have v.

Exercise 6.

6. Use the data in Exercise 16 of Sec. 10.7.

a. Use the nonparametric bootstrap to estimate the variance of the sample median.

b. How many bootstrap samples does it appear that you need in order to estimate the variance to within .005 with probability 0.95?

- $\eta(\mathbb{X}, \mathbb{F}) = E[(\hat{\theta} - \theta)^2]$ where θ is median and $\hat{\theta}$ is sample median.
- We use $\eta(\mathbb{X}^*, \hat{\mathbb{F}}) = E[(\theta^{i*} - \hat{\theta})^2]$, with θ^{i*} is sample median of bootstrap sample i. $\hat{F}$ is sample c.d.f

Exercise 7

Solution for this exercise:

So, the parametric bootstrap can proceed as follows: First, choose a large number v, and for $i = 1, \ldots, v$, simulate $(\overline{X}^{*(i)}, \overline{Y}^{*(i)}, S_X^{2*(i)}, S_Y^{2*(i)})$ where all four random variables are independent with the following distributions:

- $\overline{X}^{*(i)}$ has the normal distribution with mean 0 and variance $\hat{\sigma}_1^2/m$.
- $\overline{Y}^{*(i)}$ has the normal distribution with mean 0 and variance $\hat{\sigma}_2^2/n$.
- $S_X^{2*(i)}$ is $\hat{\sigma}_1^2$ times a random variable having the χ^2 distribution with $m-1$ degrees of freedom.
- $S_Y^{2*(i)}$ is $\hat{\sigma}_2^2$ times a random variable having the χ^2 distribution with $n-1$ degrees of freedom.

Then compute

$$U^{(i)} = \frac{(m+n-2)^{1/2}(\overline{X}^{*(i)} - \overline{Y}^{*(i)})}{\left(\frac{1}{m} + \frac{1}{n}\right)^{1/2} \left(S_X^{2*(i)} + S_Y^{2*(i)}\right)^{1/2}}$$

for each i. Our simulation approximation to the bootstrap estimate of the probability of type I error for the usual two-sample t test would be the proportion of simulations in which $|U^{(i)}| > c$.

Exercise 9.

c. To compute simulation standard errors of the bootstrap estimate of the variance of the sample correlation (in a.), we base on example 12.2.10 to compute simulation standard error of sample variance.

Exercis 10.

a. Similar to Example 12.6.4
b. Similar to Example 12.6.5

Exercise 11.

- $\eta(\mathbb{X}^*, \hat{\mathbb{F}}) = E[\theta^* - \hat{\theta}], \hat{F} = c.d.f F_n$