Homework 3

Deadline: 2024.5.24 (금) 23:59

Nonparametric Statistics 2024 Spring

Note: Use R to do computation. Provide/attach your R code at the end of your homework. Submit your answers rounded up to three decimal places.

Question 1. Bootstrap Confidence Interval for the Mean

The following 21 observations are drawn from the population distribution F which is unknown.

```
X = c(8.16, 8.47, 8.63, 9.11, 9.25, 9.45, 9.47, 9.49, 9.91, 9.99, 10.00,
      10.18, 10.33, 10.50, 10.99, 11.46, 11.52, 11.55, 11.73, 12.20, 12.20)
```

Answer the following questions using the $B = 2,000$ bootstrap samples.

(a) Construct the 95% Normality-based bootstrap confidence interval (CI) for the mean. Use the bias-correction version.

(b) Use the residual method to construct the 95% bootstrap CI for the mean.

(c) Use the BCA method to construct the 95% bootstrap CI for the mean.

(d) Use the t-pivot method to construct the 95% bootstrap CI for the mean.

Question 2. Bootstrap CI for the median.

Now, you are going to use the same dataset in Question 1, but make inference about the median, not the mean. Use $B = 2,000$.

(a) Use the normal approximation, residual and BCA methods to construct the 95% bootstrap CI for the median.

(b) Now, we want to use the t-pivot method,

$$\frac{\hat{\theta}_{med} - \theta}{\hat{se}(\hat{\theta}_{med})} \sim H.$$

We know that the distribution of the sample median can be approximated by a normal distribution ($H \approx N(0,1)$). However, $se(\hat{\theta}_{med})$ is a function of the density f and the variance σ^2. Suppose the population distribution is normally distributed, $F = N(\mu, \sigma^2)$. When $n = 2m+1$, the sample median is approximately normally distributed as,

$$\hat{\theta}_{med} \sim N(\mu, \frac{\pi\sigma^2}{4m}).$$

Based on the above approximation result, use the t-pivot method to construct the 95% bootstrap CI.

(c) Suppose we don't have any information on F. In this case, we need to use the second-level bootstrap samples to estimate $\hat{se}(\hat{\theta}^{(b)}_{med})$. Set the number of the second-level bootstrap samples as $B_{second} = 100$. Use the t-pivot method, construct the 95% bootstrap CI for the median.

Question 3. The validity of the bootstrap procedure

Let $\{X_1, \dots, X_n\}$ be a sample from the (unknown) distribution F with mean μ and variance $\sigma^2 < \infty$. Suppose we are interested in $\theta = |\mu|$. Define $\hat{\theta}_n = |\bar{X}_n|$. We want to decide whether the bootstrap procedure fails or not when $\mu = 0$. Consider the following facts:

- $\sqrt{n}(\bar{X}_n - \mu) \overset{D}{\to} N(0, \sigma^2)$ by the central limit theorem (CLT).
- $\sqrt{n}(\bar{X}^{(b)}_n - \bar{X}_n) \overset{D}{\to} N(0, \sigma^2)$.

(a) Define $T_n = \sqrt{n}(\hat{\theta}_n - \theta) = \sqrt{n}(|\bar{X}_n| - |\mu|)$. Show that when $\mu = 0$, $T_n \overset{D}{\approx} |Z|$, $Z \sim N(0, \sigma^2)$ meaning that the sampling distribution of T_n is asymptotically given by a truncated normal distribution.

(b) Now, we consider $T^*_n = \sqrt{n}(|\bar{X}^{(b)}_n| - |\bar{X}_n|)$, where $\bar{X}^{(b)}_n$ is computed from the bootstrap sample $\{X^{(b)}_1, \dots, X^{(b)}_n\}$. Show that when $\mu = 0$, $T^*_n \overset{D}{\approx} |\sqrt{n}Z^* + \sqrt{n}\bar{X}_n| - |\sqrt{n}\bar{X}_n|$ where $Z^* \sim N(0, \sigma^2)$.

(c) Suppose $n = 20, \sigma^2 = 1$ and $\bar{X}_n \in \{0.3, 0.03, 0.003, 0.0003\}$. Using the Monte Carlo (MC) approximation, compare the distribution T_n to the distribution T^*_n for various values of $\bar{X}_n$. Discuss whether you can say $T_n \overset{D}{\approx} T^*_n$ or equivalently whether the bootstrap is valid. Justify your answer.