Homework 3

刘慧航 SA18017026

2019 年 9 月 30 日

Problem 1 考虑 *bootstrap* 包里的 *scor* 考试成绩数据. 此数据是 *88* 名考生 *5* 门课程的考试成绩:*mechanics, vectors, algebra, analysis, statistics.* 前两门为闭卷考试, 其余三门为开卷考试. 在 *scor* 数据下, 记协方差矩阵为 Σ, 其特征根为 $\lambda_1 > \cdots > \lambda_5 > 0$. 则

$$\theta = \frac{\lambda_1}{\sum_{i=1}^{5} \lambda_i}$$

表示了主成分中第一主成分对方差的解释比例. 现在记 $\theta = \frac{\lambda_1}{\sum_{i=1}^{5} \lambda_i}$ 为样本协方差矩阵 $\hat{\Sigma}$ 的特征根.

（*1*）. 分别使用 *Bootstrap* 方法和 *Jackknife* 方法估计 θ 的估计

$$\hat{\theta} = \frac{\hat{\lambda}_1}{\sum_{i=1}^{5} \hat{\lambda}_i}$$

的偏差和标准差。

（*2*）. 计算在 *(1)* 中的 *Bootstrap* 重复样本下 θ 的 95% 百分位数置信区间和 *BCa* 置信区间.

Solution

(1) 使用 R 中的 boostrap 包, 及如下代码计算 $\hat{\theta}$ 的偏差和标准误差.

```
n <- nrow(scor)
cov.hat <- cov(scor)
eigenvalue.hat <- eigen(cov.hat)$values
theta.hat <- eigenvalue.hat[1]/sum(eigenvalue.hat)
```

```
B <- 50
theta.b <- numeric(B)
for (i in 1:B) {
    i <- sample(1:n, size = n, replace = TRUE)
    cov.b <- cov(scor[i,])
    eigenvalue.b <- eigen(cov.b)$values
    theta.b[i] <- eigenvalue.b[1]/sum(eigenvalue.b)
}
bias.b <- mean(theta.b-theta.hat)
se.b <- sd(theta.b)
print(bias.b)
print(se.b)
```

得到 $bias()\hat{\theta})$ 以及 $se(\hat{\theta})$ 分别为 0.05944855 和 0.04929686.

使用 Jackknife 函数来估计 $\hat{\theta}$ 的偏差和标准误差, 代码如下

```
theta.jack <- numeric(n)
for(i in 1:n){
  cov.jack <- cov(scor[-i,])
  eigenvalue.jack <- eigen(cov.jack)$values
  theta.jack[i] <- eigenvalue.jack[1]/sum(eigenvalue.jack)
}
bias.jack <- (n-1)*(mean(theta.jack)-theta.hat)
print(bias.jack)
se.jack <- sqrt((n-1) *mean((theta.jack - mean(theta.jack))^2))
print(se.jack)
```

得到 $bias()\hat{\theta})$ 以及 $se(\hat{\theta})$ 分别为 0.001069139 和 0.04955231.

(2) 使用 bootstrap 和 boot 包, 运行如下代码

```
data(scor,package="bootstrap")
theta.boot<-function(scor,i){
```

```
    eigenvalue<-eigen(cov(scor[i,],scor[i,]))$values
    eigenvalue[1]/sum(eigenvalue)
}
boot.obj<-boot(scor, theta.boot,R=2000)
print(boot.ci(boot.obj,type = c("perc","bca")))
```

得到 θ 的 95% 百分位数置信区间和 BCa 置信区间分别为 $(0.5219, 0.7068)$, $(0.5219, 0.7068)$.

Problem 2 设 $X_1, \ldots, X_n$ 表示不同的样本值, $X_1^*, \ldots, X_n^*$ 为一个 *Bootstrap* 抽样, 令 $X_1^*, \ldots, X_n^*$, 求期望 $E(\bar{X}^*|X_1, \ldots, X_n)$, $Var(\bar{X}^*|X_1, \ldots, X_n)$ 以及 $E(\bar{X}^*)$ 和 $Var(\bar{X}^*)$.

Solution

$$E\left[\overline{X}^*|X_1, \cdots, X_n\right] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^*|X_1, \cdots, X_n\right] = E\left[X_1^*|X_1, \cdots, X_n\right] = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\begin{aligned}
E\left[\left(\overline{X}^*\right)^2|X_1, \cdots, X_n\right] &= \frac{1}{n^2}E\left[\left(\sum_{i=1}^{n} X_i^\star\right)^2|X_1, \cdots, X_n\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} E\left[(X_i^\star)^2|X_1, \cdots, X_n\right] \\
&= \frac{1}{n}E\left[(X_1^*)^2|X_1, \cdots, X_n\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} X_i^2
\end{aligned}$$

$$\begin{aligned}
\operatorname{var}\left(\overline{X}^\star|X_1, \cdots, X_n\right) &= E\left[\left(\overline{X}^\star\right)^2|X_1, \cdots, X_n\right] - \left(E\left[\overline{X}^\star|X_1, \cdots, X_n\right]\right)^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2
\end{aligned}$$

再由双期望公式得

$$E\left[\overline{X}^*\right] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = EX$$

$$\begin{aligned}
\operatorname{var}\left(\overline{X}^*\right) &= E\left[\operatorname{var}\left(\overline{X}^* | X_1, \cdots, X_n\right)\right] + \operatorname{var}\left(E\left[\overline{X}^* | X_1, \cdots, X_n\right]\right) \\
&= E\left[\frac{1}{n^2}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] + \operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} E\left[X_i^2\right] - \frac{1}{n^2} E\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] + \frac{1}{n^2}\operatorname{var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \operatorname{var}(X)/n
\end{aligned}$$