EXERCISE 3

WEIYU LI

1. Consider the *scor* dataset in the package *bootstrap*, which is an 88×5 matrix. Denote its covariance matrix as Σ, and its eigenvalues $\lambda_1 > \ldots > \lambda_5 > 0$. Then

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

represents the explained proportion of variance by the first principal component. Denote $\hat{\lambda}_1 > \ldots > \hat{\lambda}_5$ the eigenvalues of the sample covariance matrix $\hat{\Sigma}$.

(1) **Use Bootstrap and Jackknife to estimate the bias and standard error of the estimation of θ, that is,**

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

Here's an example of the R code:

```
set.seed(0)
install.packages('boot')
library(boot)

lambda_hat <- eigen(cov(scor))$values
theta_hat <- lambda_hat[1] / sum(lambda_hat)
B <- 200 # number of bootstrap samples
n <- nrow(scor) # number of rows (data size)

# Bootstrap
func <- function(dat, index){
# input: dat, data; index, a vector of the bootstrap index
# output: theta, the estimated theta using the bootstrap sample
  x <- dat[index,]
  lambda <- eigen(cov(x))$values
  theta <- lambda[1] / sum(lambda)
  return(theta)
}
bootstrap_result <- boot(
  data = cbind(scor$mec, scor$vec, scor$alg, scor$ana, scor$sta),
  statistic = func, R = B)
```

Date: 2019/09/30.
liweiyu@mail.ustc.edu.cn.

```
theta_b <- bootstrap_result$t
bias_boot <- mean(theta_b) - theta_hat
# the estimated bias of theta_hat, using bootstrap
se_boot <- sqrt(var(theta_b))
# the estimated standard error (se) of theta_hat, using bootstrap

# Jackknife
theta_j <- rep(0, n)
for (i in 1:n) {
  x <- scor [-i,]
  lambda <- eigen(cov(x))$values
  theta_j[i] <- lambda[1] / sum(lambda)
  # the i-th entry of theta_j is the i-th "leave-one-out" estimation of theta
}
bias_jack <- (n - 1) * (mean(theta_j) - theta_hat)
# the estimated bias of theta_hat, using jackknife
se_jack <- (n - 1) * sqrt(var(theta_j) / n)
# the estimated se of theta_hat, using jackknife

# print the answers
bias_boot
se_boot
bias_jack
se_jack
```

This example gives the result:

Method	Bias	SE
Bootstrap	0.005884122	0.0469498
Jackknife	0.001069139	0.04955231

□

(2) **Compute the 95% percentile confidence interval and BCa confidence interval of θ, using the Bootstrap samples in (1).** The R code is: (continued)

```
boot.ci(bootstrap_result, conf = 0.95, type = c('perc', 'bca'))
```

whose output is

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 200 bootstrap replicates

CALL :
boot.ci(boot.out = bootstrap_result, conf = 0.95, type = c("perc",
    "bca"))

Intervals :
Level     Percentile            BCa
95%   ( 0.5352,  0.7016 )   ( 0.5256,  0.7004 )
```

```
Calculations and Intervals on Original Scale
Some percentile intervals may be unstable
Some BCa intervals may be unstable
```

□

2. Let $X_1, \dots, X_n$ be different sample values, $X_1^*, \dots, X_n^*$ is a Bootstrap sampling and let $\bar{X}^* = \frac{1}{n}\sum_i X_i^*$. Compute $E\left(\bar{X}^*|X_1, \dots, X_n\right)$, $Var\left(\bar{X}^*|X_1, \dots, X_n\right)$, $E\left(\bar{X}^*\right)$ and $Var\left(\bar{X}^*\right)$.

Solve. Notice that $X_1^*, \dots, X_n^* \overset{i.i.d.}{\sim} Unif\{X_1, \dots, X_n\}$, then

$$E(X_1^*|X_1, \dots, X_n) = \frac{1}{n}\sum_i X_i := \bar{X},$$
$$E\left(X_1^{*2}|X_1, \dots, X_n\right) = \frac{1}{n}\sum_i X_i^2 := \overline{X^2}.$$

Thus,

$$E\left(\bar{X}^*|X_1, \dots, X_n\right) = \frac{1}{n}\sum_i E\left(X_i^*|X_1, \dots, X_n\right) = \bar{X}, \tag{1}$$
$$Var\left(\bar{X}^*|X_1, \dots, X_n\right) = \frac{1}{n}\left[\overline{X^2} - (\bar{X})^2\right], \tag{2}$$

and taking expectations on $X_1, \dots, X_n$ gives that

$$E\left(\bar{X}^*\right) = E\left[E\left(\bar{X}^*|X_1, \dots, X_n\right)\right] = E\left(\bar{X}\right) = EX, \tag{3}$$
$$\begin{aligned} Var\left(\bar{X}^*\right) =& Var\left[E\left(\bar{X}^*|X_1, \dots, X_n\right)\right] + E\left[Var\left(\bar{X}^*|X_1, \dots, X_n\right)\right] \\ =& \frac{1}{n}Var(X) + \frac{1}{n}\frac{n-1}{n}Var(X) = \frac{2n-1}{n^2}Var(X), \end{aligned} \tag{4}$$

where $X_1, \dots, X_n \sim X,\ i.i.d.$. □