DATA130004: Homework 8

董晴园 14300680173

2017.12.16

1. Exercise 7.1: Compute a jackknife estimate of the bias and the standard error of the correlation statistic in Example 7.2.

```
## Jackknife Estimate of Bias
> data(law, package = "bootstrap")
> n <- nrow(law)
> y <- law$LSAT
> z <- law$GPA
> theta.hat <- cor(y,z)
> print (theta.hat)
[1] 0.7763745

# compute the jackknife replicates, leave-one-out estimates
> theta.jack <- numeric(n)
> for (i in 1:n)
+  theta.jack[i] <- cor(y[-i], z[-i])
> bias <- (n - 1) * (mean(theta.jack) - theta.hat)
> print(bias)  #jackknife estimate of bias
[1] -0.006473623

## Jackknife Estimate of s.e.
> se <- sqrt((n-1) *
+              mean((theta.jack - mean(theta.jack))^2))
> print(se)
[1] 0.1425186
```

2. Exercise 7.3: Obtain a bootstrap t confidence interval estimate for the correlation statistic in Example 7.2 (`law` data in `bootstrap`).

```
boot.t.ci <-
  function(x, B = 500, R = 100, level = .95, statistic){
    #compute the bootstrap t CI
    x <- as.matrix(x);  n <- nrow(x)
    stat <- numeric(B); se <- numeric(B)
    boot.se <- function(x, R, f) {
      #local function to compute the bootstrap
      #estimate of standard error for statistic f(x)
      x <- as.matrix(x); m <- nrow(x)
      th <- replicate(R, expr = {
        i <- sample(1:m, size = m, replace = TRUE)
        f(x[i, ])
      })
      return(sd(th))
    }
    for (b in 1:B) {
```

```r
      j <- sample(1:n, size = n, replace = TRUE)
      y <- x[j, ]
      stat[b] <- statistic(y)
      se[b] <- boot.se(y, R = R, f = statistic)
    }
    stat0 <- statistic(x)
    t.stats <- (stat - stat0) / se
    se0 <- sd(stat)
    alpha <- 1 - level
    Qt <- quantile(t.stats, c(alpha/2, 1-alpha/2), type = 1)
    names(Qt) <- rev(names(Qt))
    CI <- rev(stat0 - Qt * se0)
  }

# bootstrap t C.I. for law data correlation
data(law, package = "bootstrap")
dat <- cbind(law$LSAT, law$GPA)
stat <- function(dat) {
  cor(dat[, 1], dat[, 2]) }
ci <- boot.t.ci(dat, statistic = stat, B=2000, R=200)
print(ci)
      2.5%      97.5%
-0.1845940  0.9808244
```

So, the bootstrap t confidence interval is $[-0.1846, 0.9808]$.

3. Consider a p-dimensional normal distribution $X = (Y, Z)^\top$ with two partitions $Y \in \mathbb{R}^q, Z \in \mathbb{R}^{p-q}, 0 < q < p$. Correspondingly, the mean of X is $\mu = (\mu_Y, \mu_Z)^\top$ and the covariance of X is

$$\Sigma = \begin{pmatrix} \Sigma_{YY} & \Sigma_{YZ} \\ \Sigma_{ZY} & \Sigma_{ZZ} \end{pmatrix}.$$

   (a) Now derive the conditional distribution of Z given Y.
   Hint: make a non-singular transformation AX where

$$A = \begin{pmatrix} I_q & 0 \\ -\Sigma_{ZY}\Sigma_{YY}^{-1} & I_{p-q} \end{pmatrix}.$$

   First compute AX:

$$AX = \begin{pmatrix} I_q & 0 \\ -\Sigma_{ZY}\Sigma_{YY}^{-1} & I_{p-q} \end{pmatrix} \begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} Y \\ -\Sigma_{ZY}\Sigma_{YY}^{-1}Y + Z \end{pmatrix}$$

   We can prove that Y and $-\Sigma_{ZY}\Sigma_{YY}^{-1}Y + Z$ are uncorrelated, and because they are jointly normally distributed, they are independent simultaneously:

$$\begin{aligned}
Cov(Y, -\Sigma_{ZY}\Sigma_{YY}^{-1}Y + Z) &= E\big[Y(-\Sigma_{ZY}\Sigma_{YY}^{-1}Y + Z)\big] - E(Y)E(-\Sigma_{ZY}\Sigma_{YY}^{-1}Y + Z) \\
&= -\Sigma_{ZY}\Sigma_{YY}^{-1}E(Y^2) + E(YZ) + -\Sigma_{ZY}\Sigma_{YY}^{-1}\big[E(Y)\big]^2 - E(Y)E(Z) \\
&= -\Sigma_{ZY}\Sigma_{YY}^{-1}\big[E(Y^2) - E^2(Y)\big] + E(YZ) - E(Y)E(Z) \\
&= -\Sigma_{ZY}\Sigma_{YY}^{-1}\Sigma_{YY} + \Sigma_{ZY} \\
&= 0
\end{aligned}$$

   Since they are independent and jointly normally distributed, $(-\Sigma_{ZY}\Sigma_{YY}^{-1}Y + Z) \sim N(\mu^*, \Sigma^*)$, where $\mu^* = \mu_Z - \Sigma_{ZY}\Sigma_{YY}^{-1}\mu_Y$, $\Sigma^* = \Sigma_{ZZ} - \Sigma_{ZY}\Sigma_{YY}^{-1}\Sigma_{YZ}$. And because $J(A^{-1}) = J(A) = |A| = 1$,

$$f(X) = \frac{f(AX)}{|A|} = f(Y)f(-\Sigma_{ZY}\Sigma_{YY}^{-1}Y + Z).$$

$$\therefore \quad f(Z|Y) = \frac{f(Y)F(-\Sigma_{ZY}\Sigma_{YY}^{-1}Y + Z)}{f(Y)} = f(-\Sigma_{ZY}\Sigma_{YY}^{-1}Y + Z)$$
$$= \frac{1}{\sqrt{(2\pi)^{p-q}|\Sigma^*|}} \exp\left(-\frac{1}{2}(\mathbf{z} - \mu^*)^\top \Sigma^{*-1}(\mathbf{z} - \mu^*)\right),$$

where $\mu^* = \mu_Z - \Sigma_{ZY}\Sigma_{YY}^{-1}\mu_Y$, $\Sigma^* = \Sigma_{ZZ} - \Sigma_{ZY}\Sigma_{YY}^{-1}\Sigma_{YZ}$. So $Z|Y$ follows multivariate normal distribution as well.

(b) Restate the result when assuming $q = 1$, i.e., $Z|Y$ is a random variable conditioning with a $p - 1$ dimensional random vector. This result is useful in next lecture.

The result will be quite similar. If given $Y = y$, $Z|Y$ still follows multivariate normal distribution $N(\mu^*, \Sigma^*)$, where $\mu^* = \mu_Z + \frac{y-\mu_Y}{\sigma_Y^2}\Sigma_{ZY}$, $\Sigma^* = \Sigma_{ZZ} - \frac{1}{\sigma_Y^2}\Sigma_{ZY}\Sigma_{YZ}$.

$$f(Z|Y) = \frac{1}{\sqrt{(2\pi)^{p-1}|\Sigma^*|}} \exp\left(-\frac{1}{2}(\mathbf{z} - \mu^*)^\top \Sigma^{*-1}(\mathbf{z} - \mu^*)\right)$$