Advanced Statistical Methods HW8

2021-21116 Taeyoung Chang

11/22/2021

Exercise 10.4

4. Verify formula (10.38) for the number of distinct bootstrap samples.

ure 10.3). This rapidly becomes impractical. The number of distinct bootstrap samples for n points turns out to be

$$\binom{2n-1}{n}. \tag{10.38}$$

Given n data points, we want to figure out number of distinct tuples

$$(x_1, \cdots, x_n)$$

satisfying $x_1 + \cdots + x_n = n$ and $x_i \in \{0, 1, \cdots, n\}$ for each $i = 1, \cdots, n$
It is a problem equivalent to the following : for n equally shaped balls , we are going to separate those balls into n ordered groups , with group of zero ball being allowed .
The cardinality of the first group will be equal to x_1 and the cardinality of the second group will be equal to x_2 and so on.
Hence, it is a problem of locating separation bars between those n number of balls. Since we need to separate those balls into n number of ordered groups, we need $n-1$ separation bars. For example, if $n = 8$ and balls are represented by $\circ$ and separation bars are represented by $|$ then

$$\circ \mid \circ \mid \circ\ \circ \mid \mid \mid \circ \mid \circ \mid \circ\ \circ$$

the above example represents the case of

$$(x_1, x_2, \cdots, x_n) = (1, 1, 2, 0, 0, 1, 1, 2)$$

Hence , the number of distinct bootstrap samples for n points can be calculated by

$$\binom{n + (n-1)}{n} = \binom{2n-1}{n}$$

Exercise 10.5

5. A normal theory least squares model (7.28)–(7.30) yields $\hat{\beta}$ (7.32). Describe the parametric bootstrap estimates for the standard errors of the components of $\hat{\beta}$.

Linear regression, perhaps the most widely used estimation technique, is based on a version of $\hat{\boldsymbol{\mu}}^{\text{MLE}}$. In the usual notation, we observe an n-dimensional vector $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ from the linear model

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{\epsilon}. \tag{7.28}$$

Here $\boldsymbol{X}$ is a known $n \times p$ *structure matrix*, β is an unknown p-dimensional parameter vector, while the *noise vector* $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)'$ has its components uncorrelated and with constant variance σ^2,

$$\boldsymbol{\epsilon} \sim (\mathbf{0}, \sigma^2 \boldsymbol{I}), \tag{7.29}$$

where $\boldsymbol{I}$ is the $n \times n$ identity matrix. Often $\boldsymbol{\epsilon}$ is assumed to be multivariate normal,

$$\boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}), \tag{7.30}$$

From the model above given as

$$\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I})$$

and the ordinary least squares estimates of β is derived as

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Also, the estimates for variance is calculated by MSE

$$\hat{\sigma}^2 = \frac{1}{n-p}(\mathbf{y} - \mathbf{X}\hat{\beta})^T(\mathbf{y} - \mathbf{X}\hat{\beta})$$

For the parametric boostrap, we generate bootstrap sample $\mathbf{y}_1^*, \cdots, \mathbf{y}_B^*$ from

$$\mathbf{y}^* \sim N(\mathbf{X}\hat{\beta}, \hat{\sigma}^2\mathbf{I})$$

Then, the corresponding bootstrap samples for $\hat{\beta}$ is computed by

$$\hat{\beta}^{*(1)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}_1^*, \cdots, \ \hat{\beta}^{*(B)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}_B^*$$

From this bootstrap sample $\hat{\beta}^*$'s , we can derive parametric bootstrap estimate for the standard errors of the components of $\hat{\beta}$ as the following :

$$\hat{\text{se}}_{boot}(\hat{\beta}_1) = \text{sd}\big(\hat{\beta}_1^{*(1)}, \cdots, \hat{\beta}_1^{*(B)}\big) \ , \ \cdots \ , \ \hat{\text{se}}_{boot}(\hat{\beta}_p) = \text{sd}\big(\hat{\beta}_p^{*(1)}, \cdots, \hat{\beta}_p^{*(B)}\big)$$

where $\text{sd}(x_1, \cdots, x_n) = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2}$

Exercise 10.7

7. Verify formula (10.70).

> For the sample mean $\bar{x}$, the jackknife yields exactly the usual variance estimate (1.2), $\sum_i (x_i - \bar{x})^2/(n(n-1))$, while the ideal bootstrap estimate $(B \to \infty)$ gives
> $$\sum_{i=1}^{n}(x_i - \bar{x})^2/n^2. \tag{10.70}$$

First, we shall show that for $\hat{\theta} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$,

$$\hat{\mathrm{var}}_{jack}(\hat{\theta}) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n(n-1)}$$

Note that

$$\hat{\mathrm{var}}_{jack}(\hat{\theta}) = \frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2$$

For $\hat{\theta} = \overline{x}$,

$$\begin{aligned}
\hat{\theta}_{(i)} &= \frac{n\overline{x} - x_i}{n-1} \\
\hat{\theta}_{(\cdot)} &= \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(i)} = \frac{1}{n}\sum_{i=1}^{n}\frac{n\overline{x} - x_i}{n-1} = \frac{1}{n}\frac{n^2\overline{x} - n\overline{x}}{n-1} = \overline{x} \\
\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)} &= \frac{n\overline{x} - x_i}{n-1} - \overline{x} = \frac{\overline{x} - x_i}{n-1} \\
\hat{\mathrm{var}}_{jack}(\hat{\theta}) &= \frac{n-1}{n}\sum_{i=1}^{n}\frac{(\overline{x} - x_i)^2}{(n-1)^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n(n-1)}
\end{aligned}$$

Next, we shall prove that for the original estimate $\hat{\theta} = S(P_0)$ and resampling estimate $\hat{\theta}^* = S(P)$ with resampling vector P, there is a linear approximation $S_L(P)$ of $S(P)$ such that $S(P_{(i)}) = S_L(P_{(i)})$. Here, P is by definition a vector of nonnegative weights summing to 1 and $P_{(i)}$ is a resampling vector corresponding to i-th jackknife value $\hat{\theta}_{(i)}$, which is given by $\frac{1}{n-1}(1, 1, \cdots, 1, 0, 1, \cdots, 1)$. Also, P_0 is defined as $\frac{1}{n}(1, 1, \cdots, 1)$ We shall define $S_L(P)$ by the following :

$$\begin{aligned}
a &= \hat{\theta}_{(\cdot)} \\
\mathbf{b} &= (n-1)(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}) \\
S_L(P) = a - \mathbf{b}^T P &= \hat{\theta}_{(\cdot)} - (n-1)\sum_{i=1}^{n} p_i(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}) \\
&= n\hat{\theta}_{(\cdot)} - (n-1)\sum_{i=1}^{n} p_i\hat{\theta}_{(i)}
\end{aligned}$$

Observe that $S_L(P)$ is linear in P . Now we shall check that $S(P_{(i)}) = S_L(P_{(i)})$ for each $i = 1, \cdots, n$

$$S_L(P_{(i)}) = n\hat{\theta}_{(\cdot)} - (n-1)\sum_{j\neq i}\frac{1}{n-1}\hat{\theta}_{(j)} = \sum_{j=1}^{n}\hat{\theta}_{(j)} - \sum_{j\neq i}\hat{\theta}_{(j)} = \hat{\theta}_{(i)} = S(P_{(i)})$$

Now, we should consider what ideal bootstrap estimate is. For booootstrap estimate $\hat{\theta}^* = S(P^*)$, ideal bootstrap estimate of variance of $\hat{\theta}$ can be written as $Var(S(P^*))$ where P^* denotes the bootstrap resampling

vector with the form $\frac{1}{n}(N_1, \cdots, N_n)$ and $nP^* \sim Multinomial(n, P_0)$
Here, by the property of multinomial distribution, we have

$$Var(nP^*) = n\left[\text{Diag}(P_0) - P_0P_0^T\right]$$

Now, we shall claim that

$$Var(S_L(P^*)) = \frac{n-1}{n}\hat{\text{var}}_{jack}(\hat{\theta})$$

In above, we've shown that $S_L(P^*)$ is a linear approximation of $S(P^*)$ given as $S_L(P^*) = a - \mathbf{b}^TP^*$. Observe that randomness only lies in P^* for $a - \mathbf{b}^TP^*$. Thus ,

$$\begin{aligned}
Var(S_L(P^*)) = Var(a - \mathbf{b}^TP^*) = Var(\mathbf{b}^TP^*) = \mathbf{b}^TVar(P^*)\mathbf{b} &= \frac{1}{n^2}\mathbf{b}^TVar(nP^*)\mathbf{b} \\
&= \frac{1}{n^2}\mathbf{b}^Tn\left[\text{Diag}(P_0) - P_0P_0^T\right]\mathbf{b} \\
&= \frac{1}{n}\{\mathbf{b}^T\text{Diag}(P_0)\mathbf{b} - (P_0^T\mathbf{b})^2\} \\
&= \frac{1}{n}\mathbf{b}^T\text{Diag}(P_0)\mathbf{b} \quad \because \ P_0^T\mathbf{b} = \frac{n-1}{n}\sum_{i=1}^n(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}) = 0 \\
&= \frac{(n-1)^2}{n^2}\sum_{i=1}^n(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2 = \frac{n-1}{n}\hat{\text{var}}_{jack}(\hat{\theta})
\end{aligned}$$

Hence, we've shown our claim above.
Note that $\hat{\theta} = \overline{x}$ can be written as $\hat{\theta} = S(P_0)$ with $S(P) = \sum_{i=1}^n p_ix_i = \mathbf{x}^TP$ so that $S(P)$ is linear in P. Furthermore, in this case $S_L(P)$ agrees with $S(P)$.

$$S_L(P) = a - \mathbf{b}^TP = \hat{\theta}_{(\cdot)} - (n-1)\sum_{i=1}^n p_i(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}) = \overline{x} - (n-1)\sum_{i=1}^n p_i\frac{\overline{x} - x_i}{n-1} = \overline{x} - \sum_{i=1}^n p_i(\overline{x} - x_i) = \sum_{i=1}^n p_ix_i = S(P)$$

Hence, combining all the results above, the ideal bootstrap variance estimate of the sample mean $\overline{x}$ is given as

$$Var(S(P^*)) = \frac{n-1}{n}\cdot\frac{\sum_{i=1}^n(x_i - \overline{x})^2}{n(n-1)} = \frac{1}{n^2}\sum_{i=1}^n(x_i - \overline{x})^2$$

Exercise 10.9

9. A survey in a small town showed incomes $x_1, x_2, \ldots, x_m$ for men and $y_1, y_2, \ldots, y_n$ for women. As an estimate of the differences,

$$\hat{\theta} = \text{median}\{x_1, x_2, \ldots, x_m\} - \text{median}\{y_1, y_2, \ldots, y_n\}$$

was computed.

(a) How would you use nonparametric bootstrapping to assess the accuracy of $\hat{\theta}$?

(b) Do you think your method makes full use of the bootstrap replications?

(a) First, by resampling m samples among $x_1, \cdots, x_m$, we get bootstrap median $med(X)^*$. By repeating this B times , we have $med(X)^{*1}, \cdots, med(X)^{*B}$. Similarly , by resampling n samples among $y_1, \cdots, y_n$, we get bootstrap median $med(Y)^*$ and by repeating this B times, we have $med(Y)^{*1}, \cdots, med(Y)^{*B}$. With large enough B, say $B = 200$, we have B number of θ^* values derived by

$$\hat{\theta}^{*1} = med(X)^{*1} - med(Y)^{*1}\ ,\ \cdots\ ,\ \hat{\theta}^{*B} = med(X)^{*B} - med(Y)^{*B}$$

Then we have nonparametric bootstrap estimates for standard error of $\hat{\theta}$ as

$$\hat{\text{se}}_{boot}(\hat{\theta}) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}^{*b} - \hat{\theta}^{*\cdot}\right)^2} \quad \text{where} \quad \hat{\theta}^{*\cdot} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*b}$$

(b) To make full use of the bootstrap replications, we can consider a parametric bootstrap. Since the survey has implemented in a small town, it is expected that m and n are not so big. Hence, it may be useful to take advantage of the power of parametric inference. For example, assume parametric model $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$. Then, derive MLE of $\hat{\mu}_x$, $\hat{\mu}_y$, $\hat{\sigma}_x^2$, $\hat{\sigma}_y^2$ using observed data $x_1, \cdots, x_m, y_1, \cdots, y_n$. Now, generate bootstrap sample from $X^* \sim N(\hat{\mu}_x, \hat{\sigma}_x^2)$ and $Y^* \sim N(\hat{\mu}_y, \hat{\sigma}_y^2)$. The rest of the procedure is same for the nonparametric bootstrapping above.

Exercise 11.1

1. We observe $y \sim \lambda G_{10}$ to be $y = 20$. Here λ is an unknown parameter while G_{10} represents a gamma random variable with 10 degrees of freedom ($y \sim G(10, \lambda)$ in the notation of Table 5.1). Apply the Neyman construction as in Figure 11.1 to find the confidence limit endpoints $\hat{\lambda}(0.025)$ and $\hat{\lambda}(0.975)$.

Our model is $Y|\lambda \sim \Gamma(10, \lambda)$ and the observed value is $y = 20$. We will define some necessary notations.

$f_\lambda(y) = \frac{1}{\Gamma(10)\lambda^{10}} y^9 \exp(-y/\lambda)$: pdf of $y|\lambda$

$q_\alpha(f_\lambda) = \alpha$-quantile of y for given λ *i.e.* $P_\lambda(Y \le q_\alpha(f_\lambda)) = \alpha$

For given $\alpha \in (0,1)$, $q_\alpha(f_\lambda)$ is a function of λ

$I_\lambda(y) = I\{q_{\frac{\alpha}{2}}(f_\lambda)) \le y \le q_{1-\frac{\alpha}{2}}(f_\lambda))\}$

$C(y) = \{\lambda : I_\lambda(y) = 1\}$

Coverage probability : $P_\lambda(\lambda \in C(Y)) = P_\lambda(I_\lambda(Y) = 1) = P_\lambda(q_{\frac{\alpha}{2}}(f_\lambda)) \le Y \le q_{1-\frac{\alpha}{2}}(f_\lambda))) = 1 - \alpha$

Now, we shall claim that for given $\alpha \in (0,1)$, $q_\alpha(f_\lambda)$ is an increasing function of λ.
It is enough to show that $Y|\lambda_2$ is stochastically larger than $Y|\lambda_1$ whenever $\lambda_1 \le \lambda_2$ i.e. $P_{\lambda_1}(Y \ge r) \le P_{\lambda_2}(Y \ge r)$ for any $r > 0$ provided $\lambda_1 \le \lambda_2$ Take arbitrary $0 < \lambda_1 < \lambda_2$. Write $\lambda = \lambda_1$. Then there is a constant $c > 1$ such that $\lambda_2 = c\lambda$. Take $r > 0$. Then the following holds true.

$$P_{\lambda_1}(Y \ge r) = P_\lambda(Y \ge r) = \frac{1}{\Gamma(10)\lambda^{10}}\int_r^\infty y^9 \exp(-y/\lambda)\,dy$$
$$P_{\lambda_2}(Y \ge r) = P_{c\lambda}(Y \ge r) = \frac{1}{\Gamma(10)c^{10}\lambda^{10}}\int_r^\infty y^9 \exp(-y/c\lambda)\,dy$$
$$= \frac{1}{\Gamma(10)c^{10}\lambda^{10}}\int_{r/c}^\infty c^9 z^9 \exp(-z/\lambda)\,c\,dz \quad \because z = \frac{y}{c}\ ,\ dz = \frac{1}{c}dy$$
$$= \frac{1}{\Gamma(10)\lambda^{10}}\int_{r/c}^\infty z^9 \exp(-z/\lambda)\,dz = P_\lambda\big(Y \ge \frac{r}{c}\big)$$
$$\Rightarrow\ P_{\lambda_2}(Y \ge r) = P_{c\lambda}(Y \ge r) = P_\lambda\big(Y \ge \frac{r}{c}\big) \ge P_\lambda(Y \ge r) = P_{\lambda_1}(Y \ge r)$$
$$\therefore P_{\lambda_1}(Y \ge r) \le P_{\lambda_2}(Y \ge r)$$

We've shown that $Y|\lambda_2$ is stochastically larger than $Y|\lambda_1$ whenever $\lambda_1 \le \lambda_2$
Hence if we denote the cdf of $y|\lambda$ as F_λ , then

$$1 - F_{\lambda_1}(r) \le 1 - F_{\lambda_2}(r)\ ,\ F_{\lambda_1}(r) \ge F_{\lambda_2}(r)\ ,\ F_{\lambda_1}^{-1}(\alpha) \le F_{\lambda_2}^{-1}(\alpha)\ ,\ q_\alpha(f_{\lambda_1}) \le q_\alpha(f_{\lambda_2})$$

for given $r > 0$ and $\alpha \in (0,1)$ provided $\lambda_1 \le \lambda_2$. Therefore $q_\alpha(f_\lambda)$ is an increasing function of λ for fixed $\alpha \in (0,1)$. Hence $\lambda \in C(y)$ is satisfied for some closed interval $[\lambda_l(y), \lambda_u(y)]$ where

$$y = q_{1-\frac{\alpha}{2}}(f_{\lambda_l(y)}) \quad \text{and} \quad y = q_{\frac{\alpha}{2}}(f_{\lambda_u(y)})$$

We should find such $\lambda_l(y), \lambda_u(y)$ such that

$$F_{\lambda_l(y)}(y) = 1 - \frac{2}{\alpha} \quad \text{and} \quad F_{\lambda_u(y)}(y) = \frac{\alpha}{2}$$

where F_λ is a cdf of $\Gamma(10, \lambda)$ and $y = 20$ is observed.

```
pgamma(20, shape=10, scale = 1) # lambda=1
```

```
## [1] 0.9950046
```

```
pgamma(20, shape=10, scale = 1.5) # lambda=1.5
```

```
## [1] 0.8550948
```

```
# Since F_\lambda(y) is decreasing in \lambda , \lambda_L must lie b.w. 1 and 1.5

candidate=seq(from=1, to=1.5, length=1000)
p_L=0
lambda_L=0
for(i in 1:length(candidate)){
  lambda_L[i] = candidate[i]
  p_L[i] = pgamma(20, shape=10, scale = candidate[i])
}
i_L = which.min(abs(p_L-0.975)) # 1- 2/alpha = 0.975
lambda_L[i_L]
```

```
## [1] 1.170671
```

```
abs(p_L[i_L]-0.975) < 1e-5
```

```
## [1] TRUE
```

```
pgamma(20, shape=10, scale = 4)  # lambda=4
```

```
## [1] 0.03182806
```

```
pgamma(20, shape=10, scale = 4.5)  # lambda=4.5
```

```
## [1] 0.01584119
```

```
# Since F_\lambda(y) is decreasing in \lambda , \lambda_U must lie b.w. 4 and 4.5

candidate=seq(from=4, to=4.5, length=1000)
p_U=0
lambda_U=0
for(i in 1:length(candidate)){
  lambda_U[i] = candidate[i]
  p_U[i] = pgamma(20, shape=10, scale = candidate[i])
}
i_U = which.min(abs(p_U-0.025))  # 2/alpha = 0.025
lambda_U[i_U]
```

```
## [1] 4.170671
```

```
abs(p_U[i_U]-0.025) < 1e-5
```

```
## [1] TRUE
```

Therefore, the confidence limit endpoints are computed as

$$\hat{\lambda}(0.025) = 1.1707 \quad \text{and} \quad \hat{\lambda}(0.975) = 4.1707$$

Exercise 11.3

3. Suppose $\widehat{G}$ in (11.33) was perfectly normal, say $\widehat{G} \sim \mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$. What does $\hat{\theta}_{\mathrm{BC}}(\alpha)$ reduce to in this case, and why does this make intuitive sense?

$$p_0 = \# \left\{ \hat{\theta}^{*b} \le \hat{\theta} \right\} \Big/ B \tag{11.31}$$

(an estimate of (11.29)), and define the *bias-correction value*

$$z_0 = \Phi^{-1}(p_0), \tag{11.32}$$

where Φ^{-1} is the inverse function of the standard normal cdf. The BC level-α confidence interval endpoint is defined to be

$$\hat{\theta}_{\mathrm{BC}}[\alpha] = \hat{G}^{-1}\left[\Phi\left(2z_0 + z^{(\alpha)}\right)\right], \tag{11.33}$$

where $\hat{G}$ is the bootstrap cdf (11.16) and $z^{(\alpha)} = \Phi^{-1}(\alpha)$ (11.25).

Suppose $\hat{G} \sim N(\hat{\mu}, \hat{\sigma}^2)$, perfectly normal with $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}(x_i - \overline{x})^2$.
Note that if $F_{\mu,\sigma}$ is a cdf of $N(\mu, \sigma^2)$ then $F_{\mu,\sigma}(x) = \Phi(\frac{x-\mu}{\sigma})$. Hence $F^{-1}_{u,\sigma}(\alpha) = \sigma\Phi^{-1}(\alpha) + \mu$. Why?

$$F_{\mu,\sigma}(X \le \sigma\Phi^{-1}(\alpha) + \mu) = F_{\mu,\sigma}\big(\frac{X-\mu}{\sigma} \le \Phi^{-1}(\alpha)\big) = \Phi(\Phi^{-1}(\alpha)) = \alpha \quad \forall\, \alpha \in (0,1)$$

Thus , we have

$$\hat{\theta}_{BC}(\alpha) = \hat{G}^{-1}(\Phi(2z_0 + z_{(\alpha)})) = \hat{\sigma}\Phi^{-1}(\Phi(2z_0 + z_{(\alpha)})) + \hat{\mu} = \hat{\sigma}(2z_0 + z_{(\alpha)}) + \hat{\mu}$$

$\hat{\theta}_{BC}(\alpha)$ can be reduced to $\hat{\sigma}(2z_0 + z_{(\alpha)}) + \hat{\mu}$. If $z_0 = 0$, then $\hat{\theta}_{BC}(\alpha) = \hat{\sigma}z_{(\alpha)} + \hat{\mu}$. This is just a normal α-quantile of assumed bootstrap distribution $N(\hat{\mu}, \hat{\sigma})$, which makes sense because when $z_0 = 0$, we can think that there is no bias in bootstrap distribution.
However, if $z_0 > 0$, then we can think that there is downward bias in bootstrap distribution so that adjusting upward is required. Hence $\hat{\theta}_{BC}(\alpha)$ is adjusted upward by $2\hat{\sigma}z_0$ from the original normal α-quantile. On the other hand, if $z_0 < 0$, then we can think there is upward bias in bootstrap distribution so that adjusting downward is necessary. Thus $\hat{\theta}_{BC}(\alpha)$ is adjusted downward by $2\hat{\sigma}|z_0|$ from the original normal α-quantile.

Exercise 11.5

5. Suppose $\hat{\theta} \sim \text{Poisson}(\theta)$ is observed to equal 16. Without employing simulation, compute the 95% central BCa interval for θ. (You can use the good approximation $z_0 = a = 1/(6\hat{\theta}^{1/2})$.)

The *BCa method* ("bias-corrected and accelerated") takes its level-α confidence limit to be

$$\hat{\theta}_{\mathrm{BCa}}[\alpha] = \hat{G}^{-1}\left[\Phi\left(z_0 + \frac{z_0 + z^{(\alpha)}}{1 - a(z_0 + z^{(\alpha)})}\right)\right]. \tag{11.39}$$

$\hat{\theta} \sim Poi(\theta)$ and $\hat{\theta} = 16$ is observed. By the hint , we shall plug in $z_0 = a = \frac{1}{6\hat{\theta}^{1/2}} = \frac{1}{24}$ on formula 11.39 with $\alpha = 0.025$ and 0.975 .

For bootstrap distribution $\hat{G}$, from $\hat{\theta} \sim Poi(\theta)$, we expect $\hat{\theta}^* \sim Poi(\hat{\theta})$. Thus, without employing simulation, we can plug in $\hat{G}$ = cdf of $Poi(\hat{\theta})$

```
theta.hat = 16
z0 = 1/(6*sqrt(theta.hat))
a = 1/(6*sqrt(theta.hat))
alpha = 0.05
L = pnorm(z0+ (z0+qnorm(alpha/2)) / (1-a*(z0+qnorm(alpha/2))))
U = pnorm(z0+ (z0+qnorm(1-alpha/2)) / (1-a*(z0+qnorm(1-alpha/2))))
BCa = c(qpois(L, lambda=theta.hat) ,qpois(U, lambda=theta.hat))
BCa
```

```
## [1]  9 26
```

Thus, 95% central BCa interval for θ is $(9, 26)$

Exercise 11.6

6. Use the R program **`bcajack`** (available with its help file from **`efron.web.stanford.edu`** under "Talks") to find BCa confidence limits for the student score eigenratio statistic as in Figure 10.2.

First, bring R program `bcajack` from efron.web.stanford.edu - 2018 : Supplement files for bcajack

Next, load the student score data.

```
data=read.table("https://web.stanford.edu/~hastie/CASI_files/DATA/student_score.txt", header=T)
```

To take advantage of `bcajack` function created by Efron, we need to create a function `EigenRatio` which calculates the ratio $\frac{\text{largest eigenvalue}}{\text{sum of all eigenvalues}}$ of correlation matrix for input dataset.

```
EigenRatio<-function(X){
  R = cor(X)
  result = max(eigen(R, only.values=TRUE)$values) / sum(eigen(R, only.values=TRUE)$values)
  return(result)
}
```

Then, we can use `bcajack` function with `EigenRatio`

```
library(matlib) # To read a `len` function in `bcajack` , we need this library
set.seed(123)
bca_result = bcajack(x = data , B = 1000 , func = EigenRatio ,  alpha= c(0.025, 0.975), m=20, catj = 0)
```

```
## {1}{2}
```

```
bca_result
```

```
## $call
## bcajack(x = data, B = 1000, func = EigenRatio, m = 20, alpha = c(0.025,
##     0.975), catj = 0)
##
## $lims
##       bcalims jacksd standard   pct
## 0.025   0.516    NaN    0.542 0.028
## 0.975   0.820    NaN    0.843 0.978
##
## $stats
##      thet sdboot    z0     a sdjack
```

```
## est 0.693  0.077 -0.07 0.051  0.081
## jsd 0.000  0.003   NaN 0.000  0.000
##
## $B.mean
## [1] 1000.0000000    0.6873815
```

From the result above, for $\theta =$ eigenratio , BCa confidence limit endpoints $\hat{\theta}_{BCa}(0.025)$ and $\hat{\theta}_{BCa}(0.975)$ are given as $(0.516, 0.820)$