Advanced Statistical Methods Hw6

Do Hyup Shin

10/24/2021

Problem 7.2

In table 7.1, suppose the MLE batting averages were based on 180 at-bats for each player, rather than 90. What would the JS column look like?

Solution

There are two methods in the book. I'll show tables for n = 180 cases.

First table is based on normal assumption method and second table is based on arcsine method.

```
library(knitr)
bball <- read.csv("https://web.stanford.edu/~hastie/CASI_files/DATA/baseball.txt", sep = " ")

#1 Using normal transformation
n <- nrow(bball)
num <- 180
true_p <- bball$TRUTH
mle_p <- bball$MLE
sig_p  <- mean(mle_p)/num
s_mle <- sum((mle_p - mean(mle_p))^2)

p_js <- mean(mle_p) + (1 - 15*sig_p/s_mle)*(mle_p - mean(mle_p))


mse_js <- sum((p_js - true_p)^2)
mse_mle <- sum((mle_p - true_p)^2)

bball[,"JS"] <- p_js
bball <- bball[, c(1,2,4,3)]
kable(bball, caption = "baseball player 180 at-bats(normal method) ",
      align=c("c","c","c","c"))
```

Table 1: baseball player 180 at-bats(normal method)

Player	MLE	JS	TRUTH
1	0.345	0.3047351	0.298
2	0.333	0.2980222	0.346
3	0.322	0.2918687	0.222
4	0.311	0.2857151	0.276
5	0.289	0.2734081	0.263
6	0.289	0.2734081	0.273
7	0.278	0.2672545	0.303

Player	MLE	JS	TRUTH
8	0.255	0.2543881	0.270
9	0.244	0.2482345	0.230
10	0.233	0.2420810	0.264
11	0.233	0.2420810	0.264
12	0.222	0.2359275	0.210
13	0.222	0.2359275	0.256
14	0.222	0.2359275	0.269
15	0.211	0.2297740	0.316
16	0.211	0.2297740	0.226
17	0.200	0.2236204	0.285
18	0.145	0.1928528	0.200

```r
#2 Using arcsine transformation
x <- 2*sqrt(num+0.75)*asin(sqrt((num*mle_p+0.375)/(num+0.75)))
mu <- 2*sqrt(num+0.75)*asin(sqrt((num*true_p+0.375)/(num+0.75)))
s_x <- sum((x - mean(x))^2)

mu_js <- mean(x) + (1 - 15/s_x)*(x - mean(x))


p_arcs_js <- 1/num*{(num+0.75) * (sin(mu_js/(2*sqrt(num+0.5))))^2 - 0.375}

bball[,"JS"] <- p_arcs_js
bball[,"x"] <- x

kable(bball,
      caption = "baseball player 180 at-bats(arcsine method) ",
      align=c("c","c","c","c","r"))
```

Table 2: baseball player 180 at-bats(arcsine method)

Player	MLE	JS	TRUTH	x
1	0.345	0.3142896	0.298	16.89893
2	0.333	0.3063570	0.346	16.55968
3	0.322	0.2990731	0.222	16.24603
4	0.311	0.2917749	0.276	15.92958
5	0.289	0.2771249	0.263	15.28723
6	0.289	0.2771249	0.273	15.28723
7	0.278	0.2697672	0.303	14.96076
8	0.255	0.2542951	0.270	14.26494
9	0.244	0.2468457	0.230	13.92505
10	0.233	0.2393586	0.264	13.57999
11	0.233	0.2393586	0.264	13.57999
12	0.222	0.2318294	0.210	13.22927
13	0.222	0.2318294	0.256	13.22927
14	0.222	0.2318294	0.269	13.22927
15	0.211	0.2242528	0.316	12.87236
16	0.211	0.2242528	0.226	12.87236
17	0.200	0.2166233	0.285	12.50865
18	0.145	0.1774267	0.200	10.56016

Problem 7.3

In table 7.1, calculate the JS column based on (7.20).

Solution

By using problem 7.2 solution, we can easily find the james stein estimator under normal assumption.

```r
library(knitr)
bball <- read.csv("https://web.stanford.edu/~hastie/CASI_files/DATA/baseball.txt", sep = " ")

#1 Using normal transformation
n <- nrow(bball)
num <- 90
true_p <- bball$TRUTH
mle_p <- bball$MLE
sig_p  <- mean(mle_p)/num
s_mle <- sum((mle_p - mean(mle_p))^2)

p_js <- mean(mle_p) + (1 - 15*sig_p/s_mle)*(mle_p - mean(mle_p))


mse_js <- sum((p_js - true_p)^2)
mse_mle <- sum((mle_p - true_p)^2)

bball[,"JS"] <- p_js
bball <- bball[, c(1,2,4,3)]
kable(bball, caption = "baseball player 90 at-bats(normal method) ",
      align=c("c","c","c","c"))
```

Table 3: baseball player 90 at-bats(normal method)

Player	MLE	JS	TRUTH
1	0.345	0.2644703	0.298
2	0.333	0.2630444	0.346
3	0.322	0.2617373	0.222
4	0.311	0.2604303	0.276
5	0.289	0.2578161	0.263
6	0.289	0.2578161	0.273
7	0.278	0.2565091	0.303
8	0.255	0.2537761	0.270
9	0.244	0.2524691	0.230
10	0.233	0.2511620	0.264
11	0.233	0.2511620	0.264
12	0.222	0.2498550	0.210
13	0.222	0.2498550	0.256
14	0.222	0.2498550	0.269
15	0.211	0.2485479	0.316
16	0.211	0.2485479	0.226
17	0.200	0.2472409	0.285
18	0.145	0.2407056	0.200

Problem 7.5

Your brother-in-law's favorite player, number 4 in Table 7.1, is batting .311 after 90 at-bats, but JS predicts only 0.272. He says that this is due to the lousy 17 other players, who didn't have anything to do with number 4's results and are averaging only 0.250. How would you answer him?

Solution

Actually, the james-stein estimator is shrinkage estimator. The shrinkage estimate does not work well when it is really good or bad. As the brother-in-law argues in the problem, the JS estimator of number 4 player is much lower than the actual value because other 17 players.

Problem 2

Show that the Bayes risk of James-Stein estimator (M = 0 case) is NB + $\frac{2}{A+1}$ (see note page 3).

Solution

Let $\mu_i \overset{iid}{\sim} N(M, A)$ and $x_i|\mu_i \overset{iid}{\sim} N(\mu_i, 1)$. Assume that $M = 0$. This means $\mu_i \overset{iid}{\sim} N(0, A)$. Define $S = \sum_{i=1}^{N} x_i^2$ and $B = \frac{A}{A+1}$. Then, we know that $\mu_i|x_i \overset{iid}{\sim} N(Bx_i, B)$, $x_i \overset{iid}{\sim} N(0, A+1)$ and $\frac{S}{A+1} \sim \chi^2(N) = Gamma(\frac{N}{2}, 2)$. Then, $E(\frac{S}{A+1}) = N$ and $E(\frac{A+1}{S}) = \frac{1}{N-2} \quad \because \frac{A+1}{S} \sim InverseGamma(\frac{N}{2}, 2)$.

Also, we know that the bayes estimator of μ_i is $\hat{\mu}_i^B = E(\mu_i|x_i) = Bx_i$ and james-stein estimator of μ_i is $\hat{\mu}_i^{JS} = (1 - \frac{N-2}{S})x_i$.

we'll show that $E_{X,\mu}(\left\|\hat{\mu}^{JS} - \mu\right\|^2) = NB + \frac{2}{A+1}$.

Since we know that $E(X) = E(E(X|Y)) \quad \forall X, Y$ random variables which is double expectation theorem, we apply this for $\hat{\mu}_i^{JS}$ each $i = 1, \ldots, N$.

$$
\begin{aligned}
E((\hat{\mu}_i^{JS} - \mu_i)^2|X) &= E((\hat{\mu}_i^{JS} - \hat{\mu}_i^B + \hat{\mu}_i^B - \mu_i)^2|X) \\
&= E((\hat{\mu}_i^{JS} - \hat{\mu}_i^B)^2|X) + E((\hat{\mu}_i^B - \mu_i)^2|X) + 2E((\hat{\mu}_i^{JS} - \hat{\mu}_i^B)(\hat{\mu}_i^B - \mu_i)|X)
\end{aligned}
$$

Define $C = E(\hat{\mu_i}^{JS} - \hat{\mu_i}^B)(\hat{\mu_i}^B - \mu_i)|X)$. We'll show that $C = 0$.

$$
\begin{aligned}
C &= E((\hat{\mu}_i^{JS} - \hat{\mu}_i^B)(\hat{\mu}_i^B - \mu_i)|X) \\
&= (\hat{\mu}_i^{JS} - \hat{\mu}_i^B)E((\hat{\mu}_i^B - \mu_i)|X) \quad \because (\hat{\mu}_i^{JS} - \hat{\mu}_i^B) \text{ is function of X} \\
&= (\hat{\mu}_i^{JS} - \hat{\mu}_i^B)(\hat{\mu}_i^B - E(\mu_i|X)) \quad \because \hat{\mu}_i^B \text{ is function of X} \\
&= (\hat{\mu}_i^{JS} - \hat{\mu}_i^B)(\hat{\mu}_i^B - \hat{\mu}_i^B) = 0
\end{aligned}
$$

Then, we can simplify for MSE of $\hat{\mu}_i^{JS}$ conditional X.

$$
\begin{aligned}
E((\hat{\mu}_i^{JS} - \mu_i)^2|X) &= E((\hat{\mu}_i^{JS} - \hat{\mu}_i^B)^2|X) + E((\hat{\mu}_i^B - \mu_i)^2|X) + 2E((\hat{\mu}_i^{JS} - \hat{\mu}_i^B)(\hat{\mu}_i^B - \mu_i)|X) \\
&= E((\hat{\mu}_i^{JS} - \hat{\mu}_i^B)^2|X) + E((\hat{\mu}_i^B - \mu_i)^2|X) \\
&= E((1 - \frac{N-2}{S} - B)^2 x_i^2|X) + Var(\mu_i|X) \quad \because \hat{\mu}_i^B = E(\mu_i|x_i) \\
&= (1 - \frac{N-2}{S} - B)^2 x_i^2 + B \quad \because (1 - \frac{N-2}{S} - B)^2 x_i^2 \text{ is function of X}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E(\left\|\hat{\mu}^{JS}-\mu\right\|^2) &= E(E(\left\|\hat{\mu}^{JS}-\mu\right\|^2|X)) \\
&= E(\sum_{i=1}^{N}((1-\frac{N-2}{S}-B)^2x_i^2))+NB \\
&= E((1-\frac{N-2}{S}-B)^2S)+NB \\
&= E((1-B)^2S+(\frac{N-2}{S})^2S-2(1-B)\frac{N-2}{S}S)+NB \\
&= \frac{1}{(A+1)^2}E(S)+(N-2)E(\frac{N-2}{S})-\frac{2(N-2)}{A+1}+NB \\
&= \frac{1}{(A+1)^2}(A+1)N+\frac{N-2}{A+1}-\frac{2(N-2)}{A+1}+NB \\
&= NB+\frac{2}{A+1}
\end{aligned}
$$

Problem 3

Let $\hat{\mu}_i$ be the ith coordinate of the JS-estimator in the setting of p. 93 (of the textbook). Compare the risk of $\hat{\mu}_i$ with that of the MLE of μ_i.

Solution

In this case, $M \neq 0$. So similar to problem 2, we can show the above problem.

Let $\mu_i \overset{iid}{\sim} N(M,A)$ and $x_i|\mu_i \overset{iid}{\sim} N(\mu_i,1)$. Define $S=\sum_{i=1}^{N}(x_i-\bar{x})^2$ and $B=\frac{A}{A+1}$ where $\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i$. Then, we know that $\mu_i|x_i \overset{iid}{\sim} N(M+B(x_i-M),B)$, $x_i \overset{iid}{\sim} N(M,A+1)$ and $\frac{S}{A+1} \sim \chi^2(N-1) = Gamma(\frac{N-1}{2},2)$. Then, $E(\frac{S}{A+1})=N-1$ and $E(\frac{A+1}{S})=\frac{1}{N-3}$ $\quad \because \frac{A+1}{S} \sim InverseGamma(\frac{N-1}{2},2)$.

Also, we know that the bayes estimator of μ_i is $\hat{\mu}_i^B=E(\mu_i|x_i)=M+B(x_i-M)$ and james-stein estimator of μ_i is $\hat{\mu}_i^{JS}=\bar{x}+(1-\frac{N-3}{S})(\bar{x}-x_i)$. Let $\hat{B}=1-\frac{N-3}{S}$

we'll show that $E_{X,\mu}(\left\|\hat{\mu}^{JS}-\mu\right\|^2)=NB+\frac{3}{A+1}$.

Since we know that $E(X)=E(E(X|Y)) \quad \forall X,Y$ random variables which is double expectation theorem, we apply this for $\hat{\mu}_i^{JS}$ each $i=1,\ldots,N$.

$$
\begin{aligned}
E((\hat{\mu}_i^{JS}-\mu_i)^2|X) &= E((\hat{\mu}_i^{JS}-\hat{\mu}_i^B+\hat{\mu}_i^B-\mu_i)^2|X) \\
&= E((\hat{\mu}_i^{JS}-\hat{\mu}_i^B)^2|X)+E((\hat{\mu}_i^B-\mu_i)^2|X)+2E((\hat{\mu}_i^{JS}-\hat{\mu}_i^B)(\hat{\mu}_i^B-\mu_i)|X)
\end{aligned}
$$

Define $C=E(\hat{\mu_i}^{JS}-\hat{\mu_i}^B)(\hat{\mu_i}^B-\mu_i)|X)$. We'll show that $C=0$.

$$
\begin{aligned}
C &= E((\hat{\mu}_i^{JS}-\hat{\mu}_i^B)(\hat{\mu}_i^B-\mu_i)|X) \\
&= (\hat{\mu}_i^{JS}-\hat{\mu}_i^B)E((\hat{\mu}_i^B-\mu_i)|X) \quad \because (\hat{\mu}_i^{JS}-\hat{\mu}_i^B) \text{ is function of X} \\
&= (\hat{\mu}_i^{JS}-\hat{\mu}_i^B)(\hat{\mu}_i^B-E(\mu_i|X)) \quad \because \hat{\mu}_i^B \text{ is function of X} \\
&= (\hat{\mu}_i^{JS}-\hat{\mu}_i^B)(\hat{\mu}_i^B-\hat{\mu}_i^B)=0
\end{aligned}
$$

Then, we can simplify for MSE of $\hat{\mu}_i^{JS}$ conditional X.

$$
\begin{aligned}
E((\hat{\mu}_i^{JS}-\mu_i)^2|X) &= E((\hat{\mu}_i^{JS}-\hat{\mu}_i^{B})^2|X) + E((\hat{\mu}_i^{B}-\mu_i)^2|X) + 2E((\hat{\mu}_i^{JS}-\hat{\mu}_i^{B})(\hat{\mu}_i^{B}-\mu_i)|X) \\
&= E((\hat{\mu}_i^{JS}-\hat{\mu}_i^{B})^2|X) + E((\hat{\mu}_i^{B}-\mu_i)^2|X) \\
&= E((\bar{x}+\hat{B}(x_i-\bar{x})-M-B(x_i-\bar{X}))^2|X) + Var(\mu_i|X) \quad \because \hat{\mu}_i^{B} = E(\mu_i|x_i) \\
&= E(((1-B)(\bar{x}-M)+(\hat{B}-B)(x_i-\bar{x}))^2|X) + B \\
&= ((1-B)(\bar{x}-M)+(\hat{B}-B)(x_i-\bar{x}))^2 + B \quad \because ((1-B)(\bar{x}-M)+(\hat{B}-B)(x_i-\bar{x}))^2 \text{ is function of X}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E(\left\|\hat{\mu}^{JS}-\mu\right\|^2) &= E(E(\left\|\hat{\mu}^{JS}-\mu\right\|^2|X)) \\
&= E(\sum_{i=1}^{N}((1-B)(\bar{x}-M)+(\hat{B}-B)(x_i-\bar{x}))^2) + NB \\
&= E(N(1-B)^2(\bar{x}-M)^2 + (\hat{B}-B)^2 S + 2(1-B)(\hat{B}-B)(\bar{x}-M)\sum_{i=1}^{N}(x_i-\bar{x})) + NB \\
&= N(1-B)^2 Var(\bar{x}) + E((\hat{B}-B)^2 S) + 0 + NB \quad \because \sum_{i=1}^{N}(x_i-\bar{x}) = 0 \\
&= N\frac{1}{(1+A)^2}\frac{A+1}{N} + E((1-\frac{N-3}{S}-B)^2 S) + NB \because \bar{x} \sim N(M, \frac{A+1}{N}) \\
&= \frac{1}{A+1} + E((1-B)^2 S) + E((\frac{N-3}{S})^2 S) - 2E((1-B)\frac{N-3}{S}S) + NB \\
&= \frac{1}{A+1} + \frac{N-1}{A+1} + \frac{N-3}{A+1} - 2\frac{N-3}{A+1} + NB \because E(\frac{S}{A+1}) = N-1, \quad E(\frac{A+1}{S}) = \frac{1}{N-3} \\
&= \frac{1}{A+1} + \frac{N-1}{A+1} + \frac{N-3}{A+1} - 2\frac{N-3}{A+1} + NB \\
&= \frac{3}{A+1} + NB
\end{aligned}
$$

The mle of μ_i is $\hat{\mu}_i^{MLE} = x_i$, $E((\hat{\mu}_i^{MLE}-\mu_i)^2) = 1$. $E(\left\|\hat{\mu}^{MLE}-\mu\right\|^2) = N$. Thus, if $N \geq 3$, $E(\left\|\hat{\mu}^{MLE}-\mu\right\|^2) \leq E(\left\|\hat{\mu}^{JS}-\mu\right\|^2)$. But, we cannot assure that for each $\hat{\mu}_i^{JS}$ is better than $\hat{\mu}_i^{MLE}$. This means that there might be for some i such that $E(\hat{\mu}_i^{MLE}-\mu_i)^2 \nleq E(\left\|\hat{\mu}^{JS}-\mu\right\|^2)$.