

# HW5

Zequi.Yu

2022-10-30

## Homework 5

### Question 1:

Set seed of the random number generator to be 12345 using `set.seed(12345)`. Generate 100 random numbers from exponential distribution with mean=1. Let  $s^2$  be the sample variance, an estimator of the population variance (6 pts).

```
set.seed(12345)
SampleSize1 <- 100
MySample1 <- rexp(SampleSize1, 1)
```

1. Use bootstrap method to estimate the variance of  $s^2$ .

```
set.seed(12345)
B <- 100000
Samplevariances1 <- numeric(B)
for(i in 1:B){
  Samplevariance <- sample(MySample1,100,replace=TRUE)
  Samplevariances1[i] <- var(Samplevariance)
}
print(sprintf("The estimation of the variance of s^2 using bootstrap is %.3f", var(Samplevariances1)))
```

```
## [1] "The estimation of the variance of s^2 using bootstrap is 0.137"
```

2. Using Monte-Carlo method to generate 1000 random samples of size 100, compute variance of  $s^2$  evaluated based on the 1000 random samples, and compare it with results from part (a).

```
set.seed(12345)
Sample.num <- 1000
Samplevariances2 <- numeric(Sample.num)
for(i in 1:Sample.num){
  x <- rexp(100,1)
  Samplevariances2[i] <- var(x)
}
print(sprintf("The estimation of the variance of s^2 using MC is %.3f.", var(Samplevariances2)))
```

```
## [1] "The estimation of the variance of s^2 using MC is 0.085."
```

It is much smaller than the result in (a), it has higher efficiency.

### Question 2:

Use the same seed as in 1, generate a random sample with n=15 random observations from exponential distribution with mean=1 (6 pts).

```
set.seed(12345)
SampleSize2 <- 15
MySample2 <- rexp(SampleSize2, 1)
```

1. Calculate the sample median, which is an estimator of the population median.

```
SampleMedian <- median(MySample2)
print(sprintf("The sample median %.3f", SampleMedian))
```

```
## [1] "The sample median 0.455"
```

2. Use bootstrap (nonparametric, with B=1000) methods to estimate the variance of the estimator for the population median.

```
set.seed(12345)
B <- 1000
SampleMedians <- numeric(B)
for(i in 1:B){
  SampleMedian <- sample(MySample2,15,replace=TRUE)
  SampleMedians[i] <- median(SampleMedian)
}
print(sprintf("The variance of the estimator for the population median is %.3f using bootstrap", var(SampleMedians)))
```

```
## [1] "The variance of the estimator for the population median is 0.044 using bootstrap"
```

3. use Monte Carlo method, e.g. generate 1000 samples of size 15 to estimate the true variance of the median estimator. Compare and comment your results.

```
set.seed(12345)
Sample.num = 1000
SampleMedians2 <- numeric(Sample.num)
for(i in 1:Sample.num){
  x <- rexp(15,1)
  SampleMedians2[i] <- median(x)
}
print(sprintf("The variance of the estimator for the population median is %.3f using MC", var(SampleMedians2)))
```

```
## [1] "The variance of the estimator for the population median is 0.068 using MC"
```

Bootstrap works better when the estimator is a location parameter.

Question 3:

Patients with advanced terminal cancer of the stomach and breast were treated with ascorbate in an attempt to prolong survival [87]. Table 9.4 shows survival times (days). Work with the data on the log scale. (a.) Use the bootstrap t method to construct 95% confidence intervals for the mean survival time of each group.

```
t1 <- log(c(25,42,45,46,51,103,124,146,340,396,412,876,1112))
t2 <- log(c(24,40,719,727,791,1166,1235,1581,1804,3460,3808))

locator1 <- mean(t1)
locator2 <- mean(t2)
sigma1 <- sqrt(var(t1)/length(t1))
sigma2 <- sqrt(var(t2)/length(t2))
B <- 1000000
MST1s <- numeric(B)
MST2s <- numeric(B)
set.seed(12345)
for(i in 1:B){
```

```

MST1 <- sample(t1,length(t1),replace=TRUE)
MST2 <- sample(t2,length(t2),replace=TRUE)
M1 <- mean(MST1)
M2 <- mean(MST2)
MST1s[i] <- (M1 - locator1)/sqrt(var(MST1)/length(MST1))
MST2s[i] <- (M2 - locator2)/sqrt(var(MST2)/length(MST2))
}

q1 <- quantile(MST1s, c(0.025,0.975))
q2 <- quantile(MST2s, c(0.025,0.975))
bounds1 <- exp(c(locator1 - sigma1*q1[[2]], locator1 - sigma1*q1[[1]]))
bounds2 <- exp(c(locator2 - sigma2*q2[[2]], locator2 - sigma2*q2[[1]]))
names(bounds1) <- c("2.5%", "97.5")
names(bounds2) <- c("2.5%", "97.5")
bounds1

##      2.5%      97.5
## 70.05245 330.21379

bounds2

##      2.5%      97.5
## 71.6183 1648.3476

```