344HW5

Shay Lebovitz

5/19/2020

Problem 1

```
set.seed (12345)
n <- 100
x <- rexp (n, 1)
s_squared <- var(x)
s_squared
```

```
## [1] 1.265815
```

```
B <- 1000
vars <- numeric (B)
for (i in 1:B) {
  y <- sample(x, 100, replace=TRUE)
  vars[i] = var(y)
}

var(vars)
```

```
## [1] 0.1345509
```

We see that the sample variance is 1.266. Using the Bootstrap method, we estimate the variance of the sample variance and receive 0.135.

Now using normal Monte Carlo method:

```
set.seed (12345)
m <- 1000
vars2 <- numeric(1000)
for (i in 1:m) {
  obs <- rexp (100, 1)
  vars2[i] <- var(obs)
}
var(vars2)
```

```
## [1] 0.08511443
```

We see that the Monte Carlo estimate is 0.085, fairly close to the bootstrap estimate.

Problem 2

```
set.seed (12345)
n <- 15
y <- rexp (n, 1)
sample_median <- median(y)
sample_median
```

```
## [1] 0.4554105
```

```
z_median = numeric(1000)
B <- 1000
for (i in 1:B) {
  z <- sample (y, 15, replace=TRUE)
  z_median[i] <- median(z)
}

bootstrap_median_var <- var(z_median)
bootstrap_median_var
```

```
## [1] 0.04664743
```

We see that the sample median is 0.455. From the bootstrap analysis, we see that the variance of the estimate is 0.0466.

Now for normal Monte Carlo Analysis

```
set.seed (12345)
m <- 1000
MC_median <- numeric(1000)
for (i in 1:m) {
  obs <- rexp (15, 1)
  MC_median[i] = median (obs)
}
var(MC_median)
```

```
## [1] 0.06828018
```

The Monte Carlo analysis gives a variance of 0.0683, fairly close to the Bootstrap estimate.

Problem 9.5(a)

```
stomach <- c(25, 42, 45, 46, 51, 103, 124, 146, 340, 396, 412, 876, 1112)
log_stomach = log(stomach)
breast <- c(24, 40, 719, 727, 791, 1166, 1235, 1581, 1804, 3460, 3808)
log_breast = log(breast)
stomach_mean = mean (log_stomach)
breast_mean = mean (log_breast)
stomach_sd = sd(log_stomach)
breast_sd = sd(log_breast)

#stomach data
s = numeric (B)
for (i in 1:B) {
  samp = sample (log_stomach, 13, replace = TRUE)
  s[i] = (mean(samp) - stomach_mean)/(sd(samp)/sqrt(13))
}

#breast data
b = numeric (B)
for (i in 1:B) {
  samp2 = sample (log_breast, 11, replace = TRUE)
  b[i] = (mean(samp2) - breast_mean)/(sd(samp2)/sqrt(11))
}

stomach_CI = exp(stomach_mean - quantile(s, c(0.025, 0.975))*stomach_sd/sqrt(13))
```

```
breast_CI = exp(breast_mean -  quantile(b, c(0.025, 0.975))*breast_sd/sqrt(11))
stomach_CI
```

```
##      2.5%     97.5%
## 332.43537  71.52293
```

```
breast_CI
```

```
##     2.5%    97.5%
## 1533.837   78.464
```

We see that a 95% confidence interval for the stomach cancer survival times is [71.5, 332.4]. For breast cancer survival times, the confidence interval is [78.5, 1533.8]. Breast cancer has a much larger interval.