ST 502 HW 3 Extra Problem 1

Bruce Campbell

January 29, 2017

Extra Problem 1:

Suppose we are collected data that we believe comes from a normal distribution. The Method of Moments estimators are

$$\mu_{MOM} = \bar{Y}$$

and

$$\hat{\sigma} = S_B^2$$

Our sample of size $n = 30$ gave us $y = 11.3$ and $s2_B = 6.41$

a) Estimate the standard error of both method of moments estimators using a parametric bootstrap.

The SE of μ_{MOM} is given by the standard deviation of our estimator $\hat{\mu} = \bar{Y}$. We know that $Var(\bar{Y}) = \frac{\sigma_Y^2}{n}$ We don't know the true μ here so we can approximate the SE from our sample point estimate $y = 11.3$ as

$$SE_{\hat{Y}} \approx \sqrt{\frac{s_2}{n}} \approx 0.4622$$

The distribution of $\frac{(n-1)S_2}{\sigma^2}$ is the chi-square distribution with $n-1$ degrees of freedom χ^2_{n-1}. From this we can calculate the variance of $\hat{\sigma}$. First we note that $Var(\chi^2_{n-1}) = 2(n-1)$ and remembering that $Var(aY) = a^2 Var(Y)$ we see that

$$Var(\frac{\sigma^2}{n-1} \times \frac{(n-1)S_2}{\sigma^2}) = (\frac{\sigma^2}{n-1})^2 Var(\frac{(n-1)S_2}{\sigma^2}) = (\frac{\sigma^2}{n-1})^2 * 2(n-1)$$

Again, we don't know the true value of σ so we approximate the SE from our point estimate

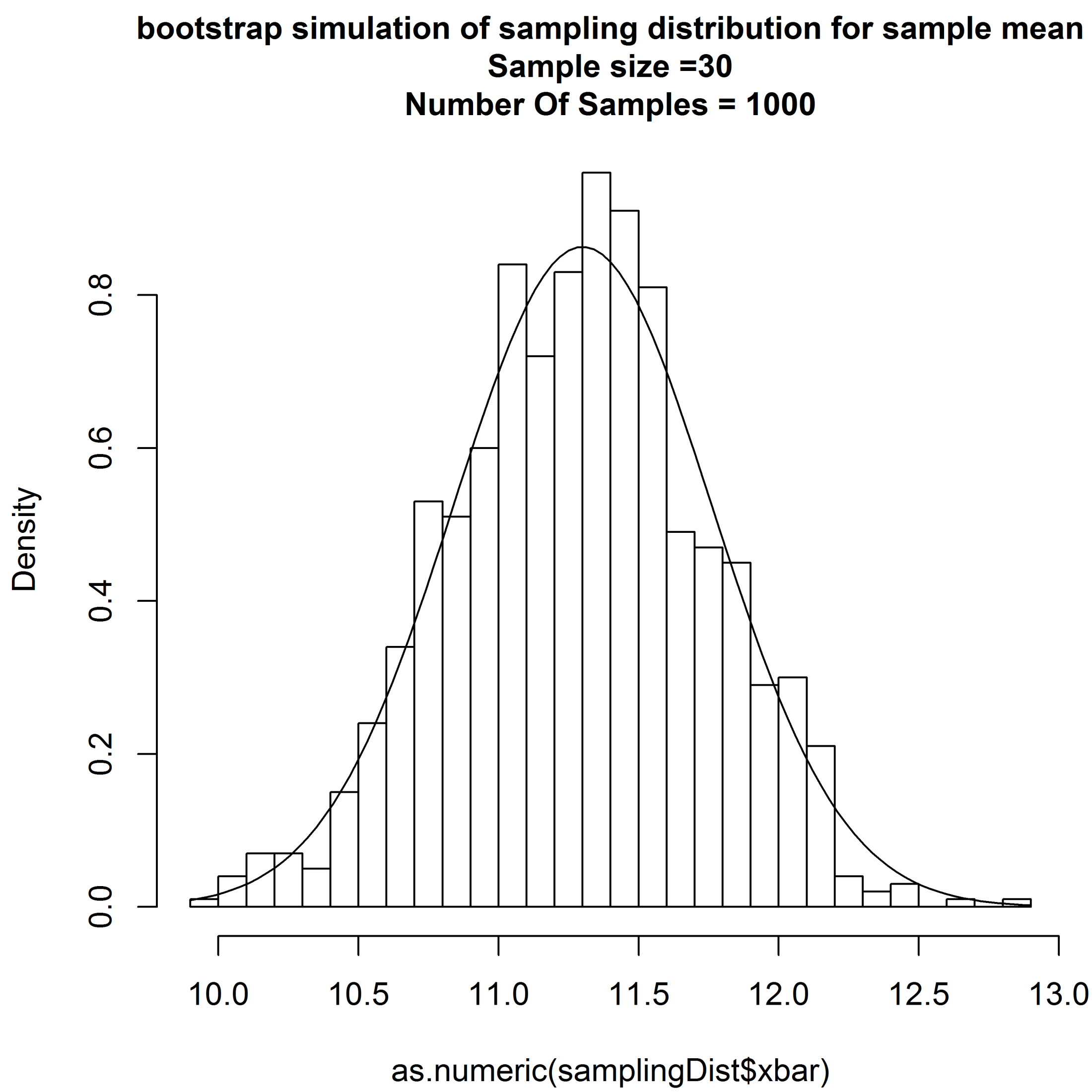
$$SE_{S_2} \approx \sqrt{(\frac{s_2}{n-1})^2 * 2(n-1)} = \sqrt{}((6.41/(30-1))^2 * 2 * (30-1)) \approx 1.683$$

Sometimes we don't have closed form expressios for the SE of the estimator and we must resort to the boostrap to estimate the SE. The bootsrap proceedure is to use the point estimates to simulate a collection SRS's from a distribution with the point estimates as parameters. We can use the empirical distribution of the boostrap estimate to appoximate properties of the sampling distribution of the estimator.

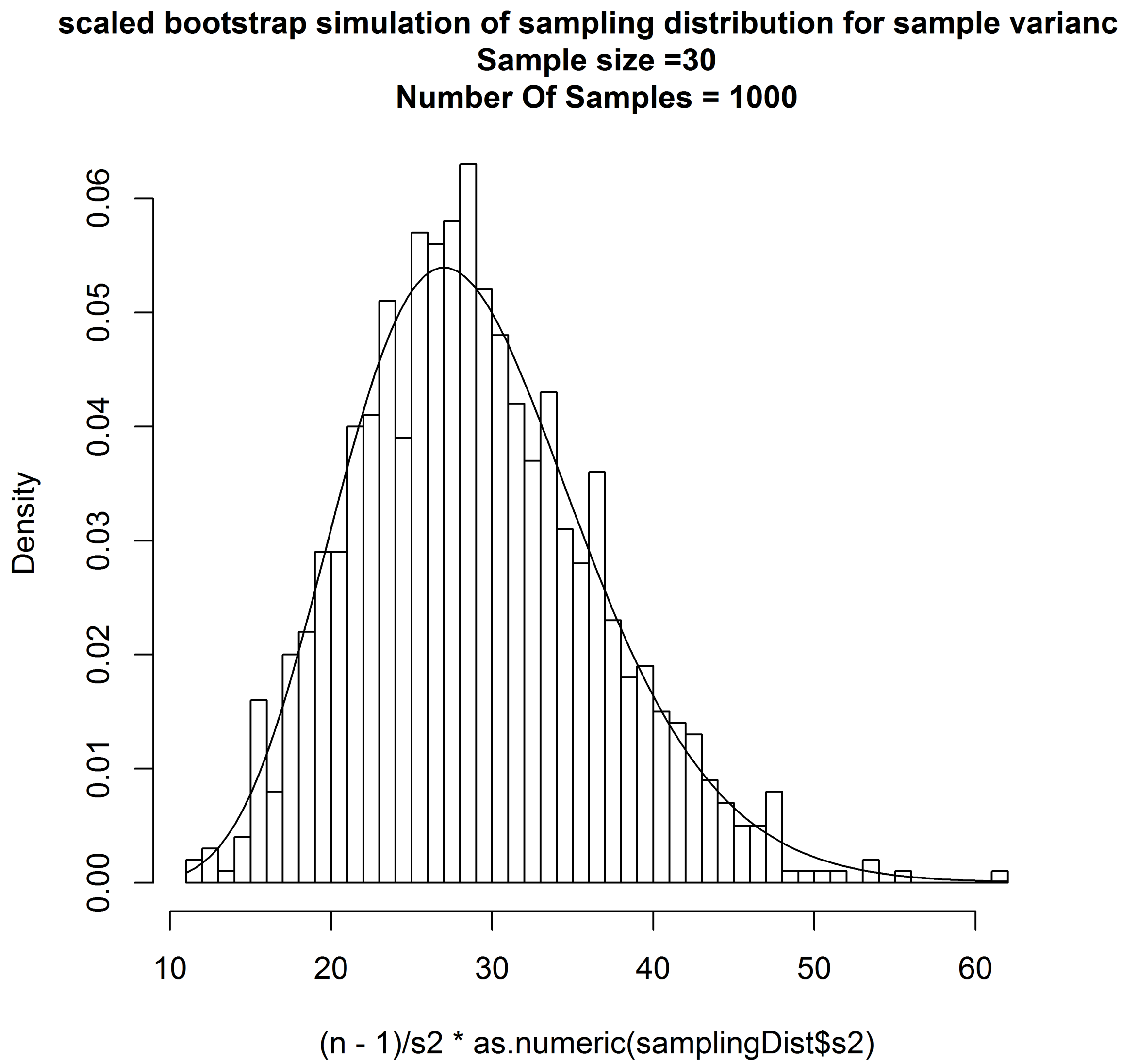
Below we make plots of the boostrap sample for variance. For the plots we scale those values and the CI's by $(n-1)/S_2$ so that we can overlay the χ^2_{n-1} distribution.

```
library(pander)
set.seed(123)
sampleSize <- 30
n <- 30
numberOfSamples <- 1000
yhat <- 11.3
s2 <- 6.41
```

```
samplingDist <- data.frame(matrix(ncol = 2, nrow = numberOfSamples))
colnames(samplingDist) <- c("xbar", "s2")
for (i in 1:numberOfSamples) {
    srs <- rnorm(sampleSize, yhat, sqrt(s2))
    sample_mean <- mean(srs)
    samplingDist[i, 1] = sample_mean
    sample_s2 <- var(srs)
    samplingDist[i, 2] <- sample_s2
}
titleString <- c("bootstrap simulation of sampling distribution for sample mean",
    paste("Sample size =", sampleSize, sep = ""), paste("Number Of Samples = ",
        numberOfSamples, sep = ""))
hist(as.numeric(samplingDist$xbar), 40, freq = FALSE, main = titleString, cex.main = 1)
curve(dnorm(x, mean = yhat, sd = 0.4622), add = TRUE)
```



```
titleString <- c("scaled bootstrap simulation of sampling distribution for sample variance",
    paste("Sample size =", sampleSize, sep = ""), paste("Number Of Samples = ",
        numberOfSamples, sep = ""))
hist((n - 1)/s2 * as.numeric(samplingDist$s2), 40, freq = FALSE, main = titleString,
    cex.main = 1)
curve(dchisq(x, sampleSize - 1), add = TRUE)
```




```
pander(data.frame(SE = sd(samplingDist$xbar)), caption = "SE of bootstrap for estimator of mean")
```

Table 1: SE of bootstrap for estimator of mean

SE
0.4495

```r
pander(data.frame(SE = sd(samplingDist$s2)), caption = "SE of bootstrap for estimator of variance")
```

Table 2: SE of bootstrap for estimator of variance

SE
1.646

We see that our bootstrap estimations are close to the approximate calculations above.

b) Create 95% Basic Percentile Confidence intervals using the parametric bootstrap for both μ and σ^2.

Sample Mean CI

The 95% CI for the sample mean is

$$\bar{Y} \pm z(\frac{\alpha}{2})\frac{\sqrt(S_2)}{\sqrt(n)}$$

or

$$\bar{Y} \pm S_2 t_{n-1}(\frac{\alpha}{2})\frac{1}{\sqrt(n)}$$

I must confess that I'm not sure yet how the second CI is derived - I'll revisit this. Let's try both

```r
s2 <- 6.41
n <- 30
yhat <- 11.3

z_alpha <- qnorm(0.95, 0, 1)
leftCI <- yhat - sqrt(s2) * z_alpha * 1/sqrt(n)
rightCI <- yhat + sqrt(s2) * z_alpha * 1/sqrt(n)


pander(data.frame(left = leftCI, right = rightCI), caption = "sample mean 95% CI - Approximated from the
```

Table 3: sample mean 95% CI - Approximated from theoretical sampling distribution

left	right
10.54	12.06

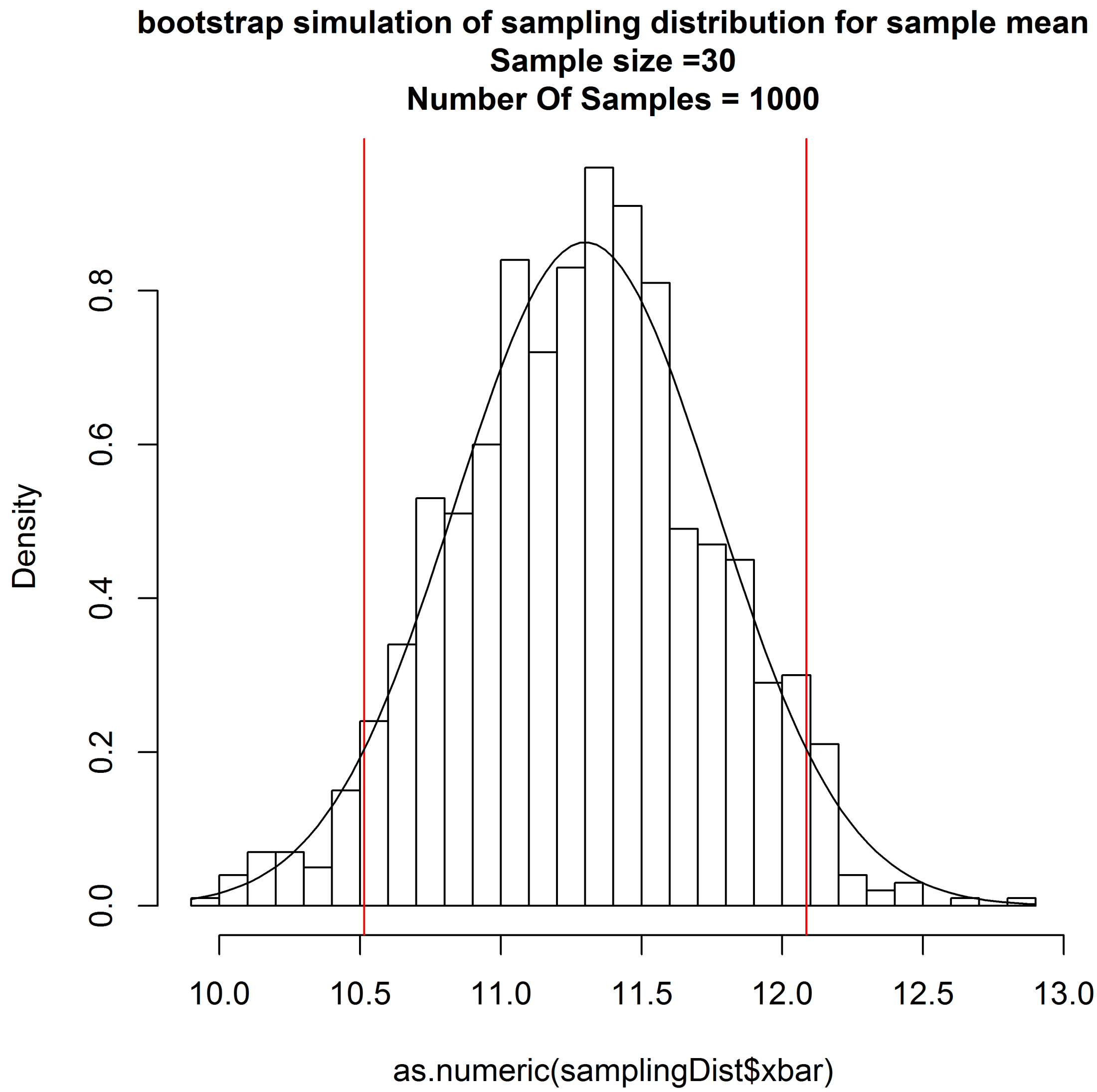
```r
t_alpha <- qt(0.95, n - 1)
leftCI <- yhat - sqrt(s2) * t_alpha * 1/sqrt(n)
rightCI <- yhat + sqrt(s2) * t_alpha * 1/sqrt(n)
pander(data.frame(left = leftCI, right = rightCI), caption = "sample mean 95% CI -  Approximated from MI
```

Table 4: sample mean 95% CI - Approximated from ML sampling distribution

left	right
10.51	12.09

We see the two CI's calculated are very close in value. Here we plot the CI bounds as calculated from the approximation to the theoretical sampling distribution on the bootstrap sampling distribution. We see there is good agreement.

```
titleString <- c("bootstrap simulation of sampling distribution for sample mean",
    paste("Sample size =", sampleSize, sep = ""), paste("Number Of Samples = ",
        numberOfSamples, sep = ""))
hist(as.numeric(samplingDist$xbar), 40, freq = FALSE, main = titleString, cex.main = 1)
abline(v = leftCI, col = "red")
abline(v = rightCI, col = "red")

curve(dnorm(x, mean = yhat, sd = 0.4622), add = TRUE)
```



Now let's calculate the 95% CI for the sample mean using the quantiles provided by the empirical distribution function of the bootsrap sampling distribution.

```
# This is one way. empCDF <- ecdf(samplingDist$xbar) empCDF(12.222)

# This is a better way
rightCI <- quantile(samplingDist$xbar, 0.975)
leftCI <- quantile(samplingDist$xbar, 0.025)


pander(data.frame(left = as.numeric(leftCI), right = as.numeric(rightCI)), caption = "sample mean 95% CI
```

Table 5: sample mean 95% CI from bootstrap quantiles

left	right
10.41	12.13

Sample Variance CI

The 95% CI for σ^2 is given by

$$(\frac{n\hat{S}_2}{\chi^2_{n-1}(\frac{\alpha}{2})}, \frac{n\hat{S}_2}{\chi^2_{n-1}(1-\frac{\alpha}{2})})$$

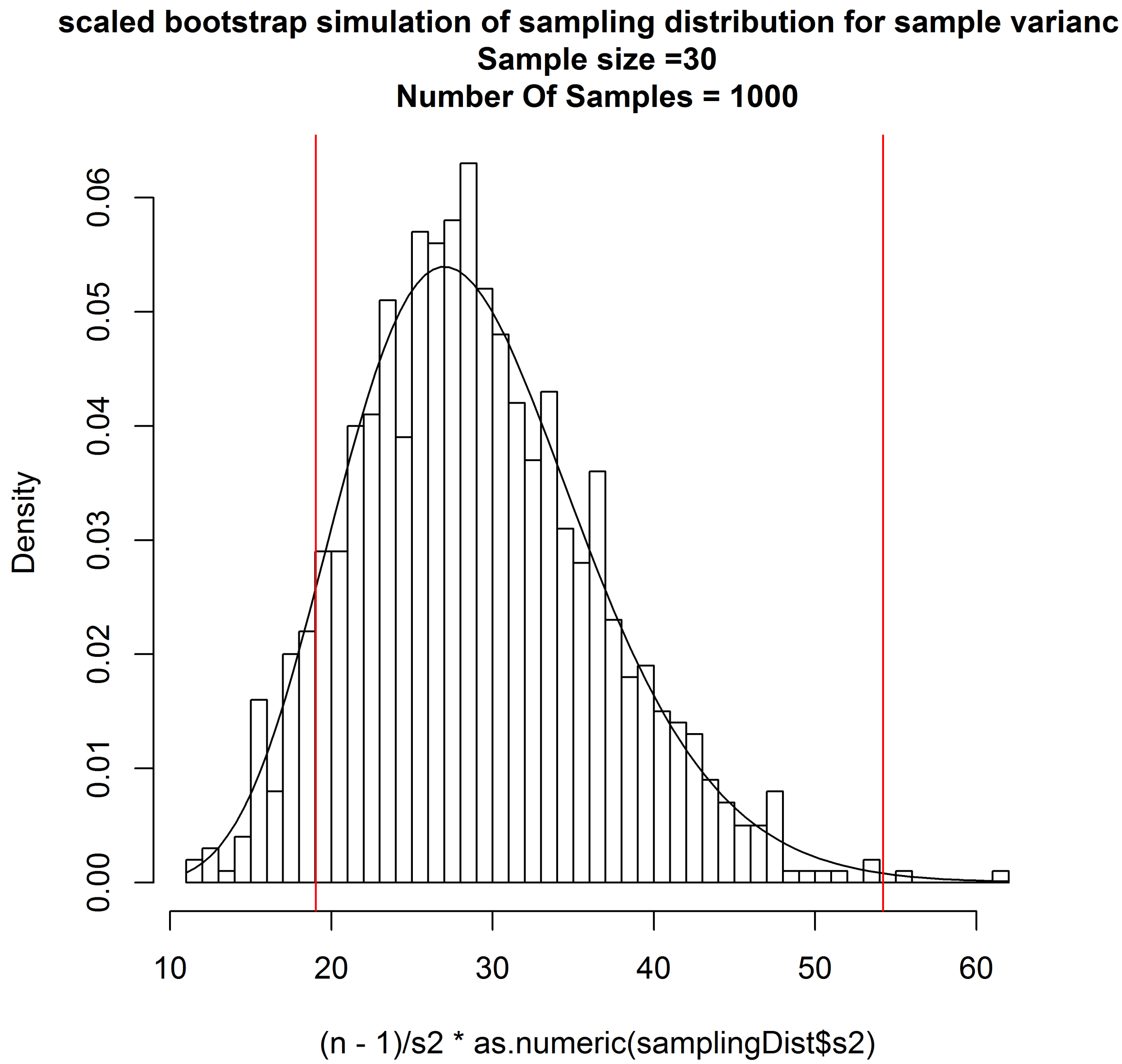
Where $\alpha = 0.05$. We can use the point estimate S__2 in forming these or we can try to use the quantiles of the bootstrap sample.

```
s2 <- 6.41
n <- 30


chi_left <- qchisq(0.025, n - 1)
chi_right <- qchisq(0.975, n - 1)

leftCI <- n * s2/chi_left
rightCI <- n * s2/chi_right

pander(data.frame(left = leftCI, right = rightCI), caption = "sample variance 95% CI - Approximated fro
```

Table 6: sample variance 95% CI - Approximated from theoretical sampling distribution

left	right
11.98	4.206

```
titleString <- c("scaled bootstrap simulation of sampling distribution for sample variance",
    paste("Sample size =", sampleSize, sep = ""), paste("Number Of Samples = ",
        numberOfSamples, sep = ""))
hist((n - 1)/s2 * as.numeric(samplingDist$s2), 40, freq = FALSE, main = titleString,
    cex.main = 1)
curve(dchisq(x, sampleSize - 1), add = TRUE)
abline(v = (n - 1)/s2 * leftCI, col = "red")
abline(v = (n - 1)/s2 * rightCI, col = "red")
```



Now let's calculate the 95% CI for the sample mean using the quantiles provided by the empirical distribution function of the bootsrap sampling distribution.

```
rightCI <- quantile(samplingDist$s2, 0.975)
leftCI <- quantile(samplingDist$s2, 0.025)


pander(data.frame(left = as.numeric(leftCI), right = as.numeric(rightCI)), caption = "sample variance 95
```

Table 7: sample variance 95% CI from bootstrap quantiles

left	right
3.535	9.961