Statisical Inference Project

James Brink

September 18, 2014

Part 1: Simulation Exercises

In this part we will create a simulation of an exponential distribution with the rexp() function in R. To do this we will do 1000 simulations of the mean of 40 exponentials using a constant lambda of 0.2 for all of our simulations.

First we will simulate our data and populate a data frame called "meanData" with the means from our simulations.

```
##set variables
set.seed(7)
lambda <- 0.2
n <- 40
sims <- 1000

##simulate means and insert into data frame
meanData <- sapply(1:sims, function(x) {mean(rexp(n, lambda))})
meanData <- as.data.frame(meanData)
names(meanData)[1] <- "x"
head(meanData)
```

```
##       x
## 1 4.968
## 2 5.185
## 3 3.801
## 4 3.992
## 5 4.465
## 6 4.694
```

Now we can look at some of the calculated attributes of our simulated distribution and compare them to their theorhetical values.

```
##the center of our simulated distribution
simDistMean <- mean(meanData$x)
simDistMean
```

```
## [1] 4.983
```

```
##the theorhetical center of our distribution
theoDistMean <- 1/lambda
theoDistMean
```

```
## [1] 5
```

We see that the distribution that we simulated is centered around 4.9833, very close to the theorhetical center of the distribution which is 5.

Next we examine the variability of our simulation distribution and compare it to the theorhetical variablity.

```
##the standard deviation of our simulated distribution
simDistSd <- sd(meanData$x)
simDistSd
```

```
## [1] 0.7611
```

```
##the theorhetical center of our distribution
theoDistSd <- (1/lambda)/sqrt(40)
theoDistSd
```

```
## [1] 0.7906
```

```
##the standard deviation of our simulated distribution
simDistVar <- var(meanData$x)
simDistVar
```

```
## [1] 0.5793
```

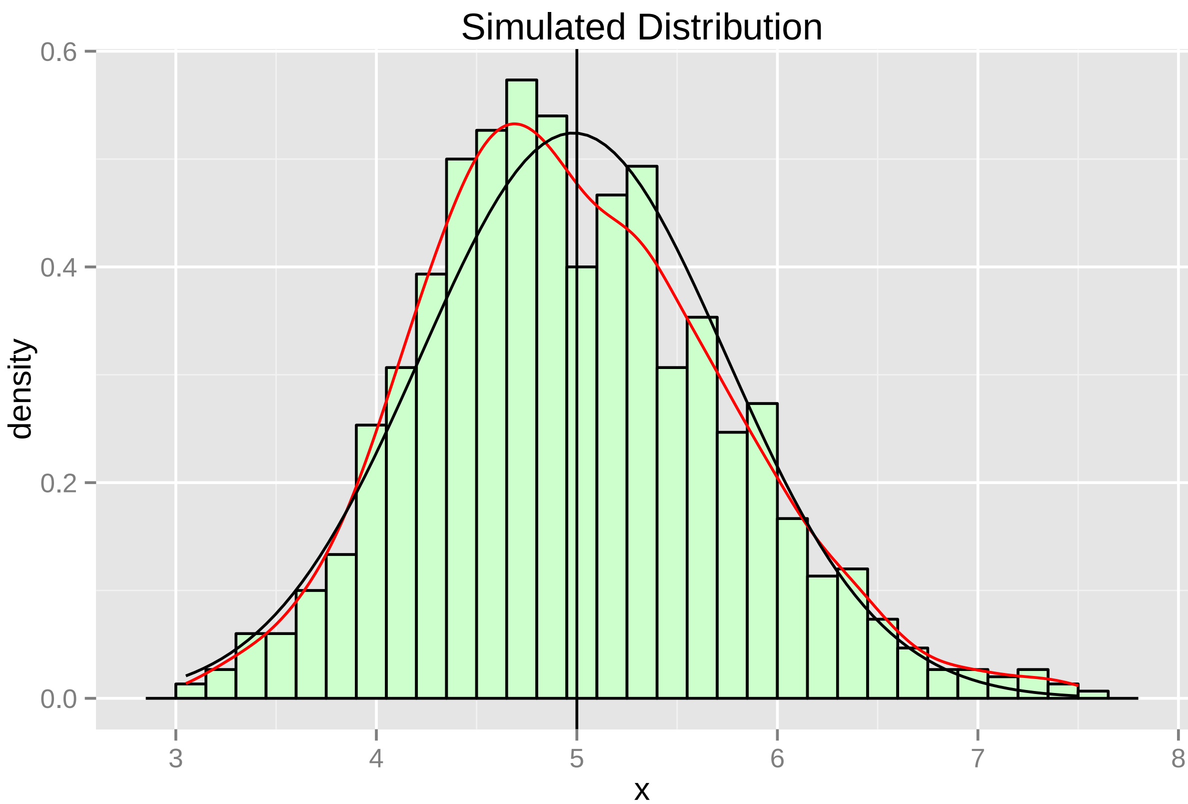
```
##the theorhetical center of our distribution
theoDistVar <- theoDistSd^2
theoDistVar
```

```
## [1] 0.625
```

Our variability measures are close to our the theoreticals as well. We see that the distribution that we simulated is has a standard deviation of 0.7611 and a variance of 0.5793, while the theorhetical standard deviation of the distribution is 0.7906 and the theorhetical variance is 0.625.

Below we examine the simulated data using a histogram and then plot a density line over top of it in red and a density line for a normal distribution over top as well in black. We can see that the simulated exponential distribution is approximately normal. We also show the simulated distribution mean as a black vertical line.

```
library(ggplot2)
g <- ggplot(meanData, aes(x=x))
g <- g + geom_histogram(aes(y=..density..), fill="#CCFFCC", color="black", binwidth=0.15)
g <- g + stat_density(color="red", geom="line", position="identity")
g <- g + stat_function(fun = dnorm, arg = list(sd = simDistSd, mean = simDistMean))
g <- g + geom_vline(aes(xintercept=theoDistMean))
g <- g + ggtitle("Simulated Distribution")
g
```



Lastly we can evaluate the coverage of +/- 1.96 standard deviations on the simulated distribution and see if it is similar to that of a normal distribution. We know that +/- 1.96 standard deviations away from the mean provides us with a 95% confidence interval on a normal distribution, lets evaluate the coverage +/-1.96 standard deviations away from our simulated mean.

```
cuts <- theoDistMean + c(-1,1)*1.96*((1/lambda)/sqrt(40))
coverage <- length(meanData[meanData$x > cuts[1] & meanData$x < cuts[2],]) / nrow(meanData)
coverage
```

```
## [1] 0.957
```

We see that 95.7% of our simulated sample means fall within 1.96 standard deviations of the population mean. This is very close to the 95% we know to be true in a normal distribution.