Exponential Distribution Simulation

Tony Bredehoeft

September 21, 2014

In this assignment we were asked to focus on two things: simulation of the exponential distribution and evaluating the ToothGrowth data set provided in R. I'll start with the simulation of the exponential distribution and highlight some properties of the simulation and how they relate to their theoretical counterparts.

I'll start by running 1000 simulations of the exponential distribution, each with 40 values and a lambda of 0.2. The theoretical mean and standard deviation of the exponential distribution are both 1/lambda, meaning the theoretical mean and standard deviation are both 5 in our simulation.

```
set.seed(1)
exp.data <- data.frame()
i = 1
while(i <= 1000){
  exp.data <- rbind(exp.data,rexp(40,.2))
  i = i + 1
}
```

Now we have our 1000 simulations of the exponential distribution in a new data frame. Next we'll look at the sample means from each of our simulations, take the mean of these sample means and compare that value to the theoretical mean of 5.

```
mean(rowMeans(exp.data))
```

```
## [1] 4.99
```

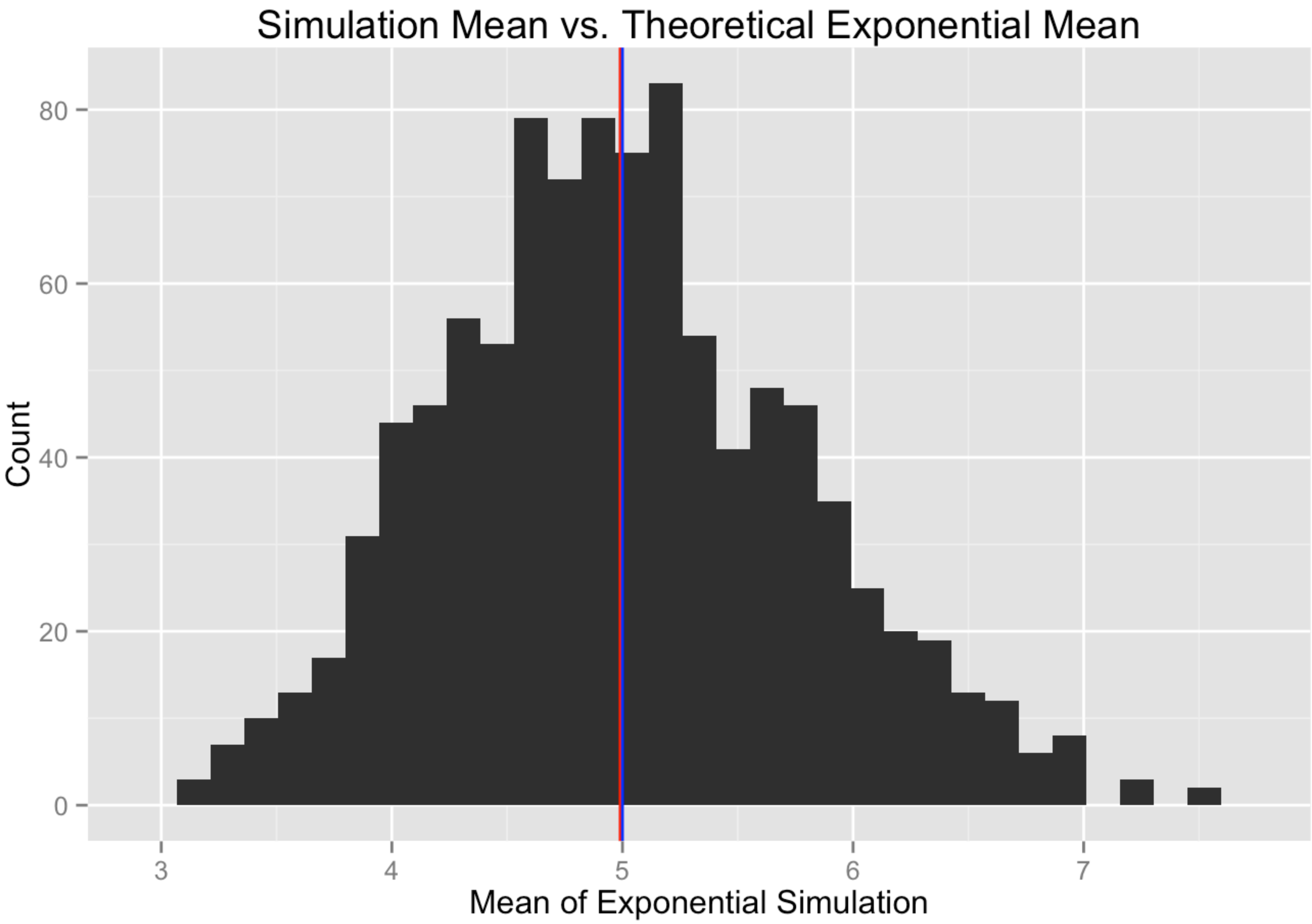
Here we can see the mean value provided by our simulations is 4.99, which is very close to the theoretical value.

Next I'll plot a histogram of the sample means provided by each simulation, and show how the simulation mean compares to the theoretical mean of 5.

Note that the theoretical mean is the blue line and the mean of the sample means is the red line below.

```
library(ggplot2)
ggplot(exp.data,aes(x=rowMeans(exp.data))) +
  geom_histogram() +
  geom_vline(xintercept = mean(rowMeans(exp.data)),color='red') +
  geom_vline(xintercept = 5,color='blue') +
  xlab('Mean of Exponential Simulation') +
  ylab('Count') +
  ggtitle('Simulation Mean vs. Theoretical Exponential Mean')
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



We can see that both the simulation and theoretical means are very close (in fact the lines overlap) and that the distribution of the sample means approximates a normal distribution.

Next let's take a look at how the variance of the sample means compares to the theoretical variance.

```
var(rowMeans(exp.data))
```

```
## [1] 0.6111
```

```
(1/.2)^2/40
```

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

Here we can see that the variance of the sample means of our simulation is slightly less than the theoretical variance provided by the formula (1/lambda)^2/n.

Now we want to evaluate the coverage provided by our simulation. As in the class lectures, for each simulation in our 1000 simulations I will calculate the confidence interval for the sample mean.

Next I will determine what percentage of these confidence intervals contain the theoretical mean of 5. Given we are looking at a 95% confidence interval for the sample means, the coverage value should be around .95.

```
coverage <- apply(exp.data,1,function(x) {
  ll <- mean(x) - 1.96*5/sqrt(40)
  ul <- mean(x) + 1.96*5/sqrt(40)
  mean(ll<5 & ul>5)
})
sum(coverage)/1000
```

```
## [1] 0.949
```

Here we can see that 0.949 of our sample confidence intervals (95%) contained the true population mean.