Simulating Exponentials to Investigate CLT

jcollingosch

February 21, 2015

Overview

We will use simulations of random exponential variables in order to investigate the claims of the Central Limit Theorem. For this example we are going to set the rate parameter for the exponential distribution, lambda, to be 0.2. The mean and standard deviation of the exponential distribution is 1/lambda. Therefore, our **theoretical distribution has a mean and standard devation of 5 and a variance of 25**. Below we will compare these values to the sampling distribution.

```
# set up parameters
lambda = 0.2
mu = 1/lambda
sd = 1/lambda
var = 1/lambda^2
```

Simulations

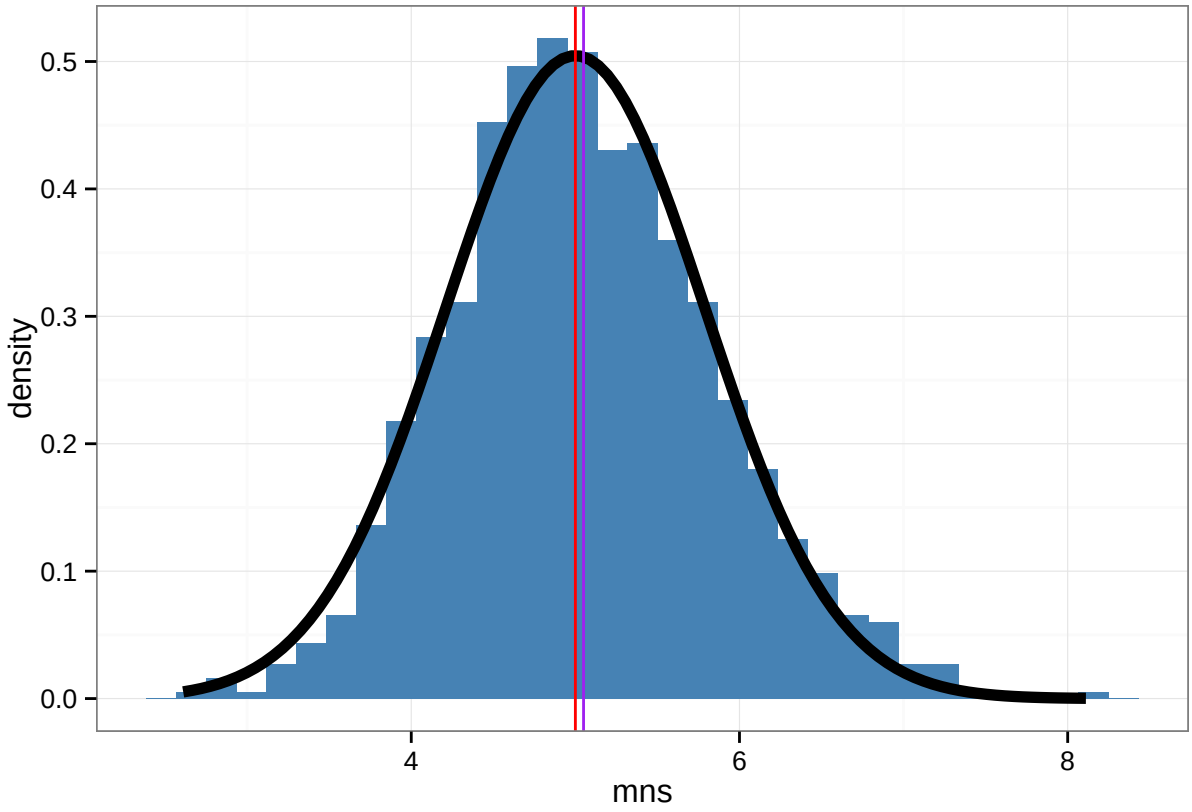
We will take the average of 40 exponentials 1000 times to build our sampling distribution. Then we will explore the relationship between the sampling distribution and the theoretical distribution.

```
set.seed(333)
n = 40
mns = NULL
for(i in 1:1000) {
  mns <- c(mns, mean(rexp(n = n, rate = lambda)))
}

sample_mu <- round(mean(mns),2)
sample_var <- round(var(mns),2)
```

```
library("ggplot2")

ggplot(data.frame(mns), aes(mns)) +
  geom_histogram(aes(y = ..density..), fill="steelblue") +
  stat_function(fun = dnorm,
                args = list(mean=mu,sd=sd/sqrt(n)),
                size = 2) +
  geom_vline(xintercept = mu, color="red") +
  geom_vline(xintercept = sample_mu, color="purple") +
  theme_bw()
```



Sample Mean vs Theoretical Mean

The sample mean of this sampling distribution is: 5.05 compared to a theoretical mean of 5. The figure above shows a histogram of the 1000 samples. The red line highlights the center of the sampling distribution centered at the sample mean. Overlaying the histogram is a normal density with a mean and standard deviation of 5. The purple line above highlights the center of the normal density at a value of 5. This matches what we learned in class that the sample mean in this case is **consistent** as it converges to the theoretical mean.

Sample Mean vs Theoretical Mean

The sample variance of this sampling distribution is: 0.64 compared to a theoretical variance of $25/40 = 1$. The variance from the exponential distribution with rate parameter of 0.2 is 25 and we know that the variance of the sampling distribution should be the variance/n. Therefore, in this case we show that the variance of the sampling distribution is much smaller than the original distribution and is approximately 25/40.

Distribution

As shown in the figure above, the sampling distribution (blue histogram) closely mirrors the normal distribution (black line) with parameters mean = 5 sd = 5.