Statistical Inference Course Project - Part 1

FChemin

November 2, 2016

Overview

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Comparison of mean of the distribution

First, we would like show the theoretical mean of the distribution and compare it to the sample mean of the distribution.

```
set.seed(1234)

lambda <- 0.2
n <- 40
mns <- NULL

for (i in 1:1000) mns <- c(mns, mean(rexp(n, lambda)))

sample_mean <- mean(mns)
theoretical_mean <- 1/lambda
```

The sample mean is equal to 4.9742388 and the theoritical mean is equal to 5. That's pretty close!

Comparison of variance of the distribution

Secondly, we would like to show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

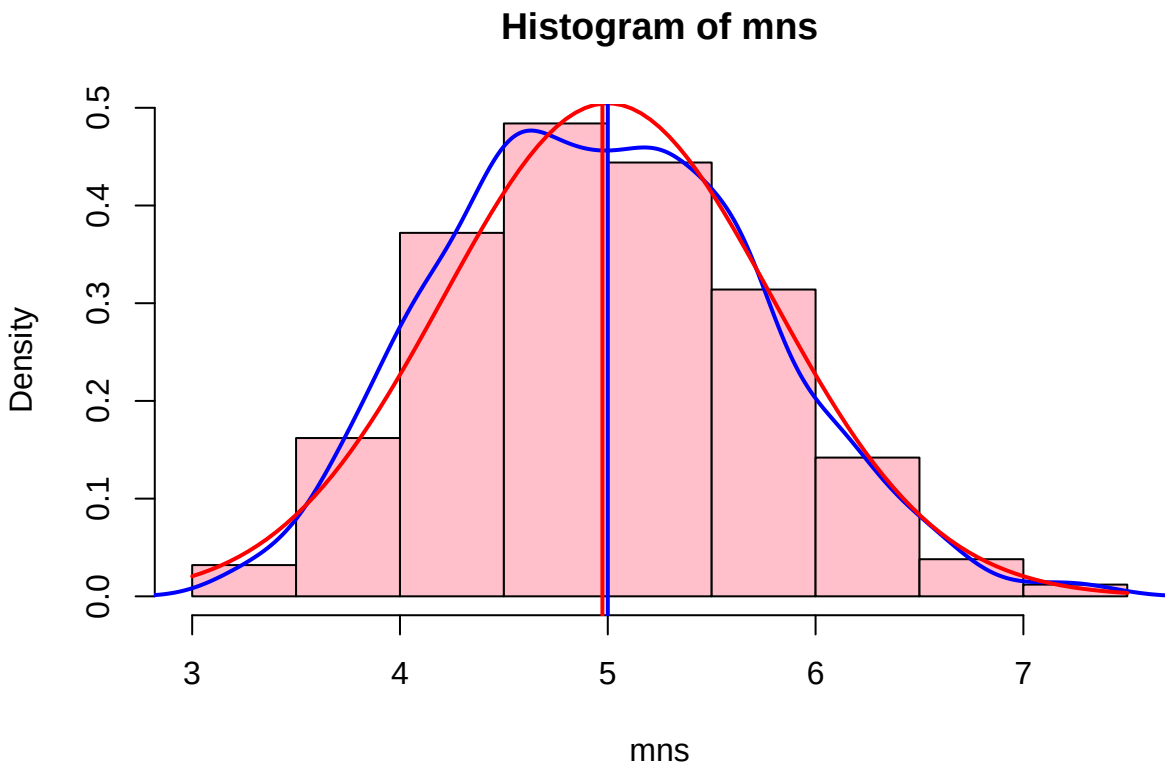
```
theoretical_var <- ((1/lambda)/sqrt(n))^2

sample_var <- var(mns)
```

The sample variance is equal to 0.5706551 and the theoritical variance is equal to 0.625. That's pretty close!

Proof of normality of the distribution

Now, we must show that the distribution is approximately normal.

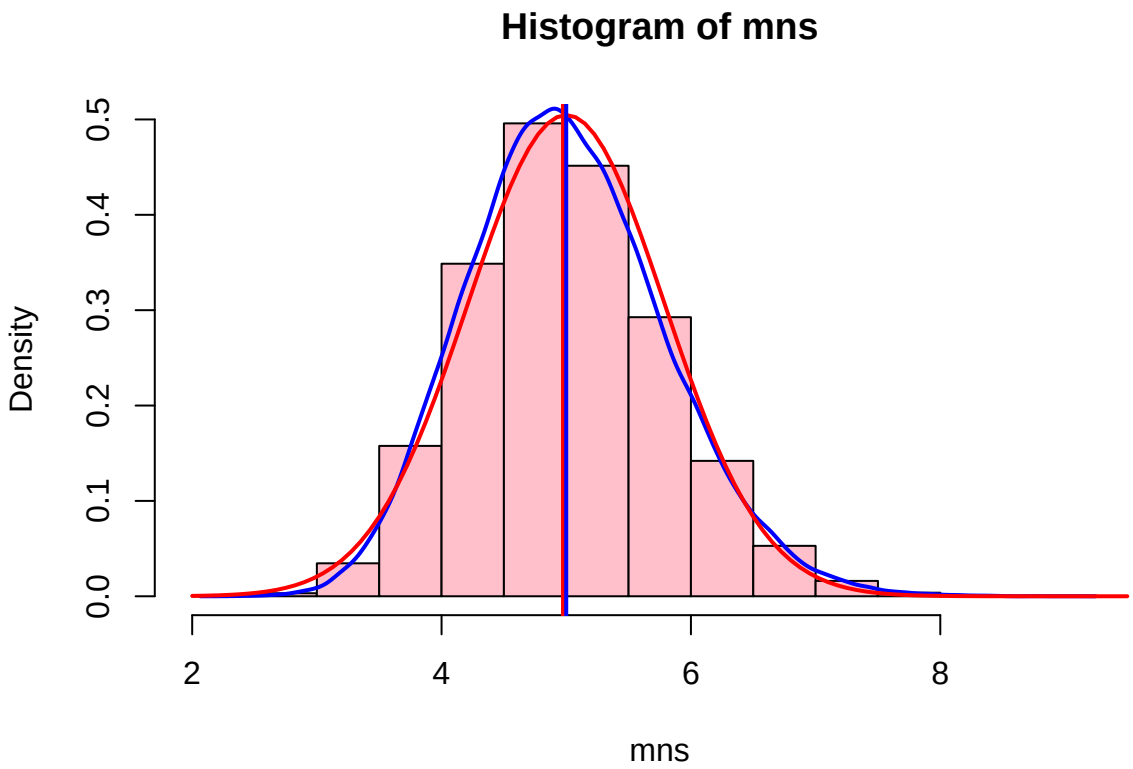
```r
hist(mns, freq = FALSE, col = "pink", breaks = 10)
lines(density(mns), col="blue", lw=2)
curve(dnorm(x, mean = 5, sd = 5/sqrt(n)), add = TRUE, col = "red", lw=2)
abline(v = sample_mean, col="red", lw=2)
abline(v = 5, col="blue", lw = 2)
```



In this plot, we compared our sample (1000 averages of 40 random exponentials) of the exponential distribution to the normal distribution. In blue, the sample distribution mean and density curve and in red, the ones for a normal distribution. That's pretty close!

What would happens if we take a greater sample of our exponential distribution?

```r
set.seed(314)

for (i in 1:100000) mns <- c(mns, mean(rexp(n, lambda)))

hist(mns, freq = FALSE, col = "pink", breaks = 10)
lines(density(mns), col="blue", lw=2)
curve(dnorm(x, mean = 5, sd = 5/sqrt(n)), add = TRUE, col = "red", lw=2)
abline(v = sample_mean, col="red", lw=2)
abline(v = 5, col="blue", lw = 2)
```



In this plot, we compared our sample (100000 averages of 40 random exponentials) of the exponential distribution to the normal distribution. In blue, the sample distribution mean and density curve and in red, the ones for a normal distribution. **We're getting closer and closer to the normal distribution as our sample size is increasing**.