Comparing Exponential Distribution with CLT

Mohan

January 24, 2015

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. For this assignment lambda will be 0.2 for all of the simulations. We will investigate the distribution of averages of 40 exponentials for 1000 simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

1. Show the sample mean and compare it to the theoretical mean of the distribution.

```
lambda = 0.2
n = 40
sim = 1:1000
set.seed(12345)
sampleMeans <- data.frame(x = sapply(sim, function(x) {mean(rexp(n, lambda))}))
head(sampleMeans)
```

```
##       x
## 1 5.554
## 2 4.671
## 3 4.256
## 4 4.614
## 5 5.304
## 6 3.627
```

```
sampleMean <- mean(sampleMeans$x)
theoreticalMean <- 1/lambda
sampleMean
```

```
## [1] 4.972
```

```
theoreticalMean
```

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

Sample mean: 4.972

Theoretical mean: 5

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
variance <- sd(sampleMeans$x)
theoreticalVariance <- (1/lambda)/sqrt(40)
variance
```

```
## [1] 0.7716
```

```
theoreticalVariance
```

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

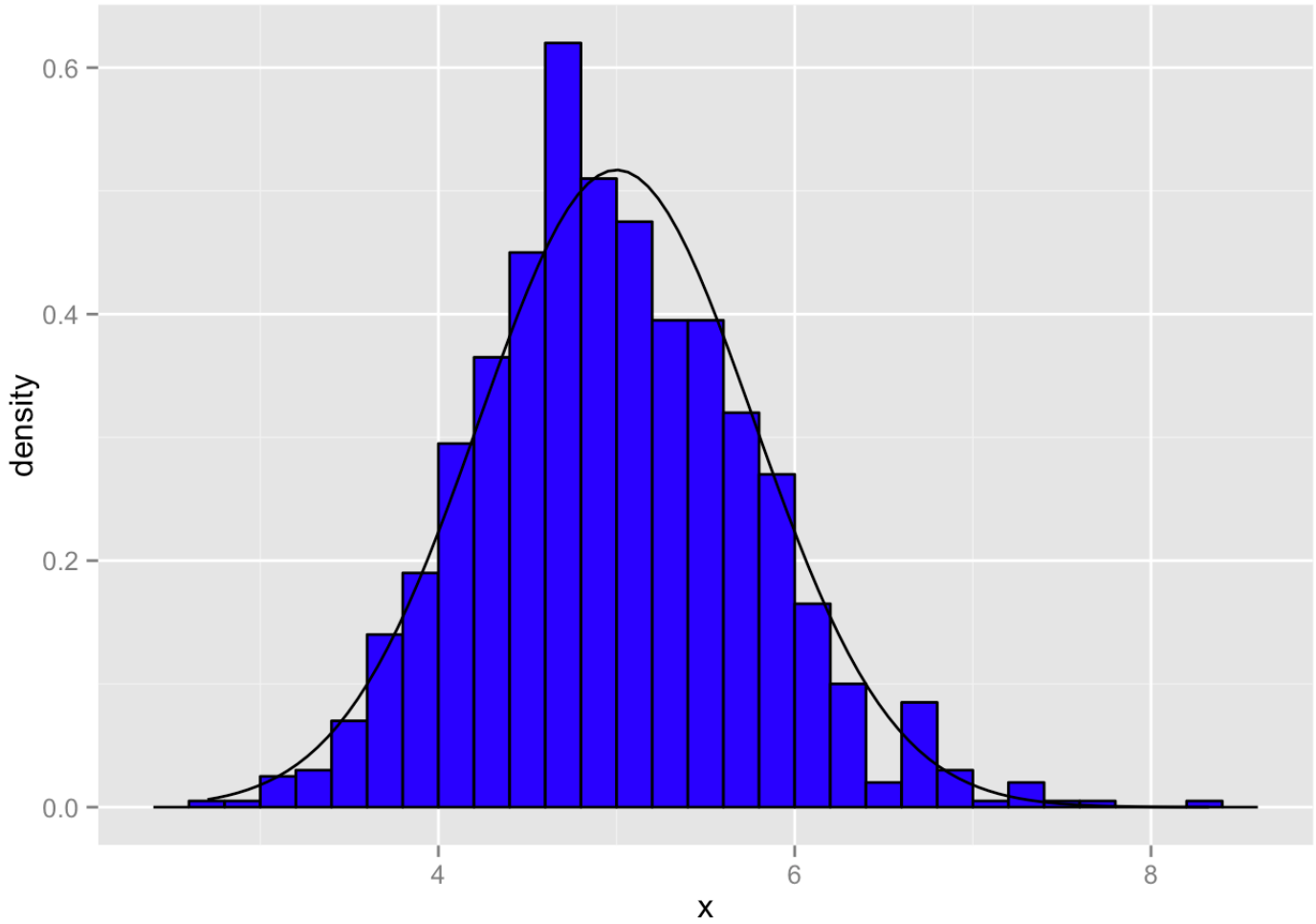
Variance of the distibution: 0.7716

Theoretical Variance : 0.7906

3. Show that the distribution is approximately normal.

Below is a histogram plot of the means of the 1000 simulations of rexp(n, lambda). It is overlaid with a normal distribution with mean 5 and variance 0.7716.

```
library(ggplot2)
ggplot(data = sampleMeans, aes(x = x)) +
    geom_histogram(aes(y=..density..), fill = I('blue'),
                   binwidth = 0.20, color = I('black')) +
    stat_function(fun = dnorm, arg = list(mean = theoreticalMean, sd = variance))
```



The distribution of our simulations appears normal.

Focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

```
sampleMean + c(-1,1)*1.96*variance/sqrt(nrow(sampleMeans))
```

```
## [1] 4.924 5.020
```