Statistical Inference Coursera Project

Nick Horne

25 September 2017

Overview

In this project you 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 1000 simulations.

```
# Seed for reproducable research
set.seed(2211)

# Taken from question
lambda <- 0.2
simSize <- 1000
distSize <- 40

# Get our simulations
simulations <- replicate(simSize, rexp(distSize, lambda))
```

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

```
# Apply means on the columns
means <- apply(simulations, 2, mean)

# Sample mean
sampleMean <- mean(means)
sampleMean
```

```
## [1] 4.9931
```

```
# Theoretical mean
theoreticalMean <- 1/lambda
theoreticalMean
```

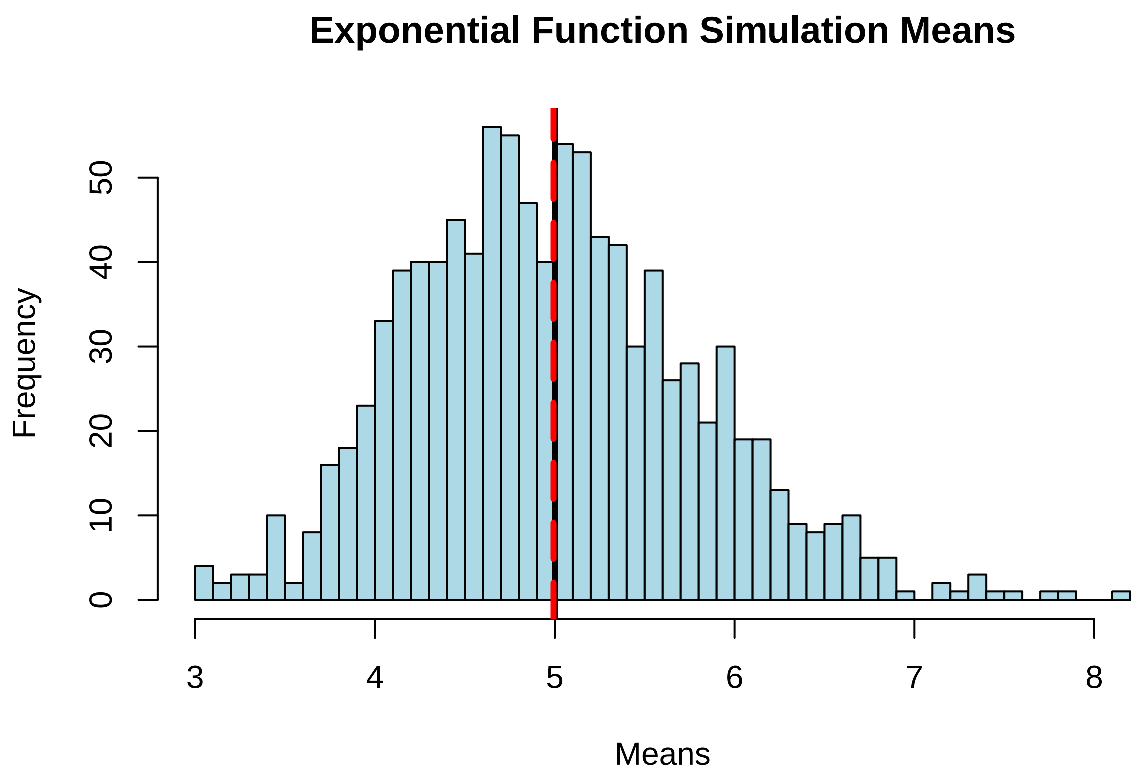
```
## [1] 5
```

Our sample mean 4.993 is very close to the theoretical mean 5. This small difference can also be seen on the graph below:

```
# Draw the means on the graph
hist(means,
     breaks=distSize,
     col="lightblue",
     main="Exponential Function Simulation Means",
     xlab="Means")

# Add the theoretical mean
abline(v=theoreticalMean, lwd=3)
```

```r
# Overlay the sample mean
abline(v=sampleMean, lwd=3, col="red", lty=5)
```



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

```r
# Sample variance
sampleVariance <- var(means)
sampleVariance
```

```
## [1] 0.6353648
```

Population Variance $= \frac{\sum_{i=1}^{n}(x-\mu)^2}{n}$

```r
# Theoretical variance
theoreticalVariance <- ((1/lambda) ^ 2) / distSize
theoreticalVariance
```

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

Our sample variance 0.635 is close to the theoretical variance 0.625 also.

3. Show that the distribution is approximately normal.

The distrubition of avarages is fairly close. This can be seen on the graph with the sample mean density line (red dotted) being visible close to the normal line. This is due to the central limit theorum which states that:

A sufficiently large sample size from a population with a finite level of variance, the mean of all samples from the same population will be approximately equal to the mean of the population.

```
# Plot the density
hist(means,
     probability=TRUE,
     breaks=distSize,
     col="lightblue",
     main="Exponential Function Simulation Means",
     xlab="Means")

x <- seq(min(means), max(means), length=100)
y <- dnorm(x, mean=1/lambda, sd=(1/lambda/sqrt(distSize)))
lines(x, y, lwd=3)

# Overlay the density of the means
lines(density(means), lwd=3, col="red", lty=5)
```