Statistical Inference Course Project - Part 1

Ken Wood

8/19/2020

Introduction

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 the expression `rexp(n, lambda)` where `lambda` is the rate parameter. The mean of the exponential distribution is `1/lambda` and the standard deviation is also `1/lambda`. We will set $\lambda = 0.2$ for all of the simulations and we will investigate the distribution of averages of 40 exponentials. Note that we will need to do a thousand simulations.

Our goal is to illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. Specifically, we will:

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution of the means is approximately normal.

In deliverable 3, we will 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.

Analysis

First, let's plot the exponential distribution for 1000 observations using $\lambda = 0.2$. The theoretical mean for this distribution is $\mu = 1/\lambda = 1/0.2 = 5$. The variance, $\sigma^2 = 1/\lambda^2 = 25$.

```
exp_dist <- rexp(1000,rate = 0.2)
hist(exp_dist,breaks=20,col="blue",xlab = "x",main="Histogram of an Exponential Distribution - 1000 Obs
abline(v=5,col="red",lwd=3)
text(10, 220, "mean(x) = 5",col="red")
```

Histogram of an Exponential Distribution – 1000 Observations
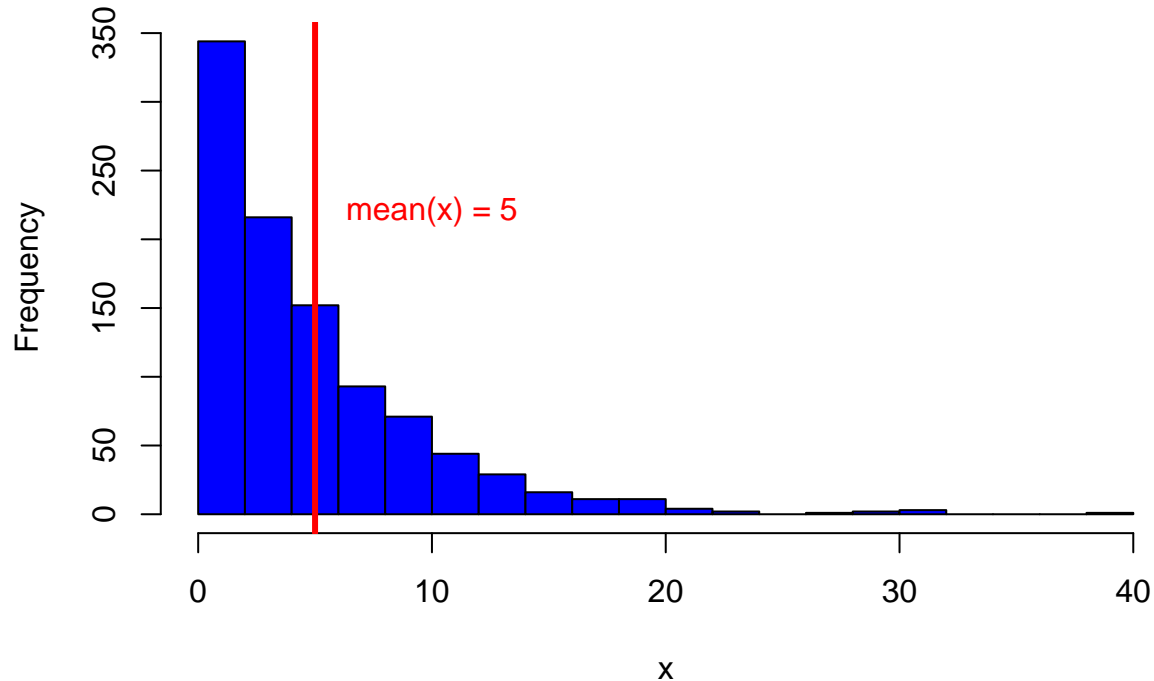

```r
mean(exp_dist) # get mean of distribution
```

```
## [1] 4.799936
```

```r
var(exp_dist)   # get variance of distribution
```

```
## [1] 23.40338
```

By using the `rexp` function in R, the simulations produce the above histogram with sample $\mu = 5$ and $\sigma^2 = 25$. If we were to increase number of simulations, we would get simulated values even closer to the theoretical ones.

Now, we will simulate 1000 trials of the exponential distribution using 40 observations for each trial and plot a histogram of the means for each trial.
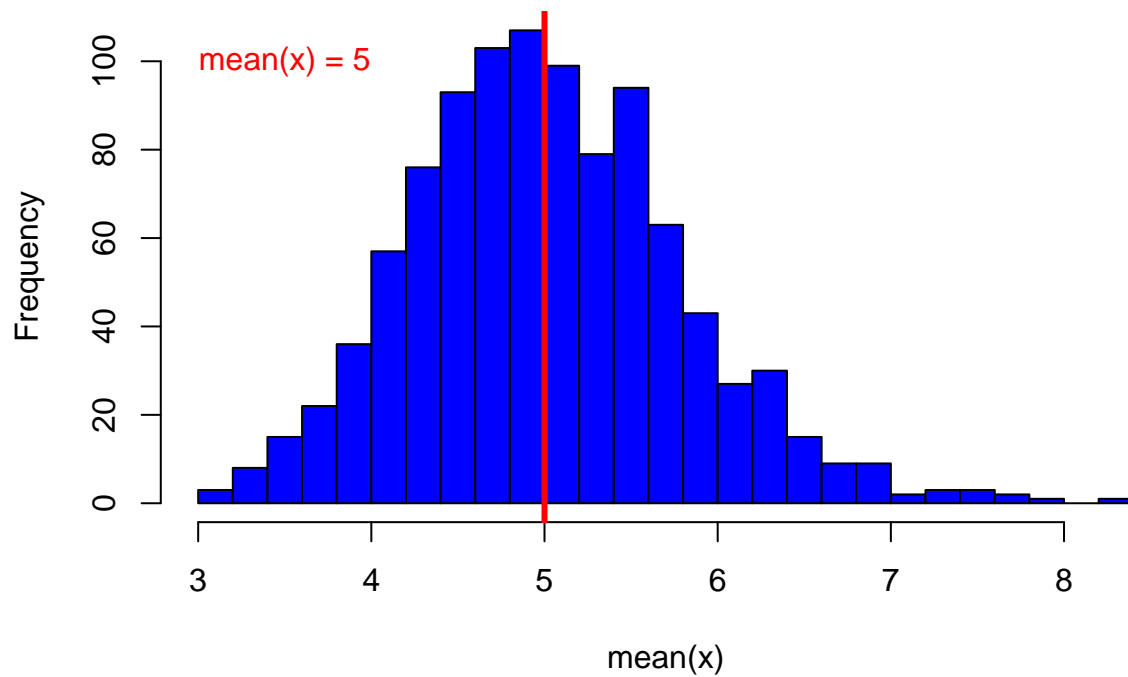
```r
mnsexp = NULL
for (i in 1 : 1000) mnsexp = c(mnsexp, mean(rexp(40,rate=0.2)))
hist(mnsexp,breaks=20,col="blue",xlab="mean(x)",main="Exponential Distribution Means - 1000 Trials of 40
abline(v=5,col="red",lwd=3)
text(3.5,100, "mean(x) = 5",col="red")
```

Exponential Distribution Means – 1000 Trials of 40 Observations


```r
mean(mnsexp) # get mean of mean distribution
```

```
## [1] 5.016336
```

```r
var(mnsexp)   # get variance of mean distribution
```

```
## [1] 0.6073751
```

We see from the histogram that the means are normally distributed with theoretical mean $\mu = 1/\lambda = 5$ and theoretical variance $\sigma^2/n = 1/(\lambda^2 n) = (25/40) = 0.625$. Our simulated values are quite close the theoretical ones.