Statistical Inference Course Project Part 1

Jessica

01/02/2021

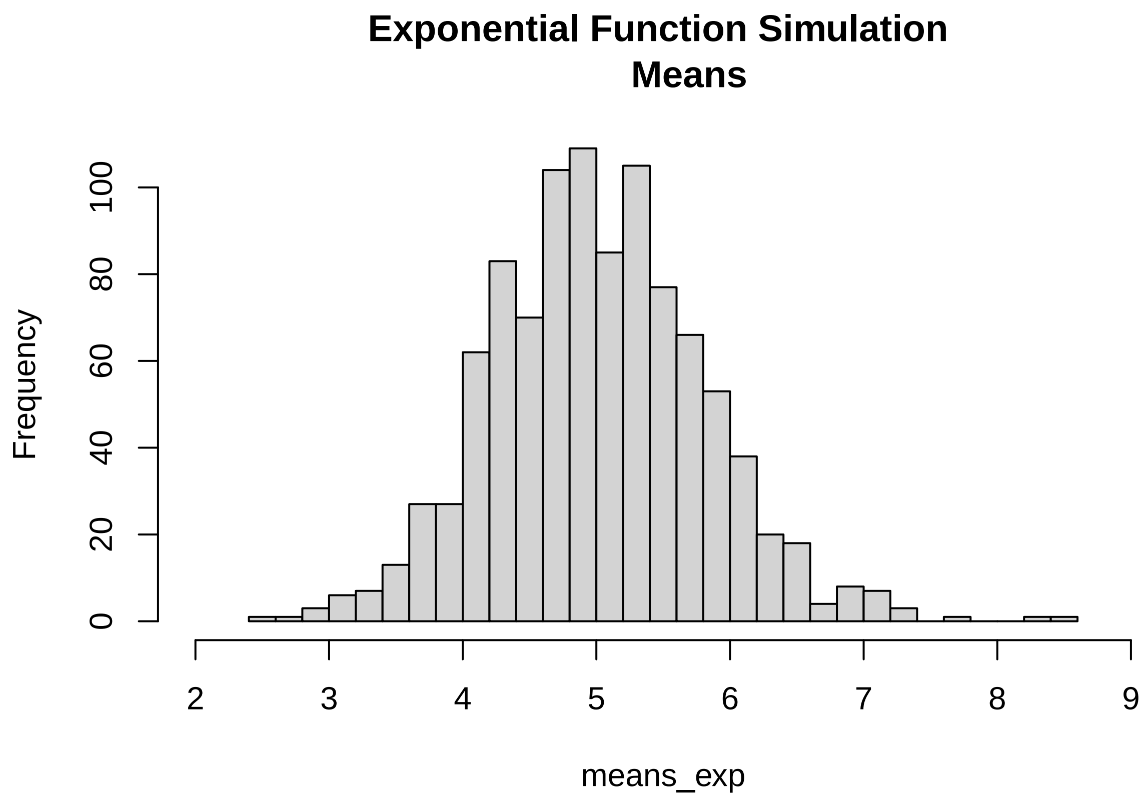
Overview:

- We will investigate the exponential distribution in R and compare it with the Central Limit Theorem. We will investigate the distribution of means of 40 exponentials setting lambda = 0.2 for all simulations.

Simulations:

- The code here runs 1,000 simulations of the distribution of averages of 40 exponentials. Lambda is set at 0.02 according to the course project instructions.
- We created a "means" vector which contains the mean of each simulation i:1000.
- We plotted the means of the simulated exponential distributions in a histogram.

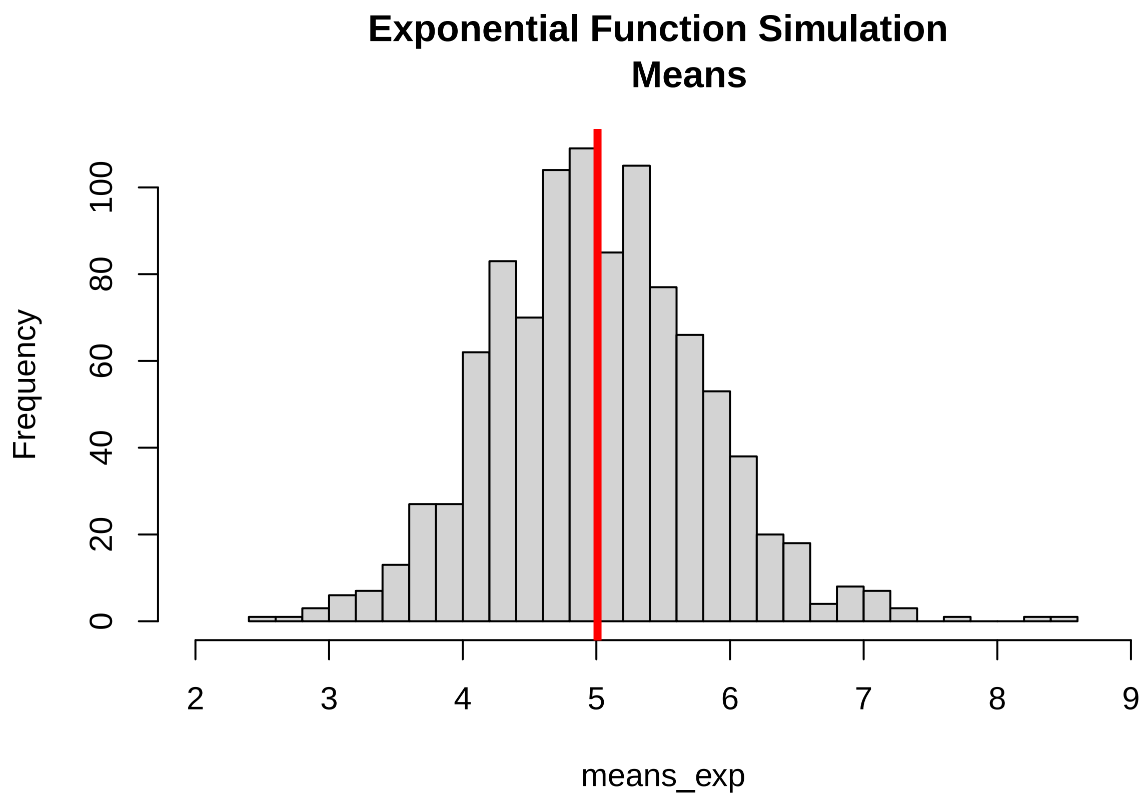
```
# set seed for reproducability
set.seed(2021)

# Set sampling values as described in the project instructions
lambda <- 0.2   # lambda

n <- 40         # number of exponentials

sims <- 1000    # number of simulations

#Run simulations
sim_exp <- replicate(sims, rexp(n, lambda))

#Calc the means of the exponential simulations
means_exp <- apply(sim_exp, 2, mean)

#Histogram of the means
hist(means_exp, breaks=40, xlim = c(2,9), main="Exponential Function Simulation
     Means")
```



Sample Mean versus Theoretical Mean:

- The mean or expected value of an exponentially distributed random variable X with rate parameter λ is given by: $E[X] = 1/\lambda$.
- Therefore, the expected mean of an exponentially distributed random variables with rate parameter $0.02 = 1/0.2 = 5$.

```r
# Histogram of the means
hist(means_exp, breaks=40, xlim = c(2,9), main="Exponential Function Simulation
     Means")

# Vertical line at mean
abline(v = mean(means_exp), lwd = "4", col = "red")
```




```r
# Calculate the mean of the simulation means
mean(means_exp)
```

```
## [1] 5.008639
```

- **The sample mean is 5.01, which is very close to the theoretical mean of 5.**

Sample Variance versus Theoretical Variance:

- The variance of X is given by $Var[X] = 1/\lambda^2/n$
- The standard deviation of X is given by $1/\lambda/\sqrt{n}$

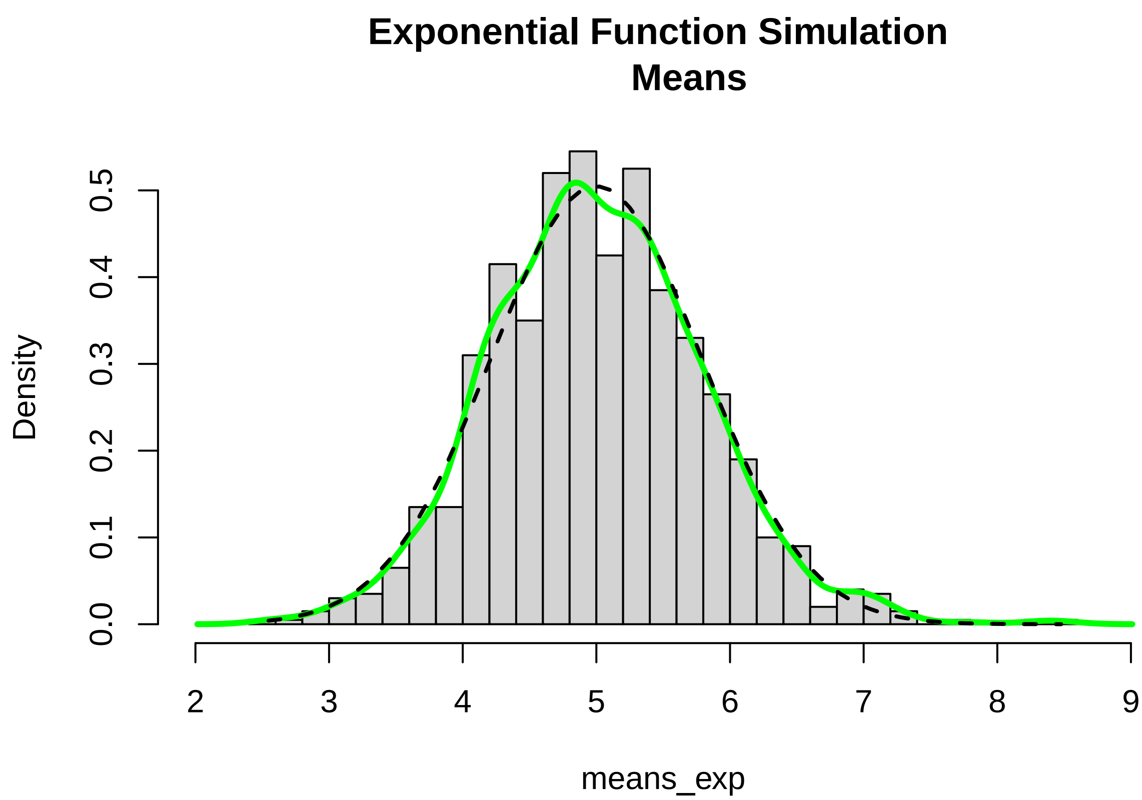
```r
# Calculate theoretical variance
theor_var <- round(1/.2^2/(n), 2)
sample_var <- round(var(means_exp), 2)

theor_sd <- round(1/.2/sqrt(n), 2)
sample_sd <- round(sd(means_exp), 2)
```

- **The theoretical variance is 0.62 and the sample variance is 0.63.**
- **The theoretical standard deviation is 0.79 and the sample standard deviation is 0.79**
- The sample variance and standard deviation are almost identical to the theoretical variance and standard deviation.

Distribution:

- Here we will investigate if the exponetial distribution is approximately normal.
- We know based on the Cetnral Limit Theorem that the means of the sample simulations should be normally distributed.

```
# Histogram of the sample means with a density line
hist(means_exp, prob = TRUE, breaks=40, xlim = c(2,9), main="Exponential Function Simulation
     Means")
lines(density(means_exp), lwd = 3, col = "green")

# Overlay a line that represents the normal distribution
x <- seq(min(means_exp), max(means_exp), length = 2*n)
y <- dnorm(x, mean = 1/lambda, sd = sqrt(((1/lambda)/sqrt(n))^2))
lines(x,y, pch = 20, lwd = 2, lty = 2)
```



- This graph shows that the distribution of the means of the sampled exponential distrituions (green line) appear to be normally distributed (dotted line).