

Statistical Inference Course Project Part 1

Gaurab Kundu

2022-10-28

Part 1: Simulation Exercise

Overview

In this part of the project I 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. I will investigate the distribution of averages of 40 exponentials. I will be doing a thousand simulations.

I will illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

Questions

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 is approximately normal.

In point 3, I 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.

Simulations

It is good practice to set the seed of R's random number generator before simulating random observations so that the simulations can be reproduced.

```
set.seed(1234)
```

Set the rate parameter (`lambda`) equal to 0.2, set the number of exponentials equal to 40, and set the number of simulations equal to 1000 as instructed in the rubric.

```
lambda <- 0.2  
n <- 40  
sim <- 1000
```

Simulate the 40 exponentials.

```
simExp <- replicate(sim, rexp(n, lambda))
```

Calculate the mean of the simulated exponentials.

```
meanExp <- apply(simExp, 2, mean)
```

Question 1: Sample Mean Versus Theoretical Mean

Calculate and compare the sample mean to the theoretical mean. The mean of the exponential distribution is equal to $1/\lambda$. In our case, λ is equal to 0.2. Thus, the theoretical mean should be approximately 5.

```
sampMean <- mean(meanExp)
sampMean
```

```
## [1] 4.974239
```

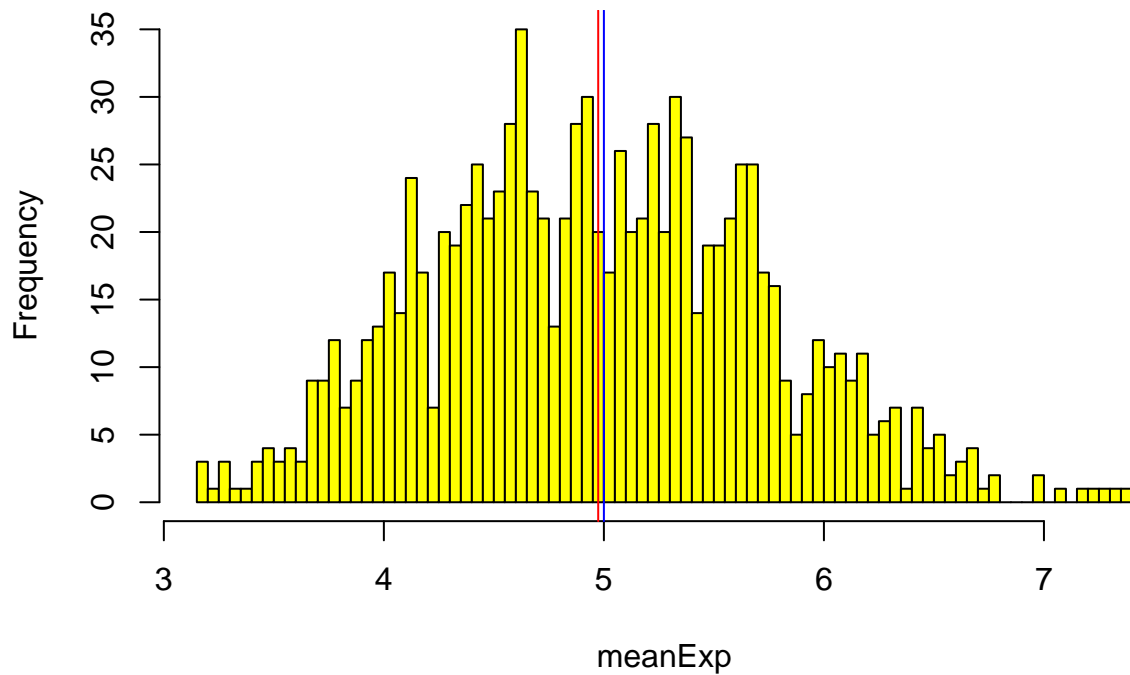
```
theoMean <- 1/lambda
theoMean
```

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

The below figure depicts a histogram of the simulated exponential sample means. The red vertical line indicates the sample mean and the blue vertical line indicates the theoretical mean.

```
hist(meanExp, main = "Simulated Exponential Sample Means", col = "yellow", breaks = 100)
abline(v = sampMean, col = "red")
abline(v = theoMean, col = "blue")
```

Simulated Exponential Sample Means



The R code above shows that the sample mean is equal to 4.974 which is very close to the theoretical mean of 5.

Question 2: Sample Variance Versus Theoretical Variance

Calculate and compare the sample variance to the theoretical variance. The standard deviation of the exponential distribution is equal to $(1/\lambda)/\sqrt{n}$. Square the standard deviation to calculate the variance.

```
sampSd <- sd(meanExp)
sampVar <- sampSd^2
sampVar
```

```
## [1] 0.5706551
```

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

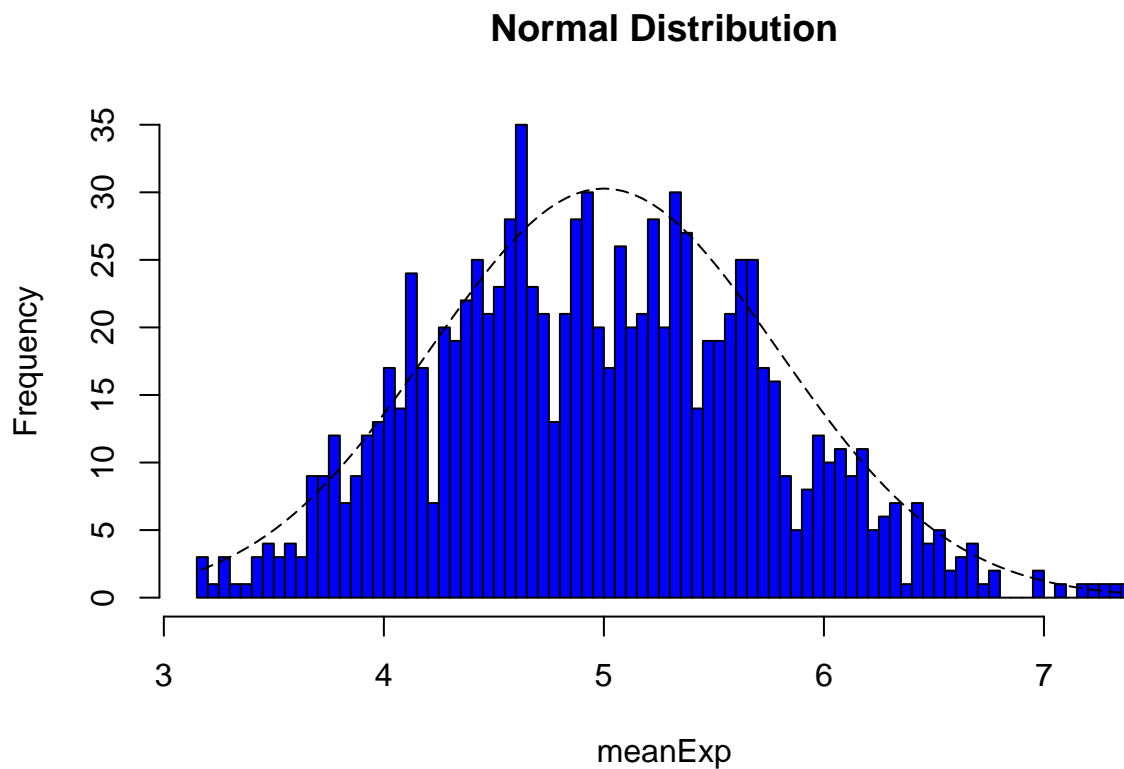
```
## [1] 0.625
```

The R code above shows that the sample variance is equal to 0.571 which is very close to the theoretical variance of 0.625.

Question 3: Distribution

Finally, I'll investigate whether the exponential distribution is approximately normal. Because of the Central Limit Theorem, the means of the sample simulations should follow a normal distribution.

```
hist(meanExp, main = "Normal Distribution", col = "blue", breaks = 100)
xfit <- seq(min(meanExp), max(meanExp), length = 100)
yfit <- dnorm(xfit, mean = 1/lambda, sd = (1/lambda)/sqrt(n))
lines(xfit, yfit*60, lty = 5)
```



As the above figure shows, the distribution of means of the sampled exponential distributions appear to follow a normal distribution. This is due to the Central Limit Theorem. It is noteworthy that increasing the number of samples from 1000, would cause the distribution to align even more closely to a normal distribution.