Statistical Inference Course Project - Part 1

pvv88

Overview

The first part of statistical inference course project is to explore the exponential distribution in R and compare it with the Central Limit Theorem. This will involved simulation to explore inference and perform some statistical inferential works.

Simulation

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 a thousand simulations.

Load libraries and set lambda, exponentials, seed, and simulation variables

```
library(knitr)
library(ggplot2)

set.seed(200)
lambda = 0.2  # lambda
expo <- 40    # no of exponentials
noofsim <- 1000 # no of simulations
```

Run simulations and calculate mean of exponentials

```
#sim_exp <- replicate(noofsim, rexp(expo,lambda))
sim_exp <- matrix(rexp(noofsim*expo, rate=lambda), noofsim, expo)
mean_exp <- rowMeans(sim_exp)
```

Sample Mean versus Theoretical Mean

```
sample_mean <- round(mean(mean_exp), 2) #Sample Mean
sample_mean
```

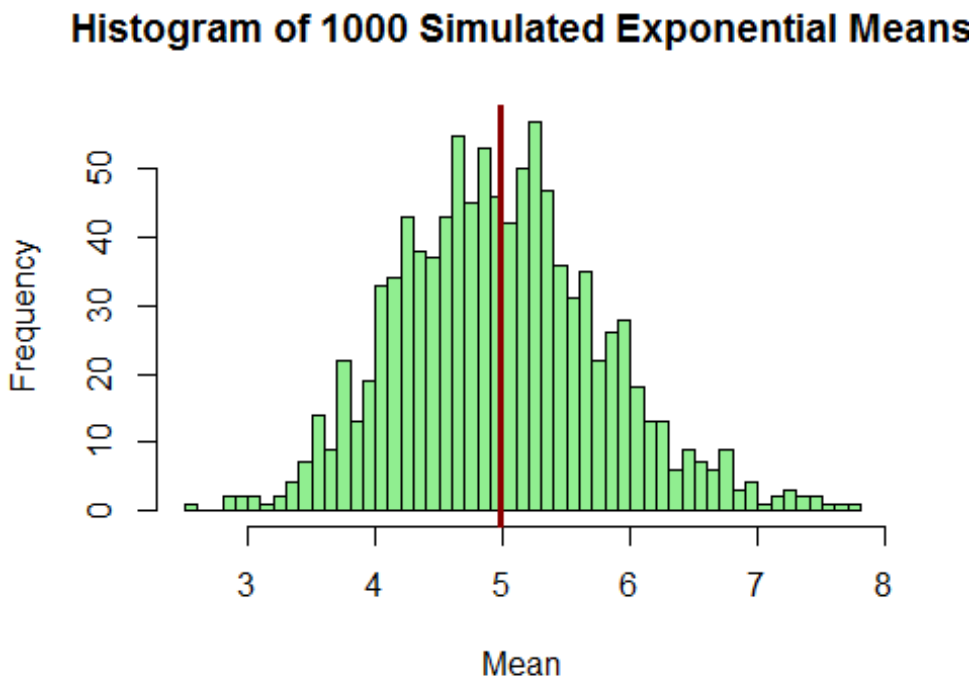
```
## [1] 4.98
```

```
1/lambda #Theoretical Mean
```

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

```
# plot the histogram of averages
hist(mean_exp, breaks = 40, xlab = "Mean", main = "Histogram of 1000
Simulated Exponential Means", col = "lightgreen")
```

```
# theoretical center of distribution
abline(v = sample_mean, col = "darkred", lwd = 3)
```



The distribution of sample means is centered at 4.98 which is very close to the theoretical center of the distribution at 5.

Sample Variance versus Theoretical Variance

```
# Sample Standard Deviation and Variance
sd(mean_exp)^2 # sample variance

## [1] 0.6580709

# Theoretical Standard Deviation and() Variance
((1/lambda)/sqrt(expo))^2 # theoretical variance

## [1] 0.625
```

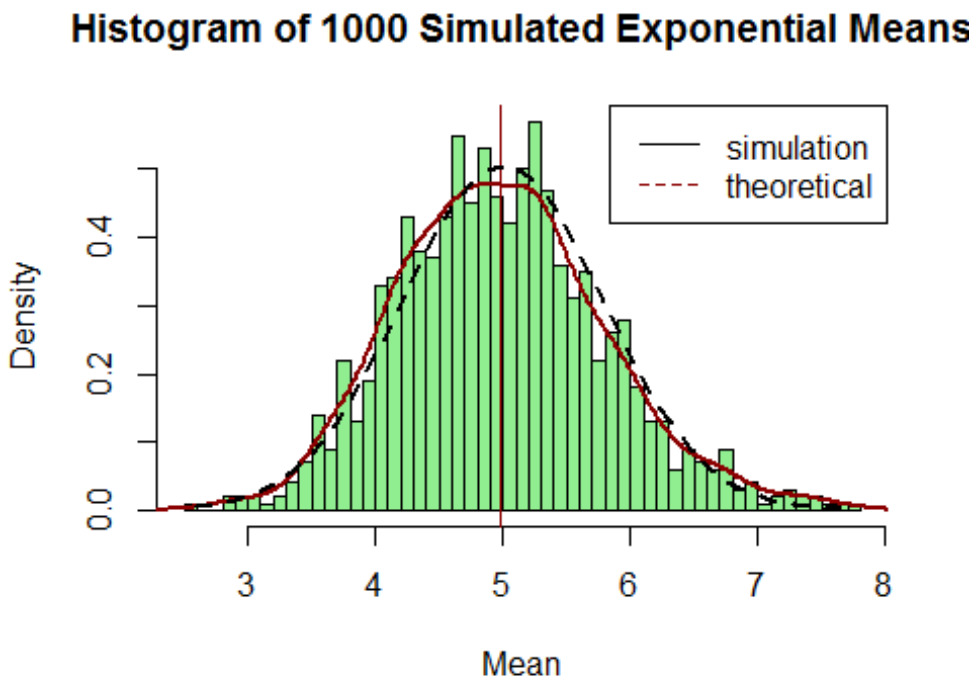
Sample variance at 0.6580709 is close to the Theoretical Variance at 0.625.

Distribution

```
# plot the histogram of averages
hist(mean_exp, prob=TRUE,breaks = 40, xlab = "Mean", main = "Histogram of
1000 Simulated Exponential Means", col = "lightgreen")
# theoretical center of distribution
abline(v = sample_mean, col = "darkred")
# Add the Theoretical Normal Distribution Line
xline <- seq(min(mean_exp), max(mean_exp), length = 100)
```

```r
yline <- dnorm(xline, mean = 1/lambda, sd = 1/lambda/sqrt(expo))
lines(density(mean_exp), lwd=2, col="darkred")
# add legend
legend('topright', c("simulation", "theoretical"), lty=c(1,2), col=c("black",
"darkred"))

# Normal distribution line creation
x <- seq(min(mean_exp), max(mean_exp), length=2*expo)
y <- dnorm(x, mean=1/lambda, sd=sqrt(((1/lambda)/sqrt(expo))^2))
lines(x, y, pch=22, col="black", lwd=2, lty = 2)
```



As the graph shows, The dotted line normal distribution curve is very close to our sampled curve, which is the dark red line. This indicate that the distribution of averages of 40 exponentials follow a normal distribution..

Conclusion

Based from the plots, the density computed using histogram and normal density plotted with theoretical mean and variance values almost overlap each other. We can conclude that the distribution of the sample means follows the normal distribution, as stated in the Central Limit Theorem.