Part 1: Simulation Exercise

Prasanti

Overview:

The purpose of this data analysis is to 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. The lambda will be set to 0.2 for all of the simulations. This investigation will compare the distribution of averages of 40 exponentials over 1000 simulations.

Simulations:

Set the simulation variables lambda, exponentials, and seed.

```
set.seed(5555)
lambda = 0.2
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
sample_means_exp <- apply(sim_exp, 2, mean)
```

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

```
# determine the mean of our sample means
sample_mean <- mean(sample_means_exp)


#The theoretical mean of an exponential distribution is 1/lambda. lambda is 0.2. Therefore, the theoret

theory_mean <- 1 / lambda

print(paste("Theoretical Mean: ", theory_mean))
```

```
## [1] "Theoretical Mean:  5"
```

```
print(paste("Sample Mean: ", sample_mean))
```

```
## [1] "Sample Mean:  4.97660899171929"
```

The code above shows that there is only a slight difference between the simulations sample mean and the exponential distribution theoretical mean.

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

```
sample_variance <- var(sample_means_exp)
theory_variance<- (lambda * sqrt(n))^-2
print(paste("Theoretical Variance: ", theory_variance))
```

```
## [1] "Theoretical Variance:  0.625"
```

```r
print(paste("Sample variance: ", sample_variance))
```

```
## [1] "Sample variance:  0.630414938233065"
```

There is only a slight difference between the simulations sample variance and the exponential distribution theoretical variance.

3. Show that the distribution is approximately normal.

Plot the density histogram for the 1000 simulations and overlay with the theoretical normal distribution.

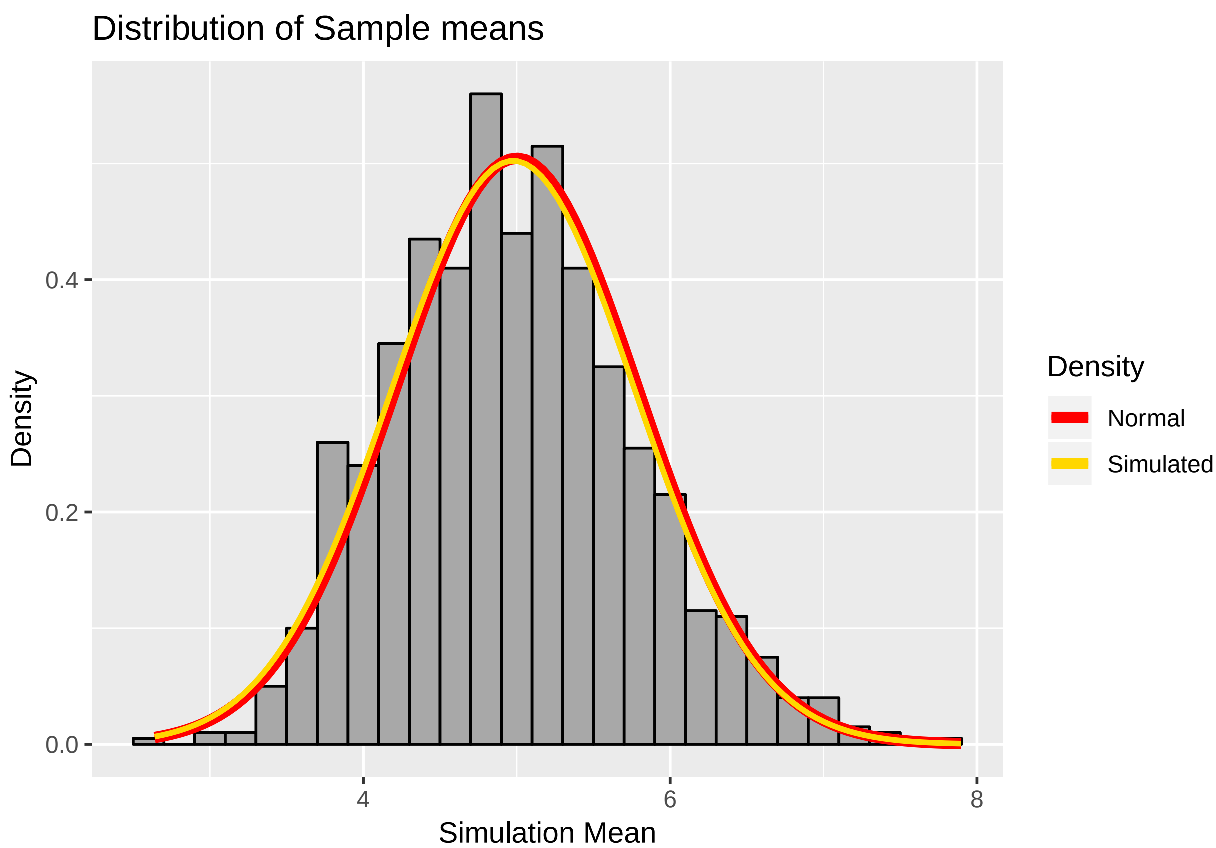
```r
library(ggplot2)
data <- data.frame(sample_means_exp)

m <- ggplot(data, aes(x =sample_means_exp))
m <- m + geom_histogram(aes(y=..density..),
                        color="black",
                        fill = "grey66",
                        binwidth=0.2
                        )

m <- m + stat_function(fun=dnorm,
                       args =list(
                                mean=theory_mean ,
                                sd=sqrt(theory_variance)
                                )
                       ,size=2
                       ,aes(color = "Normal"))

m <- m + stat_function(fun=dnorm,
                       args =list(
                                mean=sample_mean ,
                                sd=sqrt(sample_variance)
                                )
                       ,size=1
                       ,aes(color = "Simulated"))
m <- m + scale_colour_manual("Density", values = c("red", "gold1"))
m <- m + labs(title="Distribution of Sample means", x="Simulation Mean", y = "Density")


m
```



As shown in the above plot, the distribution of means of our sampled exponential distributions appear to follow a normal distribution, due to the Central Limit Theorem.