Statistical Inference Course Project (Part 1)

Annette Spithoven

3-11-2019

1. Overview

For the first part of the Statistical Inference Course Project the exponential distribution will be investigated and compared with the Central Limit Theorem.The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Lambda is set to 0.2 for all of the simulations. The distribution of averages of 40 exponentials is investigated over a thousand simulations.

2. Simulations

```
## setting seed to make simulations reproducable
set.seed(2019)

## setting predefined sample values
lambda <- 0.2
n_exponentials <- 40 ## number of exponentials
sim <- 1000 ## number of simulations
```

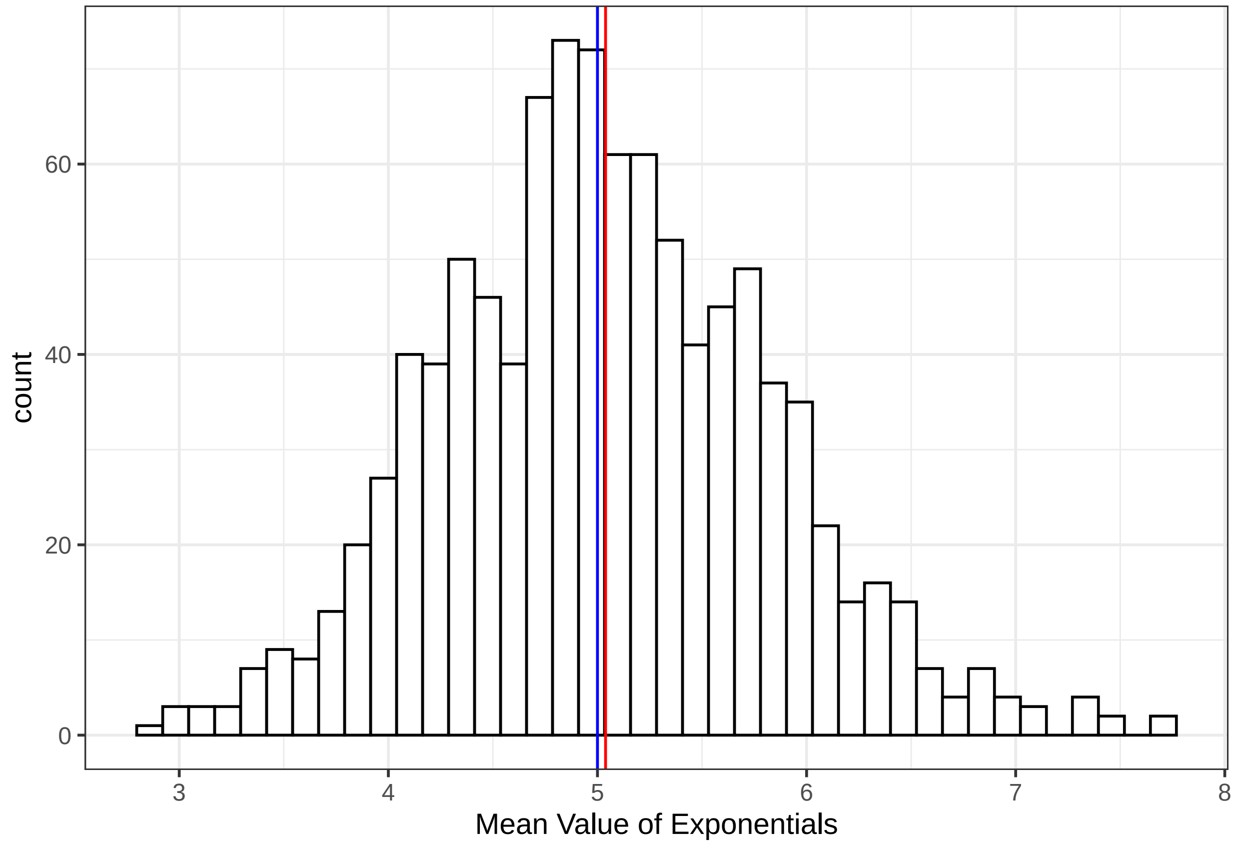
Using the predifined sample values (see the assignment explanation above) we generate 1000 simulations with 40 exponentials.

```
simulations <- replicate(sim, rexp(n_exponentials,lambda))

## calculating the mean of the exponentials
mean_sim <- data.frame(Means = colMeans(simulations))
```

Sample Mean versus Theoretical Mean

The plot below gives a graphical representation of the mean values of the simulation exponentials. The red line represents the sample mean (5.0387553), the blue line represents the theoretical mean (5).

```
## Calculating the sample mean
sample_mean <- mean(mean_sim$Means)
theoretical_mean <- 1/lambda

ggplot(mean_sim, aes(x = Means)) +
  geom_histogram(aes(y=..count..),
                 bins = 40,
                 color="black",
                 fill="white") +
  labs(main = "Means of 1000 Simulations with 40 Exponentials",
     x = "Mean Value of Exponentials") +
  ## add vertical line for sample mean
  geom_vline(xintercept = sample_mean,
            col = "red") +
  ## add vertical line for the theoretical mean
  geom_vline(xintercept= theoretical_mean,
            col = "blue" )+
  theme_bw()
```



There is slight difference between the simulation sample mean and the theoretical expected mean.

Sample Variance versus Theoretical Variance

The theoretical and sample variance were calculated

```
## Calculating the sample variance
sample_variance <- var(mean_sim$Means)
## Calculating the theoretical variance
theoretical_variance <- ((1/lambda)/sqrt(n_exponentials))^2

cbind(sample_variance, theoretical_variance)
```

```
##      sample_variance theoretical_variance
## [1,]       0.6253269                0.625
```

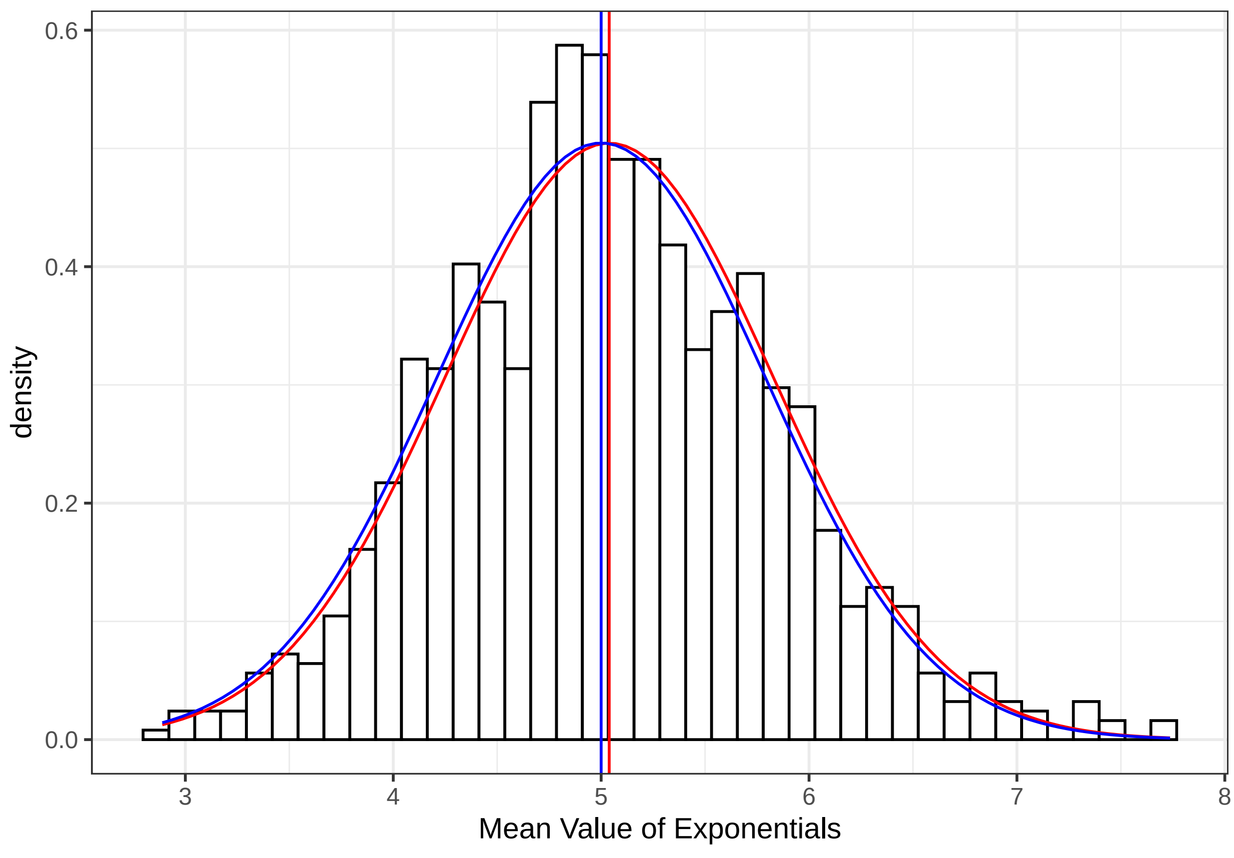
The difference between the two variances is small ($-3.2692681 \times 10^{-4}$).

Distribution

The plot below shows the density distribution for the sample as well as the theoretical distribution.

```
ggplot(mean_sim, aes(x = Means)) +
  geom_histogram(aes(y=..density..),
                 bins = 40,
                 color="black",
                 fill="white") +
  labs(main = "Histogram of 1000 Simulations with 40 Exponentials",
     x = "Mean Value of Exponentials") +
```

```r
## add sample distribution
stat_function(fun = dnorm,
              args = list(mean = sample_mean,
                          sd = sqrt(sample_variance)),
              col = "red")+
## add vertical line for sample mean
geom_vline(xintercept = sample_mean,
           col = "red") +
## add theoretical distribution
stat_function(fun=dnorm,
              args=list( mean=theoretical_mean,
                         sd=sqrt(theoretical_variance)),
              color = "blue") +
## add vertical line for the theoretical mean
geom_vline(xintercept= theoretical_mean,
           col = "blue" )+
theme_bw()
```



As can be seen in the plot, the distributions are very close to each other.