Statistical Inference Project

Jency magdalene J

14/05/2020

OverView

In this project we 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, mean of exponential distribution is 1/lambda, standard deviation is also 1/lambda.

Set lambda = 0.2 for all of the simulations. We will investigate the distribution of averages of 40 exponentials.

Simulations

We are going to set the simulation variables lambda, exponentials, and seed.After Setting run the simulation,

```
ECHO=TRUE
set.seed(1000)
lambda = 0.2
exponential = 40
simulated_mean = NULL
for (i in 1 : 1000) simulated_mean = c(simulated_mean, mean(rexp(exponential, lambda)))
```

Sample Mean versus Theoretical Mean

Here We are going to calculate the sample mean and compare it to the theoretical mean of the distribution.

Sample mean

```
mean(simulated_mean)
```

```
## [1] 4.986963
```

Theoretical mean

```
1/lambda
```

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

As we can see above there is only a slight difference between both the calculations.

Sample Variance versus Theoretical Variance

Here We are going to calculate the sample Variance and compare it to the theoretical Variance of the distribution.

Sample Variance

```
var(simulated_mean)
```

```
## [1] 0.654343
```

Theoretical Variance

```
(lambda * sqrt(exponential))/2
```

```
## [1] 0.6324555
```

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

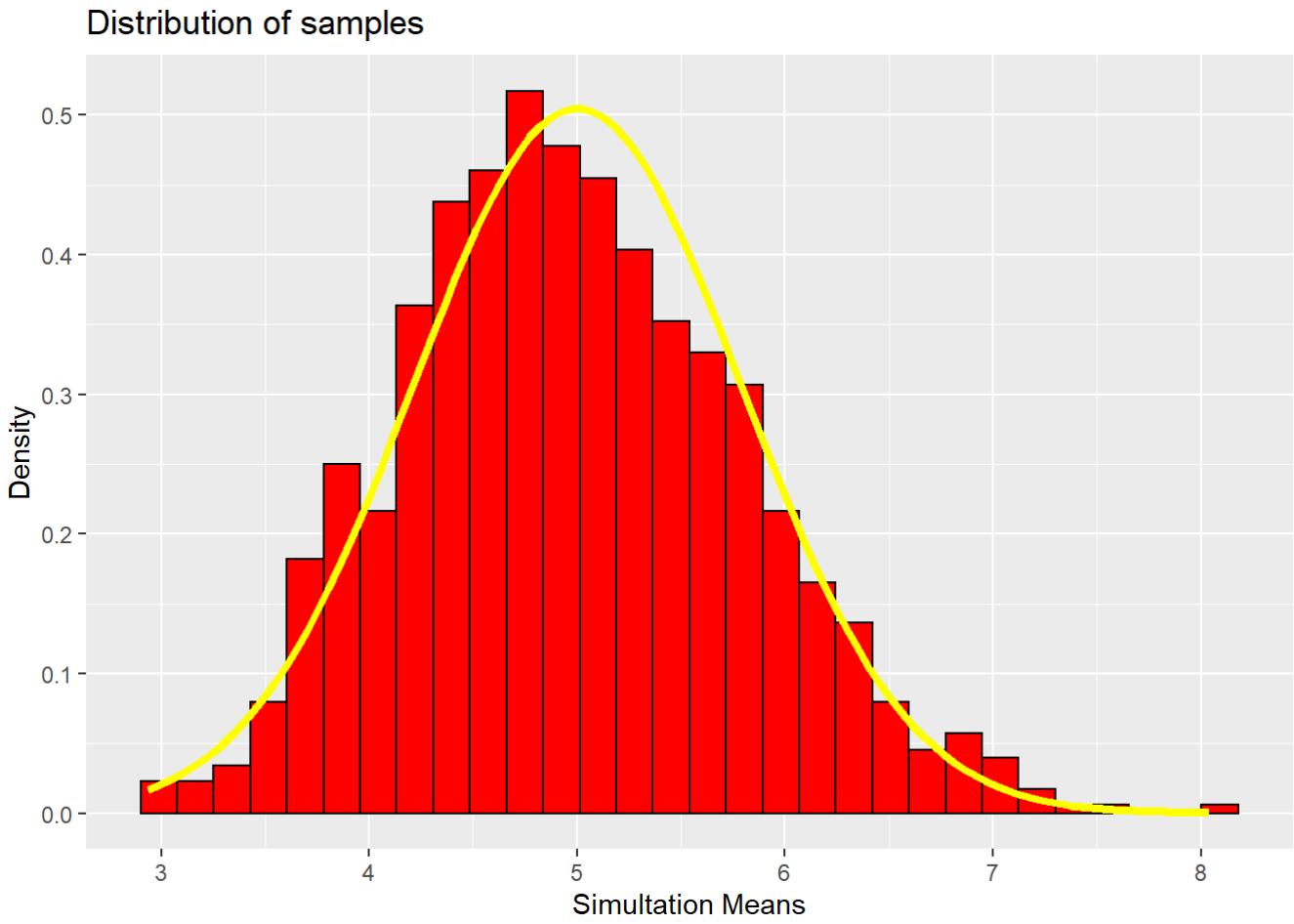
Distribution

This is a density histogram of the 1000 simulations

```
library(ggplot2)
mydata <- data.frame(simulated_mean)
m <- ggplot(mydata, aes(x =simulated_mean))
m <- m + geom_histogram(aes(y=..density..), colour="black",fill = "red")
m <- m + labs(title = "Distribution of samples", x = "Simultation Means", y = "Density")

m <- m + stat_function(fun = dnorm, args = list(mean = 1/lambda,  sd=(lambda*sqrt(exponentia
l))^-1), colour = "yellow", size = 1.5)
m
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The “yellow” line shows the normal curve formed by the the theoretical mean and standard deviation