Statistical Inference Project1 - Simulation Exercise

Sanat Das

22/6/2017

Overview

The objective of this project is to investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R using function 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. As per the instructions provided for this project, lambda = 0.2 for all of the simulations. The objective is to investigate the distribution of averages of 40 instances of the exponential fucntion. We will need to do a thousand simulations.

Using simulation and associated explanatory text, need to illustrate the properties of the distribution of the mean of 40 exponential instances to:

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.

Simulation

Loading Required Libraries and setting variables

```
library(ggplot2)
```

```
set.seed(1234)
lambda <- 0.2
n <- 40
num_simulations <- 1000
simulations <- matrix(rexp(num_simulations*n, rate=lambda),num_simulations)
simulations_mean <- apply(simulations,1,mean)
```

Sample Mean vs. Theoretical Mean

```
theoretical_mean <- 1/lambda
print(paste("Thepretical mean:",theoretical_mean ))
```

```
## [1] "Thepretical mean: 5"
```

```
sample_mean <- round(mean(simulations_mean),3)
print(paste("Sample mean based on the simulations:",sample_mean))
```

```
## [1] "Sample mean based on the simulations: 4.974"
```

```
diff_mean <- abs(sample_mean - theoretical_mean)
print(paste("difference between sample mean and theoretical mean is :", diff_mean))
```

```
## [1] "difference between sample mean and theoretical mean is : 0.0259999999999998"
```

Sample Variance vs. Theoretical Variance

```
theoretical_var <- ((1/lambda)^2)/n;
theoretical_sd <- (1/lambda)/sqrt(n);
print(paste("Theoretical variance = ",theoretical_var));
```

```
## [1] "Theoretical variance =  0.625"
```

```
print(paste("Sample variance based on simulations = ",round(var(simulations_mean ),3)));
```

```
## [1] "Sample variance based on simulations =  0.595"
```

```
print(paste("Theoretical Standard Deviation = ",round(theoretical_sd,3)));
```

```
## [1] "Theoretical Standard Deviation =  0.791"
```

```
print(paste("Actual Standard Deviation based on silmulations = ",round(sd(simulations_mean),3)));
```

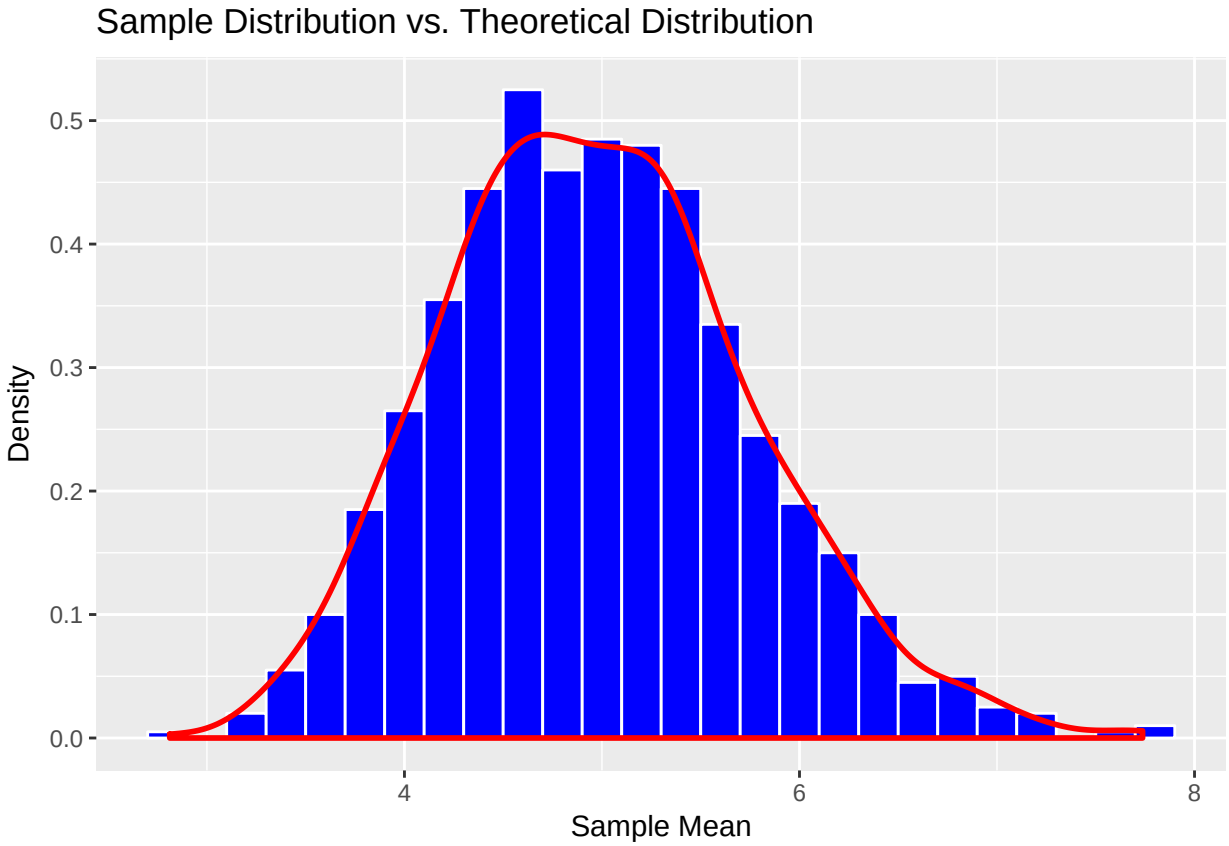
```
## [1] "Actual Standard Deviation based on silmulations =  0.771"
```

As per the above result, there is very minimal difference between sample variance based on simulations and theoretcal variance.

Distribution

To show that the distribution is approximately normal.

```
plotdata <- data.frame(simulations_mean);
ggplot(data=plotdata,aes(x=simulations_mean)) + geom_histogram(binwidth=0.2,aes(y=..density..),colour="
```



Conclusion

As per the above graph, the sampling distribution of the mean with 40 observations and lambda=0.2 is approximately normally distrbuted.