Statistical_inference_CLT

kalai

12/9/2019

Overview

This document visits the exponential distribution in R and compares its distribution with the Central Limit Theorem. rexp(n,lambda) is used to simulate the exponential distribution in R, lambda denoting the rate parameter. The mean and standard deviation of exponential distribution is 1/lambda. lambda=0.2 is assumed for all simulations. The distributions of averages of 40 exponentials are analyzed.

The objectives of this document would be to

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Loading the required libraries

```
library(knitr)
library(data.table)
library(ggplot2)
```

Initializing the variables as defined in the problem statement.

```
n <- 40 # sample size
lambda <- 0.2 # lambda for rexp
ns <- 1000 # number of simulations
quantile <- 1.96 # 95th % quantile for Confidence Interval
set.seed(500) # set the seed value for reproducibility
```

Create a matrix with 1000 simulations for 40 samples in the exponential distribution.

```
Simulation <- matrix(rexp(n * ns, rate = lambda), ns)
```

Calculate the averages across the 40 values for each of the 1000 simulations

```
Means <- rowMeans(Simulation)
```

Comparison of sample and theoritical means of the distribution

Sample mean

```
sample_mean <- mean(Means) # Mean of sample means
sample_mean
```

```
## [1] 5.010562
```

Theoretical Mean

```
t_mean <- 1 / lambda # Theoretical Mean
t_mean
```

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

The mean of sample mean is 5.010562 and the theoretical mean is 5. It can be seen that the means are close to each other.

Comparison of sample and theoretical variance

Sample variance

```
sample_Var <- var(Means)
sample_Var
```

```
## [1] 0.6003934
```

Theoretical Variance

```
theoritical_Var  <- (1 / lambda)^2 / (n)
theoritical_Var
```

```
## [1] 0.625
```

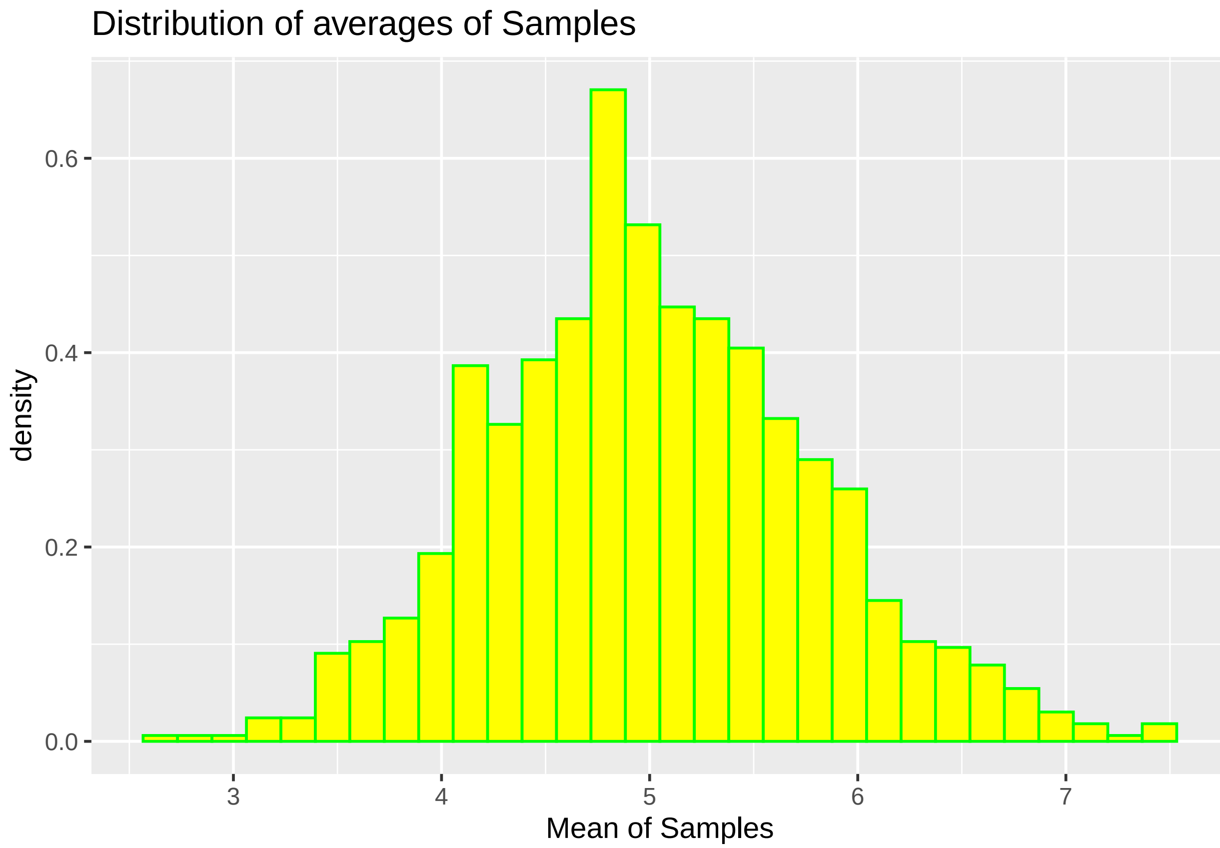
The theoretical variance is 0.625 and the variance in the sample is 0.6003934

Show that distribution is normal

```
plotdata <- data.frame(Means)
g <- ggplot(plotdata, aes(x =Means))
g <- g + geom_histogram(aes(y=..density..), colour="green",
                        fill = "yellow")
g <- g + labs(title = "Distribution of averages of Samples", x = "Mean of Samples", y = "density")

g
```

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



Comparison of standard deviation

Sample standard deviation computation

```
SD <- sd(Means)
SD
```

```
## [1] 0.7748506
```

Theoretical standard deviation computation

```
t_SD <- 1/(lambda * sqrt(n))
t_SD
```

```
## [1] 0.7905694
```

The standard deviation for sample means is 0.7748506 and theoritical means is 0.7905694. Again they are too close

To show that the distribution is approximately normal

The confidence interval is compared to prove that the distribution is indeed normal. If the difference is small, then we can conclude that the distribution tends to be normal.

computation of sample Confidence Interval

```
SCI <- round (mean(Means) + c(-1,1)*1.96*sd(Means)/sqrt(n),3)
SCI
```

```
## [1] 4.770 5.251
```

computation of theoretical Confidence Interval

```
TCI <- t_mean + c(-1,1) * 1.96 * sqrt(theoritical_Var)/sqrt(n)
TCI
```

```
## [1] 4.755 5.245
```

The thoeretical confidence interval is 4.755 5.245 and the confidence interval for sample means is 4.770 5.251

They are very close,proving that the distribution tends to be normal with large sample size.

Conclusion

The distribution adheres to the Central Limit Theorem