

Part 1: Simulation of Exponential Distribution

Alexander Kuznetsov

05/23/2018

Overview

The purpose of this report is to simulate exponential distribution with rate parameter λ set to 0.2, investigate distribution of sample means and compare it to the result predicted by Central Limit Theorem. Ultimately, distribution of the sample means is to be proven to be normal. Sample size in all simulations is set to 40 and the number of simulations is set to 1000.

Exponential Distribution

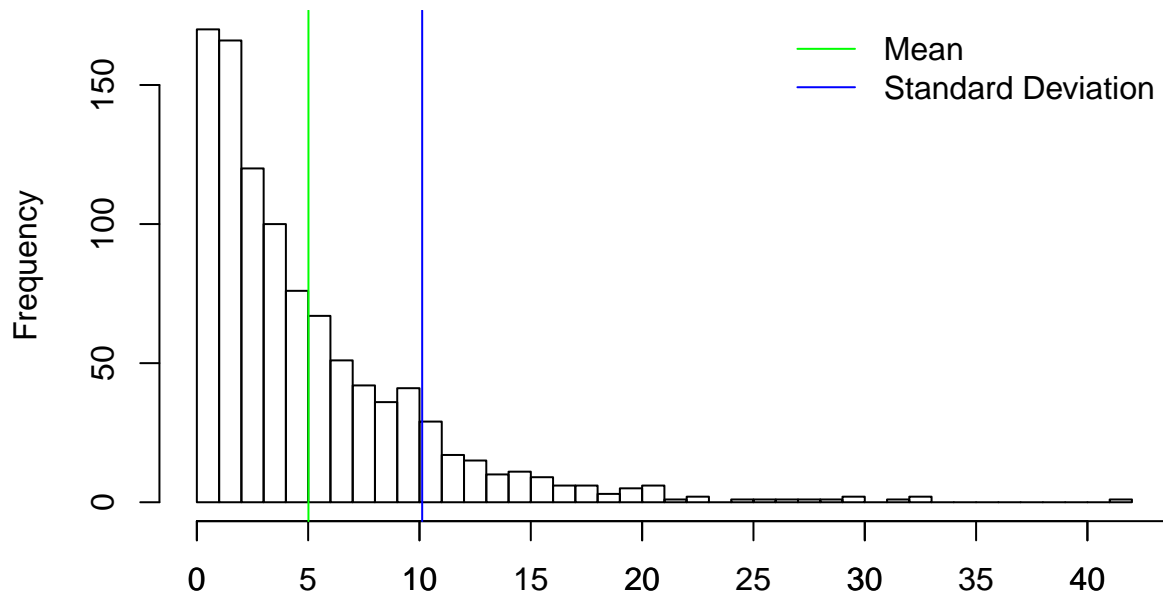
Simulations

Following code is to conduct 1000 simulations of the exponential (Poisson) distribution with the rate $\lambda = 0.2$, calculate its mean and standard deviation. Finally, the results of the simulation are going to be plotted showing mean and standard deviation values.

```
library(knitr)
opts_chunk$set(tidy.opts=list(width.cutoff=75),tidy=TRUE)

set.seed(5)
sim <- rexp(1000, 0.2)
mn <- mean(sim)
sdev <- sd(sim)
hist(sim, breaks = 40, main = "Exponential Distribution", xlab = "")
axis(1, at = seq(0, 50, by = 5))
abline(v = mn, col = "green")
abline(v = mn + sdev, col = "blue")
legend("topright", c("Mean", "Standard Deviation"), col = c("green", "blue"),
      lty = 1:1, box.lty = 0)
```

Exponential Distribution



Sample Mean vs Theoretical Mean

Theoretical value for the mean of the exponential distribution is $1/\lambda = 5$. As can be seen from the result below, mean of 1000 exponential simulations yields values very close to theoretical.

```
## [1] "Mean value for simulated exponential distribution:"  
## [1] 5.013411
```

Sample Variance vs Theoretical Variance

Theoretical value for standard deviation is also 5. This number can be confirmed with the value from the simulations.

```
## [1] "Standard deviation of the simulated exponential distribution:"  
## [1] 5.108262
```

Variance is equal to standard deviation squared and results in numbers close to 25. Before moving on to simulations of mean distribution, one can estimate standard error of the mean for a sample of 40 exponentials using $SE = \sigma/\sqrt{n}$ where $n=40$ is the size of sample.

```
se <- sdev/sqrt(40)  
se
```

```
## [1] 0.8076871
```

Table below summarizes results of theoretical statistical values and calculated values from simulation of exponential distribution.

Statistics	Theoretical value	Calculated value
Mean	5	5.01
St. Dev.	5	5.11

Distribution of the Means for Exponentials

Simulations

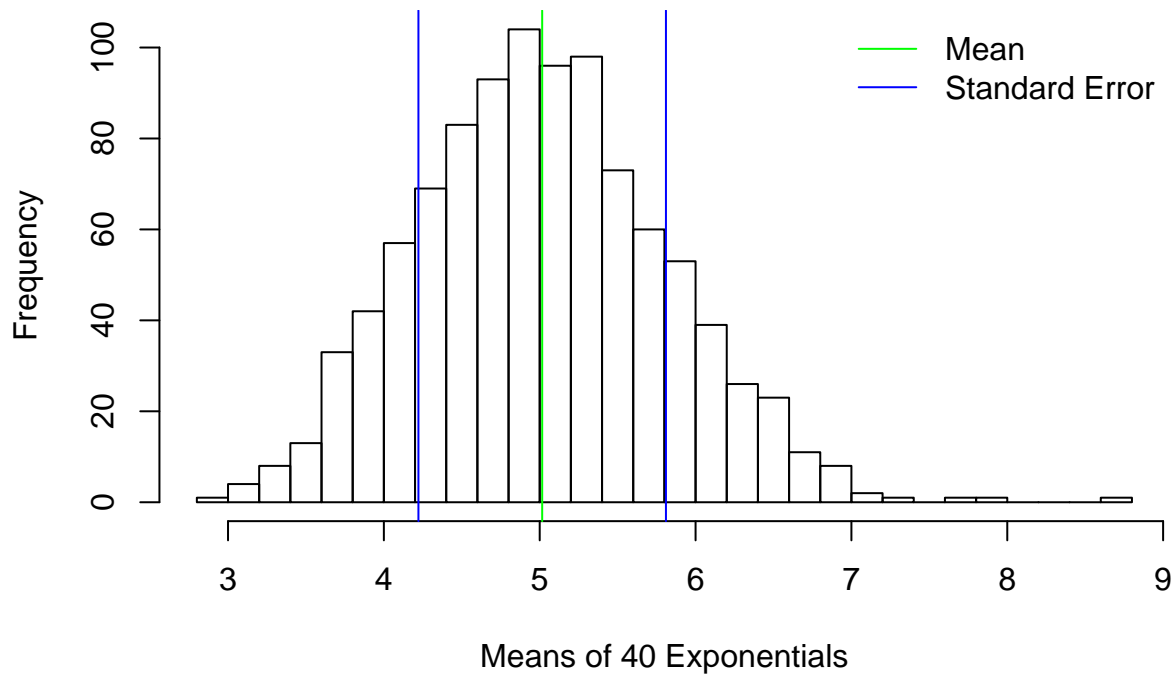
Following code simulates 40 exponentials and calculates their mean 1000 times. It also produces the estimate for the standard deviation in the distribution of the means. This value is also known as standard error of the mean.

```
set.seed(19031983)
mns <- NULL
for (i in 1:1000) mns = c(mns, mean(rexp(40, 0.2)))
mn <- mean(mns)
ser <- sd(mns)
```

Figure below shows distribution of the means with average value and one standard deviation.

```
hist(mns, breaks = 40, main = "Distribution of Averages of 40 Simulated Exponentials",
     xlab = "Means of 40 Exponentials")
abline(v = mn, col = "green")
abline(v = mn + ser, col = "blue")
abline(v = mn - ser, col = "blue")
legend("topright", c("Mean", "Standard Error"), col = c("green", "blue"), lty = 1:1,
     box.lty = 0)
```

Distribution of Averages of 40 Simulated Exponentials



Simulated Mean vs Theoretical Mean

Distribution of the mean values simulated above is centered around mean value stored in variable “mn” and reported below

```
mn
```

```
## [1] 5.015873
```

This value is very close to the theoretical value of 5 for exponential distribution with rate $\lambda = 0.2$. This is in good agreement with Central Limit Theorem which states that sample means tend to be close to the mean value of the population.

Simulated Variance vs Calculated Variance

Standard error of the means was estimated above using standard deviation for the simulated exponential distribution given that the sample size is equal to 40. This value is:

```
## [1] 0.8076871
```

This calculated value for the standard error of the means is also in good agreement with the standard deviation obtained from the simulation of the distribution of mean values. Standard error from the simulated distribution of the means is stored in variable “ser” and is reported below.

```
ser
```

```
## [1] 0.7940927
```

Conversely, standard deviation σ (and variance) for the exponential distribution can be calculated with standard error SE of the mean given that sample size n is 40: $\sigma = SE\sqrt{n}$

```
stdev <- ser * sqrt(40)
stdev
```

```
## [1] 5.022283
```

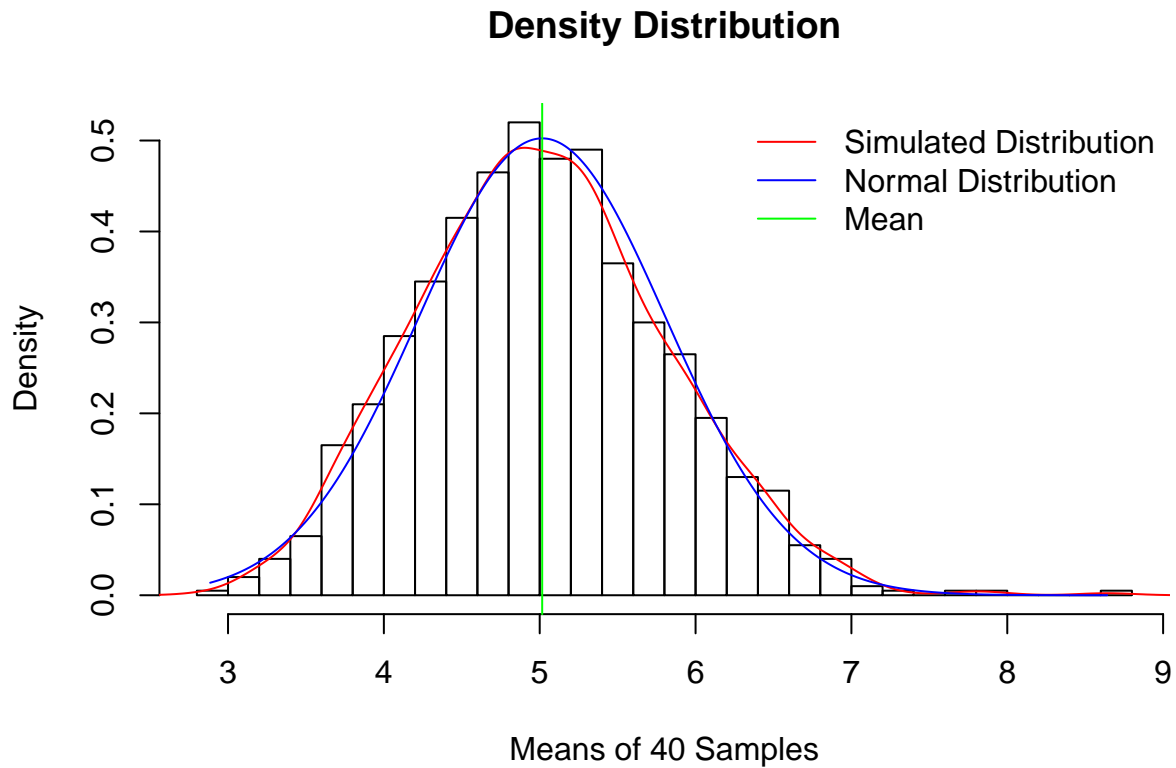
Results of theoretical and calculated values for mean and standard deviation are shown in the table below.

Statistics	Theoretical value	Calculated value
Mean	5	5.02
St. Dev.	5	5.02

Distribution

The code below is used to plot simulated distribution of the means for 40 exponentials as probability density. Red curve overlaying the distribution chart is the density distribution. Blue curve shows normal distribution with the same mean and standard deviation as the one simulated earlier in this report. Green vertical line shows the mean value for the simulated distribution of the means.

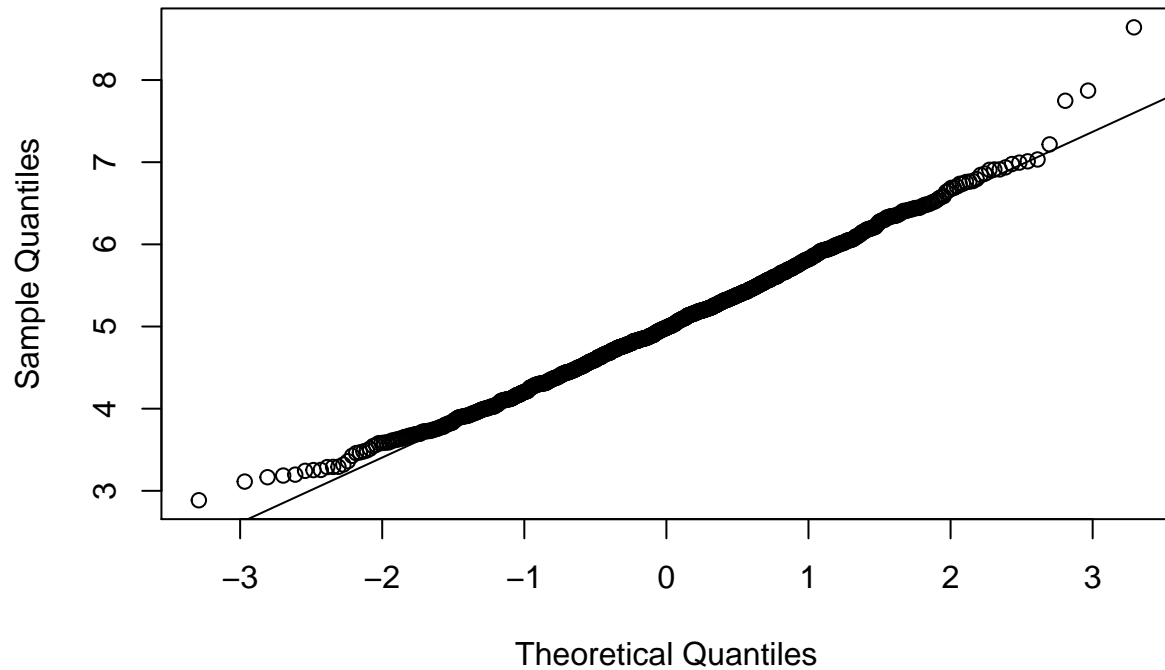
```
hist(mns, breaks = 40, prob = TRUE, xlab = "Means of 40 Samples", main = "Density Distribution")
lines(density(mns), col = "red")
lines(seq(min(mns), max(mns), length = 1000), dnorm(x = seq(min(mns), max(mns),
    length = 1000), mn, ser), col = "blue", type = "l")
abline(v = mn, col = "green")
legend("topright", c("Simulated Distribution", "Normal Distribution", "Mean"),
    col = c("red", "blue", "green"), lty = 1:1:1, box.lty = 0)
```



As can be seen from the figure above, simulated distribution is close to the normal distribution centered around 5 with standard deviation of ~ 0.8 . Another way to test for the normality of the distribution is to display q-q plot. In this plot, quantiles from simulated distribution of the means are set against the quantiles of normal distribution. Distribution equivalent to the normal would follow the straight line (q-q line). Distribution of the means follow q-q line for the most part of the range except for values located in the tails of the distribution.

```
qqnorm(mns)
qqline(mns)
```

Normal Q-Q Plot



Conclusions

It was demonstrated in this report that simulated exponential distribution in R yields mean and variance values very close to theoretical. Simulated distribution of the means for exponentials centers around the theoretical mean value as predicted by Central Limit Theorem. Its standard deviation is the same as calculated standard error of the means obtained from simulated exponential distribution. Distribution of the mean values is the same as normal distribution.