Part1_Simulation_Exercise

Aleksander Petrovskii

June 8, 2017

Overview

The purpose of this data analysis is to investigate the exponential distribution and compare it to the Central Limit Theorem.
Set lambda will be set to 0.2 for all of the simulations. This investigation will compare the distribution of averages of 40 exponentials over 1000 simulations.
In this exercise we shall:
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.

Simulations

Set and run simulation

Let's run series of 1000 simulations to create a data set for comparison to theory. For this analysis, the lambda will be set to 0.2 for all of the simulations. This investigation will compare the distribution of averages of 40 exponentials over 1000 simulations.

```
  lambda = 0.2
  exp    = 40
  sim    = 1000
  set.seed(1966)
```

Run simulation:

```
sim.means = NULL
for (i in 1 : 1000) sim.means = c(sim.means, mean(rexp(exp, lambda)))
```

Sample mean against theoretical mean (Question 1)

Sample mean

Calculating the sample mean.

```
sample.mean <- mean(sim.means)
sample.mean
```

```
## [1] 5.033011
```

Theoretical mean

The theoretical mean of an exponential distribution is 1/lambda and...

```
theor.mean <- 1/lambda
theor.mean
```

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

Compare

Show difference between the simulations sample mean and the exponential distribution theoretical mean.

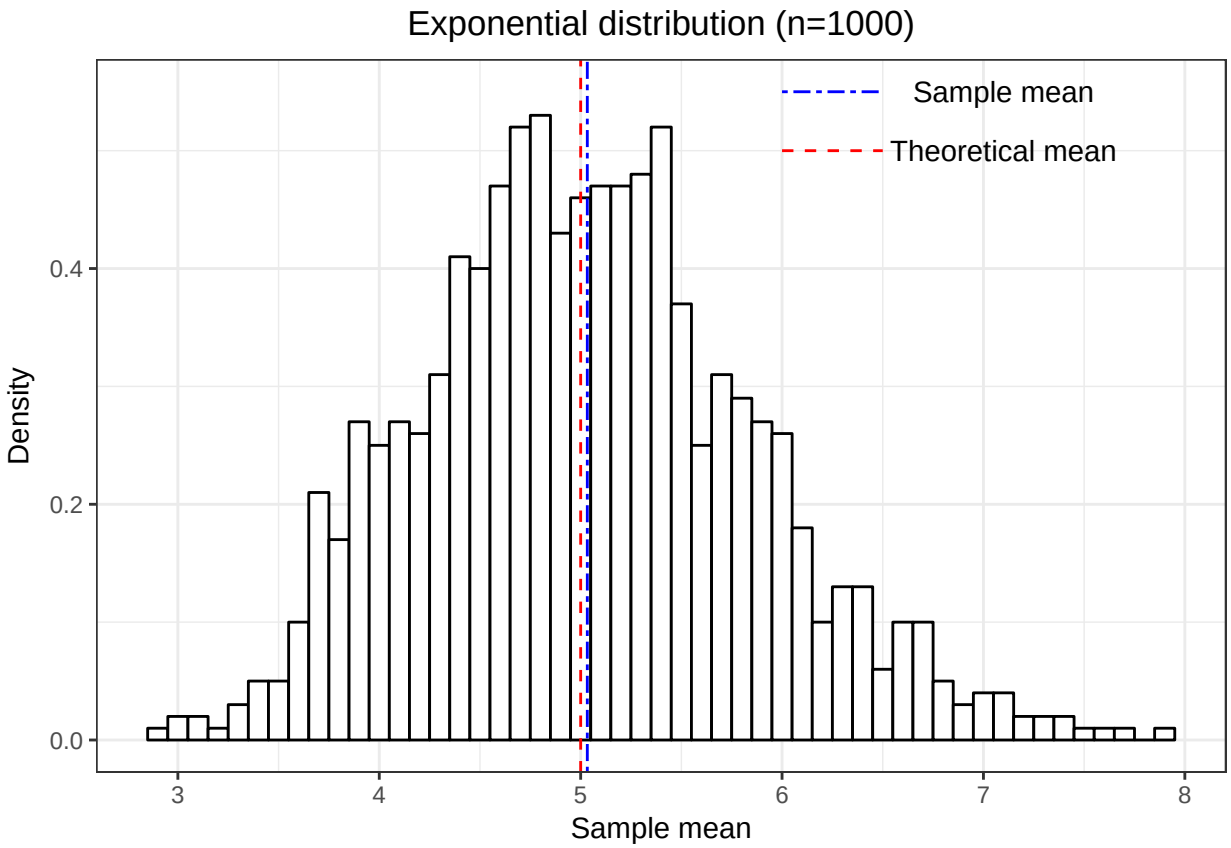
```
abs(mean(sim.means)-theor.mean)
```

```
## [1] 0.03301148
```

Show where the distribution is centered at and compare it to the theoretical center of the distribution.

This graph show that center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution.

```
data <- as.data.frame(sim.means)
ggplot(data, aes(x=sim.means, y = ..density..) ) +
  theme_bw() +
  geom_histogram(binwidth=0.1, color='black', fill='white') +
  geom_vline(xintercept = sample.mean, color='blue', size=0.5, linetype="twodash") +
  geom_vline(xintercept = theor.mean, color='red', size=0.5, linetype="dashed")+
  xlab('Sample mean') +
  ylab('Density') +
  ggtitle("Exponential distribution (n=1000)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  annotate("segment", x = 6, xend = 6.5, y = 0.55, yend = 0.55, color='blue', size=0.5,linetype="twodash
  annotate("text", x = 7.1, y = 0.55, label = "Sample mean" )+
  annotate("segment", x = 6, xend = 6.5, y = 0.5, yend = 0.5, color='red',  size=0.5,linetype="dashed")
  annotate("text", x = 7.1, y = 0.5, label = "Theoretical mean" )
```



Sample Variance vs Theoretical Population Variance (Question 2)

We will compare the variance present in the sample means of the 1000 simulations to the theoretical varience of the population.The variance of the sample means estimates the variance of the population by using the varience of the 1000 entries in the means vector times the sample size, 40.

Sample variance

Calculate sample variance

```
sample.var<- var(sim.means)
sample.var
```

```
## [1] 0.6735214
```

Theoretical variance

Calculate sample variance theoretical variance as expotential distribution

```
theoretical.var <-(lambda * sqrt(exp))^-2
theoretical.var
```

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

Compare

Show difference between simulations sample variance and exponential distribution theoretical variance.

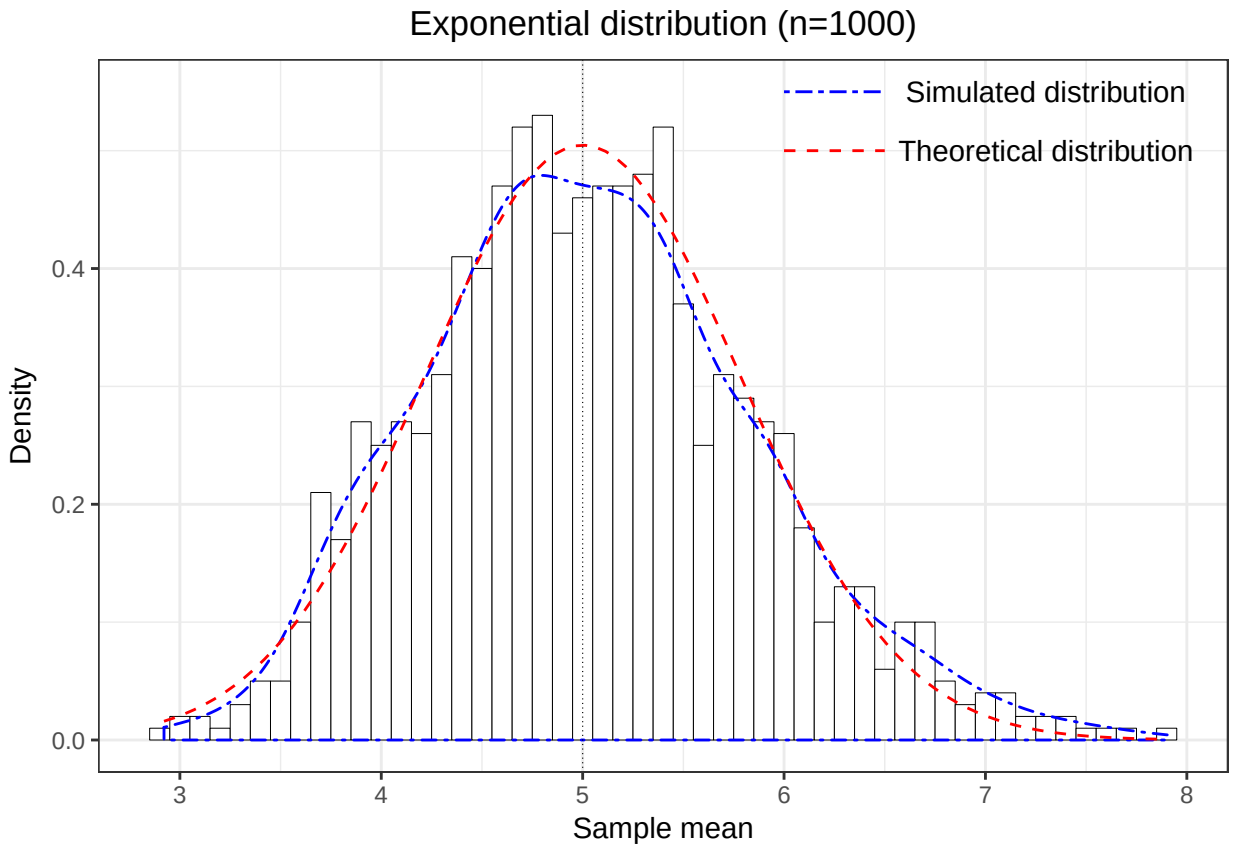
```
abs(sample.var - theoretical.var)
```

```
## [1] 0.0485214
```

Show that the distribution is approximately normal (Question 3)

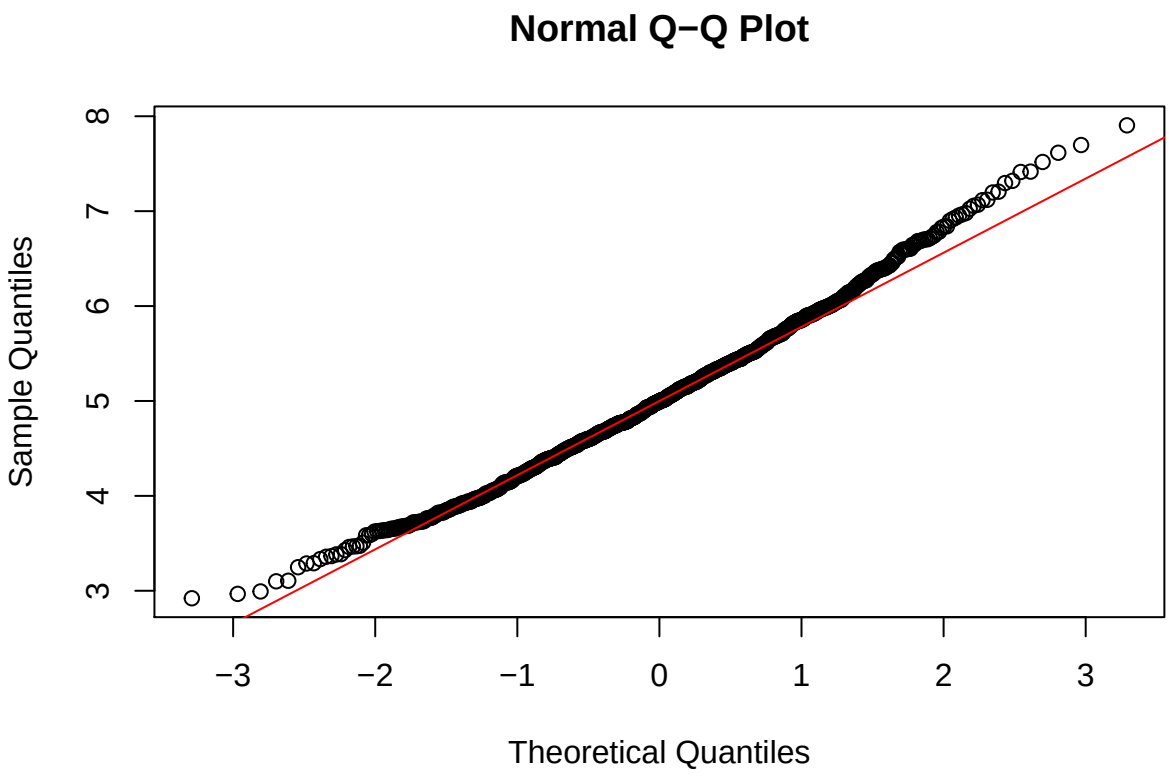
The figure above show how close the sample mean's dispersion is to the dispersion of a standard normal curve with the same theoretical mean and theoretical variance

```
data <- as.data.frame(sim.means)
ggplot(data, aes(x=sim.means) ) +
  theme_bw() +
  geom_histogram(aes(y=..density..), binwidth=0.1, color='black', fill='white', size=0.1) +
  geom_density(color="blue", linetype="twodash")  +
  stat_function(fun=dnorm, color="red", linetype="dashed", args=list(mean=lambda^-1, sd=(lambda*sqrt(ex
  xlab('Sample mean') +
  ylab('Density') +
  geom_vline(xintercept = theor.mean, color='black', size=0.2, linetype="dotted")+
  ggtitle("Exponential distribution (n=1000)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  annotate("segment", x = 6, xend = 6.5, y = 0.55, yend = 0.55, color='blue', size=0.5,linetype="twodas
  annotate("text", x = 7.3, y = 0.55, label = "Simulated distribution" )+
  annotate("segment", x = 6, xend = 6.5, y = 0.5, yend = 0.5, color='red',  size=0.5, linetype="dashed"
  annotate("text", x = 7.3, y = 0.5, label = "Theoretical distribution" )
```



Plot below suggests the theortical quanties matches closely with actual quantites. The theoretical quantiles also match closely with the actual quantiles. These evidences prove that the distribution is approximately normal.

```
qqnorm(sim.means,  main ="Normal Q-Q Plot")
qqline(sim.means, col = "red")
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
3
4
5
6
7
8
−3
−2
−1
0
1
2
3