Exponential Distribution in R vs Central Limit Theorem

Author: Siqi Zhu

Overview

This projrct attempt to answer the following questions:

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

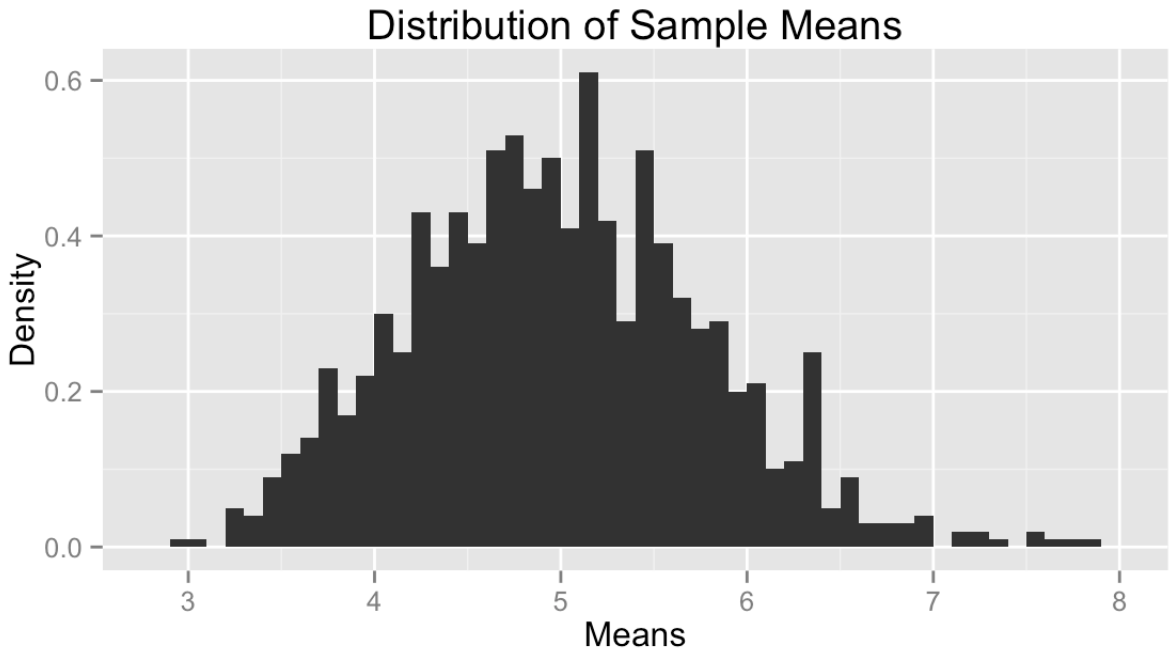
```
# load R package
library(ggplot2)

# set lambda, n, collect
lambda = 0.2
n = 40
collect = 1:1000

# set a seed in order to reproduce this project
set.seed(0815)

# calculate the mean of each collection
means = data.frame(mean = sapply(collect, function(x) {mean(rexp(n, lambda))}))

# graph the density of each mean with ggplot2
ggplot(means, aes(x = mean)) +
        geom_histogram(binwidth = 0.1, aes(y = ..density..)) +
        labs(title="Distribution of Sample Means", x="Means", y="Density")
```



It seems the mean of this distribution is around 5, let's check it out!

Sample mean & theoretical mean

```
# the theoretical mean
m_t = 1 / lambda
m_t
```

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

```
# the sample mean
m_s = mean(means$mean)
m_s
```

```
## [1] 4.983762
```

As we can see, sample mean and theoretical mean are very close!

Sample variance & theoretical variance

```
# theoretical variance
v_t = ((1 / lambda) / sqrt(n)) ^ 2
v_t
```

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

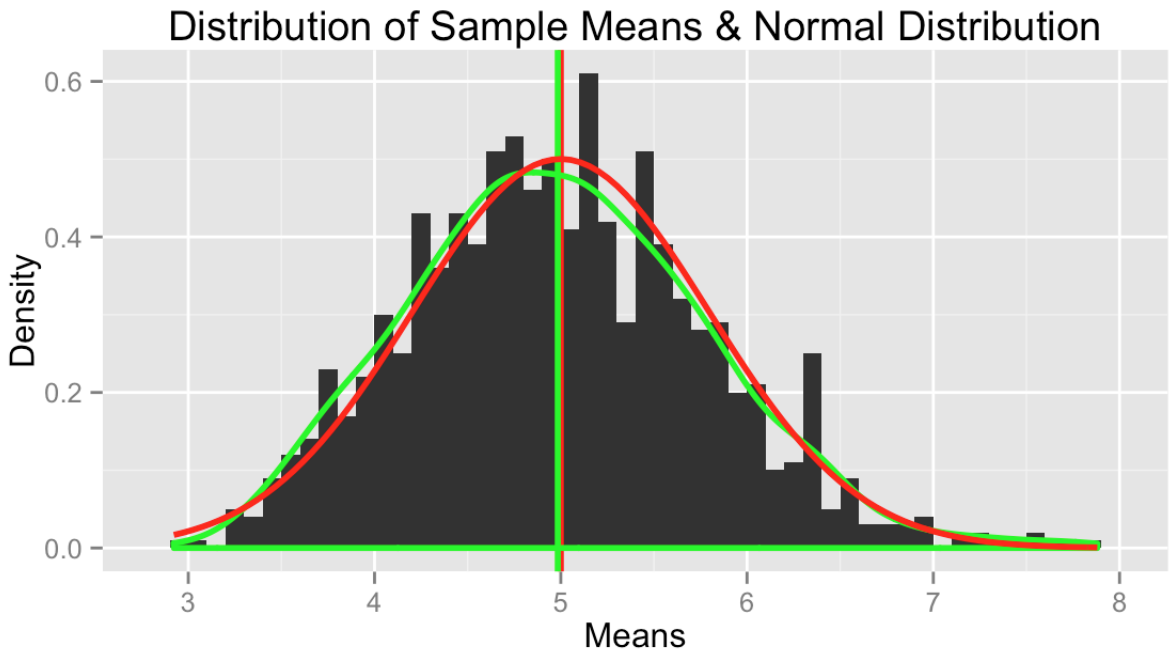
```
# sample variance
v_s = var(means$mean)
v_s
```

```
## [1] 0.6360483
```

Sample variance and theoretical variance are almost the same!

Compare this distribution with normal distribution

```
ggplot(means, aes(x = mean)) +
        geom_histogram(binwidth = 0.1, aes(y = ..density..)) +
        # draw a vertical line to highlight the thoretical mean
        geom_vline(xintercept = m_t,  colour="red", size = 1) +
        # deaw a vertical line to highlight the distribution mean
        geom_vline(xintercept = m_s,  colour="green", size = 1) +
        # draw the density of sample means
        geom_density(colour ="green", size = 1) +
        # draw a normal distribution
        stat_function(fun = dnorm, arg = list(mean = m_t , sd = sqrt(v_s)), colour
= "red", size = 1) +
        labs(title = "Distribution of Sample Means & Normal Distribution", x = "Me
ans", y = "Density")
```



According to the graph above, the distribution of the simulate mean is approaching a normal distribution!