Part1 Exponential Distribution vs. Central Limit Theorem

By Zoey Phay

Overview

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with 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. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should 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

Library required

```
library(ggplot2)
```

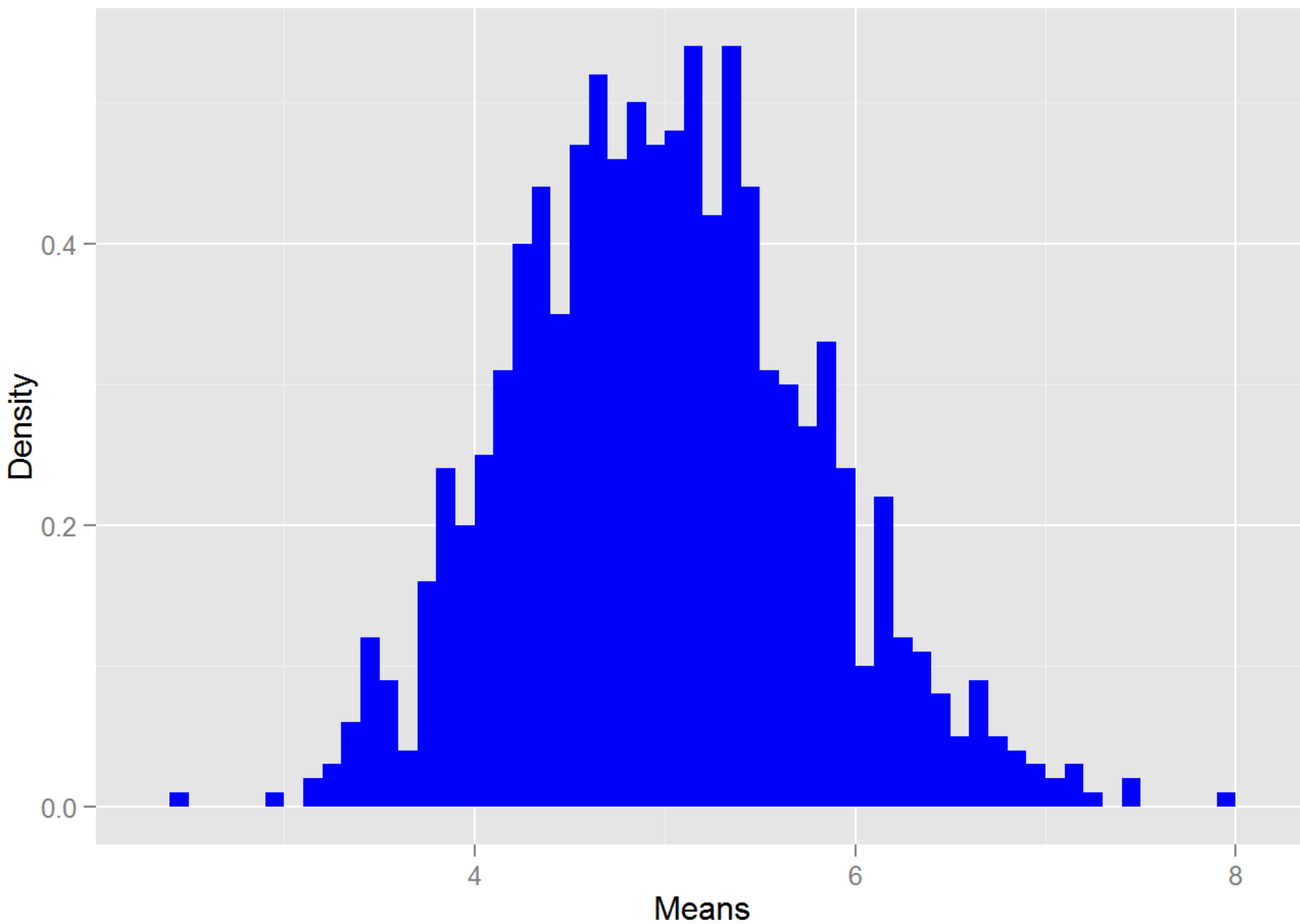
```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
# Set lambda = 0.2 for all of the simulations
lambda = 0.2

# Investigate 40 exponentials...
n = 40

# ...for 1000 simulations
nsims = 1:1000

# setting seed for reproducing the data
set.seed(876)

# gathering means..
means <- data.frame(x = sapply(nsims, function(x) {mean(rexp(n, lambda))}))

head(means)
```

```
##          x
## 1 4.129788
## 2 5.302187
## 3 4.699978
## 4 4.332842
## 5 5.726389
## 6 4.387737
```

Let's graph it..

```
ggplot(data = means, aes(x = x)) +
  geom_histogram(fill = "blue",binwidth=0.1, aes(y=..density..)) +
  labs(x="Means") +
  labs(y="Density")
```



Question 1: Sample Mean vs Theoretical Mean

Expected Mean of Exponential Distribution of lamda:

```
simmu = 1 / lambda
print(simmu)
```

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

Sample Mean of Exponential Distributions (nsims = 1:1000;n = 40) :

```
simmean <- mean(means$x)
print(simmean)
```

```
## [1] 4.991893
```

Question 2: Sample Variance vs Theoretical Variance

Expected Stardard Deviation

```
simexpsd <- (1/lambda)/sqrt(n)
print(simexpsd)
```

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

Variance of this Expected Stardard Deviation

```
simexpvar <- simexpsd^2
print(simexpvar)
```

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

Simulated Standard Deviation

```
simsd <- sd(means$x)
print(simsd)
```

```
## [1] 0.78538
```

Variance of this Simulated Stardard Devation

```
simvar <- var(means$x)
print(simvar)
```

```
## [1] 0.6168217
```

Results for Question 1 and for Question 2 are very close.

Question 3: Showing the distribution is approximately normal

From the graph, the distribution of the simulated means (blue) approaches the normal distribution (red) and that their means (blue and red vertical lines) approach each other too

```
ggplot(data = means, aes(x = x)) +
  geom_histogram(binwidth=0.1, aes(y=..density..), fill = I('#8A8A8A'),) +
  stat_function(fun = dnorm, arg = list(mean = simmu , sd = simsd), colour = "red", size=2) +
  geom_vline(xintercept = simmu, size=1, colour="red") +
  geom_density(colour="blue", size=2) +
  geom_vline(xintercept = simmean, size=1, colour="blue") +
  labs(x="Means") +
  labs(y="Density")
```