Simulation of Exponential Distribution using R

Carlos Hernández

25/9/2020

Contents

```
set.seed(1)
library(ggplot2)
theme_set(theme_bw())
```

Overview

In this project we 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.

To achieve this we will investigate the distribution of averages of 40 exponentials with lambda 0.2, replicated a thousand times.

```
lambda <- 0.2
n <- 40
```

Definitions

Exponential Distribution   The exponential distribution describes the arrival time of a randomly recurring independent event sequence. If   is the mean waiting time for the next event recurrence, its probability density function is:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Here is a graph of the exponential distribution with mean = 1.

1

```
data <- data.frame(value = c(t(rexp(1000, rate = 1))))

ggplot(data, aes(x=value)) +
  geom_histogram(aes(y=..density..),binwidth=.25, col="black", fill="lightblue")+
  labs(title= "Exponential distribution with mean = 1",
      caption="Produced by Carlos Hernández") +
  xlab("x") +
  ylab("y")
```

Exponential distribution with mean = 1

Produced by Carlos Hernández
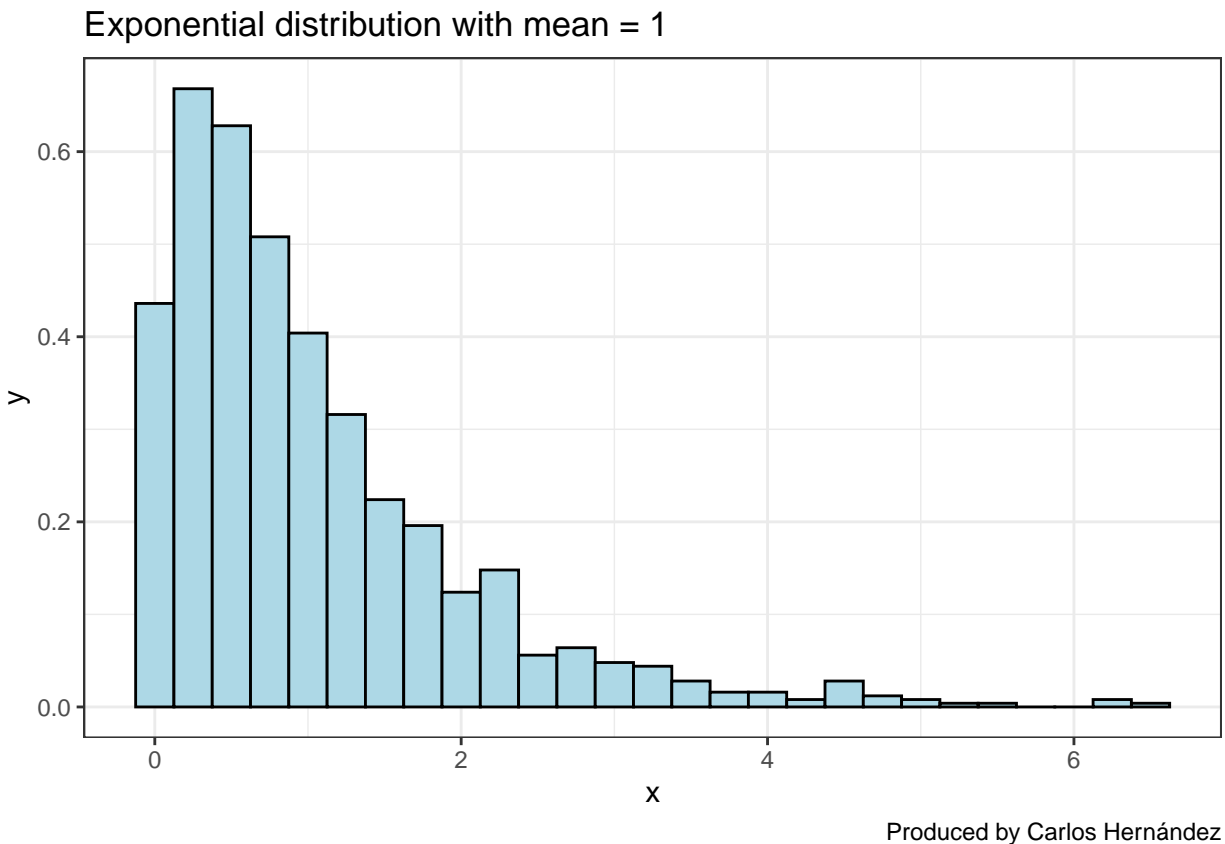
Central Limit Theorem   The central limit theorem states that if you have a population with mean   and standard deviation   and take sufficiently large random samples from the population, then the distribution of the sample means will be approximately normally distributed.

This will hold true regardless of whether the source population is normal or skewed, provided the sample size is sufficiently large (usually n > 30). If the population is normal, then the theorem holds true even for samples smaller than 30. In fact, this also holds true even if the population is binomial, provided that min(np, n(1-p))> 5, where n is the sample size and p is the probability of success in the population. This means that we can use the normal probability model to quantify uncertainty when making inferences about a population mean based on the sample mean.

Simulation

First we are going to simulate our exponential distribution, and replicated it one thousand times.

```
expData <- replicate(1000, rexp(n, lambda)) # replicate 1000 times
expData <- data.frame(value = c(t(expData))) # convert to data frame

# plot
ggplot(expData, aes(x=value)) +
  geom_histogram(aes(y=..density..), binwidth=.8,colour="black", fill="lightblue") +
    labs(title= "Exponential distribution with lambda = 0.2 and 40 observations",
         subtitle = "Replicated 1000 times",
         caption="Produced by Carlos Hernández") +
  xlab("x") +
  ylab("exp(x)")
```

Exponential distribution with lambda = 0.2 and 40 observations
Replicated 1000 times
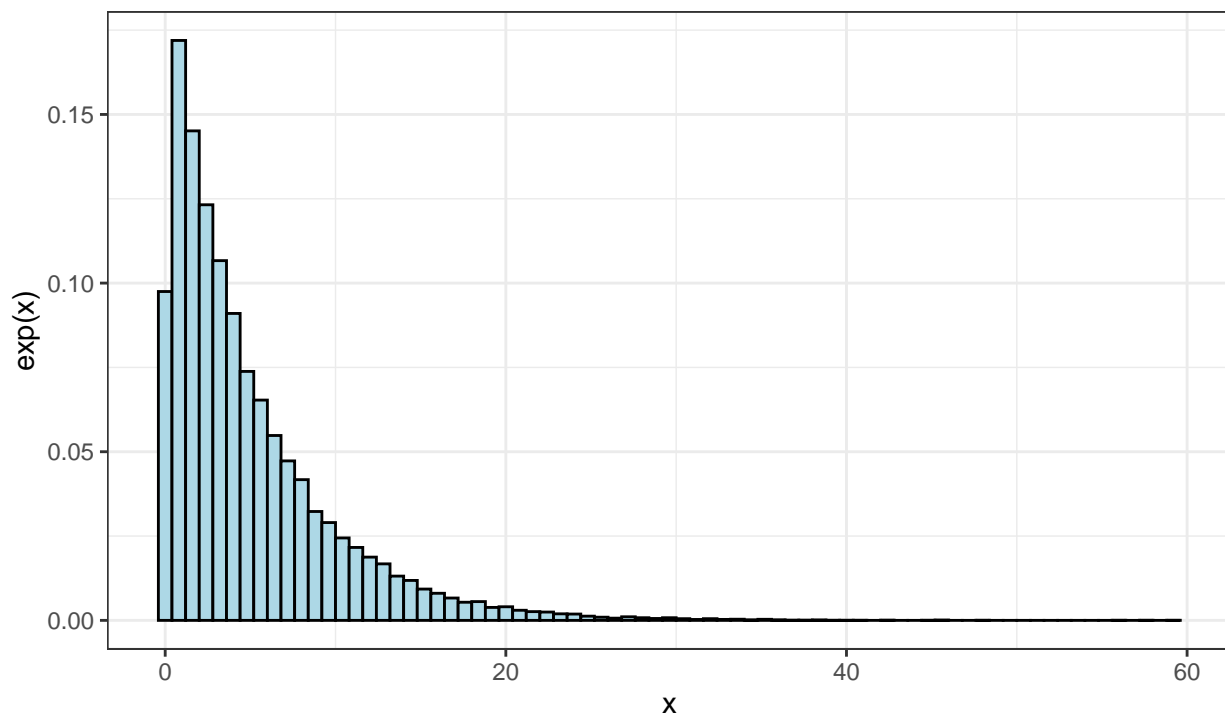
Produced by Carlos Hernández

According to the Central Limit Theorem, if we take simple random samples from the population and compute the mean for each of the samples, the distribution of sample means should be approximately normal.

Thus, we can calculate the mean of the distribution one thousand times and save its value in a data frame and the resulted distribution should be approximately normal.

```
data <- replicate(1000, mean(rexp(n,lambda)))
data <- data.frame(value = c(t(data)), size = 40)

ggplot(data, aes(x=value)) +
  geom_histogram(aes(y=..density..),binwidth=.25, col="black", fill="lightblue") +
  labs(title= "Average of 40 random exponential distribution",
       subtitle = "Replicated 1000 times",
       caption="Produced by Carlos Hernández") +
```

```
  xlab("x") +
  ylab("mean")
```

## Average of 40 random exponential distribution
Replicated 1000 times
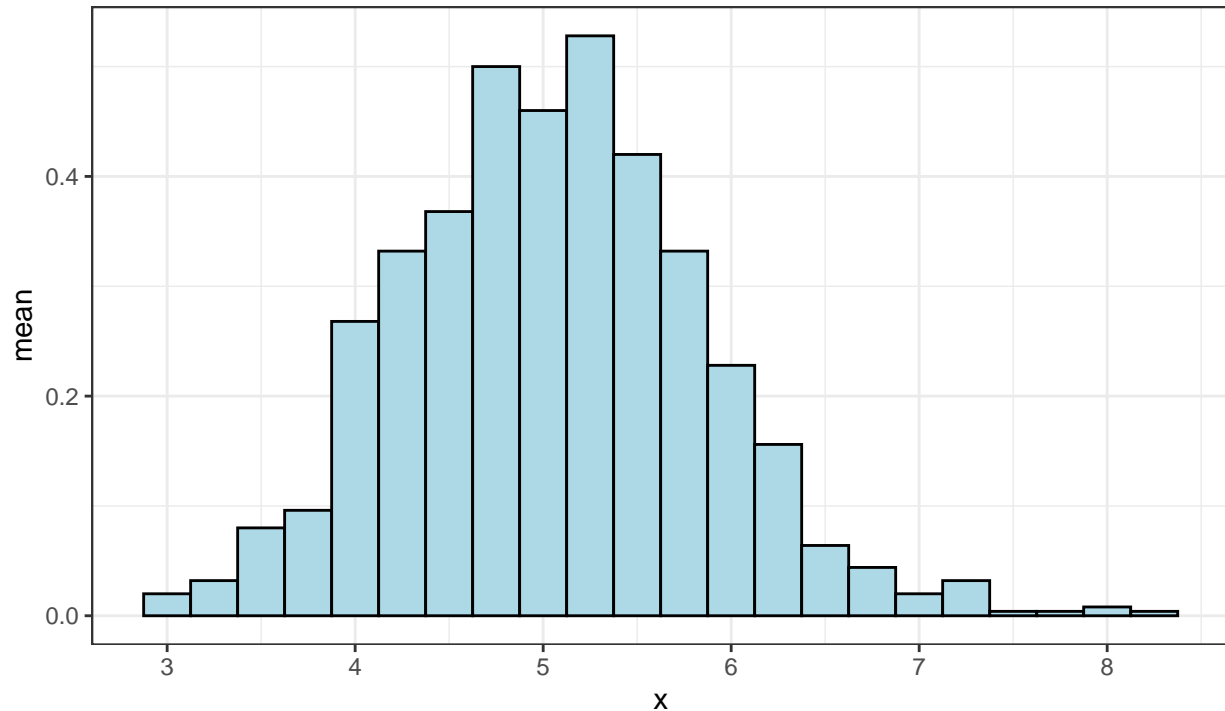
Sample Mean versus Theoretical Mean

We know that the theoretical mean is given by 1/lambda, so we can calculate it and compare it with the sample mean.

```
theoretical_mu <- 1/lambda # calculate theoretical mean
sample_mu <-mean(data$value) # calculate experimental mean

ggplot(data, aes(x=value)) +
  stat_function(fun=dnorm,
                color="black",
                args=list(mean=mean(data$value),
                sd=sd(data$value)))+
  geom_vline(xintercept = theoretical_mu, colour="red") +
  geom_text(aes(x=theoretical_mu-.25,
                label="\nTheoretical mean", y=.2),
            colour="red", angle=90, text=element_text(size=11)) +
  geom_vline(xintercept = sample_mu, colour="green")+
  geom_text(aes(x=sample_mu+.05, label="\nSample mean", y=.2),
            colour="green", angle=90) +
  labs(title= "Theoretical mean vs sample mean",
```

```
        caption="Produced by Carlos Hernández") +
xlab("x") +
ylab("y")
```

Theoretical mean vs sample mean
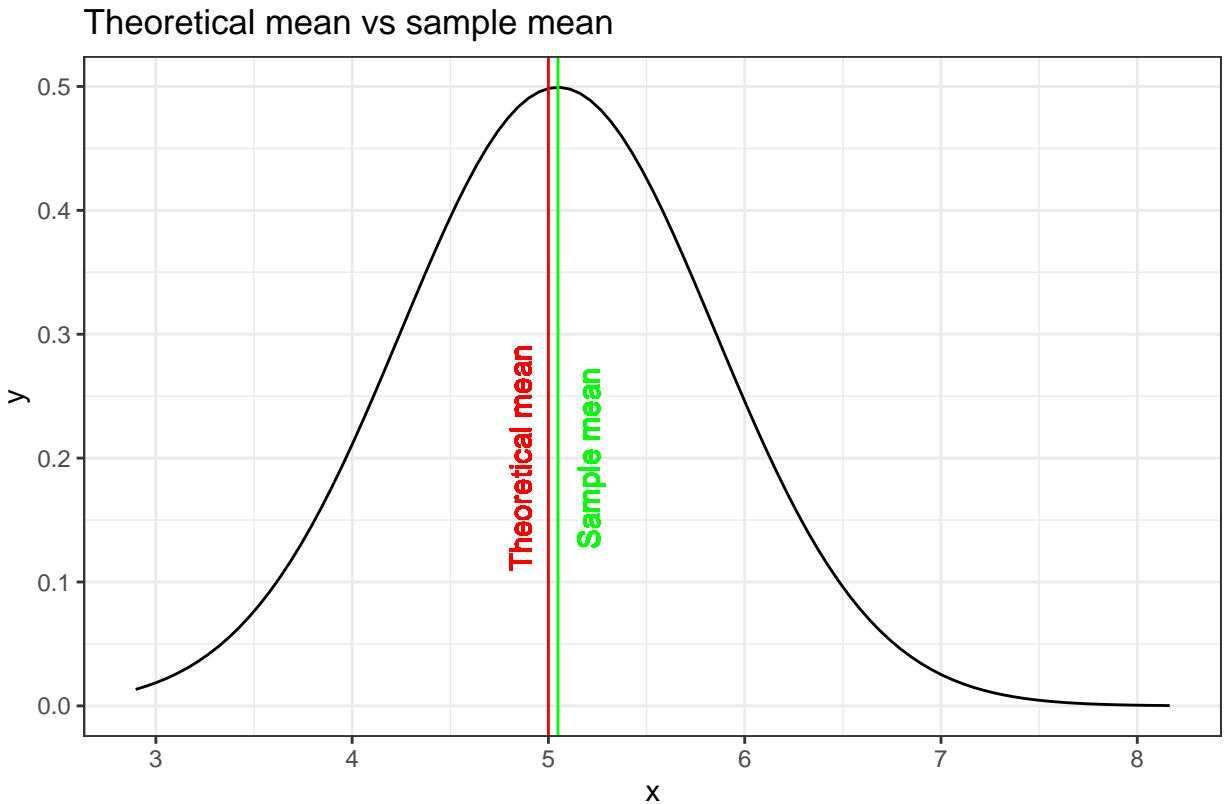
Produced by Carlos Hernández

The theoretical mean is: 5 and the sample mean is: 5.05.

Sample Variance versus Theoretical Variance

Also, we know that the theoretical variance is given by 1/lambda^2, so we can calculate it and compare it with the sample variance.

```
theoretical_variance <- 1/(n * lambda^2)
sample_variance <- round(var(data$value),3)

ggplot(data, aes(x=value)) +
  stat_function(fun=dnorm, color="black", args=list(mean=mean(data$value), sd=sd(data$value)))+
  geom_vline(xintercept = sample_mu, colour="gray", linetype="dashed")+
  geom_vline(xintercept = theoretical_mu, colour="gray", linetype="dashed")+
  geom_segment(aes(x = sample_mu, y = 0.36, xend =sample_mu +
                    sample_variance, yend = 0.36), colour="green") +
  geom_segment(aes(x = theoretical_mu - theoretical_variance, y = 0.35,
                    xend =theoretical_mu, yend = 0.35), colour="red") +
  labs(title= "Theoretical variance vs sample variance",
       caption="Produced by Carlos Hernández") +
  geom_text(aes(x=sample_mu+.55, label="\nSample variance", y=.42), colour="green") +
```

```
geom_text(aes(x=theoretical_mu-.65, label="\nTheoretical variance", y=.33), colour="red") +
xlab("x") +
ylab("y")
```

Theoretical variance vs sample variance
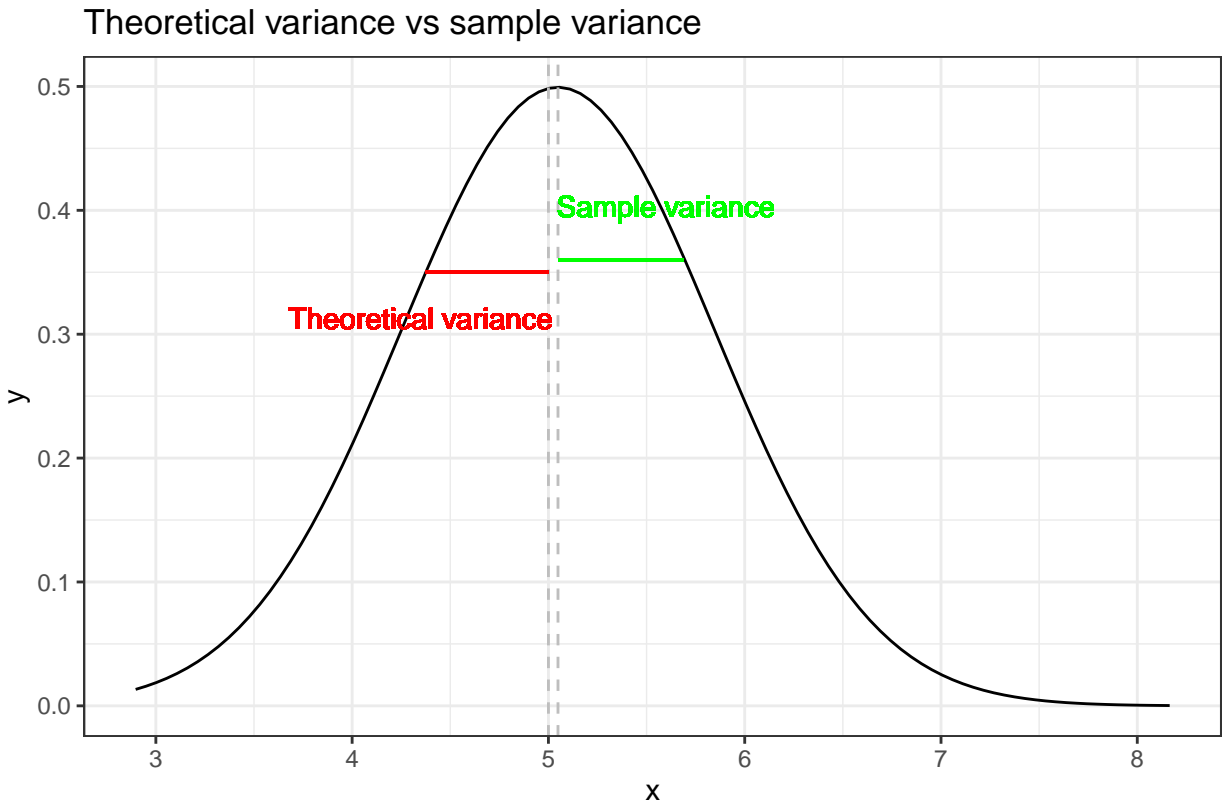
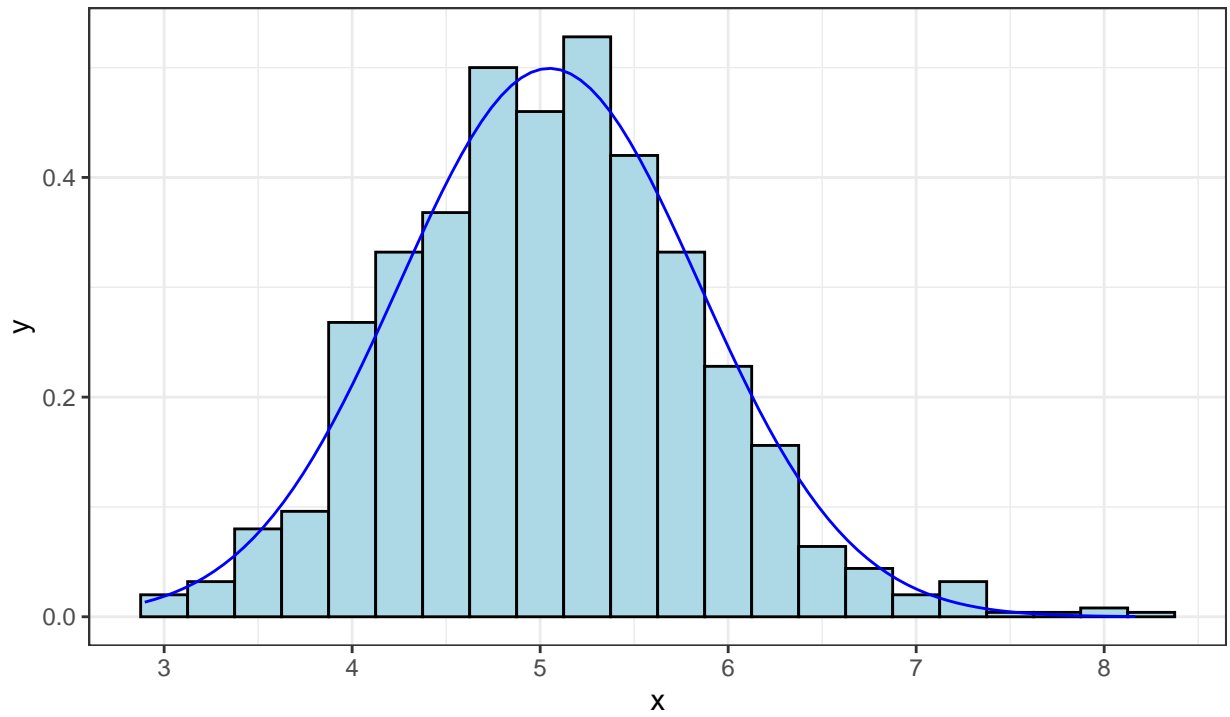Produced by Carlos Hernández

The theoretical variance is: 0.625 and the sample variance is: 0.638.

Distribution: the distribution is approximately normal

In order to show graphically that this distribution is normally distributed, we can normalize it by its mean and standard deviation. This transforms it into a Standard Normal Distribution, N(0,1). So let's do that normalization, plot the histogram of the transformed distribution, and show how it compares to an exact Standard Normal Distribution, by overlaying it:

```
ggplot(data, aes(x=value)) +
  geom_histogram(aes(y=..density..),binwidth=.25, col="black", fill="lightblue") +
  stat_function(fun=dnorm,
                color="blue",
                args=list(mean=mean(data$value),
                          sd=sd(data$value)))+
  labs(title= "Average of 40 random exponential distribution",
       subtitle = "Replicated 1000 times",
       caption="Produced by Carlos Hernández") +
  xlab("x") +
  ylab("y")
```

Average of 40 random exponential distribution
Replicated 1000 times