

Simulation Project

Juan Mari Sebastian Carino

October 20, 2018

Overview

The author will simulate 1000 observations and try to investigate the exponential distribution in R and compare these to the Central Limit Theorem. Under this simulation project, the author will show the sample mean, sample variance, and the histogram of distribution and compare it with the theoretical mean, theoretical variance, and uniform distribution.

Simulation

In the simulation process, I will generate 10000 observations under the Exponential distribution. Before dwelling to the simulation, let us define Exponential distribution. Exponential distribution is the probability distribution that describes time between events in a Poisson process. Poisson process gives you a way to find probabilities for random points in time for a process. It can tell you when one of these random points in time will likely happen. Exponential distribution is mostly used for testing product reliability such as how much time will go by before a major hurricane hits the Atlantic Seaboard or how long will the transmission in my car last before it breaks.

Adapted from statistics How To:

(<https://www.statisticshowto.datasciencecentral.com/exponential-distribution/>)

```
# Load the library
```

```
library(ggplot2)
```

```
# Summarizing the variables
```

```
n <- 10000
```

```
lambda <- 0.2
```

```
msd <- 1/lambda
```

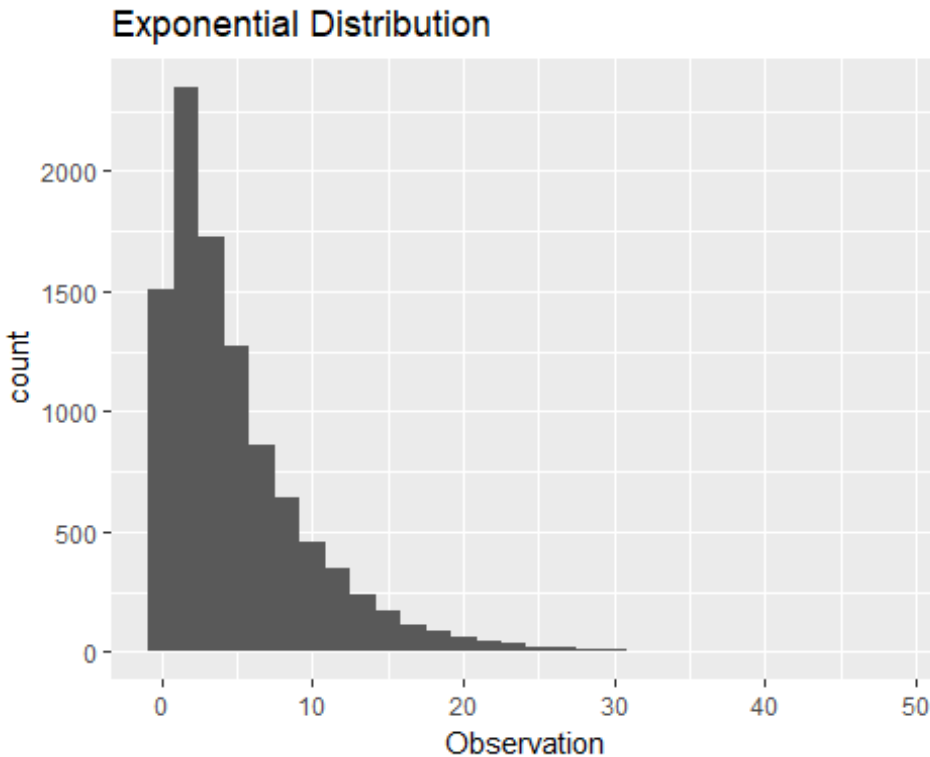
```
obs <- rexp(n, lambda)
```

```
# Setting up the illustration
```

```
g <- qplot(obs)
```

```
g + xlab("Observation") + ggtitle("Exponential Distribution")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



It can be observed from the graph that the Exponential Distribution is downward. This graph shows the probability density function where the x-axis is the values of 1000 observations and y-axis is the count or frequency. The rate we used which is 0.20 is constant meaning the events or observations occur at this constant rate.

Sample Mean vs Theoretical Mean

Now, we are going to compare the sample mean of our observations to the theoretical mean. Recall that the mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. In our simulation, λ is 0.20. The sample mean of our simulation is 5.0836374. The theoretical mean is 5 which is the mean of exponential distribution $1/\lambda$. Our mean is close enough to the theoretical mean proving that the simulation is close to the true population.

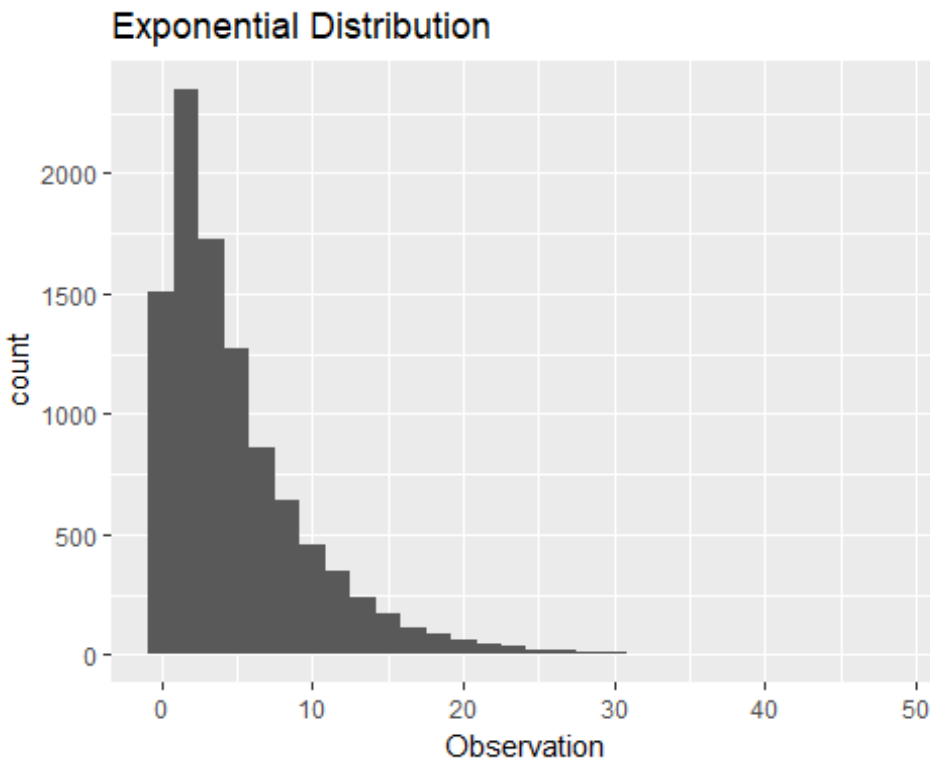
Sample Variance vs Theoretical Variance

Next, we are going to compare our sample variance and theoretical variance. Recall that the standard deviation is $1/\lambda$ similar to the mean. Then, the variance is square of the standard deviation or mean. In our simulation, the sample variance is 25.8433687 and the theoretical variance is 25. Like our sample mean, our sample variance is close to the theoretical variance.

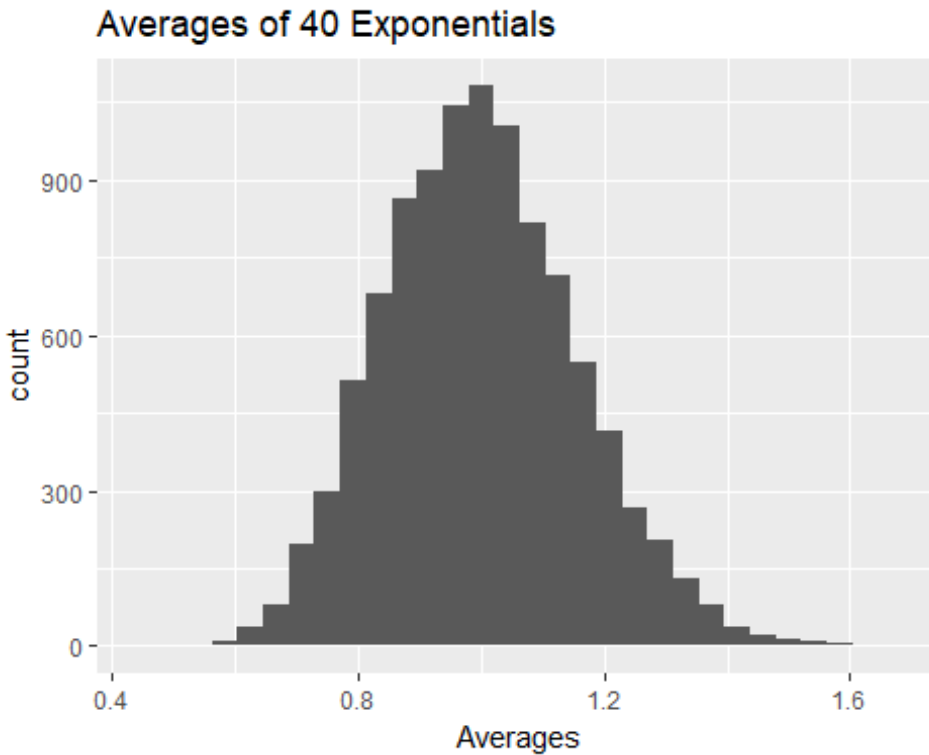
Distribution

Lastly, we will show that the distribution of our simulation is normal. First, we say that the distribution is normal if it follows the Gaussian distribution (with bell-shaped curve).

```
# Simulation
g <- qplot(obs)
g + xlab("Observation") + ggtitle("Exponential Distribution")
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
# Distribution of a large collection of averages of 40 exponentials
mns = NULL
for(i in 1:n) mns = c(mns, mean(rexp(40)))
g <- qplot(mns)
g + xlab("Averages") + ggtitle("Averages of 40 Exponentials")
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



In the first graph, it can be seen that the distribution of our simulation is not a bell-shaped curve figure because the exponential distribution has constant rate which is λ (or rate) in which the curve of our graph decreases. As we increase our λ , we also decrease the mean of the distribution. However, when we average 40 exponentials, the figure becomes more of like a bell-curve shape. Averaging means that we are calculating the midpoint of 40 exponentials. Hence, when we graph the averages of 40 exponentials, the graph of 40 exponentials in 1000 observation becomes similar to the bell curve or Gaussian distribution. This is consistent to the Central Limit Theorem in which it states that the distribution of averages of iid variables (properly normalized) becomes that of a standard normal as the sample size increases.