Investigate the Exponential Distribution in R and compare it with the Central Limit Theorem

Charanjeet Dhawan

16 September 2018

Overview

Here we will investigate the exponential distribution and compare it to the Central Limit Theorem. 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.

Simulation

Set the simulation variables λ, exponentials, seed and number of simulation(noSim)

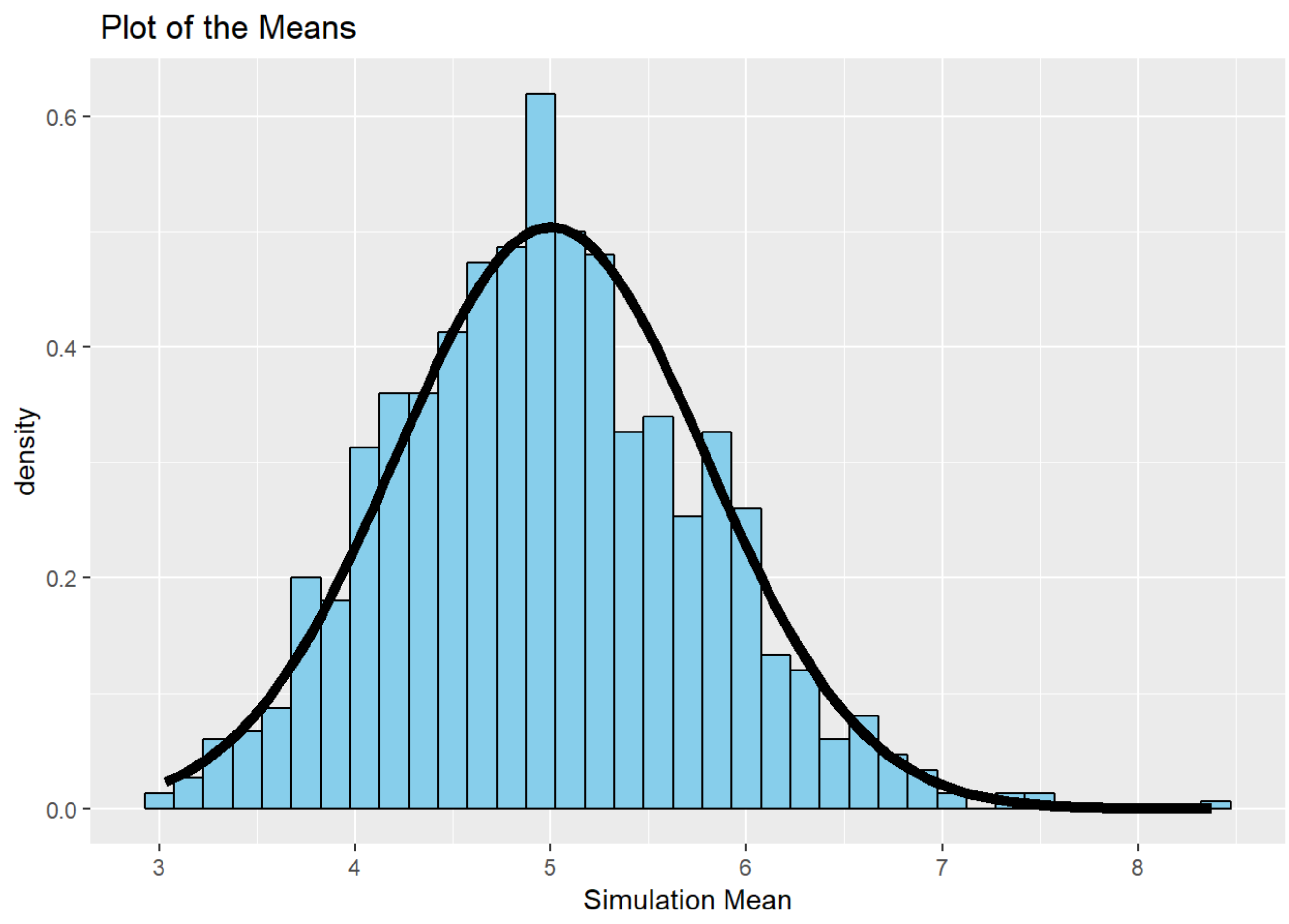
```
ECHO=TRUE
set.seed(12345)
lambda = 0.2
exponentials = 40
noSim = 1000

exponentialDistributions <- matrix(data = rexp(exponentials*noSim, lambda), nrow = noSim)
exponentialDistributionMeans <- data.frame(means = apply(exponentialDistributions, 1, mean))
```

```
library(ggplot2)

# Plot the Means

ggplot(data = exponentialDistributionMeans, aes(x = means)) +
    geom_histogram(aes(y= ..density..), binwidth=0.15, fill = "skyblue", color = "black") +
    stat_function(fun = dnorm, args = list(mean = lambda^-1, sd = (lambda*sqrt(exponentials))^-1), size = 2) +
    labs(title =" Plot of the Means", x = "Simulation Mean")
```



Q.1 Theoretical Means VS. Sample Means

The expected mean μ of a exponential distribution of rate λ is

$\mu = \frac{1}{\lambda}$

```
mu <- 1/lambda
mu
```

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

Let $\bar{X}$ be the average sample mean of 1000 simulations of 40 randomly sampled exponential distributions.

```
meanOfMeans <- mean(exponentialDistributionMeans$means)
meanOfMeans
```

```
## [1] 4.971972
```

Comparison

Here we can see the expected mean and the avarage sample mean are very close.

```
round(abs( meanOfMeans - mu), 4)
```

```
## [1] 0.028
```

Q.2 Theoretical Variance VS. Sample Variance

The expected standard deviation σ of a exponential distribution of rate λ is

$\sigma = \frac{1/\lambda}{\sqrt{no.ofexponentials}}$

```
sd <- sd <- 1/lambda/sqrt(exponentials)
sd
```

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

The variance Var of standard deviation σ is

$Var = \sigma^2$

```
Var <- sd^2
Var
```

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

Let Var_x be the variance of the average sample mean of 1000 simulations of 40 randomly sampled exponential distribution, and σ_x the corresponding standard deviation.

```
sd_x <- sd(exponentialDistributionMeans$means)
sd_x
```

```
## [1] 0.7847246
```

```
Var_x <- var(exponentialDistributionMeans$means)
Var_x
```

```
## [1] 0.6157926
```

Comparison

Here we can see that there is only a slight difference between the simulations sample variance and the exponential distribution theoretical variance.

```
round(abs(Var_x - Var),4)
```

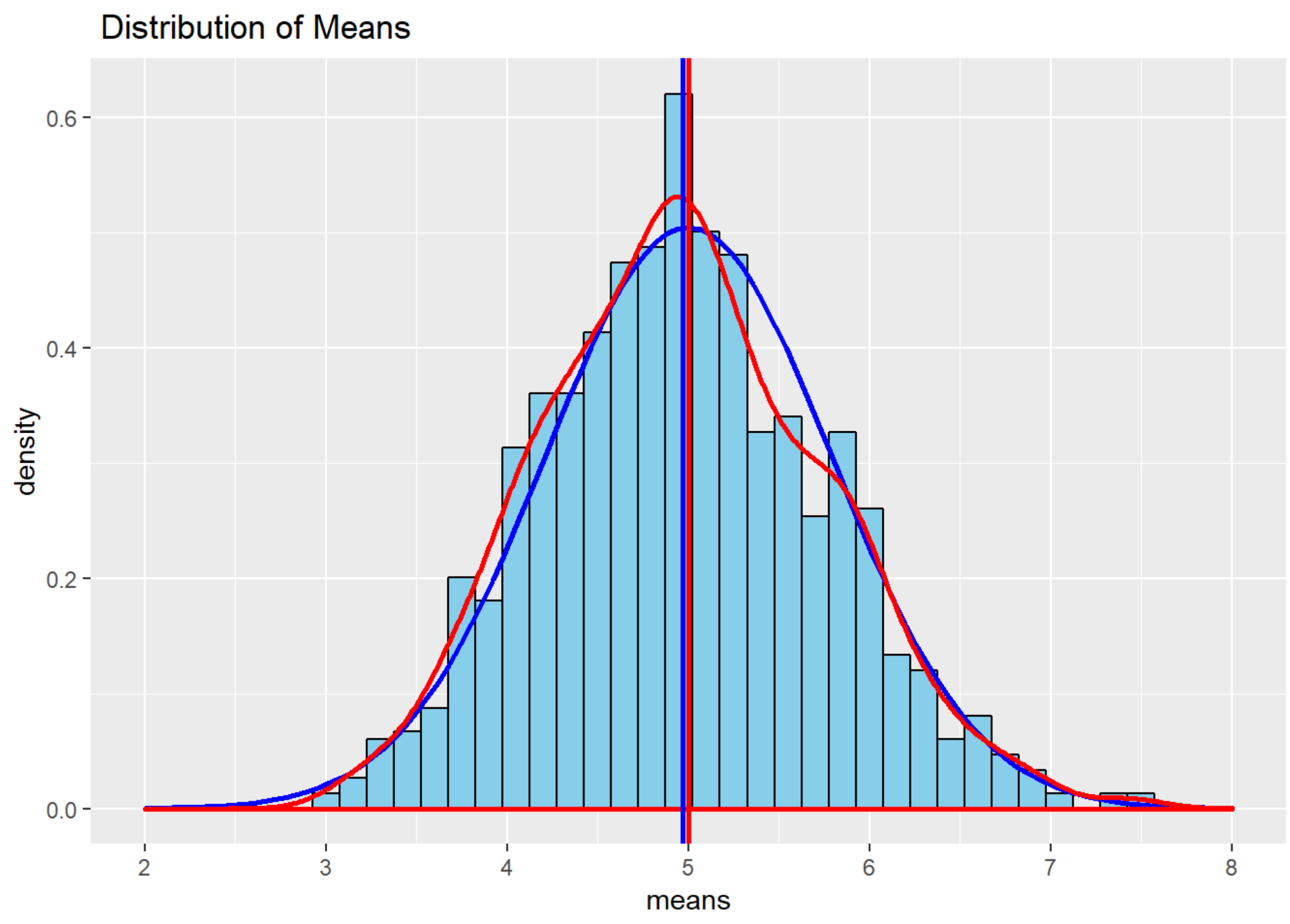
```
## [1] 0.0092
```

Q.3 Distribution

Comparing the population means & standard deviation with a normal distribution of the expected values. Added lines for the calculated and expected means

```
library(ggplot2)
options(warn = -1)

ggplot(data = exponentialDistributionMeans, aes(x = means)) +
  geom_histogram(binwidth=0.15, aes(y=..density..), fill = "skyblue", color = "black") +
  stat_function(fun = dnorm, args = list(mean = lambda^-1 , sd = (lambda*sqrt(exponentials))^-1), colour = "blu
e", size=1) +
  geom_density(color = "red",size = 1) +
  geom_vline(xintercept = mu, size=1, colour="red") +
  geom_vline(xintercept = meanOfMeans, size=1, colour="blue") +
  scale_x_continuous(breaks=seq(mu-3,mu+3,1), limits=c(mu-3,mu+3)) +
  labs(title =" Distribution of Means", x = "means")
```



Here we can see from the graph, the calculated distribution of means of random sampled exponantial distributions, overlaps quite nice with the normal distribution with the expected values based on the given lamba.

So we can say that the distribution is approximately normal.