Comparative Analysis of Exponential Distributions and Properties of the Central Limit Theorem

J Faleiro

April 17, 2016

Overview

This is an investigation of the exponential distribution through a basic comparison with the theoretical expectations of the Central Limit Theorem.

Exponential Distributions

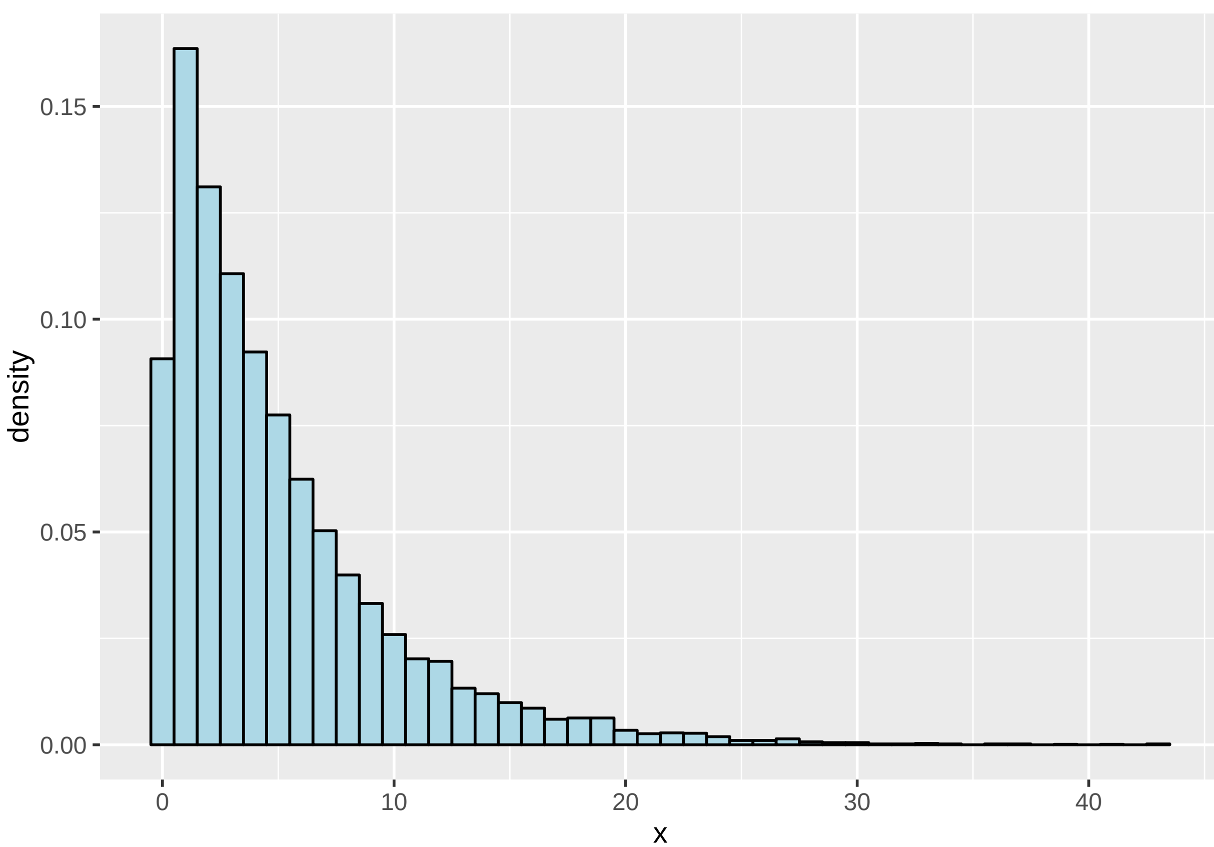
The exponential distribution can be simulated with n and λ as arguments where λ is the rate parameter.

The mean and the standard deviation of a exponential distribution are given by $\frac{1}{\lambda}$. The probability density function for a exponential distribution is given by:

$f(x) = \frac{1}{\lambda}e^{-x/\lambda}$ for $x >= 0$, or

$f(x) = 0$ for $x < 0$

```
lambda <- 0.2
x <- rexp(10000, 0.2)
ggplot(data=data.frame(x=x), aes(x=x)) +
    geom_histogram(aes(y= ..density..), binwidth=1, fill='lightblue', colour='black')
```



According to theory, the mean and the standard deviation should be close to $\frac{1}{\lambda}$:

```
c(1/lambda, mean(x), sd(x))
```

```
## [1] 5.000000 5.034855 4.984569
```

Central Limit Theorem

The Central Limit Theorem simply states that "if all possible random samples of size n are drawn from a population with mean μ_0 and a standard deviation σ_0, then as n becomes larger, the sampling distribution of sample means becomes approximately normal $N(\mu_0, \frac{{\sigma_0}^2}{n})$".

Simulations

The simulation will investigate the distribution of averages of **40** exponentials and **100**, **1000** and **10000** simulations of random exponential distributions with $\lambda = 0.2$:

```
simulate <-function(exponentials, simulations, lambda) {
    simulation <- NULL
    for (i in 1:simulations) {
        simulation <- c(simulation, mean(rexp(exponentials, lambda)))
    }
    simulation
}
n <- 40
mns100 <- simulate(exponentials=n, simulations=100, lambda=lambda)
mns1000 <- simulate(exponentials=n, simulations=1000, lambda=lambda)
mns10000 <- simulate(exponentials=n, simulations=10000, lambda=lambda)
```

According to the Central Limit Theory, the mean should match a normal distribution $N(\mu_0, \frac{{\sigma_0}^2}{n})$ where μ_0 is the population mean, σ_0 is the population standard deviation and n the sample size.

Sample Mean vs Theoretical Mean

According to the Central Limit Theorem, the mean of the sum of averages should match the population mean $\frac{1}{\lambda}$, and here are how the theoretical and empirical value through simulations compare

```
c(1/lambda, mean(mns100), mean(mns1000), mean(mns10000))
```

```
## [1] 5.000000 5.038085 5.043552 5.003331
```

We can see that the mean of the sum of means is approximatelly the theoretical mean of the distribution, and the higher n the closer it gets to the theoretical expected value.

Sample Variance vs Theoretical Variance

The Central Limit Theorem states that the variance of the sum of means is $\frac{{\sigma_0}^2}{n}$ where σ_0 is the population standard deviation (what for a exponential distribution is $\frac{1}{\lambda}$) and n the sample size.

With that, the theoretical standard deviation of the sum of the means of a exponential distribution will be then $\frac{1}{\lambda\sqrt{n}}$:

```r
c((1/lambda)/sqrt(n), sd(mns100), sd(mns1000), sd(mns10000))
```

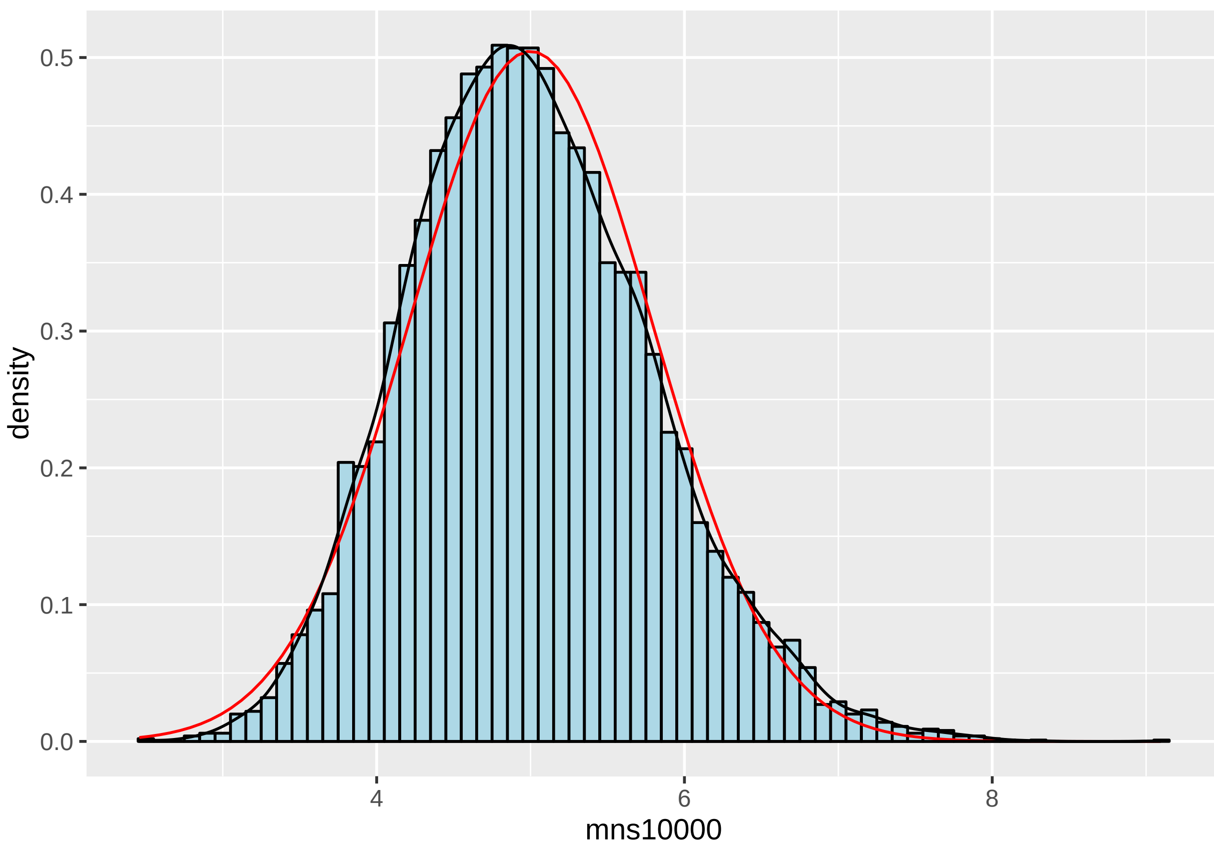
```
## [1] 0.7905694 0.7548780 0.7715328 0.7945705
```

We can see that the variance of the normal distribution of sum of means is matches the expected theoretical variance of the distribution, and the higher n the closer it gets to the theoretical expected value.

Distribution

As we increase n we will have the distribution of the sum of means matching more and more a normal distribution $N(\frac{1}{\lambda}, \frac{1/\lambda^2}{n})$ and we can visually inspect that by overlapping all distributions:

```r
ggplot(data=data.frame(x=mns10000), aes(x=mns10000)) +
    geom_histogram(aes(y= ..density..), binwidth=0.1, fill='lightblue', colour='black') +
    stat_function(geom='line', fun=dnorm, args=list(mean=1/lambda, sd=(1/lambda)/sqrt(n)), colour='red')
    geom_density(colour='black')
```



As a final synthesis of this analysis, we can see that if we plot a normal approximation (in black) to the distribution of the sum of means for $n = 10000$ in lightblue, it almost overlaps a normal distribution $N(\frac{1}{\lambda}, \frac{1/\lambda^2}{n})$ in red.