

Investigation of Exponential Distribution

Anuj Parashar

7 July 2016

Overview

In this report, we will be investigating the exponential distribution which is characterised by the rate parameter λ . We will be doing number of simulations on the distribution of averages of exponentials. Finally, the result will be compared against the CLT which will show that the distribution becomes more gaussian (normal) as the number of simulation increases.

Understanding Exponential Distributions

The exponential distribution can be simulated in R with $\text{rexp}(n, \lambda)$ where n is the sample size and λ is the rate parameter.

The λ is set to 0.2 for all the simulations. Also, we are taking average of 40 exponentials and going to perform 1000 simulation. So:

```
n <- 40          # Sample Size
lambda <- 0.2     # Rate Parameter
nosim <- 1000     # Number of Simulations
```

It's also known that the population mean and the standard deviation of the exponential distribution is $1/\lambda$. Let's calculate these values as well:

```
pop_mean <- 1/lambda # Population Mean (Theoretical)
pop_sd <- 1/lambda # Population Standard deviation
pop_var <- (1/lambda**2)/n # Population Variance (Assuming population size = sample size = 40)
```

```
## Population Mean: 5
```

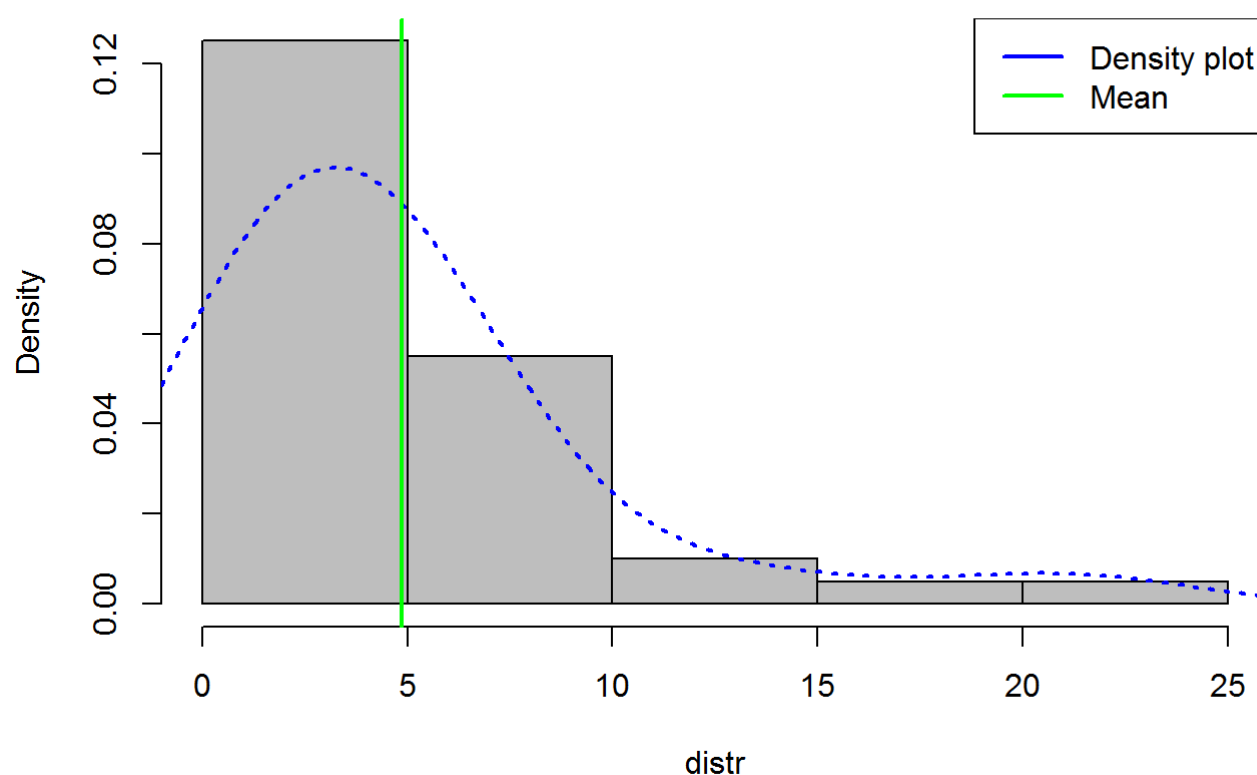
```
## Population Standard Deviation: 5
```

```
## Population Variance: 0.625
```

To see the exponential distribution in action, let's plot a histogram and it's density curve:

```
set.seed(1) # For reproducibility
distr <- rexp(n, lambda)
hist(distr, prob = TRUE, col="grey")
lines(density(distr, adjust=2), col="blue", lwd=2, lty="dotted")
abline(v = mean(distr), col = "green", lwd = 2)
legend(x = "topright", c("Density plot", "Mean"), col = c("blue", "green"), lwd = c(2, 2))
```

Histogram of distr



```
## Sample (n=40) Mean = 4.860372
```

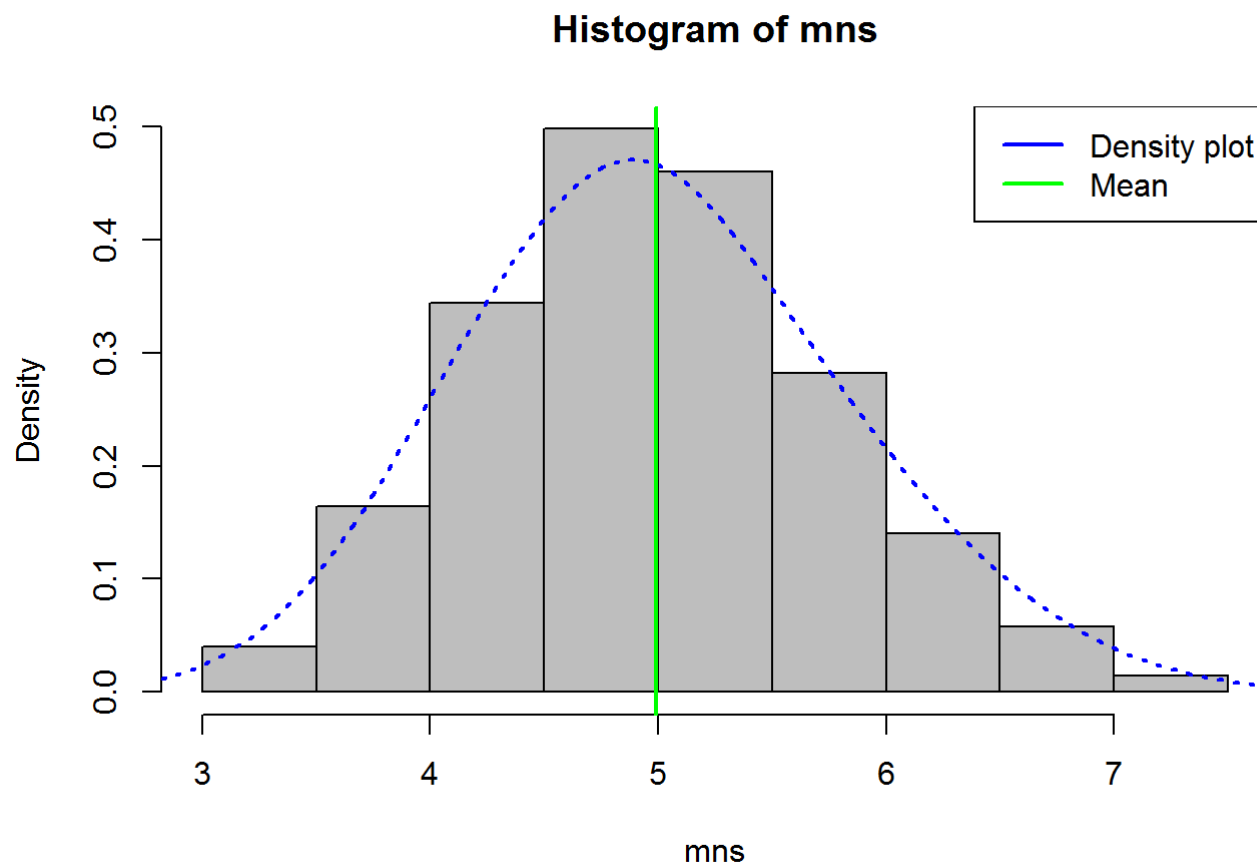
```
## Sample Standard Deviation = 4.860372
```

```
## Sample Median: 3.691892
```

Simulating distribution of mean of exponentials

We now move to the next step where we perform 1000 simulations on the distribution of mean of exponentials of sample size 40. By Central Limit Theorem (CLT), this simulated distribution should be more like a normal distribution which sample mean converging to the actual population mean. let's do this now:

```
mns = NULL # Vector of sample means for different simulations
for (i in 1 : nosim) mns <- c(mns, mean(rexp(n, lambda)))
hist(mns, prob = TRUE, col="grey")
lines(density(mns, adjust=2), col="blue", lwd=2, lty="dotted")
abline(v = mean(mns), col = "green", lwd = 2)
legend(x = "topright", c("Density plot", "Mean"), col = c("blue", "green"), lwd = c(2, 2))
```



```
## Sample Mean of the Simulated Distribution (average of exponentials):  4.989337
```

```
## Sample Standard Deviation of the Simulated Distribution:  4.989337
```

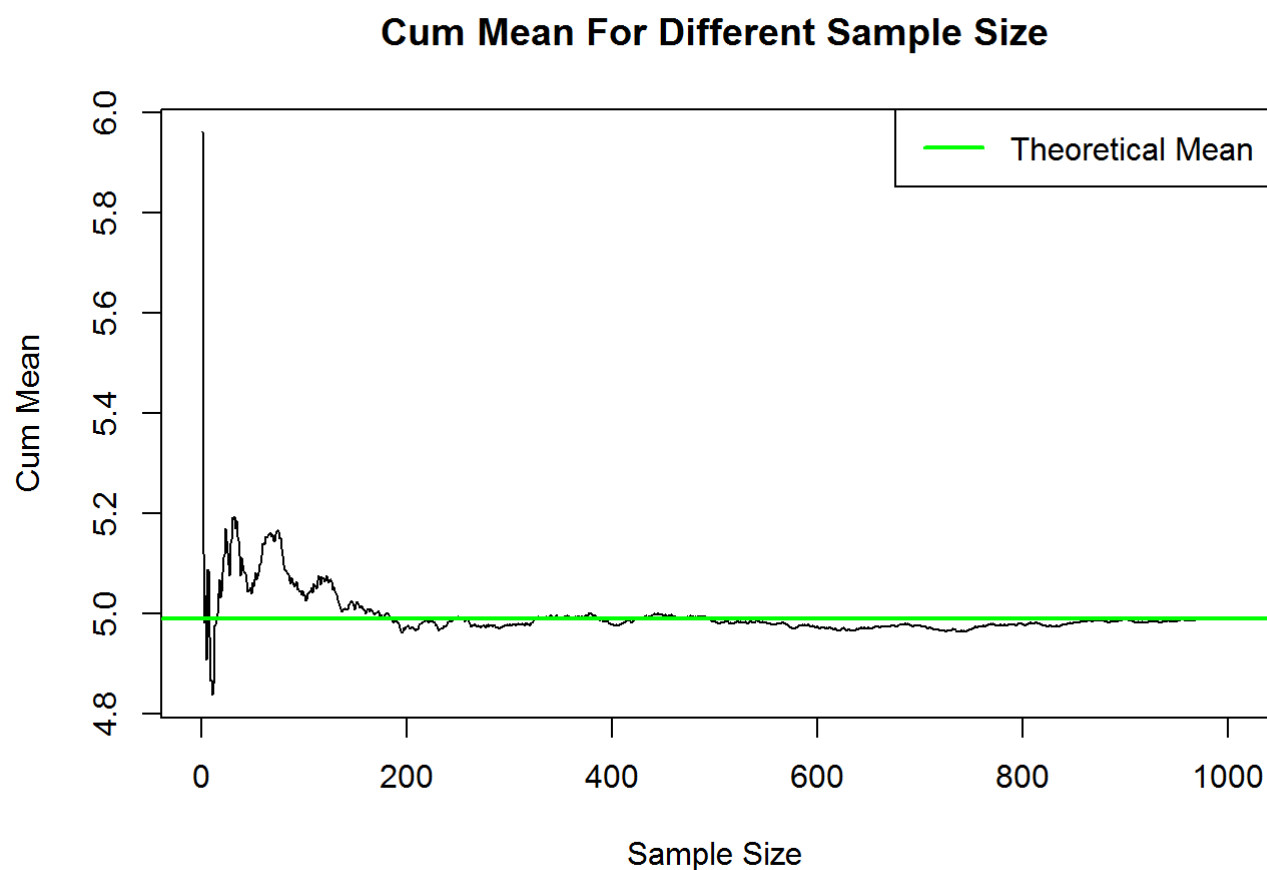
```
## Sample Median: of the Simulated Distribution 4.950169
```

This looks far more gaussian, we will empirically verify this in a while.

Sample Mean versus Theoretical Mean

Another way of looking at these distributions is to plot the change in sample mean as the number of simulation increases while keeping the sample size and rate constant. Again, the power of R comes handy in form of `cumsum()` function which can be divided by the sample range for create a vector to hold average for each range. let's do that:

```
cum_mean <- cumsum(mns)/(1:nosim)
plot(x = cum_mean, type = "l", main = "Cum Mean For Different Sample Size", xlab = "Sample Size", ylab = "Cum Mean")
abline(h = mean(mns), col = "green", lwd = 2)
legend(x = "topright", c("Theoretical Mean"), col = c("green"), lwd = c(2, 2))
```



It's clearly evident that as the sample size increases, the sample mean converges to the theoretical mean.

Sample Variance versus Theoretical Variance

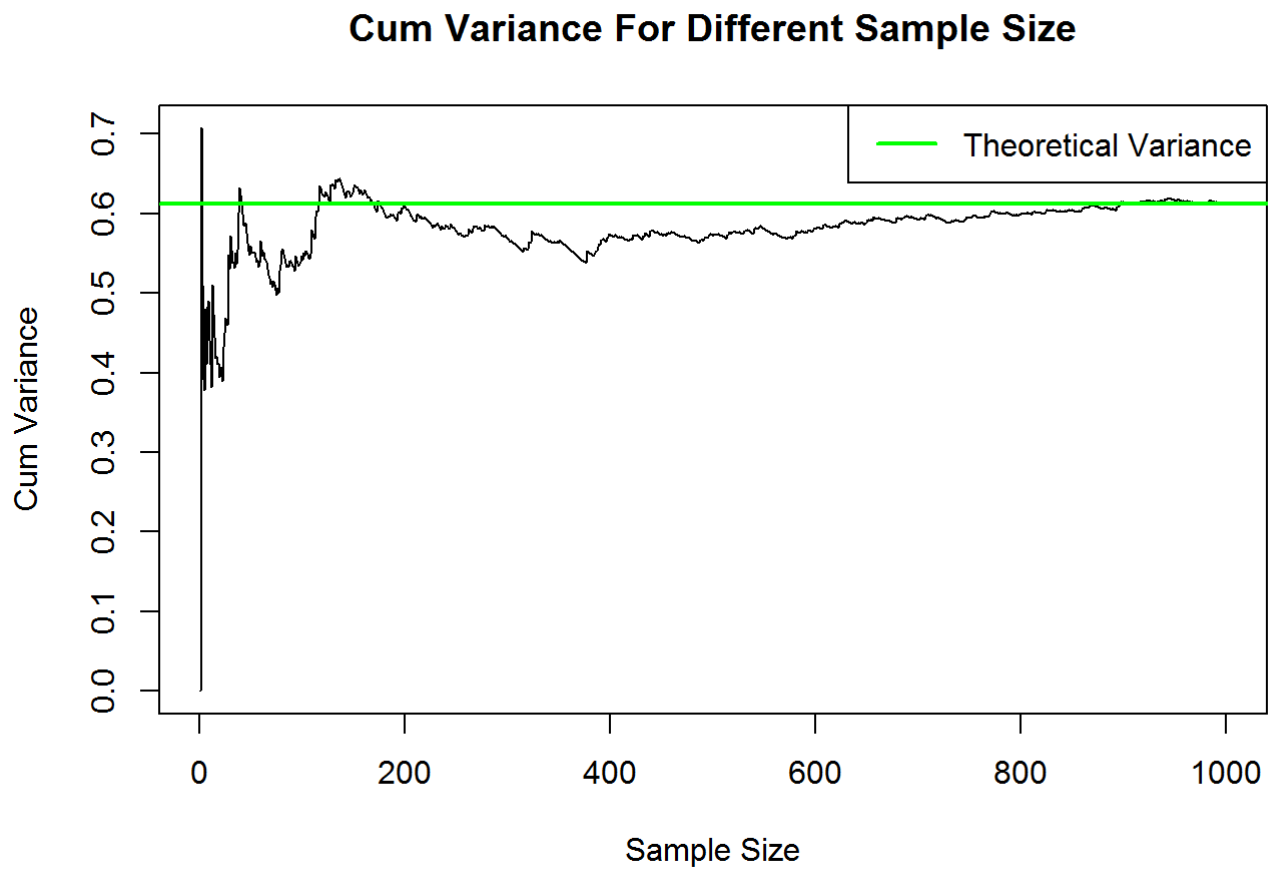
A similar graph can be plotted to show the correlation between that sample variance and the theoretial variance. By virtue of CTL, we already know that the sample variance converges to theoreical variance for large sample size. Let's see it graphically:

We already know that for any random variable X with mean μ the variance of X is defined as:

$$Var(X) = E[(X - \mu)^2] = E[X^2] - E[X]^2$$

Let's create a vector and plot it against the sample size based on this:

```
cum_var <- cumsum(mns**2)/(1:nosim) - cum_mean**2
plot(x = cum_var, type = "l", main = "Cum Variance For Different Sample Size", xlab = "Sample Size", ylab = "Cum Variance")
abline(h = var(mns), col = "green", lwd = 2)
legend(x = "topright", c("Theoretical Variance"), col = c("green"), lwd = c(2, 2))
```



No surprises, huh!

Checking the normality of the resulting distribution

Let's revisit the part where we plotted the distribution of averages of means. The following observations can be made about it:

1. By looking at the histogram and it's density function which is forming a bell curve (charaterizing a normal distribution), it is evident that the resulting distribution is approximately normal.
2. We also know that for a normal distribution, it's mean and median are equal. In this case:

sample mean ~= sample median (calculated above)

Based on the above facts, we can safely assume that the resulting distribution is approximatel normal.