# An investigation of the exponential distribution

*Danny Andersen*

*17 October 2015*

## Overview

The R stats library contains a number of statistical distributions. This paper investigates the characteristics of the exponential distribution, whose probability density function is given by:

$$f(x) = \lambda e^{-\lambda x}$$

for $x \geq 0$.

The intent of this paper is to show for the exponential distribution that the sample mean and variance is equivalent to the population mean and variance. It will also show that it obeys the Central Limit Theorem (CLT) that states the mean and variance of a sample mean will be approximately normally distributed.

## Simulations

If we take a random sample of 40 values from an exponential population, the mean and variance of this sample set will be a random approximation of the true mean and variance of an exponential population.

Therefore if we repeat this simulation 1000 times, we should find the mean of all of the sample means to be very close to the theoretical mean of $1/\lambda$ and the variance very close to the theoretical variance of $1/\lambda^2$. The R code below shows how the simulations are created. Note that the random seed is set to ensure repeatability of these results.

```r
means <- NULL
vars <- NULL
nosim <- 1000
n <- 40
lambda <- 0.2
set.seed(1234)
#Run 1000 simulations of 40 samples, taking the mean of each one
for (i in 1:nosim) {
    sample <- rexp(n, lambda)
    means <- c(means, mean(sample))
    vars <- c(vars, var(sample))
}
```
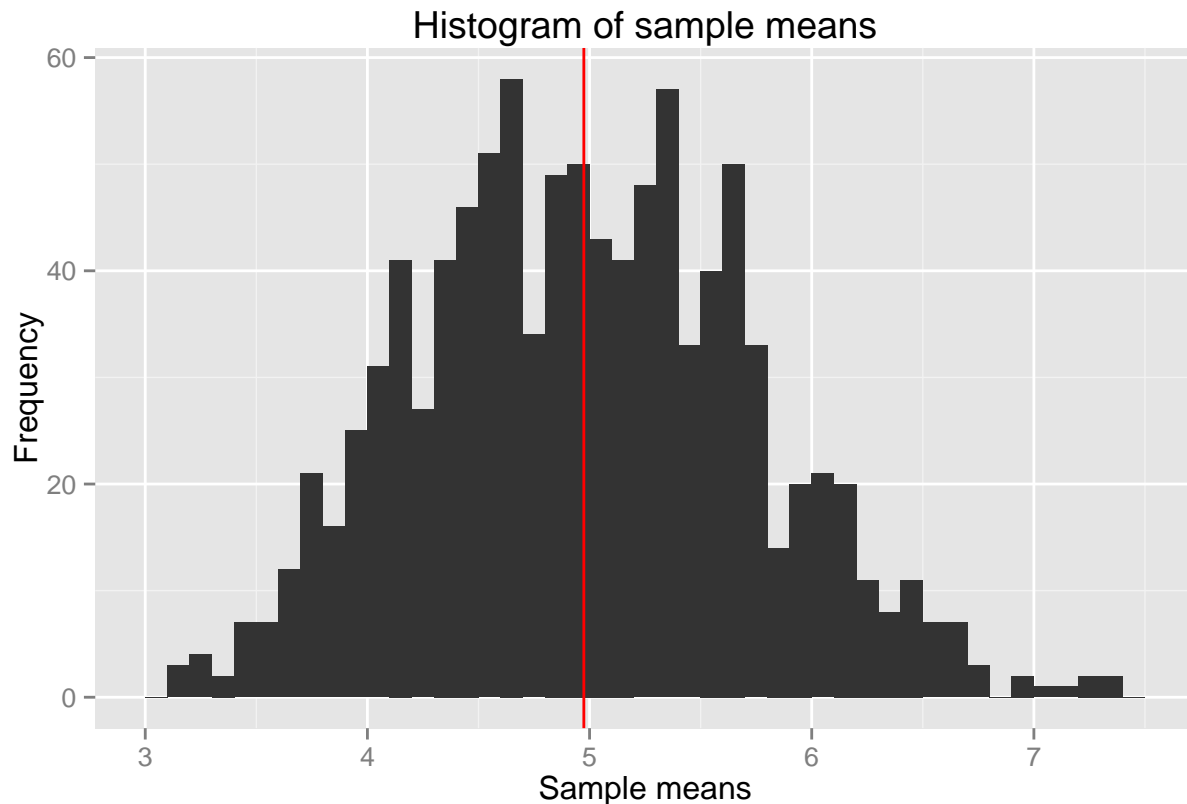
## Sample mean vs Theoretical mean

From this simulated data, we can calculate the average of all of the sample means, which according to the Central Limit Theorem, should approximate to the theoretical mean of $1/\lambda$.

```r
sampleMean <- mean(means)
mu <- 1/lambda
acc <- abs(sampleMean - mu)*100/mu
```

This gives us an overall sample mean of 4.97, which is very close to the theoretical mean of 5, i.e. within 0.52% .

The plot below shows a histogram of all the simulated 1000 sample means, with a red line showing the centre of the distribution, i.e. the mean.

## Histogram of sample means



The variance of the sample mean is $\sigma^2/n$, whose estimate is given by $S^2/n$.

```
meanVar <- (1/lambda^2)/n
sampleMeanVar <- var(means)
accMeanVar <- abs(sampleMeanVar - meanVar)*100/sampleMeanVar
```

In the simulations above the variance of the sample mean should theoretically be 0.62, whereas it was calculated to be 0.57, which is within 9.52%

## Sample variance vs Theoretical variance

From this simulated data, we can also calculate the variance of all of the sample variances, which according to the Central Limit Theorem, should approximate to the theoretical variance of lambda.

```
sampleVariance <- mean(vars)
sigmaSq <- (1/lambda)^2
accVar <- abs(sampleVariance - sigmaSq)*100/sigmaSq
```

This gives us an average sample variance of 24.38, which is very close to the theoretical variance of 25, i.e. within 2.49% . It is interesting to note that this is not as close as the sample mean is to the theoretical mean.

## Comparison with the standard normal distribution

The central limit theorem states that the distribution of iid variables becomes that of a standard normal as the sample size increases. This can be shown by scaling the sample mean to that of a uniform distribution by substracting the population mean and dividing by the standard deviation of the mean, i.e.

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

If we do this for all of our generated sample means:

```
#Scale generated means
sigma <- sqrt(sigmaSq)
scaler <- function(x) { (x-mu)/(sigma/sqrt(n)) }
smeans <- vapply(means, FUN=scaler, FUN.VALUE=c(1))
scaled_mu <- mean(smeans)
scaled_sigma <- sd(smeans)
```

we obtain a scaled mean of -0.033 which is very close to the normal distribution mean of 0 and a standard deviation of 0.956, which is close to the expected value of 1.

If we plot the histogram of the scaled means and compare it with the uniform distribution (shown as the black line) we find that there is a good match:
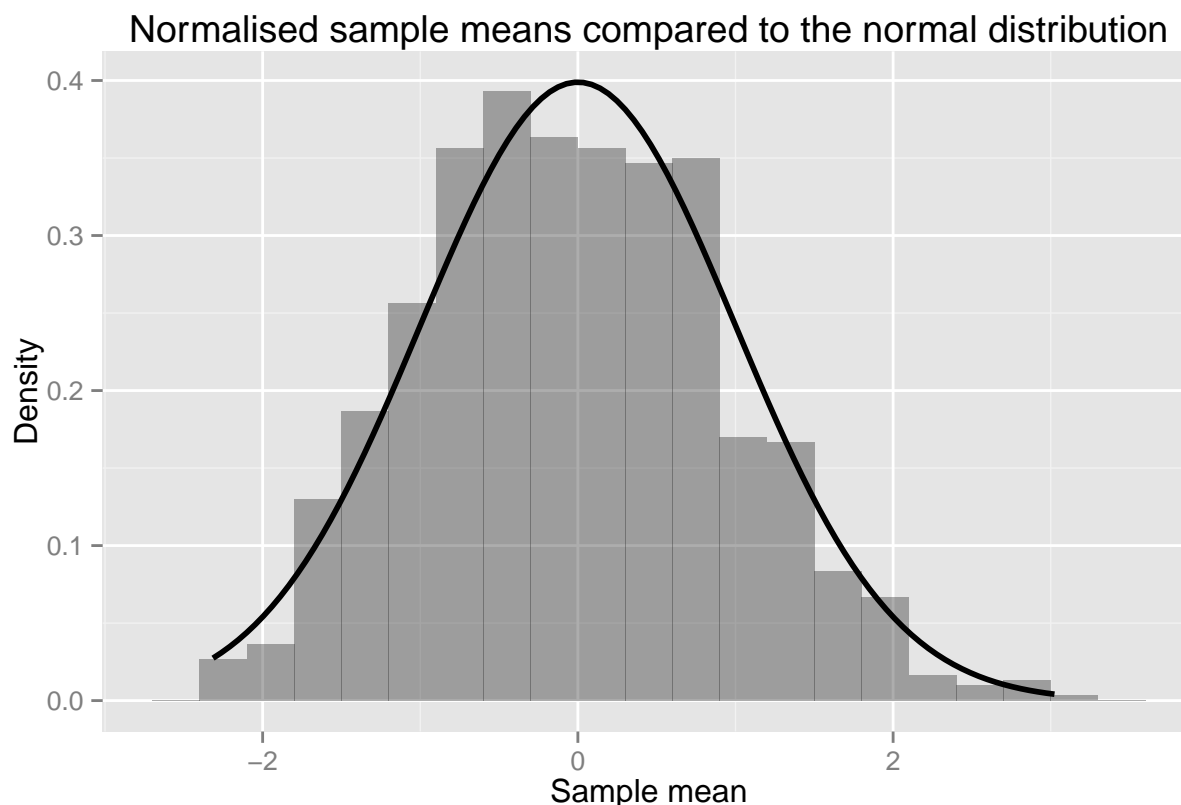


This demonstrates that the distribution of sample means closely resembles that of a normal distribution, as per the Central limit theorem.