# Exponential Distribution and the Central Limit Theorem

## Introduction

The Central Limit Theorem (CLT) lays the foundation for all statistical analysis and inference. Consider a population with a distribution described by mean `mu` and standard deviation `s`. The CLT says that any size `n` sample of independent, identically distributed random variables from that population will follow a normal distribution centered around `mu`, with variance approxiamtely `s/n`. This is a powerful tool because it makes no assumptions on the distribution of the population, only that we are working with independent and identically distributed samples, we can then make inferences about the population mean and variance.

## Exponential Distribution

The exponential distribution is a continuous probability model that describes the time between events in a Poisson process. It is described by a rate, `lambda`, which indicates the frequency of counts per unit of time. Take a look at a sample density function of the normal distribution in *Figure.1*.

The Exponential distribution has a special property which is that its mean and standard deviation are both equal to `1/lambda`. For the next few examples, lets consider an exponential distribution parametrized by `lambda = .2`. We are going to take repeated samples of size `40` from this exponential distribution and use these simulations to confirm a few things about the CLT.

## Expected Values

Let's use the CLT to test how accurate our samples are at uncovering the true mean of our population. Recal that in this specific distribution, `lambda = .2`, we know that the true mean of our population is `1/.2 = 5`. Let's first take a look at one sample of size 40.

```
lambda <- .2
n <- 40

truemu <-1/lambda
truesd <- 1/lambda
sample <- rexp(n, lambda)
data.frame("Sample Mean" = mean(sample), "True Mean" = truemu)
```

```
##   Sample.Mean True.Mean
## 1    3.565115         5
```

So we see in *Figure.2* that our 40-sized sample gives us a rough representation of the population's true mean, but it could be much more accurate.

## Variance

We can do a very similar analysis on the variance of our distribution. We will use the same sample of size 40 from before. Take a look at the variance of our original sample and how that differs from the what we know the population variance to be, `(1/.2)^2`.

```
truesd <- 1/lambda
data.frame("Sample Std.Dev" = sd(sample), "True Std.Dev" = truesd)
```

```
##   Sample.Std.Dev True.Std.Dev
## 1       4.258313            5
```

So, again, we see that with a sample size of 40, our estimate of the population standard deviation is close, but again, we can do better.

## Distribution

Remember that the mean of our 40-sized sample is itslef a random variable and the CLT tells us that this sample mean follows a normal distribution centered at the true mean and that the variance of the distribution is equal to `s^2/n`. So, lets simulate 1000 of these random variables by taking the average of 40 samples of an exponential distribution. We will then test to see if the empirical findings of the sample mean distribution matches what the CLT tells us we should find.

```
nosim <- 1000
samplemus <- apply(matrix(rexp(nosim*n, lambda), nosim), 1, mean)
truesamplesd <- sd(samplemus)/sqrt(n)


data.frame("Mean of 1000 Samples" = mean(samplemus), "True Mean" = truemu)
```

```
##   Mean.of.1000.Samples True.Mean
## 1             5.017265         5
```

```
data.frame("Std. Dev. of 1000 Samples" = sd(samplemus), "True Std. Dev." = truesamplesd)
```

```
##   Std..Dev..of.1000.Samples True.Std..Dev.
## 1                  0.790857      0.1250455
```

To show that the distribution of sample averages is normal, let's look at a plot of our 1,000 simulations seen in *Figure.3*. Note that the distribution of our sample means is roughly normal, and the peak or center of that distribution seems to be located at 5, which is the true mean of our exponential distribution.

**Appendix**

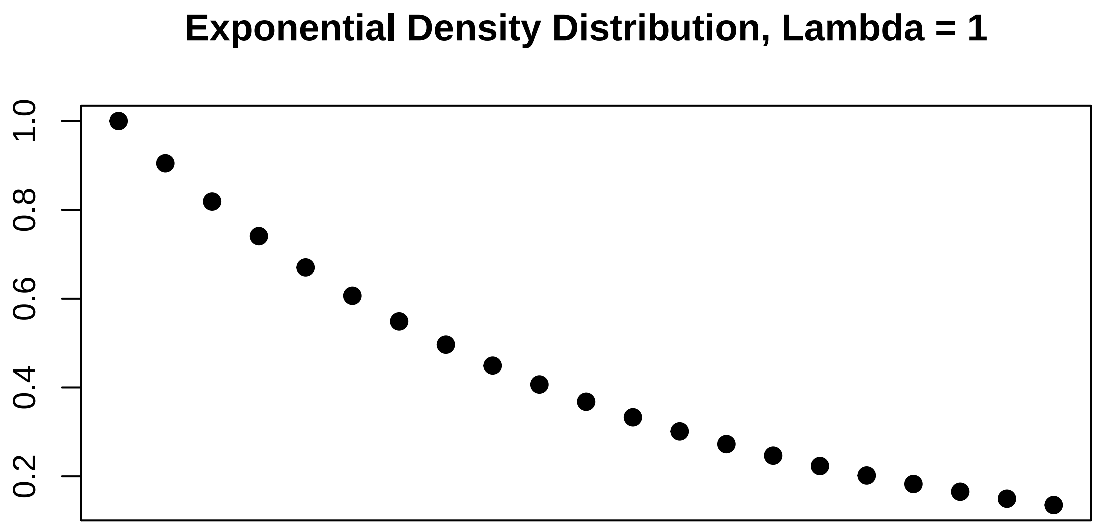


Figure.1

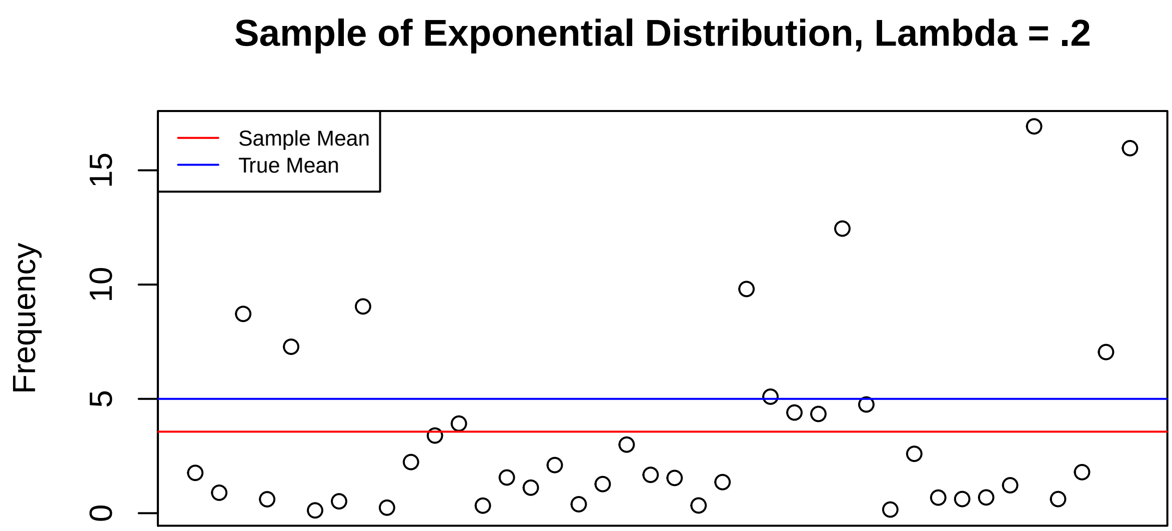


Figure.2

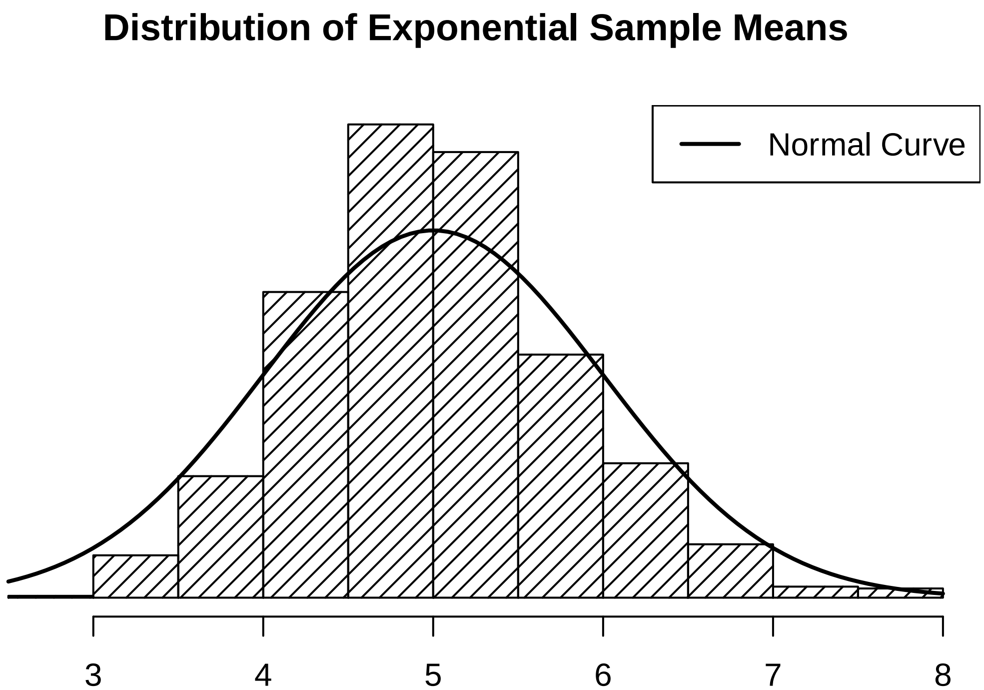
Distribution of Exponential Sample Means
Normal Curve
3
4
5
6
7
8

Figure.3