# An Examination of the Exponential Distribution

*Clinton Olson*

*May 7, 2016*

## Overview

In this report, we examine the properties of the Exponential distribution and compare it with the Central Limit Theorem (CLT). In particular, we will look at the distribution of the averages of 40 samples from the exponential distribution through simulation and compare this with the expectation under CLT that repeated sampling will yield a distribution of averages that is approximately normal.

## Simulations

We conduct our simulations by taking 40 samples from exponential distribution and repeating that process 1000 times. We will then look at the behavior of the distribution of the averages of each of these sample sets as we increase the number of simulations.

It is simple to perform the simulations. We will simply take the necessary number of samples, 40,000, and package them into a matrix that is 1000 X 40, where each row will represent a set of 40 samples from the exponential distribution.

```
set.seed(1)
exponentials <- matrix(rexp(40000, rate = .2), nrow = 1000, ncol = 40)
```

With these samples in hand, we can now perform our comparisons.

## Comparison

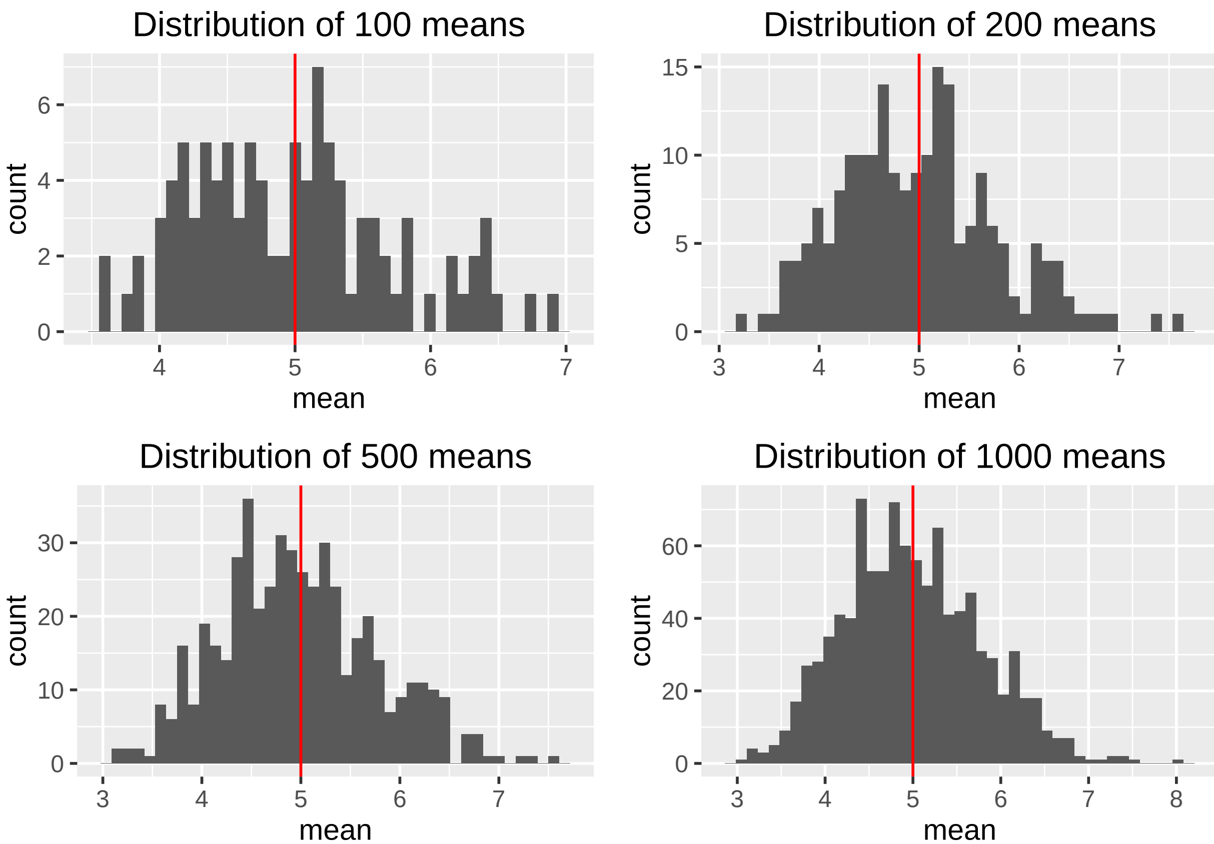
### Mean: Sample vs. Theoretical

The theoretical mean for the exponential distribution is 1/lamda. Here we are using a lambda of 0.2 for our simulations so we should expect that the mean of our averages should converge on that value as the number of our simulations increases. Lets take a look how the distribution of means changes as we increase our number of simulations. We can see numerically that this is the case:

```
means <- apply(exponentials, 1, mean)
c(mean(means[1:100]), mean(means[1:200]), mean(means[1:500]), mean(means))
```

```
## [1] 4.970711 4.986703 4.973884 4.990025
```

We can see that the mean of the averages converges to the theoretical value of 5 with increasing numbers of simulations. Lets look at the plots as well.




As expected, the distribution of averages becomes more peaked and centered around the true mean of 5 for increasing numbers of simulations.

### Variance: Sample vs. Theoretical

We can perform the same comparisons for the variance in the data as well. The theoretical variance of the exponential is 1 / lambda^2, so we should expect to see a convergence around the value of 0.625 since the variance of our mean estimate will be (1/lambda^2) / n.

```
c(var(means[1:100]), var(means[1:200]), var(means[1:500]), var(means))
```

```
## [1] 0.5574100 0.6240786 0.6161219 0.6177072
```

Again, we can see that the distributions become more centered around the true value of the variance with increasing number of samples.

### Distribution: Comparing to the Gaussian

The CLT tells us that as our number of simulations of mean and variance samples from the exponential distribution increases, it should be approximately normal. We can see this is true by examining the above plots. As the number of samples of averages increases, the distributions of averages becomes more normal in appearance.

# Appendix

## Code for plots

```
# plots for mean distributions
p1 <- ggplot() + geom_histogram(aes(means[1:100]), bins = 40) +
    labs(x = "mean", y = "count", title = "Distribution of 100 means") +
    geom_vline(xintercept=5, col='red')

p2 <- ggplot() + geom_histogram(aes(means[1:200]), bins = 40) +
    labs(x = "mean", y = "count", title = "Distribution of 200 means") +
    geom_vline(xintercept=5, col='red')

p3 <- ggplot() + geom_histogram(aes(means[1:500]), bins = 40) +
    labs(x = "mean", y = "count", title = "Distribution of 500 means") +
    geom_vline(xintercept=5, col='red')

p4 <- ggplot() + geom_histogram(aes(means), bins = 40) +
    labs(x = "mean", y = "count", title = "Distribution of 1000 means") +
    geom_vline(xintercept=5, col='red')

grid.arrange(p1,p2,p3,p4, ncol=2)
```