# Statistical Inference of the Exponential Distribution

*Andrew J. Dyck*

## Overview

A course project for the Coursera Statistical Inference course. The following report is organized into the following sections:

1. Simulations
2. Sample Mean Vs. Theoretical Mean
3. Sample Variance Vs. Theoretical Variance
4. Simulation Distributions

## Simulations

In the following simulation exercise, we'll explore sample error by simulating 40 draws from the exponential distribution. Throughout the exercise we'll use $\lambda = 0.2$ when drawing from the distribution.

The following R code sets up these parameters as well as calculates the theoretical mean and variance for this distribution. The theoretical mean is defined as $\bar{x} = \lambda^{-1}$, and the theoretical variance is defined as $\sigma^2 = \lambda^{-2}$.

```
lambda <- 0.2
n <- 40
TheoreticalMean <- lambda^(-1)
TheoreticalVariance <- lambda^(-2)
```

The following R code performs the mean of 1000 random samples of 40 observations from the exponential distribution with the parameters defined above. The simulation results are saved to an R object for analysis in the next section.

```
SingleSample <- rexp(n, lambda)
ExponentialMeanSimulations <- sapply(1:1000, function(x) mean(rexp(n, lambda)))
ExponentialVarianceSimulations <- sapply(1:1000, function(x) var(rexp(n, lambda)))
```

## Sample Mean Vs. Theoretical Mean

After a single 40 random draw sample of the exponential distribution, the mean of this sample is 5.46, compared to the theoretical mean of 5.

A random draw of 40 observations is a fairly small sample, and this is the main reason that this sample mean differs from the theoretical mean.

After taking the mean of 1000 draws of 40 observations from the exponential distribution, the average mean is 5, compared to the theoretical mean of 5. Our sample mean is now much closer to the theoretical mean.

## Sample Variance Vs. Theoretical Variance

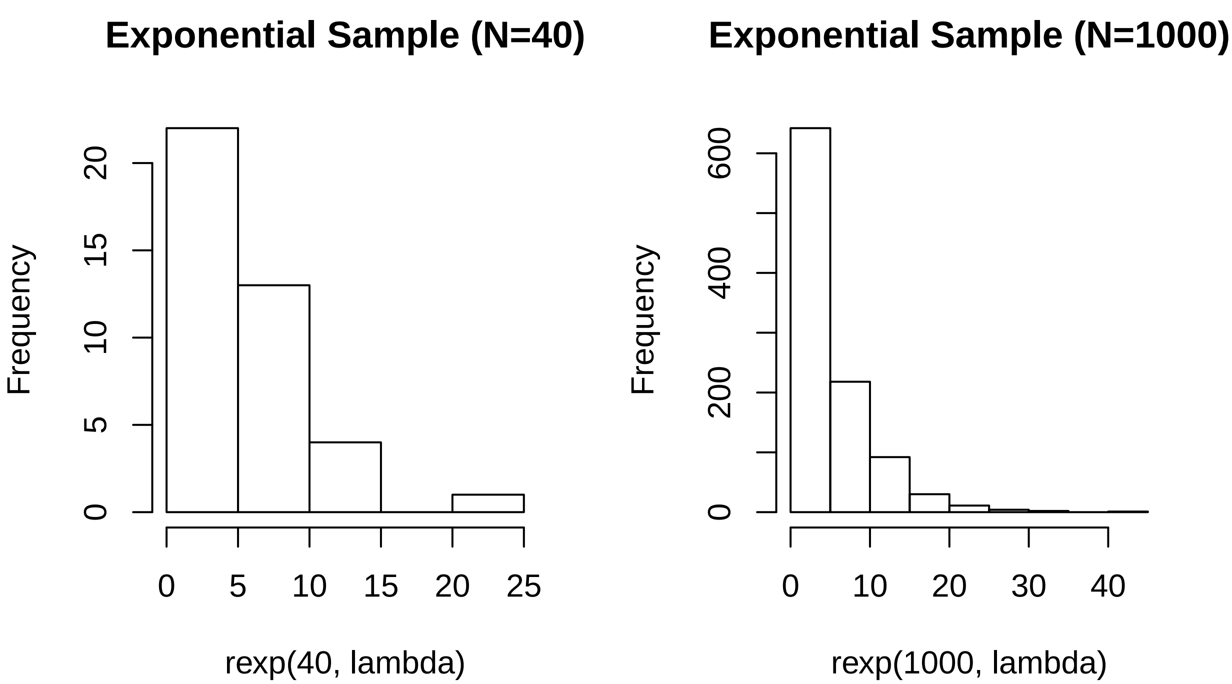
Similar to the investigation above, we can take a look at how sample variance compares to the theoretical variance. In the single sample of 40 observations, the variance is 22.75, compared to the theoretical variance of 25.

Again, in a similar fashion to how we investigated the mean above via simulation, we can see that as we take random draws from the same distribution, the average sample variance approaches the theoretical variance. After 1000 loops through our simulation exercise, the average sample variance is 25.28 compared to the theoretical variance of 25 – much closer than a single sample.

### Simulation Distributions

This exercise provides some insight into how the CLT works in practise. As we increase sample size, for example from 40 to 1000, we can expect the sample distribution to approach the theoretical exponential distribution. For example, consider the image below with N=40 and N=1000.

```
par(mfrow=c(1,2))
hist(rexp(40, lambda), main='Exponential Sample (N=40)')
hist(rexp(1000, lambda), main='Exponential Sample (N=1000)')
```




Now, consider the distribution of the average of the sample mean and variance as simulated in part one of this document below. A blue reference line displaying the standard normal density is used to show that the distribution of sample means and variances are approximately normal.

```
par(mfrow=c(1,2))
x <- ExponentialMeanSimulations
h <- hist(x, main = 'Distribution of Sample Mean')
xfit<-seq(min(x),max(x),length=40)
yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col="blue", lwd=2)

x <- ExponentialVarianceSimulations
h <- hist(x, main = 'Distribution of Sample Variance')
xfit<-seq(min(x),max(x),length=40)
yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))
```

```
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col="blue", lwd=2)
```