Comparison of the Exponential Distribution with the Central Limit Theorem

Thomson Kneeland

March 31, 2016

```
knitr::opts_chunk$set(fig.height=3.5,echo=TRUE, warning=FALSE, message=FALSE)
```

Overview

This paper will examine the exponential distribution, simulating a set of random population draws and comparing the output with the Central Limit Theorem. Theoretical expectation of the Central Limit Theorem is that given enough iterations, our sample data output should approach a normal distribution.

Simulations

The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. Accordingly, the mean of the exponential distribution is 1/lambda and the standard deviation is also 1/lambda. We will set lambda = 0.2 for all of the simulations. We will examine the averages (means) of a data set of 40 exponentials, iterating 1,000 simulations for comparison.

```
set.seed(1000)  ## set seed for study reproducibility
lambda <-.2  ## set rate parameter
sdtheo <- 1/lambda   ## calculate theoretical standard deviation
meantheo <- 1/lambda  ## set theoretical mean
n <- 40  ## set number of samples in each data set
```

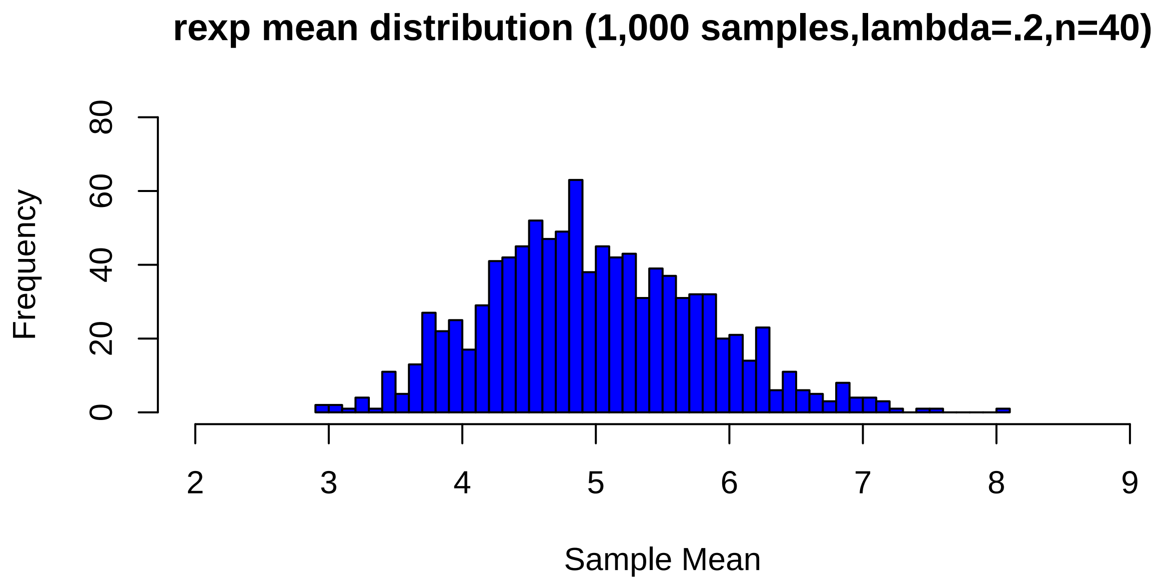
Calculate a set of 1,000 simulation means with the above parameters and examine the first 10.

```
means = NULL
for (i in 1 : 1000) means = c(means, mean(rexp(40,rate=lambda)))
head(means,10)
```

```
##  [1] 4.514222 5.050788 3.252216 3.916899 4.898008 3.677283 6.047866
##  [8] 5.193525 4.941730 4.697506
```

Viewing the first 10 sample means calculated, we can see that values appear to relatively clustered from 3-6. Our theoretical and expected mean is 1/lambda = 5, which lies within this range. Let's view the set of 1,000 means as a frequency histogram for a more detailed view of the data.

```
hist(means,col="blue",breaks=40,main="rexp mean distribution (1,000 samples,lambda=.2,n=40)",
     xlab="Sample Mean",xlim=c(2,9),ylim=c(0,80))
```



Comparison of Sample Mean and Theoretical Mean

We have set our rate parameter (lambda) to be .2 and calculated our theoretical expected mean from this (theoretical mean=1/lambda). Let's now calculate the mean of the collection of 1,000 sample means we have obtained and compare with the theoretical mean.

```
mean(means)  ## sample mean
```

```
## [1] 4.986963
```

```
meantheo  ## theoretical mean
```

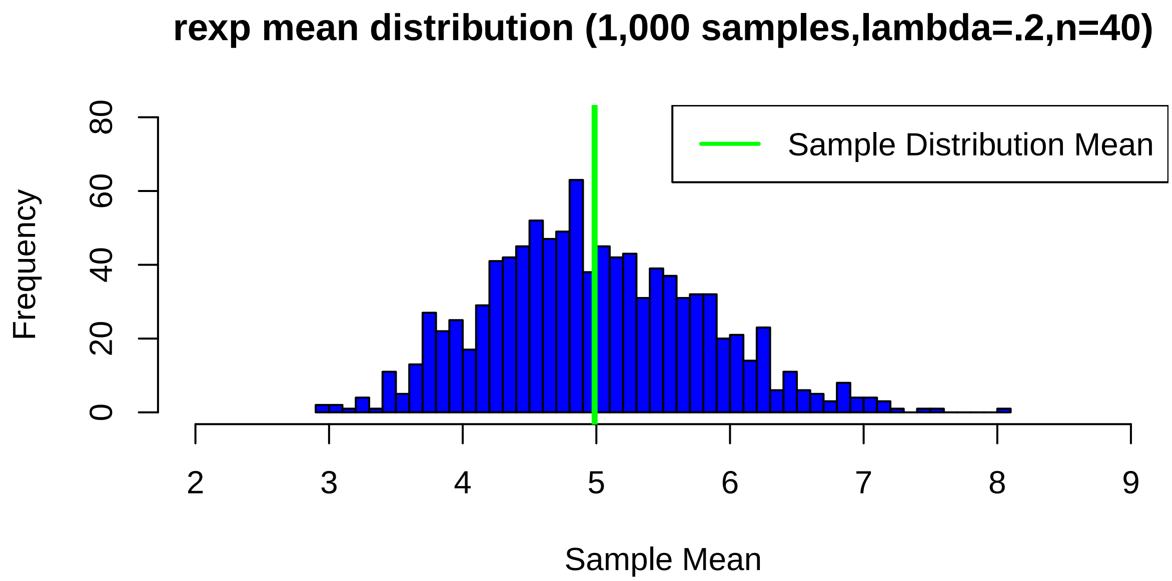
```
## [1] 5
```

```
(meantheo-mean(means))  ## calculate difference between theoretical means and sample means
```

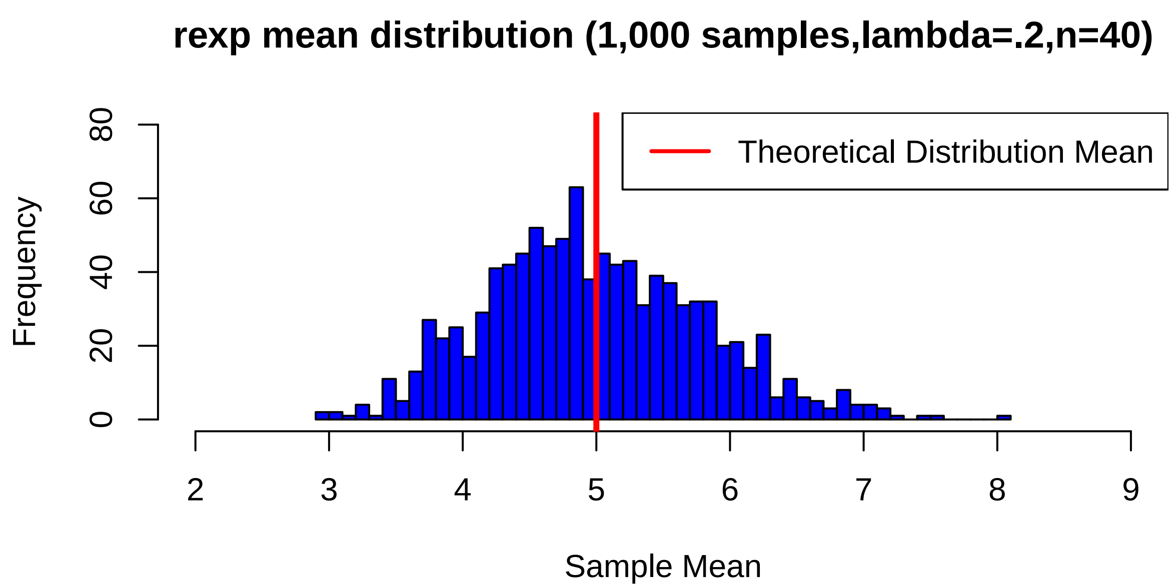
```
## [1] 0.01303661
```

Our sample mean only differs from the theoretical mean by only .013; our observations support our expectations! Let's examine these visually within the context of our dataset.

```
hist(means,col="blue",breaks=40,main="rexp mean distribution (1,000 samples,lambda=.2,n=40)",
     xlab="Sample Mean",xlim=c(2,9),ylim=c(0,80))
legend('topright', c("Sample Distribution Mean"), lty=1,lwd=2, col="green")
abline(v=mean(means),lwd=3,col="green") ## add sample mean to graph
```




```
hist(means,col="blue",breaks=40,main="rexp mean distribution (1,000 samples,lambda=.2,n=40)",
     xlab="Sample Mean",xlim=c(2,9),ylim=c(0,80))
legend('topright', c("Theoretical Distribution Mean"), lty=1,lwd=2, col="red")
abline(v=meantheo,lwd="3",col="red")  ## add theoretical mean to graph
```



Visually examining the two histograms, we find that our sample means does match our theoretical expected mean.

Comparison of Sample Variance with Theoretical Variance

The standard deviation of an exponential function is defined as 1/lambda. Accordingly, our theoretical variance would be sdˆ2/n or (1/lambda)ˆ2/n, accounting for our sample size of n. Let's calculate and compare the theoretical variance with that of our sample set.

Theoretical Variance

```
theovariance <- (sdtheo)^2/n  ## calculate theoretical variance
theovariance
```

```
## [1] 0.625
```

Sample Variance

```
samplevariance = (sd(means))^2  ## calculate sample variance
samplevariance
```

```
## [1] 0.654343
```

```
theovariance-samplevariance ##calculate difference between theoretical and sample variances
```

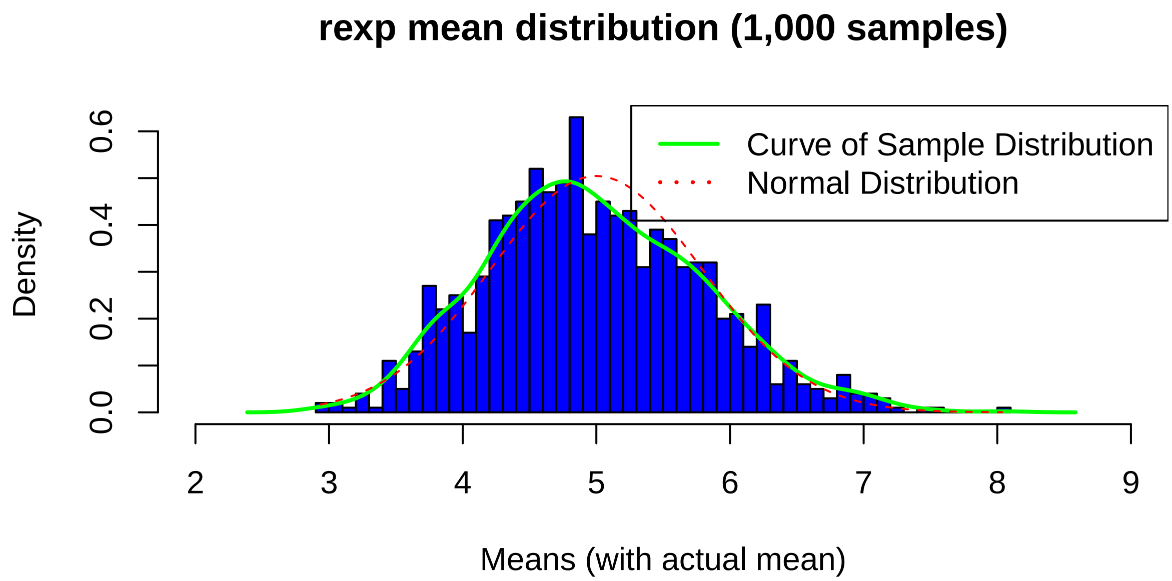
```
## [1] -0.02934304
```

The sample variance differs by only .029, supporting our theoretical expectations.

Compare Our Sample Distribution to a Normal Distribution

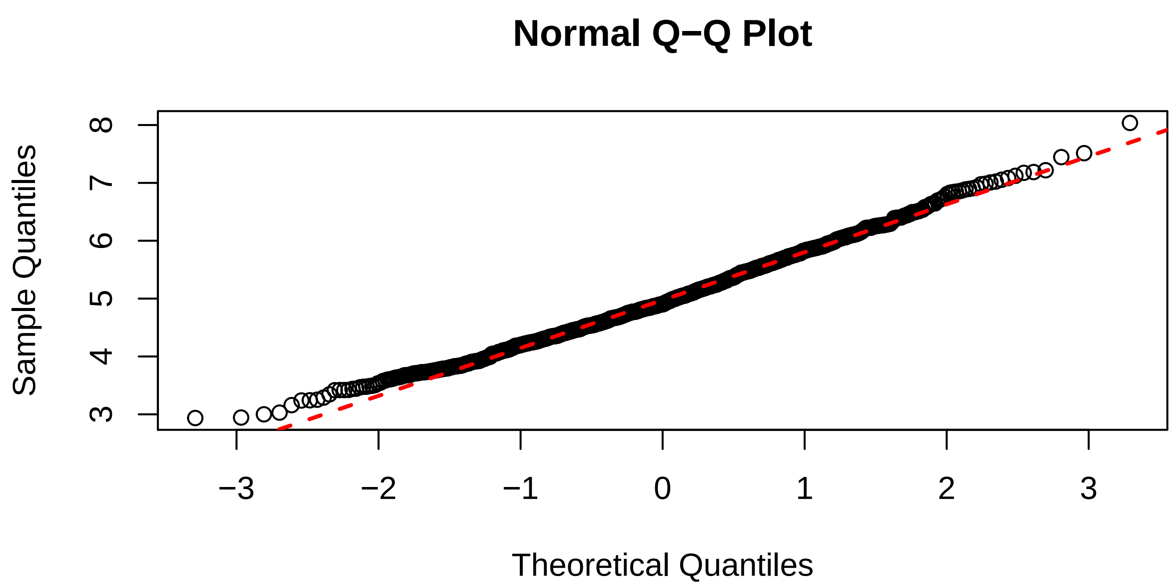
Does our sample data support the Central Limit Theorem after 1,000 iterations? If so, our exponential distribution should be approaching that of a normal distribution, with a Gaussian curve. We will fit an approximate distribution curve of the sample data along with a hypothetical normal distribution with the same mean and standard deviation/variance onto our graph for comparison. The argument prob=TRUE insures that the vertical axis uses a relative density scale, for comparison to a normal distribution density.

```
hist(means,prob=TRUE,col="blue",breaks=40,main="rexp mean distribution (1,000 samples)",
     xlab="Means (with actual mean)",xlim=c(2,9))
lines(density(means), lwd="2", col="green") ## fit approximate curve of sample data distribution
xfit <- seq(min(means), max(means), length=100)
yfit <- dnorm(xfit, mean=meantheo, sd=(sdtheo/sqrt(n)))
lines(xfit, yfit, pch=22, col="red", lty=2)  ## fit curve of normal distribution on graph
legend('topright', c("Curve of Sample Distribution","Normal Distribution"), lty=c(1,3),lwd=2,
       col=c("green","red"))
```



It would appear that the 1,000 sample exponential distribution is approaching a normal distribution, supporting the Central Limit Theorem. We can also assess this using the Quantile-Quantile Plot, which we can use to examine the normality of our data. The closer our data is to a normal distribution, the closer our data should be to the "normal distribution line" plotted.

```
qqnorm(means)
qqline(means,lwd=2,lty=2,col="red")
```



It appears here as well that our sample distribution is approaching the normal distribution. Based on both of these comparisons, it would be safe to say that our 1,000 sample set is approximately normal.