

An Application of the Central Limit Theorem to the Exponential Distribution

Kevin E. D'Elia

March 5, 2016

Overview

The purpose of this project is to create a sampling distribution of a statistic calculated on the exponential distribution and show how the Central Limit Theorem (CLT) applies. As one of the most important theorems in statistics, the CLT says that if you take many repeated samples from a population, then calculate the averages (or sum) of each one, the collection of those averages will be normally distributed. In other words, for large n , $\bar{X} \sim N(\mu, \sigma^2/n)$.

Simulation

The exponential distribution can be simulated in R with **rexp(n, lambda)** where λ is the *rate* parameter. The mean μ of the exponential distribution is $\frac{1}{\lambda}$ and the standard deviation σ is also $\frac{1}{\lambda}$. A requirement of this report is to set $\lambda = 0.2$. The sample size **n** is 40 and the number of repetitions is **1000**.

As with any good simulation, the first step is to set the seed for the random number generator. This is done so that the results generated by the random number engine will be consistent across all invocations of the code, thus ensuring that the results are reproducible by other interested parties.

```
set.seed(10) # for reproducible research
```

Next, the parameters for the simulation are defined:

```
lambda <- 0.2 # rate parameter  
n <- 40 # sample size  
runs <- 1000 # number of experiments to run
```

Now the sampling distribution of the means of 40 random exponentials can be generated using the following code:

```
xbar <- NULL  
for (i in 1:runs) { xbar <- c(xbar, mean(rexp(n, lambda))) }
```

The distribution can be evaluated along the following criteria:

1. Comparison of the sample mean to the theoretical mean
2. Comparison of the sample variance to the theoretical variance
3. Comparison of the sampling distribution to a normal distribution

Sample Mean versus Theoretical Mean:

The theoretical means of the distribution, as noted above, is $\frac{1}{\lambda}$. Computing that for the specified parameters gives:

```
1/lambda
```

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

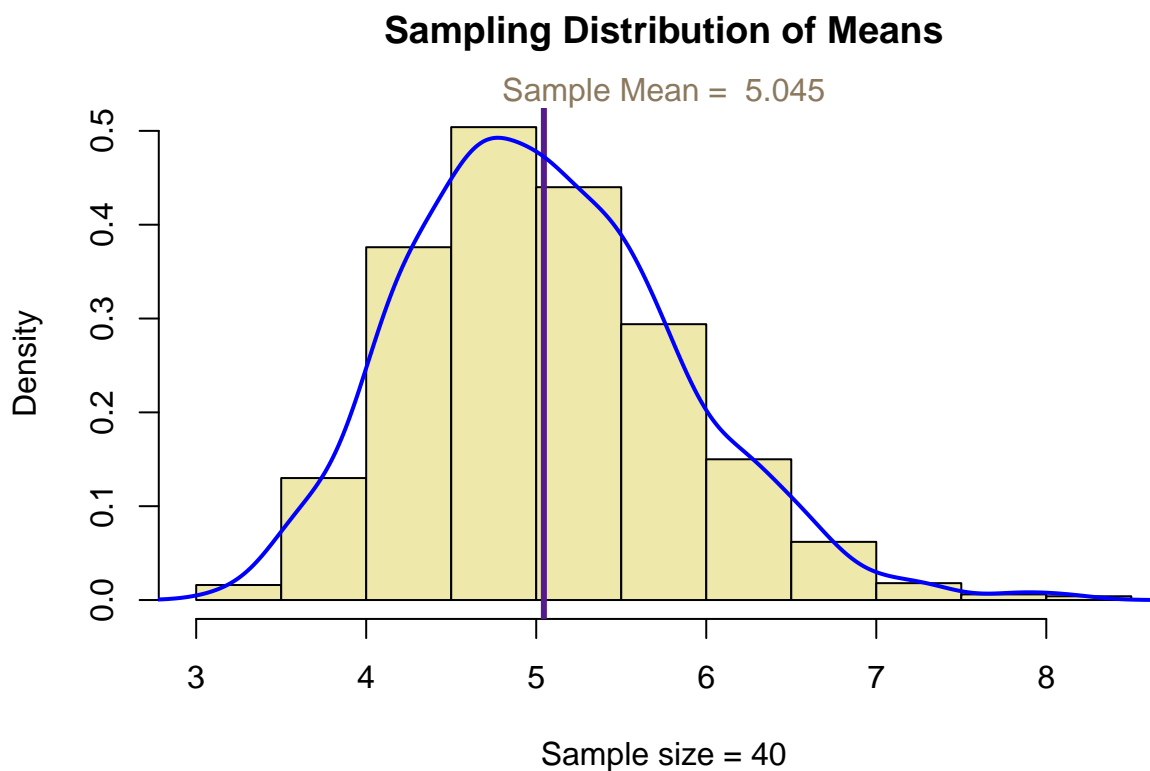
How does the sample mean compare to this value? Calculating the mean of the distribution gives the answer:

```
mean(xbar)
```

```
## [1] 5.04506
```

So, it appears that the sample mean is indeed very close to the value predicted by the theory. Graphically, this can be shown as:

```
main.title <- "Sampling Distribution of Means"
hist(xbar, prob = TRUE, xlab = "Sample size = 40", main = main.title, col = "palegoldenrod")
abline(v = mean(xbar), col = "purple4", lwd = 3)
mtext(paste("Sample Mean = ", round(mean(xbar), 3)), col = "navajowhite4")
lines(density(xbar), col="blue", lwd=2)
```



Sample Variance versus Theoretical Variance:

The theoretical variance of the exponential distribution is $\text{Var}(\bar{x}) = \frac{\sigma^2}{n}$. Using the parameters for the simulation, this evaluates to:

```
1/(n * lambda^2)
```

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

Comparing this value to the variance of the sampling distribution gives the following result:

```
round(var(xbar), 3)
```

```
## [1] 0.637
```

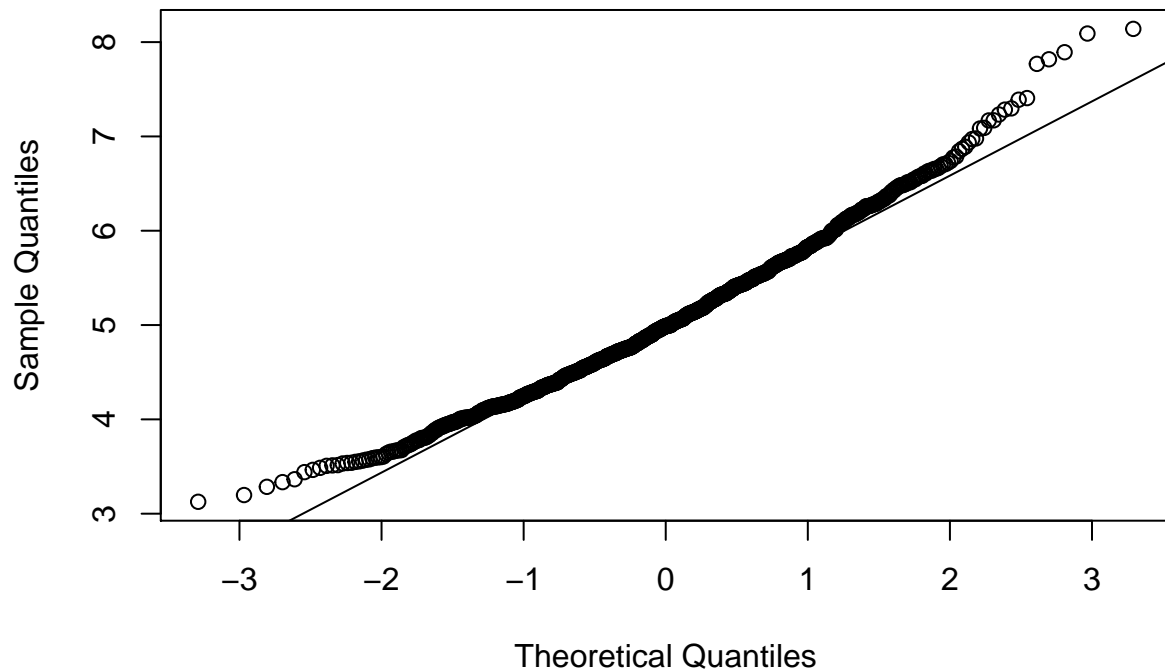
Again, as with the comparison between the theoretical and sample means, the values for the variance are very close.

Distribution

The density curve drawn atop the histogram shows the relatively normal curvature of the distribution. The balance point, or the *mean of the sample means*, appears to be reasonably close to the mean of the distribution from which the samples were drawn. So, it can be concluded that the sample mean is a good estimator of the population mean.

An additional graphic which shows how normally distributed the data might be is the **Q-Q plot**, also known as a *quantile-quantile plot*.

```
qqnorm(xbar, main = NULL); qqline(xbar)
```



The graph shows a majority of the data along the regression line, with minor skewing of the data at both ends equally above the line. This is an indicator of normally distributed data.