# The Exponential Distribution and the CLT

*Linda Kukolich*

*February 10, 2015*

This report will examine the exponential distribution and the Central Limit Theorem (CLT). According to the notes given to us in Statistical Inference: "the CLT states that the distribution of averages of *independent and identically distributed* (iid) variables (properly normalized) becomes that of a standard normal as the sample size increases". This report will show that given 1000 sets of 40 samples, the CLT holds.

---

## The Exponential Distribution

The density of the exponential distribution is $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$

This gives a mean of $1/\lambda$ and a variance of $1/\lambda^2$ (See the appendix for a derivation.)

The variance of the sample mean will be $\frac{1}{n\lambda^2}$ Which gives us a standard error of $\frac{1}{\lambda\sqrt{n}}$.

## Simulation

First, let's do our simulation, 40 sets of 1000 samples of an exponential distribution, with a lambda of 0.2.

```
# Set the seed, to ensure repeatability
set.seed(1220)
# Set our constants and parameters of our distributions
nosim <- 1000
n <- 40
lambda <- .2
s_theory <- m_theory <- 1/lambda
se_theory <- s_theory/sqrt(n)    # standard deviation of the sample mean
# sample the distribution 40*1000 times
data <- matrix(rexp(n*nosim, lambda), nosim)
m <- apply(data, 1, function(p) sum(p)/n)
s <- apply(data, 1, function(p) sqrt(sum((p - m)^2)/(n-1)))
```

# Distribution of averages of samples

The central limit theorem works with "averages of iid variables ... as the sample size increases". We are working with 1000 sets of 40 samples. While we saw some variability in the 40 samples, looking at 1000 such sets ought to "average things out".

```
# Take the means of the 1000 rows of data
means <- apply(data, 1, mean)
# Calculate our theoretical distribution over the domain of our plot
x <- seq(0, 9, length.out=50)
y <- dnorm(x, mean=m_theory, sd=se_theory)
```

# Histogram of the means of 40 samples, with a normal distribution, $N(\frac{1}{\lambda}, \frac{1}{\lambda\sqrt{40}})$
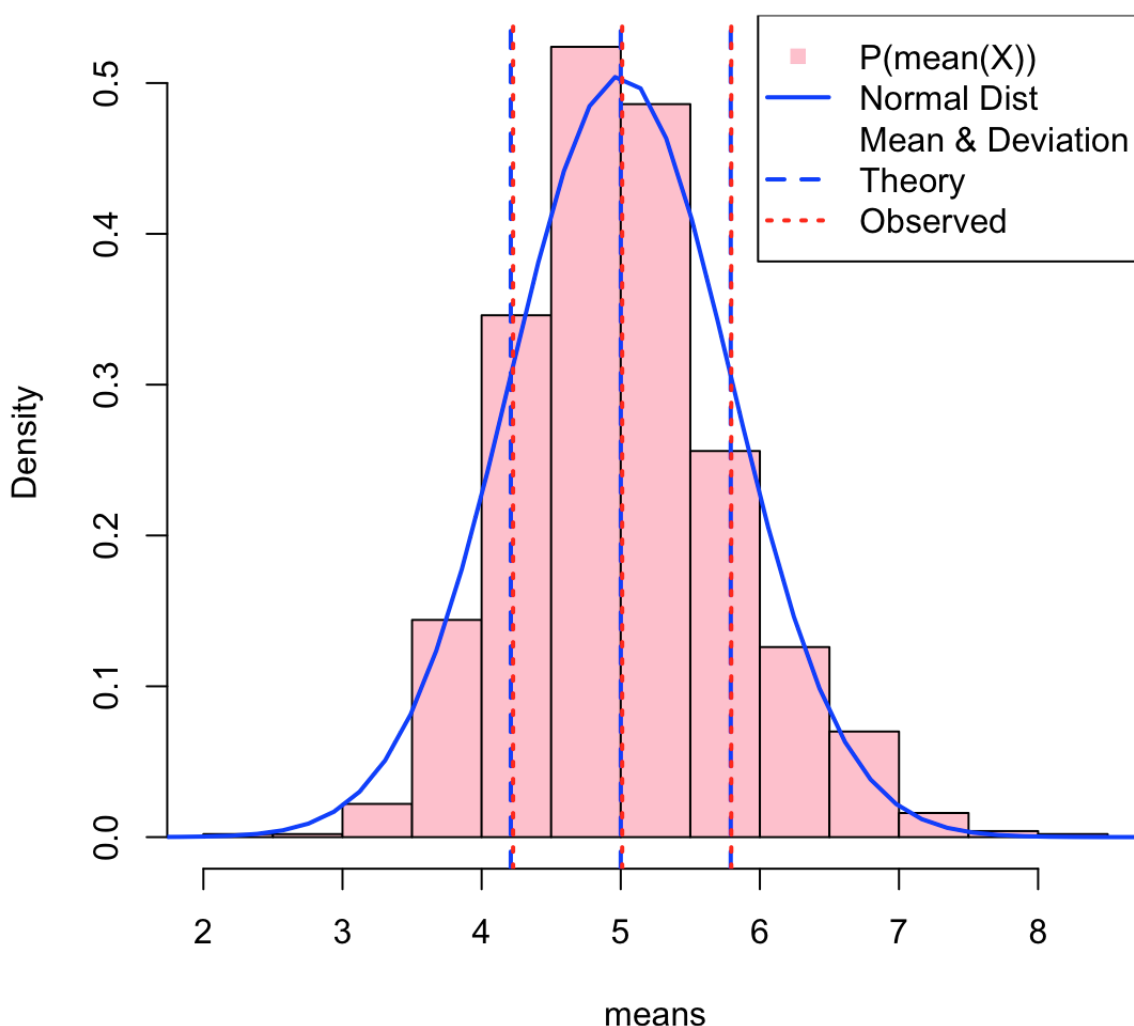


In the figure above, the pink boxes show P(X), a normalized histogram of the sample means from each of the 1000 sets of 40 samples. The solid blue line shows a normal distribution. The blue dashed lines show the mean of this distribution ($1/\lambda = 5$) and one standard deviation above and below that mean ($1/\lambda \pm 1/(\lambda\sqrt{40} = 5 \pm 0.79$). These marks for the mean and standard error can be hard to see on the plot

because the measured sample mean (5.01) and the square root of the variance of the sample mean (0.78) are also shown in red, nearly on top of the dashed blue theoretical lines. These values are given in the table below, for clarity.

|  | Mean | Standard Error |
| --- | --- | --- |
| Theory | 5 | 0.791 |
| Measured | 5.01 | 0.784 |

It would be nice to calculate the actual error between the normalized histrogram we show and the theoretical distribution. We can eyeball it and see that it matches, but that doesn't give us a number that would let us compare different theoretical models. Sadly, this task is beyond my current abilities in R. For now, the curve matches nicely, the mean matches nicely, and the standard error, as given by the variance of the sample mean, matches nicely. All in all, a win for the Central Limit Theorem.

# Appendix

## Derivation of Mean and Variance of Exponential Distribution

The density of the exponential distribution is $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$

## Mean

$$
\begin{aligned}
E[X] &= \int_0^\infty x\lambda e^{-\lambda x}\, dx \\
&= -xe^{-\lambda x}\Big|_0^\infty - \int_0^\infty -e^{-\lambda x}\, dx \\
&= (0 - 0) - \frac{1}{\lambda}e^{-\lambda x}\Big|_0^\infty \\
&= (0 - 1)\frac{-1}{\lambda} \\
&= 1/\lambda
\end{aligned}
$$

## Variance

$$E[X^2] - E[X]^2 = \int_0^\infty x^2 \lambda e^{-\lambda x}\,dx - \frac{1}{\lambda^2}$$

$$= x^2(-e^{-\lambda x})\Big|_0^\infty + \int_0^\infty 2x e^{-\lambda x}\,dx - \frac{1}{\lambda^2}$$

$$= (0-0) + \left[-\frac{2}{\lambda}xe^{-\lambda x}\right]_0^\infty + \frac{2}{\lambda}\int_0^\infty e^{-\lambda x}\,dx - \frac{1}{\lambda^2}$$

$$= (0-0) - \frac{2}{\lambda}\frac{1}{\lambda}e^{-\lambda x}\Big|_0^\infty - \frac{1}{\lambda^2}$$

$$= -\frac{2}{\lambda^2}(0-1) - \frac{1}{\lambda^2}$$

$$= \frac{1}{\lambda^2}$$

# Distribution of individual samples

Let us examine one of sets. That is, what does 1000 samples from an exponential distribution look like? The shape of the histogram ought to match the underlying exponential distribution.

## Histogram of 40 samples, with $\lambda e^{-\lambda x}$

```
x <- seq(0, 35, length.out=50)
y <- lambda * exp(-lambda*x)
```

**Histogram of One Set of 40 Samples**

Legend:
- ■ P(X)
- ── Exponential Dist
- – – 1/lambda = 5
- ····· mean(X) = 4.9
- ····· +- sqrt(s(X)) 23.1
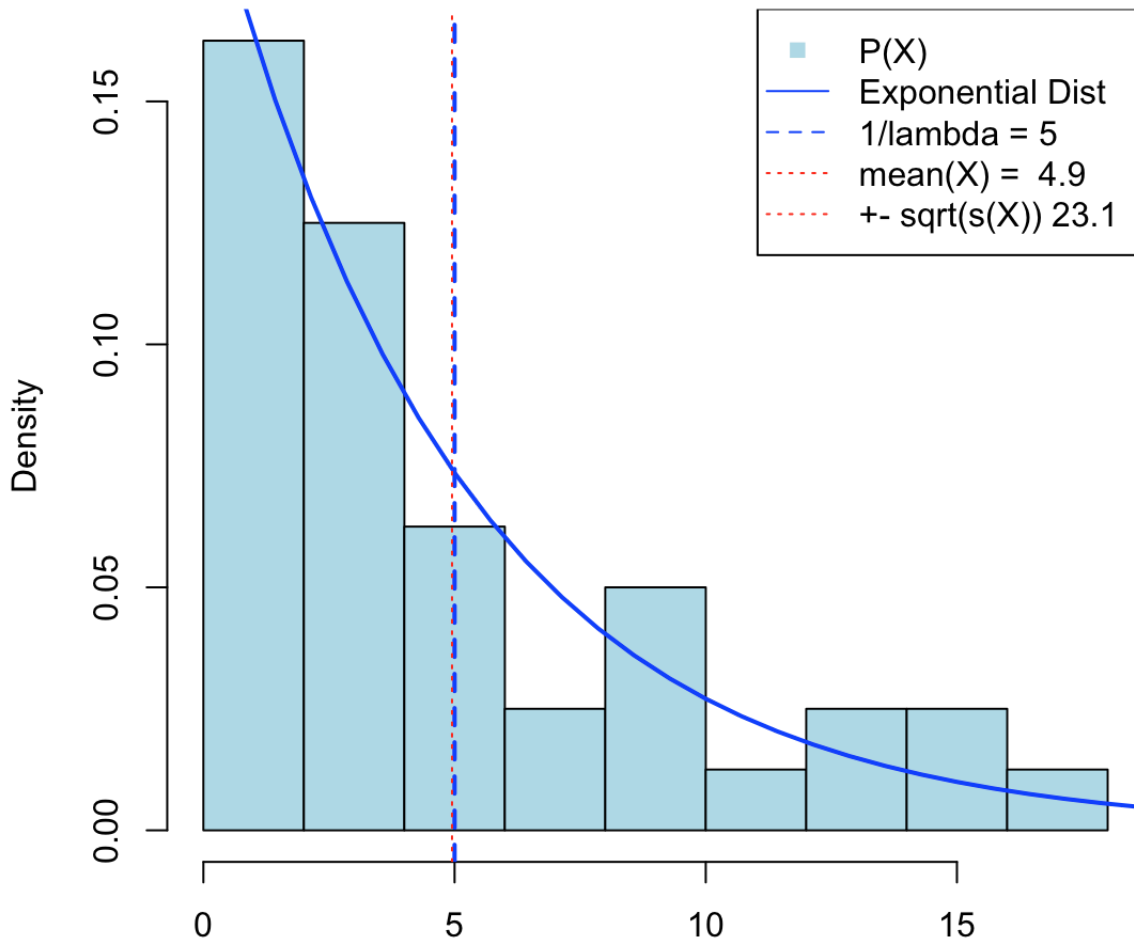
The line of the function $\lambda e^{-\lambda x}$ is a good match for the histogram, which has been normalized to give probabilities rather than counts. The dotted red line marking the sample mean for this set is near 5, which is the mean of the underlying distribution.

Out of curiosity, do all the histograms look similar? Do they all appear to match the shape of the underlying distribution?

# Histograms of the first 16 sets of 40 samples
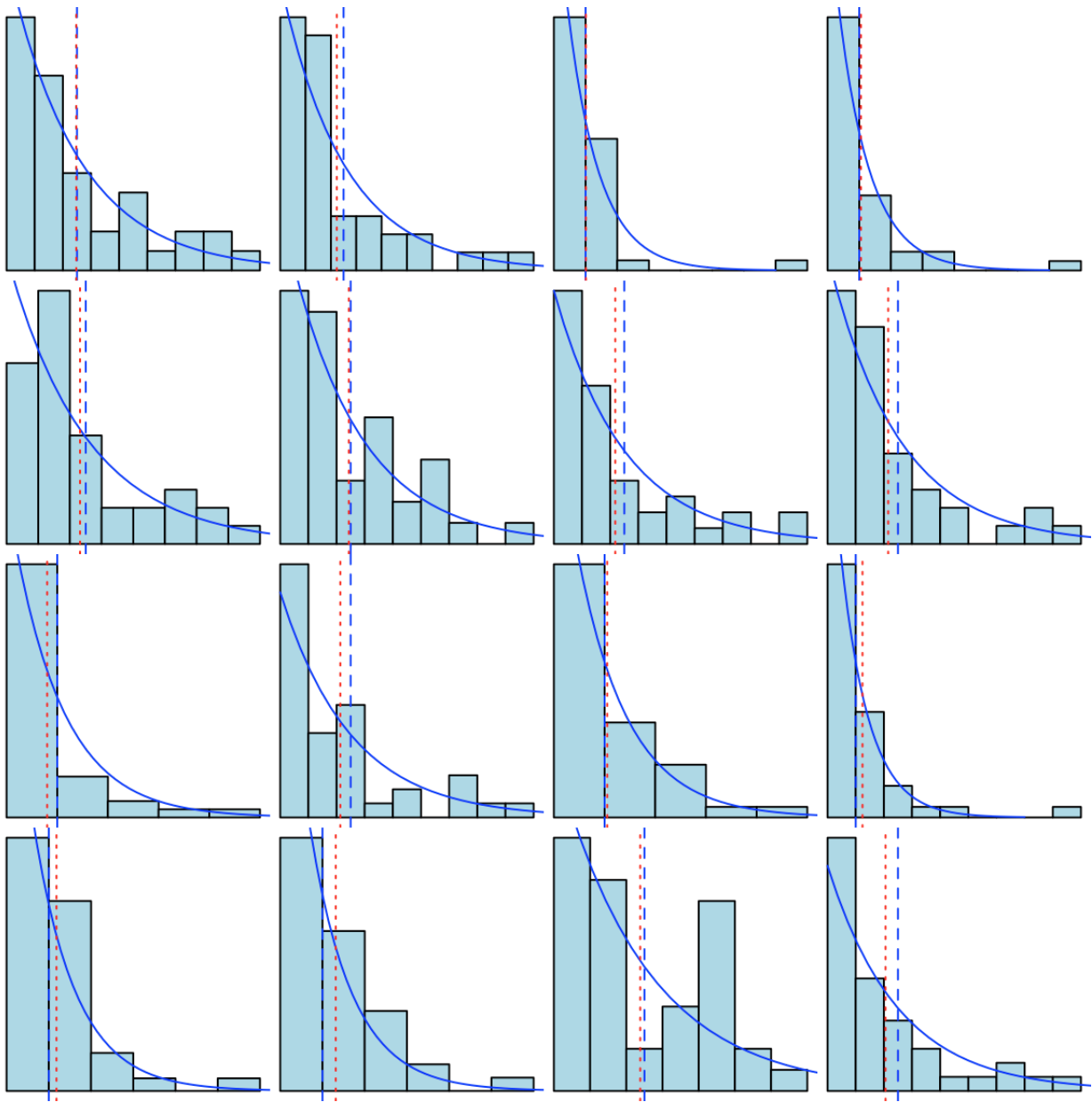
The theory line seems to be a reasonable match for the histogram probability boxes, though some match better than others. The dotted red lines showing the sample means are near the dashed blue lines for the theoretical mean of 5, though again, some match better than others.