Exploring Exponential Distributions and the Central Limit Theorem

Rafi Kurlansik

Monday, August 17, 2015

Overview

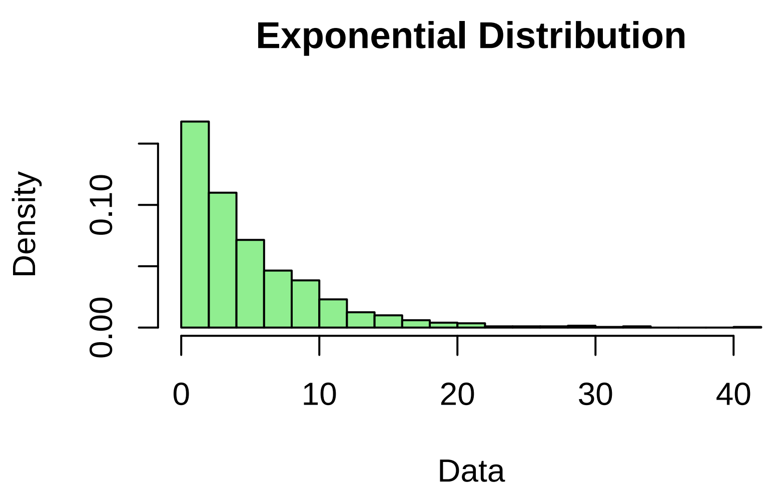
In this paper we will use R to explore the Central Limit Theorem as applied to an exponential distribution. First we will review the theoretical definitions of an exponential distribution. Then we will run a simulation to generate data that can be used to create a sampling distribution. Once we have our sampling distribution, we will **a)** *compare the sample mean to the theoretical mean,* **b)** *compare the sample variance to the theoretical variance,* and **c)** *show that the sampling distribution is essentially normal.*

The Exponential Distribution

The probability density function of an exponential distribution is $f(x) = \lambda e^{-\lambda x}$. Other important definitions for this distribution include the mean, $E[X] = \frac{1}{\lambda}$, the variance, $\sigma^2 = \frac{1}{\lambda^2}$, and the standard deviation, $\sigma = \sqrt{\frac{1}{\lambda^2}}$. In this case, the standard deviation is equal to the mean, so we can simply say $\sigma = \frac{1}{\lambda} = E[X] = \mu$.

To begin our investigation, let's set the seed to ensure reproducible results, generate some random exponentials, and build a plot. For this assignment we will be setting λ to 0.2.

```
set.seed(5)
expnts <- rexp(1000, 0.2)
```



Let's look at some of the features of this distribution!

```r
summary(expnts)
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##  0.00678  1.45200  3.43300  5.01300  7.01100 41.83000
sd(expnts)
## [1] 5.108262
var(expnts)
## [1] 26.09434
```

Standard Deviation $\sigma = 5.108$, Variance $\sigma^2 = 26.094$, Mean $\mu = 5.013$. Notice σ is roughly equal to μ in this case. This supports what was theoretically shown above - in an exponential distribution, the standard deviation is equal to the mean. Figure 1 in the appendix illustrates this.

The Sampling Distribution of the Sample Mean

Let's run the simulation 1000 more times using the same seed, each time taking a sample of 40 values and storing their mean in a new variable, called `smplmns`. Here, the equations for the variance and standard deviation are slightly different: $\sigma_m^2 = \frac{\sigma^2}{N}$, and $\sigma_m = \frac{\sigma}{\sqrt{N}}$. We will store the results of these formulas in variables `sample_variance` and `sample_std`.

We can use the `sapply()` function to run it many times, and the `sample()` function to grab our samples. We will want to replace the samples so we don't run out of them!

```r
set.seed(5)
expnts <- rexp(1000, 0.2)
smplmns <- sapply(1:1000, function(x) mean(sample(expnts, 40, replace = T)))
summary(smplmns)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.865   4.452   4.967   5.024   5.510   8.928
sample_variance <- (sd(expnts)^2)/40
sample_std <- (sd(expnts))/sqrt(40)
```

Conclusions

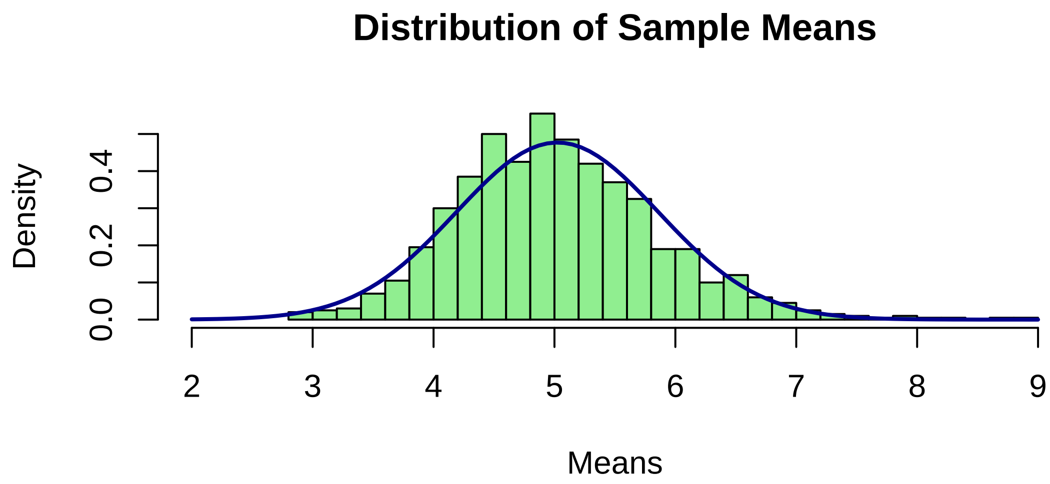
The data speaks! Before we look at pretty plots, lets compare the sample mean to the theoretical mean.

As we stated at the beginning, the mean for an exponential distribution is $E[X] = \frac{1}{\lambda}$. In this case, $\lambda = 0.2$, so $\frac{1}{\lambda} = \frac{1}{0.2} = 5$ is our theoretical mean. The sampling distribution mean is 5.024, *strongly suggesting that the mean of the distribution of sample means is roughly equivalent to the theoretical mean of the underlying distribution.* I suspect that if we took 10,000 sample means the numbers would be even closer to equal. Figure 2 in the appendix provides a visual representation of this relationship.

Next, lets compare the sample variance to the theoretical variance. For the sampling distribution, the population standard deviation of 5 and sample size of 40 gives us theoretical variance $\sigma_m^2 = \frac{\sigma^2}{N} = \frac{5^2}{40} = 0.625$, and standard deviation $\sigma_m = \frac{\sigma}{\sqrt{N}} = \frac{5}{\sqrt{40}} = 0.790$. If we look at our sample values,

```r
sample_variance
## [1] 0.6523585
sample_std
## [1] 0.8076871
```

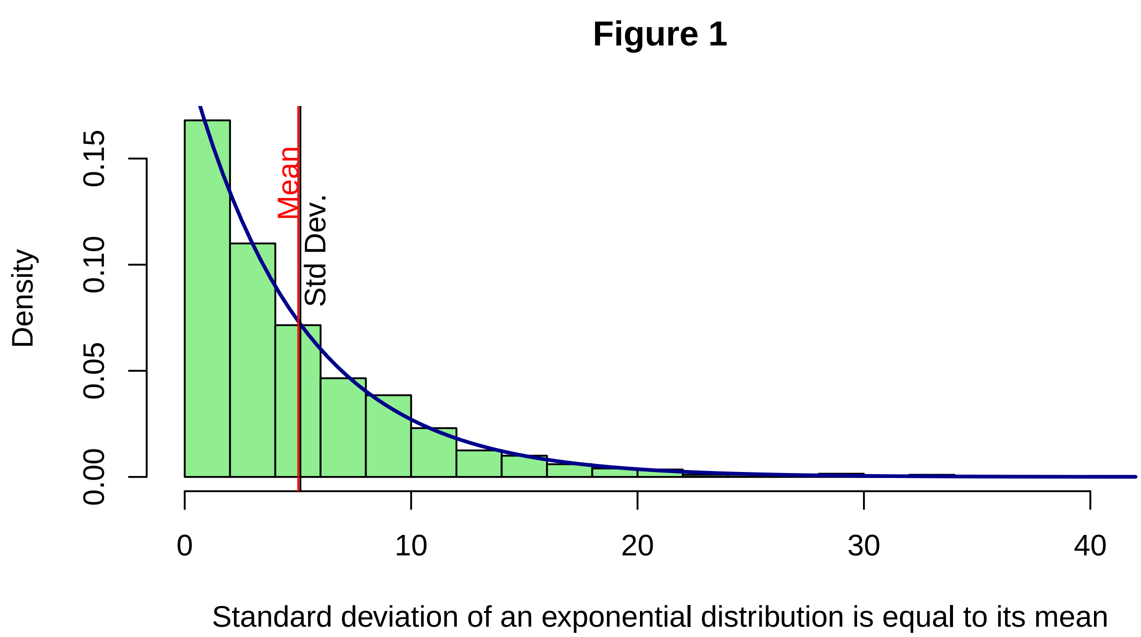
We can see that they come very close to our theoretical values!



To finish, after plotting the means of our samples from an exponential distribution 1000 times, we have what appears to essentially be a normal distribution (in light green). The overlay of a probability density function for a normal distribution (here in dark blue) confirms this. *With this plot we have shown the sampling distribution to be approximately normal.*

In summary, we have shown via simulation that the both the sampling mean and variance are approximately equal to their theoretical counterparts, and that the sampling distribution of the sample mean is approximately normal.

Appendix



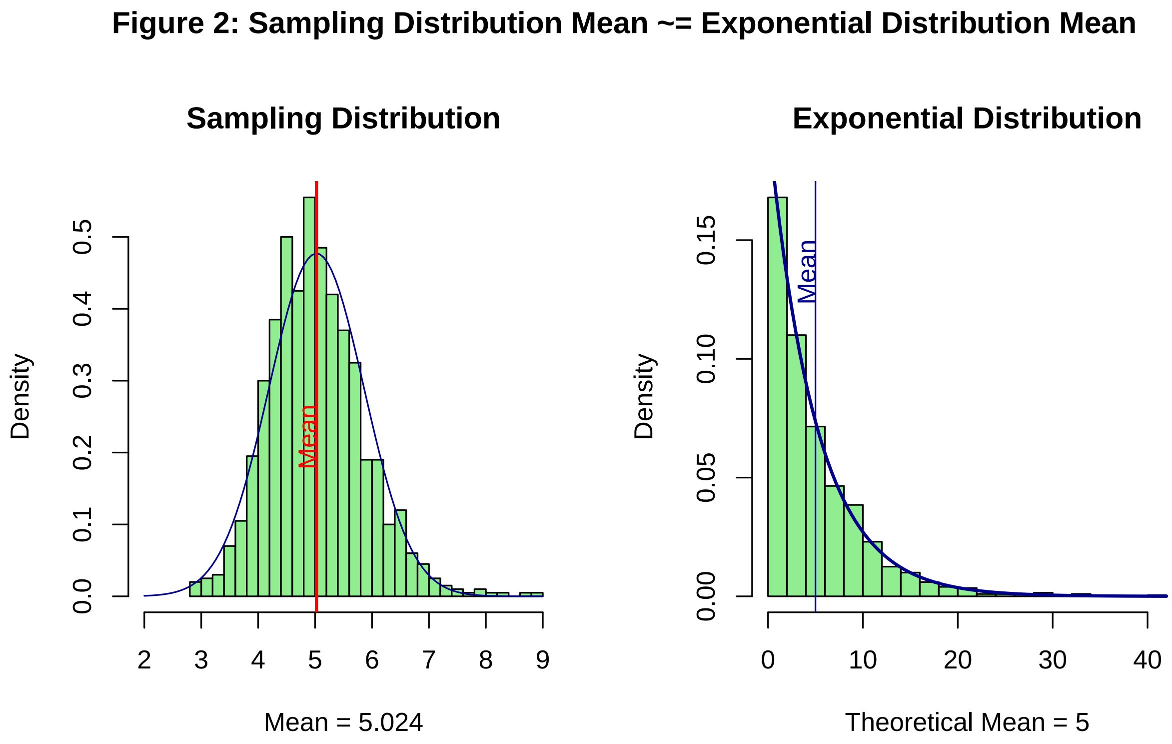


Figure 2: Sampling Distribution Mean ~= Exponential Distribution Mean
Sampling Distribution
Density
0.0
0.1
0.2
0.3
0.4
0.5
Mean
2
3
4
5
6
7
8
9
Mean = 5.024
Exponential Distribution
Density
0.00
0.05
0.10
0.15
Mean
0
10
20
30
40
Theoretical Mean = 5