Exploration of the Exponential Distribution and the Central Limit Theorem on Mean and Variance

Bill Martersteck

August 07, 2016

Contents

Overview

In this project, we will explore how the Exponential Distribution behaves relative to the Central Limit Theorem. We will look at the mean and the variance of large numbers of samples and compare against the theoretical versions of these measures.

Exponential Probability over Various λ

For this excercise, we will only consider the distribution with $\lambda = 0.2$.
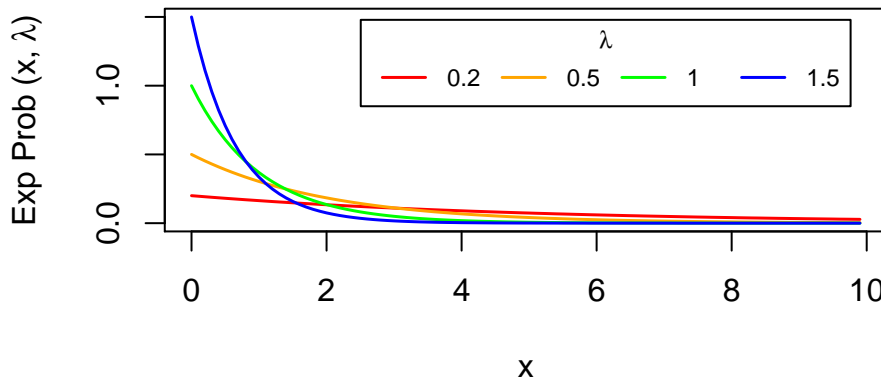
Simulations

For this exercise, we will be investigating how a large number of iterations of the mean and variance of a set of samples from the exponential distribution behave relative to the theortical mean and variance for this distribution. If the Central Limit Theorem holds for this distribution, then the sample means and variances should approximate the theoretical mean and variance of the distribution.

```
l = .2
numsims = 1000
samples = 40
```

For this analysis, we will select 40 samples to calculate the mean and variance and do this 1000 times. With the selection of random variables, we will calculate a vector of the means of these samples and a vector of the variances of these samples for further analysis. Each vector (mns and vars) will be of length 1000.

```
RandExp <- matrix(rexp(numsims * samples, rate = l), nrow = numsims)
mns <- apply(RandExp, 1, mean)
vars <- apply(RandExp, 1, var)
```

Sample Mean versus Theoretical Mean

The Exponential Distribution has a mean (μ) of $1/\lambda$. An exponential random variable from a distribution with $lambda = 0.2$, would expect to have a mean of $\mu = 1/0.2 = 5$. We calculate our sample mean and sample standard deviation from the mns vector by:

```
popMean <- 1/l
sampleMean <- mean(mns)
sampleStdDev <- sd(mns)
c(popMean, sampleMean, sampleStdDev)
```

```
## [1] 5.0000000 5.0137716 0.8019486
```

Now lets plot the 1000 means that we just created each from 40 samples along with the sample mean and the population mean.

```
clrs = c("red", "blue")
lgnd = c("Sample", "Population")
hist(mns, breaks = 40, include.lowest = TRUE,
     main = "Mean of 40 Exp Random Variables ("~lambda~"= 0.2) 1000 times",
     xlab = "Mean of 40 Exponential Random Variables", ylab = "Freq of the 1000 Means")
abline(v = popMean, col = "red")
abline(v = sampleMean, col = "blue")
legend("topright", legend = lgnd, col = clrs, title = "Means",
       lwd = 1.5, inset = .05, cex = .75)
```
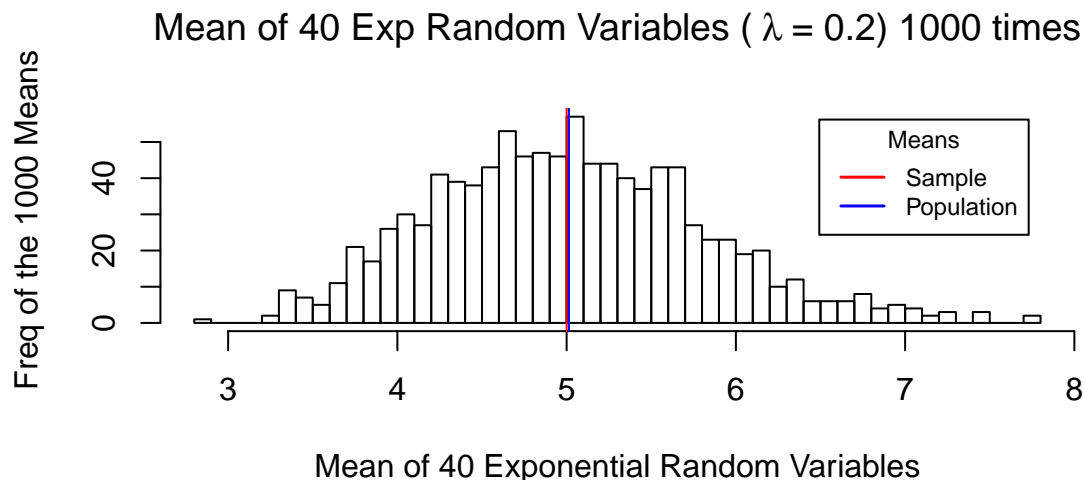
Given that the blue and red lines (showing the sample mean of 5.0137716 and population mean of 5 respectivly) are very close to each other, we can surmise that the means of 40 samples of random variable trends towards a normally distributed population mean.

Lets create a hypothesis test to provide further evidence.

1. Come up with the Null Hypothesis:

$$H_0 : \mu = 1/\lambda = 1/0.2 = 5$$
$$H_a : \mu \neq 5$$

2. The test will be done at a 5% significance level:

$$\alpha = .05$$

3. Determine $\pm z_{a/2}$:

$$\pm z_{a/2} = \pm z_{.025} = \pm qnorm(.025) = \pm 1.96$$

4. Compute the value of the test statistic:
   From above, we have the population mean ($popMean = 5$), sample mean ($sampleMean = 5.0137716$) and sample standard deviation ($sampleStdDev = 0.8019486$). Now we have everything to calculate the test statistic:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

```
z = (sampleMean - popMean)/(sampleStdDev/sqrt(numsims))
z
```

```
## [1] 0.5430487
```

5. Conclusion
   Since the test statistic, $z = 0.5430487$, falls well within the "Do Not Reject" region of ± 1.96, there is not sufficient evidence to reject H_0.

Sample Variance versus Theoretical Variance

The Exponential Distribution has a variance (σ^2) of $1/\lambda^2$. A random variable from an exponential distribution with $lambda = 0.2$ would expect to have a variance of $\sigma^2 = 1/0.2^2 = 25$. We want to analyse whether the variance of our sample of random exponentials is what we would expect to see from the exponential distribution.

For our analysis, we need the population mean, the sample mean and the sample standard deviation. These will be used to calculate the test statistic (we can use the sample standard deviation since n is relatively large). These calculations are:

```
popMean = 1/(l^2)
sampleMean = mean(vars)
sampleStdDev = sd(vars)
c(popMean, sampleMean, sampleStdDev)
```

```
## [1] 25.00000 25.38414 11.62161
```

Now lets plot the 1000 variances that we created each from 40 samples along with the sample mean and the population mean.

```r
hist(vars, breaks = 40, include.lowest = TRUE,
     main = "Variance of 40 Exp Random Variables ("~lambda~"= 0.2) 1000 times",
     xlab = "Variance of 40 Exponential Random Variables",
     ylab = "Freq of the 1000 Variances")
abline(v = popMean, col = "red")
abline(v = sampleMean, col = "blue")
legend("topright", legend = lgnd, col = clrs, title = "Means",
       lwd = 1.5, inset = .05, cex = .75)
```
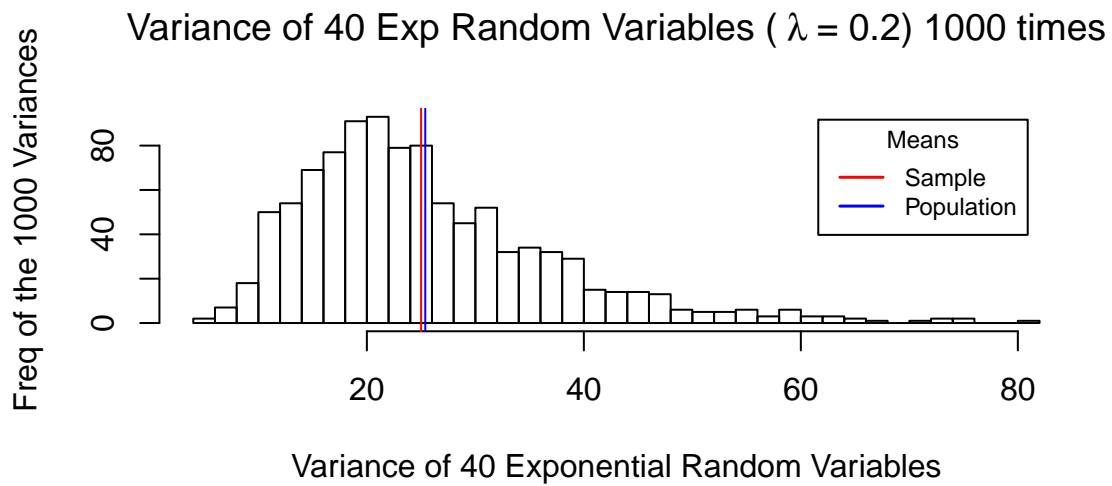

Given that the blue and red lines (showing the mean of the sample variances of 1000 iterations and expected population variance of 25 respectivly) are very close to each other, we can surmise that the variance of 40 samples of random variable trends towards a normally distributed population mean under large numbers of samples.

Lets create a hypothesis test to provide further evidence.

1. Come up with the Null Hypothesis:

$$H_0 : \sigma^2 = 1/\lambda^2 = 1/0.2^2 = 25$$
$$H_a : \sigma^2 \neq 25$$

2. The test will be done at a 5% significance level:

$$\alpha = .05$$

3. Determine $\pm z_{a/2}$:

$$\pm z_{a/2} = \pm z_{.025} = \pm qnorm(.025) = \pm 1.96$$

4. Compute the value of the test statistic:
   From above, we have the population mean ($popMean = 25$) and mean of the sample variances ($sampleMean = 25.3841376$) and the sample standard deviation ($sampleStdDev = 11.6216106$.

Now we have everything to calculate the test statistic:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

```
z = (sampleMean - popMean)/(sampleStdDev/sqrt(numsims))
z
```

```
## [1] 1.045251
```

5. Conclusion

Since the test statistic, $z = 1.0452507$, falls well within the "Do Not Reject" region of ± 1.96, there is not sufficient evidence to reject H_0.
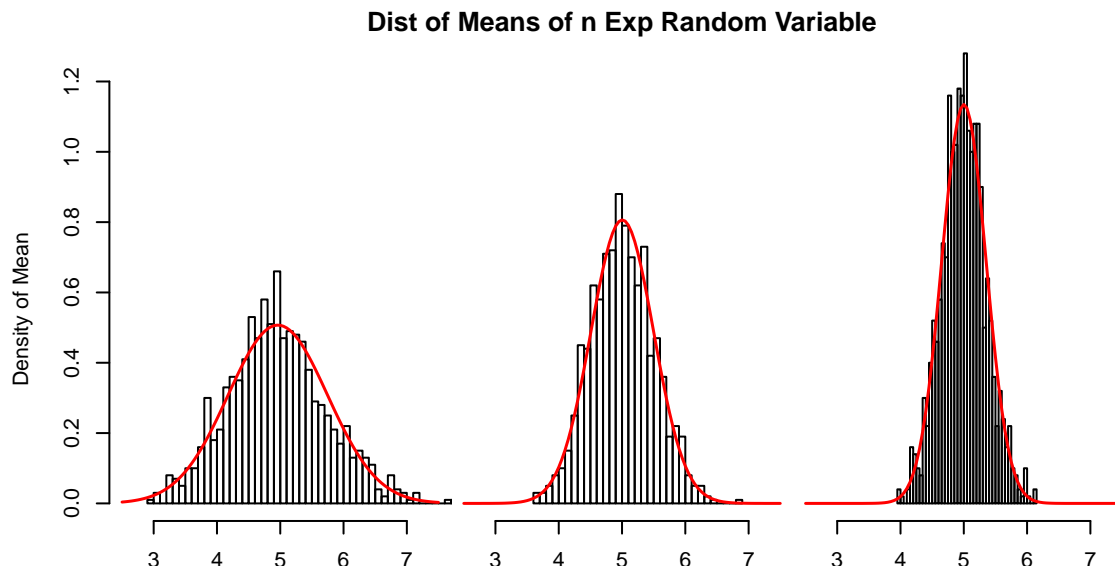
Distribution

All our analysis has been done with 1000 iterations of 40 random variables. The above histograms show the distribution looks normal. As we increase the number of samples, the distributions should become more and more normal if the Central Limit Theorem holds with the difference of the means going to zero. For the sample mean and sample variance, lets increase the number of samples and see that the distribution trends more and more normal. We will increase the number of samples to 40, 100, and 200 and plot the graphs against a normal distribution.

```
samples = c(40, 100, 200)
```

Plot the histogram of the density of the means.

```
par(mfrow=c(1,length(samples)), mar = c(2,0,0,0), oma = c(0,5,2,0))
for (i in samples) {
        RandExp <- matrix(rexp(numsims * i, rate = 1), nrow = numsims)
        mns <- apply(RandExp, 1, mean)
        x <- seq(2.5, 7.5, length.out = 20)
        hist(mns, breaks = 40, freq = FALSE, xlim = c(2.5, 7.5), ylim = c(0, 1.25),
                yaxt = 'n', xlab = paste("Num Samples =", i), main = NULL)
        curve(dnorm(x, mean = mean(mns), sd = sd(mns)), add = TRUE, lwd = 1.5, col = "red")
}
title(main = "Dist of Means of n Exp Random Variable", ylab = "Density of Mean", outer = TRUE )
axis(2, outer = TRUE )
```

As the number of samples increases, the standard deviation decreases and it appears that more the distribution is more normal.

Now we plot the density of the variances of these random variables.

```r
par(mfrow=c(1,length(samples)), mar = c(2,0,0,0), oma = c(0,5,2,0))
for (i in samples) {
        RandExp <- matrix(rexp(numsims * i, rate = 1), nrow = numsims)
        vars <- apply(RandExp, 1, var)
        x <- seq(2.5, 7.5, length.out = 20)
        hist(vars, breaks = 40, freq = FALSE, xlim = c(5, 45), ylim = c(0, .1),
             yaxt = 'n', xlab = paste("Num Samples =", i), main = NULL)
        curve(dnorm(x, mean = mean(vars), sd = sd(vars)), add = TRUE, lwd = 1.5, col = "red")
}
title(main = "Dist of Variance of n Exp Random Variable", ylab = "Density of Variance", outer = TRUE )
axis(2, outer = TRUE)
```
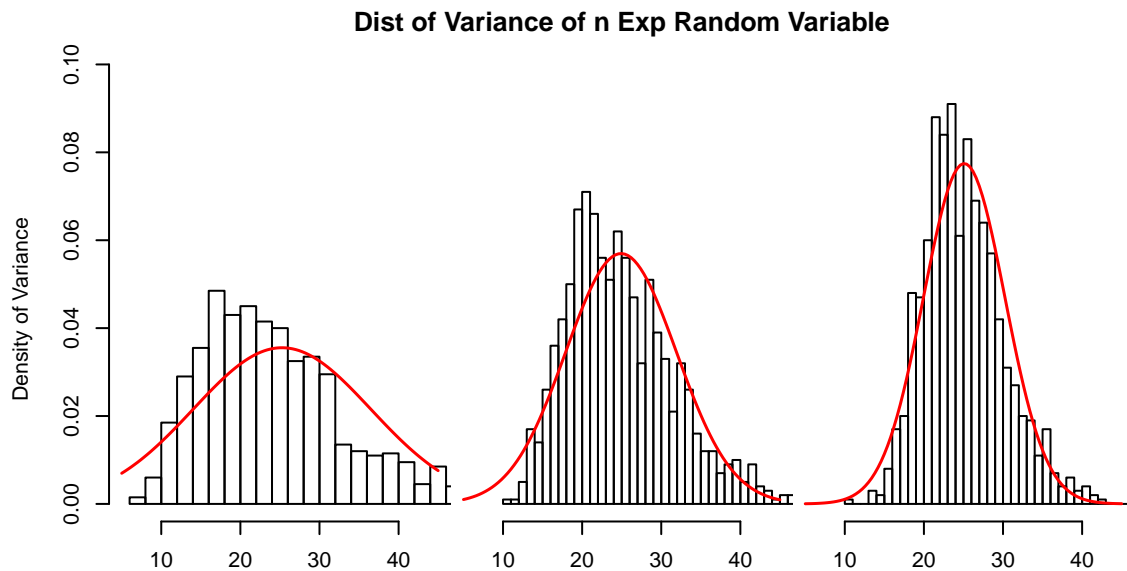

As the number of samples increases, it appears that the distribution of the variances trends more and more normal.