Statistical Inference Course Project – Simulation

Jerry Dumblauskas

April 10, 2016

This is an investigation the exponential distribution in R compared to the Central Limit Theorem. The exponential distribution was simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution was set to 1/lambda and the standard deviation was also 1/lambda. lambda = 0.2 for all of the simulations. I investigated the distribution of averages of 40 exponentials. Note, a thousand simulations were done.

I'll show 3 things:

1. Sample mean compared to theoretical mean
2. Variance of sample compared to theoretical variance
3. Investigate the distribution

Sample Mean v Theoretical Mean

```
set.seed(666773)
sims = 1000;
n = 40;
lambda = 0.2
means <- vector("numeric")
means_sum <- vector("numeric")
means_cum <- vector("numeric")

for (i in 1:sims) { means[i] <- mean(rexp(n, lambda))}
means_sum[1] <- means[1]
for (i in 2:sims) { means_sum[i] <- means_sum[i-1] + means[i] }
for (i in 1:sims) { means_cum[i] <- means_sum[i]/i }

##The sample means converged to:
print( means_cum[sims])
```

```
## [1] 4.983085
```

As the number increases the sample value moves to the theoretical value

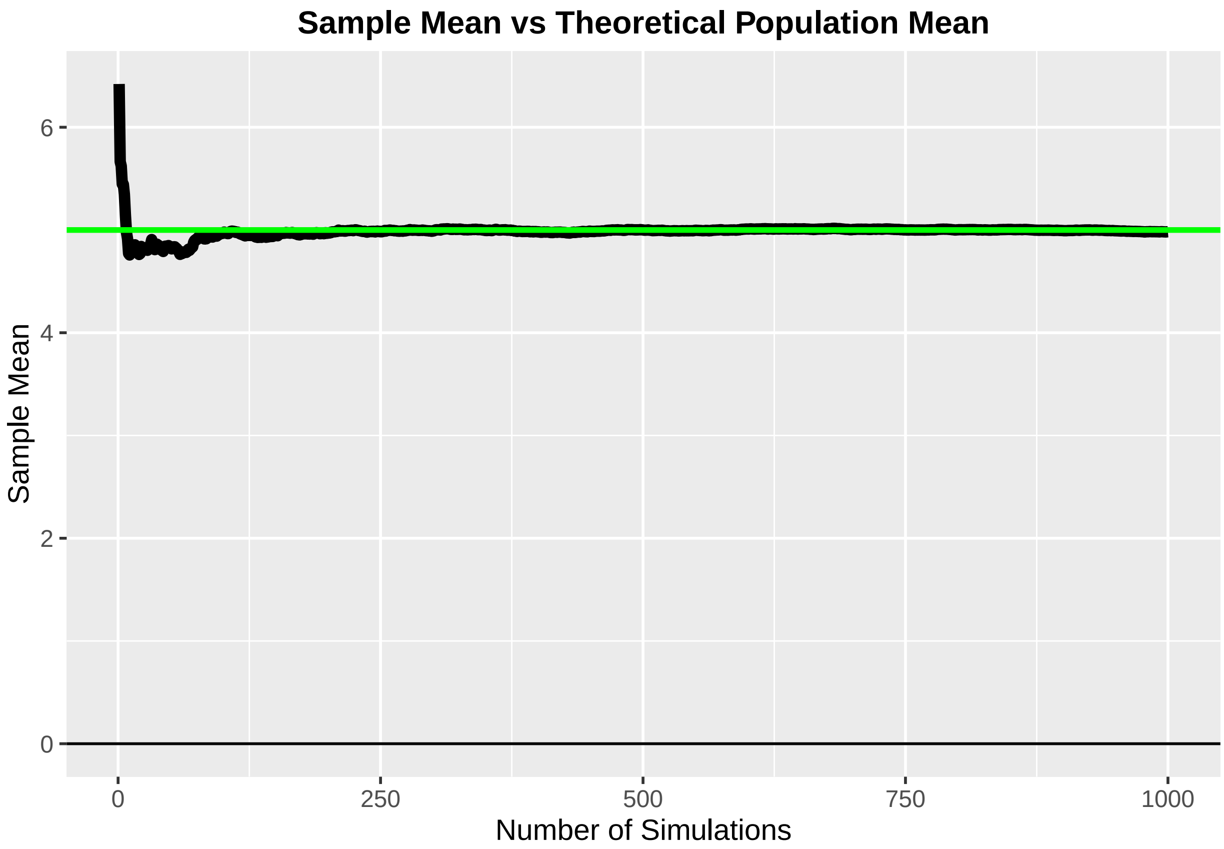
The theoretical value is:

```
print (1/lambda)
```

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

Plotting looks like:

```
library(ggplot2)
plt <- ggplot(data.frame(x = 1:sims, y = means_cum), aes(x = x, y = y))
plt <- plt + geom_hline(yintercept = 0) + geom_line(size = 2)
plt <- plt + geom_abline(intercept = 1 / lambda, slope = 0, color = "green", size = 1)
plt <- plt + theme(plot.title = element_text(size=12, face="bold", vjust=2, hjust=0.5))
plt <- plt + labs(title="Sample Mean vs Theoretical Population Mean")
plt <- plt + labs(x = "Number of Simulations", y = "Sample Mean")
print(plt)
```



Sample Variance vs Theoretical Population Variance

We now turn our attention to the variance. We will compare the variance present in the sample means of the 1000 simulations to the theoretical varience of the population.

The variance of the sample means estimates the variance of the population by using the varience of the 1000 entries in the means vector times the sample size, 40. (i.e sig=var(sm)*N)

Sample Variance

```
print (var(means)*n)
```

```
## [1] 25.63709
```

Theoretical Variance

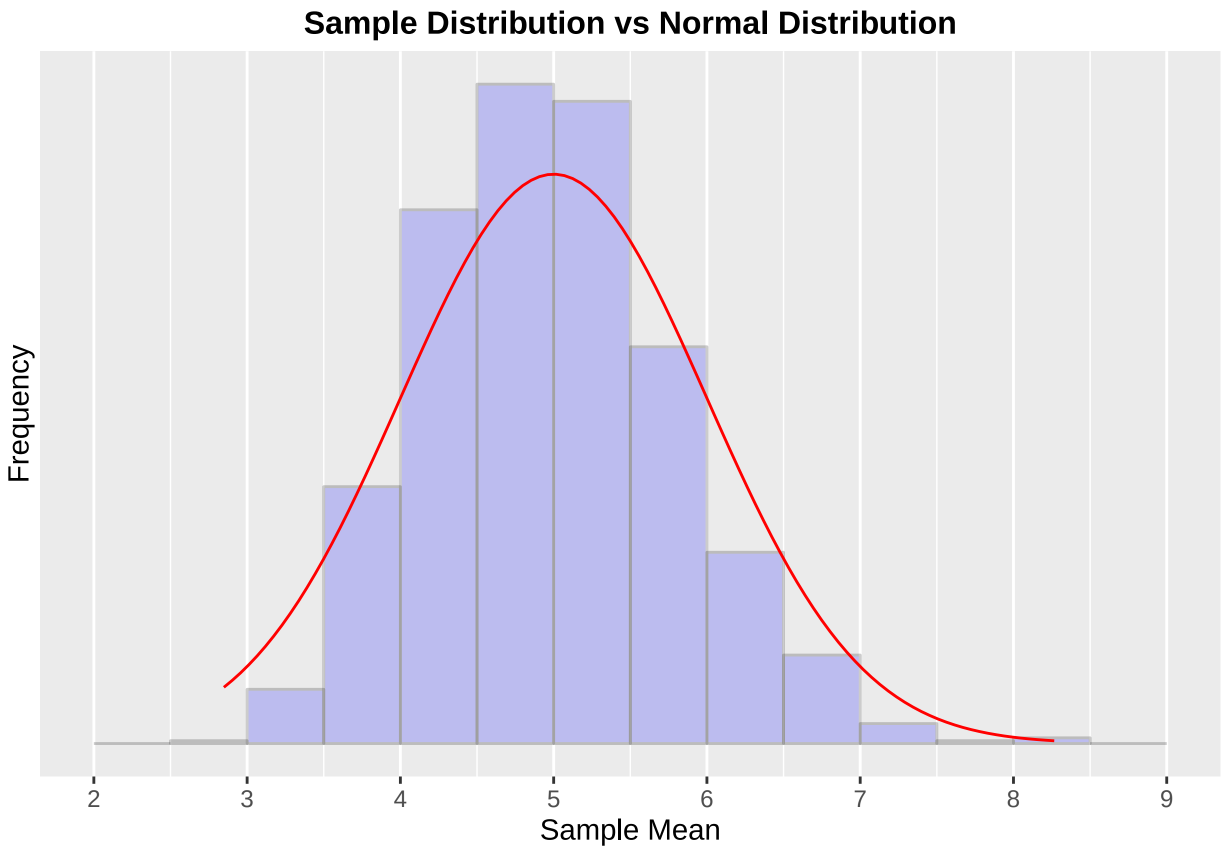
```
print ((1/lambda)^2)
```

```
## [1] 25
```

Distribution

The sample means are distributed normally around the population mean of 5. This can be easily seen by overlaying a normal curve on the histogram of the sample means.

```
library(ggplot2)
plt <- ggplot(data.frame(x = means), aes(x = x))
plt <- plt + geom_histogram(position="identity", fill="blue", color="black", alpha=0.2,
                       binwidth=0.5, aes(y= ..density..))
plt <- plt + stat_function(fun = dnorm, colour = "red", args=list(mean=5))
plt <- plt + scale_x_continuous(breaks=c(1, 2, 3, 4, 5, 6, 7, 8, 9))
plt <- plt + scale_y_continuous(breaks=c())
plt <- plt + theme(plot.title = element_text(size=12, face="bold", vjust=2, hjust=0.5))
plt <- plt + labs(title="Sample Distribution vs Normal Distribution")
plt <- plt + labs(x = "Sample Mean", y = "Frequency")
print(plt)
```



Conclusion

The sample mean closely matches the theoretical mean. The sample variance of the mean also tracks quite closely to the theoretical variance. The more samples we get the closer the density distribution be to the normal distribution bell curve.