Statistical Inference Final Project

By James Lamb

I. Theoretical Expectations

General Exponential Distribution

This project explores the asymptotic properties of the exponential distribution. To begin, I present some theoretical facts about the distribution here (courtesy of Ryerson University).

- Probability Density: For $x < 0$, $\text{PDF}_{\text{exp}} = 0$. For $x > 0$, $\text{PDF}_{\text{exp}} = \lambda e^{-x\lambda}$
- Cumulative Density: For $x < 0$, $\text{CDF}_{\text{exp}} = 0$. For $x > 0$, $\text{CDF}_{\text{exp}} = 1 - e^{-x\lambda}$
- Mean: $\text{E[X]} = \frac{1}{\lambda}$
- Variance: $\text{Var[X]} = \sigma^2 = \frac{1}{\lambda^2}$
- Standard Deviation: $\text{sd[X]} = \sigma = \frac{1}{\lambda}$

Expectations with Specific Simulation Parameters

In this project, I conduct simulations of the exponential distribution using R programming and the following parameters.

- $\lambda = 0.2$
- Random pulls per simulation $(n) = 40$
- Simulations $= 10000$

The simulation exercise will examine the distribution of averages of 40 draws from the exponential distribution. Given the above parameters and theoretical framework, the distribution of averages is expected to have the following properties:

- Centered near $\frac{1}{0.2} = 5.00$
- Variance of $\frac{\sigma^2}{n} = \frac{1}{n \cdot \lambda^2} = \frac{1}{0.2^2 \cdot 40} = 0.625$
- Standard deviation of $\frac{\sigma}{\sqrt{n}} = \frac{1}{\lambda} \cdot \frac{1}{\sqrt{n}} = \frac{1}{0.2\sqrt{40}} \approx 0.791$

II. Simulation Results

Running the Simulation

To begin, I simulated 10000 samples of 40 draws from an exponential distribution with the parameters described in the previous section.

```r
n <- 40
simnum <- 10000
lambda <- 0.2
simdata <- matrix(replicate(simnum,rexp(n,rate=lambda)),n,simnum) #simulate the random pulls
sim_means <- apply(simdata, 2, mean) #take column means
```

Simulation Results: Mean and Variance

The table below shows results of the simulation compared to theoretical expectations.

```
results_table <- data.frame(Theoretical=rep(0,3), Simulated=rep(0,3)) #initialize the data frame
rownames(results_table) <- c("Mean", "Variance", "Std. Deviation")
results_table[,1] = c(5.000, 0.6250, round(sqrt(0.625),4))
results_table[1,2] = round(mean(sim_means),4)
results_table[2,2] = round(var(sim_means),4)
results_table[3,2] = round(sd(sim_means),4)
results_table
```

```
##                Theoretical Simulated
## Mean                5.0000    5.0002
## Variance            0.6250    0.6175
## Std. Deviation      0.7906    0.7858
```

These results suggest that the distribution of averages of 10000 samples of 40 random draws from the exponential distribution *does* converge toward theoretical expectations. The mean and variance of the distribution of simulated means are very close to their expected theoretical values (within 2%).

Simulation Results: Normality

Though the mean and variance of the distribution of simulated means are close to their expected theoretical values, it does not appear that the means are normally distributed. Figure 1 shows a histogram of the simulated means. It is centered very close to the thoeretical mean of 5.00, but has a bit of excess kurtosis and positive (i.e. "right hand") skewness. These characteristics are used to create the Jarque-Bera statistic, which in this case clearly indicates that the null hypothesis of normality can be rejected with high confidence.

Simulation Results: Confidence Intervals

The project directions for this section of the report were unclear. In this section, I work under the assumption that that professor Caffo would like to see what percentage of the 95% confidence intervals from the 10000 means of n=40 draws contain the theoretical mean (5.00). Consider the following:

```
sim_vars <- apply(simdata,2,var)
sim_sd <- apply(simdata,2,sd)
sim_ll <- sim_means - 1.96*(sim_sd/sqrt(n))
sim_ul <- sim_means + 1.96*(sim_sd/sqrt(n))
coverage <- mean(sim_ll < (1/lambda) & sim_ul > (1/lambda))
noquote(paste("Coverage: ", round(coverage,2)*100, "%", sep=""))
```

```
## [1] Coverage: 92%
```

The estimated coverage is less than but close to 95%, consistent with the evidence from the previous discussion regarding the normality of this distribution. Namely, the distribution of 10000 means of 40 random draws from the exponential distribution with lambda=0.2 is close to the normal distribution but not quite normal.

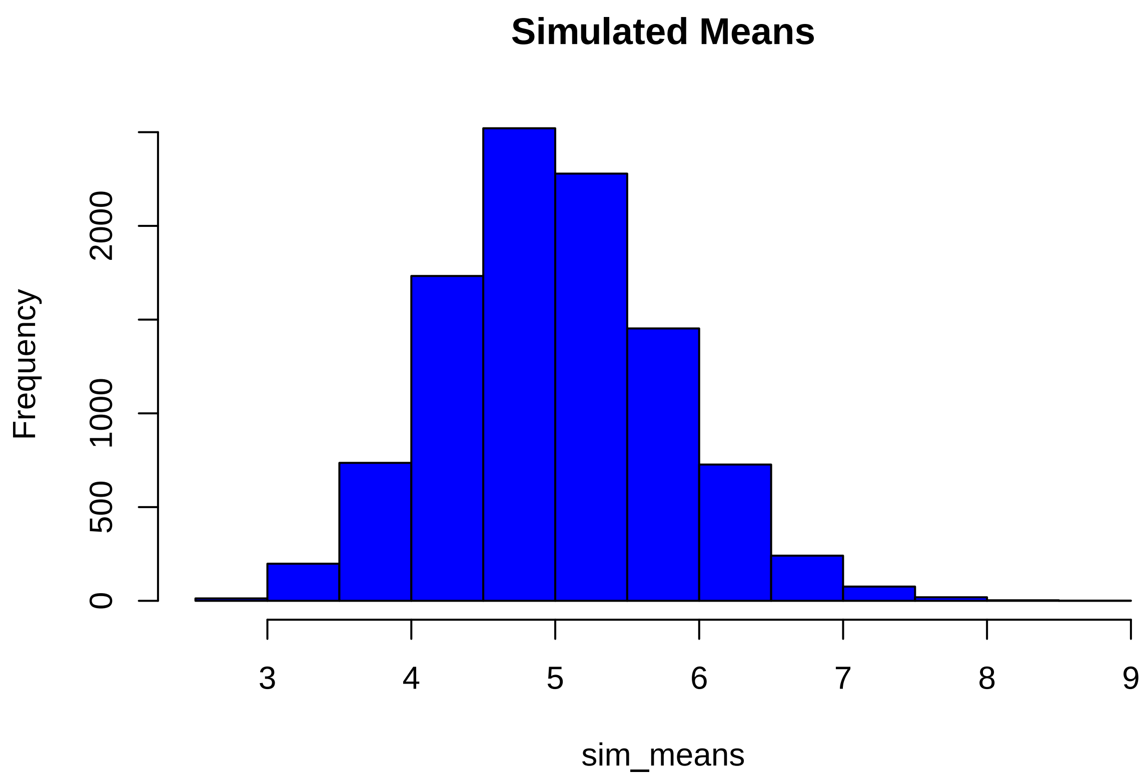
Conclusion

In this brief inferential analysis, I find some evidence that for 10,000 simulations of 40 random draws from an exponential distribution with $\lambda = 0.2$, the distribution of means of those draws converges to a close approximation of the theoretically expected values. However, the distribution of means remains stuboornly non-normally, with some excess kurtosis and a long right-hand-side tail (i.e. positive skewness.

The R markdown document with the underlying code used to generate this report can be found at this GitHub repo.

Appendix: Plots and Regression Results

Figure 1. Distribution of Simulated Averages




```
## [1] Jarque-Bera Test | JB Stat 140.68 p-value 0
```

```
## [1] Excess Kurtosis = 0.13731 | Skewness = 0.28218
```

References

[1] Pete at R-bloggers, "Using apply, sapply, lapply in R", *R-bloggers*, December 2012. Retrieved from: http://bit.ly/XTUKkm

[2] R-core team et. al, "Normality Tests fBasics", *inside-R*, 2014. Retrieved from: http://bit.ly/ZzE6rA

[3] Uknown Author, "Chap6: Poisson Process", Ryerson University, 2014. Retrieved from: http://bit.ly/1mo8f7k.