Demonstrating normal distribution models using exponential distribution means

Michael Fierro

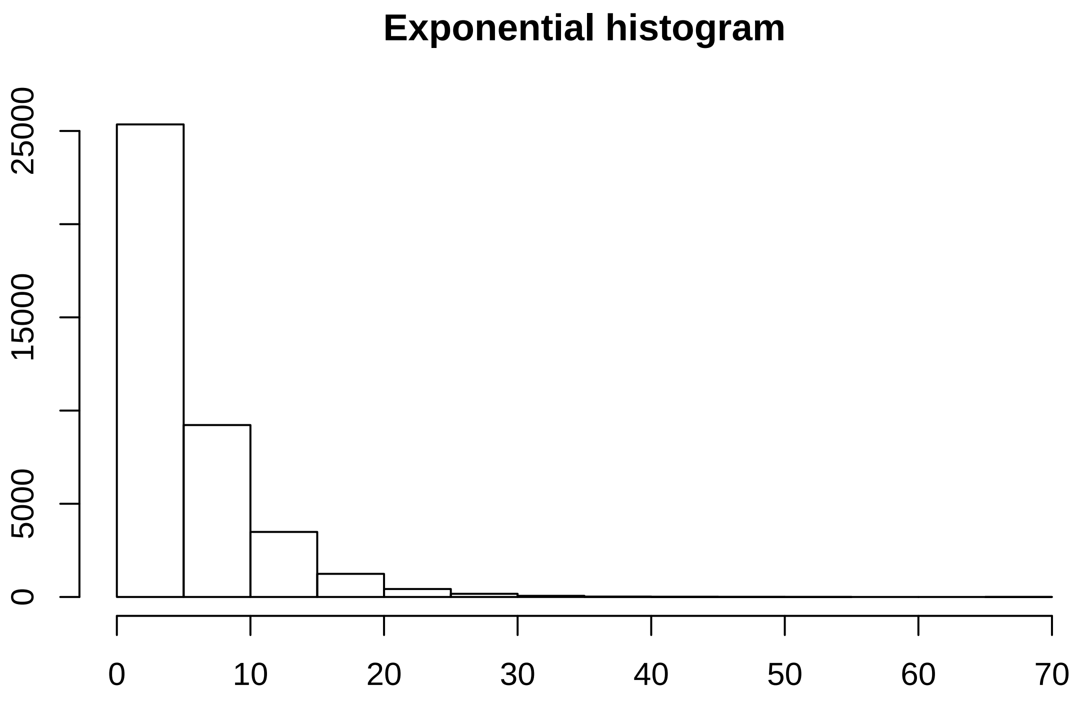
April 16, 2016

Synopsis

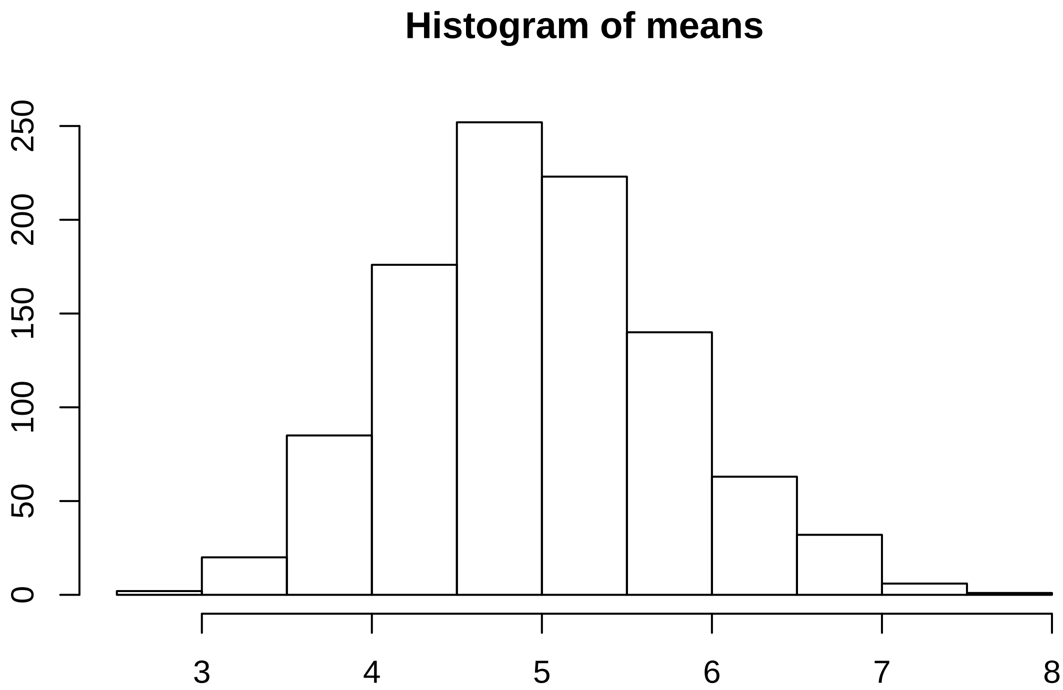
This study is a simulation of sampling data and applying observations from a sample to represent the larger population. This study will use a sample of averages of an exponential distribution, and will determine how the means, variance and distribution of that sample compares to theoretical values.

Report

By its definition, an exponential distribution is **not** a normal distribution, and as such, it does not make a Gaussian histogram. Instead, it is weighted, going from the highest values at 0 and flattening out to tail along the X axis, as shown with the histogram:



However, one would expect a sample of mean values to follow a normal distribution, and the histogram below shows that is exactly what happens:



Means

We can now compare the sample and theoretical data:

```
## [1] "Sample mean: 4.974"
```

```
## [1] "Theoretical mean: 5"
```

As expected, the sample mean closely approximates the theoretical mean, being within 0.026. Further samples would bring this even closer, as predicted by the Central Limit Theorem (CLT).

Variance

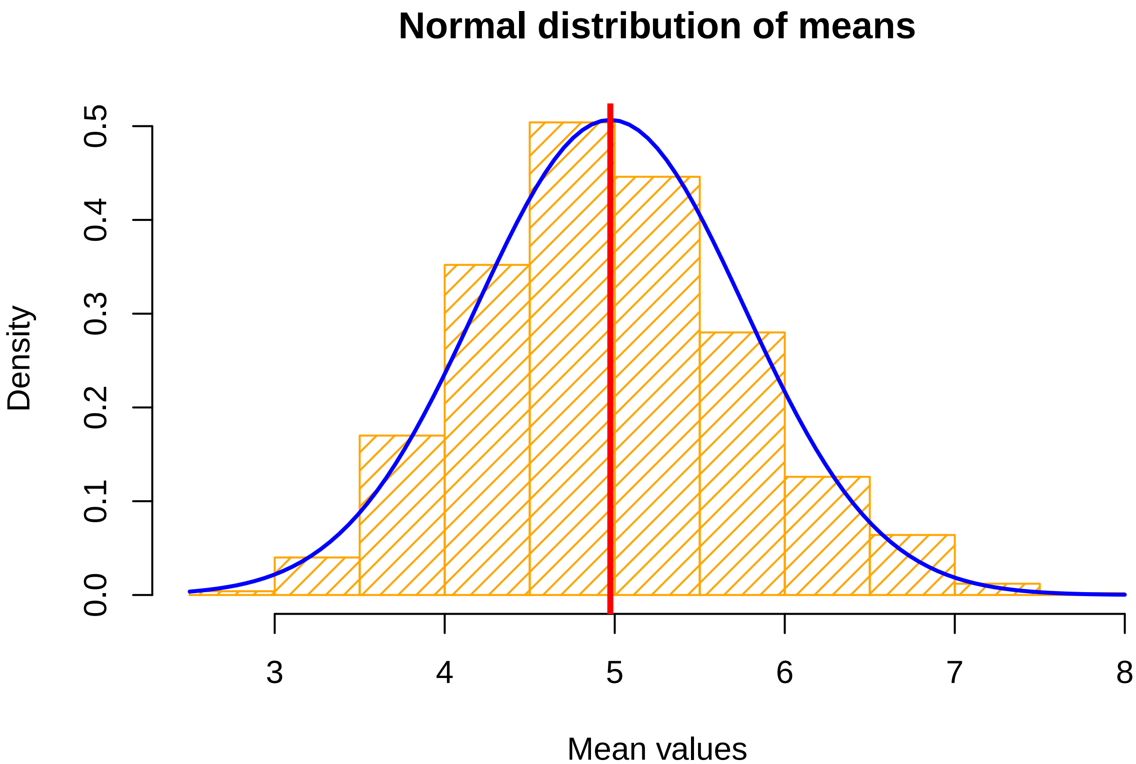
The same thing holds true with variances for sample and theoretical data:

```
## [1] "Sample variance: 0.62"
```

```
## [1] "Theoretical variance: 0.625"
```

Again, the difference between the variances is negligible and falls perfectly in line with the CLT.

Distribution



The graph plotted above shows a histogram of the means and shows that the sample is very close to a Normal Distribution. One way to prove this is to compare the confidence intervals for the sample and theoretical values. Fortunately, the previous two sections have given us all of the data we need to complete the calculations:

```
## Sample confidence levels:      4.73 5.218
```

```
## Theoretical confidence levels: 4.755 5.245
```

Once again, the values approximately match between the sample and theoretical measurements, with a very slight distinction between the two sets. This would seem to indicate that the sample distribution is Normal.

Appendix

Code snippets used to generate this report:

```
set.seed(2291)
lambda_class_1 <- .2
n_sims_class_1 <- 1000
n_obs_class_1 <- 40

initdata <- matrix(rexp(n_sims_class_1 * n_obs_class_1, rate=lambda_class_1),n_sims_class_1,n_obs_class_
initMeanData <- rowMeans(initdata)
```

```r
hist(initdata)
```

```r
hist(initMeanData)
```

```r
sample_mean_class_1 <- round(mean(initMeanData),3)
theor_mean_class_1 <- round(1/lambda_class_1,3)
paste("Sample mean:", sample_mean_class_1)
paste("Theoretical mean:", theor_mean_class_1)
```

Compute means

```r
sample_var_class_1 <- round(var(initMeanData),3)
theor_var_class_1 <- round(1/lambda_class_1^2/n_obs_class_1,3)
paste("Sample variance:", sample_var_class_1)
paste("Theoretical variance:", theor_var_class_1)
```

Compute variances

```r
hist(initMeanData, prob=TRUE, density = 15, col="orange")
curve(dnorm(x, mean=sample_mean_class_1, sd=sqrt(sample_var_class_1)), add=TRUE, col="blue", yaxt="n", l
abline(v = 4.974, untf = FALSE, lwd=3, col="red")
```

Plot normal distribution graph

```r
sample_conf_levels_class_1 <- round (sample_mean_class_1 + c(-1, 1) * 1.96 * sd(initMeanData) / sqrt(n_
theor_conf_levels_class_1 <- round (theor_mean_class_1 + c(-1, 1) * 1.96 * sqrt(theor_var_class_1) / sq
cat("Sample confidence levels:     ", sample_conf_levels_class_1)
cat("Theoretical confidence levels:", theor_conf_levels_class_1)
```

Compute confidence levels