Inference Analysis

Andrew Hope

April 6, 2018

Inference Analysis With an Exponential Distribution

Overview

This analysis will demonstrate inferential statistics techniques to evaluate sampling mean and variance for a large sample with an underlying exponential distribution. The analysis will begin by creating a simulated dataset, and will then calculate and compare sampling statistics. A final section provides plots to better convey the conclusions.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.3
```

Simulations

Create a set of 1000 random samples from an exponential distribution with lambda equal to 0.2. Each sample has a size n = 40. Begin by generating all of the needed data points. In this analysis we round the number to 4 decimal places for the sake of keeping the data simpler and cleaner.

```
r <- round(rexp(40000, rate = 0.2), 4)
```

Move the values into a matrix to create 1000 samples of size n = 40.

```
rmat <- matrix(r, 1000, 40)
head(rmat, 3)
```

```
##        [,1]   [,2]    [,3]   [,4]   [,5]   [,6]   [,7]   [,8]    [,9]
## [1,] 0.0688 0.2309 11.8643 1.9834 1.7215 1.6308 9.4865 1.0068  1.1169
## [2,] 1.3132 2.5991 16.0785 3.0040 1.9379 1.5586 1.1880 0.3826 15.2322
## [3,] 0.5820 0.9639  0.1234 0.5143 1.7711 9.5237 4.7410 1.3867  5.9174
##       [,10]  [,11]  [,12]  [,13]  [,14]  [,15]  [,16]   [,17]  [,18]
## [1,] 1.4986 2.3087 6.7699 6.7995 2.0134 3.8728 0.1034 14.0805 3.5016
## [2,] 3.4151 2.8605 3.5573 0.2810 4.7891 2.2559 5.5313 17.2180 7.7712
## [3,] 7.6438 5.1306 1.4529 5.5292 0.8029 1.0756 0.5515 23.1904 0.1509
##       [,19]  [,20]  [,21]   [,22]  [,23]  [,24]  [,25]  [,26]  [,27]
## [1,] 6.6569 9.3918 5.9027 15.9961 4.6892 0.5167 9.2882 1.0613 5.4805
## [2,] 0.7769 0.7904 2.4412  7.5401 0.6609 3.0339 1.2981 2.8424 7.9843
## [3,] 1.2306 5.2664 2.3770  1.2867 6.1054 7.6082 1.2349 3.0488 0.6645
##       [,28]   [,29]   [,30]  [,31]  [,32]  [,33]   [,34]   [,35]  [,36]
## [1,] 7.2730  3.4850 10.9592 4.0000 3.5633 0.5092  3.7904  4.9192 2.2318
## [2,] 5.3055 13.3322  2.1134 2.7604 0.7047 9.1409 17.3898  1.1841 0.1598
## [3,] 5.5290 12.5244 11.5298 0.3209 3.3494 0.6285  3.6496 10.2970 2.6536
##        [,37]  [,38]  [,39]  [,40]
## [1,]  6.6917 2.3241 1.3586 0.0652
## [2,]  3.7313 0.6614 5.2897 1.5439
## [3,] 11.6123 6.1909 7.4858 0.9361
```

Sample Mean vs. Theoretical Mean

Find the mean for each of the 1000 rows, and take the mean of means to find the sampling mean.

```
sampling <- apply(rmat, 1, mean)
m <- mean(sampling)
m
```

```
## [1] 4.994441
```

The Central Limit Theorm allows us to conclude that this sampling mean is approximately equal to the theoretical mean.

The theoretical mean of an exponential distribution is equal to 1/lambda.

```
1/0.2
```

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

Indeed, our sampling mean is very close to the theoretical mean.

Sample Variance vs. Thoretical Variance

Calculate the population variance.

```
r.var <- var(r)
r.var
```

```
## [1] 25.04168
```

This should be close to the theoretical variance of an exponential distribution, which is equal to 1/lambda^2.

```
1/0.2^2
```

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

Calculate the variance of the 1000 sample means to find the sampling variance.

```
v <- var(sampling)
v
```

```
## [1] 0.5657834
```

This variance of the sample means should be approximately equal to its theoretical equivalent, (1/lambda^2)/n

```
(1/0.2^2)/40
```

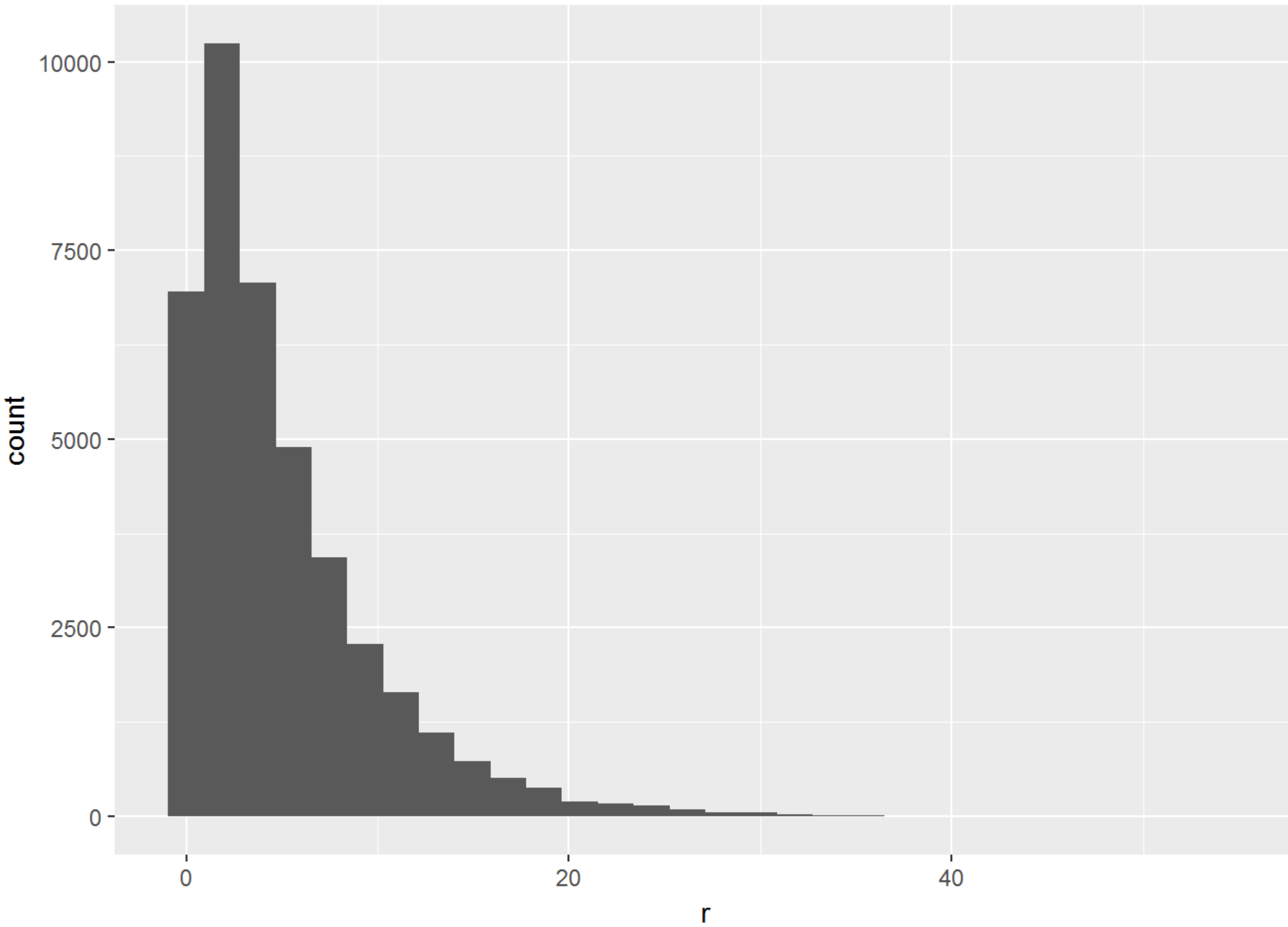
```
## [1] 0.625
```

Distribution

There is a significant difference between the distribution of an exponential random variable, and the distribution of means of a set of samples of exponential random variables.

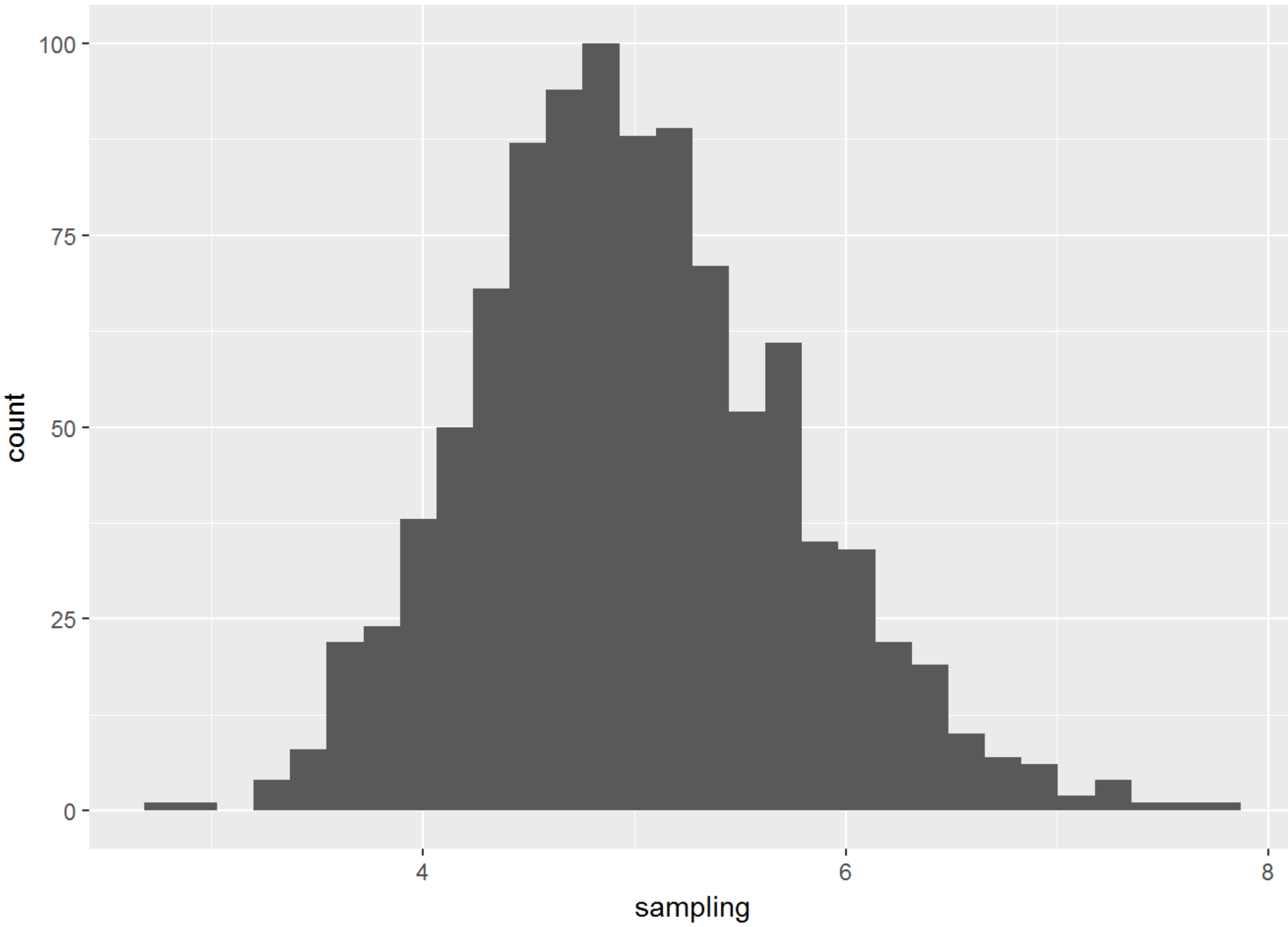
First, look at a distribution of a single exponential random variable.

```
qplot(r, bins = 30)
```



Now, look at the distribution of the 1000 mean values from the simulated data set.

```
qplot(sampling, bins = 30)
```



This demonstrates the Central Limit Threorem. The distribution of a sample’s statistic, across a set of samples, will resemble a normal distribution.