

Comparing the Exponential Distribution with the Central Limit Theorem

Michael Connolly

Saturday, May 23, 2015

The exponential distribution is a well-known probability distribution that describes the time between events in a Poisson process. The defining parameter of this distribution is λ , which is the constant average rate of the event, which must be greater than 0.

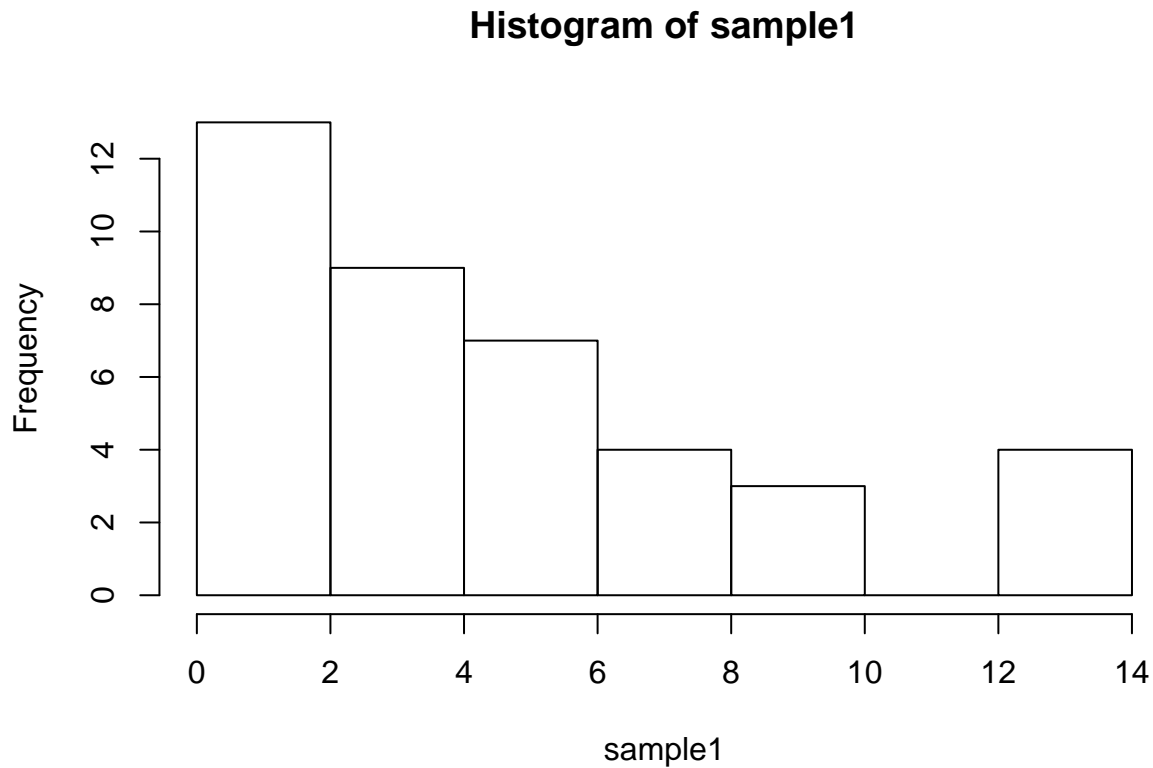
An example of such a distribution would be the time until a radioactive particle decays. Or, while walking down the street of New York City, the length of time it takes for a fellow pedestrian to smile at you.

The Central Limit Theorem (also known as CLT) is also a well-known concept in statistics, stating that the mean of a large enough number of independent variables will be approximately normally distributed.

As we have discussed in the course Statistical Inference on Coursera.com, it turns out that the average mean of a sample of a given population (say, the average value of 40 samples) is itself an independent random variable. And, that this particular random variable will conform to the CLT, i.e., be normally distributed. And, the more data points you add, the more normal it will look.

This is what one exponential distribution looks like, using a λ of 0.2, and a sample size of 40.

```
sample_size <- 40
lambda <- 0.2
sample1 <- rexp(sample_size, lambda)
hist(sample1)
```

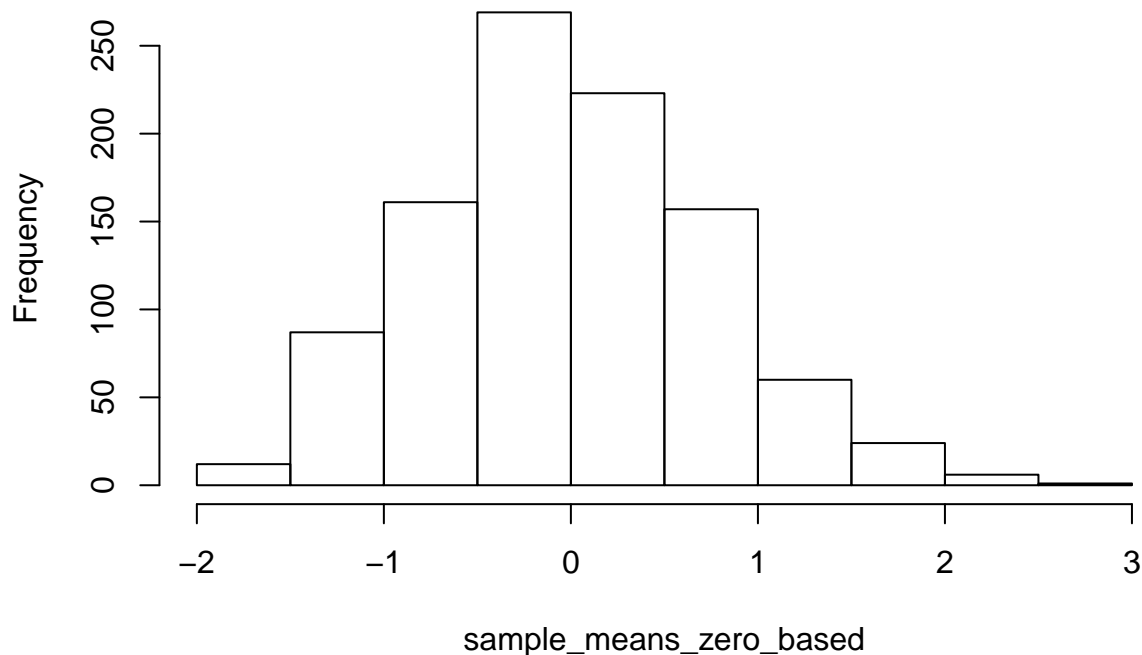


Note that the expected mean of an exponential distribution is $(1 / \lambda)$, which, in this case, equals 5. The variance is $(1 / \lambda^2)$, or 25, and the standard deviation is equal to the mean, or, 5.

This is what the distribution of the average (i.e., mean) of 1,000 simulations of the exponential distribution, which each simulation also using λ 0.2 and a sample size of 40. In order to center the distribution at zero, I have subtracted the mean from each simulation average.

```
sample_means = NULL
sample_means_zero_based = NULL
sample_size <- 40
lambda <- 0.2
variance <- (1 / (lambda^2))
expected_mean <- 5.0
standard_deviation <- expected_mean
simulation_count <- 1000
for (i in 1 : simulation_count) {
  sample <- rexp(sample_size, lambda)
  sample_mean = mean(sample)
  sample_means_zero_based = c(sample_means_zero_based, (sample_mean - expected_mean))
  sample_means = c(sample_means, sample_mean)
}
hist(sample_means_zero_based)
```

Histogram of sample_means_zero_based



So, did the distribution of the averages of these 1,000 simulations end up being normally distributed, as predicted? It most certainly looks like it. The distribution is clearly centered at the mean, as expected, and is shaped like a normal distribution.

However, the variance is not inline with predictions. This sample distribution would have an expected variance of (σ^2 / n) , where σ is the standard deviation of the population, and n is the sample size (in this case, 1000). That is not the variance that is calculated from the data. The variance of the data remains much wider. This implies that exponential distributions do not have their average variances get closer to zero as their population sizes increase.

```
expected_var_mean = variance / simulation_count
calculated_var_mean = var(sample_means)
expected_var_mean
```

```
## [1] 0.025
```

```
calculated_var_mean
```

```
## [1] 0.5686026
```

```
library(pdflatex)
```