

An Investigation of the Exponential Distribution

Mark E. Drummond

2015-05-23

Contents

Overview	1
Requirements	1
Simulations	1
Results	2
Expectations	2
Sample Mean vs. Theoretical Mean	3
Sample Variance vs. Theoretical Variance	3

Overview

We investigate the properties of the [Exponential distribution](#) and compare it with the expected theoretical results expressed by the [Central Limit Theorem](#).

The Central Limit Theorem tells us that:

“...given a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from the same population will be approximately equal to the mean of the population. Furthermore, all of the samples will follow an approximate normal distribution...” – http://www.investopedia.com/terms/c/central_limit_theorem.asp

In simple terms: Take independent random samples of size n from a population. As the sample size n and/or the number of samples increases, the distribution of the sample means will approach a Normal distribution. See below for additional explanation.

Requirements

1. Show the distribution of sample means and compare it to the theoretical mean of the distribution.
2. Show the distribution of sample variances and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Simulations

The exponential distribution can be simulated in R using the `rexp(n , $lambda$)` function where n is the sample size and $lambda$ (λ) is the rate parameter. Both the mean and the standard deviation of the exponential distribution are $\frac{1}{\lambda}$.

We run 1,000 simulations of the exponential distribution, using a sample size n of 40 and a rate parameter λ of 0.2. For each simulation we record the mean, variance, and standard deviation of the simulated data and store all the results in data frame for further analysis.

```
n <- 40          # Sample size
lambda <- 0.2    # Rate parameter
number_of_simulations <- 1000

mu <- 1 / lambda # Theoretical mean
sigma <- 1 / lambda # Theoretical standard deviation

sample_means <- NULL
sample_variances <- NULL

for (experiment in 1 : number_of_simulations) {
  sample <- rexp(n = n, rate = lambda)
  sample_means <- c(sample_means, mean(sample))
  sample_variances <- c(sample_variances, var(sample))
}

sample_statistics <- data.frame(cbind(sample_means, sample_variances))
names(sample_statistics) <- c("sample_means", "sample_variances")

summary(sample_statistics)
```

```
##   sample_means  sample_variances
##   Min.   :2.960   Min.    : 5.865
##   1st Qu.:4.461   1st Qu.:17.312
##   Median :4.984   Median :23.326
##   Mean   :5.025   Mean    :25.463
##   3rd Qu.:5.543   3rd Qu.:30.800
##   Max.   :7.911   Max.    :83.263
```

Results

Expectations

The Central Limit Theorem tells us that the distribution of sample means tends towards the Normal distribution as the sample size and the number of samples (simulations) increases, regardless of the distribution of the original data, assuming the samples are independent and identically distributed (iid).

In other words, as we increase the sample size (n) and/or the number of simulations, we expect the distribution of the sample means to tend toward the normal, despite our samples being drawn from the Exponential distribution. Given enough simulations we expect the distribution of these sample means to be approximating Normally distributed.

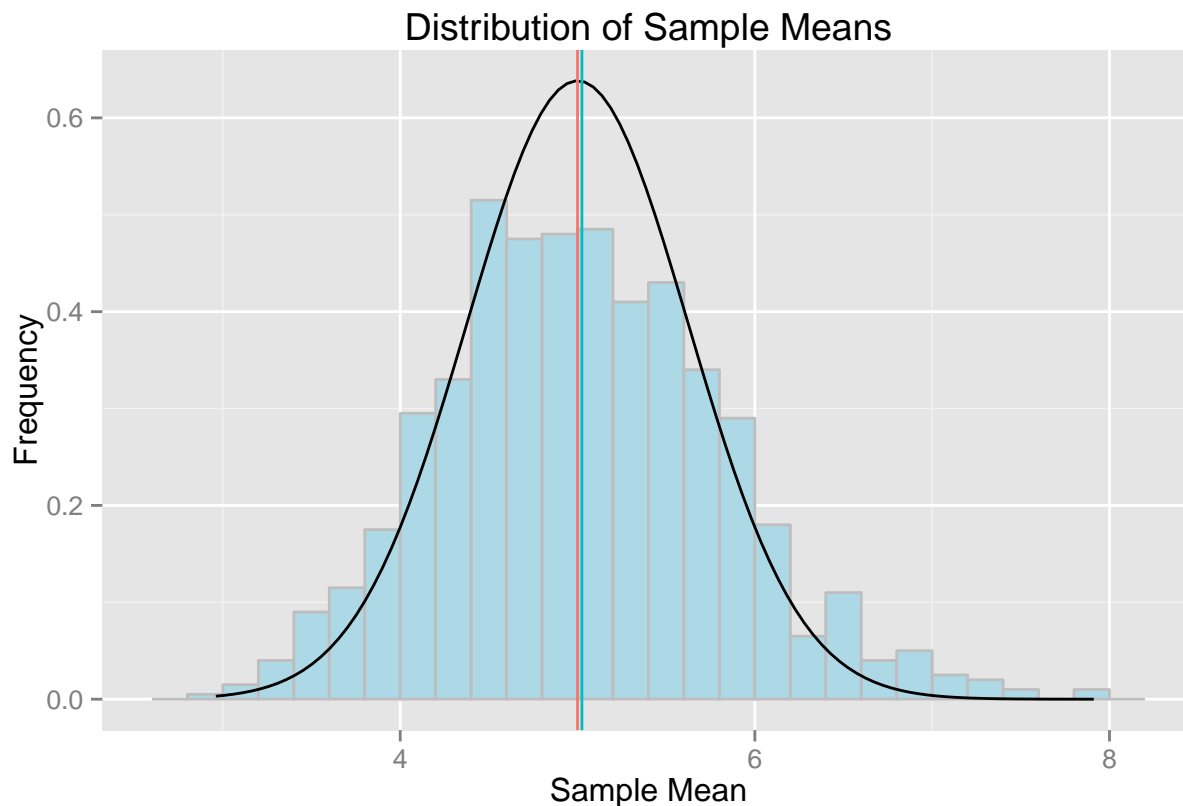
In this specific case we expect

1. The sample means to be approximately normally distributed with the mean of the samples approximately equal to the theoretical mean of the Exponential distribution.
2. The standard deviation of the samples to be approximately equal to the theoretical standard deviation of the Exponential distribution.
3. The variance of the samples to be approximately equal to the theoretical variance of the Exponential distribution.

Sample Mean vs. Theoretical Mean

```
library(ggplot2)
```

```
x_bar <- mean(sample_statistics$sample_means)
ggplot(data = sample_statistics, aes(x = sample_statistics$sample_means)) +
  geom_histogram(aes(y = ..density..),
    binwidth = 0.2, fill = "lightblue", colour = "grey") +
  stat_function(fun = dnorm, args = list(mean = mu, sd = sigma^2/n)) +
  geom_vline(aes(xintercept = mu, colour = "blue")) +
  geom_vline(aes(xintercept = x_bar, colour = "red")) +
  ggtitle("Distribution of Sample Means") +
  xlab("Sample Mean") + ylab("Frequency")
```

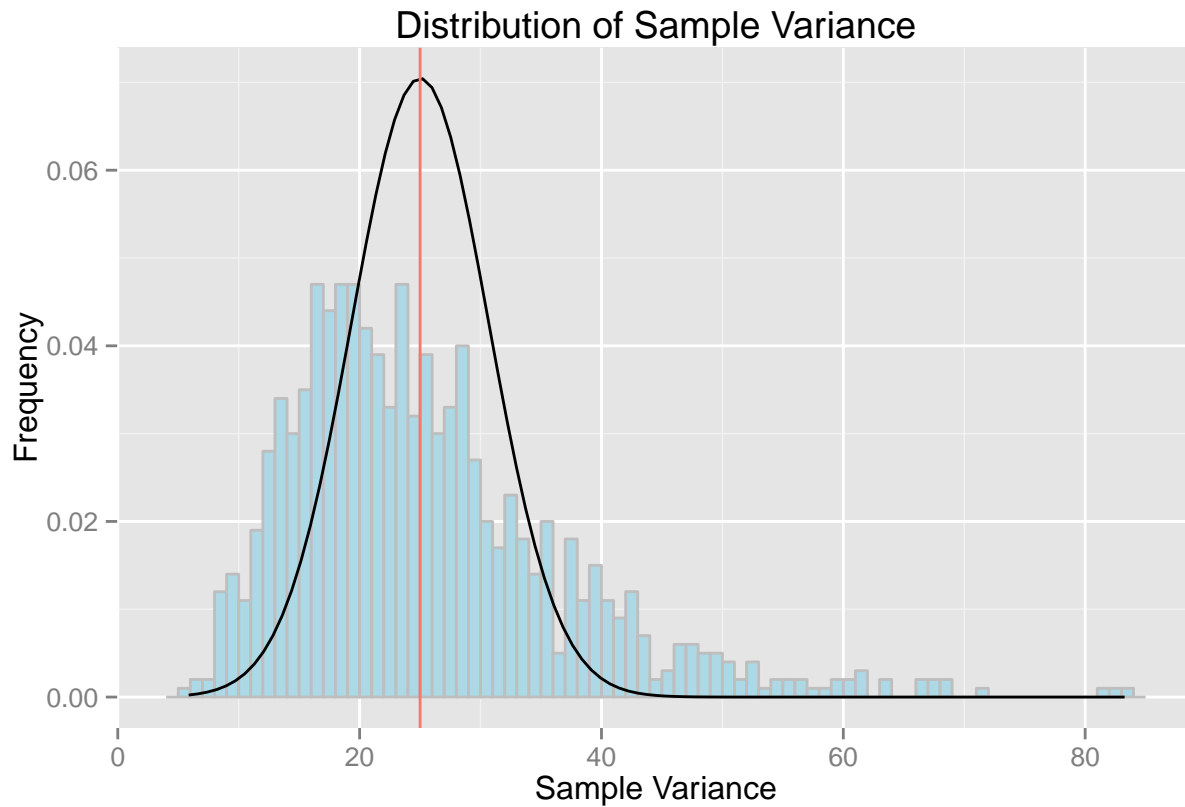


As we can see from the plot, the distribution of the sample means approximates a Normal distribution with the mean of the sample means (red vertical line) very close to the theoretical mean (blue vertical bar).

Sample Variance vs. Theoretical Variance

```
ggplot(data = sample_statistics, aes(x = sample_statistics$sample_variances)) +
  geom_histogram(aes(y = ..density..),
    binwidth = 1, fill = "lightblue", colour = "grey") +
```

```
stat_function(fun = dnorm, args = list(mean = sigma^2, sd = sqrt(2*sigma^4/(n-1)))) +
geom_vline(aes(xintercept = sigma^2, colour = "blue")) +
ggtitle("Distribution of Sample Variance") +
xlab("Sample Variance") + ylab("Frequency")
```



As with the sample means, here we see the sample variances tending toward the Normal distribution.