An Investigation of the Exponential Distribution

Mark E. Drummond

2015-05-23

Contents

Overview

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

Requirements

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

Theory

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 independant and identically distributed (iid).

In other words: 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 dis.

We expect:

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

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 and the variance of the simulated data and store all the results in data frame for further analysis.

```r
# Clear the working environment
rm(list = ls())
# We set the seed for consistent results. Comment this line to allow for a
# random seed each time the code is run.
set.seed(19647)
```

```r
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")

str(sample_statistics)
```

```
## 'data.frame':    1000 obs. of  2 variables:
##  $ sample_means    : num  5.96 5.11 3.94 3.67 3.93 ...
##  $ sample_variances: num  36.19 26.33 23.04 14.93 9.89 ...
```

```r
summary(sample_statistics)
```

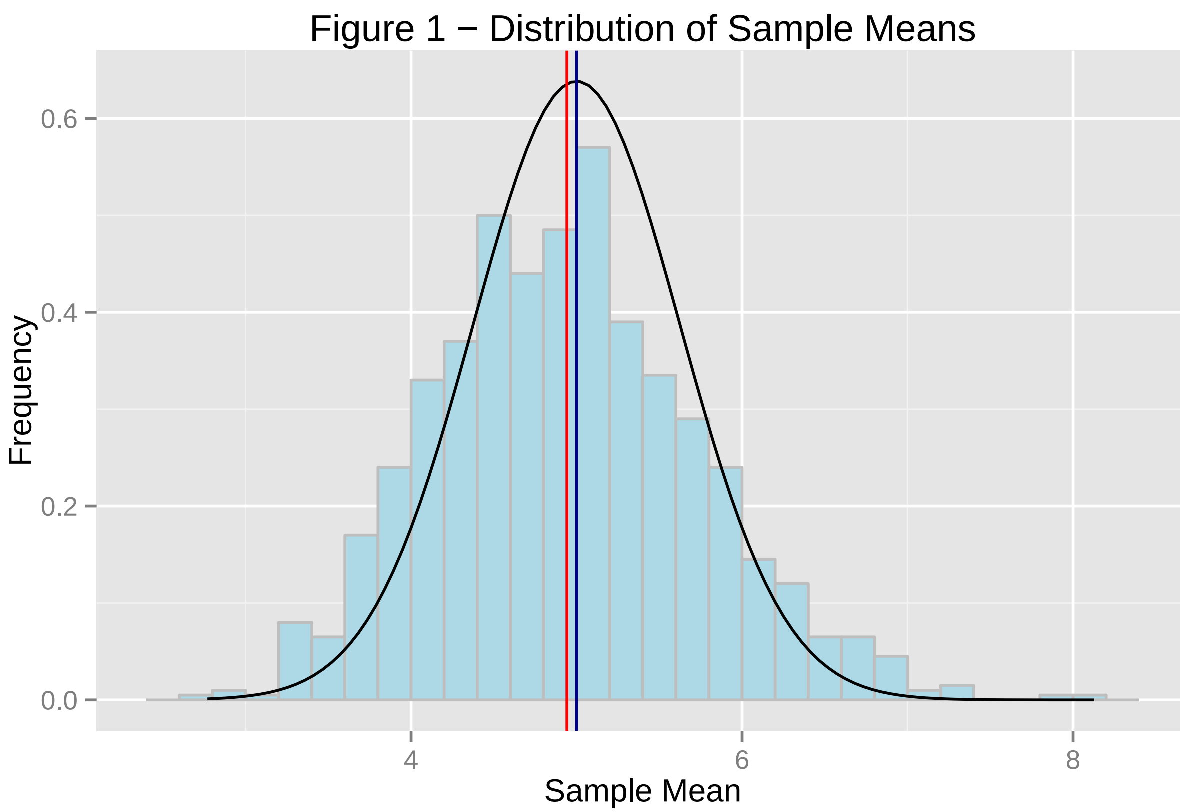
```
##   sample_means   sample_variances
##  Min.   :2.769   Min.   : 5.737
##  1st Qu.:4.381   1st Qu.:16.931
##  Median :4.926   Median :22.106
##  Mean   :4.941   Mean   :24.474
##  3rd Qu.:5.463   3rd Qu.:29.834
##  Max.   :8.128   Max.   :90.427
```

Results

Sample Mean vs. Theoretical Mean

```
library(ggplot2)
```

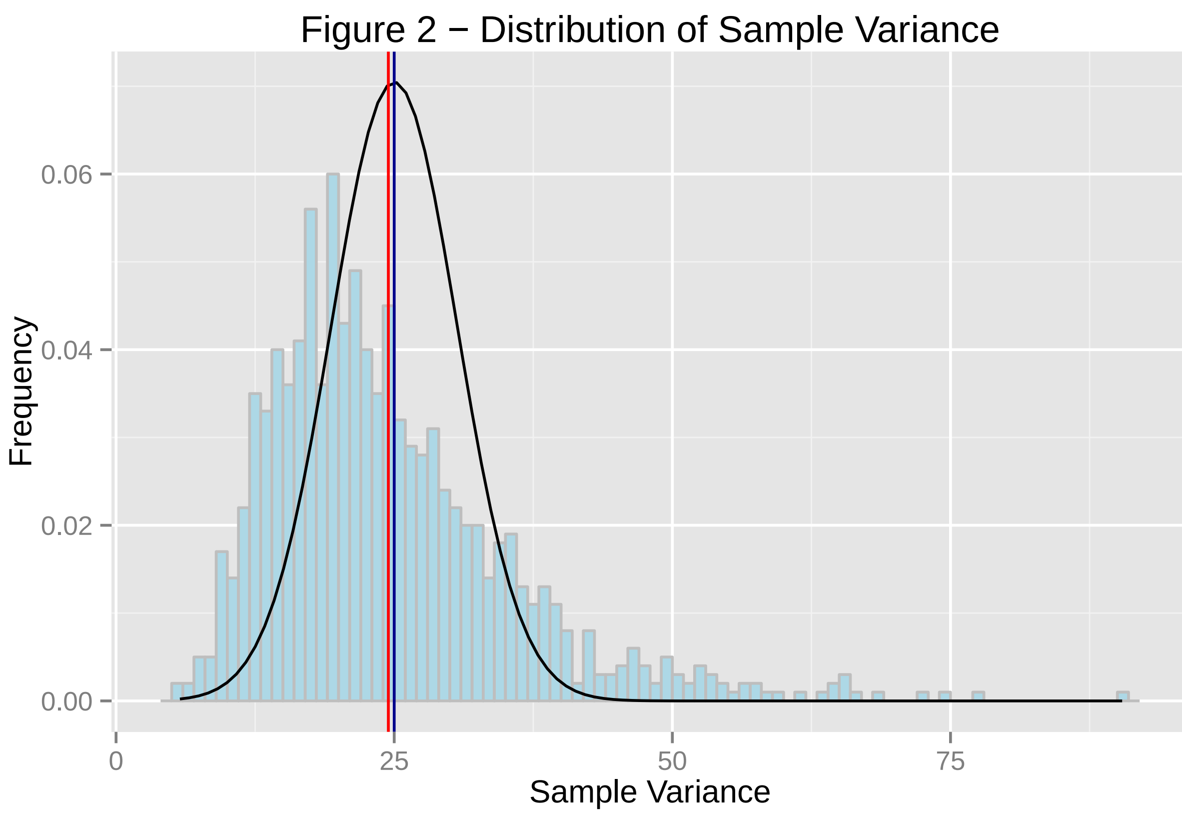
```
x_bar <- mean(sample_statistics$sample_means)
ggplot(data = sample_statistics, aes(x = sample_statistics$sample_means)) +
  geom_histogram(aes(y = ..density..),
                 binwidth = 0.2, fill = "lightblue", colour = "grey") +
  stat_function(fun = dnorm, args = list(mean = mu, sd = sigma^2/n)) +
  geom_vline(xintercept = mu, colour = "darkblue") +
  geom_vline(xintercept = x_bar, colour = "red") +
  ggtitle("Figure 1 - Distribution of Sample Means") +
  xlab("Sample Mean") + ylab("Frequency")
```



As we can see from the plot, the mean of the sample means (red veritical line) very close to the theoretical mean (blue vertical bar), as expected.

Sample Variance vs. Theoretical Variance

```r
S <- mean(sample_statistics$sample_variances)
ggplot(data = sample_statistics, aes(x = sample_statistics$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(xintercept = sigma^2, colour = "darkblue") +
  geom_vline(xintercept = S, colour = "red") +
  ggtitle("Figure 2 - Distribution of Sample Variance") +
  xlab("Sample Variance") + ylab("Frequency")
```



As with the sample means, here we see the mean of the sample variances (red vertical line) approximates the theoretical variance (blue vertical line).

Distribution

As we can see from the Figure 1, the distribution of the sample means approximates a Normal distribution.