Application of Central Limit Theorem to Exponential Distribution - Example

Amade A.

December 2, 2015

Overview

This paper describes the application of Central Limit Theorem to the distribution of means of exponential distribution. The emphasis is put on empirical presentation of a problem, with a random sample of 1000 means and comparing their distribution with the theoretical distribution of the mean and theoretical variance.

Simulations

```
n <- 1000
sampleSize <- 40
lambda <- 0.2
```

For the purpose of presentation, random sample of **1,000** means of exponential distribution needs to be generated. Each mean is calculated based on **40** values from exponential distribution with parameter λ of **0.2**. In turn, in order to obtain the sample, **40,000** random numbers need to be generated.

The following code generates the sample.

```
set.seed(2015)
means <- apply(matrix(rexp(n*sampleSize, rate = lambda),n,sampleSize),1,mean)
```

Sample Mean versus Theoretical Mean

The sample mean can be calculated with the following code.

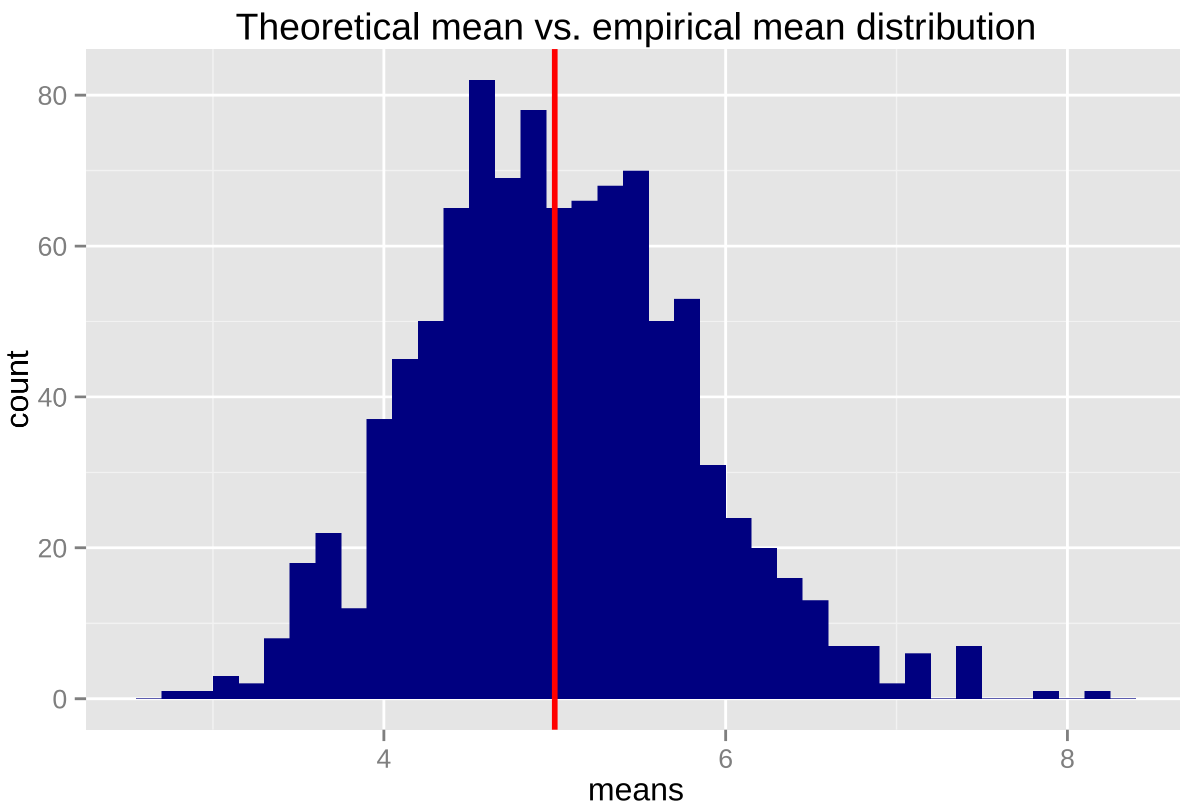
```
sampleMean <- mean(means)
sampleMean
```

```
## [1] 5.011563
```

Theoretical mean of a sample coming from exponential distribution equals $\frac{1}{\lambda}$. Assuming λ of 0.2, theoretical mean is $\frac{1}{\lambda} = \frac{1}{0.2} = 5$.

The best way to assess the magnitude of a difference in exmpirical and theoretical means is looking at a graph, which shows the histogram of expirical means.

```
library(ggplot2)
m <- ggplot() + aes(x=means)
m <- m + geom_histogram(fill = "navy", binwidth = .15)
m <- m + geom_vline(xintercept = (1/lambda), colour = "red", size = 1)
m + ggtitle("Theoretical mean vs. empirical mean distribution")
```



Theoretical sample mean is shown as red line. It can be clearly seen on the graph, that the means are centered around its value (5). In relative terms, sample mean is only 0.23% different from theoretical mean (calculated as $abs(\frac{sample\ mean - theoretical\ mean}{theoretical\ mean})$).

Sample Variance versus Theoretical Variance

Theoretical variance of the sampling distribution of the mean is calculated as:

$\sigma_M^2 = \frac{\sigma^2}{N}$

where σ^2 is a theoretical variance of the distribution and N is a sample size. In the case of exponential distribution, theoretical variance equals $\frac{1}{\lambda^2}$. With $\lambda = 0.2$ and $N = 40$, theoretical variance of the sampling distribution of the mean equals

```
theorVar <- (1/lambda^2)/(sampleSize)
theorVar
```

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

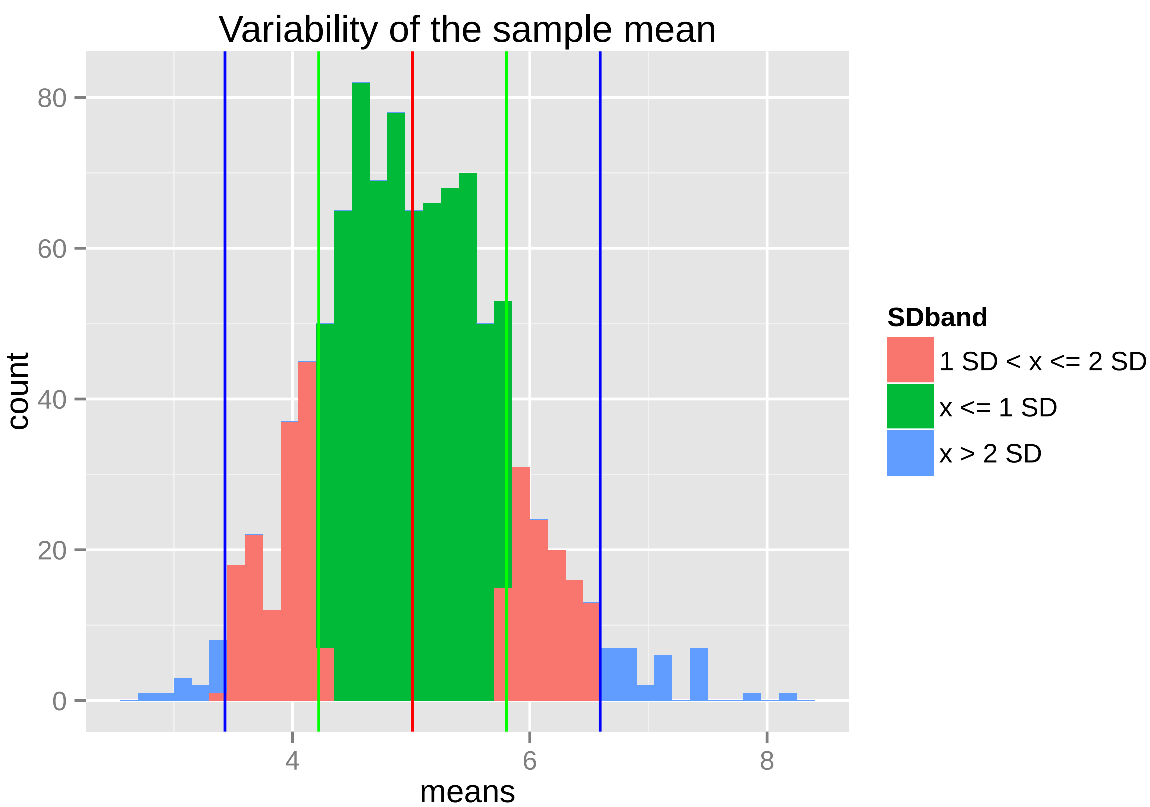
Unbiased estimator of the population variance, based on the sample data, calculated using *var()* function equals 0.641. It is 2.57% different from theoretical variance.

Variability of the sample mean is shown on the graph below.

```
SDmarker <- function(x) {
    if (x > 2*sqrt(theorVar)){
        "x > 2 SD"
    } else if (x > sqrt(theorVar)){
```

```
        "1 SD < x <= 2 SD"
    } else {
        "x <= 1 SD"
    }
}
SDband <- as.factor(sapply(abs(means - sampleMean),SDmarker))

m <- ggplot() + aes(x=means, fill = SDband)
m <- m + geom_histogram(binwidth = .15)
m <- m + geom_vline(xintercept = sampleMean, colour = "red", size = .5)
m <- m + geom_vline(xintercept = sampleMean + c(-1,1)*sqrt(theorVar),
                    colour = "green", size = .5)
m <- m + geom_vline(xintercept = sampleMean + c(-2,2)*sqrt(theorVar),
                    colour = "blue", size = .5)
m + ggtitle("Variability of the sample mean")
```

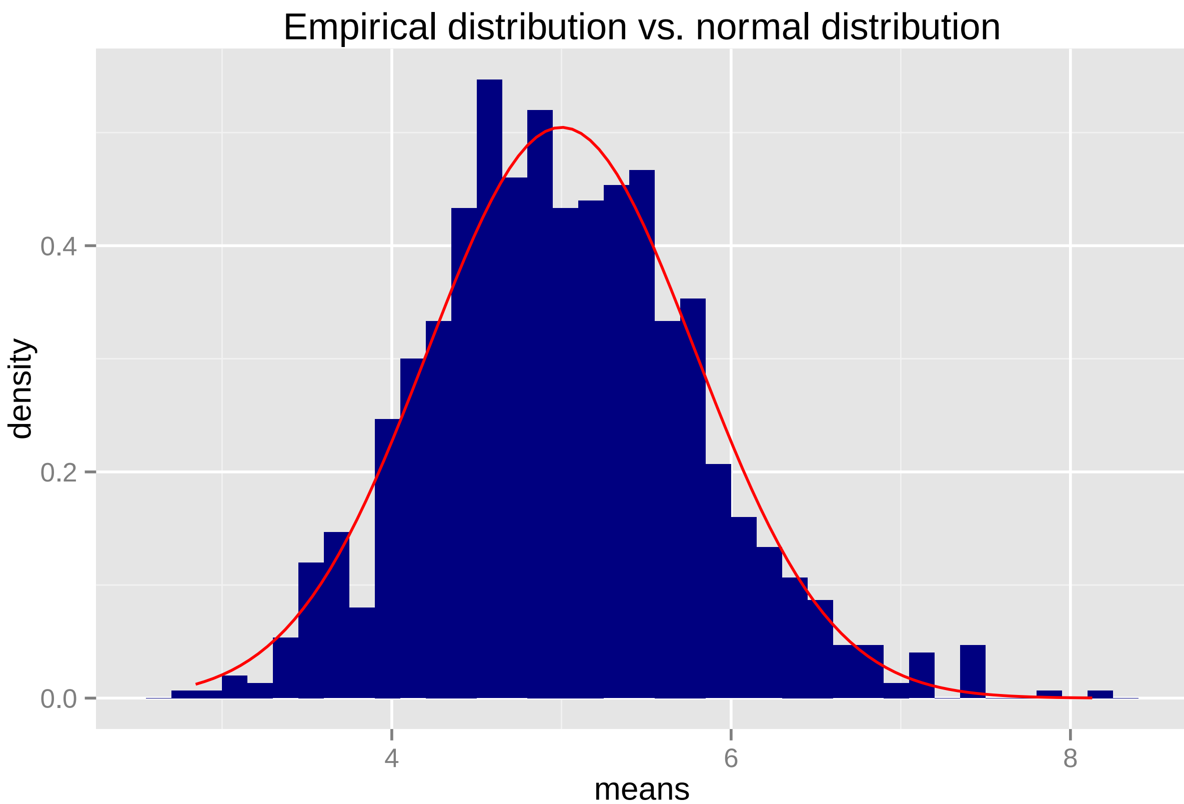


Sample mean (5.011563) is marked in red. Area marked in pink, represents the means with values within 1 standard deviation (its theoretical value) form the sample mean. Area in green - those between 1 and 2 standard deviations from the mean. Means, which fell more than 2 standard deviations from the mean are represented by the light blue area.

Distribution

In order to test the assumption of normality of the sample mean distribution, a histogram of empirical values will be compared with a normal distribution with a mean and variance equal theoretical mean and theoretical variance of the analyzed sampling distribution - $N(5, 0.625)$. Line representing normal distribution is shown in red.

```
m <- ggplot() + aes(x=means)
m <- m + geom_histogram(aes(y = ..density..), binwidth = .15, fill = "navy")
m <- m + stat_function(fun = dnorm,
                       args = list(mean = 1/lambda, sd = sqrt(theorVar)), colour = "red")
m + ggtitle("Empirical distribution vs. normal distribution")
```



Based on the visual comparison, it can be concluded that the empirical distribution is approximately normal.

Sources

1. LaTEX code generator