Exponential Distribution and CLT

Author: Joohyun Kwon

Synopsis

We will investigate the exponential distribution in R and compare it with the Central Limit Theorem. We run 1000 simulations on generating 40 random numbers from a exponential distribution with lamda value = 0.2. The mean and variance of these samples will be calculated and then compared to the theoretical value.

Simulation

We will run 1000 simulation. At each simulation, 40 random numbers will be generated from exponential distribution. These generated number will be saved as a row in a matrix, called sim. So the matrix will have 1000 rows and 40 colums

```
n <- 40
lambda <- 0.2
n.of.sim <- 1000
sim <- matrix(sapply(1:n.of.sim, function(i){rexp(n,lambda)}), n.of.sim,n)
dim(sim)
```

```
## [1] 1000   40
```

We calculate the means of each row and save the vector of means to variable means. Then the mean of the vector (sample mean) and the standard deviation (standard deviation of sample mean) will be calculated from means vector. The square value of the standard deviation is the variance of sample mean.

```
means <- apply(sim, 1, mean)

mean.of.sim = mean(means)
sd.of.sim = sd(means)
pop.sd = sd.of.sim * sqrt(n)

# sample mean
round(mean.of.sim,2)
```

```
## [1] 5.02
```

```
# variance of sample mean
round(sd.of.sim^2,2)
```

```
## [1] 0.67
```

$$\sigma^2 = S^2 * n$$

,where σ^2: population variance, S^2: variance of sample mean and n: sample size.

This sample mean (mean.of.sim) is the mean of mean values of simulated samples. The variance of population devided by sample size, n, is the variance of the sample mean. So the variance of population can be calculated by above equation.

Simulated Result

- **Sample mean: 5.02**
- **Populatin Variance: 26.7**

Theoretical Mean ($1/\lambda$) and Variance ($(1/\lambda)^2$)

- **Theoretical Mean: 5**
- **Theoretical Variance: 25**

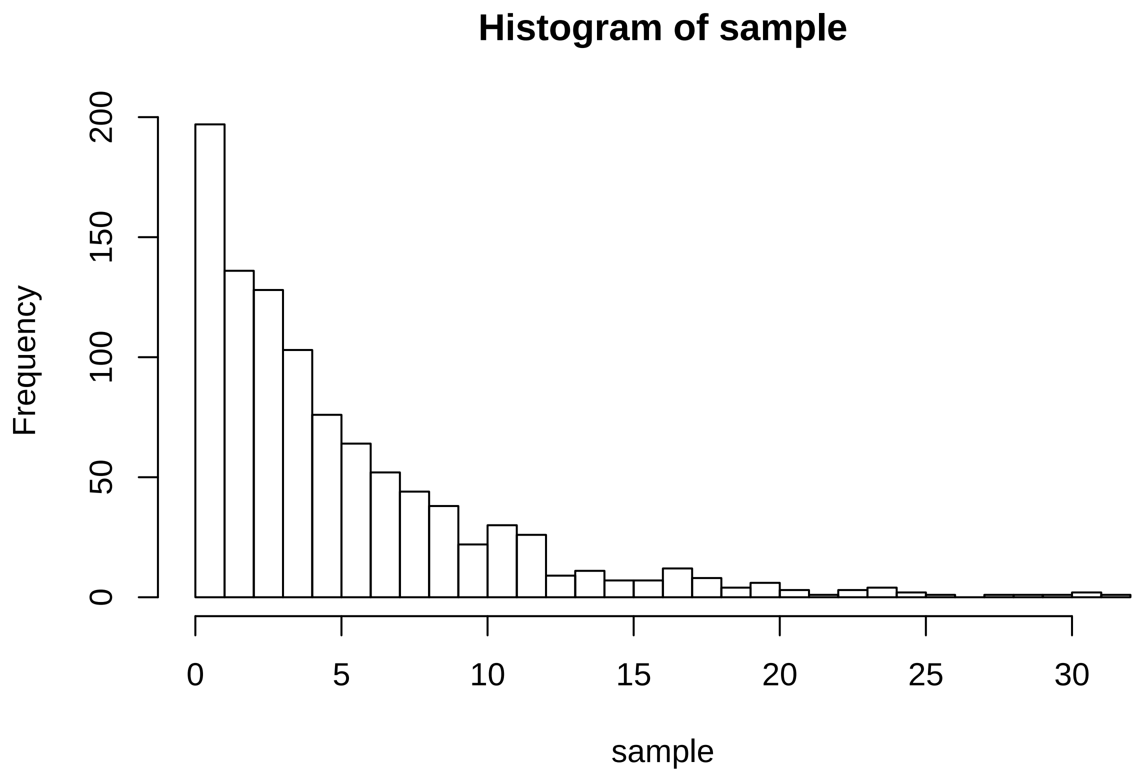
Conclusion:

Population mean and variance acquired by 1000 simulation (with 40 samplings) are pretty close to theoretical mean and variance.

Plot Comparison of Random Sample and Sample Mean Distributions

Histogram of Random Sample

```r
sample <- rexp(1000,0.2)
hist(sample, breaks=40)
```

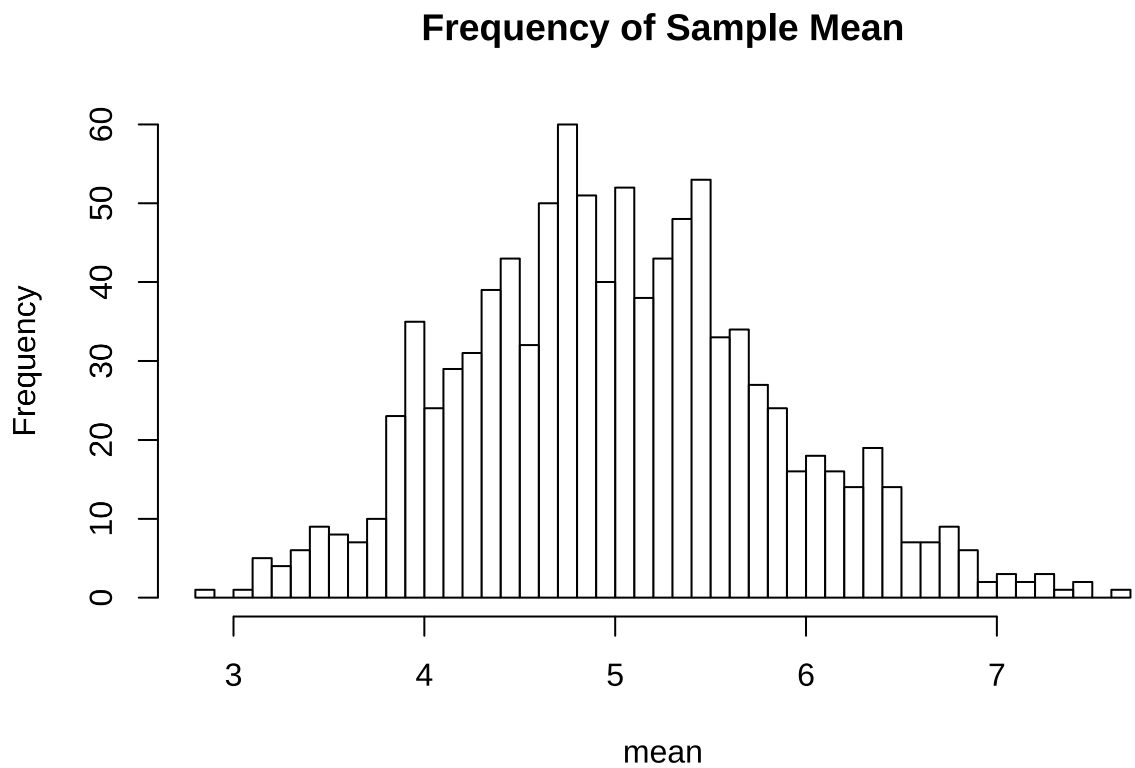


This distribution is not normal. It shows an exponential distribution.

However, the distribution of sample mean shows a normal distribution - it is approximately bell shaped and symmetrical along the middle vertical line.

Histogram of Sample Mean

```
hist(means, breaks = 40, xlab="mean", main="Frequency of Sample Mean")
```



Density Distribution of Sample Mean VS. Normal Distribution

The theoretical mean of normal distribution should be the population mean and the standard deviation should be square root of the theoretical variance devided by n.

$$S = \sigma/\sqrt(n)$$

We made a histrogram from the density of above sample mean. Then we overlaid on it a line plot of the density function with the theoretical mean and standard deviation.

The plot of theoretical density function is in red line.

```
df <- data.frame(means)
library(ggplot2)
g <- ggplot(df, aes(x=means))
g <- g + geom_histogram(aes(y=..density..), bins=40, col="black", fill="darkgreen", alpha=.3)
g <- g + geom_vline(xintercept = mean.of.sim, col = "blue", lwd=1)
g <- g + geom_text(aes(x=mean.of.sim,y=0, label=paste("Sample Mean: ", as.character(round(mean.of.sim,2
                   angle=90, col="darkred", vjust = -.3, hjust=-0.1 )
g <- g + xlab("mean")

#Plot Normal Distribution with theoretical mean an sd
g <- g + stat_function(fun = dnorm,
                       args = list(mean=1/lambda, sd=1/(lambda*sqrt(n))), col="red", lwd=1)
```

```r
g <- g + ggtitle("Density - Simulated vs Theoretical")
g
```