StatsProject

George Liu

August 18, 2015

Overview

In this report, exponential distribution sampling simulation will be used to explore the relationship between sample mean and population mean. Further, the result will be used to test against the Central Limit Theorem.

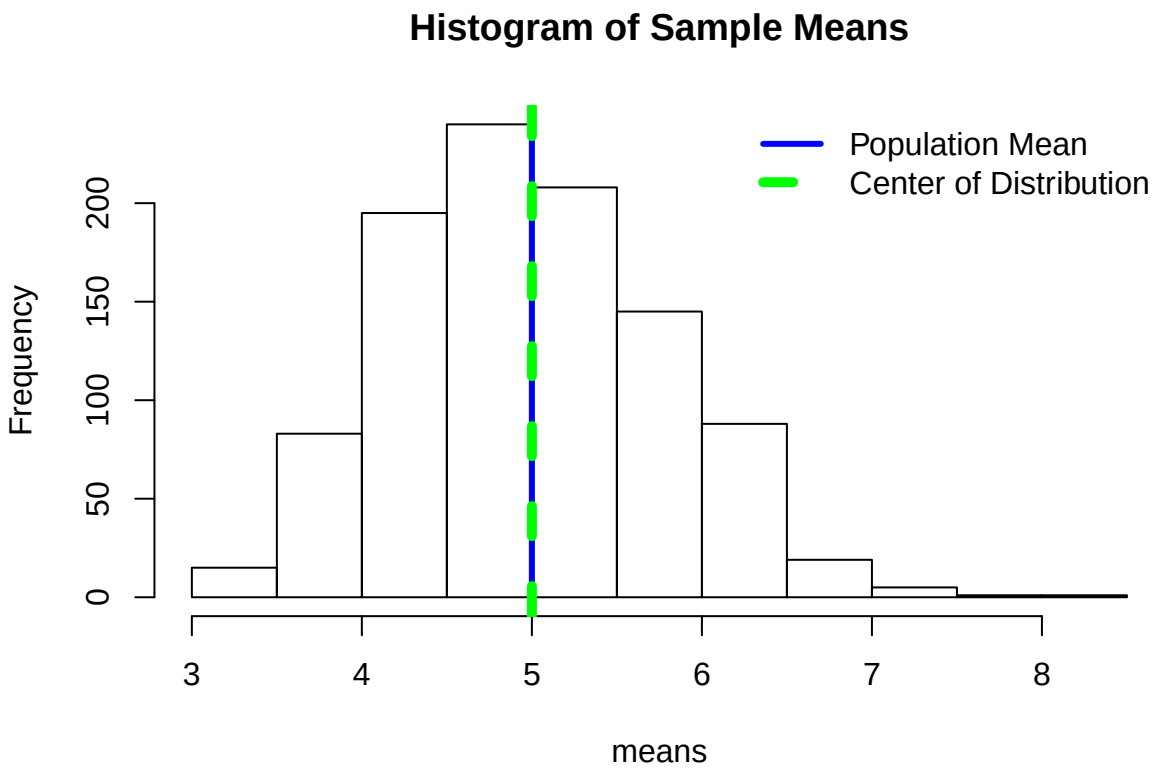
Simulations

```
# set all the parameters
lambda <- 0.2
mu <- 1 / lambda
sigma <- 1 / lambda
sample.size <- 40
sample.times <- 1000
# generate 40k random values that form the population, store them in a 1000 * 40 matrix "pop.matrix"
set.seed(1)
pop.matrix <- matrix(rexp(sample.size * sample.times, lambda), 1000, 40)
```

Sample Mean versus Theoretical Mean

In the plot below, sample mean's distribution is shown with the theoretical mean (population mean) highlighted in blue. As we can see, the the sample mean is centered at the theorectical mean of 5 which is mu (1/lambda = 1/.2 = 5). This validates the LLN and CLT.

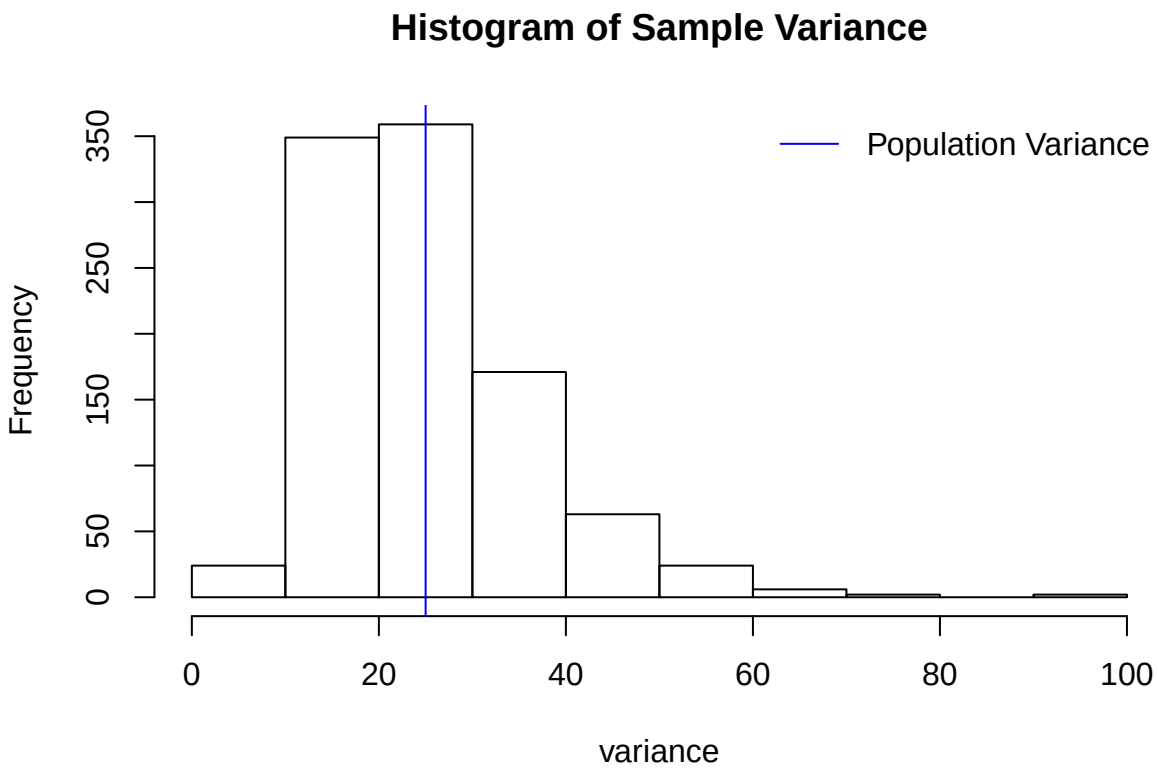
```
#calculate the sample means and plot
means <- apply(pop.matrix, 1, mean)
hist(means, main = "Histogram of Sample Means")
abline(v = mu, col = "blue", lwd = 3)
abline(v = mu, col = "green",lty = 2, lwd = 5)
legend("topright", lty = c("solid", "dashed"), legend = c("Population Mean", "Center of Distribution"),
```



Sample Variance versus Theoretical Variance

The figure below shows the sample variance distribution. Also, the theorectical variance (population variance) is highlighted in blue. The distribtion skews to the right, while the center of gravity still lies at the population variance.

```
#calculate the sample variance and plot
variance <- apply(pop.matrix, 1, var)
pop.var <- sigma ^ 2
hist(variance, main = "Histogram of Sample Variance")
abline(v = pop.var, col = "blue")
legend("topright", lty = "solid", legend = "Population Variance", col = "blue", bty = "n")
```

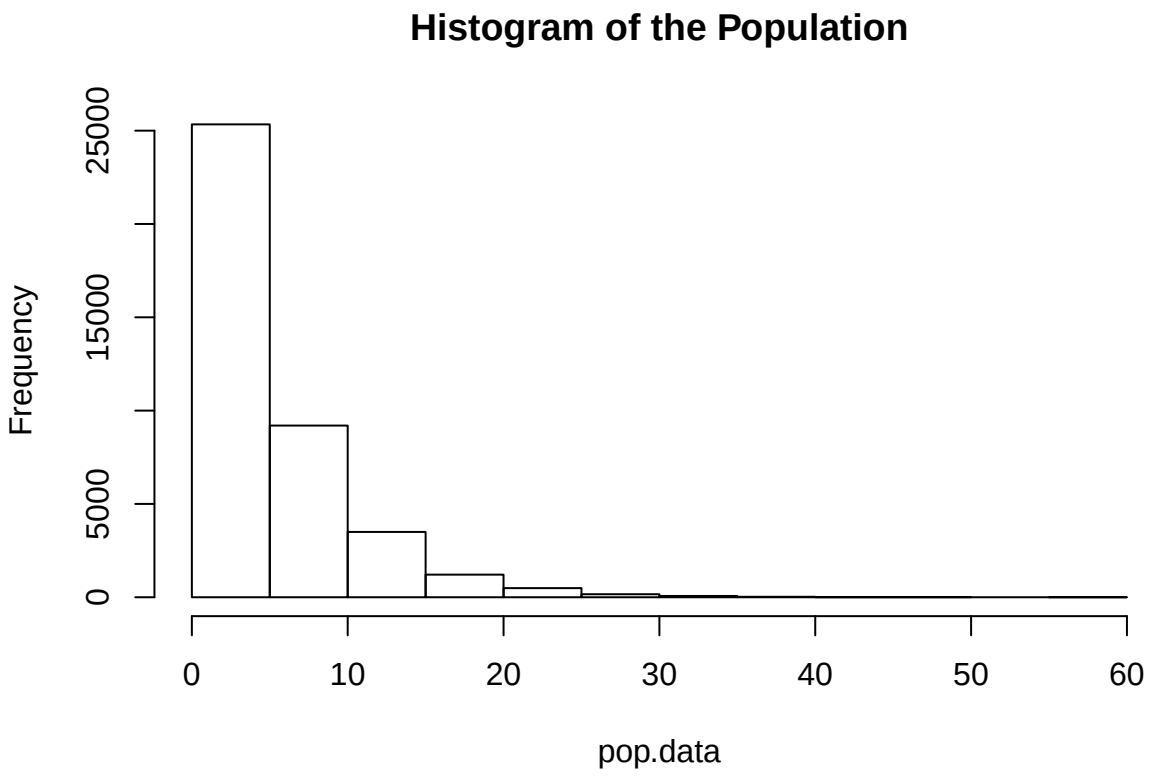


Distribution

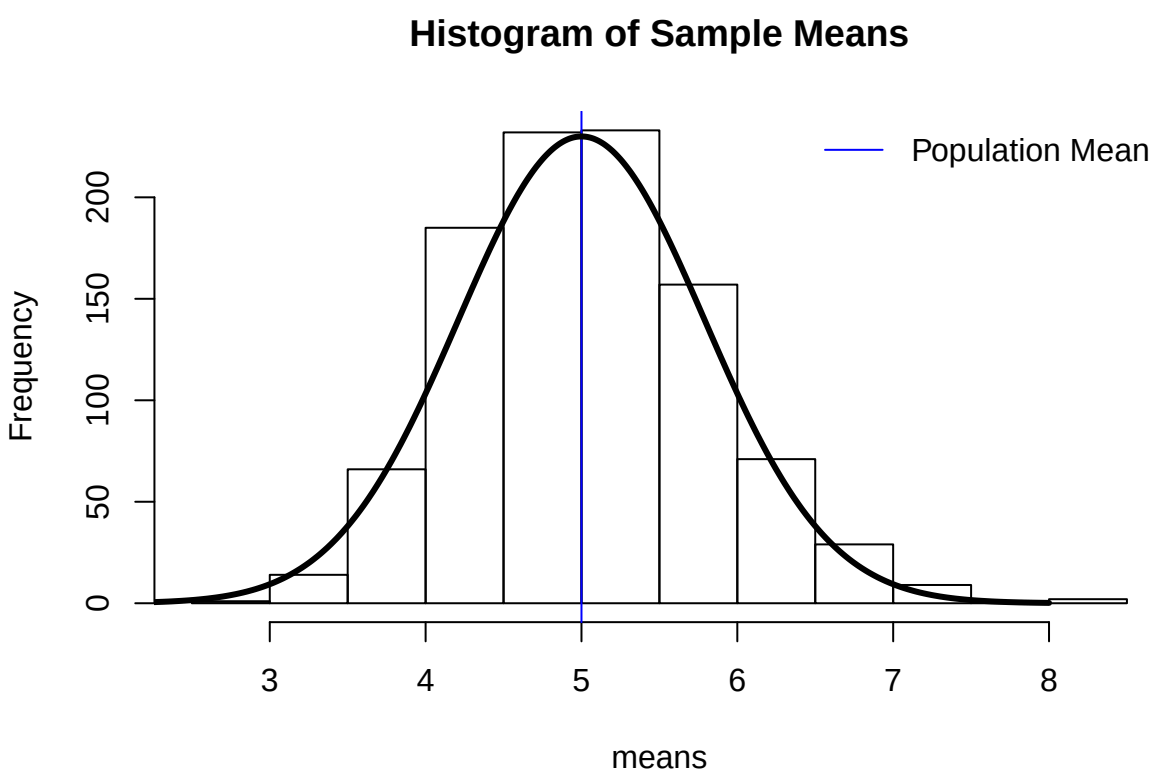
In the following plot, first, the population distribution is plotted, then the sample means' distribution is shown.

The population distribution is heavily skewed to the right. However,the sample means' distribution centered at the population mean and resembles a bell curve. On top of the sample mean distribution, a normal curve is plotted. Clearly the distribution of the sample means fit a normal distribution. This is also what the CLT tells us.

```
# grab the original generated population data and plot its distribution
set.seed(1)
pop.data <- rexp(sample.size * sample.times, lambda)
hist(pop.data, main = "Histogram of the Population")
```




```r
# plot again the sample mean distribution
pop.matrix <- matrix(rexp(sample.size * sample.times, lambda), 1000, 40)
means <- apply(pop.matrix, 1, mean)
hist(means, main = "Histogram of Sample Means")
abline(v = mu, col = "blue")
legend("topright", lty = "solid", legend = "Population Mean", col = "blue", bty = "n")
# add a normal curve for comparison
x <- seq(2, 8, length.out = 1000)
k <- dnorm(5, mean = 5, sd = sigma/sqrt(40))
y <- dnorm(x, mean = 5, sd = sigma/sqrt(sample.size)) * 230/k
points(x, y, type = "l", lwd = 3)
```


Histogram of Sample Means
Population Mean
Frequency
0
50
100
150
200
3
4
5
6
7
8
means