Part 1: Simulation Exercise

John B Cheadle

August 9, 2017

Overview

This report explores the exponential distribution in R, denoted by the function rexp(n, lambda) where lambda is the rate parameter and n is the number of observations, and compares it with the **Central Limit Theorem (CLT)**. Here we will compare the theoretical mean and variance of the exponential distribution to the simulated (sample) mean and variance. Finally, we'll comment on the distribution of the population and determine whether it is approximately normal.

Simulations

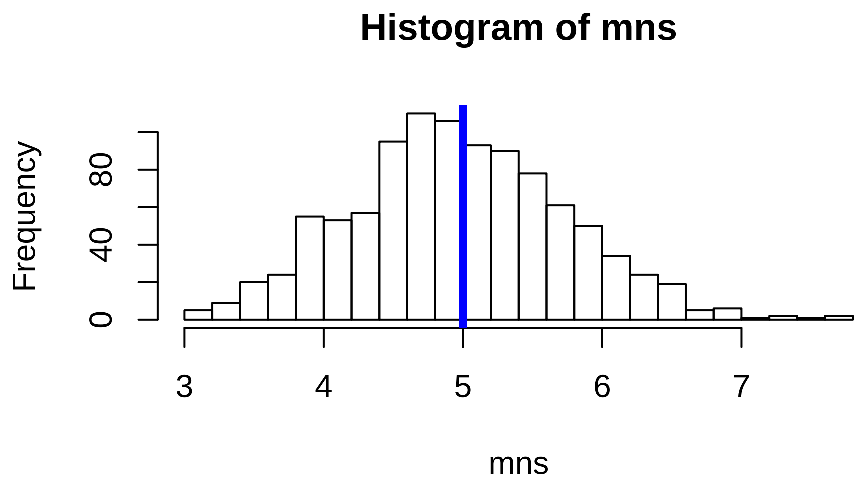
Here we explore the distribution of the mean and variance of 40 exponentials in 1000 simulations, where lambda (the rate parameter) is 0.2

```
lambda = 0.2
n = 40
sims = 1000

## Distribution of the mean of 40 exponentials in 1000 simulations
mns = NULL
for (i in 1:1000) {
    mns = c(mns, mean(rexp(n,lambda)))
}

## Distribution of the variance of 40 exponentials in 1000 simulations
vars = NULL
for (i in 1:1000) {
    vars = c(vars, var(rexp(n,lambda)))
}
```

```
## Plotting histogram with a line at the population mean
hist(mns,breaks = 20)
abline(v = 5,lwd=4, col = "blue")
```



Sample Mean vs. Theoretical Mean

Here the rate parameter lambda, the rate parameter, as 0.2. The theoretical mean for the exponential distribution is 1/lambda, or $1/0.2 = \mathbf{5.0}$.

We previously performed 1000 simulations of the mean of 40 exponentials, and stored these in the variable mns.

```
## Theoretical Mean
1/lambda
```

```
## [1] 5
```

```
## Mean of the averages of exponential distribution
mean(mns)
```

```
## [1] 4.968031
```

As described in the CLT, we see that the mean of the averages of the simulated exponential distributions is centered around the theoretical mean of that distribution with our large sample size.

Sample Variance vs. Theoretical Variance

Similar to the means, with 1000 simulations of the variance of 40 exponentials, our sample variance is going to be concentrated around our population variance, defined as sd^2 or $(1/\text{lambda})\hat{}2 = \mathbf{25.0}$

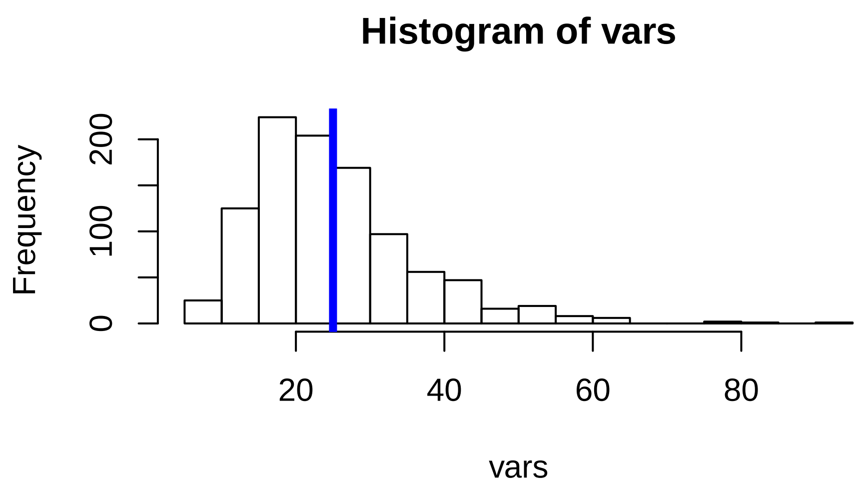
```
## Population Variance
sd = 1/lambda
pop_var = sd^2
pop_var
```

```
## [1] 25
```

```
## Sample Variance (sd^2) of the averages of exponential distribution
mean(vars)
```

```
## [1] 25.06907
```

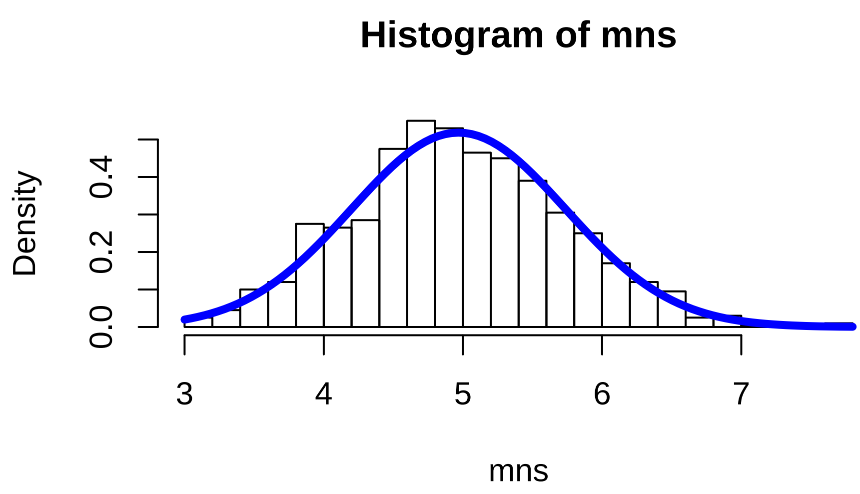
```r
## Plotting histogram with a line at the population variance
hist(vars, breaks=15)
abline(v = 25,lwd=4, col = "blue")
```



Distribution

The distribution of the simulated averages of the exponential are approximately normal. To show this, we overlay a normal curve using the dnorm() function (note that we have to change our y-axis from 'frequency' to 'density' in order to appropriately scale the y-axis).

```r
hist(mns, prob=TRUE, breaks=20)
curve(dnorm(x, mean=mean(mns), sd=sd(mns)), add=TRUE, col="blue", lwd=4)
```



The histogram and normal distribution are nearly congruous, showing clearly that our simulated exponential distribution is Gaussian in nature.