Central Limit Theorem Simulations

Alim Ray

9/26/2015

Overview

This report explores the Cental Limit Theorem by running simulations using the exponential distribution. The sample means of the simulations are plotted and shown to fix the normal distribution.

Simulation

Below is code that runs the exponential distribution 1,000 times. Cumulative means for the 1,00 simulations are plotted and shown to converge to the theorectical mean 1/where = 0.2.

```
library(ggplot2)

simulations <- function( nosim = 1000 ) {
  n <- 40
    lambda <- 0.2
    mns <- c()
  vrs <- c()
    for (i in 1 : nosim) {
        sim <- rexp(n, lambda)
        mns <- c( mns, mean(sim) )
        vrs <- c( vrs, var(sim) )
  }
  list( mns, vrs )
}

simulationGraphs <- function(nosim = 1000) {
  simulationValues <- simulations( nosim )
  mns <- simulationValues[[1]]
  vrs <- simulationValues[[2]]
    #Means plot
    plotdata <- cumsum(mns)/1:nosim
    meansGraph <- ggplot( data.frame( x = 1:nosim, y = plotdata), aes (x = x, y = 
y) ) + geom_line() + geom_hline(yintercept = 5) + ggtitle( "Cumulative Means Simul
ations")

    #Variance plot
  varGraph <- ggplot( data.frame( v = vrs), aes ( x = v ) ) + geom_histogram() + g
eom_vline(xintercept = 25, aes( colour = "red")) + ggtitle( "Variance Histogram")

    list( meansGraph, varGraph, mns, vrs )
}
```

Sample mean vs. Theoretical mean

The Central Limit Theorem (CLT) states that sample means follow a normal distribution. Therefore, since we know the exponential distribution has a mean of

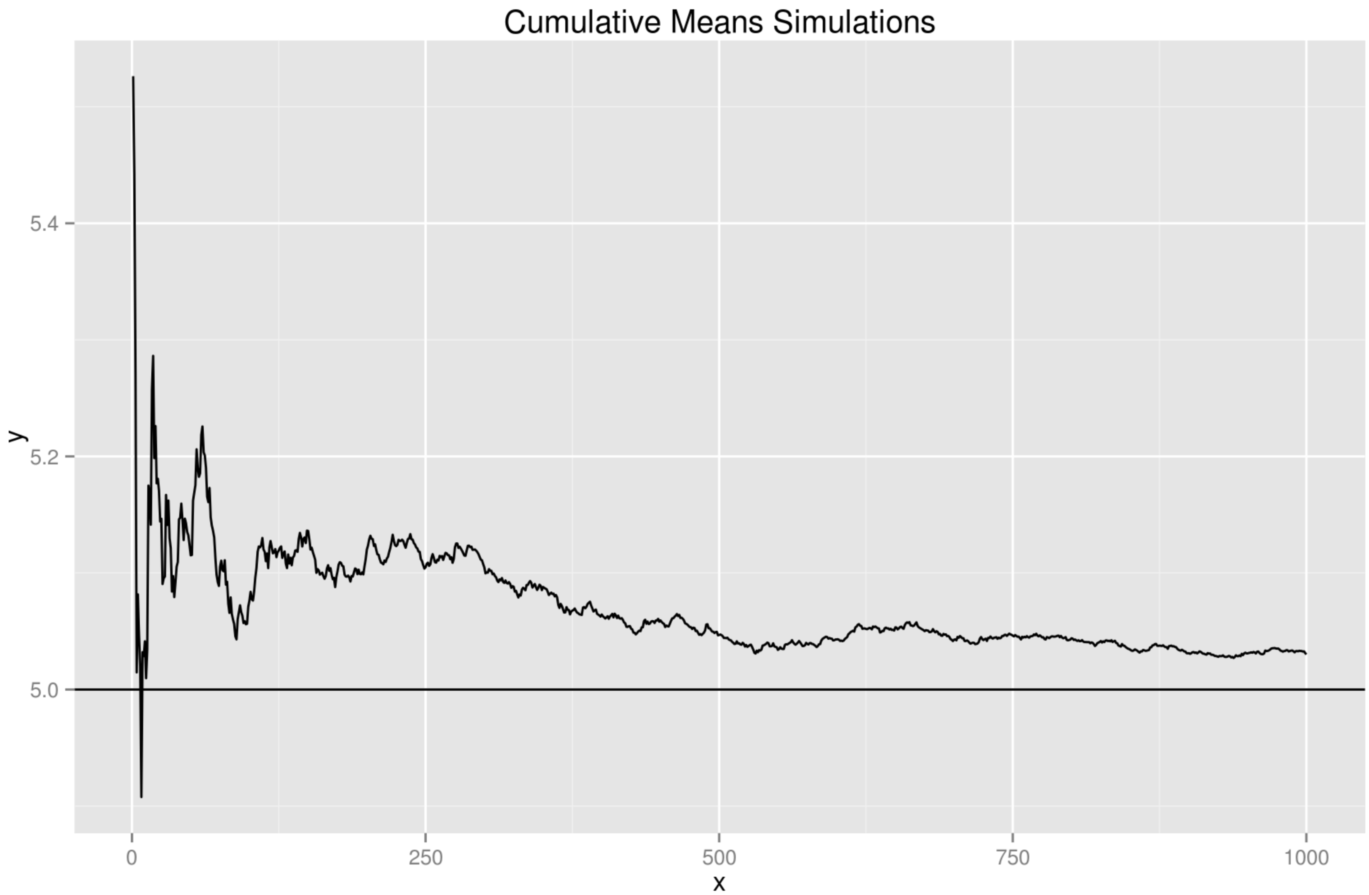
$$\frac{1}{\lambda}$$

. The theoretical mean for this simulation is

$$\frac{1}{0.2}$$

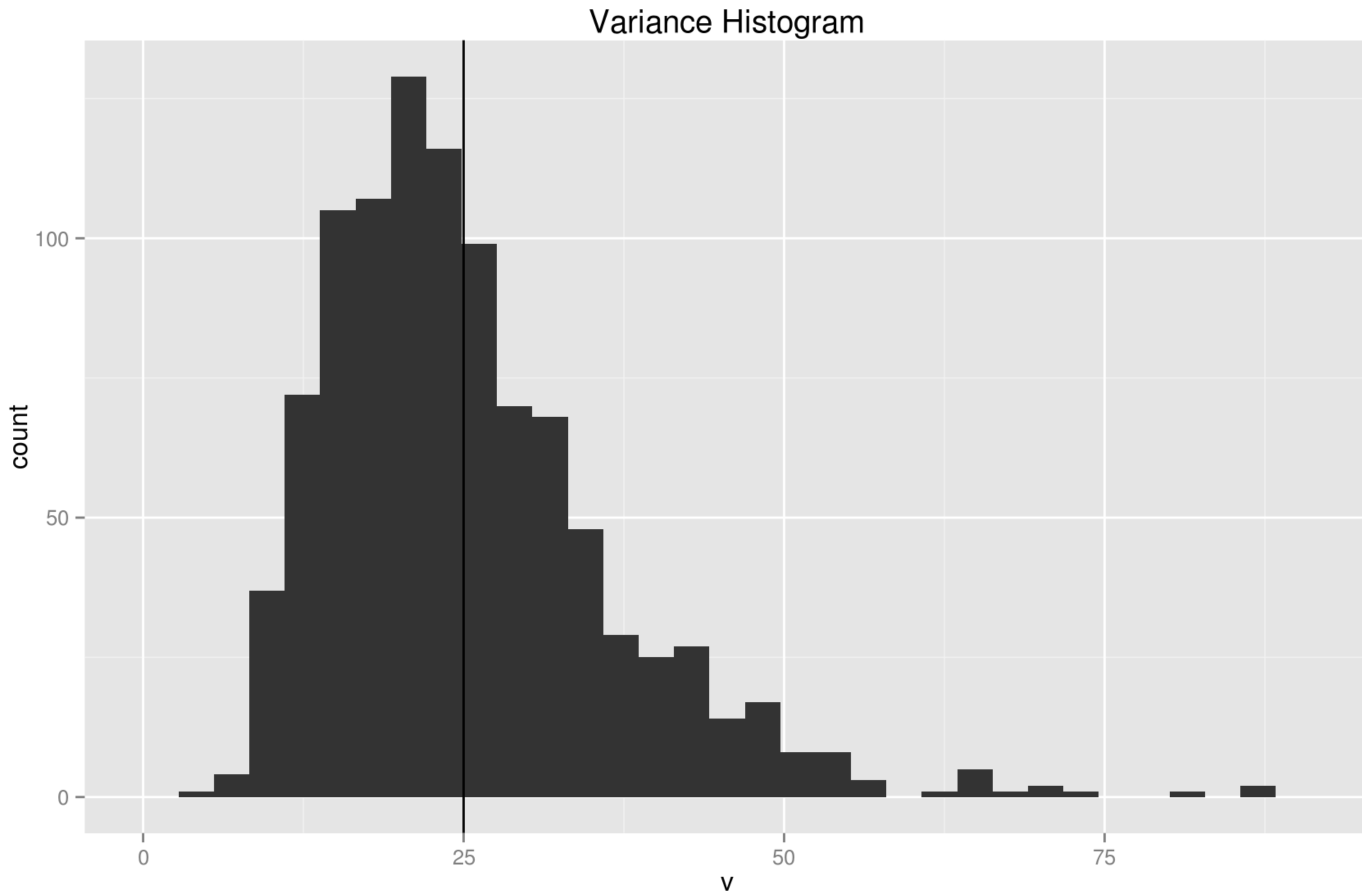
or 5. There is a horizontal line at 5 on the simulation graph representing the theoretical mean. The graph shows, as we increase the number of simulations, the closer we get to the theoretical mean.

Below is the graph returned by the code above.



Sample variance vs. Theoretical variance

The theoretical standard deviation is also 1/lambda so the theorectical variance (1/0.2)^2 or 25. The variance histogram (below) of the 1000 simulations is centered around 25.



Distritbution

The CLT states that the formula

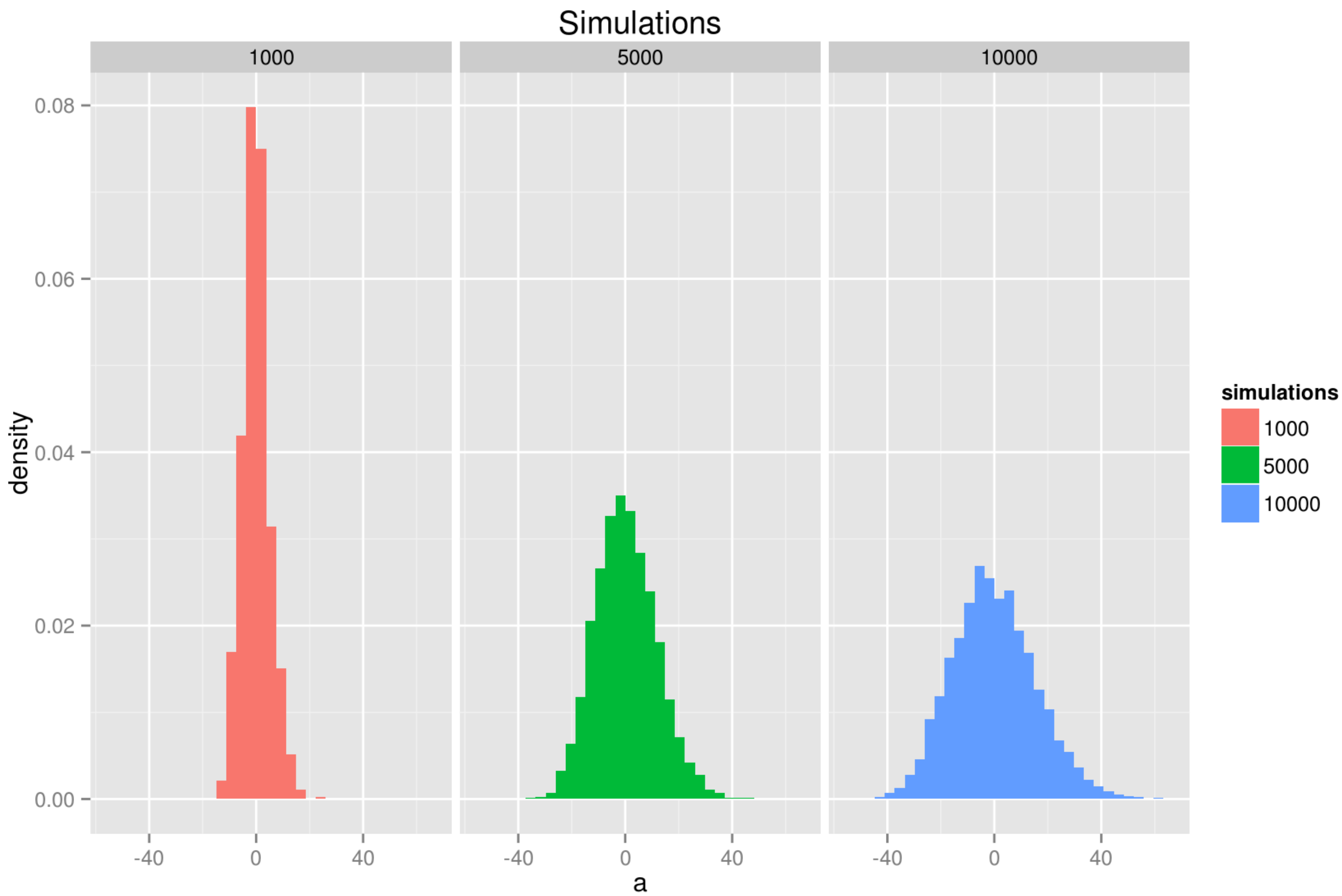
$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

is normal. Using 5 as the theoretical mean and 5 as the theoretical standard deviation, we will run n simulations, subtract 5 and divide by 5/.

We will do this for 1,000, 5,000, and 10,000 simulations and plot the results. Each graph fits a normal distribution.

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## function(x, n) { (x -5) * sqrt(n) / 5}
```



Conclusion

The graphs in this report demonstrate the CLT applies to the sample means for the exponental distribution.