A comparison of the exponential distribution and the Central Limit Theorem

Izaak Jephson

20/01/2020

Overview

This report investgates the properties of the exponential distribution by comparing the distribuiton of the means and variances of 1000 simulated sets of 40 exponentially distributed random variables to the theoretical mean and variance of the distribution. As expected from the Central Limit Theorem, the means of these samples are normally distributed around the theoretical mean of the distribution.

Simulation

```
# Simulate 1000 sets of 40 exponential random variables.

set.seed(56)
lambda <- 0.2
nexp <- 40
nsim <- 1000
sim <- replicate(n = nsim, expr = rexp(nexp, lambda), simplify = FALSE)

# Take mean and variance of each of these simulated sets.

means <- sapply(sim, mean)
vars <- sapply(sim, var)

simsum <- tibble(mean = means, var = vars)

# Take mean and variance of sample means.

mean <- mean(means)
variance <-  var(means)
```

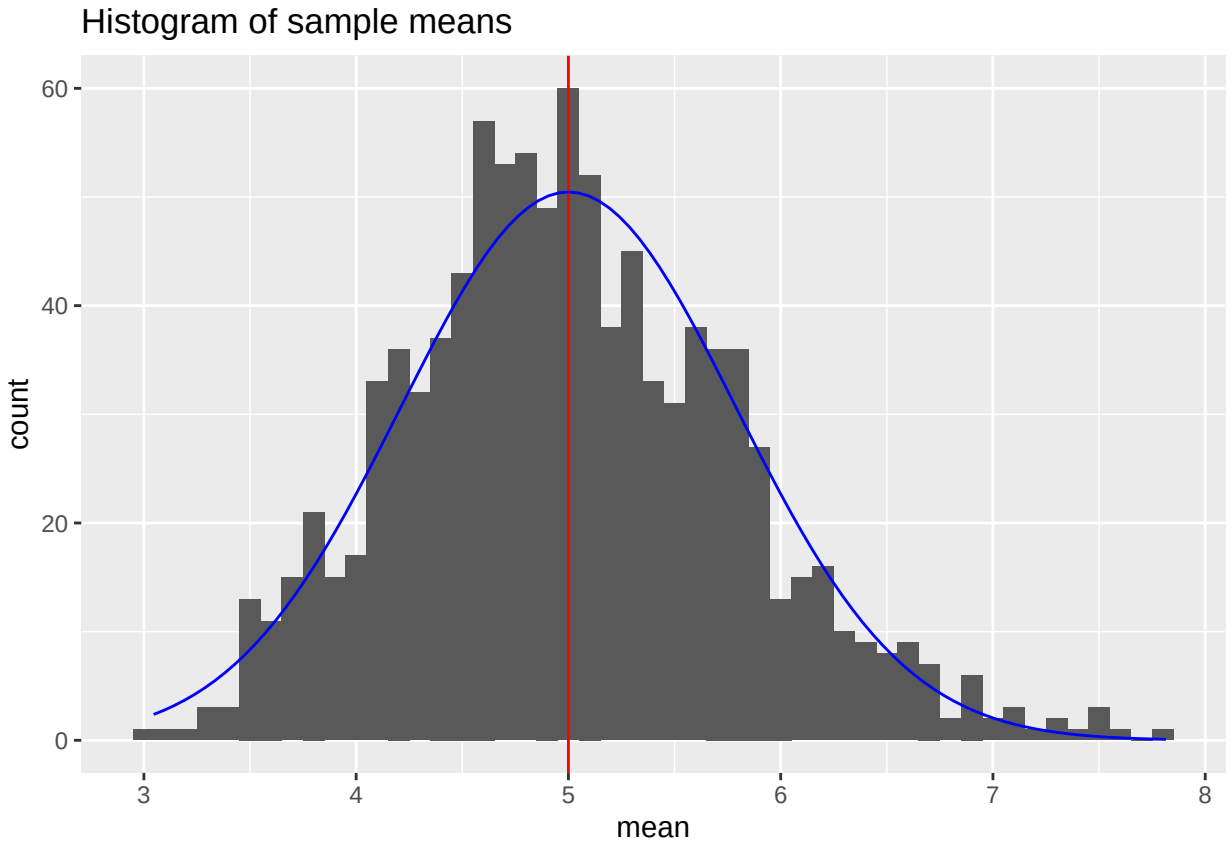
1000 sets of 40 exponential random variables were simulated. These had $\lambda = 0.2$. The means and variances of each of these sets were taken. The mean and variance of the sample means was aso taken.

Sample mean vs theoretical mean

```
# Plot histogram of sample means and overlay normal distribution.

bw <- 0.1

g <- ggplot(data = simsum) + geom_histogram(mapping = aes(mean), binwidth = bw) +
        geom_vline(aes(xintercept = 1/lambda), colour = "red") +
        stat_function(fun = function(x)
                dnorm(x, mean = 1/lambda, sd = (1/lambda)/sqrt(nexp)) * nsim * bw,
                colour = "blue") +
        labs(title = "Histogram of sample means")
g
```



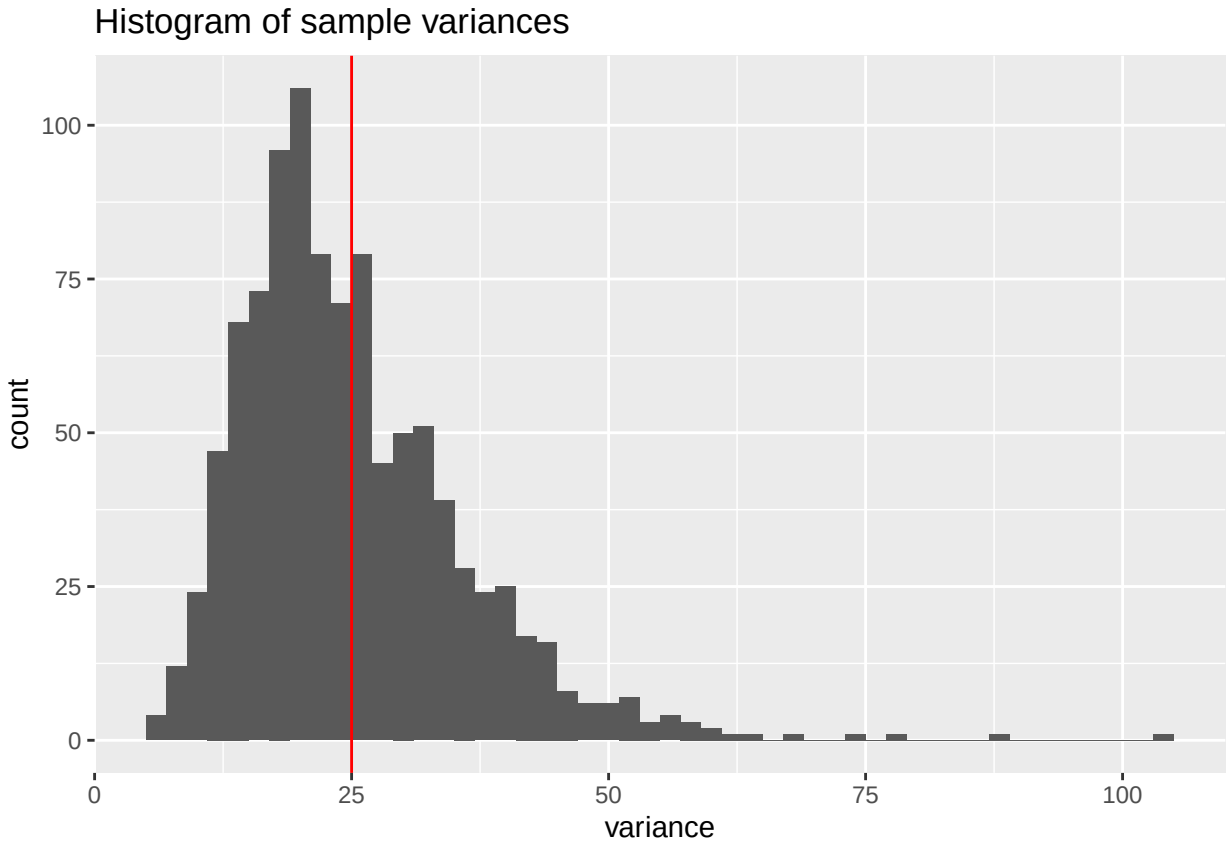
It can be seen from the plot above that the sample means are distributed about the theoretical mean of the exponential distribution which is plotted in red on the graph above. A plot of the normal distribution with the theoretical mean of the exponential distribution and variance predicted by the CLT is overlaid in blue for comparison. It can be seen that the distribution of the sample means ifs approximately normal as prediced by the CLT. The mean of the sample means is 5.0094 which is approximately the theoretical mean of the distribution. The variance of the sample means is 0.6246 which is approximately the variance predicted by the CLT of $\sigma^2/n = 0.625$. As shown by the equation above, as sample size increases, we would expect the variance of the means to reduce.

Sample variance vs theoretical variance

```
# Plot histogram of sample variance.

varbw <- 2

g <- ggplot(data = simsum) + geom_histogram(mapping = aes(var), binwidth = varbw) +
        geom_vline(aes(xintercept = 1/lambda^2), colour = "red")+
        labs(title = "Histogram of sample variances", x = "variance")
g
```



The sample variance is approximately centred around the theoretical variance of the distribution which is $1/\lambda^2 = 25$. This is as expected. As the sample size increases, we would expect this to converge to the theoretical variance of the ditribution.