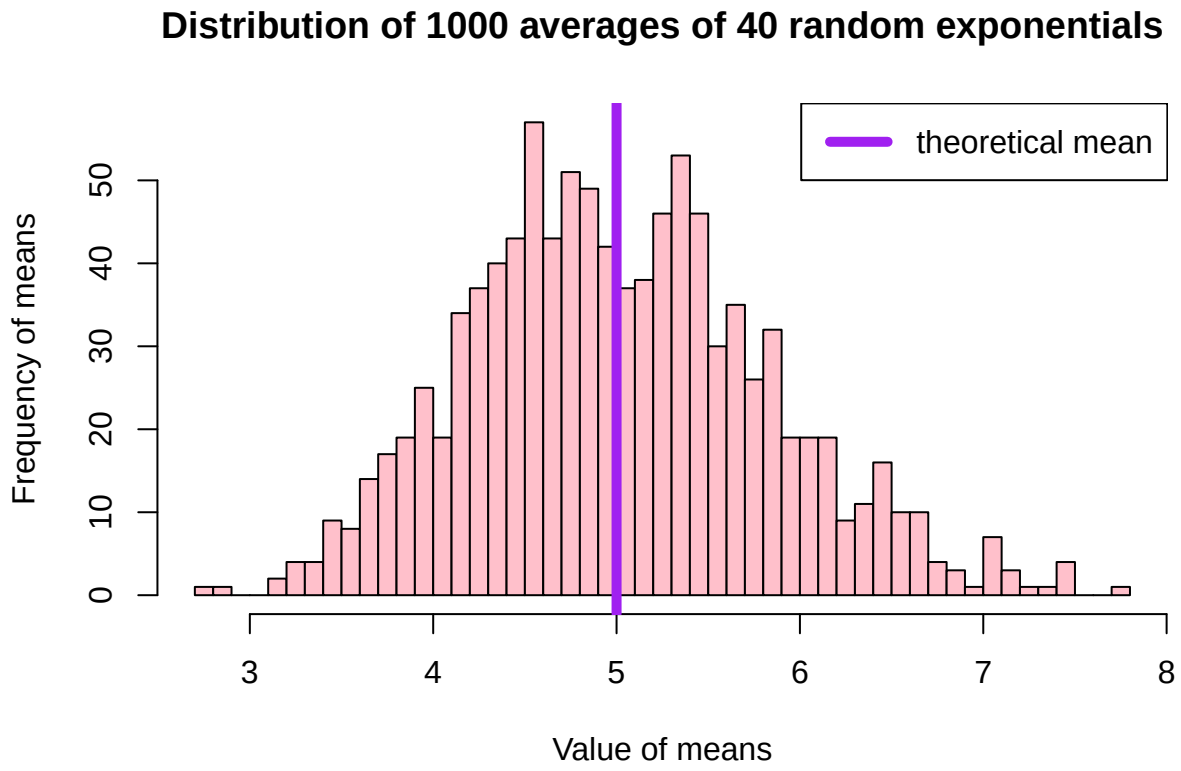
Statistical Inference Week 4 Project

Joaquin Arango

12/23/2020

Question 1: Show the sample mean and compare it to the theoretical mean of the distribution.

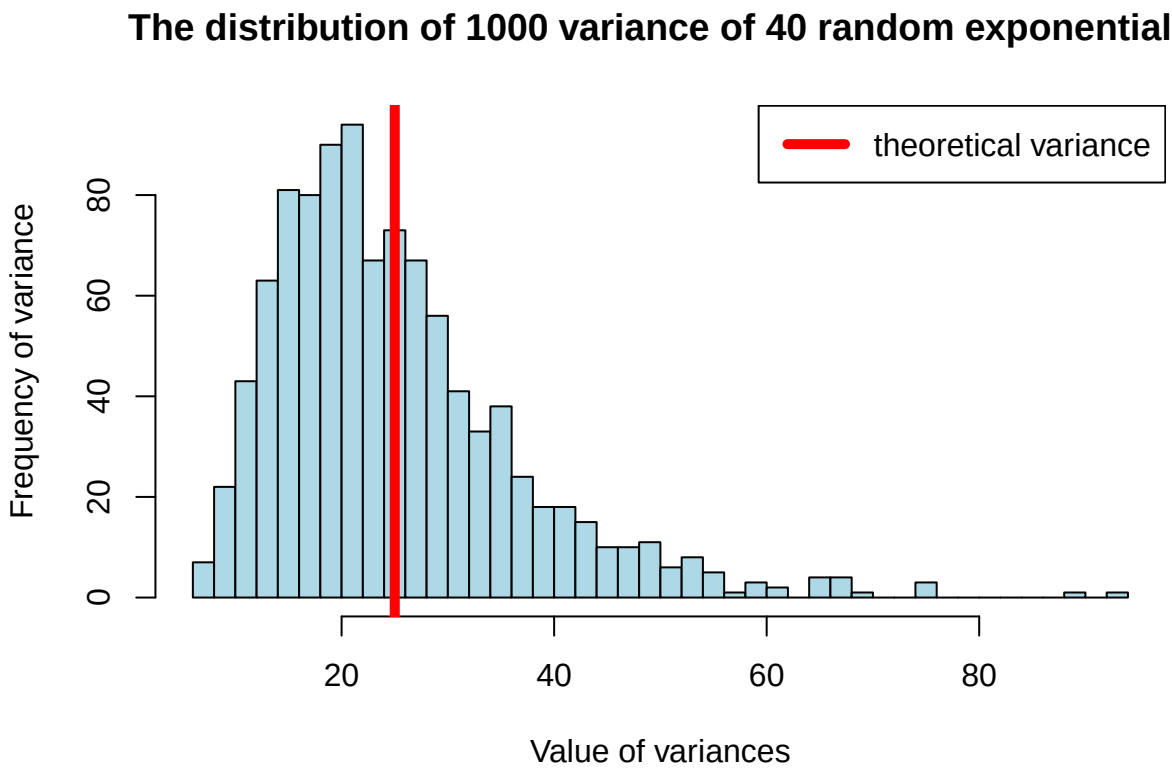
```
knitr::opts_chunk$set(echo = TRUE)
lambda <- 0.2
sim_Data <- matrix(rexp(1000*40, lambda), nrow = 1000, ncol = 40)
distMean <- apply(sim_Data, 1, mean)
hist(distMean, breaks = 50,
     main = "Distribution of 1000 averages of 40 random exponentials", xlab = "Value of means",
     ylab = "Frequency of means", col = "pink")
abline(v = 1/lambda, lty = 1, lwd = 5, col = "purple")
legend("topright", lty = 1, lwd = 5, col = "purple", legend = "theoretical mean")
```



The simulated sample means are normally distributed with a center very close to the theoretical mean.

Question 2: Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

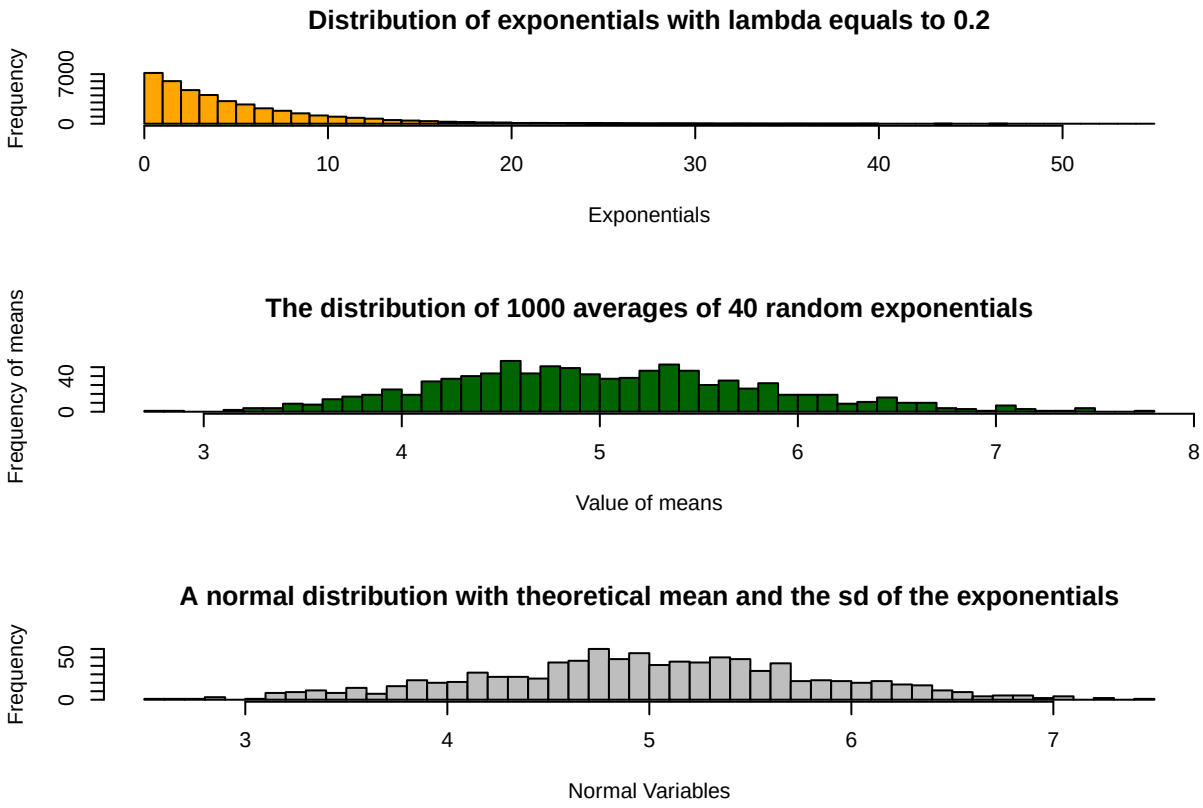
```r
distVar <- apply(sim_Data, 1, var)
hist(distVar, breaks = 50, main = "The distribution of 1000 variance of 40 random exponentials",
     xlab = "Value of variances", ylab = "Frequency of variance", col="light blue")
abline(v = (1/lambda)^2, lty = 1, lwd = 5, col = "red")
legend("topright", lty = 1, lwd = 5, col = "red", legend = "theoretical variance")
```



The simulated sample variances are almost normally distributed with a center near the theoretical variance

Question 3: Show that distribution is approximately normal.

```r
par(mfrow = c(3, 1))
hist(sim_Data, breaks = 50, main = "Distribution of exponentials with lambda equals to 0.2", xlab ="Exp
hist(distMean, breaks = 50, main = "The distribution of 1000 averages of 40 random exponentials", xlab =
simNorm <- rnorm(1000, mean = mean(distMean), sd = sd(distMean))
hist(simNorm, breaks = 50, main="A normal distribution with theoretical mean and the sd of the exponent
```



The first histogram is the distribution of the exponential with lambda equals to 0.2. The second histogram is the distribution of 1000 averages of 40 random exponential. The third histogram is a normal distribution with mean and standard deviation equal to the second histogram's.

Comparing the first with the second histogram, we can see the distribution becomes normal as the means were taken from each groups, just like the central limit theorem. Comparing the second and the third histogram, we see the distribution of the means is close to a normal distribution with the same mean and standard deviation.