A study of sample mean and variance statistics for the exponential distribution as convergent estimates for population parameters using R simulation

Jean-Remy Behaeghel

September 16, 2016

Overview

In this report we will draw 1000 samples of 40 exponentials. We will evaluate the sample mean and sample variance of those sample against the theoretical mean and variance. Lastly we will compare the distribution of the samples and variance means against a normal distribution and attempt to verify the central limit theorem. We assume iid random vairables and that the sample mean and sample variance of iid random variables are consistent;they converge to the population mean and population variance.

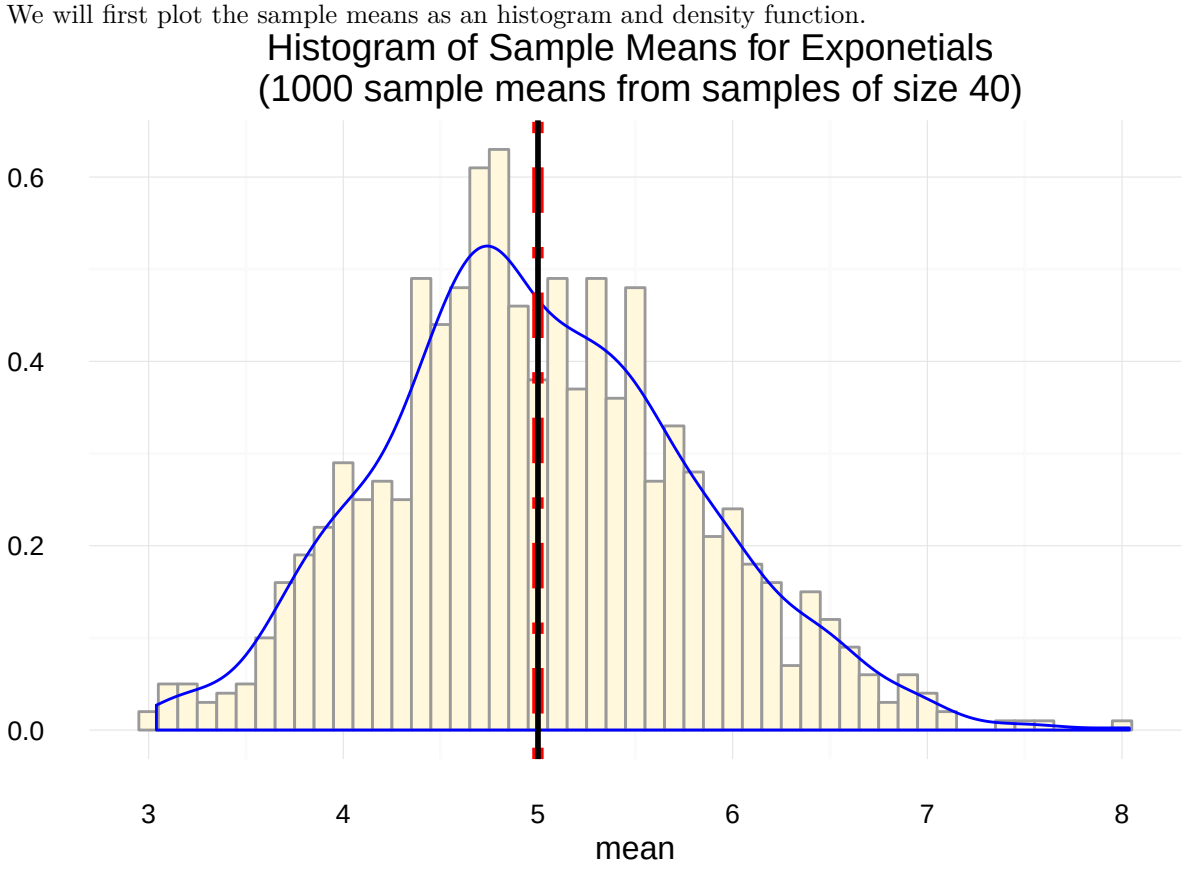
Simulation

The simulation code is straight forward, we will draw 40 exponentials using `rexp` with a parameter $\lambda = 0.2$ and calculate its mean, repeat the draw and sample mean calculation a 1,000 times. Each calculated mean will be stored in a result vector. We can then use the result vector for the remainder of the exercise. The simulation vectore will contains 1000 samples of size 40 of exponentials.

```
set.seed(100)  # So we can repeat the experience with the same reuslts
lambda <- 0.2  #Parameter for the exponential distribution
n <- 40  #sample size
nsims <- 1000  # number of simulations to run
mns <- NULL  # initiating variables
vars <- NULL  # initiating variables
for (i in 1:nsims) {
    sam <- rexp(n, lambda)
    mns <- c(mns, mean(sam))
    vars <- c(vars, var(sam))
}  #running the simulation
dfsim <- data.frame(mns, vars)  #create a data frame with two columns,
# one for the means and one for the vars
colnames(dfsim) <- c("mns", "vars")
```

Sample mean and theoretical mean

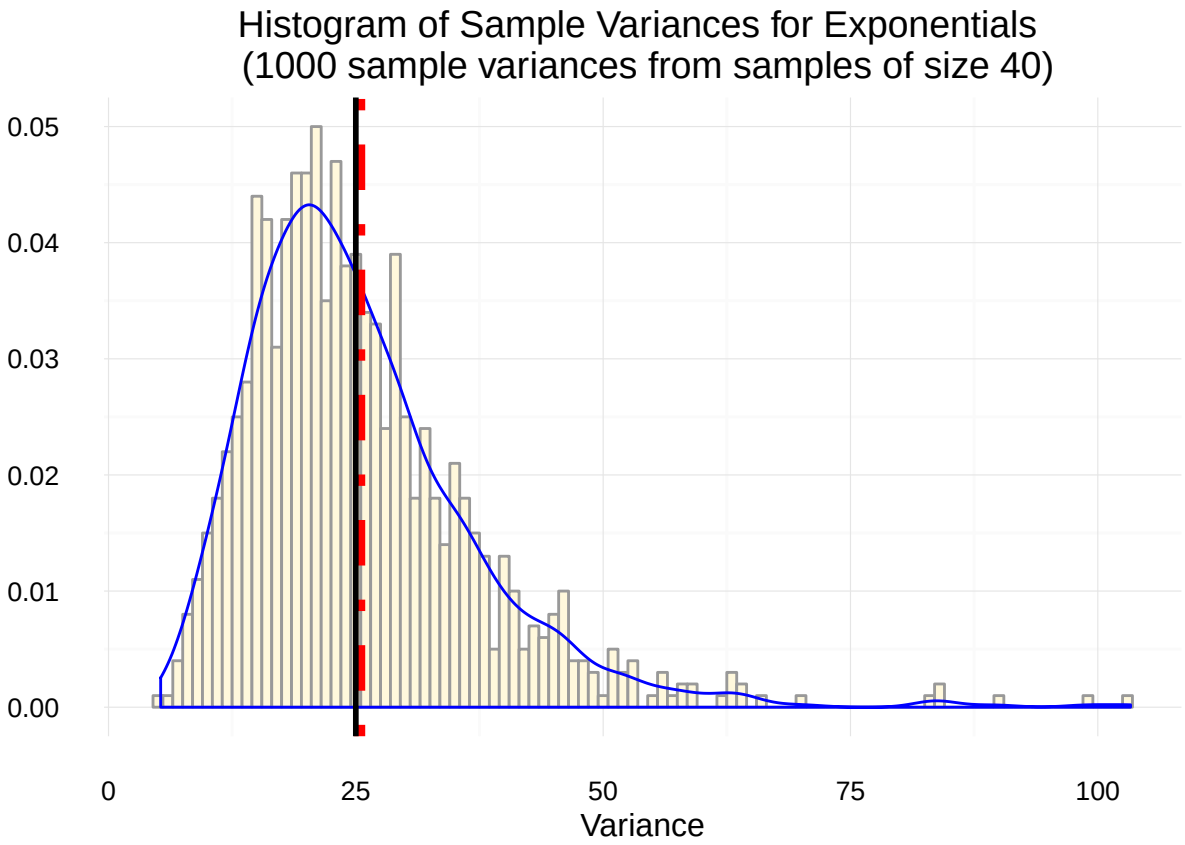
We will first plot the sample means as an histogram and density function.



The dotted dash line represents the mean of the sample means and the solid black line represent the theorectical mean $1/\lambda$. In the graph they apperar very close to each other. This is confirmed by the numerical values: sample mean is 4.9997019 where the therorical mean is 5. This result is supporting the law of large numbers conclusions. The sample mean is a random variable which expected value (mean) converge to the mean of the population.

The density plot overlaying the histogram tends to take a symetric bell shape curve centered around the mean of the sample means. This would indicate a normal distribution of the sample mean random variable as predicted by the center limit theorem.

Sample variance and theoretical variance



The dotted dash line represents the mean of the sample variances and the solid black line represent the theorectical variance $1/\lambda$. In the graph they apperar very close to each other. This is confirmed by the numerical values: sample variance is 25.3728703 where the therorical variance is 25. This result is supporting the law of large numbers conclusions. The variance is a random variable which expected value (mean) converge to the value it is estimating, in this case the variance.

The density plot overlaying the histogram for the variancemay eventually converge to a bell shape curve but with the limited number of simulations we have ran, it's still has a pretty long right tail. However, the density is centered around the mean and seems to be symetric. This should trend toward a normal curve with a higher number of simulations.