Properties of the distribution of the mean and variance of 40 exponentials

Nathalie Descusse-Brown

January 29, 2019

Overview

The work presented here aims at identifying the characteritics of the distribution of randomly generated exponentials. In order to do so we first perform a number of simulations with the goal to produce 1000 samples of 40 random exponentials. We know from the Central Limit theorem that the distribution of the sample mean should be approaching the normal distribution if the sample size is large enough, and this is a theory we will investigate through our work in order to confirm this assumption.

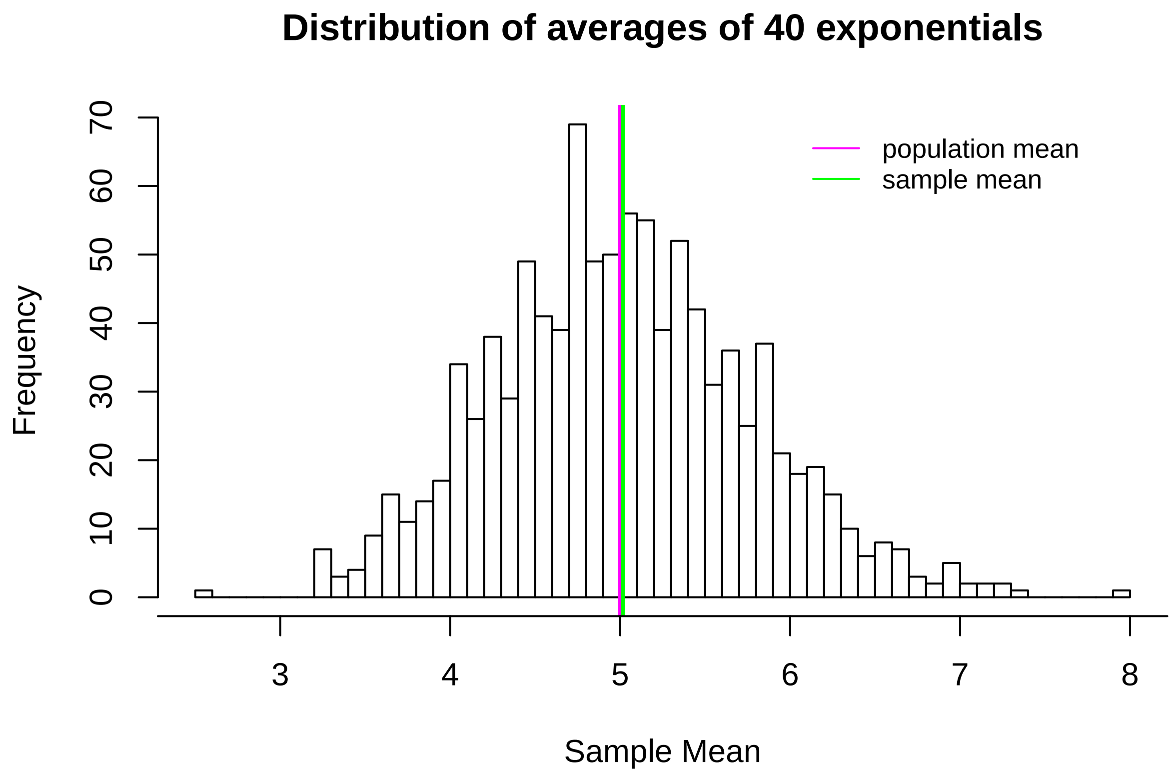
Simulations

First we will generate the 1000 samples of 40 random exponentials with a rate parameter lambda of 0.2. For each of these samples we will calculate the mean and the variance. We will then compare the means and variances distributions to the population mean (which the theory tells us is 1/lamba, or in our case 5) and population variance (which is 1/lambda^2, or 25 in the present case).

```r
library(knitr)
set.seed(2)
lambda <- 0.2
rexpsims <- rexp(40000,lambda)
rexpsimsm <- matrix(rexpsims,1000,40)
meansexp <- apply(rexpsimsm,1,mean)
varexp <- apply(rexpsimsm,1,var)
```

Sample Mean versus Theoretical Mean

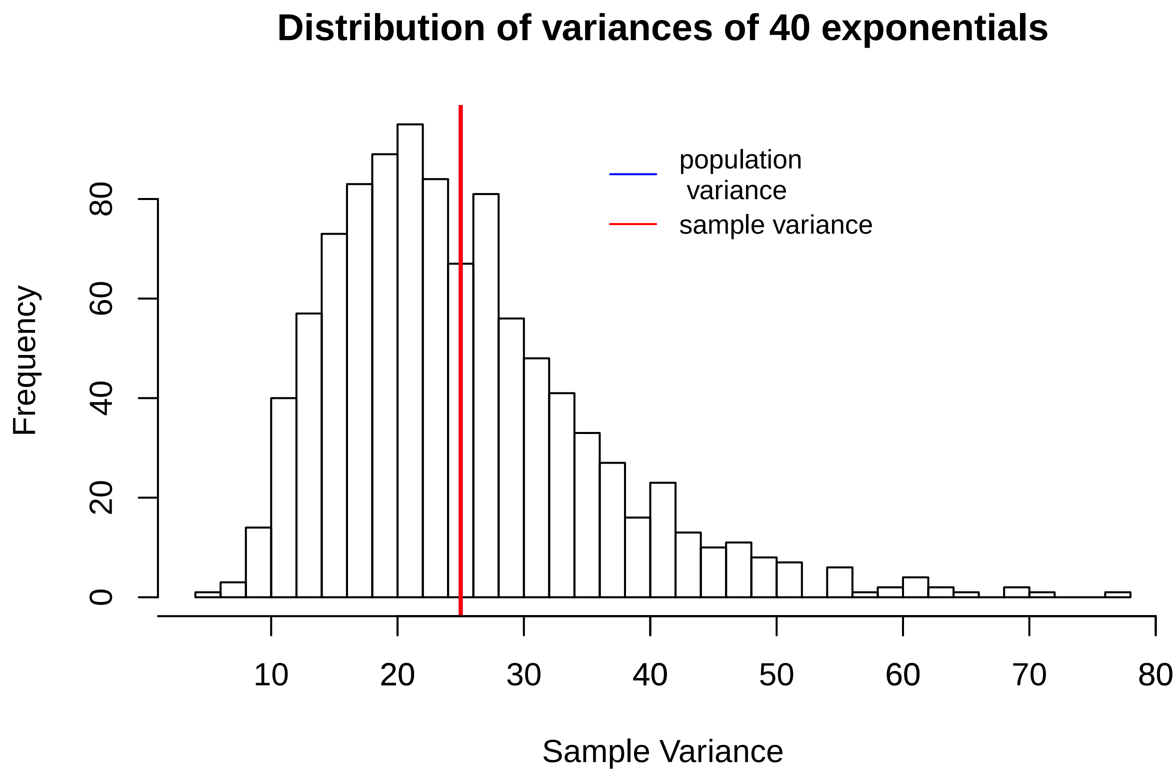
We know from theory that the sample mean should be a pretty good approximation of the population mean. In this instance the population mean is 1/lambda = 5, so we would expect our sample mean to be pretty close to 5. This is something we can verify by plotting the distribution of averages we previously simulated and its mean.



We have also calculated the actual mean of this distribution and found it to be `5.0163562`, so a pretty good estimate of the population mean as we can see.

Sample Variance versus Theoretical Variance

We also know that the mean of the distribution of sample variances should be a pretty good approximation of the population variance (1/lambda^2 = 25). Let us verify this by by plotting the distribution of variances we previously simulated and its mean.



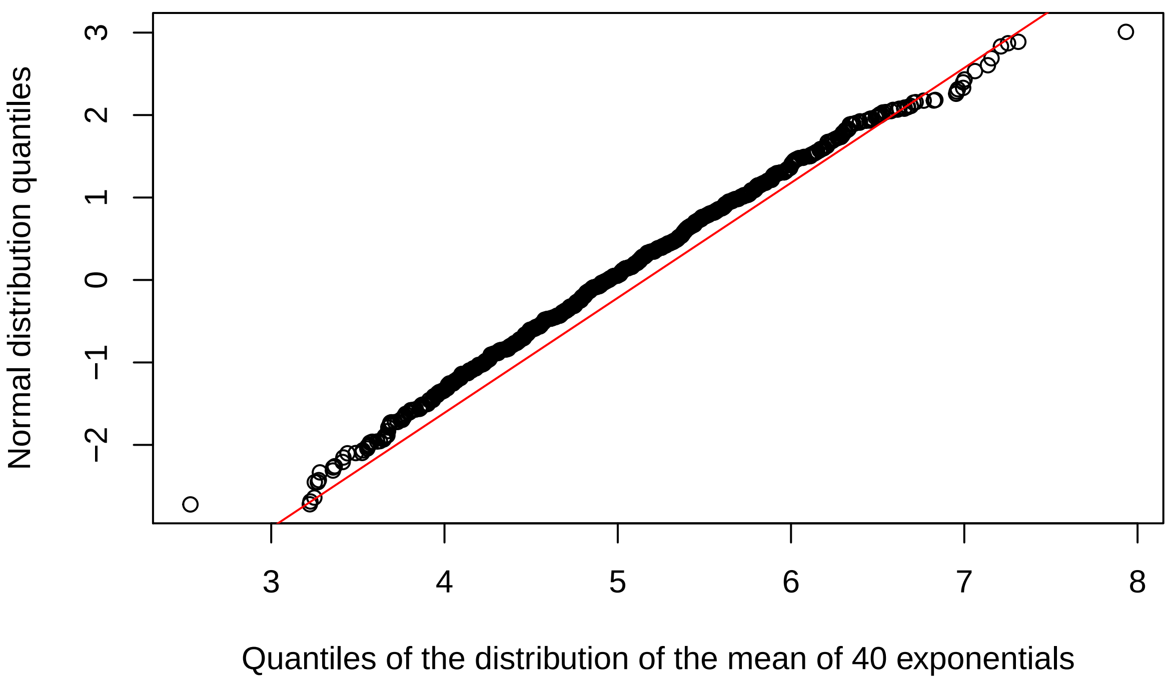
We have also calculated the actual mean of this distribution and found it to be `25.0024311`, so a pretty good estimate of the population variance as we can see.

Distribution

The theory tells us that the distribution of the sample mean and the distribution of sample variance should both be normally distributed. The above plots seem to indicate that the sample mean is indeed normally distributed but there seems to be some skewness in the distribution of the sample variance.

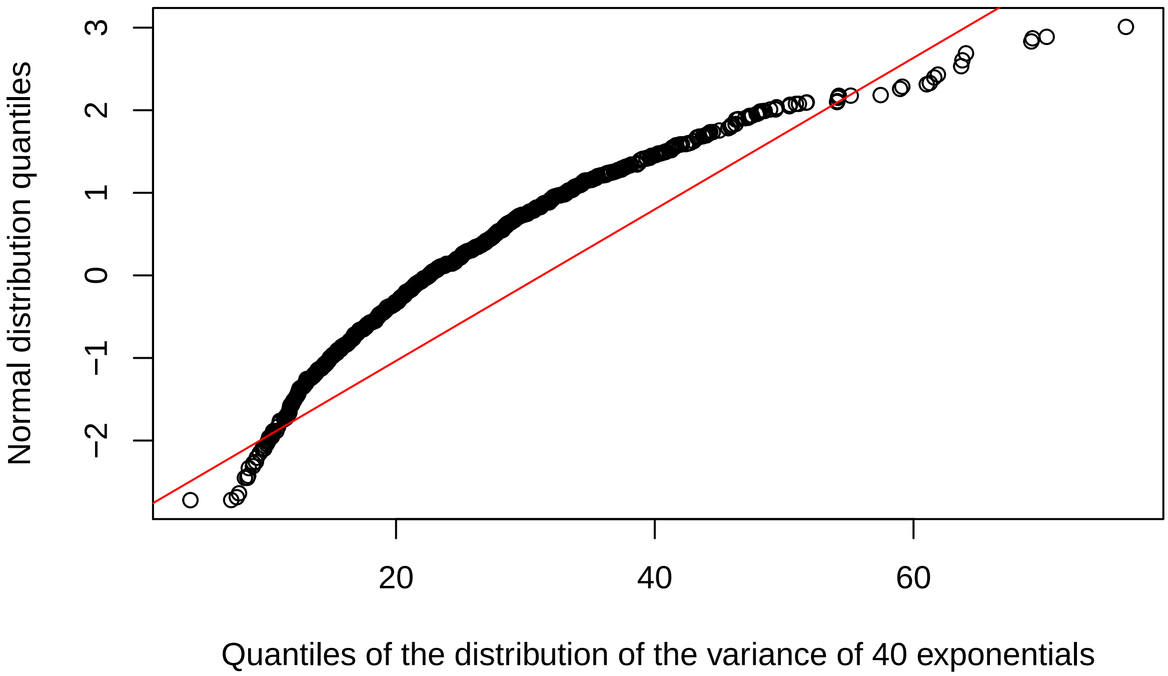
This is something we can verify by using a qqplot, which will plot normal distribution quantiles again the quantiles of our actual distribution. If our distribution is approximately normal then the qqplot points should all approximately fit on a straight line.

Sample Mean



As we can see from the above plot, the linear model fits the data reasonably well, which helps us conclude that the distribution of sample means is approximately normal.

Sample Variance



The above plot demonstrates that the distribution of sample variances is positively skewed so does not follow a normal distribution.