Exploring the exponential distribution using simulations

John Lee

21 November 2015

Overview

The exponential distribution is the probability distribution that describes the time between events in a Poisson process, a process in which events occur continuously and independently at a constant average rate. This report explores the properties of samples of this distribution. Running 1000 simulations, sampling 40 times each simulation we explore the behavior of the means of these samples.

Introduction

A histogram of 10,000 samples from the exponential distribution (with a rate parameter of 0.2, which will be used for the remainder of the report) provides insight into the shape of this probability distribution. The rate parameter of the distribution, often denoted by the greek letter λ, can be used to calculate, among other properties of the distribution, the mean $(1/\lambda)$, the variance $(1/\lambda^2)$ and the median $(\ln(2)/\lambda)$. The mean of the samples in Supplimentary Figure 1 (shown in blue) is 5.02, very close to the population mean of 5. This accuracy is expected as 10000 samples of the distribution were taken. If instead we sample only 40 times, our estimate of the true population mean becomes more variable. This report describes the distribution of the means of such randomly sampled sets of 40 numbers from the exponential distribution.

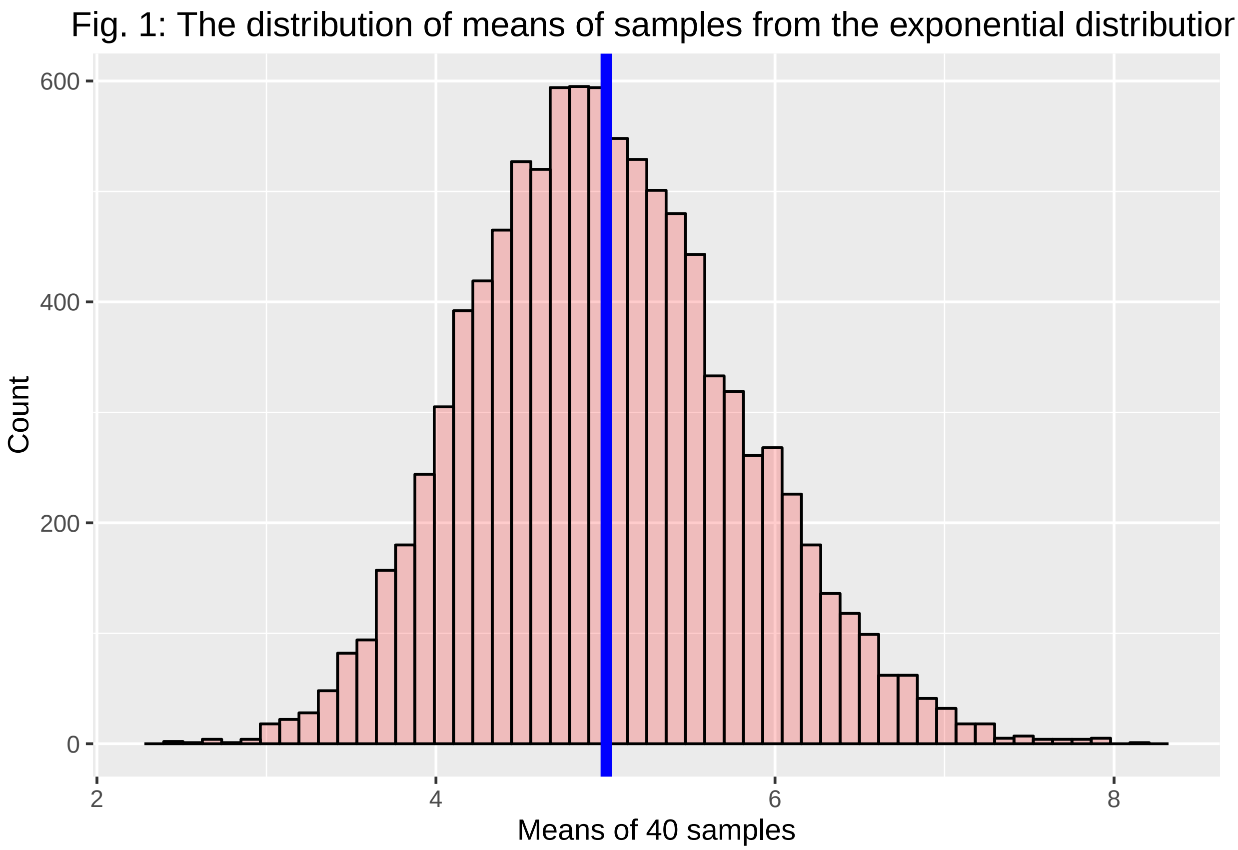
The properties of the distribution of the mean of 40 exponentials.

1000 simulations were run sampling 40 times from the exponential distribution each time. This was performed by using the 'rexp' function in R:

```
num_sims <- 10000
n <- 40
rate_var <- 0.2
set.seed(10)
vals <- matrix(rexp(num_sims * n, rate = rate_var),num_sims)
averages <-  apply(vals, 1, mean)
```

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

The distribution of the mean of 40 exponentials is depicts in Figure 1. The mean of this distribution (depicted by the blue line) is 5, close to the true mean of 5 of the original exponential distribution (calculated as $1/\lambda$). Figure 1 shows that the distribution of these means no longer approximates the original sampled distribution but instead appears more like a normal distribution (see subsection 3).



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

The theoretical variance of a population of sample means is σ^2/n, which in the case of sampling from the exponential distribution becomes $1/(\lambda^2$ x n). With a rate of 0.2 and with 40 samples used for each mean we would expect the variance to be 0.625 :

```
theoretical_variance_of_means <- 1/ (0.2^2 * 40)
```

This is similar to the variance of 0.61 calculated from the 1000 sample means generated from the simulations:

```
variance_of_simulated_means <- var(averages)
```

3. Show that the distribution is approximately normal.

As mentioned before the distribution of sample means appears normal. In Figure 2a, the histogram of the means are once again plotted but this time standardised and normalised. A smooth density estimate using the ggplot2 plotting library estimates the probability density function of the sample means (depicted as a red line). Additionally a gaussian curve is superimposed on the plot (black) to observe the similarity. In Figure 3, a normal Q-Q plot also shows the normality of this distribution. It can be seen that some deviations exist in the tails but overall we can observe the central limit theorem in action. The distribution of samples from the exponential distribution (see Supplimentary Figure 1) is no longer observed when the means of a number of samples are assessed. With a sufficient number of samples, in this case 40, the distribution of the means can be described by a normal distribution instead of the original distribution used for sampling.

Fig. 2: Sample means exhibit a gaussian distribution

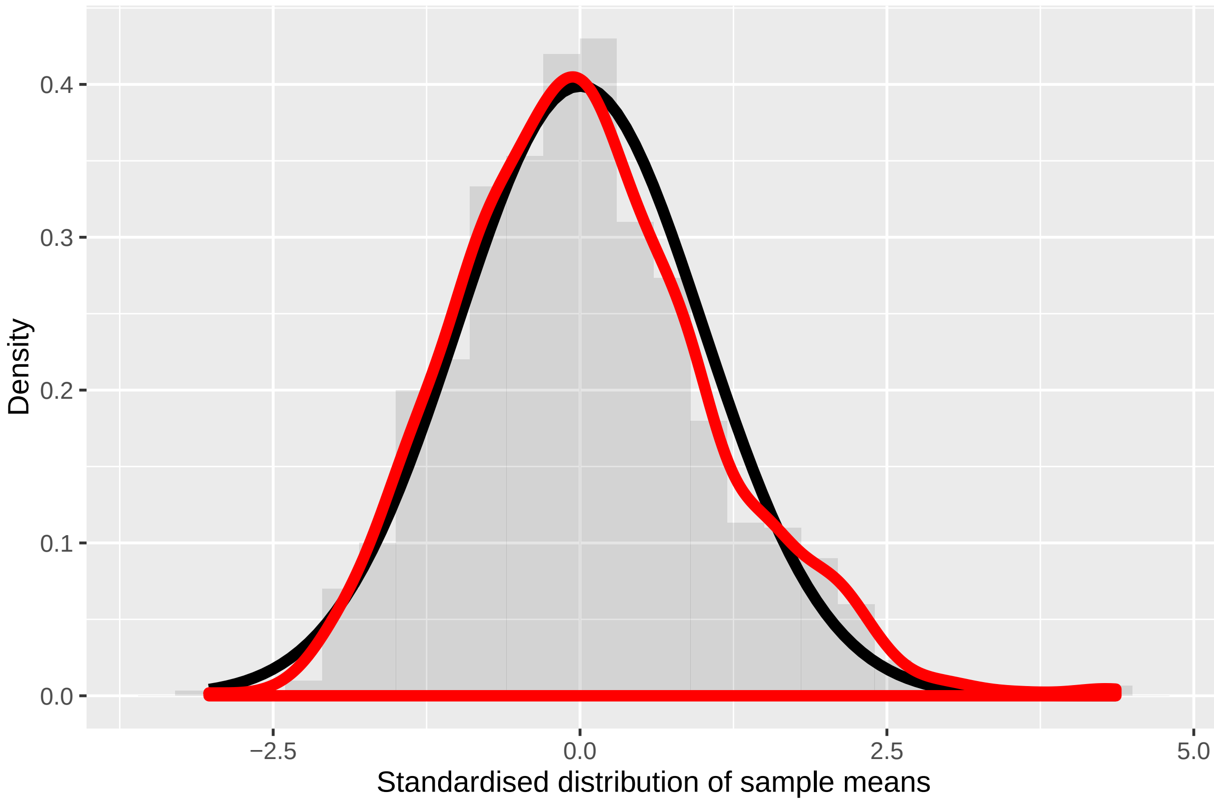
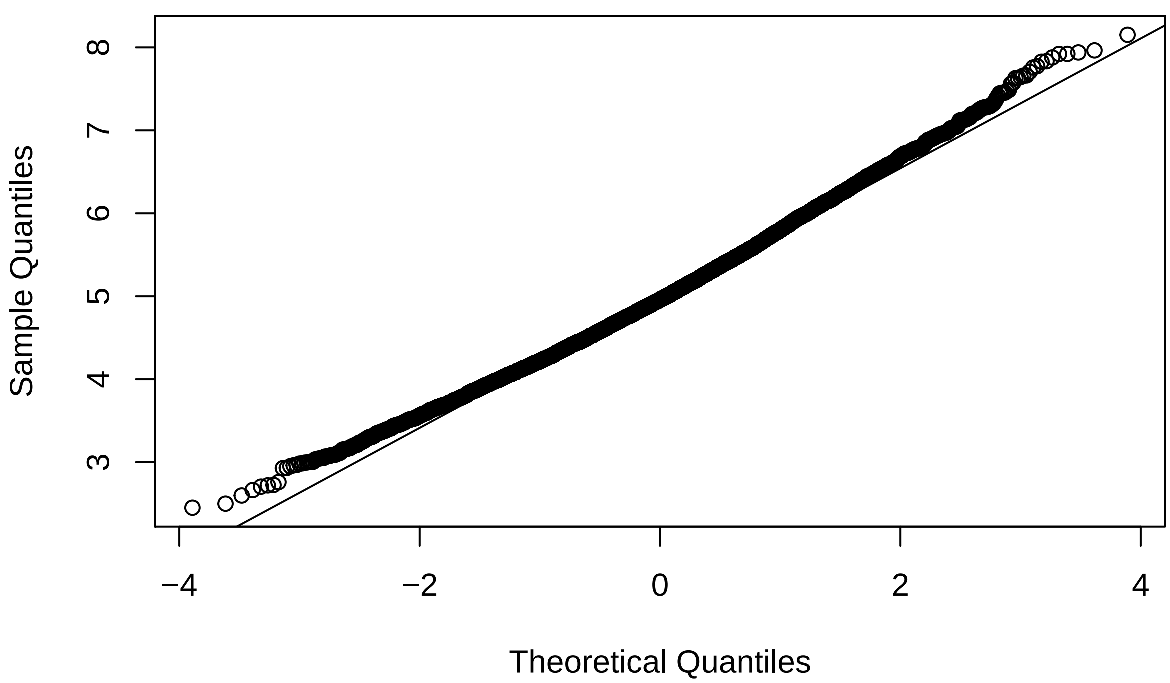
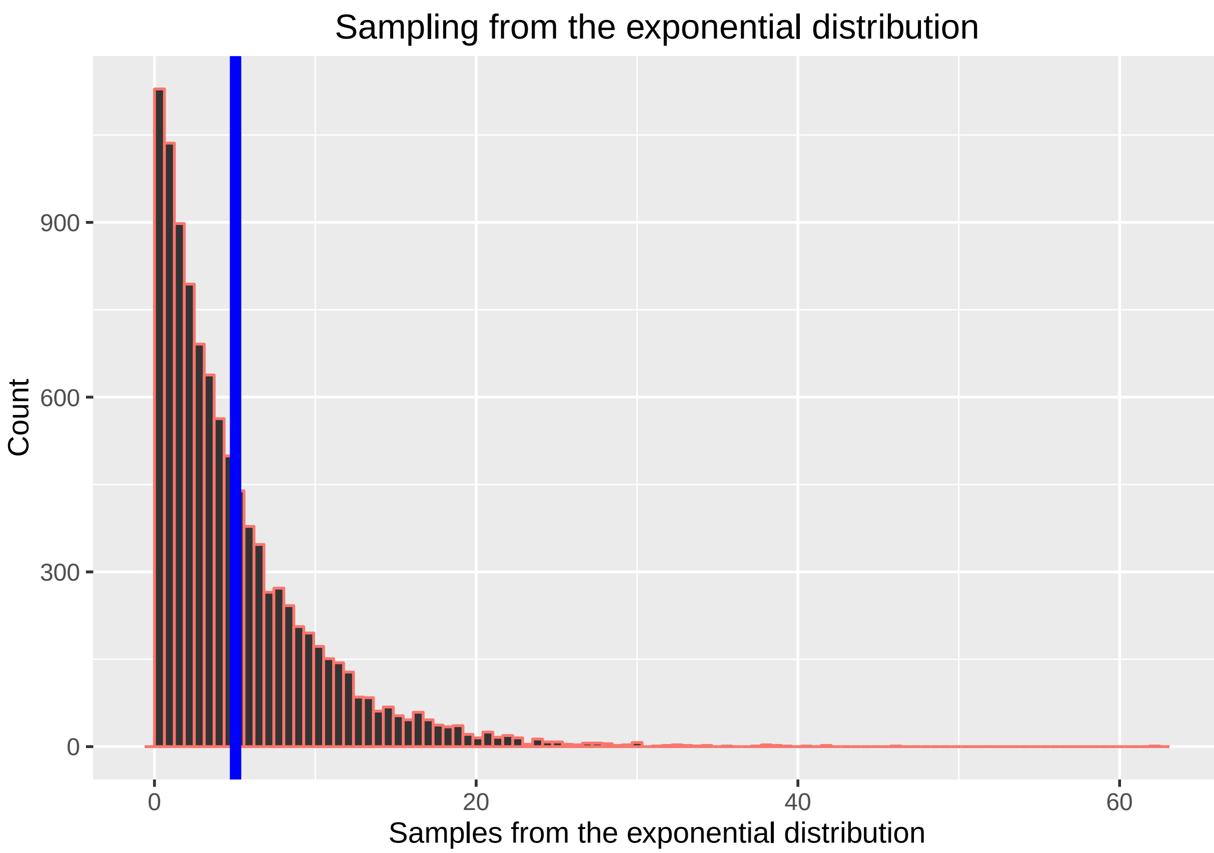


Fig. 3 : Normal Q−Q plot of means



Appendix


Sampling from the exponential distribution
Count
900
600
300
0
0
20
40
60
Samples from the exponential distribution