Central Limit Theorem in Exponential disrtibution

Polina Sklyarevsky

The following report consists of four parts: In part 1, it shows a simulation of an Exponential distribution (where lambda = 0.2). In part 2, it shows a distribution of averages of samples from an Exponential distribution. In part 3, it shows how the distribution of averages of samples (generated in part 2) and the original distribution (generated in oart 1) are centered around the same mean value, and have a similar variance. In part 4, it shows how the distribution of means approaches (but doesn't exactly become) a normal distribution. The Central Limit Theorem suggests that as the number of simulations increase, this distribution will come close to normal. To prove that empirically, a larger sinulation is probably needed.

1. Simulating the Exponential Distribution

1000 observations from the exponential distribution at rate 0.2 are randomaly generated. (See Figure 1 in the appendix).

2. Sampling from the Exponential Distribution

1000 averages of observations from the exponential distribution (n = 40) at rate 0.2 are randomaly generated (see figure 2 in the appendix).

3. Sample vs. Theoretical distribution

3a. Comparing Means

The theoretical mean of an exponential distribution is $\mu = 1/\lambda = 5$ ($\lambda = 0.2$). The sample mean is calculated from the distribution generated in part 2.

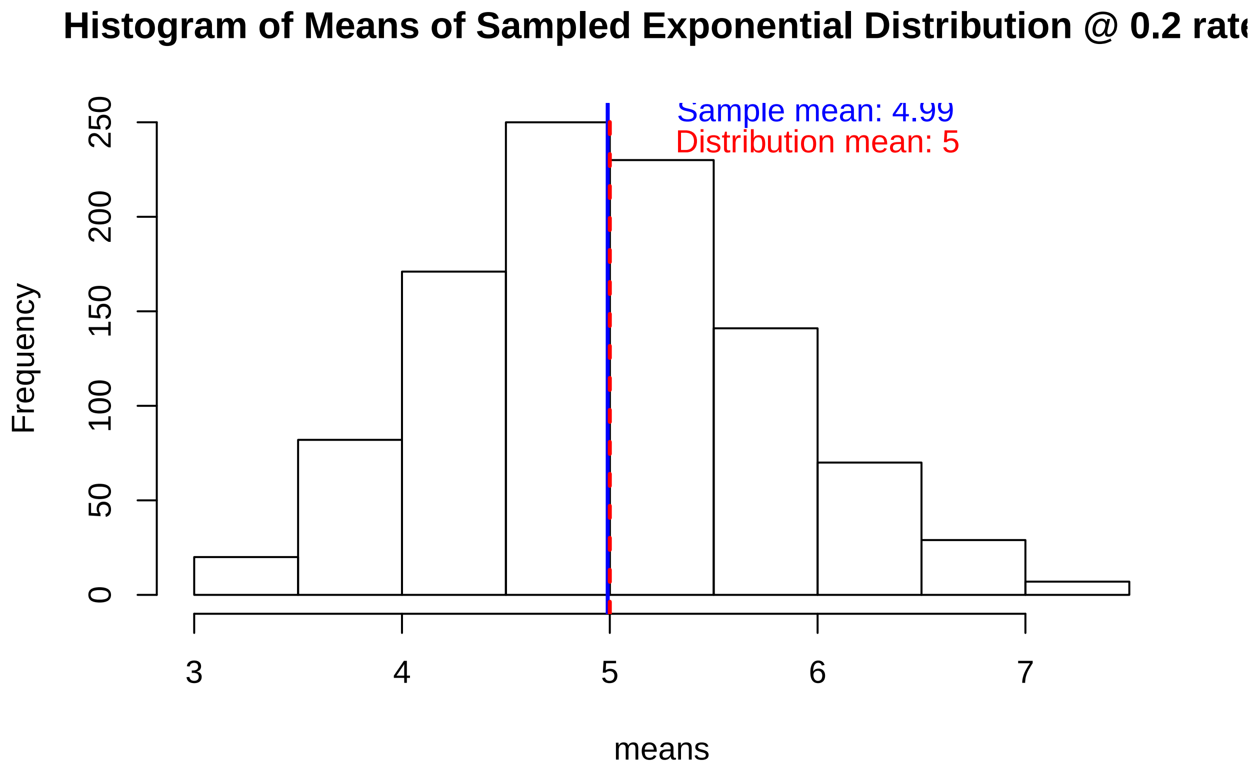

Figure 3: This figure shows a histogram of the sampled distribution (same as figure 2). Both means - the theoretical mean of the exponential distribution (red dotted line), and the sample mean (blue line) are layed over it. It can be seen vlearly that both distributions are centered around the theoretical mean. The mean values are shown for convenience.

3b. Comparing Variance

The theoretical variance of the exponential distribution is $\sigma = 1/\lambda/\sqrt{n}$. Therefore the Theoretical variance is the power of that: $1/\lambda^2/n$ (where $\lambda = 0.2$, $n = 40$). The sample variance is calculated from the distribution generated in part 2. Comparing the two values, it can be seen that they are very close:

```
## [1] "Theoretical variance is: 0.62"
```

```
## [1] "Sample variance is:: 0.61"
```

4. Normality of Distribution?

The CLT suggests that averages of samples from the Exponential distribution will approach zero as the number of simulations grow. Looking at the distribution suggests the averages sampled from the exponential distribution are approaching normal. Although more formal normality tests show that normality is not yet reached. Some more simulations might bring the sampled distributions more closer to normality.

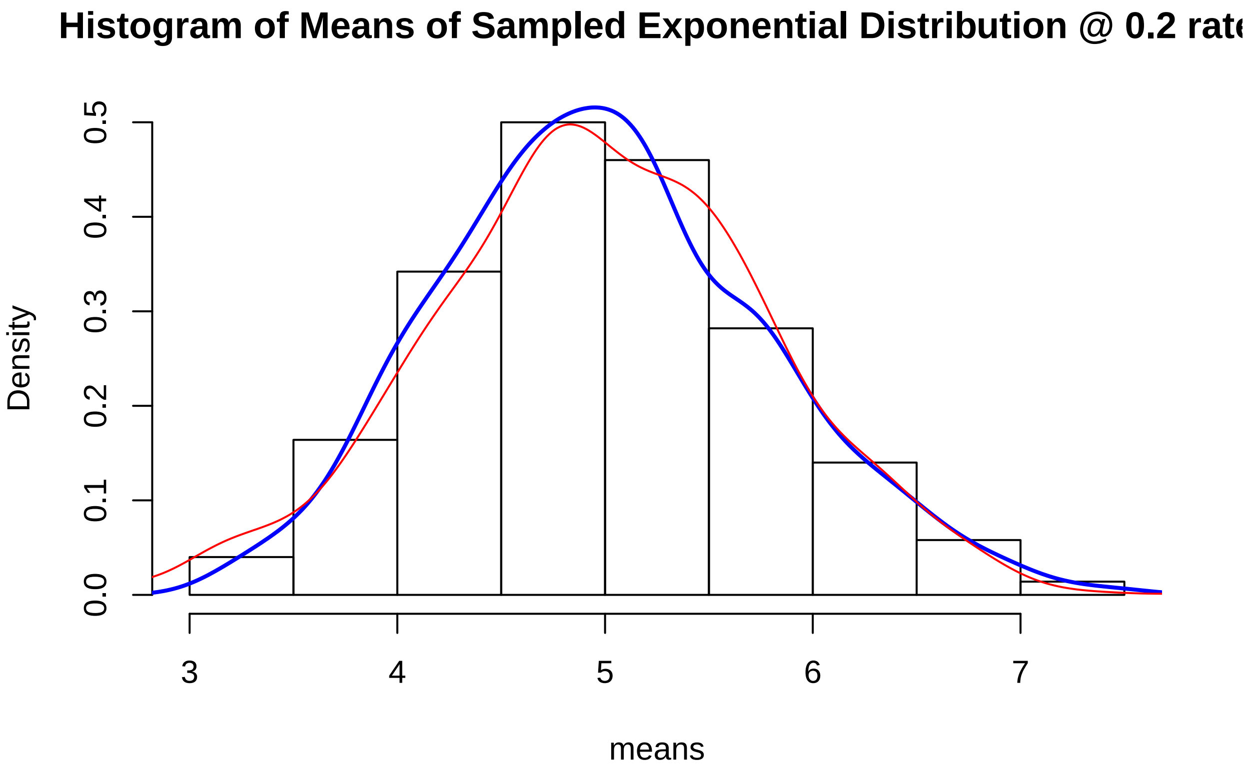

Figure 4: This figure shows a histogram of the sampled distribution (same as figurs 2 and 3). Overlayed are the densities of the sample distribution (in blue) and the normal distribution with the same mean and variance (in red). It can be seen that the lines are pretty close.

A Q-Q plot also shows closeness to normality, with some variation in the left tail (see figure 5 in the appendix). This suggests closeness to normal but perhaps not formal normality. A Jarque-Berra test is performed (where H0: non-normality, H1: normality). H0 could not be rejected ($P < 0.01$). Therefore, the sample is not completely normal in terms of skeweness/kurtosis, even if it looks close to normality to the naked eye, and expected to be normal according to CLT (see Appendix for full test output). A bigger simulation is suggested to achieve normality.

Appendix

Code

All code can be found on Github

Figures

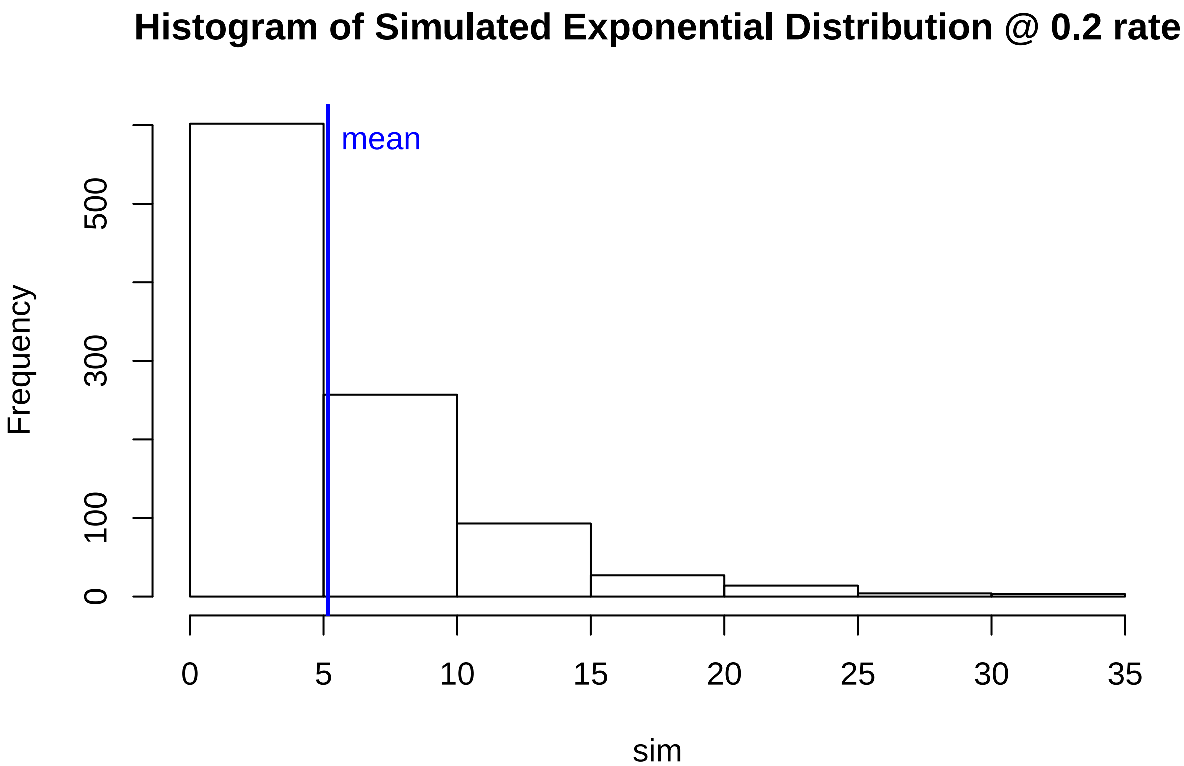

Figure 1: This figure shows a histogram of 1000 exponentioals. It's mean is marked in blue.

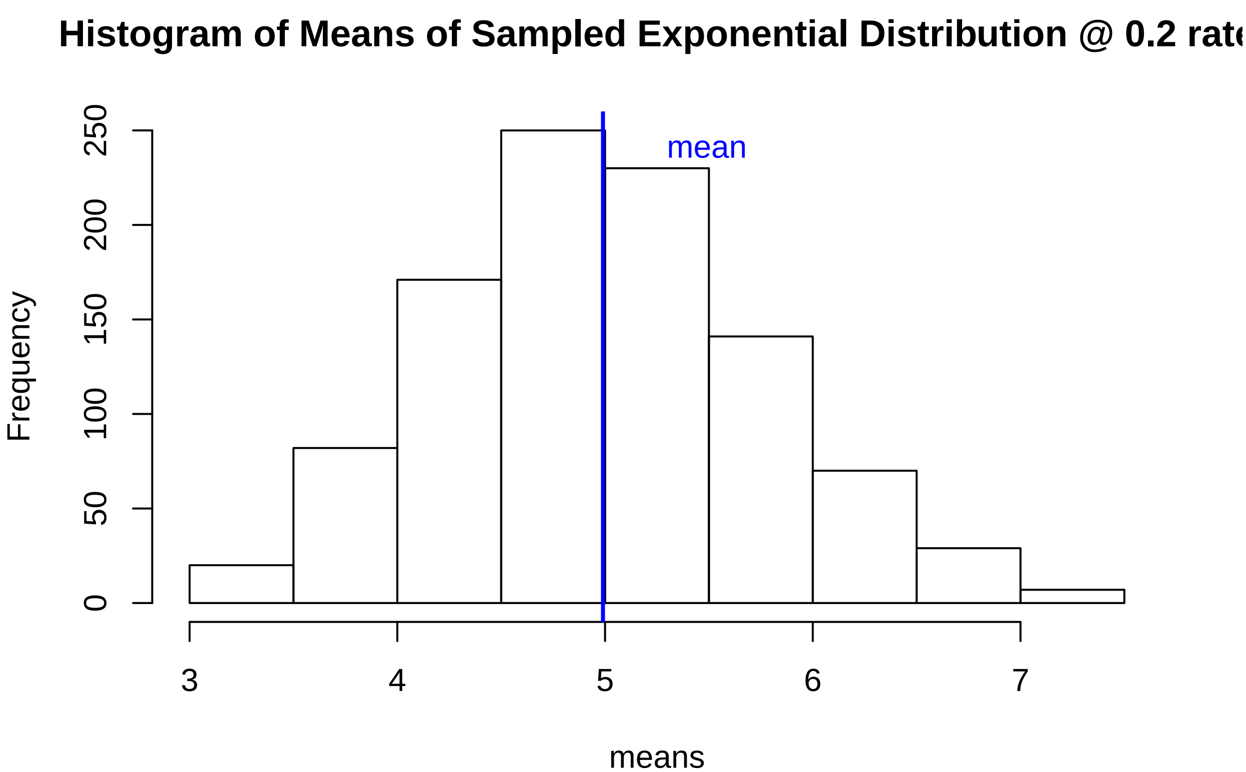

Figure 2: This figure shows a histogram of 1000 averages of sampled exponentials (each of size n = 40). It's mean is marked in blue.

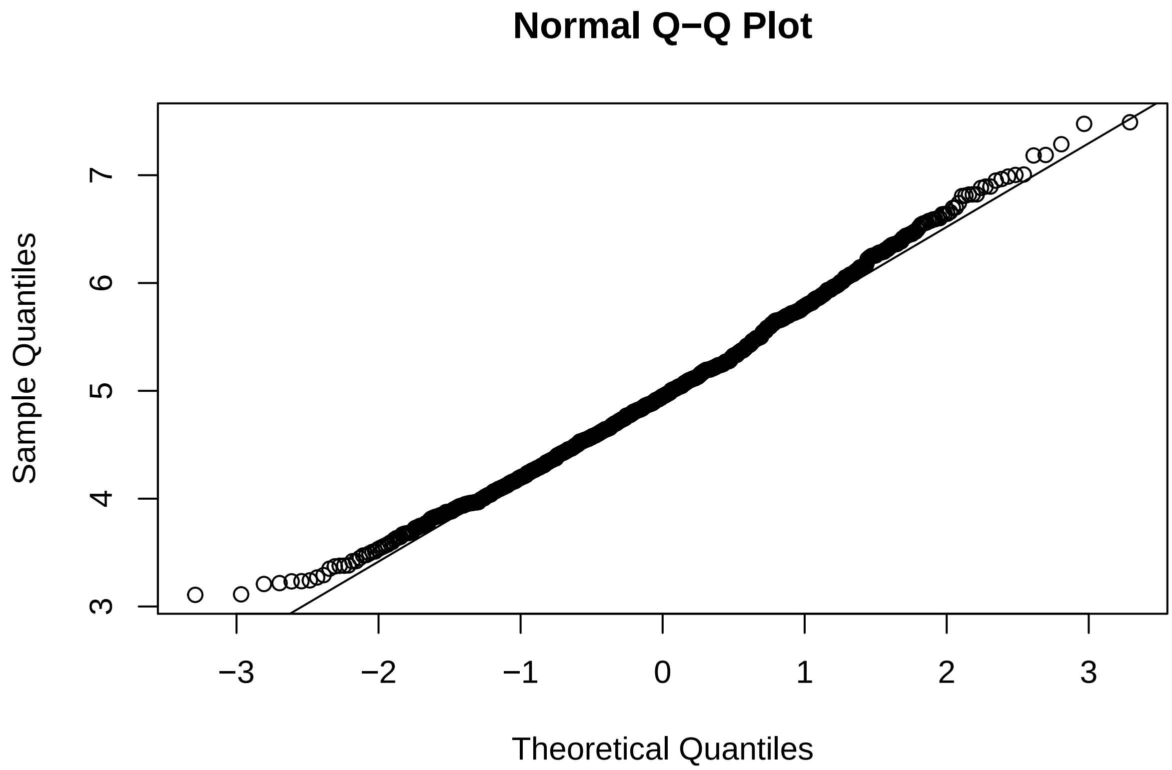

Figure 5: The Q-Q plot shows that the sampled distribution is close to normal (although it is more tailed on both sides).

Tests

Jarque-Berra Normality test results:

```
##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 13.6938
##   P VALUE:
##     Asymptotic p Value: 0.001063
##
## Description:
##  Thu Jun 22 12:21:07 2017 by user:
```