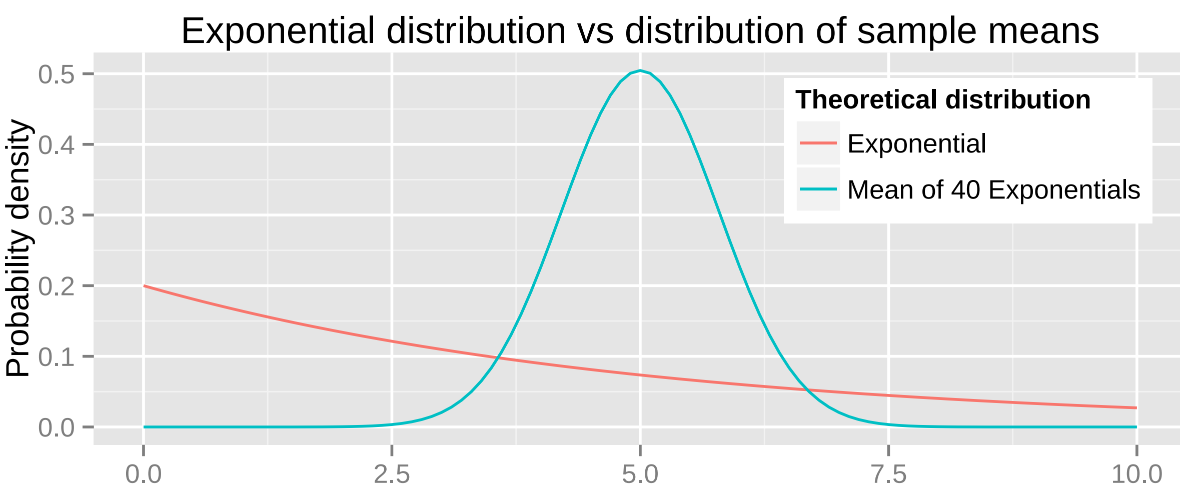
Application of the Central Limit Theorem to the exponential distribution

Chris Lill

Monday, June 15, 2015

Overview

This course project investigates the Central Limit Theorem, and it's application to the mean of observations from the exponential distribution. This is particularly interesting when you look at the difference between the exponential distribution, and the fact that distribution of its sample means will converge to a normal distribution for large samples.



Simulations

This investigation created a distribution of 1000 means, stored in a vector named means. Each of these means was generated from 40 observations from an exponential distribution with a lambda value of 0.2. This should give a distribution large enough to illustrate the Central Limit Theorem.

The following R code was used:

```
set.seed(1234)
collection.size <- 40
lambda <- 0.2
number.simulations <- 1000

means = NULL
for (i in 1 : number.simulations) {
    collection <- rexp(collection.size, lambda)
    means = c(means, mean(collection))
}
```

Sample Mean versus Theoretical Mean

In this analysis, the distribution is represented by 1000 observations of the mean of 40 exponentials. The sample mean of this distribution can be calculated in R.

$$SampleMean = mean(means) = 4.974$$

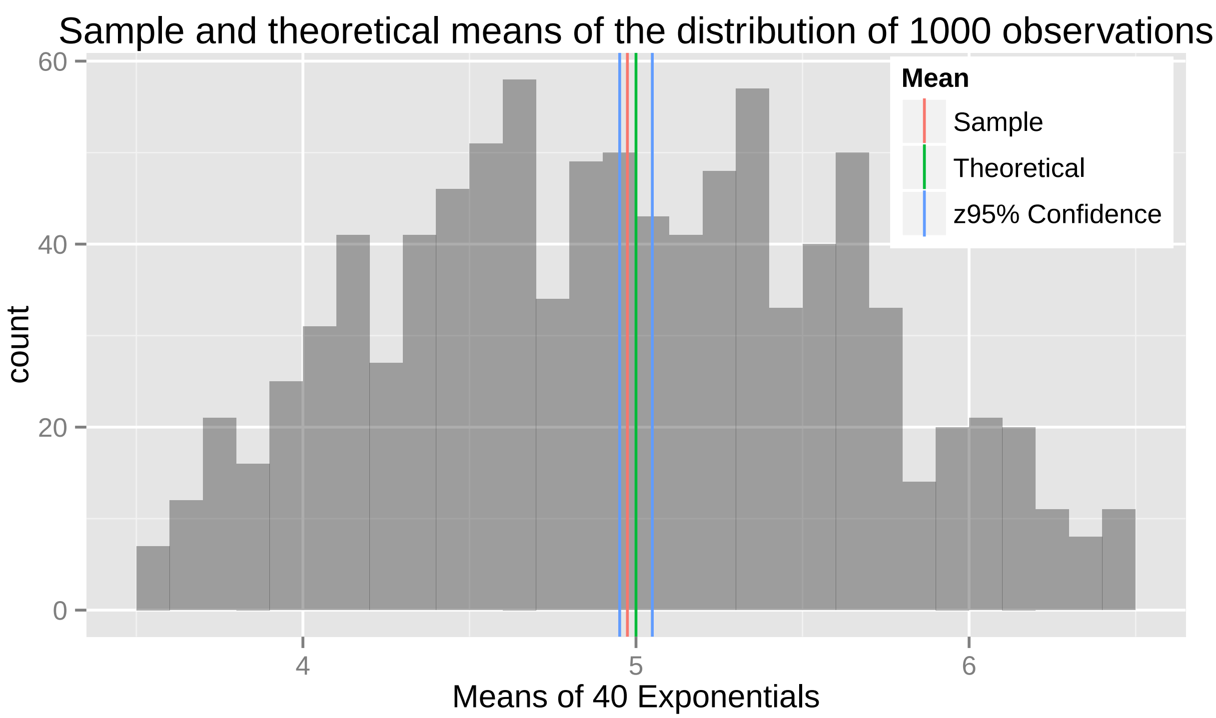
The theoretical mean of the distribution is the same as the mean of the underlying exponential distribution. The mean of the exponential distribution is 1/lambda.

$$TheoreticalMean = \frac{1}{\lambda} = \frac{1}{0.2} = 5.000$$

The 95% confidence interval for the theoretical mean of the distribution can be calculated using the standard error of the distribution (of 1000 observations). This in turn can be calculated from the standard error of 40 observations from the exponential distribution.

$$ConfidenceInterval = \frac{1}{\lambda} \pm 1.96 \frac{1}{\lambda\sqrt{40 \times 1000}} = 4.951, 5.049$$

The sample mean lies within the 95% confidence interval of the distribution, therefore this supports the validity of the Central Limit Theorem for distributions of 1000 observations.



Sample Variance versus Theoretical Variance

The sample variance of the distribution can be calculated in R by using the var() function.

$$SampleVariance = var(means) = 0.571$$

The theoretical variance is equal to the standard error of the mean of 40 exponentials. The standard deviation of the exponential distribution is $1/\lambda$.

$$TheoreticalVariance = \frac{1}{\lambda^2 \times 40} = 0.625$$

There is a 9% difference between these two values. For a more meaningful comparison a normal distribution with the sample mean and sample variance has been added in blue to the plot below. This shows that the sample variance is a good approximation to the theoretical variance.

Distribution

The histogram below shows the probability densities for the sample and theoretical distributions.

There are statistical methods for testing whether a distribution is approximately normal, by calculating their skew (0.208) and kurtosis (-0.259). This can be calculated in R using the stat.desc function from the pastecs library. However testing the significance of these values isn't appropriate for large samples because the small standard error makes the measure inaccurate.

Instead, a simple inspection of the distribution is sufficient to conclude that the sample distribution is approximately normal. This is a very good indication that the Central Limit Theorem holds true for large samples of means from the exponential distribution, and that a normal distribution is a good approximation for samples >= 1000.