Central Limit Theorem and the Exponential Distribution

Chris O'Brien

13 June 2015

Contents

Overview

In this analysis we set out to investigate the Exponential Distribution (Further Info) and how samples pulled from the distribution can help us understand the theoretical mean and variance.

We also investigate how the means of our samples are distributed, whether they are normally distributed and how this relates to the Central Limit Theorem.

Simulations

In order to really test out the assumptions of the Central Limit Theorem and Law of Large Numbers we ran simulations taking 10, 1000 and 100000 of 40 samples from the Exponential Distribution.

We also take a cumulative mean over 4000 samples to see how the sample mean varies as we add more samples to our mean.

We should expect to see that as we increase the number of samples that we get a better estimation of the theoretical mean and variance.

Sample Mean versus Theoretical Mean

We know that the theoretical mean of the exponential distribution is λ^{-1}, as our simulations the choice of lambda was 0.2 the theoretical mean should be 5.

Theoretical Mean

$$0.2^{-1} = 5$$

From figure 1 we can see that as we add samples and calculate the mean that it converges to the theoretical mean. This demonstrates the Law of Large Numbers that states as we increase the amount of samples their average will converge to the theoretical mean.

While mean of any one sample can vary wildly, as you can see from figure 2 depending on the sample size we have means ranging from 2.264 to 9.506, we know from the Central Limit Theorem that the distribution of our averages becomes that of standard normal and it's own mean is an estimator of the population mean (in this case the known theoretical mean of 5).

From figure 3 we can see that as we increase the number of simulations that the density becomes closer to looking like standard normal (which we touch on later). The mean of these averages also comes closer to the theoretical as we increase the number of simulations.

Coming back to figure 2 we can see that the means from 10, 1000 and 100000 simulations are 4.899, 5.001 and 5.000 respectively, which confirms what we know of the Central Limit Theorem.

Sample Variance versus Theoretical Variance

Similar to our analysis of the theoretical mean we know that the theoretical variance can be calculated by λ^{-2}, again, as our choice of lambda in this instance was 0.2 the theoretical variance should be 25.

Theoretical Variance

$$0.2^{-2} = 25$$

Each individual sample can have a wildly different value, sometimes quite far from the true variance. In our simulations from figure 2 we see that they can range anywhere from 3.678 to 161.756.

The average of the sample variances is an estimator for the population variance and as with the example of the mean as we increase the number (see figure 4) of simulations of sample size 'n' we come closer to the true theoretical variance of 25. From figure 2 the average variance for 10, 1000 and 100000 simulations are 25.649, 25.463 and 25.009, almost exactly our theoretical value.

Plotting the distributions of the variances we see again that as we add more simulations the distributions become closer and close to resembling standard normal.

Distribution

Our distribution in this instance is drawn from the Exponential Density that has a density function as shown in figure 5. Our averages of each sample are an estimator for the red line in this graph, the theoretical mean of 5. The distribution of our averages should lie around this number and so as we increase the numbers of sample of size 'n' (or 'n' for that matter) that we take our average distribution become more standard normal.

The standard normal distribution is characterised by having $\mu = 0$ and $\sigma = 1$. We know from the Central Limit Theorem that as we increase our simulation count that the distribution of the average should become standard normal.

From figure 6 we see that when we normalize our 100,000 simulations to 0, the standard deviations of the normalized values is 0.799 a value very close to 1.

Plotting the density functions of our simulation counts against the standard normal distribution it is possible to see that the more simulations we add the closer we approach a standard normal distribution.

Appendix

Simulation

.

```
# Start by setting the seed. This is important for reproducibility
set.seed(100)
lambda <- 0.2
exp10m <- matrix(rexp(400, lambda), ncol = 40)
exp10 <- data.table(mean=rowMeans(exp10m),
                    var=apply(exp10m, FUN=var, MARGIN = 1))
exp1000m <- matrix(rexp(40000, lambda), ncol = 40)
exp1000 <- data.table(mean=rowMeans(exp1000m),
                    var=apply(exp1000m, FUN=var, MARGIN = 1))
exp100000m <- matrix(rexp(4000000, lambda), ncol = 40)
exp100000 <- data.table(mean=rowMeans(exp100000m),
                     var=apply(exp100000m, FUN=var, MARGIN = 1))
cumMeans<-data.table(mean=cumsum(rexp(4000,0.2))/(1:4000), samples=(1:4000))
summaryFrame<-data.table(sim10mean=exp10$mean, sim10var=exp10$var,
                         sim1000mean=exp1000$mean, sim1000var=exp1000$var,
                         sim100000mean=exp100000$mean, sim100000var=exp100000$var)
```

Sample Mean versus Theoretical Mean

Figure 1 .

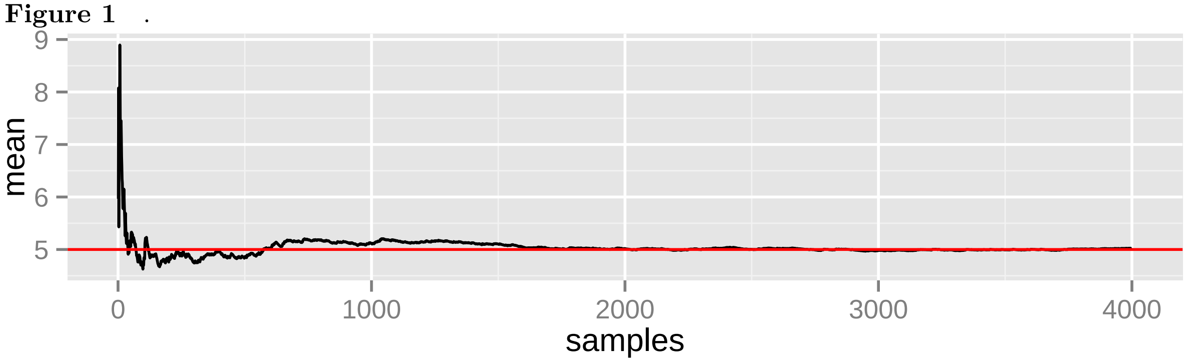


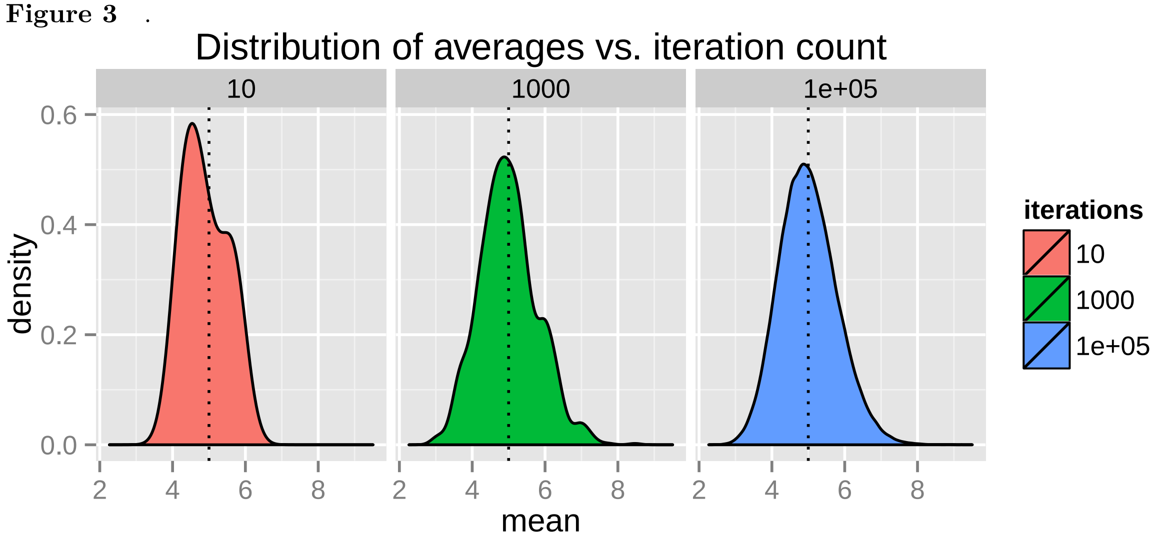
Figure 2 .

```
summary(summaryFrame)
```

```
##    sim10mean        sim10var       sim1000mean      sim1000var
##  Min.   :4.150   Min.   : 9.432   Min.   :2.966   Min.   :  4.847
##  1st Qu.:4.401   1st Qu.:15.373   1st Qu.:4.474   1st Qu.: 17.382
##  Median :4.789   Median :20.388   Median :4.951   Median : 22.824
##  Mean   :4.899   Mean   :25.649   Mean   :5.001   Mean   : 25.436
##  3rd Qu.:5.526   3rd Qu.:36.014   3rd Qu.:5.473   3rd Qu.: 30.946
##  Max.   :5.821   Max.   :51.542   Max.   :8.467   Max.   :103.604
##  sim100000mean    sim100000var
##  Min.   :2.264   Min.   :  3.678
##  1st Qu.:4.448   1st Qu.: 17.196
```

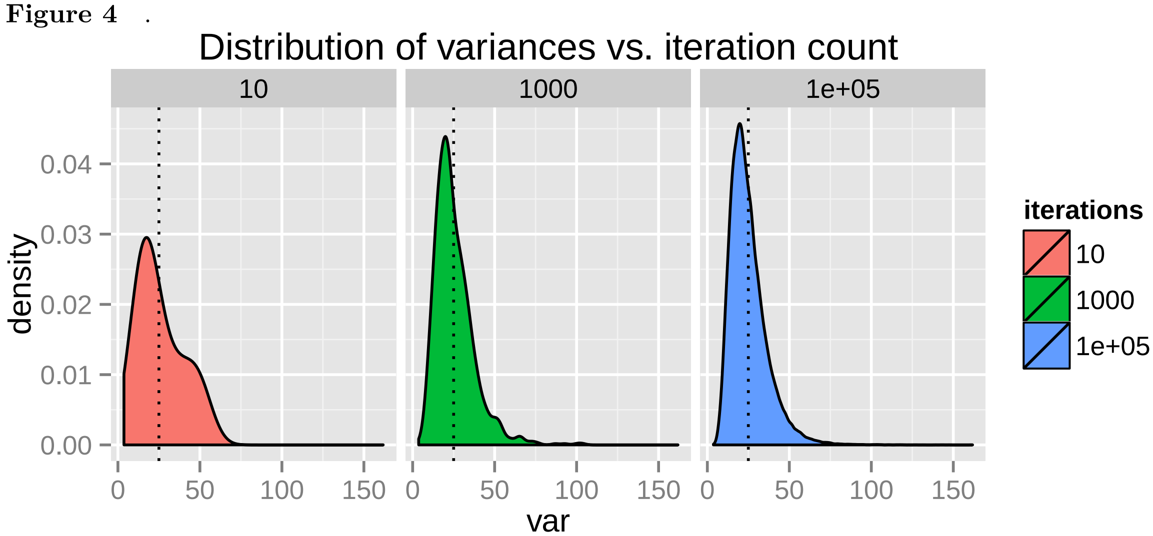
```
##  Median :4.957   Median : 22.818
##  Mean   :5.000   Mean   : 25.009
##  3rd Qu.:5.506   3rd Qu.: 30.373
##  Max.   :9.506   Max.   :161.756
```

Figure 3 .



Sample Variance versus Theoretical Variance

Figure 4 .



Distribution

Figure 5 .

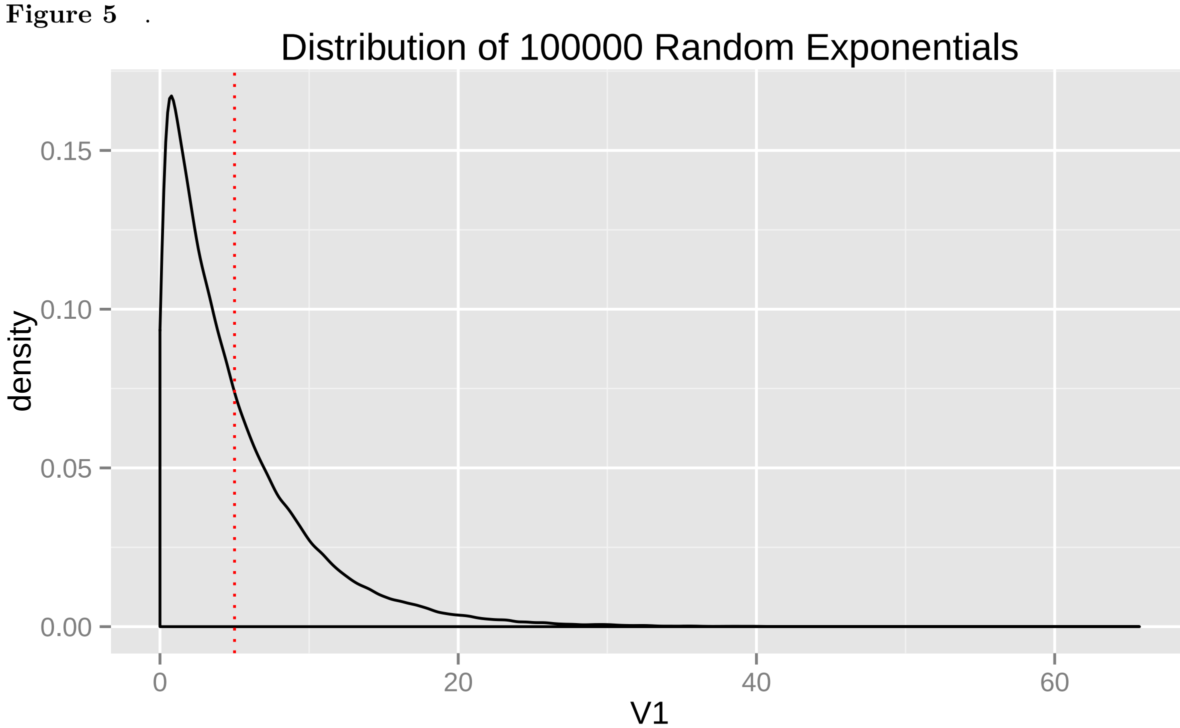


Figure 6 .