Central Limit Theorem with Exponential Distributions

Arkosnato Neogy

Sunday 26 April 2015

In this report we shall investigate the truth of the Central Limit through repeated simulation on samples drawn from an exponential distribution. We shall check the distributions of the sample mean and variance, and see whether they are approximately normally distributed while being centred about the population values of mean and variance that they are trying to estimate.

For an exponential distribution the key parameter is the rate parameter, usually called lambda. For our simulations we shall use a lambda of 0.2.

```
lambda= 0.2
```

Using this we create out dataset, by drawing 40 samples for an exponential distribution with lambda = 0.2 for a 1000 sets. We then find the mean and standard deviation for each set of 40 values thus creating 1000-sized vectors.

```
samps= 40
sims= 1000
x= matrix(data= rexp(samps*sims, lambda), nrow= sims, ncol= samps)
dim(x)
```

```
## [1] 1000   40
```

```
mns= apply(x,1,mean)
sds= apply(x,1,sd)
length(mns)
```

```
## [1] 1000
```

```
length(sds)
```

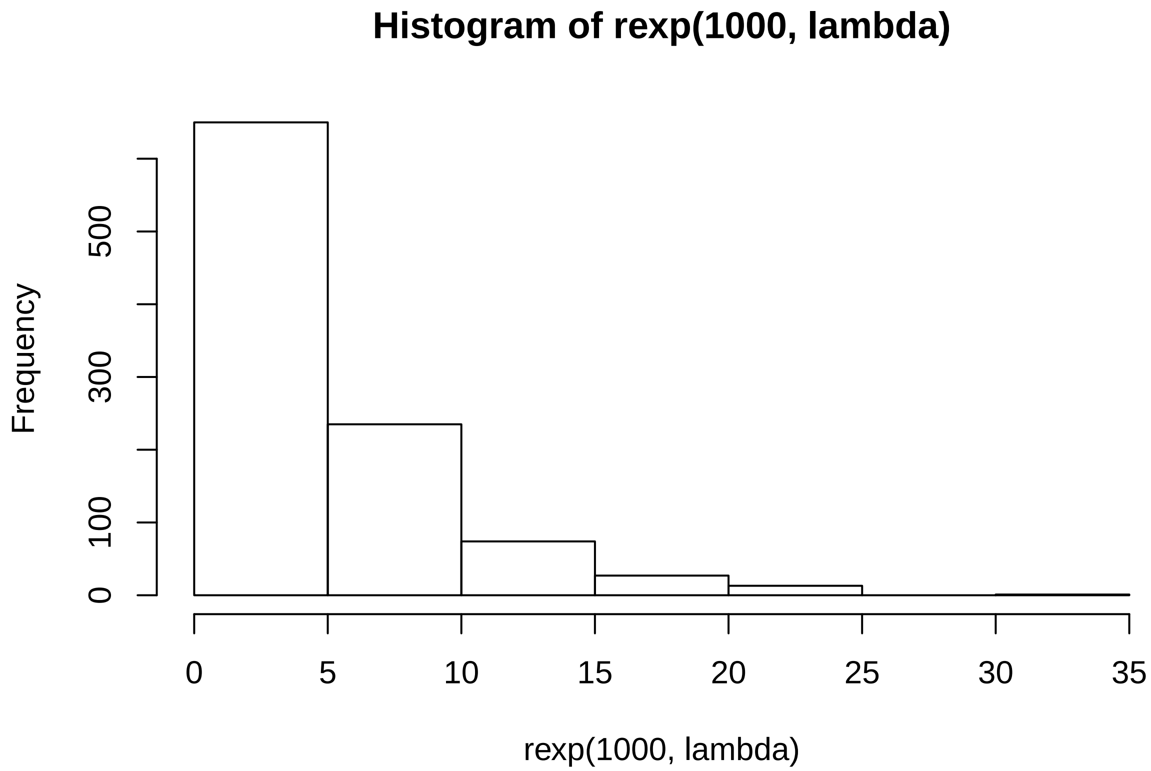
```
## [1] 1000
```

First up, we check if the mean of the 1000-sized vectors we have thus obtained matches the population parameter we are trying to estimate. We check for both the mean and the variance :

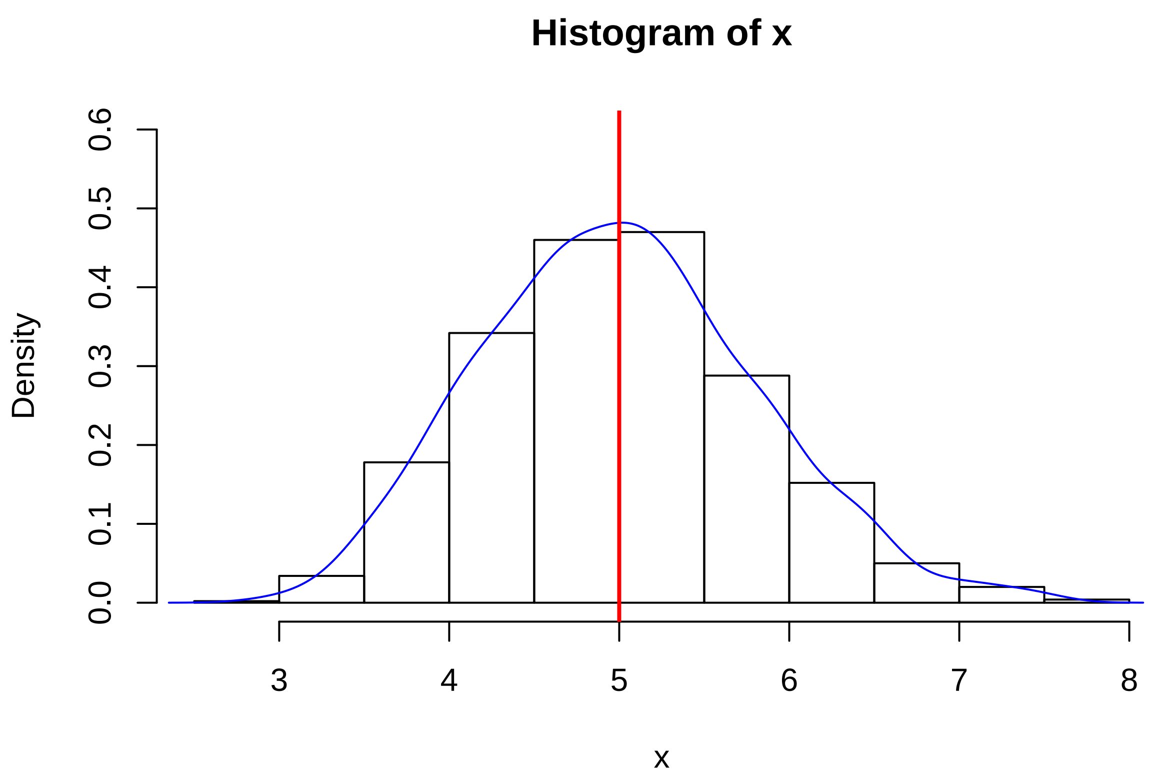
```
## [1] "Mean of the sample distribution of means is  5.001 while the theoretical mean is  5"
```

```
## [1] "Mean of the sample distribution of variance is  23.804641 while the theoretical variance is  25
```

That sounds encouraging! Let us now look at the distributions themselves. For starters, lets take a look at the distribution of 1000 samples from an exponential distribution with lambda= 0.2



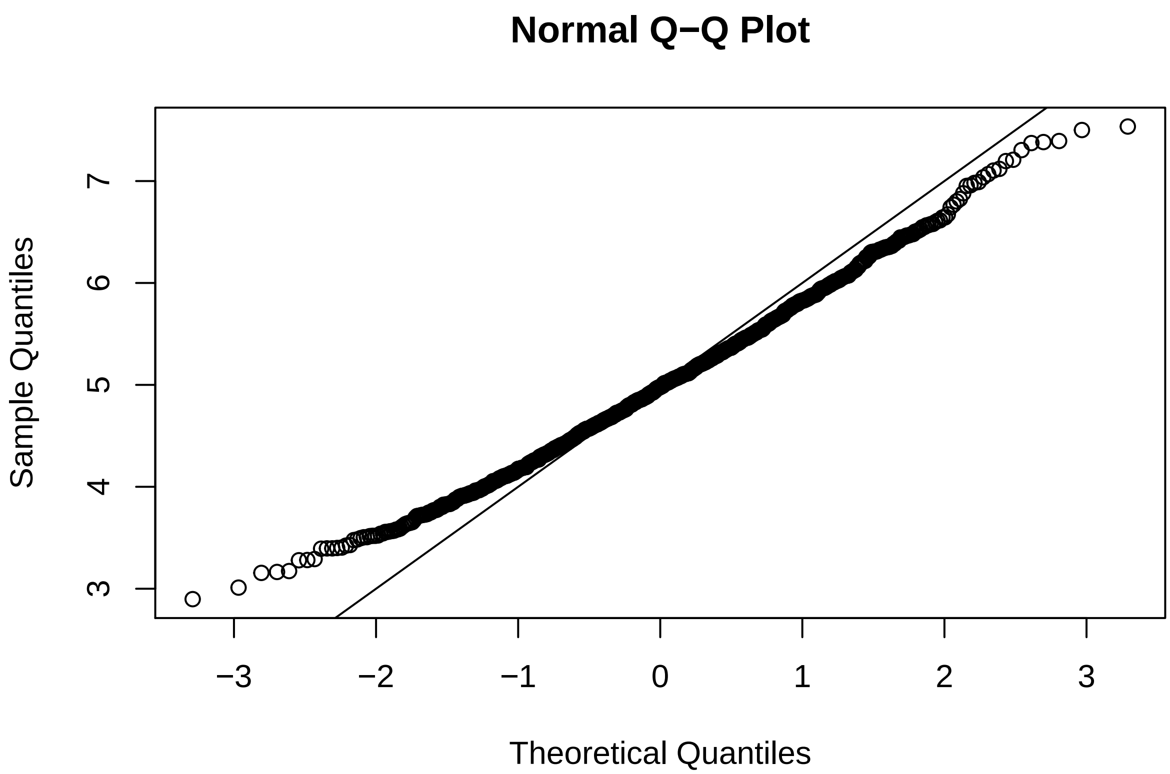
Now let us look at the distribution of the mean of the 40 samples that we drew for 1000 times.



The histogram here is shown in probablity rather than frequency, which jst means that the y-axis is normalized by 1000. We did this so that we could overlay the distribution of the values on roughly the same scale, as shown in blue. This is very different from the exponential distribution we started off with, and while its not very Gaussian looking, it is roughly close to it. Note too that the distribution is roughly centered at the expected value of 5!

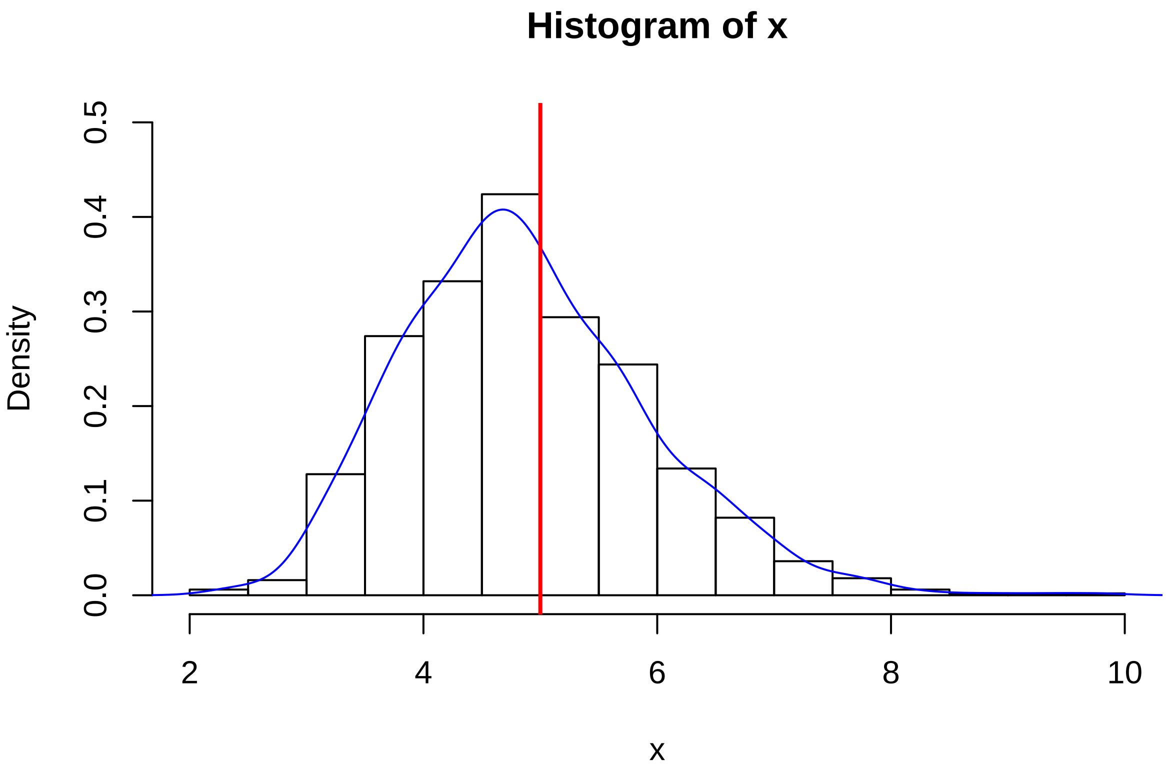
A good way of checking whether a set of value is normal or not is to look at its Q-Q plot, and see how close it looks to a straight line. Lets try that for the values we got for the means.

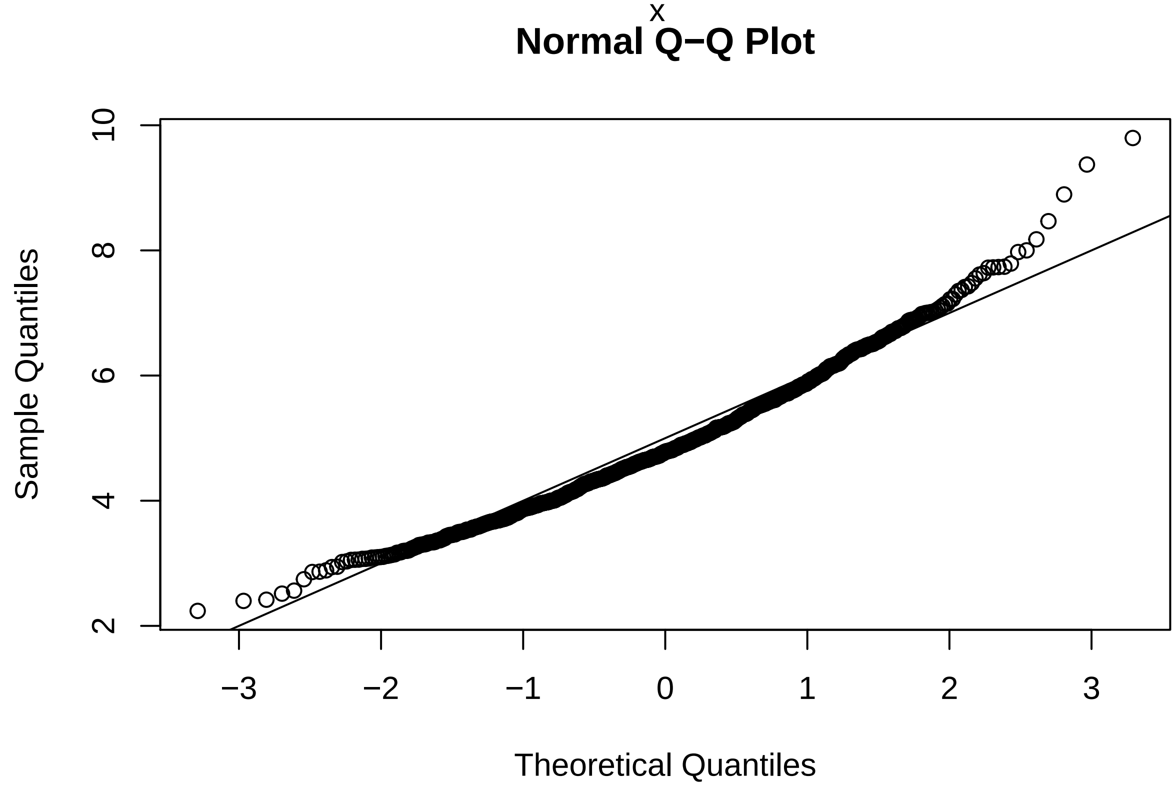
```
qqnorm(mns)
abline(5,1)
```



We have overlaid the expected straight line (passing through the (0,5) point with a slope of 1), and it looks close though not exactly perfect.

We repeat the same exercise with the standard deviation now.





Once again we see a roughly normal distribution centred about the expected population standard deviation of 5.

In view of the analysis done above we can claim the CLT to hold true!