Demonstration of LLN and CLT using the exponential distribution

Unmesh Phadke

22nd July,2015

This a report in which we will simulate the exponential distribution and compare it with the central limit theorem. We will first plot a histogram for the exponential distribution with 40 samples with the parameter lambda equal to 0.2. Then we will plot a histogram for the averages of 40 exponentials for 1000 simulations.

Now comes the actual simulation:

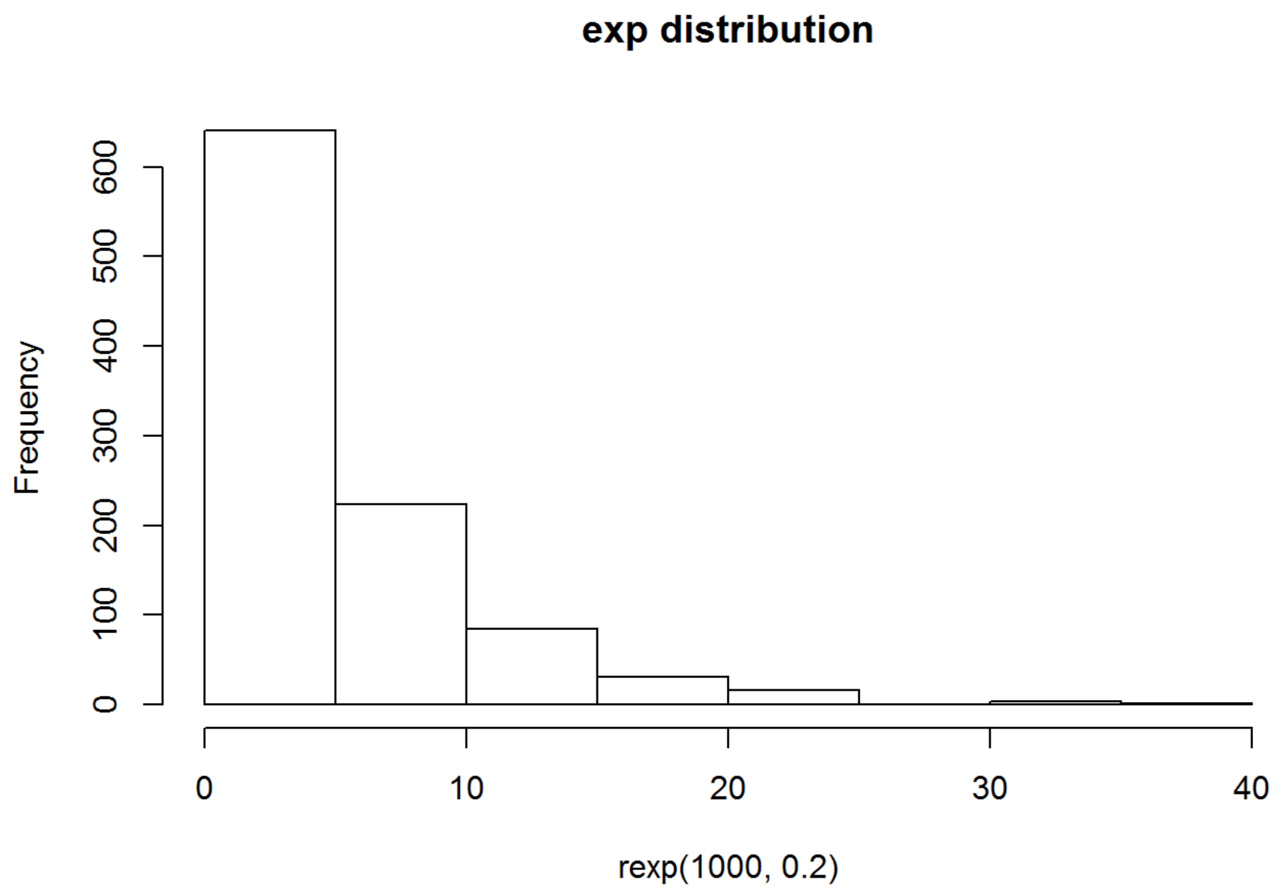
So, first we will deal with just tje exponential distribution.

We set the parameter lambda

```
lambda=0.2
```

We now plot the histogram for this distribution.

```
hist(rexp(1000,0.2),main="exp distribution")
```



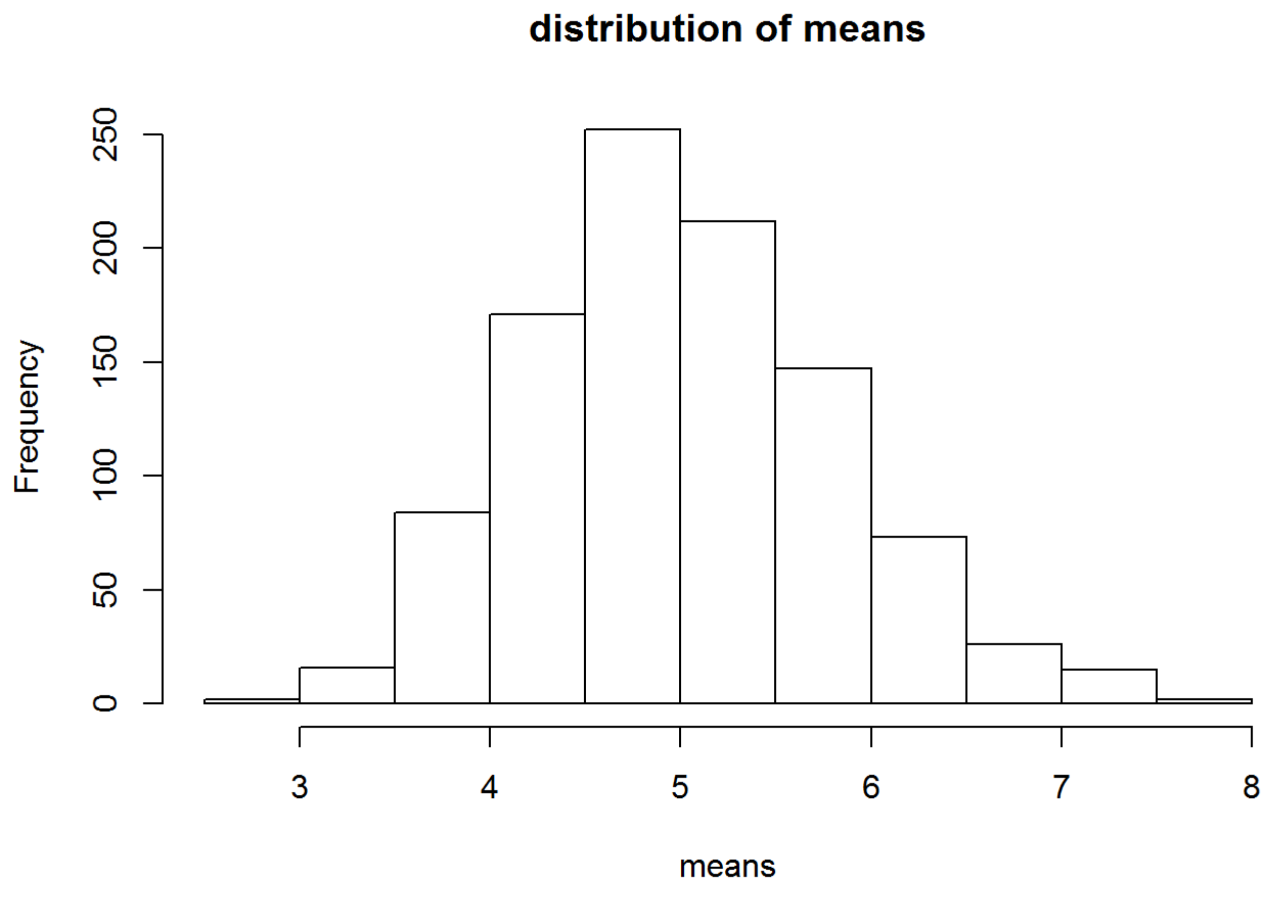
```
#We will create 40 random exponentials and plot the histogram.
```

Now, to demonstrate CLT and LLN we will have to obtain the distribution of the averages of 40 exponentials.

```
means=NULL  #The means vector will contain the mean for each simulation

#The for loop for the simulations
for (i in 1:1000){
  means=c(means,mean(rexp(40,lambda)))
  #At each step, we append the mean of 40 exponentials to the means vector
}
```

Now we will plot the histogram for the distribution for the averages of 40 exponentials.



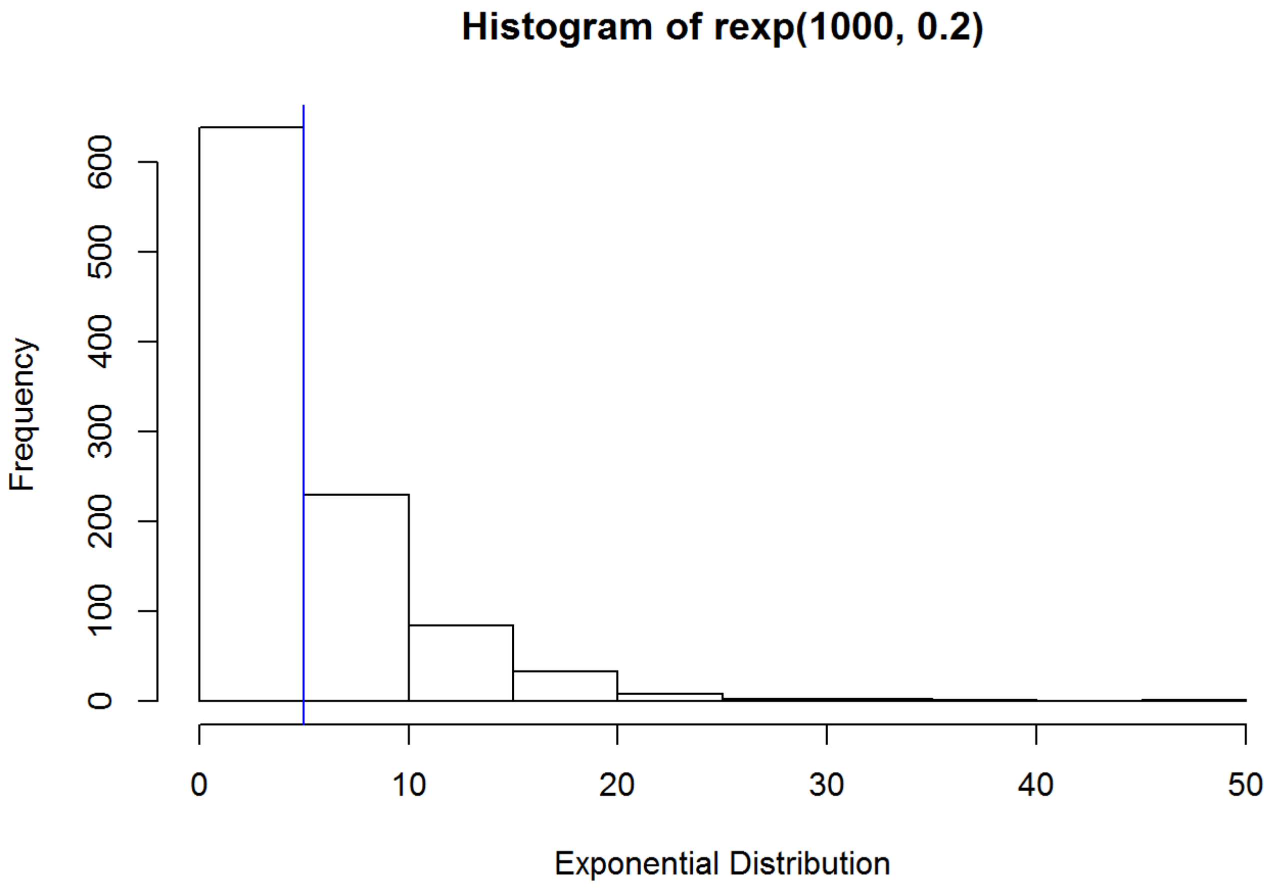
“Sample mean vs theoretical mean”

The parameter lambda for the exponential distribution was set to be equal to 0.2. The mean of the exponential distribution is equal to 1/lambda.THis is the theoretical mean.

```
theoretical_mean<- 1/lambda
theoretical_mean
```

```
## [1] 5
```

The following is the plot of our exponential distribution with the theoretical mean shown in blue.

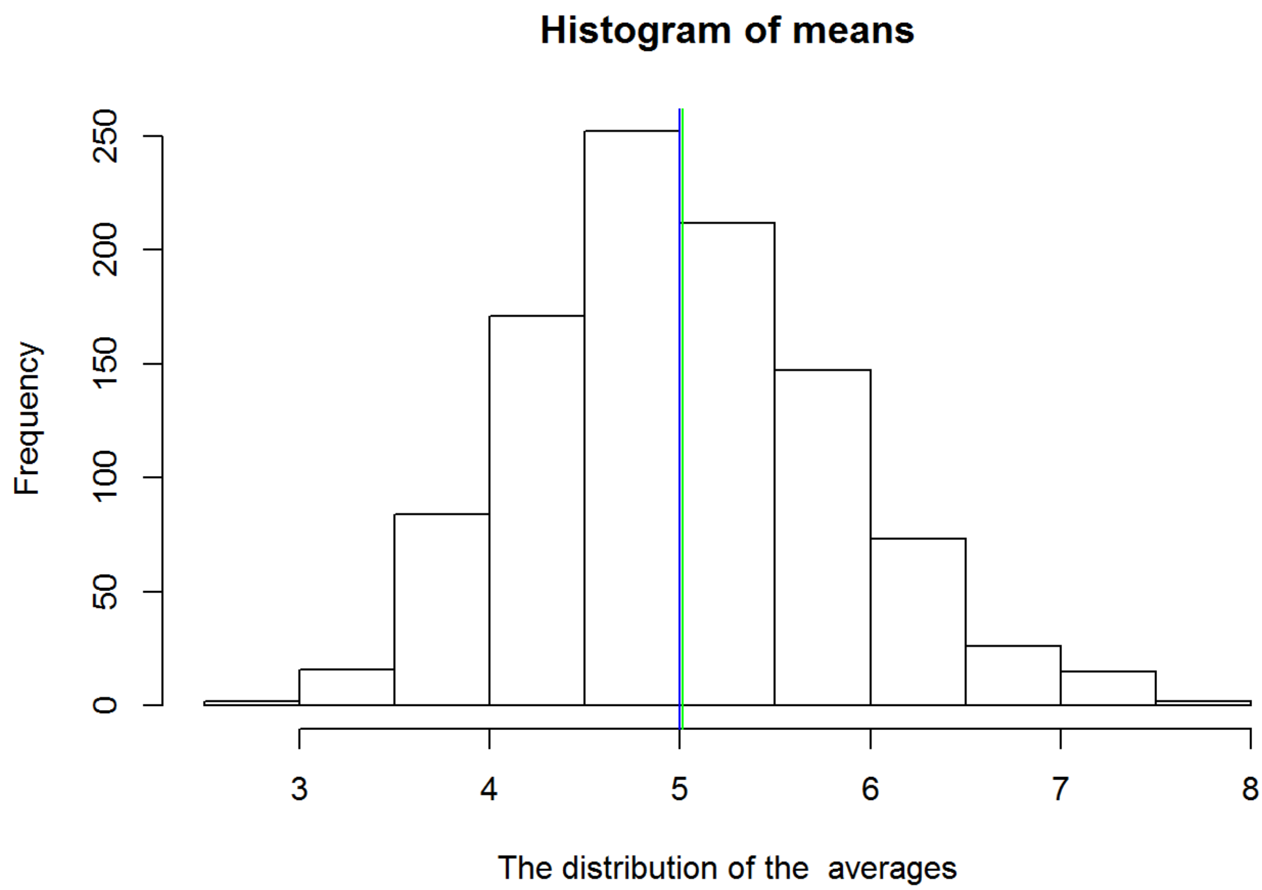


Our sample for the distribution of the averages is the ‘means’ vector. So, the sample mean is the mean of the ’means’vector.

```
sample_mean<-mean(means)
sample_mean
```

```
## [1] 5.013364
```

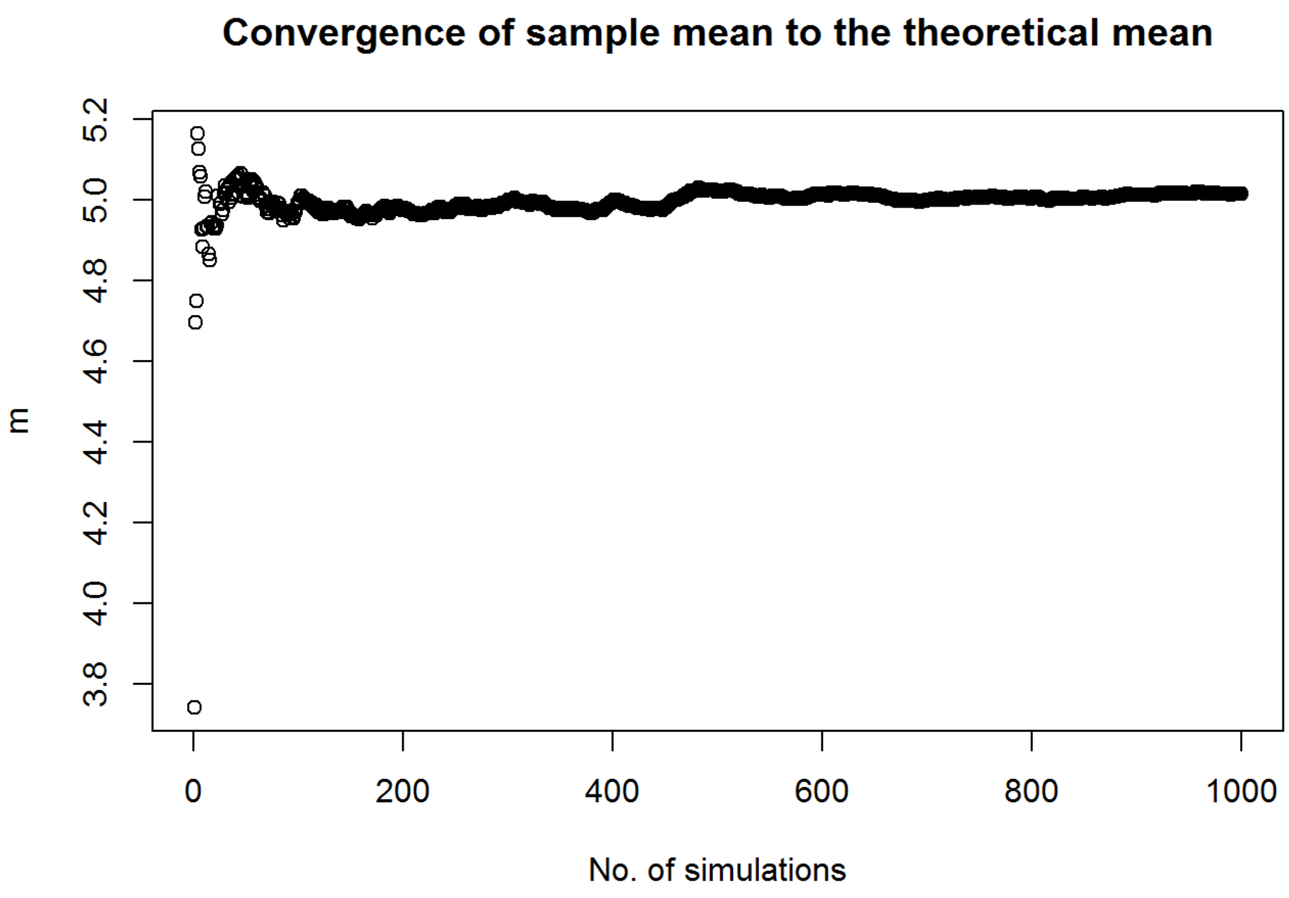
The following is the plot for the distribution of the averages of 40 exponentials with the sample mean shown in green and the theoretical mean shown in green.



It is clear that the sample mean and the theoretical mean are pretty close. This is an example of the Law of Large Numbers(LLN) which essentially says that the average of a distribution of averages tends to become equal to the value it is estimating as the number of samples becomes large.

We can make an interesting plot to demonstrate the LLN. we plot the sample mean as a function of the number of simulations and observe how it converges to the theoretical mean.

```
m=NULL
for (i in 1:1000){
  m=c(m,mean(means[1:i]))
}
x=1:1000
plot(x,m,main="Convergence of sample mean to the theoretical mean",xlab="No. of simulat
ions")
```



---

title: “Sample variance vs Theoretical Variance”

The standard deviation and the mean of the exponential distribution are equal and given by 1/lambda.This is the theoretical variance.

```
theoretical_variance=(1 /lambda)^2
theoretical_variance
```

```
## [1] 25
```

The sample variance is the variance of the data in the means vector.

```
sample_variance<-var(means)
sample_variance
```

```
## [1] 0.6797081
```

By the CLT(Central Limit Theorem) the mean of large number of iid random variables is a normal random variable with mean equal to the mean of the random variable while the variance is equal to the original variance divided by the number of random variables(40 in our case). The following calculations verify the result .

```
sample_variance
```

```
## [1] 0.6797081
```

```
theoretical_variance/40
```

```
## [1] 0.625
```

```
# Both the above numbers are close will get closer if we go on increasing the number of
random variables used to calculate the mean.
```

---

title:“Comparing the two distributions”

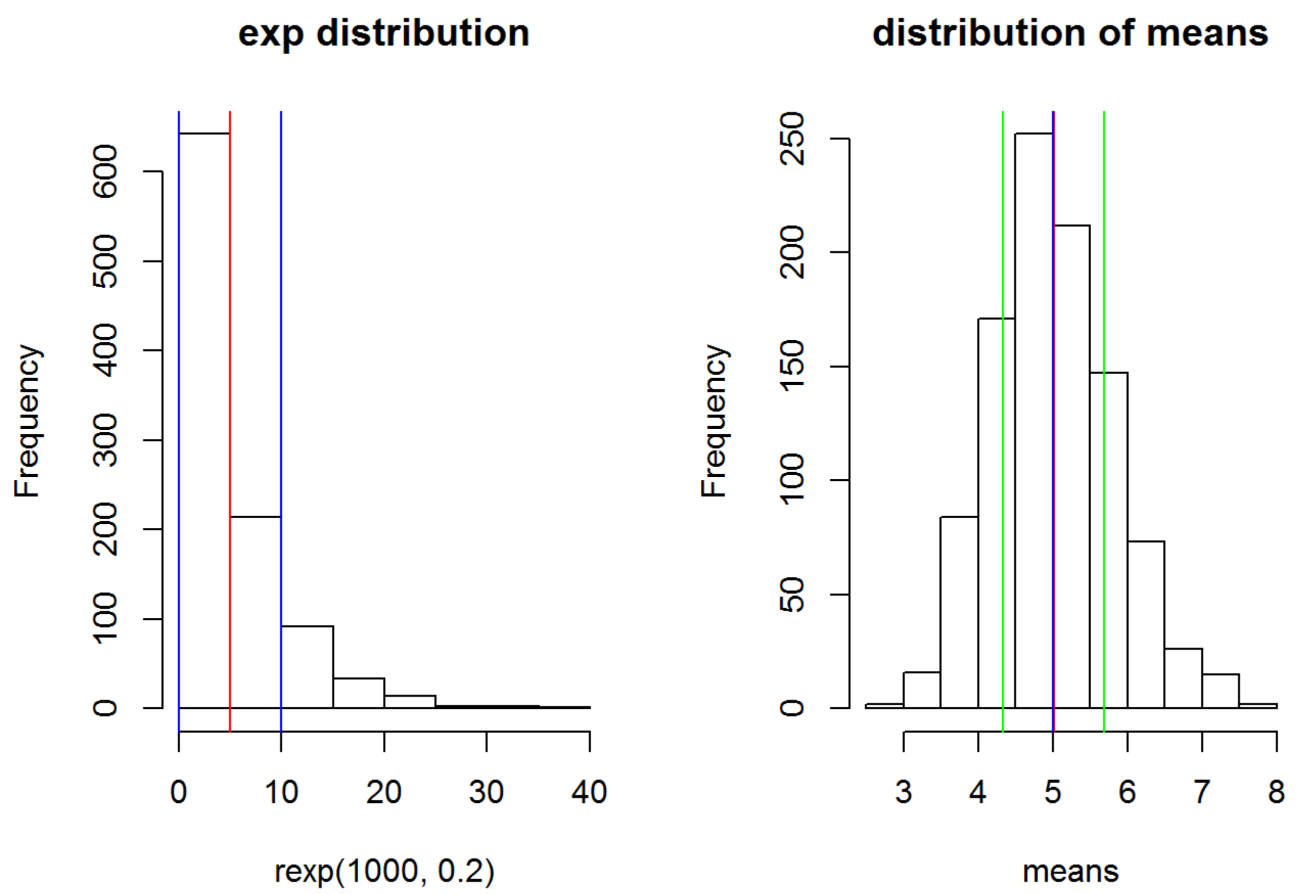
We will now compare the two distributions using direct plots. The red line shows the sample mean and the blue one shows the theoretical mean.

```
## Warning in plot.window(xlim, ylim, "", ...): "binwidth" is not a graphical
## parameter
```

```
## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "binwidth" is not a graphical parameter
```

```
## Warning in axis(1, ...): "binwidth" is not a graphical parameter
```

```
## Warning in axis(2, ...): "binwidth" is not a graphical parameter
```



In the above plot for the distribution of the means, the green lines represent distances of one standard deviation form the sample mean. The blue line represents the sample mean and the red line represents the theoretical mean.

Similarly, in the histogram for the exponential distribution, the red line is the theoretical mean and the two blue lines represent the interval contained within one standard deviation from the mean.

As seen above, clearly the histogram for the distribution of averages looks very Gaussian(normal) with a peak which is very near to the sample mean.