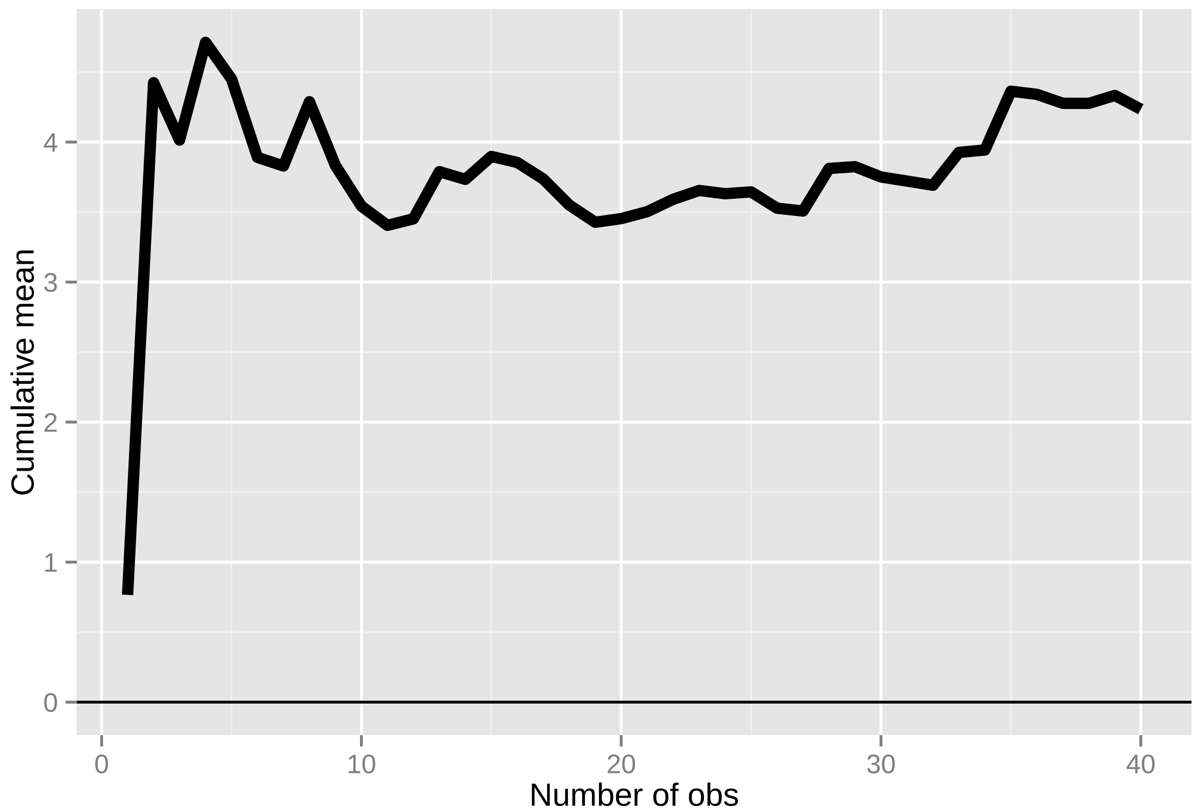
Statistical Inference Project

Pramit Choudhary

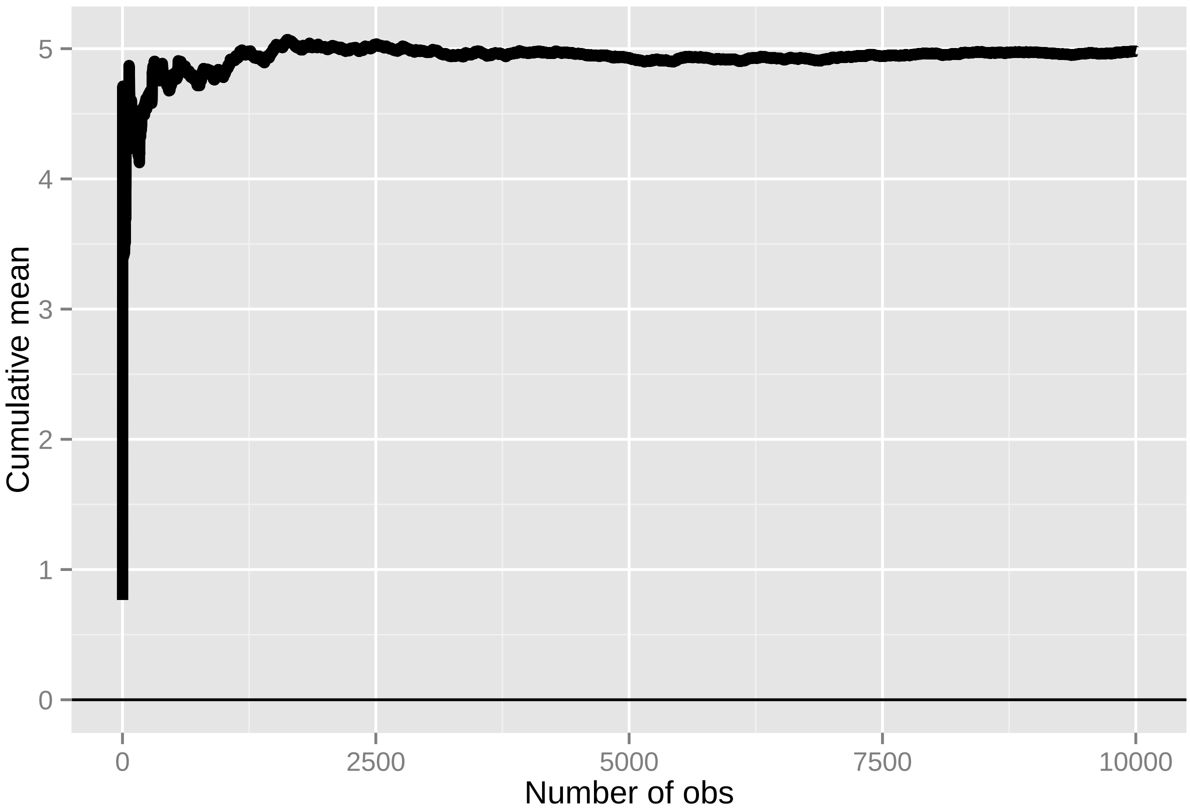
Study the random distribution of an exponential function

- Q1. Show the sample mean and compare it to the theoretical mean of the distribution. Computation of Sample Mean

```
set.seed(23)
n <- 40; #sample size of 40
means <- cumsum(rexp(n,0.2)) / (1  : n); library(ggplot2)
g <- ggplot(data.frame(x = 1 : n, y = means), aes(x = x, y = y))
g <- g + geom_hline(yintercept = 0) + geom_line(size = 2)
g <- g + labs(x = "Number of obs", y = "Cumulative mean")
g
```




```
set.seed(23)
# Sample size of 1000
n <- 10000; #sample size of 1000
means <- cumsum(rexp(n,0.2)) / (1  : n); library(ggplot2)
g <- ggplot(data.frame(x = 1 : n, y = means), aes(x = x, y = y))
g <- g + geom_hline(yintercept = 0) + geom_line(size = 2)
g <- g + labs(x = "Number of obs", y = "Cumulative mean")
g
```



The theoretical mean of the distribution, as stated in the question, is 1/0.2 = 5 and the sample mean for the above 2 examples is mean(rexp(n,0.2)) = 4.807

- Q2.Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
#Sample variance:
mns = NULL
for (i in 1 : 1000)   mns = c(mns, mean(rexp(40,0.2)))
var(mns)
```

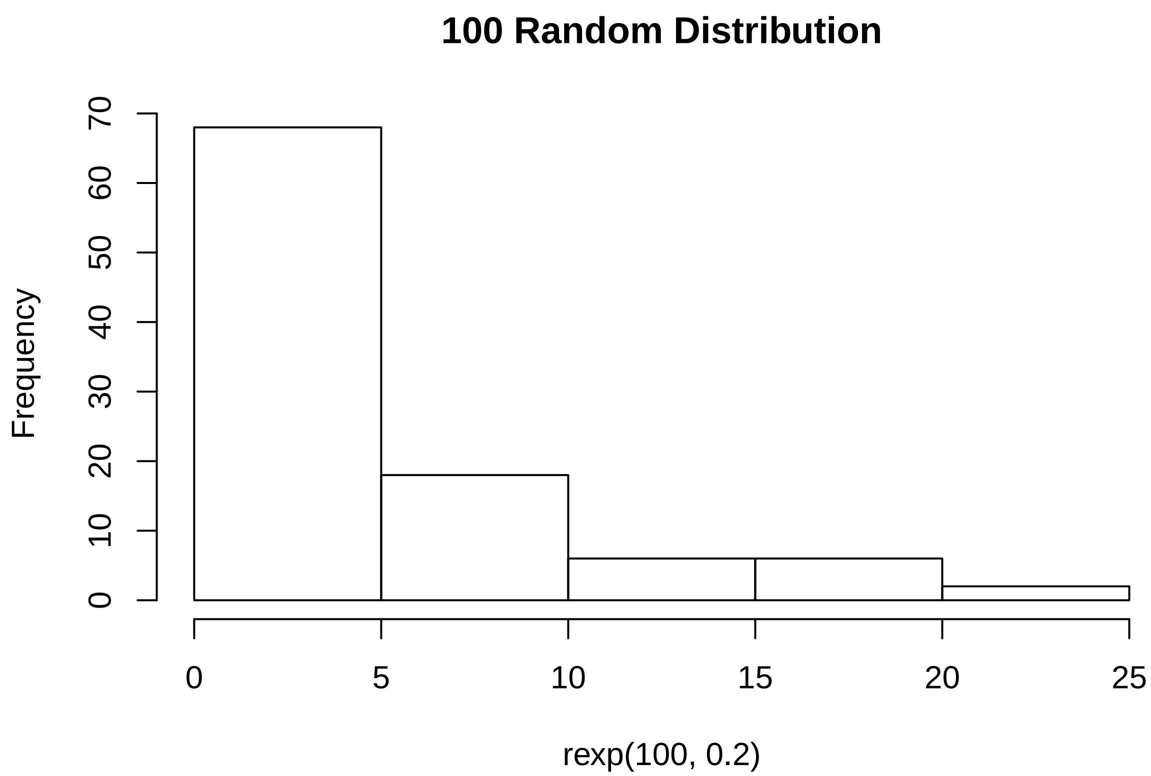
```
## [1] 0.6465578
```

Variation in the sample variance is shown in the above computation and theorectical variance can be computed as SD^2/n = 5^2/40 = 0.625

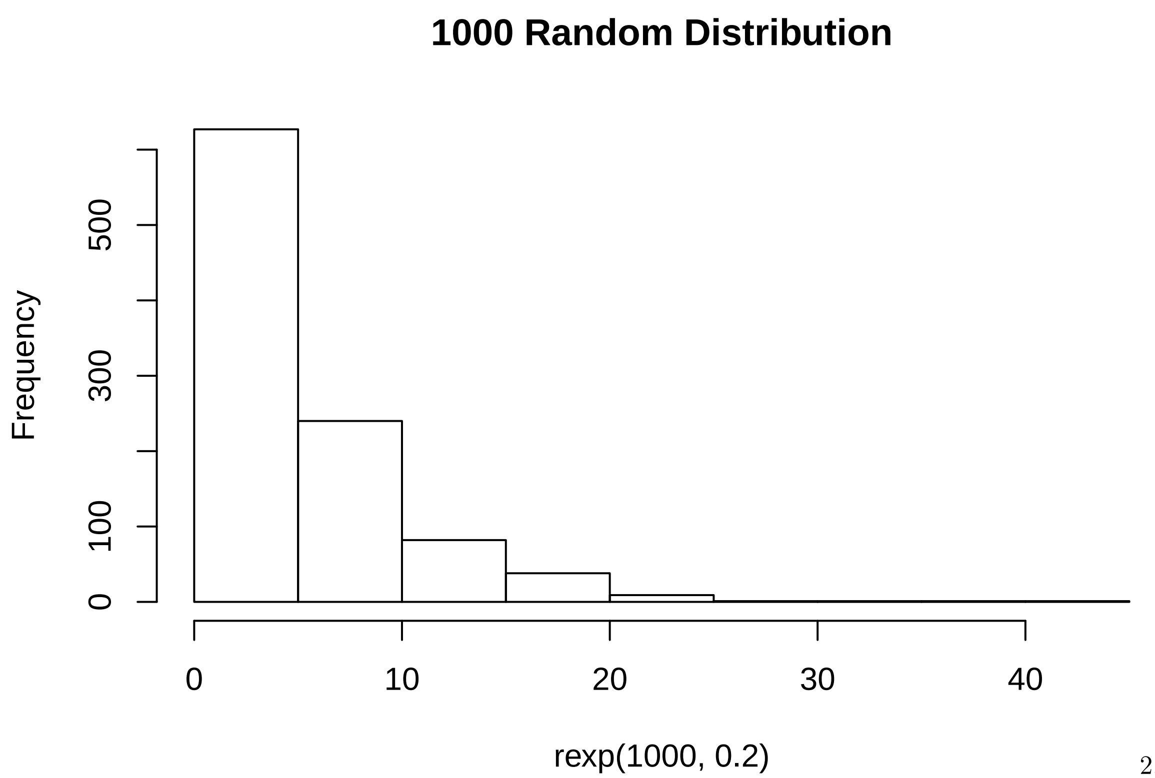
- Q3. Show that the distribution is approximately normal.

1. Random exponential distribution

```
hist(rexp(100,0.2), main="100 Random Distribution")
```

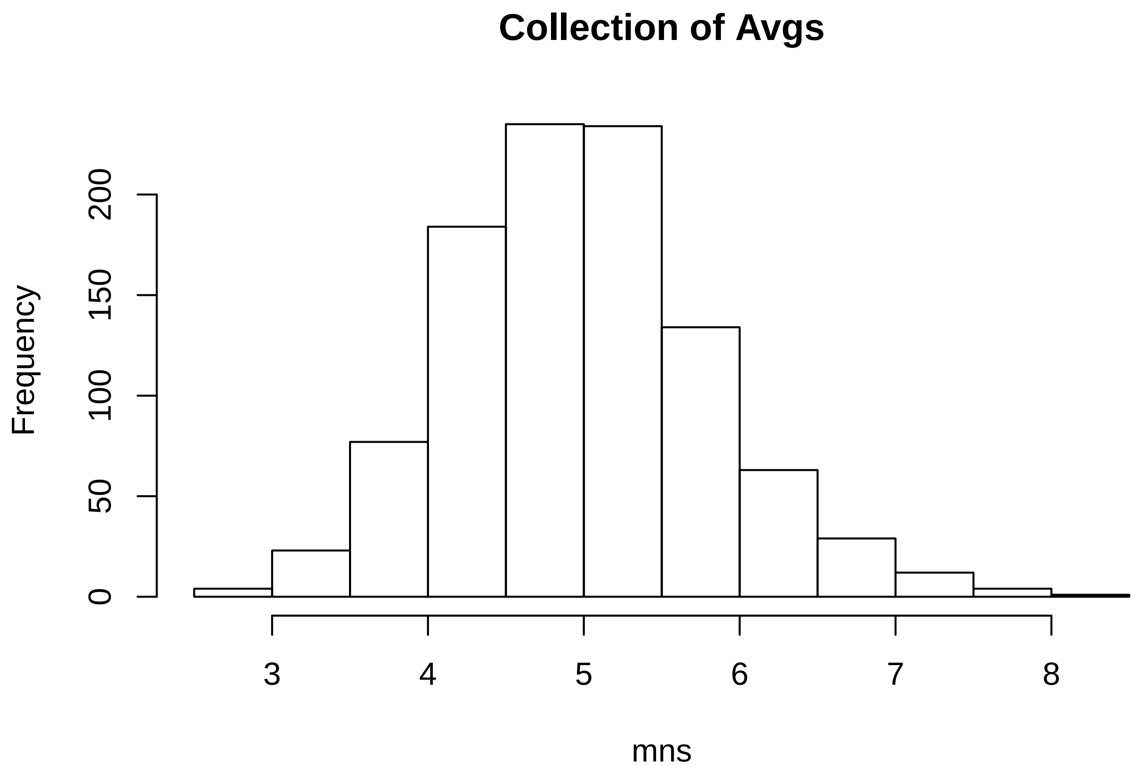



```
hist(rexp(1000,0.2), main="1000 Random Distribution")
```



2.Distribution of a large collection of averages of 40 exponentials.

```r
mns = NULL
for (i in 1 : 1000)
  mns = c(mns, mean(rexp(40,0.2)))
hist(mns, main="Collection of Avgs")
```



Observation: As one can see in the random distribution of the exponential function, the frequency of the values generated between 0-10 is pretty high compared to other values. If one increases the size of 'n' in the random generation, say 100000, the values starts to congregate more between 0-5. This behavior follows the universal Power Law which states that such a functional behavior is seen when the two related quantity vary as a power of another.

In-regards, to the collection of averages of 40 such samples. The distribution starts to follow the Central Limit Theorem. CLT states that as 'n' gets larger the distribution of the difference between sample average and its expected value when multiplied by sqrt(n), follows a normal distrbution with mean 'mu' and variance 'sigma^2/n'. As per the current use-case, mu is 1/0.2=5 and variance 5^2/40 = approx 0.5.