Sampling Distribution

Aanish Singla

1 April 2016

Overview: In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution is simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. We will investigate the distribution of averages of 40 exponentials. Note that we will do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

Part 1:

Now look at the Theoretical mean. Mean = 1/lambda, which in this case is $1/.2 = 5$

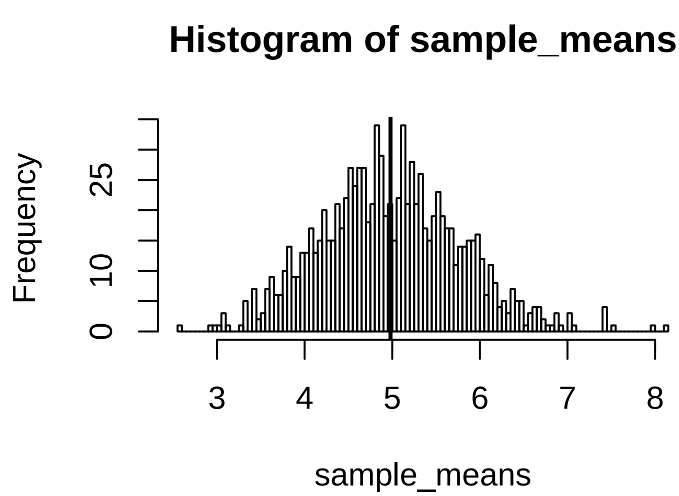
Now lets generate the distribution of means of 1000 samples of exp distribution of size 40 and lambda - 0.2

```
sample_means = NULL
  for (i in 1 : 1000)
    sample_means = c(sample_means, mean(rexp(40,0.2)))
mean(sample_means)
```

```
## [1] 4.980309
```

Lets see this visually

```
hist(sample_means, breaks=100)
  abline(v=mean(sample_means), lwd=2)
```



Pretty close to the theoretical value of the mean!

Part 2:

Now lets simulate the variance of the means. Theoretical Standard error of the mean = sigma/sqrt(n), which is 5/sqrt(1000)

```
5/sqrt(1000)
```

```
## [1] 0.1581139
```

Since variance is square of standard error of mean, Variance is

```
(5/sqrt(1000))^2
```

```
## [1] 0.025
```

Now lets calculate the variance of the sample that we have:

```
Var <- (sd(sample_means)/sqrt(1000))
Var
```

```
## [1] 0.02534528
```

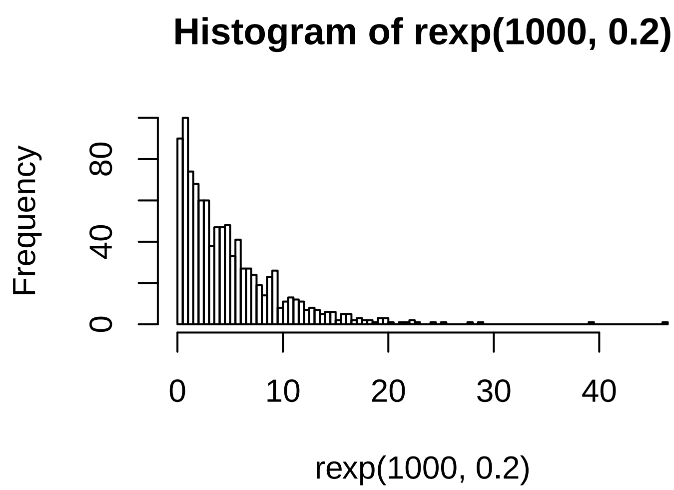
Pretty close to the theoretical value of the variance!

Part 3:

Lets see if our results comply with central limit theorem

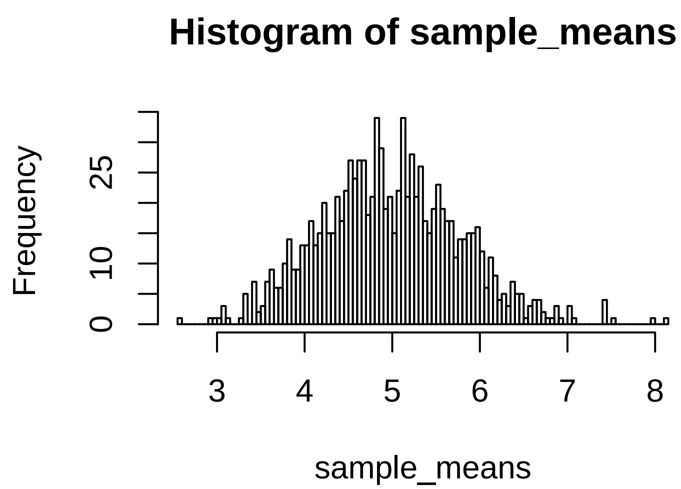
Distribution of large collection (1000) of random exponentials looks as follows

```
hist(rexp(1000,0.2), breaks=100)
```



Now lets plot the distribution of sample means of 40 exponentials

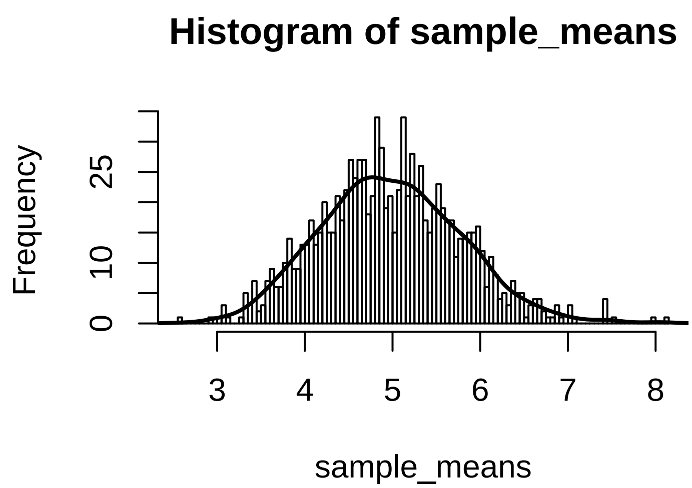
```
myhist <- hist(sample_means, breaks=100)
```



Overlaying a normal curve:

```
multiplier <- myhist$counts / myhist$density
mydensity <- density(sample_means)
mydensity$y <- mydensity$y * multiplier[1]

plot(myhist)
lines(mydensity, lwd=2)
```



Viola! central limit theorem works :-)