Statistical Inference Project, Part 1

Hoang Tam Vo

In this part of the project, I study the exponential distribution in R and compare it with the Central Limit Theorem. Specifically, in R, the exponential distribution can be simulated with `rexp(n, lambda)` where `lambda` is the rate parameter. I illustrates via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s.

Simulation

In this study, `lambda` is set to 0.2 for all of simulations. A thousand of simulations are done and the distribution of averages of 40 exponentials are investigated.

```
lambda = 0.2
n = 40
nsms = 1:1000
set.seed(26042015)
means <- data.frame(x = sapply(nsms, function(x) {mean(rexp(n, lambda))}))
head(means)
```

```
##       x
## 1 4.599
## 2 4.149
## 3 6.244
## 4 4.772
## 5 4.282
## 6 6.434
```

Questions 1: Sample Mean versus Theoretical Mean

The average sample mean of 1000 simulations of 40 randomly sampled exponential distributions:

```
mean(means$x)
```

```
## [1] 5.002
```

The expected mean of an exponential distribution of rate `lambda` is:

```
expectedMean <- 1/lambda
expectedMean
```

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

As can be seen, the expected mean and the avarage sample mean are very close.

Questions 2: Sample Variance versus Theoretical Variance

The standard deviation of the average sample mean of 1000 simulations of 40 randomly sampled exponential distribution:

```
sd(means$x)
```

```
## [1] 0.781
```

The variance of the average sample mean of 1000 simulations of 40 randomly sampled exponential distribution:

```
var(means$x)
```

```
## [1] 0.61
```

The expected standard deviation of an exponential distribution of rate `lambda` is:

```
(1/lambda)/sqrt(n)
```

```
## [1] 0.7906
```

The expected variance of an exponential distribution of rate `lambda` is:

```
((1/lambda)/sqrt(n))^2
```

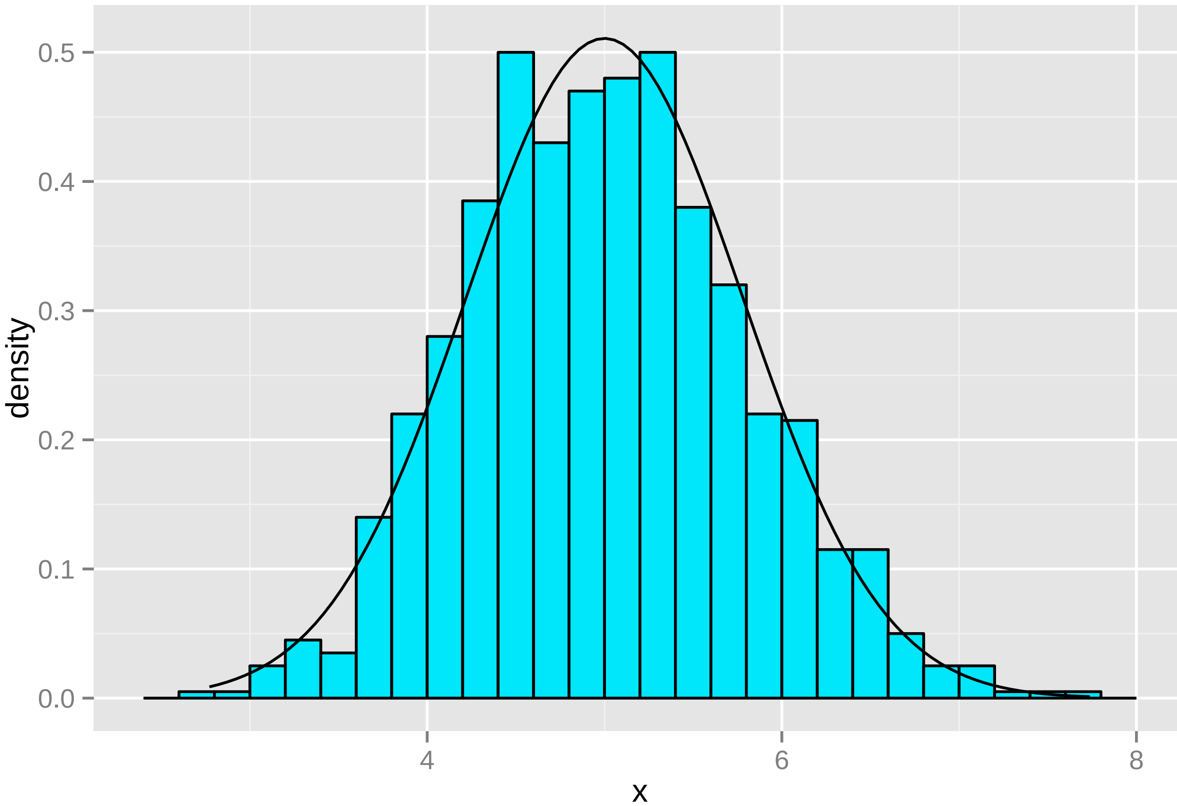
```
## [1] 0.625
```

As can be seen, the standard deviations are very close, while variances are also pretty close as variance is the square of the standard deviation.

Question 3: Show that the distribution is approximately normal

A histogram plot of the means of the 1000 simulations of rexp(n, lambda) is shown in the below figure. In addtion, it is overlaid with a normal distribution with mean 5 and standard deviation 0.781.

```
library(ggplot2)
ggplot(data = means, aes(x = x)) +
    geom_histogram(aes(y=..density..), fill = I('#00e6fa'),
                   binwidth = 0.20, color = I('black')) +
    stat_function(fun = dnorm, arg = list(mean = 5, sd = sd(means$x)))
```



As can be seen, the distribution of our simulations is approximately normal.