# Statistical Inference Course Project: Part 1

*Cheng-Han Yu*

*July 16, 2015*

## Simulation study

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with `rexp(n, lambda)` where lambda ($\lambda$) 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. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

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

**1. Show the sample mean and compare it to the theoretical mean of the distribution.**

***Solution***

```
lambda <- 0.2
n <- 40
nsim <- 1000
set.seed(123)
sim_sample <- matrix(rexp(n*nsim, rate = lambda), nrow = nsim, ncol = n)
sim_sample_mean <- apply(sim_sample, 1, mean)
mean(sim_sample)
```

```
## [1] 5.011911
```

```
(mu <- 1 / lambda)
```

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

We create a simulation sample called `sim_sample` each of size 40 drawn from exponential distribution with rate parameter $\lambda = 2$. Seed is set at 123 to make the simulation reproducible. Object `sim_sample_mean` is the sample of sample mean ($\bar{X}$) from the simulation sample. The sample mean of $\bar{X}$ is 5.0119, which is pretty close to the theoretical mean, $\mu = 1/\lambda = 5$. The estimated and the theoretical means correspond to the green and red vertical lines in the figure below.

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

***Solution***

```
var(sim_sample_mean)
```

```
## [1] 0.6088292
```

```
(sig2 <- (1 / lambda ^ 2) / n )
```

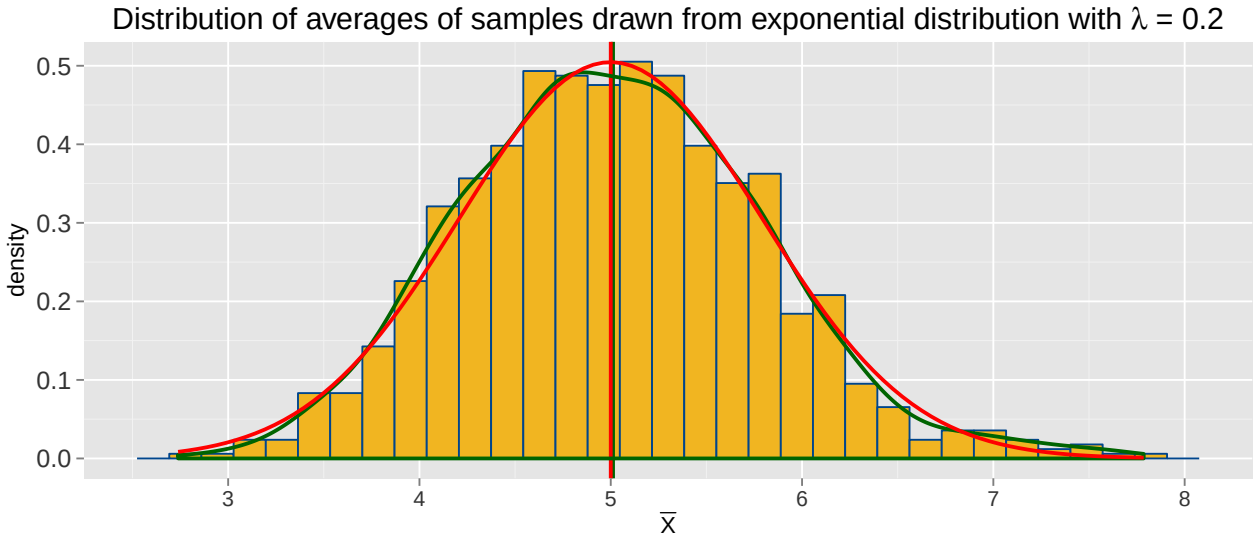
```
## [1] 0.625
```

The sample variance of $\bar{X}$ is 0.6088, which is also close the theoretical variance, $\sigma^2 = \frac{1/\lambda^2}{n} = 0.625$.

**3. Show that the distribution is approximately normal.**

***Solution***

```
library(ggplot2)
ggplot(data.frame(sim_sample_mean), aes (x = sim_sample_mean)) +
    geom_histogram(aes(y = ..density..), colour = "#00458c", fill = "#F1B521") +
    geom_vline(xintercept = c(mean(sim_sample_mean), 5), size = c(1,1),
               colour = c("darkgreen", "red")) +
    xlab(expression(bar(X))) + geom_density(colour = "darkgreen", size = 1) +
    ggtitle(expression(paste("Distribution of averages of samples drawn from exponential distribution w
    theme(axis.text.x = element_text(colour = "grey20", size = 12),
          axis.text.y = element_text(colour = "grey20", size = 14),
          axis.title.x = element_text(size = 13, hjust = .5),
          axis.title.y = element_text(size = 13, hjust = .5, vjust = 1),
          title = element_text(size = 14, hjust = .5)) +
    scale_x_continuous(breaks = c(seq(0, 8, 1))) +
    stat_function(fun = dnorm, arg = list(mean = 5, sd = sqrt(sig2)),
                  colour = "red", size = 1)
```




The above figure shows the distribution of the sample mean. It is approximately normal. The red density curve corresponds to $N(5, 0.625)$ density and the green one is the estimated density from the sample. This example validates the Centrai Limit Theorem.