

Project Title: Investigation of exponential distribution using central limit theorem (CLT).

Author name: Samiul Azam

Overview:

In this project, my target is to investigate the exponential distribution and the central limit theorem (CLT) using simulation of Rstudio. The simulation will show that the theoretical statistics (mean and variance) of exponential distribution can be achieved approximately by drawing 40 random exponential values for 1000 times and making a distribution of means of these samples. Moreover, using the theory of CLT, I have shown that the distribution is approximately normal.

Description of the simulation:

In one step, I generate 40 random exponential values for lambda 0.2, and calculate mean and standard deviation of these 40 values. I repeat this step for 1000 times and store all 1000 mean and standard deviation values in a linear matrix. Later, I utilize these matrices to generate necessary distributions (Figure 1 and Figure 2).

I also show the distribution of exponential values using 1000 randomly generated exponential values (Figure 3).

The R code for the simulation is provided in the Appendix section.

Comparison of Sample mean and the theoretical mean:

In Figure 1, I show the distribution of means of 40 exponentials (after 1000 simulations). The yellow vertical line is the intercept at the sample mean. At the top of the graph, I also provided the theoretical mean (5.0) and standard error (0.7906), as well as sample mean (4.992) and standard error (0.7951). We can see that the simulated sample mean and standard error are very close to the equivalent theoretical values.

Comparison of Sample variance and the theoretical variance:

In Figure 2, I show the distribution of std-devs of 40 exponentials (after 1000 simulations). The red vertical line is the intercept at the sample std-dev (mean of all std-devs). At the top of the graph, I also provided the theoretical std-dev (5.0) and sample std-dev (4.8691). We can see that the simulated sample std-dev is very close to the equivalent theoretical value.

Distribution of a large collection of random exponentials:

In Figure 3, I show the distribution of 1000 randomly generated exponential values. Clearly it shows the property of an exponential curve. The mean of these values is shown as a red vertical line in the Figure 3. The mean is 5.1106 which is very close to the theoretical mean 5.0. Most of the values are around the average value. However, few values are far away from the average. Figure 1 shows the distribution of mean of 40 exponential values where the average is 4.992 and shows approximately normal distribution shape (due to central limit theorem).

Justification of normal distribution:

Over 1000 simulation all the mean values are actually an iid random variable. According to central limit theorem, distribution of this iid variable becomes normal distribution when n is sufficiently large. Moreover the distribution is approximately $N(\text{theoretical mean}, \text{theoretical SE})$. Refer to Figure 1, we see that the smooth version of the distribution (black curve) is very close to normal distribution which has approximately same statistics as theoretical statistics. In short, the distribution of means of 40 exponentials behave as predicted by the central limit theorem.

Appendix

R code:

```
library(ggplot2)
lambda <- 0.2
num_obs <- 40

# Theoretical mean, std-dev and SE of exp distribution
theo_mean <- round(1/lambda,4)
theo_stddev <- round(1/lambda,4)
theo_SE <- round((1/lambda)/sqrt(num_obs),4)

# Initialization of simulation
sim_mean = NULL
sim_std = NULL
num_of_simulation = 1000

# Simulation iterations
for (i in 1 : num_of_simulation)
{
  val <- rexp(num_obs, lambda)
  sim_mean = c(sim_mean, mean(val))
  sim_std = c(sim_std, sd(val))
}

# Exploratory statistics of distribution of the 1000 means
mean_of_means = mean(sim_mean)
SE_of_means = sd(sim_mean)

# Plotting of distribution of the 1000 means
png("g1.png", width=700, height=480)
g <- ggplot(data.frame(sim_mean), aes(x = sim_mean))
g <- g + geom_histogram(fill = "salmon", binwidth = 0.1, aes(y = ..density..), color = "black")
g <- g + geom_density(size = 1.5)
g <- g + geom_vline(xintercept = mean_of_means, size = 1, color = "yellow")
g <- g + labs(title = paste('Sample mean = ', round(mean_of_means,4), ', Theoretical mean = ', theo_mean, ', Sample SE = ', round(SE_of_means,4), ', Theoretical SE = ', theo_SE))
g <- g + xlab("Distribution of means of 40 exponentials (after 1000 simulations)")
g <- g + ylab("Density")
print(g)
dev.off()

# Exploratory statistics of distribution of the 1000 std-devs
mean_of_stds = mean(sim_std)

# Plotting of distribution of the 1000 std-devs
png("g2.png", width=700, height=480)
g <- ggplot(data.frame(sim_std), aes(x = sim_std))
g <- g + geom_histogram(fill = "gray", binwidth = 0.2, aes(y = ..density..), color = "black")
g <- g + geom_density(size = 1.5)
g <- g + geom_vline(xintercept = mean_of_stds, size = 1, color = "red")
g <- g + labs(title = paste('Sample std-dev (average of all std-dev) = ', round(mean_of_stds,4), ', Theoretical std-dev = ', theo_stddev))
```

```

g <- g + xlab("Distribution of standard-deviations of 40 exponentials (after 1000 simulations)")
g <- g + ylab("Density")
print(g)
dev.off()

# Values of 1000 random exponentials
sim_exp = rexp(1000, lambda)
mean_of_exp = mean(sim_exp)

# Plotting of distribution of 1000 random exponential values
png("g3.png", width=700, height=480)
g <- ggplot(data.frame(sim_exp), aes(x = sim_exp))
g <- g + geom_histogram(fill = "cyan", binwidth = 0.5, aes(y = ..density..), color = "black")
g <- g + geom_density(size = 1)
g <- g + geom_vline(xintercept = mean_of_exp, size = 1, color = "red")
g <- g + labs(title = paste('Mean = ', round(mean_of_exp,4)))
g <- g + xlab("Distribution of a 1000 random exponentials values")
g <- g + ylab("Density")
print(g)
dev.off()

```

Figure 1: Distribution of means of 40 exponentials (after 1000 simulations).

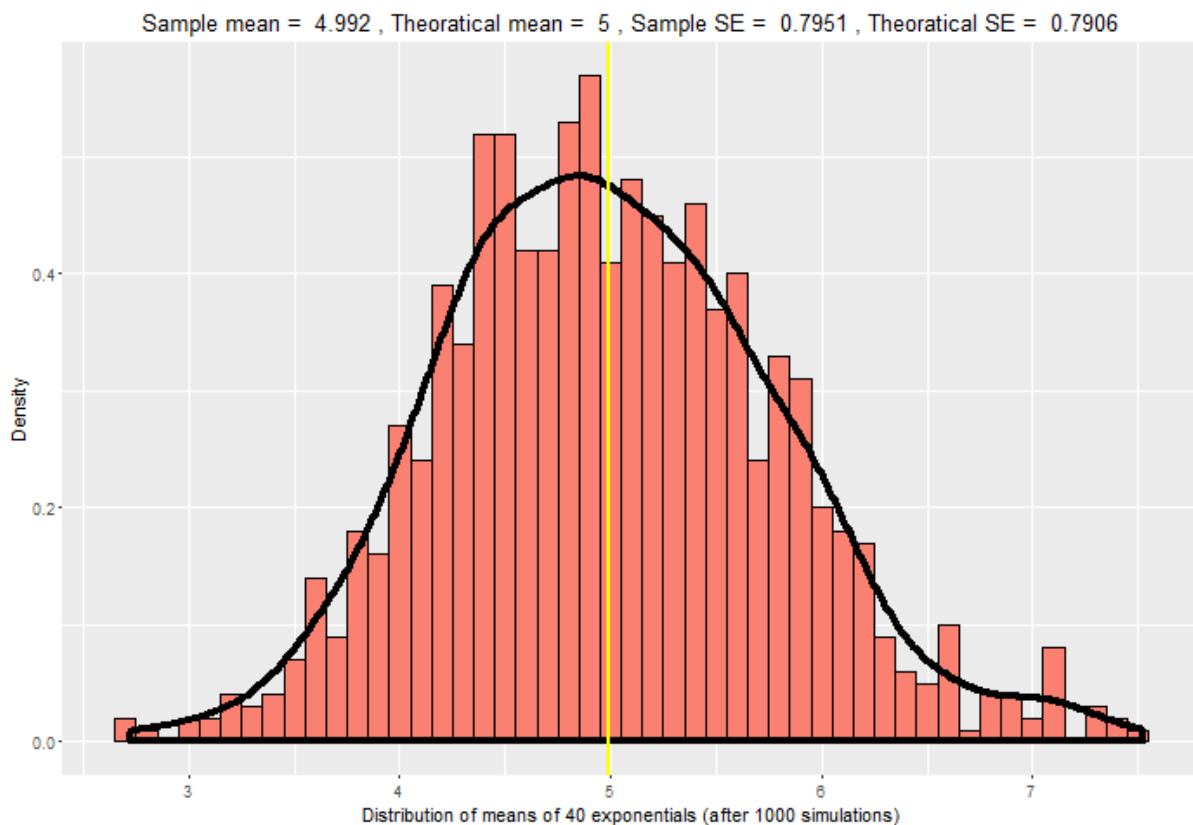


Figure 2: Distribution of standard deviations of 40 exponentials (after 1000 simulations).

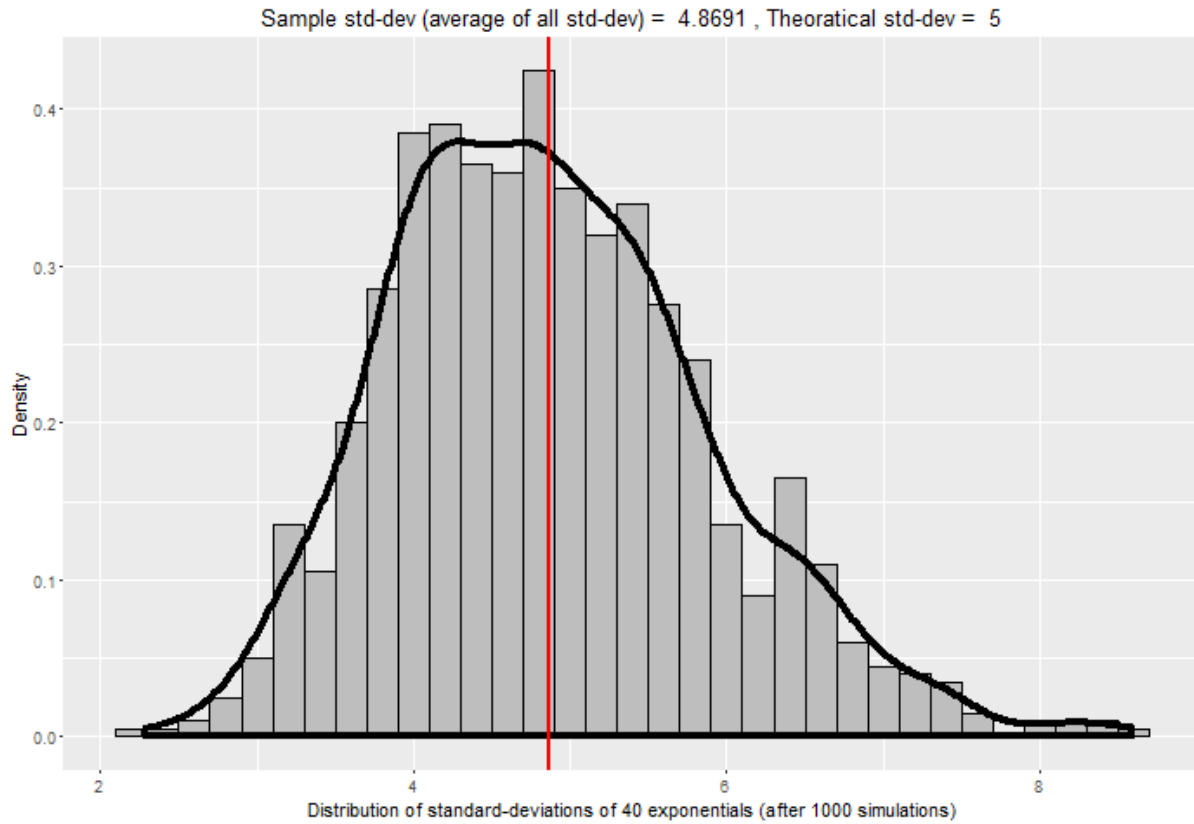


Figure 3: Distribution of 1000 random exponential values.

