# Programming Assigment 1-1

# Central Limit Theorem Applied to Exponential Distribution

## Author: Zdravka Cankova

---

### Overview

The purpose of this project was to examine the distribution of averages of 40 exponential distributions, and compare it to the Central Limit Theorem (CLT). This theorem states that when sampling probability distributions, as the sample size increases, the resuling distribution of the sample averages will be approximately normal, regardless of the type of distribution we are sampling from. Thus, we expected to find that the distribution of averages of 40 exponentials would be approximately normal.

### Data Simulations

We collected 40 samples (n = 40) from an exponential distribution with lambda = 0.2 and calculated the average. We repeated this process 1000 times to obtain a distribution of the average value of 40 exponentials (result assigned to variable *means*). Note that the seed was set to 1 to ensure the presented results are reproducible.

```
n <- 40
lambda <- 0.2

set.seed(1)
means = NULL
for (i in 1 : 1000) {means = c(means, mean(rexp(n, lambda)))}
```

The observed mean was calculated as the average value of *means*. The theoretical mean was calculated as 1/lambda.

```
obs_mean <- mean(means)
theor_mean <- 1/lambda
```

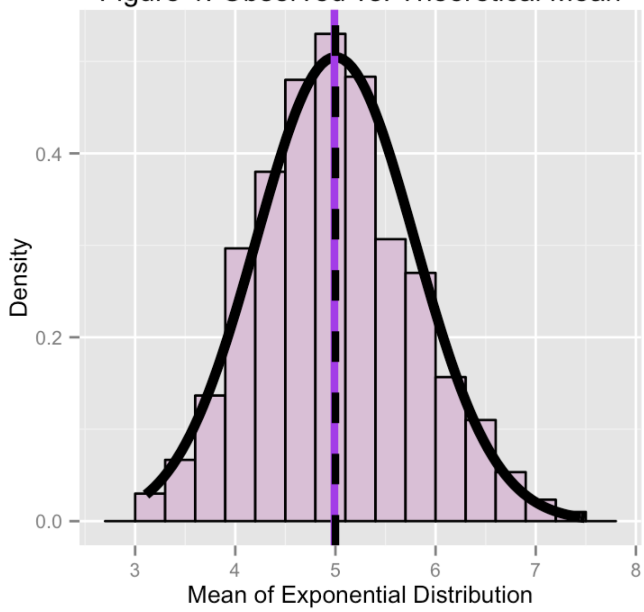
The observed population variance was calculated from the squared standard error (sd(means)^2) and the sample size n using the equation sigma^2 = n*StErr^2. The theoretical variance was calculated as (1/lambda)^2.

```
obs_var <- n*sd(means)^2
theor_var <- (1/lambda)^2
```

### Observed Mean vs. Theoretical Mean

Figure 1 shows a histogram of *means* with a purple line indicating the observed mean value of this distribution. Overlayed is the corresponding normal distribution, according to the CLT, with a black line designating the theoretical mean. The code used to generate this figure is shown in Appendix A.

Figure 1: Observed vs. Theoretical Mean

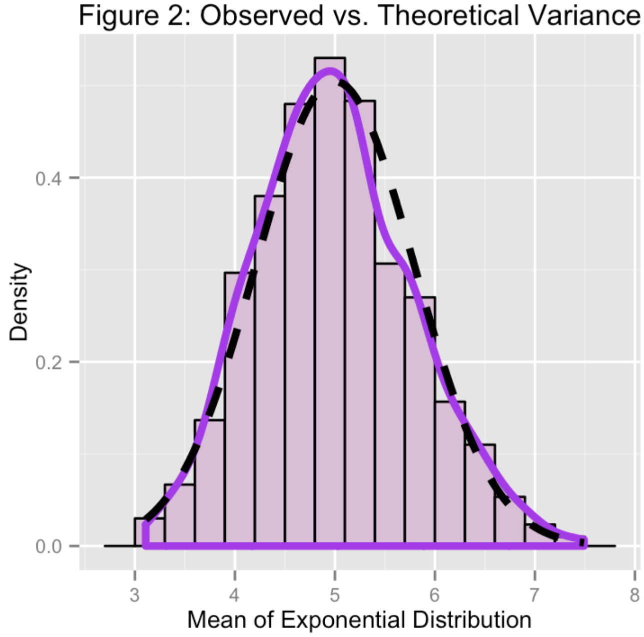



As you can see from the figure, the purple and black lines overlap closely, indicating that the **observed mean of 4.9900252** approaches the **theoretical mean of 5**, as predicted by the CLT.

## Observed Variance vs. Theoretical Variance

Figure 2 below again shows the histogram of averages of 40 exponentials, this time with a purple line outlining the observed distribution density. The corresponding normal distribution is shown in black. The code used to generate this figure is shown in Appendix B.

Figure 2: Observed vs. Theoretical Variance

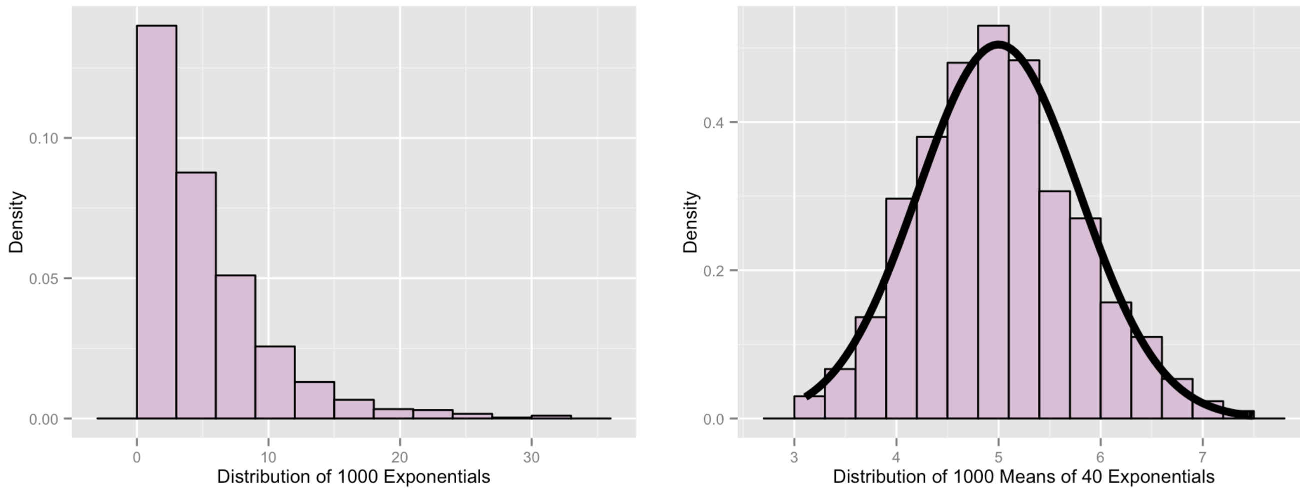



As we can see in the figure, the observed distribution shape (purple line) is the approximately the same height and width as the normal distribution (dashed black line). This indicates that the **observed variance of 24.4446587** is very similar to the **theoretical variance of 25**.

## The Distribution of the Mean of 40 Exponentials Is Approximately Normal

It is important to note the difference between the distribution of a large number (1000) samples drawn from an exponential distribution and the distribution of a large number of averages of 40 random samples from an exponential distribution. This is illustrated in Figure 3 (code in Appendix C).

Figure 3: Distribution of 1000 Exponentials vs. Distribution of 1000 Averages of 40 Exponentials




The left panel shows a histogram of 1000 random draws from an exponential distribution. As expected, this histogram looks like an exponential curve. However, let us instead take 40 random draws from an exponential distribution, then obtain the average, and repeat this process until we have 1000 averages. If we plot the histogram of these 1000 averages, we obtain the graph in the right panel of Figure 3. This distribution is no longer exponential, but appears bell-shaped, characteristic of a normal distribution. This is in agreement with the CLT - the resulting histogram is a good approximation of a normal distribution (shown in black) with a mean of 1/lambda, and variance of (1/lambda)^2.

## Appendix A: Code used to generate Figure 1

```
library(ggplot2)
library(grid)
library(gridExtra)

means_df <- data.frame(mean = means)
Figure1 <- ggplot(means_df, aes(x = mean))
Figure1 + geom_histogram(binwidth = 0.3,
                   colour = "black",
                   fill = "thistle",
                   aes(y = ..density..)) +
      stat_function(fun = dnorm,
                    arg = list(mean = 1/lambda, sd = 1/(lambda*sqrt(n))),
                    size = 2) +
      geom_vline(xintercept = obs_mean, colour = "purple", size = 1.5) +
      geom_vline(xintercept = theor_mean, linetype = "dashed", size = 1.5) +
      labs(x = "Mean of Exponential Distribution",
           y = "Density",
           title = "Figure 1: Observed vs. Theoretical Mean") +
      theme(axis.text = element_text(size = 8),
            axis.title = element_text(size = 10),
            plot.title = element_text(size = 12))
```

## Appendix B: Code used to generate Figure 2

```
Figure2 <- ggplot(means_df, aes(x = mean))
Figure2 + geom_histogram(binwidth = 0.3,
                   colour = "black",
                   fill = "thistle",
                   aes(y = ..density..)) +
      geom_density(colour = "purple", size = 1.5) +
      stat_function(fun = dnorm,
                    arg = list(mean = 1/lambda, sd = 1/(lambda*sqrt(n))),
                    size = 1.5,
                    linetype = "dashed") +
      labs(x = "Mean of Exponential Distribution",
           y = "Density",
           title = "Figure 2: Observed vs. Theoretical Variance") +
      theme(axis.text = element_text(size = 8),
            axis.title = element_text(size = 10),
            plot.title = element_text(size = 12))
```

## Appendix C: Code used to generate Figure 3

```
set.seed(1)

exp_sample <- data.frame(exp = rexp(1000, 0.2))

Figure3A <- ggplot(exp_sample, aes(x = exp)) +
      geom_histogram(binwidth = 3,
                   colour = "black",
                   fill = "thistle",
                   aes(y = ..density..)) +
      labs(x = "Distribution of 1000 Exponentials", y = "Density") +
      theme(axis.text = element_text(size = 8),
            axis.title = element_text(size = 10))

Figure3B <- ggplot(means_df, aes(x = mean)) +
      geom_histogram(binwidth = 0.3,
                   colour = "black",
                   fill = "thistle",
                   aes(y = ..density..)) +
      stat_function(fun = dnorm,
                    arg = list(mean = 1/lambda, sd = 1/(lambda*sqrt(n))),
                    size = 2) +
      labs(x = "Distribution of 1000 Means of 40 Exponentials", y = "Density") +
      theme(axis.text = element_text(size = 8),
            axis.title = element_text(size = 10))

Figure3 <- grid.arrange(Figure3A, Figure3B, ncol = 2,
                        main = textGrob("Figure 3: Distribution of 1000 Exponentials vs. Dist
ribution of 1000 Averages of 40 Exponentials", gp=gpar(fontsize=12)))
```