Statistical Inference Course Project

Overview: In this project we 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 is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

We illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials:

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Before running the simulations, we load any libraries we need:

```
library(ggplot2)
library(knitr)
```

Simulations: We set lambda = 0.2 for all of the simulations and investigate the distribution of averages of 40 exponentials. We perform a thousand simulations.

```
lambda <- 0.2
nexp <- 40
sims <- 1000
set.seed(1349)

# We perform a thousand simulated averages of 40 exponentials
exp_sim <- matrix(rexp(sims*nexp, rate=lambda), sims, nexp)

# For each of the thousand simulations, calculate the mean of the 40 values
means_sim <- rowMeans(exp_sim)

x_bar <- mean(means_sim)

x_bar
```

```
## [1] 4.995376
```

Sample Mean versus Theoretical Mean: Notice that the sample mean x_bar is very close to the theoretical mean 1/lambda = 1/0.2 = 5.

Sample Variance versus Theoretical Variance: The theoretical variance of the distribution is (1/lambda)^2/n = (1/(0.2)^2)/40 = 0.625.

The sample variance is given as follows:

```r
var(means_sim)
```

```
## [1] 0.6154561
```

The sample variance is close to the theoretical variance.

Distribution: Via figures and text, explain how one can tell the distribution is approximately normal.

To see that the sample distribution of the sampled means is approximately normal, we plot both the histogram and the normal distribution.

```r
# plot the histogram of the means
hist_data <- data.frame(means_sim)

hist <- ggplot(hist_data, aes(x = means_sim)) +
        geom_histogram(aes(y=..density..), colour="black", fill = "#2200FF") +
        xlab("Sample Mean") + ylab("Density") +
        ggtitle("Sample Means Drawn from Exponential Distribution with lambda =0.2") +
        stat_function(aes(x = c(2, 8)), fun = dnorm, color = '#FFFF00', args = list(mean = 5, sd = sqrt

hist
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```