title: "Comparison of the exponential distribution and the central limit theorem " author: "Sumant" date: "29/08/2020"

output: html_document

```{r echo=FALSE, message=FALSE, warning=FALSE} library(ggplot2)

```
Overview

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem (CLT).

The exponential distribution can be simulated in R with the `rexp(n, lambda)` function where lambda λ represents the rate pa

According the CLT if a sample consists of at least 30 independent observations and the data are not strongly skewed, then the distri

In general, a sampling distribution represents the distribution of the point estimates based on samples of a fixed size from a certa

In this analysis we will show that the sampling distribution of the mean of an exponential distribution with $n = 40$ observations a

Comparison of the sample mean and the theoretical mean of the distribution

In the following we will draw 1000 samples of size 40 from an $Exp(\frac{1}{0.2}, \frac{1}{0.2})$ distribution. For each of the 1000

According to the CLT we would expect that each single mean of those 1000 means is already approximately $\frac{1}{\lambda} = \frac{

We will check if this is the case.

```{r}
set.seed(1234)

exp_sample_means <- NULL
for(i in 1:1000){
  exp_sample_means <- c(exp_sample_means, mean(rexp(40, 0.2)))
}
mean(exp_sample_means)
```

$\bar{x}$ in our case is `r round(mean(exp_sample_means), 2)` which is very close to the mean of the theoretical distribution namely $\mu = \frac{1}{0.2} = 5$.

Comparison of the sample variance with the theoretical distribution

According to the CLT we would expect that the variance of the sample of the 1000 means is approximately $\frac{\frac{1}{0.2^{2}}}{40} = 0.625$.

```{r} var(exp_sample_means)

```

s^{2} in our case is `r round(var(exp_sample_means), 2)` which is close to the variance of the theoretical distribution we mention

Showing that the sample distribution is approximately normal

In order to demonstrate that the sample distribution of the 1000 sampled means is approximately normal we will plot the corresponden

```{r fig.height=4, fig.width=4}
data <- as.data.frame(exp_sample_means)
ggplot(data, aes(x = exp_sample_means)) +
  geom_histogram(binwidth = 0.4, color = 'black', fill = 'white', aes(y = ..density..)) +
  stat_function(aes(x = c(2, 8)), fun = dnorm, color = 'red',
                args = list(mean = 5, sd = sqrt(0.625))) +
  xlab('Sample mean') +
  ylab('Density') +
  ggtitle('Comparison of the sample distribution\n and the theoretical distribution')
```

s^{2} in our case is `r round(var(exp_sample_means), 2)` which is close to the variance of the theoretical distribution we mentioned above.

Showing that the sample distribution is approximately normal

In order to demonstrate that the sample distribution of the 1000 sampled means is approximately normal we will plot the correspondent histogram and overlay it with the density function from the theoretical sampling distribution which is $N(\frac{1}{0.2}, \frac{\frac{1}{0.2}}{\sqrt{40}})$ distributed.