Title: Comparison Exponential Distribution in R with Central Limit Theorem (CLT)

Author: midham76

Date: 22 October, 2015

output: pdf_document

Overview

This project 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 parameter. The mean of an exponential distribution is $\mu = \frac{1}{\lambda}$ and the standard deviation is $\sigma = \frac{1}{\lambda}$.

Analysis will show that the sampling distribution of the mean of an exponential distribution with $n = 40$ observations and $\lambda = 0.2$ is indeed approximately $N(\frac{1}{0.2}, \frac{\frac{1}{0.2}}{\sqrt{40}})$ distributed.

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

The following will draw 1000 samples of size 40 from an $Exp(\frac{1}{0.2}, \frac{1}{0.2})$ distribution. For each of the 1000 samples, mean will be calculated. Theoretically, this the same as drawing a single sample of size 1000 from the corresponding sampling distribution with $N(\frac{1}{0.2}, \frac{\frac{1}{0.2}}{\sqrt{40}})$.

According to the CLT, it is expected that each single mean of those 1000 means is already approximately $\frac{1}{\lambda} = \frac{1}{0.2} = 5$.

```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)
```

```
## [1] 4.974239
```

$\bar{x}$ in our case is 4.97 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 , it is expected 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)
```

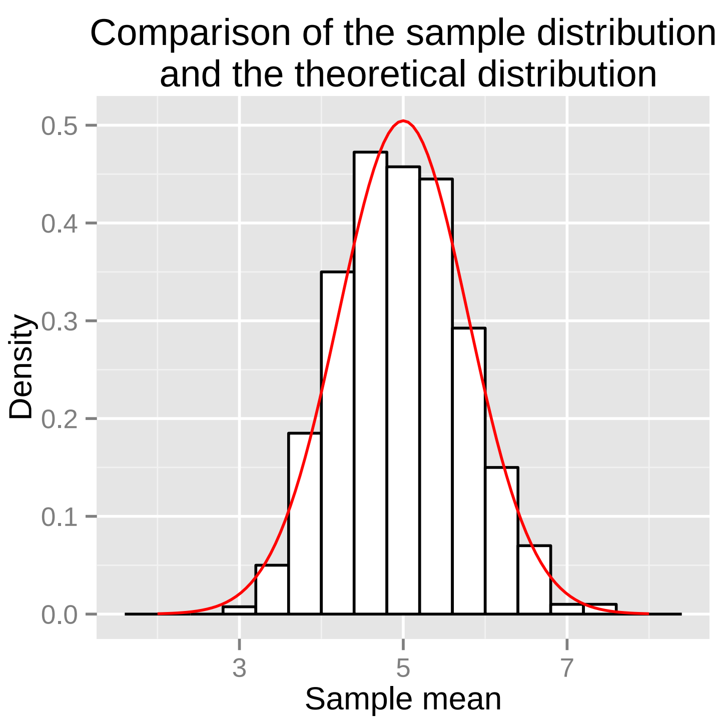
```
## [1] 0.5706551
```

s^2 in our case is 0.57 which is close to the variance of the theoretical distribution we mentioned above.

Showing that the sample distribution is approximately normal

To demonstrate that the sample distribution of the 1000 sampled means is approximately normal, correspondent histogram will be plotted 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.

```r
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')
```



Conclusions

In this analysis, it showed that the sampling distribution of the mean of an exponential distribution with $n = 40$ observations and $\lambda = 0.2$ is approximately $N(\frac{1}{0.2}, \frac{\frac{1}{0.2}}{\sqrt{40}})$ distributed.