Comparison of the exponential distribution and the Central Limit Theorem

Cindy Chen

Oct 20, 2015

Statistical Inference Project

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 'rexp(n, λ)' where lambda λ is the rate parameter.

The mean of exponential distribution (μ) is $1/\lambda$ and the standard deviation (σ) is also $1/\lambda$.

According the CLT if a sample consists of at least 30 independent observations and the data are not strongly skewed, then the distribution of the sample mean is approximattely: $\bar{x}_n \sim N(\mu, \frac{\sigma}{\sqrt{n}})$.

In this analysis, we will show that the sampling distribution of the mean of an exponential distribution with $n = 40$ observations and $\lambda = 0.2$. The $Exp(\frac{1}{0.2}, \frac{1}{0.2})$ distribution then is approximately $N(\frac{1}{0.2}, \frac{\frac{1}{0.2}}{\sqrt{40}})$.

We will investigate the distribution of averages of 40 exponentials and we will need to do 1000 simulations.

Simulations:

Here is the R code for 1000 simulations:

```
set.seed(123)
# The distribution of 1000 averages of size 40 from the exponential distribution
exp_sample_means <- NULL
for(i in 1:1000) {
  exp_sample_means <- c(exp_sample_means, mean(rexp(40, 0.2)))
}
```

Sample Mean versus Theoretical Mean:

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 samples we will calculate the mean. (Then we will get 1000 means)

Theoretically, this is the same as drawing a single sample of size 1000 from the corresponding sampling distribution with $N(\frac{1}{0.2}, \frac{\frac{1}{0.2}}{\sqrt{40}})$.

$\bar{x}$ in our case is 5.0119113 which is very close to the mean of the theoretical distribution namely $\mu = \frac{1}{0.2} = 5$.

Sample Variance versus Theoretical Variance:

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

```
var(exp_sample_means)
```

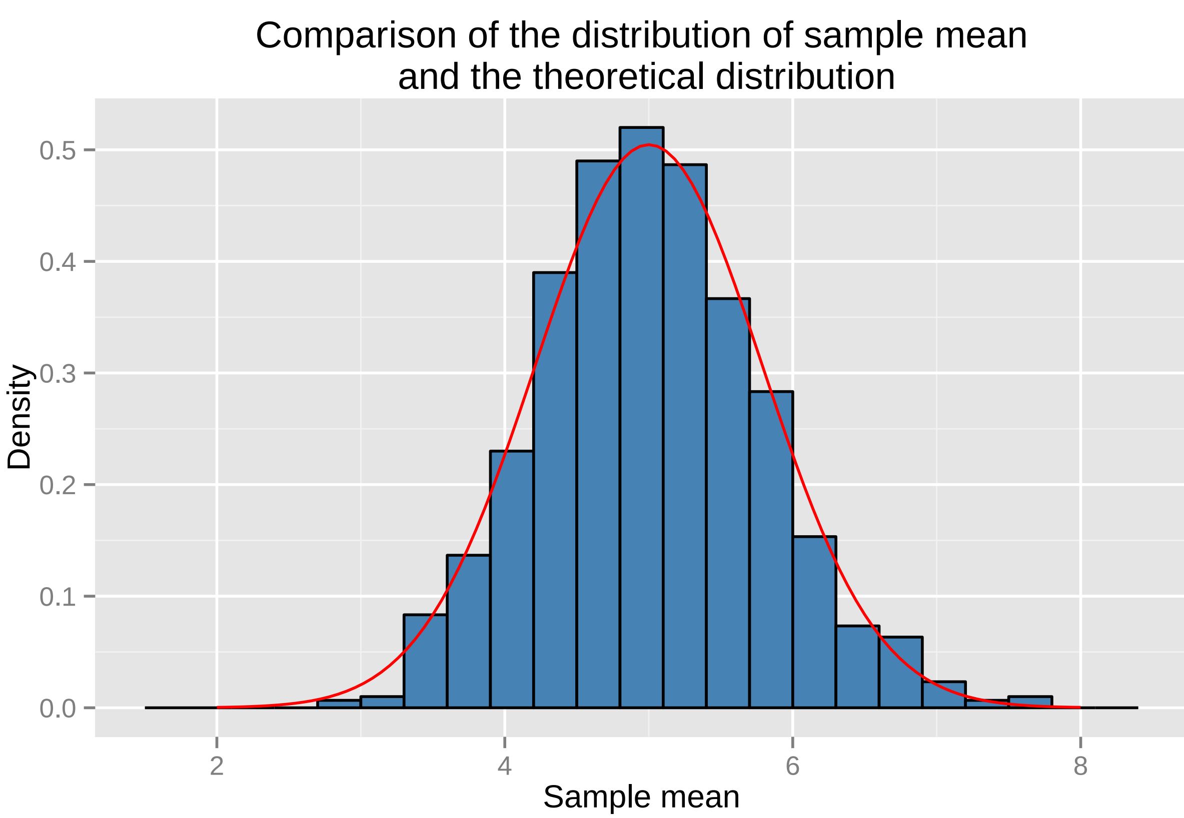
```
## [1] 0.6004928
```

s^2 in our case is 0.6004928, which is close to the variance of the theoretical distribution 0.625.

Showing the sample mean distribution is approximately normal:

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

```
library(ggplot2)
data <- as.data.frame(exp_sample_means)
ggplot(data, aes(x = exp_sample_means)) +
  geom_histogram(binwidth = 0.3, color = 'black', fill = 'steelblue', 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 distribution of sample mean\n and the theoretical distribution')
```



Conclusions

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

Thus, we showed:
1. The sample mean is close to the theoretical mean of the distribution.
2. The variable of the sample mean distribution is close to the theoretical variance of the distribution.
3. The sample mean distribution is approximately normal.