Statical inteferance assingment 1

2023-03-02

Demonstration of Central Limit theorem Using Exponential Distribution

Overview

the purpose of this analysis is to investigate the exponential distribution and compare the same to the Central Limit Theorem.The parameter(i.e., lambda) will be set to 0.2 for all the simulations. This analysis will compare the distribution of averages of 40 exponentials and a total of thousand simulation .

Simulation

The exponential distribution is simulated with , where lambda is the rate parameter. The mean of exponential distribution and standard deviation is 1/lambda. Lambda is set at 0.2 for all simulations.

Sample exponential distribution

```
set.seed(2021)
nosam <- 1000
n <- 40
lambda <- 0.2
samdata <- matrix(rexp(nosam*n,rate = lambda),nosam)

sam_mean <- rowMeans(samdata)

samdata_mean <- mean(sam_mean)
samdata_sd <- sd(sam_mean)
samdata_var <- var(sam_mean)

print(samdata_mean)
```

```
## [1] 5.008639
```

```
print(samdata_sd)
```

```
## [1] 0.788257
```

```
print(samdata_var)
```

```
## [1] 0.6213491
```

Theoretical exponential distribution

Theoretical mean of distribution

```
t_mean <- 1/lambda
print(t_mean)
```

```
## [1] 5
```

standard deviation of distribution

```
t_sd <- (1/lambda)*(1/sqrt(n))

t_var <- t_sd^2
print(t_sd)
```

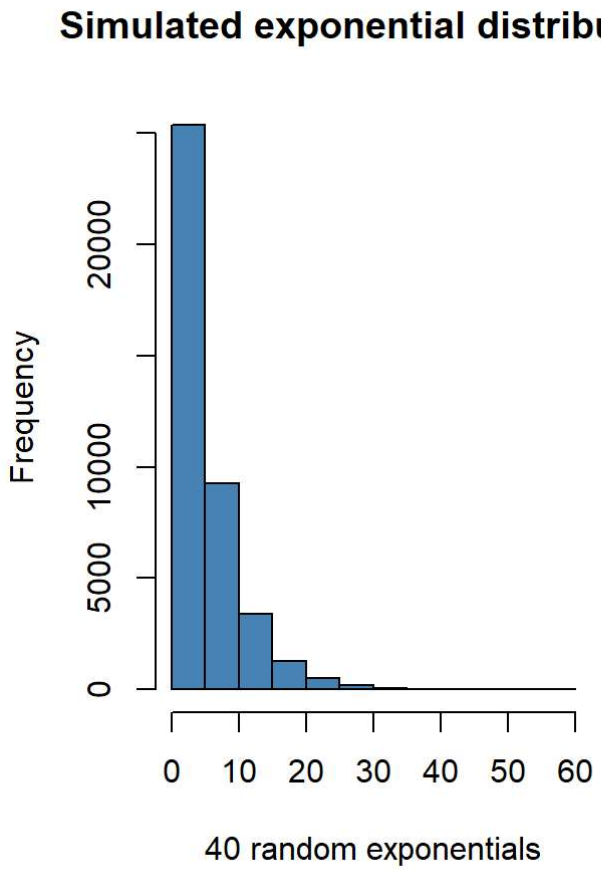
```
## [1] 0.7905694
```

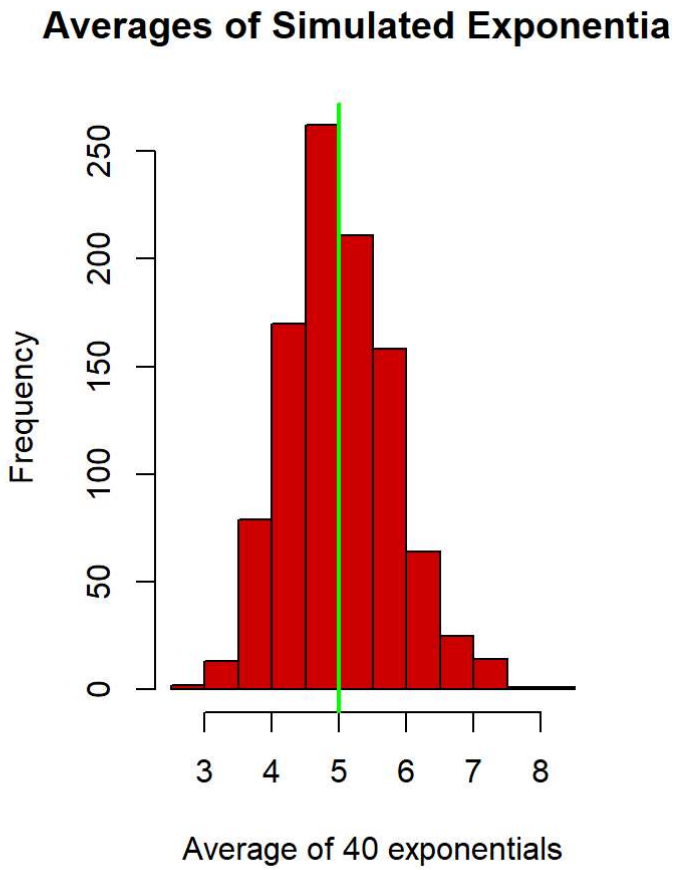
```
print(t_var)
```

```
## [1] 0.625
```

Histogram of sample exponential distribution vs Averages of simulated exponentials.

```
par(mfrow = c(1, 2))
hist(samdata,
     col = "steelblue",
     main = "Simulated exponential distribution",
     xlab = "40 random exponentials")
hist(sam_mean,
     col = "red3",
     main = "Averages of Simulated Exponentials",
     xlab = "Average of 40 exponentials")
abline(v = t_mean, col = "green", lwd = 2)
```





Sample mean vs theoretical mean

The sample mean of the exponential distribution is centered at 5.008 whereas the theoretical mean, 1/lambda is 5

The difference between the sample and theoretical mean is 0.0086

Sample variance vs theoretical variance

The sample Variance is is 0.621, which is very close to the theoretical variance, 0.625.

The difference between them is 0.0037 .

Distribution

Given below is the density histogram of the 1000 simulations. Added over the same plot is the normal distribution with mean (lambda ^ -1) and variance of (t_sd^2) which are the theoretical parameter values for the normal distribution occuring due to the simulations.

```
library(ggplot2)

samdata_mean <- data.frame(sam_mean)
ggplot(samdata_mean, aes(sam_mean)) +
    geom_histogram(
        binwidth = .3,
        fill = "steelblue",
        color = "black",
        aes(y = ..density..)
    ) +
    geom_density(color = "blue", lwd = 1) +
    labs(title = "Distribution of Random Exponential Values with 1000 simulations",
         x = "Average of 40 Exponentials", y = "Density") +
    stat_function(
        fun = dnorm,
        args = list(mean = t_mean, sd = t_sd),
        color = "red",
        lwd = 1
    ) +
    theme_bw()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
```

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
```