Statistical Inference Course Project

VR SREEGANESH

Statistical Inference Course Project

Overview

This project is to simulate the Centre Limit TheoremThis project is concerned with a simulation exercise that exhibits an example of Centre Limit Theorem taking effect.

Sample Mean vs Theoretical Mean

Here, we take the average of 40 exponentials and take their mean. This experiment/simulation is done 1000 times and a histogram is plotted. The theoretical mean, since the lambda is fed as 0.2 and the mean is given by 1/lamda, is 5. We see from the histogram that the distribution, which greatly resembles the normal distribution, is centred at 5. The blue line shows the theoretical line. We see that the theoretical mean and the mean of the samples are close to each other.

From the following code we see that the mean of the distribution lies within the 5% confidence interval of the theoretical mean.

```
library(ggplot2)

lambda <- 0.2
sample_mean = NULL
for (i in 1:1000) sample_mean = c(sample_mean, mean(rexp(40,lambda)))
sample_mean_df <- as.data.frame(sample_mean)

print(paste("The centre of the distribution is ", mean(sample_mean)))
```

```
## [1] "The centre of the distribution is  5.01240865445639"
```

```
print(paste("The theoretical mean is ", 1/lambda))
```

```
## [1] "The theoretical mean is  5"
```

```
print(paste("The distance between the centre of distribution and theoretical mean is ",
            mean(sample_mean)- 1/lambda))
```

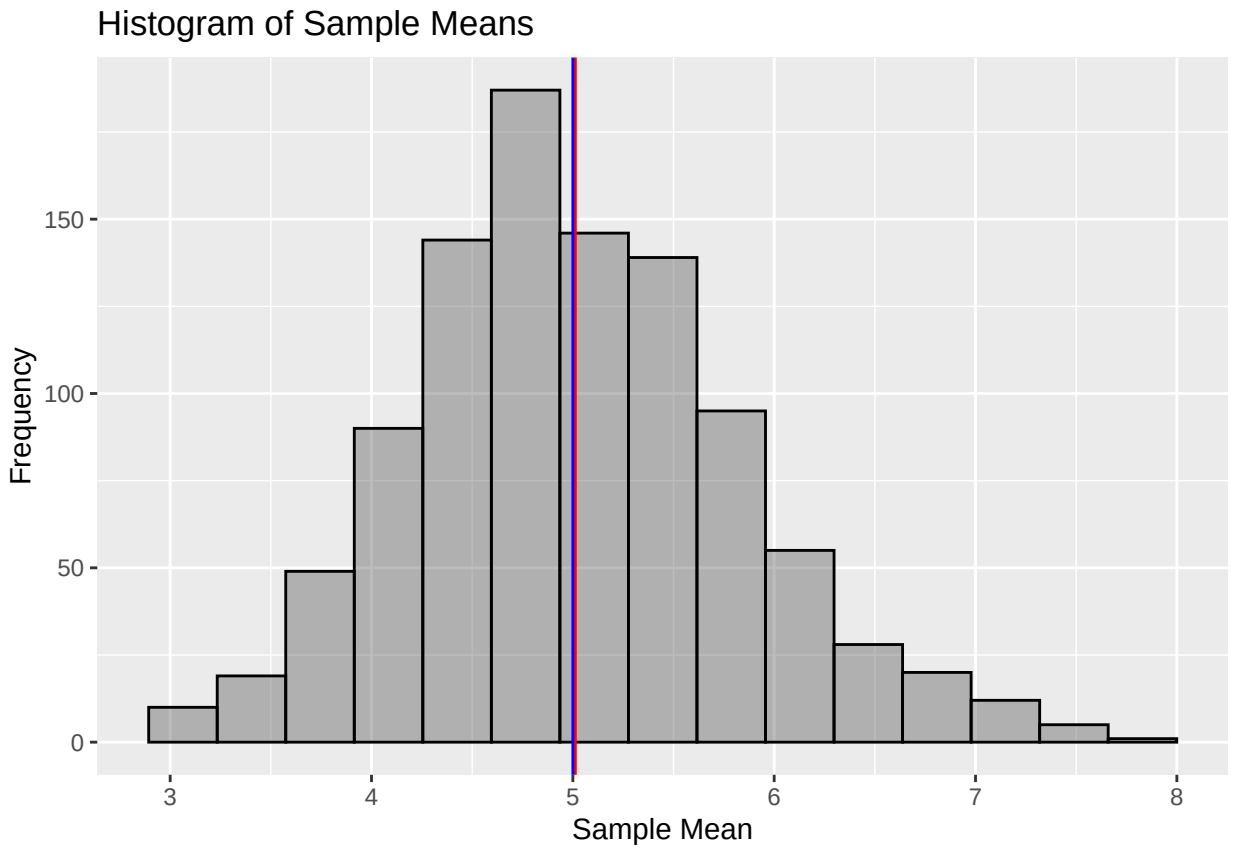
```
## [1] "The distance between the centre of distribution and theoretical mean is  0.0124086544563937"
```

```
print(paste("The confidence interval is from ",
            5 + -1*qnorm(.975)*(1/lambda)/sqrt(1000),
            "to ", 5 + qnorm(.975)*(1/lambda)/sqrt(1000)))
```

```
## [1] "The confidence interval is from  4.69010248384772 to  5.30989751615228"
```

We can see that the centre of the distribution lies within the 5 percent (two sided) confidence interval of the theoretical mean.

```
ggplot(sample_mean_df, aes(sample_mean)) + geom_histogram(alpha = 0.4, bins = 15, col = "black") +
        geom_vline(xintercept = mean(sample_mean), colour = "red") +
        geom_vline(xintercept = 1/lambda, colour = "blue") +
        ggtitle("Histogram of Sample Means") + xlab("Sample Mean") + ylab("Frequency")
```



Sample Variance vs Theoretical Variance

Here, we take the variance of 40 exponentials and we take their mean. This experiment/simulation is done 40 times and a histogram is plotted. The theoretical variance, since the lambda is fed as 0.2, is given by ((1/lambda)^2)/n.

```
sample_var = NULL
n <- 40
lambda <- 0.5
for (i in 1:1000) sample_var = c(sample_var, var(rexp(40,lambda)))
sample_var_df <- as.data.frame(sample_var)

print(paste("The theoretical variance is ", ((1/lambda)^2)))
```

```
## [1] "The theoretical variance is  4"
```

```
print(paste("The variance of the distribution is ", mean(sample_var)))
```

```
## [1] "The variance of the distribution is  3.92201144046191"
```

```
print(paste("The difference betweeen the variance of the distribution is ",
            ((1/lambda)^2) - mean(sample_var)))
```

```
## [1] "The difference betweeen the variance of the distribution is  0.0779885595380927"
```

Distribution

We see from the histograms that the curve represents a normal distribution. The red line represents a line fitted through the distribution of the sample means while the green line represents a normal distribution with mean as 5 and standard deviation equal to the standard error of the samples.

```
ggplot(sample_mean_df, aes(sample_mean)) +
        geom_histogram(aes(y = ..density..), bins = 30, fill = "turquoise", col = "black") +
        geom_density(colour = "red", size = 1) +
        stat_function(fun = dnorm, colour = "green", args = list(mean = 5, 5/sqrt(40)))
```