Central Limit Theorem with Exponential Distribution

thenuv

25 July 2018

Overview

This project is to understand the relationship between Exponential distribution with Central limit theory (CLT). In this we take group of random variables and simulate a thousand values to prove that CLT would result in normal distribution as against a single group. Mean and Standard deviation for exponential distribution is taken as 1/lambda. comparision is done for bellow. - **Mean** - **Variance** - **Distribution**

Simulation

The exponential distribution is simulated with *rexp* function. Data is simulated with 40 random exponential values and similar set is simulated 1000 times. Lambda is set as 0.2.

Compute Stats by Theory

```
n <- 40
lambda <- .2
samplesize <- 1000

mean_by_theory <- 1/lambda   # Mean value derived
variance_by_theory <- mean_by_theory ^2 / samplesize   # Variance derivied
ci_by_theory<-mean_by_theory +c(-1,1)*qnorm(.975)*sqrt(variance_by_theory)/sqrt(samplesize)
```

Compute Stats by Simulation

```
set.seed(4)
simulation_df <- matrix(rexp(n * samplesize, lambda), n, samplesize)
simulation_df <- as.data.frame(simulation_df)

sim_mean <- rowMeans(simulation_df)
simulation_df$sim_mean <-  rowMeans(simulation_df)

mean_by_Simulation <- mean(sim_mean)   # Mean for Simulated Average
variance_by_simulation <- var(sim_mean)   # Variance for Simulated data
ci_by_simulation <-mean(sim_mean)+c(-1,1)*qnorm(.975)*sqrt(var(sim_mean))/sqrt(samplesize)
```

Mean Comparision

Theoretically calculated by **1/lambda** as in the variable *mean_by_theory*. While the simulated mean of averages is calculated as in the variable *mean_by_Simulation*. Refer the plot 2 bellow which has the mean in the vertical blue line. They are approximately same.

- Mean by Theory : **5**
- Mean for Simulated data : **5.0252889**

Variance Comparision

Theoretical variance is derived with the formula **mean^2/sampleSize** as in *variance_by_theory*. Simulated variance is calculated as in the variable on R code above for *variance_by_simulation*. They values computed are close as bellow.

- Variance by Theory : **0.025**
- Vairance for Simulated data : **0.0265329**

Distribution

Confindence Interval

The CI for theory is calculated using the formula as in variable *ci_by_theory* for 2 tailed 95% interval. The simulated value is as per the variable *ci_by_simulation*. its observed the values are similar.

- Confidence Interval by Theory : **4.9902002, 5.0097998**.
- Confidence Interval by Simulation : **5.0151931, 5.0353847**.

Distribution Histogram

The first plot bellow shows the histogram for 1 set of 40 random exponential values. As by theory the negative exponential theory can be observed with the density curve bellow. This is the expected trend for this distribution.

The second plot represent the mean distribution of simulated averages. We could observe from the density curve that distribution is almost normal. This proves the Central limit theorem is applicable for exponential distribution which converges to Normal distribution.

```
hist(simulation_df[,1], col="grey", freq=FALSE,
    xlab = "Simulated Values", ylab="Lambda", main ="Histogram of Random Exponential")
lines(density(simulation_df[,1]), col="brown", lwd=2)
lines(density(simulation_df[,1], adjust=2), lty="dotted")
```

Histogram of Random Exponential
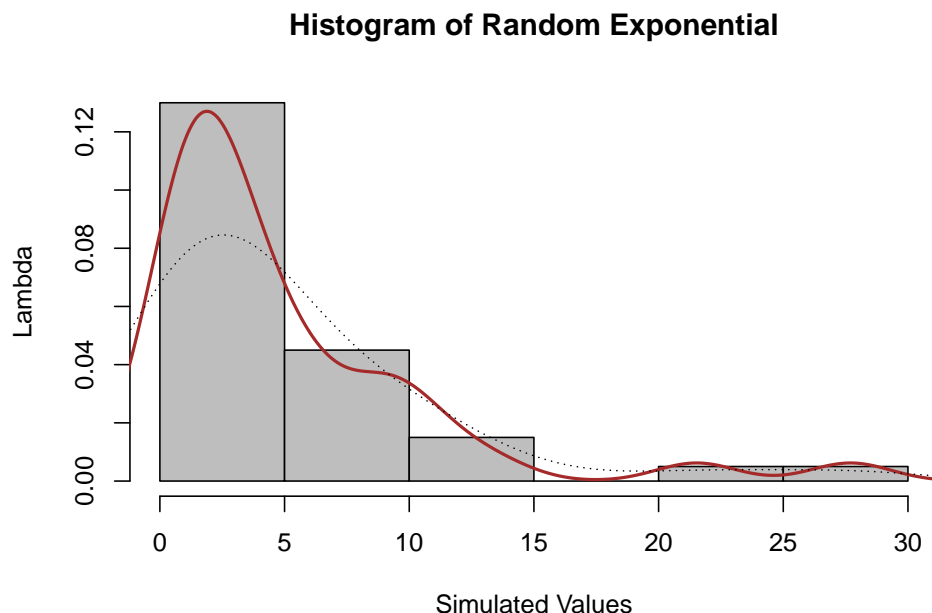
```r
hist(simulation_df$sim_mean, breaks = seq(4,6, by = .1), col = "khaki",
    xlab = "Simulated Mean", main ="Histogram of Simulated Mean")
abline(v=mean(simulation_df$sim_mean), col="blue", lwd=2)
lines(density(simulation_df$sim_mean), col ="darkgreen", lwd=3 )
lines(density(simulation_df$sim_mean, adjust=2), lty="dotted", lwd=3)
```

Histogram of Simulated Mean