Statistical Inference part 1

Astha

2024-09-30

Overview

This analysis aims to explore the exponential distribution and compare it to the Central Limit Theorem. The parameter (lambda) will be fixed at 0.2 for all simulations. We will examine the distribution of averages derived from 40 exponential random variables.

Simulations

First, we will set the seed and parameter values. The code will then generate a distribution of averages from 40 exponential random variables across 1000 simulations.

```
lambda <- 0.2
nexp <- 40

average <- NULL
for(i in 1:1000)
  average <- c(average, mean(rexp(nexp, lambda)))
```

Sample Mean versus Theoretical Mean

Theoretical mean distribution Both the theoretical and sample means are nearly identical, with only a minor difference between the two. ‘`lambda ^ -1 mean(average)` ## Sample Variance versus Theoretical Variance Again it can be seen that both the theoretical as well as sample variance are approximately same with a very small difference between them.

‘`(lambda * sqrt(nexp)) ^ -2 var(average)`

##Distribution The density histogram below illustrates the results from the 1000 simulations. Overlaying this histogram is a normal distribution with a mean of (lambda ˆ -1) and a variance of ((lambda*sqrt(nexp)) ˆ -2), which represent the theoretical parameters for the normal distribution resulting from the simulations. As shown in the plot, the normal density curve aligns closely with the histogram, demonstrating that the distribution of averages from a large sample of exponential variables approximates a normal distribution.

```
g <- ggplot(data.frame(column = average), aes(x = column))
g <- g + geom_histogram(aes(y = after_stat(density)), binwidth = 0.2, fill = 'pink', color = 'black')
g <- g + stat_function(fun = dnorm, args = list(mean = lambda^-1, sd=(lambda*sqrt(nexp))^-1), size=2)
g <- g + labs(title = "Distribution of Exponentials", x = "Simulation Means", y = "Density")
g
```