Statistical Inference Course Project1

2023-08-31

Statistical Inference Course Project

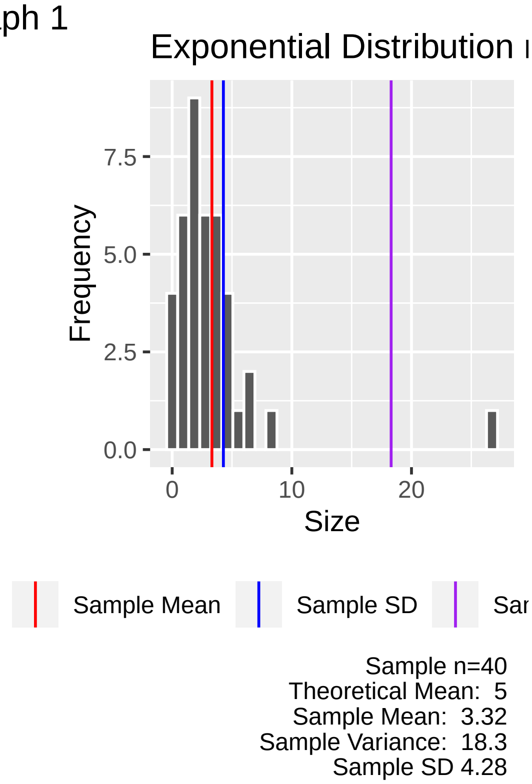
Introduction: The Exponential distribution or negative exponential is the probability distribution of the time between events in a Poisson point process; In which events occur continuously and independently at a constant average rate. One property of the exponential distribution is that it can be viewed as a continuous analogue of the geometric distribution.

The central limit theorem is one of the most useful concepts in statistics. The theorem is all about drawing samples of a finite size n from a population. The theorem states that if one collects samples of a large enough sample size n , and calculates each sample's mean (or sum), the shape of the histogram of those means (or sums) approximates a normal bell shape. The usefulness of the central limit theorem is due to the fact, that it does not matter what the distribution of the original population is, the distribution of sample means and the sums tend to follow the normal distribution.

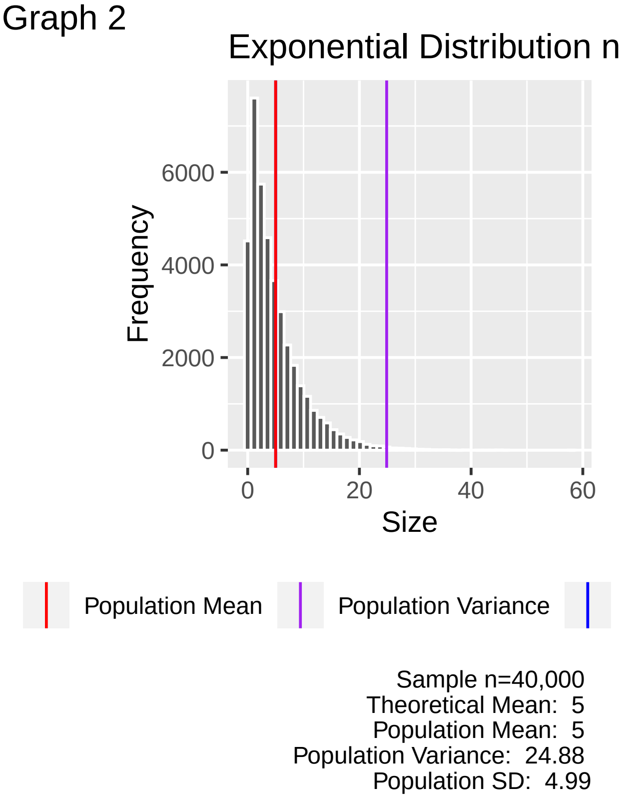
Hence, when conducting 1000 simulations with a sample size of 40 drawn from the exponential distribution, the resulting distribution of sample means is inclined to exhibit a resemblance to the normal distribution, as outlined below.

Results - graphs

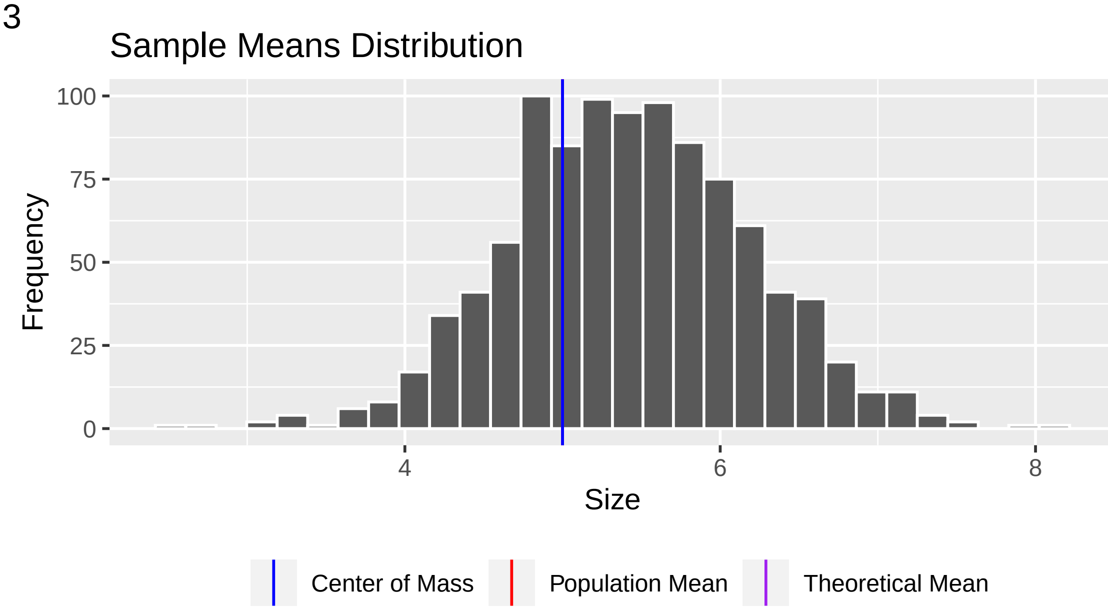
Graph 1



Graph 2

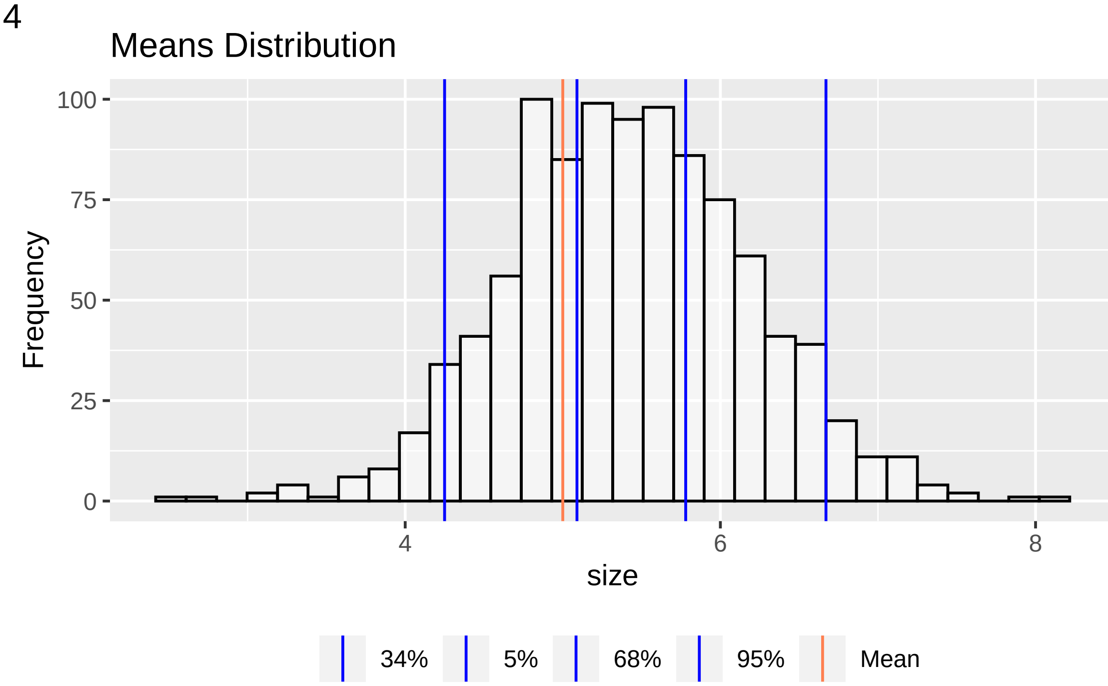


Graph 3



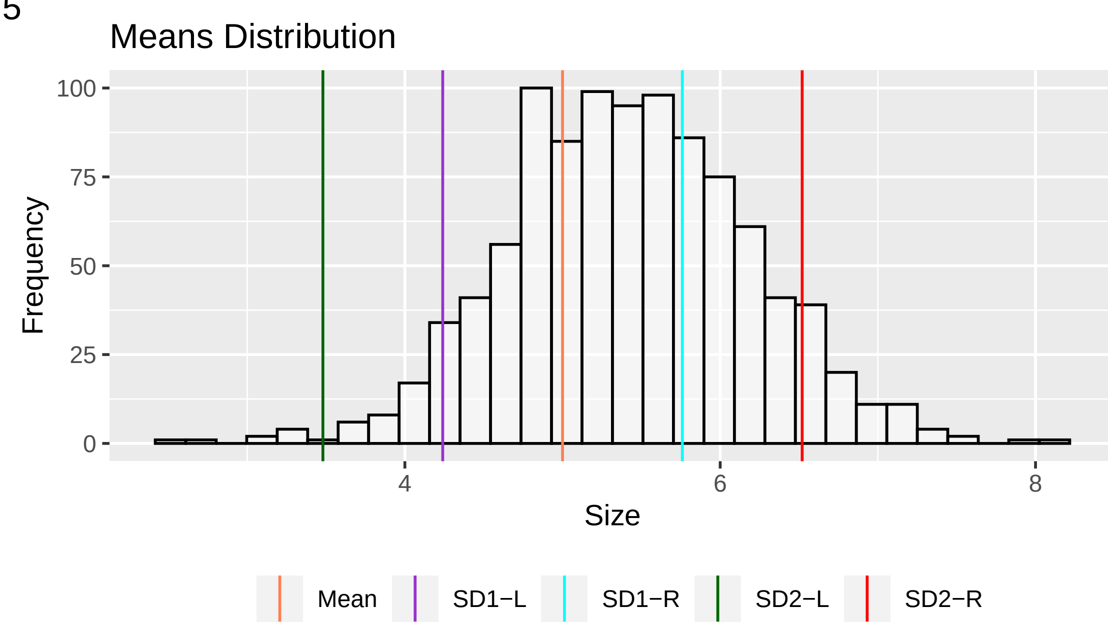
PopulationMean: 5
Theoretical mean: 5
Means Distribution Mean: 5
Variability: 32.09
The Samples Mean Distribution follows the Central Limit Theorem
which states that as the sample size increases, the distribution
of sample means approaches a normal distribution.

Graph 4



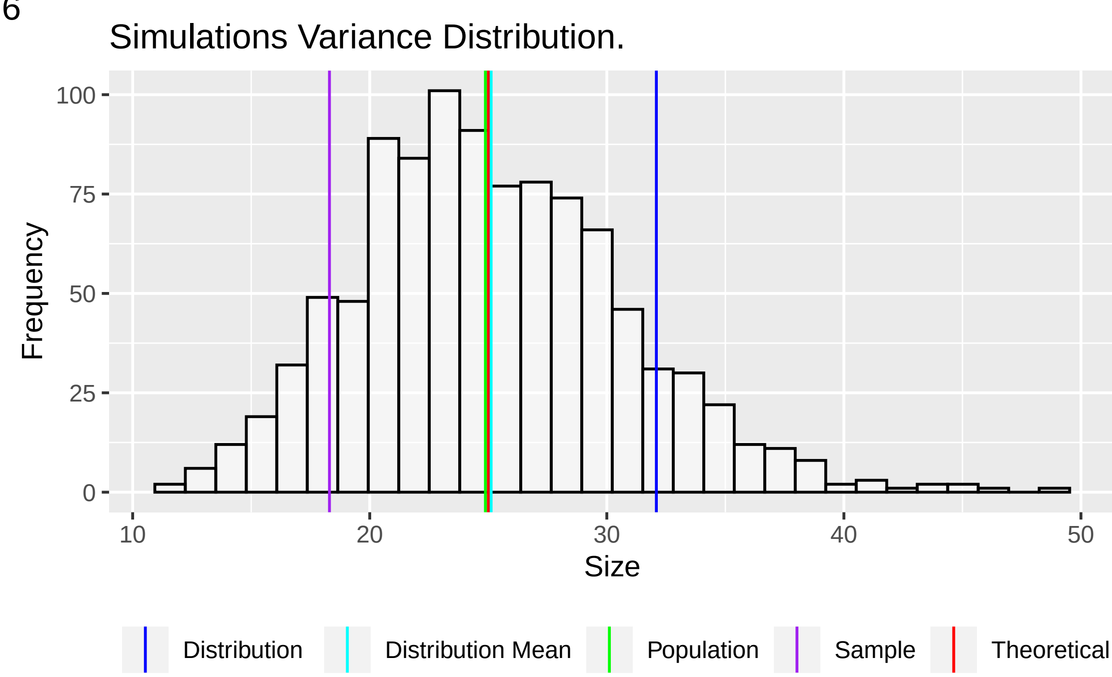
Mean: 5
5% : 4.25 34% : 5.09 68% : 5.78 95% : 6.67
SD: 0.76
Variability: 32.09

Graph 5



Mean: 5
SD: 0.76
2SD−R 6.52   1SD−R 5.76
2SD−L 3.48   1SD−L 4.24
Theoretical SD 5
Variability: 32.09

Graph 6



Distribution Variability: 32.09
Sample Variability: 18.3
Theoretical Variability: 25
Distribution Mean: 25.12
Population Variance 24.88

Discussion

As the Central Limit Theorem (CLT) states, regardless of the shape of the original population distribution, as the sample size increases, the sample means will approach a normal distribution as long as the sample size is sufficiently large. Graph 1 represents the behavior of one simulation n=40, and its mean is 3.32. In contrast, graph 2 represents the population n=40,000; its mean is 5, precisely the theoretical mean value. Graph 3 shows the mean of 1000 simulations n=40, which is also 5. The behavior of graphs 1, 2, and 3 represent CLT as the sample size increases. In Graph Four, the distribution shows that the distribution quantiles are not symmetrical around the mean as opposed to the normal distribution. Additionally, Graph 5 shows that the standard deviations are not within the limits of the quantiles that characterize the normal distribution. As seen in graphs 4 and 5, while the shapes of the exponential and normal distributions are quite different, there is a similarity in the concept of the "bell curve." A symmetric bell-shaped curve famously characterizes the normal distribution. On the other hand, the exponential distribution is skewed, with a rapidly dropping curve on the left side and a long tail on the right, as seen in Graph 2. Graph 6 represents the variance distribution, where the mean of the distribution is very close to the theoretical variance, 25.12 and 25, respectively. This corresponds with the book "Statistical Inference" on page 44, "The distribution of the sample variance is centered at what it's estimating."

Sources

- https://ggplot2.tidyverse.org/
- Statistical Inference for data science by Brian Caffo
- https://rmarkdown.rstudio.com/authoring_quick_tour.html#Markdown_Basics
- https://ggplot2-book.org/