Homework 9

Law of Large Numbers

The **law of large numbers** states that as the sample size increases, the sample mean tends to converge towards the population mean:

$$\lim_{n \to \infty} \overline{X} = \mu$$

This principle is fundamental in statistics and has applications in many fields, including **cybersecurity**. For example, in network anomaly detection, the analysis of large amounts of data can help identify unusual behaviour that could indicate malicious activity.

Simulation and Distribution of Sample Variances

In the code, several random samples are generated, each with a number of intervals and associated probabilities. For each sample, the sample mean and variance are calculated. Subsequently, the following are analysed:

- **The distribution of the sample variances**: observing how the variances of the samples vary from each other.
- **The mean of the sample variances**: by comparing it with the variance of the theoretical distribution.

Relationship with Theoretical Distribution

It is observed that:

- **Average of Sample Averages** $(\overline{X})$: tends to converge towards the population mean (μ) as the number of samples increases, in accordance with the law of large numbers.
- **Sample Mean Variance** $(\overline{S^2})$: provides an unbiased estimate of the population variance (σ^2).
- **Variance of Sample Averages**: decreases as sample size increases, indicating that sample averages become more concentrated around the population mean.
- **Variance of Sample Variances**: reflects the variability of sample variances; tends to decrease with larger samples.

Applications and Importance

This simulation provides an understanding of how sample statistics behave in practice and how they relate to theoretical parameters. In the context of **cybersecurity**, such concepts are crucial for:

- **Security Data Analysis**: understanding variability in data collected by monitoring systems.
- **Anomaly Detection**: using statistics to identify significant deviations from normal behaviour.

- **Risk Assessment**: estimate the probability and impact of safety events based on sample data.