Homework 7

Differences between Empirical Mean and Empirical Variance on Different Samples

In the programme provided, we simulate the generation of m independent samples, each with a number of intervals n and a certain total number of occurrences. For each sample, the **empirical mean** and the **empirical variance** are calculated. Next, the **average of means** and the **average of variances** are calculated over all m samples.

Empirical Average of a Single Sample

For a single sample i, the empirical mean $\bar{X}_i$ is calculated as:

$$\bar{X}_i = \frac{1}{N_i} \sum_{j=1}^{n} x_j \cdot f_{ij}$$

where:

- x_j is the value associated with the interval j,
- f_{ij} is the frequency (number of occurrences) of the interval j in sample i,
- $N_i = \sum_{j=1}^{n} f_{ij}$ is the total number of occurrences in sample i.

Empirical Variance of a Single Sample

The empirical variance S_i^2 for sample i is calculated as:

$$S_i^2 = \frac{1}{N_i} \sum_{j=1}^{n} f_{ij} \cdot (x_j - \bar{X}_i)^2$$

Mean of Averages and Mean of Variances

After calculating the means and variances for each sample, we obtain:

- **Average of Averages**:

$$\bar{X} = \frac{1}{m} \sum_{i=1}^{m} \bar{X}_i$$

- **Average of Variances**:

$$S^2 = \frac{1}{m} \sum_{i=1}^{m} S_i^2$$

Differences and Interpretation

Empirical Average vs Average of Averages

- **Empirical Average ($\bar{X}_i$)**: is the average calculated over a single sample. It may vary between samples due to data variability.
- **Average of Averages ($\bar{X}$)**: is the average of the averages of all samples. It tends to get closer to the theoretical expected value as m increases, due to the **Large Numbers Law**.

Empirical Variance vs. Average of Variances

- **Empirical Variance (S_i^2)**: measures the dispersion of data within a single sample compared to its mean.
- **Average of Variances (S^2)**: represents the average of the variances of all samples, providing a more stable estimate of the overall variability.

Demonstration of Convergence

According to the **Law of Large Numbers**, as m increases, the mean of the sample averages $\bar{X}$ converges to the expected value μ of the population:

$$\lim_{m \to \infty} \bar{X} = \mu$$

Similarly, the mean of the sample variances S^2 converges to the variance σ^2 of the population:

$$\lim_{m \to \infty} S^2 = \sigma^2$$

Conclusion

The fundamental difference between statistics calculated on a single sample and those averaged over m samples is the **stability** of the estimates:

- The **averages** and **variances** of individual samples can be affected by random fluctuations.
- The **average of averages** and **average of variances** provide more reliable estimates of actual population parameters.

This concept emphasises the importance of using multiple samples to obtain more precise statistical estimates by exploiting the fundamentals of probability theory.