

Lecture Summary: Empirical Distribution and Descriptive Statistics

Source: Lecture 7.2.pdf

Key Points

- **Empirical Distribution:**

- Computed from a set of iid samples X_1, X_2, \dots, X_n .
- Definition:

$$\hat{P}(t) = \frac{\#(X_i = t)}{n},$$

where $\#(X_i = t)$ is the number of samples equal to t .

- Represents a PMF derived directly from sample data.

- **Example of Empirical Distribution:**

- For 20 samples: 1, 1, 0, 1, 0, \dots , 0, range $\{0, 1\}$.
- Empirical probabilities:

$$\hat{P}(0) = \frac{8}{20}, \quad \hat{P}(1) = \frac{12}{20}.$$

- Another set: 1, 2, 0, 3, \dots , range $\{0, 1, 2, 3\}$.
- Probabilities:

$$\hat{P}(0) = \frac{6}{20}, \quad \hat{P}(1) = \frac{6}{20}, \quad \hat{P}(2) = \frac{5}{20}, \quad \hat{P}(3) = \frac{3}{20}.$$

- **Properties:**

- Changes with the sampling instance.
- Depends on the actual data, not just the underlying distribution.
- Is a random quantity due to the variability in samples.

- **Descriptive Statistics:**

- Provide a summary of the empirical distribution.
- Common measures:

- * **Mean:**

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

- * **Variance:**

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}.$$

- Connected to the underlying distribution.

- **Key Results:**

- Expected value of the sample mean:

$$\mathbb{E}[\bar{X}] = \mu,$$

where μ is the distribution mean.

- Variance of the sample mean:

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

- Expected value of sample variance:

$$\mathbb{E}[S^2] = \sigma^2.$$

- **Sample Proportion:**

- Defined for an event A :

$$S(A) = \frac{\#(X_i \in A)}{n}.$$

- Expected value and variance:

$$\mathbb{E}[S(A)] = P(A), \quad \text{Var}(S(A)) = \frac{P(A)(1 - P(A))}{n}.$$

Simplified Explanation

Empirical Distribution: Derived from sample data, showing the frequency of outcomes. **Descriptive Statistics:** Summarize the data, including mean, variance, and proportions.

Key Insights: 1. Sample mean converges to the true mean as n increases. 2. Sample variance reflects the true variance. 3. Sample proportions approximate event probabilities.

Conclusion

In this lecture, we:

- Defined empirical distributions and computed examples.
- Connected sample mean, variance, and proportions to underlying distributions.
- Highlighted their use in analyzing and understanding data.

Empirical and descriptive statistics provide simple yet powerful tools for connecting sample data to underlying distributions.