Maximum Likelihood Estimation

Introduction to Joint Distributions

Suppose we have data generated from observations of n random variables $X_1, X_2, \ldots, X_n$. We can describe their behavior in terms of their joint distribution:

- **Identically Distributed**: Each variable follows the same distribution.
- **Independent**: The outcome of one variable does not affect the others.
- **Independently and Identically Distributed (i.i.d.)**: Combines both conditions above.

Examples of Joint Distributions

A **joint distribution** represents multiple random variables within the same mathematical framework.

Example 1: Flipping a coin and rolling a die (independently). Possible outcomes include $\{H, 4\}$ or $\{T, 2\}$, etc.

Example 2: The heights (in inches) of three brothers might be $\{76, 56, 40\}$ or $\{67, 70, 68\}$. (Discuss which set seems more likely and why, considering normal height ranges for various ages.)

Formalizing Joint Distributions

We can express the joint density function of independent variables as a product of their individual density functions:

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\theta) = \prod_{i=1}^{n} f_{X_i}(x_i|\theta)$$

For i.i.d. data, this simplifies further to:

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

Introduction to Maximum Likelihood Estimation (MLE)

MLE is a method to estimate the parameters of a distribution. It seeks the parameter values that maximize the likelihood of observing the given data.
Example: Consider you have scores of 90%, 85%, and 95% on practice exams. How probable are these scores if your true ability corresponds to a 60% average score? Or a 100% average score? Using MLE, we can estimate the most likely true ability level.

Likelihood Function

The **Likelihood function**, denoted as $L(\theta|x_1, \ldots, x_n)$, represents the probability of observing data given certain parameters:

$$L(\theta|x_1, \ldots, x_n) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

The MLE of θ, denoted as $\hat{\theta}$, maximizes this likelihood function:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ L(\theta|x_1, \ldots, x_n)$$

Solving the Maximum Likelihood Estimation

To find the MLE, we often need to solve a system of equations derived from maximizing the likelihood function, which typically requires calculus or multivariate calculus for multiple parameters. For the normal distribution, the MLE solutions for the mean and variance are intuitive:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$