Lecture Summary: Maximum Likelihood Estimation (MLE)

Lecture: 9.6 - Estimator Design - Maximum Likelihood

Source: Lecture 8.6.pdf

Key Points

- **Introduction to Maximum Likelihood:**
  - MLE is a method for estimating parameters of a distribution by maximizing the likelihood function.
  - The likelihood represents how probable the observed data is, given the parameter values.
  - MLE uses the assumption that the data are iid samples.
- **Likelihood Function:**
  - For iid samples $X_1, X_2, \ldots, X_n$, with PDF or PMF $f_X(x;\theta)$, the likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i;\theta).$$

  - Example:
    * For $X \sim N(\mu, \sigma^2)$, $L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$.
  - The likelihood is a function of the parameters θ and not the data.
- **Log-Likelihood:**
  - Instead of maximizing $L(\theta)$, maximize $\log L(\theta)$ (simplifies calculations):

$$\log L(\theta) = \sum_{i=1}^{n} \log f_X(x_i;\theta).$$

  - Log transformation converts products to sums, making differentiation easier.
- **MLE Procedure:**
  1. Write the likelihood function $L(\theta)$.
  2. Take the logarithm to get $\log L(\theta)$.
  3. Differentiate $\log L(\theta)$ with respect to θ.
  4. Set the derivative to zero and solve for θ.
  5. Verify that the solution maximizes the likelihood.
- **Examples:**
  1. **Bernoulli(p):**

- Likelihood:

$$L(p) = p^w (1-p)^{n-w},$$

where w is the number of successes.

- Log-likelihood:

$$\log L(p) = w \log p + (n-w) \log(1-p).$$

- MLE:

$$\hat{p}_{\text{MLE}} = \frac{w}{n}.$$

2. **Normal(μ, σ^2):**
   - Likelihood:

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}.$$

   - Log-likelihood:

$$\log L(\mu, \sigma) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2.$$

   - MLE:

$$\hat{\mu}_{\text{MLE}} = \bar{x}, \quad \hat{\sigma}^2_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

- **Key Properties of MLE:**
  - **Consistency:** As $n \to \infty$, the MLE converges to the true parameter value.
  - **Asymptotic Normality:** For large n, the MLE is approximately normal.
  - **Efficiency:** MLE achieves the lowest possible variance among unbiased estimators (under certain conditions).

- **Applications:**
  - Bernoulli trials: Estimating success probability.
  - Normal distribution: Estimating mean and variance.
  - Broad applicability in statistical modeling and machine learning.

Simplified Explanation

Key Idea: MLE finds parameter values that make the observed data most likely.

Steps: 1. Define the likelihood function. 2. Take the log of the likelihood. 3. Differentiate, set to zero, and solve for parameters.

Examples: - For Bernoulli trials, $\hat{p} = \frac{\text{successes}}{\text{total trials}}$. - For normal data, $\hat{\mu} = \text{mean}$, $\hat{\sigma}^2 = \text{variance}$.

Why It Matters: MLE is widely used for parameter estimation due to its theoretical soundness and practicality.

Conclusion

In this lecture, we:

- Defined MLE and its key concepts.
- Demonstrated its application to Bernoulli and normal distributions.
- Highlighted its properties and significance in statistical analysis.

MLE is a cornerstone of statistical inference, enabling robust parameter estimation across diverse applications.