PABNA UNIVERSITY OF SCIENCE AND TECHNOLOGY

Faculty of Engineering and Technology

Department of Information and Communication Engineering

Course name: Engineering Statistics
Course Code: STAT-2201
Assignment Topic: Principle of Maximum Likelihood & MLE Illustration from Binomial

Submitted By:

MD. Junaid
Roll: 220618
Session: 2021-22
2nd Year 2nd Semester
Department of ICE, PUST

Submitted To:

Dr. Md. Sarwar Hosian
Associate Professor
Department of ICE
Pabna University of Science and Technology
Pabna, Bangladesh

Date of Submission: 23-04-2025

Maximum Likelihood Estimation: Illustration with the Binomial Distribution

Introduction to Maximum Likelihood Estimation (MLE)

Maximum Likelihood Estimation (MLE) is a statistical method used to estimate the parameters of a probability distribution by maximizing the likelihood function. The likelihood function measures the probability of observing the given data under different values of the unknown parameter(s). The MLE chooses the parameter value that makes the observed data most probable.

Likelihood Function

Suppose we have a sample of n observations $x_1, x_2, \ldots, x_n$ drawn from a distribution with probability density function (or probability mass function for discrete distributions) $f(x_i|\theta)$, where θ is the parameter to estimate. The likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Since products can be cumbersome, we often work with the **log-likelihood function**:

$$\ell(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i|\theta)$$

The MLE of θ, denoted $\hat{\theta}$, is the value that maximizes $L(\theta)$ or equivalently $\ell(\theta)$. This is typically found by taking the derivative of the log-likelihood with respect to θ, setting it equal to zero, solving for θ, and verifying that the solution is a maximum (e.g., checking the second derivative).

MLE for the Binomial Distribution

The Binomial distribution models the number of successes in n independent trials, each with a success probability p. We derive the MLE for the parameter p based on observed data.

Binomial Distribution Recap

For a Binomial random variable $X \sim \text{Binomial}(n, p)$, the probability mass function (PMF) is:

$$P(X = x|n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

where x is the number of successes ($x = 0, 1, \ldots, n$), n is the number of trials, p is the probability of success, and $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ is the binomial coefficient.

Our goal is to estimate p using MLE based on observed data.

Scenario

Suppose we have a single observation $X = k$ from a Binomial distribution with known n and unknown p. The likelihood function for $X = k$ is:

$$L(p) = P(X = k|n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

Write the Likelihood Function

Since $\binom{n}{k}$ is a constant, we focus on the part involving p:

$$L(p) \propto p^k (1-p)^{n-k}$$

Compute the Log-Likelihood

Take the natural logarithm of the likelihood function:

$$\ell(p) = \ln L(p) = \ln \left[\binom{n}{k} p^k (1-p)^{n-k} \right]$$

Using properties of logarithms:

$$\ell(p) = \ln \binom{n}{k} + \ln(p^k) + \ln\left((1-p)^{n-k}\right)$$

$$= \ln \binom{n}{k} + k \ln p + (n-k) \ln(1-p)$$

Since $\ln \binom{n}{k}$ is a constant, we focus on:

$$\ell(p) = k \ln p + (n-k) \ln(1-p)$$

Differentiate the Log-Likelihood

Take the derivative of the log-likelihood with respect to p:

$$\frac{d}{dp}\ell(p) = \frac{d}{dp}\left[k \ln p + (n-k) \ln(1-p)\right]$$

First term: $\frac{d}{dp}(k \ln p) = k \cdot \frac{1}{p} = \frac{k}{p}$. Second term: $\frac{d}{dp}\left[(n-k)\ln(1-p)\right] = (n-k) \cdot \frac{1}{1-p} \cdot (-1) = -\frac{n-k}{1-p}$. Thus:

$$\frac{d}{dp}\ell(p) = \frac{k}{p} - \frac{n-k}{1-p}$$

Set the Derivative to Zero

Solve for p:

$$\frac{k}{p} - \frac{n-k}{1-p} = 0$$

$$\frac{k}{p} = \frac{n-k}{1-p}$$

Cross-multiply:

$$k(1-p) = (n-k)p$$

$$k - kp = np - kp$$

$$k = np$$

$$p = \frac{k}{n}$$

Thus, the MLE for p is:

$$\hat{p} = \frac{k}{n}$$

Verify the Maximum

Compute the second derivative to confirm a maximum:

$$\frac{d^2}{dp^2}\ell(p) = \frac{d}{dp}\left(\frac{k}{p} - \frac{n-k}{1-p}\right)$$

First term: $\frac{d}{dp}\left(\frac{k}{p}\right) = k \cdot \left(-\frac{1}{p^2}\right) = -\frac{k}{p^2}$. Second term: $\frac{d}{dp}\left(-\frac{n-k}{1-p}\right) = -(n-k) \cdot (-1) \cdot (1-p)^{-2} \cdot (-1) = -\frac{n-k}{(1-p)^2}$. So:

$$\frac{d^2}{dp^2}\ell(p) = -\frac{k}{p^2} - \frac{n-k}{(1-p)^2}$$

Since both terms are negative for $0 < p < 1$, the second derivative is negative, confirming that $\hat{p} = \frac{k}{n}$ is a maximum (assuming $0 < k < n$).

Boundary Check

If $k = 0$ or $k = n$, then $\hat{p} = 0$ or $\hat{p} = 1$. Evaluate the likelihood at boundaries:

- If $k = 0$, $L(p) = (1-p)^n$, maximized as $p \to 0$, so $\hat{p} = 0$.
- If $k = n$, $L(p) = p^n$, maximized as $p \to 1$, so $\hat{p} = 1$.

Illustration with Examples

Example 1

Suppose you flip a coin 10 times ($n = 10$) and observe 7 heads ($k = 7$). Assume the coin flips follow a Binomial distribution with unknown p. Find the MLE of p.

Solution:

$$\hat{p} = \frac{k}{n} = \frac{7}{10} = 0.7$$

The MLE for the probability of heads is $\hat{p} = 0.7$.

The likelihood function is:

$$L(p) = \binom{10}{7} p^7 (1-p)^{10-7} = 120 p^7 (1-p)^3$$

Evaluate at different p:

- For $p = 0.5$: $L(0.5) = 120 \cdot (0.5)^7 \cdot (0.5)^3 = 120 \cdot 0.0078125 \cdot 0.125 = 0.1171875$.
- For $p = 0.7$: $L(0.7) = 120 \cdot (0.7)^7 \cdot (0.3)^3 \approx 0.2666$.
- For $p = 0.9$: $L(0.9) = 120 \cdot (0.9)^7 \cdot (0.1)^3 \approx 0.0574$.

The likelihood is maximized around $p = 0.7$.
The log-likelihood is:

$$\ell(p) = \ln 120 + 7 \ln p + 3 \ln(1 - p)$$

Example 2

Suppose a quality control test involves inspecting 20 items ($n = 20$), and 12 are found to be defective ($k = 12$). Assume the defects follow a Binomial distribution with unknown defect probability p. Find the MLE of p.

Solution:

$$\hat{p} = \frac{k}{n} = \frac{12}{20} = 0.6$$

The MLE for the defect probability is $\hat{p} = 0.6$.
The likelihood function is:

$$L(p) = \binom{20}{12} p^{12} (1 - p)^{20-12} = \binom{20}{12} p^{12} (1 - p)^8$$

Evaluate at different p:

- For $p = 0.4$: $L(0.4) = \binom{20}{12} \cdot (0.4)^{12} \cdot (0.6)^8 \approx 0.000466$.
- For $p = 0.6$: $L(0.6) = \binom{20}{12} \cdot (0.6)^{12} \cdot (0.4)^8 \approx 0.003853$.
- For $p = 0.8$: $L(0.8) = \binom{20}{12} \cdot (0.8)^{12} \cdot (0.2)^8 \approx 0.000132$.

The likelihood peaks at $p = 0.6$.

General Case: Multiple Binomial Observations

For m independent observations $X_1, X_2, \ldots, X_m$, where $X_i \sim \text{Binomial}(n_i, p)$, the likelihood is:

$$L(p) = \prod_{i=1}^{m} \binom{n_i}{x_i} p^{x_i} (1-p)^{n_i - x_i}$$
$$\propto p^{\sum x_i} (1-p)^{\sum (n_i - x_i)}$$

Log-likelihood:

$$\ell(p) = \text{constant} + \left(\sum x_i\right) \ln p + \left(\sum (n_i - x_i)\right) \ln(1-p)$$

Differentiating and solving yields:

$$\hat{p} = \frac{\sum x_i}{\sum n_i}$$

Example 3

Consider three experiments: $n_1 = 10, x_1 = 7$, $n_2 = 5, x_2 = 4$, $n_3 = 8, x_3 = 6$. Compute the MLE for p:

$$\hat{p} = \frac{7+4+6}{10+5+8} = \frac{17}{23} \approx 0.739$$

Example 4

Suppose four experiments yield: $n_1 = 15, x_1 = 9$, $n_2 = 12, x_2 = 8$, $n_3 = 20, x_3 = 14$, $n_4 = 10, x_4 = 5$. Compute the MLE for p:

$$\hat{p} = \frac{9+8+14+5}{15+12+20+10} = \frac{36}{57} \approx 0.632$$

Properties of the MLE

- **Consistency**: As the sample size increases, $\hat{p}$ converges to the true p.
- **Asymptotic Normality**: For large samples, $\hat{p}$ is approximately normal with mean p and variance $\frac{p(1-p)}{n}$.
- **Invariance**: The MLE of $g(p)$ is $g(\hat{p})$.

Practical Notes

Confidence Intervals

Using asymptotic normality, the standard error is $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, and a 95% CI is:

$$\hat{p} \pm 1.96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

For Example 1 ($n = 10, k = 7$):

$$\hat{p} = 0.7, \quad \text{SE} = \sqrt{\frac{0.7 \cdot 0.3}{10}} \approx 0.1449$$

$$\text{CI} = 0.7 \pm 1.96 \cdot 0.1449 \approx (0.416, 0.984)$$

For Example 2 ($n = 20, k = 12$):

$$\hat{p} = 0.6, \quad \text{SE} = \sqrt{\frac{0.6 \cdot 0.4}{20}} \approx 0.1095$$

$$\text{CI} = 0.6 \pm 1.96 \cdot 0.1095 \approx (0.385, 0.815)$$

Conclusion

The MLE for the Binomial parameter p is the sample proportion, $\hat{p} = \frac{k}{n}$ for a single observation or $\hat{p} = \frac{\sum x_i}{\sum n_i}$ for multiple observations. The derivation involves maximizing the likelihood function through differentiation of the log-likelihood.