Maximum Likelihood Estimators

Maximum Likelihood Estimation (MLE) is a method used in statistics to estimate the parameters of a probability distribution by **maximizing the likelihood function**.

In simpler terms, MLE finds the parameter values that make the **observed data most likely** under a given statistical model.

Step-by-Step Understanding of MLE

1. **Define the Likelihood Function**
    - Suppose we have a dataset $X = \{x_1, x_2, ..., x_n\}$ that follows a probability distribution with an unknown parameter θ.
    - The likelihood function is the **joint probability of observing the given data,** assuming they are independent:

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

      Here, $f(x_i|\theta)$ is the probability density function (PDF) or probability mass function (PMF) depending on whether the data is continuous or discrete.
2. **Take the Log-Likelihood**
    - Since the likelihood function is a **product**, it's easier to work with the **log-likelihood function**:

      $\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i|\theta)$
3. **Differentiate and Solve for θ**
    - Find the value of θ that **maximizes** the log-likelihood by **taking its derivative** and setting it to zero:

      $\frac{d\ell(\theta)}{d\theta} = 0$
    - Solve for θ, which gives the **MLE estimate** $\hat{\theta}$.

MLE for different distributions

MLE for Bernoulli Parameter p

A Bernoulli distribution models a binary outcome (success/failure, 1/0, heads/tails). It is parameterized by p, which represents the probability of success. The goal of MLE is to find the best estimate of p given observed data.

Step 1: Define the Probability Mass Function (PMF)

The **Bernoulli distribution** has the following PMF:

$$P(X = x|p) = p^x (1-p)^{1-x}, \quad \text{for } x \in \{0, 1\}$$

Where:

- p = probability of success (X=1)
- 1 - p = probability of failure (X=0)

Step 2: Write the Likelihood Function

Suppose we observe **n independent Bernoulli trials** with outcomes:

$$X_1, X_2, ..., X_n$$

The **likelihood function** is the joint probability of observing the given data:

$$L(p) = \prod_{i=1}^{n} P(X_i|p)$$

Since each X_i follows a Bernoulli distribution:

$$L(p) = \prod_{i=1}^{n} p^{X_i} (1-p)^{1-X_i}$$

Expanding the product:

$$L(p) = p^{\sum X_i} (1-p)^{n - \sum X_i}$$

Where:

- $\sum X_i$ is the total number of successes in the sample.

Step 3: Take the Log-Likelihood

Since it's easier to work with logs, we take the **log-likelihood function**:

$$\ell(p) = \log L(p)$$

$$\ell(p) = \sum X_i \log p + (n - \sum X_i) \log(1-p)$$

Step 4: Differentiate and Solve for p

To find the **maximum likelihood estimate (MLE) of p**, we take the derivative with respect to p:

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

Setting this derivative to zero:

$$\frac{\sum X_i}{p} = \frac{n - \sum X_i}{1-p}$$

Solving for p:

$p(n - \sum X_i) = (1-p) \sum X_i$

$n = \sum X_i$

$\hat{p} = \frac{\sum X_i}{n}$

Step 5: Interpretation of $\hat{p}$

The **MLE estimate of p** is simply:

$$\hat{p} = \frac{\text{Number of successes}}{\text{Total trials}}$$

Which is the **sample proportion of successes**.

MLE for Poisson Parameter λ

A **Poisson distribution** models the number of occurrences of an event in a fixed time or space interval, given a constant average rate λ.

The goal of MLE is to find the best estimate of λ based on observed data.

Step 1: Define the Probability Mass Function (PMF)

The **Poisson distribution** is given by:

$$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, ...$$

Where:

- λ = expected number of occurrences in a given interval
- x = observed number of occurrences

Step 2: Write the Likelihood Function

Suppose we observe **n independent Poisson-distributed data points**:

$X_1, X_2, ..., X_n$

Since each X_i follows a Poisson distribution, the **likelihood function** is:

$$L(\lambda) = \prod_{i=1}^{n} P(X_i|\lambda)$$

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{X_i}}{X_i!}$$

$$L(\lambda) = e^{-n\lambda}\lambda^{\sum X_i} \prod_{i=1}^{n} \frac{1}{X_i!}$$

Since $\prod \frac{1}{X_i!}$ does not depend on λ, we ignore it when maximizing.

Step 3: Take the Log-Likelihood

Taking the natural log of the likelihood:

$$\ell(\lambda) = \log L(\lambda)$$

$$\ell(\lambda) = \log\left(e^{-n\lambda}\lambda^{\sum X_i}\right)$$

$$\ell(\lambda) = -n\lambda + \sum X_i \log\lambda$$

Using **logarithm properties**:

1. $\log(ab) = \log a + \log b$
2. $\log(e^x) = x$
3. $\log(a^b) = b \log a$

Step 4: Differentiate and Solve for λ

To maximize $\ell(\lambda)$, take its derivative with respect to λ:

$$\frac{d\ell(\lambda)}{d\lambda} = -n + \frac{\sum X_i}{\lambda}$$

Setting it to **zero**:

$$-n + \frac{\sum X_i}{\lambda} = 0$$

$$\lambda = \frac{\sum X_i}{n}$$

Step 5: Interpretation of $\hat{\lambda}$

The **MLE estimate of** λ is simply the **sample mean**:

$$\hat{\lambda} = \frac{1}{n} \sum X_i$$

This means that **the best estimate for λ is just the average number of occurrences per interval**.

MLE for Normal Population

A **normal distribution** is characterized by two unknown parameters:

- μ **(mean)** → the central tendency
- σ^2 **(variance)** → the spread of the data

The goal of **MLE** is to find the best estimates $\hat{\mu}$ **and** $\hat{\sigma}^2$ given observed data.

Step 1: Define the Probability Density Function (PDF)

The **normal distribution** has the following **probability density function (PDF)**:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where:

- μ is the **mean** (location parameter)
- σ^2 is the **variance** (scale parameter)

- x is an observation from the population

Step 2: Write the Likelihood Function

Suppose we observe **n independent samples** from a normal distribution:

$X_1, X_2, ..., X_n \sim N(\mu, \sigma^2)$

Since the data points are **independent**, the **likelihood function** is the **product** of their individual probabilities:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i|\mu, \sigma^2)$$

Substituting the normal PDF:

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

Taking the product:

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

Step 3: Take the Log-Likelihood Function

Since logarithms simplify products into sums, we take the **log-likelihood**:

$$\ell(\mu, \sigma^2) = \log L(\mu, \sigma^2)$$

Applying logarithm properties:

$$\ell(\mu, \sigma^2) = \log\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}\right)$$

Splitting terms:

$$\ell(\mu, \sigma^2) = n\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \log e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$$

Using $\log a^b = b \log a$ and $\log e^x = x$

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

Step 4: Differentiate and Solve for MLEs

To maximize the log-likelihood, take **partial derivatives** with respect to μ and σ^2.

▼ Finding $\hat{\mu}$

Differentiate $\ell(\mu, \sigma^2)$ with respect to μ:

$$\frac{\partial \ell}{\partial \mu} = -\frac{1}{2\sigma^2} \cdot 2 \sum (x_i - \mu)$$

$$= -\frac{1}{\sigma^2} \sum (x_i - \mu)$$

Setting this to zero:

$$\sum (x_i - \mu) = 0$$

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

Thus, the **MLE estimate of μ** is simply the **sample mean**:

$\hat{\mu} = \bar{X}$

▼ Finding $\hat{\sigma}^2$

Now differentiate $\ell(\mu, \sigma^2)$ with respect to σ^2:

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum (x_i - \mu)^2$$

Setting this to zero:

$$-\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum (x_i - \mu)^2 = 0$$

Multiplying by $2\sigma^4$:

$$-n\sigma^2 + \sum(x_i - \mu)^2 = 0$$

Solving for σ^2:

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

Thus, the **MLE estimate of σ^2** is:

$$\hat{\sigma}^2 = \frac{1}{n}\sum(x_i - \bar{X})^2$$

Which is the **sample variance with denominator n** (not n−1, which is used in unbiased estimation).

The final MLE Estimate

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum X_i$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum(X_i - \bar{X})^2$$

$$\hat{\sigma} = \left[\frac{\sum(X_i - \bar{X})^2}{n}\right]^{1/2}$$

- $\hat{\mu}$ is the **sample mean** (same as in standard statistics).
- $\hat{\sigma}^2$ is the **MLE estimate for variance** (divided by n, not n−1).

MLE for Uniform Distribution\Estimating the Mean of a Uniform Distribution

Problem Setup

Suppose we have a **Uniform distribution**:

$X \sim \mathrm{Uniform}(0, \theta)$

meaning:

- The random variable X is equally likely to take any value between 0 and θ.
- The probability density function (pdf) is:

$$f(x;\theta) = \begin{cases} \frac{1}{\theta}, & \text{if } 0 \le x \le \theta \\ 0, & \text{otherwise} \end{cases}$$

We observe a **sample** of n values: $x_1, x_2, \ldots, x_n$.

The goal is:

🔵 **Estimate the unknown parameter θ using MLE.**

Step 1: Write the Likelihood Function

Since the observations are independent, the **likelihood** is the product of the individual densities:

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

Since $f(x_i;\theta) = \frac{1}{\theta}$ for each $x_i \in [0,\theta]$, the likelihood becomes:

$$L(\theta) = \begin{cases} \theta^{-n}, & \text{if } 0 \le x_i \le \theta \text{ for all } i \\ 0, & \text{otherwise} \end{cases}$$

✅ In short: $L(\theta) = \theta^{-n}$ **if** $\theta \ge \max(x_1, \ldots, x_n)$, otherwise 0.

Step 2: Find the MLE

- $L(\theta) = \theta^{-n}$ **decreases** as θ increases.
- We want to **maximize** $L(\theta)$, so we should pick the **smallest possible** θ that still satisfies:

$\theta \ge \max(x_1, \ldots, x_n)$

Thus, the MLE is:

$$\boxed{\hat{\theta} = \max(x_1, x_2, \ldots, x_n)}$$

🔵 **Intuition**:

The estimated upper bound θ should be just enough to include the largest observed sample point.

Final Answer:

The **MLE** for θ in $\mathrm{Uniform}(0, \theta)$ is:

$$\boxed{\hat{\theta} = \max(x_1, x_2, \ldots, x_n)}$$
