

Lecture 17 Maximum Likelihood Estimation (MLE)

BIO210 Biostatistics

Xi Chen

Spring, 2023

School of Life Sciences

Southern University of Science and Technology



南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

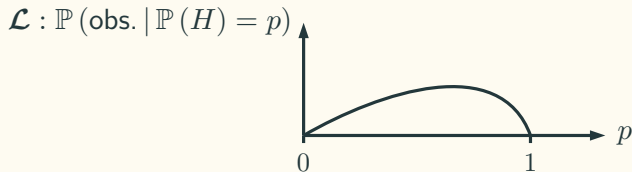
Intuition over MLE

Experiment: A coin, with an unknown $\mathbb{P}(H) = p$, was flipped 10 times. The outcome is $HHHTHHHTHH$.

Question: What is your best guess for p ?

Thinking: Given the data/observation we have, what values should p take such that our data/observation is most likely to occur ?

Aim: find the value that **maximise our chance of observing the data**, and use that value as our best guess/estimate for p .



Estimators of Parameters

- **Parameter space Ω :** the set of all possible values of a parameter θ or of a vector of parameters $(\theta_1, \theta_2, \theta_3, \dots, \theta_k)$ is called the parameter space.
 - Bernoulli: $\theta = p, \Omega = \{p \mid 0 \leq p \leq 1\}$
 - Binomial: $\theta_1 = n, \theta_2 = p, \Omega = \{(n, p) \mid n = 2, 3, \dots, \text{a finite number}; 0 \leq p \leq 1\}$
 - Poisson: $\theta = \lambda, \Omega = \{\lambda \mid \lambda \geq 0\}$
 - Normal (Gaussian): $\theta_1 = \mu, \theta_2 = \sigma^2, \Omega = \{(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 \geq 0\}$
- We refer to an estimator of a parameter θ as $\hat{\theta}$. An estimator $\hat{\theta}$ of a parameter θ is unbiased if $\mathbb{E}[\hat{\theta}] = \theta$. For example, $\hat{\mu} = \bar{X}$ is an unbiased estimator for μ .

Maximum Likelihood Estimation (MLE)

- **Maximum likelihood estimation (MLE)** is a technique used for estimating the parameters of a given distribution, using some observed data.
- Introduced by **R.A. Fisher** in 1912.
- MLE can be used to estimate parameters using a limited sample of the population, by finding particular values so that the observation is **the most likely result to have occurred**.

Maximum Likelihood Estimation (MLE)

Formal definition

Let $x_1, x_2, x_3, \dots, x_n$ be observations from n **i.i.d** random variables $(X_1, X_2, X_3, \dots, X_n)$ drawn from a probability distribution f_0 , where f_0 is known to be from a family of distributions \mathcal{f} that depend on some parameters θ . For example, f_0 could be known to be from the family of normal distributions \mathcal{f} , which depend on parameters μ and σ^2 , and $x_1, x_2, x_3, \dots, x_n$ would be observations from f_0 . The goal of MLE is to maximise the **likelihood function**:

$$\begin{aligned}\mathcal{L}(\theta; x_1, x_2, x_3, \dots, x_n) &= f(x_1, x_2, x_3, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) \\ &= f(x_1; \theta) \cdot f(x_2; \theta) \cdots f(x_n; \theta)\end{aligned}$$

The log-likelihood function:

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

Maximum Likelihood Estimation (MLE): Example 1

- Other notation: $\mathcal{L}(\theta|x_1, x_2, x_3, \dots, x_n) = f(x_1, x_2, x_3, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta)$

- **Example 1:** A coin, with an unknown $\mathbb{P}(H) = p$, was flipped 10 times. The outcome is $HHHTHHHTHH$. What is the MLE for p ?

- 1. Specify the parameter - $\theta : p$
- 2. Specify the parameter space - $\Omega : \{p \mid 0 \leq p \leq 1\}$
- 3. Write out the probability function - $p_X(k) = \begin{cases} p & , \text{ when } k = 1 \\ 1 - p & , \text{ when } k = 0 \end{cases}$
- 4. Write out the likelihood function:

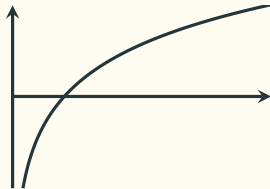
$$\begin{aligned}\mathcal{L}(p; 1110111011) &= f(1110111011; p) = \prod_{i=1}^{10} f(x_i; p) \\ &= f(1; p) \cdot f(1; p) \cdot f(1; p) \cdot f(0; p) \cdot f(1; p) \cdot f(1; p) \cdot f(1; p) \cdot f(0; p) \cdot f(1; p) \cdot f(1; p) \\ &= p \cdot p \cdot p \cdot (1 - p) \cdot p \cdot p \cdot p \cdot (1 - p) \cdot p \cdot p = p^8(1 - p)^2\end{aligned}$$

Maximum Likelihood Estimation (MLE): Example 2

- **Example 2 A more generalised case of coin flipping:** A (possibly unfair) coin is flipped m times, and k heads are observed. Let $\mathbb{P}(H) = p$. What is the MLE for p ?

- 1. $\theta : (n, p)$
- 2. $\Omega : \{(n, p) \mid n = m, 0 < p < 1\}$
- 3. $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$
- 4.

$$\mathcal{L}(n, p; k) = f(k; n, p) = \binom{m}{k} p^k (1-p)^{m-k}$$



$$\begin{aligned}\ell = \ln \mathcal{L} &= \ln \binom{m}{k} p^k (1-p)^{m-k} \\ &= \ln \binom{m}{k} + k \ln p + (m-k) \ln (1-p)\end{aligned}$$

What value should p take to maximise ℓ ?

$$\text{Let } \frac{d\ell}{dp} = 0 \Rightarrow \hat{p} = \frac{k}{m}$$

Maximum Likelihood Estimation (MLE): Example 3

- **Example 3 DNA synthesis errors:** The genetic material is copied and synthesised by DNA polymerase. One high-fidelity DNA polymerase, *Pfu*, originally isolated from the hyperthermophilic archae *Pyrococcus furiosus*, is believed to have very low error rate. Assume the errors generated by *Pfu* follow a Poisson distribution with λ mutations per 10^6 base pairs (Mb). We have examined n newly synthesised DNA fragments and observed that the number of mutations per Mb is $k_1, k_2, k_3, \dots, k_n$. What is the MLE for λ ?
 - 1. $\theta : \lambda$
 - 2. $\Omega : \{\lambda \mid \lambda > 0\}$
 - 3. $p_x(k) = \frac{\lambda^k}{k!} e^{-\lambda}$
 - 4. $\mathcal{L}(\lambda; k_1, k_2, \dots, k_n) = f(k_1, k_2, \dots, k_n; \lambda) = \prod_{i=1}^n \frac{\lambda^{k_i}}{k_i!} e^{-\lambda}$

Maximum Likelihood Estimation (MLE): Example 4

- **Example 4** *Pou5f1* expression in embryonic stem cells: Let the random variable X be the expression values of *Pou5f1*. We know $X \sim \mathcal{N}(\mu, \sigma^2)$, but μ and σ^2 are unknown. Now we have sequenced n cells, and the expressions of *Pou5f1* in those cells are x_1, x_2, \dots, x_n , respectively. What is the MLE for the parameters of this normal distribution?
 - 1. $\theta : \mu, \sigma^2$
 - 2. $\Omega : \{(\mu, \sigma^2) \mid -\infty < \mu < +\infty, \sigma^2 \geq 0\}$
 - 3. $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
 - 4. $\mathcal{L}(\mu, \sigma^2; x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n; \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$

Probability vs. Likelihood

$$\mathcal{L}(\theta; x_1, x_2, x_3, \dots, x_n) = f(x_1, x_2, x_3, \dots, x_n; \theta)$$



the likelihood of the parameter(s) θ
taking certain values given that a bunch
of data x_1, x_2, \dots, x_n are observed.



the joint probability mass/density
of observing the data x_1, x_2, \dots, x_n
with model parameter(s) θ .

from [Wolfram](#):

Likelihood is the hypothetical probability that an event that has already occurred would yield a specific outcome. The concept differs from that of a probability in that a **probability** refers to the occurrence of future events, while a **likelihood** refers to past events with known outcomes.

Advantages and Disadvantages of MLE

Advantages:

- Intuitive and straightforward to understand.
- If the model is correctly assumed, the MLE is efficient (meaning small variance or mean squared error).
- Can be extended to do other useful things.

Disadvantages:

- Relies on assumptions of a model (need to know the PMF/PDF).
- Sometimes difficult or impossible to solve the derivative of \mathcal{L} or ℓ .
- Sometimes leads to the wrong or biased conclusions