Lecture 17 Maximum Likelihood Estimation (MLE)

BIO210 Biostatistics

Xi Chen

Spring, 2022

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

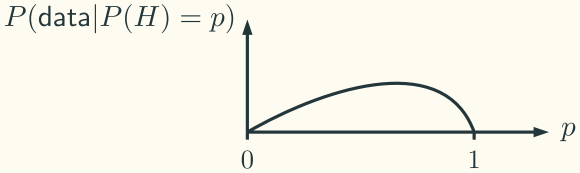
Intuition over MLE

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

Question: What is you 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 occurr ?

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 $(\theta_1, \theta_2, \theta_3, ..., \theta_k)$ is called the parameter space.

- Bernoulli: $\theta = p$, $\Omega = \{0 \leqslant p \leqslant 1\}$
- Binomial: $\theta_1 = n, \theta_2 = p$, $\Omega = \{(n, p) \mid n = 2, 3, ..., \text{a finite number}; 0 \leqslant p \leqslant 1\}$
- Poisson: $\theta = \lambda$, $\Omega = \{\lambda = 0, 1, 2, 3, 4, ...\}$
- Normal (Gaussian): $\theta_1 = \mu, \theta_2 = \sigma^2$, $\Omega = \{(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 \geqslant 0\}$

- We refer to an estimator of a parameter θ as $\hat{\theta}$. An estimator $\hat{\theta}$ of a parameter θ is unbiased if $E[\hat{\theta}] = \theta$. 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 $x_1, x_2, x_3, ..., x_n$ be observations from n **i.i.d** random variables $(X_1, X_2, X_3, ..., X_n)$ drawn from a probability distribution f_0, where f_0 is known to be from a family of distributions $\boldsymbol{f}$ that depend on some paramters θ. For example, f_0 could be known to be from the family of normal distributions $\boldsymbol{f}$, which depend on parameters μ and σ, and $x_1, x_2, x_3, ..., x_n$ would be observations from f_0. The goal of MLE is to maximise the likelihood function:

$$\boldsymbol{\mathcal{L}} = f(x_1, x_2, x_3, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdots f(x_n; \theta)$$

The log-likelihood function:

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

Maximum Likelihood Estimation (MLE): Example 1

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

- **Example 1**: A (possibly unfair) coin is flipped 100 times, and 61 heads are observed. The coin either has probability of 1/3, 1/2 or 2/3 of obtaining a head each time it is flipped. Which of the three is the MLE?

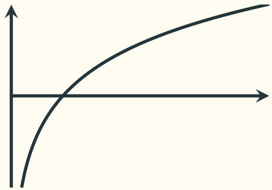
- 1. $\theta : (n, p)$
- 2. $\boldsymbol{\Omega} : \{(n, p) \mid n = 100,\ p \in \{\frac{1}{3}, \frac{1}{2}, \frac{2}{3}\}\}$
- 3. $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$
- 4. $\boldsymbol{\mathcal{L}} = f(61; n, p)$

$$f(61; n, p) = \binom{100}{61} p^{61} (1-p)^{39} = \begin{cases} \approx 9.6 \times 10^{-9} & \text{, if } p = \frac{1}{3} \\ \approx 0.007 & \text{, if } p = \frac{1}{2} \\ \approx 0.04 & \text{, if } p = \frac{2}{3} \end{cases}$$

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 $P(H) = p$. What is the MLE for p?

- 1. $\theta : (n, p)$
- 2. $\mathbf{\Omega} : \{(n, p) \mid n = m, 0 \leqslant p \leqslant 1\}$
- 3. $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$
- 4. $\boldsymbol{\mathcal{L}} = f(k; n, p) = \binom{m}{k} p^k (1-p)^{m-k}$



$$\boldsymbol{\ell} = \ln \boldsymbol{\mathcal{L}} = \ln \binom{m}{k} p^k (1-p)^{m-k}$$

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

What value should p take to maximise $\boldsymbol{\ell}$?

Let $\dfrac{d\boldsymbol{\ell}}{dp} = 0 \;\Rightarrow\; \hat{p} = \dfrac{k}{m}$

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 nubmer of mutations per Mb is $k_1, k_2, k_3, ..., k_n$. What is the MLE for λ?

- 1. $\theta : \lambda$
- 2. $\boldsymbol{\Omega} : \{\lambda \mid \lambda \geqslant 0\}$
- 3. $p_x(k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$
- 4. $\boldsymbol{\mathcal{L}} = f(k_1, k_2, ..., k_n; \lambda) = \prod_{i=1}^{n} \dfrac{\lambda^{k_i}}{k_i!} e^{-\lambda}$

Maximum Likelihood Estimation (MLE): Example 3 Solution

$$\mathcal{L} = f(k_1, k_2, ..., k_n; \lambda) = \frac{\lambda^{k_1}}{k_1!}e^{-\lambda} \cdot \frac{\lambda^{k_2}}{k_2!}e^{-\lambda} \cdot \frac{\lambda^{k_3}}{k_3!}e^{-\lambda} \cdots \frac{\lambda^{k_n}}{k_n!}e^{-\lambda} = \frac{\lambda^{k_1+k_2+\cdots+k_n}}{k_1!k_2!\cdots k_n!}e^{-n\lambda}$$

$$\begin{aligned}\ell &= \ln \frac{\lambda^{k_1+k_2+\cdots+k_n}}{k_1!k_2!\cdots k_n!}e^{-n\lambda} = \ln \lambda^{k_1+k_2+\cdots+k_n} - \ln(k_1!k_2!\cdots k_n!) + \ln e^{-n\lambda} \\ &= (k_1 + k_2 + \cdots + k_n)\ln\lambda - n\lambda - \ln(k_1!k_2!\cdots k_n!)\end{aligned}$$

Now, we let:

$$\frac{d\ell}{d\lambda} = 0 \Rightarrow \frac{k_1 + k_2 + \cdots + k_n}{\lambda} - n = 0 \Rightarrow \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} k_i$$

Maximum Likelihood Estimation (MLE): Example 4

- **Example 4 $Oct4$ expression in embryonic stem cells**: Let the random variable $\boldsymbol{X}$ be the expression values of $Oct4$. We know $\boldsymbol{X} \sim \mathcal{N}(\mu, \sigma)$, but we μ and σ are unknown. Now we have sequenced 5 cells, and the expressions of $Oct4$ in those cells are $3.0,\ 3.5,\ 2.5,\ 3.2,\ 2.8$, respectively. What is the MLE for the parameters of this normal distribution?

- 1. $\theta : \mu, \sigma$
- 2. $\boldsymbol{\Omega} : \{(\mu, \sigma) \mid -\infty < \mu < +\infty,\ \sigma \geqslant 0\}$
- 3. $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- 4. $\boldsymbol{\mathcal{L}} = f(3.0,\ 3.5,\ 2.5,\ 3.2,\ 2.8; \mu, \sigma) = \prod_{i=1}^{5} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$

Maximum Likelihood Estimation (MLE): Example 4 Solution

$$
\begin{aligned}
\mathcal{L} &= f(3.0,\ 3.5,\ 2.5,\ 3.2,\ 2.8; \mu, \sigma) \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{(3.0-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(3.5-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(2.5-\mu)^2}{2\sigma^2}} \\
&\cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(3.2-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(2.8-\mu)^2}{2\sigma^2}} \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^5 \cdot e^{-\frac{(3.0-\mu)^2+(3.5-\mu)^2+(2.5-\mu)^2+(3.2-\mu)^2+(2.8-\mu)^2}{2\sigma^2}}
\end{aligned}
$$

Maximum Likelihood Estimation (MLE): Example 4 Solution

$$\boldsymbol{\ell} = \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^5 + \ln e^{-\dfrac{5\mu^2 - 30\mu + 45.58}{2\sigma^2}}$$

$$= 5\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{5\mu^2 - 30\mu + 45.58}{2\sigma^2} = -\frac{5}{2\sigma^2}\cdot\mu^2 + \frac{15}{\sigma^2}\cdot\mu - \frac{45.58}{2\sigma^2} + 5\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right)$$

$$= -5\cdot\ln\sigma - \frac{5\mu^2 - 30\mu + 45.58}{2}\cdot\sigma^{-2} - 5\ln\sqrt{2\pi}$$

Now, we can find out the values of μ and σ by letting:

$$\frac{\partial\boldsymbol{\ell}}{\partial\mu} = 0 \text{ and } \frac{\partial\boldsymbol{\ell}}{\partial\sigma} = 0 \Rightarrow \hat{\mu} = 3 \text{ and } \hat{\sigma} = 0.116$$