

Emory University

**MATH 362 Mathematical Statistics II**

Learning Notes

Jiuru Lyu

January 30, 2024

**Contents**

<b>1</b>	<b>Estimation</b>	<b>2</b>
1.1	Introduction . . . . .	2
1.2	The Method of Maximum Likelihood and the Method of Moments . . . . .	3

# 1 Estimation

## 1.1 Introduction

**Definition 1.1.1 (Model).** A *model* is a distribution with certain parameters.

**Example 1.1.2** The normal distribution:  $N(\mu, \sigma^2)$ .

**Definition 1.1.3 (Population).** The *population* is all the objects in the experiment.

**Definition 1.1.4 (Data, Sample, and Random Sample).** *Data* refers to observed value from sample. The *sample* is a subset of the population. A *random sample* is a sequence of independent, identical (*i.i.d.*) random variables.

**Definition 1.1.5 (Statistics).** *Statistics* refers to a function of the random sample.

**Example 1.1.6** The sample mean is a function of the sample:

$$\bar{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n).$$

**Example 1.1.7 Central Limit Theorem**

We randomly toss  $n = 200$  fair coins on the table. Calculate, using the central limit theorem, the probability that at least 110 coins have turned on the same side.

$$\bar{X} = \frac{X_1 + \cdots + X_{200}}{200} \stackrel{\text{CLT}}{\sim} N(\mu, \sigma^2),$$

where

$$\mu = \mathbf{E}(\bar{X}) = \frac{\sum_{i=1}^{200} \mathbf{E}(X_i)}{200},$$

$$\sigma^2 = \mathbf{Var}(\bar{X}) = \mathbf{Var}\left(\frac{X_1 + \cdots + X_{200}}{200}\right) = \frac{\sum_{i=1}^{200} \mathbf{Var}(X_i)}{200^2}.$$

**Definition 1.1.8 (Statistical Inference).** The process of *statistical inference* is defined to be the process of using data from a sample to gain information about the population.

**Example 1.1.9** Goals in statistical inference

1. **Definition 1.1.10 (Estimation).** To obtain values of the parameters from the data.
2. **Definition 1.1.11 (Hypothesis Testing).** To test a conjecture about the parameters.
3. **Definition 1.1.12 (Goodness of Fit).** How well does the data fit a given distribution.
4. Linear Regression

**1.2 The Method of Maximum Likelihood and the Method of Moments**

**Example 1.2.1** Given an unfair coin, or  $p$ -coin, such that

$$X = \begin{cases} 1 & \text{head with probability } p, \\ 0 & \text{tail with probability } 1 - p. \end{cases}$$

How can we determine the value  $p$ ?

**Solution 1.**

1. Try to flip the coin several times, say, three times. Suppose we get HHT.
2. Draw a conclusion from the experiment.

**Key idea:** The choice of the parameter  $p$  should be the value that maximizes the probability of the sample.

$$\mathbf{P}(X_1 = 1, X_2 = 1, X_3 = 0) = \mathbf{P}(X_1 = 1)\mathbf{P}(X_2 = 1)\mathbf{P}(X_3 = 0) = p^2(1 - p) := f(p).$$

Solving the optimization problem  $\max_{p>0} f(p)$ , we find it is most likely that  $p = \frac{2}{3}$ . This method is called the *likelihood maximization method*. □

**Definition 1.2.2 (Likelihood Function).** For a random sample of size  $n$  from the discrete (or continuous) pdf  $p_X(k; \theta)$  (or  $f_Y(y; \theta)$ ), the *likelihood function*,  $L(\theta)$ , is the product of the pdf evaluated at  $X_i = k_i$  (or  $Y_i = y_i$ ). That is,

$$L(\theta) := \prod_{i=1}^n p_X(k_i; \theta) \quad \text{or} \quad L(\theta) := \prod_{i=1}^n f_Y(y_i; \theta).$$

**Definition 1.2.3 (Maximum Likelihood Estimate).** Let  $L(\theta)$  be as defined in Definition 1.2.2. If  $\theta_e$  is a value of the parameter such that  $L(\theta_e) \geq L(\theta)$  for all possible values of  $\theta$ , then we call  $\theta_e$  the *maximum likelihood estimate* for  $\theta$ .

**Theorem 1.2.4 The Method of Maximum Likelihood**

Given random samples  $X_1, \dots, X_N$  and a density function  $p_X(x)$  (or  $f_X(x)$ ), then we have the likelihood function defined as

$$\begin{aligned} L(\theta) &= p_X(X; \theta) = \mathbf{P}(X_1, X_2, \dots, X_N) \\ &= \mathbf{P}(X_1)\mathbf{P}(X_2) \cdots \mathbf{P}(X_N) && [\text{independent}] \\ &= \prod_{i=1}^N p_X(X_i; \theta) && [\text{identical}] \end{aligned}$$

Then, the maximum likelihood estimate for  $\theta$  is given by

$$\theta^* = \arg \max_{\theta} L(\theta),$$

where

$$L\left(\arg \max_{\theta} L(\theta)\right) = L^*(\theta) = \max_{\theta} L(\theta).$$

**Example 1.2.5** Consider the Poisson distribution  $X = 0, 1, \dots$ , with  $\lambda > 0$ . Then, the pdf is given by

$$p_X(k, \lambda) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots$$

Given data  $k_1, \dots, k_n$ , we have the likelihood function

$$L(\lambda) = \prod_{i=1}^n p_X(X = k_i; \lambda) = \prod_{i=1}^n e^{-\lambda} = \frac{\lambda^{k_i}}{k_i!} = e^{-n\lambda} \frac{\lambda^{\sum k_i}}{k_1! \cdots k_n!}$$

Then, to find the maximum likelihood estimate of  $\lambda$ , we need to  $\max_{\lambda} L(\lambda)$ . That is to solve

$$\frac{\partial L(\lambda)}{\partial \lambda} = 0 \text{ and } \frac{\partial^2 L(\lambda)}{\partial \lambda^2} < 0.$$

**Example 1.2.6** Waiting Time.

Consider the exponential distribution  $f_Y(y) = \lambda e^{-\lambda y}$  for  $y \geq 0$ . Find the MLE  $\lambda_e$  of  $\lambda$ .

**Solution 2.**

The likelihood function of the exponential distribution is given by

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda y_i} = \lambda^n \exp \left( -\lambda \sum_{i=1}^n y_i \right).$$

Now, define

$$\ell(\lambda) = \ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^n y_i.$$

To optimize  $\ell(\lambda)$ , we compute

$$\frac{d}{d\lambda} \ell(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n y_i \stackrel{set}{=} 0$$

So,

$$\frac{n}{\lambda} = \sum_{i=1}^n y_i \implies \lambda_e = \frac{n}{\sum_{i=1}^n y_i} =: \frac{1}{\bar{y}},$$

where  $\bar{y}$  is the sample mean. □

**Example 1.2.7** Given the exponential distribution  $f_Y(y) = \lambda e^{-\lambda y}$  for  $y \geq 0$ . Find the MLE of  $\lambda^2$ .

**Solution 3.**

Define  $\tau = \lambda^2$ . Then,  $\lambda = \sqrt{\tau}$ , and so

$$f_Y(y) = \sqrt{\tau} e^{-\sqrt{\tau} y}, \quad y \geq 0.$$

Then, the likelihood function becomes

$$L(\tau) = \prod_{i=1}^n f_Y(y) = \tau^{\frac{n}{2}} \exp \left( -\sqrt{\tau} \sum_{i=1}^n y_i \right).$$

Similarly, after maximization, we find

$$\tau_e = \frac{1}{(\bar{y})^2}.$$

□

**Theorem 1.2.8 Invariant Property for MLE**

Suppose  $\lambda_e$  is the MLE of  $\lambda$ . Define  $\tau := h(\lambda)$ . Then,  $\tau_e = h(\lambda_e)$ .

**Proof 4.** In this proof, we will prove the case when  $h$  is a one-to-one function. The case of  $h$  being a many-to-one function is beyond the scope of this course.

Suppose  $h(\cdot)$  is a one-to-one function. Then,  $\lambda = h^{-1}(\tau)$  is well-defined. Then,

$$\max_{\lambda} L(\lambda; y_1, \dots, y_n) = \max_{\tau} L(h^{-1}(\tau); y_1, \dots, y_n) = \max_{\tau} L(\tau; y_1, \dots, y_n).$$

■

**Example 1.2.9** Waiting Time with an unknown Threshold.

Let  $\lambda = 1$  in exponential but there is an unknown threshold  $\theta$ , that, is  $f_Y(y) = e^{-(y-\theta)}$  for  $y \geq \theta$ ,  $\theta > 0$ .

**Solution 5.**

Note that the likelihood function is given by

$$\begin{aligned} L(\theta; y_1, \dots, y_n) &= \prod_{i=1}^n f_Y(y_i) = \exp \left( - \sum_{i=1}^n (y_i - \theta) \right), \quad y_i \geq \theta, \theta > 0 \\ &= \exp \left( - \sum_{i=1}^n (y_i - \theta) \right) \cdot \mathbb{1}_{[y_i \geq \theta, \theta > 0]}, \end{aligned}$$

where

$$\mathbb{1}_{x \in A} = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

Using order statistics,

$$\begin{aligned} L(\theta) &= \exp \left( - \sum_{i=1}^n (y_i - \theta) \right) \cdot \mathbb{1}_{[y_{(n)} \geq y_{(n-1)} \geq \dots \geq y_{(1)} \geq \theta, \theta > 0]} \\ &= \exp \left( - \sum_{i=1}^n y_i + n\theta \right) \mathbb{1}_{[y_{(n)} \geq \dots \geq y_{(1)} \geq \theta, \theta > 0]}. \end{aligned}$$

So, we know  $\theta \leq y_{(1)} = y_{\min}$ .

To maximize the likelihood function, we want to maximize  $-\sum y_i + n\theta$ . That is, to maximize  $\theta$ , as  $\theta \leq y_{\min}$ , it must be that  $\theta_{\max} = y_{\min}$ . Therefore, the MLE is  $\theta^* = y_{\min}$ .  $\square$

**Example 1.2.10** Suppose  $Y_1, \dots, Y_n \sim \text{Uniform}[0, a]$ . That is,  $f_Y(y; a) = \frac{1}{a}$  for  $y \in [0, a]$ . Find MLE  $a_e$  of  $a$ .

**Solution 6.**

Note that

$$\begin{aligned} f_Y(y; a) &= \frac{1}{a} \cdot \mathbb{1}_{\{y \in [0, a]\}} \\ &= \frac{1}{a} \cdot \mathbb{1}_{\{0 \leq y_{(1)} \leq \dots \leq y_{(n)} \leq a\}} \end{aligned} \quad \text{where } y_{(1)} = \min y_i \text{ and } y_{(n)} = \max y_i$$

Then,

$$L(a) = \frac{1}{a^n} \mathbb{1}_{\{0 \leq y_{(1)} \leq \dots \leq y_{(n)} \leq a\}}$$

To maximize  $L(a)$ , we want to minimize  $a^n$ . Since  $a \geq y_{(n)}$ , it must be that  $a_e = y_{(n)}$ . Here, we call  $a_e = y_{(n)}$  an *estimate*, and  $\widehat{a_{\text{MLE}}} = Y_{(n)}$  an *estimator*.  $\square$

**Example 1.2.11 MLE that Does Not Exist**

Suppose  $f_Y(y; a) = \frac{1}{a}$ ,  $y \in [0, a)$ . Find the MLE.

**Solution 7.**

The likelihood function is the same:

$$L(a) = \frac{1}{a^n} \mathbb{1}_{\{0 \leq y_{(1)} \leq \dots \leq y_{(n)} < a\}}.$$

However, since  $[0, a)$  is not a closed set, the optimization problem  $\max_{a \in [0, a)} L(a)$  does not have a solution. Hence, the estimate does not exist.  $\square$

**Remark 1.1** MLE may not be unique all the time.

**Example 1.2.12 Multiple MLE Values**

Suppose  $X_1, \dots, X_n \sim \text{Uniform}\left[a - \frac{1}{2}, a + \frac{1}{2}\right]$ , where  $f_X(x; a) = 1$ ,  $x \in \left[a - \frac{1}{2}, a + \frac{1}{2}\right]$ . Find the MLE.

**Solution 8.**

In the indicator function notation, we can rewrite the pdf to be

$$f_X(x; a) = \mathbb{1}_{\{a - \frac{1}{2} \leq x \leq a + \frac{1}{2}\}} = \mathbb{1}_{\{a - \frac{1}{2} \leq x_{(1)} \leq \dots \leq x_{(n)} \leq a + \frac{1}{2}\}}.$$

So, the likelihood function will be

$$L(a) = \prod_{i=1}^n f_x(x_i; a) = \begin{cases} 1, & a \in \left[ x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2} \right] \\ 0, & \text{otherwise.} \end{cases}$$

So, the  $L(a)$  will be maximized whenever  $a \in \left[ x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2} \right]$ . Therefore, MLE can be any value in the range  $\left[ x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2} \right]$ . Say,

$$a_e = x_{(n)} - \frac{1}{2} \quad \text{or} \quad a_e = x_{(1)} + \frac{1}{2} \quad \text{or} \quad a_e = \frac{x_{(n)} - \frac{1}{2} + x_{(1)} + \frac{1}{2}}{2} = \frac{x_{(n)} + x_{(1)}}{2}.$$

□

### Theorem 1.2.13 MLE for Multiple Parameters

In general, we have the likelihood function  $L(\theta)$ , where  $\theta = (\theta_1, \dots, \theta_p)$ . To find the MLE, we need

$$\frac{\partial L(\theta)}{\partial \theta_i} = 0 \quad i = 1, \dots, p,$$

and the Hessian matrix

$$\left( \frac{\partial^2 L(\theta)}{\partial \theta_i \partial \theta_j} \right)_{i,j=1,\dots,p} := \begin{pmatrix} \frac{\partial^2 L(\theta)}{\partial \theta_1^2} & \cdots & \frac{\partial^2 L(\theta)}{\partial \theta_1 \partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 L(\theta)}{\partial \theta_p \partial \theta_1} & \cdots & \frac{\partial^2 L(\theta)}{\partial \theta_p^2} \end{pmatrix}$$

should be negative definite.

### Example 1.2.14 MLE for Multiple Parameters: Normal Distribution

Suppose  $Y_1, \dots, Y_n \sim N(\mu, \sigma)$ . Then,

$$f_{Y_i}(u; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y_i - \mu)^2 / (2\sigma^2)}.$$

Find the MLE for  $\mu$  and  $\sigma$ .

**Solution 9.**



The likelihood function will be

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

Then, we define

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

Set

$$\begin{cases} \frac{\partial \ell(\mu, \sigma)}{\partial \mu} = 0 & \textcircled{1} \\ \frac{\partial \ell(\mu, \sigma)}{\partial \sigma} = 0 & \textcircled{2} \end{cases}$$

From ①, we have

$$\begin{aligned} \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) &= 0 \\ \sum_{i=1}^n y_i &= n\mu \implies \boxed{\mu_e = \frac{\sum y_i}{n} = \bar{y}} \end{aligned}$$

From ②, by the invariant property of MLE, we instead set

$$\begin{aligned} \frac{\partial \ell(\mu, \sigma)}{\partial \sigma^2} &= 0 \\ -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2} \left( \frac{1}{\sigma^2} \right)^2 \sum_{i=1}^n (y_i - \mu)^2 &= 0 \\ \frac{1}{2\sigma^2} \left( -n + \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right) &= 0 \\ -n\sigma^2 + \sum_{i=1}^n (y_i - \mu)^2 &= 0 \quad (\mu_e = \bar{y}) \\ \sum_{i=1}^n (y_i - \bar{y})^2 &= n\sigma^2 \\ \sigma_e^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 &\implies \boxed{\sigma_e = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}} \end{aligned}$$

□