

# STAT 653 - Notes

## Introduction to Mathematical Statistics

September 20, 2025

### Contents

1	Statistical Model	1
2	The Likelihood Function	3
3	Identifiability of Statistical Models	4
4	Maximum Likelihood	5

## 1 Statistical Model

**Example.** A coin is tossed  $n$  times. The data available is  $X = (X_1, X_2, \dots, X_n)$ , where  $X_i \in \{0, 1\}$ . The assumptions are:

1. outcomes are independent.
2.  $P(X_i = 1) = \theta \in \Theta$  where  $\theta$  is an unknown parameter and  $\Theta$  is the parameter space. In this case  $\Theta = [0, 1]$ .

We need to estimate  $\theta$  based on the data  $X = (X_1, X_2, \dots, X_n)$ , where  $X_i$  are random variables before the experiment is conducted.

So we need to find an estimator  $T(X_1, X_2, \dots, X_n)$  of  $\theta \in \Theta$ .

### Possible Estimators

1.  $T_1 := T_1(X_1, X_2, \dots, X_n) = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$

**Remark.** (a)  $\mathbb{E}(T_1) = \mathbb{E}(\bar{X}_n) = \mathbb{E}(X_1) = \theta$  for all  $\theta \in \Theta$  then  $T_1$  is unbiased estimator of  $\theta$ .

(b)  $\lim_{n \rightarrow \infty} P(|\bar{X}_n - \theta| > \epsilon) = 0$  for all  $\epsilon > 0$ .

**Definition.** In general, if  $\lim_{n \rightarrow \infty} P(|T(X_1, \dots, X_n) - \theta| \epsilon) = 0$  for all  $\epsilon > 0$  and for all  $\theta \in \Theta$ , then we call  $T(X_1, \dots, X_n)$  **consistent**.

2.  $T_2(X_1, \dots, X_n) := X_1$ , where  $X_1 \in \{0, 1\}$ . Then  $\mathbb{E}(T_2) = \mathbb{E}(X_1) = \theta$  for all  $\theta \in \Theta$ .

$T_2$  is unbiased but is not consistent.

3.

$$T_3 := T_3(X_1, \dots, X_n) \\ = \sqrt{\frac{1}{\lfloor \frac{n}{2} \rfloor} \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} X_{2i} X_{2i-1}}$$

$T_3$  is biased because

$$\mathbb{E}(T_3) \leq \sqrt{\frac{1}{\lfloor \frac{n}{2} \rfloor} \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} X_{2i} X_{2i-1}} \\ = \theta \quad \forall \theta \in \Theta$$

**Example.** Suppose  $X_1, X_2, \dots, X_n$  are independent and have uniform $[0, \theta]$ , where  $\theta \in \Theta = \mathbb{R}_+$ . So  $\Theta = \{\theta : \theta > 0\}$ .

### Possible Estimators

1.  $T_1(X_1, \dots, X_n) = 2\bar{X}_n$
2.  $T_2(X_1, \dots, X_n) = X_{(n)}$  (max)
3.  $T_3(X_1, \dots, X_n) = c_n X_{(n)}$

Correct the max by a constant so it is unbiased.

**Example.** We want to receive a shipment of oranges and suspect that part of them rot off. To check the shipment we draw a random sample without replacement of size  $n$  from the shipment (population) of size  $N$ .

Let  $\theta$  be the proportion of bad oranges in the population. So  $\Theta = \{\frac{0}{N}, \frac{1}{N}, \dots, \frac{N}{N}\}$ .

Let

$$X_i = \begin{cases} 0 & \text{if good} \\ 1 & \text{if bad} \end{cases}$$

for  $i = 1, 2, \dots, n$  and let  $X = (X_1, X_2, \dots, X_n)$ .

Let  $T_1(X) = \sum_{i=1}^n X_i$ . Then  $T_1$  has a hypergeometric distribution. So

$$P_\theta(X_1 = k) = \frac{\binom{N\theta}{k} \binom{N-N\theta}{n-k}}{\binom{N}{n}}$$

for  $k \in \{\max(0, n - (N - N\theta), \dots, \min(n, N\theta))\}$

## 2 The Likelihood Function

$$X \sim P_\theta, \quad \theta \in \Theta$$

We have 2 cases for now (discrete and continuous):

(R1)  $P_\theta$  is defined by a joint pdf  $f_X(x; \theta)$  for all  $\theta \in \Theta$ .

(R2)  $P_\theta$  is defined by a joint pmf  $P(X = x; \theta)$  for all  $\theta \in \Theta$ .

**Definition.** Let  $P_\theta$ ,  $\theta \in \Theta$  be a model satisfying (R1) or (R2). Then the function

$$L(x; \theta) = \begin{cases} f_X(x; \theta) & \text{if (R1)} \\ P(X = x; \theta) & \text{if (R2)} \end{cases}.$$

**Example.** Not (R1) and not (R2).

Let

$$X \sim N(\theta, 1) \quad \theta \in \Theta = \mathbb{R}.$$

We observe  $Y = \max(0, X)$ ,

$$Y = \begin{cases} 0 & \text{if } X \leq 0 \\ X & \text{if } X > 0 \end{cases} = XI(X > 0)$$

where  $I(\cdot)$  is the indicator function.

$$F_\theta(t) = P(Y \leq t) \text{ for all } t \in \mathbb{R}.$$

**Example.** Back to oranges example where  $X = (X_1, X_2, \dots, X_n)$  is the data and

$\Theta = \{\frac{0}{N}, \frac{1}{N}, \dots, \frac{N}{N}\}$ . Let  $T(X) = \sum_{i=1}^n X_i$ . Then

$$\begin{aligned} L(x; \theta) &= P_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= P_\theta \left( X_1 = x_1, X_2 = x_2, \dots, X_n = x_n, T(X) = \sum_{i=1}^n x_i \right) \\ &= P_\theta \left( T(X) = \sum_{i=1}^n x_i \right) P \left( X_1 = x_1, \dots, X_n = x_n \mid T(X) = \sum_{i=1}^n x_i \right). \end{aligned}$$

Now define  $K_n = \sum_{i=1}^n x_i$ . For example, if  $n = 5$  and we observed  $(1, 0, 0, 1, 1)$  then

$$K = \sum_{i=1}^5 x_i = 3.$$

Since there are 10 possibilities for which entries are 1 versus 0,  $\binom{5}{3} = 10$ . Because all possible combinations of 1 and 0 are possible we can use symmetry to calculate the probability of any particular sequence of 1 and 0 as  $1/\binom{5}{3}$ . We use this reasoning below to derive the expression on the right.

Then

$$L(x; \theta) = \frac{\binom{N\theta}{K_n} \binom{N-N\theta}{n-K_n}}{\binom{N}{n}} \times \frac{1}{\binom{n}{K_n}}.$$

### 3 Identifiability of Statistical Models

**Definition.** Let  $X \sim P_\theta$ ,  $\theta \in \Theta$ . A model  $P_\theta$ ,  $\theta \in \Theta$  is identifiable if for any pair  $(\theta, \theta')$  such that  $\theta \neq \theta'$  and  $\theta, \theta' \in \Theta$ , then  $P_\theta \neq P_{\theta'}$ .

**Remark.** This means that there is an event  $A$ , such that  $P_\theta(A) \neq P_{\theta'}$  where  $\theta \neq \theta'$ .

R(1) For  $\theta \neq \theta'$ ,  $f(x; \theta) \neq f(x; \theta')$  for any neighborhood of  $x$  (an open ball  $B(x, r)$  centered at  $x$ ).

By open ball we mean  $B(x, r) = \{y : |x - y| < \epsilon\}$  where  $|v| = (\sum_{i=1}^n v_i^2)^{1/2}$  (euclidean norm).

R(2) Discrete support, for some  $x$   $P_\theta(X = x) \neq P_{\theta'}(X = x)$  where  $\theta \neq \theta'$ .

**Example.** Suppose we observe  $X_1, X_2, \dots, X_n$  where  $X_i = \theta \cdot Z_i \sim N(0, \theta^2)$  and  $Z_i \sim N(0, 1)$  and  $\theta \in \Theta = \mathbb{R} \setminus \{0\}$ .

If  $\theta_1 = 1 \neq -1 = \theta_2$ , then

$$L(x_1, x_2, \dots, x_n; \theta = 1) = L(x_1, x_2, \dots, x_n; \theta = -1)$$

for any  $x = (x_1, \dots, x_n)$ .

**Result.** The model  $\{P_\theta, \theta \in \Theta\}$  is identifiable if there exists a statistic  $T(X)$  ( $X \sim P_\theta, \theta \in \Theta$ ) where expectation is a one-to-one function of  $\theta \in \Theta$ , i.e., such that

$$\forall(\theta, \theta'), \quad \theta \neq \theta' \implies \mathbb{E}_\theta(T(X)) \neq \mathbb{E}_{\theta'}(T(X)) \quad (1)$$

*Proof.* We use proof by contradiction. Suppose that (1) holds, but there exists  $\theta \neq \theta'$  such that  $P_\theta = P_{\theta'}$ . If so, then  $\mathbb{E}_\theta(T(X)) = \mathbb{E}_{\theta'}(T(X))$ , which contradicts (1).  $\square$

In the previous example,  $\theta = 1$ ,  $\theta' = -1$ .

**Example.** Let  $X_1, X_2, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bernoulli}(\theta)$  where  $\theta \in \Theta = [0, 1]$ . We will show that  $\theta$  is identifiable using the definition and also the above result.

Let  $\theta$  and  $\theta'$  be arbitrary and suppose  $\theta \neq \theta'$  and  $\theta', \theta \in \Theta$ . Also suppose  $X = (1, 1, \dots, 1)$ . Then

$$\begin{aligned} P_\theta(X_1, X_2, \dots, X_n) &= \theta^n \\ P_{\theta'}(X_1 = 1, \dots, X_n) &= (\theta')^n. \end{aligned}$$

Since  $\theta \in [0, 1]$  then  $\theta^n \neq (\theta')^n$  and the model is identifiable.

Now take a statistic  $T(X_1, \dots, X_n) = X_1$  (or we could take  $T(X_1, \dots, X_n) = \sum_{i=1}^n X_i$  or  $T(X_1, \dots, X_n) = \sum_{i=1}^n \bar{X}_n$ ).

For any  $(\theta, \theta') \in \Theta$ , if  $\theta \neq \theta'$  then  $\mathbb{E}_\theta(\bar{X}_n) = \theta \neq \theta' = \mathbb{E}_{\theta'}(\bar{X}_n)$ . Then by the above result the model is identifiable.

## 4 Maximum Likelihood

**Example.** Let  $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{uniform}[0, \theta]$ . Find the MLE for  $\theta$ .