STAT 653 - Notes
Introduction to Mathematical Statistics

September 19, 2025

Contents

1 Statistical Model **1**

2 The Likelihood Function **3**

3 Maximum Likelihood **4**

1 Statistical Model

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

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

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

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

Possible Estimators

1. $T_1 := T_1(X_1, X_2, \ldots, X_n) = \overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$

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

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

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

2. $T_2(X_1, \ldots, X_n) := X_1$, 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 <u>consistent</u>.

3.

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

   T_3 is biased because

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

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

<u>**Possible Estimators**</u>

1. $T_1(X_1, \ldots, X_n) = 2\overline{X}_n$
2. $T_2(X_1, \ldots, X_n) = X_{(n)}$ (max)
3. $T_3(X_1, \ldots, X_n) = c_n X_{(n)}$
   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 θ be the proportion of bad oranges in the population. So $\Theta = \{\frac{0}{N}, \frac{1}{N}, \ldots, \frac{N}{N}\}$.

Let

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

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

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

2 The Likelihood Function

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

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

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

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

Definition. Let P_θ, $\theta \in \Theta$ 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

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

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

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

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

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

Example. Back to oranges example where $X = (X_1, X_2, \ldots, X_n)$ is the data and $\Theta = \{\frac{0}{N}, \frac{1}{N}, \ldots, \frac{N}{N}\}$. Let $T(X) = \sum_{i=1}^{n} X_i$. Then

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

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

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

Then

$$L(x; \theta) = \frac{\binom{N\theta}{K_n}\binom{N-N\theta}{n-K_n}}{\binom{N}{n}} \cdot \left[\frac{1}{\binom{n}{K_n}} \right].$$

3 Maximum Likelihood

Example. Let $X_1, \ldots, X_n \overset{iid}{\sim} uniform[0, \theta]$. Find the MLE for θ.