§5.2 Estimating Parameters (Maximum Likelihood)

Statistics Question: Suppose a random variable X has a **pdf** with unknown parameters. We need to estimate the unknown parameter using a random sample of n observations.

Example 1. Suppose we have a biased coin with unknown probability $\theta = P(\text{head})$. We know that it satisfies the Bernoulli distribution $Bernoulli(\theta)$.

We toss it 10 times, and get $HTHHTHHTHT$ (6 head, 4 tails).

It is natural to estimate $\theta = 6/10$.

What is the theory behind the estimation? The method is called **Maximum likelihood.** The best estimation for θ from the sample data(observations), is to make maximize the "likelihood" of getting the sample.

Definition.

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from a discrete **pdf** $p_X(x;\theta)$, or a continuous **pdf** $f_X(x;\theta)$.
The **likelihood function** $L(\theta)$ is

$$L(\theta) = \prod_{i=1}^{n} p_X(x_i;\theta) \qquad \text{for discrete case}$$

or

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i;\theta) \qquad \text{for continuous case}$$

The likelihood function $L(\theta)$ is a function with variable θ. The purpose is to find a value $\theta = \theta_e$ maximizing the function $L(\theta)$.

Example 1 (Continue).

$$L(\theta) = P(H)P(T)P(H)P(H)P(T)\cdots P(T) = \theta^6(1-\theta)^4$$

We want to maximize $L(\theta)$. (By Calculus 1, use derivative to find critical points)

$$\frac{dL(\theta)}{d\theta} = 6\theta^5(1-\theta)^4 - 4\theta^6(1-\theta)^3 = \theta^5(1-\theta)^3(6-10\theta) = 0$$

So, $\theta_e = 6/10 = 0.6$ maximize the likelihood function $L(\theta)$.

> **Remark:** $\ln(L(\theta))$ and $L(\theta)$ have the same critical points. Sometimes, it is easy to use $\ln(L(\theta))$ to find critical points.

As in Example 1,

$$\ln(L(\theta)) = \ln(\theta^6(1-\theta)^4) = 6\ln(\theta) + 4\ln(1-\theta).$$

So, $\dfrac{\ln(L(\theta))}{d\theta} = \dfrac{6}{\theta} - \dfrac{4}{1-\theta} = 0$. Hence, the critical point is $\theta_e = 6/10$.

Example 2. (Practice) Suppose we have a biased die with unknown probability of resulting "6", $\theta = P(\text{"6"})$.

We toss it 8 times, and get $* \, 6 * * * 6 * *$ (where $*$ means not 6). It is natural to estimate $\theta = 2/8$.

Use the method of maximum likelihood to verify your answer.

Example 3. Let X be the number of incoming calls in an hour. Recall that X fits a Poisson distribution

$$p_X(k;\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}, \text{ for } k = 0, 1, 2, ...$$

Here, λ is the unknown parameter.

Use the method of maximum likelihood to estimate λ based on $n = 4$ observations $X_1 = k_1,\ X_2 = k_2,\ X_3 = k_3,\ X_4 = k_4$.

Example 4. Suppose we have a biased (unfair) coin with $P(head) = p$. Let X be the number of tosses until we get a head. So, the **pdf** of X is the geometric distribution:

$$p_X(k) = (1-p)^{k-1}p.$$

Suppose we do the experiments n times and get $X_1 = k_1, X_2 = k_2, ..., X_n = k_n$. Use the method of maximum likelihood to estimate p.

Example 5. Suppose we have a distribution

$$f_X(x) = \theta e^{-\theta(x+1)} \quad \text{for } x \geq -1.$$

Find the maximum likelihood estimator for θ based on a sample $X_1 = x_1$, $X_2 = x_2$, ... , $X_n = x_n$.

Example 6. The **pdf** of a uniform distribution on $[0, \theta]$ is given by

$$f_X(x) = \frac{1}{\theta} \quad \text{for } x \in [0, \theta].$$

Take a sample $X_1 = x_1$, $X_2 = x_2$, ... , $X_n = x_n$. Use the method of maximum likelihood to estimate θ.

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i; \theta) = \frac{1}{\theta^n}$$

This is a decreasing function, so it does not have a maximum.

We know that $x_i \leq \theta$ for all $i = 1, 2, ..., n$. Hence $\theta \geq \max(x_1, x_2, \cdots, x_n)$.

So, if we want to maximize $L(\theta)$, we need to choose $\theta_e = \max(x_1, x_2, \cdots, x_n)$