Example :

Suppose that, in a particular city, airport A handles 60% of all airline traffic, and airport B handles 40% of all airline traffic. The detection rates for weapons at the airports A and B are 0.9 and 0.8, respectively.

(1) If a passenger at one of the airports is found to be carrying a weapon through the boarding gate, what is the probability that the passenger is using airport A?

(2) If 30 passengers at one of the airports are found to be carrying a weapon through the boarding gate and they are independent, what is the probability that at least 17 of them are using the airport A

Solution:

(1) Define

A : passenger uses airport A.

B : passenger uses airport B.

D : a weapon is detected.

According to the information, we have

$$\mathbb{P}(D|A) = 0.9, \;\; \mathbb{P}(A) = 0.6, \;\; \mathbb{P}(B) = 0.4, \;\; \mathbb{P}(D|B) = 0.8$$

By Bayes' rule, we have

$$\mathbb{P}(A|D) = \frac{\mathbb{P}(A)\mathbb{P}(D|A)}{\mathbb{P}(A)\mathbb{P}(D|A) + \mathbb{P}(B)\mathbb{P}(D|B)} = 0.627$$

Let X be the number of passengers carrying a weapon and using airport A. According to the information, the success probability $p = 62.7\%$

(I) $np = 30(0.627) = 18.81 > 5$ **and** $nq = 30(0.373) = 11.19 > 5$**.**

(II) $\mu = np = 18.81$, $\sigma = \sqrt{npq} = \sqrt{30(0.627)(0.373)} = 2.648$

(III) The question is just to find the probability $\mathbb{P}(X \geq 17)$. Ac-

cording to (I), we have

$$\mathbb{P}(X \geq a) = \mathbb{P}(Z \geq \frac{a - 0.5 - np}{\sqrt{npq}})$$

$$\mathbb{P}(X \geq 17) = \mathbb{P}(Z \geq \tfrac{17-0.5-18.81}{2.648}) = \mathbb{P}(Z \geq -0.872)$$
$$= \mathbb{P}(-\infty \leq Z \leq 0.872) = 0.8078$$

Starting from this chapter, we begin to talk about statistics. In statistics, we often need to find a probability distribution to fit the observed data distribution. Then, we can make inference about the random experiment. One typical problem of such kind is that we know for sure that the data has one distribution, but we don't know the parameters of the distribution. For example, we know that the data has normal distribution $N(\mu, \sigma)$. However, we don't know what are the μ and σ. An obvious way is using the data to estimate them. Now, we discuss one popular parameter estimation method which is called Maximum Likelihood estimation. Before discussing this method, we need to introduce some basic definitions.
Definition: A sequence of independent rv's $X_1, \cdots, X_n$ which are defined on a common probability space $(\Omega, \mathbb{P})$ and they have identical distribution is called a random sample of size n. An outcome or observation of $X_1, \cdots, X_n$, denoted by $x_1, \cdots, x_n$, is called a sample outcome or simply a sample.
Example: Consider an experiment of flipping a fair coin 5 times. Define

$$X_i = \begin{cases} 1 & \text{if observing head at the } i^{th} \text{ flipping} \\ 0 & \text{if observing tail at the } i^{th} \text{ flipping} \end{cases}$$

Then, $X_1, \cdots, X_5$ is a random sample. $1, 0, 0, 1, 1$ is a sample outcome.
In order to be easier to understand the idea. Let us start by a simple example which can show the idea clearly.
Example: In a box, there are 100 balls which are identical except their colors. Each ball has either white color or black color. We know that 99 balls have same color and the rest one ball has a different color. However, we don't know which color the 99 balls have. Now, we randomly draw a ball. If the observed ball is white, can you guess which color the 99 balls have? Why?
Solution: Certainly we will guess that the 99 balls have white color. The reason is as follows: If in the box there are 99 white balls and only one black ball, the probability of drawing a white ball is much bigger than the probability of drawing a black ball. In other words, a white ball has much bigger chance to be drawn. Inversely, observed one should have bigger chance to be drawn. Thus, we should find the estimator(s) of the parameter(s) such that the sample outcomes have biggest probability to be observed. This is just the basic idea of the maximum likelihood estimation.
Definition: Let $X_1, \cdots, X_n$ be a random sample from $f(x, \theta)$, where θ is an unknown parameter. The likelihood function, $L(\theta)$, is defined as follows:

$$L(\theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

where $x_1, \cdots, x_n$ is the sample outcomes. Thus, maximum likelihood method is to find a $\hat{\theta}$ which maximizes $L(\theta)$. The $\hat{\theta}$ is called the estimator of parameter θ.
Example: In a random experiment that a biased coin is flipped

n **times. The outcomes is** $(k_1, \cdots, k_n)$**, where**

$$k_i = \begin{cases} 1 & \textbf{If a head was observed in the } i^{th} \textbf{ flipping} \\ 0 & \textbf{If a tail was observed in the } i^{th} \textbf{ flipping} \end{cases}$$

Estimate p**, the probability of observing a head by the MLE.**
Solution:Define

$$X_i = \begin{cases} 1 & \textbf{If a head comes up in the } i^{th} \textbf{ flipping} \\ 0 & \textbf{If a tail comes up in the } i^{th} \textbf{ flipping} \end{cases}$$

Note that p **is the probability of observing a head. Thus,**

$$\begin{aligned} \mathbb{P}(X_1 = k_1, \cdots, X_n = k_n) &= p^{k_1}(1-p)^{(1-k_1)} \cdots p^{k_n}(1-p)^{(1-k_n)} \\ &= p^k(1-p)^{(n-k)}, \end{aligned}$$

where $k = \sum_{i=1}^n k_i$**. Since observed should have biggest probability, we should find** p **such that** $\mathbb{P}(X_1 = k_1, \cdots, X_n = k_n)$ **is maximized. To this purpose, from calculus, we know that the maximum point is the ones such that the derivative is equal to zero.**

$$\frac{d}{dp}[p^k(1-p)^{(n-k)}] = k[p^{(k-1)}(1-p)^{(n-k)}] + (k-n)[p^k(1-p)^{(n-k-1)}$$

This equation can be reduced to

$$k(1-p) + (k-n)p = 0$$

which has solution k/n**. Thus, we should use** k/n **to estimate the parameter** p**.**

Remark: We know that $f(p) = p^k(1-p)^{(n-k)}$ **attains its maximum at** p_0 **if and only if** $\ln f(p)$ **attains its maximum at** p_0**. Since**

$$\frac{d}{dp}\ln[p^k(1-p)^{(n-k)}] = \frac{k}{p} - \frac{n-k}{1-p}$$

and

$$\frac{d^2}{dp^2}\ln[p^k(1-p)^{(n-k)}] = -\frac{k}{p^2} - \frac{n-k}{(1-p)^2} \leq 0.$$

Set

$$\frac{d}{dp}\ln[p^k(1-p)^{(n-k)}] = 0.$$

We get $p = k/n$**. By calculus,** $f(p)$ **attains its maximum at** $p_0 = k/n$**.**

Example: Consider a problem of estimation of total number of fish, say n**, in a lake. Design a strategy, then using the maximum likelihood method to estimate** n**.**

Solution: We randomly catch 200 **fish. After marking each fish, we return all fish to the lake. After one day, we catch fish** 100 **times successively with replacement. We found that there are** 1 **fish with the marks and** 99 **without the marks. Since there are** 200 **fish with marks, the probability to catch a marked fish is** $200/n$**. This is a binomial experiment with successive probability** $p = 200/n$ **and failure probability** $1-p$**. The probability that catching** 100 **times, there are** 1 **marked fish is**

$$C_1^{100}p^1(1-p)^{100-1} = C_1^{100}p(1-p)^{99}$$

According the idea of maximum likelihood estimation, we should find n **or** p **which maximizes the probability**

$$C_1^{100}p(1-p)^{99}$$

or p **maximizes** $f(p) := p(1-p)^{99}$**. Let**

$$\frac{d}{dp}f(p) = (1-p)^{99} - 99p(1-p)^{98} = 0$$

we get

$$(1-p)^{99} = 99p(1-p)^{98}$$

or $1-p=99p$**. Thus,** $p=1/100$ **and** $n=20,000$**.**