Maximum Likelihood Estimation (MLE)

Suppose we have a random sample $X_1, X_2, \ldots, X_n$ whose assumed probability distribution depends on some unknown parameter θ.

For binomial distribution; unknown parameters n, p

For Poisson distribution; unknown parameter λ

For Geometric distribution unknown parameter p

For Normal distribution; unknown parameters μ, σ^2

Maximum Likelihood Estimation (MLE)

Suppose we have a random sample $X_1, X_2, \dots, X_n$ whose assumed probability distribution depends on some unknown parameter θ.

Our primary goal here will be to find a point estimator u,

such that $u(x_1, x_2, \dots, x_n)$ is

a "good" point estimate of θ,

where $x_1, x_2, \dots, x_n$ are the observed values of the random sample.

i.e., our goal will be to find a good estimate of θ,

using the data $x_1, x_2, \dots, x_n$

that we obtained from our specific random sample.

Maximum Likelihood Estimation (MLE)

Maximum Likelihood Estimation (MLE) is a technique used for estimating the parameters of a given distribution, using some observed data.

For example, if a population is known to follow a normal distribution but the mean and variance are unknown,

Colle height of student (m & v)
university

the mean and variance are unknown, MLE can be used to estimate them using a limited sample of the population, by finding particular values of the mean and variance so that the observation is the most likely result to have occurred.

Maximum Likelihood Estimation (MLE)

Example: Suppose the weights of randomly selected Indian female college students are normally distributed with unknown mean μ and standard deviation σ. A random sample of 10 Indian female college students yielded the following weights (in pounds):

115 122 130 127 149 160 152 138 149 180

(MLE)

Using this, we identify the likelihood function and the maximum likelihood estimator of μ.

Maximum Likelihood Estimation (MLE)

It seems reasonable that a good estimate of the unknown parameter θ would be the value of θ that maximizes the probability, that is, the likelihood of getting the data we observed

(this is the reason, why we called as likelihood function)

Maximum Likelihood Estimation (MLE)

Definition

Let $x_1, x_2, \ldots, x_n$ be observations from n independent and identically distributed random variables drawn from a Probability Distribution that depends on some parameters θ.

The goal of MLE is to maximize the likelihood function:

for

$$L = f(x_1, x_2, \ldots, x_n | \theta) = \prod_{i=1}^{n} f(x_i | \theta)$$

joint dens

r.v

$P(A|B)$

events

$f(x_1|\theta) \cdot \times f(x_2|\theta) \cdot \ldots f(x_n|\theta)$

Maximum Likelihood Estimation (MLE)

Since logarithm is a non-decreasing function, so for maximizing L, it is equivalently correct to maximize $\log L$, i.e.,

$$\frac{1}{L}\frac{dL}{d\theta} = 0 \Rightarrow \frac{d\log L}{d\theta} = 0$$

In other words, the log-likelihood function is easier to work with:

$$\log L = \sum_{i=1}^{n} \log f(x_i|\theta)$$

Distribution can be Discrete or Continuous

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

Maximum Likelihood Estimation (MLE)

For Discrete case

The simplest case is when both the distribution and the parameter space (the possible values of the parameters) are discrete, meaning that there are a finite number of possibilities for each.

In this case, the MLE can be determined by explicitly trying all possibilities.

Maximum Likelihood Estimation (MLE)

Example: An unfair coin is flipped 100 times, and 61 heads are observed. The coin either has probability $\frac{1}{3}, \frac{1}{2}$, or $\frac{2}{3}$ of flipping a head each time it is flipped. Which of the three is the MLE?

Solution: Here the distribution is the binomial distribution with $n = 100$.

$$P\left(H = 61 \mid p = \frac{1}{3}\right) = \binom{100}{61}\left(\frac{1}{3}\right)^{61}\left(\frac{2}{3}\right)^{39} \approx 9.6 \times 10^{-9}$$

$$P\left(H = 61 \mid p = \frac{1}{2}\right) = \binom{100}{61}\left(\frac{1}{2}\right)^{61}\left(\frac{1}{2}\right)^{39} = 0.007$$

$$P\left(H = 61 \mid p = \frac{2}{3}\right) = \binom{100}{61}\left(\frac{2}{3}\right)^{61}\left(\frac{1}{3}\right)^{39} = 0.040$$

p.m.f.

$$\boxed{\begin{array}{c} P(x) = \binom{n}{x} p^x (1-p)^{n-x}\,; \\ 0 \le p \le 1 \\ x = 0,1,2,\ldots,n\,; \end{array}}$$

Since $P\left(H = 61 \mid p = \frac{2}{3}\right)$ is maximum and hence MLE is $p = \frac{2}{3}$

Maximum Likelihood Estimation (MLE)

Example: An unfair coin is flipped 100 times, and 61 heads are observed. What is the MLE when nothing is previously known about the coin?

Solution: Since the distribution follow is Binomial distribution, with parameter p. Here $n = 100$. The likelihood function (MLE) is

$$P(H = 61|p) = \binom{100}{61} p^{61}(1-p)^{39}$$

For maximization

$$\frac{d}{dp} P(H = 61|p) = 0$$

$$\Rightarrow \binom{100}{61}\left[61p^{60}(1-p)^{39} - 39p^{61}(1-p)^{38}\right] = 0$$

$$\Rightarrow p^{60}(1-p)^{38}(61 - 100p) = 0$$

$$\Rightarrow p = 0, \frac{61}{100}, 1$$

Thus, the likelihoods are

$$P(H = 61|p = 0) = 0$$

$$P\left(H = 61 \middle| p = \frac{61}{100}\right) = \binom{100}{61}\left(\frac{61}{100}\right)^{61}\left(\frac{39}{100}\right)^{39}$$

$$P(H = 61|p = 1) = 0$$

Since $P\left(H = 61 \middle| p = \frac{61}{100}\right)$ is maximum and hence $p = \frac{61}{100}$ is the MLE.

Maximum Likelihood Estimation (MLE)

Example: For a random sample $x_1, x_2, \ldots, x_n$. Assume that x_i's are independent Bernoulli random variables of the students picking a course of Statistics with unknown parameter p, find the maximum likelihood estimator of p, the proportion of students who select Statistics subject.

Solution: Define a Bernoulli random variable as

$x_i = 1$; if a randomly selected student selects a Statistics subject

$x_i = 0$; if a randomly selected student does not select a Statistics subject

Our target is to find the MLE of p

The p.m.f. of Bernoulli random variable x_i is

$$f(x_i) = p^{x_i}(1-p)^{1-x_i}$$

Maximum Likelihood Estimation (MLE)

The likelihood function L is defined as

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

$$= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$= p^{\sum x_i}(1-p)^{\sum(1-x_i)}$$

$$\Rightarrow \log L = \sum x_i \log p + \sum(1-x_i)\log(1-p)$$

$$= \log p \sum x_i + (n - \sum x_i)\log(1-p)$$

Bernoulli

Meow

p

To maximize the L, we have

$$\frac{d}{dp}\log L = 0 \Rightarrow \frac{1}{p}\sum x_i + (n-\sum x_i)\left(-\frac{1}{1-p}\right) = 0$$

$$\Rightarrow \frac{\sum x_i}{p} = \frac{n-\sum x_i}{1-p}$$

$$\Rightarrow \sum x_i - p\sum x_i = np - p\sum x_i$$

$$\Rightarrow p = \frac{\sum x_i}{n}$$

Further, $\underbrace{\frac{d^2}{dp^2}}\log L < 0$ ✓

Thus,

an estimator of p is $\frac{\sum x_i}{n}$

Maximum Likelihood Estimation (MLE)

Example: For a random sample $x_1, x_2, \dots, x_n$. Assume that x_i's are independent Binomial random variables with unknown parameter p, find the maximum likelihood estimator of p.

Solution: For binomial distribution, we have

$$p(x_i) = \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i} \quad ; \quad x_i = 0,1,2,\dots,n\,; p \in [0,1]$$

The likelihood function L is defined as

$$L = \prod_{i=1}^{n} p(x_i|\theta) = \prod_{i=1}^{n} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i}$$

Our target is to find the MLE of p

Maximum Likelihood Estimation (MLE)

The likelihood function L is defined as

$$L = \prod_{i=1}^{n} p(x_i|\theta)$$

$$= \prod_{i=1}^{n} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i}$$

$$\Rightarrow \log L = \sum_{i=1}^{n} \left[\log\binom{n}{x_i} + \log p^{x_i} + \log(1-p)^{n-x_i}\right]$$

$$\Rightarrow \log L = \sum_{i=1}^{n} \left[\log\binom{n}{x_i} + x_i \log p + (n-x_i)\log(1-p)\right]$$

$$\Rightarrow \log L = \sum_{i=1}^{n} \log\binom{n}{x_i} + \log p \sum_{i=1}^{n} x_i + \log(1-p) \sum_{i=1}^{n} (n-x_i)$$

To maximize L, we have

$$\frac{d}{dp}\log L = 0 \Rightarrow \frac{1}{p}\Sigma x_i - \frac{1}{1-p}\Sigma(n-x_i) = 0$$

$$\Rightarrow \frac{1}{p}\Sigma x_i - \frac{n^2}{1-p} + \frac{1}{1-p}\Sigma x_i = 0$$

$$\Rightarrow \frac{1}{p(1-p)}\Sigma x_i = \frac{n^2}{1-p}$$

$$\Rightarrow \frac{1}{p}\Sigma x_i = n^2$$

$$\Rightarrow p = \frac{\Sigma x_i}{n^2}$$

$$\frac{d^2}{dp^2}\log L = -\frac{1}{p^2}\Sigma x_i + \frac{1}{(1-p)^2}\Sigma(n-x_i) < 0$$

$$\boxed{\text{Hence, the MLE of } p \text{ is } \frac{\Sigma x_i}{n^2}}$$

Maximum Likelihood Estimation (MLE)

What is the likelihood?

The likelihood function measures the extent to which the data provide support for different values of the parameter. It indicates how likely it is that a particular population will produce a sample.

Maximum Likelihood Estimation (MLE) is a frequentist approach for estimating the parameters of a model given some observed data. The general approach for using MLE is:

- Observe some data.
- Write down a model for how we believe the data was generated.
- Set the parameters of our model to values which maximize the likelihood of the parameters given the data.

Prob vs Likelihood

Imagine we have some data generated from a Gaussian distribution with a variance of 4, but we don't know the mean. I like to think of MLE as taking the Gaussian, sliding it over all possible means, and choosing the mean which causes the model to fit the data best. … where we see the maximum of the log-likelihood occur at a mean …