Estimation

INTRODUCTION

The field of statistical inference consists of those methods used to make decisions or to draw conclusions about a **population.** These methods utilize the information contained in a **sample** from the population in drawing conclusions. This chapter begins our study of the statistical methods used for inference and decision making.

Statistical inference may be divided into two major areas: **parameter estimation** and **hypothesis testing .** As an example of a parameter estimation problem, suppose that a structural engineer is analyzing the tensile strength of a component used in an automobile chassis. Since variability in tensile strength is naturally present between the individual components because of differences in raw material batches, manufacturing processes, and measurement procedures (for example), the engineer is interested in estimating the mean tensile strength of the components. In practice, the engineer will use sample data to compute a number that is in some sense a reasonable value (or guess) of the true mean. This number is called a **point estimate .**We will see that it is possible to establish the precision of the estimate.

Now consider a situation in which two different reaction temperatures can be used in a chemical process, say and . The engineer conjectures that results in higher yields than does . Statistical hypothesis testing is a framework for solving problems of this type. In this case, the hypothesis would be that the mean yield using temperature is

Formulation the problem of point estimation :

Let X ~ $f(x)$, where $f(x)$ depends upon an unknown parameter θ , we want to estimate or a point estimate of θ .

Our problem is how to find (define) the statistic $Y=u(x_1, x_2, ..., x_n)$, or is how to find the real number $y=u(x_1, x_2, ..., x_n)$ such that y will be a good point estimator of θ .

Definition 4-1 : A **point estimate** of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic Y . The statistic Y is called the **point estimator .**

If we sign to the parameter by θ , then the **point estimate of** θ will be denoted by $\hat{\theta}$.

The method of point estimation :

Maximum likelihood estimation (MLE) :

It is highly desirable to have a method that is generally applicable to the construction of statistical estimators that have "good" properties. In this section we present an important method for finding estimators of parameters proposed by geneticist/statistician Sir Ronald A. Fisher around 1922 called the method of maximum likelihood. Even though the method of moments is intuitive and easy to apply, it usually does not yield "good" estimators. The method of maximum likelihood is intuitively appealing, because we attempt to find the values of the true parameters that would have most likely produced the data that we in fact observed. For most cases of practical interest, the performance of maximum likelihood estimators is optimal for large enough data. This is one of the most versatile methods for fitting parametric statistical models to data. First, we define the concept of a likelihood function .

Definition 4-2 : Let $f(x_1, x_2, ..., x_n, \theta)$, $\theta \in Y \subset R^k$, be the joint probability (or density) function of n random variables $X_1, \ldots, X_n$ with sample values $x_1, \ldots, x_n$. The **likelihood function** of the sample is given by

L $(\theta ; x_1, x_2, ..., x_n) = f(x_1, x_2, ..., x_n ; \theta)$, [= L(θ), in a briefer notation].

We emphasize that L is a function of θ for fixed sample values.

If $X_1, \ldots, X_n$ are discrete iid random variables with probability function p(x, θ), then, the likelihood function is given by

$$L(\theta) = P(X_1 = x_1, \ldots\ldots, X_n = x_n)$$

$$= \prod_{i=1}^{n} P(X_i = x_i)$$ **,(by multiplication rule for independent random variables)**

and in the continuous case, if the density is $f(x,\theta)$, then the likelihood function is

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

It is important to note that the likelihood function, although it depends on the observed sample values x = (x_1, . . . , x_n), is to be regarded as a function of the parameter θ.

In the discrete case, L(θ; x_1, . . . , x_n) gives the probability of observing x = (x_1, . . . , x_n), for a given θ. Thus, the likelihood function is a statistic, depending on the observed sample x = (x_1, . . . , x_n) .

Example 4-1 : Let X_1, . . . , X_n be iid N(μ, σ^2) random variables. Let x_1, . . . , x_n be the sample values . Find the likelihood function .

Solution :

The density function for the normal variable is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\ \sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right).$$

Hence, the likelihood function is

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\ \sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2} \right) = \frac{1}{(2\pi)^{\frac{n}{2}}\ \sigma^n} \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2} \right)$$

Example 4-2 :

Suppose X_1, . . . , X_n are a random sample from a geometric distribution with parameter p , $0 \le p \le 1$. Find MLE p^.

Solution :

For the geometric distribution , the pmf is given by

$f(x, p) = p(1-p)^{x-1}$, $0 \le p \le 1$, x = 1, 2, 3, ...

Hence, the likelihood function is

$$L(p) = \prod_{i=1}^{n} p(1-p)^{x_i-1} = p^n (1-p)^{\sum_{i=1}^{n} x-1} .$$

Taking the natural logarithm of L(p),

$$\ln L = n \ln p + \sum_{i=1}^{n}(x_i - 1)\ln(1-p) .$$

Taking the derivative with respect to p, we have

d ln L

$$\frac{d \ln L}{dp} = n\frac{1}{p} - \sum_{i=1}^{n}(x_i - 1)\frac{1}{1-p} .$$

Equating d ln L(p) /dp to zero, we have

$$n\frac{1}{p} - \sum_{i=1}^{n}(x_i - 1)\frac{1}{1-p} = 0$$

Solving for p,

$$n\frac{1}{p} - \sum_{i=1}^{n}(x_i - 1)\frac{1}{1-p} = 0 \Rightarrow n\frac{1}{p} = \sum_{i=1}^{n}(x_i - 1)\frac{1}{1-p} \Rightarrow \frac{1-p}{p} = \frac{\sum_{i=1}^{n} x_i - n}{n}$$

$$\Rightarrow \frac{1}{p} - 1 = \frac{\sum_{i=1}^{n} x_i}{n} - 1 \Rightarrow \frac{1}{p} = \frac{\sum_{i=1}^{n} x_i}{n} \Rightarrow p = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{X}}$$

Thus, we obtain a maximum likelihood estimator of p as

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{X}}$$

Example 4-3 :
Suppose $X_1, \ldots, X_n$ are random samples from a Poisson distribution with parameter λ . Find MLE $\hat{\lambda}$.

Solution

We have the probability mass function

$$p(x) = e^{-\lambda}\frac{\lambda^{x}}{x!} \quad , \quad x = 0, 1, 2, \ldots, \lambda > 0 .$$

Hence, the likelihood function is

$$L(\lambda) = \prod_{i=1}^{n}\left(e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}\right) = e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} .$$

Then, taking the natural logarithm, we have

$$\ln L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right).\ln\lambda - \sum_{i=1}^{n}\ln x_i!$$

and differentiating with respect to λ results in

$$\frac{d\ln L(\lambda)}{d\lambda} = -n + \left(\sum_{i=1}^{n} x_i\right)\frac{1}{\lambda}$$

and $\frac{d\ln L(\lambda)}{d\lambda} = 0$ implies $-n + \left(\sum_{i=1}^{n} x_i\right)\frac{1}{\lambda} = 0$

That is, $\lambda = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{X}$

Hence, the MLE of λ is

$$\hat{\lambda} = \overline{X}$$

Example 4-4 :
Find the MLE of a Bernoulli parameter .
Solution
We have the probability mass function of Bernoulli distribution is

$$p(x) = p^{x}(1-p)^{1-x} \quad , \quad x = 0, 1$$

Hence, the likelihood function is

$$L(p) = \prod_{i=1}^{n} \left(p^{x_i}(1-p)^{1-x_i} \right) = p^{\sum_{i=1}^{n} x_i} (1-p)^{n-\sum_{i=1}^{n} x_i} \quad .$$

Then, taking the natural logarithm, we have

$$\ln L(p) = \sum_{i=1}^{n} x_i \ln p + \left(n - \sum_{i=1}^{n} x_i \right) \ln(1-p)^{n-\sum_{i=1}^{n} x_i}$$

and differentiating with respect to p results in

$$\frac{d \ln L(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i \right)}{1-p}$$

and $\frac{d \ln L(p)}{dp} = 0$, implies $\frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i \right)}{1-p} = 0$

That is, $p = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{X}$

Hence, the MLE of p is

$$\hat{p} = \overline{X}$$

Example4-5 : Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$ random variables. Let $x_1, \ldots, x_n$ be the sample values . If both μ and σ^2 are unknown , find MLE μ and σ^2 .

Solution :

The density function for the normal variable is given by

$$f(x)=\frac{1}{\sqrt{2\pi}\ \sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Hence, the likelihood function is

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\ \sigma}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)=\frac{1}{(2\pi)^{\frac{n}{2}}\ \sigma^n}\exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right)$$

For MLE of μ

$$L(\mu,\sigma^2)=\frac{1}{(2\pi)^{\frac{n}{2}}\ \sigma^n}\exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right)$$

Then, taking the natural logarithm, we have

$$\ln L(\mu,\sigma^2)=\ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}\ \sigma^n}\right)-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

$$\frac{\partial L(\mu,\sigma^2)}{\partial\mu}=2\frac{\sum_{i=1}^{n}(x_i-\mu)}{2\sigma^2}=\frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2}$$

and $\frac{\partial L(\mu,\sigma^2)}{\partial\mu}=0$, implies

$$\frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2}=0\Rightarrow\sum_{i=1}^{n}(x_i-\mu)=0\Rightarrow\sum_{i=1}^{n}x_i=n\mu\Rightarrow\mu=\frac{\sum_{i=1}^{n}x_i}{n}=\overline{X}$$

Hence, the MLE of μ is

$$\hat{\mu}=\overline{X}$$

Now , For MLE of σ^2

$$L(\mu,\sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}}\ \sigma^n} \exp\left(-\frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{2\sigma^2} \right)$$

Then, taking the natural logarithm, we have

$$\ln L(\mu,\sigma^2) = \ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}\ \sigma^n} \right) - \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{2\sigma^2}$$

and differentiating with respect to μ results in

$$\frac{\partial L(\mu,\sigma^2)}{\partial \sigma^2} = \quad = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4}$$

and $\frac{\partial L(\mu,\sigma^2)}{\partial \sigma^2} = 0$, implies

$$-\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4} = 0 \Rightarrow n = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2} \Rightarrow \sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} = S'^2$$

Hence, the MLE of σ^2 is

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} = S'^2$$