CSL003P1M : Probability and Statistics Lecture 38 (Estimation-I (Point Estimates))

Sumit Kumar Pandey

December 08, 2021

Estimator

Statistic

A statistic is a random variable whose value is determined by the sample data.

Estimator

Any statistic used to estimate the value of an unknown parameter θ is called an estimator of θ.

Estimate

The observed value of the estimator is called the estimate.

Example-1

The usual **estimator** of the **mean (parameter)** of a normal population, based on a sample $X_1, \ldots, X_n$ from that population, is the **sample mean** $\bar{X} = \sum_i X_i/n$.

If a sample of size 3 yields the data $X_1 = 2$, $X_2 = 3$, $X_3 = 4$, then the estimate of the population mean, resulting from the estimator $\bar{X}$, is the value 3.

Estimator and Estimates

- Suppose that the random variables $X_1, \ldots, X_n$ whose joint distribution is assumed given except for an unknown parameter θ, are to be observed.
- The problem of interest is to use the observed values to estimate θ.

Example-2

For example, the X_i's might be independent, exponential random variables each having the same unknown mean θ. In this case, the joint density function of the random variables would be given by

$$
\begin{aligned}
& f(x_1, x_2, \ldots, x_n) \\
= \;& f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n) \\
= \;& \frac{1}{\theta} e^{-x_1/\theta} \frac{1}{\theta} e^{-x_2/\theta} \cdots \frac{1}{\theta} e^{-x_n/\theta}, \quad 0 < x_i < \infty, i = 1, \ldots, n \\
= \;& \frac{1}{\theta^n} \exp\left\{ -\sum_{i=1}^{n} x_i/\theta \right\}, \quad 0 < x_i < \infty, i = 1, \ldots, n
\end{aligned}
$$

and the objective would be to estimate θ from the observed data $X_1, X_2, \ldots, X_n$.

Point Estimates

- We present the maximum likelihood method for determining estimators of unknown parameters.
- The estimators so obtained are called **point estimates**, because they specify a single quantity as an estimate of θ.

Maximum Likelihood Estimator (MLE)

- A particular type of estimator, known as **maximum likelihood estimator**, is widely used in statistics.
- It is obtained by reasoning as follows:
  - let $f(x_1, \ldots, x_n|\theta)$ denote the joint probability mass function of the random variables $X_1, X_2, \ldots, X_n$ when they are discrete, and let it be their joint probability density function when they are jointly continuous random variables.
  - Because θ is assumed unknown, we also write f as a function of θ.
  - Now since $f(x_1, \ldots, x_n|\theta)$ represents the likelihood that the values $x_1, x_2, \ldots, x_n$ will be observed when θ is the true value of the parameter, it would seem that a reasonable estimate of θ would be that value yielding the largest likelihood of the observed values.

Maximum Likelihood Estimate

The maximum likelihood estimate $\hat{\theta}$ is defined to be that value of θ maximizing $f(x_1, \ldots, x_n|\theta)$ where $x_1, \ldots, x_n$ are the observed values. The function $f(x_1, \ldots, x_n|\theta)$ is often referred to as the likelihood function of θ.

In determining the maximizing value of θ, it is often useful to use the fact that $f(x_1, \ldots, x_n|\theta)$ and $\log[f(x_1, \ldots, x_n|\theta)]$ have their maximum at the same value of θ. Hence we may also obtain $\hat{\theta}$ by maximizing $\log[f(x_1, \ldots, x_n|\theta)]$.

Suppose that n independent trials, each of which is a success with probability p, are performed. What is the maximum likelihood estimator of p?

Solution: The data consist of the values of $X_1, \ldots, X_n$ where

$$X_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{otherwise} \end{cases}$$

Now,

$$P\{X_i = 1\} = p = 1 - P\{X_i = 0\}$$

which can be succinctly expressed as

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

MLE of a Bernoulli Parameter

Hence, by the assumed independence of trials, the likelihood (that is, the joint probability mass function) of the data is given by

$$
\begin{aligned}
f(x_1,\ldots,x_n|p) &= P\{X_1 = x_1,\ldots,X_n = x_n|p\} \\
&= p^{x_1}(1-p)^{1-x_1}\cdots p^{x_n}(1-p)^{1-x_n} \\
&= p^{\sum_1^n x_i}(1-p)^{n-\sum_1^n x_i}, \quad x_i = 0,1, \quad i = 1,\ldots,n
\end{aligned}
$$

To determine the value of p that maximizes the likelihood, first take logs to obtain

$$\log f(x_1,\ldots,x_n|p) = \left(\sum_{i=1}^n x_i\right)\log p + \left(n - \sum_{i=1}^n x_i\right)\log(1-p)$$

differentiation yields

$$\frac{d}{dp}\log f(x_1,\ldots,x_n|p) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p}$$

Upon equating to zero, we obtain the maximum likelihood estimate $\hat{p}$ statisfies

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

or,

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

MLE of a Poisson Parameter

Suppose $X_1, \ldots, X_n$ are independent Poisson random variables each having mean λ. Determine the maximum likelihood estimator of λ.

Solution: The likelihood function is given by

$$\begin{aligned} f(x_1, \ldots, x_n|\lambda) &= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdots \frac{e^{-\lambda}\lambda^{x_n}}{x_n!} \\ &= \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{x_1! \ldots x_n!} \end{aligned}$$

Thus,

$$\log f(x_1, \ldots, x_n|\lambda) = -n\lambda + \sum_{i=1}^{n} x_i \log \lambda - \log c$$

where $c = \prod_{i=1}^{n} x_i!$ does not depend on λ.

Differentiation yields

$$\frac{d}{d\lambda} \log f(x_1, \ldots, x_n|\lambda) = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}$$

Upon equating to zero, we obtain the maximum likelihood estimate $\hat{\lambda}$ statisfies

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Suppose $X_1, \ldots, X_n$ are independent, normal random variables each with unknown mean μ and unknown standard deviation σ. Determine the maximum likelihood estimator of μ and σ.

Solution:

$$\begin{aligned} f(x_1, \ldots, x_n | \mu, \sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(x_i - \mu)^2}{2\sigma^2}\right] \\ &= \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sigma^n} \exp\left[\frac{-\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2}\right] \end{aligned}$$

MLE in a Normal Population

Thus,

$$\log f(x_1,\ldots,x_n|\mu,\sigma) = \frac{-n}{2}\log(2\pi) - n\log\sigma - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

Differentiation yields

$$\frac{\partial}{\partial\mu}\log f(x_1,\ldots,x_n|\mu,\sigma) = \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2}$$

$$\frac{\partial}{\partial\sigma}\log f(x_1,\ldots,x_n|\mu,\sigma) = \frac{-n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3}$$

MLE in a Normal Population

Equating to zero, we get

$$\hat{\mu} = \sum_{i=1}^{n} x_i / n$$

and

$$\hat{\sigma} = \left[\frac{\sum_{i=1}^{n} (x_i - \hat{\mu})^2}{n} \right]^{1/2}$$

Hence, the maximum likelihood estimate of μ and σ are given, respectively, by

$$\bar{X} \text{ and } \left[\sum_{i=1}^{n}(X_i - \bar{X})^2/n\right]^{1/2}$$

It should be noted that the maximum likelihood estimator of the standard deviation σ differs from the sample deviation

$$S = \left[\sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)\right]^{1/2}$$

Estimating the Mean of a Uniform Distribution

Suppose $X_1, \ldots, X_n$ constitute a sample from a uniform distribution on $(0, \theta)$, where θ is unknown. Determine the maximum likelihood estimator of θ.

Solution:

$$f(x_1, x_2, \ldots, x_n|\theta) = \begin{cases} \dfrac{1}{\theta^n} & 0 < x_i < \theta, \quad i = 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

The density is maximized by choosing θ as small as possible. Since θ must be at least as large as all of the observed values x_i, it follows that the smallest possible choice of θ is equal to $\max(x_1, x_2, \ldots, x_n)$. Hence the maximum likelihood estimator of θ is

$$\hat{\theta} = \max(X_1, X_2, \ldots, X_n)$$

So, the maximum likelihood estimator of $\theta/2$, the mean of the distribution, is $\max(X_1, X_2, \ldots, X_n)/2$.

Some Methods of Evaluating Estimators

1. Mean Squared Error (MSE)
2. Loss Function Optimality

Note: Actually, mean squared error method is a special case of loss function optimality method.

Mean Squared Error

Mean Squared Error

The mean squared error (MSE) of an estimator W of a parameter θ is the function $E_\theta[(W-\theta)^2]$.

$$E_\theta[(W-\theta)^2] = Var_\theta(W) + (E_\theta[W]-\theta)^2 = Var_\theta(W) + (\text{Bias}_\theta(W))^2$$

where $\text{Bias}_\theta(W)$ is defined as follows:

Bias of a point estimate

The bias of a point estimate W of a parameter θ is the difference between the expected value of W and θ; that is,
$\text{Bias}_\theta(W) = E_\theta[W] - \theta$

Unbiased Estimator

Unbiased Estimator

An estimator whose bias is identically (in θ) equal to 0 is called unbiased and satisfies $E_\theta[W] = \theta$ for all θ.

For an unbiased estimator, we have

$$E_\theta[(W - \theta)^2] = Var_\theta(W)$$

and so, if an estimator is unbiased, its MSE is equal to its variance.

Example-3

Let $X_1, \ldots, X_n$ be independent and identically distributed (iid) normal random variables with parameters μ and σ^2. The statistics $\bar{X}$ and S^2 are both unbiased estimators since

$$E[\bar{X}] = \mu, \quad E[S^2] = \sigma^2, \quad \text{for all } \mu \text{ and } \sigma^2$$

(This is true without the normality assumption). The MSEs of these estimators are given by

$$E[(\bar{X} - \mu)^2] = Var(\bar{X}) = \frac{\sigma^2}{n},$$
$$E[(S^2 - \sigma^2)^2] = Var(S^2) = \frac{2\sigma^4}{n-1}. \text{ (Why?)}$$

The MSE of $\bar{X}$ remains σ^2/n even if the normality assumption is dropped. However, the above expression for S^2 does not remain the same if the normality assumption is relaxed.

Example-4

An alternative estimator for σ^2 is the maximum likelihood estimator $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n}S^2$. It is straightforward to calculate

$$E[\hat{\sigma}^2] = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}\sigma^2,$$

so $\hat{\sigma}^2$ is a biased estimator of σ^2.

Thank You