Chapter 2
POINT ESTIMATION

2.1 Introduction

Statistical inference is the process by which information from sample data is used to draw conclusions about the population from which the sample was selected. The techniques of statistical inference can be divided into two major areas:

parameter estimation and ***hypothesis testing***. There are two types of parameter estimation, namely, point estimation and interval estimation. This chapter treats point estimation.

Consider a random sample of size n from a population with pdf (or pmf) $f(x\,;\,\theta)$. The term random sample may refer either to the set of i.i.d. (**i**ndependent **i**dentically **d**istributed) random variables, $X_1, X_2, \ldots, X_n$, or to the observed data $x_1, x_2, \ldots, x_n$.

Definition 2.1: Statistic

A function of the random sample, T= $\mathbf{t}(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$, that does not depend on any unknown parameters is called a **statistic.**

A statistic is also a random variable, the distribution of which depends on the distribution of a random sample and on the form of the function $\mathbf{t}(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$. The distribution of a statistic is referred to as a ***sampling distribution***.

Example 2.1

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from a normal distribution, $\mathbf{N(\mu,\sigma^2)}$, then the sample mean $\mathbf{\overline{X} = \frac{1}{n}\sum X_i}$ is an example of a statistic. The sample variance $\mathbf{S^2 = \frac{1}{n-1}\sum (X_i - \overline{X})^2}$ provides another example of a statistic.

A point estimate of a population parameter is a single numerical value of a statistic that corresponds to that parameter. That is, the point estimate is a unique selection for the value of an unknown parameter. More precisely, if **X** is a r.v. with probability distribution $\mathbf{f(x\,;\,\theta)}$, characterized by the unknown parameter $\boldsymbol{\theta}$, and if $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ is a random sample of size n from X, then the statistic $\hat{\theta} = \mathbf{t}(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ corresponding to $\boldsymbol{\theta}$ is called the estimator of θ.

Definition 2.2

A **statistic** T= $\mathbf{t}(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$, that is used to estimate the unknown parameter $\boldsymbol{\theta}$ is called an ***estimator*** of $\boldsymbol{\theta}$, and an observed value of the statistic, ***t***= $\mathbf{t}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ is called an ***estimate*** of $\boldsymbol{\theta}$.

Estimation problems occur frequently in real life. We often need to estimate:

- the mean $\boldsymbol{\mu}$ of a single population
- the variance $\boldsymbol{\sigma}^2$ (or standard deviation $\boldsymbol{\sigma}$) of a single population
- the proportion **p** of items in a population that belong to a class of interest
- the difference in means of two populations, $\boldsymbol{\mu_1 - \mu_2}$
- the difference in two population proportions, $\mathbf{p_1 - p_2}$

Reasonable point estimates of these parameters are as follows:

- ♦ for μ, the estimate is $\hat{\mu} = \overline{\mathbf{X}}$, the sample mean
- ♦ for σ^2, the estimate is $\hat{\sigma}^2 = \mathbf{S}^2$, the sample variance
- ♦ for p, the estimate is $\hat{\mathbf{p}} = \mathbf{X}/\mathbf{n}$, the sample proportion, where X is the number of items in a random sample of size n that belong to the class of interest
- ♦ for $\boldsymbol{\mu_1}$- $\boldsymbol{\mu_2}$, the estimate is $\hat{\mu}_1 - \hat{\mu}_2 = \overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2$ the difference between the sample means of two independent random samples
- ♦ for $\mathbf{p_1}$ - $\mathbf{p_2}$, the estimate is $\hat{\mathbf{p}}_1 - \hat{\mathbf{p}}_2$ the difference between two sample proportions computed from two independent random samples

2.2 Some Methods of Estimation

In some cases reasonable point estimators can be found on the basis of intuition, but various general methods have been developed for deriving estimators.

(I) The Method of Moments

Suppose that **X** is a r.v. with p.d.f. (or p.m.f.) $\mathbf{f}(\mathbf{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_r)$ characterized by r unknown parameters. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample of size n from X, and define the first r sample moments about the origin as

$$\mathbf{m}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \quad , \qquad \mathbf{k = 1,2, \ldots, r} \tag{2.1}$$

The population moments $\boldsymbol{\mu_k'}$ will, in general, be functions of the r unknown parameters $\{\boldsymbol{\theta}'\text{s}\}$. Equating sample moments and population moments will yield r simultaneous equations in r unknowns {the θ_i's}; that is.

$$\mu_k = m_k \quad , \quad k = 1, 2, \ldots, r \tag{2.2}$$

The solution to equation (2.2), denoted by $\hat{\theta}_1 , \hat{\theta}_2 , \cdots , \hat{\theta}_r$ yields the moment estimators of $\theta_1, \theta_2, ..., \theta_r$.

Example 2.2

Let **X** be uniformly distributed on the interval (α, 1). Given a random sample of size n, use the method of moments to obtain a formula for estimating the parameter α.

Solution

To find an estimator of α by the method of moments, we note that the first population moment about zero is

$$\mu_1 = E(X) = \int_{-\infty}^{\infty} x \, f(x;\alpha) \, dx = \int_{\alpha}^{1} x \, . \, \frac{1}{1-\alpha} \, dx == \frac{1}{1-\alpha}\left[\frac{x^2}{2}\right]_{\alpha}^{1} = \frac{1+\alpha}{2}$$

The first sample moment is just

$$m_1 = \overline{X}$$

Therefore,

$$m_1 = \mu_1 \;\Rightarrow\; \overline{X} = \frac{1+\hat{\alpha}}{2} \;\Rightarrow\; \hat{\alpha} = 2\overline{X} - 1$$

Example 2.3

Given a random sample of size n from a Poisson population, use the method of moments to obtain a formula for estimating the parameter λ.

Solution

The p.m.f. of the Poisson distribution with parameter λ is given by

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

The first population moment about zero is

$$\mu_1 = E(x) = \lambda$$

The first sample moment is

$$m_1 = \overline{X}$$

From (2.2), we obtain

$$m_1 = \mu_1 \;\Rightarrow\; \overline{X} = \hat{\lambda}$$

which has the solution

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

Example 2.4

Given a random sample of size n from a **N(μ, σ²)** population, use the method of moments to obtain formulas for estimating the parameters **μ** and **σ²**.

Solution

The normal p.d.f. with parameters **μ** and **σ²** is given by

$$\mathbf{f(x;\mu,\sigma^2)} = \frac{1}{\sqrt{2\pi}\,\sigma}\,\mathbf{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , \ -\infty < x < \infty$$

The first two population moments about zero are

$$\mu_1 = E(x) = \mu \ , \qquad \mu_2 = E(x^2) = \sigma^2 + \mu^2$$

The first two sample moments are

$$m_1 = \overline{X} \qquad \text{and} \qquad m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

From (2.2), we obtain

$$m_1 = \mu_1 \ \Rightarrow \overline{X} = \hat{\mu}$$

$$m_2 = \mu_2 \ \Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \hat{\sigma}^2 + \hat{\mu}^2$$

which have the solution

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

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum (X_i - \overline{X})^2$$

(II) The Method of Maximum Likelihood

One of the best methods of obtaining a point estimator is the method of maximum likelihood. Suppose that X is a r.v. with p.d.f. (or p.m.f.) f(x ; θ), where θ is a single unknown parameter. Let $x_1, x_2, \dots, x_n$ be a random sample of size n. Then the likelihood function represents the joint pdf (or pmf) of the random sample, i.e.

$$L(\theta) = f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^{n} f(x_i;\theta) = f(x_1;\theta)\ \dots\ f(x_n;\theta) \qquad \textbf{(2.3)}$$

Note that the likelihood function is now a function of the unknown parameter **θ** only. The maximum likelihood estimator (MLE) of **θ** is the value of **θ** that maximizes the likelihood function **L(θ)**. Essentially, the maximum likelihood estimator $\hat{\theta}$ is the value of $\boldsymbol{\theta}$ that maximizes the probability of occurrence of the sample results i.e.

$$\mathbf{f(x_1, x_2, \ldots, x_n; \hat{\theta}) = \max_{\theta} f(x_1, x_2, \ldots, x_n; \theta)}$$

If $\mathbf{L(\theta)}$ is differentiable, then the MLE will be a solution of the equation (ML equation)

$$\mathbf{\frac{d}{d\theta} L(\theta)\Big|_{\theta=\hat{\theta}} = 0} \qquad \textbf{(2.4)}$$

If one or more solutions of (2.4) exist, it should be verified which ones, if any, maximize $\mathbf{L(\theta)}$. Note also that any value of $\boldsymbol{\theta}$ that maximizes $\mathbf{L(\theta)}$ will also maximizes the log-likelihood, $\mathbf{\ln L(\theta)}$, so for computational convenience the alternate form of (2.4),

$$\mathbf{\frac{d}{d\theta} \ln\{L(\theta)\}\Big|_{\theta=\hat{\theta}} = 0} \qquad \textbf{(2.5)}$$

will often be used

Example 2.5

If $x_1, x_2, \ldots, x_n$ are the values of a random sample from a Bernoulli population. Find the MLE of its parameter θ.

Solution

The p.m.f. of the Bernoulli distribution is (Binomial with n = 1)

$$\mathbf{f(x;\theta) = \theta^x (1-\theta)^{1-x} \qquad for \qquad x = 0, 1}$$

The likelihood function is

$$\mathbf{L(\theta) = \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \left\{\theta^{x_i} (1-\theta)^{1-x_i}\right\} = \prod_{i=1}^{n} \theta^{x_i} \prod_{i=1}^{n} (1-\theta)^{1-x_i}}$$

$$\mathbf{= \theta^{\Sigma x_i} \; (1-\theta)^{n-\Sigma x_i}}$$

and the log-likelihood function is

$$\mathbf{L^*(\theta) = \ln\{L(\theta)\} = \sum x_i \ln(\theta) + (n - \Sigma x_i) \ln(1-\theta)}$$

The ML equation is

$$\mathbf{\frac{dL^*(\theta)}{d\theta} = 0 \;\Rightarrow\; \frac{\Sigma x_i}{\hat{\theta}} - \frac{n - \Sigma x_i}{1-\hat{\theta}} = 0}$$

which has the solution $\mathbf{\hat{\theta} = \overline{X}}$.

Example 2.6

Let $X_1, X_2, \ldots, X_n$ be a random sample from an exponential population, find the MLE for the parameter θ.

Solution

The pdf of the exponential distribution with parameter θ is

$$\mathbf{f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}} \quad \theta > 0 \ , \ \mathbf{x} > 0$$

The likelihood function is

$$\mathbf{L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \left\{ \frac{1}{\theta} e^{-x_i/\theta} \right\} = \frac{1}{\theta^n} e^{-\Sigma x_i/\theta}}$$

and the log-likelihood function is

$$\mathbf{L^*(\theta) = \ln\big(L(\theta)\big) = -n \ln(\theta) - \frac{\Sigma x_i}{\theta}}$$

The ML equation is

$$\mathbf{\frac{dL^*(\theta)}{d\theta}} = 0 \ \Rightarrow \ -\frac{\mathbf{n}}{\hat{\theta}} + \frac{\mathbf{\Sigma x_i}}{\hat{\theta}^2} = 0$$

which has the solution

$$\hat{\theta} = \overline{\mathbf{X}}$$

Generalization

The method of maximum likelihood can be used in situations where there are several unknown parameters, say $\boldsymbol{\theta_1, \theta_2, \ldots, \theta_k}$, to estimate. In such cases, the likelihood function is a function of the k unknown parameters $\boldsymbol{\theta_1, \theta_2, \ldots, \theta_k}$ and the MLE's $\{\hat{\theta}_\mathbf{i}\}$ would be found by equating the k first partial derivatives of the likelihood function or its logarithm, i.e. the MLE's $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_\mathbf{k}$ of the parameters $\boldsymbol{\theta_1, \theta_2, \ldots, \theta_k}$ are the solutions of the equations

$$\frac{\partial}{\partial \boldsymbol{\theta}_1} \mathbf{L}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_\mathbf{k}) = 0$$

$$\frac{\partial}{\partial \boldsymbol{\theta}_2} \mathbf{L}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_\mathbf{k}) = 0$$

$$\vdots$$

$$\frac{\partial}{\partial \boldsymbol{\theta}_\mathbf{k}} \mathbf{L}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_\mathbf{k}) = 0$$

In this case it may also be easier to work with the logarithm of the likelihood $\mathbf{L^*}$.

Example 2.7

If $X_1, X_2, \ldots, X_n$ constitute a random sample from a normal population with the mean μ and the variance σ^2, find the joint maximum likelihood estimates of these two parameters.

Solution

The pdf of the normal distribution with parameters μ & σ^2 is

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Hence, the likelihood function is

$$L(\mu,\sigma^2) = \prod_{i=1}^{n} f(x_i;\mu,\sigma^2) = \left(\frac{1}{2\pi}\right)^{n/2}\left(\frac{1}{\sigma^2}\right)^{n/2} e^{-\frac{\Sigma(x_i-\mu)^2}{2\sigma^2}}$$

and the log-likelihood function is

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

The ML equations are

$$\frac{\partial L^*(\mu,\sigma^2)}{\partial \mu} = 0 \;\Rightarrow\; 2\frac{\Sigma(x_i-\hat{\mu})}{2\hat{\sigma}^2} = 0 \;\Rightarrow\; \hat{\mu} = \overline{X}$$

$$\frac{\partial L^*(\mu,\sigma^2)}{\partial \sigma^2} = 0 \;\Rightarrow\; -\frac{n}{2}\frac{1}{\hat{\sigma}^2} + \frac{\Sigma(x_i-\hat{\mu})^2}{2\hat{\sigma}^4} = 0 \;\Rightarrow\; \hat{\sigma}^2 = \frac{1}{n}\Sigma(x_i-\hat{\mu})^2$$

which have the solution

$$\hat{\mu} = \overline{X} \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{1}{n}\Sigma(x_i-\bar{x})^2 = S_n^2$$

It should be observed that we did not show that $\hat{\sigma}$ is a maximum likelihood estimate of σ, only that $\hat{\sigma}^2$ is a MLE of σ^2. However, it can be shown that the maximum likelihood estimators have the invariance property that is if $\hat{\theta}$ is the MLE of θ and a function given by $g(\theta)$ is continuous, then $g(\hat{\theta})$ is also the MLE of $g(\theta)$.

Invariance Property of MLE'S

Let $\hat{\theta} = t(X_1, X_2, \ldots, X_n)$ be the MLE of θ, where θ is assumed one-dimensional, and $\tau(\theta)$ is a single function with a single-valued inverse (one-to-one), then the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

2.3 Properties of Estimators

There may be several different potential point estimators for a parameter. For example, if we wish to estimate the mean of a random variable, we might consider the sample mean, the sample median, or perhaps the average of the smallest and largest observations in the sample as point estimators. In order to decide which point estimator of a particular parameter is the best one to use, we need to examine their statistical properties and develop some criteria for comparing estimators. There are several properties of estimators that would appear to be desirable, such as unbiasdness, minimum variance and sufficiency.

A desirable property of an estimator is that should be "close" in some sense to the true value of the unknown parameter. A useful measure of goodness or closeness of an estimator $\hat{\theta} = t(X_1, X_2, \ldots, X_n)$ of θ is what is called the ***mean-squared error*** of the estimator.

Definition 2.3 Mean-Squared Error

If $\hat{\theta} = \mathbf{t}(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ is an estimator of θ, then the mean squared error (MSE) of $\hat{\theta}$ is given by

$$\mathbf{MSE}(\hat{\theta}) = \mathbf{E}(\hat{\theta} - \theta)^2 \qquad \mathbf{(2.6)}$$

Thus, if $\hat{\theta}_1$ and $\hat{\theta}_2$ are two different estimators for the same parameter θ, and let $\mathbf{MSE}(\hat{\theta}_1)$ and $\mathbf{MSE}(\hat{\theta}_2)$ be the mean square errors of $\hat{\theta}_1$ and $\hat{\theta}_2$, then we prefer the estimator with the smaller **MSE**, i.e we say that $\hat{\theta}_1$ is better than $\hat{\theta}_2$ if

$$\mathbf{MSE}(\hat{\theta}_1) < \mathbf{MSE}(\hat{\theta}_2)$$

a- Unbiased Estimators

An estimator $\hat{\theta}$ is said to be an unbiased estimator of the parameter θ if

$$\mathbf{E}(\hat{\theta}) = \theta \qquad \mathbf{(2.7)}$$

That is $\hat{\theta}$ is an unbiased estimator of θ if "on the average" its values are equal to θ. Note that this is equivalent to requiring that the mean of the sampling distribution of $\hat{\theta}$ is equal to θ.

The quantity

$$\mathbf{b}(\hat{\theta}) = \mathbf{E}(\hat{\theta}) - \theta$$

is called the ***bias*** of the estimator $\hat{\theta}$.

Theorem 2.3

The mean squared error can be written as follows,

$$\text{MSE}(\hat{\theta}) = \text{var}(\hat{\theta}) + (\text{bias})^2 \tag{2.8}$$

Proof

$$\text{MSE}(\hat{\theta}) = E\left[\left\{\hat{\theta} - E\left(\hat{\theta}\right)\right\} - \left\{\theta - E\left(\hat{\theta}\right)\right\}\right]^2$$

$$= E\left[\hat{\theta} - E\left(\hat{\theta}\right)\right]^2 + \left[\theta - E\left(\hat{\theta}\right)\right]^2 - 2E\left\lfloor\left\{\hat{\theta} - E\left(\hat{\theta}\right)\right\}\left\{\theta - E\left(\hat{\theta}\right)\right\}\right\rfloor$$

$$= E\left[\hat{\theta} - E\left(\hat{\theta}\right)\right]^2 + \left[\theta - E\left(\hat{\theta}\right)\right]^2$$

$$= \text{var}(\hat{\theta}) + [b(\hat{\theta})]^2$$

Since the third term is

$$2\left\{\theta - E\left(\hat{\theta}\right)\right\} E\left\lfloor\hat{\theta} - E\left(\hat{\theta}\right)\right\rfloor = 0$$

That is, the mean square error of $\hat{\theta}$ is equal to the variance of the estimator plus the squared bias. If $\hat{\theta}$ is an unbiased estimator of θ, then the mean square error of $\hat{\theta}$ is equal to the variance of $\hat{\theta}$.

Example 2.8

If X has the binomial distribution with the parameters n and p, show that the sample proportion, $\hat{p} = \dfrac{X}{n}$ is an unbiased estimator of p.

Solution

Since $E(X) = np$, then $E(\hat{p}) = E\left(\dfrac{X}{n}\right) = \dfrac{np}{n} = p$

Example 2.9

Suppose that X is a r.v. with mean μ and variance σ^2. Let $X_1, X_2, \ldots, X_n$ be a random sample from X. Show that the sample mean $\overline{X}$ and the sample variance S^2 are unbiased estimators of μ and σ^2 respectively.

Solution

Since

$$E(\overline{X}) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}n\mu = \mu$$

Therefore $\overline{X}$ is an unbiased estimator for μ.

Now, $\mathbf{E(S^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]}$

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

$$= \frac{1}{n-1}E\left[\sum_{i=1}^{n}\left(X_i - \mu\right)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu)\sum_{i=1}^{n}\left(X_i - \mu\right)\right]$$

Since

$$\sum_{i=1}^{n}\left(X_i - \mu\right) = \sum_{i=1}^{n}X_i - n\mu = n(\bar{X} - \mu)$$

then,

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

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}E\left(X_i - \mu\right)^2 - nE(\bar{X} - \mu)^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}\mathrm{var}\left(X_i\right) - n\,\mathrm{var}(\bar{X})\right]$$

$$= \frac{1}{n-1}\left[n\sigma^2 - n\frac{\sigma^2}{n}\right]$$

$$= \frac{1}{n-1}\sigma^2(n-1)$$

$$= \sigma^2$$

Note:

- **Since** $X_1, X_2, \dots, X_n$ are independent (random sample), then

$$\mathrm{var}(\bar{X}) = \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

- The MLE of σ^2, namely,

$$S_n^2 = \frac{1}{n}\sum(X_i - \bar{X})^2$$

is not unbiased estimator for σ^2, since

$$E(S_n^2) = \frac{n-1}{n}E(S_n^2) = (1 - \tfrac{1}{n})\sigma^2$$

However, we note that

$$E(S_n^2) = (1-\tfrac{1}{n})\sigma^2 \rightarrow \sigma^2 \quad \text{as} \quad n \rightarrow \infty$$

therefore, S_n^2 is said to be ***asymptotically unbiased*** estimator for σ^2.

b- Minimum variance unbiased Estimator (MVUE)

We note that if the estimator $\hat{\theta}$ is unbiased for θ, the mean square error reduces to the variance of the estimator $\hat{\theta}$. Within the class of unbiased estimators, we would like to find the estimator that has the smallest variance. Such an estimator is called a minimum variance unbiased estimator.

Definition 2.4

Let $X_1, X_2, \ldots X_n$ be a random sample of size n from $f(x;\theta)$. An estimator $\hat{\theta}$ of θ is called a ***minimum variance unbiased estimator*** **(MVUE)** of θ if

1- $\hat{\theta}$ is unbiased for θ, and

2- for any other unbiased estimator $\tilde{\theta}$ of θ, $\mathbf{Var}(\tilde{\theta}) \geqslant \mathbf{Var}(\hat{\theta})$ for all possible values of θ.

In some cases, lower bounds can be derived for the variance of unbiased estimators. If an unbiased estimator can be found that attains such a lower bound, then it follows that the estimator is a MVUE.

C- Efficiency

The mean square error is an important criterion for comparing two estimators. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two estimators of the parameter θ, and let $MSE(\hat{\theta}_1)$ and $MSE(\hat{\theta}_2)$ be the mean square errors of $\hat{\theta}_1$ and $\hat{\theta}_2$. Then, the relative efficiency of $\hat{\theta}_2$ to $\hat{\theta}_1$ is defined as.

$$\mathbf{eff}\ (\hat{\theta}_2/\hat{\theta}_1) = \frac{\mathbf{MSE}\ (\hat{\theta}_1)}{\mathbf{MSE}\ (\hat{\theta}_2)} \tag{2.9}$$

If this relative efficiency is less than one, we would conclude that $\hat{\theta}_1$ is a more efficient estimator of θ than $\hat{\theta}_2$, in the sense that it has smaller mean square error.

For example, suppose that we wish to estimate the mean μ of a population. We have a random sample of n observations $X_1, X_2, \ldots, X_n$ and we wish to compare two possible. Estimators for μ: the sample mean $\overline{X}$ and a single observation from the sample, say X_1. Not that both $\overline{X}$ and X_1 are unbiased estimators of μ; consequently, the mean square error of both estimators is simply the variance. For the sample mean, we have

$$\mathbf{MSE}\,(\overline{\mathbf{X}}) = \mathbf{var}\ (\overline{\mathbf{X}}) = \frac{\boldsymbol{\sigma}^2}{\mathbf{n}}$$

where σ^2 is the population variance; for an individual observation, we have

$$\mathbf{MSE\ (X_1) = var\ (X_1) = \boldsymbol{\sigma}^2}$$

Therefore, the relative efficiency of X_1 to $\overline{X}$ is

$$\mathbf{eff}\ \ (\mathbf{X}_1 / \overline{\mathbf{X}}) = \frac{\mathbf{MSE}\ \overline{\mathbf{X}}}{\mathbf{MSE}\,(\mathbf{X}_1)} = \frac{1}{\mathbf{n}}$$

Since $(1 / n) < 1$ for sample sizes $n \geq 2$ we would conclude that the sample mean $\overline{X}$ is more efficient estimator of μ than as single observation X_1.

Note: The MVUE is sometimes called the ***most efficient estimator***.

D- Consistency

The estimator $\hat{\theta}$ is called a consistent estimator of the parameter θ if and only if

$$\mathbf{MSE}(\hat{\theta}) = \mathbf{E}(\hat{\theta} - \theta)^2 \to \mathbf{0} \ \text{ as } \mathbf{n} \to \infty$$

That is $\hat{\theta}$ is unbiased (or $\mathbf{E}[\hat{\theta}] \to \theta$ as $\mathbf{n} \to \infty$) and var($\hat{\theta}$)$\to\theta$ as $\mathbf{n} \to \infty$.

Note that consistency is an asymptotic property, that is, a limiting property of an estimator. Informally, consistency means that when n is sufficiently large, we can be practically certain that the error made with a consistent estimator well be as small as we can.

Based on the previous examples, since $\hat{\mathbf{p}} = \frac{\mathbf{X}}{\mathbf{n}}$ is unbiased estimator for p and

$$\mathbf{var}(\hat{\mathbf{p}}) = \mathbf{var}\left(\frac{\mathbf{X}}{\mathbf{n}}\right) = \frac{\mathbf{np(1-p)}}{\mathbf{n}^2} = \frac{\mathbf{p(1-p)}}{\mathbf{n}} \to \mathbf{0} \ \text{ as } \mathbf{n} \to \infty$$

Then $\hat{\mathbf{p}} = \frac{\mathbf{X}}{\mathbf{n}}$ is consistent estimator for p.

Similarly $\overline{X}$ is a consistent estimators of μ, since $\overline{X}$ is unbiased and

$$\mathbf{var}(\overline{\mathbf{X}}) = \frac{\boldsymbol{\sigma}^2}{\mathbf{n}} \to \mathbf{0} \ \text{ as } \mathbf{n} \to \infty$$

<+><+><+><+><+><+><+><+><+>

EXERCISES

[1] Find method of moments estimators (MME's) of θ based on a random sample $X_1, \ldots, X_n$ from each of the following pdf's:

a- $f(x;\theta) = \theta x^{\theta-1}$; $0 < x < 1$, zero otherwise; $\theta > 0$.

b- $f(x;\theta) = (\theta + 1)x^{-\theta-2}$; $1 < x$, zero otherwise; $\theta > 0$.

c- $f(x;\theta) = \theta^2 x e^{-\theta x}$; $0 < x$, zero otherwise; $\theta > 0$.

[2] Find maximum likelihood estimators (MLE's) for θ based on a random sample of size n for each of the pdf's in problem [1].

[3] Find the MLE for θ based on a random sample of size n from a distribution with pdf

$$f(x;\theta) = \begin{cases} 2\theta^2 x^{-3} & x \geq \theta \\ 0 & x < \theta \;;\; \theta > 0 \end{cases}$$

[4] Let $X_1, X_2, \ldots, X_n$ be random sample from a geometric distribution

$$f(x;\theta) = \theta(1-\theta)^{x-1} \qquad \text{for} \quad x = 1, 2, 3, \ldots.$$

Find a formula for estimating θ by using,

a- the method of moments, **b-** the method of maximum likelihood.

[5] Let $X_1, X_2, \ldots, X_n$ be a random sample from a geometric distribution, X ~ GEO(p). Find the MLE's of the following quantities:

a- $E(X) = 1/p$. **b-** $Var(X) = (1-p)/p^2$.

c- $P[X > k] = (1-p)^k$ for arbitrary $k = 1, 2, \ldots$

(Hint: Use the invariance property of MLE's).

[6] If $X_1, X_2, \ldots, X_n$ constitute a random sample from a population given by the p.d.f.

$$f(x;\theta) = \begin{cases} \frac{1}{\theta^2} x e^{-x/\theta} & x > 0 \;,\; \theta > 0 \\ 0 & \text{otherwise} \end{cases}$$

a- Find the maximum likelihood estimator $\hat{\theta}$ for the parameter θ.

b- Show that the method of moments give the same estimator $\hat{\theta}$ for θ.

c- Prove that $\hat{\theta}$ is unbiased and consistent estimator for θ.

(**Hint:** $\int_0^\infty x^m e^{-x/\theta} d\theta = m!\,\theta^{m+1}$ for any +ve integer m)

[7] If $X_1, X_2, \ldots, X_n$ are a random sample from a population given by

$$f(x;\beta)=\begin{cases}\dfrac{1}{2\beta^3}\,x^2\,e^{-x/\beta} & , \quad x>0\\ 0 & , \quad o.w.\end{cases}$$

a- Find the maximum likelihood estimator $\hat{\beta}$ for the parameter β.

b- Show that the method of moments gives the same estimator $\hat{\beta}$ for β.

c- Prove that $\hat{\beta}$ is the minimum variance unbiased estimator for β.

[8] If $X_1, X_2, \dots, X_n$ are a random sample from the Poisson distribution

$$f(x;\theta)=\frac{e^{-\theta}\,\theta^x}{x!} \quad , \quad x=0,1,2,\dots.$$

a- Find the maximum likelihood estimator for the parameter θ.

b- Prove that $\hat{\theta}$ is the minimum variance unbiased estimator for θ.

[9] Let $X_1, \dots, X_n$ be a random sample from EXP(θ), and define $\hat{\theta}_1=\overline{X}$ and $\hat{\theta}_2=n\overline{X}/(n+1)$.

(a) Find the variances of $\hat{\theta}_1$ and $\hat{\theta}_2$.

(b) Find the MSE's of $\hat{\theta}_1$ and $\hat{\theta}_2$.

(c) Compare the variances of $\hat{\theta}_1$ an $\hat{\theta}_2$ for n = 2.

(d) Compare the MSE's of $\hat{\theta}_1$ and $\hat{\theta}_2$ for n = 2.

[10] Let X_1, X_2 and X_3 be a random sample from a population having mean μ and variance σ^2. Consider the following estimators:

$$\hat{\mu}_1=\frac{2X_1+X_2-X_3}{2} \quad \& \quad \hat{\mu}_2=\frac{3X_1+2X_2-X_3}{4}$$

compare these two estimators. Which do you prefer? Why?

[11] Suppose that $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimators of the parameter θ. We know that $E(\hat{\theta}_1)=\theta$, $var(\hat{\theta}_1)=10$ and $E(\hat{\theta}_2)=\theta/2$, $var(\hat{\theta}_2)=4$ which estimator is "best" ? In what sense it is best?

[12] Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two estimators of θ. The estimator $\hat{\theta}_2$ is said to be more efficient than $\hat{\theta}_1$ if

a. $var(\hat{\theta}_1)>var(\hat{\theta}_2)$

b. $MSE(\hat{\theta}_1)>MSE(\hat{\theta}_2)$

c. $E(\hat{\theta}_1)>E(\hat{\theta}_2)$

d. None of the above

[13] Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of θ. The estimator $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ if

a. $E(\hat{\theta}_1^2) > E(\hat{\theta}_2^2)$ **b.** $E(\hat{\theta}_1^2) < E(\hat{\theta}_2^2)$ **c.** $E(\hat{\theta}_1) > E(\hat{\theta}_2)$ **d.** $E(\hat{\theta}_1) < E(\hat{\theta}_2)$

[14] Suppose that $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$ are three estimators of the parameter θ. We know that $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta, E(\hat{\theta}_3) \neq \theta$, $var(\hat{\theta}_1) = 12$, $var(\hat{\theta}_2) = 10$ and $E(\hat{\theta}_3 - \theta)^2 = 6$. Then the most efficient estimator between them is:

a. $\hat{\theta}_1$ **b.** $\hat{\theta}_2$ **c.** $\hat{\theta}_3$ **d.** None of the above.

[15] Let X be a random variable with mean μ and variance σ^2. Given two independent random samples of size 30 and 50 with sample means $\bar{X}_1$ and $\bar{X}_2$, respectively. Show that

$$\bar{X} = \alpha\bar{X}_1 + (1-\alpha)\bar{X}_2$$

is an unbiased estimator of μ. Find the value of α that minimizes $var(\bar{X})$. Let $\hat{\mu} = \frac{\bar{X}_1 + \bar{X}_2}{2}$ 1be another estimator for μ, compare these two estimators. Which do you prefer? Why?

<+><+><+><+><+><+><+><+><+>