STATISTICAL METHODS IN ECONOMICS-II

LESSON: POINT ESTIMATION

LESSON DEVELOPER: KAMLESH AGGARWAL AND NIDHI AGGARWAL

COLLEGE/DEPARTMENT: DEPARTMENT OF ECONOMICS, SPM COLLEGE ABD NATA SUNDARI COLLEGE UNIVERSITY OF DELHI

TABLE OF CONTENTS

Reference: *Jay L .Devore : Probability and Statistics for Engineering and the Sciences, 8th Edition.*

POINT ESTIMATION

Learning Objectives

The need to obtain estimates of relevant population parameters in business and economics can be given by numerous examples, e.g. , a marketing organization may be interested in estimates of average income in a metropolitan area; a production department may desire an estimate of the percentage of defective articles produced by a new production process; or a bank may want an estimate of average interest rates on mortgages in a certain section of the country. In all of these cases, it is very costly or simply impossible to study complete universe to get the required information. Further, in such cases, exact accuracy is not required and estimates derived from sample data would probably provide appropriate information to meet the demands of the practical situation. After completing study of this chapter you will become familiar with such statistical estimation procedures which provide us with the means of obtaining estimates of population parameters with desired degree of precision. You will be able to choose the most appropriate value of a parameter (or the values of several parameters) for a given situation from a possible set of alternatives, as we will discuss various desirable properties of estimators and develop the concept of sampling distribution of statistic and standard error.

Two different types of estimates of population parameters are of interest: 'point estimates' and 'interval estimates'. Suppose we say that the average height of female students in XYZ College is 5.28 feet, we are giving a point estimate. If, on the other hand, we say that the height is 5.28 $\pm$ 0.02 feet, that is, the height lies between 5.26 and 5.30 feet, we are giving an interval estimate. In this chapter we will concentrate on point estimates.

1. General Concepts of Point Estimation

When we use the value of a statistic to estimate a population parameter, we call this **point estimation** and we refer to the value of the statistic as a **point estimate** of the parameter. Correspondingly, we refer to the statistics themselves as **point estimators**. For example, sample mean,$\bar{X}$, may be used as a point estimator of population mean,μ , in which case $\bar{x}$is a point estimate of this parameter. Similarly, sample variance,S^2,may be used as a point estimator of population variance, σ^2, in which case s^2 is a point estimate of this parameter. These estimates are called point estimates because in each case a single number or a single point on the real axis, is used to estimate the parameter.

Now we will explain that estimators themselves are random variables. Usually we describe a sample of size n by the values x_1 , x_2 ... x_n of the random variables X_1, X_2, ... X_n. If sampling is with replacement, X_1, X_2, ... X_n would be independent, identically distributed random variables having probability distribution $f(x)$. Their joint distribution would then be

P(X_1= x_1, X_2= x_2,........, X_n= x_n) = $f(x_1)$ $f(x_2)$..... $f(x_n)$

Now we can use the sample values x_1 , x_2 ... x_n to compute some statistic (mean, variance etc.) and use this as an estimate of population parameter. Algebraically, a statistic for a sample of size n can be defined as a function of the random variables X_1, X_2, ... X_n , i.e., g(X_1, X_2, ... X_n).The function g(X_1, X_2, ... X_n), that is any statistic, is another random variable, whose values can be represented by g(x_1 , x_2 ... x_n). The same holds true if we have more than one sample. Suppose we take two samples of heights of m male students and n female students at a particular university. We represent sample values by x_1 , x_2 ... x_m and y_1,y_2, ... y_n respectively. The difference between the two sample mean heights is $\bar{x}$-$\bar{y}$, and is the sensible statistic for estimating μ_1- μ_2 , the difference between the two population mean heights. Now the statistic $\bar{x}$ - $\bar{y}$ is a linear combination of two random variables X_1, X_2, ... X_m and Y_1,Y_2, ... Y_n and so itself is a random variable.

Since estimators are random variables, one of the key problems of point estimation is to study their sampling distributions to make a comparison among different estimators. For instance, when we estimate the variance of a population on the basis of a random sample, we can hardly expect that the value of S^2 we get will actually equal σ^2,but it would be reassuring, at least, to know whether we can expect it to be close. Similarly, suppose we draw a random sample of size n from a normal population with mean value μ . Now sample arithmetic mean is a natural statistic for estimating μ. However, median of the population, average of the two extreme observations in the population and k% trimmed mean are also equal to μ , since normal distributions are symmetric. So we can consider any of the following estimators for μ:

(a) Estimator =$\bar{X}$=Arithmetic Mean

(b) Estimator =$\tilde{X}$=Median

(c) Estimator =$\bar{X}_e = \frac{[\min(X_i)+\max(X_i)]}{2}$= the average of the two extreme observations in the sample

(d) Estimator =$\bar{X}_{tr(k)}$= the k% trimmed mean(discard the smallest and largest k% of the sample and then average)

Each one of the estimators (a) - (d) are reasonable point estimators of μ. Since each uses a different measure of the center of the sample to estimate μ so each estimator will give a different estimate for μ.

Example 1 : Consider the accompanying 20 observations on weights of six year old children.

24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
27.98	28.04	28.28	28.49	28.50	28.87	29.11	29.13	29.50	30.88

We assume that the distribution of weights is normal with mean μ. So we can consider $\bar{X}$, $\tilde{X}$, $\bar{X}_e and \bar{X}_{tr(10)}$ as the point estimators for μ. The estimates are 27.793, 27.960, 27.670 and 27.838 respectively. So each estimator is giving a different estimate for μ.

Which of these estimates is closest to the true value? We cannot answer this without knowing the true value of θ(in which case estimation is unnecessary).Questions that can be answered are, "which estimator, when used on other samples of X_i's will tend to produce estimates closest to the true value, which will expose us to the smallest risk, which will give us the most information at the lowest cost and so forth?"To decide which estimator is most appropriate in a given situation, various statistical properties of estimators can be used.

2. Desirable Properties of Point Estimators

The particular properties of estimators that we shall discuss are unbiasedness, efficiency and consistency.

2.1 Unbiased Estimators

In real world there are no perfect estimators that always give the right answer. Thus, it would seem reasonable that an estimator should do so at least on the average, i.e.its expected value should equal the parameter that it is supposed to estimate. If this is the case, the estimator is said to be unbiased, otherwise, it is said to be biased. In other words, if we repeatedly draw random samples from the same population and calculate same statistic for each sample, then the value of statistic will be different for different samples due to sampling fluctuations but the expected or mean value of this statistic should be equal to true parameter value.

Definition: Suppose we denote a point estimator by $\hat{\theta}$. Then $\hat{\theta}$ is an unbiased estimator of the true parameter value θ , if expected value of $\hat{\theta}$ is equal to the true parameter value of θ for every possible value of θ. If this does not hold true then $\hat{\theta}$ is a biased estimator of θ. The difference between the expected value of $\hat{\theta}$ and θ is called the bias of $\hat{\theta}$. It should be noted that expected value means only the arithmetic mean and not any other measure of central value like median, mode etc. of the distribution of $\hat{\theta}$.
In figure 1 below we picture the distributions of biased and unbiased estimators.

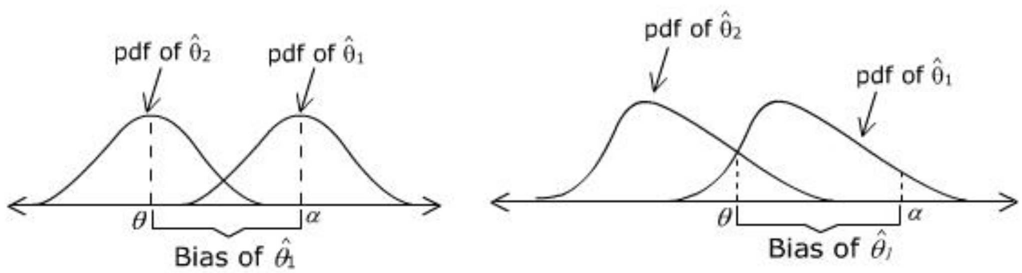

Figure 1. The pdf's of a biased estimator $\hat{\theta}_1$and an unbiased estimator $\hat{\theta}_2$ for a parameter θ.

In figure 1, the sampling distribution of $\hat{\theta}_2$ is centered at the true parameter value θ i.e.E($\hat{\theta}_2$)= θ while the sampling distribution of $\hat{\theta}_1$is centered at α i.e.E($\hat{\theta}_1$)= α. So $\hat{\theta}_2$is an unbiased estimator of θ and $\hat{\theta}_1$ is a biased estimator of θ and bias of $\hat{\theta}_1$=(α- θ).

One may feel that, it is necessary to know the true parameter value to see whether an estimator is biased or unbiased. This is not usually the case because unbiasedness is a general property of the estimator's sampling distribution-where it is centered-which is typically not dependent on any particular parameter value.The following examples will illustrate this:

Example 2: If X, the number of sample successes, is a binomial random variable with parameters n and p, then the sample proportion, $\hat{p} = \frac{X}{n}$ is an unbiased estimator of p irrespective of the true value of p.

Proof: E($\hat{p}$)=E$\left(\frac{X}{n}\right)$=$\frac{1}{n}$E(X) = $\frac{1}{n}(np) = p$

Hence the distribution of the estimator $\hat{p}$ will be centered at the true value p.

Example 3: Let X_1, X_2,-----X_n be a random sample from a normal population with mean μ and variance σ^2, then the estimator $\bar{X}=\frac{\sum X_i}{n}$ is an unbiased estimator of μ while $S^2=\frac{\sum(X_i-\bar{X})^2}{n}$ is a biased estimator of σ^2.

Proof: Since X_1, X_2,-----X_n are random variables having the same distribution as the population, which has mean μ, we have

E(X_i)=μ for i=1,2,........n

Then since the sample mean is defined as $\bar{X}=\frac{X_1+X_2+-----+X_n}{n}$

We have as required

$$E(\bar{X})=\frac{1}{n}[E(X_1)+E(X_2)+\cdots+E(X_n)]=\frac{1}{n}(n\mu)=\mu$$

Hence $\bar{X}$ is an unbiased estimator of μ irrespective of the true value of μ.

However

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

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

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

Then, since $E[(X_i-\mu)^2]=\sigma^2$(given) and $E[(\bar{X}-\mu)^2]=\frac{\sigma^2}{n}$ as shown below

$$we\ have\ \bar{X}=\frac{X_1}{n}+\frac{X_2}{n}+\cdots+\frac{X_n}{n}.\ Then\ since\ X_1, X_2,-----X_n\ are\ independent\ \{hence\ cov(X_i, X_j=0)\}$$

$$and\ have\ variance\ \sigma^2, we\ have\ Var(\bar{X})=\frac{1}{n^2}Var(X_1)+\cdots+\frac{1}{n^2}Var(X_n)=n\left(\frac{1}{n^2}\sigma^2\right)=\frac{\sigma^2}{n}$$

it follows that $E(S^2)=\frac{1}{n}\left[\sum_{i=1}^{n}\sigma^2-n\frac{\sigma^2}{n}\right]=\frac{(n-1)}{n}\sigma^2$

which is very nearly σ^2 only for large values of n(say, n≥30). The desired unbiased estimator is defined by

$\hat{S}^2=\frac{n}{n-1}S^2=\frac{\sum(X_i-\bar{X})^2}{n-1}$ so that $E(\hat{S}^2)=\sigma^2$

Again $\hat{S}^2$ is an unbiased estimator of σ^2 irrespective of the true parameter value.

It can be noted that we have divided the sum of squared deviations by (n-1) instead of n. The reason for this is that by definition we should have taken deviations from μ rather than $\bar{x}$. But we do not know the value of μ so we have to take deviations from $\bar{x}$. Since x_is will always be closer to $\bar{x}$ than to μ so the sum of squared deviations is underestimating the true sum of squared deviations.

Proof:

Denote $\sum(x_i - c)^2$ by L. Now L will be minimised when its first derivative with respect to c is zero and second derivative with respect to c is positive. Differentiating with respect to c we get

$$\frac{\partial L}{\partial C} = 2\sum(x_i - c)(-1) = 0$$

$$\Longrightarrow \sum x_i = nc$$

$$\Longrightarrow \frac{\sum x_i}{n} = c = \bar{x}$$

$$\frac{\partial^2 L}{\partial C^2} = 2n > 0$$

Hence $\sum(x_i - c)^2$ is minimum for $c = \bar{x}$. So if $\mu \neq \bar{x}$ then $\sum(x_i - \mu)^2 > \sum(x_i - \overline{x})^2$.

In order to make a correction for this underestimation we divide by (n-1) rather than n.

Now we will discuss two basic difficulties associated with the concept of unbiasedness. One difficulty associated with the concept of unbiasedness is that it may not be retained under functional transformations, i.e. if $\hat{\theta}$ is an unbiased estimator of θ, it does not necessarily follow that $g(\hat{\theta})$ is an unbiased estimator of $g(\theta)$. For example, although $\hat{S}^2$ is an unbiased estimator of σ^2 but $\hat{S}$ is not an unbiased estimator of σ. Taking the square root messes up the property of unbiasedness. Second difficulty associated with the concept of unbiasedness is that unbiased estimators are not necessarily unique. The following example will illustrate this:

Example 4: Suppose y is approximately proportional to x, that is, $y \cong \beta x$ for some value β. So for any fixed x, Y is a random variable having mean value βx. That is, we assume that the mean value of Y is related to x by a line passing through (0,0) but that the observed value of Y will typically deviate from this line. Now we can consider any of the following three estimators of β.

(1) $\hat{\beta} = \frac{1}{n}\sum\frac{Y_i}{x_i}$

(2) $\hat{\beta} = \frac{\sum Y_i}{\sum x_i}$

(3) $\hat{\beta} = \frac{\sum x_i Y_i}{\sum x_i^2}$

We can show that all three are unbiased.

(1) $E\left(\frac{1}{n}\sum\frac{Y_i}{x_i}\right) = \frac{1}{n}\sum\frac{E(Y_i)}{x_i} = \frac{1}{n}\sum\frac{\beta x_i}{x_i} = \frac{1}{n}\sum\beta = \frac{n\beta}{n} = \beta$

(2) $E\left(\frac{\sum Y_i}{\sum x_i}\right) = \frac{1}{\sum x_i} E(\sum Y_i) = \frac{1}{\sum x_i}(\sum \beta x_i) = \frac{1}{\sum x_i}\beta(\sum x_i) = \beta$

(3) $E\left(\frac{\sum x_i Y_i}{\sum x_i^2}\right) = \frac{1}{\sum x_i^2} E(\sum x_i Y_i) = \frac{1}{\sum x_i^2}\{\sum x_i E(Y_i)\} = \frac{1}{\sum x_i^2}(\sum x_i \beta x_i) = \frac{1}{\sum x_i^2}\beta(\sum x_i^2) = \beta$

Similarly, if X_1, X_2,-----X_n is a random sample from a normal distribution with mean μ, then $\bar{X}$,$\tilde{X}$ and trimmed mean with any percentage are all unbiased estimators of μ.

So the principle of unbiasedness (preferring an unbiased estimator to a biased one) cannot be invoked to select an estimator. What we now need is a criterion for choosing among unbiased estimators.

$\hat{\theta}$ is said to be an unbiased estimator of θ if E$(\hat{\theta})$= θ

2.2 Efficient Estimators

Suppose there are more than one unbiased estimators of θ. Then the question arises which one is the best. To answer this question we look at the spreads of the distributions about θ of various unbiased estimators and select the one which has least spread. In figure 2 below we have shown probability density functions (pdf's)of two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$.

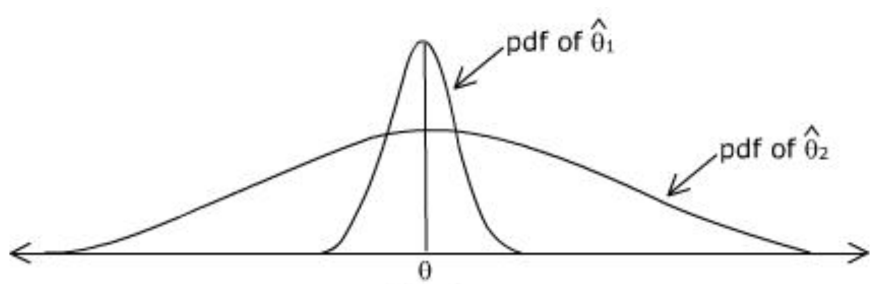

Fig. 2

Figure 2: Graphs of the pdf's of two different unbiased estimators

It can be seen that both $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of θ as pdf of each is centered at θ, but $\hat{\theta}_2$ has more spread as compared to $\hat{\theta}_1$. So we select $\hat{\theta}_1$. $\hat{\theta}_1$ is also called **minimum variance unbiased estimator (MVUE)** of θ as it has least variance among all unbiased estimators of θ.

Example 5: For a normal population, the sampling distributions of the mean and median both have the same mean, namely, the population mean. So both are unbiased estimators.

However, the variance of the sampling distribution of mean is equal to $\frac{\sigma^2}{n}$which is smaller than that of the variance of sampling distribution of median which is equal to $\sigma^2 \frac{\pi}{2n}$=$\frac{1.57\sigma^2}{n}$

Therefore, the mean provides a more efficient estimate than the median and the efficiency of the median relative to the mean is approximately

$$\frac{var\,(\bar{X})}{var\,(\tilde{X})} = \frac{\sigma^2/_n}{1.57\sigma^2/_n} = \frac{1}{1.57}$$

or about 64%. It means that mean requires only 64% as many observations as the median to estimate μ with the same reliability.

However, if a choice is to be made among different estimators on the basis of efficiency criterion, it is quite possible that sometimes a biased estimator is preferable to MVUE as e.g. in figure 3 given below.

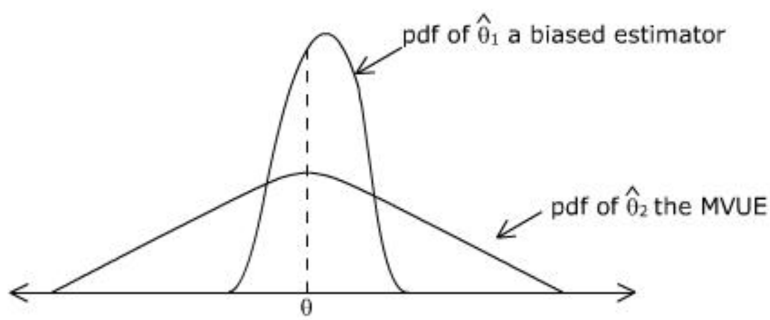

Figure 3: A biased estimator that is preferable to the MVUE.

So we choose $\hat{\theta}_1$, a biased estimator, as it has smaller spread as compared to $\hat{\theta}_2$ which is MVUE.

However, if $\hat{\theta}$ is not an unbiased estimator of a given parameter θ, we judge its merits and make efficiency comparisons on the basis of the expected or **mean squared error (MSE)**, E$\left[(\hat{\theta}-\theta)^2\right]$,instead of the variance of $\hat{\theta}$. If $\hat{\theta}$ is unbiased, then MSE($\hat{\theta}$)= V($\hat{\theta}$), but in general MSE($\hat{\theta}$)= V($\hat{\theta}$)+$(\text{bias})^2$.So if $\hat{\theta}_1$is a biased estimator of θ and $\hat{\theta}_2$ is an unbiased estimator of θ, then we should compare V($\hat{\theta}_2$) with MSE($\hat{\theta}_1$)to make efficiency comparisons.

Question 1: Show that S^2 is a biased but more efficient estimator of population variance σ^2, as compared to $\hat{S}^2$ where S^2=$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ and $\hat{S}^2$ =$\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$

Solution: We have already proved that S^2 is a biased estimator of population variance σ^2 while $\hat{S}^2$ is an unbiased estimator. Now to make efficiency comparisons we will have to compare MSE(S^2) with MSE($\hat{S}^2$). Now since $\hat{S}^2$ is an unbiased estimator so
MSE($\hat{S}^2$) = Var($\hat{S}^2$) while MSE(S^2)=Var(S^2)+$(bias)^2$.
It can be shown that

$$\text{Var}(\hat{S}^2)=\frac{2\sigma^4}{n-1}$$

Since by definition Var($\hat{S}^2$)=$\frac{1}{(n-1)^2}$Var[$\sum(X_i-\bar{X})^2$], it means that

$$\frac{1}{(n-1)^2}\text{Var}[\sum(X_i-\bar{X})^2]=\frac{2\sigma^4}{n-1} \text{and Var}[\sum(X_i-\bar{X})^2]=2\sigma^4(n-1)$$

$$\text{So Var}(S^2)=\text{Var}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=\frac{1}{n^2}2\sigma^4(n-1)$$

Now bias of S^2 is equivalent to E(S^2) - σ^2=$\frac{n-1}{n}\sigma^2$ - σ^2=- $\frac{\sigma^2}{n}$

Hence MSE(S^2)=$\frac{1}{n^2}2\sigma^4(n-1)+\left(-\frac{\sigma^2}{n}\right)^2$=$\frac{2\sigma^4 n-\sigma^4}{n^2}$

Comparing MSE of the two estimators we get

MSE(S^2)- MSE($\hat{S}^2$)=$\frac{2\sigma^4 n-\sigma^4}{n^2}$ - $\frac{2\sigma^4}{n-1}$ = $\frac{\sigma^4(1-3n)}{n^2(n-1)}$<0(since the numerator is always negative and the denominator is always positive).

So MSE($\hat{S}^2$) > MSE(S^2). It means that although S^2 is a biased estimator of population variance σ^2but it is an efficient estimator as compared to unbiased estimator$\hat{S}^2$.Hence, whether we choose S^2 or $\hat{S}^2$ as an estimator of σ^2will depend on whether unbiasedness or efficiency is more important in a particular situation.

Among all estimators of θ, the efficient estimator is one that has minimum mean squared error (MSE), E$(\hat{\theta}-\theta)^2$. If V($\hat{\theta}_1$) $<$ V($\hat{\theta}_2$) where E($\hat{\theta}_1$) = E($\hat{\theta}_2$) = θ then $\hat{\theta}_1$ is efficient.

2.3 Consistent Estimators

Clearly, one would in practice prefer to have estimates that are both efficient and unbiased, but this is not always possible. So the general practice is to consider all unbiased and asymptotically unbiased estimators of θ and select the one that has minimum variance among these. The reason is that sometimes we want to be assured that, at least for large n, the estimators will take on values which are very close to the respective parameters. This concept of closeness is generalized in the following definition of consistency.

Definition: If $\hat{\theta}$ is an unbiased or asymptotically unbiased estimator of the parameter θ and variance $\hat{\theta} \to 0$ as $n \to \infty$, then $\hat{\theta}$ is a consistent estimator of θ. Informally the definition says that when n is sufficiently large, we can be practically certain that the error made with a consistent estimator will be less than any small pre-assigned positive constant.

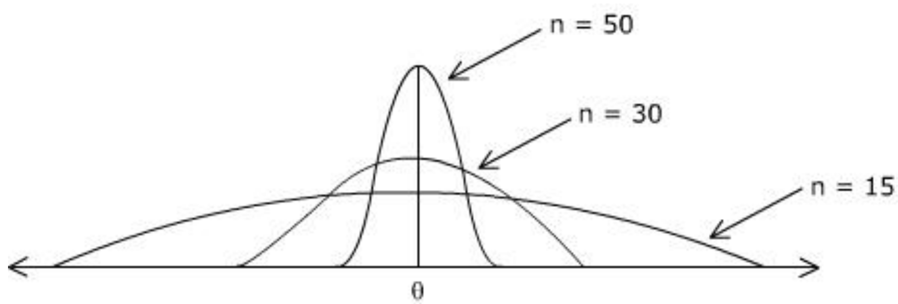

Figure 4 : variance $\hat{\theta} \to 0$ as $n \to \infty$

Note that consistency is an asymptotic property, namely, a limiting property of an estimator. There may be more than one unbiased estimators which are consistent but there can be only one minimum variance unbiased estimator.

Question 2: Show that $\frac{X}{n}$ is a consistent estimator of the binomial parameter θ.

Solution: Since $\frac{X}{n}$ is an unbiased estimator of θ, it remains only to be shown that Var($\frac{X}{n}$)→0 as n→ ∞

$V(\frac{X}{n}) = V(X)/n^2 = npq/n^2 = pq/n$ which tends to zero as $n \to \infty$ as desired.

Question 3: Show that $\bar{X}$ is a consistent estimator of the mean μ of a normal population which has a finite variance.

Solution: Since we have already shown that $\bar{X}$ is an unbiased estimator of μ, it remains only to be shown that Var($\bar{X}$)→0 as n→ ∞

By definition Var$\bar{X}$=E$[(\bar{X} - \mu)^2] = \frac{\sigma^2}{n}$ (as already shown in example 3)

which tends to zero as $n \to \infty$ as desired.

The statistic $\hat{\theta}$ is a consistent estimator of the parameter θ if and only if for each positive constant c, $\lim_{n\to\infty} P(|\hat{\theta} - \theta| \geq c)$ = 0

Now having discussed the desirable properties of point estimators, it is also important to know as to what are the main factors which decide whether an estimator possesses these properties? The most important factor is the sampling distribution of the estimator. However, the sampling distribution of the estimator depends on the distribution of the population from which the sample is drawn. In particular,

1) If we draw a random sample from a normal population, then $\bar{X}$ is the best among the four estimators ($\bar{X}$, $\tilde{X}$, $\bar{X}_e$ and $\bar{X}_{tr(k)}$), since its variance is least among all unbiased estimators.

2) If we draw a random sample from a Cauchy distribution,

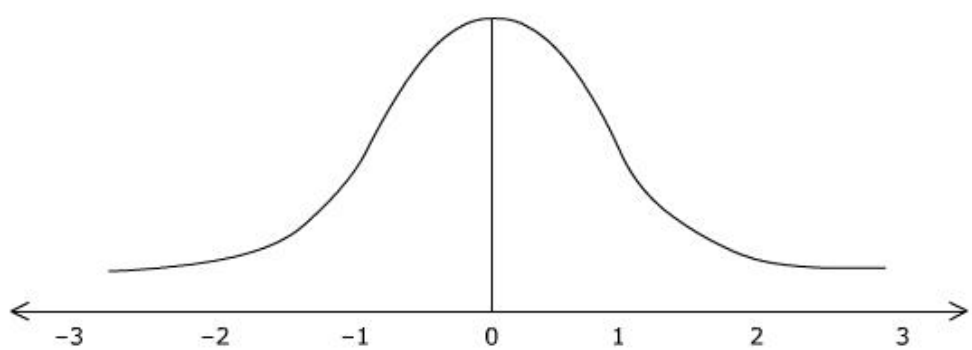

Figure 5 : Cauchy Distribution

then $\bar{X}$ and $\bar{X}_e$ are bad estimators for μ, while $\tilde{X}$ is reasonably good. $\bar{X}$ is bad as it is very sensitive to extreme observations, and due to heavy tails of the Cauchy distribution it is very likely that a few such observations appear in any sample.

3) If we draw a random sample from a uniform distribution, then $\bar{X}_e$ is the best estimator. $\bar{X}_e$ is very sensitive to extreme observations but such observations are unlikely to appear in any sample as uniform distribution does not have any tails.

4) The trimmed mean is not best in any of these three situations. However it is quite good in all three. Hence $\bar{X}_{tr(k)}$ with small trimming percentage is called a **robust estimator** i.e. one that performs reasonably well for a wide variety of population distributions.

So both i.e. distribution of population and sampling distribution of estimator are important to decide which estimator is best for a given situation.

3. Precision of the Estimate

Whenever we are making an inference about a population parameter on the basis of sample statistic, we are also interested in, as to, how much it is reliable. The best indicator is standard error of the relevant estimator which we can denote by $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$. It is the size of an average deviation between $\hat{\theta} and\, \theta$. If we use estimated values of some unknown parameters, then we call it estimated standard error and denote it by $\hat{\sigma}_{\hat{\theta}}$ or by $S_{\hat{\theta}}$.

Example 6: Let X_1 ,X_2,-----X_n be a random sample from a normal population, then the standard error of $\hat{\mu}$=$\bar{X}$is given by$\sigma_{\bar{X}}$=$\frac{\sigma}{\sqrt{n}}$. If, we do not know, the value of σ then we can substitute the estimate $\hat{\sigma}$=s into $\sigma_{\bar{X}}$ to obtain the estimated standard error $\hat{\sigma}_{\bar{X}} = s_{\bar{X}}$.

Question 4: Find out the standard error of sample proportion $\hat{p} = \frac{X}{n}$ where X is a binomial random variable with parameters n and p.

Solution: $\sigma_{\hat{p}} = \sqrt{V(\frac{X}{n})} = \sqrt{V(X)/n^2} = \sqrt{npq/n^2} = \sqrt{pq/n}$ .Since p and q are unknown so we substitute $\hat{p} = \frac{x}{n}$ and $\hat{q} = \frac{n-x}{n}$ into $\sigma_{\hat{p}}$ yielding the estimated standard error $\hat{\sigma}_{\hat{p}} = \sqrt{\hat{p}\hat{q}/n}$.

We can also use the standard error of the estimator used to convert point estimate into interval estimate.

Suppose sample size is large, then distribution of point estimator $\hat{\theta}$ will be approximately normal and we can be reasonably confident that the true value of θ would lie within approximately two standard errors of $\hat{\theta}$. Thus point estimate $\hat{\theta}$ translates to the interval estimate $\hat{\theta} \pm 2\, S.E_{\hat{\theta}}$.

If $\hat{\theta}$ is unbiased but distribution is not normal, then we can be reasonably confident that the true value of θ would lie within approximately four standard error of $\hat{\theta}$. In summary, the standard error tells us roughly within what distance of $\hat{\theta}$ we can expect the true value of θ to lie.

4. Methods of Point Estimation

As we have seen in this chapter, there can be many different ways (estimators) of estimating a parameter of a population. Further different estimators have various desirable properties to varying degrees. Therefore, it would seem desirable to have some general methods that yield estimators with reasonable desirable properties. Here we will discuss two such methods, the **method of moments**, which is historically one of the oldest methods and the **method of maximum likelihood**. Although maximum likelihood estimators are generally preferable to moment estimators because of certain efficiency properties, they often require significantly more computation than do moment estimators.

4.1 The Method of Moments

Let $X_1, X_2, \text{-----} X_n$ be a random sample from a pmf or pdf $f(x)$. For k=1,2,3,......, the kth population moment, or kth moment of the distribution $f(x)$, is E(X^k).The kth sample moment is denoted by m'_k ; symbolically,$m'_k=\frac{1}{n}\sum_{i=1}^{n} X_i^k$

Thus the first population moment is E(X)=μ, and the first sample moment is $\frac{\sum X_i}{n}$ = $\bar{X}$. The second population and sample moments are E(X^2) and $\frac{\sum Xi^2}{n}$ respectively.

The method of moments consists of equating the first few moments of a population to the corresponding moments of a sample, thus getting as many equations as are needed to solve for the unknown parameters of the population.

Thus the method of moments consists of solving the system of equations

$m'_k=\mu'_k$ k=1,2-----p

for the p parameters of the population.

Example 7: If we want to estimate the parameter p of the binomial distribution when n is known, then the system of equations we have to solve is $m_1' = \mu_1'$

Since $\mu_1' = \mathrm{np}$ so $m_1' = \mathrm{np}$

Hence $\hat{p} = \frac{m_1'}{n}$

If both n and p are unknown, then the system of equations we shall have to solve is

$m_1' = \mu_1'$ and $m_2' = \mu_2'$

Since $\mu_1' = \mathrm{np}$ and $\mu_2' = \mathrm{npq}+(\mu_1')^2$

we get

$m_1'=\mathrm{np}$ and $m_2' = \mathrm{npq} + (\mathrm{np})^2$

and solving these two equations for n and p, we find the estimates of the two parameters of the binomial distribution.

Since $m_2' = \text{npq} + (\text{np})^2$

$$\Rightarrow \quad m_2' = m_1'\text{q} + (m_1')^2$$

$$\Rightarrow \quad \frac{m_2' - (m_1')^2}{m_1'} = \hat{q} = (1-\hat{p})$$

$$\Rightarrow \hat{p} = 1 - \frac{m_2' - (m_1')^2}{m_1'}$$

Similarly, since $m_1' = \text{np}$

$$\Rightarrow \hat{n} = \frac{m_1'}{\hat{p}} = \frac{m_1'}{1 - \frac{m_2' - (m_1')^2}{m_1'}}$$

$$= \frac{m_1'}{\frac{m_1' - m_2' - (m_1')^2}{m_1'}}$$

$$= \frac{(m_1')^2}{m_1' - m_2' - (m_1')^2}$$

Question 5: Given a random sample of size n from a uniform population with β=1, use the method of moments to obtain a formula for estimating the parameter α.

Solution: The equation that we shall have to solve is $m_1' = \mu_1'$, where $m_1' = \bar{x}$ and

$\mu_1' = \frac{\alpha+\beta}{2} = \frac{\alpha+1}{2}$. Thus, $\bar{x} = \frac{\alpha+1}{2}$ and we can write the estimate of α as $\hat{\alpha} = 2\bar{x} - 1$.

4.2 The Method of Maximum Likelihood

The method of maximum likelihood looks at the values of a random sample and then chooses as our estimate of the unknown population parameter, the value for which the probability of obtaining the observed data is a maximum. The principle on which the method of maximum likelihood is based can be understood with the following example.

Example 8: Suppose Mr X receives five letters on some particular day, but unfortunately one of them gets misplaced before he has a chance to open it. If among the remaining four letters three contain credit-card billings and the other one does not, what might be a good estimate of k, the total number of credit-card billings among the five letters received? Clearly k must be three or four. Assuming that each letter had the same chance of being misplaced, we find that the probability of the observed data is

$$\frac{\binom{3}{3}\binom{2}{1}}{\binom{5}{4}} = \frac{2}{5} \text{ for } k=3$$

and

$$\frac{\binom{4}{3}\binom{1}{1}}{\binom{5}{4}} = \frac{4}{5} \text{ for } k=4$$

Therefore, if we choose as our estimate of k the value that maximizes the probability of getting the observed data, we obtain k=4. We call this estimate a maximum likelihood estimate and the method by which it was obtained is called the method of maximum likelihood.

In the general case, if the observed sample values are x_1 , x_2 ,...... x_n,we can write in the discrete case

P($X_1 = x_1$, $X_2 = x_2$,......,$X_n = x_n$)= $f(x_1 , x_2 ,...... x_n ; \theta)$ which is just the value of the joint probability distribution of the random variables X_1 , X_2 ,...... X_n at the sample point (x_1 , x_2 ,...... x_n). Since the sample values have been observed and are therefore fixed numbers, we regard $f(x_1 , x_2 ,...... x_n ; \theta)$ as the value of a function of the parameter θ,referred to as the **likelihood function** L(θ). A similar definition applies when the random sample comes from a continuous population, but in that case $f(x_1 , x_2 ,...... x_n ; \theta)$ is the value of the joint probability density at the sample point (x_1 , x_2 ,...... x_n). The method of maximum likelihood consists of maximizing the likelihood function with respect to θ, and we refer to the value of θ which maximizes the likelihood function as the maximum likelihood estimate of θ.To maximize L(θ)= $f(x_1 , x_2 ,...... x_n ; \theta)$ we take the derivative of L(θ) with respect to θ and set it equal to zero.

The method is capable of generalization. In case there are several parameters, we take the partial derivatives with respect to each parameter, set them equal to zero, and solve the resulting equations simultaneously. Moreover if we draw a large sample from a population which has well specified distribution function then maximum likelihood estimate of any parameter θ will be approximately MVUE i.e. it will be approximately unbiased and approximately have least variance.

Question 6: Given x"successes" in n trials, find the maximum likelihood estimator of the parameter θ of the binomial distribution.

Solution: To find the value of θ which maximizes

L(θ)=b(x;n, θ)=$\binom{n}{x}\theta^{x}(1-\theta)^{n-x}$, it will be convenient to make use of the fact that the value of θ which maximizes L(θ) will also maximize

$$\ln L(\theta)=\ln\binom{n}{x}+x\cdot\ln\theta+(n-x)\cdot\ln(1-\theta)$$

Thus we get $\frac{d[\ln L(\theta)]}{d\theta}=\frac{x}{\theta}-\frac{n-x}{1-\theta}$

and, equating this derivative to 0 and solving for θ, we find that the likelihood function has a maximum at $\theta=\frac{x}{n}$. Hence the maximum likelihood estimator of the parameter θ of the binomial distribution is $\hat{\theta}=\frac{X}{n}$.

Question 7: Suppose that n observations, X_1 , X_2 ,...... X_n are made from a normally distributed population. Find

(a) the maximum likelihood estimate of the mean if variance is known but mean is unknown

(b) the maximum likelihood estimate of the variance if mean is known but variance is unknown.

Solution:

(a) Since f(x_k,μ) = $\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x_k-\mu)^2}/2\sigma^2$

we have

(1) L = f(x_1,μ)........ f(x_n,μ) = $(2\pi\sigma^2)^{-n/2}\, e^{-\sum(x_k-\mu)^2}/2\sigma^2$

Therefore,

(2) ln L = $-\frac{n}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum(x_k-\mu)^2$

Taking the partial derivative with respect to μ yields

(3) $\frac{1}{L}\frac{\partial L}{\partial\mu}=\frac{1}{\sigma^2}\sum(x_k-\mu)$

Setting $\frac{\partial L}{\partial\mu}$ = 0 gives

(4) $\sum(x_k-\mu)$= 0 i.e. $\sum x_k-\mathrm{n}\mu$ = 0

or

(5) $\mu=\frac{\sum x_k}{n}$

Therefore the maximum likelihood estimate is the sample mean.

(b) Since f(x_k,σ^2) = $\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x_k-\mu)^2}/2\sigma^2$

we have

(1) L = f(x_1,μ)........ f(x_n,μ) = $(2\pi\sigma^2)^{-n/2}\ e^{-\sum(x_k-\mu)^2/2\sigma^2}$

Therefore,

(2) ln L = -$\frac{n}{2}$ ln$(2\pi\sigma^2)$ - $\frac{1}{2\sigma^2}\sum(x_k-\mu)^2$

Taking the partial derivative with respect to σ^2 yields

(3) $\frac{1}{L}\frac{\partial L}{\partial\sigma^2}$ = -$\frac{n}{2\sigma^2}$ +$\frac{1}{2(\sigma^2)^2}\sum(x_k-\mu)^2$

Setting $\frac{\partial L}{\partial\sigma^2}$ = 0 gives

$\sigma^2 = \frac{\sum(x_k-\mu)^2}{n}$

Question 8: Prove that the maximum likelihood estimate of the parameter α of a population having density function: $\frac{2}{\alpha^2}(\alpha - x)$,0< $x < \alpha$, for a sample of unit size is $2x$, x being the sample value. Show also that the estimate is biased.

Solution: Sample of unit size $\Longrightarrow$ $n = 1$

likelihood function L(α) = $\frac{2}{\alpha^2}(\alpha - x)$ = f(x,α)

logL(α) = log2 - log α^2+ log($\alpha - x$)

= log2 - 2log α + log($\alpha - x$)

Differentiating w.r.t. α we get

$\frac{d}{d\,\alpha}[L(\alpha)]$= - $\frac{2}{\alpha}$ + $\frac{1}{(\alpha - x)}$

$\frac{d^2}{d\,\alpha^2}[L(\alpha)]$ = $\frac{2}{\alpha^2}$ - $\frac{1}{(\alpha-x)^2}$

For maxima or minima $\frac{d}{d\,\alpha}[L(\alpha)]$ = 0

∴ - $\frac{2}{\alpha}$ + $\frac{1}{(\alpha-x)}$ = 0 $\Longrightarrow \frac{2}{\alpha} = \frac{1}{(\alpha-x)} \Longrightarrow 2\,\alpha - 2x = \alpha \Longrightarrow \alpha = 2\,x$

When $\alpha = 2x$,

$$\frac{d^2}{d\,\alpha^2}[L(\alpha)] = \frac{2}{4x^2} - \frac{1}{(x)^2} = \frac{-1}{2x^2} < 0\ i.e. \frac{d^2}{d\,\alpha^2}[L(\alpha)]\ is\ -ve$$

∴Maximum likelihood estimator of α is given by $\hat{\alpha} = 2x$

E($\hat{\alpha}$) = E($2X$) = 2$\int_0^\alpha \frac{2}{\alpha^2}x(\alpha - x)dx = \frac{4}{\alpha^2}\left|\frac{\alpha x^2}{2} - \frac{x^3}{3}\right|_0^\alpha = \frac{2}{3}\alpha$

Since E($\hat{\alpha}$) ≠ α, $\hat{\alpha} = 2x$ is not an unbiased estimate of α.

Practice Questions:

Q.1 Assuming that the population is normal, give examples of estimators (or estimates) which are

(a) unbiased and efficient

(b) unbiased and inefficient

(c) biased and inefficient.

Q.2 Show that $\bar{X}$ is a minimum variance unbiased estimator of the mean μ of a normal population.

Q.3 If $\widehat{\theta}$ is an estimator of a parameter θ, its bias is given by b=E($\widehat{\theta}$)- θ. Show that $E\left[(\hat{\theta}-\theta)^2\right]=V(\hat{\theta})+(b)^2$.

Q.4 If $\hat{\theta}_1$and $\hat{\theta}_2$ are unbiased estimators of the same parameter θ, what condition must be imposed on the constants k_1 and k_2 so that $k_1\hat{\theta}_1+ k_2\hat{\theta}_2$ is also an unbiased estimator of θ ?

Q.5 Suppose that we use the largest value of a random sample of size n to estimate the parameter θ of the population.

$f(x)= \frac{1}{\theta}$ $\quad for\ 0 < x < \theta$

= 0 Otherwise

Check whether this estimator is (a) unbiased and (b) consistent.

Q.6 Show that for a random sample from a normal population, the sample variance $\hat{S}^2$ is a consistent estimator of σ^2 where $\hat{S}^2 =\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$.

Q.7 In estimating the mean μ of a normal population on the basis of a random sample of size 2n+1, what is the efficiency of the median relative to the mean?

Q.8 If x_1 , x_2 ,...... $x_n$are the values of a random sample of size n from a population having the density

$f(x;\theta) = 2(\theta-x)/\,\theta^2$ $\quad for\ 0 < x < \theta$

= 0 otherwise

find an estimator for θ by the method of moments.

Q.9 Let X_1,...X_n be a random sample from a gamma distribution with parameters α and β.

- **a.** Derive the equations whose solutions yield the maximum likelihood estimators of $\alpha$and β. Do you think they can be solved explicitly ?
- **b.** Show that the mle of $\mu = \alpha\beta$ is $\hat{\mu}= \bar{X}$.

Q.10 Among N independent random variables having identical binomial distribution with the parameters θ and n=2, n_0 take on the value zero, $n_1$take on the value one, and n_2 take on the value two. Find an estimate of θ using

(a) the method of moments

(b) the method of maximum likelihood.