Chapter 6

Point Estimation

Given a parameter of interest, such as a population mean μ or population proportion p, the objective of point estimation is to use a sample to compute a number that represents in some sense a good guess for the true value of the parameter.

6.1 Some General Concepts of Point Estimation

When discussing general concepts and methods of inference, it is convenient to have a generic symbol for the parameter of interest. We will use the Greek letter θ for this purpose.

A **point estimate** of a parameter θ is a single number that can be regarded as a sensible value for θ. A point estimate is obtained by selecting a suitable statistic and computing its value from the given sample data. The selected statistic is called the **point estimator** of θ, often denoted by $\hat{\theta}$.

Given a random sample $x_1, \cdots, x_n$ from a population with unknown parameter μ and variance σ^2, then

- the sample mean $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is a point estimate of μ
- the sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right)$ is a point estimate of σ^2

If we write $\hat{\theta} = \theta +$ error of estimation
then an accurate estimator would be one resulting in small estimation errors, so that estimated values will be near the true value.

A point estimator $\hat{\theta}$ is said to be an **unbiased estimator** of θ if $E(\hat{\theta}) = \theta$. If $\hat{\theta}$ is not unbiased, the difference $E(\hat{\theta}) - \theta$ is called the **bias** of $\hat{\theta}$.

That is, $\hat{\theta}$ is unbiased if its sampling distribution is always centered at the true value of parameter θ. When choosing among several different estimators of θ, select one that is unbiased.

Examples of unbiased estimators

If $X_1, \cdots, X_n$ is a random sample from a distribution with mean μ and variance σ^2

- Sample mean $\overline{X}$ is an unbiased estimator of μ.
- If the distribution is continuous and symmetric, sample median $\widetilde{X}$ and any trimmed mean $\overline{X}_{tr(100\alpha)}$ are also unbiased estimators of μ.
- Sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n\overline{X}^2\right)$ is an unbiased estimator of σ^2.

If $X \sim Bin(n, p)$, then the sample proportion of successes $\hat{p} = \frac{X}{n}$ is an unbiased estimator of p.

Among all estimators of θ that are unbiased, choose the one that has minimum variance. The resulting $\hat{\theta}$ is called the **minimum variance unbiased estimator (MVUE)** of θ.

One of the triumphs of mathematical statistics has been the development of methodology for identifying the MVUE in a wide variety of situations. The most important result of this type for our purposes concerns estimating the mean μ of a normal distribution.

Let $X_1, \cdots, X_n$ be a random sample from a normal distribution with parameters μ and σ.

- Sample mean $\hat{\mu} = \overline{X}$ is the MVUE of μ.
- Sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is the MVUE estimator of σ^2.

The **standard error** of an estimator $\hat{\theta}$ is its standard deviation $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$. It is the magnitude of a typical or representative deviation between an estimate and the value of θ.

If the standard error itself involves unknown parameters whose values can be estimated, substitution of these estimates into $\sigma_{\hat{\theta}}$ yields the **estimated standard error** ($\hat{\sigma}_{\hat{\theta}}$ or $s_{\hat{\theta}}$) of the estimator.

<u>Exercise 6.2</u> A sample of 20 students who had recently taken elementary statistics yielded the following information on brand of calculator owned (T = Texas Instruments, H = Hewlett Packard, C = Casio, S = Sharp):

T	T	H	T	C	T	T	S	C	H
S	S	T	H	C	T	T	T	H	T

(a) Estimate the true proportion of all such students who own a Texas Instruments calculator.

(b) Of the 10 students who owned a TI calculator, 4 had graphing calculators. Estimate the proportion of students who do not own a TI graphing calculator.

<u>Exercise 6.10</u> Using a long rod that has length μ, you are going to lay out a square plot in which the length of each side is μ. Thus the area of the plot will be μ^2. However, you do not know the value of μ, so you decide to make n independent measurements $X_1, X_2, \cdots, X_n$ of the length. Assume that each X_i has mean μ and variance σ^2.

(a) Show that $\overline{X}^2$ is not an unbiased estimator for μ^2.

(b) For what value of k is the estimator $\overline{X}^2 - kS^2$ unbiased for μ^2.

6.2 Methods of Point Estimation

The Method of Moments

Let $X_1, \cdots, X_n$ be a random sample from a pmf or pdf $f(x)$. For $k = 1, 2, \cdots$

- The k**th population moment** or k**th moment of the distribution** is $E(X^k)$.
- The k**th sample moment** is $\frac{1}{n}\sum_{i=1}^{n} X_i^k$.

Let $X_1, \cdots, X_n$ be a random sample from a distribution with pmf or pdf $p(x; \theta_1, \cdots, \theta_m)$, where $\theta_1, \cdots, \theta_m$ are parameters whose values are unknown. Then the **moment estimators** $\hat{\theta}_1, \cdots, \hat{\theta}_m$ are obtained by equating the first m sample moments to the corresponding first m population moments and solving for $\theta_1, \cdots, \theta_m$.

The basic idea of this method is to equate certain sample characteristics, such as the mean, to the corresponding population expected values. Then solving these equations for unknown parameter values yields the estimators.

<u>Example</u> Suppose $X_1, \cdots, X_n$ is a random sample from a Bernoulli distribution with parameter p. That is, each X_i takes the value 1 with probability p and the value 0 with probability $1-p$. Find the moment estimator of p. Is the moment estimator unbiased?

<u>Example</u> Suppose $X_1, \cdots, X_n$ is a random sample from a normal distribution with parameters μ and σ. Find the moment estimators of μ and σ^2. Are they unbiased?

Maximum Likelihood Estimation

Let $X_1, \cdots, X_n$ have joint pmf or pdf

$$f(x_1, \cdots, x_n; \theta_1, \cdots, \theta_m) = f(x_1, \cdots, x_n; \boldsymbol{\Theta})$$

where the parameters $\boldsymbol{\Theta} = \{\theta_1, \cdots, \theta_m\}$ have unknown values.

When $x_1, \cdots, x_n$ are the observed sample values and $L = f(x_1, \cdots, x_n; \boldsymbol{\Theta})$ is regarded as a function of $\boldsymbol{\Theta} = \{\theta_1, \cdots, \theta_m\}$, it is called the **likelihood function**.

The **maximum likelihood estimates (mle's)** $\hat{\boldsymbol{\Theta}} = \{\hat{\theta}_1, \cdots, \hat{\theta}_m\}$ are those values of the θ_i's that maximize the likelihood function.

When the X_i's are substituted in place of the x_i's, the **maximum likelihood estimators** result.

The likelihood function tells us how likely the observed sample is as a function of the possible parameter values. Maximizing the likelihood gives the parameter values for which the observed sample is most likely to have been generated — that is, the parameter values that "agree most closely" with the observed data.

Notes on finding the mle

1. If $X_1, \cdots, X_n$ is a random sample, i.e. X_i's are independent and identically distributed (iid). Because of independence, the likelihood function L is a product of the individual pmf's or pdf's.

2. Finding $\mathbf{\Theta}$ to maximize $\ln(L)$ is equivalent to maximizing L itself. In statistics, taking the logarithm frequently changes a product to a sum, which is easier to work with.

$$\begin{aligned}
\ln(xy) &= \ln(x) + \ln(y) \\
\ln(x/y) &= \ln(x) - \ln(y) \\
\ln(x^y) &= y \ln(x)
\end{aligned}$$

3. To find the values of θ_i's that maximize $\ln(L)$, we must take the partial derivatives of $\ln(L)$ or with respect to each θ_i, equate them to zero, and solve the equations for θ_i's. This solution is $\hat{\mathbf{\Theta}} = \{\hat{\theta}_1, \cdots, \hat{\theta}_m\}$, the mle.

Example Suppose $X_1, \cdots, X_n$ is a random sample from a Bernoulli distribution with parameter p. That is, each X_i takes the value 1 with probability p and the value 0 with probability $1 - p$. Find the mle of p. Is the mle unbiased?

Example Suppose $X_1, \cdots, X_n$ is a random sample from a normal distribution with parameters μ and σ. Find the mle's of μ and σ^2. Are they unbiased?

Exercise 6.22 Let X denote the proportion of allotted time that a randomly selected student spends working on a certain aptitude test. Suppose the pdf of X is

$$f(x;\theta) = \begin{cases} (\theta+1)x^{\theta} & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

where $-1 < \theta$. A random sample of ten students yields data $x_1 = .92, x_2 = .79, x_3 = .90, x_4 = .65, x_5 = .86, x_6 = .47, x_7 = .73, x_8 = .97, x_9 = .94$, and $x_{10} = .77$.

(a) Use the method of moments to obtain an estimator of θ, and then compute the estimate for the data.

(b) Obtain the maximum likelihood estimator of θ, and then compute the estimate for the given data.