For example, if you toss a coin 100 times and want to find out the probability of obtaining 60 heads, you put $p = 0.5$, $n = 100$ and $x = 60$ in the above formula. Computer routines exist to evaluate such probabilities.

You can see how the binomial distribution is a generalization of the Bernoulli distribution.

The Poisson Distribution

A random X variable is said to have the Poisson distribution if its PDF is:

$$f(X) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad \text{for } x = 0, 1, 2, \ldots, \lambda > 0$$

The Poisson distribution depends on a single parameter, λ. A distinguishing feature of the Poisson distribution is that its variance is equal to its expected value, which is λ. That is,

$$E(X) = \text{var}(X) = \lambda$$

The Poisson model, as we saw in the chapter on nonlinear regression models, is used to model rare or infrequent phenomena, such as the number of phone calls received in a span of, say, 5 minutes, or the number of speeding tickets received in a span of an hour, or the number of patents received by a firm, say, in a year.

A.7 Statistical Inference: Estimation

In Section A.6 we considered several theoretical probability distributions. Very often we know or are willing to assume that a random variable X follows a particular probability distribution but do not know the value(s) of the parameter(s) of the distribution. For example, if X follows the normal distribution, we may want to know the value of its two parameters, namely, the mean and the variance. To estimate the unknowns, the usual procedure is to assume that we have a **random sample** of size n from the known probability distribution and use the sample data to estimate the unknown parameters.[5] This is known as the **problem of estimation.** In this section, we take a closer look at this problem. The problem of estimation can be broken down into two categories: point estimation and interval estimation.

Point Estimation

To fix the ideas, let X be a random variable with PDF $f(x; \theta)$, where θ is the parameter of the distribution (for simplicity of discussion only, we are assuming that there is only one unknown parameter; our discussion can be readily generalized). Assume that we know the functional form—that is, we know the theoretical PDF, such as the t distribution—but do not know the value of θ. Therefore, we draw a random sample of size n from this known PDF and then develop a function of the sample values such that

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

provides us an estimate of the true θ. $\hat{\theta}$ is known as a **statistic,** or an **estimator,** and a particular numerical value taken by the estimator is known as an **estimate.** Note that $\hat{\theta}$ can be

[5] Let $X_1, X_2, \ldots, X_n$ be n random variables with joint PDF $f(x_1, x_2, \ldots, x_n)$. If we can write

$$f(x_1, x_2, \ldots, x_n) = f(x_1)\, f(x_2) \cdots f(x_n)$$

where $f(x)$ is the common PDF of each X, then $x_1, x_2, \ldots, x_n$ are said to constitute a random sample of size n from a population with PDF $f(x_n)$.

treated as a random variable because it is a function of the sample data. $\hat{\theta}$ provides us with a rule, or formula, that tells us how we may estimate the true θ. Thus, if we let

$$\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \bar{X}$$

where $\bar{X}$ is the sample mean, then $\bar{X}$ is an estimator of the true mean value, say, μ. If in a specific case $\bar{X} = 50$, this provides an *estimate of* μ. The estimator $\hat{\theta}$ obtained previously is known as a **point estimator** because it provides only a single (point) estimate of θ.

Interval Estimation

Instead of obtaining only a single estimate of θ, suppose we obtain two estimates of θ by constructing two estimators $\hat{\theta}_1(x_1, x_2, \ldots, x_n)$ and $\hat{\theta}_2(x_1, x_2, \ldots, x_n)$, and say with some confidence (i.e., probability) that the interval between $\hat{\theta}_1$ and $\hat{\theta}_2$ includes the true θ. Thus, in interval estimation, in contrast with point estimation, we provide a range of possible values within which the true θ may lie.

The key concept underlying interval estimation is the notion of the **sampling,** or **probability distribution, of an estimator.** For example, it can be shown that if a variable X is normally distributed, then the sample mean $\bar{X}$ is also normally distributed with mean $= \mu$ (the true mean) and variance $= \sigma^2/n$, where n is the sample size. In other words, the sampling, or probability, distribution of the estimator $\bar{X}$ is $\bar{X} \sim N(\mu, \sigma^2/n)$. As a result, if we construct the interval

$$\bar{X} \pm 2\frac{\sigma}{\sqrt{n}}$$

and say that the probability is approximately 0.95, or 95 percent, that intervals like it will include the true μ, we are in fact constructing an interval estimator for μ. Note that the interval given previously is random since it is based on $\bar{X}$, which will vary from sample to sample.

More generally, in interval estimation we construct two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, both functions of the sample X values, such that

$$\Pr(\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2) = 1 - \alpha \qquad 0 < \alpha < 1$$

That is, we can state that the probability is $1 - \alpha$ that the interval from $\hat{\theta}_1$ to $\hat{\theta}_2$ contains the true θ. This interval is known as a **confidence interval** of size $1 - \alpha$ for θ, $1 - \alpha$ being known as the **confidence coefficient.** If $\alpha = 0.05$, then $1 - \alpha = 0.95$, meaning that if we construct a confidence interval with a confidence coefficient of 0.95, then in repeated such constructions resulting from repeated sampling we shall be right in 95 out of 100 cases if we maintain that the interval contains the true θ. When the confidence coefficient is 0.95, we often say that we have a 95 percent confidence interval. In general, if the confidence coefficient is $1 - \alpha$, we say that we have a $100(1 - \alpha)\%$ confidence interval. Note that α is known as the **level of significance,** or the probability of committing a Type I error. This topic is discussed in Section A.8.

EXAMPLE 24

Suppose that the distribution of height of men in a population is normally distributed with mean $= \mu$ inches and $\sigma = 2.5$ inches. A sample of 100 men drawn randomly from this population had an average height of 67 inches. Establish a 95 percent confidence interval for the mean height ($= \mu$) in the population as a whole.

As noted, $\bar{X} \sim N(\mu, \sigma^2/n)$, which in this case becomes $\bar{X} \sim N(\mu, 2.5^2/100)$. From Table D.1 one can see that

$$\bar{X} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}$$

EXAMPLE 24
(*Continued*)

covers 95 percent of the area under the normal curve. Therefore, the preceding interval provides a 95 percent confidence interval for μ. Plugging in the given values of $\bar{X}$, σ, and n, we obtain the 95 percent confidence interval as

$$66.51 \le \mu \le 67.49$$

In repeated such measurements, intervals thus established will include the true μ with 95 percent confidence. A technical point may be noted here. Although we can say that the probability that the random interval $[\bar{X} \pm 1.96(\sigma/\sqrt{n})]$ includes μ is 95 percent, we *cannot* say that the probability is 95 percent that the particular interval (66.51, 67.49) includes μ. Once this interval is fixed, the probability that it will include μ is either 0 or 1. What we can say is that if we construct 100 such intervals, 95 out of the 100 intervals will include the true μ; we cannot guarantee that one particular interval will necessarily include μ.

Methods of Estimation

Broadly speaking, there are three methods of parameter estimation: (1) least squares (LS), (2) maximum likelihood (ML), and (3) method of moments (MOM) and its extension, the generalized method of moments (GMM). We have devoted considerable time to illustrate the LS method. In Chapter 4 we introduced the ML method in the regression context. But the method is of much broader application.

The key idea behind the ML is the **likelihood function.** To illustrate this, suppose the random variable X has PDF $f(X,\theta)$ which depends on a single parameter θ. We know the PDF (e.g., Bernoulli or binomial) but do not know the parameter value. Suppose we obtain a random sample of nX values. The joint PDF of these n values is:

$$g(x_1, x_2, \ldots, x_n; \theta)$$

Because it is a random sample, we can write the preceding joint PDF as a product of the individual PDF as

$$g(x_1, x_2, \ldots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)$$

The joint PDF has a dual interpretation. If θ is known, we interpret it as the joint probability of observing the given sample values. On the other hand, we can treat it as a function of θ for given values of $x_1, x_2, \ldots, x_n$. On the latter interpretation, we call the joint PDF the **likelihood function (LF)** and write it as

$$L(\theta; x_1, x_2, \ldots, x_n) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)$$

Observe the role reversal of θ in the joint probability density function and the likelihood function.

The ML estimator of θ is that value of θ that maximizes the (sample) likelihood function, L. For mathematical convenience, we often take the log of the likelihood, called the **log-likelihood function (log *L*).** Following the calculus rules of maximization, we differentiate the log-likelihood function with respect to the unknown and equate the resulting derivative to zero. The resulting value of the estimator is called the **maximum-likelihood estimator.** One can apply the second-order condition of maximization to assure that the value we have obtained is in fact the maximum value.

In case there is more than one unknown parameter, we differentiate the log-likelihood function with respect to each unknown, set the resulting expressions to zero, and solve them simultaneously to obtain the values of the unknown parameters. We have already shown this for the multiple regression model (see Chapter 4, Appendix 4A.1).

EXAMPLE 25 Assume that the random variable X follows the Poisson distribution with the mean value of λ. Suppose $x_1, x_2, \ldots, x_n$ are independent Poisson random variables each with mean λ. Suppose we want to find out the ML estimator of λ. The likelihood function here is:

$$L(x_1, x_2, \ldots, x_n; \lambda) = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}\cdots\frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

$$= \frac{e^{-n\lambda}\lambda^{\Sigma x_i}}{x_1!x_2!\cdots x_n!}$$

This is a rather unwieldy expression, but if we take its log, it becomes

$$\log(x_1, x_2, \ldots, x_n; \lambda) = -n\lambda + \sum x_i \log\lambda - \log c$$

where $\log c = \prod x_i!$. Differentiating the preceding expression with respect to λ, we obtain $(-n + (\sum x_i)/\lambda)$. By setting this last expression to zero, we obtain $\lambda_{ml} = (\sum x_i)/n = \bar{X}$, which is the ML estimator of the unknown λ.

The Method of Moments

We have given a glimpse of MOM in Exercise 3.4 in the so-called **analogy principle** in which the sample moments try to duplicate the properties of their population counterparts. The generalized method of moments (GMM), which is a generalization of MOM, is now becoming more popular, but not at the introductory level. Hence we will not pursue it here.

The desirable statistical properties fall into two categories: small-sample, or finite-sample, properties and large-sample, or asymptotic, properties. Underlying both of these sets of properties is the notion that an estimator has a sampling, or probability, distribution.

Small-Sample Properties

Unbiasedness

An estimator $\hat{\theta}$ is said to be an unbiased estimator of θ if the expected value of $\hat{\theta}$ is equal to the true θ; that is,

$$E(\hat{\theta}) = \theta$$

or

$$E(\hat{\theta}) - \theta = 0$$

If this equality does not hold, then the estimator is said to be biased, and the bias is calculated as

$$\text{bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Of course, if $E(\hat{\theta}) = \theta$—that is, $\hat{\theta}$ is an unbiased estimator—the bias is zero.

Geometrically, the situation is as depicted in Figure A.8. In passing, note that unbiasedness is a property of repeated sampling, not of any given sample: Keeping the sample size fixed, we draw several samples, each time obtaining an estimate of the unknown parameter. The average value of these estimates is expected to be equal to the true value if the estimator is unbiased.

Minimum Variance

$\hat{\theta}_1$ is said to be a minimum-variance estimator of θ if the variance of $\hat{\theta}_1$ is smaller than or at most equal to the variance of $\hat{\theta}_2$, which is any other estimator of θ. Geometrically, we have

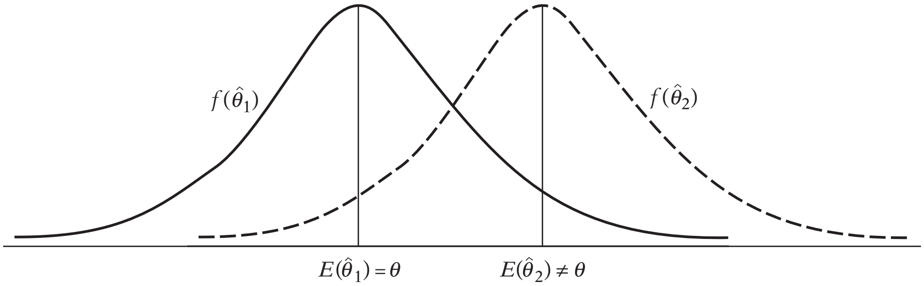


FIGURE A.8
Biased and unbiased estimators.

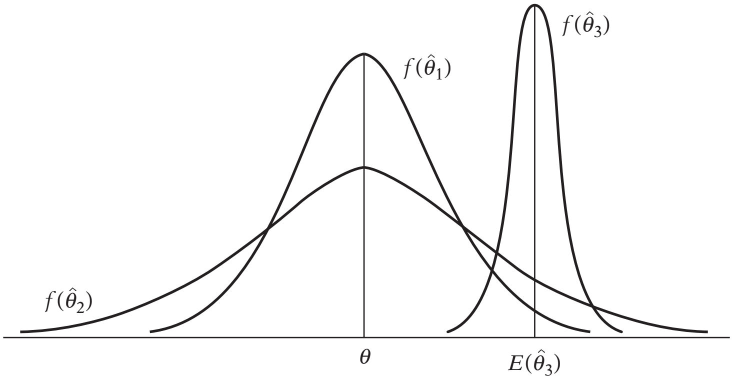


FIGURE A.9
Distribution of three estimators of θ.

Figure A.9, which shows three estimators of θ, namely $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\theta}_3$, and their probability distributions. As shown, the variance of $\hat{\theta}_3$ is smaller than that of either $\hat{\theta}_1$ or $\hat{\theta}_2$. Hence, assuming only the three possible estimators, in this case $\hat{\theta}_3$ is a minimum-variance estimator. But note that $\hat{\theta}_3$ is a biased estimator (why?).

Best Unbiased, or Efficient, Estimator

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two *unbiased* estimators of θ, and the variance of $\hat{\theta}_1$ is smaller than or at most equal to the variance of $\hat{\theta}_2$, then $\hat{\theta}_1$ is a **minimum-variance unbiased,** or **best unbiased,** or **efficient, estimator.** Thus, in Figure A.9, of the two unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, $\hat{\theta}_1$ is best unbiased, or efficient.

Linearity

An estimator $\hat{\theta}$ is said to be a linear estimator of θ if it is a linear function of the sample observations. Thus, the sample mean defined as

$$\bar{X} = \frac{1}{n}\sum X_i = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

is a linear estimator because it is a linear function of the X values.

Best Linear Unbiased Estimator (BLUE)

If $\hat{\theta}$ is linear, is unbiased, and has minimum variance in the class of all linear unbiased estimators of θ, then it is called a **best linear unbiased estimator,** or **BLUE** for short.

Minimum Mean-Square-Error (MSE) Estimator

The MSE of an estimator $\hat{\theta}$ is defined as

$$\text{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

This is in contrast with the variance of $\hat{\theta}$, which is defined as

$$\operatorname{var}(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2$$

The difference between the two is that $\operatorname{var}(\hat{\theta})$ measures the dispersion of the distribution of $\hat{\theta}$ around its mean or expected value, whereas MSE($\hat{\theta}$) measures dispersion around the true value of the parameter. The relationship between the two is as follows:

$$\begin{aligned}
\text{MSE}(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 \\
&= E[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2 \\
&= E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2 + 2E[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta] \\
&= E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2 \qquad \text{since the last term is zero}^6 \\
&= \operatorname{var}(\hat{\theta}) + \text{bias}(\hat{\theta})^2 \\
&= \text{variance of } \hat{\theta} \textit{ plus } \text{square bias}
\end{aligned}$$

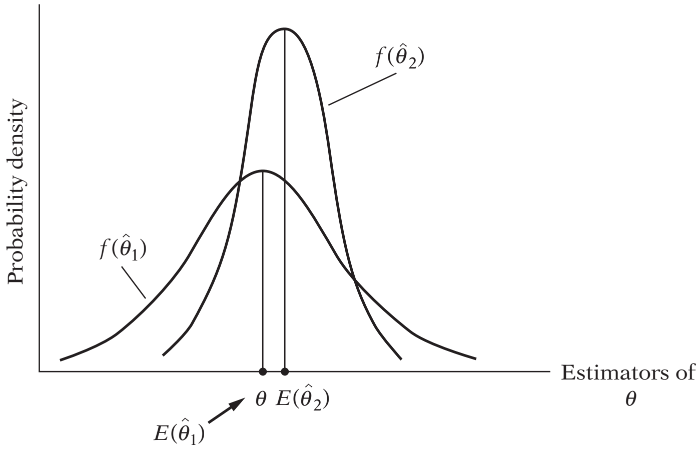
Of course, if the bias is zero, MSE($\hat{\theta}$) = var($\hat{\theta}$).

The minimum MSE criterion consists in choosing an estimator whose MSE is the least in a competing set of estimators. But notice that even if such an estimator is found, there is a tradeoff involved—to obtain minimum variance you may have to accept some bias. Geometrically, the situation is as shown in Figure A.10. In this figure, $\hat{\theta}_2$ is slightly biased, but its variance is smaller than that of the unbiased estimator $\hat{\theta}_1$. In practice, however, the minimum MSE criterion is used when the best unbiased criterion is incapable of producing estimators with smaller variances.

Large-Sample Properties

It often happens that an estimator does not satisfy one or more of the desirable statistical properties in small samples. But as the sample size increases indefinitely, the estimator possesses several desirable statistical properties. These properties are known as the **large-sample,** or **asymptotic, properties.**

FIGURE A.10
Tradeoff between bias and variance.



[6] The last term can be written as $2\{[E(\hat{\theta})]^2 - [E(\hat{\theta})]^2 - \theta E(\hat{\theta}) + \theta E(\hat{\theta})\} = 0$. Also note that $E[E(\hat{\theta}) - \theta]^2 = [E(\hat{\theta}) - \theta]^2$, since the expected value of a constant is simply the constant itself.

Asymptotic Unbiasedness

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

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$

where $\hat{\theta}_n$ means that the estimator is based on a sample size of n and where lim means limit and $n \to \infty$ means that n increases indefinitely. In words, $\hat{\theta}$ is an asymptotically unbiased estimator of θ if its expected, or mean, value approaches the true value as the sample size gets larger and larger. As an example, consider the following measure of the sample variance of a random variable X:

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

It can be shown that

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

where σ^2 is the true variance. It is obvious that in a small sample S^2 is biased, but as n increases indefinitely, $E(S^2)$ approaches true σ^2; hence it is asymptotically unbiased.

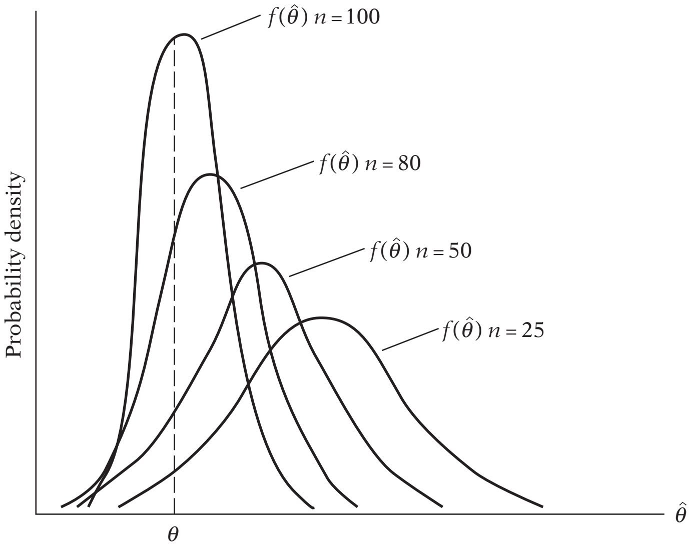
Consistency

$\hat{\theta}$ is said to be a consistent estimator if it approaches the true value θ as the sample size gets larger and larger. Figure A.11 illustrates this property.

In this figure we have the distribution of $\hat{\theta}$ based on sample sizes of 25, 50, 80, and 100. As the figure shows, $\hat{\theta}$ based on $n = 25$ is biased since its sampling distribution is not centered on the true θ. But as n increases, the distribution of $\hat{\theta}$ not only tends to be more closely centered on θ (i.e., $\hat{\theta}$ becomes less biased) but its variance also becomes smaller. If in the limit (i.e., when n increases indefinitely) the distribution of $\hat{\theta}$ collapses to the single point θ, that is, if the distribution of $\hat{\theta}$ has zero spread, or variance, we say that $\hat{\theta}$ is a **consistent estimator** of θ.

FIGURE A.11

The distribution of $\hat{\theta}$ as sample size increases.



More formally, an estimator $\hat{\theta}$ is said to be a consistent estimator of θ if the probability that the absolute value of the difference between $\hat{\theta}$ and θ is less than δ (an arbitrarily small positive quantity) approaches unity. Symbolically,

$$\lim_{n\to\infty} P\{|\hat{\theta}-\theta| < \delta\} = 1 \qquad \delta > 0$$

where P stands for probability. This is often expressed as

$$\operatorname*{plim}_{n\to\infty} \hat{\theta} = \theta$$

where plim means probability limit.

Note that the properties of unbiasedness and consistency are conceptually very different. The property of unbiasedness can hold for any sample size, whereas consistency is strictly a large-sample property.

A *sufficient condition* for consistency is that the bias and variance both tend to zero as the sample size increases indefinitely.[7] Alternatively, a sufficient condition for consistency is that the MSE($\hat{\theta}$) tends to zero as n increases indefinitely. (For MSE[$\hat{\theta}$], see the discussion presented previously.)

EXAMPLE 26

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Show that the sample mean $\bar{X}$ is a consistent estimator of μ.

From elementary statistics it is known that $E(\bar{X}) = \mu$ and $\operatorname{var}(\bar{X}) = \sigma^2/n$. Since $E(\bar{X}) = \mu$ regardless of the sample size, it is unbiased. Moreover, as n increases indefinitely, $\operatorname{var}(\bar{X})$ tends toward zero. Hence, $\bar{X}$ is a consistent estimator of μ.

The following rules about probability limits are noteworthy.

1. *Invariance (Slutsky property).* If $\hat{\theta}$ is a consistent estimator of θ and if $h(\hat{\theta})$ is any continuous function of $\hat{\theta}$, then

$$\operatorname*{plim}_{n\to\infty} h(\hat{\theta}) = h(\theta)$$

What this means is that if $\hat{\theta}$ is a consistent estimator of θ, then $1/\hat{\theta}$ is also a consistent estimator of $1/\theta$ and that $\log(\hat{\theta})$ is also a consistent estimator of $\log(\theta)$. Note that this property does not hold true of the expectation operator E; that is, if $\hat{\theta}$ is an unbiased estimator of θ (that is, $E[\hat{\theta}] = \theta$), it is *not true* that $1/\hat{\theta}$ is an unbiased estimator of $1/\theta$; that is, $E(1/\hat{\theta}) \neq 1/E(\hat{\theta}) \neq 1/\theta$.

2. If b is a constant, then

$$\operatorname*{plim}_{n\to\infty} b = b$$

That is, the probability limit of a constant is the same constant.

3. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are consistent estimators, then

$$\operatorname{plim}(\hat{\theta}_1 + \hat{\theta}_2) = \operatorname{plim}\hat{\theta}_1 + \operatorname{plim}\hat{\theta}_2$$

$$\operatorname{plim}(\hat{\theta}_1\hat{\theta}_2) = \operatorname{plim}\hat{\theta}_1 \operatorname{plim}\hat{\theta}_2$$

$$\operatorname{plim}\left(\frac{\hat{\theta}_1}{\hat{\theta}_2}\right) = \frac{\operatorname{plim}\hat{\theta}_1}{\operatorname{plim}\hat{\theta}_2}$$

[7] More technically, $\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$ and $\lim_{n\to\infty} \operatorname{var}(\hat{\theta}_n) = 0$.

The last two properties, in general, do not hold true of the expectation operator E. Thus, $E(\hat{\theta}_1/\hat{\theta}_2) \neq E(\hat{\theta}_1)/E(\hat{\theta}_2)$. Similarly, $E(\hat{\theta}_1\hat{\theta}_2) \neq E(\hat{\theta}_1)E(\hat{\theta}_2)$. If, however, $\hat{\theta}_1$ and $\hat{\theta}_2$ are independently distributed, $E(\hat{\theta}_1\hat{\theta}_2) = E(\hat{\theta}_1)E(\hat{\theta}_2)$, as noted previously.

Asymptotic Efficiency

Let $\hat{\theta}$ be an estimator of θ. The variance of the asymptotic distribution of $\hat{\theta}$ is called the **asymptotic variance** of $\hat{\theta}$. If $\hat{\theta}$ is consistent and its asymptotic variance is smaller than the asymptotic variance of all other consistent estimators of θ, $\hat{\theta}$ is called **asymptotically efficient.**

Asymptotic Normality

An estimator $\hat{\theta}$ is said to be asymptotically normally distributed if its sampling distribution tends to approach the normal distribution as the sample size n increases indefinitely. For example, statistical theory shows that if $X_1, X_2, \ldots, X_n$ are independent normally distributed variables with the same mean μ and the same variance σ^2, the sample mean $\bar{X}$ is also normally distributed with mean μ and variance σ^2/n in small as well as large samples. But if the X_i are independent with mean μ and variance σ^2 but are not necessarily from the normal distribution, then the sample mean $\bar{X}$ is asymptotically normally distributed with mean μ and variance σ^2/n; that is, as the sample size n increases indefinitely, the sample mean tends to be normally distributed with mean μ and variance σ^2/n. That is in fact the central limit theorem discussed previously.

A.8 Statistical Inference: Hypothesis Testing

Estimation and hypothesis testing constitute the twin branches of classical statistical inference. Having examined the problem of estimation, we briefly look at the problem of testing statistical hypotheses.

The problem of hypothesis testing may be stated as follows. Assume that we have an rv X with a known PDF $f(x;\theta)$, where θ is the parameter of the distribution. Having obtained a random sample of size *n,* we obtain the point estimator $\hat{\theta}$. Since the true θ is rarely known, we raise the question: Is the estimator $\hat{\theta}$ "compatible" with some hypothesized value of θ, say, $\theta = \theta^*$, where θ^* is a specific numerical value of θ? In other words, could our sample have come from the PDF $f(x;\theta) = \theta^*$? In the language of hypothesis testing $\theta = \theta^*$ is called the **null** (or maintained) **hypothesis** and is generally denoted by H_0. The null hypothesis is tested against an **alternative hypothesis,** denoted by H_1, which, for example, may state that $\theta \neq \theta^*$. (*Note:* In some textbooks, H_0 and H_1 are designated by H_1 and H_2, respectively.)

The null hypothesis and the alternative hypothesis can be **simple** or **composite.** A hypothesis is called *simple* if it specifies the value(s) of the parameter(s) of the distribution; otherwise it is called a *composite* hypothesis. Thus, if $X \sim N(\mu, \sigma^2)$ and we state that

$$H_0\colon \mu = 15 \qquad \text{and} \qquad \sigma = 2$$

it is a simple hypothesis, whereas

$$H_0\colon \mu = 15 \qquad \text{and} \qquad \sigma > 2$$

is a composite hypothesis because here the value of σ is not specified.

To test the null hypothesis (i.e., to test its validity), we use the sample information to obtain what is known as the **test statistic.** Very often this test statistic turns out to be the point estimator of the unknown parameter. Then we try to find out the *sampling,* or