and the sample variance, $V(\bar{x}) = s^2$, appear as reasonable estimates of the population mean and variance respectively. The random variables $\bar{\mathbf{x}}$ and $\mathbf{s}^2$ may be used as *estimators* of the population mean $\mu$ and population variance $\sigma^2$ respectively.

In the general sampling situation, suppose you are sampling from a population induced by a random variable whose distribution is completely unknown, or concerning which you have only partial information, not sufficient to specify it completely. Let $\theta$ represent an unknown parameter of the distribution. This might be the population mean, the population standard deviation, or some other number which, if known, would either specify the distribution completely, or provide further information about it. An *estimate*, $\hat{\theta}$, of $\theta$, is a function of the sample values, or of the sample point $(x_1, x_2, \cdots, x_n)$,

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

which you are willing to use in guessing the unknown value of $\theta$. The corresponding *estimator* of $\theta$ is the same function, evaluated at the sample random variables:

$$\hat{\boldsymbol{\theta}} = \hat{\theta}(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n).$$

The estimator is itself a random variable.

### 47. Unbiased Estimators.

▶ ***Definition.*** *The estimator* $\hat{\boldsymbol{\theta}}$ *is* unbiased *if* $E(\hat{\boldsymbol{\theta}}) = \theta$.

For example, the sample mean is an unbiased estimator of the population mean, since $E(\bar{\mathbf{x}}) = \mu$ (page 121). *The sample variance is not an unbiased estimator of the population variance.*

We have

$$E(\mathbf{s}^2) = E\left[\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i^2\right) - \bar{\mathbf{x}}^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^n E(\mathbf{x}_i^2) - E(\bar{\mathbf{x}}^2). \qquad \text{(Corollary G, page 86)}$$

Now for any random variable $\mathbf{y}$, we have

$$E(\mathbf{y}^2) = V(\mathbf{y}) + [E(\mathbf{y})]^2. \qquad \text{(page 88)}$$

In particular, for each $i$,

$$E(\mathbf{x}_i^2) = V(\mathbf{x}_i) + [E(\mathbf{x}_i)]^2 = V(\mathbf{x}) + [E(\mathbf{x})]^2 = \sigma^2 + \mu^2,$$

since each sample random variable has the same distribution as the random variable $\mathbf{x}$ which induced the population. Also

$$E(\bar{\mathbf{x}}^2) = V(\bar{\mathbf{x}}) + [E(\bar{\mathbf{x}})]^2 = \frac{\sigma^2}{n} + \mu^2. \qquad \text{(page 122)}$$



Therefore

$$E(\mathbf{s}^2) = \frac{1}{n}\sum_{i=1}^n (\sigma^2 + \mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right)$$

$$= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n}\sigma^2,$$

so that $E(\mathbf{s}^2) \neq \sigma^2$, and the sample variance is not an unbiased estimator of the population variance $\sigma^2$. On the other hand,

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

so that the statistic

$$\frac{n}{n-1}\mathbf{s}^2 = \frac{1}{n-1}\sum_{i=1}^n [\mathbf{x}_i - \bar{\mathbf{x}}]^2$$

is an unbiased estimator of the population variance.

It would be natural to expect some divergence of opinion as to the importance of such a concept as "unbiasedness." We list here some of the desirable properties an estimator possesses by virtue of being unbiased.

1. To state that an estimator is unbiased is to state that there is a measure of central tendency, the mean, of the distribution of the estimator, which is equal to the population parameter. This is simply the definition of unbiasedness. An equally appealing property, however, from this point of view, might, for example, be that the *median* of the estimator be equal to the population parameter.

2. For many unbiased estimators one can conclude, by applying the law of large numbers, that when the sample size is large the estimator is likely to be near the population parameter. However, this is the property of consistency, discussed below; and the argument here is not primarily in favor of unbiasedness, but in favor of consistency. For example, the unbiased estimator, $\frac{n}{n-1}\mathbf{s}^2$, of the population variance has this property; but so also does the sample variance $\mathbf{s}^2$ itself.

3. An important advantage from the point of view of the development of the theory of statistical inference is that in many respects unbiased estimators are simpler to deal with. One can, for example, at least when sampling from populations of a certain wide class, exhibit a lower bound on the variance of *unbiased* estimators, and show that this minimum is actually attained by maximum likelihood estimators (to be defined below). In effect, the class of possible estimators of a given parameter is so large it is impossible to say much about it. A restricting principle is required before one can penetrate deeply in a study of the behavior of estimators. To insist on unbiasedness is one way of restricting the class of estimators so as to render it more manageable.

by Theorem K and Corollary N, pages 88, 89. Again, since each $\mathbf{x}_i$ has the same distribution as $\mathbf{x}$, we have $V(\mathbf{x}_i) = V(\mathbf{x})$ for each $i$, so that

$$V(\bar{\mathbf{x}}) = \frac{1}{n^2}\sum_{i=1}^{n} V(\mathbf{x}) = \frac{1}{n^2}\cdot nV(\mathbf{x}) = \frac{1}{n}V(\mathbf{x}).$$

We shall often use the symbol $\sigma$ or $\sigma_x$ for the standard deviation of $\mathbf{x}$, so that $V(\mathbf{x}) = \sigma^2$ or $\sigma_x{}^2$; then

▶ $$V(\bar{\mathbf{x}}) = \frac{\sigma^2}{n}, \text{ or } \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}.$$

We recall that the variance of a random variable in some sense measures the compactness or the dispersion of the probability distribution of the random variable about its mean or expectation: a large variance means at least moderate probabilities associated with values distant from the mean, while a small variance means a high probability that the random variable will assume a value near its mean. We observe that for large $n$ the variance of $\bar{\mathbf{x}}$ is very small, and accordingly suspect that if $n$ is large the probability that $\bar{\mathbf{x}}$ will assume a value near its mean will be high. Essentially, this is the content of the law of large numbers, though we shall state it first in more general form. A theorem which gives us Chebyshev's Inequality will be used in the proof; this inequality is also useful for other purposes.

**44. Chebyshev's Inequality.** Chebyshev's Inequality applies to any random variable at all, provided it has an expectation and a variance.

▶ ***Theorem A.*** *Let* $\mathbf{y}$ *be a random variable with expectation* $E(\mathbf{y}) = \mu$ *and variance* $V(\mathbf{y}) = \sigma^2$. *Then if* $\epsilon$ *is any positive number,* $Pr\{|\mathbf{y} - \mu| \geq \epsilon\} \leq \sigma^2/\epsilon^2$.

$Pr\{|\mathbf{y} - \mu| \geq \epsilon\}$ may be read: "the probability that $\mathbf{y}$ will differ from $\mu$ by at least $\epsilon$."

*Proof for a continuous random variable* $\mathbf{y}$: By the definition of variance, page 51, we have

$$\sigma^2 = V(\mathbf{y}) = \int_{-\infty}^{\infty}(y - \mu)^2 g(y)\,dy,$$

where $\mu = E(\mathbf{y})$ and where $g(y)$ is the probability density function of $\mathbf{y}$. Further,

$$\sigma^2 = \int_{-\infty}^{\infty}(y - \mu)^2 g(y)\,dy$$

$$= \int_{|y-\mu|\geq\epsilon}(y - \mu)^2 g(y)\,dy + \int_{|y-\mu|<\epsilon}(y - \mu)^2 g(y)\,dy,$$



where $\int_{|y-\mu|\geq\epsilon}$ denotes the integral over the range of values of $y$ for which $|y - \mu| \geq \epsilon$, and $\int_{|y-\mu|<\epsilon}$ denotes the integral over the remainder of the range of $y$. (The last equation could also have been written

$$\int_{-\infty}^{\infty}(y - \mu)^2 g(y)\,dy$$

$$= \int_{\mu+\epsilon}^{\infty}(y - \mu)^2 g(y)\,dy + \int_{-\infty}^{\mu-\epsilon}(y - \mu)^2 g(y)\,dy + \int_{\mu-\epsilon}^{\mu+\epsilon}(y - \mu)^2 g(y)\,dy.)$$

Since the integrand is non-negative, the last integral is non-negative, and we have

$$\sigma^2 \geq \int_{|y-\mu|\geq\epsilon}(y - \mu)^2 g(y)\,dy.$$

In the integral in the right member, the integrand is always at least as large as $\epsilon^2 g(y)$ over the indicated range of integration, so that

$$\sigma^2 \geq \epsilon^2 \int_{|y-\mu|\geq\epsilon} g(y)\,dy.$$

But the integral of the probability density function of a random variable over a range of values is just the probability that the random variable will assume a value in that range (page 40). Therefore

$$\int_{|y-\mu|\geq\epsilon} g(y)\,dy = Pr\{|\mathbf{y} - \mu| \geq \epsilon\},$$

and the last inequality gives us

$$\sigma^2 \geq \epsilon^2 Pr\{|\mathbf{y} - \mu| \geq \epsilon\},$$

or

$$Pr\{|\mathbf{y} - \mu| \geq \epsilon\} \leq \sigma^2/\epsilon^2.$$

EXAMPLE 1. In the example on page 97 you have a large lot of items and wish, by random sampling, to estimate the fraction which are defective. We can use Chebyshev's inequality to find a number $N$ with the property that if you take a sample of size $N$ or larger, the probability will be at least, say, 0.95 that the average number of defectives in your sample will differ from the true proportion in the lot by less than, say, 0.1. If we set up the population as we did earlier (pp. 97 ff.), the random variable $\mathbf{x}$, "number of defectives at an elementary event," induces on $R$ a probability space in which the elementary event 1 has probability $p$, the probability of a defective, and the elementary event 0 has probability $q = 1 - p$. If $\mathbf{x}_i$ is

▶ ***Theorem K.*** *If $a$ is a real number and* $\mathbf{x}$ *a random variable with variance* $V(\mathbf{x})$, *then* $V(a\mathbf{x}) = a^2V(\mathbf{x})$.

We leave the proof as an exercise.

We have in the following theorem an alternative formula for the variance which is often more convenient than that given in the definition.

▶ ***Theorem L.***

$$V(\mathbf{x}) = E(\mathbf{x}^2) - [E(\mathbf{x})]^2.$$

*Proof.* We have

$$\begin{aligned} V(\mathbf{x}) &= E[\mathbf{x} - E(\mathbf{x})]^2 = E(\mathbf{x}^2 - 2\mathbf{x}E(\mathbf{x}) + [E(\mathbf{x})]^2) \\ &= E(\mathbf{x}^2) - E[2\mathbf{x}E(\mathbf{x})] + E([E(\mathbf{x})]^2) \quad \text{by Corollary G} \\ &= E(\mathbf{x}^2) - 2E(\mathbf{x})E(\mathbf{x}) + [E(\mathbf{x})]^2 \quad \text{by Theorems C and D} \\ &\qquad \text{(note that } E(\mathbf{x}) \text{ is a constant);} \end{aligned}$$

$$V(\mathbf{x}) = E(\mathbf{x}^2) - [E(\mathbf{x})]^2.$$

The formula of Theorem L is often useful in furnishing an expression for $E(\mathbf{x}^2)$:

▶
$$E(\mathbf{x}^2) = V(\mathbf{x}) + [E(\mathbf{x})]^2.$$

▶ ***Theorem M.*** *If* $\mathbf{x},\mathbf{y}$ *are* independent *random variables with variances* $V(\mathbf{x})$ *and* $V(\mathbf{y})$, *then*

$$V(\mathbf{x} + \mathbf{y}) = V(\mathbf{x}) + V(\mathbf{y}).$$

*Proof.*

$$\begin{aligned} V(\mathbf{x} + \mathbf{y}) &= E[(\mathbf{x} + \mathbf{y}) - E(\mathbf{x} + \mathbf{y})]^2 \quad \text{by definition;} \\ &= E([\mathbf{x} - E(\mathbf{x})] + [\mathbf{y} - E(\mathbf{y})])^2 \\ &= E([\mathbf{x} - E(\mathbf{x})]^2 + 2[\mathbf{x} - E(\mathbf{x})][\mathbf{y} - E(\mathbf{y})] \\ &\qquad + [\mathbf{y} - E(\mathbf{y})]^2) \\ &= E[\mathbf{x} - E(\mathbf{x})]^2 + 2E([\mathbf{x} - E(\mathbf{x})][\mathbf{y} - E(\mathbf{y})]) \\ &\qquad + E[\mathbf{y} - E(\mathbf{y})]^2 \end{aligned}$$

by Corollary G.

Now $\mathbf{x} - E(\mathbf{x})$ is a function of the random variable $\mathbf{x}$, namely, $\mathbf{x}$ minus a constant, and similarly $\mathbf{y} - E(\mathbf{y})$ is a function of the random variable $\mathbf{y}$. By the remark on page 80, $\mathbf{x} - E(\mathbf{x})$ and $\mathbf{y} - E(\mathbf{y})$ are independent, since $\mathbf{x}$ and $\mathbf{y}$ are independent. By Theorem H, then,

$$\begin{aligned} E([\mathbf{x} - E(\mathbf{x})][\mathbf{y} - E(\mathbf{y})]) &= E[\mathbf{x} - E(\mathbf{x})]E[\mathbf{y} - E(\mathbf{y})] \\ &= [E(\mathbf{x}) - E(\mathbf{x})][E(\mathbf{y}) - E(\mathbf{y})] \quad \text{by Corollary F,} \\ &= 0. \end{aligned}$$



Therefore

$$\begin{aligned} V(\mathbf{x} + \mathbf{y}) &= E[\mathbf{x} - E(\mathbf{x})]^2 + E[\mathbf{y} - E(\mathbf{y})]^2 \\ &= V(\mathbf{x}) + V(\mathbf{y}), \quad \text{by definition.} \end{aligned}$$

Application of the principle of mathematical induction to Theorem M yields the following corollary:

▶ ***Corollary N.*** *The variance of the sum of a finite number of* independent *random variables is the sum of their variances.*

PROBLEMS

**1.** Let $\mathbf{x}$ be a discrete random variable assuming the value 1 with probability $p$ and the value 0 with probability $q = 1 - p$. Find $E(\mathbf{x})$, $V(\mathbf{x})$.

**2.** A cube has one spot on each of four sides, two spots on each of the other two sides. Find the mean (expectation) and variance of (a) the number of spots that will show on top when it is tossed to the ground; (b) the total number of spots showing when 6 such cubes are tossed.

**3.** If $\mathbf{x}$ is the random variable described in Problem 1, what are the values assumed, and with what probabilities, by the random variables $2\mathbf{x}$, $2\mathbf{x} - 1$, $\mathbf{x}^2$? Find the expectation and variance of each.

**4.** Let $\mathbf{x},\mathbf{y},\mathbf{z}$ be independent random variables with $E(\mathbf{x}) = E(\mathbf{y}) = 2$, $E(\mathbf{z}) = -3$, $V(\mathbf{x}) = 1$, $V(\mathbf{y}) = V(\mathbf{z}) = 2$. Find (a) $E(\mathbf{x} + \mathbf{y} + \mathbf{z})$; (b) $E[\mathbf{x}(\mathbf{y} + \mathbf{z})]$; (c) $V(3\mathbf{y} + \mathbf{z})$. (d) Which, if any, of your answers depend essentially on the independence of the random variables?

**5.** If $n$ is a positive integer, and if each of the $n$ random variables $\mathbf{x}_1$, $\mathbf{x}_2, \cdots, \mathbf{x}_n$ has the same distribution (same probability function) as the random variable $\mathbf{x}$ in Problem 1, find (a) $E(\sum_{i=1}^n \mathbf{x}_i)$; (b) $E\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\right)$.

**6.** If the random variables $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ in Problem 5 are independent, find (a) $V(\sum_{i=1}^n \mathbf{x}_i)$; (b) $V\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\right)$.

**7.** If $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ is any combined random variable, express $E(\sum_{i=1}^n \mathbf{x}_i)$ and $E\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\right)$ in terms of $E(\mathbf{x}_1)$, $E(\mathbf{x}_2)$, $\cdots$, $E(\mathbf{x}_n)$ (assuming they exist).

**8.** Express $V(\sum_{i=1}^n \mathbf{x}_i)$ and $V\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\right)$ in terms of $V(\mathbf{x}_1)$, $V(\mathbf{x}_2)$, $\cdots$, $V(\mathbf{x}_n)$ (assuming they exist) if $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ are independent.

**9.** What are the expectation and variance of the sum of the numbers appearing on two dice? on $n$ dice?

**10.** Prove Theorem C.

**11.** Prove Theorem D for a continuous random variable $\mathbf{x}$.

▶ ***Theorem C.*** *If a is a constant, then*

$$E(a) = a.$$

You may supply the proof.

▶ ***Theorem D.*** *If a is a real number (constant) and* **x** *is a random variable with expectation* $E(\mathbf{x})$, *then* $E(a\mathbf{x}) = aE(\mathbf{x})$.

*Proof for discrete* **x**. On setting $\varphi(x) \equiv ax$ in Theorem A, we find that

$$E(a\mathbf{x}) = \sum_x ax\,f(x) = a\sum_x x\,f(x) = aE(\mathbf{x}).$$

You may supply the proof for a continuous random variable **x**.

▶ ***Theorem E.*** *If* **x** *and* **y** *are random variables with expectations* $E(\mathbf{x})$ *and* $E(\mathbf{y})$, *then* $E(\mathbf{x} + \mathbf{y}) = E(\mathbf{x}) + E(\mathbf{y})$.

*Proof for a continuous combined random variable* $(\mathbf{x},\mathbf{y})$. By Theorem B, with $\varphi(x,y) \equiv x + y$, we have

$$\begin{aligned} E(\mathbf{x} + \mathbf{y}) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x + y)\,f(x,y)\,dx\,dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,f(x,y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\,f(x,y)\,dx\,dy \\ &= \int_{-\infty}^{\infty} x\,dx \int_{-\infty}^{\infty} f(x,y)\,dy + \int_{-\infty}^{\infty} y\,dy \int_{-\infty}^{\infty} f(x,y)\,dx \\ &= \int_{-\infty}^{\infty} x\,f(x)\,dx + \int_{-\infty}^{\infty} y\,g(y)\,dy \\ &= E(\mathbf{x}) + E(\mathbf{y}), \end{aligned}$$

since

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{and} \quad g(y) = \int_{-\infty}^{\infty} f(x,y)\,dx.$$

We leave the proof for discrete random variables to you.

▶ ***Corollary F.*** *If a is a real number (constant) and if* **x** *is a random variable with expectation* $E(\mathbf{x})$, *then*

$$E(\mathbf{x} + a) = E(\mathbf{x}) + a.$$

This theorem is easily proved directly, using Theorem A; but it is also a corollary of Theorems E and C, obtained on replacing **y** by a random variable assuming the constant value $a$ with probability 1.

▶ ***Corollary G.*** *The expectation of a finite sum of random variables is the sum of their expectations.*

This follows from Theorem E with the aid of mathematical induction.



▶ ***Theorem H.*** *If* **x**,**y** *are* independent *random variables with expectations* $E(\mathbf{x})$ *and* $E(\mathbf{y})$, *then*

$$E(\mathbf{xy}) = E(\mathbf{x})E(\mathbf{y}).$$

*Proof for continuous random variables.* By Theorem B, with $\varphi(x,y) = xy$, we have

$$E(\mathbf{xy}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,f(x,y)\,dx\,dy.$$

Since **x** and **y** are independent, $f(x,y) = f(x)g(y)$, so that

$$\begin{aligned} E(\mathbf{xy}) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,f(x)g(y)\,dx\,dy \\ &= \int_{-\infty}^{\infty} x\,f(x)\,dx \int_{-\infty}^{\infty} y\,g(y)\,dy = E(\mathbf{x})E(\mathbf{y}). \end{aligned}$$

▶ ***Corollary I.*** *The expectation of the product of a finite number of independent random variables is the product of their expectations.*

Again, mathematical induction yields this corollary to Theorem H.

We now recall the definition of variance. If **x** is a discrete random variable with probability function $f(x) = Pr\{\mathbf{x} = x\}$, then

$$V(\mathbf{x}) = \sum_x [x - E(\mathbf{x})]^2 f(x).$$

If **x** is a continuous random variable with probability density function $f(x)$, then

$$V(\mathbf{x}) = \int_{-\infty}^{\infty} [x - E(\mathbf{x})]^2 f(x)\,dx.$$

In either case, we see from Theorem A that

▶ $$V(\mathbf{x}) = E[\mathbf{x} - E(\mathbf{x})]^2.$$

▶ ***Theorem J.*** *If a is a real number, and* **x** *a random variable with variance* $V(\mathbf{x})$, *then* $V(a) = 0$ *and* $V(\mathbf{x} + a) = V(\mathbf{x})$.

*Proof.* By Theorem C, $E(a) = a$, so that

$$V(a) = E[a - E(a)]^2 = E(a - a)^2 = 0.$$

Also

$$\begin{aligned} V(\mathbf{x} + a) &= E[\mathbf{x} + a - E(\mathbf{x} + a)]^2, \\ &= E(\mathbf{x} + a - E(\mathbf{x}) - a)^2 && \text{by Corollary F,} \\ &= E[\mathbf{x} - E(\mathbf{x})]^2 = V(\mathbf{x}) && \text{by definition.} \end{aligned}$$

AN INTRODUCTION

TO

# *Mathematical Statistics*

H. D. BRUNK, *University of Missouri*

Ginn and Company

*Boston New York Chicago Atlanta Dallas*
*Palo Alto Toronto London*