CHAPTER 3

Mathematical Expectation

Definition of Mathematical Expectation

A very important concept in probability and statistics is that of the *mathematical expectation*, *expected value*, or briefly the *expectation*, of a random variable. For a discrete random variable X having the possible values $x_1, \ldots, x_n$, the expectation of X is defined as

$$E(X) = x_1 P(X = x_1) + \cdots + x_n P(X = x_n) = \sum_{j=1}^{n} x_j P(X = x_j) \tag{1}$$

or equivalently, if $P(X = x_j) = f(x_j)$,

$$E(X) = x_1 f(x_1) + \cdots + x_n f(x_n) = \sum_{j=1}^{n} x_j f(x_j) = \sum x f(x) \tag{2}$$

where the last summation is taken over all appropriate values of x. As a special case of (2), where the probabilities are all equal, we have

$$E(X) = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{3}$$

which is called the *arithmetic mean*, or simply the *mean*, of $x_1, x_2, \ldots, x_n$.

If X takes on an infinite number of values $x_1, x_2, \ldots$, then $E(X) = \sum_{j=1}^{\infty} x_j f(x_j)$ provided that the infinite series converges absolutely.

For a continuous random variable X having density function $f(x)$, the expectation of X is defined as

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx \tag{4}$$

provided that the integral converges absolutely.

The expectation of X is very often called the *mean* of X and is denoted by μ_X, or simply μ, when the particular random variable is understood.

The mean, or expectation, of X gives a single value that acts as a representative or average of the values of X, and for this reason it is often called a *measure of central tendency.* Other measures are considered on page 83.

EXAMPLE 3.1 Suppose that a game is to be played with a single die assumed fair. In this game a player wins \$20 if a 2 turns up, \$40 if a 4 turns up; loses \$30 if a 6 turns up; while the player neither wins nor loses if any other face turns up. Find the expected sum of money to be won.

Let X be the random variable giving the amount of money won on any toss. The possible amounts won when the die turns up $1, 2, \ldots, 6$ are $x_1, x_2, \ldots, x_6$, respectively, while the probabilities of these are $f(x_1), f(x_2), \ldots, f(x_6)$. The probability function for X is displayed in Table 3-1. Therefore, the expected value or expectation is

$$E(X) = (0)\left(\frac{1}{6}\right) + (20)\left(\frac{1}{6}\right) + (0)\left(\frac{1}{6}\right) + (40)\left(\frac{1}{6}\right) + (0)\left(\frac{1}{6}\right) + (-30)\left(\frac{1}{6}\right) = 5$$

Table 3-1

x_j	0	+20	0	+40	0	−30
$f(x_j)$	1/6	1/6	1/6	1/6	1/6	1/6

It follows that the player can expect to win \$5. In a fair game, therefore, the player should be expected to pay \$5 in order to play the game.

EXAMPLE 3.2 The density function of a random variable X is given by

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

The expected value of X is then

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_0^2 x\left(\frac{1}{2}x\right)dx = \int_0^2 \frac{x^2}{2}\,dx = \frac{x^3}{6}\bigg|_0^2 = \frac{4}{3}$$

Functions of Random Variables

Let X be a discrete random variable with probability function $f(x)$. Then $Y = g(X)$ is also a discrete random variable, and the probability function of Y is

$$h(y) = P(Y = y) = \sum_{\{x|g(x)=y\}} P(X = x) = \sum_{\{x|g(x)=y\}} f(x)$$

If X takes on the values $x_1, x_2, \ldots, x_n$, and Y the values $y_1, y_2, \ldots, y_m$ $(m \le n)$, then $y_1h(y_1) + y_2h(y_2) + \cdots + y_mh(y_m) = g(x_1)f(x_1) + g(x_2)f(x_2) + \cdots + g(x_n)f(x_n)$. Therefore,

$$\begin{aligned} E[g(X)] &= g(x_1)f(x_1) + g(x_2)f(x_2) + \cdots + g(x_n)f(x_n) \\ &= \sum_{j=1}^{n} g(x_j)f(x_j) = \sum g(x)f(x) \end{aligned} \tag{5}$$

Similarly, if X is a continuous random variable having probability density $f(x)$, then it can be shown that

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx \tag{6}$$

Note that (5) and (6) do not involve, respectively, the probability function and the probability density function of $Y = g(X)$.

Generalizations are easily made to functions of two or more random variables. For example, if X and Y are two continuous random variables having joint density function $f(x, y)$, then the expectation of $g(X, Y)$ is given by

$$E[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)f(x, y)\,dx\,dy \tag{7}$$

EXAMPLE 3.3 If X is the random variable of Example 3.2,

$$E(3X^2 - 2X) = \int_{-\infty}^{\infty} (3x^2 - 2x)f(x)\,dx = \int_0^2 (3x^2 - 2x)\left(\frac{1}{2}x\right)dx = \frac{10}{3}$$

Some Theorems on Expectation

Theorem 3-1 If c is any constant, then

$$E(cX) = cE(X) \tag{8}$$

Theorem 3-2 If X and Y are any random variables, then

$$E(X + Y) = E(X) + E(Y) \tag{9}$$

Theorem 3-3 If X and Y are independent random variables, then

$$E(XY) = E(X)E(Y) \tag{10}$$

Generalizations of these theorems are easily made.

The Variance and Standard Deviation

We have already noted on page 75 that the expectation of a random variable X is often called the *mean* and is denoted by μ. Another quantity of great importance in probability and statistics is called the *variance* and is defined by

$$\mathrm{Var}(X) = E[(X - \mu)^2] \tag{11}$$

The variance is a nonnegative number. The positive square root of the variance is called the *standard deviation* and is given by

$$\sigma_X = \sqrt{\mathrm{Var}(X)} = \sqrt{E[(X - \mu)^2]} \tag{12}$$

Where no confusion can result, the standard deviation is often denoted by σ instead of σ_X, and the variance in such case is σ^2.

If X is a discrete random variable taking the values $x_1, x_2, \ldots, x_n$ and having probability function $f(x)$, then the variance is given by

$$\sigma_X^2 = E[(X - \mu)^2] = \sum_{j=1}^{n}(x_j - \mu)^2 f(x_j) = \sum (x - \mu)^2 f(x) \tag{13}$$

In the special case of (13) where the probabilities are all equal, we have

$$\sigma^2 = [(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_n - \mu)^2]/n \tag{14}$$

which is the variance for a set of n numbers $x_1, \ldots, x_n$.

If X takes on an infinite number of values $x_1, x_2, \ldots$, then $\sigma_X^2 = \sum_{j=1}^{\infty}(x_j - \mu)^2 f(x_j)$, provided that the series converges.

If X is a continuous random variable having density function $f(x)$, then the variance is given by

$$\sigma_X^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty}(x - \mu)^2 f(x)\,dx \tag{15}$$

provided that the integral converges.

The variance (or the standard deviation) is a measure of the *dispersion*, or *scatter*, of the values of the random variable about the mean μ. If the values tend to be concentrated near the mean, the variance is small; while if the values tend to be distributed far from the mean, the variance is large. The situation is indicated graphically in Fig. 3-1 for the case of two continuous distributions having the same mean μ.

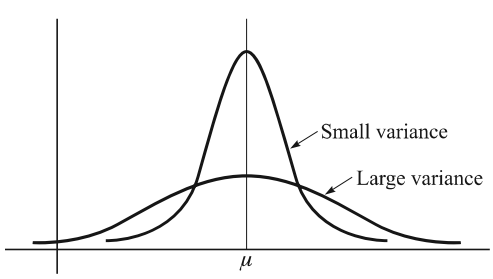

Fig. 3-1

EXAMPLE 3.4 Find the variance and standard deviation of the random variable of Example 3.2. As found in Example 3.2, the mean is $\mu = E(X) = 4/3$. Then the variance is given by

$$\sigma^2 = E\left[\left(X - \frac{4}{3}\right)^2\right] = \int_{-\infty}^{\infty}\left(x - \frac{4}{3}\right)^2 f(x)\,dx = \int_0^2\left(x - \frac{4}{3}\right)^2\left(\frac{1}{2}x\right)dx = \frac{2}{9}$$

and so the standard deviation is $\sigma = \sqrt{\frac{2}{9}} = \frac{\sqrt{2}}{3}$

Note that if X has certain *dimensions* or *units*, such as *centimeters* (cm), then the variance of X has units cm^2 while the standard deviation has the same unit as X, i.e., cm. It is for this reason that the standard deviation is often used.

Some Theorems on Variance

Theorem 3-4

$$\sigma^2 = E[(X - \mu)^2] = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2 \tag{16}$$

where $\mu = E(X)$.

Theorem 3-5 If c is any constant,

$$\text{Var}(cX) = c^2\,\text{Var}(X) \tag{17}$$

Theorem 3-6 The quantity $E[(X - a)^2]$ is a minimum when $a = \mu = E(X)$.

Theorem 3-7 If X and Y are independent random variables,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \qquad \text{or} \qquad \sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y \tag{18}$$

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) \qquad \text{or} \qquad \sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y \tag{19}$$

Generalizations of Theorem 3-7 to more than two independent variables are easily made. In words, the variance of a sum of independent variables equals the sum of their variances.

Standardized Random Variables

Let X be a random variable with mean μ and standard deviation $\sigma\,(\sigma > 0)$. Then we can define an associated *standardized random variable* given by

$$X^* = \frac{X - \mu}{\sigma} \tag{20}$$

An important property of X^* is that it has a mean of zero and a variance of 1, which accounts for the name *standardized,* i.e.,

$$E(X^*) = 0, \qquad \text{Var}(X^*) = 1 \tag{21}$$

The values of a standardized variable are sometimes called *standard scores,* and X is then said to be expressed in *standard units* (i.e., σ is taken as the unit in measuring $X - \mu$).

Standardized variables are useful for comparing different distributions.

Moments

The *rth moment of a random variable X about the mean* μ, also called the *rth central moment*, is defined as

$$\mu_r = E[(X - \mu)^r] \tag{22}$$

where $r = 0, 1, 2, \ldots$. It follows that $\mu_0 = 1$, $\mu_1 = 0$, and $\mu_2 = \sigma^2$, i.e., the second central moment or second moment about the mean is the variance. We have, assuming absolute convergence,

$$\mu_r = \sum (x - \mu)^r f(x) \qquad \text{(discrete variable)} \tag{23}$$

$$\mu_r = \int_{-\infty}^{\infty} (x - \mu)^r f(x)\, dx \qquad \text{(continuous variable)} \tag{24}$$

The *rth moment of X about the origin*, also called the *rth raw moment*, is defined as

$$\mu'_r = E(X^r) \tag{25}$$

where $r = 0, 1, 2, \ldots$, and in this case there are formulas analogous to (23) and (24) in which $\mu = 0$.

The relationship between these moments is given by

$$\mu_r = \mu'_r - \binom{r}{1}\mu'_{r-1}\mu + \cdots + (-1)^j\binom{r}{j}\mu'_{r-j}\mu^j + \cdots + (-1)^r\mu'_0\mu^r \tag{26}$$

As special cases we have, using $\mu'_1 = \mu$ and $\mu'_0 = 1$,

$$\begin{aligned} \mu_2 &= \mu'_2 - \mu^2 \\ \mu_3 &= \mu'_3 - 3\mu'_2\mu + 2\mu^3 \\ \mu_4 &= \mu'_4 - 4\mu'_3\mu + 6\mu'_2\mu^2 - 3\mu^4 \end{aligned} \tag{27}$$

Moment Generating Functions

The *moment generating function* of X is defined by

$$M_X(t) = E(e^{tX}) \tag{28}$$

that is, assuming convergence,

$$M_X(t) = \sum e^{tx} f(x) \qquad \text{(discrete variable)} \tag{29}$$

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx \qquad \text{(continuous variable)} \tag{30}$$

We can show that the Taylor series expansion is [Problem 3.15(a)]

$$M_X(t) = 1 + \mu t + \mu'_2\frac{t^2}{2!} + \cdots + \mu'_r\frac{t^r}{r!} + \cdots \tag{31}$$

Since the coefficients in this expansion enable us to find the moments, the reason for the name *moment generating function* is apparent. From the expansion we can show that [Problem 3.15(b)]

$$\mu'_r = \frac{d^r}{dt^r}M_X(t)\bigg|_{t=0} \tag{32}$$

i.e., μ'_r is the rth derivative of $M_X(t)$ evaluated at $t = 0$. Where no confusion can result, we often write $M(t)$ instead of $M_X(t)$.

Some Theorems on Moment Generating Functions

Theorem 3-8 If $M_X(t)$ is the moment generating function of the random variable X and a and b ($b \neq 0$) are constants, then the moment generating function of $(X + a)/b$ is

$$M_{(X+a)/b}(t) = e^{at/b}M_X\left(\frac{t}{b}\right) \tag{33}$$

Theorem 3-9 If X and Y are independent random variables having moment generating functions $M_X(t)$ and $M_Y(t)$, respectively, then

$$M_{X+Y}(t) = M_X(t)\,M_Y(t) \tag{34}$$

Generalizations of Theorem 3-9 to more than two independent random variables are easily made. In words, the moment generating function of a sum of independent random variables is equal to the product of their moment generating functions.

Theorem 3-10 **(Uniqueness Theorem)** Suppose that X and Y are random variables having moment generating functions $M_X(t)$ and $M_Y(t)$, respectively. Then X and Y have the same probability distribution if and only if $M_X(t) = M_Y(t)$ identically.

Characteristic Functions

If we let $t = i\omega$, where i is the imaginary unit, in the moment generating function we obtain an important function called the *characteristic function.* We denote this by

$$\phi_X(\omega) = M_X(i\omega) = E(e^{i\omega X}) \tag{35}$$

It follows that

$$\phi_X(\omega) = \sum e^{i\omega x} f(x) \qquad \text{(discrete variable)} \tag{36}$$

$$\phi_X(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} f(x)\,dx \qquad \text{(continuous variable)} \tag{37}$$

Since $|e^{i\omega x}| = 1$, the series and the integral always converge absolutely.

The corresponding results (31) and (32) become

$$\phi_X(\omega) = 1 + i\mu\omega - \mu_2'\frac{\omega^2}{2!} + \cdots + i^r\mu_r'\frac{\omega^r}{r!} + \cdots \tag{38}$$

where

$$\mu_r' = (-1)^r i^r \frac{d^r}{d\omega^r}\phi_X(\omega)\bigg|_{\omega=0} \tag{39}$$

When no confusion can result, we often write $\phi(\omega)$ instead of $\phi_X(\omega)$.

Theorems for characteristic functions corresponding to Theorems 3-8, 3-9, and 3-10 are as follows.

Theorem 3-11 If $\phi_X(\omega)$ is the characteristic function of the random variable X and a and b ($b \neq 0$) are constants, then the characteristic function of $(X + a)/b$ is

$$\phi_{(X+a)/b}(\omega) = e^{ai\omega/b}\phi_X\left(\frac{\omega}{b}\right) \tag{40}$$

Theorem 3-12 If X and Y are independent random variables having characteristic functions $\phi_X(\omega)$ and $\phi_Y(\omega)$, respectively, then

$$\phi_{X+Y}(\omega) = \phi_X(\omega)\,\phi_Y(\omega) \tag{41}$$

More generally, the characteristic function of a sum of independent random variables is equal to the product of their characteristic functions.

Theorem 3-13 **(Uniqueness Theorem)** Suppose that X and Y are random variables having characteristic functions $\phi_X(\omega)$ and $\phi_Y(\omega)$, respectively. Then X and Y have the same probability distribution if and only if $\phi_X(\omega) = \phi_Y(\omega)$ identically.

An important reason for introducing the characteristic function is that (37) represents the *Fourier transform* of the density function $f(x)$. From the theory of Fourier transforms, we can easily determine the density function from the characteristic function. In fact,

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega x}\phi_X(\omega)\,d\omega \tag{42}$$

which is often called an *inversion formula*, or *inverse Fourier transform.* In a similar manner we can show in the discrete case that the probability function $f(x)$ can be obtained from (36) by use of *Fourier series*, which is the analog of the Fourier integral for the discrete case. See Problem 3.39.

Another reason for using the characteristic function is that it always exists whereas the moment generating function may not exist.

Variance for Joint Distributions. Covariance

The results given above for one variable can be extended to two or more variables. For example, if X and Y are two continuous random variables having joint density function $f(x, y)$, the means, or expectations, of X and Y are

$$\mu_X = E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x, y)\,dx\,dy, \qquad \mu_Y = E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x, y)\,dx\,dy \tag{43}$$

and the variances are

$$\begin{aligned} \sigma_X^2 &= E[(X - \mu_X)^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)^2 f(x, y)\,dx\,dy \\ \sigma_Y^2 &= E[(Y - \mu_Y)^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - \mu_Y)^2 f(x, y)\,dx\,dy \end{aligned} \tag{44}$$

Note that the marginal density functions of X and Y are not directly involved in (43) and (44).

Another quantity that arises in the case of two variables X and Y is the *covariance* defined by

$$\sigma_{XY} = \mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \tag{45}$$

In terms of the joint density function $f(x, y)$, we have

$$\sigma_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y)\,dx\,dy \tag{46}$$

Similar remarks can be made for two discrete random variables. In such cases (43) and (46) are replaced by

$$\mu_X = \sum_x\sum_y xf(x, y) \qquad \mu_Y = \sum_x\sum_y yf(x, y) \tag{47}$$

$$\sigma_{XY} = \sum_x\sum_y (x - \mu_X)(y - \mu_Y) f(x, y) \tag{48}$$

where the sums are taken over all the discrete values of X and Y.

The following are some important theorems on covariance.

Theorem 3-14
$$\sigma_{XY} = E(XY) - E(X)E(Y) = E(XY) - \mu_X\mu_Y \tag{49}$$

Theorem 3-15 If X and Y are independent random variables, then

$$\sigma_{XY} = \mathrm{Cov}(X, Y) = 0 \tag{50}$$

Theorem 3-16
$$\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\,\mathrm{Cov}(X, Y) \tag{51}$$

or
$$\sigma_{X\pm Y}^2 = \sigma_X^2 + \sigma_Y^2 \pm 2\sigma_{XY} \tag{52}$$

Theorem 3-17
$$|\sigma_{XY}| \le \sigma_X\sigma_Y \tag{53}$$

The converse of Theorem 3-15 is not necessarily true. If X and Y are independent, Theorem 3-16 reduces to Theorem 3-7.

Correlation Coefficient

If X and Y are independent, then $\text{Cov}(X, Y) = \sigma_{XY} = 0$. On the other hand, if X and Y are completely dependent, for example, when $X = Y$, then $\text{Cov}(X, Y) = \sigma_{XY} = \sigma_X\sigma_Y$. From this we are led to a *measure of the dependence* of the variables X and Y given by

$$\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \tag{54}$$

We call ρ the *correlation coefficient*, or *coefficient of correlation.* From Theorem 3-17 we see that $-1 \leq \rho \leq 1$. In the case where $\rho = 0$ (i.e., the covariance is zero), we call the variables X and Y *uncorrelated.* In such cases, however, the variables may or may not be independent. Further discussion of correlation cases will be given in Chapter 8.

Conditional Expectation, Variance, and Moments

If X and Y have joint density function $f(x, y)$, then as we have seen in Chapter 2, the conditional density function of Y given X is $f(y|x) = f(x, y)/f_1(x)$ where $f_1(x)$ is the marginal density function of X. We can define the *conditional expectation*, or *conditional mean*, of Y given X by

$$E(Y|X = x) = \int_{-\infty}^{\infty} yf(y|x)\,dy \tag{55}$$

where "$X = x$" is to be interpreted as $x < X \leq x + dx$ in the continuous case. Theorems 3-1 and 3-2 also hold for conditional expectation.

We note the following properties:

1. $E(Y|X = x) = E(Y)$ when X and Y are independent.
2. $E(Y) = \displaystyle\int_{-\infty}^{\infty} E(Y|X = x)f_1(x)\,dx.$

It is often convenient to calculate expectations by use of Property 2, rather than directly.

EXAMPLE 3.5 The average travel time to a distant city is c hours by car or b hours by bus. A woman cannot decide whether to drive or take the bus, so she tosses a coin. What is her expected travel time?

Here we are dealing with the joint distribution of the outcome of the toss, X, and the travel time, Y, where $Y = Y_{\text{car}}$ if $X = 0$ and $Y = Y_{\text{bus}}$ if $X = 1$. Presumably, both Y_{car} and Y_{bus} are independent of X, so that by Property 1 above

$$E(Y|X = 0) = E(Y_{\text{car}}|X = 0) = E(Y_{\text{car}}) = c$$

and

$$E(Y|X = 1) = E(Y_{\text{bus}}|X = 1) = E(Y_{\text{bus}}) = b$$

Then Property 2 (with the integral replaced by a sum) gives, for a fair coin,

$$E(Y) = E(Y|X = 0)P(X = 0) + E(Y|X = 1)P(X = 1) = \frac{c + b}{2}$$

In a similar manner we can define the *conditional variance* of Y given X as

$$E[(Y - \mu_2)^2|X = x] = \int_{-\infty}^{\infty} (y - \mu_2)^2 f(y|x)\,dy \tag{56}$$

where $\mu_2 = E(Y|X = x)$. Also we can define the rth *conditional moment* of Y about any value a given X as

$$E[(Y - a)^r|X = x] = \int_{-\infty}^{\infty} (y - a)^r f(y|x)\,dy \tag{57}$$

The usual theorems for variance and moments extend to conditional variance and moments.

Chebyshev's Inequality

An important theorem in probability and statistics that reveals a general property of discrete or continuous random variables having finite mean and variance is known under the name of *Chebyshev's inequality.*

Theorem 3-18 **(Chebyshev's Inequality)** Suppose that X is a random variable (discrete or continuous) having mean μ and variance σ^2, which are finite. Then if ϵ is any positive number,

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \tag{58}$$

or, with $\epsilon = k\sigma$,

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{59}$$

EXAMPLE 3.6 Letting $k = 2$ in Chebyshev's inequality (59), we see that

$$P(|X - \mu| \geq 2\sigma) \leq 0.25 \qquad \text{or} \qquad P(|X - \mu| < 2\sigma) \geq 0.75$$

In words, the probability of X differing from its mean by more than 2 standard deviations is less than or equal to 0.25; equivalently, the probability that X will lie within 2 standard deviations of its mean is greater than or equal to 0.75. This is quite remarkable in view of the fact that we have not even specified the probability distribution of X.

Law of Large Numbers

The following theorem, called the *law of large numbers*, is an interesting consequence of Chebyshev's inequality.

Theorem 3-19 **(Law of Large Numbers):** Let $X_1, X_2, \ldots, X_n$ be mutually independent random variables (discrete or continuous), each having finite mean μ and variance σ^2. Then if $S_n = X_1 + X_2 + \cdots + X_n (n = 1, 2, \ldots)$,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0 \tag{60}$$

Since S_n/n is the arithmetic mean of $X_1, \ldots, X_n$, this theorem states that the probability of the arithmetic mean S_n/n differing from its expected value μ by more than ϵ approaches zero as $n \to \infty$. A stronger result, which we might expect to be true, is that $\lim_{n\to\infty} S_n/n = \mu$, but this is actually false. However, we can prove that $\lim_{n\to\infty} S_n/n = \mu$ *with probability one.* This result is often called the *strong law of large numbers*, and, by contrast, that of Theorem 3-19 is called the *weak law of large numbers.* When the "law of large numbers" is referred to without qualification, the weak law is implied.

Other Measures of Central Tendency

As we have already seen, the mean, or expectation, of a random variable X provides a measure of central tendency for the values of a distribution. Although the mean is used most, two other measures of central tendency are also employed. These are the *mode* and the *median.*

1. **MODE.** The *mode* of a discrete random variable is that value which occurs most often or, in other words, has the greatest probability of occurring. Sometimes we have two, three, or more values that have relatively large probabilities of occurrence. In such cases, we say that the distribution is *bimodal*, *trimodal*, or *multimodal*, respectively. The mode of a continuous random variable X is the value (or values) of X where the probability density function has a relative maximum.

2. **MEDIAN.** The *median* is that value x for which $P(X < x) \leq \frac{1}{2}$ and $P(X > x) \leq \frac{1}{2}$. In the case of a continuous distribution we have $P(X < x) = \frac{1}{2} = P(X > x)$, and the median separates the density curve into two parts having equal areas of $1/2$ each. In the case of a discrete distribution a unique median may not exist (see Problem 3.34).

Percentiles

It is often convenient to subdivide the area under a density curve by use of ordinates so that the area to the left of the ordinate is some percentage of the total unit area. The values corresponding to such areas are called *percentile values*, or briefly *percentiles.* Thus, for example, the area to the left of the ordinate at x_α in Fig. 3-2 is α. For instance, the area to the left of $x_{0.10}$ would be 0.10, or 10%, and $x_{0.10}$ would be called the *10th percentile* (also called the *first decile*). The median would be the *50th percentile* (or *fifth decile*).

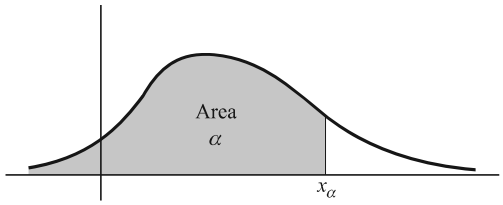

Fig. 3-2

Other Measures of Dispersion

Just as there are various measures of central tendency besides the mean, there are various measures of dispersion or scatter of a random variable besides the variance or standard deviation. Some of the most common are the following.

1. **SEMI-INTERQUARTILE RANGE.** If $x_{0.25}$ and $x_{0.75}$ represent the 25th and 75th percentile values, the difference $x_{0.75} - x_{0.25}$ is called the *interquartile range* and $\frac{1}{2}(x_{0.75} - x_{0.25})$ is the *semi-interquartile range.*
2. **MEAN DEVIATION.** The *mean deviation* (M.D.) of a random variable X is defined as the expectation of $|X - \mu|$, i.e., assuming convergence,

$$\text{M.D.}(X) = E[|X - \mu|] = \sum |x - \mu| f(x) \qquad \text{(discrete variable)} \tag{61}$$

$$\text{M.D.}(X) = E[|X - \mu|] = \int_{-\infty}^{\infty} |x - \mu| f(x)\,dx \qquad \text{(continuous variable)} \tag{62}$$

Skewness and Kurtosis

1. **SKEWNESS.** Often a distribution is not symmetric about any value but instead has one of its tails longer than the other. If the longer tail occurs to the right, as in Fig. 3-3, the distribution is said to be *skewed to the right*, while if the longer tail occurs to the left, as in Fig. 3-4, it is said to be *skewed to the left*. Measures describing this asymmetry are called *coefficients of skewness*, or briefly *skewness*. One such measure is given by

$$\alpha_3 = \frac{E[(X - \mu)^3]}{\sigma^3} = \frac{\mu_3}{\sigma^3} \tag{63}$$

The measure σ_3 will be positive or negative according to whether the distribution is skewed to the right or left, respectively. For a symmetric distribution, $\sigma_3 = 0$.

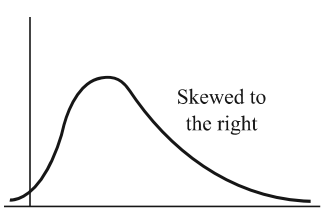

Fig. 3-3

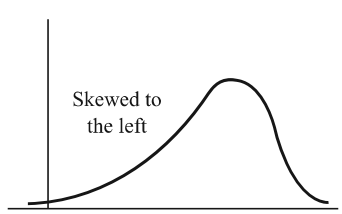

Fig. 3-4

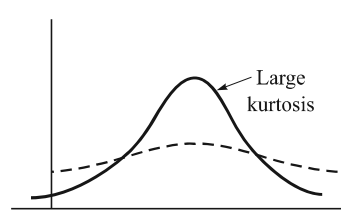

Fig. 3-5

2. **KURTOSIS.** In some cases a distribution may have its values concentrated near the mean so that the distribution has a large peak as indicated by the solid curve of Fig. 3-5. In other cases the distribution may be

relatively flat as in the dashed curve of Fig. 3-5. Measures of the degree of peakedness of a distribution are called *coefficients of kurtosis*, or briefly *kurtosis*. A measure often used is given by

$$\alpha_4 = \frac{E[(X-\mu)^4]}{\sigma^4} = \frac{\mu_4}{\sigma^4} \tag{64}$$

This is usually compared with the normal curve (see Chapter 4), which has a coefficient of kurtosis equal to 3. See also Problem 3.41.

SOLVED PROBLEMS

Expectation of random variables

3.1. In a lottery there are 200 prizes of \$5, 20 prizes of \$25, and 5 prizes of \$100. Assuming that 10,000 tickets are to be issued and sold, what is a fair price to pay for a ticket?

Let X be a random variable denoting the amount of money to be won on a ticket. The various values of X together with their probabilities are shown in Table 3-2. For example, the probability of getting one of the 20 tickets giving a \$25 prize is $20/10{,}000 = 0.002$. The expectation of X in dollars is thus

$$E(X) = (5)(0.02) + (25)(0.002) + (100)(0.0005) + (0)(0.9775) = 0.2$$

or 20 cents. Thus the fair price to pay for a ticket is 20 cents. However, since a lottery is usually designed to raise money, the price per ticket would be higher.

Table 3-2

x (dollars)	5	25	100	0
$P(X = x)$	0.02	0.002	0.0005	0.9775

3.2. Find the expectation of the sum of points in tossing a pair of fair dice.

Let X and Y be the points showing on the two dice. We have

$$E(X) = E(Y) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + \cdots + 6\left(\frac{1}{6}\right) = \frac{7}{2}$$

Then, by Theorem 3-2,

$$E(X+Y) = E(X) + E(Y) = 7$$

3.3. Find the expectation of a discrete random variable X whose probability function is given by

$$f(x) = \left(\frac{1}{2}\right)^x \qquad (x = 1, 2, 3, \ldots)$$

We have

$$E(X) = \sum_{x=1}^{\infty} x\left(\frac{1}{2}\right)^x = \frac{1}{2} + 2\left(\frac{1}{4}\right) + 3\left(\frac{1}{8}\right) + \cdots$$

To find this sum, let $\quad S = \frac{1}{2} + 2\left(\frac{1}{4}\right) + 3\left(\frac{1}{8}\right) + 4\left(\frac{1}{16}\right) + \cdots$

Then $\quad \frac{1}{2}S = \qquad \frac{1}{4} + 2\left(\frac{1}{8}\right) + 3\left(\frac{1}{16}\right) + \cdots$

Subtracting, $\quad \frac{1}{2}S = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = 1$

Therefore, $S = 2$.

3.4. A continuous random variable X has probability density given by

$$f(x) = \begin{cases} 2e^{-2x} & x > 0 \\ 0 & x \le 0 \end{cases}$$

Find (a) $E(X)$, (b) $E(X^2)$.

(a)
$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_0^{\infty} x(2e^{-2x})\,dx = 2\int_0^{\infty} xe^{-2x}dx$$
$$= 2\left[(x)\left(\frac{e^{-2x}}{-2}\right) - (1)\left(\frac{e^{-2x}}{4}\right)\right]\Bigg|_0^{\infty} = \frac{1}{2}$$

(b)
$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = 2\int_0^{\infty} x^2 e^{-2x}dx$$
$$= 2\left[(x^2)\left(\frac{e^{-2x}}{-2}\right) - (2x)\left(\frac{e^{-2x}}{4}\right) + (2)\left(\frac{e^{-2x}}{-8}\right)\right]\Bigg|_0^{\infty} = \frac{1}{2}$$

3.5. The joint density function of two random variables X and Y is given by

$$f(x, y) = \begin{cases} xy/96 & 0 < x < 4,\ 1 < y < 5 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) $E(X)$, (b) $E(Y)$, (c) $E(XY)$, (d) $E(2X + 3Y)$.

(a)
$$E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x, y)\,dx\,dy = \int_{x=0}^{4}\int_{y=1}^{5} x\left(\frac{xy}{96}\right)dx\,dy = \frac{8}{3}$$

(b)
$$E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x, y)\,dx\,dy = \int_{x=0}^{4}\int_{y=1}^{5} y\left(\frac{xy}{96}\right)dx\,dy = \frac{31}{9}$$

(c)
$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (xy)f(x, y)\,dx\,dy = \int_{x=0}^{4}\int_{y=1}^{5} (xy)\left(\frac{xy}{96}\right)dx\,dy = \frac{248}{27}$$

(d)
$$E(2X + 3Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (2x + 3y)f(x, y)\,dx\,dy = \int_{x=0}^{4}\int_{y=1}^{5} (2x + 3y)\left(\frac{xy}{96}\right)dx\,dy = \frac{47}{3}$$

Another method

(c) Since X and Y are independent, we have, using parts (a) and (b),

$$E(XY) = E(X)E(Y) = \left(\frac{8}{3}\right)\left(\frac{31}{9}\right) = \frac{248}{27}$$

(d) By Theorems 3-1 and 3-2, pages 76–77, together with (a) and (b),

$$E(2X + 3Y) = 2E(X) + 3E(Y) = 2\left(\frac{8}{3}\right) + 3\left(\frac{31}{9}\right) = \frac{47}{3}$$

3.6. Prove Theorem 3-2, page 77.

Let $f(x, y)$ be the joint probability function of X and Y, assumed discrete. Then

$$E(X + Y) = \sum_x\sum_y (x + y)f(x, y)$$
$$= \sum_x\sum_y xf(x, y) + \sum_x\sum_y yf(x, y)$$
$$= E(X) + E(Y)$$

If either variable is continuous, the proof goes through as before, with the appropriate summations replaced by integrations. Note that the theorem is true whether or not X and Y are independent.

3.7. Prove Theorem 3-3, page 77.

Let $f(x, y)$ be the joint probability function of X and Y, assumed discrete. If the variables X and Y are independent, we have $f(x, y) = f_1(x) f_2(y)$. Therefore,

$$\begin{aligned} E(XY) &= \sum_x \sum_y xyf(x, y) = \sum_x \sum_y xyf_1(x)f_2(y) \\ &= \sum_x \left[xf_1(x) \sum_y yf_2(y) \right] \\ &= \sum_x [(xf_1(x)E(y)] \\ &= E(X)E(Y) \end{aligned}$$

If either variable is continuous, the proof goes through as before, with the appropriate summations replaced by integrations. Note that the validity of this theorem hinges on whether $f(x, y)$ can be expressed as a function of x multiplied by a function of y, for all x and y, i.e., on whether X and Y are independent. For dependent variables it is not true in general.

Variance and standard deviation

3.8. Find (a) the variance, (b) the standard deviation of the sum obtained in tossing a pair of fair dice.

(a) Referring to Problem 3.2, we have $E(X) = E(Y) = 1/2$. Moreover,

$$E(X^2) = E(Y^2) = 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + \cdots + 6^2\left(\frac{1}{6}\right) = \frac{91}{6}$$

Then, by Theorem 3-4,

$$\operatorname{Var}(X) = \operatorname{Var}(Y) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

and, since X and Y are independent, Theorem 3-7 gives

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) = \frac{35}{6}$$

(b)
$$\sigma_{X+Y} = \sqrt{\operatorname{Var}(X + Y)} = \sqrt{\frac{35}{6}}$$

3.9. Find (a) the variance, (b) the standard deviation for the random variable of Problem 3.4.

(a) As in Problem 3.4, the mean of X is $\mu = E(X) = \frac{1}{2}$. Then the variance is

$$\begin{aligned} \operatorname{Var}(X) &= E[(X - \mu)^2] = E\left[\left(X - \frac{1}{2}\right)^2\right] = \int_{-\infty}^{\infty} \left(x - \frac{1}{2}\right)^2 f(x)\,dx \\ &= \int_0^{\infty} \left(x - \frac{1}{2}\right)^2 (2e^{-2x})\,dx = \frac{1}{4} \end{aligned}$$

Another method
By Theorem 3-4,

$$\operatorname{Var}(X) = E[(X - \mu)^2] = E(X^2) - [E(X)]^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

(b)
$$\sigma = \sqrt{\operatorname{Var}(X)} = \sqrt{\frac{1}{4}} = \frac{1}{2}$$

3.10. Prove Theorem 3-4, page 78.

We have

$$\begin{aligned} E[(X-\mu)^2] &= E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

3.11. Prove Theorem 3-6, page 78.

$$\begin{aligned} E[(X-a)^2] &= E[\{(X-\mu) + (\mu - a)\}^2] \\ &= E[(X-\mu)^2 + 2(X-\mu)(\mu - a) + (\mu - a)^2] \\ &= E[(X-\mu)^2] + 2(\mu - a)E(X-\mu) + (\mu - a)^2 \\ &= E[(X-\mu)^2] + (\mu - a)^2 \end{aligned}$$

since $E(X-\mu) = E(X) - \mu = 0$. From this we see that the minimum value of $E[(X-a)^2]$ occurs when $(\mu - a)^2 = 0$, i.e., when $a = \mu$.

3.12. If $X^* = (X-\mu)/\sigma$ is a standardized random variable, prove that (a) $E(X^*) = 0$, (b) $\mathrm{Var}(X^*) = 1$.

(a)
$$E(X^*) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}[E(X-\mu)] = \frac{1}{\sigma}[E(X) - \mu] = 0$$

since $E(X) = \mu$.

(b)
$$\mathrm{Var}(X^*) = \mathrm{Var}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}E[(X-\mu)^2] = 1$$

using Theorem 3-5, page 78, and the fact that $E[(X-\mu)^2] = \sigma^2$.

3.13. Prove Theorem 3-7, page 78.

$$\begin{aligned} \mathrm{Var}(X+Y) &= E[\{(X+Y) - (\mu_X + \mu_Y)\}^2] \\ &= E[\{(X-\mu_X) + (Y-\mu_Y)\}^2] \\ &= E[(X-\mu_X)^2 + 2(X-\mu_X)(Y-\mu_Y) + (Y-\mu_Y)^2] \\ &= E[(X-\mu_X)^2] + 2E[(X-\mu_X)(Y-\mu_Y)] + E[(Y-\mu_Y)^2] \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) \end{aligned}$$

using the fact that

$$E[(X-\mu_X)(Y-\mu_Y)] = E(X-\mu_X)E(Y-\mu_Y) = 0$$

since X and Y, and therefore $X - \mu_X$ and $Y - \mu_Y$, are independent. The proof of (*19*), page 78, follows on replacing Y by $-Y$ and using Theorem 3-5.

Moments and moment generating functions

3.14. Prove the result (26), page 79.

$$\begin{aligned} \mu_r &= E[(X-\mu)^r] \\ &= E\left[X^r - \binom{r}{1}X^{r-1}\mu + \cdots + (-1)^j\binom{r}{j}X^{r-j}\mu^j \right. \\ &\qquad \left. + \cdots + (-1)^{r-1}\binom{r}{r-1}X\mu^{r-1} + (-1)^r\mu^r\right] \end{aligned}$$

$$= E(X^r) - \binom{r}{1}E(X^{r-1})\mu + \cdots + (-1)^j\binom{r}{j}E(X^{r-j})\mu^j$$

$$+ \cdots + (-1)^{r-1}\binom{r}{r-1}E(X)\mu^{r-1} + (-1)^r\mu^r$$

$$= \mu'_r - \binom{r}{1}\mu'_{r-1}\mu + \cdots + (-1)^j\binom{r}{j}\mu'_{r-j}\mu^j$$

$$+ \cdots + (-1)^{r-1}r\mu^r + (-1)^{-r}\mu^r$$

where the last two terms can be combined to give $(-1)^{r-1}(r-1)\mu^r$.

3.15. Prove (a) result (31), (b) result (32), page 79.

(a) Using the power series expansion for e^u (3., Appendix A), we have

$$M_X(t) = E(e^{tX}) = E\left(1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \cdots\right)$$

$$= 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!}E(X^3) + \cdots$$

$$= 1 + \mu t + \mu'_2\frac{t^2}{2!} + \mu'_3\frac{t^3}{3!} + \cdots$$

(b) This follows immediately from the fact known from calculus that if the Taylor series of $f(t)$ about $t = a$ is

$$f(t) = \sum_{n=0}^{\infty} c_n(t-a)^n$$

then

$$c_n = \frac{1}{n!}\frac{d^n}{dt^n}f(t)\bigg|_{t=a}$$

3.16. Prove Theorem 3-9, page 80.

Since X and Y are independent, any function of X and any function of Y are independent. Hence,

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E(e^{tX}e^{tY}) = E(e^{tX})E(e^{tY}) = M_X(t)M_Y(t)$$

3.17. The random variable X can assume the values 1 and -1 with probability $\frac{1}{2}$ each. Find (a) the moment generating function, (b) the first four moments about the origin.

(a)

$$E(e^{tX}) = e^{t(1)}\left(\frac{1}{2}\right) + e^{t(-1)}\left(\frac{1}{2}\right) = \frac{1}{2}(e^t + e^{-t})$$

(b) We have

$$e^t = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \cdots$$

$$e^{-t} = 1 - t + \frac{t^2}{2!} - \frac{t^3}{3!} + \frac{t^4}{4!} - \cdots$$

Then (1)

$$\frac{1}{2}(e^t + e^{-t}) = 1 + \frac{t^2}{2!} + \frac{t^4}{4!} + \cdots$$

But (2)

$$M_X(t) = 1 + \mu t + \mu'_2\frac{t^2}{2!} + \mu'_3\frac{t^3}{3!} + \mu'_4\frac{t^4}{4!} + \cdots$$

Then, comparing (1) and (2), we have

$$\mu = 0, \qquad \mu'_2 = 1, \qquad \mu'_3 = 0, \qquad \mu'_4 = 1, \ldots$$

The odd moments are all zero, and the even moments are all one.

3.18. A random variable X has density function given by

$$f(x) = \begin{cases} 2e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Find (a) the moment generating function, (b) the first four moments about the origin.

(a)
$$M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$
$$= \int_0^{\infty} e^{tx}(2e^{-2x})\,dx = 2\int_0^{\infty} e^{(t-2)x}\,dx$$
$$= \left.\frac{2e^{(t-2)x}}{t-2}\right|_0^{\infty} = \frac{2}{2-t}, \quad \text{assuming } t < 2$$

(b) If $|t| < 2$ we have

$$\frac{2}{2-t} = \frac{1}{1-t/2} = 1 + \frac{t}{2} + \frac{t^2}{4} + \frac{t^3}{8} + \frac{t^4}{16} + \cdots$$

But
$$M(t) = 1 + \mu t + \mu_2' \frac{t^2}{2!} + \mu_3' \frac{t^3}{3!} + \mu_4' \frac{t^4}{4!} + \cdots$$

Therefore, on comparing terms, $\mu = \frac{1}{2}$, $\mu_2' = \frac{1}{2}$, $\mu_3' = \frac{3}{4}$, $\mu_4' = \frac{3}{2}$.

3.19. Find the first four moments (a) about the origin, (b) about the mean, for a random variable X having density function

$$f(x) = \begin{cases} 4x(9 - x^2)/81 & 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

(a)
$$\mu_1' = E(X) = \frac{4}{81}\int_0^3 x^2(9 - x^2)\,dx = \frac{8}{5} = \mu$$
$$\mu_2' = E(X^2) = \frac{4}{81}\int_0^3 x^3(9 - x^2)\,dx = 3$$
$$\mu_3' = E(X^3) = \frac{4}{81}\int_0^3 x^4(9 - x^2)\,dx = \frac{216}{35}$$
$$\mu_4' = E(X^4) = \frac{4}{81}\int_0^3 x^5(9 - x^2)\,dx = \frac{27}{2}$$

(b) Using the result (27), page 79, we have

$$\mu_1 = 0$$
$$\mu_2 = 3 - \left(\frac{8}{5}\right)^2 = \frac{11}{25} = \sigma^2$$
$$\mu_3 = \frac{216}{35} - 3(3)\left(\frac{8}{5}\right) + 2\left(\frac{8}{5}\right)^3 = -\frac{32}{875}$$
$$\mu_4 = \frac{27}{2} - 4\left(\frac{216}{35}\right)\left(\frac{8}{5}\right) + 6(3)\left(\frac{8}{5}\right)^2 - 3\left(\frac{8}{5}\right)^4 = \frac{3693}{8750}$$

Characteristic functions

3.20. Find the characteristic function of the random variable X of Problem 3.17.

The characteristic function is given by

$$E(e^{i\omega X}) = e^{i\omega(1)}\left(\frac{1}{2}\right) + e^{i\omega(-1)}\left(\frac{1}{2}\right) = \frac{1}{2}(e^{i\omega} + e^{-i\omega}) = \cos\omega$$

using Euler's formulas,

$$e^{i\theta} = \cos\theta + i\sin\theta \qquad e^{-i\theta} = \cos\theta - i\sin\theta$$

with $\theta = \omega$. The result can also be obtained from Problem 3.17(a) on putting $t = i\omega$.

3.21. Find the characteristic function of the random variable X having density function given by

$$f(x) = \begin{cases} 1/2a & |x| < a \\ 0 & \text{otherwise} \end{cases}$$

The characteristic function is given by

$$E(e^{i\omega X}) = \int_{-\infty}^{\infty} e^{i\omega x} f(x)\,dx = \frac{1}{2a}\int_{-a}^{a} e^{i\omega x}\,dx$$

$$= \frac{1}{2a}\frac{e^{i\omega x}}{i\omega}\bigg|_{-a}^{a} = \frac{e^{ia\omega} - e^{-ia\omega}}{2ia\omega} = \frac{\sin a\omega}{a\omega}$$

using Euler's formulas (see Problem 3.20) with $\theta = a\omega$.

3.22. Find the characteristic function of the random variable X having density function $f(x) = ce^{-a|x|}$, $-\infty < x < \infty$, where $a > 0$, and c is a suitable constant.

Since $f(x)$ is a density function, we must have

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

so that

$$c\int_{-\infty}^{\infty} e^{-a|x|}\,dx = c\left[\int_{-\infty}^{0} e^{-a(-x)}\,dx + \int_{0}^{\infty} e^{-a(x)}\,dx\right]$$

$$= c\frac{e^{ax}}{a}\bigg|_{-\infty}^{0} + c\frac{e^{-ax}}{-a}\bigg|_{0}^{\infty} = \frac{2c}{a} = 1$$

Then $c = a/2$. The characteristic function is therefore given by

$$E(e^{i\omega X}) = \int_{-\infty}^{\infty} e^{i\omega x} f(x)\,dx$$

$$= \frac{a}{2}\left[\int_{-\infty}^{0} e^{i\omega x}e^{-a(-x)}\,dx + \int_{0}^{\infty} e^{i\omega x}e^{-a(x)}\,dx\right]$$

$$= \frac{a}{2}\left[\int_{-\infty}^{0} e^{(a+i\omega)x}\,dx + \int_{0}^{\infty} e^{-(a-i\omega)x}\,dx\right]$$

$$= \frac{a}{2}\frac{e^{(a+i\omega)x}}{a+i\omega}\bigg|_{-\infty}^{0} + a\frac{e^{-(a-i\omega)x}}{-(a-i\omega)}\bigg|_{0}^{\infty}$$

$$= \frac{a}{2(a+i\omega)} + \frac{a}{2(a-i\omega)} = \frac{a^2}{a^2+\omega^2}$$

Covariance and correlation coefficient

3.23. Prove Theorem 3-14, page 81.

By definition the covariance of X and Y is

$$\begin{aligned}
\sigma_{XY} = \operatorname{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\
&= E[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y] \\
&= E(XY) - \mu_X E(Y) - \mu_Y E(X) + E(\mu_X\mu_Y) \\
&= E(XY) - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y \\
&= E(XY) - \mu_X\mu_Y \\
&= E(XY) - E(X)E(Y)
\end{aligned}$$

3.24. Prove Theorem 3-15, page 81.

If X and Y are independent, then $E(XY) = E(X)E(Y)$. Therefore, by Problem 3.23,

$$\sigma_{XY} = \text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$$

3.25. Find (a) $E(X)$, (b) $E(Y)$, (c) $E(XY)$, (d) $E(X^2)$, (e) $E(Y^2)$, (f) Var (X), (g) Var (Y), (h) Cov (X, Y), (i) ρ, if the random variables X and Y are defined as in Problem 2.8, pages 47–48.

(a)
$$E(X) = \sum_x \sum_y xf(x, y) = \sum_x x\left[\sum_y f(x, y)\right]$$
$$= (0)(6c) + (1)(14c) + (2)(22c) = 58c = \frac{58}{42} = \frac{29}{21}$$

(b)
$$E(Y) = \sum_x \sum_y yf(x, y) = \sum_y y\left[\sum_x f(x, y)\right]$$
$$= (0)(6c) + (1)(9c) + (2)(12c) + (3)(15c) = 78c = \frac{78}{42} = \frac{13}{7}$$

(c)
$$E(XY) = \sum_x \sum_y xyf(x, y)$$
$$= (0)(0)(0) + (0)(1)(c) + (0)(2)(2c) + (0)(3)(3c)$$
$$+ (1)(0)(2c) + (1)(1)(3c) + (1)(2)(4c) + (1)(3)(5c)$$
$$+ (2)(0)(4c) + (2)(1)(5c) + (2)(2)(6c) + (2)(3)(7c)$$
$$= 102c = \frac{102}{42} = \frac{17}{7}$$

(d)
$$E(X^2) = \sum_x \sum_y x^2 f(x, y) = \sum_x x^2\left[\sum_y f(x, y)\right]$$
$$= (0)^2(6c) + (1)^2(14c) + (2)^2(22c) = 102c = \frac{102}{42} = \frac{17}{7}$$

(e)
$$E(Y^2) = \sum_x \sum_y y^2 f(x, y) = \sum_y y^2\left[\sum_x f(x, y)\right]$$
$$= (0)^2(6c) + (1)^2(9c) + (2)^2(12c) + (3)^2(15c) = 192c = \frac{192}{42} = \frac{32}{7}$$

(f)
$$\sigma_X^2 = \text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{17}{7} - \left(\frac{29}{21}\right)^2 = \frac{230}{441}$$

(g)
$$\sigma_Y^2 = \text{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{32}{7} - \left(\frac{13}{7}\right)^2 = \frac{55}{49}$$

(h)
$$\sigma_{XY} = \text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{17}{7} - \left(\frac{29}{21}\right)\left(\frac{13}{7}\right) = -\frac{20}{147}$$

(i)
$$\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{-20/147}{\sqrt{230/441}\sqrt{55/49}} = \frac{-20}{\sqrt{230}\sqrt{55}} = -0.2103 \text{ approx.}$$

3.26. Work Problem 3.25 if the random variables X and Y are defined as in Problem 2.33, pages 61–63.

Using $c = 1/210$, we have:

(a)
$$E(X) = \frac{1}{210}\int_{x=2}^{6}\int_{y=0}^{5}(x)(2x + y)\,dx\,dy = \frac{268}{63}$$

(b)
$$E(Y) = \frac{1}{210}\int_{x=2}^{6}\int_{y=0}^{5}(y)(2x + y)\,dx\,dy = \frac{170}{63}$$

(c)
$$E(XY) = \frac{1}{210}\int_{x=2}^{6}\int_{y=0}^{5}(xy)(2x + y)\,dx\,dy = \frac{80}{7}$$

(d) $$E(X^2) = \frac{1}{210}\int_{x=2}^{6}\int_{y=0}^{5}(x^2)(2x+y)\,dx\,dy = \frac{1220}{63}$$

(e) $$E(Y^2) = \frac{1}{210}\int_{x=2}^{6}\int_{y=0}^{5}(y^2)(2x+y)\,dx\,dy = \frac{1175}{126}$$

(f) $$\sigma_X^2 = \mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{1220}{63} - \left(\frac{268}{63}\right)^2 = \frac{5036}{3969}$$

(g) $$\sigma_Y^2 = \mathrm{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{1175}{126} - \left(\frac{170}{63}\right)^2 = \frac{16{,}225}{7938}$$

(h) $$\sigma_{XY} = \mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{80}{7} - \left(\frac{268}{63}\right)\left(\frac{170}{63}\right) = -\frac{200}{3969}$$

(i) $$\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{-200/3969}{\sqrt{5036/3969}\sqrt{16{,}225/7938}} = \frac{-200}{\sqrt{2518}\sqrt{16{,}225}} = -0.03129 \text{ approx.}$$

Conditional expectation, variance, and moments

3.27. Find the conditional expectation of Y given $X = 2$ in Problem 2.8, pages 47–48.

As in Problem 2.27, page 58, the conditional probability function of Y given $X = 2$ is

$$f(y\,|\,2) = \frac{4+y}{22}$$

Then the conditional expectation of Y given $X = 2$ is

$$E(Y\,|\,X = 2) = \sum_y y\left(\frac{4+y}{22}\right)$$

where the sum is taken over all y corresponding to $X = 2$. This is given by

$$E(Y\,|\,X = 2) = (0)\left(\frac{4}{22}\right) + 1\left(\frac{5}{22}\right) + 2\left(\frac{6}{22}\right) + 3\left(\frac{7}{22}\right) = \frac{19}{11}$$

3.28. Find the conditional expectation of (a) Y given X, (b) X given Y in Problem 2.29, pages 58–59.

(a) $$E(Y\,|\,X = x)\int_{-\infty}^{\infty} y f_2(y\,|\,x)\,dy = \int_0^x y\left(\frac{2y}{x^2}\right)dy = \frac{2x}{3}$$

(b) $$E(X\,|\,Y = y) = \int_{-\infty}^{\infty} x f_1(x\,|\,y)\,dx = \int_y^1 x\left(\frac{2x}{1-y^2}\right)dx$$

$$= \frac{2(1-y^3)}{3(1-y^2)} = \frac{2(1+y+y^2)}{3(1+y)}$$

3.29. Find the conditional variance of Y given X for Problem 2.29, pages 58–59.

The required variance (second moment about the mean) is given by

$$E[(Y-\mu_2)^2\,|\,X = x] = \int_{-\infty}^{\infty}(y-\mu_2)^2 f_2(y\,|\,x)\,dy = \int_0^x\left(y - \frac{2x}{3}\right)^2\left(\frac{2y}{x^2}\right)dy = \frac{x^2}{18}$$

where we have used the fact that $\mu_2 = E(Y\,|\,X = x) = 2x/3$ from Problem 3.28(a).

Chebyshev's inequality

3.30. Prove Chebyshev's inequality.

We shall present the proof for continuous random variables. A proof for discrete variables is similar if integrals are replaced by sums. If $f(x)$ is the density function of X, then

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty}(x-\mu)^2 f(x)\,dx$$

Since the integrand is nonnegative, the value of the integral can only decrease when the range of integration is diminished. Therefore,

$$\sigma^2 \geq \int_{|x-\mu| \geq \epsilon} (x-\mu)^2 f(x)\,dx \geq \int_{|x-\mu| \geq \epsilon} \epsilon^2 f(x)\,dx = \epsilon^2 \int_{|x-\mu| \geq \epsilon} f(x)\,dx$$

But the last integral is equal to $P(|X-\mu| \geq \epsilon)$. Hence,

$$P(|X-\mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

3.31. For the random variable of Problem 3.18, (a) find $P(|X-\mu| > 1)$. (b) Use Chebyshev's inequality to obtain an upper bound on $P(|X-\mu| > 1)$ and compare with the result in (a).

(a) From Problem 3.18, $\mu = 1/2$. Then

$$P(|X-\mu| < 1) = P\left(\left|X - \frac{1}{2}\right| < 1\right) = P\left(-\frac{1}{2} < X < \frac{3}{2}\right)$$

$$= \int_0^{3/2} 2e^{-2x}\,dx = 1 - e^{-3}$$

Therefore
$$P\left(\left|X - \frac{1}{2}\right| \geq 1\right) = 1 - (1 - e^{-3}) = e^{-3} = 0.04979$$

(b) From Problem 3.18, $\sigma^2 = \mu_2' - \mu^2 = 1/4$. Chebyshev's inequality with $\epsilon = 1$ then gives

$$P(|X-\mu| \geq 1) \leq \sigma^2 = 0.25$$

Comparing with (a), we see that the bound furnished by Chebyshev's inequality is here quite crude. In practice, Chebyshev's inequality is used to provide estimates when it is inconvenient or impossible to obtain exact values.

Law of large numbers

3.32. Prove the law of large numbers stated in Theorem 3-19, page 83.

We have
$$E(X_1) = E(X_2) = \cdots = E(X_n) = \mu$$
$$\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \cdots = \mathrm{Var}(X_n) = \sigma^2$$

Then
$$E\left(\frac{S_n}{n}\right) = E\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n}[E(X_1) + \cdots + E(X_n)] = \frac{1}{n}(n\mu) = \mu$$

$$\mathrm{Var}(S_n) = \mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = n\sigma^2$$

so that
$$\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\mathrm{Var}(S_n) = \frac{\sigma^2}{n}$$

where we have used Theorem 3-5 and an extension of Theorem 3-7.

Therefore, by Chebyshev's inequality with $X = S_n/n$, we have

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}$$

Taking the limit as $n \to \infty$, this becomes, as required,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0$$

Other measures of central tendency

3.33. The density function of a continuous random variable X is

$$f(x) = \begin{cases} 4x(9 - x^2)/81 & 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the mode. (b) Find the median. (c) Compare mode, median, and mean.

(a) The mode is obtained by finding where the density $f(x)$ has a relative maximum. The relative maxima of $f(x)$ occur where the derivative is zero, i.e.,

$$\frac{d}{dx}\left[\frac{4x(9-x^2)}{81}\right] = \frac{36-12x^2}{81} = 0$$

Then $x = \sqrt{3} = 1.73$ approx., which is the required mode. Note that this does give the maximum since the second derivative, $-24x/81$, is negative for $x = \sqrt{3}$.

(b) The median is that value a for which $P(X \le a) = 1/2$. Now, for $0 < a < 3$,

$$P(X \le a) = \frac{4}{81}\int_0^a x(9-x^2)\,dx = \frac{4}{81}\left(\frac{9a^2}{2} - \frac{a^4}{4}\right)$$

Setting this equal to $1/2$, we find that

$$2a^4 - 36a^2 + 81 = 0$$

from which

$$a^2 = \frac{36 \pm \sqrt{(36)^2 - 4(2)(81)}}{2(2)} = \frac{36 \pm \sqrt{648}}{4} = 9 \pm \frac{9}{2}\sqrt{2}$$

Therefore, the required median, which must lie between 0 and 3, is given by

$$a^2 = 9 - \frac{9}{2}\sqrt{2}$$

from which $a = 1.62$ approx.

(c)
$$E(X) = \frac{4}{81}\int_0^3 x^2(9-x^2)\,dx = \frac{4}{81}\left(3x^3 - \frac{x^5}{5}\right)\Bigg|_0^3 = 1.60$$

which is practically equal to the median. The mode, median, and mean are shown in Fig. 3-6.

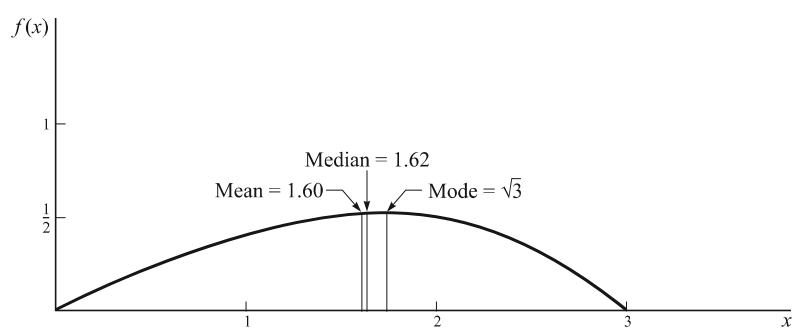

Fig. 3-6

3.34. A discrete random variable has probability function $f(x) = 1/2^x$ where $x = 1, 2, \ldots$. Find (a) the mode, (b) the median, and (c) compare them with the mean.

(a) The mode is the value x having largest associated probability. In this case it is $x = 1$, for which the probability is $1/2$.

(b) If x is any value between 1 and 2, $P(X < x) = \frac{1}{2}$ and $P(X > x) = \frac{1}{2}$. Therefore, *any number* between 1 and 2 could represent the median. For convenience, we choose the midpoint of the interval, i.e., $3/2$.

(c) As found in Problem 3.3, $\mu = 2$. Therefore, the ordering of the three measures is just the reverse of that in Problem 3.33.

Percentiles

3.35. Determine the (a) 10th, (b) 25th, (c) 75th percentile values for the distribution of Problem 3.33.

From Problem 3.33(b) we have

$$P(X \le a) = \frac{4}{81}\left(\frac{9a^2}{2} - \frac{a^4}{4}\right) = \frac{18a^2 - a^4}{81}$$

(a) The 10th percentile is the value of a for which $P(X \le a) = 0.10$, i.e., the solution of $(18a^2 - a^4)/81 = 0.10$. Using the method of Problem 3.33, we find $a = 0.68$ approx.

(b) The 25th percentile is the value of a such that $(18a^2 - a^4)/81 = 0.25$, and we find $a = 1.098$ approx.

(c) The 75th percentile is the value of a such that $(18a^2 - a^4)/81 = 0.75$, and we find $a = 2.121$ approx.

Other measures of dispersion

3.36. Determine, (a) the semi-interquartile range, (b) the mean deviation for the distribution of Problem 3.33.

(a) By Problem 3.35 the 25th and 75th percentile values are 1.098 and 2.121, respectively. Therefore,

$$\text{Semi-interquartile range} = \frac{2.121 - 1.098}{2} = 0.51 \text{ approx.}$$

(b) From Problem 3.33 the mean is $\mu = 1.60 = 8/5$. Then

$$\begin{aligned}
\text{Mean deviation} = \text{M.D.} = E(|X - \mu|) &= \int_{-\infty}^{\infty} |x - \mu| f(x)\,dx \\
&= \int_0^3 \left|x - \frac{8}{5}\right| \left[\frac{4x}{81}(9 - x^2)\right] dx \\
&= \int_0^{8/5} \left(\frac{8}{5} - x\right)\left[\frac{4x}{81}(9 - x^2)\right] dx + \int_{8/5}^3 \left(x - \frac{8}{5}\right)\left[\frac{4x}{81}(9 - x^2)\right] dx \\
&= 0.555 \text{ approx.}
\end{aligned}$$

Skewness and kurtosis

3.37. Find the coefficient of (a) skewness, (b) kurtosis for the distribution of Problem 3.19.

From Problem 3.19(b) we have

$$\sigma^2 = \frac{11}{25} \qquad \mu_3 = -\frac{32}{875} \qquad \mu_4 = \frac{3693}{8750}$$

(a) Coefficient of skewness $= \alpha_3 = \dfrac{\mu_3}{\sigma^3} = -0.1253$

(b) Coefficient of kurtosis $= \alpha_4 = \dfrac{\mu_4}{\sigma^4} = 2.172$

It follows that there is a moderate skewness to the left, as is indicated in Fig. 3-6. Also the distribution is somewhat less peaked than the normal distribution, which has a kurtosis of 3.

Miscellaneous problems

3.38. If $M(t)$ is the moment generating function for a random variable X, prove that the mean is $\mu = M'(0)$ and the variance is $\sigma^2 = M''(0) - [M'(0)]^2$.

From (32), page 79, we have on letting $r = 1$ and $r = 2$,

$$\mu_1' = M'(0) \qquad \mu_2' = M''(0)$$

Then from (27)

$$\mu = M'(0) \qquad \mu_2 = \sigma^2 = M''(0) - [M'(0)]^2$$

3.39. Let X be a random variable that takes on the values $x_k = k$ with probabilities p_k where $k = \pm 1, \ldots, \pm n$. (a) Find the characteristic function $\phi(\omega)$ of X, (b) obtain p_k in terms of $\phi(\omega)$.

(a) The characteristic function is

$$\phi(\omega) = E(e^{i\omega X}) = \sum_{k=-n}^{n} e^{i\omega x_k} p_k = \sum_{k=-n}^{n} p_k e^{ik\omega}$$

(b) Multiply both sides of the expression in (a) by $e^{-ij\omega}$ and integrate with respect to ω from 0 to 2π. Then

$$\int_{\omega=0}^{2\pi} e^{-ij\omega}\phi(\omega)\,d\omega = \sum_{k=-n}^{n} p_k \int_{\omega=0}^{2\pi} e^{i(k-j)\omega}\,d\omega = 2\pi p_j$$

since

$$\int_{\omega=0}^{2\pi} e^{i(k-j)\omega}\,d\omega = \begin{cases} \left.\dfrac{e^{i(k-j)\omega}}{i(k-j)}\right|_0^{2\pi} = 0 & k \neq j \\ 2\pi & k = j \end{cases}$$

Therefore,

$$p_j = \frac{1}{2\pi}\int_{\omega=0}^{2\pi} e^{-ij\omega}\phi(\omega)\,d\omega$$

or, replacing j by k,

$$p_k = \frac{1}{2\pi}\int_{\omega=0}^{2\pi} e^{-ik\omega}\phi(\omega)\,d\omega$$

We often call $\sum_{k=-n}^{n} p_k e^{ik\omega}$ (where n can theoretically be infinite) the *Fourier series* of $\phi(\omega)$ and p_k the *Fourier coefficients.* For a continuous random variable, the Fourier series is replaced by the Fourier integral (see page 81).

3.40. Use Problem 3.39 to obtain the probability distribution of a random variable X whose characteristic function is $\phi(\omega) = \cos\omega$.

From Problem 3.39

$$\begin{aligned} p_k &= \frac{1}{2\pi}\int_{\omega=0}^{2\pi} e^{-ik\omega}\cos\omega\,d\omega \\ &= \frac{1}{2\pi}\int_{\omega=0}^{2\pi} e^{-ik\omega}\left[\frac{e^{i\omega} + e^{-i\omega}}{2}\right]d\omega \\ &= \frac{1}{4\pi}\int_{\omega=0}^{2\pi} e^{i(1-k)\omega}\,d\omega + \frac{1}{4\pi}\int_{\omega=0}^{2\pi} e^{-i(1+k)\omega}\,d\omega \end{aligned}$$

If $k = 1$, we find $p_1 = \frac{1}{2}$; if $k = -1$, we find $p_{-1} = \frac{1}{2}$. For all other values of k, we have $p_k = 0$. Therefore, the random variable is given by

$$X = \begin{cases} 1 & \text{probability } 1/2 \\ -1 & \text{probability } 1/2 \end{cases}$$

As a check, see Problem 3.20.

3.41. Find the coefficient of (a) skewness, (b) kurtosis of the distribution defined by the *normal curve*, having density

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \quad -\infty < x < \infty$$

(a) The distribution has the appearance of Fig. 3-7. By symmetry, $\mu_1' = \mu = 0$ and $\mu_3' = 0$. Therefore the coefficient of skewness is zero.

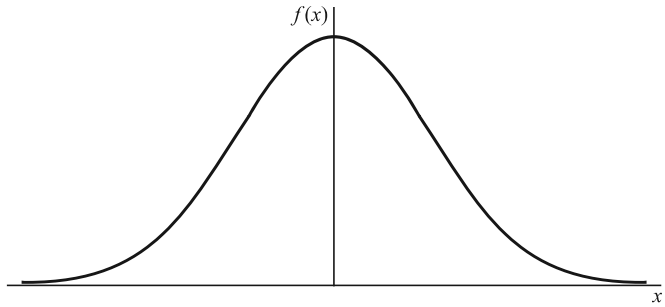

Fig. 3-7

(b) We have

$$\mu_2' = E(X^2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2 e^{-x^2/2}\,dx = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} x^2 e^{-x^2/2}\,dx$$

$$= \frac{2}{\sqrt{\pi}}\int_0^{\infty} v^{1/2} e^{-v}\,dv$$

$$= \frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right) = \frac{2}{\sqrt{\pi}}\cdot\frac{1}{2}\Gamma\left(\frac{1}{2}\right) = 1$$

where we have made the transformation $x^2/2 = v$ and used properties of the gamma function given in (2) and (5) of Appendix A. Similarly we obtain

$$\mu_4' = E(X^4) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^4 e^{-x^2/2}\,dx = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} x^4 e^{-x^2/2}\,dx$$

$$= \frac{4}{\sqrt{\pi}}\int_0^{\infty} v^{3/2} e^{-v}\,dv$$

$$= \frac{4}{\sqrt{\pi}}\Gamma\left(\frac{5}{2}\right) = \frac{4}{\sqrt{\pi}}\cdot\frac{3}{2}\cdot\frac{1}{2}\Gamma\left(\frac{1}{2}\right) = 3$$

Now

$$\sigma^2 = E[(X-\mu)^2] = E(X)^2 = \mu_2' = 1$$

$$\mu_4 = E[(X-\mu)^4] = E(X^4) = \mu_4' = 3$$

Thus the coefficient of kurtosis is

$$\frac{\mu_4}{\sigma^4} = 3$$

3.42. Prove that $-1 \le \rho \le 1$ (see page 82).

For any real constant c, we have

$$E[\{Y - \mu_Y - c(X-\mu)\}^2] \ge 0$$

Now the left side can be written

$$E[(Y-\mu_Y)^2] + c^2E[(X-\mu_X)^2] - 2cE[(X-\mu_X)(Y-\mu_Y)] = \sigma_Y^2 + c^2\sigma_X^2 - 2c\sigma_{XY}$$

$$= \sigma_Y^2 + \sigma_X^2\left(c^2 - \frac{2c\sigma_{XY}}{\sigma_X^2}\right)$$

$$= \sigma_Y^2 + \sigma_X^2\left(c^2 - \frac{\sigma_{XY}}{\sigma_X^2}\right)^2 - \frac{\sigma_{XY}^2}{\sigma_X^2}$$

$$= \frac{\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2}{\sigma_X^2} + \sigma_X^2\left(c - \frac{\sigma_{XY}}{\sigma_X^2}\right)^2$$

In order for this last quantity to be greater than or equal to zero for every value of c, we must have

$$\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2 \geq 0 \quad \text{or} \quad \frac{\sigma_{XY}^2}{\sigma_X^2\sigma_Y^2} \leq 1$$

which is equivalent to $\rho^2 \leq 1$ or $-1 \leq \rho \leq 1$.

SUPPLEMENTARY PROBLEMS

Expectation of random variables

3.43. A random variable X is defined by $X = \begin{cases} -2 & \text{prob. } 1/3 \\ 3 & \text{prob. } 1/2. \\ 1 & \text{prob. } 1/6 \end{cases}$ Find (a) $E(X)$, (b) $E(2X + 5)$, (c) $E(X^2)$.

3.44. Let X be a random variable defined by the density function $f(x) = \begin{cases} 3x^2 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$.

Find (a) $E(X)$, (b) $E(3X - 2)$, (c) $E(X^2)$.

3.45. The density function of a random variable X is $f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$.

Find (a) $E(X)$, (b) $E(X^2)$, (c) $E[(X - 1)^2]$.

3.46. What is the expected number of points that will come up in 3 successive tosses of a fair die? Does your answer seem reasonable? Explain.

3.47. A random variable X has the density function $f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$. Find $E(e^{2X/3})$.

3.48. Let X and Y be independent random variables each having density function

$$f(u) = \begin{cases} 2e^{-2u} & u \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) $E(X + Y)$, (b) $E(X^2 + Y^2)$, (c) $E(XY)$.

3.49. Does (a) $E(X + Y) = E(X) + E(Y)$, (b) $E(XY) = E(X)E(Y)$, in Problem 3.48? Explain.

3.50. Let X and Y be random variables having joint density function

$$f(x, y) = \begin{cases} \frac{3}{5}x(x + y) & 0 \leq x \leq 1, 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) $E(X)$, (b) $E(Y)$, (c) $E(X + Y)$, (d) $E(XY)$.

3.51. Does (a) $E(X + Y) = E(X) + E(Y)$, (b) $E(XY) = E(X)E(Y)$, in Problem 3.50? Explain.

3.52. Let X and Y be random variables having joint density

$$f(x, y) = \begin{cases} 4xy & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) $E(X)$, (b) $E(Y)$, (c) $E(X + Y)$, (d) $E(XY)$.