CHAPTER 8. RANDOM VARIABLES

If we have a probabilistic experimental procedure with a numerical outcome, we say that we have a random variable. The value of the numerical outcome for a particular trial of the experiment is called the value of the random variable for that trial. For example, if our experimental procedure is to roll two dice, and if we take, as our numerical outcome, the sum appearing, we have a random variable X whose possible values are $\{2,3,\ldots,11,12\}$. If we roll two dice and take, as our numerical outcome, the smallest number appearing on either die, we have a different random variable Y with $\{1,2,\ldots,6\}$ as possible values. (In this case, the random variables X and Y are both associated with the same physical experiment.) We shall usually use a capital letter to indicate a random variable and a corresponding small letter to indicate specific individual values of that random variable.

The possible values of a random variable X can be thought of as forming a sample space S_X. If we choose a probability distribution for the sample space S_E of the underlying probabilistic experiment, this distribution on S_E determines a probability distribution on S_X. The probability distribution on the sample space S_X is called the probability distribution of the random variable X.

For example, the random variable Y above has $S_Y = \{1,2,3,4,5,6\}$, and if we assume the equiprobable distribution in the underlying sample space S_E of 36 possible throws of two dice, then the distribution for Y is given in the following table:

y	1	2	3	4	5	6
P (Y = y)	11/36	9/36	7/36	5/36	3/36	1/36

Given a random variable X and a distribution for X, we can think of S_X and the given distribution as forming a probability space. We sometimes speak of this probability space (and hence of the given distribution) as a model for the random variable X.

If two random variables Z_1 and Z_2 arise from separate and independent physical experiments, we say that the random variables are independent. If they arise from the same physical experiment, but all events describable in terms of values of Z_1 (in terms of events in S_{Z_1}) are independent (as events in the probability space S_E of the underlying experiment) of all events describable in terms of values of Z_2 (in terms of events in S_{Z_2}), we also say that the random variable Z_1 and Z_2 are independent. In the experiment of rolling one red die and one blue die, let X_1 be the number on the red die and X_2 be the number on the blue die. Let $X = X_1 + X_2$ be the sum on both dice. Then X_1 and X_2 are independent, but X_1 and X are not independent (because, for example, $P(X_1 = 1 \text{ and } X = 12) = 0 \neq 1/6^3 = P(X_1 = 1)P(X = 12)$). Later in this chapter, we shall at independence of random variables in more detail.

Note. If we are given a sample space S_E for a physical experiment, then any function f which maps S_E into the real

numbers defines a random variable, and if we call that random variable X, then S_X must be $f(S_E)$, the image of S_E under f. Conversely, if we are given a random variable, we have a numerical value associated with each possible outcome of the physical experiment; hence we have a function f which maps S_E into the real numbers. We thus reach the following more formal definition of random variable: a <u>random variable</u> is a mapping from the sample space of a physical experiment into the real numbers.

For example, consider the sample space S_E for rolling two dice (one red and one blue). This space has thirty-six points which we can label $(1,1),(1,2),\ldots,(6,6)$. We let (x_1,x_2) represent a point in S_E. Various random variables can be defined on S_E. The function $f(x_1,x_2) = x_1$ determines the random variable X_1 = number on red die. The function $g(x_1,x_2) = x_2$ determines the random variable X_2 = number on blue die. The function $h(x_1,x_2) = x_1 + x_2$ determines the random variable X = sum appearing on both dice. The function $k(x_1,x_2) = \min(x_1,x_2)$ determines the random variable Y = smallest number on either die. In each case, the random variable is given by a function which maps S_E, the sample space of the underlying physical experiment, into the real numbers.

As soon as we choose a model for the underlying physical experiment by assigning a probability function to the sample space S_E, we immediately have a probability distribution for each random variable on S_E. For example, as we have seen, if we use the equiprobable measure for the case of two dice, this gives the distribution

x	2	3	4	...	12
$P(X = x)$	1/36	2/36	3/36	...	1/36

for X and the distribution

y	1	2	3	4	5	6
$P(X = y)$	11/36	9/36	7/36	5/36	3/36	1/36

for Y.

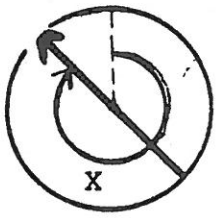
Often, in considering a given random variable X, we will focus our attention entirely on X and will ignore other possible random variables for the same physical experiment. As we have already noted above, we can then think of S_X, the set of possible values of X, as our sample space and we can think of each probability distribution for X as a probability function on this sample space. Hence we can think and speak of the different possible distributions for a random variable X as possible <u>models</u> for X. Random variables are also sometimes called <u>variates</u> or <u>random variates</u>.

<u>Continuous random variables</u>. If we have a random variable X, and if we use a set of models for X in which each of the possible probability distributions for X gives a continuous sample space in the sense of Chapter 2, we say that we have a <u>continuous</u> random variable. Recall that in such a case, a probability distribution is given by a <u>probability density function</u> $f(x)$ such that $f(x) \geq 0$ and $\int_{S_X} f(x)dx = 1$.

Furthermore, for any a and b in S_X, $P(a \leq X \leq b) = \int_a^b f(x)dx$. Continuous random variables are of special importance and usefulness in mathematical statistics. (When we use a set of models for a given random variable X in which each of the possible distributions gives a discrete probability space in the sense of Chapter 2, we say that we have a discrete random variable. The sum appearing on two dice is a discrete random variable.)

Examples of continuous random variables.

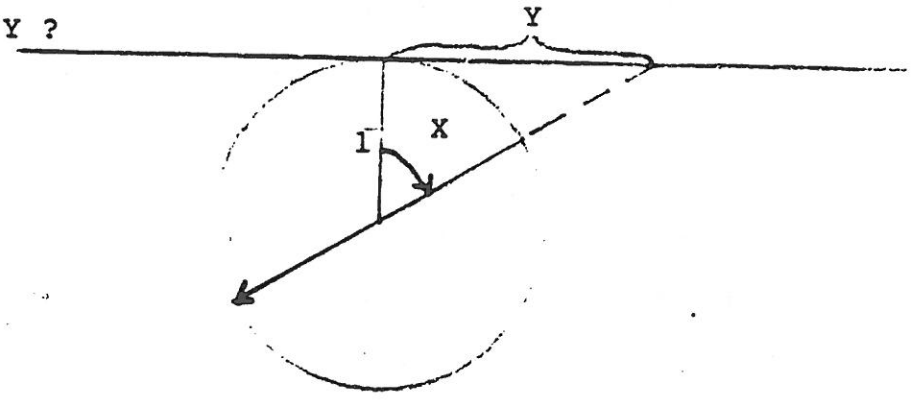
1. A balanced pointer on a circular dial is spun. It comes to rest and the clockwise angle the pointer makes with a fixed reference point on the circle is measured. Let the random variable X be the value of this angle in radians (with $0 \leq X < 2\pi$).



It is often useful, in a case like this, to think of the random variable as continuous (even though, in actual fact, we cannot measure the angle beyond a certain discrete level of accuracy). In this case, we would think of the interval $[0,2\pi]$ as the set S_X of possible values. What would be an appropriate density function? If the physical circumstances lead us to believe that no portion of the dial is favored over any other portion of the same size, we would take the constant function $f(x) = \frac{1}{2\pi}$ on the interval $[0,2\pi]$ as the probability density function. Note that $\int_0^{2\pi} f(x)dx = 1$ as desired. f is called the uniform

density function on the given interval. More generally, for a random variable X with S_X = the interval $[a,b]$, the constant function $\frac{1}{b-a}$ is called the uniform density function for the given random variable. Footnote. Strictly speaking, in the example of the pointer, we should take the sample space to be the set of x such that $0 \leq x < 2\pi$. Adding the point 2π to give the full interval $[0,2\pi]$ does not affect the value of any calculated probability, since, in a continuous distribution, the probability of any individual sample point is zero.

2. A balanced needle is spun on a circular dial of radius 1. When it comes to rest, the line of the needle is extended until it intersects a fixed numerical scale tangent to the circle. We take the origin of the scale as the point of tangency. The random variable Y is defined to be the position of this intersection point on the scale. What is an appropriate density function for Y ?



Let the random variable X be the angle indicated in the figure, where X takes values in the interval $[-\frac{\pi}{2},\frac{\pi}{2}]$. We assume for X the uniform density function $f(x) = \frac{1}{\pi}$ on this interval. Since $P(0 \leq Y \leq a) = P(0 \leq X \leq \tan^{-1} a) = \int_0^{\tan^{-1} a} \frac{1}{\pi}\, dx = \frac{1}{\pi} \tan^{-1} a$, we have

$$P(0 \leq Y \leq a) = \frac{1}{\pi} \tan^{-1} a = \int_0^a g(y)\, dy,$$

where $g(y)$ is the unknown density function for Y. Differentiating, we have, by the fundamental theorem of calculus,

$$g(y) = \frac{1}{\pi(1+y^2)}.$$

This density function gives a probability distribution known as the Cauchy distribution.

3. A random variable with the standard normal curve $\phi(x)$ taken to be its density function is called a standard normal variable. A random variable whose density is taken to be standard normal except for a possible change of origin and a possible linear change of scale is called a normal variable. The general form of the density function for a normal variable is

$$N(x;a,b) = \frac{1}{b}\,\phi\,\left(\frac{x-a}{b}\right) = \frac{1}{b\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2} \quad \text{with } b > 0$$

(The appearance of b in the denominator of the coefficient insures that $\int_{-\infty}^{\infty} f(x)\,dx = 1$.) Normal variables have a major role in mathematical statistics. A later chapter will be devoted entirely to their study.

As we have already noted, the distribution of a random variable serves as a model for the experiment of observing a value of that random variable in exactly the same way that probability spaces in general serve as models for experiments. Indeed, we have already used distributions as models. In the case of a binomial experiment, for example, the number of successes x can be viewed as a random variable, and what we have called <u>binomial distributions</u> are possible distributions for that random variable. Even when the specific distribution of a random variable is unknown, we may be able to decide on some limited set of possible distributions for the random variable (such as the set of all binomial distributions for some fixed n, for example). This set of distributions then becomes our set of possible models. For example, under certain circumstances, we might take the set of all normal distributions as our universe (without deciding on particular values for the parameters a and b in example 3 above).

<u>The median of a random variable</u>. Let a <u>continuous</u> random variable X be given <u>with a distribution</u>. There must be some value m such that $P(X < m) = P(X > m) = 1/2$. We call this value the <u>median</u> of the given distribution. (In special cases where the density function is zero over some central portion of its range, there may be more than one such value m; then the set of such m forms a finite interval, and we take the midpoint of this interval to be the <u>median</u>.) The median of a continuous distribution serves to indicate the middle or center (in the evident intuitive sense given by the definition of median) of that probability distribution.

If we have decided (or are assuming) that a certain distribution is correct for a certain random variable, then we sometimes speak of the median of that distribution as the median of the random variable. When we speak of the median of a random variable, where we have not specified a probability distribution for that random variable, we are referring to the median of a hypothetical, unknown, correct model for the random variable.

In the examples above, the median of the uniform variable on $[0,2\pi]$ is π; the median of the uniform variable on $[a,b]$ is $\frac{a+b}{2}$; the median of the Cauchy distribution is 0; the median of a standard normal variable is 0; and the median of the normal variable with the density $\frac{1}{b}\phi(\frac{x-a}{b})$ is a.

Random variables associated with an experimental procedure. Recall that it is possible for several different random variables to be associated with a single experimental procedure. For example, (1) if our procedure is to roll a red die and a blue die and if we take X_1 = number on red die, X_2 = number on blue die, and Y = sum on both dice, we then have three different random variables associated with this procedure. For another example, (2) if our procedure is to choose a student at random from a certain class and to measure the student's height in centimeters and weight in kilograms, then X = height and Y = weight are two random variables associated with that procedure. For a third example, (3) take a random variable X and let our experimental procedure be to make five successive independent observations of X. Let $X_1, X_2, \ldots, X_5$ be the five

successive values obtained in a single trial of this larger procedure. Then $X_1,\ldots,X_5$ are random variables associated with this procedure. For a fourth example, (4) let X and Y be any two given random variables, each with its own experimental procedure. Define a new combined procedure as follows: first, carry out one trial of the procedure for X, and then carry out a separate independent trial of the procedure for Y. Then X and Y can be viewed as two random variables associated with this single combined procedure.

If we have several random variables associated with an experimental procedure, then the values of some of the variables may be linked or related to values of the others. For example in (1), Y cannot be 11 unless X_1 is, at the same time, either 5 or 6. Similarly in (2), larger values of X will tend to be associated with larger values of Y. In (3), however, particular values of X_1 will not be related to particular values of X_2 (obtained in the same trial of the full procedure). Similarly, in (4) particular values of X will not be related to particular values of Y (obtained in the same trial of the combined procedure).

Joint distributions. We now consider the case of two random variables X and Y (associated with some experimental procedure). (The case of three or more variables will be similar.) How can we give a probability model for such a pair of random variables? We can do this by letting S_{XY} be the set of all possible pairs (x,y) of simultaneous values for X and Y, and then giving

a probability distribution on the set S_{XY}. In the discrete case, we have a probability function p which assigns a probability value $p_{(x,y)}$ to each pair (x,y) in S_{XY}. In the continuous case, we have a density function $h(x,y)$ on S_{XY}. p is called a joint probability function. h is called a joint density function. As in any probability space, we must have $\sum_{(x,y)\in S_{XY}} p_{(x,y)} = 1$ and $\iint_{S_{XY}} h(x,y)\,dx\,dy = 1$. (Note. This is our first mention and use of a continuous probability space in more than 1 dimension.)

Example. Let $S_{X_1X_2}$ be the set of all pairs $(1,1)(1,2),\ldots,(6,6)$ of values for a roll of two dice. Then the usual model (for fair dice) is got by taking the joint probability function to be: $p_{(x_1,x_2)} = \frac{1}{36}$ for all (x_1,x_2).

Probabilities are calculated from a joint distribution in the same way as from the distribution for a single random variable. That is to say, we treat S_{XY} as a sample space. If event R is a subset of S_{XY}, then

$$P(R) = \sum_{(x,y) \text{ in } R} p_{(x,y)}$$

or

$$P(R) = \iint_R h(x,y)\,dx\,dy.$$

Independence. As we have noted earlier in this chapter, if we ar given two random variables X and Y associated with a single experimental procedure, then we say that X and Y are independent in some given model (joint distribution) if every

event (in S_{XY}) definable in terms of X-values alone is independent of every event definable in terms of Y-values alone. The following facts are immediate consequences of this definition.

Let X and Y be given discrete random variables associated with a single experimental procedure. Let p_x and p_y be the individual distributions for X and Y respectively, and let $p_{(x,y)}$ be the joint distribution. Then X and Y are independent if and only if $p_{(x,y)} = p_x p_y$ for all x, y. Similarly, for continuous random variables X and Y, let f(x) and g(y) be the individual density functions, and let h(x,y) be the joint density function. Then X and Y are independent if and only if $h(x,y) = f(x)g(y)$ for all x, y. (These facts follow directly from the multiplication law for independent events.)

Functions of random variables. Given an experimental procedure and given several random variables associated with that procedure, then any chosen function of the given random variables may be used to define a new random variable also associated with that procedure. The distribution of the new random variable will be determined by the joint distribution for the given random variables.

For example, in the case of two dice, if X_1 and X_2 are given, then we can choose the function $f(X_1,X_2) = X_1 + X_2$ and define a new random variable $Y = X_1 + X_2$. The joint distribution for X_1 and X_2 then determines the distribution for Y. Similarly, we can form other random variables such as $2X_1$, X_1X_2, X_1^2, etc. The distribution for $W = X_1^2$, for example, would be given by

w	1	4	9	16	25	36
p_w	1/6	1/6	1/6	1/6	1/6	1/6

,

if we assume the joint probability function $p_{(x_1,x_2)} = \frac{1}{36}$ for all (x_1,x_2).

Expectation. Given a random variable X and a particular distribution for X, we can define the following quantity (where S_X is the set of possible values of X):

$$E_X = \sum_{x \in S_X} x p_x \qquad \text{(for discrete X);}$$

$$E_X = \int_{S_X} x f(x)\,dx \qquad \text{(for continuous X).}$$

This quantity is called the expectation or mean of the distribution for X. It is also sometimes abbreviated as E(X) or as μ_X. If we have decided (or are assuming) that a certain distribution is correct for a certain random variable, then we sometimes speak of the expectation of that distribution as the expectation of the random variable. When we speak of the expectation of a random variable, where we have not specified a distribution for that random variable, we are referring to the expectation of a hypothetical unknown, correct model for the random variable. As part of the above definition, we require in the infinite discrete case that the infinite series be absolutely convergent, and in the continuous case that the integral be absolutely convergent. This requirement is necessary in order to obtain useful facts and laws about E_X, as we shall see in a later chapter.

Example. Let X be the number on a single die. Assume X has the distribution of a fair die. What is E_X? From the definition we get

$$E_X = \sum_{x \in S_X} x p_x = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} = 3.5.$$

Distributions for random variables do not always have expectation. This occurs in infinite discrete cases when the series defining E_X is divergent or conditionally convergent, and in continuous cases when the integral defining E_X is divergent or conditionally convergent.

Examples. You play a game in which your opponent tosses a coin until he gets heads. If he gets heads for the first time on his n^{th} roll, you pay him 2^n dollars. Let X = your loss. What is E_X? Here, $E_X = \sum x p_x = 2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + 8 \cdot \frac{1}{8} + \ldots =$ $1 + 1 + 1 + \ldots$ and is not defined. Let Y have the Cauchy distribution. Then

$$E_Y = \int_{-\infty}^{\infty} \frac{y \, dy}{\pi(1+y^2)} ,$$

and again this is undefined since the integral is not absolutely convergent (since $\lim_{L \to \infty} \int_0^L \frac{y \, dy}{\pi(1+y)^2} = \infty$).

Example. Let X have the normal distribution $f(x) = N(x;a,b)$. It is easy to show that the integral $\int_{-\infty}^{\infty} x f(x) dx$ is absolutely convergent. It follows, from the symmetry of $f(x)$ about $x = a$, that $E_X = a$.

In what follows, we shall usually assume that all random variables under consideration have means. Certain well-known and useful laws of probability and statistics can be proved for such variables. If a random variable does not have a mean, then it may fail to obey these laws.

Algebraic laws for E_X. We shall now present some basic facts about expectation. Proofs for these facts will be given at the end of the chapter.

(I) Let a be a constant, let X be a random variable, and let Y be the new random variable defined by $Y = aX$. Then

$$E_Y = aE_X .$$

(II) Let X and Y be random variables (not necessarily independent) associated with some given experimental procedure. Let Z be the new random variable defined by $Z = X + Y$. Then

$$E_Z = E_X + E_Y .$$

(III) Let X and Y be random variables associated with some given experimental procedure. Assume that X and Y are independent. Let Z be the new random variable defined by $Z = XY$. Then

$$E_Z = E_X E_Y .$$

(IV) Let X be a given random variable and let $k(x)$ be some given function. We define the new random variable $Y = k(X)$. Then

$$E_Y = \sum_{x \in S_X} k(x) p_x \qquad \text{(discrete case)}$$

$$= \int_{S_X} k(x) f(x)\, dx. \qquad \text{(continuous case)}$$

Note that (I) is a special case of (IV). Note also that (II) does not require the assumption of independence.

Example. Fact (II) is especially powerful and useful. Consider an experiment where two decks of cards are shuffled and then laid out face up in parallel. We say that a match occurs at a given position if the same card appears in that position in each deck. Let X = total number of matches. What is E_X ? Define $X_1, X_2, \ldots, X_{52}$ by

$$X_k = \begin{cases} 1 & \text{if match occurs at } k^{th} \text{ position} \\ 0 & \text{if no match at } k^{th} \text{ position} \end{cases}.$$

Then, for each k, $P(X_k = 1) = \frac{1}{52}$. Hence $E_{X_k} = 1 \cdot \frac{1}{52} + 0 \cdot \frac{51}{52} = \frac{1}{52}$. But $X = X_1 + X_2 + \ldots + X_{52}$. By (II), $E_X = E_{X_1} + \ldots + E_{X_{52}} = \frac{1}{52} + \ldots + \frac{1}{52} = 1$. A direct solution to this problem (avoiding (II)) would require calculating the distribution for X. This calculation is much longer than the above calculation of E_X using (II).

Usefulness and significance of expectation. The expectation of a random variable (under a chosen model) has an obvious intuitive meaning: it represents the average value of the random variable in the long run if we assume that observed relative frequencies will agree exactly with probabilities in the chosen model. The expectation therefore serves to locate or represent "typical values" of the random variable. The median

of a continuous random variable (defined in Chapter 15) also serves this latter purpose. The median has the advantage that it always exists, while the expectation may not exist. As we shall see below and in later chapters, however, the expectation is more useful for certain theoretical and computational purposes, because of the algebraic laws which it obeys.

The expectation is also useful for analyzing bets and gambling games. Consider a gambling game where, on a single play, a bettor wins w or loses ℓ. In Chapter 1, we saw that the average winnings per play, after n plays is given by

$$wf_n - \ell(1 - f_n)$$

where f_n is the relative frequency of wins in those n plays. We now define a random variable X, <u>for a single play of the game</u>, as follows:

$$X = \begin{cases} w, & \text{if bettor wins;} \\ -\ell, & \text{if better loses .} \end{cases}$$

If we take a probability model in which p is the probability of winning and q = 1-p is the probability of losing, then we have

$$E_X = wp + (-\ell)q = wp - \ell(1 - p) .$$

If this probability model is a good model, then for successive independent plays of the game, we can expect f_n to become close to p as n increases. Hence we can expect the average winnings per play to become close to E_X.

<u>Expectation</u> can also be used for more complicated bets where the size of payoffs to the bettor (both positive and negative)

may vary depending upon the particular outcome of the game. (Here a negative payoff means a loss.) If we define a random variable X by

$$X = \text{payoff on a single play},$$

then E_X again provides an estimate of expected average winnings per play. It thus provides a numerical measure of how favorable (or unfavorable) a given bet may happen to be.

Examples. In the casino game of chuck-a-luck, the bettor chooses an integer k from the set {1, ..., 6}. The casino then rolls three dice. Let

$$x = \text{number of dice upon which k appears.}$$

Then x can be 0, 1, 2, or 3. If x = 0, the bettor loses 1 dollar. If x > 0, the bettor wins x dollars. What should the bettor expect as average winnings per play? Let

$$X = \text{payoff on a single play.}$$

The binomial formula yields:

$$P(x = 0) = \binom{3}{0}\left(\frac{5}{6}\right)^3 = \frac{125}{216};$$

$$P(x = 1) = \binom{3}{1}\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2 = \frac{75}{216};$$

$$P(x = 2) = \binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right) = \frac{15}{216};$$

$$P(x = 3) = \binom{3}{3}\left(\frac{1}{6}\right)^3 = \frac{1}{216}.$$

Therefore, $E_X = -1\left(\frac{125}{216}\right) + 1\left(\frac{75}{216}\right) + 2\left(\frac{15}{216}\right) + 3\left(\frac{1}{216}\right)$

$$= \frac{-125 + 75 + 30 + 3}{216} = -\frac{17}{216} = -.079$$

for an expected loss of almost 8 cents per play.

In the game of North American roulette, on the other hand, a bet on a single number wins 35 dollars (if the bettor's chosen number appears) and loses 1 dollar (if the chosen number does not appear). Let

$$X = \text{payoff on a single play.}$$

From the description of the game in Chapter 1, we see that the probability distribution for X is given by:

x	35	-1
p_x	1/38	37/38

Hence we have

$$E_X = -1(\tfrac{37}{38}) + 35(\tfrac{1}{38}) = -\tfrac{2}{38} = .053 \ .$$

Thus the expected loss per play at chuck-a-luck is about 50% greater than at North American roulette. (The reader may verify that it is about three times as great as the expected loss per play at European roulette.) At first glance, chuck-a-luck sometimes appears to bettors to be a fair or even favorable game. For this reason, it has occurred widely and in various guises as a sucker bet.

Variance. Given a random variable X and a particular distribution for X, we can also define the following quantity:

$$V_X = \sum_{x \in S_X} (x-\mu_X)^2 p_x \qquad \text{(for discrete X);}$$

$$V_X = \int_{S_X} (x-\mu_X)^2 f(x)\,dx. \qquad \text{(for continuous X).}$$

This quantity is called the variance of the distribution for X. The quantity $\sqrt{V_X}$ is called the standard deviation of the distribution for X, and is written as σ_X. Hence the variance of X is sometimes written as σ_X^2. If we have decided (or are assuming) that a certain distribution is correct for a certain random variable, then we sometimes speak of the variance of that distribution as the variance of the random variable. When we speak of the variance of a random variable, where we have not specified a distribution for that random variable, we are referring to the variance of a hypothetical, unknown correct model for the random variable. (The question of absolute convergence does not arise for V_X, since the terms of the series (for the infinite discrete case) and the values of the integrand (for the continuous case) are always positive.)

Example. Let X be the number appearing on a single die. Assume X has the distribution of a fair die. What is V_X ? From the definition we get

$$V_X = \sum (x-\mu_X)^2 p_x = \sum (x-3.5)^2 p_x$$

$$= (\tfrac{5}{2})^2 \tfrac{1}{6} + (\tfrac{3}{2})^2 \tfrac{1}{6} + (\tfrac{1}{2})^2 \tfrac{1}{6} + (\tfrac{1}{2})^2 \tfrac{1}{6} + (\tfrac{3}{2})^2 \tfrac{1}{6} + (\tfrac{5}{2})^2 \tfrac{1}{6}$$

$$= \frac{35}{12} = 2.92.$$

Hence we have $\sigma_X = \sqrt{\frac{35}{12}} = 1.71.$

A random variable must have a mean in order to have a variance, but it is possible for a random variable to have a mean without having a variance. This occurs when the series or integral defining V_X is divergent even though the series or integral defining E_X is absolutely convergent. In what follows we shall usually assume that all random variables under consideration have both means and variances. We shall see below and in a later chapter that certain useful and important laws of probability and statistics can be proved for random variables that possess variances. If a random variable does not possess a variance, it may fail to obey these laws.

Algebraic laws for V_X. We now present some basic facts about variance. Proofs for these facts will be given at the end of the chapter.

(V) Let a be a constant, let X be a random variable, and let Y be the new random variable defined by $Y = aX$. Then

$$V_Y = a^2 V_X .$$

(VI) For any random variable Y, let $Z = Y^2$. Then we also write E_Z as E_{Y^2}. For any random random variable X we can prove

$$V_X = E_{(X-E_X)^2} = E_{X^2} - (E_X)^2 .$$

(VII) Let X and Y be random variables associated with some given experimental procedure. Assume that X and Y are independent. Let Z be the new random variable defined by $Z = X + Y$. Then

$$V_Z = V_X + V_Y .$$

(VIII) Let a be a constant, let X be a random variable, and let Y be the new random variable defined by $Y = X + a$. Then

$$V_Y = V_X .$$

Fact (VI) is often useful in calculating variance. For example, in the case of a single fair die, we have the simple calculation $E_{X^2} = (1+4+9+16+25+36)\frac{1}{6} = 91/6$, and hence we have $V_X = E_{X^2} - \mu_X^2 = \frac{91}{6} - (\frac{7}{2})^2 = \frac{35}{12}$.

Let X be a given random variable with a given distribution. Assume that X possesses both mean and variance. We can then define the new random variable

$$Y = \frac{X - \mu_X}{\sigma_X} .$$

Y is called the <u>standardized form</u> of X and is sometimes written as S^X. Applying (I) and (II), we have

$$E_Y = \frac{1}{\sigma_X} E(X - \mu_X) = \frac{1}{\sigma_X} (\mu_X - \mu_X) = 0 ;$$

and applying (V) and (VIII), we have

$$V_Y = \frac{1}{\sigma_X^2} V_{(X-\mu_X)} = \frac{1}{\sigma_X^2} V_X = 1 .$$

The distribution for Y must have the same shape as the distribution for X, except for a linear change of scale which results in $E_Y = 0$ and $V_Y = 1$.

<u>Usefulness and importance of variance</u>. The variance of a random variable (under a chosen model) has an obvious intuitive

meaning: it is the mean of the squared deviation of the variable from its mean, and hence serves to measure the spread of the values of the random variable. Other natural measures of spread also exist such as the interquartile range for (the distribution of) a continuous random variable. (The interquartile range of X is $q' - q''$ where $P(X > q') = 1/4$ and $P(X > q'') = 3/4$.) As we have already noted, the variance is especially useful in both theory and computation because of the algebraic laws which it obeys.

Further and more specific significance for the variance is given in the following elementary theorem, which connects the standard deviation to the probability of certain deviations from the mean. The theorem holds for all random variables, X which possess both mean and variances, and is known as Chebychev's inequality. It is formulated in terms of the standardized variable $\frac{X-\mu_X}{\sigma_X}$.

Theorem. For any distribution for a random variable X for which both $E_X = \mu_X$ and $V_X = \sigma_X^2$ exist and for any $t > 0$,

$$P\left(\frac{|X-\mu_X|}{\sigma_X} \geq t\right) \leq \frac{1}{t^2}$$

(or, equivalently,

$$P(|X - \mu_X| \geq t\sigma_X) \leq \frac{1}{t^2} .)$$

We prove the result for the case where X is a discrete random variable. We abbreviate μ_X as μ and σ_X as σ. Beginning with the

definition of V_X and some chosen $t > 0$, we have

$$V_X = \sigma^2 = \sum_{x \in S_X} (x - \mu)^2 p_x \geq \sum_{|x-\mu| \geq t\sigma} (x - \mu)^2 p_x$$

$$\geq \sum_{|x-\mu| \geq t\sigma} (t\sigma)^2 p_x = t^2\sigma^2 P(|X - \mu| \geq t\sigma) .$$

Dividing through by σ^2, we have

$$1 \geq t^2 P(|X - \mu| \geq t\sigma)$$

or

$$P(|X - \mu| \geq t\sigma) \leq 1/t^2 .$$

This completes the proof. (The proof for a continuous random variable is similar, with $\int (x-\mu)^2 f(x)\,dx$ in place of $\sum (x-\mu)^2 p_x$.)

Examples. (a) What are the mean and variance of X where X has, as its distribution, the binomial distribution for n trials with individual success probability p ? We consider first a single trial with probability p. Let Y be the number of successes in a single trial. Then

$$E_Y = 1 \cdot p + 0 \cdot q = p .$$

Also,

$$E_{Y^2} = 1 \cdot p + 0 \cdot q = p .$$

Hence, by (VI), we have

$$V_Y = p - p^2 = p(1 - p) = pq.$$

Using laws (II) and (VII), we immediately have

$$E_X = np$$

$$V_X = npq .$$

Then

$$\sigma_X = \sqrt{npq} \ ,$$

and the standardized variable for X is

$$Z = \frac{X - np}{\sqrt{npq}}$$

(The normal approximation theorem of Chapter 6 tells us that the distribution for Z approaches the form of a standard normal curve as $n \to \infty$.)

(b) What are the mean and variance of X where X has as its distribution the Poisson distribution with parameter value m? Then

$$E_X = \sum_{x=0}^{\infty} xp(x;m) = \sum_{0}^{\infty} \frac{xe^{-m}m^x}{x!} = \sum_{1}^{\infty} \frac{xe^{-m}m^x}{x!}$$

$$= m \sum_{1}^{\infty} \frac{e^{-m}m^{x-1}}{(x-1)!} = m \sum_{0}^{\infty} \frac{e^{-m}m^x}{x!} = m.$$

To get V_X, we use (vi) and first calculate E_{X^2}. We have

$$E_{X^2} = \sum_{0}^{\infty} \frac{x^2e^{-m}m^x}{x!} = \sum_{1}^{\infty} \frac{x^2e^{-m}m^x}{x!}$$

$$= \sum_{1}^{\infty} \frac{xe^{-m}m^x}{(x-1)!} = \sum_{0}^{\infty} \frac{(x+1)e^{-m}m^{x+1}}{x!}$$

$$= m[\sum_{0}^{\infty} \frac{xe^{-m}m^x}{x!} + \sum_{0}^{\infty} \frac{e^{-m}m^x}{x!}] = m(E_X+1)$$

$$= m[m+1] = m^2 + m.$$

Therefore $V_X = m^2 + m - m^2 = m$, and $\sigma_X = \sqrt{m}$. (Again we know from normal approximation (see Chapter 7) that the distribution of the standardized variable $Y = \frac{X-m}{\sqrt{m}}$ approaches the form of a standard normal curve as $m \to \infty$.)

(c) What are the mean and variance of X when X is a continuous random variable with $X \geq 0$ and the exponential density function $f(x) = me^{-mx}$, for some $m > 0$. Then

$$E_X = \int_0^\infty xme^{-mx}\, dx.$$

Evaluating by integration by parts, we obtain

$$E_X = 1/m.$$

To get V_X, we first calculate E_{X^2}. We have

$$E_{X^2} = \int_0^\infty x^2 me^{-mx}\, dx.$$

Evaluating, we obtain

$$E_{X^2} = \frac{2}{m^2}.$$

Hence $V_X = E_{X^2} - (E_X)^2 = \frac{2}{m^2} - \frac{1}{m^2} = \frac{1}{m^2}$, and $\sigma_X = \frac{1}{m}$.
(In this case the standardized variable $Y = \frac{X - \frac{1}{m}}{\frac{1}{m}}$ has the density function $g(y) = e^{-y}$ for all values of $m > 0$.)

(d) What are the mean and variance of X, where X is a normal variable with

$$f(x) = \frac{1}{b\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2} ?$$

Carrying out the integrals, we can show that $E_X = a$ and $V_X = b^2$. (It is not difficult to show that the standardized variable $Z = \frac{X-a}{b}$ then has the standard normal curve as its density function.)

Comment. We shall see in Chapter 17 that it is often reasonable to assume that a given random variable is normal. In that case, the relationship between mean and variance given by Chebychev's inequality can be stated in a much stronger form, for we can show, from tables for the standard normal curve, that for a normal variable X:

$$P(|X-\mu_X| \geq t\sigma_X) \approx \begin{cases} 0.32 & \text{for } t = 1, \\ 0.05 & \text{for } t = 2, \\ 0.003 & \text{for } t = 3. \end{cases}$$

(Here Chebychev's inequality only gives the values 1, 0.25, and 0.11 respectively.)

Sample mean. Let X be a random variable given by some experimental procedure. We define a new and larger procedure as follows: make n successive and independent observations of X. We then consider the new random variables $X_1, \ldots, X_n$, where X_i = the i^{th} observed value in this new and larger procedure. We define a new random variable, associated with this larger procedure, as follows.

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

$\bar{X}_n$ is called the sample mean for n independent trials of the original random variable X. The sample mean is traditionally denoted by

$$\bar{x} = \bar{X}_n$$

We shall commonly use this traditional notation. In this notation, the subscript "n" is always assumed to exist even though we do not bother to write it. The notation $\bar{x}$ is one of the few exceptions, in traditional statistics notation, to the rule that a random variable is always denoted by a capital letter.

Evidently, the distribution of $\bar{x}$ will depend on the distribution of X. What can be said about the mean and variance of $\bar{x}$? We show the following.

Theorem. $E_{\bar{x}} = E_X$;

$V_{\bar{x}} = \frac{1}{n} V_X$.

Proof. For any random variable Y, we introduce the alternative notations

$$E(Y) = E_Y$$

and

$$V(Y) = V_Y \quad .$$

We can now write

$$E_{\bar{x}} = E(\bar{x}) = E\left(\frac{X_1 + \ldots + X_n}{n}\right)$$

$$= \frac{1}{n} E(X_1 + \ldots + X_n) \text{ by (I)},$$

$$= \frac{1}{n} (E(X_1) + \ldots + E(X_n)) \text{ by (II)},$$

$$E_{\bar{x}} = E(\bar{x}) = \frac{1}{n}\,(nE(X)), \text{ since } E(X) = E(X_1) = \ldots = E(X_n$$

$$= E_X \ .$$

Similarly, we have

$$V_{\bar{x}} = V(\bar{x}) = V\left(\frac{X_1 + \ldots + X_n}{n}\right)$$

$$= \frac{1}{n^2}\, V(X_1 + \ldots + X_n) \text{ by (V)},$$

$$= \frac{1}{n^2}\, (V(X_1) + \ldots + V(X_n)) \text{ by (VII)},$$

$$= \frac{1}{n^2}\, (nV(X)), \text{ since } V(X) = V(X_1) = \ldots = V(X$$

$$= \frac{1}{n}\, V_X \ .$$

The weak law of large numbers. The above results for $E_{\bar{x}}$ and $V_{\bar{x}}$, in conjunction with Chebychev's inequality, show us that if a random variable possesses both mean and variance, then the probability distribution for $\bar{x}$ becomes more and more narrowly concentrated about the value E_X as n is increased. This result is known as the weak law of large numbers. It can be stated as follows.

Theorem. Let X be a random variable which possesses both mean and variance. Let $\mu = E_X$. Then for any $\varepsilon > 0$, no matter how small,

$$\lim_{n\to\infty} P(|\bar{x} - \mu| < \varepsilon) = 1 \ .$$

Proof. Chebychev's inequality asserts that for any random variable Y and any $t > 0$,

$$P(|Y - \mu_Y| \geq t\,\sigma_p) \leq 1/t^2 .$$

Applying this to $Y = \bar{x}$ and noting from the previous theorem that

$$\mu_Y = E_{\bar{x}} = E_X = \mu ,$$

and

$$\sigma_Y = \sigma_{\bar{x}} = \frac{1}{\sqrt{n}}\,\sigma_X ,$$

we have

$$P(|\bar{x} - \mu| \geq \frac{t\sigma_X}{\sqrt{n}}) \leq 1/t^2 .$$

Hence

$$P(|\bar{x} - \mu| < \frac{t\sigma_X}{\sqrt{n}}) \geq 1 - 1/t^2 .$$

For any given $\varepsilon > 0$, take $t \equiv \frac{\varepsilon\sqrt{n}}{\sigma_X}$. Then

$$P(|\bar{x} - \mu| < \varepsilon) \geq 1 - \frac{\sigma_X^2}{\varepsilon^2 n} .$$

Hence

$$\lim_{n\to\infty} P(|\bar{x} - \mu| < \varepsilon) = 1 .$$

Central Limit Theorem for $\bar{x}$. We have presented the weak law of large numbers because of its easy, direct proof and its simple, intuitive content. There is, however, a stronger and deeper result which gives more information about the distribution of $\bar{x}$ and which which we shall normally use. This result is a generalization of the De Moivre-Laplace (normal approximation) theorem for Bernoulli trials and is a special case of an even more general result known as the Central Limit Theorem. (The

more general result will be considered in a later chapter.)

> Central Limit Theorem for $\bar{x}$. If a random variable X possesses both mean μ and variance σ^2, then, as we take larger and larger values of n, the distribution of $\bar{x}$ becomes more and more nearly normal with mean μ and variance σ^2/n.

(If we apply the Central Limit Theorem for $\bar{x}$ to the random variable for a single Bernoulli trial,

$$X = \begin{cases} 1 & , \text{ with probability } p, \\ 0 & , \text{ with probability } q = 1 - p, \end{cases}$$

we get the previously studied De Moivre-Laplace normal approximation theorem.)

We shall not give a proof for the Central Limit Theorem for $\bar{x}$ (though the proof is not difficult), and we shall not discuss the rate at which the distribution for $\bar{x}$ approaches a normal distribution (though this information is important in applications). We shall, however, make various applications of the Central Limit Theorem for $\bar{x}$ and shall usually assume, for such applications, that when $n \geq 50$, the approximation yields two-decimal-place accuracy for probabilities of the form $P(a \leq \bar{x} \leq b)$. (In fact, we can often get two-decimal accuracy when $n \geq 10$, as we have seen in the special case of the De Moivre-Laplace Theorem with $p = q = 1/2$. The rate of approach to normality depends upon how symmetrical the distribution for X is, about the value E_X. If the distribution for E_X is sufficiently asymmetrical, the rate can be much slower than $n \geq 50$.)

The Central Limit Theorem for $\bar{x}$ holds no matter what shape the original distribution of X has, provided that X has mean and variance. If X does not have mean and variance, then both the Central Limit Theorem for $\bar{x}$ and the weak law of large numbers may fail for X. For example, if X is a continuous random variable with the Cauchy distribution, then the distribution for $\bar{x}$ can be shown to be identical (for all n) with the original distribution for X. There is thus no change in the shape of the distribution for $\bar{x}$ as n increases, and hence there is no progressively narrower concentration of the distribution for $\bar{x}$ about any value.

The Central Limit Theorem has profound theoretical and practical implications. For this reason, and because of its extreme generality, it is often viewed as the single most important result in all of probability theory. We shall present a more general version of the Central Limit Theorem in a later chapter and shall discuss its implications further at that time. For the moment, we note that the Central Limit Theorem for $\bar{x}$ gives us a theoretical counterpart to weak stability of relative frequencies (discussed in Chapter 1), that it extends this weak stability idea from relative frequencies to sample means in general, and that it gives us a simple and powerful way to calculate relevant probabilities. We illustrate this in the following example.

Example. We return to the game of chuck-a-luck described earlier in this chapter. Recall that if Y is the payoff to the bettor on a single play, then the distribution of Y is given by

y	-1	1	2	3
p_y	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

We have already calculated $E_Y = -\frac{17}{216} = -.079$. We can also calculate

$$V_Y = E_{Y^2} - (E_Y)^2 = \frac{269}{216} - (\frac{17}{216})^2 = 1.239 \qquad \text{(by (VIII))},$$

and hence we have $\sigma_Y = \sqrt{1.239} = 1.113$. Let us now look at this situation from the point of view of the casino. Let X = payoff to casino on a single bet. Then X = -Y, and we immediately have

$$E_X = .079$$

and

$$\sigma_X = 1.113 .$$

Let $X_1, \ldots, X_n$ represent the payoffs to the casino from n different bets of one dollar each. We assume that $X_1, \ldots, X_n$ may be treated as independent random variables. We ask the following question: what is the smallest value of n such that the probability is at least .99 that the net return to the casino is positive? We can use the Central Limit Theorem for $\bar{x}$ to get an immediate answer. We have

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{n}} \sigma_X = \frac{1.113}{\sqrt{n}} .$$

Hence $\frac{\bar{x} - .079}{1.113/\sqrt{n}}$ is a standardized variable and, by the Central Limit Theorem, it has a distribution which approaches the standard normal curve. Note that the net return is positive provided that $\bar{x} > 0$.

We have $\bar{x} > 0 \iff \frac{\bar{x} - .079}{1.113/\sqrt{n}} > - \frac{.079}{1.113/\sqrt{n}}$. Thus we seek to have

$$P\left(\frac{\bar{x} - .079}{1.113/\sqrt{n}} > - \frac{.079\sqrt{n}}{1.113}\right) = .99$$

where $\frac{\bar{x} - .079}{1.113/\sqrt{n}}$ is standard normal. From the table for the standard normal curve, we see that this occurs when

$$\frac{.079\sqrt{n}}{1.113} = 2.326 .$$

Solving for n, we have

$$\sqrt{n} = \frac{(2.326)(1.113)}{.079}$$

$$\text{or } n = 1074.$$

Thus chuck-a-luck is virtually certain to provide a positive return to the casino by the time that 1000 bets have occurred. (Note. What about bets that are placed by different bettors on the same roll of the dice? The payoffs to the casino on these bets are no longer independent as random variables. Since the Central Limit Theorem only applies to independent random variables, does this mean that the analysis just given ceases to apply? In fact, it is possible to show that provided the different bets on the same roll are on different numbers, then a guaranteed (.99 probability) return to the casino must occur at least as soon as it would occur if all the bets were given by independent random variables (i.e., it must occur for a value of n no larger than the value given by the Central Limit Theorem. We omit details.)

Proofs for facts (I) - (VIII). We give proofs for the case where the random variables are continuous and are defined on the interval $(-\infty,\infty)$. Proofs for bounded continuous variables and for discrete variables are essentially the same.

We begin with some simple, though not immediately obvious, facts about density functions and joint density functions.

(1) Let random variables X and Y have the joint density $h(x,y)$. Then we have, from the fact that $P(R) = \iint_R h(x,y)\,dx\,dy$:

$$P(a \leq X \leq b) = \int_a^b \int_{-\infty}^{\infty} h(x,y)dydx$$

This shows us that $f(x) = \int_{-\infty}^{\infty} h(x,y)dy$ must be the density for the random variable X. Similarly $g(y) = \int_{-\infty}^{\infty} h(x,y)dx$ is the density for Y. ($f(x)$ and $g(y)$ are sometimes called the marginal probability densities for the joint probability density $h(x,y)$.)

(2) Let X and Y have the joint density $h(x,y)$, and let Z be a new random variable defined as $Z = X + Y$. We obtain a density function for Z as follows (from $h(x,y)$):

$$P(z_0 \leq x+y \leq z_0+\Delta z) = \int_{-\infty}^{\infty} \int_{z_0-y}^{z_0+\Delta z-y} h(x,y)dxdy$$

$$\approx \int_{-\infty}^{\infty} h(z_0-y,y)\Delta z \, dy$$

(since $\int_{z_0-y}^{z_0+\Delta z-y} h(x,y)dx \approx h(z_0-y,y)\Delta z$ by the definition of definite integral)

$$\approx \Delta z \int_{-\infty}^{\infty} h(z_0-y,y)dy$$

Hence the density function for Z is $j(z)=\int_{-\infty}^{\infty} h(z-y,y)dy$, by the definition of density function.

(3) This now gives us the following simple formula for the expectation of $Z=X+Y$ (in terms of $h(x,y)$). By definition, and from (2) above, we have:

$$E_Z = \int_{-\infty}^{\infty} zj(z)dz = \int_{-\infty}^{\infty} z\int_{-\infty}^{\infty} h(z-y,y)dydz$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} zh(z-y,y)dydz$$

Making the change of variables $\begin{cases} z=x+y \\ y=y \end{cases}$, we get $E_Z=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)h(x,y)dxdy$. (Note that the Jacobian=1 for this change.)

We now prove our basic facts in the following order: (II), (IV), (I), (III), (VI), (V), (VII) and (VIII).

Proof of (II). This crucial result follows from the last formula in (3) above, and from the formulas in (1) for f(x) and g(y). We have:

$$E_{X+Y} = E_Z = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)h(x,y)dxdy$$

$$= \int_{-\infty}^{\infty} x\int_{-\infty}^{\infty} h(x,y)dydx + \int_{-\infty}^{\infty} y\int_{-\infty}^{\infty} h(x,y)dxdy$$

$$= \int_{-\infty}^{\infty} xf(x)dx + \int_{-\infty}^{\infty} yg(y)dy = E_X + E_Y$$

Proof of (IV). We consider the special case where $k(x)$ is a strictly increasing (and therefore one-one) function. The more general proof is essentially the same but has added technical features to take care of the case where k is not one-one. We assume that $Y=k(X)$ and that X has density $f(x)$. Let $g(y)$ be the density for Y. Then

$$E_Y = \int_{-\infty}^{\infty} yg(y)dy \approx \sum y_i g(y_i)\Delta y_i \quad (i=\ldots,-2,-1,0,1,2,\ldots) \quad .$$

Let x_i and Δx_i be a decomposition corresponding to the y_i and Δy_i. That is to say, $y_i=k(x_i)$, $\Delta y_i=y_{i+1}-y_i$, and $\Delta x_i=x_{i+1}-x_i$. Then

$$g(y_i)\Delta y_i = P(y_i \leq Y \leq y_i+\Delta y_i) = P(x_i \leq X \leq x_i+\Delta x_i)=f(x_i)\Delta x_i \quad .$$

Hence $E_Y \approx \sum y_i g(y_i)\Delta y_i \approx \sum k(x_i)f(x_i)\Delta x_i$.

$$\therefore E_Y = \int_{-\infty}^{\infty} k(x)f(x)dx \quad .$$

Proof of (I). Apply (IV) to the special case $Y=aX$. We have $E_{aX}=\int_{-\infty}^{\infty} axf(x)dx =$ $a\int_{-\infty}^{\infty} xf(x)dx=aE_X$.

Proof of (III). Result (IV) can also be proved for joint distributions. It says that if $Z=k(X,Y)$ for some given function k, and if X and Y have the joint density $h(x,y)$, then

$$E_Z = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k(x,y)h(x,y)dxdy \ .$$ (This is a more general form of (3) above.)

The proof is similar to the proof for (IV). Applying this fact to the case $Z=XY$, we ge

$$E_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyh(x,y)dxdy \ .$$

Now if X and Y are independent, we know that

$$h(x,y) = f(x)g(y) \ .$$

Hence $E_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x)g(y)dxdy = \int_{-\infty}^{\infty} xf(x)dx\int_{-\infty}^{\infty} yg(y)dy = E_XE_Y$.

Proof of (VI). By definition, $V_X = \int_{-\infty}^{\infty} (x-\mu)^2f(x)dx$ where $\mu=E_X$. From (IV), this immediately gives that $V_X = E_{(X-\mu)^2}$.

Furthermore, $(X-\mu)^2=X^2-2\mu X+\mu^2$. Hence

$V_X=E_{X^2-2\mu X+\mu^2} =E_{X^2}-2\mu E_X+\mu^2$ (where we make use of (I) and (II)). Hence

$V_X=E_{X^2}-2\mu^2+\mu^2=E_{X^2}-\mu^2$.

Proof of (V). Using (IV) and (VI), we have

$$V_{aX}=E_{(aX-\mu_{aX})^2}=E_{a^2(X-\mu_X)^2}=a^2E_{(X-\mu_X)^2}=a^2V_X \ .$$

Proof of (VII). By (VI), $V_{X+Y}=E_{(X+Y)^2}-(\mu_X+\mu_Y)^2=E_{X^2}+2E_{XY}+E_{Y^2}-\mu_X{}^2-2\mu_X\mu_Y-\mu_Y{}^2$

$$=V_X+V_Y+2(E_{XY}-\mu_X\mu_Y) \quad .$$

If X and Y are independent, then, by (III), $E_{XY}-\mu_X\mu_Y=0$, and we have the desired result.

Proof of (VIII). $V_{X+a} = E_{(X+a-\mu_{X+a})^2}$, by (VI)

$= E_{(X+a-\mu_X-a)^2}$, by (II)

$= E_{(X-\mu_X)^2} = V_X$, by (VI) .

EXERCISES FOR CHAPTER 8

8-1. Four identical light bulbs are temporarily removed from their sockets and placed in a box. The bulbs are then taken at random from the box and put back in the sockets. What is the expected number of bulbs that will be replaced in their original sockets?

8-2. A fair coin is fairly tossed until a head appears. If the first head appears on an odd-numbered trial, the player receives 10 cents; otherwise, he pays 15 cents. What is the expected value of his winnings?

8-3. The random variable X takes the values $-1, 0$, and 1 with probabilities 0.3, 0.2, and 0.5, respectively. Find (a) the mean μ, (b) the variance σ^2, (c) the standard deviation σ.

8-4. X is a random variable which takes on values in the interval $[0,1]$ and which has the density function $f(x)=cx$ for $0 \leq x \leq 1$. Find the constant c and the mean and variance of X.

8-5. Using the formulas

$$1 + 2 + 3 + \ldots + n = \frac{n(n+1)}{2}$$

$$1^2 + 2^2 + 3^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

show that the mean and variance of a random variable that takes the values $1, 2, 3, \ldots, n$, each with probability $1/n$, are

$$\mu = \frac{n+1}{2}, \quad \sigma^2 = \frac{n^2 - 1}{12} .$$

8-6. A regular tetrahedron is a symmetrical solid with four faces. The faces are numbered 1, 2, 3, 4, and the tetrahedron is rolled on the floor. Let the random variable X be the number on the bottom face after the tetrahedron is rolled. Use the result of Problem 5 to find the mean, variance, and standard deviation of X.

8-7. Random variables X and Y are independent and have standard deviations 7 and 24. Let Z = X-Y. What is the standard deviation of Z?

8-8. Find the median of the continuous random variable X whose density function is $f(x) = me^{-mx}$, $x \geq 0$, where m is some given positive real number. (See pages 55-57 in Chapter 1.)

8-9. Consider the game of European roulette as described in Problem 1-4, and recall the special treatment given to bets on red or black. Consider a bet of one dollar on red. Let X be the bettor's winnings on a single play.

(a) Find E_X.

(b) Find V_X.

(c) Use the Central Limit Theorem to find the smallest value of n such that the probability is at least 97.5% that the bettor has a net loss after n successive bets.

8-10. Bets on odd ({1,3,5, ..., 35}) or even ({2,4,6, ..., 36}) in European roulette do not receive the same special treatment that is given to bets on red or black. For bets on odd or on even, $\ell = w = 1$. Consider a bet on odd and let X be the bettors winnings on a single play.

(a), (b), (c) as in Problem 8-9.

8-11. Consider a bet on the single number seventeen at European roulette. Let X be the bettors winnings on a single play.

(a), (b), (c) as in Problem 8-9.

8-12. Consider a bet of one dollar on pass in the game of craps. (See page 29 in Chapter 1 and Exercises 1-5 and 4-10.) Let X be the bettors winnings on a single bet. Assume that P(pass) = .493.

(a), (b), (c) as in Problem 8-9.

8-13. You are told that for a certain sucker bet, the probability is 5/6 that the bettor is behind after 100 bets.

(a) How many bets are needed to get a probability of .975 that the bettor is behind?

(b) How many bets to get a probability of .999 that the bettor is behind?