Chapter 7

Multiple Discrete Random Variables

7.1 Introduction

In Chapter 5 we introduced the concept of a discrete random variable as a mapping from the sample space $\mathcal{S} = \{s_i\}$ to a countable set of real numbers (either finite or countably infinite) via a mapping $X(s_i)$. In effect, the mapping yields useful *numerical* information about the outcome of the random phenomenon. In some instances, however, we would like to measure more than just one attribute of the outcome. For example, consider the choice of a student at random from a population of college students. Then, for the purpose of assessing the student's health we might wish to know his/her height, weight, blood pressure, pulse rate, etc. All these measurements and others are used by a physician for a disease risk assessment. Hence, the mapping from the sample space of college students to the important measurements of height and weight, for example, would be $H(s_i) = h_i$ and $W(s_i) = w_i$, where H and W represent the height and weight of the student selected. In Table 4.1 we summarized a hypothetical set of probabilities for heights and weights. The table is a two-dimensional array that lists the probabilities $P[H = h_i \text{ and } W = w_j]$. This information can also be displayed in a three-dimensional format as shown in Figure 7.1, where we have associated the center point of each interval of height and weight given in Table 4.1 with the probability displayed. These probabilities were termed *joint probabilities*. In this chapter we discuss the case of *multiple random variables*. For example, the height and weight could be represented as a 2×1 *random vector*

$$\begin{bmatrix} H \\ W \end{bmatrix}$$

and as such, its value is located in the plane (also called R^2). We will initially describe the simplest case of two random variables but all concepts are easily ex-

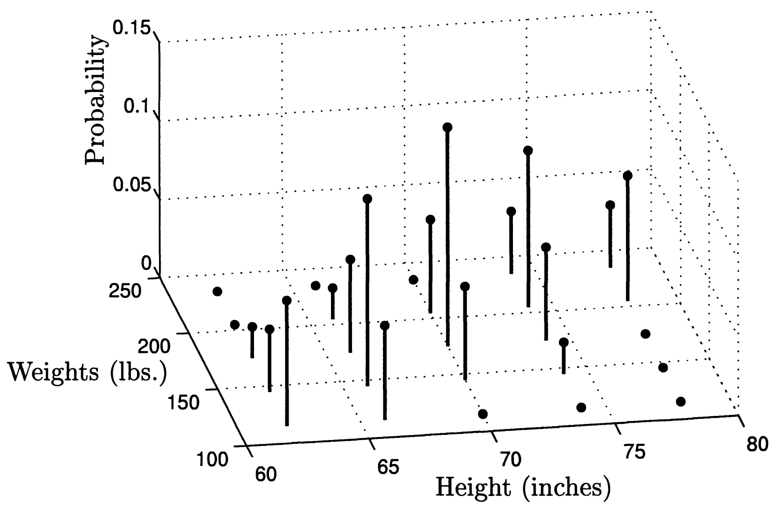


Figure 7.1: Joint probabilities for heights and weights of college students.

tended to any finite number of random variables (see Chapter 9 for this extension). As we will see throughout our discussions, the new and very important concept will be the *dependencies* between the multiple random variables. Questions such as "Can we predict a person's height from his weight?" naturally arise and can be addressed once we extend our description of a single random variable to multiple random variables.

7.2 Summary

The concept of jointly distributed discrete random variables is illustrated in Figure 7.2. Two random variables can be thought of as a random vector and assigned a joint PMF $p_{X,Y}[x_i, y_j]$ as described in Section 7.3, and which has Properties 7.1 and 7.2. The joint PMF may be obtained if the probabilities on the original experimental sample space is known by using (7.2), and is illustrated in Example 7.1. Once the joint PMF is specified, the probability of any event concerning the random variables is determined via (7.3). The marginal PMFs of the two random variables, which are the probabilities of each random variable taking on its possible values, is obtained from the joint PMF using (7.5) and (7.6). However, the joint PMF is not uniquely determined from the marginal PMFs. The joint CDF is defined by (7.7) and evaluated using (7.8). It has the usual properties as summarized via Properties 7.3–7.6. Random variables are defined to be independent if the probabilities of all the joint events can be found as the product of the probabilities of the single events. If the random variables are independent, then the joint PMF factors as in (7.11). Given a joint PMF, independence can be established by determining if the PMF factors. Conversely, if we know the random variables are independent, and we are given the marginal PMFs, then the joint PMF is found as the product of the marginals. The joint PMF of a transformed vector random variable is given by (7.12) and illustrated in Example 7.6. The PMF for the sum of two independent discrete random variables can be found using (7.22) or via characteristic functions using (7.24). The expected value of a function of two random variables is found from (7.28). Also, the variance of the sum of two random variables is given by (7.33) and involves the covariance, which is defined by (7.34). The interpretation of the covariance is given in Section 7.8 and is seen to provide a quantification of the knowledge of the outcome of one random variable on the probability of the other. Independent random variables have a covariance of zero, but the converse is not true. In Section 7.9 linear prediction of one random variable based on observation of another random variable is explored. The optimal linear predictor is given by (7.41). A variation of this prediction equation results in the important parameter called the correlation coefficient (7.43). It quantifies the relationship of one random variable with another. However, a nonzero correlation does not indicate a causal relationship. The joint characteristic function is introduced in Section 7.10 and is defined by (7.45) and evaluated by (7.46). It is shown to provide a convenient means of determining the PMF for a sum of independent random variables. In Section 7.11 a method to simulate a random vector is described. Also, methods to estimate joint PMFs, marginal PMFs, and other quantities of interest are given. Finally, in Section 7.12 an application of the methods of the chapter to disease risk assessment is described.

7.3 Jointly Distributed Random Variables

We consider two discrete random variables that will be denoted by X and Y. As alluded to in the introduction, they represent the functions that map an outcome of an experiment s_i to a value in the plane. Hence, we have the mapping

$$\begin{bmatrix} X(s_i) \\ Y(s_i) \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$$

for all $s_i \in \mathcal{S}$. An example is shown in Figure 7.2 in which the experiment consists of the simultaneous tossing of a penny and a nickel. The outcome in the sample space $\mathcal{S}$ is represented by a TH, for example, if the penny comes up tails and the nickel comes up heads. Explicitly, the mapping is

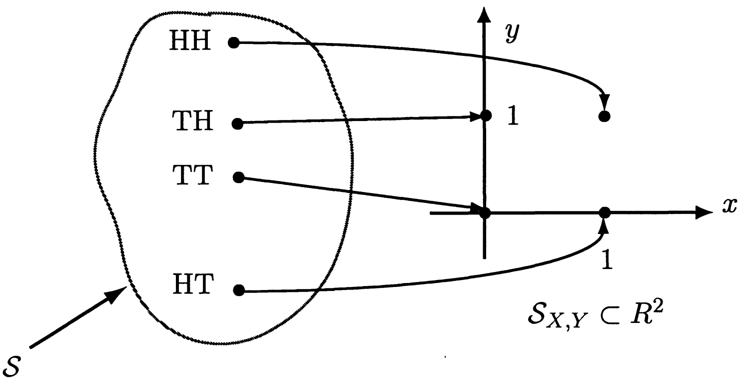


Figure 7.2: Example of mapping for jointly distributed discrete random variables.

$$\begin{bmatrix} X(s_i) \\ Y(s_i) \end{bmatrix} = \begin{cases} \begin{bmatrix} 0 \\ 0 \end{bmatrix} & \text{if } s_i = \text{TT} \\ \begin{bmatrix} 0 \\ 1 \end{bmatrix} & \text{if } s_i = \text{TH} \\ \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \text{if } s_i = \text{HT} \\ \begin{bmatrix} 1 \\ 1 \end{bmatrix} & \text{if } s_i = \text{HH}. \end{cases}$$

Two random variables that are defined on the *same sample space* $\mathcal{S}$ are said to be *jointly distributed.* In this example, the random variables are also *discrete* random variables in that the possible values (which are actually 2×1 vectors) are countable. In this case there are just four vector values. These values comprise the sample space which is the subset of the plane given by

$$\mathcal{S}_{X,Y} = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}.$$

We can also refer to the two random variables as the single *random vector* $[X\, Y]^T$, where T denotes the vector transpose. Hence, we will use the terms *multiple random variables* and *random vector* interchangeably. The values of the random vector will be denoted either by (x, y), which is an ordered pair or a point in the plane, or by $[x\, y]^T$, which denotes a two-dimensional vector. These notations will be synonomous.

The size of the sample space for discrete random variables can be finite or countably infinite. In the example of Figure 7.2, since X can take on 2 values, denoted by $N_X = 2$, and Y can take on 2 values, denoted by $N_Y = 2$, the total number of elements in $\mathcal{S}_{X,Y}$ is $N_X N_Y = 4$. More generally, if X can take on values in $\mathcal{S}_X = \{x_1, x_2, \ldots, x_{N_X}\}$ and Y can take on values in $\mathcal{S}_Y = \{y_1, y_2, \ldots, y_{N_Y}\}$, then the random vector can take on values in

$$\mathcal{S}_{X,Y} = \mathcal{S}_X \times \mathcal{S}_Y = \{(x_i, y_j) : i = 1, 2, \ldots, N_X; j = 1, 2, \ldots, N_Y\}$$

for a total of $N_{X,Y} = N_X N_Y$ values. This is shown in Figure 7.3 for the case of $N_X = 4$ and $N_Y = 3$. The notation $A \times B$, where A and B are sets, denotes a *cartesian product set.* It consists of all ordered pairs (a_i, b_j), where $a_i \in A$ and $b_j \in B$. If either $\mathcal{S}_X$ or $\mathcal{S}_Y$ is countably infinite, then the random vector will also have a countably infinite set of values.

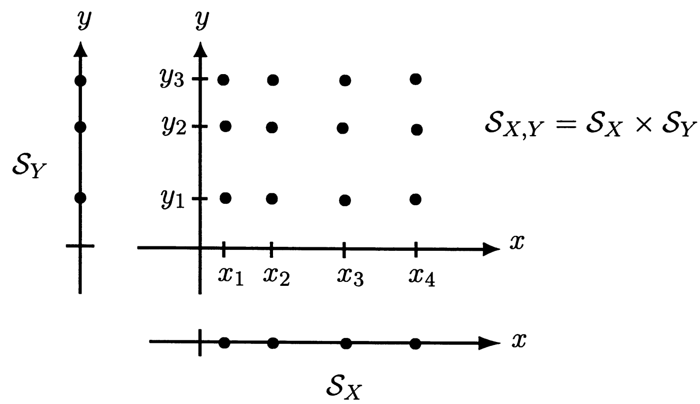


Figure 7.3: Example of sample space for jointly distributed discrete random variables.

Just as we defined the PMF for a single discrete random variable in Chapter 5 as $p_X[x_i] = P[X(s) = x_i]$, we can define the *joint PMF* (or sometimes called the *bivariate* PMF) as

$$p_{X,Y}[x_i, y_j] = P[X(s) = x_i, Y(s) = y_j] \qquad i = 1, 2, \ldots, N_X; j = 1, 2, \ldots, N_Y.$$

Note that the set of all outcomes s for which $X(s) = x_i, Y(s) = y_j$ is the same as the set of outcomes for which

$$\begin{bmatrix} X(s) \\ Y(s) \end{bmatrix} = \begin{bmatrix} x_i \\ y_j \end{bmatrix}$$

so that for the random vector to have the value $[x_i \, y_j]^T$, both $X(s) = x_i$ *and* $Y(s) = y_j$ must be satisfied. Thus, the *comma* used in the statement $X(s) = x_i, Y(s) = y_j$ is to be read as "and". An example of the joint PMF for students' heights and weights is given in Figure 7.1 in which we set X = height and Y = weight and the vertical axis represents $p_{X,Y}[x_i, y_j]$. To verify that a set of probabilities as in Figure 7.1 can be viewed as a joint PMF we need only verify the usual properties of probability. Assuming N_X and N_Y are finite, these are:

Property 7.1 – Range of values of joint PMF

$$0 \leq p_{X,Y}[x_i, y_j] \leq 1 \qquad i = 1, 2, \ldots, N_X; j = 1, 2, \ldots, N_Y.$$

□

Property 7.2 – Sum of values of joint PMF

$$\sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} p_{X,Y}[x_i, y_j] = 1$$

□

and similarly for a countably infinite sample space. For the coin toss example of Figure 7.2 we require that

$$\begin{array}{c} 0 \leq p_{X,Y}[0,0] \leq 1 \\ 0 \leq p_{X,Y}[0,1] \leq 1 \\ 0 \leq p_{X,Y}[1,0] \leq 1 \\ 0 \leq p_{X,Y}[1,1] \leq 1 \end{array}$$

and

$$\sum_{i=0}^{1} \sum_{j=0}^{1} p_{X,Y}[i,j] = 1.$$

Many possibilities exist. For two fair coins that do not interact as they are tossed (i.e., they are independent) we might assign $p_{X,Y}[i,j] = 1/4$ for all i and j. For two coins that are weighted but again do not interact with each other as they are tossed, we might assign

$$p_{X,Y}[i,j] = \begin{cases} (1-p)^2 & i = 0, j = 0 \\ (1-p)p & i = 0, j = 1 \\ p(1-p) & i = 1, j = 0 \\ p^2 & i = 1, j = 1 \end{cases}$$

if each coin has a probability of heads of p. It is easily shown that the joint PMF satisfies Properties 7.1 and 7.2 for any $0 \leq p \leq 1$. In obtaining these values for the joint PMF we have used the concept of equivalent events, which allows us to determine probabilities for events defined on $\mathcal{S}_{X,Y}$ from those defined on the original sample space $\mathcal{S}$. For example, since the events TH and $(0, 1)$ are equivalent as seen in Figure 7.2, we have that

$$\begin{aligned} p_{X,Y}[0,1] &= P[X(s)=0, Y(s)=1] \\ &= P[\{s_i : X(s_i)=0, Y(s_i)=1\}] && \text{(equivalent event in } \mathcal{S}) \\ &= P[s_i = \text{TH}] && \text{(mapping is one-to-one)} \\ &= (1-p)p && \text{(independence)} \end{aligned}$$

where we have assumed independence of the penny and nickel toss subexperiments as described in Section 4.6.1.

In general, the procedure to determine the joint PMF from the probabilities defined on $\mathcal{S}$ depends on whether the random variable mapping is one-to-one or many-to-one. For a one-to-one mapping from $\mathcal{S}$ to $\mathcal{S}_{X,Y}$ we have

$$\begin{aligned} p_{X,Y}[x_i, y_j] &= P[X(s)=x_i, Y(s)=y_j] \\ &= P[\{s : X(s)=x_i, Y(s)=y_j\}] \\ &= P[\{s_k\}] \end{aligned} \tag{7.1}$$

where it is assumed that s_k is the only solution to $X(s) = x_i$ and $X(s) = y_j$. For a many-to-one transformation the joint PMF is found as

$$p_{X,Y}[x_i, y_j] = \sum_{\{k: X(\mathcal{S}_k)=x_i, Y(\mathcal{S}_k)=y_j\}} P[\{s_k\}]. \tag{7.2}$$

This is the extension of (5.1) and (5.2) to a two-dimensional random vector. An example follows.

Example 7.1 – Two dice toss with different colored dice

A red die and a blue die are tossed. The die that yields the larger number of dots is chosen. If they both display the same number of dots, the red die is chosen. The numerical outcome of the experiment is defined to be 0 if the blue die is chosen and 1 if the red die is chosen, along with its corresponding number of dots. The random vector is therefore defined as

$$\begin{aligned} X &= \begin{cases} 0 & \text{blue die chosen} \\ 1 & \text{red die chosen} \end{cases} \\ Y &= \text{number of dots on chosen die.} \end{aligned}$$

The outcomes of the experiment can be represented by (i, j) where $i = 0$ for blue, $i = 1$ for red, and j is the number of dots observed. What then is $p_{X,Y}[1,3]$, for example? To determine this we first list all outcomes in Table 7.1 for each number of dots observed on the red and blue dice. It is seen that the mapping is many-to-one. For example, if the red die displays 6 dots, then the outcome is the same, which is $(1,6)$, for all possible blue outcomes. To determine the desired value of the PMF,

	blue=1	blue=2	blue=3	blue=4	blue=5	blue=6
red=1	$(1,1)$	$(0,2)$	$(0,3)$	$(0,4)$	$(0,5)$	$(0,6)$
red=2	$(1,2)$	$(1,2)$	$(0,3)$	$(0,4)$	$(0,5)$	$(0,6)$
red=3	$(1,3)$	$(1,3)$	$(1,3)$	$(0,4)$	$(0,5)$	$(0,6)$
red=4	$(1,4)$	$(1,4)$	$(1,4)$	$(1,4)$	$(0,5)$	$(0,6)$
red=5	$(1,5)$	$(1,5)$	$(1,5)$	$(1,5)$	$(1,5)$	$(0,6)$
red=6	$(1,6)$	$(1,6)$	$(1,6)$	$(1,6)$	$(1,6)$	$(1,6)$

Table 7.1: Mapping of outcomes in $\mathcal{S}$ to outcomes in $\mathcal{S}_{X,Y}$. The outcomes of (X,Y) are (i,j), where i indicates the color of the die with more dots (red=1, blue=0), j indicates the number of dots on that die.

we assume that each outcome in $\mathcal{S}$ is equally likely and therefore is equal to 1/36. Then, from (7.2)

$$\begin{aligned} p_{X,Y}[1,3] &= \sum_{\{k:X(\mathcal{S}_k)=1,Y(\mathcal{S}_k)=3\}} P[\{s_k\}] \\ &= \sum_{\{k:X(\mathcal{S}_k)=1,Y(\mathcal{S}_k)=3\}} \frac{1}{36} \\ &= \frac{3}{36} = \frac{1}{12} \end{aligned}$$

since there are three outcomes of the experiment in $\mathcal{S}$ that map into $(1,3)$. They are (red=3,blue=1), (red=3,blue=2), and (red=3,blue=3).

◇

In general, as in the case of a single random variable we can use the joint PMF to compute probabilities of all events defined on $\mathcal{S}_{X,Y} = \mathcal{S}_X \times \mathcal{S}_Y$. For the event $A \subset \mathcal{S}_{X,Y}$, the probability is

$$P[(X,Y) \in A] = \sum_{\{(i,j):(x_i,y_j)\in A\}} p_{X,Y}[x_i,y_j]. \tag{7.3}$$

Once we have knowledge of the joint PMF, we no longer need to retain the underlying sample space $\mathcal{S}$ of the experiment. All our probability calculations can be made concerning values of (X,Y) by using (7.3).

7.4 Marginal PMFs and CDFs

If the joint PMF is known, then the PMF for X, i.e., $p_X[x_i]$, and the PMF for Y, i.e., $p_Y[y_j]$, can be determined. These are termed the *marginal PMFs*. Consider first the determination of $p_X[x_i]$. Since $\{X = x_i\}$ does not specify any particular value for Y, the event $\{X = x_i\}$ is equivalent to the joint event $\{X = x_i, Y \in \mathcal{S}_Y\}$.

To determine the probability of the latter event we assume the general case of a countably infinite sample space. Then, (7.3) becomes

$$P[(X,Y) \in A] = \sum_{\substack{i=1 \\ \{(i,j):(x_i,y_j)\in A\}}}^{\infty} \sum_{j=1}^{\infty} p_{X,Y}[x_i, y_j]. \tag{7.4}$$

Next let $A = \{x_k\} \times \mathcal{S}_Y$, which is illustrated in Figure 7.4 for $k = 3$. Then, we have

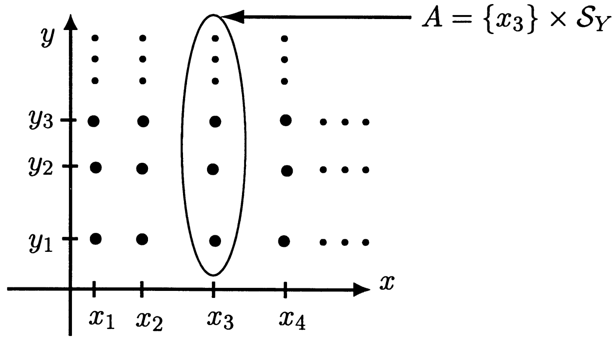


Figure 7.4: Determination of marginal PMF value $p_X[x_3]$ from joint PMF $p_{X,Y}[x_i, y_j]$ by summing along y direction.

$$\begin{aligned} P[(X,Y) \in \{x_k\} \times \mathcal{S}_Y] &= P[X = x_k, Y \in \mathcal{S}_Y] \\ &= P[X = x_k] \\ &= p_X[x_k] \end{aligned}$$

so that from (7.4) with $i = k$ only

$$p_X[x_k] = \sum_{j=1}^{\infty} p_{X,Y}[x_k, y_j] \tag{7.5}$$

and is obtained for $k = 3$ by summing the probabilities along the column shown in Figure 7.4. The terminology "marginal" PMF originates from the process of summing the probabilities along each column and writing the results in the margin (below the x axis), much the same as the process for computing the marginal probability discussed in Section 4.3. Likewise, by summing along each row or in the x direction we obtain the marginal PMF for Y as

$$p_Y[y_k] = \sum_{i=1}^{\infty} p_{X,Y}[x_i, y_k]. \tag{7.6}$$

In summary, we see that *from the joint PMF we can obtain the marginal PMFs.* Another example follows.

Example 7.2 – Two coin toss

As before we toss a penny and a nickel and map the outcomes into a 1 for a head and a 0 for a tail. The random vector is (X, Y), where X is the random variable representing the penny outcome and Y is the random variable representing the nickel outcome. The mapping is shown in Figure 7.2. Consider the joint PMF

$$p_{X,Y}[i,j] = \begin{cases} \frac{1}{8} & i=0, j=0 \\ \frac{1}{8} & i=0, j=1 \\ \frac{1}{4} & i=1, j=0 \\ \frac{1}{2} & i=1, j=1. \end{cases}$$

Then, the marginal PMFs are given as

$$\begin{aligned} p_X[i] &= \sum_{j=0}^{1} p_{X,Y}[i,j] = \begin{cases} \frac{1}{8}+\frac{1}{8}=\frac{1}{4} & i=0 \\ \frac{1}{4}+\frac{1}{2}=\frac{3}{4} & i=1 \end{cases} \\ p_Y[j] &= \sum_{i=0}^{1} p_{X,Y}[i,j] = \begin{cases} \frac{1}{8}+\frac{1}{4}=\frac{3}{8} & j=0 \\ \frac{1}{8}+\frac{1}{2}=\frac{5}{8} & j=1. \end{cases} \end{aligned}$$

As expected, $\sum_{i=0}^{1} p_X[i] = 1$ and $\sum_{j=0}^{1} p_Y[j] = 1$. We could also have arranged the joint PMF and marginal PMF values in a table as shown in Table 7.2. Note that

	$j=0$	$j=1$	$p_X[i]$
$i=0$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
$i=1$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$
$p_Y[j]$	$\frac{3}{8}$	$\frac{5}{8}$	

Table 7.2: Joint PMF and marginal PMF values for Examples 7.2 and 7.4.

the marginal PMFs are found by summing across a row (for p_X) or a column (for p_Y) and are written in the "margins".

◇

Joint PMF cannot be determined from marginal PMFs.

Having obtained the marginal PMFs from the joint PMF, we might suppose we could reverse the process to find the joint PMF from the marginal PMFs. However, this is *not possible* in general. To see why, consider the joint PMF summarized in Table 7.3. The marginal PMFs are the same as the ones shown in Table 7.2. In

	$j=0$	$j=1$	$p_X[i]$
$i=0$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{4}$
$i=1$	$\frac{5}{16}$	$\frac{7}{16}$	$\frac{3}{4}$
$p_Y[j]$	$\frac{3}{8}$	$\frac{5}{8}$	

Table 7.3: Joint PMF values for "caution" example.

fact, *there are an infinite number of joint PMFs that have the same marginal PMFs.* Hence,

$$\text{joint PMF} \Rightarrow \text{marginal PMFs}$$

but

$$\text{marginal PMFs} \not\Rightarrow \text{joint PMF}.$$

A joint cumulative distribution function (CDF) can also be defined for a random vector. It is given by

$$F_{X,Y}(x,y) = P[X \le x, Y \le y] \tag{7.7}$$

and can be found explicitly by summing the joint PMF as

$$F_{X,Y}(x,y) = \sum_{\{(i,j):x_i \le x, y_j \le y\}}\sum p_{X,Y}[x_i, y_j]. \tag{7.8}$$

An example is shown in Figure 7.5, along with the joint PMF. The *marginal* CDFs can be easily found from the joint CDF as

$$\begin{aligned} F_X(x) &= P[X \le x] = P[X \le x, Y < \infty] = F_{X,Y}(x, \infty) \\ F_Y(y) &= P[Y \le y] = P[X < \infty, Y \le y] = F_{X,Y}(\infty, y). \end{aligned}$$

The joint CDF has the usual properties which are:

Property 7.3 – Range of values

$$0 \le F_{X,Y}(x,y) \le 1$$

□

Property 7.4 – Values at "endpoints"

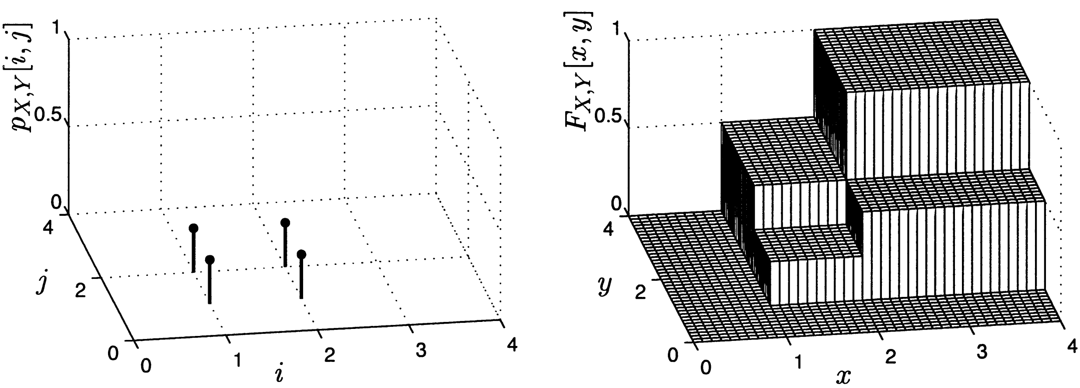
$$\begin{aligned} F_{X,Y}(-\infty, -\infty) &= 0 \\ F_{X,Y}(\infty, \infty) &= 1 \end{aligned}$$

□

Property 7.5 – Monotonically increasing
$F_{X,Y}(x, y)$ monotonically increases as x and/or y increases.

□

Property 7.6 – "Right" continuous
As expected, the joint CDF takes the value *after* the jump. However, in this case the jump is a line discontinuity as seen, for example, in Figure 7.5b. *After* the jump means as we move in the northeast direction in the x-y plane.



(a) Joint PMF

(b) Joint CDF

Figure 7.5: Joint PMF and corresponding joint CDF.

□

The reader is asked to verify some of these properties in Problem 7.17. Finally, to recover the PMF we can use

$$p_{X,Y}[x_i, y_j] = F_{X,Y}(x_i^+, y_j^+) - F_{X,Y}(x_i^+, y_j^-) - F_{X,Y}(x_i^-, y_j^+) + F_{X,Y}(x_i^-, y_j^-). \quad (7.9)$$

The reader should verify this formula for the joint CDF shown in Figure 7.5b. In particular, consider the joint PMF at the point $(x_i, y_j) = (2, 2)$ to see why we need four terms.

7.5 Independence of Multiple Random Variables

Consider the experiment of tossing a coin and then a die. The outcome of the coin toss is denoted by X and equals 0 for a tail and 1 for a head. The outcome for the die is denoted by Y, which takes on the usual values $1, 2, 3, 4, 5, 6$. In determining the probability of the random vector (X, Y) taking on a value, there is no reason to believe that the probability of $Y = y_j$ should depend on the outcome of the coin toss. Likewise, the probability of $X = x_i$ should not depend on the outcome of the die toss (especially since the die toss occurs at a later time). We expect that these two events are *independent*. The formal definition of *independent random variables* X and Y is that they are independent if *all* the joint events on $\mathcal{S}_{X,Y}$ are independent. Mathematically X and Y are independent random variables if for all events $A \subset \mathcal{S}_X$ and $B \subset \mathcal{S}_Y$

$$P[X \in A, Y \in B] = P[X \in A]P[Y \in B]. \tag{7.10}$$

The probabilities on the right-hand-side of (7.10) are defined on $\mathcal{S}_X$ and $\mathcal{S}_Y$, respectively (see Figure 7.3 for an example of the relationship of $\mathcal{S}_X, \mathcal{S}_Y$ to $\mathcal{S}_{X,Y}$). The utility of the independence property is that the probabilities of joint events may be reduced to probabilities of "marginal events" (defined on $\mathcal{S}_X$ and $\mathcal{S}_Y$), which are always easier to determine. Specifically, if X and Y are independent random variables, then it follows from (7.10) that

$$p_{X,Y}[x_i, y_j] = p_X[x_i]p_Y[y_j] \tag{7.11}$$

as we now show. If $A = \{x_i\}$ and $B = \{y_j\}$, then the left-hand-side of (7.10) becomes

$$\begin{aligned} P[X \in A, Y \in B] &= P[X = x_i, Y = y_j] \\ &= p_{X,Y}[x_i, y_j] \end{aligned}$$

and the right-hand-side of (7.10) becomes

$$P[X \in A]P[Y \in B] = p_X[x_i]p_Y[y_j].$$

Hence, *if X and Y are independent random variables, the joint PMF factors into the product of the marginal PMFs.* Furthermore, the converse is true—*if the joint PMF factors, then X and Y are independent.* To prove the converse assume that the joint PMF factors according to (7.11). Then for all A and B we have

$$\begin{aligned} P[X \in A, Y \in B] &= \sum_{\{i:x_i \in A\}} \sum_{\{j:y_j \in B\}} p_{X,Y}[x_i, y_j] && \text{(from (7.3))} \\ &= \sum_{\{i:x_i \in A\}} \sum_{\{j:y_j \in B\}} p_X[x_i]p_Y[y_j] && \text{(assumption)} \\ &= \sum_{\{i:x_i \in A\}} p_X[x_i] \sum_{\{j:y_j \in B\}} p_Y[y_j] \\ &= P[X \in A]P[Y \in B]. \end{aligned}$$

We now illustrate the concept of independent random variables with some examples.

Example 7.3 – Two coin toss – independence
Assume that we toss a penny and a nickel and that as usual a tail is mapped into a 0 and a head into a 1. If all outcomes are equally likely or equivalently the joint PMF is given in Table 7.4, then the random variables must be independent. This is

	$j=0$	$j=1$	$p_X[i]$
$i=0$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
$i=1$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
$p_Y[j]$	$\frac{1}{2}$	$\frac{1}{2}$	

Table 7.4: Joint PMF and marginal PMF values for Example 7.3.

because we can factor the joint PMF as

$$p_{X,Y}[i,j] = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = p_X[i]p_Y[j]$$

for all i and j for which $p_{X,Y}[i,j]$ is nonzero. Furthermore, the marginal PMFs indicate that each coin is fair since $p_X[0] = p_X[1] = 1/2$ and $p_Y[0] = p_Y[1] = 1/2$.

◇

Example 7.4 – Two coin toss – dependence
Now consider the same experiment but with a joint PMF given in Table 7.2. We see that $p_{X,Y}[0,0] = 1/8 \neq (1/4)(3/8) = p_X[0]p_Y[0]$ and hence X and Y cannot be independent. If two random variables are not independent, they are said to be *dependent.*

◇

Example 7.5 – Two coin toss – dependent but fair coins
Consider the same experiment again but with the joint PMF given in Table 7.5. Since $p_{X,Y}[0,0] = 3/8 \neq (1/2)(1/2) = p_X[0]p_Y[0]$, X and Y are *dependent.* However,

	$j=0$	$j=1$	$p_X[i]$
$i=0$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{2}$
$i=1$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{2}$
$p_Y[j]$	$\frac{1}{2}$	$\frac{1}{2}$	

Table 7.5: Joint PMF and marginal PMF values for Example 7.5.

by examining the marginal PMFs we see that the coins are in some sense fair since $P[\text{heads}] = 1/2$, and therefore we might conclude that the random variables were independent. This is incorrect and underscores the fact that the marginal PMFs do not tell us much about the joint PMF. The joint PMF of Table 7.4 also has the same marginal PMFs but there X and Y were independent.

◇

Finally, note that if the random variables are independent, the joint CDF factors as well. This is left as an exercise for the student (see Problem 7.20). Intuitively, if X and Y are independent random variables, then knowledge of the outcome of X does not change the probabilities of the outcomes of Y. This means that we cannot predict Y based on knowing that $X = x_i$. Our best predictor of Y is just $E[Y]$, as described in Example 6.3. When X and Y are dependent, however, we can improve upon the predictor $E[Y]$ by using the knowledge that $X = x_i$. How we actually do this is described in Section 7.9.

7.6 Transformations of Multiple Random Variables

In Section 5.7 we have seen how to find the PMF of $Y = g(X)$ if the PMF of X is given. It is determined using

$$p_Y[y_i] = \sum_{\{j:g(x_j)=y_i\}} p_X[x_j].$$

We need only sum the probabilities of the x_j's that map into y_i. In the case of two discrete random variables X and Y that are transformed into $W = g(X,Y)$ and $Z = h(X,Y)$, we have the similar result

$$p_{W,Z}[w_i, z_j] = \sum\sum_{\left\{(k,l):\substack{g(x_k,y_l)=w_i \\ h(x_k,y_l)=z_j}\right\}} p_{X,Y}[x_k, y_l] \qquad i = 1,2,\ldots,N_W; j = 1,2,\ldots,N_Z \tag{7.12}$$

where N_W and/or N_Z may be infinite. An example follows.

Example 7.6 – Independent Poisson random variables

Assume that the joint PMF is given as the product of the marginal PMFs, where each marginal PMF is a Poisson PMF. Then,

$$p_{X,Y}[k,l] = \exp[-(\lambda_X + \lambda_Y)]\frac{\lambda_X^k \lambda_Y^l}{k!\,l!} \qquad k = 0,1,\ldots; l = 0,1,\ldots \tag{7.13}$$

Note that $X \sim \text{Pois}(\lambda_X)$, $Y \sim \text{Pois}(\lambda_Y)$, and X and Y are independent random variables. Consider the transformation

$$\begin{aligned} W &= g(X,Y) = X \\ Z &= h(X,Y) = X + Y. \end{aligned} \tag{7.14}$$

The possible values of W are those of X, which are $0, 1, \ldots$, and the possible values of Z are also $0, 1, \ldots$. According to (7.12), we need to determine all (k, l) so that

$$\begin{aligned} g(x_k, y_l) &= w_i \\ h(x_k, y_l) &= z_j. \end{aligned} \tag{7.15}$$

But x_k and y_l can be replaced by k and l, respectively, for $k = 0, 1, \ldots$ and $l = 0, 1, \ldots$. Also, w_i and z_j can be replaced by i and j, respectively, for $i = 0, 1, \ldots$ and $j = 0, 1, \ldots$. The transformation equations become

$$\begin{aligned} g(k, l) &= i \\ h(k, l) &= j \end{aligned}$$

which from (7.14) become

$$\begin{aligned} i &= k \\ j &= k + l. \end{aligned}$$

Solving for (k, l) for the *given* (i, j) desired, we have that $k = i$ and $l = j - i \geq 0$, which is the only solution. Note that from (7.13) the joint PMF for X and Y is nonzero only if $l = 0, 1, \ldots$. Therefore, we must have $l \geq 0$ so that $l = j - i \geq 0$. From (7.12) we now have

$$\begin{aligned} p_{W,Z}[i, j] &= \sum_{\substack{k=0 \\ \{(k,l):k=i,l=j-i\geq 0\}}}^{\infty} \sum_{l=0}^{\infty} p_{X,Y}[k, l] \\ &= p_{X,Y}[i, j - i]u[i]u[j - i] \end{aligned} \tag{7.16}$$

where $u[n]$ is the discrete unit step sequence defined as

$$u[n] = \begin{cases} 0 & n = \ldots, -2, -1 \\ 1 & n = 0, 1, \ldots \end{cases}.$$

Finally, we have upon using (7.13)

$$p_{W,Z}[i, j] = \exp[-(\lambda_X + \lambda_Y)] \frac{\lambda_X^i \lambda_Y^{j-i}}{i!(j - i)!} u[i]u[j - i] \tag{7.17}$$

$$= \exp[-(\lambda_X + \lambda_Y)] \frac{\lambda_X^i \lambda_Y^{j-i}}{i!(j - i)!} \quad \begin{array}{l} i = 0, 1, \ldots \\ j = i, i + 1, \ldots \end{array}. \tag{7.18}$$

◊

Use the discrete unit step sequence to avoid mistakes.

As we have seen in the preceding example, the discrete unit step sequence was introduced to designate the region of the w-z plane over which $p_{W,Z}[i,j]$ is nonzero. A common mistake in problems of this type is to disregard this region and assert that the joint PMF given by (7.18) is nonzero over $i = 0, 1, \ldots; j = 0, 1, \ldots$. Note, however, that the transformation will generally change the region over which the new joint PMF is nonzero. It is as important to determine this region as it is to find the analytical form of $p_{W,Z}$. To avoid possible errors it is advisable to replace (7.13) at the outset by

$$p_{X,Y}[k,l] = \exp[-(\lambda_X + \lambda_Y)]\frac{\lambda_X^k \lambda_Y^l}{k!l!}u[k]u[l].$$

Then, the use of the unit step functions will serve to keep track of the nonzero PMF regions before and after the transformation. See also Problem 7.25 for another example.

We sometimes wish to determine the PMF of $Z = h(X,Y)$ only, which is a transformation from (X,Y) to Z. In this case, we can use an *auxiliary* random variable. That is to say, we add another random variable W so that the transformation becomes a transformation from (X,Y) to (W,Z) as before. We can then determine $p_{W,Z}[w_i, z_j]$ by once again using (7.12), and then p_Z, which is the marginal PMF, can be found as

$$p_Z[z_j] = \sum_{\{i:w_i \in \mathcal{S}_W\}} p_{W,Z}[w_i, z_j]. \tag{7.19}$$

As we have seen in the previous example, we will first need to solve (7.15) for x_k and y_l. To facilitate the solution we usually define a simple auxiliary random variable such as $W = X$.

Example 7.7 – PMF for sum of independent Poisson random variables (continuation of previous example)

To find the PMF of $Z = X + Y$ from the joint PMF given by (7.13), we use (7.19) with $W = X$. We then have $\mathcal{S}_W = \mathcal{S}_X = \{0, 1, \ldots\}$ and

$$\begin{aligned} p_Z[j] &= \sum_{i=0}^{\infty} p_{W,Z}[i,j] && \text{(from (7.19))} \quad (7.20) \\ &= \sum_{i=0}^{\infty} \exp[-(\lambda_X + \lambda_Y)]\frac{\lambda_X^i \lambda_Y^{j-i}}{i!(j-i)!}u[i]u[j-i] && \text{(from (7.17))} \end{aligned}$$

and since $u[i] = 1$ for $i = 0, 1, \ldots$ and $u[j-i] = 1$ for $i = 0, 1, \ldots, j$ and $u[j-i] = 0$ for $i > j$, this reduces to

$$p_Z[j] = \sum_{i=0}^{j} \exp[-(\lambda_X + \lambda_Y)] \frac{\lambda_X^i \lambda_Y^{j-i}}{i!(j-i)!} \qquad j = 0, 1, \ldots.$$

Note that Z can take on values $j = 0, 1, \ldots$ since $Z = X + Y$ and both X and Y take on values in $\{0, 1, \ldots\}$. To evaluate this sum we can use the binomial theorem as follows:

$$\begin{aligned} p_Z[j] &= \exp[-(\lambda_X + \lambda_Y)] \frac{1}{j!} \sum_{i=0}^{j} \frac{j!}{(j-i)!i!} \lambda_X^i \lambda_Y^{j-i} \\ &= \exp[-(\lambda_X + \lambda_Y)] \frac{1}{j!} \sum_{i=0}^{j} \binom{j}{i} \lambda_X^i \lambda_Y^{j-i} \\ &= \exp[-(\lambda_X + \lambda_Y)] \frac{1}{j!} (\lambda_X + \lambda_Y)^j \qquad \text{(use binomial theorem)} \\ &= \exp(-\lambda) \frac{\lambda^j}{j!} \qquad \text{(let } \lambda = \lambda_X + \lambda_Y\text{)} \end{aligned}$$

for $j = 0, 1, \ldots$. This is recognized as a Poisson PMF with $\lambda = \lambda_X + \lambda_Y$. By this example then, we have shown that if $X \sim \text{Pois}(\lambda_X)$, $Y \sim \text{Pois}(\lambda_Y)$, and X and Y are independent, then $X + Y \sim \text{Pois}(\lambda_X + \lambda_Y)$. This is called the *reproducing* PMF property. It is also extendible to any number of independent Poisson random variables that are added together.

◇

The formula given by (7.20) when we let $p_{W,Z}[i,j] = p_{X,Y}[i, j-i]$ from (7.16) is valid for the PMF of the sum of any two discrete random variables, whether they are independent or not. Summarizing, if X and Y are random variables that take on integer values from $-\infty$ to $+\infty$, then $Z = X + Y$ has the PMF

$$p_Z[j] = \sum_{i=-\infty}^{\infty} p_{X,Y}[i, j-i]. \tag{7.21}$$

This result says that we should sum all the values of the joint PMF such that the x value, which is i, and the y value, which is $j - i$, sums to the z value of j. In particular, if the random variables are *independent*, then since the joint PMF must factor, we have the result

$$p_Z[j] = \sum_{i=-\infty}^{\infty} p_X[i] p_Y[j-i]. \tag{7.22}$$

But this summation operation is a *discrete convolution* [Jackson 1991]. It is usually written succinctly as $p_Z = p_X \star p_Y$, where $\star$ denotes the convolution operator. This result suggests that the use of Fourier transforms would be a useful tool since a convolution can be converted into a simple multiplication in the Fourier domain. We have already seen in Chapter 6 that the Fourier transform (defined with a $+j$) of a PMF $p_X[k]$ is the characteristic function $\phi_X(\omega) = E[\exp(j\omega X)]$. Therefore, taking the Fourier transform of both sides of (7.22) produces

$$\phi_Z(\omega) = \phi_X(\omega)\phi_Y(\omega) \tag{7.23}$$

and by converting back to the original sequence domain, the PMF becomes

$$p_Z[j] = \mathcal{F}^{-1}\left\{\phi_X(\omega)\phi_Y(\omega)\right\} \tag{7.24}$$

where $\mathcal{F}^{-1}$ denotes the inverse Fourier transform. An example follows.

Example 7.8 – PMF for sum of independent Poisson random variables using characteristic function approach

From Section 6.7 we showed that if $X \sim \text{Pois}(\lambda)$, then

$$\phi_X(\omega) = \exp\left[\lambda(\exp(j\omega) - 1)\right]$$

and thus using (7.23) and (7.24)

$$\begin{aligned} p_Z[j] &= \mathcal{F}^{-1}\left\{\exp\left[\lambda_X(\exp(j\omega) - 1)\right]\exp\left[\lambda_Y(\exp(j\omega) - 1)\right]\right\} \\ &= \mathcal{F}^{-1}\left\{\exp\left[(\lambda_X + \lambda_Y)(\exp(j\omega) - 1)\right]\right\}. \end{aligned}$$

But the characteristic function in the braces is that of a Poisson random variable. Using Property 6.5 we see that $Z \sim \text{Pois}(\lambda_X + \lambda_Y)$. The use of characteristic functions for the determination of the PMF for a sum of independent random variables has *considerably* simplified the derivation.

◇

In summary, if X and Y are independent random variables with integer values, then the PMF of $Z = X + Y$ is given by

$$\begin{aligned} p_Z[k] &= \mathcal{F}^{-1}\left\{\phi_X(\omega)\phi_Y(\omega)\right\} \\ &= \int_{-\pi}^{\pi} \phi_X(\omega)\phi_Y(\omega)\exp(-j\omega k)\frac{d\omega}{2\pi}. \end{aligned} \tag{7.25}$$

When the sample space $\mathcal{S}_{X,Y}$ is finite, it is sometimes possible to obtain the PMF of $Z = g(X,Y)$ by a direct calculation, thus avoiding the need to use (7.19). The latter requires one to first find the transformed joint PMF $p_{W,Z}$. To do so we

1. Determine the finite sample space $\mathcal{S}_Z$.

2. Determine which sample points (x_i, y_j) in $\mathcal{S}_{X,Y}$ map into each $z_k \in \mathcal{S}_Z$.

3. Sum the probabilities of those (x_i, y_j) sample points to yield $p_Z[z_k]$.

Mathematically, this is equivalent to

$$p_Z[z_k] = \sum\sum_{\{(i,j):z_k=g(x_i,y_j)\}} p_{X,Y}[x_i, y_j]. \tag{7.26}$$

An example follows.

Example 7.9 – Direct computation of PMF for transformed random variable, $Z = g(X, Y)$

Consider the transformation of the random vector (X, Y) into the scalar random variable $Z = X^2 + Y^2$. The joint PMF is given by

$$p_{X,Y}[i,j] = \begin{cases} \frac{3}{8} & i = 0, j = 0 \\ \frac{1}{8} & i = 1, j = 0 \\ \frac{1}{8} & i = 0, j = 1 \\ \frac{3}{8} & i = 1, j = 1. \end{cases}$$

To find the PMF for Z first note that (X, Y) takes on the values $(i, j) = (0,0), (1,0), (0,1), (1,1)$. Therefore, Z must take on the values $z_k = i^2 + j^2 = 0, 1, 2$. Then from (7.26)

$$\begin{aligned} p_Z[0] &= \sum\sum_{\{(i,j):0=i^2+j^2\}} p_{X,Y}[i,j] \\ &= \sum_{i=0}^{0}\sum_{j=0}^{0} p_{X,Y}[i,j] \\ &= p_{X,Y}[0,0] = \frac{3}{8} \end{aligned}$$

and similarly

$$\begin{aligned} p_Z[1] &= p_{X,Y}[0,1] + p_{X,Y}[1,0] = \frac{2}{8} \\ p_Z[2] &= p_{X,Y}[1,1] = \frac{3}{8}. \end{aligned}$$

◇

7.7 Expected Values

In addition to determining the PMF of a function of two random variables, we are frequently interested in the average value of that function. Specifically, if $Z = g(X, Y)$, then by definition its expected value is

$$E[Z] = \sum_i z_i p_Z[z_i]. \tag{7.27}$$

To determine $E[Z]$ according to (7.27) we need to first find the PMF of Z and then perform the summation. Alternatively, by a similar derivation to that given in Appendix 6A, we can show that a more direct approach is

$$E[Z] = \sum_i \sum_j g(x_i, y_j) p_{X,Y}[x_i, y_j]. \tag{7.28}$$

To remind us that we are using $p_{X,Y}$ as the averaging PMF, we will modify our previous notation from $E[Z]$ to $E_{X,Y}[Z]$, where of course, Z depends on X and Y. We therefore have the useful result that the expected value of a function of two random variables is

$$E_{X,Y}[g(X,Y)] = \sum_i \sum_j g(x_i, y_j) p_{X,Y}[x_i, y_j]. \tag{7.29}$$

Some examples follow.

Example 7.10 – Expected value of a sum of random variables

If $Z = g(X,Y) = X + Y$, then

$$\begin{aligned} E_{X,Y}[X+Y] &= \sum_i \sum_j (x_i + y_j) p_{X,Y}[x_i, y_j] \\ &= \sum_i \sum_j x_i p_{X,Y}[x_i, y_j] + \sum_i \sum_j y_j p_{X,Y}[x_i, y_j] \\ &= \sum_i x_i \underbrace{\sum_j p_{X,Y}[x_i, y_j]}_{p_X[x_i]} + \sum_j y_j \underbrace{\sum_i p_{X,Y}[x_i, y_j]}_{p_Y[y_j]} \quad \text{(from (7.6))} \\ &= E_X[X] + E_Y[Y] \quad \text{(definition of expected value).} \end{aligned}$$

Hence, the expected value of a sum of random variables is the sum of the expected values. Note that we now use the more descriptive notation $E_X[X]$ to replace $E[X]$ used previously.

◇

Similarly

$$E_{X,Y}[aX + bY] = aE_X[X] + bE_Y[Y]$$

and thus as we have seen previously for a single random variable, *the expectation $E_{X,Y}$ is a linear operation.*

Example 7.11 – Expected value of a product of random variables

If $g(X,Y) = XY$, then

$$E_{X,Y}[XY] = \sum_i \sum_j x_i y_j p_{X,Y}[x_i, y_j].$$

We cannot evaluate this further without specifying $p_{X,Y}$. If, however, X and Y are independent, then since the joint PMF factors, we have

$$\begin{aligned} E_{X,Y}[XY] &= \sum_i \sum_j x_i y_j p_X[x_i] p_Y[y_j] \\ &= \sum_i x_i p_X[x_i] \sum_j y_j p_Y[y_j] \\ &= E_X[X] E_Y[Y]. \end{aligned} \tag{7.30}$$

More generally, we can show by using (7.29) that if X and Y are independent, then (see Problem 7.30)

$$E_{X,Y}[g(X)h(Y)] = E_X[g(X)] E_Y[h(Y)]. \tag{7.31}$$

◇

Example 7.12 – Variance of a sum of random variables

Consider the calculation of $\operatorname{var}(X+Y)$. Then, letting $Z = g(X,Y) = (X + Y - E_{X,Y}[X+Y])^2$, we have

$$\begin{aligned} &\operatorname{var}(X+Y) \\ &= E_Z[Z] && \text{(definition of variance)} \\ &= E_{X,Y}[g(X,Y)] && \text{(from (7.28))} \\ &= E_{X,Y}[(X+Y-E_{X,Y}[X+Y])^2] \\ &= E_{X,Y}[[(X-E_X[X]) + (Y-E_Y[Y])]^2] \\ &= E_{X,Y}[(X-E_X[X])^2 + 2(X-E_X[X])(Y-E_Y[Y]) \\ &\qquad + (Y-E_Y[Y])^2] \\ &= E_X[(X-E_X[X])^2] + 2E_{X,Y}[(X-E_X[X])(Y-E_Y[Y])] \\ &\qquad + E_Y[(Y-E_Y[Y])^2] && \text{(linearity of expectation)} \\ &= \operatorname{var}(X) + 2E_{X,Y}[(X-E_X[X])(Y-E_Y[Y])] + \operatorname{var}(Y) && \text{(definition of variance)} \end{aligned}$$

where we have also used $E_{X,Y}[g(X)] = E_X[g(X)]$ and $E_{X,Y}[h(Y)] = E_Y[h(Y)]$ (see Problem 7.28). The cross-product term is called the *covariance* and is denoted by $\operatorname{cov}(X,Y)$ so that

$$\operatorname{cov}(X,Y) = E_{X,Y}[(X-E_X[X])(Y-E_Y[Y])]. \tag{7.32}$$

Its interpretation is discussed in the next section. Hence, we finally have that the variance of a sum of random variables is

$$\operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X,Y). \tag{7.33}$$

Unlike the expected value or mean, *the variance of a sum is not in general the sum of the variances.* It will only be so when $\text{cov}(X, Y) = 0$. An alternative expression for the covariance is (see Problem 7.34)

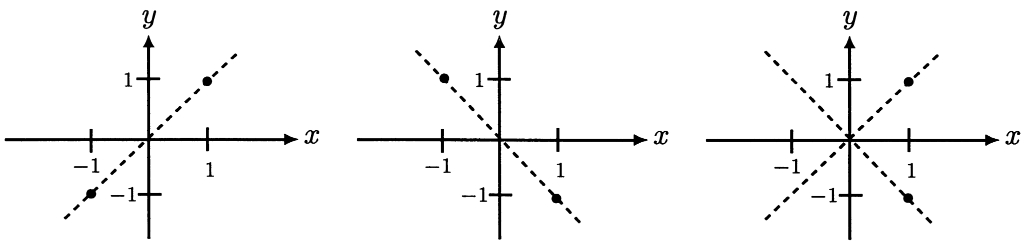
$$\text{cov}(X, Y) = E_{X,Y}[XY] - E_X[X]E_Y[Y] \tag{7.34}$$

which is analogous to Property 6.1 for the variance.

◇

7.8 Joint Moments

Joint PMFs describe the probabilistic behavior of two random variables completely. At times it is important to answer questions such as "If the outcome of one random variable is a given value, what can we say about the outcome of the other random variable? Will it be about the same or have the same magnitude or have no relationship to the other random variable?" For example, in Table 4.1, which lists the joint probabilities of college students having various heights and weights, there is clearly some type of relationship between height and weight. It is our intention to quantify this type of relationship in a succinct and meaningful way as opposed to a listing of probabilities of the various height-weight pairs. The concept of the covariance allows us to accomplish this goal. Note from (7.32) that the covariance is a joint *central* moment. To appreciate the information that it can provide we refer to the three possible joint PMFs depicted in Figure 7.6. The possible values of each joint PMF are shown as solid circles and each possible outcome has a probability of 1/2. In Figure 7.6a if $X = 1$, then $Y = 1$, and if $X = -1$, then $Y = -1$. The relationship



(a) $p_{X,Y}[-1,-1] = p_{X,Y}[1,1] = 1/2$ (b) $p_{X,Y}[1,-1] = p_{X,Y}[-1,1] = 1/2$ (c) $p_{X,Y}[1,1] = p_{X,Y}[1,-1] = 1/2$

Figure 7.6: Joint PMFs depicting different relationships between the random variables X and Y.

is $Y = X$. Note, however, that we cannot determine the value of Y until after the experiment is performed and we are told the value of X. If $X = x_1$, then we know that $Y = X = x_1$. Likewise, in Figure 7.6b we have that $Y = -X$ and so if $X = x_1$, then $Y = -x_1$. However, in Figure 7.6c if $X = 1$, then Y can equal either $+1$ or -1. On the average if $X = 1$, we will have that $Y = 0$ since $Y = \pm 1$ with equal probability. To quantify these relationships we form the product XY, which can take on the values $+1$, -1, and ± 1 for the joint PMFs of Figures 7.6a, 7.6b, and 7.6c, respectively. To determine the value of XY *on the average* we define the *joint moment* as $E_{X,Y}[XY]$. From (7.29) this is evaluated as

$$E_{X,Y}[XY] = \sum_i \sum_j x_i y_j p_{X,Y}[x_i, y_j]. \tag{7.35}$$

The reader should compare the joint moment with the usual moment for a single random variable $E_X[X] = \sum_i x_i p_X[x_i]$. For the joint PMFs of Figure 7.6 the joint moment is

$$\begin{aligned} E_{X,Y}[XY] &= \sum_{i=1}^{2} \sum_{j=1}^{2} x_i y_j p_{X,Y}[x_i, y_j] \\ &= (1)(1)\frac{1}{2} + (-1)(-1)\frac{1}{2} = 1 \qquad \text{(for PMF of Figure 7.6a)} \\ &= (1)(-1)\frac{1}{2} + (-1)(1)\frac{1}{2} = -1 \qquad \text{(for PMF of Figure 7.6b)} \\ &= (1)(-1)\frac{1}{2} + (1)(1)\frac{1}{2} = 0 \qquad \text{(for PMF of Figure 7.6c)} \end{aligned}$$

as we might have expected.

In Figure 7.6a note that $E_X[X] = E_Y[Y] = 0$. If they are not zero, as for the joint PMF shown in Figure 7.7 in which $E_{X,Y}[XY] = 2$, then the joint moment will

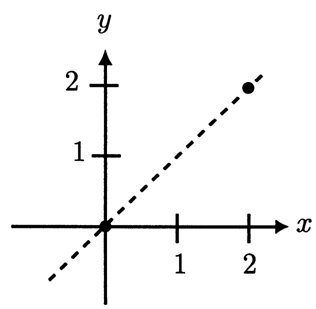


Figure 7.7: Joint PMF for nonzero means with equally probable outcomes.

depend on the values of the means. It is seen that even though the relationship $Y = X$ is preserved, the joint moment has changed. To nullify this effect of having nonzero means influence the joint moment it is more convenient to use the *joint central moment*

$$E_{X,Y}[(X - E_X[X])(Y - E_Y[Y])] \tag{7.36}$$

which will produce the desired +1 for the joint PMF of Figure 7.7. This quantity is recognized as the *covariance* of X and Y so that we denote it by $\text{cov}(X,Y)$. As we have just seen, the covariance may be positive, negative, or zero. Note that the covariance is a measure of how the random variables *co*vary with respect to each other. If they vary in the same direction, i.e., both positive or negative at the same time, then the covariance will be positive. If they vary in opposite directions, the covariance will be negative. This explains why $\text{var}(X+Y)$ may be greater than $\text{var}(X)+\text{var}(Y)$, for the case of a positive covariance. Similarly, the variance of the sum of the random variables will be less than the sum of the variances if the covariance is negative.

If X and Y are independent random variables, then from (7.31) we have

$$\begin{aligned}\text{cov}(X,Y) &= E_{X,Y}[(X-E_X[X])(Y-E_Y[Y])] \\ &= E_X[X-E_X[X]]E_Y[Y-E_Y[Y]] = 0. \end{aligned} \tag{7.37}$$

Hence, independent random variables have a covariance of zero. This also says that for independent random variables the *variance of the sum of random variables is the sum of the variances*, i.e., $\text{var}(X+Y) = \text{var}(X)+\text{var}(Y)$ (see (7.33)). However, the covariance may still be zero even if the random variables are not independent – the converse is not true. Some other properties of the covariance are given in Problem 7.34.

⚠ **Independence implies zero covariance but zero covariance does not imply independence.**

Consider the joint PMF which assigns equal probability of 1/4 to each of the four points shown in Figure 7.8. The joint and marginal PMFs are listed in Table 7.6.

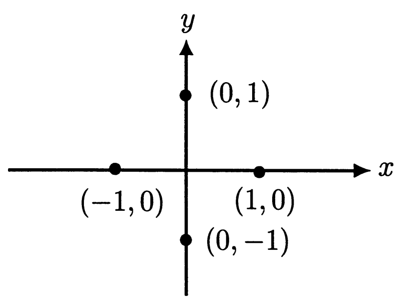


Figure 7.8: Joint PMF of random variables having zero covariance but that are dependent.

For this joint PMF the covariance is zero since

$$E_X[X] = -1\left(\frac{1}{4}\right) + 0\left(\frac{1}{2}\right) + 1\left(\frac{1}{4}\right) = 0$$

	$j=-1$	$j=0$	$j=1$	$p_X[i]$
$i=-1$	0	$\frac{1}{4}$	0	$\frac{1}{4}$
$i=0$	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$
$i=1$	0	$\frac{1}{4}$	0	$\frac{1}{4}$
$p_Y[j]$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	

Table 7.6: Joint PMF values.

and thus from (7.34)

$$\begin{aligned}\operatorname{cov}(X,Y) &= E_{X,Y}[XY] \\ &= \sum_{i=-1}^{1}\sum_{j=-1}^{1} ij p_{X,Y}[i,j] = 0\end{aligned}$$

since either x or y is always zero. However, X are Y are dependent because $p_{X,Y}[1,0] = 1/4$ but $p_X[1]p_Y[0] = (1/4)(1/2) = 1/8$. Alternatively, we may argue that the random variables must be dependent since Y can be predicted from X. For example, if $X = 1$, then surely we must have $Y = 0$.

More generally the joint k-lth moment is defined as

$$E_{X,Y}[X^kY^l] = \sum_i \sum_j x_i^k y_j^l p_{X,Y}[x_i, y_j] \tag{7.38}$$

for $k = 1, 2, \ldots; l = 1, 2, \ldots$, when it exists. The joint k-lth central moment is defined as

$$E_{X,Y}[(X - E_X[X])^k(Y - E_Y[Y])^l] = \sum_i \sum_j (x_i - E_X[X])^k (y_j - E_Y[Y])^l p_{X,Y}[x_i, y_j] \tag{7.39}$$

for $k = 1, 2, \ldots; l = 1, 2, \ldots$, when it exists.

7.9 Prediction of a Random Variable Outcome

The covariance between two random variables has an important bearing on the predictability of Y based on knowledge of the outcome of X. We have already seen in Figures 7.6a,b that Y can be perfectly predicted from X as $Y = X$ (see Figure 7.6a) or as $Y = -X$ (see Figure 7.6b). These are extreme cases. More generally, we seek a predictor of Y that is *linear* (actually affine) in X or

$$\hat{Y} = aX + b$$

where the "hat" indicates an estimator. The constants a and b are to be chosen so that "on the average" the observed value of $\hat{Y}$, which is $ax + b$ if the experimental outcome is (x, y), is close to the observed value of Y, which is y. To determine these constants we shall adopt as our measure of closeness the mean square error (MSE) criterion described previously in Example 6.3. It is given by

$$\text{mse}(a,b) = E_{X,Y}[(Y - \hat{Y})^2]. \tag{7.40}$$

Note that since the predictor $\hat{Y}$ depends on X, we need to average with respect to X and Y. Previously, we let $\hat{Y} = b$, not having the additional information of the outcome of another random variable. It was found in Example 6.3 that the *optimal* value of b, i.e., the value that minimized the MSE, was $b_{\text{opt}} = E_Y[Y]$ and therefore $\hat{Y} = E_Y[Y]$. Now, however, we presume to know the outcome of X. With the additional knowledge of the outcome of X we should be able to find a better predictor. To find the optimal values of a and b we minimize (7.40) over a and b. Before doing so we simplify the expression for the MSE. Starting with (7.40)

$$\begin{aligned}
\text{mse}(a,b) &= E_{X,Y}[(Y - aX - b)^2] \\
&= E_{X,Y}[(Y - aX)^2 - 2b(Y - aX) + b^2] \\
&= E_{X,Y}[Y^2 - 2aXY + a^2X^2 - 2bY + 2abX + b^2] \\
&= E_Y[Y^2] - 2aE_{X,Y}[XY] + a^2E_X[X^2] - 2bE_Y[Y] + 2abE_X[X] + b^2.
\end{aligned}$$

To find the values of a and b that minimize the function $\text{mse}(a,b)$, we determine a stationary point by partial differentiation. Since the function is quadratic in a and b, this will yield the minimizing values of a and b. Using partial differentiation and setting each partial derivative equal to zero produces

$$\begin{aligned}
\frac{\partial \text{mse}(a,b)}{\partial a} &= -2E_{X,Y}[XY] + 2aE_X[X^2] + 2bE_X[X] = 0 \\
\frac{\partial \text{mse}(a,b)}{\partial b} &= -2E_Y[Y] + 2aE_X[X] + 2b = 0
\end{aligned}$$

and rearranging yields the two simultaneous linear equations

$$\begin{aligned}
E_X[X^2]a + E_X[X]b &= E_{X,Y}[XY] \\
E_X[X]a + b &= E_Y[Y].
\end{aligned}$$

The solution is easily shown to be

$$\begin{aligned}
a_{\text{opt}} &= \frac{E_{X,Y}[XY] - E_X[X]E_Y[Y]}{E_X[X^2] - E_X^2[X]} = \frac{\text{cov}(X,Y)}{\text{var}(X)} \\
b_{\text{opt}} &= E_Y[Y] - a_{\text{opt}}E_X[X] = E_Y[Y] - \frac{\text{cov}(X,Y)}{\text{var}(X)}E_X[X]
\end{aligned}$$

so that the optimal linear prediction of Y given the outcome $X = x$ is

$$\begin{aligned}\hat{Y} &= a_{\text{opt}}x + b_{\text{opt}} \\ &= \frac{\text{cov}(X,Y)}{\text{var}(X)}x + E_Y[Y] - \frac{\text{cov}(X,Y)}{\text{var}(X)}E_X[X]\end{aligned}$$

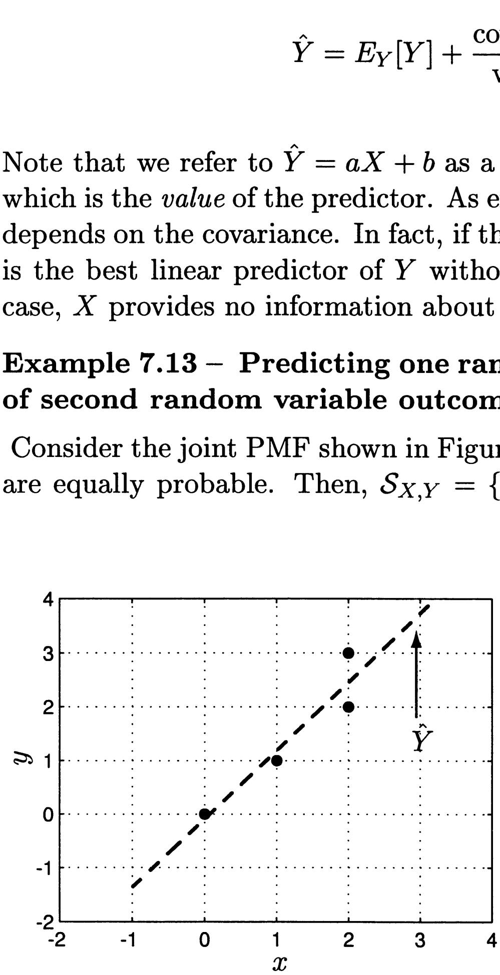
or finally

$$\hat{Y} = E_Y[Y] + \frac{\text{cov}(X,Y)}{\text{var}(X)}(x - E_X[X]). \tag{7.41}$$

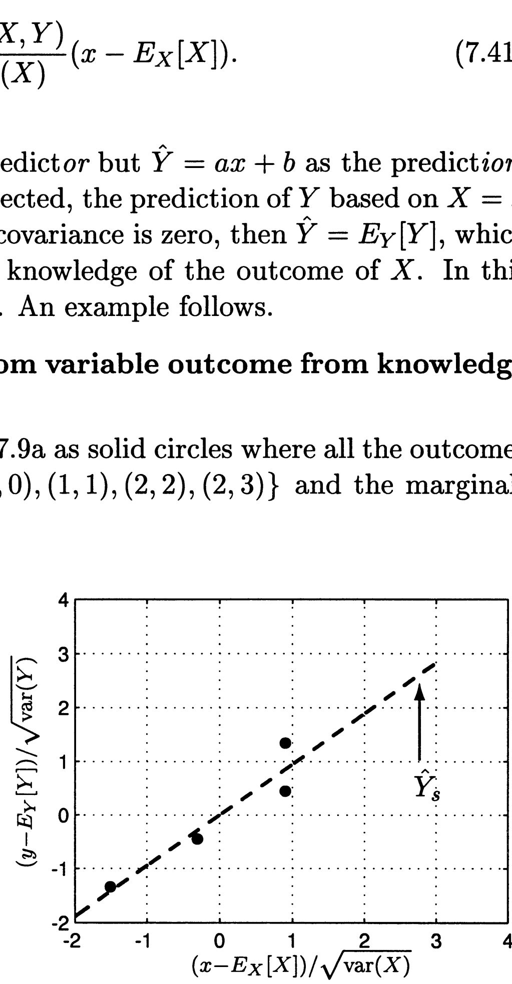
Note that we refer to $\hat{Y} = aX + b$ as a predict*or* but $\hat{Y} = ax + b$ as the predict*ion*, which is the *value* of the predictor. As expected, the prediction of Y based on $X = x$ depends on the covariance. In fact, if the covariance is zero, then $\hat{Y} = E_Y[Y]$, which is the best linear predictor of Y without knowledge of the outcome of X. In this case, X provides no information about Y. An example follows.

Example 7.13 – Predicting one random variable outcome from knowledge of second random variable outcome

Consider the joint PMF shown in Figure 7.9a as solid circles where all the outcomes are equally probable. Then, $\mathcal{S}_{X,Y} = \{(0,0),(1,1),(2,2),(2,3)\}$ and the marginals



(a) Nonstandardized X and Y



(b) Standardized X and Y

Figure 7.9: Joint PMF (shown as solid circles having equal probabilities) and best linear prediction of Y when $X = x$ is observed (shown as dashed line).

are found by summing along each direction to yield

$$p_X[i] = \begin{cases} \frac{1}{4} & i=0 \\ \frac{1}{4} & i=1 \\ \frac{1}{2} & i=2 \end{cases}$$

$$p_Y[j] = \begin{cases} \frac{1}{4} & j=0 \\ \frac{1}{4} & j=1 \\ \frac{1}{4} & j=2 \\ \frac{1}{4} & j=3. \end{cases}$$

As a result, we have from the marginals that $E_X[X] = 5/4$, $E_Y[Y] = 3/2$, $E_X[X^2] = 9/4$, and $\mathrm{var}(X) = E_X[X^2] - E_X^2[X] = 9/4 - (5/4)^2 = 11/16$. From the joint PMF we find that $E_{X,Y}[XY] = (0)(0)1/4 + (1)(1)1/4 + (2)(2)1/4 + (2)(3)1/4 = 11/4$, which results in $\mathrm{cov}(X,Y) = E_{X,Y}[XY] - E_X[X]E_Y[Y] = 11/4 - (5/4)(3/2) = 7/8$. Substituting these values into (7.41) yields the best linear prediction of Y as

$$\begin{aligned} \hat{Y} &= \frac{3}{2} + \frac{7/8}{11/16}\left(x - \frac{5}{4}\right) \\ &= \frac{14}{11}x - \frac{1}{11} \end{aligned}$$

which is shown in Figure 7.9a as the dashed line. The line shown in Figure 7.9a is referred to as a *regression line* in statistics. What do you think would happen if the probability of $(2,3)$ were zero, and the remaining three points had probabilities of 1/3?

◇

The reader should be aware that we could also have predicted X from $Y = y$ by interchanging X and Y in (7.41). Also, we note that if $\mathrm{cov}(X,Y) = 0$, then $\hat{Y} = E_Y[Y]$ or $X = x$ provides no information to help us predict Y. Clearly, this will be the case if X and Y are independent (see (7.37)) since independence of two random variables implies a covariance of zero. However, even if the covariance is zero, the random variables can still be dependent (see Figure 7.8) and so prediction should be possible. This apparent paradox is explained by the fact that in this case we must use a *nonlinear* predictor, not the simple linear function $aX + b$ (see Problem 8.27).

The optimal linear prediction of (7.41) can also be expressed in *standardized form*. A *standardized random variable* is defined to be one for which *the mean is zero and the variance is one*. An example would be a random variable that takes on the values ± 1 with equal probability. Any random variable can be standardized by subtracting the mean and dividing the result by the square root of the variance to form

$$X_s = \frac{X - E_X[X]}{\sqrt{\mathrm{var}(X)}}$$

(see Problem 7.42). For example, if $X \sim \text{Pois}(\lambda)$, then $X_s = (X - \lambda)/\sqrt{\lambda}$, which is easily shown to have a mean of zero and a variance of one. We next seek the best linear prediction of the *standardized* Y based on a *standardized* $X = x$. To do so we define the standardized predictor based on a standardized $X_s = x_s$ as

$$\hat{Y}_s = \frac{\hat{Y} - E_Y[Y]}{\sqrt{\text{var}(Y)}}.$$

Then from (7.41), we have

$$\frac{\hat{Y} - E_Y[Y]}{\sqrt{\text{var}(Y)}} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(Y)\text{var}(X)}}\frac{x - E_X[X]}{\sqrt{\text{var}(X)}}$$

and therefore

$$\hat{Y}_s = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}x_s. \tag{7.42}$$

Example 7.14 – Previous example continued

For the previous example we have that

$$\begin{aligned} x_s &= \frac{x - 5/4}{\sqrt{11/16}} \\ \hat{Y}_s &= \frac{\hat{Y} - 3/2}{\sqrt{5/4}} \end{aligned}$$

and

$$\frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{7/8}{\sqrt{(11/16)(5/4)}} \approx 0.94$$

so that

$$\hat{Y}_s = 0.94x_s$$

and is displayed in Figure 7.9b.

◊

The factor that scales x_s to produce $\hat{Y}_s$ is denoted by

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} \tag{7.43}$$

and is called the *correlation coefficient.* When X and Y have $\rho_{X,Y} \neq 0$, then X and Y are said to be *correlated.* If, however, the covariance is zero and hence $\rho_{X,Y} = 0$, then the random variables are said to be *uncorrelated.* Clearly, independent random variables are always uncorrelated, but not the other way around. Using the correlation coefficient allows us to express the best linear prediction in its standardized form as $\hat{Y}_s = \rho_{X,Y}x_s$. The correlation coefficient has an important property in that it is always less than one in magnitude. In the previous example, we had $\rho_{X,Y} \approx 0.94$.

Property 7.7 – Correlation coefficient is always less than or equal to one in magnitude or $|\rho_{X,Y}| \leq 1$.

Proof: The proof relies on the Cauchy-Schwarz inequality for random variables. This inequality is analogous to the usual one for the dot product of Euclidean vectors $\mathbf{v}$ and $\mathbf{w}$, which is

$$|\mathbf{v} \cdot \mathbf{w}| \leq ||\mathbf{v}|| \, ||\mathbf{w}||$$

where $||\mathbf{v}||$ denotes the length of the vector. Equality holds if and only if the vectors are collinear. Collinearity means that $\mathbf{w} = c\mathbf{v}$ for c a constant or the vectors point in the same direction. For random variables V and W the Cauchy-Schwarz inequality says that

$$|E_{V,W}[VW]| \leq \sqrt{E_V[V^2]}\sqrt{E_W[W^2]} \tag{7.44}$$

with equality if and only if $W = cV$ for c a constant. See Appendix 7A for a derivation. Thus letting $V = X - E_X[X]$ and $W = Y - E_Y[Y]$, we have

$$\begin{aligned} |\rho_{X,Y}| &= \frac{|\text{cov}(X,Y)|}{\sqrt{\text{var}(X)\text{var}(Y)}} \\ &= \frac{|E_{V,W}[VW]|}{\sqrt{E_V[V^2]}\sqrt{E_W[W^2]}} \leq 1 \end{aligned}$$

using (7.44). Equality will hold if and only if $W = cV$ or equivalently if $Y - E_Y[Y] = c(X - E_X[X])$, which is easily shown to imply that (see Problem 7.45)

$$\rho_{X,Y} = \begin{cases} 1 & \text{if } Y = aX + b \text{ with } a > 0 \\ -1 & \text{if } Y = aX + b \text{ with } a < 0 \end{cases}$$

for a and b constants.

□

Note that when $\rho_{X,Y} = \pm 1$, Y can be perfectly predicted from X by using $Y = aX + b$. See also Figures 7.6a and 7.6b for examples of when $\rho_{X,Y} = +1$ and $\rho_{X,Y} = -1$, respectively.

⚠ **Correlation between random variables does not imply a causal relationship between the random variables.**

A frequent misapplication of probability is to assert that two quantities that are correlated ($\rho_{X,Y} \neq 0$) are such because one causes the other. To dispel this myth consider a survey in which all individuals older than 55 years of age in the U.S. are asked whether they have ever had prostate cancer and also their height in inches. Then, for each height in inches we compute the average number of individuals per

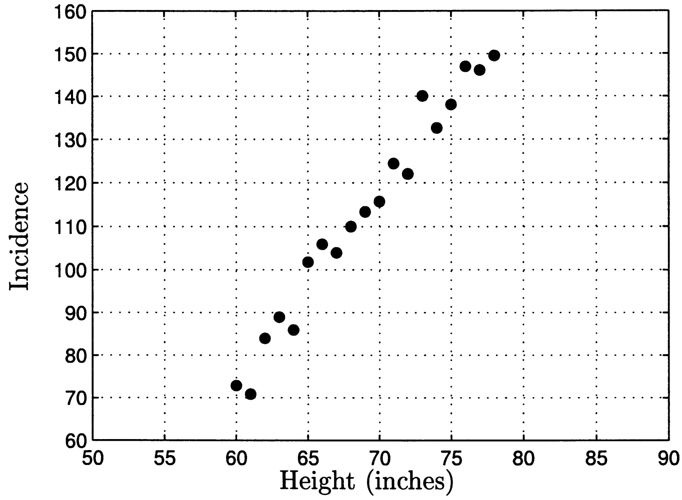


Figure 7.10: Incidence of prostate cancer per 1000 individuals older than age 55 versus height.

1000 who have had cancer. If we plot the average number, also called the incidence of cancer, versus height, a typical result would be as shown in Figure 7.10. This indicates a strong positive correlation of cancer with height. One might be tempted to conclude that growing taller causes prostate cancer. This is of course nonsense. What is actually shown is that segments of the population who are tall are associated with a higher incidence of cancer. This is because the portion of the population of individuals who are taller than the rest are predominately male. Females are not subject to prostate cancer, as they have no prostates! In summary, correlation between two variables only indicates an *association*, i.e., if one increases, then so does the other (if positively correlated). No physical or causal relationship need exist.

7.10 Joint Characteristic Functions

The characteristic function of a discrete random variable was introduced in Section 6.7. For two random variables we can define a *joint* characteristic function. For the random variables X and Y it is defined as

$$\phi_{X,Y}(\omega_X, \omega_Y) = E_{X,Y}[\exp[j(\omega_X X + \omega_Y Y)]]. \tag{7.45}$$

Assuming both random variables take on integer values, it is evaluated using (7.29) as

$$\phi_{X,Y}(\omega_X, \omega_Y) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} p_{X,Y}[k,l] \exp[j(\omega_X k + \omega_Y l)]. \tag{7.46}$$

It is seen to be the two-dimensional Fourier transform of the two-dimensional sequence $p_{X,Y}[k,l]$ (note the use of $+j$ as opposed to the more common $-j$ in the exponential). As in the case of a single random variable, the characteristic function can be used to find moments. In this case, the joint moments are given by the formula

$$E_{X,Y}[X^m Y^n] = \frac{1}{j^{m+n}} \left. \frac{\partial^{m+n} \phi_{X,Y}(\omega_X, \omega_Y)}{\partial \omega_X^m \partial \omega_Y^n} \right|_{\omega_X=\omega_Y=0}. \tag{7.47}$$

In particular, the first joint moment is found as

$$E_{X,Y}[XY] = - \left. \frac{\partial^2 \phi_{X,Y}(\omega_X, \omega_Y)}{\partial \omega_X \partial \omega_Y} \right|_{\omega_X=\omega_Y=0}.$$

Another important application is to finding the PMF for the sum of two independent random variables. This application is based on the result that if X and Y are independent random variables, the joint characteristic function factors due to the property $E_{X,Y}[g(X)h(Y)] = E_X[g(X)]E_Y[h(Y)]$ (see (7.31)). Before deriving the PMF for the sum of two independent random variables, we prove the factorization result, and then give a theoretical application. The factorization of the characteristic function follows as

$$\begin{aligned}
\phi_{X,Y}(\omega_X, \omega_Y) &= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} p_{X,Y}[k,l] \exp[j(\omega_X k + \omega_Y l)] \\
&= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} p_X[k] p_Y[l] \exp[j\omega_X k] \exp[j\omega_Y l] \qquad \text{(joint PMF factors)} \\
&= \sum_{k=-\infty}^{\infty} p_X[k] \exp[j\omega_X k] \sum_{l=-\infty}^{\infty} p_Y[l] \exp[j\omega_Y l] \\
&= \phi_X(\omega_X)\phi_Y(\omega_Y). \qquad \text{(definition of characteristic function for single random variable)}
\end{aligned} \tag{7.48}$$

The converse is also true—if the joint characteristic function factors, then X and Y are independent random variables. This can easily be shown to follow from the inverse Fourier transform relationship. As an application of the converse result, consider the tranformed random variables $W = g(X)$ and $Z = h(Y)$, where X and Y are independent. We prove that W and Z are independent as well, which is to say *functions of independent random variables are independent.* To do so we show that the joint characteristic function factors. The joint characteristic function of the transformed random variables is

$$\phi_{W,Z}(\omega_W, \omega_Z) = E_{W,Z}[\exp[j(\omega_W W + \omega_Z Z)]].$$

But we have that

$$\begin{aligned}
\phi_{W,Z}(\omega_W, \omega_Z) &= E_{X,Y}[\exp[j(\omega_W g(X) + \omega_Z h(Y))]] && \text{(slight extension of (7.28))} \\
&= E_X[\exp(j\omega_W g(X))]E_Y[\exp(j\omega_Z h(Y))] && \text{(same argument as used to yield (7.31))} \\
&= E_W[\exp(j\omega_W W)]E_Z[\exp(j\omega_Z Z)] && \text{(from (6.5))} \\
&= \phi_W(\omega_W)\phi_Z(\omega_Z) && \text{(definition)}
\end{aligned}$$

and hence W and Z are independent random variables. As a general result, we can now assert that *if X and Y are independent random variables, then so are $g(X)$ and $h(Y)$ for any functions g and h.*

Finally, consider the problem of determining the PMF for $Z = X + Y$, where X and Y are independent random variables. We have already solved this problem using the joint PMF approach with the final result given by (7.22). By using characteristic functions, we can simplify the derivation. The derivation proceeds as follows.

$$\begin{aligned}
\phi_Z(\omega_Z) &= E_Z[\exp(j\omega_Z Z)] && \text{(definition)} \\
&= E_{X,Y}[\exp(j\omega_Z(X+Y)] && \text{(from (7.28) and (7.29))} \\
&= E_{X,Y}[\exp(j\omega_Z X)\exp(j\omega_Z Y)] \\
&= E_X[\exp(j\omega_Z X)]E_Y[\exp(j\omega_Z Y)] && \text{(from (7.31))} \\
&= \phi_X(\omega_Z)\phi_Y(\omega_Z).
\end{aligned}$$

To find the PMF we take the inverse Fourier transform of $\phi_Z(\omega_Z)$, replacing ω_Z by the more usual notation ω, to yield

$$\begin{aligned}
p_Z[k] &= \int_{-\pi}^{\pi} \phi_X(\omega)\phi_Y(\omega)\exp(-j\omega k)\frac{d\omega}{2\pi} \\
&= \sum_{i=-\infty}^{\infty} p_X[i]p_Y[k-i]
\end{aligned}$$

which agrees with (7.22). The last result follows from the property that the Fourier transform of a convolution sum is the product of the Fourier transforms of the individual sequences.

7.11 Computer Simulation of Random Vectors

The method of generating realizations of a two-dimensional discrete random vector is nearly identical to the one-dimensional case. In fact, if X and Y are independent, then we generate a realization of X, say x_i, according to $p_X[x_i]$ and a realization of Y, say y_j, according to $p_Y[y_j]$ using the method of Chapter 5. Then we concatenate the realizations together to form the realization of the vector random variable as (x_i, y_j). Furthermore, independence reduces the problems of estimating a joint PMF, a joint CDF, etc. to that of the one-dimensional case. The joint PMF, for example, can be estimated by first estimating $p_X[x_i]$ as $\hat{p}_X[x_i]$, then estimating $p_Y[y_j]$ as $\hat{p}_Y[y_j]$, and finally forming the estimate of the joint PMF as $\hat{p}_{X,Y}[x_i, y_j] = \hat{p}_X[x_i]\hat{p}_Y[y_j]$.

When the random variables are not independent, we need to generate a realization of (X, Y) *simultaneously* since the value obtained for X is dependent on the value obtained for Y and vice versa. If both $\mathcal{S}_X$ and $\mathcal{S}_Y$ are finite, then a simple procedure is to consider each possible realization (x_i, y_j) as a single outcome with probability $p_{X,Y}[x_i, y_j]$. Then, we can apply the techniques of Section 5.9 directly. An example is given next.

Example 7.15 – Generating realizations of jointly distributed random variables

Assume a joint PMF as given in Table 7.7. A simple MATLAB program to generate

	$j=0$	$j=1$
$i=0$	$\frac{1}{8}$	$\frac{1}{8}$
$i=1$	$\frac{1}{4}$	$\frac{1}{2}$

Table 7.7: Joint PMF values for Example 7.15.

a set of M realizations of (X, Y) is given below.

```
for m=1:M
  u=rand(1,1);
  if u<=1/8
    x(m,1)=0;y(m,1)=0;
  elseif u>1/8&u<=1/4
    x(m,1)=0;y(m,1)=1;
  elseif u>1/4&u<=1/2
    x(m,1)=1;y(m,1)=0;
  else
    x(m,1)=1;y(m,1)=1;
  end
end
```

Once the realizations are available we can estimate the joint PMF and marginal

PMFs as

$$\begin{aligned}
\hat{p}_{X,Y}[i,j] &= \frac{\text{Number of outcomes equal to } (i,j)}{M} && i=0,1; j=0,1 \\
\hat{p}_X[i] &= \hat{p}_{X,Y}[i,0] + \hat{p}_{X,Y}[i,1] && i=0,1 \\
\hat{p}_Y[j] &= \hat{p}_{X,Y}[0,j] + \hat{p}_{X,Y}[1,j] && j=0,1
\end{aligned}$$

and the joint moments are estimated as

$$\widehat{E_{X,Y}[X^k Y^l]} = \frac{1}{M}\sum_{m=1}^{M} x_m^k y_m^l$$

where (x_m, y_m) is the mth realization. Other quantities of interest are discussed in Problems 7.49 and 7.51.

◇

7.12 Real-World Example – Assessing Health Risks

An increasingly common health problem in the United States is obesity. It has been found to be associated with many life-threatening illnesses, especially diabetes. One way to define what constitutes an obese person is via the body mass index (BMI) [CDC 2003]. It is computed as

$$\text{BMI} = \frac{703W}{H^2} \tag{7.49}$$

where W is the weight of the person in pounds and H is the person's height in inches. BMIs greater than 25 and less than 30 are considered to indicate an overweight person, and 30 and above an obese person [CDC 2003]. It is of great importance to be able to estimate the PMF of the BMI for a population of people. For example, in Chapter 4 we displayed a table of the joint probabilities of heights and weights for a hypothetical population of college students. For this population we would like to know the probability or percentage of obese persons. This percentage of the population would then be at risk for developing diabetes. To do so we could first determine the PMF of the BMI and then determine the probability of a BMI of 30 and above. From Table 4.1 or Figure 7.1 we have the joint PMF for the random vector (H, W). To find the PMF for the BMI we note that it is a function of H and W or in our previous notation, we wish to determine the PMF of $Z = g(X, Y)$, where Z denotes the BMI, X denotes the height, and Y denotes the weight. The solution follows immediately from (7.26). One slight modification that we must make in order to fit the data of Table 4.1 into our theoretical framework is to replace the height and weight intervals by their midpoint values. For example, in Table 4.1 the probability of observing a person with a height between $5'8''$ and $6'$ and a weight of between 130 and 160 lbs. is 0.06. We convert these intervals so that we can say that the probability of a person having a height of $5'10''$ and a weight of 145 lbs. is 0.06. Next to determine the PMF we first find the BMI for each height and weight using (7.49), rounding the result to the nearest integer. This is displayed in Table 7.8.

	W_1 115	W_2 145	W_3 175	W_4 205	W_5 235
H_1 $5'2''$	21	27	32	37	43
H_2 $5'6''$	19	23	28	33	38
H_3 $5'10''$	16	21	25	29	34
H_4 $6'2''$	15	19	22	26	30
H_5 $6'6''$	13	17	20	24	27

Table 7.8: Body mass indexes for heights and weights of hypothetical college students.

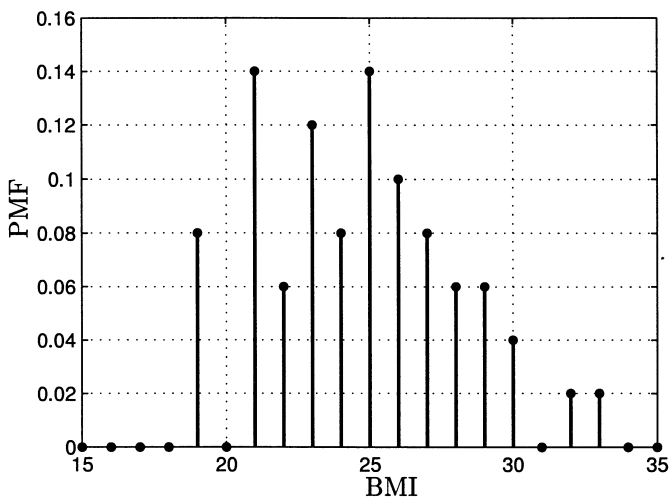


Figure 7.11: Probability mass function for body mass index of hypothetical college population.

Then, we determine the PMF by using (7.26). For example, for a BMI = 21, we require from Table 7.8 the entries $(H, W) = (5'2'',115)$ and $(H, W) = (5'10'',145)$. But from Table 4.1 we see that

$$\begin{aligned} P[H = 5'2'', W = 115] &= 0.08 \\ P[H = 5'10'', W = 145] &= 0.06 \end{aligned}$$

and therefore $P[\text{BMI} = 21] = 0.14$. The other values of the PMF of the BMI are found similarly. This produces the PMF shown in Figure 7.12. It is seen that the probability of being obese as defined by the BMI (BMI ≥ 30) is 0.08. Stated another way 8% of the population of college students are obese and so are at risk for diabetes.

References

CDC, "Nutrition and Physical Activity," Center for Disease Control, http://www.cdc.gov/nccdphp/dnpa/bmi/bmi-adult-formula.htm, 2003.

Jackson, L.B., *Signals, Systems, and Transforms*, Addison-Wesley, Reading, MA, 1991.

Problems

7.1 (w) A chess piece is placed on a chessboard, which consists of an 8×8 array of 64 squares. Specify a numerical sample space $\mathcal{S}_{X,Y}$ for the location of the chess piece.

7.2 (w) Two coins are tossed in succession with a head being mapped into a $+1$ and a tail being mapped into a -1. If a random vector is defined as (X, Y) with X representing the mapping of the first toss and Y representing the mapping of the second toss, draw the mapping. Use Figure 7.2 as a guide. Also, what is $\mathcal{S}_{X,Y}$?

7.3 (⌣) (w) A woman has a penny, a nickel, and a dime in her pocket. If she chooses two coins from her pocket in succession, what is the sample space $\mathcal{S}$ of possible outcomes? If these outcomes are next mapped into the values of the coins, what is the numerical sample space $\mathcal{S}_{X,Y}$?

7.4 (w) If $\mathcal{S}_X = \{1, 2\}$ and $\mathcal{S}_Y = \{3, 4\}$, plot the points in the plane comprising $\mathcal{S}_{X,Y} = \mathcal{S}_X \times \mathcal{S}_Y$. What is the size of $\mathcal{S}_{X,Y}$?

7.5 (w) Two dice are tossed. The number of dots observed on the dice are added together to form the random variable X and also differenced to form Y. Determine the possible outcomes of the random vector (X, Y) and plot them in the plane. How many possible outcomes are there?

7.6 (f) A two-dimensional sequence is given by

$$p_{X,Y}[i, j] = c(1 - p_1)^i (1 - p_2)^j \qquad i = 1, 2, \ldots; j = 1, 2, \ldots$$

where $0 < p_1 < 1$, $0 < p_2 < 1$, and c is a constant. Find c to make $p_{X,Y}$ a valid joint PMF.

7.7 (f) Is

$$p_{X,Y}[i,j] = \left(\frac{1}{2}\right)^{i+j} \qquad i = 0,1,\ldots; j = 0,1,\ldots$$

a valid joint PMF?

7.8 (⌣̈) (w) A single coin is tossed twice. A head outcome is mapped into a 1 and a tail outcome into a 0 to yield a numerical outcome. Next, a random vector (X, Y) is defined as

$$\begin{aligned} X &= \text{outcome of first toss} + \text{outcome of second toss} \\ Y &= \text{outcome of first toss} - \text{outcome of second toss.} \end{aligned}$$

Find the joint PMF for (X, Y), assuming the outcomes (x_i, y_j) are equally likely.

7.9 (f) Find the joint PMF for the experiment described in Example 7.1. Assume each outcome in $\mathcal{S}$ is equally likely. How can you check your answer?

7.10 (⌣̈) (f) The sample space for a random vector is $\mathcal{S}_{X,Y} = \{(i,j) : i = 1,2,3,4,5; j = 1,2,3,4\}$. If the outcomes are equally likely, find $P[(X,Y) \in A]$, where $A = \{(i,j) : 1 \leq i \leq 2; 3 \leq j \leq 4\}$.

7.11 (f) A joint PMF is given as $p_{X,Y}[i,j] = (1/2)^{i+j}$ for $i = 1,2,\ldots; j = 1,2,\ldots$. If $A = \{(i,j) : 1 \leq i \leq 3; j \geq 2\}$, find $P[A]$.

7.12 (f) The values of a joint PMF are given in Table 7.9. Determine the marginal PMFs.

	$j=0$	$j=1$	$j=2$
$i=0$	$\frac{1}{8}$	0	$\frac{1}{4}$
$i=1$	0	$\frac{1}{8}$	$\frac{1}{4}$
$i=2$	$\frac{1}{8}$	0	$\frac{1}{8}$

Table 7.9: Joint PMF values for Problem 7.12.

7.13 (⌣̈) (f) If a joint PMF is given by

$$p_{X,Y}[i,j] = p^2(1-p)^{i+j-2} \qquad i = 1,2,\ldots; j = 1,2,\ldots$$

find the marginal PMFs.

7.14 (f) If a joint PMF is given by $p_{X,Y}[i,j] = 1/36$ for $i = 1,2,3,4,5,6; j = 1,2,3,4,5,6$, find the marginal PMFs.

7.15 (w) A joint PMF is given by

$$p_{X,Y}[i,j] = c\binom{10}{j}\left(\frac{1}{2}\right)^{10} \qquad i = 0,1; j = 0,1,\ldots,10$$

where c is some unknown constant. Find c so that the joint PMF is valid and then determine the marginal PMFs. Hint: Recall the binomial PMF.

7.16 (⌣̈) (w) Find another set of values for the joint PMF that will yield the same marginal PMFs as given in Table 7.2.

7.17 (t) Prove Properties 7.3 and 7.4 for the joint CDF by relying on the standard properties of probabilities of events.

7.18 (w) Sketch the joint CDF for the joint PMF given in Table 7.2. Do this by shading each region in the x-y plane that has the same value.

7.19 (⌣̈) (w) A joint PMF is given by

$$p_{X,Y}[i,j] = \begin{cases} \frac{1}{4} & (i,j) = (0,0) \\ \frac{1}{4} & (i,j) = (1,1) \\ \frac{1}{4} & (i,j) = (1,0) \\ \frac{1}{4} & (i,j) = (1,-1) \end{cases}$$

Are X and Y independent?

7.20 (t) Prove that if the random variables X and Y are independent, then the joint CDF factors as $F_{X,Y}(x,y) = F_X(x)F_Y(y)$.

7.21 (t) If a joint PMF is given by

$$p_{X,Y}[i,j] = \begin{cases} a & (i,j) = (0,0) \\ b & (i,j) = (0,1) \\ c & (i,j) = (1,0) \\ d & (i,j) = (1,1) \end{cases}$$

where of course we must have $a+b+c+d = 1$, show that a necessary condition for the random variables to be independent is $ad = bc$. This can be used to quickly assert that the random variables are not independent as for the case shown in Table 7.5.

7.22 (f) If $X \sim \text{Ber}(p_X)$ and $Y \sim \text{Ber}(p_Y)$, and X and Y are independent, what is the joint PMF?

7.23 (⌣) **(w)** If the joint PMF is given as

$$p_{X,Y}[i,j] = \binom{10}{i}\binom{11}{j}\left(\frac{1}{2}\right)^{21} \qquad i = 0,1,\ldots,10; j = 0,1,\ldots,11$$

are X and Y independent? What are the marginal PMFs?

7.24 (t) Assume that X and Y are discrete random variables that take on all integer values and are independent. Prove that the PMF of $Z = X - Y$ is given by

$$p_Z[l] = \sum_{k=-\infty}^{\infty} p_X[k]p_Y[k-l] \qquad l = \ldots,-1,0,1,\ldots$$

by following the same procedure as was used to derive (7.22). Note that the transformation from (X,Y) to (W,Z) is one-to-one. Next show that if X and Y take on nonnegative integer values only, then

$$p_Z[l] = \sum_{k=\max(0,l)}^{\infty} p_X[k]p_Y[k-l] \qquad l = \ldots,-1,0,1,\ldots \quad .$$

7.25 (f) Using the result of Problem 7.24 find the PMF for $Z = X - Y$ if $X \sim \text{Pois}(\lambda_X)$, $Y \sim \text{Pois}(\lambda_Y)$, and X and Y are independent. Hint: The result will be in the form of infinite sums.

7.26 (w) Find the PMF for $Z = \max(X,Y)$ if the joint PMF is given in Table 7.5.

7.27 (⌣) **(f)** If $X \sim \text{Ber}(1/2)$, $Y \sim \text{Ber}(1/2)$, and X and Y are independent, find the PMF for $Z = X + Y$. Why does the width of the PMF increase? Does the variance increase?

7.28 (t) Prove that $E_{X,Y}[g(X)] = E_X[g(X)]$. Do X and Y have to be independent?

7.29 (t) Prove that

$$E_{X,Y}[ag(X) + bh(Y)] = aE_X[g(X)] + bE_Y[h(Y)].$$

7.30 (t) Prove (7.31).

7.31 (t) Find a formula for $\text{var}(X - Y)$ similar to (7.33). What can you say about the relationship between $\text{var}(X + Y)$ and $\text{var}(X - Y)$ if X and Y are uncorrelated?

7.32 (f) Find the covariance for the joint PMF given in Table 7.4. How do you know the value that you obtained is correct?

7.33 (⌣) **(f)** Find the covariance for the joint PMF given in Table 7.5.

7.34 (t) Prove the following properties of the covariance:

$$\begin{aligned}
\text{cov}(X,Y) &= E_{X,Y}[XY] - E_X[X]E_Y[Y] \\
\text{cov}(X,X) &= \text{var}(X) \\
\text{cov}(Y,X) &= \text{cov}(X,Y) \\
\text{cov}(cX,Y) &= c\,[\text{cov}(X,Y)] \\
\text{cov}(X,cY) &= c\,[\text{cov}(X,Y)] \\
\text{cov}(X,X+Y) &= \text{cov}(X,X) + \text{cov}(X,Y) \\
\text{cov}(X+Y,X) &= \text{cov}(X,X) + \text{cov}(Y,X)
\end{aligned}$$

for c a constant.

7.35 (t) If X and Y have a covariance of $\text{cov}(X,Y)$, we can transform them to a new pair of random variables whose covariance is zero. To do so we let

$$\begin{aligned}
W &= X \\
Z &= aX + Y
\end{aligned}$$

where $a = -\text{cov}(X,Y)/\text{var}(X)$. Show that $\text{cov}(W,Z) = 0$. This process is called *decorrelating the random variables.* See also Example 9.4 for another method.

7.36 (f) Apply the results of Problem 7.35 to the joint PMF given in Table 7.5. Verify by direct calculation that $\text{cov}(W,Z) = 0$.

7.37 (⌣̈) (f) If the joint PMF is given as

$$p_{X,Y}[i,j] = \left(\frac{1}{2}\right)^{i+j} \qquad i = 1,2,\ldots; j = 1,2,\ldots$$

compute the covariance.

7.38 (⌣̈) (f) Determine the minimum mean square error for the joint PMF shown in Figure 7.9a. You will need to evaluate $E_{X,Y}[(Y - ((14/11)X - 1/11))^2]$.

7.39 (t,f) Prove that the minimum mean square error of the optimal linear predictor is given by

$$\text{mse}_{\text{min}} = E_{X,Y}[(Y - (a_{\text{opt}}X + b_{\text{opt}}))^2] = \text{var}(Y)\left(1 - \rho^2_{X,Y}\right).$$

Use this formula to check your result for Problem 7.38.

7.40 (⌣̈) (w) In this problem we compare the prediction of a random variable with and without the knowledge of a second random variable outcome. Consider the joint PMF shown below. First determine the optimal linear prediction of Y

	$j=0$	$j=1$
$i=0$	$\frac{1}{8}$	$\frac{1}{4}$
$i=1$	$\frac{1}{4}$	$\frac{3}{8}$

Table 7.10: Joint PMF values for Problem 7.40.

without any knowledge of the outcome of X (see Section 6.6). Also, compute the minimum mean square error. Next determine the optimal linear prediction of Y based on the knowledge that $X = x$ and compute the minimum mean square error. Plot the predictions versus x in the plane. How do the minimum mean square errors compare?

7.41 (⌣) **(w,c)** For the joint PMF of height and weight shown in Figure 7.1 determine the best linear prediction of weight based on a knowledge of height. You will need to use Table 4.1 as well as a computer to carry out this problem. Does your answer seem reasonable? Is your prediction of a person's weight if the height is 70 inches reasonable? How about if the height is 78 inches? Can you explain the difference?

7.42 (f) Prove that the transformed random variable

$$\frac{X - E_X[X]}{\sqrt{\text{var}(X)}}$$

has an expected value of 0 and a variance of 1.

7.43 (⌣) **(w)** The linear prediction of one random variable based on the outcome of another becomes more difficult if noise is present. We model noise as the addition of an uncorrelated random variable. Specifically, assume that we wish to predict X based on observing $X + N$, where N represents the noise. If X and N are both zero mean random variables that are uncorrelated with each other, determine the correlation coefficient between $W = X$ and $Z = X + N$. How does it depend on the power in X, which is defined as $E_X[X^2]$, and the power in N, also defined as $E_N[N^2]$?

7.44 (w) Consider $\text{var}(X + Y)$, where X and Y are correlated random variables. How is the variance of a sum of random variables affected by the correlation between the random variables? Hint: Express the variance of the sum of the random variables using the correlation coefficient.

7.45 (f) Prove that if $Y = aX + b$, where a and b are constants, then $\rho_{X,Y} = 1$ if $a > 0$ and $\rho_{X,Y} = -1$ if $a < 0$.

7.46 (⌣̈) **(w)** If $X \sim \text{Ber}(1/2)$, $Y \sim \text{Ber}(1/2)$, and X and Y are independent, find the PMF for $Z = X + Y$. Use the characteristic function approach to do so. Compare your results to that of Problem 7.27.

7.47 (w) Using characteristic functions prove that the binomial PMF has the reproducing property. That is to say, if $X \sim \text{bin}(M_X, p)$, $Y \sim \text{bin}(M_Y, p)$, and X and Y are independent, then $Z = X + Y \sim \text{bin}(M_X + M_Y, p)$. Why does this make sense in light of the fact that a sequence of independent Bernoulli trials can be used to derive the binomial PMF?

7.48 (⌣̈) **(c)** Using the joint PMF shown in Table 7.7 generate realizations of the random vector (X, Y) and estimate its joint and marginal PMFs. Compare your estimated results to the true values.

7.49 (⌣̈) **(c)** For the joint PMF shown in Table 7.7 determine the correlation coefficient. Next use a computer simulation to generate realizations of the random vector (X, Y) and estimate the correlation coefficient as

$$\hat{\rho}_{X,Y} = \frac{\frac{1}{M}\sum_{m=1}^{M} x_m y_m - \bar{x}\bar{y}}{\sqrt{\left(\frac{1}{M}\sum_{m=1}^{M} x_m^2 - \bar{x}^2\right)\left(\frac{1}{M}\sum_{m=1}^{M} y_m^2 - \bar{y}^2\right)}}$$

where

$$\begin{aligned} \bar{x} &= \frac{1}{M}\sum_{m=1}^{M} x_m \\ \bar{y} &= \frac{1}{M}\sum_{m=1}^{M} y_m \end{aligned}$$

and (x_m, y_m) is the mth realization.

7.50 (w,c) If $X \sim \text{geom}(p)$, $Y \sim \text{geom}(p)$, and X and Y are independent, show that the PMF of $Z = X + Y$ is given by

$$p_Z[k] = p^2(k-1)(1-p)^{k-2} \qquad k = 2, 3, \ldots.$$

To avoid errors use the discrete unit step sequence. Next, for $p = 1/2$ generate realizations of Z by first generating realizations of X, then generating realizations of Y and adding each pair of realizations together. Estimate the PMF of Z and compare it to the true PMF.

7.51 (w,c) Using the joint PMF given in Table 7.5 determine the covariance to show that it is nonzero and hence X and Y are correlated. Next use the procedure of Problem 7.35 to determine transformed random variables W and

Z that are uncorrelated. Verify that W and Z are uncorrelated by estimating the covariance as

$$\widehat{\operatorname{cov}(W,Z)} = \frac{1}{M}\sum_{m=1}^{M} w_m z_m - \bar{w}\bar{z}$$

where

$$\begin{aligned}\bar{w} &= \frac{1}{M}\sum_{m=1}^{M} w_m \\ \bar{z} &= \frac{1}{M}\sum_{m=1}^{M} z_m\end{aligned}$$

and (w_m, z_m) is the mth realization. Be sure to generate the realizations of W and Z as $w_m = x_m$ and $z_m = ax_m + y_m$, where (x_m, y_m) is the mth realization of (X, Y).

Appendix 7A

Derivation of the Cauchy-Schwarz Inequality

The Cauchy-Schwarz inequality was given by

$$|E_{V,W}[VW]| \leq \sqrt{E_V[V^2]}\sqrt{E_W[W^2]} \tag{7A.1}$$

with equality holding if and only if $W = cV$, for c a constant. To prove this, we first note that for all $\alpha \neq 0$ and $\beta \neq 0$

$$E_{V,W}[(\alpha V - \beta W)^2] \geq 0. \tag{7A.2}$$

If we let

$$\begin{aligned} \alpha &= \sqrt{E_W[W^2]} \\ \beta &= \sqrt{E_V[V^2]} \end{aligned}$$

then we have that

$$\begin{aligned} E_{V,W}[(\sqrt{E_W[W^2]}V - \sqrt{E_V[V^2]}W)^2] &\geq 0 \\ E_{V,W}[E_W[W^2]V^2 - 2\sqrt{E_W[W^2]}\sqrt{E_V[V^2]}VW + E_V[V^2]W^2] &\geq 0 \\ E_W[W^2]E_V[V^2] - 2\sqrt{E_W[W^2]}\sqrt{E_V[V^2]}E_{V,W}[VW] + E_V[V^2]E_W[W^2] &\geq 0 \end{aligned}$$

since $E_{V,W}[g(W)] = E_W[g(W)]$, etc. , which results in

$$E_W[W^2]E_V[V^2] - \sqrt{E_W[W^2]}\sqrt{E_V[V^2]}E_{V,W}[VW] \geq 0.$$

Dividing by $E_W[W^2]E_V[V^2]$ produces

$$1 - \frac{E_{V,W}[VW]}{\sqrt{E_W[W^2]}\sqrt{E_V[V^2]}} \geq 0$$

or finally, upon rearranging terms we have that

$$\frac{E_{V,W}[VW]}{\sqrt{E_V[V^2]}\sqrt{E_W[W^2]}} \leq 1$$

or

$$E_{V,W}[VW] \leq \sqrt{E_V[V^2]}\sqrt{E_W[W^2]}.$$

By replacing the negative sign in (7A.2) by a positive sign and proceeding in an identical manner, we will obtain

$$-E_{V,W}[VW] \leq \sqrt{E_V[V^2]}\sqrt{E_W[W^2]}$$

and hence combining the two results yields the desired inequality. To determine when the equal sign will hold, we note that

$$E_{V,W}[(\alpha V - \beta W)^2] = \sum_{v_i}\sum_{w_j}(\alpha v_i - \beta w_j)^2 p_{V,W}[v_i, w_j]$$

which can only equal zero when $(\alpha v_i - \beta w_j)^2 = 0$ for all i and j since $p_{V,W}[v_i, w_j] > 0$. Thus, for equality to hold we must have

$$\alpha v_i = \beta w_j \qquad \text{all } i \text{ and } j$$

which is equivalent to requiring

$$\alpha V = \beta W$$

or finally dividing by β (asssumed not equal to zero), we obtain the condition for equality as

$$W = \frac{\alpha}{\beta}V = cV$$

for c a constant.