2

Discrete Random Variables

2.1 Definitions

Chapter 1 defines a probability model. It begins with a *physical* model of an experiment. An experiment consists of a procedure and observations. The set of all possible observations, S, is the sample space of the experiment. S is the beginning of the *mathematical* probability model. In addition to S, the mathematical model includes a rule for assigning numbers between 0 and 1 to sets A in S. Thus for every $A \subset S$, the model gives us a probability $P[A]$, where $0 \leq P[A] \leq 1$.

In this chapter and for most of the remainder of the course, we will examine probability models that assign numbers to the outcomes in the sample space. When we observe one of these numbers, we refer to the observation as a *random variable*. In our notation, the name of a random variable is always a capital letter, for example, X. The set of possible values of X is the *range* of X. Since we often consider more than one random variable at a time, we denote the range of a random variable by the letter S with a subscript which is the name of the random variable. Thus S_X is the range of random variable X, S_Y is the range of random variable Y, and so forth. We use S_X to denote the range of X because the set of all possible values of X is analogous to S, the set of all possible outcomes of an experiment.

A probability model always begins with an experiment. Each random variable is related directly to this experiment. There are three types of relationships.

1. The random variable is the observation.

Example 2.1 The experiment is to attach a photo detector to an optical fiber and count the number of photons arriving in a one microsecond time interval. Each observation is a random variable X. The range of X is $S_X = \{0, 1, 2, \ldots\}$. In this case, S_X, the range of X, and the sample space S are identical.

2. The random variable is a function of the observation.

Example 2.2 The experiment is to test six integrated circuits and after each test observe whether the circuit is accepted (a) or rejected (r). Each observation is a sequence of six letters where each letter is either a or r. For example, $s_8 =$ aararaa. The sample space S consists of the 64 possible sequences. A random variable related to this experiment is N, the number of accepted circuits. For outcome s_8, $N = 5$ circuits are accepted. The range of N is $S_N = \{0, 1, \ldots, 6\}$.

3. The random variable is a function of another random variable.

Example 2.3 In Example 2.2, the net revenue R obtained for a batch of six integrated circuits is \$5 for each circuit accepted minus \$7 for each circuit rejected. (This is because for each bad circuit that goes out of the factory, it will cost the company \$7 to deal with the customer's complaint and supply a good replacement circuit.) When N circuits are accepted, $6 - N$ circuits are rejected so that the net revenue R is related to N by the function

$$R = g(N) = 5N - 7(6 - N) = 12N - 42 \text{ dollars.} \tag{2.1}$$

Since $S_N = \{0, \ldots, 6\}$, the range of R is

$$S_R = \{-42, -30, -18, -6, 6, 18, 30\}. \tag{2.2}$$

If we have a probability model for the integrated circuit experiment in Example 2.2, we can use that probability model to obtain a probability model for the random variable. The remainder of this chapter will develop methods to characterize probability models for random variables. We observe that in the preceding examples, the value of a random variable can always be derived from the outcome of the underlying experiment. This is not a coincidence. The formal definition of a random variable reflects this fact.

Definition 2.1 ***Random Variable***

A ***random variable*** *consists of an experiment with a probability measure* $P[\cdot]$ *defined on a sample space* S *and a function that assigns a real number to each outcome in the sample space of the experiment.*

This definition acknowledges that a random variable is the result of an underlying experiment, but it also permits us to separate the experiment, in particular, the observations, from the process of assigning numbers to outcomes. As we saw in Example 2.1, the assignment may be implicit in the definition of the experiment, or it may require further analysis.

In some definitions of experiments, the procedures contain variable parameters. In these experiments, there can be values of the parameters for which it is impossible to perform the observations specified in the experiments. In these cases, the experiments do not produce random variables. We refer to experiments with parameter settings that do not produce random variables as *improper experiments*.

Example 2.4 The procedure of an experiment is to fire a rocket in a vertical direction from the Earth's surface with initial velocity V km/h. The observation is T seconds, the time elapsed until the rocket returns to Earth. Under what conditions is the experiment improper?

....

At low velocities, V, the rocket will return to Earth at a random time T seconds that depends on atmospheric conditions and small details of the rocket's shape and weight. However, when $V > v^* \approx 40{,}000$ km/hr, the rocket will not return to Earth. Thus, the experiment is improper when $V > v^*$ because it is impossible to perform the specified observation.

On occasion, it is important to identify the random variable X by the function $X(s)$ that maps the sample outcome s to the corresponding value of the random variable X. As needed, we will write $\{X = x\}$ to emphasize that there is a set of sample points $s \in S$ for which $X(s) = x$. That is, we have adopted the shorthand notation

$$\{X = x\} = \{s \in S | X(s) = x\} \tag{2.3}$$

Here are some more random variables:

- A, the number of students asleep in the next probability lecture;
- C, the number of phone calls you answer in the next hour;
- M, the number of minutes you wait until you next answer the phone.

Random variables A and C are *discrete* random variables. The possible values of these random variables form a countable set. The underlying experiments have sample spaces that are discrete. The random variable M can be any nonnegative real number. It is a *continuous random variable*. Its experiment has a continuous sample space. In this chapter, we study the properties of discrete random variables. Chapter 3 covers continuous random variables.

Definition 2.2 ***Discrete Random Variable***

*X is a **discrete** random variable if the range of X is a countable set*

$$S_X = \{x_1, x_2, \ldots\}.$$

The defining characteristic of a discrete random variable is that the set of possible values can (in principle) be listed, even though the list may be infinitely long. By contrast, a random variable Y that can take on *any* real number y in an interval $a \leq y \leq b$ is a *continuous random variable*.

Definition 2.3 ***Finite Random Variable***

*X is a **finite random variable** if the range is a finite set*

$$S_X = \{x_1, x_2, \cdots, x_n\}.$$

Often, but not always, a discrete random variable takes on integer values. An exception is the random variable related to your probability grade. The experiment is to take this course and observe your grade. At Rutgers, the sample space is

$$S = \left\{F, D, C, C^+, B, B^+, A\right\}. \tag{2.4}$$

The function $G(\cdot)$ that transforms this sample space into a random variable, G, is

$$\begin{array}{llll} G(F)=0, & G(C)=2, & G(B)=3, & G(A)=4, \\ G(D)=1, & G(C^+)=2.5, & G(B^+)=3.5. & \end{array} \tag{2.5}$$

G is a finite random variable. Its values are in the set $S_G = \{0, 1, 2, 2.5, 3, 3.5, 4\}$. Have you thought about why we transform letter grades to numerical values? We believe the principal reason is that it allows us to compute averages. In general, this is also the main reason for introducing the concept of a random variable. Unlike probability models defined on arbitrary sample spaces, random variables allow us to compute averages. In the mathematics of probability, averages are called *expectations* or *expected values* of random variables. We introduce expected values formally in Section 2.5.

Example 2.5 Suppose we observe three calls at a telephone switch where voice calls (v) and data calls (d) are equally likely. Let X denote the number of voice calls, Y the number of data calls, and let $R = XY$. The sample space of the experiment and the corresponding values of the random variables X, Y, and R are

	Outcomes	ddd	ddv	dvd	dvv	vdd	vdv	vvd	vvv
	$P[\cdot]$	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
Random	X	0	1	1	2	1	2	2	3
Variables	Y	3	2	2	1	2	1	1	0
	R	0	2	2	2	2	2	2	0

Quiz 2.1 *A student takes two courses. In each course, the student will earn a B with probability* 0.6 *or a C with probability* 0.4, *independent of the other course. To calculate a grade point average (GPA), a B is worth 3 points and a C is worth 2 points. The student's GPA is the sum of the GPA for each course divided by 2. Make a table of the sample space of the experiment and the corresponding values of the student's GPA, G.*

2.2 Probability Mass Function

Recall that a discrete probability model assigns a number between 0 and 1 to each outcome in a sample space. When we have a discrete random variable X, we express the probability model as a probability mass function (PMF) $P_X(x)$. The argument of a PMF ranges over all real numbers.

Definition 2.4 ***Probability Mass Function (PMF)***
*The **probability mass function** (PMF) of the discrete random variable X is*

$$P_X(x) = P[X = x]$$

Note that $X = x$ is an event consisting of all outcomes s of the underlying experiment for which $X(s) = x$. On the other hand, $P_X(x)$ is a function ranging over all real numbers x. For any value of x, the function $P_X(x)$ is the probability of the event $X = x$.

Observe our notation for a random variable and its PMF. We use an uppercase letter (X in the preceding definition) for the name of a random variable. We usually use the corresponding lowercase letter (x) to denote a possible value of the random variable. The notation for the PMF is the letter P with a subscript indicating the name of the random variable. Thus $P_R(r)$ is the notation for the PMF of random variable R. In these examples, r and x are just dummy variables. The same random variables and PMFs could be denoted $P_R(u)$ and $P_X(u)$ or, indeed, $P_R(\cdot)$ and $P_X(\cdot)$.

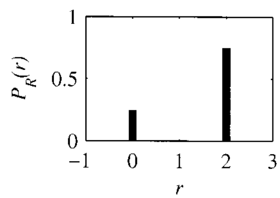
We graph a PMF by marking on the horizontal axis each value with nonzero probability and drawing a vertical bar with length proportional to the probability.

Example 2.6 From Example 2.5, what is the PMF of R?

From Example 2.5, we see that $R = 0$ if either outcome, DDD or VVV, occurs so that

$$P[R = 0] = P[DDD] + P[VVV] = 1/4. \tag{2.6}$$

For the other six outcomes of the experiment, $R = 2$ so that $P[R = 2] = 6/8$. The PMF of R is



$$P_R(r) = \begin{cases} 1/4 & r = 0, \\ 3/4 & r = 2, \\ 0 & \text{otherwise}. \end{cases} \tag{2.7}$$

Note that the PMF of R states the value of $P_R(r)$ for every real number r. The first two lines of Equation (2.7) give the function for the values of R associated with nonzero probabilities: $r = 0$ and $r = 2$. The final line is necessary to specify the function at all other numbers. Although it may look silly to see "$P_R(r) = 0$ otherwise" appended to almost every expression of a PMF, it is an essential part of the PMF. It is helpful to keep this part of the definition in mind when working with the PMF. Do not omit this line in your expressions of PMFs.

Example 2.7 When the basketball player Wilt Chamberlain shot two free throws, each shot was equally likely either to be good (g) or bad (b). Each shot that was good was worth 1 point. What is the PMF of X, the number of points that he scored?

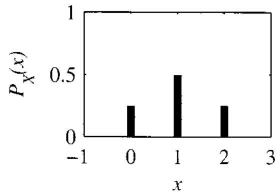
There are four outcomes of this experiment: gg, gb, bg, and bb. A simple tree diagram indicates that each outcome has probability $1/4$. The random variable X has three possible values corresponding to three events:

$$\{X = 0\} = \{bb\}, \qquad \{X = 1\} = \{gb, bg\}, \qquad \{X = 2\} = \{gg\}. \tag{2.8}$$

Since each outcome has probability $1/4$, these three events have probabilities

$$P[X = 0] = 1/4, \qquad P[X = 1] = 1/2, \qquad P[X = 2] = 1/4. \tag{2.9}$$

We can express the probabilities of these events as the probability mass function



$$P_X(x) = \begin{cases} 1/4 & x = 0, \\ 1/2 & x = 1, \\ 1/4 & x = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{2.10}$$

The PMF contains all of our information about the random variable X. Because $P_X(x)$ is the probability of the event $\{X = x\}$, $P_X(x)$ has a number of important properties. The following theorem applies the three axioms of probability to discrete random variables.

Theorem 2.1 *For a discrete random variable X with PMF $P_X(x)$ and range S_X:*

(a) For any x, $P_X(x) \geq 0$.

(b) $\sum_{x \in S_X} P_X(x) = 1$.

(c) For any event $B \subset S_X$, the probability that X is in the set B is

$$P[B] = \sum_{x \in B} P_X(x).$$

Proof All three properties are consequences of the axioms of probability (Section 1.3). First, $P_X(x) \geq 0$ since $P_X(x) = P[X = x]$. Next, we observe that every outcome $s \in S$ is associated with a number $x \in S_X$. Therefore, $P[x \in S_X] = \sum_{x \in S_X} P_X(x) = P[s \in S] = P[S] = 1$. Since the events $\{X = x\}$ and $\{X = y\}$ are disjoint when $x \neq y$, B can be written as the union of disjoint events $B = \bigcup_{x \in B} \{X = x\}$. Thus we can use Axiom 3 (if B is countably infinite) or Theorem 1.4 (if B is finite) to write

$$P[B] = \sum_{x \in B} P[X = x] = \sum_{x \in B} P_X(x). \tag{2.11}$$

Quiz 2.2 *The random variable N has PMF*

$$P_N(n) = \begin{cases} c/n & n = 1, 2, 3, \\ 0 & \text{otherwise.} \end{cases} \tag{2.12}$$

Find

(1) The value of the constant c	*(2) $P[N = 1]$*
(3) $P[N \geq 2]$	*(4) $P[N > 3]$*

2.3 Families of Discrete Random Variables

Thus far in our discussion of random variables we have described how each random variable is related to the outcomes of an experiment. We have also introduced the probability mass function, which contains the probability model of the experiment. In practical applications, certain families of random variables appear over and over again in many experiments. In each family, the probability mass functions of all the random variables have the same mathematical form. They differ only in the values of one or two parameters. This enables us to study in advance each family of random variables and later apply the knowledge we gain to specific practical applications. In this section, we define six families of discrete random variables. There is one formula for the PMF of all the random variables in a family. Depending on the family, the PMF formula contains one or two parameters. By assigning numerical values to the parameters, we obtain a specific random variable. Our nomenclature for a family consists of the family name followed by one or two parameters in parentheses. For example, *binomial* (n, p) refers in general to the family of binomial random variables. *Binomial* $(7, 0.1)$ refers to the binomial random variable with parameters $n = 7$ and $p = 0.1$. Appendix A summarizes important properties of 17 families of random variables.

Example 2.8 Consider the following experiments:

- Flip a coin and let it land on a table. Observe whether the side facing up is heads or tails. Let X be the number of heads observed.
- Select a student at random and find out her telephone number. Let $X = 0$ if the last digit is even. Otherwise, let $X = 1$.
- Observe one bit transmitted by a modem that is downloading a file from the Internet. Let X be the value of the bit (0 or 1).

All three experiments lead to the probability mass function

$$P_X(x) = \begin{cases} 1/2 & x = 0, \\ 1/2 & x = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.13}$$

Because all three experiments lead to the same probability mass function, they can all be analyzed the same way. The PMF in Example 2.8 is a member of the family of *Bernoulli* random variables.

Definition 2.5 ***Bernoulli*** (p) ***Random Variable***

*X is a **Bernoulli** (p) random variable if the PMF of X has the form*

$$P_X(x) = \begin{cases} 1-p & x = 0 \\ p & x = 1 \\ 0 & \textit{otherwise} \end{cases}$$

where the parameter p is in the range $0 < p < 1$.

In the following examples, we use an integrated circuit test procedure to represent any experiment with two possible outcomes. In this particular experiment, the outcome r, that a circuit is a reject, occurs with probability p. Some simple experiments that involve tests of integrated circuits will lead us to the *Bernoulli*, *binomial*, *geometric*, and *Pascal* random variables. Other experiments produce *discrete uniform* random variables and

Poisson random variables. These six families of random variables occur often in practical applications.

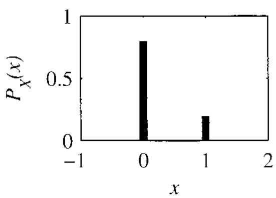
Example 2.9 Suppose you test one circuit. With probability p, the circuit is rejected. Let X be the number of rejected circuits in one test. What is $P_X(x)$?

Because there are only two outcomes in the sample space, $X = 1$ with probability p and $X = 0$ with probability $1 - p$.

$$P_X(x) = \begin{cases} 1-p & x=0 \\ p & x=1 \\ 0 & \text{otherwise} \end{cases} \tag{2.14}$$

Therefore, the number of circuits rejected in one test is a Bernoulli (p) random variable.

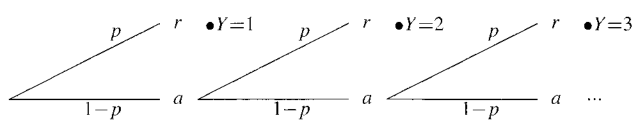
Example 2.10 If there is a 0.2 probability of a reject,



$$P_X(x) = \begin{cases} 0.8 & x=0 \\ 0.2 & x=1 \\ 0 & \text{otherwise} \end{cases} \tag{2.15}$$

Example 2.11 In a test of integrated circuits there is a probability p that each circuit is rejected. Let Y equal the number of tests up to and including the first test that discovers a reject. What is the PMF of Y?

The procedure is to keep testing circuits until a reject appears. Using a to denote an accepted circuit and r to denote a reject, the tree is



From the tree, we see that $P[Y = 1] = p$, $P[Y = 2] = p(1-p)$, $P[Y = 3] = p(1-p)^2$, and, in general, $P[Y = y] = p(1 - p)^{y-1}$. Therefore,

$$P_Y(y) = \begin{cases} p(1-p)^{y-1} & y = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2.16}$$

Y is referred to as a *geometric random variable* because the probabilities in the PMF constitute a geometric series.

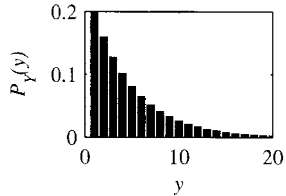
Definition 2.6 ***Geometric*** (p) ***Random Variable***

X is a ***geometric*** (p) *random variable if the PMF of X has the form*

$$P_X(x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

where the parameter p is in the range $0 < p < 1$.

Example 2.12 If there is a 0.2 probability of a reject,



$$P_Y(y) = \begin{cases} (0.2)(0.8)^{y-1} & y = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{2.17}$$

Example 2.13 Suppose we test n circuits and each circuit is rejected with probability p independent of the results of other tests. Let K equal the number of rejects in the n tests. Find the PMF $P_K(k)$.

....

Adopting the vocabulary of Section 1.9, we call each discovery of a defective circuit a *success*, and each test is an independent trial with success probability p. The event $K = k$ corresponds to k successes in n trials, which we have already found, in Equation (1.18), to be the binomial probability

$$P_K(k) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{2.18}$$

K is an example of a *binomial random variable.*

Definition 2.7 ***Binomial*** (n, p) ***Random Variable***

X is a ***binomial*** (n, p) *random variable if the PMF of X has the form*

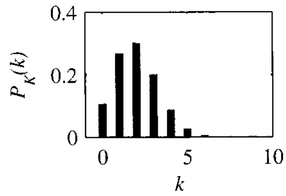
$$P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where $0 < p < 1$ *and n is an integer such that* $n \geq 1$.

We must keep in mind that Definition 2.7 depends on $\binom{n}{x}$ being defined as zero for all $x \notin \{0, 1, \ldots, n\}$.

Whenever we have a sequence of n independent trials each with success probability p, the number of successes is a binomial random variable. In general, for a binomial (n, p) random variable, we call n the number of trials and p the success probability. Note that a Bernoulli random variable is a binomial random variable with $n = 1$.

Example 2.14 If there is a 0.2 probability of a reject and we perform 10 tests,



$$P_K(k) = \binom{10}{k}(0.2)^k(0.8)^{10-k}. \tag{2.19}$$

Example 2.15 Suppose you test circuits until you find k rejects. Let L equal the number of tests. What is the PMF of L?

For large values of k, the tree becomes difficult to draw. Once again, we view the tests as a sequence of independent trials where finding a reject is a success. In this case, $L = l$ if and only if there are $k-1$ successes in the first $l-1$ trials, *and* there is a success on trial l so that

$$P[L = l] = P\Bigg[\underbrace{k-1 \text{ rejects in } l-1 \text{ attempts}}_{A}, \underbrace{\text{success on attempt } l}_{B}\Bigg] \tag{2.20}$$

The events A and B are independent since the outcome of attempt l is not affected by the previous $l-1$ attempts. Note that $P[A]$ is the binomial probability of $k-1$ successes in $l-1$ trials so that

$$P[A] = \binom{l-1}{k-1}p^{k-1}(1-p)^{l-1-(k-1)} \tag{2.21}$$

Finally, since $P[B] = p$,

$$P_L(l) = P[A]\,P[B] = \binom{l-1}{k-1}p^k(1-p)^{l-k} \tag{2.22}$$

L is an example of a *Pascal* random variable.

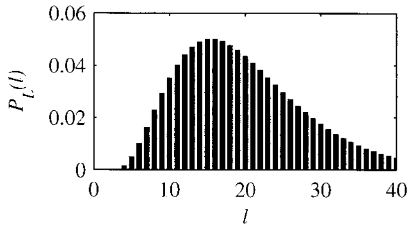
Definition 2.8 ***Pascal*** (k, p) ***Random Variable***
*X is a **Pascal** (k, p) random variable if the PMF of X has the form*

$$P_X(x) = \binom{x-1}{k-1}p^k(1-p)^{x-k}$$

where $0 < p < 1$ and k is an integer such that $k \geq 1$.

For a sequence of n independent trials with success probability p, a Pascal random variable is the number of trials up to and including the kth success. We must keep in mind that for a Pascal (k, p) random variable X, $P_X(x)$ is nonzero only for $x = k, k+1, \ldots$. Mathematically, this is guaranteed by the extended definition of $\binom{x-1}{k-1}$. Also note that a geometric (p) random variable is a Pascal $(k = 1, p)$ random variable.

Example 2.16 If there is a 0.2 probability of a reject and we seek four defective circuits, the random variable L is the number of tests necessary to find the four circuits. The PMF is



$$P_L(l) = \binom{l-1}{3}(0.2)^4(0.8)^{l-4}. \qquad (2.23)$$

Example 2.17 In an experiment with equiprobable outcomes, the random variable N has the range $S_N = \{k, k+1, k+2, \cdots, l\}$, where k and l are integers with $k < l$. The range contains $l - k + 1$ numbers, each with probability $1/(l-k+1)$. Therefore, the PMF of N is

$$P_N(n) = \begin{cases} 1/(l-k+1) & n = k, k+1, k+2, \ldots, l \\ 0 & \text{otherwise} \end{cases} \qquad (2.24)$$

N is an example of a *discrete uniform* random variable.

Definition 2.9 ***Discrete Uniform*** (k, l) ***Random Variable***

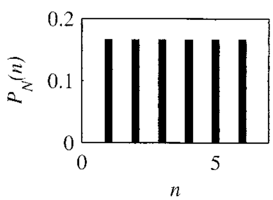
*X is a **discrete uniform** (k, l) random variable if the PMF of X has the form*

$$P_X(x) = \begin{cases} 1/(l-k+1) & x = k, k+1, k+2, \ldots, l \\ 0 & \textit{otherwise} \end{cases}$$

where the parameters k and l are integers such that $k < l$.

To describe this discrete uniform random variable, we use the expression "X is uniformly distributed between k and l."

Example 2.18 Roll a fair die. The random variable N is the number of spots that appears on the side facing up. Therefore, N is a discrete uniform $(1, 6)$ random variable and



$$P_N(n) = \begin{cases} 1/6 & n = 1, 2, 3, 4, 5, 6 \\ 0 & \text{otherwise.} \end{cases} \qquad (2.25)$$

The probability model of a Poisson random variable describes phenomena that occur randomly in time. While the time of each occurrence is completely random, there is a known average number of occurrences per unit time. The Poisson model is used widely in many fields. For example, the arrival of information requests at a World Wide Web server, the initiation of telephone calls, and the emission of particles from a radioactive source are often modeled as Poisson random variables. We will return to Poisson random variables many times in this text. At this point, we consider only the basic properties.

Definition 2.10 ***Poisson*** (α) ***Random Variable***

*X is a **Poisson** (α) random variable if the PMF of X has the form*

$$P_X(x) = \begin{cases} \alpha^x e^{-\alpha}/x! & x = 0, 1, 2, \ldots, \\ 0 & \text{otherwise,} \end{cases}$$

where the parameter α is in the range $\alpha > 0$.

To describe a Poisson random variable, we will call the occurrence of the phenomenon of interest an *arrival*. A Poisson model often specifies an average rate, λ arrivals per second and a time interval, T seconds. In this time interval, the number of arrivals X has a Poisson PMF with $\alpha = \lambda T$.

Example 2.19 The number of hits at a Web site in any time interval is a Poisson random variable. A particular site has on average $\lambda = 2$ hits per second. What is the probability that there are no hits in an interval of 0.25 seconds? What is the probability that there are no more than two hits in an interval of one second?

In an interval of 0.25 seconds, the number of hits H is a Poisson random variable with $\alpha = \lambda T =$ (2 hits/s) $\times$ (0.25 s) $=$ 0.5 hits. The PMF of H is

$$P_H(h) = \begin{cases} 0.5^h e^{-0.5}/h! & h = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2.26}$$

The probability of no hits is

$$P[H = 0] = P_H(0) = (0.5)^0 e^{-0.5}/0! = 0.607. \tag{2.27}$$

In an interval of 1 second, $\alpha = \lambda T =$ (2 hits/s) $\times$ (1s) $=$ 2 hits. Letting J denote the number of hits in one second, the PMF of J is

$$P_J(j) = \begin{cases} 2^j e^{-2}/j! & j = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2.28}$$

To find the probability of no more than two hits, we note that $\{J \le 2\} = \{J = 0\} \cup \{J = 1\} \cup \{J = 2\}$ is the union of three mutually exclusive events. Therefore,

$$P[J \le 2] = P[J = 0] + P[J = 1] + P[J = 2] \tag{2.29}$$
$$= P_J(0) + P_J(1) + P_J(2) \tag{2.30}$$
$$= e^{-2} + 2^1 e^{-2}/1! + 2^2 e^{-2}/2! = 0.677. \tag{2.31}$$

Example 2.20 The number of database queries processed by a computer in any 10-second interval is a Poisson random variable, K, with $\alpha = 5$ queries. What is the probability that there will be no queries processed in a 10-second interval? What is the probability that at least two queries will be processed in a 2-second interval?

The PMF of K is

$$P_K(k) = \begin{cases} 5^k e^{-5}/k! & k = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2.32}$$

Therefore $P[K = 0] = P_K(0) = e^{-5} = 0.0067$. To answer the question about the 2-second interval, we note in the problem definition that $\alpha = 5$ queries $= \lambda T$ with $T = 10$ seconds. Therefore, $\lambda = 0.5$ queries per second. If N is the number of queries processed in a 2-second interval, $\alpha = 2\lambda = 1$ and N is the Poisson (1) random variable with PMF

$$P_N(n) = \begin{cases} e^{-1}/n! & n = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2.33}$$

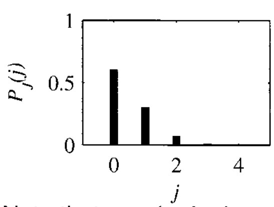
Therefore,

$$P[N \geq 2] = 1 - P_N(0) - P_N(1) = 1 - e^{-1} - e^{-1} = 0.264. \tag{2.34}$$

Note that the units of λ and T have to be consistent. Instead of $\lambda = 0.5$ queries per second for $T = 10$ seconds, we could use $\lambda = 30$ queries per minute for the time interval $T = 1/6$ minutes to obtain the same $\alpha = 5$ queries, and therefore the same probability model.

In the following examples, we see that for a fixed rate λ, the shape of the Poisson PMF depends on the length T over which arrivals are counted.

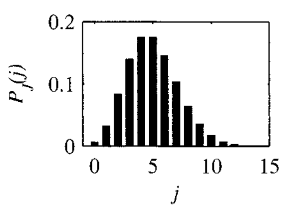
Example 2.21 Calls arrive at random times at a telephone switching office with an average of $\lambda = 0.25$ calls/second. The PMF of the number of calls that arrive in a $T = 2$-second interval is the Poisson (0.5) random variable with PMF



$$P_J(j) = \begin{cases} (0.5)^j e^{-0.5}/j! & j = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{2.35}$$

Note that we obtain the same PMF if we define the arrival rate as $\lambda = 60 \cdot 0.25 = 15$ calls per minute and derive the PMF of the number of calls that arrive in $2/60 = 1/30$ minutes.

Example 2.22 Calls arrive at random times at a telephone switching office with an average of $\lambda = 0.25$ calls per second. The PMF of the number of calls that arrive in any $T = 20$-second interval is the Poisson (5) random variable with PMF



$$P_J(j) = \begin{cases} 5^j e^{-5}/j! & j = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{2.36}$$

Each property of Theorem 2.2 has an equivalent statement in words:

(a) Going from left to right on the x-axis, $F_X(x)$ starts at zero and ends at one.

(b) The CDF never decreases as it goes from left to right.

(c) For a discrete random variable X, there is a jump (discontinuity) at each value of $x_i \in S_X$. The height of the jump at x_i is $P_X(x_i)$.

(d) Between jumps, the graph of the CDF of the discrete random variable X is a horizontal line.

Another important consequence of the definition of the CDF is that the difference between the CDF evaluated at two points is the probability that the random variable takes on a value between these two points:

Theorem 2.3 *For all $b \geq a$,*

$$F_X(b) - F_X(a) = P[a < X \leq b].$$

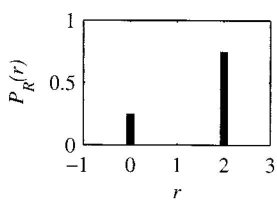
Proof To prove this theorem, express the event $E_{ab} = \{a < X \leq b\}$ as a part of a union of disjoint events. Start with the event $E_b = \{X \leq b\}$. Note that E_b can be written as the union

$$E_b = \{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\} = E_a \cup E_{ab} \tag{2.37}$$

Note also that E_a and E_{ab} are disjoint so that $P[E_b] = P[E_a] + P[E_{ab}]$. Since $P[E_b] = F_X(b)$ and $P[E_a] = F_X(a)$, we can write $F_X(b) = F_X(a) + P[a < X \leq b]$. Therefore $P[a < X \leq b] = F_X(b) - F_X(a)$.

In working with the CDF, it is necessary to pay careful attention to the nature of inequalities, strict ($<$) or loose ($\leq$). The definition of the CDF contains a loose (less than or equal) inequality, which means that the function is continuous from the right. To sketch a CDF of a discrete random variable, we draw a graph with the vertical value beginning at zero at the left end of the horizontal axis (negative numbers with large magnitude). It remains zero until x_1, the first value of x with nonzero probability. The graph jumps by an amount $P_X(x_i)$ at each x_i with nonzero probability. We draw the graph of the CDF as a staircase with jumps at each x_i with nonzero probability. The CDF is the upper value of every jump in the staircase.

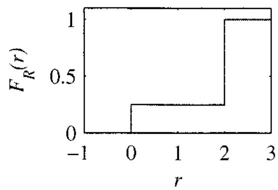
Example 2.23 In Example 2.6, we found that random variable R has PMF



$$P_R(r) = \begin{cases} 1/4 & r = 0, \\ 3/4 & r = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{2.38}$$

Find and sketch the CDF of random variable R.

From the PMF $P_R(r)$, random variable R has CDF



$$F_R(r) = P[R \le r] = \begin{cases} 0 & r < 0, \\ 1/4 & 0 \le r < 2, \\ 1 & r \ge 2. \end{cases} \tag{2.39}$$

Keep in mind that at the discontinuities $r = 0$ and $r = 2$, the values of $F_R(r)$ are the upper values: $F_R(0) = 1/4$, and $F_R(2) = 1$. Math texts call this the *right hand limit* of $F_R(r)$.

Consider any finite random variable X with possible values (nonzero probability) between $x_{\min}$ and $x_{\max}$. For this random variable, the numerical specification of the CDF begins with

$$F_X(x) = 0 \qquad x < x_{\min},$$

and ends with

$$F_X(x) = 1 \qquad x \ge x_{\max}.$$

Like the statement "$P_X(x) = 0$ otherwise," the description of the CDF is incomplete without these two statements. The next example displays the CDF of an infinite discrete random variable.

Example 2.24 In Example 2.11, let the probability that a circuit is rejected equal $p = 1/4$. The PMF of Y, the number of tests up to and including the first reject, is the geometric $(1/4)$ random variable with PMF

$$P_Y(y) = \begin{cases} (1/4)(3/4)^{y-1} & y = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2.40}$$

What is the CDF of Y?

Y is an infinite random variable, with nonzero probabilities for all positive integers. For any integer $n \ge 1$, the CDF is

$$F_Y(n) = \sum_{j=1}^{n} P_Y(j) = \sum_{j=1}^{n} \frac{1}{4}\left(\frac{3}{4}\right)^{j-1}. \tag{2.41}$$

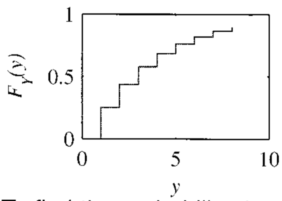
Equation (2.41) is a geometric series. Familiarity with the geometric series is essential for calculating probabilities involving geometric random variables. Appendix B summarizes the most important facts. In particular, Math Fact B.4 implies $(1-x)\sum_{j=1}^{n} x^{j-1} = 1 - x^n$. Substituting $x = 3/4$, we obtain

$$F_Y(n) = 1 - \left(\frac{3}{4}\right)^n. \tag{2.42}$$

The complete expression for the CDF of Y must show $F_Y(y)$ for all integer *and noninteger* values of y. For an integer-valued random variable Y, we can do this in a simple way using the *floor function* $\lfloor y \rfloor$, which is the largest integer less than or equal to y. In particular, if $n \le y < n - 1$ for some integer n, then $n = \lfloor y \rfloor$ and

$$F_Y(y) = P[Y \le y] = P[Y \le n] = F_Y(n) = F_Y(\lfloor y \rfloor). \tag{2.43}$$

In terms of the floor function, we can express the CDF of Y as

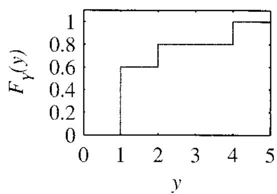


$$F_Y(y) = \begin{cases} 0 & y < 1, \\ 1 - (3/4)^{\lfloor y \rfloor} & y \ge 1. \end{cases} \tag{2.44}$$

To find the probability that Y takes a value in the set $\{4, 5, 6, 7, 8\}$, we refer to Theorem 2.3 and compute

$$P[3 < Y \le 8] = F_Y(8) - F_Y(3) = (3/4)^3 - (3/4)^8 = 0.322. \tag{2.45}$$

Quiz 2.4 *Use the CDF $F_Y(y)$ to find the following probabilities:*



(1) $P[Y < 1]$ (2) $P[Y \le 1]$

(3) $P[Y > 2]$ (4) $P[Y \ge 2]$

(5) $P[Y = 1]$ (6) $P[Y = 3]$

2.5 Averages

The average value of a collection of numerical observations is a *statistic* of the collection, a single number that describes the entire collection. Statisticians work with several kinds of averages. The ones that are used the most are the *mean*, the *median*, and the *mode*.

The mean value of a set of numbers is perhaps the most familiar. You get the mean value by adding up all the numbers in the collection and dividing by the number of terms in the sum. Think about the mean grade in the mid-term exam for this course. The median is also an interesting typical value of a set of data.

The median is a number in the middle of the set of numbers, in the sense that an equal number of members of the set are below the median and above the median.

A third average is the mode of a set of numbers. The mode is the most common number in the collection of observations. There are as many or more numbers with that value than any other value. If there are two or more numbers with this property, the collection of observations is called *multimodal*.

Example 2.25 For one quiz, 10 students have the following grades (on a scale of 0 to 10):

$$9, 5, 10, 8, 4, 7, 5, 5, 8, 7 \tag{2.46}$$

Find the mean, the median, and the mode.

The sum of the ten grades is 68. The mean value is 68/10 = 6.8. The median is 7 since there are four scores below 7 and four scores above 7. The mode is 5 since that score occurs more often than any other. It occurs three times.

Example 2.25 and the preceding comments on averages apply to observations collected by an experimenter. We use probability models with random variables to characterize experiments with numerical outcomes. A *parameter* of a probability model corresponds to a statistic of a collection of outcomes. Each parameter is a number that can be computed from the PMF or CDF of a random variable. The most important of these is the *expected value* of a random variable, corresponding to the mean value of a collection of observations. We will work with expectations throughout the course. Corresponding to the other two averages, we have the following definitions:

Definition 2.12 ***Mode***

*A **mode** of random variable X is a number x_{mod} satisfying $P_X(x_{\text{mod}}) \geq P_X(x)$ for all x.*

Definition 2.13 ***Median***

*A **median**, x_{med}, of random variable X is a number that satisfies*

$$P[X < x_{\text{med}}] = P[X > x_{\text{med}}]$$

If you read the definitions of *mode* and *median* carefully, you will observe that neither the mode nor the median of a random variable X need be unique. A random variable can have several modes or medians.

The expected value of a random variable corresponds to adding up a number of measurements and dividing by the number of terms in the sum. Two notations for the expected value of random variable X are $E[X]$ and μ_X.

Definition 2.14 ***Expected Value***

*The **expected value** of X is*

$$E[X] = \mu_X = \sum_{x \in S_X} x P_X(x).$$

Expectation is a synonym for expected value. Sometimes the term *mean value* is also used as a synonym for expected value. We prefer to use mean value to refer to a *statistic* of a set of experimental outcomes (the sum divided by the number of outcomes) to distinguish it from expected value, which is a *parameter* of a probability model. If you recall your studies of mechanics, the form of Definition 2.14 may look familiar. Think of point masses on a line with a mass of $P_X(x)$ kilograms at a distance of x meters from the origin. In this model, μ_X in Definition 2.14 is the center of mass. This is why $P_X(x)$ is called probability *mass* function.

To understand how this definition of expected value corresponds to the notion of adding up a set of measurements, suppose we have an experiment that produces a random variable X and we perform n independent trials of this experiment. We denote the value that X takes on the ith trial by $x(i)$. We say that $x(1), \ldots, x(n)$ is a set of n sample values of X. Corresponding to the average of a set of numbers, we have, after n trials of the experiment, the sample average

$$m_n = \frac{1}{n} \sum_{i=1}^{n} x(i). \tag{2.47}$$

Each $x(i)$ takes values in the set S_X. Out of the n trials, assume that each $x \in S_X$ occurs N_x times. Then the sum (2.47) becomes

$$m_n = \frac{1}{n} \sum_{x \in S_X} N_x x = \sum_{x \in S_X} \frac{N_x}{n} x. \tag{2.48}$$

Recall our discussion in Section 1.3 of the relative frequency interpretation of probability. There we pointed out that if in n observations of an experiment, the event A occurs N_A times, we can interpret the probability of A as

$$P[A] = \lim_{n \to \infty} \frac{N_A}{n} \tag{2.49}$$

This is the relative frequency of A. In the notation of random variables, we have the corresponding observation that

$$P_X(x) = \lim_{n \to \infty} \frac{N_x}{n}. \tag{2.50}$$

This suggests that

$$\lim_{n \to \infty} m_n = \sum_{x \in S_X} x P_X(x) = E[X]. \tag{2.51}$$

Equation (2.51) says that the definition of $E[X]$ corresponds to a model of doing the same experiment repeatedly. After each trial, add up all the observations to date and divide by the number of trials. We prove in Chapter 7 that the result approaches the expected value as the number of trials increases without limit. We can use Definition 2.14 to derive the expected value of each family of random variables defined in Section 2.3.

Theorem 2.4 *The Bernoulli (p) random variable X has expected value $E[X] = p$.*

Proof $E[X] = 0 \cdot P_X(0) + 1 P_X(1) = 0(1-p) + 1(p) = p$.

Example 2.26 Random variable R in Example 2.6 has PMF

$$P_R(r) = \begin{cases} 1/4 & r = 0, \\ 3/4 & r = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{2.52}$$

What is $E[R]$?

$$E[R] = \mu_R = 0 \cdot P_R(0) + 2P_R(2) = 0(1/4) + 2(3/4) = 3/2. \tag{2.53}$$

Theorem 2.5 *The geometric (p) random variable X has expected value $E[X] = 1/p$.*

Proof Let $q = 1 - p$. The PMF of X becomes

$$P_X(x) = \begin{cases} pq^{x-1} & x = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2.54}$$

The expected value $E[X]$ is the infinite sum

$$E[X] = \sum_{x=1}^{\infty} x P_X(x) = \sum_{x=1}^{\infty} xpq^{x-1}. \tag{2.55}$$

Applying the identity of Math Fact B.7, we have

$$E[X] = p \sum_{x=1}^{\infty} xq^{x-1} = \frac{p}{q} \sum_{x=1}^{\infty} xq^x = \frac{p}{q} \frac{q}{1-q^2} = \frac{p}{p^2} = \frac{1}{p}. \tag{2.56}$$

This result is intuitive if you recall the integrated circuit testing experiments and consider some numerical values. If the probability of rejecting an integrated circuit is $p = 1/5$, then on average, you have to perform $E[Y] = 1/p = 5$ tests to observe the first reject. If $p = 1/10$, the average number of tests until the first reject is $E[Y] = 1/p = 10$.

Theorem 2.6 *The Poisson (α) random variable in Definition 2.10 has expected value $E[X] = \alpha$.*

Proof

$$E[X] = \sum_{x=0}^{\infty} x P_X(x) = \sum_{x=0}^{\infty} x \frac{\alpha^x}{x!} e^{-\alpha}. \tag{2.57}$$

We observe that $x/x! = 1/(x-1)!$ and also that the $x = 0$ term in the sum is zero. In addition, we substitute $\alpha^x = \alpha \cdot \alpha^{x-1}$ to factor α from the sum to obtain

$$E[X] = \alpha \sum_{x=1}^{\infty} \frac{\alpha^{x-1}}{(x-1)!} e^{-\alpha}. \tag{2.58}$$

Next we substitute $l = x - 1$, with the result

$$E[X] = \alpha \underbrace{\sum_{l=0}^{\infty} \frac{\alpha^l}{l!} e^{-\alpha}}_{1} = \alpha. \tag{2.59}$$

We can conclude that the marked sum equals 1 either by invoking the identity $e^{\alpha} = \sum_{l=0}^{\infty} \alpha^l / l!$ or by applying Theorem 2.1(b) to the fact that the marked sum is the sum of the Poisson PMF over all values in the range of the random variable.

In Section 2.3, we modeled the number of random arrivals in an interval of length T by a Poisson random variable with parameter $\alpha = \lambda T$. We referred to λ as *the average rate* of arrivals with little justification. Theorem 2.6 provides the justification by showing that $\lambda = \alpha / T$ is the expected number of arrivals per unit time.

The next theorem provides, without derivations, the expected values of binomial, Pascal, and discrete uniform random variables.

Theorem 2.7

(a) For the binomial (n, p) random variable X of Definition 2.7,

$$E[X] = np.$$

(b) For the Pascal (k, p) random variable X of Definition 2.8,

$$E[X] = k/p.$$

(c) For the discrete uniform (k, l) random variable X of Definition 2.9,

$$E[X] = (k + l)/2.$$

In the following theorem, we show that the Poisson PMF is a limiting case of a binomial PMF. In the binomial model, n, the number of Bernoulli trials grows without limit but the expected number of trials np remains constant at α, the expected value of the Poisson PMF. In the theorem, we let $\alpha = \lambda T$ and divide the T-second interval into n time slots each with duration T/n. In each slot, we assume that there is either one arrival, with probability $p = \lambda T/n = \alpha/n$, or there is no arrival in the time slot, with probability $1 - p$.

Theorem 2.8 *Perform n Bernoulli trials. In each trial, let the probability of success be α/n, where $\alpha > 0$ is a constant and $n > \alpha$. Let the random variable K_n be the number of successes in the n trials. As $n \to \infty$, $P_{K_n}(k)$ converges to the PMF of a Poisson (α) random variable.*

Proof We first note that K_n is the binomial $(n, \alpha n)$ random variable with PMF

$$P_{K_n}(k) = \binom{n}{k} (\alpha/n)^k \left(1 - \frac{\alpha}{n}\right)^{n-k}. \tag{2.60}$$

For $k = 0, \ldots, n$, we can write

$$P_K(k) = \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\alpha^k}{k!} \left(1 - \frac{\alpha}{n}\right)^{n-k}. \tag{2.61}$$

Notice that in the first fraction, there are k terms in the numerator. The denominator is n^k, also a product of k terms, all equal to n. Therefore, we can express this fraction as the product of k fractions each of the form $(n-j)/n$. As $n \to \infty$, each of these fractions approaches 1. Hence,

$$\lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} = 1. \tag{2.62}$$

Furthermore, we have

$$\left(1 - \frac{\alpha}{n}\right)^{n-k} = \frac{\left(1 - \frac{\alpha}{n}\right)^n}{\left(1 - \frac{\alpha}{n}\right)^k}. \tag{2.63}$$

As n grows without bound, the denominator approaches 1 and, in the numerator, we recognize the identity $\lim_{n\to\infty}(1 - \alpha/n)^n = e^{-\alpha}$. Putting these three limits together leads us to the result that for any integer $k \geq 0$,

$$\lim_{n\to\infty} P_{K_n}(k) = \begin{cases} \alpha^k e^{-\alpha}/k! & k = 0, 1, \ldots \\ 0 & \text{otherwise,} \end{cases} \tag{2.64}$$

which is the Poisson PMF.

Quiz 2.5 *The probability that a call is a voice call is $P[V] = 0.7$. The probability of a data call is $P[D] = 0.3$. Voice calls cost 25 cents each and data calls cost 40 cents each. Let C equal the cost (in cents) of one telephone call and find*

(1) The PMF $P_C(c)$ *(2) The expected value $E[C]$*

2.6 Functions of a Random Variable

In many practical situations, we observe sample values of a random variable and use these sample values to compute other quantities. One example that occurs frequently is an experiment in which the procedure is to measure the power level of the received signal in a cellular telephone. An observation is x, the power level in units of milliwatts. Frequently engineers convert the measurements to decibels by calculating $y = 10\log_{10} x$ dBm (decibels with respect to one milliwatt). If x is a sample value of a random variable X, Definition 2.1 implies that y is a sample value of a random variable Y. Because we obtain Y from another random variable, we refer to Y as a *derived random variable*.

Definition 2.15 ***Derived Random Variable***

*Each sample value y of a **derived random variable** Y is a mathematical function $g(x)$ of a sample value x of another random variable X. We adopt the notation $Y = g(X)$ to describe the relationship of the two random variables.*

Example 2.27 The random variable X is the number of pages in a facsimile transmission. Based on experience, you have a probability model $P_X(x)$ for the number of pages in each fax you send. The phone company offers you a new charging plan for faxes: \$0.10 for the first page, \$0.09 for the second page, etc., down to \$0.06 for the fifth page. For all faxes between 6 and 10 pages, the phone company will charge \$0.50 per fax. (It will not accept faxes longer than ten pages.) Find a function $Y = g(X)$ for the charge in cents for sending one fax.

The following function corresponds to the new charging plan.

$$Y = g(X) = \begin{cases} 10.5X - 0.5X^2 & 1 \le X \le 5 \\ 50 & 6 \le X \le 10 \end{cases} \tag{2.65}$$

You would like a probability model $P_Y(y)$ for your phone bill under the new charging plan. You can analyze this model to decide whether to accept the new plan.

In this section we determine the probability model of a derived random variable from the probability model of the original random variable. We start with $P_X(x)$ and a function $Y = g(X)$. We use this information to obtain $P_Y(y)$.

Before we present the procedure for obtaining $P_Y(y)$, we address an issue that can be confusing to students learning probability, which is the properties of $P_X(x)$ and $g(x)$. Although they are both functions with the argument x, they are entirely different. $P_X(x)$ describes the probability model of a random variable. It has the special structure prescribed in Theorem 2.1. On the other hand, $g(x)$ can be any function at all. When we combine them to derive the probability model for Y, we arrive at a PMF that also conforms to Theorem 2.1.

To describe Y in terms of our basic model of probability, we specify an experiment consisting of the following procedure and observation:

Sample value of $Y = g(X)$
`Perform an experiment and observe an outcome` s.
`From` s, `find` x, `the corresponding value of` X.
`Observe` y `by calculating` $y = g(x)$.

This procedure maps each experimental outcome to a number, y, that is a sample value of a random variable, Y. To derive $P_Y(y)$ from $P_X(x)$ and $g(\cdot)$, we consider all of the possible values of x. For each $x \in S_X$, we compute $y = g(x)$. If $g(x)$ transforms different values of x into different values of y ($g(x_1) \neq g(x_2)$ if $x_1 \neq x_2$) we have simply that

$$P_Y(y) = P[Y = g(x)] = P[X = x] = P_X(x) \tag{2.66}$$

The situation is a little more complicated when $g(x)$ transforms several values of x to the same y. In this case, we consider all the possible values of y. For each $y \in S_Y$, we add the probabilities of all of the values $x \in S_x$ for which $g(x) = y$. Theorem 2.9 applies in general. It reduces to Equation (2.66) when $g(x)$ is a one-to-one tranformation.

Theorem 2.9 *For a discrete random variable X, the PMF of $Y = g(X)$ is*

$$P_Y(y) = \sum_{x:g(x)=y} P_X(x).$$

If we view $X = x$ as the outcome of an experiment, then Theorem 2.9 says that $P[Y = y]$

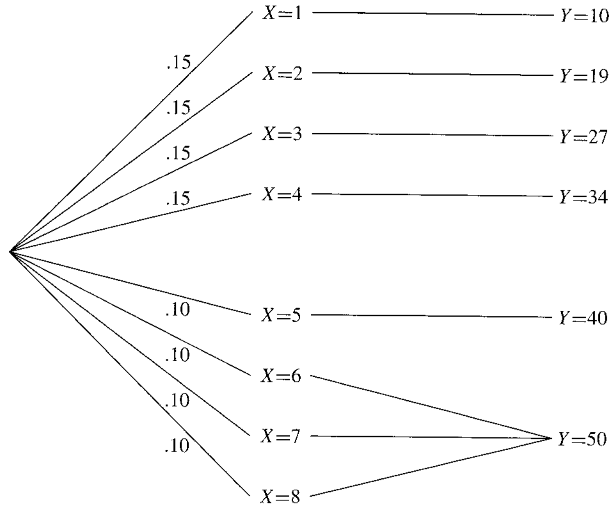


Figure 2.1 The derived random variable $Y = g(X)$ for Example 2.29.

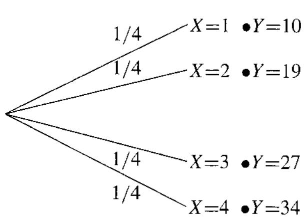
equals the sum of the probabilities of all the outcomes $X = x$ for which $Y = y$.

Example 2.28 In Example 2.27, suppose all your faxes contain 1, 2, 3, or 4 pages with equal probability. Find the PMF and expected value of Y, the charge for a fax.

From the problem statement, the number of pages X has PMF

$$P_X(x) = \begin{cases} 1/4 & x = 1, 2, 3, 4, \\ 0 & \text{otherwise.} \end{cases} \qquad (2.67)$$

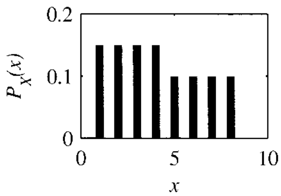
The charge for the fax, Y, has range $S_Y = \{10, 19, 27, 34\}$ corresponding to $S_X = \{1, 2, 3, 4\}$. The experiment can be described by the following tree. Here each value of Y results in a unique value of X. Hence, we can use Equation (2.66) to find $P_Y(y)$.



$$P_Y(y) = \begin{cases} 1/4 & y = 10, 19, 27, 34, \\ 0 & \text{otherwise.} \end{cases} \qquad (2.68)$$

The expected fax bill is $E[Y] = (1/4)(10 + 19 + 27 + 34) = 22.5$ cents.

Example 2.29 Suppose the probability model for the number of pages X of a fax in Example 2.28 is



$$P_X(x) = \begin{cases} 0.15 & x = 1,2,3,4 \\ 0.1 & x = 5,6,7,8 \\ 0 & \text{otherwise} \end{cases} \tag{2.69}$$

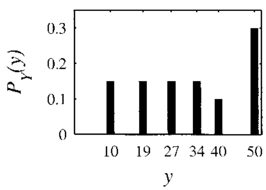
For the pricing plan given in Example 2.27, what is the PMF and expected value of Y, the cost of a fax?

....

Now we have three values of X, specifically $(6, 7, 8)$, transformed by $g(\cdot)$ into $Y = 50$. For this situation we need the more general view of the PMF of Y, given by Theorem 2.9. In particular, $y_6 = 50$, and we have to add the probabilities of the outcomes $X = 6$, $X = 7$, and $X = 8$ to find $P_Y(50)$. That is,

$$P_Y(50) = P_X(6) + P_X(7) + P_X(8) = 0.30. \tag{2.70}$$

The steps in the procedure are illustrated in the diagram of Figure 2.1. Applying Theorem 2.9, we have

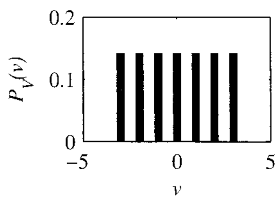


$$P_Y(y) = \begin{cases} 0.15 & y = 10, 19, 27, 34, \\ 0.10 & y = 40, \\ 0.30 & y = 50, \\ 0 & \text{otherwise.} \end{cases} \tag{2.71}$$

For this probability model, the expected cost of sending a fax is

$$E[Y] = 0.15(10 + 19 + 27 + 34) + 0.10(40) + 0.30(50) = 32.5 \text{ cents.} \tag{2.72}$$

Example 2.30 The amplitude V (volts) of a sinusoidal signal is a random variable with PMF

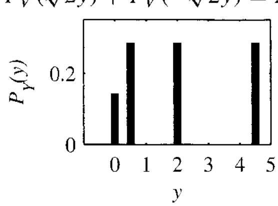


$$P_V(v) = \begin{cases} 1/7 & v = -3, -2, \ldots, 3 \\ 0 & \text{otherwise} \end{cases} \tag{2.73}$$

Let $Y = V^2/2$ watts denote the average power of the transmitted signal. Find $P_Y(y)$.

....

The possible values of Y are $S_Y = \{0, 0.5, 2, 4.5\}$. Since $Y = y$ when $V = \sqrt{2y}$ or $V = -\sqrt{2y}$, we see that $P_Y(0) = P_V(0) = 1/7$. For $y = 1/2, 2, 9/2$, $P_Y(y) = P_V(\sqrt{2y}) + P_V(-\sqrt{2y}) = 2/7$. Therefore,



$$P_Y(y) = \begin{cases} 1/7 & y = 0, \\ 2/7 & y = 1/2, 2, 9/2, \\ 0 & \text{otherwise.} \end{cases} \tag{2.74}$$

Quiz 2.6 *Monitor three phone calls and observe whether each one is a voice call or a data call. The random variable N is the number of voice calls. Assume N has PMF*

$$P_N(n) = \begin{cases} 0.1 & n = 0, \\ 0.3 & n = 1, 2, 3, \\ 0 & \text{otherwise}. \end{cases} \tag{2.75}$$

Voice calls cost 25 cents each and data calls cost 40 cents each. T cents is the cost of the three telephone calls monitored in the experiment.
(1) Express T as a function of N. *(2) Find $P_T(t)$ and $E[T]$.*

2.7 Expected Value of a Derived Random Variable

We encounter many situations in which we need to know only the expected value of a derived random variable rather than the entire probability model. Fortunately, to obtain this average, it is not necessary to compute the PMF or CDF of the new random variable. Instead, we can use the following property of expected values.

Theorem 2.10 *Given a random variable X with PMF $P_X(x)$ and the derived random variable $Y = g(X)$, the expected value of Y is*

$$E[Y] = \mu_Y = \sum_{x \in S_X} g(x) P_X(x)$$

Proof From the definition of $E[Y]$ and Theorem 2.9, we can write

$$E[Y] = \sum_{y \in S_Y} y P_Y(y) = \sum_{y \in S_Y} y \sum_{x:g(x)=y} P_X(x) = \sum_{y \in S_Y} \sum_{x:g(x)=y} g(x) P_X(x), \tag{2.76}$$

where the last double summation follows because $g(x) = y$ for each x in the inner sum. Since $g(x)$ transforms each possible outcome $x \in S_X$ to a value $y \in S_Y$, the preceding double summation can be written as a single sum over over all possible values $x \in S_X$. That is,

$$E[Y] = \sum_{x \in S_X} g(x) P_X(x) \tag{2.77}$$

Example 2.31 In Example 2.28,

$$P_X(x) = \begin{cases} 1/4 & x = 1, 2, 3, 4, \\ 0 & \text{otherwise}, \end{cases} \tag{2.78}$$

and

$$Y = g(X) = \begin{cases} 10.5X - 0.5X^2 & 1 \le X \le 5, \\ 50 & 6 \le X \le 10. \end{cases} \tag{2.79}$$

What is $E[Y]$?

Applying Theorem 2.10 we have

$$E[Y] = \sum_{x=1}^{4} P_X(x)\, g(x) \tag{2.80}$$

$$= (1/4)[(10.5)(1) - (0.5)(1)^2] + (1/4)[(10.5)(2) - (0.5)(2)^2] \tag{2.81}$$

$$+ (1/4)[(10.5)(3) - (0.5)(3)^2] + (1/4)[(10.5)(4) - (0.5)(4)^2] \tag{2.82}$$

$$= (1/4)[10 + 19 + 27 + 34] = 22.5 \text{ cents}. \tag{2.83}$$

This of course is the same answer obtained in Example 2.28 by first calculating $P_Y(y)$ and then applying Definition 2.14. As an exercise, you may want to compute $E[Y]$ in Example 2.29 directly from Theorem 2.10.

From this theorem we can derive some important properties of expected values. The first one has to do with the difference between a random variable and its expected value. When students learn their own grades on a midterm exam, they are quick to ask about the class average. Let's say one student has 73 and the class average is 80. She may be inclined to think of her grade as "seven points below average," rather than "73." In terms of a probability model, we would say that the random variable X points on the midterm has been transformed to the random variable

$$Y = g(X) = X - \mu_X \quad \text{points above average.} \tag{2.84}$$

The expected value of $X - \mu_X$ is zero, regardless of the probability model of X.

Theorem 2.11 *For any random variable X,*

$$E[X - \mu_X] = 0.$$

Proof Defining $g(X) = X - \mu_X$ and applying Theorem 2.10 yields

$$E[g(X)] = \sum_{x \in S_X} (x - \mu_X) P_X(x) = \sum_{x \in S_X} x P_X(x) - \mu_X \sum_{x \in S_X} P_X(x). \tag{2.85}$$

The first term on the right side is μ_X by definition. In the second term, $\sum_{x \in S_X} P_X(x) = 1$, so both terms on the right side are μ_X and the difference is zero.

Another property of the expected value of a function of a random variable applies to linear transformations.[1]

Theorem 2.12 *For any random variable X,*

$$E[aX + b] = aE[X] + b.$$

This follows directly from Definition 2.14 and Theorem 2.10. A linear transformation is

[1] We call the transformation $aX + b$ linear although, strictly speaking, it should be called affine.

essentially a scale change of a quantity, like a transformation from inches to centimeters or from degrees Fahrenheit to degrees Celsius. If we express the data (random variable X) in new units, the new average is just the old average transformed to the new units. (If the professor adds five points to everyone's grade, the average goes up by five points.)

This is a rare example of a situation in which $E[g(X)] = g(E[X])$. *It is tempting, but usually wrong, to apply it to other transformations.* For example, if $Y = X^2$, it is usually the case that $E[Y] \neq (E[X])^2$. Expressing this in general terms, it is usually the case that $E[g(X)] \neq g(E[X])$.

Example 2.32 Recall that in Examples 2.6 and 2.26, we found that R has PMF

$$P_R(r) = \begin{cases} 1/4 & r = 0, \\ 3/4 & r = 2, \\ 0 & \text{otherwise}, \end{cases} \tag{2.86}$$

and expected value $E[R] = 3/2$. What is the expected value of $V = g(R) = 4R + 7$?

From Theorem 2.12,

$$E[V] = E[g(R)] = 4E[R] + 7 = 4(3/2) + 7 = 13. \tag{2.87}$$

We can verify this result by applying Theorem 2.10. Using the PMF $P_R(r)$ given in Example 2.6, we can write

$$E[V] = g(0)P_R(0) + g(2)P_R(2) = 7(1/4) + 15(3/4) = 13. \tag{2.88}$$

Example 2.33 In Example 2.32, let $W = h(R) = R^2$. What is $E[W]$?

Theorem 2.10 gives

$$E[W] = \sum h(r)P_R(r) = (1/4)0^2 + (3/4)2^2 = 3. \tag{2.89}$$

Note that this is not the same as $h(E[W]) = (3/2)^2$.

Quiz 2.7 *The number of memory chips M needed in a personal computer depends on how many application programs, A, the owner wants to run simultaneously. The number of chips M and the number of application programs A are described by*

$$M = \begin{cases} 4 & \textit{chips for 1 program}, \\ 4 & \textit{chips for 2 programs}, \\ 6 & \textit{chips for 3 programs}, \\ 8 & \textit{chips for 4 programs}, \end{cases} \qquad P_A(a) = \begin{cases} 0.1(5-a) & a = 1, 2, 3, 4, \\ 0 & \textit{otherwise}. \end{cases} \tag{2.90}$$

(1) What is the expected number of programs $\mu_A = E[A]$?

(2) Express M, the number of memory chips, as a function $M = g(A)$ of the number of application programs A.

(3) Find $E[M] = E[g(A)]$. Does $E[M] = g(E[A])$?

2.8 Variance and Standard Deviation

In Section 2.5, we describe an average as a typical value of a random variable. It is one number that summarizes an entire probability model. After finding an average, someone who wants to look further into the probability model might ask, "How typical is the average?" or, "What are the chances of observing an event far from the average?" In the example of the midterm exam, after you find out your score is 7 points above average, you are likely to ask, "How good is that? Is it near the top of the class or somewhere near the middle?" A measure of dispersion is an answer to these questions wrapped up in a single number. If this measure is small, observations are likely to be near the average. A high measure of dispersion suggests that it is not unusual to observe events that are far from the average.

The most important measures of dispersion are the standard deviation and its close relative, the variance. The variance of random variable X describes the difference between X and its expected value. This difference is the derived random variable, $Y = X - \mu_X$. Theorem 2.11 states that $\mu_Y = 0$, regardless of the probability model of X. Therefore μ_Y provides no information about the dispersion of X around μ_X. A useful measure of the likely difference between X and its expected value is the expected absolute value of the difference, $E[|Y|]$. However, this parameter is not easy to work with mathematically in many situations, and it is not used frequently.

Instead we focus on $E[Y^2] = E[(X - \mu_X)^2]$, which is referred to as $\text{Var}[X]$, the variance of X. The square root of the variance is σ_X, the standard deviation of X.

Definition 2.16 ***Variance***

*The **variance** of random variable X is*

$$\text{Var}[X] = E\left[(X - \mu_X)^2\right].$$

Definition 2.17 ***Standard Deviation***

*The **standard deviation** of random variable X is*

$$\sigma_X = \sqrt{\text{Var}\,[X]}.$$

It is useful to take the square root of $\text{Var}[X]$ because σ_X has the same units (for example, exam points) as X. The units of the variance are squares of the units of the random variable (exam points squared). Thus σ_X can be compared directly with the expected value. Informally we think of outcomes within $\pm\sigma_X$ of μ_X as being in the center of the distribution. Thus if the standard deviation of exam scores is 12 points, the student with a score of $+7$ with respect to the mean can think of herself in the middle of the class. If the standard deviation is 3 points, she is likely to be near the top. Informally, we think of sample values within σ_X of the expected value, $x \in [\mu_X - \sigma_X, \mu_X + \sigma_X]$, as "typical" values of X and other values as "unusual."

Because $(X - \mu_X)^2$ is a function of X, Var$[X]$ can be computed according to Theorem 2.10.

$$\text{Var}[X] = \sigma_X^2 = \sum_{x \in S_X} (x - \mu_X)^2 P_X(x). \tag{2.91}$$

By expanding the square in this formula, we arrive at the most useful approach to computing the variance.

Theorem 2.13

$$\text{Var}[X] = E\left[X^2\right] - \mu_X^2 = E\left[X^2\right] - (E[X])^2$$

Proof Expanding the square in (2.91), we have

$$\begin{aligned}
\text{Var}[X] &= \sum_{x \in S_X} x^2 P_X(x) - \sum_{x \in S_X} 2\mu_X x P_X(x) + \sum_{x \in S_X} \mu_X^2 P_X(x) \\
&= E[X^2] - 2\mu_X \sum_{x \in S_X} x P_X(x) + \mu_X^2 \sum_{x \in S_X} P_X(x) \\
&= E[X^2] - 2\mu_X^2 + \mu_X^2
\end{aligned}$$

We note that $E[X]$ and $E[X^2]$ are examples of *moments* of the random variable X. Var$[X]$ is a *central moment* of X.

Definition 2.18 ***Moments***

For random variable X:

(a) *The nth* ***moment*** *is $E[X^n]$.*

(b) *The nth* ***central moment*** *is $E[(X - \mu_X)^n]$.*

Thus, $E[X]$ is the *first moment* of random variable X. Similarly, $E[X^2]$ is the *second moment*. Theorem 2.13 says that the variance of X is the second moment of X minus the square of the first moment.

Like the PMF and the CDF of a random variable, the set of moments of X is a complete probability model. We learn in Section 6.3 that the model based on moments can be expressed as a *moment generating function*.

Example 2.34 In Example 2.6, we found that random variable R has PMF

$$P_R(r) = \begin{cases} 1/4 & r = 0, \\ 3/4 & r = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{2.92}$$

In Example 2.26, we calculated $E[R] = \mu_R = 3/2$. What is the variance of R?

In order of increasing simplicity, we present three ways to compute Var$[R]$.

- From Definition 2.16, define

$$W = (R - \mu_R)^2 = (R - 3/2)^2 \tag{2.93}$$

The PMF of W is

$$P_W(w) = \begin{cases} 1/4 & w = (0 - 3/2)^2 = 9/4, \\ 3/4 & w = (2 - 3/2)^2 = 1/4, \\ 0 & \text{otherwise.} \end{cases} \tag{2.94}$$

Then

$$\operatorname{Var}[R] = E[W] = (1/4)(9/4) + (3/4)(1/4) = 3/4. \tag{2.95}$$

- Recall that Theorem 2.10 produces the same result without requiring the derivation of $P_W(w)$.

$$\operatorname{Var}[R] = E\left[(R - \mu_R)^2\right] \tag{2.96}$$

$$= (0 - 3/2)^2 P_R(0) + (2 - 3/2)^2 P_R(2) = 3/4 \tag{2.97}$$

- To apply Theorem 2.13, we find that

$$E\left[R^2\right] = 0^2 P_R(0) + 2^2 P_R(2) = 3 \tag{2.98}$$

Thus Theorem 2.13 yields

$$\operatorname{Var}[R] = E\left[R^2\right] - \mu_R^2 = 3 - (3/2)^2 = 3/4 \tag{2.99}$$

Note that $(X - \mu_X)^2 \geq 0$. Therefore, its expected value is also nonnegative. That is, for any random variable X

$$\operatorname{Var}[X] \geq 0. \tag{2.100}$$

The following theorem is related to Theorem 2.12

Theorem 2.14

$$\operatorname{Var}[aX + b] = a^2 \operatorname{Var}[X]$$

Proof We let $Y = aX + b$ and apply Theorem 2.13. We first expand the second moment to obtain

$$E\left[Y^2\right] = E\left[a^2X^2 + 2abX + b^2\right] = a^2 E\left[X^2\right] + 2ab\mu_X + b^2. \tag{2.101}$$

Expanding the right side of Theorem 2.12 yields

$$\mu_Y^2 = a^2\mu_X^2 + 2ab\mu_x + b^2. \tag{2.102}$$

Because $\operatorname{Var}[Y] = E[Y^2] - \mu_Y^2$, Equations (2.101) and (2.102) imply that

$$\operatorname{Var}[Y] = a^2 E\left[X^2\right] - a^2\mu_X^2 = a^2(E\left[X^2\right] - \mu_X^2) = a^2 \operatorname{Var}[X]. \tag{2.103}$$

If we let $a = 0$ in this theorem, we have $\operatorname{Var}[b] = 0$ because there is no dispersion around the expected value of a constant. If we let $a = 1$, we have $\operatorname{Var}[X + b] = \operatorname{Var}[X]$ because shifting a random variable by a constant does not change the dispersion of outcomes around the expected value.

Example 2.35 A new fax machine automatically transmits an initial cover page that precedes the regular fax transmission of X information pages. Using this new machine, the number of pages in a fax is $Y = X + 1$. What are the expected value and variance of Y?

The expected number of transmitted pages is $E[Y] = E[X] + 1$. The variance of the number of pages sent is $\text{Var}[Y] = \text{Var}[X]$.

If we let $b = 0$ in Theorem 2.12, we have $\text{Var}[aX] = a^2 \text{Var}[X]$ and $\sigma_{aX} = a\sigma X$. Multiplying a random variable by a constant is equivalent to a scale change in the units of measurement of the random variable.

Example 2.36 In Example 2.30, the amplitude V in volts has PMF

$$P_V(v) = \begin{cases} 1/7 & v = -3, -2, \ldots, 3, \\ 0 & \text{otherwise.} \end{cases} \tag{2.104}$$

A new voltmeter records the amplitude U in millivolts. What is the variance of U?

Note that $U = 1000V$. To use Theorem 2.14, we first find the variance of V. The expected value of the amplitude is

$$\mu_V = 1/7[-3 + (-2) + (-1) + 0 + 1 + 2 + 3] = 0 \text{ volts.} \tag{2.105}$$

The second moment is

$$E\left[V^2\right] = 1/7[(-3)^2 + (-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2 + 3^2] = 4 \text{ volts}^2 \tag{2.106}$$

Therefore the variance is $\text{Var}[V] = E[V^2] - \mu_V^2 = 4$ volts^2. By Theorem 2.14,

$$\text{Var}[U] = 1000^2 \text{Var}[V] = 4{,}000{,}000 \text{ millivolts}^2. \tag{2.107}$$

The following theorem states the variances of the families of random variables defined in Section 2.3.

Theorem 2.15

(a) If X is Bernoulli (p), then $\text{Var}[X] = p(1 - p)$.

(b) If X is geometric (p), then $\text{Var}[X] = (1 - p)/p^2$.

(c) If X is binomial (n, p), then $\text{Var}[X] = np(1 - p)$.

(d) If X is Pascal (k, p), then $\text{Var}[X] = k(1 - p)/p^2$.

(e) If X is Poisson (α), then $\text{Var}[X] = \alpha$.

(f) If X is discrete uniform (k, l), then $\text{Var}[X] = (l - k)(l - k + 2)/12$.

Quiz 2.8 *In an experiment to monitor two calls, the PMF of N the number of voice calls, is*

$$P_N(n) = \begin{cases} 0.1 & n = 0, \\ 0.4 & n = 1, \\ 0.5 & n = 2, \\ 0 & \text{otherwise}. \end{cases} \qquad (2.108)$$

Find

(1) The expected value $E[N]$

(2) The second moment $E[N^2]$

(3) The variance $\text{Var}[N]$

(4) The standard deviation σ_N

2.9 Conditional Probability Mass Function

Recall from Section 1.5 that the conditional probability $P[A|B]$ is a number that expresses our new knowledge about the occurrence of event A, when we learn that another event B occurs. In this section, we consider event A to be the observation of a particular value of a random variable. That is, $A = \{X = x\}$. The conditioning event B contains information about X but not the precise value of X. For example, we might learn that $X \leq 33$ or that $|X| > 100$. In general, we learn of the occurrence of an event B that describes some property of X.

Example 2.37 Let N equal the number of bytes in a fax. A conditioning event might be the event I that the fax contains an image. A second kind of conditioning would be the event $\{N > 10{,}000\}$ which tells us that the fax required more than $10{,}000$ bytes. Both events I and $\{N > 10{,}000\}$ give us information that the fax is likely to have many bytes.

The occurrence of the conditioning event B changes the probabilities of the event $\{X = x\}$. Given this information and a probability model for our experiment, we can use Definition 1.6 to find the conditional probabilities

$$P[A|B] = P[X = x|B] \qquad (2.109)$$

for all real numbers x. This collection of probabilities is a function of x. It is the *conditional probability mass function* of random variable X, given that B occurred.

Definition 2.19 ***Conditional PMF***
Given the event B, with $P[B] > 0$, the ***conditional probability mass function*** *of X is*

$$P_{X|B}(x) = P[X = x|B].$$

Here we extend our notation convention for probability mass functions. The name of a PMF is the letter P with a subscript containing the name of the random variable. For a conditional PMF, the subscript contains the name of the random variable followed by a vertical bar followed by a statement of the conditioning event. The argument of the function is usually the lowercase letter corresponding to the variable name. The argument is a dummy variable. It could be any letter, so that $P_{X|B}(x)$ is the same function as $P_{X|B}(u)$. Sometimes we write the function with no specified argument at all, $P_{X|B}(\cdot)$.

In some applications, we begin with a set of conditional PMFs, $P_{X|B_i}(x), i = 1, 2, \ldots, m$, where $B_1, B_2, \ldots, B_m$ is an event space. We then use the law of total probability to find the PMF $P_X(x)$.

Theorem 2.16 *A random variable X resulting from an experiment with event space $B_1, \ldots, B_m$ has PMF*

$$P_X(x) = \sum_{i=1}^{m} P_{X|B_i}(x)\, P[B_i].$$

Proof The theorem follows directly from Theorem 1.10 with A denoting the event $\{X = x\}$.

Example 2.38 Let X denote the number of additional years that a randomly chosen 70 year old person will live. If the person has high blood pressure, denoted as event H, then X is a geometric ($p = 0.1$) random variable. Otherwise, if the person's blood pressure is regular, event R, then X has a geometric ($p = 0.05$) PMF with parameter. Find the conditional PMFs $P_{X|H}(x)$ and $P_{X|R}(x)$. If 40 percent of all seventy year olds have high blood pressure, what is the PMF of X?

. . . .

The problem statement specifies the conditional PMFs in words. Mathematically, the two conditional PMFs are

$$P_{X|H}(x) = \begin{cases} 0.1(0.9)^{x-1} & x = 1, 2, \ldots \\ 0 & \text{otherwise,} \end{cases} \tag{2.110}$$

$$P_{X|R}(x) = \begin{cases} 0.05(0.95)^{x-1} & x = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2.111}$$

Since H, R is an event space, we can use Theorem 2.16 to write

$$P_X(x) = P_{X|H}(x)\, P[H] + P_{X|R}(x)\, P[R] \tag{2.112}$$

$$= \begin{cases} (0.4)(0.1)(0.9)^{x-1} + (0.6)(0.05)(0.95)^{x-1} & x = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2.113}$$

When a conditioning event $B \subset S_X$, the PMF $P_X(x)$ determines both the probability of B as well as the conditional PMF:

$$P_{X|B}(x) = \frac{P[X = x, B]}{P[B]}. \tag{2.114}$$

Now either the event $X = x$ is contained in the event B or it is not. If $x \in B$, then $\{X = x\} \cap B = \{X = x\}$ and $P[X = x, B] = P_X(x)$. Otherwise, if $x \notin B$, then $\{X = x\} \cap B = \phi$ and $P[X = x, B] = 0$. The next theorem uses Equation (2.114) to calculate the conditional PMF.

Theorem 2.17

$$P_{X|B}(x) = \begin{cases} \dfrac{P_X(x)}{P[B]} & x \in B, \\ 0 & otherwise. \end{cases}$$

The theorem states that when we learn that an outcome $x \in B$, the probabilities of all $x \notin B$ are zero in our conditional model and the probabilities of all $x \in B$ are proportionally higher than they were before we learned $x \in B$.

Example 2.39 In the probability model of Example 2.29, the length of a fax X has PMF

$$P_X(x) = \begin{cases} 0.15 & x = 1, 2, 3, 4, \\ 0.1 & x = 5, 6, 7, 8, \\ 0 & \text{otherwise.} \end{cases} \tag{2.115}$$

Suppose the company has two fax machines, one for faxes shorter than five pages and the other for faxes that have five or more pages. What is the PMF of fax length in the second machine?

Relative to $P_X(x)$, we seek a conditional PMF. The condition is $x \in L$ where $L = \{5, 6, 7, 8\}$. From Theorem 2.17,

$$P_{X|L}(x) = \begin{cases} \dfrac{P_X(x)}{P[L]} & x = 5, 6, 7, 8, \\ 0 & \text{otherwise.} \end{cases} \tag{2.116}$$

From the definition of L, we have

$$P[L] = \sum_{x=5}^{8} P_X(x) = 0.4. \tag{2.117}$$

With $P_X(x) = 0.1$ for $x \in L$,

$$P_{X|L}(x) = \begin{cases} 0.1/0.4 = 0.25 & x = 5, 6, 7, 8, \\ 0 & \text{otherwise.} \end{cases} \tag{2.118}$$

Thus the lengths of long faxes are equally likely. Among the long faxes, each length has probability 0.25.

Sometimes instead of a letter such as B or L that denotes the subset of S_X that forms the condition, we write the condition itself in the PMF. In the preceding example we could use the notation $P_{X|X\geq 5}(x)$ for the conditional PMF.

Example 2.40 Suppose X, the time in integer minutes you must wait for a bus, has the uniform PMF

$$P_X(x) = \begin{cases} 1/20 & x = 1, 2, \ldots, 20, \\ 0 & \text{otherwise.} \end{cases} \tag{2.119}$$

Suppose the bus has not arrived by the eighth minute, what is the conditional PMF of your waiting time X?

Let A denote the event $X > 8$. Observing that $P[A] = 12/20$, we can write the conditional PMF of X as

$$P_{X|X>8}(x) = \begin{cases} \dfrac{1/20}{12/20} = \dfrac{1}{12} & x = 9, 10, \ldots, 20, \\ 0 & \text{otherwise.} \end{cases} \tag{2.120}$$

Note that $P_{X|B}(x)$ is a perfectly respectable PMF. Because the conditioning event B tells us that all possible outcomes are in B, we rewrite Theorem 2.1 using B in place of S.

Theorem 2.18

(a) For any $x \in B$, $P_{X|B}(x) \geq 0$.

(b) $\sum_{x \in B} P_{X|B}(x) = 1$.

(c) For any event $C \subset B$, $P[C|B]$, the conditional probability that X is in the set C, is

$$P[C|B] = \sum_{x \in C} P_{X|B}(x).$$

Therefore, we can compute averages of the conditional random variable $X|B$ and averages of functions of $X|B$ in the same way that we compute averages of X. The only difference is that we use the conditional PMF $P_{X|B}(\cdot)$ in place of $P_X(\cdot)$.

Definition 2.20 ***Conditional Expected Value***

*The **conditional expected value** of random variable X given condition B is*

$$E[X|B] = \mu_{X|B} = \sum_{x \in B} x P_{X|B}(x).$$

When we are given a family of conditional probability models $P_{X|B_i}(x)$ for an event space $B_1, \ldots, B_m$, we can compute the expected value $E[X]$ in terms of the conditional expected values $E[X|B_i]$.

Theorem 2.19 *For a random variable X resulting from an experiment with event space $B_1, \ldots, B_m$,*

$$E[X] = \sum_{i=1}^{m} E[X|B_i] P[B_i].$$

Proof Since $E[X] = \sum_x x P_X(x)$, we can use Theorem 2.16 to write

$$E[X] = \sum_x x \sum_{i=1}^{m} P_{X|B_i}(x) P[B_i] \tag{2.121}$$

$$= \sum_{i=1}^{m} P[B_i] \sum_x x P_{X|B_i}(x) = \sum_{i=1}^{m} P[B_i] E[X|B_i]. \tag{2.122}$$

For a derived random variable $Y = g(X)$, we have the equivalent of Theorem 2.10.

Theorem 2.20 *The conditional expected value of $Y = g(X)$ given condition B is*

$$E[Y|B] = E[g(X)|B] = \sum_{x \in B} g(x) P_{X|B}(x).$$

It follows that the conditional variance and conditional standard deviation conform to Definitions 2.16 and 2.17 with $X|B$ replacing X.

Example 2.41 Find the conditional expected value, the conditional variance, and the conditional standard deviation for the long faxes defined in Example 2.39.

$$E[X|L] = \mu_{X|L} = \sum_{x=5}^{8} x P_{X|L}(x) = 0.25 \sum_{x=5}^{8} x = 6.5 \text{ pages} \tag{2.123}$$

$$E\left[X^2|L\right] = 0.25 \sum_{x=5}^{8} x^2 = 43.5 \text{ pages}^2 \tag{2.124}$$

$$\text{Var}[X|L] = E\left[X^2|L\right] - \mu_{X|L}^2 = 1.25 \text{ pages}^2 \tag{2.125}$$

$$\sigma_{X|L} = \sqrt{\text{Var}[X|L]} = 1.12 \text{ pages} \tag{2.126}$$

Quiz 2.9 *On the Internet, data is transmitted in packets. In a simple model for World Wide Web traffic, the number of packets N needed to transmit a Web page depends on whether the page has graphic images. If the page has images (event I), then N is uniformly distributed between 1 and 50 packets. If the page is just text (event T), then N is uniform between 1 and 5 packets. Assuming a page has images with probability 1/4, find the*

(1) conditional PMF $P_{N|I}(n)$ *(2) conditional PMF $P_{N|T}(n)$*

(3) PMF $P_N(n)$ *(4) conditional PMF $P_{N|N\le 10}(n)$*

(5) conditional expected value $E[N|N \le 10]$ *(6) conditional variance $\text{Var}[N|N \le 10]$*

2.10 MATLAB

For discrete random variables, this section will develop a variety of ways to use MATLAB. We start by calculating probabilities for any finite random variable with arbitrary PMF $P_X(x)$. We then compute PMFs and CDFs for the families of random variables introduced in Section 2.3. Based on the calculation of the CDF, we then develop a method for generating random sample values. Generating a random sample is a simple simulation of a experiment that produces the corresponding random variable. In subsequent chapters, we will see that MATLAB functions that generate random samples are building blocks for the simulation of more complex systems. The MATLAB functions described in this section can be downloaded from the companion Web site.

PMFs and CDFs

For the most part, the PMF and CDF functions are straightforward. We start with a simple finite discrete random variable X defined by the set of sample values $S_X = \{s_1, \ldots, s_n\}$ and corresponding probabilities $p_i = P_X(s_i) = P[X = s_i]$. In MATLAB, we represent the sample space of X by the vector $\mathbf{s} = \begin{bmatrix} s_1 & \cdots & s_n \end{bmatrix}'$ and the corresponding probabilities by the vector $\mathbf{p} = \begin{bmatrix} p_1 & \cdots & p_n \end{bmatrix}'$.[2] The function `y=finitepmf(sx,px,x)` generates the probabilities of the elements of the m-dimensional vector $\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}'$. The output is $\mathbf{y} = \begin{bmatrix} y_1 & \cdots & y_m \end{bmatrix}'$ where $y_i = P_X(x_i)$. That is, for each requested x_i, `finitepmf` returns the value $P_X(x_i)$. If x_i is not in the sample space of X, $y_i = 0$.

Example 2.42 In Example 2.29, the random variable X, the number of pages in a fax, has PMF

$$P_X(x) = \begin{cases} 0.15 & x = 1, 2, 3, 4, \\ 0.1 & x = 5, 6, 7, 8, \\ 0 & \text{otherwise.} \end{cases} \tag{2.127}$$

Write a MATLAB function that calculates $P_X(x)$. Calculate the probability of x_i pages for $x_1 = 2$, $x_2 = 2.5$, and $x_3 = 6$.

The MATLAB function `fax3pmf(x)` implements $P_X(x)$. We can then use `fax3pmf` to calculate the desired probabilities:

```
function y=fax3pmf(x)
s=(1:8)';
p=[0.15*ones(4,1); 0.1*ones(4,1)];
y=finitepmf(s,p,x);
```

```
» fax3pmf([2 2.5 6])'
ans =
   0.1500   0   0.1000
```

We also can use MATLAB to calculate PMFs for common random variables. Although a PMF $P_X(x)$ is a scalar function of one variable, the easy way that MATLAB handles vectors makes it desirable to extend our MATLAB PMF functions to allow vector input. That is, if `y=xpmf(x)` implements $P_X(x)$, then for a vector input `x`, we produce a vector output `y` such that `y(i)=xpmf(x(i))`. That is, for vector input $\mathbf{x}$, the output vector $\mathbf{y}$ is defined by $y_i = P_X(x_i)$.

Example 2.43 Write a MATLAB function `geometricpmf(p,x)` to calculate $P_X(x)$ for a geometric (p) random variable.

[2] Although column vectors are supposed to appear as columns, we generally write a column vector $\mathbf{x}$ in the form of a transposed row vector $\begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}'$ to save space.

```
function pmf=geometricpmf(p,x)
%geometric(p) rv X
%out: pmf(i)=Prob[X=x(i)]
x=x(:);
pmf= p*((1-p).^(x-1));
pmf= (x>0).*(x==floor(x)).*pmf;
```

In `geometricpmf.m`, the last line ensures that values $x_i \notin S_X$ are assigned zero probability. Because `x=x(:)` reshapes `x` to be a column vector, the output `pmf` is always a column vector.

Example 2.44 Write a MATLAB function that calculates the PMF of a Poisson (α) random variable.

For an integer x, we could calculate $P_X(x)$ by the direct calculation

```
px= ((alpha^x)*exp(-alpha*x))/factorial(x)
```

This will yield the right answer as long as the argument x for the factorial function is not too large. In MATLAB version 6, `factorial(171)` causes an overflow. In addition, for $a > 1$, calculating the ratio $a^x/x!$ for large x can cause numerical problems because both a^x and $x!$ will be very large numbers, possibly with a small quotient. Another shortcoming of the direct calculation is apparent if you want to calculate $P_X(x)$ for the set of possible values $\mathbf{x} = [0, 1, \ldots, n]$. Calculating factorials is a lot of work for a computer and the direct approach fails to exploit the fact that if we have already calculated $(x-1)!$, we can easily compute $x! = x \cdot (x-1)!$. A more efficient calculation makes use of the observation

$$P_X(x) = \frac{a^x e^{-a}}{x!} = \frac{a}{x} P_X(x-1). \tag{2.128}$$

The `poissonpmf.m` function uses Equation (2.128) to calculate $P_X(x)$. Even this code is not perfect because MATLAB has limited range.

```
function pmf=poissonpmf(alpha,x)
%output: pmf(i)=P[X=x(i)]
x=x(:); k=(1:max(x))';
ip=[1;((alpha*ones(size(k)))./k)];
pb=exp(-alpha)*cumprod(ip);
  %pb= [P(X=0)...P(X=n)]
pmf=pb(x+1); %pb(1)=P[X=0]
pmf=(x>=0).*(x==floor(x)).*pmf;
  %pmf(i)=0 for zero-prob x(i)
```

Note that `exp(-alpha)=0` for `alpha` > 745.13. For these large values of `alpha`, `poissonpmf(alpha,x)` will return zero for all `x`. Problem 2.10.8 outlines a solution that is used in the version of `poissonpmf.m` on the companion website.

For the Poisson CDF, there is no simple way to avoid summing the PMF. The following example shows an implementation of the Poisson CDF. The code for a CDF tends to be more complicated than that for a PMF because if x is not an integer, $F_X(x)$ may still be nonzero. Other CDFs are easily developed following the same approach.

Example 2.45 Write a MATLAB function that calculates the CDF of a Poisson random variable.

MATLAB **Functions**

PMF	CDF	Random Sample
`finitepmf(sx,p,x)`	`finitecdf(sx,p,x)`	`finiterv(sx,p,m)`
`bernoullipmf(p,x)`	`bernoullicdf(p,x)`	`bernoullirv(p,m)`
`binomialpmf(n,p,x)`	`binomialcdf(n,p,x)`	`binomialrv(n,p,m)`
`geometricpmf(p,x)`	`geometriccdf(p,x)`	`geometricrv(p,m)`
`pascalpmf(k,p,x)`	`pascalcdf(k,p,x)`	`pascalrv(k,p,m)`
`poissonpmf(alpha,x)`	`poissoncdf(alpha,x)`	`poissonrv(alpha,m)`
`duniformpmf(k,l,x)`	`duniformcdf(k,l,x)`	`duniformrv(k,l,m)`

Table 2.1 MATLAB functions for discrete random variables.

```
function cdf=poissoncdf(alpha,x)
%output cdf(i)=Prob[X<=x(i)]
x=floor(x(:));
sx=0:max(x);
cdf=cumsum(poissonpmf(alpha,sx));
  %cdf from 0 to max(x)
okx=(x>=0);%x(i)<0 -> cdf=0
x=(okx.*x);%set negative x(i)=0
cdf= okx.*cdf(x+1);
  %cdf=0 for x(i)<0
```

Here we present the MATLAB code for the Poisson CDF. Since a Poisson random variable X is always integer valued, we observe that $F_X(x) = F_X(\lfloor x \rfloor)$ where $\lfloor x \rfloor$, equivalent to `floor(x)` in MATLAB, denotes the largest integer less than or equal to x.

Example 2.46 Recall in Example 2.19 that a website has on average $\lambda = 2$ hits per second. What is the probability of no more than 130 hits in one minute? What is the probability of more than 110 hits in one minute?

Let M equal the number of hits in one minute (60 seconds). Note that M is a Poisson (α) random variable with $\alpha = 2 \times 60 = 120$ hits. The PMF of M is

$$P_M(m) = \begin{cases} (120)^m e^{-120}/m! & m = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2.129}$$

```
» poissoncdf(120,130)
ans =
    0.8315
» 1-poissoncdf(120,110)
ans =
    0.8061
```

The MATLAB solution shown on the left executes the following math calculations:

$$P[M \le 130] = \sum_{m=0}^{130} P_M(m) \tag{2.130}$$

$$P[M > 110] = 1 - P[M \le 110] \tag{2.131}$$

$$= 1 - \sum_{m=0}^{110} P_M(m) \tag{2.132}$$

Generating Random Samples

So far, we have generated distribution functions, PMFs or CDFs, for families of random variables. Now we tackle the more difficult task of generating sample values of random variables. As in Chapter 1, we use `rand()` as a source of randomness. Let $R =$ `rand(1)`. Recall that `rand(1)` simulates an experiment that is equally likely to produce any real number in the interval [0, 1]. We will learn in Chapter 3 that to express this idea in mathematics, we say that for any interval $[a, b] \subset [0, 1]$,

$$P[a < R \leq b] = b - a. \tag{2.133}$$

For example, $P[0.4 < R \leq 0.53] = 0.13$. Now suppose we wish to generate samples of discrete random variable K with $S_K = \{0, 1, \ldots\}$. Since $0 \leq F_K(k-1) \leq F_K(k) \leq 1$, for all k, we observe that

$$P[F_K(k-1) < R \leq F_K(k)] = F_K(k) - F_K(k-1) = P_K(k) \tag{2.134}$$

This fact leads to the following approach (as shown in pseudocode) to using `rand()` to produce a sample of random variable K:

Random Sample of random variable K

Generate $R =$ `rand(1)`
Find k^* such that $F_K(k^*-1) < R \leq F_K(k^*)$
Set $K = k^*$

A MATLAB function that uses `rand()` in this way simulates an experiment that produces samples of random variable K. Generally, this implies that before we can produce a sample of random variable K, we need to generate the CDF of K. We can reuse the work of this computation by defining our MATLAB functions such as `geometricrv(p,m)` to generate m sample values each time. We now present the details associated with generating binomial random variables.

Example 2.47 Write a MATLAB function that generates m samples of a binomial (n, p) random variable.

```
function x=binomialrv(n,p,m)
% m binomial(n,p) samples
r=rand(m,1);
cdf=binomialcdf(n,p,0:n);
x=count(cdf,r);
```

For vectors `x` and `y`, the function `c=count(x,y)` returns a vector `c` such that `c(i)` is the number of elements of `x` that are less than or equal to $y(i)$.

In terms of our earlier pseudocode, $k^* =$ `count(cdf,r)`. If `count(cdf,r)` $=$ 0, then $r \leq P_X(0)$ and $k^* = 0$.

Generating binomial random variables is easy because the range is simply $\{0, \ldots, n\}$ and the minimum value is zero. You will see that the MATLAB code for `geometricrv()`, `poissonrv()`, and `pascalrv()` is slightly more complicated because we need to generate enough terms of the CDF to ensure that we find k^*.

Table 2.1 summarizes a collection of functions for the families of random variables introduced in Section 2.3. For each family, there is the `pmf` function for calculating values

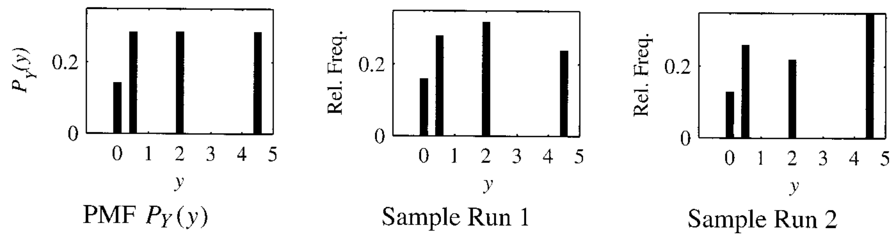


Figure 2.2 The PMF of Y and the relative frequencies found in two sample runs of `voltpower(100)`. Note that in each run, the relative frequencies are close to (but not exactly equal to) the corresponding PMF.

of the PMF, the `cdf` function for calculating values of the CDF, and the `rv` function for generating random samples. In each function description, `x` denotes a column vector $\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}'$. The `pmf` function output is a vector $\mathbf{y}$ such that $y_i = P_X(x_i)$. The `cdf` function output is a vector $\mathbf{y}$ such that $y_i = F_X(x_i)$. The `rv` function output is a vector $\mathbf{X} = \begin{bmatrix} X_1 & \cdots & X_m \end{bmatrix}'$ such that each X_i is a sample value of the random variable X. If $m = 1$, then the output is a single sample value of random variable X.

We present an additional example, partly because it demonstrates some useful MATLAB functions, and also because it shows how to generate relative frequency data for our random variable generators.

Example 2.48 Simulate $n = 1000$ trials of the experiment producing the power measurement Y in Example 2.30. Compare the relative frequency of each $y \in S_Y$ to $P_Y(y)$.

In `voltpower.m`, we first generate n samples of the voltage V. For each sample, we calculate $Y = V^2/2$.

```
function voltpower(n)
v=duniformrv(-3,3,n);
y=(v.^2)/2;
yrange=0:max(y);
yfreq=(hist(y,yrange)/n)';
pmfplot(yrange,yfreq);
```

As in Example 1.47, the function `hist(y,yrange)` produces a vector with jth element equal to the number of occurences of `yrange(j)` in the vector `y`. The function `pmfplot.m` is a utility for producing PMF bar plots in the style of this text.

Figure 2.2 shows the corresponding PMF along with the output of two runs of `voltpower(100)`.

Derived Random Variables

MATLAB can also calculate PMFs and CDFs of derived random variables. For this section, we assume X is a finite random variable with sample space $S_X = \{x_1, \ldots, x_n\}$ such that $P_X(x_i) = p_i$. We represent the properties of X by the vectors $\mathbf{s}_X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}'$ and $\mathbf{p}_X = \begin{bmatrix} p_1 & \cdots & p_n \end{bmatrix}'$. In MATLAB notation, `sx` and `px` represent the vectors $\mathbf{s}_X$ and $\mathbf{p}_X$.

For derived random variables, we exploit a feature of `finitepmf(sx,px,x)` that allows the elements of `sx` to be repeated. Essentially, we use (`sx`, `px`), or equivalently $(\mathbf{s}, \mathbf{p})$, to represent a random variable X described by the following experimental procedure:

Finite PMF

Roll an n-sided die such that side i has probability p_i.
If side j appears, set $X = x_j$.

A consequence of this approach is that if $x_2 = 3$ and $x_5 = 3$, then the probability of observing $X = 3$ is $P_X(3) = p_2 + p_5$.

Example 2.49

```
» sx=[1 3 5 7 3];
» px=[0.1 0.2 0.2 0.3 0.2];
» pmfx=finitepmf(sx,px,1:7);
» pmfx'
ans =
   0.10 0 0.40 0 0.20 0 0.30
```

The function `finitepmf()` accounts for multiple occurrences of a sample value. In particular,

`pmfx(3)=px(2)+px(5)=0.4.`

It may seem unnecessary and perhaps even bizarre to allow these repeated values. However, we see in the next example that it is quite convenient for derived random variables $Y = g(X)$ with the property that $g(x_i)$ is the same for multiple x_i. Although the next example was simple enough to solve by hand, it is instructive to use MATLAB to do the work.

Example 2.50 Recall that in Example 2.29 that the number of pages X in a fax and the cost $Y = g(X)$ of sending a fax were described by

$$P_X(x) = \begin{cases} 0.15 & x = 1, 2, 3, 4, \\ 0.1 & x = 5, 6, 7, 8, \\ 0 & \text{otherwise}, \end{cases} \qquad Y = \begin{cases} 10.5X - 0.5X^2 & 1 \le X \le 5, \\ 50 & 6 \le X \le 10. \end{cases}$$

Use MATLAB to calculate the PMF of Y.

```
%fax3y.m
sx=(1:8)';
px=[0.15*ones(4,1); ...
       0.1*ones(4,1)];
gx=(sx<=5).* ...
    (10.5*sx-0.5*(sx.^2))...
   + ((sx>5).*50);
sy=unique(gx);
py=finitepmf(gx,px,sy);
```

The vector `gx` is the mapping $g(x)$ for each $x \in S_X$. In `gx`, the element 50 appears three times, corresponding to $x = 6$, $x = 7$, and $x = 8$. The function `sy=unique(gx)` extracts the unique elements of `gx` while `finitepmf(gx,px,sy)` calculates the probability of each element of `sy`.

Conditioning

MATLAB also provides the `find` function to identify conditions. We use the `find` function to calculate conditional PMFs for finite random variables.

Example 2.51 Repeating Example 2.39, find the conditional PMF for the length X of a fax given event L that the fax is long with $X \geq 5$ pages.

```
sx=(1:8)';
px=[0.15*ones(4,1);...
      0.1*ones(4,1)];
sxL=unique(find(sx>=5));
pL=sum(finitepmf(sx,px,sxL));
pxL=finitepmf(sx,px,sxL)/pL;
```

With random variable X defined by `sx` and `px` as in Example 2.50, this code solves this problem. The vector `sxL` identifies the event L, `pL` is the probability $P[L]$, and `pxL` is the vector of probabilities $P_{X|L}(x_i)$ for each $x_i \in L$.

Quiz 2.10 *In Section 2.5, it was argued that the average*

$$m_n = \frac{1}{n}\sum_{i=1}^{n} x(i)$$

of samples $x(1), x(2), \ldots, x(n)$ *of a random variable* X *will converge to* $E[X]$ *as* n *becomes large. For a discrete uniform* $(0, 10)$ *random variable* X*, we will use* MATLAB *to examine this convergence.*

(1) For 100 *sample values of* X*, plot the sequence* $m_1, m_2, \ldots, m_{100}$*. Repeat this experiment five times, plotting all five* m_n *curves on common axes.*

(2) Repeat part (a) for 1000 *sample values of* X*.*

Chapter Summary

With all of the concepts and formulas introduced in this chapter, there is a high probability that the beginning student will be confused at this point. Part of the problem is that we are dealing with several different mathematical entities including random variables, probability functions, and parameters. Before plugging numbers or symbols into a formula, it is good to know what the entities are.

- *The random variable* X transforms outcomes of an experiment to real numbers. Note that X is the name of the random variable. A possible observation is x, which is a number. S_X is the range of X, the set of all possible observations x.
- *The PMF* $P_X(x)$ is a function that contains the probability model of the random variable X. The PMF gives the probability of observing any x. $P_X(\cdot)$ contains our information about the randomness of X.
- *The expected value* $E[X] = \mu_X$ and the variance $\text{Var}[X]$ are numbers that describe the entire probability model. Mathematically, each is a property of the PMF $P_X(\cdot)$. The expected value is a typical value of the random variable. The variance describes the dispersion of sample values about the expected value.

- *A function of a random variable* $Y = g(X)$ transforms the random variable X into a different random variable Y. For each observation $X = x$, $g(\cdot)$ is a rule that tells you how to calculate $y = g(x)$, a sample value of Y.

  Although $P_X(\cdot)$ and $g(\cdot)$ are both mathematical functions, they serve different purposes here. $P_X(\cdot)$ describes the randomness in an experiment. On the other hand, $g(\cdot)$ is a rule for obtaining a new random variable from a random variable you have observed.

- *The Conditional PMF* $P_{X|B}(x)$ is the probability model that we obtain when we gain partial knowledge of the outcome of an experiment. The partial knowledge is that the outcome $x \in B \subset S_X$. The conditional probability model has its own expected value, $E[X|B]$, and its own variance, $\mathrm{Var}[X|B]$.

Problems

Difficulty: Easy Moderate Difficult Experts Only

2.2.1 The random variable N has PMF

$$P_N(n) = \begin{cases} c(1/2)^n & n = 0, 1, 2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the value of the constant c?

(b) What is $P[N \leq 1]$?

2.2.2 For random variables X and R defined in Example 2.5, find $P_X(x)$ and $P_R(r)$. In addition, find the following probabilities:

(a) $P[X = 0]$

(b) $P[X < 3]$

(c) $P[R > 1]$

2.2.3 The random variable V has PMF

$$P_V(v) = \begin{cases} cv^2 & v = 1, 2, 3, 4, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the value of the constant c.

(b) Find $P[V \in \{u^2 | u = 1, 2, 3, \cdots\}]$.

(c) Find the probability that V is an even number.

(d) Find $P[V > 2]$.

2.2.4 The random variable X has PMF

$$P_X(x) = \begin{cases} c/x & x = 2, 4, 8, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the value of the constant c?

(b) What is $P[X = 4]$?

(c) What is $P[X < 4]$?

(d) What is $P[3 \leq X \leq 9]$?

2.2.5 In college basketball, when a player is fouled while not in the act of shooting and the opposing team is "in the penalty," the player is awarded a "1 and 1." In the 1 and 1, the player is awarded one free throw and if that free throw goes in the player is awarded a second free throw. Find the PMF of Y, the number of points scored in a 1 and 1 given that any free throw goes in with probability p, independent of any other free throw.

2.2.6 You are manager of a ticket agency that sells concert tickets. You assume that people will call three times in an attempt to buy tickets and then give up. You want to make sure that you are able to serve at least 95% of the people who want tickets. Let p be the probability that a caller gets through to your ticket agency. What is the minimum value of p necessary to meet your goal.

2.2.7 In the ticket agency of Problem 2.2.6, each telephone ticket agent is available to receive a call with probability 0.2. If all agents are busy when someone calls, the caller hears a busy signal. What is the minimum number of agents that you have to hire to meet your goal of serving 95% of the customers who want tickets?

2.2.8 Suppose when a baseball player gets a hit, a single is twice as likely as a double which is twice as likely as a triple which is twice as likely as a home run. Also, the player's batting average, *i.e.*, the probability the player gets a hit, is 0.300. Let B denote the number of bases touched safely during an at-bat. For example, $B = 0$ when the player makes an out,

$B = 1$ on a single, and so on. What is the PMF of B?

2.2.9 When someone presses "SEND" on a cellular phone, the phone attempts to set up a call by transmitting a "SETUP" message to a nearby base station. The phone waits for a response and if none arrives within 0.5 seconds it tries again. If it doesn't get a response after $n = 6$ tries the phone stops transmitting messages and generates a busy signal.

(a) Draw a tree diagram that describes the call setup procedure.

(b) If all transmissions are independent and the probability is p that a "SETUP" message will get through, what is the PMF of K, the number of messages transmitted in a call attempt?

(c) What is the probability that the phone will generate a busy signal?

(d) As manager of a cellular phone system, you want the probability of a busy signal to be less than 0.02 If $p = 0.9$, what is the minimum value of n necessary to achieve your goal?

2.3.1 In a package of M&Ms, Y, the number of yellow M&Ms, is uniformly distributed between 5 and 15.

(a) What is the PMF of Y?

(b) What is $P[Y < 10]$?

(c) What is $P[Y > 12]$?

(d) What is $P[8 \leq Y \leq 12]$?

2.3.2 When a conventional paging system transmits a message, the probability that the message will be received by the pager it is sent to is p. To be confident that a message is received at least once, a system transmits the message n times.

(a) Assuming all transmissions are independent, what is the PMF of K, the number of times the pager receives the same message?

(b) Assume $p = 0.8$. What is the minimum value of n that produces a probability of 0.95 of receiving the message at least once?

2.3.3 When you go fishing, you attach m hooks to your line. Each time you cast your line, each hook will be swallowed by a fish with probability h, independent of whether any other hook is swallowed. What is the PMF of K, the number of fish that are hooked on a single cast of the line?

2.3.4 Anytime a child throws a Frisbee, the child's dog catches the Frisbee with probability p, independent of whether the Frisbee is caught on any previous throw. When the dog catches the Frisbee, it runs away with the Frisbee, never to be seen again. The child continues to throw the Frisbee until the dog catches it. Let X denote the number of times the Frisbee is thrown.

(a) What is the PMF $P_X(x)$?

(b) If $p = 0.2$, what is the probability that the child will throw the Frisbee more than four times?

2.3.5 When a two-way paging system transmits a message, the probability that the message will be received by the pager it is sent to is p. When the pager receives the message, it transmits an acknowledgment signal (ACK) to the paging system. If the paging system does not receive the ACK, it sends the message again.

(a) What is the PMF of N, the number of times the system sends the same message?

(b) The paging company wants to limit the number of times it has to send the same message. It has a goal of $P[N \leq 3] \geq 0.95$. What is the minimum value of p necessary to achieve the goal?

2.3.6 The number of bits B in a fax transmission is a geometric ($p = 2.5 \cdot 10^{-5}$) random variable. What is the probability $P[B > 500{,}000]$ that a fax has over 500,000 bits?

2.3.7 The number of buses that arrive at a bus stop in T minutes is a Poisson random variable B with expected value $T/5$.

(a) What is the PMF of B, the number of buses that arrive in T minutes?

(b) What is the probability that in a two-minute interval, three buses will arrive?

(c) What is the probability of no buses arriving in a 10-minute interval?

(d) How much time should you allow so that with probability 0.99 at least one bus arrives?

2.3.8 In a wireless automatic meter reading system, a base station sends out a wake-up signal to nearby electric meters. On hearing the wake-up signal, a meter transmits a message indicating the electric usage. Each message is repeated eight times.

(a) If a single transmission of a message is successful with probability p, what is the PMF of N, the number of successful message transmissions?

(b) I is an indicator random variable such that $I = 1$ if at least one message is transmitted successfully; otherwise $I = 0$. Find the PMF of I.

2.3.9 A Zipf $(n, \alpha = 1)$ random variable X has PMF

$$P_X(x) = \begin{cases} c(n)/x & x = 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

The constant $c(n)$ is set so that $\sum_{x=1}^{n} P_X(x) = 1$. Calculate $c(n)$ for $n = 1, 2, \ldots, 6$.

2.3.10 A radio station gives a pair of concert tickets to the sixth caller who knows the birthday of the performer. For each person who calls, the probability is 0.75 of knowing the performer's birthday. All calls are independent.

(a) What is the PMF of L, the number of calls necessary to find the winner?

(b) What is the probability of finding the winner on the tenth call?

(c) What is the probability that the station will need nine or more calls to find a winner?

2.3.11 In a packet voice communications system, a source transmits packets containing digitized speech to a receiver. Because transmission errors occasionally occur, an acknowledgment (ACK) or a nonacknowledgment (NAK) is transmitted back to the source to indicate the status of each received packet. When the transmitter gets a NAK, the packet is retransmitted. Voice packets are delay sensitive and a packet can be transmitted a maximum of d times. If a packet transmission is an independent Bernoulli trial with success probability p, what is the PMF of T, the number of times a packet is transmitted?

2.3.12 Suppose each day (starting on day 1) you buy one lottery ticket with probability 1/2; otherwise, you buy no tickets. A ticket is a winner with probability p independent of the outcome of all other tickets. Let N_i be the event that on day i you do *not* buy a ticket. Let W_i be the event that on day i, you buy a winning ticket. Let L_i be the event that on day i you buy a losing ticket.

(a) What are $P[W_{33}]$, $P[L_{87}]$, and $P[N_{99}]$?

(b) Let K be the number of the day on which you buy your first lottery ticket. Find the PMF $P_K(k)$.

(c) Find the PMF of R, the number of losing lottery tickets you have purchased in m days.

(d) Let D be the number of the day on which you buy your jth losing ticket. What is $P_D(d)$? Hint: If you buy your jth losing ticket on day d, how many losers did you have after $d - 1$ days?

2.3.13 The Sixers and the Celtics play a best out of five playoff series. The series ends as soon as one of the teams has won three games. Assume that either team is equally likely to win any game independently of any other game played. Find

(a) The PMF $P_N(n)$ for the total number N of games played in the series;

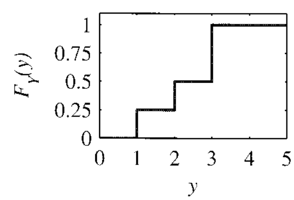
(b) The PMF $P_W(w)$ for the number W of Celtic wins in the series;

(c) The PMF $P_L(l)$ for the number L of Celtic losses in the series.

2.3.14 For a binomial random variable K representing the number of successes in n trials, $\sum_{k=0}^{n} P_K(k) = 1$. Use this fact to prove the binomial theorem for any $a > 0$ and $b > 0$. That is, show that

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

2.4.1 Discrete random variable Y has the CDF $F_Y(y)$ as shown:



Use the CDF to find the following probabilities:

(a) $P[Y < 1]$

(b) $P[Y \le 1]$

(c) $P[Y > 2]$

(d) $P[Y \ge 2]$

(e) $P[Y = 1]$

(f) $P[Y = 3]$

(g) $P_Y(y)$

2.4.2 The random variable X has CDF

$$F_X(x) = \begin{cases} 0 & x < -1, \\ 0.2 & -1 \le x < 0, \\ 0.7 & 0 \le x < 1, \\ 1 & x \ge 1. \end{cases}$$

(a) Draw a graph of the CDF.

(b) Write $P_X(x)$, the PMF of X. Be sure to write the value of $P_X(x)$ for all x from $-\infty$ to ∞.

2.4.3 The random variable X has CDF

$$F_X(x) = \begin{cases} 0 & x < -3, \\ 0.4 & -3 \le x < 5, \\ 0.8 & 5 \le x < 7, \\ 1 & x \ge 7. \end{cases}$$

(a) Draw a graph of the CDF.

(b) Write $P_X(x)$, the PMF of X.

2.4.4 Following Example 2.24, show that a geometric (p) random variable K has CDF

$$F_K(k) = \begin{cases} 0 & k < 1, \\ 1 - (1-p)^{\lfloor k \rfloor} & k \ge 1. \end{cases}$$

2.4.5 At the One Top Pizza Shop, a pizza sold has mushrooms with probability $p = 2/3$. On a day in which 100 pizzas are sold, let N equal the number of pizzas sold before the first pizza with mushrooms is sold. What is the PMF of N? What is the CDF of N?

2.4.6 In Problem 2.2.8, find and sketch the CDF of B, the number of bases touched safely during an at-bat.

2.4.7 In Problem 2.2.5, find and sketch the CDF of Y, the number of points scored in a 1 and 1 for $p = 1/4$, $p = 1/2$, and $p = 3/4$.

2.4.8 In Problem 2.2.9, find and sketch the CDF of N, the number of attempts made by the cellular phone for $p = 1/2$.

2.5.1 Let X have the uniform PMF

$$P_X(x) = \begin{cases} 0.01 & x = 1, 2, \ldots, 100, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find a mode x_{mod} of X. If the mode is not unique, find the set X_{mod} of all modes of X.

(b) Find a median x_{med} of X. If the median is not unique, find the set X_{med} of all numbers x that are medians of X.

2.5.2 Voice calls cost 20 cents each and data calls cost 30 cents each. C is the cost of one telephone call. The probability that a call is a voice call is $P[V] = 0.6$. The probability of a data call is $P[D] = 0.4$.

(a) Find $P_C(c)$, the PMF of C.

(b) What is $E[C]$, the expected value of C?

2.5.3 Find the expected value of the random variable Y in Problem 2.4.1.

2.5.4 Find the expected value of the random variable X in Problem 2.4.2.

2.5.5 Find the expected value of the random variable X in Problem 2.4.3.

2.5.6 Find the expected value of a binomial ($n = 4$, $p = 1/2$) random variable X.

2.5.7 Find the expected value of a binomial ($n = 5$, $p = 1/2$) random variable X.

2.5.8 Give examples of practical applications of probability theory that can be modeled by the following PMFs. In each case, state an experiment, the sample space, the range of the random variable, the PMF of the random variable, and the expected value:

(a) Bernoulli

(b) Binomial

(c) Pascal

(d) Poisson

Make up your own examples. (Don't copy examples from the text.)

2.5.9 Suppose you go to a casino with exactly \$63. At this casino, the only game is roulette and the only bets allowed are red and green. In addition, the wheel is fair so that $P[\text{red}] = P[\text{green}] = 1/2$. You have the following strategy: First, you bet \$1. If you win the bet, you quit and leave the casino with \$64. If you lose, you then bet \$2. If you win, you quit and go home. If you lose, you bet \$4. In fact, whenever you lose, you double your bet until either you win a bet or you lose all of your money. However, as soon as you win a bet, you quit and go home. Let Y equal the amount of money that you take home. Find $P_Y(y)$ and $E[Y]$. Would you like to play this game every day?

2.5.10 Let binomial random variable X_n denote the number of successes in n Bernoulli trials with success probability p. Prove that $E[X_n] = np$. Hint: Use the fact that $\sum_{x=0}^{n-1} P_{X_{n-1}}(x) = 1$.

2.5.11 Prove that if X is a nonnegative integer-valued random variable, then

$$E[X] = \sum_{k=0}^{\infty} P[X > k].$$

2.6.1 Given the random variable Y in Problem 2.4.1, let $U = g(Y) = Y^2$.

(a) Find $P_U(u)$.

(b) Find $F_U(u)$.

(c) Find $E[U]$.

2.6.2 Given the random variable X in Problem 2.4.2, let $V = g(X) = |X|$.

(a) Find $P_V(v)$.

(b) Find $F_V(v)$.

(c) Find $E[V]$.

2.6.3 Given the random variable X in Problem 2.4.3, let $W = g(X) = -X$.

(a) Find $P_W(w)$.

(b) Find $F_W(w)$.

(c) Find $E[W]$.

2.6.4 At a discount brokerage, a stock purchase or sale worth less than \$10,000 incurs a brokerage fee of 1% of the value of the transaction. A transaction worth more than \$10,000 incurs a fee of \$100 plus 0.5% of the amount exceeding \$10,000. Note that for a fraction of a cent, the brokerage always charges the customer a full penny. You wish to buy 100 shares of a stock whose price D in dollars has PMF

$$P_D(d) = \begin{cases} 1/3 & d = 99.75, 100, 100.25, \\ 0 & \text{otherwise.} \end{cases}$$

What is the PMF of C, the cost of buying the stock (including the brokerage fee).

2.6.5 A source wishes to transmit data packets to a receiver over a radio link. The receiver uses error detection to identify packets that have been corrupted by radio noise. When a packet is received error-free, the receiver sends an acknowledgment (ACK) back to the source. When the receiver gets a packet with errors, a negative acknowledgment (NAK) message is sent back to the source. Each time the source receives a NAK, the packet is retransmitted. We assume that each packet transmission is independently corrupted by errors with probability q.

(a) Find the PMF of X, the number of times that a packet is transmitted by the source.

(b) Suppose each packet takes 1 millisecond to transmit and that the source waits an additional millisecond to receive the acknowledgment message (ACK or NAK) before retransmitting. Let T equal the time required until the packet is successfully received. What is the relationship between T and X? What is the PMF of T?

2.6.6 Suppose that a cellular phone costs \$20 per month with 30 minutes of use included and that each additional minute of use costs \$0.50. If the number of minutes you use in a month is a geometric random variable M with expected value of $E[M] = 1/p = 30$ minutes, what is the PMF of C, the cost of the phone for one month?

2.7.1 For random variable T in Quiz 2.6, first find the expected value $E[T]$ using Theorem 2.10. Next, find $E[T]$ using Definition 2.14.

2.7.2 In a certain lottery game, the chance of getting a winning ticket is exactly one in a thousand. Suppose a person buys one ticket each day (except on the leap year day February 29) over a period of fifty years. What is the expected number $E[T]$ of winning tickets in fifty years? If each winning ticket is worth \$1000, what is the expected amount $E[R]$ collected on these winning tickets? Lastly, if each ticket costs \$2, what is your expected net profit $E[Q]$?

2.7.3 Suppose an NBA basketball player shooting an uncontested 2-point shot will make the basket with probability 0.6. However, if you foul the shooter, the shot will be missed, but two free throws will be awarded. Each free throw is an independent Bernoulli trial with success probability p. Based on the expected number of points the shooter will score, for what values of p may it be desirable to foul the shooter?

2.7.4 It can take up to four days after you call for service to get your computer repaired. The computer company charges for repairs according to how long you have to wait. The number of days D until the service technician arrives and the service charge C, in dollars, are described by

$$P_D(d) = \begin{cases} 0.2 & d = 1, \\ 0.4 & d = 2, \\ 0.3 & d = 3, \\ 0.1 & d = 4, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$C = \begin{cases} 90 & \text{for 1-day service,} \\ 70 & \text{for 2-day service,} \\ 40 & \text{for 3-day service,} \\ 40 & \text{for 4-day service.} \end{cases}$$

(a) What is the expected waiting time $\mu_D = E[D]$?

(b) What is the expected deviation $E[D - \mu_D]$?

(c) Express C as a function of D.

(d) What is the expected value $E[C]$?

2.7.5 For the cellular phone in Problem 2.6.6, express the monthly cost C as a function of M, the number of minutes used. What is the expected monthly cost $E[C]$?

2.7.6 A new cellular phone billing plan costs \$15 per month plus \$1 for each minute of use. If the number of minutes you use the phone in a month is a geometric random variable with mean $1/p$, what is the expected monthly cost $E[C]$ of the phone? For what values of p is this billing plan preferable to the billing plan of Problem 2.6.6 and Problem 2.7.5?

2.7.7 A particular circuit works if all 10 of its component devices work. Each circuit is tested before leaving the factory. Each working circuit can be sold for k dollars, but each nonworking circuit is worthless and must be thrown away. Each circuit can be built with either ordinary devices or ultrareliable devices. An ordinary device has a failure probability of $q = 0.1$ while an ultrareliable device has a failure probability of $q/2$, independent of any other device. However, each ordinary device costs \$1 while an ultrareliable device costs \$3. Should you build your circuit with ordinary devices or ultrareliable devices in order to maximize your expected profit $E[R]$? Keep in mind that your answer will depend on k.

2.7.8 In the New Jersey state lottery, each \$1 ticket has six randomly marked numbers out of $1, \ldots, 46$. A ticket is a winner if the six marked numbers match six numbers drawn at random at the end of a week. For each ticket sold, 50 cents is added to the pot for the winners. If there are k winning tickets, the pot is divided equally among the k winners. Suppose you bought a winning ticket in a week in which $2n$ tickets are sold and the pot is n dollars.

(a) What is the probability q that a random ticket will be a winner?

(b) What is the PMF of K_n, the number of other (besides your own) winning tickets?

(c) What is the expected value of W_n, the prize you collect for your winning ticket?

2.7.9 If there is no winner for the lottery described in Problem 2.7.8, then the pot is carried over to the next week. Suppose that in a given week, an r dollar pot is carried over from the previous week and $2n$ tickets sold. Answer the following questions.

(a) What is the probability q that a random ticket will be a winner?

(b) If you own one of the $2n$ tickets sold, what is the mean of V, the value (i.e., the amount you win) of that ticket? Is it ever possible that $E[V] > 1$?

(c) Suppose that in the instant before the ticket sales are stopped, you are given the opportunity to buy one of each possible ticket. For what values (if any) of n and r should you do it?

2.8.1 In an experiment to monitor two calls, the PMF of N, the number of voice calls, is

$$P_N(n) = \begin{cases} 0.2 & n = 0, \\ 0.7 & n = 1, \\ 0.1 & n = 2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find $E[N]$, the expected number of voice calls.

(b) Find $E[N^2]$, the second moment of N.

(c) Find $\text{Var}[N]$, the variance of N.

(d) Find σ_N, the standard deviation of N.

2.8.2 Find the variance of the random variable Y in Problem 2.4.1.

2.8.3 Find the variance of the random variable X in Problem 2.4.2.

2.8.4 Find the variance of the random variable X in Problem 2.4.3.

2.8.5 Let X have the binomial PMF

$$P_X(x) = \binom{4}{x}(1/2)^4.$$

(a) Find the standard deviation of the random variable X.

(b) What is $P[\mu_X - \sigma_X \le X \le \mu_X + \sigma_X]$, the probability that X is within one standard deviation of the expected value?

2.8.6 The binomial random variable X has PMF

$$P_X(x) = \binom{5}{x}(1/2)^5.$$

(a) Find the standard deviation of X.

(b) Find $P[\mu_X - \sigma_X \le X \le \mu_X + \sigma_X]$, the probability that X is within one standard deviation of the expected value.

2.8.7 Show that the variance of $Y = aX + b$ is $\text{Var}[Y] = a^2\,\text{Var}[X]$.

2.8.8 Given a random variable X with mean μ_X and variance σ_X^2, find the mean and variance of the *standardized random variable*

$$Y = \frac{1}{\sigma_X}(X - \mu_X).$$

2.8.9 In packet data transmission, the time between successfully received packets is called the interarrival time, and randomness in packet interarrival times is called *jitter*. In real-time packet data communications, jitter is undesirable. One measure of jitter is the standard deviation of the packet interarrival time. From Problem 2.6.5, calculate the jitter σ_T. How large must the successful transmission probability q be to ensure that the jitter is less than 2 milliseconds?

2.8.10 Let random variable X have PMF $P_X(x)$. We wish to guess the value of X before performing the actual experiment. If we call our guess $\hat{x}$, the expected square of the error in our guess is

$$e(\hat{x}) = E\left[(X - \hat{x})^2\right]$$

Show that $e(\hat{x})$ is minimized by $\hat{x} = E[X]$.

2.8.11 Random variable K has a Poisson (α) distribution. Derive the properties $E[K] = \text{Var}[K] = \alpha$. Hint: $E[K^2] = E[K(K-1)] + E[K]$.

2.8.12 For the delay D in Problem 2.7.4, what is the standard deviation σ_D of the waiting time?

2.9.1 In Problem 2.4.1, find $P_{Y|B}(y)$, where the condition $B = \{Y < 3\}$. What are $E[Y|B]$ and $\text{Var}[Y|B]$?

2.9.2 In Problem 2.4.2, find $P_{X|B}(x)$, where the condition $B = \{|X| > 0\}$. What are $E[X|B]$ and $\text{Var}[X|B]$?

2.9.3 In Problem 2.4.3, find $P_{X|B}(x)$, where the condition $B = \{X > 0\}$. What are $E[X|B]$ and $\text{Var}[X|B]$?

2.9.4 In Problem 2.8.5, find $P_{X|B}(x)$, where the condition $B = \{X \neq 0\}$. What are $E[X|B]$ and $\text{Var}[X|B]$?

2.9.5 In Problem 2.8.6, find $P_{X|B}(x)$, where the condition $B = \{X \geq \mu_X\}$. What are $E[X|B]$ and $\text{Var}[X|B]$?

2.9.6 Select integrated circuits, test them in sequence until you find the first failure, and then stop. Let N be the number of tests. All tests are independent with probability of failure $p = 0.1$. Consider the condition $B = \{N \geq 20\}$.

(a) Find the PMF $P_N(n)$.

(b) Find $P_{N|B}(n)$, the conditional PMF of N given that there have been 20 consecutive tests without a failure.

(c) What is $E[N|B]$, the expected number of tests given that there have been 20 consecutive tests without a failure?

2.9.7 Every day you consider going jogging. Before each mile, including the first, you will quit with probability q, independent of the number of miles you have already run. However, you are sufficiently decisive that you never run a fraction of a mile. Also, we say you have run a marathon whenever you run at least 26 miles.

(a) Let M equal the number of miles that you run on an arbitrary day. What is $P[M > 0]$? Find the PMF $P_M(m)$.

(b) Let r be the probability that you run a marathon on an arbitrary day. Find r.

(c) Let J be the number of days in one year (not a leap year) in which you run a marathon. Find the PMF $P_J(j)$. This answer may be expressed in terms of r found in part (b).

(d) Define $K = M - 26$. Let A be the event that you have run a marathon. Find $P_{K|A}(k)$.

2.9.8 In the situation described in Example 2.29, the firm sends all faxes with an even number of pages to fax machine A and all faxes with an odd number of pages to fax machine B.

(a) Find the conditional PMF of the length X of a fax, given the fax was sent to A. What are the conditional expected length and standard deviation?

(b) Find the conditional PMF of the length X of a fax, given the fax was sent to B and had no more than six pages. What are the conditional expected length and standard deviation?

2.10.1 Let X be a binomial (n, p) random variable with $n = 100$ and $p = 0.5$. Let E_2 denote the event that X is a perfect square. Calculate $P[E_2]$.

2.10.2 Write a MATLAB function `x=faxlength8(m)` that produces m random sample values of the fax length X with PMF given in Example 2.29.

2.10.3 For $m = 10$, $m = 100$, and $m = 1000$, use MATLAB to find the average cost of sending m faxes using the model of Example 2.29. Your program input should have the number of trials m as the input. The output should be $\overline{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$ where

Y_i is the cost of the ith fax. As m becomes large, what do you observe?

2.10.4 The Zipf $(n, \alpha = 1)$ random variable X introduced in Problem 2.3.9 is often used to model the "popularity" of a collection of n objects. For example, a Web server can deliver one of n Web pages. The pages are numbered such that the page 1 is the most requested page, page 2 is the second most requested page, and so on. If page k is requested, then $X = k$.

To reduce external network traffic, an ISP gateway caches copies of the k most popular pages. Calculate, as a function of n for $1 \leq n \leq 1{,}000$, how large k must be to ensure that the cache can deliver a page with probability 0.75.

2.10.5 Generate n independent samples of a Poisson ($\alpha = 5$) random variable Y. For each $y \in S_Y$, let $n(y)$ denote the number of times that y was observed. Thus $\sum_{y \in S_Y} n(y) = n$ and the relative frequency of y is $R(y) = n(y)/n$. Compare the relative frequency of y against $P_Y(y)$ by plotting $R(y)$ and $P_Y(y)$ on the same graph as functions of y for $n = 100, n = 1000$ and $n = 10{,}000$. How large should n be to have reasonable agreement?

2.10.6 Test the convergence of Theorem 2.8. For $\alpha = 10$, plot the PMF of K_n for $(n, p) = (10, 1)$, $(n, p) = (100, 0.1)$, and $(n, p) = (1000, 0.01)$ and compare against the Poisson (α) PMF.

2.10.7 Use the result of Problem 2.4.4 and the Random Sample Algorithm on Page 89 to write a MATLAB function `k=geometricrv(p,m)` that generates m samples of a geometric (p) random variable.

2.10.8 Find n^*, the smallest value of n for which the function `poissonpmf(n,n)` shown in Example 2.44 reports an error. What is the source of the error? Write a MATLAB function `bigpoissonpmf(alpha,n)` that calculates `poissonpmf(n,n)` for values of n much larger than n^*. Hint: For a Poisson (α) random variable K,

$$P_K(k) = \exp\left(-\alpha + k\ln(\alpha) - \sum_{j=1}^{k} \ln(j)\right).$$