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 = aaraaa. 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

Definition
Definition

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 $P_X(\cdot)$.

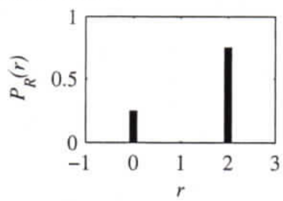
We graph a PMF by marking on the horizontal axis each value with nonzero probability and drawing a vertical bar with length proportional to the probability.

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?

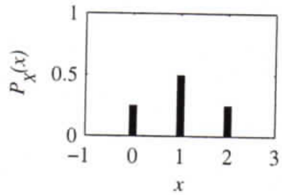
There are four outcomes of this experiment: gg, gb, bg, and bb. A simple tree diagram indicates that each outcome has probability $1/4$. The random variable X has three possible values corresponding to three events:

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

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.

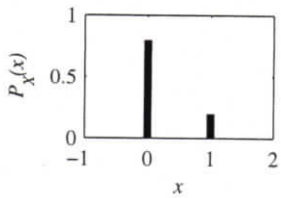
Example 2.9 Suppose you test one circuit. With probability p, the circuit is rejected. Let X be the number of rejected circuits in one test. What is $P_X(x)$?

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.

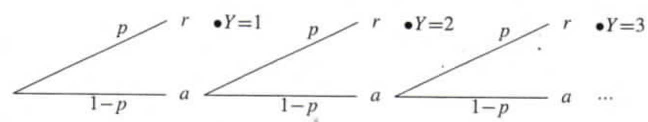
Example 2.10 If there is a 0.2 probability of a reject,



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

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.

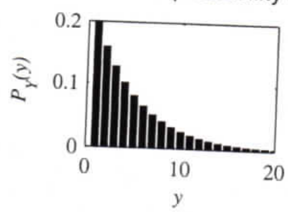
Definition 2.6 **Geometric (p) Random Variable**

*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*

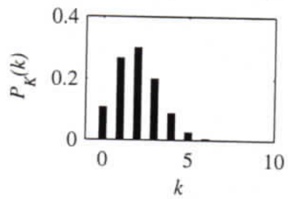
$$P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

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***

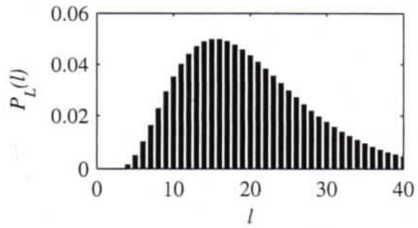
*X is a **Pascal** (k, p) random variable if the PMF of X has the form*

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

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



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

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

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

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

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

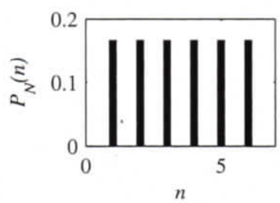
X is a ***discrete uniform*** *(k, l) random variable if the PMF of X has the form*

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

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



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

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 $\alpha = \lambda T = (2 \text{ hits/s}) \times (0.25 \text{ s}) = 0.5$ 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, $\alpha = \lambda T = (2 \text{ hits/s}) \times (1\text{s}) = 2$ 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?

Exa

Exa

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

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

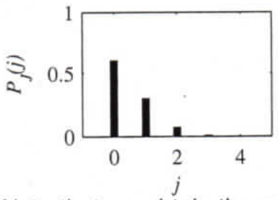
Therefore,

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

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.

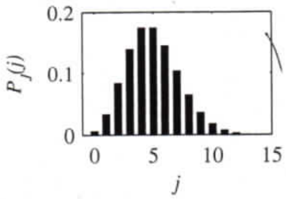
Example 2.21 Calls arrive at random times at a telephone switching office with an average of $\lambda = 0.25$ calls/second. The PMF of the number of calls that arrive in a $T = 2$-second interval is the Poisson (0.5) random variable with PMF



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

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



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

Quiz 2.3

Each time a modem transmits one bit, the receiving modem analyzes the signal that arrives and decides whether the transmitted bit is 0 or 1. It makes an error with probability p, independent of whether any other bit is received correctly.

(1) *If the transmission continues until the receiving modem makes its first error, what is the PMF of X, the number of bits transmitted?*

(2) *If $p = 0.1$, what is the probability that $X = 10$? What is the probability that $X \geq 10$?*

(3) *If the modem transmits 100 bits, what is the PMF of Y, the number of errors?*

(4) *If $p = 0.01$ and the modem transmits 100 bits, what is the probability of $Y = 2$ errors at the receiver? What is the probability that $Y \leq 2$?*

(5) *If the transmission continues until the receiving modem makes three errors, what is the PMF of Z, the number of bits transmitted?*

(6) *If $p = 0.25$, what is the probability of $Z = 12$ bits transmitted?*

2.4 Cumulative Distribution Function (CDF)

Like the PMF, the CDF of a discrete random variable contains complete information about the probability model of the random variable. The two functions are closely related. Each can be obtained easily from the other.

Definition 2.11 ***Cumulative Distribution Function (CDF)***

*The **cumulative distribution function** (CDF) of random variable X is*

$$F_X(x) = P[X \leq x].$$

For any real number x, the CDF is the probability that the random variable X is no larger than x. All random variables have cumulative distribution functions but only discrete random variables have probability mass functions. The notation convention for the CDF follows that of the PMF, except that we use the letter F with a subscript corresponding to the name of the random variable. Because $F_X(x)$ describes the probability of an event, the CDF has a number of properties.

Theorem 2.2 *For any discrete random variable X with range $S_X = \{x_1, x_2, \ldots\}$ satisfying $x_1 \leq x_2 \leq \ldots$,*

(a) $F_X(-\infty) = 0$ *and* $F_X(\infty) = 1$.

(b) *For all* $x' \geq x$, $F_X(x') \geq F_X(x)$.

(c) *For* $x_i \in S_X$ *and* ϵ, *an arbitrarily small positive number,*

$$F_X(x_i) - F_X(x_i - \epsilon) = P_X(x_i).$$

(d) $F_X(x) = F_X(x_i)$ *for all x such that* $x_i \leq x < x_{i+1}$.

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 $b \geq a$,*

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

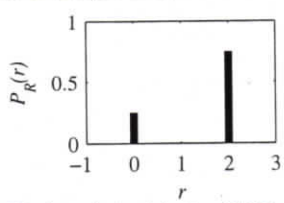
Proof To prove this theorem, express the event $E_{ab} = \{a < X \leq b\}$ as a part of a union of disjoint events. Start with the event $E_b = \{X \leq b\}$. Note that E_b can be written as the union

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

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.

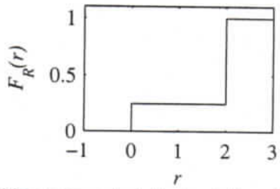
Example 2.23 In Example 2.6, we found that random variable R has PMF



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

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

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

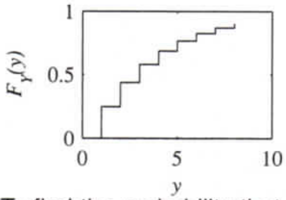
Equation (2.41) is a geometric series. Familiarity with the geometric series is essential for calculating probabilities involving geometric random variables. Appendix B summarizes the most important facts. In particular, Math Fact B.4 implies $(1-x)\sum_{j=1}^{n} x^{j-1} = 1 - x^n$. Substituting $x =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

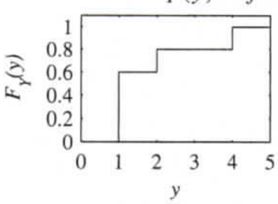


$$F_Y(y) = \begin{cases} 0 & y < 1, \\ 1 - (3/4)^{\lfloor y \rfloor} & y \ge 1. \end{cases} \tag{2.44}$$

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 ($g(x_1) \ne g(x_2)$ if $x_1 \ne x_2$) 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 $P[Y = y]$

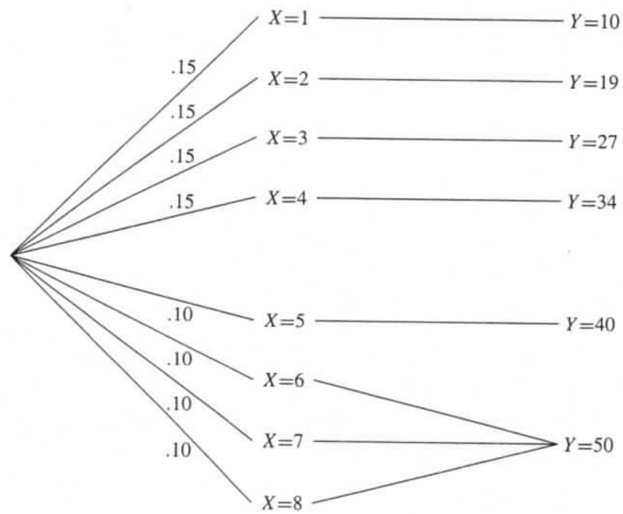


Figure 2.1 The derived random variable $Y = g(X)$ for Example 2.29.

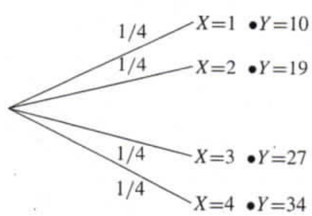
equals the sum of the probabilities of all the outcomes $X = x$ for which $Y = y$.

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

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

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 $P_Y(y)$.



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

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

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*

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

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

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

What is $E[Y]$?

Theorem 2.11

Theorem 2.1

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*

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

(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 $E[M] = g(E[A])$?