Chapter 2

Discrete Random Variables

2.1 INTRODUCTION

Thus far we have treated the sample space as the set of all possible outcomes of a random experiment. Some examples of sample spaces we have considered are

$$S_1 = \{0, 1\},$$
$$S_2 = \{(0,0), (0,1), (1,0), (1,1)\},$$
$$S_3 = \{\text{success, failure}\}.$$

Some experiments yield sample spaces whose elements are numbers, but some other experiments do not yield numerically valued elements. For mathematical convenience, it is often desirable to associate one or more numbers (in addition to probabilities) with each possible outcome of an experiment. Such numbers might naturally correspond, for instance, to the cost of each experimental outcome, the total number of defective items in a batch, or the time to failure of a component.

Through the notion of *random variables*, this and the following chapter extend our earlier work to develop methods for the study of experiments whose outcomes may be described numerically. Besides this convenience, random variables also provide a more compact description of an experiment than the finest grain description of the sample space. For example, in the inspection of manufactured products, we may be interested only in the total number of defective items and not in the nature of the defects; in a sequence of Bernoulli trials, we may be interested only in the number of successes and not in the

Probability and Statistics with Reliability, Queuing, and Computer Science Applications, Second Edition. Kishor S. Trivedi.

Companion Website: www.wiley.com/go/trivedi/probabilityandstatistics2e

actual sequence of successes and failures. The notion of random variables provides us the power of abstraction and thus allows us to discard unimportant details in the outcome of an experiment. Virtually all serious probabilistic computations are performed in terms of random variables.

2.2 RANDOM VARIABLES AND THEIR EVENT SPACES

A **random variable** is a rule that assigns a numerical value to each possible outcome of an experiment. The term "random variable" is actually a misnomer, since a random variable X is really a function whose domain is the sample space S, and whose range is the set of all real numbers, $\Re$. The set of all values taken by X, called the *image of* X, will then be a subset of the set of all real numbers.

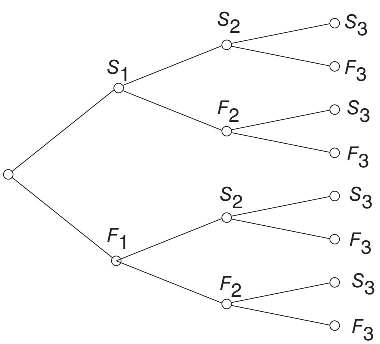
Definition (Random Variable). A random variable X on a sample space S is a function $X : S \rightarrow \Re$ that assigns a real number $X(s)$ to each sample point $s \in S$.

Example 2.1

As an example, consider a random experiment defined by a sequence of three Bernoulli trials. The sample space S consists of eight triples of 0s and 1s. We may define any number of random variables on this sample space. For our example, define a random variable X to be the total number of successes from the three trials.

The tree diagram of this sequential sample space is shown in Figure 2.1, where S_n and F_n respectively denote a success and a failure on the nth trial, and the probability of success, p, is equal to 0.5. The value of random variable X assigned to each sample point is also included.

If the outcome of one performance of the experiment were $s = (0, 1, 0)$, then the resulting experimental value of the random variable X is 1, that is $X(0, 1, 0) = 1$. Note that two or more sample points might give the same value for X (i.e., X may



Sample points	$P(s)$	$X(s)$
111	0.125	3
110	0.125	2
101	0.125	2
100	0.125	1
011	0.125	2
010	0.125	1
001	0.125	1
000	0.125	0

Figure 2.1. Tree diagram of a sequential sample space

not be a one-to-one function), but that two different numbers in the range cannot be assigned to the same sample point (i.e., X is a well-defined function). For example

$$X(1,0,0) = X(0,1,0) = X(0,0,1) = 1.$$

$\sharp$

A random variable partitions its sample space into a mutually exclusive and collectively exhaustive set of events. Thus for a random variable X and a real number x, we define the event A_x [commonly called the **inverse image** of the set $\{x\}$] to be the subset of S consisting of all sample points s to which the random variable X assigns the value x:

$$A_x = \{s \in S | X(s) = x\}.$$

It is clear that $A_x \cap A_y = \emptyset$ if $x \neq y$, and that

$$\bigcup_{x \in \Re} A_x = S$$

(see problem 1 at the end of this section). Thus the collection of events A_x for all x defines an **event space**. We may find it more convenient to work in this event space (rather than the original sample space), provided our only interest in performing the experiment has to do with the resulting experimental value of random variable X. The notation $[X = x]$ will be used as an abbreviation for the event A_x. Thus

$$[X = x] = \{s \in S | X(s) = x\}.$$

In Example 2.1 the random variable X defines four events:

$$\begin{aligned} A_0 &= \{s \in S | X(s) = 0\} = \{(0,0,0)\}, \\ A_1 &= \{(0,0,1), (0,1,0), (1,0,0)\}, \\ A_2 &= \{(0,1,1), (1,0,1), (1,1,0)\}, \\ A_3 &= \{(1,1,1)\}. \end{aligned}$$

For all values of x outside the image of X (i.e., values of x other than 0,1,2,3), A_x is the null set. The resulting event space contains four event points (see Figure 2.2). For a sequence of n Bernoulli trials, the event space defined by X will have $(n+1)$ points, compared with 2^n sample points in the original sample space!

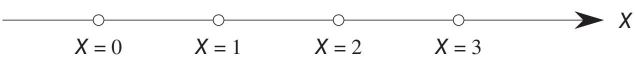


Figure 2.2. Event space for three Bernoulli trials

The random variable discussed in our example could take on values from a set of discrete numbers and hence, the image of the random variable is either finite or countable. Such random variables, known as **discrete random variables**, are the subject of this chapter, while continuous random variables are discussed in the next chapter. A random variable defined on a discrete sample space will be discrete, while it is possible to define a discrete random variable on a continuous sample space. For instance, for a continuous sample space S, the random variable defined by, say, $X(s) = 4$ for all $s \in S$ is discrete.

Problems

1. Given a discrete random variable X, define the event A_x by

$$A_x = \{s \in S \mid X(s) = x\}.$$

Show that the family of events $\{A_x\}$ defines an event space.

2.3 THE PROBABILITY MASS FUNCTION

We have defined the event A_x as the set of all sample points $\{s \mid X(s) = x\}$. Consequently

$$\begin{aligned} P(A_x) &= P([X = x]) \\ &= P(\{s \mid X(s) = x\}) \\ &= \sum_{X(s)=x} P(s). \end{aligned}$$

This formula provides us with a method of computing $P(X = x)$ for all $x \in \Re$. Thus we have defined a function with its domain consisting of the event space of the random variable X, and with its range in the closed interval [0,1]. This function is known as the **probability mass function** (pmf) or the **discrete density function** of the random variable X, and will be denoted by $p_X(x)$. Thus

$$\begin{aligned} p_X(x) &= P(X = x) \\ &= \sum_{X(s)=x} P(s) \\ &= \text{probability that the value of the random variable} \\ &\quad\ X \text{ obtained on a performance of the experiment is} \\ &\quad\ \text{equal to } x. \end{aligned}$$

It should be noted that the argument x of the pmf $p_X(x)$ is a dummy variable, hence it can be changed to any other dummy variable y with no effect on the definition.

The following properties hold for the pmf:

(p1) $0 \le p_X(x) \le 1$ for all $x \in \Re$. This must be true, since $p_X(x)$ is a probability.

(p2) Since the random variable assigns some value $x \in \Re$ to each sample point $s \in S$, we must have

$$\sum_{x \in \Re} p_X(x) = 1.$$

(p3) For a discrete random variable X, the set $\{x \mid p_X(x) \neq 0\}$ is a finite or countably infinite subset of real numbers (this set is defined to be the image of X). Let this set be denoted by $\{x_1, x_2, \ldots\}$. Then property (p2) may be restated as

$$\sum_{i} p_X(x_i) = 1.$$

A real-valued function $p_X(x)$ defined on $\Re$ is the pmf of some random variable X provided that it satisfies properties (p1) to (p3). Continuing with Example 2.1, we can easily obtain $p_X(x)$ for $x = 0, 1, 2, 3$ from the preceding definitions:

$$p_X(0) = \tfrac{1}{8},$$
$$p_X(1) = \tfrac{3}{8},$$
$$p_X(2) = \tfrac{3}{8},$$
$$p_X(3) = \tfrac{1}{8}.$$

Check that all the properties listed above hold. This pmf may be visualized as a bar histogram drawn over the event space for the random variable (see Figure 2.3).

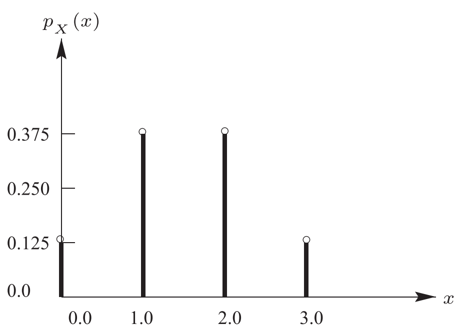


Figure 2.3. Histogram of pmf for Example 2.1

Example 2.2

Returning to the example of a wireless cell with five channels from Chapter 1, and defining the random variable $X =$ the number of available channels, we have

$$p_X(0) = \tfrac{1}{32},\ p_X(1) = \tfrac{5}{32},\ p_X(2) = \tfrac{10}{32},$$
$$p_X(3) = \tfrac{10}{32},\ p_X(4) = \tfrac{5}{32},\ p_X(5) = \tfrac{1}{32}.$$

♯

2.4 DISTRIBUTION FUNCTIONS

So far we have restricted our attention to computing $P(X = x)$, but often we may be interested in computing the probability of the set $\{s \mid X(s) \in A\}$ for some subset A of $\Re$ other than a one-point set. It is clear that

$$\{s \mid X(s) \in A\} = \bigcup_{x_i \in A} \{s \mid X(s) = x_i\}. \tag{2.1}$$

Usually this event is denoted as $[X \in A]$ and its probability by $P(X \in A)$. If $-\infty < a < b < \infty$ and A is an interval with endpoints a and b, say, $A = (a, b)$, then we usually write $P(a < X < b)$ instead of $P[X \in (a, b)]$. Similarly, if $A = (a, b]$, then $P(X \in A)$ will be written as $P(a < X \le b)$. The semiinfinite interval $A = (-\infty, x]$ will be of special interest and in this case we denote the event $[X \in A]$ by $[X \le x]$.

If $p_X(x)$ denotes the pmf of random variable X, then, from equation (2.1), we have

$$P(X \in A) = \sum_{x_i \in A} p_X(x_i).$$

Thus in Example 2.2, the probability that two or fewer channels will be available may now be evaluated quite simply as

$$\begin{aligned} P(X \le 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= p_X(0) + p_X(1) + p_X(2) \\ &= \tfrac{1}{32} + \tfrac{5}{32} + \tfrac{10}{32} \\ &= \tfrac{16}{32} \\ &= \tfrac{1}{2}. \end{aligned}$$

The function $F_X(t), -\infty < t < \infty$, defined by

$$\begin{aligned} F_X(t) &= P(-\infty < X \le t) \\ &= P(X \le t) \\ &= \sum_{x \le t} p_X(x) \end{aligned} \tag{2.2}$$

is called the **cumulative distribution function** (CDF) or the **probability distribution function** or simply the **distribution function** of the random variable X. We will omit the subscript X whenever no confusion arises. It follows from this definition that

$$\begin{aligned} P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a). \end{aligned}$$

If X is an integer-valued random variable, then

$$F(t) = \sum_{-\infty < x \leq \lfloor t \rfloor} p_X(x)$$

where $\lfloor t \rfloor$ denotes the greatest integer less than or equal to t (also known as the **floor** of t).

Several properties of $F_X(x)$ follow directly from its definition.

(F1) $0 \leq F(x) \leq 1$ for $-\infty < x < \infty$. This follows because $F(x)$ is a probability.

(F2) $F(x)$ is a monotone increasing function of x; that is, if $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$. This follows by first observing that the interval $(-\infty, x_1]$ is contained in the interval $(-\infty, x_2]$ whenever $x_1 \leq x_2$ and hence

$$P(-\infty < X \leq x_1) \leq P(-\infty < X \leq x_2).$$

That is, $F(x_1) \leq F(x_2)$.

(F3) $\lim_{x \to -\infty} F(x) = 0$, and $\lim_{x \to \infty} F(x) = 1$. If the random variable X has a finite image, then $F(x) = 0$ for all x sufficiently small and $F(x) = 1$ for all x sufficiently large.

(F4) $F(x)$ has a positive jump equal to $p_X(x_i)$ at $i = 1, 2, \ldots$, and in the interval $[x_i, x_{i+1})$ $F(x)$ has a constant value. Thus

$$F(x) = F(x_i) \quad \text{for } x_i \leq x < x_{i+1}$$

and

$$F(x_{i+1}) = F(x_i) + p_X(x_{i+1}).$$

It can be shown that any function $F(x)$ satisfying properties (F1)–(F4) is the distribution function of some discrete random variable.

We note that distribution functions of discrete random variables grow only by jumps, whereas the distribution functions of continuous random variables are continuous functions and hence have no jumps. A random variable X is said to be of **mixed type** if its distribution function has both jumps as well as continuous growth. In most practical situations, the random variable is

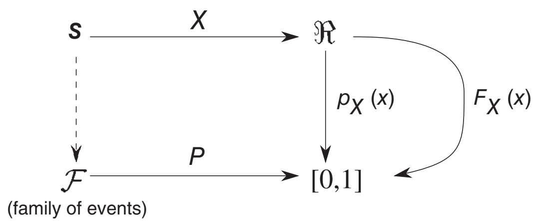


Figure 2.4. Domain and range of P, X, pmf, and CDF

either discrete or continuous. Therefore, we will study only these two cases in detail. The domains and ranges of the four functions (the probability measure, the random variable X, the pmf, and the CDF) we have studied so far are summarized in Figure 2.4.

The cumulative distribution function contains most of the interesting information about the underlying probability system and will be used extensively. Often the concepts of sample space, event space, and probability measure, which are fundamental in building the theory of probability, will fade into the background, and functions such as the distribution function or the probability mass function become the most important entities. It is important, nevertheless, to keep this background in mind. You will often see the statement "Let X be a discrete random variable with pmf p_X," with no reference made to the underlying probability space. We can always construct an appropriate space, as follows. Take $S = \Re$; $X(s) = s$, for $s \in S$; $\mathcal{F}$ = union of the inverse images of A_x of all the subsets x pertaining to the set of real numbers $\Re$ and

$$P(A) = \sum_{x \in A} p_X(x)$$

for a subset, A, of $\Re$. In this case, the event space of the random variable X is identical to the sample space defined above. Similarly, the statement, "Let X be a discrete random variable with the CDF F," always makes sense.

Example 2.3

The CDF of the running example of the sequence of three Bernoulli trials is shown in Figure 2.5. The properties (F1)—(F4) above are easily seen to hold.

♯

2.5 SPECIAL DISCRETE DISTRIBUTIONS

In many theoretical and practical problems, several probability mass functions appear frequently enough that they are worth exploring here.

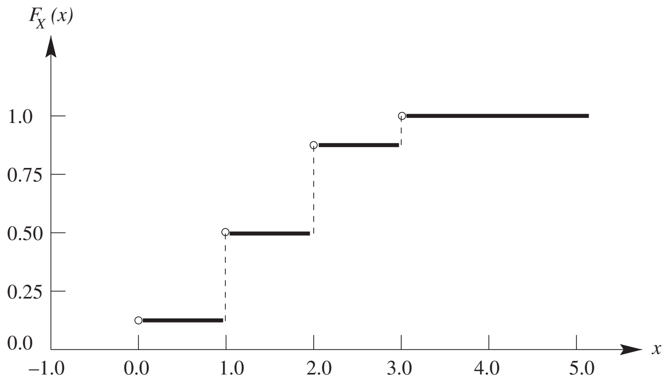


Figure 2.5. CDF of three Bernoulli trials

2.5.1 The Bernoulli pmf

The Bernoulli pmf is the density function of a discrete random variable X having 0 and 1 as its only possible values; it originates from the experiment consisting of a single Bernoulli trial. It is given by

$$\begin{aligned} p_X(0) &= p_0 = P(X=0) = q, \\ p_X(1) &= p_1 = P(X=1) = p, \end{aligned}$$

where $p+q=1$. The corresponding CDF is given by (see Figure 2.6)

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ q & \text{for } 0 \leq x < 1, \\ 1 & \text{for } x \geq 1. \end{cases}$$

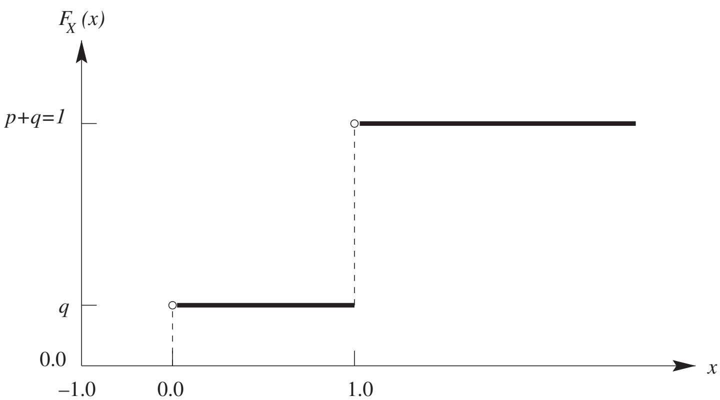


Figure 2.6. CDF of Bernoulli random variable

2.5.2 The Binomial pmf

To generate the Bernoulli pmf, we considered a single Bernoulli trial. Now we consider a sequence of n independent Bernoulli trials with the probability of success equal to p on each trial. Let Y_n denote the number of successes in n trials. The domain of the random variable Y_n is all the n-tuples of 0s and 1s, and the image is $\{0, 1, \ldots, n\}$. The value assigned to a sample point (an n-tuple) by Y_n simply corresponds to the number of 1s in the n-tuple. As was shown in Section 1.12, the pmf of Y_n is

$$\begin{aligned} p_k &= P(Y_n = k) \\ &= p_{Y_n}(k) \\ &= \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{for } 0 \le k \le n, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{2.3}$$

This equation gives the probability of k successes in $n$independent trials of an experiment that has probability p of success on each trial. One of the more important discrete densities in probability theory, this is called the **binomial density** with parameters n and p, often denoted by $b(k; n, p)$. An example of $b(k; 3, 0.5)$ was presented earlier in this chapter (see Figure 2.3).

It is easily verified using the binomial theorem that

$$\begin{aligned} \sum_{i=0}^{n} p_i &= \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \\ &= [p + (1-p)]^n \\ &= 1. \end{aligned}$$

This is the reason for the term *binomial* pmf. We often refer to a random variable Y_n having a binomial pmf by saying that Y_n has a **binomial distribution** (with parameters n and p if we want to be more precise). Similar phraseology will be used for other random variables having a named density. The distribution function of a binomial random variable will be denoted by $B(t; n, p)$ and is given by

$$\begin{aligned} B(t; n, p) &= F_{Y_n}(t) \\ &= \sum_{i=0}^{\lfloor t \rfloor} \binom{n}{i} p^i (1-p)^{n-i}. \end{aligned} \tag{2.4}$$

The binomial distribution is applicable whenever a series of trials is made satisfying the following conditions:

1. Each trial has exactly two mutually exclusive outcomes, usually labeled "success" and "failure."
2. The probability of "success" on each trial is a constant, denoted by p. The probability of "failure" is $q = 1 - p$.
3. The outcomes of successive trials are mutually independent.

A typical situation in which these conditions will apply (at least approximately) occurs when several components are selected at random (with replacement) from a large batch of components and examined to see if there are any defective components (i.e., failures). The number of defectives in a sample of size n is a random variable, denoted by Y_n, which is binomially distributed.

These assumptions constitute what is called a *binomial model*, which is a typical example of a mathematical model in that it attempts to describe a physical situation in mathematical terms. Models such as these depend on one or more **parameters** that govern their behavior. The binomial model has two parameters, n and p. If the values of model parameters are known, then it is relatively easy to evaluate the probabilities of the events of interest.

We emphasize that the three properties listed above are **assumptions** and need not always hold. We may wish to analyze empirically observed data, and may hypothesize that the assumptions of the binomial model (or any other such model) hold. This hypothesis needs to be tested and can be either rejected or accepted on the basis of the test. Hypothesis testing is discussed in Chapter 10.

Example 2.4

As an example of binomial distribution, consider a plant manufacturing VLSI (very large-scale integrated circuit) chips, 10% of which are expected to be defective. The quality control procedure consists of counting the number of defective chips in a sample of size 35. Suppose after 800 applications of this procedure we find that our experience is reflected in the following table. Although we do not expect exactly 10% defectives every time, are the observations consistent with our hypothesis that 10% are defective?

Number of defects	*Number of samples showing this number of defects*	*Fraction (of 800 samples) showing this number of defects*
0	11	0.01375
1	95	0.11875
2	139	0.17375
3	213	0.26625
4	143	0.17875

(*continued overleaf*)

Number of defects	*Number of samples showing this number of defects*	*Fraction (of 800 samples) showing this number of defects*
5	113	0.14125
6	49	0.06125
7	27	0.03375
8	6	0.00750
9	4	0.00500
10	0	0.00000
	800	1.00000

This situation is typical of those fitting a binomial model. "Success" is finding a defective chip, and we are counting the number of successes in 35 trials. Since the probability of success is $p = 0.1$, the observed fraction defective should be close to the binomial pmf:

$$b(k; 35, 0.1) = \binom{35}{k} \cdot 0.1^k \cdot 0.9^{(35-k)}.$$

The observed data and the binomial pmf are compared in the following table as well as in Figure 2.7. In Chapter 10 we will study statistical tests that will allow us to quantify the goodness of fit of the data presented above to the binomial model.

k = *defects/sample*	*Data*	$b(k; 35, 0.1)$
0	0.01375	0.0250
1	0.11875	0.0974
2	0.17375	0.1839
3	0.26625	0.2248
4	0.17875	0.1998
5	0.14125	0.1376
6	0.06125	0.0765
7	0.03375	0.0352
8	0.00750	0.0137
9	0.00500	0.0046
10	0.00000	0.0013
11	0.00000	0.0003
12	0.00000	0.0000

♯

Example 2.5

The number of surviving components, Y_n, out of a given number of n identical and independent components has a binomial distribution $B(k; n, R)$, where R is

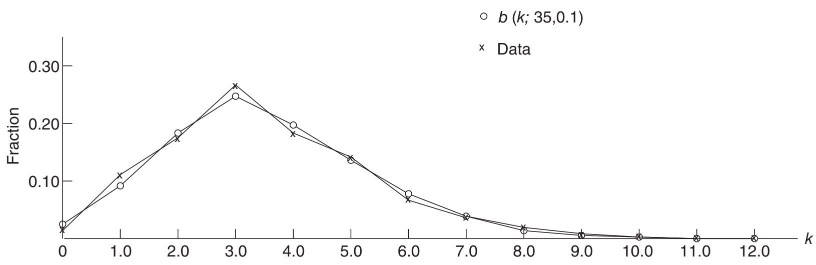


Figure 2.7. Comparing the model pmf with data of Example 2.4

the reliability of a single component. Thus the reliability of an k-out-of-n system is given by

$$
\begin{aligned}
R_{k|n} &= P(\text{“}k \text{ or more components have not failed”}) \\
&= 1 - \sum_{i=0}^{k-1} p_{Y_n}(i) \\
&= 1 - F_{Y_n}(k-1) \\
&= \sum_{i=k}^{n} \binom{n}{i} R^i (1-R)^{(n-i)}. \qquad (2.5)
\end{aligned}
$$

♯

Example 2.6

While transmitting binary digits through a communication channel, the number of digits received correctly, C_n, out of n transmitted digits has a binomial distribution $B(k; n, p)$, where p is the probability of successfully transmitting one digit. The probability of exactly i errors is given by

$$P_e(i) = p_{C_n}(n-i) = \binom{n}{i} p^{(n-i)} (1-p)^i,$$

and thus the probability of an error-free transmission is given by:

$$P_e(0) = p^n.$$

♯

Example 2.7

Now consider the logical link control (LLC) and medium access control (MAC) protocol of a wireless communication system [DEME 1999]. When an LLC frame is passed to the MAC layer, it is segmented into n MAC blocks of fixed size and these n blocks are transmitted through the radio channel separately. Assume that the automatic repeat request (ARQ) scheme is applied in case an error occurred during the transmission. Let $P_c(k)$ denote the probability that after the $(k-1)$st

MAC retransmission there are MAC blocks in error that are corrected by the kth MAC retransmission. We are interested in the pmf of K, which is the number of LLC transmissions required for the error free transmission of n MAC blocks.

Assume that the probability of successful transmission of a single block is $p(>0)$. Then the probability $P_c(1)$ that all n MAC blocks of an LLC frame are received error-free at the first transmission is equal to

$$P_c(1) = p^n,$$

where we assume that the transmission of MAC blocks are statistically independent events. To calculate $P_c(2)$, we note that $1-(1-p)^2$ is the probability that a given MAC block is successfully received after two MAC retransmissions. Then, $(1-(1-p)^2)^n$ is the probability that the LLC frame is correctly received within one or two transmissions. This yields

$$P_c(2) = [1-(1-p)^2]^n - p^n.$$

Following the above approach, the general form of $P_c(k)$ is

$$P_c(k) = \left[1-(1-p)^k\right]^n - \left[1-(1-p)^{k-1}\right]^n.$$

From above equation we have

$$\lim_{k\to\infty} P_c(k) = \lim_{k\to\infty} \left\{\left[1-(1-p)^k\right]^n - \left[1-(1-p)^{k-1}\right]^n\right\} = 0 \tag{2.6}$$

and

$$\sum_{k=1}^{\infty} \left\{\left[1-(1-p)^k\right]^n - \left[1-(1-p)^{k-1}\right]^n\right\} = 1.$$

♯

Example 2.8

Consider taking a random sample of 10 VLSI chips from a very large batch. If no chips in the sample are found to be defective, then we accept the entire batch; otherwise we reject the batch. The number of defective chips in a sample has the pmf $b(k; 10, p)$, where p denotes the probability that a randomly chosen chip is defective. Thus

$$\begin{aligned} P(\text{“No defectives”}) &= (1-p)^{10} \\ &= \text{probability that a batch is accepted.} \end{aligned}$$

If $p = 0$, the batch is certain to be accepted and if $p = 1$, the batch will certainly be rejected. The expression for the probability of acceptance is plotted in Figure 2.8.

♯

The student should not be misled by these examples into thinking that quality control problems can always be solved by simply plugging numbers into the binomial pmf.

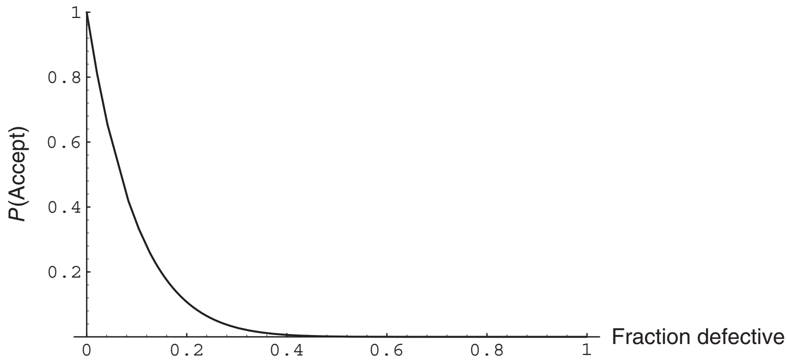


Figure 2.8. Probability of acceptance versus fraction defective

Example 2.9 (Simpson's Reversal Paradox)

Consider two shipments (labeled I and II) of VLSI chips from each of the two manufacturers A and B. Suppose that the proportion of defectives among the four shipments are as follows:

		Manufacturer A	*Manufacturer* B
Shipment	*I*	600 good 500 defective	400 good 300 defective
Shipment	*II*	300 good 600 defective	500 good 900 defective

On inspecting shipment I, the quality control engineer will find

$$\begin{aligned} & P(\text{selecting a defective chip from A}) = \tfrac{5}{11} \\ > \ & P(\text{selecting a defective chip from B}) = \tfrac{3}{7}. \end{aligned}$$

Inspection of shipment II yields

$$\begin{aligned} & P(\text{selecting a defective chip from A}) = \tfrac{6}{9} \\ > \ & P(\text{selecting a defective chip from B}) = \tfrac{9}{14}. \end{aligned}$$

The engineer will presumably conclude that the manufacturer B is sending better chips than the manufacturer A. Suppose, however, that the engineer mixes the two shipments from A together and similarly for B. A subsequent test leads him to a reverse conclusion since

$$\begin{aligned} & P(\text{selecting a defective chip from A}) = \tfrac{11}{20} \\ < \ & P(\text{selecting a defective chip from B}) = \tfrac{12}{21}. \end{aligned}$$

The problem here is that we are tempted to add the fractions $\frac{5}{11} + \frac{6}{9}$ and compare the sum with $\frac{3}{7} + \frac{9}{14}$; unfortunately, what is called for is adding numerators and adding denominators, which is *not* the way we add fractions.

$\sharp$

When n becomes very large, computation using the binomial formula becomes unmanageable. In the limit as n approaches infinity, it can be shown that

$$b(k; n, p) \simeq \frac{1}{\sqrt{2\pi \; npq}} \cdot e^{-(k-np)^2/(2npq)}. \tag{2.7}$$

This is known as the *Laplace* (or *normal*) *approximation* to the binomial pmf and the agreement between the two formulas depends on the values of n and p. Take $n = 5$ and $p = 0.5$, then

k	$b(k; 5, 0.5)$	*Laplace approximation to* $b(k; 5, 0.5)$
0	0.03125	0.02929
1	0.15625	0.14507
2	0.31250	0.32287
3	0.31250	0.32287
4	0.15625	0.14507
5	0.03125	0.02929

As p moves away from 0.5, larger values of n are needed. Larson [LARS 1974] suggests that for $n \geq 10$, if

$$\frac{9}{n+9} \leq p \leq \frac{n}{n+9},$$

then the Laplace formula provides a good approximation to the binomial pmf.

Other authors give different advice concerning when to use the normal approximation. For example, Schader and Schmid [SCHA 1989] compared the maximum absolute error in computing the cumulative binomial distribution function using the normal approximation with a continuity correction (see Example 3.7). They consider the two rules for determining whether this approximation should be used: np and $n(1-p)$ are both greater than 5, and $np(1-p) > 9$. Their conclusion is that the relationship between the maximum absolute error and p is approximately linear when considering the smallest possible sample sizes to satisfy the rules. For more information, refer to Leemis and Trivedi [LEEM 1996], Chapter 10 of this book and the Poisson pmf section in this chapter. Yet another approximation to the binomial pmf is the Poisson pmf, which we will study later.

Owing to the importance of the binomial distribution, the binomial CDF

$$B(k;n,p) = \sum_{i=0}^{k} b(i;n,p)$$

has been tabulated for $n = 2$ to $n = 49$ by the National Bureau of Standards [NBS 1950] and for $n = 50$ to $n = 100$ by Romig [ROMI 1953]. [In Appendix C, we have tabulated $B(k;n,p)$ for $n = 2$ to 20]. For larger values of n, we recommend the use of the Mathematica [WOLF 1999] built-in function **Binomial[n,k]**. Three different possible shapes of binomial pmf's are illustrated in Figures 2.9–2.11. If $p = 0.5$, the bar chart of the binomial pmf is **symmetric**, as in Figure 2.9. If $p < 0.5$, then a **positively skewed** binomial pmf is obtained (see Figure 2.10), and if $p > 0.5$ a **negatively skewed** binomial pmf is obtained (see Figure 2.11). Here, a bar chart is said to be positively skewed if the long "tail" is on the right, and it is said to be negatively skewed if the long "tail" is on the left.

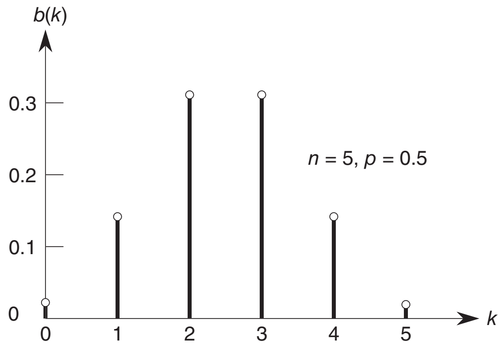


Figure 2.9. Symmetric binomial pmf

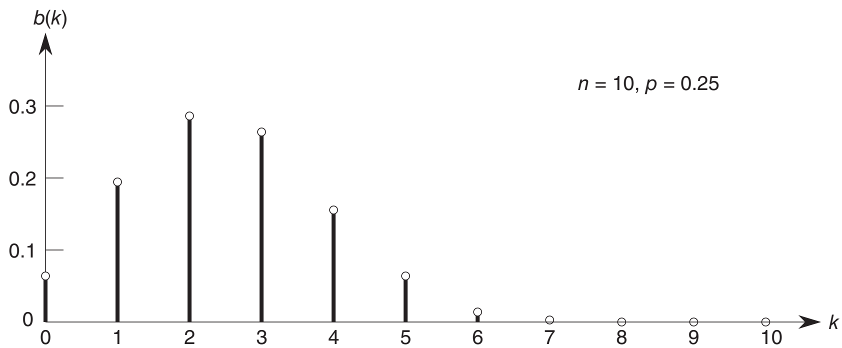


Figure 2.10. Positively skewed binomial pmf

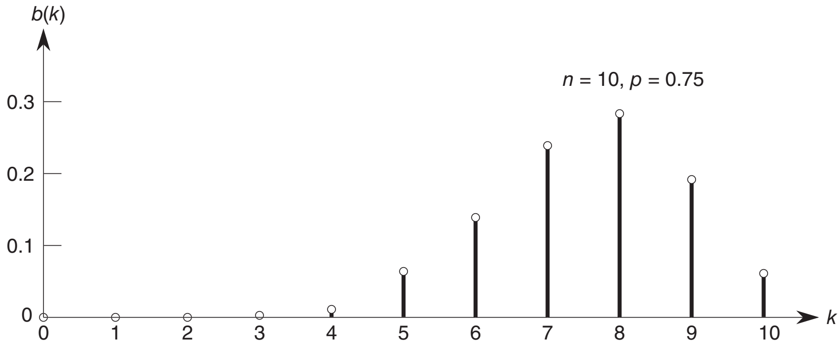


Figure 2.11. Negatively skewed binomial pmf

2.5.3 The Geometric pmf

Once again we consider a sequence of Bernoulli trials, but instead of counting the number of successes in a fixed number n of trials, we count the number of trials until the first "success" occurs. If we let 0 denote a failure and let 1 denote a success then the sample space of this experiment consists of the set of all binary strings with an arbitrary number of 0s followed by a single 1:

$$S = \{0^{i-1}1 | i = 1, 2, 3, \ldots\}.$$

Note that this sample space has a countably infinite number of sample points. Define a random variable Z on this sample space so that the value assigned to the sample point $0^{i-1}1$ is i. Thus Z is the number of trials up to and including the first success. Therefore, Z is a random variable with image $\{1, 2, \ldots\}$, which is a countably infinite set. To find the pmf of Z, we note that the event $[Z = i]$ occurs if and only if we have a sequence of $i - 1$ failures followed by one success. This is a sequence of independent Bernoulli trials with the probability of success equal to p. Hence, we have

$$\begin{aligned} p_Z(i) &= q^{i-1}p \\ &= p(1-p)^{i-1} \qquad \text{for } i = 1, 2, \ldots, \end{aligned} \tag{2.8}$$

where $q = 1 - p$. By the formula for the sum of a geometric series, we have

$$\begin{aligned} \sum_{i=1}^{\infty} p_Z(i) &= \sum_{i=1}^{\infty} pq^{i-1} \\ &= \frac{p}{1-q} \\ &= \frac{p}{p} \\ &= 1. \end{aligned}$$

Any random variable Z with the image $\{1, 2, \ldots\}$ and pmf given by a formula of the form of equation (2.8) is said to have a **geometric distribution**, and the function given by (2.8) is termed a **geometric pmf** with parameter p. The distribution function of Z is given by

$$F_Z(t) = \sum_{i=1}^{\lfloor t \rfloor} p(1-p)^{i-1}$$

$$= 1 - (1-p)^{\lfloor t \rfloor} \qquad \text{for } t \geq 0. \tag{2.9}$$

Graphs of the geometric pmf for two different values of parameter p are sketched in Figure 2.12.

The random variable Z counts the total number of trials up to and including the first success. We are often interested in counting the number of failures before the first success. Let this number be called the random variable X with the image $\{0, 1, 2, \ldots\}$. Clearly, $Z = X + 1$. The random variable X is said to

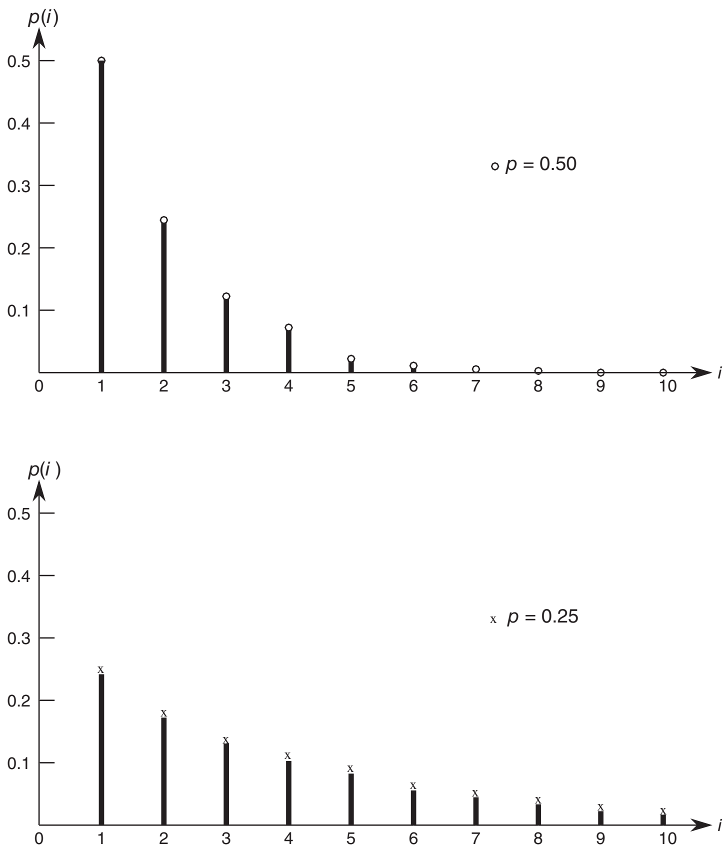


Figure 2.12. Geometric pmf

have a **modified geometric pmf**, specified by

$$p_X(i) = p(1-p)^i \qquad \text{for } i = 0, 1, 2, \ldots. \tag{2.10}$$

The distribution function of X is given by

$$\begin{aligned} F_X(t) &= \sum_{i=0}^{\lfloor t \rfloor} p(1-p)^i \\ &= 1 - (1-p)^{\lfloor t+1 \rfloor} \qquad \text{for } t \geq 0. \end{aligned} \tag{2.11}$$

The geometric (and modified geometric) distribution is encountered in some problems in queuing theory. Following are several examples where this distribution occurs:

1. A series of components is made by a certain manufacturer. The probability that any given component is defective is a constant p, which does not depend on the quality of the previous components. The probability that the ith item is the first defective one is given by formula (2.8).
2. Consider the scheduling of a computer system with a fixed time slice (see Figure 18). At the end of a time slice, the program would have completed execution with a probability p; thus there is a probability $q = 1 - p > 0$ that it needs to perform more computation. The pmf of the random variable denoting the number of time slices needed to complete the execution of a program is given by formula (2.8), if we assume that the operation of the computer satisfies the usual independence assumptions.
3. Consider the following program segment consisting of a **while** loop:

   while $\neg\, B$ **do** S

   Assume that the Boolean expression B takes the value **true** with probability p and the value **false** with probability q. If the successive tests on B are independent, then the number of times the body (or the statement group S) of the loop is executed will be a random variable having a modified geometric distribution with parameter p.
4. With the assumptions as in Example 3 above, consider a **repeat** loop:

   repeat S **until** B

   The number of times the body of the **repeat** loop is executed will be a geometrically distributed random variable with parameter p.

The geometric distribution has an important property, known as the **memoryless property**. Furthermore, it is the only discrete distribution with this property. To illustrate this property, consider a sequence of Bernoulli trials and let Z represent the number of trials until the first success. Now assume that we have observed a fixed number n of these trials and found them all to be failures. Let Y denote the number of additional trials that must be performed until the first success. Then $Y = Z - n$, and the conditional probability is

$$\begin{aligned} q_i &= P(Y = i|Z > n) \\ &= P(Z - n = i|Z > n) \\ &= P(Z = n + i|Z > n) \\ &= \frac{P(Z = n + i \text{ and } Z > n)}{P(Z > n)} \end{aligned}$$

by using the definition of conditional probability. But for $i = 1, 2, 3, \ldots$, $Z = n + i$ implies that $Z > n$. Thus the event $[Z = n + i \text{ and } Z > n]$ is the same as the event $[Z = n + i]$. Therefore

$$\begin{aligned} q_i &= P(Y = i|Z > n) \\ &= \frac{P(Z = n + i)}{P(Z > n)} \\ &= \frac{p_Z(n + i)}{1 - F_Z(n)} \\ &= \frac{pq^{n+i-1}}{1 - (1 - q^n)} \\ &= \frac{pq^{n+i-1}}{q^n} \\ &= pq^{i-1} \\ &= p_Z(i). \end{aligned}$$

Thus we see that, conditioned on $Z > n$, the number of trials remaining until the first success, $Y = Z - n$, has the same pmf as Z had originally. If a run of failures is observed in a sequence of Bernoulli trials, we need not "remember" how long the run was to determine the probabilities for the number of additional trials needed until the first success. The proof that any discrete random variable Z with image $\{1, 2, 3, \ldots\}$ and having the memoryless property must have the geometric distribution is left as an exercise.

2.5.4 The Negative Binomial pmf

To obtain the geometric pmf, we observed the number of trials until the first success in a sequence of Bernoulli trials. Now let us observe the number of trials until the r^{th} success, and let T_r be the random variable denoting this number. It is clear that the image of T_r is $\{r, r+1, r+2, \ldots\}$. To compute $p_{T_r}(n)$, define the events:

$$\begin{aligned} A &= \text{“}T_r = n\text{.”} \\ B &= \text{“Exactly } r-1 \text{ successes occur in } n-1 \text{ trials.”} \\ C &= \text{“The } n\text{th trial results in a success.”} \end{aligned}$$

Then clearly

$$A = B \cap C$$

and the events B and C are independent. Therefore

$$P(A) = P(B)P(C).$$

To compute $P(B)$, consider a particular sequence of $n-1$ trials with $r-1$ successes and $n-1-(r-1) = n-r$ failures. The probability associated with such a sequence is $p^{r-1}q^{n-r}$ and there are $\binom{n-1}{r-1}$ such sequences. Therefore

$$P(B) = \binom{n-1}{r-1} p^{r-1} q^{n-r}.$$

Now since $P(C) = p$,

$$\begin{aligned} p_{T_r}(n) &= P(T_r = n) \\ &= P(A) \\ &= \binom{n-1}{r-1} p^r q^{n-r} \\ &= \binom{n-1}{r-1} p^r (1-p)^{n-r}, \qquad n = r, r+1, r+2, \ldots. \end{aligned}$$

Using some combinatorial identities [KNUT 1997; p. 57], an alternative form of this pmf can be established:

$$p_{T_r}(n) = p^r \binom{-r}{n-r} (-1)^{n-r} (1-p)^{n-r}, \quad n = r, r+1, r+2, \ldots. \tag{2.12}$$

This pmf is known as the **negative binomial pmf**, and although we derived it assuming an integral value of r, any positive real value of r is allowed (of course the interpretation of r as a number of successes is no longer applicable). Quite clearly, if we let $r = 1$ in the formula (2.12), then we get the geometric pmf.

To verify that $\sum_{n=r}^{\infty} p_{T_r}(n) = 1$, we recall that the Taylor series expansion of $(1-t)^{-r}$ for $-1 < t < 1$ is

$$(1-t)^{-r} = \sum_{n=r}^{\infty} \binom{-r}{n-r} (-t)^{n-r}.$$

Substituting $t = 1 - p$, we have

$$p^{-r} = \sum_{n=r}^{\infty} \binom{-r}{n-r} (-1)^{n-r} (1-p)^{n-r},$$

which gives us the required result.

As in the case of the geometric distribution, there is a modified version of the negative binomial distribution. Let the random variable Z denote the number of failures before the occurrence of the r^{th} success. Then Z is said to have the **modified negative binomial** distribution with the pmf:

$$p_Z(n) = \binom{n+r-1}{r-1} p^r (1-p)^n, \qquad n \geq 0. \tag{2.13}$$

The pmf in equation (2.13) reduces the modified geometric pmf when $r = 1$.

2.5.5 The Poisson pmf

Let us consider another problem related to the binomial distribution. Suppose that we are observing the arrival of jobs to a large database server for the time interval $(0, t]$. It is reasonable to assume that for a small interval of duration Δt the probability of a new job arrival is $\lambda \cdot \Delta t$, where λ is a constant that depends upon the user population of the database server. If Δt is sufficiently small, then the probability of two or more jobs arriving in the interval of duration Δt may be neglected. We are interested in calculating the probability of k jobs arriving in the interval of duration t.

Suppose that the interval $(0, t]$ is divided into n subintervals of length t/n, and suppose further that the arrival of a job in any given interval is independent of the arrival of a job in any other interval. Then for a sufficiently large n, we can think of the n intervals as constituting a sequence of Bernoulli trials with the probability of success $p = \lambda t/n$. It follows that the probability of k arrivals in a total of n intervals each with a duration t/n is approximately given by

$$b\left(k; n, \frac{\lambda t}{n}\right) = \binom{n}{k} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

Since the assumption that the probability of more than one arrival per interval can be neglected is reasonable if and only if t/n is very small, we will take the limit of the above pmf as n approaches ∞. Now

$$\begin{aligned}
b\left(k; n, \frac{\lambda t}{n}\right) &= \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!n^k} (\lambda t)^k \cdot \left(1 - \frac{\lambda t}{n}\right)^{(n-k)} \\
&= \frac{n}{n} \cdot \frac{n-1}{n} \cdot \cdots \cdot \frac{n-k+1}{n} \cdot \frac{(\lambda t)^k}{k!} \cdot \left(1 - \frac{\lambda t}{n}\right)^{-k} \cdot \left(1 - \frac{\lambda t}{n}\right)^n.
\end{aligned}$$

We are interested in what happens to this expression as n increases, because then the subinterval width approaches zero, and the approximation involved gets better and better. In the limit as n approaches infinity, the first k factors approach unity, the next factor is fixed, the next approaches unity, and the last factor becomes

$$\lim_{n\to\infty}\left\{\left[1-\frac{\lambda t}{n}\right]^{-n/(\lambda t)}\right\}^{-\lambda t}.$$

Setting $-\lambda t/n = h$, this factor is

$$\left[\lim_{h\to 0}(1+h)^{1/h}\right]^{-\lambda t} = e^{-\lambda t},$$

since the limit in the brackets is the common definition of e. Thus, the binomial pmf approaches

$$\frac{e^{-\lambda t}(\lambda t)^k}{k!}, \qquad k = 0, 1, 2, \ldots.$$

Now replacing λt by a single parameter α, we get the well-known Poisson pmf:

$$f(k;\alpha) = e^{-\alpha}\frac{\alpha^k}{k!}, \qquad k = 0, 1, 2, \ldots. \tag{2.14}$$

Thus the Poisson pmf can be used as a convenient approximation to the binomial pmf when n is large and p is small:

$$\binom{n}{k}p^k q^{n-k} \simeq e^{-\alpha}\frac{\alpha^k}{k!}, \qquad \text{where } \alpha = np.$$

An acceptable rule of thumb is to use the Poisson approximation for binomial probabilities if $n \geq 20$ and $p \leq 0.05$. The table that follows compares $b(k; 5, 0.2)$ and $b(k; 20, 0.05)$ with $f(k; 1)$. Observe that the approximation is better in the case of larger n and smaller p.

k	$b(k; 5, 0.2)$	$b(k; 20, 0.05)$	$f(k; 1)$
0	0.328	0.359	0.368
1	0.410	0.377	0.368
2	0.205	0.189	0.184
3	0.051	0.060	0.061

There are other recommendations from different authors concerning which values of n and p are appropriate. Normal approximation was introduced earlier in this chapter as one way to approximate the binomial pmf. It is useful when n is large and $p \approx 1/2$. The Poisson approximation, although less popular, is good for large values of n and small values of p.

Besides errors in computing binomial probabilities, if the approximation is used in parameter estimation as in Chapter 10, the effect of approximation on the confidence interval should also be considered when using these approximations [LEEM 1996].

Example 2.10

A manufacturer produces VLSI chips, 1% of which are defective. Find the probability that in a box containing 100 chips, no defectives are found.

Since $n = 100$ and $p = 0.01$, the required answer is

$$\begin{aligned} b(0; 100, 0.01) &= \binom{100}{0} \cdot 0.01^0 \cdot 0.99^{100} \\ &= 0.99^{100} \\ &= 0.366. \end{aligned}$$

Using the Poisson approximation, $\alpha = 100 \cdot 0.01 = 1$, and the required answer is

$$\begin{aligned} f(0; 1) &= e^{-1} \\ &= 0.3679. \end{aligned}$$

♯

It is easily verified that the probabilities from equation (2.14) are nonnegative and sum to 1:

$$\begin{aligned} \sum_{k=0}^{\infty} f(k; \alpha) &= \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} e^{-\alpha} \\ &= e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} \\ &= e^{-\alpha} \cdot e^{\alpha} \\ &= 1. \end{aligned}$$

The probabilities $f(k; \alpha)$ are easy to calculate, starting with

$$f(0; \alpha) = e^{-\alpha}$$

and using the recurrence relation

$$f(k+1; \alpha) = \frac{\alpha f(k; \alpha)}{k+1}. \tag{2.15}$$

For very large values of α, special care is necessary to avoid numerical problems in computing Poisson pmf. Fox and Glynn have published an algorithm that is recommended for this purpose [FOX 1988].

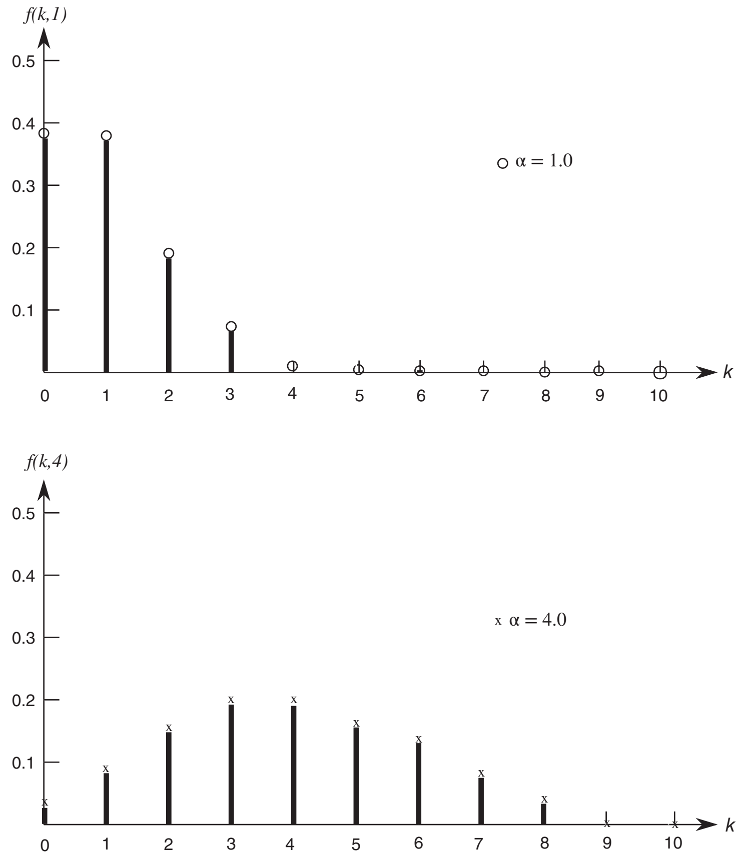


Figure 2.13. Poisson pmf

The Poisson probabilities have been tabulated [PEAR 1966]for $\alpha = 0.1$ to 15, in the increments of 0.1 (in Appendix C, we have tabulated the Poisson CDF). In Figure 2.13, we have plotted the Poisson pmf with parameters $\alpha = 1$ and $\alpha = 4$. Note that this pmf is positively skewed; in fact, it can be shown that the Poisson pmf is positively skewed for any $\alpha > 0$.

Apart from its ability to approximate a binomial pmf, the Poisson pmf is found to be useful in many other situations. In reliability theory, it is quite reasonable to assume that the probability of k components malfunctioning within an interval of time t in a system with a large number of components is given by the Poisson pmf (here λ is known as the *component failure rate*):

$$f(k; \lambda t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \qquad k = 0, 1, 2, \ldots . \tag{2.16}$$

In studying systems with congestion (or queuing), we find that the number of jobs arriving, the number of jobs completing service, or the number of messages transmitted through a communication channel in a fixed interval of time is approximately Poisson distributed.

2.5.6 The Hypergeometric pmf

We have noted earlier that the binomial pmf is obtained while "sampling with replacement." The hypergeometric pmf is obtained while "sampling without replacement." Let us select a random sample of m components from a box containing n components, d of which are known to be defective. For the first component selected, the probability that it is defective is given by d/n, but for the second selection it remains d/n only if the first component selected is replaced. Otherwise, this probability is $(d-1)/(n-1)$ or $d/(n-1)$ depending on whether or not a defective component was selected in the first drawing. Thus the assumption of a constant probability of success, as in a sequence of Bernoulli trials, is not satisfied.

We are interested in computing the **hypergeometric pmf**, $h(k; m, d, n)$, defined as the probability of choosing k defective components in a random sample of m components, chosen without replacement, from a total of n components, d of which are defective. The sample space of this experiment consists of $\binom{n}{m}$ sample points. The k defectives can be selected from d defectives in $\binom{d}{k}$ ways, and the $m-k$ non-defective components may be selected from $n-d$ non-defectives in $\binom{n-d}{m-k}$ ways. Therefore, the whole sample of m components with k defectives can be selected in $\binom{d}{k} \cdot \binom{n-d}{m-k}$ ways. Assuming an equiprobable sample space, the required probability is

$$h(k; m, d, n) = \frac{\dbinom{d}{k} \cdot \dbinom{n-d}{m-k}}{\dbinom{n}{m}}, \qquad k = 0, 1, 2, \ldots, \ \min\{d, m\}. \tag{2.17}$$

Example 2.11

Compute the probability of obtaining three defectives in a sample of size 10 taken without replacement from a box of twenty components containing four defectives.

We are required to compute

$$\begin{aligned} h(3; 10, 4, 20) &= \frac{\dbinom{4}{3} \cdot \dbinom{16}{7}}{\dbinom{20}{10}} \\ &= \frac{4 \cdot 11,440}{184,756} \\ &= 0.247678. \end{aligned}$$

If we were to approximate this probability using a binomial distribution with $n = 10$ and $p = 4/20 = 0.20$, we will get $b(3; 10, 0.20) = 0.2013$, a considerable underestimate of the actual probability.

♯

Example 2.12

Return to the TDMA (time division multiple access) wireless system example from Chapter 1 [SUN 1999], where the base transceiver system of each cell has n base repeaters [also called *base radio* (BR)]. Each base repeater provides m time-division-multiplexed channels.

A base repeater is subject to failure. Suppose the channels are allocated randomly to the users. Denote the total number of talking channels in the whole system as k when the failure occurs. Then the probability that i talking channels reside in the failed base repeater, is given by $p_i = h(i; k, m, mn)$.

♯

Example 2.13

A software reliability growth model for estimating the number of residual faults in the software after testing phase based on hypergeometric distribution has been proposed [TOHM 1989].

During the testing phase a software is subjected to a sequence of test instances t_i, $i = 1, 2, \ldots n$. Faults detected by each test instance are assumed to have been removed without introducing new faults before the next test instance is exercised. Assume that the total number of faults initially introduced into the software is m. The test instance t_i senses w_i initial faults out of m initial faults. The sensitization of the faults is distinguished from the detection of the faults in the following way. The number of faults detected by the first test instance t_1 is obviously w_1. However, the number of faults detected by t_2 is not necessarily w_2, because some faults may be removed already in t_1. Similarly, faults detected by t_3 are those that are not yet sensed by t_1 and t_2.

If the number of faults detected by t_i is denoted by N_i then the cumulative number of faults detected by test instances from t_1 to t_i is given by the random variable:

$$C_i = \sum_{j=1}^{i} N_j;$$

that is, the number of faults still remaining undetected in the software after the test instance t_i is $m - C_i$. It follows that the probability that k faults are detected by the test instance t_{i+1} given that c_i faults are detected by test instances t_1 through t_i is

$$P(N_{i+1} = k) = h(k; w_{i+1}, m - c_i, m) = \frac{\binom{m - c_i}{k}\binom{c_i}{w_{i+1} - k}}{\binom{m}{w_{i+1}}}.$$

♯

In situations where the sample size m is small compared to the lot size n, the binomial distribution provides a good approximation to the hypergeometric distribution; that is, $h(k; m, d, n) \simeq b(k; m, d/n)$ for large n.

2.5.7 The Discrete Uniform pmf

Let X be a discrete random variable with a finite image $\{x_1, x_2, \ldots, x_n\}$. One of the simplest pmf's to consider in this case is one in which each value in the image has equal probability. If we require that $p_X(x_i) = p$ for all i, then, since

$$1 = \sum_{i=1}^{n} p_X(x_i) = \sum_{i=1}^{n} p = np,$$

it follows that

$$p_X(x_i) = \begin{cases} \frac{1}{n} & x_i \text{ in the image of } X, \\ 0 & \text{otherwise.} \end{cases}$$

Such a random variable is said to have a **discrete uniform distribution**. This distribution plays an important role in the theory of random numbers and its applications to discrete event simulation. In the average-case analysis of programs, it is often assumed that the input data are uniformly distributed over the input space.

Note that the concept of uniform distribution cannot be extended to a discrete random variable with a countably infinite image, $\{x_1, x_2, \ldots\}$. The requirements that $\sum_i p_X(x_i) = 1$ and $p_X(x_i) = \text{constant}$ (for $i = 1, 2, \ldots$) are incompatible.

If we let X take on the values $\{1, 2, \ldots, n\}$ with $p_X(i) = 1/n, 1 \le i \le n$, then its distribution function is given by

$$F_X(x) = \sum_{i=1}^{\lfloor x \rfloor} p_X(i)$$
$$= \frac{\lfloor x \rfloor}{n}, \qquad 1 \le x \le n.$$

A graph of this distribution with $n = 10$ is given in Figure 2.14.

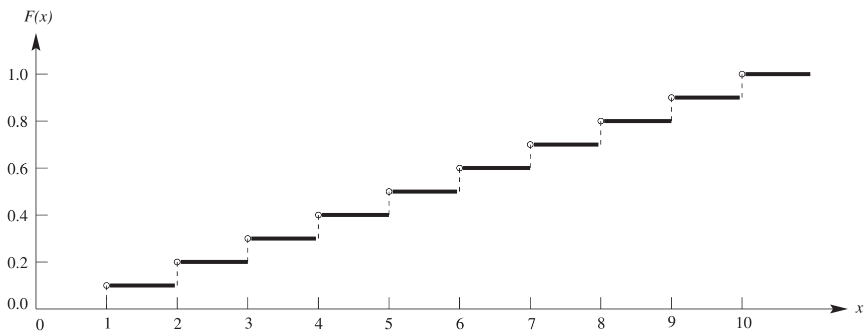


Figure 2.14. Discrete uniform distribution