4

Discrete random variables

The sample space associated with an experiment, together with a probability function defined on all its events, is a complete probabilistic description of that experiment. Often we are interested only in certain features of this description. We focus on these features using *random variables*. In this chapter we discuss *discrete* random variables, and in the next we will consider *continuous* random variables. We introduce the Bernoulli, binomial, and geometric random variables.

4.1 Random variables

Suppose we are playing the board game "Snakes and Ladders," where the moves are determined by the sum of two independent throws with a die. An obvious choice of the sample space is

$$\begin{aligned}\Omega &= \{(\omega_1, \omega_2) \,:\, \omega_1,\, \omega_2 \in \{1, 2, \ldots, 6\}\,\} \\ &= \{(1,1), (1,2), \ldots, (1,6), (2,1), \ldots, (6,5), (6,6)\}.\end{aligned}$$

However, as players of the game, we are *only* interested in the sum of the outcomes of the two throws, i.e., in the value of the function $S : \Omega \to \mathbb{R}$, given by

$$S(\,\omega_1, \omega_2\,) = \omega_1 + \omega_2 \quad \text{for } (\omega_1, \omega_2) \in \Omega.$$

In Table 4.1 the possible results of the first throw (top margin), those of the second throw (left margin), and the corresponding values of S (body) are given. Note that the values of S are constant on lines perpendicular to the diagonal. We denote the event that the function S attains the value k by $\{S = k\}$, which is an abbreviation of "the subset of those $\omega = (\omega_1, \omega_2) \in \Omega$ for which $S(\,\omega_1, \omega_2\,) = \omega_1 + \omega_2 = k$," i.e.,

$$\{S = k\} = \{(\omega_1, \omega_2) \in \Omega \,:\, S(\,\omega_1, \omega_2) = k\,\}.$$

Table 4.1. Two throws with a die and the corresponding sum.

	ω_1					
ω_2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

QUICK EXERCISE 4.1 List the outcomes in the event $\{S = 8\}$.

We denote the probability of the event $\{S = k\}$ by

$$\mathrm{P}(S = k)\,,$$

although formally we should write $\mathrm{P}(\{S = k\})$ instead of $\mathrm{P}(S = k)$. In our example, S attains only the values $k = 2, 3, \ldots, 12$ with positive probability. For example,

$$\mathrm{P}(S = 2) = \mathrm{P}(\,(1, 1)\,) = \frac{1}{36},$$

$$\mathrm{P}(S = 3) = \mathrm{P}(\,\{(1, 2), (2, 1)\}\,) = \frac{2}{36},$$

while

$$\mathrm{P}(S = 13) = \mathrm{P}(\,\emptyset\,) = 0,$$

because 13 is an "impossible outcome."

QUICK EXERCISE 4.2 Use Table 4.1 to determine $\mathrm{P}(S = k)$ for $k = 4, 5, \ldots, 12$.

Now suppose that for some other game the moves are given by the maximum of two independent throws. In this case we are interested in the value of the function $M : \Omega \to \mathbb{R}$, given by

$$M(\,\omega_1, \omega_2\,) = \max\{\omega_1, \omega_2\} \quad \text{for } (\omega_1, \omega_2) \in \Omega.$$

In Table 4.2 the possible results of the first throw (top margin), those of the second throw (left margin), and the corresponding values of M (body) are given. The functions S and M are examples of what we call discrete random variables.

DEFINITION. Let Ω be a sample space. A *discrete random variable* is a function $X : \Omega \to \mathbb{R}$ that takes on a finite number of values $a_1, a_2, \ldots, a_n$ or an infinite number of values $a_1, a_2, \ldots$.

Table 4.2. Two throws with a die and the corresponding maximum.

	ω_1					
ω_2	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

In a way, a discrete random variable X "transforms" a sample space Ω to a more "tangible" sample space $\tilde{\Omega}$, whose events are more directly related to what you are interested in. For instance, S transforms $\Omega = \{(1,1), (1,2), \ldots, (1,6), (2,1), \ldots, (6,5), (6,6)\}$ to $\tilde{\Omega} = \{2, \ldots, 12\}$, and M transforms Ω to $\tilde{\Omega} = \{1, \ldots, 6\}$. Of course, there is a price to pay: one has to calculate the probabilities of X. Or, to say things more formally, one has to determine the *probability distribution* of X, i.e., to describe how the probability mass is *distributed* over possible values of X.

4.2 The probability distribution of a discrete random variable

Once a discrete random variable X is introduced, the sample space Ω is no longer important. It suffices to list the possible values of X and their corresponding probabilities. This information is contained in the *probability mass function* of X.

DEFINITION. The *probability mass function* p of a discrete random variable X is the function $p : \mathbb{R} \to [0, 1]$, defined by

$$p(a) = \mathrm{P}(X = a) \qquad \text{for } -\infty < a < \infty.$$

If X is a discrete random variable that takes on the values $a_1, a_2, \ldots$, then

$$p(a_i) > 0, \quad p(a_1) + p(a_2) + \cdots = 1, \qquad \text{and } p(a) = 0 \text{ for all other } a.$$

As an example we give the probability mass function p of M.

a	1	2	3	4	5	6
$p(a)$	1/36	3/36	5/36	7/36	9/36	11/36

Of course, $p(a) = 0$ for all other a.

The distribution function of a random variable

As we will see, so-called continuous random variables cannot be specified by giving a probability mass function. However, the *distribution function* of a random variable X (also known as the *cumulative distribution function*) allows us to treat discrete and continuous random variables in the same way.

> DEFINITION. The *distribution function* F of a random variable X is the function $F : \mathbb{R} \to [0, 1]$, defined by
>
> $$F(a) = \mathrm{P}(X \leq a) \quad \text{for } -\infty < a < \infty.$$

Both the probability mass function and the distribution function of a discrete random variable X contain all the probabilistic information of X; the *probability distribution* of X is determined by either of them. In fact, the distribution function F of a discrete random variable X can be expressed in terms of the probability mass function p of X and vice versa. If X attains values $a_1, a_2, \ldots$, such that

$$p(a_i) > 0, \quad p(a_1) + p(a_2) + \cdots = 1,$$

then

$$F(a) = \sum_{a_i \leq a} p(a_i).$$

We see that, for a discrete random variable X, the distribution function F jumps in each of the a_i, and is constant between successive a_i. The height of the jump at a_i is $p(a_i)$; in this way p can be retrieved from F. For example, see Figure 4.1, where p and F are displayed for the random variable M.

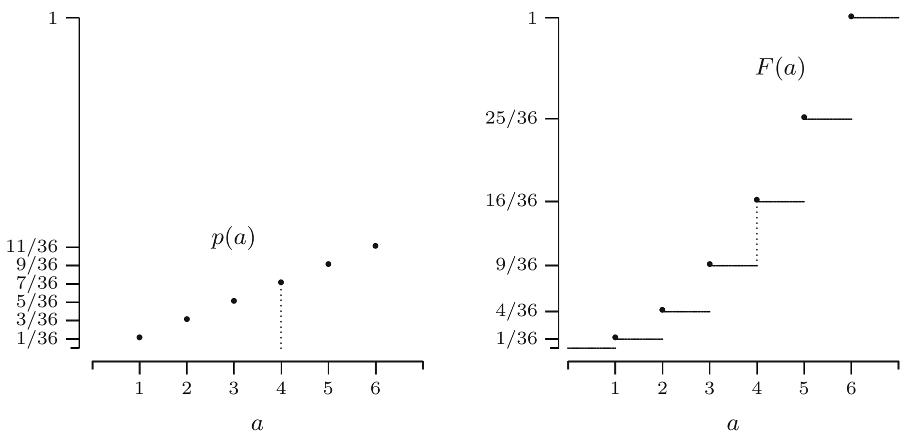

Fig. 4.1. Probability mass function and distribution function of M.

We end this section with three properties of the distribution function F of a random variable X:

1. For $a \leq b$ one has that $F(a) \leq F(b)$. This property is an immediate consequence of the fact that $a \leq b$ implies that the event $\{X \leq a\}$ is contained in the event $\{X \leq b\}$.
2. Since $F(a)$ is a probability, the value of the distribution function is always between 0 and 1. Moreover,

$$\lim_{a \to +\infty} F(a) = \lim_{a \to +\infty} \mathrm{P}(X \leq a) = 1$$
$$\lim_{a \to -\infty} F(a) = \lim_{a \to -\infty} \mathrm{P}(X \leq a) = 0.$$

3. F is right-continuous, i.e., one has

$$\lim_{\varepsilon \downarrow 0} F(a + \varepsilon) = F(a).$$

   This is indicated in Figure 4.1 by bullets. Henceforth we will omit these bullets.

Conversely, any function F satisfying 1, 2, and 3 is the distribution function of some random variable (see Remarks 6.1 and 6.2).

Quick exercise 4.3 Let X be a discrete random variable, and let a be such that $p(a) > 0$. Show that $F(a) = \mathrm{P}(X < a) + p(a)$.

There are many discrete random variables that arise in a natural way. We introduce three of them in the next two sections.

4.3 The Bernoulli and binomial distributions

The Bernoulli distribution is used to model an experiment with only two possible outcomes, often referred to as "success" and "failure", usually encoded as 1 and 0.

> Definition. A discrete random variable X has a *Bernoulli distribution* with parameter p, where $0 \leq p \leq 1$, if its probability mass function is given by
>
> $$p_X(1) = \mathrm{P}(X = 1) = p \quad \text{and} \quad p_X(0) = \mathrm{P}(X = 0) = 1 - p.$$
>
> We denote this distribution by $Ber(p)$.

Note that we wrote p_X instead of p for the probability mass function of X. This was done to emphasize its dependence on X and to avoid possible confusion with the parameter p of the Bernoulli distribution.

Consider the (fictitious) situation that you attend, completely unprepared, a multiple-choice exam. It consists of 10 questions, and each question has four alternatives (of which only one is correct). You will pass the exam if you answer six or more questions correctly. You decide to answer each of the questions in a random way, in such a way that the answer of one question is not affected by the answers of the others. What is the probability that you will pass?

Setting for $i = 1, 2, \ldots, 10$

$$R_i = \begin{cases} 1 & \text{if the } i\text{th answer is correct} \\ 0 & \text{if the } i\text{th answer is incorrect,} \end{cases}$$

the number of correct answers X is given by

$$X = R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8 + R_9 + R_{10}.$$

Quick exercise 4.4 Calculate the probability that you answered the first question correctly and the second one incorrectly.

Clearly, X attains only the values $0, 1, \ldots, 10$. Let us first consider the case $X = 0$. Since the answers to the different questions do not influence each other, we conclude that the events $\{R_1 = a_1\}, \ldots, \{R_{10} = a_{10}\}$ are independent for every choice of the a_i, where each a_i is 0 or 1. We find

$$\begin{aligned} \mathrm{P}(X = 0) &= \mathrm{P}(\text{not a single } R_i \text{ equals } 1) \\ &= \mathrm{P}(R_1 = 0, R_2 = 0, \ldots, R_{10} = 0) \\ &= \mathrm{P}(R_1 = 0)\,\mathrm{P}(R_2 = 0) \cdots \mathrm{P}(R_{10} = 0) \\ &= \left(\frac{3}{4}\right)^{10}. \end{aligned}$$

The probability that we have answered exactly one question correctly equals

$$\mathrm{P}(X = 1) = \frac{1}{4} \cdot \left(\frac{3}{4}\right)^9 \cdot 10,$$

which is the probability that the answer is correct times the probability that the other nine answers are wrong, times the number of ways in which this can occur:

$$\begin{aligned} \mathrm{P}(X = 1) = \;& \mathrm{P}(R_1 = 1)\,\mathrm{P}(R_2 = 0)\,\mathrm{P}(R_3 = 0) \cdots \mathrm{P}(R_{10} = 0) \\ & + \mathrm{P}(R_1 = 0)\,\mathrm{P}(R_2 = 1)\,\mathrm{P}(R_3 = 0) \cdots \mathrm{P}(R_{10} = 0) \\ & \quad \vdots \\ & + \mathrm{P}(R_1 = 0)\,\mathrm{P}(R_2 = 0)\,\mathrm{P}(R_3 = 0) \cdots \mathrm{P}(R_{10} = 1). \end{aligned}$$

In general we find for $k = 0, 1, \ldots, 10$, again using independence, that

$$\mathrm{P}(X = k) = \left(\frac{1}{4}\right)^k \cdot \left(\frac{3}{4}\right)^{10-k} \cdot C_{10,k},$$

which is the probability that k questions were answered correctly times the probability that the other $10-k$ answers are wrong, times the number of ways $C_{10,k}$ this can occur.

So $C_{10,k}$ is the number of different ways in which one can choose k correct answers from the list of 10. We already have seen that $C_{10,0} = 1$, because there is only one way to do everything wrong; and that $C_{10,1} = 10$, because each of the 10 questions may have been answered correctly.

More generally, if we have to choose k different objects out of an ordered list of n objects, and the order in which we pick the objects matters, then for the first object you have n possibilities, and no matter which object you pick, for the second one there are $n-1$ possibilities. For the third there are $n-2$ possibilities, and so on, with $n-(k-1)$ possibilities for the kth. So there are

$$n(n-1)\cdots(n-(k-1))$$

ways to choose the k objects.

In how many ways can we choose three questions? When the order matters, there are $10\cdot 9\cdot 8$ ways. However, the order in which these three questions are selected does *not* matter: to answer questions 2, 5, and 8 correctly is the same as answering questions 8, 2, and 5 correctly, and so on. The triplet $\{2,5,8\}$ can be chosen in $3\cdot 2\cdot 1$ different orders, all with the same result. There are six permutations of the numbers 2, 5, and 8 (see page 14).

Thus, compensating for this six-fold overcount, the number $C_{10,3}$ of ways to correctly answer 3 questions out of 10 becomes

$$C_{10,3} = \frac{10\cdot 9\cdot 8}{3\cdot 2\cdot 1}.$$

More generally, for $n \geq 1$ and $1 \leq k \leq n$,

$$C_{n,k} = \frac{n(n-1)\cdots(n-(k-1))}{k(k-1)\cdots 2\cdot 1}.$$

Note that this is equal to

$$\frac{n!}{k!\,(n-k)!},$$

which is usually denoted by $\binom{n}{k}$, so $C_{n,k} = \binom{n}{k}$. Moreover, in accordance with $0! = 1$ (as defined in Chapter 2), we put $C_{n,0} = \binom{n}{0} = 1$.

Quick exercise 4.5 Show that $\binom{n}{n-k} = \binom{n}{k}$.

Substituting $\binom{10}{k}$ for $C_{10,k}$ we obtain

$$\mathrm{P}(X = k) = \binom{10}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{10-k}.$$

Since $\mathrm{P}(X \geq 6) = \mathrm{P}(X = 6) + \cdots + \mathrm{P}(X = 10)$, it is now an easy (but tedious) exercise to determine the probability that you will pass. One finds that $\mathrm{P}(X \geq 6) = 0.0197$. It pays to study, doesn't it?!

The preceding random variable X is an example of a random variable with a binomial distribution with parameters $n = 10$ and $p = 1/4$.

DEFINITION. A discrete random variable X has a *binomial distribution* with parameters n and p, where $n = 1, 2, \ldots$ and $0 \leq p \leq 1$, if its probability mass function is given by

$$p_X(k) = \mathrm{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, \ldots, n.$$

We denote this distribution by $Bin(n, p)$.

Figure 4.2 shows the probability mass function p_X and distribution function F_X of a $Bin(10, \frac{1}{4})$ distributed random variable.

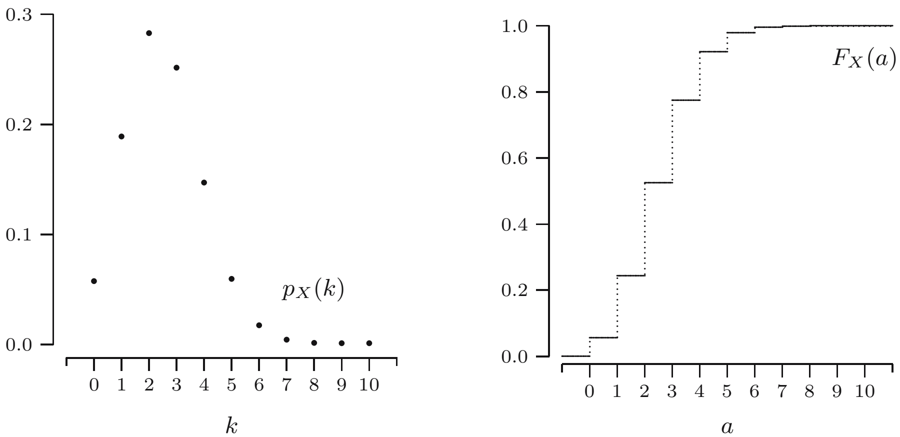

Fig. 4.2. Probability mass function and distribution function of the $Bin(10, \frac{1}{4})$ distribution.

4.4 The geometric distribution

In 1986, Weinberg and Gladen [38] investigated the number of menstrual cycles it took women to become pregnant, measured from the moment they had decided to become pregnant. We model the number of cycles up to pregnancy by a random variable X.

Assume that the probability that a woman becomes pregnant during a particular cycle is equal to p, for some p with $0 < p \leq 1$, independent of the previous cycles. Then clearly $\mathrm{P}(X = 1) = p$. Due to the independence of consecutive cycles, one finds for $k = 1, 2, \ldots$ that

$$\begin{aligned} \mathrm{P}(X = k) &= \mathrm{P}(\text{no pregnancy in the first } k-1 \text{ cycles, pregnancy in the } k\text{th}) \\ &= (1-p)^{k-1}p. \end{aligned}$$

This random variable X is an example of a random variable with a geometric distribution with parameter p.

DEFINITION. A discrete random variable X has a *geometric distribution* with parameter p, where $0 < p \leq 1$, if its probability mass function is given by

$$p_X(k) = \mathrm{P}(X = k) = (1-p)^{k-1}\, p \quad \text{for } k = 1, 2, \ldots .$$

We denote this distribution by $Geo(p)$.

Figure 4.3 shows the probability mass function p_X and distribution function F_X of a $Geo(\frac{1}{4})$ distributed random variable.

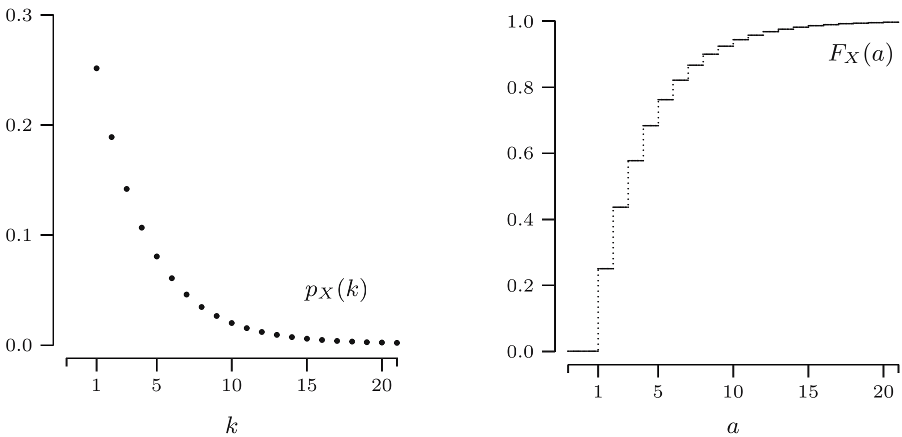

Fig. 4.3. Probability mass function and distribution function of the $Geo(\frac{1}{4})$ distribution.

QUICK EXERCISE 4.6 Let X have a $Geo(p)$ distribution. For $n \geq 0$, show that $\mathrm{P}(X > n) = (1-p)^n$.

The geometric distribution has a remarkable property, which is known as the *memoryless property*.[1] For $n, k = 0, 1, 2, \ldots$ one has

$$\mathrm{P}(X > n + k \mid X > k) = \mathrm{P}(X > n)\,.$$

We can derive this equality using the result from Quick exercise 4.6:

$$\begin{aligned}\mathrm{P}(X > n + k \mid X > k) &= \frac{\mathrm{P}(\{X > k + n\} \cap \{X > k\})}{\mathrm{P}(X > k)} \\ &= \frac{\mathrm{P}(X > k + n)}{\mathrm{P}(X > k)} = \frac{(1-p)^{n+k}}{(1-p)^k} \\ &= (1-p)^n = \mathrm{P}(X > n)\,.\end{aligned}$$

4.5 Solutions to the quick exercises

4.1 From Table 4.1, one finds that

$$\{S = 8\} = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}.$$

4.2 From Table 4.1, one determines the following table.

k	4	5	6	7	8	9	10	11	12
$\mathrm{P}(S = k)$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

4.3 Since $\{X \leq a\} = \{X < a\} \cup \{X = a\}$, it follows that

$$F(a) = \mathrm{P}(X \leq a) = \mathrm{P}(X < a) + \mathrm{P}(X = a) = \mathrm{P}(X < a) + p(a).$$

Not very interestingly: this also holds if $p(a) = 0$.

4.4 The probability that you answered the first question correctly and the second one incorrectly is given by $\mathrm{P}(R_1 = 1, R_2 = 0)$. Due to independence, this is equal to $\mathrm{P}(R_1 = 1)\,\mathrm{P}(R_2 = 0) = \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}$.

4.5 Rewriting yields

$$\binom{n}{n-k} = \frac{n!}{(n-k)!\,(n-(n-k))!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}.$$

[1] In fact, the geometric distribution is the only discrete random variable with this property.

4.6 There are two ways to show that $\mathrm{P}(X > n) = (1-p)^n$. The easiest way is to realize that $\mathrm{P}(X > n)$ is the probability that we had "no success in the first n trials," which clearly equals $(1-p)^n$. A more involved way is by calculation:

$$\begin{aligned}\mathrm{P}(X > n) &= \mathrm{P}(X = n+1) + \mathrm{P}(X = n+2) + \cdots \\ &= (1-p)^n p + (1-p)^{n+1} p + \cdots \\ &= (1-p)^n p \left(1 + (1-p) + (1-p)^2 + \cdots\right).\end{aligned}$$

If we recall from calculus that

$$\sum_{k=0}^{\infty} (1-p)^k = \frac{1}{1-(1-p)} = \frac{1}{p},$$

the answer follows immediately.

4.6 Exercises

4.1 ⊞ Let Z represent the number of times a 6 appeared in two independent throws of a die, and let S and M be as in Section 4.1.

a. Describe the probability distribution of Z, by giving either the probability mass function p_Z of Z or the distribution function F_Z of Z. What type of distribution does Z have, and what are the values of its parameters?
b. List the outcomes in the events $\{M = 2, Z = 0\}$, $\{S = 5, Z = 1\}$, and $\{S = 8, Z = 1\}$. What are their probabilities?
c. Determine whether the events $\{M = 2\}$ and $\{Z = 0\}$ are independent.

4.2 Let X be a discrete random variable with probability mass function p given by:

a	-1	0	1	2
$p(a)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$

and $p(a) = 0$ for all other a.

a. Let the random variable Y be defined by $Y = X^2$, i.e., if $X = 2$, then $Y = 4$. Calculate the probability mass function of Y.
b. Calculate the value of the distribution functions of X and Y in $a = 1$, $a = 3/4$, and $a = \pi - 3$.

4.3 ⊡ Suppose that the distribution function of a discrete random variable X is given by

$$F(a) = \begin{cases} 0 & \text{for } a < 0 \\ \frac{1}{3} & \text{for } 0 \le a < \frac{1}{2} \\ \frac{1}{2} & \text{for } \frac{1}{2} \le a < \frac{3}{4} \\ 1 & \text{for } a \ge \frac{3}{4}. \end{cases}$$

Determine the probability mass function of X.

4.4 You toss n coins, each showing heads with probability p, independently of the other tosses. Each coin that shows tails is tossed again. Let X be the total number of heads.

a. What type of distribution does X have? Specify its parameter(s).

b. What is the probability mass function of the total number of heads X?

4.5 A fair die is thrown until the sum of the results of the throws exceeds 6. The random variable X is the number of throws needed for this. Let F be the distribution function of X. Determine $F(1)$, $F(2)$, and $F(7)$.

4.6 ⊡ Three times we randomly draw a number from the following numbers:

$$1 \quad 2 \quad 3.$$

If X_i represents the ith draw, $i = 1, 2, 3$, then the probability mass function of X_i is given by

a	1	2	3
$\mathrm{P}(X_i = a)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

and $\mathrm{P}(X_i = a) = 0$ for all other a. We assume that each draw is independent of the previous draws. Let $\bar{X}$ be the average of X_1, X_2, and X_3, i.e.,

$$\bar{X} = \frac{X_1 + X_2 + X_3}{3}.$$

a. Determine the probability mass function $p_{\bar{X}}$ of $\bar{X}$.

b. Compute the probability that exactly two draws are equal to 1.

4.7 ⊡ A shop receives a batch of 1000 cheap lamps. The odds that a lamp is defective are 0.1%. Let X be the number of defective lamps in the batch.

a. What kind of distribution does X have? What is/are the value(s) of parameter(s) of this distribution?

b. What is the probability that the batch contains no defective lamps? One defective lamp? More than two defective ones?

4.8 ⊡ In Section 1.4 we saw that each space shuttle has six O-rings and that each O-ring fails with probability

$$p(t) = \frac{e^{a+b\cdot t}}{1+e^{a+b\cdot t}},$$

where $a = 5.085$, $b = -0.1156$, and t is the temperature (in degrees Fahrenheit) at the time of the launch of the space shuttle. At the time of the fatal launch of the *Challenger*, $t = 31$, yielding $p(31) = 0.8178$.

a. Let X be the number of failing O-rings at launch temperature 31°F. What type of probability distribution does X have, and what are the values of its parameters?

b. What is the probability $\mathrm{P}(X \geq 1)$ that at least one O-ring fails?

4.9 For simplicity's sake, let us assume that all space shuttles will be launched at 81°F (which is the highest recorded launch temperature in Figure 1.3). With this temperature, the probability of an O-ring failure is equal to $p(81) = 0.0137$ (see Section 1.4 or Exercise 4.8).

a. What is the probability that during 23 launches no O-ring will fail, but that at least one O-ring will fail during the 24th launch of a space shuttle?

b. What is the probability that no O-ring fails during 24 launches?

4.10 ⊞ Early in the morning, a group of m people decides to use the elevator in an otherwise deserted building of 21 floors. Each of these persons chooses his or her floor independently of the others, and—from our point of view—completely at random, so that each person selects a floor with probability 1/21. Let S_m be the number of times the elevator stops. In order to study S_m, we introduce for $i = 1, 2, \ldots, 21$ random variables R_i, given by

$$R_i = \begin{cases} 1 & \text{if the elevator stops at the } i\text{th floor} \\ 0 & \text{if the elevator does not stop at the } i\text{th floor.} \end{cases}$$

a. Each R_i has a $Ber(p)$ distribution. Show that $p = 1 - \left(\frac{20}{21}\right)^m$.

b. From the way we defined S_m, it follows that

$$S_m = R_1 + R_2 + \cdots + R_{21}.$$

Can we conclude that S_m has a $Bin(21, p)$ distribution, with p as in part **a**? Why or why not?

c. Clearly, if $m = 1$, one has that $\mathrm{P}(S_1 = 1) = 1$. Show that for $m = 2$

$$\mathrm{P}(S_2 = 1) = \frac{1}{21} = 1 - \mathrm{P}(S_2 = 2),$$

and that S_3 has the following distribution.

a	1	2	3
$\mathrm{P}(S_3 = a)$	1/441	60/441	380/441

4.11 You decide to play monthly in two different lotteries, and you stop playing as soon as you win a prize in one (or both) lotteries of at least one million euros. Suppose that every time you participate in these lotteries, the probability to win one million (or more) euros is p_1 for one of the lotteries and p_2 for the other. Let M be the number of times you participate in these lotteries until winning at least one prize. What kind of distribution does M have, and what is its parameter?

4.12 ⊡ You and a friend want to go to a concert, but unfortunately only one ticket is still available. The man who sells the tickets decides to toss a coin until heads appears. In each toss heads appears with probability p, where $0 < p < 1$, independent of each of the previous tosses. If the number of tosses needed is odd, your friend is allowed to buy the ticket; otherwise you can buy it. Would you agree to this arrangement?

4.13 ⊞ A box contains an unknown number N of identical bolts. In order to get an idea of the size N, we randomly mark one of the bolts from the box. Next we select at random a bolt from the box. If this is the marked bolt we stop, otherwise we return the bolt to the box, and we randomly select a second one, etc. We stop when the selected bolt is the marked one. Let X be the number of times a bolt was selected. Later (in Exercise 21.11) we will try to find an estimate of N. Here we look at the probability distribution of X.

a. What is the probability distribution of X? Specify its parameter(s)!

b. The drawback of this approach is that X can attain any of the values $1, 2, 3, \ldots$, so that if N is large we might be sampling from the box for quite a long time. We decide to sample from the box in a slightly different way: after we have randomly marked one of the bolts in the box, we select at random a bolt from the box. If this is the marked one, we stop, otherwise we randomly select a second bolt (we do *not* return the selected bolt). We stop when we select the marked bolt. Let Y be the number of times a bolt was selected.
Show that $\mathrm{P}(Y = k) = 1/N$ for $k = 1, 2, \ldots, N$ (Y has a so-called *discrete uniform* distribution).

c. Instead of randomly marking one bolt in the box, we mark m bolts, with m smaller than N. Next, we randomly select r bolts; Z is the number of marked bolts in the sample.
Show that

$$\mathrm{P}(Z = k) = \frac{\binom{m}{k}\binom{N-m}{r-k}}{\binom{N}{r}}, \qquad \text{for} \quad k = 0, 1, 2, \ldots, r.$$

(Z has a so-called *hypergeometric* distribution, with parameters m, N, and r.)

4.14 We throw a coin until a head turns up for the second time, where p is the probability that a throw results in a head and we assume that the outcome of each throw is independent of the previous outcomes. Let X be the number of times we have thrown the coin.

a. Determine $P(X = 2)$, $P(X = 3)$, and $P(X = 4)$.

b. Show that $P(X = n) = (n - 1)p^2(1 - p)^{n-2}$ for $n \geq 2$.