

Random Variables



2.1 Random Variables

It frequently occurs that in performing an experiment we are mainly interested in some functions of the outcome as opposed to the outcome itself. For instance, in tossing dice we are often interested in the sum of the two dice and are not really concerned about the actual outcome. That is, we may be interested in knowing that the sum is seven and not be concerned over whether the actual outcome was (1, 6) or (2, 5) or (3, 4) or (4, 3) or (5, 2) or (6, 1). These quantities of interest, or more formally, these real-valued functions defined on the sample space, are known as *random variables*.

Since the value of a random variable is determined by the outcome of the experiment, we may assign probabilities to the possible values of the random variable.

Example 2.1 Letting X denote the random variable that is defined as the sum of two fair dice; then

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

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

$$P\{X = 4\} = P\{(1, 3), (2, 2), (3, 1)\} = \frac{3}{36},$$

$$P\{X = 5\} = P\{(1, 4), (2, 3), (3, 2), (4, 1)\} = \frac{4}{36},$$

$$P\{X = 6\} = P\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\} = \frac{5}{36},$$

$$\begin{aligned}
P\{X = 7\} &= P\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} = \frac{6}{36}, \\
P\{X = 8\} &= P\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\} = \frac{5}{36}, \\
P\{X = 9\} &= P\{(3, 6), (4, 5), (5, 4), (6, 3)\} = \frac{4}{36}, \\
P\{X = 10\} &= P\{(4, 6), (5, 5), (6, 4)\} = \frac{3}{36}, \\
P\{X = 11\} &= P\{(5, 6), (6, 5)\} = \frac{2}{36}, \\
P\{X = 12\} &= P\{(6, 6)\} = \frac{1}{36}
\end{aligned} \tag{2.1}$$

In other words, the random variable X can take on any integral value between two and twelve, and the probability that it takes on each value is given by Equation (2.1). Since X must take on one of the values two through twelve, we must have

$$1 = P\left\{\bigcup_{i=2}^{12} \{X = n\}\right\} = \sum_{n=2}^{12} P\{X = n\}$$

which may be checked from Equation (2.1). ■

Example 2.2 For a second example, suppose that our experiment consists of tossing two fair coins. Letting Y denote the number of heads appearing, then Y is a random variable taking on one of the values 0, 1, 2 with respective probabilities

$$\begin{aligned}
P\{Y = 0\} &= P\{(T, T)\} = \frac{1}{4}, \\
P\{Y = 1\} &= P\{(T, H), (H, T)\} = \frac{2}{4}, \\
P\{Y = 2\} &= P\{(H, H)\} = \frac{1}{4}
\end{aligned}$$

Of course, $P\{Y = 0\} + P\{Y = 1\} + P\{Y = 2\} = 1$. ■

Example 2.3 Suppose that we toss a coin having a probability p of coming up heads, until the first head appears. Letting N denote the number of flips required, then assuming that the outcome of successive flips are independent, N is a random variable taking on one of the values $1, 2, 3, \dots$, with respective probabilities

$$\begin{aligned}
P\{N = 1\} &= P\{H\} = p, \\
P\{N = 2\} &= P\{(T, H)\} = (1 - p)p, \\
P\{N = 3\} &= P\{(T, T, H)\} = (1 - p)^2 p, \\
&\vdots \\
P\{N = n\} &= P\{(\underbrace{T, T, \dots, T}_{n-1}, H)\} = (1 - p)^{n-1} p, \quad n \geq 1
\end{aligned}$$

As a check, note that

$$\begin{aligned}
 P\left(\bigcup_{n=1}^{\infty}\{N = n\}\right) &= \sum_{n=1}^{\infty} P\{N = n\} \\
 &= p \sum_{n=1}^{\infty} (1-p)^{n-1} \\
 &= \frac{p}{1 - (1-p)} \\
 &= 1
 \end{aligned}$$

■

Example 2.4 Suppose that our experiment consists of seeing how long a battery can operate before wearing down. Suppose also that we are not primarily interested in the actual lifetime of the battery but are concerned only about whether or not the battery lasts at least two years. In this case, we may define the random variable I by

$$I = \begin{cases} 1, & \text{if the lifetime of battery is two or more years} \\ 0, & \text{otherwise} \end{cases}$$

If E denotes the event that the battery lasts two or more years, then the random variable I is known as the *indicator* random variable for event E . (Note that I equals 1 or 0 depending on whether or not E occurs.) ■

Example 2.5 Suppose that independent trials, each of which results in any of m possible outcomes with respective probabilities p_1, \dots, p_m , $\sum_{i=1}^m p_i = 1$, are continually performed. Let X denote the number of trials needed until each outcome has occurred at least once.

Rather than directly considering $P\{X = n\}$ we will first determine $P\{X > n\}$, the probability that at least one of the outcomes has not yet occurred after n trials. Letting A_i denote the event that outcome i has not yet occurred after the first n trials, $i = 1, \dots, m$, then

$$\begin{aligned}
 P\{X > n\} &= P\left(\bigcup_{i=1}^m A_i\right) \\
 &= \sum_{i=1}^m P(A_i) - \sum_{i < j} P(A_i A_j) \\
 &\quad + \sum_{i < j < k} P(A_i A_j A_k) - \dots + (-1)^{m+1} P(A_1 \dots A_m)
 \end{aligned}$$

Now, $P(A_i)$ is the probability that each of the first n trials results in a non- i outcome, and so by independence

$$P(A_i) = (1 - p_i)^n$$

Similarly, $P(A_i A_j)$ is the probability that the first n trials all result in a non- i and non- j outcome, and so

$$P(A_i A_j) = (1 - p_i - p_j)^n$$

As all of the other probabilities are similar, we see that

$$\begin{aligned} P\{X > n\} &= \sum_{i=1}^m (1 - p_i)^n - \sum_{i < j} (1 - p_i - p_j)^n \\ &\quad + \sum_{i < j < k} (1 - p_i - p_j - p_k)^n - \dots \end{aligned}$$

Since $P\{X = n\} = P\{X > n - 1\} - P\{X > n\}$, we see, upon using the algebraic identity $(1 - a)^{n-1} - (1 - a)^n = a(1 - a)^{n-1}$, that

$$\begin{aligned} P\{X = n\} &= \sum_{i=1}^m p_i (1 - p_i)^{n-1} - \sum_{i < j} (p_i + p_j) (1 - p_i - p_j)^{n-1} \\ &\quad + \sum_{i < j < k} (p_i + p_j + p_k) (1 - p_i - p_j - p_k)^{n-1} - \dots \quad \blacksquare \end{aligned}$$

In all of the preceding examples, the random variables of interest took on either a finite or a countable number of possible values.* Such random variables are called *discrete*. However, there also exist random variables that take on a continuum of possible values. These are known as *continuous* random variables. One example is the random variable denoting the lifetime of a car, when the car's lifetime is assumed to take on any value in some interval (a, b) .

The *cumulative distribution function* (cdf) (or more simply the *distribution function*) $F(\cdot)$ of the random variable X is defined for any real number b , $-\infty < b < \infty$, by

$$F(b) = P\{X \leq b\}$$

In words, $F(b)$ denotes the probability that the random variable X takes on a value that is less than or equal to b . Some properties of the cdf F are

* A set is countable if its elements can be put in a one-to-one correspondence with the sequence of positive integers.

- (i) $F(b)$ is a nondecreasing function of b ,
- (ii) $\lim_{b \rightarrow \infty} F(b) = F(\infty) = 1$,
- (iii) $\lim_{b \rightarrow -\infty} F(b) = F(-\infty) = 0$.

Property (i) follows since for $a < b$ the event $\{X \leq a\}$ is contained in the event $\{X \leq b\}$, and so it must have a smaller probability. Properties (ii) and (iii) follow since X must take on some finite value.

All probability questions about X can be answered in terms of the cdf $F(\cdot)$. For example,

$$P\{a < X \leq b\} = F(b) - F(a) \quad \text{for all } a < b$$

This follows since we may calculate $P\{a < X \leq b\}$ by first computing the probability that $X \leq b$ (that is, $F(b)$) and then subtracting from this the probability that $X \leq a$ (that is, $F(a)$).

If we desire the probability that X is strictly smaller than b , we may calculate this probability by

$$\begin{aligned} P\{X < b\} &= \lim_{h \rightarrow 0^+} P\{X \leq b - h\} \\ &= \lim_{h \rightarrow 0^+} F(b - h) \end{aligned}$$

where $\lim_{h \rightarrow 0^+}$ means that we are taking the limit as h decreases to 0. Note that $P\{X < b\}$ does not necessarily equal $F(b)$ since $F(b)$ also includes the probability that X equals b .

2.2 Discrete Random Variables

As was previously mentioned, a random variable that can take on at most a countable number of possible values is said to be *discrete*. For a discrete random variable X , we define the *probability mass function* $p(a)$ of X by

$$p(a) = P\{X = a\}$$

The probability mass function $p(a)$ is positive for at most a countable number of values of a . That is, if X must assume one of the values x_1, x_2, \dots , then

$$\begin{aligned} p(x_i) &> 0, & i = 1, 2, \dots \\ p(x) &= 0, & \text{all other values of } x \end{aligned}$$

Since X must take on one of the values x_i , we have

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

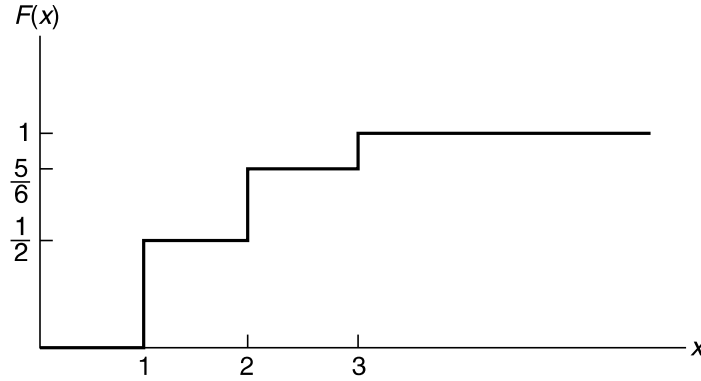


Figure 2.1 Graph of $F(x)$.

The cumulative distribution function F can be expressed in terms of $p(a)$ by

$$F(a) = \sum_{\text{all } x_i \leq a} p(x_i)$$

For instance, suppose X has a probability mass function given by

$$p(1) = \frac{1}{2}, \quad p(2) = \frac{1}{3}, \quad p(3) = \frac{1}{6}$$

then, the cumulative distribution function F of X is given by

$$F(a) = \begin{cases} 0, & a < 1 \\ \frac{1}{2}, & 1 \leq a < 2 \\ \frac{5}{6}, & 2 \leq a < 3 \\ 1, & 3 \leq a \end{cases}$$

This is graphically presented in Figure 2.1.

Discrete random variables are often classified according to their probability mass functions. We now consider some of these random variables.

2.2.1 The Bernoulli Random Variable

Suppose that a trial, or an experiment, whose outcome can be classified as either a “success” or as a “failure” is performed. If we let X equal 1 if the outcome is a success and 0 if it is a failure, then the probability mass function of X is given by

$$\begin{aligned} p(0) &= P\{X = 0\} = 1 - p, \\ p(1) &= P\{X = 1\} = p \end{aligned} \tag{2.2}$$

where p , $0 \leq p \leq 1$, is the probability that the trial is a “success.”

A random variable X is said to be a *Bernoulli* random variable if its probability mass function is given by Equation (2.2) for some $p \in (0, 1)$.

2.2.2 The Binomial Random Variable

Suppose that n independent trials, each of which results in a “success” with probability p and in a “failure” with probability $1 - p$, are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a *binomial* random variable with parameters (n, p) .

The probability mass function of a binomial random variable having parameters (n, p) is given by

$$p(i) = \binom{n}{i} p^i (1 - p)^{n-i}, \quad i = 0, 1, \dots, n \quad (2.3)$$

where

$$\binom{n}{i} = \frac{n!}{(n-i)!i!}$$

equals the number of different groups of i objects that can be chosen from a set of n objects. The validity of Equation (2.3) may be verified by first noting that the probability of any particular sequence of the n outcomes containing i successes and $n - i$ failures is, by the assumed independence of trials, $p^i(1 - p)^{n-i}$. Equation (2.3) then follows since there are $\binom{n}{i}$ different sequences of the n outcomes leading to i successes and $n - i$ failures. For instance, if $n = 3, i = 2$, then there are $\binom{3}{2} = 3$ ways in which the three trials can result in two successes. Namely, any one of the three outcomes (s, s, f) , (s, f, s) , (f, s, s) , where the outcome (s, s, f) means that the first two trials are successes and the third a failure. Since each of the three outcomes (s, s, f) , (s, f, s) , (f, s, s) has a probability $p^2(1 - p)$ of occurring the desired probability is thus $\binom{3}{2}p^2(1 - p)$.

Note that, by the binomial theorem, the probabilities sum to one, that is,

$$\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^n \binom{n}{i} p^i (1 - p)^{n-i} = (p + (1 - p))^n = 1$$

Example 2.6 Four fair coins are flipped. If the outcomes are assumed independent, what is the probability that two heads and two tails are obtained?

Solution: Letting X equal the number of heads (“successes”) that appear, then X is a binomial random variable with parameters $(n = 4, p = \frac{1}{2})$. Hence, by Equation (2.3),

$$P\{X = 2\} = \binom{4}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = \frac{3}{8}$$

■

Example 2.7 It is known that any item produced by a certain machine will be defective with probability 0.1, independently of any other item. What is the probability that in a sample of three items, at most one will be defective?

Solution: If X is the number of defective items in the sample, then X is a binomial random variable with parameters $(3, 0.1)$. Hence, the desired probability is given by

$$P\{X = 0\} + P\{X = 1\} = \binom{3}{0}(0.1)^0(0.9)^3 + \binom{3}{1}(0.1)^1(0.9)^2 = 0.972 \quad \blacksquare$$

Example 2.8 Suppose that an airplane engine will fail, when in flight, with probability $1 - p$ independently from engine to engine; suppose that the airplane will make a successful flight if at least 50 percent of its engines remain operative. For what values of p is a four-engine plane preferable to a two-engine plane?

Solution: Because each engine is assumed to fail or function independently of what happens with the other engines, it follows that the number of engines remaining operative is a binomial random variable. Hence, the probability that a four-engine plane makes a successful flight is

$$\begin{aligned} & \binom{4}{2}p^2(1-p)^2 + \binom{4}{3}p^3(1-p) + \binom{4}{4}p^4(1-p)^0 \\ &= 6p^2(1-p)^2 + 4p^3(1-p) + p^4 \end{aligned}$$

whereas the corresponding probability for a two-engine plane is

$$\binom{2}{1}p(1-p) + \binom{2}{2}p^2 = 2p(1-p) + p^2$$

Hence the four-engine plane is safer if

$$6p^2(1-p)^2 + 4p^3(1-p) + p^4 \geq 2p(1-p) + p^2$$

or equivalently if

$$6p(1-p)^2 + 4p^2(1-p) + p^3 \geq 2 - p$$

which simplifies to

$$3p^3 - 8p^2 + 7p - 2 \geq 0 \quad \text{or} \quad (p-1)^2(3p-2) \geq 0$$

which is equivalent to

$$3p - 2 \geq 0 \quad \text{or} \quad p \geq \frac{2}{3}$$

Hence, the four-engine plane is safer when the engine success probability is at least as large as $\frac{2}{3}$, whereas the two-engine plane is safer when this probability falls below $\frac{2}{3}$. ■

Example 2.9 Suppose that a particular trait of a person (such as eye color or left handedness) is classified on the basis of one pair of genes and suppose that d represents a dominant gene and r a recessive gene. Thus a person with dd genes is pure dominance, one with rr is pure recessive, and one with rd is hybrid. The pure dominance and the hybrid are alike in appearance. Children receive one gene from each parent. If, with respect to a particular trait, two hybrid parents have a total of four children, what is the probability that exactly three of the four children have the outward appearance of the dominant gene?

Solution: If we assume that each child is equally likely to inherit either of two genes from each parent, the probabilities that the child of two hybrid parents will have dd , rr , or rd pairs of genes are, respectively, $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{2}$. Hence, because an offspring will have the outward appearance of the dominant gene if its gene pair is either dd or rd , it follows that the number of such children is binomially distributed with parameters $(4, \frac{3}{4})$. Thus the desired probability is

$$\binom{4}{3} \left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right)^1 = \frac{27}{64}$$

Remark on Terminology If X is a binomial random variable with parameters (n, p) , then we say that X has a binomial distribution with parameters (n, p) .

2.2.3 The Geometric Random Variable

Suppose that independent trials, each having probability p of being a success, are performed until a success occurs. If we let X be the number of trials required until the first success, then X is said to be a *geometric* random variable with parameter p . Its probability mass function is given by

$$p(n) = P\{X = n\} = (1 - p)^{n-1}p, \quad n = 1, 2, \dots \quad (2.4)$$

Equation (2.4) follows since in order for X to equal n it is necessary and sufficient that the first $n - 1$ trials be failures and the n th trial a success. Equation (2.4) follows since the outcomes of the successive trials are assumed to be independent.

To check that $p(n)$ is a probability mass function, we note that

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

2.2.4 The Poisson Random Variable

A random variable X , taking on one of the values $0, 1, 2, \dots$, is said to be a *Poisson* random variable with parameter λ , if for some $\lambda > 0$,

$$p(i) = P\{X = i\} = e^{-\lambda} \frac{\lambda^i}{i!}, \quad i = 0, 1, \dots \quad (2.5)$$

Equation (2.5) defines a probability mass function since

$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1$$

The Poisson random variable has a wide range of applications in a diverse number of areas, as will be seen in Chapter 5.

An important property of the Poisson random variable is that it may be used to approximate a binomial random variable when the binomial parameter n is large and p is small. To see this, suppose that X is a binomial random variable with parameters (n, p) , and let $\lambda = np$. Then

$$\begin{aligned} P\{X = i\} &= \frac{n!}{(n-i)! i!} p^i (1-p)^{n-i} \\ &= \frac{n!}{(n-i)! i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1) \cdots (n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i} \end{aligned}$$

Now, for n large and p small

$$\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \quad \frac{n(n-1) \cdots (n-i+1)}{n^i} \approx 1, \quad \left(1 - \frac{\lambda}{n}\right)^i \approx 1$$

Hence, for n large and p small,

$$P\{X = i\} \approx e^{-\lambda} \frac{\lambda^i}{i!}$$

Example 2.10 Suppose that the number of typographical errors on a single page of this book has a Poisson distribution with parameter $\lambda = 1$. Calculate the probability that there is at least one error on this page.

Solution:

$$P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - e^{-1} \approx 0.633$$

■

Example 2.11 If the number of accidents occurring on a highway each day is a Poisson random variable with parameter $\lambda = 3$, what is the probability that no accidents occur today?

Solution:

$$P\{X = 0\} = e^{-3} \approx 0.05 \quad \blacksquare$$

Example 2.12 Consider an experiment that consists of counting the number of α -particles given off in a one-second interval by one gram of radioactive material. If we know from past experience that, on the average, 3.2 such α -particles are given off, what is a good approximation to the probability that no more than two α -particles will appear?

Solution: If we think of the gram of radioactive material as consisting of a large number n of atoms each of which has probability $3.2/n$ of disintegrating and sending off an α -particle during the second considered, then we see that, to a very close approximation, the number of α -particles given off will be a Poisson random variable with parameter $\lambda = 3.2$. Hence the desired probability is

$$P\{X \leq 2\} = e^{-3.2} + 3.2e^{-3.2} + \frac{(3.2)^2}{2}e^{-3.2} \approx 0.382 \quad \blacksquare$$

2.3 Continuous Random Variables

In this section, we shall concern ourselves with random variables whose set of possible values is uncountable. Let X be such a random variable. We say that X is a *continuous* random variable if there exists a nonnegative function $f(x)$, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

$$P\{X \in B\} = \int_B f(x) dx \quad (2.6)$$

The function $f(x)$ is called the *probability density function* of the random variable X .

In words, Equation (2.6) states that the probability that X will be in B may be obtained by integrating the probability density function over the set B . Since X must assume some value, $f(x)$ must satisfy

$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x) dx$$