

### 3.3: Bernoulli and Binomial Distributions

In this section and the next two, we introduce *families* of common discrete probability distributions, i.e., probability distributions for discrete random variables. We refer to these as "families" of distributions because in each case we will define a probability mass function by specifying an explicit formula, and that formula will incorporate a constant (or set of constants) that are referred to as **parameters**. By specifying values for the parameter(s) in the pmf, we define a specific probability distribution for a specific random variable. For each family of distributions introduced, we will list a set of defining characteristics that will help determine when to use a certain distribution in a given context.

#### Bernoulli Distribution

Consider the following example.

##### Example 3.3.1

Let  $A$  be an event in a sample space  $S$ . Suppose we are only interested in whether or not the outcome of the underlying probability experiment is in the specified event  $A$ . To track this we can define an **indicator random variable**, denoted  $I_A$ , given by

$$I_A(s) = \begin{cases} 1, & \text{if } s \in A, \\ 0, & \text{if } s \in A^c. \end{cases}$$

In other words, the random variable  $I_A$  will equal 1 if the resulting outcome is in event  $A$ , and  $I_A$  equals 0 if the outcome is not in  $A$ . Thus,  $I_A$  is a discrete random variable. We can state the probability mass function of  $I_A$  in terms of the probability that the resulting outcome is in event  $A$ , i.e., the probability that event  $A$  occurs,  $P(A)$ :

$$\begin{aligned} p(0) &= P(I_A = 0) = P(A^c) = 1 - P(A) \\ p(1) &= P(I_A = 1) = P(A) \end{aligned}$$

In Example 3.3.1, the random variable  $I_A$  is a *Bernoulli random variable* because its pmf has the form of the *Bernoulli probability distribution*, which we define next.

##### Definition 3.3.1

A random variable  $X$  has a **Bernoulli distribution** with parameter  $p$ , where  $0 \leq p \leq 1$ , if it has only two possible values, typically denoted 0 and 1. The probability mass function (pmf) of  $X$  is given by

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

The cumulative distribution function (cdf) of  $X$  is given by

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - p, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases} \quad (3.3.1)$$

In Definition 3.3.1, note that the defining characteristic of the Bernoulli distribution is that it models random variables that have only two possible values. As noted in the definition, the two possible values of a Bernoulli random variable are usually 0 and 1. In the typical application of the Bernoulli distribution, a value of 1 indicates a "success" and a value of 0 indicates a "failure", where "success" refers that the event or outcome of interest. The parameter  $p$  in the Bernoulli distribution is given by the probability of a "success". In Example 3.3.1, we were interested in tracking whether or not event  $A$  occurred, and so that is what a "success" would be, which occurs with probability given by the probability of  $A$ . Thus, the value of the parameter  $p$  for the Bernoulli distribution in Example 3.3.1 is given by  $p = P(A)$ .

##### Exercise 3.3.1

Derive the general formula for the cdf of the Bernoulli distribution given in Equation 3.3.1.

##### Hint

First find  $F(0)$  and  $F(1)$ .

### Answer

Recall that the only two values of a Bernoulli random variable  $X$  are 0 and 1. So, first, we find the cdf at those two values:

$$\begin{aligned} F(0) &= P(X \leq 0) = P(X = 0) = p(0) = 1 - p \\ F(1) &= P(X \leq 1) = P(X = 0 \text{ or } 1) = p(0) + p(1) = (1 - p) + p = 1 \end{aligned}$$

Now for the other values, a Bernoulli random variable will never be negative, so  $F(x) = 0$ , for  $x < 0$ . Also, a Bernoulli random variable will always be less than or equal to 1, so  $F(x) = 1$ , for  $x \geq 1$ . Lastly, if  $x$  is in between 0 and 1, then the cdf is given by

$$F(x) = P(X \leq x) = P(X = 0) = p(0) = 1 - p, \text{ for } 0 \leq x < 1.$$

## Binomial Distribution

To introduce the next family of distributions, we use our continuing example of tossing a coin, adding another toss.

### Example 3.3.2

Suppose we toss a coin three times and record the sequence of heads ( $h$ ) and tails ( $t$ ). Supposing that the coin is fair, each toss results in heads with probability 0.5, and tails with the same probability of 0.5. Since the three tosses are mutually independent, the probability assigned to any outcome is  $0.5^3$ . More specifically, consider the outcome  $hth$ . We could write the probability of this outcome as  $(0.5)^2(0.5)^1$  to emphasize the fact that two heads and one tails occurred. Note that there are two other outcomes with two heads and one tails:  $hht$  and  $thh$ . Recall from [Example 2.1.2](#) in [Section 2.1](#), that we can count the number of outcomes with two heads and one tails by counting the number of ways to select positions for the two heads to occur in a sequence of three tosses, which is given by  $\binom{3}{2}$ . In general, note that  $\binom{3}{x}$  counts the number of possible sequences with exactly  $x$  heads, for  $x = 0, 1, 2, 3$ .

We generalize the above by defining the discrete random variable  $X$  to be the number of heads in an outcome. The possible values of  $X$  are  $x = 0, 1, 2, 3$ . Using the above facts, the pmf of  $X$  is given as follows:

$$\begin{aligned} p(0) &= P(X = 0) = P(\{ttt\}) = \frac{1}{8} = \binom{3}{0}(0.5)^0(0.5)^3 \\ p(1) &= P(X = 1) = P(\{htt, tht, tth\}) = \frac{3}{8} = \binom{3}{1}(0.5)^1(0.5)^2 \\ p(2) &= P(X = 2) = P(\{hht, hth, thh\}) = \frac{3}{8} = \binom{3}{2}(0.5)^2(0.5)^1 \\ p(3) &= P(X = 3) = P(\{hhh\}) = \frac{1}{8} = \binom{3}{3}(0.5)^3(0.5)^0 \end{aligned} \tag{3.3.2}$$

In the above, the fractions in orange are found by calculating the probabilities directly using equally likely outcomes (note that the sample space  $S$  has 8 outcomes, see [Example 2.1.1](#)). In each line, the value of  $x$  is highlighted in red so that we can see the pattern forming. For example, when  $x = 2$ , we see in the expression on the right-hand side of Equation 3.3.2 that "2" appears in the binomial coefficient  $\binom{3}{2}$ , which gives the number of outcomes resulting in the random variable equaling 2, and "2" also appears in the exponent on the first 0.5, which gives the probability of two heads occurring.

The pattern exhibited by the random variable  $X$  in Example 3.3.2 is referred to as the *binomial distribution*, which we formalize in the next definition.

### Definition 3.3.2

Suppose that  $n$  independent trials of the same probability experiment are performed, where each trial results in either a "success" (with probability  $p$ ), or a "failure" (with probability  $1 - p$ ). If the random variable  $X$  denotes the total number of successes in the  $n$  trials, then  $X$  has a **binomial distribution** with parameters  $n$  and  $p$ , which we write  $X \sim \text{binomial}(n, p)$ . The probability mass function of  $X$  is given by

$$p(x) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad \text{for } x = 0, 1, \dots, n. \tag{3.3.3}$$

In Example 3.3.2, the independent trials are the three tosses of the coin, so in this case we have parameter  $n = 3$ . Furthermore, we were interested in counting the number of heads occurring in the three tosses, so a "success" is getting a heads on a toss, which occurs with probability 0.5 and so parameter  $p = 0.5$ . Thus, the random variable  $X$  in this example has a binomial(3, 0.5) distribution and applying the formula for the binomial pmf given in Equation 3.3.3 when  $x = 2$  we get the same expression on the right-hand side of Equation 3.3.2:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \Rightarrow p(2) = \binom{3}{2} 0.5^2 (1-0.5)^{3-2} = \binom{3}{2} 0.5^2 0.5^1$$

In general, we can connect binomial random variables to Bernoulli random variables. If  $X$  is a binomial random variable, with parameters  $n$  and  $p$ , then it can be written as the sum of  $n$  independent *Bernoulli* random variables,  $X_1, \dots, X_n$ . (Note: We will formally define *independence* for random variables later, in Chapter 5.) Specifically, if we define the random variable  $X_i$ , for  $i = 1, \dots, n$ , to be 1 when the  $i^{th}$  trial is a "success", and 0 when it is a "failure", then the sum

$$X = X_1 + X_2 + \dots + X_n$$

gives the total number of success in  $n$  trials. This connection between the binomial and Bernoulli distribution will be useful in a later section.

One of the main applications of the binomial distribution is to model population characteristics as in the following example.

### Example 3.3.3

Consider a group of 100 voters. If  $p$  denotes the probability that a voter will vote for a specific candidate, and we let random variable  $X$  denote the number of voters in the group that will vote for that candidate, then  $X$  follows a binomial distribution with parameters  $n = 100$  and  $p$ .

---

This page titled 3.3: Bernoulli and Binomial Distributions is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Kristin Kuter](#).