

## 3.1: Discrete Distributions

### Basic Theory

#### Definitions and Basic Properties

As usual, our starting point is a random experiment modeled by a probability space  $(S, \mathcal{S}, \mathbb{P})$ . So to review,  $S$  is the set of outcomes,  $\mathcal{S}$  the collection of events, and  $\mathbb{P}$  the probability measure on the sample space  $(S, \mathcal{S})$ . We use the terms *probability measure* and *probability distribution* synonymously in this text. Also, since we use a general definition of random variable, every probability measure can be thought of as the probability distribution of a random variable, so we can always take this point of view if we like. Indeed, most probability measures naturally have random variables associated with them.

Recall that the sample space  $(S, \mathcal{S})$  is *discrete* if  $S$  is countable and  $\mathcal{S} = \mathcal{P}(S)$  is the collection of all subsets of  $S$ . In this case,  $\mathbb{P}$  is a *discrete distribution* and  $(S, \mathcal{S}, \mathbb{P})$  is a *discrete probability space*.

For the remainder of our discussion we assume that  $(S, \mathcal{S}, \mathbb{P})$  is a discrete probability space. In the picture below, the blue dots are intended to represent points of positive probability.

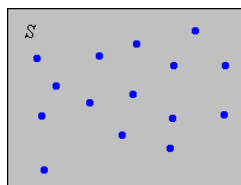


Figure 3.1.1: A discrete distribution

It's very simple to describe a discrete probability distribution with the function that assigns probabilities to the individual points in  $S$ .

The function  $f$  on  $S$  defined by  $f(x) = \mathbb{P}(\{x\})$  for  $x \in S$  is the *probability density function* of  $\mathbb{P}$ , and satisfies the following properties:

1.  $f(x) \geq 0$ ,  $x \in S$
2.  $\sum_{x \in S} f(x) = 1$
3.  $\sum_{x \in A} f(x) = \mathbb{P}(A)$  for  $A \subseteq S$

Proof

These properties follow from the axioms of a probability measure.

1.  $f(x) = \mathbb{P}(\{x\}) \geq 0$  since probabilities are nonnegative.
2.  $\sum_{x \in S} f(x) = \sum_{x \in S} \mathbb{P}(\{x\}) = \mathbb{P}(S) = 1$  by the countable additivity axiom.
3.  $\sum_{x \in A} f(x) = \sum_{x \in A} \mathbb{P}(\{x\}) = \mathbb{P}(A)$  for  $A \subseteq S$  again, by the countable additivity axiom.

Property (c) is particularly important since it shows that a discrete probability distribution is completely determined by its probability density function. Conversely, any function that satisfies properties (a) and (b) can be used to construct a discrete probability distribution on  $S$  via property (c).

A nonnegative function  $f$  on  $S$  that satisfies  $\sum_{x \in S} f(x) = 1$  is a (discrete) *probability density function* on  $S$ , and then  $\mathbb{P}$  defined as follows is a probability measure on  $S$ .

$$\mathbb{P}(A) = \sum_{x \in A} f(x), \quad A \subseteq S \quad (3.1.1)$$

Proof

1.  $\mathbb{P}(A) = \sum_{x \in A} f(x) \geq 0$  since  $f$  is nonnegative.
2.  $\mathbb{P}(S) = \sum_{x \in S} f(x) = 1$  by property (b)
3. Suppose that  $\{A_i : i \in I\}$  is a countable, disjoint collection of subsets of  $S$ , and let  $A = \bigcup_{i \in I} A_i$ . Then

$$\mathbb{P}(A) = \sum_{x \in A} f(x) = \sum_{i \in I} \sum_{x \in A_i} f(x) = \sum_{i \in I} \mathbb{P}(A_i) \quad (3.1.2)$$

Note that since  $f$  is nonnegative, the order of the terms in the sum do not matter.

Technically,  $f$  is the density of  $\mathbb{P}$  relative to counting measure  $\#$  on  $S$ . The technicalities are discussed in detail in the advanced section on absolute continuity and density functions.

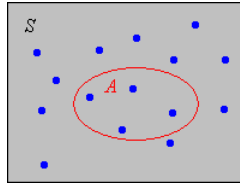


Figure 3.1.2: A discrete distribution is completely determined by its probability density function.

The set of outcomes  $S$  is often a countable subset of some larger set, such as  $\mathbb{R}^n$  for some  $n \in \mathbb{N}_+$ . But not always. We might want to consider a random variable with values in a deck of cards, or a set of words, or some other discrete population of objects. Of course, we can always map a countable set  $S$  one-to-one into a Euclidean set, but it might be contrived or unnatural to do so. In any event, if  $S$  is a subset of a larger set, we can always extend a probability density function  $f$ , if we want, to the larger set by defining  $f(x) = 0$  for  $x \notin S$ . Sometimes this extension simplifies formulas and notation. Put another way, the “set of values” is often a convenience set that includes the points with positive probability, but perhaps other points as well.

Suppose that  $f$  is a probability density function on  $S$ . Then  $\{x \in S : f(x) > 0\}$  is the *support set* of the distribution.

Values of  $x$  that maximize the probability density function are important enough to deserve a name.

Suppose again that  $f$  is a probability density function on  $S$ . An element  $x \in S$  that maximizes  $f$  is a *mode* of the distribution.

When there is only one mode, it is sometimes used as a measure of the *center* of the distribution.

A discrete probability distribution defined by a probability density function  $f$  is equivalent to a *discrete mass distribution*, with total mass 1. In this analogy,  $S$  is the (countable) set of point masses, and  $f(x)$  is the mass of the point at  $x \in S$ . Property (c) in (2) above simply means that the mass of a set  $A$  can be found by adding the masses of the points in  $A$ .

But let's consider a probabilistic interpretation, rather than one from physics. We start with a basic random variable  $X$  for an experiment, defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Suppose that  $X$  has a discrete distribution on  $S$  with probability density function  $f$ . So in this setting,  $f(x) = \mathbb{P}(X = x)$  for  $x \in S$ . We create a new, compound experiment by conducting independent repetitions of the original experiment. So in the compound experiment, we have a sequence of independent random variables  $(X_1, X_2, \dots)$  each with the same distribution as  $X$ ; in statistical terms, we are *sampling* from the distribution of  $X$ . Define

$$f_n(x) = \frac{1}{n} \# \{i \in \{1, 2, \dots, n\} : X_i = x\} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i = x), \quad x \in S \quad (3.1.3)$$

Note that  $f_n(x)$  is the *relative frequency* of outcome  $x \in S$  in the first  $n$  runs. Note also that  $f_n(x)$  is a random variable for the compound experiment for each  $x \in S$ . By the law of large numbers,  $f_n(x)$  should converge to  $f(x)$ , in some sense, as  $n \rightarrow \infty$ . The function  $f_n$  is called the *empirical probability density function*, and it is in fact a (random) probability density function, since it satisfies properties (a) and (b) of (2). Empirical probability density functions are displayed in most of the simulation apps that deal with discrete variables.

It's easy to construct discrete probability density functions from other nonnegative functions defined on a countable set.

Suppose that  $g$  is a nonnegative function defined on  $S$ , and let

$$c = \sum_{x \in S} g(x) \quad (3.1.4)$$

If  $0 < c < \infty$ , then the function  $f$  defined by  $f(x) = \frac{1}{c} g(x)$  for  $x \in S$  is a discrete probability density function on  $S$ .

Proof

Clearly  $f(x) \geq 0$  for  $x \in S$ . also

$$\sum_{x \in S} f(x) = \frac{1}{c} \sum_{x \in S} g(x) = \frac{c}{c} = 1 \quad (3.1.5)$$

Note that since we are assuming that  $g$  is nonnegative,  $c = 0$  if and only if  $g(x) = 0$  for every  $x \in S$ . At the other extreme,  $c = \infty$  could only occur if  $S$  is infinite (and the infinite series diverges). When  $0 < c < \infty$  (so that we can construct the probability density function  $f$ ),  $c$  is sometimes called the *normalizing constant*. This result is useful for constructing probability density functions with desired functional properties (domain, shape, symmetry, and so on).

### Conditional Densities

Suppose again that  $X$  is a random variable on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and that  $X$  takes values in our discrete set  $S$ . The distribution of  $X$  (and hence the probability density function of  $X$ ) is based on the underlying probability measure on the sample space  $(\Omega, \mathcal{F})$ . This measure could be a conditional probability measure, conditioned on a given event  $E \in \mathcal{F}$  (with  $\mathbb{P}(E) > 0$ ). The probability density function in this case is

$$f(x | E) = \mathbb{P}(X = x | E), \quad x \in S \quad (3.1.6)$$

Except for notation, no new concepts are involved. Therefore, all results that hold for discrete probability density functions in general have analogies for conditional discrete probability density functions.

For fixed  $E \in \mathcal{F}$  with  $\mathbb{P}(E) > 0$  the function  $x \mapsto f(x | E)$  is a discrete probability density function on  $S$ . That is,

1.  $f(x | E) \geq 0$  for  $x \in S$ .
2.  $\sum_{x \in S} f(x | E) = 1$
3.  $\sum_{x \in A} f(x | E) = \mathbb{P}(X \in A | E)$  for  $A \subseteq S$

Proof

This is a consequence of the fact that  $A \mapsto \mathbb{P}(A | E)$  is a probability measure on  $(\Omega, \mathcal{F})$ . The function  $x \mapsto f(x | E)$  plays the same role for the conditional probability measure that  $f$  does for the original probability measure  $\mathbb{P}$ .

In particular, the event  $E$  could be an event defined in terms of the random variable  $X$  itself.

Suppose that  $B \subseteq S$  and  $\mathbb{P}(X \in B) > 0$ . The conditional probability density function of  $X$  given  $X \in B$  is the function on  $B$  defined by

$$f(x | X \in B) = \frac{f(x)}{\mathbb{P}(X \in B)} = \frac{f(x)}{\sum_{y \in B} f(y)}, \quad x \in B \quad (3.1.7)$$

Proof

This follows from the previous theorem.  $f(x | X \in B) = \mathbb{P}(X = x, X \in B) / \mathbb{P}(X \in B)$ . The numerator is  $f(x)$  if  $x \in B$  and is 0 if  $x \notin B$ .

Note that the denominator is simply the [normalizing constant](#) for  $f$  restricted to  $B$ . Of course,  $f(x | B) = 0$  for  $x \in B^c$ .

### Conditioning and Bayes' Theorem

Suppose again that  $X$  is a random variable defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and that  $X$  has a discrete distribution on  $S$ , with probability density function  $f$ . We assume that  $f(x) > 0$  for  $x \in S$  so that the distribution has support  $S$ . The versions of the law of total probability and Bayes' theorem given in the following theorems follow immediately from the corresponding results in the section on Conditional Probability. Only the notation is different.

**Law of Total Probability.** If  $E \in \mathcal{F}$  is an event then

$$\mathbb{P}(E) = \sum_{x \in S} f(x) \mathbb{P}(E | X = x) \quad (3.1.8)$$

Proof

Note that  $\{\{X = x\} : x \in S\}$  is a countable partition of the sample space  $\Omega$ . That is, these events are disjoint and their union is the entire sample space  $\Omega$ . Hence

$$\mathbb{P}(E) = \sum_{x \in S} \mathbb{P}(E \cap \{X = x\}) = \sum_{x \in S} \mathbb{P}(X = x) \mathbb{P}(E | X = x) = \sum_{x \in S} f(x) \mathbb{P}(E | X = x) \quad (3.1.9)$$

This result is useful, naturally, when the distribution of  $X$  and the conditional probability of  $E$  given the values of  $X$  are known. When we compute  $\mathbb{P}(E)$  in this way, we say that we are *conditioning* on  $X$ . Note that  $\mathbb{P}(E)$ , as expressed by the formula, is a weighted average of  $\mathbb{P}(E | X = x)$ , with weight factors  $f(x)$ , over  $x \in S$ .

**Bayes' Theorem.** If  $E \in \mathcal{F}$  is an event with  $\mathbb{P}(E) > 0$  then

$$f(x | E) = \frac{f(x) \mathbb{P}(E | X = x)}{\sum_{y \in S} f(y) \mathbb{P}(E | X = y)}, \quad x \in S \quad (3.1.10)$$

**Proof**

Note that the numerator of the fraction on the right is  $\mathbb{P}(X = x) \mathbb{P}(E | X = x) = \mathbb{P}(\{X = x\} \cap E)$ . The denominator is  $\mathbb{P}(E)$  by the previous theorem. Hence the ratio is  $\mathbb{P}(X = x | E) = f(x | E)$ .

Bayes' theorem, named for Thomas Bayes, is a formula for the conditional probability density function of  $X$  given  $E$ . Again, it is useful when the quantities on the right are known. In the context of Bayes' theorem, the (unconditional) distribution of  $X$  is referred to as the *prior* distribution and the conditional distribution as the *posterior* distribution. Note that the denominator in Bayes' formula is  $\mathbb{P}(E)$  and is simply the [normalizing constant](#) for the function  $x \mapsto f(x) \mathbb{P}(E | X = x)$ .

## Examples and Special Cases

We start with some simple (albeit somewhat artificial) discrete distributions. After that, we study three special parametric models—the [discrete uniform distribution](#), [hypergeometric distributions](#), and [Bernoulli trials](#). These models are very important, so when working the computational problems that follow, try to see if the problem fits one of these models. As always, be sure to try the problems yourself before looking at the answers and proofs in the text.

### Simple Discrete Distributions

Let  $g$  be the function defined by  $g(n) = n(10 - n)$  for  $n \in \{1, 2, \dots, 9\}$ .

1. Find the probability density function  $f$  that is [proportional](#) to  $g$  as in .
2. Sketch the graph of  $f$  and find the mode of the distribution.
3. Find  $\mathbb{P}(3 \leq N \leq 6)$  where  $N$  has probability density function  $f$ .

**Answer**

1.  $f(n) = \frac{1}{165} n(10 - n)$  for  $n \in \{1, 2, \dots, 9\}$
2. mode  $n = 5$
3.  $\frac{94}{165}$

Let  $g$  be the function defined by  $g(n) = n^2(10 - n)$  for  $n \in \{1, 2, \dots, 10\}$ .

1. Find the probability density function  $f$  that is [proportional](#) to  $g$ .
2. Sketch the graph of  $f$  and find the mode of the distribution.
3. Find  $\mathbb{P}(3 \leq N \leq 6)$  where  $N$  has probability density function  $f$ .

**Answer**

1.  $f(n) = \frac{1}{825} n^2(10 - n)$  for  $n \in \{1, 2, \dots, 9\}$
2. mode  $n = 7$
3.  $\frac{428}{825}$

Let  $g$  be the function defined by  $g(x, y) = x + y$  for  $(x, y) \in \{1, 2, 3\}^2$ .

1. Sketch the domain of  $g$ .

- Find the probability density function  $f$  that is **proportional** to  $g$ .
- Find the mode of the distribution.
- Find  $\mathbb{P}(X > Y)$  where  $(X, Y)$  has probability density function  $f$ .

Answer

- $f(x, y) = \frac{1}{36}(x + y)$  for  $(x, y) \in \{1, 2, 3\}^2$
- mode  $(3, 3)$
- $\frac{2}{9}$

Let  $g$  be the function defined by  $g(x, y) = xy$  for  $(x, y) \in \{(1, 1), (1, 2), (1, 3), (2, 2), (2, 3), (3, 3)\}$

- Sketch the domain of  $g$ .
- Find the probability density function  $f$  that is **proportional** to  $g$ .
- Find the mode of the distribution.
- Find  $\mathbb{P}[(X, Y) \in \{(1, 2), (1, 3), (2, 2), (2, 3)\}]$  where  $(X, Y)$  has probability density function  $f$ .

Answer

- $f(x, y) = \frac{1}{25}xy$  for  $(x, y) \in \{(1, 1), (1, 2), (1, 3), (2, 2), (2, 3), (3, 3)\}$
- mode  $(3, 3)$
- $\frac{3}{5}$

Consider the following game: An urn initially contains one red and one green ball. A ball is selected at random, and if the ball is green, the game is over. If the ball is red, the ball is returned to the urn, another red ball is added, and the game continues. At each stage, a ball is selected at random, and if the ball is green, the game is over. If the ball is red, the ball is returned to the urn, another red ball is added, and the game continues. Let  $X$  denote the length of the game (that is, the number of selections required to obtain a green ball). Find the probability density function of  $X$ .

Solution

Note that  $X$  takes values in  $\mathbb{N}_+$ . Using the multiplication rule for conditional probabilities, the PDF  $f$  of  $X$  is given by

$$f(1) = \frac{1}{2} = \frac{1}{1 \cdot 2}, \quad f(2) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{2 \cdot 3}, \quad f(3) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{4} = \frac{1}{3 \cdot 4} \quad (3.1.11)$$

and in general,  $f(x) = \frac{1}{x(x+1)}$  for  $x \in \mathbb{N}_+$ . By partial fractions,  $f(x) = \frac{1}{x} - \frac{1}{x+1}$  for  $x \in \mathbb{N}_+$  so we can check that  $f$  is a valid PDF:

$$\sum_{x=1}^{\infty} \left( \frac{1}{x} - \frac{1}{x+1} \right) = \lim_{n \rightarrow \infty} \sum_{x=1}^n \left( \frac{1}{x} - \frac{1}{x+1} \right) = \lim_{n \rightarrow \infty} \left( 1 - \frac{1}{n+1} \right) = 1 \quad (3.1.12)$$

## Discrete Uniform Distributions

An element  $X$  is chosen at random from a finite set  $S$ . The distribution of  $X$  is the *discrete uniform distribution* on  $S$ .

- $X$  has probability density function  $f$  given by  $f(x) = 1/\#(S)$  for  $x \in S$ .
- $\mathbb{P}(X \in A) = \#(A)/\#(S)$  for  $A \subseteq S$ .

Proof

The phrase *at random* means that all outcomes are equally likely.

Many random variables that arise in sampling or combinatorial experiments are transformations of uniformly distributed variables. The next few exercises review the standard methods of sampling from a finite population. The parameters  $m$  and  $n$  are positive integers.

Suppose that  $n$  elements are chosen at random, with replacement from a set  $D$  with  $m$  elements. Let  $\mathbf{X}$  denote the ordered sequence of elements chosen. Then  $\mathbf{X}$  is uniformly distributed on the Cartesian power set  $S = D^n$ , and has probability density function  $f$  given by

$$f(\mathbf{x}) = \frac{1}{m^n}, \quad \mathbf{x} \in S \quad (3.1.13)$$

Proof

Recall that  $\#(D^n) = m^n$ .

Suppose that  $n$  elements are chosen at random, without replacement from a set  $D$  with  $m$  elements (so  $n \leq m$ ). Let  $\mathbf{X}$  denote the ordered sequence of elements chosen. Then  $\mathbf{X}$  is uniformly distributed on the set  $S$  of permutations of size  $n$  chosen from  $D$ , and has probability density function  $f$  given by

$$f(\mathbf{x}) = \frac{1}{m^{(n)}}, \quad \mathbf{x} \in S \quad (3.1.14)$$

Proof

Recall that the number of permutations of size  $n$  from  $D$  is  $m^{(n)}$ .

Suppose that  $n$  elements are chosen at random, without replacement, from a set  $D$  with  $m$  elements (so  $n \leq m$ ). Let  $\mathbf{W}$  denote the unordered set of elements chosen. Then  $\mathbf{W}$  is uniformly distributed on the set  $T$  of combinations of size  $n$  chosen from  $D$ , and has probability density function  $f$  given by

$$f(\mathbf{w}) = \frac{1}{\binom{m}{n}}, \quad \mathbf{w} \in T \quad (3.1.15)$$

Proof

Recall that the number of combinations of size  $n$  from  $D$  is  $\binom{m}{n}$ .

Suppose that  $X$  is uniformly distributed on a finite set  $S$  and that  $B$  is a nonempty subset of  $S$ . Then the conditional distribution of  $X$  given  $X \in B$  is uniform on  $B$ .

Proof

From (7), the conditional probability density function of  $X$  given  $X \in B$  is

$$f(x | B) = \frac{f(x)}{\mathbb{P}(X \in B)} = \frac{1/\#(S)}{\#(B)/\#(S)} = \frac{1}{\#(B)}, \quad x \in B \quad (3.1.16)$$

## Hypergeometric Models

Suppose that a *dichotomous* population consists of  $m$  objects of two different types:  $r$  of the objects are type 1 and  $m - r$  are type 0. Here are some typical examples:

- The objects are persons, each either *male* or *female*.
- The objects are voters, each either a *democrat* or a *republican*.
- The objects are devices of some sort, each either *good* or *defective*.
- The objects are fish in a lake, each either *tagged* or *untagged*.
- The objects are balls in an urn, each either *red* or *green*.

A sample of  $n$  objects is chosen at random (without replacement) from the population. Recall that this means that the samples, either ordered or unordered are equally likely. Note that this probability model has three parameters: the population size  $m$ , the number of type 1 objects  $r$ , and the sample size  $n$ . Each is a nonnegative integer with  $r \leq m$  and  $n \leq m$ . Now, suppose that we keep track of order, and let  $X_i$  denote the type of the  $i$ th object chosen, for  $i \in \{1, 2, \dots, n\}$ . Thus,  $X_i$  is an indicator variable (that is, a variable that just takes values 0 and 1).

$\mathbf{X} = (X_1, X_2, \dots, X_n)$  has probability density function  $f$  given by

$$f(x_1, x_2, \dots, x_n) = \frac{r^{(y)}(m-r)^{(n-y)}}{m^{(n)}}, \quad (x_1, x_2, \dots, x_n) \in \{0, 1\}^n \text{ where } y = x_1 + x_2 + \dots + x_n \quad (3.1.17)$$

Proof

Recall again that the ordered samples are equally likely, and there are  $m^{(n)}$  such samples. The number of ways to select the  $y$  type 1 objects and place them in the positions where  $x_i = 1$  is  $r^{(y)}$ . The number of ways to select the  $n - y$  type 0 objects and place them in the positions where  $x_i = 0$  is  $(m - r)^{(n-y)}$ . Thus the result follows from the multiplication principle.

Note that the value of  $f(x_1, x_2, \dots, x_n)$  depends only on  $y = x_1 + x_2 + \dots + x_n$ , and hence is unchanged if  $(x_1, x_2, \dots, x_n)$  is permuted. This means that  $(X_1, X_2, \dots, X_n)$  is exchangeable. In particular, the distribution of  $X_i$  is the same as the distribution of  $X_1$ , so  $\mathbb{P}(X_i = 1) = \frac{r}{m}$ . Thus, the variables are *identically distributed*. Also the distribution of  $(X_i, X_j)$  is the same as the distribution of  $(X_1, X_2)$ , so  $\mathbb{P}(X_i = 1, X_j = 1) = \frac{r(r-1)}{m(m-1)}$ . Thus,  $X_i$  and  $X_j$  are not independent, and in fact are negatively correlated.

Now let  $Y$  denote the number of type 1 objects in the sample. Note that  $Y = \sum_{i=1}^n X_i$ . Any counting variable can be written as a sum of indicator variables.

$Y$  has probability density function  $g$  given by.

$$g(y) = \frac{\binom{r}{y} \binom{m-r}{n-y}}{\binom{m}{n}}, \quad y \in \{0, 1, \dots, n\} \quad (3.1.18)$$

1.  $g(y-1) < g(y)$  if and only if  $y < t$  where  $t = (r+1)(n+1)/(m+2)$ .
2. If  $t$  is not a positive integer, there is a single mode at  $\lfloor t \rfloor$ .
3. If  $t$  is a positive integer, then there are two modes, at  $t-1$  and  $t$ .

Proof

Recall again that the unordered samples of size  $n$  chosen from the population are equally likely. By the multiplication principle, the number of samples with exactly  $y$  type 1 objects and  $n - y$  type 0 objects is  $\binom{m}{y} \binom{m-r}{n-y}$ . The total number of samples is  $\binom{m}{n}$ .

1. Note that  $g(y-1) < g(y)$  if and only if  $\binom{r}{y-1} \binom{m-r}{n+1-y} < \binom{r}{y} \binom{m-r}{n-y}$ . Writing the binomial coefficients in terms of factorials and canceling terms gives  $g(y-1) < g(y)$  if and only if  $y < t$ , where  $t$  is given above.
2. By the same argument,  $f(y-1) = f(y)$  if and only if  $y = t$ . If  $t$  is not an integer then this cannot happen. Letting  $z = \lfloor t \rfloor$ , it follows from (a) that  $g(y) < g(z)$  if  $y < z$  or  $y > z$ .
3. If  $t$  is a positive integer, then by (b),  $g(t-1) = g(t)$  and by (a)  $g(y) < g(t-1)$  if  $y < t-1$  and  $g(y) < g(t)$  if  $y > t$ .

The distribution defined by the probability density function in the last result is the *hypergeometric distributions* with parameters  $m$ ,  $r$ , and  $n$ . The term *hypergeometric* comes from a certain class of special functions, but is not particularly helpful in terms of remembering the model. Nonetheless, we are stuck with it. The set of values  $\{0, 1, \dots, n\}$  is a *convenience set*: it contains all of the values that have positive probability, but depending on the parameters, some of the values may have probability 0. Recall our convention for binomial coefficients: for  $j, k \in \mathbb{N}_+$ ,  $\binom{k}{j} = 0$  if  $j > k$ . Note also that the hypergeometric distribution is unimodal: the probability density function increases and then decreases, with either a single mode or two adjacent modes.

We can extend the hypergeometric model to a population of three types. Thus, suppose that our population consists of  $m$  objects;  $r$  of the objects are type 1,  $s$  are type 2, and  $m - r - s$  are type 0. Here are some examples:

- The objects are voters, each a *democrat*, a *republican*, or an *independent*.
- The objects are cicadas, each one of three species: *tredecula*, *tredecassini*, or *tredecim*.
- The objects are peaches, each classified as *small*, *medium*, or *large*.
- The objects are faculty members at a university, each an *assistant professor*, or an *associate professor*, or a *full professor*.

Once again, a sample of  $n$  objects is chosen at random (without replacement). The probability model now has four parameters: the population size  $m$ , the type sizes  $r$  and  $s$ , and the sample size  $n$ . All are nonnegative integers with  $r + s \leq m$  and  $n \leq m$ . Moreover, we now need two random variables to keep track of the counts for the three types in the sample. Let  $Y$  denote the number of type 1 objects in the sample and  $Z$  the number of type 2 objects in the sample.

$(Y, Z)$  has probability density function  $h$  given by

$$h(y, z) = \frac{\binom{r}{y} \binom{s}{z} \binom{m-r-s}{n-y-z}}{\binom{m}{n}}, \quad (y, z) \in \{0, 1, \dots, n\}^2 \text{ with } y + z \leq n \quad (3.1.19)$$

Proof

Once again, by the multiplication principle, the number of samples of size  $n$  from the population with exactly  $y$  type 1 objects,  $z$  type 2 objects, and  $n - y - z$  type 0 objects is  $\binom{r}{y} \binom{s}{z} \binom{m-r-s}{n-y-z}$ . The total number of samples of size  $n$  is  $\binom{m}{n}$ .

The distribution defined by the density function in the last exercise is the *bivariate hypergeometric distribution* with parameters  $m$ ,  $r$ ,  $s$ , and  $n$ . Once again, the domain given is a convenience set; it includes the set of points with positive probability, but depending on the parameters, may include points with probability 0. Clearly, the same general pattern applies to populations with even more types. However, because of all of the parameters, the formulas are not worth remembering in detail; rather, just note the pattern, and remember the combinatorial meaning of the binomial coefficient. The hypergeometric model will be revisited later in this chapter, in the section on joint distributions and in the section on conditional distributions. The hypergeometric distribution and the multivariate hypergeometric distribution are studied in detail in the chapter on Finite Sampling Models. This chapter contains a variety of distributions that are based on discrete uniform distributions.

### Bernoulli Trials

A *Bernoulli trials sequence* is a sequence  $(X_1, X_2, \dots)$  of independent, identically distributed indicator variables. Random variable  $X_i$  is the outcome of trial  $i$ , where in the usual terminology of reliability, 1 denotes *success* while 0 denotes *failure*. The process is named for Jacob Bernoulli. Let  $p = \mathbb{P}(X_i = 1) \in [0, 1]$  denote the success parameter of the process. Note that the indicator variables in the [hypergeometric model](#) satisfy one of the assumptions of Bernoulli trials (identical distributions) but not the other (independence).

$\mathbf{X} = (X_1, X_2, \dots, X_n)$  has probability density function  $f$  given by

$$f(x_1, x_2, \dots, x_n) = p^y (1-p)^{n-y}, \quad (x_1, x_2, \dots, x_n) \in \{0, 1\}^n, \text{ where } y = x_1 + x_2 + \dots + x_n \quad (3.1.20)$$

Proof

By definition,  $\mathbb{P}(X_i = 1) = p$  and  $\mathbb{P}(X_i = 0) = 1 - p$ . Equivalently,  $\mathbb{P}(X_i = x) = p^x (1-p)^{1-x}$  for  $x \in \{0, 1\}$ . The formula for  $f$  then follows by independence.

Now let  $Y$  denote the number of successes in the first  $n$  trials. Note that  $Y = \sum_{i=1}^n X_i$ , so we see again that a complicated random variable can be written as a sum of simpler ones. In particular, a counting variable can always be written as a sum of indicator variables.

$Y$  has probability density function  $g$  given by

$$g(y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y \in \{0, 1, \dots, n\} \quad (3.1.21)$$

1.  $g(y-1) < g(y)$  if and only if  $y < t$ , where  $t = (n+1)p$ .
2. If  $t$  is not a positive integer, there is a single mode at  $\lfloor t \rfloor$ .
3. If  $t$  is a positive integer, then there are two modes, at  $t-1$  and  $t$ .

Proof

From the previous result, any particular sequence of  $n$  Bernoulli trials with  $y$  successes and  $n-y$  failures has probability  $p^y (1-p)^{n-y}$ . The number of such sequences is  $\binom{n}{y}$ , so the formula for  $g$  follows by the additivity of probability.

1. Note that  $g(y-1) < g(y)$  if and only if  $\binom{n}{y-1} p^{y-1} (1-p)^{n-y+1} < \binom{n}{y} p^y (1-p)^{n-y}$ . Writing the binomial coefficients in terms of factorials and canceling gives  $g(y-1) < g(y)$  if and only if  $y < t$  where  $t = (n+1)p$ .
2. By the same argument,  $g(y-1) = g(y)$  if and only if  $y = t$ . If  $t$  is not an integer, this cannot happen. Letting  $z = \lfloor t \rfloor$ , it follows from (a) that  $g(y) < g(z)$  if  $y < z$  or  $y > z$ .
3. If  $t$  is a positive integer, then by (b),  $g(t-1) = g(t)$  and by (a)  $g(y) < g(t-1)$  if  $y < t-1$  and  $g(y) < g(t)$  if  $y > t$ .



The distribution defined by the probability density function in the last theorem is called the *binomial distribution* with parameters  $n$  and  $p$ . The distribution is unimodal: the probability density function at first increases and then decreases, with either a single mode or two adjacent modes. The binomial distribution is studied in detail in the chapter on Bernoulli Trials.

Suppose that  $p > 0$  and let  $N$  denote the trial number of the first success. Then  $N$  has probability density function  $h$  given by

$$h(n) = (1-p)^{n-1}p, \quad n \in \mathbb{N}_+ \quad (3.1.22)$$

The probability density function  $h$  is decreasing and the mode is  $n = 1$ .

Proof

For  $n \in \mathbb{N}_+$ , the event  $\{N = n\}$  means that the first  $n - 1$  trials were failures and trial  $n$  was a success. Each trial results in failure with probability  $1 - p$  and success with probability  $p$ , and the trials are independent, so  $\mathbb{P}(N = n) = (1 - p)^{n-1}p$ . Using geometric series, we can check that

$$\sum_{n=1}^{\infty} h(n) = \sum_{n=1}^{\infty} p(1-p)^{n-1} = \frac{p}{1-(1-p)} = 1 \quad (3.1.23)$$

The distribution defined by the probability density function in the last exercise is the *geometric distribution* on  $\mathbb{N}_+$  with parameter  $p$ . The geometric distribution is studied in detail in the chapter on Bernoulli Trials.

### Sampling Problems

In the following exercises, be sure to check if the problem fits one of the general models above.

An urn contains 30 red and 20 green balls. A sample of 5 balls is selected at random, without replacement. Let  $Y$  denote the number of red balls in the sample.

1. Compute the probability density function of  $Y$  explicitly and identify the distribution by name and parameter values.
2. Graph the probability density function and identify the mode(s).
3. Find  $\mathbb{P}(Y > 3)$ .

Answer

1.  $f(0) = 0.0073$ ,  $f(1) = 0.0686$ ,  $f(2) = 0.2341$ ,  $f(3) = 0.3641$ ,  $f(4) = 0.2587$ ,  $f(5) = 0.0673$  Hypergeometric with  $m = 50$ ,  $r = 30$ ,  $n = 5$
2. mode:  $y = 3$
3.  $\mathbb{P}(Y > 3) = 0.3260$

In the ball and urn experiment, select sampling without replacement and set  $m = 50$ ,  $r = 30$ , and  $n = 5$ . Run the experiment 1000 times and note the agreement between the empirical density function of  $Y$  and the probability density function.

An urn contains 30 red and 20 green balls. A sample of 5 balls is selected at random, *with* replacement. Let  $Y$  denote the number of red balls in the sample.

1. Compute the probability density function of  $Y$  explicitly and identify the distribution by name and parameter values.
2. Graph the probability density function and identify the mode(s).
3. Find  $\mathbb{P}(Y > 3)$ .

Answer

1.  $f(0) = 0.0102$ ,  $f(1) = 0.0768$ ,  $f(2) = 0.2304$ ,  $f(3) = 0.3456$ ,  $f(4) = 0.2592$ ,  $f(5) = 0.0778$  Binomial with  $n = 5$ ,  $p = 3/5$
2. mode:  $y = 3$
3.  $\mathbb{P}(Y > 3) = 0.3370$

In the ball and urn experiment, select sampling with replacement and set  $m = 50$ ,  $r = 30$ , and  $n = 5$ . Run the experiment 1000 times and note the agreement between the empirical density function of  $Y$  and the probability density function.

A group of voters consists of 50 democrats, 40 republicans, and 30 independents. A sample of 10 voters is chosen at random, without replacement. Let  $X$  denote the number of democrats in the sample and  $Y$  the number of republicans in the sample.

1. Give the probability density function of  $X$ .
2. Give the probability density function of  $Y$ .
3. Give the probability density function of  $(X, Y)$ .
4. Find the probability that the sample has at least 4 democrats and at least 4 republicans.

Answer

1.  $g(x) = \frac{\binom{50}{x} \binom{70}{10-x}}{\binom{120}{10}}$  for  $x \in \{0, 1, \dots, 10\}$  This is the hypergeometric distribution with parameters  $m = 120$ ,  $r = 50$  and  $n = 10$ .
2.  $h(y) = \frac{\binom{40}{y} \binom{80}{10-y}}{\binom{120}{10}}$  for  $y \in \{0, 1, \dots, 10\}$  This is the hypergeometric distribution with parameters  $m = 120$ ,  $r = 40$  and  $n = 10$ .
3.  $f(x, y) = \frac{\binom{50}{x} \binom{40}{y} \binom{30}{10-x-y}}{\binom{120}{10}}$  for  $(x, y) \in \{0, 1, \dots, 10\}^2$  with  $x + y \leq 10$ . This is the bivariate hypergeometric distribution with parameters  $m = 120$ ,  $r = 50$ ,  $s = 40$  and  $n = 10$ .
4.  $\mathbb{P}(X \geq 4, Y \geq 4) = \frac{15\,137\,200}{75\,597\,113} \approx 0.200$

The Math Club at Enormous State University (ESU) has 20 freshmen, 40 sophomores, 30 juniors, and 10 seniors. A committee of 8 club members is chosen at random, without replacement to organize  $\pi$ -day activities. Let  $X$  denote the number of freshman in the sample,  $Y$  the number of sophomores, and  $Z$  the number of juniors.

1. Give the probability density function of  $X$ .
2. Give the probability density function of  $Y$ .
3. Give the probability density function of  $Z$ .
4. Give the probability density function of  $(X, Y)$ .
5. Give the probability density function of  $(X, Y, Z)$ .
6. Find the probability that the committee has no seniors.

Answer

1.  $f_X(x) = \frac{\binom{20}{x} \binom{80}{8-x}}{\binom{100}{8}}$  for  $x \in \{0, 1, \dots, 8\}$  This is the hypergeometric distribution with parameters  $m = 100$ ,  $r = 20$ , and  $n = 8$ .
2.  $f_Y(y) = \frac{\binom{40}{y} \binom{60}{8-y}}{\binom{100}{8}}$  for  $y \in \{0, 1, \dots, 8\}$  This is the hypergeometric distribution with parameters  $m = 100$ ,  $r = 40$ , and  $n = 8$ .
3.  $f_Z(z) = \frac{\binom{30}{z} \binom{70}{8-z}}{\binom{100}{8}}$  for  $z \in \{0, 1, \dots, 8\}$  This is the hypergeometric distribution with parameters  $m = 100$ ,  $r = 30$ , and  $n = 8$ .
4.  $f_{X,Y}(x, y) = \frac{\binom{20}{x} \binom{40}{y} \binom{40}{8-x-y}}{\binom{100}{8}}$  for  $(x, y) \in \{0, 1, \dots, 8\}^2$  with  $x + y \leq 8$ . This is the bivariate hypergeometric distribution with parameters  $m = 100$ ,  $r = 20$ ,  $s = 40$  and  $n = 8$ .
5.  $f_{X,Y,Z}(x, y, z) = \frac{\binom{20}{x} \binom{40}{y} \binom{30}{z} \binom{10}{8-x-y-z}}{\binom{100}{8}}$  for  $(x, y, z) \in \{0, 1, \dots, 8\}^3$  with  $x + y + z \leq 8$ . This is the tri-variate hypergeometric distribution with parameters  $m = 100$ ,  $r = 20$ ,  $s = 40$ ,  $t = 30$ , and  $n = 8$ .
6.  $\mathbb{P}(X + Y + Z = 8) = \frac{156\,597\,013}{275\,935\,140} \approx 0.417$

## Coins and Dice

Suppose that a coin with probability of heads  $p$  is tossed repeatedly, and the sequence of heads and tails is recorded.

1. Identify the underlying probability model by name and parameter.
2. Let  $Y$  denote the number of heads in the first  $n$  tosses. Give the probability density function of  $Y$  and identify the distribution by name and parameters.
3. Let  $N$  denote the number of tosses needed to get the first head. Give the probability density function of  $N$  and identify the distribution by name and parameter.

Answer

1. Bernoulli trials with success parameter  $p$ .
2.  $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$  for  $k \in \{0, 1, \dots, n\}$ . This is the binomial distribution with trial parameter  $n$  and success parameter  $p$ .
3.  $g(n) = p(1-p)^{n-1}$  for  $n \in \mathbb{N}_+$ . This is the geometric distribution with success parameter  $p$ .

Suppose that a coin with probability of heads  $p = 0.4$  is tossed 5 times. Let  $Y$  denote the number of heads.

1. Compute the probability density function of  $Y$  explicitly.
2. Graph the probability density function and identify the mode.
3. Find  $\mathbb{P}(Y > 3)$ .

Answer

1.  $f(0) = 0.0778$   $f(1) = 0.2592$   $f(2) = 0.3456$   $f(3) = 0.2304$   $f(4) = 0.0768$   $f(5) = 0.0102$
2. mode:  $k = 2$
3.  $\mathbb{P}(Y > 3) = 0.0870$

In the binomial coin experiment, set  $n = 5$  and  $p = 0.4$ . Run the experiment 1000 times and compare the empirical density function of  $Y$  with the probability density function.

Suppose that a coin with probability of heads  $p = 0.2$  is tossed until heads occurs. Let  $N$  denote the number of tosses.

1. Find the probability density function of  $N$ .
2. Find  $\mathbb{P}(N \leq 5)$ .

Answer

1.  $f(n) = (0.8)^{n-1} 0.2$  for  $n \in \mathbb{N}_+$
2.  $\mathbb{P}(N \leq 5) = 0.67232$

In the negative binomial experiment, set  $k = 1$  and  $p = 0.2$ . Run the experiment 1000 times and compare the empirical density function with the probability density function.

Suppose that two fair, standard dice are tossed and the sequence of scores  $(X_1, X_2)$  recorded. Let  $Y = X_1 + X_2$  denote the sum of the scores,  $U = \min\{X_1, X_2\}$  the minimum score, and  $V = \max\{X_1, X_2\}$  the maximum score.

1. Find the probability density function of  $(X_1, X_2)$ . Identify the distribution by name.
2. Find the probability density function of  $Y$ .
3. Find the probability density function of  $U$ .
4. Find the probability density function of  $V$ .
5. Find the probability density function of  $(U, V)$ .

Answer

We denote the PDFs by  $f$ ,  $g$ ,  $h_1$ ,  $h_2$ , and  $h$  respectively.

1.  $f(x_1, x_2) = \frac{1}{36}$  for  $(x_1, x_2) \in \{1, 2, 3, 4, 5, 6\}^2$ . This is the uniform distribution on  $\{1, 2, 3, 4, 5, 6\}^2$ .
2.  $g(2) = g(12) = \frac{1}{36}$ ,  $g(3) = g(11) = \frac{2}{36}$ ,  $g(4) = g(10) = \frac{3}{36}$ ,  $g(5) = g(9) = \frac{4}{36}$ ,  $g(6) = g(8) = \frac{5}{36}$ ,  $g(7) = \frac{6}{36}$
3.  $h_1(1) = \frac{11}{36}$ ,  $h_1(2) = \frac{9}{36}$ ,  $h_1(3) = \frac{7}{36}$ ,  $h_1(4) = \frac{5}{36}$ ,  $h_1(5) = \frac{3}{36}$ ,  $h_1(6) = \frac{1}{36}$
4.  $h_2(1) = \frac{1}{36}$ ,  $h_2(2) = \frac{3}{36}$ ,  $h_2(3) = \frac{5}{36}$ ,  $h_2(4) = \frac{7}{36}$ ,  $h_2(5) = \frac{9}{36}$ ,  $h_2(6) = \frac{11}{36}$
5.  $h(u, v) = \frac{2}{36}$  if  $u < v$ ,  $h(u, v) = \frac{1}{36}$  if  $u = v$  where  $(u, v) \in \{1, 2, 3, 4, 5, 6\}^2$  with  $u \leq v$

Note that  $(U, V)$  in the last exercise could serve as the outcome of the experiment that consists of throwing two standard dice if we did not bother to record order. Note from the previous exercise that this random vector does not have a uniform distribution when the dice are fair. The mistaken idea that this vector *should* have the uniform distribution was the cause of difficulties in the early development of probability.

In the dice experiment, select  $n = 2$  fair dice. Select the following random variables and note the shape and location of the probability density function. Run the experiment 1000 times. For each of the following variables, compare the empirical density function with the probability density function.

1.  $Y$ , the sum of the scores.
2.  $U$ , the minimum score.
3.  $V$ , the maximum score.

In the die-coin experiment, a fair, standard die is rolled and then a fair coin is tossed the number of times showing on the die. Let  $N$  denote the die score and  $Y$  the number of heads.

1. Find the probability density function of  $N$ . Identify the distribution by name.
2. Find the probability density function of  $Y$ .

Answer

1.  $g(n) = \frac{1}{6}$  for  $n \in \{1, 2, 3, 4, 5, 6\}$  This is the uniform distribution on  $\{1, 2, 3, 4, 5, 6\}$
2.  $h(0) = \frac{63}{384}$ ,  $h(1) = \frac{120}{384}$ ,  $h(2) = \frac{90}{384}$ ,  $h(3) = \frac{64}{384}$ ,  $h(4) = \frac{29}{384}$ ,  $h(5) = \frac{8}{384}$ ,  $h(6) = \frac{1}{384}$

Run the die-coin experiment 1000 times. For the number of heads, compare the empirical density function with the probability density function.

Suppose that a bag contains 12 coins: 5 are fair, 4 are biased with probability of heads  $\frac{1}{3}$ ; and 3 are two-headed. A coin is chosen at random from the bag and tossed 5 times. Let  $V$  denote the probability of heads of the selected coin and let  $Y$  denote the number of heads.

1. Find the probability density function of  $V$ .
2. Find the probability density function of  $Y$ .

Answer

1.  $g(1/2) = 5/12$ ,  $g(1/3) = 4/12$ ,  $g(1) = 3/12$
2.  $h(0) = 5311/93312$ ,  $h(1) = 16315/93312$ ,  $h(2) = 22390/93312$ ,  $h(3) = 17270/93312$ ,  $h(4) = 7355/93312$ ,  $h(5) = 24671/93312$

Compare the [die-coin experiment](#) with the [bag of coins experiment](#). In the first experiment, we toss a coin with a *fixed* probability of heads a *random* number of times. In second experiment, we effectively toss a coin with a *random* probability of heads a *fixed* number of times. In both cases, we can think of starting with a binomial distribution and *randomizing* one of the parameters.

In the coin-die experiment, a fair coin is tossed. If the coin lands tails, a fair die is rolled. If the coin lands heads, an ace-six flat die is tossed (faces 1 and 6 have probability  $\frac{1}{4}$  each, while faces 2, 3, 4, 5 have probability  $\frac{1}{8}$  each). Find the probability density function of the die score  $Y$ .

Answer

$$f(y) = 5/24 \text{ for } y \in \{1, 6\}, f(y) = 7/24 \text{ for } y \in \{2, 3, 4, 5\}$$

Run the coin-die experiment 1000 times, with the settings in the previous exercise. Compare the empirical density function with the probability density function.

Suppose that a standard die is thrown 10 times. Let  $Y$  denote the number of times an ace or a six occurred. Give the probability density function of  $Y$  and identify the distribution by name and parameter values in each of the following cases:

1. The die is fair.

2. The die is an ace-six flat.

Answer

1.  $f(k) = \binom{10}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{10-k}$  for  $k \in \{0, 1, \dots, 10\}$ . This is the binomial distribution with trial parameter  $n = 10$  and success parameter  $p = \frac{1}{3}$
2.  $f(k) = \binom{10}{k} \left(\frac{1}{2}\right)^{10}$  for  $k \in \{0, 1, \dots, 10\}$ . This is the binomial distribution with trial parameter  $n = 10$  and success parameter  $p = \frac{1}{2}$

Suppose that a standard die is thrown until an ace or a six occurs. Let  $N$  denote the number of throws. Give the probability density function of  $N$  and identify the distribution by name and parameter values in each of the following cases:

1. The die is fair.
2. The die is an ace-six flat.

Answer

1.  $g(n) = \left(\frac{2}{3}\right)^{n-1} \frac{1}{3}$  for  $n \in \mathbb{N}_+$ . This is the geometric distribution with success parameter  $p = \frac{1}{3}$
2.  $g(n) = \left(\frac{1}{2}\right)^n$  for  $n \in \mathbb{N}_+$ . This is the geometric distribution with success parameter  $p = \frac{1}{2}$

Fred and Wilma takes turns tossing a coin with probability of heads  $p \in (0, 1)$ : Fred first, then Wilma, then Fred again, and so forth. The first person to toss heads wins the game. Let  $N$  denote the number of tosses, and  $W$  the event that Wilma wins.

1. Give the probability density function of  $N$  and identify the distribution by name.
2. Compute  $\mathbb{P}(W)$  and sketch the graph of this probability as a function of  $p$ .
3. Find the conditional probability density function of  $N$  given  $W$ .

Answer

1.  $f(n) = p(1-p)^{n-1}$  for  $n \in \mathbb{N}_+$ . This is the geometric distribution with success parameter  $p$ .
2.  $\mathbb{P}(W) = \frac{1-p}{2-p}$
3.  $f(n | W) = p(2-p)(1-p)^{n-2}$  for  $n \in \{2, 4, \dots\}$

The *alternating coin tossing game* is studied in more detail in the section on The Geometric Distribution in the chapter on Bernoulli trials.

Suppose that  $k$  players each have a coin with probability of heads  $p$ , where  $k \in \{2, 3, \dots\}$  and where  $p \in (0, 1)$ .

1. Suppose that the players toss their coins at the same time. Find the probability that there is an *odd man*, that is, one player with a different outcome than all the rest.
2. Suppose now that the players repeat the procedure in part (a) until there is an odd man. Find the probability density function of  $N$ , the number of rounds played, and identify the distribution by name.

Answer

1. The probability is  $2p(1-p)$  if  $k = 2$ , and is  $kp(1-p)^{k-1} + kp^{k-1}(1-p)$  if  $k > 2$ .
2. Let  $r_k$  denote the probability in part (a).  $N$  has PDF  $f(n) = (1-r_k)^{n-1} r_k$  for  $n \in \mathbb{N}$ , and has the geometric distribution with parameter  $r_k$ .

The *odd man out* game is treated in more detail in the section on the Geometric Distribution in the chapter on Bernoulli Trials.

## Cards

Recall that a *poker hand* consists of 5 cards chosen at random and without replacement from a standard deck of 52 cards. Let  $X$  denote the number of spades in the hand and  $Y$  the number of hearts in the hand. Give the probability density function of each of the following random variables, and identify the distribution by name:

1.  $X$
2.  $Y$
3.  $(X, Y)$

Answer

1.  $g(x) = \frac{\binom{13}{x} \binom{39}{5-x}}{\binom{52}{5}}$  for  $x \in \{0, 1, 2, 3, 4, 5\}$  This is the hypergeometric distribution with population size  $m = 52$ , type parameter  $r = 13$ , and sample size  $n = 5$
2.  $h(y) = \frac{\binom{13}{y} \binom{39}{5-y}}{\binom{52}{5}}$  for  $y \in \{0, 1, 2, 3, 4, 5\}$  This is the same hypergeometric distribution as in part (a).
3.  $f(x, y) = \frac{\binom{13}{x} \binom{13}{y} \binom{26}{5-x-y}}{\binom{52}{5}}$  for  $(x, y) \in \{0, 1, 2, 3, 4, 5\}^2$  with  $x + y \leq 5$ . This is a bivariate hypergeometric distribution.

Recall that a *bridge hand* consists of 13 cards chosen at random and without replacement from a standard deck of 52 cards. An *honor card* is a card of denomination ace, king, queen, jack or 10. Let  $N$  denote the number of honor cards in the hand.

1. Find the probability density function of  $N$  and identify the distribution by name.
2. Find the probability that the hand has no honor cards. A hand of this kind is known as a Yarborough, in honor of Second Earl of Yarborough.

Answer

1.  $f(n) = \frac{\binom{20}{n} \binom{32}{13-n}}{\binom{52}{13}}$  for  $n \in \{0, 1, \dots, 13\}$  This is the hypergeometric distribution with population size  $m = 52$ , type parameter  $r = 20$  and sample size  $n = 13$ .
2. 0.00547

In the most common *high card point system* in bridge, an ace is worth 4 points, a king is worth 3 points, a queen is worth 2 points, and a jack is worth 1 point. Find the probability density function of  $V$ , the point value of a random bridge hand.

## Reliability

Suppose that in a batch of 500 components, 20 are defective and the rest are good. A sample of 10 components is selected at random and tested. Let  $X$  denote the number of defectives in the sample.

1. Find the probability density function of  $X$  and identify the distribution by name and parameter values.
2. Find the probability that the sample contains at least one defective component.

Answer

1.  $f(x) = \frac{\binom{20}{x} \binom{480}{10-x}}{\binom{500}{10}}$  for  $x \in \{0, 1, \dots, 10\}$  This is the hypergeometric distribution with population size  $m = 500$ , type parameter  $r = 20$ , and sample size  $n = 10$ .
2.  $\mathbb{P}(X \geq 1) = 1 - \frac{\binom{480}{10}}{\binom{500}{10}} \approx 0.3377$

A plant has 3 assembly lines that produce a certain type of component. Line 1 produces 50% of the components and has a defective rate of 4%; line 2 produces 30% of the components and has a defective rate of 5%; line 3 produces 20% of the components and has a defective rate of 1%. A component is chosen at random from the plant and tested.

1. Find the probability that the component is defective.
2. Given that the component is defective, find the conditional probability density function of the line that produced the component.

Answer

Let  $D$  the event that the item is defective, and  $f(\cdot | D)$  the PDF of the line number given  $D$ .

1.  $\mathbb{P}(D) = 0.037$
2.  $f(1 | D) = 0.541, f(2 | D) = 0.405, f(3 | D) = 0.054$

Recall that in the standard model of *structural reliability*, a system consists of  $n$  components, each of which, independently of the others, is either working or failed. Let  $X_i$  denote the state of component  $i$ , where 1 means working and 0 means failed. Thus, the state vector is  $\mathbf{X} = (X_1, X_2, \dots, X_n)$ . The system as a whole is also either working or failed, depending only on the states of the components. Thus, the state of the system is an indicator random variable  $U = u(\mathbf{X})$  that depends on the states of the components according to a *structure function*  $u : \{0, 1\}^n \rightarrow \{0, 1\}$ . In a *series system*, the system works if and only if every component works. In a *parallel system*, the system works if and only if at least one component works. In a *k out of n system*, the system works if and only if at least  $k$  of the  $n$  components work.

The *reliability* of a device is the probability that it is working. Let  $p_i = \mathbb{P}(X_i = 1)$  denote the reliability of component  $i$ , so that  $\mathbf{p} = (p_1, p_2, \dots, p_n)$  is the vector of component reliabilities. Because of the independence assumption, the system reliability depends only on the component reliabilities, according to a *reliability function*  $r(\mathbf{p}) = \mathbb{P}(U = 1)$ . Note that when all component reliabilities have the same value  $p$ , the states of the components form a sequence of  $n$  [Bernoulli trials](#). In this case, the system reliability is, of course, a function of the common component reliability  $p$ .

Suppose that the component reliabilities all have the same value  $p$ . Let  $\mathbf{X}$  denote the state vector and  $Y$  denote the number of working components.

1. Give the probability density function of  $\mathbf{X}$ .
2. Give the probability density function of  $Y$  and identify the distribution by name and parameter.
3. Find the reliability of the  $k$  out of  $n$  system.

Answer

1.  $f(x_1, x_2, \dots, x_n) = p^y(1-p)^{n-y}$  for  $(x_1, x_2, \dots, x_n) \in \{0, 1\}^n$  where  $y = x_1 + x_2 + \dots + x_n$
2.  $g(k) = \binom{n}{y} p^y (1-p)^{n-y}$  for  $y \in \{0, 1, \dots, n\}$  This is the binomial distribution with trial parameter  $n$  and success parameter  $p$ .
3.  $r(p) = \sum_{i=k}^n \binom{n}{i} p^i (1-p)^{n-i}$

Suppose that we have 4 independent components, with common reliability  $p = 0.8$ . Let  $Y$  denote the number of working components.

1. Find the probability density function of  $Y$  explicitly.
2. Find the reliability of the parallel system.
3. Find the reliability of the 2 out of 4 system.
4. Find the reliability of the 3 out of 4 system.
5. Find the reliability of the series system.

Answer

1.  $g(0) = 0.0016, g(1) = 0.0256, g(2) = 0.1536, g(3) = g(4) = 0.4096$
2.  $r_{4,1} = 0.9984$
3.  $r_{4,2} = 0.9729$
4.  $r_{4,3} = 0.8192$
5.  $r_{4,4} = 0.4096$

Suppose that we have 4 independent components, with reliabilities  $p_1 = 0.6, p_2 = 0.7, p_3 = 0.8$ , and  $p_4 = 0.9$ . Let  $Y$  denote the number of working components.

1. Find the probability density function of  $Y$ .
2. Find the reliability of the parallel system.
3. Find the reliability of the 2 out of 4 system.
4. Find the reliability of the 3 out of 4 system.
5. Find the reliability of the series system.

Answer

1.  $g(0) = 0.0024, g(1) = 0.0404, g(2) = 0.2144, g(3) = 0.4404, g(4) = 0.3024$
2.  $r_{4,1} = 0.9976$
3.  $r_{4,2} = 0.9572$

4.  $r_{4,3} = 0.7428$
5.  $r_{4,4} = 0.3024$

## The Poisson Distribution

Suppose that  $a > 0$ . Define  $f$  by

$$f(n) = e^{-a} \frac{a^n}{n!}, \quad n \in \mathbb{N} \quad (3.1.24)$$

1.  $f$  is a probability density function.
2.  $f(n-1) < f(n)$  if and only if  $n < a$ .
3. If  $a$  is not a positive integer, there is a single mode at  $\lfloor a \rfloor$ .
4. If  $a$  is a positive integer, there are two modes at  $a-1$  and  $a$ .

Proof

1. Recall from calculus, the exponential series

$$\sum_{n=0}^{\infty} \frac{a^n}{n!} = e^a \quad (3.1.25)$$

Hence  $f$  is a probability density function.

2. Note that  $f(n-1) < f(n)$  if and only if  $\frac{a^{n-1}}{(n-1)!} < \frac{a^n}{n!}$  if and only if  $1 < \frac{a}{n}$ .
3. By the same argument,  $f(n-1) = f(n)$  if and only if  $a = n$ . If  $a$  is not a positive integer this cannot happen. Hence, letting  $k = \lfloor a \rfloor$ , it follows from (b) that  $f(n) < f(k)$  if  $n < k$  or  $n > k$ .
4. If  $a$  is a positive integer, then  $f(a-1) = f(a)$ . From (b),  $f(n) < f(a-1)$  if  $n < a-1$  and  $f(n) < f(a)$  if  $n > a$ .

The distribution defined by the probability density function in the previous exercise is the *Poisson distribution* with parameter  $a$ , named after Simeon Poisson. Note that like the other named distributions we studied above ([hypergeometric](#) and [binomial](#)), the Poisson distribution is unimodal: the probability density function at first increases and then decreases, with either a single mode or two adjacent modes. The Poisson distribution is studied in detail in the Chapter on Poisson Processes, and is used to model the number of “random points” in a region of time or space, under certain ideal conditions. The parameter  $a$  is proportional to the size of the region of time or space.

Suppose that the customers arrive at a service station according to the Poisson model, at an average rate of 4 per hour. Thus, the number of customers  $N$  who arrive in a 2-hour period has the Poisson distribution with parameter 8.

1. Find the modes.
2. Find  $\mathbb{P}(N \geq 6)$ .

Answer

1. modes: 7, 8
2.  $\mathbb{P}(N > 6) = 0.8088$

In the Poisson experiment, set  $r = 4$  and  $t = 2$ . Run the simulation 1000 times and compare the empirical density function to the probability density function.

Suppose that the number of flaws  $N$  in a piece of fabric of a certain size has the Poisson distribution with parameter 2.5.

1. Find the mode.
2. Find  $\mathbb{P}(N > 4)$ .

Answer

1. mode: 2
2.  $\mathbb{P}(N > 4) = 0.1088$

Suppose that the number of raisins  $N$  in a piece of cake has the Poisson distribution with parameter 10.



1. Find the modes.
2. Find  $\mathbb{P}(8 \leq N \leq 12)$ .

Answer

1. modes: 9, 10
2.  $\mathbb{P}(8 \leq N \leq 12) = 0.5713$

### A Zeta Distribution

Let  $g$  be the function defined by  $g(n) = \frac{1}{n^2}$  for  $n \in \mathbb{N}_+$ .

1. Find the probability density function  $f$  that is **proportional** to  $g$ .
2. Find the mode of the distribution.
3. Find  $\mathbb{P}(N \leq 5)$  where  $N$  has probability density function  $f$ .

Answer

1.  $f(n) = \frac{6}{\pi^2 n^2}$  for  $n \in \mathbb{N}_+$ . Recall that  $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$
2. Mode  $n = 1$
3.  $\mathbb{P}(N \leq 5) = \frac{5269}{600\pi^2}$

The distribution defined in the previous exercise is a member of the *zeta* family of distributions. Zeta distributions are used to model sizes or ranks of certain types of objects, and are studied in more detail in the chapter on Special Distributions.

### Benford's Law

Let  $f$  be the function defined by  $f(d) = \log(d+1) - \log(d) = \log\left(1 + \frac{1}{d}\right)$  for  $d \in \{1, 2, \dots, 9\}$ . (The logarithm function is the base 10 *common logarithm*, not the base  $e$  *natural logarithm*.)

1. Show that  $f$  is a probability density function.
2. Compute the values of  $f$  explicitly, and sketch the graph.
3. Find  $\mathbb{P}(X \leq 3)$  where  $X$  has probability density function  $f$ .

Answer

1. Note that  $\sum_{d=1}^9 f(d) = \log(10) = 1$ . The sum collapses.

2.	$d$	1	2	3	4	5	6	7	8	9
	$f(d)$	0.3010	0.1761	0.1249	0.0969	0.0792	0.0669	0.0580	0.0512	0.0458

3.  $\log(4) \approx 0.6020$

The distribution defined in the previous exercise is known as *Benford's law*, and is named for the American physicist and engineer Frank Benford. This distribution governs the leading digit in many real sets of data. Benford's law is studied in more detail in the chapter on Special Distributions.

### Data Analysis Exercises

In the M&M data, let  $R$  denote the number of red candies and  $N$  the total number of candies. Compute and graph the empirical probability density function of each of the following:

1.  $R$
2.  $N$
3.  $R$  given  $N > 57$

Answer

We denote the PDF of  $R$  by  $f$  and the PDF of  $N$  by  $g$

1.	$r$	3	4	5	6	8	9	10	11	12	14	15	20
----	-----	---	---	---	---	---	---	----	----	----	----	----	----

$f(r)$	$\frac{1}{30}$	$\frac{3}{30}$	$\frac{3}{30}$	$\frac{2}{30}$	$\frac{4}{30}$	$\frac{5}{30}$	$\frac{2}{30}$	$\frac{1}{30}$	$\frac{3}{30}$	$\frac{3}{30}$	$\frac{3}{30}$	$\frac{1}{30}$
--------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------

2. $n$	50	53	54	55	56	57	58	59	60	61
$g(n)$	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{4}{30}$	$\frac{4}{30}$	$\frac{3}{30}$	$\frac{9}{30}$	$\frac{3}{30}$	$\frac{2}{30}$	$\frac{2}{30}$

3. $r$	3	4	6	8	9	11	12	14	15
$f(r \mid N > 57)$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{2}{16}$

In the Cicada data, let  $G$  denotes gender,  $S$  species type, and  $W$  body weight (in grams). Compute the empirical probability density function of each of the following:

1.  $G$
2.  $S$
3.  $(G, S)$
4.  $G$  given  $W > 0.20$  grams.

Answer

We denote the PDF of  $G$  by  $g$ , the PDF of  $S$  by  $h$  and the PDF of  $(G, S)$  by  $f$ .

1.  $g(0) = \frac{59}{104}$ ,  $g(1) = \frac{45}{104}$
2.  $h(0) = \frac{44}{104}$ ,  $h(1) = \frac{6}{104}$ ,  $h(2) = \frac{54}{104}$
3.  $f(0, 0) = \frac{16}{104}$ ,  $f(0, 1) = \frac{3}{104}$ ,  $f(0, 2) = \frac{40}{104}$ ,  $f(1, 0) = \frac{28}{104}$ ,  $f(1, 1) = \frac{3}{104}$ ,  $f(1, 2) = \frac{14}{104}$
4.  $g(0 \mid W > 0.2) = \frac{31}{73}$ ,  $g(1 \mid W > 0.2) = \frac{42}{73}$

This page titled [3.1: Discrete Distributions](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.