Chapter 1

Elementary Probability

1.1 Introduction

Definition A *finite probability space* is a pair $(\mathcal{S}, P)$, in which $\mathcal{S}$ is a finite non-empty set and $P : \mathcal{S} \to [0, 1]$ is a function satisfying $\sum_{s\in\mathcal{S}} P(s) = 1$.

When the space $(\mathcal{S}, P)$ is fixed in the discussion, we will call $\mathcal{S}$ the *set of outcomes* of the space (or of the experiment with which the space is associated—see the discussion below), and P the *probability assignment* to the (set of) outcomes.

The "real" situations we are concerned with consist of an *action*, or *experiment*, with a finite number of mutually exclusive possible outcomes. For a given action, there may be many different ways of listing the possible outcomes, but all acceptable lists of possible outcomes satisfy this test: whatever happens, it shall be the case that one and only one of the listed outcomes will have occurred.

For instance, suppose that the experiment consists of someone jumping out of a plane somewhere over Ohio (parachute optional). Assuming that there is some way of defining the "patch upon which the jumper lands," it is possible to view this experiment as having infinitely many possible outcomes, corresponding to the infinitely many patches that might be landed upon. But we can collect these infinitely many possibilities into a finite number of different categories which are, one would think, much more interesting and useful to the jumper and everyone else concerned than are the undifferentiated infinity of fundamental possible outcomes. For instance, our finite list of possibilities might look like: (1) the jumper lands on some power line(s); (2) the jumper lands in a tree; . . . ; (n) none of the above (in case we overlooked a possibility).

Clearly there are infinitely many ways to make a finite list of outcomes of this experiment. How would you, in practice, choose a list? That depends on your concerns. If the jumper is a parachutist, items like "lands in water" should probably be on the list. If the jumper is a suicidal terrorist carrying an atom bomb, items like "lands within 15 miles of the center of Cincinnati" might well be on the list. There is some art in the science of parsing the outcomes to suit your interest. Never forget the constraint that one and only one of the listed outcomes will occur, whatever happens. For instance, it is unacceptable to have "lands in water" and "lands within 15 miles of the center of Cincinnati" on the same list, since it is possible for the jumper to land in water within 15 miles of the center of Cincinnati. (See Exercise 1 at the end of this section.)

Now consider the definition at the beginning of this section. The set $\mathcal{S}$ is interpretable as the finite list of outcomes, or of outcome categories, of whatever experiment we have at hand. The function P is, as the term *probability assignment* suggests, an assessment or measure of the likelihoods of the different outcomes.

There is nothing in the definition that tells you how to provide yourself with $\mathcal{S}$ and P, given some actual experiment. The jumper-from-the-airplane example is one of a multitude that show that there may be cause for debate and occasion for subtlety even in the task of listing the possible outcomes of the given experiment. And once the outcomes are listed to your satisfaction, how do you arrive at a satisfactory assessment of the likelihoods of occurrence of the various outcomes? That is a *long* story, only the beginnings of which will be told in this chapter. There are plenty of people—actuaries, pollsters, quality-control engineers, market analysts, and epidemiologists—who make their living partly by their sophistication in the matter of assigning probabilities; assigning probabilities in different situations is the problem at the center of applied statistics.

To the novice, it may be heartening to note that two great minds once got into a confused dispute over the analysis of a very simple experiment. Before the description of the experiment and the dispute, we interject a couple of comments that will be referred to throughout this chapter.

Experiments with outcomes of equal likelihood

If it is judged that the different outcomes are equally likely, then the condition $\sum_{s\in\mathcal{S}} P(s) = 1$ forces the probability assignment $P(s) = \frac{1}{|\mathcal{S}|}$ for all $s \in \mathcal{S}$, where $|\mathcal{S}|$ stands for the size of $\mathcal{S}$, also known as the number of elements in $\mathcal{S}$, also known as the cardinality of $\mathcal{S}$.

Coins. In this chapter, each coin shall have two sides, designated “heads” and “tails,” or H and T, for short. On each flip, toss, or throw of a coin, one of these two “comes up.” Sometimes H will be an abbreviation of the phrase “heads comes up,” and T similarly. Thus, in the experiment of tossing a coin once, the only reasonable set of outcomes is abbreviable $\{H, T\}$.

A *fair* coin is one for which the outcomes H and T of the one-toss experiment are equally likely—i.e., each has probability $1/2$.

The D’Alembert-Laplace controversy

D’Alembert and Laplace were great mathematicians of the 18th and 19th centuries. Here is the experiment about the analysis of which they disagreed: a fair coin is tossed twice.

The assumption that the coin is fair tells us all about the experiment of tossing it once. Tossing it twice is the next-simplest experiment we can perform with this coin. How can controversy arise? Consider the question: what is the probability that heads will come up on each toss? D'Alembert's answer: $1/3$. Laplace's answer: $1/4$.

D'Alembert and Laplace differed right off in their choices of sets of outcomes. D'Alembert took $\mathcal{S}_D = \{$both heads, both tails, one head and one tail$\}$, and Laplace favored $\mathcal{S}_L = \{HH, HT, TH, TT\}$ where, for instance, HT stands for "heads on the first toss, tails on the second." Both D'Alembert and Laplace asserted that the outcomes in their respective sets of outcomes are equally likely, from which assertions you can see how they got their answers. Neither provided a convincing justification of his assertion.

We will give a plausible justification of one of the two assertions above in Section 1.3. Whether or not the disagreement between D'Alembert and Laplace is settled by that justification will be left to your judgment.

Exercises 1.1

1. Someone jumps out of a plane over Ohio. You are concerned with whether or not the jumper lands in water, and whether or not the jumper lands within 15 miles of the center of Cincinnati, and with nothing else. [Perhaps the jumper carries a bomb that will not go off if the jumper lands in water, and you have relatives in Cincinnati.]

   Give an acceptable list of possible outcomes, as short as possible, that will permit discussion of your concerns. [Hint: the shortest possible list has length 4.]

2. Notice that we can get D'Alembert's set of outcomes from Laplace's by "amalgamating" a couple of Laplace's outcomes into a single outcome.

   More generally, given any set $\mathcal{S}$ of outcomes, you can make a new set of outcomes $\widehat{\mathcal{S}}$ by *partitioning* $\mathcal{S}$ into non-empty sets $P_1, \dots, P_m$ and setting $\widehat{\mathcal{S}} = \{P_1, \dots, P_m\}$. [To say that subsets $P_1, \dots, P_m$ of $\mathcal{S}$ *partition* $\mathcal{S}$ is to say that $P_1, \dots, P_m$ are pairwise disjoint, i.e., $\emptyset = P_i \cap P_j$, $1 \le i < j \le m$, and cover $\mathcal{S}$, i.e., $\mathcal{S} = \bigcup_{i=1}^m P_i$. Thus, "partitioning" is "dividing up into non-overlapping parts."]

   How may different sets of outcomes can be made, in this way, from a set of outcomes with four elements?

1.2 Events

Throughout, $(\mathcal{S}, P)$ will be some finite probability space. An *event* in this space is a subset of $\mathcal{S}$. If $E \subseteq \mathcal{S}$ is an event, the *probability of* E, denoted $P(E)$, is

$$P(E) = \sum_{s \in E} P(s).$$

Some elementary observations:

(i) if $s \in \mathcal{S}$, then $P(\{s\}) = P(s)$;

(ii) $P(\emptyset) = 0$;

(iii) $P(\mathcal{S}) = 1$;

(iv) if the outcomes in $\mathcal{S}$ are equally likely, then, for each $E \subseteq \mathcal{S}$, $P(E) = |E|/|\mathcal{S}|$.

Events are usually described in plain English, by a sentence or phrase indicating what happens when that event occurs. The *set* indicated by such a description consists of those outcomes that satisfy, or conform to, the description. For instance, suppose an urn contains red, green, and yellow balls; suppose that two are drawn, without replacement. We take $\mathcal{S} = \{rr, rg, ry, gr, gg, gy, yr, yg, yy\}$, in which, for instance, rg is short for "a red ball was drawn on the first draw, and a green on the second." Let $E =$ "no red ball was drawn." Then $E = \{gg, gy, yg, yy\}$.

Notice that the verbal description of an event need not refer to the set $\mathcal{S}$ of outcomes, and thus may be represented differently as a set of outcomes for different choices of $\mathcal{S}$. For instance, in the experiment of tossing a coin twice, the event "both heads and tails came up" is a set consisting of a single outcome in D'Alembert's way of looking at things, but consists of two outcomes according to Laplace. (Notice that the event "tails came up on first toss, heads on the second" is not an admissible event in D'Alembert's space; this does not mean that D'Alembert was *wrong*, only that his analysis is insufficiently fine to permit discussion of certain events associable with the experiment. Perhaps he would argue that distinguishing between the tosses, labeling one "the first" and the other "the second," makes a different experiment from the one he was concerned with.)

Skeptics can, and should, be alarmed by the "definition" above of the probability $P(E)$ of an event E. If an event E has a description that makes no reference to the set $\mathcal{S}$ of outcomes, then E should have a probability that does not vary as you consider different realizations of E as a subset of different outcome sets $\mathcal{S}$. Yet the probability of E is "defined" to be $\sum_{s \in E} P(s)$, which clearly involves $\mathcal{S}$ and P. This "definition" hides an assertion that deserves our scrutiny. The assertion is that, however an event E is realized as a subset of a set $\mathcal{S}$ of outcomes, if the probability assignment P to $\mathcal{S}$ is "correct," then the number $\sum_{s \in E} P(s)$ will be "correct," the "correct" probability of E by any "correct" assessment.

Here is an argument that seems to justify the equation $P(E) = \sum_{s \in E} P(s)$ in all cases where we agree that P is a correct assignment of probabilities to the elements of $\mathcal{S}$, and that there is a correct à priori probability $P(E)$ of the event E, realizable as a subset of $\mathcal{S}$. Let $\widehat{\mathcal{S}} = (\mathcal{S} \setminus E) \cup \{E\}$. That is, we are forming a new set of outcomes by amalgamating the outcomes in E into a single outcome, which we will denote by E. What shall the correct probability assignment $\widehat{P}$ to $\widehat{\mathcal{S}}$ be? $\widehat{P}(E)$ ought to be the sought-after $P(E)$, the correct probability of E. Meanwhile, the outcomes of $\mathcal{S} \setminus E$ are indifferent to our changed view of the experiment; we should have $P(s) = \widehat{P}(s)$ for $s \in \mathcal{S} \setminus E$. Then $1 = \sum_{s \in \widehat{S}} \widehat{P}(s) = \widehat{P}(E) + \sum_{s \in \mathcal{S} \setminus E} P(s) = P(E) + (1 - \sum_{s \in E} P(s))$, which implies the desired equation.

Definition Events E_1 and E_2 in a probability space $(\mathcal{S}, P)$ are *mutually exclusive* in case $P(E_1 \cap E_2) = 0$.

In common parlance, to say that two events are mutually exclusive is to say that they cannot *both* happen. Thus, it might seem reasonable to define E_1 and E_2 to be mutually exclusive if and only if $E_1 \cap E_2 = \emptyset$, a stronger condition than $P(E_1 \cap E_2) = 0$. It will be convenient to allow outcomes of experiments that have zero probability just because ruling out such outcomes may require a lengthy verbal digression or may spoil the symmetry of some array. In the service of this convenience, we define mutual exclusivity as above.

Example Suppose an urn contains a number of red and green balls, and exactly one yellow ball. Suppose that two balls are drawn, *without* replacement. If, as above, we take $\mathcal{S} = \{rr, rg, ry, gr, gg, gy, yr, yg, yy\}$, then the outcome yy is impossible. However we assign probabilities to $\mathcal{S}$, the only reasonable probability assignment to yy is zero. Thus, if $E_1 =$ "a yellow ball was chosen on the first draw" and $E_2 =$ "a yellow ball was chosen on the second draw," then E_1 and E_2 are mutually exclusive, even though $E_1 \cap E_2 = \{yy\} \neq \emptyset$.

Why not simply omit the impossible outcome yy from $\mathcal{S}$? We may, for some reason, be performing this experiment on different occasions with different urns, and, for most of these, yy may be a possible outcome. It is a great convenience to be able to refer to the same set $\mathcal{S}$ of outcomes in discussing these different experiments.

Some useful observations and results As heretofore, $(\mathcal{S}, P)$ will be a probability space, and E, F, E_1, E_2, etc., will stand for events in this space.

1.2.1 *If $E_1 \subseteq E_2$, then $P(E_1) \leq P(E_2)$.*

1.2.2 *If E_1 and E_2 are mutually exclusive, and $F_1 \subseteq E_1$ and $F_2 \subseteq E_2$, then F_1 and F_2 are mutually exclusive.*

1.2.3 *If $E_1, \ldots, E_m$ are pairwise mutually exclusive (meaning E_i and E_j are mutually exclusive when $1 \leq i < j \leq m$), then*

$$P(\bigcup_{i=1}^{m} E_i) = \sum_{i=1}^{m} P(E_i).$$

For a "clean" proof, go by induction on m. For an instructive proof, first consider the case in which $E_1, \ldots, E_m$ are pairwise disjoint.

1.2.4 *If $E \subseteq F \subseteq \mathcal{S}$, then $P(F \setminus E) = P(F) - P(E)$.*

Proof: Apply 1.2.3 with $m = 2$, $E_1 = E$ and $E_2 = F \setminus E$. □

1.2.5 $P(E \cup F) + P(E \cap F) = P(E) + P(F)$.

Proof: Observe that $E \cup F = (E \cap F) \cup (E \setminus (E \cap F)) \cup (F \setminus (E \cap F))$, a union of pairwise disjoint events. Apply 1.2.3 and 1.2.4. □

1.2.6 $P(E) + P(\mathcal{S} \setminus E) = 1$.

Proof: This is a corollary of 1.2.4. □

[When propositions are stated without proof, or when the proof is merely sketched, as in 1.2.4, it is hoped that the student will supply the details. The order in which the propositions are stated is intended to facilitate the verification. For instance, proposition 1.2.2 follows smoothly from 1.2.1.]

Example Suppose that, in a certain population, 40% of the people have red hair, 25% tuberculosis, and 15% have both. What percentage has neither?

The experiment that can be associated to this question is: choose a person "at random" from the population. The set $\mathcal{S}$ of outcomes can be identified with the population; the outcomes are equally likely if the selection process is indeed "random."

Let R stand for the event "the person selected has red hair," and T for the event "the person selected has tuberculosis." As subsets of the set of outcomes, R and T are the sets of people in the population which have red hair and tuberculosis, respectively. We are given that $P(R) = 40/100$, $P(T) = 25/100$, and $P(R \cap T) = 15/100$. Then

$$\begin{aligned} P(\mathcal{S} \setminus (R \cup T)) &= 1 - P(R \cup T) \qquad \text{[by 1.2.6]} \\ &= 1 - [P(R) + P(T) - P(R \cap T)] \qquad \text{[by 1.2.5]} \\ &= 1 - [\frac{40}{100} + \frac{25}{100} - \frac{15}{100}] = \frac{50}{100}. \end{aligned}$$

Answer: 50% have neither.

Exercises 1.2

1. In a certain population, 25% of the people are small and dirty, 35% are large and clean, and 60% are small. What percentage are dirty?
2. In the experiment of tossing a fair coin twice, let us consider Laplace's set of outcomes, $\{HH, HT, TH, TT\}$. We do not know how to assign probabilities to these outcomes as yet, but surely HT and TH ought to have the same probability, and the events "heads on the first toss" and "heads on the second toss" each ought to have probability $1/2$.

   Do these considerations determine a probability assignment?
3. In a certain population, 30% of the people have acne, 10% have bubonic plague, and 12% have cholera. In addition,

8% have acne and bubonic plague,
7% have acne and cholera,
4% have bubonic plague and cholera,
and 2% have all three diseases.

What percentage of the population has none of the three diseases?

1.3 Conditional probability

Definition Suppose that $(\mathcal{S}, P)$ is a finite probability space, $E_1, E_2 \subseteq \mathcal{S}$, and $P(E_2) \neq 0$. The *conditional probability* of E_1, *given* E_2, is $P(E_1 \mid E_2) = P(E_1 \cap E_2)/P(E_2)$.

Interpretation. You may as well imagine that you were not present when the experiment or action took place, and you received an incomplete report on what happened. You learn that event E_2 occurred (meaning the outcome was one of those in E_2), and nothing else. How shall you adjust your estimate of the probabilities of the various outcomes and events, in light of what you now know? The definition above proposes such an adjustment. Why this? Is it valid?

Justification. Supposing that E_2 has occurred, let's make a new probability space, $(E_2, \widehat{P})$, taking E_2 to be the new set of outcomes. What about the new probability assignment, $\widehat{P}$? We assume that the new probabilities $\widehat{P}(s)$, $s \in E_2$, are *proportional* to the old probabilities $P(s)$; that is, for some number r, we have $\widehat{P}(s) = rP(s)$ for all $s \in E_2$.

This might seem a reasonable assumption in many specific instances, but is it universally valid? Might not the knowledge that E_2 has occurred change our assessment of the relative likelihoods of the outcomes in E_2? If some outcome in E_2 was judged to be twice as likely as some other outcome in E_2 before the experiment was performed, must it continue to be judged twice as likely as the other after we learn that E_2 has occurred? We see no way to convincingly demonstrate the validity of this “proportionality” assumption, nor do we have in mind an example in which the assumption is clearly violated. We shall accept this assumption with qualms, and forge on.

Since $\widehat{P}$ is to be a probability assignment, we have

$$1 = \sum_{s \in E_2} \widehat{P}(s) = r \sum_{s \in E_2} P(s) = rP(E_2),$$

so $r = 1/P(E_2)$. Therefore, the probability that E_1 has occurred, given that E_2 has occurred, *ought* to be given by

$$P(E_1|E_2) = \widehat{P}(E_1 \cap E_2) = \sum_{s \in E_1 \cap E_2} \widehat{P}(s) = r \sum_{s \in E_1 \cap E_2} P(s) = \frac{P(E_1 \cap E_2)}{P(E_2)}.$$

End of justification.

Application to multi-stage experiments

Suppose that we have in mind an experiment with two stages, or sub-experiments. (Examples: tossing a coin twice, drawing two balls from an urn.) Let $x_1, \ldots, x_n$ denote the possible outcomes at the first stage, and $y_1, \ldots, y_m$ at the second. Suppose that the probabilities of the first-stage outcomes are known: say

$$P(\text{“}x_i \text{ occurs at the first stage”}) = p_i, \quad i = 1, \ldots, n.$$

Suppose that the probabilities of the y_j occurring are known, whenever it is known what happened at the first stage. Let us say that

$$P(\text{“}y_j \text{ occurs at the 2nd stage, supposing } x_i \text{ occurred at the first”}) = q_{ij}.$$

Now we will consider the full experiment. We take as the set of outcomes the set of ordered pairs

$$\begin{aligned}\mathcal{S} &= \{x_1, \ldots, x_n\} \times \{y_1, \ldots, y_m\}\\ &= \{(x_i, y_j); i \in \{1, \ldots, n\}, j \in \{1, \ldots, m\}\},\end{aligned}$$

in which (x_i, y_j) is short for the statement "x_i occurred at the first stage, and y_j at the second." What shall the probability assignment P be?

Let $E_i =$ "x_i occurred at the first stage" $= \{(x_i, y_1), \ldots, (x_i, y_m)\}$, for each $i \in \{1, \ldots, n\}$, and let $F_j =$ "y_j occurred at the second stage" $= \{(x_1, y_j), \ldots, (x_n, y_j)\}$, for each $j \in \{1, \ldots, m\}$. Even though we do not yet know what P is, we supposedly know something about the probabilities of these events. What we know is that $P(E_i) = p_i$, and $P(F_j \mid E_i) = q_{ij}$, for each i and j. Therefore,

$$q_{ij} = P(F_j \mid E_i) = \frac{P(E_i \cap F_j)}{P(E_i)} = \frac{P(\{(x_i, y_j)\})}{p_i}$$

which implies

$$P((x_i, y_j)) = P(\{(x_i, y_j)\}) = p_i q_{ij}.$$

We now know how to assign probabilities to outcomes of two stage experiments, given certain information about the experiment that might, plausibly, be obtainable a priori, from the description of the experiment. To put it simply, you *multiply*. (But *what* do you multiply?) Something similar applies to experiments of three or more stages. It is left to you to formulate what you do to assign probabilities to the outcomes of such experiments.

Examples

1.3.1 An urn contains 5 red, 12 green, and 8 yellow balls. Three are drawn without replacement.

(a) What is the probability that a red, a green, and a yellow ball will be drawn?

(b) What is the probability that the last ball to be drawn will be green?

Solution and discussion. If we identify the set of outcomes with the set of all sequences of length three of the letters r, g, and y, in the obvious way (e.g., ryy stands for a red ball was drawn first, then two yellows), then we will have 27 outcomes; no need to list them all. The event described in part (a) is

$$E = \{rgy, ryg, gry, gyr, yrg, ygr\}.$$

These are assigned probabilities, thus:

$$P(rgy) = \frac{5}{25} \cdot \frac{12}{24} \cdot \frac{8}{23}, \quad P(ryg) = \frac{5}{25} \cdot \frac{8}{24} \cdot \frac{12}{23},$$

etc. How are these obtained? Well, for instance, to see that $P(rgy) = \frac{5}{25} \cdot \frac{12}{24} \cdot \frac{8}{23}$, you reason thus: on the first draw, we have 25 balls, all equally likely to be drawn (it is presumed from the description of the experiment), of which 5 are red, hence a probability of $\frac{5}{25}$ of a red on the first draw; *having drawn* that red, there is then a $\frac{12}{24}$ probability of a green on the second draw, and *having drawn* first a red, then a green, there is probability $\frac{8}{23}$ of a yellow on the third draw. Multiply.

In this case, we observe that all six outcomes in E have the same probability assignment. Thus the answer to (a) is $6 \cdot \frac{5}{25} \cdot \frac{12}{24} \cdot \frac{8}{23} = \frac{24}{115}$.

For (b), we cleverly take a different set of outcomes, and ponder the event

$$F = \{NNg, Ngg, gNg, ggg\},$$

where N stands for "not green." We have

$$\begin{aligned} P(F) &= \frac{13}{25} \cdot \frac{12}{24} \cdot \frac{12}{23} + \frac{13}{25} \cdot \frac{12}{24} \cdot \frac{11}{23} + \frac{12}{25} \cdot \frac{13}{24} \cdot \frac{11}{23} + \frac{12}{25} \cdot \frac{11}{24} \cdot \frac{10}{23} \\ &= \frac{12}{25}\Big[\frac{1}{23 \cdot 24}(13 \cdot (12 + 11) + (13 + 10) \cdot 11)\Big] \\ &= \frac{12}{25}\Big[\frac{1}{23 \cdot 24} 23 \cdot 24\Big] = \frac{12}{25}, \end{aligned}$$

which is, interestingly, the probability of drawing a green on the first draw. Could we have foreseen the outcome of this calculation, and saved ourselves some trouble? It is left to you to decide whether or not the probability of drawing a green on draw number k, $1 \le k \le 25$, when we are drawing without replacement, might depend on k.

1.3.2 A room contains two urns, A and B. A contains nine red balls and one green ball; B contains four red balls and four green balls. The room is darkened, a man stumbles into it, gropes about for an urn, draws two balls without replacement, and leaves the room.

(a) What is the probability that both balls will be red?

(b) Suppose that one ball is red and one is green: what is the probability that urn A now contains only eight balls?

Solutions. (a) Using obvious and self-explanatory abbreviations,

$$P(\text{“both red”}) = P(Arr) + P(Brr) = \frac{1}{2}\frac{9}{10}\frac{8}{9} + \frac{1}{2}\frac{4}{8}\frac{3}{7}$$

(b) We calculate

$$P(\text{“Urn } A \text{ was the urn chosen”} \mid \text{“One ball is red and one green”})$$
$$= \frac{P(\{Arg, Agr\})}{P(\{Arg, Agr, Brg, Bgr\})} = \frac{\frac{1}{2}\frac{9}{10}\frac{1}{9} + \frac{1}{2}\frac{1}{10}\frac{9}{9}}{\frac{1}{2}\frac{9}{10}\frac{1}{9} + \frac{1}{2}\frac{1}{10}\frac{9}{9} + \frac{1}{2}\frac{4}{8}\frac{4}{7} + \frac{1}{2}\frac{4}{8}\frac{4}{7}} = \frac{7}{27}.$$

Notice that this result is satisfyingly less than $1/2$, the a priori probability that urn A was chosen.

Exercises 1.3

1. An urn contains six red balls, five green balls, and three yellow balls. Two are drawn without replacement. What is the probability that at least one is yellow?
2. Same question as in 1, except that *three* balls are drawn without replacement.
3. Same question as in 1, except that the drawing is *with* replacement.
4. An actuary figures that for a plane of a certain type, there is a 1 in 100,000 chance of a crash somewhere during a flight from New York to Chicago, and a 1 in 150,000 chance of a crash somewhere during a flight from Chicago to Los Angeles.

   A plane of that type is to attempt to fly from New York to Chicago and then from Chicago to Los Angeles.

   (a) What is the probability of a crash somewhere along the way? [Please do *not* use your calculator to convert to a decimal approximation.]
   (b) Suppose that you know that the plane crashed, but you know nothing else. What is the probability that the crash occurred during the Chicago-L.A. leg of the journey?
5. Urn A contains 11 red and seven green balls, urn B contains four red and one green, and urn C contains two red and six green balls. The three urns are placed in a dark room. Someone stumbles into the room, gropes around, finds an urn, draws a ball from it, lurches from the room, and looks at the ball. It is green. What is the probability that it was drawn from urn A?

   In this experiment, what is the probability that a red ball will be chosen? What is the proportion of red balls in the room?
6. Who was right, D'Alembert or Laplace? Or neither?
7. What is the probability of heads coming up exactly twice in three flips of a fair coin?

8. Let p_i and q_{ij} be as in the text preceding these exercises. What is the value of $\sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_{ij}$?
9. On each of three different occasions, a fair coin is flipped twice.
   (a) On the first occasion, you witness the second flip, but not the first. You see that heads comes up. What is the probability that heads came up on both flips?
   (b) On the second occasion, you are not present, but are shown a video of one of the flips, you know not which; either is as likely as the other. On the video, heads comes up. What is the probability that heads came up on both flips?
   (c) On the third occasion, you are not present; a so-called friend teases you with the following information, that heads came up at least once in the two flips. What is the probability that heads came up on both flips?

*10. For planes of a certain type, the actuarial estimate of the probability of a crash during a flight (including take-off and landing) from New York to Chicago is p_1; from Chicago to L.A., p_2.

In experiment #1, a plane of that type is trying to fly from New York to L.A., with a stop in Chicago. Let a denote the conditional probability that, if there is a crash, it occurs on the Chicago-L.A. leg of the journey.

In experiment #2, two different planes of the fatal type are involved; one is to fly from N.Y. to Chicago, the other from Chicago to L.A. Let b denote the conditional probability that the Chicago-L.A. plane crashed, if it is known that at least one of the two crashed, and let c denote the conditional probability that the Chicago-L.A. plane crashed, if it is known that exactly one of the two crashed.

Express a, b, and c in terms of p_1 and p_2, and show that $a \le c \le b$ for all possible p_1, p_2.

1.4 Independence

Definition Suppose that $(\mathcal{S}, P)$ is a finite probability space, and $E_1, E_2 \subseteq \mathcal{S}$. The events E_1, E_2 are *independent* if and only if

$$P(E_1 \cap E_2) = P(E_1)P(E_2).$$

1.4.1 *Suppose that both $P(E_1)$ and $P(E_2)$ are non-zero. The following are equivalent:*

(a) E_1 *and* E_2 *are independent;*

(b) $P(E_1 \mid E_2) = P(E_1)$;

(c) $P(E_2 \mid E_1) = P(E_2)$.

Proof: Left to you. $\square$

The intuitive meaning of independence should be fairly clear from the proposition; if the events have non-zero probability, then two events are independent if and only if the occurrence of either has no influence on the likelihood of the occurrence of the other. Besides saying that two events are independent, we will also say that one event is independent of another.

We shall say that two *stages*, say the ith and jth, $i \neq j$, of a multi-stage experiment, are independent if and only if for any outcome x at the ith stage, and any outcome y at the jth stage, the *events* "x occurred at the ith stage" and "y occurred at the jth stage" are independent. This means that no outcome at either stage will influence the outcome at the other stage. It is intuitively evident, and can be proven, that when two stages are independent, any two events whose descriptions involve only those stages, respectively, will be independent. See Exercises 2.2.5 and 2.2.6.

Exercises 1.4

1. Suppose that E and F are independent events in a finite probability space $(\mathcal{S}, P)$. Show that

   (a) E and $\mathcal{S} \setminus F$ are independent;
   (b) $\mathcal{S} \setminus E$ and F are independent;
   (c) $\mathcal{S} \setminus E$ and $\mathcal{S} \setminus F$ are independent.

2. Show that each event with probability 0 or 1 is independent of every event in its space.

3. Suppose that $\mathcal{S} = \{a, b, c, d\}$, all outcomes equally likely, and $E = \{a, b\}$. List all the events in this space that are independent of E.

4. Suppose that $\mathcal{S} = \{a, b, c, d, e\}$, all outcomes equally likely, and $E = \{a, b\}$. List all the events in this space that are independent of E.

5. Urn A contains 3 red balls and 1 green ball, and urn B contains no red balls and 75 green balls. The action will be: select one of urns A, B, or C at random (meaning they are equally likely to be selected), and draw one ball from it.

   How many balls of each color should be in urn C, if the event "a green ball is selected" is to be independent of "urn C is chosen," and urn C is to contain as few balls as possible?

6. Is it possible for two events to be both independent and mutually exclusive? If so, in what circumstances does this happen?

7. Suppose that $\mathcal{S} = \{a, b, c, d\}$, and these outcomes are equally likely. Suppose that $E = \{a, b\}$, $F = \{a, c\}$, and $G = \{b, c\}$. Verify that E, F, and G

are pairwise independent. Verify that

$$P(E \cap F \cap G) \neq P(E)P(F)P(G).$$

Draw a moral by completing this sentence: just because $E_1, \ldots, E_k$ are pairwise independent, it does not follow that ________________. What if $E_1, \ldots, E_k$ belong to the distinct, pairwise independent *stages* of a multistage experiment?

1.5 Bernoulli trials

Definition Suppose that $n \geq 0$ and k are integers; $\binom{n}{k}$, read as "n-choose-k" or "the binomial coefficient n, k," is the number of different k-subsets of an n-set (a set with n elements).

It is hoped that the reader is well-versed in the fundamentals of the binomial coefficients $\binom{n}{k}$, and that the following few facts constitute a mere review.

Note that $\binom{n}{k} = 0$ for $k > n$ and for $k < 0$, by the definition above, and that $\binom{n}{0} = \binom{n}{n} = 1$ and $\binom{n}{1} = n$ for all non-negative integers n. Some of those facts force us to adopt the convention that $0! = 1$, in what follows and forever after.

1.5.1 *For* $0 \leq k \leq n$, $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$. *If* $1 \leq k$, $\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!}$.

1.5.2 $\binom{n}{k} = \binom{n}{n-k}$.

1.5.3 $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$.

Definition An *alphabet* is just a non-empty finite set, and a *word of length n* over an alphabet A is a sequence, of length n, of elements of A, written without using parentheses and commas; for instance, 101 is a word of length 3 over $\{0, 1\}$.

1.5.4 *Suppose* α *and* β *are distinct symbols. Then* $\binom{n}{k} = |\{w;\, w$ *is a word of length* n *over* $\{\alpha, \beta\}$ *and* α *appears exactly* k *times in* $w\}|$.

1.5.5 *For any numbers* α, β,

$$(\alpha + \beta)^n = \sum_{k=0}^{n} \binom{n}{k} \alpha^k \beta^{n-k}.$$

1.5.6 $2^n = \sum_{k=0}^{n} \binom{n}{k}$.

Of the propositions above, 1.5.1 and 1.5.4 are the most fundamental, in that each of the others can be seen to follow from these two. However, 1.5.2 and 1.5.3 also have "combinatorial" proofs that appeal directly to the definition of $\binom{n}{k}$.

Proposition 1.5.5 is the famous binomial theorem of Pascal; it is from the role of the $\binom{n}{k}$ in this proposition that the term "binomial coefficient" arises.

Definitions A *Bernoulli trial* is an experiment with exactly two possible outcomes. A *sequence of independent Bernoulli trials* is a multi-stage experiment in which the stages are the same Bernoulli trial, and the stages are independent.

Tossing a coin is an example of a Bernoulli trial. So is drawing a ball from an urn, if we are distinguishing between only two types of balls. If the drawing is *with* replacement (and, it is understood, with mixing of the balls after replacement), a sequence of such drawings from an urn is a sequence of independent Bernoulli trials.

When speaking of some unspecified Bernoulli trial, we will call one possible outcome Success, or S, and the other Failure, or F. The distinction is arbitrary. For instance, in the Bernoulli trial consisting of a commercial DC-10 flight from New York to Chicago, you can let the outcome "plane crashes" correspond to the word Success in the theory, and "plane doesn't crash" to Failure, or the other way around.

Another way of saying that a sequence of Bernoulli trials (of the same type) is independent is: the probability of Success does not vary from trial to trial. Notice that if the probability of Success is p, then the probability of Failure is $1-p$.

1.5.7 Theorem *Suppose the probability of Success in a particular Bernoulli trial is p. Then the probability of exactly k successes in a sequence of n independent such trials is*

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

Proof: Let the set of outcomes of the experiment consisting of the sequence of n independent Bernoulli trials be identified with $\{S, F\}^n$, the set of all sequences of length n of the symbols S and F. If u is such a sequence in which S appears exactly k times, then, by what we know about assigning probabilities to the outcomes of multi-stage experiments, we have

$$P(u) = p^k (1-p)^{n-k}.$$

Therefore,

$$\begin{aligned} P(\text{“exactly } k \text{ successes”}) &= \sum_{\substack{S \text{ occurs exactly} \\ k \text{ times in } u}} P(u) \\ &= |\{u; S \text{ occurs exactly } k \text{ times in } u\}| p^k (1-p)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k}, \end{aligned}$$

by 1.5.4

In case $k < 0$ or $k > n$, when there are no such u, the truth of the theorem follows from the fact that $\binom{n}{k} = 0$. □

Example Suppose an urn contains 7 red and 10 green balls, and 20 balls are drawn with replacement (and mixing) after each draw. What is the probability that (a) exactly 4, or (b) at least 4, of the balls drawn will be red?
Answers:

(a) $\binom{20}{4}(\frac{7}{17})^4(\frac{10}{17})^{16} = \frac{20\cdot 19\cdot 18\cdot 17}{4\cdot 3\cdot 2}(\frac{7}{17})^4(\frac{10}{17})^{16}$.

(b)

$$\sum_{k=4}^{20}\binom{20}{k}(\frac{7}{17})^k(\frac{10}{17})^{20-k} = 1-\sum_{k=0}^{3}\binom{20}{k}(\frac{7}{17})^k(\frac{10}{17})^{20-k}$$
$$= 1-[(\frac{10}{17})^{20}+20(\frac{7}{17})(\frac{10}{17})^{19}+\frac{20\cdot 19}{2}(\frac{7}{17})^2(\frac{10}{17})^{18}$$
$$+\frac{20\cdot 19\cdot 18}{3\cdot 2}(\frac{7}{17})^3(\frac{10}{17})^{17}].$$

Observe that, in (b), the second expression for the probability is much more economical and evaluable than the first.

Exercises 1.5

1. An urn contains five red, seven green, and three yellow balls. Nine are drawn, *with* replacement. Find the probability that
   (a) exactly six of the balls drawn are green;
   (b) at least two of the balls drawn are yellow;
   (c) at most four of the balls drawn are red.
2. In eight tosses of a fair coin, find the probability that heads will come up
   (a) exactly three times;
   (b) at least three times;
   (c) at most three times.
3. Show that the probability of heads coming up exactly n times in $2n$ flips of a fair coin decreases with n.

*4. Find a simple representation of the polynomial $\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}$.

1.6 An elementary counting principle

1.6.1 *Suppose that, in a k-stage experiment, for each i, $1 \le i \le k$, whatever may have happened in stages preceding, there are exactly n_i outcomes possible at the ith stage. Then there are $\Pi_{i=1}^{k} n_i$ possible sequences of outcomes in the k stages.*

The proof can be done by induction on k. The word "preceding" in the statement above may seem to some to be too imprecise, and to others to be too precise, introducing an assumption about the order of the stages that need not be introduced. I shall leave the statement as it is, with "preceding" to be construed, in applications, as the applier deems wise.

The idea behind the wording is that the possible outcomes at the different stages may depend on the outcomes at other stages (those "preceding"), but whatever has happened at those other stages upon which the list of possible outcomes at the ith stage depends, the list of possible outcomes at the ith stage will always be of the same length, n_i.

For instance, with $k = 2$, suppose stage one is flipping a coin, and stage two is drawing a ball from one of two urns, RWB, which contains only red, white, and blue balls, and BGY, which contains only blue, green, and yellow balls. If the outcome at stage 1 is H, the ball is drawn from RWB; otherwise, from BGY. In this case, there are five possible outcomes at stage 2; yet $n_3 = 3$, and there are $2 \cdot 3 = 6$ possible sequences of outcomes of the experiment, namely HR, HW, HB, TB, TG, and TY, in abbreviated form. If, say, the second urn contained balls of only two different colors, then this two-stage experiment would not satisfy the hypothesis of 1.6.1.

In applying this counting principle, you think of a way to make, or construct, the objects you are trying to count. If you are lucky and clever, you will come up with a k-stage construction process satisfying the hypothesis of 1.6.1, and each object you are trying to count will result from exactly one sequence of outcomes or choices in the construction process. [But beware of situations in which the objects you want to count each arise from more than one construction sequence. See Exercise 4, below.]

For instance, to see that $|A_1 \times \cdots \times A_k| = \prod_{i=1}^{k} |A_i|$, when $A_1, \dots, A_k$ are sets, you think of making sequences $(a_1, \dots, a_k)$, with $a_i \in A_i$, $i = 1, \dots, k$, by the obvious process of first choosing a_1 from A_1, then a_2 from A_2, etc. For another instance, the number of different five-card hands dealable from a standard 52-card deck that are full houses is $13\binom{4}{3}12\binom{4}{2}$. [Why?]

Exercises 1.6

1. In the example above involving an urn, if the first urn contains balls of three different colors, and the second contains balls of two different colors, how many different possible sequences of outcomes are there in the two stages of the experiment?

2. How many different words of length ℓ are there over an alphabet with n letters? What does this have to do with 1.6.1?

3. In a certain collection of 23 balls, 15 are red and 8 are blue. How many different 7-subsets of the 23 balls are there, with 5 red and 2 blue balls?

4. How many different five-card hands dealable from a standard deck are "two-pair" hands?

1.7* On drawing without replacement

Suppose that an urn contains x red balls and y green balls. We draw n without replacement. What is the probability that exactly k will be red?

The set of outcomes of interest is identifiable with the set of all sequences, of length n, of the symbols r and g. We are interested in the event consisting of all such sequences in which r appears exactly k times.

We are not in the happy circumstances of Section 1.5, in which the probability of a red would be unchanging from draw to draw, but we do enjoy one piece of good fortune similar to something in that section: the different outcomes in the event "exactly k reds" all have the same probability. (Verify!) For $0 < k < n \le x + y$, that probability is

$$\frac{x(x-1)\cdots(x-k+1)y(y-1)\cdots(y-(n-k)+1)}{(x+y)(x+y-1)\cdots(x+y-n+1)} = \frac{x!\,y!\,(x+y-n)!}{(x-k)!\,(y-(n-k))!\,(x+y)!};$$

this last expression is only valid when $k \le x$ and $n-k \le y$, which are necessary conditions for the probability to be non-zero. Under those conditions, the last expression is valid for $k = 0$ and $k = n$, as well.

By the same reasoning as in the proof of the independent Bernoulli trial theorem, 1.5.7, invoking 1.5.4, we have that

$$\begin{aligned} &P(\text{“exactly } k \text{ reds”}) \\ &\quad = \binom{n}{k}\frac{x(x-1)\cdots(x-k+1)y(y-1)\cdots(y-(n-k)+1)}{(x+y)\cdots(x+y-n+1)} \\ &\quad = \binom{n}{k}\frac{x!\,y!\,(x+y-n)!}{(x-k)!\,(y-(n-k))!\,(x+y)!} \end{aligned}$$

for $0 < k < n$, provided we understand this last expression to be zero when $k > x$ or when $n - k > y$.

There is another way of looking at this experiment. Instead of drawing n balls one after the other, suppose you just reach into the urn and scoop up n balls all at once. Is this really different from drawing the balls one at a time, when you look at the final result? Supposing the two to be the same, we can take as the set of outcomes all n-subsets of the $x + y$ balls in the urn. Surely no n-subset is more likely than any other to be scooped up, so the outcomes are equiprobable, each with probability $1/\binom{x+y}{n}$ (provided $n \le x + y$). How many of these outcomes are in the event "exactly k reds are among the n balls selected?" Here is a two-stage method for forming such outcomes: first take a k-subset of the x red balls in the urn, then an $(n-k)$-subset of the y green balls (and put them together to make an n-set). Observe that different outcomes at either stage result in different n-sets, and that every n-set of these balls with exactly k reds is the result of one run of the two stages. By 1.6.1, it follows that

$$|\text{“exactly } k \text{ reds”}| = \binom{x}{k}\binom{y}{n-k}.$$

Thus $P(\text{“exactly } k \text{ reds”}) = \binom{x}{k}\binom{y}{n-k}/\binom{x+y}{n}$, provided $n \le x + y$.

Exercises 1.7

1. Verify that the two different ways of looking at drawing without replacement give the same probability for the event “exactly k balls are red.”
2. Suppose an urn contains 10 red, 13 green, and 4 yellow balls. Nine are drawn, without replacement.
   (a) What is the probability that exactly three of the nine balls drawn will be yellow?
   (b) Find the probability that there are four red, four green, and one yellow among the nine balls drawn.
3. Find the probability of being dealt a flush (all cards of the same suit) in a five-card poker game.

1.8 Random variables and expected, or average, value

Suppose that $(\mathcal{S}, P)$ is a finite probability space. A *random variable* on this space is a function from $\mathcal{S}$ into the real numbers, $\mathbb{R}$. If $X : \mathcal{S} \to \mathbb{R}$ is a random variable on $(\mathcal{S}, P)$, the *expected* or *average value* of X is

$$E(X) = \sum_{u \in \mathcal{S}} X(u)P(u).$$

Random variables are commonly denoted X, Y, Z, or with subscripts: X_1, $X_2, \cdots$. It is sometimes useful, as in 1.8.2, to think of a random variable as a measurement on the outcomes, and the average value as the average of the measurements. The average value of a random variable X is sometimes denoted $\overline{X}$; you may recall that a bar over a letter connotes the arithmetic average, in elementary statistics.

The word “expected” in “expected value” requires interpretation. As numerous examples show (see 1.8.1, below), a random variable which takes only integer values can have a non-integer expected value. The point is that the expected value is not necessarily a possible value of the random variable, and thus is not really necessarily to be “expected” in any running of the experiment to which $(\mathcal{S}, P)$ is associated.

Examples

1.8.1 The experiment consists of n independent Bernoulli trials, with probability p of success on each trial. Let $X =$ "number of successes." Finding $E(X)$ in these circumstances is one of our goals; for now we'll have to be content with a special case.

Sub-example: $n = 3$, $p = 1/2$; let the experiment be flipping a fair coin three times, and let "success" be "heads." With $X =$ "number of heads," we have

$$\begin{aligned} E(X) &= X(HHH)P(HHH) + X(HHT)P(HHT) \\ &\quad + \cdots + X(TTT)P(TTT) \\ &= \frac{1}{8}[3+2+2+1+2+1+1+0] = \frac{12}{8} = \frac{3}{2}. \end{aligned}$$

1.8.2 We have some finite population (voters in a city, chickens on a farm) and some measurement M that can be applied to each individual of the population (blood pressure, weight, hat size, length of femur, ...). Once units are decided upon, $M(s)$ is a pure real number for each individual s. We can regard M as a random variable on the probability space associated with the experiment of choosing a member of the population "at random." Here $\mathcal{S}$ is the population, the outcomes are equiprobable (so P is the constant assignment $1/|\mathcal{S}|$), and M assigns to each outcome (individual of the population) whatever real number measures that individual's M-measure. If $\mathcal{S} = \{s_1, \ldots, s_n\}$, then $E(M) = (M(s_1) + \cdots + M(s_n))/n$, the arithmetic average, or mean, of the measurements $M(s), s \in \mathcal{S}$.

1.8.3 Let an urn contain 8 red and 11 green balls. Four are drawn *without* replacement. Let X be the number of green balls drawn. What is $E(X)$?

Note that we have at our disposal two different views of this experiment. In one view, the set of outcomes is the set of ordered quadruples of the symbols r and g; i.e., $\mathcal{S} = \{rrrr, rrrg, \ldots, gggg\}$. In the other view, an *outcome* is a four-subset of the 19 balls in the urn.

Taking the first view, we have

$$\begin{aligned} E(X) &= 0 \cdot \frac{8\cdot 7\cdot 6\cdot 5}{19\cdot 18\cdot 17\cdot 16} + 1\cdot 4\cdot \frac{8\cdot 7\cdot 6\cdot 11}{19\cdot 18\cdot 17\cdot 16} \\ &\quad + 2\cdot \binom{4}{2}\frac{8\cdot 7\cdot 11\cdot 10}{19\cdot 18\cdot 17\cdot 16} + 3\cdot 4\frac{8\cdot 11\cdot 10\cdot 9}{19\cdot 18\cdot 17\cdot 16} \\ &\quad + 4\cdot \frac{11\cdot 10\cdot 9\cdot 8}{19\cdot 18\cdot 17\cdot 16} \end{aligned}$$

Verify: $E(X) = 4\cdot\frac{11}{19}$. Verify: the same answer is obtained if the other view of this experiment is taken. [Use 1.8.4, below, to express $E(X)$.]

1.8.4 Theorem *Suppose X is a random variable on $(\mathcal{S}, P)$. Then $E(X) = \sum_{x\in\mathbb{R}} xP(X = x)$.*

Here, "$X = x$" is the event $\{u \in \mathcal{S}; X(u) = x\} = X^{-1}(\{x\})$. Note that $X^{-1}(\{x\}) = \emptyset$ unless $x \in \operatorname{ran}(X)$. Since $\mathcal{S}$ is finite, so is $\operatorname{ran}(X)$. If $\operatorname{ran}(X) \subseteq \{x_1, \ldots, x_r\}$, and $x_1, \ldots, x_r$ are distinct, then

$$\sum_{x \in \mathbb{R}} x P(X = x) = \sum_{k=1}^{r} x_k P(X = x_k).$$

Proof: Let $x_1, \ldots, x_r$ be the possible values that X might take; i.e., $\operatorname{ran}(X) \subseteq \{x_1, \ldots, x_r\}$. Let, for $k \in \{1, \ldots, r\}$, $E_k =$ "$X = x_k$" $= X^{-1}(\{x_k\})$. Then $E_1, \ldots, E_r$ partition $\mathcal{S}$. Therefore,

$$\begin{aligned} E(X) = \sum_{u \in \mathcal{S}} X(u) P(u) &= \sum_{k=1}^{r} \sum_{u \in E_k} X(u) P(u) \\ &= \sum_{k=1}^{r} x_k \sum_{u \in E_k} P(u) = \sum_{k=1}^{r} x_k P(E_k). \end{aligned}$$ □

1.8.5 Corollary *The expected or average number of successes, in a run of n independent Bernoulli trials with probability p of success in each, is*

$$\sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k}.$$

Proof: The possible values that $X =$ "number of successes" might take are $0, 1, \ldots, n$, and $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, for $k \in \{0, 1, \ldots, n\}$, by 1.5.7. □

1.8.6 Theorem *If $X_1, \ldots, X_n$ are random variables on $(\mathcal{S}, P)$, and $a_1, \ldots, a_n$ are real numbers, then $E(\sum_{k=1}^{n} a_k X_k) = \sum_{k=1}^{n} a_k E(X_k)$.*

Proof:

$$\begin{aligned} E\Big(\sum_{k=1}^{n} a_k X_k\Big) &= \sum_{u \in \mathcal{S}} \Big(\sum_{k=1}^{n} a_k X_k\Big)(u) P(u) \\ &= \sum_{u \in \mathcal{S}} \sum_{k=1}^{n} a_k X_k(u) P(u) \\ &= \sum_{k=1}^{n} a_k \sum_{u \in \mathcal{S}} X_k(u) P(u) = \sum_{k=1}^{n} a_k E(X_k). \end{aligned}$$ □

1.8.7 Corollary *The expected, or average, number of successes, in a run of n independent Bernoulli trials with probability p of success in each, is np.*

Proof: First, let $n = 1$. Let X be the "number of successes." By definition and by the hypothesis,

$$E(X) = 0 \cdot (1-p) + 1 \cdot p = 1 \cdot p.$$

Now suppose that $n > 1$. Let $X =$ "number of successes." For $k \in \{1, \ldots, n\}$, let $X_k =$ "number of successes on the kth trial." By the case already done, we have that $E(X_k) = p, k = 1, \ldots, n$. Therefore,

$$E(X) = E(\sum_{k=1}^{n} X_k) = \sum_{k=1}^{n} E(X_k) = \sum_{k=1}^{n} p = np. \qquad \square$$

1.8.8 Corollary *For $0 \le p \le 1$, and any positive integer n,*

$$\sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np.$$

Proof: This follows from 1.8.5 and 1.8.7, provided there exists, for each $p \in [0, 1]$, a Bernoulli trial with probability p of success. You are invited to ponder the existence of such trials (problem 1 at the end of this section).

Alternatively, using a sharper argument based on an elementary theorem about polynomials, the conclusion here (and the result called for in problem 2 at the end of this section) follows from 1.8.5 and 1.8.7 and the existence of such trials for only $n+1$ distinct values of p. Therefore, the result follows from the existence of such trials for rational numbers p between 0 and 1. If $p = s/b$, where s and b are positive integers and $s < b$, let the Bernoulli trial with probability p of Success be: draw a ball from an urn containing s Success balls and $b - s$ Failure balls. $\square$

1.8.9 Corollary *Suppose k balls are drawn, without replacement, from an urn containing x red balls and y green balls, where $1 \le k \le x + y$. The expected number of red balls to be drawn is $kx/(x + y)$.*

Proof: It is left to the reader to see that the probability of a red being drawn on the ith draw, $1 \le i \le k$, is the same as the probability of a red being drawn on the first draw, namely $x/(x + y)$. [Before the drawing starts, the various $x + y$ balls are equally likely to be drawn on the ith draw, and x of them are red.] Once this is agreed to, the proof follows the lines of that of 1.8.7. $\square$

Exercises 1.8

1. Suppose $0 \le p \le 1$. Describe a Bernoulli trial with probability p of Success. You may suppose that there is a way of "picking a number at random" from a given interval.
2. Suppose that n is a positive integer. Find a simple representation of the polynomial $\sum_{k=1}^{n} k \binom{n}{k} x^k (1-x)^{n-k}$.

*3. State a result that is to 1.8.9 as 1.8.8 is to 1.8.7.

4. An urn contains 4 red and 13 green balls. Five are drawn. What is the expected number of reds to be drawn if the drawing is

   (a) with replacement?

(b) without replacement?

5. For any radioactive substance, if the material is not tightly packed, it is thought that whether or not any one atom decays is independent of whether or not any other atom decays. [We are all aware, I hope, that this assumption fails, in a big way, when the atoms are tightly packed together.]

   For a certain mystery substance, let p denote the probability that any particular atom of the substance will decay during any particular 6-hour period at the start of which the atom is undecayed. A handful of this substance is left in a laboratory, in spread-out, unpacked condition, for 24 hours. At the end of the 24 hours it is found that approximately $1/10$ of the substance has decayed. Find p, approximately. [Hint: let n be the number of atoms, and think of the process of leaving the substance for 24 hours as n independent Bernoulli trials. Note that p here is not the probability of Success, whichever of two possibilities you choose to be Success, but that probability is expressible in terms of p. Assume that the amount of substance that decayed was approximately the "expected" amount.]

6. An actuary reckons that for any given year that you start alive, you have a 1 in 6,000 chance of dying during that year.

   You are going to buy \$100,000 worth of five-year term life insurance. However you pay for the insurance, let us define the payment to be *fair* if the life insurance company's *expected gain* from the transaction is *zero*.

   (a) In payment plan number one, you pay a single premium at the beginning of the five-year-coverage period. Assuming the actuary's estimate is correct, what is the fair value of this premium?
   (b) In payment plan number two, you pay five equal premiums, one at the beginning of each year of the coverage, provided you are alive to make the payment. Assuming the actuary's estimate is correct, what is the fair value of this premium?

7. (a) Find the expected total showing on the upper faces of two fair dice after a throw.
   (b) Same question as in (a), except use backgammon scoring, in which doubles count quadruple. For instance, double threes count 12.

1.9 The Law of Large Numbers

1.9.1 Theorem *Suppose that, for a certain Bernoulli trial, the probability of Success is p. For a sequence of n independent such trials, let X_n be the random variable "number of Successes in the n trials." Suppose $\epsilon > 0$. Then*

$$P\Big(\Big|\frac{X_n}{n} - p\Big| < \epsilon\Big) \to 1 \text{ as } n \to \infty.$$

For instance, if you toss a fair coin 10 times, the probability that $|\frac{X_{10}}{10} - \frac{1}{2}| < \frac{1}{100}$, i.e., that the number of heads, say, is between 4.9 and 5.1, is just the probability that heads came up exactly 5 times, which is $\binom{10}{5}\frac{1}{2^{10}} = \frac{63}{256}$, not very close to 1. Suppose the same coin is tossed a thousand times; the probability that $|\frac{X_{1000}}{1000} - \frac{1}{2}| < \frac{1}{100}$ is the probability that between 490 and 510 heads appear in the thousand flips. By approximation methods that we will not go into here, this probability can be shown to be approximately 0.45. The probability of heads coming up between 4900 and 5100 times in 10,000 flips is around 0.95.

For a large number of independent Bernoulli trials of the same species, it is plausible that the *proportion* of Successes "ought" to be near p, the probability of Success on each trial. The Law of Large Numbers, 1.9.1, gives precise form to this plausibility. This theorem has a purely mathematical proof that will not be given here.

Theorem 1.9.1 is not the only way of stating that the proportion of Successes will tend, with high probability, to be around p, the probability of Success on each trial. Indeed, Feller [18] maligns this theorem as the weakest and least interesting of the various laws of large numbers available. Still, 1.9.1 is the best-known law of large numbers, and 'twill serve our purpose.

Exercises 1.9

*1. With n, ϵ, p, and X_n as in 1.9.1, express $P(|\frac{X_n}{n} - p| \le \epsilon)$ explicitly as a sum of terms of the form $\binom{n}{k}p^k(1-p)^{n-k}$. You will need the symbols $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$, which stand for "round up" and "round down," respectively.

2. An urn contains three red and seven green balls. Twenty are drawn, with replacement. What is the probability of exactly six reds being drawn? Of five, six, or seven reds being drawn?

3. An urn contains an unknown number of red balls, and 10 balls total. You draw 100 balls, with replacement; 42 are red. What is your best guess as to the number of red balls in the urn?

*4. A pollster asks some question of 100 people, and finds that 42 of the 100 give "favorable" responses. The pollster estimates from this result that (probably) between 40 and 45 percent of the total population would be "favorable" on this question, if asked.

   Ponder the similarities and differences between this inference and that in problem 3. Which is more closely analogous to the inference you used in doing Exercise 1.8.5?