

Chapter 1

Basic Concepts

In this chapter we will introduce the basic terminology of probability theory. The notions of independence, distribution, and expected value will be studied in more detail later, but it is hard to discuss examples without them, so we introduce them quickly here.

1.1 Outcomes, Events, Probability

The subject of probability can be traced back to the 17th century when it arose out of the study of gambling games. As we will see, the range of applications extends beyond games into business decisions, insurance, law, medical tests, and the social sciences. The stock market, “the largest casino in the world,” cannot do without it. The telephone network, call centers, and airline companies with their randomly fluctuating loads could not have been economically designed without probability theory. To quote Pierre Simon, Marquis de Laplace from several hundred years ago:

“It is remarkable that this science, which originated in the consideration of games of chance, should become the most important object of human knowledge . . . The most important questions of life are, for the most part, really only problems of probability.”

In order to address these applications, we need to develop a language for discussing them. Here and throughout the book **bold face type** indicates a term that is being defined. An **experiment** is an activity or procedure that produces distinct, well-defined possibilities called **outcomes**.

Example 1.1. If our experiment is to roll one die, then there are six outcomes corresponding to the number that shows on the top. The set of all outcomes in this case is $\{1, 2, 3, 4, 5, 6\}$. It is called the **sample space** and is usually denoted by Ω (capital Omega). Symmetry dictates that all outcomes are equally likely so each has probability $1/6$.

Example 1.2. Things get a little more interesting when we roll two dice. If we suppose, for convenience, that they are red and green, then we can write the outcomes of this experiment as (m, n) , where m is the number on the red die and n is the number on the green die. To visualize the set of outcomes it is useful to make a little table:

(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

There are $36 = 6 \cdot 6$ outcomes since there are 6 possible numbers to write in the first slot and for each number written in the first slot there are 6 possibilities for the second.

The goal of probability theory is to compute the probability of various events of interest. Intuitively, an event is a statement about the outcome of an experiment. The formal definition is: An **event** is a subset of the sample space. For example, “the sum of the two dice is 8” translates into the set $A = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$. Since this event contains 5 of the 36 possible outcomes its probability $P(A) = 5/36$.

For a second example, consider $B =$ “there is at least one six.” B consists of the last row and last column of the table, so it contains 11 outcomes and hence has probability $P(B) = 11/36$. In general the probability of an event C concerning the roll of two dice is the number of outcomes in C divided by 36.

Axioms of Probability Theory

Let \emptyset be the **empty set**, i.e., the event with no outcomes. We assume that the reader is familiar with the basic concepts of set theory such as **union** ($A \cup B$ the outcomes in either A or B) and **intersection** ($A \cap B$, the outcomes in both A and B).

Abstractly, a **probability** is a function that assigns numbers to events, which satisfies the following assumptions:

- (i) For any event A , $0 \leq P(A) \leq 1$.
- (ii) If Ω is the sample space then $P(\Omega) = 1$.
- (iii) If A and B are **disjoint**, i.e., the intersection $A \cap B = \emptyset$ then

$$P(A \cup B) = P(A) + P(B)$$

- (iv) If A_1, A_2, \dots is an infinite sequence of **pairwise disjoint events** (i.e., $A_i \cap A_j = \emptyset$ when $i \neq j$) then

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

These assumptions are motivated by the **frequency interpretation of probability**, which states that if we repeat an experiment a large number of times then the fraction of times the event A occurs will be close to $P(A)$. To be precise, if we let $N(A, n)$ be the number of times A occurs in the first n trials then

$$P(A) = \lim_{n \rightarrow \infty} \frac{N(A, n)}{n} \quad (1.1)$$

In Chapter 6 we will see this result is a theorem called the **law of large numbers**. For the moment, we will use this interpretation of $P(A)$ to motivate the definition.

Given (1.1), properties (i) and (ii) are clear: the fraction of times that a given event A occurs must be between 0 and 1, and if Ω has been defined properly (recall that it is the set of ALL possible outcomes), the fraction of times something in Ω happens is 1. To explain (iii), note that if the events A and B are disjoint then

$$N(A \cup B, n) = N(A, n) + N(B, n)$$

since $A \cup B$ occurs if either A or B occurs but it is impossible for both to happen. Dividing by n and letting $n \rightarrow \infty$, we arrive at (iii).

Property (iii) implies that (iv) holds for a finite number of events, but for infinitely many events, the last argument breaks down, and this is a new assumption. Not everyone believes that assumption (iv) should be used. However, without (iv) the theory of probability becomes much more difficult and less useful, so we will impose this assumption and not apologize further for it. In many cases the sample space is finite so (iv) is not relevant anyway.

Example 1.3. Suppose we pick a letter at random from the word TENNESSEE. What is the sample space Ω and what probabilities should be assigned to the outcomes?

The sample space $\Omega = \{T, E, N, S\}$. To describe the probability it is enough to give the values for the individual outcomes, since (iii) implies that $P(A)$ is the sum of the probabilities of the outcomes in A . Since there are nine letters in TENNESSEE the probabilities are $P(\{T\}) = 1/9$, $P(\{E\}) = 4/9$, $P(\{N\}) = 2/9$, and $P(\{S\}) = 2/9$.

Having introduced a number of definitions, we will now derive some basic properties of probabilities and illustrate their use.

Property 1. Let A^c be the **complement** of A , i.e., the set of outcomes not in A , then

$$P(A) = 1 - P(A^c) \quad (1.2)$$

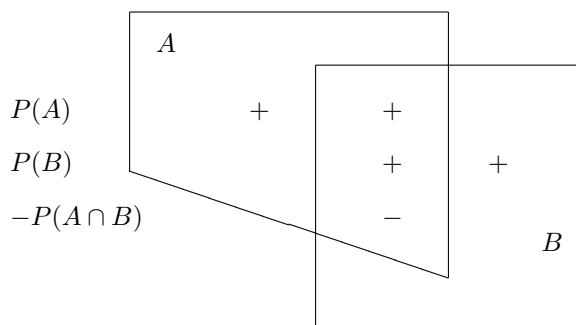
Proof. Let $A_1 = A$ and $A_2 = A^c$. Then $A_1 \cap A_2 = \emptyset$ and $A_1 \cup A_2 = \Omega$ so (iii) implies $P(A) + P(A^c) = P(\Omega) = 1$ by (ii). Subtracting $P(A)$ from each side of the equation gives the result. \square

This formula is useful because sometimes it is easier to compute the probability of A^c . For an example, consider $A =$ “at least one six.” In this case $A^c =$ “no six.” There are $5 \cdot 5$ outcomes with no six, so $P(A^c) = 25/36$ and $P(A) = 1 - 25/36 = 11/36$, as we computed before.

Property 2. For any events A and B ,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (1.3)$$

Proof by picture. We can explain this result by drawing a picture:



Intuitively, $P(A) + P(B)$ counts $A \cap B$ twice so we have to subtract $P(A \cap B)$ to make the net number of times $A \cap B$ is counted equal to 1. \square

Proof. To prove this result we note that by assumption (ii)

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \\ P(B) &= P(B \cap A) + P(B \cap A^c) \end{aligned}$$

Adding the two equations and subtracting $P(A \cap B)$:

$$P(A) + P(B) - P(A \cap B) = P(A \cap B) + P(A \cap B^c) + P(B \cap A^c) = P(A \cup B)$$

which gives the desired equality. \square

To illustrate property 2, let $A =$ “the red die shows six,” and $B =$ “the green die shows six.” In this case $A \cup B =$ “at least one 6” and $A \cap B = \{(6, 6)\}$, so we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36}$$

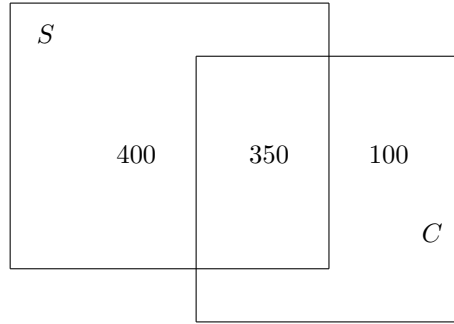
The same principle applies to counting outcomes in events.

Example 1.4. A survey of 1000 students revealed that 750 owned stereos, 450 owned cars, and 350 owned both. How many own either a car or a stereo?

Given a set A , we use $|A|$ to denote the number of points in A . The reasoning that led to (1.3) tells us that

$$|S \cup C| = |S| + |C| - |S \cap C| = 750 + 450 - 350 = 850$$

We can confirm this by drawing a picture:



Property 3. *Monotonicity.* If $A \subset B$, i.e., any outcome in A is also in B , then

$$P(A) \leq P(B) \quad (1.4)$$

Proof. A and $A^c \cap B$ are disjoint, with union B , so assumption (iii) implies $P(B) = P(A) + P(A^c \cap B) \geq P(A)$ by (i). \square

We write $A_n \uparrow A$ if $A_1 \subset A_2 \subset \dots$ and $\cup_{i=1}^{\infty} A_i = A$. We write $A_n \downarrow A$ if $A_1 \supset A_2 \supset \dots$ and $\cap_{i=1}^{\infty} A_i = A$.

Property 4. *Monotone limits.* If $A_n \uparrow A$ or $A_n \downarrow A$ then

$$\lim_{n \rightarrow \infty} P(A_n) = P(A) \quad (1.5)$$

Proof. Let $B_1 = A_1$ and for $i \geq 2$, let $B_i = A_i \cap A_{i-1}^c$. The events B_i are disjoint, with $\cup_{i=1}^{\infty} B_i = A$, so (iv) implies

$$P(A) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(B_i) = \lim_{n \rightarrow \infty} P(A_n)$$

by (iii) since B_i , $1 \leq i \leq n$, are disjoint and their union is A_n .

To prove the second result, let $B_i = A_i^c$. We have $B_n \uparrow A^c$ so by (1.5) and (1.2), $\lim_{n \rightarrow \infty} P(B_n) = 1 - P(A)$. Since $P(B_n) = 1 - P(A_n)$ the desired result follows. \square

1.2 Flipping coins, the World Series

Even simpler than rolling a die is flipping a coin, which produces one of two outcomes, called “Heads” (H) or “Tails” (T). If we flip two coins there are four outcomes

		HT		
	HH	TH	TT	
heads	2	1	0	
probability	1/4	1/2	1/4	

Flipping three coins there are eight possibilities:

		HHT	TTH		
	HHH	HTH	THT	TTT	
		THH	HTT		
heads	3	2	1	0	
probability	1/8	3/8	3/8	1/8	

Our next problem concerns flipping four to seven coins:

Example 1.5. World Series. In this baseball event, the first team to win four games wins the championship. Obviously, the series may last 4, 5, 6, or 7 games. However, a fan who wants to buy a ticket would like to know what are the probabilities of each of these outcomes.

Ignoring potential complicating factors like the advantage of playing at home or psychological factors that make the outcome of one game affect the next one, we suppose that the games are decided by tossing a fair coin to determine if team A or team B wins.

Four games. There are two possible ways this can happen: A wins all four games or B wins all four games. There are $2 \cdot 2 \cdot 2 \cdot 2 = 16$ possible outcomes and these are 2 of them so $P(4) = 2/16 = 1/8$.

Five games. Here and in the next case we will compute the probability that A wins in the specified number of games and then multiply by 2. There are four possible outcomes

$$BAAAA, ABAAA, AABAA, AAABA$$

$AAAAB$ is not possible since in that case the series would have ended in four games. There are $2^5 = 32$ outcomes so $P(5) = 2 \cdot 4/32 = 1/4$.

Six games. In the next section we will learn systematic ways of doing this, but for now we will compute the probabilities by enumerating the possibilities:

$$\begin{array}{l} BBAAAA \quad ABBAAB \quad AABBAAB \quad AAABBA \\ BABAAA \quad ABABAA \quad AABABA \\ BAABAA \quad ABAABA \\ BAAABA \end{array}$$

The first column corresponds to outcomes in which B wins the first game, the second one to outcomes in which the first game B wins is the second game, etc. We then move the remaining win for B through its possibilities. There are 10 outcomes out of $2^6 = 64$ total, so remembering to multiply by 2 to account for the ways B can win in six games, $P(6) = 2 \cdot 10/64$.

Seven games. The analysis from the previous case becomes even messier here, so we instead observe that the probabilities for the four possible outcomes must add up to one, so

$$P(7) = 1 - P(4) - P(5) - P(6) = 1 - \frac{2}{16} - \frac{4}{16} - \frac{5}{16} = \frac{5}{16}$$

As mentioned earlier, we are ignoring things that many fans think are important to determining the outcomes of the games, so our next step is to compare the probabilities just calculated with the observed frequencies of the duration of best of seven series in three sports. The numbers in parentheses give the number of series in our sample.

Games	4	5	6	7
Probabilities	0.125	0.25	0.3125	0.3125
World Series (94)	0.181	0.224	0.224	0.372
Stanley Cup (74)	0.270	0.216	0.230	0.284
NBA finals (57)	0.122	0.228	0.386	0.263

To determine whether or not the data agrees with predictions, statisticians use a **chi-squared statistic**:

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

where o_i is the number of observations in category i and e_i is what the model predicts. The details of the test are beyond the scope of this book so we just quote the results: the Stanley Cup data is very unusual (the probability of a chi-squared score this large or larger has probability $p < 0.01$) due to the larger than expected number of four game series. The World Series data does not fit the model well, but is not very unusual ($p > 0.05$). On the other hand, the NBA finals data looks like what we should expect to see. The excess of six game series can be due just to chance.

Example 1.6. Birthday problem. There are 30 people at a party. Someone wants to bet you \$10 that there are two people with exactly the same birthday. Should you take the bet?

To pose a mathematical problem, we ignore Feb. 29 which only comes in leap years, and suppose that each person at the party picks their birthday at random from the calendar. There are 365^{30} possible outcomes for that experiment. The number of outcomes in which all the birthdays are different is

$$365 \cdot 364 \cdot 363 \cdots 336$$

since the second person must avoid the first person's birthday, the third the first two birthdays and so on until the 30th person must avoid the 29 previous birthdays. Let D be the event that all birthdays are different. Dividing the number of outcomes in which all the birthdays are different by the total number of outcomes, we have

$$P(D) = \frac{365 \cdot 364 \cdot 363 \cdots 336}{365^{30}} = 0.293684$$

In words, only about 29% of the time all the birthdays are different, so you will lose the bet 71% of the time.

At first glance it is surprising that the probability of two people having the same birthday is so large, since there are only 30 people compared with 365 days on the calendar. Some of the surprise disappears if you realize that there are $(30 \cdot 29)/2 = 435$ pairs of people who are going to compare their birthdays. Let p_k be the probability that k people all have different birthdays. Clearly, $p_1 = 1$ and $p_{k+1} = p_k(365 - k)/365$. Using this recursion it is easy to generate the values of p_k for $1 \leq k \leq 40$.

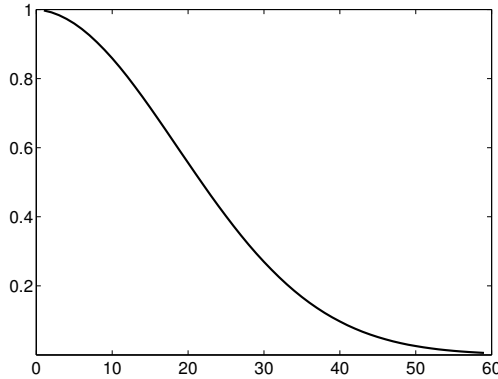


Figure 1.1: Birthday coincidence probability as a function of group size.

The graph shows the trends, but to get precise values a table is better:

1	1.00000	11	0.85886	21	0.55631	31	0.26955
2	0.99726	12	0.83298	22	0.52430	32	0.24665
3	0.99180	13	0.80559	23	0.49270	33	0.22503
4	0.98364	14	0.77690	24	0.46166	34	0.20468
5	0.97286	15	0.74710	25	0.43130	35	0.18562
6	0.95954	16	0.71640	26	0.40176	36	0.16782
7	0.94376	17	0.68499	27	0.37314	37	0.15127
8	0.92566	18	0.65309	28	0.34554	38	0.13593
9	0.90538	19	0.62088	29	0.31903	39	0.12178
10	0.88305	20	0.58856	30	0.29368	40	0.10877

A graph is less precise but conveys more qualitative information:

1.3 Independence

Intuitively, two events A and B are independent if the occurrence of A has no influence on the probability of occurrence of B . The formal definition is: A and B are **independent** if

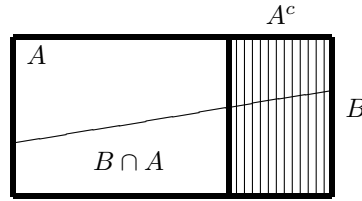
$$P(A \cap B) = P(A)P(B)$$

To make the connection between the two definitions, we need to introduce the notion of conditional probability, which will be discussed in more detail in Chapter 3.

Suppose we are told that the event A with $P(A) > 0$ occurs. Then the sample space is reduced from Ω to A and the probability that B will occur given that A has occurred is

$$P(B|A) = P(B \cap A)/P(A) \quad (1.6)$$

To explain this formula, note that (i) only the part of B that lies in A can possibly occur, and (ii) since the sample space is now A , we have to divide by $P(A)$ to make $P(A|A) = 1$.



Suppose A and B are independent. In this case $P(A \cap B) = P(A)P(B)$, so

$$P(B|A) = \frac{P(A)P(B)}{P(A)} = P(B)$$

or in the words of the intuitive definition of independence, “the occurrence of A has no influence on the probability of the occurrence of B .”

Turning to concrete examples, in each case it should be clear that the intuitive definition is satisfied, so we will only check the formal one.

- Flip two coins. Let A = “the first coin shows Heads,” and B = “the second coin shows Heads.” $P(A) = 1/2$, $P(B) = 1/2$, $P(A \cap B) = 1/4$.
- Roll two dice. Let A = “the first die shows 5,” and B = “the second die shows 2.” $P(A) = 1/6$, $P(B) = 1/6$, $P(A \cap B) = 1/36$.
- Pick a card from a standard deck of 52 cards. Let A = “the card is an ace,” and B = “the card is a spade.” $P(A) = 1/13$, $P(B) = 1/4$, $P(A \cap B) = 1/52$.

Two examples of events that are not independent are

Example 1.7. Draw two cards from a deck. Let A = “the first card is a spade,” and B = “the second card is a spade.” Then $P(A) = 1/4$ and $P(B) = 1/4$, but

$$P(A \cap B) = \frac{13 \cdot 12}{52 \cdot 51} < \left(\frac{1}{4}\right)^2$$

To derive the answer, we note that there are $52 \cdot 51$ outcomes for drawing two cards from the deck when keep track of the order they are drawn, while there are $13 \cdot 12$ outcomes that result in two spades. A second approach is to note that we have a probability of $P(A) = 13/52$ of getting a spade the first time and, if we succeed then the probability on the second draw is $P(B|A) = 12/51$. Rearranging the definition of conditional probability in (1.6)

$$P(A \cap B) = P(A)P(B|A)$$

Thus, these two events are not independent, since getting a spade the first time reduces the fraction of spades in the deck and makes it harder to get a spade the second time.

Example 1.8. Roll two dice. Let A = “the sum of the two dice is 9,” and B = “the first die is 2.” $A = \{(6, 3), (5, 4), (4, 5), (3, 6)\}$, so $P(A) = 4/36$. $P(B) = 1/6$, but $P(A \cap B) = 0$ since $(2, 7)$ is impossible.

In general, if A and B are disjoint events that have positive probability, they are not independent since $P(A)P(B) > 0 = P(A \cap B)$.

There are two ways of extending the definition of independence to more than two events. A_1, \dots, A_n are said to be **pairwise independent** if for each $i \neq j$, $P(A_i \cap A_j) = P(A_i)P(A_j)$, that is, each pair is independent. A_1, \dots, A_n are said to be **independent** if for any $1 \leq i_1 < i_2 < \dots < i_k \leq n$ we have

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$$

If we flip n coins and let A_i = “the i th coin shows Heads,” then the events A_i are independent since $P(A_i) = 1/2$ and $P(A_{i_1} \cap \dots \cap A_{i_k}) = 1/2^k$. We have already seen an example of events that are pairwise independent but not independent:

Example 1.9. Let A = “Alice and Betty have the same birthday,” B = “Betty and Carol have the same birthday,” C = “Carol and Alice have the same birthday.” Each pair of events is independent but the three are not.

Since there are 365 ways two girls can have the same birthday out of 365^2 possibilities (as in Example 1.6, we are assuming that leap year does not exist and that all the birthdays are equally likely), $P(A) = P(B) = P(C) = 1/365$. Likewise, there are 365 ways all three girls can have the same birthday out of 365^3 possibilities, so

$$P(A \cap B) = \frac{1}{365^2} = P(A)P(B)$$

i.e., A and B are independent. Similarly, B and C are independent and C and A are independent, so A , B , and C are pairwise independent. The three events A , B , and C are not independent, however, since $A \cap B = A \cap B \cap C$ and hence

$$P(A \cap B \cap C) = \frac{1}{365^2} \neq \left(\frac{1}{365}\right)^3 = P(A)P(B)P(C)$$

Example 1.10. Roll three dice. Let A = “the numbers on the first and second add to 7,” B = “the numbers on the second and third add to 7,” C = “the numbers on the third and first add to 7.” Again each pair of events is independent but the three are not.

To check that A and B are independent, note that no matter what the values i and j are on the first two dice there is probability $1/6$ that the third die is $7 - j$. Similar arguments show that B and C are independent and A and C are independent.

To show that the three events are not independent we note that $A \cap B \cap C = \emptyset$. Let i, j, k be the values for the three dice. We claim that if A and B occur then $i + k$ is even so C is impossible. To check this note that if j is odd then i and k are even, while if j is even, i and k are odd.

The last two examples are somewhat unusual. However, the moral of the story is that to show several events are independent, you have to check more than just that each pair is independent.

1.4 Distributions

A **random variable** is a numerical value determined by the outcome of an experiment. Four examples:

- Roll two dice and let $X =$ the sum of the two numbers that appear.
- Roll a die until a 4 appears and let $X =$ the number of rolls we need.
- Flip a coin 10 times and let $X =$ the number of Heads we get.
- Draw 13 cards out of a deck and let $X =$ the number of Hearts we get.

In these cases X is a **discrete random variable**. That is, there is a finite or countable sequence of possible values. In contrast, the height of a randomly chosen person, or the time they spent waiting for the bus this morning are **continuous random variables**, since the value could be any positive real number. In this book we will primarily be concerned with discrete random variables. Continuous random variables will not make an appearance until Chapter 5.

The **distribution** of a discrete random variable, is described by giving the value of $P(X = x)$ for all values of x . In each case, we will only give the values of $P(X = x)$ when $P(X = x) > 0$. The other values we do not mention are 0. We begin with the first two examples above.

Example 1.11. Roll two dice. Let $X =$ the sum of the two numbers that appear.

(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

Using the table of outcomes, it is easy to see:

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

To check this, note that the outcomes with a given sum are diagonal lines in the square. For example, the four outcomes in bold are the ones for which the sum is 5.

Example 1.12. Geometric distribution. If we repeat an experiment with probability p of success until a success occurs then the number of trials required N has

$$P(N = n) = (1 - p)^{n-1}p \quad \text{for } n = 1, 2, \dots$$

In words, N has a geometric distribution with parameter p , a phrase we will abbreviate as $N = \text{geometric}(p)$.

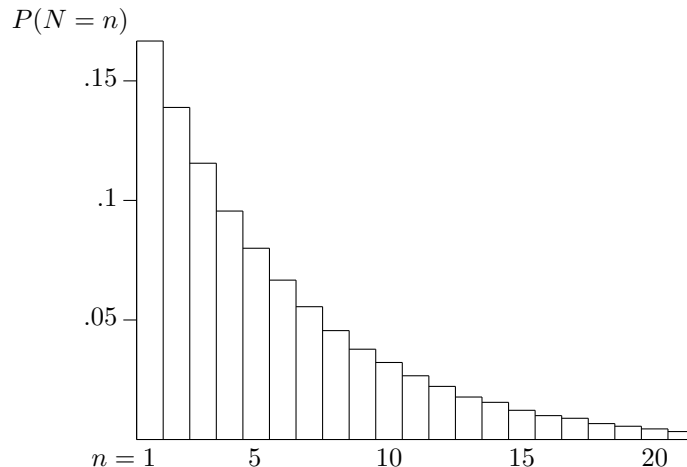


Figure 1.2: Geometric(1/6) distribution

To check the formula note that in order to first have success on trial n , we must have $n - 1$ failures followed by a success, which has probability $(1 - p)^{n-1}p$. In the example at the beginning of the section, success is rolling a 4, so $p = 1/6$. The next graph gives a picture of the distribution in this case.

For an example of the use of the geometric distribution, we consider

Example 1.13. Birthday problem, II. How large must the group be so that there is a probability > 0.5 that someone will have the same birthday as you do?

In our first encounter with the birthday problem it was surprising that the size needed to have two people with the same birthday was so small. This time the surprise goes in the other direction. Assuming 365 equally likely birthdays, a naive guess is that 183 people will be enough. However in a group of n people the probability all will fail to have your birthday is $(364/365)^n$. Setting this equal to 0.5 and solving,

$$n = \frac{\log(0.5)}{\log(364/365)} = \frac{-0.69314}{-0.0027435} = 252.7$$

So we need 253 people. The “problem” is that many people in the group will have the same birthday, so the number of different birthdays is smaller than the size of the group.

Example 1.14. Astragali. Board games involving chance were known in Egypt, 3000 years B.C. The element of chance needed for these games was at

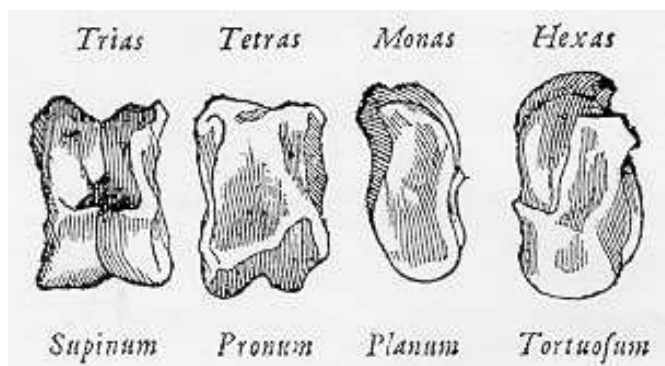


Figure 1.3: Astragali

first provided by tossing astragali, the ankle bones of sheep. These bones could come to rest on only four sides, the other two sides being rounded. The upper side of the bone, broad and slightly convex, counted four (Tetras); the opposite side, broad and slightly concave counted three (Trias); the lateral side, flat and narrow, one (Monas), and the opposite narrow lateral side, which is slightly hollow, six (Hexas).

The outcomes of this experiment are $\Omega = \{1, 3, 4, 6\}$. There is no reason to suppose that all four sides have the same probability, so our model will have probabilities for the four outcomes $p_1, p_3, p_4, p_6 \geq 0$ that have $p_1 + p_3 + p_4 + p_6 = 1$. To define the probability of an event A we let

$$P(A) = \sum_{i \in A} p_i$$

In words, we add up the probabilities of the outcomes in A . With a little thought we see that any probability with a finite set of outcomes has this form.

Example 1.15. English letter frequencies. In text written in English the 26 letters in the alphabet occur with the following frequencies

E	13.0%	H	3.5%	W	1.6%
T	9.3%	L	3.5%	V	1.3%
N	7.8%	C	3.0%	B	0.9%
R	7.7%	F	2.8%	X	0.5%
O	7.4%	P	2.7%	K	0.3%
I	7.4%	U	2.7%	Q	0.3%
A	7.3%	M	2.5%	J	0.2%
S	6.3%	Y	1.9%	Z	0.1%
D	4.4%	G	1.6%		

From this it follows that vowels (A,E,I,O,U) are used $(7.3+13.0+7.4+7.4+2.7) = 37.8\%$ of the time.

Example 1.16. Scrabble. In this game, there are 100 tiles with the following distribution. The first number after the letter is its point value, the second is the number of tiles.

E	1	12	U	1	4	V	4	2
A	1	9	D	2	4	W	4	2
I	1	9	G	2	3	Y	4	2
O	1	8	B	3	2	K	5	1
N	1	6	C	3	2	J	8	1
R	1	6	M	3	2	X	8	1
T	1	6	P	3	2	Q	10	1
L	1	4	F	4	2	Z	10	1
S	1	4	H	4	2	blank	0	2

In Scrabble, vowels are $12 + 9 + 9 + 8 + 4 = 42\%$ of the letters. The number of points on a randomly chosen tile has the following distribution:

0	1	2	3	4	5	8	10
.02	.68	.07	.08	.10	.01	.02	.02

Example 1.17. Benford's law is named after the late Dr. Frank Benford, a physicist at General Electric Company. In 1938 he noticed that pages of logarithms corresponding to numbers starting with 1 were dirtier and more worn than other pages. He examined 20,229 data sets, and developed a prediction about the number observed distribution of first digits:

$$p_k = \log_{10}(k+1) - \log_{10}(k) \quad \text{for } k = 1, 2, \dots, 9$$

This is a probability distribution because $p_k \geq 0$ and $\sum_{k=1}^9 p_k = \log_{10}(10) - \log_{10}(1)$. The numerical values of the probabilities are:

1	2	3	4	5	6	7	8	9
.3010	.1761	.1249	.0969	.0792	.0669	.0580	.0512	.0458

Some of the many examples that are supposed to follow Benford's law are: census populations of 3259 counties, 308 numbers from Reader's digest, areas of 335 rivers, 342 addresses of *American Men of Science*. The next table compares the percentages of the observations in the first five categories to Benford's law:

	1	2	3	4	5
Census	33.9	20.4	14.2	8.1	7.2
Reader's Digest	33.4	18.5	12.4	7.5	7.1
Rivers	31.0	16.4	10.7	11.3	7.2
Benford's Law	30.1	17.6	12.5	9.7	7.9
Addresses	28.9	19.2	12.6	8.8	8.5

The fits are far from perfect, but in each case Benford's law matches the general shape of the observed distribution.

Why should Benford's law hold? It is mathematical fact that the first digits of 2^n follow Benford's law. This can be established by showing that the fractional parts of $\log_{10}(2^m)$ for $1 \leq m \leq n$ (i.e., the part of the number to the right of the decimal point) are approximately uniformly distributed on $(0,1)$.

There are two good general explanations. (i) If the Dow Jones average or some other statistic grows at rate r , i.e., $f(t) = Ce^{rt}$, then the amount of time it takes for the first digit to change from k to $k + 1$ is $(\ln(k + 1) - \ln(k))/r$. When we divide by the total time to go from 1 to 10, the result is Benford's law. (ii) Ted Hill proved (see *Statistical Science* 10, 354–363) that if the first digit distribution does not depend on the units used then Benford's law is the only possibility.

1.5 Expected Value

The **expected value** of X , or **mean** of X is defined to be

$$EX = \sum_x xP(X = x) \quad (1.7)$$

In words, we multiply each possible value by its probability and sum.

Example 1.18. Roulette. If you play roulette and bet \$1 on black then you win \$1 with probability $18/38$ and you lose \$1 with probability $20/38$, so the expected value of your winnings X is

$$EX = 1 \cdot \frac{18}{38} + (-1) \cdot \frac{20}{38} = \frac{-2}{38} = -0.0526$$

The expected value has a meaning much like the frequency interpretation of probability. Suppose X_1, \dots, X_n are independent and have the same distribution as X , that is,

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X = x_1) \cdots P(X = x_n)$$

then, when n is large, the average of the values we have observed, $(X_1 + \cdots + X_n)/n$, will be close to EX with high probability. This result is called the **law of large numbers**, and will be proved in Chapter 6.

In the roulette example, if we let X_i be your winnings on the i th play then this says that $(X_1 + \cdots + X_n)/n$ will be close to -0.0526 . In words, in the long run you will lose about 5.26 cents per play.

Example 1.19. Roll one die. Let X be the number that appears on the die. $P(X = x) = 1/6$ for $x = 1, 2, 3, 4, 5, 6$, so

$$EX = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3\frac{1}{2}$$

In this case the expected value is just the average of the six possible values.

To deal with more than one die we use the following fact, which will be proved in Chapter 6.

Theorem 1.1. *If X_1, \dots, X_n are random variables then*

$$E(X_1 + \cdots + X_n) = EX_1 + \cdots + EX_n \quad (1.8)$$

From this and Example 1.19, it follows that if we roll two dice the expected value of the sum is 7. As another example, let $X_i = 1$ if the i th flip of a coin is heads and 0 otherwise. $X_1 + \cdots + X_n$ is the number of heads in n tosses. Since $EX_i = 1/2$ for a fair coin, the last result implies that

$$E(X_1 + \cdots + X_n) = nEX_i = n/2$$

Example 1.20. Scrabble. As we computed in Example 1.16, the point value of a randomly chosen scrabble tile has the following distribution:

value	0	1	2	3	4	5	8	10
prob.	.02	.68	.07	.08	.10	.01	.02	.02

The expected value is

$$= .68 + .14 + .24 + .4 + .05 + .16 + .2 = 1.87$$

This means that when we draw seven letters to start the game the average numbers of points on our rack will be $7(1.87) = 13.09$. Of course, on any play of the game the number of points may be more or less than 13. The law of large numbers implies that if we kept records for a large number of games then the average we have seen on our first draws will be close to the expected value.

Example 1.21. Fair division of a bet on an interrupted game. Pascal and Fermat were sitting in a café in Paris playing the simplest of all games, flipping a coin. Each had put up a bet of 50 francs and the first to get 10 points wins. Fermat was winning 8 points to 7 when he received an urgent message that a friend was sick, and he must rush to his home town of Toulouse. The carriage man who delivered the message offered to take him, but only if he would leave immediately. Later in correspondence between the two men, the problem arose: how should the money bet (100 francs) be divided?

Fermat came up with the reasonable idea that the fraction of the stakes that each receives should be equal to the probability it would have won the match. In the case under consideration, it is easier to calculate the probability that Pascal (P) wins. In order for Pascal to win by a score of 10-8 he must win three flips in a row: PPP , an event of probability $1/8$. To win 10-9 he can lose one flip but not the last one: $FPFP$, $PFPP$, $PPFP$, which has probability $3/16$. Adding the two we see that Pascal wins with probability $5/16$ and should get that fraction of the total bet, i.e., $(5/16)(100) = 31.25$.

Example 1.22. Deal or No Deal. In this TV game show there are 26 briefcases with amounts of money indicated in the first, third, and fifth columns of the next table. You pick one briefcase then pick five others to open. At that point they offer you an amount of money. If you take it the game ends. If not then you open more briefcases. The numbers in the second, fourth, and sixth columns are the rounds on which I opened those briefcases when I played the game online at nbc.com:

0.01	2	300	3	75,000	4
1	1	400	4	100,000	2
5	2	500	3	200,000	
10		750	3	300,000	2
25	2	1000	9	400,000	5
50	4	5000	1	500,000	3
75	5	10,000	1	750,000	7
100	8	25,000	6	1,000,000	1
200	1	50,000	1		

The expected value of the money in the briefcase you pick is 131,477. The next table gives the offers I got from the online game compared with the expected value after each of the rounds.

1	25,866	117,660
2	35,158	122,074
3	30,492	120,970
4	46,446	152,910
5	48,806	162,685
6	64,675	190,222
7	21,620	50,277
8	40,872	67,003
9	62,003	100,005

Notice that in all cases the offer is considerably less than the expected value. After the ninth round there are two briefcases left, one with 200,000 and one with 10. If I were playing for real I might have taken the 62,003 for sure but I stuck with the higher expected value and won 200,000.

Example 1.23. Geometric distribution. When $P(N = n) = (1 - p)^{n-1}p$, for $n = 1, 2, 3, \dots$ we have $EN = 1/p$.

This answer is intuitive. We have a probability p of success on each trial, so in n trials we have an average of np successes and if we want $np = 1$ we need $n = 1/p$. To get this from the definition, we begin with the sum of the geometric series

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$$

and differentiate with respect to x to get

$$\sum_{k=0}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$$

Dropping the $k = 0$ term from the left since it is 0 and setting $x = 1 - p$

$$\sum_{k=1}^{\infty} k(1-p)^{k-1} = \frac{1}{p^2}$$

Multiplying each side by p we have

$$\sum_{k=1}^{\infty} kP(N = k) = \frac{1}{p}$$

Example 1.24. China's one child policy. In order to limit the growth of its population, the Chinese government decided to limit a family to having just one child. An alternative that was suggested was the "one son" policy: as long as a woman has only female children she is allowed to have more children.

One concern voiced about this policy was that no family would have more than one son, but many families would have several girls. This concern leads to our question: How would the one son policy affect the ratio of male to female births?

To simplify the problem we assume that a family will keep having children until it has a male child. Assuming that male and female children are equally likely and the sexes of successive children are independent, the total number of children has a geometric distribution with success probability $p = 1/2$, so by the previous example the expected number of children is $1/p = 2$. There is always one male child, so the expected number of female children is $2 - 1 = 1$.

Does this continue to hold if some families stop before they have a male child? Consider for simplicity the case in which a family will stop when they have a male child or a total of three children. There are four outcomes

$$\begin{aligned}P(M) &= 1/2 \\P(FM) &= 1/4 \\P(FFM) &= 1/8 \\P(FFF) &= 1/8\end{aligned}$$

The average number of male children is $1/2 + 1/4 + 1/8 = 7/8$ while the average number of female children is $1(1/4) + 2(1/8) + 3(1/8) = 7/8$.

The last calculation makes the equality of the expected values look like a miracle, but it is not, and the claim holds true if a family with k female children continues with probability p_k and stops with probability $1 - p_k$. To explain this intuitively, if we replace M by $+1$ and F by -1 , then childbirth is a fair game. For the stopping rules under consideration the average winnings when we stop have mean 0, i.e., the expected number of male children equals the expected number of female children.

1.6 Moments, Variance

In this section we will be interested in the expected values of various functions of random variables. The most important of these are the variance and the standard deviation which give an idea about how spread out the distribution is. The first basic fact we need in order to do computations is

Theorem 1.2. *If X has a discrete distribution and $Y = r(X)$ then*

$$EY = \sum_x r(x)P(X = x) \quad (1.9)$$

Proof. $P(Y = y) = \sum_{x:r(x)=y} P(X = x)$. Multiplying both sides by y and summing gives

$$\begin{aligned} EY &= \sum_y y P(Y = y) = \sum_y y \sum_{x:r(x)=y} P(X = x) \\ &= \sum_y \sum_{x:r(x)=y} r(x)P(X = x) = \sum_x r(x)P(X = x) \end{aligned}$$

since the double sum is just a clumsy way of summing over all the possible values of x . \square

If $r(x) = x^k$, $E(X^k)$ is the k th **moment** of X . When $k = 1$ this is the first moment or mean of X .

Example 1.25. Suppose X is the result of rolling one die. Compute EX^2 .

$$EX^2 = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6} = 15.1666$$

To prepare for our next topic we need the following properties:

$$E(X + b) = EX + b \quad E(aX) = aEX \quad (1.10)$$

In words, if we add 5 to a random variable then we add 5 to its expected value. If we multiply a random variable by 3 we multiply its expected value by 3.

Proof. For the first one, we note that

$$\begin{aligned} E(X + b) &= \sum_x (x + b) P(X = x) \\ &= \sum_x xP(X = x) + \sum_x bP(X = x) = EX + b \end{aligned}$$

The second is easier:

$$E(aX) = a \sum_x xP(X = x) = aEX \quad \square$$

If $EX^2 < \infty$ then the **variance** of X is defined to be

$$\text{var}(X) = E(X - EX)^2$$

To illustrate this concept, we will consider some examples. But first, we need a formula that enables us to compute $\text{var}(X)$ more easily.

$$\text{var}(X) = EX^2 - (EX)^2 \quad (1.11)$$

Proof. Letting $\mu = EX$ to make the computations easier to see, we have

$$\text{var}(X) = E(X - \mu)^2 = E\{X^2 - 2\mu X + \mu^2\} = EX^2 - 2\mu EX + \mu^2$$

by (1.10) and the facts that $E(-2\mu X) = -2\mu EX$, $E(\mu^2) = \mu^2$. Substituting $\mu = EX$ now gives the result. \square

The reader should note that EX^2 means the expected value of X^2 and in the proof $E(X - \mu)^2$ means the expected value of $(X - \mu)^2$. When we want the square of the expected value we will write $(EX)^2$. This convention is designed to cut down on parentheses.

The variance measures how spread-out the distribution of X is. To begin to explain this statement, we will show that

$$\text{var}(X + b) = \text{var}(X) \quad \text{var}(aX) = a^2 \text{var}(X) \quad (1.12)$$

In words, the variance is not changed by adding a constant to X , but multiplying X by a multiplies the variance by a^2 .

Proof. If $Y = X + b$ then the mean of Y , $\mu_Y = \mu_X + b$ by (1.10), so

$$\text{var}(X + b) = E\{(X + b) - (\mu_X + b)\}^2 = E\{X - \mu_X\}^2 = \text{var}(X)$$

If $Y = aX$ then $\mu_Y = a\mu_X$ by (1.10), so

$$\text{var}(aX) = E\{(aX - a\mu_X)^2\} = a^2 E\{X - \mu_X\}^2 = a^2 \text{var}(X) \quad \square$$

The scaling relationship (1.12) shows that if X is measured in feet then the variance is measured in feet². This motivates the definition of the **standard deviation** $\sigma(X) = \sqrt{\text{var}(X)}$, which is measured in the same units as X and has a nicer scaling property:

$$\sigma(aX) = |a|\sigma(X) \quad (1.13)$$

We get the absolute value here since $\sqrt{a^2} = |a|$.

Example 1.26. Roll one die and let X be the resulting number. Find the variance and standard deviation of X .

Examples 1.19 and 1.25 tell us that $EX = 7/2$ and $EX^2 = 91/6$, so

$$\text{var}(X) = EX^2 - (EX)^2 = \frac{91}{6} - \frac{49}{4} = \frac{105}{36} = 2.9166$$

and $\sigma(X) = \sqrt{\text{var}(X)} = 1.7078$. The standard deviation $\sigma(X)$ gives the size of the “typical deviation from the mean.” To explain this, we note that the deviation from the mean

$$|X - \mu| = \begin{cases} 0.5 & \text{when } X = 3, 4 \\ 1.5 & \text{when } X = 2, 5 \\ 2.5 & \text{when } X = 1, 6 \end{cases}$$

so $E|X - \mu| = 1.5$. The standard deviation $\sigma(X) = \sqrt{E|X - \mu|^2}$ is a slightly less intuitive way of averaging the deviations $|X - \mu|$ but, as we will see later, is one that has nicer properties.

Example 1.27. Scrabble. As we computed in Example 1.16, the point value of a randomly chosen scrabble tile has the following distribution:

value	0	1	2	3	4	5	8	10
prob.	.02	.68	.07	.08	.10	.01	.02	.02

In Example 1.20 we have computed that $EX = 1.87$.

$$EX^2 = .68 + 4(.07) + 9(.08) + 16(.10) + 25(.01) + 64(.02) + 100(.02) = 9.06$$

so the variance is $9.06 - (1.87)^2 = 5.5631$ and the standard deviation is $\sqrt{5.5631} = 2.35$.

Example 1.28. Geometric distribution. Suppose $N = \text{geometric}(p)$. That is, $P(N = n) = (1 - p)^{n-1}p$ for $n = 1, 2, \dots$ and 0 otherwise. Compute the variance and standard deviation of N .

To compute the variance, we begin, as in Example 1.23 by observing that

$$\sum_{n=0}^{\infty} x^n = (1 - x)^{-1}$$

Differentiating this identity twice and noticing that the $n = 0$ term in the first derivative is 0 gives

$$\sum_{n=1}^{\infty} nx^{n-1} = (1 - x)^{-2} \quad \sum_{n=1}^{\infty} n(n-1)x^{n-2} = 2(1 - x)^{-3}$$

Setting $x = 1 - p$ gives

$$\sum_{n=1}^{\infty} n(1 - p)^{n-1} = p^{-2} \quad \sum_{n=1}^{\infty} n(n-1)(1 - p)^{n-2} = 2p^{-3}$$

Multiplying both sides by p in the first case and $p(1-p)$ in the second, we have

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

$$E\{N(N-1)\} = \sum_{n=1}^{\infty} n(n-1)(1-p)^{n-1}p = 2p^{-2}(1-p)$$

From this it follows that

$$EN^2 = E\{N(N-1)\} + EN = \frac{2-2p}{p^2} + \frac{1}{p} = \frac{(2-p)}{p^2}$$

$$\text{var}(N) = EN^2 - (EN)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{(1-p)}{p^2}$$

Taking the square root we see that $\sigma(X) = \sqrt{1-p}/p$.

Example 1.29. New York Yankees 2004 salaries. Salaries are in units of M, millions of dollars per year and, for convenience, have been truncated at the thousands place.

A. Rodriguez	21.726	D. Jeter	18.6
M. Mussina	16	K. Brown	15.714
J. Giambi	12.428	B. Williams	12.357
G. Sheffield	12.029	M. Rivera	10.89
J. Posada	9	J. Vazquez	9
J. Contreras	9	J. Olerud	7.7
H. Matsui	7	S. Karsay	6
E. Loazia	4	T. Gordon	3.5
P. Quantrill	3	K. Lofton	2.985
J. Lieber	2.7	T. Lee	2
G. White	1.925	F. Heredia	1.8
R. Sierra	1	M. Cairo	.9
J. Falherty	.775	T. Clark	.75
E. Wilson	.7	O. Hernandez	.5
D. Osborne	.45	C.J. Nitowski	.35
J. DePaula	.302	B. Crosby	.301

The total team salary is 183.355 M. Dividing by 32 players gives a mean of 6.149 M dollars. The second moment is 73.778 M² so the variance is 73.778 - (6.149)² = 35.961M² and the standard deviation is 5.996 M.

1.7 Exercises

Basic definitions

1. A man receives presents from his three children, Allison, Betty, and Chelsea. To avoid disputes he opens the presents in a random order. What are the possible outcomes?
2. Suppose we pick a number at random from the phone book and look at the last digit. (a) What is the set of outcomes and what probability should be assigned to each outcome? (b) Would this model be appropriate if we were looking at the first digit?
3. Two students arrive late for a math final exam with the excuse that their car had a flat tire. Suspicious, the professor says “each one of you write down on a piece of paper which tire was flat. What is the probability that both students pick the same tire?”
4. Suppose we roll a red die and a green die. What is the probability the number on the red die is larger ($>$) than the number on the green die?
5. Two dice are rolled. What is the probability (a) the two numbers will differ by 1 or less, (b) the maximum of the two numbers will be 5 or larger?
6. If we flip a coin 5 times, what is the probability that the number of heads is an even number (i.e., divisible by 2)?
7. The 1987 World Series was tied at two games a piece before the St. Louis Cardinals won the fifth game. According to the Associated Press, “The number of history support the Cardinals and the momentum they carry. Whenever the series has been tied 2-2 the team that won the fifth game won the series 71% of the time.” If momentum is not a factor and each team has a 50% chance of winning each game, what the probability that the game 5 winner will win the series?
8. Two boys are repeatedly playing a game that they each have probability $1/2$ of winning. The first person to win five games wins the match. What is the probability that Al will win if (a) he has won 4 games and Bobby has won 3; (b) he leads by a score of 3 games to 2?
9. 20 families live in a neighborhood. 4 have 1 child, 8 have 2 children, 5 have 3 children, and 3 have 4 children. If we pick a child at random what is the probability they come from a family with 1, 2, 3, 4 children?
10. In Galileo’s time people thought that when three dice were rolled, a sum of 9 and a sum of 10 had the same probability since each could be obtained in 6 ways:

$$9 : 1 + 2 + 6, 1 + 3 + 5, 1 + 4 + 4, 2 + 2 + 5, 2 + 3 + 4, 3 + 3 + 3$$

$$10 : 1 + 3 + 6, 1 + 4 + 5, 2 + 4 + 4, 2 + 3 + 5, 2 + 4 + 4, 3 + 3 + 4$$

Compute the probabilities of these sums and show that 10 is a more likely total than 9.

11. Suppose we roll three dice. Compute the probability that the sum is (a) 3, (b) 4, (c) 5, (d) 6, (e) 7, (f) 8 (g) 9 (h) 10.
12. In a group of students, 25% smoke cigarettes, 60% drink alcohol, and 15% do both. What fraction of students have at least one of these bad habits?
13. In a group of 320 high school graduates, only 160 went to college but 100 of the 170 men did. How many women did not go to college?
14. In the freshman class, 62% of the students take math, 49% take science, and 38% take both science and math. What percentage takes at least one science or math course?
15. 24% of people have American Express cards, 61% have Visa cards and 8% have both. What percentage of people have at least one credit card?
16. Suppose $\Omega = \{a, b, c\}$, $P(\{a, b\}) = 0.7$, and $P(\{b, c\}) = 0.6$. Compute the probabilities of $\{a\}$, $\{b\}$, and $\{c\}$.
17. Suppose A and B are disjoint with $P(A) = 0.3$ and $P(B) = 0.5$. What is $P(A^c \cap B^c)$?
18. Given two events A and B with $P(A) = 0.4$ and $P(B) = 0.7$, what are the maximum and minimum possible values for $P(A \cap B)$?

Independence

19. Suppose we draw two cards out of a deck of 52. Let $A =$ “the first card is an Ace,” and $B =$ “the second card is a spade.” Are A and B independent?
20. A family has three children, each of whom is a boy or a girl with probability $1/2$. Let $A =$ “there is at most 1 girl,” $B =$ “the family has children of both sexes.” (a) Are A and B independent? (b) Are A and B independent if the family has four children?
21. Suppose we roll a red and a green die. Let $A =$ “the red die shows a 2 or a 5,” $B =$ “the sum of the two dice is at least 7.” Are A and B independent?
22. Roll two dice. Let $A =$ “the sum is even,” and $B =$ “the sum is divisible by 3,” i.e., $B = \{3, 6, 9, 12\}$. Are A and B independent?
23. Roll two dice. Let $A =$ “the first die is odd,” $B =$ “the second die is odd,” and $C =$ “the sum is odd.” Show that these events are pairwise independent but not independent.
24. Nine children are seated at random in three rows of three desks. Let $A =$ “Al and Bobby sit in the same row,” $B =$ “Al and Bobby both sit at one of the four corner desks.” Are A and B independent?

25. Two students, Alice and Betty, are registered for a statistics class. Alice attends 80% of the time, Betty 60% of the time, and their absences are independent. On a given day, what is the probability (a) at least one of these students is in class, (b) exactly one of them is there?
26. Let A and B be two independent events with $P(A) = 0.4$ and $P(A \cup B) = 0.64$. What is $P(B)$?
27. Three students each have probability $1/3$ of solving a problem. What is the probability at least one of them will solve the problem?
28. Three independent events have probabilities $1/4$, $1/3$, and $1/2$. What is the probability exactly one will occur?
29. Three missiles are fired at a target. They will hit it with probabilities 0.2, 0.4, and 0.6. Find the probability that the target is hit by (a) three, (b) two, (c) one, (d) no missiles.
30. Three couples that were invited to dinner will independently show up with probabilities 0.9, $8/9$, and 0.75. Let N be the number of couples that show up. Calculate the probability $N = 3, 2, 1, 0$.
31. A college student takes 4 courses a semester for 8 semesters. In each course she has a probability $1/2$ of getting an A. Assuming her grades in different courses are independent, what is the probability she will have at least one semester with all A's?
32. When Al and Bob play tennis, Al wins a set with probability 0.7 while Bob wins with probability 0.3. What is the probability Al will be the first to win (a) two sets, (b) three sets?
33. Chevalier de Mere made money betting that he could "roll at least one 6 in four tries." When people got tired of this wager he changed it to "roll at least one double 6 in 24 tries" but then he started losing money. Compute the probabilities of winning these two bets.
34. Samuel Pepys wrote to Isaac Newton: "What is more likely, (a) one 6 in 6 rolls of one die or (b) two 6's in 12 rolls?" Compute the probabilities of these events.

Distributions

35. Suppose we roll two dice and let X and Y be the two numbers that appear. Find the distribution of $|X - Y|$.
36. Suppose we roll three tetrahedral dice that have 1, 2, 3, and 4 on their four sides. Find the distribution for the sum of the three numbers.
37. We roll two six-sided dice, one with sides 1,2,2,3,3,4 and the other with sides 1,3,4,5,6,8. What is the distribution of the sum?
38. How many children should a family plan to have so that the probability of having at least one child of each sex is at least 0.95?

39. How many times should a coin be tossed so that the probability of at least one head is at least 99%?

Expected Value

40. You want to invent a gambling game in which a person rolls two dice and is paid some money if the sum is 7, but otherwise he loses his money. How much should you pay him for winning a \$1 bet if you want this to be a fair game, that is, to have expected value 0?

41. A bet is said to carry 3 to 1 odds if you win \$3 for each \$1 you bet. What must the probability of winning be for this to be a fair bet?

42. A lottery has one \$100 prize, two \$25 prizes, and five \$10 prizes. What should you be willing to pay for a ticket if 100 tickets are sold?

43. In a popular gambling game, three dice are rolled. For a \$1 bet you win \$1 for each six that appears (plus your dollar back). If no six appears you lose your dollar. What is your expected value?

44. A roulette wheel has slots numbered 1 to 36 and two labeled with 0 and 00. Suppose that all 38 outcomes have equal probability. Compute the expected values of the following bets. In each case you bet one dollar and when you win you get your dollar back in addition to your winnings. (a) You win \$1 if one of the numbers 1 through 18 comes up. (b) You win \$2 if the number that comes up is divisible by 3 (0 and 00 do not count). (c) You win \$35 if the number 7 comes up.

45. In the Las Vegas game Wheel of Fortune, there are 54 possible outcomes. One is labeled "Joker," one "Flag," two "20," four "10," seven "5," fifteen "2," and twenty-four "1." If you bet \$1 on a number you win that amount of money if the number comes up (plus your dollar back). If you bet \$1 on Flag or Joker you win \$40 if that symbol comes up (plus your dollar back). What bets have the best and worst expected value here?

46. Sic Bo is an ancient Chinese dice game played with 3 dice. One of the possibilities for betting in the game is to bet "big." For this bet you win if the total X is 11, 12, 13, 14, 15, 16, or 17, except when there are three 4's or three 5's. On a \$1 bet on big you win \$1 plus your dollar back if it happens. What is your expected value?

47. Five people play a game of "odd man out" to determine who will pay for the pizza they ordered. Each flips a coin. If only one person gets Heads (or Tails) while the other four get Tails (or Heads) then he is the odd man and has to pay. Otherwise they flip again. What is the expected number of tosses needed to determine who will pay?

48. A man and wife decide that they will keep having children until they have one of each sex. Ignoring the possibility of twins and supposing that each trial is independent and results in a boy or girl with probability $1/2$, what is the expected value of the number of children they will have?

49. An unreliable clothes dryer dries your clothes and takes 20 minutes with probability 0.6, buzzes for 4 minutes and does nothing with probability 0.4. If we assume that successive trials are independent and that we patiently keep putting our money in to try to get it to work, what is the expected time we need to get our clothes dry?

Moments, Variance

50. A random variable has $P(X = x) = x/15$ for $x = 1, 2, 3, 4, 5$ and 0 otherwise. Find the mean and variance of X .

51. Find the mean and variance of the number of games in the World Series. Recall that it is won by the first team to win four games and assume that the outcomes are determined by flipping a coin.

52. Suppose we pick a month at random from a non-leap year calendar and let X be the number of days in the month. Find the mean and variance of X .

53. The Elm Tree golf course in Cortland, NY is a par 70 layout with 3 par fives, 5 par threes, and 10 par fours. Find the mean and variance of par on this course.

54. In a group of five items, two are defective. Find the distribution of N the number of draws we need to find the first defective item. Find the mean and variance of N .

55. Can we have a random variable with $EX = 3$ and $EX^2 = 8$?

56. Suppose $P(X \in \{1, 2, 3\}) = 1$, and $EX = 2.5$. What are the smallest and largest possible values for the variance?