1

Basic Concepts

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

1.1 Outcomes, events, and 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 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. Euclidean geometry begins with the notions of point and line. The corresponding basic object of probability is an **experiment**: an activity or procedure that produces distinct, well-defined possibilities called **outcomes**. (Here and throughout the book **boldface type** indicates a term that is being defined.)

Example 1.1 If our experiment is to roll one die then there are 6 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.

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 small 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 6." 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.

1.1.1 Axioms of probability theory

Let $\emptyset$ be the **empty set**, that is, 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 $P(\Omega) = 1$.

Cambridge University Press
978-0-521-86756-6 — Elementary Probability for Applications
Rick Durrett
Excerpt
More Information

(iii) If A and B are **disjoint**, that is, the intersection $A \cap B = \emptyset$, then

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

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

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

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

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

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

Given (1.1), assumptions (i) and (ii) are clear: the fraction of times 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) then 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 $n \to \infty$, we arrive at (iii).

Assumption (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 impose this assumption and do 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 $P(\{S\}) = 2/9$.

1.1.2 Basic properties of *P*(*A*)

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

Property 1. *Let* A^c *be the* **complement** *of* A*, that is, the set of outcomes not in* A*. Then*

$$P(A) = 1 - P(A^c) \tag{1.2}$$

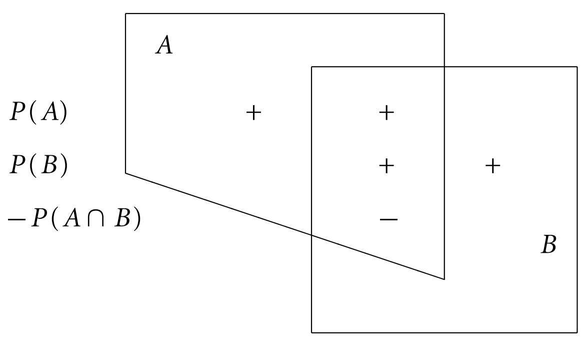
Proof. Let $A_1 = A$ and $A_2 = A^c$. Then $A_1 \cap A_2 = \emptyset$ and $A_1 \cup A_2 = \Omega$, so (iii) implies $P(A) + P(A^c) = P(\Omega) = 1$ by (ii). Subtracting $P(A)$ from each side of the equation gives the result. □

This formula is useful because sometimes it is easier to compute the probability of A^c. For an example, consider $A =$ "at least one 6." In this case $A^c =$ "no 6." There are $5 \cdot 5$ outcomes with no 6, 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*,*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{1.3}$$

Proof by 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. □

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

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

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. □

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

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

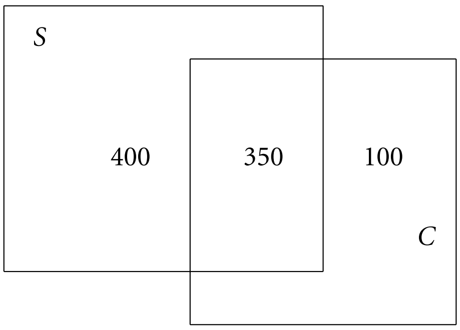
The same principle applies to counting outcomes in events.

Example 1.4 A survey of 1,000 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 $A \subset B$, that is, any outcome in A is also in B, then*

$$P(A) \leq P(B) \tag{1.4}$$

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). □

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

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

$$\lim_{n\to\infty} P(A_n) = P(A) \tag{1.5}$$

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

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

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), $\lim_{n\to\infty} P(B_n) = 1 - P(A)$. Since $P(B_n) = 1 - P(A_n)$, the desired result follows. □

1.2 Flipping coins and 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 4 outcomes:

	HH	HT TH	TT
Heads	2	1	0
Probability	1/4	1/2	1/4

Flipping three coins there are 8 possibilities:

	HHH	HHT HTH THH	TTH THT HTT	TTT
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.

Here, we are assuming that the two teams are equally matched and ignoring potential complicating factors such as the advantage of playing at home or psychological factors that make the outcome of one game affect the next one. In short, we suppose that the games are decided by tossing a fair coin to determine whether 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 compute the probability that A wins in the specified number of games and then multiply by 2. There are 4 possible outcomes:

$$BAAAA, \quad ABAAA, \quad AABAA, \quad AAABA$$

$AAAAB$ 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 learn systematic ways of doing this, but for now we compute the probabilities by enumerating the possibilities:

$$\begin{array}{llll} BBAAAA & ABBAAA & AABBAA & AAABBA \\ BABAAA & ABABAA & AABABA & \\ BAABAA & ABAABA & & \\ BAAABA & & & \end{array}$$

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, $P(6) = 2 \cdot 10/64 = 5/16$.

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 1, 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 in 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 agree 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 are very unusual (the probability of a chi-square 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 do not fit the model well, but are not very unusual ($p > 0.05$). On the other hand, the NBA finals data look like what we 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 2 people with exactly the same birthday. Should you take the bet?

To pose a mathematical problem we ignore February 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 this experiment. The number of outcomes in which all the birthdays are different is

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

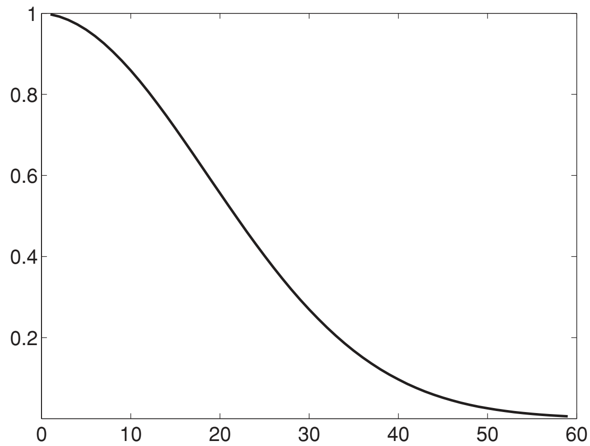
since the second person must avoid the first person's birthday, the third the first two birthdays, and so on, until the 30th person must avoid the 29 previous birthdays. Let D be the event that all birthdays are different. Dividing the number of outcomes in which all the birthdays are different by the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

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

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

Cambridge University Press
978-0-521-86756-6 — Elementary Probability for Applications
Rick Durrett
Excerpt
More Information

At first glance it is surprising that the probability of 2 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 $1 \leq k \leq 40$.



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

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

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

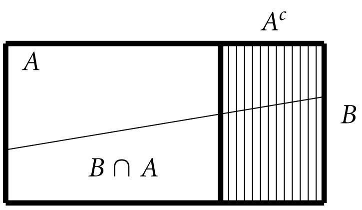
Cambridge University Press
978-0-521-86756-6 — Elementary Probability for Applications
Rick Durrett
Excerpt
More Information

To make the connection between the two definitions, we need to introduce the notion of conditional probability, which is 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

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \tag{1.6}$$

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 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

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

