5
Combinatorial Probability

5.1 Events and Probabilities

Probability theory is one of the most important areas of mathematics from the point of view of applications. In this book we do not attempt to introduce even the most basic notions of probability theory; our only goal is to illustrate the importance of combinatorial results about Pascal's Triangle by explaining a key result in probability theory, the Law of Large Numbers. To do so, we have to talk a little about what probability is.

If we make an observation about our world, or carry out an experiment, the outcome will always depend on chance (to a varying degree). Think of the weather, the stock market, or a medical experiment. Probability theory is a way of modeling this dependence on chance.

We start with making a mental list of all possible outcomes of the experiment (or observation, which we don't need to distinguish). These possible outcomes form a set S. Perhaps the simplest experiment is tossing a coin. This has two outcomes: H (heads) and T (tails). So in this case $S = \{H, T\}$. As another example, the outcomes of throwing a die form the set $S = \{1, 2, 3, 4, 5, 6\}$. In this book we assume that the set $S = \{s_1, s_2, \ldots, s_k\}$ of possible outcomes of our experiment is finite. The set S is often called a *sample space*.

Every subset of S is called an *event* (the event that the observed outcome falls in this subset). So if we are throwing a die, the subset $E = \{2, 4, 6\} \subseteq S$ can be thought of as the event that we throw an even number. Similarly,

the subset $L = \{4, 5, 6\} \subseteq S$ corresponds to the event that we throw a number larger than 3.

The intersection of two subsets corresponds to the event that both events occur; for example, the subset $L \cap E = \{4, 6\}$ corresponds to the event that we throw a better-than-average number that is also even. Two events A and B (i.e., two subsets of S) are called *exclusive* if they never occur at the same time, i.e., $A \cap B = \emptyset$. For example, the event $O = \{1, 3, 5\}$ that the outcome of tossing a die is odd and the event E that it is even are exclusive, since $E \cap O = \emptyset$.

5.1.1 What event does the union of two subsets corresponds to?

So let $S = \{s_1, s_2, \ldots, s_k\}$ be the set of possible outcomes of an experiment. To get a probability space we assume that each outcome $s_i \in S$ has a "probability" $\mathsf{P}(s_i)$ such that

(a) $\mathsf{P}(s_i) \geq 0$ for all $s_i \in S$,

and

(b) $\mathsf{P}(s_1) + \mathsf{P}(s_2) + \cdots + \mathsf{P}(s_k) = 1$.

Then we call S, together with these probabilities, a *probability space*. For example, if we toss a "fair" coin, then $\mathsf{P}(H) = \mathsf{P}(T) = \frac{1}{2}$. If the dice in our example is of good quality, then we will have $\mathsf{P}(i) = \frac{1}{6}$ for every outcome i.

A probability space in which every outcome has the same probability is called a *uniform probability space*. We shall only discuss uniform spaces here, since they are the easiest to imagine and they are the best for the illustration of combinatorial methods. But you should be warned that in more complicated modeling, nonuniform probability spaces are very often needed. For example, if we are observing whether a day is rainy or not, we will have a 2-element sample space $S = \{\text{RAINY}, \text{NONRAINY}\}$, but these two will typically *not* have the same probability.

The probability of an event $A \subseteq S$ is defined as the sum of probabilities of outcomes in A, and is denoted by $\mathsf{P}(A)$. If the probability space is uniform, then the probability of A is

$$\mathsf{P}(A) = \frac{|A|}{|S|} = \frac{|A|}{k}.$$

5.1.2 Prove that the probability of any event is at most 1.

5.1.3 What is the probability of the event E that we throw an even number with the die? What is the probability of the event $T = \{3, 6\}$ that we toss a number that is divisible by 3?

5.1.4 Prove that if A and B are exclusive, then $\mathsf{P}(A) + \mathsf{P}(B) = \mathsf{P}(A \cup B)$.

5.1.5 Prove that for any two events A and B,
$$\mathsf{P}(A \cap B) + \mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B).$$

5.2   Independent Repetition of an Experiment

Let us repeat our experiment n times. We can consider this as a single big experiment, and a possible outcome of this repeated experiment is a sequence of length n, consisting of elements of S. Thus the sample space corresponding to this repeated experiment is the set S^n of such sequences. Consequently, the number of outcomes of this "big" experiment is k^n. We consider every sequence equally likely, which means that we consider it a uniform probability space. Thus if $(a_1, a_2, \ldots, a_n)$ is an outcome of the "big" experiment, then we have

$$\mathsf{P}(a_1, a_2, \ldots, a_n) = \frac{1}{k^n}.$$

As an example, consider the experiment of tossing a coin twice. Then $S = \{H, T\}$ (heads, tails) for a single coin toss, and so the sample space for the two coin tosses is $\{HH, HT, TH, TT\}$. The probability of each of these outcomes is $\frac{1}{4}$.

This definition intends to model the situation where the outcome of each repeated experiment is independent of the previous outcomes, in the everyday sense that "there cannot possibly be any measurable influence of one experiment on the other." We cannot go here into the philosophical questions that this notion raises; all we can do is to give a mathematical definition that we can check, using examples, that it correctly expresses the informal notion above.

A key notion in probability is *independence* of events. Informally, this means that information about one event (whether or not it occurred) does not influence the probability of the other. Formally, two events A and B are *independent* if $\mathsf{P}(A \cap B) = \mathsf{P}(A)\mathsf{P}(B)$.

Consider again the experiment of tossing a coin twice. Let A be the event that the first toss is heads; let B be the event that the second toss is heads. Then we have $\mathsf{P}(A) = \mathsf{P}(HH) + \mathsf{P}(HT) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$, similarly $\mathsf{P}(B) = \frac{1}{2}$, and $\mathsf{P}(A \cap B) = \mathsf{P}(HH) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}$. Thus A and B are independent events (as they should be).

As another example, suppose that we toss a coin and simultaneously throw a die. The event H that we toss heads has probability $\frac{1}{2}$. The event K that we see 5 or 6 on the die has probability $\frac{1}{3}$. The event $H \cap K$ that we see heads on the coin and 5 or 6 on the die has probability $\frac{1}{6}$, since out of the 12 possible outcomes (H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6) two will have this property. So

$$\mathsf{P}(H \cap K) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = \mathsf{P}(H) \cdot \mathsf{P}(E),$$

and thus the events H and K are independent.

Independence of events is a mathematical notion and it does not necessarily mean that they have physically nothing to do with each other. If

$E = \{2, 4, 6\}$ is the event that the result of throwing a dice is even, and $T = \{3, 6\}$ is the event that it is a multiple of 3, then the event E and the event T are independent: we have $E \cap T = \{6\}$ (the only possibility to throw a number that is even and divisible by 3 is to throw 6), and hence

$$\mathsf{P}(E \cap T) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = \mathsf{P}(E)\mathsf{P}(T).$$

5.2.1 Which pairs of the events E, O, T, L are independent? Which pairs are exclusive?

5.2.2 Show that $\emptyset$ is independent of every event. Is there any other event with this property?

5.2.3 Consider an experiment with sample space S repeated n times ($n \geq 2$). Let $s \in S$. Let A be the event that the first outcome is s, and let B be the event that the last outcome is s. Prove that A and B are independent.

5.2.4 How many people do you think there are in the world who have the same birthday as their mother? How many people have the same birthday as their mother, father, and spouse?

5.3   The Law of Large Numbers

In this section we study an experiment that consists of n independent coin tosses. For simplicity, assume that n is even, so that $n = 2m$ for some integer m. Every outcome is a sequence of length n, in which each element is either H or T. A typical outcome would look like this:

$$HHTTTHTHTTHTHHHHTHTT$$

(for $n = 20$).

The *Law of Large Numbers* says that if we toss a coin many times, the number of "heads" will be about the same as the number of "tails". How can we make this statement precise? Certainly, this will not *always* be true; one can be extremely lucky or unlucky, and have a winning or loosing streak of arbitrary length. Also, we can't claim that the number of heads is equal to the number of tails; only that they are very likely to be close:

> *Flipping a coin n times, the probability that the percentage of heads is between 49% and 51% tends to 1 as n tends to ∞.*

The statement remains true if we replace 49% by 49.9% and 51% by 50.1%, or indeed by any two numbers strictly less 50% and larger than 50%, respectively. We can state this as a theorem, which is the simplest form of the Law of Large Numbers:

Theorem 5.3.1 *Fix an arbitrarily small positive number ϵ. If we flip a coin n times, the probability that the fraction of heads is between $0.5 - \epsilon$ and $0.5 + \epsilon$ tends to 1 as n tends to ∞.*

This theorem says, for example, that flipping a coin n times, the probability that the number of heads is between 49% and 51% is at least 0.99, if n is large enough. But how large must n be for this to hold? If $n = 49$ (which may sound pretty large) the number of heads can *never* be in this range; there are simply no integers between 49% of 49 (24.01) and 51% of 49 (24.99). How much larger does n have to be to assure that the number of heads is in this range for the majority of outcomes? This is an extremely important question in the statistical analysis of data: we want to know whether a deviation from the expected value is statistically significant.

Fortunately, much more precise formulations of the Law of Large Numbers can be made; one of these we can prove relatively easily, based on what we already know about Pascal's triangle. This proof will show that the Law of Large Numbers is not a mysterious force, but a simple consequence of the properties of binomial coefficients.

Theorem 5.3.2 *Let $0 \le t \le m$. Then the probability that out of $2m$ coin tosses, the number of heads is less than $m - t$ or larger than $m + t$, is at most $e^{-t^2/(m+t)}$.*

To illustrate the power of this theorem, let's go back to our earlier question: *How large should n be in order that the probability that the number of heads is between 49% and 51% is at least 0.99?* We want $m - t$ to be 49% of $n = 2m$, which means that $t = m/50$. The theorem says that the probability that the number of heads is not in this interval is at most $e^{-t^2/(m+t)}$. The exponent here is

$$-\frac{t^2}{m+t} = -\frac{\left(\frac{m}{50}\right)^2}{m + \frac{m}{50}} = -\frac{m}{2550}.$$

We want $e^{-m/2550} < 0.01$; taking the logarithm and solving for m, we get $m \ge 11744$ suffices. (This is pretty large, but, after all, we are talking about the "Law of Large Numbers.")

Observe that m is in the exponent, so that if m increases, the probability that the number of heads is outside the given interval drops very fast. For example, if $m = 1,000,000$, then this probability is less than 10^{-170}. Most likely, over the lifetime of the universe it never happens that out of a million coin tosses less than 49% or more than 51% are heads.

Normally, we don't need such a degree of certainty. Suppose that we want to make a claim about the number of heads with 95% certainty, but we would like to narrow the interval into which it falls as much as possible. In other words, we want to choose the smallest possible t so that

the probability that the number of heads is less than $m - t$ or larger than $m + t$ less than 0.05. By Theorem 5.3.2, this will be the case if

$$e^{-t^2/(m+t)} < 0.05.$$

(This is only a sufficient condition; if this holds, then the number of heads will be between $m - t$ and $m + t$ with probability at least 0.95. Using more refined formulas, we would find a slightly smaller t that works.) Taking the logarithm, we get

$$-\frac{t^2}{m+t} < -2.996.$$

This leads to a quadratic inequality, which we could solve for t; but it should suffice for this discussion that $t = 2\sqrt{m} + 2$ satisfies it (which is easy to check). So we get an interesting special case:

> With probability at least 0.95, the number of heads among $2m$ coin tosses is between $m - 2\sqrt{m} - 2$ and $m + 2\sqrt{m} + 2$.

If m is very large, then $2\sqrt{m} + 2$ is much smaller than m, so we get that the number of heads is very close to m. For example, if $m = 1{,}000{,}000$ then $2\sqrt{m} = 2{,}002 \approx 0.002m$, and so it follows that with probability at least 0.95, the number of heads is within $\frac{1}{5}$ of a percent of $m = n/2$.

It is time now to turn to the proof of Theorem 5.3.2.

Proof. Let A_k denote the event that we toss exactly k heads. It is clear that the events A_k are mutually exclusive. It is also clear that for every outcome of the experiment, exactly one of the A_k occurs.

The number of outcomes for which A_k occurs is the number of sequences of length n consisting of k heads and $n - k$ tails. If we specify which of the n positions are heads, we are done. This can be done in $\binom{n}{k}$ ways, so the set A_k has $\binom{n}{k}$ elements. Since the total number of outcomes is 2^n, we get the following:

$$\mathsf{P}(A_k) = \frac{\binom{n}{k}}{2^n}.$$

What is the probability that the number of heads is far from the expected, which is $m = n/2$; say, it is less than $m - t$ or larger than $m + t$, where t is any positive integer not larger than m? Using Exercise 5.1.4, we see that the probability that this happens is

$$\frac{1}{2^{2m}} \left(\binom{2m}{0} + \binom{2m}{1} + \cdots + \binom{2m}{m-t-1} + \binom{2m}{m+t+1} + \cdots \right.$$
$$\left. + \binom{2m}{2m-1} + \binom{2m}{2m} \right).$$

Now we can use Lemma 3.8.2, with $k = m - t$, and get that

$$\binom{2m}{0} + \binom{2m}{1} + \cdots + \binom{2m}{m-t-1} < 2^{2m-1} \binom{2m}{m-t} \bigg/ \binom{2m}{m}.$$

By (3.9), this can be bounded from above by

$$2^{2m-1}e^{-t^2/(m+t)}.$$

By the symmetry of Pascal's triangle, we also have

$$\binom{2m}{m+t+1} + \cdots + \binom{2m}{2m-1} + \binom{2m}{2m} < 2^{2m}e^{-t^2/(m+t)}.$$

Hence we get that the probability that we toss either fewer than $m - t$ or more than $m + t$ heads is less than $e^{-t^2/(m+t)}$. This proves the theorem. $\square$

5.4   The Law of Small Numbers and the Law of Very Large Numbers

There are two further statistical "laws" (half serious): the *Law of Small Numbers* and the *Law of Very Large Numbers*.

The first one says that if you look at small examples, you can find many strange or interesting patterns that do not generalize to larger numbers. Small numbers exhibit only a small number of patterns, and looking at various properties of small numbers, we are bound to see coincidences. For example, "every odd number is a prime" is true for 3, 5 and 7 (and one may be tempted to say that it is also true for 1, which is even "simpler" than primes: instead of two divisors, it has only one). Of course, this fails for 9.

Primes are strange (as we'll see) and in their irregular sequence, many strange patterns can be observed, which than fail if we move on to larger numbers. A dramatic example is the formula $n^2 - n + 41$. This gives a prime for $n = 0, 1, \ldots, 40$, but for $n = 41$ we get $41^2 - 41 + 41 = 41^2$, which is not a prime.

Fibonacci numbers are not as strange as primes: We have seen many interesting properties of them, and derived an explicit formula in Chapter 4. Still, one can make observations for the beginning of the sequence that do not remain valid if we check them far enough. For example, Exercise 4.3.4 gave a (false) formula for the Fibonacci numbers, namely $\lceil e^{n/2-1} \rceil$, which was correct for the first 10 positive integers n. There are many formulas that give integer sequences, but these sequences can start only so many ways: we are bound to find different sequences that start out the same way.

So the moral of the "Law of Small Numbers" is that to make a mathematical statement, or even to set up a mathematical conjecture, it is not enough to observe some pattern or rule, because you can only observe small instances and there are many coincidences for these. There is nothing wrong with making conjectures in mathematics, generalizing facts observed in special cases, but even a conjecture needs some other justification (an imprecise

argument, or a provable special case). A theorem, of course, needs much more: an exact proof.

The Law of Very Large Numbers says that strange coincidences can also be observed if we look at large sets of data. A friend of ours says, "I know two people who were both born on Christmas day. They complain that they get only one set of presents.... That's really strange. Are there many more people born on Christmas day than on other days?" No, this is not the explanation. The probability that a person is born on Christmas day is 1/365 (let's ignore leap years), so if you know, say, 400 people, then you can expect 1 or 2 of them to have a birthday on Christmas. Of course, you probably don't remember the birthdays of most people you know; but you are likely to remember those who complain about not getting enough presents!

Would you find it strange, even spooky, if somebody had the same birthday as his/her mother, father, and spouse? But if you have solved Exercise 5.2.4, you know that we have probably about 40 or so such people in the world, and probably a couple of them in the United States.

This is a fertile area for the tabloids and also for believers in the paranormal. We had better leave it at that.

Review Exercises

5.4.1 We throw a die twice. What is the probability that the sum of the points is 8?

5.4.2 Choose an integer uniformly from the set $\{1, 2, 3, \ldots, 30\}$. Let A be the event that it is divisible by 2; let B be the event that it is divisible by 3; let C be the event that it is divisible by 7.

(a) Determine the probabilities of A, B, and C.

(b) Which of the pairs (A, B), (B, C), and (A, C) are independent?

5.4.3 Let A and B be independent events. Express the probability $\mathsf{P}(A \cup B)$ in terms of the probabilities of A and B.

5.4.4 We select a subset X of the set $S = \{1, 2, \ldots, 100\}$ randomly and uniformly (so that every subset has the same probability of being selected). What is the probability that

(a) X has an even number of elements;

(b) both 1 and 100 belong to X;

(c) the largest element of S is 50;

(d) S has at most 2 elements.

5.4.5 We flip a coin n times $(n \geq 1)$. For which values of n are the following pairs of events independent?

(a) The first coin flip was heads; the number of all heads was even.

(b) The first coin flip was head; the number of all heads was more than the number of tails.

(c) The number of heads was even; the number of heads was more than the number of tails.