# 3 Conditional Probability & Independence

## 3.1 Conditional Probability

Suppose we have two events *E*, *F*. It is reasonable to ask;

How likely is it for *E* to occur if we know that *F* has already occurred?

**Example 3.1.** You roll two dice, one after the other. The sample space may therefore be taken to be  $S = \{(x, y) : 1 \le x, y \le 6\}$ . Let *E*, *F* be the events

- $E = \{(x, y) : x + y \ge 9\} = \{(3, 6), (4, 5), \dots, (6, 6)\}$  "the sum of the values is at least nine.
- $F = \{(5, y), (x, 5) : 1 \le x, y \le 6\} =$  "one of the dice is a five.

By simple counting, we have

$$\mathbb{P}(E) = \frac{4+3+2+1}{36} = \frac{10}{36} = \frac{5}{18}, \qquad \mathbb{P}(F) = \frac{11}{36}$$

Suppose you roll the dice; one lands out of sight but the visible die is a five: what is the probability that the sum is at least nine? For this new experiment, the set of possible outcomes (the sample space!) is now *F*, and we need only count how many of these have a sum of at least nine; we therefore want the event

$$E \cap F = \{(5,4), (5,5), (5,6), (4,5), (6,5)\}$$

The required probability is therefore  $\frac{|E \cap F|}{|F|} = \frac{5}{11}$ .

In general we make this a definition: the idea is that the probability of *E* happening *given F*, is the probability of both events happening relative to that of *F* itself.

**Definition 3.2.** Suppose *E*, *F* are events and that  $\mathbb{P}(F) \neq 0$ . The *conditional probability* that *E* occurs given that *F* has already occurred ('the probability of *E* given *F*') is

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$$

To revisit the example in this language,  $\mathbb{P}(E|F) = \frac{5/36}{11/36} = \frac{5}{11}$ . The other conditional probability represents the chance that one of the dice is a five given that the sum is at least nine: this would be

$$\mathbb{P}(F|E) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(E)} = \frac{5/36}{10/36} = \frac{1}{2}$$

As in the example, when *S* is a finite set all of whose singleton outcomes are equally likely, this approach should seem intuitive. We simply view *F* as a new sample space and count how many times the event  $E \cap F$  occurs relative to this new sample space:

$$\mathbb{P}(E|F) = \frac{|E \cap F|}{|F|} = \frac{|E \cap F|}{|S|} \cdot \frac{|S|}{|F|} = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$$

More generally, recall that we tend to interpret the probability  $\mathbb{P}(F)$  as being the (approximate) proportion of times *F* occurs if we run the experiment a very large number *n* of times:

$$\mathbb{P}(F) \approx \frac{\text{number of times } F \text{ occurs in } n \text{ trials}}{n}$$

Over n trials, it follows that

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} \approx \frac{\text{number of times both } E \text{ and } F \text{ occur}}{\text{number of times } F \text{ occurs}}$$

The intuition is that this approximation should become an equality as  $n \to \infty$ . Before proceeding to examples, there is one other obvious thing to check.

**Theorem 3.3.** Let *F* be an event in the sample space *S*. If  $\mathbb{P}(F) \neq 0$ , then  $\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$  defines a probability measure on *S*. More specifically, it satisfies the axioms of probability:

- 1.  $0 \leq \mathbb{P}(E|F) \leq 1$ ,
- 2.  $\mathbb{P}(S|F) = 1$ ,
- 3. If  $E_1, E_2, \ldots$  are mutually exclusive, then  $\mathbb{P}(\bigcup E_i | F) = \sum \mathbb{P}(E_i | F)$

*Proof.* 1. Since  $F = (E \cup E^{\mathsf{C}}) \cap F = (E \cap F) \cup (E^{\mathsf{C}} \cap F)$  is a union of mutually exclusive events,

$$\mathbb{P}(F) = \mathbb{P}(E \cap F) + \mathbb{P}(E^{\mathsf{C}} \cap F) \ge \mathbb{P}(E \cap F) \implies \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} \le 1$$

2. 
$$\mathbb{P}(S|F) = \frac{\mathbb{P}(S \cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}(F)}{\mathbb{P}(F)} = 1.$$

3. This follows from the distributive law

$$\left(\bigcup E_i\right)\cap F=\bigcup(E_i\cap F)$$

If the  $E_I$  are mutually exclusive, so plainly are the events  $E_i \cap F$ . But then

$$\mathbb{P}(\bigcup E_i|F) = \frac{\mathbb{P}(\bigcup(E_i \cap F))}{\mathbb{P}(F)} = \sum \frac{\mathbb{P}(E_i \cap F)}{\mathbb{P}(F)} = \sum \mathbb{P}(E_i|F)$$

**Example 3.4.** A student estimates that she has a 50% chance of getting an A in Ordinary Differential Equations (ODE), a 40% chance in geometry and a 30% chance of getting an A in both.

1. If the student receives an A in geometry, what is the probability that she will also do so in ODE?

Let *O*, *G* be, respectively, the events that the student receives an A in ODE and geometry. We therefore want

$$\mathbb{P}(O|G) = \frac{\mathbb{P}(O \cap G)}{\mathbb{P}(G)} = \frac{0.3}{0.4} = 75\%$$

2. The student decides to take only one of the classes and decides which by a coin flip. What is the probability that she gets an A in geometry?

This time let *E* be the event that the student chooses geometry and *A* that the student gets an A in whichever class she chooses. This time we require

$$\mathbb{P}(E \cap A) = \mathbb{P}(E)\mathbb{P}(A|E) = \frac{1}{2} \cdot \frac{4}{10} = 20\%$$

This last illustrates a general rule for conditional probabilities:

**Theorem 3.5 (Multiplication Rule).** Given events  $E_i$  with  $\mathbb{P}(E_1) \neq 0$ , we have

 $\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2|E_1)$ 

More generally, if  $\mathbb{P}(E_1 \cap \cdots \cap E_{n-1}) \neq 0$ , then<sup>1</sup>

$$\mathbb{P}(E_1 \cap \cdots \cap E_n) = \mathbb{P}(E_1) \mathbb{P}(E_2 | E_1) \mathbb{P}(E_3 | E_1 \cap E_2) \cdots \mathbb{P}(E_n | E_1 \cap \cdots \cap E_{n-1})$$

The general formula is easily checked by multiplying out the right hand side and cancelling:

$$\mathbb{P}(E_1 \cap \dots \cap E_n) = \mathbb{P}(E_1) \cdot \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_1)} \cdot \frac{\mathbb{P}(E_1 \cap E_2 \cap E_3)}{\mathbb{P}(E_1 \cap E_2)} \cdots \frac{\mathbb{P}(E_1 \cap \dots \cap E_n)}{\mathbb{P}(E_1 \cap \dots \cap E_{n-1})}$$

The multiplication rule makes for an intuitive method of calculation which feels akin to the fundamental principle of counting.

**Examples 3.6.** 1. You estimate your chance of shooting an unpressured basket at 90%. You enter a competition where the goal is to shoot 5 baskets in a row. Due to nerves, your chance of shooting a basket is reduced by 10% for each previous successful shot. What is the probability that you win the competition?

Let  $E_1, \ldots, E_5$  be the events "shoot the i<sup>th</sup> basket." We therefore want

$$\mathbb{P}(E_1 \cap \dots \cap E_5) = \mathbb{P}(E_1) \mathbb{P}(E_2 | E_1) \mathbb{P}(E_3 | E_1 \cap E_2) \cdots \mathbb{P}(E_5 | E_1 \cap E_2 \cap E_3 \cap E_4)$$
  
=  $\frac{9}{10} \cdot \frac{8}{10} \cdot \frac{7}{10} \cdot \frac{6}{10} \cdot \frac{5}{10} = \frac{15120}{100000} = 15.12\%$ 

2. Recall the hat-matching problem with *n* participants. We use the multiplication rule to obtain the probability that exactly *k* people match their own hat.

To keep ourselves straight, here are all 24 = 4! equally likely arrangements of hats *a*, *b*, *c*, *d* for n = 4 people *A*, *B*, *C*, *D*; we group these according to all the ways that 4, 2, 1 or no people claim their own hat (note that if three claim their own hat, the fourth must also!).

 A
 a
 a
 a
 d
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 b
 c
 d
 c
 d
 c
 d
 c
 d
 c
 d
 c
 d
 d
 c
 d
 d
 c
 d
 d
 c
 d
 d
 c
 d
 d
 c
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d
 d

<sup>1</sup>You may prefer the look of this expression if you use the convention of juxtaposition for intersection; i.e.

 $<sup>\</sup>mathbb{P}(E_1 \cdots E_n) = \mathbb{P}(E_1) \mathbb{P}(E_2 | E_1) \mathbb{P}(E_3 | E_1 E_2) \cdots \mathbb{P}(E_n | E_1 \cdots E_{n-1})$ 

Consider the events  $E_1, \ldots, E_n$  where  $E_i$  is the event of the *i*<sup>th</sup> person claims their own hat. Plainly

$$\mathbb{P}(E_1) = \frac{1}{n}, \quad \mathbb{P}(E_2|E_1) = \frac{1}{n-1}, \quad \mathbb{P}(E_3|E_1 \cap E_2) = \frac{1}{n-2}, \dots$$

from which

$$\mathbb{P}(E_1 \cap \dots \cap E_k) = \frac{1}{n(n-1)\dots(n+1-k)} = \frac{(n-k)!}{n!}$$
(\*)

In the context of the concrete example:

- $\mathbb{P}(E_1) = \frac{1}{4} = \frac{6}{24}$  corresponds to the six *a*'s in the 1<sup>st</sup> row.
- $\mathbb{P}(E_2|E_1) = \frac{1}{3} = \frac{2}{6}$  corresponds to the two *b*'s *underneath* the six *a*'s in columns 1 & 2.
- $\mathbb{P}(E_3|E_1 \cap E_2) = \frac{1}{2}$  corresponds to the single *c* in column 1.
- $\mathbb{P}(E_4|E_1 \cap E_2 \cap E_3) = 1$  corresponds to the single *d* in column 1.

At this point,<sup>*a*</sup> given that the first *k* people have their own hat, our earlier discussion tells us that the probability of none of the remaining n - k people getting their own hat is  $\sum_{j=0}^{n-k} \frac{(-1)^j}{j!}$ . The probability that *exactly* the first *k* people getting their own hat is therefore

$$\mathbb{P}\left(\bigcap_{i=1}^{k} E_{i} \cap \bigcap_{i=k+1}^{n} E_{i}^{\mathsf{C}}\right) = \mathbb{P}\left(\bigcap_{i=1}^{k} E_{i}\right) \mathbb{P}\left(\bigcap_{i=k+1}^{n} E_{i}^{\mathsf{C}} \mid \bigcap_{i=1}^{k} E_{i}\right) = \frac{(n-k)!}{n!} \sum_{j=0}^{n-k} \frac{(-1)^{j}}{j!}$$

Since there are  $\binom{n}{k}$  subsets of size *k*, the probability that exactly *k* people claim their own hat is

$$\mathbb{P}(\text{Exactly } k \text{ claim own hat}) = \binom{n}{k} \frac{(n-k)!}{n!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}$$

As previously, note that this converges to  $\frac{e^{-1}}{k!}$  as  $n \to \infty$ , in line with the fact that  $\sum_{k=0}^{\infty} \frac{1}{k!} = e$ . Again, we verify that this matches our complete data when n = 4:

•  $\mathbb{P}(k=0) = \frac{1}{0!} \sum_{j=0}^{4} \frac{(-1)^j}{j!} = 1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} = \frac{9}{24}$ 

• 
$$\mathbb{P}(k=1) = \frac{1}{1!} \sum_{j=0}^{3} \frac{(-1)^j}{j!} = 1 - 1 + \frac{1}{2} - \frac{1}{6} = \frac{1}{3} = \frac{8}{24}$$

• 
$$\mathbb{P}(k=2) = \frac{1}{2!} \sum_{j=0}^{2} \frac{(-1)^j}{j!} = \frac{1}{2} \left( 1 - 1 + \frac{1}{2} \right) = \frac{1}{4} = \frac{6}{24}$$

• 
$$\mathbb{P}(k=3) = \frac{1}{3!} \sum_{j=0}^{1} \frac{(-1)^j}{j!} = \frac{1}{6} (1-1) = 0$$

• 
$$\mathbb{P}(k=4) = \frac{1}{4!} \sum_{j=0}^{0} \frac{(-1)^j}{j!} = \frac{1}{24}$$

<sup>a</sup>It is *really* tempting to continue the process via

$$\mathbb{P}(E_{k+1}^{\mathsf{C}}|E_1\cap\cdots\cap E_k)=\frac{n-k-1}{n-k},\ldots$$

We leave it to the exercises to explain why this *doesn't* work!

## 3.2 Bayes' Formula

Recall that complementary events are mutually exclusive  $F \cap F^{\mathsf{C}} = \emptyset$ . If follows that we may decompose any event *E* into mutually exclusive events *dependent* on the occurrence or non-occurrence of *F*:  $E = (E \cap F) \cup (E \cap F^{\mathsf{C}})$ . This should be entirely intuitive.

**Example 3.7.** Let *E* be the event "Get an A overall in analysis," and *F* the event "Score 40/50 on the midterm." We plainly have two mutually exclusive events which correspond to all possible ways to obtain an A:

- $E \cap F$  "Get an A overall and score 40/50 on the midterm"
- $E \cap F^{\mathsf{C}}$  "Get an A overall and score < 40 on the midterm"

Suppose that  $\mathbb{P}(F) = 20\%$  of students score  $\geq 40$  on the midterm. Moreover, suppose that an overall A is obtained by

- $\mathbb{P}(E|F) = 75\%$  of students who get  $\geq 40$  on the midterm,
- $\mathbb{P}(E|F^{\mathsf{C}}) = 15\%$  of students who get < 40 on the midterm.

These are conditional probabilities, whence we can use the multiplication rule to obtain the probability of a student obtaining an overall A

$$\mathbb{P}(E) = \mathbb{P}(E \cap F) + \mathbb{P}(E \cap F^{\mathsf{C}}) \qquad (\text{mutually exclusive events})$$

$$= \mathbb{P}(F)\mathbb{P}(E|F) + \mathbb{P}(F^{\mathsf{C}})\mathbb{P}(E|F^{\mathsf{C}}) \qquad (\text{mutually exclusive events})$$

$$= \mathbb{P}(F)\mathbb{P}(E|F) + (1 - \mathbb{P}(F))\mathbb{P}(E|F^{\mathsf{C}}) \qquad (\text{complementary events } F, F^{\mathsf{C}})$$

$$= \frac{1}{5} \cdot \frac{3}{4} + \frac{4}{5} \cdot \frac{3}{20} = \frac{15 + 12}{100} = 27\%$$

Now suppose a student gets an overall A in the course. What is the probability that they scored at least 40 on the midterm?

We use the multiplication rule again to compute

$$\mathbb{P}(F|E) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(E)} = \frac{\mathbb{P}(F)\mathbb{P}(E|F)}{\mathbb{P}(E)} = \frac{\frac{1}{5} \cdot \frac{3}{4}}{\frac{27}{100}} = \frac{5}{9} \approx 55.5\%$$

This approaches are completely general:

**Theorem 3.8 (Bayes' Theorem).** If *E*, *F* are any events for which the conditional probabilities make sense, then

- 1. (Law of total probability)  $\mathbb{P}(E) = \mathbb{P}(F)\mathbb{P}(E|F) + (1 \mathbb{P}(F))\mathbb{P}(E|F^{\mathsf{C}})$
- 2. (Bayes' formula)  $\mathbb{P}(F|E) = \frac{\mathbb{P}(F)\mathbb{P}(E|F)}{\mathbb{P}(E)}$

Bayes' formula permits us to switch the order of conditional probabilities. Take extra care with this, since answers are often counter-intuitive, particularly when data is socially loaded and biases come in to play.

**Example 3.9.** Suppose we have the following data:

- 1% of a population is a millionaire.
- 80% of millionaires drive a luxury car.
- 20% of the entire population drives a luxury car.

If you see a luxury car, what is the probability that the driver is a millionaire?

The gut interpretation is that it is *very likely*, since such a large majority of millionaires drive a luxury car. This is to get the implication the wrong way round: if *E*, *F* are, respectively, the events "person is a millionaire" and "person drives a luxury car," then the probability we want is

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E)\mathbb{P}(F|E)}{\mathbb{P}(F)} = \frac{0.01 \cdot 0.8}{0.2} = 4\%$$

This is still small, since the proportion of millionaires in the overall population is also small.

Problems such as this are very common in medical situations. Suppose a patient is tested for a particular medical condition. Consider two events:

- *C* "The patient has the *condition*"
- *T* "The patient *tests* positive."

If a patient is tested, there are four possible mutually exclusive events, sometimes given names

If a patient takes a test and it comes back positive, they want to know how likely it is that they have the condition. The important values are the following:

- $\mathbb{P}(T|C)$  (Sensitivity): if a patient has condition, how likely are they to test positive?
- $\mathbb{P}(T^{\mathsf{C}}|C^{\mathsf{C}})$  (Specificity): if a patient hasn't got the condition, how likely are they to test negative?

**Example 3.10.** A person tests positive for a disease which is known to be present in 1% of the population. If the test correctly detects 90% of cases and 91% of negatives, how likely<sup>*a*</sup> is it that the person has the disease?

We know  $\mathbb{P}(C) = 1\%$ ,  $\mathbb{P}(T|C) = 90\%$ , and  $\mathbb{P}(T^{\mathsf{C}}|C^{\mathsf{C}}) = 91\%$  and wish to compute

$$\mathbb{P}(C|T) = \frac{\mathbb{P}(C \cap T)}{\mathbb{P}(T)} = \frac{\mathbb{P}(C)\mathbb{P}(T|C)}{\mathbb{P}(T|C)\mathbb{P}(C) + \mathbb{P}(T|C^{\mathsf{C}})\mathbb{P}(C^{\mathsf{C}})}$$
$$= \frac{\mathbb{P}(C)\mathbb{P}(T|C)}{\mathbb{P}(T|C)\mathbb{P}(C) + (1 - \mathbb{P}(T^{\mathsf{C}}|C^{\mathsf{C}}))(1 - \mathbb{P}(C))}$$
$$= \frac{1 \cdot 90}{90 \cdot 1 + 9 \cdot 99} = \frac{90}{981} = \frac{10}{109} \approx 9.17\%$$

where we used the fact was written using percentages.

<sup>&</sup>quot;See https://www.bbc.com/news/magazine-28166019here for how this exact problem is routinely misunderstood by experts who should know better!

#### **Conditioning on Mutually Exclusive Events**

Both parts of Theorem 3.8 extend naturally to finitely many events:

**Corollary 3.11.** Suppose the sample space is a union of mutually exclusive events  $S = \bigcup_{i=1}^{n} F_i$ .

1. (Law of total probability) The probability of an event *E* is a weighted average of the probabilities that *E* occurs given each of the mutually exclusive events *F*<sub>i</sub>:

$$\mathbb{P}(E) = \mathbb{P}\left(\bigcup_{i=1}^{n} E \cap F_i\right) = \sum_{i=1}^{n} \mathbb{P}(E \cap F_i) = \sum_{i=1}^{n} \mathbb{P}(E|F_i)\mathbb{P}(F_i)$$

This method of computation is known as conditioning on the events F<sub>i</sub>.

2. (Bayes' formula)  $\mathbb{P}(F_j|E) = \frac{\mathbb{P}(E \cap F_j)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E|F_j)\mathbb{P}(F_j)}{\sum_{i=1}^n \mathbb{P}(E|F_i)\mathbb{P}(F_i)}$ 

The special case of the Theorem was where  $S = F \cup F^{\mathsf{C}}$ ; i.e.  $F_2 = F_1^{\mathsf{C}}$ .

**Examples 3.12.** 1. A bin contains three types of flashlight: types I, II, III. Suppose the following data is known regarding these lights.

Туре	Number in bin	chance of lasting at least 24 hours
Ι	50	60%
II	100	70%
III	25	90%

(a) What is the probability that a randomly chosen flashlight lasts at least 24 hours?

Let *E* be "the chosen light lasts at least 24 hours." We condition on the three mutually exclusive events  $F_i$  "choose a flashlight of type i."

$$\mathbb{P}(E) = \mathbb{P}(E|F_{\rm I})\mathbb{P}(F_{\rm I}) + \mathbb{P}(E|F_{\rm II})\mathbb{P}(F_{\rm II}) + \mathbb{P}(E|F_{\rm III})\mathbb{P}(F_{\rm III})$$
$$= \frac{6}{10} \cdot \frac{2}{7} + \frac{7}{10} \cdot \frac{4}{7} + \frac{9}{10} \cdot \frac{1}{7} = \frac{7}{10} = 70\%$$

(b) If a randomly chosen flashlight lasts at least 24 hours, we find the probability that it was a flashlight of each type. For each *j*,

$$\mathbb{P}(F_j|E) = \frac{\mathbb{P}(F_j)\mathbb{P}(E|F_j)}{\mathbb{P}(E)} = \begin{cases} \frac{10}{7} \cdot \frac{12}{70} = \frac{12}{49} \approx 24.5\% & \text{for type I} \\ \frac{10}{7} \cdot \frac{28}{70} = \frac{28}{49} \approx 57.1\% & \text{for type II} \\ \frac{10}{7} \cdot \frac{9}{70} = \frac{9}{49} \approx 18.4\% & \text{for type III} \end{cases}$$

Note that the three conditional probabilities sum to 1, as they must!

(c) Suppose the chosen flashlight does not last 24 hours. Find the probability that the choice was not type I. This time we want

$$\mathbb{P}(F_1^{\mathsf{C}}|E^{\mathsf{C}}) = 1 - \mathbb{P}(F_1|E^{\mathsf{C}}) = 1 - \frac{\mathbb{P}(F_1)\mathbb{P}(E^{\mathsf{C}}|F_1)}{\mathbb{P}(E^{\mathsf{C}})} = 1 - \frac{\mathbb{P}(F_1)(1 - \mathbb{P}(E|F_1))}{1 - \mathbb{P}(E)}$$
$$= 1 - \frac{\frac{7}{7} \cdot \frac{4}{10}}{3/10} = \frac{13}{21} \approx 61.9\%$$

2. Recall the example of the problem of points from the introduction, where player A is 3–2 up in a first-to-five coin flip. If the game continues until someone wins, the outcome tree is as follows.

$$3-2$$

$$A = 2$$

$$A = 3$$

$$A = 4$$

$$A = 4$$

$$A = 5$$

Remember that the probability of any particular scoreline depends on which *column* it lies in; each time you move one column to the right, the outcomes become half as likely.

Our approach in the introduction was essentially to use conditional probabilities. Here it is a little more formally. Let  $F_n$  be the event "the game finishes in n coin flips". The possible finishes are scored. Since the outcome of each coin-flip is equally likely, we easily count the probabilities of the mutually exclusive events

$$\mathbb{P}(F_7) = \frac{1}{2^2} = \frac{1}{4}, \qquad \mathbb{P}(F_8) = \frac{3}{2^3} = \frac{3}{8}, \qquad \mathbb{P}(F_9) = \frac{6}{2^4} = \frac{3}{8}$$

Let *E* be the event "Player A wins." Conditioning on the events  $F_i$ , we obtain

$$\mathbb{P}(E) = \mathbb{P}(E|F_7)\mathbb{P}(F_7) + \mathbb{P}(E|F_8)\mathbb{P}(F_8) + \mathbb{P}(E|F_9)\mathbb{P}(F_9)$$
  
=  $1 \cdot \frac{1}{4} + \frac{2}{3} \cdot \frac{3}{8} + \frac{3}{6} \cdot \frac{3}{8} = \frac{1}{4} + \frac{1}{4} + \frac{3}{16} = \frac{11}{16}$ 

This isn't the best way to solve the problem since finding these ingredients becomes much harder for longer games. Recall that we solved the problem in a different way in chapter 2.

This last example also illustrates a commonly used concept.

**Definition 3.13.** The *odds* of an event *E* occurring are 
$$\frac{\mathbb{P}(E)}{\mathbb{P}(E^{\mathsf{C}})} = \frac{\mathbb{P}(E)}{1 - \mathbb{P}(E)}$$

In the previous example, the odds of player A winning are  $\frac{11}{5}$ . The odds might instead be written 11 : 5 and read "11 to 5," encapsulating the idea that if the game were repeated 16 = 11 + 5 times, player A would expect to win 11 times and lose 5 times.

We finish this section with an example of how the odds change on receipt of new information. This is another straightforward application of Bayes' formula:

$$\frac{\mathbb{P}(E|F)}{\mathbb{P}(E^{\mathsf{C}}|F)} = \frac{\mathbb{P}(F|E)\mathbb{P}(E)}{\mathbb{P}(F)} \cdot \frac{\mathbb{P}(F)}{\mathbb{P}(F|E^{\mathsf{C}})\mathbb{P}(E^{\mathsf{C}})} = \frac{\mathbb{P}(E)}{\mathbb{P}(E^{\mathsf{C}})} \cdot \frac{\mathbb{P}(F|E)}{\mathbb{P}(F|E^{\mathsf{C}})}$$

If a new event *F* occurs, the odds change by multiplication by the relative chance of *F* occurring given *E* and its complement.

**Example 3.12.1 (cont)** The odds that a drawn flashlight lasts 24 hours are  $\frac{\mathbb{P}(E)}{\mathbb{P}(E^{C})} = \frac{7}{3}$ . Now suppose we know that the chosen light is *not* of type I. This new information increases the odds

$$\frac{\mathbb{P}(E|F_{\mathrm{I}}^{\mathsf{C}})}{\mathbb{P}(E^{\mathsf{C}}|F_{\mathrm{I}}^{\mathsf{C}})} = \frac{\mathbb{P}(E)}{\mathbb{P}(E^{\mathsf{C}})} \cdot \frac{\mathbb{P}(F_{\mathrm{I}}^{\mathsf{C}}|E)}{\mathbb{P}(F_{\mathrm{I}}^{\mathsf{C}}|E^{\mathsf{C}})} = \frac{\mathbb{P}(E)}{\mathbb{P}(E^{\mathsf{C}})} \cdot \frac{\mathbb{P}(F_{\mathrm{II}}|E) + \mathbb{P}(F_{\mathrm{III}}|E)}{\mathbb{P}(F_{\mathrm{I}}^{\mathsf{C}}|E^{\mathsf{C}})}$$
$$= \frac{7}{3} \cdot \frac{40/49}{13/21} = \frac{7}{3} \cdot \frac{120}{91} = \frac{40}{13}$$

which makes sense, since the type I flashlights are the least likely to last a full 24 hours.

**Warning! Gambling Odds** In the real world, probabilities are often quoted in odds but written *backwards* relative to the mathematical definition. Colloquially, a "million to one" event *E* occurs once for every million times it does not; that is

$$\frac{\mathbb{P}(E)}{1 - \mathbb{P}(E)} = \frac{1}{1000000} \iff \mathbb{P}(E) = \frac{1}{1000001}$$

This approach is especially prevalent in gambling, for instance,

The grey horse has odds of 5-2 (or 5/2)

Quoting odds this way helps punters understand what they really care about: *winnings* (profit). If you risk \$2 on a 5–2 horse and it wins, you'll get back your \$2 stake plus winnings of \$5.

To unpack this, imagine the race were run 7 = 5 + 2 times, then the grey horse would be expected to win *twice*. The probability *p* of the grey horse winning therefore satisfies

$$\frac{p}{1-p} = \frac{2}{5} \iff 5p = 2 - 2p \iff p = \frac{2}{7}$$

The expected outcome of betting \$2 on this race seven times would be therefore be net zero:

- Lose five times for a total loss of -\$10.
- Win twice for a total profit of \$10.

Alternatively, we can view the problem symmetrically. You put \$2 in the pot and the bookmaker \$5; if the grey horse wins, you take the \$7 pot; if it loses, the bookmaker takes the \$7. Each of you risked different amounts and had different potential winnings. Indeed the bookmaker's odds were precisely the opposite at 2–5.

## 3.3 Independent events

It is sometimes the case that new information does not change a probability; that  $\mathbb{P}(E|F) = \mathbb{P}(E)$ . This immediately implies  $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$ . We make the latter our definition.

**Definition 3.14.** Events *E*, *F* are *independent* if  $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$ , and *dependent* otherwise. More generally, a collection of events *E<sub>n</sub>* is *independent* if every finite subcollection satisfies

 $\mathbb{P}(E_{n_1} \cap \cdots \cap E_{n_k}) = \mathbb{P}(E_{n_1}) \cdots \mathbb{P}(E_{n_k})$ 

If an experiment is consists of *n* independent identical subexperiments, then we call each subexperiment a *trial*.

- **Examples 3.15.** 1. The events *E* "roll a 4 on a fair die" and *F* "roll an even number" are dependent since  $\mathbb{P}(E|F) = \frac{1}{3} \neq \frac{1}{6} = \mathbb{P}(E)$ . Equivalently  $\mathbb{P}(E \cap F) = \mathbb{P}(E) = \frac{1}{6} \neq \frac{1}{18} = \mathbb{P}(E)\mathbb{P}(F)$ 
  - 2. Let *K*, *C* be the events "deal a King" and "deal a club" from a shuffled pack of cards. These events are independent since

$$\mathbb{P}(K \cap C) = \frac{1}{52} = \frac{4}{52} \cdot \frac{1}{4} = \mathbb{P}(K)\mathbb{P}(C)$$

3. Suppose we roll a fair die three times. Since a die has no memory, each roll is a separate experiment. Any events which concern only the values on separate dice are therefore independent. Consider the events  $E_{ij}$ ; "the *i*<sup>th</sup> die roll is a *j*". If the sample space is the set  $S = \{(a, b, c) : 1 \le a, b, c \le 6\}$ , then

$$E_{1x} = \{(x, b, c) : 1 \le b, c \le 6\}$$
 and  $\mathbb{P}(E_{1x}) = \frac{6}{6^3} = \frac{1}{36}$ 

Plainly

$$\mathbb{P}(E_{1x} \cap E_{2y}) = \mathbb{P}\{(x, y, c) : 1 \le c \le 6\} = \frac{6}{6^3} = \frac{1}{6^2} = \mathbb{P}(E_{1x})\mathbb{P}(E_{2y})$$

The same holds for any pair of events from  $E_{1x}$ ,  $E_{2y}$  and  $E_{3z}$ . Moreover,

$$\mathbb{P}(E_{1x} \cap E_{2y} \cap E_{3z}) = \mathbb{P}(\{(x, y, z)\}) = \frac{1}{6^3} = \mathbb{P}(E_{1x})\mathbb{P}(E_{2y})\mathbb{P}(E_{3z})$$

We conclude that the events  $E_{1x}$ ,  $E_{2y}$  and  $E_{3z}$  are independent.

- 4. Roll two fair dice and consider the following events and their probabilities
  - *E* "the sum is 7" has  $\mathbb{P}(E) = \frac{6}{36} = \frac{1}{6}$
  - *F* "the first die is a 3" has  $\mathbb{P}(F) = \frac{1}{6}$
  - *G* "the second die is a 5" has  $\mathbb{P}(G) = \frac{1}{6}$

Plainly *F*, *G* are independent since the two dice rolls have nothing to do with each other. *E* and *F* are also independent, since

$$\mathbb{P}(E \cap F) = \frac{|\{(3,4)\}|}{36} = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(E)\mathbb{P}(F)$$

*E* and *G* are independent similarly. However  $\mathbb{P}(E \cap F \cap G) = 0 \neq \mathbb{P}(E)\mathbb{P}(F)\mathbb{P}(G)$ , so the three events are dependent.

Warnings: The definition of independence is purely *formulaic*, so you must be very careful applying intuition related to its meaning in *English*. In particular:

• In the last example, *F* and *G* are independent *because* they are outcomes of entirely unrelated experiments; this is our intuition about independence, but it is incorrect. The experiments for *E* and *F* are related, but the events are still independent! This is, however, merely a quirk of numbers: if *E* were changed to "the sum is 8," then the events *E*, *F* are now *dependent* 

$$\mathbb{P}(E \cap F) = \frac{1}{36} \neq \frac{5}{36} \cdot \frac{1}{6} = \mathbb{P}(E)\mathbb{P}(F)$$

• *Independence* and *mutual exclusivity* ( $E \cap F = \emptyset$ ) are not the same thing! If *E*, *F* are mutually exclusive events with non-zero probabilities, then

 $\mathbb{P}(E)\mathbb{P}(F) \neq 0 = \mathbb{P}(\emptyset) = \mathbb{P}(E \cap F)$ 

Mutually exclusive events are essentially never independent!

Independent events combine naturally; for instance, if E, F are independent, so also are E and  $F^{C}$ :

$$\mathbb{P}(E \cap F^{\mathsf{C}}) = \mathbb{P}(E) - \mathbb{P}(E \cap F) = \mathbb{P}(E) - \mathbb{P}(E)\mathbb{P}(F) = \mathbb{P}(E)(1 - \mathbb{P}(F)) = \mathbb{P}(E)\mathbb{P}(F^{\mathsf{C}})$$

**Example 3.16.** Independent trials of rolling a fair die are performed. What is the probability of rolling a 1 before rolling a 5 or 6?

Let  $E_n$  be the event that the first n - 1 rolls are either 2, 3 or 4, and that the n<sup>th</sup> roll is a 1. Since the outcome of each roll is an independent trial, we have

$$\mathbb{P}(E_n) = \left(\frac{3}{6}\right)^{n-1} \cdot \frac{1}{6}$$

Since the events  $E_n$  are mutually exclusive, it follows that the required probability is

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(E_n) = \frac{1}{6} \sum_{n=1}^{\infty} \frac{1}{2^{n-1}} = \frac{1}{6} \cdot \frac{1}{1-1/2} = \frac{1}{3}$$

where we used the basic geometric series formula.

Alternatively, we could compute by letting *E* be the event that a 1 occurs before a 5 or 6 and conditioning on the outcome of the  $1^{st}$  trial: consider events

- *F* "the first trial is a 1"
- *G* "the first trial is a 2, 3 or 4"
- *H* "the first trial is a 5 or 6"

Then

$$\mathbb{P}(E) = \mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|G)\mathbb{P}(G) + \mathbb{P}(E|H)\mathbb{P}(H) = 1 \cdot \frac{1}{6} + \mathbb{P}(E) \cdot \frac{1}{2} + 0 \cdot \frac{1}{3}$$
$$\implies \mathbb{P}(E) = \frac{1}{3}$$

We used the fact that  $\mathbb{P}(E|G) = \mathbb{P}(E)$  since, if the first trial does not produce a conclusion, it is as if it never happened.

The example illustrates the following.

**Lemma 3.17.** Suppose *A*, *B* are mutually exclusive outcomes of a single trial. If independent trials are conducted, then the probability of *A* occurring before *B* is

$$\frac{\mathbb{P}(A)}{\mathbb{P}(A) + \mathbb{P}(B)} = \frac{\mathbb{P}(A)}{\mathbb{P}(A \cup B)}$$

This should be intuitive; we care only about the outcomes which end the experiment (A or B), and count those which produce the desired result (A). Note that this only works because the outcomes A, B of each trial were mutually exclusive.<sup>2</sup>

*Proof.* Let *E* be the event "*A* occurs before *B*." We condition on the outcome of the first trial, namely the mutually exclusive events

- *A*<sub>1</sub> "the first trial has outcome *A*"
- *B*<sub>1</sub> "the first trial has outcome *B*"
- $C = A_1^{\mathsf{C}} \cap B^{\mathsf{C}}$  "the first trial's outcome is neither A nor B"

We have  $\mathbb{P}(E|A) = 1$ ,  $\mathbb{P}(E|B) = 0$  and  $\mathbb{P}(E|C) = \mathbb{P}(E)$ , since if neither of the desired outcomes occurs in the first trial, it is as if it never happened. Now compute

$$\mathbb{P}(E) = \mathbb{P}(E|A_1)\mathbb{P}(A_1) + \mathbb{P}(E|B_1)\mathbb{P}(B_1) + \mathbb{P}(E|C)\mathbb{P}(C)$$
  
= 1 \cdot \mathbb{P}(A) + 0\mathbb{P}(B) + \mathbb{P}(E)(1 - \mathbb{P}(A) - \mathbb{P}(B))

from which the result is read off.

#### Introduction to the Binomial Distribution

Suppose we have a (biased) coin whose probability of landing heads is *p*, and suppose that separate coin flips are independent trials. The probability of tossing three heads, two tails, then one head is

$$\mathbb{P}(HHHTTH) = p^3(1-p)^2p = p^4(1-p)^2$$

Plainly it does not matter in which order the heads and tails appear; the probability of tossing four heads and two tails from six trials is therefore  $\binom{6}{4}p^4(1-p)^2$ , where the binomial coefficient counts how many subsets of the six trials have precisely four heads. We can ask this question more generally:

**Definition 3.18 (Binomial Distribution).** We perform *n* independent trials, each with probability of success *p*. If *X* is the number of successes from the *n* trials, then

$$\mathbb{P}\{X=k\} = \binom{n}{k} p^k (1-p)^{n-k}$$

We say that *X* is a *binomially distributed random variable*.

<sup>&</sup>lt;sup>2</sup>The problem is more one of English rather than Mathematics: if *A* and *B* could occur simultaneously ( $\mathbb{P}(A \cap B) \neq 0$ , which requires  $A \cap B \neq \emptyset$ ), what does it mean for *A* to happen *before B*? Should we really be asking for  $A \cap B^{\mathsf{C}}$  to happen before *B*? But these new events are now mutually exclusive, so the lemma applies!

To return to the previous example, the probability of rolling two 1's in five trials is

$$\mathbb{P}\{X=2\} = {\binom{5}{2}} \left(\frac{1}{6}\right)^2 \left(1-\frac{1}{6}\right)^{5-2} = 10 \cdot \frac{5^3}{6^5} = \frac{625}{3888} \approx 16.1\%$$

We'll return to the binomial distribution more formally in the next chapter. For the present, we return (again!) to Fermat & Pascal's problem of points.

**Example 3.19.** Let  $P_{a,b}$  denote the probability of *a* successes before *b* failures: i.e. the probability that player *A* wins the game. This requires *at least a* successes in the first a + b - 1 trials; If we let *X* be the number of successes in the first a + b - 1 trials, then the binomial distribution tells us that<sup>3</sup>

$$P_{a,b} = \sum_{k=a}^{a+b-1} \mathbb{P}\{X=k\} = \sum_{k=a}^{a+b-1} \binom{a+b-1}{k} p^k (1-p)^{a+b-1-k}$$

**The Gambler's Ruin Problem** We finish the chapter with another famous, and related, problem. We start small.

**Example 3.20.** Ava and Bruno roll a fair die. On a roll of 1 or 2, Ava pays Bruno \$1; on a 3, 4, 5 or 6, Bruno pays \$1 to Ava. If Ava starts with \$2, Bruno with \$8, and they repeat until one player has \$10, find the probability that Ava wins.

Ava has probability  $p = \frac{2}{3}$  of gaining \$1 on each roll. Since the rolls (trials) are independent, the probability that Ava wins depends only on how much money she *currently* has, not on what happened previously. Let

 $P_k$  be the probability that Ava wins given that she currently has k

We condition on the outcome of the next coin toss:

$$P_k = \frac{2}{3}P_{k+1} + \frac{1}{3}P_{k-1}, \qquad 1 \le k \le 9$$

with initial conditions  $P_0 = 0$  and  $P_{10} = 1$ . There is a sneaky trick to this; rewrite symmetrically

$$P_{k+1} - P_k = \frac{1}{2}(P_k - P_{k-1})$$

which becomes a simple recurrence for a new variable  $R_k := P_k - P_{k-1}$ ;

$$R_{k+1} = \frac{1}{2}R_k, \qquad 1 \le k \le 9, \qquad R_1 = P_1 - P_0 = P_1$$

We therefore have  $R_k = P_1 \left(\frac{1}{2}\right)^{k-1}$ , so we can compute  $P_k$  by summing a geometric sequence

$$P_{k} = (P_{k} - P_{k-1}) + (P_{k-1} - P_{k-2}) + \dots + (P_{1} - P_{0}) + P_{0} = \sum_{j=1}^{k} R_{j} = P_{1} \sum_{j=1}^{k} \frac{1}{2^{j-1}}$$
$$= P_{1} \cdot \frac{1 - (\frac{1}{2})^{k}}{1 - \frac{1}{2}} = 2P_{1} \left[ 1 - \left(\frac{1}{2}\right)^{k} \right]$$

<sup>3</sup>This is much easier than conditioning on the outcome of the first toss and solving a recurrence

$$P_{a,b} = \mathbb{P}(H)P_{a-1,b} + \mathbb{P}(T)P_{a,b-1} = pP_{a-1,b} + (1-p)P_{a,b-1}$$

subject to initial conditions  $P_{a,0} = 0$  and  $P_{0,b} = 1$ . This is similar to how we analyzed the problem in the introduction.

Using the initial condition  $P_{10} = 1$ , we obtain the solution

$$P_k = \frac{1 - \left(\frac{1}{2}\right)^k}{1 - \left(\frac{1}{2}\right)^{10}}$$

The answer to our original problem is that Ava wins with probability

$$P_2 = \frac{2^8(2^2 - 1)}{2^{10} - 1} = \frac{256 \cdot 3}{1023} = \frac{256}{341} \approx 75.1\%$$

It is perhaps surprising that the result is so lopsided, given that Ava started with far less money. In fact,  $P_1 = \frac{512}{1023} \approx 50.01\%$ , so even if Bruno only needs one more dollar, the outcome is finely balanced!

The general problem can be stated, and solved similarly.

**Theorem 3.21 (Gambler's Ruin).** *A and B play a game with the following rules.* 

- *A* starts with *k* points and *B* with *N k* points.
- Each round is an independent trial; A pays B one point with probability p; B pays A one point with probability q = 1 p.
- The game ends when one player has all N points.

The probability of player A winning the game is

$$P_k = \begin{cases} \frac{1 - (\frac{q}{p})^k}{1 - (\frac{q}{p})^N} & \text{if } p \neq \frac{1}{2} \\ \frac{k}{N} & \text{if } p = \frac{1}{2} \end{cases}$$

The example had N = 10 and  $p = \frac{2}{3}$ , whence  $\frac{q}{p} = \frac{1}{2}$ .

*Proof.* If  $p \neq \frac{1}{2}$ , then  $\frac{q}{p} \neq 1$  and the argument proceeds exactly as in the example; the recurrence becomes symmetric via

$$P_k = pP_{k+1} + qP_{k-1} \iff P_{k+1} - P_k = \frac{q}{p}(P_k - P_{k-1}), \quad 1 \le k \le N-1$$

This is solved in the same way by summing a geometric sequence and applying the initial conditions  $P_0 = 0$  and  $P_N = 1$ .

If  $p = q = \frac{1}{2}$ , then  $\frac{q}{p} = 1$  and the recurrence is

$$P_{k+1} - P_k = P_k - P_{k-1} \implies P_k = (P_k - P_{k-1}) + \dots + (P_1 - P_0) = k(P_1 - P_0) = kP_1$$

The initial condition  $P_N = 1$  forces  $P_1 = \frac{1}{N}$  and thus the result.