

## Chapter 1

# Probability: Basic concepts and theorems

The mathematical formalism of quantum mechanics is a probability calculus. The probability algorithms it places at our disposal—state vectors, wave functions, density matrices, statistical operators—all serve the same purpose, which is to calculate the probabilities of measurement outcomes. That's reason enough to begin by putting together what we already know and what we need to know about probabilities.

### 1.1 The principle of indifference

Probability is a measure of likelihood ranging from 0 to 1. If an event has a probability equal to 1, it is certain that it will happen; if it has a probability equal to 0, it is certain that it will not happen; and if it has a probability equal to  $1/2$ , then it is as likely as not that it will happen.

Tossing a fair coin yields heads with probability  $1/2$ . Casting a fair die yields any given natural number between 1 and 6 with probability  $1/6$ . These are just two examples of the *principle of indifference*, which states:

*If there are  $n$  mutually exclusive and jointly exhaustive possibilities (or possible events), and if we have no reason to consider any one of them more likely than any other, then each possibility should be assigned a probability equal to  $1/n$ .*

Saying that events are *mutually exclusive* is the same as saying that at most one of them happens. Saying that events are *jointly exhaustive* is the same as saying that at least one of them happens.

## 1.2 Subjective probabilities versus objective probabilities

There are two kinds of situations in which we may have no reason to consider one possibility more likely than another. In situations of the first kind, there are objective matters of fact that would make it certain, *if we knew them*, that a particular event will happen, but we don't know any of the relevant matters of fact. The probabilities we assign in this case, or whenever we know some but not all relevant facts, are in an obvious sense *subjective*. They are *ignorance* probabilities. They have everything to do with our (lack of) *knowledge* of relevant facts, but nothing with the *existence* of relevant facts. Therefore they are also known as *epistemic* probabilities.

In situations of the second kind, there are no objective matters of fact that would make it certain that a particular event will happen. There may not even be objective matters of fact that would make it more likely that one event will occur rather than another. There isn't any relevant fact that we are ignorant of. The probabilities we assign in this case are neither subjective nor epistemic. They deserve to be considered *objective*. Quantum-mechanical probabilities are essentially of this kind.

Until the advent of quantum mechanics, all probabilities were thought to be subjective. This had two unfortunate consequences. The first is that probabilities came to be thought of as something *intrinsically* subjective. The second is that something that was not a probability at all—namely, a *relative frequency*—came to be called an “objective probability.”

## 1.3 Relative frequencies

Relative frequencies are useful in that they allow us to measure the likelihood of possible events, at least approximately, provided that trials can be repeated under conditions that are identical in all relevant respects. We obviously cannot measure the likelihood of heads by tossing a single coin. But since we can toss a coin any number of times, we can count the number  $N_H$  of heads and the number  $N_T$  of tails obtained in  $N$  tosses and calculate the fraction  $f_N^H = N_H/N$  of heads and the fraction  $f_N^T = N_T/N$  of tails. And we can expect the difference  $|N_H - N_T|$  to increase significantly slower than the sum  $N = N_H + N_T$ , so that

$$\lim_{N \rightarrow \infty} \frac{|N_H - N_T|}{N_H + N_T} = \lim_{N \rightarrow \infty} |f_N^H - f_N^T| = 0. \quad (1.1)$$

In other words, we can expect the relative frequencies  $f_N^H$  and  $f_N^T$  to tend to the probabilities  $p_H$  of heads and  $p_T$  of tails, respectively:

$$p_H = \lim_{N \rightarrow \infty} \frac{N_H}{N}, \quad p_T = \lim_{N \rightarrow \infty} \frac{N_T}{N}. \quad (1.2)$$

#### 1.4 Adding and multiplying probabilities

Suppose you roll a (six-sided) die. And suppose you win if you throw either a 1 or a 6 (no matter which). Since there are six equiprobable outcomes, two of which cause you to win, your chances of winning are  $2/6$ . In this example it is appropriate to *add* probabilities:

$$p(1 \vee 6) = p(1) + p(6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}. \quad (1.3)$$

The symbol  $\vee$  means “or.” The general rule is this:

**Sum rule.** Let  $\mathcal{W}$  be a set of  $w$  mutually exclusive and jointly exhaustive events (for instance, the possible outcomes of a measurement), and let  $\mathcal{U}$  be a subset of  $\mathcal{W}$  containing a smaller number  $u$  of events:  $\mathcal{U} \subset \mathcal{W}$ ,  $u < w$ . The probability  $p(\mathcal{U})$  that one of the events  $e_1, \dots, e_u$  in  $\mathcal{U}$  takes place (no matter which) is the sum  $p_1 + \dots + p_u$  of the respective probabilities of these events.

One nice thing about relative frequencies is that they make a rule such as this virtually self-evident. To demonstrate this, let  $N$  be the total number of trials—think coin tosses or measurements. Let  $N_k$  be the total number of trials with outcome  $e_k$ , and let  $N(\mathcal{U})$  be the total number of trials with an outcome in  $\mathcal{U}$ . As  $N$  tends to infinity,  $N_k/N$  tends to  $p_k$  and  $N(\mathcal{U})/N$  tends to  $p(\mathcal{U})$ . But

$$\frac{N(\mathcal{U})}{N} = \frac{N_1 + \dots + N_u}{N} = \frac{N_1}{N} + \dots + \frac{N_u}{N}, \quad (1.4)$$

and in the limit  $N \rightarrow \infty$  this becomes

$$p(\mathcal{U}) = p_1 + \dots + p_u. \quad (1.5)$$

Suppose now that you roll two dice. And suppose that you win if your total equals 12. Since there are now  $6 \times 6$  equiprobable outcomes, only one of which causes you to win, your chances of winning are  $1/(6 \times 6)$ . In this example it is appropriate to *multiply* probabilities:

$$p(6 \wedge 6) = p(6) \times p(6) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}. \quad (1.6)$$

The symbol  $\wedge$  means “and.” Here is the general rule:

**Product rule.** The *joint probability*  $p(e_1 \wedge \dots \wedge e_v)$  of  $v$  *independent* events  $e_1, \dots, e_v$  (that is, the probability with which *all* of them happen) is the product of the probabilities  $p(e_1), \dots, p(e_v)$  of the individual events.

It must be stressed that the product rule only applies to independent events. Saying that two events  $a, b$  are *independent* is the same as saying that the probability of  $a$  is independent of whether or not  $b$  happens, and *vice versa*.

As an illustration of the product rule for two independent events, let  $a_1, \dots, a_J$  be mutually exclusive and jointly exhaustive events (think of the possible outcomes of a measurement of a variable  $A$ ), and let  $p_1^a, \dots, p_J^a$  be the corresponding probabilities. Let  $b_1, \dots, b_K$  be a second such set of events with corresponding probabilities  $p_1^b, \dots, p_K^b$ . Now draw a  $1 \times 1$  square with coordinates  $x, y$  ranging from 0 to 1. Partition it horizontally into  $J$  strips of respective width  $p_j^a$ . Partition it vertically into  $K$  strips of respective width  $p_k^b$ . You now have a square partitioned into  $J \times K$  rectangles with respective areas  $p_j^a \times p_k^b$ . Since a joint measurement of  $A$  and  $B$  is equivalent to throwing a dart in such a way that it hits a random position  $(x, y)$  within the square, the joint probability  $p(a_j \wedge b_k)$  equals the corresponding area.

**Problem 1.1.** *We have seen that the probability of obtaining a total of 12 when rolling a pair of dice is  $1/36$ . What is the probability of obtaining a total of (a) 11, (b) 10, (c) 9?*

**Problem 1.2.**  $(*)^1$  *In 1999, Sally Clark was convicted of murdering her first two babies, which died in their sleep of sudden infant death syndrome. She was sent to prison to serve two life sentences for murder, essentially on the testimony of an “expert” who told the jury it was too improbable that two children in one family would die of this rare syndrome, which has a probability of  $1/8,500$ . After over three years in prison, and five years of fighting in the legal system, Sally was cleared by a Court of Appeal, and another two and a half years later, the “expert” pediatrician Sir Roy Meadow was found guilty of serious professional misconduct. Amazingly, during the trial nobody raise the objection that an expert pediatrician was not likely to be an expert statistician. Meadow had argued that the probability of two sudden infant deaths in the same family was  $(1/8,500) \times (1/8,500) = 1/72,250,000$ . Explain why he was so terribly wrong.*

<sup>1</sup>A star indicates that a solution or a hint is provided in Appendix A.

### 1.5 Conditional probabilities and correlations

If the events  $a_j$  and  $b_k$  are *not* independent, we must distinguish between *marginal probabilities*, which are assigned to the possible outcomes of either measurement without taking account of the outcome of the other measurement, and *conditional probabilities*, which are assigned to the possible outcomes of either measurement depending on the outcome of the other measurement. If  $a_j$  and  $b_k$  are not independent, their joint probability is

$$p(a_j \wedge b_k) = p(b_k|a_j) p(a_j) = p(a_j|b_k) p(b_k), \quad (1.7)$$

where  $p(a_j)$  and  $p(b_k)$  are marginal probabilities, while  $p(b_k|a_j)$  is the probability of  $b_k$  conditional on the outcome  $a_j$  and  $p(a_j|b_k)$  is the probability of  $a_j$  conditional on the outcome  $b_k$ . This gives us the useful relation

$$p(b|a) = \frac{p(a \wedge b)}{p(a)}. \quad (1.8)$$

Another useful rule is

$$p(a) = p(a|b) p(b) + p(a|\bar{b}) p(\bar{b}), \quad (1.9)$$

where  $b$  and  $\bar{b}$  are two mutually exclusive and jointly exhaustive events. (To obtain  $\bar{b}$  is to obtain any outcome other than  $b$ .) The validity of this rule is again readily established with the help of relative frequencies. We obviously have that

$$\frac{N(a)}{N} = \frac{N(a \wedge b)}{N} + \frac{N(a \wedge \bar{b})}{N} = \frac{N(a \wedge b)}{N(b)} \frac{N(b)}{N} + \frac{N(a \wedge \bar{b})}{N(\bar{b})} \frac{N(\bar{b})}{N}, \quad (1.10)$$

where  $N$  is the number of joint measurements of two variables, one with the possible outcome  $a$  and one with the possible outcome  $b$ . In the limit  $N \rightarrow \infty$ ,  $N(a)/N$  (the left-hand side of Eq. 1.10) tends to the marginal probability  $p(a)$ , while the right-hand side of this equation tends to the right-hand side of Eq. (1.9), as will be obvious from a glance at Eq. (1.8).

An important concept is that of (probabilistic) correlation. Two events  $a, b$  are *correlated* just in case that  $p(a|b) \neq p(a|\bar{b})$ . Specifically,  $a$  and  $b$  are *positively* correlated if  $p(a|b) > p(a|\bar{b})$ , and they are *negatively* correlated if  $p(a|b) < p(a|\bar{b})$ . Saying that  $a$  and  $b$  are *independent* is thus the same as saying that they are *uncorrelated*, in which case  $p(a|b) = p(a|\bar{b}) = p(a)$ .

**Problem 1.3.** (\*) Let's Make a Deal was a famous game show hosted by Monty Hall. In it a player was to open one of three doors. Behind one door there was the Grand Prize (for example, a car). Behind the other doors

there were booby prizes (say, goats). After the player had chosen a door, the host opened a different door, revealing a goat, and offered the player the opportunity of choosing the other closed door. Should the player accept the offer or should he stick with his first choice? Does it make a difference?

**Problem 1.4.** (\*) Which of the following statements do you think is true? (i) Event A happens more frequently because it is more likely. (ii) Event A is more likely because it happens more frequently.

**Problem 1.5.** (\*) Suppose we have a 99% accurate test for a certain disease. And suppose that a person picked at random from the population tests positive. What is the probability that this person actually has the disease?

## 1.6 Expectation value and standard deviation

Another two important concepts associated with a probability distribution are the *expected/expectation value* (or *mean*) and the *standard deviation* (or *root mean square deviation from the mean*).

The expected value associated with the measurement of an observable with  $K$  possible outcomes  $v_k$  and corresponding probabilities  $p(v_k)$  is

$$\langle v \rangle \stackrel{\text{Def}}{=} \sum_{k=1}^K p(v_k) v_k. \quad (1.11)$$

Note that the expected value doesn't have to be one of the possible outcomes. The expected value associated with the roll of a die, for instance, equals 3.5.

To calculate the *rms* deviation from the mean,  $\Delta v$ , we first calculate the squared deviations from the mean,  $(v_k - \langle v \rangle)^2$ , then we calculate their mean, and finally we take the root:

$$\Delta v = \sqrt{\sum_{k=1}^K p(v_k)(v_k - \langle v \rangle)^2}. \quad (1.12)$$

The standard deviation of a random variable  $V$  with possible values  $v_k$  is an important measure—albeit not the only one—of the variability or spread of  $V$ .

**Problem 1.6.** (\*) Calculate the standard deviation for the sum obtained by rolling two dice.