Chapter 1
Basic Terminology

Abstract The concepts of random experiment, outcomes, sample space and events are introduced, and basic combinatorics (variations, permutations, combinations) is reviewed, leading to the exposition of fundamental properties of probability. A discussion of conditional probability is offered, followed by the definition of the independence of events and the derivation of the total probability and Bayes formulas.

1.1 Random Experiments and Events

A physics experiment can be envisioned as a process that maps the initial state (input) into the final state (output). Of course we wish such an experiment to be *non-random:* during the measurement we strive to control all external conditions—input data, the measurement process itself, as well as the analysis of output data—and justly expect that each repetition of the experiment with an identical initial state and in equal circumstances will yield the same result.

In a *random experiment*, on the other hand, it *may* happen that multiple repeats of the experiment with the same input and under equal external conditions will end up in different outputs. The main feature of a random experiment is therefore our inability to uniquely predict the precise final state based on input data. We rather ask ourselves about the *frequency of occurrence* of a specific final state with respect to the number of trials. That is why this number should be as large as possible: we shall assume that, in principle, a random experiment can be repeated infinitely many times.

A specific output of a random experiments is called an *outcome*. An example of an outcome is the number of photons measured by a detector, e.g. 12. The set of all possible outcomes of a random experiment is called the *sample space*, S. In the detector example, the sample space is the set $S = \{0, 1, 2, \ldots\}$. Any subset of the sample space is called an *event*. Individual outcomes are *elementary* events. Elementary events can be joined in *compound events:* for example, the detector sees more than 10 photons (11 or 12 or 13, and so on) or sees 10 photons and less than 20 neutrons simultaneously.

S. Širca, *Probability for Physicists*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-31611-6_1

The events, elementary or compound, are denoted by letters $A, B, C, \ldots$ The event that occurs in all repetitions of the experiment—or can be assumed to occur in all future tries—is called a *certain* or *universal event* and is denoted by U. The event that does not occur in any repetition of the experiment is called an *impossible event*, denoted by $\varnothing$ or $\{\,\}$. The relations between events can be expressed in the language of set theory. Take two events A and B and consider the possibility that at least one of them occurs: this eventuality is called the *sum of events* and is denoted by

$$A \cup B.$$

Summing events is commutative and associative: we have $A \cup B = B \cup A$ and $(A \cup B) \cup C = A \cup (B \cup C)$. The sum of two events can be generalized: the event that at least one of the events $A_1, A_2, \ldots, A_n$ occurs, is

$$A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{k=1}^{n} A_k.$$

The event that both A and B occur simultaneously, is called the *product of events A* and B. It is written as

$$A \cap B$$

or simply

$$AB.$$

For each event A one obviously has $A\varnothing = \varnothing$ and $AU = A$. The product of events is also commutative and associative; it holds that $AB = BA$ and $(AB)C = A(BC)$. The compound event that all events $A_1, A_2, \ldots, A_n$ occur simultaneously, is

$$A_1 A_2 \ldots A_n = \bigcap_{k=1}^{n} A_k.$$

The addition and multiplication are related by the distributive rule $(A \cup B)C = AC \cup BC$. The event that A occurs but B does not, is called the *difference of events* and is denoted by

$$A - B.$$

(In general $A - B \neq B - A$.) The events A and B are *exclusive* or *incompatible* if they can not occur simultaneously, that is, if

$$AB = \varnothing.$$

The events A are B *complementary* if in each repetition of the experiment precisely one of them occurs: this implies

$$AB = \varnothing \quad \text{and} \quad A \cup B = U.$$

An event complementary to event A is denoted by $\bar{A}$. Hence, for any event A,

$$A\bar{A} = \varnothing \quad \text{and} \quad A \cup \bar{A} = U.$$

Sums of events in which individual pairs of terms are mutually exclusive, are particularly appealing. Such sums are denoted by a special sign:

$$A \cup B \stackrel{\text{def.}}{=} A + B \quad \Leftrightarrow \quad A \cap B = \{\,\}.$$

Event sums can be expressed as sums of incompatible terms:

$$A_1 \cup A_2 \cup \cdots \cup A_n = A_1 + \bar{A}_1 A_2 + \bar{A}_1 \bar{A}_2 A_3 + \cdots + \left(\bar{A}_1 \bar{A}_2 \ldots \bar{A}_{n-1} A_n\right). \tag{1.1}$$

The set of events

$$\{A_1, A_2, \ldots, A_n\} \tag{1.2}$$

is called the *complete set of events*, if in each repetition of the experiment precisely one of the events contained in it occurs. The events from a complete set are all possible ($A_i \neq \varnothing$), pair-wise incompatible ($A_i A_j = \varnothing$ for $i \neq j$), and their sum is a certain event: $A_1 + A_2 + \cdots + A_n = U$, where n may be infinite.

Example There are six possible outcomes in throwing a die: the sample space is $S = \{1, 2, 3, 4, 5, 6\}$. The event A of throwing an odd number—the compound event consisting of outcomes $\{1\}$, $\{3\}$ or $\{5\}$—corresponds to $A = \{1, 3, 5\}$, while for even numbers $B = \{2, 4, 6\}$. The sum of A and B exhausts the whole sample space; $A \cup B = S = U$ implies a certain event. The event of throwing a 7 is impossible: it is not contained in the sample space at all. ◁

Example A coin is tossed twice, yielding either head (h) or tail (t) in each toss. The sample space of this random experiment is $S = \{\text{hh}, \text{ht}, \text{th}, \text{tt}\}$. Let A represent the event that in two tosses we get at least one head, $A = \{\text{hh}, \text{ht}, \text{th}\}$, and let B represent the event that the second toss results in a tail, thus $B = \{\text{ht}, \text{tt}\}$. The event that at least one of A and B occurs (i.e. A or B or both) is

$$A \cup B = \{\text{hh}, \text{ht}, \text{th}, \text{tt}\}.$$

We got $A \cup B = S$ but that does not hold in general: if, for example, one would demand event B to yield two heads, $B = \{\text{hh}\}$, one would obtain $A \cup B = \{\text{hh}, \text{ht}, \text{th}\} = A$. The event of A and B occurring simultaneously is

$$A \cap B = AB = \{\text{ht}\}.$$

This implies that A and B are *not* exclusive, otherwise we would have obtained $AB = \{\ \} = \varnothing$. The event that A occurs but B does *not* occur is

$$A - B = A \cap \bar{B} = \{\text{hh, ht, th}\} \cap \{\text{hh, th}\} = \{\text{hh, th}\}.$$

The complementary event to A is $\bar{A} = S - A = \{\text{tt}\}$. ◁

The sample spaces in the above examples are discrete. An illustration of a continuous one can be found in thermal implantation of ions into quartz (SiO_2) in the fabrication of chips. The motion of ions in the crystal is diffusive and the ions penetrate to different depths: the sample space for the depths over which a certain concentration profile builds up is, say, the interval $S = [0, 1]\,\mu$m.

1.2 Basic Combinatorics

1.2.1 Variations and Permutations

We perform m experiments, of which the first has n_1 possible outcomes, the second has n_2 outcomes for each outcome of the first, the third has n_3 outcomes for each outcome of the first two, and so on. The number of possible outcomes of all m experiments is

$$n_1 n_2 n_3 \ldots n_m.$$

If $n_i = n$ for all i, the number of all possible outcomes is simply

$$n^m.$$

Example A questionnaire contains five questions with three possible answers each, and ten questions with five possible answers each. In how many ways the questionnaire can be filled out if exactly one answer is allowed for each question? By the above formulas, in no less than $3^5 5^{10} = 2373046875$ ways. ◁

What if we have n different objects and are interested in how many ways (that is, *variations*) m objects from this set can be reshuffled, paying attention to their *ordering*? The first object can be chosen in n ways. Now, the second one can only be chosen from the reduced set of $n - 1$ objects, ..., and the last object from the remaining $n - m + 1$. The number of variations is then

$$n(n-1)\cdots(n-m+1) = \frac{n!}{(n-m)!} = {}_nV_m = (n)_m. \qquad (1.3)$$

The symbol on the right is known as the Pochammer symbol.

Example The letters A, B, C and D ($n = 4$) can be assembled in groups of two ($m = 2$) in $4!/2! = 12$ ways: {AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, DC}. Note that in this procedure, ordering is crucial: AB does not equal BA. ◁

A special case of (1.3) is $m = n$ when variations are called *permutations:* the number of permutations of n objects is

$$n(n-1)(n-2)\cdots 3\cdot 2\cdot 1 = n! = P_n.$$

Speaking in reverse, $n!$ is the number of all permutations of n objects, while (1.3) is the number of *ordered* sub-sequences of length m from these n objects.

Example We would like to arrange ten books (four physics, three mathematics, two chemistry books and a dictionary) on a shelf such that the books from the same field remain together. For each possible arrangement of the fields we have $4!\,3!\,2!\,1!$ options, while the fields themselves can be arranged in $4!$ ways, hence there are a total of $1!\,2!\,3!\,4!\,4! = 6912$ possibilities. ◁

We are often interested in the permutations of n objects, n_1 of which are of one kind and indistinguishable, n_2 of another kind ..., n_m of the mth kind, while $n = n_1 + n_2 + \cdots + n_m$. From all $n!$ permutations the indistinguishable ones $n_1!$, $n_2!$ ... must be removed, hence the required number of permutations is $n!/(n_1!\,n_2!\cdots n_m!)$ and is denoted by the *multinomial symbol:*

$$\frac{n!}{n_1!\,n_2!\dots n_m!} = {}_nP_{n_1,n_2,\dots,n_m} = \binom{n}{n_1, n_2, \dots, n_m}. \tag{1.4}$$

1.2.2 Combinations Without Repetition

In how many ways can we arrange n objects into different groups of m objects if the ordering is irrelevant? (For example, the letters A, B, C, D and E in groups of three.) Based on previous considerations leading to (1.3) we would expect $n(n-1)\cdots(n-m+1)$ variations. But in doing this, equal groups would be counted multiple ($m!$) times: the letters A, B and C, for example, would form $m! = 3! = 6$ groups ABC, ACB, BAC, BCA, CAB and CBA, in which the letters are just mixed. Thus the desired number of arrangements—in this case called *combinations of mth order among n elements without repetition*—is

$$\frac{n(n-1)\cdots(n-m+1)}{m!} = \frac{n!}{(n-m)!\,m!} = {}_nC_m = \frac{{}_nV_m}{P_m} = \binom{n}{m}. \tag{1.5}$$

The symbol at the extreme right is called the *binomial symbol*. It can not hurt to recall its parade discipline, the *binomial formula*

$$(x+y)^n = \sum_{m=0}^{n} \binom{n}{m} x^{n-m} y^m. \tag{1.6}$$

1.2.3 Combinations with Repetition

In combinations *with repetition* we allow the elements to appear multiple times, for example, in combining four letters (A, B, C and D) into groups of three, where not only triplets with different elements like ABC or ABD, but also the options AAA, AAB and so on should be counted. The following combinations are allowed:

AAA, AAB, AAC, AAD, ABB, ABC, ABD, ACC, ACD, ADD,
BBB, BBC, BBD, BCC, BCD, BDD, CCC, CCD, CDD, DDD.

In general the *number of combinations of* m*th order among* n *elements with repetition* is

$$\frac{(n+m-1)!}{(n-1)!\,m!} = \binom{n+m-1}{m}. \tag{1.7}$$

In the example above ($n = 4$, $m = 3$) one indeed has $6!/(3!\,3!) = 20$.

1.3 Properties of Probability

A random experiment always leaves us in doubt whether an event will occur or not. A measure of probability with which an event may be expected to occur is its relative frequency. It can be calculated by applying "common sense", i.e. by dividing the number of chosen ("good") events A to occur, by the number of all encountered events: in throwing a die there are six possible outcomes, three of which yield odd numbers, so the relative frequency of the event A = "odd number of points" should be $P(A) = \text{good}/\text{all} = 3/6 = 0.5$. One may also proceed pragmatically: throw the die a thousand times and count, say, 513 odd and 487 even outcomes. The empirical relative frequency of the odd result is therefore $513/1000 = 0.513$. Of course this value will fluctuate if a die is thrown a thousand times again, and yet again—to 0.505, 0.477, 0.498 and so on. But we have reason to believe that after many, many trials the value will stabilize at the previously established value of 0.5.

We therefore define the probability $P(A)$ of event A in a random experiment as the value at which the relative frequency of A usually stabilizes after the experiment has been repeated many times[1] (see also Appendix A). Obviously

$$0 \le P(A) \le 1.$$

The probability of a certain event is one, $P(U) = 1$. For any event A we have

$$P(A) + P(\bar{A}) = 1,$$

hence also $P(\varnothing) = 1 - P(U) = 0$: the probability of an impossible event is zero. For arbitrary events A and B the following relation holds:

$$P(A \cup B) = P(A) + P(B) - P(AB). \tag{1.8}$$

For exclusive events, $AB = \varnothing$ and the equation above reduces to

$$P(A + B) = P(A) + P(B),$$

which can be generalized for pair-wise exclusive events as

$$P(A \cup B \cup C \cup \cdots) = P(A) + P(B) + P(C) + \cdots .$$

To generalize (1.8) to multiple events one only needs to throw a glance at (1.1): for example, with three events A, B and C we read off

$$\begin{aligned} A \cup B \cup C &= A + \bar{A}B + \bar{A}\bar{B}C \\ &= A + (U - A)B + (U - A)(U - B)C \\ &= A + B + C - AB - AC - BC + ABC, \end{aligned}$$

therefore also

$$\begin{aligned} &P(A \cup B \cup C) \\ &\quad = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC). \end{aligned} \tag{1.9}$$

Example (Adapted from [3].) In the semiconductor wafer production impurities populate the upper layers of the substrate. In the analysis of 1000 samples one finds a large concentration of impurities in 113 wafers that were near the ion source during the process, and in 294 wafers that were at a greater distance from it. A low concentration is found in 520 samples from near the source and 73 samples that were farther away. What is the probability that a randomly selected wafer was near the source during the production (event N), or that it contains a large concentration of impurities (event L), or both?

---

[1] This is the so-called *frequentist approach* to probability, in contrast to the *Bayesian approach:* an introduction to the latter is offered by [2].

We can answer the question by carefully counting the measurements satisfying the condition: $P(N \cup L) = (113 + 294 + 520)/1000 = 0.927$. Of course, (1.8) leads to the same conclusion: the probability of N is $P(N) = (113 + 520)/1000 = 0.633$, the probability of L is $P(L) = (113 + 294)/1000 = 0.407$, while the probability of N and L occurring simultaneously—they are not exclusive!—is $P(NL) = 113/1000 = 0.113$, hence

$$P(N \cup L) = P(N) + P(L) - P(NL) = 0.633 + 0.407 - 0.113 = 0.927.$$

Ignoring the last term, $P(NL)$, is a frequent mistake which, however, is easily caught as it leads to probability being greater than one. ◁

Example (Adapted from [4].) A detector of cosmic rays consists of nine smaller independent sub-detectors all pointing in the same direction of the sky. Suppose that the probability for the detection of a cosmic ray shower (event E) by the individual sub-detector—the so-called detection efficiency—is $P(E) = \varepsilon = 90\%$. If we require that the shower is seen by all sub-detectors simultaneously (nine-fold coincidence, Fig. 1.1 (left)), the probability to detect the shower (event X) is

$$P(X) = \big(P(E)\big)^9 \approx 0.387.$$

The sub-detectors can also be wired in three triplets, where a favorable outcome is defined by at least one sub-detector in the triplet observing the shower. Only then a

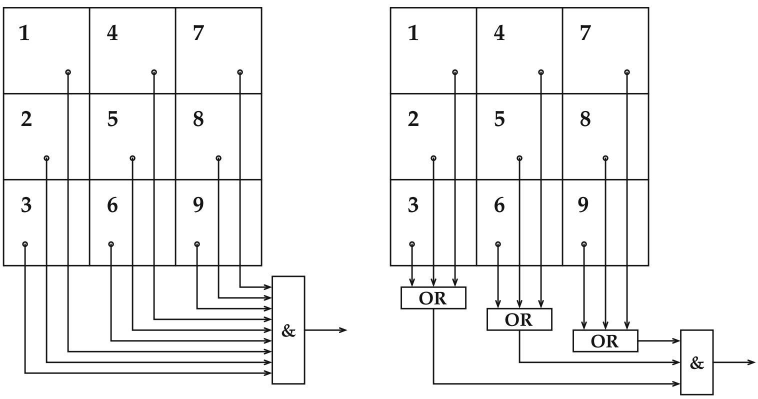


Fig. 1.1 Detector of cosmic rays. [Left] Sub-detectors wired in a nine-fold coincidence. [Right] Triplets of sub-detectors wired in a three-fold coincidence

triple coincidence is formed from the three resulting signals (Fig. 1.1 (right)). In this case the total shower detection probability is

$$P(X) = \left(P(E_1 \cup E_2 \cup E_3)\right)^3 = \left(3\varepsilon - 3\varepsilon^2 + \varepsilon^3\right)^3 \approx 0.997,$$

where we have used (1.9). ◁

1.4 Conditional Probability

Let A be an event in a random experiment (call it ‘first’) running under a certain set of conditions, and $P(A)$ its probability. Imagine another event B that may occur in this or another experiment. What is the probability $P'(A)$ of event A if B is interpreted as an additional condition for the first experiment? Because event B modifies the set of conditions, we are now actually performing a new experiment differing from the first one, thus we generally expect $P(A) \neq P'(A)$. The probability $P'(A)$ is called the *conditional probability* of event A *under the condition* B or *given event* B, and we appropriately denote it by $P(A|B)$. This probability is easy to compute: in n repetitions of the experiment with the augmented set of conditions B occurs n_B times, while $A \cap B$ occurs n_{AB} times, therefore

$$P(A|B) = \lim_{n\to\infty} \frac{n_{AB}/n}{n_B/n} = \frac{P(AB)}{P(B)}.$$

The conditional probability for A given B ($P(B) \neq 0$) is therefore computed by dividing the probability of the simultaneous event, $A \cap B$, by $P(B)$. Obviously, the reverse is also true:

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

Both relations can be merged into a single statement known as the *theorem on the probability of the product of events* or simply the *product formula:*

$$P(AB) = P(B|A)P(A) = P(A|B)P(B). \tag{1.10}$$

The first part of the equation can be verbalized as follows: the probability that A and B occur simultaneously equals the product of probabilities that A occurs first, and the probability that B occurs, given that A has already occurred. (The second part proceeds analogously.)

The theorem can be generalized to multiple events. Let $A_1, A_2, \ldots, A_n$ be arbitrary events and let $P(A_1 A_2 \ldots A_n) > 0$. Then

$$P(A_1 A_2 \ldots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \ldots P(A_n|A_1 A_2 \ldots A_{n-1}). \tag{1.11}$$

Perhaps the essence becomes even clearer if we reverse the ordering of the factors and digest the formula from right to left:

$$P(A_n \dots A_2 A_1) = P(A_n|A_{n-1} \dots A_2 A_1) \dots P(A_3|A_2 A_1) P(A_2|A_1) P(A_1).$$

Example What is the probability that throwing a die yields a number of spots which is less than four *given that* the number is odd? Let A mean "odd number of spots" ($P(A) = 1/2$), and B "the number of spots less than four" ($P(B) = 1/2$). If A and B occur *simultaneously*, the probability of the compound event can be inferred from the intersection of sets A and B in Fig. 1.2 (left): it is

$$P(AB) = \tfrac{2}{6} = \tfrac{1}{3},$$

since only elements $\{1, 3\}$ inhabit the intersection, while the complete sample space is $\{1, 2, 3, 4, 5, 6\}$. But this is not yet the answer to our question! We are interested in the probability of B once A ("the condition") has already occurred: this implies that the sample space has shrunk to $\{1, 3, 5\}$ as shown in Fig. 1.2 (right). From this reduced space we need to pick the elements that fulfill the requirement B: they are $\{1, 3\}$ and therefore

$$P(B|A) = \tfrac{2}{3}.$$

Equation (1.10) says the same: $P(B|A) = P(AB)/P(A) = \frac{1}{3}\big/\frac{1}{2} = \frac{2}{3}$. We can imagine that the unconditional probability $P(B) = 1/2$ has increased to $P(B|A) = 2/3$ by the additional *information* that the throw yields an odd number. ◁

Example A box in our cellar holds 32 bottles of wine, eight of which are spoiled. We randomly select four bottles from the box for today's dinner. What is the probability that not a single one will be spoiled?

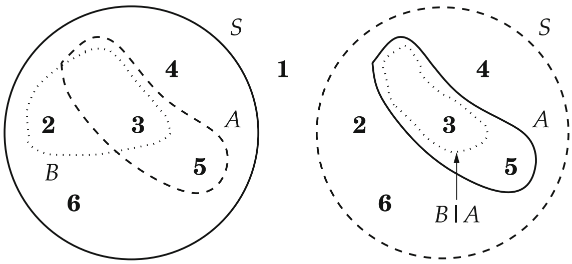


Fig. 1.2 The conditional probability in throwing a die. [Left] The probability of events A and B occurring *simultaneously* corresponds to the intersection of the sets $\{1, 3, 5\}$ and $\{1, 2, 3\}$ within the complete sample space S. [Right] The condition A first isolates the set $\{1, 3, 5\}$ from the complete S. The conditional probability of B given A corresponds to the fraction of the elements in this set that also fulfill the requirement B

This can be solved in two ways. The first method is to apply the product formula by considering that with each new bottle fetched from the box, both the total number of bottles and the number of spoiled bottles in it are reduced by one. Let A_i denote the event that the ith chosen bottle is good, and A the event that all four bottles are fine. The probability of the first bottle being good is $P(A_1) = 24/32$. This leaves 31 bottles in the box, 23 of which are good, hence the probability of the second bottle being intact is $P(A_2|A_1) = 23/31$. Analogously $P(A_3|A_1A_2) = 22/30$ and $P(A_4|A_1A_2A_3) = 21/29$ for the third and fourth bottle, respectively. Formula (1.11) then gives

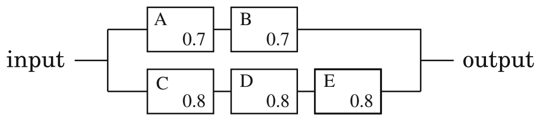
$$\begin{aligned} P(A) = P(A_1A_2A_3A_4) &= P(A_1)P(A_2|A_1)P(A_3|A_1A_2)P(A_4|A_1A_2A_3) \\ &= \frac{24}{32}\frac{23}{31}\frac{22}{30}\frac{21}{29} \approx 0.2955. \end{aligned}$$

The second option is to count the number of ways in which 24 *good* bottles can be arranged in four places: it is equal to $24!/(4!\,20!)$ (see (1.5)). But this number must be divided by the number of *all* possible combinations of bottles in four places, which is $32!/(4!\,28!)$. The probability of four bottles being good is then

$$P(A) = \frac{24!}{4!\,20!}\frac{4!\,28!}{32!} \approx 0.2955.$$

◁

Example An electric circuit has five independent elements with various degrees of reliability—probabilities that an element functions—shown in the figure.



What is the probability that the circuit works (transmits signals from input to output) and the probability that A does not work, given that the circuit works?

Let us denote the event "element A works" by A (and analogously for the elements B, C, D and E). The circuit works (event V) when the elements A and B work *or* the elements C, D and E work *or* all five of them work, hence

$$\begin{aligned} P(V) = P(AB \cup CDE) &= P(AB) + P(CDE) - P(ABCDE) \\ &= P(A)P(B) + P(C)P(D)P(E) \\ &\quad -P(A)P(B)P(C)P(D)P(E) \\ &= (0.7)^2 + (0.8)^3 - (0.7)^2(0.8)^3 = 0.75112, \end{aligned}$$

where we have used (1.8). The probability that A has failed (event $\bar{A}$), given that the circuit works, is obtained by the following consideration, noting that $XY = X \cap Y$. We first calculate the probability that A does not work while the circuit as a whole works. If A has failed, then the bottom branch of the circuit *must* work. But even if A is inoperational, *two* options remain for B: it either works or it does not. Thus $\bar{A} \cap V = \left[\left(\bar{A} \cap B\right) \cup \left(\bar{A} \cap \bar{B}\right)\right] \cap (C \cap D \cap E)$. Thus the conditional probability we have been looking for is

$$P\left(\bar{A}|V\right) = \frac{P\left(\bar{A}V\right)}{P(V)} = \frac{P\left[\left(\bar{A}B \cup \bar{A}\bar{B}\right)\left(CDE\right)\right]}{P(V)} = \frac{\left[P\left(\bar{A}B\right) + P\left(\bar{A}\bar{B}\right)\right] \cdot P(CDE)}{P(V)}$$
$$= \frac{[(1-0.7)0.7 + (1-0.7)^2] \cdot (0.8)^3}{0.75112} = 0.2045,$$

where we have used $\bar{A} \cap B \cap \bar{A} \cap \bar{B} = \{\,\}$, since $A \cap \bar{A} = B \cap \bar{B} = \{\,\}$. ◁

1.4.1 Independent Events

If events A and B are *independent*, the probability that A occurs (or does not occur) is independent of whether we have any information on B (and vice-versa), hence

$$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B).$$

According to (1.10), the probability that such events occur simultaneously equals the product of probabilities of them occurring individually:

$$P(AB) = P(A)P(B). \tag{1.12}$$

When more than two events are involved, independence must be defined more carefully. The events in the set

$$\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$$

are *mutually* or *completely independent* if, for every combination $(i_1, i_2, \ldots, i_k)$ of kth order without repetition $(k = 2, 3, \ldots, n)$ among the numbers $1, 2, \ldots, n$, it holds that

$$P\left(A_{i_1}A_{i_2}\ldots A_{i_k}\right) = P\left(A_{i_1}\right)P\left(A_{i_2}\right)\ldots P\left(A_{i_k}\right). \tag{1.13}$$

When $k = n$ this system of equations has the form

$$P(A_1A_2\ldots A_n) = P(A_1)P(A_2)\ldots P(A_n),$$

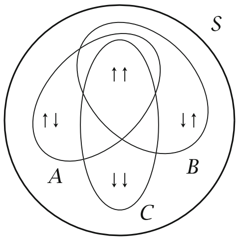
which is a special case of (1.11); when $k = 2$, the leftover of (1.13) is simply

$$P(A_iA_j) = P(A_i)P(A_j).$$

If (1.12) applies to any pair of events in $\mathcal{A}$, we say that such events are *pair-wise independent*, but this is still a far cry from *mutual* (complete) independence! There are 2^n combinations without repetition among n elements (see (1.6) with $x = y = 1$). One of them corresponds to the empty set, while there are n combinations of the first order, as we learn from (1.5). The system above therefore imposes $2^n - n - 1$ conditions that must be fulfilled by the events from $\mathcal{A}$ in order for them to be mutually independent. In the special case $n = 3$ there are four such conditions:

$$\begin{aligned} P(A_1A_2) &= P(A_1)P(A_2), \\ P(A_1A_3) &= P(A_1)P(A_3), \\ P(A_2A_3) &= P(A_2)P(A_3), \\ P(A_1A_2A_3) &= P(A_1)P(A_2)P(A_3). \end{aligned}$$

This important distinction between pair-wise and mutual independence is discussed in the following Example.



Example The spin in a quantum system can have two projections: $+\frac{1}{2}$ (spin “up”, $\uparrow$) or $-\frac{1}{2}$ (spin “down”, $\downarrow$). The orientation of the spin is measured twice in a row. We make the following event assignments: event A means “spin $\uparrow$ in the first measurement”, event B is “spin $\uparrow$ in the second measurement”, and event C is “both measurements show the same projection”. The sample space for the measured pairs of orientations is $S = \{\uparrow\uparrow, \uparrow\downarrow, \downarrow\uparrow, \downarrow\downarrow\}$, while the chosen three events correspond to its subsets $A = \{\uparrow\uparrow, \uparrow\downarrow\}$, $B = \{\uparrow\uparrow, \downarrow\uparrow\}$ and $C = \{\uparrow\uparrow, \downarrow\downarrow\}$ shown in the Figure. We immediately obtain the probabilities

$$P(A) = P(B) = P(C) = \tfrac{2}{4} = \tfrac{1}{2},$$

as well as

$$P(AB) = P(AC) = P(BC) = \tfrac{1}{4} \quad \text{and} \quad P(ABC) = \tfrac{1}{4}.$$

Since

$$P(AB) = P(A)P(B) = P(AC) = P(A)P(C) = P(BC) = P(B)P(C),$$

events A, B and C are pair-wise independent. On the other hand,

$$P(ABC) = \tfrac{1}{4} \neq \tfrac{1}{8} = P(A)P(B)P(C),$$

so the events are *not* mutually independent. ◁

1.4.2 Bayes Formula

When an event A occurs under different, mutually exclusive conditions, and we know the conditional probabilities of A given all these conditions, we can also calculate the unconditional probability of A. The two-condition case is illustrated by the following classic insurance-company example.

Example An insurance company classifies the drivers into those deemed less (85%) and those more accident-prone (15%). These are two mutually exclusive 'conditions'—call them B and $\bar{B}$—that exhaust all options, as there is no third class, thus $P(B) = 0.85$, $P(\bar{B}) = 0.15$. On average, a first-class driver causes a crash every 10 years, and the second-class driver once in 5 years. Let A denote the event of an accident, regardless of its cause. The probability for a first-tier driver to cause a crash within a year is $P(A|B) = 1/10$, while it is $P(A|\bar{B}) = 1/5$ for the second-tier driver. What is the probability that a new customer will cause an accident within the first year? Since for any A and B, $A = (A \cap B) \cup (A \cap \bar{B})$, we also have

$$P(A) = P(AB) + P(A\bar{B}),$$

while from (1.10) it follows that

$$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B}). \tag{1.14}$$

Statistically, the company may therefore expect the probability of

$$P(A) = 0.1 \cdot 0.85 + 0.2 \cdot 0.15 = 0.115$$

for a newly insured driver to cause an accident within a year. ◁

Equation (1.14) is a sort of weighted average over both driver classes, where the weights depend on conditions B and $\bar{B}$. Suppose that there are more such mutually exclusive conditions: we then prefer to call them *assumptions* or *hypotheses* and denote them by H_i: we have H_1 or H_2 ... or H_n, exhausting all possibilities. The set of all H_i constitutes a complete set defined by (1.2), hence

$$P(A) = P(AH_1) + P(AH_2) + \cdots + P(AH_n).$$

Applying the left-hand side of (1.10) to each term separately yields the so-called *total probability formula*

$$P(A) = P(A|H_1)P(H_1) + P(A|H_2)P(H_2) + \cdots + P(A|H_n)P(H_n), \qquad (1.15)$$

illustrated in Fig. 1.3.

Let us recall (1.10) once more, this time in its second part, whence one reads off $P(H_i|A)P(A) = P(A|H_i)P(H_i)$ or

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{P(A)}.$$

The denominator of this expression is given by (1.15) and the final result is the famous Bayes formula [5]

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{P(A|H_1)P(H_1) + \cdots + P(A|H_n)P(H_n)}, \quad i = 1, 2, \ldots, n. \qquad (1.16)$$

A random experiment may repeatedly yield events A, but the events H_i conditioning A—with corresponding probabilities $P(H_i)$—occurred *prior to* A. The quantities $P(H_i)$ are therefore called *prior probabilities* since they are, in principle, known in advance. In contrast, the left side of the Bayes formula gives the probability that the hypothesis H_i is valid with respect to the *later* outcome A. The conditional probability $P(H_i|A)$ is called *posterior*, since it uses the present outcome A to specify the probability of H_i occurring prior to A. This is why the Bayes formula is also known as the *theorem on probability of hypotheses.*

Example A company decides to manufacture cell-phones by using processor chips of different suppliers. The first type of chip is built into 70%, the second into 20%, and the third into 10% of devices. A randomly chosen device contains chip i (event C_i) with probability $P(C_i)$, where $P(C_1) = 0.7$, $P(C_2) = 0.2$ and $P(C_3) = 0.1$: these are the known prior probabilities. Some chips are unreliable, causing the devices to malfunction. The probability that a cell-phone breaks down (event A), given

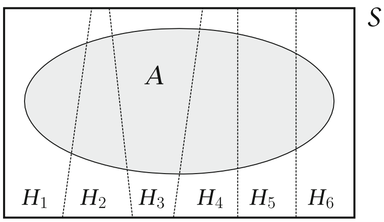


Fig. 1.3 Illustration of the total probability formula. Mutually exclusive conditions or hypotheses H_i are disjoint sets that partition the sample space S and therefore also an arbitrary event A

that it contains chip i, is $P(A|C_i)$. The manufacturer establishes $P(A|C_1) = 0.01$, $P(A|C_2) = 0.03$ and $P(A|C_3) = 0.05$.

We go to a store and buy a cell-phone made by this company. It breaks down immediately (event A at this very moment). What is the probability that it was manufactured (event C_i in the past) in the factory installing type-i chips ($i = 1, 2, 3$)? We are looking for the posterior probabilities $P(C_i|A)$ given by the Bayes formula. Its denominator contains $P(A) = \sum_{i=1}^{3} P(A|C_i)P(C_i) = 0.01 \cdot 0.7 + 0.03 \cdot 0.2 + 0.05 \cdot 0.1 = 0.018$, which is common to all three cases—and this is the probability that the cell-phone breaks down. This leads to

$$P(C_1|A) = \frac{P(A|C_1)P(C_1)}{P(A)} = \frac{0.01 \cdot 0.7}{0.018} \approx 38.9\%,$$
$$P(C_2|A) = \frac{P(A|C_2)P(C_2)}{P(A)} = \frac{0.03 \cdot 0.2}{0.018} \approx 33.3\%,$$
$$P(C_3|A) = \frac{P(A|C_3)P(C_3)}{P(A)} = \frac{0.05 \cdot 0.1}{0.018} \approx 27.8\%.$$

Of course we also have $P(C_1|A) + P(C_2|A) + P(C_3|A) = 1$. ◁

1.5 Problems

1.5.1 Boltzmann, Bose–Einstein and Fermi–Dirac Distributions

(Adapted from [1].) Imagine a system of n particles in which the state of each particle is described by p values (components of the position vector or linear momentum, spin quantum number, and so on). Each particle state can be represented by such a p-plet, which is a point in p-dimensional space. The state of the whole system is uniquely specified by a n-plet of such points.

Let us divide the phase space into N ($N \geq n$) cells. The state of the system is described by specifying the distribution of states among the cells. We are interested in the probability that a given cell is occupied by the prescribed number of particles. Consider three options: ① The particles are distinguishable, each cell can be occupied by an arbitrary number of particles, and all such distributions are equally probable. We then say that the particles “obey” Boltzmann statistics: an example of such a system are gas molecules. ② The particles are *indistinguishable*, but the cells may still be occupied by arbitrary many particles and all such distributions are equally probable. This is the foundation of Bose–Einstein statistics obeyed by particles with integer spins (bosons), e.g. photons. ③ The particles are indistinguishable, each cell may accommodate *at most one particle* due to the Pauli principle [6]. All distributions are equally probable. This case refers to the Fermi–Dirac statistics applicable to particles with half-integer spins (fermions), e.g. electrons, protons and neutrons.

✎ Let A_k be the event that there are precisely k particles ($0 \le k \le n$) in a certain cell, regardless of their distribution in other cells. ① Each of the n particles can be put into any of the N cells, even if other particles are already sitting there. All particles can therefore be arranged in N^n ways and this is the number of all possible outcomes. How many correspond to event A_k? Into the chosen cell one can pour k particles in $\binom{n}{k}$ ways, while the remaining $n-k$ particles can be arranged into the other $N-1$ cells in $(N-1)^{n-k}$ ways. Event A_k therefore accommodates $\binom{n}{k}(N-1)^{n-k}$ outcomes, thus

$$P(A_k) = \binom{n}{k}(N-1)^{n-k}\frac{1}{N^n} = \binom{n}{k}\left(\frac{1}{N}\right)^k\left(1-\frac{1}{N}\right)^{n-k}.$$

② Since particles are indistinguishable and each cell is allowed to swallow an arbitrary number of particles, the number of all possible distributions equals the number of combinations of nth order among N elements with repetition (1.7), i.e. $\binom{N+n-1}{n}$. How many are acceptable for A_k? Event A_k occurs precisely when k particles are selected for a given cell—since they are indistinguishable, this can be accomplished in one way only—while the remaining $n-k$ are distributed among $N-1$ cells, which amounts to combinations of order $n-k$ among $N-1$ elements with repetition, i.e. $\binom{N+n-k-2}{n-k}$. It follows that

$$P(A_k) = \binom{N+n-k-2}{n-k}\bigg/\binom{N+n-1}{n}.$$

③ Since at most one particle is allowed to occupy any single cell, all possible distributions can be counted by choosing n cells out of N and putting one particle into every one of them: this can be accomplished in $\binom{N}{n}$ ways. How many of them correspond to event A_k? For $k > 1$ there are none, while for $k = 0$ or $k = 1$ there are as many ways as one can arrange $n-k$ particles over $N-1$ cells, which is $\binom{N-1}{n-k}$. Therefore

$$P(A_k) = \binom{N-1}{n-k}\bigg/\binom{N}{n} = \begin{cases} 1-\dfrac{n}{N} & ;\ k=0, \\ \dfrac{n}{N} & ;\ k=1, \end{cases}$$

while $P(A_k) = 0$ for $k > 1$. Figure 1.4 (left) shows the probabilities $P(A_k)$ for all three distributions in the case $N = 15$, $n = 5$, while Fig. 1.4 (right) shows the Boltzmann and the Bose–Einstein distribution in the case $N = 100$, $n = 10$.

1.5.2 Blood Types

The fractions of blood types O, A, B and AB in the whole population are

$$\mathrm{O}:\ 44\%,\quad \mathrm{A}:\ 42\%,\quad \mathrm{B}:\ 10\%,\quad \mathrm{AB}:\ 4\%.$$

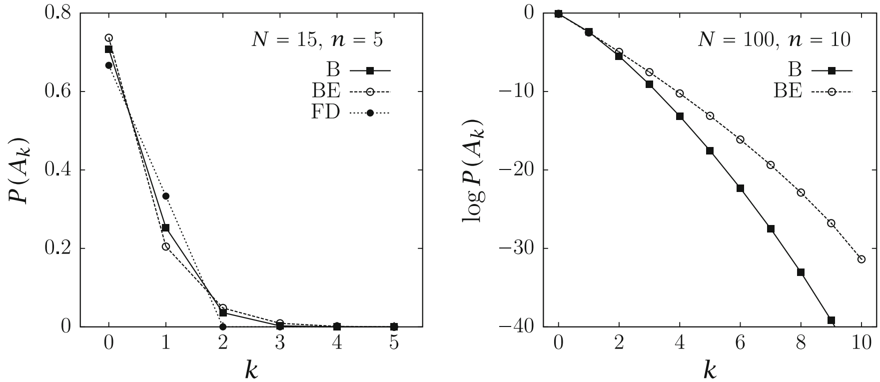


Fig. 1.4 The probability of finding k particles in any chosen cell of a N-cell phase space flooded with n particles, in the case of Boltzmann (B), Bose–Einstein (BE) and Fermi–Dirac (FD) statistics. [Left] $N = 15, n = 5$. The sum of all probabilities within a given distribution of course equals 1, as it is obvious e.g. in the case of the Fermi–Dirac distribution: $P(A_0) = 1 - \frac{5}{15} = \frac{2}{3}, P(A_1) = \frac{5}{15} = \frac{1}{3}$. [Right] $N = 100, n = 10$

① Two persons are picked at random from the population. What is the probability of their having the same blood type, and what is the probability that their types differ? ② We pick four people from the same population. What is the probability that precisely k ($k = 1, 2, 3, 4$) blood types will be found among them?

✎ Let us replace the letter notation O, A, B, AB by indices 1, 2, 3, 4, and let P_i denote the probability that a person has blood type i ($i = 1, 2, 3, 4$). ① All possible pairs are $\{i, i\}$, $i = 1, 2, 3, 4$, each having probability P_i^2, therefore $P = P_1^2 + P_2^2 + P_3^2 + P_4^2 = 0.44^2 + 0.42^2 + 0.1^2 + 0.04^2 = 0.3816$. The complementary event has probability $1 - P = 0.6184$ which, in a more arduous manner, can be computed as:

$$\begin{aligned} 1 - P &= P_1(P_2 + P_3 + P_4) + P_2(P_1 + P_3 + P_4) + P_3(P_1 + P_2 + P_4) \\ &\quad + P_4(P_1 + P_2 + P_3) \\ &= 2\big[P_1(P_2 + P_3 + P_4) + P_2(P_3 + P_4) + P_3 P_4\big] = 0.6184. \end{aligned}$$

② Let $P(k)$ denote the probability that precisely k blood types will be found in the chosen four. For $k = 1$ the quartets are $\{i, i, i, i\}$, $i = 1, 2, 3, 4$, hence $P(1) = \sum_i P_i^4 = 0.44^4 + 0.42^4 + 0.1^4 + 0.04^4 = 0.0687$. For $k = 2$ we use (1.4) to obtain the number of possible combinations in samples of the form $\{i, j, j, j\}$ ($i \neq j$), which is $N_{13} = 4!/(1!\,3!) = 4$, and the number of combinations in samples of the form $\{i, i, j, j\}$ ($i \neq j$), which is $N_{22} = 4!/(2!\,2!) = 6$. We get

$$P(2) = N_{13} \sum_{i \neq j} P_i^1 P_j^3 + N_{22} \sum_{i<j} P_i^2 P_j^2 = 0.3665 + 0.2308 = 0.5973.$$

The calculation for $k = 3$ is tedious and is best avoided by calculating the probability for $k = 4$, which is $P(4) = 4! \cdot P_1 P_2 P_3 P_4 = 0.0177$, and accumulating all previously computed $P(k)$ into the complementary event: $P(3) = 1 - P(1) - P(2) - P(4) = 0.3163$.

1.5.3 Independence of Events in Particle Detection

Two detectors are used to detect charged particles with different parities (mirror symmetries of their wave-functions): pions (π^+ and π^-) and kaons (K^+ and K^-), all possessing negative parity, as well as protons (p), deuterons (d) and ^{3}He and ^{4}He nuclei, all of which have positive parities. Assume that all particles appear with equal frequencies and assign indices $\{1, 2, 3, 4, 5, 6, 7, 8\}$ to types $\{\pi^+, \pi^-, K^+, K^-, p, d, {}^3\text{He}, {}^4\text{He}\}$. Let A denote the event that the first detector has seen a negative-parity particle. Let B denote the event that the second detector has detected a positive-parity particle, and suppose that

$$P(A) = P(A|B) = \tfrac{4}{8} = \tfrac{1}{2},$$
$$P(B) = P(B|A) = \tfrac{4}{8} = \tfrac{1}{2}.$$

Let C denote the event that both detectors observe particles with equal parities. Are events A, B and C (pair-wise or mutually) independent?

✎ There are 64 equally probable outcomes (i, j) in an experiment where the first and second detector detect particles i and j, respectively; 16 of them are pion-kaon combinations fulfilling condition C:

$$(1,1),\ (1,2),\ (1,3),\ (1,4),\ (2,1),\ (2,2),\ (2,3),\ (2,4),$$
$$(3,1),\ (3,2),\ (3,3),\ (3,4),\ (4,1),\ (4,2),\ (4,3),\ (4,4),$$

as well as 16 combinations of atomic nuclei,

$$(5,5),\ (5,6),\ (5,7),\ (5,8),\ (6,5),\ (6,6),\ (6,7),\ (6,8),$$
$$(7,5),\ (7,6),\ (7,7),\ (7,8),\ (8,5),\ (8,6),\ (8,7),\ (8,8),$$

thus $P(C) = (16 + 16)/64 = \frac{1}{2}$. Suppose that the first detector has seen a negative-parity particle and has thereby imposed condition A: then C occurs if the second detector also reports a negative-parity particle (probability 1/2), implying $P(C|A) = 1/2$. Analogously we conclude $P(C|B) = 1/2$, and finally

$$P(C) = P(C|A) = P(C|B) = \tfrac{1}{2}.$$

We conclude that A, B and C are pair-wise but not mutually independent since $P(ABC) = P(A)P(B)P(C)$ does not hold true. Our calculation shows that $P(A)P(B)P(C) = \frac{1}{8}$, while $A \cap B \cap C$ is an impossible event: if there is a negative-parity particle in the first detector and a positive-parity particle in the second one, we can not have the same parity in both detectors, thus $P(ABC) = 0$.

How do these considerations change if the detectors are inefficient in detecting heavier nuclei (^{3}He and ^{4}He)? Do events A, B and C remain independent? How does the result change in physically more sensible circumstances in which the number of pions exceeds the number of kaons by a factor of 100?

1.5.4 Searching for the Lost Plane

The authorities believe that an airliner has been lost in one of the three regions R_i $(i = 1, 2, 3)$ in which the crash has occurred with equal probability, $P(R_i) = 1/3$. Let P_i denote the probability that the plane search in region i will locate the plane that actually does lie in i. Calculate the conditional probability that the plane crashed in region i, given that the search in region 1 was unsuccessful!

✎ Let R_i $(i = 1, 2, 3)$ denote the event that the plane went down in region i, and N the event that the search in region 1 was unsuccessful. Bayes formula for $i = 1$ gives

$$P(R_1|N) = \frac{P(NR_1)}{P(N)} = \frac{P(N|R_1)P(R_1)}{\sum_{i=1}^{3} P(N|R_i)P(R_i)}$$
$$= \frac{(1-P_1)\frac{1}{3}}{(1-P_1)\frac{1}{3} + 1\cdot\frac{1}{3} + 1\cdot\frac{1}{3}} = \frac{1-P_1}{3-P_1},$$

while for $i = 2$ and $i = 3$ one gets

$$P(R_i|N) = \frac{P(N|R_i)P(R_i)}{P(N)} = \frac{1\cdot\frac{1}{3}}{(1-P_1)\frac{1}{3} + 1\cdot\frac{1}{3} + 1\cdot\frac{1}{3}} = \frac{1}{3-P_1}, \quad i = 2, 3,$$

where we have exploited the fact that the search in region 1 *must* be unsuccessful if the plane lies in region 2 or 3, hence $P(N|R_2) = P(N|R_3) = 1$. For example, if $P_1 = 0.7$, the probability that the plane is in region 1—given that it has *not* been found in it—is $0.3/2.3 \approx 13\%$. Note that $\sum_i P(R_i|N) = 1$.

1.5.5 The Monty Hall Problem ⋆

In the Monty Hall TV show with three boxes (adapted from [7, 8]) one box contains the car keys while the remaining boxes are empty. When the contestant picks one of the boxes (e.g. box 1), Monty Hall (MH) tells her: "I'll make you a favor and open one of the remaining boxes that *does not contain the keys* (e.g. 2). Thus the keys are either in your chosen box or in box 3, so the probability of your winning the car has increased from 1/3 to 1/2." The contestant (C) responds: "I've changed my mind. I prefer to pick box 3 instead of box 1."

① Is Monty's claim correct? What is the probability of the contestant winning the car if she changes her mind following Monty's disclosure, and what is her chance of winning if she insists on her initial choice? ② Suppose that the contestant has been playing this game for a long time and knows that different boxes have different probabilities of containing the keys, e.g. 50, 40 and 10% for boxes 1, 2 and 3. What is the most promising strategy in this case?

✎ Two observations are crucial: MH *knows* which box contains the keys and obviously does not wish to reveal it; he opens one of the two remaining boxes at random and with equal probability. The answer to ① can then be obtained by simple counting of possible outcomes shown in Table 1.1: 'W' means that the contestant 'wins', 'L' means 'loses'. (All information is contained in the first three rows since the rest consists just of cyclic permutations.) The probability of C winning the car when insisting on the initial choice is 1/3. The probability of winning the car after having changed her mind is 2/3. Consequently, Monty's claim is false.

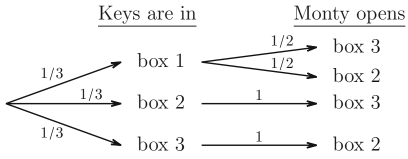
The problem can be approached from another, more intuitive viewpoint [9]. Suppose C decides to *always* switch. If she chooses an empty box, she can not lose: MH is then obliged to open the other empty box, so, by switching, C gets the only remaining box—the one containing the keys. C loses only if she initially chooses the box with the keys. Whether this strategy of "perpetual switching" works depends only on the initial choice of the empty box (probability 2/3) or the box containing the keys (probability 1/3).

Table 1.1 Possible outcomes in the Monty Hall contest

Keys are in	C picks	MH opens	Outcome	C switches	Outcome
1	1	2 or 3	W	1 for 3 or 2	L
1	2	3	L	2 for 1	W
1	3	2	L	3 for 1	W
2	1	3	L	1 for 2	W
2	2	1 or 3	W	2 for 3 or 1	L
2	3	1	L	3 for 2	W
3	1	2	L	1 for 3	W
3	2	1	L	2 for 3	W
3	3	1 or 2	W	3 for 2 or 1	L

Both contestant's strategies are shown: "C picks" means the one and only choice of the box, while "C switches" means that the contestant selects a different box after Monty's disclosure

Conditional probability offers yet another vantage point. Suppose that C chooses box 1 while the keys are in box 2 (event A, $P(A) = 1/3$). MH opens box 3 (event B). The graph



then tells us that $P(B|A) = 1$ and $P(B) = \frac{1}{3}\frac{1}{2} + \frac{1}{3}1 = \frac{1}{2}$, hence, by Bayes formula,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

In two thirds of the cases the keys are in the remaining box, so C doubles her $1/3$ chance of winning by switching. The same conclusion can be reached by analyzing the sample space in which the events are not equally probable. Denote all possible outcomes by (i, j), where i is the box containing the keys, j is the box opened by MH, and let P_{ij} denote the corresponding probability for such an outcome. When we shall later become familiar with the concept of random variables, all these values will be merged into the expression

$$X \sim \begin{pmatrix} (i,j) & \cdots \\ P_{ij} & \cdots \end{pmatrix} = \begin{pmatrix} (1,3) & (2,3) & (1,2) & (3,2) \\ 1/6 & 1/3 & 1/6 & 1/3 \end{pmatrix} \tag{1.17}$$

which we shall read as: "The discrete variable X is distributed such that the probability of outcome $(1, 3)$ is $P_{13} = 1/6$, the probability of outcome $(2, 3)$ is $P_{23} = 1/3$, and so on." In a compact manner, however, we can write down the sample space with the probability values attached as subscripts:

$$S = \left\{(1,3)_{1/6},\ (2,3)_{1/3},\ (1,2)_{1/6},\ (3,2)_{1/3}\right\}.$$

In this notation, $A = \left\{(2,3)_{1/3}\right\}$ and $B = \left\{(1,3)_{1/6}, (2,3)_{1/3}\right\}$. Since $A \subset B$, we have

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{6}} = \frac{2}{3}.$$

Monty's tempting offer to increase the contestant's probability of winning to $1/2$ is based on the wrong assumption that the remaining two possible events $(1, 3)$ and $(2, 3)$ are equally probable—i.e. that the sample space after the condition A has been imposed is $\left\{(1,3)_{1/2},\ (2,3)_{1/2}\right\}$—leading to the wrong result $P(A|B) = 1/2$.

The best strategy for ② is: C should choose the *least* probable box (box 3); when MH reveals an empty box, C should switch. In this case C will win 90% of the time.

Table 1.2 Conditional probabilities in a diagnostic test that can be negative when the disease is absent (specificity $\mathcal{R}$), negative in spite of the disease (false negative), positive with no disease (false positive) or positive with the disease present (sensitivity $\mathcal{O}$)

	Disease absent	Disease present
Negative test	$P(L\vert\overline{D}) = \mathcal{R}$	$P(L\vert D) = 1 - \mathcal{O}$
Positive test	$P(H\vert\overline{D}) = 1 - \mathcal{R}$	$P(H\vert D) = \mathcal{O}$

The limiting case that box 3 *never* holds the keys is also covered: MH reveals the other empty box so, by switching, C always wins.

1.5.6 Bayes Formula in Medical Diagnostics

We have fallen ill with fever and visit a doctor. Recently he has read some news on the west Nile virus that, on average, infects one person per million. He draws a blood sample for a test that has a positive outcome in $\mathcal{O} = 99\%$ of the cases where the disease is actually present (the so-called *sensitivity* of the test), and a negative outcome in $\mathcal{R} = 95\%$ of the cases where the disease is not present (the so-called *specificity* of the test). The test of our blood comes out positive. ① What is the probability that we are actually infected by the virus? ② Analyze the more general case of a disease probed by a larger number of tests or exhibiting multiple symptoms.

✎ Let us denote the positive outcome of the test by H ("high titer") and negative by L ("low titer") and write the corresponding conditional probabilities in Table 1.2. Now just read it carefully. ① The probability that the test is positive and the disease (D) is in fact present, is indeed $P(\text{high titer}|\text{infected}) = P(H|D) = \mathcal{O} = 99\%$. But the probability that we are actually infected by the virus, given the test was positive, is $P(D|H)$, and can be computed by using the Bayes formula (1.16):

$$P(D|H) = \frac{P(H|D)P(D)}{P(H|D)P(D) + P(H|\overline{D})P(\overline{D})} = \frac{\mathcal{O}P(D)}{\mathcal{O}P(D) + (1-\mathcal{R})P(\overline{D})},$$

where we have used $P(\text{low titer}|\text{not infected}) = P(L|\overline{D}) = \mathcal{R}$ and thus, due to the complementarity of H and L, $P(H|\overline{D}) = 1 - P(L|\overline{D}) = 1 - \mathcal{R}$. But the numerator also contains the prior probability that, as a random member of the population, we catch the disease at all, which is $P(D) = 10^{-6}$. This results in a very small probability

$$\begin{aligned} P(D|H) &= \frac{0.99 \times 10^{-6}}{0.99 \times 10^{-6} + 0.05\,(1 - 10^{-6})} \\ &= \frac{9.9 \times 10^{-7}}{9.9 \times 10^{-7} + 0.04999995} \approx 1.98 \times 10^{-5}. \end{aligned}$$

② When a disease manifests itself in several symptoms or tests ($S = S_1 \cap S_2 \cap \cdots \cap S_m$), the posterior probability for the disease is still given by the Bayes formula

$$P(D|S) = \frac{P(S|D)P(D)}{P(S)} = \frac{P(S_1 S_2 \ldots S_m|D)P(D)}{P(S_1 S_2 \ldots S_m)},$$

but it becomes useless in practical cases. Namely, for a specific disease D_j from a set of n diseases and a single symptom S one would have the expression

$$P(D_j|S) = \frac{P(S|D_j)P(D_j)}{\sum_{k=1}^{n} P(S|D_k)P(D_k)},$$

which becomes much more complex by adding new symptoms. With each new symptom S_m added to the previous set of symptoms $S_1, S_2, \ldots, S_{m-1}$, one would have to compute

$$P\big(D_j \big| S_1 S_2 \ldots S_m\big) = \frac{P\big(S_m \big| D_j S_1 S_2 \ldots S_{m-1}\big)\, P\big(D_j \big| S_1 S_2 \ldots S_{m-1}\big)}{\sum_{k=1}^{n} P\big(S_m \big| D_k S_1 S_2 \ldots S_{m-1}\big)\, P\big(D_k | S_1 S_2 \ldots S_{m-1}\big)}.$$

For a diagnostic system incorporating, say, 50 diseases and 500 symptoms that may occur individually or collectively in any of these diseases, we would require the data on $n \cdot 2^m = 50 \cdot 2^{500} \approx 10^{152}$ conditional probabilities. In the so-called *naive Bayes approach* one therefore frequently assumes that the symptoms are independent in the sense that

$$P(S_i|S_j) = P(S_i), \quad P(S_i|DS_j) = P(S_i|D).$$

The first equation states that the probability for S_i to appear in a part of the population that also exhibits symptom S_j is equal to the probability of S_i appearing in the whole population. The second approximation says that the probability of S_i appearing in a part of the population that has the disease D and some other symptom S_j, is equal to the probability of S_i appearing in *all* persons having the disease D. These simplifications allow us to operate with far fewer conditional probabilities $P(S_i|D_j)$—only $m \cdot n = 50 \cdot 500 = 25{,}000$ in the example above—expressing the probability of S_i given the presence of the disease D_j [10]:

$$P\big(D_j \big| S_1 S_2 \ldots S_m\big) \approx \frac{\prod_{i=1}^{m} P(S_i|D_j)P(D_j)}{\prod_{i=1}^{m} \sum_{k=1}^{n} P(S_i|D_k)P(D_k)}.$$

Inevitably, the assumption of symptom independence is quite coarse: given the presence of the disease, the probability of two symptoms appearing simultaneously is larger than the product of probabilities of individual symptoms. (If we have a headache and know it was caused by the flu, we will most likely develop a sore throat as well.)

1.5.7 One-Dimensional Random Walk ⋆

(Adapted from [1].) A particle moves along the real axis, starting at the origin ($x = 0$). Consecutive random collisions uniformly spaced in time send it one step to the left (-1) or to the right ($+1$) with probabilities $1/2$ either way. ① What is the probability that after $2n$ collisions the particle will return to $x = 0$ without ever meandering into the $x < 0$ region? Five random walks are shown for illustration in Fig. 1.5 (left). For example, walk number 3 that has always remained at $x \geq 0$ and has terminated at $x = 0$ after 100 collisions is "acceptable". ② Verify your result by a computer simulation. (Random walks will be discussed more generally in Sects. 6.7 and 6.8.)

✎ Each random walk is a consequence of $2n$ collisions. Each collision shifts the particle to the left ($x \mapsto x - 1$) or to the right ($x \mapsto x + 1$), thus the number of all possible walks is 2^{2n}. Let A be the event that the particle returns to the origin after $2n$ collisions, and B the event that the particle does not wander to $x < 0$ during $2n$ collisions. We are looking for the probability $P(AB)$, where $P(AB) = P(B|A)P(A)$.

① Let us first determine $P(A)$. From 2^{2n} possible and equally probable walks only those are acceptable for event A that end up at $(2n, 0)$, like the walk in Fig. 1.6 denoted by the full line. In all of them the particle has experienced n unit kicks to the left and n unit kicks to the right. The number of all such walks can be calculated by counting all possible ways of choosing n collisions that result in a left (or right) shift, from the total $2n$ collisions. There are $\binom{2n}{n}$ such ways, therefore

$$P(A) = \frac{1}{2^{2n}}\binom{2n}{n}.$$

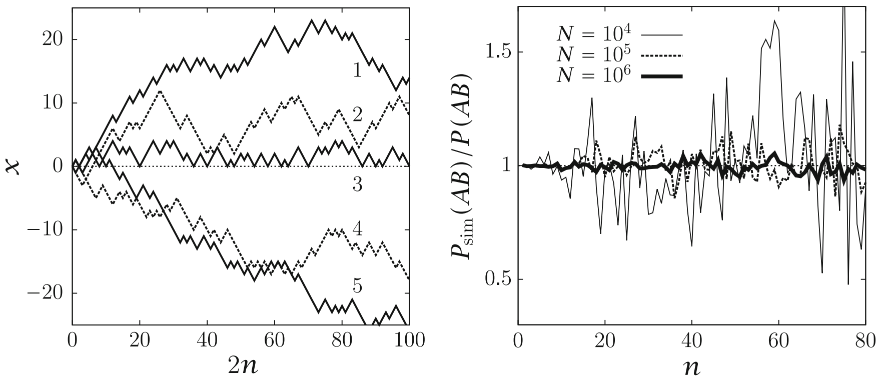


Fig. 1.5 [Left] Five one-dimensional random walks with 100 time steps. We are looking for the fraction of the walks that terminate at the origin (event A) and never blunder to $x < 0$ (condition B), as in walk number 3 shown here. [Right] The ratio between the simulated and theoretical expectation value for event AB

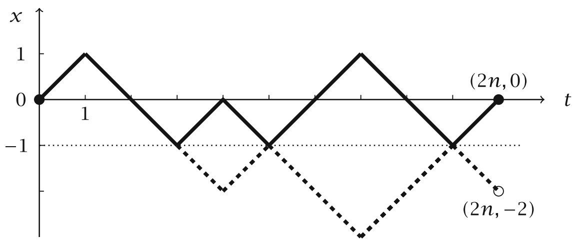


Fig. 1.6 A random walk that enters the region $x < 0$ after a certain time—still returning to the origin after $2n$ steps—and its mirror image from that moment on

From the walks ending up at $(2n, 0)$ and thereby fulfilling condition A, we should disregard those that fluctuate to $x < 0$ if we wish to satisfy condition B. How do we count such occurrences? For each such walk (from the very moment it has crossed the boundary and reached the point $x = -1$) we imagine a new walk, which is the mirror image of the remainder of the previous walk across the $x = -1$ axis (dashed line in Fig. 1.6). The new walk certainly terminates at $(2n, -2)$ and is therefore composed of $n-1$ right and $n+1$ left shifts. Hence, under condition A, $\binom{2n}{n+1}$ do not fulfill B, while $\binom{2n}{n} - \binom{2n}{n+1}$ do. This implies that

$$P(B|A) = \frac{\binom{2n}{n} - \binom{2n}{n+1}}{\binom{2n}{n}}.$$

The probability we have been looking for is therefore

$$P(AB) = P(B|A)P(A) = \frac{1}{2^{2n}}\left[\binom{2n}{n} - \binom{2n}{n+1}\right] = \frac{1}{2^{2n}(n+1)}\binom{2n}{n}. \tag{1.18}$$

② You do not trust this calculation? Let us try to check it by a simple computer simulation. For each n chosen in advance, start with a particle at the origin, then randomly add $+1$ or -1 to its current position and write down its final coordinate after $2n$ steps. A walk that ends up at $x = 0$ and has never erred into $x < 0$ is counted as "good". If for each n we repeat N walks, we may expect that the ratio of the good walks and all attempted walks will approach the calculated probability (1.18) in the limit $N \to \infty$. Let us denote this simulated probability by $P_{\mathrm{sim}}(AB)$. Figure 1.5 (right) shows the ratio between $P_{\mathrm{sim}}(AB)$ and the theoretical $P(AB)$ as a function of the walk duration n for three different numbers N of how many times the simulation was re-run. Apparently our calculation was correct: with increasing N the ratio does stabilize near 1. The thick line in the figure still looks wiggly? It is! Recall that for $n = 80$ there are $2^{160} \approx 10^{48}$ all possible walks, while we have performed only a million of them at each n.

References

1. R. Jamnik, *Verjetnostni račun* (Mladinska knjiga, Ljubljana, 1971)
2. P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge University Press, Cambridge, 2005)
3. D.C. Montgomery, G.C. Runger, *Applied Statistics and Probability for Engineers*, 5th edn. (Wiley, New York, 2010)
4. A.G. Frodesen, O. Skjeggestad, H. Tøfte, *Probability and Statistics in Particle Physics* (Universitetsforlaget, Bergen, 1979)
5. T. Bayes, An essay toward solving a problem in the doctrine of chances. Philos. Trans. **53**, 370 (1763)
6. J.J. Brehm, W.J. Mullin, *Introduction to the Structure of Matter* (Wiley, New York, 1989)
7. S. Selvin, A problem in probability. Am. Stat. **29**, 67 (1975)
8. S. Selvin, On the Monty Hall problem. Am. Stat. **29**, 134 (1975)
9. M.A. Carlton, Pedigrees, prizes, and prisoners: the misuse of conditional probability. J. Stat. Educ. **13**(2) (2005)
10. S. Schwartz, J. Baron, J. Clarke, A casual Bayesian model for the diagnosis of appendicitis, *Conference on Uncertainty in Artificial Intelligence (UAI-86)* (Elsevier Science Publishers, Amsterdam, 1986), p. 423

http://www.springer.com/978-3-319-31609-3

Probability for Physicists
Sirca, S.
2016, XIV, 415 p. 142 illus., 29 illus. in color., Hardcover
ISBN: 978-3-319-31609-3