Cambridge University Press
978-1-108-41585-9 — Introduction to Probability
David F. Anderson , Timo Seppäläinen , Benedek Valkó
Excerpt
More Information

1

Experiments with random outcomes

The purpose of probability theory is to build mathematical models of experiments with random outcomes and then analyze these models. A random outcome is anything we cannot predict with certainty, such as the flip of a coin, the roll of a die, the gender of a baby, or the future value of an investment.

1.1. Sample spaces and probabilities

The mathematical model of a random phenomenon has standard ingredients. We describe these ingredients abstractly and then illustrate them with examples.

Definition 1.1. These are the ingredients of a probability model.

- The **sample space** Ω is the set of all the possible outcomes of the experiment. Elements of Ω are called **sample points** and typically denoted by ω.
- Subsets of Ω are called **events**. The collection of events in Ω is denoted by $\mathcal{F}$. ♣
- The **probability measure** (also called **probability distribution** or simply **probability**) P is a function from $\mathcal{F}$ into the real numbers. Each event A has a probability $P(A)$, and P satisfies the following axioms.
  (i) $0 \le P(A) \le 1$ for each event A.
  (ii) $P(\Omega) = 1$ and $P(\varnothing) = 0$.
  (iii) If $A_1, A_2, A_3, \dots$ is a sequence of pairwise disjoint events then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \tag{1.1}$$

The triple $(\Omega, \mathcal{F}, P)$ is called a **probability space**. Every mathematically precise model of a random experiment or collection of experiments must be of this kind.

The three axioms related to the probability measure P in Definition 1.1 are known as *Kolmogorov's axioms* after the Russian mathematician Andrey Kolmogorov who first formulated them in the early 1930s.

Cambridge University Press
978-1-108-41585-9 — Introduction to Probability
David F. Anderson , Timo Seppäläinen , Benedek Valkó
Excerpt
More Information

A few words about the symbols and conventions. Ω is an upper case omega, and ω is a lower case omega. $\varnothing$ is the empty set, that is, the subset of Ω that contains no sample points. The only sensible value for its probability is zero. *Pairwise disjoint* means that $A_i \cap A_j = \varnothing$ for each pair of indices $i \neq j$. Another way to say this is that the events A_i are *mutually exclusive*. Axiom (iii) says that the probability of the union of mutually exclusive events is equal to the sum of their probabilities. Note that rule (iii) applies also to finitely many events.

Fact 1.2. If $A_1, A_2, \dots, A_n$ are pairwise disjoint events then

$$P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n). \tag{1.2}$$

Fact 1.2 is a consequence of (1.1) obtained by setting $A_{n+1} = A_{n+2} = A_{n+3} = \cdots = \varnothing$. If you need a refresher on set theory, see Appendix B.

Now for some examples.

Example 1.3. We flip a fair coin. The sample space is $\Omega = \{\mathtt{H}, \mathtt{T}\}$ ($\mathtt{H}$ for heads and $\mathtt{T}$ for tails). We take $\mathcal{F} = \{\varnothing, \{\mathtt{H}\}, \{\mathtt{T}\}, \{\mathtt{H}, \mathtt{T}\}\}$, the collection of all subsets of Ω. The term "fair coin" means that the two outcomes are equally likely. So the probabilities of the singletons $\{\mathtt{H}\}$ and $\{\mathtt{T}\}$ are

$$P\{\mathtt{H}\} = P\{\mathtt{T}\} = \tfrac{1}{2}.$$

By axiom (ii) in Definition 1.1 we have $P(\varnothing) = 0$ and $P\{\mathtt{H}, \mathtt{T}\} = 1$. Note that the "fairness" of the coin is an assumption we make about the experiment. ▲

Example 1.4. We roll a standard six-sided die. Then the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. Each sample point ω is an integer between 1 and 6. If the die is fair then each outcome is equally likely, in other words

$$P\{1\} = P\{2\} = P\{3\} = P\{4\} = P\{5\} = P\{6\} = \tfrac{1}{6}.$$

A possible event in this sample space is

$$A = \{\text{the outcome is even}\} = \{2, 4, 6\}. \tag{1.3}$$

Then

$$P(A) = P\{2, 4, 6\} = P\{2\} + P\{4\} + P\{6\} = \tfrac{1}{2}$$

where we applied Fact 1.2 in the second equality. ▲

Some comments about the notation. In mathematics, sets are typically denoted by upper case letters A, B, etc., and so we use upper case letters to denote events. Like A in (1.3), events can often be expressed both in words and in mathematical symbols. The description of a set (or event) in terms of words or mathematical symbols is enclosed in braces $\{\ \}$. Notational consistency would seem to require that the probability of the event $\{2\}$ be written as $P(\{2\})$. But it seems unnecessary to add the parentheses around the braces, so we simplify the expression to $P\{2\}$ or $P(2)$.

Example 1.5. (Continuation of Examples 1.3 and 1.4) The probability measure P contains our assumptions and beliefs about the phenomenon that we are modeling.

If we wish to model a flip of a biased coin we alter the probabilities. For example, suppose we know that heads is three times as likely as tails. Then we define our probability measure P_1 by $P_1\{\mathtt{H}\} = \frac{3}{4}$ and $P_1\{\mathtt{T}\} = \frac{1}{4}$. The sample space is again $\Omega = \{\mathtt{H}, \mathtt{T}\}$ as in Example 1.3, but the probability measure has changed to conform with our assumptions about the experiment.

If we believe that we have a loaded die and a six is twice as likely as any other number, we use the probability measure $\widetilde{P}$ defined by

$$\widetilde{P}\{1\} = \widetilde{P}\{2\} = \widetilde{P}\{3\} = \widetilde{P}\{4\} = \widetilde{P}\{5\} = \tfrac{1}{7} \quad \text{and} \quad \widetilde{P}\{6\} = \tfrac{2}{7}.$$

Alternatively, if we scratch away the five from the original fair die and turn it into a second two, the appropriate probability measure is

$$Q\{1\} = \tfrac{1}{6},\ Q\{2\} = \tfrac{2}{6},\ Q\{3\} = \tfrac{1}{6},\ Q\{4\} = \tfrac{1}{6},\ Q\{5\} = 0,\ Q\{6\} = \tfrac{1}{6}. \qquad \blacktriangle$$

These examples show that to model different phenomena it is perfectly sensible to consider different probability measures on the same sample space. Clarity might demand that we distinguish different probability measures notationally from each other. This can be done by adding ornaments to the P, as in P_1 or $\widetilde{P}$ (pronounced "P tilde") above, or by using another letter such as Q. Another important point is that it is perfectly valid to assign a probability of zero to a nonempty event, as with Q above.

Example 1.6. Let the experiment consist of a roll of a pair of dice (as in the games of Monopoly or craps). We assume that the dice can be distinguished from each other, for example that one of them is blue and the other one is red. The sample space is the set of pairs of integers from 1 through 6, where the first number of the pair denotes the number on the blue die and the second denotes the number on the red die:

$$\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}.$$

Here (a, b) is a so-called *ordered pair* which means that outcome $(3, 5)$ is distinct from outcome $(5, 3)$. (Note that the term "ordered pair" means that order matters, not that the pair is in increasing order.) The assumption of fair dice would dictate equal probabilities: $P\{(i, j)\} = \frac{1}{36}$ for each pair $(i, j) \in \Omega$. An example of an event of interest would be

$$D = \{\text{the sum of the two dice is } 8\} = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$$

and then by the additivity of probabilities

$$P(D) = P\{(2,6)\} + P\{(3,5)\} + P\{(4,4)\} + P\{(5,3)\} + P\{(6,2)\}$$
$$= \sum_{(i,j):i+j=8} P\{(i,j)\} = 5 \cdot \tfrac{1}{36} = \tfrac{5}{36}. \qquad \blacktriangle$$

Example 1.7. We flip a fair coin three times. Let us encode the outcomes of the flips as 0 for heads and 1 for tails. Then each outcome of the experiment is a sequence of length three where each entry is 0 or 1:

$$\Omega = \{(0,0,0), (0,0,1), (0,1,0), \dots, (1,1,0), (1,1,1)\}. \tag{1.4}$$

This Ω is the set of ordered triples (or 3-tuples) of zeros and ones. Ω has $2^3 = 8$ elements. (We review simple counting techniques in Appendix C.) With a fair coin all outcomes are equally likely, so $P\{\omega\} = 2^{-3}$ for each $\omega \in \Omega$. An example of an event is

$$B = \{\text{the first and third flips are heads}\} = \{(0,0,0), (0,1,0)\}$$

with

$$P(B) = P\{(0,0,0)\} + P\{(0,1,0)\} = \tfrac{1}{8} + \tfrac{1}{8} = \tfrac{1}{4}. \qquad \blacktriangle$$

Much of probability deals with repetitions of a simple experiment, such as the roll of a die or the flip of a coin in the previous two examples. In such cases *Cartesian product spaces* arise naturally as sample spaces. If $A_1, A_2, \dots, A_n$ are sets then the Cartesian product

$$A_1 \times A_2 \times \cdots \times A_n$$

is defined as the set of ordered n-tuples with the ith element from A_i. In symbols

$$A_1 \times A_2 \times \cdots \times A_n = \{(x_1, \dots, x_n) : x_i \in A_i \text{ for } i = 1, \dots, n\}.$$

In terms of product notation, the sample space of Example 1.6 for a pair of dice can be written as

$$\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$$

while the space for three coin flips in Example 1.7 can be expressed as

$$\Omega = \{0,1\} \times \{0,1\} \times \{0,1\} = \{0,1\}^3.$$

1.2. Random sampling

Sampling is choosing objects randomly from a given set. It can involve repeated choices or a choice of more than one object at a time. Dealing cards from a deck is an example of sampling. There are different ways of setting up such experiments which lead to different probability models. In this section we discuss three sampling mechanisms that lead to equally likely outcomes. This allows us to compute probabilities by counting. The required counting methods are developed systematically in Appendix C.

Before proceeding to sampling, let us record a basic fact about experiments with equally likely outcomes. Suppose the sample space Ω is a finite set and let $\#\Omega$ denote the total number of possible outcomes. If each outcome ω has the same probability then $P\{\omega\} = \frac{1}{\#\Omega}$ because probabilities must add up to 1. In this case probabilities of events can be found by counting. If A is an event that consists of elements $a_1, a_2, \dots, a_r$, then additivity and $P\{a_i\} = \frac{1}{\#\Omega}$ imply

$$P(A) = P\{a_1\} + P\{a_2\} + \cdots + P\{a_r\} = \frac{\#A}{\#\Omega}$$

where we wrote $\#A$ for the number of elements in the set A.

Fact 1.8. If the sample space Ω has finitely many elements and each outcome is equally likely then for any event $A \subset \Omega$ we have

$$P(A) = \frac{\#A}{\#\Omega}. \tag{1.5}$$

Look back at the examples of the previous section to check which ones were of the kind where $P\{\omega\} = \frac{1}{\#\Omega}$.

Remark 1.9. (Terminology) It should be clear by now that random outcomes do not have to be equally likely. (Look at Example 1.5 in the previous section.) However, it is common to use the phrase "an element is chosen at random" to mean that all choices are equally likely. The technically more accurate phrase would be "chosen *uniformly* at random." Formula (1.5) can be expressed by saying "when outcomes are equally likely, the probability of an event equals the number of favorable outcomes over the total number of outcomes." ▲

We turn to discuss sampling mechanisms. An *ordered sample* is built by choosing objects one at a time and by keeping track of the order in which these objects were chosen. After each choice we either replace (put back) or discard the just chosen object before choosing the next one. This distinction leads to *sampling with replacement* and *sampling without replacement*. An *unordered sample* is one where only the identity of the objects matters and not the order in which they came.

We discuss the sampling mechanisms in terms of an urn with numbered balls. An urn is a traditional device in probability (see Figure 1.1). You cannot see the contents of the urn. You reach in and retrieve one ball at a time without looking. We assume that the choice is uniformly random among the balls in the urn.

Figure 1.1. Three traditional mechanisms for creating experiments with random outcomes: an urn with balls, a six-sided die, and a coin.

Sampling with replacement, order matters

Suppose the urn contains n balls numbered $1, 2, \ldots, n$. We retrieve a ball from the urn, record its number, and put the ball back into the urn. (Putting the ball back into the urn is the *replacement* step.) We carry out this procedure k times. The outcome is the ordered k-tuple of numbers that we read off the sampled balls. Represent the outcome as $\omega = (s_1, s_2, \ldots, s_k)$ where s_1 is the number on the first ball, s_2 is the number on the second ball, and so on. The sample space Ω is a Cartesian product space: if we let $S = \{1, 2, \ldots, n\}$ then

$$\Omega = \underbrace{S \times S \times \cdots \times S}_{k \text{ times}} = S^k = \{(s_1, s_2, \ldots, s_k) : s_i \in S \text{ for } i = 1, \ldots, k\}. \qquad (1.6)$$

How many outcomes are there? Each s_i can be chosen in n different ways. By Fact C.5 from Appendix C we have

$$\#\Omega = \underbrace{n \cdot n \cdots n}_{k \text{ times}} = n^k.$$

We assume that this procedure leads to equally likely outcomes, hence the probability of each k-tuple is $P\{\omega\} = n^{-k}$.

Let us illustrate this with a numerical example.

Example 1.10. Suppose our urn contains 5 balls labeled 1, 2, 3, 4, 5. Sample 3 balls with replacement and produce an ordered list of the numbers drawn. At each step we have the same 5 choices. The sample space is

$$\Omega = \{1, 2, 3, 4, 5\}^3 = \{(s_1, s_2, s_3) : \text{each } s_i \in \{1, 2, 3, 4, 5\}\}$$

and $\#\Omega = 5^3$. Since all outcomes are equally likely, we have for example

$$P\{\text{the sample is } (2,1,5)\} = P\{\text{the sample is } (2,2,3)\} = 5^{-3} = \frac{1}{125}.$$ ▲

Repeated flips of a coin or rolls of a die are also examples of sampling with replacement. In these cases we are sampling from the set $\{\mathtt{H}, \mathtt{T}\}$ or $\{1, 2, 3, 4, 5, 6\}$.

(Check that Examples 1.6 and 1.7 are consistent with the language of sampling that we just introduced.)

Sampling without replacement, order matters

Consider again the urn with n balls numbered $1, 2, \dots, n$. We retrieve a ball from the urn, record its number, and put the ball aside, in other words not back into the urn. (This is the *without replacement* feature.) We repeat this procedure k times. Again we produce an ordered k-tuple of numbers $\omega = (s_1, s_2, \dots, s_k)$ where each $s_i \in S = \{1, 2, \dots, n\}$. However, the numbers $s_1, s_2, \dots, s_k$ in the outcome are distinct because now the same ball cannot be drawn twice. Because of this, we clearly cannot have k larger than n.

Our sample space is

$$\Omega = \{(s_1, s_2, \dots, s_k) : \text{each } s_i \in S \text{ and } s_i \neq s_j \text{ if } i \neq j\}. \tag{1.7}$$

To find $\#\Omega$, note that s_1 can be chosen in n ways, after that s_2 can be chosen in $n - 1$ ways, and so on, until there are $n - k + 1$ choices remaining for the last entry s_k. Thus

$$\#\Omega = n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) = (n)_k. \tag{1.8}$$

Again we assume that this mechanism gives us equally likely outcomes, and so $P\{\omega\} = \frac{1}{(n)_k}$ for each k-tuple ω of distinct numbers. The last symbol $(n)_k$ of equation (1.8) is called the descending factorial.

Example 1.11. Consider again the urn with 5 balls labeled 1, 2, 3, 4, 5. Sample 3 balls without replacement and produce an ordered list of the numbers drawn. Now the sample space is

$$\Omega = \{(s_1, s_2, s_3) : \text{each } s_i \in \{1, 2, 3, 4, 5\} \text{ and } s_1, s_2, s_3 \text{ are all distinct}\}.$$

The first ball can be chosen in 5 ways, the second ball in 4 ways, and the third ball in 3 ways. So

$$P\{\text{the sample is } (2,1,5)\} = \frac{1}{5 \cdot 4 \cdot 3} = \frac{1}{60}.$$

The outcome $(2, 2, 3)$ is not possible because repetition is not allowed. ▲

Another instance of sampling without replacement would be a random choice of students from a class to fill specific roles in a school play, with at most one role per student.

If $k = n$ then our sample is a random ordering of all n objects. Equation (1.8) becomes $\#\Omega = n!$. This is a restatement of the familiar fact that a set of n elements can be ordered in $n!$ different ways.

Sampling without replacement, order irrelevant

In the previous sampling situations the order of the outcome was relevant. That is, outcomes $(1, 2, 5)$ and $(2, 1, 5)$ were regarded as distinct. Next we suppose that we do not care about order, but only about the set $\{1, 2, 5\}$ of elements sampled. This kind of sampling without replacement can happen when cards are dealt from a deck or when winning numbers are drawn in a state lottery. Since order does not matter, we can also imagine choosing the entire set of k objects at once instead of one element at a time.

Notation is important here. The ordered triple $(1, 2, 5)$ and the set $\{1, 2, 5\}$ must not be confused with each other. Consequently *in this context* we must not mix up the notations ( ) and { }.

As above, imagine the urn with n balls numbered $1, 2, \dots, n$. Let $1 \le k \le n$. Sample k balls without replacement, but record only which balls appeared and not the order. Since the sample contains no repetitions, the outcome is a subset of size k from the set $S = \{1, 2, \dots, n\}$. Thus

$$\Omega = \{\omega \subset S : \#\omega = k\}.$$

(Do not be confused by the fact that an outcome ω is itself now a set of numbers.) The number of elements of Ω is given by the binomial coefficient (see Fact C.12 in Appendix C):

$$\#\Omega = \frac{n!}{(n-k)!k!} = \binom{n}{k}.$$

Assuming that the mechanism leads to equally likely outcomes, $P\{\omega\} = \binom{n}{k}^{-1}$ for each subset ω of size k.

Another way to produce an unordered sample of k balls without repetitions would be to execute the following three steps: (i) randomly order all n balls, (ii) take the first k balls, and (iii) ignore their order. Let us verify that the probability of obtaining a particular selection $\{s_1, \dots, s_k\}$ is $\binom{n}{k}^{-1}$, as above.

The number of possible orderings in step (i) is $n!$. The number of favorable orderings is $k!(n-k)!$, because the first k numbers must be an ordering of $\{s_1, \dots, s_k\}$ and after that comes an ordering of the remaining $n - k$ numbers. Then from the ratio of favorable to all outcomes

$$P\{\text{the selection is } \{s_1, \dots, s_k\}\} = \frac{k!(n-k)!}{n!} = \frac{1}{\binom{n}{k}},$$

as we expected.

The description above contains a couple of lessons.

(i) There can be more than one way to build a probability model to solve a given problem. But a warning is in order: once an approach has been chosen, it must be followed consistently. Mixing up different representations will surely lead to an incorrect answer.

Cambridge University Press
978-1-108-41585-9 — Introduction to Probability
David F. Anderson , Timo Seppäläinen , Benedek Valkó
Excerpt
More Information

(ii) It may pay to introduce additional structure into the problem. The second approach introduced order into the calculation even though in the end we wanted an outcome without order.

Example 1.12. Suppose our urn contains 5 balls labeled 1, 2, 3, 4, 5. Sample 3 balls without replacement and produce an unordered set of 3 numbers as the outcome. The sample space is

$$\Omega = \{\omega : \omega \text{ is a 3-element subset of } \{1, 2, 3, 4, 5\}\}.$$

For example

$$P(\text{the sample is } \{1{,}2{,}5\}) = \frac{1}{\binom{5}{3}} = \frac{2!3!}{5!} = \frac{1}{10}.$$

The outcome $\{2, 2, 3\}$ does not make sense as a set of three numbers because of the repetition. ▲

The fourth alternative, sampling with replacement to produce an unordered sample, does not lead to equally likely outcomes. This scenario will appear naturally in Example 6.7 in Chapter 6.

Further examples

The next example contrasts all three sampling mechanisms.

Example 1.13. Suppose we have a class of 24 children. We consider three different scenarios that each involve choosing three children.

(a) Every day a random student is chosen to lead the class to lunch, without regard to previous choices. What is the probability that Cassidy was chosen on Monday and Wednesday, and Aaron on Tuesday?

This is sampling with replacement to produce an ordered sample. Over a period of three days the total number of different choices is 24^3. Thus

$$P\{(\text{Cassidy, Aaron, Cassidy})\} = 24^{-3} = \frac{1}{13{,}824}.$$

(b) Three students are chosen randomly to be class president, vice president, and treasurer. No student can hold more than one office. What is the probability that Mary is president, Cory is vice president, and Matt treasurer?

Imagine that we first choose the president, then the vice president, and then the treasurer. This is sampling without replacement to produce an ordered sample. Thus

$$P\{\text{Mary is president, Cory is vice president, and Matt treasurer}\}$$
$$= \frac{1}{24 \cdot 23 \cdot 22} = \frac{1}{12{,}144}.$$

Suppose we asked instead for the probability that Ben is either president or vice president. We apply formula (1.5). The number of outcomes in which Ben ends up as president is $1 \cdot 23 \cdot 22$ (1 choice for president, then 23 choices for vice president, and finally 22 choices for treasurer). Similarly the number of ways in which Ben ends up as vice president is $23 \cdot 1 \cdot 22$. So

$$P\{\text{Ben is president or vice president}\} = \frac{1 \cdot 23 \cdot 22 + 23 \cdot 1 \cdot 22}{24 \cdot 23 \cdot 22} = \frac{1}{12}.$$

(c) A team of three children is chosen at random. What is the probability that the team consists of Shane, Heather and Laura?

A team means here simply a set of three students. Thus we are sampling without replacement to produce a sample without order.

$$P(\text{the team is }\{\text{Shane, Heather, Laura}\}) = \frac{1}{\binom{24}{3}} = \frac{1}{2024}.$$

What is the probability that Mary is on the team? There are $\binom{23}{2}$ teams that include Mary since there are that many ways to choose the other two team members from the remaining 23 students. Thus by the ratio of favorable outcomes to all outcomes,

$$P\{\text{the team includes Mary}\} = \frac{\binom{23}{2}}{\binom{24}{3}} = \frac{3}{24} = \frac{1}{8}.$$ ▲

Problems of unordered sampling without replacement can be solved either with or without order. The next two examples illustrate this idea.

Example 1.14. Our urn contains 10 marbles numbered 1 to 10. We sample 2 marbles without replacement. What is the probability that our sample contains the marble labeled 1? Let A be the event that this happens. However we choose to count, the final answer $P(A)$ will come from formula (1.5).

Solution with order. Sample the 2 marbles in order. As in (1.8), $\#\Omega = 10{\cdot}9 = 90$. The favorable outcomes are all the ordered pairs that contain 1:

$$A = \{(1,2),(1,3),\dots,(1,10),(2,1),(3,1),\dots,(10,1)\}$$

and we count $\#A = 18$. Thus $P(A) = \frac{18}{90} = \frac{1}{5}$.

Solution without order. Now the outcomes are subsets of size 2 from the set $\{1,2,\dots,10\}$ and so $\#\Omega = \binom{10}{2} = \frac{9\cdot 10}{2} = 45$. The favorable outcomes are all the 2-element subsets that contain 1:

$$A = \{\{1,2\},\{1,3\},\dots,\{1,10\}\}.$$

Now $\#A = 9$ so $P(A) = \frac{9}{45} = \frac{1}{5}$.

Both approaches are correct and of course they give the same answer. ▲