Basic Probability

P. Reany

November 5, 2025

Abstract

In this article we'll investigate the most basic probability theory we can encounter. The sample spaces are finite, rather than infinite.

We must use time as a tool, not as a couch.
— John F. Kennedy

1 Getting started

When an **experiment** is performed, the result is referred to as an **outcome**. For example, when we flip a coin, we accept that only one of two outcomes can result: either a heads or a tail. Generally, the outcome space can be two or more outcomes, but, for our humble purposes here, only a finite number of outcomes. This is a special case in probability, which can allow an infinite number of outcomes. But this time we're going to go the simple route this time.

Now we need to introduce a bunch of definitions, and couch everything in set theory. So, the set of all possible outcomes of an experiment is called the **sample space**, which is represented by the Greek letter Ω. The number of elements in the sample space is represented as $|\Omega|$, which is the cardinality of the set.

For simplicity, we will order the elements in the sample space. Now, we must assign the probability of the ith element, s_i, $p(s_i)$. Our first rule:

$$0 < p(s_i) < 1\,. \tag{1}$$

We don't allow the probability of $p(s_i)$ to be zero because this outcome s_i would never happen — so there's not much reason to include it in the sample space. Also, we don't allow the probability of $p(s_i)$ to be unity because this event s_i would be the only outcome that could happen.

Let's assume that there are N possible outcomes to the experiment.

$$|\Omega| = N\,. \tag{2}$$

Then, our second rule:

$$\sum_{i=1}^{N} p(s_i) = 1\,. \tag{3}$$

If the probabilities of all the outcomes are equally likely, then then the probability of a single outcome is given as

$$p(s_i) = \frac{1}{N} \quad \text{for all } i\,, \tag{4}$$

which is consistent with (3).

2 The Event Concept

In probability theory an **event** is mathematically a subset of the sample space. So, we need a means to assign probabilities to events. We're not going to specify the exact probabilities to events at this time because how this is done will depend on the nature of the experiment being performed. However, we can at this time lay out certain rules that the probabilities of events have to be obeyed.

Let E_1 and E_2 be disjoint subsets of the sample space. Then our first rule is

$$p(E_1 \cup E_2) = p(E_1) + p(E_2)\,. \tag{5}$$

Now we need to get a bit more sophisticated. We have to account for every possible event. If we collect every possible event into a set all its own, that set has a name, called the **power set**[1] of the sample space.

Believe it or not, one of the elements in the power set is the empty set $\emptyset$. Another is the set Ω. We assign the probability of Ω to be unity: $p(\Omega) = 1$. We assign the probability of 0 to the empty set, that is: $p(\emptyset) = 0$.

The smallest event one can have is a single outcome of the experiment, of its N possible discrete outcomes. Let's assume for the moment that we have ordered our finite list of outcomes, $s_1, \ldots, s_N$. Call the probability of the ith element p_i. Then

$$\sum_{i=1}^{N} p(s_i) = \sum_{i=1}^{N} p_i = 1\,. \tag{6}$$

Now, in a given experiment the outcomes do not need to be all of equal likelihood. For example, in the experiment of the toss of a die, we often assume that the 6 outcomes are equally likely for the some purpose of analysis of a so-called 'fair die', though it's quite unlikely that any real die will have equally likely outcomes for all its face numbers (on a single roll). But whatever the individual probabilities are, they have to add to unity:

$$p(1) + p(2) + p(3) + p(4) + p(5) + p(6) = 1\,. \tag{7}$$

[1] This term comes from set theory.

3 Independent Events

Say we design a simple experiment that records the results of flipping a coin two times. The outcomes can be recorded as ordered pairs (H = head, T = tail):

$$\Omega = \{(\mathrm{H,H}),\ (\mathrm{H,T}),\ (\mathrm{T,H}),\ (\mathrm{T,T})\}\,. \tag{8}$$

If you have good reason to believe that the second coin flip is in no way influenced by the first flip, then you can consider the two flips as independent of each other. This state of affairs has profound mathematical consequences for calculation. Let the probability that a flip produces a head as $P(\mathrm{H})$ and a tail as $P(\mathrm{T})$. Now, if these consecutive results are independent of each other then

$$p(\mathrm{H} \cap \mathrm{T}) = p((\mathrm{H,T})) = p(\mathrm{H})p(\mathrm{T})\,, \tag{9}$$

where neither $p(\mathrm{H})$ nor $p(\mathrm{T})$ is zero.

Going into this, we accepted as obvious that a combination of first head and then tail is possible. If we express this in terms of 'half events', we write

$$(\mathrm{H}, \cdot\,) \cap (\,\cdot\,,\mathrm{T}) \neq \emptyset\,, \tag{10}$$

where we can think of '(H, $\cdot$)' and '($\cdot$,T)' as half events. What (10) tells us is that a head followed by a tail is doable, for otherwise that would contradict (9), since in that case $p(\emptyset) = 0$, causing a disagreement between the LHS and the RHS.

Let's state the active rule behind this more generally: Let A and B be events. Then

$$P(A \cap B) = P(A)P(B)\,, \text{ provided that } A \cap B \neq \emptyset\,. \tag{11}$$

But how are we to determine if two events are independent of each other? I had a lengthy conversation with Copilot on how this should best be done to teach new students about probability. This is what I came up with:

> "Two events A and B are said to be independent of each other if there does not exist a causal or functional relationship between them and $A \cap B \neq \emptyset$." And so I suppose they'll have to rely on intuition and common sense to make this evaluation, unless they are explicitly told that the events are independent.

However, to this Copilot replied that I should provide a more epistemologically based description, as follows:

> "Two events A and B are said to be independent if knowing that one occurs does not change the likelihood of the other. In practical terms, this means there is no causal or functional connection between them. Furthermore, for the notion to be applicable, $A \cap B \neq \emptyset$; that is, they must be able to occur together."

4 An Example Problem

Sofia and Tess will each randomly choose one of the 10 integers from 1 to 10. What is the probability that neither integer chosen will be the square of the other?

(A) 0.64 (B) 0.72 (C) 0.81 D) 0.90 (E) 0.95

5 The Preparation

In probability, the Sofia-Tess situation is an 'experiment'. Each time it is performed results in a particular outcome of a pair of so-called 'random choices', which I'll represent as an ordered pair: (Sofia's pick, Tess's pick). On each experiment, each subject has ten possible numbers to choose from, and since the choices are made randomly, that means we assign a 1 in 10 chance for a any given integer from 1 to 10 to be picked. So, for example, for the specific outcome (2,5), we assign the probability of 1/100 because the probability that Sofia chose the 2 in a given experiment is 1/10 and the probability that Tess chose a 5 is also 1/10. Now the probability of a joint outcome is the product of the individual outcomes when they are independent of each other.

The union of all possible outcomes of an experiment is called the 'sample space', which is often denoted by the Greek letter Ω. In the case of the Sofia-Tess experiment, Ω is given as

$$\Omega = \{(1,1),(1,2),\ldots,(2,1),(2,2),\ldots,(10,1),(10,2),\ldots,(10,10)\}\,. \tag{12}$$

Of course, each possible pair is only listed once. The cardinality of the set Ω is 100. A subset of the sample space is called an 'event.'

Generally speaking, when all the outcomes are equally likely, the probability of an event is the ratio of the cardinality of the event set divided by the cardinality of the sample space. As an example, what is the probability that Sofia and Tess both choose the same number? Okay, this is easy: The event space E is

$$E = \{(1,1),(2,2),\ldots,(10,10)\}\,, \tag{13}$$

with cardinality 10. Hence, the probability of this event, $P(E)$, is

$$P(E) = \frac{10}{100} = 0.1\,. \tag{14}$$

What's the probability that any particular event will lie within the sample space? It's a certainty that it will because that's what the sample space is — the collection of all possible outcomes. In other words, there is no such thing as an outcome that is not contained in the sample space. Hence

$$P(\Omega) = 1\,. \tag{15}$$

Let E_1 and E_2 be two disjoint subsets of the sample space. Then

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)\,. \tag{16}$$

Now, for an important corollary: Let E and $\bar{E}$ be two disjoint subsets of Ω that together contain all possible outcomes of an experiment. In other words,

$$\Omega = E \cup \bar{E}\,. \tag{17}$$

Then

$$P(\Omega) = P(E) + P(\bar{E})\,, \tag{18}$$

or

$$1 = P(E) + P(\bar{E})\,. \tag{19}$$

Now, in some experiments, we may be asked to solve for the probability of E, which may be difficult to collect all of its elements. If in this case, it's a lot easier to collect the elements of E's set complement $\bar{E}$,[2] then we have the option of solving for $P(E)$ by first solving first for $P(\bar{E})$, and then using the formula derived from (19):

$$P(E) = 1 - P(\bar{E})\,. \tag{20}$$

6 The Solution

From the Preparation Section, we've already determined the sample space Ω for our experiment (12), which has cardinality 100.

The event space E that we're interested in is the set Ω minus all pairs of the form (x, x^2) or (x^2, x). Do you see it? It's the sample space with all pairs removed from it in which one value is the square of the other. This is one of those cases in which it's simpler to regard the set complement of E:

$$\bar{E} = \big\{(x, x^2), (x^2, x) \,|\, x, x^2 \in [1..10]\big\}\,. \tag{21}$$

So, let's list it out!

$$\bar{E} = \{(1,1), (2,4), (3,9), (4,2), (9,3)\}\,. \tag{22}$$

This set has cardinality 5. Hence

$$P(\bar{E}) = \frac{5}{100} = 0.05\,. \tag{23}$$

And then, substituting this value into (20), we have that

$$P(E) = 1 - P(\bar{E}) = 1 - 0.05 = 0.95\,. \tag{24}$$

Therefore, the answer is (E).

[2] Let C be the disjoint union of subsets A and B. Then, B is the set **complement** of A in C, and likewise, A is the set **complement** of B in C

7 Appendix:

Use induction to show that for m mutually disjoint events $E_1, \ldots, E_m$,

$$p(E_1 \cup E_2 \cup \cdots \cup E_m) = p(E_1) + p(E_2) + \cdots + p(E_m)\,. \tag{25}$$

Proof:

Let A and B be disjoint subsets of the sample space. Then our first rule on events is, again,

$$p(A \cup B) = p(A) + p(B)\,, \tag{26}$$

and this becomes our base case for an induction proof.

Lemma:

If A is a set, disjoint from each set in $\{B_1, B_2, \ldots, B_{m-1}\}$, then A is disjoint from $B_1 \cup B_2 \cup \ldots \cup B_{m-1}$. Assume not.

Then

$$A \cap (B_1 \cup b_2 \cup \ldots \cup B_{m-1}) \neq \emptyset\,. \tag{27}$$

This implies that there exists $x \in A$ and $x \in B_i$ for some $i \in [1..m-1]$. But this implies that A and B_i are not disjoint. Contradiction.

Assume the 'inductive hypothesis':

$$p(E_1 \cup E_2 \cup \cdots \cup E_{m-1}) = p(E_1) + p(E_2) + \cdots + p(E_{m-1})\,. \tag{28}$$

Next, let

$$E_1 \cup E_2 \cup \cdots \cup E_{m-1} = A \quad \text{and} \quad E_m = B\,, \tag{29}$$

then, employing the associativity of set union,

$$\begin{aligned} p(E_1 \cup E_2 \cup \cdots \cup E_{m-1} \cup E_m) &= p((E_1 \cup E_2 \cup \cdots \cup E_{m-1}) \cup E_m) \\ &= p(A \cup B) \\ &= p(A) + p(B) \\ &= p(E_1) + p(E_2) + \cdots + p(E_{m-1}) + p(E_m)\,, \end{aligned} \tag{30}$$

where the last step used the inductive hypothesis.