

Probability spaces.

We are going to give a mathematical definition of **probability space**. We will first make some remarks which will motivate this definition.

A fundamental notion in probability theory is that of an **experiment**. An experiment is an activity which can be repeated any number of times, each repetition of which has an **outcome**. **We require that information about outcomes of past performances of the experiment provides no information about future outcomes of performances of the experiment.** The set of outcomes of the experiment is called the **sample space** of the experiment. The points of the sample space, which are outcomes, are sometimes called **sample points**. (It is possible that the experiment have only one outcome so that the sample space has only one member, but this situation is probabilistically trivial.) Let us give two examples.

For the first example, suppose I take a coin out of my pocket, flip it three times and observe for each flip whether it came up heads or tails. A natural sample space for this experiment would be the set of ordered triples

$$\{(H, H, H), (H, H, T), (H, T, H), (H, T, T), (T, H, H), (T, H, T), (T, T, H), (T, T, T)\}.$$

Other sample spaces are possible. You may encode the outcomes as bit patterns, taking 1 for heads and 0 for tails, and ending up with

$$\{(1, 1, 1), (1, 1, 0), (1, 0, 1), (1, 0, 0), (0, 1, 1), (0, 1, 0), (0, 0, 1), (0, 0, 0)\},$$

or you could use the the numbers with the foregoing binary representations ending up with

$$\{7, 6, 5, 4, 3, 2, 1, 0\}.$$

But wait — maybe you don’t care about the pattern of heads and tails. You might be interested in the sum of the angles with the vertical made by the plane of the coin the instant it strikes the table. This would be considered a different experiment. So the emphasis is on the outcome more than the activity which produced it.

As a second example, consider a barrel full of batteries of all different sizes and shapes. You pick a battery out of the barrel (I didn’t and won’t say how, but that matters if we’re interested in probabilities) and we measure its voltage on my voltmeter. If you do that, the outcome will be a real number between 0 and 15. Now you can argue that there are lots of real numbers between 0 and 15 that will never arise as outcomes, like $\sqrt{2}$ times 10^{-3000} for example, and you’ll be right. As often happens in science, we are replacing some physical reality by an idealization.

In probability theory one associates with a sample space a family of subsets of the sample space the members of which are called **events**. In this course, for all practical purposes, *every* subset of the sample space will be an event. It turns out that there are serious technical **and** intuitive problems with this, but these are beyond the scope of the course and we will not go into them much. A natural event associate to the first example is “more heads than tails” which is

$$\{(H, H, H), (H, H, T), (H, T, H), (T, H, H)\}.$$

A natural event associated with the second event is

$$(1.2, 1.8).$$

If an outcome, which corresponds to a battery picked out of the barrel, is in this event then I might reasonably put the battery in my flashlight, assuming the battery is a “D” battery and my flashlight takes “D” batteries.

Now we come to the key concept. Fix an experiment with sample space S and fix an event E . What is the probability

$$P(E)$$

of E ? Here is a very good nonmathematical definition. Perform the experiment indefinitely. Now that’s impossible but imagine doing it anyway. This results in a sequence

$$s_1, s_2, \dots, s_n, \dots$$

of outcomes, or sample points if you prefer that terminology. (We're using s 's here because "sample" starts with s .) By the way, the repetitions of the experiment should be **independent**, whatever that means. At the very least, you shouldn't rig the experiment in any way that allows to obtain **any** information about the n -th outcome until the experiment has been performed for the n -th time. Given such a sequence s of outcomes we let

$$\nu(E, n)$$

be the number of $i \in \{1, \dots, n\}$ such that $s_i \in E$ and **define** the probability of the event E to be

$$P(E) = \lim_{n \rightarrow \infty} \frac{\nu(E, n)}{n}.$$

Now it is not at all clear that this limit exists. But let's imagine that it does. There is an outstanding result in probability theory called the strong law of large numbers that says that the limit nearly always (whatever that means) does exist. Let us make some observations. First,

$$0 \leq P(E) \leq 1.$$

next, the probability of the empty event (no sample points) is zero and the probability of the certain event (all sample points) is one. Finally, suppose E_1, \dots, E_N are disjoint or **mutually exclusive** events. Then

$$\nu(\cup_{i=1}^N E_i, n) = \sum_{i=1}^N \nu(E_i, n).$$

Passing to the limit, we find the basic identity

$$P(\cup_{i=1}^N E_i) = \sum_{i=1}^N P(E_i).$$

We have just described the **relative frequency** interpretation of probability. This is discussed in the book in some detail. The book also discusses some other ways of thinking about probability. I find them incomprehensible. There **are** other good ways of thinking about probability but we will not go into them.

Now we're ready to make the formal definition of a probability space.

Definition. A **probability space** is an ordered triple

$$(S, \mathcal{E}, P)$$

where

S is a set,

\mathcal{E} is a family of subsets of S and

P is a function on \mathcal{E} taking values in $[0, 1]$

such that the following conditions hold:

$$\emptyset \in \mathcal{E};$$

$$S \in \mathcal{E};$$

$$\text{if } E \in \mathcal{E} \text{ then } S \sim E \in \mathcal{E};$$

$$\text{if } \mathcal{C} \text{ is a countable family of sets in } \mathcal{E} \text{ then } \bigcup \mathcal{C} \in \mathcal{E};$$

$$P(\emptyset) = 0;$$

$$P(S) = 1 \text{ and}$$

if \mathcal{C} is a countable disjointed family of sets in \mathcal{E} then

$$P(\bigcup \mathcal{C}) = \sum_{E \in \mathcal{C}} P(E).$$

Here is a typical way one builds a probability space. Start with a finite set S . Next, let

$$p : S \rightarrow [0, 1]$$

be such that

$$\sum_{s \in S} p(s) = 1.$$

Let \mathcal{E} be the family of all subsets of S . (There are 2^N of them, right?) For each $E \in \mathcal{E}$ let

$$P(E) = \sum_{s \in E} p(s).$$

Then (S, \mathcal{E}, P) is a probability space.

For example, if the coin in the example above is fair it would be reasonable to take S to be the set of ordered triples with entries H or T and to take p to be $\frac{1}{8}$ on each of the 8 sample points. The probability of the event “more heads than tails” would then be $\frac{1}{2}$.