CSCI2244-Randomness and Computation
Lecture 1: Probability Spaces

January 22, 2019

Some important words.

- probability space
- sample space
- outcome
- event
- discrete space (finite or infinite)
- continuous space
- probability mass function (= probability distribution)
- probability function
- mutually exclusive events

A *probability space* is a mathematical model of a random experiment (*e.g.,* flipping a coin, rolling dice, throwing a dart at a dart board, picking a person out of a crowd and asking for their birthday...). Formally, it is a pair (S, P). The first component is S called the *sample space*, the second component P is called the *probability function.* We will take these up in turn.

1 Sample Space

Mathematically S is just a set. We think of it as the set of *outcomes* of the experiment.

Example 1 Toss a coin. $S = \{H, T\}$. $|S| = 2$.

Example 2 Toss N coins in succession. An outcome is the resulting sequence of N H's and T's.

$$S = \underbrace{\{H, T\} \times \cdots \times \{H, T\}}_{N \text{ times}}. \text{ (cartesian product)}$$

$$|S| = 2^N.$$

For our in-class experiment, $N = 200$, so $|S| = 2^{200} \approx 10^{60}$, far beyond the capacity of humans or machines to count.

A subset of the sample space is called an *event*. For example, in this experiment, with $N = 4$, there are 16 outcomes. The event 'there are at least two consecutive H's' is the set

$$E = \{HHHH, HHHT, HHTH, HHTT, THHH, THHT, TTHH\}.$$

Example 3 Roll a single six-sided die

$$S = \{1, 2, 3, 4, 5, 6\}$$
$$|S| = 6.$$

Example 4 Roll two six-sided dice.

Two reasonable choices here to model the outcome of this experiment. Either

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

with $|S| = 36$, or

$$S = \{(i, j) : 1 \leq i \leq j \leq 6\},$$

with $|S| = 21$.

The first treats 'a one and a two' as a different outcome from 'a two and a one', the second treats these as the same outcome. Either one can correctly be used as a model for this experiment, but the second is preferred, as we will see shortly when we discuss probability functions.

Example 5 Flip a coin repeatedly until you get heads. The outcome is the number of tosses you made:

$$S = \{1, 2, 3, \ldots\} = \mathbf{Z}^+ \text{(set of positive integers)}.$$

S is a *countably infinite* set. This is a *discrete infinite* sample space. The sample spaces we saw earlier were *finite* sample spaces, which are also discrete spaces.

Example 6 Spin a spinner whose circumference is labeled by real numbers between 0 and 1 (including 0, not including 1). Here the outcome is the point on the circumference that the spinner stops on. S is the half-open interval

$$[0, 1) = \{x \in \mathbf{R} : 0 \leq x < 1\}.$$

(See Figure 1.) This is a *continuous* sample space

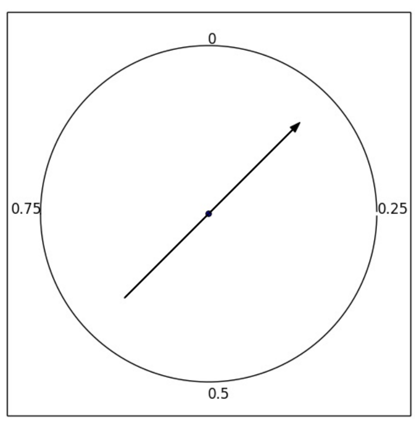

Figure 1: A spinner. Outcomes are the points on the circumference of the circle.

Example 7 Throw a dart at a circular dart board one foot in radius. Here the outcome is the point on the board that the dart hits.

$$S = \{(x, y) \in \mathbf{R} \times \mathbf{R} : x^2 + y^2 \leq 1\}.$$

This is also a continuous sample space.

The fundamental dichotomy is between discrete (either finite or infinite) and continuous spaces. Continuous spaces require different mathematics, and we will take them up later. For the next several weeks, we will deal only with discrete spaces.

Example 8 Again, throw a dart at a circular dart board one foot in radius. Only now treat the possible outcomes as *(a)* the dart hits the bulls-eye (the central disk six inches in radius); *(b)* the dart hits the dart board outside of the bullseye; *(c)* the dart misses the board entirely. This is now a discrete sample space with three elements. See Figure 2.

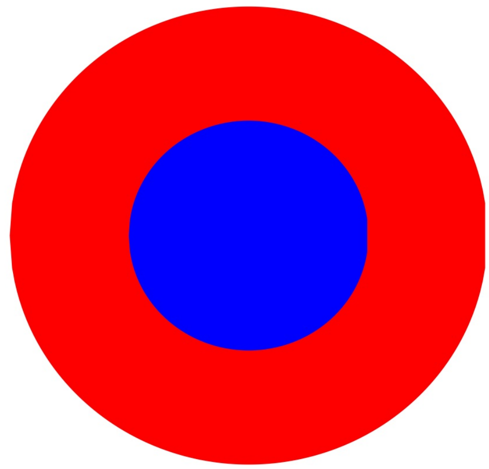

Figure 2: A dartboard with a bullseye. We can treat the sample space as a continuous space (outcomes are individual points) or a finite discrete space (outcomes are the three regions of the plane–bullseye, ring around the bullseye, and the set of points not on the board).

2 Probability mass function

aka PMF, probability distribution. This assigns a measure of likelihood–called the *probability* of the outcome– to each outcome in sample space.

Formally,

$$P : S \to [0, 1].$$

$[0, 1]$ denotes the *closed interval* $\{x \in \mathbf{R} : 0 \leq x \leq 1\}$. So this just says that for each $s \in S$, $P(s)$ is nonnegative and no more than 1.

One more requirement:

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

Example 9 For the coin toss, we set $P(0) = P(1) = \frac{1}{2}$. This models a 'fair coin'. Likewise, a fair die is modeled by $P(i) = \frac{1}{6}$ for all $i \in \{1, 2, 3, 4, 5, 6\}$. These are instances of *uniform distributions* on finite sample spaces, in which

$$P(s) = \frac{1}{|S|}$$

for all $s \in S$.

Example 10 For a sequence of N consecutive tosses of a fair coin, we can reason as follows. We take $N = 3$ just for the sake of concreteness, but the argument is the same for any N: If we repeat the experiment many times, then for about half of the trials, we will get heads on the first toss; then for half of the trials that gave this result, we will get tails on the second toss, so that all in all about $1/4$ of the trials will give HT for the first two tosses. And half of *those* trials will give tails on the third toss, so that roughly $1/8$ of the trials will give the sequence HTT. Thus when we model this experiment we set $P(HTT) = 1/8$. The identical reasoning applies to any sequence of three tosses, so that every outcome has probability $1/8$, meaning that this is a uniform PMF. In general, for a sequence of N tosses, we would assign every outcome probability 2^{-N}.

We can illustrate the reasoning using tree diagrams like the one in Figure 3 (see the caption).

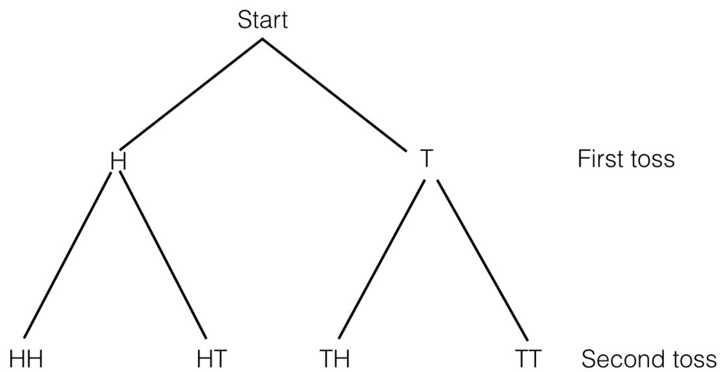

Figure 3: Tree diagram illustrating two successive tosses of a coin. For a fair coin, we reach each of the nodes on the first level in about half the trials, and on the next toss, we reach each of the four leaves of three in about one-quarter of the trials, so we model the experiment by a uniform distribution, with a probability of $1/4$ for each of the four outcomes. Such diagrams are very flexible–we can use them to deduce probabilities even for unfair coins, or a sample space in which the probability of heads on the second toss depends on what happened on the first toss.

Example 11 What about the roll of two fair dice? We can reason as above (and even illustrate it with a tree diagram): imagine that one of the dice is black with white dots and the other white with black dots. Approximately one-sixth of the time, the black die will come up 2, and approximately one-sixth of the times this happens, the white die will come up 3. So we should expect that a 2 on the black die and a 3 on the white die will occur about $\frac{1}{36}$ of the times we perform the experiment.

By the same reasoning, a 2 on the white die and a 3 on the black die should occur about $\frac{1}{36}$ of the times we perform the experiment, and likewise a 2 on both dice should occur about $\frac{1}{36}$ of the time. Thus ‘a 2 and a 3’ is twice as likely as ‘a pair of 2’s’. (This is borne out by actually performing the experiment.)

So if we want to model the roll of two fair dice, it is simplest to use

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

because the uniform distribution on S, assigning a probability $\frac{1}{36}$ to each outcome, models the behavior we described above. Observe that we *could* model the same behavior using the sample space

$$S = \{(i, j) : 1 \leq i \leq j \leq 6\},$$

but we would require a non-uniform distribution, setting, for example, $P((2, 2)) = \frac{1}{36}$ and $P((2, 3)) = \frac{1}{18}$.

Example 12 What if we had an unfair die in which 4,5,6 were each twice as likely as 1,2,3? We can find the PMF by setting $P(4) = P(5) = P(6) = x$. Then $P(1) = P(2) = P(3) = 2x$. Since the probabilities must add to 1, we get

$$3 \cdot 2x + 3x = 1,$$

so

$$x = \frac{1}{9}.$$

Example 13 A bin contains 500 black beans, 300 navy beans, and 200 pinto beans. You reach in without looking and pull out a bean. You can view the sample space as the set $\{\text{black}, \text{navy}, \text{pinto}\}$. If no one of the thousand beans is more likely to be puled out than any other, we have $P(\text{black}) = 0.5$, $P(\text{navy}) = 0.3$, $P(\text{pinto}) = 0.2$.

Alternatively, you can treat the sample space as $S = \{1, 2, \ldots, 1000\}$, where beans 1 to 500 are black, *etc.,* and use the uniform distribution. This is a different probability space but it models the same behavior.

Example 14 'Flip repeatedly until heads'.

With a fair coin, T,H should be equally likely for the first roll, TH,TT equally likely for the second roll, so we should get $P(1) = \frac{1}{2}$, $P(2) = \frac{1}{4}$, and generally, $P(i) = 2^{-i}$. We need to check that

$$\sum_{i \in \mathbf{Z}^+} P(i) = 1.$$

This is an infinite series, so its sum is the limit of partial sums. Just a brief digression on how we compute this: The sum of a finite *geometric* series

$$\sum_{i=0}^{n} r^i,$$

where $r \neq 1$, is given by the formula

$$\sum_{i=0}^{n} r^i = \frac{1 - r^{n+1}}{1 - r}.$$

If $-1 < r < 1$ then

$$\lim_{n \to \infty} r^{n+1} = 0,$$

so

$$\sum_{i=0}^{\infty} r^i = \frac{1}{1 - r}.$$

In particular, with $r = \frac{1}{2}$ we have

$$\sum_{i=0}^{\infty} 2^{-i} = \frac{1}{1 - \frac{1}{2}} = 2,$$

so

$$\sum_{i=1}^{\infty} 2^{-i} = 1,$$

since we have just cut off the first term 1. This is the property we required. By the way, in general, you cannot rearrange the terms of an infinite series any way you like and keep the sum the same, but you can if the terms are all positive. So it does not matter how we order the elements of our discrete infinite sample space.

We cannot have a uniform distribution on a discrete infinite sample space. Do you see why?

The PMF does not make sense for a continuous sample space (do you see why?) That is why we will need fundamentally different mathematics to study continuous spaces.

3 Probability Function

Definition. The probability function, which we also denote by P, assigns a value in the interval $[0, 1]$ to each *event* over the sample space. So formally, P is a function

$$P : \mathcal{P}(S) \to [0, 1].$$

Here $\mathcal{P}(S)$ denotes the *power set* of S, that is, the set of all subsets of S.

To define the probability function, we just extend the PMF:

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

In the case where the PMF is uniform, this simplifies to

$$P(E) = \frac{|E|}{|S|}.$$

Example 15 In the two-dice example 'rolling two sixes' is the event $\{(6,6)\}$ consisting of a single outcome. 'rolling an eight' is the event $\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$. If the experiment consists of tossing three coins, then 'at least two consecutive heads ' is the event $\{HHH, HHT, THH\}$. In the case of fair coins and fair dice, we use the uniform distribution, so the probabilities assigned to these events are respectively, $\frac{1}{36}, \frac{5}{36}, \frac{3}{8}$

Basic properties. The following basic properties of the probability function all follow easily from the definition:

- $P(E) \geq 0$ for every event E.
- $P(S) = 1$.
- If $E \cap F = \emptyset$ then
$$P(E \cup F) = P(E) + P(F).$$
- More generally, if $E_1, E_2, \ldots$ are pairwise disjoint events (meaning $E_i \cap E_j = \emptyset$ whenever $i \neq j$, then
$$P(\cup_i E_i) = \sum_i P(E_i).$$
The sum can be finite or infinite.
- For any events $E_1, E_2, \ldots,$
$$P\big(\bigcup_i E_i\big) \leq \sum_i P(E_i).$$
- For any event E, $P(\bar{E}) = 1 - P(E)$.

Set talk: disjoint sets (empty intersection). Probability talk: mutually exclusive events. *They mean the same thing.*