Probability Theory & Statistics

Sample Spaces & Events

Sample Spaces

Sample spaces are the set of all possible outcomes of an experiment. The following are a few examples of sample spaces

- the set of all 52 cards in a card deck: $S_1 = \{A\heartsuit, A\diamondsuit, \ldots\}$
- the set of tomorrow's weather conditions: $S_2 = \{Sunny,\ Overcast, \ldots\}$
- the set of all combinations of 2-flip coin experiments: $S_3 = \{HH, HT, TT, TH\}$
- the set of all possible temperatures (in Kelvin) in Bedford: $S_4 = \{T \in \mathbb{R}, T \geq 0\}$

The last example uses *set builder notation* which is mathematical shorthand notation for saying "all of the positive decimal temperatures".[1]

Events

Events are subsets of sample spaces. A possible subset from S_3 is the event "at least one head is observed", that is, $A_H = \{HH, HT, TH\} \subset S_3$. We say that A_H *occurred* if the actual outcome of an experiment is in A_H. If we perform an experiment and flip a coin twice, observing that the first outcome is a head, and the second outcome is a tail, the event "at least one head is observed" has occurred.

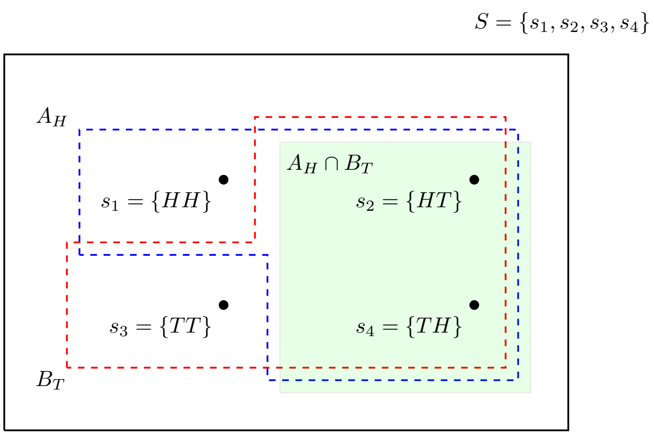

Figure 1: Sample space S of 2-flip coin outcomes s_j, and the events A_H: "at least one head is observed" in blue, B_T: "at least one tail is observed" in red, and their union $A_H \cap B_T$ "at least one head *and* one tail is observed" in green.

[1] Decimals with an infinite precision.

Finite Sample Spaces

Finite sample spaces and their *outcomes,* s_j, can be visualized as marbles, with events, collections of marbles. It's useful to think of a sample space as a bag that contains marbles that each represent an outcome in an experiment. Performing the experiment corresponds to reaching into the bag and choosing a marble, where the chance of choosing any given marble is the same or *equally likely.*[2] We will see later that this can be generalized to smaller or larger marbles where their *mass* makes them more or less likely to be chosen.

Figure 1 is the sample space of a 2-flip coin experiment. The enclosing boxes are events, collections of outcomes that are subsets of the sample space corresponding to "at least 1 head is observed" or "at least 1 tail is observed" observed in the 2-flip sequence. The shaded box is the event that at least 1 head *and* 1 tail is observed,

$$A_H \cap B_T = \{HT, TH\}. \tag{1}$$

The symbol $\cap$ is used to denote the *intersection* of (the outcomes of) events, the outcomes shared by two events.

Naïve Probability

Definition

The *naïve probability* that an event A occurs (on a finite sample space) is the number of outcomes favourable to A divided by the total number of outcomes. Using the same 2-flip coin example, the naïve probability of event A_H, that "at least 1 head is observed," can be calculated as

$$P(A_H) = \frac{|A_H|}{|S_3|} = \frac{|\{HH, HT, TH\}|}{|\{HH, HT, TT, TH\}|} = 3/4. \tag{2}$$

The *bar notation* $|\cdot|$, denotes the *size of a set*, for example $A = \{HH, TT\}$ has $|A| = 2$. Relating the naïve probability to the bag of marbles interpretation, assigns a mass $1/|S|$ to each of the marbles, and a total mass $P(S) = 1$.

Limitations

Naïve probability is often a good model to start with, for example:

- When there is symmetry, e.g., physical symmetry like in a coin, or choosing a card at random from a deck (no single card prefers to be picked over any other card)
- By design – it's a model assumption

But it has limitations

[2] The source of the randomness is *random sampling.*

- There is no difference between the probability of "there is life on Mars" and "there is intelligent life on Mars". (We would expect that the second statement is far less likely, but it is still $\frac{1}{2}$ by the naïve definition of probability)
- Sample spaces with an infinite number of possible outcomes (possible temperature in Bedford tomorrow) have 0 probability of any single outcome occurring since $|S_4| = \infty$

Conditioning

Events that share possible outcomes (they *intersect*) change the probabilities associated with them if either of the events occur (we will later formally define this as "dependence", but it's more useful to think of the events as containing information about each other).

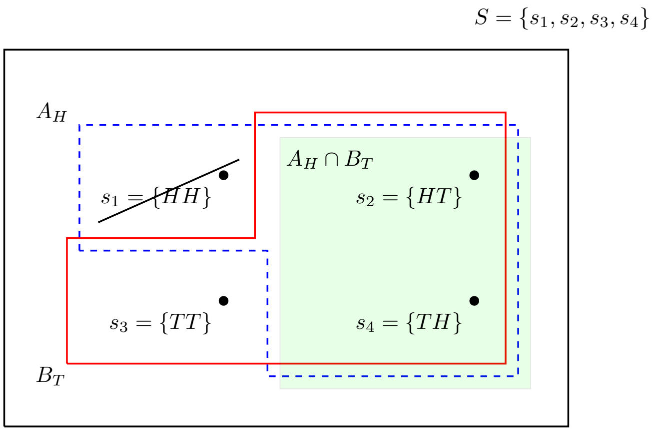

Figure 2: A 2-flip coin experiment outcome space after observing B_T: "at least 1 tail is observed", means its complement the event $\overline{B_T}$: "no tails are observed" cannot occur.

For a 2-flip coin experiment, if the event "at least 1 tail is observed" occurs, it's impossible for the outcome $s_1 = \{HH\}$ to occur. The probability of the event "at least 1 head is observed" is then changed (it is less likely).

We define the *conditional probability*, i.e., the probability of an event A occurring given an event B has occurred, as

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \tag{3}$$

To see why this is true, consider the 2-flip coin experiment again. Upon obtaining the information that "at least 1 tail" has been observed in an actual experiment – conditioning on B_T, we can remove the marbles from the sample space S_3 that aren't in B_T since they are incompatible with the knowledge that B_T has occurred.

Elements "not in" any given set are denoted with the *complement notation* $\overline{A}$. In the 2-flip coin experiment, the complement of "at least 1 tail" is "no tails",

$$\overline{B_T} = \{HH\} \tag{4}$$

Having removed this event from the sample space, the total mass of the marbles that remain in the event A_H, "at least 1 head", is $P(A_H \cap B_T) = 2/4 = 1/2$. But only 3 outcomes are left in the sample space totalling a mass 3/4, and each outcome must be equally likely, i.e., for a probability to be valid, the total mass of marbles must sum to 1. To achieve this, we *renormalize* the sample space by dividing through by the probability of the event B_T

$$P(A_H|B_T) = \frac{P(A_H \cap B_T)}{P(B_T)} = \frac{2/4}{3/4} = 2/3. \tag{5}$$

This assigns the correct "mass" to the event A_H conditional on the occurrence of the event B_T.

Discussion

Random Variables

How are events related to random variables and distributions?

Random variables are functions, $X(\cdot)$, that map elements in the outcome space to the real number line, i.e., $X(s_j) = x$, for $x \in \mathbb{R}$. Distributions provide the "blueprint" for the probability of events like $\{X = x\}$.

Practice

Finite Sample Spaces

1a) 2 coins were flipped and at least one of them landed heads. What is the probability that both coins landed heads?

1b) Imagine m coins flipped at the same time and their outcomes recorded. What is the probability that all the coins land heads, given at least one of them landed heads?

2a) Show that the probability of the complement of an event A can be calculated from the naïve probability as

$$P\left(\overline{A}\right) = 1 - P(A). \tag{6}$$

2b) Consider the sample space of 52 playing cards, S_1. How many possible events (subsets of S_3) are there?

Conditioning

3a) Using the definition of conditional probability, derive *Bayes' Theorem* that relates the probability of event A given B has occurred with the probability of B given A has occurred

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \tag{7}$$

3b) Show that $P(A|A) = 1$.