Probability

Probability is the language we use to model uncertainty. If we are to develop a proper understanding of such experimental results, we need to be able to understand the randomness underlying them. We shall briefly consider the following terminologies:

1 Introduction and Definitions

Probability is the language we use to model uncertainty. If we are to develop a proper understanding of such experimental results, we need to be able to understand the randomness underlying them. We shall briefly consider the following terminologies.

Sample spaces

Probability theory is used as a model for situations for which the outcomes occur randomly. Generically, such situations are called *experiments*, and the set of all possible outcomes of the experiment is known as the *sample space* corresponding to an experiment. The sample space is usually denoted by S, and a generic element of the sample space (a possible outcome) is denoted by s. The sample space is chosen so that exactly one outcome will occur. The size of the sample space is *finite*, *countably infinite* or *uncountably infinite*.

Event

Trial is an act performed of experiment. *Outcomes* is a result realized from the trial. A subset of the sample space (a collection of possible outcomes) is known as an *event*.

Example. When a coin is tossed twice, Sample Space, $S = \{HH, HT, TH, TT\}$. Define the event A as: at least one head is observed. We have $A = \{HH, HT, TH\}$.

Two events A and B are *disjoint* or *mutually exclusive* if they cannot both occur. That is, their intersection in empty $A \cap B = \emptyset$.

2 Probability axioms

Now that we have a good mathematical framework for understanding events in terms of sets, we need a corresponding framework for understanding probabilities of events in terms of sets. The real valued function $P(\cdot)$ is a *probability measure* if it acts on subsets of S and obeys the following axioms:

1. $P(S) = 1$
2. If $A \subseteq S$ then $P(A) \geq 0$
3. If A and B are disjoint ($A \cap B = \emptyset$) then $P(A \bigcup B) = P(A) + P(B)$.

Repeated use of Axiom 3 gives the more general result that if $A_1, A_2, ..., A_n$ are mutually disjoint, then

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

Indeed, we will assume further that the above result holds even if we have a countably infinite collection of disjoint events ($n = \infty$).

Example. Suppose that a fair coin is thrown twice, and the results recorded. The sample space is

$$S = \{HH, HT, TH, TT\}.$$

Let us assume that each outcome is equally likely — that is, each outcome has a probability of $1/4$. Let A denote the event *head on the first toss*, and B denote the event *head on the second toss*. In terms of sets,

$$A = \{HH, HT\}, B = \{HH, TH\}.$$

So $P(A) = \frac{n(A)}{n(S)} = \frac{2}{4} = \frac{1}{2}$ and similarly $P(B) = 1/2$. If we are interested in the event $C = A \bigcup B$ we can work out its probability using from the set definition as $P(C) = \frac{n(C)}{n(S)} = \frac{n(A \bigcup B)}{4} = \frac{n(\{HH,HT,TH\})}{4} = \frac{3}{4}$.

3 Conditional probability

We now have a way of understanding the probabilities of events, but so far we have no way of *modifying* those probabilities when certain events occur. For this, we need an extra axiom which can be justified under any of the interpretations of probability. The axiom defines the *conditional probability* of A given B, written $P(A|B)$ as

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

for $P(B) > 0$. Note that we can only condition on events with positive probability.

Under the classical interpretation of probability, we can see that if we are told that B has occurred, then all outcomes in B are equally likely, and all outcomes not in B have zero probability — so B is the new sample space. The number of ways that A can occur is now just the number of ways $A \cap B$ can occur, and these are all equally likely. Consequently we have

$$P(A|B) = \frac{n(A \cap B)}{n(B)} = \frac{n(A \cap B)/n(S)}{n(B)/n(S)} = \frac{P(A \cap B)}{P(B)}.$$

Because conditional probabilities really just correspond to a new probability measure defined on a smaller sample space, they obey all of the properties of "ordinary" probabilities. For example, we have

$$\begin{aligned} P(B|B) &= 1 \\ P(\emptyset|B) &= 0 \\ P(A \cup C|B) &= P(A|B) + P(C|B), \ for\ A \cap B = \emptyset \end{aligned}$$

and so on. The definition of conditional probability simplifies when one event is a special case of the other. If $A \subseteq B$, then $A \cap B = A$ so

$$P(A|B) = \frac{P(A)}{P(B)}, \ for\ A \subseteq B.$$

Example. A die is rolled and the number showing recorded. Given that the number rolled was even, what is the probability that it was a six?

Let E denote the event "even" and F denote the event "a six". Clearly $F \subseteq E$, so

$$P(F|E) = \frac{P(F)}{P(E)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

The multiplication rule

The formula for conditional probability is useful when we want to calculate $P(A|B)$ from $P(A \cap B)$ and $P(B)$. However, more commonly we want to know $P(A \cap B)$ and we know $P(A|B)$ and $P(B)$. A simple rearrangement gives us the multiplication rule.

$$P(A \cap B) = P(B) \times P(A|B)$$

Example. Two cards are dealt from a deck of 52 cards. What is the probability that they are both Aces?

We now have three different ways of computing this probability. First, let's use conditional probability. Let A_1 be the event "first card an Ace" and A_2 be the event "second card an Ace". $P(A_2|A_1)$ is the probability of a second Ace. Given that the first card has been drawn and was an Ace, there are 51 cards left, 3 of which are Aces, so $P(A_2|A_1) = 3/51$. So,

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1) \times P(A_2|A_1) \\ &= \frac{4}{52} \times \frac{3}{51} = \frac{1}{221}. \end{aligned}$$

The multiplication rule generalizes to more than two events. For example, for three events we have

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \bigcap A_2).$$

Independence

For some events A and B, knowing that B has occurred will not alter the probability of A, so that $P(A|B) = P(A)$. When this is so, the multiplication rule becomes

$$P(A \cap B) = P(A)P(B)$$

and the events A and B are said to be independent events. Independence is a very important concept in probability theory, and is used a lot to build up complex events from simple ones. Do not confuse the independence of A and B with the exclusivity of A and B — they are entirely different concepts. If A and B both have positive probability, then they cannot be both independent and exclusive (exercise).

When it is clear that the occurrence of B can have no influence on A, we will assume independence in order to calculate $P(A \cap B)$. However, if we can calculate $P(A \cap B)$ directly, we can check the independence of A and B by seeing if it is true that

$$P(A \cap B) = P(A)P(B).$$

Example. A playing card is drawn from a pack. Let A be the event "an Ace is drawn" and let C be the event "a Club is drawn". Are the events A and C exclusive? Are they independent?

A and C are clearly not exclusive, since they can both happen — when the Ace of Clubs is drawn. Indeed, since this is the only way it can happen, we know that $\mathrm{P}(A \cap C) = 1/52$. We also know that $P(A) = 1/13$ and that $P(C) = 1/4$. Now since

$$\begin{aligned} P(A)P(C) &= \frac{1}{13} \times \frac{1}{4} \\ &= \frac{1}{52} \\ &= P(A \cap C) \end{aligned}$$

we know that A and C are independent. Of course, this is intuitively obvious — you are no more or less likely to think you have an Ace if someone tells you that you have a Club.

4 Bayes Theorem

Partitions

A *partition* of a sample space is simply the decomposition of the sample space into a collection of mutually *exclusive* events with positive probability. That is, $\{B_1, ..., B_n\}$ form a partition of S if

- $S = B_1 \cup B_2 \cup \cdots \cup B_n = \bigcup_{i=1}^{n} B_i$,
- $B_i \cap B_j = \emptyset$, for any $i \neq j$,
- $P(B_i) > 0$, for any i.

Example. A card is randomly drawn from the pack. The events $\{C, D, H, S\}$ (Club, Diamond, Heart, Spade) form a partition of the sample space, since one and only one will occur, and all can occur.

Theorem of total probability

Suppose that we have a partition $\{B_1, ..., B_n\}$ of a sample space,S. Suppose further that we have an event A. Then A can be written as the disjoint union

$$A = (A \cap B_1) \cup \cdots \cup (A \cap B_n),$$

and so the probability of A is given by

$$\begin{aligned} P(A) &= P((A \cap B_1) \cup \cdots \cup (A \cap B_n)) \\ &= P(A \cap B_1) + \cdots + P(A \cap B_n), \quad \textit{by Axiom III} \\ &= P(A|B_1)P(B1) + \cdots + P(A|B_n)P(B_n), \qquad \textit{by the multiplication rule} \\ &= \sum_{i=1}^{n} P(A|B_i)P(B_i). \end{aligned}$$

Exercise. Craps is a game played with a pair of dice. A player plays against a banker. The player throws the dice and notes the sum.

- If the sum is 7 or 11, the player wins, and the game ends (a natural).
- If the sum is 2, 3 or 12, the player looses and the game ends (a crap).
- If the sum is anything else, the sum is called the players *point*, and the player keeps throwing the dice until his sum is 7, in which case he looses, or he throws his *point* again, in which case he wins.

What is the probability that the player wins?

Bayes Theorem

From the multiplication rule, we know that

$$P(A \cap B) = P(B)P(A|B)$$

and that

$$P(A \cap B) = P(A)P(B|A)$$

so clearly

$$P(B)P(A|B) = P(A)P(B|A),$$

and so

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

This is known as *Bayes Theorem*, and is a very important result in probability, as it tells us how to "turn conditional probabilities around" — that is, it tells us how to work out $P(A|B)$ from $P(B|A)$, and this is often very useful.

Example. Example A clinic offers you a free test for a very rare, but hideous disease. The test they offer is very reliable. If you have the disease it has a 98% chance of giving a positive result, and if you don't have the disease, it has only a 1% chance of giving a positive result. You decide to take the test, and find that you test positive — what is the probability that you have the disease?

Let A be the event "test positive" and D be the event "you have the disease". We know that $P(A|D) = 0.98$ and that $P(A|D^c) = 0.01$. We want to know P(D|P), so we use Bayes' Theorem.

$$\begin{aligned} P(D|A) &= \frac{P(A|D)P(D)}{P(A)} \\ &= \frac{P(A|D)P(D)}{P(A|D)P(D) + P(A|D^c)P(D^c)} \\ &= \frac{0.98P(D)}{0.98P(D) + 0.01\,(1 - P(D))}. \end{aligned}$$

So we see that the probability you have the disease given the test result depends on the probability that you had the disease in the first place. This is a rare disease, affecting only one in ten thousand people, so that $P(D) = 0.0001$. Substituting this in gives

$$P(D|A) = \frac{0.98 \times 0.0001}{0.98 \times 0.0001 + 0.01 \times 0.9999} \simeq 0.01.$$

So, your probability of having the disease has increased from 1 in 10,000 to 1 in 100, but still isn't that much to get worried about! Note the crucial difference between $P(A|D)$ and $P(D|A)$.

Bayes Theorem for partitions

Another important thing to notice about the above example is the use of the theorem of total probability in order to expand the bottom line of Bayes Theorem. In fact, this is done so often that Bayes Theorem is often stated in this form.

Suppose that we have a partition $\{B_1, ..., B_n\}$ of a sample space S. Suppose further that we have an event A, with $P(A) > 0$. Then, for each B_j , the probability of B_j given A is

$$\begin{aligned} P(B_j|A) &= \frac{P(A|B_j)P(B_j)}{P(A)} \\ &= \frac{P(A|B_j)P(B_j)}{P(A|B_1)P(B_1) + ... + P(A|B_n)P(B_n)} \\ &= \frac{P(A|B_j)P(B_j)}{\sum\limits_{i=1}^{n} P(A|B_i)P(B_i)}. \end{aligned}$$

In particular, if the partition is simply $\{B, B^c\}$, then this simplifies to

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}.$$