Introduction to Probability

We are familiar with “propositions” (also called “statements” or “assertions”). These are statements that have definite true or false values. For example:

- This is lecture hall 6
- We are studying MTH102
- It is cold in January in Mohali

However, there are similar statements about which we are not so definite:

- It will be cold in March in Mohali
- Z likes Mathematics
- X will attend all lectures in MTH202

Most of such statements are about things in the *future* which have not yet happened, so we cannot be certain that they will happen. Other statements are *subjective* (depend on the observer).

We would like to extend the notion of true/false to such statements (to the extent possible) as well. To do this, we can replace true by “certain” and give it the value 1, and we can replace false by “unlikely” and give it the value 0. To various shades of certainity we can give values between 0 and 1. (At this point it is worth pointing out the distinction between “improbable” and “impossible”[1]).

In order to apply this approach, let us first recall the basics of the calculus of propositions (which was probably introduced in MTH102 and also in MTH101). We will then see how this calculus is to be modified to work with probabilities.

Propositional Calculus

We will denote propositions by capital letters of the Roman alphabet. Propositions can be combined to produce new propositions.

- $A \wedge B$ represents the combination A ‘and’ B.
- $A \vee B$ represents the combination A ‘or’ B.
- $\neg A$ represents the negation of A (‘not’ A).

[1] In “Leave it to Psmith” by P G Wodehouse, Psmith says: ‘Comrade Spiller, never confuse the unusual the impossible.’

These are the basic operations of “Boolean Algebra”. If we use 0 for ‘false’ and 1 for ‘true’ as suggested above, then we have the identities:

- $A \wedge A = A = A \vee A$.
- $\neg(A \wedge B) = (\neg A) \vee (\neg B)$.
- $\neg(\neg A) = A$.
- $A \wedge (\neg A) = 0$.
- $A \vee (\neg A) = 1$.
- $A \wedge (B \wedge C) = (A \wedge B) \wedge C$.
- $A \vee (B \vee C) = (A \vee B) \vee C$.
- $A \vee B = (B \vee A)$.
- $A \wedge B = (B \wedge A)$.
- $A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$.

In your course (IDC102) on electronics you must have encountered ‘nand’ gates which are defined by $A \dagger B = \neg(A \wedge B)$. You must have also learned that all other operations can be obtained from this by using the identities:

- $\neg(A) = A \dagger A$
- $A \vee B = (A \dagger A) \dagger (B \dagger B)$
- $A \wedge B = (A \dagger B) \dagger (A \dagger B)$

A more algebraic approach is to define the ‘xor’ (exclusive or) of two prpositions:

- $A \oplus B$ represents the combination A ‘or’ B but ‘not both’.

This can be written as $A \oplus B = (A \vee B) \wedge \neg(A \wedge B)$. This If we use the suggestive notation $A \otimes B$ in place of $A \wedge B$, then the above rules simplify into the familiar algebraic rules:

- $A \oplus B = B \oplus A$.
- $A \otimes B = B \otimes A$.
- $A \otimes (B \oplus C) = A \otimes B \oplus A \otimes C$.
- $A \oplus 0 = A$
- $A \otimes 1 = A$.
- $A \otimes 0 = 0$.

Along with these familiar rules, we have the slightly unfamiliar rules $A \oplus A = 0$ and $A \otimes A = A$.

Probability

The subsets of a set with the composition rules of intersection and union behave very similarly to propositions.

- $A \cap B$ represents the intersection of A and B. This is similar to the $\wedge$ operation on propositions.
- $A \cup B$ represents the union of A and B. It is similar to the $\vee$ operation on propositions.
- A^c represents the complement of a set. It is similar to the negation operation on propositions.

It is an exercise to check that the above identities hold good when we make the corresponding replaces.

The reason for this similarity is not hard to find. We can think of a subset A as the set of elements x so that a certain property P is satisfied by x. We can also think of each consistent assignment of truth or falsity of each proposition as an "instance of reality". The subset of those instances where the proposition A is true is the subset associated with A.

We can now extend this to "possible realities" or "sample points". Each experimental observation results in a "sample point". The subset associated with a "prediction" is the collection of sample points where this prediction holds. In probability theory, these predictions (or the associated set of sample points) are called "events" and we assign a probability to each event. For this reason (and for conventional reasons) we will use the set-theoretic notation of unions, intersections and complements from now on (and not the notation of $\wedge$, $\vee$ and $\neg$). The basic rules of probability are:

- The probability of the "empty" event is 0, i. e. $P(\emptyset) = 0$.
- If $A \subset B$ then $P(A) \leq P(B)$.
- The "excluded middle" law, $P(A) + P(A^c) = 1$.
- The addition law $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

We use Ω to denote the "universe" of all possible results of the experiment; it is also called the "sample space" and we have $\Omega = (\emptyset)^c$. It follows that $P(\Omega) = 1$. Moreover, since any A is a subset of Ω, we obtain an important rule (which we should never forget!):

$$0 \leq P(A) \leq 1$$

Any calculation that purports to give a probability where the answer does not satisfy this is *obviously* wrong!

Bayes Rule

While giving the basic rules governing probability, we have said nothing about how to assign probabilities other than to say that any such assignment should be consistent with the rules.

In practice, we assign probabilities based on information about events that has already been gathered.

Let's take a specific and common example. When we flip a coin, we have no information about whether it will come heads or tails. So we can assign an equal probability (of half!) to each event since we do not expect it to stand on its edge! Similarly, when we first go to a new city, we can assign an equal probability of finding the food nice or not nice!

However, the two events behave differently after we have made a number of observations.

In the case of coin flips, we generally have the feeling (especially if the coin andthe person flipping it has changed!) that the knowledge of one 100 coin flips gives us no information about the result of the 101st coin flip.

On the other hand after eating in the mess for 100 days, we have a rather good idea about whether we will like the food on the 101st day or not!

We use the notion $P(A|B)$ to denote the probability that the event A occurs if we are *given* that B has occurred. Bayes rule is:

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

We *could* treat this as "defining" $P(A|B)$ provided that we note that $P(A|B)$ is not defined if $P(B) = 0$. However, it can also be seen as a way that we determine $P(A \cap B)$. (Intelligence can be seen as the capacity to gather information and convert it into conditional probabilites!)

Let H_n denote the probability of heads on then n-th coin toss, then the above statement about coin tosses can be written as:

$$P(H_{n+1}|E) = P(H_{n+1})$$

where E denotes an event that represents the sequence of H_r's (heads) and H_r^c (tails) that occur on the previous n tosses ($r \leq n$).

We see that this implies that $P(H_{n+1} \cap E) = P(H_{n+1})P(E)$. We generalise this and define A and B to be *independent* events if $P(A \cap B) = P(A)P(B)$. Equivalently, idendependence

Mathematical Summary

We can approach probability as a *formal* theory without worrying about its interpretation. This can help us avoid confusion during calculations. This formal structure is given below.

In probability theory we study the algebra of events, which is an algebra of subsets of the sample space Ω with the usual operations of union, intersection and complement.

To each event A we assign a probability $P(A)$, which is a number between 0 and 1. This satisfies the following rules:

- The probability of the "empty" event is 0, i. e. $P() = 0$.
- If $A \subset B$ then $P(A) \leq P(B)$.
- The "excluded middle" law, $P(A) + P(A^c) = 1$.
- The addition law $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

If B is an event for which $P(B) > 0$ then we define $P(A|B)$ by the identity $P(A|B)P(B) = P(A \cap B)$. When $P(B) = 0$ we can define $P(A|B) = P(A)$.

We say that A is idependent of B if $P(A|B) = P(A)$ or equivalently if $P(A \cap B) = P(A)P(B)$.