

Laws of Probability

Probability Theory studies the assignment of probabilities to Events. Events are statements like propositions for which it is not immediately clear whether they are true or false. For example:

- It will be cold in March in Mohali
- Arul likes Mathematics
- Beena 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. However, we can also assign probabilities to statements about events in the past as we are no longer sure that they happened!

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 certainty we can give values between 0 and 1. (At this point it is worth pointing out the distinction between “improbable” and “impossible”¹).

Events

The *algebra* of events will be a Boolean Algebra just like the algebra of propositions. In particular, we will have the notions:

- $A \wedge B$ is the event that both A and B will occur
- $A \vee B$ is the event that at least one of A and B will occur
- A^c which the event that A will not occur

By defining $A \otimes B = A \wedge B$ and $A \oplus B = (A \vee B) \wedge (A \wedge B)^c$, we require that the algebraic system obtained is a Boolean algebra. For completeness, we will also require the notion of an event Ω which is the union of all possible events and ϕ as the intersection of all possible events. (Since $(\vee_i A_i)^c = \wedge_i A_i^c$ we can see that $\Omega^c = \phi$.)

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

Laws of Probability

To each event A we assign a probability $P(A)$ which is a real number between 0 and 1 called the probability of (occurrence of) the event A . This satisfies the following laws:

- The probability of the universal event is 1, i. e. $P(\Omega) = 1$.
- If $A \subset B$ then $P(A) \leq P(B)$.
- The “excluded middle” law, $P(A) + P(A^c) = 1$.
- The addition law $P(A \wedge B) = P(A) + P(B) - P(A \vee B)$.

We have $\Omega = (\phi)^c$. It follows that $P(\phi) = 0$. 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!

We note that

$$(A \wedge B) \wedge (A \wedge B^c) = \phi \text{ and } (A \wedge B) \vee (A \wedge B^c) = A$$

and so it follows that $P(A) = P(A \wedge B) + P(A \wedge B^c)$.

Mutual Exclusivity

The above formulas and rules do not really give too much of a handle on the probability $P(A \wedge B)$ or $P(A \vee B)$ in terms of $P(A)$ and $P(B)$. However, there are some cases when we are *certain* that A and B cannot happen together. In other words $P(A \wedge B) = 0$. Such events A and B are called “mutually exclusive”. The terminology indicates that if A occurs, it excludes B from occurring and vice versa.

For mutually exclusive events, we have $P(A \vee B) = P(A) + P(B)$. Conversely, when this holds, we have $P(A \wedge B) = 0$ by the law of addition of probabilities.

The simplest case then this holds is when $B = A^c$ since, in this case, $A \wedge A^c = \phi$ so $P(A \wedge A^c) = 0$ and equivalently $P(\Omega) = 1 = P(A) + P(A^c)$.

Another important realisation is that *if* $P(A) + P(B) > 1$, then A and B *cannot* be mutually exclusive since $P(A \vee B) \leq 1$ which means that $P(A \wedge B) > 0$ by the law of addition of probabilities.

A simple case of exclusive events would be the following in the case of the flip of a single coin. The event A would be the assertion that this particular flip resulted in a Head and the event B would be the assertion that this particular flip resulted in a Tail. The probability of getting both a Head *and* a Tail in the same flip is evidently 0 (under normal conditions!).

Conditional Probability

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 laws of probability!

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 appear to us differently after we have made a number of observations.

In the case of coin flips, we generally have the feeling (especially if the coin and the 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. We can also think of B as the data gathered about the universe and A as the prediction based on this data. In that case $P(A|B)$ can be seen as the probability of correct-ness of our prediction A given the data B already gathered. (Intelligence can be seen as the capacity to gather information and convert it into conditional probabilities!)

Note that it does not make sense to think about $P(A|B)$ when $P(B) = 0$. Basically, how can we determine the probability of an event A occurring, when the improbable ($B!$) has already happened. We will see a more mathematical justification below.

Since $P(A|B)$ treats B as "background data", we can think of it as another way of assigning probability to the same events. Hence, it satisfies:

- $0 \leq P(A|B) \leq 1$.
- $P(\Omega|B) = 1$.

- If $A \subset A'$ then $P(A|B) \leq P(A'|B)$.
- The “excluded middle” law, $P(A|B) + P(A^c|B) = 1$.
- The addition law $P(A \wedge A'|B) = P(A|B) + P(A'|B) - P(A \vee A'|B)$.

Kolmogorov’s Formula and Independence

A very important rule that allows us to link conditional probability to the probability of individual events is Kolmogorov’s formula:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} \text{ if } P(B) > 0$$

We can also write this as

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

We *could* treat this as “defining” $P(A|B)$ provided that we note that $P(A|B)$ is not defined if $P(B) = 0$. In fact, if $P(B) = 0$, then $P(A \cap B) = 0$ since $A \cap B$ is contained in B . Thus, $P(A|B)$ is like $0/0$ and is undefined. Conceptually, we can also see this it as a way to *determine* $P(A \wedge B)$ using $P(A|B)$ when $P(B) > 0$.

As a special case, we note that $P(A|\Omega) = P(A)$. This *might* seem confusing since one might believe that $P(A)$ is the probability when we have *no* information. However, we note that saying that Ω has occurred gives no information since we *know* that $P(\Omega) = 1$, so it *has* to happen!

There are cases where $P(A|B) = P(A)$ and $P(B|A) = P(B)$. In other words, where the knowledge that the event B has occurred tells us nothing about the event A and vice versa. In this case we have $P(A \cap B) = P(A)P(B)$. If the latter identity holds, we say that A and B are *independent*. Consequently, we have

$$P(A \vee B) = P(A) + P(B) - P(A)P(B) \text{ when } A \text{ and } B \text{ are independent}$$

Deciding whether events are independent is not always easy in a practical situation. In fact, ensuring that certain physical quantities can be measured independently is impossible according to quantum mechanics! However, in the cases we will study it will mostly be evident.

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 an algebra of events, which is a Boolean Algebra.

To each event A we assign a probability $P(A)$, which is a number between 0 and 1. More generally, for events A and B , we define $P(A|B)$ which is a number between 0 and 1.

This satisfies the following rules:

- 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 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

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 independent of B if $P(A \cap B) = P(A)P(B)$.