Lecture 2: Event relations, Kolmogorov's axioms of probability

1 Event relations

There are also a few basic relations among different events. Let A, B be two events.

(i) *Inclusion*. We denote $A \subseteq B$ if A is a subset of B (A happens implies B happens).
(ii) *Disjointness*. We say that A, B are *disjoint*, if A and B have no common outcomes (i.e. $A \cap B = \emptyset$). Similarly, a sequence of events $A_1, A_2, \cdots$ are said to be *mutually disjoint* if no two of them have common outcomes, i.e. if $A_i \cap A_j = \emptyset$ for any $i \neq j$.
(iii) *Exhaustiveness*. We say that A, B are *exhaustive*, if they contain all possible outcomes (i.e. $A \cup B = \Omega$). Similarly, a sequence of events $A_1, A_2, \cdots$ are said to be exhaustive, if $\cup_{n=1}^{\infty} A_n = \Omega$.

Having introduced that many concepts, all of them are identical to the corresponding concepts in set theory. Here we are simply rephrasing them using the language of outcomes and events. We often draw a diagram called the *Venn diagram* to help illustrating the above relations and operations. The following figure gives a few such examples:

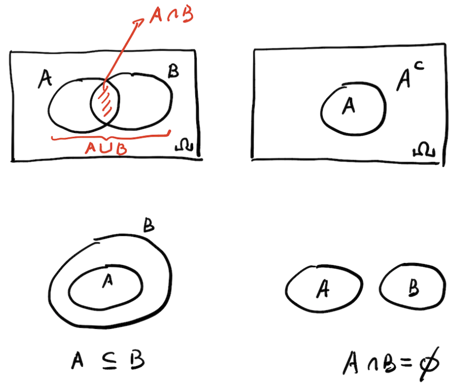

Figure 1: Venn diagrams.

Example 1.1. In the experiment of throwing a die, we have the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$. Consider the following events:

$$A = \{1, 3, 5\},\ B = \{2, 4, 6\},\ C = \{1, 2, 4, 6\},\ D = \{2, 4\}.$$

The following relations are straight forward:

(i) A, B are exhaustive and disjoint;
(ii) A, C are exhaustive but not disjoint;
(iii) A, D are disjoint but not exhaustive.

Example 1.2. Let A be an event. Then A, A^c are disjoint (since $A \cap A^c = \emptyset$) and exhaustive (since $A \cup A^c = \Omega$).

Since events are subject to set relations and operations, we immediately have the following two important properties inherited from set theory. They are very useful for understanding event operations and constructing new events.

Distributive laws:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

De Morgan's laws:

$$(A \cup B)^c = A^c \cap B^c,$$
$$(A \cap B)^c = A^c \cup B^c.$$

Showing two events are equal

There is a standard element-wise way to show that two events A, B are equal. It consists of two basic steps:

Step one. Let ω be an arbitrary outcome in A. Show that it belongs to B.
Step two. Let ω be an arbitrary outcome in B. Show that it belongs to A.

We illustrate this method by showing the first distributive law

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

Step one. Suppose that ω belongs to the left hand side. Then "$\omega \in A$" and "$\omega \in B$ or C". If $\omega \in B$, we know that it belongs to A and B, thus belong to the right hand side. Similarly, if $\omega \in C$, it belongs to A and C, thus also belonging to the right hand side.
Step two. Now suppose that ω belongs to the right hand side. Then $\omega \in A \cap B$ or $\omega \in A \cap C$. If it is the first case, since belonging to B implies belonging to $B \cup C$, we conclude that $\omega \in A$ and $B \cup C$, namely it belongs to the left hand side. The second case is similar.

It is also helpful to draw the following Venn diagram to illustrate the above distributive law (as well as the other three laws):

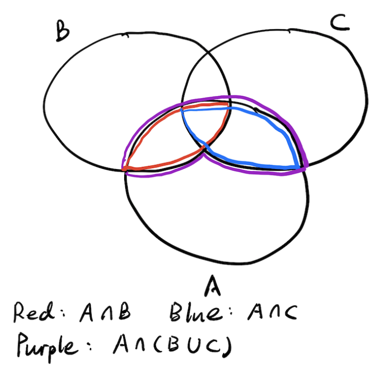

Figure 2: Venn diagram for the first distributive law.

But please keep in mind that the Venn diagram is *not* considered as a precise argument for establishing event relations – it is only used for intuitive demonstrative purpose.

The above element-wise argument is tedious and not enlightening. In many situations, it is often simpler to directly apply the distributive and De Morgan's laws rather than performing the element-wise argument.

2 Assigning probabilities: Kolmogorov's axiomatic approach

As we have mentioned earlier, a primary task of probability theory is to give meaning to probabilities of events. This is about assigning a function $A \mapsto \mathbb{P}(A)$ on a given class of events, so that $\mathbb{P}(A)$ represents the "probability/likelihood" of the event A heuristically.

Before Kolmogorov

One naive approach to define the probability of a given event, is to perform the random experiment independently and repeatedly for a large amount of times and observe the relative frequency for the occurrence of the event A. For instance, if we toss a fair coin for many times, we may observe that the frequency of seeing a Head is relatively stabilised around 0.5. It is then natural to *define* that the probability of having a Head is 0.5.

However, this approach is not satisfactory for several reasons. Firstly, it is unrealistic to repeatedly perform a random experiment for a large amount of times subject to identical conditions. Some experiments are even non-performable in practice and can only exist theoretically (for instance, the event that Betelgeuse will explode into supernova in the 22nd century). Secondly, returning to the coin tossing example, it is not entirely convincing that the relative frequency is stabilising at a theoretical number, and why it is 0.5 rather than 0.49999. Even if the number 0.5 is "theoretically" reasonable, the way of "proving" this (known as the *law of large numbers*) relies critically on a precise mathematical formulation of probability which is missing in the first place.

Kolmogorov's axiomatic approach to probability

In the 1930s, Kolmogorov introduced an axiomatic approach to define probabilities. It was a milestone for modern probability theory, turning the subject into a rigorous branch of mathematics for the first time.

The philosophy of Kolmogorov's approach can be summarised as follows. Recall that, the formal symbol $A \mapsto \mathbb{P}(A)$ represents a "probability function" acting on a given family of events, so that for each event A the quantity $\mathbb{P}(A)$ is interpreted as the "probability" of A. Instead of specifying this function precisely, Kolmogorov first postulated several fundamental properties that any such "probability function" should satisfy. This set of rules are so natural to expect and are thus treated as the *axioms* for all generic probability functions. Many interesting properties of probability functions can then be obtained from these axioms solely based on logical deduction. These axioms do not give a way of specifying a particular probability function, and any function satisfying these axioms is considered to be mathematically legal. It is then a practical matter to construct a particular one (subject to the axioms), which corresponds to the reality that we are considering.

There are three axioms in Kolmogorov's framework, which are all heuristically convincing.

The first axiom asserts that, the probability of an event is always non-negative. Mathematically, this is saying that $\mathbb{P}(A) \geqslant 0$ for all event A.

The second axiom asserts that, the certain event Ω has probability 1. Mathematically, $\mathbb{P}(\Omega) = 1$.

The last axiom is concerned with certain additivity property. The naive version is that, for any pair of disjoint events A and B, one should have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

This property clearly extends to finitely many events:

$$A_1, \cdots, A_n \text{ mutually disjoint} \implies \mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n).$$

However, for both theoretical and practical reasons, this "finite additivity property" is not strong enough for our use, and we often need to consider a sequence of events $A_1, A_2, \cdots$. The correct extension of this property is the following so-called *countable additivity*, which is the third axiom of Kolmogorov: for any sequence of mutually disjoint events $A_1, A_2, \cdots, A_n, \cdots$, one has

$$\mathbb{P}\big(\cup_{n=1}^{\infty} A_n\big) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

The need for considering a sequence of events is reflected from the following simple example.

Example 2.1. Throw a die repeatedly until forever. Consider the event

$$A : \text{a Six eventually appears.}$$

To analyse this event, for each $n \geqslant 1$ we define A_n to be the event that "a Six appears for the first time on the n-th toss". Then the sequence of events $A_1, A_2, \cdots$ are mutually disjoint, and we have the relation

$$A = \cup_{n=1}^{\infty} A_n.$$

The countable additivity axiom then allows us to compute the probability of A via

$$\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

This is particularly helpful because $\mathbb{P}(A_n)$ is easy to calculate. Indeed, under suitable assumptions, it is not hard to convince ourselves that

$$\mathbb{P}(A_n) = (\frac{5}{6})^{n-1} \times \frac{1}{6}.$$

As a result, we will have

$$\mathbb{P}(A) = \sum_{n=1}^{\infty} (\frac{5}{6})^{n-1} \times \frac{1}{6} = 1.$$

This example tells us that, an event with probability one may not necessarily be the certain event Ω (Surprise?)! In fact, in this example, since we are throwing the die repeatedly, a generic outcome should be given by a sequence $(x_1, x_2, \cdots, x_n, \cdots)$ where the n-th component x_n represents the result on the n-th toss. Therefore, the sample space is given by

$$\Omega = \{\omega = (x_1, x_2, \cdots) : x_n = 1, 2, \cdots, 6 \text{ for each } n\}.$$

It is clear that, any outcome $\omega = (x_1, x_2, \cdots)$ not containing a Six in its components is not an element of A. As a result, we see that $A \neq \Omega$.

To summarise, *Kolmogorov's three axioms of probability* are stated as follows. Let Ω be a given sample space. A *probability function* (also commonly known as a probability measure) $\mathbb{P}(\cdot)$ is a real valued function defined on a given class of events, such that:

Axiom 1 [Positivity]. $\mathbb{P}(A) \geqslant 0$ for every event A.
Axiom 2 [Unitarity]. $\mathbb{P}(\Omega) = 1$.
Axiom 3 [Countable Additivity]. For any sequence of mutually disjoint events $A_1, A_2, \cdots$, we have

$$\mathbb{P}\big(\cup_{n=1}^{\infty} A_n\big) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Remark 2.1. For deep theoretical reasons, it is often not reasonable to consider every subset of Ω as a legal event, and one needs to be very careful about specifying a given class of events on which probability functions are considered. This point is beyond the scope of the current subject, and we do not need to worry about it at this level.

It is important to point out that, any function satisfying Kolmogorov's three axioms is considered as a legal probability function. The axioms do not provide a concrete way of *constructing* a probability function. It is a main task for probabilists to construct suitable probability functions *subject to the axioms*, that correspond well to the real world problems that are under consideration.

Some basic properties of a probability function

One nice thing about the axiomatic approach, which is also the nature of all major mathematical areas, is that all the interesting properties of a probability function will follow from the three axioms by *logical deduction*. As a result, our understanding is no longer based on empirical evidence but on precise mathematical analysis.

We list several basic properties of a probability function. It is important to note that, these are properties *deduced* from Kolmogorov's three axioms, rather than another set of presumed hypotheses/axioms.

Property 1. $\mathbb{P}(\emptyset)=0$.

Proof. We can trivially write

$$\emptyset=\emptyset\cup\emptyset\cup\emptyset\cup\cdots.$$

Note that the sequence of events $\emptyset,\emptyset,\cdots$ are mutually disjoint. Therefore, by using Axiom 3, we have

$$\mathbb{P}(\emptyset)=\mathbb{P}(\emptyset\cup\emptyset\cup\emptyset\cup\cdots)=\mathbb{P}(\emptyset)+\mathbb{P}(\emptyset)+\mathbb{P}(\emptyset)+\cdots. \tag{2.1}$$

Since $\mathbb{P}(\emptyset)$ is a real number, the only possibility to make the above identity valid is that $\mathbb{P}(\emptyset)=0$. Indeed, if $\mathbb{P}(\emptyset)\neq 0$, the right hand side of (2.1) is infinite while the left hand side is finite, which is absurd.

Property 2 [Finite additivity]. Let $A_1,\cdots,A_n$ be a collection of mutually disjoint events. Then

$$\mathbb{P}(A_1\cup\cdots\cup A_n)=\mathbb{P}(A_1)+\cdots+\mathbb{P}(A_n).$$

Proof. We can form a sequence of mutually disjoint events by keeping adding $\emptyset$:

$$A_1,A_2,\cdots,A_n,\emptyset,\emptyset,\emptyset,\cdots.$$

This does not change the union of the events. According to Axiom 3 and Property 1 we just established, we have

$$\begin{aligned}
\mathbb{P}(A_1 \cup \cdots \cup A_n) &= \mathbb{P}(A_1 \cup \cdots \cup A_n \cup \emptyset \cup \emptyset \cup \cdots) \\
&= \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n) + \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) + \cdots \\
&= \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n) + 0 + 0 + \cdots \\
&= \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n).
\end{aligned}$$

More properties will be discussed in the next lecture.