Sets, Relations and Probability. Part IA Formal Methods.

Lecture V, *Intro to Probability*, 1st March.

Christopher J. Masterman (cm789@cam.ac.uk, christophermasterman.com)

We're now going to leave an explicit discussion of set theory and relations behind and, for Lectures V–VIII, look at probability. (The set-theoretic tools we discussed are still going to come in very handy, though!)

1. Outcomes Spaces and Fields

1.1. Probability crops up everywhere in philosophy: in understanding confirmation in the philosophy of science, degrees of belief in decision theory, causation in metaphysics, and so on. There's also a whole host of philosophical issues that concern the nature of probability itself. (Here, we will largely work around *those* issues.) Outside of philosophy, most of our reasoning concerns matters about which we are uncertain. So, for a variety of reasons, it looks like we should make sure we have a sharp understanding of probability.

1.2. At first glance, we assign probabilities to a variety of different types of thing: events, e.g., Labour winning the next election; claims or sentences, e.g., 'Labour wins in 2024' is likely true; and propositions, e.g., it is highly probable *that* Labour wins the next general election. There's a debate about what sort of thing we should apply probabilities to. Here, we'll take one side: events, understood as sets of outcomes.

1.3. We define probability relative to a certain set of total possible outcomes. To start, then, we define the so-called *outcome space*—sometimes called the *reference set* or *sample space*.

> **THE OUTCOME SPACE**: The *outcome space* is the set of all possible outcomes. Call this V.

What is in the outcome space? This will depend on what we are interested in calculating the probability of. For instance, if we wanted to work out the probability of a die landing a certain number, the relevant outcome space could be described as $\{1, 2, 3, 4, 5, 6\}$, where 1 is the outcome of the die landing on a 1 and so forth.

1.4. But we are not just interested in the probability of the die landing 1, we are also interested in calculating the probability that the die lands on *an odd number*, or *an even number*. That is, we are interested in the probability of certain *sets* of outcomes, what we will call *events*. We can represent the *event* that the die lands on an *odd* number as the set $\{1, 3, 5\}$, i.e., the set of outcomes from the outcome space where the die lands odd. In general, for an outcome space V, the set of possible *events* is called the *field*.

> **FIELD OF V**: For any outcome space V, the set of all events F_V is the *field* of V, where $F_V = \mathcal{P}(V)$.

1. A random number generator gives either 1, 2, or 3. Let the outcome space $V = \{1, 2, 3\}$. Field F_V:

   $$F_V = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}, \varnothing\}$$

   Here, the singletons, e.g., $\{1\}, \{2\}$, and $\{3\}$, represent the event of the generator giving 1, 2, or 3, respectively. $\{1, 2\}$ can be thought of as the event of it *either* giving a 1 *or* a 2.

Question: What do $\{1, 2, 3\}$ and $\varnothing$ represent in F_V above? Here, $\{1, 2, 3\}$ represents the *certain* event, e.g., the event that the generator gives *either* a 1, a 2, or a 3 would be represented by this. Here, $\varnothing$ represents the *impossible* event, e.g., the event of it giving both a 1 and a 2 would be represented by this. Generally, for any outcome space V, the field F_V will have in it both the certain event (V) and the impossible event ($\varnothing$).

1.5. One final point about fields: fields are, what we call, *closed under intersection, union and complement*. All *this* means is that if $X \in F_V$ and $Y \in F_V$, then $X \cap Y \in F_V$ and $X \cup Y \in F_V$, and if $X \in F_V$, then $V - X \in F_V$. In what follows, we'll call $V - X$ simply $\overline{X}$. Think of it loosely as the 'negation' of the event X.

2. Kolmogorov Axioms

2.1. Enough about outcome spaces and fields. Now, let's look at how we define probability—or, more precisely, a *probability function*. A probability function is a function which takes any event in F_V and assigns it a number (the probability of that event). A probability function is not, however, just any old such function.

2.2. In fact, we already have an inkling of what numbers a probability function should be assigning some events. The *certain* event in F_V, i.e., V itself, is certain, so it should be a assigned a probability of 1. And the *impossible* event in V_F, i.e., $\varnothing$, is impossible, so it should get a probability of 0. Saying a bit more, a *probability* function Pr on a field F_V has to satisfy the following axioms, for any $X, Y \in F_V$

(Axiom 1) $Pr(V) = 1$

(Axiom 2) $Pr(X) \geq 0$

(Axiom 3) If $X \cap Y = \varnothing$, then $Pr(X \cup Y) = Pr(X) + Pr(Y)$

(1)–(3) are called the **Kolmogorov Axioms**. Though a minimal characterisation of a probability function on F_v, we can use these to establish some perhaps familiar facts about probability.

THEOREM 1. *For any F_V, probability function Pr on F_v, and $X \in F_V$: $Pr(X) + Pr(\overline{X}) = 1$.*

Proof. Since $\overline{X} = (V - X)$, $Pr(X) + Pr(\overline{X}) = Pr(X) + Pr(V - X)$. Since $X \cap (V - X) = \varnothing$, from Axiom 3 and then Axiom 1, it follows that $Pr(X) + Pr(V - X) = Pr(X \cup (V - X)) = Pr(V) = 1$. So, $Pr(X) + Pr(\overline{X}) = 1$. □

THEOREM 2. *For any F_V, probability function Pr on F_V: $Pr(\varnothing) = 0$.*

Proof. From Theorem 1, $Pr(\varnothing) + Pr(\overline{\varnothing}) = 1$. By Axiom 1, since $\overline{\varnothing} = V - \varnothing = V$, $Pr(\varnothing) = 0$. □

3. Assigning and Calculating Specific Probabilities.

3.1. The axioms tell us how probability functions must generally behave, but they tell us little about which probabilities we should assign to particular events. To do this, we first assign probabilities to the outcomes. How should we do this? Here, we assign probabilities to outcomes based on the *frequency* of such an event. This is often known as a *frequentist interpretation* of probability. Now, there are *philosophical* issues with this interpretation—how do we assign probabilities to outcomes which have never happened? why should we think that the frequency so far observed will continue? But for our purposes, it suits.

3.2. Applying this to a specific case, take a fair six-sided die. What probability should we assign to the event that it lands on a six? Well, we assign that outcome $1/6$ because it has come up about one-sixth of the time, or similar dice have come up one-sixth of the time. (No surprise there, it is a *fair* die.) More precisely, if we let our set of outcomes $V = \{1, 2, 3, 4, 5, 6\}$, then our probability function Pr defined on F_V will be such that $Pr(\{6\}) = 1/6$. Likewise, for the event of it landing one, of it landing two, of it landing three, etc.:

$$Pr(\{1\}) = Pr(\{2\}) = Pr(\{3\}) = Pr(\{4\}) = Pr(\{5\}) = Pr(\{6\}) = \frac{1}{6}.$$

3.3. We can now calculate the probability of other events, based on the probability we have assigned the outcomes. For instance, how would we calculate the probability that a fair six-sided die lands on either a one or a three? Well, the relevant subset of our total set of outcomes V is $\{1, 3\}$. After all, this set represents the event that *either* the die lands on one *or* on three. Now, since $\{1\} \cap \{3\} = \varnothing$, from Axiom 3:

$$Pr(\{1,3\}) = Pr(\{1\} \cup \{3\}) = Pr(\{1\}) + Pr(\{3\}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

What about the probability of other events, e.g., the probability that the die lands on an odd number? In this case, the relevant subset of our total set of outcomes V is $\{1, 3, 5\}$—the set representing the event that *either* the die lands on one, or on three, or on five. Again, since $\{1\} \cap \{3\} \cap \{5\} = \varnothing$, from Axiom 3:

$$Pr(\{1,3,5\}) = Pr(\{1\} \cup \{3\} \cup \{5\}) = (Pr(\{1\}) + Pr(\{3\}) + Pr(\{5\})) = \left(\frac{1}{6} + \frac{1}{6} + \frac{1}{6}\right) = \frac{1}{2}$$

4. Conditional Probability

4.1. We define the conditional probability of one event X given an event Y, which we write as $Pr(X \mid Y)$, as:

CONDITIONAL PROBABILITY: $Pr(X|Y) = \dfrac{Pr(X \cap Y)}{Pr(Y)}$, where $Pr(Y) > 0$.

Conditional probability is useful. You may want to just calculate the probability that someone has brown eyes. You may want to just calculate the probability that someone has brown hair. However, you may also want to calculate the probability that someone has brown eyes, given that they have brown hair.

4.2. For instance, given a set of total outcomes from last time $V = \{1, 2, 3, 4, 5, 6\}$, let's calculate the probability that the die lands on two, *given* that it lands even. The relevant subset of V for the event that the die lands on two is $\{2\}$ and the relevant subset of V for the event that the die lands on an even number is $\{2, 4, 6\}$. So:

$$Pr(\{2\} \mid \{2,4,6\}) = \frac{Pr(\{2\} \cap \{2,4,6\})}{Pr(\{2,4,6\})} = \frac{Pr(\{2\})}{Pr(\{2,4,6\})}$$

Now, $Pr(\{2\}) = \dfrac{1}{6}$. Given that $\{2\} \cap \{4\} \cap \{6\} = \varnothing$, from Axiom 3 it follows that:

$$Pr(\{2,4,6\}) = Pr(\{2\} \cup \{4\} \cup \{6\}) = (Pr(\{2\}) + Pr(\{4\}) + Pr(\{6\})) = \left(\frac{1}{6} + \frac{1}{6} + \frac{1}{6}\right) = \frac{1}{2}$$

Thus: $\dfrac{Pr(\{2\})}{Pr(\{2,4,6\})} = \dfrac{1/6}{1/2} = \dfrac{1}{3}$. So, the probability of the die landing on two, given that it landed even is $1/3$.

4.3. We use the notion of a conditional probability to define when two events are **probabilistically independent**. For any two $X, Y \in F_V$, X and Y are **independent** if and only if $Pr(X \mid Y) = Pr(X)$. We say that any two events which are not independent are **dependent**. For example, the event of rolling an even number and the event of rolling a two are dependent events, since the probability of rolling a two given that you rolled an even number is not the same as the probability of rolling a two, as shown above. However, the event of rolling a two and the event of rolling *any* number (either 1, 2, 3, 4, 5, or 6, i.e., the event V) are independent:

$$Pr(\{2\} \mid V) = \frac{Pr(\{2\} \cap V)}{Pr(V)} = \frac{Pr(\{2\})}{1} = Pr(\{2\}).$$