SETS, RELATIONS AND PROBABILITY LECTURE 5

1. We can now move on to **probability**. There are two important things that you will eventually want to learn about: the probability calculus, and the interpretation of that calculus, and this course is mostly about the first. The nature of this distinction will become clearer when you have tried to use the calculus for yourself.

2. What are the things to which we assign probability? If you say that the probability that the Conservatives win the next election is 40%, are you saying it of an event, a proposition, a sentence, or a set? Here we shall say that it is sets of outcomes. In some ways it is natural to regard a Conservative victory as a set of outcomes. For it includes the outcome that they win by one seat, the outcome that they win by 100 seats etc., and these are different. It isn't fixed a priori *how* finely one should divide possibilities into outcomes but is usually settled by the context of the investigation, as we shall see.

3. We begin by defining the universe of possibilities that we wish to consider. This is a set V of **outcomes**, and V is called the **reference set** or **sample space**. For example, if we are throwing a die, there are 6 possible outcomes, and V = {1, 2, 3, 4, 5, 6}.

4. But there are many more than 6 possible *sets* of outcomes or **events**. We might want to know the probability of an even number, for example, or of: a 1 or a 6. We may write these situations as {2, 4, 6} and {1, 6} respectively. The **field** F contains all such sets: F = $\wp$(V). Note in particular that it contains the certain event V = {1, 2, 3, 4, 5, 6} and the impossible event ∅. Note also that if X ∈ F and Y ∈ F then X ∩ Y ∈ F and X U Y ∈ F. For any X ∈ F I'll write X* or ¬X for V - X i.e. the set of all outcomes other than those that are members of X.

5. Now *probability*, or better a *probability function*, is a function Pr that assigns numbers to elements of F and which obeys the following axioms for any X, Y ∈ F:

   (i) Pr (V) = 1
   (ii) Pr (X) ≥ 0
   (iii) If X ∩ Y = ∅ then Pr (X ∪ Y) = Pr (X) + Pr (Y)

   Axioms (i)-(iii) are known as the **Kolmogorov axioms** (in fact (iii) is slightly simplified in ways that are unimportant for present purposes).

6. Given these axioms we can show that Pr (X) + Pr (¬X) = 1. For clearly X ∩ ¬X = ∅. Hence by (iii), Pr (X U ¬X) = Pr (V) = Pr (X) + Pr (¬X). But by (i), Pr (V) = 1. So Pr (X) + Pr (¬X) = 1. We can also use the axioms to show that Pr (∅) = 0: since ∅ ∩ V = ∅, Pr (∅ ∪ V) = Pr (∅) + Pr (V) by (iii); but ∅ ∪ V = V, so Pr (∅) = 0.

7. Probabilities may now be assigned to events as follows. First we assign probabilities to the outcomes themselves, or rather to their singletons; logic does not tell us what these are. You might think they could be settled a priori on the basis that if there is no reason to choose between two options then they are equally likely. Unfortunately that principle is inconsistent.

8. The simplest way to justify an assignment of probabilities to basic events is via frequencies. The reason we assign probability 1 in 6 to the die's coming up 6 is *not* the symmetry of the die but the fact that when it was tossed, or when dice like it have been tossed, they in fact came up 6 about one-sixth of the time. Of course even this procedure is in a way irrational, as Hume showed: experience no more makes the future likely than it makes it certain. But it is consistent. So if you're throwing a die, we say that the probability of a 1 is 1/6 i.e. Pr ({1}) = 1/6. By similar reasoning, if you have a pack of cards, the probability of drawing the King of Clubs is 1/52 i.e. Pr ({KC}) = 1/52.

9. We can now work out the probability of other sets of events. What is the probability that the card is the ace of hearts or an even spade? Let T be the subset of V consisting of those events. Then T = {2S, 4S, 6S, 8S, 10S, AH}. The probability we want is Pr (T). But we know by (iii) that Pr (T) = Pr ({2S}) + Pr ({4S}) + Pr ({6S}) + Pr ({8S}) + Pr ({10S}) + Pr ({AH}) = 6/52 = 3/26.

10. We next define **conditional probability** as follows:

    $\Pr(A \mid B) =_{def.} \Pr(A \cap B) / \Pr(B)$.

    '$\Pr(A \mid B)$' is pronounced 'the probability of A given B'. It is undefined when Pr (B) = 0. This has a natural statistical interpretation: if Pr (A) measures the proportion of A's within a population V then $\Pr(A \mid B)$ measures the proportion of A's *amongst the B's* in that population.

11. For instance: what is the probability that a die comes up 2, given that it comes up even? We start off by taking V = {1, 2, 3, 4, 5, 6}. The outcome that it is even is E = {2, 4, 6}. The outcome that it comes up 2 is T={2}. Then the axiom tells us that $\Pr(T \mid E) = \Pr(T \cap E) / \Pr(E)$. But $T \cap E = T$ since $T \subseteq E$. Hence $\Pr(T \cap E) = \Pr(T) = 1/6$. And Pr (E) = 1/2. So $\Pr(T \mid E) = (1/6)/(1/2) = 1/3$.

12. Finally, if $X, Y \in F$ then X and Y are called **probabilistically independent** for a given Pr if $\Pr(X \mid Y) = \Pr(X)$; otherwise they are **probabilistically dependent** (and both relations are symmetric). E.g. drawing a king first time and drawing a club first time are independent events. But drawing a king first time and drawing a picture card then are *dependent*. This corresponds to the intuitive distinction between events that do / don't 'make a difference' to another event; but statistical probabilistic dependence can obtain between events that stand in no *causal* relation (although they may have a common cause).