SETS, RELATIONS AND PROBABILITY LECTURE 5

1. The simplest way to justify an assignment of probabilities to the basic events is in terms of frequencies. The reason that we assign a probability of 1 in 6 to the die's coming up 6 is *not* the symmetry of the die but the fact that when it has been tossed, or when dice like it have been tossed, it has or they have in fact come up 6 about one-sixth of the time. Of course even this procedure is in a sense irrational, as Hume showed: experience no more makes the future likely than it makes it certain. But at least it does not lead to inconsistency. So if you are 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. $\Pr(\{KC\}) = 1/52$.

2. 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 $\Pr(T) = \Pr(\{2S\}) + \Pr(\{4S\}) + \Pr(\{6S\}) + \Pr(\{8S\}) + \Pr(\{10S\}) + \Pr(\{AH\}) = 6/52 = 3/26$.

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

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

   '$\Pr(A \mid B)$' 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.

4. 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 $\Pr(T \mid E) = (1/6)/(1/2) = 1/3$.

5. Finally, if $X, Y \in F$ then X and Y are said to be **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, and those that don't, 'make a difference' to another event; but statistical probabilistic dependence can obtain between events that stand in no *causal* relation.

6. We have used this formula to calculate conditional probabilities, but we can also use it the other way around. That is, if we are given $\Pr(A \mid B)$ and $\Pr(B)$, we can calculate $\Pr(A \cap B)$ by means of the formula:

   $$\Pr(A \cap B) = \Pr(A \mid B) \times \Pr(B)$$

7. For instance, suppose we know that: exactly 50% of the population is male and that 30% of males have blue eyes. Then what is the probability that a randomly drawn person is male *and* blue-eyed? Letting M say that the person is male and B say that he/she is blue-eyed, we are asking after Pr ($B \cap M$):

   (i) Pr (M) = 0.5
   (ii) Pr ($B \mid M$) = 0.3
   (iii) Pr ($B \cap M$) / Pr (M) = Pr ($B \mid M$)
   (iv) Pr ($B \cap M$) / 0.5 = 0.3
   (v) Pr ($B \cap M$) = 0.15

8. Suppose that we take probability to measure one's confidence in a proposition. Then the **Bayesian** claim is that after you learn some new bit of evidence E, your new confidence in any hypothesis H is your old Pr ($H \mid E$). For instance, in the preceding example, your confidence that a randomly drawn person has blue eyes, once you learn that he is male, should be 30%.

9. One crucial application of the Bayesian thesis is to cases where one knows Pr (H), Pr (E) and Pr ($E \mid H$) and then learns that E is true. How confident should one *then* be that H is true? The answer is Pr ($H \mid E$), and we can calculate this using the following rule:

   Bayes's Theorem (first version): Pr ($H \mid E$) = Pr ($E \mid H$) $\times$ Pr (H) / Pr (E)

   What this means is e.g. that if some hypothesis either predicts, or makes highly likely, some observation that was very unlikely in advance, then actually making that observation is very strong evidence for the hypothesis.

10. For instance, suppose that I hold some bizarre conspiracy theory H according to which the end of the world will occur on 1 March 2016, and one consequence of which is (E) that tomorrow there will be earthquakes all over the world. We might expect Pr (E) and Pr (H) both to be very small today, and Pr ($E \mid H$) = 1 (make sure you are clear why). But if we do observe E tomorrow, then Bayes's Theorem implies that Pr (H) should now rise enormously, in fact that it should increase by a factor 1/Pr (E).

11. Again, consider the Monty Hall puzzle. You (A) or one of two other prisoners (B, C) will be shot tomorrow. You ask the guard to tell you which of B and C will *not* be shot (or to mention one at random if neither will). The guard tells you truthfully that B will not be shot. How confident should you now be that you will not be shot? Letting H say that you will be shot, and E say that the guard mentioned B, we have Pr ($E \mid H$) = 0.5, Pr (H) = 1/3 and Pr (E) = 0.5. The formula therefore gives Pr ($H \mid E$) = 1/3: the information should make you twice as confident that C will be shot as that you will.