Probability: A Quick Introduction

Curtis Brown

February 11, 2002

1 Axioms

Everything we need to know about probability can be derived from three axioms and one definition, together with first-order logic. Let us begin by looking at the three axioms. (I am more or less following the presentation in Curd & Cover, *Philosophy of Science: The Central Issues*, pp. 627-632. However, there are some notational differences, and I provide proofs that Curd and Cover leave out.)

Axiom 1 $0 \leq P(A) \leq 1$

This first axiom simply declares that all probabilities are between zero and one, inclusive. Any real number between zero and one can be a probability. This axiom is in a sense arbitrary; we could allow probabilities to range between any numbers we choose (for example, between 0 and 100, as we do with percentages). But it is convenient to keep them between zero and 1.

Axiom 2 *If* $\models A$*, then* $P(A) = 1$.

The notation here needs a little explanation. Logicians write $\Gamma \models \phi$ to indicate that a sentence ϕ is a logical consequence of a set of sentences Γ. A special case is $\emptyset \models \phi$, where the set of sentences Γ has no members; this is also simply written $\models \phi$, with nothing on the left-hand-side of the '$\models$' symbol. This means that ϕ is a sentence that is logically valid, i.e. a sentence that must be true no matter what. So, where C&C write that the probability of a necessary truth is 1, we write that if $\models A$ then $P(A) = 1$, meaning that if A is logically necessary, then $P(A)$ is 1.

This seems only reasonable. If we have a sentence that simply *must* be true no matter what, then surely we must give it the highest possible probability, which by Axiom 1 is 1.

Axiom 3 (Special Addition Rule) *If* $\models \neg(A \wedge B)$*, then* $P(A \vee B) = P(A) + P(B)$.

Notation: '$\wedge$' is the symbol for conjunction, and can be read as 'and'; '$\vee$' is the symbol for disjunction, and can be read as 'or'. It should be noted that this is what is sometimes called *inclusive or*, meaning that a sentence '$A \vee B$' is true if A is true, or B is true, or both. This third axiom perhaps seems a little less like a mere definition than the other two. But it is still pretty obvious. If it cannot happen that A and B both occur, i.e. that A and B are mutually exclusive, then the probability that one or the other will occur is just the probability of the first plus the probability of the second. For example, if you throw a die, it cannot come up both 1 and 5; so the probability of getting either a 1 or a 5 is $P(1) + P(5) = 1/6 + 1/6 = 1/3$.

2 Derived Rules

Those are all the axioms we need (though we will later add a definition of conditional probability that could be treated as an additional axiom). From these three axioms, we can derive a number of additional rules that are useful in calculating probabilities.

Derived Rule 1 (Negation Rule) $P(\neg A) = 1 - P(A)$

Notation: '$\neg$' is the symbol for negation, and can be read "it is not the case that" or simply "not". How can we prove the Negation Rule from the axioms we already have? Here is one way: it is a logical necessity that either A or $\neg A$, i.e. $\models (A \vee \neg A)$. So, by Axiom 2, $P(A \vee \neg A) = 1$. By Axiom 3, $P(A \vee \neg A) = P(A) + P(\neg A)$. So we know that

$$P(A) + P(\neg A) = 1$$

Subtracting $P(\neg A)$ from both sides gives us the Negation Rule.

Derived Rule 2 (Equivalence Rule) If $A \models B$ *and* $B \models A$*, then* $P(A) = P(B)$.

The Equivalence Rule could be seen as a special case of the following rule, the Implication Rule. But it turns out that it is useful to prove the Equivalence Rule first, and then use this in the derivation of the Implication Rule.

Here is one way to derive the Equivalence Rule. Suppose that A and B are logically equivalent, that is, that $A \models B$ and $B \models A$. We want to show that $P(A) = P(B)$.

Since $A \models B$, we know that $\models (\neg A \vee B)$. (A must be either true or false. If A is true, then B must be true, and in that case the disjunction is true. On the other hand, if A is false, the $\neg A$ is true, making the first disjunct true, and hence making the whole disjunction also true. So either way, the disjunction is true.)

So, by Axiom 2, $P(\neg A \vee B) = 1$. However, since A and B are equivalent, the two disjuncts are mutually exclusive. So, by the Special Addition Rule, $P(\neg A \vee B) = P(A) + P(B)$. So we have

$$P(\neg A) + P(B) = 1$$

and therefore

$$P(B) = 1 - P(\neg A) = 1 - (1 - P(A)) = P(A)$$

— establishing the Equivalence Rule.

Derived Rule 3 (Implication Rule) *If* $A \models B$*, then* $P(B) \geq P(A)$*.*

This one is a little tricky, in my view. Suppose that $A \models B$. Notice, to begin with, that B is equivalent to $(B \wedge \neg A) \vee (B \wedge A)$. ($A$ must be either true or false, so to say B is true is to say that either it is true and A is false, or it is true and A is true.)

However, since $A \models B$, $B \wedge A$ is equivalent to A: if A is true, then B must also be true. So $(B \wedge \neg A) \vee (B \wedge A)$ is equivalent to $(B \wedge \neg A) \vee A$, which therefore is also equivalent to B.

Since B is equivalent to $(B \wedge \neg A) \vee A$, we have, by the Equivalence Rule,

$$P((B \wedge \neg A) \vee A) = P(B)$$

But the two disjuncts of the left-hand side above are mutually exclusive, so the Special Addition rule applies, giving us

$$P(B \wedge \neg A) + P(A) = P(B)$$

By Axiom 1, $0 \leq P(B \wedge \neg A) \leq 1$, so we have $P(B) \leq P(A)$, as desired.

Derived Rule 4 (General Addition Rule) $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

We first note that $(A \vee B)$ is logically equivalent to $(A \wedge B) \vee (A \wedge \neg B) \vee (\neg A \wedge B)$. (One way to see this is to see that there are exactly three ways for $A \vee B$ to be true: if A and B are both true; if A is true and B is false (so that $\neg B$ is true); and if A is false (so $\neg A$ is true) and B is true. By the Equivalence Rule, then, $P(A \vee B) = P((A \wedge B) \vee (A \wedge \neg B) \vee (\neg A \wedge B))$. Since the three disjuncts are all mutually exclusive, we have (by the Special Addition Rule)

$$P(A \vee B) = P(A \wedge B) + P(A \wedge \neg B) + P(\neg A \wedge B)$$

Keep that fact in mind while we pursue two further facts. First, A is logically equivalent to $(A \wedge B) \vee (A \wedge \neg B)$. So, by the Equivalence Rule together with the Special Addition Rule,

$$P(A) = P(A \wedge B) + P(A \wedge \neg B)$$

Similarly, B is logically equivalent to $(A \wedge B) \vee (\neg A \wedge B)$. So

$$P(B) = P(A \wedge B) + P(\neg A \wedge B)$$

Adding together the last two displayed equations, we get

$$P(A) + P(B) = P(A \wedge B) + P(A \wedge B) + P(A \wedge \neg B) + P(\neg A \wedge B)$$

Subtracting $P(A \wedge B)$ from both sides gives us

$$P(A) + P(B) - P(A \wedge B) = P(A \wedge B) + P(A \wedge \neg B) + P(\neg A \wedge B)$$

Now, the right-hand side of the above equation is the same as the right-hand side of the first displayed equation. So the left-hand sides are also equal, giving us

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

as desired!

3 Conditional Probability

Now we introduce a definition of conditional probability. We represent the conditional probability that A is true, given that B, with the notion $P(A|B)$ (read "probability of A given B").

Definition 1 $P(A|B) = P(A \wedge B)/P(B)$

We are treating this as simply a definition. But it is worth seeing why it is a *reasonable* definition. Curd & Cover give a nice graphical illustration. But many other simple examples could be given. For instance, suppose my friend draws a card from a well-shuffled deck, and tells me that he has drawn a queen. I want to know the probability that he has drawn the Queen of Clubs, *given that* he has drawn a queen. Intuitively, it seems that this should be 1/4, since there are four queens in the deck, and only one of them is the Queen of Clubs. The definition gives us this intuitive result.

Intuitively, $P(A \wedge B)$, that is, of getting the Queen of Clubs *and* getting a queen, is exactly the same as $P(A)$, getting the Queen of Clubs, namely 1/52. The probability of getting a queen, $P(B)$, is $4/52 = 1/13$. So $P(A \wedge B)/P(B)$ is 1/52 divided by 1/13, which is $13/52 = 1/4$, as desired.

Of course, we could also write the above definition $P(A \wedge B) = P(A|B) \times P(B)$, and it is useful in this form as well.

4 Bayes' Theorem

Bayes' Theorem has achieved notoriety in a number of different disciplines, and is associated with a number of controversial issues. However, the theorem itself is simply a straightforward theorem of probability theory, one that follows almost immediately from the material developed above.

The definition of conditional probability tells us that $P(A|B) = P(A \wedge B)/P(B)$. By the Equivalence Rule, $P(A \wedge B) = P(B \wedge A)$. Now, $P(B \wedge A) = P(B|A) \times P(A)$ (simply interchanging A's and B's on both sides of an earlier equation). Substituting this in, we have $P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$. Usually, however, instead of A and B, Bayes' Theorem is presented with T and E (for "theory" and "evidence"), giving us:

$$P(T|E) = \frac{P(E|T) \times P(T)}{P(E)}$$

This is a simple form of Bayes' Theorem. We can complicate it by noting that E is equivalent to $(E \wedge T) \vee (E \wedge \neg T)$. Since the two disjuncts are mutually exclusive, we have $P(E) = P(E \wedge T) + P(E \wedge \neg T)$. If we expand both of the right-hand probabilities, we get the usual version of Bayes' Theorem:

$$P(T|E) = \frac{P(E|T) \times P(T)}{P(E|T) \times P(T) + P(E|\neg T) \times P(\neg T)}$$

Actually, we can generalize a bit further. In the above equation, we considered two mutually exclusive and exhaustive ways that E could be true, namely $E \wedge T$ and $E \wedge \neg T$. Using T and $\neg T$ is a very simple way to obtain mutually exclusive and exhaustive possibilities, but not the only way. Suppose we have a larger set of possibilities, $T_1, T_2, \ldots, T_n$. Suppose further that the T_i are mutually exclusive, so that no more than one of them can be true, and that they are exhaustive, so that at least one of them must be true. Then E is equivalent to $(E \wedge T_1) \vee (E \wedge T_2) \vee \ldots \vee (E \wedge T_n)$. Since the disjuncts here are mutually exclusive, we can use the Special Addition rule to calculate their probability. So we get the following generalized version of Bayes' Theorem, where we are calculating the probability of a particular theory T_k out of a collection of alternatives $T_1, T_2, \ldots, T_n$:

$$P(T_k|E) = \frac{P(E|T_k) \times P(T_k)}{\sum_{i=0}^{n} P(E|T_i) \times P(T_i)}$$

This generalized version can be useful if we are evaluating a theory in relation to a number of competitors. (The competitors must be mutually exclusive; if they are, we can guarantee that they are also exhaustive by making T_n simply "none of the above.") However, for our purposes, we will be concerned only with the simple case in which we are considering only two alternatives, T and $\neg T$.