Some Basics on Logic and Probability

Richard Johns, September 2006.

1. Logical Relations

Logic is concerned especially with certain relations between propositions, known as logical relations, the most important of which are *consistency* and *entailment*. These are hard to define with precision, but the basic idea can be conveyed by the following.

Definitions (i) Propositions A and B are consistent iff it is possible for them both to be true (together).

(ii) A entails B just in case, necessarily, if A is true then so is B.

Some examples might help. For someone knowledgeable about geography, the proposition that Alice was born in Coquitlam is consistent with her having been born in B.C., but inconsistent with the claim that she was born in Idaho. The claim that Alice was born in Coquitlam is, however, logically consistent with the proposition that she *grew up* in Idaho. Moreover, we note that if Alice was born in Coquitlam, then she *must* have been born in B.C., so that her being born in Coquitlam *entails* that she was born in B.C. It does not entail, of course, that she grew up in Idaho, as she might have grown up somewhere else, such as in Lethbridge.

Logical relations impose constraints on what it is proper to believe. If A is inconsistent with B, for example, then it is not proper to believe both of them together. Since they cannot both be true, and beliefs are supposed to be true, one should not believe them both. It is foolish to believe that Alice was born in Coquitlam and also in Idaho. If A entails B, then it is legitimate to proceed from belief that A holds to belief that B is also true. Thus we sometimes describe the fact that A entails B by saying that *B follows from A*, or that B may be deduced, or inferred, from A. For example, if one believes that $x + 2 = 6$, then one may proceed to the conclusion that $x = 4$.

Definition If A entails B, then we shall say that A is *stronger than* B.

Note that, since every proposition entails itself, this definition says that A is stonger than A. Don't worry about this. Just think of "stronger than" as "stronger than or equal to", just as we write "greater than" as "≥". If A is stronger than B, then A gives all the information that B gives, and possibly more information on top of that. Suppose you start off knowing that B is true. If you later learn that A is true, your previous knowledge that B becomes redundant, as it is "contained" in the knowledge that A. People sometimes say that the old knowledge that B is "screened off" by the new knowledge that A. For example, the knowledge that Alice was born in B.C. is screened off, or made redundant, by the knowledge that she was born in Coquitlam.

If A is stronger than B, then we shall say that B is weaker than A. Note that, for most pairs of propositions, neither entails the other, and so neither is stronger (or weaker) than the other.

During this section of the course it will be convenient to use some logical symbols. We use upper-case, italic, Latin letters A, B, C, etc. to stand for propositions. We then have the following abbreviations:

A entails B	$A \Rightarrow B$
It is not the case that A	$\neg A$
Either A or B is the case (or both)	$A \vee B$
Both A and B are the case	$A \& B$ or $A \wedge B$

If A entails B and B entails A as well, i.e. $A \Rightarrow B$ and $B \Rightarrow A$, then we shall say that A and B are *equivalent*. This will be written as $A \Leftrightarrow B$.

For convenience, we define O as the "empty" proposition that says nothing, so that $A \wedge O \Leftrightarrow A$, and $A \vee O \Leftrightarrow O$, for any A. We can then define a logical truth, or logical necessity, as a proposition A such that $O \Rightarrow A$.

2. States of Knowledge

The notion of a state of knowledge, or *epistemic state*, is very useful in understanding probability. The reason for this is that the probability of an event or proposition depends on what knowledge you have. If I learn that the Canucks' opening game is against Detroit, then the probability that they will win is reduced (for me). If I learn that they are playing the Calgary Flames, then a win is more likely.

An epistemic state captures everything I know at a particular instant of time. It includes not only the things (propositions) that I know for sure, but also things that I have a pretty good idea about. For this reason, it is perhaps more accurately described as a state of *rational belief* rather than of knowledge. Consider, for instance, my knowledge of physics. Most of the current theories of physics, such a relativity, quantum mechanics and so forth, are uncertain, or fallible. Future experiments may show that they need to be revised. Nonetheless, our knowledge of physics is still valuable and important, even though it is uncertain.

One's epistemic state changes with time, as one learns more about the world. The simplest kind of change is where one learns (for sure) that some proposition *A* is true, which one was previously unsure about. This is a simple addition to, or expansion of, one's knowledge. Such an expansion occurs, for instance, when one performs an experiment for the first time. Prior to the experiment one does not know what the result will be, but after the experiment one does know this, so that's one's knowledge is increased.

We will need some more symbols to describe epistemic states. In general, an epistemic state will be given some symbol based on '*K*', for '*K*nowledge', e.g. *K*, *K*', and so on. Using the concept of an epistemic state, we can think of propositions as "information vectors" that carry you from one state to a better one, in which you have more knowledge. If one starts in the epistemic state *K*, and then learns that *A* is true, then one's new state will be written $K+A$. The idea of an epistemic state may seem strange at first, but it is very simple and should soon seem obvious.

Suppose that, in the epistemic state *K*, you know for sure that all reptiles lay eggs. The proposition *A* says that Clive is a reptile, and *B* says that Clive lays eggs. Then *A* does not entail *B*, as one cannot infer *B* from *A* alone. Within the epistemic state *K*, however, one *can* infer *B* from *A*, as one knows that all reptiles lay eggs. Thus, although *A* does not entail *B* absolutely, the entailment does hold *within K*. We can write this relation as $A \Rightarrow_K B$.

It is a similar story for the other logical relations of consistency and equivalence. Two statements can be consistent in themselves, and yet be inconsistent within an epistemic state *K*. For example, the statements "Clive is a reptile" and "Clive does not lay eggs" are consistent for a sufficiently ignorant person, but inconsistent within the state *K* above.

If a proposition *A* is known for certain in *K*, then of course adding *A* to *K* will make no difference at all. Thus the case where *A* is 'contained' in *K*, or absolutely certain in *K*, can be expressed as $K+A = K$.

3. Epistemic Probability

We know that scientists are almost never able to prove their theories conclusively. The best they can do is show that their theories are "probably" true, given all of the available evidence. What do we mean by "probably" here?

Well, the probability of a theory, in this sense, depends on the evidence available to you. If two scientists are working with different sets of data, then a single theory might be likely according to one, and unlikely according to the other. These probabilities are therefore *subjective*, in the sense that they can vary from one subject (person) to another. Sometimes these probabilities are called "subjective probabilities". Other people call them "knowledge probabilities", or "ignorance probabilities", as they depend on one's knowledge or ignorance. Philosophers call them *epistemic* probability, as "epistemic" means "pertaining to knowledge".

One's degree of belief in a proposition may vary with time, as one acquires knowledge. For example, at the start of a one-mile race Fred looks tired and sluggish. After the first lap he's in last place, and there's an expression of pain on his face. At this point, as I watch from the stands, my epistemic probability for him winning is quite low, perhaps 0.01. Then, half way through the race, he seems to be warming up. He looks relaxed, his stride is more fluid, and he's moved up to ninth place. The probability of Fred winning is now much higher, perhaps 0.2. At the bell, Fred is in third place, just behind the leaders, and looks as fresh as a daisy. I've seen Fred run like this before, and am now pretty sure he'll win. The probability is now 0.95. Unfortunately, just as he moves into first place, coming off the final bend, Fred trips over his laces and crashes to the ground. His probability of winning is now effectively 0.

In other words, a person's epistemic probability for a proposition depends on his epistemic state. If one's epistemic state expands, as one acquires new information, then one's epistemic probabilities can change. Moreover, one's epistemic probability should depend *only* on one's epistemic state – nothing else should make a difference. Owning shares in Philip Morris should not affect one's degree of belief that smoking is harmful, for instance. If two people are in the same epistemic state (which is rather unlikely) then they should have the same epistemic probabilities.

Our notation for epistemic probability is then as follows:

Definition $P_K(A)$ is the epistemic probability for *A* within the state *K*.

Epistemic probability is not the only kind of probability. It is now (almost) generally agreed that there are two basic kinds of probability, which (although they are related) need to be kept quite separate. Confusing these two kinds of probability is one of the most basic errors in this subject. The second kind is called *physical probability*, or *chance*. Fortunately the nature of physical probability doesn't concern us here. It's rather mysterious, as it has exactly the same structure as epistemic probability, but it seems to exist objectively, in the physical world, independently of us humans.

4. How is Probability Measured?

Length is (often) measured in metres, time is measured in seconds. What is the unit of probability? The unit of probability is a proposition that is absolutely certain, i.e. true beyond all doubt, such as '2 + 2 = 4'. Such a proposition has probability 1 by definition, just as the platinum metre rod in Paris had a length of 1 metre. Other propositions, that are less than certain, have lower probabilities, but how are actual values assigned?

Various methods have been used to assign probabilities, but my favourite is to use the values of gambles. Take a type of object that has value, say a loonie ($1). Let a *gamble*, [$1 if *A*], be a legal contract that will pay the owner $1 if a particular proposition *A* is true, but which is worth nothing if *A* is false.

If *A* is absolutely certain, as certain as '2+2=4', then intuitively the gamble [$1 if *A*] is worth exactly the same as $1 itself. If *A* is less than certain, on the other hand, [$1 if *A*] is worth less than $1 in the sense that you would prefer to have $1 over [$1 if *A*].

Consider, for example, that a fair coin is to be tossed once, and *heads* says that the coin will land heads. How much is [$1 if *heads*] worth? Intuitively it should be worth the same as [$1 if *tails*], since the coin is fair. But what is a fair price for buying *both* gambles together, i.e. the bundle {[$1 if *heads*], [$1 if *tails*]}? The fair price is surely $1, since the bundle will pay exactly $1 no matter how the coin lands – excluding *edge* as a possibility! Notice then that we have created two other objects, of equal value, which together are worth $1. Such objects can be defined as having value $0.5.

To cut a long story short, a value scale can be created in this way using any object of value, and once this is done the (epistemic) probability of a proposition can be defined as the value of a gamble involving that proposition.

Definition $P_K(A)$ = the fair price of [$1 if *A*], for someone in epistemic state *K*.

5. The Axioms of Probability

We saw in Section 1 that logic imposes some constraints on belief – if *A* and *B* are inconsistent, for example, then one should not believe both. In this section we shall find that logic also imposes constraints on *degrees* of belief, i.e. epistemic probabilities. These constraints are called the *axioms* of probability. There are four axioms, the first three of which are known as the *Kolmogorov* axioms, after the famous Russian mathematician A. N. Kolmogorov.

Axiom 1 $P_K(A) \geq 0$, for all epistemic states *K*.

Axiom 2 If $K{+}A = K$, then $P_K(A) = 1$.

Axiom 3 If *A* is inconsistent with *B* then $P_K(A \vee B) = P_K(A) + P_K(B)$.

Axiom 1 is clear from the fact that possessing the gamble [$1 if *A*] cannot cost you anything, and so canno be worth less than $0. Axiom 2 is just as straightforward. If *A* is known for sure within *K*, then having the contract [$1 if *A*] is no different from having $1. Axiom 3 is more tricky, but may be demonstrated as follows. If *A* is inconsistent with *B*, then the two contracts [$1 if *A*] and [$1 if *B*] cannot both yield money, as in that case both *A* and *B* would be true. It follows that having *both* contracts together is equivalent to having just the single contract [$1 if $A \vee B$]. (In each case, you get $1 if either *A* or *B* is true.) Now, the value of two separate contracts is the sum of the values of each one, so that Axiom 3 must hold.

The fourth axiom is the most interesting. Since it involves what is called *conditional probability*, we shall first define this.

Definition $P_K(A \mid B)$, which we read as "the probability of *A* *given* *B*", is $P_{K+B}(A)$.

In other words, the probability of *A* given *B* is the probability of *A* in an expanded epistemic state, in which (100% certain) knowledge of *B* has been added. Conditional probability enables us to measure the impact of new evidence upon the credibility of a scientific theory. If the old probability of a hypothesis H was $P_K(H)$, then upon receipt of new evidence E our epistemic state becomes K+E, and so the new probability of H is $P_{K+E}(H)$, i.e. $P_K(H \mid E)$.

The fourth axiom relates conditional probability to normal probabilities, as follows.

Axiom 4 If $P_K(B) \neq 0$, then $P_K(A \mid B) = \dfrac{P_K(A \,\&\, B)}{P_K(B)}$.

Axiom 4 is also known as the 'Principle of Conditionalisation'. Sometimes it is taken to be the *definition* of conditional probability, but this is a mistake. It requires demonstration, although the proof is tricky and is not included here.

It should be noted that probabilistic reasoning, like all reasoning, is not simply a matter of following rules. The axioms, though useful, do not by themselves get us very far. Reasoning is an art, and involves intuition, skill and good judgment. In particular, reasoning involves making judgments about symmetry and irrelevance, according to the following principles. It should be noted that these principles are not further axioms, since the notions of symmetry and irrelevance cannot be rigorously defined. They are guidelines. One learns to use them properly through practice.

Symmetry Principle If A and B are symmetric within K, then $P_K(A) = P_K(B)$.

Irrelevance Principle If A is irrelevant to B within K, i.e. A provides no information (relative to K) about whether or not B might be true, then $P_K(B \mid A) = P_K(B)$.

6. Examples

Suppose we know that a coin is lying on a table top, but don't know which face is showing. What is the probability that it is heads? Let K be our knowledge here. Since the coin is symmetric with respect to its two faces, heads and tails, we have that $P_K(\text{heads}) = P_K(\text{tails})$, by symmetry. Also, since heads and tails are inconsistent (they cannot both be true), Axiom 3 tells us that $P_K(\text{heads} \vee \text{tails}) = P_K(\text{heads}) + P_K(\text{tails})$. Finally, since we know (within K) that (heads $\vee$ tails) is true, Axiom 2 tells us that $P_K(\text{heads} \vee \text{tails}) = 1$. (We sometimes describe this by saying that heads and tails are *jointly exhaustive*,

within K. Together, they exhaust all the possibilities that K allows.) We then have:

$P_K(\text{heads}) + P_K(\text{tails}) = 1$

Thus, $P_K(\text{heads}) + P_K(\text{heads}) = 1$

So $P_K(\text{heads}) = 1/2$.

It might seem like a lot of work for such an obvious result! It is important to understand the logical basis of such facts, however. More generally, if there are n alternatives $A_1, A_2, ..., A_n$, that are pairwise inconsistent, pairwise symmetric, and jointly "exhaustive" (i.e. one of them has to be true), then each alternative has probability $1/n$, i.e. $P_K(A_i) = 1/n$, for every i.

Applying this to the case of a six-sided die resting on a table, we find that the probability of each face being the upper face is 1/6. If the faces are numbered 1 to 6, then what is the probability of getting an even number? Let the random variable **X** represent the score on the die. Using Axiom 3 we get:

$$\begin{aligned} P_K(\mathbf{X} \text{ is even}) &= P_K(\mathbf{X}{=}2 \vee \mathbf{X}{=}4 \vee \mathbf{X}{=}6) \\ &= P_K(\mathbf{X}{=}2) + P_K(\mathbf{X}{=}4) + P_K(\mathbf{X}{=}6) \\ &= 1/6 + 1/6 + 1/6 \\ &= 1/2. \end{aligned}$$

Suppose we find out that the score is even. Within this new epistemic state, what is the probability of **X**=2? Intuitively one feels that it should be 1/3, since there are 3 possible values that are even. Let us check this, using Axiom 4.

$$\begin{aligned} P_K(\mathbf{X} = 2 \mid \mathbf{X} \text{ is even}) &= \frac{P_K(\mathbf{X} = 2 \,\&\, \mathbf{X} \text{ is even})}{P_K(\mathbf{X} \text{ is even})} \\ &= \frac{P_K(\mathbf{X} = 2)}{P_K(\mathbf{X} \text{ is even})} \\ &= \frac{\frac{1}{6}}{\frac{1}{2}} = \tfrac{1}{3}. \end{aligned}$$

5. Useful Theorems

1. $P_K(\neg A) = 1 - P_K(A)$.

2. If $K + \neg A = K$ (i.e. A is surely false in K) then $P_K(A) = 0$.

3. If $A \Leftrightarrow_K B$, then $P_K(A) = P_K(B)$.

4. If $A \Rightarrow_K B$, then $P_K(A) \leq P_K(B)$. (Note that stronger statements are *less* probable!)

5. $0 \leq P_K(A) \leq 1$.

6. $P_K(A \,\&\, B) = P_K(A) \times P_K(B \mid A)$. (This is the *general product rule*.)

7. If A and B are mutually irrelevant within K, then $P_K(A \,\&\, B) = P_K(A) \times P_K(B)$. (This is the *product rule for independent propositions*.)

8. $P_K(A \vee B) = P_K(A) + P_K(B) - P_K(A \,\&\, B)$

(This rule *always* holds, regardless of whether or not A and B are inconsistent within K.)

9. If $P_K(A) = 1$, then $P_K(\mathrm{A} \mid \mathrm{B}) = 1$ as well, provided that $P_K(B)$ exists and is > 0.

10. Bayes's theorem: $$P_K(H \mid E) = \frac{P_K(E \mid H) P_K(H)}{P_K(E)}.$$

11. The theorems of total probability:

Let $O \Rightarrow_K (A_1 \vee A_2 \vee ... \vee A_n)$, so that the propositions $A_1, A_2, ..., A_n$ are jointly exhaustive in K, and also let $A_i \Rightarrow_K \neg A_j$ for $i \neq j$, so that the A_i are pairwise inconsistent in K, then

1. $P_K(B) = P_K(B \,\&\, A_1) + P_K(B \,\&\, A_2) + ... + P_K(B \,\&\, A_n)$.

2. $P_K(B) = P_K(B \mid A_1)P_K(A_1) + P_K(B \mid A_2)P_K(A_2) + ... + P_K(B \mid A_n)P_K(A_n)$.