Bayes' Theorem and the Philosophy of Science

Curtis Brown

February 14, 2002

1 Bayes's Theorem

My previous handout, introducing probability theory, concluded with the following version of Bayes's Theorem:

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

Some quick terminology: $P(T|E)$ is known as the *posterior probability* of T. The idea is that $P(T|E)$ represents the probability assigned to T *after* taking into account a new piece of evidence, E. $P(T)$ is known as the *prior probability* of T, the probability it is assigned prior to taking E into account. $P(E|T)$ is known as the *likelihood* of T (for better or worse — probably worse!). With these preliminaries in hand, it is time to see why Bayes's theorem is of philosophical interest.

2 Objective and Subjective Probabilities

The terminology just listed provides some insight into why Bayes's theorem has been viewed as important for the philosophy of science. The idea is that it can give us insight into the relationship between a new piece of evidence, E, and our acceptance of a theoretical hypothesis T to which E is relevant.

Viewed simply as an assertion about objective probability, Bayes's Theorem is completely unobjectionable. It follows very quickly from the axioms of probability theory together with the definition of conditional probability, so there is simply no avoiding it. However, its application to the relationship between theory and evidence requires a larger step than may be immediately obvious. As long as we have objective probabilities to work with, Bayes's Theorem expresses a necessary relationship between them. However, it is not at all clear that we can calculate anything like an objective probability, in this sense, for a scientific theory. Probability calculations make the most obvious sense when we are dealing with a clearly delimited set of possible outcomes of a particular event, such as throwing a die or drawing one or more cards from a shuffled deck. Nothing much like this seems to be available when we are considering how likely it is that a scientific theory is true.

The large leap that "Bayesians" in the philosophy of science make, then, involves two main assumptions: first, that we can distinguish between different *degrees of belief* or *subjective probabilities*, and second, that, for a rational person, these subjective probabilities should conform to the axioms of probability theory.

It does seem clear at least that we have different degrees of belief or confidence in various propositions. We are absolutely certain of some things, for instance that either grass is green or grass is not green, while we believe other things with less confidence, for instance that a friend repaid the five bucks he borrowed several weeks ago. It is less clear that we can actually assign a numerical value to the degree to which we believe a certain proposition.

The Bayesian holds that, at least for a perfectly rational person, we can determine subjective probabilities for particular propositions by determining the odds the person would be willing to accept in a bet on that proposition. If you offer me 4-1 odds that Bush will not be reelected, and I accept, then I must think that there is at least a 20

The assumption of perfect rationality may be very unrealistic, however! Consider the following admission by William Safire, in a February 7, 2002 column in the *New York Times*:

> In a column last month I offered my "morning line" on the potential Democratic presidential candidates (Tom Daschle, 4 to 1; Joe Biden, 5 to 1; Al Gore, 2 to 1; and seven others at various odds).
>
> "What Safire doesn't seem to realize," wrote Dan Seligman in Forbes magazine, "is that odds translate into percentage probabilities (e.g., 4-1 means the guy has a 20% chance) and that his probabilities add up to 168%. Alas, mutually exclusive contingencies cannot have probabilities adding up to more than 100%."
>
> That's bad news for Gore, whose chances I have just dropped to 3 to 1, and to Joe Lieberman, now a 12-to-1 long shot. Scratch Biden. O.K. down at the starting gate? The feeling is pari-mutuel.

Safire's admission points up the reason for thinking that subjective probabilities should conform to the axioms of probability theory. Suppose he had offered the following response to Seligman: "Of course it's true that the *objective probabilities* of mutually exclusive possibilities cannot add up to more than 100%. But we can't determine objective probabilities of candidates being nominated. I was merely offering my *subjective probabilities* or degrees of belief, and, since the axioms of probability theory don't necessarily hold of degrees of belief, there's no reason my belief in mutually exclusive outcomes can't add up to more than 100%." What would be wrong with this response? The problem is that, if he were then willing to back up his assessments with dollars, someone who accepted bets on all the candidates at the odds Safire gave would be in a situation in which he would win money from Safire no matter who was nominated! (A collection of bets is "book," and a collection of bets which guarantees that you will lose money no

matter what happens is called a "Dutch book." The argument that subjective probabilities should conform to the laws of probability theory because otherwise one would always be susceptible to bets they were guaranteed to lose is called the "Dutch book argument.")

3 Calculating Posterior Probabilities

The Bayesian, then, is interested in using Bayes's Theorem to help determine how a rational person would modify their degree of belief in a theoretical hypothesis in light of a new piece of evidence. If we are to apply the theorem to calculate $P(T|E)$, we need to already have values for several other subjective probabilities.

3.1 Likelihood and Prior Probability

The numerator of the right-hand side of the theorem is $P(E|T) \times P(T)$. The "likelihood" $P(E|T)$ will usually be fairly straightforward. It is the probability of the new piece of evidence if the theory is true. In many cases the theory (together with presupposed background information about initial conditions and the like) will deductively imply that E will occur. (Consider the flagpole and the shadow: we hypothesize that the flagpole is a certain height, and then check the length of the shadow. If we presuppose that the sun is at a particular angle, and that light travels in straight lines, then our theoretical hypothesis deductively implies that the shadow will be a particular length. The probability $P(E|T)$ that the shadow will be the predicted length if the flagpole is the hypothesized height is 1.) In other cases, when the hypothesis involves statistical generalizations, we should still be able to give a precise value less than 1 to $P(E|T)$.

In a sense $P(T)$, the prior probability of T, is also unproblematic. It simply represents the degree to which one was convinced of T prior to observing E. This value *could* be the result of calculating posterior probabilities on the basis of many earlier pieces of evidence, so that as each new piece of evidence is collected, the current "prior probability" is simply the "posterior probability" from the last round of evidence-gathering. On the other hand, we may not have gathered any evidence yet; in that case, $P(T)$ may simply be a hunch about the plausibility of the hypothesis.

3.2 Washing Out of the Priors

The idea that $P(T)$ could be based on a mere hunch may seem unsettling. After all, different people may have very different hunches about the truth of a theory, and so may begin the process with very different values for $P(T)$! In a way, however, this does not matter very much. This is because of the phenomenon of the *washing out of the priors*. If you and I begin with very different evaluations of T, but we agree on $P(E|T)$ and $P(E|\neg T)$, then our posterior probabilities will get closer and closer to each other the more evidence we investigate. In the long run, we will end up with the same assessment of T even if we started out with very different guesses.

3.3 Another Likelihood: $P(E|\neg T)$

We have considered the probabilities in the numerator; now let's look at the denominator. This is the sum of two values. The first, $P(E|T) \times P(T)$, is just the same as the numerator, so that is taken care of already. The other value is $P(E|\neg T) \times P(\neg T)$. Well, we already have $P(T)$, our prior probability; by the Negation Rule, $P(\neg T)$ is just $1 - P(T)$. So the only remaining probability is $P(E|\neg T)$. This is actually a very important value. It represents an assessment of how probable the evidence is if the theory is *not* true. Clearly, the more probable the evidence is if the theory is false, the less the evidence does to support the theory! And again, it seems that in most cases it will be possible to reach a reasonable assessment of this value.

4 Applying Bayes's Theorem

Now let us consider some examples to see how Bayes's Theorem can help to organize and systematize a number of observations we have already made about the relation between theory and evidence.

4.1 Confirmation by Positive Instances

4.1.1 Marbles

We remember that a fairly simple model of confirmation held that a general law is confirmed by its positive instances; for example, the theory that all ravens are black is supported by each new black raven we find.

We can find some support for this intuitively plausible idea in Bayes's Theorem. Under normal circumstances, positive instances will in fact support a hypothesized generalization.

Suppose someone tells me all the 500 marbles in a bag are black. I haven't seen any of them, and I am agnostic to begin with about whether this is true are not. I draw out a marble and look at it: it's black. How should this affect my belief in the generalization?

$P(T)$ is .5, the same as $P(\neg T)$, representing my initial agnosticism. $P(E|T)$ is 1, since if all the marbles in the bag are black, then it must be the case that the particular marble I have just examined is black. We only need one further value, namely $P(E|\neg T)$. This is trickier, and in fact unless I have some further background information, it is not clear exactly how I could assign a value to it. But it could be like this: I know that the bag was purchased at Wal-Mart, and I know that they sell two kinds of bags of marbles, namely bags of all-black marbles, and bags of multicolored marbles of which 20% are black. In that case, if $\neg T$ is the case, I would expect that the probability of observing a black marble is .2.

That's all I need to apply Bayes's Theorem. I plug in these values, and my new posterior probability is

$$P(T|E) = \frac{1 \times .5}{(1 \times .5) + (.2 \times .5)} = .83$$

Wow! That went up by a lot. But it's reasonable that it should, because the odds of getting a black marble unless the bag was all-black were rather low. If I draw another marble and repeat the exercise, with .83 as my new prior probability, $P(T|E)$ can be expected to increase dramatically once again. Let's see:

$$P(T|E) = \frac{1 \times .83}{(1 \times .83) + (.2 \times .17)} = .96$$

Compare this with what would have happened if $P(E|\neg T)$ were greater — for example, if the marble bags at Wal-Mart were either all-black or half-black. Then $P(T|E)$ would have risen to .67 after the first marble and to .80 after the second marble.

So we observe two things about confirmation by positive instances, on this model. First, under normal circumstances, positive instances *will* confirm a generalization, as we might expect. Second, the *extent* to which a positive instance confirms a generalization depends on the situation, and especially on how likely the positive instance is if the generalization is not true.

4.1.2 ESP

Consider an extreme case. My sister once bought me a book entitled something like *Test Your ESP*. This book contained some standard tests for ESP. For example, someone might deal a card face-down from a well-shuffled deck with equal numbers of five different kinds of cards. You attempt to determine what sort of card it is without looking at the face, and then check to see whether you were correct, and then repeated this procedure over a number of trials. But the book had an interesting twist concerning the interpretation of the results: if you scored better than chance, you probably had ESP, for obvious reasons; if you scored worse than chance, you probably had ESP, and were apparently unconsciously attempting to get the answers wrong, perhaps to conceal your abilities; and if you scored exactly what chance would predict, you probably had ESP, because scoring *exactly* what chance would predict is actually very unlikely. One way to look at this is to say that the theory predicted that if you had ESP you would score either above chance, or below chance, or at chance. But of course these are the only three possibilities!

So now we check to see whether this prediction comes true. Lo and behold, it does. What should the posterior probability of the hypothesis that you have ESP be, following the book's procedure? Well, let's say that you are initially agnostic again, so $P(T) = P(\neg T) = .5$. The theory (plus background information) predicts the observed result with certainty, so $P(E|T) = 1$. On the other hand, the observed result would have been obtained in any case, since it is logically necessary, so $P(E|\neg T) = 1$. Plugging in these values, we find that

$$P(T|E) = \frac{1 \times .5}{(1 \times .5) + (1 \times .5)} = .5$$

In this case, the posterior probability of .5 is completely unchanged from the prior probability! Evidence which was bound to occur even if the theory is false does nothing to confirm the theory, even if it is a positive instance of the generalization. (Notice that this is similar to a variety of cases in which a theory "predicts" something that is likely to happen in any case: if my holistic cure works, then you will get over your cold in a few days; if birth control pills prevent pregnancy in men, then Mr. Jones will not get pregnant while taking the pills; and so on.)

4.2 Falsification

Popper stressed that, although a theory can never be proven to be true, it can be decisively disconfirmed if it makes a prediction that turns out to be false. Bayes's Theorem accommodates this insight. Return to the bag of marbles. The generalization that all the marbles are black predicts that the next marble will be black (which implies that it will not be yellow, white, or any other color incompatible with being black). We draw a marble from the bag and find that it is yellow. In this case, $P(E|T) = 0$. So, assuming we were agnostic to begin with, and that 20% of the marbles in the bags that are not all-black are yellow, we get this:

$$P(E|T) = \frac{0 \times .5}{(0 \times .5) + (.2 \times .5)} = 0$$

Of course, no matter what the values of $P(T)$ and $P(E|\neg T)$ are, the final value will be zero if $P(E|T) = 0$. So Popper is partially vindicated; in a certain sense falsification really does deliver a decisive result in a way that confirmation never does.

4.3 Conclusive Confirmation?

It may seem that conclusive confirmation is sometimes possible. Return to the marbles example for a moment. Suppose that I am still testing the hypothesis that I have an all-black bag of marbles. I am convinced that the bag is either all-black or else 20% black. The bag contains 500 marbles, and I have already drawn out 100 black marbles. I draw one more out, and discover that it too is black. In this case, $P(E|\neg T) = 0$, since according to my background information the bag is either all black or 20% black, and if it is 20% black then it cannot have 101 black marbles. Assuming $P(E|T) = 1$, and assuming any prior probability we please – let's say .002 just for vividness – we get the following value for $P(T|E)$:

$$P(T|E) = \frac{1 \times .002}{(1 \times .002) + (0 \times .998)} = 1$$

So no matter how low my prior probability for the theory was, I end up with a posterior probability of 1! Isn't this precisely the conclusive confirmation that Popper denies?

The problem here is just that we are dealing with an issue of known and finite size. It's true that in this case we can conclusively verify a hypothesis (at least given our background information). However, most scientific issues are not like this. Under normal

circumstances we cannot verify that all ravens are black, for instance, by examining every raven, since many ravens have not even been born yet.

4.4 Background Information

How can we square the observation that in the case of falsification we get a posterior probability of 0 with the point made by Kuhn and many others that in general we do not regard apparently refuting evidence as conclusive? Notice that our version of Bayes's Theorem is somewhat oversimplified in that it makes no reference to background information (for example, in the flagpole case, the information that light travels in straight lines and that the sun is at a certain angle). This is why Wesley Salmon, in his writings on Bayesianism, makes the role of such background information explicit. Representing background information by B, his version of Bayes's Theorem is:

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

Notice that on this version, the posterior probability we arrive at is $P(T|E \wedge B)$. Just because this value is 0 does not show that our degree of belief in T must be 0; we could instead give up our belief in B. And in fact this will often be the rational thing to do.

4.5 Ravens

What about the ravens paradox? Does Bayesianism have anything to offer here? I think that it does (along the lines suggested by Horwich). Consider a case similar to the ravens case but in which we can use some semi-real data. Suppose I visit a small campus – perhaps my alma mater, St. Olaf College – and after seeing large numbers of blond, blue-eyed students with names like Christiansen, Rolvaag, and Ytterboe, I reach a tentative conclusion that all the students at the college are of Norwegian descent. I can try to confirm this by examining students to see whether they are of Norwegian descent. $P(E|T) = 1$, the prior probability let's say is .5, and let's say that $P(E|\neg T) = .9$, because I am pretty sure that most of the students are Norwegian, so that the chances that the next one I observe will be Norwegian are pretty good even if my hypothesis is wrong. So what effect does one more Norwegian student have on my assessment of my hypothesis?

$$P(T|E) = \frac{1 \times .5}{(1 \times .5) + (.9 \times .5)} = .526$$

It didn't go up very much, because the odds that the next student would be Norwegian were pretty good even if they aren't *all* Norwegian. But there was some discernable movement in the positive direction.

Now consider attempting to confirm the hypothesis by finding non-Norwegian non-Oles. $P(E|T)$ and $P(T)$ are unchanged. But what is the probability that the next non-Norwegian will be a non-Ole if my theory is false? Well, there are six billion or so non-Norwegians in the world and only around 3,000 Oles, so we should expect that the odds

that the next non-Norwegian will be a non-Ole are around $5,999,997,000/6,000,000,000$, i.e. .9999995. Let's plug these values in:

$$P(T|E) = \frac{1 \times .5}{(1 \times .5) + (.9999995 \times .5)} = .5000001$$

So we've confirmed the hypothesis a little, but not much! It seems to me that this may well be the correct thing to say about the ravens paradox: it's true, as the model of confirmation by positive instances holds, that positive instances confirm a generalization (unless they are equally likely if the theory is false). And it is true that what confirms a generalization also confirms anything logically equivalent to that generalization. What is missing from the traditional confirmation-by-positive-instances account is any sense of the *degree* of confirmation. When we take this into account, we see that the suspicion that, for example, non-Norwegian non-Oles do not confirm the hypothesis that all Oles are Norwegians is partially vindicated. They do confirm the hypothesis a little, but it is *so* little that it is negligible!

5 Objections to Bayesianism

Some objections to Bayesianism, with my own tentative views about what they show.

5.1 Computationally Expensive

The approach looks plausible when we examine small examples. When we consider actually trying to use a Bayesian approach to modifying belief in the face of new evidence, for instance in an AI program that must update a knowledge base in light of new information, it quickly becomes evident that the computational resources required are enormous. Alternative, less computationally expensive approaches have been considered, for example the Dempster-Schafer theory of evidence.

However, it seems to me that this does not diminish the value of Bayesianism as a theoretical idealization that can help us to understand the nature of the relation between theory and evidence.

5.2 Doesn't Explain Simplicity

Another objection to the view is that it does not explain why simplicity is a virtue. (Glymour makes this claim, suggesting that a Bayesian approach will not tell us what is wrong with a "DeOccamized theory," i.e. a theory with some nonfunctioning junk thrown in.)

Suggested response: we should distinguish between two sorts of virtues a scientific theory may have. There is an extremely important *empirical virtue*, namely being confirmed by the evidence. But there are also *nonempirical* or *pragmatic* virtues, such as simplicity, elegance, scope, and conservatism. To explain why the pragmatic virtues are desirable, we need different sorts of arguments.

(For example, what is so good about conservatism, that is, staying close to what we already believe, given that the world may be radically different than we think it is? One answer is simply that, even if the correct theory is radically different from the one we have, the most efficient way to discover it is likely to be by a succession of small changes to our existing theory. Simply leaping far from our own theory to radically different alternatives *could*, through great good fortune, land us at the correct theory, but it is more likely to lead us to theories that are wildly false!)

5.3 Confirming while Reducing Probability?

van Fraassen suggests that evidence which reduced the probability of a theory could still have the effect of making it more worthy of belief if it increased the difference in probability between the theory and its near rivals.

This is interesting, but nothing Bayes's theorem cannot accommodate. We can use the generalized version of the theorem to simultaneously evaluate the impact of evidence on a theory and several rivals. van Fraassen's point is important, and counts against a simple measure of the degree to which evidence confirms a theory (e.g. as the ratio of the posterior to the prior probability). But Bayes's Theorem itself can quite comfortably accommodate a conception of confirmation that involves comparing the posterior probability of one theory with those of its rivals.

5.4 Problem of Old Evidence

What is the problem? Bayesians interpret Bayes's Theorem as concerned with subjective probabilities, i.e. degrees of belief, not objective probabilities. Any evidence I already know about should have a subjective probability of 1, regardless of what it is conditionalized on. Since the denominator of Bayes's Theorem is just an expansion of $P(E)$, the denominator must be one, which means that $P(T|E) = P(E|T)P(T)$, and since $P(E|T)$ is 1 also, this means that $P(T|E) = P(T)$: that is, "old evidence" can have no effect whatsoever on the degree to which I believe a theory.

However, this doesn't seem right to many people. It does seem that, if a theory predicts evidence I knew about beforehand, this adds some credibility to the theory. Can Bayesianism account for this?

5.5 Dependence of Confirmation on Available Theories

It is true that how much evidence confirms a particular theory depends on what its rivals are (this is related to the previous point). We always test the theories we have thought of, not all the logically possible theories we haven't thought of. But it seems to me a virtue of Bayesianism that it calls this to our attention.