2002

A Very Brief Primer on Bayesian Methods in Evidence

Richard D. Friedman
University of Michigan Law School, rdfrdman@umich.edu

A Very Brief Primer on Bayesian Methods in Evidence

Richard D. Friedman, University of Michigan

I have been asked to write an extremely short explanation of the Bayesian approach to evidentiary issues, for the benefit of those who regard themselves as probabilistically challenged. Although the application of Bayesian probability to evidence has generated a good deal of debate, its use as a heuristic device should not be particularly controversial.

Evidence concerns propositions that are uncertain. Accordingly, some concept of probability must play a role. Standards of persuasion, such as "more likely than not" and "beyond a reasonable doubt" are clearly probabilistic, and the definition of relevant evidence, as expressed in Fed. R. Evid. 401, is explicitly probabilistic. The standard probability calculus expresses the probability of a proposition as a number ranging from 0 (for impossibility) to 1 (for certainty). The best interpretation of a probability statement, most Bayesians would say, is as a subjective assessment of one's level of confidence that the given proposition is true. Here's a simple example. Recently, I ran a 5K race with my 9-year-old daughter. She was aiming to come in under 30 minutes. Partway through the last mile, she asked me – this is the truth – "Can I break half an hour? Is there better than a 40% chance?" (Is there a genetically transmitted affinity for probability?) The question made sense, even though the race was a non-recurring event, and she was either going to break the 30-minute barrier or not. (She did, with 14 seconds to spare.)

A central part of Bayesian probability theory is Bayes' Rule (or Theorem) named for an 18th century English clergyman. The Rule is not all there is to the theory, but I will concentrate here on it, because it is particularly useful in evaluating trace evidence, and it is in that context that thinking explicitly about probability can often be most helpful. I use the term trace evidence broadly, to include any evidence that assertedly arose from the condition at issue. It could be fingerprints or DNA or tire treadmarks, but it could also be testimony. Typically, the proponent of the evidence contends that the evidence arose as a result of one hypothetical condition; the opponent may contend that it arose as a result of another hypothetical condition.

Take a DNA case: A spot of blood was found at the scene of a crime. A test indicates that this blood has the same DNA profile as the accused's blood. Call this evidence of a matching profile E. The prosecution contends that E arose as a result of the accused having been at the scene of the crime and having bled there. Call that hypothesis H1. The accused may offer various counter-hypotheses (and is not limited to one), but let's suppose that he focuses on the hypothesis, which we can call H2, that the blood was left by someone else who happened to have the same profile. Now suppose an expert testifies that *if* the defendant left the crime stain, there would almost certainly be matching profiles. That is, $P(E|H1)$, or the probability of E given H1, is very close to 1. Furthermore, the expert testifies that if the defendant did *not* leave the crime stain, the probability that the person who did leave it would have a profile matching the accused's is very small, 1 in a billion. Thus, $P(E|H2)$ is 1 in 1 billion. Now, prosecutors sometimes argue from such evidence, "The probability that someone other than the accused left this blood is 1 in a billion." But that is a fallacy – in fact, it is a version of what is often called the prosecutor's fallacy. The expert has told us the probability of E given H2, but the prosecutor has incorrectly equated it with the probability of H2 given E; she has *transposed the conditional.* The two probabilities may be very different. Consider this classic example: The probability of the hypothesis that an animal has four legs given the evidence that it is a cow is very high, nearly 1. But the probability of the hypothesis that an animal is a cow given the evidence that it has four legs is quite low.

So our prosecutor has argued fallaciously, but we *do* want to reach an assessment of the probability of the competing hypotheses given the evidence. Here is where Bayes' Rule comes in handy. Before presenting the Rule, I will introduce a few other concepts. A familiar concept is odds. Odds are another way of stating probabilities. The odds of a proposition are the probability that the proposition is true divided by the probability that the proposition is not true. Thus, a probability of .5 corresponds to odds of 1:1 – what are sometimes called even odds. A probability of .25 corresponds to odds of 1:3. The *prior odds* of a proposition are the odds as assessed without taking into account the particular evidence in issue, in this case the DNA match. The *posterior odds* are the odds once we take that evidence into account; they are the bottom line that we are trying to assess. Finally, the *likelihood ratio* of a piece of evidence with respect to a hypothesis is a fraction, the numerator of which is the probability that the evidence would arise if the proposition were true and the denominator of which is the probability that the evidence would arise if the proposition were not true. In our DNA case, the likelihood ratio is approximately 1 billion; the numerator is close to 1 and the denominator is 1/1,000,000,000.

Now we are ready for Bayes' Rule. (I won't include a derivation here.) Expressed most simply, the Rule declares that the posterior odds equal the prior odds times the likelihood ratio. Thus, if the likelihood ratio is greater than 1, the evidence makes the proposition more probable, if it less than 1 the evidence makes the proposition less probable, and if it is exactly 1 the evidence is irrelevant to the proposition. Put another way, *all other things being equal*, the posterior odds will be greater

- *the greater are the posterior odds..* Thus, suppose that in our DNA case before hearing the DNA evidence the jury thought it was about as likely as not that the accused was the source of the DNA – that is, the prior odds were 1:1, or simply 1. Then the posterior odds equal the likelihood ratio of 1 billion, meaning that it is overwhelmingly probable that the accused is the source of the crime scene stain. If, on the other hand, before hearing the DNA evidence the jury thought the accused was no more likely than anyone else in the world to be the source of the crime scene stain, then the DNA evidence, though very powerful, will not make it probable that the accused is in fact the source. Note that in a case like this, the DNA expert really should have nothing to say about what the prior odds are; the expert has nothing useful to offer on that point, which is a jury matter.

- *the greater is numerator of the likelihood ratio.* One factor making DNA evidence powerful is that a match is so probable if the accused is the source of the crime scene stain; that is, the numerator of the likelihood ratio is close to 1. But a high numerator does not in itself make the evidence powerful. Say we have a diagnostic test that yields a positive result in 99% of the cases in which a patient has a given disease. If the test also yields a positive result in 98% of the cases in which the patient does *not* have the disease, it's a virtually worthless test; a positive result has hardly any probative value. Similarly, a low denominator does not in itself mean that the evidence is not powerful. Suppose a smoking pistol with the accused's fingerprints is found by the side of a shooting victim, and defense counsel argues, "This helps my client if anything; it's so unlikely that he would have shot the victim and been dumb enough to leave the pistol with his fingerprints right there." But all that argument addresses is the numerator of the likelihood ratio. Even though the numerator is small, if (as seems likely) the denominator is much smaller – it is highly improbable that we would have smoking-pistol-with-accused's-fingerprints-by-the-side-of-gunshot-victim if the accused did *not* shoot the victim – then the evidence cuts against the accused.

- *the lower is the denominator of the likelihood ratio.* What really gives the DNA evidence power is that the probability of the profiles matching if the accused was not the source of the DNA is so small. Compare old-fashioned blood typing. If the crime scene stain has O+ blood, and that is the accused's blood type, that is some evidence favoring the prosecution, but it has nowhere near the power of the DNA test. The numerator of the likelihood ratio is close to 1, as in the DNA case, but the denominator is substantial, about .4. So the evidence more than doubles the odds in favor of the prosecution – but if the odds were tiny beforehand they will still be tiny. DNA evidence, by contrast, can transform what previously appeared improbable to overwhelmingly probable.

Bayesian thinking does not require numerification. We all think in terms of magnitude without necessarily assigning numbers; "unlikely," for example, is a non-numerical expression of probability. In thinking about how probable a proposition is in light of a given piece of evidence, it is often helpful to think: How probable did this proposition appear before I knew of the evidence? Is it more probable that this evidence would arise if the proposition were true than if it were not true? If you are asking those questions, then you may be thinking like a Bayesian without even knowing it.