Bayes' Theorem for Intelligence Analysis

More on probability — I

Jack Zlotnick

The intelligence interest in probability theory stems from the probabilistic character of customary intelligence judgment. Intelligence analysis must usually be undertaken on the basis of incomplete evidence. Intelligence conclusions are therefore characteristically hedged by such words and phrases as "very likely," "possibly," "may," "better than even chance," and other qualifiers.

This manner of allowing for more than one possibility leaves intelligence open to the charge of acting the oracle whose prophecies seek to cover all contingencies. The apt reply to this charge is that intelligence would do poor service by overstating its knowledge. The very best that intelligence can do is to make the most of the evidence without making more of the evidence than it deserves. The best recourse is often to address the probabilities.

The professional focus on probabilities has led to some in-house research on possible intelligence applications of Bayes' Theorem. At the time of my participation in this research, I was an analyst in the Central Intelligence Agency, which sponsored the scholarship but took no position of its own on the issues under study. My personal views on these issues, as elaborated in the following pages, have no official character.

The Bayesian Approach

Bayes' Theorem in its odds-likelihood form served participants in our test program as their diagnostic rule for appraising new evidence. The odds-likelihood formulation of Bayes' Theorem is the equation

R is the revised estimate of the odds favoring one hypothesis over another — the estimate of the odds after consideration of the latest item of evidence. P is the prior estimate of the odds — the odds before consideration of the latest item of evidence. There is no escaping some starting estimate of P. However, after the starting estimate was in hand, the participating analysts offered no judgments about P. It was a value carried forward in machine memory from previous analysis. R, the result of the mathematical processing, was what went back into machine memory to become the value of P used in consideration of the next item of evidence. The participating analysts offered judgments only about L, the likelihood ratio.

The likelihood ratio was the analyst's evaluation of the diagnosticity of an item of evidence. Evidence is diagnostic when the chances of its appearing are different if one hypothesis is true than if another hypothesis is true. Suppose intelligence is asked to estimate the comparative merits of two hypotheses — one of imminent war, the other of no imminent war. The estimate is to be expressed in terms of the odds favoring or disfavoring the war hypothesis. The latest evidence is deployment of foreign troops to a border area. Is the deployment deemed to be say two times more likely if the war hypothesis is true than if the no-war hypothesis is true? Then the evidence is certainly diagnostic. The value of L, a judgment of the analyst communicated to the machine processor, would in this case be the fraction $^2/_1$.

Three principal features of Bayesian method distinguish it from conventional intelligence analysis. The first is that the intelligence analyst is required to quantify judgments which he does not ordinarily express in numerical terms. This requirement to quantify probabilistic judgment is the feature that perhaps draws most of the critical fire against the Bayesian approach in intelligence analysis. A debating point

of the critics is that analysts are bound to disagree in their opinions of the exact figure that should represent the diagnostic value of an item of evidence. The Bayesian rebuttal is that disagreement among analysts is just as much a characteristic of traditional method and is no less serious for being implicit rather than explicit in the analysis. The critic returns to the debate by observing that the typical analyst, being a verbal and not a mathematical man, finds it inordinately difficult to express his degree of belief to the precision implied by a numerical value. The partisan of Bayes, for his part, takes the position that people have been quantifying probabilistic judgments since the beginning of time — whenever they offered or accepted betting odds on the outcome of any doubtful issue.

The second distinguishing feature of Bayesian method is that the analyst does not take the available evidence as given and draw therefrom his conclusions about the relative merits of opposing hypotheses. He rather postulates, by turns, the truth of each hypothesis, addressing himself only to the likelihood that each item of evidence would appear, first under the assumption that one hypothesis is true and then under the assumption that another hypothesis is true. The analyst is under no ego-supporting need to hold to positions previously taken on the merits of the respective hypotheses; he does not feel called upon to reinforce his self-esteem by reaffirmation of opinions previously put on the record.

The third distinctive feature of Bayesian method is that the analyst makes his judgments about the bits and pieces of evidence. He does not sum up the evidence as he would have to do if he had to judge its meaning for final conclusions. The mathematics does the summing up, telling the analyst in effect: "If these are your readings of the individual items of evidence, then this is the conclusion that follows." The research findings of some Bayesian psychologists seem to show that people are generally better at appraising a single item of evidence than at drawing inferences from the body of evidence considered in the aggregate. If these are valid findings, then the Bayesian approach calls for the intelligence analyst to do what he can do best and to leave all the rest to the incorruptible logic of a dispassionate mathematics.

The Bayesian approach was not studied with any idea of its replacing other approaches in intelligence analysis. The responsibility of intelligence is to depict, as best it can, the current and prospective state of international affairs. The intelligence estimate is a closely reasoned

analysis of such important matters of interest as the top political leadership of a foreign country, evolving popular attitudes in that country, changing force structures in its military establishment, its levels of scientific achievement, and the hard choices it is making in allocation of resources to the guns and butter sectors of the economy. The intelligence estimate is sketched in all the lights and shadows of descriptive, narrative, and interpretive commentary. This task is not reducible to terse statement of the odds favoring one particular hypothesis over another.

There are, however, areas of intelligence analysis where Bayes' Theorem might well complement other approaches. One crucially important area is that of strategic warning — the analysis directed to uncovering any pattern of activity by a foreign power suggestive of a major and imminent threat to US security interests. The patterns of events leading to Pearl Harbor in 1941 and to the Communist invasion of South Korea in 1950 are cases in point. Strategic warning analysis focuses primarily on just the problem that Bayes' Theorem addresses — the odds favoring one hypothesis (say imminent attack) over another hypothesis (no imminent attack).

The Research Task

One way to test the usefulness of Bayes' Theorem for intelligence analysis is to replay intelligence history. This means going back to international crises of years past. It means assembly of the evidence which was available before the outcomes of the crises were known. It means reading the old intelligence estimates and other studies in order to find out how the analysts of the day interpreted the evidence. It means assignment of L values — likelihood ratios — that honestly reflect these analyst evaluations of the evidence at the time and not our present hindsight knowledge.

Another way to test Bayes' Theorem is on current inflows of evidence. The advantage of this kind of testing is that hindsight knowledge does not intrude; Bayes' Theorem is pitted fairly and squarely against the conventional modes of analysis. Offsetting this advantage for honest research, however, is a disabling disadvantage.

The disadvantage derives from the very nature of the hypotheses at interest in strategic warning. The alternative hypotheses are commonly of two types. One stipulates continuation of the status quo. The other stipulates sudden change from the status quo. Usually the situation today is pretty much what it is going to be a week from today. The status quo hypothesis, in other words, usually turns out to be the true one in strategic warning analysis. But the main test of strategic warning effectiveness is the capability to give forewarning of the sudden changes that occasionally do occur in the status quo. The intelligence interest in Bayes' Theorem is primarily in how well the Bayesian approach to strategic warning would meet this main test of performance in situations of general surprise, without chronic resort to cry-wolf false alarms. Unfortunately, intelligence research cannot be speeded up by focus on the particular current issues which will turn into occasions of intelligence surprise. If intelligence could pick out in advance the issues on which it was going to be surprised, it would by definition never be surprised, and it would have no interest in the possible contributions of Bayes' Theorem to improved analysis.

The outlook, then, is that many tests of Bayes' Theorem on current inflows of evidence will be needed to get the few interesting occasions that show Bayesian performance in circumstances of general intelligence surprise. And just a few interesting examples are not enough to make the case for or against the Bayesian approach, which may do better than conventional method sometimes and not as well other times. A large enough sample of interesting examples is needed to justify confident findings of comparative performance on the average.

The results of the testing so far have been interesting enough to make a good case for further testing of Bayes' Theorem in intelligence analysis. Among the interesting results has been an uncovering of problem areas that flank the path of intelligence analysis and that are not very easily outflanked.

The Life-Span of Evidence

One such problem area has been called nonstationarity. In situations of nonstationarity, that is, when hypotheses are being effectively altered by

the passage of time, evidence will have a limited life-span. An intelligence hypothesis about current Soviet policy is not exactly the same hypothesis on January 15 that it is on February 15. The date has changed, so the hypothesis is to a degree different; and evidence back in January which had a certain bearing on the hypothesis of what, was then current Soviet policy does not have the same bearing on the hypothesis of what is current Soviet policy a month later.

Consider, for example, some evidence which was available to intelligence and to the public at large in the summer of 1962, before photographic confirmation was received of missiles in Cuba that could reach targets deep in the United States. Soviet leaders gave public assurances during this period that the expanding military aid to Cuba was for defensive purposes only. Now an analyst's appraisal of this kind of assurance will depend partly on how honorable or dishonorable he believes Communists to be. But whatever his views about the honor of Communists, he would certainly not consider any government's assurances to constitute a commitment for all eternity. Governments do make new decisions and reconsider old ones. This amounts to saying that the diagnostic value of evidence bearing on hypotheses about current government policy tends to erode over time. A mathematical logic for strategic warning analysis has to be attentive to this erosion. Perhaps the analyst can specify the expected rate of erosion when he first encounters an item of evidence. If he cannot or prefers not to, the Bayesian approach does not quite attain the mechanistic ideal that would require of the analyst only his one-time attention to each item of incoming evidence. The analyst instead finds himself looking back from time to time at his whole body of past evidence, to consider whether its diagnostic value, as recorded in machine memory, is still valid and not out-dated.

Causal Evidence

Another problem area spotlighted in the testing is the occasional reversal in cause and effect relationship between hypotheses and data. The disease generates the symptoms of the disease, and so the physician can infer the disease from the symptoms. Similarly in his surveillance of the Soviet scene, the intelligence analyst in Washington

can infer from Soviet actions a good deal about Soviet policy. But the analyst also has his eye cocked for relevant data other than Soviet actions, data which have less a derivative than a causal relationship to Soviet policy. I draw again on the Cuban missile crisis of 1962 for my historical example.

On several occasions that year, President Kennedy publicly warned that the United States would take a grave view of strategic missile emplacements in Cuba. How would a Bayesian analyst evaluate President Kennedy's warnings for their relevance to opposing hypotheses about Soviet missile shipments to Cuba? If the analyst were a mechanical, uncritical Bayesian, he would say to himself: "President Kennedy is more likely to issue these statements if the hypothesis of imminent Soviet missile shipments to Cuba is true than if the hypothesis of no such missile shipments to Cuba is true. My L in the Bayesian equation R=PL is greater than $^1/_1$, and so my mathematics works out to an increase in the odds favoring the missile hypothesis."

Well, the analyst in this case is surely not reasoning as President Kennedy reasoned. The President no doubt felt that the clear communication of American concern would either have no effect on Moscow or, hopefully, would dissuade the Soviet leadership from shipping strategic missiles to Cuba. He thought, in other words, that his statements would tend to reduce, not increase, the odds favoring the missile hypothesis.

The complication for the Bayesian analyst is the causal character of President Kennedy's statements. Soviet actions are direct derivatives of Soviet policy. President Kennedy's statements were not. They were important primarily for the chance that they would affect, not reflect, Soviet policy.

It can be shown that, in principle, Bayes' Theorem is as applicable to causal evidence as to derivative evidence. In practice, Bayes' Theorem often offers slippery ground to the analyst appraising causal evidence. In practice, the analyst does better by putting a little sand in his tracks. He gets better mental traction in this case by making a direct judgment about the impact of the causal evidence on the comparative merits of his hypotheses. He says to himself: "If the odds were even-money in favor of the missile hypothesis before receipt of the causal evidence, what would the odds be now after receipt of this evidence?" When the prior odds are even-money (that is, $^1/_1$), the revised odds equate to the

likelihood ratio, according to the Bayesian equation R = PL. So, by making a direct judgment of revised odds following a stipulation of even-money prior odds, the analyst obtains an effective likelihood ratio to give the computer.

This is an approach which respects the mathematics of Bayes but does violence to the spirit of Bayes. One of the attractive features about Bayesian method in its pristine purity is that the analyst need address himself to the merits of the hypotheses only at the very beginning of his analysis. In principle, he does not thereafter reaffirm his first opinion, admit to a change in opinion, or criticize anybody else's opinion on the subject. He is supposed to make a judgment, instead, of quite another sort, a judgment about the evidence which postulates the truth of each hypothesis in turn, a judgment which does not involve him again in debate about the merits of each hypothesis. His encounters with causal evidence, however, often do not allow him to keep quite this detachment from the hypotheses. He finds himself addressing R, not L.

Catch-All Hypotheses

Another problem area encountered in our research has been examined in Bayesian literature as the nonindependence issue. Nonindependence enters into analysis as a complicating feature when the likelihood ratio — the L value of an item of evidence — is affected by the previous pattern of evidence.

Nonindependence is an arcane subject to analysts who are new to probability mathematics, mainly perhaps because items of evidence which are independent if one hypothesis is assumed true can be nonindependent if another hypothesis is taken as true. Analysis is easier when items of evidence are independent (or to put it more properly, conditionally independent) — that is to say, when the likelihoods of their being received do depend on which hypothesis is assumed true but when these conditional likelihoods hold regardless of the previous pattern of evidence. Intelligence analysts have their way of reaching for conditional independence, whether or not they have ever heard of the nonindependence issue. They reach for a new hypothesis to do service for some hypothesis that no longer seems suitable as originally worded.

Such an unsuitable hypothesis could be the one postulating continuation of the status quo in the strategic warning problem. This catch-all hypothesis can be divided into two or more subhypotheses (and it can be divided different ways into different sets of subhypotheses). For an illustrative example, take any case in history of a big power threatening its much smaller neighbor and finally invading the little country when threats alone did not avail.

Suppose the invasion is preceded by reports that the big power is moving its troops toward the border. Considered later in time from the vantage point of hindsight, the troop movements certainly would seem to be strong evidence, which ought to have tipped the odds substantially in favor of the invasion hypothesis. But the analyst of the day would probably find himself reflecting on at least two relevant subhypotheses of the no-invasion hypothesis. Subhypothesis A might be that the big power will not invade the little country but will apply very strong pressures — psychological, political, and other — just short of military invasion. Subhypothesis B might be that the big power will neither invade nor apply other extremes of pressure against the little country.

Now the analyst using Bayes' Theorem introduces an initial opinion about the hypotheses when he begins his analysis. He must similarly introduce an opinion about the subhypotheses if he comes to make them explicit elements in his analysis. By the time he receives the reports of troop movements, the previous evidence will have inclined him to the opinion that subhypothesis A — strong pressures against the little country — is the only reasonable interpretation of the noinvasion hypothesis. The events leading up to the troop movements (the grim warnings, the shrill propaganda, the military alerts) will constitute such virtual contradiction of subhypothesis B — no extremes of pressure — as to give it a near-zero probability. If this is the analyst's view, then the troop movements toward the border must seem almost as likely under the no-invasion hypothesis as under the invasion hypothesis. His L is just about $1/1$. His Bayesian approach has done virtually nothing to change his current odds.

This undiagnostic character of incoming evidence near the climax of international crises may seem novel to novices; it is familiar enough to experienced intelligence analysts. The more experienced they are, the more rueful they are likely to be in their recollections of evidence that was ambiguous to contemporaneous vision but became telling in

retrospective inquiries.

False Evidence

Perhaps the most difficult problem area is the suspect character of some evidence. The intelligence analyst gets his information in accounts from sources of varying reliability. He does not know for sure which accounts to believe and which to disbelieve. So he has to appraise his evidence, not only for its bearing on the hypotheses, but also for its probability of being accurate. The estimated probability of accuracy will enter into the analysis and will affect final results.

Unfortunately, an analyst's opinion about a report's probable accuracy or inaccuracy will be influenced by his current opinion about the hypotheses. Does he find it hard to give credence to reports from Cuban refugees who claim to have seen objects resembling medium range missiles near Havana? If he is skeptical, it may well be because he finds it hard to give credence to the hypothesis that the USSR will do anything so foolish as to ship such missiles to Cuba. So once again, we have a case of information not doing the work which critics later, in all the wisdom of hindsight, will say it should have done.

The Research Promise

My exposition of these problem areas is not meant to imply that they muddle only the Bayesian approach; they plague — with fine impartiality all types of intelligence analysis — traditional method as well as Bayesian method, verbal logic as well as mathematical logic. Traditional method also must cope with the eroding diagnostic value of past evidence as it recedes into history. Traditional method also finds it harder to draw probabilistic conclusions about the state of the world from causal evidence than from derivative evidence. Traditional method also sometimes explains away evidence that can be explained away by a favored subhypothesis of a catch-all hypothesis. Traditional method also has to contend with the implausibility of evidence that is not in

character with the climate of prevailing opinion.

My purpose in expanding on the problem areas is to show that much of the difficulty in intelligence analysis is not the difficulty to which the Bayesian approach is addressed. The Bayesian approach seeks to insulate analysis from frailties of logic in aggregating the evidence. The working world of intelligence, however, is concerned not only about possible inconsistency in everyday thinking between the conclusion drawn from the body of evidence considered as a whole and the conclusion that should logically follow from judgments about the evidence considered item by item. Intelligence views with concern also the possibilities of mistaken judgments about individual items of evidence. The intelligence pragmatist is wistful about evidence which almost speaks for itself, evidence to which most people will attribute much the same probability values because the values can be documented by, say, actuarial statistics or other such extrinsic authority. The pragmatist feels that an increase in the amount of this kind of evidence would do more to help men reach sound conclusions than could any formal logic — Bayesian or other — for reasoning from uncertain propositions about the evidence.

Conceding this point, the Bayesian responds that intelligence must still do the best it can with what it has. In a world of fallible judgments about evidence, the Bayesian approach is not a path to perfection; it can be at best only a path to improvement. The promise of the research on Bayesian method is a mathematical logic to which intelligence can have recourse for substantiating or contradicting the verbalizations of the traditional analysis. When the different approaches lead to discrepant conclusions, intelligence should perhaps undertake to rethink, recalculate, and if possible reconcile. The research interest at this time should be to find out whether such a Bayesian cross-check on other reasoning would significantly improve the quality of analysis.

Footnotes

* Paper presented at the Conference on The Diagnostic Process, Ann Arbor, Michigan, 18 June 1970.

Posted: May 08, 2007 08:33 AM