Bayesian Analysis for Intelligence: Some Focus on the Middle East

APPROVED FOR RELEASE 1994
CIA HISTORICAL REVIEW PROGRAM
2 JULY 96

~~SECRET~~

More on analytical techniques

Nicholas Schweitzer

The job of intelligence is becoming increasingly more complex, partly because of changes in the international order, but primarily because there is ever more information available. The following measures of this "information explosion" are often cited, in popular works like Alvin Toffler's *Future Shock* and in more sober publications like the Harvard Business Review:

> — scientific and technical information is generated at the rate of 6,000,000 pages per year;
> — the U.S. government alone publishes more than 100,000 reports per year, not to mention more than 450,000 articles, books, and papers;
> — the world's output of books is approximately 1,000 titles per day.

At times the analysts liken the intelligence process to a funnel, in which they are the narrow neck which has to read and assimilate everything that passes through. The problem is that the top of the funnel has grown larger by the year, while the neck has remained the same.

How does one cope with this information flow? What techniques can aid the intelligence analyst in assessing data more efficiently? How can the various individuals, offices, and agencies better coordinate their efforts? How can we improve our capabilities to perceive the gossamer without rending it, and to react to the ephemeral before it passes? One obvious answer is computer assistance in selecting, filing, retrieving, and coordinating data, as well as in editing and producing the end reports of this process. The SAFE system and others under development in the Agency are promising responses to these needs. Another answer may lie in the use of more sophisticated methods of data analysis, often developed in academia and often dependent on computer processing for data correlation or other statistical computations, such as modeling and simulation.

It is obvious to the most casual observer, however, that such analysis will take us only a very short distance in understanding the politics of the Middle East. Even if one knows the military, economic, and social capabilities of the major countries in the Middle East, the individual personalities of leaders such as President Asad, President Sadat, and Prime Minister Rabin, along with the political pressures which buffet them, make the prediction of events one of the most Sisyphean of human endeavors. Hans Morgenthau in *Politics Among Nations* categorically states that "the first lesson the student of international politics must learn and never forget is that the complexities of international affairs make simple solutions and trustworthy prophecies impossible."

The more experience we gain in attempting to apply "methodological" analysis to complex political-military problems such as the near future of the Middle East, the more hopelessly doomed to failure such mechanistic models seem. They will continue to be so until the model employed is at least as complex and flexible as that which a human expert is capable of working with in his own head, and even that mental model — as we are so painfully aware — is far from adequate to encompass the complexities of such a situation.

What I shall describe here is one very modest method of adding certain benefits of probability and statistics to traditional analysis. The central technique employed in the reports was Bayesian inference,* but equal credit for the project's success must go to an adaptation of the Delphi technique. Over the past three years, OPR has applied these techniques to the study of various situations involving a potential for hostilities, including the likelihood of a major North Vietnamese offensive during the dry season of 1974 (the year before the debacle), the prospect of Sino-Soviet hostilities from summer 1974 to the present, and the likelihood of Arab-Israeli hostilities from autumn 1974 to June 1976, when Arab-Israeli issues took a back seat to the internal problems of Lebanon. The result has been three series of reports, each coordinated by OPR and issued on a periodic basis, and all well-received by their readers.

In the course of our projects, we have recognized many limitations and have harbored certain reservations about the applicability of the method; such limitations and reservations are discussed in Appendix 2. In many ways, the Middle East was the most complex of our studies, and our analyses were correspondingly less straightforward, less satisfying, and less impressive than the other two. By the same token, however, this situation presented us with the greatest challenges and pushed us against the limits of the design. The examples used in this paper have therefore been drawn from that problem.

II. Description of the Techniques

Bayesian Analysis

The statistical formula which forms the basis for our analysis bears the name of the Reverend Thomas Bayes, who was the first to express in precise quantitative form this particular mode of inductive inference. His work, entitled "An Essay towards Solving a Problem in the Doctrine of Chances," was read posthumously in 1763 before the Royal Society, of which he was a fellow. The most accessible copy of it appears in *Biometrika*, Dec. 1958, pp. 293-315. It is a tool of statistical inference, used to deduce the probabilities of various hypothetical causes from the observation of a real event. It also provides a convenient method for recalculating those probabilities in the light of a continuing flow of new events. Reduced to its simplest form — and it is by no means a difficult formula to begin with — the "rule of Bayes" states that the probability of an underlying cause (hypothesis) equals its previous probability multiplied by the probability that the observed event was caused by that hypothesis. Once the probabilities are assigned, which is the difficult part, the mathematics are as simple as A=B x C. (A more detailed explanation of the mathematics involved is given in Appendix 1.)

One of the classic uses of this type of induction is described by Frederick Mosteller and David Wallace in "Inference in an Authorship Problem," in the *Journal of the American Statistical Association*, June 1963, pp. 275-309. This was an investigation to resolve the disputed authorship of twelve of the Federalist papers (49-58, 62 and 63), based upon the frequency per 1000 words of various non-contextual words such as *upon, also, though, although, while* and *whilst*. The rule of Bayes was used to derive the probability of Hamiltonian or Madisonian authorship for the papers, based upon observed frequencies of the tell-tale words in *The Federalist* and in other examples of their writing. The results, which are considered a classic example of the overwhelmingly strong inference which can attend this type of analysis, were in consonance with most contemporary assessments, assigning all of the disputed papers to Madison.

In the field of intelligence, a perfect application for Bayesian inference has been developed to identify military units and installations which are seen in photography. For example, groups of weapons, buildings, and local improvements may be observed, and the type of military unit — such as an infantry regiment or a motorized rifle battalion — may be immediately obvious. Often, however, certain typical features may be absent, or certain extraneous pieces of equipment may confuse the identification. In such cases, the inferential strength of the rule of Bayes has frequently been able to cut through the noise of a few contradictory items to assign a high probability to a particular type of unit.

An explanation of how this is done will serve as a useful lead-in to the use of Bayes for political analysis. The starting point, the data base from which inference is drawn, is a set of probabilities for seeing certain identifying features in different units. For example, whereas a group of ten tanks may be a common sight in a motorized rifle battalion, ten tanks may very rarely be seen in an infantry regiment. Thus the following probabilities might be assigned: 90 percent that such a group of tanks would be seen if the unit is a motorized rifle battalion, and 10 percent that the tanks would be seen if the unit is in fact an infantry battalion. These probabilities, called the "objective probabilities," can be derived either from historical observation or from expert opinion; the early application of this technique drew upon the knowledgeable guesses of analysts in this field; these probabilities were later supplemented by a study of known units. When the probabilities associated with a whole range of identifying features and equipment are aggregated using Bayesian analysis, the type of unit under study often emerges clearly from the noise.

This paper describes the application of Bayes to political analysis, an even more complex field in which there are no objective probabilities of events, and in which the historical examples of previous conduct are ambiguous or inapplicable. The political, economic, strategic, and social events of the world are imperfectly understood and difficult to measure. For example, what is the difference in significance if Shimon Peres says "the Israeli military movements are strictly precautionary," rather than "the Israeli military movements are strictly defense"? Which is more probable if Israel is seriously considering a preemptive attack? How much more probable? Would we be able to derive some form of objective probability from a catalog of previous uses of those phrases and previous attacks? And how probable are such statements even if no attack is in the wind? The interpretation of public statements and troop movements is difficult, to be sure, but it is still our job to interpret them. Hence to use a technique like Bayes it is necessary to turn to expert judgments expressed quantitatively. The values assigned are educated guesses and are imprecise, but they provide a starting point and at least a rough basis for comparison and analysis. Our experience suggests that it is relatively easy to induce analysts accustomed to qualitative expressions of probability to shift to numerical assessments. If we can face the challenge of assigning probabilities to such events, Bayesian analysis can allow us to squeeze a little more information from the data we do receive. One danger, however, is the ever-present tendency to attribute more precision to a number than is warranted, and it must continually be stressed that the numbers are only approximate.

It is only proper at this point to mention that individuals in the Agency have investigated the utility of this technique for many years, and that much of the acceptance of our current efforts is due to earlier experiments applying Bayes to the analysis of historical intelligence situations. As Mark Twain said, "Habit is habit, and not to be flung out of the window by any man, but coaxed downstairs a step at a time.

Delphi

The Delphi technique was developed by the RAND Corporation in the late 1940s as a systematic approach to soliciting, improving, and combining expert opinions on a subject. Its major points are

> (1) formulation of the problem under investigation in quantitative terms,
>
> (2) interrogation of experts through questionnaire or interview,
>
> (3) controlled iteration, in which the results are usually presented statistically and the anonymity of individuals is preserved.

It has been used by government, business, contract R&D organizations like RAND, and academic institutions in hundreds of studies, often as a method of forecasting scientific and technological progress. It tends to break down barriers between disciplines and to stimulate creative thought through cross-fertilization from related and unrelated technical fields. It also tends to elicit ideas from experts in a setting which enjoys some of the benefits of large groups without the difficulties of group dynamics and personal competition. As will become apparent in the next section, these major components of Delphi were all incorporated in the Bayesian analyses.

A historical footnote at this point may lend credence to the premise that mere expert opinion can be used to investigate the labyrinthine corridors of international affairs. A study by Frank Klingberg in 1937 analyzed responses from 220 persons who were judged to be knowledgeable about world affairs. They were asked to rate from 0 to 100 the probability of war within ten years for 88 pairs of states. The results, as reported by Quincy Wright in *A Study of War* (pp. 338-340 in the abridged version) were highly correlated with the orientation and sequence of entry of states in World War 11. Wright's critique of the study was that "predictive results of some value for a few years ahead can be obtained from an analysis of expert opinions."

The Specific Adaptation of the Techniques

In creating a workable vehicle for intelligence analysis, there has been much adaptation. We have faced the realities of working with individuals who often are under pressure, and we have tried to develop a genuinely useful procedure without being Procrustean. There is often a conflict between pure theory and applied engineering, and although we do our best to satisfy both, we tend to favor the engineer.

The actual procedure for the reports is a periodic routine. On the first day of the period, each of a number of participating analysts submits the items of evidence he or she has seen since the last round which relate in any way to possible hostilities in the Middle East. The submission is in the form of one or two sentences summarizing the item, along with the date, source, and the classification, for example,

> The Egyptian war minister visited naval forces in Alexandria on 11 February. He asked officers and other members of the naval forces to continue their vigilance and to prepare to face any sudden military situation. (Cairo Radio, 12 February 1976, unclassified)

The choice of data is left entirely to the analyst, who is instructed to include anything he considers relevant, and to exclude what can be judged to be irrelevant. There is surprisingly little overlap in what is submitted, with considerable diversity the rule. Later the same day, a coordinator consolidates the items, resolving differences of wording, emphasis, and meaning, and returns the complete list of items to all participants. By the following day, the analysts working individually evaluate the items and return the numerical assessments. Cartography and printing usually take two more days, so the reports are three or more days old when distributed.

One of the central features of our studies is the use of a group of analysts rather than a single expert. This more than anything else influences the data-gathering process, the format of the publication, and the actual production procedure. The reasons for this approach are:

> — to bring to the exercise a range of expertise beyond the experience of any single analyst;
> — to supply a richer mix of evidence on the questions by asking each analyst to contribute anything he or she

considers important. As most political, military, or strategic intelligence problems are reflected in a host of areas, such varied inputs as propaganda analysis, photographic interpretation, and logistic calculations are useful;
— to provide a balance of expertise in which the effects of organizational and individual bias are minimized.

It is an accepted fact that different analysts will tend to place greater reliance on different types and sources of intelligence. The consolidated list of intelligence items, which is circulated to all participants without identifying the contributors, provides an opportunity for each analyst to call an item to the attention of his colleagues. To avoid time-consuming group meetings, the problems of scheduling, and group dynamics effects, each analyst works on the probabilistic assessments individually and relays them to the coordinator.

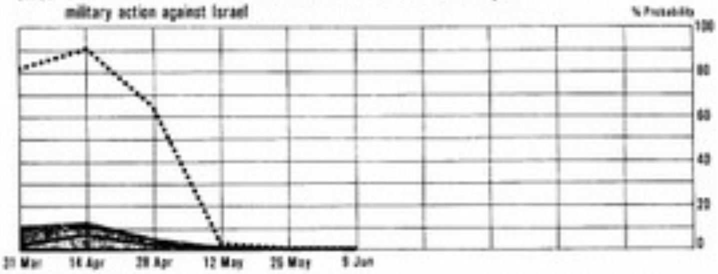
Another feature of the studies is that each periodic report actually contains the intelligence items identified and used by the participants, with only a paragraph or two of composed text on the principal trends during the period. No attempt is made to formulate or coordinate a lengthy textual analysis of the situation. This allows the reader of the reports:

— to see the basic evidence rather than just a summary and hence to understand better the analysts' assessments;
— to make his own direct assessments if he so desires, or just to keep up with the topic by viewing the evidence regularly;
— to maintain a concise chronology of the situation.

The ability to portray the results of the analysis graphically was one of the strongest arguments for using a quantitative method like Bayes, and the graphs in the publication have been well-received. The probabilities of the various types of hostilities (the hypothesized events) are immediately visible on a broken-line chart (Figure 1). This conveys much information at a glance, and seems to represent an advance in communication over traditional methods of reporting, especially in illustrating trends far more concisely and vividly than do words. In addition, the range of estimates around the central measure shows clearly and concisely how much disagreement there is. It is just possible that much of the success of the reports is due more to this informative brevity than to the validity of the estimative technique.

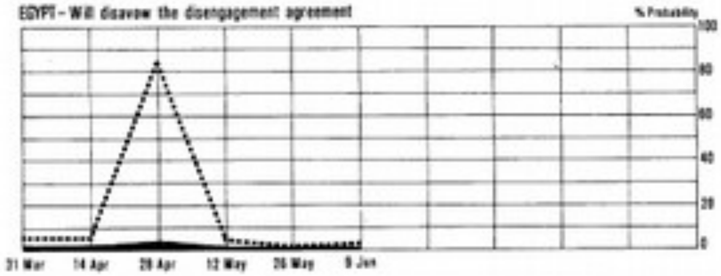
The final technique, also presentational, is the listing of all participants by name and office. This visibility is pleasing to the analysts, who normally endure extreme anonymity in their work. In fact, the interest evinced by both participants and readers over this feature has been so great that a supplementary graphic was developed which allows the participants to be visibly identified with an individual line. This graphic reflects their hunch opinions on a separate but related question (Figure 2). The individual identification is not permitted in the main questions because we have found a slight tendency to want to manipulate the results if an analyst feels uneasy with the direction his assessment is taking. Incidentally, although this chart is totally nonmethodological, it performs well as a barometer of change; once each individual chooses his position on the chart — as a hawk, a dove, or a middle-of-the-roader — the ups and downs are fairly consistent across the board. The trends are clear from month to month, and there is no need for all participants to agree on a single number.

SYRIA – Either alone or in concert with other Arab States – will initiate major military action against Israel



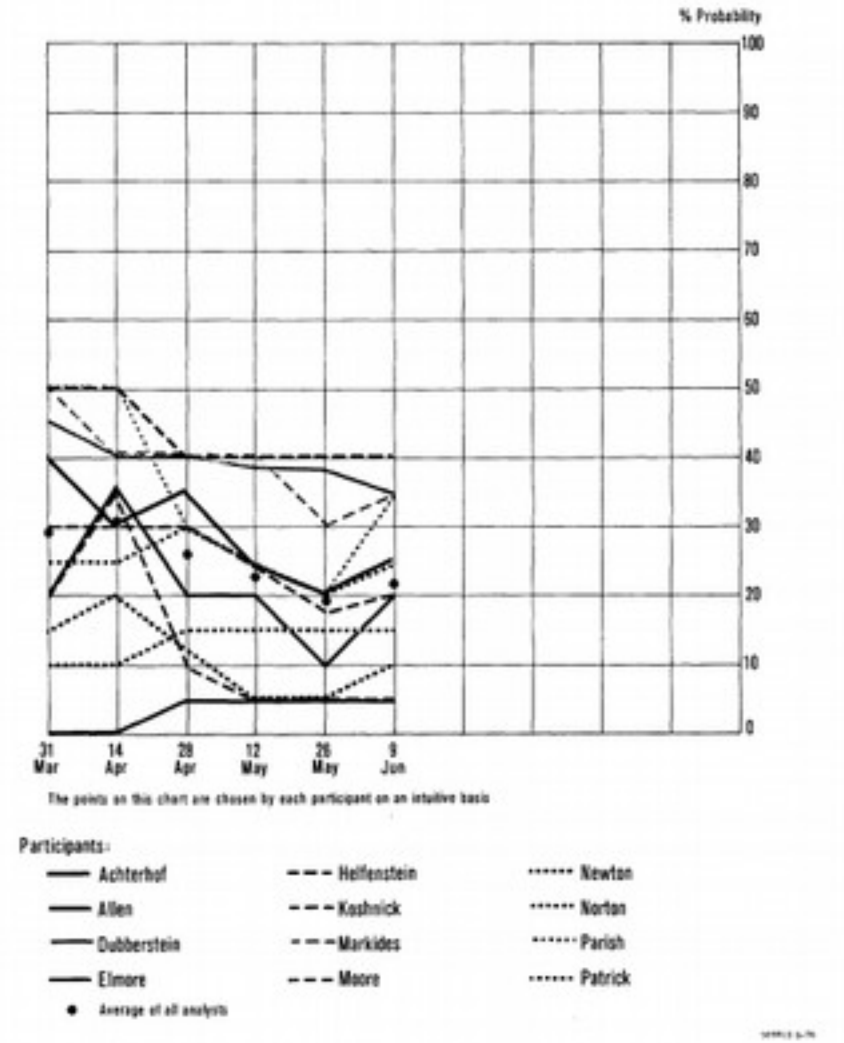
ISRAEL – Will launch an attack against one or more Arab states

EGYPT – Will disavow the disengagement agreement



— — Lowest estimate by any of the analysts
—— Average of all analysts (median)
***** Highest estimate by any of the analysts

Most deeply shaded area shows inter-quartile range of opinion



III. A Middle East Example

The Starting Point

The following description will catalog the questions which were investigated, and illustrate how the probabilities are calculated. The project also required extensive preparation for the proper use of bureaucratic and human resources, but these matters are beyond the scope of this paper.

After much discussion of what questions would be relevant, approachable, and of interest to our government audience, four scenarios or hypotheses were set out:

— No major hostilities are planned by Syria, Israel, or Egypt within 30 days;
— Syria — either alone or in concert with other Arab states — plans to initiate major military action against Israel within 30 days;
— Israel plans to launch an attack against one or more Arab states within 30 days;
— Egypt plans to disavow the disengagement agreement within 30 days.

These four may not be mutually exclusive, but we have treated them as such for the purpose of calculating probabilities. We also know that we have not exhausted all the possibilities. Foreign policy analysts are all too familiar with the words of the elder von Moltke: "Gentleman, I notice that there are always three courses open to the enemy, and that he usually takes the fourth."

At the inception of the exercise, each participating analyst assigned a set of probabilities to these four hypotheses, based upon his understanding of the situation up to that time; these were the best starting estimates available. The sum of the probabilities had to equal 1, or 100%; that is, it was assumed that one of the four had to occur within 30 days. Generally, the hypothesis of no hostilities within 30 days was assigned a probability of from .7 to .95.

Subsequently, these estimates were changed, by assessing the evidence — from open and classified sources — in terms of each of the hypotheses and calculating the new probabilities according to the rule of Bayes. There is an independent set of figures maintained for each analyst, which is charted over time to show changes and trends.

Evaluation of a Sample Intelligence Item

For example, let us simplify the calculations by assuming only two hypotheses, that Israel is planning to launch a major offensive against Syria within 30 days, and that she is not, and further assume that an analyst has assigned the following probabilities to these hypotheses:

Probability that Israel is planning to launch a major offensive against Syria in 30 days — 10% or .1

Probability that Israel is not planning such an offensive — 90% or .9

Also assume that the following item arrives and that the analyst assigns probabilities to such a report surfacing, first assuming that Israel is planning an attack, and second assuming that she is not:

"Israeli Finance Minister Rabinowitz stated that the nation's economic situation is one of war and scarcity, not one of peace and prosperity." (Jerusalem Radio, 20 February, unclassified)

Probability that this would be said if Israel is planning to launch a major offensive against Syria within 30 days — 99% or .99

Probability that this would be said if Israel is not planning such an offensive — 80% or .8

This information can be used to revise the probabilities that each hypothesis is true by using the Bayesian formula. The formula itself and the complete calculations are given in Appendix 1.

Revised probability that Israel is planning to launch a major offensive against Syria within 30 days — .12

Revised probability that Israel is not planning such an offensive — .88

Notice that the two prior probabilities added to 1, or 100%, and that the revised figures also equal 1, even though the conditional event probabilities do not.

As this is a recursive process, in which a succession of events are assessed, the revised probabilities become the prior probabilities for calculating the effect of the next item, and the final set of probabilities for a period become the starting point for the next period's assessments.

IV. Conclusions

Utility as a Predictor of Events

The Bayesian method upon completion results in an archive of evidence, evaluations, and predictions which lend themselves to various forms of evaluation. The main criterion for evaluation is the accuracy of prediction, although this may not be as straightforward as it seems. Because of the myriad variables in the prediction equation, an event may occur which was only ten percent probable the day before, or an event which was scheduled to occur may fail to materialize. Thus there have been times of great uncertainty during our reporting periods when the probability of certain hostilities rose, only to fall back again later. Does this mean that the high probability of the event was somehow in error? Rather it would seem to mean that at the time the event could very well have occurred if other factors had coincided; the evaluation cannot really be considered "wrong."

Generally, our studies have successfully predicted non-events. That is, they showed that the evidence did not support any of the positive hypotheses of hostilities, and none of them in fact occurred during the period studied. In such a case, the point to be noted is how early the evaluations moved away from an indeterminate figure toward a strong probability of no hostilities. It has been our experience that the Bayesian calculations show this movement earlier than the analyst's intuitive judgment would. Until one of the positive hypotheses actually occurs during the course of a Bayesian exercise, it is difficult to know the predictive value of the technique. If such a positive event does take place, it would be possible to conduct a much more searching evaluation. What were the earliest indicators? What evidence was missing, overlooked or misperceived? When did the trend lines signal a significant alteration in the situation? How did the Bayesian assessment compare with other intelligence assessments?

Other Benefits to Participants and Readers

There is no magic and no inherent wisdom in Bayes. In simplest terms, the Bayesian technique consists of a statistical formula and a procedure for its use. It is an organizing device which allows an analyst to use his expert understanding of a situation to assess the likelihood of various hypotheses about an intelligence problem, and to evaluate fragments of evidence in terms of those hypotheses. The Bayesian formula then aggregates those numbers mathematically, rather than by the nonrigorous logic of human induction, into an overall set of probabilities. This has the following advantages:

— More information can be extracted from the available data because the technique allows each piece of evidence, central or marginal, to add its weight to the final assessment in a systematic way; thus, a number of small items can outweigh a large one, and the probabilities are not at the mercy of the most recent or most visible item.
— The procedure provides a reproducible sequence of steps for arriving at the final figures; a disagreement among analysts can thus often be seen to be a disagreement over the meaning of certain items rather than an unresolvable difference of opinion.
— The formulation of the questions forces the analyst to consider alternative explanations of the facts he sees, thus loosening the bonds of established opinions. In other words, he is asked to look at how well the evidence explains hypotheses other than the one he has already decided is most likely.
— The use of quantified judgments allows the results of the analysis to be displayed on a numerical scale, rather than through the use of terms such as "probable," "likely," "unlikely," or that gem "possible." In addition, the work of more than one analyst can be portrayed in graphic form, with ranges and averages.
— The formal procedure has been shown to be less conservative than analysts' informal opinions, and to drive the probabilities away from 50-50 faster and farther than the analysts' overall subjective judgments do. This is often initially unsettling for the analysts, but most have admitted that they later agreed with the assessment.
— The mere fact that a team of experts is asked to assess

periodically the evidence on an important intelligence question provides managers of intelligence production with a degree of assurance that the question is indeed being monitored effectively.

Applicability of the Technique

The starting point for any investigation, whether in intelligence or in an academic setting, must always be the careful formulation of the relevant questions. The Bayesian technique has definite limitations, and it can only be applied where certain criteria are met:

— The question must lend itself to formulation in mutually exclusive categories, such as war versus no war, or the development of a nuclear capability versus no nuclear development. If various overlapping possibilities enter into the picture, such as limited border harassment or the development of a purely peaceful nuclear capability, the results of any Bayesian formulation may be suspect.

— The question must be expressed as a specific set of hypothetical outcomes. The Bayesian approach would be useless as a predictor of "the pattern of future Middle East relations." The question would at the least have to be re-cast in terms of specific alternatives, that is, a set of scenarios of Middle East developments. In this process, however, there would be a danger that the question would be so simplified as to render any answer irrelevant and uninteresting.

— There should be a fairly rich flow of data which are at least peripherally related to the question. For example, for the question of nuclear development, data on all related materials and processes would be relevant. If information is sparse, the technique is very sensitive to each item and may be less reliable.

— The question must revolve around the type of activity which produces preliminary signs and is not largely a chance or random event. For example, it would be fruitless to attempt to predict which military leaders will be in Cairo on a particular day. Bayesian analysis reacts only to preparations for and indicators of the hypothesized outcomes.

The Future of Such Analysis

The Office of Political Research has found this technique to be a useful adjunct to traditional analysis. It is the frequently-voiced opinion of various readers of the Bayesian reports that they are thought-provoking and represent an advance in communications over traditional methods. There is nevertheless a healthy respect and a continued need for the traditional analysis of complex problems which are beyond the limited scope of Bayes. Most of the research and writing in our office, the rest of the Agency, and the rest of the intelligence community will continue in the traditional mode, but we shall supplement it when appropriate with Bayesian analysis and other "new" methods.

Appendix 1. The Statistical Basis for the Technique

The rule of Bayes is a statistical identity, derivable from the laws of intersection of sets and the definitions of conditional probability and mutually exclusive events. Most statistics textbooks contain the derivation, one source being Miller and Freund, *Probability and Statistics for Engineers*, pp. 29-32.

In symbols, the rule is $P(Hi/E) = \frac{P(Hi) \times P(E/Hi)}{\sum_{i=1}^{n} (P(Hi) \times P(E/Hi))}$

where

E is an event, an "item" of intelligence;
H is an hypothesis, a hypothetical cause of events;
Hi is one of a set of n mutually exclusive hypotheses;
P(Hi) is the starting, or "prior," probability of an hypothesis;
P(E/Hi) is the probability of an event given Hi, of an event occurring, given a particular underlying cause;

P(Hi/E) is the probability of an hypothesis given E, the "revised" probability of an hypothesis, given that a particular event has occurred.

In words, it says that,

given an analyst's starting probabilities P(Hi) — his intuitive feeling for the likelihoods of a set of more or less mutually exclusive hypotheses, and

given his assessments P(E/Hi) of how likely an event would be if each of the hypotheses were true, then

the updated version of the hypotheses themselves P(Hi/E) can be calculated in a straightforward fashion.

The procedure is also recursive; if there are more than one event to be assessed, the updated or revised probabilities of the hypotheses from this round become the starting probabilities for the next round.

The example which was given in section III is here shown calculated in full, using the Bayesian formula:

$$P(H1)=.1$$
$$P(H2)=.2$$

$$P(E/H1)=.99$$
$$P(E/H2)=.8$$

$$P(H1/E) = \frac{P(H1) \times P(E/H1)}{\sum_{i=1}^{2} (P(Hi) \times P(E/Hi))} = \frac{.1 \times .99}{.099 + .72} = \frac{.099}{.819} = .12$$

$$P(H2/E) = \frac{P(H2) \times P(E/H2)}{\sum_{i=1}^{2} (P(Hi) \times P(E/Hi))} = \frac{.9 \times .8}{.099 + .72} = \frac{.72}{.819} = .88$$

Appendix 2. Limitations of the Method

Limited Applicability — The first and foremost reservation in the use of this technique, as noted earlier, is that it is applicable only to certain types of questions. They must be capable of definition as a set of fairly distinct outcomes or hypotheses. Also, the procedure involving many analysts, cartographic plates, and finished printing is too cumbersome to use on crisis questions. It is certainly possible, however, for the technique to be further adapted, either through computer assistance in routing and printing, or by eliminating the printing overhead and the complexity of operating with many analysts.

Data Problems — There is the problem of identifying which evidence is relevant; that is, whether certain peripheral items should be included, and, if included, whether they should carry less weight than the other items. We have delegated that decision to the analysts. After all, they are the experts, and their frequent disagreement over items shows that objective measures of relevance would be virtually impossible to devise.

Very little editorial judgment is imposed by the coordinator in the process of consolidating evidence, and any item which appears to be even marginally related is included for evaluation. Nevertheless, each analyst is then allowed to ignore any item he considers irrelevant. This gives the participants great leeway over what they rate, but insures that they at least see the evidence and make an explicit decision on its relevance. Furthermore, if a participant sees two or more items as overlapping, he is asked to rate only one of them.

Related to this is the problem of source reliability. A report may come from an unreliable source, it may be subject to other interpretations, or it may be a misleading fabrication. Although some methodologists have suggested that each analyst assign a numerical measure of source reliability along with each item, to be incorporated into the calculations as a weight, we have avoided placing this extra burden on the analyst by requesting that he internalize this requirement and assign probabilities which reflect how much faith he places in each item. If an analyst understands the process and rates items thoughtfully, he can cause items of greater salience and reliability to have a greater effect on the calculations. This is because the effect of an item increases as the range of probabilities assigned to it increases.

Another related problem is that of negative evidence, or "the dog that barked in the night-time," from "Silver Blaze," a Sherlock Holmes story in which the singular event was that the dog *did not* bark in the night-time. This refers to the fact that the absence of any positive evidence may in itself be highly indicative, and the journalistic bias toward reporting events rather than non-events compounds the situation. That is, we tend to get news only of events or changes, whereas the fact that the status quo is being maintained may be quite significant, and there is no way for the analyst to rate this. We recognize the problem, and feel that it is at least partially solved by including the following item whenever it appears necessary: "How likely is it that only these events would occur (and be seen) if hypothesis 1 is true?", " ... if hypothesis 2 is true," etc.

Problems over Time — There are difficulties in the use of the method in a project continuing over many months. First, the questions probably require some reference to a time period (explicit or implicit) in which the hypotheses are to manifest themselves; that is, whether they will occur within 30 days, or a year, or five years. As a project such as this continues, the timeframe must either contract or move forward. Contraction would occur if there is a fixed date in the future which limits the possibilities, such as the development of a nuclear potential by 1978. In this case, the passage of time and the reduction of the period remaining may itself be of significance, and a coordinator may choose to include an item to that effect for evaluation, Moving the timeframe forward occurs when the question is of the probability of events within the next 30 days, etc. This is looking through a "sliding window," and the approach raises the problem of retaining or discarding data which were evaluated months earlier with regard to an earlier frame of possibilities. Our solution has been to drop earlier evaluations and recalculate the probabilities each time using only a fixed timespan of evidence multiplied against the original intuitive probabilities. The intuitive starting probabilities are also updated at intervals.

Problems with Numbers — There are also two numerical problems. The first is that a probability of zero is unacceptable mathematically, not to mention analytically. If any conditional probability is evaluated at zero, the probability of the related hypothesis becomes zero, and no amount of other evidence can rejuvenate the probability. Thus any evaluation of zero should be replaced by a very small number. The second problem is more profound, being the way individual analysts handle probabilities. It has been our experience that some people think easily in probabilities, others have to work at it every time, and a few need constant attention and retraining to overcome a distorted or unrealistic feeling for probabilities. The only solution for this problem, aside from a careful initial choice and subsequent replacement, is constant attention to the analysts' assessments and frequent retraining using illustrative items of evidence.

Manipulation — Finally, there is the problem of conscious manipulation. An analyst may assign his probabilities in a manner which reflects a pre-determined goal rather than unbiased judgment. Although we have found this to be quite rare, nevertheless it does occur. In our early studies, the participants were identified with the Bayesian trend lines, and there were occasions of manipulation. Avoiding disciplinary solutions, we have almost entirely circumvented the problem by identifying the participants on the supplementary intuitive graphs and not on the Bayesian charts. This allows them to express strongly-held personal opinions in a forum designed for that purpose, and the methodological purity of the Bayesian calculations is increased greatly.

*See Jack Zlotnick, "Bayes Theorem for Intelligence Analysis," and Charles E. Fisk, "Conventional and Bayesian Methods for Intelligence Warning," Studies In Intelligence XVI/2.

~~SECRET~~

Posted: May 08, 2007 08:46 AM