The Sino-Soviet Border Dispute: A Comparison of the Conventional and Bayesian Methods for Intelligence Warning

More on probability — II

Charles E. Fisk

Problems of "indications analysis" or "intelligence warning" are essentially questions of how to assign probabilities to hypotheses of interest. For example, a problem of indications analysis occurred in August 1969 when two hypotheses arose; namely, the conjecture (H_1) that within the next month the USSR would attempt to destroy China's nascent nuclear capabilities, and the alternative hypothesis (H_2) that such an attack would not occur.

A method of indications analysis is a rule for eliciting probability judgments from intelligence analysts, and alternative methods for this purpose have been studied within the Agency since 1967.[1] The usual and most direct method is simply that of asking analysts to make either verbal or numerical probability judgments about hypotheses of interest. As an alternative to the conventional approach, the so-called Bayesian method does not require analysts to assign probabilities to the main hypotheses of interest; instead, analysts are asked to specify values for certain "conditional" probabilities, from which one can infer judgments about the main hypotheses.

It has been argued[2] that the Bayesian method is better than the conventional approach to problems of intelligence warning. This article will illustrate the two alternatives, and will then explain the results of an experiment that was designed to test the assertion of the Bayesian method's superiority.

The Conventional Method of Intelligence Warning

The conventional approach to intelligence warning begins when a set of hypotheses first comes under active scrutiny. For example, during August 1969 several intelligence officers warned that the USSR would probably launch a major attack against China within the next month.

This warning spawned various hypotheses, two of which were (H_1) that the USSR would begin the offensive during September 1969, and (H_2) that there would be no attack.

For a large class of hypotheses, the problem of indications analysis remains essentially the same: certain Agency officials must first elicit from qualified analysts judgments about the hypotheses, and then must synthesize these judgments into a warning. The officials obviously cannot pore over every bit of evidence observed by each analyst, so analysts must focus and summarize their views.

Generally, then, the first step in the conventional method involves the gathering of either verbal or numerical probability estimates. On 30 August 1969, for example, each of six senior analysts from six Agency offices was asked to estimate the probability of the war hypothesis H_1. Their estimates — i.e., values for $P(H_1)$ — appear in Table 1. As time passes, further estimates are elicited, and previous warnings are either amplified or damped on the basis of the new estimates. Clearly, then, a key question is how an official ought to elicit probabilities from analysts. The conventional approach suggests that an official should simply ask analysts to state the probabilities whenever the official wants to reconsider his warning.

Table 1	
Analyst	The Probability of H1 on 30 August 1969*
A	.20
B	.85

C	.40
D	.25
E	.35
F	.20

*The symbol H, denotes the hypothesis that during September 1969, the USSR would launch a nuclear attack against China.

As part of an experiment that was designed to compare the conventional method with an alternative system (the Bayesian) for eliciting probabilities, each of the six analysts mentioned above was asked on 5 September 1969 to re-estimate the probability that a Sino-Soviet war would erupt before 29 September. On 12 September the analysts were asked again, and so forth for each week in September. As a result of this process of questioning, each analyst produced an "intuitive" probability track such as the one shown in Figure 1. Each point on the illustrated track denotes the best probability judgment that Analyst D could offer after reading the all-source intelligence available to him.

On the basis of a considerable amount of research involving simulated questions of intelligence warning, however, Edwards,[3] Zlotnick,[4] and other proponents[5] of the Bayesian method for eliciting probabilities would argue that the sequence of estimates shown in Figure 1 was not the best sequence that Analyst D could have specified. They claim that an official who had asked "the right questions" could have obtained from Analyst D — and from each of the other analysts — a better sequence of probabilities. This alternative method of questioning will be explained in the following section.
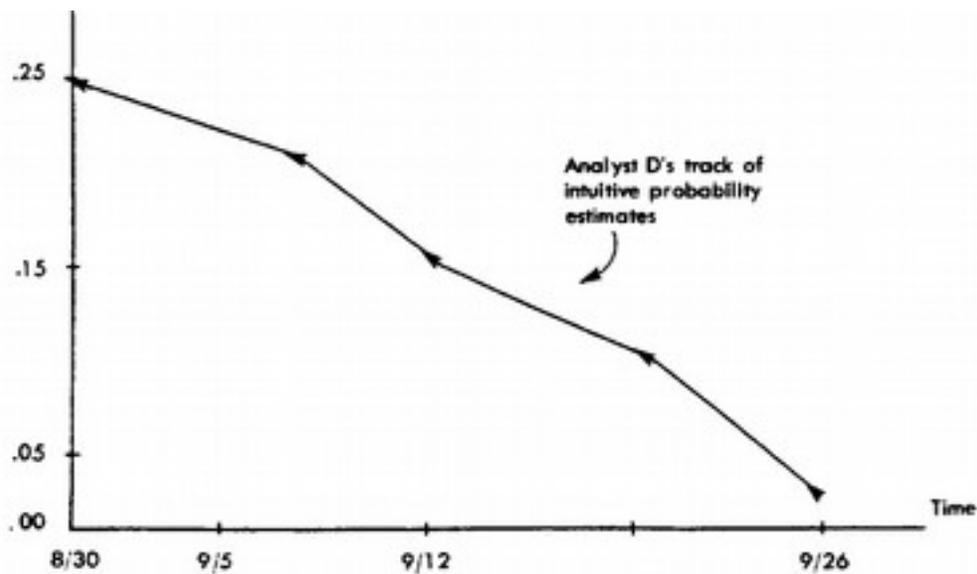
Probabilities Stated by Analyst D

Figure 1.

The Bayesian Method of Intelligence Warning

There is no unique "Bayesian method": dozens of systems, each slightly different from its predecessors, have been proposed and tested on simulated problems of intelligence warning. Most of these systems, however, involve substantially similar steps. The steps taken in the Sino-Soviet Experiment to obtain from each of the six analysts a Bayesian track that could be compared with the analyst's intuitive track are as follows:

(a) On 30 August 1969 each of the six analysts was asked to estimate a value for $P(H_1)$, which at that time denoted the probability that the war hypothesis HI was true. This first step duplicated the first step in the conventional method discussed above, so each analyst's estimate for $P(H_1)$ appeared as in Table 1.

(b) In contrast to the conventional method, on 5 September the Bayesian approach did not require the analysts to re-estimate $P(H_1)$. Instead, each analyst was asked to list the major events whose occurrence during the previous week had influenced his opinion about the war hypothesis. For example, during the week

Analyst D might have observed that no men in the Soviet reserve army had been called for active duty. This event of "no calls" could have been denoted by E_1. And, since Analyst D might have believed that a call-up during the previous week would precede the event of a Soviet attack in September, the event E_1 might have lowered his intuitive probability judgment concerning the chance of war. Similarly, E_2 might have denoted the event of no increases in Soviet propaganda against the Chinese, and so forth for other events that an analyst might have thought relevant to the war hypothesis H_1.

(c) A majority of the analysts listed virtually the same set of relevant events, although some analysts' views had been influenced by events that other analysts had not listed. From the separate lists, a master event list was compiled, such that the events E_1, E_2, ... on the master list exhibited two properties; namely, (i) each event proposed by each analyst was reflected in the master list; and (ii) each master event was, roughly speaking, independent of each other master events.[6]

(d) When the master list had been compiled on 5 September 1969, some of the analysts asserted that certain events suggested by other analysts had not actually occurred. Such differences over raw intelligence were recorded as each analyst estimated a probability of occurrence for each of the events E_1, E_2, ... on the master list.

(e) In addition to specifying probabilities of occurrence, each analyst estimated various conditional probabilities on 5 September. For example, with respect to the event no reserve calls during E_1 of the previous week, Analyst D was asked to specify a value for $P(E_1 \mid H_1)$, which denotes the probability that E1 would have occurred, given the assumption that the war hypothesis (H_1) was true. Moreover, Analyst D was asked to estimate $P(E_1 \mid H_2)$, the probability of E_1 on the assumption that the no-war hypothesis (H_2) was true. For each of the other events on the master list Analyst D specified a similar set of conditional probabilities, as did each of the other analysts.

(f) A modified version of Bayes' Theorem was then used on 5 September 1969 to calculate for each analyst a "revised" probability of war.[7] This probability was called an analyst's Bayesian estimate, and was plotted on the same graph as his intuitive probability. Thus for Analyst D in particular, on 5 September 1969 the two probability tracks shown in Figure 2 had

been obtained — one track by the conventional method, and one by the Bayesian approach.

(g) On 12 September 1969 the Bayesian procedure outlined above was repeated, with the exception that the "prior" probabilities used in the revision process were the Bayesian probabilities of war that had been obtained on 5 September 1969. Thus after two weeks, a typical analyst's probability tracks appeared as in Figure 3.

(h) After the Bayesian procedure had been repeated at weekly intervals during September, the Bayesian tracks derived from conditional probabilities specified by Analysts A, B, and D appeared as in Figure 4. The Bayesian and intuitive tracks compiled for Analysts C, E, and F resembled the tracks shown for A and D, in the sense that for five of the six analysts, the Bayesian track always fell below the intuitive track.

A Criterion for Comparing Probability Estimates

A criterion for comparing methods of probability elicitation can be illustrated with reference to Figure 3. In retrospect, we know that the hypothesis H_1 was false: Russia did not attack China. Thus if an analyst's "probability tracks" had actually appeared as in Figure 3, then on 12 September 1969 an official would have acted more wisely on the basis of the Bayesian sequence of estimates. In other words, if one had been forced to gamble according to either the Bayesian or the intuitive tracks shown in Figure 3, one would in retrospect have preferred the Bayesian sequence.
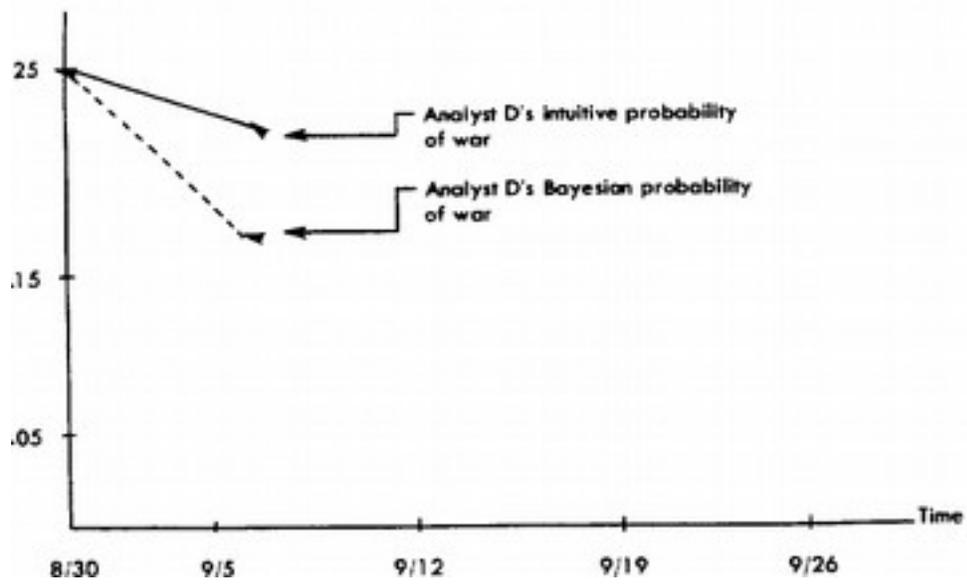
Probabilities Stated by Analyst D

Figure 2.

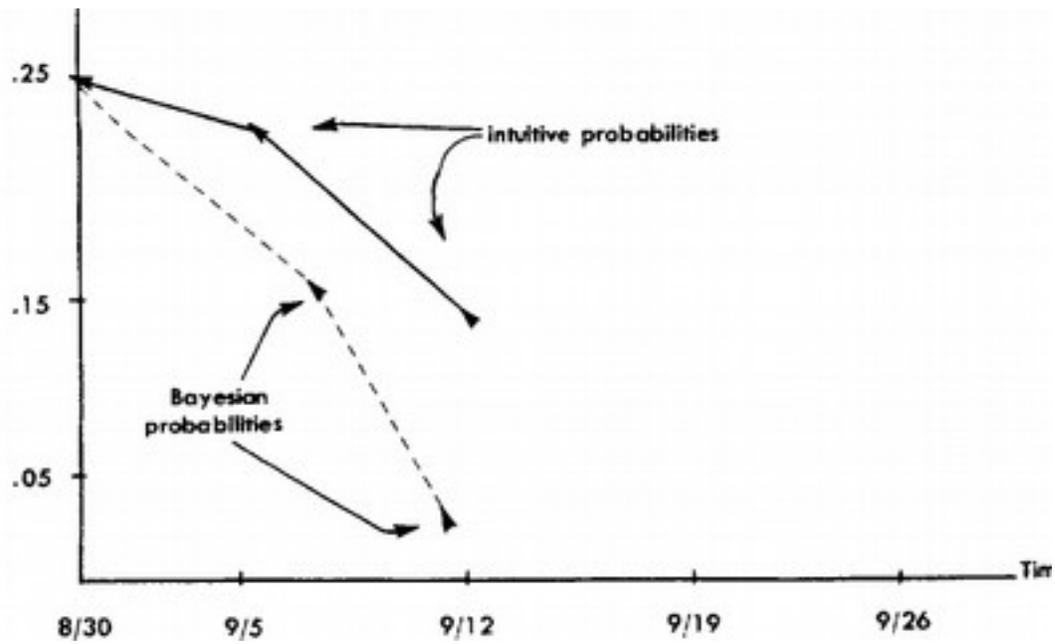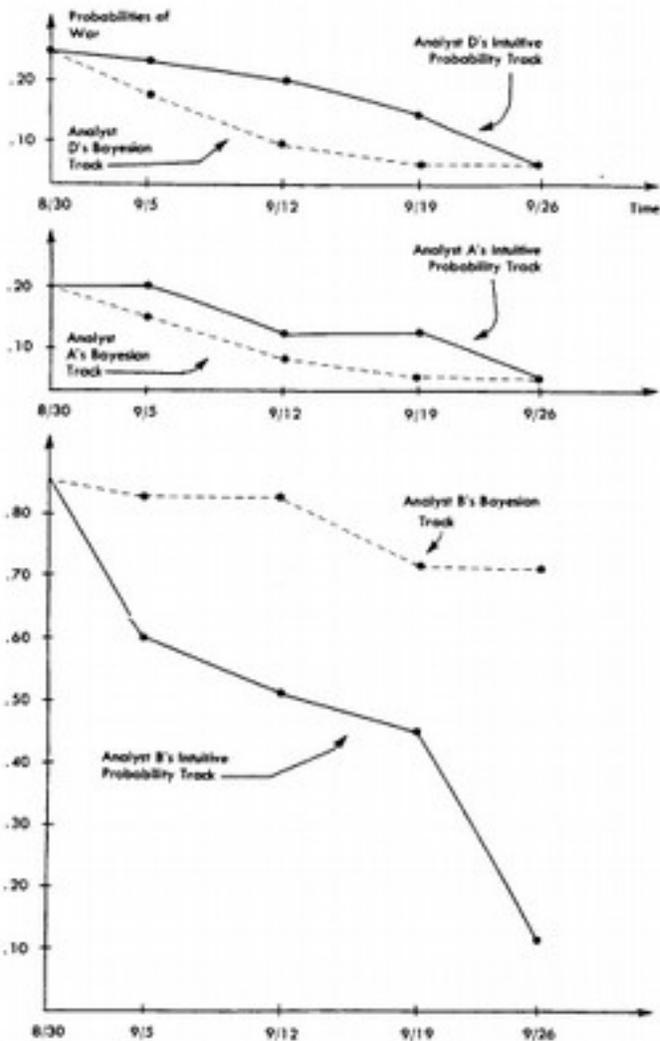*Probabilities Stated by Analyst D*

Figure 3.

Figure 4.

Of course, if Russia had attacked China, and if a typical analyst's probability tracks had appeared as in Figure 3, then one would have preferred to have acted according to the analyst's intuitive track. But according to the advocates of Bayesian analysis, such a preference for an intuitive track will seldom occur: if Russia had attacked, then — according to the Bayesian proponents — prior to the attack the Bayesian track for a typical analyst would have been above his intuitive track, such that in retrospect the Bayesian method would again have been preferred. As is evident in Figure 4, Analyst B proved to be an exception to this assertion: his Bayesian track always fell above his sequence of intuitive estimates.

This criterion of "retrospective superiority" has served as the basis for dozens of experiments[8] in which researchers have compared the Bayesian method with alternative techniques for eliciting probabilities, and in most cases the Bayesian approach has triumphed. But there is

no firmly established analytical justification for the method. Bayes' Theorem is a mathematical truism, but there are no axioms from which one can infer that repeated applications of the theorem to conditional probabilities specified by analysts will yield superior intelligence warnings. Thus, in the fall of 1969, it was of considerable interest to review the Bayesian method's effectiveness in the context of the actual intelligence problem posed by the chance of a Sino-Soviet war.

The Sino-Soviet Experiment

As explained above, the six analysts met at weekly intervals during September 1969 in order to re-estimate the probability of the war hypothesis H_1, and to specify the conditional probability estimates that were processed according to the Bayesian method. In October 1969 (when the war hypothesis H_1 was known to have been false) the probability tracks derived from the two methods were compared as in Figure 4. The primary result was that for five of the six analysts, the Bayesian track had always been below the intuitive sequence of probabilities. Thus in retrospect, an official would have preferred to have acted according to the Bayesian estimates, rather than according to the analysts' best intuitive judgments concerning the war hypothesis.

An Evaluation of the Bayesian Method

Several results of general interest emerged from the Sino-Soviet experiment. First of all, when the experiment began the analysts differed widely in their views concerning the chance of a war; but the reasons for their differences were murky at best.

A typical argument between two analysts would arise when one would accuse the other of having ignored certain crucial facts in estimating the likelihood of war. The accused would then respond that he had indeed considered all relevant information, and that his estimate was based on facts that other analysts had overlooked. Such arguments were difficult

to evaluate, since there was no record of who had considered what, or of how each analyst's probability estimate had evolved over time.

Once the Sino-Soviet experiment had begun, however, one could easily determine the relative importance that an analyst had assigned to any given event. For example, it was evident from Analyst B's conditional probability estimates that he had considered the event of Kosygin's visit in September 1969 to Peking as being irrelevant to the war hypothesis. In contrast, Analyst E had regarded the meeting as a profound indicator that war would not occur. The issue of whether Analyst B exercised good judgment in this respect remains an open question; but at least his assessment of the Peking trip had been recorded and could be evaluated.

Thus the Bayesian approach provided a kind of accounting system for intelligence analysis. If such a system were implemented for other questions of indications analysis, a significant class of disagreements among analysts might be resolved. And to the extent that such disagreements would persist, an official who must synthesize warnings on the basis of analysts' estimates could discern and evaluate causes for the disagreements.

A second contribution of the accounting system was the fact that after the system's inception, the analysts definitely did consider the same relevant events. In particular, Analyst E wrote the following review of the experiment.

> In the case of Office E, interchanges with other offices are usually on an unsystematic ad hoc basis. The Bayesian experiment afforded an opportunity to bring these interchanges into focus on a systematic basis. Its particular merit lies in the manner in which participants are led to identify the factors influencing their estimates and to present these for critical review by others approaching the question from varying angles. I would emphasize the value of focus, though perhaps no less valuable is the exposure of participants to lines of analysis — as one analyst noted — of which they are dimly if at all aware.

Similarly, Analyst C wrote:

The meeting was a useful forum for the interplay of ideas and the exchange of information which might otherwise not occur. Interchanges would take place in the absence of such a meeting; but they would be limited because of their bilateral nature (in most cases).

In summary, an improved system of accounting for analytical judgments is needed. Although it cannot be said categorically that the Bayesian method excels as a forecasting device, the Sino-Soviet experiment indicates that it might provide a means for such accounting.

Footnotes

1 Two examples of these studies are *A Mathematical Model for Intelligence Warning* (Intelligence Report No. 1396/67, November 1967), and *Bayes' Theorem in the Korean War* (Intelligence Report No. 0605/68, July 1968). For references to various studies done outside the Agency, see *A Bibliography of Research on Behavioral Decision Processes* by Ward Edwards (University of Michigan, Human Performance Center, Memorandum Report No. 7, January 1969).

2 A detailed exposition of this argument is offered by Ward Edwards et al in "Probabilistic Information Processing Systems: Design and Evaluation," *IEEE Transactions on Systems Science and Cybernetics* (Vo. SSC-4, No. 3) September 1968. Further expositions have been put forth by Jack Zlotnick in "A Theorem for Prediction," *Studies in Intelligence* (Vol. 11, No. 4) Fall 1967.

3 Ibid.

4 *Ibid.*

5 See the bibliography cited in Footnote 1.

6 The notion of independence can be illustrated as follows: suppose that Analyst D has listed the event of "a high-level diplomatic probe by the USSR to ascertain probable US reactions to a Sino-Soviet war," while

Analyst E has listed "a war-related contact between US and Soviet officials." These two events clearly refer to the same thing, so the master list would contain only one event referring to a diplomatic probe. In some cases, however, the two properties of inclusiveness and independence were difficult to achieve in compiling the master list.

7 This method of calculating revised probabilities is sometimes called a "roll-back" procedure. See *Applied Statistical Decision Theory* by H. Raiffa and R. Schlaifer (Harvard Business School, Division of Research, 1961).

8 See the references cited above on page 53.

Posted: May 08, 2007 08:33 AM