J. R. Statist. Soc. B (1992)
54, *No.* 1, *pp.* 285–299

Modelling Probabilistic Agent Opinion

By MIKE WEST† and JO CROSSE

Duke University, Durham, USA

[Received August 1989. Revised November 1990]

SUMMARY

A single agent (an individual, expert or model) provides a decision maker with probabilistic information partially or completely describing the agent's opinion about a collection of uncertain quantities or events. This paper discusses ways in which the decision maker may model the agent's opinion to provide rules for updating his own probability for a related event or random quantity of particular interest. Concepts discussed include the relevance of the agent's information and experience, the accord between the agent and decision maker in terms of common or conflicting information and calibration of probability assessments. New theory develops and extends that of Genest and Schervish, requiring only a partial specification of the decision maker's prior over the agent's opinion. Several illustrative examples are developed.

Keywords: AGENT OPINION ANALYSIS; BAYESIAN UPDATING; CALIBRATION; COMBINING PROBABILITY FORECASTS; EXPERT OPINION

1. INTRODUCTION

Applied inference typically involves the use of information from a variety of sources that may be quite disparate in nature, including data from separate surveys, observational or experimental studies, inferences from formal or informal statistical analyses used in different subareas of a problem by different investigators under possibly quite widely varying circumstances, and purely subjective judgments of ostensibly informed individuals, or experts. Problems of considerable practical importance to personal and corporate decision makers rely on the coherent and efficient combination and synthesis of such information sources, and this leads to the need for appropriate statistical methodology for such synthesis. Several researchers have written on approaches to various problems in this area, notably on the analysis of 'expert' opinion and the assessment and combination of forecasts from collections of individuals and models (Dawid, 1987; French, 1980, 1985; Genest and Schervish, 1985; Genest and Zidek, 1986; Lindley, 1983, 1985, 1988; Morris, 1983; Schervish, 1984; West, 1984, 1988, 1992; Winkler, 1981). The general problem area includes more specific issues such as calibration and adjustment of subjective judgments (DeGroot and Fienberg, 1983; French, 1986; Lindley, 1982; Schervish, 1984), data and forecast aggregation (West and Harrison, 1989), group decision-making (Genest and Zidek (1986) and references therein, and West (1984)) and overlapping and dependent information sources (Lindley, 1988; Winkler, 1981).

This paper describes new and general results for the analysis of agent opinion by a decision maker. The term 'agent' is used to denote any individual, group of

†*Address for correspondence*: Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27706, USA.

 0035-9246/92/54285 $2.00

individuals acting collectively in their inferences, statistical (or other) model or generally any process by which uncertain inferences are generated. Such an agent provides inferences about uncertain quantities of interest to the decision maker. The problems of agent opinion analysis concern the assessment by the decision maker of inferences received from one or more agents and ways in which the decision maker may use such inferences in revising his beliefs about uncertain quantities of interest. Attention is restricted to a single agent, the development being based on generalization and extension of a result of Genest and Schervish (1985).

Initially, attention is restricted to a single event. It is convenient to suppose that I am the decision maker. A fundamental concept in such problems is that any form of inferential information provided by the agent be viewed as data informing on the quantity of interest. Typical of models developed from this viewpoint are French (1980) and Lindley (1988), in which I have access to the agent's probability distribution for the event; denote the event by E, with complement $\bar{E}$. In medical diagnosis, E may indicate the presence of a disease, with the agent an informed clinician or technician, or a diagnostic test. In other examples, such as time series forecasting, E is an unrealized future outcome and the agent an informed forecaster or forecasting model. My prior probability is $p=P[E]$. The problem of agent opinion analysis may be summarized as follows:

> I am to learn the agent's probability forecast f for E and may use it in forming a revised view about E. How should I do this?

Formally, I must determine my posterior probability $p_f=P[E|f]$. In practice, this is only needed for the particular f observed, thus providing the solution to my problem in that particular case, but a general model requires that it be specified as a function of all possible values of f.

Two distinct approaches to calculating the required posterior appear in the above references. Firstly, I could directly model p_f as a function of f and of my prior probability p. For example, I could simply take a weighted average $p_f = \omega f + (1-\omega)p$ and, consequently, $P[\bar{E}|f] = 1 - p_f = \omega(1-f) + (1-\omega)(1-p)$. The weight ω reflects my view about the relevance of the agent's forecast and the information on which it is based, and would depend on various features of the particular problem and past experience with the agent. If the agent has been successful in forecasting similar events in the past, giving eventual outcomes high probability, then I am likely to weight the agent's opinion highly. The second, alternative, approach has more in common with standard statistical modelling, with f being viewed as data informing on E. Bayes's theorem provides the required posterior probabilities; namely, for all f $(0 \leqslant f \leqslant 1)$, $p_f \propto p\, p(f|E)$ and $1 - p_f \propto (1-p)\,p(f|\bar{E})$. Modelling is involved in the terms $p(f|E)$ and $p(f|\bar{E})$, my subjective probability densities for the random quantity f $(0 \leqslant f \leqslant 1)$, conditional on the occurrence or non-occurrence of E. Together with p, this provides one way of specifying my views about how the agent's stated probabilities relate to E, based on any information I have about the agent's performance in forecasting in the past with similar or related events, and the relevance of the agent's available information in the context of the current events. These densities may also depend on my prior probability p, relating the agent's probability to my own.

Models of this sort are discussed by many of the earlier referenced authors (e.g. French, Lindley, Schervish, Winkler, West). They follow the usual prior-to-posterior approach to computing the conditional probabilities $p(E|f)$ and $p(\bar{E}|f)$ of the full joint distribution over E and f. The modelling focus is on the specification of the necessary conditional distributions for $(f|E)$ and $(f|\bar{E})$. Any model should allow the decision maker to express beliefs about the expertise, honesty and calibration of the agent, and about the relationships between the decision maker and the agent. The problems of assessing and quantifying suitable model forms are usually very difficult. See comments in Genest and Schervish (1985) and Lindley (1985, 1988) for a thorough discussion. Genest and Schervish (1985) recognized these difficulties and developed what may be referred to as a robust Bayesian approach based on a partial specification of the distribution of f and E. They proved a remarkable result that provides the starting point for development here. Their result is as follows.

Theorem 1 (theorem 2.1 of Genest and Schervish (1985)). Suppose that I specify my joint prior distribution over $\{E, f\}$ only partially, providing values of $p = P[E]$ and my prior mean for f, namely $\mu = E[f]$, and assume that the marginal distribution for f has full support [0, 1]. Then my posterior probability $p_f = P[E|f]$ has the form of a linear function of f,

$$p_f = p + \lambda(f - \mu), \tag{1}$$

for some constant λ depending on p and μ but not on f.

This simple result must be interpreted with care. In particular, it does not imply that p_f is of the stated form for any particular joint prior that happens to be consistent with the given values of p and μ (as simple examples will show). Rather, the interpretation reads as follows: for any particular joint distribution let $g(f) = P[E|f]$ be the deducible conditional probability for E, for any f in [0, 1]. For many such joint distributions, $g(f)$ will be a complicated non-linear function of f, as in Lindley's models (Lindley (1988), for example). Though not explicitly noted in their paper, the result of Genest and Schervish (1985) takes account of uncertainty about the functional form $g(\ )$ that results from the partial specification of the joint model. Without further specifications, my posterior probability p_f is the expectation, conditional on f, of $g(f)$, namely $p_f = E[g(f)|f]$, where the expectation is now with respect to my posterior distribution of the function $g(\ )$ given f. It transpires, as proven indirectly by Genest and Schervish, that this posterior expectation, over all possible functional forms consistent with a joint prior having the stated values of p and μ, is simply a linear function. What the result does not provide is the value of the linear coefficient λ.

This result, remarkable for its robustness and simplicity, is far reaching. It is reinterpreted and illustrated later and extended to problems in which the agent provides collections of probabilities for events related to E, such as discrete distribution functions for related variables. Further extensions provide novel updating results for problems in which the event E is replaced by a (discrete or) continuous random quantity, and the agent provides information partially or completely describing his distribution for that, or other, related random quantities.

2. MODELLING AGENT OPINION

In the Genest and Schervish result summarized in theorem 1, consider the choice of the linear coefficient λ. Genest and Schervish note that I may choose λ by directly assessing p_f at one particular value of f, requiring that I decide, in advance, on my posterior probability were I to learn that the agent's took that particular value. This is perhaps most easily considered by taking f to the extreme of 0 or 1, considering how I would react were I to learn that the agent actually believed in E or $\bar{E}$. At $f=1$, $p_1 = p+\lambda(1-\mu)$. Fixing p_1 implies $\lambda = (p_1-p)/(1-\mu)$. Similarly, $\lambda = (p-p_0)/\mu$. The full support assumption for my prior for f implies that μ lies strictly between 0 and 1, so these values are well defined. Also, the argument can be followed by using any other two values of f, such as values close to, though not necessarily equal to, 0 and 1 (i.e. p_ϵ and $p_{1-\epsilon}$ for small positive ϵ), to determine the result. If p and μ are specified, direct assessment of both p_0 and p_1 provides a coherency check, calculating λ in two different ways. A practically important point, not noted by Genest and Schervish, is that I may feel it easier to assess directly my responses p_0 and p_1 to extreme agent probabilities (or to p_ϵ and $p_{1-\epsilon}$ for small positive ϵ) than to assess the prior expectation μ. If this is so (as I believe is often the case), then the problem can be reformulated, as follows.

Define probabilities α and β via

$$\begin{aligned} \alpha &= p_1 = P[E|f=1], \\ \beta &= 1 - p_0 = P[\bar{E}|f=0]. \end{aligned} \tag{2}$$

Suppose that α and β are specified directly. It is easily deduced that $\mu = (p-p_0)/(p_1-p_0) = (p+\beta-1)/(\alpha+\beta-1)$, and $\lambda = p_1 - p_0 = \alpha + \beta - 1$. As a result, the updating equation (1) may be rewritten as follows.

Corollary 1. My posterior probability has the alternative representation

$$p_f = \alpha f + (1-\beta)(1-f). \tag{3}$$

Genest and Schervish require p, μ and some means of determining λ, whereas equation (3) requires only the probabilities α and β. Furthermore, since $p = E[p_f]$, the expectation being with respect to my prior for f, then $p = \alpha\mu + (1-\beta)(1-\mu)$. Given p, α and β, this relation allows μ to be determined.

More generally, Genest and Schervish prove a similar result to the effect that, given only p and the expectation $\mu = E[\pi(f)]$ for a known and bounded function $\pi(f)$ of the random quantity f, then, again, $p_f = p + \lambda\{\pi(f)-\mu\}$ for some constant λ (not necessarily the same as that above). This has an obvious application to cases in which I may initially transform f, using historical data on observed events and the associated agent forecasts to (frequency-) recalibrate f (e.g. French (1986) and West and Mortera (1987); see also the comments in Section 3). $\pi(f)$ is then a specified (frequency) recalibration function, and the updating formula may be rewritten as $p_f = p + \lambda\{\pi(f)-\mu\} = \alpha\pi(f) + (1-\beta)\{1-\pi(f)\}$, where now $\mu = E[\pi(f)]$, $\alpha = P[E|\pi(f)=1]$ and $\beta = P[\bar{E}|\pi(f)=0]$. My posterior probability is the prior value p corrected by a term proportional to $\pi(f)-\mu$, the difference between the observed value of $\pi(f)$ and its prior expectation μ. The coefficient $\lambda=\alpha+\beta-1$ takes larger values when $\alpha+\beta$ is large. If $\pi(f)$ exceeds expectation, the agent viewing the event more likely than expected, my probability increases if, and only if, $\alpha > 1 - \beta$ or, naturally enough, $P[E|f=1] > P[E|f=0]$. Otherwise, it decreases.

Here is an example in which the above ingredients μ, α and β are assessed through reference to information available in the context of the problem. In particular, this involves consideration of relevant information that the agent and I have in common, and of the particular features of a problem about which the agent's information is deemed informative.

Example 1. Let E indicate the occurrence of a disease in a given patient, $\bar{E}$ indicating absence of the disease. Suppose a particular symptom, or class of symptoms or test results, is used in diagnosis; occurrence of the symptoms, denoted by S, indicates that the patient belongs to a high risk group, whereas $\bar{S}$ indicates membership of a low risk group. I determine that the patient belongs to a population among which the disease is evidenced with chances $P[E|S] = 0.35$ and $P[E|\bar{S}] = 0.05$; thus, given the risk group classification S or $\bar{S}$, residual uncertainty is described by the rates of 35% for high risk patients and 5% for low risk patients. Suppose further that the symptoms are evident in 20% of the population, so that $P[S] = 0.2$ and $P[\bar{S}] = 0.8$. My prior for E is now implicitly determined; it follows immediately that $p = 0.35 \times 0.2 + 0.05 \times 0.8 = 0.11$, an overall disease incidence rate of 11%. Were I to learn the risk group classification, this would update to either 0.35 (under S) or 0.05 (under $\bar{S}$). This risk group classification represents ideal, but currently unattainable, information.

I am to consult the agent, an informed clinician, whose judgments I believe are also largely based on S, and I assume that his views are relevant only in so far as they provide me with additional information about S. Thus I assume that E given S (or $\bar{S}$) is conditionally independent of f, or $P[E|S, f] = P[E|S] = 0.35$ and $P[E|\bar{S}, f] = P[E|\bar{S}] = 0.05$, for all f. I can then write

$$p_f = P[E|S,f]\ P[S|f] + P[E|\bar{S},f]\ P[\bar{S}|f] = 0.35P[S|f] + 0.05P[\bar{S}|f]. \quad (4)$$

Following discussion with the agent, I hold the view that his experience is with similar patients but from a more disease-prone subset of the population, the incidence rates being 60% for high risk (S) individuals and 5% for low risk ($\bar{S}$). With the agent's probabilities denoted by $P_A[\ ]$ (and restricting the use of unsubscripted $P[\ ]$ for my own probabilities), this means that $P_A[E|S] = 0.6$ and $P_A[E|\bar{S}] = 0.05$. Thus, because of different information, I believe the agent to rather overestimate the strength of S as an indicator of the disease for the current patient, although his probabilities are acceptable for patients in the population subset of his experience. Finally, the agent agrees with me on the incidence rate of S, $P_A[S] = P[S] = 0.2$.

Thus I expect the agent to hold the view that the incidence rate of the disease is

$$P_A[E|S]\ P_A[S] + P_A[E|\bar{S}]\ P_A[\bar{S}] = 0.6 \times 0.2 + 0.05 \times 0.8 = 0.16.$$

Assuming that the agent's probability assessment will honestly and accurately reflect his experience and information (with no need for recalibration), this is my expected value for the forecast, $\mu = 0.16$.

Consider now my assessment of $\alpha = P[E|f=1]$. From equation (4), substituting $f=1$ leads to $\alpha = 0.35\ P[S|f=1] + 0.05\ P[\bar{S}|f=1]$. Since we agree on the prior probability of S, $P[S] = P_A[S] = 0.2$, and since $\{f=1\}$ is equivalent to the occurrence E in the agent's opinion, then, applying Bayes's theorem, $P[S|f=1] \propto 0.2\ P_A[E|S] = 0.2 \times 0.6 = 0.12$ and, similarly, $P[\bar{S}|f=1] \propto 0.8\ P_A[E|S] = 0.8 \times 0.05 = 0.04$; after normalization, $P[S|f=1] = 0.75$. Applying the same argument to $(S|f=0)$ leads directly to $P[S|f=0] = 0.095$. Substituting these values in equation (4) then gives $\alpha = p_1 = 0.35 \times 0.75 + 0.05 \times 0.25 = 0.275$ and $\beta = 1 - p_0 =$

$1-(0.35\times0.095+0.05\times0.905) = 0.922$. (As a consistency check, equation (3) gives $p = P_\mu = 0.275 \times 0.16 + 0.078 \times 0.84 = 0.11$, as required.)

Then, on learning the agent's prediction f for disease of the patient, my updated view is given by corollary 1 as $p_f = 0.195f + 0.08$. The extremes are $p_0 = 0.08$ if the agent perfectly classifies the patient as healthy and $p_1=0.275$ if he classifies the patient as diseased. (These probabilities are now marginal with respect to the risk group classification S ($\overline{S}$); were S, or $\overline{S}$, to become known, the agent's forecast would be irrelevant under the above assumptions, my posterior probability being given by 0.35 to 0.05 respectively.)

In a second example, and in contrast with the first, μ and λ are directly assessed by reference to my opinion about the agent's likely information sources.

Example 2. Suppose that I model E as one of a sequence of Bernouilli trials with uncertain success probability θ, and my prior distribution for θ is beta, $\theta \sim \mathscr{B}\,[cp; c(1-p)]$, with mean p and variance $p(1-p)/(c+1)$, for some $c > 0$. Then $E[P[E|\theta]] = E[\theta] = p$. Were I to learn θ, then the agent's opinion would be irrelevant. His views are of interest only in so far as they provide information about the success probability θ.

Suppose that I further believe the agent's opinion to be largely based on his having observed a sequence of such trials in the past, independent of information in my experience, the (uncertain) number of trials being k, and having some (uncertain) number of successes y. Then, if I assume that the effects of any other agent information are negligible, standard Bayesian analysis of Bernouilli trials data leads me to believe the agent's prior for θ to be approximately beta, $\mathscr{B}\,[y; k-y]$. Hence I hold the view that the agent's forecast for E will be $f = P_A[E] = E[P_A[E]\theta]] = y/k$. Suppose that I further believe this experience to be similar to my own in the sense that $\mu = E[f] = p$.

Under the above assumption that our information sets are independent, full knowledge of the agent's data $\{y, k\}$ would lead me to update my prior for θ directly to the posterior $(\theta|y,k) \sim \mathscr{B}\,[cp+y; c(1-p) + k - y]$, or $\mathscr{B}\,[cp + kf; c(1-p) + k(1-f)]$, having posterior mean $P[E|y, k] = E[\theta|y,k] = (cp+kf)/(c+k)$. Of course, I learn only f, not y and k; taking expectations with respect to my beliefs about the uncertain number of trials k in the agent's experience, and noting that $p=\mu$, it follows that $p_f = p + \lambda(f-\mu)$ where $\lambda = E[k/(c+k)|f]$. Generally, knowledge of f would be uninformative about the precision k of the agent's information, so that $\lambda = E[k/(c+k)]$, the expectation being with respect to my prior for k. If I knew k, then λ simply reflects the precision of the agent's information relative to my own. If I believe the agent to have a large amount of experience (in terms of independent trials) relative to my own prior information, then λ will be near 1. If, however, I have rather precise prior information about θ with c large relative to my prior expectation of k, then λ is small and the agent's opinion only marginally affects my view of E. The agent's opinion is naturally more highly relevant in cases when I am vague about θ.

3. GENERAL COMMENTS

Before proceeding to develop and extend these basic results, some discussion of general issues is given in this simple framework.

Opinion models should allow the expression of a range of (my) beliefs about the relevance of the agent's information and experience, and how such experience relates to the current problem of interest. This will be judged on the basis of information about past experience with similar and related forecasting problems, consideration of just how closely the agent's past relates to the present context and judgments about the extent and nature of relevant information that we share. Generally, I should more highly value the opinions of an agent whose past accords closely with the environment and context of the current problem than otherwise. In medical diagnosis, for example, I will judge such accord on the basis of past experience of the agent in diagnosis with similar patients having similar symptoms. If the current case is similar in these terms to cases in the past experience of the agent, and the agent is experienced in such diagnoses, then I will tend to weight the agent's forecast positively. However, it may be that the current problem differs, in my opinion, from such past experiences of the agent. For example, I may have additional relevant data, such as other test results and background clinical information, that I believe the agent does not share. Under such circumstances, the agent's opinion should have less influence on my posterior views. The construction through probabilities α and β allows flexibility here, as illustrated in examples 1 and 2 earlier.

Decision makers must consider potential overstatement or understatement of probabilities by agents, whether due to actual biases, attempts to mislead or persuade, perceptual problems or inexperience in probability assessment. These issues are really separate from those concerning subject matter relevance, involving notions of calibration. Schervish (1984) remarks on this point, stating that

> 'One problem . . . is that of separating the judgement of *calibration* for an expert from a determination of the *dependence* between the expert and the decision maker, or more appropriately, the information held in common by the decision maker and expert'.

Of the various notions of calibration, the empirical concept of frequency calibration (e.g. DeGroot and Fienberg (1983) and French (1986)) is perhaps the most widely discussed. Consider the diagnosis example. Suppose that the agent has past experience in assessing similar events—diagnosing this particular disease—under similar conditions over a (long) period of time. Suppose also that, for all probabilities f (or, at least, all those ever stated by the agent), the relative frequency of cases on which the events occurred is some number $\pi(f)$. Then the agent is said to be frequency calibrated with respect to such events if $\pi(f)=f$. Otherwise, the agent is mis-calibrated. Of what relevance is this in the agent opinion problem or, more generally, to a Bayesian decision maker? Well,

(a) if I accept that the event is suitably similar to events in the past experience of the agent and that his probability assessment is similarly obtained (with no attempt to misquote f to mislead, for example) and
(b) if I know $\pi(f)$ for all f,

then, on learning f, I should transform directly to the recalibrated probability $\pi(f)$. This has been mentioned in Section 2. In practice, I may be prepared to assume that (a) applies, at least approximately, to a past sequence of events, and then $\pi(f)$ can be approximately assessed or estimated on the basis of the observed forecasts and outcomes over such a past sequence. Viewed as a function of f, $\pi(f)$ provides a (non-linear and possibly non-monotonic) map from the stated forecast to a calibrated scale;

$\pi(f)$ is a (frequency) recalibration function (e.g. West and Mortera (1987)). Although the requirements in (a) may be difficult to satisfy in any given problem, it is reasonable to attempt to recalibrate forecasts based on any relevant historical information about performance and also to account for any anticipated biases, such as over extreme assessments of probabilities due to inexperience in elicitation, and so forth. This ties in directly with the usual notion of frequency calibration, quite separately from issues of subject matter expertise. The recalibrated probability $\pi(f)$ is now to be used in place of f in updating my view about E, and the more general result of Genest and Schervish applies.

The model allows for extreme agent opinion, permitting and appropriately accommodating zero probabilities. The agent may state either $f=0$ or $f=1$ (perhaps following recalibration). Such statements are made in practice, and my reactions to them depend on what I believe to lie behind such extreme judgments. Some possibilities (not exhaustive), each requiring a different response, are that

(a) the agent may know the outcome, or believe (perhaps mistakenly) the same,
(b) the agent may be constrained to provide 0–1 forecasts, as with witnesses in certain courts, or be reluctant or unable to consider more refined judgments on a probability scale, or
(c) 0 or 1 may result from probability assessment on a crude numerical scale, probabilities being quoted to the first decimal place.

The resulting extreme posterior probabilities are α and $1-\beta$. Dawid (1987) discusses related ideas. There the agent is a witness in court and E corresponds to guilt of a defendant, $\bar{E}$ denoting innocence. $f=1$ is the stated belief of the agent that the defendant is guilty, $f=0$ that the defendant is innocent. Dawid is concerned with cases in which the witness holds a degenerate opinion, stating either $f=0$ or $f=1$.

On the subject of expertise, suppose that I believe the agent to be an expert in the area under study, the current problem relating closely to those in his experience. Suppose also that he has all the relevant information available to me and has appropriately utilized it in forming inferential statements. It is then appropriate to set $\alpha=\beta=1$, defining the notion of personal judgment as to what makes the agent a subject matter expert. As a consequence, my posterior probability is $p_f=\pi(f)$, the recalibrated version of f. If, in addition, he is expert in probability assessment (and honest etc.) then $\pi(f)=f$ and I simply accept the stated opinion. This agrees with the definition of probability calibration in Lindley (1982), discussed by Schervish (1984). Notice how the two concepts, expertise and calibration, are neatly separated in the updating; the former is measured by α and β, and corresponds to the linear form of p_f as a function of $\pi(f)$; the latter is responsible for any non-linearity as a function of f. Of course, I cannot simultaneously view two or more agents to be probability (re)calibrated, in this sense, unless I believe that their probabilities will coincide; any model for more than one agent should recognize this.

Finally, there is an obvious opportunity for formal learning about the probabilities α and β, and the calibration function $\pi(f)$, if relevant past data are available. $\pi(f)$ should model any non-linearity of p_f as a function of f. Otherwise, p_f will be a linear function of f. West and Mortera (1987) consider issues related to model-based assessment of $\pi(f)$ in forecasting binary time series. There E represents the outcome of a single observation in a time series, and the approach is taken in the context of a subject matter expert agent, corresponding to $\alpha=\beta=1$. Thus, given a sequence of outcomes and forecasts, $\pi(f)$ may be estimated. Given information about the probability calibration characteristics of an agent, an estimated calibration function may then be assumed in the general problem with α and β not 1, representing limits on the agent's subject matter expertise. With suitable data in any specific problem, α and β may be estimated too, although this is not explicitly considered further here. Related issues are developed in West (1992).

4. DISCRETE AGENT PROBABILITIES

Suppose now that the agent's opinion comes in the form of $n \geqslant 1$ probabilities for exclusive events possibly related to E, denoted by $E_1, \ldots, E_n$. The sample spaces of E and the E_i may be quite different. Define $E_{n+1} = \overline{(\cup_{i=1}^n E_i)}$ so that the events $E_1, \ldots, E_{n+1}$ form a partition. The agents states probabilities f_i for E_i $(i=1, \ldots, n)$, with $f_{n+1} = 1 - \Sigma_{i=1}^n f_i$.

Example 3. As a specific example that provides context and is of major interest later, let Y be a real-valued random quantity and E be the event $\{Y \leqslant y\}$ for some known number y. For a given partition of the real line defined by intervals $[q_{i-1}, q_i)$ $(i=1, \ldots, n+1)$, for some known numbers $q_0 = -\infty < q_1 < \ldots < q_n < q_{n+1} = \infty$, define with $E_i = \{q_{i-1} \leqslant Y < q_i\}$. Then f_i is the agent's probability that $Y \in [q_{i-1}, q_i)$. If the agent is assumed to have a continuous distribution for Y, then the q_i provide a collection of quantiles under that distribution, a partial specification.

Generally, $\mathbf{f} = (f_1, \ldots, f_n)$ is viewed by me, the decision maker, as the realized values of random quantities in the simplex $S_n = \{\mathbf{f}: 0 \leqslant f_i \leqslant 1, \Sigma_{i=1}^n f_i \leqslant 1\}$. Following Genest and Schervish (1985), suppose that I specify a model for E and $\mathbf{f}$ jointly only partially, providing only the values $p = P[E]$ and the prior expectations $\boldsymbol{\mu} = E[\mathbf{f}]$, an n-vector in S_n. As with Genest and Schervish (1985), these prior expectations are assumed consistent with my full joint distribution, and though only a partial specification of this joint distribution turn out to be sufficient to determine the form of my posterior probability $p_{\mathbf{f}} = P[E|\mathbf{f}]$ as a function of $\mathbf{f}$. The derivation of this key result is based on results in Genest and Schervish (1985), involving only a very minor change to their theorem 3.2. The key mathematical ingredients are provided in lemma 1 below, and the import of the result in theorem 2. Although the mathematics of the results is essentially just that of Genest and Schervish, the statistical meaning of the results is rather different. Genest and Schervish are concerned with n agents each providing a single probability for the common event A; we are concerned with a single agent providing n probabilities for exclusive events related to E.

Lemma 1. Let S_n be the simplex $S_n = \{\mathbf{f}: 0 \leqslant f_i \leqslant 1, \Sigma_{i=1}^n f_i \leqslant 1\}$. Let $\boldsymbol{\mu}$ be a fixed vector in S_n, and let $\Delta\boldsymbol{\mu}$ denote all n-dimensional distribution functions G with support S_n and having mean vector $\boldsymbol{\mu}$. If k is a real-valued Lebesgue measurable function on S_n such that

$$\int_{S_n} k(\mathbf{f})\, \mathrm{d}G(\mathbf{f}) = p \qquad \text{for all } \mathbf{f} \in \Delta\boldsymbol{\mu},$$

then $k(\mathbf{f}) = p + \Sigma_{i=1}^n \lambda_i (f_i - \mu_i)$ for some real numbers λ_i, $i=1, \ldots, n$.

Proof. The result is essentially as in lemma 3.1 in Genest and Schervish, except that the support here is S_n and not $[0, 1]^n$. Clearly, $S_n \subset [0, 1]^n$, so the proof from

Genest and Schervish (1985) should apply to this restricted space. In their conclusion, Genest and Schervish also point out that their lemmas and theorems apply equally well to restricted spaces. All that is needed therefore is to check that the proof of lemma 3.1 in Genest and Schervish (1985) applies with the restriction. This is easily seen to be the case. Their result applies for $\mathbf{f}$ defined on the n-dimensional unit cube, and we are simply restricting attention to those having zero mass outside the n-dimensional simplex S_n. Since Genest and Schervish (1985) do not rely on the functions having non-zero support over the entire n-dimensional cube, it only remains to show that the restricted space is non-empty. Dirichlet distributions are obvious examples that satisfy the requirements of their theorem proof, and the result holds.

Using this lemma, the following result is immediate and is analogous to theorem 3.2 of Genest and Schervish (1985).

Theorem 2 (application of Genest and Schervish (1985)). With initial specifications $p = P[E]$ and $\boldsymbol{\mu} = E[\mathbf{f}]$ consistent with a complete joint distribution for $\mathbf{f}$ and E, and assuming that the distribution of $\mathbf{f}$ has support S_n, my posterior probability $p_{\mathbf{f}} = P[E|\mathbf{f}]$ has the form

$$p_{\mathbf{f}} = p + \sum_{i=1}^{n} \lambda_i(f_i - \mu_i)$$

for some λ_i $(i = 1, \ldots, n)$, depending on p and $\boldsymbol{\mu}$ but not on $\mathbf{f}$.

The posterior probability is thus a linear combination of p and the agent's probabilities, with a weight λ_i on the agent's probability f_i. These weights determine the extent and nature of the effect of the agent's views on mine. Their values remain unknown until further features of my distribution over $\mathbf{f}$ and E are specified. The following result provides reinterpretation and a route to such a specification.

Theorem 3. In the framework of theorem 2, the posterior $p_{\mathbf{f}}$ may be written as

$$p_{\mathbf{f}} = \sum_{i=1}^{n+1} \Pi_i f_i$$

where $\Pi_1, \ldots, \Pi_{n+1}$ are the conditional probabilities $\Pi_i = P[E|f_i = 1]$ $(i = 1, \ldots, n+1)$.

Proof. Writing $\delta = \Sigma_{i=1}^{n} \lambda_i \mu_i$, we have $p_{\mathbf{f}} = p - \delta + \Sigma_{i=1}^{n} \lambda_i f_i$. Now, for $i = 1, \ldots, n$, let $\mathbf{e}_i$ be the vector $\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ where the unit element is in the ith position. Then $\mathbf{f} = \mathbf{e}_i$ if the agent is of the opinion that E_i holds, so that $f_i = 1$ and $f_j = 0$ for $i \neq j$. Defining probabilities $\Pi_i = P[E|\mathbf{f} = \mathbf{e}_i] = P[E|f_i = 1]$, it follows that, for each $i = 1, \ldots, n$, $\Pi_i = p^*(\mathbf{e}_i) = p - \delta + \lambda_i$ or $\lambda_i = \Pi_i - (p - \delta)$. Also, $\mathbf{f} = \mathbf{0}$ implies that the agent believes E_{n+1} to hold so that $\Pi_{n+1} = p^*(\mathbf{0}) = p - \delta$. Then, for $i = 1, \ldots, n$, $\lambda_i = \Pi_i - \Pi_{n+1}$. Substituting these expressions for the λ_i in the formula for $p_{\mathbf{f}}$ and replacing $p - \delta$ with Π_{n+1} leads to

$$p_{\mathbf{f}} = \sum_{i=1}^{n} \Pi_i f_i + \Pi_{n+1}\left(1 - \sum_{i=1}^{n} f_i\right)$$

and the result follows.

The result shows how the weights λ_i, determining my posterior probability, may be related to the extreme probabilities Π_i that I would adopt were I to learn that the agent believed one of the events E_i to hold with probability 1. Discussion of this representation is given below, in application and extension to problems where E is just one of many events defining the decision maker's distribution function for an uncertain quantity, and the E_i similarly relate to a distribution function of the agent for a related random quantity.

5. AGENT DISTRIBUTION FUNCTIONS

5.1. *General Framework*

Suppose the event E to be a cumulative probability under my distribution for some random quantity Y. Y is real valued, discrete or continuous (or mixed), and my distribution function is denoted by $P(y)$ $(-\infty < y < \infty)$. The event E is defined by $E = \{Y \leqslant y\}$ for some specified value y. The E_i remain a set of exclusive and exhaustive related events in some sample space. Of particular, though not exclusive, interest are cases when the E_i partition the sample space of Y, so that $\mathbf{f}$ (partially) summarizes the agent's own distribution for Y (West, 1988).

Example 4. Y may be the value of a financial indicator, such as a particular exchange rate, at some specified future time, say one month hence. With $n=1$, E_1 may represent the occurrence of a particular economic event, such as a change in interest rates, or a political announcement that I expect to influence financial markets. I consult the agent about the likelihood of this event and use the stated opinion to revise my views indirectly about whether or not Y will exceed the specified threshold y.

Example 5. Generalizing example 4, suppose that I consult the agent for his opinion about the future level X of a related indicator that I believe to be of use in forecasting Y (i.e. an independent or predictor variable). I request agent probabilities for the event $E_i = \{q_{i-1} \leqslant X < q_i\}$ $(i=1, \ldots, n+1)$, for some chosen points $q_0 = -\infty < q_1 < \ldots < q_n < q_{n+1} = \infty$. The data received from the agent are the probabilities $f_i = F(q_i) - F(q_{i-1})$ $(i=1, \ldots, n+1)$ under his forecast distribution $F(\ )$ for X.

Example 6. In the framework of example 5, suppose that $X=Y$, so that I consult the agent for his opinion about the future level Y directly. This is a more typical opinion analysis problem, similar to those considered in Lindley (1988) and West (1988). In the general framework of Section 4, we have $E = \{Y \leqslant y\}$ and $E_i = \{q_{i-1} \leqslant Y < q_i\}$ $(i=1, \ldots, n+1)$.

In each example, theorems 2 and 3 apply directly. To calculate $p_{\mathbf{f}}$, I must specify prior expectations of the f_i, in addition to p. In example 4, $\mu=\mu_1$ is my prior expectation of the agent's probability of the related economic event or political announcement. In examples 5 and 6, the μ_i are the expected values of the probabilities that the agent will assign to the intervals $[q_{i-1}, q_i)$. By varying the specified quantity y that determines the event E, I can completely determine my posterior distribution function for the random quantity Y based on prior information plus the agent's probabilities $\mathbf{f}$. We have the following result, deduced directly from theorems 2 and 3.

Theorem 4. My posterior distribution function for Y given $\mathbf{f}$, namely $P(y|\mathbf{f}) = \Pr[Y \leqslant y|\mathbf{f}]$ $(-\infty < y < \infty)$, is given by

$$P(y|\mathbf{f}) = P(y) + \sum_{i=1}^{n} \lambda_i(y)(f_i - \mu_i)$$

for some quantities $\lambda_i(y)$ depending on y. An equivalent representation is

$$P(y|\mathbf{f}) = \sum_{i=1}^{n+1} \Pi_i(y) f_i$$

where, for each $i = 1, \ldots, n+1$, $\Pi_i(y) = \Pr[Y \leqslant y | f_i = 1]$.

Calculation of the posterior distribution requires preliminary specification of the quantities $\lambda_i(y)$ or, equivalently, the $\Pi_i(y)$. In terms of the latter, $\Pi_i(y)$ is my posterior probability that $Y \leqslant y$ were I to learn that the agent believed E_i to hold. In example 6, for example, this is the probability that $Y \leqslant y$ were the agent to believe that $q_{i-1} \leqslant Y < q_i$; compare Dawid (1987). These quantities may be more easily appreciated and assessed in cases when the underlying random quantity Y is continuous, the results then being essentially special cases of those in the following section.

5.2. *Agent Distribution Function and Limiting Results*

Suppose that Y is a continuous random quantity and that my prior distribution function $P(y)$ is continuous with respect to the Lesbesgue measure, having density function $p(y)$ $(-\infty < y < \infty)$. Consider agent information provided in terms of a collection of percentage points from the agent's distribution $F(x)$ for a related random quantity X; suppose that $F(x)$ is continuous with density function $f(x)$ $(-\infty < x < \infty)$. This is just as in example 5 with additional assumptions of continuity. The agent's information set is denoted by H_n, given by

$$H_n = \{f_1, \ldots, f_{n+1}: f_i = F(q_i) - F(q_{i-1}), i = 1, \ldots, n+1\} \tag{5}$$

for some specified quantities $q_0 = -\infty < q_1 < \ldots < q_n < q_{n+1} = \infty$. As in West (1988), we assume the q_i to be given and the model is based on my view of the f_i as random quantities given the q_i; further discussion of this issue appears in West (1992).

We can now apply theorem 4. I must specify my prior expectation $\mu_i = E[f_i]$ for each i. Given the assumed additional structure, that the f_i are derived from the underlying, unknown distribution function $F(\ )$ of the agent, it is necessary for consistency that the μ_i be similarly defined. Suppose, therefore, that the μ_i are determined as

$$\mu_i = M(q_i) - M(q_{i-1}) \qquad (i = 1, \ldots, n+1), \tag{6}$$

where $M(x)$ is a known distribution function specified by me, the decision maker. This is simply my prior expectation of the agent's distribution function $F(x)$; for all x, $E[F(x)] = M(x)$, where the expectation is taken with respect to my prior distribution over the uncertain distribution function $F(\ )$. This expectation is all that I am required to specify at this stage; in particular, a completely specified prior over distribution function space is not necessary. Only a collection of probabilities under $M(\ )$, sufficient to determine the μ_i in equation (6), is really required. However, we intend to examine the consequences of this specification for any n and any set of points q_i and consistency is maintained by imposing equation (6) for a prespecified $M(\ )$. Since $F(\ )$ is continuous, then $M(\ )$ is too, having a density function $m(\ )$. Specifying $m(\ )$ directly leads to the μ_i via

$$\mu_i = \int_{q_{i-1}}^{q_i} m(x)\,\mathrm{d}x.$$

Now, theorem 4 applies for any n and any collection of quantiles q_i for the given value of n. Once the f_i have been stated by the agent, the information set H_n is observed and my resulting posterior distribution is obtained. We shall work now with the more interpretable expression in terms of the probabilities $\Pi_i(y)$. If these are specified directly, then the updating problem is solved without further inputs from me as the decision maker. This raises the question of assessment, in particular, of how the $\Pi_i(y)$ are related, which is addressed in the next theorem.

For any given n, the information set H_n can be viewed as providing a discrete approximation to the true agent distribution function $F(\;)$. If n is large and the q_i distinct, then the $\Pi_i(y)$ may provide a rather good discrete representation of $F(\;)$. If n were to increase without bound while the q_i remain distinct and become everywhere dense, then H_n would approach the limiting information set

$$H = \lim_{n\to\infty} (H_n) = \{F(x),\ -\infty < x < \infty\}. \tag{7}$$

Thus, using a limiting argument, we can address the problem of how I update on learning $F(\;)$, or equivalently, the density $f(\;)$. The following theorem provides the result.

Theorem 5. Under the continuity assumptions of this section, suppose that I learn the full agent distribution function $F(x)$, obtaining information H in equation (7). Then my posterior distribution for $(Y|H)$ is given by

$$P(y|H) = \int_{-\infty}^{\infty} \Pi(y|x)\,f(x)\,\mathrm{d}x \qquad (-\infty < y < \infty),$$

where, for each x, $\Pi(y|x)$ is my posterior distribution were I to learn that the agent believed the random quantity X to take the value x with probability 1.

Proof. Choose any sequence of information sets H_n $(n=1,2,\ldots,)$ such that $-\infty = q_0 < q_1 < \ldots < q_n < q_{n+1} = \infty$ for all n. Then $H = \lim_{n\to\infty}(H_n)$ with these specified quantiles determining H_n in equation (5). By the mean value theorem, and for each i, we have $f_i = f(x_i)\delta_i$, where $\delta_i = q_i - q_{i-1}$ and x_i is some number between q_{i-1} and q_i. As $n \to \infty$ under the assumptions of the theorem, the intervals $[q_{i-1}, q_i)$ collapse about a point since $|q_i - q_{i-1}| \to 0$, and so the probabilities $\Pi_i(y)$ have limiting values $\Pi(y|q_i)$, namely my posterior probability for $Y \leqslant y$ were the agent to state that $X = q_i$ with probability 1. Hence, applying theorem 4,

$$P(y|\mathbf{f}) = \sum_{i=1}^{n+1} \Pi_i(y)\,f(x_i)\delta_i \to \int_{-\infty}^{\infty} \Pi(y|x)\,f(x)\,\mathrm{d}x,$$

as $n \to \infty$.

The result here can be alternatively expressed as

$$P(y|H) = P(y) + \int_{-\infty}^{\infty} \Pi(y|x)\{f(x) - m(x)\}\,\mathrm{d}x;$$

the posterior is the prior plus a correction term that, for each y, is an average distance measure between the observed agent density $f(\ )$ and the expected version $m(\ )$. In terms of densities, theorem 5 may be written as

$$p(y|H) = \int_{-\infty}^{\infty} \pi(y|x)\, f(x)\, \mathrm{d}x \qquad (-\infty < y < \infty),$$

where, for each x, $\pi(y|x)$ is the density of the distribution $\Pi(y|x)$. This is my posterior distribution were I to learn that the agent believed the random quantity X to take the value x with probability 1. This applies whatever X may be. In particular, if $Y=X$, $\Pi(y|x)$ is my posterior for Y at the point y were I to learn that the agent's distribution for Y was degenerate at the point x. Though such states of information, involving agent distributions that are degenerate, are usually hypothetical, consideration of just how I would adjust my views about Y in the light of such extreme agent information is not typically a difficult exercise and, if performed, provides the posterior based on any agent distribution by applying theorem 5. An example illustrates how this may be done simply by using familiar parametric families of distributions.

Example 5. Consider the case in which $X=Y$, the agent providing a forecast distribution $F(\ )$ for the quantity Y. Suppose that my prior for Y is normal with known mean a and variance A, denoted by $Y \sim \mathcal{N}[a; A]$. Suppose that I expect the agent to hold similar views, sharing similar information and background experiences, and processing such information in similar ways, and also that the stated distributional information accurately and honestly represents the agent's true opinions. As a result, I take $M(\ )$ as $\mathcal{N}[a; A]$ too. One possible distribution $\Pi(y|x)$ consistent with these priors is the normal $(Y|x) \sim \mathcal{N}[a+r(x-a); (1-r^2)A]$ for some correlation r. If $r > 0$, the agent is viewed as in positive accord with the decision maker, so that were the agent to believe in $Y=x$ with probability 1 the above posterior represents a shift in location to higher values than the the prior mean a if x exceeds a, otherwise to lower values. The magnitude of r is a measure of my belief in the expertise of the agent, the posterior variance $(1-r^2)A$ decreasing towards 0 as r increases. At $r=1$, the posterior degenerates at $Y=x$ reflecting a belief that the agent is expert; in such a case, if the agent were to believe $Y=x$, then so would I on learning x in this case.

Suppose now that the agent provides the information set H in equation (7), the density function $f(\ )$. The following points are immediate from theorem 5. Firstly, if $f(\ )$ has mean c, whatever the global form may be, then my posterior mean is just $E[Y|H] = a + r(c-a)$, a linear function of the two point estimates a and c of Y. Secondly, if $f(\ )$ has mean c and variance C, whatever the global form may be, then my posterior variance is just $V[Y|H] = (1-r^2)A + r^2C$, a linear function of the two variances A and C. Obviously, a normal agent distribution having these moments c and C implies that $p(Y|H)$ is normal with the above mean and variance. This is a special case of examples in West (1992), where further development and variants on this and other examples are given.

ACKNOWLEDGEMENTS

The comments of the reviewers on previous versions of this paper led to substantial revisions. Mike West acknowledges support of the National Science Foundation, under grant DMS-8903842. Jo Crosse acknowledges support from the Science and Engineering Research Council, under a research studentship.

REFERENCES

Dawid, A. P. (1987) The difficulty about conjunction. *Statistician*, **36**, 91–98.

DeGroot, M. H. and Fienberg, S. E. (1983) The comparison and evaluation of forecasters. *Statistician*, **32**, 14–22.

French, S. (1980) Updating of belief in the light of someone else's opinion. *J. R. Statist. Soc.* A, **143**, 43–48.

——(1985) Group consensus probability distributions: a critical survey. In *Bayesian Statistics 2* (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith). Amsterdam: North-Holland.

——(1986) Calibration and the expert problem. *Mangmt Sci.*, **32**, 315–321.

Genest, C. and Schervish, M. J. (1985) Modelling expert judgements for Bayesian updating. *Ann. Statist.*, **13**, 1198–1212.

Genest, C. and Zidek, J. V. (1986) Combining probability distributions: a critique and an annotated bibliography. *Statist. Sci.*, **1**, 114–148.

Lindley, D. V. (1982) The improvements of probability judgements. *J. R. Statist. Soc.* A, **145**, 117–126.

——(1983) Reconciliation of probability distributions. *Ops Res.*, **31**, 806–886.

——(1985) Reconciliation of discrete probability distributions. In *Bayesian Statistics 2* (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith). Amsterdam: North-Holland.

——(1988) The use of probability statements. In *Accelerated Life Tests and Experts' Opinion in Reliability* (eds C. A. Clarotti and D. V. Lindley). Amsterdam: North-Holland.

Morris, P. A. (1983) An axiomatic approach to expert resolution. *Mangmt Sci.*, **29**, 24–32.

Schervish, M. J. (1984) Combining expert judgements. *Technical Report 294*. Department of Statistics, Carnegie Mellon University, Pittsburgh.

West, M. (1984) Bayesian aggregation. *J. R. Statist. Soc.* A, **147**, 600–607.

——(1988) Modelling expert opinion. In *Bayesian Statistics 3* (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith). Oxford: Oxford University Press.

——(1992) Modelling agent forecast distributions. *J. R. Statist. Soc.* B, **54** in the press.

West, M. and Harrison, P. J. (1989) *Bayesian Forecasting and Dynamic Models*, ch. 15. New York: Springer.

West, M. and Mortera, J. (1987) Bayesian models and methods for binary time series. In *Probability and Bayesian Statistics* (ed. R. Viertl). New York: Plenum.

Winkler, R. L. (1981) Combining probability distributions from dependent information sources. *Mangmt Sci.*, **27**, 479–488.