PROBABILITY FORECASTING AND MODELLING EXPERT OPINION

by

Mike West
University of Warwick

Technical Report #87-3

Department of Statistics
Purdue University

February 1987

Invited paper for the Third International Meeting on Bayesian Statistics in Altea, Spain, June 1987. Written whilst visiting Department of Statistics, Purdue University, February 1987.

◇

ABSTRACT

I consider problems of Bayesian information processing in which data consist of forecasts from individuals, "experts" or models. The basic concepts have been developed and extensively used by Lindley in his works on reconciliation of probabilities. Here a new class of models is introduced to deal with expert distributions of essentially arbitrary form, although the focus is on continuous distributions. The models provide methods for processing forecast information in terms of full, continuous distributions or densities, and partial information in terms of collections of quartiles. The latter use of such models is appropriate in contexts where forecasts are given in terms of simple point forecasts, with or without uncertainty measures, or histograms. The models are illustrated in special, practically useful cases.

1. INTRODUCTION

In a series of papers, most recently Lindley (1987) and references therein, Lindley has identified and developed the basic ingredients of the Bayesian approach to information processing when the information obtained consists of the statements of individuals. As a simple example, suppose I am considering the purchase of pesetas for a trip to Spain and my decision to buy or not today depends primarily on what the exchange rate is likely to be at some future time, say the day before I leave. Denote this uncertain quantity by Y. I have a view about Y and also consult a colleague who provides me with his forecast distribution for Y. This provides me with additional information that I should treat as data, processing it in more or less standard ways, to obtain my revised beliefs about the exchange rate. In order to do this I require a probability model for the stated distribution of my colleague, the "expert" providing his opinion in this example, conditional on each possible future value of Y. This probability model provides the likelihood for Y used in updating my prior opinion, via Bayes' Theorem, to process the expert information. The development of appropriate models is the central, technical problem in this area, and the focus of this paper. Note that the same principles apply to a variety of problems involving the assessment and use of information from forecasting models and bureaux, and other sources.

Lindley's models provide for cases in which the random quantity of interest, Y, is discrete. In this paper, the focus is on the wider class of problems in which expert, or other, opinion may be obtained about continuous random quantities in terms of:

(a) fully specified distribution or density functions;

(b) point estimates, such as medians, alone;

(c) collections of percentiles, such as median and quartiles, or deciles;

(d) histograms as discrete approximations to continuous distributions.

Relative to full information on the expert distribution, cases (b), (c) and (d) represent partial knowledge. It is clearly vital in practice that such cases be considered. It is common practice in some areas of forecasting, for example, for simple point forecasts, with or without uncertainty measures, to be quoted with no reference to a global forecast distribution. In addition, it is often (or rather, always) difficult to elicit a full distribution with which an expert is totally comfortable, whereas a small collection of quantiles may be perfectly acceptable as a partial description.

In Section 2, the case of an event indicator Y is considered, and concepts underlying the basic approach developed by Lindley outlined. Even in that case, there is a need for partial information processing, such as with upper and lower bounds on expert probabilities. Section 3 develops the fundamental model for predictive distributions. A key ingredient is the focus on the quantile function of the expert as data, rather than on the distribution function directly. This model is shown to provide simple, interpretable likelihoods for the quantity Y based on expert information provided in any of the forms above. Cases (b), (c) and (d) are considered together in Section 4, that of full information (a) in Section 5. Some final discussion and examples are given in Section 6.

2. BASIC CONCEPTS IN THE EVENT CASE

To fix ideas suppose Y is binary and that my prior probability that $Y = 1$ is p. My consulted expert provides the probability f which comprises my only additional piece of information, $H = \{f\}$. My problem is to construct the model defining the density or mass function $p(H|Y)$, for each possible value $Y = 0$ or 1. Lindley's models (Lindley, 1987) suppose that $p(H|Y) = p(f|Y)$ is the density of the random quantity f following a logistic normal distribution. Another possibility is Beta, such as

$$(f|Y) \sim B[\delta\alpha_Y, \delta(1-\alpha_Y)], \qquad (Y = 0, 1),$$

where $\delta > 0$ is a precision parameter and, for each Y, $\alpha_Y = E[f|Y]$ is my expectation of the expert's forecast. Clearly I view the expert to positively accord with reality if $\alpha_1 > \alpha_0$, and that expertise increases with $\alpha_1 - \alpha_0$. This model provides densities

$$p(f|Y) = \frac{\Gamma(\delta)}{\Gamma(\delta\alpha_Y)\Gamma(\delta[1-\alpha_Y])} f^{\delta\alpha_Y - 1}(1-f)^{\delta(1-\alpha_Y)-1},$$

for $0 < f < 1$, that form the likelihood function for updating to my posterior probability $p^* = P(Y = 1|H) = P(Y = 1|f)$. On the log-odds scale, routine calculations lead to

$$\log\left(\frac{p^*}{1-p^*}\right) = \log\left(\frac{p}{1-p}\right) + k + \delta(\alpha_1 - \alpha_0)\log\left(\frac{f}{1-f}\right),$$

where k involves δ, α_1 and α_0 via gamma functions.

This result is analogous to those in Lindley's normal models; my posterior log-odds are obtained by adding a linear function of the expert's log-odds to my prior log-odds. If the expert is vague in the sense that $f = 0.5$, there will still typically be a correction due to the constant k. Only in very special cases is k zero, namely those in which $\alpha_1 + \alpha_0 = 1$. Specializing even further, the multiplier $\delta(\alpha_1 - \alpha_0)$ being unity leads to $p^* \alpha pf$ and $1 - p^* \alpha (1-p)(1-f)$, so that the expert's forecast is itself the likelihood. An example, discussed below, is the case $\alpha_1 = 2/3$, $\alpha_0 = 1/3$ and $\delta = 3$. In such cases, if I am vague initially with $p = 0.5$, then $p^* = f$ and I adopt the expert's opinion.

Such models allow the processing of expert opinion in terms of bounds on f. This can be viewed as partial information, or censoring of the data f, and is particularly appropriate if the expert feels unhappy with further refinement of his statement beyond $H = \{f_l \leq f \leq f_u\}$, for lower and upper bounds f_l and f_u, respectively. The connections with the "robust Bayesian" viewpoint (Berger, 1984) are evident. Based on the Beta, or any other, model for f, the relevant likelihood components are now given by

$$p(H|Y) = \int_{f_l}^{f_u} p(f|Y)df, \qquad (Y = 0, 1),$$

providing a general solution to the censoring problem. In the very special case above, with $\alpha_1 = 2/3$, $\alpha_0 = 1/3$ and $\delta = 3$, this leads to posterior $p^* \alpha\, p\, \overline{f}$ and $1 - p^* \alpha (1-p)(1-\overline{f})$ where $\overline{f} = (f_l + f_u)/2$, although such simplistic cases are likely to be rare in practice. It should be stressed that this form of $p(H|Y)$ is only appropriate when it is not known in advance that f is to be censored in this way, and when the censoring mechanisms (ie. the reasons why the expert provides only bounds) provide no information about Y. Otherwise, alternative models would consider the bivariate distribution of $(f_l, f_u|Y)$, for each Y, if it were known in advance that the expert were to provide only bounds, or if the expert took an upper and lower probabilistic view of Y, such as Walley (1987).

3. MODELLING EXPERT DISTRIBUTIONS

The ideas above extend to cases in which Y takes values in a discrete set. Lindley (1985) proposes what are essentially multivariate logistic normal models for discrete probabilities within a general framework. The case of continuous Y has also been considered by Lindley (1983, 1987) when it is assumed that the expert distribution lies in a parametrised family. In the former reference, for example, it is assumed that the expert states $Y \sim N[\mu, \sigma^2]$ and then the model defines a joint distribution for the parameters μ and σ^2, conditional on possible outcomes Y. Similar ideas are discussed by Harrison (1985), in the context of time series forecasting. The assumption of a particular, parametric form is, however, very restrictive; here models are developed for essentially arbitrary forms.

Suppose Y to be real-valued and that the expert is to provide information about his distribution function $F(Y)$. Assume that $F(Y)$ is monotonically increasing over the real line and differentiable with density $f(Y)$. The first step in constructing a model is to specify the anticipated form of the expert distribution, conditional on all possible values of Y. With hypothesised value Y, suppose the anticipated form to be described by a distribution function $M_Y(X)$, for all real X, referred to as the *target* distribution function conditional on true value Y. These target distributions play essentially the roles of the means α_Y in the event case of Section 2, catering for anticipated biases and lack of calibration in the expert's statements. Generally, all the dependence on Y is modelled via the targets. Given Y, the value of the expert distribution $F(X)$ at any point X is a random quantity distributed about the target value $M_Y(X)$.

Example 3.1 If the target distribution is that of $(X|Y) \sim N[c+Y, 1]$, then the expert is expected to state a normal forecast distribution with unit variance and a location bias, or anticipated point forecast error, of c units. The expert is viewed as unbiased if $c = 0$ since then the anticipated location of $F(X)$ is the true value Y.

Rather than considering $F(X)$ directly, the focus in modelling is on the inverse function, namely the *quantile* function

$$Q(U) = F^{-1}(U), \qquad (0 \le U \le 1).$$

From the assumptions about $F(X)$ it follows that $Q(U)$ is monotonically increasing on [0,1], tending to $\pm\infty$ at the end-points, and also differentiable. Now $Q(U)$ maps the unit interval onto the real line so that, for each Y, the compound function

$$M_Y[Q(U)], \qquad (0 \le U \le 1),$$

is a distribution function over the unit interval; since, for me, the expert quantile function is random, then so is the compound distribution above. If the target distribution suitably models the dependence on Y and has a form as anticipated of $F(X)$, then, for each U, $M_Y[Q(U)]$ will have a distribution with location near U. The model for the quantile function can now be specified indirectly by modelling the compound distribution over the unit interval and then transforming. For fixed U, eliciting the quantile $Q(U)$ from the expert leads to the quantity $M_Y[Q(U)]$ being a random probability, and the Beta model of Section 2 may be applied. Considering U to vary over the unit interval suggests a Dirichet model as the natural extension.

Definition 1.

(a) For integer $n > 1$, let $\underline{U}_n = (U_1, \ldots, U_{n-1})$ be any fixed values defining the partition of the unit interval

$$0 = U_0 < U_1 < \ldots < U_{n-1} < U_n = 1.$$

(b) Define the corresponding quantiles of the expert distribution by $q_t = Q(U_t)$, $t = 0, \ldots, n$, so that

$$-\infty = q_0 < q_1 < \ldots < q_{n-1} < q_n = \infty.$$

Let $\underline{q}_n = (q_1, \ldots, q_{n-1})$.

(c) Let $A(U)$ be a known, continuous distribution function over the unit interval, having a density $\alpha(U)$, and set

$$a_t = A(U_t) - A(U_{t-1}), \qquad (t = 0, \ldots, n).$$

Thus a_t is the probability assigned to the interval (U_{t-1}, U_t) by $A(U)$. Let $\underline{a}_n = (a_1, \ldots, a_{n-1})$ and δ be a known, positive number.

(d) For any fixed Y, define the probabilities $\underline{\pi}_n = (\pi_1, \ldots, \pi_{n-1})$ via

$$\pi_t = M_Y(q_t) - M_Y(q_{t-1}), \qquad t = 1, \ldots, n-1,$$

and let $\pi_n = 1 - (\pi_1 + \ldots + \pi_{n-1})$. Note that $\underline{\pi}_n$ depends on Y although this is not made explicit in the notation. These probabilities are random, giving the masses allocated to the intervals (U_{t-1}, U_t) by the random distribution $M_Y[Q(U)]$ over the unit interval.

Assumption 1. $\underline{\pi}_n$ follows a Dirichlet distribution with mean $\underline{a}_n$ and precision parameter δ, having density

$$p(\underline{\pi}_n|Y) = p(\underline{\pi}_n) = \Gamma(\delta) \prod_{t=1}^{n} \frac{\pi_t^{\delta a_t - 1}}{\Gamma(\delta a_t)},$$

over the $(n-1)$-dimensional simplex. Note again that $\underline{\pi}_n$ depends on Y; this Dirichlet model is defined conditional on Y. However, since neither δ nor A depends on Y, the distribution is independent of Y. Transforming from the quantile function to that compounded with the target distribution is a pivotal device to obtain a distribution that does not involve Y. Since the assumption holds for all n and any partition $\underline{U}_n$ of the unit interval, then $M_Y[Q(U)]$ is a Dirichlet process. Some comments on this are in order. Firstly, the fact that a Dirichlet process is discrete with probability one implies that the model involves a discrete approximation, of essentially indeterminable accuracy, to a continuous problem. In the model analysis below, this feature is of little consequence due essentially to the use of a discretisation of the quantile function Q from the outset. The likelihood for Y given Q is constructed as the limiting form of a sequence of likelihoods from discrete approximations to Q, ie. histograms. A second, related feature is the implied negative correlation between probabilities $\underline{\pi}_n$ that precludes the incorporation of smoothness assumptions. The implied distribution for Q has, however, qualitatively the right form of dependence structure between quantiles. This is illustrated further in the next section. Here simply note that any two quantiles $q_t = Q(U_t)$ and $q_s = Q(U_s)$ are positively correlated, the correlation decreases as $|U_t - U_s|$ increases and tends to unity as $|U_t - U_s|$ tends to zero.

A key feature of the model is the mean distribution function $A(U)$ for $M_Y[Q(U)]$. The choice of target M_Y is assumed to provide for the general form of F anticipated, and cater for all dependence on Y. Stochastic variation away from target is modelled and controlled by the precision parameter δ, whilst $A(U)$ may be used to cater for minor systematic departures from anticipated form. A uniform mean $A(U) = U$ will often be appropriate, implying satisfaction with the target as capturing the relevant features anticipated. An example serves to illustrate the use of alternative forms. Suppose, as in Example 3.1, that the target distribution is unit variance normal with mean Y, but that it is recognised that the expert may state a heavier tailed distribution, such as Cauchy. If that happens then the compound distribution $M_Y[Q(U)]$ will be lighter tailed than uniform. Use of a mean function $A(U)$ that is essentially uniform across the central part of the unit interval but that has lighter tails will lead to a discounting of the contribution made to the likelihood by quantiles in the tails of F. This feature stems directly from the focus on the quantile function of the expert rather than the distribution directly; the positioning of the tails of F is unknown whilst those of Q lie near 0 and 1.

4. EXPERT OPINION: COLLECTIONS OF QUANTILES

Forecast statements are often given in terms of summaries of distributions, such as point forecasts with simple uncertainty measures. As in Section 2, this can be viewed as a form of censoring, providing only a partial specification of $F(Y)$. The model developed above provides a relatively easily calculated likelihood in cases when this partial information consists of selected percentage points, or quantiles, of $F(Y)$. Suppose the expert provides quantiles $\underline{q}_n$ as in Definition 1. Let $m_Y(X)$ be the density of the target distribution $M_Y(X)$, for each X. The following result now holds.

Theorem 1. Under Assumption 1, the density function for the random quantiles

$\underline{q}_n = (q_1, \ldots, q_{n-1})$, conditional on Y, is given by

$$p(\underline{q}_n|Y) = c[1 - M_Y(q_{n-1})]^{\delta a_n - 1} \prod_{t=1}^{n-1} [M_Y(q_t) - M_Y(q_{t-1})]^{\delta a_t - 1} m_Y(q_t)$$

for $-\infty < q_1 < \ldots < q_{n-1} < \infty$, where c is the constant $c = \Gamma(\delta)/\prod_{t=1}^{n}\Gamma(\delta a_t)$.

Proof. Directly by transformation from $\underline{\pi}_n$ to $(\underline{q}_n|Y)$ using the defining relationships in Assumption 1. Note that the Jacobian is simply given by

$$\left|\frac{d\underline{\pi}_n}{d\underline{q}_n'}\right| = \prod_{t=1}^{n-1} m_Y(q_t).$$

Theorem 1 provides the joint density of any collection of expert quantiles. Some insight into the form of this density and the implied relationships amongst quantiles can be obtained in the context of Example 3.1.

Example 4.1 Take target distributions $(X|Y) \sim N[Y, 1]$ corresponding to unbiased, unit variance normal expert, and suppose we consider the particular case of $Y = 0$, so that the target is standard normal. Suppose $\delta = 5$ and $A(U) = U$, a uniform mean distribution. Consider two expert quantiles, $\mu = q_1 = Q(0.5)$, the median, and $q = q_2 = Q(0.75)$, the upper quartile. The marginal distributions of each and their bivariate distribution follow from Theorem 1. The marginal for μ is symmetric and unimodal at zero, the true value. That of q has mode approximately 0.73, close to the value 0.67 of the upper quartile of the standard normal target. To explore the joint structure, Figure 1 displays the conditional density of $(q|\mu)$ for $\mu = -2, 0$ and 2. Clearly this density is zero for $q < \mu$. As μ decreases to negative values, the conditional distribution of q flattens out with mode tending quickly to zero. As μ takes larger, positive values, the conditional distribution for q becomes highly skewed, concentrating near the value of μ conditioned on as μ moves away from the target value of zero. Also displayed is the marginal density of q.

As a likelihood for Y given $\underline{q}_n$ the form in Theorem 1 has two components: one from the Dirichlet involving the product of the probabilities $\underline{\pi}_n$; the other given by the product of densities $m_Y(q_t)$. The latter is just what would be obtained if the quantiles were treated as a random sample from the target distribution given Y. The former provides correction for the postioning of the quantiles under the target distribution determind by the $\underline{U}_n$, and the implied dependence.

5. EXPERT OPINION: FULL DISTRIBUTION

Consider the generation of the expert's quantiles q_t in the previous Sections. Letting n tend to infinity with the grid points U_t remaining distinct leads to $Q(U)$ being evaluated almost everywhere. Assuming continuity implies that $Q(U)$ is fully observed, hence so is the inverse $F(Y)$. Thus the likelihood for Y based on the full expert distribution is obtainable as the limiting form, if it exists, of the likelihood from a discrete approximation as in Theorem 1. An easy way to do this is simply to take $U_t = t/n$ and this is done here.

Let H_n denote the information set $H_n = \{\underline{q}_n\}$ where the quantiles are as in Definition 1 but now with $U_t = t/n$ for each t. Denote full information by H, so that

$$H = \lim_{n \to \infty} \underline{H}_n = \{Q(U);\ 0 < U < 1\}.$$

The following result now holds.

Theorem 2. As $n \to \infty$, $H_n \to H$ and $p(\underline{q}_n|Y) = p(H_n|Y) \to p(H|Y)$ where the limiting likelihood has the form

$$p(H|Y)\ \alpha\ \exp\{-\delta D(Y)\}$$

as a function of Y, with

$$D(Y) = \int_{-\infty}^{\infty} \alpha[F(x)] f(x) \log\left[\frac{f(x)}{m_Y(x)}\right] dx,$$

whenever the integral exists for all Y.

Proof. The density of $M_Y[Q(U)]$ over $0 \leq U \leq 1$ is just the derivative $m_Y[Q(U)]/f[Q(U)]$. Hence

$$M_Y(q_t) - M_Y(q_{t-1}) = \frac{m_Y(q_t^*)}{f(q_t^*)}(U_t - U_{t-1})$$

where $q_t^* = Q(U_t^*)$ for some U_t^* between U_{t-1} and U_t. As n tends to infinity, the contribution to the likelihood from the last interval $(U_{n-1}, 1]$ is negligible compared to the rest of the likelihood, and so, for large n,

$$p(\underline{q}_n|Y) \simeq c \prod_{t=1}^{n-1} \left[\frac{m_Y(q_t^*)}{f(q_t^*)}\right]^{\delta a_t - 1} m_Y(q_t)$$

as a function of Y. Now $a_t = \alpha(\hat{U}_t)(U_t - U_{t-1})$ where $\hat{U}_t$ lies between U_{t-1} and U_t, and it follows that

$$\log[p(\underline{q}_n|Y)] \simeq \sum_{t=1}^{n-1} \left\{\log[m_Y(q_t)] + (\delta a_t - 1)\log\left[\frac{m_Y(q_t^*)}{f(q_t^*)}\right]\right\} + k$$

for some constant k. Now, since $U_t - U_{t-1} = 1/n$,

$$\log[p(\underline{q}_n|Y)] - k - \sum_{t=1}^{n-1} \left\{\log[m_Y(q_t)] + \left[\frac{\delta\alpha(U_t)}{n} - 1\right]\log\left[\frac{m_Y(q_t)}{f(q_t)}\right]\right\}$$

tends to zero as n tends to infinity. The sum in this expression may be written as

$$\frac{\delta}{n}\sum_{t=1}^{n-1} \alpha(U_t)\log\left[\frac{m_Y(q_t)}{f(q_t)}\right] + \text{ terms not involving } Y,$$

the first term of which has the limiting value

$$\delta \int_0^1 \alpha(U)\log\left[\frac{m_Y[Q(U)]}{f[Q(U)]}\right] du$$

if this integral exists. Then, transforming to $X = Q(U)$ so that $U = F(X)$, this integral is given by $-\delta D(Y)$ where

$$D(Y) = \int_{-\infty}^{\infty} \alpha[F(x)]f(x)\log\left[\frac{f(x)}{m_Y(x)}\right] dx.$$

Thus $\log[p(\underline{q}_n|Y)] \to -\delta D(Y) +$ constant as $n \to \infty$ and so, asymptotically, $\{\underline{q}_n\} \to H$ and $p(H|Y) \;\alpha\; \exp\{-\delta D(Y)\}$ as stated.

Corollary. If my prior distribution for Y has density $p(Y)$, then fully observing the expert distribution as stated leads to posterior

$$p(Y|H) \;\alpha\; p(Y)\exp\{-\delta D(Y)\}.$$

The function $D(Y)$ determining the likelihood is a generalized *divergence* measure; it measures the discrepancy between the target density $m_Y(X)$ and the stated expert density $f(X)$, for each Y. $D(Y)$ is always non-negative, being zero for all Y if and only if $f(X) = m_Y(X)$ for all X. Thus, as the value of Y varies, a large divergence leads to a small likelihood $p(H|Y)$; conversely, if $f(X)$ and $m_Y(X)$ are close in the sense of small divergence, then $p(H|Y)$ is large. Some special cases and examples appear in Section 6 below. Here note the special case in which $A(U) = U$, $0 \leq U \leq 1$, so that $\alpha(U) = 1$. This implies that, given the finite data $\underline{q}_n$, $E[\pi_t] = a_t = 1/n$ in the Dirichlet distribution of Assumption 1. The implication for $p(H|Y)$ is that $D(Y)$ is the well-known Kullback-Leibler directed divergence

$$\int_{-\infty}^{\infty} f(x) \log \left[\frac{f(x)}{m_Y(x)} \right] dx.$$

Note that, as mentioned in the proof, $D(Y)$ is assumed to exist for all Y. Some discussion of this appears in Section 6 below.

6. DISCUSSION AND EXAMPLES

Some general comments are in order before proceeding to examples. Sections 3, 4 and 5 detail the model that allows a variety of forms of expert opinion to be processed. If the full distribution is made available, Section 5 shows how a generalized divergence measure between the stated density and the target, $m_Y(X)$ for each Y, determines the likelihood. Given only collections of percentage points as in Section 4, the likelihood clearly shows that the global form of expert distribution is irrelevant; only the values of the chosen quantiles appear there, naturally weighted with the Beta form for probabilities under the target model, the values $M_Y(q_t)$, and the target density $m_Y(q_t)$.

The choice of the Dirichlet precision and the target distributions $M_Y(X)$ will typically depend on previous experience with the expert, and may be estimated based on such experience, although this is not considered here. The distribution $A(U)$ must also be specified; often $A(U) = U$ will be suitable. It leads, in particular, to the Kullback-Leibler based likelihood from the full distribution. This choice is consistent with a view that the random quantiles q_t obtained in Section 3 are to be treated equally; that is, the expert's assessment of his quantile function/distribution function is as sound in the tails as it is in the center. To model the commonly held view that tail behavior is generally difficult to determine and so q_1 and q_{n-1}, for example, are more likely to be subject to assessment error than, say, q_7 and q_8, alternative forms for $A(U)$ can be specified. It is clear from the form of $D(Y)$ in Theorem 1 that assessments in the tails will be discounted if the density $\alpha(U)$ decays rapidly as U tends to 0 or 1. As an extreme example, a "trimmed" assessment, ignoring the expert distribution below 5% and above 95% probabilities whilst treating the rest of the range consistently, can be modelled with

$$\alpha(U) = \begin{cases} 1, & 0.05 < u < 0.95; \\ 0, & \text{otherwise.} \end{cases}$$

Finally note that these features, and the forms of likelihood, derive directly from the initial focus on $Q(U)$ rather than $F(Y)$ as providing the data. This parallels experiences with elicitation where it has often been found that quantiles are more easily understood and elicited from subjects than probability distributions directly.

And now for some examples. In each of the examples, $A(U) = U$ so that $D(Y)$ is the usual, Kullback-Leibler divergence measure.

Example 6.1 The target distribution $M_Y(X)$ is that of

$$(X|Y) \sim N[c + Y, W],$$

a normal distribution with mean $c + Y$ and variance W. The constant c is an expected forecast bias; if $c = 0$ then the expert is viewed as unbiased in the sense that his forecast distribution is essentially expected to be centered at the hypothesized value Y. Three information sets are considered: the median of $F(Y)$ alone; the median plus quantiles; and the full distribution. In the first two, the global form of $F(Y)$ is irrelevant. In the third case, suppose the expert actually states a distribution, of *any* form, with mean f and variance V. It is easily shown that, using the Kullback-Leibler divergence in Section 4, the likelihood is

$$p(H|Y) \alpha \exp\left\{-\frac{\delta}{2W}(f - c - Y)^2\right\}.$$

The likelihood is the same as would be obtained from an ad-hoc model in which the point forecast f is viewed directly as a random quantity to be modelled, having a normal distribution $(f|Y) \sim N[c+Y, W/\delta]$. Such methods are used in Lindley (1983) and Harrison (1985); the current approach thus provides a foundation for such methods. That the likelihood does not depend on the spread of the forecast distribution of the expert is a rather surprising feature of the model, suggesting a need for refinement. This is a general feature arising from the assumption of target distributions that depend on Y essentially only in location parameters. Note, however, that had the expert provided a non-normal distribution with infinite variance then the results would be rather different. In fact in such a case with $F(Y)$ Cauchy for defiteness, the Kullback-Leibler divergence does not exist. This highlights the need for discounting of the tails of $F(Y)$ using $\alpha(U) \neq 1$, decaying to zero as U tends to 0 and 1. Generally the Kullback-Leibler divergence will exist only when the tails of $M_Y(X)$ are heavier than those of $F(X)$, for all Y. Since this cannot typically be ensured before observing the expert distribution, a weighting function $\alpha(U)$ decaying in the tails is essential if the likelihood is to exist. It is always possible, for example, to ensure a finite divergence using $\alpha(U)$ constant over most of the unit interval but zero for $U < \epsilon$ and $U > 1 - \epsilon$ where ϵ is a very small, positive quantity. It is also clear, however, that with $\alpha(U) = 1$ the likelihood based on any finite collection of quantiles from the Cauchy distribution is perfectly well-defined and appropriate, so that a discrete approximation to Q may be used in such cases.

Figure 2 shows the likelihoods for the three forms of information: median alone, $f = 0$; median $f = 0$ and quantiles ± 0.67 (coinciding with those of a unit normal distribution for illustration); and full information with forecast distribution having mean $f = 0$ and finite variance. The likelihoods have each been normalised to integrate to unity over the interval ± 3 for each comparison. They thus coincide with posterior densities relative to a prior $p(Y)$ being uniform over that interval. In this special case, the three forms of information can be viewed as an increasingly informative hierarchy. The model assumes $c = 0$, $W = 1$ and $\delta = 5$. The effects of increasing information are apparent in Figure 2.

Example 2. Suppose $Y > 0$ is the survival or failure time of a patient or test component and that the target distribution $M_Y(X)$ is gamma,

$$(X|Y) \sim G[b, bc/Y],$$

with density

$$m_Y(X) \; \alpha \; Y^{-b} X^{b-1} \exp\{-bcX/Y\}, \qquad (X > 0),$$

as a function of both X and Y. Under $M_Y(X)$, $E[X|Y] = Y/c$ so that c is a multiplicative bias; $c = 1$ implies an unbiased target analogous to that in the previous example where the bias was additive. Suppose that the expert actually states a distribution, of any form such that $E[log(X)] < \infty$, having mean f. Then the Kullback-Leibler based likelihood is easily seen to be given by

$$p(H|Y) \; \alpha \; Y^{-\delta b} \exp\{-\delta bcf/Y\}.$$

This is a form that is analogous to that provided by an ad-hoc model in which the point forecast f is directly model as $(f|Y) \sim G[\delta b, \delta bc/Y]$; i.e. with the bias correction c and an extra scaling δ.

Example 3. The above examples are each special cases of the following, exponential family class models. Suppose that the target distribution has density

$$m_Y(X) = h(X, \phi) \exp\{\phi[X\mu_Y - a(\mu_Y)]\}$$

for some location parameter μ_Y (for each Y), precision $\phi > 0$, and known functions $a(\cdot)$ and $h(\cdot, \cdot)$. Note that this distribution has mean $E[X|Y] = a'(\mu_Y)$. Suppose that the expert distribution is such that $E[X] = a'(f)$ for some f, and that $E[log\{h(X, \phi\}] < \infty$. The Kullback-Leibler based likelihood is easily derived as

$$p(H|Y) \; \alpha \; \exp\{\delta\phi[a'(f)\mu_Y - a(\mu_Y)]\}.$$

This is a form analogous to that provided by an ad-hoc model in which the point forecast $E[X] = a'(f)$ of the expert is directly modelled as coming from a distribution of the form $M_Y(\cdot)$, but with precision $\delta\phi$.

Note that in each case the expert distribution appears only through the mean. The comments on this point in Example 1 are relevant generally. Thus, again, partial expert opinion is processed, now in terms of the mean rather than quantiles. This is a very special setup, however, and is not the case if the precision ϕ of the target model depends on Y; then calculation of the Kullback-Leibler divergence requires other features of F to be available.

ACKNOWLEDGEMENTS

I am grateful to Mark Schervish, Dennis Lindley and Jim Berger for their comments on the work reported here. An original draft was written whilst visiting Department of Statistics at Purdue University.

REFERENCES

[1] BERGER, J.O. (1984). The robust Bayesian viewpoint. In *Robustness of Bayesian analyses*, (ed. J.B. Kadane). Elsevier Science.

[2] HARRISON, P.J. (1985). The aggregation and combination of forecasts. Warwick Research Report 68.

[3] LINDLEY, D.V. (1983). Reconciliation of probability distributions. *Oper. Res.* **31**, 806-880.

[4] LINDLEY, D.V. (1985). Reconciliation of discrete probability distributions. In 1 *Bayesian Statistics 2* (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith). North Holland, Amsterdam, and Valencia University Press.

[5] LINDLEY, D.V. (1987). Using probabilities. Unpublished manuscript.

[6] WALLEY, P. (1987). Conflict and uncertainty. Unpublished manuscript.