Modelling Expert Opinion

MIKE WEST
University of Warwick*

SUMMARY
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 provide methods for the processing of forecast information in terms of full, continuous distributions or densities, and partial information in terms of collections of quantiles. 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.

Keywords: EXPERT OPINION; DIRICHLET DISTRIBUTION; QUANTILES.

1. INTRODUCTION
In a series of papers, most recently Lindley (1988) 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 pesos 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 quantiles, or deciles;
(d) histograms as discrete approximations to continuous distributions.

* Now at Duke University, North Carolina, USA.
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, 1988) 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

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

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 $\alpha_Y > \alpha_0$, and that expertise increases with $\alpha_Y - \alpha_0$. 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

\[ \log \left( \frac{p^*}{1 - p^*} \right) = \log \left( \frac{p}{1 - p} \right) + \delta (\alpha_Y - \alpha_0) \log \left( \frac{f}{1 - f} \right), \]

where $\delta$ involves $\delta$, $\alpha_Y$ and $\alpha_0$ 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 $\alpha_Y + \alpha_0 = 1$. Specializing even further, the multiplier $\delta (\alpha_Y - \alpha_0)$ being unity leads to $p^* \propto pf$ and $1 - p^* \propto (1 - p)(1 - f)$, so that the expert’s forecast is itself the likelihood. An example, discussed below, is the case, $\alpha_Y = 0.5$, $\alpha_0 = 0.5$ 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 opinions 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
Modelling Expert Opinion

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 (1982, 1988) when it is assumed that the expert distribution lies in a parameterized 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 \( \mu \) and \( \sigma^2 \), conditional on possible outcomes \( Y \). The assumption of a particular, parametric form is, however, rather restrictive; it would be nice to have models catering for relatively general forms and partial information about the distribution. A first step towards this goal is explored here.

Suppose \( Y \) to be real-valued and that the expert is to provide information about its distribution function for \( Y \), namely \( F(.) \). For clairty of notation in what follows, use \( X \) as the argument of \( F \). Assume that \( F(X) \) is monotonically increasing over the real line and differentiable with density \( f(X) \). Rather than considering \( F(X) \) directly, the focus is switched to the inverse of \( F(X) \), namely the quantile function

\[
Q(U) = F^{-1}(U), \quad (0 \leq U \leq 1).
\]

From the assumptions about \( F(X) \) it follows that \( Q(U) \) is monotonically increasing on \([0, 1]\), tending to \(+\infty\) at the end-points, and also differentiable. Knowledge of the distribution function is equivalent to that of the quantile function and, in view of the earlier comments about partial information in terms of quantiles, it is natural to consider the latter directly. To develop a model for the distribution, individual points on the quantile function are considered, starting with the median of \( F(X) \), \( \mu = Q(0.5) \).

The first step is to specify the distribution for \( \mu \) conditional on each possible value of \( Y \), via the density \( p(\mu|Y) \). Again the basic ideas in Lindley (1988, Section 17) are helpful; there Lindley considers the mean (assumed to exist) rather than the median of \( F(X) \) and uses a normal model for the mean for each \( Y \). The focus on the median as a point forecast links more directly to the quantile function and, in this context, \( \mu \) always exists and is unique. Choice of the distribution for \( (\mu|Y) \) will depend on prior information about the relationship between median point forecasts from the expert/model, and actual outcomes. Clearly in some contexts there will exist substantial relevant information on which to base this model and the parameters defining it, although these considerations are not explored further here.
For each \( Y \), let

\[ M_Y(X), \quad (-\infty < X < \infty), \]

be a monotone, continuous distribution function over the real line. \( M_Y(X) \) is chosen to model the anticipated form of the distribution of the median \( \mu \) conditional on true value \( Y \). This plays a role analogous to that of the mean \( \alpha_Y \) in the event case model of Section 2, catering for biases and lack of calibration of the expert median forecast.

**Example 3.1.** A particular example is the normal model in which \( M_Y(X) \) is the normal distribution \( (X|Y) \sim N(c + Y, W) \). Often it may be suitable to assume that median point forecasts are related to outcomes via such a normal model, possibly after transformation; i.e. that forecast errors \( \mu - Y \) (with \( Y \) observed therefore fixed) are normally distributed with mean (bias) \( c \) units and variance \( W \). The empirical study of Pratt and Schlaifer (1985) provides a case study in which this form of relationship is suggested. A possibly useful refinement is to consider also a dependence of the variance \( W \) on \( Y \) in some cases. Note that the expert is viewed as median unbiased if \( c = 0 \) since then the anticipated location of \( p(\mu|Y) \) is the true value \( Y \).

If the median \( \mu \) is assumed to follow the distribution \( M_Y(\mu) \) conditional on \( Y \), it follows that

\[ \mu = M_Y^{-1}(\pi), \]

where \( \pi \) is a uniform random quantity on the unit interval, \( \pi \sim U[0,1] \). In Example 3.1, \( \mu = c + Y + W^{1/2} \Phi^{-1}(\pi) \) where \( \Phi \) is the standard normal c.d.f. This model is completely analogous to that for the forecast mean in Lindley (1988, Section 17), and as such it suffers from the sensitivity to the assumption of the particular parametric form of \( M_Y \). To alleviate this, the distribution of \( \mu \) can be modelled as coming from a neighbourhood of the chosen \( M_Y \) by allowing the distribution of \( \pi \) above to be something other than uniform. Thus the following assumption is made.

**Assumption 1.** Conditional on \( Y \) and the specified distribution function \( M_Y(\cdot) \),

\[ \mu = M_Y^{-1}(\pi), \]

where \( \pi \) is a Beta random quantity,

\[ \pi \sim B(\delta/2, \delta/2), \]

for some precision parameter \( \delta > 0 \), not depending on \( Y \).

Under this assumption, the Beta random quantity can be written as \( \pi = M_Y(\mu) = M_Y(Q(0.5)) \). Thus \( \pi \) is the value at \( U = 0.5 \) of the compound distribution function \( M_Y(Q(U)) \) over the unit interval. Note that this is a random distribution function (for me) since \( Q(U) \) is a random quantity for all \( U \). Note that \( E[\pi] = 0.5 \) for all \( \delta \) so that the location of the distribution of \( \mu \) is essentially that determined by \( M_Y \). If \( \delta = 2 \) then the distribution of \( \mu \) is just the original \( M_Y \). Otherwise \( p(\mu|Y) \) is more or less diffuse than the density of \( M_Y \) according to whether \( \delta \) is less than or greater than 2. In the normal case of Example 3.1, \( \mu \) has a symmetric distribution with median \( c + Y \) for all \( \delta \).

Consider now extending the model to a general point on the forecast quantile function,

\[ q = Q(U), \]

for given \( U \) in \([0, 1] \). Take \( U = 0.75 \) as an example so that \( q \) is the upper quantile. Clearly \( q > \mu \) and therefore can be expressed as

\[ q = M_Y^{-1}(\theta), \]
for some $\theta$ such that $\pi < \theta \leq 1$. Following the reasoning for the median, the obvious extension is to take $\theta$ as $\text{distribution over the unit interval}$ with location near 0.75. Referring back to any quantile $q = Q(U)$, the immediate extension of the model for the median is given by assuming $q = Q(U) = M_Y^{-1}(\theta)$ where

$$\theta \sim B\{U, 5(1-U)\},$$

for each $U$. Considering $U$ to vary over the unit interval lead to a Dirichlet model as the natural extension. The following definition formalises this discussion. From here on, the function $M_Y(\cdot)$ is referred to as the $\text{larger distribution}$ of the model for the quantiles, conditional on $U$.

**Definition 1.**

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

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

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

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

Let $\xi = (\xi_1, \ldots, \xi_{n-1})$.

(c) For any fixed $Y$, define the probabilities $\pi_n = (\pi_1, \ldots, \pi_{n-1})$ via

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

and let $\pi_n = 1 - (\pi_1 + \cdots + \pi_{n-1})$. Note that $\pi_n$ 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 2.** Let $\xi_n$ be the $n$-vector $(1/n, \ldots, 1/n)$. Then $\pi_n$ follows a Dirichlet distribution with mean $\xi_n$ and precision parameter $\delta$, having density

$$p(\pi_n | Y) = \pi^\delta \prod_{t=1}^n \Gamma(\delta/n),$$

over the $(n-1)$-dimensional simplex.

Note again that $\pi_n$ depends on $Y$; this Dirichlet model is defined conditional on $Y$. However, since neither $\delta$ nor $\xi_n$ depends on $Y$, the distribution is independent of $Y$. Transforming the quantiles to $\pi_n$ involving the target distribution $M_Y(\cdot)$ is essentially a pivotal device to obtain a distribution that does not involve $Y$. Since the assumption holds for all $n$ and any partition $U_n$ of the unit interval, $M_Y[Q(U)]$ 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 of a sequence of likelihoods from discrete approximations to $Q$ i.e. histograms. A second, related feature is the implied negative correlation between probabilities $\pi_n$ that precludes the incorporation of smoothness assumptions. The implied distribution for $Q$ has, however, qualitatively the right form of dependence structure between quantiles due to ordering. Thus any two quantiles $q_1 = Q(U_1)$ and $q_2 = Q(U_2)$ are positively correlated, the correlation decreases as $|U_1 - U_2|$ increases and tends to unity as $|U_1 - U_2|$ tends to zero.
4. EXPERT OPINION: COLLECTIONS OF QUANTILES

Forecast statements are often 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 \( F(X) \). The model developed above provides a relatively easily calculated likelihood in cases when this partial information consists of selected percentage points, or quantiles, of \( F(X) \). Suppose the expert provides quantiles \( q_n \), as in Definition 2. Let \( m_Y(X) \) be the density of the target distribution \( M_Y(X) \), for each \( Y \). The following result now holds.

Theorem 1. Under Assumption 2, the density function for the random quantiles \( q_n \), conditional on \( Y \), is given by

\[
p(q_n|Y) = c (1 - M_Y(q_n-1))^{1/n-1} \prod_{i=1}^{n-1} (M_Y(q_i) - M_Y(q_{i-1}))^{i/n-1} m_Y(q_i)
\]

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

**Proof.** Directly by transformation from \( \pi_n \) to \( (q_n|Y) \) using the defining relationships in Assumption 1. Note that the Jacobian is simply given by

\[
\left| \frac{dn_n}{dq_n} \right| = \prod_{i=1}^{n-1} m_Y(q_i).
\]

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

Example 4.1. Take target distributions \( (X|Y) \sim N(Y, 1) \), so that the expert median \( \mu \) has a distribution symmetric about \( Y \), unimodal if \( \delta > 2 \). Thus the expert is viewed as median unbiased. Consider the particular case of \( Y = 0 \), so that the target is standard normal. Suppose in addition that \( \delta = 5 \). Consider two expert quantiles, \( \mu = q_1 = Q(0.5) \), the median, and \( \mu = q_2 = Q(0.75) \), the upper quantile. The marginal distributions of each and their bivariate distribution follow from Theorem 1. The marginal for \( \mu \) is symmetric and unimodal at zero, the true value. That of \( q \) has mode approximately 0.68, close to the value 0.67, the upper quantile of the standard normal target. To explore the joint structure, Figure 1 displays the marginal densities of \( (q|\mu) \) for \( \mu = -2, 0 \) and 2. Clearly this density is zero for \( q < \mu \). As \( \mu \) decreases to negative values, the conditional distribution of \( q \) flattens out with mode tending quickly to zero. As \( \mu \) takes larger, positive values, the conditional distribution for \( q \) becomes highly skewed, concentrating near the value of \( \mu \) conditioned on as \( \mu \) moves away from the target value of zero. Also displayed is the marginal density of \( q \).

As a likelihood for \( Y \) given \( q_n \), the form in Theorem 1 has two components: one from the Dirichlet involving the product of the probabilities \( \pi_n \), the other given by the product of densities \( m_Y(q_i) \). 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 positioning of the quantiles under the target distribution determined by \( \pi_n \), and the implied dependence.

A generalisation of some practical importance (to be explored later), involves taking the same Dirichlet based model but allowing for different means of the \( \pi_n \) probabilities. This extension is described as follows.
3F QUANTILES

Distributions, such as point forecasts be viewed as a form of censoring, model developed above provides a
tial information consists of selected expert provides quantiles \( q_i \) as in
bution \( M_Y(X) \), for each \( Y \). The

\[
M_Y(q_{i-1})^{1/n-1} m_Y(q_i)
\]

c = \( \Gamma(\delta)/\prod_{i=1}^n \Gamma(\delta/n) \).

The defining relationships in Ass-
on of expert quantiles. Some insight
among quantiles can be obtained
so that the expert median \( \mu \) has
as the expert is viewed as median
target is standard normal. Suppose
that = \( Q(0.5) \), the median, and
utions of each and their bivariate
ymmetric and unimodal at zero,
ose to the value 0.67, the upper
x structure, Figure 1 displays the
is density is zero for \( q < \mu \). As \( \mu
 flattens out with mode tending
itional distribution for \( q \) becomes
ed on as \( \mu \) moves away from the
of \( \mu \).
has two components: one from
the other given by the product of
of the quantiles were treated as
former provides correction for a
determined by the \( U_i \), and the
xplored later), involves taking the
ans of the \( \pi_i \) probabilities. This

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Marginal density \( p(q) \) and conditional densities \( p(q|m) \).}
\end{figure}

**Assumption 2a.**

(a) Let \( A(U) \) be a known, continuous and monotone distribution function over the unit interval, having a density function \( a(U) \).

(b) In the framework of Definition 1, redefine the \( n \)-vector \( a_n \) as \( a_n = (a_1, \ldots, a_n) \) where
\[
a_i = A(U_i) - A(U_{i-1}), \quad (i = 1, \ldots, n).
\]

Then, under Assumption 2 with this modification, it follows that \( x_n \) has a Dirichelet distribution with mean \( a_n \) and precision parameter \( \delta \), having density

\[
p(x_n | Y) = p(x_n) = \Gamma(\delta) \prod_{i=1}^n x_i^{\delta a_i - 1} / \Gamma(\delta a_i),
\]

over the \( (n - 1) \)-dimensional simplex.

**Theorem 1a.** Under Assumption 2a, the density function for the random quantiles \( q_{ni} \), conditional on \( Y \), is given by

\[
p(q_{ni} | Y) = c [1 - M_Y(q_{ni-1})^{\delta a_{ni-1}} \prod_{i=1}^{n-1} (M_Y(q_i) - M_Y(q_{i-1}))^{\delta a_i - 1} m_Y(q_i)]
\]

for \(-\infty < q_1 < \cdots < q_{n-1} < \infty\), where \( c \) is the constant \( c = \Gamma(\delta)/\prod_{i=1}^n \Gamma(\delta a_i) \).

**Proof.** Directly by transformation as in Theorem 1.

Why is this generalisation of interest? Note first that the original model in Assumption 2 and Theorem 1 is the special case when \( a_i = 1/n \) for all \( i \), corresponding to the uniform distribution \( A(U) = U \). The choice of target \( M_Y \) is assumed to cater for all dependence on \( Y \) in the distribution of the median in Assumption 1, and the other quantiles in Assumption 2a.
Stochastic variation away from the target is modelled and controlled by the precision parameter \( \delta \), 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 4.1, that the target distribution is unit variance normal with mean \( \mu \), but that in recognition of the fact that \( \delta \) is small there is some small chance that the median may be more accurately modelled using 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. Further discussion of this appears in the illustrations of Section 6.

5. EXPERT OPINION: FULL DISTRIBUTION

Consider the generation of the expert’s quantiles \( q_{n} \) in Definition 1. Letting \( n \) tend to infinity with the grid points \( U_{1} \) remaining distinct leads to \( Q(U) \) being evaluated almost everywhere. Assuming continuity implies that \( Q(U) \) is fully observed hence is the inverse \( F(X) \). 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 \( \hat{U}_{i} = t/n \) and this is done here.

Let \( H_{n} \) denote the information set \( H_{n} = \{ q_{n} \} \) where the quantiles are as in Definition 1 but now with \( U_{i} = t/n \) for each \( t \). Denote full information by \( H \), so that

\[
H = \lim_{n \to \infty} H_{n} = \{ Q(U); 0 < U < 1 \}.
\]

Recall that \( F(X) \) has density \( f(X) \), \( M_{Y}(X) \) has density \( m_{Y}(X) \) and \( A(U) \) of Assumption 2a has density \( \alpha(U) \). The following result now holds.

**Theorem 2.** Suppose the model is as defined in Assumption 2a. Then, as \( n \to \infty, H_{n} \to H \) and \( p(q_{n} | Y) = p(H_{n} | Y) \to p(H | Y) \) where the limiting likelihood has the form

\[
p(H | Y) \propto \exp(-\delta D(Y))
\]

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 \( 0 < U < 1 \) is just the derivative which is easily seen to be given by \( m_{Y}(Q(U))/f(Q(U)) \). Hence

\[
M_{Y}(U) - M_{Y}(U_{i-1}) = \frac{m_{Y}(U_{i})}{f(U_{i})} (U_{i} - U_{i-1})
\]

where \( q_{i} = Q(U_{i}) \) for some \( U_{i} \) between \( U_{i-1} \) and \( U_{i} \). 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 \( n \),

\[
p(q_{n} | Y) \approx c \prod_{i=1}^{n-1} \left[ \frac{m_{Y}(U_{i})}{f(U_{i})} \right]^{d_{i}-1} m_{Y}(U_{i})
\]
Modelling Expert Opinion

as a function of \( Y \). Thus, for some constant \( k \),

\[
\log[p(q_{n}|Y)] = \sum_{i=1}^{n-1} \left\{ \log[m_Y(q_{i})] + \frac{\delta q_{i}}{n} \right\} + k.
\]

Now \( a_{i} = \alpha(\hat{U}_{i})(U_{i} - U_{i-1}) \) where \( \hat{U}_{i} \) lies between \( U_{i-1} \) and \( U_{i} \), and so, since \( U_{i} - U_{i-1} = 1/n \),

\[
\log[p(q_{n}|Y)] - k = \sum_{i=1}^{n-1} \left\{ \log[m_Y(q_{i})] + \frac{\delta q_{i}}{n} \right\} + \left[ \log[m_Y(q_{i})] - \log[f(q_{i})] \right]
\]

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

\[
\frac{\delta}{n} \sum_{i=1}^{n-1} \alpha(\hat{U}_{i}) \log \left[ \frac{m_Y(q_{i})}{f(q_{i})} \right] + \text{ terms not involving } Y,
\]

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 \( \log[p(q_{n}|Y)] \rightarrow -\delta D(Y) + \text{ constant as } n \rightarrow \infty \) and so, asymptotically, \( \{q_{n}\} \rightarrow H \) and \( p(Y|H) \propto \exp(-\delta D(Y)) \) as stated.

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

\[
p(Y|H) \propto p(Y) \exp(-\delta D(Y)).
\]

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 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 special case in which \( \alpha(U) = U \), \( 0 \leq U \leq 1 \), so that \( \alpha(U) = 1 \). This implies that, given the finite data \( q_{n}, k \), \( k/q_{i} = a_{i} = 1/n \) as in the original Dirichlet distribution of Assumption 2. 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 \( M_Y(q_i) \), and the target density \( m_Y(q_i) \).

The choice of the Dirichlet precision and the 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 \( q_i \) obtained in Section 3 are to be treated equally; that is, the expert’s assessment of his quantile function/distribution function is just sound in the tails as it is in the centre. To model the commonly held view that tail behavior is generally difficult to determine and so \( q_i \) and \( q_{n-1} \) for example, are more likely to be subject to assessment error than, say, \( q_1 \) and \( q_0 \), 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 modeled with

\[
\alpha(U) \propto \begin{cases} 
1, & 0.05 < U < 0.95; \\
0, & \text{otherwise.}
\end{cases}
\]

Finally note that these features, and the forms of likelihood, derive directly from the initial focus on \( Q(U) \) rather than \( P(Y) \) 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

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

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

\[
p(H|Y) \propto \exp \left\{ -\frac{\delta}{2W} (f - \theta - Y)^2 \right\}.
\]

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 \( f(Y) \sim N(\theta + Y, W/\delta) \). Such methods are used in Lindley (1985); the current approach provides further foundation. 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
Modelling Expert Opinion

There are examples. Sections 3, 4, and 5 present a tutorial opinion to be processed. If the full generalized divergence measure between a probability function and a set of probability functions $M_Y(X)$ is to be processed, then it is necessary to determine the likelihood of the different quantiles appearing, under the target model, the values of $A(U)$ will typically depend on the data. This is the case with $F(Y)$ Cauchy, for example, the Kullback-Leibler divergence does not exist. This highlights the need for discounting the tails of $F(Y)$ using $a(U) \neq 1$, decay $a(U)$ in the tails. Generally the Kullback-Leibler divergence will exist only when $a(U)$ is smaller than the behaviour of $F(Y)$ for all $Y$. Since this cannot be ensured before observing the expert distribution, a weighting function $a(U)$ decay in the tails, such as $q_{0.5}$, is essential if the likelihood is to exist. It is always possible, for example, to ensure a finite divergence using $a(U)$ constant over most of the unit interval but zero for $U < c$ and $U > 1 - c$ where $c$ is a very small, positive quantity. It is also clear, however, that with $a(U) = 1$ 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. Likelihoods $p(Y|X)$](image)

Figure 2 shows the likelihoods for the three forms of information: median alone, $f = 0$ and $f = 1$ and $f = 0.67$ (coinciding with those of a unit normal distribution for illustration); and full information vs forecast distribution, $f = 0$ and finite variance. The likelihoods have been normalised to integrate to unity over the interval $0$ to $f$. The likelihoods do not depend on the spread of the forecast distribution is rather surprising feature of the model, suggesting a need for refinement. One refinement that is clearly necessary in practice is to allow for uncertainty about the target distribution generally, and $W$ in particular in this example, and also about $\delta$. Uncertainty about $W$ is represented in terms of a prior distribution which would allow for learning. Although this is not pursued further here, studying past predictions from the expert model will lead to an informed prior for $W$ and the other components. Sequential forecasting of time series is a natural candidate application area in which such information about predictive ability is sequentially obtained.

An alternative refinement stems from the recognition that the strong (indeed, almost, sufficient) dependence of the likelihood on the location of $F(X)$ alone derives from the

\[ c - Y \right] \]
use of a neighbourhood of the normal target distribution. Use of a target more disposed to recognising spread in the quantiles, i.e. a heavier-tailed form (not strongly unimodal), provides a simple and appropriate alternative. As an example, compare the above normal model with that in which the target is replaced by the Cauchy with unit scale parameter, \( M_Y(X) = 0.5 + \arctan(X - Y) \), denoted by \( C(Y, 1) \). Suppose that data \( H \) consists of the quantiles \( Q(0.05), Q(0.10), \ldots, Q(0.95) \) taken from the expert distribution \( P(X) = \Phi(X/\beta) \), a normal forecast distribution with zero mean and variance 4. The likelihoods for \( Y \) from the normal and Cauchy models are displayed in Figure 3. Note that these are similar to the limiting forms that would be obtained from Theorem 2a given the full information about \( P \). Clearly the Cauchy based model accounts for the unanticipated spread of \( P \) much more than the normal based model; unlike the latter model, the limiting form of the likelihood in the former will be dependent on the spread. An illustration of the response of these two models to spread amongst number of quantiles appears in Figure 4. Here the likelihoods from normal and Cauchy models are based on data \( H \) comprising the median \( \mu = 0 \) and the upper quantile \( q = 3 \) of \( P \). As with the example illustrated in Figure 3, the data are consistent with \( Y = 0 \) but also with a much more diffuse expert distribution than anticipated by the target. The more appropriate likelihood from the Cauchy model accounts for the surprise value of the data in a way that the normal model cannot.

![Figure 3. Comparison of normal and Cauchy models.](image)

**Example 6.2.** Suppose \( Y > 0 \) is the survival or failure time of a patient or test component and that the distribution \( M_Y(X) \) is gamma,

\[
(X|Y) \sim \text{Gamma}(b, d|Y),
\]

with density

\[
m_Y(X) \propto Y^{-d}X^{b-1} \exp \{-dX/Y \}, \quad (X > 0),
\]
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
Modelling Expert Opinion

Information:

\[ N = \{ \text{median} = 0 \text{ and } \text{upper quartile} = 3 \} \]

Normal: \( N \sim N(\mu, \sigma) \)

Cauchy: \( C \sim C(Y, 1) \)

---

Normal

-----

Cauchy

\[ \text{Information: } \]

\[ H = \{ \text{median} = 0 \text{ and } \text{upper quartile} = 3 \} \]

\[ N \sim N(\mu, \sigma) \]

\[ C \sim C(Y, 1) \]

\[ \text{Normal} \]

-----

\[ \text{Cauchy} \]

---

\[ \text{Cauchy models.} \]

The time of a patient or test component

\[ (X > 0) \]

\[ Y = Y/e \] so that \( e \) is a multiplicative factor in the previous example where there is a distribution, of any form such that

\[ E[\log(X)] < \infty, \text{ having mean } f \]. \]

Then the Kullback-Leibler based likelihood is easily seen to be given by

\[ p(H|Y) \propto Y^{-\delta} \exp\{-\delta \log f / Y\}. \]

This is a form that is analogous to that provided by the ad-hoc model in which the point forecast \( f \) is directly modelled as \( (f(Y) \sim G(\delta X, \delta \phi / Y); \text{i.e. with the bias correction } \phi \text{ and an extra scaling } \delta. \]

**Example 6.3.** 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 \( \mu_Y \) (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 \( E[X] = f \) for some \( f \), and that \( E[\log h(X, \phi)] < \infty \). The Kullback-Leibler based likelihood is easily derived as

\[ p(H|Y) \propto \exp\{\delta \phi[f \mu_Y - a(\mu_Y)]\}. \]

This is a form analogous to that provided by an ad-hoc model in which the point forecast \( E[X] = f \) 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. Thus again, partial expert opinion is processed, now in terms of the mean rather than quantiles. The extensive discussion of this feature, and the types of refinement possible to alleviate it, in the normal case of Example 6.1 are clearly relevant more generally. The models require exploration to identify the effects of refinements such as: (i) the use of non-uniform functions \( \lambda(U) \); (ii) the use of alternative target distributions (as with the Cauchy replacement for normality) to those in the exponential family that lead to this strong dependence on location parameters of \( F \) and (iii) prior distributions for parameters of the target distribution to provide learning.
ACKNOWLEDGEMENTS

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

REFERENCES


DISCUSSION

R. J. OWEN (University of Wales, Aberystwyth)

The question of how to deal formally with the opinions of persons is an important area for Bayesians. A central feature is the apparent inability of people to say what they really know. Their statements may be incomplete, biased or even ‘in negative accord with reality’. Dr. West has made here a valuable contribution to this problem area in providing a way of using all the information stated for prediction of a future variable Y. This new development both allows Y to be continuous and copes with an expert stating his distribution either completely or incompletely in quantile form. Elicitation of quantiles is of course often a more realistic proposition.

The key to the new development is the Dirichlet model for ‘my’ distribution of the expert’s statement conditional on Y. This gives, for fixed ‘statement’ and generic Y, the likelihood function in Y.

Lindley (1983, 1988) has considered the case of discrete Y in detail and continuous Y (Lindley, 1988) when attention is restricted to stating location and scale measures.

It is perhaps worth stating two related problems which are not covered by this paper. Firstly, there is the given problem but without the implicit assumption that the expert is passively stating what he believes and is not contriving to influence the conclusion beyond this. In such a situation a game theoretic element would enter the analysis. Secondly, there is the problem of determining a consensus where ‘I’ am no longer the executive and instead the expert and ‘I’ must together reach a conclusion.

Some interesting features may be illustrated by the event case (binary Y) where the probability f is stated by the expert. The constant k, which depends only on the model for the likelihood function and has the same sign as 1 - a0 - a1, corrects for possible assessed bias of the expert. If a0 + a1 le 1 and a1 > 1 (representing an expert viewed as being in positive accord with reality though biased towards Y = 0) then for large 8 and any p, the resultant p* is near 1 or 0 according as f > or < 1. Thus, in this case, the pooling of the probabilities p and f is very far from linear. To see that this is sensible, observe that when 8 is large the expert is virtually expected to specify Y (though he would be unaware of this) by a value of f near a0 or a1. The normative approach has taken this extreme case in its stride!

Also in the binary case with large 8, a value of f near neither a0 nor a1 could lead to my questioning the assumed model for the likelihood function. The question of responding
Modelling Expert Opinion

nts

n Lindley for discussion of the work visiting the Department of Statistics at

s of Bayesian analysis, (J. B. Kadane, ed.).


elevar Life Tests and Experts’ Opinion in

of judgements on probability distributions: a

D. V. Lindley and A. P. M. Smith,

abilities. London: Chapman and Hall. (To

As of persons is an important area for

people to say what they really know.

negative accord with reality’. Dr. 

lem area in providing a way of using 

iable Y. This new development both 

ing his distribution either completely 

is of course often a more realistic 

at model for ‘my’ distribution of the 

xed ‘statement’ and generic Y, the 

crete Y in detail and continuous Y 

ication and scale measures, 

which are not covered by this paper, 

licit assumption that the expert is 

fluence the conclusion beyond 

nter the analysis. Secondly, there 

 recognizing an expert viewed as being in 

= 0) then for large $\delta$ and any $p$, the 

Thus, in this case, the pooling of the 

is this sensitive, observe that when $\delta$ 

ough he would be unaware of this) by 

taken this extreme case in its strident 

ear neither $\alpha_0$ nor $\alpha_1$ could lead to 

ction. The question of responding 

formally to extremely surprising utterances of the expert is considered in Example 6.1 where a 

Cauchy alternative to the normal target distribution is seen to alleviate the difficulty. This is a

smooth model allowing for surprising statements and I wonder whether a tractable alternative 

would be a ‘discontinuous’ model giving a different functional form for the likelihood function 

when there is a sufficiently surprising statement by the expert.

Generally the modelling and elicitation of both the target distribution and expert’s state- 

ment is delicate and the sensitivity of the subsequent pooling with respect to these choices 

deserves scrutiny. However, sensitivity requirements would typically be softened when the 

ultimate purpose is the choice of a decision.

The examples in the final section show that, when the target distribution belongs to a 

subset of the exponential family, the pooling depends on a (mildly restricted) expert distribution 

only through its mean. Consideration of the Cauchy alternative target distribution in Example 

6.1 indicates that the property is less than universal. It would be useful to know how 

nearly the property holds when the target distribution is ‘near’ the exponential class; for 

instance in Example 6.1 a Student’s r target distribution with large degrees of freedom would 

be illuminating. Generally there is a related set of questions concerning when (for which class 

of target distribution?) a particular partial specification of the expert distribution is sufficient. 

This could influence both modelling the target distribution and eliciting the expert distribution.

McConway (1981) focussed on a marginalization property to be satisfied in this con- 

text. The procedures here do not satisfy that property; however, Lindley (1985) has argued 

convincingly against marginalization being satisfied.

A possible refinement would be desirable in cases where ‘my’ knowledge and that of 

the expert overlap. This has been considered by French (1980) and Lindley (1988) with other 

models. In this paper consider, for example, the case where $Y$ is binary and $f$ is stated, then 

one kind of dependence of knowledge may be incorporated by allowing $\alpha Y$ to depend on $p$.

Further possible future developments are mentioned by Dr. West. Perhaps of most 

interest would be the one which would allow for uncertainty about the target distribution so 

that learning about the expert could be incorporated when he is consulted repeatedly.

Another potential generalization would be to the case of several experts. Here the possi- 

bility of knowledge overlap would also arise.

REPLY TO THE DISCUSSION

First let me thank Dr. Owen for his comments and discussion. He raises several issues 

concerning features of the particular models in my paper, and others concerned with the area 

more generally. Suggestions for further development, including multi-expert problems and 

learning about model parameters, are accepted without comment and will, hopefully, lead to 

further work in the area.

Several points relate essentially to the choice of target distribution $M_f(\cdot)$. The problem 

of “surprising” data $H$ is a common one in parametric models, with no particularly new features 

here. Standard considerations of model robustness, outlier accommodation and the use of 

predictive distributions (either $p(H)$ or $p(H|Y)$ when $Y$ is observed) apply. What is certainly 

needed first, however, is an appreciation of the nature and extent of the dependence of the 

likelihood on features of $H$ for any chosen model; Section 6, and notably Example 6.1, is 

devoted to this sort of study, comparing models based on normal and Cauchy targets. Figure 

4 highlights the differences for a data set $H$ “surprising” under the normal model. The effects 

here are, of course, to be anticipated in view of the well-known, related features of normal 

vs. Cauchy based models for sampling distributions and/or priors in standard applications 

(see, eg. O’Hagan, in these Proceedings). As Dr. Owen suggests, there is a need for wider 

theoretical study of the likelihoods, and apparently the most promising approach is via the 

limiting case and exploration of divergences (although this, limiting case is only part the story).
The question of whether a partial specification of \(F(\cdot)\) suffices to determine the divergence is intriguing; my guess is that this is only the case in particular, exponential family based models, as in Example 6.3, although I have no theoretical results here. Let me reaffirm that I find this particular feature of exponential family models to be rather unsatisfactory as a general rule, preferring models that reflect some features of the spread of uncertainty in \(F(\cdot)\) rather than just the location. From a practical viewpoint, however, the appropriateness of particular models must be addressed in the context of application. In the applied study of Pratt and Schlaifer (1985), for example, the authors conclude (from a data analytic viewpoint) that the median forecasts are essentially sufficient for the outcomes, further specification of \(F(\cdot)\) via quantities providing relatively little additional information.

Dr. Owen raises three issues drawn from the wider field of aggregation of probabilities/distributions. What about the consensus issue, when "the expert and I must together reach a conclusion", and game theoretic considerations? Well, my paper concerns only how I, or any other individual, may view the statements of others. The general idea may apply, of course, within a group environment. Each individual models what he sees, having scope to adjust for perceived or anticipated biases in the statements of others, and therefore allowing for and counteracting possible attempts to "influence the conclusion". On the aggregation issue, the Bayesian framework concerns individuals; group probabilities do not exist. For further discussion of this, see West (1984) and French (1985). On the issue of marginalisation, I have nothing to add to Dr. Owen's support for the conclusions of Lindley (1988).

A final noteworthy point raised by Dr. Owen concerns "cases where my knowledge and that of the expert overlap". Clearly the models in my paper allow for dependence of \(p(H|Y)\) on \(p(Y)\), my prior, although this is not specifically mentioned nor made explicit. In connection with this, let me note a slightly alternative approach that obviates this problem, and the need for explicit specification of \(p(Y)\) itself. Begin by switching the focus in modelling from the forecast median \(\mu\) to the outcome \(Y\). To be specific, take the example in which \((\mu|Y) \sim N[\mu,Y]\) (or any other location model). This can be re-written as \(Y - \mu \sim N[0,W]\). If we interpret this as a specification of \(p(Y|\mu)\) rather than, as originally intended, of \(p(\mu|Y)\), then, given \(H = \{\mu\}\), the posterior has been specified directly: \(p(Y|H) = p(Y|\mu)\) where \((Y|\mu) \sim N[\mu,W]\). Thus \(p(Y)\) is not required. Refining specifications of \(F(\cdot)\) to include further quantities \(g\), giving data \(H^* = \{\mu,g\}\), then

\[
p(Y|H^*) \propto p(Y|\mu)p(\mu|Y),
\]

with the second term following directly from the paper. See Pratt and Schlaifer (1985, and in reply to discussion by Lindley) for related ideas. This is simply an alternative approach to specifying (some features of) the joint distribution for \((Y,H)\) which is, after all, the central objective of work in this area.

REFERENCES IN THE DISCUSSION

