MODELING EXPERT OPINION: LIKELIHOODS UNDER INCOMPLETE PROBABILISTIC SPECIFICATION

by

Alan E. Gelfand

Bani K. Mallick

Dipak K. Dey

TECHNICAL REPORT No. 464

DECEMBER 9, 1992

Prepared Under Contract

N00014-92-J-1264 ((NR-042-267)

FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA 94305-4065

MODELING EXPERT OPINION: LIKELIHOODS UNDER INCOMPLETE PROBABILISTIC SPECIFICATION

by

Alan E. Gelfand

Bani K. Mallick

Dipak K. Dey

TECHNICAL REPORT No. 464

DECEMBER 9, 1992

Prepared Under Contract

N00014-92-J-1264 ((NR-042-267)

FOR THE OFFICE OF NAVAL RESEARCH

Professor Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA 94305-4065

Modeling Expert Opinion: Likelihoods Under Incomplete Probabilistic Specification

by

Alan E. Gelfand

Bani K. Mallick

and

Dipak K. Dey*

University of Connecticut

Abstract

Expert opinion is often sought with regard to unknowns in a decision–making setting. Our presumption is that such opinion is elicited as an incomplete probabilistic specification either in the form of probability assignments to fixed intervals or in the form of selected quantiles. We present likelihoods for such specification which arise through random mixtures of Beta distributions. We presume that a supra Bayesian presides over the opinion collection resulting in the posterior distribution as the mechanism for pooling opinion. The models are applied to opinion collected regarding points per game for participants in the 1991 NBA championship basketball series.

1. Introduction

Expert opinion is often sought with regard to unknowns in a decision–making setting in order to improve the quality of the decision making process. We presume that this opinion regards a univariate unknown denoted by θ, that it may be collected from several experts and that it is probabilistic in nature.

In this context, to date, most work, assumes that, at the individual level, an expert fully supplies a probability measure for θ. The literature on elicitation of a probability measure is, by now, substantial. See Kahneman, Slovic and Tversky (1982) for a readable review and Kadane et. al. (1980) for implementation suggestions. A naive approach is to insist that the individual probability measures for θ are members of a standard parametric family whence the expert need only supply its parameters. This seems far too restrictive. In fact full specification of a distribution for a continuous parameter θ seems to be more than we can reliably expect even an expert to provide. Rather we assume that each expert expresses belief regarding the unknown θ in a partial or incomplete way. That is, either probabilities are provided for a small collection of disjoint exhaustive intervals in the domain of θ or a small set of quantiles for the distribution of θ are provided.

One might seek incomplete specification in terms of the first few moments of the distribution of θ (Genest and Schervish, 1985) but moments seem less intuitive than probabilities or quantiles hence less reliably collected.

It is noteworthy that we have traded one modeling problem for another in that we now need to develop appropriate probabilistic models for the randomly collected incomplete expert opinion itself. This problem is the subject of our work and we address it primarily at the individual level. Our work builds upon that of Lindley (1985) and West (1988). Modeling the joint distribution of opinions is, in principle, no harder as we show. In our illustrative example we do have the simplification of independently collected opinion.

We adopt the supra Bayesian approach as our mechanism for combining expert opinions. The literature on normative approaches for the formation of aggregate opinion is substantial. See, for instance, the survey articles of French (1985), Genest and Zidek (1986) and Chatterjee and Chatterjee (1987). Many of the papers in this area presume an axiomatic specification of a set of properties which the aggregation mechanism must obey and then deduce the class of pooling functions meeting these properties. We do not engage in philosophical debate regarding selection of a pooling recipe. We are drawn to the supra Bayesian stance in that it adopts Bayes' rule as the pooling operator naturally producing combined opinion as the posterior distribution. Often there is an implicit external decision–maker who has his/her prior opinions, who gathers the experts' opinions and who is capable of calibrating the relative quality of each expert's opinion. We seek to help this decision maker specify a likelihood for this collection of opinions after which the Bayesian paradigm would enable the desired pooling.

We recognize that many decision making situations would not be of this type making required supra Bayesian specification difficult and arguing against its use. The supra Bayesian approach has its roots in Winkler (1968). The name was coined by Keeney and Raiffa (1976). Genest and Zidek (1986) provide insightful additional discussion.

Lindley (1985) and West (1988) articulate very clearly the key issues in the likelihood specification. In particular, specification of probabilities of disjoint exhaustive sets results in a discrete distribution for θ. Lindley essentially takes a multivariate normal distribution for the logits of these probabilities. If the supra Bayesian specifies a prior for θ which is also a discrete distribution over these sets, then Lindley observes that the posterior or combined opinion updates these prior probabilities in a linear fashion on the log scale. In the case of specification of quantiles, under a continuous distribution for θ, West develops a likelihood emanating from a Dirichlet process which implicitly determines the distribution for θ. Our contribution provides alternative, richer and thus possibly preferable, classes of likelihoods for opinions of each of the above types. Our starting point is the family of mixtures of Beta distributions, a dense collection of integrable functions on

[0,1] (see e.g., Diaconis and Ylvisaker, 1985). In section 2 we formalize our problem developing appropriate notation. In section 3 we consider likelihoods under discretized distributions for θ. In section 4 we consider likelihoods under quantile specification for θ. In both cases so-called missing data or marginal likelihoods emerge. In section 5 we describe extension to multiple experts. In Section 6 we show that computation required to develop posteriors from such likelihoods can be handled using the Gibbs sampler (see, e.g., Gelfand and Smith, 1990). In section 7 we offer an example involving opinions on team point totals in professional basketball games. In particular several experts were asked to supply information of the above sort with regard to the distribution of points per game they anticipated for the participants in the 1991 NBA championship basketball series. We present a synthesis of this opinion. The key features of our approach are summarized in section 8.

2. Notation and Preliminaries

Let us assume that θ is univariate and that its domain, Θ, is an interval in R^1. Consider a partition of Θ into k fixed sets determined by the points $a_0<a_1<\cdots<a_k$ where $a_0 = \inf\{\theta \in \Theta\}$, $a_k = \sup\{\theta \in \Theta\}$ and let $I_j = (a_{j-1}, a_j)$. An expert supplies a vector $\mathbf{p}^T = (p_1,\cdots,p_k)$ where p_j is the expert's opinion regarding the chance that $\theta \in I_j$. A collection of N expert opinions results in vectors $\mathbf{p}_1,\cdots,\mathbf{p}_N$ which we assemble into a $k\times N$ matrix P $= (\mathbf{p}_1,\cdots,\mathbf{p}_N)$. Also consider a set of k ordered probabilities $0<\alpha_1<\alpha_2<\cdots<\alpha_k<1$ denoted by $\boldsymbol{\alpha}$. Let $\mathbf{q}^T=(q_1,\cdots,q_k)$ where q_j is the expert's opinion as to the $\alpha_j{}^{th}$ quantile of θ's distribution. Of course $q_1<q_2<\cdots<q_k$ with $q_1>a_0$, $q_k<a_k$. Quantile specification can also include the often used "smallest, middle, largest" elicitation taking these as, for instance, the .005, .5, .995 quantiles respectively. Again with a collection of N experts we would obtain vectors $\mathbf{q}_1,\cdots,\mathbf{q}_N$ which we assemble into a $k\times N$ matrix $Q = (\mathbf{q}_1,\cdots,\mathbf{q}_N)$.

What sort of prior information will the supra Bayesian provide? It is possible that he/she may give a fully specified distribution for θ whose density we will denote by $f(\theta)$. It may be more appropriate to assume that the supra Bayesian offers the same sort of information that the expert does. In the case where a **p** is supplied by the expert we assume that the supra Bayesian also provides a vector of probabilities for the sets I_j which we denote by ρ. In the case where **q** is supplied the supra Bayesian provides a vector of quantiles γ for the same collection of α's. Hence there are four types of likelihoods which we shall consider. At the individual level these are: (i) $L(\mathbf{p}|\theta)$, (ii) $L(\mathbf{p}|\theta \in I_j)$, $j=1,\cdots,k$, (iii) $L(\mathbf{q}|\theta)$ and (iv) $L(\mathbf{q}|\theta \in (\gamma_{j-1},\gamma_j))$, $j=1,\cdots,k$. At the group level P replaces **p**, Q replaces **q**.

Note that likelihoods must be specified with regard to events for θ, in particular events whose probabilities can be computed using the supra Bayesian's prior opinion. In other words $L(\mathbf{p}|\rho)$ or $L(\mathbf{q}|\gamma)$ are not meaningful. Moreover the Bayesian synthesis updates the prior densities only for these events. In the case of (i) and (iii) the combined opinion yields a fully specified posterior density of for θ. In case (ii) revised probabilities for the intervals result. In case (iv) revised probabilities are associated with the γ_i but unfortunately no updated quantiles can be extracted. Such quantiles can be gotten under (iii).

An alternative approach to handle the incomplete specification engendered in **p** or **q** is to assume that the expert's opinion is modeled by a parametric exponential family having, in the case of **p**, a k–1 dimensional parameter or, in the case of **q**, a k dimensional parameter. Specification of **p** or of **q** then yields a system of equations which, in principle, determines the member of the family i.e. the expert's distribution. Our approach is nonparametric avoiding the selection of the family as well as possible difficulties in solving the system of equations.

One might suspect that, since θ is univariate, computing concerns are not an issue. In fact, the likelihoods introduced in Sections 3.2 and 4.2 are expressed as high dimensional integrals over "missing data". Computation of posterior distributions under such likelihoods is handled using the Gibbs sampler introduced as a Bayesian computing tool in

Gelfand and Smith (1990). Details are supplied in Section 6.

3. Likelihoods for discrete probability specification

In this section we focus on likelihoods when $\mathbf{p}$'s are collected i.e. the cases of $L(\mathbf{p}|\theta)$ and $L(\mathbf{p}|\theta \in I_j)$. Note that appropriate behavior for such likelihoods requires that when θ is in, say I_j, p_j should tend to be large but, as θ moves further away from I_j, p_j should tend to be smaller. In section 3.1 we describe a natural first attempt at modeling these likelihoods using the Dirichlet process. After noting several criticisms of this version we review Lindley's (1985) alternative. In section 3.2 we present a very broad class of likelihoods arising from mixture distributions.

3.1. Dirichlet process model; Lindley's model

Consider case (i) of section 2, $L(\mathbf{p}|\theta)$. The most natural *specification* for $L(\mathbf{p}|\theta)$ would be to assume that $\mathbf{p}$ is induced by some underlying probability measure G on (a_0, a_k). A well discussed mechanism for generating probability measures is the Dirichlet process (see Ferguson, 1973 or Antoniak, 1974). In the present case we need to draw G randomly given θ so we assume $G|\theta \sim DP(\lambda; G_{0,\theta})$ where the continuous distribution $G_{0,\theta}$ is the mean of G and λ is a precision parameter. For convenience we work with cdf's rather than probability measures i.e. $G_{0,\theta}(y)$ is the probability assigned to the interval (a_0,y). The induced distribution for $\mathbf{p}$, $\mathbf{p}|\theta$, is a Dirichlet distribution, $D(\lambda; \beta(\theta))$, where $\beta(\theta) = (\beta_1(\theta),\cdots,\beta_k(\theta))$ with $\beta_j(\theta) = (G_{0,\theta}(a_j) - G_{0,\theta}(a_{j-1}))$. We see that the precision parameter λ may be interpreted as reflecting (the supra Bayesian's) confidence in the expert's opinion. Small λ, i.e., small precision implies that this expert's $\mathbf{p}$ will tend to be less informative about θ than under large λ.

How shall we choose $G_{0,\theta}$? Consider $G_{0,\theta}(\cdot) = G_0(\cdot - \theta)$. Then for any fixed a, $G_{0,\theta}(a)$ decreases in θ. Hence $P(p_1 \leq b|\theta)$ increases in θ while $P(p_k \leq b|\theta)$ decreases in θ so that the likelihood exhibits appropriate behavior for extreme intervals. In addition suppose we take G_0 to be a distribution which is symmetric about 0 and assume that a_j are equal spaced, i.e., $a_j - a_{j-1}$ is constant $j = 2, 3, \cdots, k-1$. Then for p_j, $j = 2, \cdots, k-1$, $P(p_j \leq b \mid \theta)$ will increase as θ moves away from $m_j = (a_j + a_{j-1})/2$ in either direction, again appropriate behavior for the likelihood. In considering choices for G_0, e.g., normal, t, logistic, we can assume that the scale parameter $\sigma = 1$ or else σ and λ will not be identifiable in $p \mid \theta$. Then, the more heavy–tailed G_0 is the more probability will tend to be placed in the extreme intervals.

This likelihood is immediately adapted to the case (ii), $L(p \mid \theta \in I_j)$. The conditional distribution $p \mid \theta \in I_j$ is interpreted as $\underset{\sim}{p} \mid \theta = m_j$, i.e., $D(\lambda, \beta^{(j)})$ where $\beta^{(j)} \equiv (\beta_1^{(j)}, \cdots, \beta_k^{(j)})$ with $\beta_i^{(j)} = G_0(a_i - m_j) - G_0(a_{i-1} - m_j)$ with m_j as above and $m_1 < a_1$, $m_k > a_k$. (In fact, m_j can be an arbitrary point in (a_{j-1}, a_j)). This likelihood again exhibits proper behavior.

The Dirichlet process for generating random distributions has simplicity in its favor but may be criticized as follows. First, the only G's which can be generated under this process are discrete, an unappealing limitation. Secondly, with regard to the induced likelihood for $\underset{\sim}{p}$, as observed in discussion by Lindley (1988), probabilities for all intervals, hence for adjacent intervals, will be negatively correlated, i.e., $\text{cov}(p_i, p_{i+1} \mid \theta) < 0$. This reflects unsatisfying behavior for the likelihood in that if, for instance, θ encourages large p_i it should also encourage large p_{i-1} and p_{i+1}. Of course since $\Sigma p_j = 1$ at least some correlations amongst the p_i must be negative regardless of how the likelihood is specified. Moreover $k > 3$ is required in order that it be possible for all adjacent correlations to be positive.

To attempt to alleviate this problem Lindley (1985) suggests converting the probabilities to the log scale and using multivariate normal models. Most simply, if we let the ℓ_i be baseline logits, i.e., $\ell_i = \log p_i/p_k$, $i = 1, \cdots, k-1$, Lindley assumes that $\boldsymbol{\ell} = (\ell_1, \cdots, \ell_{k-1})$ given θ (or $\theta \in I_j$) is modeled as a multivariate normal whence covariances can be set as desired. Of course the covariances induced for **p** need not agree in sign with those of the corresponding $\boldsymbol{\ell}$ (Consider the case where all covariances amongst the ℓ_i are assumed nonnegative.) Moreover the collection of normal models for ℓ results in a limited collection of models for $p|\theta$ (additive logistic normal models in the terminology of Aitchison, 1986). In this regard see also Bernardo's (1985) discussion to Lindley's paper. Thus Lindley's formulation is not totally satisfying. We would be remiss however if we failed to articulate the advantages of Lindley's likelihood – mathematical convenience, ease of interpretation and a linear pooling mechanism on the log scale.

3.2 Mixture of Betas model

We now consider another class of likelihoods for $L(p|\theta)$ and $L(p|\theta \in I_j)$ which addresses the criticism of the previous section. We first remind the reader that any continuous density on [0,1] can be arbitrarily well approximated by a mixture of Beta densities. See, e.g., Diaconis and Ylvisaker (1985 p. 136) for a formal statement and proof. The order of the approximation is $r^{-\frac{1}{2}}$ where r is the number of terms in the mixture. Inversion by a fixed continuous c.d.f. yields a random distribution G on (a_0, a_k). Other mixture classes could be used as well. For instance mixtures of gamma densities can arbitrarily well approximate any continuous distribution on R^+. See Diaconis and Ylvisaker for more general discussion. The important point is the use of mixing.

Suppose we denote an r–component mixture of Beta densities by $\sum_{\ell=1}^{r} w_\ell \, Be(u|\eta_{1\ell}, \eta_{2\ell})$ where $\underset{\sim}{w} = (w_1,\ldots,w_r)$ is a vector on the r–dimensional simplex. Let $\eta_\ell = (\eta_{1\ell}, \eta_{2\ell})$ and $\underset{\sim}{\eta} = (\eta_1,\ldots,\eta_r)$. If $G_{0,\theta}(\cdot) = G_0(\cdot - \theta)$ is a specified c.d.f. such as in section 3.1 and if U is a drawn at random from this mixture, then consider the random variable $Y = G_{0,\theta}^{-1}(U) = \theta + G_0^{-1}(U)$. The distribution of Y, say $G(\cdot)|\mathbf{w}, \underset{\sim}{\eta}, \theta$, has the form

$$G(\cdot)|\mathbf{w}, \underset{\sim}{\eta}, \theta = P(Y \leq \cdot \;|\mathbf{w}, \underset{\sim}{\eta}, \theta) = P(U \leq G_0(\cdot - \theta)|\mathbf{w}, \underset{\sim}{\eta}). \tag{1}$$

A vector **p** arises from G as in the previous section. From (1)

$$p_i = P(U \in [G_0(a_{i-1} - \theta), G_0(a_i - \theta)] \mid \mathbf{w}, \underset{\sim}{\eta}) = \Sigma\, w_\ell\, h_{i\ell}(\eta_\ell, \theta) \tag{2}$$

where $h_{i\ell}(\eta_\ell, \theta)$ is the area under the $Be(u|\eta_{1\ell}, \eta_{2\ell})$ density between $G_0(a_{i-1} - \theta)$ and $G_0(a_i - \theta)$, i.e., the difference between two incomplete Beta functions. From (2)

$$p = H(\eta, \theta)\mathbf{w} \tag{3}$$

when $H(\eta, \theta)$ is a k×r matrix having $(i, \ell)^{th}$ entry $h_{i\ell}(\eta_\ell, \theta)$. Hence p arises as a linear transformation **w**. Note that the columns of $H(\eta, \theta)$ sum to 1.

In order to develop a likelihood $L(p, \theta)$ we need to introduce a random mechanism for p given θ. Suppose we assume that r, the "denseness" measure of our family of mixture distributions and η are specified but that w is random, i.e., $\sum_{\ell=1}^{r} w_\ell\, Be(u|\eta_{1\ell}, \eta_{2\ell})$ is an r component random mixture of fixed Beta densities. If w is random then G is and thus p is as well. The density of w denoted by $f_w(\cdot)$ is on the r-dimensional simplex. The book of Aitchison (1986) is devoted almost entirely to discussion of distributions on simplexes; in our illustration example we chose f_w to be a Dirichlet. Note that given r, for any set of $Be(u|\eta_{1\ell}, \eta_{2\ell})$, vectors p arising under (3) do not span the simplex. For instance clearly, $\min_j h_{ij}(\eta_j, \theta) \le p_i \le \max_j h_{ij}(\eta_j, \theta)$ with further constraints arising due to the restriction of **w** to the r–dimensional simplex. To remedy this problem, suppose we constrain or modify sampled opinion vectors p such that $p_i \ge \epsilon$ for all i. This is not much of a restriction; it only requires that every interval is assigned at least some minimum probability. But then, for r large enough, there will be an interval for θ in which $L(\mathbf{p}, \theta) > 0$. Practically, this means choosing r large relative to k and taking $L(p; \theta) = 0$ when θ is too extreme to permit p to be observed. Lastly, how shall we specify η? The result of Diaconis and Ylvisaker is not constructive in this regard. In practice, we have chosen the η_ℓ to yield a set of Beta densities which "cover" [0, 1]. For a given r, a flexible choice is $\eta_{1\ell} = \delta\{\epsilon(r+1-i) + (1-\epsilon)i\}$, $\eta_{2\ell} = \delta(\epsilon i, (1-\epsilon)(r+1-i)\}$. We suppress η in the subsequent notation.

We now obtain $L(p; \theta)$. Since $\Sigma p_i = 1$, $\Sigma w_\ell = 1$ we rewrite (3) as

$$\mathbf{p}^{(1)} = H^{(1)}(\theta)\mathbf{w}^{(1)} + \mathbf{h}_r^{(1)}(\theta) \tag{4}$$

where $p^{(1)T} = (p_1,\cdots p_{k-1})$, $w^{(1)T} = (w_1,\cdots w_{r-1})$. In (4), if we write $H(\theta) = (h_1(\theta)\cdots h_r(\theta))$ and let $h_\ell^{(1)}(\theta)$ be $h_\ell(\theta)$ with the last row omitted, then $H^{(1)}(\theta) = (h_1^{(1)}(\theta) - h_r^{(1)}(\theta), \cdots, h_{r-1}^{(1)}(\theta) - h_r^{(1)}(\theta))$. If we let $Z = w_1^{(1)}$ where $w_1^{(1)T} = (w_1, \cdots w_{r-k})$ then consider the linear transformation from $w^{(1)}$ into $(Z, p^{(1)})$, i.e.,

$$\begin{bmatrix} Z \\ p^{(1)} - h_r^{(1)}(\theta) \end{bmatrix} = A(\theta)\, w^{(1)} \tag{5}$$

where, if $H^{(1)}(\theta) = (H_1^{(1)}(\theta), H_2^{(1)}(\theta))$, $H_1^{(1)}$ being $(k-1) \times (r-k)$, $H_2^{(1)}$ being $(k-1) \times (k-1)$ then

$$A(\theta) = \begin{bmatrix} I_{r-k} & O \\ H_1^{(1)}(\theta) & H_2^{(1)}(\theta) \end{bmatrix}.$$

Thus, if $\pi^{(1)}(\theta) = p^{(1)} - h_r^{(1)}(\theta)$,

$$f(Z, p^{(1)} \mid \theta) = f_w\left(A^{-1}(\theta)\begin{bmatrix} Z \\ \pi^{(1)}(\theta) \end{bmatrix}\right) \cdot \left| \, |A(\theta)|^{-1} \right| \tag{6}$$

where $|A(\theta)| = |H_2^{(1)}(\theta)|$ and $A^{-1}(\theta) = \begin{bmatrix} I_{r-k} & O \\ -H_2^{(1)-1}(\theta)H_1(\theta) & H_2^{(1)-1}(\theta) \end{bmatrix}$. Finally

$$L(p;\theta) = \int_{C(p;\theta)} f(Z,p^{(1)} \mid \theta)dZ \tag{7}$$

where $C(p;\theta)$ denotes the restriction of Z resulting from (5) given p and θ with $w^{(1)}$ constrained to the r–dimensional simplex. We are assuming that p and θ are compatible i.e. $L(p,\theta) > 0$ whence from (7) the set $C(p;\theta)$ has positive Lebesgue measure. More precisely we have

$$0 \le A^{-1}(\theta)\begin{bmatrix} Z \\ \pi^{(1)}(\theta) \end{bmatrix} \text{ and } 1_{r-1}^T\, A^{-1}(\theta)\begin{bmatrix} Z \\ \pi^{(1)}(\theta) \end{bmatrix} \le 1$$

which simplify to

$$Z \ge 0, \quad H_2^{(1)-1}(\theta)\, H_1^{(1)}(\theta)Z \le H_2^{(1)-1}(\theta)\pi^{(1)}(\theta)$$

and

$$1^T_{r-k}Z - 1^T_{k-1}H_2^{(1)-1}(\theta)\, H_1^{(1)}(\theta)Z + 1^T_{k-1}H_2^{(1)-1}(\theta)\pi^{(1)}(\theta) \le 1. \tag{8}$$

We thus note that restriction to $C(\mathbf{p},\theta)$ provides linear constraints on $\mathbf{Z}$. The likelihood in (7) may be viewed as a "missing data" or marginalized likelihood. Computation of the posterior for θ under (7) with specification of the supra Bayesian's prior can be carried out using the Gibbs sampler. Details are provided in section 6. As in section 3.1, if ~~we seek~~ $L(\mathbf{p}|\theta \in I_j)$ is interpreted as $L(\underline{p}|m_j)$.

From (3) note that $P(p_1 \le b|\theta) = P(\Sigma\, w_\ell\, h_{1\ell}(\theta) \le b)$. As θ increases $h_{1\ell}(\theta)$ decreases for each ℓ. Thus the random variable $\Sigma\, w_\ell\, h_{1\ell}(\theta_1)$ is greater than the random variable $\Sigma\, w_\ell h_{1\ell}(\theta_2)$ if $\theta_1 < \theta_2$ and so $P(p_1 \le b|\theta)$ increases in θ. Similarly $P(p_k \le b|\theta)$ decreases in θ. The likelihood in (7) behaves appropriately for extreme intervals. Precise description of the behavior of $P(p_j \le b|\theta)$ as a function of θ depends upon the choice of c_ℓ and d_ℓ and on G_0. For the aforementioned choices of c_ℓ and d_ℓ, since each $h_{i\ell}(\theta)$ as a function of θ will necessarily increase to a unique maximum and then decrease, eventually, $P(p_j \le b|\theta)$ will increase as θ grows large or small, again desired behavior for the likelihood.

Finally, the mean of $\mathbf{p}$, $E(\mathbf{p}|\theta) = H(\theta)\, E(\mathbf{w})$ and the covariance matrix for $\mathbf{p}$, $\Sigma_p(\theta) = H(\theta)\, \Sigma_w\, H(\theta)^T$ where Σ_w is the covariance matrix of $\mathbf{w}$. Similar forms arise in the case of $L(\mathbf{p}|\theta \in I_j)$. Clearly the covariance structure available for $\mathbf{p}$ under (3) is much richer than under the Dirichlet model.

4. Likelihoods under quantile specification

In this section we develop likelihood functions for cases (iii) and (iv) of section 2, i.e., $L(\mathbf{q}|\theta)$ and $L(\mathbf{q}|\theta \in (\gamma_{j-1},\gamma_j))$ respectively. How should these likelihoods behave in order to be sensible? They should be such that, with increasing θ, q_i tends to increase and such that any pair q_i, q_j are positively correlated. In section 4.1 we review West's (1988) approach based upon the Dirichlet process while in section 4.2 we start with random mixtures of Beta densities to create these likelihoods.

4.1. West's model

Recall that, for the set of probabilities $0<\alpha_1<\alpha_2\cdots<\alpha_k<1$, the expert specifies a vector of quantiles $\mathbf{q} = (q_1,\cdots q_k)$. West proposes a density for $\mathbf{q}$ given θ which arises as follows. Suppose we think of $\mathbf{q}$ as being the quantiles of a random distribution G given θ. Suppose we again consider a baseline c.d.f. $G_{0,\theta}(\cdot) = G_0(\cdot - \theta)$ as in section 3. Let F be a distribution (c.d.f.) derived from a Dirichlet process on [0,1]. Then the random distribution, G given θ is defined implicitly through $F^{-1}(\cdot) = G_{0,\theta}(G^{-1}(\cdot))$. In particular, if π_j is the α_j^{th} quantile of F i.e. $\pi_j=F^{-1}(\alpha_j)$, $j=1,\cdots,k$ then $q_j = G_{0,\theta}^{-1}(\pi_j) = G_0^{-1}(\pi_j) + \theta$. Observe that the Dirichlet process which generates F need not be specified; rather we need only an ordered Dirichlet distribution say $D(\mathcal{E})$ which generates $\boldsymbol{\pi} = (\pi_1,\cdots,\pi_k)$. Then $\mathbf{q}$ is obtained from $\boldsymbol{\pi}$ by inverting $G_{0,\theta}$. The resultant likelihood is

$$L(\mathbf{q}|\theta)\ \alpha\ (1-G_{0,\theta}(q_1))^{\mathcal{E}_{k+1}} \prod_{i=2}^{k} (G_{0,\theta}(q_i)-G_{0,\theta}(q_{i-1}))^{\mathcal{E}_i}. \tag{9}$$

Though not mentioned by West, we can immediately modify (9) to a likelihood for $\mathbf{q}$ given $\theta \in (\gamma_{j-1}, \gamma_j)$ by replacing $G_{0,\theta}$ with G_{0,n_j} where the n_j are selected analogously to the m_j in section 3.1.

West comments that the likelihood (9) behaves appropriately. In fact since $P(q_i<b(\theta) = P(G_0^{-1}(\pi_i) + \theta < b)$ decreases in θ, q_j tends to increase in θ. Moreover, $\mathrm{cov}(q_i,q_j) = \mathrm{cov}(G_0^{-1}(\pi_i), G_0^{-1}(\pi_j)) > 0$ if and only if $\mathrm{cov}(\pi_i,\pi_j) > 0$. Since the π_i, π_j arise from $\mathcal{D}(\mathcal{E})$ we can calculate that, if $\tilde{\mathcal{E}}_i = \sum_{\ell=1}^{i} \mathcal{E}_\ell$, $\tilde{\mathcal{E}}_j = \sum_{\ell=i+1}^{j} \mathcal{E}_\ell$,

$$\mathrm{cov}(\pi_i, \pi_j) = \frac{\tilde{\mathcal{E}}_i \left[\sum_{\ell=1}^{k+1} \mathcal{E}_\ell - \tilde{\mathcal{E}}_i - \tilde{\mathcal{E}}_j\right]}{\left[\sum_{\ell=1}^{k+1} \mathcal{E}_\ell\right]^2 \left[\sum_{\ell=1}^{k+1} \mathcal{E}_\ell + 1\right]} > 0.$$

4.2. Mixture of Betas model

Here we parallel the development of section 3.2. A distribution on [0,1] which is a random mixture of Beta distributions is selected and then inverted to a distribution on (a_0, a_k). Then q_i is taken as the α_i^{th} quantile of the latter distribution. In particular, if $U|\mathbf{w} \sim \Sigma\, w_\ell\, Be(u|\eta_{1\ell}, \eta_{2\ell})$ and φ_i is the α_i^{th} quantile of $U|\mathbf{w}$, $i=1,2,\cdots,k$, analogous to section 4.1, let $q_i = G_{0,\theta}^{-1}(\varphi_i) = G_0^{-1}(\varphi_i) + \theta$. Since $\mathbf{w}$ is random so is $\varphi = (\varphi_1,\cdots,\varphi_k)$ and hence $\mathbf{q}$. The q_i are the quantiles of the distribution of the variable $Y = G_{0,\theta}^{-1}(U)$ given $\mathbf{w}$. This random distribution is given in (1). Using obvious notation, we have $\varphi = G_0(\mathbf{q}-\theta 1_k)$ where $G_0(\mathbf{q}-\theta 1_k)^T = (G_0(q_1-\theta),\cdots,G_0(q_k-\theta))$.

What is the likelihood $L(\mathbf{q}|\theta)$? Given $\mathbf{w}$, $\alpha_i = \int_0^{\varphi_i} \Sigma\, w_\ell\, Be(u|\eta_{1\ell}, \eta_{2\ell})du$ $= \sum_{\ell=1}^{r} w_\ell\, b_\ell(\varphi_i)$ where $b_\ell(\varphi_i) = \int_0^{\varphi_i} Be(u|\eta_{1\ell}, \eta_{2\ell})du$, $i=1,2,\cdots,k$. Hence

$$\alpha = B(\varphi)\mathbf{w} \tag{10}$$

where $B(\varphi)$ is $k \times r$ with $B(\varphi)_{i\ell} = b_\ell(\varphi_i)$. Analogous to (4) we rewrite (10) as

$$\alpha = B^{(1)}(\varphi)\, \mathbf{w}^{(1)} + \mathbf{b}_r(\varphi) \tag{11}$$

where, if $B(\varphi) = (\mathbf{b}_1(\varphi),\cdots \mathbf{b}_r(\varphi))$, $B^{(1)}(\varphi) = (\mathbf{b}_1(\varphi) - \mathbf{b}_r(\varphi),\cdots \mathbf{b}_{r-1}(\varphi) - \mathbf{b}_r(\varphi))$.

Again we introduce missing variables $\mathbf{Z} = (w_1,\cdots w_{r-1-k})$ and consider the transformation form $\mathbf{w}^{(1)}$ to $(\mathbf{Z},\varphi)$, i.e.,

$$\begin{bmatrix}\mathbf{Z} \\ \alpha-\mathbf{b}_r(\varphi)\end{bmatrix} = D(\varphi)\mathbf{w}^{(1)},\ \mathbf{w}^{(1)} = D^{-1}(\varphi)\begin{bmatrix}\mathbf{Z} \\ \alpha-\mathbf{b}_r(\varphi)\end{bmatrix} \tag{12}$$

where, if $B^{(1)}(\varphi) = (B_1^{(1)}(\varphi), B_2^{(1)}(\varphi))$, $B_1^{(1)}$ being $k\times(r-1-k)$, $B_2^{(1)}$ being $(r-1-k) \times (r-1-k)$ then

$$D(\varphi) = \begin{bmatrix} I_{r-1-k} & O \\ B_1^{(1)}(\varphi) & B_2^{(1)}(\varphi)\end{bmatrix},\ D^{-1}(\varphi) = \begin{bmatrix} I_{r-1-k} & O \\ -B_2^{(1)-1}(\varphi)B_1^{(1)}(\varphi) & B_2^{(1)-1}(\varphi)\end{bmatrix}.$$

Thus, if $\tau(\varphi) = \boldsymbol{\alpha} - \mathbf{b}_r(\varphi)$, $f(\mathbf{Z},\varphi) = f_w(D^{-1}(\varphi)\begin{bmatrix}\mathbf{Z} \\ \tau(\varphi)\end{bmatrix}) \cdot \left|J_{\mathbf{w}}(1)_{\to(\mathbf{Z},\varphi)}\right|$

and therefore

$$f(\mathbf{Z},\mathbf{q}|\,\theta) = f_w(D^{-1}(G_0(\mathbf{q}-\theta 1_k))\begin{bmatrix}\mathbf{Z} \\ \tau(\varphi)\end{bmatrix}) \cdot$$

$$\left|J_{\mathbf{w}}(1)_{\to(\mathbf{Z},\varphi)}\right|_{\varphi=G_0(\mathbf{q}-\theta 1_k)} \cdot \prod_{j=1}^{k} \frac{dG_0(q_j-\theta)}{dq_j}. \tag{13}$$

Finally

$$L(\mathbf{q},\theta) = \int_{C(\mathbf{q},\theta)} f(\mathbf{Z},\mathbf{q}|\theta)\, d\mathbf{Z} \tag{14}$$

where $C(\mathbf{q},\theta)$ denotes the restriction of $\mathbf{Z}$ resulting from (12) given $\mathbf{q}$ and θ with $\mathbf{w}^{(1)}$ constrained to the k–dimensional simplex. We are assuming that $\mathbf{q}$ and θ are compatible i.e., $L(\mathbf{q},\theta) > 0$ whence $C(\mathbf{q},\ \theta)$ has positive Lebesgue measure. Analogous to (8) these become

$$\mathbf{Z} \geq 0,\ B_2^{(1)-1}(\mathbf{q}-\theta 1_k)\ B_1^{(1)}(\mathbf{q} - \theta 1_k))\ \mathbf{Z} \leq B_2^{(1)-1}(\mathbf{q}-\theta 1_k)\ \tau(\varphi) \tag{15}$$

and

$$1^T_{r-1-k}\,\mathbf{Z} - 1_k^T\, B_2^{(1)-1}(\mathbf{q} - \theta_1^{(1)} 1_k) + B_1^{(1)}(\mathbf{q} - \theta 1)\mathbf{Z} + 1_k^T\, B_2(\mathbf{q} - \theta 1_k)\ \tau(\varphi) \leq 1.$$

Importantly, the restriction to $C(\mathbf{q},\theta)$ still provides linear constraints on $\mathbf{Z}$.

The Jacobian matrix in (13), $J_{\mathbf{w}}(1)_{\to(\mathbf{Z},\varphi)}$, has entries $\partial w_i/\partial Z_j$ and $\partial w_i/\partial \varphi_j$ which appear to be straightforward to calculate using (12). The difficulty is that $D^{-1}(\varphi)$ can not be obtained explicitly so that it can not be differentiated appropriately with respect to φ_j and then evaluated at $G_0(\mathbf{q}\ \ -\theta 1_k)$. In section 6 we show how the calculation of J can be simplified using an implicit function theorem.

As in (7), the likelihood in (14) may be viewed as a "missing data" or marginalized likelihood. Computation of the posterior for θ under (14) with specification of the supra Bayesian's prior can be carried out using the Gibbs sampler. Details are provided in section 6. Modification for $L(\mathbf{q} | \theta \in (\gamma_{j-1}, \gamma_j))$ replaces θ in (14) with n_j.

Turning to the behavior of (14), from (10), $P(q_i < b | \theta) = P(\varphi_i < G_0(b-\theta))$ which decreases in θ so q_j tends to increase in θ. Clearly cov $(q_i, q_j) > 0$ if and only if $\mathrm{cov}(\varphi_i, \varphi_j) > 0$. For a given α_i, α_j and two $\mathbf{w}$ vectors say $\mathbf{w}_1$ and $\mathbf{w}_2$, let $\varphi_i^{(1)}$, $\varphi_j^{(1)}$ be the quantiles arising from $\mathbf{w}_1$, $\varphi_i^{(2)}$, $\varphi_j^{(2)}$arising from $\mathbf{w}_2$. Consider $\varphi_j^{(2)} - \varphi_j^{(1)} = (\varphi_i^{(2)} - \varphi_i^{(1)}) + (\varphi_i^{(1)} - \varphi_j^{(1)}) + (\varphi_j^{(2)} - \varphi_i^{(2)})$. For $\alpha_j - \alpha_i$ small, by the continuity of the $b_\ell(\varphi)$, the second and third terms will be small and offsetting. Thus $\mathrm{sgn}(\varphi_j^{(2)} - \varphi_j^{(1)})$ will tend to be the same as sgn $(\varphi_i^{(2)} - \varphi_i^{(1)})$ whence averaging over $\mathbf{w}$ will result in $\mathrm{cov}(\varphi_i, \varphi_j) > 0$ for close quantiles.

5. Several experts

Typically in the elicitation of expert opinion more than one source is tapped. Suppose that N opinions are collected. Often the supra Bayesian will obtain independent evaluations of the distribution of θ whence we obtain the likelihoods

$$L(\mathbf{P} | \theta) = \prod_{i=1}^{N} L(p_i | \theta)$$

$$L(\mathbf{Q} | \theta) = \prod_{i=1}^{N} L(q_i | \theta) \, . \tag{16}$$

Dependence amongst opinions may arise in a variety of ways. For instance, individual opinion may be solicited after group deliberation or perhaps the opinion of the t^{th} expert is obtained after he/she has been shown the opinion of the previous t-1 experts, $t=2,\cdots,N$.

Is there a convenient way to extend the likelihoods (7) and (14) to accommodate dependence? The most straightforward solution is to consider $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N)$ to be a collection of N dependent vectors each on the simplex, having joint density $f_W(\mathbf{W})$. In the case of **P** we consider, for each p_i, a transformation as in (5), $\begin{bmatrix} Z_i \\ p_i^{(1)} \end{bmatrix} = A(\theta)w_i^{(1)}$.

Extending (6) in the obvious way results in

$$L(P|\theta) = \int_{C(P,\theta)} f(Z_1,p_1^{(1)},Z_2,p_2^{(1)},\cdots Z_N,p_N^{(1)}|\theta) \prod_{i=1}^{N} dZ_i \tag{17}$$

where $C(P,\theta)$ denotes the collection of restrictions, $Z_i \in C(p_i, \theta)$, $i = 1,2,\cdots,N$. Modification for $L(P|\theta \in I_j)$ is obvious. Using (3) if $\Sigma_w^{(i,j)}$ denotes the covariance between w_i and w_j then $cov(p_i, p_j) = H(\theta)\, \Sigma_w^{(i,j)}\, H^T(\theta)$.

In the case of $\underset{\sim}{Q}$ we consider the associated matrix $\Phi_{k\times N} = (\varphi_1,\cdots\varphi_N)$ where $\varphi_i = G_0(q_i - \theta 1_k)$. For each φ_i we envision a transformation as in (12), $\begin{bmatrix} Z_i \\ \tau(\varphi_i) \end{bmatrix} = D(\varphi_i)w_i^{(1)}$.

Extending (13) in the obvious way results in

$$L(Q|\theta) = \int_{C(Q,\theta)} F(Z_1,q_1,Z_2,q_2,\cdots Z_N,q_N|\theta) \prod_{i=1}^{N} dz_i \tag{18}$$

where $C(Q,\theta)$ denotes the collection of restriction, $Z_i \in C(q_i,\theta)$, $i=1,2,\cdots N$. Modification for $L(Q|\theta \in (\gamma_{j-1},\gamma_j))$ is obvious. It is not possible to explicitly obtain $cov(q_i, q_j)$.

Convenient mechanisms for producing correlated vectors $w_1, \cdots, w_N$ include (i) creating $V_1, \cdots, V_N$ correlated multivariate normal vectors and then transforming to the simplex by an inverse logit transformation or (ii) creating correlated Dirichlet vectors by using common gamma variables in defining them.

6. Computational Issues

To carry out the Bayesian synthesis we first require specification of the supra Bayesian's prior. If the supra Bayesian supplies a complete prior distribution for θ, say $f(\theta)$, the likelihood×prior form is either

$$L(p|\theta)\, f(\theta) \ \text{ or } \ L(q|\theta)\, f(\theta). \tag{19}$$

If the supra Bayesian provides a vector ρ of probabilities for the intervals I_j the associated prior is in fact a multinomal trial, i.e., the prior probability of the event $\theta \in I_j$ is ρ_j. Similarly, if the supra Bayesian provides a vector of quantiles γ, the associated prior is again a multinomial trial where the probability of the event $\theta \in (\gamma_{j-1}, \gamma_j)$ is $\alpha_j - \alpha_{j-1}$. Now the likelihood×prior form is either

$$L(\mathbf{p} \mid \theta \in I_j) \cdot \rho_j \quad \text{or} \quad L(\mathbf{q} \mid \theta \in (\gamma_{j-1}, \gamma_j)) \cdot (\alpha_j - \alpha_{j-1}). \tag{20}$$

For (19), under the likelihoods (7) and (14) respectively the posterior distribution for θ takes the not too promising form of a ratio of high dimensional integrals. For (20) all integrals with respect to $\mathbf{Z}$ remain but integration with respect to θ is replaced by summation over j. Notice, however, that the *joint* posterior distribution of θ and $\mathbf{Z}$ given $\mathbf{p}^{(1)}$ is proportional to $f(\mathbf{Z}, \mathbf{p}^{(1)} \mid \theta) \cdot f(\theta)$ and we seek the *marginal* posterior of θ given $\mathbf{p}^{(1)}$. Similarly, the *joint* posterior distribution of θ and $\mathbf{Z}$ given $\mathbf{q}$ is proportional to $f(\mathbf{Z}, \mathbf{q} \mid \theta) \cdot f(\theta)$ and we seek the *marginal* posterior of θ given $\mathbf{q}$.

The Gibbs sampler, introduced as a Bayesian computing tool in Gelfand and Smith (1990) and Gelfand et al (1990), is a Markov chain Monte Carlo approach for producing samples from the joint posterior distribution hence from a desired marginal posterior. In the context of (19) we would obtain θ_j^*, $j = 1, \cdots, m$ as a sample from $f(\theta \mid \mathbf{p})$ or from $f(\theta \mid \mathbf{q})$ respectively. Appropriate summaries of this sample (kernel density estimates, moments, quantiles, etc.) provide estimates of desired features of the posterior of θ. In the context of (20) we obtain frequencies with which $\theta \in I_j$ or $\theta \in (\gamma_{j-1}, \gamma_j)$. These frequencies are converted to proportions to update ρ_j or $\alpha_j - \alpha_{j-1}$ respectively. We do not describe the Gibbs sampler here, referring the reader to the aforementioned papers. Rather we clarify the complete conditional densities which must be sampled in order to implement the Gibbs sampler.

Consider $f(\theta, \mathbf{Z} \mid \mathbf{p}^{(1)}) \propto f(\mathbf{Z}, \mathbf{p}^{(1)} \mid \theta) \cdot f(\theta)$. To simplify matters we discretize θ to a fine grid. This converts the interval of θ's compatible with $\mathbf{p}$, i.e., $\{\theta: L(\mathbf{p}, \theta) > 0\}$ to a discrete set say $\boldsymbol{\theta}_{\mathbf{p}}$. For each $\theta \in \boldsymbol{\theta}_{\mathbf{p}}$ the set $C(\mathbf{p}, \theta)$ has positive measure. The Gibbs sampler simulates realizations of a Markov chain on this (θ, Z) space. If r is large enough or the domain of θ is small enough, the chain will be irreducible and convergence of the sampler is assured.

Consider sampling Z. Suppose, as we do in the example of section 7, that $f_w(\mathbf{w})$ is taken to be a Dirichlet density i.e. $f_w(\mathbf{w}) \propto \prod_{i=1}^{r} w_i^{\eta_i}$. Then, from (6) $f(Z_i | Z_{-i}, p, \theta)$, the complete conditional distribution for Z_i is a product of terms restricted to a cross sectional set of Z_i arising from $C(p, \theta)$ given Z_{-i}. That is, if θ, p and all Z's are specified save Z_i, then apart from $Z_i \geq 0$, (8) imposes k+1 constraints on Z_i. These constraints are readily obtained and result in restriction of Z_i to a nonempty interval. We sample Z_i restricted to this interval using an approximate inverse c.d.f. method.

To sample θ given Z and p we evaluate $f(Z, p^{(1)} | \theta)) \cdot f(\theta)$ at Z and p for each θ in Θ_p. We then draw θ from the resulting discrete distribution on Θ_p. Iterating the Gibbs sampler using m parallel replications results in $(\theta_j^{*(t)}, Z_j^{*(t)})$, $j=1,\cdots m$ at the t^{th} iteration. Convergence is assessed using suggestions in Gelfand et al (1990) and the resulting set θ_j^*, $j=1,\cdots m$ is retained to summarize the posterior of θ given p. Developing samples from the posterior $f(\theta, Z | q)$ is handled in the same manner using (15) and (13) in place of (8) and (6) respectively. Implementing the Gibbs sampler for (20) is also handled similarly. Since the specifications $\theta \in I_j$, $\theta \in (\gamma_{j-1}, \gamma_j)$ are treated by setting $\theta = m_j$, $\theta = n_j$ respectively a very coarse grid for θ results.

Implementation of the Gibbs Sampler as described above requires repeated evaluation of (6) and (8) at specified θ, Z and p or of (13) and (15) at specified θ, Z and q. For (6) and (8) only calculation of incomplete Beta functions is required to obtain $H(\theta)$ hence $H^{(1)}(\theta)$, $H_1^{(1)}(\theta)$ and $H_2^{(1)}(\theta)$. For (13) and (15) again computation of $B(\varphi)$ only requires calculation of incomplete Beta functions. Because k is not large computation of $|H_2^{(1)}|$, $H_2^{(1)-1}$, $|B_2^{(1)}|$ and $B_2^{(1)-1}$ should present no problem. However, as r increases, for any ℓ, the ℓ^{th} and $\ell+1^{st}$ columns of $H(\theta)$ and of $B(\varphi)$ will become almost identical. To avoid having $H_2^{(1)}$ and $B_2^{(1)}$ which are essentially singular, in practice we define them by using roughly equally spaced columns from H and B rather than using consecutive ones.

Let us consider computation of $J_{w^{(1)}\to(Z,\varphi)}$ at a given Z and $\varphi = G_0(q-\theta 1_k)$. From (12) we see that

$$J_{w^{(1)}\to(Z,\varphi)} = \begin{bmatrix} I_{r-1-k} & O \\ -B_2^{(1)-1}(\varphi)B_1^{(1)}(\varphi) & E(\varphi) \end{bmatrix} \tag{21}$$

where $E(\varphi)$ is a k×k matrix such that $(E(\varphi))_{ij} = \dfrac{\partial w_{2i}^{(1)}}{\partial \varphi_j}$. Here $w^{(1)T} = (w_1^{(1)}\ w_2^{(1)})$ where $w_1^{(1)}$ is (r-1-k)×1 and $w_2^{(1)}$ is k×1. The relationship $B_2^{(1)}(\varphi)w_2^{(1)} + B_1^{(1)}(\varphi)Z - \tau(\varphi) = 0$ suggests the use of an implicit function theorem to calculate the $(E(\varphi))_{ij}$. In particular

$$\frac{\partial w_{2i}^{(1)}}{\partial \varphi_j} = -\frac{|B_{2(i,j)}^{(1)}(\varphi)|}{|B_2^{(1)}(\varphi)|} \tag{22}$$

where $B_{2(i,j)}^{(1)}(\varphi)$ is the matrix $B_2^{(1)}(\varphi)$ with the i^{th} column replaced by the vector S_j where S_j is a k×1 vector having all zeroes except in the j^{th} row where the entry is

$$s_j = \sum_{\ell=1}^{r} w_\ell \, Be(\varphi_j | \eta_{1\ell}, \eta_{2\ell}). \tag{23}$$

In (23) for a given Z and φ we obtain $w^{(1)}$ hence the w_ℓ while $Be(\varphi_j | \eta_{1\ell}, \eta_{2\ell})$ is merely a Beta density evaluated at (φ_j). From (21) and (22) we see that, regardless of r, in evaluating $|J_{w^{(1)}\to(Z,\varphi)}|$ we need never work with larger than a k×k determinant.

We see that computation is somewhat easier with **p** than with **q**. However in our experience the **q**'s are more reliably elicited than the **p**'s. Several of the experts struggled to comfortably assign to the intervals probabilities which summed to 1.

7. An illustrative example

Our example involves the development of a prior for the number of points that the participants in the 1991 National Basketball Association championship series would score in a series game. The two contestants were the Chicago Bulls and the Los Angeles Lakers. As the supra Bayesian's opinion for a given team we took the distribution of points scored by the team in each of the 82 regular season games of the 1990–91 season. The data were graciously supplied to us in summary form by the Bull and Laker organizations. Histograms for these point totals are provided for each team in Figure 1. Encouraged by individual normal plots of these samples we took $f(\theta)$ to be N(110.22, 192.10) for the Bulls, N(106.30, 151.78) for the Lakers based upon $\bar{X}$ and S^2 for the respective samples. Table 1 provides, for each team, the proportions of these point totals in the five intervals ≤ 90, 91–100, 101–110, 111–120, ≥ 121. Table 1 also provides, for each team, the 25^{th}, 50^{th} and 75^{th} quantiles of these point totals. These are the supra Bayesians ρ and γ vectors.

The experts included faculty, students and basketball players (not mutually exclusive sets). Their opinions were collected independently during the two day window in May 1991 between the end of the semi–final series (so that the participants in the championship series were identified) and the start of the championship series. We have a total of N = 5 experts and their opinions are presented in Table 2.

Our analysis is based exclusively on likelihoods of form (7) (with modification for the case $\theta \in I_j$) and (14) (with modification to the case $\theta \in (\gamma_{j-1}, \gamma_j)$) for each of the 5 independent judges. For (7) we have k=5, $m_j = 75+10j$, $j=1,\cdots,5$ and $G_0 = N(0, (10)^2)$ suggesting a roughly 60 point range between high and low point total. We set r=10 and took $\delta=1$, $\epsilon=.002$ for Chicago, $\delta=1$, $\epsilon=.002$ for Los Angeles with, for both teams, $f_w(w) = D(2,2,\cdots,2)$. The updated priors for θ are shown in the form of histograms in figures 2a and 2b. The updated probabilities for the intervals ≤ 90, 91–100, etc. are given in Table 3 and may be compared with Table 1. For (14) we have k=3. We chose the n_j to correspond to the $(Z_{j-1})/8$ quantile of the samples in Figure 1, resulting in $n_1 = 95$ $n_2 =$

105 n_3 = 113 n_4 = 129 for the Bulls with n_1 = 92 n_2 = 103 n_3 = 111 n_4 = 120 for the Lakers. We again set r=10 and took δ=1, ϵ=.004 for Chicago, δ=1, ϵ=.004 for Los Angeles with again $f_w(w) = D(2,2,\cdots,2)$. The resulting updated priors for θ are given in figures 2c and 2d. The updated probabilities associated with γ_1, γ_2 and γ_3 are .142, .661, and 1 for the Bulls with .166, .7 and .9766 for the Lakers. The championship series ultimately went five games with scores given in Table 4.

8. Summary

We have supplied a framework within which to model and aggregate expert opinion. Our development, though more demanding than previous ones, is attractive in providing

(i) a likelihood for expert opinion appropriate for the most reliably elicited forms of opinion

(ii) a rich family of likelihoods for such opinion

(iii) immediate extension to modeling collections of expert opinion

(iv) compatibility with the supra Bayesian's knowledge about the process and about the experts

(v) a natural pooling mechanism for synthesizing all the information.

References

Aitchison, J. (1986). The Statistical Analysis of Compositional Data. London/New York: Chapman and Hall.

Antoniak, C.E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* 2, 1152–1174.

Bernardo, J. (1985). Discussion of "Reconciliation of discrete probability distributions". In: Bayesian Statistics 2, eds. J.M. Bernardo et. al. 388–389, North Holland, Amsterdam.

Chatterjee, S. and Chatterjee, S. (1987). On combining expert opinions, American

Journal of Mathematical and Management Sciences, 7, (3 and 4) 271–295.

Devroye, L. (1986). *Non–Uniform Random Variate Generation*, New York, Springer–Verlag.

Diaconis, P. and Ylvisaker, D. (1985). Quantifying prior opinion. In: Bayesian Statistics 2, eds. J.M. Bernardo et. al. 133–156. North Holland, Amsterdam.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* 1, 209–230.

French, S. (1985). Group consensus probability distributions: a critical survey. In: *Bayesian Statistics* 2, eds. J.M. Bernardo et. al. 183–201, North Holland, Amsterdam.

Gelfand, A. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. J. Amer. Statist. Assoc. 85, 398–409.

Gelfand, A., Hills, S.E., Racine, A. and Smith A.F.M. Illustration of Bayesian inference in normal data models using Gibbs sampling. J. Amer. Statist. Assoc. 85, 972–985.

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

Genest C. and Zidek, J. (1986). Combining probability distributions: a critique and an annotated bibliography (with discussion). *Statistical Science* 1, 114–148.

Kadane, J., Dickey, J., Winkler, R., Smith, W., and Peters S. (1980). Interactive elicitation of opinion for a normal linear model. J. Amer. Statist. Assoc. 75, 845–854.

Kahneman, D., Slovic, P., and Tversky, A. (1982)*Judgment Under Uncertainty: Heuristics and Biases.* Cambridge University Press, New York.

Keeney, R.L., and Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs.* New York: John Wiley & Sons.

Lindley, D.V. (1985). Reconciliation of discrete probability distributions. In: Bayesian Statistics, 2, eds. J.M. Bernardo, et. al. 375–390, North Holland, Amsterdam.

West, M. (1988). Combining expert opinion. In: Bayesian Statistics, 3, eds. J.M. Bernardo, et. al. 375–390, North Holland, Amsterdam.

Winkler, R. (1968). The consensus of subjective probability distributions, Management Sci., 15, 61–75.

Table 1: Summary of point totals for Chicago and for Los Angeles for the 82 regular season games of the 1990–91 season.

	Intervals					Quantiles		
	≤90	90–100	101–110	111–120	>120	1st	2nd	3rd
Chicago	.073	.122	.378	.207	.220	101.75	108.00	118.50
Los Angeles	.110	.195	.305	.293	.097	99.00	107.00	115.00

Chicago $\bar{X}$ (prior mean) = 110.22 S^2 (prior variance) = 192.10

Los Angeles $\bar{X}$ (prior mean) = 106.30 S^2 (prior variance) = 151.78

Table 2: Probability and quantile elicitation of experts

Expert	Intervals ≤90	90–100	101–110	111–120	>120		Quantiles 1st	2nd	3rd
Chicago:									
1	.05	.15	.50	.25	.05		102	108	117
2	.106	.394	.394	.099	.007		94.6	100	105.4
3	.05	.20	.45	.20	.10		98	104	112
4	.10	.18	.35	.30	.07		95	105	113
5	.05	.15	.30	.30	.20		90	105	115
pmin	.05	.15	.30	.099	.006	qmin	90	100	105.4
pmax	.106	.394	.50	.30	.20	qmax	102	108	117

Los Angeles:

Expert	≤90	90–100	101–110	111–120	>120		1st	2nd	3rd
1	.05	.15	.40	.35	.05		102	109	117
2	.354	.455	.174	.016	.001		87.6	93.0	98.4
3	.05	.10	.45	.25	.15		103	107	115
4	.13	.28	.32	.22	.05		92	100	110
5	.1	.3	.3	.2	.1		85	100	110
pmin	.05	.1	.174	.016	.001	qmin	85	93	98.4
pmax	.354	.455	.45	.35	.15	qmax	103	109	117

Table 3: Updated probabilities for intervals

	≤90	90–100	101–110	111–120	>120
	Intervals				
Chicago	0	.21	.52	.27	0
Los Angeles	0	.27	.47	.26	0

Chicago posterior mean = 107.5 posterior variance = 22.34

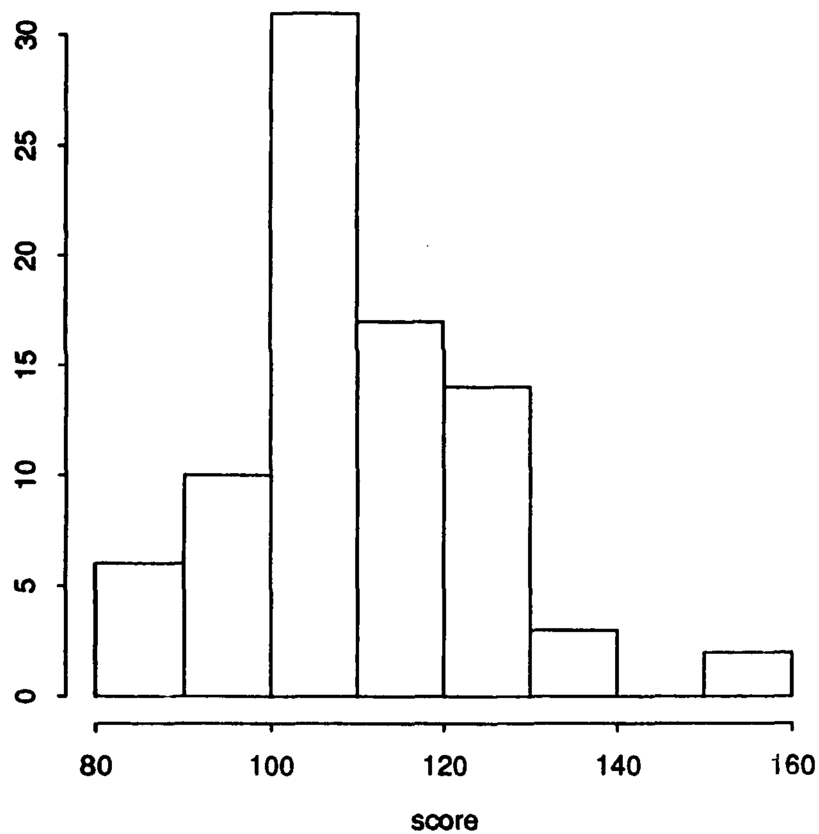
Los Angeles posterior mean = 105.8 posterior variance = 18.55

Table 4: Scores for the championship series

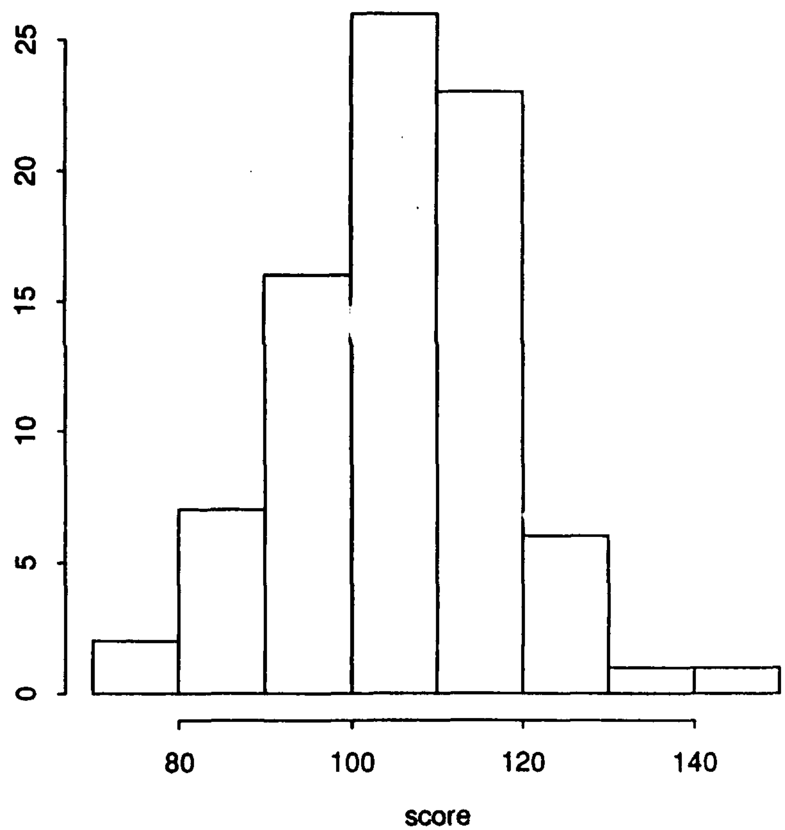
Game	Chicago	Los Angeles
1	91	93
2	107	86
3	104	96
4	97	82
5	108	101

Chicago wins series 4–1.

Figure 1: Histogram of points per game for Chicago and Los Angeles, 1990-91 NBA Season

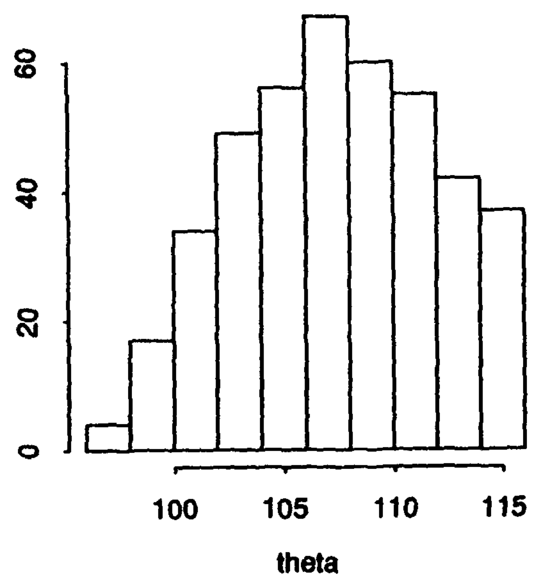


1 a) Chicago

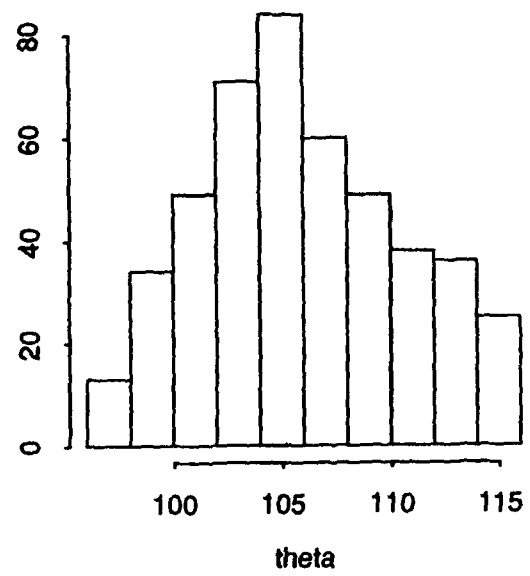


1 b) Los Angeles

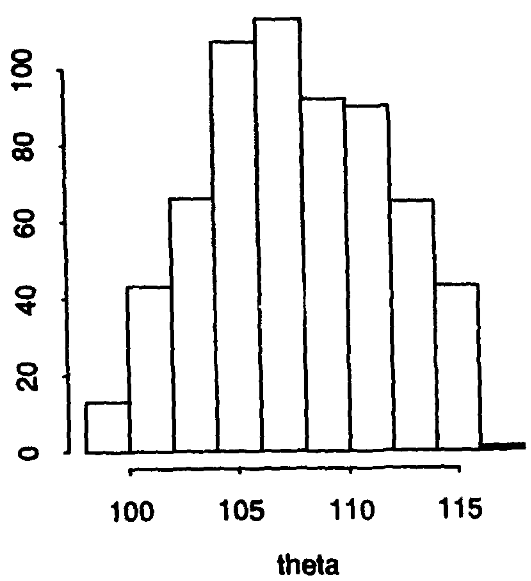
Figure 2: Updated priors for θ using probabilities (2a, 2b), using quantiles (2c, 2d)



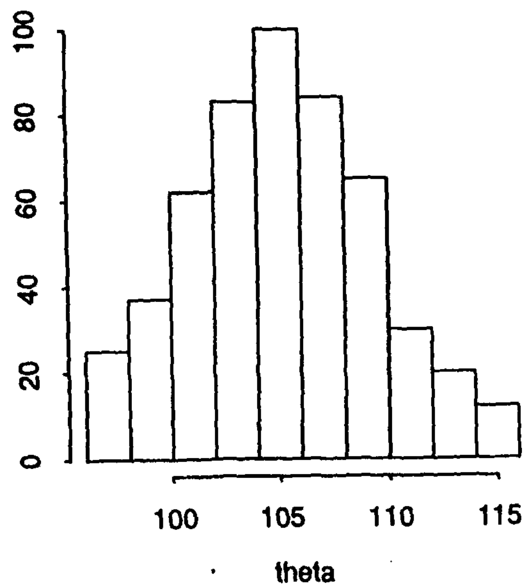
2 a) Chicago



2 b) Los Angeles



2 c) Chicago



2 d) Los Angeles

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 464	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Modeling Expert Opinion: Likelihoods Under Incomplete Probabilistic Specification		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Alan E. Gelfand, Bani K. Mallick, Dipak K. Dey		8. CONTRACT OR GRANT NUMBER(s) N0025-92-J-1264
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305-4065		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 111		12. REPORT DATE December 9, 1992
		13. NUMBER OF PAGES 29
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Reverse Side

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Abstract

Expert opinion is often sought with regard to unknowns in a decision–making setting. Our presumption is that such opinion is elicited as an incomplete probabilistic specification either in the form of probability assignments to fixed intervals or in the form of selected quantiles. We present likelihoods for such specification which arise through random mixtures of Beta distributions. We presume that a supra Bayesian presides over the opinion collection resulting in the posterior distribution as the mechanism for pooling opinion. The models are applied to opinion collected regarding points per game for participants in the 1991 NBA championship basketball series.