Beliefs about beliefs, a theory for Stochastic Assessments of Subjective Probabilities

J.M. DICKEY
University College of Wales, Aberystwyth *

SUMMARY

Parameterized families of subjective probability distributions can be used to great advantage to model beliefs of experts, especially when such models include dependence on concomitant variables. In one such model, probabilities of simple events can be expressed in loglinear form. In another, a generalization of the multivariate t distribution has concomitant variables entering linearly through the location vector. Interactive interview methods for assessing this second model and matrix extensions thereof were given in recent joint work of the author with A.P. Dawid, J.B. Kadane and others. In any such verbal assessment method, elicited quantiles must be fitted by subjective probability models. The fitting requires the use of a further probability model for errors of elicitation. This paper gives new theory relating the form of the distribution of elicited probabilities and elicited quantiles to the form of the subjective probability distribution. The first and second order moment structures are developed to permit generalized least squares fits.

Keywords: SUBJECTIVE PROBABILITY; PROBABILITY MODELING; PROBABILITY ASSESSMENT; PROBABILITY ELICITATION; STOCHASTIC ELICITATIONS.

1. SUBJECTIVE PROBABILITY MODELS

Mathematically, subjective probability models resemble the more familiar sampling theory models. The usual Kolmogorov axioms will be satisfied by a probability mass or density function, which is nonnegative, integrates to unity and is otherwise well behaved.[1]. The distinguishing

* Present affiliation: State University of New York, Albany

[1] Some writers on subjective probability prefer to work in terms of finitely additive probabilities. This distinction is not material to the present paper.

characteristic of a subjective distribution, then, is not some mathematical property, but rather its use to describe a person's state of mind or subjective uncertainty concerning particular events or quantities of interest. Although a realistic subjective probability distribution for a future sample typically has the mathematical property of exchangeability, or another weakened version of the i.i.d. or related property, my emphasis here is on the distinction according to interpretation or use.

Models in general can be classified as fixed or parametric. A *fixed* probability model is a single probability distribution for a scalar or vector random quantity, or a single distribution-valued function of concomitant variables. In the case of subjective probability, the distribution would be said to be *conditional* on the information in the concomitant variables. A *parametric* model, on the other hand, is a class of models indexed by one or more parameters, whose values serve to specify corresponding fixed models in the class. The difference between a concomitant variable and a parameter, for subjective probability, is that a parameter is used to indicate a class of fixed models merely for mathematical convenience. A parameter may fail to have any interpretation as real information. We shall refer to the parametric and fixed forms of the following subjective probability models.

Model 1. Loglinear odds for an event. A person may be uncertain regarding the occurence of a particular event of interest, say for a dichotomous variable $y = 0,1$, the event $y = 1$. Conditionally on the vector of concomitant variables $\mathbf{x}$, his probability is said to take the loglinear form

$$\begin{aligned} p &= \mathrm{Prob}\{y=1\} \\ &= e^{u}/(1 + e^{u}), \end{aligned} \tag{1.1}$$

where for $\mathbf{x} = (x_1,\ldots,x_r)'$ and $\mathbf{b} = (b_1,\ldots,b_r)'$,

$$u = \mathbf{x}'\mathbf{b} = x_1b_1 + \ldots + x_rb_r. \tag{1.2}$$

Inversely, $u = \ell n\{p/(1\text{-}p)\}$. The corresponding parametric model has the vector of parameters $\mathbf{b}$.

Model 2. Location-scale density for a continuous quantity, with linear location and gathered elliptical symmetry. The person may be uncertain about a particular continuous quantity y. His probability distribution is modeled in location-scale form. Suppose it has a density $p\ (y)$, expressible in terms of some special standardized density f,

$$p\,(y) = f\left(\frac{y-m}{c}\right)/c. \tag{1.3}$$

It is as if there were a standard random quantity z having density f, for which

$$y = m + cz. \tag{1.4}$$

The parameters are m and c.

In the presence of the vector of concomitant variables $\mathbf{x} = (x_1,\ldots,x_r)'$ the conditional distribution of y has the linear-form location, for $\mathbf{b} = (b_1,\ldots,b_r)'$,

$$m = \mathbf{x}'\mathbf{b}. \tag{1.5}$$

Then $\mathbf{b}$ would become the parameter, instead of m. Of course, c too could depend on $\mathbf{x}$ (and it will in an important case to be introduced).

This model can be usefully extended in various ways. Writing the concomitant vector as an arbitrary function of more elementary variables $\mathbf{h}$, $\mathbf{x} = \mathbf{x}(\mathbf{h})$, one has the notion of a subjective response surface. This complements the theory of objective response surfaces as traditionally used in the optimization of industrial processes. The surface ordinate $m(\mathbf{h}) = \mathbf{x}(\mathbf{h})'\mathbf{b}$ would represent a subjective location for the response y, as opposed to an ideal long-term mean response. The location $m(\mathbf{h})$ can serve as a subjective point prediction, while the scale parameter c expresses the amount of predictive uncertainty.

Opinion concerning samples in time can be modeled by replacing the scalars y, m, z by vectors $\mathbf{y}$, $\mathbf{m}$, $\mathbf{z}$; the scale parameter c becomes a matrix C; and if concomitant variables are present, $\mathbf{x}'$ should be replaced by a matrix X whose row vectors are point values for $\mathbf{x}_i'$

$$\mathbf{m} = X\mathbf{b}. \tag{1.6}$$

Equations (1.3) through (1.5) then hold again as written with the given replacements. Equation (1.3) for example becomes

$$p\,(y) = f\,\{C^{-1}(\mathbf{y} - X\mathbf{b})\}\{\det(C)\}^{-1}, \tag{1.7}$$

if we assume the matrix C is nonsingular.

Bruce Hill (1969) and A.P. Dawid (1977, 1978) have investigated the property of spherical symmetry, in which the distribution of $\mathbf{z}$ is invariant under rotations. If $A\mathbf{z}$ would have the same distribution as $\mathbf{z}$ for any orthogonal matrix A, then the distribution of $\mathbf{y} = \mathbf{m} + C\mathbf{z}$ depends on the scale matrix C only through the product,

$$W = CC'. \tag{1.8}$$

An example of such a location-scale model is the multivariate Student family

$$\mathbf{y} \sim \text{Student}_d(\mathbf{m}, W), \tag{1.9}$$

where $\mathbf{z} \sim \text{Student}_d(\mathbf{0}, \mathrm{I})$ means that $\mathbf{z}$ can be represented as the product of a standard normal vector and the independent random quantity $(d/\chi_d^2)^{1/2}$. Kadane *et al* (1978) have developed such models for subjective probability modeling.

The multivariate Student distribution (1.9) has the property that the density of $\mathbf{y}$ depends on $\mathbf{y}$ only through the positive definite quadratic form $(\mathbf{y}\text{-}\mathbf{m})'\,W^{-1}(\mathbf{y}\text{-}\mathbf{m})$, and it strictly decreases in this quadratic form. We shall refer to *any* distribution which has these properties as *gathered and elliptically symmetric*. Much of the work here will apply with full force to a general gathered elliptically symmetric distribution with linear location. The main advantage of such models is that they can be maximized in their coefficients vector by the method of generalized least squares. We write for such a model in analogy to (1.9), for $\mathbf{y} = \mathbf{m} + C\mathbf{z}$,

$$\mathbf{y} \sim F\,(\mathbf{m}, W), \tag{1.10}$$

where $\mathbf{z}$ has the standard distribution $\mathbf{z} \sim F\,(\mathbf{0}, \mathrm{I})$.

Matrix-variate extensions of such models are also available for opinion about multivariate responses sampled at various concomitant points (Dawid, Dickey and Kadane, 1979).

Subjective probability models, such as the models introduced here, are important for situations where there is not a large amount of proper statistical data available and expert opinions must be used for planning experiments or other decision making. Such models are indispensible when there is little or no proper data. Expert opinion is already used extensively now without formal modeling. The intention is that probability models can bring order into expert-opinion processes. The general scientific method urges observation and experimentation where feasible, and samples can be planned and analyzed using subjective probability. But these models are also useful in situations where statistical methods would not be applied.

Modelling of beliefs has the following types of use:

1. Clarification of belief, during the modeling or assesment process.
2. Communication. More precise expression of opinion.
3. Comparison and possible pooling of experts' opinions.
4. Decision; e.g. coherent decision (criterion of maximum utility).
5. Planning of experiments (e.g. criterion of maximum expected value of sample information).
6. Analysis of experimental or observational data. Updating of opinion by probability conditioning.

Given a joint probability distribution for observed data and some uncertain quantities of interest, such as future data, opinion is coherently updated to account for the observed data by the usual probability conditioning in the joint distribution. For example, in the joint distribution (1.6) for $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2')'$, $p(\mathbf{y}_2 | \mathbf{y}_1) = p(\mathbf{y}_1, \mathbf{y}_2) / \int p(\mathbf{y}_1, \mathbf{y}_2) d\mathbf{y}_2$. In the multivariate Student case (1.9),

$$\mathbf{y}_2 | \mathbf{y}_1 \sim \text{Student}_{d+r_1}\{A(\mathbf{y}_1), B(\mathbf{y}_1)\}, \qquad (1.11)$$

where

$$A(\mathbf{y}_1) = \mathbf{m}_2 + W_{21} W_{11}^- (\mathbf{y}_1 - \mathbf{m}_1) \qquad (1.12)$$
$$B(\mathbf{y}_1) = (1 + r_1/d)^{-1}\{1 + d^{-1}(\mathbf{y}_1 - m_1)' W_{11}^- (\mathbf{y}_1 - \mathbf{m}_1)\} \cdot (W_{22} - W_{21} W_{11}^- W_{12}) \quad ,$$

We have partitioned $\mathbf{m}$ and W conformably to $\mathbf{y}$; W_{11}^- is a generalized inverse; and $r_1 = \text{rank } W_{11}$. (Of course, what is actually meant by this in practice is conditioning on a small positive-probability interval for $\mathbf{y}_1$.)

Note that there has been no need to mention Bayes' theorem. It is only in the special case that $p(\mathbf{y})$ is a mixture of sampling models that Bayes' theorem arises. That is, if $p(\mathbf{y}) = \int p(\mathbf{y} | \theta) p(\theta) d\theta$ in terms of an i.i.d. sampling model $p(\mathbf{y} | \theta)$ with an unknown parameter θ subject to the prior distribution $p(\theta)$, then

$$p(\mathbf{y}_2 | \mathbf{y}_1) = \int p(\mathbf{y}_2 | \theta) p(\theta | \mathbf{y}_1) d\theta \, , \qquad (1.13)$$

where the posterior distribution in the integrand is obtained by Bayes' theorem,

$$p(\theta \mid \mathbf{y}_1) = p(\mathbf{y}_1 \mid \theta) p(\theta) / \int p(\mathbf{y}_1 \mid \theta) p(\theta)\, d\theta \tag{1.14}$$

A special case of our multivariate Student model (1.9) can be viewed as a subjective average of the familiar normal-linear-regression sampling models in which

$$\mathbf{y} \mid \beta, \sigma \sim \text{Normal}(X\beta; \sigma^2 \mathrm{I}). \tag{1.15}$$

If β and σ have the usual conjugate prior distribution,

$$\beta \mid \sigma \sim \text{Normal}(\mathbf{b}; \sigma^2 N^{-1}) \tag{1.16}$$
$$\sigma^2 \sim s^2(d/\chi^2_d),$$

then the corresponding prior-predictive distribution for $\mathbf{y}$ is just the multivariate Student distribution (1.9) with the special parameter values,

$$\mathbf{m} = X\mathbf{b}, \quad W = s^2(X' N^{-1} X + \mathrm{I}). \tag{1.17}$$

The usual Bayesian updating equations for opinion regarding β and σ (Raiffa and Schlaifer, 1961) lead to the same posterior predictive distribution as (1.11) with (1.17). But, of course, our form (1.11) is much more general.

We state again, for emphasis, that a mixture of sampling models is a special case. In the multivariate Student prevision, a special form of the parameter W is implied (1.17), special in the sense that W is then the sum of a scalar matrix and a matrix of rank fixed relative to the sample size (dimensionality of $\mathbf{y}$).

2. THE PROBLEM OF ASSESSMENT

Just as in any mathematical modeling situation, a person who wishes to model his beliefs by probability is faced with the problem of specifying his model. This can be broken down into the subproblems of determining a parametric model and assessing a fixed model within a given parametric model. We treat the latter type of problem here. In practice, the full specification may proceed by an iteration alternating between tasks of the two types.

We assume that the assessor subjectively specifies *aspects* of the model. Aspects may include: probability values; quantiles; moments; even parameters themselves. Typically, he will overdetermine the model by assessing more aspects than are required to fix the model mathematically. That is, his assessed aspects will be logically contradictory under the model, and some kind of fit must be performed. The extent to which they contradict each other can help indicate the degree of suitability of the given parameterized model.

I should like to emphasize here that subjective probability modeling is like any other type of mathematical modeling, in that diagnostic checks are necessary to see whether the chosen parametric model is adequate for the real situation being modeled. Loglinear odds and gathered elliptically symmetric models are here claimed to be widely useful, but like any parametric model, they cannot be universal. (No model is ever exactly true). The main argument for considering them is that they are tractable and allow a wide variety of opinion structures.

We envisage the assessment process as an aspect-specification and fitting cycle:

1. Specify new aspects
2. Fit model to specified aspects
3. Diagnostic checking
4. Change aspects or change parametric model, and go to 1; or stop.

Interactive computer programs for such a process for models of our second type (1.9), (1.17), are reported in Kadane *et al* (1978) and Dickey and Price (1979). This previous work, however, is informal, in using convenient but arbitrary methods for step 2. The present paper attempts to meet the need for reasonable formal criteria and methods for fitting subjective probability models to specified aspects.

A question of interpretation may be of particular interest at this point. The aspect specifications and the model aimed at are both conceived as subjective entities in the sense of being merely expressions of personal opinions, rather than properties of real-world objects or processes. The reader may appreciate, however, that much of the development here would also apply to situations where an underlying probability model, which a person is trying to assess, is considered to have its own objective existence (say, the long term frequency of failure for a particular type of component in an operating nuclear power plant). Then the aspect specifications could be conceived as subjective *estimates* of the objective aspects.

Contexts of the latter sort resemble in many ways the traditional sampling context in which *both* the model and the aspect specifications are objective. That is, data *drawn from* the model are used to form statistics, which then estimate aspects of the model. This resemblance will receive further discussion latter. For the present we merely point out the logical distinction between data *concerning* a model and data *drawn from* a model. The former concept is the more general.

3. STOCHASTIC ASSESSMENT MODELS

We postulate two models, in general. First, the *belief-model* or subjective probability model, denoted p, say a probability mass or density function $p\ (y)$ for the uncertain quantity y. This is the underlying true fixed model, the object of the assessment. It is *true* in the sense of exactly describing the given expert's personal belief, and it takes the form of a probability distribution, possibly conditional on concomitant variables. Aspects, functions of this model, are denoted,

$$u_1, u_2, \ldots, u_n. \tag{3.1}$$

Denote the vector $\mathbf{u} = (u_1, \ldots, u_n)'$. Then $\mathbf{u} = \mathbf{u}(p)$.

Strictly speaking, for the aspects to be functions, the model p would need to be seen as a member of a class of models, such as the class of all distributions for y on the given range. For another example, if the model is parameterized by $\mathbf{a}$, then $\mathbf{u} = \mathbf{u}(\mathbf{a})$. Typically, this function is invertible on a subrange of $\mathbf{u}$ values. For these values, then, the model p would be identified (in the mathematical sense) by $\mathbf{u}$.

The expert assesses values for the aspects,

$$u_1^*, u_2^*, \ldots, u_n^*. \tag{3.2}$$

In vector form, write $\mathbf{u}^* = (u_1^*, \ldots, u_n^*)'$. The second category of model is the *assessment model*, denoted q. This is a probability mass or density function $q\ (\mathbf{u}^*)$ for the random assessments $\mathbf{u}^*$, which depends on the true model p. Whereas $u_1, \ldots, u_n$ "concern" p, $u_1^*, \ldots, u_n^*$ are "drawn from" q. We assume that the dependence of q on p comes only through $\mathbf{u}$, and hence write for given p,

$$q(\mathbf{u}^*) = g\ (\mathbf{u}^*; \mathbf{u}). \tag{3.3}$$

This is a new use for the concept of probability. (See, however, Lindley, Tversky and Brown, 1979). On the one hand, q models the subjective belief of the expert concerning his own belief p. On the other hand, a sample $\mathbf{u}^*$ drawn from q is actually available for analysis, and $\mathbf{u}^*$ can be analysed in any of the ways a statistician would ordinarily work with data drawn from a distribution. The assessment probability for $\mathbf{u}^*$ (3.3) depends on $\mathbf{u}$, and hence on p. Thus, in the case of a parameterized belief model p, the assessment likelihood for the belief parameter $\mathbf{a}$ can be written

$$\ell_q(\mathbf{a}) = q\,(\mathbf{u}^*)_{u=u(a)} = g\,(\mathbf{u}^*; \mathbf{u}(\mathbf{a})\,). \tag{3.4}$$

Consequently, familiar Bayesian or likelihood methods can now be used to make inference concerning p through $\mathbf{a}$. In particular, one can estimate the belief model p by maximizing the assessment likelihood $\ell_q(\mathbf{a})$ (3.4). (The frequentist justifications for maximum likelihood are well known; Bayesians might justify it as an approximate posterior mode).

Lindley, Tversky and Brown (1979) postulate a further probability model in order to carry out Bayesian inference concerning p. For them p itself would be random under a further "prior" distribution.

Example. Assessment likelihood having linear location and gathered elliptical symmetry. Consider the following useful structures for q and p, respectively, in terms of a standard gathered elliptically symmetric distribution $G\,(\mathbf{0},\mathbf{I})$,

1. $\mathbf{u}^* \mid \mathbf{u} \sim_q G\,(\mathbf{u}, V)$
2. $\mathbf{u} = L\mathbf{a}$.

The assessment likelihood in this case would be maximized by the generalized-least-squares estimate,

$$\hat{\mathbf{a}} = (L'V^{-1}L)^{-1}(L'V^{-1}\mathbf{u}^*). \tag{3.5}$$

One usually sees the estimate (3.5) justified by the Gauss-Markov theorem in terms of variance and bias. It was derived here by maximum likelihood. This structure would include the usual normal linear model, to which both such "justifications" apply. Variance, bias, and other moments may fail to exist, however, for more general G.

Note that in the present example very little has yet been stated concerning the belief model p; merely, that some aspects of p are linearly related to some parameters in p. Nothing yet has been assumed regarding the interpretation of $\mathbf{u}$ or $\mathbf{a}$. In a special case of some interest, the object $\mathbf{y}$ of the belief would follow a related subjective-probability model,

3. $\mathbf{y} \mid \mathbf{u} \sim_p F\,(\mathbf{u}, W)$,

where, for example, $F(0,1)$ is the same standard distribution as $G(0,1)$. We shall discuss later a possible relevance for taking the matrices V and W to be proportional.

We turn in the following sections to theoretical considerations relating assessment models to the belief models previously given. Particular location and scale structures will be motivated for assessment models q, for use of the generalized least squares estimate (3.5).

4. ASSESSING THE PROBABILITY OF AN EVENT

For aspects, consider the linear logodds of equation (1.1), $u_i = \ell n \{p_i/(1-p_i)\} = \mathbf{x}_i'\mathbf{b}$, $i = 1,\ldots,n$. The expert could assess either u_i or p_i, but we retain the notation in which the logodds are treated as the aspects. In practice, one might prefer to assess p_i directly and then transform to an assessment of u_i. Expanding both the transformation and its inverse about the point $p = \frac{1}{2}$ yields

$$u = 2\{p - (1-p)\} + \tfrac{2}{3}\{p - (1-p)\}^3 + \ldots$$
$$p = \tfrac{1}{2} + \tfrac{1}{4}u - \tfrac{1}{48}u^3 + \ldots \qquad (4.1)$$

Both second-order terms vanish, and so the transformation is approximately linear for moderate probabilities.

Assuming that assessments u^*, p^* are related similarly to u, p, that is by $u^* = \ell n\{p^*/(1-p^*)\}$, we have that unbiasedness of p^* is approximately equivalent to unbiasedness of u^*. Hence we assume for the first moment of u^*:

Assumption 4.1

$$Eu^* = u. \qquad (4.2)$$

Cox (1958) uses the somewhat weaker assumption of a constant bias for u_i^* in the context of subjective estimation of objective probabilities.

We discuss the second moment at length.

It is clear that very small or very large probabilities are assessed with smaller absolute errors than moderate probabilities. We shall argue here for the proportionality

$$\mathrm{Var}(p^*) \propto p(1-p) \qquad (4.3)$$

Justification (a). A constant coefficient of variation S.D. $(p^*)/E(p^*)$ would express the idea that the errors in assessment are proportional, in their standard deviation, to the true value p. This seems more reasonable than a constant standard deviation for small probabilities, but perhaps overly optimistic in that such probabilities are notoriously difficult to assess. It would also be unrealistic for large probabilities in not having the standard deviation there smaller than at moderate probabilities. A reasonable compromise which meets all of the above points is to consider the new ratio,

$$S.D.(p^*)/\{Ep^*)(1 - Ep^*)\}^{1/2}. \tag{4.4}$$

This will be constant under (4.3) for unbiased p^*.

Justification (b). If p^* is Beta distributed under the assessment model, then $\mathrm{Var}(p^*) \propto (Ep^*)(1 - Ep^*)$, which again yields (4.3) in the unbiased case.

Justification (c). The variance within the subjective-probability model is $\mathrm{Var}(y) = p\,(1 - p)$. We shall argue, below, for the case of continuous y, that assessment variance is proportional to belief variance. By (mere) analogy here, $\mathrm{Var}\,(p^*) \propto \mathrm{Var}(y) = p\,(1 - p)$.

Considering now the second moment of u^*, we have, to first order,

$$u^* - u = (du/dp) \cdot (p^* - p). \tag{4.5}$$

Hence, $\mathrm{Var}(u^*) \doteq (du/dp)^2\,\mathrm{Var}(p^*) \propto \{p\,(1-p)\}^{-2}\{p\,(1-p)\} \propto \{p\,(1-p)\}^{-1}$, by (4.3). This motivates the following:

Assumption 4.2. $\mathrm{Var}(u^*)$ is proportional to $\{p\,(1-p)\}^{-1} = e^{-u} + 2 + e^{u}$.

To use the moment structure of Assumptions 4.1 and 4.2 to fit a loglinear odds model to assessed aspects will require iteration, because the variance is a function of the mean. Further assumptions would also be needed regarding the covariances.

5. ASSESSING QUANTILES OF A LOCATION-SCALE MODEL

For a continuous random quantity y define the πth quantile $(0 < \pi < 1)$ as the number y_π satisfying

$$P\,\{y \leq y_\pi\} = \pi\,. \tag{5.1}$$

In the problem of assessing a simple location-scale model (1.3) consider as aspects u_i, the quantiles y_{π_i} for given probabilitiy values π_i, $i = 1,\ldots,n$. A linear relation holds between the quantiles of y and the corresponding quantiles of the standard random quantity z,

$$y_{\pi_i} = m + cz_{\pi_i}\,. \tag{5.2}$$

Hence given the assessed quantiles $y^*_{\pi_i}$, $i = 1,...,n$, a natural method to use to estimate m and c is to fit the straight line (5.2) to the "data", z_{π_i}, $y^*_{\pi_i}$, $i = 1,...,n$. This was proposed by I.J. Good (1978) as a method of reconciling subjective quantiles from several experts in the normal case.

An appealing fitting method to use here is generalized least squares (3.5), and a candidate for the required error-covariance structure will be developed below. Garthwaite and Dickey (1979) study properties of the "bisection" method, a special case of this method when bisection is used for assessing location-scale parameters. In the bisection method, particular quantiles are elicited as medians of distributions conditioned on subintervals.

In the more general multivariate location-scale model with linear form location, $\mathbf{m} = X\mathbf{b}$ (1.7), it might seem reasonable to fit this form for $\mathbf{b}$ after assessing a single quantile of y_i at each point $\mathbf{x}_i$ (y conditional on $\mathbf{x} = \mathbf{x}_i$), where the row vectors $\mathbf{x}_i'$, $i = 1,...,n$, comprise the matrix X. These quantiles $y_{\pi_i}(\mathbf{x}_i)$ would all be assessed for the same probability value say $\pi_i \equiv 1/2$, an appealing value to use in the elliptically symmetric model (1.9), for which the coordinates of $\mathbf{m}$ are the medians of the coordinates of $\mathbf{y}$. One would fit the linear relation, in this case,

$$y_{.50}\,(\mathbf{x}_i) = \mathbf{x}_i'\,\mathbf{b} \tag{5.3}$$

to the "data", $\mathbf{x}_i$, $y^*_{.50}\,(\mathbf{x}_i)$, $i = 1,...,n$. Again, generalized least squares will require an error-covariance structure.

5.1 *Sample quantiles as estimates of quantiles*

Subjective assessment of quantiles may be preferrable to the assessment of probabilities of intervals or half-lines, because an expert may find it more meaningful to weigh against each other quantities having the same units as the unknown y, relative to a fixed probability, rather than comparing candidate probability values. But how accurate are such quantile assessments? Perhaps a clue is available from the analogous problem of estimating the quantiles of a traditional population by the quantiles of a sample drawn from the population. There is, of course, no *necessary* connection between this and our problem of assessing subjective probability quantiles.

Denote a population by p, or $p\,(y)$. Denote its πth quantile by y_π, and the corresponding quantile of an independent sample from p by y^*_π. Then for large samples, the asymptotic distribution of y^*_π is normal with mean and variance,

$$E(y^*_\pi) = y_\pi$$
$$\mathrm{Var}(y^*_\pi) = \nu^{-1}\pi(1 - \pi)/p\,(y_\pi)^2, \tag{5.4}$$

where ν denotes the sample size. Indeed, the joint distribution of the sample quantiles $y^*_{\pi_i}$ at several probability values π_i is asymptotically multivariate normal with the covariance structure,

$$\mathrm{Cov}(y^*_{\pi_1}, y^*_{\pi_2}) = \nu^{-1}\pi_1(1 - \pi_2)/\{p\,(y_{\pi_1})p\,(y_{\pi_2})\} \tag{5.5}$$

for $\pi_1 \leq \pi_2$ (Mosteller, 1946).

So sample quantiles are asymptotically unbiased; and in the case of a location-scale family $p(y) = f\{(y-m)/c\}/c$, the variances and covariances will be proportional to the squared population scale parameter,

$$\mathrm{Cov}(y^*_{\pi_1}, y^*_{\pi_2}) = [\nu^{-1}\pi_1(1-\pi_1)/\{f(z_{\pi_1})f(z_{\pi_2})\}]c^2 \tag{5.6}$$

That is, if the distribution being estimated has a variance, the sample quantiles will be distributed with an asymptotic variance proportional to it,

$$\mathrm{Var}(y^*_\pi) \propto \mathrm{Var}(y) \tag{5.7}$$

We shall argue for an analog of this principal in the next section.

In unpublished work Michael Cain has derived assessment fitting procedured for the linear model (1.9), (1.17) using the moment structure of sample quantiles, following a suggestion by J.B. Kadane.

5.2 *Assessed quantiles*

Returning to the general notion of assessed quantiles y^*_π having a distribution q conditional on the quantiles y_π of the distribution p of y, define the cumulative distribution function P for y, at any value y^o,

$$P(y^o) = \mathrm{Prob}\{y \leq y^o\} = \int_{-\infty}^{y^o} p(y)dy. \tag{5.8}$$

Then, of course, $\pi = P(y_\pi)$.

Transforming the assessment, define the quantity,

$$\pi^* = P(y^*_\pi) = \int_{-\infty}^{y^*_\pi} p(y)dy. \tag{5.9}$$

In practice, π^* will not be available in numerical form, depending as it does on the model p. But still, π^* is a mathematically well defined random quantity and has a distribution induced by the assessment distribution q, and we can discuss the behavior of π^* relative to the "true" value π.

A main idea of this paper is that the induced distribution of π^* promises to be insensitive to the model p; at any rate, less sensitive than the distribution of the assessed quantile y_π^* itself. The quantity π^* represents the amount of "true" probability included to the left of y_π^*. The assessment errors in y_π^* could be expected to be large if the integrand $p(y)$ of (5.9) is small in the vicinity of y_π, and small if $p(y)$ is large there. The less believable a region is, the more difficult it is to assess a quantile within it, and visa-versa. This would have the effect of stabilizing the distribution of π^* in its dependence on the local behavior of $p(y)$. We consider small errors in y_π^*, and hence small errors in π^*.

Assumption 5.1. π^* is unbiased:

$$E(\pi^*) = \pi. \tag{5.10}$$

Now, take the linear expansion of the cumulative,

$$\pi^* \doteq \pi + p(y_\pi)(y_\pi^* - y_\pi) \tag{5.11}$$

which yields, together with Assumption 5.1.

Consequence 5.2. For small assessment errors, y_π^* is unbiased:

$$E(y_\pi^*) \doteq y_\pi. \tag{5.12}$$

Assumption 5.3. The model $p(y)$ is parameterized as a location-scale family, $y = m + cz$, where z has a known distribution with density $f(z)$ and unit variance. Hence, $p(y_\pi) = f(z_\pi)/\{\mathrm{Var}(y)\}^{1/2}$. (If one makes the assumption that the assessment model $q(y_\pi^*)$ is also of location-scale form, then no moments need exist, and one can read locations and squared-scale parameters for the means and variances throughout this section.)

In spirit similar to Assumption 5.1, we have,

Assumption 5.4. Var (π^*) is constant in m and c.

Consequence 5.5. Proportionality of variances (scales).

$$\mathrm{Var}(y_\pi^*) \doteq \mathrm{Var}(\pi^*)/p(y_\pi)^2 = \{\mathrm{Var}(\pi^*)/f(z_\pi)^2\} \cdot \mathrm{Var}(y) \tag{5.13}$$

Under an assumption analogous to the constant modified coefficient of variation in the linear log odds problem (4.4), we obtain a more explicit form for the dependence on the quantile probability value π, as follows.

Assumption 5.6. Constant moments ratio

$$\text{S.D.}\,(\pi^*)/\{(\mathrm{E}\pi^*)(1-\mathrm{E}\pi^*)\}^{1/2} = \mathrm{K} \quad . \tag{5.14}$$

Consequence 5.7.

$$\operatorname{Var}(y_\pi^*) \doteq K^2\pi(1-\pi)/p(y_\pi)^2 = \{K^2\pi(1-\pi)/f(z_\pi)^2\}\cdot \operatorname{Var}(y). \tag{5.15}$$

This exhibits an even closer resemblance than (5.13) to the variance of a sample quantile (5.4). It is tempting here to speculate that the covariance of assessed quantiles might have an analogous resemblance to the covariance of sample quantiles,

$$\operatorname{Cov}(y_{\pi_1}^*, y_{\pi_2}^*) = \{\mathrm{K}^2\pi_1(1-\pi_2)/f(z_{\pi_1})f(z_{\pi_2})\}\operatorname{Var}(y) \tag{5.16}$$

for $\pi_1 \leq \pi_2$. The corresponding correlation coefficient would be the same as in the sample quantile case, namely $[\{\pi_1/(1-\pi_1)\}/\{\pi_2/(1-\pi_2)\}]^{1/2}$.
This correlation approaches unity as $\pi_2-\pi_1 \rightarrow 0$, and so our stochastic assessment model is "smooth" in the assessment of neighboring quantiles. (For sample quantiles, of course, such a limiting operation makes no sense).

We turn finally to the distribution of median assessments $y_{50}^*(\mathbf{x}_i)$ in the gathered elliptically symmetric location-scale model with linear location $y_{50}(\mathbf{x}_i) = \mathbf{x}_i'\mathbf{b}$, $i = 1,\ldots,n$. The essential property for the discussion here is that a set of jointly distributed assessed quantities $y_{50}^*(\mathbf{x}_i)$ have variances proportional to the corresponding jointly distributed observables y_i at $\mathbf{x}_i$. Denote the vectors having these two sets of coordinates, respectively, by $\mathbf{y}_{50}^*$ and $\mathbf{y}$. We extend this property in the following,

Assumption 5.8. In the coordinate system of the principal components of $\mathbf{y}$, the vectors $\mathbf{y}_{50}^*$ and $\mathbf{y}$ again have coordinates with proportional variances: Write $\eta = \mathbf{A}\mathbf{y}$ and $\zeta = \mathbf{A}\mathbf{y}_{50}^*$. Assume that for some orthogonal matrix A, both $\operatorname{Var}(\eta) = \operatorname{Diag}(\tau_1^2, \ldots, \tau_n^2)$ and $\operatorname{Var}(\zeta_i) = k\tau_i^2$, $i = 1, \ldots, n$.

Assumption 5.9. Quantities uncorrelated in the belief model correspond to quantities uncorrelated in the assessment model: Assume $\operatorname{Cov}(\zeta_i, \zeta_j) = 0$, $i \neq j$.

Clearly then, the matrices $\operatorname{Var}(\eta) = k \operatorname{Var}(\zeta)$, and hence,

Consequence 5.10. Proportionality of covariance matrices:

$$\operatorname{Var}(\mathbf{y}_{50}^*) = k \operatorname{Var}(\mathbf{y}) \tag{5.17}$$

To use this moments structure for a fit on the linear location will require, of course, a separate assessment and fitting procedure for the scale matrix, or an iteration alternating between the scale and the location.

6. DISCUSSION

We have argued theoretically for particular forms of probability model for the behavior of subjectively assessed aspects of a probability model of belief or frequency. Such a model of assessment behavior would be essentially *descriptive* in its interpretation, rather than *normative* as the underlying belief model. As such, its suitability should be investigated experimentally. Do errors in assessing belief behave as advertized; or is another stochastic model more realistic; or can better descriptions be given in deterministic form?

One difficulty to be met in the experimental study of assessment models is that of establishing the underlying belief model. Assessments are measurements on beliefs, and to study the distribution of assessment errors would seem to require working in controlled conditions where the "true" opinion values are known, that is, known to the experimenter but not known precisely to the person whose opinions are being assessed. This seems hardly likely for underlying *subjective* probabilities. The subjective estimation of *objective* probabilities is another story, and perhaps experiments on this problem can be extrapolated in their implications to the former problem. Of course, sophisticated statistical methods are also available for inferring the distribution of errors without knowing the underlying "true" values, though typically, this will require additional structural assumptions.

A more fundamental difficulty must, however, be addressed here, and that is that underlying belief models may fail to exist in any realistic sense. In his second philosophy, following Ramsey, Wittgenstein (1953) dealt devastingly with all kinds of logical constructs invented to describe the human mind. A mind's "perceptions" of its own "mental states" (including beliefs) was a favorite target of his. Such logical constructs seem to exhibit what DeFinetti (1974, p.22) calls "the inveterate tendency of savages to objectivize and mythologize everything; a tendency that, unfortunately, has been, and is, favoured by many more philosophers than have struggled to free us from it".

My purpose in this paper is to investigate a framework that may be of use in practice, in the sense that the subjective probability models eventually fixed by an assessment-and-fitting cycle will be found useful. The suspicion remains that the model produced many depend strongly on the assessment method (Hogarth 1975). A person's opinions are not coherent (probabilistic) to begin with, but only as he makes deliberate use of the normative theory of subjective probability. Stochastic assessment models may help provide ways of using the normative theory.

ACKNOWLEDGEMENT

Research supported by a grant from the Social Science Research Council, United Kingdom.

REFERENCES

COX, D.R. (1958). Two further appplications of a model for binary regression. *Biometrika* **45**, 562-65.

DAWID, A.P. (1978). Extendability of spherical matrix distributions. *J. of Multivariate Analysis*, **8**, 559-566.

— (1977). Spherical matrix distributions and a multivarite model. *J. Roy. Statist. Soc., B.* **39** 254-261.

DAWID, A.P., DICKEY, J.M. and KADANE, J.B. (1979). Matrix *t* and multivariate *t* assessment. *Tech. Rep.,* Department of Statistics, University College of Wales. Aberystwyth.

DE FINETTI, B. (1974). *Theory of Probability, Vol. 1.* New York: Wiley.

DICKEY, J.M. and PRICE, D.E.(1979). Interactive methods for choice of linear subjective probability models. *Tech. Rep.* Statistics Department, University College of Wales, Aberystwyth.

GARTHWAITE, P.H. and DICKEY, J.M. (1979). Assessment by bisection methods for a location--scale family. *Tech. Rep.* Department of Statistics, University College of Wales. Aberystwyth.

GOOD, I.J. (1978). On the combination of judgements concerning quantiles of a distribution with potential application to the estimation of mineral resources. *Tech. Rep.* Department of Statistics, Virginia Polytechnic Institute and State University.

HILL, B.M. (1969). Foundations for the theory of least squares. *J. Roy. Statist. Soc. B.* **31**, 89-97.

HOGARTH, R.M. (1975). Cognitive processes and the assessment of subjective probability distributions. *J. Amer. Statist. Assoc.,* **70**, 271-94.

KADANE, J.B., DICKEY, J.M., WINKLER, R.L., SMITH, W.S. and PETERS, S.C. (1978). Interactive elicitation of opinion for a normal linear model. *Tech. Rep.* **150**, Department of Statistics, Carnegie-Mellon University. *J. Amer. Statist. Assoc.* Dec. issue, 1980.

LINDLEY, D.V., TVERSKY, A., BROWN, R.V. (1979). On the reconciliation of probability assessments. *J. Roy. Statist. Soc. A.* **142**, 146-180 (with discussion).

MOSTELLER, F. (1946). On some useful "inefficient" statistics. *Ann. Math. Statist.* **17**, 377-408.

RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory.* Boston: Graduate School of Business, Harvard University.

WITTGENSTEIN, L. (1953). *Philosophical Investigations.* Oxford: Basil Blackwell.