Bayesian inference in group judgment formulation and decision making using qualitative controlled feedback

S.J. PRESS
University of California, Riverside

SUMMARY

This paper considers the problem of making statistical inferences about group judgments and group decisions using Qualitative Controlled Feedback, from the Bayesian point of view. The qualitative controlled feedback procedure was first introduced by Press (1978), for a single question of interest. The procedure is first reviewed here including the extension of the model to the multiple question case. We develop a model for responses of the panel on each stage. Many questions are treated simultaneously and an autoregressive model is developed for explaining the responses of the group members as a function of the feedback. The errors are assumed to follow a matrix intraclass covariance structure. Marginal and conditional posterior distributions of the regression coefficient vector are found in both small and large samples. The broadly defined generic family of multidimensional Student-*t* distributions is found to play a major role in the results.

Keywords: BAYESIAN, MULTIVARIATE, GROUP DECISION MAKING, QUALITATIVE CONTROLLED FEEDBACK.

1. INTRODUCTION

Group judgment formulation and decision making using qualitative controlled feedback (QCF) was introduced in Press (1978). The work was extended to the multivariate case of many questions in Press (1980). In this paper we carry the work further by adopting the Bayesian point of view and developing the posterior distribution of the coefficient vector that relates individual responses of group members to explanatory variables.

The methodology was originally conceived in order to study how the U.S. Air Force might be reorganized. We will motivate the procedure, however, in a different context.

Suppose, for examples, a city planning bureau would like to resolve some public policy issues that are of importance to the city in various ways. They would like to determine how to allocate the resources in their budget so that "appropriate" funding is devoted to police, fire, and other municipal services, consistent with environmental considerations, political considerations, economic feasibility, engineering and scientific constraints, and perhaps other factors as well. These factors affect most people in some, possibly indirect, way, and no one person is likely to be knowledgable in all related areas.

It is decided to adopt a QCF procedure to assist the policy makers in generating the factors that argue for one allocation over another. A sample of panelists is taken from the city population; the panel members are each given a survey instrument that includes a battery of questions.

The survey instrument could be administered by mail, by telephone, by on-line computer, or whatever. The data collection protocol of QCF requires that each panelist respond to the questions independently of all other panelists, and without any panelists knowing the identity of any other panelists. Thus, the social pressures of face-to-face confrontation in a room, perhaps at the expense of logical reasoning, are avoided.

In applying a QCF procedure, each respondent is typically asked to answer a set of basic questions. In addition, the subject is asked to provide distinct reasons for each answer that will help justify the subject's answers. He will usually also be asked to answer some subsidiary questions that will serve to provide demographic and attitudinal information about the degree of expertise of the subject, his likely institutional biases, etc.

An intermediary is asked to collect all the answers. This person then forms a merged composite of the reasons provided by the panel for the answer to each question asked. This merging can be carried out with the aid of a computer editor. That is, in some situations this step may be carried out mechanically (if most reasons are listed in advance, panelists can check them off and a computer can talley them). Reasons can be coded and classified into some intrinsically orthogonal set (many reasons are probably just paraphrases of one another). The end product generated is a composite of reasons corresponding to each pair of questions and answers.

The composites of reasons are now presented to each panelist in a simple form (such as a checklist). Each panelist is then asked to answer the same set of questions a second time, only now, the panelist is exposed to the reasoning used by all other panelists. The numerical responses given by the other panelists are not provided for any subjects, nor do they receive any other data, such as sample group mean vectors. The composites of reasons are the only data fed back. As a result, the second stage response of a panelist is likely to differ from his first stage response only because he feels he has ignored some arguments used by other panelists. Note that panelists are not told the proportion of panelists who gave a particular reason; a panelist does not have any basis for deciding how much to weight each reason, in his own thinking, other than by adopting his own weighting system according to his own perceptions of value and importance.

This procedure is repeated until the process stabilizes, in the sense that respondents are not changing their responses very much from stage to stage.

There is room, however, for manipulation of the outcome by a devious intermediary who might misrepresent the composite fed back to the panel on each stage. This effect can be minimized by using a group of intermediaries to accomplish the task of forming a composite of reasons.

Earlier research involving group decision making and judgment formulation, and the effects of social interaction pressures, is summarized in Press, 1978. In Section 2 we develop a model for studying the relationships between responses to the questions, and the rationale the panel feels is most important to explain the answers. The model can also be used for predicting the next round's responses (in many situations, for economic or other reasons, it may be difficult or undesirable to carry out the process for one more stage).

The multiple question model is treated in greater detail from a sampling theory viewpoint in Press (1980). The methodology was applied to study a real problem in Press (1979b). Section 3 presents several distinct developments that provide methods for making Bayesian inferences useful for predicting the next round's responses. Finally, Section 4 provides a summary and conclusions.

2. MULTIPLE QUESTION MODEL

2.1 *First Stage*

Let $z_{in}(j)$ denote the numerical response of subject i, on stage n, to question j; $i = 1,2,\ldots,N$; $j = 1,2,\ldots,q$. Let $\mathbf{F}_n$ denote the totality of information obtained on stage n and feed back to each panelist at the beginning of stage $(n+1)$. Let $F^{(n)}$ denote the n-vector (F_j). Finally, let $\mathbf{X}{:}N\mathrm{x}r$ denote a regressor matrix of explanatory variables observed for the N panelists (these are answers to subsidiary questions). Take $F^{(0)} \equiv \mathbf{0}$.

For the first stage model we adopt a simple regression with uncorrelated errors (subjects respond independently on the first stage). Accordingly, assume

$$\mathbf{z}_1(j)\,|\,\mathbf{X} = X\beta(j) + \mathbf{u}_1(j),$$
$$E(\mathbf{u}_1) = \mathbf{0},\ \mathrm{var}[\mathbf{u}_1(j)] = \sigma_1^2(j)\mathbf{I}_N$$

where:

$$\mathbf{z}_1(j) = [z_{11}(j),\ldots,z_{N1}(j)]',\ \mathbf{u}_1 = [u_{11}(j),\ldots,u_{N1}(j)]'$$

$u_{i1}(j)$ denotes an error term, and $\beta(j)$ denotes an rx1 vector of unknown coefficients. For convenience, take

$$\underset{(Nxq)}{\mathbf{V}} = [\underset{(Nx1)}{\mathbf{u}_1(1)}, \ldots, \underset{(Nx1)}{\mathbf{u}_1(q)}], \qquad \underset{(qxN)}{\mathbf{V}'} = [\underset{(qx1)}{\mathbf{v}_1}, \ldots, \underset{(qx1)}{\mathbf{v}_N}],$$

and assume

$$E(\mathbf{v}_i\mathbf{v}_j') = \begin{cases} \Phi_j^* & i=j \\ \mathbf{0}, & i \neq j \end{cases} .$$

If

$$\underset{(Nxq)}{Z_1} \equiv [\underset{(Nx1)}{z_1(1)}, \ldots, \underset{(Nx1)}{z_1(q)}], \qquad \underset{(rxq)}{\mathbf{B}} \equiv [\underset{(rx1)}{\beta(1)}, \ldots, \underset{(rx1)}{\beta(q)}]$$

the model may be written in the compact form

$$\underset{(Nxq)}{\mathbf{Z}_1} = \underset{(Nxr)}{\mathbf{X}} \; \underset{(rxq)}{\mathbf{B}} + \underset{(Nxq)}{\mathbf{V}}, \tag{1}$$

where:

$$E(\mathbf{V}) = \mathbf{0}, \operatorname{cov}(\mathbf{v}_i, \mathbf{v}_j) = \mathbf{0}, i \neq j, \operatorname{var}(\mathbf{v}_i) = \Phi^* \quad .$$

The model of course represents a classical multivariate regression. The Gauss-Markov estimator of **B** is therefore

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_1 \quad . \tag{2}$$

2.2 *Feedback Stages ($n \geq 2$)*

For later stages, beyond the first, the model must change. This is because the composites of reasons fed back to each respondent cause their responses to be mutually correlated. Since they all get the same feedback, however, their responses on the next stage are likely to be *similarly* correlated (homogeneous, or intraclass correlation structure). Moreover, their answers on stage two are likely to be related to their answers on stage one. Adopt the autoregressive model

$$\triangle z_{in}(j) \equiv [z_{in}(j) \,|\, \mathbf{F}^{(n-1)}] - [z_{i,n-1}(j) \,|\, \mathbf{F}^{(n-2)}]$$

$$= \sum_{\alpha=1}^{R_{n-1}^{(j)}} c_i^{(\alpha)}(j)\,[1-\delta_{i,n-1}^{(\alpha)}(j)]\; p_n^{(\alpha)}(j) + u_{in}(j), \tag{3}$$

where $R_n(j)$ denotes the number of distinct reasons given by the panel (this is the number of reasons in the composite) for the answer to question j, on stage n; $\delta_{i,n}^{(\alpha)}$ is unity or zero, depending upon whether or not respondent i records reason α for his answer to question j, on stage n; $c_i^{(\alpha)}(j)$ is an unknown constant of proportionality (to be estimated); and $p_n^{(\alpha)}(j)$ denotes the proportion of respondents who record reason α for question j, on stage n (this will be interpreted as the weight or importance the panel gives to this reason). Note that even though the panel members do not know $p_n^{(\alpha)}(j)$, it can nevertheless be used in our model since the intermediary knows it or can compute it.

The model in eqn. (3) may be interpreted as follows; $\triangle z_{in}(j)$ represents the change in response for subject i, on question j, from stage (n-1) to stage (n). This change results from an incremental effect attributable to each reason (linear combination of effects). If the subject gave that reason on the last stage, there is of course no effect, while if he didn't give it, the effect is proportional to the importance of the reason (as measured by the proportion of panelists who gave the reason).

2.3 *Error Structure* ($n \geq 2$)

Define

$$\underset{(q\text{x}1)}{\mathbf{u}_{in}} = [u_{in}(1), \ldots, u_{in}(q)]'$$

and assume

$$(1)\ E(\mathbf{u}_{in}) = \mathbf{0},$$

$$(2)\ \text{var}(\mathbf{u}_{in}) = \underset{(q\text{x}q)}{\Sigma_n},$$

$$(3)\ \text{cov}(\mathbf{u}_{in}, \mathbf{u}_{jm}) = \begin{cases} \underset{(q\text{x}q)}{\Lambda_n}, & i \neq j, \quad n = m. \\ \mathbf{0}, & n \neq m \end{cases}.$$

For compactness, let

$$\underset{(Nq\text{x}1)}{\mathbf{u}_n} \equiv [\mathbf{u}'_{1n}, \ldots, \mathbf{u}'_{Nn}]',$$

Then, $E(\mathbf{u}_n) = \mathbf{0}$, and

$$\operatorname{var}(\mathbf{u}_n) = \underset{(NqxNq)}{\Omega_n} = \begin{pmatrix} \Sigma_n & \cdot & & \Lambda_n \\ & & \cdot & \\ \Lambda_n & & \cdot & \Sigma_n \end{pmatrix}$$

Ω_n is seen to be a matrix intraclass covariance matrix. Some of its properties are given, e.g., in Press, 1972, pp. 21, 48, 49 and in Press, 1979a. The assumption of equal diagonal blocks in Ω_n means we are assuming multivariate homoscedasticity. All off-diagonal elements of the qxq blocks are assumed to be identical (Λ_n). We are therefore assuming that in many situations it is reasonable to expect that the panel will be constituted with members who are sufficiently homogeneous in background so that a pattern of homogeneous correlation is reasonable.

2.4 *Transformations to Canonical Form*

Let

$$\underset{(q\mathrm{x}1)}{\triangle \mathbf{z}_{in}} = [\triangle z_{in}(1),\ldots,\triangle z_{in}(q)]' \quad ,$$

and assume

$$c_i^{(\alpha)}(j) = \underset{(1\mathrm{x}r)}{\mathbf{x}_i'} \ \underset{(r\mathrm{x}1)}{\mathbf{a}_\alpha(j)} \quad , \tag{4}$$

where $\mathbf{x}_i$ denotes the (rx1) vector of explanatory variables for subject i, and $\mathbf{a}_\alpha(j)$ denotes an (rx1) vector of unkown weights. For compactness, let

$$\underset{[R_{n-1}(j)\mathrm{x}1]}{\mathbf{c}_i(j)} = [c_i^{(1)}(j),\ldots,c_i^{(R_{n-1}(j))}(j)]' \quad ,$$

and

$$\underset{[rR_{n-1}(j)\mathrm{x}1]}{\mathbf{a}^{(n-1)}(j)} = [a_i(j),\ldots,a'_{R_{n-1}(j)}(j)]' \quad ,$$

so that

$$\mathbf{c}_i(j) = (I \oplus x_i')\, a^{(n-1)}(j) \quad ,$$

where $\oplus$ denotes the direct product. We next combine all the observable explanatory data into one matrix. Define

$$\underset{[rR_{n-1}(j)\mathrm{x}1]}{\mathbf{w}_{in}(j)} = (\mathbf{I} \oplus \mathbf{x}_i)\, \mathbf{t}_{in}(j)$$

where:

$$t_{in}^{(\alpha)}(j) \equiv [1-\delta_{in-1}^{(\alpha)}(j)]\, p_n^{(\alpha)}(j) \quad ,$$

$$\underset{[1\mathrm{x}R_{n-1}(j)]}{\mathbf{t}_{in}'(j)} \equiv [t_{in}^{(1)}(j),\ldots,t_{in}^{(R-1(j))}(j)] \quad ,$$

and define

$$\underset{(q\mathrm{h}_{n-1})}{\mathbf{W}_{in}} = \begin{pmatrix} w_{in}'(1) & \cdot & & 0 \\ & & \cdot & \\ 0 & \cdot & \cdot & w_{in}'(q) \end{pmatrix} \quad ,$$

and

$$\underset{(h_{n-1}\mathrm{x}1)}{\mathbf{a}^{(n-1)}} = [a^{(n-1)}{}'(1),\ldots,a^{(n-1)}{}'(q)]' \quad ,$$

where:

$$h_{n-1} \equiv r \sum^{q} R_{n-1}(j), \quad n \geq 2 \quad .$$

The model now becomes

$$\underset{(q\mathrm{x}1)}{\triangle \mathbf{z}_{in}} = \underset{(q\mathrm{x}h_{n-1})}{\mathbf{W}_{in}} \times \underset{(h_{n-1}\mathrm{x}1)}{\mathbf{a}^{(n-1)}} + \underset{(q\mathrm{x}1)}{\mathbf{u}_{in}} \quad . \tag{5}$$

Combining all subjects, (5) becomes

$$\underset{(Nq\mathrm{x}1)}{\triangle \mathbf{z}_{n}} = \underset{(Nq\mathrm{x}h_{n-1})}{\mathbf{W}_{n}} \times \underset{(h_{n-1}\mathrm{x}1)}{\mathbf{a}^{(n-1)}} + \underset{(Nq\mathrm{x}1)}{\mathbf{u}_{n}} \quad , \tag{6}$$

where:

$$\triangle \mathbf{z}_n = (\triangle \mathbf{z}_{1n}',\ldots,\triangle \mathbf{z}_{Nn}')', \quad \mathbf{W}_n = (W_{1n}', \ldots, \mathbf{W}_{Nn}')' \quad ,$$

Iterating over the n stages gives

$$\underset{(Nq\mathrm{x}1)}{\mathbf{z}} \equiv \mathbf{z}_n - \mathbf{z}_1 = \underset{(Nq\mathrm{x}h)}{\mathbf{W}} \cdot \underset{(h\mathrm{x}1)}{\mathbf{a}} + \underset{(Nq\mathrm{x}1)}{\mathbf{u}} \quad , \tag{7}$$

where for $h = \sum_{j=1}^{n-1} h_j \quad n \geq 2$,

$$\underset{(h\text{x}1)}{\mathbf{a}} \equiv (\mathbf{a}^{(1)\prime}, \ldots, \mathbf{a}^{(n-1)\prime})' \quad ,$$

and

$$\underset{(Nq\text{x}h)}{\mathbf{W}} = (\mathbf{W}_2, \ldots, \mathbf{W}_n), \quad \underset{(Nq\text{x}1)}{\mathbf{u}} \equiv \sum_{j=1}^{n} \mathbf{u}_j \quad .$$

The transformed error vector in (7) satisfies

$$E(\mathbf{u}) = \mathbf{0}, \text{var}(\mathbf{u}) = \Omega = \begin{pmatrix} \Sigma & & \Lambda \\ & \ddots & \\ \Lambda & & \Sigma \end{pmatrix} , \tag{8}$$

where

$$\Sigma \equiv \sum_{j=2}^{n} (\Sigma_j), \qquad \Lambda \equiv \sum_{j=2}^{n} (\Lambda_j) \quad .$$

3. BAYESIAN INFERENCE

In this section we examine the unknown coefficient vector in the model defined by (7) and (8), from the Bayesian point of view. Four different approaches will be taken. First we will examine the coefficient vector conditional on the error covariance matrix. Then, we will develop an approximate conditional Bayesian estimator which is useful when samples are large. This approach ignores the intraclass structure of the covariance matrix and is useful for cases where the intraclass structure cannot be assumed. Next, in subsection 3, we will use the intraclass covariance structure when we develop the marginal posterior distribution of the coefficients. The result is complicated, and so a large sample solution is found. In the final subsection we develop a result which is useful in small samples.

3.1 *Known Covariance matrix*

From (7) and (8) it follows that under the assumption of normality on u, the density of the response vector (likelihood function) given the parameters and explanatory variables, is

$$p(\mathbf{z} \mid \mathbf{W},\mathbf{a},\Omega) \propto |\Omega|^{-1/2} \exp\{(-1/2)[(\mathbf{z}-\mathbf{Wa})'\Omega^{-1}(\mathbf{z}-\mathbf{Wa})]\} \quad . \tag{9}$$

Hence, if we adopt a vague prior for **a** (assuming Ω is known), its density is given by

$p(\mathbf{a}) \propto$ constant,

so that the posterior density is given (from Bayes theorem) by

$$p(\mathbf{a} \mid \mathbf{z},\mathbf{W},\Omega) \propto \exp\{(-1/2)[(\mathbf{z}-\mathbf{Wa})'\Omega^{-1}(\mathbf{z}-\mathbf{Wa})]\} \quad . \tag{10}$$

Note that we are using the common Bayesian convention of using the symbol $p(\cdot)$ to denote a generic density; the densities differ from one another according to the arguments and conditioning variables used.

Define the generalized least squares (and maximum likelihood) estimator

$$\bar{\mathbf{a}}(\Omega) = (\mathbf{W}'\Omega^{-1}\mathbf{W})^{-1}\mathbf{W}'\Omega^{-1}\mathbf{z} \quad . \tag{11}$$

Completing the square in the exponent in (10) shows that

$$p(\mathbf{a} \mid \mathbf{z},\mathbf{W},\Omega) \propto \exp\{(-1/2)[(\mathbf{a}-\bar{\mathbf{a}}(\Omega))'(\mathbf{W}'\Omega^{-1}\mathbf{W})(\mathbf{a}-\bar{\mathbf{a}}(\Omega))]\} \quad ,$$

so that

$$(\mathbf{a} \mid \mathbf{z},\mathbf{W},\Omega) \sim N[\bar{\mathbf{a}}(\Omega), (\mathbf{W}'\Omega^{-1}\mathbf{W})^{-1}] \quad . \tag{12}$$

That is, conditional on Ω, a posteriori, and adopting a vague prior on **a**, **a** is normally distributed, centered at the MLE, with precision matrix $(\mathbf{W}'\Omega^{-1}\mathbf{W})$.

We remark in passing that $\bar{\mathbf{a}}(\Omega)$ is the same estimator found from a frequentist point of view in Press (1979a).

3.2 *Large Sample Estimator*

One approximate large sample Bayesian estimator of a may be found (when Ω is unknown) by using the result obtained conditional on Ω, and then replacing Ω by a consistent estimator. This approach follows the spirit used in the frequentist analysis.

Suppose $\tilde{\Omega}$ is a consistent estimator of Ω (for unknown Ω). Then, the approximate posterior distribution of **a** is

$$(\mathbf{a} \mid \mathbf{z},\mathbf{W},\Omega) = N[\mathbf{a}(\tilde{\Omega}),(\mathbf{W}'\tilde{\Omega}^{-1}\mathbf{W})^{-1}] \quad .$$

A consistent estimator, $\tilde{\Omega}$, is developed in Press (1980). Thus, in large samples,

$$\mathbf{a}(\widetilde{\Omega}) \cong \mathbf{a}(\Omega) \quad ,$$

and $\mathbf{a}$ is approximately normally distributed.

3.3 *Marginal Distribution of* a

In this subsection we find Bayesian estimators based upon the marginal posterior distribution of a. The likelihood in (9) is equivalent to

$$(\mathbf{z} \mid \mathbf{W},\mathbf{a},\Omega) \sim N(\mathbf{Wa},\Omega) \quad , \tag{13}$$

where

$$\Omega = \begin{pmatrix} \Sigma & & \Lambda \\ & \ddots & \\ \Lambda & & \Sigma \end{pmatrix} \quad .$$

The posterior density of $\mathbf{a}$ is found by first reducing (13) to canonical form; then adopting a prior for the canonical form parameters, and finally applying Bayes theorem.

Define the orthogonal matrix $\Gamma = \Gamma_0 \oplus \mathbf{I}_q$, where Γ_0 denotes an orthogonal matrix of order N whose first row has equal elements. Then it is straightforward to check (see Press, 1979b, Theorem 5) that if

$$\Omega_0 \equiv \Gamma\Omega\Gamma' \quad ,$$

Ω_0 is block diagonal of the form

$$\Omega_0 = \begin{pmatrix} \Sigma_1 & & \mathbf{0} \\ & \Sigma_2 \ddots & \\ \mathbf{0} & & \Sigma_2 \end{pmatrix} ,$$

,

$\Sigma_1 = \Sigma + (N\text{-}1)\Lambda$, $\Sigma_2 = \Sigma\text{-}\Lambda$. Accordingly, define $\mathbf{z}^* = \Gamma\mathbf{z}$, $\mathbf{w}^* = \Gamma\mathbf{W}$. Then,

$$(\mathbf{z}^* \mid \mathbf{w}^*,\mathbf{a},\Omega_0) \sim N(\mathbf{w}^*\mathbf{a},\Omega_0) \quad . \tag{14}$$

We now view eqn. (14) as the canonical form of the problem and adopt $(\Sigma_1,\Sigma_2,\mathbf{a})$ as the canonical parameter set. Equivalently, if

$$\underset{(Nq\times 1)}{\mathbf{z}^*} \equiv (\underset{(1\times q)}{\mathbf{z}_1^{*\prime}},\ldots,\underset{(1\times q)}{\mathbf{z}_N^{*\prime}})' \quad , \qquad \underset{(Nq\times h)}{\mathbf{w}^*} \equiv (\underset{(h\times q)}{\mathbf{w}_1^{*\prime}},\ldots,\underset{(h\times q)}{\mathbf{w}_N^{*\prime}})' \quad ,$$

the canonical problem is the following:

$$(\mathbf{z}_1^* | \mathbf{w}_1^*, \mathbf{a}, \Sigma_1) \sim N(\mathbf{w}_1^* \mathbf{a}, \Sigma_1) \quad ,$$

$(\mathbf{z}_1^*, \ldots, \mathbf{z}_N^*)$ are independent and

$$(\mathbf{z}_j^* | \mathbf{w}_j^*, \mathbf{a}, \Sigma_2) \sim N(\mathbf{w}_j^* \mathbf{a}, \Sigma_2) \quad ,$$

for all $j = 2, \ldots, N$. A fundamental difficulty at this point is that Σ_1 depends on the sample size N (since $\Sigma_1 = \Sigma + (N\text{-}1)\Lambda$). To circumvent this difficulty we will seek a Bayesian solution to our problem which ignores one data point, namely, $\mathbf{z}_1^*$, and then we will seek a large sample solution, so that the loss of the one data point will be irrelevant.

Accordingly, we consider the joint posterior density

$$p(\mathbf{a}, \Sigma_2 | \hat{\mathbf{z}}, \hat{\mathbf{w}}) \propto \frac{p'(\mathbf{a}, \Sigma_2)}{|\Sigma|^{(N-1)/2}} e^{-(1/2) tr \Sigma_2^{-1} B} \quad ,$$

where $p'(\mathbf{a}, \Sigma_2)$ denotes the joint prior density of $\mathbf{a}$ and Σ_2,

$$\mathbf{B} \equiv \Sigma_{j=2}^{N} (\mathbf{w}_j^* \mathbf{a} - \mathbf{z}_j^*)(\mathbf{w}_j^* \mathbf{a} - z_j^*)'$$

and

$$\hat{\mathbf{z}} \equiv (\mathbf{z}_2^{*\prime}, \ldots, \mathbf{z}_N^{*\prime})', \hat{\mathbf{w}} \equiv (\mathbf{w}_2^{*\prime}, \ldots, \mathbf{w}_N^{*\prime})' \quad .$$

It is interesting to note that the sample covariance among the $(\mathbf{z}_2^*, \ldots, \mathbf{z}_N^*)$ vectors follows a non-central Wishart distribution.

Adopt the prior density

$$p'(\mathbf{a}, \Sigma_2) = p_1'(\mathbf{a}) p_2'(\Sigma_2) \quad ,$$

where:

$$p_1'(\mathbf{a}) \propto \text{constant},$$

$$p_2'(\Sigma_2) \propto \frac{1}{|\Sigma_2|^{n_0/2}} e^{-1/2 tr \Sigma_2^{-1} G}, \Sigma_2 > 0 \quad .$$

That is, the prior density of $\mathbf{a}$ is vague, and the prior density of Σ_2 is inverted Wishart. Note that $(\mathbf{G}, \mathbf{n}_0)$ are assumed to be known hyperparameters. The posterior density now becomes

$$p(\mathbf{a}, \Sigma_2 | \hat{\mathbf{z}}, \hat{\mathbf{w}}) \propto \frac{1}{|\Sigma_2|^{(N+n_0-1)/2}} e^{-(1/2) tr \Sigma_2^{-1} (B+G)} \quad .$$

The marginal posterior density of $\mathbf{a}$ is found by integrating the joint density of $(\mathbf{a},\Sigma_2)$ with respect to Σ_2. Because of the known form of the inverted Wishart density, we readily effect the required integrarion and find

$$p(\mathbf{a}\,|\,\hat{\mathbf{z}},\hat{\mathbf{w}}) \propto \frac{1}{|\mathbf{G}+\Sigma_{j=2}^{N}(\mathbf{w}_j^{*}\mathbf{a}-\mathbf{z}_j^{*})(\mathbf{w}_j^{*}\mathbf{a}-\mathbf{z}_j^{*})'\,|^{\nu/2}}, \tag{15}$$

where $\upsilon = N+n_0-q-2$. The posterior density in eqn. (15) is in the matrix-T family, but is quite complicated analytically. It could always be evaluated numerically, of course, but we seek instead a large sample approximation. An alternative approach will be developed for obtaining simple Bayesian results in small samples.

Large Sample Approximation

Let $\Phi \equiv (\mathbf{w}_2^{*}\mathbf{a}-\mathbf{z}_2^{*},\ldots, \mathbf{w}_N^{*}\mathbf{a}-\mathbf{z}_N^{*})$. Then, eqn. (15) becomes

$$p(\mathbf{a}\,|\,\hat{\mathbf{z}},\hat{\mathbf{w}}) \propto |\mathbf{G}+\Phi\Phi'\,|^{-\nu/2} \propto |\mathbf{I}+\Phi'\mathbf{G}^{-1}\Phi\,|^{-\nu/2} \quad ,$$

or

$$p(\mathbf{a}\,|\,\hat{\mathbf{z}},\hat{\mathbf{w}}) \propto \exp\{(-\upsilon/2)\log|\mathbf{I}_{N-1}+\Phi'\mathbf{G}^{-1}\Phi\,|\} \quad .$$

Let $(\lambda_1, \ldots, \lambda_{N-1})$ denote the latent roots of $\Phi'\mathbf{G}^{-1}\Phi$, and let $\mathbf{D} = \mathrm{diag}\,(\lambda_1,\ldots,\lambda_{N-1})$. Then

$$\begin{aligned} p(\mathbf{a}\,|\,\hat{\mathbf{z}},\hat{\mathbf{w}}) &\propto \exp\{(-\upsilon/2)\log|\mathbf{I}_{N-1}+\mathbf{D}_\lambda|\} \\ &= \exp\{(-\upsilon/2)\log\Pi_1^{N-1}(1+\lambda_i)\} \\ &= \exp\{(-\upsilon/2)\ \Sigma_1^{N-1}\log(1+\lambda_i)\} \quad . \end{aligned}$$

It will be shown shortly that λ_i decreases with sample size, N. Thus, for N sufficiently large, $|\lambda_i| \ll 1$, so that $\log(1+\lambda_i) \cong \lambda_i$. Then,

$$\begin{aligned} p(\mathbf{a}\,|\,\hat{\mathbf{z}},\hat{\mathbf{w}}) &\propto \{(-\upsilon/2)\,\Sigma_1^{N-1}\lambda_i\} \\ &= \exp\{(-\upsilon/2)\,\mathrm{tr}(\Phi'\mathbf{G}^{-1}\Phi)\} \\ &= \exp\{(-\upsilon/2)\,\Sigma_{j=2}^{N}(\mathbf{w}_j^{*}\mathbf{a}-\mathbf{z}_j^{*})'\mathbf{G}^{-1}(\mathbf{w}_j^{*}\mathbf{a}-\mathbf{z}_j^{*})\} \quad . \end{aligned}$$

Each term in the exponent is a quadratic form in $\mathbf{a}$. Combining terms gives

$$\begin{aligned} p(\mathbf{a}\,|\,\hat{\mathbf{z}},\hat{\mathbf{w}}) \propto \exp\{-(\tfrac{1}{2})[\mathbf{a}'(\Sigma_2^N\mathbf{w}_j^{*}{}'(\mathbf{G}/\upsilon)^{-1}\mathbf{w}_j^{*})\mathbf{a} - 2(\Sigma_2^N\mathbf{z}_j^{*}{}'(\mathbf{G}/\upsilon)^{-1}\mathbf{w}_j^{*})\,\mathbf{a} \\ + (\Sigma_2^N\mathbf{z}_j^{*}{}'(\mathbf{G}/\upsilon)^{-1}\mathbf{z}_j^{*})]\} \quad . \end{aligned}$$

To simplify, complete the square in $\mathbf{a}$ to get

$$p(\mathbf{a} \mid \hat{\mathbf{z}}, \hat{\mathbf{w}}) \propto \exp\{(-\tfrac{1}{2})[(\mathbf{a}-\alpha)' F(\mathbf{a}-\alpha)]\} \quad ,$$

where:

$$\alpha \equiv \mathbf{F}^{-1}\mathbf{b}, \; \mathbf{b} \equiv \Sigma_2^N \mathbf{w}_j^{*\prime} + (\mathbf{G}/v)^{-1}\mathbf{z}_j^*; \; \mathbf{F} \equiv \Sigma_2^N \mathbf{w}_j^{*\prime}\,(\mathbf{G}/v)^{-1}\,\mathbf{w}_j^* \quad .$$

That is, a posteriori, in large samples,

$$\mathbf{a} \sim N(\alpha, \mathbf{F}^{-1}) \quad . \tag{16}$$

The only unfinished item remaining in this large sample approximation is to show that the latent roots of $\Phi' \mathbf{F}^{-1} \Phi$ go to zero with increasing sample size. The matrix $\Phi' \mathbf{F}^{-1} \Phi \equiv (r_{ij})$ where

$$r_{ij} \equiv (\mathbf{w}_i^* \mathbf{a} - \mathbf{z}_i^*)' \mathbf{G}^{-1} (\mathbf{w}_j^* \mathbf{a} - \mathbf{z}_j^*) \quad .$$

But

$$\underset{(Nqxh)}{\mathbf{w}^*} \equiv (\mathbf{w}_1^{*\prime}, \ldots, \mathbf{w}_N^{*\prime})' = (\underset{(NxN)}{\Gamma_0} \oplus \mathbf{I}_q)\mathbf{w} \quad ,$$

where $\mathbf{w}$ only changes in dimension with increasing N. But Γ_0 is an orthogonal matrix each of whose elements is of order $N^{-1/2}$. So r_{ij} is of order N^{-1}. So its latent roots must vanish as $N \to \infty$.

Remark (1):

We note that since $\Sigma_1 = \Sigma + (N-1)\Lambda$, as N gets large, Σ_1 becomes very large, so $\mathbf{z}_1^*$ is less and less informative as $N \to \infty$. As a result, ignoring this observation is no great loss in large samples.

Remark (2):

The large sample Bayesian result shows that the elements of the regression coefficient vector $\mathbf{a}$ are, for large N, jointly normally distributed, so that inferences about particular coefficients are readily made.

Remark (3):

The large sample Bayesian result just found is meaningful when the number of subjects on the panel is large; the number of feedback stages may still be small.

3.4 *Small Samples*

To obtain a Bayesian result useful in small or moderate samples we adopt a different point of view than that used in subsection 3.3. Our approach now is to first ignore the (possibly) intraclass covariance structure in the likelihood function, but to recapture the structure in the prior distribution.

We begin with eqn. (14),

$$(\mathbf{z}^* \mid \mathbf{w}^*, \mathbf{a}, \Omega_0) \sim N(\mathbf{w}^*\mathbf{a}, \Omega_0) \quad . \tag{14}$$

Thus, the posterior distribution of $(\mathbf{a}, \Omega_0)$ is

$$p(\mathbf{a}, \Omega_0 \mid \mathbf{z}^*, \mathbf{w}^*) \propto \frac{p'(\mathbf{a}, \Omega_0)}{|\Omega_o|^{1/2}} \exp\{(-1/2)\mathrm{tr}\Omega_0^{-1}\mathbf{H}\} \quad , \tag{17}$$

where

$$\mathbf{H} \equiv (\mathbf{z}^* - \mathbf{w}^*\mathbf{a})(\mathbf{z}^* - \mathbf{w}^*\mathbf{a})' \quad ,$$

and $p'(\mathbf{a}, \Omega)$ denotes the prior density. Note that we are ignoring the intraclass structure of Ω_0 at this point.

For the prior density, assume $p'(\mathbf{a}, \Omega_0) = p_1'(\mathbf{a}) p_2'(\Omega_0)$, and

$$p_1'(\mathbf{a}) \propto \text{constant},$$

$$p_2'(\Omega_0) \propto \frac{1}{|\Omega_0|^{m/2}} \exp\{(-1/2)\mathrm{tr}\Omega_0^{-1}\mathbf{M}\} \quad ,$$

where $(m, \mathbf{M})$ are assumed to be known hyperparameters, $\mathbf{M} > 0$.

The joint posterior density becomes

$$p(\mathbf{a}, \Omega_0 \mid \mathbf{z}^*, \mathbf{w}^*) \propto \frac{1}{|\Omega_0|^{(m+1)/2}} \exp\{(-1/2)\mathrm{tr}\Omega_0^{-1}(\mathbf{M} + \mathbf{H}\} \quad .$$

In Ω_0 this expression is the kernel of an inverted Wishart distribution so it is readily integrated to give the marginal density

$$p(\mathbf{a} \mid \mathbf{z}^*, \mathbf{w}^*) \propto \frac{1}{|\mathbf{M} + \mathbf{H}|^{(m-Nq)/2}},$$

or

$$p(\mathbf{a} \mid \mathbf{z}^*, \mathbf{w}^*) \propto \frac{1}{\{1 + (\mathbf{w}^*\mathbf{a} - \mathbf{z})' M^{-1} (\mathbf{w}^*\mathbf{a} - \mathbf{z}^*)\}^{(m-Nq)/2}} .$$

Completing the square in $\mathbf{a}$ gives

$$p(\mathbf{a}\,|\,\mathbf{z}^*,\mathbf{w}^*) \propto \frac{1}{\{\upsilon^* + (\mathbf{a}-\alpha^*)'\mathbf{Q}^{-1}(\mathbf{a}-\alpha^*)\}^{(\nu^*+h)/2}}, \tag{18}$$

where:

$$\alpha^* \equiv (\mathbf{w}^{*\prime}\mathbf{M}^{-1}\mathbf{w}^*)^{-1}(\mathbf{w}^{*\prime}\mathbf{M}^{-1}\mathbf{z}^*), \ \upsilon^* \equiv m\text{-}Nq\text{-}h \quad ,$$

$$\mathbf{Q}^{-1} \equiv (\mathbf{w}^{*\prime}\mathbf{M}^{-1}\mathbf{w}^*)\upsilon^*/\beta^* \quad ,$$

$$\beta^* \equiv 1 + \mathbf{z}^{*\prime}\mathbf{M}^{-1}\mathbf{z}^* - \alpha^{*\prime}\mathbf{w}^{*\prime}\mathbf{M}^{-1}\mathbf{w}^*\alpha^* \quad .$$

That is, **a** follows an *h-dimensional Student* t-density with mean α^*, and υ^* degrees of freedom. Then,

$$E(\mathbf{a}\,|\,\mathbf{z}^*,\mathbf{w}^*) = \alpha^* = (\mathbf{w}^{*\prime}\mathbf{M}^{-1}\mathbf{w}^*)^{-1}(\mathbf{w}^{*\prime}\mathbf{M}^{-1}\mathbf{z}^*) \quad .$$
$$\text{var}(\mathbf{a}\,|\,\mathbf{z}^*,\mathbf{w}^*) = (\upsilon^*/\upsilon^*\text{-}2)\mathbf{Q} \quad ,$$
$$\text{var}(\mathbf{a}\,|\,\mathbf{z}^*,\mathbf{w}^*) = (\beta^*/(\upsilon^*\text{-}2))\,(\mathbf{w}^{*\prime}\mathbf{M}^{-1}\mathbf{w}^*)^{-1} \quad . \tag{19}$$

Discussion of Prior

The mean and variance of the inverted Wishart distribution are well known (see e.g. Press, 1972, p. 111). Therefore

$$E(\Omega_0) = \mathbf{M}/(m\text{-}2Nq\text{-}2) \quad .$$

But if we subjectively believe that

$$\Omega_0 = \begin{pmatrix} \Sigma_1 & & \mathbf{0} \\ & \Sigma_2 \ddots & \\ \mathbf{0} & & \Sigma_2 \end{pmatrix},$$

we should take

$$\mathbf{M} \equiv \begin{pmatrix} \mathbf{M}_1 & & 0 \\ & \mathbf{M}_2 \ddots & \\ 0 & & \mathbf{M}_2 \end{pmatrix} \quad , \mathbf{M}_1 > 0, \mathbf{M}_2 > 0.$$

Then,

$$E(\Sigma_1) \equiv \frac{\mathbf{M}_1}{m\text{-}2Nq\text{-}2}, \qquad E(\Sigma_2) = \frac{\mathbf{M}_2}{m\text{-}2Nq\text{-}2},$$

and $E(\Omega_0) = \mathbf{0}$ for all elements of Ω_0 not in the block diagonal elements. Moreover, if $\Omega_0 \equiv (\omega_{\alpha\beta})$, for all (α,β) not in the block diagonal elements

$$\text{var}(\omega_{\alpha\beta}) = \frac{m_{\alpha\alpha}m_{\beta\beta}}{(m\text{-}2Nq\text{-}1)(m\text{-}2Nq\text{-}2)(m\text{-}2Nq\text{-}4)},$$

where $\mathbf{M} \equiv (m_{\alpha\beta})$. Note that $\text{var}(\omega_{\alpha\beta})$ is of order m^3; that is, $\text{var}(\omega_{\alpha\beta})$ goes to zero with increasing m^3. We can always choose m large enough so that all elements off the block diagonal elements of Ω_0 are centered at zero, with very small variance. Note from eqn. (19) that $\text{var}(\mathbf{a}|\mathbf{z}^*,\mathbf{w}^*)$ goes to zero with increasing υ^* (which is linear in m). By selecting $(\mathbf{M}_1,\mathbf{M}_2)$ appropriately, and choosing m sufficiently large this prior distribution will be sufficiently rich to accommodate many classes of subjective information.

This type of prior is not recommended for the general case, since the structure of the prior distribution is too restrictive(1). Our reasoning is that although elements of Ω_0 not in the blocks on the main diagonal are centered at zero with arbitrarily small variance, because there are only two parameters in the inverted Wishart distribution, viz. $(\mathbf{M},\mathrm{m})$, the elements of Ω_0 that do lie in the main diagonal blocks are simultaneously constrained in all of their moments (by taking m large). Such constraints may not always be desirable. For the general case, an alternative prior for Ω_0 which is richer in parameters is recommended. We propose such a prior below.

Generalized Prior Distribution

A generalized family of Wishart type distributions was introduced by Roux, 1971. The generalization includes hypergeometric functions of matrix argument. A form of the associated density which widens the parameter spaces is given (for a general *pds* matrix $\hat{\mathbf{X}}$) by

$$f(\hat{\mathbf{X}}) = c|\hat{\mathbf{X}}|^{\gamma-(q+1)/2} \exp\{-\text{tr}(\mathbf{J}\hat{\mathbf{X}})\}\, {}_{r}{}^{*}F_{q}{}^{*}(\delta;\eta;\mathbf{JRJ}\hat{\mathbf{X}}) \quad , \tag{20}$$

for $\hat{\mathbf{X}}$: $q\text{x}q$, $\hat{\mathbf{X}}>0$, $\delta \equiv (\delta_1,\ldots,\delta_r{}^*)'$, $\eta \equiv (\eta_1,\ldots,\eta_q{}^*)'$, $\mathbf{J}$: $q\text{x}q$, $\mathbf{R}$: $q\text{x}q$, $\mathbf{J}>0$, $\mathbf{R}>0$, and ${}_{r}{}^{*}F_{q}{}^{*}(\cdot)$ denotes the generalized hypergeometric function of matrix argument (see Constantine, 1963). The normalizing constant is given by

$$c = \frac{|\mathbf{J}|^{\gamma}}{\Gamma_q(\gamma)\,{}_{(r^*+1)}F_q{}^*(\delta,\gamma;\eta;\mathbf{JR})},$$

---

(1) The restrictiveness of the structural form of the Inverted Wishart distribution has already been noted by Rothenberg, 1963, in a different context (see References).

where $\Gamma_q(\gamma)$ denotes a q-dimensional gamma function. The parameters (δ_i,η_j), $i=1,\ldots,r^*$, $j=1,\ldots,q^*$, are restricted to take those values for which $f(\hat{\mathbf{X}})$ is positive.

Now let $\Omega_0 \equiv \hat{\mathbf{X}}^{-1}$, replace q by Nq (the dimension of Ω_0), and transform the density in (20) to yield the generalized inverted Wishart density

$$p_2'(\Omega_0) = \frac{c}{|\Omega_0|^{\gamma+(Nq+1)/2}} \exp\{-\mathrm{tr}(\mathbf{J}\Omega_0^{-1})\}\, {}_{r^*}F_{q^*}(\delta;\eta;\mathbf{JRJ}\Omega_0^{-1}) \quad . \tag{21}$$

Using eqn. (21) in eqn. (17), with $p'(\mathbf{a},\Omega_0) \propto p_2'(\Omega_0)$, gives the joint posterior density

$$p(\mathbf{a},\Omega_0|\mathbf{z}^*,\mathbf{w}^*) \propto |\Omega_0|^{-(2\gamma+Nq+2)/2} \exp\{(-\tfrac{1}{2})\mathrm{tr}\Omega_0^{-1}(\mathbf{H}+2\mathbf{J})\}$$

$$\bullet\ {}_{r^*}F_{q^*}(\delta;\mathbf{h};\mathbf{JRJ}\Omega_0^{-1}). \tag{22}$$

The marginal posterior density of $\mathbf{a}$ is found by integrating (22) with respect to Ω_0. The integration is carried out by reference to eqn. (21), using its normalizing constant. The result is

$$p(\mathbf{a}|\mathbf{z}^*,\mathbf{w}^*) \propto \frac{1}{|2\mathbf{J}+(\mathbf{z}^*\text{-}\mathbf{w}^*\mathbf{a})(\mathbf{z}^*\text{-}\mathbf{w}^*\mathbf{a})'|^{\gamma+1/2}}$$

$$\bullet\ ({}_{r^*+1)}F_{q^*}(\delta,\gamma+\tfrac{1}{2};\eta;\mathbf{JRJ}(2\mathbf{J}+\mathbf{H})^{-1}) \quad , \tag{23}$$

where: $\mathbf{H} \equiv (\mathbf{z}^*\text{-}\mathbf{w}^*\mathbf{a})(\mathbf{z}^*\text{-}\mathbf{w}^*\mathbf{a})'$. If we identify $\mathbf{M} \equiv 2\mathbf{J}$, $2\gamma \equiv m\text{-}Nq\text{-}1$, and take $\mathbf{R} \equiv \mathbf{0}$, it is immediately seen that the result obtained in (18), for the inverted Wishart prior, is a special case of eqn. (23). This result, however, has the advantage of being richer in parameters and can therefore accommodate a much greater variety of types of subjective information. Inferences about $\mathbf{a}$, however, are more complicated, and will require the use of zonal polynomial tables in order to evaluate the hypergeometric functions in (23) (see James and Parkhurst, 1974). The parameters of the hypergeometric functions are selected so as to satisfy the block diagonal structure of Ω_O.

4. CONCLUSIONS AND SUMMARY

The qualitative controlled feedback process of forming group judgments and making decisions has been examined from a Bayesian viewpoint. The group responses to many questions was modeled as an autoregressive process with coefficient vector $\mathbf{a}$.

It was shown that if the error covariance matrix, Ω, is known, the posterior distribution of **a** is normal, and centered at the generalized LSE. In large samples, if Ω is unknown, a consistent estimator may be used to make conditional inferences about **a**.

Bayesian inferences can also be made marginally, without reference to Ω. Assuming intraclass covariance structure, the marginal posterior distribution of **a** was shown to be, approximately, a complicated member of the matrix **T** family of distributions. We developed a normal distribution approximation which is very useful in large samples, however. For small sample situations involving the intraclass covariance structured situation we developed a posterior multivariate Student t-density for **a**. This result although useful for many situations is somewhat restrictive in the types of prior information it will accommodate. A more general result was obtained using generalized inverted Wishart distribution priors. The result is more complicated to use, however.

Finally, note that the entire QCF process is subjective in nature. It is therefore not surprising that inferences about the relationship between the responses of the panel members, and their individual characteristics and judgmental behavior regarding the reasons other panelists give, would depend heavily upon the nature and quantity of the prior information available.

ACKNOWLEDGMENT

I would like to thank Dr. Hiroshi Watanabe for his careful reading of the manuscript. I am also grateful to Professors J.R. Dunsmore and S. Geisser for helpful comments.

REFERENCES

CONSTANTINE, A.G. (1963). Some Non-central Distribution Problems in Multivariate Analysis, *Ann. Math. Statist.*, **34** 1270-1285.

JAMES, A.T. and A.M. PARKHURST (1974). *Selected Tables in Mathematical Statistics*, Vol. **2**. (H.L. Harter and D.B. Owen, eds.) California: Institute of Mathematical Statistics.

PRESS, S. JAMES (1972). *Applied Multivariate Analysis*. New York: Holt, Rinehart and Winston

— (1978). Qualitative Controlled Feedback for Forming Group Judgments and Making Decisions. *J. Amer. Statist. Assoc.* **73**, 526-535.

— (1980). Multivariate Group Judgments by Qualitative Controlled Feedback, In *Multivariate Analysis V*, (P.R. Krishnaiah, ed.) New York: North Holland, 581-591.

— (1979a). Matrix Intraclass Covariance matrices with Applications in Agriculture, *Tech. Rep.* **49**. Department of Statistics, University of California, Riverside.

— M.W. ALI, and E. YANG (1979b). An empirical Study of a New Method for Forming Group Judgments: Qualitative Controlled Feedback, *Technological Forecasting and Social Change*, **15**, 171-189.

ROTHENBERG, T.J. (1963) A Bayesian Analysis of Simultaneous Equation Systems, *Tech. Rep.* **6315**, Econometric Institute, Rotterdam.

ROUX, J.J.J. (1971) On Generalized Multivariate Distributions, *South African Statist*, **5**, 91-100.