DETERMINING OPTIMAL WEIGHTS FOR EXPERT JUDGMENT

Bernard Grofman and Scott L. Feld

1. INTRODUCTION

Consider a group of N members. Let $u_i(\ell)$ be the Von Neumann–Morgenstern utility of the ith individual for alternative ℓ. Define a group utility function $u(\ell)$ as a weighted additive function of the individual utilities:

$$u(\ell) = \sum_{i=1}^{N} a_i u_i(\ell). \tag{1}$$

The function specified in (1) can be regarded as a Bergson–Samuelson social welfare function (SWF; Samuelson, 1977). A number of authors have investigated the properties of such an additive SWF and shown it to have various desirable attributes. (See discussion in Bodily, 1979, p. 1036; Harsanyi, 1955; Stone, 1961; Keeney and Kirkwood, 1975; Keeney, 1976; Morris, 1974, 1977; Lehrer and Wagner, 1981; but cf. Bordley, 1980a, b, and essay in this volume.)

Various authors have proposed mechanisms for assigning the weights in Eq. (1) (see, e.g., Theil, 1963, 1969, 1970; Roskamp and McMeekin, 1970), of which Harsanyi's (1955) weighted average proposal is the best known.

In this paper we shall be interested in the question of how the a_i values in Eq. (1) ought to be assigned. If the members of the group are "experts" whose advice is sought, then we have the problem of optimally pooling expert judgments. We shall review three different but related approaches; that of Kreweras–DeGroot–Bodily, that of Mirkin, and that of Shapley–Grofman–Nitzan–Paroush.[1]

2. BODILY ALGORITHM

Bodily (1979) offers a delegation process in which each committee member i assigns voting weights w_{ij}, $0 \leq w_{ij} \leq 1$, to each of his fellow committee members j, such that $\sum_j w_{ij} = 1$ for all i. This gives rise to a delegation matrix W which can be regarded as a Markov chain. For W ergodic, it is well known that there exists an eigenvector **a** such that $\mathbf{a}W = \mathbf{a}$ (see, e.g., Feller, 1957). Bodily (1979) proposes that the a_i values in this eigenvector be used as the weights in Eq. (1). The weight w_{ij} reflects i's judgment of how large a role j's utility function should play in the group decision making. The ith entry a_i of the unit eigenvector **a** reflects a group "consensus" as to the relative weights to be given each member's preferences. As far as we are aware, the first author to propose this method of assigning decision weights was Kreweras (1965). De Groot (1974) also has independently proposed this method. It appears to be well known in the Soviet Union (Mirkin, 1979; also Cherous'ko, 1972; Bruk and Burkov, 1972; and other references cited in Mirkin). The Kreweras–De Groot–Bodily algorithm has also been used as a way of determining the relative skill levels of players in a round-robin tournament (Berge, 1975; Mirkin, 1979, pp. 150–153).

3. MIRKIN ALGORITHM

One practical difficulty with the Kreweras–De Groot–Bodily method is that it requires individuals to make assessments of the relative competence of their fellows. Even if these assessments are taken via secret ballots, individuals may have some reluctance to make such judgments. Mirkin (1979) has proposed an ingenious way to bypass direct assessments. We can motivate his model best in the context of a situation in which a group of individuals is asked to assess the utility of alternatives with respect to some vaguely stated but generally agreed upon standard such as the "public interest." Assume, further, that individuals may differ in their "expertise" with respect to assessing this criterion (see, e.g., Einhorn et al., 1977). Mirkin (1979, p. 153) asserts that "one often thinks" that competence "should be evaluated with respect to the level of consistency of... evaluations with those of the majority." Mirkin (1979) does not give any theory to provide support to his "one often thinks," but he does present a simple algorithm to assign weights to individuals based on their concordance with the views of their fellows. His algorithm is very similar in spirit to the Kreweras–De Groot–Bodily approach. Let K be the number of alternatives to be evaluated. We define u as a $K \times N$ matrix with $u_i(j)$ as its ijth entry.

Mirkin (1979, p. 155) proposes that we assign weights such that

$$\mathbf{q}\boldsymbol{\lambda}UU^t = \mathbf{q}. \tag{2}$$

Note that UU^t is always a square matrix. It need not however be a stochastic matrix. The vector $\boldsymbol{\lambda}$ is a normalizing factor: $\lambda_i = \sum_j (UU^t)_{ij}$, which turns it into a stochastic matrix. For the method to work we also require UU^t to be nondecomposable. This will be true if and only if U is nondecomposable (Mirkin, 1979, p. 156).

For the example in Bodily (1979, p. 1039) we have the following:

$$U = \textit{individual} \begin{cases} 1 \\ 2 \\ 3 \end{cases} \overset{\textit{project}}{\overset{\begin{matrix} 1 & 2 & 3 \end{matrix}}{\begin{bmatrix} .6 & .3 & .1 \\ .3 & .6 & .1 \\ .3 & .3 & .4 \end{bmatrix}}}; \tag{3}$$

$$U^t = \textit{project} \begin{cases} 1 \\ 2 \\ 3 \end{cases} \overset{\textit{individual}}{\overset{\begin{matrix} 1 & 2 & 3 \end{matrix}}{\begin{bmatrix} .6 & .3 & .3 \\ .3 & .6 & .3 \\ .1 & .1 & .4 \end{bmatrix}}}; \tag{4}$$

$$UU^t = \textit{individual} \begin{cases} 1 \\ 2 \\ 3 \end{cases} \overset{\textit{individual}}{\overset{\begin{matrix} 1 & 2 & 3 \end{matrix}}{\begin{bmatrix} .46 & .37 & .31 \\ .37 & .46 & .31 \\ .31 & .31 & .34 \end{bmatrix}}}; \tag{5}$$

$$\frac{UU^t}{\boldsymbol{\lambda}} = \begin{bmatrix} .40 & .32 & .27 \\ .32 & .40 & .27 \\ .32 & .32 & .36 \end{bmatrix}. \tag{6}$$

Solving matrix (6) for its asymptotic value we obtain

$$\mathbf{q} = (.35, .35, .30).$$

Bodily (1979, p. 1039) posited a delegation matrix W:

$$W = \begin{bmatrix} 0 & .4 & .6 \\ 1 & 0 & 0 \\ .5 & .5 & 0 \end{bmatrix}. \tag{7}$$

Solving the asymptotic vector for that matrix we obtain $\mathbf{a} = (.44, 3.0, .26)$. When two individuals "think alike" in the way that they evaluate alternatives, then they may also be more likely to give each other high evaluations—in which case the Kreweras–De Groot–Bodily and the Mirkin procedures will give very similar results.

The procedure in Eq. (2) for obtaining a vector $\mathbf{q}$ of competence coefficients is, as Mirkin (1979, p. 156) points out, theoretically well grounded. The elements b_{ij} of the matrix UU^t are the dot products of the evaluation vectors of individuals i and j. These quantities "express the degree of closeness of evaluations since two individuals that have high evaluations of the same objects (hence also low evaluations of the same objects) will have larger values of b_{ij}" (Mirkin, 1979, p. 156, with some change in wording). The iterative process does not affect the basic nature of this relationship. Moreover, we can show that $\mathbf{q}$ has optimally captured the relationships between the individual judgments in that $\mathbf{q}$ is proportional to what is called in statistics the *principal inner factor* (Mirkin, 1979, p. 156).

Consider any symmetric matrix B whose coefficients characterize interrelationships between elements 1, 2, . . . , N:

> Let component t_i of vector $t = (t_1, \ldots, t_N)$ increase in value as it becomes more related to the other elements of t. Then the quantity $t_{ij} = t_i t_j$ characterizes the degree of closeness between i and j . . . The less that matrix $T = t_{ij}$ differs from B, the better vector t estimates the relationships between the elements in B. The vector t for which the difference between T and B attains its minimum over all t is . . . the principal inner vector. (Mirkin, 1979, pp. 156–157)

Because $\mathbf{q}$ is proportional to the principal inner factor, we know that the competence coefficients are determined by the same factor that determines the degree of similarity between experts.[2] Mirkin (1979, p. 155) points out that, using his method, we can also directly solve for the *utility* values. First, we find $\mathbf{x}$ such that

$$\mathbf{x}\gamma UU^t = \mathbf{x}. \tag{8}$$

Then we let

$$u(\ell) = x_\ell. \tag{9}$$

Here γ is, as before, a normalizing factor, in this case given by $\gamma_i = \sum_j (U^tU)_{ij}$. Using Eq. (9), we obtain $\mathbf{x} = (.38, .35, .26)$ for our earlier example. This choice of the first alternative as the most favored one corresponds to that obtained by the Kreweras–De Groot–Bodily method (see Bodily, 1979, p. 1040). It is also the choice that would arise from a weighted majority voting rule applied to the matrix WU, where

$$\mathrm{WU} = \begin{bmatrix} .30 & .44 & .28 \\ .60 & .30 & .10 \\ .45 & .45 & .10 \end{bmatrix}, \qquad (10)$$

since the weighted utility sum is maximum in column 1.

4. SHAPLEY–GROFMAN–NITZAN–PAROUSH DECISION RULE

The weights associated with the Bayesian optimal group decision rule, that is, with $\log[p_i/(1-p_i)]$, where p_i is the judgmental competence of the ith individual, seem to us to be the most satisfying answer to the question of what weights should be used in constructing a SWF—at least in contexts where there is general agreement as to the (perhaps fuzzy) criterion of choice but differing individual assessments of the value of the alternatives with respect to that criterion (see Shapley and Grofman, 1984; Grofman and Owen, this volume). A natural objection to use of Bayesian optimal weights is that they are unknown—and we cannot use weights which we have no way of measuring. To this difficulty we offer four possible solutions. First, we can use the Kreweras–De Groot–Bodily-delegation process to obtain a "group concensus" assessment of the p_i values of each of the group members and then insert those p_i values in Eq. (3) on page 95 to estimate the Bayesian optimal weights, rather than using the p_i values themselves as the weights.[3] Second, we can use the Mirkin (1979) procedure to generate for each individual a concordance value with the group consensus rankings, which we can then convert to a p_i (and thence to a w_i value), rather than simply taking the $\mathbf{q}$ value itself to provide the optimum weightings as does Mirkin (1979).

Third, in a fashion similar to that in the Mirkin (1979) procedure, we can construct for each individual a concordance value with the group majority choice on each pairwise comparison and use that aggregate concordance figure as an estimate of p_i to be substituted in Eq. (3) on page 95 to find the Bayesian optimal weights. In certain simple cases, this estimate is not unreasonable (Grofman, Owen, and Feld, 1983; Feld and Grofman, 1984).

Finally, if experts have established a "track record" of *past* performance, we can assign p_i values based on this track record and then use Eq. (3) on page 95. When such over-time data is available, this is the method which we recommend. It is both simple and sensible. Of course, we must be careful that the new questions being raised are sufficiently similar to those previously dealt with as to make that previous performance a useful predictor.

ACKNOWLEDGMENTS

This research was supported by NSF Grant No. SES 80-07915. We are indebted to helpful conversations with Lloyd Shapley and Guillermo Owen. Errors are the sole responsibility of the authors. The manuscript was typed and edited by Lillian White and the staff of the Word Processing Center, School of Social Sciences, University of California, Irvine, and by Sue Pursche. We would also like to thank Laurel Eaton and Sue Pursche for bibliographic assistance.

NOTES

1. We shall not deal with interactive techniques for arriving at group judgment such as Delphi and Span (see Helmer, 1963; Dalkey and Helmer, 1963; Brown and Helmer, 1964; MacKinnon, 1966a, b; Dalkey, 1969a, b, c, 1970; Gustafson et al., 1973; Del Becq et al., 1975; MacKinnon and Anderson, 1976; Sackman, 1974; cf. Einhorn, 1972, 1974; Morris, 1977; McCarthy et al., 1980).
2. Mirkin (1979, p. 157) goes on to note that "Of course, $\sum_{i,j} (b_{ij} - t_{ij})^2$ may be large, so not that all relations are 'linearly' described by one factor." When this occurs it would indicate inconsistency in the evaluations of the experts. For this case of multiple dimensions of evaluation, Mirkin (1979, pp. 157–163) discusses alternative techniques and shows how the ideas discussed above can be related to factor analysis, and how UU^t can be normalized so as to become a matrix of correlation coefficients. We shall not, however, pursue these notions further in this paper. Rather, we shall deal with models where one underlying factor, that of judgmental skill, is at the root of the similarities and differences in individual evaluations of alternatives and in competence assessments [see also Bodily, 1979, p. 1040, Eq. (7)]. We shall also neglect complications caused by multifaceted problems where different experts may each have only a "piece" of the puzzle (see Lorge and Solomon, 1955).
3. If it is reasonable to suppose that individuals judge the competence of others by the extent to which others are thought likely to agree with oneself, then the matrix of weights generated by the Kreweras-De Groot-Bodily procedure is a variance-covariance matrix whose eigenvector has ith entry equal to $2p_i - 1$ (Batchelder and Romney, this volume).

R24