COWLES COMMISSION DISCUSSION PAPER: STATISTICS NO. 386

The Quadratic Team. I.*

Roy Radner

August 2, 1954

Best decision rules and the value of information for many decision makers with a common quadratic payoff function.

---

* Research undertaken by the Cowles Commission for Research in Economics under contract Nonr-358(01), NR 047-006 with the Office of Naval Research.

Introduction

The Statistical Decision Problem, as developed in recent years, is concerned with an individual who makes a decision and is thereupon rewarded according to his choice and the prevailing state of Nature. As an extension of this we might consider several individuals each deciding about something different, but receiving a common reward as a result of their joint choice. Prof. Marschak has called such a group a team, to emphasize that the group members have a common goal, and distinguish it from a group whose members have conflicting interests and whose study is therefore threatened with all the uncertainties of the general theory of games. This discussion of the team will be devoted to the implications of the fact that different decisions may be based on different information.

In the one person problem part of the decision made by the individual may be concerned with obtaining information about the state of nature. We suppose that ultimately the individual is to take some action a and that his payoff will be $u(a,x)$ where x denotes the state of Nature, which is random. Which action is chosen may depend upon the information $y = \eta(x)$, so that $a = \alpha(y)$, where α is called the decision function. For any given function η, which we will call the information structure, we assume that the individual will choose his decision function α so as to maximize the expected value *

$$E\, u(\alpha(y),x)$$

provided, of course, that he knows the probability distribution of x. In general, however, the information structure will cost something; we will assume that this cost $C(\eta)$ can simply be subtracted from the expected payoff above, so that the complete decision problem for the individual is:

* The symbol E will denote expected value.

"choose an information structure η and a decision function α which maximizes the net expected payoff $E\, u(\alpha(y),x) - C(\eta)$"

A few words may be necessary at this point, to warn the reader that the definition of "state of Nature" in any problem will usually be relative to the set of alternative information structures being considered. For example, some cookbooks say that a roast of beef is done (rare) when the inside temperature of the roast reaches 140 degrees. Thus one might first say that for the decision about when to take the roast out of the oven, the relevant state of nature is the internal temperature of the meat. But if one's information about this temperature consists of the reading on a cheap thermometer, then immediately one must include the error of the thermometer in the description of Nature. On the other hand, one's information might be the length of time the roast has been in the oven, the weight of the roast, etc., in which case these variables would have to be included in the state of nature, together with some statements about the (statistical) relationship between them and the temperature.

For the <u>team problem</u> there will be several action variables a_i and as many (in general different) information variables y_i, where $y_i = \eta_i(x)$, and as many decision functions α_i. The set of functions η_i together constitute the information structure. The expected net payoff for the team is thus

$$E\, u(\alpha_1(y_1),\ldots, \alpha_m(y_m), x) - C(\eta_1,\ldots,\eta_m) \;.$$

If we call $a = (a_1,\ldots,a_m)$ the team action, $\alpha = (\alpha_1,\ldots,\alpha_m)$ the team decision function, etc., we see that the team can be formally reduced to an individual. In fact we may say that <u>the team problem is an individual decision problem in which the action variable has many components, each of which may be made to depend upon different information</u>.

If $\hat{\alpha}$ is the best team decision function for a given η, then we will call the expectation

$$U(\hat{\alpha},\eta) = E\, u(\hat{\alpha}_1(y_1),\ldots,\hat{\alpha}_m(y_m),\, x)$$

the value of the information structure η. Our aim here is to study methods of finding best decision functions for given η, and to investigate how the value of an information structure depends upon the properties of the payoff function $u(a_1,\ldots,a_m,x)$ and the probability distribution of the state of Nature x.

The two simplest types of payoff functions u we can consider are either linear or quadratic in the action variables a_i. In the linear case the "gross expected payoff" $E\, u(a_1(y_1),\ldots,x)$ is the sum of terms each of which depends on only one decision function α_i so that each component of η makes a contribution to the total value of the information structure which is independent of the other information components. Thus for our purposes, the quadratic case is the simplest one of real interest, and this paper will be entirely restricted to consideration of a quadratic payoff function[1/]

In the first part of this paper, we will consider the situation in which the probability distribution of the state of Nature is known. It will be shown that for any given information structure there is a unique best team decision function, which is in fact the projection, in a certain Hilbert space, of the decision function which would be best under complete information onto a subspace which represents the given information structure. From this we will be able to show how the best payoff depends upon the interactions between the different decision variables and upon the correlation between the information (in a certain sense) and the decision function

1/ This is not to say that consideration of the cost of information cannot lead to many interesting problems of interdependence in the linear case. See, for example, M. Beckmann, "On Marschak's Model of an Arbitrage Form", CCDP: Economics No. 2058.

which would be best under complete information. A necessary and sufficient condition for a best decision function is then given together with an application to the case of normally distributed information. Finally, the preceding general remarks are applied to a discussion of certain special types of information structure.

In the second part of this paper we will consider the situation in which the probability distribution of the state of Nature is unknown.

Part I: The Case of Known Distribution of the State of Nature

1. The Existence and Uniqueness of the Best Decision Function for Fixed Information Structure.

The quadratic team decision problem for fixed information structure is defined by

(i) a real n-dimensional space of decision variables $a = (a_1, \ldots, a_n)$,

(ii) a probability space X with elements x which describe the state of Nature, and a given probability measure on X.

(iii) a payoff function $u(a,x)$, which is a quadratic function of the a_i, for almost every x. Without loss of generality we can represent this quadratic function in matrix terms by:

(3) $\quad u(a,x) = - a\, Q(x)\, a' + 2r(x)\, a'$

where $Q(x)$ is a matrix and $r(x)$ is a vector whose elements are real-valued functions of x, and $E\, r(x) = 0$. We will assume that $Q(x)$ is symmetrix positive definite for almost every x and that $E\,[r_i(x)]^2$ exists for all i.

(iv) an n-tuple of information variables $y = (y_1, \ldots, y_n)$ which are defined by $y_i = \eta_i(x)$ where $\eta = (\eta_1, \ldots, \eta_n)$ is a given n-tuple of functions, called the information structure. The variables y_i can be considered to be elements of probability spaces Y_i, with probability measures which are induced by the measure on X through the functions η_i.

(v) a space $\mathcal{A}$ of n-dimensional decision functions $\alpha = (\alpha_1, \ldots, \alpha_n)$ such that α_i is a real valued function on Y_i. In general $\mathcal{A}$ will be taken to be all functions α of this form for which the expectation

$$E\,[\alpha_i(y_i)]^2$$

exists.

Given the information structure η, an optimal set α of decision functions is defined as one which maximizes

$$U(\alpha) \equiv E\,[u(\alpha_1(y_1),\ldots,\alpha_n(y_n),\ x]$$

The matrix $Q(x)$ is of central importance in our discussion for it is the thing that gives our problem its many-person character. For any given x,

$$\frac{\partial u\ (a,x)}{\partial a_i \partial a_j} = -\,q_{ij}$$

or in economic terms, the marginal contribution of one action variable a_i depends upon the value of another action variable a_j only insofar as $-q_{ij}$ is different from zero. The term $(-q_{ij})$ might be called the <u>interaction between the i and j decision variables</u>. We shall see that unless $Q(x)$ is (almost always) diagonal the consequences of a change in one information η_i will usually depend upon the nature of the other information variables η_j, and it will be one of our principal concerns to investigate this <u>interaction between information variables</u>. In the special case in which $Q(x)$ is diagonal the expected payoff becomes

$$U\ (\alpha) = \sum_i E\ [-q_{ii}(x)\ (\alpha_i(y_i))^2 + 2r_i(x)\ \alpha_i(x)]$$

$$= \sum_i U_i(\alpha_i)$$

and the problem degenerates into many "one-person" problems.

$\mathcal{A}$ is a linear subspace of the linear space $\mathcal{H}$ of all functions v from X to n-dimensional Euclidean space such that $E[v_i(x)]^2$ exists for all i. 1/

1/ Strictly speaking, it is the set of all β of the form $\beta_i(x) = \alpha_i(\eta_i(x))$ which forms this subspace of $\mathcal{H}$; we will call this $\mathcal{A}$, too.

We can define an inner product between any v and w in $\mathcal{H}$ by

$$(v,w) \equiv E\, v(x)\, Q(x)\, w(x)'$$

where Q is the matrix in equation (3) above. The norm of u is then of course

$$\| u \| = \left\{ E\, u(x)\, Q(x)\, u(x)' \right\}^{\frac{1}{2}}$$

It is easy to see that (u,v) is actually an inner product and that $\mathcal{H}$ is thus a Hilbert space. Furthermore, since L_2 of Y_i is complete (for each i), $\mathcal{A}$ is a closed linear subspace of $\mathcal{H}$.

Completing the square in equation (3) above:

$$\text{(3a)} \qquad u(a,x) = -\left(a - r(x)\, Q(x)^{-1}\right) Q(x) \left(a - r(x)\, Q(x)^{-1}\right)' + r(x)\, Q(x)^{-1}\, r(x)'$$

and the expected payoff using the decision function α is

$$U(\alpha) = -E\left[\alpha(\eta(x)) - r(x)\, Q(x)^{-1}\right] Q(x) \left[\alpha(\eta(x)) - r(x)\, Q(x)^{-1}\right]' + E\, r(x)\, Q(x)^{-1}\, r(x)'$$

or more briefly, letting $z(x) = r(x)\, Q(x)^{-1}$:

$$\text{(4)} \qquad U(\alpha) = \| z \|^2 - \| \alpha - z \|^2$$

The decision problem is to find that α in $\mathcal{A}$ which maximizes $U(\alpha)$, which according to (4) is equivalent to minimizing the distance between α and z. Since $\mathcal{A}$ is closed we immediately have the following theorem:

Theorem 1. For $\mathcal{A}$ as defined above, there exists a unique optimal decision function $\hat{\alpha}$ in $\mathcal{A}$; $\hat{\alpha}$ is the orthogonal projection of z on $\mathcal{A}$.

It follows from (4) and the Pythogoream Theorem that the best expected payoff is the square of the length of $\hat{\alpha}$, i.e.

(5) $$U(\hat{\alpha}) = \| \hat{\alpha} \|^2$$

Its maximum value is, of course, $\| z \|^2$, which occurs when z is in $\mathcal{A}$, and its minimum is zero, which occurs when z is orthogonal to $\mathcal{A}$.

If we are to say something more explicit about $\hat{\alpha}$ we must describe $\mathcal{A}$ more explicitly. One such description is by means of an orthonormal basis $\{\beta^j\}$ of $\mathcal{A}$. In this case $\hat{\alpha}$ can be described simply as

(6) $$\hat{\alpha} = \sum_j (\beta^j, z)\, \beta^j$$

$$\text{and} \quad U(\hat{\alpha}) = \sum_j (\beta^j, z)^2$$

This reduces the problem to one where the main task is to translate the description of the information structure η into the specification of an orthonormal basis of $\mathcal{A}$.

To gain further insight from this approach let us look more carefully at the special character of our inner product. For simplicity, suppose that the matrix $Q(x)$ is just a constant matrix Q. Using the fact that there is a non-singular matrix P such that

$$Q = PP'$$

we can rewrite the inner product as

$$(v,w) = E\left\{(vP)\ (wP)'\right\}$$

If we let $vP = (v_1^*, \ldots, v_n^*)$

$$wP = (w_1^*, \ldots, w_n^*)$$

Then

$$(v,w) = \sum_i E\, v_i^*\, w_i^*$$

if $Ev = Ew = 0$,
That is, the inner product of v and w is the sum of the covariances of the corresponding coordinate variables of vP and wP.

Heuristically, we might say that the best payoff for a given information structure is measured by the "total covariance" between the transformed random vector zP and the transformed subspace $\mathcal{Q}P$. Recall now the definition of z:

$$z(x) = r(x) Q^{-1}$$

As is clear from equation (3a), for every x the best possible decision is $a = z(x)$. That is, z is the correct decision function when there is full information. On the other hand, we have seen how the subspace $\mathcal{Q}$ is a complete characterization of the information structure. Hence we might say, again heuristically, that the best payoff for a given structure of information is measured by the "total covariance" between the transformed correct decision and the transformed information, the transformation being the one which diagonalizes the matrix Q of interactions.1/

[If $Q(x)$ does actually vary with x, then instead of P we have a random matrix $P(x)$ and the transformation discussed above is a random one.]

It is interesting to note that the quadratic team problem reduces to the familiar minimum mean squared error prediction problem when there is only one action variable. In that case the best prediction of $z(x)$ on the basis of the information $\eta(x)$ is the expected value of $z(x)$ given $\eta(x)$. The thing to note here is that this rule does not go over simply into the many dimensional problem, i.e., it is not true in general that the best decision $\alpha_i(\eta_i(x))$ is equal to the expected value of $z_i(x)$ given $\eta_i(x)$, unless the matrix $Q(x)$ is almost always the identity matrix. The form of the best α_i will in general depend not only on the form of the information function η_i, but on all the other η_j as well.

1/ Recall that $Ez = 0$.

2. A Necessary and Sufficient Condition for Best Decision Functions

If $\hat{\alpha}_1, \ldots, \hat{\alpha}_n$ are optimal decision functions for a given information structure η then certainly $\hat{\alpha}_1$ maximizes $U(\alpha_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_n)$ with respect to α_1. But

$$\max_{\alpha_1} U(\alpha_1, \hat{\alpha}_1, \ldots, \hat{\alpha}_n, x) = \max_{\alpha_1} E\left\{E\left[u(\alpha_1(y_1), \hat{\alpha}_2(y_2), \ldots, \hat{\alpha}_n(y_n), x) \mid y_1\right]\right\}$$
$$= E\left\{\max_{a_1} E\left[u(a_1, \alpha_2(y_2), \ldots, \hat{\alpha}_n(y_n), x) \mid y_1\right]\right\}$$

Hence for every y_1, $\hat{\alpha}_1$ must take on a value which maximizes

$$E\left\{u(a_1, \hat{\alpha}_2(y_2), \ldots, \hat{\alpha}_n(y_n), x) \mid y_1\right\}$$

with respect to a_1. Similar statements can be made for $\hat{\alpha}_2, \ldots, \hat{\alpha}_n$. This can all be summarized as follows:

<u>A necessary condition that $\hat{\alpha}_1, \ldots, \hat{\alpha}_n$ be optimal is that they simultaneously satisfy the n conditions</u>

$$\text{(M)} \qquad \hat{\alpha}_i(y_i) = \underset{a_i}{\text{max'er}}\; E\left\{u(\hat{\alpha}_1(y_1), \ldots, a_i, \ldots, \hat{\alpha}_n(y_n), x) \mid y_i\right\}, \quad i = 1, \ldots, n$$

In general (that is, for a general payoff function) condition (M) will not be sufficient to ensure a maximum. We should mention a trivial case in which it is sufficient, namely when the payoff function is a sum of terms each of which depends upon one decision variable only, i.e. $u(a_1, \ldots, a_n, x) = u_1(a_1, x) + \ldots + u_n(a_n, x)$. In this case

$$U(\alpha_1, \ldots, \alpha_n) = E\left\{u(\alpha_1(y_1), x)\right\} + \ldots + E\left\{u(\alpha_n(y_1), x)\right\}$$

and our "many-person" decision problem has degenerated into n independent "one person" problems.

The main object of this section is to prove the following:

Theorem 2. For the quadratic team problem defined in section 1, Condition (M) is both necessary and sufficient to determine the optimal decision function for any given information structure, and has a unique solution.

Proof. By Theorem 1 it will be sufficient to show that if (M) has a solution, it is unique. Suppose on the contrary that α and β are two distinct solutions of (M). Form a convex combination γ of the two.

$$\gamma_i(y_i) = \lambda\alpha_i(y_i) + (1 - \lambda)\,\beta_i(y_i), \qquad 0 \leq \lambda \leq 1$$

Taking the derivative of $U(\gamma)$ with respect to λ, (see (4)):

$$\frac{d}{d\lambda}\, U(\gamma) = \frac{d}{d\lambda}\left[\|\, z\, \|^2 - \|\, \gamma - z\, \|^2\right]$$

$$= -\frac{d}{d\lambda}\,(\gamma - z,\ \gamma - z)$$

$$= -2\,(\gamma - z,\ \frac{d}{d\eta}\,[\gamma - z])$$

$$\frac{d}{d\lambda}\, U(\gamma) = -2\,(\gamma - z,\ \alpha - \beta)$$

and

$$\frac{d^2}{d\lambda^2}\, U(\gamma) = -2\,(\alpha - \beta,\ \alpha - \beta)$$

Since $\alpha - \beta \neq 0$,

$$\text{(7)} \qquad \frac{d^2}{d\lambda^2}\, U(\gamma) < 0$$

On the other hand, we can rewrite $\frac{dU}{d\lambda}$ as:

$$\frac{dU}{d\lambda} = -2\, E \sum_j\, [\alpha_j(y_j) - \beta_j(y_j)]\, [\sum_i q_{ij}(x)\, \gamma_i(y_i) - r_j(x)]$$

$$= -2\, E \sum_j\, [\alpha_j(y_j) - \beta_j(y_j)]\, E\left\{\sum_i q_{ij}(x)\, \gamma_i(y_i) - r_j(x)]\ \middle|\ y_j\right\}$$

Consider one of the conditional expectations $E\left\{\ldots\ldots|y_j\right\}$ in the above expression. For $\lambda = 1$ this equals

$$(8) \qquad E\left\{\sum_i q_{ij}(x)\,\alpha_i(y_i) - r_j(x) \mid y_j\right\}$$

This is in fact equal to (for fixed y_j)

$$\frac{1}{2}\,\frac{\partial}{\partial\alpha_j(y_j)} \quad E\left\{\alpha(y)\,Q(x)\,\alpha(y)' - 2\,r(x)\,\alpha(y)' \mid y_j\right\}$$

But since α satisfies condition (M) this partial derivative must be zero. Hence

$$(9a) \qquad \frac{dU}{d\lambda}\Big|_{\lambda=1} = 0$$

and by a similar argument about β

$$(9b) \qquad \frac{dU}{d\lambda}\Big|_{\lambda=0} = 0$$

But (9a) and (9b) together contradict (7), which shows that there can be only one solution of condition (M). Q.E.D.

3. <u>Normality and Linearity</u>

One who is familiar with the theory of minimum variance estimation and prediction might guess that the normal distribution would have a special place in our theory when the payoff function is quadratic. Such a guess would be correct as will be shown in the present section. The main result here is essentially that if all the information variables and the given random variables which enter into the payoff function are normally distributed, and the matrix Q is constant, then the best decision functions are linear in the information variables.

We now make the assumption:

(N) (a) The information variables y_i, which are vectors, and the variables $r_i(x)$ in equation (2.1) all have a joint normal distribution. Without loss of generality we may assume that $Er_i(x) = 0$ for all i.

(b) The elements $q_{ij}(x)$ of $Q(x)$ are constant (for almost all values of x).

Theorem 2. Assumption (N) implies that for the best decision function $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_n)$ each $\hat{\alpha}_i$ is linear in y_i.

Proof.

According to Theorem 1, the present theorem will be proved if we can show that condition (M) is satisfied by a decision function $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_n)$ such that $\hat{\alpha}_i$ is a linear function of y_i. As we saw in the proof of Theorem 1, in the quadratic case condition (M) is equivalent to setting expression (8) equal to zero for each j; that is:

$$(10) \qquad \sum_i q_{ij}\, E\,[\alpha_i(y_i) \mid y_j] = E\,[r_j(x) \mid y_j], \quad \text{all } j$$

If α is linear we can represent it as

$$(11) \qquad \alpha_j(y_j) = \sum_{k=1}^{m_j} a_{jk}\, y_{jk}, \qquad j = 1, \ldots, n$$

where y_{jk} $(k = 1, \ldots, m_j)$ are the coordinates of y_j. The linearity of α_i together with the normality of the y_i's together imply that $E\,[\alpha_i(y_i) \mid y_j]$ is a linear function of y_j, the coefficients in this linear function being determined by the a_{ik} and the means and covariances of the coordinates of y_i and y_j. Similarly, $E\,[r_j(x) \mid y_j)$ is a linear function of y_j.

We may assume, without loss of generality, that for each j the coordinates y_{jk} are independent, with means zero and variances one. The correlation matrix R of all the variables $y_{11}, \ldots, y_{1m_1}, \ldots, y_{n1}, \ldots, y_{nm_n}$ can be partitioned into blocks R_{ij}, where R_{ij} is the matrix of correlations ρ_{hk}^{ij} between the variables y_{ih} and the variables y_{jk}. Note that the matrix R is non-negative semi-definite, and the blocks R_{ii} on the diagonal are m_i by m_i identity matrices.

In terms of these correlations:

$$E\,[\alpha_i(y_i) \mid y_j] = \sum_{h=1}^{m_i} a_{ih} \sum_{k=1}^{m_j} \rho_{hk}^{ij}\, y_{jk} \tag{12}$$

The conditional expectations $E\,[r_j(x) \mid y_j]$ will be of the form:

$$E\,[r_j(x) \mid y_j] = \sum_{k=1}^{m_j} r_{jk}\, y_{jk} \tag{13}$$

Substitution of (11) (12) and (13) in condition (10) gives the condition:

$$\sum_i q_{ij} \left[\sum_{h=1}^{m_i} a_{ih} \sum_{k=1}^{m_j} \rho_{hk}^{ij}\, y_{jk} \right] = \sum_{k=1}^{m} r_{jk}\, y_{jk} \quad , \qquad j = 1, \ldots, n \tag{14}$$

Condition (11) must hold for almost all values of the y_j; hence in each equation, the coefficients of corresponding variables on each side must be equal, which gives:

$$\sum_i q_{ij} \sum_{h=1}^{m_i} a_{ih}\, \rho_{hk}^{ij} = r_{jk} \qquad , \; k = 1, \ldots, m_j \quad j = 1, \ldots, n \tag{15}$$

The coefficients of the unknown numbers a_{jk} in (15) form a matrix B which is made up of blocks $B_{ij} = q_{ij}\, R_{ij}$ $(i,\, j = 1, \ldots, n)$. Our object is to show that equations (15) can be solved for the a_{jk}; this will be accomplished if we can show that the matrix B is non-singular, and this in turn will be true if B is positive definite.

Let v be any vector with $\sum_i m_i$ coordinates, and let these coordinates be denoted v_h^i, $(i = 1,\dots,n;\ h = 1,\dots,m_i)$.

$$vBv' = \sum_{i,h} \sum_{j,k} q_{ij}\, \rho_{hk}^{ij}\, v_h^i\, v_k^j \tag{16}$$

Since R is non-negative, it can be expressed as the sum $\sum_p R(p)$ of P non-negative matrices $R(p)$ each of rank one. For any non-negative matrix $c = (c_{ij})$ of rank one there is a vector $c = (c_i)$ such that $c_{ij} = c_i c_j$.1/ For each matrix $R(p)$ let $\left\{\rho_h^i\,(p)\right\}$ be the corresponding vector, i.e.

$$\rho_{hk}^{ij} = \sum_p \rho_h^i\,(p)\, \rho_k^j\,(p)$$

Substituting this in (16) gives

$$\begin{aligned} vBv' &= \sum_{ij} q_{ij} \sum_{hk} \sum_p \rho_h^i\,(p)\, \rho_k^j(p)\, v_h^i\, v_k^j \\ &= \sum_p \sum_{ij} q_{ij} \sum_h \rho_h^i(p)\, v_h^i \sum_k \rho_k^j\,(p)\, v_k^j \\ &= \sum_p \sum_{ij} q_{ij}\, w_i(p)\, w_j(p) \end{aligned}$$

where $\quad w_i(p) \equiv \sum_h \rho_h^i\,(p)\, v_h^i$

Hence $vBv' \geq 0$ for all v. Now let v be different from zero; then for some i the m_i dimensional vector $(v_1^i,\dots,v_{m_i}^i)$ is different from zero. For that i consider the sub-matrix R_{ii} of R, and the corresponding sub-matrices $R_{ii}(p)$ of the matrices R_p.

1/ Halmos, Finite Dimensional Vector Spaces, §69, Theorem 1.

$$R_{ii} = \sum_p R_{ii}(p)$$

Each $R_{ii}(p)$ is of rank one and the corresponding vector is of course $(\rho_1^i(p), \ldots, \rho_{m_i}^i(p))$. Since R_{ii} is the m_i demensional identity matrix these P vectors must span m_i dimensional space, and hence for some p

$$w_i(p) = \sum_{h=1}^{m_i} \rho_h^i(p)\, v_h^i \neq 0$$

and for that p

$$\sum_{ij} q_{ij}\, w_i(p)\, w_j(p) > 0 \text{ and hence}$$

$$vBv' > 0$$

which proves that B is positive definite. Q.E.D.

4. Some Special Information Structures

Throughout this section it will be assumed that the elements of the matrix $Q(x)$ are constants, independent of the value of x.

A. Identity and Independence

Identical Information

Since the team problem is distinguished from the simplest one-person problem by the possibility of different components of action being based on different information, it is interesting to have as a basis for comparison the value of an information structure in which all decisions are based on the same information. Let $y = \eta(x)$ be this common information variable. Condition (M) immediately gives us the best decision function as

$$\alpha(y) = E(r|y)\,Q^{-1} \tag{17}$$

and the corresponding expected payoff as

$$\begin{aligned} U(\hat{\alpha},\eta) &= E\left[E(r|y)\,Q^{-1}\,E(r|y)'\right] \\ &= ||\,E(z|y)\,||^2 \end{aligned} \tag{18}$$

Since the maximum expected payoff is obtained when $E(z|y) = z$ (see remarks following equation (5)), and equals $||\,z\,||^2$, subtracting (18) from this gives us the "loss" which is suffered by using the "incomplete" information structure η:

$$\begin{aligned} L(\eta) &= ||\,z - E(z|y)\,||^2 \\ &= E\left[(z - E(z|y))\,Q\,(z - E(z|y))'\right] \\ &= E\left\{E\left[(z - E(z|y))\,Q\,(z - E(z|y))' \mid y\right]\right\} \end{aligned} \tag{19}$$

The expression $E\left[(z - E(z|y))\,Q\,(z - E(z|y)|y\right]$ might be thought of as a conditional generalized variance of z given y.[1/] Thus the loss due to incomplete information is equal to the expected value of the conditional generalized variance of z given y.

Independent Information

At the opposite extreme from identical information is independent information; by the latter we mean an η such that the y_i are statistically independent. In this case application of condition (M) gives:

$$(20) \qquad q_{jj}\,\alpha_j(y_j) + \sum_{i \neq j} q_{ij}\, E\,\alpha_i(y_i) = E\,(r_j \mid y_j) \qquad j = 1, \ldots, n$$

One can easily verify that

$$(21) \qquad \alpha_j(y_j) = \frac{1}{q_{jj}}\, E\,(r_j \mid y_j) \qquad j = 1, \ldots, n$$

is a solution of(20) and therefore is the best team decision function. The value of η is

$$(22) \qquad U(\hat{\alpha}, \eta) = E \sum_j \frac{1}{q_{jj}} \left[E\,(r_j \mid y_j) \right]^2$$

Thus the value of an information structure with independent components is the sum of the values the components would have in "one-person" games with payoff functions

$$- q_{jj}\, a_j^2 + 2 r_j\, a_j$$

Combining Identity and Independence

It is but one step further to combine these two types of information -- identical and independent -- in the following way: Suppose that the set of action variables is partitioned into groups $\left\{ a_i \right\}_{i \in M_k}$, $k = 1, \ldots, K$; and

1/ Not to be confused with Aitken's generalized variance.

suppose that for all variables in the same group, the information is identical, whereas the different information components for the different groups are independent. If we denote by w_k the information variable which is common to the k'th group, by α^k and r^k that part of α and r, respectively, corresponding to the k'th group, and by Q_k the submatrix $[q_{ij}]$, $i,j \in M_k$, then by reasoning similar to that already used above we find the best decision function to be

$$\hat{\alpha}^k = E\,(r^k|w_k)\,Q_k^{-1} \tag{23}$$

and the value of η to be

$$\begin{aligned} U\,(\hat{\alpha},\eta) &= E\sum_k E\,(r^k|w_k)\,Q_k^{-1}\,E\,(r^k|w_k)' \\ &= E\sum_k (r^k)\,Q_k^{-1}\,(r^k)' \\ &\quad - E\sum_k \left[r^k - E\,(r^k|w_k)\right] Q_k^{-1} \left[r^k - E\,(r^k|w_k)\right]' \end{aligned} \tag{24}$$

Thus partitioning the team in this way has the effect of reducing the problem to a corresponding number of simple problems, with identical information.

<u>All or Nothing Components</u>

Suppose that the decision variables are divided into two groups such that the decisions in the first group are all based upon complete information, while the decisions in the second group are based on no information at all. Thus for $i \in I$, $\eta_i(x) = x$, and for $j \notin I$, $\eta_j(x) = y_o$, a constant independent of x. This is really a special case of the structure just discussed. Let $\bar{\alpha}$ and $\bar{r}$ refer to those components of α and r for which $i \in I$, and let $\bar{Q}$ be the corresponding submatrix of Q. Then by (23) the best decision functions are:

$$(25) \qquad \begin{cases} \bar{\alpha}(x) = \bar{r}(x)\ (\bar{Q})^{-1} \\ \alpha_j = 0 \qquad\qquad , j \notin I \end{cases}$$

and the best payoff is (from (24)

$$(26) \qquad U\ (\alpha,\eta) = E\ \bar{r}\ (\bar{Q})^{-1}\ \bar{r}\ ' = \sum_{ij \in I} \rho_{ij}\ \bar{q}^{\ ij}$$

where $\rho_{ij} = Er_i r_j$ and $\bar{q}^{ij}$ are the elements of $(\bar{Q})^{-1}$.

B. Information Subject to Error

We have already noted that for any fixed x the best decision is $a = z(x) = r(x)\ Q^{-1}$. Thus the value of the information structure

$$\eta_i(x) = z_i(x)$$

is equal to (z,z), the highest possible value. Suppose now that the information $z_i(x)$ is received with some error; what will be the effect on the best decision function and on the value of the information? If we make the assumption (N) of normality (section 3) and further assume that the errors are distributed independently of each other and of the $z_i(x)$, then an explicit answer can be given to this question.

The information structure is defined by

$$y_i = \eta_i(x) = z_i(x) + e_i$$

where the e_i are independent and each e_i is independent of each $z_j(x)$.

Further, the e_i and z_i are normally distributed. We can assume without loss of generality that the e_i and z_i have zero means. Let $\sigma_{ij} = E\, z_i\, z_i$ and $\tau_i^2 = E\, e_i^2$. By Theorem 2 each decision rule must be of the form $\alpha_i(y_i) = b_i y_i$; then the team decision functions of the form

$$\beta_i = (0, \ldots, 0,\ y_i,\ 0, \ldots, 0)$$

span the space $\mathcal{A}$ of all team decision functions. From the definitions of the β_i we get the inner products

$$(\beta_i, \beta_j) = \begin{cases} q_{ij}\ \sigma_{ij} & , \text{ if } i \neq j \\ q_{ii}\ (\sigma_{ii} + \tau_i^2), & \text{if } i = j \end{cases}$$

(27)

$$(\beta_i, z) = \sum_j q_{ij}\ \sigma_{ij}$$

If we let $F = ((f_{ij}))$ be the inverse of the matrix of inner products (β_i, β_j), then by the well known projection formula, the projection of z on $\mathcal{A}$ is

$$\hat{\alpha} = \sum_j \hat{b}_j\ \beta_j$$

where
$$\hat{b}_j = \sum_i f_{ij}\ (\beta_i, z)$$

$$= \sum_i f_{ij} \sum_k q_{ik}\ \sigma_{ik}$$

By using the definition of F, a simple computation reduces the above to

(28) $$\hat{b}_j = 1 - \sum_k f_{jk}\ q_{kk}\ \tau_k^2$$

The sum on the right of (28) is the modification of the decision function

$\alpha = z$ made necessary by the errors. Furthermore, the formula for the length of the projection gives us the corresponding expected payoff:

$$U(\hat{\alpha},\eta) = \sum_{ij} f_{ij}\,(\beta_i,z)\,(\beta_j,z)$$

$$= \sum_{ij} f_{ij} \sum_h q_{ih}\,\sigma_{ih} \sum_k q_{jk}\,\sigma_{jk}$$

which after some computation reduces to

$$(29) \qquad U(\hat{\alpha},\eta) = (z,z) - \left[\sum_h q_{hh}\,\tau_h^2 - \sum_{hk} f_{hk}\,q_{hh}\,\tau_h^2\,q_{kk}\,\tau_k^2\right]$$

The term in brackets at the right of (29) is the loss due to the presence of the error terms e_i.

Unfortunately formula (29) is not very simple, and seems to admit no easy interpretation. To dramatize the effects of interaction let us evaluate (29) in the degenerate case in which $q_{ij} = 0$ if $i \neq j$. We then have

$$(29a) \qquad U(\hat{\alpha},\eta) = (z,z) - \sum_h \frac{q_{hh}}{\frac{1}{\sigma_{hh}} + \frac{1}{\tau_h^2}}$$

C. Coding

Often information about a continuously varying variable can take only a finite number of values. We shall call such an information function a code. Since most information functions can be approximated by a code, one would not expect that there is much which can be said about coding in general which cannot also be said about the general team problem. However some special coding problems have interesting properties, and we shall illustrate this by discussing a particular two person problem.

Suppose $x = (x_1, x_2)$, where x_1 and x_2 are <u>independent</u> real random variables. It will be convenient to think of x_i as the variable corresponding to person i. Consider the class of information structures of the form

$$\eta_1(x) = (x_1, s_{2j}) \text{ if } x_2 \text{ in } X_{2j} \qquad j = 1,\ldots,m_2 \tag{30}$$

$$\eta_2(x) = (s_{1k}, x_2) \text{ if } x_1 \text{ in } X_{1k} \qquad k = 1,\ldots,m_1$$

where the s_{ij} are just m_i arbitrary symbols, and X_{2j} and X_{1k} are both partitions of the real line. In other words, each person observes his own variable, but receives only coded information about the other variable.

The payoff function considered will be

$$u(a_1x) = -a_1^2 - 2q\, a_1a_2 - a_2^2 + 2\,(x_1a_1 + x_2a_2) \tag{31}$$

Notice that we have specialized the general quadratic payoff by making $r_i(x) = x_i$. [We have also taken $q_{ii} = 1$, this can be done by choosing suitable units for a_i.] Without loss of generality we can assume that $E\, x_i = 0$.

Any decision function for person i must be of the form

$$\alpha_i(y_i) = \alpha_{ij}(x_i) \quad \text{if } x_i \text{ in } X_{ij}$$

Condition (M) thus takes the form:

$$\left\{\begin{array}{l} \alpha_{1k}(x_1) + q\, E\,[\alpha_{2j}(x_2) \mid x_1 \cdot s_{2k}] = x_1 \\[2ex] \alpha_{2j}(x_2) + q\, E\,[\alpha_{1k}(x_1) \mid x_2 \cdot s_{1j}] = x_2 \end{array}\right. \tag{32}$$

Suppose the decision functions had the form

$$\text{(33)} \qquad \begin{cases} \alpha_{1k}(x_1) = x_1 + b_{1k}^{j}, & j = j(x_1) \\ \alpha_{2j}(x_2) = x_2 + b_{2j}^{k}, & k = k(x_2) \end{cases}$$

Substituting (33) in (32) would give the condition

$$\begin{cases} b_{1k}^{j} + q\, b_{2j}^{k} = -q\, E\,(x_2 | s_{2k}) & , k = 1, \ldots, m_2 \\ q\, b_{1k}^{j} + b_{2k}^{j} = -q\, E\,(x_1 | s_{1j}) & , j = 1, \ldots, m_1 \end{cases}$$

This does have the solution

$$b_{1k}^{j} = \frac{q^2 \bar{x}_{1j} - q\, \bar{x}_{2k}}{1 - q^2}$$

$$b_{2j}^{k} = \frac{q^2\, \bar{x}_{2k} - q\, \bar{x}_{1j}}{1 - q^2}$$

where $\bar{x}_{1j} = E\,(x_1 | s_{1j})$, and hence the best decision rule for a given η is

$$\text{(35)} \qquad \begin{cases} \alpha_{1k}(x_1) = \dfrac{x_1 - q^2\, [x_1 - \bar{x}_{1j}] - q\, \bar{x}_{2k}}{1 - q^2} \\ \text{and similarly for } \alpha_2 \end{cases}$$

From (35) we get the expected gross payoff:

$$(36) \qquad U(\alpha,\eta) = \frac{E[(\bar{x}_{1j})^2 + (\bar{x}_{2k})^2]}{1 - q^2}$$

$$= \frac{Ex_1^2 - Eq^2 \operatorname{Var}(x_1|s_{1j})}{1 - q^2}$$

$$+ \frac{Ex_2^2 - Eq^2 \operatorname{Var}(x_2|s_{2k})}{1 - q^2}$$

Since $\frac{E(x_1^2 + x_2^2)}{1 - q^2}$ is the best that can be achieved under full information, the loss due to incomplete information is

$$(37) \qquad \frac{q^2}{1-q^2} \; E\,[\operatorname{Var}(x_1|s_{1j}) + \operatorname{Var}(x_2|s_{2k})]$$

It is noteworthy here that the total value of the information structure is the sum of two terms, each depending on one information component only. Thus even though $q \neq 0$ and the two information variables are not independent, there is no "interaction" between the two components of the information structure.

Formula (37) tells us that we want to minimize the average variance within the sets X_{ij}. Thus if the cost of the information structure depends only upon the number of symbols m_i and the probabilities $Pr(x_i \in X_{ij})$, the sets X_{ij} must be intervals.