COWLES FOUNDATION DISCUSSION PAPER NO. 5

Team Decision Problems*

Roy Radner

October 11, 1955

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

CHAPTER I

General Discussion of the Team

1. Problem Formulation and Summary of Results.

1.1. A Review of the Statistical Decision Problem.

The statistical decision problem is concerned with an individual who chooses a decision a from a set D of possible decisions, and is thereupon rewarded according to his choice and the prevailing state of the world x. It will be assumed here that the reward is a real number $u(a,x)$ determined by a payoff function u. In a given situation the relevant state of the world may be the outcome of some random process, as in a prediction problem, or it may be a particular probability distribution of random events, as in an estimation problem, or it may be a combination of the two. Therefore let Z be a measurable space, whose class $\mathcal{Z}$ of measurable subsets is to be interpreted as the class of random events, and let P be a set of probability measures P on Z. The states of the world x will be represented by pairs (z,p), with z in Z and p in P.

Typically, the decision maker bases his decision upon certain information about the state of the world, according to some rule or decision function. This concept will be represented in this paper as follows. Assume that there are given a field $\mathcal{P}$ of subsets of P and a field $\mathcal{D}$ of subsets of D, and let $\mathcal{X}$ be the field of subsets of X that is the Cartesian product of $\mathcal{Z}$ and $\mathcal{P}$.[1] Let $\mathcal{Y}$ be a given subfield of $\mathcal{X}$;

---

[1] See Halmos [9], p. 140.

then any $\mathcal{Y}$-measurable function from X to D will be called a <u>decision function</u> based upon the <u>information subfield</u> $\mathcal{Y}$. In some problems the decision functions may, for some external reason, be restricted to some given set A.

A probably more familiar way of representing information would be in terms of a transformation, that is, a measurable function from X to some other measurable space, say Y; decision functions would then be measurable functions from Y to D. These two ways of representing information are equivalent, and in this paper one or the other will be used, according to the nature of the particular problem being considered. For a discussion of subfields, transformations, and the closely related concept of a "statistic", see Bahadur [2], and Bahadur and Lehmann [3].

For any given decision function α and any probability measure p in P, the expected payoff is

$$U(\alpha, p) \equiv \int_Z u[\alpha(z,p), (z,p)] \, d\, p(z).$$

The problem for the decision maker is to choose a decision function from the given set A that will make $U(\alpha, p)$ large in some sense. Two approaches, commonly known as <u>Bayes</u> and <u>minimax</u>, will be considered in this paper, with emphasis on the former.

If G is a given <u>a priori</u> probability measure on P, then the <u>Bayes payoff</u> for a decision function α is defined as

$$V(\alpha, G) \equiv \int_P U(\alpha, p) \, d\, G(p),$$

and $\hat{\alpha}$ is a Bayes decision function if it maximizes the Bayes payoff on the set A. Note that as soon as the Bayes approach is adopted, the distinction between Z and P ceases to have any real significance, for once a particular a priori measure G is given, the whole probability structure of the problem can be summarized in terms of a single measure on X.

For any decision function α, and any p in P, the risk is defined as

$$\rho(\alpha, p) = \sup_{\alpha' \in A} U(\alpha', p) - U(\alpha, p).$$

A decision function $\hat{\alpha}$ is minimax in A if for all α in A,

$$\sup_{p \in P} \rho(\hat{\alpha}, p,) \leq \sup_{p \in P} \rho(\alpha, p).$$

1.2. The Team.

Suppose now that the decision variable a actually is an N-tuple $(a_1, \ldots, a_N)$ of decision variables a_i, with a_i in D_i. Suppose further that each of the component decisions is based upon different information; that is, there are N subfields $\mathcal{Y}_i$ of $\mathcal{X}$, and the decision function α is an N-tuple of functions $\{\alpha_i\}$ such that each function α_i is a $\mathcal{Y}_i$-measurable function from X to D_i.

For example, the "decision maker" may be a group of N persons, each with access to different information (because of difficulty of communication, say), each deciding about something different, but receiving a common payoff as a result of their joint decision. As a second example, the decision maker might be an individual making different decisions in successive time periods, the payoff being a function of all the decisions made over the total time period. In such a case, if the decision maker does not "forget" anything from one time period to the next, then $\mathcal{Y}_1 \subset \mathcal{Y}_2 \subset, \ldots, \subset \mathcal{Y}_N$. However, the keeping of records might be costly, so that it might be worthwile to forget something.

Although from the most general point of view the decision problem just outlined is a one-person problem, it is sometimes suggestive to talk about it in terms of the first, "many-person", example above, and this will be done in this paper. J. Marschak has called such a decision maker a team,[1/] to emphasize that there are no conflicts of interest between the members of the group. It should also be pointed out that differences of opinion (as embodied in different a priori distributions on P, for example) cannot be handled in this context, since these result formally in the same N-person game-theoretic difficulties as do conflicts of interest.

In this paper it will be assumed that for each i the set D_i of possible decisions for the i^{th} team member is a Borel measurable subset of the real line. This assumption is not quite as special as it might at first seem, for a problem in which some D_i is a Borel measurable subset of M-dimensional Euclidean space could be recast in the present framework by replacing that person by M persons, all with the same information, each

1/ See Marschak [11].

with a one-dimensional decision variable, and with their decision functions possibly subject to some joint constraint. Admittedly, although this device achieves a certain technical generality for the present framework, in practice it might sometimes lead to unnecessary complexity and awkwardness. However, all of the results in this paper can be reinterpreted in terms of vector decision variables.

1.3. Summary of Results.

The remainder of the first chapter is devoted to some of the things that can be said about the Bayes problem for the team at the level of generality of the preceding problem formulation, and consists of remarks on the value of information for different team members, on sufficient subfields for team members, and on conditions under which "person-by-person" maximization leads to a true Bayes decision function.

Chapter II explores the consequences of assuming that the payoff function **u** is quadratic in the decision variables for a.e. x. The geometry of Hilbert space is helpful in investigating the existence and uniqueness of Bayes decision functions, which can be interpreted as projections.

If, in the quadratic payoff function, the coefficients of the quadratic terms are independent of x, then the situation is even more amenable to analysis. Chapter III proceeds on this assumption, and is devoted mainly to an exploration of the role of linear decision functions in this setup. It is shown that if the a priori distribution induces a normal distribution of all the information variables and of the coefficients in the linear terms of the payoff function, then the Bayes team decision function is linear in the information variables. A team analogue of the Markoff problem is solved and the minimax properties of such solutions are investigated.

For a discussion of the case of a team with a linear payoff function and with linear constraints on the decision variables, see [12].

1.4. Acknowledgements and Historical Remarks.

The origin of the problem considered in the present paper is Marschak's work on the theory of organization (see Marschak [11]). Marschak's approach is in the spirit of the theory of games and of decision theory. This paper grew out of an attempt to analyze some of the many-person aspects of organizations that are present even in the absence of many-person game complications (i.e., conflicts of interest and differences of opinion). Initially this attempt took the form of the study of some simplified examples of organizations (see references [18]-[23]).

From discussions with Marschak and with Savage and from a study of Savage's recent book [13] I derived both encouragement to undertake and continue this study, and help in arriving at the present formulation and in working out specific problems.

2. The Value of Information in the Bayesian Problem.

2.1. Interaction Between Information Subfields.

Consider a team that has adopted the Bayes approach, with a given a priori distribution G and a given N-tuple $\mathcal{Y} = (\mathcal{Y}_1, \ldots, \mathcal{Y}_N)$ of information subfields, and suppose that the supremum

$$\sup_{\alpha} V(\alpha, G)$$

of the Bayes payoff is finite. This supremum actually depends upon $\mathcal{Y}$, and might be denoted by $v(\mathcal{Y})$. A change in $\mathcal{Y}$, say from $\mathcal{Y}$ to $\mathcal{Y}'$, would be accompanied by the change

$$v(\mathcal{Y}') - v(\mathcal{Y})$$

in the maximum Bayes payoff, which might be called the <u>value of the change in</u> $\mathcal{Y}$. (This interpretation of the above difference, of course, makes good sense only if it is appropriate to think of the expected cost of information as something to be subtracted from the expected payoff. (See Savage [13], p. 118.) Such an assumption will be made implicitly here, thus making it sensible to discuss the "value of a change of information" without explicitly considering costs.)

Correspondingly, define the <u>value of the change of a single</u> $\mathcal{Y}_i$ <u>to</u> $\mathcal{Y}_i'$ as

$$\Delta v_i(\mathcal{Y}_i', \mathcal{Y}) = v(\mathcal{Y}_1, \ldots, \mathcal{Y}_i', \ldots, \mathcal{Y}_N) - v(\mathcal{Y}).$$

As the notation suggests, this value will in general depend both upon $\mathcal{Y}_i'$

and upon all of $\mathcal{Y}$. In what circumstances will it depend only upon $\mathcal{Y}_i'$ and $\mathcal{Y}_i$? To be precise, for every i let $\mathcal{S}_i$ be a family of subfields of $\mathcal{X}$, and let $\mathcal{S}$ be the set of all $\mathcal{Y}$ such that $\mathcal{Y}_i \in \mathcal{S}_i$ for all i. If for every $\mathcal{Y} \in \mathcal{S}$, every i, and every $\mathcal{Y}_i'$ in $\mathcal{S}_i$,

$$\Delta_i \, v(\mathcal{Y}_i', \mathcal{Y})$$

depends only on $\mathcal{Y}_i'$ and $\mathcal{Y}_i$, then there will be said to be no <u>interaction between subfields, in</u> $\mathcal{S}$. In such a case it is clear that there exist functions $v_1, \ldots, v_N$ such that for all $\mathcal{Y}$ in $\mathcal{S}$

$$v(\mathcal{Y}) = \sum_i v_i(\mathcal{Y}_i) .$$

In particular, this is true if the payoff function u has the form:

(1) $$u(a,x) = \sum_i u_i(a_i,x) \quad \text{a.e.},$$

for some functions $u_1, \ldots, u_N$. A team with a payoff function of this form is, in a sense, degenerate, for the Bayes decision function $\hat{\alpha}$ is characterized by the fact that, for each i, $\hat{\alpha}_i$ maximizes the expectation $E\, u_i(\alpha_i(x),x)$, subject to α_i measurable-$\mathcal{Y}_i$.

This is the best that can be said without making any specific assumptions about u and the a priori distribution G; that is, given any (non-trivial) $\mathcal{S}$, there exist u and G such that there is interaction between subfields, in $\mathcal{S}$. Two special examples of no interaction are given below in Chapter III, Section 6.

It should be noted that if the minimax approach is used, the value of information could also be defined, in terms of the minimax risk corresponding to each $\mathcal{Y}$, but it is not true, however, that equation (1) implies lack of interaction in the minimax situation. Counterexamples can easily be constructed.

2.2. Sufficient Subfields for Team Members.

Suppose that the payoff function $u(a,x)$ is, for every fixed a, a function of p alone, and that every information subfield $\mathcal{Y}_i$ is contained in $\mathcal{Z}$.[1]

In the one-person problem, if the information available to the decision maker is represented by the subfield $\mathcal{Y}$, then a subfield $\mathcal{W} \subset \mathcal{Y}$ is said to be sufficient for P relative to $\mathcal{Y}$ if, for every $Y \in \mathcal{Y}$, the conditional probability,[2]

$$P(Y|\mathcal{W}) ,$$

can be chosen so as to be independent of p in P. If $\mathcal{W}$ is sufficient, then the use of any other subfield $\mathcal{V} \subset \mathcal{Y}$ cannot increase the expected payoff, no matter what the a priori distribution G or the payoff function u may be (as long as u satisfies the assumption made at the beginning

---

[1] Strictly speaking, "contained in the field of all sets $A \times P$ such that $A \in \mathcal{Z}$".

[2] For a discussion of conditional probability and expectation defined in terms of subfields see Doob [5], Chapter 1, Section 7-10.

of this section). (See Savage [13], Chapter 7, Section 4, and Bahadur [2].)

The concept of sufficiency, suitably modified, can also be made to apply to the team, as follows. If $\mathcal{A}$ and $\mathcal{B}$ are subfields, let $F(\mathcal{A}, \mathcal{B})$ denote the smallest subfield containing both $\mathcal{A}$ and $\mathcal{B}$. A subfield $\mathcal{W}_i \subset \mathcal{Y}_i$ will be called <u>sufficient</u> for P, relative to $\mathcal{Y}_1, \ldots, \mathcal{Y}_N$ if for every $Y_i \in \mathcal{Y}_i$, the conditional probability

$$P[Y_i \mid F(\mathcal{W}_i, \{\mathcal{Y}_j\} \quad j \neq i)]$$

can be chosen so as to be independent of p in P, and measurable-$\mathcal{W}_i$.

The reader can easily construct examples in which $\mathcal{W}_i$ is sufficient relative to $\mathcal{Y}_i$, in the one-person sense, but is not sufficient relative to $\mathcal{Y}_1, \ldots, \mathcal{Y}_N$ in the team sense.

If, however, $\mathcal{Y}_1, \ldots, \mathcal{Y}_N$ are statistically independent, then it is easy to see that the (one-person) sufficiency of $\mathcal{W}_i$ relative to $\mathcal{Y}_i$ implies the (team) sufficiency of $\mathcal{W}_i$ relative to $\mathcal{Y}_1, \ldots, \mathcal{Y}_N$. That the independence of $\mathcal{Y}_1, \ldots, \mathcal{Y}_N$ is not necessary for this implication to hold is shown by the following example.

Let $\mathcal{A}$ and $\mathcal{C}$ be independent subfields of $\mathcal{Y}$; let $\mathcal{B} \subset \mathcal{A}$ and $\mathcal{D} \subset \mathcal{C}$; and consider a 2-person team with:

$$\mathcal{Y}_1 = F(\mathcal{A}, \mathcal{D})$$

$$\mathcal{W}_1 = F(\mathcal{B}, \mathcal{D})$$

$$\mathcal{Y}_2 = \mathcal{C} .$$

It follows that $F(\mathcal{W}_1, \mathcal{Y}_2) = F(\mathcal{B}, \mathcal{C})$. If f is the characteristic function of a set in $\mathcal{Y}_1$, then according to the lemma of Chapter II, Section 3.2 below

$$E[f \mid F(\mathcal{W}_1, \mathcal{Y}_2)] = E[f \mid \mathcal{W}_1] .$$

Hence if $\mathcal{W}_1$ is sufficient relative to $\mathcal{Y}_1$, it is sufficient relative to $\mathcal{Y}_1$ and $\mathcal{Y}_2$.

It should be pointed out that the above concept of sufficiency could easily have been made to cover problems in which the payoff depends upon z, but this would have complicated the notation. On the other hand a whole-hearted Bayesian approach would have simplified the situation, as there would be no distinction between Z and P. (See Savage [13], Chapter VII, Section 4.)

3. Person-by-Person Maximization and Stationarity in the Bayes Problem.

If $\hat{\alpha}$ is a Bayes team decision function relative to a given a priori distribution, then the decision function $\hat{\alpha}_i$ for any one team member i must be best, given that every other member j uses the decision function $\hat{\alpha}_j$. Call a decision function α person-by-person maximal if α cannot be improved by changing α_i for any one i alone. Thus any Bayes decision function is person-by-person maximal. The converse is not true, however, as the following example shows.

Consider a team of two members, whose payoff function is independent of x, with contour lines as in the accompanying figure (for example,

$$u(a_1,a_2) = \min \left\{-a_1^2 - (a_2 - 1)^2, -(a_1 - 1)^2 - a_2^2 \right\} ,$$

where each decision variable a_i is restricted to the unit interval). It is easily verified that any a for which $a_1 = a_2$ is person-by-person maximal, (e.g., point P in the figure) whereas the maximum of u is attained only at $a_1 = a_2 = 1/2$. Note that u may be strictly concave in a.

Suppose the decision functions of all but one of the team members are fixed; then the problem facing that one member becomes a one-person Bayesian problem if he thinks of the actions of the other members as part of the "state of the world"; and he can therefore apply Bayes' Rule. More precisely, let $\hat{\alpha}$ be person-by-person maximal, and suppose that for each i, conditional expectations given $\mathcal{Y}_i$ are bona fide expectations; then for every i and a.e. x, $\hat{\alpha}_i(x)$ maximizes, with respect to a_i, the conditional expectation

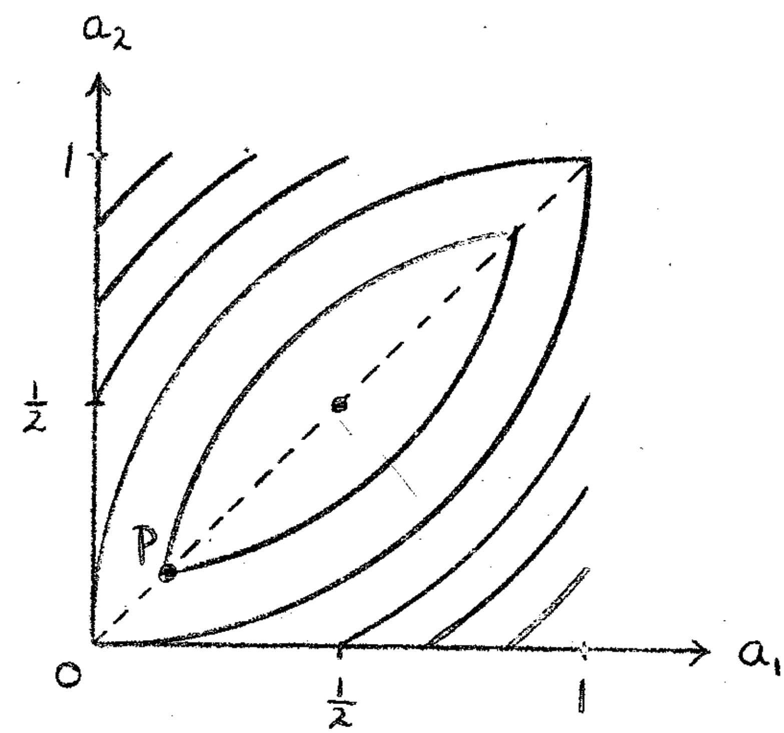

Fig. 1

$$\Phi_i(a_i,x) = E\{u[\hat{\alpha}_1(x),\dots,a_i,\dots,\hat{\alpha}_N(x),x] \mid \mathcal{Y}_i\} \quad .$$

These N simultaneous conditions constitute the team analogue of the one person Bayes rule, but as was just shown, they are not in general sufficient to determine the Bayes decision function.

If, in addition, $\Phi_i(a_i,x)$ is differentiable in a_i, then $\frac{\partial}{\partial a_i}\Phi_i(\hat{\alpha}_i(x),x) = 0$. Call a decision function α stationary if, with Φ_i defined as above,

(2) $$\frac{\partial}{\partial a_i}\Phi_i(\alpha_i(x),x) = 0 \quad \text{a.e.},$$

for every i. Thus, under suitable regularity conditions, a Bayes decision function is stationary. The following theorem gives one condition under which a stationary decision function is Bayes. This theorem will be applied in Chapters II and III.

The Bayes payoff functional $V(\alpha) = V(\alpha,G)$ will be said to be <u>locally finite</u> at α if

1. $|V(\alpha)| < \infty$,

2. *for any decision function* δ *such that* $|V(\alpha+\delta)| < \infty$, *there* exist $k_1,\dots,k_N$ all positive such that $|V(\alpha_1+h_1\delta_1,\dots,\alpha_N+h_N\delta_N)| < \infty$ for all $h_1,\dots,h_N$ for which $|h_1| \le k_1,\dots,|h_N| \le k_N$.

In what follows it is assumed that the set A from which a decision function is to be chosen is actually the set of all $\alpha = (\alpha_1,\dots,\alpha_N)$ such that α_i is $\mathcal{Y}_i$-measurable.

Theorem I.1. If

1. $u(a,x)$ is concave and differentiable in a for a.e. x,

2. $\sup_{\beta} V(\beta) < \infty$,

3. V is locally finite at α,

4. α is stationary,

then α is Bayes.

Lemma. If $f(c,x)$ is a concave function of the real variable c on the closed interval $[c', c'']$, for a.e. x, and $|Ef(c,x)| < \infty$ on $[c', c'']$, then

$$\frac{d^+}{dc} Ef(c,x) \Big|_{c = c'} = E \frac{\partial^+}{\partial c} f(c,x) \Big|_{c = c'}$$

$$\frac{d^-}{dc} Ef(c,x) \Big|_{c = c''} = E \frac{\partial^-}{\partial c} f(c,x) \Big|_{c = c''} .$$

Proof: For $c > c'$, the concavity of f implies that

$$\frac{f(c,x) - f(c',x)}{c - c'} - \frac{f(c'',x) - f(c',x)}{c'' - c'} \geq 0 \quad \text{a.e.},$$

and increases monotonically as c approaches c'. Therefore, the first part of the lemma follows from the Lebesgue monotone convergence theorem (Halmo [9], Section 27, Theorem A); the second part follows by symmetry.

Proof of the theorem:

Suppose that $V(\alpha + \delta) > -\infty$. Let

$$f(k,x) = u[\alpha_1(x) + k_1\delta_1(x), \ldots, \alpha_N(x) + k_N\delta_N(x), x] ,$$

and $F(k) \equiv Ef(k,x)$. By the assumption of regularity, F is finite in some neighborhood of 0, and it is easy to verify that F is concave. Hence, for every i, $\frac{\partial^+}{\partial k_i} F(0)$ and $\frac{\partial^-}{\partial k_i} F(0)$ are finite. Hence, by the lemma, for every i,

$$\frac{\partial^{\pm}}{\partial k_i} F(0) = E \frac{\partial^{\pm}}{\partial k_i} f(0,x) .$$

But f is differentiable in k a.e., and hence

$$\frac{\partial}{\partial k_i} F(0) = E \frac{\partial}{\partial k_i} f(0,x)$$

$$= E\ \delta_i(x) \frac{\partial}{\partial a_i} \mu[\alpha(x),x] ,$$

which last expectation is therefore finite.

On the other hand, by the lemma, applied to the condition of stationarity,

$$E\left\{ \frac{\partial}{\partial a_i} u[\alpha(x),x] \,\middle|\, \mathcal{Y}_i \right\} = 0 ;$$

hence $E \frac{\partial}{\partial a_i} u[\alpha(x),x]$ is finite, and therefore

$$E\ \delta_i(x) \frac{\partial}{\partial a_i} u[\alpha(x),x] = E[E \left\{ \delta_i(x) \frac{\partial}{\partial a_i} u[\alpha(x),x] \,\middle|\, \mathcal{Y}_i \right\}]$$

$$= E[\delta_i(x) E \left\{ \frac{\partial}{\partial a_i} u(\alpha(x),x) \quad \mathcal{Y}_i \right\}]$$

$$= 0 .$$

(See Doob [5], Chapter I, Theorem 8.3.) Hence, for every i, $\frac{\partial}{\partial k_i} F(0) = 0$, and therefore

$$\frac{d}{dt} V(\alpha + t\delta)\Big|_{t=0} = \sum_i \frac{\partial}{\partial k_i} F(0) = 0$$

(see Bonnesen and Fenchel [17], Section 13). Since $V(\alpha + t\delta)$ is a concave function of t, the theorem follows immediately.

CHAPTER II

The Team with a Quadratic Payoff Function

1. Introduction.

This chapter will explore the consequences of assuming that for every state of the world x, the payoff is a quadratic function of the team decision. Thus:

$$(1) \qquad u(a,x) = \lambda(x) + \mu(x)\, a' - aQ(x)a' \, ,$$

where a is in N-dimensional Cartesian space R^N, and for every x, $\lambda(x)$ is in R^1, $\mu(x)$ is in R^N and $Q(x)$ is an $N \times N$ symmetric matrix (λ, μ and Q all measurable). I want to consider only the situation in which, for every a.e. x, u(a,x) has a unique maximum in a; it will therefore be assumed that Q(x) is positive definite for every a.e. x.

It will be more convenient to speak in terms of loss (in the technical sense) rather than payoff. Completing the square, (1) can be rewritten as

$$u(a,x) = -[a - \tfrac{1}{2}\mu(x)\, Q^{-1}(x)]\, Q(x)\, [a - \tfrac{1}{2}\mu(x)\, Q^{-1}(x)]'$$

$$+ \lambda(x) + \tfrac{1}{2}\mu(x)\, Q^{-1}(x)\, \mu(x)' \, .$$

The best team decision for any given x is clearly

$$\gamma(x) = \tfrac{1}{2}\mu(x)\, Q^{-1}(x) \, ,$$

and the loss due to using any other decision a is therefore

$$[a - \gamma(x)]\, Q(x)\, [a - \gamma(x)]' \, .$$

The risk, or expected loss, given the team decision function α and the state of nature p, is

$$(2) \qquad \rho(\alpha, p) = E\{[\alpha(x) - \gamma(x)]\, Q(x)\, [\alpha(x) - \gamma(x)]' \mid p\} \quad .$$

For any a priori distribution G, the Bayes risk, $E\,\rho$, will be denoted by $\sigma(\alpha, G)$, or sometimes just by $\sigma(\alpha)$.

The risk function can be transformed to a certain extent without altering the problem. Let $T(x)$ be a measurable, $N \times N$-matrix valued function of x that is non-singular a.e., and such that for any decision function α, both $[\alpha(x)\, T(x)]_j$ and $[\alpha(x)\, T^{-1}(x)]_j$ are measurable-$\mathcal{Y}_j$; then the function β defined by

$$\beta(x) = \alpha(x)\, T(x)$$

is a decision function if and only if α is, and the risk function

$$E\{[\beta(x) - \gamma(x)\, T(x)]\, T^{-1}(x)\, Q(x)\, (T^{-1}(x))'\, [\beta(x) - \gamma(x)\, T(x)]'$$

defines a problem equivalent to that defined by (2).

For example, let the team members be divided into subgroups $I_1, \ldots, I_k$, and let $T(x)$ be a matrix with blocks $T_k(x)$ down the diagonal and zeros elsewhere, where $T_k(x)$ is a non-singular matrix of order equal to the size of group I_k, and measureable- $\bigcap_{k \in I_i} \mathcal{Y}_k$. In particular, $T(x)$ can be taken to be any diagonal matrix such that, for every i, $t_{ii}(x)$ is measurable- $\mathcal{Y}_i$; typically, this will be the only transformation that can be made.

2. Bayes Decisions as Projections - Their Existence and Uniqueness.

A glance at equation (3) above shows that the team problem looks something like a prediction or estimation problem, with a quadratic form replacing the one-dimensional mean squared error. It is not surprising, therefore, that the Bayes solution can be described in terms of a projection in a suitable Hilbert space. Theorem 1 below describes the set of team desision functions that have finite Bayes risk, and shows, incidentally, that it is no real restriction to assume that the coordinates of γ have finite mean square. Theorem 2 proves that if any decision function has finite Bayes risk, then the best decision exists, is unique, and can be characterized as a projection.

In what follows it is understood that there is a given a priori probability measure on P, and that all expectations, probabilities, and references to "almost everywhere" are on the basis of it. Any two functions on X that are equal a.e. will be considered equivalent, so that hereafter whenever any space of functions on X is introduced, it is to be understood that, strictly speaking, the object under discussion is really the corresponding quotient space modulo the space of functions that are zero a.e.

Lemma 1. Let H be the set of all measurable functions α from X to R^N for which

$$E\,\alpha(x)\;Q(x)\,\alpha(x)' < \infty;$$

then, under the inner product

$$(\alpha, \beta) = E\,\alpha(x)\;Q(x)\,\beta(x)' ,$$

H is a (non-trivial) Hilbert space.

Proof. The only things not immediately obvious are the completeness of H and its non-triviality. Let K be the space of all measurable β from X to R^N for which

$$\|\beta\|_K^2 \equiv E\,\beta(x)\,\beta(x)' < \infty .$$

For any sequence $\{\beta^{(n)}\}$ in K, $\|\beta^{(n)}\|_K \to 0$ implies that $E(\beta_i^{(n)})^2 \to 0$ for $i = 1,\ldots,N$; hence the Riesz-Fisher Theorem 1/ applies coordinate-wise, and K is therefore complete.

Since Q(x) is positive definite a.e., $Q(x) = S(x)S'(x)$, where S(x), a square root of Q(x), is non-singular a.e. S induces a linear transformation $\mathcal{S}$ from H to K by

$$(\mathcal{S}\alpha)(x) = \alpha(x)S(x) .$$

Furthermore,

$$\|\mathcal{S}\alpha\|_K = \|\alpha\|_H$$

and

$$\|\mathcal{S}^{-1}\beta\|_H = \|\beta\|_K ,$$

so that $\mathcal{S}$ is an isometry, and therefore H is also complete. Furthermore, K has non-zero elements (e.g., the bounded functions), and hence so does H.

1/ For a proof of this theorem, see the remarks at the end of Section 42 of Halmos [9], who, however, does not identify it with this name; or see Stone [14], p. 25.

<u>Lemma 2</u>. If the sequence $\{\alpha^{(n)}\}$ converges to α in the norm of H, then there is a subsequence that converges to α coordinatewise and pointwise a.e.

<u>Proof</u>. Let $\beta = \mathcal{S}\alpha$ and $\beta^{(n)} = \mathcal{S}\alpha^{(n)}$. As in Lemma 1, $\|\alpha^{(n)} - \alpha\|_H = \|\beta^{(n)} - \beta\|_K \longrightarrow 0$. By theorems 25.A, 22.D, and 21.B of Halmos [9], applied to the first coordinates of $\beta^{(n)}$ and β, there exists a subsequence $\{\beta^{(m)}\}$ such that $\beta_1^{(m)}(x) \longrightarrow \beta_1(x)$ a.e. Again, there exists a subsequence $\{\beta^{(k)}\}$ of $\{\beta^{(m)}\}$ such that $\beta_2^{(k)}(x) \longrightarrow \beta_2(x)$ a.e., etc. Thus there exists a subsequence $\{\beta^{(j)}\}$ of the original sequence such that $\beta_i^{(j)}(x) \longrightarrow \beta_i(x)$ a.e. for $i = 1,\ldots,N$; and hence

$$\alpha_i^{(j)}(x) = [\beta^{(j)}(x)\overset{-1}{S(x)}]_i \longrightarrow [\beta(x)\overset{-1}{S(x)}]_i = \alpha_i(x) \text{ a.e. } (i=1,\ldots,N),$$

which completes the proof.

Let A be the set of all measurable α from X to R^N such that α_i is $\mathcal{Y}_i$-measurable.

<u>Theorem II.1</u>. For any measurable γ from X to R^N, the set F of α in A for which

$$E[\alpha(x) - \gamma(x)]\, Q(x)\, [\alpha(x) - \gamma(x)]' < \infty$$

is either empty or it is the closed linear subvariety

$$A \cap (\gamma + H)$$

of the complete linear variety $(\gamma + H)$ under the distance function

$$d(\alpha, \beta) = \|(\alpha - \gamma) - (\beta - \gamma)\|_H .$$

Proof. Suppose that F is not empty. It follows from Lemma 1 that $F = A \cap (\gamma + H)$ and that $(\gamma + H)$ is complete under the given distance function; it remains to show that F is closed. Let α^o be any element of F; then $(\gamma + H) = (\alpha^o + H)$. The transformation that takes any β in $(\gamma + H)$ into $(\beta - \alpha^o)$ in H is an isometry from $(\gamma + H)$ onto H, and the image of F under this transformation is $A \cap H$. Therefore, F is closed if and only if $A \cap H$ is closed. Suppose that the sequence $\{\alpha^{(n)}\}$ in $A \cap H$ converges, in the norm of H, to α, which is therefore in H. By Lemma 2 there is a subsequence $\{\alpha^{(k)}\}$ converging to α coordinatewise and pointwise a.e.; hence each α_i is $\mathcal{Y}_i$-measurable, which puts α in A.

The Bayes problem is one of finding an element of F that is closest to γ, in the sense of the distance d. Because of the isometry between $(\gamma + H)$ and H, this problem is equivalent to the problem of finding an element of $A \cap H$ that is closest to $\gamma - \alpha^o$ in the sense of the norm of H (where α^o is as in the proof of Theorem 1). Therefore, if F is not empty, <u>it can be assumed without loss of generality that γ is in</u> H, and this will be done from now on. It then follows, of course, that $F = A \cap H$.

From this point on, the subscript H will be omitted from the symbol of the norm of a function, it being understood that the norm is in H.

As an example of a case in which F is empty, consider a problem such that A contains only constant functions, but the elements of $Q(x)$ do not have finite mean and γ is in H.

<u>Theorem II.2</u>. If F is not empty, then there is a unique team decision function $\hat{\alpha}$ that minimizes the Bayes risk $\sigma(\alpha) = \|\alpha - \gamma\|$ on F, and $\hat{\alpha}$ is the orthogonal projection of γ onto F.

<u>Proof</u>. Immediate from Theorem 1 and the minimizing property of the orthogonal projection (see Halmos [8], Theorems II.1 and II.2).

3. Stationarity.

In this section it is shown that, under a certain condition on the random matrix $Q(x)$, the hypothesis of Theorem I.1 is satisfied, and therefore that a stationary decision function is Bayes; also an example of the application of the stationarity condition is given.

3.1. The Condition of Stationarity for the Quadratic Case.

Theorem II.3. Let $r(x)$ be the smallest characteristic root of $Q(x)$ with respect to the quadratic form $\sum_i q_{ii}(x)a_i^2$, and let $r = \text{ess}\inf_x r(x)$. If $r > 0$, and if α is stationary, then α is Bayes.

Proof. The present theorem will be proved if it can be shown that the hypothesis of Theorem I.1 is satisfied.[1] The only point not immediately obvious is that the risk ρ is locally finite at α, which point is covered by the following:

Lemma. If $r > 0$, then ρ is locally finite at every α such that $\sigma(\alpha) < \infty$.

Proof. As shown in Section II.2, there is no loss of generality in assuming that $\|\gamma\|^2 \equiv E\gamma Q\gamma' < \infty$, and hence that $\sigma(\alpha) < \infty$ if and only if $\|\alpha\| < \infty$. Suppose, then, that $\|\alpha\| < \infty$ and $\|\alpha + \delta\| < \infty$; it follows by Theorem II.1, that $\|\delta\| < \infty$. For a.e. x,

[1] In interpreting Theorem I.1 the reader should keep in mind that he is now concerned with risk.

$$r(x) \sum_i q_{ii}(x)\ \delta_i^2(x) \leq \delta(x)\ Q(x)\ \delta(x)' .$$

Hence

$$r\ E \sum_i q_{ii}(x)\, \delta_i^2(x) \leq \|\delta\|^2 ,$$

$$E \sum_i q_{ii}(x)\, \delta_i^2(x) \leq \frac{1}{r} \|\delta\|^2 ,$$

and thus for every i,

$$E\ q_{ii}(x)\, \delta_i^2(x) < \infty .$$

Let $\delta^{(i)} = (0,\ldots,0, \delta_i, 0,\ldots,0)$; then for every i, $\|\delta^{(i)}\| < \infty$, and hence by Theorem II.1,

$$\|\alpha + \sum_i k_i\ \delta^{(i)}\| < \infty ,$$

for all real $k_1,\ldots,k_N$, which proves the lemma.

It might be noted that in the quadratic case the condition for stationarity becomes (see equation I.2):

$$E\left\{\sum_j q_{ij}(x)\ [\alpha_j(x) - \gamma_j(x)]\ \middle|\ \mathcal{Y}_i\right\} = 0$$

for every i and a.e. x.

3.2. Example: Sharing Independent Data

Consider a situation in which the team members make mutually independent observations, and then each member communicates some part of his observation to every other member (the same to all). Formally, let $\mathcal{U}_1, \ldots, \mathcal{U}_N$ be N statistically independent subfields of $\mathcal{X}$; for each i let $\mathcal{V}_i$ be a subfield of $\mathcal{U}_i$; let $\mathcal{V} \equiv F(\mathcal{V}_1, \ldots, \mathcal{V}_N)$ be the smallest subfield containing $\mathcal{V}_1, \ldots, \mathcal{V}_N$; and, for each i, let $\mathcal{Y}_i$ be the smallest subfield containing both $\mathcal{U}_i$ and $\mathcal{V}$.

Suppose further that the matrix Q is constant (independent of x). Let $\delta(x) \equiv \gamma(x)Q$ then the stationarity condition can be written:

$$\sum_j q_{ij} \, E[\alpha_j | \mathcal{Y}_i] = E[\delta_i | \mathcal{Y}_i], \tag{3}$$

for every i. This last set of equations will now be used to obtain the Bayes decision function, with the help of the following:

Lemma. Let $\mathcal{A}$, $\mathcal{C}$ and $\mathcal{E}$ be statistically independent subfields; let $\mathcal{B} \subset \mathcal{A}$, $\mathcal{D} \subset \mathcal{C}$; and let f be measurable $-F(\mathcal{A}, \mathcal{D}, \mathcal{E})$; then

$$E\{f | F(\mathcal{B}, \mathcal{C}, \mathcal{E})\} = E\{f | F(\mathcal{B}, \mathcal{D}, \mathcal{E})\}.$$

Proof. It is easy to show that there exists a sequence $\{f_n\}$ of functions such that, for every $S \in F(\mathcal{A}, \mathcal{D}, \mathcal{E})$, $\int_S f = \lim_n \int_S f_n$, and such that every f_n is simple on the ring $R(\mathcal{A}, \mathcal{D}, \mathcal{E})$ of sets generated by $\mathcal{A}$, $\mathcal{D}$ and $\mathcal{E}$ (use Halmos [9], Section 13, Theorem D). It follows (ibid., Section 32, Example 5) that it is sufficient to prove the lemma for all functions f that are simple on the ring $R(\mathcal{A}, \mathcal{D}, \mathcal{E})$. Hence it is sufficient to prove the lemma for all functions $f = ade$, where a, d and e are characteristic functions of sets in $\mathcal{A}$, $\mathcal{D}$ and $\mathcal{E}$, respectively.

Since $E\{f \mid F(\mathcal{B}, \mathcal{D}, \mathcal{E})\}$ is measurable - $F(\mathcal{B}, \mathcal{C}, \mathcal{E})$, it is sufficient to prove

$$\int_S E\{f \mid F(\mathcal{B}, \mathcal{D}, \mathcal{E})\} = \int_S f \tag{4}$$

for all $S \in F(\mathcal{B}, \mathcal{C}, \mathcal{E})$. Since the indefinite integral is a totally finite signed measure, it is sufficient to prove that (4) holds for all S such that $S = C \cap B \cap E$, for some $C \in \mathcal{C}$, $B \in \mathcal{B}$ and $E \in \mathcal{E}$ (see ibid., Section 13, Theorem A and Section 29, Theorem A). Equivalently, it is sufficient to prove that for all c, b and e' that are characteristic functions of sets in $\mathcal{C}$, $\mathcal{B}$, and $\mathcal{E}$, respectively,

$$E[bce' \, E\{f \mid F(\mathcal{B}, \mathcal{D}, \mathcal{E})\}] = E\, bce' f \,. \tag{5}$$

If $f = ade$ as above, then

$$E\{f \mid F(\mathcal{B}, \mathcal{D}, \mathcal{E})\} = de\, E\{a \mid \mathcal{B}\} \,,$$

because for all characteristic functions b' measurable - $\mathcal{B}$, d' measurable $\mathcal{D}$ and e' measurable $\mathcal{E}$,

$$E\{b'd'e'de\, E[a \mid \mathcal{B}]\} = E\{d'de'e\, E[b'a \mid \mathcal{B}]\}$$

$$= E\{d'de'e\}\, E\{E[b'a \mid \mathcal{B}]\} \qquad \text{(independence)}$$

$$= E\{d'de'e\}\, E\{b'a\}$$

$$= E\{d'de'e\, b'a\}. \qquad \text{(independence).}$$

Hence, proceeding to verify equation (5):

$$E[bce'\, E\{f \mid F(\mathcal{B}, \mathcal{D}, \mathcal{E})\}]$$

$$= E[bce'\, de\, E\{a \mid \mathcal{B}\}]$$

$$= E[cdee'\, E\{ab \mid \mathcal{B}\}]$$

$$= E[cdee']\, E[ab]$$

$$= E\, cdee'\, ab$$

$$= E(bce')(ade),$$

which completes the proof of the lemma.

Returning to equation (3), and recalling that $E(q_i' \mid \mathcal{Y}_i) = q_i$, one gets

$$(6) \qquad q_{ii}\, q_i = E(\delta_i \mid \mathcal{Y}_i) - \sum_{j \neq i} q_{ij}\, E(\alpha_j \mid \mathcal{Y}_i).$$

By the lemma just proved,

(7) $$E(\alpha_j | \mathcal{Y}_i) = E(\alpha_j | \mathcal{V}), \quad \text{for } j \neq i .$$

Taking the conditional expectation of (3) given $\mathcal{V}$:

$$\sum_j q_{ij} E(\alpha_j | \mathcal{V}) = E(\delta_i | \mathcal{V}) .$$

Since Q is non-singular, it follows that

(8) $$E(\alpha_j | \mathcal{V}) = \sum_k q^{kj} E(\delta_k | \mathcal{V}) ,$$

where $((q^{ij})) = Q^{-1}$. Substituting (7) and (8) into (6) gives the Bayes decision function

$$\hat{\alpha}_i = \frac{1}{q_{ii}} [E(\delta_i | \mathcal{Y}_i) - \sum_{j \neq i} q_{ij} \sum_k q^{kj} E(\delta_k | \mathcal{V})]$$

$$= \frac{1}{q_{ii}} [E(\delta_i | \mathcal{Y}_i) - E(\delta_i | \mathcal{V}) + q_{ii} \sum_k q^{ik} E(\delta_k | \mathcal{V})] .$$

Let $\tilde{\delta}_i \equiv E(\delta_i | \mathcal{Y}_i)$, and let D be the diagonal matrix with diagonal elements q_{ii}, then the Bayes decision function can be written:

(9) $$\hat{\alpha} = \tilde{\delta} D^{-1} + E(\delta | \mathcal{V})[Q^{-1} - D^{-1}] .$$

CHAPTER III

Quadratic Team with Constant Coefficients of the Quadratic Terms

1. Introduction.

It will now be assumed that the matrix Q is independent of x, i.e., is constant and known. A number of detailed results can be derived in this case, but the reader should keep in mind that this assumption represents an important loss of generality. As thus far discussed, the general quadratic payoff might be thought of as an approximation, for each x, to a smooth payoff function in the neighborhood of the best team action $\gamma(x)$ corresponding to x. If Q is constant, this means that the payoff function in the neighborhood of $\gamma(x)$ is the same for each x, which is clearly a most special circumstance.

This chapter is devoted mainly to an exploration of the role of linear decision functions in this setup. The topics discussed are: (1) normality and the linearity of Bayes decision functions; (2) "Markoff" decision functions, i.e., team analogues of minimum variance linear unbiased estimators; (3) minimax properties of Markoff decision functions; and (4) two examples of no interaction between subfields.

It will be more convenient to use the transformation, as opposed to the subfield, terminology through most of the chapter.

2. Probability Vector Spaces.

This section will set up a linear space framework for the topics discussed in the following sections. The first three parts of this section contain a slight elaboration of material presented by L. J. Savage in recent lectures on the theory of regression and analysis of variance. The fourth part contains a generalization of a known result about the Hadamard product of two non-negative semi-definite matrices.

Most of the section will read like a summary; those statements that are not clearly definitions, or for which no proof is given, can easily be proved by a reader familiar with abstract vector spaces.

2.1. The Structure of a Probability Vector Space.

Let V be a real, finite-dimensional vector space, with dual space V^*. The value of a linear functional v^* in V^* at a point v in V will be denoted by v^*v. Furthermore, let there be a fixed probability measure defined on the Borel subsets of V[1] such that

$$E(v_1^*\mathcal{V})(v_2^*\mathcal{V}) < \infty$$

for all v_1^* and v_2^* in V^*. ($\mathcal{V}$ denotes the variable of integration in V.)

Define $E\mathcal{V}$ to be that element of V for which

[1] The class of Borel subsets of V is the field generated by the class of sets of the form $\{v \mid v^*v \le c\}$, where v^* is any element of V^*, and c is any real number.

$$E(v^* \mathcal{V}) = v^*(E\mathcal{V})$$

for all v^* in V^*.

Define a non-negative product $(\ ,\ )^*$ on V^* by

$$(v_1^*, v_2^*)^* = E[v_1^*(\mathcal{V} - E\mathcal{V})][v_2^*(\mathcal{V} - E\mathcal{V})].$$

Let N^* be the set of all v^* in V^* such that $(v^*, v^*)^* = 0$. N^* is a linear subspace, and $(n^*, v^*) = 0$ for all n^* in N^* and v^* in V^*. Let M^* be any linear subspace of V^* such that V^* is the direct sum of N^* and M^*; then $(\ ,\ )^*$ is positive-definite on M^*.

Define a linear transformation C from V^* to V, as usual, by:

$$v^* C w^* = (v^*, w^*)^*,$$

for all v^* and w^* in V^*. N^* is the null space of C and $C(M^*)$ is the range of C in V. In particular, C is non-singular if and only if $(\ ,\ )^*$ is positive definite.

For any subspace S of V^*, define S^o to be the set of all v in V such that $v^* v = 0$ for all v^* in S. S^o is a subspace of V.

<u>$N^{*o} = C(V^*)$</u>.

<u>Proof</u>. v is in $C(V^*) \Longrightarrow v = C v^*$ for some v^* in $V^* \Longrightarrow$ for every n^* in N^*, $n^* v = n^* C v^* = (n^*, v^*)^* = 0 \Longrightarrow v$ is in N^{*o}. Hence $C(V^*) \subset N^{*o}$; also, these two subspaces have the same dimension.

Since $V^* = N^* \oplus M^*$, it follows that $V = N^{*o} \oplus M^{*o} = C(V^*) \oplus M^{*o}$.

Define an inner product $(\ ,\ )$ on V as follows:

If v_1 and v_2 are in $C(V^*)$, let v_1^* and v_2^* be functionals in V^* such that $v_1 = C v_1^*$, $v_2 = C v_2^*$, and let $(v_1, v_2) = (v_1^*, v_2^*)^*$ (this value is independent of the particular choice of v_1 and v_2). If v_1 is in M^{*o} and v_2 is in V, let $(v_1, v_2) = 0$.

With this definition,

$$(v_1, v_2) = E(v_1, \mathcal{V} - E\,\mathcal{V})(v_2, \mathcal{V} - E\,\mathcal{V}).$$

Note that the inner product $(\ ,\ )$ is independent of the particular choice of M^*. It is positive definite on $C(V^*)$, and, in particular, is therefore positive definite on V if and only if $(\ ,\ )^*$ is positive definite on V^*.

The space V, togehter with its measure, will be called a <u>probability vector space</u>, and the inner product $(\ ,\ )$ will be called its <u>covariance structure</u>. If $(\ ,\ )$ is positive definite, then it is natural to identify V and its dual space by means of the isomorphism, $v = C v^*$. In this case V will be called <u>non-singular</u>.

A probability vector space is called <u>normal</u> if for every v^* in V^*, $v^* \mathcal{V}$ is normally distributed.

2.2. <u>Direct Sums of Probability Vector Spaces</u>.

Let V be a probability vector space, and suppose that V is the external direct sum of the vector spaces $V_1, \ldots, V_N$; i.e., every v in V is an N-tuple $(v_1, \ldots, v_N)$, with v_i in V_i, for every i. The measure on V naturally induces a measure on each V_i. Suppose that the covariance structure $(\ ,\ )_{ii}$ in each V_i is positive definite. For every i and j define a bilinear form $(\ ,\ )_{ij}$ on V_i and V_j by

$$(v_i,v_j)_{ij} = E(v_i,\mathcal{V}_i - E\mathcal{V}_i)_{ii}(v_j,\mathcal{V}_j - E\mathcal{V}_j)_{jj}$$

and define a non-negative inner product $(\ ,\ )$ on V by

$$(v,w) = \sum_{ij} (v_i,w_j)_{ij} .$$

This inner product need not be definite. Furthermore, it is <u>not</u> true that $E(v,\mathcal{V}-E\mathcal{V})(w,\mathcal{V}-E\mathcal{V}) = (v,w)$. However, $(\ ,\ )$ does express the covariance structure in V in the following way. Since each $(\ ,\ )_{ii}$ is definite, any linear functional on V can be represented by some f in V, according to

$$f\cdot v = \sum_i (f_i,v_i)_{ii} .$$

With this particular representation of linear functionals on V, it is clear that

$$Ef\cdot(\mathcal{V} - E\mathcal{V})\, g\cdot(\mathcal{V} - E\mathcal{V}) = (f,g) .$$

For each i and j define the linear transformation R^{ij} from V_j to V_i by

$$(v_i,R^{ij}v_j)_{ii} = (v_i,v_j)_{ij} .$$

Note that R^{ii} is the identity transformation on V_i. For every i and

j, the adjoint T^* of a linear transformation T from V_j to V_i is the linear transformation from V_i to V_j defined by

$$(v_i, T v_j)_{ii} = (T^* v_i, v_j)_{jj} ,$$

for all v_i in V_i and v_j in V_j. It is clear that the adjoint of R^{ij} is R^{ji}.

For any given orthonormal coordinate systems in V_i and V_j, the matrix representing the transformation R^{ij} is the matrix of correlations between the coordinates of $\mathcal{V}_i$ and the coordinates of $\mathcal{V}_j$.

An equivalent, alternative, framework could have been set up in terms of internal subspaces $V_1, \ldots, V_N$, but this would not be as convenient for the purposes of the following sections, in which different V_i's correspond to different team members.

2.3. Conditional Expectation in Direct sums of Normal Probability Vector Spaces.

Let the normal probability vector space V be the external direct sum of the normal probability vector spaces $V_1, \ldots, V_N$. For each i define T_i from V to V_i by

$$T_i v = v_i$$

(where $v = (v_1, \ldots, v_N)$).

For each j and each v_j in V_j, there is a bona fide conditional normal probability measure on V, given $T_j \mathcal{V} = v_j$. Hence for each i and j and each v_j in V_j there is a conditional normal probability measure on V_i, given $T_j \mathcal{V} = v_j$. Translating the well known facts about conditional expectations of normally distributed variables (see, for example, Doob [5], pp. 75-76 and Chapter IV, Section 3) into the present notation, one gets the result:

$$E(\mathcal{V}_i \mid v_j) = E\mathcal{V}_i + R^{ij}(v_j - E\mathcal{V}_j) .$$

2.4. Hadamard Inner Products on Direct Sums.

Let V be the direct sum of probability vector spaces $V_1,\ldots,V_N$, and let $Q = ((q_{ij}))$ be an $N \times N$ positive definite matrix. Define:

$$H(v,w) = \sum_{ij} q_{ij}(v_i,w_j)_{ij}$$

where $v = (v_1,\ldots,v_N)$ and $w = (w_1,\ldots,w_N)$ are in V.

Lemma. If $(\ ,\ )_{ii}$ is positive definite for every i, then $H(\ ,\ )$ is a positive definite inner product on V.

Proof. $H(\ ,\ )$ is obviously bilinear. Since $(\ ,\ )$ is a non-negative semi-definite form on V, there exist vectors $e(1),\ldots,e(M)$ in V such that for all v and w in V:

$$(v,w) = \sum_m (v,e(m))\,(w,e(m)) .$$

For every i, define the linear transformation T_i' from V_i to V by

$$T_i' v_i = (0,\ldots,0,v_i,0,\ldots,0) ,$$

for all v_i in V_i. Then

$$(v_i,w_j)_{ij} = (T_i' v_i, T_j' w_j) .$$

For any v in V,

$$H(v,v) = \sum_{ij} q_{ij} (v_i,v_j)_{ij}$$

$$= \sum_{ij} q_{ij} (T_i' v_i, T_j' v_j)$$

$$= \sum_m \sum_{ij} q_{ij} (T_i' v_i, e(m)) (T_j' v_j, e(m)) \geq 0.$$

Furthermore,

$$H(v,v) = 0 \Rightarrow \text{for every } m \text{ and every } i, \ (T_i' v_i, e(m)) = 0$$

$$\Rightarrow \text{for every } i, \ \sum_m (T_i' v_i, e(m))^2 = (T_i' v_i, T_i' v_i)$$

$$= (v_i,v_i)_{ii} = 0$$

$$\Rightarrow \text{for every } i, \ v_i = 0$$

$$\Rightarrow v = 0 ,$$

which proves the lemma.

The form $H(\ , \ )$ will be called the <u>Hadamard inner product</u> on V induced by Q and $(\ , \ )$.

A special case of the lemma is the known result that the Hadamard product of two positive definite matrices is positive definite. (See Halmos [7], Theorem 2, Section 69.)

3. Normality and the Linearity of Bayes Decision Functions.

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 the theory of the team with a quadratic payoff function. Such a guess would be correct, as will be shown in the present section. The main result here is that if the a priori distribution induces a normal distribution of all the information variables and the vector $\gamma(x)$, then the Bayes decision function is linear in the information variables. An actual algorithm is given.

Under a given a priori distribution, let the set X of states of the world be a direct sum of $(N+1)$ non-singular probability vector spaces $X_0, \ldots, X_N$, and let $\gamma = (\gamma_1, \ldots, \gamma_N)$ be such that each γ_i is a linear functional on X. Suppose that for any state of the world ξ in X, member i $(i = 1, \ldots, N)$ observes only

$$\xi_i = P_i \xi \;,$$

i.e., any decision function for member i must be a measurable function on X_i.

Let $\hat{\alpha}$ be the Bayes team decision function; if the problem were now changed by replacing γ with $(\gamma - E\gamma)$, it is clear from Theorem II.2 that the Bayes decision function would now be $\hat{\alpha} - E\gamma$. It may therefore be assumed, without loss of generality, that $E\gamma = 0$.

Define the function δ by

$$\delta(x) = \gamma(x)$$

The conditional expectation $E[\delta_j(\xi) \mid T_j\xi = x_j]$ is a linear functional on X_j and can therefore be represented as $(d_j, x_j)_{jj}$, for some suitable d_j. Let $d = (d_1, \ldots, d_N)$.

Let S be the linear transformation from the external direct sum $Y \equiv \sum_{i=1}^{N} X_i$ into itself defined by

$$[S\,x]_j = \sum_i q_{ij}\, R^{ji}\, x_i \;.$$

Theorem 1. The Bayes team decision function $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_N)$ is given by

$$\hat{\alpha}_i(\xi_i) = (\hat{a}_i, \xi_i)_{ii} \;,$$

where $\hat{a} \equiv (\hat{a}_1, \ldots, \hat{a}_N) = S^{-1}\,d$.

Proof. Suppose for the moment that S^{-1} exists; then it is sufficient to show that $\hat{\alpha}$ as defind above is stationary, i.e., that

$$E\left\{ \sum_i q_{ij}\, (\hat{a}_i, \xi_i)_{ii} \,\middle|\, x_j \right\} = E\left\{ \delta_j(x) \mid x_j \right\} \;,$$

for all j and x_j. (See Section II.3.) Applying Section III.2.3, this becomes

$$\sum_i q_{ij}\, (\hat{a}_i, R^{ij}\, x_j)_{ii} = (d_j, x_j) \;,$$

or

$$\Big(\sum_i q_{ij}\, R^{ji}\, \hat{a}_i, x_j\Big)_{jj} = (d_j, x_j), \quad \text{for all } j,\, x_j \;,$$

or $$\sum_i q_{ij} R^{ji} \hat{a}_i = d_j, \text{ for all } j ,$$

or $$S \hat{a} = d ,$$

which is so.

It therefore remains to show that S is non-singular.

For any $y = (x_1,\ldots,x_N)$ in Y, define the quadratic form,

$$H(y,y) \equiv \sum_{i,j=1}^{N} q_{ij} (x_i,x_j)_{ij} = \sum_{i=1}^{N} (x_i,[S x]_i)_{ii} .$$

S is non-singular if and only if H is positive definite; but H is the Hadamard inner product on Y induced by Q and the inner product (on Y):

$$(y,z)_Y \equiv \sum_{i,j=1}^{N} (y_i,z_j)_{ij} .$$

Since $(\ , \ )_{ii}$ is positive-definite, it follows from the lemma of Section III.2.4 that H is positive definite, which completes the proof.

One way in which a normal distribution on X can arise is as follows. Let Z be a normal probability vector space that is a direct sum of N non-singular normal probability vector spaces $Z_1,\ldots,Z_N$, with known covariance structure, but with unknown mean. However, for each i let M_i be a linear subspace of Z_i, and suppose that $E\,\xi_i$ is known to lie in M_i; i.e., the set P of probability measures p on Z is the set of all normal distributions on Z that have given covariance structure and such that $E\,\xi_i$ is in M_i for each i. The set P

can therefore be represented by the direct sum $M = \sum_i M_i$, and X can be represented by the direct sum $M \oplus Z$. If the a priori distribution on M is itself normal, this induces a normal distribution on X.

<u>Example</u>. Suppose that the team as a whole has available to it a random sample of K observations from an N-variate normal distribution with unknown mean and known covariances σ_{ij}, but that for every i, member i knows only the sample values of the i^{th} coordinate, and on the basis of those values wants to estimate the corresponding mean μ_i. Suppose further that the a priori distribution specifies that $\mu_1, \ldots, \mu_N$ are independent and normally distributed with means all zero and variances $\tau_1^2, \ldots, \tau_N^2$, respectively.

In the notation of Theorem III.1 this situation can be represented as follows:

1. X_i $(i = 1, \ldots, N)$ is a K-dimensional Cartesian space with points $x_i = (x_i^1, \ldots, x_i^K)$.

2. X_o is an N-dimensional Cartesian space with points $m = (m_1, \ldots, m_N)$.

3. $X = \sum_o^N X_i$ is a normal probability vector space such that for $i, j = 1, \ldots, N$, and $h, k = 1, \ldots, K$:

$$E(\xi_i^h \mid \mu = m) = m_i, \quad E\mu_i = 0$$

$$E[(\xi_i^h - m_i)(\xi_j^k - m_j) \mid \mu = m] = \sigma_{ij}\, \delta_{hk}$$

(where δ_{hk} is the Kronecker delta),

$$E\mu_i \mu_j = \tau_i^2 \delta_{ij} .$$

4. $\gamma_i(x) = m_i, \; i = 1,\dots,N$.

It follows immediately that

$$E \xi_i^h = 0$$

$$E \xi_i^h \xi_j^k = \sigma_{ij} \delta_{hk} + \tau_i^2 \delta_{ij}$$

$$E \xi_i^h \mu_j = \tau_i^2 \delta_{ij} .$$

A slight calculation yields, for i, j = 1,...,N:

$$(x_i, y_i)_{ii} = \frac{K}{\sigma_{ii}} \left[\frac{1}{K} \sum_h (x_i^h - \bar{x}_i)(y_i^h - \bar{y}_i) + c_i \bar{x}_i \bar{y}_i\right] ,$$

where $\bar{x}_i = \frac{1}{K} \sum_h x_i^h$

$$c_i = \frac{1}{1 + \frac{K \tau_i^2}{\sigma_{ii}}} ;$$

$$(x_i, x_j)_{ij} = \frac{K \sigma_{ij}}{\sigma_{ih} \sigma_{jj}} \left[\frac{1}{K} \sum_h (x_i^h - \bar{x}_i)(x_j^h - \bar{x}_j) + c_i c_j \bar{x}_i \bar{x}_j\right]$$

(for $i \neq j$) ;

$$R^{ij} x_j = \frac{\sigma_{ij}}{\sigma_{jj}} [x_j - (1 - c_i\ c_j)\bar{x}_j f] \quad (\text{for } i \neq j) ;$$

$$d_j = \frac{q_{jj} \sigma_{jj} (1 - c_j)}{K\ c_j} f ,$$

where $f = (1, \ldots, 1)$ in Z_j .

According to Theorem III.1, the Bayes decision function $\hat{a}$ is the solution of the linear system:

$$\sum_i q_{ij} R^{ij} a_i = d_j \qquad j = 1, \ldots, N ;$$

$$q_{jj}\ a_j + \sum_{i \neq j} q_{ij} \frac{\sigma_{ji}}{\sigma_{ii}} [a_i - (1 - c_i\ c_j)\bar{a}_i\ f] = \frac{q_{jj} \sigma_{jj}(1-c_j)}{K\ c_j} f, \ j=1, \ldots, N.$$

Hence every $\hat{a}_j$ has the form $a_j^o f$, where a_j^o is a scalar, and $a^o \equiv (a_1^o, \ldots, a_N^o)$ is the solution of

$$q_{jj}\ a_j^o + \sum_{i \neq j} q_{ij} \frac{\sigma_{ji}}{\sigma_{ii}} c_i\ c_j\ a_i^o = \frac{q_{jj} \sigma_{jj} (1-c_j)}{K\ c_j} , \ j = 1, \ldots, N .$$

Let $b_j^o = \frac{K\ c_j}{jj} a_j^o$, then

$$(a_j^o\ f, x_j)_{jj} = b_j^o\ \bar{x}_j ,$$

and the corresponding linear system determining $b^o = (b_1^o, \ldots, b_N^o)$ is:

$$q_{jj}\ \sigma_{jj}\ b_j^o + c_j^2 \sum_{i \neq j} q_{ij}\ \sigma_{ij}\ b_i^o = q_{jj}\ \sigma_{jj}\ (1 - c_j) , \ j = 1, \ldots, N ;$$

$$\text{or} \quad q_{jj}\,\sigma_{jj}\,(1-c_j^2)b_j^o + \sum_i q_{ij}\,\sigma_{ij}\,b_i^o = q_{jj}\,\sigma_{jj}\,(1-c_j)\,.$$

Letting $F = ((q_{ij}\sigma_{ij}))$, $D \equiv$ the diagonal matrix with diagonal entries $q_{jj}\,\sigma_{jj}$, and $C \equiv$ the diagonal matrix with diagonal entries c_j, the last system of equations becomes

$$b^o[FC^2 + D(I-C^2)] = f\,D(I - C)\;,$$

where f now denotes $(1,\ldots,1)$ in R^N. Hence the Bayes decision function is

$$\hat{\alpha}_i(\xi_i) = b_i^o\,\bar{\xi}_i\;,$$

$$\text{where} \qquad b^o = f[I - C][I + FD^{-1}\,C^2 - C^2]^{-1}\;.$$

What happens if the a priori information is "vague", i.e., if $\tau_1^2,\ldots,\tau_N^2$ all become indefinitely large? In that case $c_1,\ldots,c_N$ all approach zero, and therefore b^o approaches f, that is, the best decision function for member i approaches the sample mean $\bar{\xi}_i$ of the values of the corresponding coordinate. (See also Example 3 of the next section.) A similar effect is produced when the sample size K becomes large.

It should not be inferred from this particular example that, in general, as the a priori information becomes more and more vague, the Bayes team decision function tends to a function that is independent of Q. The next two sections discuss an important class of decision functions that are limits of Bayes decision functions (see Lemma 2 of Section III.5).

4. Markoff Decision Functions.

This section and the next deal with a problem that is the team analogue of the one-person problem of minimum-mean-square-error linear unbiased estimation, or the "Markoff problem." In the situation to be considered, each team member observes the value of a different random vector. The covariance structure in each vector space is known, as are the covariances between the vectors of different members, but the mean of the N-tuple of vectors is known only to be in a certain linear subspace of the direct sum of the N vector spaces. Each team member wants to estimate a given linear functional of the mean of his vector, the loss function for the team as a whole being a given quadratic form (determined by Q) in the errors of the estimates. Suppose further that the team wants to use only estimators for which the risk is bounded as a function of the mean, and finally, suppose that the team members want to keep their estimators simple and therefore restrict themselves to linear estimators. A Markoff estimator for the team is one that minimizes the risk for all possible values of the mean, subject to the two conditions of bounded risk and linearity. One result of this section is that the Markoff problem for the team is, in a certain sense, equivalent to an ordinary, one-person, Markoff problem, involving all the vectors together, in which the covariances between the vectors are weighted by the corresponding elements of the matrix Q.

The requirement of bounded risk is close to the minimax principle in spirit, and, in fact, the next section shows that Markoff estimators are actually minimax under certain conditions.

Let the space Z of random events (see Section 1.1) be an external direct sum of N real, finite-dimensional vector spaces $Z_1, \dots, Z_N$; and let M be a subspace of Z. Let P be a set of probability measures p on Z such that:

1. For every p in P, Z is a probability vector space under p, such that each summand Z_i is non-singular.

2. Any two p's in P induce the same inner product on Z.

3. For any p in P, $E(\xi \mid p)$ is in M. For any m in M there exists a p in P such that $E(\xi \mid p) = m$.

The assumption of non-singularity of every Z_i does not entail any loss of generality, as will be shown later.

The set A of allowable decision functions is the set of all α such that each α_i is a (possibly non-homogeneous) linear functional on Z_i. The function γ is assumed to have the form of an N-tuple of linear functionals γ_i of $E\xi_i$. (A slightly different formulation will be considered later.) Since each Z_i is non-singular, $\gamma_i(E\xi_i)$ can be represented as $(g_i, E\xi_i)_{ii}$, and for any decision function α in A, $\alpha_i(z_i)$ can be represented as $(a_i, z_i)_{ii} + a_i^o$, for some a_i in Z_i and real a_i^o.

The risk for α is

$$E \sum_{ij} q_{ij} [(a_i, \xi_i)_{ii} + a_{io} - (g_i, E\xi_i)_{ii}][(a_j, \xi_j) + a_{jo} - (g_j, E\xi_j)_{jj}]$$

$$= \sum_{ij} q_{ij} (a_i, a_j)_{ij} + \sum_{ij} q_{ij} [(a_i - g_i, E\xi_i)_{ii} + a_{io}][(a_j - g_j, E\xi_j)_{jj} + a_{jo}]$$

The function α will be called a <u>bounded-risk decision function</u> if the above risk is bounded as $E\xi$ varies in M. The <u>Markoff problem</u> is to choose a bounded-risk function that minimizes the risk for all $E\xi$ in M.[1]

For every i, let M_i be the set of all z_i in Z_i such that for some $m = (m_1, \ldots, m_N)$ in M, $Z_i = m_i$. It is clear that M_i is a linear subspace of Z_i. Let $M_i^{\perp}$ denote the orthogonal complement of M_i in Z_i; then α is a bounded risk function if and only if, for every i,

$$(a_i - g_i) \text{ is in } M_i^{\perp},$$

as will now be shown.

Suppose, on the contrary, that for some i, there exists an m_i in M_i such that $(a_i - g_i, m_i) \neq 0$. For every $m = (m_1, \ldots, m_N)$ in M let $v(m) = (v_1(m), \ldots, v_N(m))$ be defined by

$$v_j(m) \equiv (a_j - g_j, m_j)_{jj}, \quad j = 1, \ldots, N;$$

then there exists an m^o in M such that $v(m^o) \neq 0$. Furthermore, for all real r, $r\,m^o$ is in M, and $v(rm^o) = rv(m^o)$. Now observe that the first term in the last expression for the risk does not depend upon $E\xi$, and the second term, with $E\xi = m$, is equal to

[1] The idea of replacing the familiar constraint of unbiasedness with the equivalent but intuitively more reasonable constraint of bounded risk seems to be due to L. J. Savage.

$$(v(m) + a^o)\, Q\, (v(m) + a^o)' ,$$

and hence, for $m = rm^o$ the second term is equal to

$$(v(rm^o) + a^o)\, Q\, (v(rm^o) + a^o)$$

$$= r^2 v(m^o)\, Qv(m^o)' + 2rv(m^o)\, Qa^{o'} + a^o\, Qa^{o'} ,$$

where $a^o \equiv (a_1^o, \ldots, a_N^o)$. Since Q is positive definite, $v(m^o)\, Qv(m^o)' > 0$; therefore, the above expression is unbounded (from above) in r, and hence the risk is unbounded on M.

With this established, it is evident that no bounded risk function is admissible unless $a_i^o = 0$ for all i; and whenever a bounded risk function is referred to from now on, it is to be understood that $a_i^o = 0$ for all i. The risk for such a decision function is therefore

$$\sum_{ij} q_{ij}\, (a_i, a_j)_{ij} \equiv H(a,a)$$

where $H(\; , \;)$ denotes the Hadamard inner product on Z induced by Q and $(\; , \;)$. (See Section III.2.4.)

Let N be the interval direct sum $\sum_i M_i^{\perp}$, and let $g = (g_1, \ldots, g_N)$; then a <u>is a solution of the Markoff problem if and only if it minimizes</u> $H(a,a)$ <u>subject to</u> $(a - g)$ <u>in</u> N.

Let $\tilde{M}$ be the orthogonal complement of N in Z, relative to the inner product $H(\; , \;)$. Then it is clear that a is also the solution of the ordinary Markoff problem of estimating the linear functional $H(g, E\xi)$, with Z a covariance space with inner product

$H(\ , \ )$, and with $E\zeta$ known to lie in $\tilde{M}$. As is well known,[1/] the solution of this last problem is given by

$$H(\hat{a},z) = H(g,Pz) \ ,$$

where P is the H-orthogonal projection onto $\tilde{M}$, or

$$\hat{a} = Pg.$$

Note that P is not affected by multiplying either Q or $H(\ , \ )$ by a positive constant.

As in Section III.2, for each i, define the linear transformations T_i and T_i' by:

1. If $z = (z_1,\ldots,z_N)$, $T_i z = z_i$.

2. $T_i' z_i = (0,\ldots,0,z_i,0,\ldots,0)$.

Then for the solution $\hat{a}$ of the Markoff problem,

$$\begin{aligned}
\hat{a}_i &= T_i P g \\
&= T_i P\Big(\sum_j T_j' g_j\Big) \\
&= T_i P T_i' g_i + \sum_{j \neq i} T_i P T_j' g_j \\
&= U_{ii} g_i + \sum_{j \neq i} U_{ij} g_j \ ,
\end{aligned} \qquad (1)$$

1/ See Aitken [1]. Although his result is stated in matrix notation, it can be translated into the geometric statement given here.

where $U_{ij} \equiv T_i P T_j'$. Thus

$$(\hat{a}_i, \xi_i)_{ii} = (g_i, U_{ii}^* \xi_i)_{ii} + (\sum_{j \neq i} U_{ij} g_j, \xi_i)_{ii} ,$$

where U_{ii}^* denotes the adjoint of U_{ii} (in Z_i) .

For a given i, when will it be true that there exists a transformation S_i from Z_i into itself such that, for all g in Z, the best decision function for member i is $(g_i, S_i \xi_i)_{ii}$? According to equation (1) this will be true if and only if,

$$(T_i Pg, z_i)_{ii} = (g_i, S_i z_i), \quad \text{for all } g \text{ and } z ,$$

$$\text{or} \qquad T_i Pg = S_i^* g_i, \quad \text{for all } g,$$

$$\text{or} \qquad T_i P \sum_j T_j' g_j = S_i^* g_i, \quad \text{for all } g;$$

in other words, if and only if P is completely reduced by $T_i'(Z_i)$ and $\sum_{j \neq i} T_j'(Z_j)$. (See Halmos [7], Section 28.)

In particular, the latter is true if M equals the direct sum $\sum_i M_i$, which can be seen as follows. For any z in Z, let $z = x + y$, where x is in $\tilde{M}$ and y is in N; x is the H-orthogonal projection of z onto $\tilde{M}$, since $\tilde{M}$ is by definition the H-orthogonal complement of $N \equiv \sum_i M_i^{\perp}$. On the other hand, for every i, $z_i = m_i + n_i$, where m_i is in M_i and n_i is in $M_i^{\perp}$; but $n \equiv (n_1, \ldots, n_N)$ is in $\sum_i M_i^{\perp}$, and, since $\tilde{M} = \sum_i M_i$, $m \equiv (m_1, \ldots, m_N)$ is in $\tilde{M}$. Hence $x = m$, i.e., the projection Pz of z onto $\tilde{M}$ equals $(P_1 z_1, \ldots, P_N z_N)$, where P_i is the projection, in Z_i, onto M_i.

Example 1. *If the random vectors* ξ_i *are uncorrelated, i.e., if for all* $i \neq j$, *and all* z_i *and* z_j,

$$(z_i, z_j)_{ij} = 0 \ ,$$

then it is clear that $\tilde{M} = \sum_i M_i$.

Example 2. If the matrix Q is diagonal, then $H(z,z) = \sum_i q_{ii}(z_i, z_i)_{ii}$, *and it is therefore clear that* $\tilde{M} = \sum_i M_i$.

Example 3. Suppose that all the spaces Z_i have the same dimension, that all the subspaces M_i have the same dimension, and let $\{I^{ij}\}$ be a family of linear transformations such that for every i, j, and k,

1. I^{ij} is non-singular from Z_j onto Z_i.

2. $I^{ij} I^{jk} = I^{ik}$.

3. $M_i = I^{ij}(M_j)$.

Suppose further that there exists a *symmetric non-negative semi-definite* $N \times N$ matrix $((\sigma_{ij}))$ such that, for every i and j,

$$R^{ij} = \sigma_{ij} I^{ij} \ .$$

(For such a matrix, σ_{ii} necessarily equal 1 for all i, since both R^{ii} and I^{ii} are the identity transformation *on* Z_i.) It follows that

$$\underline{\tilde{M} = \sum M_i}$$

Proof. If x is in $\sum M_i$, then for every z in $\sum M_i^{\perp}$,

$$H(x,z) = \sum_{ij} q_{ij} (x_i, z_j)_{ij}$$

$$= \sum_{ij} q_{ij} (R^{ji} x_i, z_j)_{jj}$$

$$= \sum_{ij} q_{ij} \sigma_{ji} (I^{ji} x_i, z_j)_{jj}$$

$$= 0 ,$$

because $I^{ji} x_i$ is in M_j, for every i and j.

On the other hand, x in $\tilde{M}$ implies:

For every z in $\sum M_i^{\perp}$, $\sum_{ij} q_{ij} (x_i, z_j)_{ij} = 0$,

or $$\sum_j \Big(\sum_i q_{ij} x_i, z_j \Big)_{ij} = 0 .$$

Hence, for every j, and z_j in $M_j^{\perp}$, $\Big(\sum_i q_{ij} x_i, z_j \Big) = 0$,

$$\sum_i q_{ij} (x_i, R^{ij}_{zi})_{ii} = 0 ,$$

$$\sum_i q_{ij} \sigma_{ij} (x_i, I^{ij} z_j)_{ii} = 0 ,$$

Hence, for every j, k, and z_k in $M_k^{\perp}$,

$$\sum_i q_{ij}\,\sigma_{ij}\,(x_i, T^{ij}\,T^{jk}\,z_k)_{ii} = 0\,,$$

$$\sum_i q_{ij}\,\sigma_{ij}\,(x_i, T^{ik}\,z_k)_{ii} = 0\,.$$

By the lemma of Section III.2.4, the matrix $((q_{ij}\,\sigma_{ij}))$ is positive definite, and hence for every i, k, and z_k in $M_k^{\perp}$,

$$(x_i, T^{ik}\,z_k)_{ii} = 0\,.$$

Hence, for every i and every z_i in $M_i^{\perp}$,

$$(x_i, z_i)_{ii} = 0\,,$$

i.e., x is in $\sum M_i$, which completes the proof.

Two extreme special cases of this example are obtained if either (1) the vectors ξ_i are uncorrelated ($((\sigma_{ij}))$ is the identity matrix), (see also Example 1) or (2) the vectors ξ_i are perfectly correlated ($\sigma_{ij} = 1$ for all i and j). The latter case is, of course, essentially equivalent to the situation in which all the team members observe the same random vector.

A third special case of this example is obtained if the team as a whole has available to it a random sample of observations from an N-variate distribution with unknown mean and known covariances σ_{ij}, but, for every i, member i knows only the sample values of the i^{th} coordinate. This is the same setup as that of the example of Section III.4, and if, for example, each member wants to estimate the mean of his corresponding coordinate, then he would use the sample mean for his coordinate, as in that example in the case of infinitely "vague" a priori information.

The following example describes a situation in which $\tilde{M}$ is not equal to $\sum\sum M_{ij}$. The situation is also an example of "sharing independent data" (see Section II. 3).

Example 4. Suppose there are only two team members. Let ξ_1, ξ_2, and ξ_3 be independent and identically distributed random variables with unknown mean, and variance 1, and suppose that member 1 observes $\zeta_1 \equiv (\xi_1, \xi_2)$, and member two observes $\zeta_2 \equiv (\xi_1, \xi_3)$. The covariance matrix of ζ_1 and ζ_2 together is therefore

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

Suppose that $Q = \begin{pmatrix} 1 & q \\ q & 1 \end{pmatrix}$, and that each member wants to estimate $E\xi_i$. The Markoff decision function $a = (a_{11}, a_{12}; a_{21}, a_{22})$ therefore minimizes

$$H(a,a) = \sum_{ih} a_{ih}^2 + 2q\, a_{11}\, a_{21}$$

subject to the constraints:

$$a_{11} + a_{12} = a_{21} + a_{22} = 1 .$$

The solution is easily found to be:

$$\left.\begin{aligned} a_{i1} &= \frac{1}{2+q} \\ a_{i2} &= \frac{1+q}{2+q} \end{aligned}\right\} i = 1, 2 .$$

Note that if $q = 0$, the best estimator for member i is his sample mean $\bar{\xi}_i$. If $q > 0$, then the loss is greater the higher the (positive) correlation between the errors of the two members; hence less weight is given to the variable that is observed by both.

<u>The Case of Z_i's with Non-Definite Covariance Structures</u>.

Any case in which some of the summands Z_i have covariance structures that are not definite can be reduced to the definite case, as follows.

First note that the Markoff problem, stated in the dual spaces Z_i^*, reduces to:

"Minimize $\sum_{ij} q_{ij} \, (a_i^*, a_j^*)_{ij}^*$ subject to a_i^* in $(g_i^* + M_i^o)$ for all i,"

where the bilinear forms $(\; , \;)_{ij}^*$ are defined by:

$$(z_i^*, y_j^*)_{ij}^* = E \, z_i^* \, (\xi_i - E \, \xi_i) y_j^* \, (\xi_j - E \, \xi_j) ,$$

for every i and j and every z_i^* in Z_i^* and y_j^* in Z_j^*, and where M_i^o is the annihilator (in Z_i^*) of M_i.

For every i, let N_i^* be the subspace of all z_i^* such that $(z_i^*, z_i^*)_{ii}^* = 0$ (as in Section III.2.1); let S_i^* be any other subspace of Z_i^* such that $Z^* = N_i^* \oplus S_i^*$; and let K_i be the projection onto

S_i^* along N_i^*. Then it can easily be verified that, for every i and j, and every z_i^* and z_j^*,

$$(z_i^*, z_j^*)_{ij}^* = (K_i z_i^*, K_j z_j^*)_{ij} .$$

It follows that any solution of the following problem will also be a solution of the above Markoff problem:

"Minimize $\sum_{ij} q_{ij} (b_i^*, b_j^*)_{ij}^*$ subject to b_i^* in $(K_i g_i^* + K_i(M_i^o))$ for all i."

Therefore, according to the discussion of Section III.2.1, each Z_i can be replaced by N_i^{*o}, M_i by $C_i(K_i M_i^o)$, and g_i^* by $C_i(K_i g_i^*)$, where C_i is as defined in that section.

<u>The Case of M Defined in Terms of a Parameter Space.</u>

In many problems the set of possible values of the mean $E\xi$ is defined as the image, under a linear transformation, of a "parameter space." Thus, let V be a vector space; for each i, let B_i be a linear transformation from V to Z_i, and suppose that for every possible value of $E\xi$ there is a v in V such that

$$E\xi_i = B_i v .$$

Suppose, further, that each γ_i is a linear functional on V. The risk for the linear decision function φ, given that

$$E\xi_i = B_i v, \text{ is } E \sum_{ij} q_{ij}[\varphi_i(\xi_i) - \gamma_i(v)][\varphi_j(\xi_j) - \gamma_j(v)]$$

$$= \sum_{ij} q_{ij} E[\alpha_i(\xi_i - E\xi_i)][\alpha_j(\xi_j - E\xi_j)]$$

$$+ \sum_{ij} q_{ij} [\alpha_i(E\xi_i) - \gamma_i(v)][\alpha_j(E\xi_j) - \gamma_j(v)].$$

Hence there will exist a linear bounded-risk decision function if and only if, for each i, there exists a linear functional α_i on Z_i such that, for all v in V,

$$\alpha_i(B_i v) = \gamma_i(v) ,$$

or

$$(B_i^* \alpha_i)(v) = \gamma_i(v) ,$$

where B_i^* is the adjoint of B_i. Hence a linear bounded-risk decision function exists if and only if

$$B_i^* \alpha_i = \gamma_i$$

has a solution, for every i. Note, however, that if the latter is true, then for every i there exists a g_i in Z_i, such that

$$(g_i, B_i v)_{ii} = \gamma_i(v)$$

for all v in V. Representing α_i by $\alpha_i(z_i) = (a_i, z_i)_{ii}$, it follows that $a = (a_1, \ldots, a_N)$ is a bounded-risk decision function if and only if, for every i,

$$(a_i - g_i, B_i v)_{ii} = 0$$

for every v in V, i.e., if and only if, for every i, $a_i - g_i$ is in $M_i^{\perp}$, where $M_i \equiv B_i(V)$. Note that the set M of possible values of $E\xi$ is a linear subspace of Z, and that the definition of the subspaces M_i just given coincides with the earlier definition.

5. <u>Minimax Properties of Markoff Decision Functions.</u>

In the decision situation described in the previous section, it can be shown that the Markoff decision function is minimax in the set of all team decision functions. More generally, suppose that the inner product expressing the covariance structure is known only to be one of a class of such inner products; then for certain such classes there exists a Markoff decision function, relative to one of the inner products in the class, that is minimax in the set of <u>all</u> team decision functions. Three cases of this kind will be considered in this section. Theorem 7 contains, as a special case, new results for the one-person Markoff problem.

As in the previous section, let $Z = \sum_{i=1}^{N} Z_i$ be the external direct sum of N real, finite-dimensional vector spaces Z_i, and let M be a subspace of Z. Let $\mathcal{R}$ be a set of non-negative inner products R <u>on the dual space</u> Z^*, and for each R let $\{R_{ij}\}$ be the corresponding set of bilinear functionals such that

$$R(x^*, y^*) = \sum_{ij} R_{ij}(x_i^*, y_j^*) \ .$$

The statement that Z is a direct sum of probability vector spaces "with inner product R on Z^*" will mean that

$$R_{ij}(x_i^*, y_j^*) = E[x_i^*(\xi_i - E\xi_i)][y_j^*(\xi_j - E\xi_j)] \ .$$

Given $\mathcal{R}$, let P be a set of probability measures p on Z such that:

1. For every p in P, there exists an R in $\mathcal{R}$ and an m in M such that Z is a direct sum of probability vector spaces under p, with inner product R on Z^* and mean $E(\xi | p) = m$.

2. For every R in $\mathcal{R}$ and m in M, there exists a p in P such that Z is a direct sum of normal probability vector spaces under p, with inner product R on Z^* and mean $E(\xi | p) = m$.

For any R, let H_R denote the Hadamard inner product induced by R and $Q \equiv ((q_{ij}))$. For every i, define M_i as in the previous sections, and let $N = \sum_i M_i^o$ (recall that M_i^o is the annihilator of M_i). For each i let g_i be a fixed linear functional on Z_i (where $\gamma_i(m_i) = g_i m_i$). Then according to the result of the previous section, here rephrased in terms of the dual space, $\hat{a}$ is Markoff relative to R if and only if it minimizes $H_R(a,a)$ subject to $(a - g)$ in N. Recall that multiplying R by a positive constant does not result in a change in the corresponding $\hat{a}$.

The main tool of this section is Lemma 3 below, which is an immediate consequence of the following two lemmas. The first of these is a trivial result which holds for general risk functions.

Lemma 1. If the risk $\rho(\varphi, p) \leq k$ for all p in P, and φ is minimax in a subset P' of P on which $\sup_p \rho(\varphi, p) = k$, then φ is minimax in P.

The next lemma is a generalization of the well-known result that the minimax estimate (from a random sample) of the mean of a normal distribution with known variance is the sample mean. (See Wald [15], p. 142, Hodges and Lehmann [16], Theorem 6.5, and Girshick and Savage [6].)

Lemma 2. If $\mathcal{R}$ consists of a single non-negative inner product R such that R_{ii} is positive definite for every i, and if P contains only normal distributions, then the Markoff decision function relative to R is minimax in the set of all decision functions.

Proof: The Markoff decision function has constant risk; therefore, to show that it is minimax it is sufficient to show that it is Bayes against a sequence of a priori distributions on P.

Adopting the "parameter space" formulation (see the end of the previous section), let V be a vector space and, for every i, let B_i be a linear transformation from V to Z_i such that the linear subspace of points $(B_1 v, \ldots, B_N v)$ in Z, as v ranges over V, is the set of all possible values of $E\xi$. It is clear that P can be represented by V; therefore, let R_{oo} be any positive definite inner product on V^*, the dual space of V, let r be any positive real number, and suppose that the a priori distribution on P is represented by a normal distribution on V, with mean zero and covariance structure determined by:

$$E(v^* \nu)(w^* \nu) = r\, R_{oo}(v^*, w^*) \ ,$$

for all v^* and w^* in V^*. This distribution, together with the relations

$$E \left\{ \xi_i \mid \nu = v \right\} = B_i\, v$$

$$E \left\{ z_i^*(\xi_i - B_i\, v) y_j^* (\xi_j - B_j\, v) \mid \nu = v \right\} = R_{ij}(z_i^*, y_j^*) \ ,$$

for all i and j, and all z_i^* in Z_i^* and y_j^* in Z_j^*, determine a normal distribution on the direct sum $X = V \oplus Z$, such that

$$E \xi_i = 0 ,$$

$$E(z_i^* \xi_i)(y_j^* \xi_j) = R_{ij}(z_i^*, y_j^*) + r R_{oo}(B_i^* z_i^*, B_j^* y_j^*) ,$$

$$E(z_i^* \xi_i)(v^* \nu) = r R_{oo}(B_i^* z_i^*, v^*) ,$$

for all i, j, z_i^* in Z_i^*, y_j^* in Z_j^* and v^* in V^*, where B_i^* is the adjoint of B_i.

For every i, let γ_i be the linear functional g_i in V_i^*. According to Theorem III.1, the Bayes decision function is linear. The Bayes risk for any linear decision function $a = (a_1, \ldots, a_N)$, with a_i in Z_i^*, is

$$E \sum_{ij} q_{ij}(a_i \xi_i - g_i \nu)(a_j \xi_j - g_j \nu)$$

$$= \sum_{ij} q_{ij}[R_{ij}(a_i, a_j) + r R_{oo}(B_i^* a_i, B_j^* a_j) - r R_{oo}(B_i^* a_i, g_j)$$

$$- r R_{oo}(g_i, B_j^* a_j) + r R_{oo}(g_i, g_j)]$$

$$= \sum_{ij} q_{ij} R_{ij}(a_i, a_j) + r \sum_{ij} q_{ij} R_{oo}(B_i^* a_i - g_i, B_j^* a_j - g_j) .$$

Hence minimum Bayes risk is the minimum of the above expression with respect to a. As r increases without limit, the minimum Bayes risk therefore approaches as a limit the minimum of $\sum_{ij} q_{ij} R_{ij}(a_i, a_j)$ subject to the constraint:

$$\sum_{ij} q_{ij} R_{oo}(B_i^* a_i - g_i, B_j^* a_j - g_j) = 0 .$$

By the lemma on Hadamard inner products of Section III.2.4, the constraint is satisfied if and only if

$$B_i^* a_i - g_i = 0, \quad i = 1, \ldots, N.$$

This is the condition for a linear bounded risk function; hence as r increases without limit, the Bayes risk approaches the Markoff risk as a limit, which completes the proof.

For any (homogeneous) bounded-risk linear decision function a (and therefore, for any Markoff decision function), the risk function depends upon p in P only through the corresponding R in $\mathcal{R}$, and is in fact equal to $H_R(a,a)$; because of this the risk for a bounded-risk linear decision function a will be denoted by $\rho(a,R)$.

Lemmas 1 and 2 result immediately in:

Lemma 3. If $\hat{a}$ is Markoff relative to $\hat{R}$, and if $\rho(\hat{a},R) \leq \rho(\hat{a},\hat{R})$ for all R in $\mathcal{R}$, then $\hat{a}$ is minimax among all team decision functions.

If $\hat{a}$ is minimax, and Markoff relative to $\hat{R}$, as in Lemma 3, then although there may be no least favorable a priori distribution, $\hat{R}$ is, in a sense, a least favorable covariance structure. In the following three theorems three different classes $\mathcal{R}$ are considered; in each case there is some Markoff decision function, relative to a least favorable R, that is minimax.

The next theorem is the team analogue of a theorem of Hodges and Lehmann ([16], Theorem 6.5).

Theorem 5. Let R be a fixed non-negative inner product on Z^* such that R_{ii} is definite for every i; let k be a fixed positive number; and let $\mathcal{R}$ be the set of all cR such that $0 \leq c \leq k$. Then the Markoff decision function a relative to kR is minimax. (Note that a does not depend upon the value of k.)

Proof: For any $c \leq k$,

$$\rho(a, cR) = c\, H_R(a,a) \leq k\, H_R(a,a) = \rho(a, kR) ,$$

so that Lemma 3 applies.

The next theorem deals with the situation in which the covariance structure of the random vector observed by each team member is known, but the covariances between the vectors observed by different team members are unknown.

Theorem 6. For each i, let R^o_{ii} be a fixed positive definite inner product on Z^*_i, and let $\mathcal{R}$ be the set of all non-negative inner products R on Z^* such that $R_{ii} = R^o_{ii}$ for all i. Then there exists $\hat{R}$ in $\mathcal{R}$ such that the Markoff decision function $\hat{a}$ relative to $\hat{R}$ is minimax.

Proof: According to Lemma 3, the theorem will be proved if it can be shown that the function $\rho(a,R)$ has a saddle point (minimax) for a in $(g + N)$ and R in $\mathcal{R}$.

Let $\mathcal{H}$ be the set of all H_R such that R is in $\mathcal{R}$, and let $\mathcal{A}$ be the smallest closed convex set containing all a such that a is Markoff relative to R for some R in $\mathcal{R}$. It is clear that it is sufficient to prove that the function $H(a,a)$ has a saddle point for a in $\mathcal{A}$ and H in $\mathcal{H}$. According to a paper of Kakutani [10], such a saddle point will exist if for some topologies on $\mathcal{A}$ and $\mathcal{H}$,

1. $H(a,a)$ is continuous on $\mathcal{A} \times \mathcal{H}$.

2. $\mathcal{A}$ and $\mathcal{H}$ are each convex and compact.

3. For every a^o in $\mathcal{A}$, the set of all H that maximize $H(a^o,a^o)$ on $\mathcal{H}$ is convex.

For every H^o in $\mathcal{H}$, the set of all a that minimize $H^o(a,a)$ on $\mathcal{A}$ is convex.1/

Let R^o be the inner product on Z^* defined by

$$R^o(a,a) = \sum_i R_{ii}(a_i,a_i) ;$$

1/ Although the conditions given here are not explicitly written out in the paper cited, they are implicit in the proof of Theorem 3. Kakutani's result is not the most general saddle-point theorem available, but is sufficient for this proof. For references to other literature on the subject see Debrue [4].

note that R^o is positive definite. Relative to R^o, one can define, as usual, a self-adjoint transformation H, from Z^* into itself, corresponding to each H(,) in $\mathcal{H}$, by

$$H(a,b) = R^o(a, Hb) ,$$

for all a and b in Z^*. Having made this identification between inner products and transformations, it will be convenient to think of $\mathcal{H}$ as a set of transformations. Consider $\mathcal{H}$ as embedded in a vector space, in the usual way, and define the positive definite inner product

$$H \cdot K \equiv \text{trace } (HK);$$

the resulting topology on $\mathcal{H}$ is the one that will be used in this proof. It is easily verified that $\mathcal{H}$ is convex and closed; furthermore, $\mathcal{H}$ is bounded because for any H in $\mathcal{H}$ (recall that H is positive definite),

$$\text{trace } (H^2) \leq [\text{trace } (H)]^2 ,$$

and the trace of H equals $t = \sum_i q_{ii} n_i$, where n_i is the dimension of Z_i^*. Hence $\mathcal{H}$ is compact.

The topology to be used on $\mathcal{A}$ is the one induced by R^o. For any H in $\mathcal{H}$ let r(H) denote the smallest characteristic root of H with respect to R^o. Let a^o be some fixed element of $\mathcal{A}$; then for any $H = H_R$ in $\mathcal{H}$, and for $\hat{a}$ Markoff relative to R,

$$R^o(a,a)\, r(H) \leq R^o(a,Ha) \leq R^o(a^o,Ha^o) \leq R^o(a^o,a^o) \text{ trace } (H) .$$

On the one hand, trace (H) has the same value, t, for all H in $\mathcal{H}$ (see above). On the other hand,

$$\inf_{H \text{ in } \mathcal{H}} r(H) \equiv r_o > 0 ,$$

which follows from the continuity of $r(H)$ on $\mathcal{H}$, the compactness of $\mathcal{H}$, and the fact that $r(H) > 0$ for all H in $\mathcal{H}$. It follows that

$$R^o(\hat{a},\hat{a}) \leq \frac{t}{r_o} R^o(a^o,a^o)$$

for all $\hat{a}$ that are Markoff relative to some R in $\mathcal{R}$; therefore $\mathcal{A}$ is bounded, and hence compact.

For any fixed a^o in $\mathcal{A}$, the function $-R^o(a^o,Ha^o)$ is linear on $\mathcal{H}$, which is convex; hence the set of H's that maximize this function on $\mathcal{H}$ is convex. For any fixed H^o in $\mathcal{H}$, there is only one a in $\mathcal{A}$ that minimizes $R^o(a,H^oa)$. Hence condition 3 is satisfied. $R^o(a,Ha)$ is clearly continuous on $\mathcal{A} \times \mathcal{H}$, which completes the proof of the theorem.

If nothing at all is known about R, that is, if $\mathcal{R}$ is taken to be the class of all possible inner products on Z^*, then the risk for every decision function is unbounded on $\mathcal{R}$. To get a finite minimax value, the class $\mathcal{R}$ must be "bounded" in some sense. One such sense is provided by the concept of the norm of one quadratic form with respect to another. Let R^o be a given positive definite inner product on Z^*, and for any inner product R, define the norm of R (with respect to R^o) by

$$\| R \| \equiv \max_{R^o(a,a) = 1} R(a,a) .$$

<u>Theorem</u> 7. Let $\mathcal{R}$ be the set of all R such that $\| R \| \leq k$, where

k is a given positive number, and the norm is with respect to a given R^o; then the Markoff decision function a relative to kR^o is minimax. Note that a does not depend upon the value of k, but does depend upon R^o.

(This last fact might cause some uneasiness about the application of the minimax principle in this situation. In practice, one typically is willing to grant that $\mathcal{R}$ is bounded somehow, but does not have a very precise idea about the nature of the bound, whereas this theorem shows that the minimax decision function is quite sensitive to the nature of the bound.)

<u>Proof</u>: Since Q is positive definite, there exist $\{q_i(n)\}$, $i = 1,..,N$, such that

$$q_{ij} = \sum_n q_i(n)\, q_j(n)$$

for all i and j. For a Markoff relative to kR^o, and for any R in $\mathcal{R}$,

$$\rho(a,R) = \sum_{ij} q_{ij}\, R_{ij}(a_i,a_j)$$

$$= \sum_n \sum_{ij} R_{ij}(q_i(n)a_i, q_j(n)\, a_j)$$

$$\leq \sum_n \|R\| \sum_{ij} R^o_{ij}(q_i(n)a_i, q_j(n)a_j)$$

$$= \|R\| \sum_{ij} \sum_n q_i(n) q_j(n) R^o_{ij}(a_i, a_j)$$

$$= \|R\| \sum_{ij} q_{ij} R^{o}_{ij}(a_i, a_j)$$

$$\leq k H_{R^o}(a,a)$$

$$= \rho(a, kR^{o}) .$$

Hence Lemma 3 applies to prove the theorem.

6. Two Examples of no Interaction between Subfields.

As was pointed out in Section I.2, it is only under special circumstances that the minimum Bayes risk, as a function of $\mathcal{Y}_1, \ldots, \mathcal{Y}_N$, can be expressed as a sum of functions $\hat{V}_1, \ldots, \hat{V}_N$, such that each $\hat{V}_i$ depends upon $\mathcal{Y}_i$ alone; in particular this can be done if the payoff function can be expressed as a sum of functions $u_1, \ldots, u_N$ such that for a.e. x each u_i is a function of a_i alone. In the general quadratic case this means that $Q(x)$ is diagonal a.e.

In this section two other examples of lack of interaction between subfields will be given, both of them dependent on the assumption that Q is constant.

6.1. Independent Subfields.

Suppose that there is some fixed a priori distribution on P. For each i let $\mathcal{S}_i$ be a family of subfields of $\mathcal{X}$, and let $\mathcal{S}$ be the Cartesian product $\prod_{i=1}^{N} \mathcal{S}_i$. <u>Assume that every</u> $\mathcal{Y} = (\mathcal{Y}_1, \ldots, \mathcal{Y}_N)$ <u>in</u> $\mathcal{S}$ <u>is a statistically independent set of subfields</u>. Under this assumption, for any $\mathcal{Y}$ in $\mathcal{S}$ the condition of stationarity becomes (see Section II.3):

$$q_{ii} \alpha_i + \sum_{j \neq i} q_{ij} \, E \alpha_j = E[\delta_i | \mathcal{Y}_i], \text{ for every } i,$$

(where $\delta = \gamma Q$), or

$$\alpha_i = \frac{1}{q_{ii}} \left\{ E[\delta_i | \mathcal{Y}_i] - \sum_{j \neq i} q_{ij} \, E \alpha_j \right\} , \text{ for every } i .$$

Therefore, the Bayes decision function $\hat{\alpha}$ has the form:

$$(\tilde{\delta} - c)\, D^{-1} ,$$

where $\tilde{\delta}_i = E[\delta_i \mid \mathcal{Y}_i]$, D is the diagonal matrix with diagonal entries q_{ii}, and c is some constant vector. It is easily verified that c must equal $E\,\delta(I - Q^{-1}D)$, so that

$$\hat{\alpha} = \tilde{\delta}D^{-1} - E\,\delta(D^{-1} - Q^{-1})$$

$$= \tilde{\delta}D^{-1} - E\,\delta D^{-1} + E\,\gamma .$$

Note that $E\,\hat{\alpha} = E\,\gamma$. A slight calculation shows that the Bayes risk for $\hat{\alpha}$ is

$$\hat{V}(\mathcal{Y}) = E(\hat{\alpha} - \gamma)Q(\hat{\alpha} - \gamma)^1$$

$$= \| \gamma - E\,\gamma \| - \sum_i \frac{1}{q_{ii}} \operatorname{Var}(\tilde{\delta}_i) .$$

Thus the statistical independence of $\mathcal{Y}_1, \ldots, \mathcal{Y}_N$ has the effect of splitting the team into N independent one-person teams, the i^{th} person having a payoff function

$$u_i(a_i,x) = \left(a_i - \left[\frac{\delta_i(x) - E\,\delta_i}{q_{ii}} + E\,\gamma_i \right] \right)^2$$

The generalization to the case in which the team members can be partitioned into groups, with statistical independence between groups, is obvious.

6.2. Sharing Independent Data.

Consider the example of Section II.3, with the subfields $\mathcal{U}_1, \ldots, \mathcal{U}_N$

fixed; let $\mathcal{V}$ be the set of all fields of the form $F(\mathcal{V}_1 \cup \dots \cup \mathcal{V}_N)$[1/] such that for each i, $\mathcal{V}_i$ is a subfield of $\mathcal{U}_i$; and let $\mathcal{S}$ be the set of all $\mathcal{Y} = (\mathcal{Y}_1, \dots, \mathcal{Y}_N)$ such that for some $\mathcal{V}$ in $\mathcal{V}$, $\mathcal{Y}_i = F(\mathcal{U}_i \cup \mathcal{V})$, for every i.

Let $\mathcal{W}_1, \dots, \mathcal{W}_N$ be a fixed, statistically independent N-tuple of subfields of $\mathcal{X}$ such that $\mathcal{U}_i \subset \mathcal{W}_i$, for every i, and make the assumption:

δ_i <u>is</u> $\mathcal{W}_i$<u>-measurable, for every</u> i.

It will now be shown that within $\mathcal{S}$ there is no interaction between the subfields $\mathcal{Y}_i$.

For every $\mathcal{Y}$ in $\mathcal{S}$, and for every i, it follows from the assumptions that:

$$\tilde{\delta}_i = E(\delta_i \mid \mathcal{V}_i) ,$$

$$\bar{\delta}_i = E(\delta_i \mid \mathcal{V}_i) .$$

By Section II.3, the Bayes decision function is

$$\hat{\gamma} = \tilde{\delta} D^{-1} + \bar{\delta}(Q^{-1} - D^{-1}) ,$$

where D is the diagonal matrix with diagonal entries q_{ii}, and hence (recall that $\delta Q^{-1} = \gamma$) ,

---

[1/] If $\mathcal{J}$ is a collection of subsets of X, then $F(\mathcal{J})$ denotes the smallest field containing $\mathcal{J}$.

$$\hat{\alpha} - \gamma = \tilde{\delta} D^{-1} + \bar{\delta}(Q^{-1} - D^{-1}) - \delta Q^{-1}$$

$$= (\tilde{\delta} - \bar{\delta})(D^{-1} - Q^{-1}) + (\tilde{\delta} - \delta) Q^{-1} .$$

Under the assumption of this section for, all $i \neq j$,

$$E(\tilde{\delta}_i - \bar{\delta}_i)(\tilde{\delta}_j - \tilde{\delta}_j) = E(\tilde{\delta}_i - \delta_i)(\tilde{\delta}_j - \delta_j) = 0 ,$$

and for all i and j,

$$E(\tilde{\delta}_i - \bar{\delta}_i)(\tilde{\delta}_j - \delta_j) = 0 .$$

The Bayes risk for $\hat{\alpha}$ is therefore

$$\hat{V}(\mathcal{Y}) = \sum_i \left\{ (q^{ii} - \frac{1}{q_{ii}}) \, E \, Var \, (\tilde{\delta} \mid \mathcal{V}_i) + q^{ii} \, E \, Var \, (\delta \mid \mathcal{U}_i) \right\} ,$$

where $((q^{ij})) \equiv Q^{-1}$. Hence, within $\mathcal{S}$, there is no interaction between the subfields $\mathcal{Y}_i$.

REFERENCES

[1] A. C. Aitken, "On least squares and the linear combination of observations", Proceedings of the Royal Society of Edinburgh, V. 55 (1935), 42-48.

[2] R. R. Bahadur, "Sufficiency and statistical decision functions", Annals of Mathematical Statistics, V. 25 (1954), 423-462.

[3] R. R. Bahadur and E. L. Lehmann, "Two comments on 'Sufficiency and statistical decision functions'", Annals of Mathematical Statistics, V. 26 (1955), 139-141.

[4] G. Debreu, "A social equilibrium existence theorem", Proceedings of the National Academy of Science, V. 38 (1952), 886-893.

[5] J. L. Doob, Stochastic Processes, New York: John Wiley and Sons, 1953.

[6] M. A. Girshick and L. J. Savage, "Bayes and minimax estimates for quadratic loss functions", Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley: University of California Press, 1951, 53-73.

[7] P. R. Halmos, Finite Dimensional Vector Spaces, (Annals of Mathematics Study No. 7) Princeton: Princeton University Press, 1948.

[8] P. R. Halmos, Introduction to Hilbert Spaces, New York: Chelsea, 1951.

[9] P. R. Halmos, Measure Theory, New York: Van Nostrand, 1950.

[10] S. Kakutani, "A generalization of Brouwer's fixed point theorem", Duke Mathematics Journal, V. 8 (1941), 457-459.

[11] J. Marschak, "Elements for a theory of teams", Management Science, V. 1 (1955), 127-137.

[12] R. Radner, "The linear team: an example of linear programming under uncertainty", Proceedings of a Symposium on Linear Programming, Washington: United States Government Printing Office (forthcoming).

[13] L. J. Savage, The Foundations of Statistics, New York: John Wiley and Sons, 1954.

References (Cont.)

[14] M. H. Stone, Linear Transformations in Hilbert Space, New York: American Mathematical Society, 1932.

[15] A. Wald, Statistical Decision Functions, New York: John Wiley and Son, 1950.

[16] J. L. Hodges and E. L. Lehmann, "Some problems in minimax point estimation", Annals of Mathematical Statistics, V. 21 (1950), 182-197.

[17] T. Bonnesen and W. Fenchel, Theorie der Konvexen Körper, Berlin: J. Springer, 1934.

[18] J. Marschak and D. Waterman, "On optimal communication rules for teams", Cowles Commission Discussion Paper, Economics No. 2058, November 21, 1952, (unpublished).

[19] M. Beckmann, "On Marschak's model of an arbitrage firm", Cowles Commission Discussion Paper, Economics No. 2058, November 21, 1952, (unpublished).

[20] R. Radner, "On optimal communication rules for certain types of teams", Cowles Commission Discussion Paper, Economics No. 2064, January 8, 1953, (unpublished).

[21] J. Kiefer and S. Orey, "On the arbitrage problem", Cowles Commission Discussion Paper, Economics No. 2068, February 23, 1953, (unpublished).

[22] D. Bratton, "On the arbitrage problem", Cowles Commission Discussion Paper, Economics No. 2075, May 13, 1953, (unpublished).

[23] J. Marschak and R. Radner, "The firm as a team", Econometrica, V. 22, (1954), 523 (abstract).