The Annals of Statistics
1983, Vol. 11, No. 4, 1032–1046

ELEMENTS OF MULTI-BAYESIAN DECISION THEORY[1]

BY S. WEERAHANDI AND J. V. ZIDEK

University of Sri Jayewardenapura and University of Washington

This work provides the elements of a framework for multi-Bayesian statistical decision theory. Solution concepts and criteria are presented. The relationship to Wald's theory is discussed. And two criteria for assessing group decision procedures are defined. One is based on the idea of subsampling the group, and it is found that among the proposed solution concepts only Nash's solution is optimal under subsampling as well. The other assumes the group is itself a sample from a superpopulation, and this yields an analogue of Wald's theory where the elicitation of the priors becomes part of the experimental process. Results on admissibility, minimaxity and so on found in Wald's classical theory become directly applicable in the new setting.

1. Introduction. This paper presents a framework for a normative theory of multi-Bayesian statistical decision theory. Of interest is a group of Bayesians, B, who must jointly select a, possibly randomized, decision rule δ. This group has been called a "bunch" (Fisher, 1972), a "population" (Dickey and Freeman, 1975) and a "bevy" (Dawid, 1982). The worth of δ to each Bayesian, $\beta \in B$ is determined by his prior or posterior (as appropriate) expected gain (or loss) of utility.

An obstacle to the development of a uniquely acceptable solution to the group decision problem is the lack of a generally acceptable basis for inter-Bayesian utility comparisons. Without intercomparable utilities, Arrow's celebrated impossibility theorem (Arrow, 1966) makes it unreasonable to postulate that a group will possess a group preference ordering, which completely orders the class of available δ's and thereby points to a best choice.

Section 2 is a survey which complements that of Weerahandi and Zidek (1981) of the criteria and solution concepts which have been proposed for group decision processes like that under consideration. The Nash (1950) solution is the only one presented which does not require an explicit comparison of utilities and it is easily the most distinguished. It is applied in an illustrative example where just two Bayesians seek a joint estimate of a multivariate normal mean vector under a conjugate gain-in-utility structure. The obvious answer, the average of their posterior means, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, proves to be the Nash solution when there is a sufficiently high degree of consensus, i.e., when the Mahalanobis distance, D^2, between these means is sufficiently small, i.e. $D^2 \le 4$. Otherwise a randomized rule is optimal.

In Section 3, a method of assessing, by hypothetical subsampling, a prospective δ is given and some of its implications are worked out. The acceptability of δ is tested by ascertaining if it would remain the group's choice were it known that it would be implemented by, and change the utilities of, only the members of an as yet unspecified subgroup. Only the Nash solution survives this test when the subgroup is chosen by random sampling.

Section 4 briefly describes another method of assessment. The elicited priors are regarded as observables which have, with the data, a joint sampling distribution given the true state of nature. The full implications of this view remain to be worked out, but an

Received October 1982; revised June 1983.

[1] Dedicated to Professor George Barnard.

AMS 1980 *subject classifications*. Primary 62A15, secondary 62C99.

Key words and phrases. Bayes estimate, Nash solution, multivariate normal law, statistical decision theory, utility theory, multi-person games, bargaining games, opinion pooling.

example illustrates in the uni-Bayesian case, the almost paradoxical difficulties which can arise.

The new material in this paper is found in the example of Section 2 and material of Section 3. Apart from this, the general framework, while new, merely embraces ideas which are already latent in published work. The formal links between Wald's theory and multi-Bayesianity, while more-or-less obvious, seem worth emphasizing since they do not appear to have been seriously exploited. The fact that Wald's theory is even a special case of the multi-Bayesian theory does not seem to have been previously recognized.

2. Basic components of multi-Bayesian analysis. This theory shares certain objects with Wald's. The parameter, action, and x-sample spaces, Θ, A, and X, respectively, are among these. The counterpart of the loss function is also involved.

Wald introduces a "weight function" (Wald, 1939), a non-negative function which "expresses the loss suffered by making the terminal decision..." (Wald 1950, page 8). How this loss is to be quantified is left unclear, although Wald does say (Wald, 1939, page 302) that it is based on a determination of the "relative importance of all possible errors which will depend entirely on the special purposes of his (the statistician's) investigation". This suggests that it might well be derived from the statistician's utility function and this is how it is commonly supposed to be found (c.f. Berger, 1980). We also adopt this interpretation and so regard Wald's theory as personalistic.

Let B denote a possibly infinite label set whose elements identify the Bayesians. Its elements, $\beta \in B$, may be identity tags or parameter-vectors which index β's only relevant characteristics. In the latter case, B's elements need not be distinct.

Let $u(a \mid x, \theta, \beta)$ denote β's utility for $a \in A$ given $x \in X$, $\theta \in \Theta$, $\beta \in B$. Like Berger (1980) we assume u is bounded above. In terms of u, Wald's loss function would be

$$L(a \mid x, \theta, \beta) = \max\{u(a' \mid x, \theta, \beta) \colon a' \in A\} - u(a \mid x, \theta, \beta). \tag{2.1}$$

B must choose $\delta = \delta(\cdot \mid x)$, a probability distribution on A for each $x \in X$, from the class D of all such randomized decision rules.

In Wald's theory, the performance of a procedure δ, is characterized by its risk function:

$$r(\delta \mid \theta, \beta) = \int \int L(a \mid x, \theta, \beta)\delta(da \mid x) P_{\tilde{x}}(dx \mid \theta, \beta) \tag{2.2}$$

where $P_{\tilde{x}}(\cdot \mid \theta, \beta)$ represents the sampling distribution, $\theta \in \Theta$, $\beta \in B$. The tilde in "$\tilde{x}$" is affixed to obtain the random counterpart of x. The problem of selecting δ now reduces to an examination of the various achievable risk functions. Wald's theory could just as well have been developed in terms of

$$W(\delta \mid \theta, \beta) \triangleq \int \int u(a \mid x, \theta, \beta)\delta(da \mid x) P_{\tilde{x}}(dx \mid \theta, \beta). \tag{2.3}$$

Conventional Bayesian analysis is concerned with the case $B = \{\beta\}$ and (2.3) is replaced by

$$B(\delta \mid x, \beta) \triangleq \int \int u(a \mid x, \theta, \beta)\delta(da \mid x)\pi_{\tilde{\theta}}(d\theta \mid x, \beta) \tag{2.4}$$

or

$$B(\delta \mid \beta) \triangleq \int \int B(\delta \mid x, \beta) P_{\tilde{x}}(dx \mid \beta) \tag{2.5}$$

according as x is or is not observed. In (2.4) β's posterior distribution is introduced while in (2.5) β's marginal distribution for $\tilde{x}$ appears. For simplicity, (2.5) is assumed, although the differences that arise in subsequent developments due to the choice of (2.5) instead of (2.4) are mostly formalistic.

Criteria and solution concepts. Multi-Bayesian analysis is concerned with what Bacharach calls δ's assessment profile, namely, $\beta \rightarrow B(\delta \mid \beta)$ if, as is now being assumed, x is unknown and $\beta \rightarrow B(\delta \mid x, \beta)$ otherwise (Bacharach, 1975, page 183). This profile replaces the function $\theta \rightarrow W(\delta \mid \theta, \beta)$ which is central to Wald's theory.

It is obvious that the two theories are formalistically similar, even if quite different in interpretation. The connection is, in fact, stronger. If utilities are comparable,

$$u(a \mid x, \theta, \beta) \equiv u(a \mid x, \theta) \quad \text{and} \quad P(\cdot \mid \theta, \beta) = P(\cdot \mid \theta) \tag{2.6}$$

for all θ while

$$B = \{[\theta]: \theta \in \Theta\} \tag{2.7}$$

where $[\theta]$ denotes the probability measure which is degenerate at θ, then the multi-Bayesian and Wald theories are identical.

It is not surprising then that multi-Bayesian theory shares with Wald's the need anticipated by Savage (1954, page 173) for the inclusion of randomized decision rules. Randomized rules cannot be ignored here as they are in conventional Bayesian analysis.

In general, the two theories, Wald's and multi-Bayesian, differ because utility functions are not unique. They can be specified only up to an order-preserving affine transformation because utility theory does not admit interpersonal comparisons of utility (c.f. Luce and Raiffa 1957; Jones 1980, page 180). Since only a single β is involved in the Wald-theory, this issue is irrelevant to it.

Because of the nonuniqueness of utility functions, two fundamental approaches have been taken in the development of a group decision theory. One approach, exemplified in the work of Nash (1950), admits the indeterminacy of utilities and develops solution concepts, some of which will be presented below, which do not depend on the comparison of utilities. The other approach admits such comparisons. And Savage (1954, page 172) argues that these comparisons would be possible in the case of a jury whose members, he says: "... are supposed to have common value judgements in connection with the legal matters at hand; for these are incorporated in the law as stated in the instructions of the court".

Savage (1954) introduces the multi-Bayesian decision problem into the statistical literature. Like Wald, he bases his developments on loss rather than gain, and introduces an analogue of the risk function (equation (2.2)), which he calls β's "(personal) loss":

$$L_P(\delta \mid \beta) = \max\{B(\delta' \mid \beta) : \delta' \in D\} - B(\delta \mid \beta). \tag{2.8}$$

This is the counterpart of what in the Wald framework is sometimes called "regret" (c.f. Blackwell and Girschick 1954). However, Savage declines to use that terminology on the grounds that it is "... charged with emotion ...".

By exploiting the analogy between the Wald and the multi-Bayesian problems which derives from the formal similarity of $W(\delta \mid \cdot \beta)$ (equation (2.3)) and $B(\delta \mid \cdot)$ (equations (2.4) and (2.5)), various concepts and criteria become interchangeable to an extent which is determined primarily by whether or not comparisons of utility are deemed to be possible.

Savage (1954) proposes the obvious analogue of Wald's notion of admissibility as the *group principle of admissibility.* This notion, to be referred to here as B-admissibility, is also called the *principle of* (strong) *Pareto optimality.*

Savage (1954) also defines a solution which, unlike B-admissibility, does require intercomparable utilities called the *group minimax rule.* We will call it the *B-minimax rule* and define it as the δ^* (or δ^*'s) for which

$$\max_{\beta \in B} L_P(\delta^* \mid \beta) = \min_{\delta \in D} \max_{\beta \in \mathrm{B}} L_P(\delta \mid \beta). \tag{2.9}$$

Alternatively, the B-minimax rules could be found by maximizing over D the minimum over B of $\{B(\delta \mid \beta) - c_{sa}(\beta)\}$ where

$$c_{sa}(\beta) = \max_{\delta \in D} B(\delta \mid \beta). \tag{2.10}$$

Observe that if equations (2.6) and (2.7) hold, B-admissibility and B-minimaxity, respectively, reduce to admissibility and minimaxity.

Because there is no natural origin on the range of a utility function, solution concepts like that embraced in equation (2.9) entail the creation of a benchmark, a quantity, $c(\beta)$, like that in equation (2.10) which we will call a *reference utility level* (RUL). The choice of this function is irrelevant, however, to the B-admissibility criterion.

It is also irrelevant for a solution concept which might be called *B-Bayes* (Bayes if (2.6) and (2.7) hold). Madansky (see Bacharach, 1975, page 186) exploits the Wald-multi-Bayesian connection to obtain from the work of Blackwell and Girschick (1954, page 118) certain weak principles of choice. These principles are shown to imply in their new setting if B is finite that there exists a probability distribution on B, α, (a *B-prior distribution*) such that the solutions of the group's decision problem must maximize, if x is unknown,

$$B(\delta \mid \alpha) = \int B(\delta \mid \beta)\alpha\,(d\beta). \tag{2.11}$$

The work of Harsanyi (1955, 1977) in the context of welfare economics yields (2.11) with α the uniform distribution.

All other solution concepts to be presented do require that a RUL be specified.

Various possibilities exist (Rapoport, 1970), notably:

Savage's: $c_{sa}(\beta) = \max\{B(\delta' \mid \beta): \delta' \in D\}$

Shapley's: $c_{sh}(\beta) = \beta$'s *security level* $= \min\{B(\delta' \mid \beta): \delta' \notin D\}$

Nash's: $c_N(\beta) = \beta$'s *current utility level.*

Actually, the Nash RUL (Nash, 1950) is a good deal more complicated, but it seems plausible that it would reduce to the specified quantity, the utility of the Bayesian's current assets, in the context under consideration. In general, it would represent β's utility in the event that the group was unable to find a mutually acceptable $\delta \in D$. Nash explicitly introduces the "agree to disagree" action as an allowable choice and, as well, gives each $\beta \in B$ the right to precipitate a breakdown in negotiations by insisting on this choice. In general, this conflict utility might well be less than β's current utility level because enforceable, binding threats are allowable strategies which must be specified in advance: bluffing is not allowed.

Nash's theory recognizes that an individual, β, cannot be forced by the group to agree to a choice $\delta \in D$ for which $B(\delta \mid \beta) < c_N(\beta)$. Thus the set of Nash-feasible solutions consists of those (δ's) for which $B(\delta \mid \beta) - c_N(\beta) > 0$. In contrast, the B-Bayes and B-minimax solution concepts allow the possibility that the group might choose a $\delta \in D$ which leaves individual β's with a net loss of utility.

Of the solution concepts for finite B, which are presented in the survey of Weerahandi and Zidek (1981), the most celebrated is that of Nash (1950): maximize

$$\textstyle\prod_\beta \, [B(\delta \mid \beta) - c(\beta)]_+ \tag{2.12}$$

where $c(\beta) = c_N(\beta)$ and, in general, $[x]_+ = \max\{x, 0\}$ (Nash's derivation of this product does not depend on the particular choice of $c(\cdot)$). A slight weakening of Nash's assumptions (Kalai, 1977) gives: maximize

$$P(\delta \mid \alpha) \triangleq \textstyle\prod_\beta (B(\delta \mid \beta) - c_N(\beta))_+^{\alpha(\beta)} \tag{2.13}$$

where $\alpha(\beta) > 0$ and $\sum \alpha(\beta) = 1$, but is otherwise unspecified. Evidently $\alpha(\cdot)$ is to be determined by preliminary negotiation; at least that is what the character of Kalai's proof would suggest. Shapley's solution concept agrees with Nash's in this situation (Jones, 1980) except that $c(\beta)$ is $c_{sh}(\beta)$ and not $c_N(\beta)$.

Nash's derivation of (2.12) and Kalai's of (2.13) explicitly assume that utility functions are not comparable. And this is reflected in their respective forms; the transformations

$u(a \mid x, \theta, \beta) \rightarrow a(\beta)u(a \mid \theta, x, \beta) + b(\beta)$, $a(\cdot) > 0$ will lead to exactly the same solution set. In this respect the Nash and Kalai solution concepts are generalizations of that of the conventional uni-Bayesian decision theory.

A more general class of solutions for finite B and comparable utilities,

$$D^* = \{\delta_\rho: -\infty \le \rho \le \infty\}, \tag{2.14}$$

may be obtained by maximizing

$$\{\textstyle\sum_\beta \alpha(\beta)[\Delta B(\delta \mid \beta)]^\rho\}^{1/\rho}, \tag{2.15}$$

subject to

$$\Delta B(\delta \mid \beta) > 0 \quad \text{for all} \quad \beta, \tag{2.16}$$

where $\Delta B(\delta \mid \beta) = B(\delta \mid \beta) - c(\beta)$. The cases $\rho = -\infty$, 0, and $+\infty$ are obtained in the limit.

For $\rho = 0$ (2.15) gives (2.13) (c.f. Weerahandi and Zidek, 1980). Thus, δ_0, is a Nash-Kalai rule and 0 is the only ρ-value for which the resulting solution set is independent of the scales chosen by the β's for their utilities. $\delta_{-\infty}$ is any solution which satisfies condition (2.16) and maximizes $\min\{\Delta B(\delta \mid \beta):\beta\}$. $\delta_{+\infty}$ and δ_1 are B-maximax and B-Bayes rules, respectively. $\delta_{+\infty}$ lets the group member who has potentially the most to gain be the dictator. When conditions (2.6) and (2.7) hold, the class $D^* = \{\delta_\rho: -\infty \le \rho \le \infty\}$ constitutes a set of admissible solutions to the decision problem of Wald in the case of finite Θ.

By letting $\Delta B = c_{sa} - B$, requiring $\rho \ge 1$ and minimizing (2.15), the class of solutions proposed by Yu (1973) is obtained. $\delta_{+\infty}$ is then Savage's B-minimax solution.

Other solution concepts which are given in Weerahandi and Zidek (1981) are, for brevity, not considered here. For a recent survey of related results, see White (1976).

EXAMPLE 2.1. *Multi-Bayesian estimates of the multi-normal mean.* Bayesians $\beta = i$, $i = 1, 2$ are required to estimate a vector $\boldsymbol{\theta} \in R_p$ of normal means. It will be assumed that the costs of disagreement are $c(1) = c(2) = 0$, that is, that no penalty is attached to the joint decision not to declare an estimate beyond the loss of anticipated utility. This assumption is realistic and simplifies our analysis considerably. In particular, $\Delta B = B$, since the Nash RUL is adopted. Assume x has been observed.

Denote the "agree to disagree" decision by $\hat{\theta}_0$ and assume that choosing $\hat{\theta}_0$ will lead to no change in the utility of either Bayesian, i.e. $B(\hat{\theta}_0 \mid i) = c(i) = 0$ for $i = 1, 2$.

Bayesian i is assumed to have a posterior distribution with the multivariate normal density function given by (when x has been deleted)

$$\pi(\boldsymbol{\theta} \mid i) \propto \exp[-\tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_i)^T \textstyle\sum_i^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_i)] \tag{2.17}$$

where $\sum_i$ is a specified positive definite matrix and $\boldsymbol{\theta} \in R_p$, $\boldsymbol{\theta}_i \in R_p$. As his gain-in-utility function, we take

$$u(\hat{\boldsymbol{\theta}} \mid \boldsymbol{\theta}, i) \propto \exp[-\tfrac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{W}_i^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})] \tag{2.18}$$

where W_i is a positive definite matrix of specified constants. Assume for simplicity that $W_i = \Lambda - \sum_i > 0$. Much of what follows may be generalized.

The expected gain-in-utility for Bayesian i of the estimate $\theta = \hat{\theta}$ is easily shown to be from equations (2.17) and (2.18), after a convenient rescaling and deleting x

$$B(\hat{\boldsymbol{\theta}} \mid i) = \phi\{(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}})^T \Lambda^{-1}(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}})\}$$

where $\phi(u) = \exp(-\tfrac{1}{2}u)$, $u \ge 0$. Finally,

$$B(\delta \mid i) = \int B(\hat{\boldsymbol{\theta}} \mid i)\delta\,(d\hat{\boldsymbol{\theta}}).$$

Pareto optimal estimation rules. Of fundamental interest is the set, $\mathscr{S}$, consisting of all 2-tuples, $(B(\hat{\theta} \mid 1), B(\hat{\theta} \mid 2))$ obtained by varying $\hat{\theta} \in R_p \cup \{\hat{\theta}_0\}$. The convex hull of $\mathscr{S}$, say $\overline{\mathscr{S}}$, would consist of all utility pairs which can be achieved by adopting randomized rules, i.e. $\{(B(\delta \mid 1), B(\delta \mid 2)): \delta \text{ randomized}\}$. Let $\partial \mathscr{S}$ represent $\mathscr{S}$'s boundary and $\mathscr{P}$, that portion of $\partial \mathscr{S}$ which corresponds to the elements, $\hat{\theta}$, of the class, $\mathscr{C}$, of nonrandomized rules which are B-admissible within $\mathscr{S}$. The sets $\mathscr{S}$, $\partial \mathscr{S}$, $\mathscr{P}$ and $\mathscr{C}$ will be characterized in this subsection.

For simplicity, $B(\cdot \mid i)$ is hereafter denoted by $u_i(\cdot)$.

Obviously, $\mathscr{S} \subset [0, 1]^2$ and $u_1 = 1$ is attained at $\hat{\theta} = \theta_1$ so that the corresponding u_2 is uniquely determined. However, if $u_1 = c$, $0 < c < 1$, $\hat{\theta}$ and hence u_2 is not uniquely determined; in this case, if $p = 1$, $\hat{\theta}$ may have either of two possible values, while if $p > 1$, it is merely constrained to lie on an ellipsoid. Thus, if $p = 1$, $\mathscr{S}$ is a curve, while if $p > 1$ it is a compact of $[0, 1]^2$. The precise form of $\mathscr{S}$ is of no relevance to our analysis, but our results will be more easily interpreted by referring to Figure 1.

The set of u_2-values corresponding to $u_1 = u_1(\hat{\theta}) = c$, i.e. $u_2 = u_2(\hat{\theta})$ is the cross-section of $\mathscr{S}$ at $u_1 = c$. The maximum and minimum u_2-values within this cross-section are the points of $\partial \mathscr{S}$ at $u_1 = c$. By varying c, $0 \leq c \leq 1 = u_1(\theta_1)$, these extrema generate $\partial \mathscr{S}$.

Finding the extrema of u_2 or equivalently $(\theta_2 - \hat{\theta})^T \Lambda^{-1} (\theta_2 - \hat{\theta})$ for fixed u_1 or equivalently $(\theta_1 - \hat{\theta})^T \Lambda^{-1} (\theta_1 - \hat{\theta})$ is easily accomplished using a Lagrange-multiplier argument. The result is: *$\partial \mathscr{S}$ consists of the distinguished point* $(0, 0)$ *and the utility pairs,* (u_1, u_2) *corresponding to* $\hat{\theta} = \theta_1$, $i = 1, 2$ *and the elements of*

$$\{\hat{\theta}^* = (\lambda\theta_1 + \theta_2)/(\lambda + 1): -\infty < \lambda < \infty, \lambda \neq -1\}. \tag{2.19}$$

The quantity λ which indexes the set in (2.19), is the Lagrange multiplier. It varies with u_1 and its value, as we shall see below, determines whether the u_2-extremum it determines is a maximum or a minimum.

As is easily shown, the utility pairs corresponding to the elements of (2.19), (u_1, u_2), are $u_i = \phi[\Delta_i]$, $i = 1, 2$ where $\Delta_1 = D^2(1 + \lambda)^{-2}$, $\Delta_2 = \lambda^2 \Delta_1$ and $D^2 = (\theta_1 - \theta_2)^T \Lambda^{-1} (\theta_1 - \theta_2)$.

The character of $\partial \mathscr{S}$ is now easy to determine. It is easily shown that

$$du_2/du_1 = -\lambda(u_2/u_1) \tag{2.20}$$

and

$$d^2u_2/du_1^2 = (u_2/u_1^2)(\lambda[1 + \lambda] + 2/\Delta_1') \tag{2.21}$$

where $\Delta_1' = \partial \Delta_1 / \partial \lambda = -2D^2(1 + \lambda)^{-3}$.

Observe that as $\lambda \searrow -1$, $(u_1, u_2) \to (0, 0)$. From (2.20) it follows that as λ increases from -1 to 0, u_1 and u_2 increase to $\phi[D^2]$ and 1 respectively, the values corresponding to $\hat{\theta} = \theta_2$. Then as λ continues to increase from 0 to $+\infty$, u_1 increases to 1 while u_2 decreases to $\phi[D^2]$, these latter values corresponding to the choice $\hat{\theta} = \theta_1$. As $\lambda \nearrow -1$, $(u_1, u_2) \to (0, 0)$. As λ decreases from -1 to $-\infty$, u_1 and u_2 increase to 1 and $\phi[D^2]$, respectively. Thus: *$\partial \mathscr{S}$ is a continuous parametric curve in* $[0, 1]^2$ *which consists of two branches. The upper branch increases from* $(0, 0)$ *to* $(u_1(\theta_2), u_2(\theta_2))$ *and then decreases to* $(u_1(\theta_1), u_2(\theta_1))$. *The lower branch increases from* $(0, 0)$ *to* $(u_1(\theta_1), u_2(\theta_1))$.

The Pareto optimal nonrandomized rules which are B-admissible in $\mathscr{S}$ correspond to the utility pairs in $\mathscr{P} \subset \partial \mathscr{S}$. Obviously $\mathscr{P}$ is that part of $\partial \mathscr{S}$'s upper branch which joins the utility pairs corresponding to $\hat{\theta} = \theta_1$ and $\hat{\theta} = \theta_2$. The Nash solution will be nonrandomized if the function u_2 of u_1 defined by $\mathscr{P}$, is concave. This will be the case if $d^2u_2/du_1^2 \leq 0$ on $\mathscr{P}$. Equation (2.21) implies: *$\mathscr{P}$ determines a concave function,* u_2 *of* u_1 *if and only if*

$$D^2 = (\theta_1 - \theta_2)^T \Lambda^{-1} (\theta_1 - \theta_2) \leq 4. \tag{2.22}$$

The quantity D^2 in (2.22) is a Mahalanobis distance. So the condition is intuitively natural since D^2 is a measure of the consensus between the two decision makers.

Other noteworthy features of $\mathscr{S}$ are easily derived. Let $\Delta_1 = (1 - t)^2 D^2$ and $\Delta_2 = t^2 D^2$ where $t = \lambda(1 + \lambda)^{-1}$. In terms of this new parameter, t, the lower branch of $\partial \mathscr{S}$

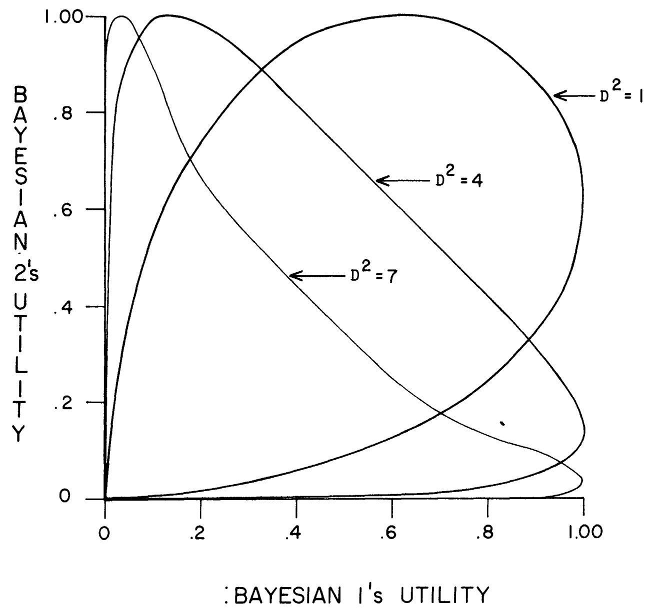

FIG 1. *Utility vectors contained within the boundary of those of nonrandomized rules for varying Mahalanobis distances,* D^2.

corresponds to $1 < t < +\infty$ and the upper, $-\infty < t < 1$ where, as before, the endpoints are achieved in the limit. The nonrandomized rules which are B-admissible in $\mathscr{S}$ correspond to utility pairs for which $0 < t < 1$. Everywhere on $\partial\mathscr{S}$, $u_1 = \exp[-\frac{1}{2}(1-t)^2D^2]$ and $u_2 = \exp[-\frac{1}{2}\ t^2D^2]$ while $du_2/du_1 = -t(1-t)^{-1} \times (u_2/u_1)$ and $d^2u_2/du_1^2 = (u_2/u_1^2) \times (1-t)^{-3}(t - t^2 - D^{-2})$, $t \neq 1, +\infty$. It follows that the lower branch of $\partial\mathscr{S}$ represents a convex function, u_2 of u_1, since $d^2u_2/du_1^2 > 0$ for $1 < t < \infty$. Similar reasoning shows that $\partial\mathscr{S}$ represents a concave function for $-\infty < t < 0$. On the Pareto boundary, as noted above, $\partial\mathscr{S}$ is concave or not according as $D^2 \leq 4$ or $D^2 > 4$. Finally, let $u = t - \frac{1}{2}$. Then $u_1 = \exp[-\frac{1}{2}(\frac{1}{2} - u)^2D^2]$ and $u_2 = \exp[-\frac{1}{2}(\frac{1}{2} + u)^2D^2]$ and the transformation $u \to -u$ leads to $u_1 \to u_2$, $u_2 \to u_1$ which implies that $\mathscr{S}$ is symmetric about the 45° radial line through the origin. This result, too, agrees with intuition. Figure 1 shows how $\mathscr{S}$ varies with the Mahalanobis distance, D^2.

Observe that each of the sets, $\mathscr{S}$, illustrated in Figure 1: (i) includes (0, 0), (ii) is symmetric about the 45° line, (iii) is below (respectively, to the left of) the line $u_2 = 1$ ($u_1 = 1$) except at their unique point of intersection which represents $\hat{\theta} = \theta_2(\theta_1)$, (iv) has a concave (respectively, convex) increasing upper (lower) boundary to the left of the point which represents $\hat{\theta} = \theta_2(\theta_1)$. As we have shown in this Section, all $\mathscr{S}$'s must have features (i) – (iv). The only qualitative difference that may exist between any two $\mathscr{S}$'s is found by inspecting their boundaries between $\hat{\theta} = \theta_1$ and $\hat{\theta} = \theta_2$. Figure 1 portrays the fact that this portion of the boundary will be a concave curve if, and only if, $D^2 \leq 4$.

Randomized decision rules. An inspection of the curve in Figure 1 for $D^2 = 7$ makes obvious the need to introduce randomized decision rules. There is a considerable divergence of preferences or (a posteriori) opinions of the two Bayesians. The best nonrandomized rule ($t = \frac{1}{2}$ in $\partial\mathscr{S}$) yields a relatively small increase (.42) in utility to either Bayesian. So a coin toss leading (approximately) to either $\hat{\theta} = \theta_1$ or $\hat{\theta} = \theta_2$ would seem preferable, for then one of the Bayesians will be well satisfied, while the other is not much worse off than if the best nonrandomized rule had been adopted. And their expected utilities would rise considerably (by .52). Thus it is mutually beneficial to co-operate and jointly adopt a randomized rule δ.

The introduction of randomized estimation rules changes the set of expected utilities to the convex hull, $\bar{\mathscr{S}}$ of $\mathscr{S}$. It should be noted that if $p > 2$ then $\mathscr{S} = \bar{\mathscr{S}}$ when $D^2 \leq 4$.

Optimal joint estimation rules. Note that among nonrandomized rules, that which maximizes $P(\delta) = P(\delta \mid \alpha)$ (see equation (2.13)), the generalized Nash product, chooses with certainty the action,

$$\hat{\boldsymbol{\theta}} = \bar{\boldsymbol{\theta}}_\alpha = \alpha_1 \boldsymbol{\theta}_1 + \alpha_2 \boldsymbol{\theta}_2 \tag{2.23}$$

which is merely a weighted average of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. However, as noted above, this will be an unsatisfactory choice if the $\boldsymbol{\theta}_i$ are widely separated, $i = 1, 2$. Thus we seek those δ^*'s for which $P(\delta^*) = \sup\{P(\delta): \delta \text{ randomized}\}$.

Recall that since $c_N(\beta) \equiv 0$ in this Section, maximizing P amounts to maximizing the function $(u_1, u_2) \longrightarrow u_1^{\alpha_1} u_2^{\alpha_2}$ over $\bar{\mathscr{S}}$, the convex hull of $\mathscr{S}$.

First consider the case where $\alpha_i = ½$, $i = 1, 2$. Obviously, $\bar{\mathscr{S}}$, the convex hull of the set of achievable utilities is symmetric, like those of the sets illustrated in Figure 1. Consequently, the mid-point of $\mathscr{S}$'s northeast boundary represents the optimal estimation rule. Either this rule is nonrandomized, i.e. the Bayesians accept the joint estimate $\bar{\boldsymbol{\theta}} = (\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2)/2$, or else it is a two-point distribution according as the boundary of $\mathscr{S}$ does or does not coincide with that of $\bar{\mathscr{S}}$. Whether or not these boundaries coincide depends on whether or not $\mathscr{S}$'s northeast boundary is convex. Thus: *If $\alpha_1 = \alpha_2 = ½$, a necessary and sufficient condition for the nonrandomized rule, choose $\bar{\boldsymbol{\theta}}$, to be optimal is that*

$$(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)^T \Lambda^{-1} (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \leq 4. \tag{2.24}$$

It remains for us to evaluate the optimal randomized rule which obtains when condition (2.24) fails. Our earlier discussion implies that this rule's support must consist of estimates of the form $\theta[t] = t\boldsymbol{\theta}_1 + (1 - t)\boldsymbol{\theta}_2$ for $t \in [0, 1]$. As well, $(u_1(\hat{\theta}[t]), u_2(\hat{\theta}[t]))$ is the reflection with respect to the line, $u_1 = u_2$, of $(u_1(\hat{\theta}[1 - t]), u_2(\hat{\theta}[1 - t]))$, i.e. $u_1(\hat{\theta}[t]) = u_2(\hat{\theta}[1 - t])$ and hence $u_1(\hat{\theta}[1 - t]) = u_2(\hat{\theta}[t])$.

The symmetry of the problem in this special case and the considerations of the previous paragraph imply that the optimal rule chooses with equal probability the estimates $\hat{\theta}[t^*]$ and $\hat{\theta}[1 - t^*]$ where t^* maximizes the function, $g(t) = u_1(\hat{\theta}[t]) + u_2(\hat{\theta}[t])$. Maximizing g is equivalent to maximizing the function f defined by $f(s) = \cosh(s) \exp[-2s^2 D^{-2}]$ where $D^2 = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)^T \Lambda^{-1} (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)$, and $s = D^2(t - ½)/2$. It is straightforward to show that if $D^2 \leq 4$, f has a unique maximum at $s = 0$. At the same time, if $D^2 > 4$, f has a local minimum at 0 and global maxima at $\bar{s}$ and $-\bar{s}$, where $\bar{s}$ is the positive root of the equation

$$\tanh(\bar{s})(\bar{s})^{-1} = 4D^{-2}. \tag{2.25}$$

Thus, in this latter case, the optimal rule chooses $(\bar{\boldsymbol{\theta}} + \boldsymbol{\delta})/2$ or $(\bar{\boldsymbol{\theta}} - \boldsymbol{\delta})/2$ with equal probability where $\boldsymbol{\delta} = 2\bar{s}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)/D^2$. This completes the analysis when $\alpha_1 = \alpha_2 = ½$.

Now drop the assumption $\alpha_1 = \alpha_2 = ½$. It is no longer true that the contours of $u_1^{\alpha_1} u_2^{\alpha_2}$, which is to be maximized over $\bar{\mathscr{S}}$, are symmetric with respect to the line $u_1 = u_2$. Consequently this case is quite complicated. It may well happen (for example, when $\alpha_1 = 1$, $\alpha_2 = 0$ and we are faced with just a conventional uni-Bayesian analysis) that the optimal rule is non-randomized, even when the north-east boundary of $\mathscr{S}$ is not concave. The following result states exactly when, in fact, this does happen: *the optimal nonrandomized rule, i.e. the rule which chooses $\bar{\boldsymbol{\theta}}_\alpha = \alpha_1 \boldsymbol{\theta}_1 + \alpha_2 \boldsymbol{\theta}_2$ with certainty, is globally optimal if, and only if, either*

(i) $D^2 = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)^T \Lambda^{-1} (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \leq 4$ *or*

(ii) $D^2 > 4$ *and* $\bar{s} \leq (D^2/2)(\text{Max}\{\alpha_1, \alpha_2\} - ½)$ *where $\bar{s}$ is the positive root of equation* (2.25).

We prove the sufficiency of the condition. Its necessity will then be obvious from this proof.

If $D^2 \leq 4$, the northeastern boundary of $\mathscr{S}$ is concave. Thus the best nonrandomized rule is globally optimal since the Pareto optimal rules, in this case, are nonrandomized rules.

Now suppose $D^2 > 4$. From equation (2.21) it follows that the portions of $\mathscr{S}$s boundary corresponding to $\hat{\theta}[t] = t\theta_1 + (1 - t)\theta_2$ for $0 \leq t \leq 1/2 - (1 - 4/D^2)^{1/2}/2$ and $1/2 + (1 - 4/D^2)^{1/2}/2 \leq t \leq 1$ are concave. Equivalently, in terms of $s = D^2(t - 1/2)/2$, these segments of the boundary correspond to $-\alpha \leq s \leq -\alpha(1 - \alpha^{-1})^{1/2}$ and $\alpha(1 - \alpha^{-1})^{1/2} \leq s \leq \alpha$, respectively, where $\alpha = D^2/4$. It is not hard to show and, in any case, it is geometrically obvious that $\bar{s}$, the positive root of (2.25) is contained in the latter segment. But $\bar{s}$ and $-\bar{s}$ correspond to the points of tangency on $\mathscr{S}$'s northeastern boundary, of the most extreme of the translates of the line $u_2 = -u_1$ which remain in contact with $\mathscr{S}$. It therefore follows that the best nonrandomized rule is optimal if (and only if) it corresponds to a value of s, say s^*, for which either $-\alpha \leq s^* \leq -\bar{s}$ or $\bar{s} \leq s^* \leq \alpha$. But $s^* = (D^2/2)(\alpha_1 - 1/2)$. Thus the latter condition is equivalent to $\bar{s} \leq (D^2/2)(\text{Max}\{\alpha_1, \alpha_2\} - 1/2)$ since $-\alpha \leq s^* \leq \alpha$ is always true. This completes the argument.

The proof of the last result indicates how to locate the optimal randomized rule in the event that its necessary and sufficient condition is violated. This rule will be represented by a point on the line segment joining the points of tangency on $\mathscr{S}$'s north-eastern boundary, which corresponds to $s = \bar{s}$ and $s = -\bar{s}$ where $\bar{s}$ is the solution of equation (2.25). Alternatively, since $s = (D^2/2)(t - 1/2)$, these points of tangency correspond to $t = \bar{t}$ and $t = 1 - \bar{t}$ where $\bar{t} = 1/2 + 2\bar{s}D^{-2}$. The line joining these points of tangency has the equation $u_1 + u_2 = u_1(\hat{\theta}[\bar{t}]) + u_2(\hat{\theta}(\bar{t}])$ as a consequence of the fact that $u_1(\hat{\theta}[t]) = u_2(\hat{\theta}[1 - t])$ for all t. This line is also tangent to a contour of the function $h(u_1, u_2) = u_1^{\alpha_1} u_2^{\alpha_2}$ and the point of tangency, say (u_1^0, u_2^0) is precisely the point at which h is maximized over $\bar{\mathscr{S}}$ (the convex hull of $\mathscr{S}$). But the tangent line to the contour of h which passes through the point (u_1^0, u_2^0) may alternatively be expressed as $\alpha_1(u_1^0)^{-1}u_1 + \alpha_2(u_2^0)^{-1} u_2 = 1$. We now have two equations for the same line and so deduce that $(u_1^0, u_2^0) = (\alpha_1, \alpha_2)\ (u_1(\hat{\theta}[\bar{t}]) + u_2(\hat{\theta}[\bar{t}]))$. Since this point is on the line segment joining the points of tangency which are described above and correspond to $t = \bar{t}$ and $t = 1 - \bar{t}$, it must be a weighted average of these points. Thus, there exists a value, say $\bar{\alpha} \in [0, 1]$, such that $\bar{\alpha}u_1(\hat{\theta}[\bar{t}]) + (1 - \bar{\alpha})u_1(\hat{\theta}[1 - \bar{t}]) = \alpha_1\{u_1(\hat{\theta}[\bar{t}]) + u_2(\hat{\theta}[\bar{t}])\}$. In fact, as is easily shown, this value is given by

$$\bar{\alpha} = \{\alpha_1 - \alpha_2 \exp[-(\bar{t} - 1/2)D^2]\}/\{1 - \exp[-(\bar{t} - 1/2)D^2]\}. \tag{2.26}$$

We may summarize our conclusions. If (and only if) the condition (i) or (ii) given above is true, the best nonrandomized rule, "choose $\bar{\theta} = \alpha_1\theta_1 + \alpha_2\theta_2$ with certainty", is optimal. If these conditions fail to hold, the optimal rule is randomized and consists of choosing the estimates $(1/2 + 2\bar{s}D^{-2})\theta_1 + (1/2 - 2\bar{s}D^{-2})\theta_2$ and $(1/2 - 2\bar{s}D^{-2})\theta_1 + (1/2 + 2\bar{s}D^{-2})\theta_2$ with probabilities $\bar{\alpha}$ and $1 - \bar{\alpha}$, respectively, where $\bar{s}$ is the positive or negative root of equation (2.25) according as $\bar{\alpha} > 1/2$ or $\bar{\alpha} < 1/2$.

3. Subsampling assessments. Other means of analyzing any proposed group decision rule are introduced in this Section.

The group, B, is still required to choose a mutually acceptable δ. However, we now suppose that δ will be implemented only by a subgroup, $s \subset B$. Only s's members will derive any change-in-utility from the use of this rule. The remainder will simply retain their current utility level, $c_N(\beta)$, $\beta \in B - s$. It is known only that $s \in S$, a specified collection of subsets of B but not which of these is to be chosen.

Since $\beta \in B$ is potentially a member of s, he retains his self-interest in the group's ultimate choice of δ. So this is a variant of the problem considered in the previous subsection, and is the subject of the present one.

There are two different possible points of view about s. It is fixed by some unknown process and is itself unknown or it is chosen at random by some well-defined procedure. If s is fixed and $S = \{\{\beta\}: \beta \in B\}$, the group's choice of δ would have to depend on a comparative analysis of its assessment profile, $B(\delta \mid \cdot)$; the problem of Section 2 reappears. Nothing is known about the general version of this problem with arbitrary S. The case of random s will be discussed below.

EXAMPLE 3.1. Here $\tilde{x}$, $\tilde{\theta}$ and β are $p \times 1$ vectors with

$$\tilde{x} \mid \theta, \beta \sim N(\theta, \Sigma), \tilde{\theta} \mid \beta \sim N(\beta, \tau) \text{ and } \beta \in R_p = B. \text{ Also } S = \{\{\beta\}: \beta \in B\}.$$

In general, β might occur with multiplicity greater than one and, in a case like this where B is infinite, this would need to be accounted for by means of a distribution on β. This issue will not arise, however, in the present analysis.

Assume, for simplicity, a quadratic utility function

$$u(\hat{\theta} \mid x, \theta, \beta) = C_0(\beta) - D(\beta)(\hat{\theta} - \theta)^T Q(\hat{\theta} - \theta)$$

where $D > 0$ and $\hat{\theta} \in A = R_p$. Then, as is easily shown, δ's assessment profile is

$$B(\delta \mid \beta) = C_1(\beta) - D(\beta)E(\hat{\beta} - \beta)^T R(\hat{\beta} - \beta),$$

the expectation here being over the marginal distribution of $\tilde{x}$ given β. Here $C_1(\cdot)$ is a certain function whose exact form is of no relevance. Also, $R = \zeta' Q \zeta$, $\hat{\beta} = \zeta^{-1}\{\hat{\theta} - \bar{\zeta}x\}$, $\zeta = \Sigma(\Sigma + \tau)^{-1}$ and $\bar{\zeta} = I - \zeta$,

If β had been specified, then the best possible choice of $\hat{\theta}$ would be the Bayesian estimator given by $\hat{\beta} = \beta$, i.e. $\hat{\theta} = \zeta\beta + \bar{\xi}x$.

But we are supposing that $s = \{\beta\}$ has not been selected and that everyone, $\beta \in B$, is eligible to be chosen. A mutually acceptable $\hat{\theta} = \hat{\theta}(x)$ is to be found before this choice is made and it cannot therefore depend on s.

Since, as is easily shown, $x \mid \beta \sim N(\beta, \Sigma + \tau)$ a naive candidate for $\hat{\beta}$ is $\hat{\beta} = \hat{\beta}_0(x) = x$, this being a B-unbiased rule. This choice then entails $\hat{\theta} = \hat{\theta}_0(x) = x$ as well.

If $p \leq 2$, a result of Gatonis (1981; Theorem 2.1) obtained in a different setting implies that $\hat{\theta}_0$ is Pareto-optimal, i.e. B-admissible. However, if $p > 2$ then $\hat{\theta}_0$ is B-inadmissible and there exist estimators $\hat{\theta}_1$, whose assessment profiles are uniformly larger for all β than that of $\hat{\theta}_0$. Any such $\hat{\theta}_1$ would be jointly preferred to $\hat{\theta}_0$. A general class of potential alternatives to $\hat{\theta}_0$ is given by Thisted (1976) (see also Brown and Zidek, 1980).

Now suppose $s \in S$ is chosen at random, according to a sampling design, $p = \{p(s): s \in S\}$, $p \geq 0$. Let us assume for the remainder of this Section that B is finite and so require that $\Sigma p(s) = 1$. The case of an infinite B remains unexplored.

Each $\beta \in B$ is concerned with $\pi(\beta) = \sum_{\beta \in s} p(s)$, his inclusion probability, $0 \leq \pi(\beta) \leq 1$. Assume $n = \Sigma\pi(\beta)$, the expected sample size, is fixed. His expected utility is, in any case, $\pi(\beta)B(\delta \mid \beta) + \bar{\pi}(\beta)c_N(\beta)$ where $\bar{\pi} = 1 - \pi$. Thus the problem of choosing δ again reduces to that considered in the last section, albeit with a different assessment profile.

The solution concepts presented in the last section not only imply a class of optimal choices for δ for fixed p but imply an optimal design as well. This optimal design yields some insight into the nature of the solution criteria and will now be computed in various cases to conclude this subsection.

To unify this discussion the analysis will focus on the criterion function given in equation (2.15). For simplicity the seemingly realistic choice $c(\beta) = c_N(\beta)$ will be adopted. Thus the appropriate utility increments are

$$\pi(\beta) \ \Delta B(\delta \mid \beta) \tag{3.1}$$

where $\Delta B(\delta \mid \beta) = B(\delta \mid \beta) - c_N(\beta)$. The problem under consideration then reduces to maximizing

$$\{\Sigma\alpha(\beta)\pi^\rho(\beta)[\Delta B(\delta \mid \beta)]^\rho\}^{1/\rho}, \quad -\infty \leq \rho \leq \infty, \tag{3.2}$$

subject to Nash-feasibility:

$$\Delta B(\delta \mid \beta) > 0 \quad \text{for all} \quad \beta \in B, \tag{3.3}$$

with $\rho = \pm \infty$ and $\rho = 0$ defined in the limit.

Consider $f(\pi) \triangleq [g(\pi)]^{1/\rho}$ where $g(\pi) = \Sigma\lambda(\beta)\pi^{\rho}(\beta)$. To maximize f on $\{\pi\colon 0 \le \pi(\beta) \le 1, \Sigma\pi(\beta) = n\}$ entails maximizing or minimizing g according as $\rho > 0$ or $\rho < 0$. But g is convex, concave or convex according as $\rho < 0$, $0 < \rho < 1$ or $1 \le \rho$. Since for $0 < \rho$, f is increasing in $\pi(\beta)$ for each β, it follows that for $1 \le \rho$, f is maximized at $\pi(\beta) = 1$ for $\beta \in s \subset B$, where $|s| = n$ and $\Sigma_s\lambda(\beta) = \max\{\Sigma\lambda_{s'}(\beta)\colon |s'| = n, s' \subset B\}$. If, however, $\rho < 1$ the situation is more complicated and Kuhn-Tucker optimization methods would need to be employed to find the optimum (c.f. Panik, 1976).

This reasoning translates into conclusions about the maximization of equation (3.2) over π, and δ. In particular, it reveals that if $\rho \ge 1$, the optimum choices of (π, δ), (π^*, δ^*), satisfy the requirements $\pi^*(\beta) = 1$, for $\beta \in s^*$ while $s^* \subset B$, $|s^*| = n$ and δ^* jointly maximize $\Sigma_s\alpha(\beta)[\Delta B(\delta \mid \beta)]^{\rho}$. For $\rho = +\infty$ this means an $s = s^* \subset B$, $|s^*| = n$ which contains the β^* for which $\alpha(\beta)\Delta B(\delta \mid \beta)$ is jointly maximized in β and δ. For $\rho < 1$, little can be said except that the problems of choosing the optimal p and δ are interlocked (except when $\rho = 0$, a case we treat separately, below) and that if $\rho < 0$, $\pi(\beta) > 0$ is required for all β.

This analysis reveals a striking feature of the Madansky-Bacharach-Harsanyi-Blackwell-Girshick (B-Bayesian) criterion which corresponds to the case $\rho = 1$ (which is shared with solutions for $\rho > 1$), under this subsampling analysis. The group of N Bayesians would defer the choice of δ and subsequent action to that subgroup of size n who had jointly the greatest possible expected personal gains of any such subgroup. That is, $N - n$ of these Bayesians would voluntarily withdraw their eligibility for membership on the decision making committee of size n.

The Kalai-Nash (i.e. optimal) solution ($\rho = 0$) is quite different from the B-Bayesian results which were just described. In this case (3.2), the quantity to be maximized in searching for the optimal solution, becomes

$$\prod [\pi(\beta)]^{\alpha(\beta)} \prod [\Delta B(\delta \mid \beta)]^{\alpha(\beta)}. \tag{3.4}$$

Therefore, the optimal choices of π and δ are made, each without regard to the other. The optimal δ's under subsampling are the same as those in the original problem treated in Section 2 without subsampling. And the optimal π, that which maximizes (3.4), is easily evaluated by the Kuhn-Tucker argument.

We conclude this Section with a derivation of the optimal π. Assume without loss of generality, that $\alpha(\beta) > 0$ for all β. Relabel these α's so that $\alpha_1 \le \cdots \le \alpha_N$ and then the π's to maintain their correspondence with the α's: $\pi_1, \cdots, \pi_N$.

Observe that

$$F(\pi) \triangleq \log \Pi(\pi_i)^{\alpha_i} \tag{3.5}$$

is a strictly concave function. Therefore, it has a unique maximum over its convex domain, $\{\pi\colon 0 \le \pi_i \le 1, \Sigma\pi_i = n\}$. Furthermore, any local maximum will be a global maximum.

Clearly, none of the constraints, $\pi_i = 0$ will be binding. So either the constraints $\Sigma\pi_i = n$ alone will be binding, in which case by Lagrange's argument the optimal values are

$$\pi_i = n\alpha_i, \quad i = 1, \cdots, N, \tag{3.6}$$

or

$$\pi_i = 1 \text{ on some subset, } J \subset B, \text{ while } \pi_i < 1 \text{ on } B - J.$$

In order that the latter prevail, F must be maximized at this π, subject to $\Sigma\pi_i = n$, $\pi = 1$ on J in which case (Lagrange's argument again)

$$\pi_i = (n - |J|)\alpha_i / \Sigma_{B-J}\alpha_j < 1 \tag{3.7}$$

on $B - J$. But for (3.7) to be a global maximum it must be a local maximum. This means that F must decline when π is perturbed to $\pi + h$ in any feasible direction, h, on $B - J$ etc., where $h = (h_1, \cdots, h_N)$. Such a direction will be feasible if, and only if, h is small and

$$h_i \leq 0 \text{ on } J \text{ and } \Sigma h = 0. \tag{3.8}$$

Thus π will be a local maximum if, and only if, for all feasible h, $F(\pi + h) \leq F(\pi)$, i.e. $0 \geq \Sigma_J \alpha_i h_i + \Sigma_{B-J}\, \alpha_i h_i/\pi_i = \Sigma_J \alpha_i h_i + (n - |J|)^{-1} \Sigma_{B-J} \alpha_j \times \Sigma_{B-J} h_i = \Sigma_J h_i(\alpha_i - (n - |J|)^{-1} \Sigma_{B-J} \alpha_j)$. Thus the maximal value of the objective function in (3.4) is attained at

$$\begin{aligned} \pi_i &= 1, & i &\in J \\ \pi_i &= (n - |J|)\alpha_i/\Sigma_{B-J}\alpha_j, & i &\notin J \end{aligned} \tag{3.9}$$

if, and only if, $[n - |J|]\alpha_i/\Sigma_{B-J}\alpha_j > 1$ or ≤ 1 according as $i \in J$ or $i \in B - J$.

To reduce this result to a more explicit form, observe that (3.9) entails $\min\{\alpha_j : j \in J\} \geq \max\{\alpha_j : j \in B - J\}$. So without loss of generality $J = \{N - |J| + 1, \cdots, N\}$.

Let $L_k = (n - k)\alpha_{N-k} - \alpha_1 - \cdots - \alpha_{N-k}$, $k = 0, \cdots, n$. Since $L_n < 0$, either $L_k \leq 0$ for $0 \leq k \leq n$ or there exists $k = |J|$ such that

$$L_{|J|-1} > 0 \quad \text{and} \quad L_{|J|} \leq 0. \tag{3.10}$$

But $L_{|J|-1} > 0$ entails $L_k > 0$, $k < |J|$. Thus, in the latter case, there exists a unique $|J|$, $1 \leq |J| \leq n$ for which (3.10) holds. The global optimum is, therefore,

$$\begin{aligned} \pi_i &= 1,\ i = N - |J| + 1, \cdots, N \\ &= (n - |J|)\alpha_i/(\alpha_1 + \cdots + \alpha_{N-|J|}), \text{ otherwise.} \end{aligned}$$

These results are summarized in

THEOREM 3.1. *Let K be 0 or the largest integer, $1 \leq K \leq n$ for which $(n - k)\alpha_{N-K} - \alpha_1 - \cdots - \alpha_{N-K} > 0$ if such an integer exists. Then*

$$\pi_i = \begin{cases} 1, & N - K + 1 \leq i \leq N \\ (n - K)\alpha_i/(\alpha_1 + \cdots + \alpha_{N-K}), & i < N - K + 1 \end{cases} \tag{3.11}$$

maximizes (3.4).

EXAMPLE 3.2. Suppose $N = 19$, $n = 10$ and $\alpha = (0.01, 0.01, \cdots, 0.01, 0.05, 0.20, 0.25, 0.35)$. Then $L_0 = 10\ (0.35) - 1 = 2.50$, $L_1 = 9\ (0.25) - (1 - 0.35) = 1.60$, $L_2 = 1.20$, $L_3 = 0.15$, $L_4 = -0.09$, $\cdots$. So $|J| = 4$, and $\pi_i = 1$ for $i \in J$, while $\pi_i = 0.4$, $i \in B - J$ for an expected sample size of $n = 10$. □

Theorem 3.1 and the formula (3.11) establish that when certain members of B are designated, by the large size of their α's as having a particularly important role in the analysis, these must be included in the sample. As for the remainder, there is at least some chance each will be included. However, utilities are assumed to be incomparable, so the importance of individuals is measured only in terms of α and not, as in the B-Bayesian case, in terms of expected utility gains as well.

4. Superpopulation assessments. Suppose B is a subset of $\mathscr{B}$, a set which will be called a superpopulation. It is assumed that $\delta = \delta(\cdot \mid x, B)$ has been specified. This Section is concerned with the evaluation of δ. It is, implicitly, about the evaluation of solution concepts like these proposed in Section 2, which are invoked in obtaining δ. Explicitly, it is concerned with the ability of groups, B, to produce effective decision rules.

In Wald's and, more generally, the frequency theory of statistics, δ's performance is assessed by considering not only what it does with these data, x, but, as well, what it would do with every other data set $\tilde{x}$ which might have been obtained, but was not. The seemingly natural counterpart of this principle here would require that we look not only at what this procedure would do for this group B but, as well, what it would do for every other group, $\tilde{B}$, that might have used it to analyze *these* data but did not. Dawid (1982) proposes a related but different notion of objectivity.

When $\mathscr{B}$ is finite $\tilde{B}$, pre-sampling (random) counterpart, has a sampling distribution which is conditional on $\mathscr{B}$ and, in turn, on θ. The existence of a B-sampling distribution may be assumed, even when $\mathscr{B}$ is infinite, as is done in the theory of survey sampling (c.f. Hajek, 1981). For these reasons and others, priors like those indexed by the elements of B may be thought of as observables and a likelihood, given these quantities and the data, may be constructed. This approach is at the basis of several recent publications (Morris, 1977; Lindley, Tversky and Brown, 1979; Steele and Zidek, 1980; French, 1980; French, 1981, and Lindley, 1982).

Choosing an assessment criterion poses a problem. In general, there cannot exist an objective, i.e. group utility function, as Arrow's theorem shows (Arrow, 1966; see also Bacharach, 1975). Such a criterion may exist in individual cases (see Savage, 1954, page 172 and Steele and Zidek, 1980) or it may be subjective, the utility function of a member of the superpopulation, the "super-Bayesian", say, or the "investigator" in the terminology of Lindley, Tversky and Brown (1979).

With a criterion selected and $(\tilde{x}, \tilde{B})$ treated as the observable, we regain the classical decision problem which may be approached via the Wald, super-Bayesian, or even multi-super-Bayesian approach, one step up on the staircase of an infinite regress. At this step, notions like S-admissibility, and S-minimaxity, for example, become relevant. We conclude with an example which illustrates the kinds of problems that arise.

EXAMPLE 3.1 (continued). To the other assumptions we add these: $C_0(\beta) \equiv C_0$, $D(\beta) \equiv D$ where C_0 and D are constants, $D > 0$, and $\tilde{\beta} \mid \theta \sim N(\theta, \Gamma)$ independently of $\tilde{x} \mid \theta \sim N(\theta, \Sigma)$. The Bayesian solution concept led to the rule $t(x \mid \beta) = \zeta\beta + \bar{\zeta}x$ in earlier analysis. The efficacy of this procedure is the issue of interest here.

Because $t(\tilde{x} \mid \tilde{\beta}) \mid \theta \sim N(\theta, \Delta)$ where $\Delta = \zeta'\Gamma\zeta + \bar{\zeta}'\Sigma\bar{\zeta}$, this problem is easily reduced to canonical form, and so may be analyzed by the methods presented earlier with this example. The results would indicate that t is S-admissible if $p \leq 2$ and S-inadmissible if $2 < p$.

The qualitative implications of this mathematical result are somewhat surprising. If $p > 2$ and $\hat{\theta} = \hat{\theta}(\tilde{x} \mid \tilde{\beta})$ needed to be specified before the arrival of $\tilde{x}$ and $\tilde{\beta}$, then the choice $\hat{\theta} = t$ would not be acceptable. This would prove perplexing to an executive who planned to employ a Bayesian consultant. The super-Bayesian would use his S-Bayes rule, not t, in any case.

Before leaving this example, it is worth noting that an extended version of the usual notion of equivariance obtains. If, for example, $\Gamma = \sigma^2 I = \tau$ and $\Sigma = I$, and the resulting value of ζ is estimated by $\hat{\zeta} = [(p-2)/\| x - \beta \|^2]I$, the James-Stein estimator is obtained: $\hat{t}(x \mid \beta) = \hat{\zeta}\beta + \hat{\bar{\zeta}}x$. It is an S-affine equivalent rule under the transformations $x \to \alpha x + b$ and $\beta \to \alpha \beta + b$ where α is orthogonal and $b \in R_p$. So the James-Stein estimator is translation equivariant in this extended sense, even though it is not in the more familiar setting.

5. Discussion. Among the various solution concepts described in Section 2, that of Nash or Kalai would seem most acceptable. Its validity does not require a comparison of utilities. In consequence it, unlike the other concepts discussed in this paper, inherits the property of the uni-Bayesian concept that its solutions remain invariant under positive affine transformations of each of the utilities involved. This equilibrium solution is implied by extremely weak assumptions and so, in any case, is normative (c.f. Weerahandi and Zidek, 1981). However, it is not known precisely how the required $\{\alpha(\beta)\}$ should be chosen. The results of Section 3 give some basis for this choice; $\{\alpha(\beta)\}$ is just the selection (inclusion) probability of Bayesian β under the (Kalai-Nash) optimal sampling scheme for selecting a random subcommittee of size 1 to implement the optimal (Kalai-Nash) procedure.

The Kalai-Nash rules, like the maximum rules, will often be randomized. As Weerahandi and Zidek (1981) point out, this is quite natural; it is a reflection of a lack of consensus. It might be expected that by increasing the amount of data, the consensus necessary to yield a nonrandomized rule would be achieved. This is not always the case. Professor L. D. Brown communicated to the second author an example where, in fact, opinions diverge as the amount of data increases.

There are a number of important differences between the multi-Bayesian and uni-Bayesian theory. The need to explicitly introduce a RUL is one. The requirement of Nash feasibility prevents the group from "ganging up" on one of its members and jointly increasing their utility by reducing his to something less than his current level. Of course, even in uni-Bayesian theory the RUL is present, but there its role may be left implicit.

There is an arbitrariness in the choice of a reference utility level (RUL). However, that of Nash which in a statistical setting, would seem to reduce to $c_N(\beta)$, the current utility level, seems natural.

With this choice of RUL, it is shown in Section 3 that the Kalai-Nash solution has the surprising property that it remains the group choice even under the random subsampling scheme introduced there.

The potential value of the model of Section 4 is unclear. There is evidence that people have difficulty making coherent judgments under uncertainty (c.f. Tversky and Kahneman, 1974). And there seems to be little empirical evidence to suggest reasonable models for the $\tilde{B}$-sampling distribution. It has not yet been applied to make a systematic evaluation of the solution concepts presented in Section 2.

Acknowledgements. This paper combines the manuscripts by Weerahandi and Zidek (1981) and Zidek (1982). The former was presented at a symposium held in Waterloo, Canada, in August 1981, to mark the sixty-fifth birthday of Professor George Barnard. We are grateful for that opportunity to present some of these ideas, and we benefited from the discussion that took place there. We owe much to referees for their comments. The penultimate version of this manuscript was prepared while the second author was on leave at the National Research Institute for the Mathematical Sciences of the CSIR, Pretoria, South Africa, which generously provided facilities. This leave of absence was made possible by a Fellowship from the Social Science and Humanities Research Council of Canada. The research itself was supported by a grant from the Natural Science and Engineering Research Council of Canada.

REFERENCES

ARROW, K. J. (1966). *Social Choice and Individual Values.* Second Edition. Wiley, New York.

BACHARACH, M. (1975). Group differences in the face of differences of opinion. *Management Science* **22** 182–191.

BERGER, J. O. (1980). *Statistical Decision Theory.* Springer, New York.

BLACKWELL, D. H. and GIRSCHICK, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

BROWN, P. J. and ZIDEK, J. V. (1980). Adaptive multivariate ridge regression. *Ann. Statist.* **8** 64–74.

DAWID, A. P. (1982). Intersubjective statistical models. In *Exchangeability in Probability and Statistics* (eds. G. Koch and F. Spizzichino). North Holland, Amsterdam, 217–232.

DICKEY, J. and FREEMAN, P. (1975). Population-distributed personal probabilities. *J. Amer. Statist. Assoc.* **70** 362–364.

FISHER, L. (1972). The behavior of Bayesians in bunches. *Amer. Statist.* **26** No. 5, 19–20.

FRENCH, S. (1980). Updating of belief in the light of someone else's opinion. *J. Roy. Statist. Soc. Ser. A* **143** 43–48.

FRENCH, S. (1981). Consensus of opinion. *European J. Oper. Res.* **7** 332–340.

GATSONIS, C. A. (1981). Estimation of the posterior density of a location parameter. Ph.D. Thesis, Cornell University.

HAJEK, J. (1981). *Sampling from a finite population.* Dekker, New York.

HARSANYI, J. C. (1977). *Rational Behavior and Bargaining Equilibrium in Games and Social Situations.* Cambridge University Press.

HARSANYI, J. C. (1955). Cardinal welfare, individual ethics and interpersonal comparisons of utility. *J. Political Economy* **63** 309–321.

JONES, A. J. (1980). *Game Theory: Mathematical Models of Conflict.* Wiley, Toronto.

KALAI, E. (1977). Nonsymmetric Nash solutions and replications of 2-person bargaining. *Int. J. Game Theory* **6** 129–133.

LINDLEY, D. V. (1982). The improvement of probability judgements. *J. Roy. Statist. Soc. Ser. A* **145** 117–126.

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

LUCE, D. and RAIFFA, H. (1957). *Games and Decisions.* Wiley, New York.

MORRIS, P. A. (1977). Combining expert opinion: a Bayesian approach. *Manag. Sci.* **23** 679–693.

NASH, J. F., JR. (1950). The bargaining problem. *Econometrica* **18** 155–162.

PANIK, M. J. (1976). *Classical Optimization: Foundations and Extensions.* American Elsevier, New York.

RAPOPORT, A. (1970). *N-Person Game Theory: Concepts and Applications.* University of Michigan Press, Ann Arbor.

SAVAGE, L. J. (1954). *The Foundations of Statistics.* Wiley, New York.

STEELE, J. M. and ZIDEK, J. V. (1980). Optimal strategies for second guessers. *J. Amer. Statist. Assoc.* **75** 596–601.

THISTED, R. A. (1976). Ridge regression, minimax estimation, and empirical Bayes methods. Tech. Rep. 28, Statistics Department, Stanford University.

TVERSKY, A. and KAHNEMAN, D. (1974). Judgement under uncertainty: heuristics and biases. *Science* **135** 1124–1131.

WALD, A. (1950). *Statistical Decision Function.* Wiley, New York.

WALD, A. (1939). Contributions to the theory of statistical estimation and testing hypothesis. *Ann. Math. Statist.* **10** 299–326.

WEERAHANDI, S. and ZIDEK, J. V. (1981). Multi-Bayesian statistical decision theory. *J. Roy. Statist. Soc. Ser. A* **144** 85–93.

WEERAHANDI, S. and ZIDEK, J. V. (1980). A characterization of the general mean. *Can. J. Statist.* **7** 83–90.

WEERAHANDI, S. and ZIDEK, J. V. (1981). Bi-Bayes estimates of the multinormal mean. Tech. Rep. 81-9, Instit. Applied Math. Statist., University of British Columbia.

WHITE, D. J. (1976). *Fundamentals of Decision Theory.* North Holland, New York.

YU, P. L. (1973). A class of solutions for group decision problem. *Management Science Ser. B* **19** 936–446.

ZIDEK, J. V. (1982). Aspects of multi-Bayesian theory. Tech. Rep. 82-11. Instit. Applied Math. Statist., University of British Columbia.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SRI JAYEWARDENAPURA
NUGEGODA, SRI LANKA

DEPARTMENT OF STATISTICS GN-22
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195