MULTI-BAYESIAN ESTIMATION THEORY

D.J. de Waal, P.C.N. Groenewald,
J.M. van Zyl, J.V. Zidek

Technical Report No. 04

July 1984

MULTI-BAYESIAN ESTIMATION THEORY

D. J. de Waal, P. C. N. Groenewald, J. M. van Zyl, J. V. Zidek

Received: Revised version: July 19, 1984

Abstract. A theory of multi-Bayesian estimation is developed for a fairly general group estimation problem. In particular, multi-Bayesian essentially complete class theorems for non-randomized rules are proved and applications to the normal distribution are given. Finally a characterization is given of the Nash estimation procedure.

1. Introduction. A general theory of multi-Bayesian estimation is presented in this paper. A group, $\mathcal{B} = \{1,\ldots,m\}$, of statisticians is required to choose a single estimate, $\hat{\theta}$, of an unknown parameter, θ, whose range of possible values is Θ. The data are assumed to be in hand (see Section 4 for a discussion of this assumption) and the statisticians must agree on a possibly randomized decision rule, δ, for choosing $\hat{\theta}$: δ depends on the data and is a probability distribution on Θ which we take to be a convex, open subset of R_p.

The value of any proposed δ to $i \in \mathcal{B}$ is determined by

AMS 1980 Subject Classification:
Key words and phrases: Multi-Bayesian estimation; Bayes estimates; Nash rules; Pareto optimal estimators.

$$B_i(\delta) = \int \int u_i(\hat{\theta}|\theta)\delta(d\hat{\theta})\, \Pi_i(d\theta). \qquad (1.1)$$

Here both integrals are over Θ. The integrand in equation (1.1), u_i, is the utility of $\hat{\theta} \in \Theta$ to $i \in B$ given θ while $\Pi_i(\cdot)$ is i's posterior distribution for θ. Assume u_i is bounded above for every i.

$\underset{\sim}{B}(\delta) = (B_1(\delta), \ldots, B_m(\delta))^T$ assumes a fundamental importance in guiding B's choice of δ and it is called δ's *assessment profile*.

Denote by D the class of all δ's. A subclass $D_0 \subset D$ is called *essentially B-complete* if for every $\delta \notin D_0$ there exists a $\delta' \in D_0$ such that $\underset{\sim}{B}(\delta) \leq \underset{\sim}{B}(\delta')$. This notion is formally equivalent to that of *essential completeness* in Wald's theory. Other analogues such as B-completeness, B-admissibility, and B-Bayesianity will emerge in the sequel but for brevity they will not be explicitly redefined in this new setting. Incidentally, B-admissibility is the same notion as (strong) *Pareto optimality* in other contexts.

Identifying essentially B-complete subclasses of D would seem to be as important here as solving the analogous problem in the Wald-setting has proven to be. Fortunately the strong formal links between the two problems may be exploited to simplify the present problem and various essentially B-completeness classes are presented in Section 2.

Of particular importance are the results concerning the essential B-completeness of D_{NR}, the class of nonrandomized procedures, $\delta = \delta_\alpha$ which are degenerate at points $\alpha = \hat{\theta}$, say, in Θ. Randomized rules are generally considered to be unsatisfactory, but as the results of Sections 2 and 3 will show, the choice of such a rule may well be inevitable if the preferences or belief's of B's members are sufficiently divergent. And these same results will indicate the degree of consensus which B must achieve to avoid the use of randomized decisions.

The theorems on B-completeness which we present in Section 2 delimit a class of decision rules from which a jointly acceptable solution to the group estimation problem might be found. Various potentially applicable solution concepts are presented and discussed in detail by Weerahandi and Zidek (1981,1983) and one of these, that of Nash (1950), is characterized in Section 3. Nash derived his solution as a normative solution to the two person bargaining problem. The domain of applicability of his theory is very broad. He formulates the bargaining problem as a game in normal (rather than extensive) form. The object of the game is the joint acceptance of the players on an agreement whose value to each player is determined by his/her Von Neumann - Morgenstern utility. The players are assumed to co-operate to the extent of complete and honest disclosure of their preferences among the possible agreements. As well it is assumed that joint randomization among agreements is feasible. Nash shows that certain weak, qualitative assumptions determine an equilibrium solution to the game and that this solution may be found by maximizing a certain functional which is defined in Section 3. While originally conceived as a solution to the bargaining problem, Nash's implicit definition of bargaining is so broad that it would seem to encompass in principle any group decision problem where joint agreement on a course of action is required. In other words, nothing about the nature of the bargaining process is assumed other than that strategic manipulation (for example, bluffing) and variable threats are excluded. And so the theory would seem to encompass the problem addressed in this paper.

Nash's solution is easily the most distinguished of various group decision concepts and it, unlike many others, does not require that the utilities of the Bayesians be compared. There is, in fact, no satisfactory general method for making such comparisons and this feature of the multi-Bayesian statistical decision problem distinguishes it from some of its formal analogues such as that of Wald, of empirical-Bayes hyperparameter estimation (c.f., Morris, 1983), of robust-Bayesianity (Berger, 1982) and of multi-criterion decision analysis (c.f. Yu, 1973).

2. Essential B-completeness. We will first present counterparts of some familiar results in the Wald-setting. Their proofs, which are simple adaptations of those of the standard results, are omitted. For details see, for example, Blackwell and Girschick (1954), Ferguson (1967), or Berger (1980). Most of this section will consist of results on the topic of its title which do not even formally resemble known results.

Consider $S = \{\underset{\sim}{B}(\delta):\delta\in D\}$, a convex subset of R_m which is *bounded above*. Assume, following Ferguson (1967) that S is also *closed from above*, i.e., $S \supset \lambda(S) \overset{\Delta}{=} \{\underset{\sim}{y}\in R_m:\{\underset{\sim}{y}\} = \overline{S} \cap \{\underset{\sim}{x}:\underset{\sim}{y} \leq \underset{\sim}{x}\}\}$ where $\overline{S}$ denotes the closure of S.

Let $\Pi = \{\underset{\sim}{\pi} = (\pi_1,\ldots,\pi_m): \pi_i \geq 0, \Sigma\pi_i = 1\}$ denote the class of all *B-priors* on B and Π_+ the subclass for which $\underset{\sim}{\pi} > 0$. A *B-Bayes rule with respect to* $\underset{\sim}{\pi}$ is any δ which maximizes $\underset{\sim}{\pi}^T \underset{\sim}{B}(\delta)$. Such δ's exist for any $\underset{\sim}{\pi} \in \Pi$ (Ferguson, 1967). Let B be the class of all such rules and B_+, the subclass corresponding to $\underset{\sim}{\pi}$'s for which $\underset{\sim}{\pi} > 0$. The closure of $\{\underset{\sim}{B}(\delta):\delta\in B_+\}$ is contained in $\{\underset{\sim}{B}(\delta):\delta\in B\}$ (Blackwell and Girshick, 1954, p.127). Let $\overline{B}_+ = \{\delta:\underset{\sim}{B}(\delta)\in \text{closure of } \{\underset{\sim}{B}(\delta'):\delta'\in B_+\}\}$

Let A be the class of all B-admissible rules and A_{NR} those nonrandomized rules which are B-admissible with respect to D_{NR}.

Since S is bounded above and closed from above, A is a minimal B-complete class (Ferguson, 1967) and $B_+ \subset A \subset \overline{B}_+ \subset B$ (Blackwell and Girshick, 1954).

If B_i is concave for every i then it is easy to show with the help of Carathéodory's Theorem (c.f., Rockafeller, 1970) that D_{NR} is essentially B-complete (See Ferguson 1967, Theorem 1, p.78). However, the assumption of concavity is too restrictive and alternatives will be presented.

In our analysis all sup's and inf's will be over Θ unless otherwise indicated. With an abuse of notation we will let $\underset{\sim}{B}(\hat{\theta}) = \underset{\sim}{B}(\delta_{\hat{\theta}})$.
The following condition is fundamental:
<u>Assumption</u>. There exists a bounded $K \subset \Theta$ such that

$$\{\hat{\theta}:\underset{\sim}{\pi}^T \underset{\sim}{B}(\hat{\theta}) = \sup \underset{\sim}{\pi}^T \underset{\sim}{B},\ \underset{\sim}{\pi}\in\Pi_+\} \subset K.$$

We now show that this assumption is valid under intuitively natural conditions. The B_i's will be called *positively affine independent* if there do not exist constants $c_i > 0$ and b_i such that

$$\sum_{i=1}^{M}(c_i B_i + b_i) = 0.$$

Lemma 2.1. *If the B_i's are bounded, positively affine independent functions for which there exists for every $\epsilon > 0$, a bounded set $K_\epsilon \subset \Theta$ such that $\underline{B}(\hat{\theta}) \le \inf \underline{B} + \underline{\epsilon}$ for every $\hat{\theta} \notin K_\epsilon$ where $\underline{\epsilon}$ is the vector whose elements are all ϵ, then the Assumption holds.*

Proof. Let $g_{\underline{\pi}} = \underline{\pi}^T \underline{B}$. Then $\zeta \overset{\Delta}{=} \inf\{\sup g_{\underline{\pi}} - \inf g_{\underline{\pi}} : \underline{\pi} \in \Pi\} > 0$. For if not there would exist $\{\underline{\pi}^{(i)}\}$ such that $\sup g_{\underline{\pi}(i)} - \inf g_{\underline{\pi}(i)} \to 0$ as $i \to \infty$. Since Π is compact these exists a $\underline{\pi}^{(*)} \in \Pi$ and subsequence $\{\underline{\pi}^{(j)}\}$ such that $\underline{\pi}^{(j)} \to \underline{\pi}^{(*)}$. Since the B_i's are bounded, $\sup g_{\underline{\pi}} - \inf g_{\underline{\pi}}$ is continuous in $\underline{\pi}$ so $g_{\underline{\pi}(*)} = d$, some constant, over Θ. But this contradicts the hypothesis that the B_i's are positively affine independent. So $\zeta > 0$.

Choose $\epsilon < \zeta$ and $K = K_\epsilon$. Then for $\hat{\theta} \notin K$, $\Sigma \pi_i B_i(\hat{\theta}) \le \epsilon + \Sigma \pi_i \inf \Sigma \pi_i B_i \le \epsilon - \zeta + \sup \Sigma \pi_i B_i < \sup \Sigma \pi_i B_i$ for all $\underline{\pi}$ so the conclusion follows.||

The proof of the following lemma is straightforward and hence omitted.

Lemma 2.2. *If $\{\hat{\theta} : B_i \ge M\}$, $i = 1, \ldots, m$ are for some M, bounded with a nonempty intersection, K, then the assumption holds.* ||

A principal result is the following:

Theorem 2.3. *If the Assumption is valid and the B_i's are twice differentiable, D_{NR} is essentially B-complete if and only if for every $\underline{\pi} \in \Pi_+$, the Hessian matrix of $\underline{\pi}^T \underline{B}$ is negative semi-definite at every point of $E \overset{\Delta}{=} \{\hat{\theta} : \delta_{\hat{\theta}} \in A_{NR}$ and $\underline{\pi}^T \nabla \underline{B}(\hat{\theta}) = 0\}$ where ∇, the gradient operator, is applied co-ordinate-wise to generate row vectors.*

Proof. The necessity is obvious. To prove sufficiency note that if $\hat{\theta} \in E$, it cannot be dominated by a randomized rule. And since E has a compact closure, the points in the closure of $\{\underline{B}(\hat{\theta}) : \hat{\theta} \in E\}$ are profiles of elements of D_{NR}. ||

"Essentially" cannot be dropped from the statement of the last Theorem. It is easy to vizualize δ's for which the Hessian matrix of Theorem 2.3 is even negative-definite at points, $\{a\}$, where $\underline{\pi}^T \underline{B}, \underline{\pi} \in \Pi_+$, attain their local maxima but where the elements of the closure of

$\{\underset{\sim}{B}(a)\}$ correspond to elements of both D_{BR} and $D - D_{BR}$.

The set E in Theorem 2.3 can be explicitly determined when

$$B_i(\hat{\theta}) = U_i(b_i(a)),\ i = 1, \ldots, m \tag{2.1}$$

where the b's are strictly concave and the U's are strictly increasing. Then $E = \{\hat{\theta} : \underset{\sim}{\pi}^T \nabla \underset{\sim}{B}(\hat{\underset{\sim}{\theta}}) = 0,\ \underset{\sim}{\pi} \in \Pi_+\}$.

Example 2.1. Weerahandi and Zidek (1983) show that when a certain class of normal posteriors and conjugate utility functions obtain, $B_i(\hat{\theta}) = \exp[-\frac{1}{2}Q(\hat{\theta} - \theta_i)^2]$, $i = 1, \ldots, n$, in the case where $p = 1$ and the preference-belief dispersion matrix is a constant, $Q > 0$. In this case it is easily shown that the Hessian of $\Sigma\pi_i B_i$, i.e., $\Sigma\pi_i d^2B_i(\hat{\theta})/d\hat{\theta}^2$, is a positive multiple of $\Sigma v_i(\theta_i - \bar{\theta})^2 - \frac{1}{2}Q^{-1}$ where $v_i = \pi_i B_i(\bar{\theta})/\Sigma\pi_r B_r(\bar{\theta})$ and $\bar{\theta} = \Sigma v_i \theta_i$. Thus D_{BR} is essentially B-complete in this example if any only if for every $\underset{\sim}{\pi} > 0$, $\Sigma v_i(\theta_i - \bar{\theta})^2 < \frac{1}{2}Q^{-1}$. But the left hand side of this inequality is obviously maximized over v by the achievable choice $v_{max} = v_{min} = \frac{1}{2}$ where the subscripts "max" and "min" denote indices at which θ_i attains a maximum and minimum respectively. Thus D_{BR} is essentially B-complete if and only if

$$(\theta_{max} - \theta_{min})^2 < \frac{1}{2}Q^{-1} . \tag{2.2}$$

This result generalizes for $p = 1$ the result of Weerhahandi and Zidek (1983) which was for the case $m = 2$.

Theorem 2.3 can be difficult to apply in specific cases since this entails the examination, in principle, of the Hessian matrix of $\Sigma\pi_i B_i$ over a broad domain, for each $\underset{\sim}{\pi}$ in a large class (Π_+). The next theorem would be easier to apply in the subclass of such cases which lie within its domain of applicability. There and in the sequel $\partial^2 f/\partial x \partial y$ will denote the second mixed derivative of f with respect to x and y.

Theorem 2.4. *Assume that $A \overset{\Delta}{=} (\partial^2 B_n/\partial B_i \partial B_j)$, an $(n - 1) \times (n - 1)$ matrix, is well defined at all points of the set* E *defined in the* Theorem 2.3. *Then D_{BR} is essentially B-complete if and only if A is negative semi-definite at all such points.*

Proof. Suppose A is positive definite at say $\hat{\theta}^0 \in E$. Then for $\hat{\theta}$'s close to $\hat{\theta}^0$,

$$B_m - B_m^o + \sum_{i=1}^{n} (B_i - B_i^o)(-\partial B_n^o / \partial B_i^o) > 0 \tag{2.3}$$

where the suffix, o, indicates evaluation at $\hat{\theta} = \hat{\theta}^o$. Note that $-\partial B_m^o / \partial B_m^o \geq 0$ for $i = 1,\ldots,n - 1$ since $\hat{\theta} = \hat{\theta}^o \in A_{NR}$. So for any pair, $\hat{\theta}^1$, $\hat{\theta}^2$, of $\hat{\theta}$'s which are sufficiently close to $\hat{\theta}^3$ there must exit an α, $0 < \alpha < 1$, for which $\alpha B_i^{\,1} + (1 - \alpha)B_i^{\,2} > B_i^{\,3}$, $i = 1,\ldots,n$ where again suffices indicate evaluation at the corresponding $\hat{\theta}$'s. Otherwise $\alpha(B_i^{\,1} - B_i^{\,3}) + (1 - \alpha)(B_i^{\,2} - B_i^{\,3}) \leq 0$ for all α, $0 < \alpha < 1$ and this contradicts equation (2.3) and establishes the necessity of the negative semi-definiteness condition.

Now suppose this condition holds for all $\hat{\theta} \in E$ and that D_{NR} is not essentially B-complete. Then there would exist $\hat{\theta} \in D_{NR}$ and $\delta \in D - D_{NR}$ such that a δ dominates $\delta_{\hat{\theta}}$. By Caratheodory's Theorem (c.f., Rockafeller, 1970), $B_i(\delta) = \sum_{j=1}^{m} \delta_j B_i^{\,j}$ for all i and some m, $2 < m \leq n + 1$, where $B_i^{\,j} = B_i(\hat{\theta}^j)$. We may assume $\delta \in A$ so $B^j \triangleq (B_1^j, \ldots, B_m^j)^T$, $j = 1,\ldots,m$ (and δ) are all B-Bayes with respect to the same $\underline{\pi} \in \Pi$. Since $m \geq 1$ there exist points at which $\underline{\pi}^T \underline{B}$ attains a local minimum. Since $\underline{B}$ is continuous, in every neighbourhood of at least one such point, $\hat{\theta}^o$ say, there exist other points for which $\underline{\pi}^T \underline{B}^3 < \underline{\pi}^T \underline{B}$. But this is easily shown to contradict the assumption of negative semi-definiteness. So D_{NR} is essentially B-complete.||

Before presenting the next example some computational formulas are presented for a relevant special case, that of equation (2.1) where the U's and b's are continuously twice differentiable. The conditions equivalent to essential B-completeness which are established in Theorems 2.3 and 2.4 involve $F_o \triangleq \{\hat{\theta} : \delta_{\hat{\theta}} < B_+\}$. There is a convenient parametrization of this set which will now be described.

Observe that $\hat{\theta} \in F_o$ if and only if

$$\underline{\pi}^T \underline{b}(\hat{\theta}) = \underline{0} \tag{2.4}$$

for some $\underline{\pi} > 0$. Let $\underline{\lambda} = \underline{\pi}/\pi_n$ and

$$\underline{0} = \underline{F}(\underline{\lambda}, \hat{\theta}) = \underline{\lambda}^T \nabla \underline{b}(\hat{\theta}). \tag{2.5}$$

The Jacobian $J = (\partial F_i/\partial \hat{a}_j)$ associated with these equations is easily shown to be $J = \Sigma\lambda_i b_i''(\hat{\theta})$ where, in general, $g''(x)$ will denote the matrix of g's second derivatives with respect to the co-ordinates of $\underset{\sim}{x}$. Since J is negative definite, its determinant is nonzero for every point $(\underset{\sim}{\lambda},\hat{\theta})$ where $\lambda_i > 0$ for all i, $\lambda_n = 1$ and $\hat{\theta} \in \Theta$. The implicit function theorem enables us to conclude that equation (2.4) has a solution, $\hat{\theta} = \theta(\underset{\sim}{\lambda})$ which has continuously differentiable co-ordinate functions. Thus $F_o = \{\theta(\underset{\sim}{\lambda}):\underset{\sim}{\lambda} > 0,\ \lambda_n = 1\}$.

The derivatives, $\partial\theta/\partial\lambda_i$, $i = 1, \ldots n - 1$ are readily computed. Observe that from equation (2.5) $F_m(\underset{\sim}{\lambda}+e_ih,\ \theta(\underset{\sim}{\lambda}+e_ih)) \cong F_m(\underset{\sim}{\lambda},\theta) + h[\partial F_m(\underset{\sim}{\lambda},\theta)/\partial\lambda_i + \nabla_\theta F_m(\lambda,\theta)(\partial\theta/\partial\lambda_i)]$,
$i = 1, \ldots, n - 1$, $m = 1, \ldots, p$ where e_i is the vector all of whose elements are 0 except the ith which is 1, and $\theta = \theta(\underset{\sim}{\lambda})$. But $\partial F_m(\lambda,\theta)/\partial\lambda_i = \partial b_i/\partial\theta_m$ and $\nabla_\theta F_m(\underset{\sim}{\lambda},\theta) = \partial\nabla b_\cdot(\theta)/\partial\theta_m$ where $b_\cdot = \underset{\sim}{\lambda}^T\underset{\sim}{b}$ so,

$$\partial\theta/\partial\lambda_i = -[b''_\cdot(\theta)]^{-1}b'_i(\theta) \qquad (2.6)$$

where $b'_i = (\nabla b_i)^T$.

To obtain the matrix A of Theorem 2.4 in a more explicit form observe that $\partial B_n/\partial B_k = \Sigma(\partial B_n/\partial\lambda_s)(\partial\lambda_s/\partial B_k)$. So

$$\partial^2 B_n/\partial B_g\partial B_h = \Sigma\Sigma(\partial^2 B_n/\partial\lambda_k\partial\lambda_l)(\partial\lambda_k/\partial B_g)(\partial\lambda_l/\partial B_h) \qquad (2.7)$$
$$+ \Sigma(\partial B_n/\partial\lambda_s)(\partial^2\lambda_s/\partial B_g\partial B_h).$$

But $\partial B_n/\partial\lambda_s = \nabla B_n\cdot\partial\theta/\partial\lambda_s = -U'_n(b'_n)^T(b''_\cdot)^{-1}(b_s)$. Also $dB_i = \sum_{j=1}^{n-1}(\partial B_i/\partial\lambda_j)d\lambda_j$, $i = 1, \ldots, n-1$ or $dB = Jd\lambda$ so $d\lambda = J^{-1}dB$, i.e., $d\lambda_i = \Sigma J^{ij}dB_j$, $i = 1, \ldots, n-1$ where $J^{-1} = (J^{ij})$. Consequently $\partial\lambda_s/\partial B_g = J^{sg}$.

To obtain the required second derivatives will require a repeated application of these methods. Note that if $n = 2$ the necessary and sufficient condition for the essential B-completeness of D_{NB} is simply that $d^2B_2/dB_1^2 \le 0$ for every $\hat{\theta} = \theta(\underset{\sim}{\lambda})$ where $\underset{\sim}{\lambda} = (\lambda_1,1)$.

Example 2.2. Here $n = 2$ and following Weerahandi and Zidek (1983), $B_i(\hat{\theta}) = U_i(b_i(\hat{\theta}))$, where $U_i(x) = \exp(x)$ and $b_i(\hat{\theta}) = -(\hat{\theta} - \theta_i)^T Q_i(\hat{\theta}-\theta_i)$

where $Q_i > 0$ is a constant matrix, $i = 1,2$. Equation (2.5) becomes $\underline{0} = (\hat{\theta}-\theta_2)^T Q_2 + \lambda_1(\theta-\theta_1)^T Q_1$ so that $\hat{\theta} = \theta(\lambda) = (Q_2+\lambda_1 Q_1)^{-1}(Q_2\theta_2+\lambda_1 Q_1\theta_1)$. For this $\hat{\theta}$, $b_1(\hat{\theta}) = -\lambda_1^2 \Sigma\Delta_i^2 D_i(D_i+\lambda_1)^{-2}$ and $b_2(\hat{\theta}) = -\Sigma\Delta_i^2 D_i^2(D_i+\lambda_1)^{-2}$ where $D = \mathrm{diag}(D_1,\ldots,D_p)$ is the diagonal matrix of $Q \hat{=} Q_2^{-1/2}Q_1Q_2^{-1/2}$, $\Delta = OQ_2^{1/2}(\theta_1-\theta_2)$, and O is the orthogonal matrix which diagonalizes Q. Using the computational strategy outlined above it is easily shown that $dB_2/dB_1 = -\lambda_1 B_2 B_1^{-1}$ and $d^2B_2/dB_1^2 = B_2(\lambda_1 B_1)^{-2}[(\lambda_1+1) - \frac{1}{2}(\lambda_1\Sigma\Delta_i^2 D_i^2(D_i+\lambda_1)^{-3})^{-1}]$. Thus a necessary and sufficient condition for essential B-completeness is

$$\sup_{\lambda_1 \geq 0}\left|\lambda_1(1+\lambda_1)\Sigma\Delta_i^2 D_i^2(D_i+\lambda_1)^{-3}\right| \leq \frac{1}{2}$$

or, in the noncanonical setting,

$$\sup_{\lambda_1 \geq 0}\lambda_1(1+\lambda_1)(\theta_1 - \theta_2)^T Q(\lambda_1)(\theta_1 - \theta_2) \leq \frac{1}{2} \qquad (2.8)$$

where $Q(\lambda_1) = Q_1R(\lambda_1)Q_2R(\lambda_1)Q_2R(\lambda_1)Q_1$ and $R(\lambda_1) = (Q_1+\lambda_1 Q_2)^{-1}$. In the special case where $Q_1 = Q_2$ equation (2.8) reduces to the result obtained by Weerahandi and Zidek (1983).

<u>3. Nash Estimation Rules</u>. Nash (1950) proposed weak, intuitively appealing conditions which an equilibrium solution to an $n = 2$ person decision problem should satisfy. He then proved that such a solution must maximize a certain simple functional. His result is easily extended to the case $n \geq 2$ where the corresponding functional is $\prod_{i=1}^{n}[B_i(\delta)-c_i]^{\alpha_i}$, $\alpha_i \geq 0$, $\Sigma\alpha_i = 1$ and here c_i represents the worth, in utility, of i's current assets. The special case $n = 2$, $\alpha_1 = \alpha_2 = 1/2$ gies Nash's functional. The required maximization is carried out over what might be called the Nash-feasible rules, $\{\delta : \underline{B}(\delta) > \underline{c}\}$. See Weerahandi and Zidek (1981, 1983) for more details.

Assume $\underline{\alpha} > \underline{0}$ without loss of generality. By Carathéodory's Theorem (c.f., Rockafeller, 1970), $\underline{B}(\delta^N) = \Sigma\delta_j\underline{B}^{(j)}$ for some δ's, $\underline{\delta} \geq 0$, $\Sigma\,\delta_j = 1$ and $\underline{B}^{(j)} = \underline{B}(\hat{\theta}^{(j)})$, where δ^N is a Nash solution and $m \leq n + 1$. This representation need not be unique.

Since δ^N maximizes the generalized Nash product and $\underset{\sim}{\alpha} > 0$, δ^N must be B-Bayes with respect to a B-prior $\underset{\sim}{\lambda} \in \Pi$.

For convenience the following notation is adopted: $\lambda \in \Theta$ and $\underset{\sim}{B}^\lambda = \underset{\sim}{B}[\lambda]$, $(x,y) = x^T y$, $d_i^\lambda = \partial B_n^\lambda / \partial B_i^\lambda$, $i = 1, \ldots, n$, $d^\lambda = (d_1^\lambda, \ldots, d_n^\lambda)$ $\Lambda = \{\lambda : \delta_\lambda \in \mathcal{B}_+\}$. H_λ will denote the tangent plane to the surface $\{B^\lambda : \lambda \in \Lambda\}$ which is located at B^λ, i.e., $H_\lambda = \{B : (B, d^\lambda) = (B^\lambda, d^\lambda)\}$. It is quite possible that $H_{\lambda_1} = H_{\lambda_2}$ when $\lambda_1 \neq \lambda_2$. Define the equivalence relation "$\sim$" by $\lambda_1 \sim \lambda_2$ if and only if $H_{\lambda_1} = H_{\lambda_2}$. Then $\lambda_1 \sim \lambda_2$ if and only if $d_i^{\lambda_1} = d_i^{\lambda_2}$ for all i and $(B^{\lambda_1}, d^{\lambda_1}) = (B^{\lambda_2}, d^{\lambda_2})$. Under this equivalence relation Λ decomposes into disjoint equivalence classes which will be indexed by ξ. That is $\Xi = \Lambda/\sim$. Let $\zeta : \Lambda \to \Xi$ denote the canonical mapping so that $\lambda_1 \sim \lambda_2$ entails $\zeta(\lambda_1) = \zeta(\lambda_2) = \xi$ say. Finally define $L_\xi = L_\lambda$, $d^\xi = d^\lambda$ and $c^\xi = (B^\lambda, d^\lambda)$ if $\zeta(\lambda) = \xi$.

If δ^N is a Nash solution, $B[\delta^N] = \sum_{j=1}^{m} \omega_j \underset{\sim}{B}^{\lambda_j}$, $m \leq n + 1$ by Carathéodory's Theorem (Rockafeller, 1970). This establishes that $\underset{\sim}{B}^{\lambda_j}$, $j = 1, \ldots, m$ are co-planar and the plane in question must be tangent to $\underset{\sim}{B}^{\lambda_j}$, $j = 1, \ldots, m$. Thus $\lambda^j \sim \lambda^i$, $j \neq i$ and H_{λ_j} is this tangent plane for every i. In summary, $B[\delta^N]$ must lie in a plane which is simultaneously tangent to S and a level surface of the function $g(x) = \Pi x_i^{\alpha_i}$. Also it must be a weighted average of agreement utility vectors corresponding to $\lambda \in \zeta^{-1}(\xi)$ for some $\xi \in \Xi$. In general we will take the minimal set of such utility vectors and so assume they are linearly independent.

To complete our analysis we will determine conditions under which δ^N corresponds to $\xi \in \Xi$ in the manner just prescribed. The tangent plane to g passing through B^0 is easily shown to be $H^0 = \{\underset{\sim}{B} : \Sigma(\alpha_i / B_i^0) B_i = 1\}$. Thus δ^N corresponds to ξ if and only if there exists $\{\omega^\lambda : \lambda \in \zeta^{-1}(\xi)\}$, $\Sigma\omega^\lambda = 1$, $\omega^\lambda > 0$ for which $B^0 = \Sigma\{\omega^\lambda \underset{\sim}{B}^\lambda : \lambda \in \zeta^{-1}(\xi)\}$ and

$$H^0 = L_\xi \tag{3.1}$$

But condition (3.1) is equivalent to $\alpha_i / B_i^0 = d_i^\xi / c^\xi$, $i = 1, \ldots, n$ which is equivalent to $(B_i^0 d_i^\xi)/(\alpha_i c^\xi) = 1$, $1 = 1, \ldots, n$ and hence to

$$\Sigma \omega^\lambda A_{i\lambda} = 1, \quad A_{i\lambda} = ({}_i^\lambda d_i^\xi)/(\alpha_i c^\xi) \tag{3.2}$$

Let $A = (A_{i\lambda}) = (A^{\lambda_1}, \ldots, A^{\lambda_m}) = (A_1^T, \ldots, A_n^T)^T$ where $m = |\zeta^{-1}(\xi)|$. Then equations (3.2) may be written as

$$\Sigma\omega^{\lambda}A^{\lambda} = \underline{1}_n = (1, \ldots, 1)^T \tag{3.3}$$

or alternatively

$$A\omega = \underline{1}_n \quad . \tag{3.4}$$

Thus δ^N corresponds to ξ if and only if there exists ω $\Sigma\omega^{\lambda} = 1$, $\omega^{\lambda} > 0$ for which equations (3.3) and (3.4) hold.

Equation (3.3) shows that $\underline{1}_n$ must be in the cone of $\{A^{\lambda_1}, \ldots, A^{\lambda_m}\}$. If, for example, $m = 1$, this condition shows that $\delta^N = \delta_{\lambda}$ if and only if $B_i^{\lambda}(\partial B_n^{\lambda}/\partial B_i^{\lambda}) = \alpha_i \Sigma B_j^{\lambda}(\partial B_n^{\lambda}/\partial B_j^{\lambda})$. Equation (3.4) implies that $\omega = (A^TA)^{-1}A^T\underline{1}_n$ which implies that unless the quantity on the right hand side of equation (3.4) satisfies $\underline{1}_m^T(A^TA)^{-1}A^T\underline{1}_n = 1$ and has non-negative coordinates, δ^N cannot have $\zeta^{-1}(\xi)$ as its support. Conversely, if these conditions are satisfied δ^N does correspond to ξ.

In general this characterization may be difficult to apply because the equivalent classes, $\zeta^{-1}(\xi)$, are too numerous to be explicitly determined. For small values of n, however, this characterization is applicable. We illustrate this in the following example.

<u>Example 3</u>. Consider the specialization of Example 2.2 obtained by taking $p = n = 2, Q_1 = I_2$, $Q_2 = I_2/2$, $\theta_1^T = (0,0)$ and $\theta_2^T = (1,1.5)$. The set S is illustrated in Figure 3.1.

Figure 3.1 Here

It is clear from Figure 3.1 that there is just a single ξ for which $\zeta^{-1}(\xi)$ contains more than one point, namely that for which $\zeta^{-1}(\xi) = \{\lambda_1, \lambda_2\}$. Whether or not δ^N has support determined by this class depends on α_1 and α_2.

We first determine α_1 and α_2 by graphical analysis. [For the details see de Waal et al, 1982]. The result is $\lambda_1 = (0.2135, 0.3203)^T$.

As well, for this class $d_1^{\xi} = 0.782, d_2^{\xi} = 1$ and $c^{\xi} = 1.04$. Thus

$$A = (A_{i\lambda_j}) = \begin{bmatrix} 0.6484/\alpha_1 & 0.05602/\alpha_1 \\ 0.3519/\alpha_2 & 0.9441/\alpha_2 \end{bmatrix}$$

so that on applying equation (3.4) we obtain

$$\omega = A^{-1}\underline{1}_2 = (1.594\alpha_1 - 0.094\alpha_2, 1.094\alpha_2 - 0.594\alpha_1)^T$$

It follows that $\omega_1 \geq 0$ and $\omega_2 \geq 0$ while $\omega_1 + \omega_2 = 1$ provided that $0.05569 \leq \alpha_1 \leq 0.6481$. Otherwise δ^N will be a nonrandomized rule, which is degenerate at

$$\begin{aligned} a &= (\alpha_1 Q_1 + \alpha_2 Q_2)^{-1}(\alpha_1 Q_1 \theta_1 + \alpha_2 Q_2 \theta_2) \\ &= (\alpha_1 + \alpha_2/2)^{-1}(\alpha_1 \theta_1 + \alpha_2 \theta_2/2) \ . \end{aligned}$$

Here we observe a situation in which δ^N may be nonrandomized even when $\mathcal{D}_{NR}$ is not B-complete.

4. Discussion. Essential B-complete class theorems and a characterization of the Nash solution are given in Sections 2 and 3, respectively when the data are in hand. The other case which is not treated here where the data are forthcoming would involve an entirely different analysis. The appropriate assessment profile would be obtained by averaged $\underset{\sim}{P}$ in equation (1.1) over the prospective data sets. The set S of feasible profiles would no longer be the convex hull of the profiles of nonrandomized rules, but an average (over the data) of such hulls. Carathéodory theorem would no longer apply and Theorems 2.3 and 2.4 would be invalid. In fact, it seems unlikely that such precise statements would be achievable judging by the corresponding developments in the Wald setting. There would be an obvious counterpart of sufficiency, B-sufficiency, and the class of δ's based on a B-sufficient statistic would be essentially B-complete.

Theorems 2.3 and 2.4 reveal that there will be a basis for a consensual choice (i.e., nonrandomized decision) if and only if a fairly strong condition is satisfied. In the application treated in Example 2.1 where a joint estimate of a (real-valued) normal mean is required these theorems yield the condition that the magnitude of the range of the n posterior means must not exceed $[Q^{-1}/2]^{1/2}$. The quantity, Q^{-1}, is $W_1 + \Sigma_i$ (see Weerahandi and Zidek, 1983) where W_i^{-1} and Σ_i^{-1}, respectively, are the preference and posterior precision parameters, $u_i(\hat{\theta}|\theta) = \exp[-\frac{1}{2}(\hat{\theta}-\theta_i)^2/W_i]$ and Π_i has a density function which is proportional to $\exp[-\frac{1}{2}(\hat{\theta}-\theta_i)^2/\Sigma_i]$, $i = 1, \ldots, n$. As the sample size increases the range in question decreases (in a stochastic sense) so a consensus would very likely exist if the sample size were sufficiently large. How large depends critically on the strengths of preference, i.e., the size of W_i. Strong preferences (small W_i) can only be overcome with large samples. The size of Σ_i is not ultimately of much relevance since it would be nearly 0 for large sample sizes. Since the spread of the prior means determine the range the required sample size would increase with n, the number of statisticians, involved.

The discussion leading to inequality (2.8) is readily shown to imply that D_{NR} is essentially B-complete in Example 2.2 if $(\theta_1-\theta_2)^T Q_2 (\theta_1-\theta_2) \leq 2$ when $Q_1 < Q_2$ and hence (by symmetry) if $(\theta_1-\theta_2)^T Q_1 (\theta_1-\theta_2) \leq 2$ when $Q_2 < Q_1$. This may be interpreted as saying, essentially, that if either one of the statisticians has much stronger preferences than the other, there is a basis for consensual choice if the remaining statistician perceives the first to agree closely with him; proximity is measured here by the appropriate Mahalanobis-like distance.

There are formal similarities between multi-Bayesian and multi-attribute decision analyses. We have chosen not to develop and use those links, preferring instead to pursue the connections with statistical decision theory. This choice is natural because of the statistical context of the problems treated by our theory.

The differentiability assumptions imposed in Section 3 are somewhat unappealing from a philosophical point of view as a referee has pointed out. Weaker qualitative assumptions would have been desirable. We do not know any such assumptions, however, and in any case this is not a matter of great practical substance since the inevitable approximations entailed in the elicitation of utilities and so on, commonly lead to choices which are made from parametric classes of convenient, smooth models.

Acknowledgement. This work was supported by the University of the Orange Free State, South Africa, and the Council for Scientific and Industrial Research, where the last author was a Visiting Scientist while on leave from the University of British Columbia. This leave was financed in part in Canada's Research Councils for the natural sciences and engineering and for the social sciences and humanities. The work benefitted from the comments of Dr. Theo Stewart.

References

[1] Berger, J.O. (1980). Statistical decision theory. New York. Springer-Verlag.

[2] Berger, J.O. (1982). The robust Bayesian viewpoint. Tech. Rep. 82-9, Dept. of Statistics, Purdue University.

[3] Blackwell, D. and Girschick, M.A. (1954). Theory of games and statisical decisions. New York. John Wiley and Sons Inc.

[4] De Waal, D.J., Groenewald, P.C.N., Van Zyl, J.M., Zidek, J.V. (1982). A randomized solution for multi-Bayes estimates of the multinormal mean. Unpublished manuscript.

[5] Ferguson, T.S. (1967). Mathematical statistics: a decision theoretical approach. New York. Academic Press.

[6] Morris, C.N. (1983). Parametric empirical Bayes inference: theory and applications. J. Amer. Statist. Assoc., 78:47-65.

[7] Nash, J.R. Jr. (1960) The bargaining problem. Econometrica, 18:155-162.

[8] Rockafellar, R.T. (1970) Convex Analysis. Princeton University Press.

[9] Weerahandi, S. and Zidek, J.V. (1981). Multi-Bayesian statistical decision theory. J. Roy. Statist. Soc., Ser.A. 144:85-93.

[10] Weerahandi, S. and Zidek, J.V. (1983). Elements of multi-Bayesian decision theory. Ann. Statist., 11.

[11] Yu, P.L. (1973). Introduction to domination structures in multicriteria decision problems. In: J.L. Cochrane and Zeleny, Eds., Multiple Criteria Decision Making. South Carolina Press, pp. 249-261.

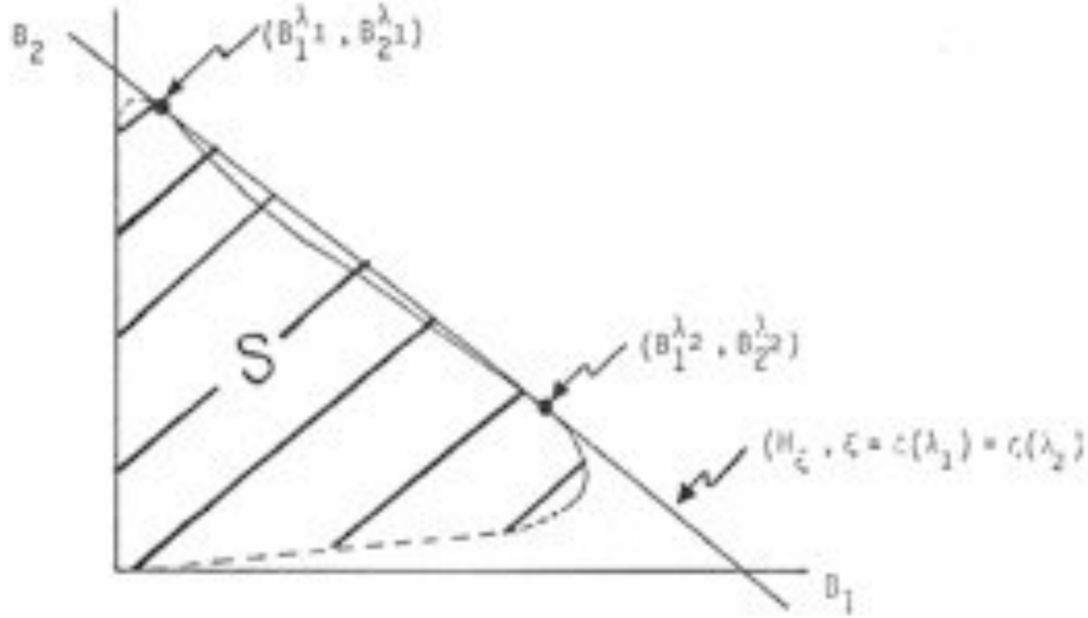

FIGURE .1. *The set of feasible utility pairs* (B_1, B_2) *for a bivariate normal mean estimation problem.*