MULTI-BAYESIANITY

James V. Zidek

Technical Report No. 05

December 1984

ABSTRACT

This report consists of a summary of three lectures given by the author in May 1983 at the invitation of the Board of Studies for Statistics, University of London. Thanks go to Dr Agnes Herzberg, Department of Mathematics, Imperial College and Professor Phil Dawid, Head, Department of Statistical Science, University College London for their part in making arrangements for these lectures. The author is also grateful to the University of London for providing an opportunity to discuss ideas rather outside the mainstream of statistics. The first of the lectures deals with the problems of combining opinions. The second is about the problem faced by a group of statistical decision makers who must compromise their individual positions to reach a joint decision. The final lecture contains two applications, the first on group-estimating the multinormal mean, and the second on optimal second-quessing.

I. MULTIBAYESIANITY: CONSENSUS OF OPINION

1. INTRODUCTION

This talk is a summary of mostly recent work on the problem of aggregating opinions (see also the surveys of Winkler 1968 and Hogarth 1977). Each of a group of individuals has information about the state, θ, of the world, like, for example, the cause of this patient's pain. How should this information be combined?

This talk will not be concerned with the related problem about how this information should be used in reaching a group-decision (about the cause of the pain, for example).

The relevant literature on this problem is extensive and omissions are inevitable. In any case, this survey will tend to focus primarily on work in the area of my own current interests.

2. QUANTIFYING DEGREES OF BELIEF

Most of the solutions to the aggregation problem require that each individual's information be quantifiable as a subjective probability distribution. This distribution assigns a number, $P(E)$, to each possible subset, E, of the set, Θ, of all possible states of the world according to the degree to which it is believed to contain the "true" state.

Such a distribution need not always exist (Kraft, Pratt, and Seidenberg 1959). Its meaning has been challenged (Tversky and Kahneman 1974). Its existence is implied by weak conditions (*c.f.* Fine 1973, de Finetti 1974 and Lindley 1982a), but to elicit it is unclear (Savage 1971). In any case, it is widely employed in theory and practice. Raiffa (1968) gives a very readable account of it.

Often $P(E) = \int_E p(\theta)\mu(d\theta)$ for some measure μ and point-function, $p \geq 0$, on the states θ of the world. Then p is a convenient stand-in for P.

Usually $P(\Theta) = 1$ but "... nothing necessitates this choice; it is a mere convention. The problem with conventions is that we may lose sight of their arbitrary origin. The unit normalization and non-negativity conventions begin to appear as substantive properties...", Fine (1973). Nothing forces the group's members to choose the same scale for their probabilities. They may even have different scales when Θ is not logically complete (for example, {Heads, Tails} versus {Heads, not Heads}), for although all probabilities are

conditional, it may not be possible to fully describe the conditioning event.

3. TAXONOMY OF SOLUTIONS.

These are either "summarizers" or "accumulators". The first are formulas for combining (pooling) the individuals' subjective distributions to obtain something typifying their views. The second accumulates the information of the individuals either through interaction (the Delphi method, etc.) or a "super-Bayesian" for whom the individual opinions serve as data.

4. INTERACTIVE APPROACHES

The individuals interact with or without feedback limitations, and through dialogue attempt to reach a consensus.

WITHOUT FEEDBACK RESTRICTIONS. This is natural and easy to adopt. It permits the free exchange of information so the range of views diminishes. The group may well be "synergistic". However, on the negative side, the interaction may induce conformity, agreement without a commensurate exchange of information. It permits strategic manipulation, bluffing, intimidation and threats, to be employed. And the group needs a strong director.

WITH FEEDBACK RESTRICTIONS. (The Delphi technique and variants; see Dalkley 1975). Open discussion is not permitted. The feedback may consist of summaries, such as group means or quantiles, for example, when the task is point estimation. Or it may consist of the subjective distributions of the other individuals. The process may be iterated. This approach can be inexpensive since the individuals need not meet. And social pressure is reduced. On the negative side, feedback is unduly restricted, information cannot be freely exchanged. Recent studies of this method cast doubt on its value (see Hogarth 1977).

FORMALIZATIONS. DeGroot (1974) postulates that $F_{in} = \sum w_{ij} F_{j,n-1}$ where F_{ik} = i's distribution after iteration k and $w_{ij} \geq 0$ are fixed weights with $\sum_j w_{ij} = 1$. Alternatively, $P_n = W P_{n-1} = W^n P_o$. If the i-th row of W^n converges to $(\pi_1, \ldots, \pi_k)$ as $n \to \infty$, "consensus" will have been reached; $\sum \pi_j P_{jo}$ is the group's distribution. Berger (1981) establishes conditions for the required convergence to take place.

In a more general model, W^n is replaced by $W_1, \ldots, W_n$. The W_i's may themselves be arrived at by a dialogue (Lehrer 1978; Chatterjee and Senata 1977). Other iterative-interaction models have been proposed and analysed (Bacharach 1979; Aumann 1976). Press (1978) offers a more statistical model for the process.

5. SUMMARIZATION PROBLEM

Suppose, given subjective probability distributions, P, Q, a summary of $\{P,Q\}$ is required.

LINEAR OPINION POOL. Stone (1961) proposed $R = \alpha P + (1-\alpha)Q$, $0 \le \alpha \le 1$ on grounds relating to a solution of Savage (1954) to the group decision problem. Obviously R is itself a probability distribution. However, it has some unexpected properties. By definition $R(A|B) = R(A \cap B)/R(A)$. It might be expected that $R(A|B) = \alpha P(A|B) + (1-\alpha)Q(A|B)$ as well. However, this is true in only exceptional circumstances since, by an elementary argument, this implies that $0 = \alpha(1-\alpha)[Q(A) - P(A)][P(B|A) - Q(B|A)]$. It is easy to find P and Q which make $\alpha = 0$ or $\alpha = 1$ the only acceptable alternatives. So without limiting R's domain, it follows that R is a dictatorship.

Incidentally, requiring that $P(A) \ne Q(A)$ and $P(B|A) \ne Q(B|A)$ implies that the underlying probability space, $(\Theta, \mathcal{B})$, be tertiary (Wagner 1982), i.e. $\mathcal{B} \ne \{\phi, D, \bar{D}, \Theta\}$ for any D.

This conditionality inconsistency is well-known (Raiffa 1968; Dalkey 1972, 1975). A related concern is R's failure to preserve independence (Wagner 1982; Genest 1983a).

But the linear pool does have appealing properties. It is simple and natural. It yields a probability distribution. McConway (1981) shows that if Θ is tertiary, then $R = \alpha P + (1-\alpha)Q$ provided R is a probability distribution and $R(A) = f(P(A), Q(A))$ for some f and all events A (or, equivalently, R has the Marginalization Property and the Zero Preservation Property (ZPP)). Genest (1983a) drops ZPP and generalizes McConway: $R = \alpha_1 P + \alpha_2 Q + (1 - \alpha_1 - \alpha_2)S$ for some probability distribution S and (possibly negative) weights α_i with $|\alpha_i| \le 1$ and $|\alpha_1 + \alpha_2| \le 1$. So now R may vary inversely with P (or Q) and it may even ignore both P and Q. Genest does give a weak condition under which $\alpha_i \ge 0$. Additional support for the linear pool is found in Bacharach (1975).

Usually densities are used to specify distributions. The linear pool has an obvious counterpart, $r = \alpha p + (1-\alpha)q$, which may formally be derived by differentiation. The required axiomatic support for its use is given by Genest (1983b).

For simplicity, only $n =$ [illegible] individuals have been considered above. All

of the results quoted in this lecture have counterparts for general n.

OBJECTIONS TO THE LINEAR POOL. A normative basis for selecting the pooling weights does not exist, although various proposals have been made (Roberts 1965; Raiffa 1968; Winkler 1968). The density-version of the pool is typically multi-modal so no clearcut choice for a jointly-preferred θ is indicated. The recipe is not Externally Bayesian, i.e. Prior-to-Posterior-Coherent in the terminologies of Madansky 1978 and Weerahandi and Zidek (1978), respectively. This means that pooling the priors and invoking Bayes rule when the data is received gives a different answer than pooling the posteriors if the weights are fixed. Whether or not this is objectionable is controversial (McConway 1981). Finally the linear pool is not scale-invariant, i.e., it does depend critically on the assumption that $P(\Theta) = Q(\Theta)$. So, for example, vague prior opinions cannot be accommodated.

Hogarth (1977) points out the relative insensitivity of the linear pool to the choice of pooling weights. This could be considered either advantageous or disadvantageous.

LOGARITHMIC OPINION POOL. This, too, can be applied to distributions or densities. In the latter case $r = c(p^{\alpha p})(q^{\alpha q})$, $\alpha p, \alpha q \ge 0$. Bacharach (1973) attributes it to someone named Hammond. For Dalkey (1975) it is a natural "summarizer": (i) it derives from the super-Bayesian approach (Morris 1977; French 1981; Lindley 1982c), (ii) has a natural conjugate-likelihood interpretation (Winkler 1968), (iii) is typically uni-modal (in density form) and less dispersed than the corresponding linear pool (Winkler 1968), (iv) it is the only Externally Bayesian method of the type $f(p(\theta), q(\theta)) \div \int f(p(\theta), q(\theta))\mu(d\theta)$ (Genest 1983b) and (v) it is scale invariant.

OBJECTIONS TO THE LOGARITHMIC OPINION POOL. Like the linear pool, it lacks a normative basis for weight selection; the super-Bayesian approach helps here. It is mathematically more complicated than the linear pool, even if it does overcome some of the other objections to the latter. Zero's are vetoes; the difficulties faced are like those which arise in conventional Bayesian theory when the likelihood and prior have different support sets.

BROADER HORIZONS. Θ will be an arbitrary carrier set, for example, $\Theta = \{\delta : \delta \in$ [illegible] $\}$ = set of available decision rules or $\Theta = (-\infty,\infty) = \{N(\mu,1), -\infty < \mu < \infty\}$. Interest will focus on [illegible]-functions ("propensity functions" except that name has already been appropriated)

$b(\theta) > 0, \theta \in \Theta$, for which $b(\theta)/b(\lambda)$ is deemed to be the relative degree of support for θ over λ. Examples: (i) $b(\theta)$ = observed significance level for the hypothesis " the true state of nature is θ", (ii) $b(\delta) = u(\delta) - u_o$ = gain-in-utility using $\delta \in \{\delta: u(\delta) - u_o > 0\}$, (iii) $b(\theta)$ = likelihood of θ, (iv) $b(\theta)$ = density of a diffuse prior, (v) $b(\theta)$ = density of a prior or a posterior. Given b_1 and b_2, the relative propensity profile for $\mu \in \Theta$ versus $\nu \in \Theta$ is $(b_1(\mu)/b_1(\nu), b_2(\mu)/b_2(\nu)) = RP(b_1, b_2; \mu, \nu)$. A prospective pooling method, $T(b_1, b_2)(\theta)$, will be required to preserve <u>Relative Propensity Consistent</u> (RPC) :

$T(b_1,b_2)(\mu_1)/T(b_1,b_2)(\nu_1) \geq T(c_1,c_2)(\mu_2)/T(c_1,c_2)(\nu_2)$ whenever $RP(b_1,b_2; \mu_1,\nu_1) \geq RP(c_1,c_2; \mu_2,\nu_2)$. It then follows, under this seemingly weak requirement, that

$T(b_1,b_2)(\theta) = C b_1^{\alpha_1} b_2^{\alpha_2}$, the logarithmic pool, if $|\Theta| \geq 3$. The argument for this is quite straightforward. Clearly

$T(b_1,b_2)(\mu)/T(b_1,b_2)(\nu) = H(b_1,b_2; \mu,\nu)$ say, is constant on RPC equivalence classes which may be indexed (if every $b : \Theta \to (0,K)$, for some K, is feasible) by $(x,y) \in R_2$. Thus $H = Q \circ RP$ for some function $Q : R^2 \to (0,\infty)$. Choose μ, ν, η and $\delta > 0$ so that $\max\{\delta, \delta \cdot x_i, \delta/y_i\} < K$ where x_i, y_i are arbitrary in $(0,\infty)$, $i = 1,2$. Choose $b_i(\mu) = \delta \cdot x_i$, $b_i(\nu) = \delta/y_i$ and $b_i(\eta) = \delta$. The problem is now easily reduced to an application of Cauchy's equation (Genest, Weerahandi and Zidek 1983). In this last cited work an alternative formula is also derived, under an external Bayesianity-like condition, with $C = 1$ and $\alpha_i = \alpha_i(\theta)$, $i = 1,2$ so that the weights in the resulting logarithmic pool become carrier-dependent. The weights for the various experts, in one application of the formula would be allowed to vary with θ.

By applying the logarithmic pool with this broader domain, one rediscovers Fisher's method for combining significance levels, Nash's product for finding the equilibria of multi-person negotiation problems and the usual rule for combining likelihoods.

6. ACCUMULATING OPINIONS VIA SUPERBAYESIANITY.

This approach has great potential but as yet few results are available. The possibility of exploiting someone else's opinion has been investigated (Lindley, Tversky and Brown 1979; French 1980, Lindley 1982b).

In a more specialized form this becomes the second-guessing problem (Steele and Zidek 1980; Hwang and Zidek 1982).

In this natural but difficult-to-implement approach, the individuals' opinions are treated as data (Winkler 1968; Morris 1977) and an individual's opinion is revised in the light of this data by Bayesian updating.

French (1982), for example, assumes $\lambda = (\ell_1,\ell_2)$ has a jointly normal sampling distribution in the view of the "super-Bayesian", where $\ell_1 = \ln\left[P(E)/P(\overline{E})\right]$ and $\ell_2 = \ln\left[Q(E)/Q(\overline{E})\right]$. This density is conditional on E and, of course, the super-Bayesian's, super-prior, $\Pi(e)$. The super-posterior, $\Pi(E|\lambda)$, is readily derived:

$$\ln\left[\Pi(E|\lambda)/\Pi(\overline{E}|\lambda)\right] = (\mu_E - \mu_{\overline{E}})^T \Sigma^{-1}(\lambda - \tfrac{1}{2}\left[\mu_E + \mu_{\overline{E}}\right]) + \ln\left[\Pi(E)/\Pi(\overline{E})\right]$$

where $\mu_E = E(\lambda|E,\Pi)$. Taking antilogarithms gives (essentially) the logarithmic opinion pool completely equipped with interpretable coefficients.

<u>OBJECTIONS TO THE SUPERBAYESIAN APPROACH.</u> There will not always be an acceptable choice for this role. Each individual can, of course, play the role himself, but this just reconstitutes the original problem in a new form. The task befalling the super-Bayesian is difficult. There is little empirical evidence to guide his choice of a likelihood. And he has to discover the prior information sets upon which the individuals' opinions are based and to model the dependence between them.

7. CONCLUDING REMARKS.

Where feasible, well-directed group interaction with unrestricted feedback would seem to be the most promising approach to consensus. The process is unlikely to produce unanimity of opinion, however, and some method of summarizing or accumulating opinion would then be called for. For the former, the logarithmic opinion pool seems to me to be the most reasonable choice. The super Bayesian approach is the only one available for the latter, but much remains to be done to put it in an implementable form. Finally, it should be remarked that the group decision problem would seem to be of greater practical importance than the consensus problem.

REFERENCES

AUMANN, R.J. (1976) Agreeing to disagree. Ann.Math.Statist., 4, 1236-1239.

BACHARACH, Michael (1975) Group decisions in the face of differences of opinion. Manag. Sci., 22, 182-191

BACHARACH, M. (1979) Normal bayesian dialogues. J.Amer.Statist.Assoc., 74, 837-846.

BERGER, Roger L. (1981) A necessary and sufficient condition for reaching a consensus using DeGroot's method. J.Amer.Statist.Assoc., 76, 415-418.

CHATTERJEE, S. and SENATA, E. (1977) Towards consensus: some theorems on repeated averaging. J.Appl.Prob., 14, 89-97.

DALKEY, N.C. (1972) An impossibility theorem for group probability functions. Santa Monica, Calif: Rand Corporation, P-4862.

DALKEY, N.C. (1975) Towards a theory of group estimation. The Delphi method: techniques and applications. (H.A. Linstone & M. Turoff, Editors) Reading, Mass: Addison-Wesley.

de FINETTI, B. (1937) La prévision: ses lois logiques, ses sources subjectives. Ann.Inst.Henri Poincaré, 7, 1-68.

de FINETTI, B. (1974) Theory of Probability (Vol.I). New York: Wiley.

DeGROOT, M.H. (1974) Reaching a consensus. J.Amer.Statist.Assoc., 69, 118-121.

EISENBERG, E. and GALE, D. (1959) Consensus of subjective probabilities: The pari-mutuel method. Ann.Math.Statist., 30, 165-168.

FRENCH, S. (1980) Updating of belief in the light of someone else's opinion. Jour.Royal.Statist.Soc. A, 143, 43-48.

FINE, T.L. (1973) Theories of probability. New York: Academic Press.

FRENCH, S. (1981) Consensus of opinion. Europ.J.Oper.Res., 7, 332-340.

GENEST, C. (1982) External bayesianity: An impossibility theorem. Tech.Rep. 82-8. Instit.Appl.Math.Statist., UBC.

GENEST, C. (1983) Towards a consensus of opinion. Ph.D. Thesis. U. of British Columbia.

GENEST, C. (1983) Pooling operators with the marginalization property. Tech.Rep. 83-4, Instit.Applied Math.Statist., U. of British Columbia.

GENEST, C., WEERAHANDI, S. and ZIDEK, J.V. (1983) Aggregating opinions through logarithmic pooling. Unpublished manuscript.

HOGARTH, R.M. (1977) Methods for aggregating opinions. Decision Making and Change in Human Affairs. (H. Jungermann & G. deZeeuw, Editors) Dordrecht, Holland: D.Reidel Publishing Co.

HWANG, J.T. and ZIDEK (1982) Limit theorems for out-guesses with mean-guided second guessing. Jour.Applied Prob.,

KRAFT, C.H., PRATT, J.W. and Seidenberg, A. (1959) Intuitive probability on finite sets. Ann.Math.Statist. 30, 408-419.

LEHRER, K. (1978) Consensus and comparison: a theory of social rationality in Foundations and Applications of Decision Theory (Eds: Hooker, C.A. *et al.*) Dordrecht: Reidel Publishing Co., 203-309.

LINDLEY, D.V. (1982a) Scoring rules and the inevitability of probability (with discussion). Int.Statist.Review, 50, 1-26.

LINDLEY, D.V. (1982b) The improvement of probability judgements. Jour.Roy.Statist.Soc. A, 145, 117-126.

LINDLEY, D.V. (1982c) Reconciliation of discrete probability distributions. Unpublished manuscript.

LINDLEY, D.V., TVERSKY, A. and BROWN, R.V. (1979) On the reconciliation of probability assessments (with discussion). J.R.Statist.Assoc. A, 142, 146-180.

MADANSKY, A. (1964) Externally bayesian groups. RAND Memo RM-4141-PR.

MADANSKY, A. (1978) Externally bayesian groups. Unpublished manuscript - University of Chicago.

McCONWAY, K.J. (1981) Marginalization and linear opinion pools. J.Amer. Statist.Assoc., 76, 410-414.

MORRIS, P.A. (1974) Decision analysis expert use. Manag.Sci., 20, 1233-1241.

MORRIS, P.A. (1977) Combining expert judgments: a bayesian approach. Manag.Sci., 23, 679-693.

PRESS, J.S. (1978) Qualitative controlled feedback for forming group judgements and making decisions. J.Amer.Statist.Assoc., 73, 526-535.

RAIFFA, H. (1968) Decision analysis: introductory lectures on choices under uncertainty. Reading, Mass: Addison-Wesley.

ROBERTS, Harry V. (1965) Probabilistic prediction. J.Amer.Statist.Assoc., 60, 50162.

SAVAGE, L.J. (1954) The foundations of statistics. New York: Wiley.

SAVAGE, L.J. (1971) Elicitation of personal probabilities and expectations. J.Amer.Statist.Assoc., 66, 783-801.

STAEL VON HOLSTEIN, C.-A. S. (1970) Assessment and evaluation of subjective probability distributions. Stockhold: Economics Research Institute at the Stockholm School of Economics.

STEELE, M.J. and ZIDEK, J.V. (1980) Optimal strategies for second guessing. Jour.Amer.Statist.Soc., 75, 596-601.

STONE, M. (1961) The linear opinion pool. Ann.Math.Statist., 32, 1339-1342.

TVERSKY, A. and KAHNEMAN, D. (1974) Judgment under uncertainty: heuristics and biases. Science, 185, 1124-1131.

WAGNER, C. (1982) Allocation, lehrer models and the consensus of probabilities. Theory and Decision, 14, 207-220.

WAGNER, C. (1982) Aggregating subjective probabilities: some limitative theorems. Unpublished manuscript.

WEERAHANDI, S. and ZIDEK, J.V. (1978) Pooling prior distributions. Tech.Rep. 78-34. Instit.Appl.Math.Statist., UBC.

WEERAHANDI, S. and ZIDEK, J.V. (1981) Multi-bayesian statistical decision theory. J.R.Statist.Soc. A, 144, 85-93.

WINKLER, R.L. (1968) The consensus of subjective probability distributions. Manag.Sci., 15, B61-B75.

II. MULTIBAYESIANITY: GROUP DECISIONS

1. INTRODUCTION

This talk is about multi-Bayesian statistical decision theory. The fundamental ingredient is a group of Bayesians, B, who have been variously called a "bunch" (Fisher 1972), a "population" (Dickey and Freeman 1975) and a "bevy" (David 1982). These Bayesians can quantify their preferences and degrees of belief as utility functions and prior/posterior distributions, respectively. I suppose they must jointly select a possibly randomized statistical decision rule, δ, by which to choose an action when the data is obtained. The worth of δ to each Bayesian, $\beta \in B$, is determined by his prior or posterior (as appropriate) expected gain (or loss) in utility.

2. BASIC COMPONENTS OF MULTI-BAYESIAN ANALYSIS

This theory shares, with Wald's, the parameter, action, and x-sample spaces, Θ, A, and X, respectively. The counterpart of the loss function is also involved.

Let B denote a possibly infinite label set whose elements identify the relevant characteristics of the Bayesians. Let $u(a|x,\theta,\beta)$ denote β's (bounded) utility for $a \in A$ given $x \in A$, $\theta \in \Theta$ $\beta \in B$. In terms of u, Wald's loss function would be

$$L(a|x,\theta,\beta) = \max\{u(a'|x,\theta,\beta):a' \in A\} - u(a|x,\theta,\beta). \qquad (2.1)$$

B's problem is that of choosing a δ: $\delta(\cdot|x)$ is a probability distribution on the (measurable subsets of A for each $x \in X$. In Wald's theory δ's risk function is

$$r(\delta|\theta,\beta) = \iint L(a\ x,\theta,\beta)\delta(da|x)P_{\tilde{x}}(dx|\theta,\beta) \qquad (2.2)$$

This theory could just as well have been developed in terms of

$$W(\delta|\theta,\beta) \triangleq \iint u(a|x,\theta,\beta)\delta(da|x)P_{\tilde{x}}(dx|\theta,\beta). \qquad (2.3)$$

Conventional Bayesian analysis is concerned with

$$B(\delta|x,\beta) \triangleq \iint u(a|x,\theta,\beta)\delta(da|x)\pi_{\tilde{\theta}}(d\theta|x,\beta) \qquad (2.4)$$

or

$$B(\delta|\beta) \triangleq \iint B(\delta|x,\beta)P_{\tilde{x}}(dx|\beta) \qquad (2.5)$$

according as x is or is not observed. For simplicity, (2.5) is assumed.

CRITERIA AND SOLUTION CONCEPTS. δ's assessment profile is $\beta \to B(\delta|\beta)$ if, as I now assume, x is unknown (Bacharach 1975, p.183). This profile replaces the function $\theta \to W(\delta|\theta,\beta)$, which is central to Wald's theory.

The Wald, multi-Bayesian connection is, in fact, stronger than mere analogy. If utilities can be compared and

$$u(a|x,\theta,\beta) \equiv u(a|x,\theta) \text{ and } P(\cdot|\theta,\beta) = P(\cdot|\theta) \qquad (2.6)$$

for all θ while

$$B = \{[\theta]:\theta \in \Theta\} \qquad (2.7)$$

where $[\theta]$ denotes the probability measure which is degenerate at θ then the multi-Bayesian and Wald theories are identical. 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, p.180).

Because of the nonuniqueness of utility functions, two fundamental approaches have been taken. One (Nash 1950), admits the indeterminacy of utilities and develops solution concepts which do not depend on the comparison of utilities. The other admits such comparisons (c.f. Bacharach 1975). And Savage (1954, p.172) argues that these comparisons would be possible in the case of a jury.

Savage (1954), like Wald, bases his theory on loss rather than gain and introduces an analogue of the risk function (equation (2.2)), β's "(personal) loss":

$$L_p(\delta|\beta) = \max\{B(\delta'|\beta):\delta'\varepsilon D\} - B(\delta|\beta). \qquad (2.8)$$

This is the counterpart of what is sometimes called "regret".

By exploiting the analogy between the Wald and the multi-Bayesian problems which derives from the formal similarity of $W(\delta|\cdot\beta)$ (equation (2.3)) and $B(\delta|\cdot)$ (equations 2.4), (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) states the analogue of "admissibility", as the <u>group principle of admissibility</u>. This notion I will call <u>B-admissibility</u>.

Savage (1954) also defines a solution which, unlike B-admissibility, does require intercomparable utilities, called the <u>group minimax rule</u>, which I will call the B-<u>minimax rule</u> (or rules), namely the δ* (or δ*'s) for which

$$\max_{\beta\varepsilon B} L_p(\overset{*}{\delta}|\beta) = \min_{\delta\varepsilon D} \max_{\beta\varepsilon B} L_p(\delta|\beta). \qquad (2.9)$$

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, $c(\beta)$, I will call a <u>reference utility level</u> (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-<u>Bayes</u> [Bayes if (2.6) and (2.7) hold]. Madansky (see Bacharach 1975, p.186) exploits the Wald-multi-Bayesian connection to show there exists a probability distribution on B, α, (a B-<u>prior distribution</u>) such that the solutions of the group's decision problem must maximize

$$B(\delta|\alpha) = \int B(\delta|\beta)\alpha(d\beta). \qquad (2.11)$$

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 : $c_{sa}(\beta) = \max\{B(\delta'|\beta):\delta'\varepsilon D\}$

Shapley's : $c_{sh}(\beta)$ = β's <u>security level</u> = $\min\{B(\delta'|\beta):\delta'\varepsilon D\}$

Nash's : $c_N(\ )$ = 's <u>current utility level</u>.

Nash explicitly introduces the "agree to disagree" action as an allowable choice and, as well, gives each $\beta\varepsilon B$ the right to precipitate a breakdown in negotiations by insisting on this choice. Nash's theory recognizes that an individual, β, cannot be forced by the group to agree to a choice $\delta\varepsilon D$ for which $B(\delta|\beta) < c_N(\beta)$. Thus the set of Nash feasible solutions consists of those (δ's) for which $B(\delta|\beta) - c_N(\beta) > 0$. In contrast, the B-Bayes and B-minimax solution concepts allow the possibility that the group might choose a $\delta\varepsilon D$ which leaves individual β's with a net loss of utility.

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

$$\prod_{\beta} \left[B(\delta|\beta) - c(\beta) \right]_{+} \tag{2.12}$$

where $c(\beta) = c_N(\beta)$ and, in general, $[x]_+ = \max(x,o)$ (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

$$P(\delta|\alpha) \triangleq \prod_{\beta} \left[B(\delta|\beta) - c_N(\beta)\right]_{+}^{\alpha(\beta)} \tag{2.13}$$

where $\alpha(\beta) \geq 0$ and $\sum \alpha(\beta) = 1$ but is otherwise unspecified. 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 transformation $u \to a u + b$ leaves the solution-set invariant as it does in conventional Bayesian decision-theory.

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

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

may be obtained by maximizing

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

subject to

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

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

Savage's minimax solution is obtained by letting c be his RUL, replacing ΔB by $-\Delta B$ and minimizing the resulting version of (2.15) when $\rho = +\infty$.

3. SUBSAMPLING ASSESSMENTS

B is still required to choose δ. However, I 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.

If s is fixed by some unknown process and is itself unknown and $S = \{\{\beta\} : \beta \in B\}$ the group's choice of δ would have to depend on a comparative analysis of its assessment profile, $B(\delta|\cdot)$; the problem of Section 2 reappears. Nothing is known about the general version of this problem with arbitrary S.

Example 3.1

Here $\tilde{x}$, $\tilde{\theta}$ and β are $p \times 1$ vectors with $\tilde{x}|\theta,\beta \sim N(\theta,\Sigma)$, $\tilde{\theta}|\beta \sim N(\beta,\tau)$ and $\beta \in R_p = B$. $S = \{\{\beta\}:\beta \in B\}$ and

$$u(a|x,\theta,\beta) = C_0(\beta) - D(\beta)(a-\theta)^T Q(a-\theta)$$

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

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

Here $C_1(\cdot)$ is a certain function whose exact form is of no relevance. Also $R = \zeta' Q \zeta$, $\hat{\beta} = \zeta^{-1}\{\delta - \bar{\zeta}x\}$, $\zeta = \Sigma(\Sigma + \tau)^{-1}$ 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. $\hat{\theta} = \zeta\beta + \bar{\zeta}x$. But we are supposing that $s = \{\beta\}$ has not been selected and that every $\beta \in B$ 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, $x|\beta \sim N(\beta, \Sigma + \tau)$ a possible 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 B-inadmissible. However, if $2 < p$ 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). END OF EXAMPLE.

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 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 $\pi(\beta)B(\delta|\beta) + \bar{\pi}(\beta)c_N(\beta)$ 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.

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:

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

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

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

subject to Nash-feasibility:

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

with $\rho = \pm\infty$ and $\rho = 0$ defined in the limit. For $1 \leq \rho$, f is maximized at $\pi(\beta) = 1$ for $\beta \in c B$, where $|s| = n$ and $\sum_s \lambda(\beta) = \max\{\sum_{s'}\lambda(\beta) : |s'| = n,\ s' c B\}$. If, however, $\rho < 1$ the situation is more complicated and Kuhn-Tucker optimization methods would need to be employed.

So if $\rho > 1$, the optimum choices of (π, δ) (π^*, δ^*), satisfy the requirements $\pi^*(\beta) = 1$, for $\beta \in s^*$ while $s^* c B$, $|s^*| = n$ and δ^* jointly maximize $\sum_s \alpha(\beta)[\Delta B(\delta|\beta)]^{\rho}$. For $\rho = +\infty$ this means an $s = s^* c\ B$, $|s^*| = n$ which contains the β^* for which $\alpha(\beta)\Delta B(\delta|\beta)$ is jointly maximized in β and δ. For $\rho < 1$, little can be said.

This analysis reveals a striking feature of the Madansky-Bacharach-Harsanyi-Blackwell-Girshick (B-Bayesian) and other solutions for $\rho \geq 1$. 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.

The Kalai-Nash solution $(\rho = 0)$ is quite different from the B-Bayesian results which were just described. In this case (3.2) becomes

$$\Pi\left[\pi(\beta)\right]^{\alpha(\beta)} \quad \Pi\left[\Delta B(\delta|B)\right]^{\alpha(\beta)} \tag{3.4}$$

Therefore the optimal choices of π and δ are made, each without regard to the other. The optimal δ's under subsampling are just the Nash solutions in the original problem. The optimal π is easily evaluated .

The maximal value of the objective function given in (3.4) is attained at

$$\begin{aligned}\pi_i &= 1, \quad i = N - J + 1, \ldots, N \\ &= (N - J)\alpha_i/(\alpha_1 + \ldots + \alpha_{N-J}), \quad \text{otherwise},\end{aligned} \tag{3.5}$$

where J is the first $k = 0, 1, \ldots, n$ for which $(n-k)\alpha_{n-k} - \alpha_1 - \ldots - \alpha_{N-k} \leq 0$ after the α_i's have been ordered $\alpha_1 \leq \ldots \leq \alpha_N$.

EXAMPLE 3.2

Suppose $N = 19$, $n = 10$ and $\alpha = (0.01, 0.01, \ldots, 0.01, 0.05, 0.20, 0.25, 0.35)$. So $J = 4$ $\pi_i = 1$ for $i \geq 16$ while $\pi_i = 0.4$, $i < 16$ for an expected sample size of $n = 10$.||

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 B-Bayesian case, in terms of expected utility gains as well.

4. SUPERPOPULATION ASSESSMENTS

Suppose B is a subset of $\mathcal{B}$, a set which will be called a superpopulation. It is assumed that $\delta = \delta(\cdot|x,B)$ has been specified.

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 be 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. This is one interpretation of the idea underlying superpopulation evaluation.

Choosing an evaluation criterion is problematical. In general there cannot exist an objective, i.e. group utility function as Arrow's celebrated theorem shows (Arrow 1966; see also Bacharach 1975). Such a criterion may, however, exist in individual cases (like that of Steele and Zidek 1980; see also Savage 1954, p.172) when equation (2.6) holds. In what might be called the "super-Bayesian" approach, the criterion would be subjective, the utility function of an individual (the "investigator" in the terminology of Lindley, Tversky and Brown 1979) selected from the superpopulation.

With a criterion selected and (x,B) treated, formally, as "data", the classical decision problem re-emerges as the next step in a potentially infinite regress.

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 $\tilde{\beta}|\theta \sim N(\theta,\Gamma)$ independently of $\tilde{x}|\theta \sim N(\theta\ \Sigma)$. The Bayesian solution concept led to the rule $t(x|\beta) = \zeta\beta + \bar{\zeta}x$ in earlier analysis. The efficacy of this procedure is the issue of interest here.

Because $t(\tilde{x}|\tilde{\beta})|\theta \sim N(\theta,\Delta)$ 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 $2 < p$ and $\hat{\theta} = \hat{\theta}(\tilde{x}|\tilde{\beta})$ 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, $\Gamma = \tau = \sigma^2\Sigma$, $\Sigma = I$ and the resulting value of ζ is estimated by $\hat{\zeta} = [(p-2)/||x-\beta||^2]I$, the James-Stein estimator is obtained: $t(x|\beta) = \hat{\zeta}\beta + \hat{\bar{\zeta}}x$. It is an S-affine equivariant rule under the transformations $x \rightarrow \alpha\ \ x + b$ and $\beta \rightarrow \alpha\ \ \beta + b$ where α is orthogonal and $b \epsilon R_p$. So the James-Stein estimator is translation equivalent in this extended sense even though it is not in the more familiar setting.

REFERENCES

Arrow, K.J. (1966). Social choice and individual values. New York: John Wiley and Sons, Inc., Second Edition.

Bacharach, M. (1975). Group differences in the face of differences of opinion. Management Science, 22, 182-191.

David, A.P. (1982). "Intersubjective statistical models" in Exchangeability in Probability and Statistics. (eds. G. Koch and F. Spizzichino) North Holland Publishing Co., 217-232.

Dickey, J. and Freeman, P. (1975). Population-distributed personal probabilities. JASA, 70, 362-364.

Fisher, L. (1972). The behavior of Bayesians in bunches. Amer. Statist., No.5, 19-20.

French, S. (1981). Consensus of opinion. European Jour. of Operational Research, 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. New York: Marcel Dekker, Inc.

Harsanyi, J.C. (1977). Rational behavior and bargaining equilibrium engames and social situations. Cambridge: Cambridge University Press.

Harsanyi, J.C. (1955). Cardinal welfare, individual ethics and interpersonal comparisons of utility. Jour. of Political Economy, 63, 309-321.

Jones, A.J. (1980). Game theory: mathematical models of conflict. Toronto: John Wiley and Sons.

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., 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. A, 142, 146-180.

Luce, D. and Raiffa, H. (1957). Games and decisions. New York: Wiley.

Nash, J.F. Jr. (1950). The bargaining problem. Econometrica, 18, 155-162.

Rapoport, A. (1970). N-person game theory: concepts and applications. Ann Arbor: University of Michigan Press.

Savage, L.J. (1954). The foundations of statistics. New York: Wiley.

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. Judgement under uncertainty: heuristics and biases. Science, 1974, 135, 1124, 1131.

Wald, A. (1950). Statistical Decision Function. New York: John Wiley and Sons, Inc.

Wald, A. (1939). Contributions to the theory of statistical estimation and testing hypothesis. Ann. Math. Statist., X, 299-326.

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

White, D.J. (1976). Fundamentals of decision theory. New York: North Holland.

Zidek, J.V. (1982) Elements of multi-Bayesian decision theory. Unpublished manuscript (with Weerahandi, S.).

III. MULTIBAYESIANITY: APPLICATIONS

1. INTRODUCTION

In previous lectures I surveyed solutions of the consensus and multi-Bayesian decision problems in the context of statistics. In this talk illustrative applications will be presented.

2. ESTIMATING THE NORMAL MEAN (Weerahandi and Zidek 1982)

I will begin by recalling Nash's solution (Nash 1950). Suppose two Bayesians, β_i, $i = 1,2$, must jointly select a (possibly randomized) decision rule, δ, for which their expected gains-in-utility are $u_i(\delta) - c_i$, $i = 1,2$. Here c_i, $i = 1,2$, is the utility of their current assets.

Any Nash solution is required to maximize $\prod_{i=1}^{2} [u_i(\delta) - c_i]$ subject to Nash-feasibility: $u_i(\delta) - c_i > 0$. This is implied by certain weak assumptions. In particular, each β is required to honestly disclose his preferences and degrees of belief and to believe the other's disclosure.

Bayesian $\beta = i$, $i = 1,2$, is assumed to have a posterior distribution with the multivariate normal density function given by

$$\pi(\underset{\sim}{\theta}|i) \propto \exp[-\tfrac{1}{2}(\underset{\sim}{\theta} - \underset{\sim}{\theta}_i)^T \underset{\sim}{\Sigma}_i^{-1}(\underset{\sim}{\theta} - \underset{\sim}{\theta}_i)] \tag{2.1}$$

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

$$u(\hat{\underset{\sim}{\theta}}|\underset{\sim}{\theta},i) \propto \exp[-\tfrac{1}{2}(\underset{\sim}{\theta} - \hat{\underset{\sim}{\theta}})^T \underset{\sim}{W}_i^{-1}(\underset{\sim}{\theta} - \hat{\underset{\sim}{\theta}})] \tag{2.2}$$

where $\underset{\sim}{W}_i$ is a positive definite matrix of constants. I assume for simplicity that $W_i = \Lambda - \sum_i > 0$ for some Λ but the results I will describe are easily generalized.

The expected gain-in-utility for Bayesian i of the estimate $\underset{\sim}{\theta} = \hat{\underset{\sim}{\theta}}$ is easily shown to be, after a convenient rescaling,

$B(\hat{\underset{\sim}{\theta}}|i) = \phi(\underset{\sim}{\theta}_i - \hat{\underset{\sim}{\theta}})^T \Lambda^{-1}(\underset{\sim}{\theta}_i - \hat{\underset{\sim}{\theta}})$ where $\phi(u) = \exp(-\tfrac{1}{2}u)$, $u \geq 0$. Finally, $B(\delta|i) = \int B(\hat{\underset{\sim}{\theta}}|i)\delta(d\hat{\underset{\sim}{\theta}}) = \Delta B(\delta|i)$.

It will be assumed, quite realistically, 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.

Denote the "agree to disagree" decision by $\hat{\theta}_o$ and assume that $B(\hat{\theta}_o|i) = c(i) = 0$ for $i = 1,2$.

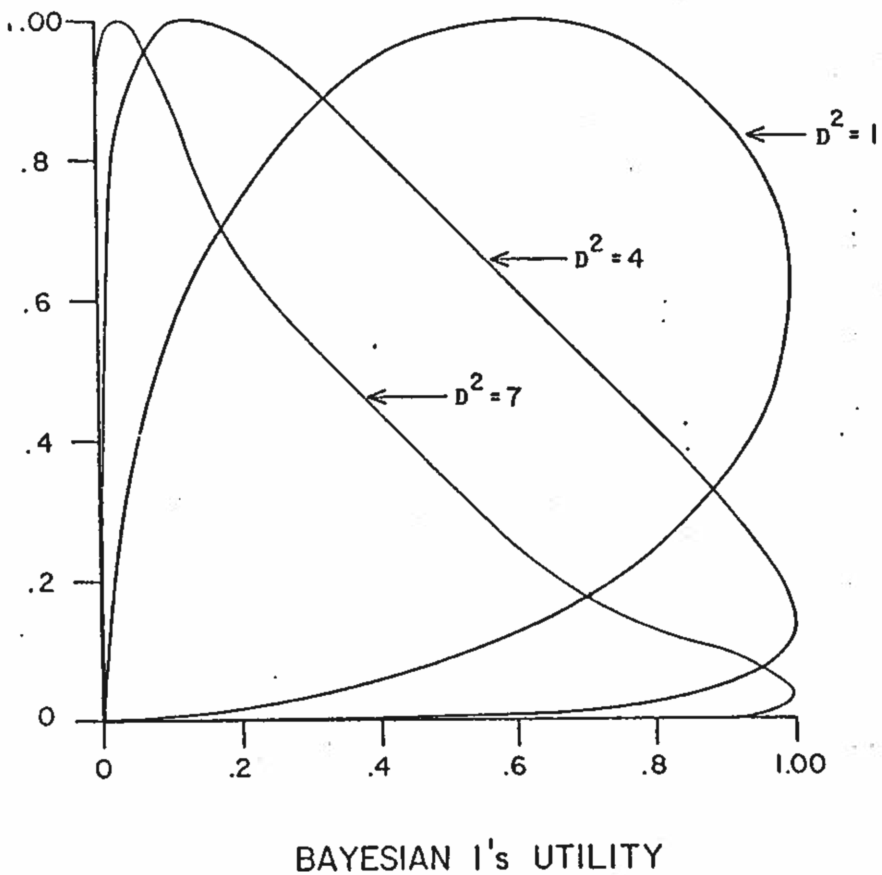
<u>B-ADMISSIBLE ESTIMATION RULES</u>. Of fundamental interest is the set S, consisting of all 2-tuples, $(B(\hat{\underset{\sim}{\theta}}|1), B(\hat{\underset{\sim}{\theta}}|2))$ obtained by varying $\hat{\underset{\sim}{\theta}} \in R_p \cup \{\hat{\theta}_o\}$. The convex hull of S, say $\bar{S}$, would consist of all utility pairs which can be achieved by adopting randomized rules, i.e. $\{(B(\delta|1), B(\delta|2)) : \delta \text{ randomized}\}$. Let δS represent S's boundary and P, that portion of δS which corresponds to the elements, $\hat{\underset{\sim}{\theta}}$, of the class, C, of nonrandomized rules which are B-admissible within S. The sets S, δS, P and C will be characterized in this subsection.

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

Obviously, $S \subset [0,1]^2$ and $u_1 = 1$ is attained at $\hat{\underset{\sim}{\theta}} = \underset{\sim}{\theta}_1$ so that the corresponding u_2 is uniquely determined. However, if $u_1 = \lambda$, $0 < \lambda < 1$, $\hat{\underset{\sim}{\theta}}$ and hence u_2 is not uniquely determined; in this case, if $p = 1$, $\hat{\underset{\sim}{\theta}}$ may have either of two possible values while, if $p > 1$, it is merely constrained to lie on an ellipsoid. Thus, if $p = 1$, S is a curve while, if $p > 1$ it is a compact subset of $[0,1]^2$. The precise form of S is of no relevance to our analysis, but our results will be more easily interpreted by referring to the Figure.

The set of u_2-values corresponding to $u_1 = u_1(\hat{\underset{\sim}{\theta}}) = c$, i.e. $u_2 = u_2(\hat{\underset{\sim}{\theta}})$ is the cross-section of S at $u_1 = c$. The maximum and

A: UTILITY VECTORS CONTAINED WITHIN THE BOUNDARY OF THOSE OF NON RANDOMIZED RULES FOR VARYING MAHALANUBIS DISTANCES, D^2.



minimum u_2-values within this cross-section are the points of ∂S at $u_1 = c$. By varying c, $0 \leq c \leq 1 = u_1(\hat{\theta}_1)$, these extrema generate ∂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. The result is: ∂S <u>consists of the distinguished point</u> $(0,0)$ <u>and the utility pairs</u>, (u_1, u_2) <u>corresponding to</u> $\hat{\theta} = \theta_i$, 1,2 <u>and the elements of</u>

$$\{\hat{\theta}^* = (1+\lambda)^{-1}(\lambda\theta_1 + \theta_2).\ -\infty < \lambda < \infty,\ \lambda \neq -1\} \tag{2.3}$$

The quantity λ which indexes the set in (2.3) 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 (2.3), (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 ∂S is now easy to determine. It is easily shown that

$$du_2/du_1 = -\lambda(u_2/u_1) \tag{2.4}$$

and

$$d^2u_2/du_1^2 = (u_2/u_1^2)(\lambda[1+\lambda] + 2/\Delta_1' \tag{2.5}$$

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

Observe that as $\lambda \to -1$, $(u_1, u_2) \to (0,0)$. From (2.4) 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 $\hat{\theta} = \hat{\theta}_2$. 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: ∂S is a continuous parametric curve in $[0,1]^2$ which consists of two branches. The upper branch increases from $(0,0)$ to $(u_1(\underset{\sim}{\theta}_2), u_2(\underset{\sim}{\theta}_2))$ and then decreases to $(u_1(\underset{\sim}{\theta}_1), u_2(\underset{\sim}{\theta}_1))$. The lower branch increases from $(0,0)$ to $(u_1(\underset{\sim}{\theta}_1), u_2(\underset{\sim}{\theta}_1))$.

The nonrandomized rules which are B-admissible in S correspond to the utility pairs in $P \subset \partial S$. Obviously, P is that part of ∂S's upper branch which joins the utility pairs corresponding to $\underset{\sim}{\hat{\theta}} = \underset{\sim}{\hat{\theta}}_1$ and $\underset{\sim}{\hat{\theta}} = \underset{\sim}{\hat{\theta}}_2$. The Nash solution will be nonrandomized if the function u_2 of u_1 defined by P, is concave. This will be the case if $d^2u_2/du_1^2 \le 0$ on P. Equation (2.5) implies: P determines a concave function, u_2 of u_1 if and only if

$$D^2 = (\underset{\sim}{\theta}_1 - \underset{\sim}{\theta}_2)^T \Lambda^{-1} (\underset{\sim}{\theta}_1 - \underset{\sim}{\theta}_2) \le 4 \qquad (2.6)$$

The quantity D^2 in (2.6) 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 S are easily derived, and these appear in Figure A.

Observe that each of the sets, S, illustrated in Figure A: (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 $\underset{\sim}{\hat{\theta}} = \underset{\sim}{\hat{\theta}}_2(\underset{\sim}{\hat{\theta}})$, (iv) has a concave (respectively, convex) increasing upper (lower) boundary to the left of the point which represents $\underset{\sim}{\hat{\theta}} = \underset{\sim}{\theta}_2(\underset{\sim}{\theta})$. All S's must have features (i) - (iv). The only qualitative difference that

may exist between any two S's is found by inspecting their boundaries between $\underset{\sim}{\hat{\theta}} = \underset{\sim}{\hat{\theta}}_1$ and $\underset{\sim}{\hat{\theta}} = \underset{\sim}{\hat{\theta}}_2$. Figure A portrays the fact that this portion of the boundary will be a concave curve if and only if $D^2 \le 4$.

RANDOMIZED DECISION RULES. An inspection of the curve in Figure A 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 ($\lambda = \frac{1}{2}$) in (2.3)) yields a relatively small increase (.42) in utility to either Bayesian. So a coin toss leading (approximately) to either $\underset{\sim}{\hat{\theta}} = \underset{\sim}{\theta}_1$ or $\underset{\sim}{\hat{\theta}} = \underset{\sim}{\theta}_2$ 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 cooperate and jointly adopt a randomized rule δ.

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

Recall that maximizing P, the Nash product, amounts to maximizing the function $(u_1,u_2) \to u_1^{\alpha_1} u_2^{\alpha_2}$ over $\bar{S}$, the convex hull of S. It can be shown that the optimal nonrandomized rule, i.e. the rule which chooses $\underset{\sim}{\bar{\theta}}_\alpha = \alpha_1\underset{\sim}{\theta}_1 + \alpha_2\underset{\sim}{\theta}_2$ with certainty, is globally optimal if and only if either

(i) $D^2 = (\underset{\sim}{\theta}_1 - \underset{\sim}{\theta}_2)^T \Lambda^{-1} (\underset{\sim}{\theta}_1 - \underset{\sim}{\theta}_2) \le 4$ or

(ii) $D^2 > 4$ and $\bar{s} \le (D^2/2)(\mathrm{Max}\{\alpha_1,\alpha_2\} - \frac{1}{2})$ where $\bar{s}$ is the positive root of

$$\tan h(s)(s)^{-1} = 4D^{-2}. \qquad (2.7)$$

When neither condition (i) or (ii) holds, the Nash solution is randomized and it consists of choosing the estimates $(\frac{1}{2} + 2\bar{s}D^{-2})\theta_1 + (\frac{1}{2} - 2\bar{s}D^{-2})\theta_2$ and $(\frac{1}{2} - 2\bar{s}D^{-2})\theta_1 + (\frac{1}{2} + 2\bar{s}D^{-2})\theta_2$ with probabilities $\bar{\alpha}$ and $1 - \bar{\alpha}$, respectively, where $\bar{s}$ is the positive or negative root of equation (2.7) according as $\bar{\alpha} > 1/2$ or $\bar{\alpha} < 1/2$, and

$$\bar{\alpha} = \left[\tfrac{1}{2}(\alpha_1 - \alpha_2)D^2/\bar{s} + 1\right] / 2 \qquad (2.8)$$

As $D^2 \to \infty$, $D^2/(4\bar{s}) \to 1$, $\bar{\alpha} \to \alpha_1$ and the Nash solution converges to the rule, "choose θ_1 and θ_2 with probabilities α_1 and α_2, respectively. On the other hand, $D^2 \to 4$ implies $\bar{s} \to 0$, and when $\bar{s} < D^2/2\ (\mathrm{Max}\{\alpha_1,\alpha_2\} - \frac{1}{2})$, the nonrandomized rule, $\alpha_1\theta_1 + \alpha_2\theta_2$, becomes optimal. The convergence of D^2 to 4 is insured by increasing the amount of data which is available.

3. SECOND GUESSING STRATEGIES

I will now present an example wherein the "second guesser" in the role of a super-Bayesian exploits his adversary by taking advantage of his special position in the scheme of things to win a guessing contest.

The structure of our guessing model can be described by a system of four p-vectors.

Target values:	$(\theta_1,\theta_2, \ldots, \theta_p)$	=	θ
First Guess:	$(X_1,X_2, \ldots, X_p)$	=	X
Second guesser's hunch:	$(Y_1,Y_2, \ldots, Y_p)$	=	Y
Second guess:	$(G_1,G_2, \ldots, G_p)$	=	G

The θ_i represent the real values to be guessed. The X_i are guesses made by the person who goes first, and all these are assumed to be available to the second guesser before he acts. The Y_i represent the second guesser's best estimate of the θ_i. Finally, the G_i are the guesses to

be announced by the second guesser. Our principal task is to determine how G should be based on X and Y.

The objective of each player is to come closer to θ than his opponent, so we begin by setting V = number of times in p the second guess is closer than the first.

Assume the existence of the joint, continuous distribution, $\theta|X,Y$ which could be purely subjective or involve "objective components". Let $\nu = (\nu_1, \ldots, \nu_p)$ be the medians of the marginal distributions, $\{\theta_i|X,Y\}$. Then Hotelling's strategy, G^ϵ, with $G_i^\epsilon = X_i + \epsilon$ or $X_i - \epsilon$ according as $X_i < \nu_i$ or not, is optimal.

A simpler, sub-optimal strategy is "hunch-guided", unlike Hotelling's which is "median guided": $\tilde{G}_i^\epsilon = X_i + \epsilon$ or $X_i - \epsilon$ according as $X_i < Y_i$, or $X_i > Y_i$. Then (Three-Quarter Theorem): if $\tilde{X} = X - \theta$ and $\tilde{Y} = Y - \theta$ are identically distributed, independent and symmetric about zero, then the hunch-guided guess has probability $\frac{3}{4}$ of winning as $\epsilon \to 0$. This gives a lower bound for the performance of the optimal strategy. This may only be from the optimistic perspective of the second-guesser. I have enjoyed considerable empirical success at second-guessing, using this strategy, however.

When $\theta|\mu \sim N(\mu 1_p, \sigma_\theta^2 I_p)$ $\mu \sim N(\mu_0, \sigma_\mu^2)$ and $X,Y|\theta,\mu \sim N(\theta(I_p,I_p),\Gamma)$,

$$\Gamma = \begin{pmatrix} \sigma_X^2 I_p & 0 \\ 0 & \sigma_Y^2 I_p \end{pmatrix},$$

the median-guided strategy reduces to a more explicit form. Under the preceding Gaussian model the Hotelling strategy is

$$\begin{aligned} G_i^* &= X_i + \epsilon \quad \text{if } X_i < \gamma\mu_0 + (1-\gamma)[\beta\bar{X} + (1-\beta)\bar{Y}] \\ &\qquad + (1-\alpha)[\beta(X_i - \bar{X}) \\ &\qquad + (1-\beta)(Y_i - \bar{Y})] \\ &= X_i - \epsilon \quad \text{otherwise,} \end{aligned}$$

where $\gamma = \sigma^{-2}(\sigma^{-2} + \sigma_X^{-2} + \sigma_Y^{-2})^{-1}$, $\quad \sigma^2 = \sigma_\theta^{\,2} + p\sigma_\mu^{\,2}$,

$\alpha = \sigma_\theta^{-2}(\sigma_\theta^{-2} + \sigma_X^{-2} + \sigma_Y^{-2})^{-1}$

and $\quad \beta = \sigma_X^{-2}(\sigma_X^{-2} + \sigma_Y^{-2})^{-1}$.

The basic part is the mixture of means, $\gamma\mu_o + (1-\gamma)[\beta\bar{X} + (1-\beta)\bar{Y}]$ which is perturbed on trial i by the "mixture" of residuals $\alpha \cdot 0 + (1-\alpha)[\beta(X_i - X) + (1-\beta)(Y_i - Y)]$. The coefficient $\beta = \sigma X^{-2}(\sigma X^{-2} + \sigma Y^{-2})^{-1}$ appearing in these mixtures is near 0, $\frac{1}{2}$, or 1 accordingly as $\sigma_X^{\,2}\sigma_Y^{-2}$ is near ∞, 1 or 0. This ratio is one natural measure of the relative abilities of the two guessers, and this interpretation is reinforced by considering the extreme cases. When $\sigma_X^{\,2}\sigma_Y^{-2} \sim \infty$ the first guess is essentially ignored, and when $\sigma_X^{\,2}\sigma_Y^{-2} \sim 0$ it is the hunch that is ignored. This last case is of particular interest since it corresponds to trying to outguess a far better informed adversary.

The strategies just derived have the drawback that they are functions of $\mu_0, \sigma_\mu^{\,2}, \sigma_\theta^{\,2}, \sigma_X^{\,2}$, and $\sigma_Y^{\,2}$. Although the magnitude of μ_0 and of the relevant variance ratios may be sufficiently understood for some applications, their exact values cannot always be specified. They may be determined empirically, however, (see Steele and Zidek 1980). In particular, if $\sigma_\mu^{\,2} = 0, \sigma_Y^{\,2} = \infty, \sigma_X^{\,2}$ is known and $\sigma_\theta^{\,2}$ unknown, the inferred optimal strategy becomes "mean-guided" : $G_i^* = X_i + \varepsilon$ or $X_i - \varepsilon$ according as $X_i < \bar{X}$ or $X_i \geq \bar{X}$.

This has intuitive appeal even outside the context of Gaussian distributions, and if p is large, the second guesser has a considerable apparent advantage (Hwang and Zidek 1982). Let G1 and G2 represent the number of wins for the first and second guessers, respectively. Assume

$X_i - \theta_i$ are i.i.d., continuous and symmetrically distributed about 0. Let $\Delta_i = \bar{\theta} - \theta_i$ and $\zeta(x) = \lim|\{i : |\Delta_i| < x\}|/p$ which, I will assume, exists. Then $G2/p \to \frac{1}{2} + \int_0^\infty \zeta(x)\,dF(x)$ (a.e) where F is the common d.f. of $X_i - \theta_i$, $i = 1, \ldots, p$. If the θ_i's are i.i.d., continuous symmetric with mean 0 and d.f. G, $G2/p \to \frac{1}{2} + \int_0^\infty G(x) - G(-x)\,dF(x)$ (a.e.) and so if $G(x) \geq F(x)$ and $\leq F(x)$ for $x \geq 0$ and ≤ 0, respectively, $G2 \to \frac{1}{2} + \beta$ (a.e.) where $\beta \geq \frac{3}{4}$. This is an analogue of the "$\frac{3}{4}$-Theorem" of Steele and Zidek (1980).

Under severe sufficient regularity conditions limiting distributions for G1 may be obtained:

(i) <u>if</u> $\theta_i = m$ <u>for all</u> i, $G1/p^{\frac{1}{2}} \to f(o)|Z|$ <u>in distribution where</u> $Z \sim N(0,1)$ <u>and</u> f <u>is the density of the</u> $X_i - \theta_i$ and

(ii) <u>if</u> $f(o) > 0$, $\left[G1 - \sum p_i\right] / \left[\sum p_i(1-p_i)\right]^{\frac{1}{2}} \to Z \sim N(0,1)$ <u>in distribution where</u> $P_i = P(0 < X_i - \theta_i < |\Delta_i|$, $i = 1, \ldots, p$ <u>provided that</u> $\sum f(\Delta_i)/p \to 0$ as $p \to \infty$

The first of these theorems applies when the θ's are concentrated at a point, the second when they are widely distributed. Their great dissimilarity suggests the general result will be quite complex. In any case, all of these results establish the great advantages which obtain for the second guesser.

Little more is known about second guessing. Pittenger (1980) indicates another direction for the theory.

Hwang, J.T. and Zidek, J.V. (1982). Limit theorems for out-guesses with mean-guided second guessing. Jour. Applied Prob. 15

Nash, J.F. Jr. (1950). The bargaining problem. Econometrica, 18, 155-162.

Pittenger, A. (1980). Success probabilities for second guessers. J. Appl. Prob. 17, 1133-1137.

Steele, J.M. and Zidek, J.V. (1980). Optimal strategies for second guessers. J. Amer. Statist. Assoc. 75, 596-601.

Weerahandi, S. and Zidek, J.V. (1982). Elements of multi-Bayesian decision theory. Unpublished manuscript.