The Annals of Statistics
1988, Vol. 16, No. 4, 1594–1608

CONTROLLING RISKS UNDER DIFFERENT LOSS FUNCTIONS: THE COMPROMISE DECISION PROBLEM[1]

BY PETER J. KEMPTHORNE

Massachusetts Institute of Technology

Controlling Bayes and/or minimax risks under possibly different loss functions is formulated as a problem faced by two or more statisticians who must compromise and agree on the use of a single decision procedure. The theory characterizing solutions to Bayes compromise problems and minimax-Bayes compromise problems is presented. In a Bayes compromise problem, Bayes risks under different prior distributions and/or loss functions are minimized simultaneously. In a minimax-Bayes compromise problem, a Bayes risk under some loss function for a given prior distribution and a maximum risk under a possibly different loss function are minimized simultaneously.

1. Introduction and summary. Suppose a group is faced with a compromise decision problem whereby its constituents must agree on the use of a single decision procedure. For example, a venture capital firm might have to decide which of several projects to finance. The partners of the firm may not have the same preferences over the projects because some partners prefer projects with regular cash flows providing a consistent income while others prefer those which delay payments until the end of the investment period for greater capital appreciation. Also, the partners' opinions may differ as to the uncertainties of projects' cash flows. Consequently, to agree on a group decision, some or all of the partners will have to compromise.

Section 2 describes the assumptions and notation for general compromise decision prolems. For a given problem, assume each individual is a statistician who can specify his loss function and choose a decision principle, such as the Bayes or minimax principle. For simplicity we assume there are just two statisticians. Because the optimal procedure for one individual is generally not optimal for another, the choice of a decision procedure by the group will require some compromise. We address the characterization of decision procedures constituting principled compromises.

In Section 3 we treat the theory of Bayes compromise problems when both apply the Bayes principle relative to different prior distributions and/or loss functions. Each statistician prefers procedures with smaller Bayes risks. The solutions to these problems characterize the class of compromise decision procedures satisfying Savage's (1954) "group principle of admissibility," a Pareto-optimality criterion; see Theorems 3.1 and 3.2.

The results of Section 3 complement those of Weerahandi and Zidek (1981, 1983), de Waal, Groenwald, von Zyl and Zidek (1983) and Zidek (1984). They

Received July 1985; revised August 1987.
[1]Research supported in part by National Science Foundation Grants MCS-82-01820, SES-84-01422 and DMS-86-05819.
AMS 1980 *subject classifications*. Primary 62C05; secondary 62C25.
Key words and phrases. Bayes inference, group decision analysis, minimax, multiple objective decision analysis.

apply the theory of n-person bargaining games to the multi-Bayesian decision problem and examine the Nash (1950) solution as a group compromise procedure. While the motivation for the Nash solution will be compelling in certain compromise problems, in others it will not. In general, the Nash solution is just one compromise procedure in the class of principled compromises—the group-admissible decision procedures.

Also, Weerahandi and Zidek (1981), page 86, report a result in an unpublished manuscript of Madansky (1978) which, if not identical, is closely related to Corollary 1 to Theorem 3.2. When Bayesians have identical loss functions, then their group decision is admissible if (and only if) they act like a single Bayesian with a prior distribution equal to a linear pooling of their individual priors. This result together with Corollary 2, concerning the analogous result interchanging loss functions and prior distributions, characterize when group-admissible behavior is consistent with acting like an individual decision-maker with a group loss function and group prior distribution, constructed by separately pooling constituent's loss functions and pooling their prior distributions.

While interesting independently, the theory of Bayes compromises provides the necessary foundation for solving minimax-Bayes compromise problems. In such problems, some of the statisticians apply the Bayes principle while others apply the frequentist, minimax principle. Section 4 gives a formal development of the relevant theory, first treated in Kempthorne (1983). As in the case of Bayes compromises, the collection of solutions to all such compromise problems characterizes the class of Pareto-optimal group decision procedures. We show that a minimaxist, engaged in a compromise problem with Bayesians, must act like a Bayesian in a Bayes compromise problem for the decision to be Pareto optimal. In such a problem, the minimaxist's prior distribution is least favorable to the compromise. We prove that this least favorable distribution has discrete structure for a large class of problems. Then it is susceptible to numerical solution; see Kempthorne (1987a).

The minimax-Bayes compromise problem provides a reasonable basis for characterizing procedures which control risks under different loss functions. While stated with reference to two statisticians with different objectives and perspectives, the problem also applies to a single statistician who addresses a decision problem and wishes to control simultaneously the maximum and Bayes risks under possibly different loss functions. The special case of the minimax-Bayes compromise problem where the two loss functions are equal is well suited to the frequentist decision theorist who has carefully specified a single loss function and who also has some prior knowledge about the unknown parameter. Not having complete confidence about his specification of a prior measure for the parameter, however, he seeks to control the maximum risk of the procedure over the entire parameter space while still achieving a small Bayes risk with respect to the prior measure. The solutions to such minimax-Bayes compromise problems incorporate prior knowledge about the parameter but are robust against gross inaccuracies in these prior beliefs.

The particular case of the minimax-Bayes compromise problems with two statisticians having the same loss function was considered first by Hodges and

Lehmann (1952). The procedures which solve such problems, "restricted Bayes solutions," are Bayes for the single loss with respect to the prior measure which is a mixture between the prior distribution of the Bayesian and another prior distribution. This is just the special case of Theorem 4.4 under the condition of Corollary 1 to Theorem 3.2.

Bickel (1983) addresses the special minimax-Bayes compromise problem of estimating the mean of a univariate normal random variable where the Bayesian's prior measure concentrates all its mass at the origin, and both losses are squared error. Bickel proves that the solution procedures which minimize the maximum squared-error risk subject to doing well at the origin are Bayes with respect to discrete prior measures. This result is a special case of Theorem 4.6(b).

Several important and popular problems in decision theory are minimax-Bayes compromise problems. Approximate solutions to minimax-Bayes compromise problems for estimating a multivariate normal mean when the loss functions are equal to squared error have been the subject of considerable research; see Efron and Morris (1971, 1972), Berger (1976, 1980, 1982a, b), Chen (1983) and George (1986). The approaches used in these papers do not approximate the prior measures characterizing the exact solutions. However, they can provide reasonably tractable procedures which perform at levels close to those of the exact solutions. Kempthorne (1986, 1987b) extends the theory of Section 4 to obtain partial characterizations of the measures specifying the exact solutions.

Regarding minimax-Bayes compromise problems where the loss functions are different, Kempthorne (1986, 1987a) uses the theory of Section 4 to characterize minimax squared-error risk preliminary test estimators of a normal mean. These estimators bound the probability of using estimates associated with a Type I error in the preliminary test. The theory of minimax-Bayes compromises applies because this probability can be expressed as a Bayes risk under a zero–one loss.

In another application Kempthorne (1988) uses the theory of minimax-Bayes compromises to characterize admissible dominating procedures of inadmissible procedures in simple decision problems with one loss function and one statistician. Dominating an inadmissible procedure is expressible as a constraint on the maximum risk for a second loss function. A class of optimal dominators consists of those procedures minimizing Bayes risks under the loss function of the original problem.

Among alternate approaches to controlling risks under different loss functions one uses a vector risk function whose components are the individual risks under each loss function. Cohen and Sackrowitz (1984) characterize admissible procedures in a general setting and Meeden and Vardeman (1985) do so for set estimation problems. Their admissible procedures can be interpreted as Bayes compromise procedures, or their limits. See Kempthorne (1985) for details. Brown (1975), using a distinctly different approach, investigates the performance of estimation procedures when the loss function is incompletely specified.

2. Components of the compromise decision problem. Let X be a random variable whose distribution is in a family indexed by the parameter θ. Suppose that the decision problem concerns drawing inferences about θ given an observation of X. Possible inferences include an estimate of θ, a confidence/credible region for θ or a prediction of a future observation of X. When the decision problem has multiple objectives, several inferences might be made collectively.

Let $\mathscr{X}$ denote the sample space and $\mathscr{F}$ a σ-field on $\mathscr{X}$ for which X is measurable. For each $\theta \in \Theta$, the parameter space, let the distribution of X be specified by $P(\cdot|\theta)$, a probability measure on $\mathscr{F}$, which is absolutely continuous with respect to μ, a σ-finite measure on $(\mathscr{X}, \mathscr{F})$.

Let $\mathscr{G}$ denote a σ-field on Θ which includes all atoms. Let A denote the collection of all inferences with σ-field $\mathscr{A}$. We assume that both A and Θ are metrisable spaces, locally compact and the union of a denumerable family of compact spaces. For many estimation problems A is equal to Θ. For testing, A is often a set of two points corresponding to "accept" and "reject." For multiobjective problems the action space can be more general, consisting of pairs or triples of simultaneous actions such as estimates of parameters, predictions of future values of X and outcomes for hypothesis tests. We note that in some decision problems A will include the action making no inference or even the action not to observe X. In compromise decision problems, the first would correspond to not compromising.

A decision procedure δ identifies an action in A for each value $X = x$ which could be observed. For any procedure δ, we assume that (i) the action it identifies given $X = x$ is described by a probability measure $\delta(\cdot|x)$ on $(A, \mathscr{A})$ and (ii) for any $A_0 \in \mathscr{A}$, $\delta(A_0|x)$ is an $\mathscr{F}$-measurable function of x.

Suppose two statisticians are presented with the problem of agreeing to use a single decision procedure from $\mathscr{D}$, the collection of all such decision procedures. Assume that the statisticians have agreed on A, the universe of available (joint) actions and that each has formulated his or her own loss function. Let $i = 1, 2$ index the statisticians and let $L_i(a, \theta)$ denote the loss incurred by statistician i for taking action a when θ is true. Assume that (i) L_i is jointly measurable in a and θ, (ii) for each $\theta \in \Theta$, L_i is a lower-semicontinuous function in a satisfying $L_i(a, \theta) \geq 0$ and (iii) if A is not compact, let $A^* = A \cup \{j\}$ denote its one-point compactification with $\lim_{a \to j} L_i(a, \theta) = \infty$ and $\lim_{a \to j} \nu(a) = 0$, where ν is a dominating measure on $(A, \mathscr{A})$ for all $\delta \in \mathscr{D}$.

For any procedure $\delta \in \mathscr{D}$ and true θ, let $R_i(\delta, \theta)$ denote the risk under loss L_i, that is,

$$R_i(\delta, \theta) = \int_{\mathscr{X}} \int_{\mathscr{A}} L_i(a, \theta) \delta(da|x) P(dx|\theta).$$

For any probability measure π on $(\Theta, \mathscr{G})$, let $r_i(\delta, \pi)$ denote the expected risk of δ with respect to π for loss L_i, that is,

$$r_i(\delta, \pi) = \int_{\Theta} R_i(\delta, \theta) \pi(d\theta).$$

Let $\overline{R}_i(\delta)$ denote $\sup_{\theta \in \Theta} R_i(\delta, \theta)$, the maximum risk of δ under loss L_i.

If statistician i is a Bayesian with prior distribution π, then he prefers those $\delta \in \mathscr{D}$ with smaller Bayes risk $r_i(\delta, \pi)$. His optimal procedure, if it exists, is δ_i^B, a Bayes procedure which minimizes the Bayes risk, i.e., $r_i(\delta_i^B, \pi) \le r_i(\delta, \pi)$, for all $\delta \in \mathscr{D}$.

If, on the other hand, statistician i is a frequentist applying the minimax principle, then he prefers procedures with smaller maximum risk under his loss. The optimal procedure is then a minimax procedure δ_i^M which, if it exists, satisfies $\overline{R}_i(\delta_i^M) \le \overline{R}_i(\delta)$, for all $\delta \in \mathscr{D}$.

The general situation we consider is when the two statisticians having chosen either the Bayes or minimax principle must agree on the use of a single decision procedure. To solve this problem, they must find a procedure which is simultaneously optimal for both, agree on a principled compromise or fail to agree. We characterize a class of admissible compromise procedures when both statisticians are Bayesians in the next section.

3. The Bayes compromise problem. Suppose that each statistician is a Bayesian. Let π_i denote the probability measure on $(\Theta, \mathscr{G})$ describing the prior beliefs about θ for the ith statistician and let δ_i^B denote the corresponding Bayes procedure whose existence is guaranteed by the lower-semicontinuity of L_i; see, for example, Le Cam (1955). In general, if either the loss functions or the prior distributions differ, then the Bayes procedures δ_1^B and δ_2^B are not the same.

Any principled compromise must be a Pareto-optimal procedure, that is, one for which no other procedure is preferred by at least one statistician and is no less preferable for both, as judged by their respective Bayes risks. This is just the group admissibility principle of Savage (1954).

We pursue the characterization of the class of Pareto-optimal compromise procedures. First, we define two related problems.

Bayes compromise problem. Find a decision procedure which

$$\text{(3.1)} \qquad \begin{aligned} &\text{minimizes} && r_i(\delta, \pi) \\ &\text{subject to} && r_j(\delta, \pi) \le K_j, \end{aligned}$$

where (i, j), the roles of the statisticians, is fixed at $(1,2)$ or $(2,1)$ and K_j is a fixed constant satisfying $r_j(\delta_j^B, \pi_j) \le K_j \le r_j(\delta_i^B, \pi_j)$.

λ-Bayes compromise problem. Find a decision procedure which minimizes

$$\text{(3.2)} \qquad \lambda r_1(\delta, \pi_1) + (1 - \lambda) r_2(\delta, \pi_2),$$

where λ is a fixed constant satisfying $0 \le \lambda \le 1$.

The first result resolves the existence of solutions to problem (3.1) and (3.2), their Pareto optimality and the relationship between Bayes compromise and λ-Bayes compromise problems. To facilitate its proof, suppose that

$$r_1\big(\delta_2^B, \pi_1\big) < \infty \quad \text{and} \quad r_2\big(\delta_1^B, \pi_2\big) < \infty.$$

Consider then Theorem 3.1.

THEOREM 3.1. (*a*) *For any Bayes compromise problem* (3.1), *a solution exists.*

(b) *If a procedure is Pareto optimal, then it solves a Bayes compromise problem.*

(c) *A necessary and sufficient condition for a procedure* δ^* *to solve the Bayes compromise problem* (3.1) *specified by* K_j *is that there exists a* λ, $0 \leq \lambda \leq 1$, *such that* δ^* *solves the* λ*-Bayes compromise problem and* δ^* *satisfies* $r_j(\delta^*, \pi_j) = K_j$ *when* $j = 2$ *and* $\lambda < 1$, *or,* $j = 1$ *and* $\lambda > 0$.

(d) *If a* λ*-Bayes compromise problem* (3.2) *is specified by* $\lambda \in (0, 1)$, *then the solution is Pareto optimal.*

PROOF. Each $r_i(\delta, \pi)$ is a lower-semicontinuous function on $\mathscr{D}$ by the lower-semicontinuity of L_i and Fatou's lemma. Consequently, the "risk set" $S = \{(r_1(\delta, \pi_1), r_2(\delta, \pi_2))), \delta \in \mathscr{D}\}$ is closed from below. The remainder of the theorem then follows from standard decision-theoretic results. See, e.g., Section 5.2 of Berger (1985). □

REMARK A. Parts (c) and (d) of the theorem almost give the converse to part (b). The difficulty lies in the possibility that when $\lambda = 0$ or $\lambda = 1$, the λ-Bayes compromise problem is solved by any Bayes procedure for the second or first statistician, respectively. There is no guarantee that Bayes procedures under L_1, π_1 have constant Bayes risk under L_2, π_2 and vice versa.

REMARK B. To each group-admissible procedure δ^*, there corresponds a point (K_1, K_2) on the "lower-left" boundary of the risk set S, where the slope of the tangent line is $-\lambda/(1 - \lambda)$. The procedure δ^* simultaneously solves the λ-Bayes compromise problem and two Bayes compromise problems: that with $(i, j) = (1, 2)$ and $K_j = K_2$ and that with $(i, j) = (2, 1)$ and $K_j = K_1$.

To characterize solutions to Bayes compromise problems, we assume that the sample space $\mathscr{X}$ and parameter space Θ are finite-dimensional Euclidean spaces and that the probability measures $P(\cdot|\theta)$ for X given θ and π_i for θ are each absolutely continuous with respect to Lebesgue measure with respective densities $p(x|\theta)$, $\pi(\theta)$. For statistician i, $i = 1, 2$, we use the notation

$$p_i(x) = \int_\Theta p(x|\theta)\pi_i(d\theta),$$

$$\pi_i(\theta|x) = \frac{p(x|\theta)\pi_i(\theta)}{p_i(x)}$$

and

$$r_i(a|x) = E_i[L_i(a, \theta)|x],$$

respectively, for the marginal density of X, the density of the posterior distribution for θ given x and the posterior risk of an action $a \in \mathscr{A}$, where E_i denotes expectation with respect to the posterior distribution for θ given x.

The next result addresses the structure of solutions to λ-Bayes compromise problems.

THEOREM 3.2. *For a λ-Bayes compromise problem specified by $\lambda \in [0,1]$,*

(a) *a measurable solution δ_λ exists and*
(b) *given $X = x$, δ_λ satisfies $\delta_\lambda(H_x|x) = 1$, where*

$$H_x = \{a\colon \lambda p_1(x) r_1(a|x) + (1 - \lambda) p_2(x) r_2(a|x) \text{ is a minimum}\}.$$

PROOF. Part (a) follows from Theorem 3 in Brown and Purves (1973). Part (b) is obvious upon interchanging the order of integration over $\mathscr{X}$ and Θ (by Fubini's theorem) in the objective function $\lambda r_1(\delta, \pi_1) + (1 - \lambda) r_2(\delta, \pi_2)$. □

One might question whether the Bayes compromise problem for two Bayesians can always be reexpressed as a decision problem for a single Bayesian with a "compromise" loss function and a "compromise" prior distribution. When the statisticians have the same loss function or the same prior distribution or both, then it can. We formalize these results in the following corollaries whose proofs are straightforward.

COROLLARY 1. *If $L_1 = L_2 = L$, then the λ-Bayes compromise problem is solved by the Bayes procedure for loss L and compromise prior distribution π, equal to the λ-mixture of π_1 and π_2, i.e., $(\pi(d\theta) = \lambda\pi_1(d\theta) + (1 - \lambda)\pi_2(d\theta)$.*

COROLLARY 2. *If $\pi_1 = \pi_2 = \pi$, then the λ-Bayes compromise problem is solved by the Bayes procedure for the compromise loss: $L = \lambda L_1 + (1 - \lambda) L_2$ and prior distribution π.*

DISCUSSION. When the statisticians have the same loss function or the same prior distribution, then any Pareto-optimal compromise is to act as one Bayesian with a particular loss and prior. This resolves how a Bayesian with a well-specified prior distribution should approach a problem in which he or she is concerned simultaneously with Bayes risks under two loss functions. A single loss function which is a weighted average of the two losses achieves the objectives characterized by both risks. This is not a suprising result since the axioms of Bayesian decision theory imply that a Bayesian acts in accordance with a single loss/utility function.

Corollary 1 relates to the problem of a single decision-maker with a well-defined loss function who seeks the assistance of an expert to specify a prior distribution and hence a Bayes procedure. Controlling the Bayes risk under his and the expert's prior distributions simultaneously is equivalent to mixing the two priors with no adjustment of widely divergent priors. The λ-Bayes compromise procedure is consistent with the decision-maker believing that his or her own prior is true with probability λ and that the expert's prior is true with probability $1 - \lambda$.

Theorems 3.1 and 3.2 together characterize the Pareto-optimal procedures available to two Bayesians who must compromise and use a single procedure.

The appropriateness of one of the versions of (3.1) or of (3.2) would depend on the group structure of the pair of Bayesians. However, it is important to note that any choice of a Pareto-optimal procedure can be motivated by either version of (3.1) or by (3.2) (recall Remark B).

Choosing a specific procedure requires setting $j = 1$ or 2 and specifying K_j, the maximal Bayes risk r_j in (3.1) or specifying λ in (3.2). The study of principles guiding the choice of the "best" Pareto-optimal procedure is beyond the scope of this article. However, we note that the Nash (1950) solution investigated by Zidek and his collaborators is likely to be appropriate in many group problems because it is invariant under linear transformations of the loss (utility) functions of the Bayesians.

4. The minimax-Bayes compromise problem. Suppose now that the two statisticians who must agree on the use of a single decision procedure consist of a Bayesian and a "minimaxist," that is, a frequentist who applies the minimax principle. We still assume that each has a loss function, L_i, satisfying the conditions given in Section 2. Let the first statistician be the minimaxist and the second a Bayesian with prior distribution π_2, as before. The optimal procedures for the first statistician are minimax with respect to the first loss function. Typically no L_1-minimax procedure would be Bayes under loss L_2 with respect to the prior distribution π_2. If the two statisticians are to agree on the use of a single procedure, then a nontrival compromise is necessary.

A principled compromise for these statisticians is to use a Pareto-optimal procedure balancing the two criteria: $\overline{R}_1(\cdot)$, maximum risk under loss L_1; and $r_2(\cdot, \pi_2)$, Bayes risk under loss L_2 for prior π_2. As in the case when both statisticians are Bayesians, we seek to characterize the class of Pareto-optimal procedures for a minimaxist and a Bayesian who must compromise. To this end, define analogs to the Bayes compromise and λ-Bayes compromise problems of Section 3 as follows.

Minimax-Bayes compromise problem. Find a decision procedure which either

$$\text{(4.1a)} \qquad \begin{aligned} &\text{minimizes} \quad \overline{R}_1(\delta) \\ &\text{subject to} \quad r_2(\delta, \pi_2) \le K_2, \end{aligned}$$

where K_2 is a finite constant satisfying $r_2(\delta_2^B, \pi_2) \le K_2 \le r_2(\delta_1^M, \pi_2)$, or

$$\text{(4.1b)} \qquad \begin{aligned} &\text{minimizes} \quad r_2(\delta, \pi_2) \\ &\text{subject to} \quad \overline{R}_1(\delta) \le K_1, \end{aligned}$$

where K_1 is a finite constant satisfying $\overline{R}_1(\delta_1^M) \le K_1 \le \overline{R}_1(\delta_2^B)$.

λ-minimax-Bayes compromise problem. Find a decision procedure which minimizes

$$\text{(4.2)} \qquad \lambda \overline{R}_1(\delta) + (1 - \lambda) r_2(\delta, \pi),$$

where λ is a fixed constant satisfying $0 \le \lambda \le 1$.

Solutions to minimax-Bayes compromise problems (4.1a) or (4.1b) exist, are Pareto optimal and their specification as solutions to related λ-minimax-Bayes compromise problems is straightforward according to Theorem 4.1.

THEOREM 4.1. (*a*) *For any minimax-Bayes compromise problem* (4.1a) *or* (4.1b), *a solution exists.*

(b) *If a procedure is Pareto optimal with respect to* $\overline{R}_1(\cdot)$ *and* $r_2(\cdot, \pi_2)$, *then it solves a minimax-Bayes compromise problem.*

(c) *A necessary and sufficient condition for a procedure* δ^* *to solve the minimax-Bayes compromise problem* (4.1a) *or* (4.1b) *specified by* K_1 *or* K_2 *is that there exists a* λ: $0 \le \lambda \le 1$, *for which* δ^* *solves the* λ-*minimax-Bayes compromise problem and* δ^* *satisfies either* $r_2(\delta^*, \pi_2) = K_2$ *or* $\overline{R}_1(\delta^*) = K_1$ *depending on whether* δ^* *solves* (4.1a) *and* $\lambda < 1$ *or* δ^* *solves* (4.1b) *and* $\lambda > 0$.

(d) *If a* λ-*minimax-Bayes compromise problem* (4.2) *is specified by* $\lambda \in (0, 1)$, *then the solution is Pareto optimal.*

PROOF. The proof is entirely analogous to the proof of Theorem 3.1 upon noting that $\overline{R}_1$ is a convex, lower-semicontinuous function on $\mathscr{D}$. □

By Theorem 4.1, any Pareto-optimal procedure can be characterized as a solution to a λ-minimax-Bayes compromise problem. We now show that any λ-minimax-Bayes compromise problem can be expressed as a minimax problem for a generalized risk function.

THEOREM 4.2. *For a given* λ: $0 \le \lambda \le 1$, *a procedure solves the* λ-*minimax-Bayes compromise problem if and only if it is a minimax procedure with respect to the generalized risk function*

$$R_\lambda(\delta, \theta) = \lambda R_1(\delta, \theta) + (1 - \lambda) r_2(\delta, \pi_2). \tag{4.3}$$

PROOF. The result is immediate upon noting that

$$\sup_\theta R_\lambda(\delta, \theta) = \lambda \overline{R}_1(\delta) + (1 - \lambda) r_2(\delta, \pi_2). \qquad \square$$

We call R_λ a generalized risk function because there is no underlying loss function depending on just the true parameter, $\theta \in \Theta$, and the action taken, $a \in A$.

To solve a λ-minimax-Bayes compromise problem then, we can apply standard minimax techniques with a generalized risk. Specifically, under additional assumptions on the decision problem, a minimax procedure can be characterized as a Bayes procedure with respect to a least favorable prior distribution or as a limit of such procedures. Wald (1950) and Brown (1976) provide extensive discussions of conditions on the decision problem for which this is true. Introductory expositions of this theory are given in the texts of Bickel and Doksum (1977), Ferguson (1967) and Berger (1985). Kempthorne (1987a) gives simple proofs for special cases of the general problem treated by Wald and Brown.

We quote a theorem proven in Kempthorne (1987a) regarding minimax problems for generalized risks and then interpret it for compromise problems. Let

$R(\cdot,\cdot)$ be a risk function on $\mathscr{D}\times\Theta$, which is lower-semicontinuous on $\mathscr{D}$ for each $\theta\in\Theta$. For any prior distribution π on Θ and procedure $\delta\in\mathscr{D}$, let $r(\delta,\pi)$ and $\bar{R}(\delta)$ denote the Bayes and maximum risks.

Let $\underline{r}(\pi)=\inf_{\delta\in\mathscr{D}} r(\delta,\pi)=r(\delta_\pi,\pi)$ denote the Bayes risk of the Bayes procedure δ_π. If it exists, a least favorable prior distribution π^* satisfies $\underline{r}(\pi^*)=\sup_{\pi\in\mathscr{P}(\Theta)}\underline{r}(\pi)$, where $\mathscr{P}(\Theta)$ denotes the class of all distributions on Θ. The existence of a least favorable prior distribution and the minimaxity of the corresponding Bayes procedure is proved under the following assumptions.

(A.1′) The risk function $R(\delta,\theta)$ is a continuous function of θ for any $\delta\in\mathscr{D}$ having everywhere finite risk.

(A.2′) If $\{\pi_i,\ i=1,2,\dots\}$ is any sequence of distributions on Θ which converge weakly to π, then for θ in any compact set, the risks of the corresponding Bayes procedures $\{R(\delta_{\pi_i},\theta),\ i=1,2,\dots\}$ converge uniformly to the risk $R(\delta_\pi,\theta)$ of the Bayes procedure with respect to π.

THEOREM 4.3. *Under assumptions* (A.1′) *and* (A.2′), *if* Θ *is a compact subset of* $\mathfrak{R}^k$, *then*

(a) *a least favorable prior distribution* π^* *exists,*

(b) *the Bayes procedure* δ_{π^*} *corresponding to the distribution* π^* *in* (a) *is minimax and*

(c) *the Bayes procedure* δ_{π^*} *is an equalizer on* π^*, *i.e.,*

$$R(\delta_{\pi^*},\theta')=\underline{r}(\pi^*),\quad \textit{for all } \theta'\in\operatorname{supp}(\pi^*).$$

PROOF. A direct proof is given in Theorems 2.1–2.3 of Kempthorne (1987a). While assumption (A.2′) is stronger than needed, it enables a proof without relying on properties of regular-convergent sequences of decision procedures. Proofs in more general contexts are given in Wald (1950) and Brown (1976). For example, it is sufficient to assume that if $\{\delta_i,\ i=1,2,\dots\}$ converges regularly to δ, then $\{R(\delta_i,\theta),\ i=1,2,\dots\}$ converges uniformly to $R(\delta,\theta)$. □

The next theorem provides a solution to the minimax-Bayes compromise problem. First, we must interpret the assumptions when $R(\cdot,\theta)$ is the generalized risk $R_\lambda(\cdot,\theta)=\lambda R_1(\cdot,\theta)+(1-\lambda)r_2(\cdot,\pi_2)$. We can restate (A.1′) and (A.2′) for a compromise problem as follows.

(A.1) The risk function for the minimaxist, $R_1(\delta,\theta)$ is a continuous function of θ for any procedure $\delta\in\mathscr{D}$ having everywhere finite risk.

(A.2) Let $\{\pi_{1,i},\ i=1,2,\dots\}$ be any sequence of distributions on Θ which converge weakly to the distribution π_1. For λ: $0\le\lambda\le1$, let $\delta_{\lambda,i}$ (respectively, δ_λ) denote the procedure which solves the λ-Bayes compromise problem when statistician 1 uses the prior distribution $\pi_{1,i}$ (respectively, π_1) and statistician 2 uses prior distribution π_2. Then

(a) $\lim_{i\to\infty} r_2(\delta_{\lambda,i},\pi_2)=r_2(\delta_\lambda,\pi)$ and

(b) $\lim_{i\to\infty} R_1(\delta_{\lambda,i},\theta)=R_1(\delta_\lambda,\theta)$ uniformly for θ in any compact set.

Now we can state Theorem 4.4.

THEOREM 4.4. *If Θ is compact and assumptions* (A.1) *and* (A.2) *are satisfied, then there exists a distribution π_1^* on Θ such that*

(a) δ_λ^*, *the solution to the λ-Bayes compromise problem, where π_1^* and π_2 are the prior distributions of the two statisticians, solves the λ-minimax-Bayes compromise problem and*

(b) *for all θ' in the support of π_1^*, the procedure δ_λ^* in* (a) *satisfies*

$$R_1(\delta_\lambda^*, \theta') = \sup_{\theta \in \Theta} R_1(\delta_\lambda^*, \theta).$$

PROOF. Part (a) follows from Theorem 4.3(a) and (b) using the generalized risk R_λ. Part (b) follows from Theorem 4.3(c), noting that R_λ and R_1 are maximized at the same θ values. □

Note that it is possible for π_1^* to equal π_2. If L_1 equals L_2, then the λ-minimax-Bayes compromise procedure would be δ_2^B, the second statistician's Bayes procedure by Corollary 1 to Theorem 3.2.

The next theorem extends Theorem 4.4 to the case of noncompact Θ. Consider Theorem 4.5.

THEOREM 4.5. *For a given λ, suppose the λ-minimax-Bayes compromise problem* (4.2) *satisfies* (A.1) *and* (A.2).

(a) *There exists a sequence of λ-Bayes compromise problems defined by a sequence of prior distribution pairs $\{(\pi_{1,n}^*, \pi_2),\ n = 1, 2, \dots\}$ (i.e., π_2 is fixed), with solutions $\{\delta_n^*,\ n = 1, 2, \dots\}$ and a procedure δ_λ^*, solving the λ-minimax-Bayes compromise problem which satisfy $\liminf_{n \to \infty} R_\lambda(\delta_n^*, \theta) \geq R_\lambda(\delta_\lambda^*, \theta)$. Furthermore, the sequence $\{\pi_{1,n}^*,\ n = 1, 2, \dots\}$ has weak limit π_1^* and is least favorable in the sense that $\{r_1(\delta_n^*, \pi_{1,n}^*),\ n = 1, 2, \dots\}$ is an increasing sequence whose limit is $\sup_{\theta \in \Theta} R_1(\delta_\lambda^*, \theta)$.*

(b) *If the weak limit π_1^* of the sequence $\{\pi_{1,n}^*,\ n = 1, 2, \dots\}$ in* (a) *is proper, then* (i) *δ_λ^* is the λ-Bayes compromise procedure corresponding to the prior distribution pair (π_1^*, π_2) and* (ii) *δ_λ satisfies $R_1(\delta_\lambda^*, \theta') = \sup_{\theta \in \Theta} R_1(\delta_\lambda^*, \theta)$ for all θ' in the support of π_1^*.*

PROOF. Let $\{\Theta_n,\ n = 1, 2, \dots\}$ be a sequence of nested compact sets which converge to Θ. For each n, consider the minimax problem for parameter space Θ_n and generalized risk function $R_\lambda(\delta, \theta) = \lambda R_1(\delta, \theta) + (1 - \lambda) r_2(\delta, \pi_2)$. For each n, let δ_n^* be the solution to the minimax problem and let $\pi_{1,n}^*$ be the least favorable distribution for which δ_n^* is Bayes. The procedure δ_n^* minimizes $\sup_{\theta \in \Theta_n} R_\lambda(\delta, \theta)$, for $\delta \in \mathscr{D}$; and it solves the λ-Bayes compromise problem when the two statisticians' prior distributions are $\pi_{1,n}^*$ and π_2.

Consider the sequence of maximum risks $\{\overline{R}_n^* = \max_{\theta \in \theta_n} R_\lambda(\delta_n^*, \theta),\ n = 1, 2, \dots\}$. Clearly $\{\overline{R}_n^*\}$ is monotonically increasing. If the sequence increases without bound, then $\inf_{\delta \in \mathscr{D}} \max_{\theta \in \Theta} R_\lambda(\delta, \theta) = \infty$ and any procedure is minimax.

So, we need only consider the case when the limit is bounded above by $C < \infty$. For any $\theta_0 \in \Theta$, if n is sufficiently large, say $n > n_0$, then $\theta_0 \in \Theta_n$ and $\delta_n^* \in \mathscr{D}_C = \{\delta\colon\ R_\lambda(\delta, \theta_0) \le C\}$. By the sequential subcompactness of $\Gamma = \{R_\lambda(\delta, \cdot)\colon \delta \in \mathscr{D}_C\}$, there exists a subsequence $\{\delta_{n_j}^*,\ j = 1,2,\dots\} \subset \mathscr{D}_C$ of the δ_n^* and a $\delta_\lambda^* \in \mathscr{D}_C$ such that $\liminf_{j\to\infty} R_\lambda(\delta_{n_j}^*, \theta) \ge R_\lambda(\delta_\lambda^*, \theta)$, for all $\theta \in \Theta$; see Theorem 8.6 on page 64 of Brown (1976).

Without loss of generality, substitute the original sequence with the subsequence $\{n_j\}$ (or a further subsequence), where π_1^* denotes the weak limit of the subsequence of $\{\pi_{1,n}^*,\ n = 1,2,\dots\}$ (the further subsequence may be necessary for the weak limit to exist).

For any n, it is easily shown that

$$\sup_{\theta\in\Theta_n} R_\lambda(\delta_\lambda^*, \theta) \le \liminf_j \left[\sup_{\theta\in\Theta_j} R_\lambda\big(\delta_j^*, \theta\big)\right]. \tag{4.4}$$

The bracketed quantity on the right-hand side of (4.4) is $\inf_\delta \sup_{\theta\in\Theta_j} R_\lambda(\delta, \theta)$ because δ_j^* is minimax for R_λ on Θ_j. Together with the fact that $\Theta_j \nearrow \Theta$ it follows that $\sup_{\theta\in\Theta} R_\lambda(\delta_\lambda^*, \theta) = \inf_\delta \sup_{\theta\in\Theta} R_\lambda(\delta, \theta)$. Thus δ_λ^* is minimax for risk R_λ and the proof of (a) is complete.

For (b), suppose first that (i) is not true. Then there exists a procedure δ_λ^{**} for which $r_\lambda(\delta_\lambda^{**}) < r_\lambda(\delta_\lambda^*)$, where $r_\lambda(\cdot)$ is the objective function for the λ-Bayes compromise problem specified by the prior distributions π_1 and π_2. Let $\delta_{\lambda,n}^{**}$ be the procedure which randomizes between δ_n^* and δ_λ^{**} with probability $\frac{1}{2}$. Then it follows that

$$\lim_{n\to\infty} \left[\lambda r_1\big(\delta_{\lambda,n}^{**}, \pi_{1,n}^*\big) + (1-\lambda) r_2\big(\delta_{\lambda,n}^{**}, \pi_2\big)\right] < r_\lambda(\delta_\lambda^*).$$

This yields a contradiction because

$$\begin{aligned} r_\lambda(\delta_\lambda^*) &= \lim_{n\to\infty} \inf_\delta \left[\lambda r_1\big(\delta, \pi_{1,n}^*\big) + (1-\lambda) r_2\big(\delta, \pi_2\big)\right] \\ &\le \lim_{n\to\infty} \left[\lambda r_1\big(\delta_{\lambda,n}^{**}, \pi_{1,n}^*\big) + (1-\lambda) r_2\big(\delta_{\lambda,n}^{**}, \pi_2\big)\right]. \end{aligned}$$

Thus δ_λ^* solves the given λ-Bayes compromise problem.

It remains to prove (ii) of part (b). Let Q denote any compact set of the parameter space where the L_1-risk of δ_λ^* does not achieve its maximum. There must exist an $\varepsilon > 0$ such that $R_1(\delta_\lambda^*, \theta') < \sup_\theta R_1(\delta_\lambda^*, \theta) - \varepsilon$, for all $\theta' \in Q$. But for sufficiently large n, $\pi_{1,n}^*(Q) = 0$ because $\pi_{1,n}^*$ supports only the points of maximum L_1-risk of δ_n^* on Θ_n which are within ε of the maximum risk of δ_λ^* on Θ. Since π_1^* is the weak limit of $\{\pi_{1,n}^*,\ n = 1,2,\dots\}$ it must be that $\pi_1^*(Q) = 0$. Because Q is arbitrary the proof is complete. □

A difficulty in applying Theorem 4.5 is that when Θ is not compact, there is no guarantee that the weak limit of $\{\pi_{1,n}^*,\ n = 1,2,\dots\}$ is indeed a proper probability distribution. Brown (1980) gives some conditions under which π_1^* is proper in the standard decision theory problem. The study of conditions ensuring that π_1^* is proper and the study of solutions when the weak limit is a subprobability measure are beyond the scope of this article. However, Kempthorne (1987b) pursues the characterization of such solutions for minimax-Bayes compromise estimation of a multivariate normal mean under squared-error loss.

The application of Theorems 4.3 and 4.5 entails determining the least favorable prior distribution for a generalized risk function. To conclude, we show that when Θ is a closed interval of $\mathfrak{R}^1$ the distribution is discrete for a large class of problems satisfying the following assumption.

(A.3) For any procedure $\delta \in \mathscr{D}$ whose L_1-risk is finite for all θ, $R_1(\delta, \theta)$ is an analytic function of θ.

Now consider

THEOREM 4.6. *For the* λ*-minimax-Bayes compromise problem of Theorem* 4.4 *or* 4.5, *suppose that* Θ *is a closed interval of* $\mathfrak{R}^1$ *and* (A.3) *is satisfied. Then either the* L_1*-risk of* δ_λ^*, *the* λ*-minimax-Bayes compromise procedure, is constant on* Θ, *or the support of the least favorable distribution* $\pi_1{}^*$ *is*

(a) *discrete and finite, if* Θ *is compact or*

(b) *discrete with no accumulation point, if* Θ *is not compact and* $\pi_1{}^*$ *is proper* [*see Theorem* 4.5(b)].

PROOF. By Theorem 4.4 or 4.5 the procedure δ_λ^* has constant L_1-risk on the support of $\pi_1{}^*$. Since $R_1(\delta_\lambda^*, \theta)$ is analytic in θ, if the support of $\pi_1{}^*$ contains an accumulation point, then the function must be constant on Θ by the identity theorem for analytic functions. Otherwise, the support of $\pi_1{}^*$ is discrete with no accumulation point. If Θ is compact, then it follows that the support must be finite. □

So, when we can show that no procedure with constant L_1-risk solves the λ-minimax-Bayes compromise problem, it must be that the distribution $\pi_1{}^*$ specifying the compromise procedure is discrete. Kempthorne (1987a) presents an algorithm for specifying such discrete distributions numerically and proves its convergence when Θ is a compact interval.

For solving minimax-Bayes compromise problems (4.1a) or (4.1b), if the λ specifying the corresponding λ-minimax-Bayes compromise problem is known, then just one minimax problem with a generalized risk function must be solved. When λ is unknown a sequence of such problems can be solved varying λ in the unit interval until the constraint of (4.1a) or (4.1b) is just satisfied. Implementing this strategy analytically may be intractable, but numerical solutions are sometimes possible; see e.g., Kempthorne (1987a, 1988).

Acknowledgments. I am grateful to the late Jack Kiefer and to Peter Bickel for stimulating discussions in the early stages of this work. I thank Arthur Dempster, John Emerson, David Hoaglin, Augustine Kong and Donald Rubin for their comments on an earlier draft. I am particularly indebted to the referees whose criticisms contributed significantly to revisions resulting in the present manuscript.

REFERENCES

BERGER, J. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.* **4** 223–226.

BERGER, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.* **8** 716–761.

BERGER, J. (1982a). Bayesian robustness and the Stein effect. *J. Amer. Statist. Assoc.* **77** 358–368.

BERGER, J. (1982b). Estimation in continuous exponential families: Bayesian estimation subject to risk restrictions and inadmissibility results. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) **1** 109–141. Academic, New York.

BERGER, J. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

BICKEL, P. J. (1983). Minimax estimation of the mean of a normal distribution subject to doing well at a point. In *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday* (M. H. Rizvi, J. S. Rustagi and D. Siegmund, eds.) 511–528. Academic, New York.

BICKEL, P. J. and DOKSUM, K. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics.* Holden-Day, San Francisco.

BROWN, L. D. (1975). Estimation with incompletely specified loss functions (the case of several location parameters). *J. Amer. Statist. Assoc.* **70** 417–427.

BROWN, L. D. (1976). Notes on statistical decision theory. Unpublished lecture notes, Cornell Univ.

BROWN, L. D. (1980). A necessary condition for admissibility. *Ann. Statist.* **8** 540–544.

BROWN, L. D. and PURVES, R. (1973). Measurable selections of extrema. *Ann. Statist.* **1** 901–912.

CHEN, S.-Y. (1983). Restricted Bayes estimation for the mean of the multivariate normal distribution. Technical Report 83-33, Dept. Statistics, Purdue Univ.

COHEN, A. and SACKROWITZ, H. D. (1984). Decision theory results for vector risks with applications. *Statist. Decisions Suppl.* **1** 159–176.

DE WAAL, D. J., GROENEWALD, P. C. N., VAN ZYL, J. M. and ZIDEK, J. V. (1984). Multi-Bayesian estimation theory. Technical Report 4, Dept. Statistics, Univ. British Columbia.

EFRON, B. and MORRIS, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators. I. The Bayes case. *J. Amer. Statist. Assoc.* **66** 807–815.

EFRON, B. and MORRIS, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators. II. The empirical Bayes case. *J. Amer. Statist. Assoc.* **67** 130–139.

FERGUSON, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* Academic, New York.

GEORGE, E. (1986). A formal Bayes multiple shrinkage estimator. *Comm. Statist. A—Theory Methods* **15** 2099–2114.

HODGES, J. L., JR. and LEHMANN, E. L. (1952). The use of previous experience in reaching statistical decisions. *Ann. Math. Statist.* **23** 396–407.

KEMPTHORNE, P. J. (1983). Minimax-Bayes compromise estimators. *Proc. Bus. and Econ. Statist. Sec. Amer. Statist. Assoc.* 563–573.

KEMPTHORNE, P. J. (1985). Controlling risks under different loss functions. Technical Report NS-540, Dept. Statistics, Harvard Univ.

KEMPTHORNE, P. J. (1986). Optimal minimax square error risk estimation of the mean of a multivariate normal distribution. *Comm. Statist. A—Theory Methods* **15** 2145–2158.

KEMPTHORNE, P. J. (1987a). Numerical specification of discrete least favorable prior distributions. *SIAM J. Sci. Statist. Comput.* **8** 171–184.

KEMPTHORNE, P. J. (1987b). Minimax-Bayes compromise estimation of a multivariate normal mean under squared-error loss. Technical Report, Sloan School of Management, M.I.T.

KEMPTHORNE, P. J. (1988). Dominating inadmissible procedures using compromise decision theory. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) **1** 381–396. Academic, New York.

LE CAM, L. (1955). An extension of Wald's theory of statistical decision functions. *Ann. Math. Statist.* **26** 69–81.

MADANSKY, A. (1978). Externally Bayesian groups. [Unpublished manuscript, Univ. Chicago, as referenced in Zidek (1984).]

MEEDEN, G. and VARDEMAN, S. (1985). Bayes and admissible set estimation. *J. Amer. Statist. Assoc.* **80** 465–471.

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

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

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

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

WEERAHANDI, S. and ZIDEK, J. (1983). Elements of multi-Bayesian decision theory. *Ann. Statist.* **11** 1032–1046.

ZIDEK, J. V. (1984). Multi-Bayesianity. Technical Report 5, Dept. Statistics, Univ. British Columbia.

SLOAN SCHOOL OF MANAGEMENT
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139