STRINGENT SOLUTIONS TO STATISTICAL DECISION PROBLEMS[1]

By Robert Cogburn

University of California, Berkeley

1. Introduction. The statistical decision problem generally involves performing some kind of experiment which yields partial and uncertain information about a parameter or an *a priori* distribution, and decision theory studies the risks resulting from basing action on the partial and uncertain nature of this information. When, implicit in the decision problem, more reliable information is available about some relevant aspects of the parameter or *a priori* distribution than about others, a natural breaking up of the parameter space or space of *a priori* distributions into equivalence classes, or "slices" (see Wesler [13]), is suggested. The equivalence classes are chosen so there is relatively good information about which equivalence class the parameter or *a priori* distribution lies in. In this situation the stringent decision functions appear as a class of conservative solutions to the problem, in the way that the minimax solution appears in the unsliced problem. The notion of a stringent solution has arisen in at least two particular contexts previously: the first in the theory of testing hypotheses, see Lehmann [7] (the proper source is Hunt and Stein [3]), the second in prediction theory, see Lehmann [6].

Section 2 introduces the concept of a stringent solution for the statistical decision problem and develops some elementary aspects of the theory of stringency.

Section 3 applies the concept of stringency to the multivariate (non-Bayesian) problem obtained by repeating a basic decision problem, and to the empirical Bayes problem (see Robbins [10]). These two problems are shown to be closely related from the viewpoint of stringency. The possibility of using stringent or nearly stringent solutions in treating the empirical Bayes problem is discussed, and then it is shown how these solutions may be used in some cases to obtain interesting solutions to the multivariate problem.

Section 4 provides two examples of applications of the ideas developed in Section 3 to estimation problems with squared error loss function. These problems were chosen to reveal a maximum amount of the relevant structure with a minimum of extraneous complication; stringency is in no way restricted to such problems.

So far as I know, the results in Section 3 venture in a hitherto unexplored area. They leave much to be desired, but I hope they may interest someone else in further exploration.

The theory set forth in this paper should be compared with alternative ap-

Received 4 September 1964; revised 26 October 1966.

[1] Prepared with the partial support of the U. S. Army Research Office, DA-ARO(D) 31-124-G-83.

proaches for the situation where partial information about a parameter or *a priori* distribution is available. See, in particular, Hodges and Lehmann [2], Wesler [13], Robbins [10] and the papers cited by Robbins for the empirical Bayes and related approaches.

2. Stringent decision functions. We will consider statistical decision problems defined in terms of the following elements:

1. A *parameter space* Θ. The elements of Θ correspond to "states of nature."
2. An *experiment space* $\mathfrak{X}$, together with a σ-algebra of subsets $\mathcal{S}$. Each $\theta \varepsilon \Theta$ determines a probability distribution P_θ on $\mathcal{S}$.
3. A *decision space* D, together with a σ-algebra of subsets $\mathfrak{D}$. *We assume all singletons are in* $\mathfrak{D}$.
4. A *loss function* L defined on $\Theta \times D$ to the non-negative real line such that $L(\theta, \cdot)$ is $\mathfrak{D}$ measurable on D for each $\theta \varepsilon \Theta$.

These elements specify the problem. In addition, we have

5. The space Δ of all functions δ on $\mathfrak{X} \times \mathfrak{D}$ to the reals such that $\delta(x, \cdot)$ is a probability distribution on $\mathfrak{D}$ for each $x \varepsilon \mathfrak{X}$ and $\delta(\cdot, C)$ is $\mathcal{S}$ measurable in x for each $C \varepsilon \mathfrak{D}$. Elements of Δ are called *decision functions.*

Of course, Δ is the space of "randomized decision functions" and "pure decision functions" appear as a special case, when the distributions $\delta(x,\cdot)$ degenerate.

6. The *risk function* R defined on $\Theta \times \Delta$ by

$$R(\theta, \delta) = \int_{\mathfrak{X}} P_\theta(dx) \int_D L(\theta, c)\delta(x, dc) = \int_{\mathfrak{X}} L(\theta, \delta, x) P_\theta(dx), \tag{2.1}$$

where

$$L(\theta, \delta, x) = \int_D L(\theta, c)\delta(x, dc). \tag{2.2}$$

If θ corresponds to the state of nature and the statistician uses decision function δ, then he incurs the risk (or average loss) $R(\theta, \delta)$.

Now let $\mathfrak{J}$ be the smallest σ-algebra in Θ such that the functions $R(\theta, \delta)$ and $P_\theta S$ are measurable in θ for all $\delta \varepsilon \Delta$ and $S \varepsilon \mathcal{S}$, respectively. Then we have

7. The space Π of all probability distributions on $\mathfrak{J}$. The elements of Π are called *a priori distributions.*
8. *The Bayes risk function* R^* defined on $\Pi \times \Delta$ by

$$R^*(\pi, \delta) = \int R(\theta, \delta)\pi(d\theta). \tag{2.3}$$

9. The *Bayes envelope function* r^* defined on Π by

$$r^*(\pi) = \inf_{\delta \epsilon \Delta} R^*(\pi, \delta). \tag{2.4}$$

A decision function δ_π is a *Bayes solution* relative to π if $R^*(\pi, \delta_\pi) = r^*(\pi)$.

10. The *restricted minimax risk* function m defined on the class of all subsets T of Θ by

$$m(T) = \inf_{\delta \epsilon \Delta} \sup_{\theta \epsilon T} R(\theta, \delta). \tag{2.5}$$

A decision function δ_T is a *restricted minimax solution relative to* T if $R(\theta, \delta) \leqq m(T)$ for all $\theta \varepsilon T$. When $T = \Theta$, we drop the word "restricted," and *we assume that the minimax risk*, $m(\Theta)$, *is finite.*

If the statistician knew beforehand that $\theta \varepsilon T$, it might be reasonable to use a restricted minimax solution relative to T for the decision problem, if one exists. At least, he can choose δ so that $R(\theta, \delta) \leqq m(T) + \epsilon$ for any $\epsilon > 0$ and all $\theta \varepsilon T$. In general, one can describe the situation where the statistician has partial, but accurate, beforehand knowledge about the parameter θ by introducing an equivalence relation $\sim$ on Θ, the statistician knowing θ up to the equivalence $\sim$. Let $\sim(\theta)$ denote the equivalence class generated by θ,

$$\sim(\theta) = \{\theta' \varepsilon \Theta : \theta' \sim \theta\}.$$

We then introduce

11. The *envelope risk function* $r_\sim$, relative to $\sim$, defined on Θ by

$$r_\sim(\theta) = m(\sim(\theta)). \tag{2.6}$$

Now this envelope function represents the best the statistician could do if he *knew* θ beforehand up to the equivalence $\sim$. When he lacks this beforehand knowledge, the random variable X may still provide some, but less than certain, information about the aspect of θ corresponding to $\sim$. A measure of the effect of this uncertainty on the risk is

12. The *excess* relative to $\sim$.

$$e_\sim = \inf_{\delta \epsilon \Delta} \sup_{\theta \epsilon \Theta} (R(\theta, \delta) - r_\sim(\theta)). \tag{2.7}$$

A decision function $\delta_\sim$ is *stringent* relative to $\sim$ if

$$R(\theta, \delta_\sim) \leqq r_\sim(\theta) + e_\sim \quad \text{for all} \quad \theta \varepsilon \Theta.$$

Note that in 12 we might have used any function r' on Θ, rather than one of the envelope functions described in 11. Then we would define the excess and stringent solutions "relative to r'." Of course, the excess is simply the minimax risk when the risk function $R(\theta, \delta)$ is replaced by $R(\theta, \delta) - r_\sim(\theta)$, and a stringent solution is a minimax solution relative to this new risk function. Thus the existing decision theory can be transferred in a trivial way to yield information about stringent solutions, and we will not have much more to say about this. However, the first two theorems are obtained by this transposition and are noted for later use.

THEOREM 2.1. *If* π *is an* a priori *distribution and* δ_π *is a Bayes solution relative to* π, *and if there exists a set* $T \varepsilon \mathfrak{I}$ *such that* $\pi(T) = 1$ *and for all* $\theta \varepsilon T$

$$R(\theta, \delta) - r_\sim(\theta) = \sup_{\theta' \varepsilon \Theta} (R(\theta', \delta) - r_\sim(\theta')), \tag{2.8}$$

then δ_π *is stringent relative to* $\sim$, *and, if* $r_\sim$ *is* $\mathfrak{I}$ *measurable on* T, *then,*

$$e_\sim = r^*(\pi) - \int_T r_\sim(\theta) \pi(d\theta). \tag{2.9}$$

Such a distribution π is "least favorable" relative to $R(\theta, \delta) - r_\sim(\theta)$ according to the usual terminology. We might call it "least informative" relative to $\sim$.

When G is a group leaving the decision problem invariant (see, for example, Wesler [13]), a natural equivalence relation on Θ is obtained by setting $\theta \sim \theta'$ if θ' is the image of θ under some member of the group acting on Θ (i.e., if θ, θ' are on the same orbit). Clearly, the problem then remains invariant under G if $R(\theta, \delta)$ is replaced by $R(\theta, \delta) - r_\sim(\theta)$. Thus the Hunt-Stein theorem and its generalizations (see, for example, Kiefer [5], Lehmann [7] and Wesler [13]) can be applied under suitable regularity conditions to conclude that, if Δ_G denotes the class of decision functions invariant under G, then

$$\inf_{\delta\epsilon\Delta_G} \sup_{\theta\epsilon\Theta} (R(\theta, \delta) - r_\sim(\theta)) = e_\sim , \tag{2.10}$$

so attention can be restricted to invariant decision functions in looking for a stringent solution. In particular, this result will hold if G is a compact group in its natural group topology (see Loomis [8]) and necessary measurability conditions are satisfied, and as a trivial case we have

THEOREM 2.2. *If G is a finite group leaving the decision problem invariant, then* (2.10) *is satisfied.*

The more interesting aspects of the theory of stringency revolve around the relation of various equivalences or envelope functions to the information at hand. Consider the information about the parameter θ available to the statistician. Given any equivalence $\sim$ on Θ, we can split this information into two parts: the information about which equivalence class θ is in (location "between" equivalence classes) and the information about the position of θ in a given equivalence class (location "within" equivalence classes). An equivalence relation will be of particular interest from the viewpoint of stringency when there is relatively accurate information about the location between the equivalence classes. The excess relative to such an equivalence should be small. We would also like the information within the equivalence classes to be relatively unimportant to the decision problem, in which case the envelope function should also be small. Of course, we cannot expect to obtain a small excess and a uniformly small envelope function for the same equivalence relation. What we can try to do is to make the envelope function small at least for some values of θ (it never exceeds the unrestricted minimax risk in any case) and keep the excess small.

A natural partial ordering is defined on the set of all equivalence relations on Θ by saying that $\approx$ *follows* $\sim$, denoted $\approx \supset \sim$, if

$$\theta \approx \theta' \Rightarrow \theta \sim \theta'.$$

In this partial ordering there is the first member, which identifies all members of Θ, and the last member, equality on Θ. As an immediate consequence of the definitions, we have

THEOREM 2.3. *If $\approx \subset \sim$, then*

$$0 \leqq r_\approx \leqq r_\sim \leqq m(\Theta) \tag{2.11}$$

and

$$0 \leqq e_\sim \leqq e_\approx \leqq m(\Theta). \tag{2.12}$$

Moreover, for any two envelope functions r_1 *and* r_2 *with corresponding excesses* e_1 *and* e_2,

$$|e_1 - e_2| \leqq \sup_{\theta\varepsilon\Theta} |r_1(\theta) - r_2(\theta)|. \tag{2.13}$$

The risk functions of all stringent procedures are uniformly bounded:

THEOREM 2.4. *If* $\delta_\sim$ *is stringent relative to* $\sim$, *then*

$$R(\theta, \delta_\sim) \leqq m(\Theta) + r_\sim(\theta) - \inf_{\theta\varepsilon\Theta} r_\sim(\theta) \leqq 2m(\Theta). \tag{2.14}$$

For, from (2.7),

$$e_\sim \leqq \inf_{\delta\varepsilon\Delta} \{\sup_{\theta\varepsilon\Theta} R(\theta, \delta) - \inf_{\theta\varepsilon\Theta} r_\sim(\theta)\} = m(\Theta) - \inf_{\theta\varepsilon\Theta} r_\sim(\theta).$$

Let $Q = \Theta/\sim$ be the quotient space of Θ relative to $\sim$, so $\sim(\theta)$ is the canonical function mapping θ into its equivalence class in Q, and suppose that for each $q \varepsilon Q$ a restricted minimax solution δ_q relative to $\{\theta \varepsilon \Theta: \sim(\theta) \varepsilon q\}$ exists. Then a natural way to find an at least approximately stringent solution would be to try to find a good "estimate" of q (that is, a function $\hat{q}$ on $\mathfrak{X}$ to Q) and then use the decision function $\delta_{\hat{q}}$. To apply our apparatus, we must assume that $\hat{q}$ is chosen so $\delta_{\hat{q}(x)}(x, C)$ is $\mathcal{S}$ measurable in x for each $C \varepsilon \mathfrak{D}$. Then we have

$$R(\theta, \delta_{\hat{q}}) \leqq r_\sim(\theta) + \int |L(\theta, \delta_{\hat{q}(x)}, x) - L(\theta, \delta_{\sim(\theta)}, x)| P_\theta(dx). \tag{2.15}$$

For example, if we can take Q to be a finite dimensional real linear space and

$$|L(\theta, \delta_{\hat{q}}, x) - L(\theta, \delta_{\sim(\theta)}, x)| \leqq A |\hat{q} - \sim(\theta)| + B |\hat{q} - \sim(\theta)|^2, \tag{2.16}$$

then

$$R(\theta, \delta_{\hat{q}}) \leqq r_\sim(\theta) + AM_\theta^{\frac{1}{2}} + BM_\theta \tag{2.17}$$

where M_θ is the mean squared error of $\hat{q}$

$$M_\theta = E_\theta |\hat{q} - \sim(\theta)|^2. \tag{2.18}$$

The notion of stringency can be applied to Bayesian problems (that is, problems where "nature employs an *a priori* distribution") by taking Π to be the parameter space in the preceding discussion. Then the restricted minimax function is defined in terms of R^* on subsets of Π, and the equivalences and corresponding envelope functions are defined on Π. In particular, the Bayes envelope is the envelope function relative to equality on Π. This device will be used in treating the empirical Bayes problem in Section 3.

3. Multivariate and empirical Bayes problems. In this section we consider decision problems obtained by repeating a basic decision problem. More precisely, if Θ, $(\mathfrak{X}, \mathcal{S})$, $(D, \mathfrak{D})$ and L are as described in Section 2, then we may consider a new problem with parameter space $\Theta^{(n)}$ and experiment space $(\mathfrak{X}^{(n)}, \mathcal{S}^{(n)})$ (the superscript (n) referring to the n-fold cartesian product of the basic space). For each $\theta^{(n)} \varepsilon \Theta^{(n)}$, we define $P_{\theta^{(n)}}$ by making $X_1, \cdots, X_n$ independent,

X_k having the distribution corresponding to the kth coordinate of $\theta^{(n)}$. That is, for $\theta^{(n)} = (\theta_1, \cdots, \theta_n)$,

$$P_{\theta^{(n)}}[X_1 \varepsilon S_1, \cdots, X_n \varepsilon S_n] = P_{\theta_1}[X_1 \varepsilon S_1] \cdots P_{\theta_n}[X_n \varepsilon S_n].$$

We take the decision space for this problem to be $(D^{(n)}, \mathfrak{D}^{(n)})$ and the loss function to be

$$L_n(\theta_1, \cdots, \theta_n; d_1, \cdots, d_n) = n^{-1} \textstyle\sum_{k=1}^n L(\theta_k, d_k). \tag{3.1}$$

A decision function is then a function, $\delta^{(n)} = (\delta_1, \cdots, \delta_n)$, where each δ_k is a function on $\mathfrak{X}^{(n)} \times \mathfrak{D}$ to the reals such that $\delta_k(x^{(n)}, \cdot)$ is a probability distribution on $\mathfrak{D}$ for each $x^{(n)} \varepsilon \mathfrak{X}^{(n)}$ and $\delta(\cdot, C)$ is $\mathcal{S}^{(n)}$ measurable in $x^{(n)}$ for each $C \varepsilon \mathfrak{D}$, and an *a priori* distribution is a probability distribution defined on a sufficiently large σ-algebra in $\Theta^{(n)}$. The risk function, obtained by integrating L_n, is then

$$R_n(\theta^{(n)}, \delta^{(n)}) = n^{-1} \textstyle\sum_{k=1}^n \int_{\mathfrak{X}} L_n(\theta_k, \delta_k, x^{(n)}) P_{\theta^{(n)}}(dx^{(n)}), \tag{3.2}$$

where

$$L_n(\theta_k, \delta_k, x^{(n)}) = \textstyle\int_D L(\theta_k, c)\delta_k(x^{(n)}, dc). \tag{3.3}$$

This problem is invariant under the group of permutations of coordinates, and we henceforth restrict attention to those procedures invariant under this group in discussing stringent or approximately stringent solutions, as justified by Theorem 2.2. (A decision function $\delta^{(n)}$ is invariant under this group if, and only if, δ_1 depends on $X_2, \cdots, X_n$ symmetrically and each coordinate δ_k may be obtained from δ_1 by interchanging X_k and X_1 in δ_1; thus an invariant $\delta^{(n)}$ is specified completely by the form of any one of its coordinates.)

We call the decision problem defined above the *n-stage multivariate problem.*

Parallel to this we will consider the *n-stage empirical Bayes problem*: the parameter space is Π and the experiment space is $(\mathfrak{X}^{(n)}, \mathcal{S}^{(n)})$. For each $\pi \varepsilon \Pi$, we define P_π on $(\mathfrak{X}^{(n)}, \mathcal{S}^{(n)})$ by

$$\begin{aligned} P_\pi[X_1 \varepsilon S_1, \cdots, X_n \varepsilon S_n] \\ = \int \cdots \int P_{\theta_1}[X_1 \varepsilon S_1] \cdots P_{\theta_n}[X_n \varepsilon S_n]\, d\pi(\theta_1) \cdots d\pi(\theta_n). \end{aligned} \tag{3.4}$$

In other words, a vector $\theta^{(n)}$ in $\Theta^{(n)}$ is selected by choosing the components of $\theta^{(n)}$ independently with distribution π, and then $(X_1, \cdots, X_n)$ is given distribution $P_{\theta^{(n)}}$. The decision space is $(D, \mathfrak{D})$, so a decision function δ_n is a function on $\mathfrak{X}^{(n)}$ to the reals such that $\delta_n(x^{(n)}, \cdot)$ is a probability distribution on $\mathfrak{D}$ for each $x^{(n)} \varepsilon \mathfrak{X}^{(n)}$ and $\delta_n(\cdot, C)$ is $\mathcal{S}^{(n)}$ measurable in $x^{(n)}$ for each $C \varepsilon \mathfrak{D}$. We then define the risk function R_n^* directly by

$$R_n^*(\pi, \delta_n) = \textstyle\int_\Theta \pi(d\theta_1) \cdots \int_\Theta \pi(d\theta_n) \int L_n(\theta_n, \delta_n, x^{(n)}) P_{\theta^{(n)}}(dx^{(n)}) \tag{3.5}$$

where $L_n(\theta_n, \delta_n, x^{(n)})$ is given by (3.3).

In particular, $R_1^*(\pi, \delta) = R^*(\pi, \delta)$ as defined by (2.3), and, in fact, the 1-stage empirical Bayes problem is simply the Bayesian problem.

It is important to note that, for every n,

$$\inf_{\delta_n} R_n^*(\pi, \delta_n) = \inf_\delta R^*(\pi, \delta) = r^*(\pi), \tag{3.6}$$

since, taking δ_n to be a function of X_n, we see that the left side is at most $r^*(\pi)$, while, given π, the use of a δ_n based on $X_1, \cdots, X_n$ to make a decision about θ_n amounts to using a randomized decision in the Bayesian problem, since $X_1, \cdots, X_{n-1}$ are independent of θ_n and X_n under P_π.

The significance of the empirical Bayes formulation is that, if π is not known, then some information about π is obtained from the sample $X_1, \cdots, X_n$. In [10], Robbins considers conditions under which a sequence of decision functions, δ_n for the n-stage empirical Bayes problem, $n = 1, 2, 3, \cdots$, exists such that, as $n \to \infty$,

$$R_n^*(\pi, \delta_n) \to r^*(\pi) \tag{3.7}$$

for every $\pi \varepsilon \Pi$. Several approaches to this problem, using the theory of stringent procedures, are possible. In particular, remembering that Π is the "parameter space" for this problem, we may take the equivalence relation on Π to be equality, so the envelope function is $r^*(\pi)$. If the excess relative to equality for the n-stage empirical Bayes problem is $\bar{e}_n$, and if $\bar{e}_n \to 0$ as $n \to \infty$, it then follows, not only that a sequence δ_n exists satisfying (3.7), but also that δ_n can be chosen so (3.7) holds uniformly in π. Of course, in many problems the convergence of $\bar{e}_n$ to 0 may not hold, or even if it does hold, the convergence may be very slow. Thus, for a given value of n, stringent procedures relative to less fine equivalences on Π may be of interest. The remarks of Robbins in [10], particularly on pages 8 to 11, obviously have a natural interpretation in terms of stringency. We may, for example, choose a sequence of equivalence relations on Π, for the n-stage empirical Bayes problems, so both the envelope functions and excesses behave nicely as $n \to \infty$. Or, perhaps more realistically in terms of what can be computed in most problems, the choice of the equivalence relation in the n-stage problem may be based on certain features of π that can be estimated easily from the sample $X_1, \cdots, X_n$. For example, we may identify those distributions π having a specified set of moments or quantiles the same.

To treat the general case of an equivalence, $\sim$, on Π, we first impose the hypothesis that for each equivalence class $V = \sim(\pi)$,

$$\inf_\delta \sup_{\pi \epsilon V} R^*(\pi, \delta) = \sup_{\pi \epsilon V} \inf_\delta R^*(\pi, \delta). \tag{3.8}$$

In other words, we assume each of the Bayesian (1-stage empirical Bayes) restricted decision problems has a value. The envelope function, $r_\sim^*$, relative to $\sim$, for the Bayesian problem is defined on Π by the left side of (3.8) and becomes

$$r_\sim^*(\pi) = \sup_{\pi' \varepsilon \sim(\pi)} r^*(\pi'). \tag{3.9}$$

Now if $r_\sim^{*,n}$ is the envelope function relative to $\sim$ for the n-stage empirical Bayes problem, then it is clear that $r_\sim^{*,n} \leqq r_\sim^*$ (consider decision functions depending only on X_n), but then (3.6) and (3.9) yield

$$r_\sim^{*,n}(\pi) = \inf_{\delta_n} \sup_{\pi' \varepsilon \sim(\pi)} R_n^*(\pi', \delta_n)$$
$$\geqq \sup_{\pi' \varepsilon \sim(\pi)} r^*(\pi') = r_\sim^*(\pi),$$

hence $r_\sim^* = r_\sim^{*,n}$ is the envelope function for the n-stage problem, as well.

As has been intimated, the n-stage multivariate and n-stage empirical Bayes problems are related. To explore this relationship, we first define a mapping τ_n from $\Theta^{(n)}$ to Π by letting $\tau_n(\theta_1, \cdots, \theta_n)$ be the distribution obtained by attaching probability n^{-1} to each of the points $\theta_1, \cdots, \theta_n$. (If a particular value θ is repeated k times in the sequence, then the total probability attached to θ is k/n.) Then we define an equivalence relation, $\sim^n$, on $\Theta^{(n)}$ by setting

$$(\theta_1, \cdots, \theta_n) \sim^n (\theta_1', \cdots, \theta_n') \Leftrightarrow \tau_n(\theta_1, \cdots, \theta_n) \sim \tau_n(\theta_1', \cdots, \theta_n').$$

The envelope function, r_n, relative to $\sim^n$, is defined on $\Theta^{(n)}$ by

$$(3.10) \qquad r_n(\theta^{(n)}) = \inf_{\delta^{(n)}} \sup_{\varphi^{(n)} \varepsilon \sim^n(\theta^{(n)})} R_n(\varphi^{(n)}, \delta^{(n)}).$$

Throughout the remainder of this section we impose the following additional hypothesis on the equivalence relation $\sim$: *for each n and equivalence class $\sim^n (\theta^{(n)})$ in $\Theta^{(n)}$*,

$$(3.11) \quad \inf_{\delta^{(n)}} \sup_{\varphi^{(n)} \varepsilon \sim^n(\delta^{(n)})} R_n(\varphi^{(n)}, \delta^{(n)}) = \sup_{\varphi^{(n)} \varepsilon \sim^n(\theta^{(n)})} \inf_{\delta^{(n)}} R(\varphi^{(n)}, \delta^{(n)}).$$

In other words, we assume that each of the n-stage multivariate restricted decision problems, as determined by $\sim^n$, has a value.

In the remainder of this section we establish that under certain conditions the envelope functions for the n-stage multivariate and n-stage empirical Bayes problems are close for large n.

The decision functions are also related. Thus the nth coordinate of a decision function for the n-stage multivariate problem can be used as a decision function for the n-stage empirical Bayes problem. Conversely, a decision function for the latter problem generates a unique invariant decision function for the former problem. The hope is that stringent or nearly stringent solutions for one of the n-stage problems will correspond to nearly stringent solutions for the other n-stage problem. Such a relationship will not be valid in general, but under certain conditions it may be used to obtain reasonable solutions to complex problems. At the end of this section we discuss the possibility of obtaining decision functions for the n-stage multivariate problem in this way, and in Section 4 we illustrate with two examples.

THEOREM 3.1. $r_n \leqq r_\sim^* \circ \tau_n$.

PROOF. Given an equivalence class T in $\Theta^{(n)}$ and an $\epsilon > 0$, let V denote the corresponding equivalence class in Π and choose a δ for the 1-stage empirical Bayes problem so that, for $\pi \varepsilon V$,

$$R^*(\pi, \delta) \leqq r_\sim^*(\pi) + \epsilon.$$

Then let $\delta^{(n)}$ be the decision function for the n-stage multivariate problem whose kth coordinate is $\delta_k = \delta(X_k)$. Applying (3.2), we have, for $\theta^{(n)} \varepsilon T$ and $\pi \varepsilon V$,

$$R_n(\theta^{(n)}, \delta^{(n)}) = n^{-1}\sum_{k=1}^{n} R(\theta_k, \delta)$$
$$= R^*(\tau_n(\theta^{(n)}), \delta) \leqq r_\sim^*(\pi) + \epsilon,$$

and the theorem follows since $\epsilon > 0$ was arbitrary.

An equivalence relation $\sim$ on Π is *preserved under convex combinations* if, whenever $\pi_1 \sim \pi_1'$, $\pi_2 \sim \pi_2'$ and $0 \leqq w \leqq 1$, it follows that

$$w\pi_1 + (1 - w)\pi_2 \sim w\pi_1' + (1 - w)\pi_2'.$$

In this case the envelope function is concave on Π since

$$\begin{aligned} r_\sim^*(w\pi_1 &+ (1 - w)\pi_2) \\ &\geqq \inf_{\delta\epsilon\Delta} \sup_{\pi_1'\sim\pi_1, \pi_2'\sim\pi_2} R^*(w\pi_1' + (1 - w)\pi_2', \delta) \\ &= \inf_{\delta\epsilon\Delta} \{w \sup_{\pi_1'\sim\pi_1} R^*(\pi_1', \delta) + (1 - w) \sup_{\pi_2'\sim\pi_2} R^*(\pi_2', \delta)\} \\ &\geqq wr_\sim^*(\pi_1) + (1 - w)r_\sim^*(\pi_2). \end{aligned}$$

An analogous property holds for the envelope functions r_n of the n-stage multivariate problem:

LEMMA 3.1. *If the equivalence relation $\sim$ is preserved under convex combinations, then*

$$(3.12) \quad (j + k)r_{j+k}(\theta_1, \cdots, \theta_{j+k}) \geqq jr_j(\theta_1, \cdots, \theta_j) + kr_k(\theta_{j+1}, \cdots, \theta_{j+k}).$$

PROOF. Let T, T' be equivalence classes in $\theta^{(j)}$, $\theta^{(k)}$, respectively. Since $\sim$ is preserved under convex combinations, it follows that

$$\tau_{j+k}(\theta^{(j)}, \theta^{(k)}) = j(j + k)^{-1}\tau_j(\theta^{(j)}) + k(j + k)^{-1}\tau_k(\theta^{(k)})$$

is in the same equivalence class in Π for every $\theta^{(j)} \varepsilon T$, $\theta^{(k)} \varepsilon T'$, where $(\theta^{(j)}, \theta^{(k)})$ denotes the vector in $\Theta^{(j+k)}$ obtained by following the coordinates of $\theta^{(j)}$ by those of $\theta^{(k)}$. Let T'' be the equivalence class in $\Theta^{(j+k)}$ so determined.

Now given $\epsilon > 0$, by (3.10) and (3.11), there exist $\varphi^{(j)}$, $\varphi^{(k)}$ such that

$$jr_j(\theta^{(j)}) + kr_k(\theta^{(k)}) - \epsilon \leqq \inf_{\delta^{(j)}} jR_j(\varphi^{(j)}, \delta^{(j)}) + \inf_{\delta^{(k)}} kR_k(\varphi^{(k)}, \delta^{(k)}).$$

But, letting $\psi^{(j+k)} = (\varphi^{(j)}, \varphi^{(k)})$,

$$\begin{aligned} \inf_{\delta^{(j)}} jR_j(\varphi^{(j)}, \delta^{(j)}) &= \inf_{\delta^{(j)}} \sum_{i=1}^{j} \int L_j(\varphi_i, \delta_i, x^{(j)})P_{\varphi^{(j)}}(dx^{(j)}) \\ &\leqq \inf_{\delta^{(j+k)}} \sum_{i=1}^{j} \int L_{j+k}(\psi_i, \delta_i, x^{(j+k)})P_{\psi^{(j+k)}}(dx^{(j+k)}). \end{aligned}$$

Since from each decision function $\delta^{(j+k)}$ a randomized decision function for the j-stage multivariate problem can be constructed by performing an experiment independent of $X_1, \cdots, X_j$ to obtain $X_{j+1}, \cdots, X_{j+k}$ distributed by the parameters $\psi_{j+1}, \cdots, \psi_{j+k}$ and then applying the first j coordinates of $\delta^{(j+k)}$ as a function of $X_1, \cdots, X_j$. Similarly,

$$\inf_{\delta^{(k)}} R_k(\varphi^{(k)}, \delta^{(k)}) \leqq \inf_{\delta^{(j+k)}} \sum_{i=j+1}^{j+k} \int L_{j+k}(\psi_i, \delta_i, x^{(j+k)})P_{\psi^{(j+k)}}(dx^{(j+k)}).$$

But then, since $\psi^{(j+k)} \varepsilon T''$,

$$jr_j(\theta^{(j)}) + kr_k(\theta^{(k)}) - \epsilon \leqq \inf_{\delta^{(j+k)}} \sum_{i=1}^{j+k} \int L_{j+k}(\psi_i, \delta_i, x^{(j+k)}) P_{\psi^{(j+k)}}(dx^{(j+k)})$$
$$\leqq (j+k) r_{j+k}(\theta^{(j)}, \theta^{(k)}),$$

and, since $\epsilon > 0$ was arbitrary, the lemma is proved.

Let $\Pi_0 = \bigcup_{n=1}^{\infty} \tau_n(\Theta^{(n)})$, so Π_0 is the set of all distributions π in Π that concentrate on a finite number of points in Θ and such that $\pi(\{\theta\})$ is a rational number for each $\theta \varepsilon \Theta$. We introduce the following regularity hypotheses, which will be used in some of the subsequent results:

(A) *For each equivalence class V in Π, there exists a finite $n(V)$ such that, for each $n \geqq n(V)$, $\tau_n(\Theta^{(n)}) \cap V$ is non-empty.*

(B) *For each equivalence class V in Π,*

$$\sup_{\pi \varepsilon V \cap \Pi_0} r^*(\pi) = \sup_{\pi \varepsilon V} r^*(\pi).$$

(C) *For each equivalence class V in Π and each integer n and each decision function δ_n for the n-stage empirical Bayes problem such that $R_n^*(\pi, \delta_n)$ is bounded,*

$$\sup_{\pi \epsilon V \cap \Pi_0} R_n^*(\pi, \delta_n) = \sup_{\pi \varepsilon V} R_n^*(\pi, \delta_n).$$

Let e_n denote the excess relative to $\sim^n$ for the n-stage multivariate problem and e_n^* denote the excess relative to $\sim$ for the n-stage empirical Bayes problem. Note that we can define r_n on equivalence classes V in Π if we set $r_n(V) = r_n(\theta^{(n)})$ if $\tau_n(\theta^{(n)}) \varepsilon V$ and $r_n(V) = 0$ if $\tau_n(\Theta^{(n)}) \cap V$ is empty. We use this device to state subsequent results.

LEMMA 3.2. *If the equivalence relation $\sim$ is preserved under convex combinations and* (A) *is valid, then* $\lim r_n(V)$ *exists for every equivalence class V in Π.*

PROOF. The argument is a minor modification of that needed in [9], p. 17, number 98. Let $j \geqq n(V)$ (see hypothesis (A)) and $(i+1)j \leqq k \leqq (i+2)j$. Then it follows from Lemma 3.1 that

$$r_{ij}(V) \geqq r_j(V),$$
$$kr_k(V) \geqq ijr_{ij}(V) + (k - ij) r_{k-ij}(V),$$

hence

$$r_k(V) \geqq (ij/k) r_j(V) \geqq (1 - 2j/k) r_j(V).$$

Letting $k \to \infty$, $\lim\inf r_k(V) \geqq r_j(V)$, and, letting $j \to \infty$, $\lim\inf r_k(V) \geqq \lim\sup r_j(V)$, proving the lemma.

LEMMA 3.3. *If the equivalence relation $\sim$ is preserved under convex combinations and* (A) *is valid, then*

$$\inf_{\delta_n} \sup_V \{\sup_{\pi \varepsilon V \cap \Pi_0} R_n^*(\pi, \delta_n) - \lim r_k(V)\} \leqq e_n, \tag{3.13}$$

where V ranges over all the equivalence classes in Π.

PROOF. Given $\epsilon > 0$, choose a decision function $\delta^{(n)}$ for the n-stage multivariate problem so that $\delta^{(n)}$ is invariant under permutations of coordinates and

$$R_n(\theta^{(n)}, \delta^{(n)}) \leqq r_n(\theta^{(n)}) + e_n + \epsilon. \tag{3.14}$$

Let δ_k be the kth coordinate of $\delta^{(n)}$, and, given V, choose $j \geqq n(V)$ and let $\pi \,\varepsilon\, \tau_j(\Theta^{(j)}) \cap V$. By (3.5),

$$R_n^*(\pi, \delta_n) = j^{-n} \textstyle\sum_{i_1=1}^{j} \cdots \sum_{i_n=1}^{j} \int L_n(\theta_{i_n}, \delta_n, x^{(n)}) P_{(\theta_{i_1}, \cdots, \theta_{i_n})}(dx^{(n)}).$$

From the invariance of $\delta^{(n)}$ under permutations of coordinates and the symmetry of the above sums, it follows that the value of the formula is unchanged if θ_{i_n} and δ_n are replaced by θ_{i_k} and δ_k, simultaneously, for any $1 \leqq k \leqq n$. Hence the integral can be replaced by

$$n^{-1} \textstyle\sum_{k=1}^{n} \int L_n(\theta_{i_k}, \delta_k, x^{(n)}) P_{(\theta_{i_1}, \cdots, \theta_{i_n})}(dx^{(n)}) = R_n((\theta_{i_1}, \cdots, \theta_{i_n}), \delta^{(n)}).$$

Applying the upper bound (3.14) on the right side and then using Lemma 3.1 repeatedly, we obtain

$$R_n^*(\pi, \delta_n) \leqq r_{nj^n}(\theta^{(nj^n)}) + e_n + \epsilon,$$

where $\theta^{(nj^n)}$ is obtained by laying the j^n vectors $(\theta_{i_1}, \cdots, \theta_{i_n})$ end to end. But then

$$\tau_{nj^n}(\theta^{(nj^n)}) = \pi.$$

Also, by Lemma 3.1, $r_{2k}(V) \geqq r_k(V)$ for each k, and, by Lemma 3.2, $\lim r_k(V)$ exists, hence

$$R_n^*(\pi, \delta_n) \leqq \lim r_k(V) + e_n + \epsilon.$$

This holds for all $\pi \,\varepsilon\, V \cap \Pi_0$, so

$$\sup_{\pi \epsilon V \cap \Pi_0} R_n^*(\pi, \delta_n) - \lim r_k(V) \leqq e_n + \epsilon.$$

Formula (3.13) follows directly from this, and the lemma is proved.

Combining Lemma 3.3, (C) and Theorem 3.1, we have

THEOREM 3.2. *If the equivalence relation $\sim$ is preserved under convex combinations and* (A) *and* (C) *are valid, then $e_n^* \leqq e_n$ for each n.*

THEOREM 3.3. *If the equivalence relation $\sim$ is preserved under convex combinations and* (A), (B) *and* (C) *are valid and $e_n \to 0$ as $n \to \infty$, then $r_n(V) \to r_\sim^*(V)$ as $n \to \infty$ for every equivalence class V in Π.*

PROOF. For each δ_n, by (C), (3.6), (3.8) and (3.9),

$$\sup_{\pi \epsilon V \cap \Pi_0} R_n^*(\pi, \delta_n) = \sup_{\pi \epsilon V} R_n^*(\pi, \delta_n) \geqq r_\sim^*(V).$$

Hence by Lemma 3.3,

$$\sup_V (r_\sim^*(V) - \lim r_k(V)) \leqq e_n .$$

Letting $n \to \infty$ we have $r_\sim^*(V) \leqq \lim r_k(V)$, and the theorem follows upon combining this inequality with Theorem 3.1.

The examples in the next section are based on estimation of Bayes solutions to obtain stringent or nearly stringent solutions to the n-stage empirical Bayes problem for large n. These decision functions are then seen to determine decision functions with nearly stringent properties for the n-stage multivariate problem.

The success of this procedure depends on two factors. First, the risk functions r_n and r_n^* must be uniformly close for large n. Theorem 3.3 suggests that this may be the case under certain conditions, but unfortunately the convergence in Theorem 3.3 is not uniform. Second, to apply this procedure in the multivariate case it is necessary to estimate the Bayes solutions to the distributions $\tau_n(\theta^{(n)})$, and the estimator must be of comparable accuracy in the multivariate case to that in the empiricial Bayes case. Fortunately the situation is often favorable here.

For example, if the Bayes solution depends on a real valued parameter ϕ and for some function f on $\mathfrak{X}$ to the real line $E_\pi f(X) = \int E_\theta f(X)\pi(d\theta) = \phi(\pi)$ for all π and this unbiased estimator $f(X)$ has finite variance, then we may use the estimator

$$\hat{\phi}_n = n^{-1} \sum_{k=1}^n f(X_k)$$

in either n-stage problem. Clearly, $\hat{\phi}_n$ is an unbiased estimator of ϕ in the n-stage empirical Bayes problem, while in the n-stage multivariate problem

$$E_{\theta(n)}\hat{\phi}_n = n^{-1} \sum_{k=1}^n E_{\theta_k} f(X_k) = \phi(\tau_n(\theta^{(n)})),$$

so again $\hat{\phi}_n$ is unbiased. The variance in the n-stage multivariate problem is

$$\mathrm{Var}_{\theta(n)}\, \hat{\phi}_n = n^{-2} \sum_{k=1}^n \mathrm{Var}_{\theta_k} f(X_k),$$

but in the n-stage empirical Bayes problem it is

$$\begin{aligned} \mathrm{Var}_\pi\, \hat{\phi}_n &= n^{-1} \int E_\theta(f(X) - \phi(\pi))^2 \pi(d\theta) \\ &= n^{-1} \int \{\mathrm{Var}_\theta f(X) + (E_\theta f(X) - \phi(\pi))^2\} \pi(d\theta). \end{aligned}$$

When $\pi = \tau_n(\theta^{(n)})$ this becomes

$$\mathrm{Var}_\pi\, \hat{\phi}_n = \mathrm{Var}_{\theta(n)}\, \hat{\phi}_n + n^{-2} \sum_{k=1}^n (E_{\theta_k} f(X) - \phi(\pi))^2,$$

so the variance of the estimator in the multivariate case is less than or equal to the variance in the empirical Bayes case.

4. Two examples. (a) Estimation of binomial p with squared error loss.

The basic problem is to estimate the binomial parameter p from the observation of X having distribution $B(j, p)$, with squared error loss. If p has *a priori* distribution π, then

$$P_\pi[X = i] = \binom{j}{i} \int p^i (1-p)^{j-i}\, d\pi(p)$$

depends only on the first j moments of π. Thus a natural equivalence relation on Π for this problem identifies those *a priori* distributions having the same first j moments. The Bayes solution, on the other hand, depends on the first $j + 1$

moments of the *a priori* distribution. In this problem we can, and do, restrict attention to non-randomized estimates.

The n-stage empirical Bayes problem for the case $j = 1$ has been discussed by Robbins in [10], but will bear some further examination here. If π is a distribution with mean μ and variance $\lambda\mu(1 - \mu)$ (where, necessarily, $0 \leqq \lambda \leqq 1$), then the Bayes solution for the $B(1, p)$ case is

$$\delta_\pi(X) = \lambda X + (1 - \lambda)\mu \tag{4.1}$$

and the Bayes envelope function is

$$r^*(\pi) = \lambda(1 - \lambda)\mu(1 - \mu). \tag{4.2}$$

Thus, setting $\pi \sim \pi'$ when π and π' have the same mean, we have

$$r_\sim^*(\pi) = \mu(1 - \mu)/4, \tag{4.3}$$

where μ is the mean of π; a distribution π with mean μ and $\lambda = \frac{1}{2}$ is least favorable for the restricted problem, and the restricted minimax solution is

$$\delta_\mu(X) = (X + \mu)/2. \tag{4.4}$$

The hypotheses of Section 3 are easily seen to be satisfied in this case.

Now in the empirical Bayes n-stage problem, as Robbins states in [10], it is natural to estimate the (unknown) value of μ by $\bar{X} = (X_1 + \cdots + X_n)/n$, and then use

$$\delta_n(X_1, \cdots, X_n) = (X_n + \bar{X})/2 \tag{4.5}$$

as an approximation to the stringent solution.

For this δ_n a little algebra will show that

$$R_n^*(\pi, \delta_n) \leqq r_\sim^*(\pi) + 3/16n \tag{4.6}$$

with the right side being obtained when π concentrates on $\{\frac{1}{2}\}$, since then $n\bar{X}$ is $B(n, \frac{1}{2})$ and

$$\begin{aligned} R_n^*(\pi, \delta_n) &= \text{Var}(\delta_n) = \tfrac{1}{4}\{\text{Var}(X_n) + 2\,\text{Cov}(X_n, \bar{X}) + \text{Var}\,\bar{X}\} \\ &= 1/16 + 3/16n. \end{aligned} \tag{4.7}$$

A solution δ_n of the form

$$\delta_n(X_1, \cdots, X_n) = w_1X_n + w_2\bar{X} + \tfrac{1}{2}w_3, \tag{4.8}$$

where

$$\begin{gathered} w_1 + w_2 + w_3 = 1 \\ nw_1^2 + 2w_1w_2 + w_2^2 - nw_3^2 = n/4 \\ (w_1 + w_2)^2 - w_3^2 + 2(n - 1)w_1 = n - 1, \end{gathered} \tag{4.9}$$

and, as $n \to \infty$,

(4.10) $w_1 = \frac{1}{2} + O(n^{-1}),\ w_2 = \frac{1}{2} - (2n^{\frac{1}{2}})^{-1} + O(n^{-1}),\ w_3 = (2n^{\frac{1}{2}})^{-1} + O(n^{-1}),$

can be shown to have risk

(4.11) $R_n^*(\pi, \delta_n) = r_\sim^*(\pi) + w_3^2/4 = r_\sim^*(\pi) + 1/16n + O(n^{-\frac{3}{2}}).$

That this δ_n is stringent can then be shown by applying Theorem 2.1, since $R_n^*(\pi, \delta_n) - r_\sim^*(\pi) =$ constant, provided we can show that δ_n is a Bayes solution for the n-stage empirical Bayes problem.

Now select the parameter μ according to a beta distribution with density of the form $c\mu^{a-1}(1 - \mu)^{a-1}$, $a > 0$, and then use an *a priori* distribution with mean μ and variance $\lambda\mu(1 - \mu)$, the value of λ being fixed, for the n-stage empirical Bayes problem. The Bayes solution with respect to this distribution on $p_1, \cdots, p_n$ is then

$$\begin{aligned} E(p_n \mid X_1, \cdots, X_n) &= E\{E(p_n \mid X_1, \cdots, X_n, \mu) \mid X_1, \cdots, X_n\} \\ &= E\{\lambda X_n + (1 - \lambda)\mu \mid X_1, \cdots, X_n\} \\ &= \lambda X_n + (1 - \lambda)E\{\mu \mid X_1, \cdots, X_n\} \\ &= \lambda X_n + (1 - \lambda)\{(n/(n + 2a))\bar{X} + (2a/(n + 2a))\tfrac{1}{2}\} \end{aligned}$$

since $n\bar{X}$ has the conditional distribution $B(n, \mu)$, given μ. The first equation in (4.9) is then satisfied, and it is easily seen that values of λ and a can be chosen so that δ_n as given by (4.8) is the Bayes solution.

In the multivariate n-stage problem, the equivalence classes induced by $\sim$ are the hyperplanes specified by

(4.12) $$\bar{p} = (p_1 + \cdots + p_n)/n = \text{constant}.$$

As suggested at the end of Section 3, for the multivariate problem we use the estimator $\delta^{(n)}$ whose kth coordinate is

(4.13) $$\delta_k(X_1, \cdots, X_n) = (X_k + \bar{X})/2.$$

Again, a little algebra will show that

(4.14) $$r_n((p_1, \cdots, p_n), \delta^{(n)}) \leqq \bar{p}(1 - \bar{p})/4 + 3/16n.$$

We have discussed the case $j = 1$ in considerable detail because it reveals the essential elements of the problem in their simplest form. For larger values of j, the restricted minimax solution given the first j moments of π can be calculated and then approximately stringent solutions can be formed, by estimating the moments of π, as in (4.5). We content ourselves with noting that, for $j = 2$, the restricted minimax solution given π has mean μ and variance $\lambda\mu(1 - \mu)$ is

(4.15) $$\delta(X) = (\lambda X + \mu)/(1 + \lambda).$$

(b) Estimation of a location parameter with squared error loss.

The basic problem is to estimate a location parameter θ from the observation $X = \theta + Z$, where Z has a distribution not depending on θ, with squared error loss. If the values of θ are unrestricted, then Z must have finite variance in order for this decision problem to have a finite minimax risk, and to fix ideas we assume Z has mean 0 and variance 1. If the characteristic function of Z does not vanish on any non-degenerate interval, then distinct *a priori* distributions for θ yield distinct distributions of X (see [12]), so from this viewpoint the natural equivalence on the space Π of *a priori* distributions of θ is equality.

On the other hand, if we take a partially non-parametric attitude by assuming that, while Z has mean 0 and variance 1, we do not know the exact distribution of Z, then we will see that a natural equivalence, $\sim$, on Π identifies those distributions having the same finite mean and variance. (Those distributions for which the variance does not exist will be considered as constituting a single equivalence class, whose members will be said to have *infinite variance.*)

Again with squared error loss we need consider only non-randomized estimators. Suppose we take δ to be a linear estimator,

$$\delta(X) = aX + b, \tag{4.16}$$

in the Bayes problem. Then

$$\begin{aligned} R^*(\pi, \delta) &= E_\pi(aZ + b - (1 - a)\theta)^2 \\ &= E(aZ + b)^2 - 2(1 - a)bE_\pi\theta + (1 - a)^2E_\pi\theta^2 \end{aligned}$$

depends only on the first two moments of Z and θ. Thus in each equivalence class of Π the risk is constant for a given linear estimator. But it is a quadratic function of a and b, and it is easily seen that, if π has mean μ and variance σ^2, then the risk is minimized by the estimator δ_{μ,σ^2} with

$$a = \sigma^2/(1 + \sigma^2), \qquad b = \mu/(1 + \sigma^2) \tag{4.17}$$

(where, if the variance is infinite, we set $a = 1$, $b = 0$), and the constant value of the risk on the equivalence class is then

$$R^*(\pi, \delta_{\mu,\sigma^2}) = \sigma^2/(1 + \sigma^2) \tag{4.18}$$

(the value being 1 if the variance is infinite).

If Z is normally distributed, then this estimator will be recognized as the Bayes solution when π is normal with mean μ and variance σ^2, thus δ_{μ,σ^2} is an admissible restricted minimax solution for this case. But the estimator δ_{μ,σ^2} will perform in the same way whatever distribution "nature picks" for Z having mean 0 and variance 1.

To consolidate these ideas, we can redefine the parameter space to be $\Theta \times \mathfrak{F}$, where $\mathfrak{F}$ is the class of all distributions with mean 0 and variance 1. The distribution $P_{\theta,F}$ of X is then obtained by setting $X = \theta + Z$, where Z is a random variable having distribution $F \in \mathfrak{F}$. *A priori* distributions are then formally defined on $\Theta \times \mathfrak{F}$, but we still define two distributions to be equivalent if the mean and variance of their marginals on Θ coincide, so in identifying equivalence classes we can restrict attention to Π, the space of marginals on Θ of the *a priori* distributions. Since nature can now choose the distribution of Z to be normal with probability 1, and then give θ a normal distribution, the estimators δ_{μ,σ^2} are admissible restricted minimax solutions for this problem and

$$r_{\sim}^*(\pi) = \sigma^2/(1 + \sigma^2). \tag{4.19}$$

Now in the n-stage empirical Bayes problem, we might estimate the mean, μ, of π by $\bar{X} = (X_1 + \cdots + X_n)/n$, and the variance σ^2, of π by $\hat{\sigma}^2 = (\sum_{k=1}^n (X_k - \bar{X})^2 - (n-1))^+/(n-1)$, where $(a)^+$ equals a if $a \geqq 0$ and 0 if $a \leqq 0$, and then set

$$\delta_n(X_1, \cdots, X_n) = (\hat{\sigma}^2 X_n + \bar{X})/(1 + \hat{\sigma}^2). \tag{4.20}$$

Similarly, in the multivariate problem, we may define the kth coordinate of $\delta^{(n)}$ by

$$\delta_k(X_1, \cdots, X_n) = (\hat{\sigma}^2 X_k + \bar{X})/(1 + \hat{\sigma}^2). \tag{4.21}$$

In [1], the asymptotic behavior of a class of estimators including these as particular cases is investigated.
It is shown there that, for each π,

$$R_n^*(\pi, \delta_n) \to r_{\sim}^*(\pi) \tag{4.22}$$

as $n \to \infty$, and similar results are obtained for the multivariate problem. Under the assumption that the fourth moment of Z is bounded by a fixed constant, the convergence in (4.22) is shown to be uniform in π, with $R_n^*(\pi, \delta_n) - r_{\sim}^*(\pi)$ bounded by a term of order $n^{-\frac{1}{2}}$.

Two other equivalence relations on Π may be considered: namely $\sim'$, identifying those π having the same mean, or $\sim''$, identifying those π having the same second moment. But even if Z is normally distributed, the envelope function for $\sim'$ is identically 1, the same as the over-all minimax risk, so the problem is unchanged. On the other hand, for $\sim''$, the estimator

$$\delta_\gamma(X) = (\gamma/(1 + \gamma))X \tag{4.23}$$

has constant risk

$$r(\pi, \delta) = \gamma/(1 + \gamma) \tag{4.24}$$

in the equivalence class of all π having second moment γ, $\gamma/(1 + \gamma)$ being the value of the corresponding envelope function in this equivalence class.

In this case estimators of the form

$$\delta_k(X_1, \cdots, X_n) = ((\sum_{j=1}^n X_j^2 - n)^+/\sum_{j=1}^n X_j^2)X_k \tag{4.25}$$

for the n-stage multivariate and empirical Bayes problem might be used as approximately stringent solutions. The asymptotic behavior of these estimators again follows from the general result in [1]. The theory of Stein [11], James and Stein [4], using estimators for the multivariate n-stage problem similar to those in (4.25), investigates the asymptotic behavior when Z is normally distributed.

These results may be considered in the context of this problem, and the theory of stringency provides a natural interpretation of their results.

We may attempt to improve the asymptotic situation for the empirical Bayes problem by defining equivalences on Π involving higher moments of the *a priori* distribution; but the restricted minimax solutions then involve higher moments of Z, and we must move closer to the "parametric" case where the distribution of Z is completely known.

In the multivariate problem, the equivalence classes in $\Theta^{(n)}$ corresponding to $\sim''$ are the surfaces of spheres centered at the origin, and the equivalence classes in $\Theta^{(n)}$ corresponding to $\sim$ are $(n - 2)$-dimensional spheres lying in $(n - 1)$-dimensional planes perpendicular to the equiangular line, having center in the equiangular line. Moreover, the envelope risk function is constant on the surfaces of cylinders whose axis is the equiangular line, and the problem is invariant under translations along the equiangular line. In [1] it is shown that equivalence relations in $\Theta^{(n)}$ identifying points on the surfaces of spheres whose centers lie in an arbitrary specified linear subspace can be treated similarly: only the first two moments of Z need be known, and, as the number of coordinates increases, the estimators have asymptotic risk $d^2/(n + d^2)$, where d is the distance of the parameter from the specified subspace. This function is constant on surfaces whose points are the same distance from the specified linear subspace. It is the function $r_{\sim}^* \circ \tau_n$.

REFERENCES

[1] Cogburn, R. (1965). On the estimation of a multivariate location parameter with squared error loss. Bernoulli (1723) Bayes (1763) Laplace (1813), Anniversary Vol. 24–29

[2] Hodges, J. L., Jr. and Lehmann, E. L. (1952). The use of previous experience in reaching statistical decisions. *Ann. Math. Statist.* **23** 396–407.

[3] Hunt, G. and Stein, C. (1946). Most stringent tests of statistical hypotheses. Unpublished.

[4] James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 361–379. Univ. of California Press.

[5] Kiefer, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Statist.* **28** 573–601.

[6] Lehmann, E. L. (1950). Notes on the theory of estimation. Mimeographed notes. University of California, Berkeley.

[7] Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[8] Loomis, L. H. (1953). *An Introduction to Abstract Harmonic Analysis.* Van Nostrand, New York.

[9] Polya, G. and Szego, G. (1925). *Aufgaben und Lehrsatze aus der Analysis, Erster Band.* Springer, Berlin.

[10] Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.

[11] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 197–206. Univ. of California Press.

[12] Teicher, H. (1961). Identifiability of mixtures. *Ann. Math. Statist.* **32** 244–248.

[13] Wesler, O. (1959). Invariance theory and a modified minimax principle. *Ann. Math. Statist.* **30** 1–20.