The Annals of Statistics
1977, Vol. 5, No. 3, 516-521

IMPROVED RATES IN THE EMPIRICAL BAYES MONOTONE MULTIPLE DECISION PROBLEM WITH MLR FAMILY[1]

BY DENNIS C. GILLILAND AND JAMES HANNAN

Michigan State University

In the context of the two-action, linear loss, exponential family multiple decision problem, Van Houwelingen (1973), (1976) has shown that faster rates of convergence are deducible for monotone empirical Bayes procedures than result from application of the bound established for general empirical Bayes procedures by Johns and Van Ryzin (1972). This note generalizes a (1973) Van Houwelingen bound to arbitrary k-action, monotone loss, MLR family multiple decision problems. An example is given to show that the result is a useful alternative to the recent Van Ryzin and Susarla (1977) multiple decision problem generalization of Johns and Van Ryzin.

1. The multiple decision problem. We consider a statistical decision problem with states θ indexing $\mathscr{P} = \{P_\theta \mid \theta \in \Theta\}$ where the P_θ are probability distributions on $(\mathscr{X}, \mathscr{B})$. In a multiple decision problem the action space is finite. Here we denote it by $A = \{1, 2, \cdots, k\}$ and we denote the loss for action a and state θ by $L_a(\theta) \geqq 0$. The class $\mathscr{T}$ of decision rules is the class of all $\mathscr{B}$-measurable mappings t into the $(k-1)$-dimensional simplex of probability distributions on A. That is, $t = (t_1, t_2, \cdots, t_k)$ where $t_a \geqq 0$ and $\sum t_a = 1$. The risk of t at θ is

$$R(\theta, t) = E_\theta(\sum t_a L_a(\theta)) . \tag{1}$$

Let $\mathscr{C}$ be a σ-field of subsets of Θ such that $P_\bullet(B)$ for all $B \in \mathscr{B}$, and $L_a(\cdot)$ for all $a \in A$ are $\mathscr{C}$-measurable. For G a distribution on $\mathscr{C}$, denote the conditional expected loss of action a given x by

$$l_a(x) = E_x(L_a) . \tag{2}$$

The Bayes risk of t at G is

$$R(G, t) = E(\sum t_a l_a) \tag{3}$$

where E is expectation with respect to P, the marginal distribution on $\mathscr{B}$. A decision rule t is Bayes with respect to G if and only if

$$\sum t_a l_a = \bigwedge l_b \qquad \text{a.e.} \quad P . \tag{4}$$

We assume the minimum Bayes risk to be finite, that is,

$$R(G) = \bigwedge_t R(G, t) = E(\bigwedge l_b) < \infty . \tag{5}$$

Received February 1976; revised November 1976.
[1] Research supported by NSF Grant No. MPS-75-07467.
AMS 1970 *subject classifications*. Primary 62C25; Secondary 62F99.
Key words and phrases. Empirical Bayes, monotone multiple decision problem, monotone likelihood ratio family.

2. Empirical Bayes. In this section we develop results which are slightly more general than some of Van Ryzin and Susarla (1977). The development is novel in that it is based on the application a.e. P of simple number sequence bounds of the Appendix.

Let $(\mathscr{X}^*, \mathscr{B}^*, P^*)$ be independent of the above $(\mathscr{X} \times \Theta, \mathscr{B} \times \mathscr{C}, \mathscr{P}, G)$. (In the traditional empirical Bayes problems $\mathscr{X}^* = \mathscr{X}^n$, $\mathscr{B}^* = \mathscr{B}^n$ and $P^* = P^n$.) Mappings t^* from $\mathscr{X}^*$ into $\mathscr{T}$ are sought so that $R(G, t^*)$ is close to $R(G)$ in some sense, for example $E^*R(G, t^*)$ close to $R(G)$. To this end let $l^* = (l_1^*, l_2^*, \cdots, l_k^*)$ be a mapping from $\mathscr{X}^*$ into the class of $(-\infty, \infty)^k$-valued $\mathscr{B}$-measurable functions and suppose that t^* is conditional Bayes with respect to l^*, that is, $\sum t_a^* l_a^* = \bigwedge l_a^*$ a.e. P. For interchanges of P and P^* we also require that l^* be jointly measurable $\mathscr{B}^* \times \mathscr{B}$. Here l^* is an estimator based on x^* of the conditional expected loss function l, and for each x^*, t^* is Bayes with respect to $l^*(x^*)$. Let $l_{ab} = l_a - l_b$ when the difference is defined, that is, not $\infty - \infty$, and let l_{ab} be arbitrary measurable otherwise.

THEOREM 1 (Van Ryzin and Susarla). *For t satisfying* (4),

$$E^*R(G, t^*) - R(G) \leqq \sum_{a=1}^k \sum_{b=1}^k E\{t_b l_{ab} P^*[l_{ab}^* \leqq 0]\} \tag{6}$$

$$\leqq \sum_{a=1}^k \sum_{b=1}^k E\{t_b |l_{ab}| P^*[|l_{ab}^* - l_{ab}| \geqq |l_{ab}|]\} . \tag{7}$$

PROOF. $R(G, t^*) = E(\sum t_a^* l_a)$ and $R(G) = E(\bigwedge l_b)$. In view of the fact that $t_b = 0$ if $l_{ab} < 0$, (6) is a consequence of (c) of the Appendix and (7) is a consequence of (d). □

Theorem 1 is slightly more general than Lemma 1 of Van Ryzin and Susarla (1977) in that arbitrary Bayes rules t and empirical Bayes rules t^* are covered rather than particular determinations.

Suppose that $l_1^*, l_2^*, \cdots, l_k^*$ is monotone decreasing and then monotone increasing, that is, $l_{a+1\,a}^* < 0 < l_{b+1\,b}^*$ implies $a < b$. Let $l_0^* = l_{k+1}^* \equiv \infty$. A t^* such that $\sum t_a^* l_a^* = \bigwedge l_b^*$ is

$$t_a^* = [l_{a\,a-1}^* < 0 \leqq l_{a+1\,a}^*, \cdots, l_{k+1\,k}^*] , \qquad a = 1, 2, \cdots, k , \tag{8}$$

where here and henceforth square brackets denote the indicator function. The linear loss multiple decision problem of Van Ryzin and Susarla (1977) leads naturally to t^* of this nature. For if $-\infty = \theta_0 < \theta_1 < \cdots < \theta_{k-1} < \theta_k = \infty$ and

$$L_a(\theta) = \sum_{b=1}^{a-1} (\theta - \theta_b)_- + \sum_{b=a}^{k-1} (\theta - \theta_b)_+ , \qquad a = 1, 2, \cdots, k , \tag{9}$$

and

$$\delta(x) = E_x(\theta) \quad \text{is finite a.e.} \quad P , \tag{10}$$

then $L_{a+1\,a}(\theta) = \theta_a - \theta$ and $l_{a+1\,a} = \theta_a - \delta$, $a = 1, 2, \cdots, k-1$. So if $l_{a+1\,a}^* = \theta_a - \delta^*$, $a = 1, 2, \cdots, k-1$ where δ^* is an estimator of the conditional expectation δ, (8) becomes

$$t_a^* = [\theta_{a-1} < \delta^* \leqq \theta_a] , \qquad a = 1, 2, \cdots, k . \tag{11}$$

Since here $\sum_1^a t_b{}^* = [\delta^* \leqq \theta_a]$ and t satisfies (4), (e) of the Appendix implies

$$R(G, t^*) - R(G) = E \sum_{a=1}^{k-1} (\delta - \theta_a)\{[\delta^* \leqq \theta_a] - [\delta \leqq \theta_a]\} \tag{12}$$

from which

$$R(G, t^*) - R(G) \leqq E \sum_{a=1}^{k-1} |\delta - \theta_a|[|\delta^* - \delta| \geqq |\delta - \theta_a|] . \tag{13}$$

Taking E^* expectation and weakening the result by the Markov inequality leads to (33) of Van Ryzin and Susarla (1977):

$$E^*R(G, t^*) - R(G) \leqq \sum_{a=1}^{k-1} E\{|\delta - \theta_a|^{1-\varepsilon} E^*|\delta^* - \delta|^\varepsilon\} \qquad \text{for all } \varepsilon > 0 . \tag{14}$$

3. Monotone case with MLR family. The monotone multiple decision problem is defined by Karlin and Rubin (1956) and also by Ferguson (1967, Chapter 6). Here Θ is a subset of the reals and the loss function has the following structure: there exist numbers $\theta_1 \leqq \theta_2 \leqq \cdots \leqq \theta_{k-1}$ in Θ such that, for $a = 1, 2, \cdots, k - 1$,

$$\begin{aligned} L_{a+1\,a}(\theta) &\geqq 0 , \quad \theta < \theta_a \\ &\leqq 0 , \quad \theta > \theta_a . \end{aligned} \tag{15}$$

The linear loss function (9) is an example of a function satisfying (15). Also suppose $\mathscr{P}$ is a monotone likelihood ratio family with densities p_θ, $\theta \in \Theta$. (Without loss of generality we take the likelihood ratio to be increasing.) The monotone procedures form an essentially complete class so that there exists a monotone procedure t which is Bayes versus the prior G, that is, there exist $-\infty = x_0 \leqq x_1 \leqq \cdots \leqq x_{k-1} \leqq x_k = \infty$ such that for $a = 1, 2, \cdots, k$ and all x,

$$t_a(x) > 0 \quad \text{implies} \quad x_{a-1} \leqq x \leqq x_a . \tag{16}$$

Let t^{**} be a monotone rule based on estimates $-\infty = x_0{}^* \leqq x_1{}^* \leqq \cdots \leqq x_{k-1}^* \leqq x_k{}^* = \infty$. For $a = 1, 2, \cdots, k - 1$, let

$$\tau_a(x) = P_{\theta_a}[X \geqq x] , \qquad x \text{ real}, \tag{17}$$

$$\alpha_a = \tau_a(x_a) , \tag{18}$$

and

$$\alpha_a{}^* = \tau_a(x_a{}^*) . \tag{19}$$

THEOREM 2. *Suppose* $l_a < \infty$ a.e. P, $a = 1, 2, \cdots, k$. *It follows that*

$$E^*R(G, t^{**}) - R(G) \tag{20}$$
$$\leqq \sum_{a=1}^{k-1} E\{|l_{a\,a+1}| P^*[|\alpha_a{}^* - \alpha_a| \geqq |\alpha_a - \tau_a|]\}$$
$$\leqq \sum_{a=1}^{k-1} E\{|l_{a\,a+1}|\,|\alpha_a - \tau_a|^{-\varepsilon}\} E^*|\alpha_a{}^* - \alpha_a|^\varepsilon \qquad \textit{for all } \varepsilon > 0 . \tag{21}$$

PROOF. Consider (e) of the Appendix and the facts $R(G, t^{**}) = E(\sum t_a{}^{**} l_a)$ and $R(G) = E(\bigwedge l_b)$. In the monotone case, if $T_a{}^{**}(x) \neq T_a(x)$ then x is between x_a and $x_a{}^*$ and therefore $\tau_a(x)$ is between α_a and $\alpha_a{}^*$, so that an interchange of E^* and E gives (20). (21) is a Markov weakening of (20). □

The proof shows an alternative bound results by replacing τ_a, α_a, $\alpha_a{}^*$ by x, x_a and $x_a{}^*$ respectively, in RHS (20). However, as Van Houwelingen (1976) remarks following his (15), the estimation of tail probabilities α_a is easier to handle than the estimation of cutpoints x_a.

Theorem 2 is a three-fold generalization of Lemma 3.2.2 of Van Houwelingen (1973) from linear to monotone loss, from continuous exponential to MLR family, and from $k = 2$ to $k \geqq 2$ acts. The specialization of (21) to linear loss is

$$(22)\qquad E^*R(G, t^{**}) - R(G) \leqq \sum_{a=1}^{k-1} E\{|\theta_a - \delta|\,|\alpha_a - \tau_a|^{-\varepsilon}\}E^*|\alpha_a{}^* - \alpha_a|^\varepsilon \qquad \text{for all} \quad \varepsilon > 0\,.$$

Here the estimation errors $\alpha_a{}^* - \alpha_a$ are isolated in the bound whereas the Van Ryzin and Susarla bound (14) isolates $\delta^* - \delta$. Since the monotone rules are essentially complete in the monotone multiple decision problem with MLR family, (22) is then a possible useful alternative to (14).

The next section looks at an example used by Van Houwelingen (1973) to show in the $k = 2$ case that faster rates can follow from (22) than from (14).

4. Example. Let P_θ be $N(\theta, 1)$, $\theta \in \Theta = (-\infty, \infty)$ and let the loss be the linear loss (9). This is a monotone multiple decision problem with MLR family so the monotone decision rules are essentially complete. Here a version of the conditional expectation of θ given x is

$$(23)\qquad \delta(x) = x - \frac{p'(x)}{p(x)}\,, \qquad -\infty < x < \infty$$

where p is the marginal density of x and p' its derivative. Let $\delta^*(x) = x - Dp_n(x)/p_n(x)$ where p_n and Dp_n are estimators for p and p' based on $x^* = (x_1, x_2, \cdots, x_n)$ with $x_1, x_2, \cdots$ i.i.d. P, the traditional empirical Bayes estimator of δ. Let t^* be given by (11) and let t^{**} be a monotone rule such that $R(\theta, t^{**}) \leqq R(\theta, t^*)$ for all θ; for example, following Karlin and Rubin (1956, Lemma 4) and Ferguson (1967, Theorem 6.1.1) take $T_a{}^{**}(x) = [x \leqq x_a{}^*]$ where $x_a{}^*$ is determined so that $P_{\theta_a}[X \leqq x_a{}^*] = E_{\theta_a} T_a{}^*$, $a = 1, 2, \cdots, k-1$.

Van Houwelingen (1973) considers only two-action problems. He applies his Theorem 3.2.1 to show (pages 42–43) that for this example and $k = 2$,

$$(24)\qquad \text{RHS (22)} = O(n^{-\varepsilon(r-1)/(2r+1)}) \qquad \text{for all} \quad 0 < \varepsilon < 2$$

provided

$$(25)\qquad G[\theta < \theta_1]G[\theta > \theta_1] > 0$$

and

$$(26)\qquad \int |\theta|^r \, dG < \infty$$

and the estimators p_n and Dp_n are based on sufficiently smooth kernels.

For $k > 2$, RHS (22) is simply a sum of $k - 1$ Van Houwelingen bounds. Since

$$(27)\qquad G[\theta < \theta_1]G[\theta > \theta_{k-1}] > 0$$

implies $G[\theta < \theta_a]G[\theta > \theta_a] > 0$ for $a = 1, 2, \cdots, k-1$, we see that replacing (25) by (27) ensures that (24) obtains if $k > 2$.

If G has finite absolute moments of all order, then by choice of kernels in the estimators p_n and Dp_n, convergence rate arbitrarily close to $O(n^{-1})$ is established for the monotone empirical Bayes procedure t^{**}. Van Houwelingen (1973, Remark 3.2.2) notes the Johns and Van Ryzin (1972) obtain only close to $O(n^{-\frac{1}{2}})$ in the $k = 2$ case while Van Ryzin and Susarla (1977, Section 4) remark that the Johns and Van Ryzin rates carry over to $k > 2$.

In this note we have not considered the question of rates deducible from (21) or (22) in general. Van Houwelingen (1973) and (1976) has considered this question relative to several bounds and two-action, linear loss, continuous exponential family. For example, Van Houwelingen (1976) shows that (26) and conditions slightly stronger than those given in (1973) imply LHS (20) $\doteq O(n^{-2(r-1)/(2r+1)} \log^2 n)$, which is a ratewise improvement over what is deducible from the family of rates (24) for given r. However, for the normal family and G with finite moments of all order, both approaches yield rates arbitrarily close to $O(n^{-1})$ by choice of kernel.

APPENDIX

Let $l = (l_1, l_2, \cdots, l_k)$ be a k-tuple of extended real numbers such that $|\bigwedge l_b| < \infty$ and suppose that $t = (t_1, t_2, \cdots, t_k)$ is a k-tuple of probabilities such that

(a) $$\sum t_a l_a = \bigwedge l_b \, .$$

Suppose that l^* and t^* are similarly defined but require that the components of l^* be real. Letting $l_{ab} = l_a - l_b$, when the difference is defined and arbitrary otherwise, it follows that

(b) $$\sum t_a{}^* l_a - \bigwedge l_b = \sum_a \sum_b t_a{}^* t_b l_{ab}$$

(c) $$\leqq \sum_a \sum_b t_b l_{ab} [l_{ab}^* \leqq 0 \leqq l_{ab}]$$

(d) $$\leqq \sum_a \sum_b t_b |l_{ab}| [|l_{ab}^* - l_{ab}| \geqq |l_{ab}|] \, .$$

Furthermore, if $|l_a| < \infty$ and $T_a{}^* = \sum_1^a t_b{}^*$, $a = 1, 2, \cdots, k$, then

(e) $$\sum t_a{}^* l_a - \bigwedge l_b = \sum_{a=1}^{k-1} l_{a\,a+1}(T_a{}^* - T_a) \, .$$

PROOF. Note that $t_b > 0$ implies $|l_b| < \infty$ so that $t_b l_{ab} = t_b(l_a - l_b)$. Using this fact, (a), and $\sum t_a{}^* = \sum t_b = 1$, we obtain (b). Since $t_a{}^* t_b > 0$ implies $l_{ab}^* \leqq 0 \leqq l_{ab}$, (c) bounds (b). A further weakening gives (d). In case the l_a are finite so that $l_{a\,a+1}$ is defined for all a, $\sum t_a{}^* l_a - \bigwedge l_b = \sum (t_a{}^* - t_a) l_a$ and summation by parts yields (e). □

Acknowledgment. The authors wish to thank the referee for a careful reading and several valuable suggestions, in particular for pointing out the representation (e) of the Appendix.

REFERENCES

FERGUSON, THOMAS S. (1967). *Mathematical Statistics—A Decision Theoretic Approach.* Academic Press, New York.

JOHNS, M. V., JR. and VAN RYZIN, J. (1972). Convergence rates for empirical Bayes two-action problems II. Continuous case. *Ann. Math. Statist.* **43** 934–947.

KARLIN, SAMUEL and RUBIN, HERMAN (1956). The theory of decision procedures for distributions with monotone likelihood ration. *Ann. Math. Statist.* **27** 272–299.

VAN HOUWELINGEN, J. C. (1973). On empirical Bayes rules for the continuous one-parameter exponential family. Ph. D. thesis, Rijksuniversiteit te Utrecht, Netherlands

VAN HOUWELINGEN, J. C. (1976). Monotone empirical Bayes tests for the continuous one-parameter exponential family. *Ann. Statist.* **4** 981–989.

VAN RYZIN, J. and SUSARLA, V. (1977). On the empirical Bayes approach to multiple decision problems. *Ann. Statist.* **5** 172–181.

DEPARTMENT OF STATISTICS AND PROBABILITY
WELLS HALL
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824