*The Annals of Stastistics*
1982, Vol. 10, No. 4, 1277–1282

# BAYES EMPIRICAL BAYES: FINITE PARAMETER CASE[1]

By Dennis C. Gilliland, John E. Boyer, Jr. and How Jan Tsao

*Michigan State University, Kansas State University and Oakridge National Laboratory*

Since Robbins (1956), a large literature has evolved treating the empirical Bayes formulation of a sequence of decision problems. In this paper we look at the formulation and the asymptotic optimality criterion as a classical iid problem and a consistency property. With this point of view and the finite state component problem, we evoke classical results on Bayes procedures to prove a complete class theorem and to establish the asymptotic optimality of Bayes empirical Bayes procedures.

**1. Empirical Bayes.** The empirical Bayes decision problem of Robbins (1956) consists of a sequence of independent repetitions of a given component decision problem. At the $n$th stage, data from the past as well as present stage are available on which to base a decision. In this paper, we consider the empirical Bayes decision problem with a general finite-state component and show that, at each stage $n$, the class of Bayes empirical Bayes procedures is complete. Moreover, we prove that, for a prior with support equal to the simplex of all conponent priors, the Bayes empirical Bayes procedure is asymptotically optimal, i.e., has risk converging to minimum risk as $n \to \infty$.

Consider the component decision problem with observation $X \sim P_\theta$ taking values in $\mathscr{X}$, parameters $\theta \in \Theta$, action space $A$, decision rules $d \in D$, loss $L(\theta, d(X)) \geq 0$, risk $R(\theta, d)$, priors $G \in \mathscr{G}$, Bayes risk $R(G, d)$, Bayes rules $d_G$ and minimum Bayes risk $R(G)$. Here $\theta$ can be anything from an index for a finite set of distributions to a distribution function to be estimated.

The standard empirical Bayes problem is usually formulated as follows. Let $(\theta_1, X_1)$, $\cdots$, $(\theta_n, X_n)$, $\cdots$ be iid with $(\theta, X)$ having distribution $G$ on $\theta$ and, conditional on $\theta$, $P_\theta$ on $X$. For each $n \geq 1$, $\mathbf{X} = (X_1, \cdots, X_n) \sim P_G^n \equiv P_G \times \cdots \times P_G$, where $P_G$ denotes the $G$ mixture of the $P_\theta$. A decision is to be made about $\theta_n$ using $\mathbf{X}$ with loss $L(\theta_n, t_n(\mathbf{X}))$. A sequence $t_n$ is referred to as an empirical Bayes decision procedure. It is said to be asymptotically optimal (a.o.) if

$$\lim_n EL(\theta_n, t_n(\mathbf{X})) = R(G) \quad \text{for all} \quad G \in \mathscr{G}.$$

Empirical Bayes problems with a great variety of components have been treated in the literature. Most work has concerned the development of ad hoc a.o. procedures.

A point of view which this paper adopts (also see Meeden, 1972, Section 4) better exposes the classical structure of this decision problem. Let $\varphi(\check{\mathbf{X}})$ denotes the $\check{\mathbf{X}}$-section of $t_n$ where $\check{\mathbf{X}} = (X_1, \cdots, X_{n-1})$ so that $E[L(\theta_n, t_n(\mathbf{X})) \mid \check{\mathbf{X}}] = R(G, \varphi(\check{\mathbf{X}}))$. We call this "loss" in the decision problem with observation $\check{\mathbf{X}} \sim P_G^{n-1}$ taking values in $\mathscr{X}^{n-1}$, parameters $G \in \mathscr{G}$, action space $D$, decision rules $\varphi \in \Phi$, loss $R(G, \varphi(\check{\mathbf{X}}))$, risk which we denote by $\mathbf{R}(G, \varphi)$, priors $\Lambda \in \mathscr{G}^*$, Bayes risk $\mathbf{R}(\Lambda, \varphi)$, Bayes rules $\varphi_\Lambda$ and minimum Bayes risk $\mathbf{R}(\Lambda)$.

Admissibility considerations for the empirical Bayes rule $t_n$ in terms of $EL(\theta_n, t_n(\mathbf{X}))$ as a function of $G$ correspond to the usual admissibility considerations for $\varphi$ in terms of its risk function $\mathbf{R}(G, \varphi)$. The Robbins a.o. property for the empirical Bayes decision rule $t_n$ is seen to be the classical mean loss consistency of the decision procedure $\varphi$. Schwartz (1965), for example, defines the weak (in prob.) and strong (a.s.) versions.

Received September 1979; revised August 1981 and February 1982.
[1] Research supported by NSF Grants MCS7703493-01 and MCS80-03001.
*AMS* 1970 *subject classifications.* Primary 62C10; secondary 62C25.
*Key words and phrases.* Bayes empirical Bayes, consistency, asymptotic optimality, complete class, Bayes.

It follows that classical results in decision theory for models with iid $P_G$-distributed observations have immediate implications for the empirical Bayes decision problem formulated in terms of the sections $\varphi$. Among these results are complete class and consistency theorems for Bayes procedures. This comment does not presently have a far-reaching impact on empirical Bayes theory because the high dimensionality of the parameter set $\mathscr{G}$ limits the applicability of the classical theorems. However, we believe it is worthwhile to set down some implications of classical Bayes theory to some simpler empirical Bayes problems, in particular, for the finite $\Theta$ component problem.

**2. Bayes empirical Bayes.** For notational convenience replace $n-1$ by $n$ and $\check{\mathbf{x}}$ by $\mathbf{x}$. An empirical Bayes decision rule $\varphi$ is Bayes w.r.t. $\Lambda \in \mathscr{G}^*$ and is denoted by $\varphi_\Lambda$ if it is a minimizer (across $\Phi$) of

$$\mathbf{R}(\Lambda, \varphi) = \int \mathbf{R}(G, \varphi)\Lambda(dG) = \int\int R(G, \varphi(\mathbf{x}))P_G^n(d\mathbf{x})\Lambda(dG). \tag{1}$$

A minimizer can be constructed pointwise (when possible) by choosing $\varphi_\Lambda(\mathbf{x})$ to minimize

$$\int R(G, \varphi(\mathbf{x}))\Lambda(dG \mid \mathbf{x}), \tag{2}$$

where $\Lambda(\cdot \mid \mathbf{x})$ is conditional probability on $G$ given $\mathbf{X} = \mathbf{x}$ in the model $G \sim \Lambda$ and, conditional on $G$, $\mathbf{X} \sim P_G^n$. Since $R(G, \varphi(\mathbf{x}))$ is linear in $G$, it follows from (2) that $\varphi_\Lambda$ is given by

$$\varphi_\Lambda(\mathbf{x}) = d_{\hat{G}} \tag{3}$$

where $\hat{G}$ is the $\Lambda(\cdot \mid \mathbf{x})$ mixture of $G$. (This notation hides the display of dependence of $\hat{G}$ on $\Lambda$ and $\mathbf{x}$.) Note that $\varphi_\Lambda$ is pointwise component Bayes relative to the induced estimator $\hat{G}$. Of course, (3) exposes how loss consistency of $\varphi_\Lambda$ can often be reduced to a question of the consistency of the estimator $\hat{G}$ for $G$.

Lindley (1961, 1971) and other advocates of Bayesian statistics have proposed the use of Bayes rules in the empirical Bayes setting. Deely and Lindley (1981) discuss Bayes empirical Bayes methods and contrast the same with the usual ad hoc empirical Bayes rules. However, from their Bayesian perspective, there is little to no interest in the frequentist risk functions $\mathbf{R}(G, \varphi)$ and asymptotic optimality.

**3. Bayes empirical Bayes – finite $\Theta$.** In this section, $\Theta = \{0, 1, \cdots, m\}$, where $m \geq 1$ and $S$ denotes the risk set in $\mathbb{R}^{m+1}$ of the component decision problem. The class of distributions $\mathscr{G}$ is the $m$-dimensional simplex of probability measures on $\Theta$, $G = (g_0, g_1, \cdots, g_m)$ and $P_G = \sum_0^m g_i P_i$. We let $\mathscr{A}$ denote the $\sigma$-field of subsets for the component observation space $\mathscr{X}$ and adopt the $S$-game point of view.

For each $n > 0$, the class of nonrandomized empirical Bayes rules is

$$\Phi = \{\varphi \mid \varphi = (\varphi^0, \varphi^1, \cdots, \varphi^m) \text{ is an } \mathbf{x}\text{-measurable mapping into } S\}. \tag{4}$$

In the no (previous) data problem, $n = 0$, $\Phi$ is taken to be the empirical Bayes action space $S$. For $\varphi \in \Phi$, $G \in \mathscr{G}$, the loss at $\mathbf{x} \in \mathscr{X}^n$ is

$$R(G, \varphi(\mathbf{x})) = G'\varphi(\mathbf{x}), \tag{5}$$

the inner product of the vectors $G$ and $\varphi(\mathbf{x})$, and the risk is

$$\mathbf{R}(G, \varphi) = \int G'\varphi(\mathbf{x})P_G^n(d\mathbf{x}). \tag{6}$$

REMARK 1. If $S$ is compact and convex, behavioral empirical Bayes rules need not be considered. For if $\varphi$ is a mapping into the probability measures on $S$, then by (5) and

Ferguson (1967, Lemma 2.7.3), it is risk equivalent to the element of $\Phi$ which maps $\mathbf{x}$ into the mean of $\varphi(\mathbf{x})$.

Let $\Lambda \in \mathscr{G}^*$, the class of all probability measures on the Borel subsets of $\mathscr{G}$. The Bayes risk of $\varphi$ w.r.t. to $\Lambda$ is

$$\mathbf{R}(\Lambda, \varphi) = \int \hat{G}'\varphi(\mathbf{x})P_{(\Lambda)}(d\mathbf{x}). \tag{7}$$

In (7), $P_{(\Lambda)}$ denotes the $\Lambda$-mixture of the $P_G^n$, $G \in \mathscr{G}$, and $\hat{G}$ is the conditional expectation

$$\hat{G} = \int G\Lambda(dG \mid \mathbf{x}). \tag{8}$$

REMARK 2. If $S$ is compact, then (7) has a minimizer $\varphi_\Lambda(\mathbf{x}) = d_{\hat{G}} \in \Phi$. The existence of a $\mathbf{x}$-measurable pointwise minimizer $\varphi_\Lambda$ of $\hat{G}'\varphi(\mathbf{x})$ is guaranteed by LeCam (1956, Theorem 3.3.2), also by Brown and Purves (1973, Corollary 1); furthermore, it is seen to be pointwise component Bayes with respect to $\hat{G}$. *If* $n = 0$, $\hat{G}$ is the mean of $\Lambda$.

REMARK 3. If $S$ is bounded, then for each $\varphi \in \Phi$, $\mathbf{R}(G, \varphi)$ is a polynomial in $G \in \mathscr{G}$. This is immediate in view of (6), the fact that $P_G = \sum_0^m g_i P_i$, and the boundedness of $\varphi$ which ensures finite coefficients.

THEOREM 1. *Suppose that* $S$ *is a compact, convex subset of* $\mathbb{R}^{m+1}$. *Then for each* $n \geq 0$, *the class of Bayes rules,*

$$\mathscr{B} = \{\varphi \in \Phi \mid \varphi \textit{ minimizes } (7) \textit{ for some } \Lambda \in \mathscr{G}^*\}, \tag{9}$$

*is complete in* $\Phi$.

PROOF. If $n = 0$, $\mathscr{B}$ consists of the component Bayes points in $S$ and is complete by Ferguson (1967, Theorem 2.10.2). Let $n \geq 1$. By Lemma 1 of the Appendix and Ferguson (1967, Theorem 2.10.3), the class $\mathscr{B}_e$ of extended Bayes rules in $\Phi$ is essentially complete in $\Phi$. Let $\varphi \in \mathscr{B}_e$. Then by definition, for every integer $m > 0$, there exists a prior distribution $\Lambda_m$ such that

$$\mathbf{R}(\Lambda_m) \leq \mathbf{R}(\Lambda_m, \varphi) \leq \mathbf{R}(\Lambda_m) + \frac{1}{m}. \tag{10}$$

Since $\mathscr{G}$ is a compact subset of $\mathbb{R}^{m+1}$, the sequence $\{\Lambda_m\}$ is tight. By the Prohorov theorem (Billingsley, 1968, Theorem 5.1), there exists a prior $\Lambda \in \mathscr{G}^*$ and a subsequence $\{\Lambda_{m'}\}$ such that $\Lambda_{m'}$ converges weakly to $\Lambda$. As seen by Remark 3, all risk functions $\mathbf{R}(G, \varphi)$ are polynomials in $G$ and, hence, are continuous in $G$. Thus, $\mathbf{R}(\Lambda) \leq \mathbf{R}(\Lambda, \varphi) = \lim \mathbf{R}(\Lambda_{m'}, \varphi) = \lim \mathbf{R}(\Lambda_{m'})$ where the last equality follows from (10). Also, $\lim \mathbf{R}(\Lambda_{m'}) \leq \lim \mathbf{R}(\Lambda_{m'}, \varphi_\Lambda) = \mathbf{R}(\Lambda, \varphi_\Lambda) = \mathbf{R}(\Lambda)$, so that $\varphi$ is seen to be Bayes with respect to $\Lambda$, i.e., $\varphi \in \mathscr{B}$. Hence $\mathscr{B}_e = \mathscr{B}$, and the Bayes rules are seen to be essentially complete. For the class of Bayes rules, essential completeness is equivalent to completeness. □

Under more restrictive conditions, $\mathscr{B}$ is minimal complete. Tsao (1980a) shows that if the probability measures $P_0, P_1, \cdots, P_m$ are mutually absolutely continuous and the Bayes component rules are unique up to risk equivalence, then $\mathscr{B}$ is minimal complete in $\Phi$.

Boyer and Gilliland (1980) give results on the relationship of compound admissibility and empirical Bayes admissibility. One simple observation follows from the continuity of the risk functions $\mathbf{R}(G, \varphi)$; namely, that Bayes rules $\varphi_\Lambda$, where $\Lambda$ has support equal to $\mathscr{G}$, are necessarily admissible. Snijders (1977) has proved a complete class theorem for the empirical Bayes problem with component $\Theta = \{0, 1\} = A$, mutually absolutely continuous $P_0, P_1$, and finite observation space $\mathscr{X}$. The rules of his complete class have a monotonicity property. Balder, Gilliland and Van Houwelingen (1981) have generalized Theorem 1 by replacing the finite $\Theta$ requirement with compactness and continuity conditions on the component structure.

THEOREM 2. *Suppose that $S$ is a compact subset of $\mathbb{R}^{m+1}$ and $P_G$ is identified by $G$. Suppose that $\Lambda \in \mathscr{G}^*$ has support equal to $\mathscr{G}$. Then the Bayes empirical Bayes procedure defined for each stage $n$ to be $\varphi_\Lambda = d_{\hat{G}}$ of Remark 2 is a.o. i.e., $\mathbf{R}(G, \varphi_\Lambda) \to R(G)$ for all $G \in \mathscr{G}$.*

PROOF. Let $G \in \mathscr{G}$ and let $V$ be any $\mathscr{G}$-neighborhood of $G$. Lemmas 2 and 3 of the Appendix verify hypotheses (ii) and (iii) of Theorem 6.1 of Schwartz (1965) in regard to $V$. Hypothesis (i) is trivially satisfied in this application. Therefore, the conclusion of the Schwartz consistency theorem obtains, i.e.,

$$\Lambda(V \mid \mathbf{X}) \to 1 \quad \text{a.s.} \quad P_G^\infty. \tag{11}$$

Let $\varepsilon > 0$ be given and let $V_\varepsilon = \{F \in \mathscr{G} \mid \| F - G \| \le \varepsilon\}$ where $\| \cdot \|$ denotes the Euclidean norm in $\mathbb{R}^{m+1}$. Then by (8) and the Minkowski integral inequality (cf. Stein, 1970, page 271),

$$\| \hat{G} - G \| = \left\| \int (F - G)\Lambda(dF \mid \mathbf{X}) \right\| \le \int \| F - G \| \Lambda(dF \mid \mathbf{X}).$$

Partitioning the range of integration into $V_\varepsilon$ and its complement, we obtain $\| \hat{G} - G \| \le \varepsilon + \sqrt{2}(1 - \Lambda(V_\varepsilon \mid \mathbf{X}))$ so that (11) with $V = V_\varepsilon$, and the fact that $\varepsilon > 0$ is arbitrary, shows that $\| \hat{G} - G \| \to 0$ a.s. $P_G^\infty$. Oaten (1972, Lemma 1) implies that

$$0 \le R(G, d_{\hat{G}}) - R(G) \le (m + 1)^{1/2} M \| \hat{G} - G \|, \tag{12}$$

where $M$ is a bound for the components of the compact set $S$, from which $\varphi_\Lambda = d_{\hat{G}}$ is seen to be a.o. □

The a.o. property of Bayes empirical Bayes procedures can be deduced from the corresponding results for the compound decision problem. Robbins (1951) first proposed the use of Bayes procedures in his compound decision problem and conjectured their asymptotic optimality. However, the results for Bayes procedures and the $m > 1$ finite state component compound problem have generally been obtained only under regularity conditions on the distributions $P_\theta$ (cf. Vardeman, 1978), and only in the case $m = 1$ have completely general results with rates been published (cf. Gilliland and Hannan, 1974; Gilliland, Hannan and Huang, 1976).

The class of Bayes empirical Bayes rules $\mathscr{B}$ has been shown to be a complete class. A Bayes empirical Bayes rule takes the form $d_{\hat{G}}$, i.e., is component Bayes with respect to the conditional expectation $\hat{G}$. The more easily computed classical empirical Bayes rules of the form $d_{\hat{G}}$, where $\hat{G}$ is an unbiased estimator based on averaging a kernel across components, are seen to be inadmissible.

The classical empirical Bayes procedures are easily shown to be a.o. based on the consistency derived from a simple law of large numbers. As we have shown, empirical Bayes procedures which are Bayes versus diffuse $\Lambda$ have the same a.o. property and have admissible risk behavior for each $n$. Tsao (1980a, b) examines questions related to their computation and compares their small to moderate $n$ risk functions with those of selected classical empirical Bayes rules.

## APPENDIX

Let $\mu$ be a measure dominating $\{P_0, P_1, \cdots, P_m\}$ and $p_i = dP_i/d\mu$, $i = 0, 1, \cdots, m$, so that the polynomial risk function $\mathbf{R}(G, \varphi)$ (cf. (6)), can be written

$$\mathbf{R}(G, \varphi) = \int \textstyle\sum_{i=0}^m g_i \varphi^i(\mathbf{x}) \prod_{j=1}^n \sum_{k=0}^m g_k p_k(x_j) \mu^n(d\mathbf{x}).$$

Without loss of generality we take $\mu = \sum_0^m P_i$ so that a.e. $\mu^n$ is equivalent to a.s. $P_G^n$ for all $G \in \mathscr{G}$. We also extend $\Phi$ to include all measurable mappings from $\mathscr{X}^n$ into $\mathbb{R}^{m+1}$ such that $\varphi(\mathbf{X}) \in S$ a.e. $\mu^n$. For $\varphi \in \Phi$ we define $f(\varphi) \in \mathbb{R}^N$ to be the vector of coefficients for the polynomial $\mathbf{R}(G, \varphi)$ based on a specified ordering of terms $g_0^{i_0} g_1^{i_1} \cdots g_m^{i_m}$, $i_0 + i_1 + \cdots + i_m = n + 1$. Finally, without loss of generality, we take the observation space $\mathscr{X}$ to be the simplex in $\mathbb{R}^{m+1}$ through reduction by the sufficient statistic $(p_0(X), p_1(X), \cdots, p_m(X))$ and take $\mathscr{A}$ to be the Borel subsets of $\mathscr{X}$.

LEMMA 1. *Let $n \geq 1$. Suppose that the component risk set $S$ is a compact, convex subset of $\mathbb{R}^{m+1}$. Then $\Phi$ is compact with respect to the topology induced by $f^{-1}$ and the open sets of $\mathbb{R}^N$, and $\mathbf{R}(G, \varphi)$ is continuous with respect to this topology for each $G \in \mathscr{G}$.*

PROOF. It suffices to prove that the range $f[\Phi]$ is a compact subset of $\mathbb{R}^N$. For metric spaces, compactness follows from sequential compactness. Let $\{f(\varphi_j)\} \subset f[\Phi]$ with the $\varphi_j$ taking values in the set $S$. The weak compactness theorem (Lehmann (1959, page 354)) applied to components of $\{\varphi_j\}$ and the diagonalization process produce a subsequence $\{\varphi_{j'}\}$ with weak limit $\varphi \in \Phi$. It follows that $f(\varphi_{j'})$ converges to $f(\varphi)$ in $\mathbb{R}^N$. □

LEMMA 2. *Suppose that $P_G$ is identified by $G$. Let $G \in \mathscr{G}$ and let $V$ be any $\mathscr{G}$-neighborhood of $G$. There is a uniformly consistent test of $P = P_G$ versus $P \in \{P_F \mid F \in \mathscr{G} - V\}$ based on $X_1, X_2, \cdots$ iid $P$.*

PROOF. The alternatives are contained in the finite union $\cup B_i$ where $B_i = \{P_F \mid F \in \mathscr{G}, |f_i - g_i| \geq \varepsilon_i\}$ and the $\varepsilon_i > 0$ are sufficiently small, $i = 0, 1, \cdots, m$. By Kraft (1955, Theorem 7), it suffices to show that for each $i$, there exists a uniformly consistent test of $P = P_G$ versus $P \in B_i$. Suppose $h = (h^0, h^1, \cdots, h^m)$ is bounded and such that $\int h^i \, dP_j = \delta_{ij}$, the Kronecher delta function; for example, $h^0, h^1, \cdots, h^m$ can be taken as a basis dual to the densities $p_0, p_1, \cdots, p_m$ in $L_2(\mu)$ as observed, e.g., by Robbins (1964). The Hoeffding bound (1963, Theorem 2) shows that the test function $[|n^{-1} \sum_{j=1}^n h^i(x_j) - g_i| > \varepsilon_i/2]$ is uniformly consistent for $P = P_G$ versus $P \in B_i$. □

For $F, G \in \mathscr{G}$ the Kullback-Leibler information number between the densities $p_G$ and $p_F$ is $KL(G, F) = -\int \ln\{p_F(x)/p_G(x)\} P_G(dx)$.

LEMMA 3. *Suppose that $P_G$ is identified by $G$. Let $\Lambda \in \mathscr{G}^*$ have support equal to $\mathscr{G}$. Let $G \in \mathscr{G}$ and let $V$ be any $\mathscr{G}$-neighborhood of $G$. Then given $\varepsilon > 0$ there exists a subset $W \subset V$ such that $\Lambda(W) > 0$ and $KL(G, F) < \varepsilon$ for $F \in W$.*

PROOF. Let $O$ be an open set in $\mathbb{R}^{m+1}$ which contains $G$ and let $V = O \cap \mathscr{G}$. For $0 < \rho < 1$ consider the rectangle

$$U_\rho = \{F \in \mathbb{R}^{m+1} \mid \rho g_i \leq f_i \leq \rho g_i + 1 - \rho, i = 0, 1, \cdots, m\}$$

and note that $G \in U_\rho$. Let $\rho$ be sufficiently close to 1 so that $-\ln \rho < \varepsilon$ and $U = U_\rho \cap \mathscr{G} \subset V$. For such a $\rho$ let

$$W_\rho = \left\{F \in \mathbb{R}^{m+1} \mid \rho g_0 < f_0 < \rho g_0 + (1 - \rho), \right.$$
$$\left. \rho g_i < f_i < \rho g_i + \frac{1}{m}(1 - \rho), i = 1, 2, \cdots, m\right\}$$

and note that $W = W_\rho \cap \mathscr{G}$ is a nonempty open set in $\mathscr{G}$ with $W \subset U \subset V$. Since $\Lambda$ has support equal to $\mathscr{G}$, $\Lambda(W) > 0$. For $F \in W$, $p_F \geq \rho p_G$ and, therefore, $KL(G, F) \leq -\ln \rho < \varepsilon$. □

## REFERENCES

BALDER, E., GILLILAND, DENNIS C. and VAN HOUWELINGEN, J. C. (1981). Completeness of Bayes empirical Bayes decision rules. Unpublished manuscript.

BILLINGLEY, PATRICK (1968). *Convergence of Probability Measures*. Wiley, New York.

BOYER, JOHN E. JR. and GILLILAND, DENNIS C. (1980). Admissibility considerations in the finite state compound and empirical Bayes decision problems. *Statistica Neerlandica* **34** 151–159.

BROWN, L. D. and PURVES, R. (1973). Measurable selections of extrema. *Ann. Statist.* **1** 902–912.

DEELY, J. J. and LINDLEY, D. V. (1981). Bayes empirical Bayes. *J. Amer. Statist. Assoc.* **76** 833–841.

FERGUSON, THOMAS (1967). *Mathematical Statistics, A Decision Theoretic Approach*. Academic, New York and London.

GILLILAND, DENNIS C. and HANNAN, JAMES (1974). The finite state compound decision problem, equivariance and restricted risk components. RM-317, Statistics and Probability, Michigan State University.

GILLILAND, DENNIS C., HANNAN, JAMES and HUANG, J. S. (1976). Asymptotic solutions to the two state component compound decision problem, Bayes versus diffuse priors on proportions. *Ann. Statist.* **4** 1101–1112.

HOEFFDING, WASSILY (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

KRAFT, CHARLES (1955). Some conditions for consistency and uniform consistency of statistical procedures. *Univ. of Calif. Publications in Statist.*

LECAM, LUCIEN (1956). Lecture Notes., Dept. Statistics, University of California at Berkeley.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

LINDLEY, D. V. (1961). The use of prior probability distributions in statistical inference and decision. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 453–468, University of California Press.

LINDLEY, D. V. (1971). Bayesian statistics, a review. *Regional Conference Series in Applied Mathematics No. 2*, SIAM, Philadelphia.

MEEDEN, GLEN (1972). Some admissible empirical Bayes procedures. *Ann. Math. Statist.* **43** 96–101.

OATEN, ALLAN (1972). Approximation to Bayes risk in compound decision problems. *Ann. Math. Statist.* **43** 1164–1184.

ROBBINS, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. Math. Statist. Probab.* 131–148. University of California Press.

ROBBINS, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 157–163. University of California Press.

ROBBINS, HERBERT (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.

SCHWARTZ, L. (1965). On Bayes' procedures. *Z. Wahrsch. verw. Gebiete* **4** 10–26.

SNIJDERS, TOM (1977). Complete class theorems for the simplest empirical Bayes decision problems. *Ann. Statist.* **5** 164–171.

STEIN, ELIAS M. (1970). *Singular Integrals and Differentiability Properties of Functions*. Princeton University Press, Princeton.

TSAO, HOW JAN (1980a). On the risk performance of Bayes empirical Bayes procedures in the finite state component case. RM-407, Statistics and Probability, Michigan State University.

TSAO, HOW JAN (1980b). On the risk performance of Bayes empirical Bayes procedures for classification between $N(-1, 1)$ and $N(1, 1)$. *Statistica Neerlandica* **34** 197–208.

VARDEMAN, STEPHEN B. (1978). Admissible solutions of finite state sequence compound decision problems. *Ann. Statist.* **6** 673–679.

DENNIS C. GILLILAND
DEPARTMENT OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824

JOHN E. BOYER, JR.
DEPARTMENT OF STATISTICS
KANSAS STATE UNIVERSITY
MANHATTAN, KANSAS 66506

HOW JAN TSAO
OAKRIDGE NATIONAL LABORATORY
M/S H-32 4500N
OAKRIDGE, TENNESSEE 37830