The Annals of Statistics
1991, Vol. 19, No. 1, 338–353

ON THE CONSISTENCY OF POSTERIOR MIXTURES AND ITS APPLICATIONS

BY SOMNATH DATTA

University of Georgia

Consider i.i.d. pairs (θ_i, X_i), $i \geq 1$, where θ_1 has an unknown prior distribution ω and given θ_1, X_1 has distribution P_{θ_1}. This setup arises naturally in the empirical Bayes problems. We put a probability (a hyperprior) on the space of all possible ω and consider the posterior mean $\hat{\omega}$ of ω. We show that, under reasonable conditions, $P_{\hat{\omega}} = \int P_\theta \, d\hat{\omega}$ is consistent in L_1. Under a identifiability assumption, this result implies that $\hat{\omega}$ is consistent in probability. As another application of the L_1 consistency, we consider a general empirical Bayes problem with compact state space. We prove that the Bayes empirical Bayes rules are asymptotically optimal.

1. Introduction. In an empirical Bayes problem [Robbins (1951, 1956)] one observes independent repetitions of a Bayes problem (the so-called component problem) where the Bayes prior is unknown. Formally, let (X_1, θ_1), $(X_2, \theta_2), \ldots$ be a sequence of i.i.d. pairs where θ_1 is distributed according to some unknown prior distribution ω and given θ_1, X_1 has distribution P_{θ_1}. The goal of an empirical Bayes problem is to estimate the component Bayes rule from the past X's and then use it with the present X to take a decision about the present θ.

The empirical Bayes problem is often closely related to the following estimation problem considered by Blum and Susarla (1977) and many others: Let $X_1, X_2, \ldots$ be i.i.d. with distribution $P_\omega = \int P_\theta \, d\omega$, where the mixing distribution ω is unknown. The problem is to estimate ω from the X's.

Note that, in the empirical Bayes problem, X's are indeed i.i.d. with distribution P_ω. Thus, once ω can be estimated from the past data, an empirical Bayes rule can be constructed by playing Bayes versus the estimated ω.

A well-known approach to solve the above problems is to put a hyperprior on the space of possible ω and to consider the posterior mean $\hat{\omega}$ of ω given the X's. A rule which plays Bayes versus $\hat{\omega}$ turns out to be a Bayes empirical Bayes rule. Balder, Gilliland and van Houwelingen (1983) proved some admissibility and complete class results for the Bayes empirical Bayes rules under the compactness of the state space.

Rolph (1968) and Meeden (1972b) established that $\hat{\omega}$ is consistent for ω as the number of X's approaches infinity. They both considered cases where the

Received September 1988; revised December 1989.
AMS 1980 *subject classifications*. Primary 62C10; secondary 62C12.
Key words and phrases. Posterior, consistency, mixing distribution, empirical Bayes, asymptotic optimality.

X's are nonnegative integer valued and used very special hyperpriors suitable for the discrete case only. In this paper, we consider X's which can take values in any measurable space. We prove that, under reasonable conditions, $\hat{\omega}$ is consistent for ω in probability whenever the selected hyperprior is sufficiently diffuse.

In the case of only finitely many P_θ's, Gilliland, Boyer and Tsao (1982) proved that the Bayes empirical Bayes rules are asymptotically optimal (see Section 4); i.e., possess good asymptotic risk behaviors, if the hyperprior is sufficiently diffuse. The only asymptotic optimality results known for the Bayes empirical Bayes rules in an infinite state space case were due to Meeden (1972a) where the component problems were (i) squared error loss estimation of a Geometric parameter and (ii) linear loss estimation of a Poisson mean. To reduce the complexity involved in establishing the asymptotic optimality of Bayes empirical Bayes rules for all possible priors, several authors considered a parametric subclass (mostly one-dimensional) of possible ω and put a hyperprior on that. Consequently, the asymptotic optimality holds for those priors only. In this paper, we prove asymptotic optimality of Bayes empirical Bayes rules for a general component problem with compact state space under reasonable conditions on the component distributions and the risk function. Similarly, a Bayes empirical Bayes rule versus a hyperprior on a compact subspace of all possible ω is asymptotically optimal for all ω in that subspace.

The above consistency and asymptotic optimality results are presented in Section 4 and treated as applications of the following key result: Let $\mathbf{P}_\omega$ denote the joint marginal distribution of the X's and $\mathbf{P}_\theta$ denote the joint conditional distribution of the X's given the θ's. For each n, let G_n stand for the empirical distribution of $\theta_1, \theta_2, \dots, \theta_n$. We prove that $P_{\hat{\omega}}$ is L_1 consistent for P_ω and conditionally L_1 consistent for P_{G_n}, uniformly in ω and $\underline{\theta}$, respectively. These results are stated in Section 3 and proved in Section 6. Section 2 gives a formal definition of $\hat{\omega}$ and interprets it in Bayesian terms. Some examples of families of distributions satisfying the assumptions of the theorems are given in Section 5.

2. Various Bayes models. Let $\{P_\theta: \theta \in \Theta\}$ be a family of probability measures on some common measurable space $\mathscr{X}$ dominated by some σ-finite measure μ. We assume that Θ is a metric space and consider the Borel σ-field on it. Suppose we have $\{p_\theta: \theta \in \Theta\}$ on $\mathscr{X}$ such that (a) $p_\theta(x)$ is jointly measurable in θ and x and (b) $\forall\ \theta$, p_θ is a density of P_θ wrt μ.

Let $\Omega = \{\omega: \omega \text{ is a probability on } \Theta\}$ with the Borel σ-field corresponding to the topology of weak convergence. For $\omega \in \Omega$, let P_ω stand for the mixture $\int P_\theta \, d\omega$ and p_ω denote its μ-density $\int p_\theta \, d\omega$.

Let Λ be a probability on Ω and n be a positive integer. For probabilities $P_1, \dots, P_n$, let $\times_{\alpha=1}^n P_\alpha$ denote their measure theoretic product. Consider the following Bayes model on $\Omega \times \Theta^n \times \mathscr{X}^n$:

(i) Bayes model: ω is distributed as Λ and given ω, $\underline{\theta}$ is distributed as $\omega^n = \times_{\alpha=1}^n \omega$ and given $\underline{\theta}$ and ω, $\underline{X}$ is distributed as $\mathbf{P}_{\underline{\theta}} = \times_{\alpha=1}^n P_{\theta_\alpha}$.

The above model gives rise to the following two marginal models:

(ii) Bayes compound model: $\underline{\theta} = \langle \theta_1, \dots, \theta_n \rangle$ is distributed as $\bar{\omega}_\Lambda^n$ and given $\underline{\theta}$, $\underline{X} = \langle X_1, \dots, X_n \rangle$ is distributed as $\mathbf{P}_{\underline{\theta}}$, where $\bar{\omega}_\Lambda^n = \Lambda \circ \omega^n$, i.e.,

$$\bar{\omega}_\Lambda^n(B_1 \times \cdots \times B_n) = \int \prod_{\alpha=1}^{n} \omega(B_\alpha)\, d\Lambda, \quad \text{for } B_1, \dots, B_n \text{ Borels of } \Theta.$$

Let $\mathbf{E}_{\underline{\theta}}$ stand for the expectation under $\mathbf{P}_{\underline{\theta}}$.

(iii) Bayes empirical Bayes model: ω is distributed as Λ and given ω, $\underline{X}$ is distributed as $\mathbf{P}_\omega = \times_{\alpha=1}^n P_\omega$.

Let $\mathbf{E}_\omega$ stand for the expectation under $\mathbf{P}_\omega$.

Let $\hat{\Lambda}$ be the posterior distribution of ω given $\langle X_1, \dots, X_n \rangle = \langle x_1, \dots, x_n \rangle$. Then $\hat{\Lambda}$ is the probability measure on Ω with density proportional to $\prod_{\alpha=1}^n p_\omega(x_\alpha)$. Let $\hat{\omega} = \hat{\Lambda} \circ \omega$.

The following interpretations of $\hat{\omega}$ are easy to prove. In the following, all conditional distributions are regular.

Proposition 1. (a) *Under model* (i) *or* (ii), *with n replaced by $n+1$, $\hat{\omega}$ is the posterior distribution of θ_{n+1} given $\langle X_1, \dots, X_n \rangle = \langle x_1, \dots, x_n \rangle$.*

(b) *Under model* (i) *or* (iii), *$\hat{\omega}$ is the posterior mean of ω given $\underline{X} = \underline{x}$ in the sense that $\forall$ Borel $B \subset \Theta$, $\hat{\omega}(B)$ is the posterior mean of $\omega(B)$ given $\underline{X} = \underline{x}$.*

Proof. Proof of (b) is same as if ω were a real parameter which is standard [e.g., Berger (1985)] in decision theory. For part (a) it is sufficient to show that

$$\begin{aligned} &\operatorname{Prob}(\theta_{n+1} \in B, \langle X_1, \dots, X_n \rangle \in A_1 \times \cdots \times A_n) \\ &\quad = \int_{A_1 \times \cdots \times A_n} \hat{\omega}(B) \left(\int \prod_1^n p_{\theta_\alpha}\, d\bar{\omega}_\Lambda^n \right) d\mu^n. \end{aligned}$$

Under model (i),

$$\begin{aligned} LHS &= \int_{(\Theta^n \times B) \times (A_1 \times \cdots \times A_n)} \prod_1^n p_{\theta_\alpha}\, d(\mu^n \times \bar{\omega}_\Lambda^{n+1}) \\ &= \int_{A_1 \times \cdots \times A_n} \left(\int_{\Theta^n \times B} \prod_1^n p_{\theta_\alpha}\, d(\Lambda \circ \omega^{n+1}) \right) d\mu^n, \end{aligned}$$

by the Fubini theorem on $\mathscr{X}^n \times \Theta^{n+1}$.

By the Fubini theorem on $\Omega \times \Theta^{n+1}$, the inner integral equals

$$\int \omega(B) \prod_1^n p_\omega\, d\Lambda = \hat{\omega}(B) \left(\int \prod_1^n p_{\theta_\alpha}\, d\bar{\omega}_\Lambda^n \right),$$

by the definition of $\hat{\omega}$ and another application of the Fubini theorem finishing the proof. □

Our main results are that under reasonable conditions on $\{p_\theta\}$ and Λ, $P_{\hat{\omega}}$ is uniformly L_1 consistent for P_{G_n} under $\mathbf{P}_\theta$ and is L_1 consistent for P_ω under $\mathbf{P}_\omega$, where G_n is the empirical distribution of $\theta_1, \dots, \theta_n$. These are presented in the next section.

3. Consistency of posterior mixture. From now on assume Θ to be separable. Then by Theorem II.6.2 of Parthasarathy (1967), Ω with the weak convergence topology can be metrized as a separable metric space.

For a measure m on a separable metric space $\mathscr{S}$, the support of m is defined to be the set

$$S_m = \bigcap \{F\colon F \text{ is closed and } m(F^c) = 0\}.$$

By expressing S_m^c as a countable union of F^c sets, it follows that $m(S_m^c) = 0$. Also note that $s \in S_m$, iff for any open set O containing s, we have, $m(O) > 0$.

For ω, ω' in Ω, let

$$\|P_\omega - P_{\omega'}\| = \int |p_\omega - p_{\omega'}|\, d\mu$$

denote the (L_1) distance between P_ω and $P_{\omega'}$. Note that this definition does not depend on the choice of μ and the μ-densities.

Consider the following assumption on the family of densities $\{p_\theta\colon \theta \in \Theta\}$. Let Λ be a probability on Ω. Let $x_+ = x \vee 0$, for $x \in \mathbb{R}$. We interpret $\log 0$ as $-\infty$.

A1. $p_\theta(x)$ is continuous in θ for each x.

A1′. $p_\omega(x)$ is continuous in ω on S_Λ for each x.

A2. With $h_{\theta'}^* = \bigvee_\theta |\log(p_\theta/p_{\theta'})|$,

$$\bigvee_\theta \int (h_\theta^* - M)_+ p_\theta \, d\mu \to 0 \quad \text{as } M \to \infty.$$

A2′. With $h_{\omega'}^* = \bigvee_{\omega \in S_\Lambda} |\log(p_\omega/p_{\omega'})|$,

$$\bigvee_{\omega \in S_\Lambda} \int (h_{\omega'}^* - M)_+ p_\omega \, d\mu \to 0 \quad \text{as } M \to \infty.$$

A2″. With h_ω^* as in A2′, $\int h_\omega^* p_\omega \, d\mu < \infty$, $\forall\, \omega \in S_\Lambda$.

We now state our main theorems. The proofs of the theorems are interesting but technical. They will be given in Section 6.

Let G_n be the empirical distribution of $\theta_1, \dots, \theta_n$.

THEOREM 3.1. *Suppose* Θ *is compact and* A1 *and* A2 *are satisfied. If* $S_\Lambda = \Omega$, *then*

$$\mathbf{E}_{\underline{\theta}} \| P_{\hat{\omega}} - P_{G_n} \| \to 0, \quad \text{uniformly in } \underline{\theta}, \text{ as } n \to \infty. \tag{3.1}$$

COROLLARY 3.1. *Under the conditions of Theorem* 3.1,

$$\mathbf{E}_{\omega} \| P_{\hat{\omega}} - P_{\omega} \| \to 0 \quad \text{as } n \to \infty, \forall\, \omega \in \Omega. \tag{3.2}$$

PROOF OF COROLLARY 3.1. For each x, $p_\theta(x)$ is continuous in θ by A1 and bounded as Θ is compact and hence $p_\omega(x)$ is continuous by the definition of weak convergence. This implies that $\omega \rightsquigarrow P_\omega$ is continuous in $\| \cdot \|$ by the Scheffé theorem. Let $\mathbf{E}$ be the joint expectation under which, for all n, $\underline{\theta} \sim \omega^n$ and given $\underline{\theta}$, $\underline{X} \sim \mathbf{P}_{\underline{\theta}}$. Then $G_n \to \omega$ a.s. ($\mathbf{E}$) by the Glivenko–Cantelli theorem implying $\|P_{G_n} - P_\omega\| \to 0$ a.s. ($\mathbf{E}$) by the continuity just noted. Hence by D.C.T., $\mathbf{E}\|P_{G_n} - P_\omega\| \to 0$. The conclusion now follows by taking ω^n expectation of (3.1), the triangle inequality and noting the fact that $\mathbf{E}\|P_{\hat{\omega}} - P_\omega\| = \mathbf{E}_\omega\|P_{\hat{\omega}} - P_\omega\|$. □

(3.1) can be viewed as a robust version of (3.2). In the next theorem we generalize the consistency result in the empirical Bayes case to a large extent. In particular, (3.2) holds under weaker conditions on the distributions.

THEOREM 3.2. (a) *Suppose,* S_Λ *is a compact subset of* Ω *and* A1′ *and* A2″ *are satisfied. Then*

$$\mathbf{E}_{\omega} \| P_{\hat{\omega}} - P_{\omega} \| \to 0 \quad \text{as } n \to \infty, \forall\, \omega \in S_\Lambda. \tag{3.3}$$

(b) *If, moreover,* A2′ *is satisfied, then the above convergence is uniform in* $\omega \in S_\Lambda$.

REMARK 3.1. If S_Λ is compact and equals Ω, then (3.3) is the same as (3.2), A1 and A1′ are equivalent and A2 and A2′ are equivalent. But A2″ is still weaker than A2.

REMARK 3.2. Assumption A2 forces the possible P_θ's to be pairwise mutually absolutely continuous. Similarly, A2″ forces two mixtures P_ω and $P_{\omega'}$ to be mutually absolutely continuous whenever ω and ω' are in the support of Λ. In the finite Θ case, it is possible to obtain the conclusion of Theorem 3.2, when $S_\Lambda = \Omega$, without these requirements from the proof of Theorem 2 of Gilliland, Boyer and Tsao (1982). Therefore it may be possible relax the assumptions of the above theorems.

4. Applications. In this section we discuss some applications of the Section 3 results in the problem of *estimating a mixing distribution* (see Section 4.1) and the *empirical Bayes* problem (see Section 4.2). For an application of Theorem 3.1 to a compound decision problem, see Datta (1988).

4.1. *Estimation of a mixing distribution.* Suppose that $\{P_\theta: \theta \in \Theta\}$ is as in Section 2. Let $X_1, \dots, X_n$ be i.i.d. sample from $P_\omega = \int P_\theta \, d\omega$ for some unknown mixing distribution $\omega \in \Omega$. The problem is to estimate ω on the basis of $X_1, \dots, X_n$. In order that this estimation problem make sense, we assume the usual identifiability condition,

$$\text{(I)} \qquad \omega \rightsquigarrow P_\omega \text{ is one-to-one on } \Omega,$$

throughout this subsection.

The following are easy consequences of the Section 3 results and the condition (I).

THEOREM 4.1. (a) *Let Θ be compact, $S_\Lambda = \Omega$ and* A1′ *and* A2″ *be satisfied. Then $\forall\, \omega \in \Omega$, $\hat{\omega} \to \omega$ in probability (P_ω).*

(b) *If, moreover,* A2′ *is satisfied, then the above convergence is uniform in ω.*

PROOF. Let d be a metric metrizing the topology of weak convergence on Ω.

The mapping $\omega \rightsquigarrow P_\omega$ is one-to-one by (I), continuous by A1′ and the Scheffé theorem, and onto its range $\mathscr{P} = \{P_\omega: \omega \in \Omega\}$. Hence, because Ω is compact, it is a homeomorphism and $\mathscr{P}$ is compact. So $P_\omega \rightsquigarrow \omega$ is uniformly continuous on $\mathscr{P}$.

So given $\varepsilon > 0$, $\exists\, \delta = \delta(\varepsilon) > 0$ such that $d(\omega, \omega') \le \varepsilon$ whenever $\|P_\omega - P_{\omega'}\| \le \delta$. Thus,

$$\mathbf{P}_\omega\{d(\hat{\omega}, \omega) > \varepsilon\} \le \mathbf{P}_\omega\{\|P_{\hat{\omega}} - P_\omega\| > \delta\} \le \delta^{-1}\mathbf{E}_\omega\|P_{\hat{\omega}} - P_\omega\| \to 0$$

as $n \to \infty$, by Theorem 3.2(a).

The last convergence is uniform in ω, if A2′ is satisfied, by Theorem 3.2(b). □

REMARK 4.1. A careful inspection of the proof shows that if we replace the assumptions of the compactness of Θ and $S_\Lambda = \Omega$ by the compactness of S_Λ alone, we still get the consistency for ω in S_Λ provided we have, in addition, $\mathbf{P}_\omega(\hat{\omega} \notin S_\Lambda) \to 0$ as $n \to \infty$, for all $\omega \in S_\Lambda$. Part (b) holds if the last convergence is uniform in $\omega \in S_\Lambda$.

REMARK 4.2. By standard arguments it follows that if d is a bounded metric metrizing the weak convergence topology on Ω then,

$$\mathbf{E}_\omega d(\hat{\omega}, \omega) \to 0 \quad \text{as } n \to \infty, \forall\, \omega,$$

if the conclusion of Theorem 4.1(a) holds, and the convergence is uniform on Ω if the conclusion of Theorem 4.1(b) holds.

REMARK 4.3. Rolph (1968) and Meeden (1972b) established the strong (a.s. $\mathbf{P}_\omega$) consistency of $\hat{\omega}$ in cases where $\mathscr{X}$ is the set of nonnegative integers. Rolph considered the case when $p_\theta(x)$ is continuous in θ and $\Theta = [0, 1]$.

Meeden extended the consistency result to the case where $p_\theta(x)$ satisfies continuity in θ plus some other smoothness conditions and $\Theta = [0, \infty)$. Both the authors considered very special priors with full support. On the other hand, our weak consistency results apply to general $\mathscr{X}$ and compact Θ.

Computation of $\hat{\omega}$. Meeden (1972b) said that it was not possible to calculate $\hat{\omega}$ in practice from the data since its expression involved infinitely many integrals each over an infinite dimensional space. Below we obtain a form of $\hat{\omega}$, which holds quite generally, involving only finitely many integrals over Θ. So if Θ is a subset of the real line, say, then it is possible to evaluate this expression, at least numerically.

For a set A, let $[A]$ denote its indicator function. From the definition of $\hat{\omega}$ and some Fubini arguments, it follows that

$$\hat{\omega}(B) = \frac{\int [\theta_{n+1} \in B] \prod_{\alpha=1}^{n} p_{\theta_\alpha}(X_\alpha)\, d(\bar{\omega}_\Lambda^{n+1})}{\int \prod_{\alpha=1}^{n} p_{\theta_\alpha}(X_\alpha)\, d(\bar{\omega}_\Lambda^{n})}, \tag{4.1}$$

for any Borel B of Θ.

Let $\omega^{\underline{\theta}_n}$ denote the posterior mean of ω given $\underline{\theta}_n = \langle \theta_1, \dots, \theta_n \rangle$ under model (i) of Section 2. Then, by repeated conditioning, it follows that

$$d\bar{\omega}_\Lambda^n(\underline{\theta}_n) = \prod_{\alpha=1}^{n} d\omega^{\underline{\theta}_{\alpha-1}}(\theta_\alpha), \qquad \forall\, n \geq 1,$$

where $\omega^{\underline{\theta}_0} = \int \omega \, d\Lambda$.
Using this in (4.1), we get

$$\hat{\omega}(B) = \frac{\int \cdots \int [\theta_{n+1} \in B] \prod_{\alpha=1}^{n} p_{\theta_\alpha}(X_\alpha) \prod_{\alpha=1}^{n+1} d\omega^{\underline{\theta}_{\alpha-1}}(\theta_\alpha)}{\int \cdots \int \prod_{\alpha=1}^{n} p_{\theta_\alpha}(X_\alpha) \prod_{\alpha=1}^{n} d\omega^{\underline{\theta}_{\alpha-1}}(\theta_\alpha)}. \tag{4.2}$$

EXAMPLE 4.1. Let $\Theta = [c, d] \subset (-\infty, \infty)$. Let γ be a finite measure on $[c, d]$ with support$(\gamma) = [c, d]$. Then for Λ = Dirichlet prior with parameter γ [see Ferguson (1973)], we have support$(\Lambda) = \Omega$ and

$$\omega^{\underline{\theta}_{i-1}} = \left(\gamma + \sum_{\alpha=1}^{i-1} \delta_{\theta_\alpha}\right) \bigg/ (\gamma[c, d] + i - 1), \qquad i \geq 1,$$

where δ_θ stands for the probability measure degenerate at θ.

4.2. *Empirical Bayes problem*. As our second application we consider the empirical Bayes problem of Robbins (1951, 1955). In this formulation we have $\{P_\theta: \theta \in \Theta\}$, i.i.d. pairs $(\theta_1, X_1), (\theta_2, X_2), \dots,$ where θ_1 is distributed as ω and given θ_1, X_1 is distributed as P_{θ_1}. $\{P_\theta: \theta \in \Theta\}$ is known but $\omega \in \Omega$ is unknown to the statistician.

At stage n, a decision $t_n = t_n(X_1, \dots, X_n)$ about θ_n has to be taken with loss $L(t_n, \theta_n)$ and risk $R_n(t_n, \omega) = \int\int L(t_n, \theta_n)\, d\mathbf{P}_{\underline{\theta}}\, d\omega^n$. The loss function L is independent of n and satisfies appropriate measurability condition. $\langle t_n: n \geq 1 \rangle$ is called an empirical Bayes rule. The standard for evaluating the empirical

Bayes rules is taken to be the component Bayes envelope

$$R(\omega) = \bigwedge_t R(t, \omega),$$

where $R(t, \omega) = \int\int L(t, \omega)\, dP_\theta\, d\omega$ and the infimum is taken over all component rule t. An empirical Bayes rule $\langle t_n \rangle$ is called asymptotically optimal (a.o. hereafter) if

$$R_n(t_n, \omega) \to R(\omega) \quad \text{as } n \to \infty, \forall\, \omega \in \Omega.$$

$\langle t_n \rangle$ is called a.o. relative to $\Omega_0 \subset \Omega$ if the above convergence takes place for all $\omega \in \Omega_0$.

We assume that for all $\omega \in \Omega$, there exists a component Bayes rule τ_ω versus ω, i.e., there exists τ_ω such that $R(\omega) = R(\tau_\omega, \omega)$.

The following theorem gives a general scheme for obtaining a.o. empirical Bayes rules. Let $\tilde{\omega} = \tilde{\omega}(X_1, \dots, X_n)$ be an estimator of ω.

THEOREM 4.2. *Suppose that* Θ *is compact and*

(i) $R(\tau_{\omega'}, \omega)$ *is continuous in* ω' *for each* ω,
(ii) $\omega \rightsquigarrow P_\omega$ *is one-to-one and continuous,*
(iii) $P_{\tilde{\omega}} \to P_\omega$ *in probability* $(\mathbf{P}_\omega)$, $\forall\, \omega$.

Then the empirical Bayes rule $\hat{t}_{n+1}(\underline{X}_{n+1}) = \tau_{\tilde{\omega}}(X_{n+1})$, $n \geq 1$ *is a.o.*

PROOF. Compactness of Θ and (i) imply that $R(\tau_{\omega'}, \omega)$ is bounded as a function of ω'. From (ii) and (iii), it follows, as in the proof of Theorem 4.1, that $\tilde{\omega} \to \omega$ in probability $(\mathbf{P}_\omega)$ $\forall\, \omega$. Hence by (i), $R(\tau_{\tilde{\omega}}, \omega) \to R(\omega)$ in probability $(\mathbf{P}_\omega)$ and hence in $L_1(\mathbf{E}_\omega)$ because of its boundedness. So

$$|R_{n+1}(\hat{t}_{n+1}, \omega) - R(\omega)| \leq \mathbf{E}_\omega |R(\tau_{\tilde{\omega}}, \omega) - R(\omega)| \to 0, \qquad \forall\, \omega. \qquad \square$$

REMARK 4.4. The following more general version of the above theorem can be proved in the same way. Let $\Omega_0 \subset \Omega$ be compact and (ii) and (iii) hold for ω in Ω_0 only. Moreover, let $R(\tau_\omega, \omega')$ be continuous in ω on Ω_0 for all $\omega' \in \Omega_0$ and $\mathbf{P}_\omega(\tilde{\omega} \notin \Omega_0) \to 0$ as $n \to \infty$, for all $\omega \in \Omega_0$. Then $\hat{t}_{n+1}(\underline{X}_{n+1}) = \tau_{\tilde{\omega}}(X_{n+1})$ is a.o. relative to Ω_0.

It can be shown that the empirical Bayes rule $\hat{t}_{n+1}(\underline{X}_{n+1}) = \tau_{\hat{\omega}}(X_{n+1})$ is the Bayes empirical Bayes rule versus Λ. The above theorem and the remark imply the following asymptotic optimality results for the Bayes empirical Bayes rules.

COROLLARY 4.1. *Let* (i) *of Theorem* 4.2, A1′, A2″ *and the condition* (I) *hold. If* $S_\Lambda = \Omega$ *is compact, then the Bayes empirical Bayes rule* $\hat{t}_{n+1}(\underline{X}_{n+1}) = \tau_{\hat{\omega}}(X_{n+1})$ *is a.o.*

PROOF. That (ii) of Theorem 4.2 holds is given by (I) and A1′ with the Scheffé theorem. (iii) with $\tilde{\omega} = \hat{\omega}$ is guaranteed by Theorem 3.2(a). The conclusion then follows by Theorem 4.2 with $\tilde{\omega} = \hat{\omega}$. □

COROLLARY 4.2. *Let S_Λ be compact, $R(\tau_\omega, \omega')$ be continuous in ω on S_Λ, for each $\omega' \in S_\Lambda$,* A1′, A2″ *and the condition* (I) *hold. If $\mathbf{P}_\omega(\hat{\omega} \notin S_\Lambda) \to 0$ as $n \to \infty$, for all $\omega \in S_\Lambda$, then the Bayes empirical Bayes rule $\hat{t}_{n+1}(\underline{X}_{n+1}) = \tau_{\hat{\omega}}(X_{n+1})$ is a.o. relative to S_Λ.*

PROOF. The conditions of Remark 4.4 are satisfied by $\Omega_0 = S_\Lambda$ and $\tilde{\omega} = \hat{\omega}$. Hence the conclusion follows by Remark 4.4. □

REMARK 4.5. Often Bayes empirical Bayes rules are admissible. For example, if the component problem is the estimation of $\phi(\theta)$ under squared error loss $L(t, \theta) = (t - \phi(\theta))^2$, ϕ being a measurable function on Θ, and $P_\theta \lll\ggg \mu$ $\forall$ θ and for some μ, then any Bayes empirical Bayes estimator is unique up to risk equivalence and hence admissible. So in this case $\langle \hat{t}_n \rangle$ is an admissible a.o. empirical Bayes rule.

5. Examples.

We list a few families of distributions which satisfy the assumptions (see Section 3) of our theorems. In these examples, Θ is compact and hence A1 and A1′ are equivalent.

EXAMPLE 5.1 (Finite Θ). Let $\Theta = \{1, 2, \dots, m\}$, for some positive integer m. In this case, A1 is trivially satisfied. A2 holds iff P_i and P_j are mutually absolutely continuous for all $1 \le i, j \le m$. Thus our Corollary 4.1 in the finite Θ case is weaker than Theorem 2 of Gilliland, Boyer and Tsao (1982) which arrives at the same conclusion without the above requirement. In this case, condition (I) amounts to the linear independence of $P_1, P_2, \dots, P_m$.

EXAMPLE 5.2 (Location family on $\mathbb{R}$). Let f be a nonnegative continuous function on $\mathbb{R}$ satisfying the following conditions.

(i) There exists $a \le b \in \mathbb{R}$, such that f is increasing on $(-\infty, a)$ and decreasing on (b, ∞),

(ii) $\int f(x)\, d\lambda(x) = 1$,

(iii) $\int |\log f(x)| f(x + c)\, d\lambda(x) < \infty$, $\forall$ $c \in \mathbb{R}$,

where λ denotes the Lebesgue measure on $\mathbb{R}$. Let $-\infty < A < B < \infty$. It is easy to check that the family of Lebesgue densities

$$p_\theta(x) = f(x - \theta), \qquad x \in \mathbb{R}, \theta \in \Theta = [A, B]$$

satisfies A1 and A2.

In this case, for any $\omega \in \Omega$, the characteristic function of P_ω is the product of the characteristic functions of f and ω. Hence (I) holds if f has a nonvanishing characteristic function.

A special case of this example is the Cauchy location family with compact location parameter.

EXAMPLE 5.3 (Exponential family). Let $\Theta = \Pi_{i=1}^{k}[a_i, b_i] \subset \mathbb{R}^k$, where Π denotes the set theoretic product and $\forall\ \theta \in \Theta$, $p_\theta(x) = c(\theta)e^{\theta' T}$, for some measurable function $T = (T_1, \dots, T_k)$: $\mathscr{X} \to \mathbb{R}^k$ and $c(\theta) = 1/\int e^{\theta' T}\, d\mu$. Further assume that

$$(5.1)\qquad \prod_{i=1}^{k}\{a_i, b_i\} \subset \text{interior of } \left\{\theta\colon \int e^{\theta' T}\, d\mu < \infty\right\}.$$

Then c is positive and continuous by (5.1) and Lemma 3.5.8 of Fabian and Hannan (1985). Hence, A1 holds and $c^* = \bigvee_{\theta \in \Theta} c(\theta) < \infty$, $|\log c|^* = \bigvee_{\theta \in \Theta} |\log c(\theta)| < \infty$. Clearly,

$$p_\theta \le c^* \sum_{d \in \Pi\{a_i, b_i\}} e^{d' T}$$

and

$$|\log p_\theta| \le |\log c|^* + \sum_{i=1}^{k} b_i |T_i|, \quad \text{for all } \theta.$$

The above inequalities immediately show that A2 hold because

$$(5.2)\qquad \int |T_i| e^{d' T}\, d\mu < \infty, \qquad \forall\ i = 1, \dots, k \text{ and } d \in \prod_{i=1}^{k}\{a_i, b_i\},$$

by Lemma 3.5.8 of Fabian and Hannan (1985).

If $k = 1$, then A1 follows by the monotone convergence theorem and (5.2) is sufficient to guarantee A2 [and (5.1) is not required].

In this example, no easy criterion for condition (I) can be given in general. However, if $k = 1$ and $T(x) = x$, then the following sufficient condition can be stated. If there exist $x_0 < x_1, x_2, \dots \in S_\mu$ such that $\sum_{i \ge 1}(x_i - x_0)^{-1} = \infty$, then (I) holds. A proof of this statement can be given using Müntz theorem [see page 384, Dunford and Schwartz (1957)].

Several examples of full support Λ in the case $\Theta = [c, d] \subset \mathbb{R}$ have been given in Datta (1988).

6. Proofs. We first introduce a few lemmas which will be used to prove Theorems 3.1 and 3.2.

For any $\omega, \omega' \in \Omega$, define

$$(6.1)\qquad \Delta_{\omega'}(\omega) = \int \log(p_{\omega'}/p_\omega)\, dP_{\omega'}.$$

Let

$$\mathscr{V}_{\omega'} = \bigvee_{\omega} \left| n^{-1} \sum_{1}^{n} \log\left(\frac{p_\omega}{p_{\omega'}}(X_\alpha)\right) + \Delta_{\omega'}(\omega) \right|, \tag{6.2}$$

$$\mathscr{V}_{\Lambda,\omega'} = \bigvee_{\omega \in S_\Lambda} \left| n^{-1} \sum_{1}^{n} \log\left(\frac{p_\omega}{p_{\omega'}}(X_\alpha)\right) + \Delta_{\omega'}(\omega) \right| \tag{6.3}$$

and

$$\mathscr{U}_\delta(\omega') = \{\Delta_{\omega'} < \delta\} \subset \Omega, \tag{6.4}$$

for $\omega' \in \Omega$, probability Λ on Ω and $\delta > 0$.

LEMMA 6.1. *For each $\delta > 0$ and $\omega' \in \Omega$,*

$$\frac{1}{2}\|P_{\hat\omega} - P_{\omega'}\| < \sqrt{2\delta} + \left[\mathscr{V}_{\Lambda,\omega'} > \frac{\delta}{4}\right] + \frac{e^{-(1/2)n\delta}}{\Lambda(\mathscr{U}_\delta(\omega'))}.$$

PROOF. By definition of $\hat\omega$,

$$\|P_{\hat\omega} - P_{\omega'}\| = \int \left| \int p_\theta \, d(\hat\Lambda \circ \omega) - p_{\omega'} \right| d\mu,$$

$$= \int \left| \int \left(\int p_\theta \, d\omega - p_{\omega'} \right) d\hat\Lambda(\omega) \right| d\mu \tag{6.5}$$

(by the Fubini theorem on $\Omega \times \Theta$)

$$\le \int \int |p_\omega - p_{\omega'}| \, d\hat\Lambda(\omega) \, d\mu = \int \|P_\omega - P_{\omega'}\| \, d\hat\Lambda(\omega).$$

For any ω, by (3.6) of Hannan (1960),

$$\tfrac{1}{2}\|P_\omega - P_{\omega'}\| \le \sqrt{\Delta_{\omega'}(\omega)}\,.$$

Clearly, the LHS above is less than or equal to 1 everywhere and, by the above, less than $\sqrt{2\delta}$ on $\mathscr{U}_{2\delta}(\omega')$. Combining this with (6.5), we get

$$\tfrac{1}{2}\|P_{\hat\omega} - P_{\omega'}\| < \sqrt{2\delta} + \hat\Lambda\big((\mathscr{U}_{2\delta}(\omega'))^c\big). \tag{6.6}$$

Since $\hat\Lambda$ has density wrt Λ proportional to $\exp(\Sigma_1^n \log p_\omega(X_\alpha))$ and $\mathscr{V}_{\Lambda,\omega'}$ is the sup norm of

$$n^{-1} \sum_{1}^{n} \log p_\omega(X_\alpha) + \Delta_{\omega'}(\omega) - n^{-1} \sum_{1}^{n} \log p_{\omega'}(X_\alpha)$$

on S_Λ, one easily gets [cf. equation (iii)′ of the addendum of Gilliland, Hannan and Huang (1976)]

$$\frac{\hat\Lambda\big((\mathscr{U}_{2\delta}(\omega'))^c\big)}{\hat\Lambda(\mathscr{U}_\delta(\omega'))} \le \frac{e^{-2n\delta + n\mathscr{V}}_{\Lambda,\omega'}}{\Lambda(\mathscr{U}_\delta(\omega'))e^{-n\delta - n\mathscr{V}}_{\Lambda,\omega'}}, \tag{6.7}$$

by bounding $\Sigma_1^n \log p_\omega(X_\alpha)$ above on $S_\Lambda \cap (\mathscr{U}_{2\delta}(\omega'))^c$ and below on $S_\Lambda \cap$

$\mathscr{U}_\delta(\omega')$. Since $\hat{\Lambda}$ is a probability, the LHS bounds $\hat{\Lambda}((\mathscr{U}_{2\delta}(\omega'))^c)$; while on the set $[\mathscr{Y}_{\Lambda,\omega'} \le \delta/4]$, the RHS is bounded by $(e^{-n\delta/2})/\Lambda(\mathscr{U}_\delta(\omega'))$. Using these in (6.6), we get the asserted bound. □

The following L_1 law of large numbers for random continuous functions will be used in the proof of Lemma 6.3. This result is a trivial generalization of Theorem A.3 of Datta (1988).

LEMMA 6.2 [Theorem A.3, Datta (1988)]. *Let (S, d) be a compact metric space and $\| \ \|$ denote the* sup norm *on $C(S)$, the space of real continuous functions on S. Let $\{Q_\nu : \nu \in \mathscr{N}\}$ be an arbitrary family of probability measures.* [*We use the measure to denote the corresponding expectation and use the superscript $^{(\nu)}$ to denote deviations of random elements from the values of their Q_ν expectations.*] *Let A_n denote the uniform expectation on $\{1, \dots, n\}$. Let $*$ denote the* (*iterated*) *operation* $\limsup_n \bigvee_\nu A_n \times Q_\nu$.

If

(i) *Under each Q_ν, for every $n \ge 1$, $H_{\nu n1}, H_{\nu n2}, \dots, H_{\nu nn}$ are independent $C(S)$ valued random elements with expectations belonging to* $\mathbb{R}^S$ $((Q_\nu H_{\nu nk})(s) = Q_\nu H_{\nu nk}(s)\ \forall\ s)$,

(ii) $$*\left(\|H_{\nu n.}^{(\nu)}\| - M\right)_+ \downarrow 0 \quad \text{as } M \uparrow \infty,$$

(iii) $\forall\ \varepsilon > 0$ *and* $s \in S$, *with* $V_{s\rho\nu nk} = \bigvee\{|[H_{\nu nk}^{(\nu)}]_s^t| : d(s,t) < \rho\}$,

$$*\left[V_{s\rho\nu n.} > \varepsilon\right] \downarrow 0 \qquad \text{as } \rho \downarrow 0,$$

then

$$*\|A_n H_{\nu n.}^{(\nu)}\| = 0.$$

PROOF. The proof is the same as that of Theorem A.3 of Datta (1988) with H_k replaced by $H_{\nu nk}$ throughout. □

REMARK 6.1. Let (ii +) and (iii +) denote (ii) and (iii), respectively, without the centerings $^{(\nu)}$. Then (ii +) implies (ii) and (ii +) and (iii +) together imply (iii). The proof of these statements can be found in Datta [Remark A.3 (1988)].

LEMMA 6.3. *Let Θ be compact and* A1 *and* A2 *hold. Then*

$$\bigvee_{\underline{\theta}} \mathbf{E}_{\underline{\theta}} \mathscr{Y}_{G_n} \to 0 \quad \text{as } n \to \infty.$$

PROOF. The conclusion readily follows by an application of Lemma 6.2 with $S = \Omega$, d equals any metric metrizing the weak convergence topology on Ω, $\mathscr{N} = \Theta^\infty$, $Q_{\underline{\theta}} = \times_{\alpha=1}^{\infty} P_{\theta_\alpha}$ for $\underline{\theta} \in \Theta^\infty$, $H_{\underline{\theta} n\alpha}(\omega) = \log(p_\omega / p_{G_n})(X_\alpha)$, $\omega \in \Omega$, $n \ge 1$, $1 \le \alpha \le n$.

(S, d) is a compact metric space by Theorem II.6.4 of Parthasarathy (1967).

A2 implies that, for each $\underline{\theta}$, $\log(p_\omega/p_{G_n})(X_\alpha)$ is finite valued for all ω, except possibly on a $\mathbf{P}_{\underline{\theta}}$ null set. Continuity of $\omega \rightsquigarrow p_\omega(x)$ follows from the continuity of $\theta \rightsquigarrow p_\theta(x)$ and its boundedness on compact Θ. Thus $H_{\underline{\theta} n \alpha}$'s satisfy (i). We verify (ii +) and (iii +) of Remark 6.1 in the present situation.

(ii +) holds because

$$*\big(\|H_{\underline{\theta} n.}\| - M\big)_+ \le \bigvee_{\theta} E_\theta\big(2h_\theta^* - M\big)_+ \downarrow 0, \quad \text{as } M \uparrow \infty,$$

by A2.

For any $\omega' \in \Omega$,

$$\bigvee_{\underline{\theta}} \mathbf{E}_{\underline{\theta}} \vee \left\{\left|H_{\underline{\theta} n\alpha}\right]_{\omega'}^{\omega}\right| : d(\omega, \omega') < \rho\right\} = \bigvee_{\theta} \int \vee \left\{\left|\log \frac{p_\omega}{p_{\omega'}}\right| : d(\omega, \omega') < \rho\right\} dP_\theta$$

decreases to 0 as $\rho \downarrow 0$ by A2, because the integrand decreases to 0 a.s. and is dominated by $2h_\theta^*$. [Use the above facts to prove convergence along any sequence $\{\theta_k\} \subset \Theta$ as $\rho = \rho_k \downarrow 0$.] This completes the verification of (iii +).

Also

$$\left\|A_n H_{\underline{\theta} n.}^{(\underline{\theta})}\right\| = \bigvee_{\omega} \left| n^{-1} \sum_1^n \log \frac{p_\omega}{p_{G_n}}(X_\alpha) - n^{-1} \sum_1^n \int \log \frac{p_\omega}{p_{G_n}} \, dP_{\theta_\alpha}\right| = \mathscr{V}_{G_n}.$$

Hence by Lemma 6.2, $*\mathscr{V}_{G_n} = \limsup_n \bigvee_{\underline{\theta}} \mathbf{E}_{\underline{\theta}} \mathscr{V}_{G_n} = 0$. □

LEMMA 6.4. *Let S_Λ be compact and A1′ and A2′ hold. Then*

$$\bigvee_{\omega' \in S_\Lambda} \mathbf{E}_{\omega'} V_{\Lambda, \omega'} \to 0 \quad \text{as } n \to \infty.$$

PROOF. We apply Lemma 6.2 with $S = S_\Lambda$, d equals any metric metrizing the weak convergence topology on Ω, $\mathscr{N} = S_\Lambda$, $Q_{\omega'} = \times_{\alpha=1}^{\infty} P_{\omega'}$, $\omega' \in S_\Lambda$, $H_{\omega' n \alpha}(\omega) = \log(p_\omega/p_{\omega'})(X_\alpha)$ for all $1 \le \alpha \le n$. Then

$$\left\|A_n H_{\omega' n.}^{(\omega')}\right\| = \bigvee_{\omega \in S_\Lambda} \left| n^{-1} \sum_1^n \log \frac{p_\omega}{p_{\omega'}}(X_\alpha) - n^{-1} \sum_1^n \int \log \frac{p_\omega}{p_{\omega'}} \, dP_{\omega'}\right| = \mathscr{V}_{\Lambda, \omega'},$$

$$\omega' \in S_\Lambda.$$

(i) holds by A1′ and the fact that under $P_{\omega'}$, $\log(p_\omega/p_{\omega'})(X_\alpha)$ is finite valued for all $\omega \in S_\Lambda$.

Verifications of (ii +) and (iii +) are similar to those in the proof of the previous lemma. (Change $\theta \in \Theta$ to $\omega' \in S_\Lambda$, Ω to S_Λ and use A1′ and A2′ in place of A1 and A2.)

Hence Lemma 6.2 implies that

$$*\mathscr{V}_{\Lambda, \omega'} = \limsup_n \bigvee_{\omega' \in S_\Lambda} \mathbf{E}_{\omega'} \mathscr{V}_{\Lambda, \omega'} = 0. \ \square$$

LEMMA 6.5. *Let S_Λ be compact and* A1′ *and* A2″ *hold. Then*

$$\mathbf{E}_{\omega'}\mathscr{V}_{\Lambda,\omega'} \to 0 \quad \text{as } n \to \infty, \forall\, \omega' \in S_\Lambda.$$

PROOF. Fix $\omega' \in S_\Lambda$. The proof follows, once again, by an application of Lemma 6.2 with $S = S_\Lambda$, d equals any metric metrizing the weak convergence topology on Ω, $\mathscr{N} = \{\omega'\}$, $Q_{\omega'} = \times_{\alpha=1}^{\infty} P_{\omega'}$, $H_{\omega' n\alpha}(\omega) = \log(p_\omega/p_{\omega'})(X_\alpha)$, $\omega \in S_\Lambda$, $1 \le \alpha \le n$.

As shown before, $\|A_n H_{\omega' n}^{(\omega')}\| = \mathscr{V}_{\Lambda,\omega'}$. (i) and (ii) hold as before. Verification of (iii +) is more direct in this case since for any $\omega'' \in S_\Lambda$,

$$\int \bigvee_{\omega \in S_\Lambda} \left\{\left|\log\frac{p_\omega}{p_{\omega''}}\right| : d(\omega,\omega'') < \rho\right\} p_{\omega'}\, d\mu \downarrow 0 \quad \text{as } \rho \downarrow 0$$

by (i), A2″ and the dominated convergence theorem. □

LEMMA 6.6. (*a*) *Let* A1′ *and* A2″ *hold. Then*

$$\Lambda(\mathscr{U}_\delta(\omega')) > 0, \quad \textit{for any } \delta > 0 \textit{ and } \omega' \in S_\Lambda.$$

(b) *If, moreover, S_Λ is compact and* A2′ *holds, then*

$$\bigwedge_{\omega' \in S_\Lambda} \Lambda(\mathscr{U}_\delta(\omega')) > 0, \quad \textit{for any } \delta > 0.$$

PROOF. (a) Fix $\delta > 0$ and $\omega' \in S_\Lambda$. The continuity of the function $\omega \rightsquigarrow \int \log(p_{\omega'}/p_\omega)\, dP_{\omega'}$ on S_Λ follows by the dominated convergence theorem since the integrand is continuous as noted before and is bounded by $h_{\omega'}^*$, which is $P_{\omega'}$ integrable by A2″. So the set $\mathscr{U}_\delta(\omega') \cap S_\Lambda$ is open in S_Λ and thus equals $N \cap S_\Lambda$ for some N open in Ω. Hence

$$\Lambda(\mathscr{U}_\delta(\omega')) = \Lambda(N) > 0, \tag{6.8}$$

since $\omega' \in N \cap S_\Lambda$.

(b) Fix $\delta > 0$ and $\omega' \in S_\Lambda$. Observe that the functions Δ_{ω_n} converge to $\Delta_{\omega'}$ pointwise on S_Λ by A1′, A2′, the Scheffé theorem and the dominated convergence theorem, and hence in Λ-distribution, if $S_\Lambda \ni \omega_n \to \omega'$. Hence by a defining property of the latter convergence [see Billingsley (1968), Theorem 2.1.iv]

$$\liminf_n \Lambda(\{\Delta_{\omega_n} < \delta\}) \ge \Lambda(\{\Delta_{\omega'} < \delta\}) \quad \text{if } \omega_n \to \omega',\ \omega_n \in S_\Lambda.$$

This shows that the function $\omega' \rightsquigarrow \Lambda(\mathscr{U}_\delta(\omega'))$ is lower semicontinuous on S_Λ. Hence it attains its infimum on S_Λ because it is compact.

The proof now ends by part (a). □

PROOF OF THEOREM 3.1. Since $S_\Lambda = \Omega$, $\mathscr{V}_{\Lambda, G_n} = \mathscr{V}_{G_n}$. Fix a $\delta > 0$. Consider the $\mathbf{E}_\theta$ expectation of the bound in Lemma 6.1 with $\omega' = G_n$. Now, as $n \to \infty$, the $\mathbf{E}_{\underline{\theta}}$ expectation of the second term goes to zero by Lemma 6.3. The third term is nonrandom and goes to zero uniformly in $\underline{\theta}$ since Lemma 6.6(b) applies in view of Remark 3.1. Thus

$$\limsup_n \bigvee_{\underline{\theta}} \mathbf{E}_{\underline{\theta}} \| P_{\hat{\omega}} - P_{G_n} \| \le 2\sqrt{2\delta}\,.$$

The proof ends, $\delta > 0$ being arbitrary. □

PROOF OF THEOREM 3.2. (a) Fix $\delta > 0$. This time consider the $\mathbf{E}_\omega$ expectation of the bound in Lemma 6.1 with $\omega' = \omega$. The expectation of the second term goes to zero by Lemma 6.5 and so does the third term by Lemma 6.6(a). The proof ends, once again, $\delta > 0$ being arbitrary.

(b) Since A2′ holds, Lemma 6.4 and 6.6(b) apply in this case to conclude that the above convergences of the expectations of the second and the third terms are uniform in ω on S_Λ. This finishes the proof as before. □

Acknowledgments. Part of this paper extends a result in my Ph.D. dissertation at Michigan State University written under the guidance of Professor James F. Hannan. I would like to express my sincere thanks to Prof. Hannan for many extremely helpful discussions. Thanks are also due to an Associate Editor and three referees for their careful readings of an earlier version of the manuscript leading to the correction of errors and improvements in presentation.

REFERENCES

BALDER, E. J., GILLILAND, D. C. and VAN HOUWELINGEN, J. C. (1983). On the essential completeness of Bayes empirical Bayes decision rules. *Statist. Decisions* **1** 503–509.

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

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BLUM, J. R. and SUSARLA, V. (1977). Estimation of a mixing distribution function. *Ann. Probab.* **5** 200–209.

DATTA, S. (1988). Asymptotically optimal Bayes compound and empirical Bayes estimators in exponential families with compact parameter space. Ph.D. dissertation, Dept. Statist. and Probab., Michigan State Univ.

DUNFORD, N. and SCHWARTZ, J. T. (1957). *Linear Operator, Part 1*. Interscience, New York.

FABIAN, V. and HANNAN, J. (1985). *Introduction to Probability and Mathematical Statistics*. Wiley, New York.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

GILLILAND, D. C., BOYER, J. E. and TSAO, H. J. (1982). Bayes empirical Bayes: Finite parameter case. *Ann. Statist.* **10** 1277–1282.

GILLILAND, D. C. and HANNAN, J. (1986). The finite state compound decision problem, equivariance and restricted risk components. In *Adaptive Statistical Procedures and Related Topics* (J. Van Ryzin, ed.) *IMS Lecture Notes–Monograph Series* **8** 129–145. IMS, Hayward, Calif.

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

HANNAN, J. F. (1957). Approximation to Bayes risk in repeated play. *The Theory of Games 3*. *Ann. Math. Studies* **39** 97–139. Princeton Univ. Press.

HANNAN, J. (1960). Consistency of maximum likelihood estimation of discrete distributions. In *Contributions to Probability Statistics: Essays in Honor of Harold Hotelling* (I. Olkin, et. al., eds.) 249–257. Stanford Univ. Press.

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

MEEDEN, G. (1972b). Bayes estimation of mixing distribution, the discrete case. *Ann. Math. Statist.* **43** 1993–1999.

PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic, New York.

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

ROBBINS, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 157–164. Univ. California Press, Berkeley.

ROLPH, J. E. (1968). Bayesian estimation of mixing distributions. *Ann. Math. Statist.* **39** 1289–1302.

DEPARTMENT OF STATISTICS
UNIVERSITY OF GEORGIA
ATHENS, GEORGIA 30602