ON THE CONSISTENCY OF POSTERIOR MIXTURES AND ITS APPLICATIONS

BY SOMNATH DATTA

University of Georgia

Consider i.i.d. pairs \((\theta_i, X_i)\), \(i \geq 1\), where \(\theta_i\) has an unknown prior distribution \(\omega\) and given \(\theta_1, X_1\) has distribution \(P_\theta\). This setup arises naturally in the empirical Bayes problems. We put a probability (a hyperprior) on the space of all possible \(\omega\) and consider the posterior mean \(\hat{\omega}\) of \(\omega\). We show that, under reasonable conditions, \(P_{\hat{\omega}} = \int P_\theta \, d\omega\) 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 \((X_1, \theta_1), (X_2, \theta_2), \ldots\) be a sequence of i.i.d. pairs where \(\theta_1\) is distributed according to some unknown prior distribution \(\omega\) and given \(\theta_1, X_1\) has distribution \(P_\theta\). 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 \(\theta\).

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 \(\omega\) is unknown. The problem is to estimate \(\omega\) from the \(X\)'s.

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

A well-known approach to solve the above problems is to put a hyperprior on the space of possible \(\omega\) and to consider the posterior mean \(\hat{\omega}\) of \(\omega\) 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 admisibility 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 \(\omega\) as the number of \(X\)'s approaches infinity. They both considered cases where the
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 \( \omega \) in probability whenever the selected hyperprior is sufficiently diffuse.

In the case of only finitely many \( P_\theta \)'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 \( \omega \) 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 \( \omega \) is asymptotically optimal for all \( \omega \) 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 \( P_\omega \) denote the joint marginal distribution of the X's and \( P_\theta \) denote the joint conditional distribution of the X's given the \( \theta \)'s. For each \( n \), let \( G_n \) stand for the empirical distribution of \( \theta_1, \theta_2, \ldots, \theta_n \). We prove that \( P_\omega \) is \( L_1 \) consistent for \( P_\omega \) and conditionally \( L_1 \) consistent for \( P_{G_n} \), uniformly in \( \omega \) and \( \theta \), 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 \( \mathcal{X} \) dominated by some \( \sigma \)-finite measure \( \mu \). We assume that \( \Theta \) is a metric space and consider the Borel \( \sigma \)-field on it. Suppose we have \( \{P_\theta; \theta \in \Theta\} \) on \( \mathcal{X} \) such that (a) \( P_\theta(x) \) is jointly measurable in \( \theta \) and \( x \) and (b) \( \forall \theta, p_\theta \) is a density of \( P_\theta \) w.r.t \( \mu \).

Let \( \Omega = \{\omega; \omega \) is a probability on \( \Theta \) with the Borel \( \sigma \)-field corresponding to the topology of weak convergence. For \( \omega \in \Omega \), let \( P_\omega \) stand for the mixture \( \int P_\theta d\omega \) and \( p_\omega \) denote its \( \mu \)-density \( \int p_\theta d\omega \).

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

(i) Bayes model: \( \omega \) is distributed as \( \Lambda \) and given \( \omega \), \( \theta \) is distributed as \( \omega^n = \times_{a=1}^n \omega \) and given \( \theta \) and \( \omega \), \( X \) is distributed as \( P_{\theta} = \times_{a=1}^n P_{\theta_a} \).
The above model gives rise to the following two marginal models:

(ii) Bayes compound model: \( \theta = \langle \theta_1, \ldots, \theta_n \rangle \) is distributed as \( \bar{\omega}^n_\Lambda \) and given \( \theta, X = \langle X_1, \ldots, X_n \rangle \) is distributed as \( P_\theta \), where \( \bar{\omega}^n_\Lambda = \Lambda \circ \omega^n \), i.e.,

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

Let \( E_\theta \) stand for the expectation under \( P_\theta \).

(iii) Bayes empirical Bayes model: \( \omega \) is distributed as \( \Lambda \) and given \( \omega, X \) is distributed as \( P_\omega = \times_{\alpha=1}^{n} P_\omega \).

Let \( E_\omega \) stand for the expectation under \( P_\omega \).

Let \( \hat{\Lambda} \) be the posterior distribution of \( \omega \) given \( \langle X_1, \ldots, X_n \rangle = \langle x_1, \ldots, x_n \rangle \). Then \( \hat{\Lambda} \) is the probability measure on \( \Omega \) 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 \( \theta_{n+1} \) given \( \langle X_1, \ldots, X_n \rangle = \langle x_1, \ldots, x_n \rangle \).

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

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

\[
\text{Prob}(\theta_{n+1} \in B, \langle X_1, \ldots, X_n \rangle \in A_1 \times \cdots \times A_n)
= \int_{A_1 \times \cdots \times A_n} \hat{\omega}(B) \left( \int \prod_{\alpha=1}^{n} p_\theta \ d\bar{\omega}^n_\Lambda \right) d\mu^n.
\]

Under model (i),

\[
LHS = \int_{(\Theta^n \times B) \times (A_1 \times \cdots \times A_n)} \prod_{\alpha=1}^{n} p_{\theta_\alpha} \ d(\mu^n \times \bar{\omega}^{n+1}_\Lambda)
= \int_{A_1 \times \cdots \times A_n} \left( \int_{\Theta^n \times B} \prod_{\alpha=1}^{n} p_{\theta_\alpha} \ d(\Lambda \circ \omega^{n+1}) \right) d\mu^n,
\]

by the Fubini theorem on \( \Theta^n \times \Theta^{n+1} \).

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

\[
\int \omega(B) \prod_{\alpha=1}^{n} p_\omega \ d\Lambda = \hat{\omega}(B) \left( \int \prod_{\alpha=1}^{n} p_{\theta_\alpha} \ d\bar{\omega}^n_\Lambda \right),
\]

by the definition of \( \hat{\omega} \) and another application of the Fubini theorem finishing the proof. \( \square \)
Our main results are that under reasonable conditions on \( \{p_\theta\} \) and \( \Lambda \), \( P_\theta \) is uniformly \( L_1 \) consistent for \( P_{G_n} \) under \( P_\theta \) and is \( L_1 \) consistent for \( P_\omega \) under \( P_\omega \), where \( G_n \) is the empirical distribution of \( \theta_1, \ldots, \theta_n \). These are presented in the next section.

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

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

\[
S_m = \bigcap \{ F : F \text{ is closed and } m(F^c) = 0 \}.
\]

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 \( \omega, \omega' \) in \( \Omega \), let

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

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

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

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

A1’. \( p_\omega(x) \) is continuous in \( \omega \) on \( S_\Lambda \) for each \( x \).

A2. With \( h^*_\theta = \bigvee_\theta \log (p_\theta/p_\theta) \),

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

A2’. With \( h^*_\omega = \bigvee_{\omega \in S_\Lambda} \log (p_\omega/p_\omega) \),

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

A2”. With \( h^*_\omega \) 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 \( \theta_1, \ldots, \theta_n \).
THEOREM 3.1. Suppose $\Theta$ is compact and A1 and A2 are satisfied. If $S_\Lambda = \Omega$, then

$$E_\bar{\theta}\|P^{\bar{\theta}} - P^{G_n}\| \to 0, \text{ uniformly in } \bar{\theta}, \text{ as } n \to \infty. \tag{3.1}$$

COROLLARY 3.1. Under the conditions of Theorem 3.1,

$$E_\omega\|P^{\omega} - P_\omega\| \to 0 \text{ as } n \to \infty, \forall \omega \in \Omega. \tag{3.2}$$

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

(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, $S_\Lambda$ is a compact subset of $\Omega$ and A1' and A2'' are satisfied. Then

$$E_\omega\|P^{\omega} - P_\omega\| \to 0 \text{ as } n \to \infty, \forall \omega \in S_\Lambda. \tag{3.3}$$

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

REMARK 3.1. If $S_\Lambda$ is compact and equals $\Omega$, 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 $P_\theta$'s to be pairwise mutually absolutely continuous. Similarly, A2'' forces two mixtures $P_\omega$ and $P_\omega'$ to be mutually absolutely continuous whenever $\omega$ and $\omega'$ are in the support of $\Lambda$. In the finite $\Theta$ 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 \( X_1, \ldots, X_n \) 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 \( \omega \) on the basis of \( X_1, \ldots, X_n \). In order that this estimation problem make sense, we assume the usual identifiability condition,

\[
\omega \to P_\omega \text{ is one-to-one on } \Omega,
\]

throughout this subsection.

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

**Theorem 4.1.** (a) Let \( \Theta \) be compact, \( S_\lambda = \Omega \) and A1' and A2" be satisfied. Then \( \forall \omega \in \Omega, \hat{\omega} \to \omega \) in probability \( (P_\omega) \).

(b) If, moreover, A2' is satisfied, then the above convergence is uniform in \( \omega \).

**Proof.** Let \( d \) be a metric metrizing the topology of weak convergence on \( \Omega \).

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

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

\[
P_\omega\{d(\hat{\omega}, \omega) > \varepsilon\} \leq P_\omega\{\|P_\hat{\omega} - P_\omega\| > \delta\} \leq \delta^{-1} E_\omega\|P_\hat{\omega} - P_\omega\| \to 0
\]
as \( n \to \infty \), by Theorem 3.2(a).

The last convergence is uniform in \( \omega \), 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 \( \Theta \) and \( S_\lambda = \Omega \) by the compactness of \( S_\lambda \) alone, we still get the consistency for \( \omega \) in \( S_\lambda \) provided we have, in addition, \( P_\omega(\hat{\omega} \notin S_\lambda) \to 0 \) 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 \( \Omega \) then,

\[
E_\omega d(\hat{\omega}, \omega) \to 0 \quad \text{as } n \to \infty, \forall \omega,
\]

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

**Remark 4.3.** Rolph (1968) and Meeden (1972b) established the strong (a.s. \( P_\omega \)) consistency of \( \hat{\omega} \) in cases where \( \mathcal{X} \) is the set of nonnegative integers. Rolph considered the case when \( p_\theta(x) \) is continuous in \( \theta \) and \( \Theta = [0, 1] \).
Meeden extended the consistency result to the case where \( p_\theta(x) \) satisfies continuity in \( \theta \) 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 \( \mathcal{X} \) and compact \( \Theta \).

**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 \( \Theta \). So if \( \Theta \) 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

\[
\hat{\omega}(B) = \frac{\int \prod_{\alpha = 1}^{n+1} p_{\theta_\alpha}(X_\alpha) d(\omega_{n+1}^\Lambda)}{\int \prod_{\alpha = 1}^{n} p_{\theta_\alpha}(X_\alpha) d(\omega_n^\Lambda)},
\]

for any Borel \( B \) of \( \Theta \).

Let \( \omega_{\theta_n}^\Lambda \) denote the posterior mean of \( \omega \) given \( \theta_n = \langle \theta_1, \ldots, \theta_n \rangle \) under model (i) of Section 2. Then, by repeated conditioning, it follows that

\[
d\omega_n^\Lambda(\theta_n) = \prod_{\alpha = 1}^{n} d\omega_{\theta_n-1}(\theta_\alpha), \quad \forall \ n \geq 1,
\]

where \( \omega_{\theta_0}^\Lambda = \int \omega \ d\Lambda \).

Using this in (4.1), we get

\[
\hat{\omega}(B) = \frac{\int \cdots \int \prod_{\alpha = 1}^{n+1} p_{\theta_\alpha}(X_\alpha) d(\omega_{n+1}^\Lambda)}{\int \cdots \int \prod_{\alpha = 1}^{n} p_{\theta_\alpha}(X_\alpha) d(\omega_{n-1}^\Lambda)}.
\]

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

\[
\omega_{\theta_n}^\Lambda = \left( \gamma + \sum_{\alpha = 1}^{i-1} \delta_{\theta_\alpha} \right)^{1/(\gamma[c,d] + i - 1)}, \quad i \geq 1,
\]

where \( \delta_\theta \) stands for the probability measure degenerate at \( \theta \).

**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 \( (\theta_1, X_1), (\theta_2, X_2), \ldots \), where \( \theta_1 \) is distributed as \( \omega \) and given \( \theta_1 \), \( X_1 \) is distributed as \( P_{\theta_1} \). \( \{P_\theta: \theta \in \Theta\} \) is known but \( \omega \in \Omega \) is unknown to the statistician.

At stage \( n \), a decision \( t_n = t_n(X_1, \ldots, X_n) \) about \( \theta_n \) has to be taken with loss \( L(t_n, \theta_n) \) and risk \( R_n(t_n, \omega) = \int L(t_n, \theta_n) dP_\theta d\omega_n \). 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

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

where \( R(t, \omega) = \int L(t, \omega) dP_t d\omega \) 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 \( \tau_\omega \) versus \( \omega \), i.e., there exists \( \tau_\omega \) such that \( R(\omega) = R(\tau_\omega, \omega) \).

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

**Theorem 4.2.** Suppose that \( \Theta \) is compact and

(i) \( R(\tau_\omega, \omega) \) is continuous in \( \omega' \) for each \( \omega \),
(ii) \( \omega \to P_\omega \) is one-to-one and continuous,
(iii) \( P_\omega \to P_\omega \) in probability (\( P_\omega \)), \( \forall \omega \).

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

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

$$\left| R_{n+1}(\hat{t}_{n+1}, \omega) - R(\omega) \right| \leq \mathbb{E}_\omega \left| R(\tau_\omega, \omega) - R(\omega) \right| \to 0, \quad \forall \omega. \quad \square$$

**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 \( \omega \) in \( \Omega_0 \) only. Moreover, let \( R(\tau_\omega, \omega') \) be continuous in \( \omega \) on \( \Omega_0 \) for all \( \omega' \in \Omega_0 \) and \( P_\omega(\tilde{\omega} \in \Omega_0) \to 0 \) as \( n \to \infty \), for all \( \omega \in \Omega_0 \). Then \( \hat{t}_{n+1}(X_{n+1}) = \tau_\omega(X_{n+1}) \) is a.o. relative to \( \Omega_0 \).

It can be shown that the empirical Bayes rule \( \hat{t}_{n+1}(X_{n+1}) = \tau_\omega(X_{n+1}) \) is the Bayes empirical Bayes rule versus \( \Lambda \). 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 \( \hat{t}_{n+1}(X_{n+1}) = \tau_\omega(X_{n+1}) \) is a.o.
PROOF. That (ii) of Theorem 4.2 holds is given by (I) and A1' with the Scheffé theorem. (iii) with $\hat{\omega} = \hat{\omega}$ is guaranteed by Theorem 3.2(a). The conclusion then follows by Theorem 4.2 with $\hat{\omega} = \hat{\omega}$. □

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

PROOF. The conditions of Remark 4.4 are satisfied by $\Omega_0 = S_\Lambda$ and $\hat{\omega} = \hat{\omega}$. 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 $L(t, \theta) = (t - \phi(\theta))^2$, $\phi$ being a measurable function on $\Theta$, and $P_\theta \ll \mu \, \forall \theta$ and for some $\mu$, 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, $\Theta$ is compact and hence A1 and A1' are equivalent.

EXAMPLE 5.1 (Finite $\Theta$). Let $\Theta = \{1, 2, \ldots, m\}$, 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 $1 \leq i, j \leq m$. Thus our Corollary 4.1 in the finite $\Theta$ 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 $P_1, P_2, \ldots, P_m$.

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 $a < b \in \mathbb{R}$, such that $f$ is increasing on $(a, b)$ and decreasing on $(b, \infty)$.

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

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

satisfies A1 and A2.

In this case, for any $\omega \in \Omega$, the characteristic function of $P_\omega$ is the product of the characteristic functions of $f$ and $\omega$. 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 \( \Theta = \prod_{i=1}^{k} [a_i, b_i] \subset \mathbb{R}^k \), where \( \Pi \) denotes the set theoretic product and \( \forall \theta \in \Theta, \ p_\theta(x) = c(\theta) e^{\theta T} \), for some measurable function \( T = (T_1, \ldots, T_k) : \mathcal{X} \to \mathbb{R}^k \) and \( c(\theta) = 1/\int e^{\theta T} d\mu \). Further assume that

\[
(5.1) \quad \prod_{i=1}^{k} [a_i, b_i] \subset \text{interior of } \{ \theta : \int e^{\theta T} d\mu < \infty \}.
\]

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

\[
p_\theta \leq c^* \sum_{d \in \Pi(a_i, b_i)} e^{d \cdot T}
\]

and

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

The above inequalities immediately show that A2 hold because

\[
(5.2) \quad \int |T_i| e^{d \cdot T} d\mu < \infty, \quad \forall \ i = 1, \ldots, k \text{ and } d \in \prod_{i=1}^{k} [a_i, b_i],
\]

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 \( x_0 < x_1, x_2, \ldots \in S_\mu \) such that \( \sum_{i \geq 1} (x_i - x_0)^{-1} = \infty \), 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 \( \Lambda \) 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

\[
(6.1) \quad \Delta_\omega(\omega') = \int \log(p_{\omega'}/p_\omega) \, dP_{\omega'}.
\]
Let
\begin{equation}
\gamma_{\omega'} = \bigvee_{\omega} \left| n^{-1} \sum_{1}^{n} \log \left( \frac{p_{\omega}}{p_{\omega'}}(X_{\alpha}) \right) + \Delta_{\omega'}(\omega) \right|
\end{equation}
and
\begin{equation}
\gamma_{\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|
\end{equation}
for \( \omega' \in \Omega \), probability \( \Lambda \) on \( \Omega \) and \( \delta > 0 \).

**Lemma 6.1.** For each \( \delta > 0 \) and \( \omega' \in \Omega \),
\[
\frac{1}{2} \| P_{\hat{\omega}} - P_{\omega} \| < \sqrt{2\delta} + \left[ \frac{\gamma_{\Lambda, \omega'}}{4} \right] + \frac{e^{-(1/2)n\delta}}{\Lambda(\%_{\delta}(\omega'))}.
\]

**Proof.** By definition of \( \hat{\omega} \),
\[
\| P_{\hat{\omega}} - P_{\omega} \| = \int \left| \int p_{\theta} d(\hat{\Lambda} \circ \omega) - p_{\omega} \right| d\mu,
\]
\[
= \int \left| \int \left( \int p_{\theta} d\omega - p_{\omega} \right) d\hat{\Lambda}(\omega) \right| d\mu
\]
(\text{by the Fubini theorem on } \Omega \times \Theta)
\[
\leq \int \int |p_{\omega} - p_{\omega}| d\hat{\Lambda}(\omega) \, d\mu = \int \| P_{\omega} - P_{\omega} \| d\hat{\Lambda}(\omega).
\]

For any \( \omega \), by (3.6) of Hannan (1960),
\[
\frac{1}{2} \| P_{\omega} - P_{\omega} \| \leq \sqrt{\Delta_{\omega'}(\omega)}.
\]

Clearly, the LHS above is less than or equal to 1 everywhere and, by the above, less than \( \sqrt{2\delta} \) on \( \%_{2\delta}(\omega') \). Combining this with (6.5), we get
\begin{equation}
\frac{1}{2} \| P_{\omega} - P_{\omega} \| < \sqrt{2\delta} + \hat{\Lambda}\left(\left(\%_{2\delta}(\omega')\right)^{c}\right).
\end{equation}

Since \( \hat{\Lambda} \) has density wrt \( \Lambda \) proportional to \( \exp(\Sigma_{1}^{n} \log p_{\omega}(X_{\alpha})) \) and \( \gamma_{\Lambda, \omega'} \) 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 \( S_{\Lambda} \), one easily gets [cf. equation (iii)' of the addendum of Gilliland, Hannan and Huang (1976)]
\begin{equation}
\frac{\hat{\Lambda}\left(\left(\%_{2\delta}(\omega')\right)^{c}\right)}{\hat{\Lambda}(\%_{\delta}(\omega'))} \leq \frac{e^{-2n\delta + n\rho}}{\Lambda(\%_{\delta}(\omega'))e^{-n\delta - n\rho}},
\end{equation}
by bounding \( \Sigma_{1}^{n} \log p_{\omega}(X_{\alpha}) \) above on \( S_{\Lambda} \cap (\%_{2\delta}(\omega'))^{c} \) and below on \( S_{\Lambda} \cap (\%_{\delta}(\omega'))^{c} \).
Since $\hat{\lambda}$ is a probability, the LHS bounds $\hat{\lambda}((\mathcal{C}_d^\omega(\omega'))^c)$; while on the set $[\mathcal{C}_d^\omega, \omega \leq \delta/4]$, the RHS is bounded by $(e^{-n\delta}/L(\mathcal{C}_d^\omega(\omega'))$. 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 $\{Q_v : v \in \mathcal{V}\}$ be an arbitrary family of probability measures. [We use the measure to denote the corresponding expectation and use the superscript $^{(v)}$ to denote deviations of random elements from the values of their $Q_v$ expectations.] Let $A_n$ denote the uniform expectation on $\{1, \ldots, n\}$. Let $\| \|$ denote the (iterated) operation $\limsup_n \cap v A_n \times Q_v$.

If

(i) Under each $Q_v$, for every $n \geq 1$, $H_{v_1}, H_{v_2}, \ldots, H_{v_n}$ are independent $C(S)$ valued random elements with expectations belonging to $\mathbb{R}^S ((Q_v, H_{v_{nk}})(s) = Q_v H_{v_{nk}}(s) \forall s)$,

(ii) $\lim (\|H_{v_{nk}}^{(v)}\| - M) \downarrow 0$ as $M \uparrow \infty$,

(iii) $\forall \varepsilon > 0$ and $s \in S$, with $V_{s_{0v_{nk}}} = \vee \{H_{v_{nk}}^{(v)} \leq s, d(s, t) < \rho\}$,

$\lim \mathbb{P} \left[ V_{s_{0v_{nk}}} > \varepsilon \right] \downarrow 0$ as $\rho \downarrow 0$,

then

$\lim A_n H_{v_{nk}}^{(v)} = 0$.

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

**Remark 6.1.** Let (ii +) and (iii +) denote (ii) and (iii), respectively, without the centerings $^{(v)}$. 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 $\Theta$ be compact and $A_1$ and $A_2$ hold. Then

$\sqrt{E_\Theta \mathcal{V}_{G_n}} \to 0$ 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 $\Omega$, $\mathcal{V}(\omega) = \mathcal{X}_{\alpha = 1}^\infty P_{\theta_{\alpha}}$ for $\theta \in \Theta^\omega$, $H_{\alpha}(\omega) = \log(p_\omega/P_{G_n}(\omega))$, $\omega \in \Omega$, $n \geq 1$, $1 \leq \alpha \leq n$. 

\(\mathcal{C}_d^\omega(\omega')\) is a probability, the LHS bounds $\hat{\lambda}((\mathcal{C}_d^\omega(\omega'))^c)$; while on the set $[\mathcal{C}_d^\omega, \omega \leq \delta/4]$, the RHS is bounded by $(e^{-n\delta}/L(\mathcal{C}_d^\omega(\omega'))$. 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 $\{Q_v : v \in \mathcal{V}\}$ be an arbitrary family of probability measures. [We use the measure to denote the corresponding expectation and use the superscript $^{(v)}$ to denote deviations of random elements from the values of their $Q_v$ expectations.] Let $A_n$ denote the uniform expectation on $\{1, \ldots, n\}$. Let $\| \|$ denote the (iterated) operation $\limsup_n \cap v A_n \times Q_v$.

If

(i) Under each $Q_v$, for every $n \geq 1$, $H_{v_1}, H_{v_2}, \ldots, H_{v_n}$ are independent $C(S)$ valued random elements with expectations belonging to $\mathbb{R}^S ((Q_v, H_{v_{nk}})(s) = Q_v H_{v_{nk}}(s) \forall s)$,

(ii) $\lim (\|H_{v_{nk}}^{(v)}\| - M) \downarrow 0$ as $M \uparrow \infty$,

(iii) $\forall \varepsilon > 0$ and $s \in S$, with $V_{s_{0v_{nk}}} = \vee \{H_{v_{nk}}^{(v)} \leq s, d(s, t) < \rho\}$,

$\lim \mathbb{P} \left[ V_{s_{0v_{nk}}} > \varepsilon \right] \downarrow 0$ as $\rho \downarrow 0$,

then

$\lim A_n H_{v_{nk}}^{(v)} = 0$.

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

**Remark 6.1.** Let (ii +) and (iii +) denote (ii) and (iii), respectively, without the centerings $^{(v)}$. 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 $\Theta$ be compact and $A_1$ and $A_2$ hold. Then

$\sqrt{E_\Theta \mathcal{V}_{G_n}} \to 0$ 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 $\Omega$, $\mathcal{V}(\omega) = \mathcal{X}_{\alpha = 1}^\infty P_{\theta_{\alpha}}$ for $\theta \in \Theta^\omega$, $H_{\alpha}(\omega) = \log(p_\omega/P_{G_n}(\omega))$, $\omega \in \Omega$, $n \geq 1$, $1 \leq \alpha \leq n$. 

\(\mathcal{C}_d^\omega(\omega')\) is a probability, the LHS bounds $\hat{\lambda}((\mathcal{C}_d^\omega(\omega'))^c)$; while on the set $[\mathcal{C}_d^\omega, \omega \leq \delta/4]$, the RHS is bounded by $(e^{-n\delta}/L(\mathcal{C}_d^\omega(\omega'))$. 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 $\{Q_v : v \in \mathcal{V}\}$ be an arbitrary family of probability measures. [We use the measure to denote the corresponding expectation and use the superscript $^{(v)}$ to denote deviations of random elements from the values of their $Q_v$ expectations.] Let $A_n$ denote the uniform expectation on $\{1, \ldots, n\}$. Let $\| \|$ denote the (iterated) operation $\limsup_n \cap v A_n \times Q_v$.

If

(i) Under each $Q_v$, for every $n \geq 1$, $H_{v_1}, H_{v_2}, \ldots, H_{v_n}$ are independent $C(S)$ valued random elements with expectations belonging to $\mathbb{R}^S ((Q_v, H_{v_{nk}})(s) = Q_v H_{v_{nk}}(s) \forall s)$,

(ii) $\lim (\|H_{v_{nk}}^{(v)}\| - M) \downarrow 0$ as $M \uparrow \infty$,

(iii) $\forall \varepsilon > 0$ and $s \in S$, with $V_{s_{0v_{nk}}} = \vee \{H_{v_{nk}}^{(v)} \leq s, d(s, t) < \rho\}$,

$\lim \mathbb{P} \left[ V_{s_{0v_{nk}}} > \varepsilon \right] \downarrow 0$ as $\rho \downarrow 0$,

then

$\lim A_n H_{v_{nk}}^{(v)} = 0$.

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

**Remark 6.1.** Let (ii +) and (iii +) denote (ii) and (iii), respectively, without the centerings $^{(v)}$. 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 $\Theta$ be compact and $A_1$ and $A_2$ hold. Then

$\sqrt{E_\Theta \mathcal{V}_{G_n}} \to 0$ 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 $\Omega$, $\mathcal{V}(\omega) = \mathcal{X}_{\alpha = 1}^\infty P_{\theta_{\alpha}}$ for $\theta \in \Theta^\omega$, $H_{\alpha}(\omega) = \log(p_\omega/P_{G_n}(\omega))$, $\omega \in \Omega$, $n \geq 1$, $1 \leq \alpha \leq n$. 

(S, d) is a compact metric space by Theorem II.6.4 of Parthasarathy (1967). A2 implies that, for each \( \theta \), \( \log(p_\omega/p_{G_n})(X_\alpha) \) is finite valued for all \( \omega \), except possibly on a \( P_\theta \) null set. Continuity of \( \omega \to p_\omega(x) \) follows from the continuity of \( \theta \to p_\theta(x) \) and its boundedness on compact \( \Theta \). Thus \( H_{\theta n\alpha} \)'s satisfy (i). We verify (ii + ) and (iii + ) of Remark 6.1 in the present situation.

(ii + ) holds because

\[
*\left( \|H_{\theta n\alpha}\| - M \right) \leq \bigvee_{\theta} E_\theta(2h^*_\theta - M) \downarrow 0, \quad \text{as } M \uparrow \infty,
\]

by A2.

For any \( \omega' \in \Omega \),

\[
\bigvee_{\theta} E_\theta \cup \left\{ \left| H_{\theta n\alpha} \right|^\omega \colon d(\omega, \omega') < \rho \right\} = \bigvee_{\omega} \int \bigvee \left\{ \left| \log \frac{p_\omega}{p_{G_n}} \right| \colon d(\omega, \omega') < \rho \right\} dP_\theta
\]

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

\[
\left\| A_n H_{\theta n\alpha}^{(\omega)} \right\| = \bigvee_{\omega} \left| n^{-1} \sum_{1}^{n} \log \frac{p_\omega}{p_{G_n}}(X_\alpha) - n^{-1} \sum_{1}^{n} \int \log \frac{p_\omega}{p_{G_n}} dP_{\theta_\alpha} \right| = \gamma_{G_n}.
\]

Hence by Lemma 6.2, \( *\gamma_{G_n} = \limsup_n \bigvee_{\theta} E_\theta \gamma_{G_n} = 0 \). \( \square \)

**Lemma 6.4.** Let \( S_\Lambda \) be compact and A1' and A2' hold. Then

\[
\bigvee_{\omega' \in S_\Lambda} E_\omega V_{\Lambda, \omega} \to 0 \quad \text{as } n \to \infty.
\]

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

\[
\left\| A_n H_{\omega n\alpha}^{(\omega)} \right\| = \bigvee_{\omega' \in S_\Lambda} \left| n^{-1} \sum_{1}^{n} \log \frac{p_\omega}{p_{\omega'}}(X_\alpha) - n^{-1} \sum_{1}^{n} \int \log \frac{p_\omega}{p_{\omega'}} dP_{\omega'} \right| = \gamma_{\Lambda, \omega'},
\]

\( \omega' \in S_\Lambda \).

(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 \( \omega' \in S_\Lambda \), \( \Omega \) to \( S_\Lambda \) and use A1' and A2' in place of A1 and A2.)

Hence Lemma 6.2 implies that

\[
*\gamma_{\Lambda, \omega'} = \limsup_n \bigvee_{\omega' \in S_\Lambda} E_\omega \gamma_{\Lambda, \omega'} = 0. \quad \square
\]
Lemma 6.5. Let $S_\Lambda$ be compact and $\Lambda'$ and $\Lambda''$ hold. Then

$$E_{\omega'} \mathcal{Y}_{\Lambda, \omega'} \to 0 \quad \text{as } n \to \infty, \forall \omega' \in S_\Lambda.$$

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 $\Omega$, $\mathcal{N} = \{\omega\}$, $Q_{\omega'} = \sum_{\alpha=1}^{\infty} P_{\omega'}$, $H_{\omega'n, \alpha}(\omega) = \log(p_{\omega}/p_{\omega'})(X_{\alpha})$, $\omega \in S_\Lambda$, $1 \leq \alpha \leq n$.

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

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

by (i), $\Lambda''$ and the dominated convergence theorem. 

Lemma 6.6. (a) Let $\Lambda'$ and $\Lambda''$ hold. Then

$$\Lambda(\mathcal{Y}_{\delta}(\omega')) > 0, \quad \text{for any } \delta > 0 \text{ and } \omega' \in S_\Lambda.$$

(b) If, moreover, $S_\Lambda$ is compact and $\Lambda'$ holds, then

$$\bigwedge_{\omega' \in S_\Lambda} \Lambda(\mathcal{Y}_{\delta}(\omega')) > 0, \quad \text{for any } \delta > 0.$$

Proof. (a) Fix $\delta > 0$ and $\omega' \in S_\Lambda$. The continuity of the function $\omega \mapsto \int \log(p_{\omega}/p_{\omega'}) dP_{\omega'}$ on $S_\Lambda$ 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 $\Lambda''$. So the set $\mathcal{Y}_{\delta}(\omega') \cap S_\Lambda$ is open in $S_\Lambda$ and thus equals $N \cap S_\Lambda$ for some $N$ open in $\Omega$. Hence

$$(6.8) \quad \Lambda(\mathcal{Y}_{\delta}(\omega')) = \Lambda(N) > 0,$$

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

(b) Fix $\delta > 0$ and $\omega' \in S_\Lambda$. Observe that the functions $\Delta_{\omega_n}$ converge to $\Delta_{\omega'}$ pointwise on $S_\Lambda$ by $\Lambda'$, $\Lambda''$, the Scheffé theorem and the dominated convergence theorem, and hence in $\Lambda$-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]

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

This shows that the function $\omega' \mapsto \Lambda(\mathcal{Y}_{\delta}(\omega'))$ is lower semicontinuous on $S_\Lambda$. Hence it attains its infimum on $S_\Lambda$ because it is compact.

The proof now ends by part (a). 

Proof of Theorem 3.1. Since $S_\Lambda = \Omega$, $\mathcal{Y}_{\Lambda, G_n} = \mathcal{Y}_{G_n}$. Fix a $\delta > 0$. Consider the $E_{\rho}$ expectation of the bound in Lemma 6.1 with $\omega' = G_n$. Now, as $n \to \infty$, the $E_{\rho}$ expectation of the second term goes to zero by Lemma 6.3. The third term is nonrandom and goes to zero uniformly in $\bar{\Omega}$ since Lemma 6.6(b) applies.
in view of Remark 3.1. Thus
\[ \limsup_n \bigvee_{\omega \in \mathcal{E}} \| P_{\omega} - P_{\omega_n} \| \leq 2\sqrt{2\delta}. \]

The proof ends, \( \delta > 0 \) being arbitrary. \( \Box \)

**Proof of Theorem 3.2.** (a) Fix \( \delta > 0 \). This time consider the \( E_\omega \) 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 \( \omega \) on \( S_A \). This finishes the proof as before. \( \Box \)

**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**


