The Annals of Statistics
1991, Vol. 19, No. 1, 354–365

ASYMPTOTIC OPTIMALITY OF BAYES COMPOUND ESTIMATORS IN COMPACT EXPONENTIAL FAMILIES

BY SOMNATH DATTA

University of Georgia

The problem of finding admissible, asymptotically optimal compound rules is pursued in the infinite state case. The components involve the estimation of an arbitrary continuous transform of the natural parameter of a real exponential family with compact parameter space. We show that all Bayes estimators are admissible. Our main result is that any Bayes compound estimator versus a mixture of i.i.d. priors on the compound parameter is asymptotically optimal if the mixing hyperprior has full support.

The asymptotic optimality results are generalized to weighted squared error loss with continuous weight function and applications to some nonexponential situations are also considered.

Several examples of such hyperpriors are given and for some of them practically useful forms of the corresponding Bayes estimators are obtained.

1. Introduction. We start with some notational conventions used throughout the body of this paper. Given any vector $\underset{\sim}{x} = (x_1, \dots, x_n)$, for each $1 \le \alpha \le n$, $\underset{\sim}{x}_\alpha$ denotes the vector $(x_1, \dots, x_\alpha)$ and $\underset{\sim}{x}_\alpha^\vee$ denotes the vector $(v_1, \dots, v_{n-1})$ with $v_j = x_j$ for $j < \alpha$ and $= x_{j+1}$ for $j \ge \alpha$. Typically the letter P is used for probabilities and E for the corresponding expectations. For probabilities $P_1, \dots, P_n$, $\times_{\alpha=1}^n P_\alpha$ denotes their measure theoretic product. For the sake of clarification, dummy variables are often displayed in integrals. Also mixed mode integral expressions like $\int X(\omega)\, dP$ are used. If X is a random variable on a probability space $(\cdot, \cdot, P)$, then XP^{-1} denotes the distribution of X under P. For a bounded function f, f_* and f^* denote its infimum and supremum, respectively, over its entire domain. $\mathbb{R}$, $\mathbb{Z}$ and $\mathbb{N}$ stand for the set of reals, integers and nonnegative integers, respectively.

1.1. *The component and the compound problem.* The component problem has the structure of the usual decision theory problem, that is, we have a parameter space Θ, a family of probability measures $\{P_\theta\colon \theta \in \Theta\}$ on some common measurable space $\mathscr{X}$, an observable $\mathscr{X}$-valued random variable $X \sim P_\theta$ under θ, an action space $\mathscr{A}$, a loss function $L\colon \mathscr{A} \times \Theta \to [0, \infty)$ and decision rules t, $t\colon \mathscr{X} \to \mathscr{A}$ such that $L(t, \theta)$ is measurable for each θ, with risk $R(t, \theta) = E_\theta L(t, \theta)$.

Received September 1988; revised December 1989.

AMS 1980 *subject classifications*. Primary 62C25; secondary 62C15.

Key words and phrases. Bayes compound estimators, asympotic optimality, exponential families, prior, consistency.

The compound problem simultaneously considers a number, say n, of independent decision problems each of which is structurally identical to the above component problem and the compound loss is taken to be the average of all the component losses. Thus, for each $n \geq 1$, the compound problem can be formulated as a decision problem as follows. We have the parameter space Θ^n, the action space $\mathscr{A}^n$, observations $\underline{X} = (X_1, \dots, X_n) \sim \mathbf{P}_{\underline{\theta}} = \times_{\alpha=1}^n \mathbf{P}_{\theta_\alpha}$, $\underline{\theta} = (\theta_1, \dots, \theta_n) \in \Theta^n$, compound rules $\underline{t} = (t_1, \dots, t_n)$ with loss $L_n(\underline{t}, \underline{\theta}) = n^{-1}\Sigma_1^n L(t_\alpha, \theta_\alpha)$ and risk

$$R_n(\underline{t}, \underline{\theta}) = \mathbf{E}_{\underline{\theta}} L_n(\underline{t}, \underline{\theta}). \tag{1.1}$$

In the set compound version t_α is a function of $X_1, X_2, \dots, X_n$ for all $1 \leq \alpha \leq n$. In the sequence compound version, however, the statistician is allowed to use the data $\underline{X}_\alpha = (X_1, \dots, X_\alpha)$ up to stage α in making the αth decision and so t_α is a function of $X_1, \dots, X_\alpha$, for all $1 \leq \alpha \leq n$.

Let $\Omega = \{\omega\colon \omega$ is a probability on $\Theta\}$. For $\omega \in \Omega$, let $R(\omega)$ stand for the minimum Bayes risk versus ω in the component problem, that is

$$R(\omega) = \bigwedge_t \int R(t, \theta)\, d\omega(\theta).$$

For a traditional simple symmetric rule [i.e., $t_\alpha(\underline{x}) = t(x_\alpha)\ \forall\ 1 \leq \alpha \leq n$ for some component rule t] the compound risk is easily seen to be at least $R(G_n)$, G_n being the empirical distribution of $\theta_1, \dots, \theta_n$. Thus compound rules which attain risks asymptotically no more than $R(G_n)$ are of interest. Hannan (1957) used the term "approximation to Bayes risk" to describe such effects. For a compound rule $\underline{t}$, the difference $D_n(\underline{t}, \underline{\theta}) = R_n(\underline{t}, \underline{\theta}) - R(G_n)$ is called the modified regret of $\underline{t}$ at $\underline{\theta}$. We say that a rule $\underline{t}$ is asymptotically optimal (a.o.) if

$$\bigvee_{\underline{\theta}} D_n(\underline{t}, \underline{\theta})_+ \to 0 \quad \text{as } n \to \infty. \tag{1.2}$$

For the relation of this notion of optimality to that with a more stringent envelope in the finite Θ case, see Gilliland and Hannan (1986).

A set/sequence compound rule $\underline{t}$ is said to be admissible if for each $n \geq 1$, $R_n(\underline{t}, \underline{\theta})$ is admissible in the usual decision theoretic sense as a function of $\underline{\theta}$ in the class of set/sequence compound rules.

1.2. *Literature review and a summary.* The problem of exhibiting compound rules which are a.o. as well as admissible has been an interesting and challenging question ever since it was put forward by Robbins (1951). His featured example was decision between $N(-1, 1)$ and $N(1, 1)$. He exhibited an a.o. compound procedure and called it asymptotically subminimax by comparison with the simple symmetric minimax rule. He proposed the Bayes compound rule versus the symmetric prior uniform on proportions for his featured example and conjectured that it might have better risk behavior than his asymptotically subminimax rule. Inglis (1973) studied the asymptotic optimality of a class of admissible Bayes rules for two state components under the finiteness of the expected log densities and tacit [cf. Inglis (1979)] nonatomicity conditions for his "generalization" of the Hannan–Robbins theorem. Gilliland and Hannan (1974), which was later published in 1986, treated the more general problem of restricted risk components in the finite Θ case. They worked with a more stringent envelope and reduced the problem of asymptotic optimality to the problem of establishing the L_1 consistency of certain induced estimators. Gilliland, Hannan and Huang (1976) established that consistency in two state components for Bayes compound estimators versus certain symmetric priors including the Robbins prior. This approach yielded for them admissible rules which are a.o. with rates as good as $O(n^{-1/2})$ in the general two state component case. Vardeman (1978) successfully exploited a result by the last authors to obtain admissible a.o. sequence compound rules in the two state component case.

None of the results mentioned in the previous paragraph goes beyond the finite Θ case. Inglis (1973) attempted to prove the admissibility and asymptotic optimality of a class of Bayes compound estimators versus mixtures of i.i.d. priors in estimating a geometric parameter with compact parameter space. Unfortunately his proof of asymptotic optimality appears to contain certain serious gaps. For a discussion on this see the addendum of Gilliland and Hannan (1986).

The present work, which subsumes Inglis's example, seems to be the first successful attempt in the literature to accomplish compound admissibility and asymptotic optimality simultaneously in the nonfinite state case. Our component distributions form a one dimensional exponential family of quite general nature, whose examples include well known exponential families such as normal, exponential, geometric, Poisson and negative binomial, where the parameter space is any compact interval of the natural parameter space on which the first moment is finite. The component problem is to estimate an arbitrary continuous transform of the natural parameter under squared error loss. We note that all compound Bayes estimators in our problem are admissible. Our main result is that those Bayes versus a mixture of i.i.d. priors on the compound parameter are a.o. if the mixing hyperprior has full support.

2. Exponential family component. Our component problem is the following: $\Theta = [c, d]$, $-\infty < c < d < \infty$, $\mathscr{A} = \mathbb{R}$ and $L(a, \theta) = (a - \phi(\theta))^2$, where ϕ is any real continuous function on Θ and $\forall\ \theta \in \Theta$, P_θ admits a density p_θ wrt a common σ-finite μ on $\mathbb{R}$ given by

$$p_\theta(x) = e^{\theta x} h(\theta), \qquad x \in \mathbb{R}. \tag{2.1}$$

In addition μ is assumed to satisfy $\mu_1 \ll \mu$, where $\mu_k = \mu s_k^{-1}$ and $s_k(x) = x + k$, $x \in \mathbb{R}$; $k \in \mathbb{Z}$.

Clearly $c, d \in \tilde{\Theta} = \{\theta\colon \int e^{\theta x}\, d\mu(x) < \infty\}$. Throughout we will assume

$$\int x e^{cx}\, d\mu(x) < \infty, \qquad \int x e^{dx}\, d\mu(x) < \infty. \tag{2.2}$$

It is well known [e.g., Lehmann (1986), Theorem 2.9] that (2.2) holds if $c, d \in \operatorname{int} \tilde{\Theta}$.

For $\omega \in \Omega$, let P_ω denote the ω mix of P_θ's and p_ω denote its μ density, $p_\omega(x) = \int p_\theta(x)\, d\omega$, $x \in \mathbb{R}$.

The following consequences of our assumptions are worth noting and some will be used in the later sections.

C1. $h(\theta) = (\int e^{\theta x}\, d\mu(x))^{-1}$ and h is continuous and positive on compact Θ. Consequently, $0 < h_* \le h^* < \infty$. [Indeed $h_* = h(c) \wedge h(d)$, since $\log h$ is concave by the Hölder inequality.]

C2. For any $\theta \in \Theta$,

$$p_\theta \le (h^*/h_*)(p_c + p_d), \tag{2.3}$$

and hence any $f \in L_1(P_c) \cap L_1(P_d)$ is uniformly integrable wrt the family of probability measures $\{P_\theta,\ \theta \in \Theta\}$.

C3. Since, for any $\theta \in \Theta$,

$$h_*(e^{cx} \wedge e^{dx}) \le p_\theta(x) \le h^*(e^{cx} \vee e^{dx}), \tag{2.4}$$

and for any $\omega \in \Omega$, p_ω inherits the above bound (2.4) on p_θ, we have

$$\frac{p_\omega(x)}{p_{\omega'}(x)} \le \frac{h^*}{h_*} e^{|x|(d-c)}, \qquad x \in \mathbb{R} \tag{2.5}$$

for any $\omega, \omega' \in \Omega$.

We will use $\mathbf{P}$ for $\mathbf{P}_{\underline{\theta}} = \times_{\alpha=1}^{n} P_{\theta_\alpha}$ and $\mathbf{E}$ for the corresponding expectation. Since $\mathscr{X} = \mathbb{R}$ in our problem, we will view $X_1, \dots, X_n$ to be the coordinate functions on $\mathbb{R}^n$. On the other hand, any measurable function H on $\mathbb{R}^n$ will be viewed as the random variable $H(\underline{X})$ whenever convenient.

3. Estimators induced by priors on Ω.

3.1. *Bayes versus mixture of i.i.d. priors.* Consider Ω with the topology of weak convergence. Let $B(\Omega)$ denote the Borel σ field of Ω. Let Λ be a probability on $(\Omega, B(\Omega))$. Define the prior (for each n) $\bar{\omega}_\Lambda^n$ on Θ^n as

$$\bar{\omega}_\Lambda^n(B_1 \times \cdots \times B_n) = \int \prod_{i=1}^{n} \omega(B_i)\, d\Lambda, \tag{3.1}$$

for $B_1, \dots, B_n$ Borels of Θ. (Note that the above integral makes sense because the integrand is nonnegative and measurable. For a proof of measurability it suffices to take $n = 1$ and B_1 open. But then it follows from a defining property of weak convergence [Billingsley (1968), Theorem 2.1.iv] that $\{\omega: \omega(B_1) > \varepsilon\}$ is open and hence measurable $\forall\ \varepsilon > 0$.) Hereafter we will drop the superscript n in $\bar{\omega}_\Lambda^n$, as it will be clear from the context.

By the Fubini theorem, the α-component Bayes risk against $\bar{\omega}_\Lambda$ of an estimator $\underline{t} = (t_1, \dots, t_n)$ in our set compound problem is

$$(3.2) \qquad R(t_\alpha, \bar{\omega}_\Lambda) = \int_{\mathbb{R}^n} \left(\int_{\Theta^n} (t_\alpha - \phi(\theta_\alpha))^2 \prod_{i=1}^{n} p_{\theta_i}(x_i) \, d\bar{\omega}_\Lambda \right) d\mu^n.$$

Let $\Lambda_{\alpha n}$ denote the conditional probability measure on Ω given $\underline{X}_\alpha^\vee = \underline{x}_\alpha^\vee$. Then $\Lambda_{\alpha n}$ has density proportional to $e^{g_{\alpha n}}$ wrt Λ where $g_{\alpha n}(\omega) = \sum_{i \neq \alpha}^n \log p_\omega(x_i)$. Let $\omega_{\alpha n} = \Lambda_{\alpha n} \circ \omega$. By the Fubini theorem (on the space $\Omega \times \Theta^n$), the inner integral in (3.2) is

$$\int \int (t_\alpha - \phi(\theta_\alpha))^2 p_{\theta_\alpha}(x_\alpha) \, d\omega \, e^{g_{\alpha n}(\omega)} \, d\Lambda,$$

which then, by definitions of $g_{\alpha n}$ and $\omega_{\alpha n}$, equals

$$(3.3) \qquad \left(\int e^{g_{\alpha n}(\omega)} \, d\Lambda \right) \int (t_\alpha - \phi(\theta_\alpha))^2 p_{\theta_\alpha}(x_\alpha) \, d\omega_{\alpha n}(\theta_\alpha).$$

The compound risk being the average of the risks across the components, it follows from (3.2) and (3.3) that the estimator which plays component Bayes versus $\omega_{\alpha n}$ in the αth estimation $\forall \alpha = 1, \dots, n$ is Bayes versus $\bar{\omega}_\Lambda$ in the set compound problem. Let τ_ω denote the Bayes estimator of $\phi(\theta)$ versus a prior ω on Θ in the component problem,

$$\tau_\omega(x) = \int \phi(\theta) p_\theta(x) \, d\omega / p_\omega(x).$$

Thus the Bayes set compound estimator $\hat{\underline{t}}$ versus $\bar{\omega}_\Lambda$ is then given by

$$(3.4) \qquad \hat{t}_\alpha(\underline{x}) = \tau_{\omega_{\alpha n}}(x_\alpha), \qquad \alpha = 1, \dots, n.$$

In a similar fashion, or otherwise by relating the sequence rules with the set rules and using the above, one can prove that the Bayes sequence compound estimator versus $\bar{\omega}_\Lambda$ is given by

$$(3.5) \qquad \hat{t}_\alpha(\underline{x}_\alpha) = \tau_{\omega_{\alpha\alpha}}(x_\alpha), \qquad 1 \le \alpha \le n.$$

3.2. *Admissibility.* For each $\underline{\theta} \in \Theta^n$, $\mathbf{P}$ and μ^n are mutually absolutely continuous. Hence for any prior ζ on Θ^n, $\mathbf{P} \ll \int \mathbf{P} \, d\zeta$, the marginal distribution of $\underline{X}$, for all $\underline{\theta} \in \Theta^n$. Hence, by an immediate application of Lemma A of the Appendix, we get that for each $n \ge 1$, all Bayes estimators in our compound problems are admissible. In particular, $\hat{\underline{t}}$ of (3.4) and of (3.5) are so in the respective versions.

3.3. *Asymptotic optimality.* The support of any probability Λ on Ω is defined to be the smallest closed set with Λ probability 1. Our main result is that the Bayes compound estimators versus $\bar{\omega}_\Lambda$ are asymptotically optimal whenever Λ has full support. The proof of this result is given in the next section.

Recall that G_n stands for the empirical distribution of $\theta_1, \dots, \theta_n$. Let $\tilde{t}_\alpha(\underline{x}) = \tau_{G_n}(x_\alpha)$, $1 \le \alpha \le n$. Then the modified regret of $\hat{\underline{t}}$ at $\underline{\theta}$ is

$$(3.6) \qquad D_n(\hat{\underline{t}}, \underline{\theta}) = n^{-1} \sum_{\alpha=1}^{n} \left\{ \mathbf{E}(\hat{t}_\alpha - \phi(\theta_\alpha))^2 - \mathbf{E}(\tilde{t}_\alpha - \phi(\theta_\alpha))^2 \right\}.$$

Since $\hat{t}_\alpha, \tilde{t}_\alpha, \phi(\theta_\alpha) \in \phi[\Theta]$,

$$|D_n(\hat{\underline{t}}, \underline{\theta})| \le 2 \operatorname{diam} \phi[\Theta] n^{-1} \sum_{\alpha=1}^{n} \mathbf{E}|\hat{t}_\alpha - \tilde{t}_\alpha|. \tag{3.7}$$

THEOREM. *If* supp $\Lambda = \Omega$, *then* (a) *with* $\hat{\underline{t}}$, *the Bayes compound estimator given by* (3.4),

$$\bigvee_{\alpha \le n} \mathbf{E}|\hat{t}_\alpha - \tilde{t}_\alpha| \to 0, \quad \textit{uniformly in } \underline{\theta}, \textit{ as } n \to \infty \tag{3.8}$$

and consequently, $\hat{\underline{t}}$ *is asymptotically optimal*; (b) *the Bayes sequence compound estimator given by* (3.5) *is asymptotically optimal.*

3.4. *Examples of* Λ. We briefly discuss two examples of Λ with support Ω, as needed in the above theorem. For more details on them and some more examples see Datta (1988).

EXAMPLE A (Dirichlet process). An important class of priors on the probabilities on $\mathbb{R}$ with manageable posteriors has been introduced by Ferguson (1973). Let γ be a nonnull, finite Borel measure on $\mathbb{R}$. Then Λ is called the Dirichlet process prior with parameter γ [hereafter we write $\Lambda = \mathscr{D}(\gamma)$] if for every finite measurable partition $\{B_1, \dots, B_m\}$ of $\mathbb{R}$ the distribution of $(\omega(B_1), \dots, \omega(B_m))$ under Λ (ω is the identity function on the space of probabilities on $\mathbb{R}$) is Dirichlet with parameters $(\gamma(B_1), \dots, \gamma(B_m))$. It is well known [e.g., Ferguson (1974)] that supp $\mathscr{D}(\gamma)$ is the set of probability distributions on $\mathbb{R}$ whose support is contained in supp γ. So if we choose γ with supp $\gamma = \Theta = [c, d]$, then $\Lambda = \mathscr{D}(\gamma)$ has support Ω.

EXAMPLE B (Distributions on the moment space). Let $D = \{(\mu_1, \mu_2, \dots)$: $\mu_i = \int \theta^i \, d\omega, \forall\, i \ge 1$, for some $\omega \in \Omega\} \subset \mathbb{R}^\infty$ be the space of moment sequences of probabilities on Θ. Since any $\omega \in \Omega$ is determined by its moment sequence $\{\mu_i\} \in D$, Θ being compact, a prior on D induces a prior on Ω in the obvious way. To make the ideas precise consider Ω with the weak convergence topology and D with the product topology. Let $\boldsymbol{\mu}$ be the mapping $\omega \rightsquigarrow (\mu_1(\omega), \mu_2(\omega), \dots)$, $\mu_i(\omega) = \int \theta^i \, d\omega$, $i \ge 1$. Then $\boldsymbol{\mu}$ is 1–1, continuous, onto D and hence is a homeomorphism since Ω is compact and D is Hausdorff. Thus a prior Λ_D on $(D, \mathscr{B}(D))$ induces the prior $\Lambda = \Lambda_D \circ \boldsymbol{\mu}$ on $(\Omega, \mathscr{B}(\Omega))$. Since $\boldsymbol{\mu}$ is a homeomorphism, supp $\Lambda = \Omega$ iff supp $\Lambda_D = D$.

The structure of D for the case $\Theta = [0, 1]$ has been studied by many authors. Rolph (1968) exploited this structure to define his prior sequentially on the coordinates. His priors can be adapted to the case $\Theta = [c, d]$ by the reparametrization $\theta \rightsquigarrow (\theta - c)/(d - c)$.

Another way of putting priors on D would be to follow Rolph's approach directly for $D = D[c, d]$, which has the same structure as $D[0, 1]$. See Datta (1988) for more details.

REMARK 3.1. Datta (1988) obtained a computationally feasible form of the Bayes compound estimators. He expressed $\hat{t}_\alpha$ as a ratio of two finite dimensional integrals wrt the posterior means of ω under Λ. Kuo (1986) also obtained the same expression in Example A, for which the posterior means are known, and described a Monte Carlo method for its calculation. When P_θ is geometric, negative binomial or Poisson, Example B yields a computationally feasible form of $\hat{t}_\alpha$. Some detailed studies in the direction of finding implementable $\hat{t}_\alpha$ and appropriate numerical methods seem desirable.

3.5. *Extensions.*

1. It should be obvious that we can also treat the cases where the components are 1–1 transforms of some exponential families we have been considering. Suppose that the component distributions P_θ, $\theta \in \Theta$, are such that $\{Q_\eta: \eta \in H\}$ form one such exponential family where $Q_\eta = P_{\psi^{-1}(\eta)}T^{-1}$, T and ψ are 1–1 transformations on $\mathbb{R}$ and Θ, respectively, and ψ^{-1} is continuous. Let $\underline{X} \sim \mathbf{P}_{\underline{\theta}}$. Then $\underline{Y} \sim \mathbf{Q}_{\underline{\eta}}$ where $Y_\alpha = T(X_\alpha)$ and $\eta_\alpha = \psi(\theta_\alpha)$, $1 \le \alpha \le n$. Since T is 1–1, estimators (based on $\underline{Y}$) in the transformed problem are related in a 1–1 fashion to the estimators (based on $\underline{X}$) in the original problem. Any such two estimators have identical risk function under a common parametrization. Moreover since ψ^{-1} is continuous, ϕ remains continuous in the reparametrization η of the transformed problem. Hence the conclusion of Section 3.2 and the theorem in Section 3.3 for the transformed problem implies that the set compound estimator
$$\hat{t}_\alpha(\underline{X}) = \frac{\int \phi(\psi^{-1}(\eta)) q_\eta(T(X_\alpha))\, d\omega_{\alpha n}}{\int q_\eta(T(X_\alpha))\, d\omega_{\alpha n}}, \qquad \alpha = 1, \dots, n,$$
is admissible and a.o. for estimating $\underline{\phi}$ in the original problem, where $\omega_{\alpha n}$ is as in Section 3.1 with $g_{\alpha n}(\omega) = \sum_{i \ne \alpha}^{n} \log q_\omega(T(X_i))$. Analogous conclusions hold for the sequence compound version.
2. We can generalize the component loss to weighted squared error loss, where the weight function is positive and continuous. If $L(a, \theta) = w(\theta)(a - \phi(\theta))^2$, then a component Bayes estimator of $\phi(\theta)$ is the ratio of the corresponding Bayes estimators of $w(\theta) \cdot \phi(\theta)$ and $w(\theta)$ wrt the squared error loss. Since $w\phi$ is continuous and $w_* > 0$, the $L_1(\mathbf{E})$ case of Lemma 4.1 (see the next section) with $L = 2w^*|\phi|^*/w_*$ and two applications of the theorem for the squared error loss imply that $\mathbf{E}|\hat{t}_\alpha - \tilde{t}_\alpha| \to 0$ uniformly in α and $\underline{\theta}$ where $\hat{t}_\alpha$ and $\tilde{t}_\alpha$ refer here to the weighted loss. As before this is sufficient to conclude the asymptotic optimality of $\underline{\hat{t}}$ since (3.7) holds with w^* multiple of the RHS.
3. If we have $\mu_{-1} \ll \mu$ instead of $\mu_1 \ll \mu$, we can use the transformation $T(x) = -x$ in extension 1 above to obtain admissible, a.o. estimators.

4. Proof of asymptotic optimality. We first introduce a few lemmas which are needed for the proof of the theorem. For a set A, let $[A]$ denote its indicator function. The following Datta–Singh inequality is Lemma A.1 of Datta (1988). Hence we omit the proof.

LEMMA 4.1. *For* $\langle y, z, Y, Z, L\rangle \in \mathbb{R}^5$ *and* $z \neq 0 \leq L$,

$$|z|\left\{\left|\frac{y}{z} - \frac{Y}{Z}\right| \wedge L\right\} \leq |y - Y| + \left(\left|\frac{y}{z}\right| + L\right)|z - Z|.$$

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

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

and for a function f on $\mathbb{R}$ and any $k \in \mathbb{N}$, let $f^{(k)} = f \circ s_k$. Note that $\mu_k \ll \mu$ follows from the assumption $\mu_1 \ll \mu$.

LEMMA 4.2. *Let* $\phi(\theta) = e^{\theta k}$ *for some* $k \in \mathbb{N}$. *Then for any* ω *and* $\omega' \in \Omega$ *and* $m, m' \in (0, \infty)$,

$$\frac{h_*}{h^*} E_\theta|\tau_\omega - \tau_{\omega'}| \leq (e^{dk} - e^{ck})(P_c + P_d)\big[|\cdot| > m \text{ or } f^{(k)} > m'\big] \tag{4.1}$$
$$+ e^{(d-c)m}(2e^{dk} - e^{ck} + m')\|P_\omega - P_{\omega'}\|,$$

with $f = d\mu_k/d\mu$.

PROOF. Since τ_ω and $\tau_{\omega'} \in (e^{ck}, e^{dk})$, by C2 it follows that

$$\frac{h_*}{h^*} E_\theta|\tau_\omega - \tau_{\omega'}|\big[|\cdot| > m \text{ or } f^{(k)} > m'\big]$$

is bounded by the first term in the RHS of (4.1). To bound the expectation over the other region first observe that

$$\tau_\omega(x) = \int e^{\theta k} p_\theta(x)\, d\omega / p_\omega(x) = p_\omega^{(k)}(x)/p_\omega(x)$$

for any $\omega \in \Omega$ since $e^{\theta k} p_\theta(x) = p_\theta(x + k) = p_\theta^{(k)}(x)$. Lemma 4.1 then applies to yield

$$p_\omega|\tau_\omega - \tau_{\omega'}| \leq (2e^{dk} - e^{ck})|p_\omega - p_{\omega'}| + |p_\omega^{(k)} - p_{\omega'}^{(k)}|. \tag{4.2}$$

Since $(h_*/h^*)p_\theta \leq e^{(d-c)|\cdot|}p_\omega$ follows from (2.5), (4.2) shows that $(h_*/h^*)E_\theta|\tau_\omega - \tau_{\omega'}|[|\cdot| \leq m,\ f^{(k)} \leq m']$ is bounded by

$$e^{(d-c)m}\left\{(2e^{dk} - e^{ck})\|P_\omega - P_{\omega'}\| + \int_{f^{(k)} \leq m'} |p_\omega^{(k)} - p_{\omega'}^{(k)}|\, d\mu\right\}.$$

This completes the proof because, by the transformation theorem, the above integral wrt μ is

$$\int_{f \leq m'} |p_\omega - p_{\omega'}|\, d\mu_k = \int_{f \leq m'} |p_\omega - p_{\omega'}| f\, d\mu \leq m'\|P_\omega - P_{\omega'}\|. \qquad \square$$

LEMMA 4.3. *If* $\operatorname{supp}\Lambda = \Omega$, *then*

$$\bigvee_{\alpha \le n} \mathbf{E}\|P_{\omega_{\alpha n}} - P_{G_n}\| \to 0 \quad \textit{uniformly in } \underline{\theta} \textit{ as } n \to \infty.$$

PROOF. Replace n by $n' = n + 1$ in Section 3.1. Denote $g_{n'n'}$, $\Lambda_{n'n'}$ and $\omega_{n'n'}$ (of the second paragraph of 3.1) by g, $\hat{\Lambda}$ and $\hat{\omega}$, respectively. [$\hat{\omega}$ can be shown to be the posterior distribution of θ_{n+1} given the data $\underline{X} = (X_1, \dots, X_n)$ under the prior $\bar{\omega}_{\Lambda}^{n+1}$ on $(\theta_1, \dots, \theta_n, \theta_{n+1})$.] We will use $\hat{\omega}_{(n)}$, if necessary, to exhibit the number of arguments.

Fix $\underline{\theta} \in \Theta^n$ and $1 \le \alpha \le n$. Let $G_{n\alpha}$ denote the empirical distribution based on $\underline{\theta}_{\alpha}^{\vee}$. Then we have the following:

(i) $G_{n\alpha}$ is G_{n-1} corresponding to $\underline{\theta}_{\alpha}^{\vee} \in \Theta^{n-1}$.

(ii) Letting F: $\mathbb{R}^{n-1} \to \Omega$ be the function such that $F(\underline{X}_{n-1}) = \hat{\omega}_{(n-1)}$, it follows from the definition of $\omega_{n\alpha}$ that $\omega_{n\alpha} = F(\underline{X}_{\alpha}^{\vee})$.

(iii) Clearly $\mathbf{P}_{\underline{\theta}} \underline{X}_{\alpha}^{\vee -1} = \mathbf{P}_{\underline{\theta}_{\alpha}^{\vee}} \underline{X}_{n-1}^{-1}$.

From (i), (ii) and (iii), we get that

$$\mathbf{P}_{\underline{\theta}} \|P_{\omega_{n\alpha}} - P_{G_{n\alpha}}\|^{-1} = \mathbf{P}_{\underline{\theta}_{\alpha}^{\vee}} \|P_{\hat{\omega}_{(n-1)}} - P_{G_{n-1}}\|^{-1}.$$

Since $\underline{\theta}$ and α are arbitrary

$$\bigvee_{\underline{\theta} \in \Theta^n} \bigvee_{\alpha \le n} \mathbf{E}\|P_{\omega_{n\alpha}} - P_{G_{n\alpha}}\| \le \bigvee_{\underline{\theta} \in \Theta^{n-1}} \mathbf{E}\|P_{\hat{\omega}_{(n-1)}} - P_{G_{n-1}}\| \to 0$$

as $n \to \infty$, by Theorem 3.1 and Example 5.3 [with $k = 1$ and $T_1(x) = x$], both from Datta (1991).

Next, because $G_n - G_{n\alpha} = n^{-1}(\delta_{\theta_\alpha} - G_{n\alpha})$ with δ_{θ_α} the distribution degenerate at θ_α, the variation norm of $G_n - G_{n\alpha}$ is no more than $2n^{-1}$. Thus, by definition of p_ω and by (2.3),

$$|p_{G_n} - p_{G_{n\alpha}}| \le 2n^{-1} \bigvee_{\theta} p_\theta \le 2n^{-1} \frac{h^*}{h_*}(p_c + p_d).$$

Consequently

$$\|P_{G_n} - P_{G_{n\alpha}}\| \le 2n^{-1} \int \Big(\bigvee_{\theta} p_\theta(x) \Big)\, d\mu \le (4h^*/h_*) n^{-1}.$$

The proof now ends by the triangle inequality. □

PROOF OF THE THEOREM. (a) The second part of the assertion follows from (3.7).

For the first part, first consider the special case $\phi(\theta) = e^{\theta k}$, $k \in \mathbb{N}$. Clearly,

$$\begin{aligned} \mathbf{E}|\hat{t}_\alpha - \tilde{t}_\alpha| &= \mathbf{E}_{\underline{\theta}_{\alpha}^{\vee}} E_{\theta_\alpha} |\tau_{\omega_{n\alpha}} - \tau_{G_n}|, \\ &\le \frac{h^*}{h_*} \big\{ (e^{dk} - e^{ck})(P_c + P_d)\big[|\cdot| > m \text{ or } f^{(k)} > m' \big] \\ &\qquad + e^{(d-c)m}(2e^{dk} - e^{ck} + m')\mathbf{E}\|P_{\omega_{n\alpha}} - P_{G_n}\| \big\} \end{aligned}$$

by Lemma 4.2, where $m, m' < \infty$ are arbitrary. For each m and m', the second term of the above bound is $o(1)$ uniformly in α and $\underline{\theta}$ as $n \to \infty$ by Lemma 4.3. The first term is independent of α and $\underline{\theta}$ and can be made arbitrarily small by choosing m and m' large enough. This concludes the proof in the special case.

Next, more generally, let $\phi(\theta) = \sum_k a_k e^{\theta k}$ be a polynomial in e^{θ}. By definition and the linearity property of conditional expectation (or integral) it follows that

$$\hat{t}_\alpha = \sum_k a_k \hat{t}_\alpha^{[k]}, \qquad \tilde{t}_\alpha = \sum_k a_k \tilde{t}_\alpha^{[k]},$$

where $\hat{t}_\alpha^{[k]}$ and $\tilde{t}_\alpha^{[k]}$ are the corresponding Bayes estimators of $e^{\theta_\alpha k}$ for each k. Hence (3.8) holds since it holds with $\hat{\underline{t}}^{[k]}$ and $\tilde{\underline{t}}^{[k]}$ for each k by the previous case.

Finally for general continuous ϕ, given $\varepsilon > 0$, choose a polynomial p such that $\bigvee_\theta |\phi(\theta) - p(e^\theta)| < \varepsilon$. Then, using definitions and taking absolute values under integrals,

$$\left|\hat{t}_\alpha - \hat{t}_\alpha^{[p]}\right| < \varepsilon, \qquad \left|\tilde{t}_\alpha - \tilde{t}_\alpha^{[p]}\right| < \varepsilon$$

where $\hat{t}_\alpha^{[p]}$ and $\tilde{t}_\alpha^{[p]}$ are the corresponding Bayes estimators of $p(e^{\theta_\alpha})$ and so

$$|\hat{t}_\alpha - \tilde{t}_\alpha| \le |\hat{t}_\alpha^{[p]} - \tilde{t}_\alpha^{[p]}| + 2\varepsilon.$$

The proof is now complete by the previous case, ε being arbitrary.

(b) Let $\tilde{t}_{\alpha n}(\underline{x}_n) = \tau_{G_n}(x_\alpha)$ and $\tilde{\underline{t}} = (\tilde{t}_{1n}, \dots, \tilde{t}_{nn})$ for $1 \le \alpha \le n < \infty$ and $\tilde{\underline{t}}' = (\tilde{t}_{11}, \dots, \tilde{t}_{nn})$. Now $R_n(\tilde{\underline{t}}', \cdot\,) \le R_n(\tilde{\underline{t}}, \cdot\,)$, since its generalization [cf. inequality (8.8.) of Hannan (1957)] holds without restriction and hence $D_n(\hat{\underline{t}}, \underline{\theta})_+$ is bounded by RHS (3.7) with $\tilde{t}_\alpha$ replaced by $\tilde{t}_{\alpha\alpha}$. But since Λ has full support this bound is $o(1)$ uniformly in $\underline{\theta}$, because $\bigvee_{\underline{\theta}} \mathbf{E}|\hat{t}_n - \tilde{t}_{nn}|$ is $o(1)$ by part (a) and the fact that the limit of a convergent sequence equals its Cesaro sum. □

REMARK 4.1. In fact, in part (b) above, $\bigvee_{\underline{\theta}} |D_n(\hat{\underline{t}}, \underline{\theta})| = o(1)$. To prove this, note that a slight extension of (2.5) of Gilliland (1968) gives

$$\left|D_n(\hat{\underline{t}}, \underline{\theta})\right| \le 2 \operatorname{diam} \phi[\Theta] n^{-1} \sum_{\alpha=1}^{n} \mathbf{E}\left|\hat{t}_\alpha - \tilde{t}_{\alpha\alpha}\right| + O(n^{-1}\log n),$$

where $O(n^{-1}\log n)$ is uniform in $\underline{\theta}$. But the above proof has shown that the first term is $o(1)$ uniformly in $\underline{\theta}$.

REMARK 4.2. An interesting question seems to be how far we can relax the compactness assumption on the component parameter space. It is known that we cannot always go up to the natural parameter space. An example where no a.o. compound estimator exists is the Poisson family with unbounded parameter space. See Gilliland [(1968), Section 3.3] for a proof.

APPENDIX

Admissibility of Bayes compound estimators under squared error loss. Let $\{P_\theta\}$, $\theta \in \Theta$ be the component distributions. Consider the compound problem of estimating $\underline{\phi}$ under squared error loss $L(\underline{t}, \underline{\theta}) = n^{-1}\Sigma_{\alpha=1}^{n}(t_\alpha - \phi(\theta_\alpha))^2$ for any function ϕ on Θ. Let $\mathbf{P}_{\underline{\theta}} = \times_{\alpha=1}^{n} P_{\theta_\alpha}$ for $\underline{\theta} \in \Theta^n$. Let ζ be a prior on $\underline{\theta}$. Denote the joint distribution $\zeta \circ P_{\underline{\theta}}$ on $\langle \underline{x}, \underline{\theta} \rangle$ by Q. Then the marginal of $\underline{x}$ is $Q\underline{x}^{-1} = \int P_{\underline{\theta}}\, d\zeta$. For a function f on Θ^n let $Q_{\underline{x}} f(\underline{\theta})$ denote the class of conditional expectations of $f(\underline{\theta})$ given $\underline{x}$.

LEMMA A. *If ζ is such that $\mathbf{P}_{\underline{\theta}} \ll Q\underline{x}^{-1}$ $\forall$ $\underline{\theta} \in \Theta^n$, then every Bayes estimator versus ζ is admissible.*

PROOF. First consider the set compound case.

Fix an $\alpha \in \{1, \dots, n\}$. Then $Q(t_\alpha - \theta_\alpha)^2$ is minimal iff $t_\alpha(\underline{x}) \in Q_{\underline{x}}\phi(\theta_\alpha)$. Hence t_α is determined up to $Q\underline{x}^{-1}$ null sets and so by the assumption of the lemma has unique risk $\underline{\theta} \rightsquigarrow \int (Q_{\underline{x}}\phi(\theta_\alpha) - \phi(\theta_\alpha))^2\, d\mathbf{P}_{\underline{\theta}}$. Thus, since $\alpha \in \{1, \dots n\}$ is arbitrary, the compound Bayes estimators have the unique compound risk $\underline{\theta} \rightsquigarrow n^{-1}\Sigma_{\alpha=1}^{n} \int (Q_{\underline{x}}\phi(\theta_\alpha) - \phi(\theta_\alpha))^2\, d\mathbf{P}_{\underline{\theta}}$ and hence are admissible.

For the sequence compound case, the given condition implies that for each $\alpha \in \{1, \dots, n\}$, $\mathbf{P}_{\underline{\theta}_\alpha} \ll Q\underline{x}_\alpha^{-1}$ $\forall$ $\underline{\theta}_\alpha \in \Theta^\alpha$. Hence, by combining the intermediate results in set case with $n = \alpha$ for each α, we get that the sequence compound Bayes estimators have the unique compound risk $\underline{\theta} \rightsquigarrow n^{-1}\Sigma_{\alpha=1}^{n} \int (Q_{\underline{x}_\alpha}\phi(\theta_\alpha) - \phi(\theta_\alpha))^2\, d\mathbf{P}_{\underline{\theta}_\alpha}$ and hence are admissible. □

Acknowledgment. The present work is part of author's dissertation at Michigan State University. The author wishes to thank Professor James Hannan for his guidance throughout this research.

REFERENCES

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

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.

DATTA, S. (1991). On the consistency of posterior mixtures and its applications. *Ann. Statist.* **19** 338–353.

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.

FERGUSON, T. S. (1974). Prior distributions on space of probability measures. *Ann. Statist.* **2** 615–629.

GILLILAND, D. C. (1968). Sequential compound estimation. *Ann. Math. Statist.* **39** 1890–1904.

GILLILAND, D. C. and HANNAN, J. (1974). The finite state compound decision problem, equivariance and restricted risk components. Report 317, Dept. Statist. and Probab., Michigan State Univ.

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, Haywood, 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.

INGLIS, J. (1973). Admissible decision rules for the compound problem. Ph.D. dissertation, Dept. Statist., Stanford Univ.

INGLIS, J. (1979). Admissible decision rules for the compound decision problem: The two action two state case. *Ann. Statist.* **7** 1127–1135.

KUO, L. (1986). A note on Bayes empirical Bayes estimation by means of Dirichlet process. *Statist. Probab. Lett.* **4** 145–150.

LEHMANN, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. Wiley, 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.

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

DEPARTMENT OF STATISTICS
UNIVERSITY OF GEORGIA
ATHENS, GEORGIA 30602