The Annals of Statistics
1988, Vol. 16, No. 3, 1270–1282

ON EMPIRICAL BAYES TESTING WITH SEQUENTIAL COMPONENTS

BY ROHANA J. KARUNAMUNI

The University of Alberta

We study the empirical Bayes decision theory with an m-truncated sequential statistical decision problem as the component. An empirical Bayes sequential decision procedure is constructed for the linear loss two-action problem. Asymptotic results are presented regarding the convergence of the Bayes risk of the empirical Bayes sequential decision procedure. With sequential components, an empirical Bayes sequential decision procedure selects both a stopping rule function and a terminal decision rule function for use in the component with parameter θ.

1. Introduction. Empirical Bayes decision theory [introduced by Robbins (1956) and later developed by Robbins (1963, 1964), Johns (1957), Samuel (1963) and Johns and Van Ryzin (1971, 1972), among others] deals with a sequence of independent repetitions of a given statistical decision problem, called the component problem, where each problem in the sequence has the same unknown prior distribution G. The components to which empirical Bayes methods have been applied are, with few exceptions, the fixed sample size identical statistical decision problems. Exceptions are the varying (nonstochastic) sample size components considered by O'Bryan (1972, 1976), O'Bryan and Susarla (1977) and Susarla and O'Bryan (1975). Another exception is found in the works of Liappala (1979, 1985). In his case the varying sample sizes are random. In this paper we consider empirical Bayes decision theory with a sequential statistical decision problem as the component, and study the linear loss two action problem for a very general class of densities.

In Section 2 we introduce notation to describe the sequential component and in Section 3 we discuss the two action problem and the sequential component problem to be studied in this paper. In Section 4 we define the empirical Bayes problem and construct an empirical Bayes sequential procedure. Asymptotic results and examples are presented in Section 5.

2. Notation. Let the parameter space be the measurable space $(\Omega, \mathscr{A})$ and let $\mathscr{G}$ denote the class of all prior distributions on Ω. Let $X_1, X_2, \ldots$ be i.i.d. P_θ, $\theta \in \Omega$, where P_θ is a probability distribution on $(\mathscr{X}, \mathscr{B})$, $\mathscr{X}$ is the real line and $\mathscr{B}$ is the Borel σ-field. For $k = 1, \ldots, m$ (m is a positive integer), we write $\mathbf{x}^k = (x_1, \ldots, x_k)$, $P_\theta^k = P_\theta \times \cdots \times P_\theta$ (k terms) and let $\mathscr{B}^k$ denote the Borel σ-field in $\mathscr{X}^k = \mathscr{X} \times \cdots \times \mathscr{X}$ (k terms). Suppose that the component problem has (terminal) action space A and loss function $L \geq 0$ defined on $\Omega \times A$. Let the constant

Received January 1986; revised November 1987.

AMS 1980 *subject classifications*. Primary 62C12; secondary 62H15.

Key words and phrases. Empirical Bayes, sequential decision procedures, asymptotic optimality, asymptotic superiority.

$c \geq 0$ denote the cost per observation. For $k = 0, 1, \ldots, m$ let $\mathscr{D}^k$ denote a set of mappings δ from $\mathscr{X}^m$ into A that are constant with respect to the last $m - k$ coordinates and such that $L(\theta, \delta)$ is $\mathscr{A} \times \mathscr{B}^m$ measurable. $\mathscr{D}^0$ consists of constant functions. We will regard the domain of $\delta \in \mathscr{D}^k$ as $\mathscr{X}^k$ when it is convenient to do so, $k = 1, 2, \ldots, m$. For $k \geq 1$, let G_k denote the posterior distribution of θ given $\mathbf{x}^k$ when $\theta \sim G$, and $G \in \mathscr{G}$. Throughout this paper, we let $[E]$ denote the indicator function of a set E.

3. The two action problem. Let the parameter space Ω be a subset of the real line and let $f_\theta \geq 0$ be a density function of the distribution P_θ with respect to a given σ-finite measure μ on $(\mathscr{X}, \mathscr{B})$. To conserve the notation we will also let $f_\theta(\mathbf{x}^k)$ denote the product $f_\theta(x_1), \ldots, f_\theta(x_k)$ for $\mathbf{x}^k \in \mathscr{X}^k$, $k \geq 1$. We wish to test the hypothesis

$$H_0\colon \theta \leq \theta_0 \text{ against } H_1\colon \theta > \theta_0,$$

where $\theta_0 \in \Omega$. Consequently, the action space A consists of two actions only, that is, $A = \{a_0, a_1\}$, where a_0 and a_1 denote the actions of deciding H_0 and H_1, respectively. We assume the linear loss function

$$L(\theta, a_0) = (\theta - \theta_0)^+, \qquad L(\theta, a_1) = (\theta_0 - \theta)^+, \qquad \theta \in \Omega, \tag{3.1}$$

where $b^+ = \max\{b, 0\}$ for any real number b. We assume that the first moment of θ is finite with respect to G, where G is the prior distribution of θ and $G \in \mathscr{G}$.

Let us now derive our sequential component consisting of a terminal decision rule $\boldsymbol{\delta}(G)$ and a stopping rule $\tau(G)$ with respect to G for our testing problem. The terminal decision rule $\boldsymbol{\delta}(G)$ is defined by a finite sequence

$$(\delta_1, \ldots, \delta_m), \tag{3.2}$$

where $\delta_k \in \mathscr{D}^k$, $k = 1, 2, \ldots, m$, is a Bayes decision rule relative to G for the fixed sample size k testing problem (Ω, A, L) with $A = \{a_0, a_1\}$ and the loss function (3.1). For $k \geq 1$, δ_k can be determined as follows. Let $\delta_k(\mathbf{x}^k) =$ pr$\{$choosing $a_0 | \mathbf{X}^k = \mathbf{x}^k\}$. Then it is easy to show that a Bayes rule is provided by the nonrandomized rule [see Johns and Van Ryzin (1972)]

$$\delta_k(\mathbf{x}^k) = \begin{cases} 1, & \text{if } \alpha_k(\mathbf{x}^k) \leq 0, \\ 0, & \text{if } \alpha_k(\mathbf{x}^k) > 0, \end{cases} \tag{3.3}$$

where

$$\alpha_k(\mathbf{x}^k) = \int_\Omega (\theta - \theta_0) f_\theta(\mathbf{x}^k) G(d\theta). \tag{3.4}$$

Then it is easy to see that

$$E^* r(G_{k+1}) + c - r(G_k) = \begin{cases} \rho_k(\mathbf{x}^k)/f_k(\mathbf{x}^k), & \text{if } f_k(\mathbf{x}^k) > 0, \\ 0, & \text{if } f_k(\mathbf{x}^k) = 0, \end{cases}$$

where $r(G_k)$ denotes the minimum posterior Bayes risk with respect to G_k, E^*

denotes conditional expectation on X_{k+1} given $\mathbf{x}^k$,

$$f_k(\mathbf{x}^k) = \int_\Omega f_\theta(\mathbf{x}^k) G(d\theta) \tag{3.5}$$

and

$$\rho_k(\mathbf{x}^k) = \int_{\mathscr{X}} [\alpha_{k+1} \le 0] \alpha_{k+1} \mu(dx_{k+1}) + c f_k(\mathbf{x}^k) - [\alpha_k \le 0] \alpha_k(\mathbf{x}^k). \tag{3.6}$$

The stopping rule $\tau(G)$ is defined by a finite sequence $\tau(G) = (\tau_0, \tau_1, \dots, \tau_m)$, where $\tau_0 = 0$, $\tau_m = 1$ and, for $k = 1, \dots, m-1$, τ_k: $\mathscr{X}^k \to \{0, 1\}$ such that

$$\tau_k(\mathbf{x}^k) = \begin{cases} 1, & \text{if } \rho_k(\mathbf{x}^k) \ge 0, \\ 0, & \text{if } \rho_k(\mathbf{x}^k) < 0. \end{cases} \tag{3.7}$$

Let N be the stopping time variable of the procedure $(\tau(G), \boldsymbol{\delta}(G))$. Then $N(\mathbf{X}^m) = \min\{k | \tau_k(\mathbf{X}^k) = 1\}$ and $1 \le N \le m$. (In the empirical Bayes application, this ensures that at least one observation is made at each repetition of the component which allows for an updating of the empirical Bayes estimates.) These kinds of sequential procedures are described in detail in various books concerning decision theory, e.g., Berger (1985), Chapter 7. The risk of this procedure at G is

$$R(G) = \int_\Omega R(\theta, (\tau, \boldsymbol{\delta})) G(d\theta),$$

where

$$R(\theta, (\tau, \boldsymbol{\delta})) = \int_{\mathscr{X}^m} \sum_{k=1}^m [N = k] \{L(\theta, \delta_k) + ck\} P_\theta^m(d\mathbf{x}^m).$$

Now using the definitions of L and δ_k [see (3.1) and (3.3)], we have

$$\begin{aligned} R(G) = C_G + \sum_{k=1}^m \int_\Omega \int_{\mathscr{X}^m} &[N = k]([\alpha_k \le 0](\theta - \theta_0) + ck) \\ &\times f_\theta(\mathbf{x}^m) \mu^m(d\mathbf{x}^m) G(d\theta), \end{aligned} \tag{3.8}$$

where $C_G = \int L(\theta, a_1) G(d\theta)$. The empirical Bayes approach applied to this problem can be based upon estimation of the functions $[N = k]$ and α_k, $k = 1, 2, \dots, m$. For this purpose, it is useful to decompose and represent the indicator functions as follows. For $k = 1, \dots, m-1$, we write $[N = k] = A_k + B_k$ and $[N = m] = A_m$, where

$$\begin{aligned} A_k &= [\rho_1 < 0] \cdots [\rho_{k-1} < 0][\rho_k > 0] \quad \text{for } k = 1, \dots, m-1, \\ B_k &= [\rho_1 < 0] \cdots [\rho_{k-1} < 0][\rho_k = 0] \quad \text{for } k = 1, \dots, m-1, \\ A_m &= [\rho_1 < 0] \cdots [\rho_{m-1} < 0]. \end{aligned} \tag{3.9}$$

Thus, the Bayes risk of the sequential component $(\tau(G), \boldsymbol{\delta}(G))$ relative to G for our testing problem can be written in the form [see (3.8)]

$$\begin{aligned} R(G) = C_G &+ \sum_{k=1}^{m} \int_{\Omega}\int_{\mathscr{X}^m} A_k\{[\alpha_k \le 0](\theta - \theta_0) + ck\} f_\theta(\mathbf{x}^m)\mu^m(d\mathbf{x}^m)G(d\theta) \\ &+ \sum_{k=1}^{m-1} \int_{\Omega}\int_{\mathscr{X}^m} B_k\{[\alpha_k \le 0](\theta - \theta_0) + ck\} f_\theta(\mathbf{x}^m)\mu^m(d\mathbf{x}^m)G(d\theta). \end{aligned} \tag{3.10}$$

The indicators corresponding to boundary sets are defined by B_k in (3.9). In the empirical Bayes application, the functions ρ_k are estimated, so that a separate treatment of boundaries is important insofar as convergences are concerned.

4. Empirical Bayes problem. Suppose that the prior distribution G is unknown but fixed. Then the classical Bayes quantities (3.3)–(3.7) of Section 3 are not available to the statistician. However, suppose we are experiencing independent repetitions of the same component problem. Then applying the empirical Bayes approach introduced by Robbins (1956), we derive empirical Bayes estimates of the classical Bayes quantities (3.3)–(3.7) and, hence, obtain an empirical Bayes sequential decision procedure $\mathbf{d}^n = (\tau^n, \boldsymbol{\delta}^n)$, where τ^n is an empirical Bayes stopping rule and $\boldsymbol{\delta}^n$ is an empirical Bayes terminal decision rule.

At the nth stage of the repetitions, we will have observed the random vectors $\mathbf{X}_1^{N_1}, \dots, \mathbf{X}_{n-1}^{N_{n-1}}$ from the past $(n-1)$ repetitions of the sequential component described in Section 3, where $N_1, \dots, N_{n-1}$ are the respective stopping times of the past $(n-1)$ repetitions. In order to construct an empirical Bayes sequential decision procedure, let us suppose that $f_k(\mathbf{x}^k)$, $k \ge 1$, and $\alpha_k(\mathbf{x}^k)$, $k \ge 1$ [see (3.5) and (3.4)], can be estimated by functions

$$f_k^n(\mathbf{x}^k) = f_k^n(\mathbf{x}^k, \mathbf{X}_1^{N_1}, \dots, \mathbf{X}_{n-1}^{N_{n-1}}), \qquad k \ge 1,$$

and

$$\alpha_k^n(\mathbf{x}^k) = \alpha_k^n(\mathbf{x}^k, \mathbf{X}_1^{N_1}, \dots, \mathbf{X}_{n-1}^{N_{n-1}}), \qquad k \ge 1,$$

respectively such that a.e. $(\mu^k)\mathbf{x}^k$, $k \ge 1$,

$$f_k^n(\mathbf{x}^k) \to_P f_k(\mathbf{x}^k) \quad \text{as } n \to \infty,\ k \ge 1, \tag{4.1}$$

and

$$\alpha_k^n(\mathbf{x}^k) \to_P \alpha_k(\mathbf{x}^k) \quad \text{as } n \to \infty,\ k \ge 1, \tag{4.2}$$

where $\to_P$ denotes convergence in probability with respect to the sequence of random vectors $\{\mathbf{X}_1^{N_1}, \dots, \mathbf{X}_{n-1}^{N_{n-1}}, \dots\}$.

We now define our empirical Bayes sequential decision (EBSD) procedure $\mathbf{d}^n = (\tau^n, \boldsymbol{\delta}^n)$ as follows. (Henceforth we use the superscript n to indicate an empirical Bayes quantity.)

Let $\boldsymbol{\delta}^n$ be a finite sequence of functions $(\delta_1^n, \dots, \delta_m^n)$, where δ_k^n is such that $\delta_k^n(\mathbf{x}^k) = \mathrm{pr}\{\text{choosing } a_0 | \mathbf{X}^k = \mathbf{x}^k\}$ and, motivated by (3.3) and (3.4),

$$\delta_k^n(\mathbf{x}^k) = \begin{cases} 1, & \text{if } \alpha_k^n(\mathbf{x}^k) \le 0, \\ 0, & \text{if } \alpha_k^n(\mathbf{x}^k) > 0, \end{cases} \tag{4.3}$$

where $\alpha_k^n(\mathbf{x}^k)$ satisfies (4.2).

Let τ^n be a stopping rule consisting of a finite sequence of functions $(\tau_1^n, \dots, \tau_m^n)$, where, motivated by (3.7), $\tau_m^n = 1$ and, for $k = 1, \dots, m-1$,

$$\tau_k^n(\mathbf{x}^k) = \begin{cases} 1, & \text{if } \rho_k^n(\mathbf{x}^k) \ge 0, \\ 0, & \text{if } \rho_k^n(\mathbf{x}^k) < 0, \end{cases} \tag{4.4}$$

where

$$\begin{aligned} \rho_k^n(\mathbf{x}^k) = {} & \int_{\mathscr{X}} \delta_{k+1}^n(\mathbf{x}^{k+1}) \alpha_{k+1}^n(\mathbf{x}^{k+1}) \mu(dx_{k+1}) \\ & + c f_k^n(\mathbf{x}^k) - \delta_k^n(\mathbf{x}^k) \alpha_k^n(\mathbf{x}^k), \end{aligned} \tag{4.5}$$

and $f_k^n(\mathbf{x}^k)$ and $\alpha_k^n(\mathbf{x}^k)$ satisfy (4.1) and (4.2), respectively. Since $\tau_m^n = 1$, sampling will be stopped just after X_m has been observed if it has not been stopped earlier. For investigating the risk of the EBSD procedure it is useful to define

$$\begin{aligned} C_k^n &= [\rho_1^n < 0] \cdots [\rho_{k-1}^n < 0][\rho_k^n \ge 0] \quad \text{for } k = 1, \dots, m-1, \\ C_m^n &= [\rho_1^n < 0] \cdots [\rho_{m-1}^n < 0]. \end{aligned} \tag{4.6}$$

Then $[N^n = k] = C_k^n$ for $k = 1, \dots, m-1$ and $[N^n = m] = C_m^n$, where N^n denotes the stopping time of the EBSD procedure $\mathbf{d}^n = (\tau^n, \boldsymbol{\delta}^n)$. Note that $\sum_{k=1}^m [N^n = k] = 1$ implies $\sum_{k=1}^m C_k^n = 1$. Let $R(G, \mathbf{d}^n)$ denote the conditional Bayes risk of $\mathbf{d}^n = (\tau^n, \boldsymbol{\delta}^n)$ with respect to G. Then since C_k^n partition $\mathscr{X}^m$,

$$\begin{aligned} R(G, \mathbf{d}^n) = C_G + \sum_{k=1}^m \int_\Omega \int_{\mathscr{X}^m} & C_k^n([\alpha_k^n \le 0](\theta - \theta_0) + ck) \\ & \times f_\theta(\mathbf{x}^m) \mu^m(d\mathbf{x}^m) G(d\theta). \end{aligned} \tag{4.7}$$

The difference $ER(G, \mathbf{d}^n) - R(G)$ is treated as a measure of optimality of the sequential procedure $\mathbf{d}^n = (\tau^n, \boldsymbol{\delta}^n)$ and this motivates the following definitions.

DEFINITION 4.1. A sequence of sequential decision procedures $\{(\tau^n, \boldsymbol{\delta}^n)\}$ is said to be asymptotically risk equivalent (optimal) relative to the (optimal) sequential procedure $(\tau(G), \boldsymbol{\delta}(G))$ if $\lim_{n \to \infty} ER(G, (\tau^n, \boldsymbol{\delta}^n)) = R(G)$, where E denotes expectation with respect to the random vectors $\mathbf{X}_1^{N_1}, \dots, \mathbf{X}_{n-1}^{N_{n-1}}$.

DEFINITION 4.2. A sequence of sequential decision procedures $\{(\tau^n, \boldsymbol{\delta}^n)\}$ is said to be asymptotically superior relative to $(\tau(G), \boldsymbol{\delta}(G))$ if $\limsup_{n \to \infty} ER(G, (\tau^n, \boldsymbol{\delta}^n)) \le R(G)$.

5. Asymptotic results. In this section we compare the asymptotic behaviour of the unconditional Bayes risk $ER(G, \mathbf{d}^n)$ of the empirical Bayes sequential decision procedure $\mathbf{d}^n = (\tau^n, \boldsymbol{\delta}^n)$, defined in the previous section, with the Bayes risk $R(G)$ [see (3.10)] defined in Section 3. Before presenting the main results, we first state the following useful lemmas. Convergence of sequences of functions on $\mathscr{X}^m$ is understood to be pointwise convergence. The proofs of the lemmas are omitted. For convenience, the notations $\mathscr{X}^k$, $k \ge 1$, and Ω under the integral signs that follow are suppressed in all future discussions.

LEMMA 5.1. *Assume that* $\int |\theta| G(d\theta) < \infty$. *For* $k \ge 1$, *let the functions* $f_k^n(\mathbf{x}^k)$ *and* $\alpha_k^n(\mathbf{x}^k)$ *be defined by* (4.1) *and* (4.2), *respectively. If*

$$\int [\alpha_{k+1}^n \le 0] \alpha_{k+1}^n \mu(dx_{k+1}) \rightarrow_P \int [\alpha_{k+1} \le 0] \alpha_{k+1} \mu(dx_{k+1}) \quad \text{as } n \to \infty,\ k \ge 1, \tag{5.1}$$

then

$$\rho_k^n \rightarrow_P \rho_k \quad \text{as } n \to \infty,\ k \ge 1, \tag{5.2}$$

where ρ_k *and* ρ_k^n *are given by* (3.6) *and* (4.5), *respectively.*

LEMMA 5.2. *Let*

$$J_1^n = \sum_{j=2}^{m} \sum_{i=1}^{j-1} \int\int EC_i^n A_j \{ ([\alpha_i^n \le 0] - [\alpha_j \le 0])(\theta - \theta_0) + c(i-j) \} \times f_\theta(\mathbf{x}^m) \mu^m(d\mathbf{x}^m) G(d\theta),$$

$$J_2^n = \sum_{j=1}^{m-1} \sum_{i=j+1}^{m-1} \int\int EC_i^n A_j \{ ([\alpha_i^n \le 0] - [\alpha_j \le 0])(\theta - \theta_0) + c(i-j) \} \times f_\theta(\mathbf{x}^m) \mu^m(d\mathbf{x}^m) G(d\theta),$$

and

$$J_3^n = \sum_{i=1}^{m} \int\int EC_i^n A_i \{ ([\alpha_i^n \le 0] - [\alpha_i \le 0])(\theta - \theta_0) \} \times f_\theta(\mathbf{x}^m) \mu^m(d\mathbf{x}^m) G(d\theta).$$

If $\int |\theta| G(d\theta) < \infty$, *then* $\lim_{n \to \infty} J_i^n = 0$, $i = 1, 2, 3$.

LEMMA 5.3. *Let*

$$K_1^n = \sum_{j=2}^{m-1} \sum_{i=1}^{j-1} \int\int EC_i^n B_j \{ ([\alpha_i^n \le 0] - [\alpha_j \le 0])(\theta - \theta_0) + c(i-j) \} \times f_\theta(\mathbf{x}^m) \mu^m(d\mathbf{x}^m) G(d\theta)$$

and

$$K_2^n = \sum_{i=1}^{m-1} \int\int EC_i^n B_i \{ ([\alpha_i^n \le 0] - [\alpha_i \le 0])(\theta - \theta_0) \} f_\theta(\mathbf{x}^m) \mu^m(d\mathbf{x}^m) G(d\theta).$$

If $\int |\theta| G(d\theta) < \infty$, *then* $\lim_{n \to \infty} K_i^n = 0$, $i = 1, 2, 3$.

The following theorems give the main results of this paper concerning the asymptotic risk equivalence and the asymptotic superiority of the empirical Bayes sequential decision procedure $\mathbf{d}^n = (\tau^n, \delta^n)$ defined in the previous section. The conditions under which the theorems are stated will be discussed in the examples at the end of this section.

THEOREM 5.1. *Let* $\mathbf{d}^n = (\tau^n, \delta^n)$ *be defined by* (4.3), (4.4) *and* (4.5). *For* $k \geq 1$, *let* $f_k^n(\mathbf{x}^k)$ *and* $\alpha_k^n(\mathbf{x}^k)$ *be defined by* (4.1) *and* (4.2), *respectively, and, further, assume* $\alpha_k^n(\mathbf{x}^k)$ *satisfies* (5.1). *Let G be such that* $\int |\theta| G(d\theta) < \infty$. *Then* $\limsup_{n \to \infty} ER(G, \mathbf{d}^n) \leq R(G)$.

PROOF. By Fubini's theorem and $\int |\theta| G(d\theta) < \infty$, the difference between $ER(G, \mathbf{d}^n)$ and $R(G)$ can be written as

$$ER(G, \mathbf{d}^n) - R(G) = \sum_{i=1}^{3} J_i^n + \sum_{i=1}^{3} K_i^n, \tag{5.3}$$

where J_1^n, J_2^n, J_3^n and K_1^n, K_2^n are as defined in Lemmas 5.2 and 5.3, respectively, and

$$K_3^n = \sum_{j=1}^{m-1} \sum_{i=j+1}^{m} \int\int E\hat{C}_i^n B_j \{ ([\alpha_i^n \leq 0] - [\alpha_j \leq 0])(\theta - \theta_0) + c(i-j) \}$$
$$\times f_\theta(\mathbf{x}^m) G(d\theta) \mu^m(d\mathbf{x}^m).$$

We define

$$\hat{C}_k^n = \sum_{i=k}^{m} C_i^n \quad \text{for } k = j+1, \dots, m, \ j = 1, \dots, m-1.$$

Then observe that

$$C_k^n = \hat{C}_k^n - \hat{C}_{k+1}^n \quad \text{for } k = j+1, \dots, m-1, \ j = 1, \dots, m-1,$$
$$\hat{C}_k^n = [\rho_1^n < 0] \cdots [\rho_{k-1}^n < 0]$$

and $\hat{C}_k^n$ depends only on the first $k-1$ observations $(X_1, \dots, X_{k-1})$. Using these facts and the definition of $\hat{C}_k^n$ we can write

$$K_3^n = \sum_{j=1}^{m-1} \Bigg\{ \sum_{i=j+1}^{m} T_i(n, j) - \sum_{i=j+1}^{m-1} S_i(n, j)$$
$$+ \sum_{k=j+1}^{m} c \int E\hat{C}_k^n B_j f_{k-1}(\mathbf{x}^{k-1}) \mu^{k-1}(d\mathbf{x}^{k-1})$$
$$- \int E\hat{C}_{j+1}^n B_j [\alpha_j \leq 0] \alpha_j \mu^j(d\mathbf{x}^j) \Bigg\},$$

where

$$T_i(n, j) = \int E\hat{C}_i^n B_j[\alpha_i^n \le 0]\alpha_i \mu^i(d\mathbf{x}^i)$$

and

$$S_i(n, j) = \int E\hat{C}_{i+1}^n B_j[\alpha_i^n \le 0]\alpha_i \mu^i(d\mathbf{x}^i).$$

Routine calculations yield

$$K_3^n = \sum_{j=1}^{m-1} \{M_1(n, j) + M_2(n, j)\}, \tag{5.4}$$

where

$$\begin{aligned} M_1(n, j) = T_{j+1}(n, j) &- \int E\hat{C}_{j+1}^n B_j[\alpha_j \le 0]\alpha_j \mu^j(d\mathbf{x}^j) \\ &+ c\int E\hat{C}_{j+1}^n f_j(\mathbf{x}^j)\mu^j(d\mathbf{x}^j) \end{aligned} \tag{5.5}$$

and

$$\begin{aligned} M_2(n, j) = \sum_{i=j+2}^{m} \Big\{ & T_i(n, j) - S_{i-1}(n, j) \\ & + c\int E\hat{C}_i^n B_j f_{i-1}(\mathbf{x}^{i-1})\mu^{i-1}(d\mathbf{x}^{i-1})\Big\}. \end{aligned} \tag{5.6}$$

Now observe that $M_1(n, j)$ is equal to

$$\int E\hat{C}_{j+1}^n B_j\left\{\int[\alpha_{j+1}^n \le 0]\alpha_{j+1}\mu(dx_{j+1}) - [\alpha_j \le 0]\alpha_j + cf_j\right\}\mu^j(d\mathbf{x}^j). \tag{5.7}$$

But $0 \le B_j \le [\rho_j = 0]$ and by the definition of ρ_j [see (3.6)]

$$cf_j - [\alpha_j \le 0]\alpha_j = -\int[\alpha_{j+1} \le 0]\alpha_{j+1}\mu(dx_{j+1}) \quad \text{on } [\rho_j = 0].$$

Then from (5.7),

$$\begin{aligned} |M_1(n, j)| &\le \int E[\rho_j = 0]\left(\left|\int[\alpha_{j+1}^n \le 0]\alpha_{j+1}\mu(dx_{j+1})\right.\right. \\ &\qquad\qquad \left.\left. - \int[\alpha_{j+1} \le 0]\alpha_{j+1}\mu(dx_{j+1})\right|\right)\mu^j(d\mathbf{x}^j) \\ &\le \int E\left[|\alpha_{j+1}^n - \alpha_{j+1}| \ge |\alpha_{j+1}|\right]|\alpha_{j+1}|\mu^{j+1}(d\mathbf{x}^{j+1}). \end{aligned} \tag{5.8}$$

The r.h.s. of (5.8) goes to zero as $n \to \infty$, since $\alpha_{j+1}^n \to_P \alpha_{j+1}$ as $n \to \infty$ by applying the dominated convergence theorem (DCT). Thus $M_1(n, j) \to 0$ as $n \to \infty$.

Now it remains to consider $M_2(n, j)$ given by (5.6). Observe that the summand of $M_2(n, j)$ is equal to

$$\int E\hat{C}_i^n B_j [\alpha_i^n \le 0]\, \alpha_i \mu^i(d\mathbf{x}^i) - \int E\hat{C}_i^n B_j [\alpha_{i-1}^n \le 0]\, \alpha_{i-1}\mu^{i-1}(d\mathbf{x}^{i-1}) + \int E\hat{C}_i^n B_j c f_{i-1}(\mathbf{x}^{i-1})\mu^{i-1}(d\mathbf{x}^{i-1}),$$

$i = j + 2, \ldots, m$. The preceding sum can be written as

$$\int E\hat{C}_i^n B_j \left(\int [\alpha_i^n \le 0]\, \alpha_i \mu(dx_i) - [\alpha_{i-1}^n \le 0]\, \alpha_{i-1} + c f_{i-1} \right) \mu^{i-1}(d\mathbf{x}^{i-1}). \tag{5.9}$$

Adding and subtracting the term $[\alpha_{i-1} \le 0]\alpha_{i-1}$ in the integrand of the preceding integral (5.9), we get

$$\begin{aligned} &\int E\hat{C}_i^n B_j \left(\int [\alpha_i^n \le 0]\, \alpha_i \mu(dx_i) - [\alpha_{i-1} \le 0]\alpha_{i-1} + c f_{i-1} \right) \mu^{i-1}(d\mathbf{x}^{i-1}) \\ &\quad + \int E\hat{C}_i^n B_j ([\alpha_{i-1} \le 0]\alpha_{i-1} - [\alpha_{i-1}^n \le 0]\, \alpha_{i-1})\mu^{i-1}(d\mathbf{x}^{i-1}). \end{aligned} \tag{5.10}$$

Using the DCT we see that

$$\int E\hat{C}_i^n B_j ([\alpha_{i-1} \le 0]\alpha_{i-1} - [\alpha_{i-1}^n \le 0]\, \alpha_{i-1})\mu^{i-1}(d\mathbf{x}^{i-1})$$

goes to zero as $n \to \infty$. The first integral in (5.10) can be rewritten as

$$\begin{aligned} &\int E\hat{C}_i^n B_j [\rho_{i-1} \le 0] \Bigl(\int [\alpha_i^n \le 0]\, \alpha_i \mu(dx_i) \\ &\qquad\qquad - [\alpha_{i-1} \le 0]\alpha_{i-1} + c f_{i-1} \Bigr) \mu^{i-1}(d\mathbf{x}^{i-1}) \\ &\quad + \int E\hat{C}_i^n B_j [\rho_{i-1} > 0] \Bigl(\int [\alpha_i^n \le 0]\, \alpha_i \mu(dx_i) \\ &\qquad\qquad - [\alpha_{i-1} \le 0]\alpha_{i-1} + c f_{i-1} \Bigr) \mu^{i-1}(d\mathbf{x}^{i-1}). \end{aligned} \tag{5.11}$$

By an application of the DCT again and $0 \le \hat{C}_i^n [\rho_{i-1} > 0] \le [\rho_{i-1}^n < 0][\rho_{i-1} > 0]$, one can show that the second integral in (5.11) goes to zero as $n \to \infty$. Notice that the asymptotic behaviour of the first integral in (5.11) is equivalent to the asymptotic behaviour of the expression

$$\int E\hat{C}_i^n B_j [\rho_{i-1} \le 0] \left(\int [\alpha_i \le 0]\alpha_i \mu(dx_i) - [\alpha_{i-1} \le 0]\alpha_{i-1} + c f_{i-1} \right) \mu^{i-1}(d\mathbf{x}^{i-1}).$$

That is [see (3.6)],

$$\int E\hat{C}_i^n B_j [\rho_{i-1} \le 0]\rho_{i-1}\mu^{i-1}(d\mathbf{x}^{i-1}). \tag{5.12}$$

Expression (5.12) is nonpositive for all n. Therefore, $\limsup_{n \to \infty} ER(G, \mathbf{d}^n) \le$

$R(G)$ now follows from (5.3)–(5.12) and Lemmas 5.2 and 5.3. This completes the proof of Theorem 5.1. □

THEOREM 5.2. *Under the same assumptions as in Theorem* 5.1, *let*

$$\liminf_{n\to\infty} E[\rho_1^n < 0] \cdots [\rho_{i-1}^n < 0] = [\rho_1 < 0] \cdots [\rho_{i-1} < 0], \qquad i = 2, \dots, m. \tag{5.13}$$

Then $\lim_{n\to\infty} ER(G, \mathbf{d}^n) = R(G)$, *where* ρ_i *and* ρ_i^n *are defined by* (3.6) *and* (4.5), *respectively.*

PROOF. First observe that, by the definition of $\hat{C}_i^n$, we have

$$\begin{aligned}[\rho_1 < 0] \cdots [\rho_{i-1} < 0] &\le \liminf_{n\to\infty} E\hat{C}_i^n \\ &\le \limsup_{n\to\infty} E\hat{C}_i^n \le [\rho_1 \le 0] \cdots [\rho_{i-1} \le 0], \\ &\qquad i = 2, \dots, m.\end{aligned}$$

Since

$$E\hat{C}_i^n B_j [\rho_{i-1} \le 0] \rho_{i-1} \ge [\rho_{i-1} \le 0] \rho_{i-1}, \qquad i = j+2, \dots, m, \ j = 1, \dots, m-1,$$

and

$$\int |\rho_{i-1}| \mu^{i-1}(d\mathbf{x}^{i-1}) < \infty,$$

an application of Fatou's lemma yields

$$\begin{aligned}&\liminf_{n\to\infty} \int E\hat{C}_i^n B_j [\rho_{i-1} \le 0] \rho_{i-1} \mu^{i-1}(d\mathbf{x}^{i-1}) \\ &\qquad \ge \int \liminf_{n\to\infty} E\hat{C}_i^n B_j [\rho_{i-1} \le 0] \rho_{i-1} \mu^{i-1}(d\mathbf{x}^{i-1}).\end{aligned}$$

Now using the definition [see (3.9)] of B_j we obtain

$$\liminf_{n\to\infty} E\hat{C}_i^n B_j = 0 \quad \text{for } i = j+2, \dots, m, \ j = 1, \dots, m-1,$$

provided that

$$\liminf_{n\to\infty} E[\rho_1^n < 0] \cdots [\rho_{i-1}^n < 0] = [\rho_1 < 0] \cdots [\rho_{i-1} < 0], \qquad i = 2, \dots, m.$$

Thus, under the assumptions of Theorem 5.2, we have

$$\liminf_{n\to\infty} \int E\hat{C}_i^n B_j [\rho_{i-1} \le 0] \rho_{i-1} \mu^{i-1}(d\mathbf{x}^{i-1}) = 0, \qquad i = j+2, \dots, n, \ j = 1, \dots, m-1. \tag{5.14}$$

Hence $\liminf_{n\to\infty} ER(G, \mathbf{d}^n) \ge R(G)$ follows from (5.3)–(5.14) and Lemmas 5.2 and 5.3. Using Theorem 5.1 we then obtain $\lim_{n\to\infty} ER(G, \mathbf{d}^n) = R(G)$. □

For $i \geq 2$, by the definitions of N^n and N, we have $[N^n \geq i] = [\rho_1^n < 0] \cdots [\rho_{i-1}^n < 0]$ and $[N \geq i] = [\rho_1 < 0] \cdots [\rho_{i-1}]$. Therefore, condition (5.13) is equivalent to $\liminf_{n\to\infty} E[N^n \geq i] = [N \geq i]$, $i \geq 2$. This means that asymptotically the stopping time of the EBSD procedure is forced to behave like the stopping time used in the component. Also notice that for $i \geq 2$, $\liminf_{n\to\infty} E[N^n \geq i] \geq [N \geq i]$ holds.

COROLLARY 5.1. *If $m = 2$, then under the assumptions in Theorem 5.1, $\lim_{n\to\infty} ER(G, \mathbf{d}^n) = R(G)$; that is, $\mathbf{d}^n = (\tau^n, \delta^n)$ is asymptotically optimal relative to $(\tau(G), \delta(G))$.*

PROOF. If $m = 2$, then observe that $(\tau(G), \delta(G))$ is optimal and $R(G, \mathbf{d}^n) \geq R(G)$ for all n and G by the definition of the sequential component and the construction of the EBSD procedure $\mathbf{d}^n$. Hence $\lim_{n\to\infty} ER(G, \mathbf{d}^n) = R(G)$ follows from Theorem 5.1. □

Notice that for $m \geq 3$ the stopping rule used in the component of this paper is not the optimal stopping rule among the m-truncated procedures. Observe also that when the optimal strategy is used, $\limsup_{n\to\infty} ER(G, \mathbf{d}^n) \leq R(G)$. That is, $R(G)$ is attained from below, implying the asymptotic optimality of the empirical Bayes decision procedure. Therefore, in comparison with the results in the literature, Theorem 5.1 is not unusual, or at least not surprising.

EXAMPLE 1. Let $f_\theta(x) = \theta^{-1}[0 < x < \theta]$ for $\theta \in \Omega = (0, a]$, $0 < a < \infty$, and let G be a prior distribution on Ω. Let G_n be a sequence of distribution functions on Ω, where $G_n(\theta) = G_n(\theta, \mathbf{X}_1^{N_1}, \ldots, \mathbf{X}_{n-1}^{N_{n-1}})$ depends only on the random vectors $\mathbf{X}_1^{N_1}, \ldots, \mathbf{X}_{n-1}^{N_{n-1}}$, such that G_n converges to G in Lévy metric with probability 1. For this model, an explicit method of constructing such a sequence of distribution functions is given in Fox (1970).

The EBSD procedure $\mathbf{d}^n = (\tau^n, \delta^n)$ used in Theorems 5.1 and 5.2 is based on the functions $f_k^n(\mathbf{x}^k)$ and $\alpha_k^n(\mathbf{x}^k)$, $k \geq 1$. So we need only to find sequences of functions $\{f_k^n(\mathbf{x}^k)\}_{n\geq 1}$ and $\{\alpha_k^n(\mathbf{x}^k)\}_{n\geq 1}$ for this example.

For $k \geq 1$, we define

$$f_k^n(\mathbf{x}^k) = \int_\Omega f_\theta(\mathbf{x}^k) G_n(d\theta)$$

and

$$\alpha_k^n(\mathbf{x}^k) = \int_\Omega (\theta - \theta_0) f_\theta(\mathbf{x}^k) G_n(d\theta).$$

Then the EB estimators $f_k^n(\mathbf{x}^k)$ and $\alpha_k^n(\mathbf{x}^k)$ satisfy conditions (4.1) and (4.2), respectively. Also by an application of the generalized dominated convergence theorem, it is easy to verify condition (5.1) for these estimators.

EXAMPLE 2. Let $f_\theta(x) = e^{-\theta}\theta^x(x!)^{-1}$, $x = 0, 1, 2 \ldots$, for $\theta \in \Omega = (0, a]$, $a < \infty$. Let $m = 2$. Then observe the following facts about $f_k(\mathbf{x}^k)$, $k = 1, 2$, with

$\mathbf{x}^2 = (x_1, x_2)$ and $\mathbf{x}^1 = x_1$:

$$f_1(x_1) = (x_1!)^{-1}\int_0^a e^{-\theta}\theta^{x_1}G(d\theta), \qquad x_1 = 0, 1, 2\dots, \tag{5.15}$$

$$\begin{aligned} f_2(\mathbf{x}^2) = (x_1!x_2!)^{-1}\sum_{k=0}^{\infty}(k!)^{-1}(-1)^k \\ \times (x_1 + x_2 + k)!\,f_1(x_1 + x_2 + k). \end{aligned} \tag{5.16}$$

Recalling (3.4),

$$\alpha_k(\mathbf{x}^k) = \int_0^a(\theta - \theta_0)f_\theta(\mathbf{x}^k)G(d\theta), \qquad k = 1, 2,$$

we note that

$$\alpha_k(\mathbf{x}^k) = g_k(\mathbf{x}^k) - \theta_0 f_k(\mathbf{x}^k), \qquad k = 1, 2, \tag{5.17}$$

where $f_1(x_1)$ and $f_2(\mathbf{x}^2)$ are given by (5.15) and (5.16), respectively, and $g_1(x_1)$ and $g_2(\mathbf{x}^2)$ are given by expressions

$$g_1(x_1) = (x_1 + 1)f_1(x_1 + 1) \tag{5.18}$$

and

$$\begin{aligned} g_2(\mathbf{x}^2) = (x_1!x_2!)^{-1}\sum_{k=0}^{\infty}(k!)^{-1}(-1)^k \\ \times (x_1 + x_2 + k + 1)!\,f_1(x_1 + x_2 + k + 1). \end{aligned} \tag{5.19}$$

Notice that the first observations $X_{11}, X_{21}, \dots, X_{n1}$ from the past and present repetitions are i.i.d. with common marginal density

$$f(x) = (x!)^{-1}\int_0^a e^{-\theta}\theta^x G(d\theta).$$

Thus we define our EB estimator of $f(x)$ by

$$f^n(x) = n^{-1}\sum_{i=1}^{n} I(X_{i1} = x), \qquad x = 0, 1, 2, \dots.$$

Now EB estimators of $f_k(\mathbf{x}^k)$, $k = 1, 2$, and $g_k(\mathbf{x}^k)$, $k = 1, 2$, can be obtained by putting $f^n(x)$ in place of $f_1(x)$ and in the expressions (5.15)–(5.19), respectively, and it is easy to show that the resulting EB estimators $f_k^n(\mathbf{x}^k)$, $k = 1, 2$, and $\alpha_k^n(\mathbf{x}^k)$, $k = 1, 2$, satisfy (4.1) and (4.2), respectively, and that (5.1) holds for $\alpha_2^n(\mathbf{x}^2)$.

Acknowledgments. This paper is based on part of the author's Ph.D. dissertation written at Michigan State University. I would like to express my sincere gratitude to Professor Dennis C. Gilliland, my thesis advisor, for his help throughout this work. I am also grateful to a referee and an Associate Editor for useful suggestions.

REFERENCES

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

FOX, R. J. (1970). Estimating the empirical distribution function of certain parameter sequences. *Ann. Math. Statist.* **41** 1845–1852.

JOHNS, M. V., JR. (1957). Nonparametric empirical Bayes procedures. *Ann. Math. Statist.* **28** 649–669.

JOHNS, M. V., JR. and VAN RYZIN, J. (1971). Convergence rates for empirical Bayes two-action problems I. Discrete case. *Ann. Math. Statist.* **42** 1521–1539.

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

LAIPPALA, P. (1979). The empirical Bayes approach with floating sample size in binomial experimentation. *Scand. J. Statist.* **6** 113–118. Correction **7** 105.

LAIPPALA, P. (1985). The empirical Bayes rules with floating optimal sample size for exponential conditional distributions. *Ann. Inst. Statist. Math.* **37** 315–327.

O'BRYAN, T. (1972). Empirical Bayes results in the case of non-identical components. Ph.D. thesis, Michigan State Univ.

O'BRYAN, T. (1976). Some empirical Bayes results in the case of component problems with varying sample sizes for discrete exponential families. *Ann. Statist.* **4** 1290–1293.

O'BRYAN, T. and SUSARLA, V. (1977). Empirical Bayes estimation with non-identical components. Continuous case. *Austral. J. Statist.* **19** 115–125.

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

ROBBINS, H. (1963). The empirical Bayes approach to testing of statistical hypotheses. *Rev. Inst. Internat. Statist.* **31** 195–208.

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

SAMUEL, E. (1963). An empirical Bayes approach to the testing of certain parametric hypotheses. *Ann. Math. Statist.* **34** 1370–1385.

SUSARLA, V. and O'BRYAN, T. (1975). An empirical Bayes two-action problem with non-identical components for a translated exponential distribution. *Comm. Statist.* **4** 767–775.

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

DEPARTMENT OF STATISTICS AND
APPLIED PROBABILITY
THE UNIVERSITY OF ALBERTA
EDMONTON, ALBERTA T6G 2G1
CANADA