Ann. Inst. Statist. Math.
Vol. 42, No. 4, 637-655 (1990)

ON THE EMPIRICAL BAYES APPROACH TO MULTIPLE DECISION PROBLEMS WITH SEQUENTIAL COMPONENTS

R. J. KARUNAMUNI

Department of Statistics and Applied Probability, 434 Central Academic Building, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

(Received December 12, 1988; revised August 7, 1989)

Abstract. The empirical Bayes approach to multiple decision problems with a sequential decision problem as the component is studied. An empirical Bayes m-truncated sequential decision procedure is exhibited for general multiple decision problems. With a sequential component, an empirical Bayes sequential decision procedure selects both a stopping rule function and a terminal decision rule function for use in the component. Asymptotic results are presented for the convergence of the Bayes risk of the empirical Bayes sequential decision procedure.

Key words and phrases: Empirical Bayes procedures, asymptotic risk equivalence, asymptotic superiority, sequential components, multiple decision problems.

1. Introduction

Empirical Bayes theory as introduced by Robbins (1956) and later developed by Robbins (1964) and others (see Maritz (1970), Suzuki (1975) and Susarla (1982)) deals with a sequence of independent repetitions of a given statistical decision problem, called the component problem, where for each problem in the sequence the prior distribution is the same. The components to which the empirical Bayes approach for multiple decision problems have been applied, apart from the fixed sample size case, are varying non-random (see O'Bryan (1972)) and varying random (see Laippala (1985)) sample size cases. A notable work on the fixed sample size two-action with linear loss component is given by Johns and Van Ryzin (1971, 1972), which was generalized by Gilliland and Hannan (1977) and Van Ryzin and Susarla (1977) for general multiple decision problems. In this paper, we consider the empirical Bayes theory with a sequential statistical decision problem as the component. The idea of empirical Bayes problems with sequential components was investigated by Gilliland and Karunamuni (1988), where a finite parameter space case was considered. Furthermore, Karunamuni (1988, 1989) studied empirical Bayes linear loss two-action and squared error loss estimation problems in the context of sequential components. Here we will exhibit an empirical Bayes sequential decision (EBSD) procedure for general multiple decision problems, generalizing the work of Karunamuni (1988).

With the sequential component, an EBSD procedure selects both a stopping rule and a terminal decision rule for each problem. Therefore, the sample size for each problem in the sequence of repetitions is determined by a sequential decision procedure. The EBSD procedure, which operates with a fixed but unknown prior, utilizes the information accumulated from earlier experiences of the same decision problem in order to determine the terminal decision rule and the terminal sample size in the current decision problem.

The treatment in this article can be thought of as a generalization of the following important practical situation in which testing of two hypotheses is relevant.

Consider lots, each containing N items. In order to decide whether a lot should be accepted or not, it is customary to sample r items from it and to accept it when the number of defectives in the sample does not exceed some specified constant k_0. Let θ be the proportion of defectives in the lot. Assume that θ is a random variable which varies from lot to lot, and distributed according to a distribution function G. Let c be the constant cost for sampling an item. Let X_i, $i \geq 1$ be the observations, where $X_i = 1$ for a defective item, and $X_i = 0$ for a non-defective item. We shall assume that sample size r is small as compared to the lot size N, and thus assume that, conditional on $\theta, X_1, X_2, \ldots,$ are i.i.d. Bernoulli random variables with parameter θ (this assumption is natural in sampling inspection; see e.g., Wetherill (1975)). Let a_0 denote the action of accepting the lot and a_1 the action of rejecting it. Let the loss functions be $L(a_0, \theta) = a\theta$, $a > 0$ and $L(a_1, \theta) = b(1 - \theta)$, $b > 0$. Then the action would be to accept if $\theta < b/(a + b)$, reject it if $\theta > b/(a + b)$ and take either action when $\theta = b/(a + b)$. It seems rather reasonable here to use a sequential plan (which selects items one by one) to sample items from the lot and use an EBSD procedure in order to accept or reject a particular lot when G is unknown.

In Section 2 we introduce notation to define the sequential decision procedure to be studied in this paper, and in Section 3 we discuss the sequential component for multiple decision problems. In Section 4, an EBSD procedure is constructed. Asymptotic results and examples are given in Section 5. Most of the proofs are deferred to the Appendix and final remarks are in Section 6.

2. Notation

Let the parameter space be the measurable space $(\Omega, \mathscr{A})$ and denote the class of all prior probability distributions on Ω by $\mathscr{G}$. Conditional on θ, the observable random variables $X_1, X_2, \ldots$ are i.i.d. P_θ, where P_θ is a probability distribution on $(\chi, \mathscr{B})$, χ is the real line and $\mathscr{B}$ is the Borel σ-field. For $k = 1, \ldots, m$, $(m > 1)$, $\boldsymbol{x}^k = (x_1, \ldots, x_k) \in \chi^k$, $\boldsymbol{X}^k \sim P_\theta^k = P_\theta \times \cdots \times P_\theta$ (k terms) conditional on θ, and $\mathscr{B}^k$ denotes the Borel σ-field in χ^k. Let A be the action space in the component problem, and let L be a non-negative loss function defined on $\Omega \times A$. Let c (≥ 0) denote the cost per observation.

For $k = 1, \ldots, m$, let $\mathscr{D}^k$ denote a set of mappings δ from χ^m into A that are constant with respect to the last $m - k$ coordinates and are 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 χ^k when it is convenient to do so, $k = 1, 2, \ldots, m$. For $k = 1, 2, \ldots, m$, let G_k denote the posterior distribution of θ given $\boldsymbol{X}^k = \boldsymbol{x}^k$, where $\theta \sim G$, $G \in \mathscr{G}$. Assume such G_k exists for each $\boldsymbol{x}^k \in \chi^k$, $k = 1, \ldots, m$ and suppose that

$$r(G) = \inf\left\{\int_\Omega L(\theta, a)\, G(d\theta) \colon a \in A\right\} \tag{2.1}$$

is attained at each G. We suppose that a Bayes decision function δ_k relative to G_k attains the infimum posterior Bayes risk, i.e.,

$$r(G_k) = \inf\left\{\int_\Omega L(\theta, a)\, G_k(d\theta) \colon a \in A\right\} = \int_\Omega L(\theta, \delta_k)\, G_k(d\theta) \tag{2.2}$$

for all $\boldsymbol{x}^k \in \chi^k$, $k = 1, \ldots, m$. The situations where $r(G_k) = 0$ will not be considered.

A sequential decision procedure consists of two factors, namely a stopping rule τ and a terminal decision rule $\boldsymbol{\delta}$ (relative to G). The sequential decision procedure we consider in this paper is the following m-truncated sequential decision procedure relative to G. The stopping rule $\boldsymbol{\tau}(G)$ is the finite sequence of functions $(\tau_0, \tau_1, \ldots, \tau_m)$ where τ_0 is a constant function representing the probability of making a decision without sampling, and, for $k = 1, 2, \ldots, m$, $\tau_k \colon \chi^m \to \{0, 1\}$ is an $\boldsymbol{x}^k$-measurable function representing the conditional probability of stopping at stage k given that sampling did not stop at stages $0, 1, \ldots, k-1$, and given the observation $\boldsymbol{x}^k$. We define $\boldsymbol{\tau}(G) = (\tau_0, \tau_1, \ldots, \tau_m)$ as follows: $\tau_0 = 0$, $\tau_m = 1$, and for $k = 1, \ldots, m - 1$

$$\tau_k(\boldsymbol{x}^k) = \begin{cases} 1 & \text{if} \quad E_k r(G_{k+1}) + c - r(G_k) \geq 0 \\ 0 & \text{if} \quad E_k r(G_{k+1}) + c - r(G_k) < 0 \end{cases} \tag{2.3}$$

with E_k denoting the conditional expectation on X_{k+1} given $\boldsymbol{X}^k = \boldsymbol{x}^k$. Notice that for $m \geq 3$, the above stopping rule is not the optimal stopping rule among the m-truncated procedures, but rather, a particularly tractable sequential stopping rule. However, (2.3) defines an optimal stopping rule for $m = 2$ (the stopping rule (2.3) in literature is called the 1-step look ahead stopping rule; see, e.g., the works of Laippala (1979, 1985) and Berger (1985)). A terminal decision rule $\boldsymbol{\delta}(G)$ is defined by the finite sequence

$$(\delta_0, \delta_1, \ldots, \delta_m) \tag{2.4}$$

such that $\delta_k \in \mathscr{D}^k$, $k = 0, 1, \ldots, m$ and δ_k, $k = 0, 1, \ldots, m$, is a Bayes decision rule for the fixed sample size k problem (Ω, A, L) relative to G. If N denotes the stopping time associated with the above sequential decision procedure $(\boldsymbol{\tau}(G), \boldsymbol{\delta}(G))$, then δ_k is used on the set $\{N = k\}$, $k = 0, 1, \ldots, m$, where $\{N = k\}$ denotes the set of all $\boldsymbol{x}^k \in \chi^k$ for which $\tau_k(\boldsymbol{x}^k) = 1$ and $\tau_j(\boldsymbol{x}^j) = 0$ $(j < k)$. Since $\tau_0 = 0$, we may drop the first coordinate function and take $\boldsymbol{\tau}(G) = (\tau_1, \ldots, \tau_m)$ and $\boldsymbol{\delta}(G) = (\delta_1, \ldots, \delta_m)$.

For convenience, the notations χ^k, $k \geq 1$, and Ω under the integral signs that follow are suppressed in all future discussions. Also let $\sum_r$ and $\sum_*$ denote summations over k in $\{1, \ldots, r\}$, $r \geq 1$, and over j in $\{0, 1, \ldots, l\}$. Throughout, we let $[V]$ denote the indicator function of set V and the arguments of functions will not be exhibited whenever they are clear from the context.

3. Multiple decision problems

Assume that the action space $A = \{a_0, a_1, \ldots, a_l\}$ consists of a finite number of distinct actions, and let $L(\theta, a) \geq 0$ on $\Omega \times A$ be the loss function associated with the problem. Assume that the parameter space Ω is a subset of χ. Let $f_\theta \geq 0$ be a density function of the probability distribution P_θ with respect to a given σ-finite measure μ on $(\chi, \mathscr{B})$. Let $f_\theta(\boldsymbol{x}^k)$ denote the product of $f_\theta(x_1), \ldots, f_\theta(x_k)$ for $\boldsymbol{x}^k \in \chi^k$, $k \geq 1$.

We now define our sequential component consisting of a terminal decision rule $\boldsymbol{\delta}(G)$ and a stopping rule $\tau(G)$ relative to G, $G \in \mathscr{G}$, for these multiple decision problems (Ω, A, L), as follows: Let $\boldsymbol{\delta}(G)$ be the terminal decision rule consisting of a finite sequence of functions

$$(\delta_1, \ldots, \delta_m)\,, \tag{3.1}$$

where δ_k is a Bayes decision function with respect to G_k for the fixed sample size k decision problem based on the sample $(X_1, \ldots, X_k)$, $k = 1, \ldots, m$. We define δ_k as follows (see Van Ryzin and Susarla (1977) or Chapter 6 of Ferguson (1967)). Let $\delta_k(\boldsymbol{x}^k) = (\delta(0|\boldsymbol{x}^k), \ldots, \delta(l|\boldsymbol{x}^k))$, where $\delta(j|\boldsymbol{x}^k) =$ Pr {choosing action $a_j|\boldsymbol{X}^k$} and $\sum_* \delta(j|\boldsymbol{x}^k) = 1$. Then, for $j = 0, 1, \ldots, l$,

$$\delta(j \mid \boldsymbol{x}^k) = \begin{cases} 1 & \text{if} \quad \boldsymbol{x}^k \in S_j \\ 0 & \text{if} \quad \boldsymbol{x}^k \notin S_j \end{cases} \tag{3.2}$$

with

$$S_j = \{\boldsymbol{x}^k \mid j = \min \{t: \Delta(t, \boldsymbol{x}^k) = \min \Delta(i, \boldsymbol{x}^k)\}\} ,$$

and

$$\Delta(i, \boldsymbol{x}^k) = \int (L(\theta, a_i) - L(\theta, a_0)) f_\theta(\boldsymbol{x}^k)\, G(d\theta), \qquad i = 0, 1, \ldots, l . \tag{3.3}$$

Routine calculations yield that the minimum posterior Bayes risk with respect to G_k is given by

$$r(G_k) = (f_k(\boldsymbol{x}^k))^{-1} \left\{ \sum_* \delta(j \mid \boldsymbol{x}^k) \Delta(j, \boldsymbol{x}^k) + \int L(\theta, a_0) f_\theta(\boldsymbol{x}^k)\, G(d\theta) \right\}, \tag{3.4}$$

provided $f_k(\boldsymbol{x}^k) > 0$ (note that $r(G_k) = 0$ if $f_k(\boldsymbol{x}^k) = 0$), where

$$f_k(\boldsymbol{x}^k) = \int f_\theta(\boldsymbol{x}^k)\, G(d\theta), \qquad k \geq 1 . \tag{3.5}$$

Hence with

$$\begin{aligned} \beta_k(\boldsymbol{x}^k) = {} & \sum_* \int \delta(j \mid \boldsymbol{x}^{k+1}) \Delta(j, \boldsymbol{x}^{k+1}) \mu(dx_{k+1}) + c f_k(\boldsymbol{x}^k) \\ & - \sum_* \delta(j \mid \boldsymbol{x}^k) \Delta(j, \boldsymbol{x}^l) , \end{aligned} \tag{3.6}$$

the stopping rule (see (2.3)) $\tau(G)$ is defined by a finite sequence $(\tau_1, \ldots, \tau_m)$ which stops sampling at the first $k (k = 1, \ldots, m)$ for which $\tau_k(\boldsymbol{x}^k) = 1$, where

$$\tau_k(\boldsymbol{x}^k) = \begin{cases} 1 & \text{if} \quad \beta_k(\boldsymbol{x}^k) \geq 0 \\ 0 & \text{if} \quad \beta_k(\boldsymbol{x}^k) < 0 \end{cases} \tag{3.7}$$

for $k = 1, \ldots, m - 1$ and $\tau_m = 1$. Let N be the stopping time of the sequential procedure with the stopping rule (3.7). Then

$$N(\boldsymbol{X}^m) = \min \{k: \tau_k(\boldsymbol{X}^k) = 1\} . \tag{3.8}$$

Since $\tau_m = 1$, sampling will be stopped after X_m has been observed if it had not been stopped earlier. The Bayes risk of the sequential component with

$\boldsymbol{d} = (\tau(G), \boldsymbol{\delta}(G))$, relative to G, is

$$R(G, \boldsymbol{d}) = \int R(\theta, (\tau(G), \boldsymbol{\delta}(G)))\, G(d\theta)\,, \tag{3.9}$$

where $R(\theta, (\tau(G), \boldsymbol{\delta}(G))) = \int \sum_m [N = k]\{L(\theta, \delta_k) + ck\} P_\theta^m(d\boldsymbol{x}^m)$. From (3.2) and the equality $\sum_m [N = j] = 1$, we can write $R(G, \boldsymbol{d})$ as:

$$\begin{aligned} R(G, \boldsymbol{d}) = \sum_m \iint [N = k] \Big\{ \sum_* \delta(j \mid \boldsymbol{x}^k)(L(\theta, a_j) - L(\theta, a_0)) \\ + ck \Big\} P_\theta^m(d\boldsymbol{x}^m)\, G(d\theta) \\ + \iint L(\theta, a_0) P_\theta^m(d\boldsymbol{x}^m)\, G(d\theta)\,. \end{aligned} \tag{3.10}$$

We define

$$\begin{aligned} A_k &= [\beta_1 < 0] \cdots [\beta_{k-1} < 0][\beta_k > 0], \quad \text{for} \quad k = 1, \ldots, m-1 \\ B_k &= [\beta_1 < 0] \cdots [\beta_{k-1} < 0][\beta_k = 0], \quad \text{for} \quad k = 1, \ldots, m-1 \\ A_m &= [\beta_1 < 0] \cdots [\beta_{m-1} < 0]\,. \end{aligned} \tag{3.11}$$

Then observe that (3.7), (3.8) and (3.11) give

$$[N = k] = A_k + B_k, \quad k = 1, \ldots, m-1, \quad [N = m] = A_m \tag{3.12}$$

and

$$\sum_{m-1} (A_k + B_k) + A_m = 1\,. \tag{3.13}$$

Now use (3.12) and (3.13) to write (3.10) in the following form:

$$\begin{aligned} R(G, \boldsymbol{d}) = \sum_m \iint A_k \Big\{ \sum_* \delta(j \mid \boldsymbol{x}^k)(L(\theta, a_j) - L(\theta, a_0)) \\ + ck \Big\} P_\theta^m(d\boldsymbol{x}^m)\, G(d\theta) \\ + \sum_{m-1} \iint B_k \Big\{ \sum_* \delta(j \mid \boldsymbol{x}^k)(L(\theta, a_j) \\ - L(\theta, a_0)) + ck \Big\} P_\theta^m(d\boldsymbol{x}^m)\, G(d\theta) \end{aligned} \tag{3.14}$$

$$+\iint L(\theta, a_0) P_\theta^m(d\boldsymbol{x}^m)\, G(d\theta)\,.$$

The main feature of the sequential component is that the observations are taken one at a time, with the statistician having the option of stopping the sampling and making a decision at any time. The idea is that at every stage of the procedure statisticians should compare the posterior Bayes risk of making an immediate decision with the expected posterior Bayes risk that will be obtained if one more observation is taken.

4. Empirical Bayes sequential decision procedure

Suppose that we are experiencing independent repetitions of the same component as described in Section 3. Then, at the n-th problem of the repetitions, random vectors $\boldsymbol{X}_1^{N_1},\dots,\boldsymbol{X}_{n-1}^{N_{n-1}}$ are available to the statistician from the past $(n-1)$ repetitions of the component, where $N_1,\dots,N_{n-1}$ are the respective stopping times of the past $(n-1)$ repetitions. Let $\boldsymbol{d}^n=(\boldsymbol{\tau}^n,\boldsymbol{\delta}^n)$ denote an empirical Bayes sequential decision procedure for the multiple decision problem (Ω, A, L) when the prior distribution G is unknown, where $\boldsymbol{\tau}^n$ is an empirical Bayes stopping rule and $\boldsymbol{\delta}^n$ is an empirical Bayes terminal decision rule. Applying the empirical Bayes approach introduced by Robbins (1956, 1964), we shall construct $\boldsymbol{d}^n$ based on the past data $\boldsymbol{X}_1^{N_1},\dots,\boldsymbol{X}_{n-1}^{N_{n-1}}$ and the present data vector $\boldsymbol{X}$ at the n-th problem of the sequence. We assume that the functions $f_k^n(\boldsymbol{x}^k)=f_k^n(\boldsymbol{x}^k,\boldsymbol{X}_1^{N_1},\dots,\boldsymbol{X}_{n-1}^{N_{n-1}})$, $k\ge 1$ and $\Delta^n(j,\boldsymbol{x}^k)=\Delta^n(j,\boldsymbol{x}^k,\boldsymbol{X}_1^{N_1},\dots,\boldsymbol{X}_{n-1}^{N_{n-1}})$, $k\ge 1, j=0,1,\dots,l$ can be determined such that a.e. $(\mu^k)\boldsymbol{x}^k$, $k\ge 1$,

$$f_k^n(\boldsymbol{x}^k)\xrightarrow{P} f_k(\boldsymbol{x}^k) \quad \text{as} \quad n\to\infty \tag{4.1}$$

and

$$\Delta^n(j,\boldsymbol{x}^k)\xrightarrow{P}\Delta(j,\boldsymbol{x}^k) \quad \text{as} \quad n\to\infty\,, \tag{4.2}$$

where $f_k(\boldsymbol{x}^k)$ and $\Delta(j,\boldsymbol{x}^k)$ are given by (3.5) and (3.3), respectively, and $\xrightarrow{P}$ denotes convergence in probability with respect to the sequence of random vectors $\{\boldsymbol{X}_1^{N_1},\dots,\boldsymbol{X}_{n-1}^{N_{n-1}},\dots\}$. Assumptions similar to (4.1) and (4.2) have been used by Van Ryzin and Susarla (1977) and Gilliland and Hannan (1977) for the standard empirical Bayes multiple decision problems. Examples will be given to describe the functions f_k^n and $\Delta^n(j,\cdot)$ at the end of Section 5.

We use the superscript n to indicate an empirical Bayes quantity in what follows. Let $\boldsymbol{\delta}^n$ be the terminal decision rule consisting of a finite sequence of functions

$$(\delta_1^n,\dots,\delta_m^n)\,, \tag{4.3}$$

where

$$\delta_k^n(\boldsymbol{x}^k) = (\delta^n(0|\boldsymbol{x}^k),\ldots,\delta^n(l|\boldsymbol{x}^k)) \tag{4.4}$$

such that $\sum_* \delta^n(j|\boldsymbol{x}^k) = 1$ for all k, and for $j = 0, 1,\ldots, l$,

$$\delta^n(j|\boldsymbol{x}^k) = \begin{cases} 1 & \text{if} \quad \boldsymbol{x}^k \in S_j^n \\ 0 & \text{if} \quad \boldsymbol{x}^k \notin S_j^n , \end{cases} \tag{4.5}$$

where $S_j^n = \left\{ \boldsymbol{x}^k | j = \min \left\{ t: \Delta^n(t, \boldsymbol{x}^k) = \min_i \Delta^n(i, \boldsymbol{x}^k) \right\}\right\}$, and $\Delta^n(j, \boldsymbol{x}^k)$ satisfies (4.2) for $j = 0, 1,\ldots, l$. The stopping rule τ^n is defined by a finite sequence of functions

$$(\tau_1^n,\ldots, \tau_m^n) \tag{4.6}$$

which stops sampling for the procedure $\boldsymbol{d}^n$ at the first k $(k = 1,\ldots, m)$ for which $\tau_k^n(\boldsymbol{x}^k) = 1$, where $\tau_m^n = 1$ and, for $k = 1,\ldots, m-1$,

$$\tau_k^n(\boldsymbol{x}^k) = \begin{cases} 1 & \text{if} \quad \beta_k^n(\boldsymbol{x}^k) \geq 0 \\ 0 & \text{if} \quad \beta_k^n(\boldsymbol{x}^k) < 0 , \end{cases} \tag{4.7}$$

where the function β_k^n is defined by

$$\begin{aligned} \beta_k^n(\boldsymbol{x}^k) = {} & \sum_* \int \delta^n(j|\boldsymbol{x}^{k+1}) \Delta^n(j, \boldsymbol{x}^{k+1}) \mu(dx_{k+1}) + cf_k^n(\boldsymbol{x}^k) \\ & - \sum_* \delta^n(j|\boldsymbol{x}^k) \Delta^n(j, \boldsymbol{x}^k) , \end{aligned} \tag{4.8}$$

and f_k^n and $\Delta^n(j, \boldsymbol{x}^k)$, $j = 0, 1,\ldots, l$, satisfy (4.1) and (4.2), respectively. We define

$$C_k^n = [\beta_1^n < 0] \cdots [\beta_{k-1}^n < 0][\beta_k^n \geq 0] \qquad \text{for} \quad k = 1,\ldots, m-1 \tag{4.9}$$

and

$$C_m^n = [\beta_1^n < 0] \cdots [\beta_{m-1}^n < 0] . \tag{4.10}$$

Then note that $[N^n = k] = C_k^n$ for $k = 1,\ldots, m$, where N^n is the stopping time of the EBSD procedure $\boldsymbol{d}^n = (\tau^n, \boldsymbol{\delta}^n)$ defined above for multiple decision problems, and given by

$$N^n(\boldsymbol{X}^m) = \min \{k | \tau_k^n(\boldsymbol{X}^k) = 1\} . \tag{4.11}$$

Let $R(G, \boldsymbol{d}^n)$ denote the Bayes risk of the EBSD procedure $\boldsymbol{d}^n = (\tau^n, \boldsymbol{\delta}^n)$ with respect to G. Then it is easy to show that

$$R(G, \boldsymbol{d}^n) = \sum_m \iint C_k^n \left\{ \sum_* \delta^n(j \mid \boldsymbol{x}^k)(L(\theta, a_j) - L(\theta, a_0)) + ck \right\} P_\theta^m(d\boldsymbol{x}^m) G(d\theta) + \iint L(\theta, a_0) P_\theta^m(d\boldsymbol{x}^m) G(d\theta) . \tag{4.12}$$

The difference $ER(G, \boldsymbol{d}^n) - R(G, \boldsymbol{d})$ is used as a measure of optimality of the sequence of EBSD procedures $\boldsymbol{d}^n = (\tau^n, \boldsymbol{\delta}^n)$, with respect to $\boldsymbol{d} = (\tau, \boldsymbol{\delta})$ where $E(\cdot)$ stands for expectation operation with respect to the random vectors $\boldsymbol{X}_1^{N_1}, \ldots, \boldsymbol{X}_{n-1}^{N_{n-1}}$. (This notation will be used in the rest of the paper without further comment.) We say $\boldsymbol{d}^n = (\tau^n, \boldsymbol{\delta}^n)$ is asymptotically risk equivalent (optimal) relative to (an optimal) sequential procedure $\boldsymbol{d} = (\tau, \boldsymbol{\delta})$ if $\lim_{n \to \infty} ER(G, \boldsymbol{d}) = R(G, \boldsymbol{d})$, and asymptotically superior (a.sr.) relative to $\boldsymbol{d} = (\tau, \boldsymbol{\delta})$ if $\limsup_{n \to \infty} ER(G, \boldsymbol{d}^n) \leq R(G, \boldsymbol{d})$.

5. Asymptotic behaviour of $ER(G, (\tau^n, \delta^n))$

In this section we state the main results of this paper regarding the asymptotic optimality and the asymptotic superiority of the EBSD procedure $\boldsymbol{d}^n$ defined in the previous section. We first state and prove a lemma which is useful in proving the main theorems of this paper.

LEMMA 5.1. *Let G be such that*

$$\int L(\theta, a_j) G(d\theta) < \infty, \quad j = 0, 1, \ldots, l, \tag{5.1}$$

and let $f_k^n(\boldsymbol{x}^k)$ and $\Delta^n(j, \boldsymbol{x}^k)$ be defined by (4.1) *and* (4.2), *respectively*, $k \geq 1$, $j = 0, 1, \ldots, l$. *If $\Delta^n(j, \boldsymbol{x}^k)$ satisfies*

$$\int \delta^n(j \mid \boldsymbol{x}^{k+1}) \Delta^n(j, \boldsymbol{x}^{k+1}) \mu(dx_{k+1}) \xrightarrow{P} \int \delta(j \mid \boldsymbol{x}^{k+1}) \Delta(j, \boldsymbol{x}^{k+1}) \mu(dx_{k+1}) \tag{5.2}$$

as $n \to \infty$, $j = 0, 1, \ldots, l$, $k \geq 1$, *then*

$$\beta_k^n \xrightarrow{P} \beta_k \quad \text{as} \quad n \to \infty, \quad k \geq 1 . \tag{5.3}$$

PROOF. For fixed $\boldsymbol{x}^k$, the definitions of $\delta^n(j|\boldsymbol{x}^k)$ and $\delta(j|\boldsymbol{x}^k)$, $j = 0, 1, \ldots, l$, and (4.2) give

$$\sum_* \delta^n(j|\boldsymbol{x}^k)\Delta^n(j, \boldsymbol{x}^k) \xrightarrow{P} \sum_* \delta(j|\boldsymbol{x}^k)\Delta(j, \boldsymbol{x}^k) \quad \text{as} \quad n \to \infty .$$

Now the proof is completed by (4.1) and (5.2).

The next theorem and its corollaries discuss the asymptotic behaviour of the unconditional Bayes risk $ER(G, \boldsymbol{d}^n)$ of the EBSD procedure $\boldsymbol{d}^n$. The proofs of Theorem 5.1 and its corollaries below are given in the Appendix.

THEOREM 5.1. *Let* $\boldsymbol{d}^n = (\tau^n, \boldsymbol{\delta}^n)$ *be defined by* (4.2), (4.4), (4.5), (4.6) *and* (4.7). *Let* $f_k^n(\boldsymbol{x}^k)$ *and* $\Delta^n(j, \boldsymbol{x}^k)$ *satisfy* (4.1) *and* (4.2), *respectively*, $k \geq 1$, $j = 0, 1, \ldots, l$, *and further assume that* (5.2) *holds. Let G satisfy* (5.1). *Then*

$$\limsup_{n \to \infty} ER(G, \boldsymbol{d}^n) \leq R(G, \boldsymbol{d}) ,$$

that is, the EBSD procedure $\boldsymbol{d}^n$ *is a.sr. relative to* $\boldsymbol{d}$.

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

$$\text{(5.4)} \quad \liminf_{n \to \infty} E[\beta_1^n < 0] \cdots [\beta_{i-1}^n < 0] = [\beta_1 < 0] \cdots [\beta_{i-1} < 0], \qquad i \geq 2 ,$$

then $\lim_{n \to \infty} ER(G, \boldsymbol{d}^n) = R(G, \boldsymbol{d})$.

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

$$\text{(5.5)} \quad \iint [\beta_1 < 0] \cdots [\beta_{j-1} = 0][\beta_j = 0][\beta_{i-1} < 0] P_\theta^m(d\boldsymbol{x}^m) G(d\theta) = 0 ,$$

for $i = j + 2, \ldots, m$, $j = 1, \ldots, m - 1$, *then* $\lim_{n \to \infty} ER(G, \boldsymbol{d}^n) = R(G, \boldsymbol{d})$.

COROLLARY 5.3. *If* $m = 2$, *then* $\boldsymbol{d} = (\tau, \boldsymbol{\delta})$ *defined by* (3.7) *and* (3.2) *is optimal, and under the same assumptions as in Theorem* 5.1, $\lim_{n \to \infty} ER(G, \boldsymbol{d}^n) = R(G, \boldsymbol{d})$, *that is, the EBSD procedure* $\boldsymbol{d}^n = (\tau^n, \boldsymbol{\delta}^n)$ *is asymptotically optimal relative to* $\boldsymbol{d}$.

Since the definitions of N^n and N give $[N^n \geq i] = [\beta_1^n < 0] \cdots [\beta_{i-1}^n < 0]$ and $[N \geq i] = [\beta_1 < 0] \cdots [\beta_{i-1} < 0]$, $i \geq 2$, condition (5.4) is equivalent to saying that $\liminf_{n \to \infty} E[N^n \geq i] = [N \geq i]$, $i \geq 2$. Roughly speaking, this means that asymptotically the sample size N^n of the EBSD will be no more than the sample size N of the component. The next theorem compares the asymptotic behaviour of EB stopping time N^n with N, and shows that $\liminf_{n\to\infty} E[N^n \geq i] \geq [N \geq i]$ always, $i \geq 2$.

THEOREM 5.2. *Let N^n and N be the stopping times associated with the EBSD procedure and the sequential component, respectively. Then,* (a) $\liminf_{n\to\infty} E[N^n \geq i] \geq [N \geq i]$, $i \geq 2$, *and*

$$\text{(b)} \qquad \liminf_{n\to\infty} \iint E[N^n \geq i] P_\theta^m(d\boldsymbol{x}^m) G(d\theta) \geq \iint [N \geq i] P_\theta^m(d\boldsymbol{x}^m) G(d\theta), \qquad i \geq 2 .$$

PROOF. For $i \geq 2$, by the definitions of N^n and N, we have

$$(5.6) \qquad [\beta_1 < 0] \cdots [\beta_{i-1} < 0] \leq \liminf E[N^n \geq i] \leq \limsup_{n\to\infty} E[N^n \geq i] \leq [\beta_1 \leq 0] \cdots [\beta_{i-1} \leq 0] .$$

Part (a) of Theorem 5.2 is the first inequality of (5.6), and part (b) follows from (5.6) and Fatou's lemma.

The EBSD procedure used in Theorem 5.1 is based on the functions $f_k^n(\boldsymbol{x}^k)$ and $\Delta^n(j, \boldsymbol{x}^k)$, $j = 0, 1, \ldots, l$, $k \geq 1$. We now find sequences of functions $\{f_k^n(\boldsymbol{x}^k)\}_{n\geq 1}$ and $\{\Delta^n(j, \boldsymbol{x}^k)\}_{n\geq 1}$, $j = 0, 1, \ldots, l$, $k \geq 1$ for two examples.

Example 1. (A monotone multiple decision problem) Let $m = 2$. Let the conditional density be $f_\theta(x) = \theta^{x-1}(1-\theta)$, $x = 1, 2, \ldots$, and $0 < \theta < 1$. Let $0 = \theta_{-1} < \theta_0 < \cdots < \theta_{l-1} < \theta_l = 1$ be known. Let the action a_j correspond to deciding "the value of θ is in the interval $[\theta_{j-1}, \theta_j]$," $j = 0, 1, \ldots, l$. Let the loss function on $(0, 1) \times A$ be defined by

$$L(\theta, a_0) = \begin{cases} 0 & \text{if } \theta \leq \theta_0 \\ \sum_j (\theta - \theta_{i-1}) & \text{if } \theta_{j-1} < \theta \leq \theta_j \end{cases}$$

and $L(\theta, a_{j+1}) - L(\theta, a_j) = (\theta_j - \theta)$, $j = 0, 1, \ldots, l-1$. First, notice the following expressions for $f_1(x_1)$ and $f_2(\boldsymbol{x}^2)$ with $\boldsymbol{x}^1 = x_1$ and $\boldsymbol{x}^2 = (x_1, x_2)$.

$$(5.7) \qquad f_1(x_1) = \int_0^1 \theta^{x_1 - 1}(1-\theta) G(d\theta), \qquad x_1 = 1, 2, \ldots$$

and

(5.8) $$f_2(\boldsymbol{x}^2) = f_1(x_1 + x_2 - 1) - f_1(x_1 + x_2) .$$

With the above loss functions, observe that $L(\theta, a_i) - L(\theta, a_0) = \sum_{j=0}^{i-1} (\theta_j - \theta)$, $i = 1, 2, \ldots, l$, and thus, $\Delta(a_i, \boldsymbol{x}^k)$ is given by

(5.9) $$\Delta(a_i, \boldsymbol{x}^k) = f_k(\boldsymbol{x}^k) \sum_{j=0}^{i-1} \theta_j - i g_k(\boldsymbol{x}^k), \quad i = 1, 2, \ldots, l, \quad k = 1, 2 ,$$

where

(5.10) $$g_k(\boldsymbol{x}^k) = \int_0^1 \theta f_\theta(\boldsymbol{x}^k) G(d\theta), \quad k = 1, 2 .$$

Routine calculations yield

(5.11) $$g_1(x_1) = f_1(x_1 + 1), \quad x_1 = 1, 2, \ldots$$

and

(5.12) $$g_2(\boldsymbol{x}^2) = f_1(x_1 + x_2) - f_1(x_1 + x_2 - 1) .$$

Notice that the first observations $X_{11}, X_{21}, \ldots, X_{n1}$ from the past and present repetitions are i.i.d. with a common marginal density $f_1(x) = \int_0^1 \theta^{x-1}(1 - \theta) \cdot G(d\theta)$. Thus, we define our EB estimator of $f_1(x)$ by

$$f_1^n(x) = n^{-1} \sum_{i=1}^{n} [X_{i1} = x], \quad x = 1, 2, \ldots ,$$

and now EB estimators of $f_2(\boldsymbol{x}^2)$ and $\Delta(a_i, \boldsymbol{x}^k)$, $k = 1, 2$ are obtained by (5.8), (5.9), (5.10), (5.11) and (5.12). It is easy to verify that the above estimators $\Delta^n(i, \boldsymbol{x}^k)$ satisfy condition (5.2).

Example 2. Let $f_\theta(x) = (2\pi\tau^2)^{-1/2} \exp[-(x - \theta)^2/2\tau^2]$, $-\infty < x < \infty$, where τ^2 is known and the parameter θ is normally distributed with mean μ and variance σ^2. Assume that μ and σ^2 are fixed but unknown. Then it is clear that the functions $f_k(\boldsymbol{x}^k)$ and $\Delta(a_j, \boldsymbol{x}^k)$, $j = 1, 2, \ldots, l$, $k \geq 1$, depend on the unknown quantities μ and σ^2, and therefore it is enough to estimate μ and σ^2 consistently in the EB context using the past and present data $\boldsymbol{X}_1^{N_1}, \ldots, \boldsymbol{X}_n^{N_n}$, where the random sample sequence $\{N_i\}_{i \geq 1}$ is defined by (4.11). We define

$$\sigma^n = \max\left\{\left(\sum_{i=1}^n N_i - \frac{\sum_{i=1}^n N_i^2}{\sum_{i=1}^n N_i}\right)^{-1} \sum_{i=1}^n N_i(\bar{X}_i - \bar{\bar{X}})^2 - \tau^2, 0\right\} \tag{5.13}$$

and

$$\mu^n = \sum_{i=1}^n N_i \bar{X}_i(\tau^2 + \sigma^n N_i)^{-1} \Big/ \sum_{i=1}^n N_i(\tau^2 + \sigma^n N_i)^{-1}, \tag{5.14}$$

where $\bar{X}_i = N_i^{-1} \sum_{j=1}^{N_i} X_{ij}$, $i \geq 1$ and $\bar{\bar{X}} = \sum_{i=1}^n N_i \bar{X}_i \Big/ \sum_{i=1}^n N_i$. Monte Carlo study shows as $n \to \infty$, σ^n and μ^n approach σ^2 and μ respectively, and thus, may be useful in applications.

6. Final Remarks

In empirical Bayes decision theory, researchers have dealt only with procedures involving non-sequential components, namely, the fixed sample size and varying sample size decision problems. The present paper examines the sequential nature of the empirical Bayes procedure and introduces the stopping rule concept to empirical Bayes problems, in particular for multiple decision problems. From this paper one can see the flexibility of empirical Bayes ideas in sequential decision theory, a case which is one of the most natural to apply empirical Bayes methods. The sequential procedure used as the component in this paper is called a one-step look ahead procedure. The one-step look ahead stopping rule is much simpler than the Bayes stopping rule. The latter can be defined by backward induction argument (i.e., the decision to stop is based on a comparison of the conditional on $\boldsymbol{x}^k$, Bayes risk of stopping and making a decision with that of not stopping and playing an optimal sequential strategy from that point). Observe that when $m = 2$, the one-step look ahead stopping rule used in this article is a Bayes stopping rule. For $m > 2$, the construction of empirical Bayes procedures for the Bayes stopping rule defined by backward induction can be made with laborious calculations.

The results here naturally lead to corresponding results in other possible multiple decision component problems as well. Details of these results will be reported in a separate paper.

Acknowledgements

The author is indebted to Professor Dennis C. Gilliland, his thesis advisor, for his continued encouragement and help throughout this work, and also thanks Professor James O. Berger and the late Professor V. Susarla for several comments on an earlier version of the manuscript. Thanks are also due to Professor T. Matsunawa and the referee for their thoughtful suggestions which improved the presentation of this paper.

Appendix

The outline of the proofs of Theorem 5.1 and its corollaries are given in this Appendix. Detailed proofs can be found in Karunamuni (1986). The intuitive idea of the proof of Theorem 5.1 is to treat the cases $N^n = N$, $N^n > N$ and $N^n < N$, separately. We write the difference $ER(G, \boldsymbol{d}^n) - R(G, \boldsymbol{d})$ as a sum $\sum_{i=1}^{3} J_i^n + \sum_{i=1}^{3} K_i^n$, where the terms J_3^n and K_2^n refer to the difference on $[N^n = N]$, J_1^n and K_1^n refer to the difference on $[N^n > N]$ and finally, J_2^n and K_3^n represent the difference on $[N^n < N]$. The terms J_i^n's and K_i^n's are defined below. We will show that $\lim_{n\to\infty} J_i^n = 0$, $i = 1, 2, 3$ and $\lim_{n\to\infty} K_i^n = 0$, $i = 1, 2$ in Lemmas A.1 and A.2, respectively, below. The proof of Theorem 5.1 is completed by proving $\limsup_{n\to\infty} K_3^n \le 0$. Specifically, K_3^n represents the difference $ER(G, \boldsymbol{d}^n) - R(G, \boldsymbol{d})$ on $[N^n < N]B_j$, $j = 1, \dots, m-1$, that is, when $N^n < N$ and the sampling of the sequential component is stopped as soon as β_j reaches the boundary at zero, $j = 1, \dots, m-1$.

LEMMA A.1. *Let*

$$J_1^n = \sum_{j=2}^{m} \sum_{i=1}^{j-1} \iint EC_i^n A_j \left\{ \sum_{k=0}^{l} \delta^n(k|\boldsymbol{x}^i) g_k(\theta) - \sum_{k=0}^{l} \delta(k|\boldsymbol{x}^j) g_k(\theta) + c(i-j) \right\} P_\theta^m(d\boldsymbol{x}^m) G(d\theta),$$

$$J_2^n = \sum_{j=1}^{m-1} \sum_{i=j+1}^{m} \iint EC_i^n A_j \left\{ \sum_{k=0}^{l} \delta^n(k|\boldsymbol{x}^i) g_k(\theta) - \sum_{k=0}^{l} \delta(k|\boldsymbol{x}^j) g_k(\theta) + c(i-j) \right\} P_\theta^m(d\boldsymbol{x}^m) G(d\theta)$$

and

$$J_3^n = \sum_{i=1}^{m} \iint EC_i^n A_i \left\{ \sum_{k=0}^{l} \delta^n(k|\boldsymbol{x}^i) g_k(\theta) - \sum_{k=0}^{l} \delta(k|\boldsymbol{x}^i) g_k(\theta) \right\} \times P_\theta^m(d\boldsymbol{x}^m) G(d\theta).$$

If G is such that $\int L(\theta, a_j) G(d\theta) < \infty$, $j = 0, 1, \dots, l$, then $\lim_{n\to\infty} J_i^n = 0$, $i = 1, 2, 3$, *where* $g_j(\theta) = L(\theta, a_j) - L(\theta, a_0), j = 0, 1, \ldots, l.$

PROOF. Observe that

$$|J_1^n| \le \sum_{j=2}^{m} \sum_{i=1}^{j-1} \iint EC_i^n A_j \left\{ 2 \sum_{k=0}^{l} |g_k(\theta)| + c|i-j| \right\} P_\theta^m(d\boldsymbol{x}^m)\, G(d\theta),$$

and by the definitions of C_i^n and A_j, $C_i^n A_j \le [\beta_i^n \ge 0][\beta_i < 0]$ for $i < j$. But, from Lemma 5.1, $[\beta_i^n \ge 0][\beta_i < 0] \xrightarrow{P} 0$ as $n \to \infty$. Now use the dominated convergence theorem (DCT) and the assumptions on G to conclude that $\lim_{n\to\infty} J_1^n = 0$. The proof that $\lim_{n\to\infty} J_2^n = 0$ is similar. Notice that J_3^n can be written in the following form:

$$J_3^n = \sum_{i=1}^{m} \int EC_i^n A_i \left\{ \sum_{k=0}^{l} [\delta^n(k|\boldsymbol{x}^i) - \delta(k|\boldsymbol{x}^i)] \Delta(k, \boldsymbol{x}^i) \right\} \mu^i(d\boldsymbol{x}^i).$$

Now use $\sum_{k=0}^{l} \delta^n(k|\boldsymbol{x}^i) \Delta(a_k, \boldsymbol{x}^i) \xrightarrow{P} \sum_{k=0}^{l} \delta(k|\boldsymbol{x}^i) \Delta(k, \boldsymbol{x}^i)$ as $n \to \infty$, $\int |\Delta(j, \boldsymbol{x}^i)| \cdot \mu^i(d\boldsymbol{x}^i) < \infty$, $j = 0, 1, \ldots, l$, $i \ge 1$, and the DCT to conclude that $\lim_{n\to\infty} J_3^n = 0$.

LEMMA A.2. *Let*

$$K_1^n = \sum_{j=2}^{m-1} \sum_{i=1}^{j-1} \iint EC_i^n B_j \left\{ \sum_{k=0}^{l} \delta^n(k|\boldsymbol{x}^i) g_k(\theta) - \sum_{k=0}^{l} \delta(k|\boldsymbol{x}^j) g_k(\theta) + c(i-j) \right\} P_\theta^m(d\boldsymbol{x}^m)\, G(d\theta)$$

and

$$K_2^n = \sum_{i=1}^{m-1} \iint EC_i^n B_i \left\{ \sum_{k=0}^{l} [\delta^n(k|\boldsymbol{x}^i) - \delta(k|\boldsymbol{x}^i)] g_k(\theta) \right\} P_\theta^m(d\boldsymbol{x}^m)\, G(d\theta),$$

then under the same assumptions as in Lemma A.1, $\lim_{n\to\infty} K_i^n = 0$, $i = 1, 2$.

PROOF. The proofs of K_1^n and K_2^n are similar to the proofs of J_1^n and J_3^n, respectively.

PROOF OF THEOREM 5.1. Use Fubini theorem and $\int L(\theta, a_i)\, G(d\theta) < \infty$, $i = 0, 1, \ldots, l$, to write the difference $ER(G, \boldsymbol{d}^n) - R(G, \boldsymbol{d})$ in the following form:

(A.1) $$ER(G,\boldsymbol{d}^n) - R(G,\boldsymbol{d}) = \sum_{i=1}^{3} J_i^n + \sum_{i=1}^{3} K_i^n ,$$

where J_1^n, J_2^n, J_3^n and K_1^n, K_2^n are as defined in Lemmas A.1 and A.2, respectively, and

(A.2) $$K_3^n = \sum_{j=1}^{m-1} \sum_{i=j+1}^{m} \iint EC_i^n B_j \left\{ \sum_{k=0}^{l} \delta^n(k|\boldsymbol{x}^i) g_k(\theta) - \sum_{k=0}^{l} \delta(k|\boldsymbol{x}^j) g_k(\theta) + c(i-j) \right\} P_\theta^m(d\boldsymbol{x}^m) G(d\theta) .$$

We will show that $\limsup_{n\to\infty} K_3^n \le 0$. First we define $L(\boldsymbol{x}^i,\theta) = \sum_{k=0}^{l} \delta(k|\boldsymbol{x}^j) \cdot g_k(\theta)$, $L^n(\boldsymbol{x}^i,\theta) = \sum_{k=0}^{l} \delta^n(k|\boldsymbol{x}^i) g_k(\theta)$, and $M(\boldsymbol{x}^i) = \sum_{k=0}^{l} \delta(k|\boldsymbol{x}^i) \Delta(k,\boldsymbol{x}^i)$ and $M^n(\boldsymbol{x}^i) = \sum_{k=0}^{l} \delta^n(k|\boldsymbol{x}^i) \Delta(k,\boldsymbol{x}^i)$, $i \ge 1$. Also, let $D_k^n = \sum_{i=k}^{m} C_i^n$, $k = j+1,\dots,m$ and $j = 1,\dots,m-1$. Then $C_i^n = D_i^n - D_{i+1}^n$, $i = 1,\dots,m-1$ and $D_m^n = C_m^n$. Now observe that K_3^n can be simplified into the following form:

(A.3) $$K_3^n = \sum_{j=1}^{m-1} \left\{ \sum_{i=j+1}^{m} T^n(i,j) - \sum_{i=j+1}^{m-1} S^n(i,j) - \int EB_j D_{j+1}^n M(\boldsymbol{x}^j) \mu^j(d\boldsymbol{x}^i) + \sum_{i=j+1}^{m} U^n(i,j) \right\},$$

where

$$T^n(i,j) = \int ED_i^n B_j M^n(\boldsymbol{x}^i) \mu^i(d\boldsymbol{x}^i) ,$$

$$S^n(i,j) = \int ED_i^n B_j M^n(\boldsymbol{x}^i) \mu^i(d\boldsymbol{x}^i) \quad \text{and}$$

$$U^n(i,j) = c \int ED_i^n B_j f_{i-1}(\boldsymbol{x}^{i-1}) \mu^{i-1}(d\boldsymbol{x}^{i-1}) .$$

Now combining the $(j+1)$-th terms and writing the rest of the terms together, we get

$$K_3^n = \sum_{j=1}^{m-1} \left\{ T^n(j+1,j) - \int ED_{j+1}^n B_j M(\boldsymbol{x}^j) \mu^j(d\boldsymbol{x}^j) + U^n(j+1,j) + \sum_{i=j+2}^{m} [T^n(i,j) - S^n(i-1,j) + U^n(i,j)] \right\}.$$

The sum of $T^n(j+1,j)$ and $-\int ED^n_{j+1}B_j M(\boldsymbol{x}^j)\mu^j(d\boldsymbol{x}^j) + U^n(j+1,j)$ is equal to

$$\text{(A.4)}\quad \int ED^n_{j+1}B_j\left(\int M^n(\boldsymbol{x}^{j+1})\mu(dx_{j+1}) - M(\boldsymbol{x}^j) + cf_j(\boldsymbol{x}^j)\right)\mu^j(d\boldsymbol{x}^j)\,.$$

But, from the definition of β_j (see (3.6)), we have $-M(\boldsymbol{x}^j) + cf_j(\boldsymbol{x}^j) = -\int M(\boldsymbol{x}^{j+1})\mu(dx_{j+1})$ on $\{\beta_j = 0\}$, and $0 \le B_j \le [\beta_j = 0]$. Therefore, the absolute value of the expression (A.4) is less than or equal to $\int E[\beta_j = 0]|M^n(\boldsymbol{x}^{j+1}) - M(\boldsymbol{x}^{j+1})|\mu^{j+1}(d\boldsymbol{x}^{j+1})$ which goes to zero as $n \to \infty$, by an application of the DCT. Now for $i = j+2,\ldots,m$, $T^n(i,j) - S^n(i-1,j) + U^n(i,j)$ is equal to

$$\text{(A.5)}\quad \int ED^n_i B_j\left\{\int M^n(\boldsymbol{x}^i)\mu(dx_i) - M^n(\boldsymbol{x}^{i-1}) + cf_{i-1}(\boldsymbol{x}^{i-1})\right\}\mu^{i-1}(d\boldsymbol{x}^{i-1})\,.$$

Adding and subtracting the term $M(\boldsymbol{x}^{i-1}) + \int M(\boldsymbol{x}^i)\mu(dx_i)$ into the integrand of the integral (A.5), we get for $i = j+2,\ldots,m, j = 1,\ldots,m-1$,

$$\begin{aligned}\text{(A.6)}\quad &\int ED^n_i B_j\{M(\boldsymbol{x}^{i-1}) - M^n(\boldsymbol{x}^{i-1})\}\mu^{i-1}(d\boldsymbol{x}^{i-1})\\ &+\int ED^n_i B_j\{M^n(\boldsymbol{x}^i) - M(\boldsymbol{x}^i)\}\mu^i(d\boldsymbol{x}^i)\\ &+\int ED^n_i B_j\left\{\int M(\boldsymbol{x}^i)\mu(dx_i) - M(\boldsymbol{x}^{i-1}) + cf_{i-1}(\boldsymbol{x}^{i-1})\right\}\mu^{i-1}(d\boldsymbol{x}^{i-1})\,.\end{aligned}$$

Again the first and the second integral terms in (A.6) go to zero as $n \to \infty$, by an application of the DCT and from the definition of β_j. The third term is equal to

$$\text{(A.7)}\quad \int ED^n_i B_j\beta_{i-1}\mu^{i-1}(d\boldsymbol{x}^{i-1}) \quad \text{for} \quad i = j+2,\ldots,m, \quad j = 1,\ldots,m-1\,.$$

Expression (A.7) can be rewritten as a sum of two terms as follows:

$$\begin{aligned}\text{(A.8)}\quad &\int ED^n_i(d\boldsymbol{x}^{i-1})B_j[\beta_{i-1} \le 0]\beta_{i-1}\mu^{i-1}(d\boldsymbol{x}^{i-1})\\ &+\int ED^n_i B_j[\beta_{i-1} > 0]\beta_{i-1}\mu^{i-1}(d\boldsymbol{x}^{i-1})\,.\end{aligned}$$

The second term in (A.8) goes to zero as $n \to \infty$, by an application of the DCT, since $0 \le D^n_i[\beta_{i-1} > 0] \le [\beta^n_{i-1} > 0][\beta_{i-1} > 0]$. The first term in (A.8) is non-positive for all values of n. Thus, $\limsup_{n\to\infty}\int ED_i^n B_j[\beta_{i-1}\le 0]\beta_{i-1}\mu^{i-1}\cdot(d\boldsymbol{x}^{i-1})\le 0$, and now from (A.4)–(A.8), it follows that $\limsup_{n\to\infty} K_3^n\le 0$. This completes the proof of Theorem 5.1.

PROOF OF COROLLARY 5.1. First, notice that $|ED_i^n[\beta_{i-1}\le 0]B_j\beta_{i-1}|\le|\beta_{i-1}|$, for all n, and $\int|\beta_{i-1}|\mu^i(d\boldsymbol{x}^i)<\infty$, $i=j+2,\dots,m$, $j=1,\dots,m-1$. Now applying Fatou's lemma to $ED_i^n[\beta_{i-1}\le 0]B_j\beta_{i-1}+|\beta_{i-1}|$, we obtain

$$\begin{aligned}&\liminf_{n\to\infty}\int ED_i^n[\beta_{i-1}\le 0]B_j\beta_{i-1}\mu^{i-1}(d\boldsymbol{x}^{i-1})\\&\qquad\ge\int\liminf_{n\to\infty} ED_i^n[\beta_{i-1}\le 0]B_j\beta_{i-1}\mu^{i-1}(d\boldsymbol{x}^{i-1}),\end{aligned}$$

$i=j+2,\dots,m$, $j=1,\dots,m-1$. But $\liminf_{n\to\infty} ED_i^n B_j=0$, for $i=j+2,\dots,m$, $j=1,\dots,m-1$, when (5.4) holds. Thus, $\liminf_{n\to\infty}\int ED_i^n[\beta_{i-1}\le 0]B_j\beta_{i-1}\cdot\mu^{i-1}(d\boldsymbol{x}^{i-1})$ is equal to zero, $i=j+2,\dots,m$, $j=1,\dots,m-1$. Now $\liminf_{n\to\infty} ER(G,\boldsymbol{d}^n)\ge R(G,\boldsymbol{d})$ follows from (A.4)–(A.8), and Lemmas A.1 and A.2. Also, by Theorem 5.1, we obtain $\lim_{n\to\infty} ER(G,\boldsymbol{d}^n)=R(G,\boldsymbol{d})$.

PROOF OF COROLLARY 5.2. When (5.5) holds, observe that $\int B_j[\beta_{i-1}\le 0]\mu^{i-1}(d\boldsymbol{x}^{i-1})=0$, $i=j+2,\dots,m, j=1,\dots,m-1$. Therefore, the first term in (A.8) is equal to zero by (5.5).

PROOF OF COROLLARY 5.3. When $m=2$, by the definition of the sequential component $\boldsymbol{d}$ is optimal, and by the construction of the EBSD procedure $\boldsymbol{d}^n$, we have $R(G,\boldsymbol{d}^n)\ge R(G,\boldsymbol{d})$ for all n and G. Therefore, $\liminf_{n\to\infty} ER(G,\boldsymbol{d}^n)\ge R(G,\boldsymbol{d})$. Hence, $\lim_{n\to\infty} ER(G,\boldsymbol{d}^n)=R(G,\boldsymbol{d})$ now follows from Theorem 5.1.

REFERENCES

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis, and Methods*, Springer, New York.

Ferguson, T. S. (1967). *Mathematical Statistics. A Decision Theoretical Approach*, Academic Press, New York.

Gilliland, D. and Hannan, J. (1977). Improved rates for empirical Bayes monotone decision problems, *Ann. Statist.*, **5**, 516–521.

Gilliland, D. and Karunamuni, R. J. (1988). On empirical Bayes with sequential component, *Ann. Inst. Statist. Math.*, **40**, 187–193.

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.

Karunamuni, R. J. (1986). On the empirical Bayes approach to multiple decision problems with sequential components, Unpublished manuscript.

Karunamuni, R. J. (1988). On empirical Bayes testing with sequential components, *Ann. Statist.*, **16**, 1270–1282.

Karunamuni, R. J. (1989). On empirical Bayes sequential estimation, *Comm. Statist. A—Theory Methods*, **18**, 2533–2552.

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.

Maritz, J. S. (1970). *Empirical Bayes Methods*, Methurn, London.

O'Bryan, T. (1972). Empirical Bayes results in the case of non-identical components, Ph. D. Thesis, Michigan State University, East Lansing, Michigan.

Robbins, H. (1956). An empirical Bayes approach to statistics, *Proc. Third Berkeley on Symp. Math. Statist. Prob.*, Vol. 1, 157–163.

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

Susarla, V. (1982). Empirical Bayes theory, *Encyclopedia of Statistical Sciences*, (eds. S. Kotz and N. L. Johnson), Vol. 2, 440–503, Wiley, New York.

Suzuki, Y. (1975). On empirical Bayes models, *Proc. 40th ISI Session*, 372–385.

Van Ryzin, J. and Susarla, V. (1977). On the empirical Bayes approach to multiple decision problems, *Ann. Statist.*, **5**, 172–181.

Wetherill, G. B. (1975). *Sequential Methods in Statistics*, Chapman & Hall, London.