Ann. Inst. Statist. Math.
Vol. 48, No. 4, 711–730 (1996)

EMPIRICAL BAYES SEQUENTIAL ESTIMATION FOR EXPONENTIAL FAMILIES: THE UNTRUNCATED COMPONENT*

ROHANA J. KARUNAMUNI

Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

(Received July 20, 1995; revised January 19, 1996)

Abstract. We consider the empirical Bayes decision problem where the component problem is the sequential estimation of the mean θ of one-parameter exponential family of distributions with squared error loss for the estimation error and a cost $c > 0$ for each observation. The present paper studies the *untruncated* sequential component case. In particular, an untruncated asymptotically pointwise optimal sequential procedure is employed as the component. With sequential components, an empirical Bayes decision procedure selects both a stopping time and a terminal decision rule for use in the component with parameter θ. The goodness of the empirical Bayes sequential procedure is measured by comparing the asymptotic behavior of its Bayes risk with that of the component procedure as the number of past data increases to infinity. Asymptotic risk equivalence of the proposed empirical Bayes sequential procedure to the component procedure is demonstrated.

Key words and phrases: Empirical Bayes estimation, sequential components, asymptotically pointwise optimal, asymptotically optimal.

1. Introduction

The empirical Bayes (EB) decision problem of Robbins (1956, 1963, 1964) consists of a sequence of independent repetitions of a given component decision problem. At the n-th stage, data from the past in addition to the present stage are available with which to base a decision. Specifically, consider the component problem with observation $X \sim F_\omega$ taking values in χ, parameter $\omega \in \Omega$, action space A, decision rules $d \in D$, a loss $L(\omega, d(X)) \geq 0$, risk $R(\omega, d)$, priors $G \in \mathcal{G}$, Bayes risk $R(G, d)$, Bayes rules d_G and minimum Bayes risk $R(G)$. Here ω can represent anything, from an index for a finite set of distributions to a distribution to be estimated. The standard EB decision problem is usually formulated as follows: Let $(\omega_1, X_1), \dots, (\omega_n, X_n), \dots$ be i.i.d. with (ω, X) having distribution G on ω and,

* This research was supported in part by the Natural Sciences and Engineering Research Council of Canada under grant GP7987.

conditional on ω, F_ω on X. For each $n \geq 1$, $(X_1, \ldots, X_n) \sim F_G^n = F_G \times \cdots \times F_G$, where F_G denotes the G-mixture of the F_ω. A decision is to be made about ω_n using $(X_1, \ldots, X_n)$ with loss $L(\omega_n, t_n(X_1, \ldots, X_n))$. A sequence $\{t_n\}$ is said to be asymptotically optimal if

$$\lim_{n\to\infty} EL(\omega_n, t_n(X_1, \ldots, X_n)) = R(G) \quad \text{for all} \quad G \in \mathcal{G}.$$

Empirical Bayes problems with a wide variety of components have been treated in the literature; see, e.g., Berger (1985) and Maritz and Lwin (1989) for detailed bibliographies. The present paper deals with the sequential component case.

In recent years EB problems with sequential components have been studied by several authors including Laippala (1979, 1985), Martinsek (1987), Gilliland and Karunamuni (1988), Ghosh and Hoekstra (1989) and Karunamuni (1985, 1988, 1989, 1990). In the work of Laippala (1979, 1985), Gilliland and Karunamuni (1988) and Karunamuni (1985, 1988, 1989, 1990), the cost factor c was fixed and the risk of the empirical Bayes sequential decision (EBSD) procedure is compared with the risk of the component problem as the number of components increases. On the other hand, Martinsek (1987) and Ghosh and Hoekstra (1989) compare the risk of their respective EBSD procedures with the risk of the component procedure as c goes to zero. Their criterion is one of "asymptotic non-deficiency" (for the definition see Woodroofe (1982)). The approach of the present paper is to study the risk of the EBSD procedure as the number of components increases (the cost c is fixed), i.e., asymptotically optimal EBSD procedures are studied here. Gilliland and Karunamuni (1988) formally defined the sequential component for the *truncated* sequential decision problem, and demonstrated asymptotically optimal EBSD prodecures for some finite state components. Laippala (1979, 1985) and Karunamuni (1985, 1988, 1989, 1990) applied empirical Bayes methods for some infinite state sequential components. However, these authors considered again only the *truncated* case. In particular, they studied truncated procedures with myopic stopping rules of the "one-step-look-ahead" (OSLA) type in their respective component problems. Since truncated OSLA stopping rules are optimal or nearly optimal only in rather specialized circumstances, the Bayes risk of this procedure seems an inappropriate target to achieve. A goal in empirical Bayes theory is asymptotic optimality. For an envelope R^*, this means the construction of a sequence of stopping rules $\{N_n\}$ and decision rules $\{\delta_n\}$ where (N_n, δ_n) depends upon the data from the past as well as the present such that the risk of the sequential procedure (N_n, δ_n) converging to the envelope value R^* as the number of components goes to infinity. The facts that the Bayes risk of a truncated OSLA procedure is not an envelope risk and that the truncated procedures are not optimal in many of the most commonly considered examples are major deficiencies of the works of Laippala (1979, 1985) and Karunamuni (1985, 1988, 1989, 1990).

The purpose of the present paper is to study the *untruncated* case and to achieve the most stringent envelope risk—the Bayes risk of the optimal sequential procedure. The latter goal is achieved under some restrictive conditions, since, as in the Bayesian sequential decision problem, the construction of empirical Bayes sequential procedures that achieve the most stringent envelope is the most difficult construction. The component problem that is investigated in this paper is that of sequential estimation of the mean of a one-parameter exponential family of distributions with squared error loss for estimation error and a sampling cost $c(> 0)$ for each observation. Further, natural conjugate prior distributions on the parameter are assumed although such priors are not essential; see Remark 3.3 below. An untruncated asymptotically pointwise optimal procedure is employed as the sequential component. The next section describes the sequential component problem. The corresponding empirical Bayes sequential estimation problem is described in Section 3. The asymptotic behavior of the empirical Bayes sequential procedure, (N_n, δ_{N_n}), with respect to the component procedure, (N_c, δ_{N_c}), is given in Section 3 as well. Specifically, our main result establishes that $\lim_{n\to\infty} R_{N_n}(\pi_0) = R_{N_c}(\pi_0)$, where $R_{N_n}(\pi_0)$ and $R_{N_c}(\pi_0)$ denote Bayes risks (w.r.t. π_0) of the proposed EBSD and the component procedures, respectively. Some useful remarks and two examples are also included in Section 3. Proofs of the main results are detailed in Section 4.

2. The sequential component problem

We will use the notation and the setup of Woodroofe (1981) to describe the component problem that is the kernel of the empirical Bayes decision problem that is the subject of our investigation.

Let Ω be an interval and let $\{F_\omega : \omega \in \Omega\}$ denote a non-degenerate exponential family of probability distributions on the Borel subsets of $(-\infty, \infty)$: that is, suppose that

$$dF_\omega(x) = \exp\{\omega x - \psi(\omega)\}d\lambda(x), \qquad -\infty < x < \infty, \ \omega \in \Omega, \tag{2.1}$$

where Ω is the natural parameter space of the family. We assume that Ω is open, say $\Omega = (\underline{\omega}, \overline{\omega})$, where $-\infty \le \underline{\omega} < \overline{\omega} \le \infty$, and that λ is σ-finite. The distributions F_ω, $\omega \in \Omega$ are absolutely continuous w.r.t. λ. We denote the common closed convex support by χ, and we write χ° for the interior of χ. It is well known that if X is a r.v. with distribution F_ω, where $\omega \in \Omega$, then the mean and variance of X are $E_\omega(X) = \psi'(\omega)$ and $\mathrm{Var}_\omega(X) = \psi''(\omega)$; see, e.g., Lehmann ((1986), Section 2.7). Of special interest to us will be the mean which we write as

$$\theta = \psi'(\omega) \tag{2.2}$$

throughout.

Now let $\mathcal{G}$ be the class of conjugate prior distributions π_0 (see Diaconis and Ylvisaker (1979)) on the Borel sets of Ω; that is, suppose

$$d\pi_0(\omega) = [C(r_0, \mu_0)]^{-1} \exp\{r_0\mu_0\omega - r_0\psi(\omega)\}d\omega, \qquad \omega \in \Omega, \tag{2.3}$$

where

$$0 < C(r_0, \mu_0) = \int_\Omega \exp\{r_0\mu_0\omega - r_0\psi(\omega)\}d\omega < \infty,$$

with $r_0 > 0$ and $\mu_0 \in \chi^\circ$. Now suppose that $E[\psi''(\omega)] < \infty$, so that θ has finite variance, and consider the problem of sequentially estimating θ (see (2.2)) with squared error loss for estimation error and a cost $c > 0$ for each observation $X_1, X_2, \ldots,$ where $X_1, X_2, \ldots$ are conditionally i.i.d. with common distribution F_ω, given $\omega(\omega \sim \pi_0)$. The Bayesian sequential decision problem is then to find a stopping time t (recall that a stopping time is a r.v. t which takes values $0, 1, 2, \ldots, \infty$ and has the properties $P\{t < \infty\} = 1$ and $\{t = n\} \in \mathcal{A}_n$ for all n, where $\mathcal{A}_n = \sigma\{X_1, \ldots, X_n\}$ denotes the σ-algebra generated by $X_1, \ldots, X_n$ for $n \geq 1$ and $\mathcal{A}_o$ denotes the trivial σ-algebra) and an $\mathcal{A}_t$-measurable function $\delta_t = \delta_t(X_1, \ldots, X_t)$ for which the

$$\text{Bayes risk } (t, \delta_t) = E\{(\delta_t - \theta)^2 + ct\} \tag{2.4}$$

is minimized. It is well known that for any stopping time t, the Bayes risk is minimized by letting $\delta_t = E(\theta \mid \mathcal{A}_t)$; see, e.g., Chapter 7 of Ferguson (1967) or Berger (1985). Moreover, it follows from Theorems 4.4 and 4.5 of Chow *et al.* (1971) that an optimal stopping time, one which minimizes (2.4), exists for each $c > 0$. Although in principle the optimal stopping rule may be derived by a backward induction argument, the exact determination of the optimal stopping time appears to be rather difficult in practice. However, it is quite well known that asymptotically pointwise optimal (A.P.O.) rules defined by Bickel and Yahav (1967, 1969) are solutions to the minimization problem as $c \to 0$. (See also Woodroofe (1982) who gives stronger optimality results in the context of sequential estimation for one-parameter exponential families.) In practice then for small c, A.P.O. rules are good approximations to the optimal stopping rules. Many procedures are reported in the literature which are A.P.O. in the sense of Bickel and Yahav (1967, 1969). In this paper we use the A.P.O. myopic stopping rule of the type of Shapiro and Wardrop (1980). They suggest the stopping time

$$N_c = \inf\{k \geq 1 : U_k \leq c r_k r_{k+1}\}, \tag{2.5}$$

where

$$U_k = E[\psi''(\omega) \mid \mathcal{A}_k] = \int_\Omega \psi'' d\pi_k \tag{2.6}$$

with π_k denoting the posterior distribution of ω given $\mathcal{A}_k = \sigma\{X_1, \ldots, X_k\}$; that is,

$$d\pi_k(\omega) = [C(r_k, \mu_k)]^{-1} \exp\{r_k \mu_k \omega - r_k \psi(\omega)\} d\omega, \tag{2.7}$$

where $C(r_k, \mu_k) = \int_\Omega \exp\{r_k \mu_k \omega - r_k \psi(\omega)\} d\omega$,

$$r_k = r_0 + k, \quad \mu_k = (r_0 \mu_0 + S_k)/r_k \quad \text{and} \quad S_k = \sum_{i=1}^{k} X_i, \tag{2.8}$$

for $k \geq 1$. In this paper, we shall employ the pair

$$(N_c, \delta_{N_c}) \tag{2.9}$$

as the sequential estimation procedure for the present component problem, where N_c is defined by (2.5) and δ_{N_c} is defined by

$$\delta_{N_c} = (r_0\mu_0 + S_{N_c})/(r_0 + N_c) \tag{2.10}$$

with $S_{N_c} = \sum_{i=1}^{N_c} X_i$. We shall assume that the cost c per observation is small enough to permit the sequential procedure (2.9) to represent a resonably good approximation to the optimal (Bayes) decision procedure. Let $R_{N_c}(\pi_0)$ denote the Bayes risk (w.r.t. π_0) of the procedure (N_c, δ_{N_c}). Then,

$$R_{N_c}(\pi_0) = E\{(\delta_{N_c} - \theta)^2 + cN_c\}. \tag{2.11}$$

Let $R^*(\pi)$ denote the Bayes risk of the optimal sequential estimation procedure for the present problem. Then

$$R^*(\pi_0) = \inf\{\text{Bayes risk } (s, \delta_s) : s \text{ is any stopping time}\}, \tag{2.12}$$

where the infimum extends over all stopping times s. We call $R^*(\pi_0)$ the Bayes envelope risk of the sequential component problem. The asymptotic optimality that is typically proved in the standard empirical Bayes problem is the convergence of the empirical Bayes risk to the envelope risk $R^*(\pi_0)$.

Finally, for later use, it is important to note that $\{(U_k, \mathcal{A}_k)\}$ is a uniformly integrable martingale for which $U_k \to \psi''(\omega)$ w.p. 1 as $k \to \infty$ (see, e.g., Lemma 7.6.1 and Theorem 7.6.2 of Ash (1972)). Furthermore, in spirit of Theorem 1 of Shapiro and Wardrop (1980) we can state the following result.

THEOREM 2.1.
(i) *The stopping rule* (2.5) *is the myopic rule.*
(ii) $EN_c < \infty$ *for* $c > 0$.
(iii) *If* $(r_{k+1}r_k)^{-1}U_k$ *is non-increasing in* k, *then* (2.5) *is the optimal* (*Bayes*) *stopping rule.*

3. The empirical Bayes problem

Suppose now that the prior π_0 is not completely known, that is, r_0 or μ_0 or both are unknown, and consider the case where we have available sample information from the past. Then the EB approach of Robbins' (1956, 1963, 1964) could be utilized to approximate the (A.P.O.) sequential procedure (N_c, δ_{N_c}) defined by (2.9) based on the sample information from the past along with the present.

Suppose, for $n > 1$,

$$\boldsymbol{X}^i = (X_{i1}, \ldots, X_{iK_i}), \qquad i = 1, \ldots, (n-1)$$

denote the data available from the past, where (ω_i, X_{ij}) are i.i.d. with the same distribution as (ω, X_1), and the (ω_i, X_{ij}) are independent of ω and the current data vector $\boldsymbol{X} = (X_1, X_2, \ldots)$, $1 \le j \le K_i$, $1 \le i \le (n-1)$. If the data $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^{n-1}$ represent information available from the past experiences of the same component problem, then K_i equals the sample size N_i of the i-th experience of the problem, $1 \le i \le n-1$. If, however, one has available these data from some auxiliary information, then K_i's are some known fixed positive integers. In this case, $\boldsymbol{X}^i$'s are independent random vectors. The EB approach attempts to construct a decision procedure to estimate θ_n ($\theta_1, \ldots, \theta_{n-1}$ remains unobservable) at stage n based on the past data as well as the present data.

In order to construct our EBSD procedure, let us suppose that

$$0 < \hat{r}_0 = \hat{r}_0(\boldsymbol{X}^1, \ldots, \boldsymbol{X}^{n-1})$$

and

$$\hat{\mu}_0 = \hat{\mu}_0(\boldsymbol{X}^1, \ldots, \boldsymbol{X}^{n-1})$$

denote estimators of r_0 and μ_0 respectively, based on the past data $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^{n-1}$. Based on these estimators and motivated by (2.5) and (2.10), we define our empirical Bayes estimate of θ_n as

$$\hat{\delta}_{N_n} = \{\hat{r}_0\hat{\mu}_0 + S_{N_n}\}\{\hat{r}_0 + N_n\}^{-1} \tag{3.1}$$

and the empirical Bayes sample size N_n as

$$N_n = \inf\{k \ge 1 : \hat{U}_k \le c(\hat{r}_0 + k)(\hat{r}_0 + k + 1)\} \tag{3.2}$$

where $S_{N_n} = \sum_{i=1}^{N_n} X_i$ and $\hat{U}_k$ is defined by (cf. (2.6))

$$\hat{U}_k = \frac{1}{C(\hat{r}_k, \hat{\mu}_k)} \int_\Omega \psi''(\omega)\exp\{(\hat{r}_0\hat{\mu}_0 + S_k)\omega - \hat{r}_k\psi(\omega)\}d\omega \tag{3.3}$$

with $\hat{r}_k = \hat{r}_0 + k$, $S_k = \sum_{i=1}^k X_i$, and

$$C(\hat{r}_k, \hat{\mu}_k) = \int_\Omega \exp\{(\hat{r}_0\mu_0 + S_k)\omega - \hat{r}_k\psi(\omega)\}d\omega.$$

The empirical Bayes sample size N_n is obtained sequentially and then θ_n is estimated by $\hat{\delta}_{N_n}$ based on N_n. Sequential stopping times of the type above were initiated by Robbins (1959) and later developed by others; see, e.g., Sen (1981) and Woodroofe (1982).

Let $R_{N_n}(\pi_0)$ denote the unconditional Bayes risk (w.r.t. π_0) of the EBSD procedure $(N_n, \hat{\delta}_{N_n})$ defined by (3.1) and (3.2). Then

$$R_{N_n}(\pi_0) = E\{(\hat{\delta}_{N_n} - \theta)^2 + cN_n\}, \tag{3.4}$$

where E denotes expectation w.r.t. all of the random variables involved in defining N_n and $\hat{\delta}_{N_n}$, and w.r.t. the random variable ω. As we described earlier, the performance of the EBSD procedure $(N_n, \hat{\delta}_{N_n})$ is measured by comparing the risks $R_{N_n}(\pi_0)$ and $R_{N_c}(\pi_0)$ asymptotically, as $n \to \infty$, where $R_{N_c}(\pi_0)$ is given by

(2.11). The most natural way to compare the expressions $R_{N_n}(\pi_0)$ and $R_{N_c}(\pi_0)$ is to show that

$$\lim_{n\to\infty} R_{N_n}(\pi_0) = R_{N_c}(\pi_0),$$

i.e., the asymptotic risk equivalence of the EBSD procedure to the component sequential estimation procedure (N_c, δ_{N_c}). The above result is established in Theorem 3.1 below under some assumptions on the estimators $\hat{r}_0$ and $\hat{\mu}_0$. When (N_c, δ_{N_c}) is the optimal procedure then the asymptotic optimality of the EBSD procedure $(N_n, \hat{\delta}_{N_n})$ easily follows from Theorem 3.1. This result is established in Corollary 3.1. below.

THEOREM 3.1. *Let the pair (N_c, δ_{N_c}) be defined by* (2.5) *and* (2.10). *Let the EBSD procedure $(N_n, \hat{\delta}_{N_n})$, defined by* (3.1) *and* (3.2), *be based on the estimators $\hat{r}_0$ and $\hat{\mu}_0$ satisfying the conditions*

$$\text{(i)}\quad E(\hat{r}_0 - r_0)^2 \to 0 \qquad as \quad n \to \infty$$

and

$$\text{(ii)}\quad E(\hat{\mu}_0 - \mu_0)^2 \to 0 \qquad as \quad n \to \infty,$$

where E denotes expectation w.r.t. all of the random variables involved in defining $\hat{r}_0$ and $\hat{\mu}_0$. Suppose that $E(\psi''(\omega))^2 < \infty$. Then

$$\lim_{n\to\infty} R_{N_n}(\pi_0) = R_{N_c}(\pi_0), \tag{3.5}$$

where $R_{N_c}(\pi_0)$ and $R_{N_n}(\pi_0)$ are defined by (2.11) *and* (3.4), *respectively.*

COROLLARY 3.1. *Suppose that the assumptions of Theorem* 3.1 *hold. Moreover, suppose that $(r_{k+1}r_k)^{-1}U_k + ck$ is non-increasing in k, where U_k and r_k are defined by* (2.6) *and* (2.8), *respectively, $k \geq 1$. Then, $\lim_{n\to\infty} R_{N_n}(\pi_0) = R_{N_c}(\pi_0)$, and thus $(N_n, \hat{\delta}_{N_n})$ is asymptotically optimal.*

Remark 3.1. The EBSD procedure employed in Theorem 3.1 is based on mean square consistent estimates $\hat{\mu}_0$ and $\hat{r}_0$ of μ_0 and r_0, respectively. We now exhibit a class of such estimators constructed using the past data obtained from some auxiliary information. From (2.1), (2.2) and (2.3), we notice that

$$E(\theta) = \int_\Omega \psi'' d\pi_0 = \mu_0$$

and

$$\mathrm{Var}(\theta) = E(\theta - \mu_0)^2 = r_0^{-1} r^2$$

where $r^2 = E[\psi''(\omega)]$. Let us denote $\sigma^2 = r_0^{-1} r^2$. Then $r_0 = r^2/\sigma^2$. Now, in order to estimate μ_0 and r_0 based on the auxiliary data $\boldsymbol{X}^i = (X_{i1}, \ldots, X_{iK_i})$, $i = 1, \ldots, n-1$, we define $m = (n-1)$,

$$(\hat{r}_0)^{-1} = \max\left\{0, \left(\frac{(m-1)MSB}{(m-3)MSW} - 1\right)(m-1)g^{-1}\right\} \tag{3.6}$$

and

$$\hat{\mu}_0 = \begin{cases} \sum_{j=1}^{m}(1-\hat{B}_j)\bar{X}_{K_j} \Big/ \sum_{j=1}^{m}(1-\hat{B}_j) & \text{if } \hat{r}_0 \neq 0 \\ m^{-1}\sum_{j=1}^{m}\bar{X}_{K_j} & \text{if } \hat{r}_0 = 0, \end{cases} \tag{3.7}$$

where $\hat{B}_j = \{1+N_j(\hat{r}_0)^{-1}\}^{-1}$, $\bar{X}_{K_j} = K_j^{-1}\sum_{i=1}^{K_j} X_{ji}$, $g = \sum_{j=1}^{m} K_j - (\sum_{j=1}^{m} K_j^2)/\sum_{j=1}^{m} K_j$, $MSW = (\sum_{j=1}^{m} K_j - m)^{-1}\sum_{j=1}^{m}\sum_{i=1}^{K_j}(X_{ji} - \bar{X}_{K_j})^2$ and $MSB = (m-1)^{-1}\sum_{j=1}^{m} K_j(\bar{X}_{K_j} - \bar{\bar{X}}_m)^2$ with $\bar{\bar{X}}_m = (\sum_{j=1}^{m} K_j - m)^{-1}\sum_{j=1}^{m} K_j\bar{X}_{K_j}$. These estimators are motivated by the work of Ghosh and Lahiri (1987). Since the N_j's are fixed positive integers in the case of auxiliarly data, $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^{n-1}$ are independent random vectors, and (ω_i, X_{ij}) are i.i.d. with the same distribution as (ω, X_1). Therefore, the mean square consistency of $\hat{r}_0$ and $\hat{\mu}_0$ can be established along the lines of Ghosh and Lahiri (1987). There are a number of situations in which one may be able to observe such types of auxiliary data; see Martinsek (1987) for some applications. As one of the referees pointed out, one drawback of such an approach in an empirical Bayes setup is that one must wait until the "final problem" to use the accumulated information to construct a good stopping rule.

If $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^{n-1}$ represent the data from the past experiences of the same component problem, then of course the independence of these random vectors is lost and the consistency proofs of (3.6) and (3.7) may be rather tedious. Our Monte Carlo studies show, however, that the estimators $\hat{r}_0$ and $\hat{\mu}_0$ approach respective true values of r_0 and μ_0 as n increases. Thus, one may expect them to be useful in applications. Nevertheless, one other method (not so efficient) of constructing consistent estimators based on the past data is as follows: It is reasonable to assume that $2 \leq N_c < \infty$ in many applications, where N_c is defined by (2.5). Now re-define N_n given by (3.2) as $N_n = \inf\{k \geq 2 : \hat{U}_k \leq c(\hat{r}_0 + k)(\hat{r}_0 + k + 1)\}$, i.e., at least two independent observations are made before making a decision. As a result, the random vectors consisting only with the first two observations from the past realizations form an independent sequence, and have the same features as auxiliary data. Now, these random vectors can be used in the expressions (3.6) and (3.7) above, and the consistency can be easily established.

Remark 3.2. The condition that $(r_{k+1}r_k)^{-1}U_k + ck$ is nonincreasing in k is known as the 'monotone' total cost case. Shapiro and Wardrop (1980) considered the monotone cost case for a richer class of loss functions (including the squared error decision loss) related to the Fisher information and linear sampling cost. They illustrated a number of commonly used examples to show that the monotone cost case (and hence the optimality of the myopic rule) can be attained. Results of the present paper can easily be extended to cover such cases as well, since the Bayes estimator of θ and the myopic stopping rule of Shapiro and Wardrop are very similar to the one used here.

Remark 3.3. Asymptotic results of the present paper can be extended easily to more general families of prior distributions, those which are not conjugate priors. However, some restrictions are needed. For example, if the posterior Bayes risk, based on a fixed sample of size k, can be expanded in terms of a uniformly integrable martingale Y_k and a uniformly integrable super-martingale Z_k, namely,

$$\mathrm{Var}(\theta \mid A_k) = k^{-1}Y_k + k^{-2}Z_k, \qquad k \geq 1,$$

then the overall posterior Bayes risk due to sampling cost and estimation error is

$$R_k(c) = k^{-1}Y_k + ck + k^{-2}Z_k, \qquad k \geq 1 \tag{3.8}$$

and $R_t(c) = t^{-1}Y_t + ct + t^{-2}Z_t$ for any stopping time t. Expression (3.8), in turn, suggests the myopic stopping rule

$$t_c = \inf\{k \geq 1 : Y_k \leq ck(k+1)\}, \qquad c > 0 \tag{3.9}$$

for use in the component problem. Using the semi-martingale nature of the expansion (3.8), Rehailia (1984) showed that stopping rules of the type of Bickel and Yahav (1967) are A.P.O. as $c \to 0$. Following the arguments of Shapiro and Wardrop (1980), it may be possible to show that the stopping rule (3.9) is also A.P.O. as $c \to 0$. Details of such a result are not yet reported in the literature. The EB stopping rule corresponding to (3.9) can easily be defined and the asymptotic results will essentially be the same as those of the present paper.

Remark 3.4. The setup of the present paper falls into the category of *parametric empirical Bayes*. More specifically, it is in the sub-category of parametric empirical Bayes sequential estimation. Parametric models are becoming increasingly popular in empirical Bayes methods, due in large part to their compromising behavior between frequentist modelling and Bayesian modelling; see, e.g., Morris (1983*a*, 1983*b*). More references can be found in Berger (1985) and Maritz and Lwin (1989). One disadvantage of such models in applications is, however, the possible effects of wrong guess of the prior. Study of Bayesian analysis to possible misspecification of the prior distribution is one of *Bayesian robustness* issues, an important element in Bayesian analysis. The most commonly used technique for investigating 'prior robustness' is simply to try different reasonable priors and see what happens. This is often called *sensitivity analysis*. Though of considerable interest, investigation of such a study is beyond the scope of the present paper. There is a huge literature on Bayesian robustness; see, e.g., Berger (1985), Subsection 4.7, for methodologies and references on the subject.

Closely related to a parametric empirical Bayes procedure is the *hierarchical Bayes* procedure which models the prior distribution in stages. In the first stage, conditional on λ, parameter ω is assumed distributed according to a prior $\pi_0(\omega \mid \lambda)$. In the second stage, a prior (often improper) distribution is assigned to λ. This is an example of a two-stage prior. The idea can be generalized to multi-stage prior; see, e.g., Good (1965) and Lindley and Smith (1972). For details on hierarchical Bayes settings of sequential estimation, see Ghosh (1991), Ghosh and Hoekstra (1989) and Hoekstra (1989).

The nature of the stopping rules N_c and N_n will now be illustrated by two special cases considered by Woodroofe (1981). More examples can be provided; see, e.g., Shapiro and Wardrop (1980).

Example 1. Suppose that F_ω has density $f_\omega(x) = |\omega| \exp(\omega x)$ for $x > 0$ and $-\infty < \omega < 0$, w.r.t. Lebesgue measure. Then $\{F_\omega : -\infty < \omega < 0\}$ form an exponential family with $\psi(\omega) = \log(1/|\omega|)$; and the mean and the variance of F_ω are $\theta = \psi'(\omega) = 1/|\omega|$ and $\psi''(\omega) = 1/\omega^2 = \theta^2$ for $-\infty < \omega < 0$. The conjugate prior distributions are gamma distributions for $|\omega|$ with densities

$$\pi_{0,1}(\omega) = (\Gamma(a_0))^{-1} b_0^{a_0} |\omega|^{a_0-1} e^{+b_0\omega}, \qquad -\infty < \omega < 0,$$

with shape parameter $a_0 = r_0 + 1 > 1$ and scale parameter $b_0 = r_0/\mu_0 > 0$. Letting $S_k = \sum_1^k X_i$, one easily sees that $U_k = E[\psi''(\omega) \mid \mathcal{A}_k] = E[\theta^2 \mid \mathcal{A}_k] = (b_0 + S_k)^2/(a_0 + k - 1)(a_0 + k - 2)$. Then, for this example the stopping time N_c defined by (2.5) takes the form

$$N_{c,1} = \inf\{k \geq 1 : (r_0\mu_o^{-1} + S_k)^2 \leq c(r_0 + 1 + k)(r_0 + k)^2(r_0 + k - 1)\}.$$

It is easy to show that $E[\psi''(\omega)]^2 < \infty$ if and only if $a_0 > 4$. The corresponding EB stopping time is obtained by replacing r_0 and μ_0 by $\hat{r}_0$ and $\hat{\mu}_0$, respectively:

$$N_{n,1} = \inf\{k \geq 1 : (\hat{r}_0\hat{\mu}_0^{-1} + S_k)^2 \leq c(\hat{r}_0 + 1 + k)(\hat{r}_0 + k)^2(\hat{r}_0 + k - 1)\}.$$

The performance of $N_{n,1}$ was studied using a simulation study as n increases. Two separate cases were considered with the following specifications: (i) $a_0 = 5$, $b_0 = 5$, $c = 0.0001$, and (ii) $a_0 = 5$, $b_0 = 10$, $c = 0.01$. For the chosen sample, the values of $N_{c,1}$ for the two cases (i) and (ii) were 115 and 13, respectively. Figures 1 and 2 represent the summary of results which exhibit the behavior of $N_{n,1}$ for the cases (i) and (ii), respectively, as n ranges from 11 and 200.

Example 2. Suppose that X has a Bernoulli distribution with parameter θ, where $0 < \theta < 1$; that is, suppose that $X = 0$ and 1 with probabilities $1 - \theta$ and θ, where θ is unknown. Then the distributions of X form an exponential family with $\omega = \log\theta - \log(1-\theta)$ and $\psi(\omega) = \log(1/(1-\theta))$. The mean and the variance of X are θ and $\psi''(\omega) = \theta(1-\theta)$. The conjugate prior distributions are beta distributions for θ with densities

$$\pi_{0,2}(\theta) = \frac{\Gamma(a_0 + b_0)}{\Gamma(a_0)\Gamma(b_0)} \theta^{a_0-1}(1-\theta)^{b_0-1}, \qquad 0 < \theta < 1,$$

where $r_0 = a_0 + b_0$ and $\mu_0 = a_0/(a_0 + b_0)$. Letting $S_k = X_1 + \cdots + X_k$, we find that $U_k = E[\psi''(\omega) \mid \mathcal{A}_k] = E[\theta(1-\theta) \mid \mathcal{A}_k] = (a_0 + S_k)(b_0 + k - S_k)/(a_0 + b_0 + k - 1)(a_0 + b_0 + k)$. In this case the stopping time N_c takes the form

$$\begin{aligned} N_{c,2} = \inf\{k \geq 1 :& (r_0\mu_0 + S_k)(r_0 - r_0\mu_0 + k - S_k) \\ &\leq c(r_0 + k)^2(r_0 + k - 1)(r_0 + k + 1)\}. \end{aligned}$$

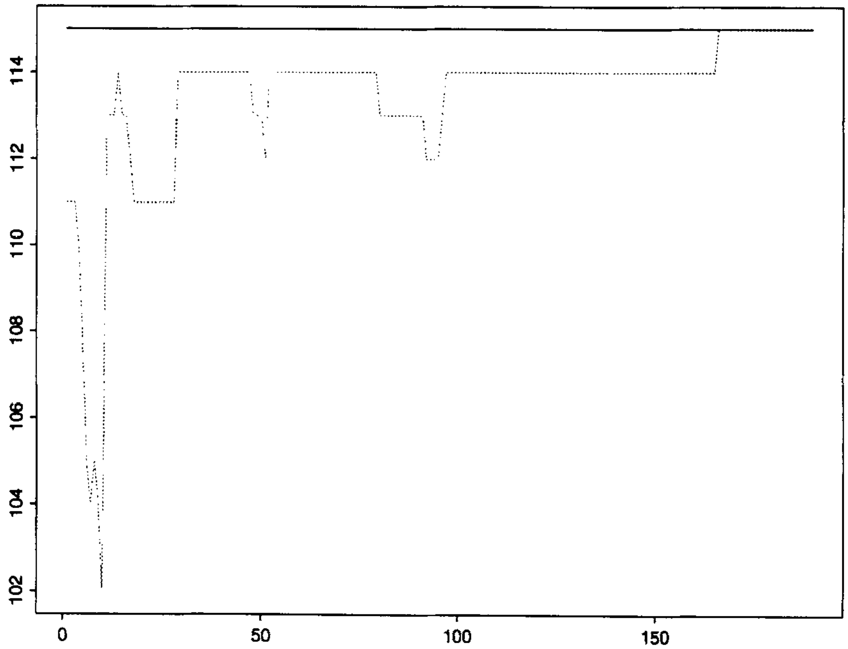

Fig. 1. Behaviors of $N_{n,1}$ (dotted line) and $N_{c,1} = 115$ (solid line) for case (i).

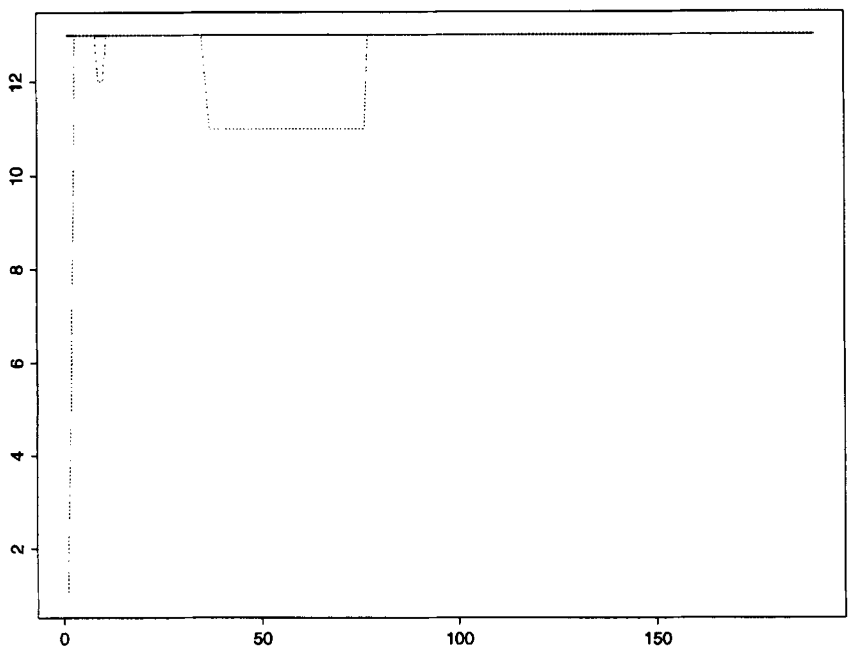

Fig. 2. Behaviors of $N_{n,1}$ (dotted line) and $N_{c,1} = 13$ (solid line) for case (ii).

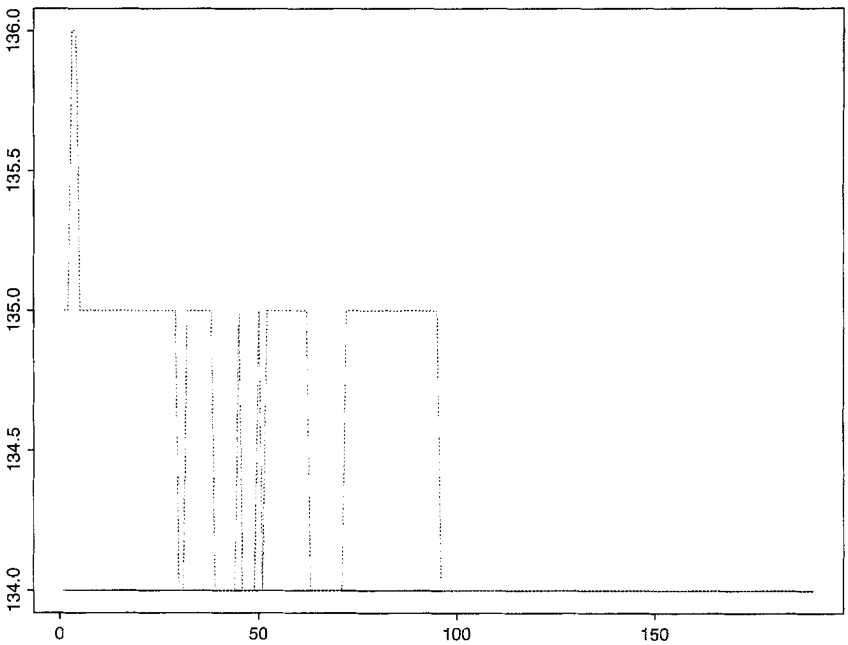

Fig. 3. Behaviors of $N_{n,2}$ (dotted line) and $N_{c,2} = 134$ (solid line) for case (a).

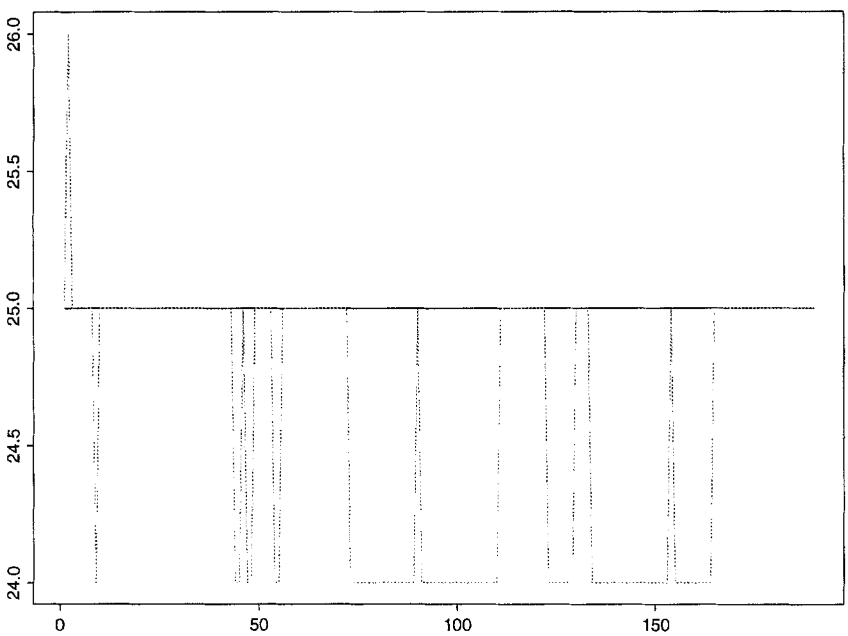

Fig. 4. Behaviors of $N_{n,2}$ (dotted line) and $N_{c,1} = 25$ (solid line) for case (b).

Note that, in this case, $E[\psi''(\omega)]^2 < \infty$ for any $a_0 > 0$ and $b_0 > 0$. The corresponding EB stopping time defined by

$$N_{n,2} = \inf\{k \geq 1 : (\hat{r}_0\hat{\mu}_0 + S_k)(\hat{r}_0 - \hat{r}_0\hat{\mu}_0 + k - S_k) \\ \leq c(\hat{r}_0 + k)^2(\hat{r}_0 + k - 1)(\hat{r}_0 + k + 1)\}.$$

Again, the behavior of $N_{n,2}$ was examined using a simulation study for two separate cases: (a) $a_0 = 5$, $b_0 = 2$, $c = 0.0001$, and (b) $a_0 = 2$, $b_0 = 3$, $c = 0.0001$. For the chosen sample, the values of $N_{c,2}$ for the two cases (a) and (b) were 134 and 25, respectively. Figures 3 and 4 summarize the results.

Overall, the simulation results indicate that the nature of convergence of empirical Bayes stopping times, N_n's, to corresponding component stopping times, N_c's, is fairly satisfactory for moderate values of n. For small values ($n < 30$), the convergence is slightly conservative.

4. Proofs

To prove Theorem 3.1 and Corollary 3.1 we need a few lemmas. For notational convenience, we let $[A]$ denote the indicator function of A and the arguments of functions will not be exhibited whenever they are clear from the context. The basic idea of the proof is to decompose the difference $R_{N_n}(\pi_0) - R_{N_c}(\pi_0)$ into portions corresponding to the cases $[N_n = N_c]$, $[N_n > N_c]$ and $[N_n < N_c]$, and then to study the asymptotic behaviors of each portion separately. By the definitions of N_c (see (2.5)), we first note that for $k \geq 1$,

$$[N_c = k] = [\alpha_1 < 0] \cdots [\alpha_{k-1} < 0][\alpha_k \geq 0], \tag{4.1}$$

where

$$\alpha_k = cr_kr_{k+1} - U_k, \tag{4.2}$$

with U_k and r_k are given by (2.6) and (2.8), respectively. Similarly, using the definition of N_n (see (3.2)) we get for $k \geq 1$,

$$[N_n = k] = [\hat{\alpha}_1 < 0] \cdots [\hat{\alpha}_{k-1} < 0][\hat{\alpha}_k \geq 0], \tag{4.3}$$

where

$$\hat{\alpha}_k = c(\hat{r}_0 + k)(\hat{r}_0 + k + 1) - \hat{U}_k, \tag{4.4}$$

with $\hat{r}_0$ is as defined in Theorem 3.1 and $\hat{U}_k$ is given by (3.3). Now observe that the difference $R_{N_n}(\pi_0) - R_{N_c}(\pi_0)$ (see (2.11) and (3.2)) can be written as

$$\begin{aligned} R_{N_n}(\pi_0) - R_{N_c}(\pi_0) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} E[N_n = i][N_c = j]\{(\hat{\delta}_i - \theta)^2 - (\delta_j - \theta)^2 + c(i - j)\} \\ &= \hat{J} + \hat{K} + \hat{L}, \end{aligned} \tag{4.5}$$

where

$$\hat{J} = \sum_{i=1}^{\infty} E[N_n = i][N_c = i]\{(\hat{\delta}_i - \theta)^2 - (\delta_i - \theta)^2\}, \tag{4.6}$$

$$\hat{K} = \sum_{j=2}^{\infty}\sum_{i=1}^{j-1} E[N_n = i][N_c = j]\{(\hat{\delta}_i - \theta)^2 - (\delta_j - \theta)^2 + c(i-j)\}, \tag{4.7}$$

and

$$\hat{L} = \sum_{j=2}^{\infty}\sum_{i=1}^{j-1} E[N_n = j][N_c = i]\{(\hat{\delta}_j - \theta)^2 - (\delta_i - \theta)^2 + c(j-i)\}, \tag{4.8}$$

with E denotes expectation w.r.t. all of the random variables involved. We will show that $\lim_{n\to\infty}\hat{J} = 0$, $\lim_{n\to\infty}\hat{K} = 0$ and $\lim_{n\to\infty}\hat{L} = 0$. Before proving these results we first give two preliminary lemmas. Lemma 4.1 is simply a version of the dominated convergence theorem w.r.t. the counting measure on positive integers.

LEMMA 4.1. *For each $n = 1, 2, \ldots$, let $\{a_1^{(n)}, a_2^{(n)}, \ldots\}$ be a sequence of real numbers. If $|a_k^{(n)}| \le b_k$, for all k and n, and $\sum_{k=1}^{\infty} b_k < \infty$, then*

$$\lim_{n\to\infty}\sum_{k=1}^{\infty} a_k^{(n)} = \sum_{k=1}^{\infty}\lim_{n\to\infty} a_k^{(n)}. \tag{4.9}$$

LEMMA 4.2. *Under the hypotheses of Theorem 3.1, $\hat{\alpha}_k \xrightarrow{P} \alpha_k$ as $n \to \infty$, for $k \ge 1$, where α_k and $\hat{\alpha}_k$ are defined by (4.2) and (4.4), respectively, and $\xrightarrow{P}$ denotes convergence in probability w.r.t. the past data, conditional on the current data.*

PROOF. Since $\hat{r}_0 \xrightarrow{P} r_0$ as $n \to \infty$, it is enough to show that $\hat{U}_k \xrightarrow{P} U_k$ as $n \to \infty$, for $k \ge 1$. Note that U_k and $C(r_k, \mu_k)$ (see circa (2.7)) are continuous functions of r_0 and μ_0 by Theorem 2.9 of Lehmann (1986). Now the result follows from $\hat{r}_0 \xrightarrow{P} r_0$ and $\hat{\mu}_0 \xrightarrow{P} \mu_0$ as $n \to \infty$. □

LEMMA 4.3. *Under the hypotheses of Theorem 3.1,*

$$\lim_{n\to\infty}\hat{J} = 0. \tag{4.10}$$

PROOF. From (4.6) we can write

$$\begin{aligned}\hat{J} = {} & \sum_{i=1}^{\infty} E[N_n = i][N_c = i](\hat{\delta}_i - \delta_i)^2 \\ & + 2\sum_{i=1}^{\infty} E[N_n = i][N_c = i](\hat{\delta}_i - \delta_i)(\delta_i - \theta).\end{aligned} \tag{4.11}$$

Now use a conditional expectation argument to show that the second term on the right hand side of (4.11) is equal to zero as follows:

$$\begin{aligned}
&E\{E([N_n = i][N_c = i](\hat{\delta}_i - \delta_i)(\delta_i - \theta) \mid \mathcal{A}_i, \boldsymbol{X}^1, \ldots, \boldsymbol{X}^{n-1})\}\\
&\quad = E[N_n = i][N_c = i](\hat{\delta}_i - \delta_i)E\{(\delta_i - \theta) \mid \mathcal{A}_i, \boldsymbol{X}^1, \ldots, \boldsymbol{X}^{n-1}\}\\
&\quad = E[N_n = i][N_c = i](\hat{\delta}_i - \delta_i)E\{(\delta_i - \theta) \mid \mathcal{A}_i\}\\
&\quad = 0,
\end{aligned}$$

since θ and the current information, $\mathcal{A}_i$, are independent of the past data $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^{n-1}$ and since $\delta_i = E(\theta \mid \mathcal{A}_i)$ for $i \geq 1$. Therefore, from (4.11) we get

$$|\hat{J}| \leq \sum_{i=1}^{\infty} E(\hat{\delta}_i - \delta_i)^2. \tag{4.12}$$

By the definitions of $\hat{\delta}_i$ and δ_i (see (2.10) and (3.1)), we obtain

$$\begin{aligned}
E(\hat{\delta}_i - \delta_i)^2 \leq{}& 3(r_0\mu_0)^2 i^{-4} E(\hat{\mu}_0 - \mu_0)^2 + 3\mu_0^2 i^{-2} E(\hat{r}_0 - r_0)^2\\
&+ 3i^{-4} ES_i^2(\hat{r}_0 - r_0)^2,
\end{aligned} \tag{4.13}$$

the preceding inequality is obtained by applying the trivial inequality $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$. Now, since S_i and $\hat{r}_0$ are independent, we obtain $i^{-4}ES_i^2(\hat{r}_0 - r_0)^2 \leq i^{-2}EX_1^2 E(\hat{r}_0 - r_0)^2$ using the inequality $ES_i^2 \leq i^2 EX_1^2$. Also $EX_1^2 < \infty$, since $E(\psi''(\omega))^2 < \infty$. Combining these facts, (4.12) and (4.13) we get

$$|\hat{J}| \leq K\left\{E(\hat{\mu}_0 - \mu_0)^2 \sum_{i=1}^{\infty} i^{-4} + E(\hat{r}_0 - r_0)^2 \sum_{i=1}^{\infty} i^{-2}\right\}, \tag{4.14}$$

where K is a positive constant. Now $\lim_{n\to\infty} \hat{J} = 0$ follows from $E(\hat{\mu}_0 - \mu_0)^2 \to 0$ and $E(\hat{r}_0 - r_0)^2 \to 0$ as $n \to \infty$. □

LEMMA 4.4. *Under the hypotheses of Theorem* 3.1,

$$\lim_{n\to\infty} \hat{K} = 0. \tag{4.15}$$

PROOF. Write $\hat{K} = \hat{K}_1 + \hat{K}_2$, where

$$\hat{K}_1 = \sum_{j=2}^{\infty}\sum_{i=1}^{j-1} E[N_n = i][N_c = j]\{(\delta_i - \theta)^2 - (\delta_j - \theta)^2 + c(i - j)\} \tag{4.16}$$

and

$$\hat{K}_2 = \sum_{j=2}^{\infty}\sum_{i=1}^{j-1} E[N_n = i][N_c = j]\{(\hat{\delta}_i - \theta)^2 - (\delta_i - \theta)^2\}. \tag{4.17}$$

Following a similar argument that we used to obtain (4.12), we can show that $|\hat{K}_2| \le \sum_{i=1}^{\infty} E(\hat{\delta}_i - \delta_i)^2$. It can be easily shown now that $\lim_{n\to\infty} \hat{K}_2 = 0$ using similar steps to show that $\lim_{n\to\infty} \hat{J} = 0$. To study the asymptotic behavior of $\hat{K}_1$, we define

$$A_k = \sum_{j=k}^{\infty} [N_c = j] = [N_c \ge k], \qquad k \ge 1. \tag{4.18}$$

Then note that A_k is $\mathcal{A}_{k-1} = \sigma\{X_1, \ldots, X_{k-1}\}$ measurable, $k \ge 1$. Using A_k we can re-arrange each term on the right hand side of (4.16):

$$\sum_{j=2}^{\infty}\sum_{i=1}^{j-1} E[N_n = i][N_c = j](\delta_i - \theta)^2 = \sum_{i=1}^{\infty} E[N_n = i]A_{i+1}(\delta_i - \theta)^2, \tag{4.19}$$

$$\begin{aligned} &\sum_{j=2}^{\infty}\sum_{i=1}^{j-1} E[N_n = i][N_c = j](\delta_j - \theta)^2 \\ &\quad = \sum_{i=1}^{\infty} E[N_n = i] \left\{ \sum_{j=i+1}^{\infty} A_j(\delta_j - \theta)^2 - \sum_{j=i+1}^{\infty} A_{j+1}(\delta_j - \theta)^2 \right\}, \end{aligned} \tag{4.20}$$

and

$$c\sum_{j=2}^{\infty}\sum_{i=1}^{j-1} E[N_n = i][N_c = j](i - j) = -c\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} E[N_n = i]A_j. \tag{4.21}$$

Combining (4.19), (4.20) and (4.21) we can rewrite (4.16) as follows:

$$\begin{aligned} \hat{K}_1 &= -\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} E[N_n = i]A_j\{(\delta_j - \theta)^2 - (\delta_{j-1} - \theta)^2 + c\} \\ &= -\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} E[N_n = i]A_j\{E((\delta_j - \theta)^2 \mid \mathcal{A}_j) \\ &\qquad\qquad - E((\delta_{j-1} - \theta)^2 \mid \mathcal{A}_{j-1}) + c\} \\ &= -\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} E[N_n = i]A_j\{U_j/r_j - U_{j-1}/r_{j-1} + c\}, \end{aligned} \tag{4.22}$$

where U_k is given by (2.6), $k \ge 1$, and the last equality in (4.22) follows from (2.6), (2.7), (2.8) and the fact that A_j is $\mathcal{A}_{j-1}$-measurable and that δ_j is the Bayes rule, $j \ge 1$. Now using the martingale property of $\{U_j\}$, (4.22) can be rewritten as

$$\begin{aligned} \hat{K}_1 &= -\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} E[N_n = i]A_j\{r_j^{-1}E(U_j \mid \mathcal{A}_{j-1}) - r_{j-1}^{-1}U_{j-1} + c\} \\ &= -\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} E[N_n = i]A_j\{r_j^{-1}U_{j-1} - r_{j-1}^{-1}U_{j-1} + c\} \\ &= \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} E[N_n = i]A_j\{r_j^{-1}r_{j-1}^{-1}U_{j-1} - c\}, \end{aligned} \tag{4.23}$$

since $r_i = r_0 + i$ for $i \geq 1$. Again use the martingale property of $\{U_j\}$, $r_i = r_0 + i$ and the fact that $A_i \geq A_{i+1}$ for $i \geq 1$ (see (4.18)) to obtain for $i \geq 1$,

$$\sum_{j=i+1}^{\infty} |E[N_n = i]A_j(r_j^{-1}r_{j-1}^{-1}U_{j-1} - c)| \leq EA_{i+1}\{aU_i + c\}, \tag{4.24}$$

where $a = \sum_{j=1}^{\infty}(j(j+1))^{-1}$. Since $A_{i+1} = [N_c \geq i+1]$ by the definition, we obtain by the definition of N_c (see (2.5))

$$\begin{aligned} \sum_{i=1}^{\infty} EA_{i+1}\{aU_i + c\} &\leq a\sum_{i=1}^{\infty} E[N_c \geq i+1]U_i + c\sum_{i=1}^{\infty} E[N_c \geq i+1] \\ &\leq a\sum_{i=1}^{\infty} E[U_i > cr_ir_{i+1}]U_i + c\sum_{i=1}^{\infty} E[N_c \geq i]. \end{aligned} \tag{4.25}$$

The second term on the right hand side of the inequality (4.25) is equal to cEN_c, which is finite (see Theorem 2.1 above). The first term can be bounded by an application of Holder's inequality followed by Markov inequality as follows:

$$\begin{aligned} \sum_{i=1}^{\infty} E[U_i > cr_ir_{i+1}]U_i &\leq \sum_{i=1}^{\infty}(E[U_i > cr_ir_{i+1}])^{1/2}(EU_i^2)^{1/2} \\ &\leq \sum_{i=1}^{\infty} \frac{(EU_i^2)^{1/2}}{c(r_ir_{i+1})}(EU_i^2)^{1/2} \\ &\leq c^{-1}\sum_{i=1}^{\infty} \frac{1}{i^2}EU_i^2 \\ &\leq c^{-1}E(\psi''(\omega))^2\sum_{i=1}^{\infty} i^{-2} < \infty, \end{aligned} \tag{4.26}$$

since $E(\psi''(\omega))^2 < \infty$ and $EU_i^2 \leq E(\psi''(\omega))^2$ by Jensen's inequality; i.e., $EU_i^2 = E\{E(\psi''(\omega) \mid \mathcal{A}_i)\}^2 \leq E(\psi''(\omega))^2$. Now apply Lemma 4.1 with

$$a_i^{(n)} = \sum_{j=i+1}^{\infty} E[N_n = i]A_i\{r_j^{-1}r_{j-1}^{-1}U_{j-1} - c\}, \qquad i \geq 1$$

and

$$b_i = EA_{i+1}\{aU_i + c\}, \qquad i \geq 1.$$

Observe that $|a_i^{(n)}| \leq b_i$ from (4.24), and $\sum_{i=1}^{\infty} b_i < \infty$ from (4.25) and (4.26). Therefore, from the conclusion of Lemma 4.1 and (4.23) we obtain

$$\begin{aligned} \lim_{n\to\infty} \hat{K}_1 &= \sum_{i=1}^{\infty} \lim_{n\to\infty} a_i^{(n)} \\ &= \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} \lim_{n\to\infty} E[N_n = i]A_j\{r_j^{-1}r_{j-1}^{-1}U_{j-1} - c\}, \end{aligned} \tag{4.27}$$

the last equality is obtained by applying Lemma 4.1 again with $b'_j = EA_j\{r_j^{-1}r_{j-1}^{-1}U_{j-1} - c\}$. (It can be easily shown that again $\sum_{j=i+1}^{\infty} b'_j < \infty$ along the same lines used to prove that $\sum_{i=1}^{\infty} b_i < \infty$.) Now it is enough to study the limit behavior of $E[N_n = i]A_j$ for $j \geq i+1$ and $i \geq 1$. By the definitions of N_n and A_j we see that for $j \geq i+1$ and $i \geq 1$,

$$E[N_n = i]A_j \leq E[\hat{\alpha}_i \geq 0][\alpha_i < 0]$$

which goes to zero as $n \to \infty$, since $\hat{\alpha}_i \xrightarrow{P} \alpha_i$ as $n \to \infty$ by Lemma 4.2. This completes the proof. □

LEMMA 4.5. *Under the hypotheses of Theorem* 3.1,

$$\lim_{n\to\infty} \hat{L} = 0. \tag{4.28}$$

PROOF. Write $\hat{L} = \hat{L}_1 + \hat{L}_2$, where

$$\hat{L}_1 = \sum_{j=2}^{\infty}\sum_{i=1}^{j-1} E[N_n = j][N_c = i]\{(\delta_j - \theta)^2 - (\delta_i - \theta)^2 + c(j-i)\} \tag{4.29}$$

and

$$\hat{L}_2 = \sum_{j=2}^{\infty}\sum_{i=1}^{j-1} E[N_n = j][N_c = i]\{(\hat{\delta}_j - \theta)^2 - (\delta_j - \theta)^2\}. \tag{4.30}$$

Again it is easy to show that $|\hat{L}_2| \leq \sum_{j=1}^{\infty} E(\hat{\delta}_j - \delta_j)^2$ and that $\lim_{n\to\infty} \hat{L}_2 = 0$; compare with (4.12) and (4.17). It is now enough to show that $\lim_{n\to\infty} \hat{L}_1 = 0$. First we define

$$\hat{D}_k = \sum_{j=k}^{\infty}[N_n = j] = [N_n \geq k], \qquad k \geq 1. \tag{4.31}$$

Making use of $\hat{D}_k$ and re-arranging terms in (4.29) we can show that (cf., (4.22))

$$\begin{aligned} \hat{L}_1 &= \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} E[N_c = i]\hat{D}_j\{(\delta_j - \theta)^2 - (\delta_{j-1} - \theta)^2 + c\} \\ &= \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} E[N_c = i]\hat{D}_j\{c - r_j^{-1}r_{j-1}^{-1}U_{j-1}\}, \\ &= \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} E[N_c \geq i][\alpha_i = 0]\hat{D}_j\{c - r_j^{-1}r_{j-1}^{-1}U_{j-1}\} \\ &\quad + \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} E[N_c \geq i][\alpha_i > 0]\hat{D}_j\{c - r_j^{-1}r_{j-1}^{-1}U_{j-1}\}. \end{aligned} \tag{4.32}$$

The first term on the right hand side of (4.32) is zero, since $E[\alpha_i = 0] = E[U_i = cr_ir_{i+1}] = 0$ because U_i is a continuous random variable (U_i is a continuous function of continuous random variable S_i by Theorem 2.9 of Lehmann (1986)). It can be shown that the limit of the second term is equal to zero by an application of Lemma 4.1 (with suitable choice of $a_i^{(n)}$ and b_i) and using the facts that, for $j \geq i+1$, $\hat{D}_j[\alpha_i > 0] \leq [\hat{\alpha}_i < 0][\alpha_i > 0] \xrightarrow{P} 0$ as $n \to \infty$. This completes the proof. □

PROOF OF THEOREM 3.1. Follows from (4.5) and Lemmas 4.3, 4.4 and 4.5. □

PROOF OF COROLLARY 3.1. Stopping rule N_c defined by (2.5) is optimal; see Theorem 2.1. Then the risk $R_{N_c}(\pi_0)$ of N_c becomes the Bayes envelope risk of the sequential component. □

Acknowledgements

The author thanks the referees for many useful comments and suggestions that helped improve the presentation. Thanks also go to Shunpu Zhang for computational assistance.

REFERENCES

Ash, Robert B. (1972). *Real Analysis and Probability*, Academic Press, New York.

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

Bickel, P. and Yahav, J. (1967). Asymptotically pointwise optimal procedures in sequential analysis, *Proc. 5th Berkeley Symp. on Math. Statist. Prob.*, VI, 401–413, University of California Press, Berkeley.

Bickel, P. and Yahav, J. (1969). An A.P.O. rule in sequential estimation with quadratic loss, *Ann. Math. Statist.*, **40**, 417–426.

Chow, Y. S., Robbins, H. and Siegmund, D. (1971). *Great Expectations: The Theory of Optimal Stopping*, Houghton-Mifflin, Boston.

Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families, *Ann. Statist.*, **6**, 269–281.

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

Ghosh, M. (1991). Hierarchical and empirical Bayes sequential estimation, *Handbook of Sequential Analysis* (eds. B. K. Ghosh and P. K. Sen), 441–458, Marcel Dekker, New York.

Ghosh, M. and Hoekstra, R. M. (1989). A.P.O. rules in heirarchical and empirical Bayes models, *Sequential Analysis*, 79–100.

Ghosh, M. and Lahiri, P. (1987). Robust empirical Bayes estimation of means from stratified samples, *J. Amer. Statist. Assoc.*, **82**, 1153–1162.

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

Good, I. J. (1965). *The Estimation of Probabilities. An Essay on Modern Bayesian Methods*, M.I.T. Press, Cambridge, Massachusetts.

Hoekstra, M. (1989). Asymptotically pointwise optimal stopping rules in multiparameter estimation, Ph.D. Dissertation, Department of Statistics, University of Florida (unpublished).

Karunamuni, R. J. (1985). Empirical Bayes with sequential components, Ph.D. dissertation, Department of Statistics and Probability, Michigan State University.

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. Theorey Methods*, **18**, 2533–2552.

Karunamuni, R. J. (1990). On the empirical Bayes approach to multiple decision problems with sequential components, *Ann. Inst. Statist. Math.*, **42**, 637–655.

Laippala, P. (1979). The empirical Bayes approach with floating sample size in binomial experimentation, *Scand. J. Statist.*, **6**, 113–118 (Correction: *ibid.* (1980), **7**, p. 105).

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

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed., Wiley, New York.

Lindlay, D. V. and Smith, A. F. M. (1972). Bayes estimators for the linear model (with discussion), *J. Roy. Statist. Soc. Ser. B*, **34**, 1–41.

Maritz, J. S. and Lwin, T. (1989). *Empirical Bayes Methods*, 2nd ed., Chapman and Hall, London.

Martinsek, A. T. (1987). Empirical Bayes methods in sequential estimation, *Sequential Anal.*, **6**, 119–137.

Morris, C. (1983*a*). Parametric empirical Bayes inference: theory and applications, *J. Amer. Statist. Assoc.*, **78**, 47–65.

Morris, C. (1983*b*). Parametric empirical Bayes confidence sets, *Scientific Inference, Data Analysis, and Robutness* (eds. G. E. P. Box, T. Lenonard and C. F. Wu), 25–50, Academic Press, New York.

Rehailia, M. E. H. (1984). Asymptotic sequential analysis on the A.P.O. rule performance, *Sequential Anal.*, **3**, 155–174.

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

Robbins, H. (1959). Sequential estimation of the mean of a normal population, *Probability and Statistics* (the H. Cramér volume), Almqvist and Wiksell, Stockholm.

Robbins, H. (1963). The empirical Bayes approach to testing statistical hypotheses, *Review of the International Statistical Institute*, **31**, 195–208.

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

Sen, P. K. (1981). *Sequential Nonparametrics*, Wiley, New York.

Shapiro, C. P. and Wardrop, R. L. (1980). Bayesian sequential estimation for one-parameter exponential families, *J. Amer. Statist. Assoc.*, **75**, 984–988.

Woodroofe, M. (1981). A.P.O. rules are asymptotically non-deficient for estimation with squared error loss, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gabiete*, **58**, 331–341.

Woodroofe, M. (1982). Nonlinear renewal theory in sequential analysis, *CBMS-NSF Regional Conf. Ser. in Appl. Math.*, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.