Journal of the Korean Statistical Society
Vol. 20, No. 1, 1991

Empirical Bayes Problem With Random Sample Size Components+

Inha Jung*

ABSTRACT

The empirical Bayes version involves "independent" repetitions(a sequence) of the component decision problem. With the varying sample size possible, these are not identical components. However, we impose the usual assumption that the parameters sequence $\theta = (\theta_1, \theta_2, \cdots)$ consists of independent G-distributed parameters where G is unknown. We assume that $G \in \mathcal{G}$, a known family of distributions. The sample size N_i and the decisin rule d_i for component i of the sequence are determined in an evolutionary way. The sample size N_1 and the decision rule $d_1 \in D_{N1}$ used in the first component are fixed and chosen in advance. The sample size N_2 and the decision rule d_2 are functions of $\underline{X}^1 = (X_{11}, \cdots, X_{1N_1})$, the observations in the first component.

In general, N_i is an integer-valued funtion of $(\underline{X}^1, \underline{X}^2, \cdots, \underline{X}^{i-1})$ and, given N_i, d_i is a D_{Ni}-valued funtion of $(\underline{X}^1, \cdots, \underline{X}^{i-1})$. The action chosen in the i-th component is $d_i(\underline{X}^i)$ which hides the display of dependence on $(\underline{X}^1, \cdots, \underline{X}^{i-1})$.

We construct an empirical Bayes decision rule for estimating normal mean and show that it is asymptotically optimal.

1. Introduction

In the usual empirical Bayes decision problem as was introduced by Robbins(1955, 1966), the component problems are i.i.d. : we are given a stochastic process(θ_1, X_1), (θ_2, X_2), $\cdots$, of i.i.d. random vectors with the interpretation that, at the ith component problem, an observation X_i has a distribution P_θ given the prameter $\theta_i = \theta$ and $\theta_1, \theta_2, \cdots$ are i.i.d. with a fixed but unknown prior distribution G in a family of distributions $\mathcal{G}$. O'Bryan(1972, 1976) introduced

+ This research was suported by Korean Science and Engineering Foundation 88-07-23-02

* Ajou University

the empirical Bayes decision prolem with non-i.i.d. components by allowing unequal nonrandom sample sizes in the component problems. Laippala(1985), whose work is motivated by O'Bryan (1976), introduced an empirical Bayes problem with nonidentical components with cost for observations and random "floating" sample sizes for the components. In Laippala(1985), "optimal" sample size is not optimal among the class of all fixed sample sizes and the proof of *Theorem 1* claiming the convergence of the "floating" sample sizes to his "optimal" sample size neglect the boundary set on which the convergence may fail.

Most of the empirical Bayes work involves identical components with the exception of the nonrandom sample size work of O'Bryan and Laippala. Karunamuni(1985, 1988) and Gilliland and Karunamuni(1988) consider the possiblity of varying stochastic sample sizes. Gilliland and Karunamuni(1988) develop the theory of finite state problems. Karunamuni(1985, 1988) studies an empirical Bayes problem with a sequentical component. He dose not treat the optimal fixed sample size problem.

In the paper we consider the empirical Bayes decision problem where the component problem includes a constant cost per observation and the option to choose in advance the total number of observations. The empirical Bayes decision approach with our component permits data accumulated over past component problems to be used in selecting both the sample size and the decision rule to be used in the current component problem. The generality introduced by allowing sample sizes that are determined stochastically makes the result more useful in applications where, typically, the choice of sample size is an option based on past data.

In *section 2*, a statistical decision problem with cost for observations is considered. Optimal fixed sample size is obtained. Estimation of the normal mean is presented as an example. In *section 3*, we develop an empirical Bayes decision problem with random sample size component. Asymptotic optimality of the empirical Bayes procedure is defined for our case. It is shown that asymptotic optimality imlpies the convergence of the sample sizes to the optimal fixed sample size. In *section 4*, under the squared error loss, we construct an asymptotic optimal empirical Bayes estimate of the normal mean with random sample size components. We assume that the variance of the conditional distribution is unknown. The class $\mathcal{G}$ is restricted to the conjugate family, family of normal priors.

2. A statistical decision problem with cost for observations

Consider a statistical decision problem with a parameter space Θ, an action space $\mathcal{A}$, a nonnegative loss funtion $L(\cdot, \cdot)$ on $\Theta \times \mathcal{A}$, unknown prior distribution G on Θ and a cost $c > 0$ per observation. Let $X_1, X_2, \cdots$ be observations which are independently and identically distriduted with a distribution P_θ given θ, taking values in a set $\mathcal{X}$, the observation space. Let D_n be the set of all decision funtions $d : \mathcal{X}^n \to \mathcal{A}$ where $\mathcal{X}^n$ is the observation space for the vector $\underline{X} = (X_1, \cdots, X_n)$. When θ is the parameter and a decision rule $d \in D_n$ is used, the decision loss plus cost for observing $\underline{X} = (X_1, \cdots, X_n)$ is

$$L(\theta, d(\underline{X})) + cn$$

, where we assume that L is integrable for all θ, n and $d \in D_n$.

Let R_n denote the risk and Bayes risk of the decision rule $d \in D_n$, i.e.,

$$R_n(\theta, d) = \int_{\mathcal{X}^n} L(\theta, d(x))\, dP_\theta^n(x) \tag{2.1}$$

$$R_n(G,\ d)=\int R_n(\theta,\ d)\,dG(\theta) \tag{2.2}$$

and let r_n denote the risk and Bayes risk of the decision rule $d \in D_n$ including cost for observations. Then

$$r_n(\theta,\ d)=R_n(\theta,\ d)+cn \tag{2.3}$$

and

$$r_n(G,\ d)=R_n(G,\ d)+cn \tag{2.4}$$

We define the minimum Bayes risk and the minimum Bayes risk plus cost in the usual way. We assume for each prior G and each $n=1,2,\cdots$ that a Bayes rule $d_G^n \in D_n$ exists. Thus,

$$\inf_{d\in D_n} R_n(G,\ d)=R_n(G,\ d_G^n) \tag{2.5}$$

Let

$$R_n(G)=R_n(G,\ d_G^n)$$

and

$$r_n(G)=R_n(G)+cn. \tag{2.6}$$

Since $R_n(G)$ is nonincreasing in n, a minimizer of $r_n(G)$ exists among the integers 1, 2, $\cdots$. We will denote a specified minimizer as $n^*=n^*(G)$ and refer to it as an optimal fixed sample size. Therefore, $r(G)=r_{n^*}(G)$ is the minimum Bayes risk in the component across all the possible sample sizes and the corresponding class of decision rules, i.e.,

$$r(G)=r_{n^*}(G)=min\{min\{r_n(G,\ d)\ :\ d\ \in D_n\}\ :\ n=1,\ 2,\ \cdots\}. \tag{2.7}$$

Moreover, note that $R_{n^*}(G)+cn^* < R_1(G)+c < \infty$ so that $n^* < (R_1(G)+c)/c < \infty$. For some components, $R_1(G)$ is a bounded funtion of $G \in \mathcal{G}$.

Example 2.1 (Estimation). Let $X_1,\ X_2,\ \cdots,X_n$ be i.i.d. $N(\theta,\ A)$ given θ and let θ have prior distrbution $G=N(\mu,\ V)$. Assume A is known. Let $\Theta=\mathcal{A}=(-\infty,\ \infty)$, $L(\theta,\ a)=(\theta-a)^2$ for $(\theta,\ a) \in \Theta\times\mathcal{A}$, and let $c>0$ denote the constant cost per observation. Then a Bayes decision function for estimating θ based on observation $\underline{X}=(X_1,\ \cdots,X_n)$ is

$$d_G(\underline{X})=\left(\frac{A}{A+nV}\right)\mu+\left(1-\frac{A}{A+nV}\right)\bar{X}_n \tag{2.8}$$

and

$$r_n(G)=\left(\frac{AV}{A+nV}\right)+cn. \tag{2.9}$$

The funtion $AV/(A+nV)+cn$ is a convex function of $n \in (-A/V,\ \infty)$ with a minimum at $\eta=(A/c)^{1/2}-A/V$. Therefore, we can define an optimal fixed sample size n^* as the positive integer minimizer of (2.9), which is related to η by

$$n^*=n^*(A,\ V)=\begin{cases}1 & \text{if } \eta<1\\ \eta & \text{if } \eta\in\{1,\ 2,\ \cdots\}\\ [\eta] \text{ or } [\eta]+1, & \text{otherwise,}\end{cases} \tag{2.10}$$

where $[\ ]$ denotes the greatest integer funtion.

3. An empirical Bayes decision problem with random sample size components

When a statistical decision problem occurs repeatedly and independently with the same unknown prior G, one can apply an empirical Bayes approach where G is estimated using data collected from previous repititions and Bayes rule with respect to the estimated G is used in the current component problem. The empirical Bayes decison approach with our component permits data accumulated over past component problems to be used in selecting both the sample size and the decision rule to be used in the current component problem. Choice of a sample size is an option and is based on past data. We impose the usual assumption that the parameter sequence$(\theta_1, \theta_2, \cdots)$ consists of independent G-distributed parameters, where G is an unknown element of the known class of distributions $\mathcal{G}$.

The sample size N_i and the decision rule d_i for the components are determined in an evolutionary way. The sample size N_1 and the decision rule $d_1 \in D_{N1}$ used in the first component are given nonrandom choices.

The sample size N_2 and the decision rule $_2$ are funtions of $\underline{X}^1=(X_{11}, \cdots, X_{1N_1})$, the observations in the first component.

The sample size N_3 and the decision rule d_3 are funtions of $(\underline{X}^1, \underline{X}^2)$. In general N_i is an integer-valued funtion of $(\underline{X}^1, \underline{X}^2, \cdots \underline{X}^{i-1})$ and, given N_i, d_i is a D_{Ni}-valued function of $(\underline{X}^1, \underline{X}^2, \cdots \underline{X}^{i-1})$.

Let $\underline{N}=(N_1, N_2, \cdots)$ and $\underline{d}=(d_1, d_2, \cdots)$. We will be concerned with the risk behavior of empirical Bayes procedures$(\underline{N}, \underline{d})$. (Here and henceforth, the term risk will refer to the expected loss plus cost for observations.) The Bayes risk for the decision about θ_i is

$$Er_{Ni}(G, d_i)=\mathrm{ER}_{Ni}(G, d_i)+cEN_i, \tag{3.1}$$

where E denotes the expectation over the earlier observations $\underline{X}^1, \underline{X}^2, \cdots, \underline{X}^{i-1}$.

Definition 3.1 If the empirical Bayes procedure $(\underline{N}, \underline{d})$ possesses the property:

$$\lim_i \mathrm{Er}_{Ni}(G\ d_i)=r(G) \text{ for all } G \in \mathcal{G}, \tag{3.2}$$

we say it is asymptotically optimal. This means that in the limit, the empirical Bayes procedure has the best possible risk behavior, i.e, achives minimum Bayes risk.

Goal of the empirical Bayes rules is achiving asymptotic optimality. All of our results concern parametric families of priors, $\mathcal{G}=\{G_\omega \mid \omega \in \Omega\}$ where Ω is a specified subset of a finite-dimensional Euclidean space R^p.

We will identify G by ω and replace G accordingly in formulas for risk, etc. Also, we will use the empirical Bayes approach wherein the prior ω is estimated, say by $\hat{\omega}$, and $\hat{n}=n^*(\hat{\omega})$ and $d_{\hat{\omega}} \in D_{\hat{n}}$ are used in defining the empirical Bayes procedure. Note that we have dropped the superscript on d_ω. The following table shows how the empirical Bayes procedure evolves using estimates at $\hat{\omega}_0$ arbitrary, $\hat{\omega}_1=\hat{\omega}_1(\underline{X}^1)$, $\hat{\omega}_2=\hat{\omega}_2(\underline{X}^1, \underline{X}^2)$, $\hat{\omega}_3=\hat{\omega}_3(\underline{X}^1, \underline{X}^2, \underline{X}^3)$, $\cdots$. The θ_1, θ_2, θ_3, $\cdots$ are i.i.d. G_ω.

Table 3.1. Empirical Bayes Procedure with Stochastically determined Sample Size

Stage	Parameter	Sample Size	Decision Rule	Observation	Estimated Prior	Risk
1	θ_1	$N_1=n^*(\hat{\omega}_0)$	$d_1=d^*_{\hat{\omega}_0}$	$\underline{X}^1$	$\omega_1(\underline{X}^1)$	$E\{L(\theta_1,\ d_1(\underline{X}^1))+cN_1\}$ $=r_{N1}(\omega,\ d_1)$
2	θ_2	$N_2=n^*(\omega_1)$	$d_2=d^*_{\hat{\omega}_1}$	$\underline{X}^2$	$\omega_2(\underline{X}^1,\ \underline{X}^2)$	$E\{L(\theta_2,\ d_2(\underline{X}^2))+cN_2\}$ $=Er_{N2}(\omega,\ d_2)$
3	θ_3	$N_3=n^*(\omega_2)$	$d_3=d^*_{\hat{\omega}_2}$	$\underline{X}^3$	$\hat{\omega}_3(\underline{X}^1,\ \underline{X}^2,\ \underline{X}^3)$	$E\{L(\theta_3,\ d_3(\underline{X}_3))+cN_3\}$ $=Er_{N3}(\omega,\ d_3)$
:	:	:	:	:	:	:

The convergence of the sequence of risks in the last column to the smallest possible risk $r(\omega)=r_{n^*}^{(\omega)}$ is the asympototic optimality property. The following remark shows how asymptotic optimality implies the convergence of the sample sizes N_i to the set of optimal fixed sample sizes.

Remark 3.1. Let $s(\omega)$ denote the set of integer minimizers of $r_n(\omega)$.

(a) If $(\underline{N},\ \underline{d})$ is asmptotically optimal at ω, then

$$P(N_i \in s(\omega))\to 1 \quad \text{as} \quad i\to\infty. \tag{3.3}$$

(b) If $r_{Ni}(\omega,\ d_i)\to R(\omega)$ a.s., then

$$P(N_i \in s(\omega),\ \text{eventually})=1 \tag{3.4}$$

This can be shown as follows. For given ω, there exists an $\epsilon>0$ such that for all $n' \notin s(\omega)$, $r_{n'}(\omega,\ d)-r(\omega)\geqslant \in$ for all $d \in D_{n'}$. On the event, $N_i \notin s(\omega)$, $r_{Ni}(\omega,\ d_i)-r(\omega)\geqslant \epsilon$ so that

$$E[r_{Ni}(\omega,\ d_i)-r(\omega)]\geqslant \in P(Ni \notin s(\omega)),$$

which yields (3.3) by letting $i\to\infty$. Since$(N_i \in s(\omega)$ i.o.$)$ implies $r_{Ni}(\omega,\ d_i)-r(\omega)\geqslant \epsilon$, i.o., (3.4) is proved.

The following lemma will be used in section 4 in establishing the asymptotic optimality property.

Lemma 3.1. For prior ω and ν, let $n=n^*(\omega)$, $m=n^*(\nu)$ be optimal fixed sample sizes and let $d^k_\omega,\ d^k_\nu \in D_k$ denote Bayes decisions with respect to ω, ν for $k=1,\ 2,\ \cdots$. Then

$$0\leqslant r_m(\omega,\ d^m_\nu)-r(\omega)\leqslant \tag{3.5}$$

$$sup_k \mid R_k(\omega,\ d^k_\nu)-R_k(\nu,\ d^k_\nu) \mid + sup_k \mid R_k(\omega,\ d^k_\omega)-R_k(\nu,\ d^k_\omega) \mid .$$

Proof. The left inequality follows from the fact that $r(\omega)$ is the minimum Bayes risk over choices $d \in D_k$ and sample sizes k. Adding and subtracting $r_m(\nu,\ d^m_\nu)$ and noting that $r_m(\nu,\ d^m_\nu)\leqslant r_n(\nu,\ d^n_\omega)$ yields

$$r_m(\omega,\ d^m_\nu)-r_n(\omega,\ d^n_\omega)\leqslant r_m(\omega,\ d^m_\nu)-r_m(\nu,\ d^m_\nu)+r_n(\nu,\ d^n_\omega)-r_n(\omega,\ d^n_\omega) \tag{3.6}$$

which together with (3.4) implies the right inequality of (3.5). Q.E.D.

The quadratic loss funtion $L(\theta,\ a)=b(\theta-a)^2$, where $b>0$, is covered by our results by factoring b out and replacing c by c/b.

Our methods cover more general cost funtions as well. If the cost funtion is c(n) and liminf $c(n) > R_1(G)$, then for any given G, $\inf\{r_n(G) \mid n=1, 2, \cdots\}$ is attained, and we can define $n^*(G)$ as the smallest minimizer. Moreover, the proof of Lemma 3.1 applies to give the same conclusion, that is, a bound for excess risk in terms of the supremum of differences in decision risks over varying sample size problems.

4. Estimation of the normal mean.

The component problem considered in this section is the one introduced in Example 2.1. Here $G=N(\mu, V)$ and, letting

$$\rho = \frac{A}{A+nV}, \tag{4.1}$$

the posterior distribution of θ given $\underline{X}=(X_1, X_2, \cdots, X_n)$ is

$$N(\rho\mu+(1-\rho)\bar{X}, \frac{A}{A+nV}) \tag{4.2}$$

With this notation, the Bayes estimator (2.8) can be written

$$d_G(\underline{X}) = \rho\mu+(1-\rho)\bar{X}. \tag{4.3}$$

The following remark parallels Remark 3.1.

Remark 4.1. For $G=N(\mu, V)$ and $G'=N(\mu', V')$,

$$R_n(G, d_{G'}) = (1-\rho')^2 \frac{A}{n} + \rho'^2 [(\mu-\mu')^2+V], \tag{4.4}$$

$$| R_n(G, d_{G'}) - R_n(G', d_{G'}) | \leqq (\mu'-\mu)^2 + | V'-V | . \tag{4.5}$$

and

$$R_n(G) = \frac{AV}{A+nV} \tag{4.6}$$

This can be shown as follows. By(4.3), $d_{G'}(\underline{X}) = \rho'\mu'+(1-\rho')\bar{X}$. Since expected squared deviation is variance plus bias squared,

$$\begin{aligned} R_n(G, d_{G'}) &= E_G E_\theta[\rho' \mu'+(1-\rho')\bar{X}-\theta]^2 \\ &= E_G\{(1-\rho')^2 \frac{A}{n} + \rho'^2 (\mu'-\theta)^2\} \\ &= (1-\rho')^2 \frac{A}{n} + \rho'^2[V+(\mu'-\mu)^2]. \end{aligned}$$

Then (4.6) follow by replacing G' by G above and using(4.1). Since

$$R_n(G', d_{G'}) = (1-\rho')^2 \frac{A}{n} + \rho'^2 V'$$

if follows that

$$R_n(G, d_{G'}) - R_n(G', d_{G'}) = \rho'^2[(\mu'-\mu)^2+(V-V')]$$

which yields(4.5).

We seek the optimal sample size n^* which mimimizes

$$r_n(G)=R_n(G)+cn=\frac{AV}{A+nV}+cn$$

among the integrs $n=1, 2, \cdots$. We consider $r_n(G)$ as a funtion of real n and the equation

$$0=\frac{d}{dn}r_n(G)=-\frac{AV^2}{(A+nV)^2}+c.$$

Its larger solution is

$$\eta=(A/c)^{1/2}-A/V. \tag{4.7}$$

We see that $r_n(G)$ is convex in $n \in (-A/V, \infty)$ and that the optimal sample size $n^*=n^*(A, V)$ is given by (4.8).

In our empirical Bayes application the variance A of the conditional distribution $N(\theta, V)$ is assumed to be unknown but is assumed to be in a given bounded interval $(0, a]$. Thus we are taking A to be a nuisance parameter. It is convenient, though not necessary, to require that at least two observations be taken in each component of the empirical Bayes problem so that the estimation of A is simple. Therefore, we will optimize sample size over choices $n=2, 3\cdots$ in defining the envelope risk. It follows that

$$n^*=n^*(A, V)=\begin{cases}1 & \text{if } \eta<2\\ \eta & \text{if } \eta \in \{2, 3, \cdots\}\\ [\eta] \text{ or } [\eta]+1, & \text{otherwise,}\end{cases} \tag{4.8}$$

where η is given in (4.7).

Since $R_2(G)=E_G E_\theta(\bar{X}-\theta)^2<A/2$, it follows as in the comment preceding Example 2.1 that $n^* \leqslant (A/2+2c)/c$. Letting M be the integer $[a/2c+1]+1$, it follows that

$$2 \leqslant n^*(A, V) \leqslant M<\infty \tag{4.9}$$

for all A and priors $G=N(\mu, V)$.

Notice that in the component problem

$$E\bar{X}=\mu, \tag{4.10}$$

$$E\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2=V+A, \tag{4.11}$$

and, provided $n \geqslant 2$,

$$E\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\bar{X})^2=A. \tag{4.12}$$

We construct a decision procedure for the empirical Bayes problem with the component described above. The unknown prior G is assumed to be from the family of normal distributions $\mathcal{G}$, the family of conjugate priors. Let $G=N(\mu, V)$, where $\mu \in (=\infty, \infty)$ and $V \in (0, \infty)$.

Let $\hat{A}_0$, $\hat{\mu}_0$ and $\hat{V}_0$ be initial nonrandom estimates of the component nuisance parameter A and parameters μ, V of the prior. Let $N_1=n^*(\hat{A}_0, \hat{V}_0)$. Then $\underline{X}^1=(X_{11}, \cdots, X_{1N_1})$ is observed in the first component. The empirical Bayes procedure that we will study is defined through sequences of estimators $\hat{A}_i$, $\hat{\mu}_i$ and $\hat{V}_i$ that are $(\underline{X}^1, \cdots, \underline{X}^i)$ measurable with

$$N_{i+1}=n^*(\hat{A}_i, \hat{V}_i), \quad i=0, 1\cdots \tag{4.13}$$

and

$$d_{i+1}(\underline{X}^{i+1})=\hat{\rho}_{i+1}\hat{\mu}_i+(1-\hat{\rho}_{i+1})Y_{i+1}, \quad i=0, 1\cdots \tag{4.14}$$

where

$$\hat{\rho}_{i+1}=\frac{\hat{A}_i}{\hat{A}_i+N_{i+1}\hat{V}_i} \quad i=0, 1\cdots \tag{4.15}$$

and

$$Y_i=\frac{1}{N_i}\sum_{j=1}^{N_i} X_{ij}, \quad i=1, 2, \cdots \tag{4.16}$$

We now define the estimators $\hat{\mu}_i$, $\hat{A}_i$ and $\hat{V}_i=1, 2, \cdots$ Motivated by (4.10) we define

$$\hat{\mu}_i=\overline{Y}_i=\frac{1}{i}\sum_{j=1}^{i} Y_j, \quad i=1, 2, \cdots \tag{4.17}$$

which is the average of the sample means for the first i components. Motivated by (4.12) we define

$$\hat{A}_i=\overline{S}_i \wedge a, \quad i=1, 2, \cdots \tag{4.18}$$

where

$$\overline{S}_i=\frac{1}{i}\sum_{j=1}^{i} S_j \tag{4.19}$$

is the average of the sample variances

$$S_j=\frac{1}{N_i-1}\sum_{k=1}^{N_j}(X_{ik}-Y_j)^2 \tag{4.20}$$

for the first i components. Finally, motivated by (4.11) we define

$$\hat{V}_i=[\hat{T}_i-\hat{A}_i]^+, \quad i=1, 2, \cdots \tag{4.21}$$

where

$$\hat{T}_i=\frac{1}{i}\sum_{j=1}^{i} T_{ji} \tag{4.22}$$

is the average of the average squared deviations from $\hat{\mu}_i=\overline{Y}_i$,

$$T_{ji}=\frac{1}{N_j}\sum_{k=1}^{N_j}(X_{jk}-\overline{Y}_i)^2. \tag{4.23}$$

In (4.23) the centerings change with i, which creates a more complicated stochastic structure than exists in (4.20). For the purpose of triangulation, we introduce

$$\bar{T}_i=\frac{1}{i}\sum_{j=1}^{i} T_j, \tag{4.24}$$

where

$$T_j=\frac{1}{N_j}\sum_{k=1}^{N_j}(X_{jk}-\mu)^2. \tag{4.25}$$

Let $\mathcal{Y}_0$ be the trivial σ-field and let $\mathcal{Y}_j=\sigma(\underline{X}_1, \underline{X}_2, \cdots, \underline{X}_j)$, $j=1, 2, \cdots$. The sample size N_j is $\mathcal{Y}_j$-measurable and we see that

$$\begin{aligned} &E(Y_j \mid \mathcal{Y}_{j-1})=\mu, \qquad j=1, 2, \cdots \\ &E(S_j \mid \mathcal{Y}_{j-1})=A, \qquad j=1, 2, \cdots \\ &E(T_j \mid \mathcal{Y}_{j-1})=V+A, \qquad j=1, 2, \cdots \end{aligned} \tag{4.26}$$

Lemma 4.1. The sequences $\hat{\mu}_i=\bar{Y}_i$, S_i and T_i are a.s. consistent for μ, A and $V+A$, respectively.

Proof. We will use (4.26) and the theorem on stability about conditional expectation used i.e. Hall and Heyde(1980, Theorem 2.19). The sequences Y_i, S_i and T_i are not bounded. However, we will find random variables Y, S and T that are square integrable and stochastically larger than their absolute values. This implies the hypothesis of Theorem 2.19 that is sufficient for the a.s. convergence.

Recall that $2 \leqslant N_i \leqslant M$, $i=1, 2, \cdots$. Consider the component problem with sample size M and observations $X_1, X_2, \cdots, X_M$. Let $Y=\Sigma \mid X_j \mid$, $S=\Sigma X_j^2$ and $T=\Sigma(X_j-\mu)^2$. From the definitions (4.16), (4.20) and (4.25) we see that Y, S and T are stochastically larger than $\mid Y_i \mid$, $\mid S_i \mid$ and $\mid T_i \mid$ $i=1, 2, \cdots$. Also $Y^2 \leqslant MS$ and, conditional on θ, the distributions of S and T are noncentral chi-square distibutions with second moments that are integrable $N(\mu, V)$. Thus, Y, S and T are square integrable. Q.E.D

Lemma 4.2. The estimator $\hat{T}_i$ and $\hat{V}_i$ are a.s. consistent for $V+A$ and V.

Proof. We have from (4.23) and (4.25) that

$$T_j-T_{ji}=(\bar{Y}_i-\mu)\,(2Y_j-\mu-\bar{Y}_i). \tag{4.27}$$

Since $Y_i=\Sigma Y_j/i$, we have from (4.22) and (4.24), that

$$\hat{T}_i-\bar{T}_i=(\bar{Y}_i-\mu)^2. \tag{4.28}$$

It follows from Lemma 4.1 that $\hat{T}_i$ is a.s. consistent for $V+A$. Using (4.21) and Lemma 4.1 it follows that $\hat{V}_i$ is consistent for V. Q.E.D.

Theorem 4.1. Let $A \leqslant a$. Then the empirical Bayes procedure $(\underline{N}, \underline{d})$ defined by (4.13)$-$(4.23) is asymptotically optimal at each $G=N(\mu, V)$.

Proof. From Lemma 3.1 and (4.5),

$$0 \leqslant r_{N+1}(G,\ d_{i+1}) - r(G) \leqslant 2(\hat{\mu}_i - \mu)^2 + 2 \mid \hat{V}_i - V \mid . \tag{4.29}$$

Let Y, T be the random variables defined in the proof of Lemma 4.1. Then for $p > 0$, $E \mid Y_i \mid^{2+p} \leqslant E(Y+1)^{2p+1} < \infty$ and $E \mid T_j \mid^{1+p} \leqslant E(T+1)^{1+p} < \infty$ for $j = 1, 2, \cdots$. Hence, the $\{\hat{\mu}_j^2\}$ and the $\{\overline{T}_j\}$ are uniformly integrable. Thus, $\{\hat{\mu}_j^2\}$ and $\{\overline{T}_j\}$ are uniformly integrable and the a.s. convergence (lemma 4.1) implies that

$$E(\hat{\mu}_i - \mu)^2 \to 0 \tag{4.30}$$

and

$$E \mid \overline{T}_i - (V+A) \mid \to 0. \tag{4.31}$$

It follows form the triangle ineguality and (4.28) that

$$\begin{aligned} \mid \hat{V}_i - V \mid &\leqslant \mid \hat{T}_i - \overline{T}_i \mid + \mid \overline{T}_i - (V+A) \mid + \mid (V+A) - (V+\hat{A}_i) \mid \\ &= (\hat{\mu}_i - \mu)^2 + \mid \overline{T}_i - (V+A) \mid + \mid \hat{A}_i - A \mid . \end{aligned} \tag{4.32}$$

The dominated convergence theorem and Lemma 4.1 imply

$$E \mid \hat{A}_i - A \mid \to 0 \tag{4.33}$$

which together with (4.29) − (4.32) establish the result. Q.E.D

References

1. Gilliland, D. and Karunamuni, R(1988). On empirical Bayes with sequential component. Annals of the Institute of Statistical Mathematics, **40**, 187-193.
2. Hall, P. and Heyde, C.C.(1980). Martingale limit theory and its application, Academic Press, New York.
3. Karunamuni, R(1985). Empirical Bayes with sequential components. Ph.D. Thesis, Dept. of Statistics & Probability, Michigan State University.
4. Karunamuni, R(1988). On empirical Bayes testing with sequential components, Annals of Statistics, **16**, 1270-1282.
5. Laippala, P.(1979). The empirical Bayes approach with floating optimal sample size in binomial experimentation. Scandinavian Journal of Statistics, **6**, 113-118 : coorection note 7 105.
6. Laippala. P.(1985). The empirical Bayes rules with floating optimal samle size for exponential conditional distributions. Annals of the Institute of Statistical Mathematics, **37**, 315-327.
7. O'Bayan, T.(1972). Empirical Bayes results in the case of non-identical components. Ph.D. Thesis, RM-306, Dept. of Statistics and Probability, Michigan State University.
8. O'Bayan, T.(1976). Some empirical Bayes results in the case of component problem with varying sample sizes for discrete exponential families. Annals of Statistics, **4**, 1290-1293.
9. Robbins, H.(1956). The empirical Bayes approach to statistics. Proc. 3rd Berkeley Symp. Math. Statist. Prob. **1**, 157-163n University of California Press.
10. Robbins, H.(1966). The empirical Bayes approach to statistical decision problems. Annals of Mathematical Statistics, **35**, 1-20.