The Annals of Statistics
1984, Vol. 12, No. 3, 1071–1078

EMPIRICAL BAYES WITH A CHANGING PRIOR

BY M. K. MARA AND J. J. DEELY

University of Georgia and University of Canterbury, New Zealand

We consider modified empirical Bayes problems in which the prior distribution of Θ at stage $n+1$ is $G^{(n+1)}(\theta)$. The Bayes optimality criterion is now given by the sequence of functionals $R(G^{(n+1)})$. The observations $X_1, \cdots, X_n$ are no longer i.i.d. so decision procedures are constructed based on modified empirical density estimates for $f_{G^{(n+1)}}(x)$. Asymptotic optimality together with asymptotic convergence rates is established for two action and estimation problems when the observations are drawn from a member of the one-parameter exponential family.

1. Introduction. The empirical Bayes approach to statistics has been formulated by Robbins (1964). One of the principal assumptions of this method is the *existence* of a prior distribution that remains *fixed* for each component problem. It is clear that any physical process generating data under the assumption of "*existence*" could change from time to time. Existence of a prior distribution implies that θ, the parameter of interest, is being generated by a distribution G, which *physically* exists and is not merely a subjective considered opinion. Hence during this generation process it is to be expected that G could change. Furthermore, in this context it is clear that the empirical Bayes model is precisely the same as that used to describe data generated in a mixture problem. For a discussion of this relationship, see Deely and Lindley (1981).

Mixture problems amenable to a model with a changing mixing distribution (i.e. the prior) occur in such fields as quality control, reliability, biometrics, education, and many others. (See for example Hoadley, 1979, Everitt and Hand, 1981, and Gupta and Huang, 1980). A specific example from quality control is given in Section 4. Whereas there are many ways this change could be described, only a deterministic linear drift in the mean of the prior is dealt with here. Also it should be mentioned that the problem considered in a series of papers, O'Bryan and Susarla (1975, 1976a, 1976b, 1977), allows changes in the form of the observations for each component problem but not the prior. Allowing the prior to change seems to be a neglected yet important area of research.

Consider the sequence $\{(\theta_i, X_i)\}$, $i = 1, 2, \cdots$, of independent random vectors where, conditional on $\theta_i = \theta$, X_i is drawn according to the density

$$f(x \mid \theta) = e^{-\theta x}\beta(\theta)h(x), \quad x \in \mathscr{X} \subset R. \tag{1.1}$$

The functions $\beta(\theta)$ and $h(x)$ are known. The random variables θ_i are unobservable with prior distribution $G^{(i)}(\theta)$, $i = 1, 2, \cdots$. It is assumed that all $G^{(i)}$'s have the

Received April 1980; revised February 1984.
AMS 1980 *subject classifications*. Primary 62C10, 62C12.
Key words and phrases. Empirical Bayes, modified optimality, criterion, rates of convergence.

same support Ω. This gives the marginal density of X_i as

$$(1.2)\qquad f_{G^{(i)}}(x) = h(x) \int_\Omega e^{-\theta x}\beta(\theta)\, dG^{(i)}(\theta).$$

It is assumed that $h(x)$ is r-times differentiable and $| h^{(r)}(x) |$ is bounded on $\mathscr{X}$.

From (1.2) it is apparent how the empirical Bayes formulation is related to mixture problems—f_G is the mixture of ingredients, $f(x \mid \theta)$, using a *mixing distribution*, G.

At the $(n + 1)$st stage ($X_1, \cdots, X_n$ have previously been observed on $f_{G^{(1)}}, \cdots, f_{G^{(n)}}$ respectively) the observation $X_{n+1} = x$ is made. The loss function for a decision rule $t(x)$ is given by $L(t(x), \theta)$. Hence for a decision rule $t_n(x)$ which depends on $X_1, \cdots, X_n$, the overall expected loss for a decision concerning θ_{n+1} is

$$(1.3)\qquad R_n(t_n, G^{(n+1)}) = E_{n+1}[L[t_n(X_{n+1}), \theta]]$$

where expectation E_j is w.r.t. $X_1, \cdots, X_j$; $j = 1, 2, \cdots$. When $G^{(n+1)}(\theta)$ is known, it is generally straightforward to obtain the Bayes procedure $t(x; G^{(n+1)})$ which has minimum expected loss given by the Bayes envelope functional $R(G^{(n+1)})$. When $G^{(n+1)}$ is unknown but nevertheless exists, the problem becomes one of trying to use the past history (i.e. observations $x_1, \cdots, x_n$) to compensate for this deficiency. That this can be done when the G's are identical is the essence of the many papers on empirical Bayes. As one would conjecture, even if the G's differ by a linear drift in their means, the past history can still be used effectively to obtain asymptotically optimal (a.o.) procedures. However, in this case the definition of asymptotic optimality has to reflect this notion of a changing prior and therefore the following definition is given:

DEFINITION. The sequence of decision rules $T = \{t_n\}$ is a.o. relative to $\{G_n\}$ if

$$(1.4)\qquad \lim_{n\to\infty}\{R_n(t_n, G^{(n+1)}) - R(G^{(n+1)})\} = 0.$$

Suppose that $G^{(i)}$ has density $g^{(i)}$ with $g^{(i)}(\theta) = g^*(\theta - \lambda_i)$ for some g^*, where $\lambda_i = E[\theta_i]$, $i = 1, 2, \cdots$. In addition, suppose θ is a location parameter for the density $f(x \mid \theta)$. (For characterization of such, see Ferguson, 1962). Without loss of generality it may be assumed that $E[X \mid \theta] = \theta$. Now define new random variables $Y_i = X_i - \lambda_i$ and note that the Y_i's are i.i.d. with mean zero and density $f_{G^*}(Y)$. In particular, we have $f_{G^{(i)}}(X_i) = f_{G^*}(Y_i)$, $i = 1, 2, \cdots$. Thus if $\{\lambda_i\}$ were a known sequence, the Y_i's are observable and the usual empirical Bayes situation is evident.

Suppose now that the λ_i's are unknown, but that they follow a linear drift $\lambda_i = az_i + b$ where a, b are unknown but $\{z_i\}$ is a known sequence. Thus writing $X_i = az_i + b + Y_i$ gives the usual form for simple linear regression with residual Y_i. Provided $\lim_{n\to\infty} \sum_{i=1}^n (z_i - \bar{z})^2 = \infty$, a, b may be consistently estimated by the usual least squares estimators a_n, b_n. In fact, defining the double sequence $\{\lambda_{i,n}\}$ of estimators of $\{\lambda_i\}$ by $\lambda_{i,n} = a_n Z_i + b_n$, $i = 1, 2, \cdots, n + 1$; $n = 1, 2, \cdots$ yields $\text{plim} \mid \lambda_{i,n} - \lambda_i \mid = 0$. Thus to find a.o. procedures in the sense of (1.4) it remains to show that the consistent estimation of λ_i's is sufficient for the usual empirical Bayes requirement that f_{G^*} be consistently estimated. The construction of this consistent estimator is given in Section 2, and in Section 3 this estimator is used to obtain a.o. procedures for the cases of hypothesis testing and estimation. Section 4 concludes with several remarks and a practical example.

2. Estimation of $f_{G^*}(Y)$. Following Parzen (1961) and Schuster (1969) define, for $j = 0, 1$,

$$f_n^{(j)}(y) = \frac{1}{nk_n^{j+1}} \sum_{i=1}^n K^{(j)}\left(\frac{y - Y_i}{k_n}\right) = \int \frac{1}{k_n} K^{(j)}\left(\frac{t - y}{k_n}\right) dF_n(t) \tag{2.1}$$

where $F_n(t)$ is the empirical c.d.f. of the Y_i's and the kernel K is a continuously differentiable function on a closed interval I in $(-\infty, \infty)$ such that $\int K(u)\, du = 1$ and $\int |u| K(u)\, du < \infty$. When $Y_1, \cdots, Y_n, Y_{n+1}$ are known, the results of Schuster (1969) guarantee consistency of the estimates (2.1). Replacing Y_i's by $Y_{i,n}$'s in (2.1) where $Y_{i,n} = X_i - \lambda_{i,n}$, define the estimators

$$f_{n,n}^{(j)}(\hat{y}) = \frac{1}{k_n} \int_I K^{(j)}\left(\frac{t - \hat{y}}{k_n}\right) dF_{n,n}(t) \tag{2.2}$$

where $F_{n,n}(t)$ is the empirical c.d.f. of $Y_{1,n}, \cdots, Y_{n,n}$ and $\hat{y} = y_{n+1,n} = x - \lambda_{n+1,n}$. Note that the "present" observation with respect to the Y_i sequence is $y = x - \lambda_{n+1}$ and hence must also be estimated, in this case with $\hat{y}$. For notational ease when $j = 0$ in (2.1) or (2.2), the superscript will be omitted. We can now prove:

LEMMA 2.1.

$$E_n | f_{n,n}(\hat{y}) = f_n(y) |^\delta = O((nk_n^4)^{-\delta/2}) \tag{2.3}$$

$$E_n | f_{n,n}^{(1)}(\hat{y}) - f_n^{(1)}(y) |^\delta = O((nk_n^6)^{-\delta/2}) \tag{2.4}$$

for $0 < \delta < 2$.

PROOF. Since K is continuous on I,

$$| f_{n,n}(\hat{y}) - f_n(y) | \leq M(nk_n^2)^{-1} | \textstyle\sum_{i=1}^n (Y_{i,n} - Y_i + Y_{n+1} - Y_{n+1,n}) |$$

where $M = \sup_{u \in I} | K^{(1)}(u) |$. Therefore for $0 < \delta \leq 2$

$$E_n | f_{n,n} - f_n |^\delta \leq \begin{cases} (M(nk_n^2)^{-1})^\delta \sum_{i=1}^n E_n | Y_{i,n} - Y_i + Y_{n+1} - y_{n+1,n} |^\delta \\ \qquad \text{for } \delta \leq 1, \text{ by the } C_\delta \text{ inequality} \\ (M(nk_n^2)^{-1})^\delta [\sum_{i=1}^n (E_n | Y_{i,n} - Y_i + Y_{n+1} - Y_{n+1,n} |^\delta)^{1/\delta}]^\delta \\ \qquad \text{for } \delta > 1, \text{ by the Minkowski inequality.} \end{cases} \tag{2.5}$$

Since $E_n | Y_{i,n} - Y_i |^2 = O(n^{-1})$, $\forall_i$, the r.h.s. of (2.5) is $O((nk_n^4)^{-\delta/2})$ as required. Equation (2.4) follows in a similar fashion.

The conditions of Lemma 2.1, together with the results of Schuster (1969) are sufficient to guarantee consistency of the estimates (2.2). Following Johns and Van Ryzin (1972) and upon choosing $k_n = O(n^{-(1/(2r+1))})$ and a kernel K with support on $I = [0, u_1]$ satisfying

$$\int_0^{u_1} u^{i+j} K^{(j)}(u)\, du = 0 \tag{2.6}$$

for $i = 1, 2, \cdots, r-1$ and $j = 0, 1$, we can establish a rate for this convergence with the following:

LEMMA 2.2.

$$E_n \mid f_{n,n}(\hat{y}) - f_{G^*}(y) \mid^{\delta} = O(n^{-((2r-3)\delta)/2(2r+1)})\{1 + (f_\varepsilon^*(y))^{\delta/2} + (q_\varepsilon^{(r)}(y))^{\delta}\} \tag{2.7}$$

$$E_n \mid f_{n,n}^{(1)}(\hat{y}) - f_{G^*}^{(1)}(y) \mid^{\delta} = O(n^{-((2r-5)\delta)/2(2r+1)})\{1 + (f_\varepsilon^*(y))^{\delta/2} + (q_\varepsilon^{(r)}(y))^{\delta}\} \tag{2.8}$$

where $0 < \varepsilon \le u_1$, $f_\varepsilon^*(y) = \sup_{0<\delta<\varepsilon}\{f_{G^*}(y + \delta)\}$ *and* $q_\varepsilon^{(r)}(y) = \sup_{0\le\delta\le\varepsilon}\{f_{G^*}^{(r)}(y)\}$.

PROOF. To verify (2.7), first note that from the C_δ inequality we have

$$\begin{aligned} &E_n \mid f_{n,n}(\hat{y}) - f_{G^*}(y) \mid^{\delta} \\ &\qquad \le C_\delta E_n \mid f_{n,n}(\hat{y}) - f_n(y) \mid^{\delta} + C_\delta E_n \mid f_n(y) - f_{G^*}(y) \mid^{\delta}. \end{aligned} \tag{2.9}$$

The first term on the r.h.s. of (2.9) is $O(n^{-(2r-3)\delta/2(2r+1)})$ by Lemma 2.1. The second term is bounded by $O(n^{-(r-1)\delta/(2r+1)})\{(f_\varepsilon^*(y))^{\delta/2} + (q_\varepsilon^{(r)}(y))^{\delta}\}$ from Theorem 3 in Johns and Van Ryzin (1972). Since $O(n^{-(2r-3)\delta/2(2r+1)})$ is the dominating rate, this yields the result. Also, (2.8) can be verified in a similar fashion.

3. Asymptotically optimal procedures.

(a) *Hypothesis testing.* Consider the test of H_0: $\theta \le \theta_0$ vs H_1: $\theta > \theta_0$ under the piecewise linear loss structure

$$L(a_i, \theta) = \begin{cases} \max(0, \theta - \theta_0), & \text{if } i = 0 \\ \max(0, \theta_0 - \theta), & \text{if } i = 1 \end{cases} \tag{3.1}$$

where a_i is the action in favor of H_i, $i = 0, 1$. At stage $n + 1$ we observe $X_{n+1} = x$ and since $f_{G_{n+1}}(x) = f_{G^*}(y)$, where $y = x - \lambda_{n+1}$, the Bayes procedure may be written as $t(y; G^*) = P$ (accept $H_0 \mid y$) with

$$t(y; G^*) = \begin{cases} 1 & \text{if } \alpha^*(y) \le 0 \\ 0 & \text{elsewhere} \end{cases} \tag{3.2}$$

where $\alpha^*(y) = (v(y) - \theta_0) f_{G^*}(y) - f_{G^*}^{(1)}(y)$ with

$$v(y) = \frac{h^{(1)}(y + \lambda_{n+1})}{h(y + \lambda_{n+1})} = \frac{h^{(1)}(x)}{h(x)}.$$

As in the usual empirical Bayes situation, the form of (3.2) suggests a natural candidate for an a.o. procedure, except in this case the "present" observation $y = x - \lambda_{n+1}$ must also be estimated. Using $\hat{y}$ as in (2.2) for the estimate of y, define $\alpha_n(\hat{y}) = (v(y) - \theta_0) f_{n,n}(\hat{y}) - f_{n,n}^{(1)}(\hat{y})$ and

$$t_n(\hat{y}) = \begin{cases} 1 & \text{if } \alpha_n(\hat{y}) \le 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{3.3}$$

Note that $\hat{y}$ does not appear in $v(y)$ since that function depends only on knowledge of x, the "present" observation with respect to the x_i sequence. (See also (3.7)). From Lemma 2.2 we have $|\alpha_n(\hat{y}) - \alpha_{G^*}(y)| \to_p 0$ which is sufficient for asymptotic optimality, in the sense of (1.4), of the procedure $T = \{t_n\}$ for all sequences for which $\lambda_n < \infty$, $\forall_n$. Using Lemma 2.2 directly, we can establish the following.

THEOREM 3.1. *Consider the hypothesis testing problem with loss* (3.1). *Define* $T = \{t_n\}$ *by* (3.3) *with* $f_{n,n}^{(j)}$ *given by* (2.2), $j = 0, 1$, *and* K *satisfying* (2.6). *If for some* $0 < \delta < 2$ *and some* $\varepsilon > 0$

$$\int |\alpha^*(y)|^{1-\delta}\{1 + [v(y)]^{\delta}\}\{1 + [f_\varepsilon^*(y)]^{\delta/2}\}\, dy < \infty \tag{3.4}$$

$$\int |\alpha^*(y)|^{1-\delta}\{1 + [v(y)]^{\delta}\}\{1 + [q_\varepsilon^{(r)}(y)]^{\delta}\}\, dy < \infty \tag{3.5}$$

then $R_n(T, G^{(n+1)}) - R(G^{(n+1)}) = O(n^{-(2r-5)\delta/2(2r+1)})$.

PROOF. From Lemma 1 of Johns and Van Ryzin (1972),

$$\begin{aligned} R_n(T, G^{(n+1)}) &- R(G^{(n+1)}) \\ &\le \int |\alpha^*(y) P(|\alpha_n(\hat{y}) - \alpha^*(y)| > |\alpha^*(y)|)\, dy \\ &\le \int |\alpha^*(y)|^{1-\delta} E_n |\alpha_n(\hat{y}) - \alpha^*(y)|^{\delta}\, dy \\ &\le C_\delta \int |\alpha^*(y)|^{1-\delta}[v(y) - \theta_0]^{\delta} E_n |f_{n,n}(\hat{y}) - f_{G^*}(y)|^{\delta}\, dy \\ &\quad + C_\delta \int |\alpha^*(y)|^{1-\delta} E_n |f_{n,n}^{(1)}(\hat{y}) - f_{G^*}^{(1)}(y)|^{\delta}\, dy. \end{aligned}$$

Using Lemma 2.2, we have the result, since the second term dominates asymptotically.

(b) *Estimation.* The Bayes procedure for the squared error loss estimation problem is $t(x; G^{(n+1)}) = t(y; G^*)$ with

$$t(y; G^*) = v(y) - (f_{G^*}^{(1)}(y)/f_{G^*}(y)). \tag{3.6}$$

In our changing prior case, a suitable empirical Bayes procedure is then

$$t_n(\hat{y}) = v(y) - (f_{n,n}^{(1)}(\hat{y})/\tilde{f}_{n,n}(\hat{y})) \tag{3.7}$$

where $\tilde{f}_{n,n}(y) = \max\{f_{n,n}(y), \eta_n\}$ and $\{\eta_n\}$ is a known sequence, $\eta_n \downarrow 0$. By modifying the results of Deely and Zimmer (1976) to the changing prior case, and appealing to Lemma 2.2, the asymptotic optimality of (3.7) is assured. In addition, however, we can establish an asymptotic rate of this convergence as follows.

THEOREM 3.2. *If for some* $\varepsilon > 0$

$$\int \left\{1 + \left[\frac{f_{G^*}^{(1)}(y)}{f_{G^*}(y)}\right]^2\right\}\{1 + f_\varepsilon^*(y) + [q_\varepsilon^{(r)}(y)]^2\} f_{G^*}(y)\, dy < \infty$$

and if there exists a $\delta > 0$ *for which*

$$\int_{\{y: f_{G^*}(y) < \eta_n\}} \left\{\frac{f_{G^*}^{(1)}(y)}{f_{G^*}(y)}\right\}^2 f_{G^*}(y)\, dy < C\eta_n^\delta$$

for some $C > 0$, *then with* $\{k_n\}$, K *as previously defined, the sequence* $T = \{t_n\}$ *given by* (3.7) *is* a.o. *in the sense of* (1.4) *with*

$$R_n(T, G^{(n+1)}) - R(G^{(n+1)}) = O(n^{-\gamma\delta})$$

where $\gamma = (2r - 5)/(2r + 1)(2 + \delta)$.

PROOF. From Lin (1975)

$$E_n \mid t_n(\hat{y}) - t(y; G^*) \mid^2$$

$$\le \eta_n^{-2} E_n \left| f_{n,n}^{(1)}(\hat{y}) - \tilde{f}_{n,n}(\hat{y}) \frac{f_{G^*}^{(1)}(y)}{f_{G^*}(y)} \right|^2$$

$$\le 2\eta_n^{-2}\left\{E_n \mid f_{n,n}^{(1)}(\hat{y}) - f_{G^*}^{(1)}(y) \mid^2 + \left|\frac{f_{G^*}^{(1)}(y)}{f_{G^*}(y)}\right|^2 E_n \mid \tilde{f}_{n,n}(\hat{y}) - f_{G^*}(y) \mid^2\right\}$$

$$\le 2\eta_n^{-2} E_n \mid f_{n,n}^{(1)}(\hat{y}) - f_{G^*}^{(1)}(y) \mid^2$$

$$+ 4\eta_n^{-2} \left|\frac{f_{G^*}^{(1)}(y)}{f_{G^*}(y)}\right|^2 E_n \mid f_{n,n}(\hat{y}) - f_{G^*}(y) \mid^2 + 4\eta_n^{-2} \left|\frac{f_{G^*}^{(1)}(y)}{f_{G^*}(y)}\right| \chi_{\{y: f_{G^*}(y) < \eta_n\}}.$$

Thus, since

$$R_n(T, G^{(n+1)}) - R(G^{(n+1)}) = \int E_n \mid t_n(\hat{y}) - t(y; G^*) \mid^2 f_{G^*}(y)\, dy$$

the result follows, noting that the final term in the above inequality is asymptotically dominant.

4. Example and remarks.

(1) Suppose a typical quality control variables sampling situation in which the process generating lots deteriorates with time. Each lot is given a true mean θ according to the process G and a sample of k items from the lot yields a mean $\bar{x}$ which is normally distributed with true mean θ. Suppose the deterioration in G over time is described by a drift in the mean λ from a lower bound b to an upper

bound $(a + b)$ as a function of time, say $\lambda_t = b + a(1 - e^{-t})$. According to the particular sampling rate, a sequence of t's is generated. Thus setting $Z_i = 1 - e^{-t_i}$ gives the linear drift model where λ_i is the mean of the process during the generation of the ith lot. The empirical Bayes estimator for the true mean of the $(n + 1)$st lot is given by (3.7).

(2) Upon strengthening the convergence of the least squares estimators given in (3.5) to a.e., it is easy to show that $P(\lim F_{n,n}(y) = F_{G^*}(y)) = 1$. Thus direct estimation of the prior distribution may be considered and in particular Theorem 2 of Robbins (1964) can be extended to the changing prior case treated in this paper.

(3) The definition of asymptotically optimal in the sense of (1.4) is equivalent to

$$\lim R(t_n, G^{(n+1)}) = R(G^{(0)})$$

where

$$G^{(0)}(\theta) = G^*(\theta - az_0 - b), \quad z_0 = \lim z_n.$$

One could then view the problem as a usual empirical Bayes problem with the same prior $G^{(0)}$ and proceed to estimate $G^{(0)}$ (or find a procedure whose risk is close to $R(G^{(0)})$) using the past history up to stage n. Whereas these two views are equivalent in the limiting case, the approach we have taken emphasizes the fact that as n increases, the difference between our "estimate" and the "present" situation, whatever it may be at the moment, is small. It also appears that a smaller number of observations is necessary to make $| R(t_n, G^{(n+1)}) - R(G^{(n+1)}) | \leq \varepsilon$ than to make $| R(t_n, G^{(n+1)}) - R(G^{(0)}) | < \varepsilon$ although precise statements are not yet available on this matter.

(4) The results given above are nonparametric inasmuch as they rely upon the kernel function estimators of a probability density function. If a parametric form is assumed for $G^{(n+1)}(\theta)$, the Bayes procedure may often be expressed directly in terms of the prior parameters. In such cases, it may be possible to remove the condition that $f(x \mid \theta)$ and $g^{(n+1)}(\theta)$ are location parameter densities. As an example, suppose $f(x \mid \theta)$ is Poisson with parameter θ and that $g^{(i)}(\theta)$ is a gamma (α_i, β_i) density, β_i known. The Bayes procedure for squared error loss estimation of θ is

$$t(x; G^{(n+1)}) = (\alpha_{n+1} + x)/(\beta_{n+1} + 1). \tag{4.1}$$

If we suppose $\alpha_i = az_i + b$, $\{z_i\}$ known, and let $\alpha_{n+1,n+1} = a_{n+1}z_{n+1} + b_{n+1}$, where a_{n+1}, b_{n+1} are the ordinary least squares estimates for a, b respectively, then defining $t_n(x)$ by replacing α_{n+1} by $\alpha_{n+1,n+1}$ in (4.1) will yield that $T = \{t_n\}$ is a.o. with a convergence rate of $O(n^{-1})$ provided only that $\mathrm{Var}(X_i) = \alpha_i(\beta_i + 1)/\beta_i^2 < \infty$.

(5) It is clear from the development given here that various questions remained

unanswered. A general theory for a drift in the mean of the prior (i.e mixing distribution) would be desirable as well as statistical tests for validating the type of drift. Tests concerning G fixed should be developed if empirical Bayes procedures are to find application. However, the practical but difficult problem that remains unsolved not only for the linear drift model but for the usual empirical Bayes model as well is the calculation of a lower bound for the number of component observations necessary to make the present risk arbitrarily close to the Bayes risk. Specifically, using the notation in this paper, given $\varepsilon > 0$, find N_ε such that $n \geq N_\varepsilon \Rightarrow | R_n(t_n, G^{(n+1)}) - R(G^{(n+1)}) | \leq \varepsilon$.

REFERENCES

DEELY, J. J. and LINDLEY, D. V. (1981). Bayes empirical Bayes. *J. Amer. Statist. Assoc.* **76** 833–841.

DEELY, J. J. and ZIMMER, W. J. (1976). Asymptotic optimality of the empirical Bayes procedure. *Ann. Statist.* **4** 576–580.

EVERITT, B. S. and HAND, H. J. (1981). Finite mixture distributions. *Monographs on Applied Probability and Statistics.* Chapman, London.

FERGUSON, T. S. (1962). Location and scale parameters in the exponential family of distributions. *Ann. Math. Statist.* **33** 986–1001.

GUPTA, S. S. and HUANG, WEN TAO (1980). On mixtures of distributions: a survey and some new results on ranking and selection. *Dept. of Statistics Mimeo Series* 79–22. Purdue University.

HOADLEY, B. (1979). An empirical Bayes approach to quality assurance. *33rd Annual Technical Conference Transactions*, American Society for Quality Control.

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

LIN, P. E. (1975). Rates of convergence in empirical Bayes estimation problems: continuous case. *Ann. Statist.* **3** 155–164.

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

O'BRYAN, T. and SUSARLA, V. (1975). An empirical Bayes estimation problem with non-identical components involving normal distributions. *Comm. Statist.* **4** 1033–1042.

O'BRYAN, T. and SUSARLA, V. (1976). Rates in the empirical Bayes estimation problems with non-identical components. Case of normal distributions. *Ann. Inst. Statist. Math.* **28** 389–397.

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

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1079.

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

SCHUSTER, E. F. (1969). Estimation of a probability density function and its derivatives. *Ann. Math. Statist.* **40** 1187–1195.

DEPARTMENT OF STATISTICS
AND COMPUTER SCIENCE
UNIVERSITY OF GEORGIA
ATHENS, GEORGIA 30602

UNIVERSITY OF CANTERBURY
DEPARTMENT OF MATHEMATICS
CHRISTCHURCH 1
NEW ZEALAND