CONVERGENCE RATES FOR EMPIRICAL BAYES
TWO-ACTION PROBLEMS: THE UNIFORM $U(0,\theta)$ DISTRIBUTION[1]

MOHAMED TAHIR
Temple University
Department of Statistics
Philadelphia, PA 19122

ABSTRACT

The purpose of this paper is to study the convergence rates of a sequence of empirical Bayes decision rules for the two-action problems in which the observations are uniformly distributed over the interval $(0,\theta)$, where θ is a value of a random variable having an unknown prior distribution. It is shown that the proposed empirical Bayes decision rules are asymptotically optimal and that the order of associated convergence rates is $O(n^{-\alpha})$, for some constant α, $0 < \alpha < 1$, where n is the number of accumulated past observations at hand.

Key words: Asymptotically optimal, Bayes risk, convergence rates, empirical Bayes decision rule, prior distribution.

AMS (MOS) subject classifications: 62C12.

1. INTRODUCTION

In situations involving sequences of similar but independent statistical decision problems, it is reasonable to formulate the component problem in the sequence as a Bayes statistical decision problem with respect to an unknown prior distribution over the parameter space, and then use the accumulated observations from the previous decision problems to improve the decision rule at each stage. This approach was first developed by Robbins [5] and was later studied in estimation and hypothesis testing problems by many authors. For example, Robbins [6] and Samuel [7] exhibit empirical Bayes rules for the two-action problems in which the distributions of the observations belong to a certain exponential family of probability distributions. Johns and Van Ryzin [1] study the convergence rates of a sequence of empirical Bayes rules which they propose for the two-action problems where the observations are members of some continuous exponential families of distributions. Also, Susarla and O'Bryan [8] and Nogami [4] consider estimation problems in which the observations are uniformly distributed over the interval $(0,\theta)$. These authors show that their empirical Bayes rules are asymptotically

optimal, in the sense that the associated Bayes risks of these rules converge to the minimum Bayes risk which would have been obtained if the prior distribution were completely known and the Bayes rule with respect to this prior distribution were used.

The objective of this paper is to investigate the convergence rate of a sequence of empirical Bayes decision rules for the two-action problems in which the observations are uniformly distributed over the interval $(0, \theta)$, where θ is a value of a random variable having an unknown prior distribution.

Let X denote the random observation of interest in the component decision problem and suppose that X has density function

$$f_\theta(x) = \tfrac{1}{\theta} I_{(0,\theta)}(x),$$

where θ is an unknown number such that $0 < \theta < b \leq \infty$, and $I(\cdot)$ denotes the indicator function. It is desired to determine a decision rule for choosing between the hypotheses

$$H_0: \theta \leq \theta_0 \text{ and } H_1: \theta > \theta_0,$$

where θ_0 is a given number such that $0 < \theta_0 < b$. Let a_0 and a_1 be the two possible actions, of which a_i is appropriate when H_i is true, $i = 0, 1$, and suppose that the loss function is of the form

$$L_0(\theta) = (\theta - \theta_0)^+$$

$$L_1(\theta) = (\theta_0 - \theta)^+ \tag{1.1}$$

for $0 < \theta < b$, where for $i = 0, 1$, $L_i(\theta)$ measures the loss incurred when action a_i is taken and θ is the true value of the parameter, and where $(u)^+ = max\{u, 0\}$.

Suppose that θ is a realization of a random variable Θ having an unknown prior distribution function G. Thus, the formulation of the problem assumes that the random observation X is conditionally uniformly distributed over $(0, \theta)$, given $\Theta = \theta$, where Θ is a random variable having an unknown prior distribution G; that, based on X, an action $a \in \{a_0, a_1\}$ is to be taken; and that if action a is taken, then the loss incurred is of the form (1.1).

Let

$$\delta(x) = P\{\text{accepting } H_0 \mid X = x\}$$

be a randomized decision rule for the above decision problem. Then the Bayes risk incurred by

using the decision rule δ with respect to the prior distribution G is given by

$$R(G,\delta) = \int\int\left[L_0(\theta)f_\theta(x)\delta(x) + L_1(\theta)f_\theta(x)(1-\delta(x))\right]dx dG(\theta)$$

$$= \int_0^b a(x)\delta(x)dx + C(G), \tag{1.2}$$

where

$$a(x) = \int_x^b \theta f_\theta(x)dG(\theta) - \theta_0 f(x) \tag{1.3}$$

$$f(x) = \int_x^b f_\theta(x)dG(\theta)$$

for $0 < x < b$ and

$$C(G) = \int_0^b L_1(\theta)dG(\theta).$$

Note that $C(G)$ is a constant which is independent of the decision rule δ. Hence, it follows from (1.2) that a Bayes decision rule, δ_G, is clearly given by

$$\delta_G(x) = \begin{cases} 1 & \text{if } a(x) \le 0 \\ 0 & \text{if } a(x) > 0. \end{cases} \tag{1.4}$$

The Bayes decision rule δ_G cannot be applied to the decision problem under study since it depends on the unknown prior distribution G. In view of this remark, an empirical Bayes approach is used with prior distributions G for which $\int_0^b \theta dG(\theta) < \infty$, to insure that the Bayes risk is always finite.

Section 2 provides a sequence of empirical Bayes rules for the decision problem described above. Section 3 establishes the asymptotic optimality of the proposed rules and investigates the convergence rates of these rules. Section 4 contains an example which illustrates the results of this paper.

2. THE PROPOSED EMPIRICAL BAYES RULES

The construction of a sequence of empirical Bayes decision rules for the decision problem described in the previous section is motivated by the following observations. First, $a(x)$ in (1.3) can be rewritten as

$$a(x) = 1 - G(x) - \theta_0 f(x), \tag{2.1}$$

by using the definition of $f_\theta(x)$. Furthermore, the conditional distribution function of X, given $\Theta = \theta$, is given by

$$F_\theta(x) = \int\limits_0^x f_\theta(t)dt = xf_\theta(x) + I_{(\theta,\infty)}(x)$$

for $0 < x < b$; so that, the marginal distribution function of X is given by

$$F(x) = \int\limits_0^b F_\theta(x)dG(\theta) = xf(x) + G(x)$$

for $0 < x < b$. Therefore,

$$a(x) = 1 - F(x) + (x - \theta_0)f(x) \tag{2.2}$$

for $0 < x < b$, by (2.1).

An empirical Bayes decision rule for the $(n+1)st$ decision problem may be obtained by first estimating $a(x)$, for each x, using the n accumulated observations from the previous problem, and then adopting an empirical Bayes decision rule which is similar, in a sense, to the Bayes rule δ_G, but which does not require the knowledge of the prior distribution G. Specifically, let $X_1,\ldots,X_n$ be the observations from n past experiences of the component decision problem, where for each $i = 1,\ldots,n$, X_i is conditionally uniformly distributed over the interval $(0,\theta_i)$, given $\Theta_i = \theta_i$, and $\Theta_1,\Theta_2,\ldots$ are independent and identically distributed ($i.i.d.$) random variables with common distribution G. Hence, $X_1,X_2,\ldots$ are $i.i.d.$ with common distribution F. Let $X_{n+1} \equiv X$ denote the current observation and let

$$F_n(x) = \tfrac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le x\}}$$

denote the natural estimator of $F(x)$. Also, since by definition of a density function,

$$f(x) = \lim_{h\to 0}\frac{F(x+h) - F(x)}{h},$$

$f(x)$ can be estimated by

$$f_n(x) = \frac{F_n(x+h_n) - F_n(x)}{h_n}$$

for $n \ge 1$, where h_n, $n \ge 1$, is a sequence of positive real numbers such that $h_n\to 0$ and $nh_n\to\infty$ as $n\to\infty$. In view of (2.2), $a(x)$ can be estimated by

$$a_n(x) = 1 - F_n(x) + (x - \theta_0)f_n(x) \tag{2.3}$$

for each x, $0 < x < b$.

In view of (1.4) and the above observations, define the empirical Bayes decision rule for the $(n+1)st$ decision problem by

$$\delta_n(x) = \begin{cases} 1 & if \quad a_n(x) \le 0 \\ 0 & if \quad a_n(x) > 0. \end{cases} \tag{2.4}$$

for each x, $0 < x < b$. Then clearly, δ_n does not depend on the unknown prior distribution G.

3. ASYMPTOTIC OPTIMALITY

Let Δ denote the class of all decision rules, and let $R^*(G)$ denote the minimum Bayes risk with respect to the prior distribution G. Then,

$$
\begin{aligned}
R^*(G) &= \inf_{\delta \in \Delta} R(G,\delta) \\
&= R(G,\delta_G) \\
&= \int_0^b a(x)\delta_G(x)dx + C(G),
\end{aligned} \tag{3.1}
$$

as in (1.2). Also, let $R_n^*(G)$ denote the Bayes risk incurred by using the empirical Bayes rule δ_n, defined by (2.4), with respect to the prior distribution G. Then,

$$
\begin{aligned}
R_n^*(G) &= E\big[R(G,\delta_n)\big] \\
&= \int_0^b a(x)E\big[\delta_n(x)\big]dx + C(G),
\end{aligned} \tag{3.2}
$$

where E denotes expectation with respect to the marginal joint distribution of $X_1,...,X_n$. Then, $R_n^*(G) \geq R^*(G)$ for all $n \geq 1$, since $R^*(G)$ is the minimum risk, and hence $R_n^*(G) - R^*(G)$ may be used as a measure of the optimality of the empirical Bayes rule δ_n.

Lemma 3.1: *For $n \geq 1$,*

$$
R_n^*(G) - R^*(G) \leq \int_0^b |a(x)| P\Big\{|a_n(x) - a(x)| > |a(x)|\Big\}dx.
$$

Proof: An application of (3.1) and (3.2), followed by (1.4) and (2.4), yields

$$
\begin{aligned}
R_n^*(G) - R^*(G) &= \int_0^b a(x)\big[P\{a_n(x) \leq 0\} - \delta_G(x)\big]dx \\
&= \int_0^b |a(x)| B(n,x)dx,
\end{aligned} \tag{3.3}
$$

where

$$
B(n,x) = \begin{cases} P\{a_n(x) > 0\} & if \quad a(x) \leq 0 \\ P\{a_n(x) \leq 0\} & if \quad a(x) > 0 \end{cases}
$$

for $0 < x < b$. The lemma now follows since $|a_n(x) - a(x)| > |a(x)|$ is implied by $a_n(x) > 0$ when $a(x) \leq 0$, and by $a_n(x) \leq 0$ when $a(x) > 0$.

Definition 3.1:   Let ν_n, $n \geq 1$ be a sequence of positive numbers such that $\nu_n \to 0$ as $n \to \infty$. A sequence of Bayes empirical estimators δ_n, $n \geq 1$, is said to be asymptotically optimal at least of order ν_n if $R_n^*(G) - R^*(G) \leq O(\nu_n)$ as $n \to \infty$.

The main result is presented next.

Theorem 3.1:   Let δ_n be as in (2.4) and suppose that f', the derivative of f, exists and is continuous. If

$$(i) \qquad \int_0^b \theta \, dG(\theta) < \infty,$$

and if for some r, $0 < r < 2$,

$$(ii) \qquad \int_0^b |a(x)|^{1-r} \left[F(x)(1 - F(x))\right]^{r/2} dx < \infty,$$

$$(iii) \qquad \int_0^b |x - \theta_0|^r |a(x)|^{1-r} \left[f_0^*(x)\right]^{r/2} dx < \infty,$$

$$(iv) \qquad \int_0^b |x - \theta_0|^r |a(x)|^{1-r} \left[f_1^*(x)\right]^r dx < \infty,$$

where $f_0^*(x) = \sup_{0 < t \leq \epsilon} f(x + t)$ and $f_1^*(x) = \sup_{0 < t \leq \epsilon} |f'(x + t)|$, for some $\epsilon > 0$, then choosing $h_n = n^{-\frac{1}{2\beta + 1}}$, for some β, $0 < \beta < 1$, yields

$$R_n^*(G) - R^*(G) \leq O(n^{-\alpha})$$

as $n \to \infty$, where $\alpha = \frac{\beta r}{2\beta + 1}$. Thus, the sequence δ_n, $n \geq 1$, is asymptotically optimal at least of order $n^{-\alpha}$ with respect to the prior distribution G.

Proof: Let $r > 0$ be given. It follows from Lemma 3.1 and Markov's inequality that

$$R_n^*(G) - R^*(G) \leq \int_0^b |a(x)|^{1-r} E|a_n(x) - a(x)|^r dx \qquad (3.3)$$

for all $n \geq 1$. Furthermore, by (2.2), (2.3), and the c_r-inequality (Loève, [3])

$$
\begin{aligned}
E|a_n(x) - a(x)|^r &= E\left|F(x) - F_n(x) + (x - \theta)[f_n(x) - f(x)]\right|^r \\
&\leq c_r E|F_n(x) - F(x)|^r + c_r^2 |x - \theta_0|^r E|f_n(x) - H(x;n)|^r \\
&\quad + c_r^2 |x - \theta_0|^r |H(x;n) - f(x)|^r
\end{aligned}
\qquad (3.4)
$$

for each $x > 0$, where $c_r = 1$ or 2^{r-1} according as $0 < r \leq 1$ or $1 \leq r < 2$ and

$$H(x;n) = \frac{F(x + h_n) - f(x)}{h_n}$$

for $x > 0$. Next, since $E[F_n(x)] = F(x)$ and $Var[F_n(x)] = \frac{1}{n}[1 - F(x)]F(x)$,

$$
\begin{aligned}
E|F_n(x) - F(x)|^r &\leq \left(Var[F_n(x)]\right)^{r/2} + \left|E[F_n(x)] - F(x)\right|^r \\
&= \left(Var[F_n(x)]\right)^{r/2} \\
&= n^{-r/2}[F(x)(1 - F(x))]^{r/2}
\end{aligned}
\tag{3.5}
$$

for all $x > 0$. Similarly,

$$
E|f_n(x) - H(x;n)|^r \leq \left(Var[f_n(x)]\right)^{r/2} \leq n^{-r/2}[f(x + x^*)]^{r/2},
\tag{3.6}
$$

where $0 < x^* < h_n$, since $E[f_n(x)] = H(x;n)$, $Var[f_n(x)] \leq \frac{1}{nh_n}H(x;n)$ and $H(x;n) = f(x + x^*)$. Finally, a Taylor's expansion of $F(x + h_n)$ about x yields

$$
H(x;n) - f(x) = \frac{1}{2}h_n f'(x + x_*),
\tag{3.7}
$$

where $0 < x_* < h_n$. Letting $h_n = n^{-\frac{1}{2\beta + 1}}$ and combining (3.3)-(3.7) yields

$$
\begin{aligned}
R_n^*(G) - R^*(G) &\leq \frac{K_1}{n^{r/2}} + \frac{K_2}{(nh_n)^{r/2}} + K_3 h_n^r \\
&= O(n^{-\alpha})
\end{aligned}
$$

as $n \to \infty$, where $\alpha = \frac{\beta r}{2\beta + 1}$ and

$$
K_1 = c_r \int_0^b |a(x)|^{1-r} [F(x)(1 - F(x))]^{r/2} dx
$$

$$
K_2 = c_r^2 \int_0^b |x - \theta_0|^r |a(x)|^{1-r} [f_0^*(x)]^{r/2} dx
$$

$$
K_3 = 2^{-r} c_r^2 \int_0^b |x - \theta_0|^r |a(x)|^{1-r} [f_1^*(x)]^r dx
$$

are all finite by assumptions (ii)-(iv) of the theorem.

4. EXAMPLE

The following example provides a class of prior distributions G to which Theorem 3.1 can be applied.

Suppose that G is the gamma distribution with density function

$$
g(\theta) = \begin{cases} s^2 \theta e^{-s\theta} & if & \theta > 0 \\ 0 & if & \theta \leq 0, \end{cases}
$$

where $s > 0$ is a given number. Then

$$
f(x) = \begin{cases} se^{-sx} & if & x > 0 \\ 0 & if & x \leq 0, \end{cases}
$$

$$F(x) = \begin{cases} 1 - e^{sx} & if \qquad x > 0 \\ 0 \quad if & \qquad x \le 0, \end{cases}$$

and

$$a(x) = \begin{cases} (1 + sx - s\theta_0)e^{-sx} & if \qquad x > 0 \\ 0 \quad if & \qquad x \le 0. \end{cases}$$

Also, $f_0^*(x) = f(x)$ and $f_1^*(x) = sf(x)$. Clearly, Assumption (i) of Theorem 3.1 is satisfied. To verify Condition (ii), let $\theta_0 > 0$ be a given number and observe that

$$\int_0^\infty |a(x)|^{1-r} \left[F(x)(1 - F(x))\right]^{r/2} dx \le \int_0^{\theta_0} |1 + sx - s\theta_0|^{1-r} dx$$

$$+ \int_{\theta_0}^\infty (1 + sx - s\theta_0)^{1-r} e^{-(1-\frac{r}{2})sx} dx$$

$$= I_1 + I_2,$$

say. Moreover,

$$I_1 \le c_r \theta_0 |1 - s\theta_0|^{1-r} + c_r(2 - r)^{-1} s^{1-r} \theta_0^{2-r},$$

by a simple integration, where $c_r = 1$ or 2^{r-1} according as $0 < r \le 1$ or $1 < r < 2$; and

$$I_2 \le \int_{\theta_0}^\infty (1 + sx)^{1-r} e^{-(1-\frac{r}{2})sx} dx < \infty$$

when $0 < r \le 1$. When $1 \le r < 2$, $(1 + sx - s\theta_0)^{1-r} \le 1$ for all $x > \theta_0$ and thus,

$$I_2 \le \int_{\theta_0}^\infty e^{-(1-\frac{r}{2})sx} dx < \infty.$$

Therefore, Condition (ii) of Theorem 3.1 is satisfied when $0 < r < 2$.

To verify Condition (iii) of Theorem 3.1, note that

$$\int_0^\infty |x - \theta_0|^r |a(x)|^{1-r} \left[f_0^*(x)\right]^{r/2} dx \le s^{r/2}\theta_0^r \int_0^{\theta_0} |1 + sx - s\theta_0|^{1-r} dx$$

$$+ s^{r/2} \int_{\theta_0}^\infty x^r |1 - s\theta_0 + sx|^{1-r} e^{-(1-\frac{r}{2})sx} dx$$

$$= s^{r/2}(J_1 + J_2), \tag{4.1}$$

say. Furthermore, by the c_r-inequality,

$$J_1 \leq c_r \theta_0^{r+1} |1 - s\theta_0|^{1-r} + c_r(2-r)^{-1} s^{1-r} \theta_0^2$$

and

$$J_2 \leq c_r |1 - s\theta_0|^{1-r} \int_{\theta_0}^{\infty} x^r e^{-(r-\frac{1}{2})sx} dx$$

$$+ c_r s^{1-r} \int_{\theta_0}^{\infty} xe^{-(1-r/2)sx} dx$$

$$< \infty. \tag{4.2}$$

Finally, to verify Condition (*iv*) of the theorem, observe that

$$\int_0^{\theta_0} |x - \theta_0|^r |a(x)|^{1-r} \left[f_1^*(x)\right]^r dx \leq s^{2r}(J_1 + J_2^*),$$

where J_1 is as in (4.1) and

$$J_2^* = \int_{\theta_0}^{\infty} x^r |1 + sx - s\theta_0|^{1-r} e^{-rsx} dx < \infty$$

as in (4.2).

REFERENCES

[1] M.V. Johns and J. Van Ryzin, "Convergence rates for empirical Bayes two-action problems II. Continuous case", *Ann. Math. Statist.* 43, (1972), pp. 934-947.

[2] R.L. Lehman, "*Testing Statistical Hypotheses*", Wiley, New York, 1959.

[3] M. Loève, "*Probability Theory*", 3rd ed. Van Nostrand, Princeton, 1963.

[4] Y. Nogami, "Convergence rates for empirical Bayes estimation in the uniform $U(0,\theta)$ distribution", *Ann. Statist.* 16, (1988), pp. 1335-1341.

[5] H. Robbins, "An empirical Bayes approach to statistics", *Proc. Third Berkeley Symp. Math. Statist.* 1, (1955), pp. 157-163.

[6] H. Robbins, "The empirical Bayes approach to statistical decision problems", *Ann. Math. Statist.* 35, (1964), pp. 1-20.

[7] E. Samuel, "An empirical Bayes approach to the testing of certain parametric hypotheses, *Ann. Math. Statist.* 34, (1963), pp. 1370-1385.

[8] V. Susarla and T. O'Bryan, "Empirical Bayes interval estimates involving uniform distributions", *Comm. Statist. A-Theory Methods* 8, (1979), pp. 385-397.