Ann. Inst. Statist. Math.
Vol. 44, No. 1, 133–140 (1992)

ON A MONOTONE EMPIRICAL BAYES TEST PROCEDURE IN GEOMETRIC MODEL*

TACHEN LIANG[1] AND S. PANCHAPAKESAN[2]

[1] *Department of Mathematics, Wayne State University, Detroit, MI 48202, U.S.A.*
[2] *Department of Mathematics, Southern Illinois University, Carbondale, IL 62901-4408, U.S.A.*

(Received November 17, 1989; revised September 17, 1990)

Abstract. A monotone empirical Bayes procedure is proposed for testing $H_0 : \theta \geq \theta_0$ against $H_1 : \theta < \theta_0$, where θ is the parameter of a geometric distribution. The asymptotic optimality of the test procedure is established and the associated convergence rate is shown to be of order $O(\exp(-cn))$ for some positive constant c, where n is the number of accumulated past experience (observations) at hand.

Key words and phrases: Bayes, empirical Bayes, hypothesis testing, geometric, antitonic and isotonic regression, asymptotic optimality, convergence rate.

1. Introduction

The empirical Bayes approach to statistical decision is typically appropriate when the same decision problem is faced repeatedly and independently. It is reasonable in such situations to formulate the component problem in the sequence as a Bayes decision problem with respect to an unknown prior distribution on the parameter space and then improve the decision rule at each stage based on the accumulated observations. This approach is due to Robbins (1956, 1964, 1983).

Many empirical Bayes rules have been shown to be *asymptotically optimal* in the sense that the risk for the n-th decision problem converges to the optimal Bayes risk which would have been obtained if the prior distribution was fully known and the Bayes rule with respect to this prior was used. However, the practical significance of the asymptotic optimality depends on the convergence rate with which the risks for the successive decision problems approach the optimal Bayes risk. Recently, Liang (1988) studied the convergence rate of a sequence of empirical Bayes rules for two-action decision problems when the distributions of the observations belong to a discrete exponential family. His treatment of this problem is different from that of Johns and Van Ryzin (1971). In the present paper, our approach is

* This research was supported in part by the NSF Grants DMS-8702620 and DMS-8717799 at Purdue University.

similar to that of Liang (1988) except that we use a different smoothing technique to obtain a sequence of monotone decision rules. Our smoothing technique is more logical in the special case of a geometric model than that of Liang (1988) for the discrete exponential family in general. This is due to the fact that in the geometric model the marginal density $f(x)$ in (1.5) is monotonically decreasing in x and our smoothing makes the estimator $f_n(x)$ reflect this property; see also Remark 2.1.

Let X be a random observation with probability function

$$f(x \mid \theta) = \theta^x(1-\theta), \qquad x = 0, 1, 2, \ldots, \qquad 0 < \theta < 1. \tag{1.1}$$

We consider testing $H_0 : \theta \geq \theta_0$ against $H_1 : \theta < \theta_0$, where $\theta_0 \in (0, 1)$ is known. Let i denote the action of deciding in favor of H_i, $i = 0, 1$. Corresponding to the true value θ of the parameter and action i, let the loss be

$$L(\theta, i) = (1-i)(\theta_0 - \theta)I_{(0,\theta_0)}(\theta) + i(\theta - \theta_0)I_{[\theta_0,1)}(\theta), \tag{1.2}$$

where $I_A(\cdot)$ denotes the indicator function of the set A. The two terms in (1.2) are, respectively, the loss due to action 0(1) when $H_1(H_0)$ is true. It is assumed that θ is the value of a random variable Θ having an unknown prior distribution $G(\theta)$.

A decision rule d is a mapping $d : x \to [0, 1]$, where x is the observed value of X and the value $d(x)$ is the probability of taking action 0 given that $X = x$. Let D be the class of all decision rules, and $r(G, d)$ denote the Bayes risk associated with each rule $d \in D$. Then $r(G) = \inf_{d \in D} r(G, d)$ is the minimum Bayes risk in the class D.

The Bayes risk associated with any rule d can be expressed in the form:

$$r(G, d) = \sum_{x=0}^{\infty}[\theta_0 - \varphi(x)]d(x)f(x) + C, \tag{1.3}$$

where

$$\varphi(x) = f(x+1)/f(x), \tag{1.4}$$

$$f(x) = \int_0^1 f(x \mid \theta)dG(\theta) \qquad \text{and} \tag{1.5}$$

$$C = \sum_{x=0}^{\infty}\int_{\theta_0}^1 (\theta - \theta_0)f(x \mid \theta)dG(\theta). \tag{1.6}$$

Note that $f(x)$ is the marginal probability function of X, and $\varphi(x)$ is the posterior mean of θ given $X = x$. Further, C is a constant independent of the rule d. From (1.3), a Bayes decision rule d_G is easily seen to be

$$d_G(x) = \begin{cases} 1 & \text{if } \varphi(x) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{1.7}$$

Since the prior G is unknown, we use the empirical Bayes approach. We propose (Section 2) a sequence of empirical Bayes rules $\{d_n^*\}$ for the testing problem and establish (Section 3) its asymptotic optimality along with its convergence rate.

2. The proposed empirical Bayes rule

Let (X_j, Θ_j), $j = 1, 2, \ldots$, be a sequence of pairs of random variables, where the X_j are observable but the Θ_j are not. Conditional on $\Theta_j = \theta$, X_j has probability function $f(x \mid \theta)$ in (1.1). It is assumed that the Θ_j are iid having unknown distribution G. Therefore, the pairs (X_j, Θ_j) are iid. Let $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ denote the n past observations and let $X_{n+1} \equiv X$ denote the current observable whose observed value is x.

Now, for each $x = 0, 1, 2, \ldots$, let

$$f_n(x) = \frac{1}{n}\sum_{j=1}^{n} I_{\{x\}}(X_j) \quad \text{and} \quad F_n(y) = \sum_{x=0}^{y} f_n(x). \tag{2.1}$$

Since $f_n(x)$ is the empirical frequency estimator of $f(x)$, we have $f_n(x) \to f(x)$ with probability one. However, $f_n(x)$ *does not* possess the monotonically decreasing property that $f(x)$ has. In order to smooth our estimator, let $\{f_n^*(x)\}$ be the antitonic regression of $\{f_n(x)\}$ with equal weights for $x = 0, 1, \ldots, M_n + 1$, where $M_n = \max(x_1, \ldots, x_n) - 1$. We take $f_n^*(x) = f_n(x) = 0$ for $x > M_n + 1$, and let $F_n^*(y) = \sum_{x=0}^{y} f_n^*(x)$. The following properties of $f_n^*(x)$ and $F_n^*(y)$ easily follow: (1) $f_n^*(x)$ is nonincreasing in x with $\sum_{x=0}^{\infty} f_n^*(x) = 1$, (2) $F_n^*(y) \geq F_n(y)$ for all y, and (3) $\sup_{y\geq 0} |F_n^*(y) - F(y)| \leq \sup_{y \geq 0} |F_n(y) - F(y)|$, where $F(y) = \sum_{x=0}^{y} f(x)$. (For details regarding isotonic (antitonic) regression, see Barlow *et al.* (1972). A proof of Property (3) is on pp. 70–72.)

Now let $\varphi_n(x) = f_n^*(x+1)/f_n^*(x)$, $x = 0, 1, \ldots, M_n$. It is intuitively appealing to use $\varphi_n(x)$ to estimate $\varphi(x) = f(x+1)/f(x)$. However, $\varphi(x)$ is increasing in x (which can easily be verified), and $\varphi_n(x)$ may not exhibit this property. So we consider an isotonic smoothing of $\{\varphi_n(x)\}$; that is, we let $\{\varphi_n^*(x)\}$ be the isotonic regression of $\{\varphi_n(x)\}$ for $x = 0, 1, \ldots, M_n$ with weights $f_n^*(x)$. Also, let $\varphi_n^*(y) = \varphi_n^*(M_n)$ for all $y > M_n$.

Some important properties of φ_n^* are:

(A1) $\varphi_n^*(x)$ is nondecreasing in x.

(A2) If we let $\psi_n(y) = \sum_{x=0}^{y} \varphi_n(x) f_n^*(x)$ and $\psi_n^*(y) = \sum_{x=0}^{y} \varphi_n^*(x) f_n^*(x)$ for $y = 0, 1, \ldots, M_n$, then $\psi_n^*(y) \leq \psi_n(y)$ for all $y = 0, 1, \ldots, M_n$.

(A3) By Theorem 2.1 of Puri and Singh (1990),

$$\varphi_n^*(0) = \min_{0 \leq i \leq M_n} [\psi_n(i)/F_n^*(i)],$$
$$\varphi_n^*(x) = \min_{x \leq i \leq M_n} [(\psi_n(i) - \psi_n^*(x-1))/(F_n^*(i) - F_n^*(x-1))],$$

$x = 1, 2, \ldots, M_n$.

(A4) $\psi_n(y) = \sum_{x=0}^{y} f_n^*(x+1) = F_n^*(y+1) - F_n^*(0)$.

Remark 2.1. The smoothing technique of Liang (1988) uses

$$\varphi_n^L(x) = \min\left[\left\{\max_{0 \leq i \leq x} \frac{f_n(i+1) + \delta_n}{f_n(i) + \delta_n}\right\}, 1\right]$$

where δ_n is a positive value such that $\delta_n = o(1)$ and is arbitrarily chosen. Our smoothing technique does not involve such sequence. Further, $\varphi_n^L(x)$ tends to overestimate $\varphi(x)$ because of the use of max and consequently to overaccept H_0. This, as well as the nonpreservation by $f_n(x)$ of the monotonicity referred to in Section 1, makes our present smoothing technique preferable.

We now define our empirical Bayes rule as:

$$d_n^*(x) = \begin{cases} 1 & \text{if } \varphi_n^*(x) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

Remark 2.2. Since $\varphi_n^*(x)$ is nondecreasing in x, it follows that for $x < y$, $d_n^*(x) \leq d_n^*(y)$; in other words, the rule $d_n^*(x)$ is *monotone.*

3. Asymptotic optimality

In order to establish the asymptotic optimality of the empirical Bayes rules $\{d_n^*\}$, we need a few preliminary definitions and results.

The expected Bayes risk of any empirical Bayes procedure d_n for our testing problem is given by:

$$r(G, d_n) = \sum_{x=0}^{\infty} [\theta_0 - \varphi(x)] E[d_n(x)] f(x) + C \tag{3.1}$$

where the expectation is taken with respect to $(X_1, \ldots, X_n)$ and C is given in (1.6). Since $r(G) = r(G, d_G)$ is the minimum Bayes risk, $\Delta(G, d_n) \equiv r(G, d_n) - r(G) \geq 0$ for all n. Thus, the difference $\Delta(G, d_n)$ is a natural measure of the optimality of the empirical Bayes rule d_n.

DEFINITION 3.1. A sequence of empirical Bayes decision rules $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotic optimal at least of order α_n relative to the (unknown) prior distribution G if $\Delta(G, d_n) \leq O(\alpha_n)$ as $n \to \infty$, where $\{\alpha_n\}$ is a sequence of positive numbers such that $\lim_{n\to\infty} \alpha_n = 0$.

Now, let $A(\theta_0) = \{x \mid \varphi(x) > \theta_0\}$ and $B(\theta_0) = \{x \mid \varphi(x) < \theta_0\}$. Define

$$M = \begin{cases} \min A(\theta_0) & \text{if } A(\theta_0) \neq \emptyset, \\ \infty & \text{otherwise,} \end{cases} \tag{3.2}$$

and

$$m = \begin{cases} \sup B(\theta_0) & \text{if } B(\theta_0) \neq \emptyset, \\ -1 & \text{otherwise,} \end{cases} \tag{3.3}$$

where $\emptyset$ denotes the empty set. By the increasing property of φ, $m \leq M$; also, $m < M$ if $A(\theta_0) \neq \emptyset$. Furthermore,

$$x \leq m \quad (y \geq M) \quad \text{if and only if} \quad \varphi(x) < \theta_0 \quad (\varphi(y) > \theta_0). \tag{3.4}$$

For our purpose, we may assume that $A(\theta_0) \neq \emptyset$ so that $m < M < \infty$.

Before we proceed to state and prove our main theorem regarding the asymptotic optimality of the empirical Bayes rule d_n^*, we obtain some intermediate results which are given below as lemmas.

LEMMA 3.1. *For* $y \geq M$,

$$H(y) \equiv [F(M) - F(y+1)] - \theta_0[F(M-1) - F(y)]$$

is nonincreasing in y*; and* $H(M)$ *is negative.*

PROOF. $H(y+1)-H(y) = \theta_0 f(y+1) - f(y+2) = f(y+1)[\theta_0 - \varphi(y+1)] < 0$ by the definition of M, which proves the first part. $H(M) = \theta_0 f(M) - f(M+1) = f(M)[\theta_0 - \varphi(M)] < 0$.

LEMMA 3.2. *For* $0 \leq y \leq m$,

$$L(y) \equiv [F(y) - F(y+1)] - \theta_0[F(y-1) - F(y)]$$

is decreasing in y*, and* $L(m)$ *is positive, where* $F(-1) \equiv 0$.

PROOF. $L(y) = \theta_0 f(y) - f(y+1) = f(y)[\theta_0 - \varphi(y)] > 0$ by the definition of m; in particular, $L(m) > 0$. Since $f(y)$ and $[\theta_0 - \varphi(y)]$ are both positive and decreasing in y, so is $L(y)$.

LEMMA 3.3. *If* $M_n \geq M$*, then for* $M \leq y \leq M_n$,

$$\psi_n(y) - \psi_n^*(M-1) \geq F_n^*(y+1) - F_n^*(M).$$

PROOF. By Property (A2) of φ_n^*, $\psi_n(y) - \psi_n^*(M-1) \geq \psi_n(y) - \psi_n(M-1) = \sum_{x=M}^{y} \varphi_n(x) f_n^*(x) = \sum_{x=M}^{y} f_n^*(x+1) = F_n^*(y+1) - F_n^*(M)$.

LEMMA 3.4. *Let* $\tau_1 = \min\{(f(M+1) - \theta_0 f(M))^2/8, -\log F(M)\}$*, where* M*, defined in* (3.2)*, is assumed to be finite. Then*

$$P\{\varphi_n^*(M) < \theta_0\} \leq O(\exp(-\tau_1 n)).$$

PROOF. Let $X_{(n)} = \max(X_1, \ldots, X_n)$. Then

$$\begin{aligned} &P\{\varphi_n^*(M) < \theta_0\} \\ &\quad = P\{\varphi_n^*(M) < \theta_0 \text{ and } X_{(n)} \leq M\} + P\{\varphi_n^*(M) < \theta_0 \text{ and } X_{(n)} > M\} \\ &\quad = P_1 + P_2, \qquad \text{say.} \end{aligned}$$

Obviously, $P_1 \leq [F(M)]^n = e^{n \log F(M)}$. Using Property (A3) of φ_n^*,

$$
\begin{aligned}
P_2 &= P\left\{ \min_{M \le y \le M_n} [(\psi_n(y) - \psi_n^*(M-1))/(F_n^*(y) - F_n^*(M-1))] < \theta_0 \right. \\
&\qquad\qquad \left. \text{and } M \le M_n \right\} \\
&\le P\left\{ \min_{M \le y \le M_n} [(F_n^*(y+1) - F_n^*(M))/(F_n^*(y) - F_n^*(M-1))] < \theta_0 \right. \\
&\qquad\qquad \left. \text{and } M \le M_n \right\}, \\
&\quad \text{by Lemma 3.3} \\
&= P\{[F_n^*(y+1) - F_n^*(M)] - \theta_0[F_n^*(y) - F_n^*(M-1)] < 0 \\
&\qquad\qquad \text{for some } M \le y \le M_n\} \\
&\le P\{T(y+1) - T(M) - \theta_0 T(y) + \theta_0 T(M-1) < H(y) \text{ for some } y \ge M\}, \\
&\quad \text{where } T(y) = F_n^*(y) - F(y) \text{ and } H(y) \text{ is given in Lemma 3.1} \\
&\le P\{T(y+1) - T(M) - \theta_0 T(y) + \theta_0 T(M-1) < H(M) \text{ for some } y \ge M\}, \\
&\quad \text{by Lemma 3.1} \\
&\le P\left\{ T(y+1) < \frac{H(M)}{4} \text{ or } T(M) > \frac{-H(M)}{4} \text{ or } \theta_0 T(y) > \frac{-H(M)}{4} \right. \\
&\qquad\qquad \left. \text{or } \theta_0 T(M-1) < \frac{H(M)}{4} \text{ for some } y \ge M \right\} \\
&\le P\left\{ \sup_{y \ge M-1} |T(y)| > \frac{-H(M)}{4} \right\}, \text{ noting that } 0 < \theta_0 < 1 \text{ and } H(M) < 0 \\
&\le P\left\{ \sup_{y \ge 0} |F_n^*(y) - F(y)| > \frac{-H(M)}{4} \right\} \\
&\le P\left\{ \sup_{y \ge 0} |F_n(y) - F(y)| > \frac{-H(M)}{4} \right\} \\
&\le d \exp\left\{ -2n \left[\frac{H(M)}{4} \right]^2 \right\}, \text{ using Lemma 2.1 of Schuster (1969)} \\
&= d \exp\{-n(f(M+1) - \theta_0 f(M))^2/8\}, \text{ where } d \text{ is some positive constant.}
\end{aligned}
$$

Combining the orders of P_1 and P_2, we get the desired result.

LEMMA 3.5. *Let* $\tau_2 = \min\{f^2(m)[\theta_0 - \varphi(m)]^2/8, -\log F(m)\}$, *where* m, *defined in* (3.3), *is assumed to be finite. Then*

$$P\{\varphi_n^*(m) \ge \theta_0\} \le O(\exp(-\tau_2 n)).$$

PROOF. $P\{\varphi_n^*(m) \ge \theta_0\} = P\{\varphi_n^*(m) \ge \theta_0 \text{ and } X_{(n)} \le m\} + P\{\varphi_n^*(m) \ge \theta_0 \text{ and } X_{(n)} > m\} = Q_1 + Q_2$, say. Obviously, $Q_1 \le [F(m)]^n = e^{n \log F(m)}$. If

$\varphi_n^*(m) \geq \theta_0$, then by properties of isotonic regression, $\varphi_n(y) \geq \theta_0$ for some $y \leq m$. Thus

$$\begin{aligned}
Q_2 &\leq P\{\varphi_n(y) \geq \theta_0 \text{ for some } y \leq m \text{ and } X_{(n)} > m\} \\
&\leq P\{f_n^*(y+1) - \theta_0 f_n^*(y) \geq 0 \text{ for some } y \leq m\} \\
&= P\{[F_n^*(y+1) - F_n^*(y)] - \theta_0[F_n^*(y) - F_n^*(y-1)] \geq 0 \text{ for some } y \leq m\} \\
&= P\{T(y+1) - T(y) - \theta_0 T(y) + \theta_0 T(y-1) \geq L(y) \text{ for some } y \leq m\}, \\
&\quad \text{where } T(y) \text{ is as defined earlier with } T(-1) \equiv 0, \text{ and} \\
&\quad L(y) \text{ is given in Lemma 3.2} \\
&\leq P\{T(y+1) - T(y) - \theta_0 T(y) + \theta_0 T(y-1) \geq f(m)[\theta_0 - \varphi(m)] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for some } y \leq m\}, \\
&\quad \text{by Lemma 3.2.}
\end{aligned}$$

Now, proceeding as in Lemma 3.4, we get $Q_2 \leq d\exp\{-nf^2(m)(\theta_0 - \varphi(m))^2/8\}$. Combining the orders of Q_1 and Q_2, we obtain the desired result.

We now state and prove our main theorem dealing with the asymptotic optimality of the test procedure d_n^*.

THEOREM 3.1. *Let $\{d_n^*\}$ be the sequence of empirical Bayes test procedures defined in* (2.2). *Then, for $M < \infty$, $r(G, d_n^*) - r(G) \leq O(\exp(-cn))$ for some positive constant c.*

PROOF. We first note that

$$\begin{aligned}
r(G, d_n^*) - r(G) &= \sum_{x=0}^{m} [\theta_0 - \varphi(x)] P\{\varphi_n^*(x) \geq \theta_0\} f(x) \\
&\quad + \sum_{x=M}^{\infty} [\varphi(x) - \theta_0] P\{\varphi_n^*(x) < \theta_0\} f(x) \\
&\leq b_1 P\{\varphi_n^*(m) \geq \theta_0\} + b_2 P\{\varphi_n^*(M) < \theta_0\},
\end{aligned}$$

using the nondecreasing property of φ_n^*, and letting $b_1 = \sum_{x=0}^{m}[\theta_0 - \varphi(x)]f(x)$ and $b_2 = \sum_{x=M}^{\infty}[\varphi(x) - \theta_0]f(x)$. Also, b_1 and b_2 are nonnegative constants because of (3.4). Using the asymptotic behavior of $P\{\varphi_n^*(m) \geq \theta_0\}$ and $P\{\varphi_n^*(M) < \theta_0\}$ obtained in Lemmas 3.4 and 3.5, the theorem is established with $c = \min(\tau_1, \tau_2)$.

REFERENCES

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*, Wiley, New York.

Johns, M. V., Jr. and Van Ryzin, J. (1971). Convergence rates for empirical Bayes two-action problems I: discrete case, *Ann. Math. Statist.*, **42**, 1521–1539.

Liang, T. (1988). On the convergence rates of empirical Bayes rules for two-action problems: discrete case, *Ann. Statist.*, **16**, 1635–1642.

Puri, P. S. and Singh, H. (1990). On recursive formulas for isotonic regression useful for statistical inference under order restrictions, *J. Statist. Plann. Inference*, **24**, 1–11.

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

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

Robbins, H. (1983). Some thoughts on empirical Bayes estimation, *Ann. Statist.*, **11**, 713–723.

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