Rates of Convergence for Empirical Bayes
Two-Action Problems: Discrete Case*

by

TaChen Liang
Wayne State University

Technical Report # 91-71C

Department of Mathematics
Wayne State University

December, 1991

* This research was supported in part by the National Science Foundation, Grant DMS-8923071 at Purdue University.

*

RATES OF CONVERGENCE FOR EMPIRICAL BAYES TWO-ACTION PROBLEMS: DISCRETE CASE

by

TaChen Liang
Department of Mathematics
Wayne State University
Detroit, MI 48202

Abstract

We investigate the convergence rates of a sequence of empirical Bayes decision rules for the two-action decision problem with the underlying distributions belonging to a class of discrete exponential family. The asymptotic optimality of the empirical Bayes rules is studied under the situation where the assumption regarding the unknown prior distribution is limited to the minimum. It is found that the sequence of the empirical Bayes rules under study is asymptotically optimal, and the order of the associated convergence rates is either $O(\exp(-cn))$ for some positive constant c or $O(n^{-2\alpha}\exp(-2n^{1-\beta}))$, $\alpha > 0$, $0 < \beta \leq 1$, depending on two different situations related to the unknown prior distributions.

AMS 1980 subject classification. 62C12.

Key words and phrases: empirical Bayes, asymptotically optimal, rates of convergence.

1. INTRODUCTION

Since Robbins (1956, 1964), empirical Bayes procedures have been extensively studied in the literature, e.g., see Samuel (1963), Johns and Van Ryzin (1971, 1972), Lin (1972, 1975), Singh (1976, 1979), Van Houwelingen (1976, 1987), Van Ryzin and Susarla (1977), Karunamuni (1988), Nogami (1988) and Liang (1988), among the many others. Many of the authors were concerned with the asymptotic optimality of the empirical Bayes procedures. They established the best possible rates of convergence of the empirical Bayes procedures based on certain assumptions regarding the behavior of the unknown prior distributions. Several empirical Bayes procedures were constructed according to the assumptions. However, since the prior distribution is unknown it is hard to verify the assumptions. Hence whether or not the concerned empirical Bayes procedures achieve or near the best possible rates of convergence is doubtful. In this sense, one may be interested in limiting the assumptions regarding the unknown prior distribution to the minimum and seeing how good the performance of the empirical Bayes procedures can still be.

In this paper, we investigate the convergence rates of a sequence of empirical Bayes decision rules for the two-action problem with the underlying distributions belonging to a class of discrete exponential family. The asymptotic optimality of the empirical Bayes decision rules is studied under the situation where the assumption regarding the unknown prior distribution is limited to the minimum. It is found that the sequence of empirical Bayes decision rules under study is asymptotically optimal, and the order of the associated convergence rates is either $O(\exp(-cn))$ for some positive constant c, or $O(n^{-2\alpha}\exp(-2n^{1-\beta})), \alpha > 0, 0 < \beta \leq 1$, depending on two different situations related to the unknown prior distribution.

2. The Two-Action Problem

Let X denote a random observation with probability function $f(x|\theta) = h(x)\theta^x\beta(\theta)$, $x = 0, 1, 2, \ldots; 0 < \theta < Q$, where $h(x) > 0$ for all $x = 0, 1, \ldots$, and where Q may be finite or infinite. Consider the problem of testing $H_0 : \theta \geq \theta_0$ against $H_1 : \theta < \theta_0$ with the loss function

$$L(\theta, i) = (1 - i)(\theta_0 - \theta)I_{(0,\theta_0)}(\theta) + i(\theta - \theta_0)I_{(\theta_0,Q)}(\theta), \tag{2.1}$$

where θ_0 is a known positive constant, $0 < \theta_0 < Q, i(i = 0, 1)$ is the action deciding in favor of H_i, and I_A is the indicator function of the set A. It is assumed that the parameter θ is a realization of a random variable Θ having an unknown prior distribution G over $(0, Q)$.

For a decision rule $d, d(x)$ is defined as the probability of taking action 0 for given $X = x$. Let $r(G, d)$ denote the Bayes risk associated with the decision rule d. Then

$$r(G, d) = \sum_{x=0}^{\infty}[\theta_0 - \varphi(x)]d(x)f(x) + C, \tag{2.2}$$

where

$$\begin{aligned} &\varphi(x) = E[\Theta|X = x] = \frac{h(x)f(x+1)}{h(x+1)f(x)} \text{ is the posterior mean of } \Theta \text{ given } X = x, \\ &f(x) = \int_0^Q f(x|\theta)dG(\theta) \text{ is the marginal probability function of } X, \text{ and} \\ &C = \int_{\theta_0}^Q (\theta - \theta_0)dG(\theta). \end{aligned} \tag{2.3}$$

We consider only those priors G such that $\int_0^Q \theta dG(\theta) < \infty$ to insure that the Bayes risk is always finite and hence the problem is meaningful. This assumption always holds true when Q is finite. For example, in a negative binomial distribution, $f(x|\theta) = \binom{r+x-1}{r-1}\theta^x(1-\theta)^r, 0 < \theta < 1$, where r is a positive integer. In such a case, $Q = 1$.

From (2.2), a Bayes decision rule, say d_G, is clearly given by

$$d_G(x) = \begin{cases} 1 & \text{if } \varphi(x) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

The minimum of Bayes risks among the class of all decision rules is: $r(G) = r(G, d_G)$.

Let $B(\theta_0) = \{x|\varphi(x) < \theta_0\}$. Define

$$m = \begin{cases} \sup B(\theta_0) & \text{if } B(\theta_0) \neq \phi, \\ -1 & \text{if } B(\theta_0) = \phi, \end{cases} \tag{2.5}$$

where ϕ denotes the empty set.

A straightforward computation leads to that the posterior mean $\varphi(x)$ is increasing in x. By the definition of $m, \varphi(x) < \theta_0$ iff $x \leq m$. Therefore, the Bayes decision rule d_G can be represented as:

$$d_G(x) = \begin{cases} 1 & \text{if } x > m, \\ 0 & \text{if } x \leq m. \end{cases} \tag{2.6}$$

When the prior distribution G is unknown, Johns and Van Ryzin (1971) and Liang (1988) have studied this two-action problem using the empirical Bayes approach. Johns and Van Ryzin (1971) have proposed some empirical Bayes decision rule and studied the corresponding asymptotic optimality under certain assumptions relating to the behavior of the tail probability of the unknown prior distribution. They established the best possible rates of convergence to be of order n^{-1} where n is the number of the accumulated past data. Liang (1988) proposed an alternative empirical Bayes decision rule, say d_n^*, and found that the rate of convergence of d_n^* is of order $\exp(-cn)$ for some positive constant c under a very weak assumption that $m < \infty$. Liang (1988) also found that the assumptions made in Johns and Van Ryzin (1971) always imply the finiteness of m. However, the case where $m = \infty$ was not discussed. The basic assumption that $\int_0^Q \theta dG(\theta) < \infty$ was used in Johns and Van Ryzin (1971) and Liang (1988).

3. The Empirical Bayes Rules and Its Aymptotic Optimality

First, we recall some property related to this decision problem. Note that the class of the probability functions $\{f(x|\theta) \mid 0 < \theta < Q\}$ has monotone likelihood ratio in x. Under the loss function (2.1), the class of monotone decision rules is essentially complete; see Berger (1985). Hence, it is natural to desire that the proposed empirical Bayes rule be monotone.

For each $j = 1, 2, \ldots$, let (X_j, Θ_j) be a pair of random variables, where X_j is observable but Θ_j is not observable. Conditional on $\Theta_j = \theta_j, X_j$ has probability function $f(x|\theta_j)$. It is assumed that $\Theta_j, j = 1, 2, \ldots$, are independently distributed with common unknown prior distribution G. Therefore, $(X_j, \Theta_j), j = 1, 2, \ldots$, are iid. Let $\underset{\sim}{X}_n = (X_1, \ldots, X_n)$ denote the n past observations and let $X_{n+1} \equiv X$ denote the present random observation.

For each $x = 0, 1, 2, \ldots$, let $f_n(x) = \frac{1}{n} \sum_{j=1}^{n} I_{\{x\}}(X_j)$. Mimicking the form of the posterior mean $\varphi(x)$, (see (2.3)), let

$$\varphi_n(x) = \frac{h(x) f_n(x+1)}{h(x+1) f_n(x) + \delta(n, x)}, \tag{3.1}$$

where

$$\text{(3.2)} \qquad \begin{cases} \delta(n,x) = 3\theta_0^{-1}[h(x) + \theta_0 h(x+1)]\varepsilon(n,\alpha,\beta), \text{ and} \\ \varepsilon(n,\alpha,\beta) = (n^{-\beta} + \alpha n^{-1} \ln\, n)^{1/2}, \ \alpha > 0, \ 0 < \beta \le 1. \end{cases}$$

One may use $\varphi_n(x)$ to estimate $\varphi(x)$ and obtain an empirical Bayes rule based on $\varphi_n(x)$. However, $\varphi_n(x)$ does not possess the increasing property. A smoothed version of $\varphi_n(x)$, say $\tilde{\varphi}_n(x)$, is defined as follows. Let

$$\text{(3.3)} \qquad \tilde{\varphi}_n(x) = (\max_{0 \le y \le x} \varphi_n(y)) \wedge Q,$$

where $a \wedge b = \min(a,b)$. Then, we propose an empirical Bayes decision rule $\tilde{d}_n$ defined as:

$$\text{(3.4)} \qquad \tilde{d}_n(X) = \begin{cases} 1 & \text{if } \tilde{\varphi}_n(X) \ge \theta_0, \\ 0 & \text{otherwise.} \end{cases}$$

Since $\tilde{\varphi}_n(x)$ is nondecreasing in x, the empirical Bayes decision rule $\tilde{d}_n$ is a monotone rule. Note that the past data $\underset{\sim}{X}_n$ is implicitly contained in the subscript n of $\tilde{\varphi}_n$ and $\tilde{d}_n$.

Consider an empirical Bayes decision rule d_n. Let $r(G, d_n)$ be the Bayes risk associated with the rule d_n. Then,

$$\text{(3.5)} \qquad r(G, d_n) = \sum_{x=0}^{\infty} [\theta_0 - \varphi(x)] E_n[d_n(x)] f(x) + C,$$

where the expectation E_n is taken with respect to $\underset{\sim}{X}_n$. Since $r(G)$ is the minimum Bayes risk, $r(G, d_n) - r(G) \ge 0$ for all n. The nonnegative regret risk $r(G, d_n) - r(G)$ is used as a measure of the optimality of the empirical Bayes decision rule d_n.

<u>Definition 3.1.</u> A sequence of empirical Bayes rules $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal of order α_n relative to the unknown prior distribution G if $r(G, d_n) - r(G) = O(\alpha_n)$ as $n \to \infty$, where $\{d_n\}$ is a sequence of positive numbers such that $\alpha_n = o(1)$.

The sequence of the empirical Bayes rules $\{\tilde{d}_n\}$ has the following asymptotic optimality.

Theorem 3.1. Let $\{\tilde{d}_n\}_{n=1}^{\infty}$ be the sequence of empirical Bayes decision rules defined previously. Assume that $\int_0^Q \theta dG(\theta) < \infty$. Then, the following results hold.

(a) If $m < \infty$, $r(G, \tilde{d}_n) - r(G) = O(\exp(-cn))$ for some positive constant c.

(b) If $m = \infty$, $r(G, \tilde{d}_n) - r(G) = O(n^{-2\alpha} \exp(-2n^{1-\beta}))$, where $\alpha > 0$, $0 < \beta \leq 1$.

Proof: Part (a) can be proved by a proof analogous to that of Theorem 2.1 of Liang (1988). Hence, the detail is omitted here. In the following, we provide the proof for part (b) only.

When $m = \infty$, by the definition of m, $\varphi(x) < \theta_0$ for all x. By the definition of $\tilde{d}_n$, direct computation leads to

$$r(G, \tilde{d}_n) - r(G) = \sum_{x=0}^{\infty} [\theta_0 - \varphi(x)] P\{\tilde{\varphi}_n(x) \geq \theta_0\} f(x). \tag{3.6}$$

Let $F(x)$ denote the marginal distribution function of the random variable X and let $F_n(x)$ be the empirical distribution based on $\underset{\sim}{X}_n$. Also, define $F(-1) = F_n(-1) \equiv 0$. Let $R_n(x) = F_n(x) - F(x)$.

From (3.1) - (3.3), by the fact that $\varphi(x) - \theta_0 < 0$ for all x, following some algebraic operation, we can obtain: For each $x = 0, 1, \ldots,$

(3.7)

$$\begin{aligned}
&\{\tilde{\varphi}_n(x) \geq \theta_0\} \\
&= \bigcup_{y=0}^{x} \{\varphi_n(y) \geq \theta_0\} \\
&= \bigcup_{y=0}^{x} \{h(y)f_n(y+1) - h(y+1)f_n(y)\theta_0 \geq \theta_0\delta(n,y)\} \\
&= \bigcup_{y=0}^{x} \{h(y)F_n(y+1) - [h(y) + \theta_0 h(y+1)]F_n(y) + \theta_o h(y+1)F_n(y-1) \geq \theta_0\delta(n,y)\} \\
&\subset \bigcup_{y=0}^{x} \{h(y)R_n(y+1) - [h(y) + \theta_o h(y+1)]R_n(y) + \theta_0 h(y+1)R_n(y-1) > \theta_0\delta(n,y)\} \\
&\subset \bigcup_{y=0}^{x} \Big(\{h(y)R_n(y+1) > \frac{\theta_0}{3}\delta(n,y)\} \cup \{\theta_0 h(y+1)R_n(y-1) > \frac{\theta_0}{3}\delta(n,y)\} \\
&\qquad \cup \{[h(y) + \theta_0 h(y+1)]R_n(y) < -\frac{\theta_0}{3}\delta(n,y)\}\Big) \\
&\subset \bigcup_{y=0}^{x} (\{R_n(y+1) > \varepsilon(n,\alpha,\beta)\} \cup \{R_n(y) < -\varepsilon(n,\alpha,\beta)\} \cup \{R_n(y-1) > \varepsilon(n,\alpha,\beta)\}) \\
&\subset \{\sup_{z\geq 0} |R_n(z)| > \varepsilon(n,\alpha,\beta)\}.
\end{aligned}$$

By Lemma 2.1 of Schuster (1969), and (3.2) and (3.7), we obtain

(3.8)

$$\begin{aligned}
P\{\tilde{\varphi}_n(x) \geq \theta_0\} &\leq P\{\sup_{z\geq 0} |R_n(z)| > \varepsilon(n,\alpha,\beta)\} \\
&\leq k \exp(-2n\varepsilon^2(n,\alpha,\beta)) \\
&= k \exp(-2n^{1-\beta} - 2\alpha \ \ln \ n) \\
&= kn^{-2\alpha} \exp(-2n^{1-\beta}).
\end{aligned}$$

Note that in (3.8), the constant k is independent of the distribution F; see Schuster (1969). Also, the upper bound at the right-hand-side of (3.8) is independent of x. Therefore, from

(3.6) and (3.8), we conclude

$$
\begin{aligned}
r(G,\tilde{d}_n) - r(G) &\le kn^{-2\alpha}\exp(-2n^{1-\beta})\sum_{x=0}^{\infty}[\theta_0 - \varphi(x)]f(x) \\
&\le k\theta_0 n^{-2\alpha}\exp(-2n^{1-\beta}) \\
&= O(n^{-2\alpha}\exp(-2n^{1-\beta})).
\end{aligned}
$$

Hence, the proof of part (b) is complete.

4. Concluding Remarks

Johns and Van Ryzin (1971) have proposed some empirical Bayes rules for this two-action problem and investigated the corresponding asymptotic optimality under certain assumptions regarding the behavior of the tail probability of the unknown prior distribution G. They established the best possible rates of convergence to be of order n^{-1}. It should be noted that the only assumption regarding the unknown prior distribution G we made in this paper is: $\int_0^Q \theta dG(\theta) < \infty$. This assumption always holds when Q is finite. This assumption insures the finiteness of the Bayes risks so that the problem under study is meaningful. In other words, Theorem 3.1 says that the convergence rate of the sequence of the empirical Bayes decision rules $\{\tilde{d}_n\}_{n=1}^{\infty}$ is of order $n^{-2\alpha}\exp(-2n^{1-\beta}), \alpha > 0$, $0 < \beta \le 1$, with no further assumptions about the unknown prior distribution G (i.e., no matter whether $m < \infty$ or not), where the values of $\alpha > 0$ and $0 < \beta \le 1$ can be chosen arbitrarily. Of course, the choice of the values of α and β may effect the performance of the empirical Bayes decision rule $\tilde{d}_n$ for small to moderate n.

The empirical Bayes decision rule $\tilde{d}_n$ is constructed in a way similar to the empirical Bayes decision rule d_n^* of Liang (1988). When $m < \infty, d_n^*$ and $\tilde{d}_n$ are asymptotically equivalent in the sense that both d_n^* and $\tilde{d}_n$ have exponential rates of convergence. However, when $m = \infty$, the rule d_n^* cannot achieve the rates of convergence as described in part (b) of Theorem 3.1. In Liang (1988), the empirical Bayes estimator for $\varphi(x)$, say $\varphi_n^*(x)$, may overestimate $\varphi(x)$, and therefore, the rule d_n^* is forced to accept H_0 biasedly. To reduce such bias, our $\varphi_n(x)$ is defined in a way such that $\varphi_n(x)$ may underestimate $\varphi(x)$. Then, we use the simple monotonizing technique by defining $\tilde{\varphi}_n(x) = (\max_{0 \leq y \leq x} \varphi_n(y)) \wedge Q$ to force it back so that the bias may be corrected.

References

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

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

Karunamuni, Rohana J. (1988). On empirical Bayes testing with sequential components. *Ann. Statist.* **16** 1270-1282.

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

Lin, P.E. (1972). Rates of convergence in empirical Bayes estimation problems: Discrete case. *Ann. Inst. Statist. Math.* **24** 319-325.

Lin, P.E. (1975). Rates of convergence in empirical Bayes estimation problems: Continuous case. *Ann. Statist.* **3** 155-164.

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

Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 157-163. Univ. California Press.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann.*

Math. Statist. **35** 1-20.

Samuel, E. (1963). An empirical Bayes approach to the testing of certain parametric hypotheses. *Ann. Meth. Statist.* **34** 1370-1385.

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

Singh, R.S. (1976). Empirical Bayes estimation with convergence rates in non-continuous Lebesque-exponential families. *Ann. Statist.* **4** 431-439.

Singh, R.S. (1979). Empirical Bayes estimation in Lebesque-exponential families with ratio near the best possible rate. *Ann. Statist.* **7** 890-902.

Van Houwelingen, J.C. (1976). Monotone empirical Bayes tests for the continuous one parameter exponential fmaily. *Ann. Statist.* **4** 981-989.

Van Houwelingen, J.C. (1987). Monotone empirical Bayes tests for uniform distributions using the maximum likelihood estimator of a decreasing density. *Ann. Statist.* **15** 875-879.

Van Ryzin, J. and Susarla, V. (1977). On the empirical Bayes approach to multiple decision problems. *Ann. Statist.* **5** 172-181.