ON EMPIRICAL BAYES TEST WITH $0 - L_i$ LOSS IN SOME NONEXPONENTIAL FAMILIES*

by

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

Technical Report # 91-33C

Department of Statistics
Purdue University

July 1991
Revised March 1992

*This research was supported in part by NSF Grants DMS–8923071 and DMS–8717799 at Purdue University.

◇

ON EMPIRICAL BAYES TEST WITH $0 - L_i$ LOSS
IN SOME NONEXPONENTIAL FAMILIES

by

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

Abstract

We study the problem of testing $H_0: \theta > \theta_0$ against $H_1: \theta \leq \theta_0$ with $0 - L_i$ loss for a truncation parameter distribution family through the nonparametric empirical Bayes approach. A monotone empirical Bayes testing procedure d_n^* is proposed. The asymptotic optimality of the empirical Bayes procedure d_n^* is investigated. It is shown that under certain regularity conditions, the associated convergence rate of the empirical Bayes procedure d_n^* is of order $O(((\ell n\ n)^{1+\delta}/n)^{2/3})$ where n is the number of accumulated experience at hand and δ is a small positive number.

AMS 1980 Subject Classification: 62C12

Keywords and phrases: Empirical Bayes, asymptotically optimal, rate of convergence, 0-L_i loss.

1. Introduction

Since Robbins (1956, 1964), empirical Bayes theory has been developed extensively in the literature; for example, see Johns and Van Ryzin (1971, 1972), Van Houwelingen (1976) and Stijnen (1982) for empirical Bayes testing, and Lin (1972, 1975), Singh (1976, 1979) for empirical Bayes estimation. Many of the authors are dealing with exponential family distributions with either linear error loss (for testing problem) or squared error loss (for estimation problem). In recent years, there appears to be a growing interest in empirical Bayes analysis for nonexponential families. Fox (1970, 1978), Wei (1983, 1985) and Nogami (1988) have considered empirical Bayes estimation for the parameter θ of a uniform distribution $\mathcal{U}(0,\theta)$ with squared error loss. Van Houwelingen (1987) and Liang (1990) have studied empirical Bayes testing procedures for a $\mathcal{U}(0,\theta)$ distribution with linear error loss. Gupta and Hsiao (1983) and Huang and Liang (1990) have developed empirical Bayes selection procedures for $\mathcal{U}(0,\theta)$ distributions. Recently, Datta (1991) studied an empirical Bayes estimation problem with squared error loss for a class of distribution having pdf $f(x|\theta) = \frac{a(x)}{A(\theta)} I_{(0,\theta)}(x)$, where $a(x) > 0$ for $x > 0$ and $A(\theta) = \int_0^\theta a(x)dx < \infty$, which includes uniform distributions $\mathcal{U}(0,\theta)$ as a special case. Also, Prasad and Singh (1990) studied an empirical Bayes estimator for the truncated location parameter in a truncated exponential distribution.

In this paper, we study an empirical Bayes testing procedure for testing $H_0: \theta > \theta_0$ against $H_1: \theta \leq \theta_0$ with $0-L_i$ loss in a distribution with pdf $f(x|\theta)$ given above. The general setup of this testing problem is described in Section 2. Motivated by the behavior of a Bayes decision procedure, a monotone empirical Bayes testing procedure d_n^* is proposed in Section 3. Finally we investigate the asymptotic optimality of the empirical Bayes procedure d_n^*. It is shown that under certain regularity conditions, the associated convergence rate of the empirical Bayes procedure d_n^* is of order $O(((\ell n\ n)^{1+\delta}/n)^{2/3})$, where n is the number of accumulated experience at hand, and δ is a small positive number.

2. The Testing Problem

Let X be a random variable having a distribution with pdf

$$f(x|\theta) = \frac{a(x)}{A(\theta)} I_{(0,\theta)}(x), \tag{2.1}$$

where $a(x) > 0$ for $x > 0$ and $A(\theta) = \int_0^\theta a(x)dx < \infty$ for all $\theta > 0$. We consider the problem of testing $H_0: \theta > \theta_0$ against $H_1: \theta \leq \theta_0$ with the $0 - L_i$ loss:

$$L(\theta, i) = (1-i)L_0 I_{(0,\theta_0]}(\theta) + iL_1 I_{(\theta_0,\infty)}(\theta), \tag{2.2}$$

where θ_0 is a known positive constant, and i denotes the action in favor of the hypothesis H_i, $i = 0, 1$ and $L_i > 0$, $i = 0, 1$. It is assumed that the parameter θ is a realization of a random variable Θ having an unknown prior distribution G over $(0, \infty)$.

A decision procedure d is defined to be a mapping from $\mathcal{X}$, the sample space of the random variable X, into the interval $[0, 1]$ such that $d(x)$ is the probability of taking action 0 when $X = x$ is observed. Let D be the class of all decision procedures. For each $d \in D$, let $r(G, d)$ denote the associated Bayes risk. Then, $r(G) \equiv \inf_{d \in D} r(G, d)$ is the minimum Bayes risk among all decision procedures in D. A decision procedure, say d_G, such that $r(G, d_G) = r(G)$ is called a Bayes procedure.

Based on the precedingly described statistical model, simple algebraic computation yields that for each $d \in D$,

$$r(G, d) = \int_{x=0}^{\infty} d(x)[L_0 G(\theta_0|x) - L_1(1 - G(\theta_0|x))]f(x)dx + C, \tag{2.3}$$

where

$$f(x) = \int_{\theta=x}^{\infty} f(x|\theta)dG(\theta) = \int_{\theta=x}^{\infty} \frac{a(x)}{A(\theta)} dG(\theta) = a(x)\psi(x),$$

$$\psi(x) = \int_{\theta=x}^{\infty} \frac{1}{A(\theta)} dG(\theta),$$

$G(\theta|x)$ is the posterior distribution of Θ given $X = x$, and

$C = L_1[1 - G(\theta_0)]$.

Straightforward computation yields

$$G(\theta|x) = \begin{cases} 0 & \text{if } \theta \leq x; \\ 1 - \frac{\psi(\theta)}{\psi(x)} & \text{if } \theta > x. \end{cases} \tag{2.4}$$

Hence,

$$r(G,d) = \int_{x=0}^{\theta_0} d(x)Q(x)a(x)dx - \int_{x=\theta_0}^{\infty} d(x)L_1 f(x)dx + C, \tag{2.5}$$

where

$$Q(x) = L_0\psi(x) - (L_0 + L_1)\psi(\theta_0). \tag{2.6}$$

It is clear that the Bayes procedure d_G can be obtained as follows:

$$d_G(X) = \begin{cases} 1 & \text{if } (X \geq \theta_0) \text{ or } (X < \theta_0 \text{ and } Q(X) \leq 0), \\ 0 & \text{otherwise.} \end{cases} \tag{2.7}$$

An Alternative Form of the Bayes Procedure d_G

Let $S = \{0 < x \leq \theta_0 | Q(x) < 0\}$ and $T = \{0 < x \leq \theta_0 | Q(x) > 0\}$. Define

$$s_0 = \begin{cases} \inf S & \text{if } S \neq \phi, \\ \theta_0 & \text{if } S = \phi; \end{cases} \tag{2.8}$$

$$t_0 = \begin{cases} \sup T & \text{if } T \neq \phi, \\ 0 & \text{if } T = \phi. \end{cases} \tag{2.9}$$

Since $\psi(x)$ is nonincreasing in x for all $x > 0$, by (2.7), $Q(x)$ is nonincreasing in x. By the definitions of s_0 and t_0 and the nonincreasing property of $Q(x)$, $t_0 \leq s_0$. We see that the Bayes procedure d_G can be presented in the following form:

$$d_G(X) = \begin{cases} 1 & \text{if } X \geq s_0; \\ \text{any} & \text{if } t_0 < X < s_0; \\ 0 & \text{if } X \leq t_0. \end{cases} \tag{2.10}$$

Note that as $t_0 < s_0$, by the definitions of s_0 and t_0, $Q(x) = 0$ for $x \in (t_0, s_0)$. In such a situation, no matter what action is taken, it does not affect the Bayes risk.

When the prior distribution G is unknown, it is not possible to apply the Bayes procedure d_G for the decision problem at hand. Suppose that the same decision problem repeats independently and certain past experience is available. In such a situation, the empirical Bayes approach can be applied.

Note that the class of densities $\{f(x|\theta)|\theta > 0\}$ has monotone likelihood ratio in x. For the $0 - L_i$ loss, all monotone decision procedures form an essentially complete class; see

Berger (1985). Therefore, we seek a monotone empirical Bayes procedure for the testing problem under study.

3. A Monotone Empirical Bayes Testing Procedure

For the empirical Bayes framework, for each $j = 1, 2, \ldots$, let (X_j, Θ_j) be a pair of random variables incurred at stage j, where X_j is observable but Θ_j is not observable. Conditional on $\Theta_j = \theta_j$, X_j has a pdf $f(x|\theta_j)$. It is assumed that (X_j, Θ_j), $j = 1, 2, \ldots$, are mutually independent and $\Theta_1, \Theta_2, \ldots$ are iid having the unknown prior distribution G. Therefore, $X_1, X_2, \ldots$ are iid with pdf $f(x)$. Suppose that we are now at stage $n+1$. So, X_{n+1} is the present random observation and $X_1, \ldots, X_n$ are the n past observations. An empirical Bayes procedure is a decision procedure for testing $H_0\colon \theta_{n+1} > \theta_0$ against $H_1\colon \theta_{n+1} \le \theta_0$ with the error loss (2.2), which can be viewed as a function of the present observation $X_{n+1} = x$ and the n past observations $X_1, \ldots, X_n$, and is denoted by $d_n(x) \equiv d_n(x; X_1, \ldots, X_n)$.

Let $Y_i = A(X_i)$, $i = 1, \ldots, n+1$. It can be shown that conditional on $\Theta_i = \theta_i$, Y_i is uniformly distributed over the interval $(0, A(\theta_i))$. Therefore, Y_i has a marginal pdf, say $h(y)$,

$$h(y) = \int_{\theta=A^{-1}(y)}^{\infty} \frac{1}{A(\theta)} dG(\theta), y > 0. \tag{3.1}$$

Hence,

$$h(A(x)) = \int_{\theta=x}^{\infty} \frac{1}{A(\theta)} dG(\theta) = \psi(x), x > 0. \tag{3.2}$$

Let $H_n(y) = \frac{1}{n} \sum_{j=0}^{n} I_{(0,y]}(Y_j)$ be the empirical distribution based on $Y_1, \ldots, Y_n$. Let $\{b_n\}$ be a sequence of positive numbers such that $\lim_{n\to\infty} b_n = 0$ and $\lim_{n\to\infty} nb_n = \infty$. For each $x > 0$ and each n, define

$$\psi_n(x) = [H_n(A(x) + b_n) - H_n(A(x))]/b_n \tag{3.3}$$

and

$$Q_n(x) = L_o\psi_n(x) - (L_0 + L_1)\psi_n(\theta_0). \tag{3.4}$$

Based on the form (2.7) of the Bayes procedure d_G, we may consider an empirical Bayes procedure d_n° defined as follows:

$$d_n^\circ(X_{n+1}) = \begin{cases} 1 & \text{if } (X_{n+1} \geq \theta_0) \text{ or } (X_{n+1} < \theta_0 \text{ and } Q_n(X_{n+1}) \leq 0); \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

However, since $Q_n(x)$ may not be nonincreasing in x, the empirical Bayes procedure d_n° is not a monotone decision procedure. In the following, we consider a monotonized version of the empirical Bayes procedure d_n°.

Let $A_n = \{0 < x < \theta_0 | Q_n(x) > 0\}$. Also let $a_n = \int_{A_n} a(x)dx$. Note that $A_n \subset (0, \theta_0)$. Therefore, $a_n \leq \int_0^{\theta_0} a(x)dx = A(\theta_0)$. Since $A(x)$ is a strictly increasing function, we have $A^{-1}(a_n) \leq \theta_0$.

We propose an empirical Bayes procedure d_n^* given as follows:

$$d_n^*(X_{n+1}) = \begin{cases} 1 & \text{if } X_{n+1} \geq A^{-1}(a_n); \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

¿From (3.6), it can be seen that d_n^* is a monotone procedure.

Let $r(G, d_n^\circ)$ and $r(G, d_n^*)$ denote the conditional Bayes risks (conditional on the past observations $\underset{\sim}{X}_n = (X_1, \ldots, X_n)$) of the empirical Bayes procedures d_n° and d_n^*, respectively. That is, from (2.5),

$$r(G, d_n^\circ) = \int_0^{\theta_0} d_n^\circ(x)Q(x)a(x)dx - \int_{\theta_0}^{\infty} L_1 f(x)dx + C, \tag{3.7}$$

and

$$r(G, d_n^*) = \int_0^{\theta_0} d_n^*(x)Q(x)a(x)dx - \int_{\theta_0}^{\infty} L_1 f(x)dx + C. \tag{3.8}$$

The following theorem provides the superiority of d_n^* to d_n°.

Theorem 3.1. The monotone empirical Bayes procedure d_n^* dominates the empirical Bayes procedure d_n° in the sense that $r(G, d_n^*) \leq r(G, d_n^\circ)$ for all $\underset{\sim}{X}_n$ for all n.

Proof: Let $C_n = (0, \theta_0) \backslash A_n$, the complement of A_n relative to $(0, \theta_0)$. Also, let $A_n^* = (0, A^{-1}(a_n))$ and $C_n^* = [A^{-1}(a_n), \theta_0)$. Hence, $C_n \cup A_n = C_n^* \cup A_n^* = (0, \theta_0)$. By the definitions of d_n° and d_n^*, for $0 < x < \theta_0$,

$$d_n^\circ(x) = \begin{cases} 0 & \text{if } x \in A_n; \\ 1 & \text{if } x \in C_n; \end{cases} \tag{3.9}$$

and

$$d_n^*(x) = \begin{cases} 0 & \text{if } x \in A_n^*; \\ 1 & \text{if } x \in C_n^*. \end{cases} \tag{3.10}$$

Therefore, from (3.7)–(3.10),

$$\begin{aligned} & r(G, d_n^\circ) - r(G, d_n^*) \\ &= \int_0^{\theta_0} d_n^\circ(x)Q(x)a(x)dx - \int_0^{\theta_0} d_n^*(x)Q(x)a(x)dx \\ &= \int_{C_n} Q(x)a(x)dx - \int_{C_n^*} Q(x)a(x)dx \\ &= [\int_{C_n \cap C_n^*} Q(x)a(x)dx + \int_{C_n \cap A_n^*} Q(x)a(x)dx] \\ &\quad - [\int_{C_n^* \cap C_n} Q(x)a(x)dx + \int_{C_n^* \cap A_n} Q(x)a(x)dx] \\ &= \int_{C_n \cap A_n^*} Q(x)a(x)dx - \int_{C_n^* \cap A_n} Q(x)a(x)dx. \end{aligned} \tag{3.11}$$

Note that for each $x_1 \in C_n \cap A_n^*$ and each $x_2 \in C_n^* \cap A_n$, by the definitions of a_n, A_n^* and C_n^*, we have $x_1 \leq A^{-1}(a_n) \leq x_2$. Since $Q(x)$ is nonincreasing in x, therefore, $Q(x_1) \geq Q(A^{-1}(a_n)) \geq Q(x_2)$. Let $q_1 = \inf_{x \in C_n \cap A_n^*} Q(x)$ and $q_2 = \sup_{x \in C_n^* \cap A_n} Q(x)$. We have

$$q_1 \geq q_2. \tag{3.12}$$

By the definition of a_n and A_n^*,

$$\int_{A_n} a(x)dx = a_n = \int_0^{A^{-1}(a_n)} a(x)dx = \int_{A_n^*} a(x)dx.$$

Hence

$$\int_{A_n \cap A_n^*} a(x)dx + \int_{A_n \cap C_n^*} a(x)dx = \int_{A_n^* \cap A_n} a(x)dx + \int_{A_n^* \cap C_n} a(x)dx$$

which implies that

$$\int_{A_n \cap C_n^*} a(x)dx = \int_{A_n^* \cap C_n} a(x)dx. \tag{3.13}$$

Now combining (3.11)–(3.13) and by the definitions of q_1 and q_2, we can obtain

$$\begin{aligned}
&r(G, d_n^\circ) - r(G, d_n^*) \\
&= \int_{C_n \cap A_n^*} Q(x)a(x)dx - \int_{C_n^* \cap A_n} Q(x)a(x)dx \\
&\geq q_1 \int_{C_n \cap A_n^*} a(x)dx - q_2 \int_{C_n^* \cap A_n} a(x)dx \\
&= (q_1 - q_2) \int_{C_n \cap A_n^*} a(x)dx \geq 0.
\end{aligned} \tag{3.14}$$

Hence the proof of the theorem is completed. □

It should be noted that the idea of monotonizing the decision procedure in a monotone decision problem so that to improve the performance of the decision procedure is known in the literature; for example, see Berger (1985).

4. Asymptotic Optimality of d_n^*

For an empirical Bayes procedure d_n, let $r(G, d_n)$ denote the associated conditional Bayes risk (conditional on the past observations $\underset{\sim}{X}_n$) and $E[r(G, d_n)]$ the associated overall Bayes risk. That is,

$$r(G, d_n) = \int_{x=0}^{\infty} d_n(x)[L_0 G(\theta_0|x) - L_1(1 - G(\theta_0|x))]f(x)dx + C,$$

and

$$E[r(G, d_n)] = \int_{x=0}^{\infty} E[d_n(x)][L_0 G(\theta_0|x) - L_1(1 - G(\theta_0|x))]f(x)dx + C,$$

where the expectation $E[d_n(x)]$ is taken with respect to $\underset{\sim}{X}_n$. Since $r(G)$ is the minimum Bayes risk, $r(G, d_n) - r(G) \geq 0$ for all $\underset{\sim}{X}_n$ and for all n. Therefore, $E[r(G, d_n)] - r(G) \geq 0$ for all n. The nonnegative regret risks $r(G, d_n) - r(G)$ and $E[r(G, d_n)] - r(G)$ can be used as measures of the performance of the empirical Bayes procedure d_n. In the following, we are concerned only with the difference $E[r(G, d_n)] - r(G)$.

Definition 4.1. A sequence of empirical Bayes procedures $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal of order $\{\alpha_n\}_{n=1}^{\infty}$ in E relative to the prior distribution G if $E[r(G, d_n)] - r(G) = O(\alpha_n)$ where $\{\alpha_n\}_{n=1}^{\infty}$ is a sequence of positive numbers such that $\alpha_n = o(1)$.

In the following, we investigate the asymptotic optimality of the sequence of the empirical Bayes procedures $\{d_n^*\}$. Without loss of generality, we may assume that $0 < t_0$, $s_0 < \theta_0$. It is also assumed that the prior distribution G satisfies the following conditions:

Condition A. There exist positive constants ε, m and M where $\varepsilon < \min(t_0, \theta_0 - s_0)$ and $m \le M$ such that

[A1]
$$m(t_0 - x) \le \int_x^{t_0} \frac{1}{A(\theta)} dG(\theta) \text{ for } x \in [t_0 - \varepsilon, t_0],$$
$$m(y - s_0) \le \int_{s_0}^{y} \frac{1}{A(\theta)} dG(\theta) \text{ for } y \in [s_0, s_0 + \varepsilon];$$

[A2]
$$\int_y^x \frac{1}{A(\theta)} dG(\theta) \le M(x - y) \text{ for } 0 < y \le x < \infty.$$

It should be noted that the constants ε, m and M depend on the prior distribution G. Also, under Condition A, $Q(t_0) = Q(s_0) = 0$.

We also assume that the function $a(x)$ has the following properties.

[P1] There is a positive constant k^* such that $a(x) \le k^*$ for all $0 < x < \theta_0$.

[P2] There exists a positive constant k such that for each x in a neighborhood of θ_0 or in a neighborhood of t_o and for $b > 0$ being very small, $A(x) + b \le A(x + kb)$.

Though the requirement of the function $a(x)$ to possess the properties [P1] and [P2] is a restriction, the class of pdfs $f(x|\theta)$ under consideration is still broad enough to cover many plausible probability distributions.

The sequence of the empirical Bayes procedures $\{d_n^*\}$ has the following asymptotic optimality.

Theorem 4.1. Suppose the prior distribution G satisfies Conditions [A1] and [A2] and the function $a(x)$ possesses the properties [P1] and [P2].

Then

$$E[r(G, d_n^*)] - r(G) = O(b_n^2) + O(\exp(-\tau n b_n^3)),$$

where $\tau = k^2 m^2 L_0^2 / \{8[MA^{-1}(\theta_0) + \theta_0^{-1}](L_0 + L_1)^2\}$. Hence letting $b_n = [(\ell n\ n)^{1+\delta}/n]^{1/3}$, where $\delta > 0$, we have

$$E[r(G, d_n^*)] - r(G) = O(((\ell n\ n)^{1+\delta}/n)^{2/3}).$$

To present a concise proof of the theorem, we introduce several useful lemmas. Since $\lim_{n\to\infty} b_n = 0$, in the following, we only consider the case where b_n is sufficiently small such that $\frac{4M(L_0+L_1)b_n}{L_0 m}\max(1, k) < \varepsilon$ where ε and k are the constants described in [A1] and [P2], respectively.

Lemma 4.1. Let $0 < b \le b_n$ and the function $a(x)$ possesses the property [P2]. Then,

(a) For each $0 < x < t_0 - 2kb_n$, $A(x) + b \le A(t_0 - kb_n)$.

(b) $A(\theta_0) + b \le A(\theta_0 + kb)$.

Proof: (a) By the increasing property of $A(\cdot)$, for $0 < x < t_0 - 2kb_n$,

$$\begin{aligned} A(x) + b &\le A(t_0 - 2kb_n) + b \\ &\le A(t_0 - 2kb_n + kb) \qquad \text{(by property [P2])} \\ &\le A(t_0 - kb_n). \qquad \text{(since } b \le b_n\text{)} \end{aligned}$$

(b) By taking $x = \theta_0$, then the result follows from [P2]. □

Lemma 4.2. Suppose that Conditions [A1] and [A2] hold. Then,

(a) $imkL_0b_n \le Q(t_0 - ikb_n) \le i\ MkL_0b_n$ for $i = 1, 2$;

(b) $-\frac{2M^2(L_0+L_1)kb_n}{m} \le Q(s_0 + \frac{2M(L_0+L_1)kb_n}{L_0 m}) \le -2Mk(L_0 + L_1)b_n$.

Proof: (a) Since $Q(t_0) = 0$,

$$\begin{aligned} Q(t_0 - ikb_n) &= Q(t_0 - ikb_n) - Q(t_0) \\ &= L_0 \int_{t_0 - ikb_n}^{t_0} \frac{1}{A(\theta)} dG(\theta). \end{aligned}$$

Under [A1], $imkb_n \leq \int_{t_0-ikb_n}^{t_0} \frac{1}{A(\theta)} dG(\theta)$ and under [A2], $\int_{t_0-ikb_n}^{t_0} \frac{1}{A(\theta)} dG(\theta) \leq i\ Mkb_n$. Hence the result follows directly.

(b) By noting that $Q(s_0) = 0$,

$$Q(s_0 + \frac{2M(L_0 + L_1)kb_n}{L_0 m}) = -L_0 \int_{s_0}^{s_0 + \frac{2M(L_0+L_1)kb_n}{L_0 m}} \frac{1}{A(\theta)} dG(\theta).$$

Under [A1], $\int_{s_0}^{s_0+\frac{2M(L_0+L_1)kb_n}{L_0 m}} \frac{1}{A(\theta)} dG(\theta) \geq \frac{2M(L_0+L_1)kb_n}{L_0}$, and under [A2] $\int_{s_0}^{s_0+\frac{2M(L_0+L_1)kb_n}{L_0 m}} \frac{1}{A(\theta)} dG(\theta) \leq \frac{2M^2(L_0+L_1)kb_n}{L_0 m}$. Hence the result follows. □

Let $H(\cdot)$ denote the distribution function of the random variable $Y_i = A(X_i)$. For each $x > 0$, define

$$Q_1(x) = L_0[H(A(x) + b_n) - H(A(x))] - (L_0 + L_1)[H(A(\theta_0) + b_n) - H(A(\theta_0))].$$

Lemma 4.3. Suppose that the function $a(x)$ possesses the property [P2] and the Conditions [A1] and [A2] hold. Then,

(a) for $0 < x < t_0 - 2kb_n$, $Q_1(x) \geq mkL_0 b_n^2$;

(b) for $s_0 + \frac{2M(L_0+L_1)kb_n}{mL_0} < x < \theta_0$, $Q_1(x) \leq -Mk(L_0 + L_1)b_n^2$.

Proof: (a) Since $Q(t_0) = 0$ and $\psi(x) = h(A(x))$, $L_o h(A(t_0)) = (L_0 + L_1)h(A(\theta_0))$. For each $x \in (0, t_0 - 2kb_n)$, by the nonincreasing property of the pdf $h(\cdot)$

$$\begin{aligned}
Q_1(x) &\geq b_n[L_0 h(A(x) + b_n) - (L_0 + L_1)h(A(\theta_0))] \\
&= b_n L_0[h(A(x) + b_n) - h(A(t_0))] \\
&\geq b_n L_0[h(A(t_0 - kb_n)) - h(A(t_0))] \\
&= b_n L_0 \int_{t_0-kb_n}^{t_0} \frac{1}{A(\theta)} dG(\theta) \\
&\geq b_n^2 L_0 mk.
\end{aligned}$$

(b) By the nonincreasing property of the pdf $h(\cdot)$,

$$H(A(x) + b_n) - H(A(x)) \leq b_n h(A(x))$$

and

$$H(A(\theta_0)+b_n)-H(A(\theta_0)) \geq b_n h(A(\theta_0)+b_n).$$

Hence,

$$\begin{aligned}
Q_1(x)-b_nQ(x) &\leq b_n(L_0+L_1)[h(A(\theta_0))-(A(\theta_0)+b_n)] \\
&\leq b_n(L_0+L_1)[h(A(\theta_0))-h(A(\theta_0+kb_n))] \\
&= b_n(L_0+L_1)\int_{\theta_0}^{\theta_0+kb_n}\frac{1}{A(\theta)}dG(\theta) \\
&\leq kM(L_0+L_1)b_n^2.
\end{aligned}$$

Then, from Lemma 4.2.b,

$$\begin{aligned}
Q_1(x) &\leq b_nQ(x)+kM(L_0+L_1)b_n^2 \\
&\leq -2kM(L_0+L_1)b_n^2+kM(L_0+L_1)b_n^2 \\
&= -kM(L_0+L_1)b_n^2. \qquad \square
\end{aligned}$$

For each $0<x<\theta_0$ and $j=1,2,\dots,n$, define

$$\begin{aligned}
T_{nj}(x) &= L_0\{I_{(A(x),A(x)+b_n]}(Y_j)-[H(A(x)+b_n)-H(A(x))\} \\
&\quad -(L_0+L_1)\{I_{(A(\theta_0),A(\theta_0)+b_n]}(Y_j)-[H(A(\theta_0)+b_n)-H(A(\theta_0))]\},
\end{aligned}$$

and let $T_n(x)=\frac{1}{n}\sum_{j=1}^{n}T_{nj}(x)$. Then $T_{nj}(x)$, $j=1,\dots,n$, are iid such that

$$|T_{nj}(x)| \leq 2L_0+L_1 \equiv T_0, \;\; E[T_{nj}(x)]=0.$$

Also,

$$\begin{aligned}
\text{Var }(T_{nj}(x)) &\leq L_0^2[H(A(x)+b_n)-H(A(x))]+(L_0+L_1)^2[H(A(\theta_0)+b_n)-H(A(\theta_0))] \\
&\leq 2(L_0+L_1)^2[H(A(x)+b_n)-H(A(x))] \qquad (4.1) \\
&\leq 2(L_0+L_1)^2h(0)b_n,
\end{aligned}$$

where the second and third inequalities in (4.1) are obtained due to the nonincreasing property of the pdf $h(\cdot)$.

Note that under Condition [A2],

$$
\begin{aligned}
h(0) &= \int_{\theta=0}^{A^{-1}(\theta_0)} \frac{1}{A(\theta)} dG(\theta) + \int_{A^{-1}(\theta_0)}^{\infty} \frac{1}{A(\theta)} dG(\theta) \\
&\le MA^{-1}(\theta_0) + h(\theta_0) \qquad (4.2) \\
&\le MA^{-1}(\theta_0) + \theta_0^{-1} < \infty.
\end{aligned}
$$

In (4.2), we use the inequality: $h(y) \le y^{-1}$ for all $y > 0$, which is guaranteed by the fact that $h(y)$ is a nonincreasing pdf.

Therefore, for each integer $s \ge 2$,

$$
\begin{aligned}
E[T_{nj}^s(x)] &\le E|T_{nj}^s(x)| \\
&\le T_0^{s-2} E|T_{nj}^2(x)| \\
&\le \frac{1}{2} T_0^{s-2} s! \text{ Var } (T_{nj}(x)).
\end{aligned}
$$

Then, by Bernstein's inequality (see page 169 of Ibragimov and Linnik (1971)), we have

Lemma 4.4. Under Condition [A2], for $0 < t \le \sqrt{n \text{ Var } (T_{n1}(x))}/(2T_0)$,

(a) $P\{T_n(x) \ge 2t\sqrt{\text{Var } (T_{n1}(x))/n}\} \le \exp(-t^2)$;

(b) $P\{T_n(x) \le -2t\sqrt{\text{Var } (T_{n1}(x))/n}\} \le \exp(-t^2)$.

Lemma 4.5. Suppose the function $a(x)$ possesses the property [P2] and the prior distribution G satisfies the Conditions [A1] and [A2]. Then for n being sufficiently large,

(a) for each $x \in (0, t_0 - 2kb_n)$,

$$P\{d_n^\circ(x) = 1\} \le \exp\{-m^2 L_0^2 k^2 n b_n^3 / \{8(L_0 + L_1)^2 [MA^{-1}(\theta_0) + \theta_0^{-1}]\}\};$$

(b) for each $x \in (s_0 + \frac{2Mk(L_0+L_1)b_n}{L_0 m}, \theta_0)$,

$$P\{d_n^\circ(x) = 0\} \le \exp\{-M^2 k^2 n b_n^3 / \{8[MA^{-1}(\theta_0) + \theta_0^{-1}]\}\}.$$

Proof: (a) For $x \in (0, t_0 - 2kb_n)$, by Lemma 4.3a, $Q_1(x) \ge mkL_0 b_n^2 \equiv c_n$, where $c_n \le \text{ Var } (T_{n1}(x))T_0^{-1}$ for n being sufficiently large. By the definition of d_n°, Lemma 4.4, (4.1) and (4.2), a straightforward computation leads to:

$$\begin{aligned}P\{d_n^\circ(x)=1\} &= P\{T_n(x)\le -Q_1(x)\}\\ &\le P\{T_n(x)\le -c_n\}\\ &\le \exp\{-nc_n^2/[4\ \mathrm{Var}\ (T_{n1}(x))]\}\\ &\le \exp\{-nc_n^2/\{8(L_0+L_1)^2[MA^{-1}(\theta_0)+\theta_0^{-1}]b_n\}\}\\ &= \exp\{-m^2k^2L_0^2nb_n^3/\{8(L_0+L_1)^2[MA^{-1}(\theta_0)+\theta_0^{-1}]\}\}.\end{aligned}$$

Part (b) can be proved similarly. Hence the detail is omitted. □

Proof of Theorem 4.1: By theorem 3.1, $E[r(G,d_n^*)]\le E[r(G,d_n^\circ)]$ for all n. Hence, it suffices to consider $E[r(G,d_n^\circ)]-r(G)$. Note that

$$0\le E[r(G,d_n^\circ)]-r(G)=B_1+B_2+B_3+B_4, \tag{4.3}$$

where

$$\begin{aligned}B_1 &= \int_0^{t_0-2kb_n} Q(x)a(x)P\{d_n^\circ(x)=1\}dx,\\ B_2 &= \int_{t_0-2kb_n}^{t_0} Q(x)a(x)P\{d_n^\circ(x)=1\}dx,\\ B_3 &= \int_{s_0}^{s_0+\frac{2kM(L_0+L_1)b_n}{L_0m}} [-Q(x)]a(x)P\{d_n^\circ(x)=0\}dx,\end{aligned}$$

and

$$B_4=\int_{s_0+\frac{2kM(L_0+L_1)b_n}{L_0m}}^{\theta_0} [-Q(x)]a(x)P\{d_n^\circ(x)=0\}dx.$$

By the nonincreasing property of $Q(\cdot)$ and Lemma 4.2.a and [P1],

$$\begin{aligned}0\le B_2 &\le \int_{t_0-2kb_n}^{t_0} Q(t_0-2kb_n)a(x)dx\\ &\le 2MkL_0b_n\int_{t_0-2kb_n}^{t_0} a(x)dx\\ &\le 4Mk^2k^*L_0b_n^2\\ &= O(b_n^2).\end{aligned} \tag{4.4}$$

Similarly,

$$
\begin{aligned}
0 \le B_3 &\le \int_{s_0}^{s_0+\frac{2kM(L_0+L_1)b_n}{L_0 m}} [-Q(s_0 + \frac{2kM(L_0+L_1)b_n}{L_o m})a(x)dx \\
&\le \frac{2M^2(L_0+L_1)kb_n}{m} \int_{s_0}^{s_0+\frac{2kM(L_0+L_1)b_n}{L_0 m}} a(x)dx) \\
&\le \frac{4M^3k^2k^*(L_0+L_1)^2 b_n^2}{m^2 L_0} \\
&= O(b_n^2).
\end{aligned} \tag{4.5}
$$

It suffices to consider B_1 and B_4.

By Lemma 4.5.a,

$$
\begin{aligned}
0 \le B_1 &= \int_0^{t_0-2kb_n} Q(x)a(x)P\{d_n^\circ(x)=1\}dx \\
&\le \exp\{-\tau n b_n^3\} \int_0^{t_0-2kb_n} a(x)Q(x)dx \\
&\le L_0 \exp\{-\tau n b_n^3\}, \\
&= O(\exp(-\tau n b_n^3)).
\end{aligned} \tag{4.6}
$$

and by Lemma 4.5.b,

$$
\begin{aligned}
0 \le B_4 &\le \exp\{-\tau n b_n^3\} \int_{s_0+\frac{2kM(L_0+L_1)b_n}{L_0 m}}^{\theta_0} [-Q(x)]a(x)dx \\
&\le \exp\{-\tau n b_n^3\}(L_0+L_1)h(A(\theta_0))A(\theta_0) \\
&= 0(\exp(-\tau n b_n^3)).
\end{aligned} \tag{4.7}
$$

Combining (4.3)–(4.7) yields

$$E[r(G, d_n^\circ)] - r(G) = O(b_n^2) + O(\exp(-\tau n b_n^3)).$$

Since $\tau(\ell n\ n)^\delta \to \infty$ as $n \to \infty$, by letting $b_n = [(\ell n\ n)^{1+\delta}/n]^{1/3}$, we have,

$$E[r(G, d_n^\circ)] - r(G) = O(((\ell n)^{1+\delta}/n)^{2/3}).$$

Hence the proof of the theorem is completed.

References

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Second edition, Springer–Verlag, New York.

Datta, S. (1991). Nonparametric empirical Bayes estimation with $O(n^{-1/2})$ rate of a truncation parameter. *Statistics & Decisions* **9**, 45–61.

Fox, R. (1970). Estimating the empiric distribution function of certain parameter sequences. *Ann. Math. Statist.* **41**, 1845–1852.

Fox, R. (1978). Solutions to empirical Bayes squared error loss estimation problems. *Ann. Statist.* **6**, 846–853.

Gupta, S. S. and Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. *J. Statist. Plann. Inference* **8**, 87–101.

Huang, W. T. and Liang, T. (1990). Selecting the best uniform distribution: An empirical Bayes approach. Unpublished Technical Report.

Ibragimov, I. A. and Linnik, Yu. V. (1971). *Independent and Stationary Sequence of Random Variables.* Wolters-Noordhoff, Groningen.

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.

Liang, T. (1990). On the convergence rates of a monotone empirical Bayes test for uniform distributions. *J. Statist. Plann. Inference* **26**, 25–34.

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 $\mathcal{U}(0, \theta)$ distribution. *Ann. Statist.* **16**, 1335–1341.

Prasad, B. and Singh, R. S. (1990). Estimation of prior distribution and empirical Bayes estimation in a nonexponential family. *J. Statist. Plann. Inference* **24**, 81–86.

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

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

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

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

Stijnen, T. (1982). A monotone empirical Bayes estimation and test for the one–parameter continuous exponential family based on spacings. *Scand. J. Statist.* **9**, 153–158.

Van Houwelingen, J. C. (1976). Monotone empirical Bayes tests for the continuous one–parameter exponential family. *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.

Wei, L. (1983). Empirical Bayes estimation of parameter with convergence rates about uniform distribution families $\mathcal{U}(0, \theta)$. *Acta Math. Appl. Sinica* **6**, 485–493.

Wei, L. (1985). Empirical Bayes estimation of location parameter with convergence rates in one–sided truncation distribution families. *Chinese Ann. Math.* **6A**, 193–202.