EMPIRICAL BAYES TEST FOR DEPENDENT DATA: DISCRETE CASE

by

TaChen Liang
Wayne State University

Technical Report #95-25C

Department of Statistics
Purdue University

May 1995

EMPIRICAL BAYES TEST FOR DEPENDENT DATA:
DISCRETE CASE*

TaChen Liang

Department of Mathematics
Wayne State University
Detroit, MI 48202

Abstract

We study a sequence of empirical Bayes tests for the two-action problem in a discrete exponential family with dependent observations $\{X_i, i = 1, 2, \ldots\}$, which is assumed to be a stationary process. Three cases of dependence are considered: (1) $\{X_i, i = 1, 2, \ldots\}$ is m-dependent; (2) $\{X_i, i = 1, 2, \ldots\}$ is a strictly stationary φ-mixing process; and (3) $\{X_i, i = 1, 2, \ldots\}$ is a strictly stationary α-mixing process. In each case, the asymptotic optimality of the empirical Bayes tests is investigated and the corresponding rate of convergence of the regret risks is established. The rates of convergence have orders of exponential type of the form $O(\exp(-c_i\psi_i(n)))$, where $c_i > 0$, $\psi_i(n) > 0$ and $\lim_{n\to\infty} \psi_i(n) = \infty$, depending on cases of dependence, and n is the number of past data at hand for a current testing problem.

AMS 1991 Subject Classification: 62C12

Keywords and phrases: asymptotic optimality, empirical Bayes, α-mixing, φ-mixing, m-dependent, rate of convergence, two-action problem.

* This research was supported in part by US Army Research Office, Grant DAAH04-95-1-0165 under the direction of Professor S. S. Gupta

1. Introduction

Let X denote a random variable arising from a discrete exponential family with probability function

$$f(x|\theta) = a(x)\beta(\theta)\theta^x, \quad x = 0, 1, 2, \ldots; \quad 0 < \theta < Q; \tag{1.1}$$

where $a(x) > 0$ for all $x = 0, 1, \ldots$, and Q may be finite or infinite, and is unknown. Consider the problem of testing H_0: $\theta > \theta_0$ versus H_1: $\theta \le \theta_0$ with respect to a control θ_0, a known positive constant. Let i, $i = 0, 1$, denote an action deciding in favor of H_i. For the parameter θ and action i, the loss function is defined to be:

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

where 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 $(0, Q)$.

Let $\mathcal{X}$ be the sample space of X. A test d is defined to be a mapping from $\mathcal{X}$ into $[0, 1]$, such that $d(x)$ is the probability of taking action 0 when $X = x$ is observed. That is, $d(x) = P\{\text{accepting } H_0 | X = x\}$. Let $r(G, d)$ denote the Bayes risk associated with the test d. Then,

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

where

$$\tau_G(x) = E[\Theta|X = x] = \frac{h(x+1)}{h(x)}: \text{ the posterior mean of } \Theta \text{ given } X = x,$$

$$f_G(x) = \int_0^Q f(x|\theta)dG(\theta) = a(x)h(x): \text{ the marginal probability function of } X,$$

$$h(x) = \int_0^Q \beta(\theta)\theta^x dG(\theta), \text{ and } C = \int_{\theta_0}^Q (\theta - \theta_0)dG(\theta).$$

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 test problem is meaningful. This assumption always holds when Q is finite. For example, in a negative binomial distribution, $f(x|\theta) = \binom{r+x-1}{x}\theta^x(1-\theta)^r$, $0 < \theta < 1$ and r is a positive integer. In such a case, $Q \le 1$.

From (1.3), a Bayes test, say d_G, is clearly given by

$$d_G(x) = \begin{cases} 1 & \text{if } \tau_G(x) > \theta_0, \\ 0 & \text{otherwise.} \end{cases} \tag{1.4}$$

The minimum of Bayes risks among the class of all tests is:

$$r(G) \equiv r(G, d_G) = \sum_{x=0}^{\infty} [\theta_0 - \tau_G(x)] d_G(x) f_G(x) + C. \tag{1.5}$$

When the prior distribution G is unknown, the two-action problem has been studied by Johns and Van Ryzin (1971) and Liang (1988), respectively, along the line of standard empirical Bayes approach of Robbins (1956, 1964). The standard empirical Bayes statistical framework for the concerned two-action problem is given as follows. Let (X_i, Θ_i), $i = 1, 2, \ldots$, be iid with (X, Θ), where X_i, $i = 1, 2, \ldots$, are observable, but Θ_i, $i = 1, 2, \ldots$, are not observable. At time $n+1$, $\underset{\sim}{X}_n = (X_1, \ldots, X_n)$ denotes the past data, X_{n+1} denotes the present random observation, and one is interested in testing $H_0\colon \theta_{n+1} > \theta_0$ versus $H_1\colon \theta_{n+1} \leq \theta_0$, with the loss (1.2), where θ_{n+1} is a realization of the random variable Θ_{n+1}. A test d_n, called as an empirical Bayes test, is a function of the present observation $X_{n+1} = x$ and the past data $\underset{\sim}{X}_n$, such that $d_n(x; \underset{\sim}{X}_n) \equiv d_n(x)$ is the probability of accepting H_0. Let $r(G, d_n | \underset{\sim}{X}_n)$ be the Bayes risk of the empirical Bayes test d_n conditioning on the past data $\underset{\sim}{X}_n$. Also, let $r(G, d_n) = E_n r(G, d_n | \underset{\sim}{X}_n)$ denote the overall Bayes risk of the empirical Bayes test d_n, where the expectation E_n is taken with respect to the probability measure generated by $\underset{\sim}{X}_n$. A sequence of empirical Bayes tests $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal relative to the prior distribution G, with rate of convergence of order $O(\alpha_n)$ if the regret risk $r(G, d_n) - r(G)$ is such that $r(G, d_n) - r(G) = O(\alpha_n)$, where $\{\alpha_n\}$ is a sequence of positive numbers satisfying $\lim_{n \to \infty} \alpha_n = 0$.

The empirical Bayes tests of Johns and Van Ryzin (1971) and Liang (1988) are along the line of the standard empirical Bayes approach. Extensions of the preceedingly described empirical Bayes statistical framework have been studied by several authors. O'Bryan (1976) studies an empirical Bayes estimation problem in which $X_1, X_2, \ldots,$ are assumed to be mutually independent, but may not be identically distributed. Karunamuni (1988) investigates some empirical Bayes sequential decision rules in which the number of observations taken at each stage is random and is determined according to a stopping rule.

In this paper, we study a sequence of empirical Bayes tests for the two-action problem based on dependent data $\{X_i, i = 1, 2, \ldots\}$ which is assumed to be a stationary process. Three cases of dependence are considered: (1) $\{(X_i, \Theta_i);\ i = 1, 2, \ldots\}$ is m-dependent; (2) $\{(X_i, \Theta_i);\ i = 1, 2, \ldots\}$ is a strictly stationary φ-mixing process; and (3) $\{(X_i, \Theta_i);\ i = 1, 2, \ldots\}$ is a strictly stationary α-mixing process. In each case, the asymptotic optimality of the empirical Bayes tests is investigated and the rate of convergence of the associated regret risks is established. The rates of convergence of the regret risks have orders of exponential type of the form $O(\exp(-c_i\psi_i(n)))$, where $c_i > 0$, $\psi_i(n) > 0$ and $\lim_{n\to\infty} \psi_i(n) = \infty$, depending on cases of independence.

At the end of this section, we briefly introduce the definitions of the two mixing processes.

(a) The process $\{(X_i, \Theta_i);\ i = 1, 2, \ldots,\}$ is a strictly stationary φ-mixing (uniformly strongly mixing) process if there exists a nonincreasing sequence of positive numbers $\{\varphi(n);\ n \geq 1\}$, with $\lim_{n\to\infty} \varphi(n) = 0$, such that for any positive integers n and t,

$$|P(A \cap B) - P(A)P(B)| \leq \varphi(n)P(A), \quad A \in \mathcal{M}_0^t,\ B \in \mathcal{M}_{t+n}^{\infty}$$

where $\mathcal{M}_u^v$ is the σ-field generated by the random vectors $\{(X_i, \Theta_i);\ u \leq i \leq v\}$.

(b) The process $\{(X_i, \Theta_i);\ i = 1, 2, \ldots,\}$ is an α-mixing process if there exists a nonincreasing sequence of positive numbers $\{\alpha(n);\ n \geq 1\}$, with $\lim_{n\to\infty} \alpha(n) = 0$, such that for any positive integers n and t,

$$|P(A \cap B) - P(A)P(B)| \leq \alpha(n),\ A \in \mathcal{M}_0^t,\ B \in \mathcal{M}_{t+n}^{\infty}.$$

2. Construction of Empirical Bayes Test

Before we go further to construct an empirical Bayes test, we recall some property related to this decision problem. Since the class of the probability function $\{f(x|\theta)|0 < \theta < Q\}$ has the monotone likelihood ratio, for the loss function (1.2), the class of monotone tests is essentially complete. Hence, it is desirable that the proposed empirical Bayes test be monotone.

We now examine the monotone behavior of the Bayes test d_G. Let

$$A = \{x | \tau_G(x) > \theta_0\} = \{x | H(x) > 0\},$$
$$B = \{x | \tau_G(x) < \theta_0\} = \{x | H(x) < 0\},$$

where $H(x) = h(x+1) - \theta_0 h(x)$. Define

$$a^* = \begin{cases} \inf A & \text{if } A \neq \phi, \\ \infty & \text{if } A = \phi; \end{cases}$$
$$b^* = \begin{cases} \sup B & \text{if } B \neq \phi, \\ -1 & \text{if } B = \phi. \end{cases}$$

Since $\tau_G(x)$ is increasing in x, $b^* \leq a^*$. Also, $y \geq a^*$ iff $\tau_G(y) > \theta_0$; $x \leq b^*$ iff $\tau_G(x) < \theta_0$. Therefore, the Bayes test d_G can be written as:

$$d_G(x) = \begin{cases} 1 & \text{if } x \geq a^*, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

By mimicking the form (2.1) of the Bayes test d_G, an empirical Bayes test is constructed as follows.

For each $x = 0, 1, 2, \ldots$, and each positive integer n, define $h_n(x) = \frac{1}{na(x)} \sum_{j=1}^{n} I_{\{x\}}(X_j)$. Note that $h_n(x)$ is an unbiased and consistent estimator of $h(x)$. Let $H_n(x) = h_n(x+1) - \theta_0 h_n(x)$. Set

$$A_n = \{x | H_n(x) > 0\}$$

and define

$$a_n = \begin{cases} \inf A_n & \text{if } A_n \neq \phi, \\ \infty & \text{if } A_n = \phi. \end{cases}$$

We then propose an empirical Bayes test d_n^*, which is defined as:

$$d_n^*(X_{n+1}) = \begin{cases} 1 & \text{if } X_{n+1} \geq a_n, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

Note that d_n^* is a monotone test. The overall Bayes risk of the test d_n^* is

$$r(G, d_n^*) = \sum_{x=0}^{\infty} [\theta_0 - \tau_G(x)] E_{n+1}[d_n^*(X_{n+1}) I_{\{x\}}(X_{n+1})] + C. \tag{2.3}$$

Since $r(G)$ is the minimum Bayes risk, $r(G, d_n^*) - r(G) \geq 0$ for all n. This nonnegative regret risk $r(G, d_n^*) - r(G)$ is used as a measure of the performance of the empirical Bayes test d_n^*. We will investigate the asymptotic optimality of d_n^*. The results are presented in the next section.

3. Asymptotic Optimality

Our main results regarding the asymptotic optimality of the empirical Bayes test d_n^* are stated in Theorems 3.1–3.3 given below.

Theorem 3.1 Assume that

(a) $\{(X_i, \Theta_i); i = 1, 2, \ldots\}$ is m-dependent;

(b) $\int_0^Q \theta dG(\theta) < \infty$; and

(c) $b^* < \infty$.

Then, $r(G, d_n^*) - r(G) = O(\exp(-c_1 n))$ for some $c_1 > 0$.

Theorem 3.2 Assume that

(a) $\{(X_i, \Theta_i); i = 1, 2, \ldots\}$ is a strictly stationary φ-mixing process with mixing coefficients $\{\varphi(i); i \geq 1\}$;

(b) $\sum_{x=0}^{\infty} \tau_G(x)[f_G(x)]^{1/t} < \infty$ for some $t > 1$; and

(c) $b^* < \infty$

Then,

(1) $r(G, d_n^*) - r(G) = O(\exp(-c_2 n/\ell(n)))$ for some $c_2 > 0$, where $\ell(n)$ is a positive integer such that $\ell(n) \leq n$ and $\lim_{n \to \infty} \ell(n) = \infty$.

(2) If one chooses $\ell(n) = [\log n]$, where $[x]$ denotes the largest integer not exceeding x, then, $r(G, d_n^*) - r(G) = O(\exp(-c_2 n/\log n))$.

Theorem 3.3 Assume that

(a) $\{(X_i, \Theta_i), i = 1, 2, \ldots\}$ is a strictly stationary α-mixing process with mixing coefficients $\{\alpha(i); i = 1, 2, \ldots\}$;

(b) $\sum_{x=0}^{\infty} \tau_G(x)[f_G(x)]^{1/t} < \infty$ for some $t > 1$; and

(c) $b^* < \infty$.

Then,

(1) $r(G, d_n^*) - r(G) \leq k_3[8\exp(-k_1 n/p(n)) + k_2\alpha(p(n))n/p(n)]^{1/s}$, where k_1, k_2, k_3 are positive values and $p(n)$ is a positive integer such that $1 \leq p(n) \leq n/2$.

(2) If $\alpha(i) \leq a\rho^i$ where $a > 0$, $0 < \rho < 1$, $i \geq 1$, by choosing $p(n) = [\sqrt{n}]$, then $r(G, d_n^*) - r(G) = O(\exp(-c_3\sqrt{n}))$ for some $c_3 > 0$.

(3) If $\alpha(i) \leq \frac{a}{i^b}$ where $a > 0$, $b > 0$ and $i \geq 1$, by choosing δ such that $1 < \delta < 1 + b$, and letting $p(n) = [n^{\delta/(1+b)}]$, then $r(G, d_n^*) - r(G) = O(n^{(1-\delta)/s})$.

Remark: Note that

$$\int_0^Q \theta dG(\theta) = E[\Theta] = E[E[\Theta|X]] = E[\tau_G(X)] = \sum_{x=0}^{\infty} \tau_G(x) f_G(x).$$

Since $0 < f_G(x) < 1$, the condition that $\sum_{x=0}^{\infty} \tau_G(x)[f(x)]^{1/t} < \infty$ for some $t > 1$ implies that $\int_0^Q \theta dG(\theta) < \infty$.

In the following, we provide examples in each of which the Theorems may be applied.

Example 1 *(The negative binomial distribution)* Suppose that

$$f(x|\theta) = \binom{r+x-1}{x} \theta^x (1-\theta)^r, \; 0 < \theta < 1,$$

where r is a positive integer, and the prior distribution is a beta distribution with the probability density function $g(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$, $0 < \theta < 1$, where $\alpha > 0$, and $\beta > 0$. Then,

$$f_G(x) = \frac{\Gamma(\alpha+\beta)\Gamma(r+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \times \frac{\Gamma(x+\alpha)}{\Gamma(x+\alpha+r+\beta)} \times \binom{r+x-1}{x} \text{ and } \tau_G(x) = \frac{x+\alpha}{x+\alpha+r+\beta}.$$

Since $\tau_G(x) \to 1$ as $x \to \infty$, for $0 < \theta_0 < 1$, $A = \{x | \tau_G(x) > \theta_0\} \neq \phi$. Therefore, $b^* < a^* = \inf A < \infty$.

Note that $f_G(x) \propto x^{-(1+\beta)}$ for sufficiently large x. Hence, for $1 < t < 1 + \beta$, $\sum_{x=0}^{\infty} \tau_G(x)[f_G(x)]^{1/t} < \infty$.

Example 2 *(The Poisson distribution)* Let $f(x|\theta) = e^{-\theta}\theta^x/x!$, $x = 0, 1, \ldots$; $\theta > 0$, and let the prior density be $g(\theta) = \beta^\alpha \theta^{\alpha-1} e^{-\beta\theta}/\Gamma(\alpha)$, where $\alpha > 0$, $\beta > 0$. Then,

$$f_G(x) = \frac{\Gamma(x+\alpha)}{x!\Gamma(\alpha)} \times \frac{\beta^\alpha}{(1+\beta)^{x+\alpha}} \text{ and } \tau_G(x) = \frac{x+\alpha}{1+\beta}.$$

Since $\tau_G(x) \to \infty$ as $x \to \infty$, for finite θ_0, $b^* < \infty$.

Note that $f(x) \propto \frac{(x+\alpha-1)\times\cdots\times(x+1)}{(1+\beta)^{x+\alpha}}$ for sufficiently large x. Hence $\sum_{x=0}^{\infty} \tau_G(x)[f_G(x)]^{1/t} < \infty$ for every $t > 1$.

4. Proof

According to (2.1), the minimum Bayes risk $r(G)$ can be written as:

$$\begin{aligned} r(G) &= \sum_{x=a^*}^{\infty} [\tau_G(x) - \theta_0] f_G(x) + C \\ &= \sum_{x=a^*}^{\infty} [\tau_G(x) - \theta_0] E[I_{\{x\}}(X_{n+1})] + C. \end{aligned} \tag{4.1}$$

From (4.1), (2.3) and the definition of b^* and a^*, we obtain

$$\begin{aligned} 0 &\le r(G,\ d_n^*) - r(G) \\ &= \sum_{x=0}^{b^*} [\theta_0 - \tau_G(x)] E_{n+1}[d_n^*(x) I_{\{x\}}(X_{n+1})] \\ &\quad + \sum_{y=a^*}^{\infty} [\tau_G(y) - \theta_0] E_{n+1}[(1 - d_n^*(y)) I_{\{y\}}(X_{n+1})] \\ &\le \sum_{x=0}^{b^*} [\theta_0 - \tau_G(x)] E_{n+1}[d_n^*(b^*) I_{\{x\}}(X_{n+1})] \\ &\quad + \sum_{y=a^*}^{\infty} [\tau_G(y) - \theta_0] E_{n+1}[(1 - d_n^*(a^*)) I_{\{y\}}(X_{n+1})] \\ &= \sum_{x=0}^{b^*} [\theta_0 - \tau_G(x)] P\{d_n^*(b^*) = 1,\ X_{n+1} = x\} \\ &\quad + \sum_{y=a^*}^{\infty} [\tau_G(y) - \theta_0] P\{d_n^*(a^*) = 0,\ X_{n+1} = y\}. \end{aligned} \tag{4.2}$$

The inequality in (4.2) is obtained by the fact that d_n^* is monotone and therefore $d_n^*(x) \le d_n^*(b^*)$ for all $x \le b^*$, and $1 - d_n^*(y) \le 1 - d_n^*(a^*)$ for all $y \ge a^*$.

By the definitions of d_n^* and a_n, $d_n^*(b^*) = 1$ iff $b^* \geq a_n$ iff $H_n(z) > 0$ for some $0 \leq z \leq b^*$. Hence,

$$\{d_n^*(b^*) = 1,\ X_{n+1} = x\} = \bigcup_{z=0}^{b^*} \{H_n(z) > 0,\ X_{n+1} = x\}. \tag{4.3}$$

Also,

$$d_n^*(a^*) = 0 \Leftrightarrow a^* < a_n \Rightarrow H_n(a^*) \leq 0. \tag{4.4}$$

Combining (4.2)– (4.4) yields that

$$\begin{aligned} 0 &\leq r(G, d_n^*) - r(G) \\ &\leq \sum_{x=0}^{b^*} [\theta_0 - z_G(x)] \left[\sum_{z=0}^{b^*} P\{H_n(z) > 0,\ X_{n+1} = x\} \right] \\ &\quad + \sum_{y=a^*}^{\infty} [\tau_G(y) - \theta_0] P\{H_n(a^*) \leq 0,\ X_{n+1} = y\}. \end{aligned} \tag{4.5}$$

Proof of Theorem 3.1 *(m-Dependence Case)*

For each $z = 0, 1, \ldots$, let

$$\begin{aligned} V_j(z) &= \frac{1}{a(z+1)} I_{\{z+1\}}(X_j) - \frac{\theta_0}{a(z)} I_{\{z\}}(X_j), \\ W_j(z) &= V_j(z) - H(z). \end{aligned}$$

Then, $H_n(z) = \frac{1}{n} \sum\limits_{j=1}^{n-m} V_j(z) + \frac{1}{n} \sum\limits_{j=n-m+1}^{n} V_j(z)$.

Note that for each $0 \leq z \leq b^*$, $H(z) < 0$. Therefore, for $0 \leq z \leq b^*$,

$$\begin{aligned} &\{H_n(z) > 0,\ X_{n+1} = x\} \\ &= \{H_n(z) - H(z) > -H(z),\ X_{n+1} = x\} \\ &\subset \bigcup_{i=1}^{m+1} C_{in}(z), \end{aligned} \tag{4.6}$$

where for each $i = 1, \ldots, m$,

$$C_{in}(z) = \{\frac{1}{n} \sum_{k=0}^{m(n)} W_{km+i}(z) > -H(z)/(m+1),\ X_{n+1} = x\},$$

and

$$C_{m+1,n}(z) = \{\frac{1}{n} \sum_{[m(n)+1]m+1}^{n} W_j(z) > -H(z)/(m+1),\ X_{n+1} = x\},$$

$$m(n) = \left[\frac{n}{m} - 2\right].$$

Note that $-\frac{\theta_0}{a(z)} - H(z) \leq W_j(z) \leq \frac{1}{a(z+1)} - H(z)$. So, $W_j(z)$, $j = 1, \ldots, n$, are bounded, identically distributed random variables. Since m is a fixed integer, $\frac{1}{n} \sum_{[m(n)+1]m+1}^{n} W_j(z)$ tends to zero as n tends to infinity. Since $-H(z)/(m+1)$ is a fixed positive value, the set $C_{m+1,n}(z)$ will become into an empty set when n is sufficiently large.

Recall that $\{X_j,\ j = 1, 2, \ldots\}$ is m-dependent. Hence, for each $i = 1, \ldots, m$, $\frac{1}{n} \sum_{k=0}^{m(n)} W_{km+i}(z)$ and X_{n+1} are independent. Also, for each i, $W_{km+i}(z)$, $k = 0, \ldots, m(n)$, are mutually independent and identically distributed.

Combining the preceding discussions and from (4.6), for n being sufficiently large so that $C_{m+1,n}(z)$ to be an empty set, one has,

$$\begin{aligned}
&P\{H_n(z) > 0, X_{n+1} = x\} \\
&\leq P\left\{\bigcup_{i=1}^{m}\left\{\frac{1}{n}\sum_{k=0}^{m(n)} W_{km+i}(z) > -H(z)/(m+1), X_{n+1} = x\right\}\right\} \\
&\leq \sum_{i=1}^{m} P\left\{\frac{1}{n}\sum_{k=0}^{m(n)} W_{km+i}(z) > -H(z)/(m+1),\ X_{n+1} = x\right\} \\
&= \sum_{i=1}^{m} P\left\{\frac{1}{n}\sum_{k=0}^{m(n)} W_{km+i}(z) > -H(z)/(m+1)\right\} f_G(x) \\
&\leq \sum_{i=1}^{m} \exp\left\{-\frac{2n^2}{(m+1)^2[m(n)+1]} H^2(z)\left[\frac{1}{a(z+1)} + \frac{\theta_0}{a(z)}\right]^{-2}\right\} f_G(x).
\end{aligned} \tag{4.7}$$

In (4.7), the last inequality is obtained by an application of Theorem 2 of Hoeffding (1963) by noting that $W_{km+i}(z)$, $k = 0, 1, \ldots, m(n)$, are iid, with mean 0 and $-\frac{\theta_0}{a(z)} - H(z) \leq W_j(z) \leq \frac{1}{a(z+1)} - H(z)$.

Next, for $z = a^*$, $H(a^*) > 0$. Following an argument analogous to the previous discussion, for n being sufficiently large, we have

$$P\{H_n(a^*) \leq 0,\ X_{n+1} = y\}$$
$$\leq m \exp\left\{ -\frac{2n^2}{(m+1)^2[m(n)+1]} H^2(a^*) \left[\frac{1}{a(a^*+1)} + \frac{\theta_0}{a(a^*)}\right]^{-2} \right\} f_G(y). \tag{4.8}$$

Let $c = \min\{H^2(z)\left[\frac{1}{a(z+1)} + \frac{\theta_0}{a(z)}\right]^{-2} | z = 0, 1, \ldots, b^* \text{ or } z = a^*\}$. Then $c > 0$. Let $b_1 = \sum_{x=0}^{b^*} [\theta_0 - \tau_G(x)] f_G(x)$ and $b_2 = \sum_{y=a^*}^{\infty} [\tau_G(y) - \theta_0] f_G(y)$. Note that, $0 \leq b_1, b_2 < \infty$ by the definitions of b^*, a^* and the assumption that $\int_0^Q \theta dG(\theta) < \infty$.

Then, substituting (4.7) and (4.8) into (4.5), we obtain: for sufficiently large n,

$$\begin{aligned}
&r(G, d_n^*) - r(G) \\
&\leq \sum_{x=0}^{b^*} [\theta_0 - \tau_G(x)] \left\{ \sum_{z=0}^{b^*} m \exp\left\{ -\frac{2n^2 c}{(m+1)^2[m(n)+1]} \right\} \right\} f_G(x) \\
&\quad + \sum_{y=a^*}^{\infty} [\tau_G(y) - \theta_0] m \exp\left\{ -\frac{2n^2 c}{(m+1)^2(m(n)+1)} \right\} f_G(y) \\
&= [(b^*+1)b_1 + b_2] m \exp\left(-\frac{2n^2}{(m+1)^2[m(n)+1]} \right) \\
&\leq [(b^*+1)b_1 + b_2] m \exp(-c_1 n) \\
&= O(\exp(-c_1 n),
\end{aligned}$$

where $c_1 = 2m(m+1)^{-2} > 0$ and the last inequality is obtained by noting that $\frac{n}{m(n)+1} \geq m$.

Hence, the proof of Theorem 3.1 is completed. □

To investigate the asymptotic optimality of the empirical Bayes tests for mixing processes, the following lemmas are introduced.

Lemma 4.1 *(Bernstein inequality for φ-mixing process)*

Let $\{U_i\}$ be a sequence of φ-mixing random variables with φ-mixing coefficients $\{\varphi(i),\ i = 1, 2, \ldots\}$, satisfying $EU_i = 0$, $|U_i| \leq d$, $E|U_i| \leq \delta$, and $E(U_i^2) \leq D$. Denote $\tilde{\varphi}(\ell) = \sum_{i=1}^{\ell} \varphi(i)$ for each $\ell = 1, 2, \ldots$. Then, for $\varepsilon > 0$,

$$P\{|\sum_{i=1}^{n} U_i| > \varepsilon\} \leq 2\exp(-\lambda\varepsilon + 3\sqrt{e}n\varphi(\ell)/\ell + 6n\lambda^2(D + 4\delta D\tilde{\varphi}(\ell))),$$

where λ and ℓ are, respectively, a positive real number and a positive integer less than or equal to n satisfying $\lambda \ell d \leq \frac{1}{4}$.

Lemma 4.1 is due to Collomb (1984) which gives a sharper bound than a similar result obtained by Bosq (1975).

Lemma 4.2 *(Bosq (1998))* Let $\{U_i\}$ be a sequence of α-mixing process with α-mixing coefficients $\{\alpha(i);\ i = 1, 2, \ldots\}$, satisfying $EU_i = 0$. Consider the following assumptions.

A1. There exist positive constants L_1 and L_2 such that

$$0 < pL_1 \leq ||U_{t+1} + \ldots + U_{t+p}||_\infty \leq pL_2 \text{ for all } t \text{ and } p \geq 1.$$

A2. There exist positive constants L_1 and L_2 such that

(a) $0 < L_1 \leq EU_t^2 \leq L_2 < \infty$ for all t.

(b) $E|U_t|^k \leq L_2^{k-2} k! EU_t^2$ for all $k \geq 3$, $t = 1, 2, \ldots$.

(1) If A1 is satisfied, and $1 \leq p(n) \leq \frac{n}{2}$, then for every $\varepsilon > 0$

$$P\{|\sum_{i=1}^{n} U_i| > n\varepsilon\} \leq 8 \exp\left(-\frac{\varepsilon^2}{25L_2^2} \times \frac{n}{p(n)}\right) + 18 \frac{L_2}{\sqrt{L_1 \varepsilon}} \frac{n}{p(n)} \alpha(p(n))$$

(2) If A2 is satisfied and $1 \leq p(n) \leq \frac{n}{2}$, then for every $\varepsilon > 0$ and $\gamma \geq 2$,

$$P\left\{|\sum_{i=1}^{n} U_i| > n\varepsilon\right\} \leq (2p(n) + 1 + \frac{1}{L_2}) \exp(-\frac{\varepsilon^2}{10L_2(5+\varepsilon)} \frac{n}{p(n)})$$
$$+ c_\gamma (1 + \frac{1}{\varepsilon})^{\beta_\gamma} n [\alpha(p(n))]^{2\beta_\gamma},$$

where $\beta_\gamma = \frac{\gamma}{2\gamma+1}$ and $c_\gamma = 11(L_2^{(\gamma-1)/\gamma} (\gamma!)^{1/\gamma} \frac{5}{2} (1 + \frac{4}{5\sqrt{L_1}}))^{\beta_\gamma}$.

Lemma 4.3 For s, $t > 0$ such that $\frac{1}{s} + \frac{1}{t} = 1$, we have

(a) For each $z = 0, 1, \ldots, b^*$,

$$P\{H_n(z) > 0,\ X_{n+1} = x\} \leq [P\{H_n(z) > 0\}]^{1/s} [f_G(x)]^{1/t};$$

(b) $P\{H_n(a^*) \le 0,\ X_{n+1} = y\} \le [P\{H_n(a^*) \le 0\}]^{1/s}[f_G(y)]^{1/t}$.

Proof: These are results of an application of Hölder inequality. □

Now from (4.5) and Lemma 4.3,

$$\begin{aligned} &r(G, d_n^*) - r(G) \\ &\le \sum_{x=0}^{b^*}[\theta_0 - \tau_G(x)][f_G(x)]^{1/t}\left\{\sum_{z=0}^{b^*}[P\{H_n(z) > 0\}]^{1/s}\right\} \\ &\quad + \sum_{y=a^*}^{\infty}[\tau_G(y) - \theta_0][f_G(y)]^{1/t}[P\{H_n(a^*) \le 0\}]^{1/s} \\ &= \sum_{x=0}^{b^*}[\theta_0 - \tau_G(x)][f_G(x)]^{1/t}[\sum_{z=0}^{b^*}P^{1/s}\{\frac{1}{n}\sum_{j=1}^{n}W_j(z) > -H(z)\}] \\ &\quad + \sum_{y=a^*}^{\infty}[\tau_G(y) - \theta_0][f_G(y)]^{1/t}P^{1/s}\{\frac{1}{n}\sum_{j=1}^{n}W_j(a^*) \le -H(a^*)\}. \end{aligned} \tag{4.9}$$

Proof of Theorem 3.2 *(φ-mixing case)*

Recall that $\{X_j,\ j = 1, 2, \ldots\}$ is a φ-mixing process. Hence, $\{W_j(z),\ j = 1, 2, \ldots,\}$ is also a φ-mixing process with the same φ-mixing coefficients $\{\varphi(i), i = 1, 2, \ldots\}$. Also, $W_j(z),\ j = 1, 2, \ldots$ are identically distributed with $E[W_j(z)] = 0$, $\eta_1(z) \le W_j(z) \le \eta_2(z)$ where $\eta_1(z) = -\frac{\theta_0}{a(z)} - H(z) < 0$, $\eta_2(z) = \frac{1}{a(z+1)} - H(z) > 0$. Let $\eta(z) = \max(|\eta_1(z)|,\ \eta_2(z))$. Then, $E(|W_j(z)|) \le \eta(z)$ and $E(|W_j(z)|^2) \le \eta^2(z)$. Then, by Lemma 4.1, for each $z = 0, 1, \ldots, b^*$,

$$\begin{aligned} &P\{\frac{1}{n}\sum_{j=1}^{n}W_j(z) > -H(z)\} \\ &\le 2\exp\{-n\lambda(-H(z)) + 3\sqrt{e}n\varphi(\ell)/\ell + n6\lambda^2[\eta^2(z) + 4\eta^3(z)\tilde{\varphi}(\ell)]\} \\ &= 2\exp\{-n[-\lambda H(z) - 3\sqrt{e}\varphi(\ell)/\ell - 6\lambda^2[\eta^2(z) + 4\eta^3(z)\tilde{\varphi}(\ell)]]\}, \end{aligned} \tag{4.10}$$

and

$$\begin{aligned} &P\{\frac{1}{n}\sum_{j=1}^{n}W_j(a^*) \le -H(a^*)\} \\ &\le 2\exp\{-n[\lambda H(a^*) - 3\sqrt{e}\varphi(\ell)/\ell - 6\lambda^2[\eta^2(a^*) + 4\eta^3(a^*)\tilde{\varphi}(\ell)]]\} \end{aligned} \tag{4.11}$$

where $\lambda > 0$ and ℓ, a positive integer, are chosen so that $\lambda\ell\eta(z) \le \frac{1}{4}$.

We choose $\ell = \ell(n)$ so that $\ell(n) \le n$, $\ell(n) \to \infty$ and $\frac{\ell(n)}{n} \to 0$ as $n \to \infty$. Let $\lambda = \lambda(n,z) = \frac{1}{4\eta(z)\ell(n)}$. For such chosen $\ell(n)$ and $\lambda(n,z)$,

$$\begin{aligned}&\lambda|H(z)| - 3\sqrt{e}\varphi(\ell)/\ell - 6\lambda^2[\eta^2(z) + 4\eta^3(z)\tilde{\varphi}(\ell)] \\ &= \frac{1}{\ell(n)}[\frac{|H(z)|}{4\eta(z)} - 3\sqrt{e}\varphi(\ell(n)) - \frac{3}{8}[\frac{1}{\ell(n)} + \frac{4\eta(z)\tilde{\varphi}(\ell(n))}{\ell(n)}]].\end{aligned} \tag{4.12}$$

Note that $\varphi(\ell(n)) \to 0$, $\frac{1}{\ell(n)} \to 0$ and $\frac{\tilde{\varphi}(\ell(n))}{\ell(n)} \to 0$ as $n \to \infty$. Hence, for sufficiently large n, we can obtain

$$\lambda|H(z)| - 3\sqrt{e}\varphi(\ell)/\ell - 6\lambda^2[\eta^2(z) + 4\eta^3(z)\tilde{\varphi}(\ell)] \ge \frac{|H(z)|}{8\eta(z)\ell(n)}. \tag{4.13}$$

Let $c_2 = \min\{\frac{|H(z)|}{8s\eta(z)}|z = 0, 1, \ldots, b^* \text{ or } z = a^*\}$. Then $c_2 > 0$. Also, let $k = (b^* + 1)2^{1/s} \sum_{x=0}^{b^*}[\theta_0 - \tau_G(x)][f_G(x)]^{1/t} + 2^{1/s} \sum_{y=a^*}^{\infty}[\tau_G(y) - \theta_0][f_G(y)]^{1/t}$. By assumption (b), $k < \infty$.

Now, plugging (4.10), (4.11) and (4.13) into (4.9), we obtain:

$$\begin{aligned}&r(G, d_n^*) - r(G) \\ &\le \sum_{x=0}^{b^*}[\theta_0 - \tau_G(x)][f_G(x)]^{1/t}[\sum_{z=0}^{b^*} 2^{1/s}\exp(-\frac{n}{\ell(n)}\frac{|H(z)|}{8s\eta(z)})] \\ &\quad + \sum_{y=a^*}^{\infty}[\tau_G(y) - \theta_0][f_G(y)]^{1/t}2^{1/s}\exp(-\frac{n}{\ell(n)}\frac{H(a^*)}{8s\eta(a^*)}) \\ &\le \sum_{x=0}^{b^*}[\theta_0 - \tau_G(x)][f_G(x)]^{1/t}(b^* + 1)2^{1/s}\exp(-\frac{c_2 n}{\ell(n)}) \\ &\quad + \sum_{y=a^*}^{\infty}[\tau_G(y) - \theta_0][f_G(y)]^{1/t}2^{1/s}\exp(-\frac{c_2 n}{\ell(n)}) \\ &= k\exp(-\frac{c_2 n}{\ell(n)}) \\ &= O(\exp(-c_2 n/\ell(n))).\end{aligned} \tag{4.14}$$

Hence, the proof of Theorem 3.2 is completed. □

Remark: In (4.14), the requirements for the choice of the integer $\ell(n)$ are: $\ell(n) \le n$ and $\ell(n) \to \infty$ as $n \to \infty$. Hence, we may choose these $\ell(n)$ which tend to infinity in a very slow speed. For example, let $\ell(n) = [\log\ n]$. Then, we have $r(G, d_n^*) - r(G) =$

$O(\exp(-c_2 n/\log\ n))$. This rate of convergence is faster than that of $O(\exp(-c_2 n^{1-\varepsilon}))$ for every $\varepsilon > 0$. In the iid case, the rate of convergence of Liang (1988) is of $O(\exp(-cn))$ for some $c > 0$.

Proof of Theorem 3.3 *(α-mixing case)*

Recall that $\{X_j, j = 1,2,\ldots\}$ is an α-mixing process. So, $\{W_j(z), j = 1,2,\ldots\}$ is also an α-mixing process with the same α-mixing coefficients $\{\alpha(i);\ i = 1,2,\ldots\}$. Also, $W_j(z),\ j = 1,2,\ldots$ are identically distributed with $EW_j(z) = 0$, $\eta_1(z) \le W_j(z) \le \eta_2(z)$. Hence, $||W_{t+1}(z)+\ldots+W_{t+p}(z)||_\infty = p\eta(z)$ where $\eta(z) = \max(|\eta_1(z)|, \eta_2(z))$, for all $t \ge 1$ and all positive integers p. Hence the assumption A1 is satisfied.

Similarly, for each $j = 1,2,\ldots$, $0 < E[W_j^2(z)] \le \eta^2(z)$ and for each integer $k \ge 3$, $E[|W_j(z)|^k] \le \eta^{k-2}(z)E[W_j^2(z)] \le \eta^{k-2}(z)k!E[W_j^2(z)]$. Hence the assumption A2 is also satisfied. Therefore, either part of Lemma 4.2 can be applied. In the following, we apply the result of Lemma 4.2 (1) only.

By Lemma 4.2 (1), for each $z = 0,1,\ldots,b^*$,

$$P\left\{\frac{1}{n}\sum_{j=1}^{n} W_j(z) > -H(z)\right\} \le 8\exp(-\frac{[H(z)]^2}{25\eta(z)} \times \frac{n}{p(n)}) + 18\sqrt{\frac{\eta(z)}{|H(z)|}\frac{n}{p(n)}}\alpha(p(n)) \tag{4.15}$$

and

$$P\{\frac{1}{n}\sum_{j=1}^{n} W_j(a^*) < -H(a^*)\} \le 8\exp\{-\frac{H^2(a^*)}{25\eta(a^*)}\frac{n}{p(n)}\} + 18\sqrt{\frac{\eta(a^*)}{H(a^*)}\frac{n}{p(n)}}\alpha(p(n)) \tag{4.16}$$

where $p(n)$ is an integer such that $1 \le p(n) \le \frac{n}{2}$.

Let

$$k_1 = \min\left\{\frac{H^2(z)}{25\eta(z)}|z = 0,1,\ldots,b^* \text{ or } z = a^*\right\} \text{ and}$$
$$k_2 = \max\{18\sqrt{\eta(z)/|H(z)|}|z = 0,1,\ldots b^* \text{ or } z = a^*\}.$$

We see that $0 < k_1, k_2 < \infty$.

Plugging (4.15) and (4.16) into (4.9), we obtain

$$
\begin{aligned}
&r(G, d_n^*) - r(G) \\
&\leq \sum_{x=0}^{b^*} [\theta_0 - \tau_G(x)][f_G(x)]^{1/t} \left[\sum_{z=0}^{b^*} [8 \exp(-k_1 \frac{n}{p(n)}) + k_2 \frac{n}{p(n)} \alpha(p(n))]^{1/s} \right] \\
&\quad + \sum_{y=a^*}^{\infty} [\tau_G(y) - \theta_0][f_G(y)]^{1/t} [8 \exp(-k_1 \frac{n}{p(n)}) + k_2 \frac{n}{p(n)} \alpha(p(n))]^{1/s} \\
&= k_3 [8 \exp(-k_1 \frac{n}{p(n)}) + k_2 \alpha(p(n)) n/p(n)]^{1/s},
\end{aligned}
\tag{4.17}
$$

where $k_3 = (b^* + 1) \sum_{x=0}^{b^*} [\theta_0 - \tau_G(x)][f_G(x)]^{1/t} + \sum_{y=a^*}^{\infty} [\tau_G(y) - \theta_0][f_G(y)]^{1/t}$.

The results of parts (2) and (3) of Theorem 3.3 can be obtained by replacing $\alpha(i)$ and $p(n)$ by the corresponding forms and following a straightforward computation. The detail is omitted. □

References

Bosq, D. (1975). Inégalité de Bernstein pour les processus stationnaires et mélangeants. *C. R. Acad. Sci., Paris, Ser. A* **290**, 711–713.

Bosq, D. (1993). Bernstein-type large deviations inequalities for partial sums of strong mixing processes. *Statistics*, **24**, 59–70.

Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. *Z. Wahrsch. Verw. Gebiete* **66**, 441–460.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.

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.

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.

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.

Robins, 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.