Statistica Sinica **8**(1998), 589-605

CONVERGENCE RATES OF EMPIRICAL BAYESIAN ESTIMATION IN A CLASS OF LINEAR MODELS

Laisheng Wei

University of Science and Technology of China

Abstract: By using kernel estimators of a multivariate density function and its partial derivatives, and the estimators of the nuisance parameters, we construct empirical Bayes (EB) estimators of parameters in a class of linear models. Under suitable moment conditions on the prior distribution, the proposed EB estimators are asymptotically optimal with rates arbitrarily close to $O(n^{-1})$.

Key words and phrases: Convergence rates, empirical Bayesian estimation, linear models.

1. Introduction

Since the empirical Bayes (EB) procedure was first proposed by Robbins (1955, 1964), it has received considerable attention in the literature. Suppose that in the pair (Y, α) of random vectors, Y is observable and parameter vector α is unobservable. The conditional distribution of Y given α is specified by the density f_α on the observation space $\mathcal{Y}$, and α has an unknown and unspecified distribution on parameter space Θ. Based on an observation on Y, the problem is to decide on α subject to a nonnegative loss function. If the prior distribution G were known, we could estimate α by the Bayes estimator $\widehat{\alpha}_G$ which achieves the minimum Bayes risk $R(G)$ with respect to (w.r.t.) G. But since G is not known, the optimal estimator (o.e.) $\widehat{\alpha}_G$ is not practically available. In the EB decision problem, we assume that the above problem has occurred independently in the past, say n times. Hence, we actually have n+1 independent pairs $(Y^{(1)}, \alpha^{(1)})$, $\ldots$, $(Y^{(n)}, \alpha^{(n)})$ and (Y, α). Our purpose is to use the information contained in the historical observation $(Y^{(1)}, \ldots, Y^{(n)})$ and the present observation Y to obtain an estimator $\widehat{\alpha}_n = \widehat{\alpha}_n(Y^{(1)}, \ldots, Y^{(n)}; Y)$ for the present parameter vector α, called the EB estimator, so that for large n it performs "nearly" as good as the unavailable o.e. $\widehat{\alpha}_G$ in the following sense: the overall Bayes risk, say R_n, of $\widehat{\alpha}_n$ approximates the minimum Bayes risk $R(G)$ attained by $\widehat{\alpha}_G$. If $\lim_{n\to\infty} R_n = R(G)$, then the EB estimators are called asymptotically optimal (a.o.). If for some $q > 0$, $R_n - R(G) = O(n^{-q})$, we will say the EB estimators are a.o. with convergence rates $O(n^{-q})$. From the above arguments, it follows that the EB approach to the statistical decision problem is applicable when the same decision problem has presented itself repeatedly and independently with a fixed but unknown prior distribution of parameters.

For EB decision problems in a linear model, Singh (1985) considered the EB estimation problem of regression coefficients in a linear regression model under known error-variance. Wei (1990) studied the EB test problem for the same model. Zhang and Wei (1994), Wei and Zhang (1995) extended Singh's work to the case of unknown error-variance; they discussed asymptotic optimality and convergence rates for the EB estimators of regression coefficients and error-variance in the above model. Recently, Wei (1995) considered asymptotic optimality of the EB estimators in a one-way analysis of variance (ANOVA) model. In this paper we study the convergence rates of EB estimators for a class of linear models.

In section 2 of this paper we derive the Bayes estimator of parameters, and in section 3 we construct the EB estimator of parameters and state the main result about convergence rates. We consider applications for the class of linear models in section 4. Finally, some necessary lemmas and the technical proof of one theorem are presented in the Appendices.

2. The Bayes Estimator of a Parameter Vector

Consider a class of linear models as follows:

$$Y_{m\times 1} = Z_{m\times q}\theta_{q\times 1} + X_{m\times a}\alpha_{a\times 1} + e_{m\times 1}, \tag{2.1}$$

where $X^\tau X$ is assumed to be non-singular, $\alpha = (\alpha_1, \ldots, \alpha_a)^\tau$ is the estimated parameter vector, $\theta = (\theta_1, \ldots, \theta_q)^\tau$ are nuisance parameters, $Y = (Y_1, \ldots, Y_m)^\tau$ is the observation vector, and $e = (e_1, \ldots, e_m)^\tau$ is an unobservable random vector, the conditional distribution of e given α is $N(0, \sigma^2 I_m)$ with σ^2 unknown. In this paper, we assume that σ^2 is bounded away from both zero and infinity, i.e., $0 < \gamma_0 \le \sigma^2 \le M_0 < \infty$, where γ_0 and M_0 are constants.

It is easy to see that linear regression model, the ANOVA and analysis of covariance (ANOCOVA) models are special cases of model (2.1).

Let the loss function be given by

$$L(d, \alpha) = \|d - \alpha\|^2, \tag{2.2}$$

where $d = (d_1, \ldots, d_a)^\tau \in \mathcal{D}$, the decision space; $\alpha \in \Theta$, the parametric space; and $\|t\|^2 = t^\tau t = \sum_{i=1}^a t_i^2$.

Suppose that the prior distribution G of α is a member of the following family

$$\mathcal{F}_\delta = \{G(\alpha) : \int_\Theta \|\alpha\|^\delta dG(\alpha) < \infty\}, \tag{2.3}$$

where $\delta \geq 2$. Then the conditional density of Y given α is

$$f(y|\alpha) = (2\pi\sigma^2)^{-\frac{m}{2}} \exp\Big\{ -\frac{1}{2\sigma^2}\|y - Z\theta - X\alpha\|^2\Big\}, \tag{2.4}$$

and the marginal density of Y is

$$f(y) = \int_\Theta f(y|\alpha)dG(\alpha). \tag{2.5}$$

By formular (2.5) and Lemma A.1 in Appendix A, we have

$$\begin{aligned}\sigma^2\frac{\partial f(y)}{\partial y} &= -\int_\Theta [y - Z\theta - X\alpha]f(y|\alpha)dG(\alpha)\\ &= -[y - Z\theta]f(y) + X\int_\Theta \alpha f(y|\alpha)dG(\alpha),\end{aligned} \tag{2.6}$$

where

$$\frac{\partial f(y)}{\partial y} = (f'_{(1)}(y), \ldots, f'_{(m)}(y))^\tau, \quad f'_{(i)}(y) = \frac{\partial f(y)}{\partial y_i},\ i = 1, \ldots, m. \tag{2.7}$$

Since $X^\tau X$ is an invertible matrix, from (2.6) we have

$$\int_\Theta \alpha f(y|\alpha)dG(\alpha) = (X^\tau X)^{-1}X^\tau\Big[(y - Z\theta)f(y) + \sigma^2\frac{\partial f(y)}{\partial y}\Big]. \tag{2.8}$$

Hence, we get the Bayes estimator of α under the loss function (2.2) as follows

$$\begin{aligned}\phi_G(y) = E(\alpha|y) &= \int_\Theta \alpha f(y|\alpha)dG(\alpha)/f(y)\\ &= (X^\tau X)^{-1}X^\tau[y - Z\theta + \sigma^2\psi(y)],\end{aligned} \tag{2.9}$$

where

$$\begin{aligned}\psi(y) &= \frac{\partial f(y)}{\partial y}/f(y) = (\psi_{(1)}(y), \ldots, \psi_{(m)}(y))^\tau,\\ \psi_{(i)}(y) &= f'_{(i)}(y)/f(y), \quad i = 1, \ldots, m.\end{aligned} \tag{2.10}$$

The minimum Bayes risk of $\phi_G(y)$ w.r.t. G is

$$R(G) = R(\phi_G, G) = E_{(Y,\alpha)}\|\phi_G(Y) - \alpha\|^2, \tag{2.11}$$

where $E_{(Y,\alpha)}$ denotes the expectation w.r.t. the joint distribution of random vector Y and α.

We know that $R(G) = \inf_{\alpha^*} R(\alpha^*, G)$, where the inf is taken over the set of all possible estimators α^* for which $R(\alpha^*, G)$ is finite. The estimator which achieves the minimum Bayes risk $R(G)$ is the Bayes estimator $\widehat{\alpha} = \phi_G$ given by (2.9). Thus $R(\widehat{\alpha}, G) = R(G)$. Notice that $R(G)$ can be exactly attained only if the prior distribution G is known and α is estimated by $\widehat{\alpha}$. Unfortunately, G is unknown and hence $\widehat{\alpha}$ is unavailable to us. This leads to the EB approach to exhibit estimators whose risks are close to $R(G)$.

3. The EB Estimation and the Main Result

In the EB framework, we make the following assumptions: Let $(Y^{(1)}, \alpha^{(1)}), \ldots, (Y^{(n)}, \alpha^{(n)})$ be independent pairs from a historical record and $(Y^{(n+1)}, \alpha^{(n+1)}) = (Y, \alpha)$ from the present experiment, with

$$Y^{(l)} = Z\theta + X\alpha^{(l)} + e^{(l)}, \quad l = 1, \ldots, n+1, \tag{3.1}$$

where the vectors $Y^{(l)}, \alpha^{(l)}, e^{(l)}$, $l = 1, \ldots, n+1$ are i.i.d. as described in (2.1). Usually $Y^{(1)}, \ldots, Y^{(n)}$ are said to be the past (or historical) samples and Y is called the present sample.

In order to obtain the EB estimator of α, we use the class of kernel functions similar to Singh (1981) and Lu (1982) to construct the kernel estimator of the multivariate density $f(y)$ and its derivatives as follows:

Let $P_i(x)$, $x \in R_1$, $i = 0, 1, \ldots, k-1$ be a class of Borel-measurable bounded functions vanishing outside (0,1) such that for each $0 \leq i \leq k-1$

$$\frac{1}{l!}\int_0^1 x^l P_i(x)dx = \begin{cases} 1, & l = i, \\ 0, & l \neq i, l = 0, 1, \ldots, k-1, \end{cases} \tag{3.2}$$

where $k \geq 2$ is an integer.

Let $K_r(u) = \prod_{i=1}^m P_{r_i}(u_i)$. It is easy to show that

$$\frac{1}{l_1! \ldots l_m!}\int_{R_m} K_r(u)(\prod_{i=1}^m u_i^{l_i})du = \begin{cases} 1, & \text{if } l_i = r_i,\ i = 1, \ldots, m, \\ 0, & \text{otherwise}, \end{cases} \tag{3.3}$$

where $u = (u_1, \ldots, u_m)^\tau \in R_m$, $r = \sum_{i=1}^m r_i$, $0 \leq r \leq k-1$, $r_i \geq 0$, $l_i \geq 0$, $i = 1, \ldots, m$ and $0 \leq \sum_{i=1}^m l_i \leq k-1$.

The kernel estimator of $f^{(r)}(y)$ (see (A.1) in the Appendix A) is defined as follows

$$f_n^{(r)}(y) = \frac{1}{nh^{m+r}}\sum_{l=1}^n K_r\Big(\frac{Y^{(l)} - y}{h}\Big), \quad r = 0, 1, \ldots, k-1, \tag{3.4}$$

where $h > 0$ and $h \to 0$ as $n \to \infty$. When $r = 0$ in (3.4) we get the following kernel estimator of $f^{(0)}(y) = f(y)$

$$f_n(y) = f_n^{(0)}(y) = \frac{1}{nh^m}\sum_{l=1}^n \Big[\prod_{t=1}^m P_0\Big(\frac{Y_t^{(l)} - y_t}{h}\Big)\Big]. \tag{3.5}$$

When $r = 1$ in (3.4) we obtain the kernel estimator of $f'_{(i)}(y)$ as follows

$$f'_{n(i)}(y) = \frac{1}{nh^{m+1}} \sum_{l=1}^{n} \Big\{ \Big[\prod_{\substack{t=1 \\ t\neq i}}^{m} P_0\Big(\frac{Y_t^{(l)} - y_t}{h}\Big) \Big] P_1\Big(\frac{Y_i^{(l)} - y_i}{h}\Big) \Big\}. \qquad (3.6)$$

$$(i = 1, \ldots, m)$$

Let

$$\psi_{n(i)}(y) = \Big[\frac{f'_{n(i)}(y)}{f_n(y)}\Big]_{n^\nu}, \ i = 1, \ldots, m, \quad [d]_L = \begin{cases} d, & \text{if } |d| \le L, \\ 0, & \text{if } |d| > L, \end{cases} \qquad (3.7)$$

where $0 < \nu < 1$ to be determined, and the estimator of $\psi(y)$ is defined by

$$\psi_n(y) = (\psi_{n(1)}(y), \ldots, \psi_{n(m)}(y))^\tau. \qquad (3.8)$$

The nuisance parameters σ^2 and θ involved in (2.1) are estimated by the past samples as follows

$$\widehat{\sigma}^2 = \begin{cases} \widehat{\sigma}_n^2, & \text{if } \gamma_0 \le \widehat{\sigma}_n^2 \le M_0, \\ M_0, & \text{if } \widehat{\sigma}_n^2 > M_0, \\ \gamma_0, & \text{if } \widehat{\sigma}_n^2 < \gamma_0, \end{cases} \qquad (3.9)$$

where γ_0 and M_0 are given in (2.1),

$$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{l=1}^{n} \widehat{\sigma}_{(l)}^2, \qquad \widehat{\sigma}_{(l)}^2 = \frac{1}{s} Y^{(l)\tau} H Y^{(l)}, \qquad (3.9a)$$

$$\widehat{\theta} = \frac{1}{n} \sum_{l=1}^{n} \widehat{\theta}_{(l)}, \qquad \widehat{\theta}_{(l)} = DY^{(l)}, \qquad (3.10)$$

and where H and D are constant matrixes such that $\widehat{\sigma}_{(l)}^2$ and $\widehat{\theta}_{(l)}$ are unbiased estimators of the nuisance parameters σ^2 and θ. Furthermore, assume that H is an idempotent matrix with rank $R(H) = s$ such that $s\widehat{\sigma}_{(l)}^2/\sigma^2$ is distributed according to χ^2 with s degrees of freedom; denote this by $s\widehat{\sigma}_{(l)}^2/\sigma^2 \sim \chi_s^2$.

Then, the EB estimator of α can be formulated as

$$\phi_n = \phi_n(y) = (X^\tau X)^{-1} X^\tau [y - Z\widehat{\theta} + \widehat{\sigma}^2 \psi_n(y)]. \qquad (3.11)$$

Throughout this paper, let E and E_* be the expectation w.r.t. the joint distribution of $(Y^{(1)}, \ldots, Y^{(n)})$ and $(Y^{(1)}, \ldots, Y^{(n)}, (Y, \alpha))$ respectively. Then, the "overall" Bayes risk of ϕ_n is

$$R_n = R_n(\phi_n, G) = E_* \|\phi_n - \alpha\|^2. \qquad (3.12)$$

By definition, if for some $q > 0$, $R_n - R(G) = O(n^{-q})$ then the convergence rates of ϕ_n are said to be of the order $O(n^{-q})$.

The main result about the convergence rates of EB estimators for the model (2.1) is given in the following theorem.

Theorem 3.1. *Let $\phi_G(y)$ be defined by* (2.9) *and $\phi_n(y)$ be as given by* (3.11) *with $h = n^{-\nu}$ and $\nu = 1/(2k+m)$. If for $\delta = [(m+\xi)\lambda/(\eta-\lambda)] \vee (2\lambda k)$ with $1/2 < \lambda < \eta < 1$ and an arbitrarily small number $\xi > 0$ such that $G(\alpha) \in \mathcal{F}_\delta$, i.e.,*

$$\int_\Theta \|\alpha\|^\delta dG(\alpha) < \infty, \tag{3.13}$$

then

$$R_n - R(G) = O\Big(n^{-\frac{2(\lambda k-1)}{2k+m}}\Big),$$

where $k \geq 2$ is a given natural number and m is the dimension of the vector Y.

In the above theorem, it is easy to see that if k is large enough and λ is chosen arbitrarily close to 1, then the convergence rates can be arbitrarily close to $O(n^{-1})$. The proof of Theorem 3.1 is given in Appendix B.

Remark 3.1. If the error-variance σ^2 is known in model (2.1), it is a special case in which the upper bound M_0 and the lower bound γ_0 of σ^2 are equal. Therefore the estimator $\widehat{\sigma}^2$ in (3.9) becomes constant, and the J_2 appearing in Appendix B turns out to be zero. In this case, the conclusion about convergence rates is still true, and the proof is a little easier than that given in Appendix B.

4. Applications

In this section, we consider EB estimation problems for several general linear models, which are particular applications of model (2.1).

4.1. The linear regression model

In model (2.1), let $Z = 0$ and suppose that the elements of the matrix X are values taken on by observations on continuous variables; then model (2.1) becomes the following linear regression model:

$$Y_{m\times 1} = X_{m\times a}\alpha_{a\times 1} + e_{m\times 1}, \tag{4.1}$$

where the components of α are regression coefficients.

The Bayes estimator and the EB estimator of α given by (2.9) and (3.11) become

$$\phi_G(y) = (X^\tau X)^{-1}X^\tau[y + \sigma^2\psi(y)], \tag{4.2}$$

$$\phi_n(y) = (X^\tau X)^{-1}X^\tau[y + \widehat{\sigma}^2\psi_n(y)], \tag{4.3}$$

where X is described as in model (4.1), $\psi(y)$ and $\psi_n(y)$ are expressed by (2.10) and (3.8), $\widehat{\sigma}^2$ is given by (3.9) in which $\widehat{\sigma}^2_{(l)}$ is defined by (3.9a) with $H = I_m - X(X^\tau X)^{-1}X^\tau$ and $s = m - a$, i.e.,

$$\widehat{\sigma}^2_{(l)} = Y^{(l)^\tau}[I_m - X(X^\tau X)^{-1}X^\tau]Y^{(l)}/(m-a). \tag{4.4}$$

It is obvious that $\widehat{\sigma}^2_{(l)}$ $(l = 1, \ldots, n)$ are i.i.d. and $(m-a)\widehat{\sigma}^2_{(l)}/\sigma^2 \sim \chi^2_{m-a}$.

By Theorem 3.1, we get the convergence rates of EB estimators for the linear regression model as follows.

Subtheorem 4.1. *Let $\phi_G(y)$ be defined by* (4.2), *and $\phi_n(y)$ be as given by* (4.3) *with $h = n^{-1/(2k+m)}$. If condition* (3.13) *is satisfied, then*

$$R_n - R(G) = O\Big(n^{-\frac{2(\lambda k-1)}{2k+m}}\Big),$$

where $k \geq 2$ is a given natural number and m is the dimension of the vector Y.

Remark 4.1. If the error-variance σ^2 is known in model (4.1), that means the upper bound M_0 and the lower bound γ_0 of σ^2 are equal; then the estimator $\widehat{\sigma}^2$ in (3.9) becomes constant. In this special case, the conclusion of Subtheorem 4.1 is similar to the main result of Singh (1985) (see Theorem 6.1 of Singh (1985)).

4.2. The one-way ANOVA model

In model (2.1), suppose that $m = ab, q = 1$ and $\theta = \mu$. Let $Y = (Y_1, \ldots, Y_m)^\tau = (Y_{11}, \ldots, Y_{1b}; \cdots; Y_{a1}, \ldots, Y_{ab})^\tau$ and $e = (e_1, \ldots, e_m)^\tau = (e_{11}, \ldots, e_{1b}; \cdots; e_{a1}, \ldots, e_{ab})^\tau$. Then, model (2.1) becomes the following balanced one-way ANOVA model:

$$Y_{m\times 1} = Z_{m\times 1}\mu_{1\times 1} + X\alpha + e_{m\times 1} = (\mathbf{l}_a \otimes \mathbf{l}_b)\mu + (I_a \otimes \mathbf{l}_b)\alpha + e, \tag{4.5}$$

where $Z = \mathbf{l}_a \otimes \mathbf{l}_b$ and $X = I_a \otimes \mathbf{l}_b$, the parameter μ stands for the global mean and $\alpha = (\alpha_1, \ldots, \alpha_a)^\tau$ satisfying the constraint $\sum_{i=1}^a \alpha_i = 0$ denotes the treatment effect of a factor, say A, the vector $\mathbf{l}_p = (1, \ldots, 1)^\tau$ stands for a p-dimensional vector with all its components being one, I_p is an identity matrix and the symbol $\otimes$ denotes the kronecker product.

The Bayes estimator and EB estimator of α given by (2.9) and (3.11) become, respectively;

$$\phi_G(y) = (X^\tau X)^{-1}X^\tau[y - Z\mu - \sigma^2\psi(y)] \tag{4.6}$$

$$\phi_n(y) = (X^\tau X)^{-1}X^\tau[y - Z\widehat{\mu} - \widehat{\sigma}^2\psi_n(y)], \tag{4.7}$$

where Z and X are given in model (4.5), $\psi(y)$ and $\psi_n(y)$ are expressed by (2.10) and (3.8), $\widehat{\mu} = \widehat{\theta}$ and $\widehat{\mu}_{(l)} = \widehat{\theta}_{(l)}$ are defined by (3.10) in which $D = (\mathbf{l}_a \otimes \mathbf{l}_b)^\tau/(ab)$, i.e.,

$$\widehat{\mu} = \frac{1}{n}\sum_{l=1}^{n}\widehat{\mu}_{(l)}, \qquad \widehat{\mu}_{(l)} = \overline{Y}^{(l)}_{..} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} Y^{(l)}_{ij}, \tag{4.8}$$

and $\widehat{\sigma}^2$ is given by (3.9) in which $\widehat{\sigma}^2_{(l)}$ is defined by (3.9a) with $s = a(b-1)$ and $H = I_a \otimes I_b - (I_a \otimes J_b)/b$, $J_b = \mathbf{l}_b \otimes \mathbf{l}_b^\tau$, i.e.,

$$\widehat{\sigma}^2_{(l)} = \frac{1}{a(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}(Y^{(l)}_{ij} - \overline{Y}^{(l)}_{i.})^2, \quad \overline{Y}^{(l)}_{i.} = \frac{1}{b}\sum_{j=1}^{b} Y^{(l)}_{ij},\ i = 1,\ldots,a. \tag{4.9}$$

It is obvious that $\widehat{\sigma}^2_{(l)}$ (l=1,...,n) are i.i.d. and $a(b-1)\widehat{\sigma}^2_{(l)}/\sigma^2 \sim \chi^2_{a(b-1)}$.

By Theorem 3.1, we obtain the following convergence rates of EB estimators for the one-way ANOVA model.

Subtheorem 4.2. *Let $\phi_G(y)$ be defined by* (4.6), *and $\phi_n(y)$ be as given by* (4.7) *with $h = n^{-1/(2k+m)}$. If condition* (3.13) *is satisfied, then*

$$R_n - R(G) = O\Big(n^{-\frac{2(\lambda k-1)}{2k+m}}\Big),$$

where $k \geq 2$ is a given natural number and m is the dimension of the vector Y.

4.3. The two-way ANOVA model

In model (2.1), let $q = b+1$, $Z = (\mathbf{l}_a \otimes \mathbf{l}_b\ \mathbf{l}_a \otimes I_b)$ and $\theta = (\mu, \beta^\tau)^\tau$, with the other assumptions the same as described in section 4.2; then, model (2.1) becomes the following two-way ANOVA model:

$$\begin{aligned} Y_{m\times 1} &= Z_{m\times(b+1)}\ \theta_{(b+1)\times 1} + X_{m\times a}\ \alpha_{a\times 1} + e_{m\times 1} \\ &= (\mathbf{l}_a \otimes \mathbf{l}_b)\mu + (I_a \otimes \mathbf{l}_b)\alpha + (\mathbf{l}_a \otimes I_b)\beta + e \qquad (4.10)\\ &= (\mathbf{l}_a \otimes \mathbf{1}_b)\mu + [I_a \otimes \mathbf{l}_b\, \mathbf{l}_a \otimes I_b]\gamma + e, \qquad (4.10a) \end{aligned}$$

where $\gamma = (\alpha^\tau, \beta^\tau)^\tau$, the vectors $\alpha = (\alpha_1,\ldots,\alpha_a)^\tau$ and $\beta = (\beta_1,\ldots,\beta_b)^\tau$ satisfying the constraint $\sum_{i=1}^{a}\alpha_i = 0$ and $\sum_{j=1}^{b}\beta_j = 0$ are the treatment effects of factors A and B respectively.

The Bayes estimator and EB estimator of α given by (2.9) and (3.11) become, respectively,

$$\begin{aligned} \phi_G(y) &= (X^\tau X)^{-1}X^\tau[y - Z\theta + \sigma^2\psi(y)] \\ &= (X^\tau X)^{-1}X^\tau[y - (\mathbf{l}_a \otimes \mathbf{l}_b)\mu - (\mathbf{l}_a \otimes I_b)\beta + \sigma^2\psi(y)], \qquad (4.11)\\ \phi_n(y) &= (X^\tau X)^{-1}X^\tau[y - Z\widehat{\theta} + \widehat{\sigma}^2\psi_n(y)] \\ &= (X^\tau X)^{-1}X^\tau[y - (\mathbf{l}_a \otimes \mathbf{l}_b)\widehat{\mu} - (\mathbf{l}_a \otimes I_b)\widehat{\beta} + \widehat{\sigma}^2\psi_n(y)], \qquad (4.12) \end{aligned}$$

where Z and X are given in (4.10), $\psi(y)$ and $\psi_n(y)$ are expressed by (2.10) and (3.8), $(\widehat{\mu}, \widehat{\beta}^\tau)^\tau = \widehat{\theta}$ and $(\widehat{\mu}_{(l)}, \widehat{\beta}^\tau_{(l)})^\tau = \widehat{\theta}_{(l)}$ are defined by (3.10) in which $D =$

$[(\mathbf{l}_a \otimes \mathbf{l}_b)/(ab),\ (\mathbf{l}_a \otimes I_b)/a - (\mathbf{l}_a \otimes \mathbf{l}_b)\mathbf{l}_b{}^\tau/(ab)]^\tau$, i.e., $\widehat{\mu}$ and $\widehat{\mu}_{(l)}$ are given by (4.8) and

$$\widehat{\beta} = \frac{1}{n}\sum_{l=1}^{n}\widehat{\beta}_{(l)}, \qquad \widehat{\beta}_{(l)} = \Big[\frac{1}{a}(\mathbf{l}_a \otimes I_b)^\tau - \frac{1}{ab}\mathbf{l}_b(\mathbf{l}_a \otimes \mathbf{l}_b)^\tau\Big]Y^{(l)}, \tag{4.13}$$

and $\widehat{\sigma}^2$ is given by (3.9) in which $\widehat{\sigma}^2_{(l)}$ is defined by (3.9a) with $s = (a-1)(b-1)$ and $H = I_a \otimes I_b - (I_a \otimes J_b)/b - (J_a \otimes I_b)/a + (J_a \otimes J_b)/(ab)$, $J_p = \mathbf{l}_p \otimes \mathbf{l}_p^\tau$, i.e.,

$$\widehat{\sigma}^2_{(l)} = \frac{1}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}(Y^{(l)}_{ij} - \overline{Y}^{(l)}_{i.} - \overline{Y}^{(l)}_{.j} + \overline{Y}^{(l)}_{..})^2, \tag{4.14}$$

with $\overline{Y}^{(l)}_{..}$ and $\overline{Y}^{(l)}_{i.}$ as given in (4.8) and (4.9), and $\overline{Y}^{(l)}_{.j} = \sum_{i=1}^{a} Y^{(l)}_{ij}/a$. It is obvious that $\widehat{\sigma}^2_{(l)}$ $(l = 1, \dots, n)$ are i.i.d. and $(a-1)(b-1)\widehat{\sigma}^2_{(l)}/\sigma^2 \sim \chi^2_{(a-1)(b-1)}$.

By Theorem 3.1, we get the convergence rates of EB estimators for the two-way ANOVA model as follows.

Subtheorem 4.3. *Let $\phi_G(y)$ be defined by* (4.11), *and $\phi_n(y)$ be as given in* (4.12) *with $h = n^{-1/(2k+m)}$. If condition* (3.13) *is satisfied, then*

$$R_n - R(G) = O\Big(n^{-\frac{2(\lambda k-1)}{2k+m}}\Big),$$

where $k \geq 2$ is a given natural number and m is the dimension of the vector Y.

Remark 4.2. In model (4.10), if we want to get the EB estimator for parameter vector β, when α is considered as nuisance parameters since $\alpha = (\alpha_1, \dots, \alpha_a)^\tau$ is a vector, then we can exchange the estimated status of α for β in section 4.3; this case is also a special form of model (2.1). Therefore, similar to section 4.3 we can construct the EB estimator of β and obtain the conclusion about convergence rates like Subtheorem 4.3.

Remark 4.3. In the two-way ANOVA model (4.10a), if we want to get the EB estimator for parameter vector $\gamma = (\alpha^\tau, \beta^\tau)^\tau$, it is easy to see that this is still a special case of model (2.1). In model (4.5), replacing α and $X = I_a \otimes \mathbf{l}_b$, with γ and $X = [I_a \otimes \mathbf{l}_b, \mathbf{l}_a \otimes I_b]$; then similar to section 4.2 we can get the EB estimator of γ and obtain the result about convergence rates like Subtheorem 4.2.

4.4. The ANOCOVA model

In model (2.1), assume that the elements of X are integers, usually 0 or 1, and the elements of Z are values take on by observations on continuous variables (also called concomitant variables). Let $\theta = \beta$, then model (2.1) becomes the following ANOCOVA model:

$$Y_{m\times 1} = X_{m\times a}\alpha_{a\times 1} + Z_{m\times q}\beta_{q\times 1} + e_{m\times 1}, \tag{4.15}$$

where the components of β are the regression coefficients, and the components of α denote the effects in the corresponding ANOVA problem derived from the ANOCOVA model. Suppose that $R(Z) = q$ and

$$\mu(X) \cap \mu(Z) = \{0\}, \tag{4.16}$$

where $\mu(B)$ stands for a generating linear subspace by the columns of matrix B.

The Bayes estimator and the EB estimator of α given by (2.9) and (3.11) become, respectively

$$\phi_G(y) = (X^\tau X)^{-1} X^\tau [y - Z\beta + \sigma^2 \psi(y)], \tag{4.17}$$

$$\phi_n(y) = (X^\tau X)^{-1} X^\tau [y - Z\widehat{\beta} + \widehat{\sigma}^2 \psi_n(y)], \tag{4.18}$$

where X and Z are described as in (4.15), $\psi(y)$ and $\psi_n(y)$ are expressed by (2.10) and (3.8), $\widehat{\beta} = \widehat{\theta}$ and $\widehat{\beta}_{(l)} = \widehat{\theta}_{(l)}$ are defined by (3.10) in which $D = (Z^\tau N Z)^{-1} Z^\tau N$, i.e.,

$$\widehat{\beta} = \frac{1}{n} \sum_{l=1}^{n} \widehat{\beta}_{(l)}, \qquad \widehat{\beta}_{(l)} = (Z^\tau N Z)^{-1} Z^\tau N Y^{(l)}, \tag{4.19}$$

and $\widehat{\sigma}^2$ is given by (3.9) in which $\widehat{\sigma}^2_{(l)}$ is defined by (3.9a) with $s = m - a - q$ and $H = N - NZ(Z^\tau N Z)^{-1} Z^\tau N$, i.e.,

$$\widehat{\sigma}^2_{(l)} = \frac{1}{m - a - q} \Big[Y^{(l)\tau} N Y^{(l)} - (Z^\tau N Y^{(l)})^\tau (Z^\tau N Z)^{-1} (Z^\tau N Y^{(l)}) \Big] \tag{4.20}$$

with $N = I - X(X^\tau X)^{-1} X^\tau$. Under assumption (4.16) we know that $Z^\tau N Z$ is an invertible matrix. It is easy to see that $\widehat{\sigma}^2_{(l)}$ $(l = 1, \ldots, n)$ are i.i.d. and $(m - a - q)\widehat{\sigma}^2_{(l)}/\sigma^2 \sim \chi^2_{m-a-q}$.

By Theorem 3.1, we obtain the following convergence rates of EB estimators for the ANOCOVA model.

Subtheorem 4.4. *Let $\phi_G(y)$ be defined by* (4.17)*, and $\phi_n(y)$ be as given in* (4.18) *with $h = n^{-1/(2k+m)}$. If condition* (3.13) *is satisfied, then*

$$R_n - R(G) = O\Big(n^{-\frac{2(\lambda k - 1)}{2k+m}}\Big),$$

where $k \geq 2$ is a given natural number and m is the dimension of the vector Y.

Acknowledgement

The author wishes to thank the referees and the Editor for their helpful suggestions and comments which greatly improved the presentation. This research was supported by the National Natural Science Foundation of China, and the Doctoral Program Foundation of the Institute of High Education.

Appendix A: The Lemmas

In order to get the Bayes estimator of α in section 2 and to prove Theorem 3.1, we need the following lemmas.

In the Appendices, suppose that c, $c_0, c_1, \ldots$ always stand for positive constants and they may denote different values even within the same expression.

Lemma A.1. *Let* $f(y|\alpha)$ *and* $f(y)$ *be given by* (2.4) *and* (2.5) *respectively. Then* $f(y)$ *has continuous* r*th order mixed partial derivatives*

$$f^{(r)}(y) = f^{(r)}(r_1, \ldots, r_m; y) = \frac{\partial^r f(y)}{\partial y_1^{r_1} \ldots \partial y_m^{r_m}}, \tag{A.1}$$

$$(r = \sum_{i=1}^{m} r_i, \quad r_i \geq 0, \quad i = 1, \ldots, m, \quad 0 \leq r \leq k)$$

which satisfies

$$f^{(r)}(y) = \int_\Theta f^{(r)}(y|\alpha) dG(\alpha), \tag{A.2}$$

where $k \geq 1$ *is natural number.*

Proof. It is obvious that $f^{(r)}(y|\alpha)$ exists, is continuous, and can be written in the form of exponential family. By Chen (1981), Theorem 1.2.1, $f^{(r)}(y)$ exists, is continuous, and has the expression (A.2). This lemma is proved.

Lemma A.2. *Suppose that* $R(G) < \infty$. *Let* ϕ_n *be an arbitrary a-vector statistic, then*

$$R_n - R(G) = E_* \|\phi_n - \phi_G\|^2.$$

Proof. See Lemma 4.1 of Singh (1985).

Lemma A.3. *Let* X, X' *be random variables,* x, x' *be real numbers, and* $L > 0$ *is a constant; then for* $0 < r \leq 2$ *we have*

$$E\Big|\Big[\frac{X'}{X} - \frac{x'}{x}\Big]_L\Big|^r \leq 2|x|^{-r}\Big\{E|X' - x'|^r + \Big(L + \Big|\frac{x'}{x}\Big|\Big)^r E|X - x|^r\Big\}.$$

Proof. See Lemma 3 of Zhao (1981); it is similar to Lemma 4.1 of Singh (1979).

Lemma A.4. *Let* $\phi_G(y)$ *and* $\psi(y)$ *be defined by* (2.9) *and* (2.10) *respectively. If* $\int_\Theta \|\alpha\|^\delta dG(\alpha) < \infty$ *for* $\delta \geq 1$, *then*

(i) $E_* \|\phi_G(Y)\|^\delta < \infty$,

(ii) $E_* \|\psi(Y)\|^\delta < \infty$.

Proof. By Jensen's inequality, we get

$$\begin{aligned} E_*\|\phi_G(Y)\|^\delta &= \int_{R_m} \|E(\alpha|y)\|^\delta f(y)dy \le \int_{R_m} E(\|\alpha\|^\delta|y)f(y)dy \\ &= \int_{R_m}\int_\Theta \|\alpha\|^\delta f(y|\alpha)dG(\alpha)dy = \int_\Theta \|\alpha\|^\delta dG(\alpha) < \infty. \end{aligned}$$

From (2.6) we know that $\psi(y) = \frac{\partial f(y)}{\partial y}/f(y) = -(y - Z\theta)/\sigma^2 + E(X\alpha|y)/\sigma^2$, and noting that $\sigma^2 \ge \gamma_0 > 0$, we have

$$\begin{aligned} E_*\|\psi(Y)\|^\delta &\le \gamma_0^{-\delta} E_*\|E(X\alpha|y) - (Y - Z\theta)\|^\delta \\ &\le c_1 E_*\|E(X\alpha|y)\|^\delta + c_2 E_*\|Y - Z\theta\|^\delta = c_1P_1 + c_2P_2. \end{aligned}$$

Let $\lambda = \max[\mathrm{root}(X^\tau X)]$, then $\|X\alpha\|^2 \le \lambda\|\alpha\|^2$. By Jensen's inequality we get

$$\begin{aligned} P_1 &= E_*\|E(X\alpha|y)\|^\delta \le E_*[E(\|X\alpha\|^\delta|y)] \\ &\le \lambda^{\delta/2}\int_{R_m}\int_\Theta \|\alpha\|^\delta f(y|\alpha)dG(\alpha)dy = \lambda^{\delta/2}\int_\Theta \|\alpha\|^\delta dG(\alpha) < \infty, \end{aligned}$$

and

$$\begin{aligned} P_2 &= E_*\|Y - Z\theta\|^\delta \\ &\le c_{21}\int_{R_m}\int_\Theta \|y - Z\theta - X\alpha\|^\delta f(y|\alpha)dG(\alpha)dy \\ &\quad + c_{22}\int_{R_m}\int_\Theta \|X\alpha\|^\delta f(y|\alpha)dG(\alpha)dy = c_{21}P_{21} + c_{22}P_{22}. \end{aligned}$$

By the fact that $E\|Y-\eta\|^\delta < \infty$ if $Y \sim N(\eta, \sum)$, we know that $P_{21} < \infty$. Similar to the proof of P_1, we have $P_{22} < \infty$, hence $P_2 < \infty$. This lemma is proved.

Lemma A.5. *Let $f^{(r)}(y)$ be given by* (A.1) *and $f_n^{(r)}(y)$ be as defined by* (3.4) *with $h = n^{-1/(2k+m)}$, $r = 0, 1$; then for $0 < \lambda \le 1$ we have*

$$E|f_n^{(r)}(y) - f^{(r)}(y)|^{2\lambda} \le cn^{-\frac{2\lambda(k-r)}{2k+m}}\Big[f^{2\lambda}(y)A^{2\lambda}(y) + f^\lambda(y)B^\lambda(y)\Big],$$

where

$$A(y) = E_{(\alpha|Y)}\Big\{\sum_{l=1}^k \Big[\|v\|^l + (mh)^l\Big]\exp(mh\|v\|/\sigma^2) \mid y\Big\}, \tag{A.3}$$

$$B(y) = E_{(\alpha|Y)}[\exp(mh\|v\|/\sigma^2)|y], \tag{A.4}$$

$v = y - Z\theta - X\alpha$, $k \ge 2$ is a given natural number and $E_{(\alpha|Y)}$ denotes the conditional expectation of α given Y.

Proof. Similar to the proof of Theorem 5.1 of Singh (1985).

Lemma A.6. *Let* $f(y)$ *be given by* (2.5)*, and* ξ *is an arbitrarily small positive number. If for* $\delta = [(m+\xi)\lambda/(1-\lambda)] \vee 1$ *with* $0 < \lambda < 1$ *such that* $\int_\Theta \|\alpha\|^\delta dG(\alpha) < \infty$*, then*

$$\int_{R_m} (f(y))^{1-\lambda} dy < \infty.$$

Proof. For $0 < \lambda < 1$ we have

$$\begin{aligned}&\int_{R_m} (f(y))^{1-\lambda} dy \\ &= \int_D (f(y))^{1-\lambda}\, dy + \int_{D^c} \|y - Z\theta\|^{-(m+\xi)\lambda} [\|y - Z\theta\|^{(m+\xi)\lambda} (f(y))^{1-\lambda}] dy \\ &= J_1 + J_2,\end{aligned}$$

where $D = \{y : \|y - Z\theta\| \le m\}$ and $D^c = R_m - D$.

It is obvious that $J_1 = \int_D (f(y))^{1-\lambda} dy < \infty$. By Holder's inequality, we have

$$\begin{aligned}J_2 &\le \Big(\int_{D^c} \|y - Z\theta\|^{-(m+\xi)} dy\Big)^{\lambda} \Big(\int_{D^c} \|y - Z\theta\|^{\frac{(m+\xi)\lambda}{1-\lambda}} f(y) dy\Big)^{1-\lambda} \\ &= J_{21}^{\lambda} J_{22}^{1-\lambda}.\end{aligned}$$

It is obvious that $J_{21} < \infty$, and

$$J_{22} \le E_* \|Y - Z\theta\|^{\frac{(m+\xi)\lambda}{1-\lambda}} \le E_* \|Y - Z\theta\|^{\delta}.$$

Similar to the proof of $P_2 < \infty$ in Lemma A.4, we know that $J_{22} < \infty$. Therefore $J_2 = J_{21}^{\lambda} J_{22}^{1-\lambda} < \infty$. This lemma is proved.

Lemma A.7. *Let* $A(y)$ *and* $B(y)$ *be given in Lemma* A.5*, and* ξ *is an arbitrarily small positive number. If for* $\delta = [(m+\xi)\lambda/(\eta - \lambda)] \vee 1$ *with* $0 < \lambda < \eta < 1$ *such that*

$$\int_\Theta \|\alpha\|^\delta dG(\alpha) < \infty, \tag{A.5}$$

then

$$\int_{R_m} [A^{2\lambda}(y) f(y) + B^{\lambda}(y) (f(y))^{1-\lambda}] dy < \infty.$$

Proof. From formula (A.3), we know that

$$\begin{aligned}J_0 &= \int_{R_m} A^{2\lambda}(y) f(y) dy = E_*[A^{2\lambda}(Y)] \\ &= E_* \Big\{ E_{(\alpha|Y)} \Big[\sum_{l=1}^{k} (\|v\|^l + (mh)^l) \exp(mh\|v\|/\sigma^2) |y \Big] \Big\}^{2\lambda},\end{aligned} \tag{A.6}$$

where $v = y - Z\theta - X\alpha$, defined as in Lemma A.5.

If $2\lambda \geq 1$, then by (A.6) and Jensen's inequality for convex functions we have

$$J_0 \leq cE_*\Big\{E_{(\alpha|Y)}\Big[\sum_{l=1}^{k}[\|v\|^{2\lambda l} + (mh)^{2\lambda l}]\exp(2\lambda mh\|v\|/\sigma^2)|y\Big]\Big\}$$
$$= c\sum_{l=1}^{k}\int_{\Theta}\int_{R_m}[\|v\|^{2\lambda l} + (mh)^{2\lambda l}]\exp\{2\lambda mh\|v\|/\sigma^2\}f(y|\alpha)dG(\alpha)dy. \quad \text{(A.7)}$$

If $0 < 2\lambda < 1$, then by (A.6) and Jensen's inequality for concave functions we get

$$J_0 = E_*[A^{2\lambda}(Y)] \leq \Big\{E_*[A(Y)]\Big\}^{2\lambda} = J^{2\lambda},$$

where

$$J = \sum_{l=1}^{k}\int_{\Theta}\int_{R_m}[\|v\|^{l} + (mh)^{l}]\exp\{mh\|v\|/\sigma^2\}f(y|\alpha)dG(\alpha)dy. \quad \text{(A.8)}$$

It is easy to see that

$$\int_{\Theta}\int_{R_m}\|v\|^{r}\exp\{cmh\|v\|/\sigma^2\}f(y|\alpha)dG(\alpha)dy < \infty, \quad \text{(A.9)}$$
$$\int_{\Theta}\int_{R_m}(mh)^{r}\exp\{cmh\|v\|/\sigma^2\}f(y|\alpha)dG(\alpha)dy < \infty, \quad \text{(A.10)}$$

where $r \geq 0$. From (A.6) to (A.10) we have

$$\int_{R_m}A^{2\lambda}(y)f(y)dy < \infty. \quad \text{(A.11)}$$

By Holder's inequality we get

$$\int_{R_m}B^{\lambda}(y)(f(y))^{1-\lambda}dy = \int_{R_m}[B^{\lambda}(y)(f(y))^{1-\eta}](f(y))^{\eta-\lambda}dy$$
$$\leq \Big[\int_{R_m}(B(y))^{\frac{\lambda}{1-\eta}}f(y)dy\Big]^{1-\eta}\Big[\int_{R_m}(f(y))^{1-\frac{\lambda}{\eta}}dy\Big]^{\eta} = Q_1^{1-\eta}Q_2^{\eta},$$

where $0 < \lambda < \eta < 1$. Similar to the proof of (A.11) we get

$$Q_1 = \int_{R_m}(B(y))^{\frac{\lambda}{1-\eta}}f(y)dy < \infty. \quad \text{(A.12)}$$

Since $0 < \lambda/\eta < 1$, by (A.5) and Lemma A.6 we have

$$Q_2 = \int_{R_m}(f(y))^{1-\frac{\lambda}{\eta}}dy < \infty. \quad \text{(A.13)}$$

Therefore

$$\int_{R_m} B^{\lambda}(y)(f(y))^{1-\lambda}dy \le Q_1^{1-\eta}Q_2^{\eta} < \infty. \tag{A.14}$$

By (A.11) and (A.14), this lemma is proved.

Appendix B: Proof of Theorem 3.1

Let $\lambda_0 = \max[\text{root}(X(X^\tau X)^{-2}X^\tau)]$. By Lemma A.4, it is obvious that $R(G) < \infty$, therefore from formulas (2.9), (3.11) and Lemma 4.2 we have

$$\begin{aligned} R_n - R(G) &= E_*\|\phi_n(Y) - \phi_G(Y)\|^2 \\ &= E_*\|(X^\tau X)^{-1}X^\tau[(\widehat{\sigma}^2\psi_n(Y) - \sigma^2\psi(Y)) - Z(\widehat{\theta} - \theta)]\|^2 \\ &\le 4[E_*\|(X^\tau X)^{-1}X^\tau Z(\widehat{\theta} - \theta)\|^2 + \lambda_0 E_*\|(\widehat{\sigma}^2 - \sigma^2)\psi(Y)\|^2 \\ &\quad + \lambda_0 E_*\|\widehat{\sigma}^2(\psi_n(Y) - \psi(Y))\|^2] = 4[J_1 + J_2 + J_3]. \end{aligned} \tag{B.1}$$

Let $\lambda_1 = \max[\text{root}(Z^\tau X(X^\tau X)^{-2}X^\tau Z)]$ and $\overline{\lambda}_1 = \max[\text{root}(X^\tau D^\tau DX)]$. Since $\widehat{\theta}_{(l)} = DY^{(l)}$ $(l = 1, \ldots, n)$, given by (3.10), are i.i.d. and unbiased by condition (3.13) and the fact that $\sigma^2 \le M_0$ we have

$$\begin{aligned} J_1 &= E_*\|(X^\tau X)^{-1}X^\tau Z(\widehat{\theta} - \theta)\|^2 \le \lambda_1 E_*\|\widehat{\theta} - \theta\|^2 = \frac{\lambda_1}{n}E_*\|\widehat{\theta}_{(1)} - \theta\|^2 \\ &= \frac{\lambda_1}{n}\text{tr}[D\,\text{Cov}\,(Y^{(1)})D^\tau] = \frac{\lambda_1}{n}\text{tr}[D^\tau D(\sigma^2 I_m + X\,\text{Cov}\,(\alpha)X^\tau)] \\ &\le cn^{-1}(M_0\text{tr}(D^\tau D) + \overline{\lambda}_1 E_*\|\alpha\|^2) \le c_1 n^{-1}, \end{aligned} \tag{B.2}$$

where $\text{tr}(A)$ denotes trace of the matrix A.

From (3.9) we know that $(\widehat{\sigma}^2 - \sigma^2)^2 \le (\widehat{\sigma}_n^2 - \sigma^2)^2$, $E(\widehat{\sigma}_n^2) = \sigma^2$ and $\text{Var}\,(\widehat{\sigma}_{(1)}^2) = 2\sigma^4/s$; therefore by Lemma A.4 and the fact that $\sigma^2 \le M_0$, we get

$$\begin{aligned} J_2 &= \lambda_0 E_*\|(\widehat{\sigma}^2 - \sigma^2)\psi(Y)\|^2 \le \lambda_0 E_{(Y,\alpha)}[\|\psi(Y)\|^2 E(\widehat{\sigma}_n^2 - \sigma^2)^2] \\ &= \lambda_0 E_{(Y,\alpha)}[\|\psi(Y)\|^2 \text{Var}\,(\widehat{\sigma}_n^2)] \le \frac{\lambda_0}{n}E_{(Y,\alpha)}[\|\psi(Y)\|^2\text{Var}\,(\widehat{\sigma}_{(1)}^2)] \\ &= \frac{2\lambda_0}{ns}\sigma^4 E_*\|\psi(Y)\|^2 \le \frac{2\lambda_0}{ns}M_0^2 E_*\|\psi(Y)\|^2 \le c_2 n^{-1}. \end{aligned} \tag{B.3}$$

Noting that $\widehat{\sigma}^2 \le M_0$ in the formula (3.9), we have

$$\begin{aligned} J_3 &= \lambda_0 E_*\|\widehat{\sigma}^2(\psi_n(Y) - \psi(Y))\|^2 \le \lambda_0 M_0^2 E_*\|\psi_n(Y) - \psi(Y)\|^2 \\ &\le c_0 \sum_{i=1}^m E_*\Big(\psi_{n(i)}(Y) - \psi_{(i)}(Y)\Big)^2 = c_0 \sum_{i=1}^m Q_i. \end{aligned} \tag{B.4}$$

First consider $Q_1 = E_*(\psi_{n(1)}(Y) - \psi_{(1)}(Y))^2 = E_{(Y,\alpha)}[E(\psi_{n(1)}(y) - \psi_{(1)}(y))^2]$. Suppose that $A_{n(1)} = \{y : y \in R_m,\ |\psi_{(1)}(y)| \le n^\nu/2\}$, $B_{n(1)} = R_m - A_{n(1)}$. If

$y \in A_{n(1)}$ then $|\psi_{n(1)} - \psi_{(1)}| \le 3n^\nu/2$; therefore by Lemma A.3 and Lemma A.5 we obtain

$$\begin{aligned}
& E(\psi_{n(1)}(y) - \psi_{(1)}(y))^2 \\
& \le \Big(\frac{3}{2}n^\nu\Big)^{2-2\lambda} E\Big|\Big[\frac{f'_{n(1)}(y)}{f_n(y)} - \frac{f'_{(1)}(y)}{f(y)}\Big]_{\frac{3}{2}n^\nu}\Big|^{2\lambda} \\
& \le cn^{2\nu(1-\lambda)}\Big(f(y)\Big)^{-2\lambda}\Big\{E|f'_{n(1)}(y) - f'_{(1)}(y)|^{2\lambda} + (2n^\nu)^{2\lambda}E|f_n(y) - f(y)|^{2\lambda}\Big\} \\
& \le cn^{-(\frac{2\lambda k}{2k+m}-2\nu)}[A^{2\lambda}(y) + B^\lambda(y)f^{-\lambda}(y)]. \qquad \text{(B.5)}
\end{aligned}$$

By Lemma A.7 we have

$$\begin{aligned}
& \int_{A_{n(1)}} E(\psi_{n(1)}(y) - \psi_{(1)}(y))^2 f(y)dy \\
& \le cn^{-(\frac{2\lambda k}{2k+m}-2\nu)}\Big[\int_{R_m} A^{2\lambda}(y)f(y)dy + \int_{R_m} B^\lambda(y)(f(y))^{1-\lambda}dy\Big] \\
& \le cn^{-(\frac{2\lambda k}{2k+m}-2\nu)}. \qquad \text{(B.6)}
\end{aligned}$$

If $y \in B_{n(1)}$ then $|\psi_{(1)}(y)| > n^\nu/2$, therefore we get $(\psi_{n(1)}(y) - \psi_{(1)}(y))^2 \le 2n^{2\nu} + 2\psi^2_{(1)}(y) \le 10\psi^2_{(1)}(y)$; thus by Holder's inequality, Markov's inequalities and Lemma A.4 we have

$$\begin{aligned}
& \int_{B_{n(1)}} E(\psi_{n(1)}(y) - \psi_{(1)}(y))^2 f(y)dy \\
& \le 10E_*\Big\{\psi^2_{(1)}(Y)I_{[|\psi_{(1)}(Y)|>\frac{1}{2}n^\nu]}\Big\} \\
& \le 10\Big\{E_*|\psi_{(1)}(Y)|^\delta\Big\}^{2/\delta}\Big\{2^\delta n^{-\nu\delta}E_*|\psi_{(1)}(Y)|^\delta\Big\}^{(\delta-2)/\delta} \\
& \le 10\Big\{E_*\|\psi(Y)\|^\delta\Big\}^{2/\delta}\Big\{2^\delta n^{-\nu\delta}E_*\|\psi(Y)\|^\delta\Big\}^{(\delta-2)/\delta} \\
& \le cn^{-\nu(\delta-2)}. \qquad \text{(B.7)}
\end{aligned}$$

Let $\delta = 2\lambda k$ with $1/2 < \lambda < 1$ and $k \ge 2$; hence $\delta > 2$. Put $\nu(2\lambda k - 2) = 2\lambda k/(2k+m) - 2\nu$; then we obtain $\nu = 1/(2k+m)$. Therefore by (B.6) and (B.7) we have

$$\begin{aligned}
Q_1 &= E_*(\psi_{n(1)}(Y) - \psi_{(1)}(Y))^2 \\
&= \int_{A_{n(1)}} E(\psi_{n(1)}(y) - \psi_{(1)}(y))^2 f(y)dy + \int_{B_{n(1)}} E(\psi_{n(1)}(y) - \psi_{(1)}(y))^2 f(y)dy \\
&\le cn^{-\frac{2(\lambda k-1)}{2k+m}}. \qquad \text{(B.8)}
\end{aligned}$$

Secondly, similar to the proof of Q_1 we have

$$Q_i \le cn^{-\frac{2(\lambda k-1)}{2k+m}} \qquad i = 2, \ldots, m. \qquad \text{(B.9)}$$

Substituting (B.8) and (B.9) into (B.4) we get

$$J_3 \leq c_0 \sum_{i=1}^{m} Q_i \leq c_3 n^{-\frac{2(\lambda k-1)}{2k+m}}. \tag{B.10}$$

Substituting (B.2), (B.3) and (B.10) into (B.1) we obtain

$$R_n - R(G) \leq 4\Big[c_1 n^{-1} + c_2 n^{-1} + c_3 n^{-\frac{2(\lambda k-1)}{2k+m}}\Big] \leq c n^{-\frac{2(\lambda k-1)}{2k+m}}.$$

This means $R_n - R(G) = O(n^{-2(\lambda k-1)/(2k+m)})$. The theorem is proved.

References

Chen, Xiru (1981). *An Introduction to Mathematical Statistics.* Science Press, Beijing.

Lu, K. L. (1982). Kernel estimators of mixed partial derivatives of a multivariate density and their convergence rate. *J. Systems Sci. Math. Sci.* **2**, 220-226.

Robbins, H. (1955). An empirical Bayes approach to statistics, Proc. Third Berkeley Symp. Math. Statist. Prob. 1, 157-163, Univ. California Press.

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

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

Singh, R. S. (1981). Speed of convergence in nonparametric estimation of a multivariate μ density and its mixed partial derivatives. *J. Statist. Plann. Inference* **5**, 287-298.

Singh, R. S. (1985). Empirical Bayes estimation in a multiple linear regression model. *Ann. Inst. Statist. Math.* **37**, 71-86.

Wei, L. S. (1990). Empirical Bayes test of regression coefficient in a multiple linear regression model. *Acta Math. Appl. Sinica* **6**, 251-262.

Wei, L. S. and Zhang, S. P. (1995). The convergence rates of empirical Bayes estimation in a multiple linear regression model. *Ann. Inst. Statist. Math.* **47**, 81-97.

Wei, L. S. (1995). The asymptotically optimal empirical Bayes estimation in one way ANOVA model. To appear.

Zhang, S. P. and Wei, L. S. (1994). The asymptotically optimal empirical Bayes estimation in a multiple linear regression model. *Appl. Math. -A Journal of Chinese Universities* **9**, 245-258.

Zhao, L. C. (1981). Empirical Bayes estimation with convergence rates about a class of discrete distribution families. *J. Math. Res. Exposition* **1**, 59-69.

Department of Statistics and Finance, University of Science and Technology of China, Hefei, Anhui 230026.

E-mail: lwei@math.ustc.edu.cn

(Received January 1996; accepted March 1997)