Ann. Inst. Statist. Math.
37 (1985), Part A, 71-86

EMPIRICAL BAYES ESTIMATION IN A MULTIPLE LINEAR REGRESSION MODEL*

R. S. SINGH

(Received Feb. 14, 1983)

Summary

Estimation of the vector $\boldsymbol{\beta}$ of the regression coefficients in a multiple linear regression $\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$ is considered when $\boldsymbol{\beta}$ has a completely unknown and unspecified distribution and the error-vector $\boldsymbol{\varepsilon}$ has a multivariate standard normal distribution. The optimal estimator for $\boldsymbol{\beta}$, which minimizes the overall mean squared error, cannot be constructed for use in practice. Using $\boldsymbol{X}$, $\boldsymbol{Y}$ and the information contained in the observation-vectors obtained from n independent past experiences of the problem, (empirical Bayes) estimators for $\boldsymbol{\beta}$ are exhibited. These estimators are compared with the optimal estimator and are shown to be asymptotically optimal. Estimators asymptotically optimal with rates near $O(n^{-1})$ are constructed.

1. Introduction

Empirical Bayes (EB) approach to a statistical problem, introduced by Robbins [22], and later developed by Johns [7], Robbins [23], [24], Samuel [25] and Johns and Van Ryzin [8], [9], among others, has recently drawn considerable attention in the literature. Suppose there is a pair (X, θ) of random variables, where X is observable and θ, called the parameter, is unobservable. The r.v. X given θ has a specified distribution P_θ on the observation space $\mathcal{X}$ and θ has a completely *unknown* and *unspecified* distribution G on the parameter space Θ. Based on an observation on X (which could be a sufficient statistic for θ), the problem is to decide about θ with respect to a nonnegative loss function. If the prior distribution G were known, the statistician would use the *Bayes procedure* ψ_G which achieves the *minimum Bayes risk*, say $R(G)$, with respect to G. (We assume such a ψ_G exists). But since

* Supported in part by a Natural Sciences and Engineering Research Council of Canada grant.

Key words and phrases: Empirical Bayes, estimation, multiple regression model, squared error loss, asymptotically optimal, rates.

G is not known, the *optimal procedure* ψ_G is not available for the use to the statistician. In the EB context we assume that the above problem, called the *component problem*, has occurred independently in the past, say n times, so that there are $n+1$ independent pairs $(X_1, \theta_1), \cdots, (X_n, \theta_n)$ and (X, θ). The object is to utilize the information contained in the past observations $(X_1, \cdots, X_n)$ and the present observation X to produce a decision rule for the present parameter θ so that for large n this rule is "nearly" as good as the unavailable optimal rule ψ_G in the sense that the overall risk, say R_n, of this rule approximates the *Bayes envelope* $R(G)$ achieved by ψ_G. If $R_n \rightarrow R(G)$ as $n \rightarrow \infty$ then the rule is called asymptotically optimal (a.o.), (Robbins [22]). If for a $\delta > 0$ $R_n - R(G) = O(n^{-\delta})$ as $n \rightarrow \infty$ we will say that the rule is a.o. with rate $O(n^{-\delta})$.

Johns and Van Ryzin [8], [9] considered EB linear loss two-action hypothesis testing problems in one parameter discrete [8] and continuous [9] exponential families and exhibited a.o. test procedures with rates near $O(n^{-1})$ in both cases. Lin [13], [14] considered EB squared error loss estimation (SELE) in the two families considered by Johns and Van Ryzin and obtained EB estimators with rates near $O(n^{-1/3})$. Yu [34] and Singh [27], [29] exhibited a.o. EB estimators in the one parameter Lebesgue exponential family with rates near $O(n^{-1/3})$, $O(n^{-2/5})$ and $O(n^{-1})$ respectively. O'Bryan and Susarla [20] considered EB estimation with varying sample sizes in component problems in the univariate normal distribution and obtained estimators a.o. with rates near $O(n^{-1/3})$. Recently Monte-Carlo simulations performed by Clemmer and Krutchkoff [2], Maritz [16], Martz and Krutchkoff [18], Griffin and Krutchkoff [5], Bennet and Martz [1] and Maritz and Lwin [17], among others, have shown how for certain priors the EB procedures often perform better compared to the usual procedures whenever there is at least one past experience of the problem.

We, in this paper, consider EB approach to the SELE problem in a multiple linear regression model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ with loss function $L(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}}) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$. Using the information contained in $(\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_n)$ from the past problems and $(\boldsymbol{Y}, \boldsymbol{X})$ from the present problem we will exhibit for each $n > 0$ two classes of estimators $\hat{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\phi}}$ for $\boldsymbol{\beta}$, one for the case when nothing is known about the support of the prior distribution and the other for the case when it is known that the prior distribution has a compact support. We show that $\hat{\boldsymbol{\phi}}$ are a.o. with rates $O(n^{-1+\eta})$ uniformly over the class of all priors satisfying certain moment conditions dependent on η, whereas $\tilde{\boldsymbol{\phi}}$ are shown to be a.o. with rates $O(n^{-1+\eta})$ uniformly over the class of all priors with compact support. Thus we have given procedures for constructing EB estimators a.o. with rates arbitrarily close to $O(n^{-1})$ in the above multiple linear regression model. As pointed out in Singh [29] EB procedures a.o. with rates $O(n^{-1})$ have not yet been exhibited in any EB problem other than that involving discrete distributions.

Martz and Krutchkoff [18] have considered EB estimation in the general linear model considered here. They have, however, not proved any consistency or asymptotic optimality of their estimators, but have shown, through a Monte-Carlo simulation, that for certain priors their EB estimators perform better than the usual least square estimator. Wind [33] has considered EB estimation of $\boldsymbol{\beta}$ when the error vector $\boldsymbol{\varepsilon}$ is assumed to have $\mathbf{0}$ mean and covariance $\sigma^2\boldsymbol{I}$ but is not assumed to take on a specific parametric form, e.g. Normal. For priors specifying their means and variances or specifying their second moments, he has exhibited restricted asymptotically optimal EB estimators of $\boldsymbol{\beta}$, that is estimators whose Bayes risks converge to the risk of the restricted minimax estimator at each component stage.

Some other important related contributions worth mentioning here are due to Stein [31], James and Stein [6], Cogburn [3], Kantor [10], Wind [32], Efron and Morris [4] and Rao [21], among others, where problems related to estimation of regression coefficient $\boldsymbol{\beta}$ or of a multivariate normal distribution mean are considered.

2. The model and preliminaries

We consider the following multiple linear regression model

$$Y=X\boldsymbol{\beta}+\boldsymbol{\varepsilon} \tag{2.1}$$

where $\boldsymbol{Y}$ is an $l\times 1$ vector of random observations, $\boldsymbol{X}$ is an $l\times p$ matrix of known constants so that $\boldsymbol{X}'\boldsymbol{X}$ is invertible, $\boldsymbol{\beta}$ is the $p\times 1$ vector of unknown regression coefficients having an *unknown* and *unspecified* prior distribution G on $(E^p, \mathscr{B}^p)$, $\mathscr{B}^p$ being the Borel-field of subsets of E^p, the p-dimentional Euclidean space, and $\boldsymbol{\varepsilon}$ is an $l\times 1$ vector of unobservable random variables. We will assume that the conditional distribution of $\boldsymbol{\varepsilon}$ given $\boldsymbol{\beta}$ is $N_p(\mathbf{0}, \sigma^2\boldsymbol{I})$, the p-variate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $\sigma^2\boldsymbol{I}$, where σ^2 is assumed to be known. The object is to estimate $\boldsymbol{\beta}$ with respect to the quadratic loss function

$$L(\boldsymbol{\beta}, \boldsymbol{\phi})=\|\boldsymbol{\beta}-\boldsymbol{\phi}\|_2^2 \tag{2.2}$$

where $\boldsymbol{\phi}$ is an estimate of $\boldsymbol{\beta}$, and for a $1\times p$ vector $\boldsymbol{t}'=(t_1,\cdots,t_p)$ and $m>0$, $\|\boldsymbol{t}\|_m=\left(\sum_{i=1}^{p}|t_i|^m\right)^{1/m}$.

For an estimator $\boldsymbol{\phi}$ of $\boldsymbol{\beta}$, let $R(\boldsymbol{\phi}, G)$ be the (Bayes) risk of $\boldsymbol{\phi}$ with respect to the prior distribution G, i.e.,

$$R(\boldsymbol{\phi}, G)=\mathrm{E}\,\|\boldsymbol{\phi}-\boldsymbol{\beta}\|_2^2 \tag{2.3}$$

where E stands for the expectation operator with respect to all the random variables involved in the expression. The *Bayes envelope* with respect to G is given by

$$R(G)=\inf_{\boldsymbol{\phi}} R(\boldsymbol{\phi}, G)$$

where the inf is taken over the set of all estimators $\boldsymbol{\phi}$ for which (2.3) exist. The estimator which achieves the Bayes envelope $R(G)$ is the *Bayes estimator*, also called *optimal estimator* (o.e.) $\boldsymbol{\phi}_G$ given by

$$\boldsymbol{\phi}_G(\boldsymbol{Y})=\mathrm{E}\,(\boldsymbol{\beta}|\boldsymbol{Y})\,. \tag{2.4}$$

Thus $R(\boldsymbol{\phi}_G, G)=R(G)$. Notice that $R(G)$ can be exactly achieved only if the prior distribution G is known and $\boldsymbol{\beta}$ is estimated by the o.e. $\boldsymbol{\phi}_G$. Unfortunately G is completely unknown, and hence $\boldsymbol{\phi}_G$ is not available to us for use. This motivates us to use the EB approach to exhibit estimators whose risks are close to $R(G)$ achived by $\boldsymbol{\phi}_G$.

3. Empirical Bayes approach

Suppose we have incurred n independent experiences of the above estimation problem, called the *component problem*, in the past. That is we have independent pairs

$$\{\boldsymbol{Y}_1, \boldsymbol{\beta}_1\}, \cdots, \{\boldsymbol{Y}_n, \boldsymbol{\beta}_n\}$$

from the past experiments, and $(\boldsymbol{Y}, \boldsymbol{\beta})$ from the present experiment, with

$$\boldsymbol{Y}_i=\boldsymbol{X}\boldsymbol{\beta}_i+\boldsymbol{e}_i\,, \qquad i=1, \cdots, n\,.$$

The vectors $\boldsymbol{Y}_i$, $\boldsymbol{\beta}_i$, $\boldsymbol{e}_i$ behave like $\boldsymbol{Y}$, $\boldsymbol{\beta}$, $\boldsymbol{e}$ described above; $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_n$ and $\boldsymbol{\beta}$, are i.i.d. according to the same unknown prior distribution G; and given $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_n$ and $\boldsymbol{\beta}$, the vectors $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$ are i.i.d. according to $N_p(\mathbf{0}, \sigma^2\boldsymbol{I})$ and $\boldsymbol{Y}_i \sim N_p(\boldsymbol{X}\boldsymbol{\beta}_i, \sigma^2\boldsymbol{I})$. We will exhibit (EB) estimators for $\boldsymbol{\beta}$ on the basis of $\boldsymbol{X}$, the past observations $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_n$ and the present observation $\boldsymbol{Y}$; and show that for large n these estimators can be considered as good as the unavailable o.e. $\boldsymbol{\phi}_G$.

In Section 4 we will show how the problem of approximation of the optimal estimator can be reduced to the problem of estimation of the marginal density of $\hat{\boldsymbol{\beta}}$, the maximum likelihood estimate of $\boldsymbol{\beta}$, and its first partial derivatives. In Section 5 we exhibit for each integer $r>1$ uniformly mean squared consistent estimators of the marginal density of $\hat{\boldsymbol{\beta}}$ and its partial derivatives. Using these estimates we will introduce two EB estimators $\hat{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\phi}}$ of $\boldsymbol{\beta}$, one for the case when nothing is known about the support of the prior distribution G of $\boldsymbol{\beta}$ and the other for the case when the support of G is a known compact subset of E^p. In Section 6 we will show that the excess risk of $\tilde{\boldsymbol{\phi}}$ over the minimum risk $R(G)$ is of order $n^{-(r-1)\gamma/(2r+p)}$ for every $0<\gamma<2$, and that of $\hat{\boldsymbol{\phi}}$ is of order $n^{-(r\gamma-2)/(2r+p)}$ for every γ in $(0, 2)$ for which $\int \|\boldsymbol{\beta}\|_1^{pr+/(2-r)} dG(\boldsymbol{\beta})$ is finite. Thus the rates of convergence can be sharpened to any desired degree by choosing appropriately a larger value of r.

4. Reduction in the problem

Recall from Section 1 that an EB estimator based on the observations from the past n experiences of the problem and the present observation is a.o. if its risk approaches to the minimum Bayes risk $R(G)$ as n gets large. The following known lemma therefore reduces the problem of searching an a.o. estimator for $\boldsymbol{\beta}$ to the one of searching a quadratic mean squared consistent estimator of the o.e. $\boldsymbol{\phi}_G$. The univariate version of the lemma is proved (under the present condition) in Singh [29].

LEMMA 4.1. *Let the prior distribution G be such that $R(G)<\infty$. Let $\boldsymbol{\phi}$ be an arbitrary p-vector statistic. Then*

$$R(\boldsymbol{\phi}, G)-R(G)=\mathrm{E}\,\|\boldsymbol{\phi}-\boldsymbol{\phi}_G\|_2^2$$

where the expectation is taken with respect to all the random variables involved in the expression.

PROOF. Let E_* stand for the expectation operator conditional on $\boldsymbol{Y}$ and all other random variables involved in $\boldsymbol{\phi}$. Then

$$\begin{aligned}\mathrm{E}_*\,[(\boldsymbol{\phi}-\boldsymbol{\beta})'(\boldsymbol{\phi}-\boldsymbol{\beta})]=(\boldsymbol{\phi}-\boldsymbol{\phi}_G)'(\boldsymbol{\phi}-\boldsymbol{\phi}_G)+\mathrm{E}_*\,\{&(\boldsymbol{\phi}_G-\boldsymbol{\beta})'(\boldsymbol{\phi}_G-\boldsymbol{\beta})\\ &+2(\boldsymbol{\phi}-\boldsymbol{\phi}_G)'(\boldsymbol{\phi}_G-\boldsymbol{\beta})\}\,.\end{aligned}$$

Since $R(G)=\mathrm{E}\,(\boldsymbol{\phi}_G-\boldsymbol{\beta})'(\boldsymbol{\phi}_G-\boldsymbol{\beta})<\infty$, the second term above is $\mathrm{E}_*\,(\boldsymbol{\phi}_G-\boldsymbol{\beta})' \cdot(\boldsymbol{\phi}_G-\boldsymbol{\beta})+2(\boldsymbol{\phi}-\boldsymbol{\phi}_G)'\,\mathrm{E}_*\,(\boldsymbol{\phi}_G-\boldsymbol{\beta})=\mathrm{E}_*\,\|\boldsymbol{\phi}_G-\boldsymbol{\beta}\|_2^2$ since $\mathrm{E}_*\,(\boldsymbol{\beta})=\boldsymbol{\phi}_G$. Thus

$$\mathrm{E}_*\,\|\boldsymbol{\phi}-\boldsymbol{\beta}\|_2^2=\|\boldsymbol{\phi}-\boldsymbol{\phi}_G\|_2^2+\mathrm{E}_*\,\|\boldsymbol{\phi}_G-\boldsymbol{\beta}\|_2^2\,.$$

Now taking expectations on both sides we get the conclusion of the lemma.

Since σ^2 is known, without loss of generality we can (and do) take $\sigma^2=1$. The least square estimator of $\boldsymbol{\beta}$ is

$$(4.1)\qquad \hat{\boldsymbol{\beta}}=\boldsymbol{\Sigma X'Y} \quad \text{where } \boldsymbol{\Sigma}=(X'X)^{-1}\text{.}$$

The estimator (4.1) is the maximum likelihood estimator (MLE) as well as the minimum variance unbiased estimator (MVUE) of $\boldsymbol{\beta}$.

To reduce the problem further we will now show that the o.e. in the EB context $\mathrm{E}(\boldsymbol{\beta}|\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_n,\boldsymbol{Y})$ is the same as the o.e. (2.4) in the component problem. Then we will show how the o.e. (2.4) can be expressed as a function of the marginal p.d.f. of $\hat{\boldsymbol{\beta}}$ and its first order partial derivatives.

Notice that the conditional distribution of $\hat{\boldsymbol{\beta}}$ given $\boldsymbol{\beta}$ is $N_p(\boldsymbol{\beta},\boldsymbol{\Sigma})$. Hence $\hat{\boldsymbol{\beta}}$ is sufficient for $\boldsymbol{\beta}$ and

$$(4.2)\qquad \boldsymbol{\psi}_G=\mathrm{E}(\boldsymbol{\beta}|\boldsymbol{Y})=\mathrm{E}(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}})=\boldsymbol{\psi}_G(\hat{\boldsymbol{\beta}})\,.$$

If $f(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta})$ denotes the conditional p.d.f. of $\hat{\boldsymbol{\beta}}$ given $\boldsymbol{\beta}$, then

$$(4.3)\qquad f(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta})=\frac{|\boldsymbol{\Sigma}|^{1/2}}{(2\pi)^{p/2}}\exp\left\{-\frac{1}{2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right\}$$

where $|\boldsymbol{\Sigma}|=\det\boldsymbol{\Sigma}$. The marginal p.d.f. of $\hat{\boldsymbol{\beta}}$ is given by

$$(4.4)\qquad f(\hat{\boldsymbol{\beta}})=\int f(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta})dG(\boldsymbol{\beta})\,.$$

For $j=1,\cdots,n$, let $\hat{\boldsymbol{\beta}}_j$ be the MLE of $\boldsymbol{\beta}_j$ in the jth experiment that we had in the past. Then

$$\hat{\boldsymbol{\beta}}_j=\boldsymbol{\Sigma X'Y}_j\,,\qquad j=1,\cdots,n$$

and $\hat{\boldsymbol{\beta}}_1,\cdots,\hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\beta}}$ are i.i.d. according to the marginal p.d.f. given by (4.4). Thus the *optimal estimator in the* EB *context* given by $\mathrm{E}(\boldsymbol{\beta}|\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_n,\boldsymbol{Y})=\mathrm{E}(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}_1,\cdots,\hat{\boldsymbol{\beta}}_n,\hat{\boldsymbol{\beta}})$ is simply $\mathrm{E}(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}})=\boldsymbol{\psi}_G$. Thus the estimators which are as good as the o.e. ψ_G in the component problem are also as good as the o.e. in the EB context.

It follows from Theorem 2.9 of Lehmann [12] that the mth order partial derivative of $f(\hat{\boldsymbol{\beta}})$ with respect to the ith component $\hat{\beta}_i$ of $\hat{\boldsymbol{\beta}}$ is

$$(4.5)\qquad \frac{\partial^m f(\hat{\boldsymbol{\beta}})}{\partial\hat{\beta}_i^m}=\int\frac{\partial^m f(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta})}{\partial\hat{\beta}_i^m}dG(\boldsymbol{\beta})$$
$$\text{for all } m=1,2,\cdots \text{ and } i=1,2,\cdots,p\,.$$

Now notice that

$$\frac{\partial f(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta})}{\partial\hat{\boldsymbol{\beta}}_i}=\begin{bmatrix}\partial f(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta})/\partial\hat{\beta}_1\\ \vdots\\ \partial f(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta})/\partial\hat{\beta}_p\end{bmatrix}=\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})f(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta})\,.$$

Therefore

$$\boldsymbol{\beta} f(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}} f(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta}) + \boldsymbol{\Sigma}(\partial f(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta})/\partial\hat{\boldsymbol{\beta}}) . \tag{4.6}$$

Hence from (4.2), (4.5) and (4.6) it follows that $\boldsymbol{\phi}_G$ can be written as

$$\boldsymbol{\phi}_G = \mathrm{E}(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}) = (f(\hat{\boldsymbol{\beta}}))^{-1} \int \boldsymbol{\beta} f(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta}) dG(\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}} + \boldsymbol{\Sigma}\boldsymbol{q}(\boldsymbol{\beta}) \tag{4.7}$$

where $\boldsymbol{q}' = (q_1, \cdots, q_p)$ and

$$q_s = f_s / f \qquad \text{and} \qquad f_s(\hat{\boldsymbol{\beta}}) = \partial f(\hat{\boldsymbol{\beta}})/\partial\hat{\beta}_s \tag{4.8}$$

with $\hat{\beta}_s$ denoting the sth component of $\hat{\boldsymbol{\beta}}$.

Thus the problem of approximating the minimum expected loss estimator $\boldsymbol{\phi}_G$ reduces to the problem of estimating the marginal p.d.f. f and its first order partial derivatives $f_1, \cdots, f_p$. The representation (4.7) of $\boldsymbol{\phi}_G$ is also noted in Martz and Krutchkoff [18] and Singh [28].

5. Proposed empirical Bayes estimators and an important result

In this section first we will exhibit estimators of the marginal p.d.f. f and its first partial derivatives f_s. We will then obtain a bound for the mean square errors of these estimates. Finally, on the basis of these estimates we will exhibit two classes of EB estimates for $\boldsymbol{\beta}$, one for the case when nothing is known about the support of the prior distribution and the other for the case when the prior distribution has compact support in E^p, the p-dimensional Euclidean space.

5.1. *Estimation of f and f_s*

For an integer $r>1$ and for $i=0, 1$, let $\mathcal{K}_i^r$ be the class of all Borel-measurable bounded functions vanishining outside $(0, 1)$ such that for K_0 in $\mathcal{K}_0^r$,

$$\int y^j K_0(y) dy = \begin{cases} 1 & \text{if } j=0 \\ 0 & \text{if } j=1, \cdots, r-1 \end{cases}$$

and for K_1 in $\mathcal{K}_1^r$,

$$\int y^j K_1(y) dy = \begin{cases} 1 & \text{if } j=1 \\ 0 & \text{if } j=0, 2, 3, \cdots, r-1 . \end{cases}$$

Let $0<h=h(n)$ be a function of n such that $h \to 0$ as $n \to \infty$. For $j= 1, \cdots, n$, let $\hat{\beta}_{j1}, \cdots, \hat{\beta}_{jp}$ denote the components of the MLE $\hat{\boldsymbol{\beta}}_j$ of $\boldsymbol{\beta}_j$. At $\boldsymbol{x}' = (x_1, \cdots, x_p)$, we estimate f by

$$\hat{f}(\boldsymbol{x})=(nh^p)^{-1}\sum_{j=1}^{n}\left\{\mathop{\times}_{i=1}^{p} K_0\left(\frac{\hat{\beta}_{ji}-x_i}{h}\right)\right\} \tag{5.1}$$

and $f_s(\boldsymbol{x})=\partial f(\boldsymbol{x})/\partial x_s$ by

$$\hat{f}_s(\boldsymbol{x})=(nh^{p+1})^{-1}\sum_{j=1}^{n}\left\{K_1\left(\frac{\hat{\beta}_{js}-x_s}{h}\right)\mathop{\times}_{\substack{i=1\\ i\neq s}}^{p}\left(K_0\left(\frac{\hat{\beta}_{ji}-x_i}{h}\right)\right)\right\}\,. \tag{5.2}$$

Estimates (5.1) and (5.2) are special cases of those given in Singh ([26], p. 37 and [30]) where the author has given nonparametric method of estimation of a mixed partial derivatives of a multivariate density. For $0<\gamma\leqq 2$, the following theorem proves the γth mean consistency of estimators $\hat{f}$ and $\hat{f}_s$ of f and f_s respectively for each $s=1,\cdots,p$. The numbers $c_1, c_2, \cdots$ below are absolute constants.

THEOREM 5.1. *For every* $0<\gamma\leqq 2$,

$$\mathrm{E}\,|\hat{f}(\boldsymbol{x})-f(\boldsymbol{x})|^{\gamma}\leqq c_1 f^{\gamma}(\boldsymbol{x})\{h^{r\gamma}B^{\gamma}(\boldsymbol{x})+(nh^{p})^{-\gamma/2}C^{\gamma/2}(\boldsymbol{x})\} \tag{5.3}$$

and

$$\mathrm{E}\,|\hat{f}_s(\boldsymbol{x})-f_s(\boldsymbol{x})|^{\gamma}\leqq c_2 f^{\gamma}(\boldsymbol{x})\{h^{(r-1)\gamma}B^{\gamma}(\boldsymbol{x})+(nh^{p+2})^{-\gamma/2}C^{\gamma/2}(\boldsymbol{x})\} \tag{5.4}$$

where, with

$$\lambda_0=\text{max. root }\boldsymbol{\Sigma}^{-1}\,. \tag{5.5}$$

$$t_1(\hat{\boldsymbol{\beta}},\boldsymbol{\beta})=\exp\left(h\lambda_0\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1\right)\sum_{j=0}^{r}\left(\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1^{j}+hp^{j}\right) \tag{5.6}$$

and

$$t_2(\hat{\boldsymbol{\beta}},\boldsymbol{\beta})=\exp\left(h\lambda_0\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1\right)(f(\hat{\boldsymbol{\beta}}))^{-1}\,, \tag{5.7}$$

B and C are given by

$$B(\boldsymbol{x})=\mathrm{E}\,[t_1(\hat{\boldsymbol{\beta}},\boldsymbol{\beta})\,|\,\hat{\boldsymbol{\beta}}=\boldsymbol{x}]$$

and

$$C(\boldsymbol{x})=\mathrm{E}\,[t_2(\hat{\boldsymbol{\beta}},\boldsymbol{\beta})\,|\,\hat{\boldsymbol{\beta}}=\boldsymbol{x}]\,.$$

Remark 5.1. Notice that $f^2(\boldsymbol{x})B^2(\boldsymbol{x})$ and $f^2(\boldsymbol{x})C(\boldsymbol{x})$ are bounded in $\boldsymbol{x}$ by a constant independent of n and G. Hence if h is taken proportional to $n^{-1/(p+2r)}$, then Theorem 5.1 proves that $\sup_{\boldsymbol{x}}\mathrm{E}\,(\hat{f}(\boldsymbol{x})-f(\boldsymbol{x}))^2\leqq$ const. $n^{-2r/(p+2r)}$ and $\sup_{\boldsymbol{x}}\mathrm{E}\,(\hat{f}_s(\boldsymbol{x})-f_s(\boldsymbol{x}))^2\leqq$ const. $n^{-2(r-1)/(p+2r)}$ uniformly for all $n\geqq 1$ and for all prior distributions.

PROOF OF THE THEOREM. We will prove (5.3). Proof of (5.4) follows analogously. Liapunov's inequality followed by c_r-inequality (Loève

[15], p. 155) gives

$$\text{(5.8)} \qquad \begin{aligned} \mathrm{E}\,|\hat{f}(\boldsymbol{x})-f(\boldsymbol{x})|^r &\leqq \{(\mathrm{E}\,\hat{f}(\boldsymbol{x})-f(\boldsymbol{x}))^2+\operatorname{var}(\hat{f}(\boldsymbol{x}))\}^{r/2} \\ &\leqq |\mathrm{E}\,\hat{f}(\boldsymbol{x})-f(\boldsymbol{x})|^r+(\operatorname{var}(\hat{f}(\boldsymbol{x})))^{r/2} . \end{aligned}$$

We now obtain bounds for each term on the right hand side of (5.8). Since $\hat{\boldsymbol{\beta}}_1,\cdots,\hat{\boldsymbol{\beta}}_n$ are i.i.d. with marginal p.d.f. f, it follows that

$$\mathrm{E}\,\hat{f}(\boldsymbol{x})=\int\cdots\int\left(\mathop{\times}_{i=1}^{p} K_0(z_i)\right)f(\boldsymbol{x}+h\boldsymbol{z})dz_1\cdots dz_p .$$

Substituting $f(\boldsymbol{x}+h\boldsymbol{z})$ by its rth order Taylor expansion about $\boldsymbol{x}$ with Lagrange-form of the remainder at the rth term and then making use of the orthogonality properties of K_0 and the fact that K_0 vanishes outside $(0, 1)$, we get

$$\text{(5.9)} \quad |\mathrm{E}\,\hat{f}(\boldsymbol{x})-f(\boldsymbol{x})|\leqq \text{const.}\, h^r \sup_{\|\boldsymbol{z}\|_1\leqq p}\left|\left\{\sum_{i_1=0}^{r}\cdots\sum_{\substack{i_p=0\\ i_1+\cdots+i_p=r}}^{r}\frac{\partial^r f(\boldsymbol{t})}{\partial t_1^{i_1}\cdots\partial t_p^{i_p}}\right|\boldsymbol{t}= (\boldsymbol{x}+h\boldsymbol{z})\right\}\right| .$$

Let $\boldsymbol{C}_i$ be the ith column of Σ^{-1}. Using Schwarz-inequality at the second step below we get for some constants $a_0, a_1,\cdots$

$$\begin{aligned} \frac{\partial^{i_k}\hat{f}(\boldsymbol{\beta}|\boldsymbol{\beta})}{\partial\hat{\beta}_i^{i_k}} &= \left|\sum_{j=0}^{i_k} a_j(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\boldsymbol{C}_i)^j f(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta})\right| \\ &\leqq(\lambda_0\vee 1)^{i_k}\max_{0\leqq j\leqq i_k}|a_j|\left(\sum_{j=0}^{i_k}\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1^j\right)f(\hat{\boldsymbol{\beta}}|\boldsymbol{\beta}) \end{aligned}$$

where λ_0 is given by (5.5). This inequality and (4.5) applied to (5.9) give

$$\text{(5.10)} \quad |\mathrm{E}\,\hat{f}(\boldsymbol{x})-f(\boldsymbol{x})|\leqq c_3 h^r\int\left(\sum_{j=0}^{r}\sup_{\|\boldsymbol{z}\|_1\leqq p}\|\boldsymbol{x}+h\boldsymbol{z}-\boldsymbol{\beta}\|_1^j f(\boldsymbol{x}+h\boldsymbol{z}|\boldsymbol{\beta})\right)dG(\boldsymbol{\beta}) .$$

Also, since $\sup_{\|\boldsymbol{z}\|_1\leqq p}|\boldsymbol{z}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\beta})|\leqq\lambda_0\|\boldsymbol{x}-\boldsymbol{\beta}\|_1$, we have

$$\text{(5.11)} \qquad \begin{aligned} &\sup_{\|\boldsymbol{z}\|_1\leqq p}\|\boldsymbol{x}+h\boldsymbol{z}-\boldsymbol{\beta}\|_1^j f(\boldsymbol{x}+h\boldsymbol{z}|\boldsymbol{\beta}) \\ &\quad\leqq 2^{j-1}(\|\boldsymbol{x}-\boldsymbol{\beta}\|_1^j+hp^j)f(\boldsymbol{x}|\boldsymbol{\beta})\exp(h\lambda_0\|\boldsymbol{x}-\boldsymbol{\beta}\|_1) . \end{aligned}$$

This last inequality applied to (5.10) gives

$$\text{(5.12)} \qquad \mathrm{E}\,|\hat{f}(\boldsymbol{x})-f(\boldsymbol{x})|\leqq\text{const.}\, f(\boldsymbol{x})B(\boldsymbol{x})$$

where $B(\boldsymbol{x})$ is as given in the theorem. Now consider var $(\hat{f}(\boldsymbol{x}))$. Since $\hat{\beta}_1,\cdots,\hat{\beta}_n$ are i.i.d. with marginal p.d.f. f; and for a r.v. T, var $(t)\leqq$ E (T^2),

$$\text{(5.3)} \qquad \operatorname{var}(\hat{f}(\boldsymbol{x})) \leqq (nh^p)^{-1} \int_0^1 \cdots \int_0^1 \left(\mathop{\times}_{i=1}^{p} K_0^2(z_i) \right) f(\boldsymbol{x}+h\boldsymbol{z}) dz_1 \cdots dz_p$$

$$\leqq \text{const.}\ (nh^p)^{-1} \int \sup_{\|\boldsymbol{z}\|_1 \leqq p} f(\boldsymbol{x}+h\boldsymbol{z} \mid \boldsymbol{\beta}) dG(\boldsymbol{\beta})$$

$$\leqq \text{const.}\ (nh^p)^{-1} f(\boldsymbol{x}) C(\boldsymbol{x})$$

where $C(\boldsymbol{x})$ is as given in the theorem, and the last inequality follows by arguments used for (5.11). Proof of (5.3) is now complete from (5.8), (5.12) and (5.13).

5.2. *Proposed* EB *estimators for* $\boldsymbol{\beta}$

Lemma 4.1, followed by Equation (4.7), shows that the problem of obtaining an a.o. estimator of $\boldsymbol{\beta}$ is essentially the problem of obtaining a quadratic-mean consistent estimator of $\boldsymbol{q}(\hat{\boldsymbol{\beta}})$ where $\boldsymbol{q}=(q_1, \cdots, q_p)$ and $q_s = f_s/f$. Having noted that $\hat{f}$ and $\hat{f}_s$, given by (5.1) and (5.2) are mean square consistent estimators of f and f_s respectively, we are now able to propose our EB estimators of $\boldsymbol{\beta}$ for the two cases indicated below.

Notation. For $b>0$ and for a number a, $[a]_b$ will stand for $-b$, a or b according as $a<-b$, $|a| \leqq b$ or $a>b$. For any vector $\boldsymbol{t}'=(t_1, \cdots, t_p)$, $[\boldsymbol{t}']_c$ means $([t_1]_c, \cdots, [t_p]_c)$.

Case 1. If nothing is known about the support of the prior distribution G, then our proposed EB estimator for $\boldsymbol{\beta}$ is

$$\text{(5.14)} \qquad \hat{\boldsymbol{\varphi}}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}} + \boldsymbol{\Sigma} \hat{\boldsymbol{q}}(\hat{\boldsymbol{\beta}})$$

where $\hat{\boldsymbol{q}}' = (\hat{q}_1, \cdots, \hat{q}_p)$, and for $s=1, \cdots, p$,

$$\text{(5.15)} \qquad \hat{q}_s(\hat{\boldsymbol{\beta}}) = [\hat{f}_s(\hat{\boldsymbol{\beta}}) / \hat{f}(\hat{\boldsymbol{\beta}})]_{h^{-1}} .$$

Case 2. If it is known that the support of G is in a compact subset $[-A, A]^p$ of E^p and A is known, then our proposed EB estimator for $\boldsymbol{\beta}$ is

$$\text{(5.16)} \qquad \tilde{\boldsymbol{\varphi}}(\hat{\boldsymbol{\beta}}) = [\hat{\boldsymbol{\beta}} + \boldsymbol{\Sigma} \tilde{\boldsymbol{q}}(\hat{\boldsymbol{\beta}})]_A$$

where $\tilde{\boldsymbol{q}}' = (\tilde{q}_1, \cdots, \tilde{q}_p)$ and for $s=1, \cdots, p$

$$\text{(5.17)} \qquad \tilde{q}_s(\hat{\boldsymbol{\beta}}) = \hat{f}_s(\hat{\boldsymbol{\beta}}) / \hat{f}(\hat{\boldsymbol{\beta}}) .$$

6. Asymptotic optimality of the EB estimators $\hat{\boldsymbol{\varphi}}$ and $\tilde{\boldsymbol{\varphi}}$ and speed of convergence

In this section we will show that the EB estimators $\hat{\boldsymbol{\varphi}}$ and $\tilde{\boldsymbol{\varphi}}$ of

$\boldsymbol{\beta}$ are asymptotically as good as the unavailable optimal estimator $\boldsymbol{\phi}_G$ whatever be the G as long as it satisfies a certain moment condition. Theorems in this section demonstrate that EB estimators a.o. with rates arbitrarily close to $O(n^{-1})$ can be constructed. Using a variation of (2.3), we will denote the risks of $\hat{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\phi}}$ by $R_n(\hat{\boldsymbol{\phi}}, G)$ and $R_n(\tilde{\boldsymbol{\phi}}, G)$ respectively.

THEOREM 6.1. *Let h be proportional to $n^{-1/(p+2r)}$. If for some $0<\gamma<2$, the prior distribution G satisfies*

$$\int \|\boldsymbol{\beta}\|_1^{p\gamma+/(2-\gamma)} dG(\boldsymbol{\beta}) < \infty \,, \tag{6.0}$$

then

$$R_n(\hat{\boldsymbol{\phi}}, G) - R(G) \leqq c_4 n^{-(r\gamma-2)/(2r+p)} \qquad \textit{for every } n \geqq 1 \tag{6.1}$$

where c_4 is independent of n and γ.

PROOF. Fix a γ in $(0, 2)$. Let λ_0^* denote the max. root of $\boldsymbol{\Sigma}$. Then Lemma 4.1 followed by (4.7) and (5.14) gives

$$\begin{aligned} R_n(\hat{\boldsymbol{\phi}}, G) - R(G) &= \mathrm{E}\,\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_G\|_2^2 \\ &= \mathrm{E}\,\|\boldsymbol{\Sigma}\hat{(}\hat{\boldsymbol{q}}(\boldsymbol{\beta}) \hat{-} \boldsymbol{q}(\boldsymbol{\beta})\|_2^2 \\ &\leqq \lambda_0^{*2} \sum_{s=1}^{p} \mathrm{E}\,[\hat{q}_s(\hat{\boldsymbol{\beta}}) - q_s(\hat{\boldsymbol{\beta}})]^2 \,. \end{aligned} \tag{6.2}$$

Temporarily, let $t \neq 0$ and t' be numbers and T and T' be random variables. Then we prove that for every $a>0$,

$$\begin{aligned} \mathrm{E}\left|\frac{t'}{t} - \left[\frac{T'}{T}\right]_a\right|^2 &\leqq 8|t'/t|^2 I(|t'/t|>a) + 16a^{2-\gamma}|t|^{-\gamma} \\ &\quad \times \{\mathrm{E}\,|T'-t'|^\gamma + 2a^\gamma\, \mathrm{E}\,|T-t|^\gamma\} \,. \end{aligned} \tag{6.3}$$

(The inequality is true for $\gamma=2$ too). To prove (6.3), notice that for numbers b_1 and b_2,

$$|b_1 - [b_2]_a| \leqq (|b_1 - b_2| \wedge 2a) I(|b_1| \leqq a) + (a + |b_1|) I(|b_1| > a) \,,$$

and by a use of c_r-inequality (Loève [15], p. 155) we get

$$|b_1 - [b_2]_a|^2 \leqq 2(2a)^{2-\gamma}(|b_1 - b_2| \wedge 2a)^\gamma I(|b_1| \leqq a) + 8|b_1|^2 I(|b_1| > a) \,.$$

This inequality followed by the lemma in the Appendix of Singh [28] gives (6.3).

Thus (6.3) in conjunction with the definitions $q_s = f_s/f$ from (4.8) and $\hat{q}_s = [\hat{f}_s/\hat{f}]_{h^{-1}}$ from (5.14) and Theorem 5.1 gives

(6.4) $$\mathrm{E}\,[\hat{q}_s(\hat{\boldsymbol{\beta}})-q_s(\hat{\boldsymbol{\beta}})]^2 \leqq c_5(D_1+D_2+D_3)$$

where, with B and C as given in Theorem 5.1,

$$D_1=\mathrm{E}\,\{|q_s(\hat{\boldsymbol{\beta}})|^2 I(|q_s(\hat{\boldsymbol{\beta}})|>h^{-1})\}$$

$$D_2=h^{r\gamma-2}\,\mathrm{E}\,\{B^{\gamma}(\hat{\boldsymbol{\beta}})\}$$

and $$D_3=n^{-\gamma/2}h^{-2-(\gamma p/2)}\,\mathrm{E}\,\{C^{\gamma/2}(\hat{\boldsymbol{\beta}})\}\,.$$

First consider D_1. Hölder-inequality followed by Markov-inequality gives for a $\delta>1$,

(6.5) $$\begin{aligned} D_1 &\leqq (\mathrm{E}\,|q_s(\hat{\boldsymbol{\beta}})|^{2\delta})^{\delta^{-1}}(\mathrm{E}\,I(|q_s(\hat{\boldsymbol{\beta}})|>h^{-1}))^{1-\delta^{-1}} \\ &\leqq h^{r\gamma-2}(\mathrm{E}\,|q_s(\hat{\boldsymbol{\beta}})|^{2\delta})^{\delta^{-1}}(\mathrm{E}\,|q_s(\hat{\boldsymbol{\beta}})|^{(r\gamma-2)/(1-\delta^{-1})})^{1-\delta^{-1}}\,. \end{aligned}$$

Notice that if $\boldsymbol{C}_s$ is the sth column of $\boldsymbol{\Sigma}^{-1}$, then

$$\left|\frac{\partial f(\hat{\boldsymbol{\beta}}\,|\,\boldsymbol{\beta})}{\partial\hat{\beta}_s}\right| = |(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{C}_s|\,f(\hat{\boldsymbol{\beta}}\,|\,\boldsymbol{\beta}) \leqq \|\boldsymbol{C}_s\|_1\,\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1 f(\hat{\boldsymbol{\beta}}\,|\,\boldsymbol{\beta})\,.$$

Thus by (4.5), $|q_s(\hat{\boldsymbol{\beta}})|=|f_s(\hat{\boldsymbol{\beta}})/f(\hat{\boldsymbol{\beta}})| \leqq \|\boldsymbol{C}_s\|_1\,\mathrm{E}\,(\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1\,|\,\hat{\boldsymbol{\beta}})$. Thus by Hölder-inequality for every $\eta\geqq 1$

$$\mathrm{E}\,|q_s(\hat{\boldsymbol{\beta}})|^{\eta} \leqq \|\boldsymbol{C}_s\|_1^{\eta}\,\mathrm{E}\,\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1^{\eta} \quad <\infty\,.$$

Similarly it can be shown that for every $0<\eta<1$, $\mathrm{E}\,|q_s(\hat{\boldsymbol{\beta}})|^{\eta}<\infty$. Hence from (6.5) we have

(6.6) $$h^{-(r\gamma-2)}D_1<\infty \qquad \text{for all } n\,.$$

Now to see that $h^{-(r\gamma-2)}D_2$ is finite, notice that for an integer $j\geqq 0$ and any $\eta\geqq 1$, Hölder inequality yields

(6.7) $$\begin{aligned} &\mathrm{E}\,\{\mathrm{E}\,(\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1^{j}\exp\,(c_6\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1)\,|\,\hat{\boldsymbol{\beta}})\}^{\eta} \\ &\leqq \mathrm{E}\,[\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1^{j\eta}\exp\,(c_6\eta\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1)]<\infty\,. \end{aligned}$$

Similarly, it can be shown that the left hand side of (6.7) for $0<\eta<1$ too is finite. Thus from the definition of $B(\hat{\boldsymbol{\beta}})$ given in Theorem 5.1, we conclude that

(6.8) $$h^{-(r\gamma-2)}D_2<\infty \qquad \text{whatever be } n\,.$$

Finally consider D_3. Notice that from (5.7)

$$\mathrm{E}\,(C^{\gamma/2}(\hat{\boldsymbol{\beta}}))=\mathrm{E}\,[(f(\hat{\boldsymbol{\beta}}))^{-\gamma/2}(\mathrm{E}\,\{\exp\,(h\lambda_0\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1)\,|\,\hat{\boldsymbol{\beta}}\}\,)^{\gamma/2}]\,.$$

Take an η such that $\gamma/2<\eta<1$. Then by Hölder inequality

(6.9) $$\mathrm{E}\,(C^{\gamma/2}(\hat{\boldsymbol{\beta}})) \leqq c_7[\mathrm{E}\,(f(\hat{\boldsymbol{\beta}}))^{-\eta}]^{\gamma/2\eta}\,.$$

Since $\mathrm{E}(f(\hat{\beta}))^{-\eta}=\int f^{1-\eta}(\hat{\beta})d\hat{\beta}$; $\mathrm{E}[(f(\hat{\beta}))^{-\eta}I(\|\hat{\beta}\|_1\leqq p)]$ is finite. According to our hypothesis on G, there exists an $\varepsilon>0$ such that $\mathrm{E}\|\beta\|_1^{p\gamma(2-\gamma)^{-1}+\varepsilon}<\infty$. Take $\xi=p-1+\varepsilon(1-\eta)\eta^{-1}$. Then by Hölder inequality

$$\mathrm{E}(f(\hat{\beta}))^{-\eta}I(\|\hat{\beta}\|\geqq p)\leqq I_1\cdot I_2$$

where

$$I_1=\left\{\int \|\hat{\beta}\|_1^{-1-\xi}I(\|\hat{\beta}\|_1>p)d\hat{\beta}\right\}^{\eta}$$

and

$$I_2=\{\mathrm{E}\|\hat{\beta}\|_1^{(1+\xi)\eta/(1-\eta)}\}^{1-\eta}\,.$$

Now by a use of the multivariate polar coordinate transformation (see Kendall and Stuart [11], pp. 246–247) and the inequalities $\|\hat{\beta}\|_2^2\leqq\|\hat{\beta}\|_1^2\leqq p\|\hat{\beta}\|_2^2$ it can be shown that $I_1<\infty$ since $\xi+1>p$. Further

$$I_2^{1/(1-\eta)}\leqq\text{const.}\,(1+\mathrm{E}(\|\beta\|_1^{(1+\xi)\eta/(1-\eta)})<\infty$$

by our hypothesis, since $(1+\xi)\eta/(1-\eta)=p\eta(1-\eta)^{-1}+\varepsilon$. Hence we conclude that

$$n^{\gamma/2}h^{2+(\gamma p/2)}D_3<\infty \qquad \text{for all } n\geqq 1\,. \tag{6.10}$$

Now the proof of the theorem is complete from (6.2), (6.4), (6.6), (6.8) and (6.10).

Remark 6.1. Notice that the moment assumption (6.0) on the prior distribution G is used in the proof of the theorem only to prove

$$\mathrm{E}(f(\hat{\beta}))^{-\eta}<\infty \qquad \text{for an } \eta \text{ in } (0, 1)\,. \tag{6.11}$$

Professor Dennis Gilliland has provided the following counter-example showing that (6.11) is not true for all G whatever be η in $(1/2, 1)$. Consider only the univariate case. It suffices to show that with $\varepsilon=1-\eta$, in $(0, 1/2]$ there exists a G for which the integral,

$$I=\int\left[\int \exp\left(-\frac{1}{2}(x-\beta)^2dG(\beta)\right)\right]^{\varepsilon}dx$$

is not finite. Consider a prior G which puts mass c/k^2 on $k=1, 2,\cdots$ where $c=\left(\sum_1^{\infty} 1/k^2\right)^{-1}$. Then

$$I=c^{\varepsilon}\int\left[\sum_{k=1}^{\infty}\frac{\exp((x-k)^2/2)}{k^2}\right]^{\varepsilon}dx\,.$$

But $\left[\sum_{k=1}^{\infty}\{\exp(-(x-k)^2/2)/k^2\}\right]^{\varepsilon}\geqq\exp(-\varepsilon(x-k)^2/2)/k^{2\varepsilon}$ for all x and k.

Hence applying this inequality to the interval $k-1/2 \leqq x \leqq k+1/2$, we see that

$$I \geqq c^{\varepsilon} \lim_{K\to\infty} \sum_{k=1}^{K} \left\{ \int_{k-1/2}^{k+1/2} \exp\left(-\varepsilon(x-k)^2/2\right) dx / k^{2\varepsilon} \right\} .$$

But since $\int_{k-1/2}^{k+1/2} \exp\left(-\varepsilon(x-k)^2/2\right) dx \geqq \exp(-\varepsilon/8)$ for all k,

$$I \geqq c^{\varepsilon} \exp(-\varepsilon/8) \lim_{K\to\infty} \sum_{k=1}^{K} 1/k^{2\varepsilon} = \infty \qquad \text{for } \varepsilon \leqq 1/2 .$$

Now we will prove the asymptotic optimality of our second EB estimator $\tilde{\boldsymbol{\phi}}$ of $\boldsymbol{\beta}$.

THEOREM 6.2. *Let h be proportional to $n^{-1/(2r+p)}$. If the support of the prior distribution G is in some known compact subset $[-A, A]^p$ of E^p and $\tilde{\boldsymbol{\phi}}$ is given by* (5.16), *then for every* $0<\gamma<2$

$$R_n(\tilde{\boldsymbol{\phi}}, G) - R(G) \leqq c_8 n^{-(r-1)\gamma/(2r+p)} \tag{6.12}$$

where c_8 is independent of n and γ.

PROOF. From (4.7) and (5.16), the ith components of $\hat{\boldsymbol{\phi}}_G(\hat{\boldsymbol{\beta}})$ and $\tilde{\boldsymbol{\phi}}(\hat{\boldsymbol{\beta}})$ can be expressed as ratios $t_i(\hat{\boldsymbol{\beta}})/f(\hat{\boldsymbol{\beta}})$ and $[T_i(\hat{\boldsymbol{\beta}})/\hat{f}(\hat{\boldsymbol{\beta}})]_A$ for $i=1, \cdots, p$. Since the support of G is in $[-A, A]^p$, t_i/f is in $[-A, A]$ for each $i=1, \cdots, p$. Therefore, by Lemma 4.1,

$$\begin{aligned} R_n(\tilde{\boldsymbol{\phi}}, G) - R(G) &= \mathrm{E}\, \|\tilde{\boldsymbol{\phi}}(\hat{\boldsymbol{\beta}}) - \hat{\boldsymbol{\phi}}_G(\hat{\boldsymbol{\beta}})\|_2^2 \\ &\leqq \sum_{i=1}^{p} \mathrm{E} \left(\left| \frac{t_i(\hat{\boldsymbol{\beta}})}{f(\hat{\boldsymbol{\beta}})} - \frac{T_i(\hat{\boldsymbol{\beta}})}{\hat{f}(\hat{\boldsymbol{\beta}})} \right| \wedge 2A \right)^2 . \end{aligned} \tag{6.13}$$

Notice that t_i is a linear combination of $f, f_1, \cdots, f_p$ and that T_i of $\hat{f}, \hat{f}_1, \cdots, \hat{f}_p$; the coefficient of $f(\hat{\boldsymbol{\beta}})$ in $t_i(\hat{\boldsymbol{\beta}})$ and that of $\hat{f}(\hat{\boldsymbol{\beta}})$ in $T_i(\hat{\boldsymbol{\beta}})$ being the ith component of $\hat{\boldsymbol{\beta}}$. Therefore, (6.13) followed by the lemma in the Appendix of Singh [28] and Theorem 5.1 here gives for $0<\gamma<2$

$$R_n(\hat{\boldsymbol{\phi}}, G) - R(G) \leqq c_9 n^{-(r-1)\gamma/(p+2r)} (\mathrm{E}\, B^{\gamma}(\hat{\boldsymbol{\beta}}) + \mathrm{E}\, C^{\gamma/2}(\hat{\boldsymbol{\beta}})) . \tag{6.14}$$

The proof of the theorem is complete since in the proof of Theorem 6.1 it is shown that $\mathrm{E}\, B^{\gamma}(\hat{\boldsymbol{\beta}})$ and $\mathrm{E}\, C^{\gamma/2}(\hat{\boldsymbol{\beta}})$ are finite constants for $0<\gamma<2$.

Remark 6.2. It may be recalled that the constants C_4 in (6.1) and C_8 in (6.12) are independent of n and r. Further, the rate result in (6.1) in uniform over the class of all priors G satisfying (6.0), and that in (6.12) is uniform over the class of all priors whose supports are a subset of $[-A, A]^p$.

For each integer $r>1$ we have exhibited two sets of EB estimators $\hat{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\phi}}$ for the general multiple linear regression model $\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{\varepsilon}$. Theorem 6.1 gives sufficient condition on the distribution G of $\boldsymbol{\beta}$ under which $\hat{\boldsymbol{\phi}}$ is a.o. with rates $O(n^{-(r\gamma-2)/(2r+p)})$ for every choice of γ in $(0, 2)$. Theorem 6.2 shows that if G has a compact support then estimators $\tilde{\boldsymbol{\phi}}$ a.o. with improved rates $O(n^{-(r-1)\gamma/(2r+p)})$ for any γ in $(0, 2)$ can be constructed. This paper, thus, suggests how to construct EB estimators a.o. with rates arbitrarily close to $O(n^{-1})$ in the regression model $\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{\varepsilon}$.

Acknowledgments

The author wishes to thank Professor James Hannan for his various helpful comments during the preparation of the first version of the manuscript. The author also wishes to thank Professor Dennis Gilliland for providing the counter-example following (6.11).

UNIVERSITY OF GUELPH

REFERENCES

[1] Bennett, Kemble G. and Martz, H. F. (1972). A continuous empirical Bayes smoothing technique, *Biometrika*, **59**, 361-368.

[2] Clemmer, B. A. and Krutchkoff, Richard G. (1968). The use of empirical Bayes estimators in a linear regression model, *Biometrika*, **55**, 525-534.

[3] Cogburn, R. (1965). On the estimation of a multivariate location parameter with squared error loss, *Bernoulli (1723), Bayes (1763) and Laplace (1813) Anniversary Volume* (eds. J. Neyman and L. LeCam), Springer-Verlag, Berlin, 24-29.

[4] Efron, B. and Morris, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators—Part II: The empirical Bayes case, *J. Amer. Statist. Ass.*, **67**, 130-139.

[5] Griffin, Barry S. and Krutchkoff, Richard G. (1971). Optimal linear estimators: an empirical Bayes version with application to the binomial distribution, *Biometrika*, **58**, 195-201.

[6] James, W. and Stein, C. (1961). Estimation with quadratic loss, *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, **1**, Univ. California Press, 361-379.

[7] Johns, M. V., Jr. (1957). Nonparametric empirical Bayes procedures, *Ann. Math. Statist.*, **28**, 649-669.

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

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

[10] Kantor, M. (1967). Estimating the mean of a multivariate normal distribution with applications to time series and empirical Bayes estimation, Ph. D. dissertation, Columbia Univ.

[11] Kendall, Maurice and Stuart, Alan. (1969). *The Advanced Theory of Statistics, 3rd ed. Vol.* I, Hafner Pub. Co., New York.

[12] Lehmann, E. L. (1959). *Testing of Statistical Hypothesis*, Wiley, New York.

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

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

[15] Loève, Michel (1963). *Probability Theory, 3rd ed.*, Van Nostrand, Princeton.

[16] Maritz, J. S. (1969). Empirical Bayes estimation for continuous distributions, *Biometrika*, **56**, 349-359.

[17] Maritz, J. S. and Lwin, T. (1975). Construction of simple empirical Bayes estimators, *J. R. Statist. Soc.*, B, **75**, 421-425.

[18] Martz, H. and Krutchkoff, R. (1969). Empirical Bayes estimators in a multiple linear regression model, *Biometrika*, **56**, 367-374.

[19] Neyman, J. (1962). Two breakthroughs in the theory of statistical decision making, *Rev. Int. Statist. Inst.*, **30**, 11-27.

[20] O'Bryan, Thomas E. and Susarla, V. (1976). Rates in the empirical Bayes estimation problem with nonidentical components: Case of normal distributions, *Ann. Inst. Statist. Math.*, **28**, 389-397.

[21] Rao, C. Radhakrishna (1975). Simultaneous estimation of parameters in different linear models and applications to biometric problems, *Biometrika*, **31**, 545-554.

[22] Robbins, Herbert (1955). An empirical Bayes approach to statistics, *Proc. 3rd Berkeley Symp. Math. Statist. Prob.*, **1**, 157-163, University California Press.

[23] Robbins, Herbert (1963). The empirical Bayes approach to the testing of statistical hypothesis, *Rev. Int. Statist. Inst.*, **31**, 195-208.

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

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

[26] Singh, R. S. (1974). Estimation of derivatives of average of μ-densities and sequence-compound estimation in exponential families, RM-318, Dept. Statist. Prob., Michigan State University.

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

[28] Singh, R. S. (1977). Applications of estimators of a density and its derivatives to certain statistical problems, *J. R. Statist. Soc.*, B, **39**, 357-363.

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

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

[31] Stein, C. (1960). Multiple regression, *Contributions to Probability and Statistics—Essays in Honor of Harold Hotelling*, Stanford University Press, 424-443.

[32] Wind, S. (1972). Stein-James estimators of a multivariate location parameter, *Ann. Math. Statist.*, **43**, 340-343.

[33] Wind, S. (1973). An empirical Bayes approach to multiple linear regression, *Ann. Statist.*, **1**, 93-103.

[34] Yu, Benito (1971). Rates of convergence in empirical Bayes two-action and estimation problems and in sequence-compound estimation problems, RM-279, Dept. Statist. Prob., Michigan State University.