数理解析研究所講究録 1380 巻 2004 年 127-139

A New Class of Generalized Bayes Minimax Ridge Regression Estimators

丸山 祐造

YUZO MARUYAMA

東京大学・空間情報科学研究センター

CENTER FOR SPATIAL INFORMATION SCIENCE, THE UNIVERSITY OF TOKYO *

1 Introduction

In this paper we consider absolute ridge regression estimators in general linear model with homogeneous normal errors. These are two main contributions a) we find a class of generalized Bayes estimators which have a particularly simple form and b) we show that we may always construct such estimators with smaller condition number than the usual least squares estimator.

Hoerl and Kennard (1970) introduced the ridge regression technique as a way to simultaneously reduce the risk and increase the numerical stability of the least squares estimator in ill-conditional problem. The risk reduction aspect of Hoerl and Kennard's method was often observed in simulations but was not theoretically justified. Strawderman (1978) looked at the problem in the context of minimaxity and produced minimax adaptive ridge-type estimators but ignored the condition number aspect of the problem. Casella (1980,1985) considered both the minimaxity and condition number aspects and gave estimators which were minimax and condition number decreasing for some but not all design matrix. Neither Strawderman or Casella gave generalized Bayes minimax estimators.

In the present paper, we propose a broad class of generalized Bayes minimax estimators which increase the numerical stability of the least squares estimator for all full rank design matrices. What is particularly noteworthy about our class of estimators is that they contain a subclass with a form (adapted to the case of unknown σ^2) which is remarkably similar to that of the estimators originally suggested in Stein (1956) for the case $\mathrm{Cov}(X) =$

*maruyama@csis.u-tokyo.ac.jp

I. In particular, our simple generalized Bayes estimators of the mean vector are of the form

$$\hat{\theta}_{SB} = (I - \alpha/\{\gamma(\alpha+1) + W\}C^{-1})X$$

where $W = X'C^{-1}D^{-1}X/S$ for some positive definite matrices C and D.

To be more precise, we start the familiar linear regression model $Y = A\beta + \epsilon$ where Y is an $n \times 1$ vector of observations, A is the known $n \times p$ design matrix of rank p, β is the $p \times 1$ vector of unknown regression coefficients, and ϵ is an $n \times 1$ vector of experimental errors. We assume ϵ has a multivariate normal distribution with mean zero and covariance matrix $\sigma^2 I$, that is, $\epsilon \sim N(0, \sigma^2 I)$.

The least squares estimator of β is $\hat{\beta} = (A'A)^{-1}A'y$. Since the covariance matrix of $\hat{\beta}$ is given by $\sigma^2(A'A)^{-1}$, the least squares estimator may not be a suitable estimator when some components of $\hat{\beta}$ or some linear combinations of $\hat{\beta}$ have a very large variance and when $A'A$ is nearly singular. Additionally $(A'A)^{-1}$ may have inflated diagonal values so that small changes in the observations produce large changes in $\hat{\beta}$. Hoerl and Kennard (1970) proposed the ridge estimator

$$\hat{\beta}_R(k) = (A'A + kI_p)^{-1}A'y \tag{1.1}$$

where k is a positive constant to ameliorate these problems. Adding the number k before inverting amounts to increasing each eigenvalue of $A'A$ by k. We will also be concerned with reducing the condition number of adaptive version of $\hat{\beta}_R$ in Section 3 and 4.

In particular, if P is the orthogonal matrix of eigenvectors of $(A'A)^{-1}$, with $d_1 \geq d_2 \geq \cdots \geq d_p$ as eigenvalues, it follows that

$$P'(A'A)^{-1}P = D, \quad P'P = I_p$$

where $D = \text{diag}(d_1, \ldots, d_p)$. Then (1.1) can be written as

$$\hat{\beta}_R(k) = P(D^{-1} + kI_p)^{-1}P'Ay. \tag{1.2}$$

The ridge estimator is more stable than $\hat{\beta}$ in the sense that the condition number of the estimator is reduced.

However, we are interested in proposing better estimator than $\hat{\beta}$ from the decision-theoretic point of view. We measure the loss in estimating β by b with loss function

$$L(b, \beta) = \sigma^{-2}(b - \beta)'(b - \beta).$$

Then risk function (mean squared error) of an estimator b is given by

$$R(b, \beta) = EL(b, \beta).$$

The least squares estimator $\hat{\beta}$ is minimax with constant risk. Therefore, b is a minimax estimator of β if and only if

$$R(b, \beta) \leq R(\hat{\beta}, \beta) = \sum d_i, \quad \text{for all } \beta.$$

Hence the search for estimators better than $\hat{\beta}$ is a search for minimax estimators.

To simplify expression and to make matters a bit clearer it is helpful to rotate the problem via the above orthogonal transformation, P, so that the covariance matrix becomes diagonal. We define $X = P'\hat{\beta}$ and $\theta = P'\beta$, which implies that $X \sim N(\theta, \sigma^2 D)$. Therefore, for $X \sim N(\theta, \sigma^2 D)$ and $S = (y - A\hat{\beta})'(y - A\hat{\beta}) \sim \sigma^2 \chi^2_n$, (independent of X), we consider the problem of estimation of θ under the loss function $(\delta - \theta)'(\delta - \theta)/\sigma^2$.

Strawderman (1978) and Casella (1980) essentially considered the class of estimators of the form

$$\hat{\theta}_R(K) = \left(I - \{I + D^{-1}K^{-1}\}^{-1}\right) X,$$

which originally came from straight generalization of (1.2), that is, the generalized ridge estimator

$$\hat{\beta}_R(K) = P(D^{-1} + K)^{-1} P' A' y$$

where $K = \text{diag}(k_1, \ldots, k_p)$. They proposed a sufficient condition for minimaxity, for adaptive estimators $\hat{\theta}_R(\hat{K})$ where $\hat{K} = \psi(X'D^{-1}X/S)\text{diag}(a_1, \ldots, a_p)$, ψ is a suitable positive function and a_i is positive for all i. Casella (1980) discussed the relationship between minimaxity and stability (in terms of lowered condition number) and pointed out that forcing ridge regression estimators to be minimax makes it difficult for them to provide the numerical stability for which they were originally intended. Casella (1985) found that, under certain conditions on the structure of the eigenvalues of the design matrix, both minimaxity and stability can be simultaneously achieved for a special case $\psi(w) \equiv w^{-1}$.

In section 2, we give a class of minimax estimators of θ (and hence, by transformation, β) somewhat broader than those of Strawderman (1978) and Casella (1980,1985). We then give a class of generalized hierarchical prior distributions on θ and σ^2 which give generalized Bayes estimators satisfying the minimaxity condition. This class generalizes (also to the class of unknown σ^2) the class of priors in Strawderman (1971), Lin and Tsai (1973), Berger (1976,1980) and Faith (1978). We further show that for certain choices of parameters in the hierarchy, the resulting estimators have the simple form indicated above. Section 3 is denoted to the study of general conditions under which an estimator competitive with $\hat{\beta}$ has increased numerical stability (i.e. decreased condition number). Section 4 is denoted to showing that we may always choose a simple generalized Bayes minimax estimator in our class which has greater numerical stability than the least squares estimator.

2 A class of minimax generalized Bayes estimators

In this section, we first give a sufficient condition for minimaxity and then use it to obtain a class of generalized Bayes minimax estimators. This class contains a sub-class

of a particularly simple form, which we hope, adds to the practical utility of our results. Our estimators are of the form

$$\hat{\theta}_\phi = \left(I - \frac{S}{X'C^{-1}D^{-1}X}\phi\left(\frac{X'C^{-1}D^{-1}X}{S}\right)C^{-1}\right)X \tag{2.1}$$

where $C = \text{diag}(c_1, \ldots, c_p)$ where $c_i \geq 1$ for any i.

First we give a sufficient condition for minimaxity.

Theorem 2.1. *$\hat{\theta}_\phi$ is minimax if $\phi'(w) \geq 0$ and*

$$0 \leq \phi(w) \leq 2(n+2)^{-1}\left(\frac{\sum(d_i/c_i)}{\max(d_i/c_i)} - 2\right).$$

Proof. The risk of $\hat{\theta}_\phi$ is given by

$$\begin{aligned} R(\theta, \sigma^2, \hat{\theta}_\phi) &= E\left[(\hat{\theta}_\phi - \theta)'(\hat{\theta}_\phi - \theta)/\sigma^2\right] \\ &= \sum d_i + E\left[\frac{S^2}{\sigma^2}\frac{\sum\{X_i^2/c_i^2\}}{(\sum\{X_i^2/(c_id_i)\})^2}\phi^2\left(\frac{\sum\{X_i^2/(c_id_i)\}}{S}\right)\right] \\ &\quad - 2E\left[\sum\frac{S}{\sigma^2}\frac{X_i}{c_i}(X_i - \theta_i)\sum\frac{X_i^2}{c_id_i}\phi\left(\frac{\sum\{X_i^2/(c_id_i)\}}{S}\right)\right]. \end{aligned} \tag{2.2}$$

Let $W = X'C^{-1}D^{-1}X/S$. For the second term in (2.2), using chi-square identity

$$E[\chi_n^2 h(\chi_n^2)] = nE[h(\chi_n^2)] + 2E[\chi_n^2 h'(\chi_n^2)]$$

(See for example Efron and Morris (1976)), we have

$$\begin{aligned} &E\left[\frac{X'C^{-2}X}{(X'C^{-1}D^{-1}X)^2}\frac{S}{\sigma^2}\left(S\phi^2\left(\frac{X'C^{-1}D^{-1}X}{S}\right)\right)\right] \\ &= E\left[\frac{X'C^{-2}X}{(X'C^{-1}D^{-1}X)^2}\left(nS\phi^2(W) + 2S\phi^2(W) - 4\phi(W)\phi'(W)X'C^{-1}D^{-1}X\right)\right] \\ &= E\left[\frac{X'C^{-2}X}{X'C^{-1}D^{-1}X}\left((n+2)\frac{\phi^2(W)}{W} - 4\phi(W)\phi'(W)\right)\right]. \end{aligned}$$

For the third term in (2.2), using the Stein identity, we have

$$\begin{aligned} &\sum E\left[\frac{1}{c_i\sigma^2}(X_i - \theta_i)X_i\left(\frac{\sum\{X_i^2/(c_id_i)\}}{S}\right)^{-1}\phi\left(\frac{\sum\{X_i^2/(c_id_i)\}}{S}\right)\right] \\ &= \sum E\left[\frac{d_i}{c_i}\left(W^{-1}\phi(W) + 2\frac{X_i^2}{c_id_iS}\left(W^{-1}\phi'(W) - W^{-2}\phi(W)\right)\right)\right] \\ &= E\left[\sum\frac{d_i}{c_i}\frac{\phi(W)}{W} + 2\frac{X'C^{-2}X}{S}\left(\frac{\phi'(W)}{W} - \frac{\phi(W)}{W^2}\right)\right]. \end{aligned}$$

Hence since $\phi'(w) \geq 0$, we have

$$
\begin{aligned}
&R(\theta, \sigma^2, \delta_{\phi,c}) \\
&\leq \sum d_i + E\left[\frac{\phi(W)}{W}\frac{X'C^{-2}X}{X'C^{-1}D^{-1}X}\left((n+2)\phi(W) - 2\sum\frac{d_i}{c_i}\frac{X'C^{-1}D^{-1}X}{X'C^{-2}X} + 4\right)\right] \\
&\leq \sum d_i + E\left[\frac{\phi(W)}{W}\frac{X'C^{-2}X}{X'C^{-1}D^{-1}X}\left((n+2)\phi(W) - 2\frac{\sum\{d_i/c_i\}}{\max\{d_i/c_i\}} + 4\right)\right] \\
&\leq \sum d_i.
\end{aligned}
$$

□

Next, consider the following generalized prior distribution:

$$
\begin{aligned}
&\theta|\lambda,\eta \sim N_p\left(0, \eta^{-1}D(\lambda^{-1}C - I)\right), \quad \text{for } \eta = \sigma^{-2}, \\
&\lambda \propto \lambda^a(1-\gamma\lambda)^b I_{[0,1/\gamma]}, \quad \text{for } \gamma \geq 1, \qquad \eta \propto \eta^e.
\end{aligned} \tag{2.3}
$$

This is a generalization of prior considered in Strawderman (1971), Lin and Tsai (1973), Berger (1976, 1980) and Faith (1978). The marginal density of X, S, λ and η is proportional to

$$
\begin{aligned}
&\int \exp\left(-\frac{\eta}{2}\sum\frac{(x_i-\theta_i)^2}{d_i} - \frac{\eta}{2}\sum\frac{\lambda}{c_i-\lambda}\frac{\theta_i^2}{d_i} - \frac{\eta s}{2}\right)\eta^{p+n/2}\lambda^{p/2} \\
&\quad \prod\left(d_i^{-1/2}(c_i-\lambda)^{-1/2}\right)\lambda^a(1-\gamma\lambda)^b\eta^e d\theta \\
&= \int \exp\left(-\frac{\eta}{2}\sum\frac{c_i}{(c_i-\lambda)d_i}(\theta_i - (1-\lambda/c_i)x_i)^2 - \frac{\eta}{2}\lambda\sum\frac{x_i^2}{c_i d_i} - \frac{\eta s}{2}\right) \\
&\quad \eta^{p+n/2+e}\lambda^{p/2+a}\prod\left(d_i^{-1/2}(c_i-\lambda)^{-1/2}\right)(1-\gamma\lambda)^b d\theta \\
&\propto \exp\left(-\frac{\eta}{2}\lambda\sum\frac{x_i^2}{c_i d_i} - \frac{\eta s}{2}\right)\eta^{p/2+n/2+e}\lambda^{p/2+a}(1-\gamma\lambda)^b.
\end{aligned} \tag{2.4}
$$

Under the loss $(\delta-\theta)'(\delta-\theta)/\sigma^2$, the generalized Bayes estimator is given by $E(\eta\theta|X,S)/E(\eta|X,S)$, which can be written, using (2.4),

$$
\hat{\theta}_{GB} = \left(I - \frac{E(\lambda\eta|X,S)}{E(\eta|X,S)}C^{-1}\right)X = \left(I - \frac{\phi_{GB}(W)}{W}C^{-1}\right)X,
$$

where $W = X'C^{-1}D^{-1}X/S$. When $p/2+n/2+e+2 > 0$,

$$
\int_0^\infty \eta^{p/2+n/2+e+1}\exp\left(-\frac{\eta}{2}\lambda\sum\frac{x_i^2}{c_i d_i} - \frac{\eta s}{2}\right)d\eta \propto (1+\lambda w)^{-p/2-n/2-e-2}, \tag{2.5}
$$

and we have

$$
\begin{aligned}
\phi_{GB}(w) = w\frac{E(\eta\lambda|X,S)}{E(\eta|X,S)} &= w\frac{\int_0^{1/\gamma}\lambda^{p/2+a+1}(1-\gamma\lambda)^b(1+w\lambda)^{-p/2-n/2-e-2}d\lambda}{\int_0^{1/\gamma}\lambda^{p/2+a}(1-\gamma\lambda)^b(1+w\lambda)^{-p/2-n/2-e-2}d\lambda} \\
&= \frac{w}{\gamma}\frac{\int_0^1 t^{p/2+a+1}(1-t)^b(1+wt/\gamma)^{-p/2-n/2-e-2}dt}{\int_0^1 t^{p/2+a}(1-t)^b(1+wt/\gamma)^{-p/2-n/2-e-2}dt},
\end{aligned} \tag{2.6}
$$

which is well-defined for $a > -p/2 - 1$ and $b > -1$. Using an identity, which is given by change of variables $t = (1+w)\lambda/(1+w\lambda)$

$$\int_0^1 \lambda^\alpha (1-\lambda)^\beta (1+w\lambda)^{-\gamma} d\lambda = \frac{1}{(w+1)^{\alpha+1}} \int_0^1 t^\alpha (1-t)^\beta \{1 - \frac{tw}{w+1}\}^{-\alpha-\beta+\gamma-2} dt,$$

we have

$$\phi_{GB}(w) = w\frac{E(\eta\lambda|X,S)}{E(\eta|X,S)} = \frac{w}{\gamma+w} \frac{\int_0^1 t^{p/2+a+1}(1-t)^b \{1-tw/(w+\gamma)\}^{n/2+e-a-1-b} dt}{\int_0^1 t^{p/2+a}(1-t)^b \{1-tw/(w+\gamma)\}^{n/2+e-a-b} dt}. \tag{2.7}$$

We have the following lemma.

Lemma 2.2. *If $b \geq 0$, $e > -p/2 - n/2 - 2$ and $-p/2 - 1 < a < n/2 + e$, we have for $\phi_{GB}(w)$ given by (2.7),*

1. *$\phi(w)$ is monotone increasing in w.*
2. *$\phi(w)/w$ is monotone decreasing in w.*
3. *$\lim_{w\to\infty} \phi(w) = (p/2+a+1)/(n/2+e-a)$.*

Proof. The proof of (i) and (ii) is straightforward using monotone likelihood ratio properties of the densities implied in (2.6) and (2.7). The proof of (iii) follows from (2.7). □

By Lemma 2.2, parts (i) and (ii) and Theorem 2.1, we have immediately the following result.

Theorem 2.3. *If $b \geq 0$, $e > -p/2 - n/2 - e - 2$ and $-p/2 - 1 < a < n/2 + e$, then $\hat{\theta}_{GB}$ is minimax provided $c_1, \ldots, c_p$ are chosen so that*

$$0 \leq \frac{p/2+a+1}{n/2+e-a} \leq \frac{2}{n+2}\left(\frac{\sum\{d_i/c_i\}}{\max\{d_i/c_i\}} - 2\right).$$

Note: If we choose $c_i = d_i/d_p$ the bound on the RHS is $2(p-2)/(n+2)$. The choice of $a = -2$ and $e = -1$ give a value of $(p-2)/(n+2)$ for the LHS and hence for $p \geq 3$ and $n \geq 1$ these choice of a and e give minimax generalized Bayes estimators for any $b \geq 0$ and $\gamma \geq 1$. As Casella (1980,1985) indicated, this choice of c_i may be poor from the point of view of the numeric stability of the estimator. We consider this point further in Section 3. In that section, γ will play a role in the stability of the estimator.

2.1 A class of simple generalized Bayes minimax estimators

When $b = n/2-a+e-1$ in equation (2.7), the expression for $\phi_{GB}(w)$ takes a particularly simple form. In this case,

$$\begin{aligned}\phi_{GB}(w) &= \frac{w}{w+\gamma}\frac{B(p/2+a+2,b+1)}{B(p/2+a+1,b+1)-\{w/(w+\gamma)\}B(p/2+a+2,b+1)} \\ &= \frac{w}{(w+\gamma)(1+1/\alpha)-w} \\ &= \frac{\alpha w}{\gamma(\alpha+1)+w} \end{aligned} \tag{2.8}$$

where $\alpha = (p/2+a+1)/(b+1) = (p/2+a+1)/(n/2+e-a)$.

Therefore our simple generalized Bayes estimator is

$$\hat{\theta}_{SB} = \left(I - \frac{\alpha}{\gamma(\alpha+1)+W}C^{-1}\right)X. \tag{2.9}$$

Hence we have the following corollary which follows immediately from Theorem 2.3.

Corollary 2.4. *$\hat{\theta}_{SB}$ given by (2.9) is minimax provided $c_1, \ldots, c_p$ is chosen so that*

$$0 < \alpha \leq \frac{2}{n+2}\left(\frac{\sum\{d_i/c_i\}}{\max\{d_i/c_i\}} - 2\right).$$

It is interesting to note that when $C = D = I_p$, our simple estimator has the form

$$\hat{\theta}_{SB} = \left(1 - \frac{\alpha}{\gamma(\alpha+1)+X'X/S}\right)X.$$

This is very closely related to Stein's (1956) initial class of estimators. He suggested that for $X \sim N(\theta, I_p)$ with $p \geq 3$, there exist estimators dominating the usual estimator X among a class of estimators of the form $\delta_{a,b} = (1 - b/(a+X'X))X$ for large a and small b. Hence our estimators may be regarded as a variant for unknown variance case.

Following Stein (1956), James and Stein (1961) showed that $\delta_{a,b}$ for $a = 0$ and $0 < b < 2(p-2)$ dominates X. Since Strawderman (1971) derived Bayes minimax estimators, many authors have proposed various minimax (generalized) Bayes estimators. However the form of these estimators is invariably complicated like our expression (2.7) above. Simple estimators $\delta_{a,b}$ have received little attention although $\delta_{a,b}$ for $a > 0$ and $0 < b < 2(p-2)$ is easily shown to be minimax by using Baranchik's (1970) condition. It seems that most statisticians have believed that generalized Bayes estimators which improve on X must have a quite complicated structure. Our result above indicates that this is not so and that generalized Bayes minimax estimators, improving on X may indeed have a very simple form.

3 Condition numbers and numerical stability

As in Casella (1985) and other papers, we use the condition number to measure numerical stability of our ridge-type estimators. This discussion focuses on the stability of estimators of β (as opposed to estimators of θ). Recall that our estimators of θ may be represented as $\hat{\theta}_\phi = (I - tC^{-1})X$ where $t = \phi(w)/w$ and $w = X'C^{-1}D^{-1}X/S$. The vector of regression parameters, β, is related to the mean vector θ through the orthogonal matrix P ($\theta = P'\beta$) and the observation vector X in Section 2 is related to the least squares estimator, $\hat{\beta}$, through $X = P'\hat{\beta}$. In this section, we are interested in studying the numerical stability of ridge-type estimators of $\hat{\beta}_\phi$, arising from our improved estimators $\hat{\theta}_\phi$ of θ through

$$\begin{aligned}\hat{\beta}_\phi = P\hat{\theta}_\phi &= P(I - tC^{-1})X \\ &= P(I - tC^{-1})P'\hat{\beta} \\ &= P(I - tC^{-1})P'(A'A)^{-1}A'y \\ &= P(I - tC^{-1})P'PDP'A'y \\ &= P(\text{diag}\{d_i^{-1}(1 - t/c_i)^{-1}\})^{-1}P'A'y \\ &= G^{-1}A'y.\end{aligned} \tag{3.1}$$

The condition number of a matrix H is defined by $\kappa(H) = \|H\|\|H^{-1}\|$ where $\|H\| = \sup_{x'x=1}(x'H'Hx)^{1/2} = \max \lambda_i$, where λ_i are the eigen values of the positive-definite matrix $H'H$. It follows that if H is a positive-definite matrix, $\kappa(H) = \kappa(H^{-1})$. As indicated in Casella (1985) (See also Belsley, Kuh and Welsch (1980)), the condition number measures the numerical sensitivity of the solution of a linear equation $\hat{\beta} = H^{-1}A'y$. In particular if $\delta\hat{\beta}$ and $\delta(A'y)$ indicate perturbations in $\hat{\beta}$ and $A'y$ respectively,

$$|\delta\hat{\beta}|/|\hat{\beta}| \leq \kappa(H)(|\delta A'y|/|A'y|),$$

where $|\cdot|$ denotes the usual Euclidean norm. For simplicity of notation, we define the condition number of an estimator of the form (3.1) $\kappa(\hat{\beta}_\phi)$ to be equal to the condition number of the matrix G^{-1}, $\kappa(G^{-1}) = \kappa(G)$, i.e. $\kappa(\hat{\beta}_\phi) = \kappa(G)$.

It follows immediately from the definition of $\kappa(G)$ that (we assume $t \leq 1$, $c_i \geq 1$)

$$\kappa(\hat{\beta}) = d_1/d_p \tag{3.2}$$

and

$$\kappa(\hat{\beta}_\phi) = \frac{\max d_i(1 - t/c_i)}{\min d_i(1 - t/c_i)}. \tag{3.3}$$

In terms of numerical stability, a smaller condition number implies greater stability.

Of course, the condition number gives in (3.3) depends on $t = \phi(w)/w$ and in particular when $w = \infty$, $t = 0$ and (3.3) reduces to (3.2). We will be interested in finding conditions

on our generalized Bayes estimator so that for all possible values of t we have inequality $\kappa(\hat{\beta}_\phi) \leq \kappa(\hat{\beta})$.

The following result allows condition number improving generalized Bayes estimators under two different conditions on $c_1, \ldots, c_p$.

Theorem 3.1. *Suppose $\phi(w)/w$ is monotone decreasing and suppose $\lim_{w\to 0} \phi(w)/w = t_0 < 1$. Then $\kappa(\hat{\beta}_\phi) \leq \kappa(\hat{\beta})$ for any $t \in [0, t_0]$ if either*

1. *if $c_p > c_1 \geq c_2 \geq \cdots \geq c_{p-1}$ and*

$$t_0 \leq \min\left(\frac{c_1 c_{p-1}(d_{p-1} - d_p)}{c_1 d_{p-1} - c_{p-1} d_p}, \frac{c_{p-1} c_p (d_1 d_{p-1} - d_p^2)}{c_p d_1 d_{p-1} - c_{p-1} d_p^2}\right)$$

or

2. *if $c_1 \leq c_2 \leq \cdots \leq c_p$ and*

$$t_0 \leq \min_{i>j}\left(\frac{c_i c_j (d_1 d_j - d_i d_p)}{c_i d_1 d_j - c_j d_i d_p}\right),$$

Proof. Suppose $c_p > c_1 \geq c_2 \geq \cdots \geq c_{p-1}$. Then $d_1(1 - t/c_1) \geq \cdots \geq d_{p-1}(1 - t/c_{p-1})$ and so

$$\max_t \frac{\max_{i=1,\ldots,p-1} d_i(1 - t/c_i)}{\min_{j=1,\ldots,p-1} d_j(1 - t/d_j)} \leq \frac{d_1(1 - t_0/c_1)}{d_{p-1}(1 - t_0/c_{p-1})}. \tag{3.4}$$

Also

$$\max_t \frac{d_1(1 - t/c_1)}{d_p(1 - t/c_p)} \leq \frac{d_1}{d_p} \tag{3.5}$$

and

$$\max_t \frac{d_p(1 - t/c_p)}{d_{p-1}(1 - t/c_{p-1})} \leq \frac{d_p(1 - t_0/c_p)}{d_{p-1}(1 - t_0/c_{p-1})}. \tag{3.6}$$

Hence if

$$\max\left(\frac{d_1(1 - t_0/c_1)}{d_{p-1}(1 - t_0/c_p)}, \frac{d_p(1 - t_0/c_p)}{d_{p-1}(1 - t_0/c_{p-1})}\right) \leq \frac{d_1}{d_p}$$

or equivalently

$$t_0 \leq \min\left(\frac{c_1 c_{p-1}(d_{p-1} - d_p)}{c_1 d_{p-1} - c_{p-1} d_p}, \frac{c_{p-1} c_p (d_1 d_{p-1} - d_p^2)}{c_p d_1 d_{p-1} - c_{p-1} d_p^2}\right)$$

we have

$$\max_t \frac{\max_i d_i(1 - t/c_i)}{\min_j d_j(1 - t/c_j)} \leq \frac{d_1}{d_p},$$

which proves part (i).

If $c_1 \leq c_2 \leq \cdots \leq c_p$, we have

$$\frac{d_i(1 - t/c_i)}{d_j(1 - t/c_j)} \leq \frac{d_i}{d_j} \quad \text{for } i < j \tag{3.7}$$

and

$$\max_t \frac{d_i(1-t/c_i)}{d_j(1-t/c_j)} \leq \frac{d_i(1-t_0/c_i)}{d_j(1-t_0/c_j)} \quad \text{for } i > j. \tag{3.8}$$

Hence if

$$\max_{i>j}\left(\frac{d_i(1-t_0/c_i)}{d_j(1-t_0/c_j)}\right) \leq \frac{d_1}{d_p}$$

or equivalently

$$t_0 \leq \min_{i>j}\left(\frac{c_ic_j(d_1d_j - d_id_p)}{c_id_1d_j - c_jd_id_p}\right),$$

we have

$$\max_t \frac{\max_i d_i(1-t/c_i)}{\min_j d_j(1-t/c_j)} \leq \frac{d_1}{d_p}$$

which proves part (ii). □

In the next section, we will show that it is always possible to find a condition improving (simple) generalized Bayes minimax estimator.

4 Minimaxity and stability

In this section, we show that the results of the previous two sections can be combined to give simple generalized Bayes minimax estimators which simultaneously reduce the condition number relative to the least squares estimator.

Note that it seems generally desirable to have $c_1 \leq \cdots \leq c_p$ since this implies that the components of X with larger variances get shrunk more. See Casella (1985) for an expanded discussion of this point.

Our first result below shows that we may find generalized Bayes minimax condition number improving estimator satisfying $c_1 \leq \cdots \leq c_p$ whenever $\sum\{d_i/d_1\} - 2 > 0$.

Theorem 4.1. *Suppose $p \geq 3$ and $\sum\{d_i/d_1\} - 2 > 0$. If $d_1 > d_2$, let η_* be the unique root such that $\sum\{d_i/d_1\}^{\eta} = 2$ and let η_{**} be any value in $(1, \eta_*)$. If $d_1 = d_2$, let η_{**} be any value > 1. Then if $c_i = (d_1/d_i)^{\eta_{**}-1}$ and $\alpha \in (0, u_+]$ for*

$$u_+ = 2(n+2)^{-1}\left(\sum\{d_i/d_1\}^{\eta_{**}} - 2\right)$$

and if

$$\gamma \geq \frac{\alpha}{\alpha+1}\max_{i>j}\left(\frac{c_id_1d_j - c_jd_id_p}{c_ic_j(d_1d_j - d_id_p)}\right), \tag{4.1}$$

the estimator $\hat{\theta}_{SB}$ of Corollary 2.4 is generalized Bayes, minimax and condition number decreasing, further $c_1 \leq \cdots \leq c_p$.

Proof. Since $(d_i/d_1)^\eta$ is strictly decreasing in η if $d_i/d_1 < 1$. There exists exactly one root η_* of $\sum(d_i/d_1)^\eta = 2$ if $d_2/d_1 < 1$ and that root is strictly larger than 1. If $d_1 = d_2$ $\sum(d_i/d_1)^\eta > 2$ for any $\eta > 0$. Hence $\eta_{**} > 1$ and $c_i = (d_1/d_i)^{\eta_{**}-1}$ is monotone non-decreasing in i. Also from Corollary 2.4 we have minimaxity provided

$$0 < \alpha \leq \frac{2}{n+2}\left(\frac{\sum\{d_i/c_i\}}{\max\{d_i/c_i\}} - 2\right)$$
$$= \frac{2}{n+2}\left(\sum\{d_i/d_1\}^{\eta_{**}} - 2\right) = u_+ \ (> 0).$$

Also by Theorem 3.1 since $c_1 \leq c_2 \leq \cdots \leq c_p$ the generalized Bayes estimator will have reduced condition number provided

$$t_0 \leq \min_{i>j}\left(\frac{c_i c_j(d_1 d_j - d_i d_p)}{c_i d_1 d_j - c_j d_i d_p}\right). \tag{4.2}$$

Since $t_0 = \alpha/\{\gamma(\alpha+1)\}$ (recall $\hat{\theta} = (I - \alpha/\{\gamma(\alpha+1)+w\}C^{-1})X$), the condition number of (4.2) is seems to be equivalent to (4.1). □

There remains the case where $\sum\{d_i/d_1\} - 2 \leq 0$. This case corresponds to the case where no spherically symmetric estimator ($c_1 = c_2 = \cdots = c_p$) and therefore no estimator with $c_1 \leq c_2 \leq \cdots \leq c_p$ can be minimax (e.g. See Bock (1975)). Our solution while less pleasing in a sense than Theorem 4.1 nevertheless allows a simple minimax generalized Bayes estimator which reduces the condition number and hence increases the stability.

Theorem 4.2. *Suppose $p \geq 3$ and $\sum\{d_i/d_1\} - 2 \leq 0$. If $p \geq 4$ let $\nu_* \in (0,1)$ be the unique solution of $\sum_{i=1}^{p-1}\{d_i/d_1\}^\nu = 2$. Let ν_{**} be any value in $[0, \nu_*)$. If $p = 3$, choose $\nu_{**} = 0$. Then if $c_i = (d_i/d_{p-1})^{1-\nu_{**}}$ for $i = 1, 2, \ldots, p-1$ and $c_p > c_1$, $0 < \alpha \leq u_-$, for*

$$u_- = 2(n+2)^{-1}\left(\sum_{i=1}^{p-1}\{d_i/d_1\}^{\nu_{**}} - 2 + \frac{c_1 d_p}{c_p d_1}\right)$$

and

$$\gamma \geq \frac{\alpha}{\alpha+1}\max\left(\frac{c_1 d_{p-1} - c_{p-1}d_p}{c_1 c_{p-1}(d_{p-1} - d_p)}, \frac{c_p d_1 d_{p-1} - c_{p-1}d_p^2}{c_{p-1}c_p(d_1 d_{p-1} - d_p^2)}\right),$$

the estimator of Corollary 2.4 is a (simple) generalized Bayes minimax condition number improving estimator.

Proof. It is easy to see as in Theorem 4.1 that ν_*, ν_{**} can be chosen as indicated. In this case, Corollary 2.4 implies minimaxity provided

$$0 < \alpha \leq \frac{2}{n+2}\left(\frac{\sum\{d_i/c_i\}}{\max\{d_i/c_i\}} - 2\right) = \frac{2}{n+2}\left(\sum_{i=1}^{p-1}\{d_i/d_1\}^{\nu_{**}} - 2 + \frac{c_1 d_p}{c_p d_1}\right) = u_- \ (> 0).$$

Theorem 3.1 (i) then implies since $c_p > c_1 \geq c_2 \geq \cdots \geq c_{p-1}$ that our estimator is condition improving if γ is chosen so that

$$\gamma \geq \frac{\alpha}{\alpha+1} \max\left(\frac{c_1 d_{p-1} - c_{p-1} d_p}{c_1 c_{p-1}(d_{p-1} - d_p)}, \frac{c_p d_1 d_{p-1} - c_{p-1} d_p^2}{c_{p-1} c_p (d_1 d_{p-1} - d_p^2)} \right).$$

□

We note that versions of Theorems 4.1 and 4.2 are valid also for the broader class of generalized Bayes minimax estimators of Theorem 2.3. We omit the straightforward details.

参 考 文 献

[1] Baranchik, A.J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.*, **41**, 642-645.

[2] Belsley, D.A., Kuh, E. and Welsch, R.E. (1980). *Regression diagnostics.* John Wiley, New York.

[3] Berger, J. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.*, **4**, 223-226.

[4] Berger, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.*, **8**, 716-761.

[5] Casella, G. (1980). Minimax ridge regression estimation. *Ann. Statist.*, **8**, 1036-1056.

[6] Casella, G. (1985). Condition numbers and minimax ridge regression estimators. *J. Amer. Statist. Assoc.*, **80**, 753-758.

[7] Bock, M.E. (1975). Minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.*, **3**, 209-218.

[8] Efron, B. and Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.*, **4**, 11-21.

[9] Faith, R.E. (1978). Minimax Bayes estimators of a multivariate normal mean. *J. Multivariate Anal.*, **8**, 372-379.

[10] Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55-68

[11] James, W. and Stein, C. (1961). Estimation of quadratic loss. *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, Vol.1, 361-379, Univ. of California Press, Berkeley.

[12] Lin, P. and Tsai, H. (1973). Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix. *Ann. Statist.*, **1**, 142-145.

[13] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. 3rd Berkeley Symp. Math. Statist. Prob.*, Vol.1,197-206, Univ. of California Press, Berkeley.

[14] Strawderman, W.E. (1971). Proper Bayes minimax estimators of multivariate normal mean. *Ann. Math. Statist.*, **42**, 385-388.

[15] Strawderman, W.E. (1978). Minimax adaptive generalized ridge regression estimators. *J. Amer. Statist. Assoc.*, **73**, 623-627.