Ann. Inst. Statist. Math.
Vol. 41, No. 3, 485-502 (1989)

SIMULTANEOUS ESTIMATION OF MEANS OF CLASSIFIED NORMAL OBSERVATIONS

TOYOAKI AKAI

Department of Mathematics, Faculty of Science and Technology, Keio University, Yokohama 223, Japan

(Received June 8, 1987; revised September 29, 1988)

Abstract. Simultaneous estimation of normal means is considered for observations which are classified into several groups. In a one-way classification case, it is shown that an adaptive shrinkage estimator dominates a Stein-type estimator which shrinks observations towards individual class averages as Stein's (1966, *Festschrift for J. Neyman*, (ed. F. N. David), 351-366, Wiley, New York) does, and is minimax even if class sizes are small. Simulation results under quadratic loss show that it is slightly better than Stein's (1966) if between variances are larger than within ones. Further this estimator is shown to improve on Stein's (1966) with respect to the Bayes risk. Our estimator is derived by assuming the means to have a one-way classification structure, consisting of three random terms of grand mean, class mean and residual. This technique can be applied to the case where observations are classified into a two-stage hierarchy.

Key words and phrases: Bayes estimator, normal means, sum of squared error loss, shrinkage estimator, Stein estimator.

1. Introduction

Let $\boldsymbol{Y} = (Y_1,\dots, Y_p)'$ have a p-dimensional multivariate normal distribution

$$\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, I)\,, \tag{1.1}$$

with unknown mean $\boldsymbol{\mu} = (\mu_1,\dots,\mu_p)'$ and the identity covariance matrix I. An estimator $\hat{\boldsymbol{\mu}}(\boldsymbol{Y}) = (\hat{\mu}_1(\boldsymbol{Y}),\dots,\hat{\mu}_p(\boldsymbol{Y}))'$ of $\boldsymbol{\mu}$ is evaluated by the risk function

$$R(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}) = E_\mu\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2\,,$$

where $\|\hat{\boldsymbol{\mu}}(\boldsymbol{Y})-\boldsymbol{\mu}\|^2=\sum_{i=1}^{p}(\hat{\mu}_i(\boldsymbol{Y})-\mu_i)^2$ and E_μ stands for averaging over the sample space with respect to the distribution (1.1).

James and Stein (1961) showed that if $p \geq 3$ the maximum likelihood estimator $\boldsymbol{Y}$ is dominated by the estimator

$$\hat{\boldsymbol{\mu}}^{SO}=(1-(p-2)/\|\boldsymbol{Y}\|^2)\boldsymbol{Y}\,.$$

Several drawbacks of $\hat{\boldsymbol{\mu}}^{SO}$ have been pointed out and many efforts have been made to improve them. One major drawback is that the region of the parameter space where the risk of $\hat{\boldsymbol{\mu}}^{SO}$ (or some other estimators of a similar type) is significantly smaller than that of $\boldsymbol{Y}$ is quite limited (see Stein (1981) and Berger (1982)).

Improvements over $\boldsymbol{Y}$ when observations are classified into several groups are considered below. Stein (1966) discussed Stein-type estimators for designs admitting a completely orthogonal analysis of variance: complete k-way classifications, Latin squares and Greco-Latin squares. Two-way classification case was studied in detail. He showed that if sample sizes are very large, the Stein-type estimator applied separately to each orthogonal subspace is approximately better than the estimator which shrinks observations towards the general average. Takeuchi (1980) proposed an estimator very similar to Stein's (1966) when observations are classified as one-way or two-way.

Haff (1978) considered the estimator of normal means in situations where means are close to each other. He assumed that means have normal distributions with intraclass correlation structure, in which diagonal elements of the same value are larger than nondiagonal elements of the same value, and derived a Bayes estimator. Replacing hyperparameters in it by the mode of posterior density, he obtained a minimax estimator, which is a modification of a Stein-type estimator, which shrinks observations towards the grand average.

From another viewpoint, Efron and Morris (1973) considered the estimation of normal means which are divided into two groups with different prior variances. They showed that the Stein estimator applied separately to the two groups is better than the estimator applied to the two combined if the prior variances are largely different, and is worse if the prior variances are almost the same. Based on this consideration, they proposed a compromise estimator and improved the Bayes risk of the two Stein estimators in a wide region of the ratios of prior variances. Berger and Dey (1983) considered the estimation of normal means which are divided into k groups with k different prior variances. It was shown that a Stein-type estimator applied to the combined group often dominates, with respect to the Bayes risk for normal or flat-tailed priors, the estimator applied separately to each group.

George (1986*a*, 1986*b*, 1986*c* and 1986*d*) considered situations where only conflicting or vague prior information is available. For example, when more than one of a broad class of minimax shrinkage estimators might be effective, he proposed new minimax estimators, called multiple shrinkage Stein estimators, which are derived from posterior means of mixture priors. An estimator of this class is an adaptive convex combination of shrinkage estimators which gives more weight to the estimator which shrinks most.

The main objective of this paper is to propose a class of new estimators in the estimation of components of $\boldsymbol{\mu}$ when observations are classified into A groups of size B, and to show their goodness. Let Y_{ij} be independently and normally distributed;

$$(1.2) \qquad Y_{ij} \sim N(\mu_{ij}, 1^2); \quad i = 1,\ldots,A; \quad j = 1,\ldots,B\,.$$

If the class means $\bar{\mu}_{i\cdot}$'s are almost the same, the Stein-type estimator which shrinks Y_{ij}'s towards the grand average,

$$\hat{\mu}_{ij}^{ST} = \bar{Y}_{\cdot\cdot} + \left(1 - (AB-3) \Big/ \left(\sum_i \sum_j (Y_{ij} - \bar{Y}_{\cdot\cdot})^2\right)\right)(Y_{ij} - \bar{Y}_{\cdot\cdot})\,,$$

where $\bar{Y}_{\cdot\cdot} = \sum_i \sum_j Y_{ij}/AB$ will have smaller risk than the Stein-type estimator which shrinks Y_{ij}'s towards individual class averages,

$$\hat{\mu}_{ij}^{SG} = \bar{Y}_{i\cdot} + \left(1 - 1 \Big/ \left(b \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\cdot})^2\right)\right)(Y_{ij} - \bar{Y}_{i\cdot})\,,$$

where $\bar{Y}_{i\cdot} = \sum_j Y_{ij}/B$ and $0 < 1/b < 2(A(B-1)-2)$. On the other hand, if $\bar{\mu}_{i\cdot}$'s are largely different, $\hat{\boldsymbol{\mu}}^{SG}$ will have smaller risk than $\hat{\boldsymbol{\mu}}^{ST}$. We construct an adaptive shrinkage estimator which behaves like $\hat{\boldsymbol{\mu}}^{ST}$ when $\bar{\mu}_{i\cdot}$'s are close, and like $\hat{\boldsymbol{\mu}}^{SG}$ when they are different.

To derive such an estimator, assume a one-way classification structure of $\boldsymbol{\mu}$;

$$(1.3) \qquad \mu_{ij} = \mu + \alpha_i + e_{ij}; \quad i = 1,\ldots,A; \quad j = 1,\ldots,B\,;$$

where $\mu \sim N(\omega, \sigma_\mu^2)$, $\alpha_i \sim N(0, \sigma_\alpha^2)$, $e_{ij} \sim N(0, \sigma_e^2)$ and these are independently distributed. A limit of the Bayes estimator as $\sigma_\mu^2 \to \infty$ is derived. We estimate $1/(B\sigma_\alpha^2 + \sigma_e^2 + 1)$ and $1/(\sigma_e^2 + 1)$ in the limit estimator by suitable statistics satisfying the order relationship. The estimator $\hat{\boldsymbol{\mu}}^{1*}$ of (2.12), which is an adaptive convex combination of $\hat{\boldsymbol{\mu}}^{ST}$ and $\hat{\boldsymbol{\mu}}^{SG}$ obtained by different estimates of hyper-parameters of variances, is proposed in Subsection 2.2. This estimator is different from George's which is a convex combination of shrinkage estimators based on unbiased estimates of risk reductions. But it is similar to Stein and Takeuchi's $\hat{\boldsymbol{\mu}}^{SS*}$ of (2.13) which may be regarded as an empirical Bayes estimator for the above prior distribution. It is shown that if $A \geq 4$ the estimator $\hat{\boldsymbol{\mu}}^{1*}$ dominates $\hat{\boldsymbol{\mu}}^{SG}$ as $\hat{\boldsymbol{\mu}}^{SS*}$ does. The estimator $\hat{\boldsymbol{\mu}}^{1*}$ is minimax even if $A = 2$ or 3, the case where $\hat{\boldsymbol{\mu}}^{SS*}$ cannot be constructed. Further it is shown that $\hat{\boldsymbol{\mu}}^{1*}$ improves on $\hat{\boldsymbol{\mu}}^{SS*}$ with respect to the Bayes risk.

In Section 2, the estimator $\hat{\boldsymbol{\mu}}^{1*}$ is proposed and analyzed as stated above. In the case where A and B are small, we calculate Monte Carlo estimates of risk functions of positive part versions of $\hat{\boldsymbol{\mu}}^{1*}$, $\hat{\boldsymbol{\mu}}^{SS*}$, $\hat{\boldsymbol{\mu}}^{ST}$ and $\hat{\boldsymbol{\mu}}^{SG}$ for different combinations of between and within variances. It is observed that $\hat{\boldsymbol{\mu}}^{1*}$ is slightly better than the others if between variances are larger than within ones, and $\hat{\boldsymbol{\mu}}^{SS*}$ is the best if within variances are larger than between ones.

In Section 3, we estimate components of $\boldsymbol{\mu}$ when observations are classified into A classes, each of which can be divided into B subclasses consisting of C components. Similarly to Section 2, an adaptive minimax estimator is obtained, which is a convex combination of Stein-type estimators which shrink observations towards the grand averages, towards individual class averages or towards individual subclass averages. Furthermore we estimate components of $\boldsymbol{\mu}$ in a special case where observations are classified into two classes, each of which can be divided into two subclasses of different sizes. This is the case where Stein and Takeuchi's estimator cannot be constructed. The estimator (3.6) is proposed and applied to the estimation of American baseball team batting averages, which was discussed by George (1986*b*). The sum of squared error loss, called actual losses, for the estimator (3.6), George's, which combines three estimators (3.7), (3.8) and (3.9) to be a competitor to the estimator (3.6), and the others are compared. The estimator (3.6) has a smaller actual loss than the others.

2. One-way classification case

Let Y_{ij} be distributed as (1.2) and let its mean μ_{ij} be distributed as (1.3). We also consider the case where μ_{ij} has the structure

$$\mu_{ij} = \alpha_i + e_{ij}\,, \tag{2.1}$$

where $\alpha_i \sim N(0, \sigma_a^2)$, $e_{ij} \sim N(0, \sigma_e^2)$ and these are independently distributed. The assumption on $\boldsymbol{\mu}$ is a special case discussed by Lindley and Smith (1972) to derive the Bayes estimator under a linear multiple regression model for the means.

2.1 *Shrinkage towards zero*

Suppose that (1.2) and (2.1) hold. Write (2.1) in the matrix form,

$$\mu = X\alpha + e , \tag{2.2}$$

where $\alpha = (\alpha_1,\dots,\alpha_A)'$ and X is an $AB \times A$ design matrix with 1 for the (i,j) elements, $B(j-1)+1 \le i \le Bj$, and 0 for the other elements. For simple derivation of the Bayes estimator, construct an AB-dimensional orthogonal matrix P, of which the first A column vectors are normalized columns of X and the remainder columns are arbitrary. Consider the orthogonal transformation

$$Z = P'Y \quad \text{and} \quad \theta = P'\mu ,$$

where $Z = (Z_1,\dots,Z_{AB})'$ and $\theta = (\theta_1,\dots,\theta_{AB})'$. The prior distribution of θ is

$$\theta \sim N_{AB}(0, \Sigma_1) , \tag{2.3}$$

where Σ_1 is a diagonal matrix, of which the first A diagonal elements are $B\sigma_\alpha^2 + \sigma_e^2$ and the others are σ_e^2. The Bayes estimator under squared error loss is

$$\hat{\theta}_i^B = \begin{cases} (1 - 1/(B\sigma_\alpha^2 + \sigma_e^2 + 1))Z_i, & \text{if} \quad 1 \le i \le A , \\ (1 - 1/(\sigma_e^2 + 1))Z_i, & \text{if} \quad A+1 \le i \le AB . \end{cases} \tag{2.4}$$

The marginal distribution of Z is

$$Z \sim N_{AB}(0, \Sigma_2) , \tag{2.5}$$

where Σ_2 is a diagonal matrix, of which the first A diagonal elements are $B\sigma_\alpha^2 + \sigma_e^2 + 1$ and the others are $\sigma_e^2 + 1$. Put $S_{11} = \sum_{i=1}^{A} Z_i^2$ and $S_{12} = \sum_{i=A+1}^{AB} Z_i^2$. Suppose that some estimators $\hat{\sigma}_1^2(S_{11}, S_{12})$ and $\hat{\sigma}_2^2(S_{11}, S_{12})$ can be used for $1/(B\sigma_\alpha^2 + \sigma_e^2 + 1)$ and $1/(\sigma_e^2 + 1)$ in (2.4), respectively. The following component estimator is obtained;

$$\hat{\theta}_i^{SB} = \begin{cases} (1 - \hat{\sigma}_1^2(S_{11}, S_{12}))Z_i, & \text{if} \quad 1 \le i \le A , \\ (1 - \hat{\sigma}_2^2(S_{11}, S_{12}))Z_i, & \text{if} \quad A+1 \le i \le AB . \end{cases} \tag{2.6}$$

Replacing $\hat{\sigma}_1^2(\cdot,\cdot)$ and $\hat{\sigma}_2^2(\cdot,\cdot)$ in $\hat{\theta}^{SB}$ by $1/aS_{11}$ and $1/bS_{12}$ (a and b are positive constants to be suitably chosen), respectively, we obtain Stein's (1966) and Takeuchi's (1980) estimator $\hat{\theta}^{SS}$ with components

$$\hat{\theta}_i^{SS} = \begin{cases} (1 - 1/aS_{11})Z_i, & \text{if} \quad 1 \le i \le A , \\ (1 - 1/bS_{12})Z_i, & \text{if} \quad A+1 \le i \le AB . \end{cases} \tag{2.7}$$

It is trivial that $\hat{\boldsymbol{\theta}}^{SS}$ dominates $\hat{\boldsymbol{\theta}}^{SG} = \boldsymbol{P}'\hat{\boldsymbol{\mu}}^{SG}$ if $0 < 1/a < A-2$ and $0 < 1/b < 2(A(B-1)-2)$. However, against the order relationship $1/(B\sigma_a^2 + \sigma_e^2 + 1) < 1/(\sigma_e^2 + 1)$, it happens that $1/aS_{11} > 1/bS_{12}$ in $\hat{\boldsymbol{\theta}}^{SS}$. Observing $1/aS_{11} > 1/bS_{12}$ suggests a violation of (2.1) or (2.3). According to ideas on estimates in Barlow *et al.* (1972) satisfying the order relationship, we obtain the estimator

$$\hat{\theta}_i^1 = \begin{cases} (1 - 1/aS_{11})Z_i I_{[aS_{11} \geq bS_{12}]}(\boldsymbol{Z}) \\ \quad + (1 - (1/a + 1/b)/(S_{11} + S_{12}))Z_i I_{[aS_{11} < bS_{12}]}(\boldsymbol{Z}) , \\ \qquad\qquad\qquad \text{if} \quad 1 \leq i \leq A , \\ (1 - 1/bS_{12})Z_i I_{[aS_{11} \geq bS_{12}]}(\boldsymbol{Z}) \\ \quad + (1 - (1/a + 1/b)/(S_{11} + S_{12}))Z_i I_{[aS_{11} < bS_{12}]}(\boldsymbol{Z}) , \\ \qquad\qquad\qquad \text{if} \quad A+1 \leq i \leq AB , \end{cases} \tag{2.8}$$

where

$$I_{[aS_{11} \geq bS_{12}]}(\boldsymbol{Z}) = \begin{cases} 1, & \text{if} \quad aS_{11} \geq bS_{12} , \\ 0, & \text{otherwise} . \end{cases}$$

Inverting $\hat{\boldsymbol{\theta}}^{SB}$, we get

$$\begin{aligned} \hat{\mu}_{ij}^{SB} &= \rho(S_{11}, S_{12})(1 - \hat{\sigma}_1^2(S_{11}, S_{12}))Y_{ij} \\ &\quad + (1 - \rho(S_{11}, S_{12}))(\overline{Y}_{i\cdot} + (1 - \hat{\sigma}_2^2(S_{11}, S_{12}))(Y_{ij} - \overline{Y}_{i\cdot})) , \end{aligned} \tag{2.9}$$

where $\rho(S_{11}, S_{12}) = \hat{\sigma}_1^2(S_{11}, S_{12})/\hat{\sigma}_2^2(S_{11}, S_{12})$. Note that $S_{11} = B\sum_{i=1}^{A} \overline{Y}_{i\cdot}^2$ and $S_{12} = \sum_i \sum_j (Y_{ij} - \overline{Y}_{i\cdot})^2$. The estimator $\hat{\boldsymbol{\mu}}^1 = \boldsymbol{P}\hat{\boldsymbol{\theta}}^1$ is a convex combination of $\hat{\boldsymbol{\mu}}^{SG}$ and a Stein-type estimator which shrinks Y_{ij}'s towards zero. The estimator $\hat{\boldsymbol{\mu}}^1$ is the same as $\hat{\boldsymbol{\mu}}^{SS} = \boldsymbol{P}\hat{\boldsymbol{\theta}}^{SS}$ in the case $\rho(\cdot, \cdot) \leq 1$, and is the same as the Stein estimator $\hat{\boldsymbol{\mu}}^{SO}$ in the case $\rho(\cdot, \cdot) > 1$.

To get conditions on the constant a such that the risk of the estimator $\hat{\boldsymbol{\mu}}^1$ is smaller than that of $\hat{\boldsymbol{\mu}}^{SG}$, we use the following lemma which is well known and can be verified by integrating by parts.

LEMMA 2.1. *Let Z be a standard normal random variable. If $f(\cdot)$ is absolutely continuous, then $E_\theta\{f(Z)(Z-\theta)\} = E_\theta\{f'(Z)\}$.*

THEOREM 2.1. *Suppose that $A \geq 3$. Then the risk of $\hat{\boldsymbol{\mu}}^1$ is uniformly smaller than that of $\hat{\boldsymbol{\mu}}^{SG}$ if $0 < 1/a < 2(A-2)$.*

PROOF. By the orthogonal transformation $\hat{\boldsymbol{\theta}}^{SG} = \boldsymbol{P}'\hat{\boldsymbol{\mu}}^{SG}$

$$\hat{\theta}_i^{SG} = \begin{cases} Z_i, & \text{if} \quad 1 \le i \le A, \\ (1 - 1/bS_{12})Z_i, & \text{if} \quad A + 1 \le i \le AB, \end{cases}$$

it suffices to show that $\Delta(\boldsymbol{\theta}) \equiv R(\hat{\boldsymbol{\theta}}^{SG}, \boldsymbol{\theta}) - R(\hat{\boldsymbol{\theta}}^1, \boldsymbol{\theta}) > 0$ under the condition of the theorem.

$$\begin{aligned} \Delta(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \Bigg\{ & \left(\sum_{i=1}^{A} \frac{2Z_i(Z_i - \theta_i)}{aS_{11}} - \frac{1}{a^2 S_{11}} \right) I_{[aS_{11} \ge bS_{12}]}(\boldsymbol{Z}) \\ & + \left(\sum_{i=1}^{AB} \frac{2c_1 Z_i(Z_i - \theta_i)}{S_{11} + S_{12}} - \frac{c_1^2}{S_{11} + S_{12}} \right) I_{[aS_{11} < bS_{12}]}(\boldsymbol{Z}) \Bigg\} \\ - E_{\boldsymbol{\theta}} \Bigg\{ & \left(\sum_{i=A+1}^{AB} \frac{2Z_i(Z_i - \theta_i)}{bS_{12}} - \frac{1}{b^2 S_{12}} \right) I_{[aS_{11} < bS_{12}]}(\boldsymbol{Z}) \Bigg\}, \end{aligned}$$

where $c_1 = 1/a + 1/b$. Applying Lemma 2.1 to the first, third and fifth terms of the above expression, we have

$$\begin{aligned} \Delta(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \Bigg\{ & \frac{1}{aS_{11}} \left(2(A-2) - \frac{1}{a} \right) I_{[aS_{11} \ge bS_{12}]}(\boldsymbol{Z}) \\ & + \left(\frac{c_1}{S_{11} + S_{12}} (2(AB-2) - c_1) - \frac{c_2}{bS_{12}} \right) I_{[aS_{11} < bS_{12}]}(\boldsymbol{Z}) \Bigg\} \\ > E_{\boldsymbol{\theta}} \Bigg\{ & \frac{1}{aS_{11}} \left(2(A-2) - \frac{1}{a} \right) I_{[aS_{11} \ge bS_{12}]}(\boldsymbol{Z}) \\ & + \frac{c_1}{S_{11} + S_{12}} \left(2A - \frac{1}{a} \right) I_{[aS_{11} < bS_{12}]}(\boldsymbol{Z}) \Bigg\}, \end{aligned}$$

where $c_2 = 2(A(B-1) - 2) - 1/b$, which is positive if $0 < 1/a < 2(A-2)$.

Even if $A = 1$ or 2 the estimator $\hat{\boldsymbol{\mu}}^1$ is minimax under the following conditions, although the estimator $\hat{\boldsymbol{\mu}}^{SS}$ cannot be constructed in that case. This estimator is a convex combination of Stein-type estimators which shrink Y_{ij}'s towards zero or towards the average $\bar{Y}_{1\cdot}$ if $A = 1$.

THEOREM 2.2. *Suppose that* $A \ge 1$. *Then* $\hat{\boldsymbol{\mu}}^1$ *is minimax if* $0 < 1/a + 1/b < 2(AB - 4)$ *and* $0 < 1/b < 2(A(B-1) - 2)$.

Note. The MLE $\boldsymbol{Y}$ is minimax with constant risk AB. The risk of $\hat{\boldsymbol{\mu}}^1$ is uniformly smaller than that of $\boldsymbol{Y}$ under the above conditions on a and b, and approaches AB as $\|\mu\| \to \infty$. Therefore $\hat{\boldsymbol{\mu}}^1$ is minimax. Throughout this paper, an estimator dominating $\boldsymbol{Y}$ is minimax by the same reason.

PROOF. Applying Lemma 2.1 to three terms of $\Delta(\boldsymbol{\theta}) \equiv R(\boldsymbol{Z}, \boldsymbol{\theta}) - R(\hat{\boldsymbol{\theta}}^1, \boldsymbol{\theta})$ in a similar way as Theorem 2.1, we have

$$\begin{aligned}\Delta(\boldsymbol{\theta}) &= E_{\boldsymbol{\theta}}\left\{\left(\frac{1}{aS_{11}}\left(2(A-2)-\frac{1}{a}\right)+\frac{c_2}{bS_{12}}\right)I_{[aS_{11}\geq bS_{12}]}(\boldsymbol{Z})\right.\\ &\qquad\left.+\frac{c_1}{S_{11}+S_{12}}(2(AB-2)-c_1)I_{[aS_{11}<bS_{12}]}(\boldsymbol{Z})\right\}\\ &\geq E_{\boldsymbol{\theta}}\left\{\frac{1}{aS_{11}}(2(AB-4)-c_1)I_{[aS_{11}\geq bS_{12}]}(\boldsymbol{Z})\right.\\ &\qquad\left.+\frac{c_1}{S_{11}+S_{12}}(2(AB-2)-c_1)I_{[aS_{11}<bS_{12}]}(\boldsymbol{Z})\right\},\end{aligned}$$

where c_1 and c_2 are defined in the proof of Theorem 2.1, which is positive if the constants a and b satisfy the conditions.

Remark 2.1. The above theorems hold for a positive part version of $\hat{\boldsymbol{\mu}}^1$. The positive part version of $\hat{\boldsymbol{\mu}}^1$ improves on $\hat{\boldsymbol{\mu}}^1$.

The estimator $\hat{\boldsymbol{\theta}}^{SS}$ of (2.7) would be a natural competitor to the estimator $\hat{\boldsymbol{\theta}}^1$ of (2.8). It is not known if $\hat{\boldsymbol{\theta}}^1$ improves $\hat{\boldsymbol{\theta}}^{SS}$ with respect to the risk. However, it is shown that the Bayes risk of $\hat{\boldsymbol{\theta}}^1$, that is, $E_B\|\hat{\boldsymbol{\theta}}^1 - \boldsymbol{\theta}\|^2$, is uniformly smaller than that of $\hat{\boldsymbol{\theta}}^{SS}$ in a special case of the following theorem. E_B denotes expectation under the prior distribution (2.3). Before stating the result, we need the following lemma.

LEMMA 2.2. *Let U be a linear space with an inner product $(\boldsymbol{x}, \boldsymbol{y})$ and a norm $\|\boldsymbol{x}\| = (\boldsymbol{x}, \boldsymbol{x})^{1/2}$ and let V be a closed convex set of U. For any $\xi \in V$ and any $\boldsymbol{x} \in U - V$, let $\boldsymbol{y}$ be the projection of $\boldsymbol{x}$ onto V: $\|\boldsymbol{x} - \boldsymbol{y}\| = \inf(\|\boldsymbol{x} - \boldsymbol{v}\|: \boldsymbol{v} \in V)$. Then $\|\xi - \boldsymbol{x}\| > \|\xi - \boldsymbol{y}\|$.*

Remark 2.2. If $U = R^n$ and $V = \{\boldsymbol{v} = (v_1, \dots, v_n): v_1 \leq v_2 \leq \cdots \leq v_n\}$, then the projection $\boldsymbol{y}$ of $\boldsymbol{x}$ onto V is an "isotonic regression" (Barlow *et al.* (1972)). The lemma means that the isotonic regression has a smaller mean squares error than the observations.

THEOREM 2.3. *Let $\hat{\boldsymbol{\theta}}^{SBR}$ be an estimator obtained by replacing $\hat{\sigma}_1^2(S_{11}, S_{12})$ and $\hat{\sigma}_2^2(S_{11}, S_{12})$ in $\hat{\boldsymbol{\theta}}^{SB}$ of (2.6) with $(\hat{\sigma}_1^2(S_{11}, S_{12})S_{11} + \hat{\sigma}_2^2(S_{11}, S_{12}) \cdot S_{12})/(S_{11} + S_{12})$ in the case $\hat{\sigma}_1^2(S_{11}, S_{12}) > \hat{\sigma}_2^2(S_{11}, S_{12})$. Then the Bayes risk of $\hat{\boldsymbol{\theta}}^{SBR}$ is uniformly smaller than that of $\hat{\boldsymbol{\theta}}^{SB}$.*

PROOF. First, we calculate the Bayes risk of $\hat{\boldsymbol{\theta}}^{SB}$.

$$E_B\|\hat{\boldsymbol{\theta}}^{SB}-\boldsymbol{\theta}\|^2 = A(1-1/(B\sigma_a^2+\sigma_e^2+1)) + A(B-1)(1-1/(\sigma_e^2+1)) + E_Z\{(1/(B\sigma_a^2+\sigma_e^2+1)-\hat{\sigma}_1^2)^2 S_{11} + (1/(\sigma_e^2+1)-\hat{\sigma}_2^2)^2 S_{12}\}\,,$$

where $\hat{\sigma}_i^2=\hat{\sigma}_i^2(S_{11},S_{12})$; $i=1$, 2, and E_Z indicates expectation under the marginal distribution (2.5). The expression in the braces is a distance between $(1/(B\sigma_a^2+\sigma_e^2+1),\ 1/(\sigma_e^2+1))$ and $(\hat{\sigma}_1^2,\hat{\sigma}_2^2)$. Hence by Lemma 2.2, the expression is reduced by replacing $(\hat{\sigma}_1^2,\hat{\sigma}_2^2)$ with its isotonic regression, $(\hat{\sigma}_0^2,\hat{\sigma}_0^2)$, $\hat{\sigma}_0^2=(\hat{\sigma}_1^2 S_{11}+\hat{\sigma}_2^2 S_{12})/(S_{11}+S_{12})$, if $\hat{\sigma}_1^2>\hat{\sigma}_2^2$.

Remark 2.3. Replacing $\hat{\sigma}_1^2(S_{11},S_{12})$ and $\hat{\sigma}_2^2(S_{11},S_{12})$ in $\hat{\boldsymbol{\theta}}^{SB}$ of (2.6) by min $(1, 1/aS_{11})$ and min $(1, 1/bS_{12})$, respectively, we get the positive part version of $\hat{\boldsymbol{\theta}}^{SS}$. From Theorem 2.3, we can construct the estimator improving on the positive part version of $\hat{\boldsymbol{\theta}}^{SS}$ with respect to the Bayes risk.

Remark 2.4. For priors different from (2.3), the estimator (2.7) may have smaller Bayes risk than the estimator (2.8). Suppose that $\boldsymbol{\theta}\sim N_{AB}(\mathbf{0},\Sigma)$, where Σ is a diagonal matrix, of which the first A diagonal elements are τ_1^2 and the others are τ_2^2. The estimator (2.7) seems to be better than (2.8).

2.2 *Shrinkage towards the grand average*

Suppose that (1.2) and (1.3) hold. That is,

$$\boldsymbol{\mu}=[\mathbf{1}_{AB}\vdots\, \boldsymbol{X}]\begin{bmatrix}\mu\\ \cdots\\ \boldsymbol{\alpha}\end{bmatrix}+\boldsymbol{e}\,,$$

where $\mathbf{1}_{AB}=(1,\dots,1)'$ and $\boldsymbol{X}$ is the same as (2.2). Construct an AB-dimensional orthogonal matrix $\boldsymbol{P}$, of which the first A columns form an orthonormal basis of the subspace spanned by the columns of $[\mathbf{1}_{AB}\vdots\,\boldsymbol{X}]$ and the remainder columns are arbitrary. Consider the orthogonal transformation

$$\boldsymbol{Z}=\boldsymbol{P}'\boldsymbol{Y} \quad\text{and}\quad \boldsymbol{\theta}=\boldsymbol{P}'\boldsymbol{\mu}\,.$$

It may be reasonable to assume that the prior on μ is vague. Hence we get a limit of the Bayes estimator as $\sigma_\mu^2\to\infty$;

$$\text{(2.10)}\qquad \hat{\theta}_i^B=\begin{cases} Z_i, & \text{if}\quad i=1\,,\\ (1-1/(B\sigma_a^2+\sigma_e^2+1))Z_i, & \text{if}\quad 2\le i\le A\,,\\ (1-1/(\sigma_e^2+1))Z_i, & \text{if}\quad A+1\le i\le AB\,.\end{cases}$$

The marginal distribution of $\boldsymbol{Z}$ is

$$\boldsymbol{Z} \sim N_{AB}(\boldsymbol{\theta}^3, \Sigma_3)\,,$$

where $\boldsymbol{\theta}^3 = (\sqrt{AB}\omega, 0, \ldots, 0)'$ and Σ_3 is a diagonal matrix, of which the first diagonal element is $AB\sigma_\mu^2 + B\sigma_a^2 + \sigma_e^2 + 1$, the second $A-1$ diagonal elements are $B\sigma_a^2 + \sigma_e^2 + 1$ and the remainders are $\sigma_e^2 + 1$. Suppose that some estimators $\hat{\sigma}_1^2(S_{11}', S_{12})$ and $\hat{\sigma}_2^2(S_{11}', S_{12})$ can be used for $1/(B\sigma_a^2 + \sigma_e^2 + 1)$ and $1/(\sigma_e^2 + 1)$ in (2.10), respectively. The following component estimator is obtained;

$$\hat{\theta}_i^{SB*} = \begin{cases} Z_i, & \text{if} \quad i = 1\,, \\ (1 - \hat{\sigma}_1^2(S_{11}', S_{12}))Z_i, & \text{if} \quad 2 \le i \le A\,, \\ (1 - \hat{\sigma}_2^2(S_{11}', S_{12}))Z_i, & \text{if} \quad A+1 \le i \le AB\,, \end{cases} \tag{2.11}$$

where $S_{11}' = \sum_{i=2}^{A} Z_i^2$ and S_{12} is defined in (2.6). Similarly to Subsection 2.1, the following reasonable estimator is considered;

$$\hat{\theta}_i^{1*} = \begin{cases} Z_i, & \text{if} \quad i = 1\,, \\ (1 - 1/aS_{11}')Z_i I_{[aS_{11}' \ge bS_{12}]}(\boldsymbol{Z}) & \\ \quad + (1 - (1/a + 1/b)/(S_{11}' + S_{12}))Z_i I_{[aS_{11}' < bS_{12}]}(\boldsymbol{Z})\,, & \\ & \text{if} \quad 2 \le i \le A\,, \\ (1 - 1/bS_{12})Z_i I_{[aS_{11}' \ge bS_{12}]}(\boldsymbol{Z}) & \\ \quad + (1 - (1/a + 1/b)/(S_{11}' + S_{12}))Z_i I_{[aS_{11}' < bS_{12}]}(\boldsymbol{Z})\,, & \\ & \text{if} \quad A+1 \le i \le AB\,, \end{cases} \tag{2.12}$$

where a and b are positive constants to be suitably chosen. Inverting $\hat{\boldsymbol{\theta}}^{SB*}$, we have

$$\begin{aligned} \hat{\mu}_{ij}^{SB*} &= \rho(S_{11}', S_{12})(\overline{Y}_{..} + (1 - \hat{\sigma}_1^2(S_{11}', S_{12}))(Y_{ij} - \overline{Y}_{..})) \\ &\quad + (1 - \rho(S_{11}', S_{12}))(\overline{Y}_{i.} + (1 - \hat{\sigma}_2^2(S_{11}', S_{12}))(Y_{ij} - \overline{Y}_{i.}))\,, \end{aligned}$$

where $\rho(\cdot,\cdot)$ is defined in (2.9). The estimator $\hat{\boldsymbol{\mu}}^{1*} = \boldsymbol{P}\hat{\boldsymbol{\theta}}^{1*}$ is a convex combination of $\hat{\boldsymbol{\mu}}^{SG}$ and the Stein-type estimator which shrinks Y_{ij}'s towards the grand average.

Conditions on the constant a such that the risk of $\hat{\boldsymbol{\mu}}^{1*}$ is smaller than that of $\hat{\boldsymbol{\mu}}^{SG}$ in the case $A \ge 4$ are given. Also even if $A = 2$ or 3 the estimator $\hat{\boldsymbol{\mu}}^{1*}$ is minimax under the following conditions.

THEOREM 2.4. *Suppose that $A \geq 4$. Then the risk of $\hat{\boldsymbol{\mu}}^{1*}$ is uniformly smaller than that of $\hat{\boldsymbol{\mu}}^{SG}$ if $0 < 1/a < 2(A-3)$.*

THEOREM 2.5. *Suppose that $A \geq 2$. Then $\hat{\boldsymbol{\mu}}^{1*}$ is minimax if $0 < 1/a + 1/b < 2(AB-5)$ and $0 < 1/b < 2(A(B-1)-2)$.*

Remark 2.5. The Bayes risk of $\hat{\boldsymbol{\mu}}^{1*}$ is uniformly smaller than that of Stein and Takeuchi's (2.13) which is defined in Subsection 2.3.

2.3 *Simulations of $\hat{\boldsymbol{\mu}}^{1*}$ and other Stein-type estimators*

We obtain Monte Carlo estimates for the case $A = 4$ and $B = 3$, of the risk of the positive part version of each of the following estimators; $\hat{\boldsymbol{\mu}}^{1*}$, $\hat{\boldsymbol{\mu}}^{ST}$ and $\hat{\boldsymbol{\mu}}^{SG}$ which are defined in Section 1, and Stein and Takeuchi's estimator $\hat{\boldsymbol{\mu}}^{SS*}$ obtained by

$$(2.13)\quad \hat{\mu}_{ijk}^{SS*} = \bar{Y}_{..} + (1 - 1/aS_{11}')(\bar{Y}_{i.} - \bar{Y}_{..}) + (1 - 1/bS_{12})(Y_{ij} - \bar{Y}_{i.})\,,$$

where $0 < 1/a < 2(A-3)$ and $0 < 1/b < 2(A(B-1)-2)$. It is trivial that $\hat{\boldsymbol{\mu}}^{SS*}$ dominates $\hat{\boldsymbol{\mu}}^{SG}$. The risk of each estimator for each choice of $\boldsymbol{\mu}$ is estimated by the average loss $\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2$ based on 10,000 independent samples of $\boldsymbol{Y} \sim N_{12}(\boldsymbol{\mu}, I)$ (the normal random deviates are generated from the FACOM SSLII routine RANN2).

As estimators S_{11}' and S_{12} follow the noncentral chi-squared distributions with $A-1$ and $A(B-1)$ degrees of freedom, respectively, we put $\sigma_B = (B\sum_i (\bar{\mu}_{i.} - \bar{\mu}_{..})^2/(A-1))^{1/2}$ and $\sigma_W = (\sum_i \sum_j (\mu_{ij} - \bar{\mu}_{i.})^2/A(B-1))^{1/2}$, where noncentralities of S_{11}' and S_{12} are $(A-1)\sigma_B^2$ and $A(B-1)\sigma_W^2$, respectively. We estimate the risk values of the above four estimators for sixteen choices of $\boldsymbol{\mu}$, varying $\sigma_B = 0.0$, 1.7888(0.4472) and $\sigma_W = 0.0$, 1.2(0.3). The values of $1/a$ and $1/b$ in $\hat{\boldsymbol{\mu}}^{1*}$ are $A-3$ and $A(B-1)-2$, respectively. The values $1/a = A-3$ and $1/b = A(B-1)-2$ are the midpoints of possible values of $1/a$ and $1/b$ in $\hat{\boldsymbol{\mu}}^{SS*}$, respectively, whose risk is known to be minimum at these values.

Table 1 shows that if $\sigma_B \geq \sigma_W$ the estimator $\hat{\boldsymbol{\mu}}^{1*}$ is slightly better than the others. It is observed that the estimator $\hat{\boldsymbol{\mu}}^{SS*}$ is better than the others when σ_B is small and σ_W is large.

3. Two-stage hierarchical classification case

Let components of $\boldsymbol{Y}$ be classified into A classes and let each class be divided into B subclasses consisting of C components. Let Y_{ijk} be independently and normally distributed;

$$(3.1)\quad Y_{ijk} \sim N(\mu_{ijk}, 1^2); \quad i = 1,\dots,A; \quad j = 1,\dots,B; \quad k = 1,\dots,C\,.$$

Table 1. The risk values of positive part versions of $\hat{\boldsymbol{\mu}}^{1*}$, $\hat{\boldsymbol{\mu}}^{SS*}$, $\hat{\boldsymbol{\mu}}^{ST}$ and $\hat{\boldsymbol{\mu}}^{SG}$ for $A = 4$ and $B = 3$.

σ_B \ σ_W	0.0000	0.3000	0.6000	0.9000	1.2000
0.0000	3.9	4.5	5.8	7.4	8.5
	3.9	4.5	5.9	7.3	8.4
	2.3	2.9	4.4	6.3	7.8
	5.3	5.9	7.3	8.7	9.8
0.4472	4.1	4.7	6.1	7.6	8.7
	4.2	4.7	6.1	7.6	8.6
	2.8	3.4	4.8	6.5	7.9
	5.3	5.9	7.3	8.7	9.8
0.8944	4.6	5.2	6.5	8.0	9.0
	4.7	5.2	6.6	8.0	9.1
	4.1	4.6	5.7	7.1	8.3
	5.3	5.9	7.3	8.7	9.8
1.3416	5.0	5.6	6.9	8.4	9.4
	5.0	5.6	7.0	8.4	9.5
	5.8	6.1	6.9	7.9	8.7
	5.3	5.9	7.3	8.7	9.8
1.7888	5.2	5.8	7.1	8.6	9.6
	5.2	5.8	7.1	8.6	9.6
	7.3	7.5	8.0	8.6	9.2
	5.3	5.9	7.3	8.7	9.8

The four lines in each entry correspond to the risk values of $\hat{\boldsymbol{\mu}}^{1*}$, $\hat{\boldsymbol{\mu}}^{SS*}$, $\hat{\boldsymbol{\mu}}^{ST}$ and $\hat{\boldsymbol{\mu}}^{SG}$, respectively. The standard error of each estimate is less than 0.05.

An adaptive shrinkage estimator will be obtained. It is shown that this estimator dominates a Stein-type estimator which shrinks Y_{ijk}'s towards individual subclass averages,

$$\hat{\mu}_{ijk}^{SC} = \bar{Y}_{ij.} + \left(1 - 1 \bigg/ \left(c \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2\right)\right)(Y_{ijk} - \bar{Y}_{ij.}) ,$$

where $0 < 1/c < 2(AB(C-1) - 2)$. Also it is shown that it is minimax even if A is small. To derive such an estimator, assume a two-stage hierarchical classification structure of $\boldsymbol{\mu}$;

$$\mu_{ijk} = \alpha_i + \beta_{ij} + e_{ijk} \quad \text{or} \quad \mu_{ijk} = \mu + \alpha_i + \beta_{ij} + e_{ijk} , \tag{3.2}$$

where $\alpha_i \sim N(0, \sigma_\alpha^2)$, $\beta_{ij} \sim N(0, \sigma_\beta^2)$, $e_{ijk} \sim N(0, \sigma_e^2)$, $\mu \sim N(\omega, \sigma_\mu^2)$ and these are independently distributed.

3.1 *Shrinkage towards zero*

Suppose that (3.1) and the first of (3.2) hold. From the matrix form of (3.2), construct an ABC-dimensional orthogonal matrix $\boldsymbol{P}$. Consider the orthogonal transformation

$$\boldsymbol{Z} = \boldsymbol{P}'\boldsymbol{Y} \quad \text{and} \quad \boldsymbol{\theta} = \boldsymbol{P}'\boldsymbol{\mu}\,,$$

where $\boldsymbol{Z} = (Z_1, \dots, Z_{ABC})'$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{ABC})'$. The Bayes estimator under squared error loss is

$$\hat{\theta}_i^B = \begin{cases} (1 - 1/(BC\sigma_\alpha^2 + C\sigma_\beta^2 + \sigma_e^2 + 1))Z_i, & \text{if} \quad 1 \le i \le A\,, \\ (1 - 1/(C\sigma_\beta^2 + \sigma_e^2 + 1))Z_i, & \text{if} \quad A + 1 \le i \le AB\,, \\ (1 - 1/(\sigma_e^2 + 1))Z_i, & \text{if} \quad AB + 1 \le i \le ABC\,. \end{cases} \tag{3.3}$$

The marginal distribution of $\boldsymbol{Z}$ is

$$\boldsymbol{Z} \sim N_{ABC}(\boldsymbol{0}, \Sigma_1)\,,$$

where Σ_1 is a diagonal matrix, of which the first A diagonal elements are $BC\sigma_\alpha^2 + C\sigma_\beta^2 + \sigma_e^2 + 1$, the second $A(B-1)$ diagonal elements are $C\sigma_\beta^2 + \sigma_e^2 + 1$ and the remainders are $\sigma_e^2 + 1$. Put $S_{21} = \sum_{i=1}^{A} Z_i^2$, $S_{22} = \sum_{i=A+1}^{AB} Z_i^2$ and $S_{23} = \sum_{i=AB+1}^{ABC} Z_i^2$. Suppose that some estimators $\hat{\sigma}_1^2(S_{21}, S_{22}, S_{23})$, $\hat{\sigma}_2^2(S_{21}, S_{22}, S_{23})$ and $\hat{\sigma}_3^2(S_{21}, S_{22}, S_{23})$ can be used for $1/(BC\sigma_\alpha^2 + C\sigma_\beta^2 + \sigma_e^2 + 1)$, $1/(C\sigma_\beta^2 + \sigma_e^2 + 1)$ and $1/(\sigma_e^2 + 1)$ in (3.3), respectively. The following component estimator is obtained;

$$\hat{\theta}_i^{SB} = \begin{cases} (1 - \hat{\sigma}_1^2(S_{21}, S_{22}, S_{23}))Z_i, & \text{if} \quad 1 \le i \le A\,, \\ (1 - \hat{\sigma}_2^2(S_{21}, S_{22}, S_{23}))Z_i, & \text{if} \quad A + 1 \le i \le AB\,, \\ (1 - \hat{\sigma}_3^2(S_{21}, S_{22}, S_{23}))Z_i, & \text{if} \quad AB + 1 \le i \le ABC\,. \end{cases} \tag{3.4}$$

Since $1/(BC\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_e^2 + 1) < 1/(C\sigma_\beta^2 + \sigma_e^2 + 1) < 1/(\sigma_e^2 + 1)$, the component estimator corresponding to the estimator (2.8) in Subsection 2.1 is considered;

(1) If $1/aS_{21} \le 1/bS_{22} \le 1/cS_{23}$,

$$\hat{\theta}_i^1 = \begin{cases} (1 - 1/aS_{21})Z_i, & \text{if} \quad 1 \le i \le A\,, \\ (1 - 1/bS_{22})Z_i, & \text{if} \quad A + 1 \le i \le AB\,, \\ (1 - 1/cS_{23})Z_i, & \text{if} \quad AB + 1 \le i \le ABC\,. \end{cases}$$

(2) If $1/cS_{23} \le 1/bS_{22}$ and $1/aS_{21} \le (1/b + 1/c)/(S_{22} + S_{23})$,

$$\hat{\theta}_i^1 = \begin{cases} (1 - 1/aS_{21})Z_i, & \text{if} \quad 1 \le i \le A\,, \\ (1 - (1/b + 1/c)/(S_{22} + S_{23}))Z_i, & \text{if} \quad A + 1 \le i \le ABC\,. \end{cases}$$

(3) If $1/cS_{23} \leq 1/bS_{22}$ and $(1/b+1/c)/(S_{22}+S_{23}) \leq 1/aS_{21}$,

$$\hat{\theta}^1_i = (1-(1/a+1/b+1/c)/(S_{21}+S_{22}+S_{23}))Z_i, \quad i=1,\dots,ABC\,.$$

(4) If $1/bS_{22} \leq 1/aS_{21}$ and $(1/a+1/b)/(S_{21}+S_{22}) \leq 1/cS_{23}$,

$$\hat{\theta}^1_i = \begin{cases} (1-(1/a+1/b)/(S_{21}+S_{22}))Z_i, & \text{if} \quad 1 \leq i \leq AB\,, \\ (1-1/cS_{23})Z_i, & \text{if} \quad AB+1 \leq i \leq ABC\,. \end{cases}$$

(5) If $1/bS_{22} \leq 1/aS_{21}$ and $1/cS_{23} \leq (1/a+1/b)/(S_{21}+S_{22})$,

$$\hat{\theta}^1_i = (1-(1/a+1/b+1/c)/(S_{21}+S_{22}+S_{23}))Z_i, \quad i=1,\dots,ABC\,.$$

The coefficients a, b and c are positive constants to be suitably chosen. Inverting $\hat{\boldsymbol{\theta}}^{SB}$, we have

$$\begin{aligned}\hat{\mu}^{SB}_{ijk} &= (\hat{\sigma}^2_1/\hat{\sigma}^2_3)(1-\hat{\sigma}^2_3)Y_{ijk} + (1-\hat{\sigma}^2_1/\hat{\sigma}^2_2)((1-\hat{\sigma}^2_3)Y_{ijk} + \hat{\sigma}^2_2\overline{Y}_{i..}) \\ &\quad + (\hat{\sigma}^2_1/\hat{\sigma}^2_2 - \hat{\sigma}^2_1/\hat{\sigma}^2_3)((1-\hat{\sigma}^2_3)Y_{ijk} + \hat{\sigma}^2_2\hat{\sigma}^2_3\overline{Y}_{ij.}/\hat{\sigma}^2_1)\,,\end{aligned}$$

where $\hat{\sigma}^2_i = \hat{\sigma}^2_i(S_{21}, S_{22}, S_{23})$; $i=1, 2, 3$. Note that $S_{21} = BC\sum_i \overline{Y}^2_{i..}$, $S_{22} = C\sum_i\sum_j(\overline{Y}_{ij.} - \overline{Y}_{i..})^2$ and $S_{23} = \sum_i\sum_j\sum_k(Y_{ijk} - \overline{Y}_{ij.})^2$. The estimator $\hat{\boldsymbol{\mu}}^1 = \boldsymbol{P}\hat{\boldsymbol{\theta}}^1$ is a convex combination of Stein-type estimators which shrink Y_{ijk}'s towards zero, towards the class average $\overline{Y}_{i..}$ or towards the subclass average $\overline{Y}_{ij.}$.

Conditions on the constant a such that the risk of $\hat{\boldsymbol{\mu}}^1$ is smaller than that of $\hat{\boldsymbol{\mu}}^{SC}$ in the case $A \geq 3$ are given. Also even if $A=1$ or 2, then $\hat{\boldsymbol{\mu}}^1$ is minimax under the following conditions.

THEOREM 3.1. *Suppose that $A \geq 3$ and $B \geq 2$. Then the risk of $\hat{\boldsymbol{\mu}}^1$ is uniformly smaller than that of $\hat{\boldsymbol{\mu}}^{SC}$ if $0 < 1/a < 2(A-2)$ and $0 < 1/b < 2(A(B-1)-2)$.*

THEOREM 3.2. *Suppose that $A \geq 1$ and $B \geq 2$. Then $\hat{\boldsymbol{\mu}}^1$ is minimax if $0 < 1/a + 1/b + 1/c < 2(ABC-6)$, $0 < 1/b < 2(A(B-1)-2)$ and $0 < 1/c < 2(AB(C-1)-2)$.*

Remark 3.1. The Bayes risk of $\hat{\boldsymbol{\mu}}^1$ is uniformly smaller than that of Stein and Takeuchi's corresponding to (2.7) in Subsection 2.1.

3.2 *Shrinkage towards the grand average*

Suppose that (3.1) and the latter of (3.2) hold. A limit of the Bayes estimator as $\sigma^2_\mu \to \infty$ is derived and the hyperparameters are estimated by suitable statistics. We get

$$\begin{aligned}\hat{\mu}_{ijk}^{SB*} &= (\hat{\sigma}_1^2/\hat{\sigma}_3^2)(\overline{Y}_{...} + (1-\hat{\sigma}_3^2)(Y_{ijk} - \overline{Y}_{...})) \\ &\quad + (1-\hat{\sigma}_1^2/\hat{\sigma}_2^2)((1-\hat{\sigma}_3^2)\overline{Y}_{i..} + \hat{\sigma}_2^2\overline{Y}_{i..}) \\ &\quad + (\hat{\sigma}_1^2/\hat{\sigma}_2^2 - \hat{\sigma}_1^2/\hat{\sigma}_3^2)((1-\hat{\sigma}_3^2)\overline{Y}_{ij.} + \hat{\sigma}_2^2\hat{\sigma}_3^2\overline{Y}_{ij.}/\hat{\sigma}_1^2)\,,\end{aligned}$$

where $\hat{\sigma}_i^2 = \hat{\sigma}_i^2(S'_{21}, S_{22}, S_{23})$; $i = 1, 2, 3$; and $S'_{21} = BC\sum_i(\overline{Y}_{i..} - \overline{Y}_{...})^2$, S_{22} and S_{23} are defined in (3.4). Suppose that $\hat{\boldsymbol{\mu}}^{1*}$ corresponds to $\hat{\boldsymbol{\mu}}^1$. The estimator $\hat{\boldsymbol{\mu}}^{1*}$ is a convex combination of Stein-type estimators which shrink Y_{ijk}'s towards the grand average, towards individual class averages or towards individual subclass averages. We obtain similar results to those in Subsection 3.1.

3.3 *Estimating American baseball team batting averages*

We estimate $\boldsymbol{\mu}$ when components of $\boldsymbol{Y}$ are classified into two classes, each of which can be divided into two subclasses of different sample sizes. It is applied to estimating American baseball team averages (a problem discussed by George (1986*b*)).

Let Y_{ijk} be independently and normally distributed;

$$Y_{ijk} \sim N(\mu_{ijk}, 1^2); \quad i = 1, 2; \quad j = 1, 2; \quad k = 1, \ldots, C_i\,,$$

under the assumption (3.2). Put $C = 2(C_1 + C_2)$. Construct a C-dimensional orthogonal matrix $\boldsymbol{P}$ and consider the orthogonal transformation

$$\boldsymbol{Z} = \boldsymbol{P}'\boldsymbol{Y} \quad \text{and} \quad \boldsymbol{\theta} = \boldsymbol{P}'\boldsymbol{\mu}\,,$$

where $\boldsymbol{Z} = (Z_1, \ldots, Z_C)'$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_C)'$. We get a limit of the Bayes estimator as $\sigma_\mu^2 \to \infty$;

$$\text{(3.5)} \qquad \hat{\theta}_i^B = \begin{cases} Z_i, & \text{if} \quad i = 1\,, \\ (1 - 1/(2D\sigma_\alpha^2 + D\sigma_\beta^2 + \sigma_e^2 + 1))Z_i, & \text{if} \quad i = 2\,, \\ (1 - 1/(C_1\sigma_\beta^2 + \sigma_e^2 + 1))Z_i, & \text{if} \quad i = 3\,, \\ (1 - 1/(C_2\sigma_\beta^2 + \sigma_e^2 + 1))Z_i, & \text{if} \quad i = 4\,, \\ (1 - 1/(\sigma_e^2 + 1))Z_i, & \text{if} \quad 5 \le i \le C\,, \end{cases}$$

where $D = 4C_1C_2/C$. The hyperparameters in (3.5) may be estimated by the following statistics; $2D\sigma_\alpha^2$ by aZ_2^2; $D\sigma_\beta^2$ by $b_1(Z_3^2 + Z_4^2)$; $C_1\sigma_\beta^2$ by $b_2(Z_3^2 + Z_4^2)$; $C_2\sigma_\beta^2$ by $b_3(Z_3^2 + Z_4^2)$ and $\sigma_e^2 + 1$ by $c\sum_{i=5}^{C} Z_i^2$. The coefficients a, b_1, b_2, b_3 and c are positive constants to be suitably chosen. Therefore, we get

$$(3.6)\qquad \hat{\theta}_i^* = \begin{cases} Z_i, & \text{if} \quad i=1\,, \\ (1-1/(aS_{31}+b_1S_{32}+cS_{33}))Z_i, & \text{if} \quad i=2\,, \\ (1-1/(b_2S_{32}+cS_{33}))Z_i, & \text{if} \quad i=3\,, \\ (1-1/(b_3S_{32}+cS_{33}))Z_i, & \text{if} \quad i=4\,, \\ (1-1/cS_{33})Z_i, & \text{if} \quad 5\le i\le C\,, \end{cases}$$

where $S_{31}=Z_2^2$, $S_{32}=Z_3^2+Z_4^2$ and $S_{33}=\sum_{i=5}^{C} Z_i^2$. Notice that Stein and Takeuchi's estimator cannot be constructed from (3.5). Inverting $\hat{\boldsymbol{\theta}}^*$, we have $\hat{\boldsymbol{\mu}}^*=\boldsymbol{P}\hat{\boldsymbol{\theta}}^*$. The estimator $\hat{\boldsymbol{\mu}}^*$ is minimax under the following conditions.

THEOREM 3.3. *Suppose that* $C>9$ *and* $1/c=C-6$. *Then* $\hat{\boldsymbol{\mu}}^*$ *is minimax if* $a\ge 1/(C-9)$, $b_1\ge 1/(C-4)$, $b_2\ge 1/(C-9)$ *and* $b_3\ge 1/(C-4)$.

In George (1986*b*) the batting averages of all 26 baseball teams through their first 300 official at bats of the 1984 season are listed. He regarded the estimator based on the team batting averages through the first 300 official at bats as the predictor $\boldsymbol{\delta}$, and the team batting averages over the remainder of the season as the true values $\boldsymbol{\mu}$. Arcsine transformations of the averages are approximated by normal distributions. American and National Leagues were regarded as the classes and East and West as the subclasses. Here we consider the following three positive part versions; the estimator which shrinks Y_{ijk}'s towards the grand average,

$$(3.7)\quad \hat{\mu}_{ijk}^{ST+}=\bar{Y}_{...}+\left(1-\left(1\wedge 23\Big/\left(\sum_i\sum_j\sum_k (Y_{ijk}-\bar{Y}_{...})^2\right)\right)\right)(Y_{ijk}-\bar{Y}_{...})\,,$$

where $a\wedge b=\min(a,b)$, the estimator which shrinks Y_{ijk}'s towards class averages,

$$(3.8)\quad \hat{\mu}_{ijk}^{SG+}=\bar{Y}_{i..}+\left(1-\left(1\wedge 22\Big/\left(\sum_i\sum_j\sum_k (Y_{ijk}-\bar{Y}_{i..})^2\right)\right)\right)(Y_{ijk}-\bar{Y}_{i..})\,,$$

and the estimator which shrinks Y_{ijk}'s towards subclass averages,

$$(3.9)\quad \hat{\mu}_{ijk}^{SC+}=\bar{Y}_{ij.}+\left(1-\left(1\wedge 20\Big/\left(\sum_i\sum_j\sum_k (Y_{ijk}-\bar{Y}_{ij.})^2\right)\right)\right)(Y_{ijk}-\bar{Y}_{ij.})\,.$$

Further we consider three estimators of George's $\hat{\boldsymbol{\mu}}^{EG}$ using $c=1, 2, 5$ in the calibration (2.7) in George (1986*b*) which combine the above estimators $\hat{\boldsymbol{\mu}}^{ST+}$, $\hat{\boldsymbol{\mu}}^{SG+}$ and $\hat{\boldsymbol{\mu}}^{SC+}$. The sum of squared error loss $\|\boldsymbol{\delta}-\boldsymbol{\mu}\|^2$, called actual losses, are computed for $\hat{\boldsymbol{\mu}}^{ST+}$, $\hat{\boldsymbol{\mu}}^{SG+}$, $\hat{\boldsymbol{\mu}}^{SC+}$, the MLE $\boldsymbol{Y}$ and a positive part version of $\hat{\mu}^*$ with $1/c = 1/20$ and the following lower limits of the constants a, b_1, b_2 and b_3; $a = 1/17$, $b_1 = 1/22$, $b_2 = 1/17$ and $b_3 = 1/22$ in Theorem 3.3. The results are shown in Table 2. However, the smallest loss ($c = 5$) among three George's is presented in Table 2. Notice that a positive part version of $\hat{\mu}^*$ has the smallest loss.

Table 2. Actual losses for positive part versions of $\hat{\mu}^*$, $\hat{\mu}^{ST}$, $\hat{\mu}^{SG}$, $\hat{\mu}^{SC}$, $\hat{\mu}^{EG}$ and Y.

δ	$\hat{\mu}^*$	$\hat{\mu}^{ST}$	$\hat{\mu}^{SG}$	$\hat{\mu}^{SC}$	$\hat{\mu}^{EG}$	Y
$\Vert\delta - \mu\Vert^2$	4.21	4.24	4.94	7.30	4.38	26.53

Acknowledgements

The author would like to thank Professor M. Sibuya and Professor Y. Washio of Keio University, and Professor N. Shinozaki of Tokyo Keizai University for their helpful suggestions and criticisms. Also the author is grateful to the referee for insightful comments.

REFERENCES

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions*, Wiley, New York.

Berger, J. (1982). Selecting a minimax estimator of a multivariate normal mean, *Ann. Statist.*, **10**, 81-92.

Berger, J. and Dey, D. (1983). Combining coordinates in simultaneous estimation of normal means, *J. Statist. Plann. Inference*, **8**, 143-160.

Efron, B. and Morris, C. (1973). Combining possibly related estimation problems, *J. Roy. Statist. Soc. Ser. B*, **35**, 379-421.

George, E. I. (1986*a*). Minimax multiple shrinkage estimation, *Ann. Statist.*, **14**, 188-205.

George, E. I. (1986*b*). Combining minimax shrinkage estimators, *J. Amer. Statist. Assoc.*, **81**, 437-445.

George, E. I. (1986*c*). A formal Bayes multiple shrinkage estimator, *Comm. Statist. A—Theory Methods*, **15**, 2099-2114.

George, E. I. (1986*d*). Multiple shrinkage generalizations of the James-Stein estimator, *Contributions to the Theory and Applications of Statistics*, Academic Press, New York.

Haff, L. R. (1978). The multivariate normal mean with intraclass correlated components: estimation of urban fire alarm probabilities, *J. Amer. Statist. Assoc.*, **60**, 806-825.

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

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model, *J. Roy. Statist. Soc. Ser. B*, **34**, 1-41.

Stein, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs, *Festschrift for J. Neyman*, (ed. F. N. David), 351-366, Wiley, New York.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution, *Ann Statist.*, **9**, 1135–1151.
Takeuchi, K. (1980). *Toukei-suri*, 50–59, Shinyosha, Tokyo (in Japanese).