Ann. Inst. Statist. Math.
Vol. 40, No. 3, 555–563 (1988)

INADMISSIBILITY OF THE UNCOMBINED TWO-STAGE ESTIMATOR WHEN ADDITIONAL SAMPLES ARE AVAILABLE

TATSUYA KUBOKAWA

Department of Mathematics, University of Tsukuba, Tsukuba, Ibaraki 305, Japan

(Received February 25, 1987; revised November 11, 1987)

Abstract. Consider the problem of constructing an estimator with a preassigned bound on the risk for a mean of a normal distribution. The paper shows that the usual two-stage estimator is improved on by combined estimators when additional samples taken from distributions with the same mean and different variances are available.

Key words and phrases: Two-stage procedure, inadmissibility, common mean.

1. Introduction

Let $X_1, X_2, \ldots$ be a sequence of mutually independent random variables, each having normal distribution $N(\mu, \sigma_1^2)$ with unknown parameters μ and σ_1^2. Given a preassigned number $B > 0$, we consider the problem of constructing estimator $\hat{\mu}$ of μ such that

$$R(\hat{\mu}) = E[(\hat{\mu} - \mu)^2] \leq B\,, \tag{1.1}$$

uniformly with respect to unknown parameters. If a sample of size n is taken, then sample mean $\overline{X}_n$ has risk $R(\overline{X}_n) = \sigma_1^2/n$. When σ_1^2 is known, the risk can satisfy the required condition (1.1) by taking $n = n^* = \sigma_1^2/B$ where, for simplicity, n^* is assumed to be an integer. However, when σ_1^2 is unknown, there does not exist any fixed sample size such that $R(\overline{X}_n) \leq B$ for all $\sigma_1^2 > 0$. Then the following two-stage estimation rule is proposed (see Rao (1973), pp. 486–487):

(i) Start with an initial sample $X_1, \ldots, X_m$ of size m (≥ 4).

(ii) Define the stopping number by

$$N = \max\left(m, \left[\frac{S_{1m}}{B(m-3)}\right] + 1\right), \tag{1.2}$$

where $S_{1m} = \sum_{j=1}^{m} (X_j - \overline{X}_m)^2$, $\overline{X}_m = \sum_{j=1}^{m} X_j/m$, and $[u]$ denotes the largest integer less than u.

(iii) Take another sample $X_{m+1},\dots, X_N$. Then estimate μ by

$$\overline{X}_N = N^{-1} \sum_{j=1}^{N} X_j ,$$

which has the risk not larger than B.

In this paper, we assume the following situation: The above two-stage sampling procedure is carried out for the principal estimation of μ, while some supplementary observations are obtained before or after its two-stage sampling, and these observations seem to be particularly distributed with the same mean μ. In this case, of great interest is to investigate whether their additional observations are available for estimation of μ, which is just our purpose.

From a decision-theoretic point of view, sequential estimation have two kinds of domination problems, that is, improving on the estimation procedure and making the stopping number smaller. The former problem has been resolved by Ghosh and Sen (1983), Takada (1984), Ghosh *et al.* (1987) and Nickerson (1987), but little is known about analytical study for the latter problem, which seems to be not easy. Here the former domination problem is discussed.

Section 2 deals with two-sample problem. In addition to the sample obtained by the above two-stage sampling rule, we assume that a random sample of fixed size l is taken from $N(\mu, \sigma_2^2)$ with unknown σ_2^2 possibly different from σ_1^2. Using information of the additional sample, we consider the class of combined estimators which have been, in the fixed sample size case, investigated by Brown and Cohen (1974), Khatri and Shah (1974), Bhattacharya (1980) and Kubokawa (1987*a*). Then the condition under which the two-stage combined estimators dominate $\overline{X}_N$ is developed. It is noted that this domination result holds when the size l is at least 3. In Section 3, we discuss k-sample problem and obtain the condition which ensures that Brown and Cohen's type estimator (1974) has a smaller risk than $\overline{X}_N$.

2. Two-sample problem

Following the stopping number N given by (1.2), sample $X_1,\dots, X_N$ is taken from $N(\mu, \sigma_1^2)$. We further assume that random sample $Y_1,\dots, Y_l$ of fixed size l is additionally taken from $N(\mu, \sigma_2^2)$. Denote $S_{1N} = \sum_{j=1}^{N} (X_j - \overline{X}_N)^2$, $\overline{Y} = l^{-1} \sum_{j=1}^{l} Y_j$ and $S_2 = \sum_{j=1}^{l} (Y_j - \overline{Y})^2$. Then it is shown that the usual two-stage estimator $\overline{X}_N$ is improved on by using information from $Y_1, \ldots, Y_l$. Let a be a positive constant, and let b_N and c_N be nonnegative functions of N such that $b_N > 0$ and $b_N \geq c_N \geq 0$. The combined estimators we consider are of the form

$$\hat{\mu}_N(a, b_N, c_N) = \overline{X}_N + \frac{a}{1 + R_N} (\overline{Y} - \overline{X}_N) , \tag{2.1}$$

where $R_N = \{b_N S_2 / l + c_N(\overline{X}_N - \overline{Y})^2\} / (S_{1N}/N)$. By the symmetry of the conditional distribution of $\overline{X}_N - \overline{Y}$ given S_{1m}, it is seen that $\hat{\mu}_N(a, b_N, c_N)$ is an unbiased estimator of μ.

THEOREM 2.1. *The combined estimator $\hat{\mu}_N(a, b_N, c_N)$ dominates the two-stage uncombined estimator $\overline{X}_N$ relative to the mean squared error loss and has a risk bounded by a preassigned positive constant B if the following conditions hold for $m \geq 4$:*

(a) *$l \geq 6$ if $c_N = 0$, or $l \geq 3$ if $c_N > 0$.*

(b) *nb_n^2 is nondecreasing in n for $n \geq m$ and $\sum_{n=m}^{\infty} (nb_n)^{-2} < \infty$ if $c_N = 0$, or b_n^2/n and c_n/b_n are nondecreasing in n for $n \geq m$ if $c_N > 0$.*

(c) *a, b_n and c_n satisfy that*

$$a \leq \min\left[1, 2 \cdot \inf_{n \geq m} \left\{ \frac{l-5}{2n - m + 1} b_n \right\} \right] \quad \text{if} \quad c_N = 0 , \tag{2.2}$$

or that

$$a \leq \min\left[1, 2 \cdot \inf_{n \geq m} \left\{ \frac{l-2}{2n - m + 1} \cdot \left(1 + \frac{(l-1)(b_n - c_n)}{(l+2)b_n} \right) c_n \right\} \right] \quad \text{if} \quad b_N \geq c_N > 0 . \tag{2.3}$$

Example. (i) Brown and Cohen's type estimator (1974) $\hat{\mu}_N(a, (N-1)/(l+2), (N-1)/(l+2))$ is better than $\overline{X}_N$ provided $l \geq 3$ and

$$a \leq \min \{1, (l-2)/(l+2)\} ,$$

by Theorem 2.1.

(ii) Khatri and Shah's type estimator (1974) $\hat{\mu}_N(1, (N+1)/(l-2), (N+1)/(l-2))$ is always better than $\overline{X}_N$ for $l \geq 3$ by Theorem 2.1.

PROOF OF THEOREM 2.1. We shall prove that $\hat{\mu}_N(a, b_N, c_N)$ is better than $\overline{X}_N$ under the mean squared error loss. Let $\rho = \sigma_2^2/\sigma_1^2$, and let $T \sim \chi_3^2$, that is, a random variable distributed as a chi-square variate with 3 degrees of freedom, being independent of (S_{1m}, S_{1n}, S_2) for any $n \geq m$. Denote $\overline{R}_N = \{(b_N N/l)S_2 + c_N \sigma_1^2 (1 + \rho N/l)T\}/S_{1N}$. Then according to Brown and Cohen (1974) and Khatri and Shah (1974), the risk of $\hat{\mu}_N(a, b_N, c_N)$ is written by

$$(2.4) \quad R(\hat{\mu}_N(a, b_N, c_N)) = R(\overline{X}_N) + a\sigma_1^2 E\left[\frac{1}{N}\left\{(1 + \rho N/l)\frac{a}{(1 + \overline{R}_N)^2} - \frac{2}{1 + \overline{R}_N}\right\}\right],$$

which gives that $R(\hat{\mu}_N(a, b_N, c_N)) \leq R(\overline{X}_N)$ if and only if the second term in the r.h.s. of (2.4) is not positive. Here, observe that for $0 < a \leq 1$, $\theta > 0$ and $X > 0$,

$$(2.5) \quad \frac{(1+\theta)a}{(1+\theta X)^2} - \frac{2}{1+\theta X} \leq \frac{a}{1+\theta a}(aX^{-2} - 2X^{-1}),$$

which follows from the inequality $2\theta X(X-a)^2 + (a\theta + 1)(1-a)X^2 + (X-a)^2 \geq 0$. Putting $\theta = \rho N/l$ and $X = \overline{R}_N l/(N\rho)$ in (2.5), we see that

$$E\left[\frac{1}{N}\left\{(1 + \rho N/l)\frac{a}{(1 + \overline{R}_N)^2} - \frac{2}{1 + \overline{R}_N}\right\}\right] \leq E\left[\frac{a}{N(1 + a\rho N/l)}\{a(\rho N/l)^2 \overline{R}_N^{-2} - 2(\rho N/l)\overline{R}_N^{-1}\}\right].$$

Hence it is sufficient to show that

$$(2.6) \quad h(\sigma_1^2, \sigma_2^2) \stackrel{\text{def}}{=} E\left[\frac{1}{N(1 + a\rho N/l)}\{(\rho N/l)^2 \overline{R}_N^{-2} - (2/a)(\rho N/l)\overline{R}_N^{-1}\}\right] \leq 0,$$

for any $\sigma_1^2, \sigma_2^2 > 0$. This is also represented as

$$(2.7) \quad h(\sigma_1^2, \sigma_2^2) = \sum_{n=m}^{\infty} \frac{1}{n(1 + a\rho n/l)} E\left[\{(\rho n/l)^2 \overline{R}_n^{-2} - \frac{2}{a}(\rho n/l)\overline{R}_n^{-1}\} I_{[N=n]}\right]$$
$$= \sum_{n=m}^{\infty} \frac{1}{n(1 + a\rho n/l)} E\left[\frac{(\rho n/l)^2}{\overline{R}_n^2} I_{[N=n]}\right]$$

$$\cdot \left\{ 1 - \frac{2}{a} \cdot \frac{E[(\rho n/l)\bar{R}_n^{-1} I_{[N=n]}]}{E[(\rho n/l)^2 \bar{R}_n^{-2} I_{[N=n]}]} \right\},$$

where for $n \geq m + 1$, $I_{[N=n]} = I_{[B(m-3)(n-1) < S_{1m} \leq B(m-3)n]}$ and $I_{[N=m]} = I_{[S_{1m} \leq B(m-3)m]}$.

To prove the inequality (2.6) by use of (2.7), given $N = n$, we express $\bar{R}_n$ by other mutually independent random variables whose distributions do not depend on unknown parameters. For $n \geq m + 1$, let $Q_{1n} = S_{1n} - S_{1m}$ and denote $U_1 = S_{1m}/\sigma_1^2$ and $U_{2n} = Q_{1n}/\sigma_1^2$. Also let $V = S_2/\sigma_2^2 + T$ and $W = T(S_2/\sigma_2^2 + T)^{-1}$. Then it is seen that U_1, U_{2n}, V and W are mutually independent, and that $U_1 \sim \chi^2_{m-1}$, $U_{2n} \sim \chi^2_{n-m}$, $V \sim \chi^2_{l+2}$ and $W \sim$ beta $\{3/2, (l-1)/2\}$, that is, a random variable having a beta distribution with parameters $\{3/2, (l-1)/2\}$. Note that $V = S_2/\sigma_2^2 + VW$. Then

$$\frac{\rho n/l}{\bar{R}_n} = \frac{U_1 + U_{2n}}{\{b_n(1 - W) + c_n W + c_n Wl/(\rho n)\}V}.$$

By using this expression, the ratio of expectations in (2.7) is rewritten as

$$\frac{E[(\rho n/l)\bar{R}_n^{-1} I_{[N=n]}]}{E[(\rho n/l)^2 \bar{R}_n^{-2} I_{[N=n]}]} = \frac{E[\{b_n(1 - W) + c_n W + c_n Wl/(\rho n)\}^{-1}]E[V^{-1}]}{E[\{b_n(1 - W) + c_n W + c_n Wl/(\rho n)\}^{-2}]E[V^{-2}]}$$
$$\cdot \frac{E[(U_1 + U_{2n})I_{[N=n]}]}{E[(U_1 + U_{2n})^2 I_{[N=n]}]}.$$

From Proof of Theorem 2.1 in Kubokawa (1987*b*), we observe that

$$\frac{E[\{b_n(1 - W) + c_n W + c_n Wl/(\rho n)\}^{-1}]}{E[\{b_n(1 - W) + c_n W + c_n Wl/(\rho n)\}^{-2}]}$$
$$\geq \frac{E[\{b_n(1 - W) + c_n W + c_n Wl/(\rho n)\}^{-1}]}{E[\{b_n(1 - W) + c_n W + c_n Wl/(\rho n)\}^{-1}\{b_n(1 - W) + c_n W\}^{-1}]}$$
$$\geq \frac{E[\{b_n(1 - W) + c_n W\}^{-1}]}{E[\{b_n(1 - W) + c_n W\}^{-2}]} \stackrel{\text{def}}{=} u(b_n, c_n; 0),$$

where $u(b_n, c_n; 0)$ is the notation corresponding to (2.2) of Kubokawa (1987*b*). In other words, the second inequality follows from Theorem 2.1 of Bhattacharya (1984) and the fact that both $\{b_n(1 - W) + c_n W\}/\{b_n(1 - W) + c_n W + c_n Wl/(\rho n)\}$ and $b_n(1 - W) + c_n W$ are nonincreasing in W for $b_n \geq c_n$. Also observe that $E[V^{-1}]/E[V^{-2}] = l - 2$. Thus from (2.7), we have

$$(2.8) \qquad h(\sigma_1^2, \sigma_2^2) \leq \sum_{n=m}^{\infty} \alpha_n\{E[(U_1 + U_{2n})^2 I_{[N=n]}]$$

$$- (2/a)(l-2)u(b_n, c_n; 0)E[(U_1 + U_{2n})I_{[N=n]}]\} ,$$

where

$$(2.9) \qquad \alpha_n = \frac{1}{n(1+a\rho n/l)} \cdot \frac{E[(\rho n/l)^2 \bar{R}_n^{-2} I_{[N=n]}]}{E[(U_1+U_{2n})^2 I_{[N=n]}]} .$$

Since U_1 and U_{2n} are independent, we see that

$$(2.10) \qquad E[(U_1+U_{2n})^2 I_{[N=n]}] = E[U_1^2 I_{[N=n]}] + 2(n-m)E[U_1 I_{[N=n]}] + (n-m+2)(n-m)E[I_{[N=n]}] .$$

Here, writing $\beta_{m-1} = 0$, $\beta_n = B(m-3)n/\sigma_1^2$, for $n \geq m$, we define $q_f(n)$ by

$$q_f(n) = P\{\beta_{n-1} < \chi_f^2 \leq \beta_n\} ,$$

for positive integer f. Then, $P(N=n) = q_{m-1}(n)$. Also, note that if $f_k(x)$ is a chi-square density with k degrees of freedom, $x^r \cdot f_k(x) = 2^r\{\Gamma(k/2+r)/\Gamma(k/2)\}f_{k+2r}$ for real r. Thus from (2.10),

$$(2.11) \qquad E[(U_1+U_{2n})^2 I_{[N=n]}] = (m+1)(m-1)q_{m+3}(n) + 2(m-1)(n-m)q_{m+1}(n) + (n-m+2)(n-m)q_{m-1}(n) .$$

Similarly,

$$(2.12) \qquad E[(U_1+U_{2n})I_{[N=n]}] = (m-1)q_{m+1}(n) + (n-m)q_{m-1}(n) .$$

From (2.8), (2.11) and (2.12), we get

$$(2.13) \quad h(\sigma_1^2, \sigma_2^2) \leq (m+1)(m-1) \cdot \sum_{n=m}^{\infty} \alpha_n \Big[q_{m+3}(n) - \frac{2}{m+1}\{(l-2)u(b_n, c_n; 0)/a - (n-m)\}q_{m+1}(n) + \frac{n-m}{(m+1)(m-1)} \cdot \{(\mathrm{n}-m+2) - 2(l-2)u(b_n, c_n; 0)/a\}q_{m-1}(n) \Big] .$$

By Lemma 2.2 of Kubokawa (1987*b*), observe that

$$u(b_n, c_n; 0) = (l-5)b_n/(l-2) \qquad \text{for} \quad c_n = 0\,,$$

$$u(b_n, c_n; 0) \geq \left\{ 1 + \frac{(l-1)(b_n - c_n)}{(l+2)b_n} \right\} c_n \qquad \text{for} \quad b_n \geq c_n > 0\,.$$

Combining these relations and the conditions (a) and (c) of Theorem 2.1 can show that

$$(l-2)u(b_n, c_n; 0)/a - (n-m) \geq (m+1)/2\,,$$
$$n - m + 2 \leq 2(l-2)u(b_n, c_n; 0)/a\,.$$

Hence from (2.13), the inequality (2.6) can be proved if

$$\sum_{n=m}^{\infty} \alpha_n\{q_{m+3}(n) - q_{m+1}(n)\} \leq 0\,. \tag{2.14}$$

Since α_n given by (2.9) is rewritten as $\alpha_n = (\rho/l)^2(1 + a\rho n/l)^{-1}(b_n^2/n)^{-1} \cdot E[\{(\rho n/l)\chi_{l-1}^2 + (c_n/b_n)(1 + \rho n/l)T\}^{-2}]$, we can show that α_n is decreasing in n from the condition (b), and that $\sum_{n=m}^{\infty} \alpha_n < \infty$. Finally, from Ghosh and Sen ((1983), p. 363), we have

$$\begin{aligned}
\sum_{n=m}^{\infty} \alpha_n\{q_{m+3}(n) - q_{m+1}(n)\} \qquad\qquad & (2.15)\\
= \sum_{n=m}^{\infty} \left\{ \sum_{j=n}^{\infty} (\alpha_j - \alpha_{j+1}) \right\}\{q_{m+3}(n) - q_{m+1}(n)\}&\\
= \sum_{j=m}^{\infty} \left[\sum_{n=m}^{j} \{q_{m+3}(n) - q_{m+1}(n)\} \right] (\alpha_j - \alpha_{j+1})&\\
= \sum_{j=m}^{\infty} \left[P\{\chi_{m+3}^2 \leq \beta_j\} - P\{\chi_{m+1}^2 \leq \beta_j\} \right] (\alpha_j - \alpha_{j+1})&\\
< 0\,,&
\end{aligned}$$

which proves (2.14). Therefore the proof of Theorem 2.1 is complete.

3. k (≥ 3) sample problem

In this section, k (≥ 3) sample problem is discussed. Following the stopping number N given by (1.2), sample $X_{11},\ldots,X_{1N}$ is taken from $N(\mu,\sigma_1^2)$. We further assume that independent random samples $(X_{21},\ldots, X_{2l_2}),\ldots,(X_{k1},\ldots,X_{kl_k})$ are additionally taken, where each X_{ij} has $N(\mu,\sigma_i^2)$. Denote $\overline{X}_{1N} = N^{-1}\sum_{j=1}^{N} X_{1j}$, $S_{1N} = \sum_{j=1}^{N} (X_{1j} - \overline{X}_{1N})^2$, $\overline{X}_i = l_i^{-1}\sum_{j=1}^{l_i} X_{ij}$ and $S_i =$

$\sum_{j=1}^{l_i}(X_{ij}-\overline{X}_i)^2$ for $i=2,\dots,k$. Let $a_2,\dots,a_k$ be positive constants, and let $b_{2N},\dots,b_{kN}$ be functions of N. Consider the combined estimators based on k samples of the form

$$\hat{\mu}_{jN}=\overline{X}_{1N}+\sum_{i=2}^{j}\phi_{iN}(\overline{X}_i-\overline{X}_{1N}) \qquad j=2,\dots,k\,, \tag{3.1}$$

where

$$\phi_{iN}=\frac{a_iS_{1N}}{S_{1N}+b_{iN}(N/l_i)S_i} \qquad i=2,\dots,k\,.$$

The unbiasedness of $\hat{\mu}_{jN}$ easily follows. This type of estimators was introduced by Bhattacharya (1980) extending a particular case [namely, $b_{iN}=(N-1)/(l_i-1)$], proposed by Brown and Cohen (1974), and was recently treated by Sugiura and Kubokawa (1986) in the problem of estimating common parameters of growth curve models.

THEOREM 3.1. *Assume that* $l_2\ge 6,\dots,l_k\ge 6$, *and that the following conditions hold for* $j=2,\dots,k$.

(a) nb_{jn}^2 *is nondecreasing in* n *for* $n\ge m$ *and* $\sum_{n=m}^{\infty}(nb_{jn})^{-2}<\infty$.

(b) $0<a_2\le\min\left[1,\,2\cdot\inf_{n\ge m}\left\{\frac{l_2-5}{2n-m+1}b_{2n}\right\}\right]$,

$$0<\frac{a_3}{1-a_2}\le\min\left[1,\,2\cdot\inf_{n\ge m}\left\{\frac{l_3-5}{2n-m+1}b_{3n}\right\}\right],$$

$$\vdots$$

$$0<\frac{a_k}{1-a_2-\cdots-a_{k-1}}\le\min\left[1,\,2\cdot\inf_{n\ge m}\left\{\frac{l_k-5}{2n-m+1}b_{kn}\right\}\right].$$

Then we have

$$R(\hat{\mu}_{kN})\le R(\hat{\mu}_{k-1,N})\le\cdots\le R(\hat{\mu}_{2N})\le R(\overline{X}_{1N})<B\,,$$

for a preassigned constant $B>0$.

PROOF. Let $C_j=a_j/(1-a_j-\cdots-a_{j-1})$; $R_{jN}=b_{jN}(S_j/l_j)/(S_{1N}/N)$; $t_{jN}=\overline{X}_{1N}+C_j(\overline{X}_j-\overline{X}_{1N})/(1+R_{jN})$, $j=2,\dots,k$. Then the result follows in the same way as in Brown and Cohen (1974) once it is noted that for each $j\ge 2$, $R(t_{jN})\le R(\overline{X}_{1N})$ in view of Theorem 2.1.

Remark. It is interesting if we could show that Shinozaki's estimator (1978) dominates $\overline{X}_{1N}$.

Acknowledgements

The author is very grateful to Professor N. Sugiura of University of Tsukuba, and Dr. Y. Takada of Kumamoto University for their helpful suggestions and valuable comments. The author is also grateful to the referees for their valuable comments, which led to improvements in the presentation of the results.

REFERENCES

Bhattacharya, C. G. (1980). Estimation of a common mean and recovery of interblock information, *Ann. Statist.*, **8**, 205–211.

Bhattacharya, C. G. (1984). Two inequalities with an application, *Ann. Inst. Statist. Math.*, **36**, 129–134.

Brown, L. D. and Cohen, A. (1974). Point and confidence estimation of a common mean and recovery of interblock information, *Ann. Statist.*, **2**, 963–976.

Ghosh, M. and Sen, P. K. (1983). On two-stage James-Stein estimator, *Sequential Anal.*, **2**, 359–367.

Ghosh, M., Nickerson, D. M. and Sen, P. K. (1987). Sequential shrinkage estimation, *Ann. Statist.*, **15**, 817–829.

Khatri, C. G. and Shah, K. R. (1974). Estimation of location parameters from two linear models under normality, *Comm. Statist. A—Theory Methods*, **3**, 647–663.

Kubokawa, T. (1987*a*). Estimation of a common mean of two normal distributions, *Tsukuba J. Math.*, **11**, 157–175.

Kubokawa, T. (1987*b*). Admissible minimax estimation of a common mean of two normal populations, *Ann. Statist.*, **15**, 1245–1256.

Nickerson, D. M. (1987). Sequential shrinkage estimation of linear regression parameters, *Sequential Anal.*, **6**, 93–117.

Rao, C. R. (1973). *Linear Statistical Inference and Its Application*, 2nd ed., Wiley, New York.

Shinozaki, N. (1978). A note on estimating the common mean of K normal distributions and Stein problem, *Comm. Statist. A—Theory Methods*, **7**, 1421–1432.

Sugiura, N. and Kubokawa, T. (1986). Estimating common parameters of growth curve models, *Ann. Inst. Statist. Math.*, **40**, 119–135.

Takada, Y. (1984). Inadmissibility of sequential estimation rule of the mean of a multivariate normal distribution, *Sequential Anal.*, **3**, 267–271.