Kumamoto J. Math
Vol. 1, 1~8 March(1988)

TWO-STAGE PROCEDURES FOR A MULTIVARIATE NORMAL DISTRIBUTION

Yoshikazu TAKADA

(Received October 1, 1987)

1. INTRODUCTION

Let $\{X_i; i \geqq 1\}$ be a sequence of independently distributed p-dimensional normal random vectors with unknown mean μ and unknown covariance matrix Σ. We consider the following two problems for the mean vector μ.

(I) Given $0<1-\alpha<1$ and $d>0$, we want to find a region R in p-dimensional Euclidean space such that $P(\mu \varepsilon R) \geqq 1-\alpha$ for all (μ, Σ) and the maximum diameter of R does not exceed 2d.

(II) Given $W>0$, we want to construct an estimator δ of μ such that $E\|\delta-\mu\|^2 \leqq W$ for all (μ, Σ), where $\|t\|^2=t't$.

It follows easily from the following theorem that there does not exist a fixed sample procedure to meet such requirements.

THEOREM 1. *Let $X_1, \ldots, X_n$ be independent and identically distributed p-dimensional random vectors with the probability density function with respect to Lebesgue measure,*

$$\lambda^{-p} f(\lambda^{-1}(x-\mu)),$$

where $\lambda>0$, f is some known function and $\theta=(\mu, \lambda)$ is unknown.
Let $L(\theta, d)=\rho(\|d-\mu\|)$ be a loss function, where $\rho \geqq 0$ is a non-decreasing function defined on $[0, \infty)$, and let $M=\mathrm{Sup}\rho(u)$, which may be infinite. Then given any $W<M$, there does not exist any estimator of μ whose risk is bounded by W for all θ.

For p=1, Lehmann [8] proved it under the assumption that f is continuous almost everywhere (cf. Example 4.1 of Singh [12]). Our proof, which is given in Section 4, is different from those of Lehmann and Singh and such an assumption is not needed.

Healy [7] constructed a confidence region of the problem (I). The method is based on the two-stage procedure of Stein [13] for the univariate case. When p=1, Stein's procedure is not asymptotically efficient (Ghosh and Mukhopadhyay [4]). In Section 2 we show that Healy's procedure is not asymptotically efficient at least for p=2, but becomes asymptoticall efficient by choosing the first sample size properly (cf. Mukhopadhyay [9]). When $\Sigma=\sigma^2 H$ with unknown $\sigma>0$ and p×p known positive definite matrix H, Mukhopadhyay and Al-Mousawi [10] considered the same problem. In Section 3 we construct an estimator of the problem (II). For the univariate case, see Rao [11] (pp. 486-487).

2. CONFIDENCE REGION

Healy [7] proposed the following two-stage procedure to the problem (I). Let n(>p) be the first sample size and

$$\bar{X}_n=n^{-1}\sum_{i=1}^{n}X_i, \qquad S_n=(n-1)^{-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)(X_i-\bar{X}_n)'.$$

Determine a constant f_n such that $P(F_n<f_n)=1-\alpha$, where F_n/p has F distribution with (p, n−p) degrees of freedom. Define the random sample size N by

$$N=Max\{n,\ [\frac{f_n(n-1)}{d^2(n-p)}\hat{\lambda}_n]+1\}, \tag{2.1}$$

where [u] denotes the largest integer less than u and $\hat{\lambda}_n$ is the largest characteristic root of S_n. Then the confidence region R_N is defined by

$$R_N=\{\mu;\ N(\bar{X}_N-\mu)'S_n^{-1}(\bar{X}_N-\mu)\leqq(n-1)(n-p)^{-1}f_n\}$$

with $\bar{X}_N=N^{-1}\sum_{i=1}^{N}X_i$. Healy [7] showed that R_N satisfies the requirements of problem (I).

In this section, we consider properties of the procedure. If Σ were known, we would propose a region R defined by

$$R=\{\mu;\ m(\bar{X}_m-\mu)'\Sigma^{-1}(\bar{X}_m-\mu)\leqq f\},$$

where m is the least integer greater than $\lambda f/d^2$, f is the 100(1−α)% point of χ^2 distribution with p degrees of freedom, λ is the largest characteristic root of Σ and $\bar{X}_m=m^{-1}\sum_{i=1}^{m}X_i$. It is easy to show that R satisfies the requirements of problem (I).

By (2.1) we have the inequality

$$\frac{f_n(n-1)}{d^2(n-p)}\hat{\lambda}_n\leqq N\leqq\frac{f_n(n-1)}{d^2(n-p)}\hat{\lambda}_n+n. \tag{2.2}$$

The following lemma, which is proved by Cacoullos and Olkin [1], is useful for the subsequent discussion.

LEMMA 1. *If Z is a $p\times p$ real random matrix with only real characteristic roots and $E(Z)=A$, then*

$$E\{\lambda_1(Z)\}\geqq\lambda_1(A),\qquad E\{\lambda_p(Z)\}\leqq\lambda_p(A),$$

where λ_1 and λ_p denote the largest and smallest characteristic roots.

Since $E(S_n)=\Sigma$, Lemma 1 implies that $E(\hat{\lambda}_n)\geqq\lambda$. Hence from the left hand side of the inequality (2.2), we obtain that

$$E(N)/c\geqq(n-1)f_n/\{(n-p)f\}, \tag{2.3}$$

where $c=\lambda f/d^2$.

From Theorem 4 of Ghosh [3], we obtain that $f_n\geqq f$ for p=2, so that by (2.3)

$$E(N)/c\geqq(n-1)/(n-p).$$

Hence if the first sample size does not depend on d,

$$\underline{lim}_{d\to 0}\ E(N)/c>1,$$

which implies that the Healy's procedure is not asymptotically efficient for at least p=2. But we conjecture that $f_n\geqq f$ for all p if α is small, so that Healy's procedure is not asymptotically efficient for all p if the first sample size n does not depend on d. But it is possible to make the procedure asymptotically efficient by letting the first sample size n depend on d (see Mukhopadhyay [9] for the univariate case).

LEMMA 2. (*i*) $\hat{\lambda}_n\to\lambda$ *a.s.* *as* $n\to\infty$.
(*ii*) $E(\hat{\lambda}_n)\to\lambda$ *as* $n\to\infty$.

PROOF. Note that $S_n\to\Sigma$ *a. s. as* $n\to\infty$. Hence (i) is obtained. Let

$$W_i=\{X_1+\ldots+X_{i-1}-(i-1)X_i\}/\{i(i-1)\}^{1/2}$$

for $i\geqq2$. Then it follows easily that $\{W_i; i\geqq2\}$ is a sequence of independently and identically distributed normal random vectors with mean zero and covariance matrix

Σ, and that for $n \geqq 2$

$$S_n = (n-1)^{-1} \sum_{i=2}^{n} W_i W_i'. \tag{2.4}$$

By (2.4) we have that

$$tr(S_n) = (n-1)^{-1} \sum_{i=2}^{n} \| W_i \|^2,$$

where $tr(S_n)$ denote the trace of S_n. Hence $\{tr(S_n); n \geqq 2\}$ becomes a reverse martingale. Using the Doob's moment inequality (Doob [2], p. 318), we have that

$$E\{\sup_n (tr(S_n))\} < \infty.$$

Since $\hat{\lambda}_n \leqq \sup_n (tr(S_n))$, by (i) and the dominated convergence theorem, we obtain (ii). This completes the proof.

THEOREM 2. *If the first sample size $n = n(d)$ is chosen such that*

$$n(d) \to \infty \text{ and } \quad d^2 n(d) \to 0 \quad \text{as } d \to 0,$$

then

$$(i)\ N/c \to 1 \quad a.s. \quad \text{as } d \to 0$$

and

$$(ii)\ E(N)/c \to 1 \qquad \text{as } d \to 0 \text{ (asymptotic efficiency).}$$

PROOF. From (2.2) we have that

$$\frac{(n-1) f_n \hat{\lambda}_n}{(n-p) f \lambda} \leqq \frac{N}{c} \leqq \frac{(n-1) f_n \hat{\lambda}_n}{(n-p) f \lambda} + \frac{d^2 n}{f \lambda}. \tag{2.5}$$

Note that $f_n \to f$ *as* $d \to 0$. Then (i) and (ii) are proved by Lemma 2.

REMARK 1. From the left hand side of the inequality (2.2),

$$N - c \geqq \frac{f_n (n-1)}{d^2 (n-p)} \hat{\lambda}_n - \frac{\lambda f}{d^2}$$

Note that $f_n \geqq f$ for $p=2$ and $E(\hat{\lambda}_n) \geqq \lambda$. Hence we have that

$$E(N-c) \geqq \frac{f_n (n-1) \lambda}{d^2 (n-p)} - \frac{\lambda f}{d^2} \geqq \frac{\lambda f (p-1)}{d^2 (n-p)},$$

from which we have that for $p=2$

$$E(N-c)\to\infty \quad as\ d\to 0$$

(cf. Gosh and Mukhopadhyay [5], p. 223).

3. BOUNDED MEAN SQUARED ERROR

In this section we consider the problem (II), that is, we construct an estimator δ of μ such that

$$E\|\delta-\mu\|^2\leqq W \tag{3.1}$$

for all (μ, Σ). Such an estimator is constructed by a two-stage procedure similar to that of Stein [13]. For the univariate case, see Rao [11] (pp. 486-487).

THEOREM 3. *Let n $(>p+2)$ be the first sample size. Define the random sample size N by*

$$N=Max\{n, [\frac{p(n-1)}{W(n-p-2)}\hat{\lambda}_n]+1\} \tag{3.2}$$

Estimate μ by $\bar{X}_N$. Then $\bar{X}_N$ satisfies (3.1)

PROOF. Note that given S_n, $\bar{X}_N$ is normally distributed with mean μ and covariance matrix Σ/N. Hence

$$E\|\bar{X}_N-\mu\|^2=tr(\Sigma)E(N^{-1}). \tag{3.3}$$

From (3.2) we have the inequality

$$\frac{p(n-1)}{W(n-p-2)}\hat{\lambda}_n\leqq N\leqq\frac{p(n-1)}{W(n-p-2)}\hat{\lambda}_n+n. \tag{3.4}$$

Then from the left hand side of the inequality (3.4), we obtain that

$$E(N^{-1})\leqq\frac{W(n-p-2)}{p(n-1)}E(\hat{\lambda}_n^{-1}). \tag{3.5}$$

Note that $E(S_n^{-1})=(n-1)(n-p-2)^{-1}\Sigma^{-1}$ (e.g. Giri [6], p. 104). Hence by Lemma 1 we have that

$$E(\hat{\lambda}_n^{-1})\leqq(n-1)(n-p-2)^{-1}\lambda^{-1},$$

so that it follows from (3. 3) and (3. 5)that

$$E\|\bar{X}_N-\mu\|^2\leqq W(p\lambda)^{-1}tr(\Sigma)\leqq W.$$

This completes the proof.

Next we consider properties of the procedure. If Σ were known, we would use $\bar{X}_m$ as an estimator of μ satisfying (3. 1), where m is the least integer greater than $c=tr(\Sigma)/W$. By the left hand side of the inequality and Lemma 1, we obtain that

$$\frac{E(N)}{c}\geqq\frac{(n-1)p\lambda}{(n-p-2)tr(\Sigma)}\geqq\frac{n-1}{n-p-2}>1.$$

Hence the procedure is not asymptotically efficient if the first sample size n does not depend on W (cf. Ghosh and Mukhopadhyay [4], p. 207).

LEMMA 3. *If the first sample size $n=n(W)$ is chosen such that*

(3. 6) $$n(W)\to\infty \text{ and } Wn(W)\to 0 \quad \text{as } W\to 0,$$

Then

$$N/c\to p\lambda/tr(\Sigma) \text{ a. s.} \quad \text{as } W\to 0$$

and

$$E(N)/c\to p\lambda/tr(\Sigma) \quad \text{as } W\to 0.$$

PROOF. It follows from (3. 4) that

$$\frac{(n-1)p\hat{\lambda}_n}{(n-p-2)tr(\Sigma)}\leqq\frac{N}{c}\leqq\frac{(n-1)p\hat{\lambda}_n}{(n-p-2)tr(\Sigma)}+\frac{nW}{tr(\Sigma)}.$$

Hence by Lemma 2 the proof is completed.

The following theorem is easily obtained from Lemma 3.

THEOREM 4. *If $\Sigma=\lambda I_p$ (I_p denotes the identity matrix) and the first sample size n satisfies (3. 6), then*

$$N/c\to 1 \quad \text{a. s.} \quad \text{as } W\to 0$$

and

$$E(N)/c\to 1 \quad \text{as } W\to 0 \text{ (asymptotic efficiency).}$$

REMARK 2. From the left hand side of the inequality (3. 4) and $E(\hat{\lambda}_n)\geqq\lambda$,

$$E(N-c) \geqq \frac{p(n-1)}{W(n-p-2)}E(\hat{\lambda}_n)-\frac{tr(\Sigma)}{W} \geqq \frac{(p+1)tr(\Sigma)}{W(n-p-2)}.$$

Hence we have that

$$E(N-c)\to\infty \qquad as\ W\to 0$$

(cf. Remark 1).

REMARK 3. From Lemma 3, it follows that the procedure is not asymptotically efficient when $\Sigma \neq \lambda I_p$. At present, we can not construct a procedure which satisfies (3.1) and is asymptotically efficient for all Σ.

4. PROOF OF THEOREM 1

Given λ, consider the following function $h_\lambda(\mu)$ of μ,

$$h_\lambda(\mu)=(2\pi)^{-p/2}\lambda^{-p}exp(-\|\mu\|^2/2\lambda^2).$$

For any estimator δ, let

$$\tau_\lambda(\delta)=\int R(\theta,\delta)h_\lambda(\mu)d\mu.$$

where

$$R(\theta,\delta)=E_\theta\rho(\|\delta(\underline{X})-\mu\|)$$

and $\underline{X}=(X_1,\ldots,X_n)$. Letting $h(\mu)=h_1(\mu)$ and $\theta_\lambda=(\lambda\mu,\lambda)$, we have that

$$\tau_\lambda(\delta)=\int R(\theta_\lambda,\delta)h(\mu)d\mu$$

$$=\int\{E_{\theta_1}\rho(\|\delta(\lambda\underline{X})-\lambda\mu\|)\}h(\mu)d\mu.$$

By the tedious calculation (cf. Theorem 1 of Takada [14]), it can be shown that

$$\varliminf_{\lambda\to\infty}\tau_\lambda(\delta)\geqq M. \tag{4.1}$$

Note that

$$M\geqq \operatorname*{Sup}_{\theta} R(\theta,\delta)\geqq \operatorname*{Sup}_{\lambda}\{\operatorname*{Sup}_{\mu} R(\theta,\delta)\}\geqq \operatorname*{Sup}_{\lambda} \tau_\lambda(\delta).$$

Hence by (4.1) we obtain that

$$\operatorname*{Sup}_{\theta} R(\theta,\delta)=M,$$

from which the proof is completed.

REFERENCES

[1] Cacoullos, T. and Olkin, I., On the bias of functions of characteristic roots of a random matrix, Biometrika, 52(1965), 87—94.

[2] Doob, J.L., Stochastic Processes, 2nd ed., Wiley, New York, 1967.

[3] Ghosh, B.K., Some monotonicity theorem for χ^2, F and t distribution with applications, J. Roy. Statist. Soc., B, 35(1973), 480—492.

[4] Ghosh, M. and Mukhopadhyay, N., On two fundamental problems of sequential estimation, Sankhyā, A, 38(1976), 203—218.

[5] Ghosh, M. and Mukhopadhyay, N., Consistency and asymptotic efficiency of two stage and sepuential estimation procedures, Sankhyā, A, 43(1981), 220—227.

[6] Giri, N.C., Multivariate Statistical Inference, Academic Press, New York, 1977.

[7] Healy, W.C.Jr., Two-sample procedures in simultaneous estimation, Ann. Math. Statist. 27(1956), 687—702.

[8] Lehmann, E.L., Notes on the theory of estimation, University of Carifornia Press, 1950.

[9] Mukhopadhyay, N., A consistent and asymptotically efficient two-stage procedure to construct fixed width confidence intervals for the mean, Metrika, 27 (1980), 281—284.

[10] Mukhopadhyay, N. and Al-Mousawi, J.S., Fixed-size confidence regions for the mean vector of a multinormal distribution, Sequential Analysis, 5(1986), 139—168.

[11] Rao, C.R., Linear Statistical Inference and Its Application, 2nd ed., Wiley, New York.

[12] Singh, R., Existence of bounded length confidence intervals, Ann, Math. Statist., 34(1963), 1474—1485.

[13] Stein, C., A two-sample test for a linear hypothesis whose power is independent of the variance, Ann. Math. Statist., 16(1945), 245—258.

[14] Takada, Y., Non-existence of fixed sample size procedures for scale families, Sequential Analysis, 5(1986), 93—101.

Department of Mathematics
Faculty of Science
Kumamoto University