

STATISTICAL INFERENCE IN TWO-STAGE SAMPLING

Makoto Aoshima

Institute of Mathematics

University of Tsukuba

Ibaraki 305-8571, Japan

Email: aoshima@math.tsukuba.ac.jp

(This article originally appeared in Japanese in *Sūgaku* **54** (4) (2000), 365–382.)

1. Introduction

Let X_1, \dots, X_n be i.i.d. random variables with the normal distribution $N(\mu, \sigma^2)$. Consider a hypothesis testing $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1$ ($\mu_0 \neq \mu_1$). Stein [S45] proved a result originally given by Dantzig [D40] that there does not exist a critical region such that the probability of error of first kind is less than the power for the testing of the mean when the scale parameter σ is unknown as long as one considers a single-stage sampling scheme in which n is fixed in advance. That argument was extended to the case of the location-scale family of distributions by Chatterjee [C91]. Consider a fixed-width confidence interval for μ or a bounded risk point estimation for μ with the squared loss. For the location-scale family of distributions with unknown scale parameter, Lehmann [L51] proved that there does not exist such an estimation for the location parameter μ as long as one considers a single-stage sampling scheme. Both these conclusions essentially follow from the fact that two distributions with different locations can be made to approach each other uniformly in measure arbitrarily closely by making the scale parameter sufficiently large. Recently, Takada [T98] gave a sufficient condition for the nonexistence of such statistical procedures in terms of the distance of distributions under a single-stage sampling scheme.

To handle with the inference described above, one is required to perform at least two stages of sampling. For most cases, two-stage sampling schemes successfully give a solution to those problems about fixed-size inference. There are cases, however, in which no sampling scheme exists, with a predetermined number of stages, that can yield fixed-size inference. Then, purely sequential sampling schemes are required. (See Farrell [F66], Blum and Rosenblatt [BR66] and Koopmans et al. [KOR64].) Moreover, there are cases in which there does not exist any sampling scheme that can guarantee fixed-size inference. (See Bahadur and Savage [BaS56] and Blum and Rosenblatt [BR69a].)

for such examples. As for necessary and sufficient conditions about existence of certain multistage or purely sequential sampling schemes for fixed-size inference, refer to Blum and Rosenblatt [BR63] and Singh [Sin63].) Zacks [Z71, Chap.10] gives a rich source of this context. This article deals with inference problems which should be solved in at least two stages of sampling.

The two-stage procedure is a statistical method originally created by Stein [S45] in 1945 to give a solution to inference problems, whose precision was fixed in advance, with the smallest number of required stages. In that procedure, the size of the second stage sample is a random variable that is defined after observing the first stage sample. The two-stage procedure has been applied to various fields of research afterwards and has given an impact on development of following methodologies: The purely sequential procedure given by Chow and Robbins [CR65] aims at reducing the average sample number asymptotically by taking a sample one by one repeatedly till it gets an approximate solution. It was also followed by the three-stage procedure given by Hall [Ha81], the accelerated sequential procedure given by Hall [Ha83], the modified accelerated sequential procedure given by Liu [Li97a], the multistage procedure given by Liu [Li97b] and the modified three-stage procedure given by Holm [Hol95, 99]. Even though development of methodology advances, the two-stage procedure does not lose brightness as an object of study. It is because of its utility. For instance, there is a case pressed for making a statistical decision with demanded precision as soon as possible. There is a case that a cost to appear when one enforces sampling becomes serious. There is a case to include the fear that changes a state of population by repeating observation. For these all affairs that can happen in a practical scene, it may be said that the two-stage procedure is the methodology that is easy to deal.

Basic principle appeared in [S45] which Stein first invented the two-stage procedure is really elegant and marvelously lucid. Actually, the later a lot of researchers have still gotten a lot of hints from this basic principle. In this article, without losing essence of the idea, the developments after birth of the two-stage procedure shall be addressed mainly on bounded risk problems. About testing problems due to the two-stage procedure, one may refer to Hewett and Spurrier [HeS83].

2. Essence of Stein's two-stage procedure

We consider inference problems on the location parameter θ of a location-scale distribution $F((x-\theta)/\xi)$. The purpose of this section is to expose essence of Stein's two-stage procedure and give help of understanding the main subject of this article after Section 3.

Let $\{X_i; i \geq 1\}$ be i.i.d. with a p.d.f. $\xi^{-1}f((x-\theta)/\xi)$ where unknown parameters $(\theta, \xi) \in R \times R^+$. Here, $f(\cdot)$ is known and continuous and its domain space may depend on θ alone.

Let $T_n = T_n(X_1, \dots, X_n)$ and $U_n = U_n(X_1, \dots, X_n)$ be estimators of θ and ξ respectively based on

samples (X_1, \dots, X_n) . Suppose that T_n and U_n satisfy the following conditions:

(a) For any fixed n (≥ 2), T_n is independent of (U_2, \dots, U_n) .

(b) (i) For some $\gamma > 0$, for a measurable function $g : R^+ \rightarrow R^+$, the distribution of $n^\gamma(T_n - \theta)/g(\xi)$ does not depend on (n, θ, ξ) ;

(ii) The distribution of $n^\gamma(T_n - \theta)/g(U_n)$ does not depend on (θ, ξ) .

Given preassigned width $2d > 0$ and confidence coefficient $1 - \alpha > 0$, let us construct a confidence interval for θ by using Stein's two-stage procedure. Having recorded X_1, \dots, X_m of a suitable size m (≥ 2), calculate U_m and define

$$N = \max\{m, [(b_m g(U_m)/d)^{1/\gamma}] + 1\}, \quad (1)$$

where $b_m > 0$ is a design constant and, here and throughout, $[u]$ will stand for the greatest integer less than u . Next, take an additional sample X_{m+1}, \dots, X_N of size $N - m$ accordingly and calculate T_N based on all the samples. Then, the fixed-width confidence interval for θ is finally obtained as $[T_N - d, T_N + d]$, keeping in mind that the lower or upper confidence limit will be modified suitably if the domain space depends on θ . Let us verify satisfaction of the probability requirement. Noting the stopping rule (1) and assumptions (a)-(b), we have

$$\begin{aligned} P(|T_N - \theta| \leq d) &= P(N^\gamma |T_N - \theta|/g(\xi) \leq N^\gamma d/g(\xi)) \\ &\geq P(m^\gamma |T_m - \theta|/g(\xi) \leq b_m g(U_m)/g(\xi)). \end{aligned}$$

So, the probability requirement is satisfied by designing b_m as a constant such that $P(m^\gamma |T_m - \theta|/g(U_m) \leq b_m) = 1 - \alpha$.

By the way, when b_m is designed as positive square root of $b_m^2 = E\{(m^\gamma |T_m - \theta|/g(U_m))^2\}$, we have the following about a risk of T_N :

$$\begin{aligned} E(|T_N - \theta|^2) &= E\{(N^\gamma |T_N - \theta|/g(\xi))^2 (g(\xi)/N^\gamma)^2\} \\ &\leq (d/b_m)^2 E\{(m^\gamma |T_m - \theta|/g(\xi))^2 (g(\xi)/g(U_m))^2\} = d^2. \end{aligned}$$

It should be noted that the risk is bounded above by the preassigned number d (> 0).

For example, suppose that X 's are from $N(\mu, \sigma^2)$. Choose $\theta = \mu$, $\xi = \sigma$ and for $n \geq 4$ let $T_n = \bar{X}_n$ ($= n^{-1} \sum_{i=1}^n X_i$), $U_n = S_n$ ($= (n-1)^{-1/2} (\sum_{i=1}^n (X_i - \bar{X}_n)^2)^{1/2}$). In this case, set $\gamma = 1/2$ and $g(x) = x$ (> 0). Then, assumptions (a)-(b) are satisfied. Finally, Stein's two-stage procedure yields the fixed-width confidence interval $[\bar{X}_N - d, \bar{X}_N + d]$ with confidence $1 - \alpha$ when b_m is designed as the upper $\alpha/2$ point of Student's t -distribution with $m - 1$ degrees of freedom (d.f.),

while the estimator \bar{X}_N has a risk bounded above by d^2 when b_m is designed as the standard deviation of the same t -distribution. As for another example, we may consider a case in which the domain space depends on the parameter to be estimated. Suppose that X 's are from a negative exponential distribution $e(\mu, \sigma)$ with the p.d.f.

$$f(x; \mu, \sigma) = \sigma^{-1} \exp\left(-\frac{x - \mu}{\sigma}\right) I(x > \mu),$$

where $I(\cdot)$ stands for the indicator function of (\cdot) . Choose $\theta = \mu$, $\xi = \sigma$ and for $n \geq 4$ let $T_n = \min(X_1, \dots, X_n)$, $U_n = (n-1)^{-1} \sum_{i=1}^n (X_i - T_n)$. In this case, set $\gamma = 1$ and $g(x) = x (> 0)$. Then, assumptions (a)-(b) are satisfied. Exchange d in the stopping rule (1) with $2d$ and define the fixed-width confidence interval for μ by $[T_N - 2d, T_N]$. That interval has confidence $1 - \alpha$ when b_m is designed as the upper α point of F -distribution with $(2, 2(m-1))$ d.f.s, while the risk of T_N is bounded above by $4d^2$ when b_m is designed as positive square root of the second order moment for the same F -distribution.

Several estimation problems for parameters such as the location parameter of a negative exponential distribution, the scale parameter of a Pareto distribution, and the mean of an inverse Gaussian distribution have the same nature as for the mean of a normal distribution in a sense of essence of Stein's two-stage procedure, therefore, it is possible to apply methodologies in this article to these problems. Refer to Ghurye [G58] and Mukhopadhyay [M82] for the details. In addition, estimation problem for regression parameters of a linear regression model, and ranking and selection problem of populations are possible to consider in a similar context.

Even if essence of the two-stage procedure in one-dimensional case of a population distribution remains in multi-dimensional case, however, handling of statistic appeared in multi-dimensional case becomes complicated. In case of problems processing several population simultaneously, it becomes further complicated. The efficiency of inference really depends on how you take in properties of the underlying distribution in methodologies for individual problem. In all problems, the two-stage procedure aims at guaranteeing requirement about a risk in authenticity (instead of approximation). With that in mind, the optimality to reduce the sample size required in inference is pursued. In Sections 3-4, a multivariate normal distribution is supposed as a population distribution so that we are not bothered by specific details peculiar to distribution. We shall give an individual inference problem a solution by using a two-stage procedure. There are several cases when assumptions (a)-(b) are not satisfied. In Section 5, about inference problems for such cases, we shall discuss robustness of a two-stage procedure, along with several other techniques to handle with such the cases.

3. For a multivariate normal population

Let $\{\mathbf{X}_i; i \geq 1\}$ be a sequence of i.i.d. random vectors with p -variate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} (> \mathbf{0})$. Let us consider the following two inference problems about the mean vector $\boldsymbol{\mu}$.

(P1) For prespecified constant $\varepsilon > 0$, find an estimator $\boldsymbol{\delta}_n = \boldsymbol{\delta}_n(X_1, \dots, X_n)$ for $\boldsymbol{\mu}$ such that $E(||\boldsymbol{\delta}_n - \boldsymbol{\mu}||^2) \leq \varepsilon$ for any $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Here and throughout, the norm is the Euclidian norm and $\|\mathbf{X}\|^2 = \mathbf{X}'\mathbf{X}$ where \mathbf{X}' stands for the transpose of a vector \mathbf{X} .

(P2) For value $d > 0$ of radius and confidence coefficient $1 - \alpha > 0$ both given beforehand, determine the sample size n such that the confidence region $CR(\boldsymbol{\delta}_n) = \{\boldsymbol{\mu} | ||\boldsymbol{\mu} - \boldsymbol{\delta}_n|| \leq d\}$ for $\boldsymbol{\mu}$ satisfies $P(\boldsymbol{\mu} \in CR(\boldsymbol{\delta}_n)) \geq 1 - \alpha$ for any $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Having recorded $\mathbf{X}_1, \dots, \mathbf{X}_n$ of size n , it is natural to consider the sample mean $\bar{\mathbf{X}}_n$ as an estimator of $\boldsymbol{\mu}$. When $\boldsymbol{\Sigma}$ is known, we have $E(||\bar{\mathbf{X}}_n - \boldsymbol{\mu}||^2) = tr(\boldsymbol{\Sigma})/n$ so that the sample size n to solve (P1) is determined as the smallest integer such that $n \geq tr(\boldsymbol{\Sigma})/\varepsilon$. As for (P2), let a be the upper α point of the chi-squared distribution function, $F_p(\cdot)$, with p d.f., i.e., $F_p(a) = 1 - \alpha$. Consider the confidence region $CR(\bar{\mathbf{X}}_n)$. When $\boldsymbol{\Sigma}$ is known, the sample size n to solve (P2) is determined as the smallest integer such that $n \geq a\lambda/d^2$ where λ denotes the maximum latent root of $\boldsymbol{\Sigma}$. However, if $\boldsymbol{\Sigma}$ is unknown, the sample size n should be a function of a sample through an estimate of $\boldsymbol{\Sigma}$. To handle with this process, the two-stage procedure comes into consideration. As for (P1), it is solved as follows.

[Two-stage procedure (P1)]

(T1) First, take a pilot sample $\mathbf{X}_1, \dots, \mathbf{X}_m$ of size m (≥ 4) and calculate $\mathbf{S}_m = (m-1)^{-1} \sum_{i=1}^m (\mathbf{X}_i - \bar{\mathbf{X}}_m)(\mathbf{X}_i - \bar{\mathbf{X}}_m)'$ as an estimate of $\boldsymbol{\Sigma}$. Define the stopping rule

$$N = \max \{m, [c_m tr(\mathbf{S}_m)/\varepsilon] + 1\}, \quad (2)$$

where $c_m = (m-1)/(m-3)$.

(T2) Next, take an additional sample $\mathbf{X}_{m+1}, \dots, \mathbf{X}_N$ of size $N-m$. By combining the initial sample and the additional sample, estimate $\boldsymbol{\mu}$ by $\bar{\mathbf{X}}_N = N^{-1} \sum_{i=1}^N \mathbf{X}_i$.

When $p = 1$, Birnbaum and Healy [BH60] showed that the above two-stage procedure solves problem (P1) and it was developed to the case when $p \geq 2$ by Kubokawa [K90]. (Refer to Rao

[Ra73, pp.486–487] as well.) As for a choice of the size m , Cohen and Sackrowitz [CS84ab] studied the Bayes decision rule with respect to suitable prior distributions of the parameters.

When supposed to have a known real number s_* such that $\text{tr}(\boldsymbol{\Sigma}) > s_*(> 0)$ for any $\boldsymbol{\Sigma} (> \mathbf{0})$, Aoshima [A00] showed that procedure (P1) is second-order efficient under the condition that $m = m(\varepsilon)$ and $\lim_{\varepsilon \rightarrow 0} m\varepsilon = s_*$ by developing the techniques given by Mukhopadhyay and Duggan [MD97, 99]. It means that the average sample size and the associated risk can be expanded up to the second order when $\varepsilon \rightarrow 0$. It would be interesting if a reduction of the sample size, required in (P1), is considered with the help of improvements on estimators together with the stopping rule by incorporating prior information about nuisance parameters.

For a specific case that $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p$, Ghosh and Sen [GS83] showed that the James-Stein type estimator

$$\boldsymbol{\delta}_N = \bar{\mathbf{X}}_N - \frac{(p-2)\hat{\sigma}_N^2}{N\|\bar{\mathbf{X}}_N\|^2}\bar{\mathbf{X}}_N, \quad \hat{\sigma}_N^2 = \sum_{i=1}^N \|\mathbf{X}_i - \bar{\mathbf{X}}_N\|^2 / \{p(N-1) + 2\}$$

dominates $\bar{\mathbf{X}}_N$ when $p \geq 3$. As for a purely sequential procedure, Ghosh et al. [GNS87] showed a similar risk dominance result. Natarajan and Strawderman [NS85] and Kubokawa and Saleh [KS94] studied the improvements of the stopping rule together with the estimator such that in the shrinkage procedures the sample size is exactly smaller than or equal to N and the shrinkage estimator is asymptotically better than $\bar{\mathbf{X}}_N$ when $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p$ with $p \geq 3$. As for the case of arbitrary $\boldsymbol{\Sigma}$, the James-Stein type estimator has not been developed.

When certain prior distributions are assumed for the parameters, the Bayes sequential estimation is pursued to minimize the Bayes risk over all stopping rules and over all estimators. Arrow et al. [ABG49] showed that in the Bayes sequential estimation problem, the optimal estimator for squared loss is given by the Bayes solution for any stopping rule. It can be shown that an optimal stopping rule exists, however it is given by the method of backward induction and it is often inaccessible. To overcome this difficulty, Bickel and Yahav [BY68] devised the APO (asymptotically pointwise optimal) rule which derives an explicit stopping rule whose Bayes risk typically is close to the Bayes risk of the optimal rule. When a conjugate distribution is supposed as a prior distribution, the APO rule has a risk of the same extent as the Bayes stopping rule asymptotically as the observation cost approaches zero. When $p = 1$, Woodroffe [W81] showed that certain stopping rules are asymptotically non-deficient as the observation cost approaches zero: That verification was extended to a multivariate case by Nagao [N97ab]. When the prior is not completely known but auxiliary data are available for estimating unknown parameters of the prior, Martinsek [M87] considered a general empirical Bayes approach which approximates the optimal rule. On the other hand, to overcome the

concerning difficulty, Alvo [Alv77, 78], Akahira and Koike [AkK96], Koike [Ko99] and among others considered a heuristic approach giving stopping rules for which the excess risk incurred over the optimum Bayes risk is bounded and possibly evaluated explicitly from the prior distribution. Further evolutions such as improvements of the inequality so that a certain distribution family achieving the bound exists successfully could be anticipated in this research.

As for (P2), it is solved as follows.

[Two-stage procedure (P2)]

(T1) First, take a pilot sample $\mathbf{X}_1, \dots, \mathbf{X}_m$ of size $m (> p)$ and calculate the maximum latent root, ℓ_m , of \mathbf{S}_m . Define the stopping rule

$$N = \max \left\{ m, \left[a_m \ell_m / d^2 \right] + 1 \right\}, \quad (3)$$

where $a_m = p(m-1)F_{p,m-p}(\alpha)/(m-p)$ with $F_{r,s}(\alpha)$ the upper α point of F distribution having d.f.s (r, s) .

(T2) Next, take an additional sample $\mathbf{X}_{m+1}, \dots, \mathbf{X}_N$ of size $N - m$. By combining the initial sample and the additional sample, define the confidence region $CR(\bar{\mathbf{X}}_N)$ with $\bar{\mathbf{X}}_N = N^{-1} \sum_{i=1}^N \mathbf{X}_i$.

Healy [He56] showed that the above two-stage procedure solves problem (P2). Aoshima [A00] suggested that the stopping rule (3) replacing a_m with smaller constant $a_m^* = pF_{p,m-1}(\alpha)$ still enables the procedure to solve problem (P2). When p is large and m is small, the reduction of N by this correction is remarkable. Under the assumptions that λ is simple and there exists a known and positive λ_* such that $\lambda > \lambda_*$ for any $\Sigma (> \mathbf{0})$, Mukhopadhyay [M99] and [A00] considered a sequence of $m = m(d)$ such that $\lim_{d \rightarrow 0} md^2 = a\lambda_*$ and studied the second-order efficiency of procedure (P2) as $d \rightarrow 0$. Recall that a denotes the upper α point of the chi-squared distribution with p d.f. As for a purely sequential procedure to problem (P2), see Srivastava [Sriv67], and refer to Woodroffe [W77] for its second-order efficiency when $p = 1$: For a multivariate case, its second-order efficiency has not been resolved fully but Dmitrienko and Govindarajulu [DB00] showed a risk boundedness.

It would be interesting to consider a James-Stein type estimator improving $CR(\bar{\mathbf{X}}_N)$ at least when $\Sigma = \sigma^2 \mathbf{I}_p$. For this challenging problem, Hwang and Casella [HC82, 84], Shinozaki [Sh89] and Takada [T98b] should be referred: They considered this issue in a single-stage sampling scheme. In addition, about an empirical Bayes confidence region, it has not been elucidated enough theoretically even in a single-stage sampling scheme.

Another multivariate version of Stein's [S45] two-stage procedure is available for (P2).

[Two-stage procedure (P2)']

(T1) First, take a pilot sample $\mathbf{X}_1, \dots, \mathbf{X}_m$ of size $m (> p)$ and calculate $tr(\mathbf{S}_m)$. Define the stopping rule

$$\tilde{N} = \max \left\{ m + p^2, \left[\tilde{a}_m tr(\mathbf{S}_m)/d^2 \right] + 1 \right\}, \quad (4)$$

where $\tilde{a}_m > 0$ is given as $u = \tilde{a}_m$ such that $\int_0^\infty v^{-1} f_p(v^{-1}u) g_{p,m}(v) dv = 1 - \alpha$ with $f_p(\cdot)$ p.d.f. of the chi-squared distribution with p d.f. and $g_{p,m}(\cdot)$ p.d.f. of $V = tr(\Sigma \mathbf{S}_m^{-1})$.

(T2) Next, take an additional sample $\mathbf{X}_{m+1}, \dots, \mathbf{X}_{\tilde{N}}$ of size $\tilde{N} - m$. By combining the initial sample and the additional sample, calculate a generalized sample mean vector $\widetilde{\mathbf{X}}_{\tilde{N}} = (tr(\mathbf{T}_1 \mathbf{Y}'), \dots, tr(\mathbf{T}_p \mathbf{Y}'))'$, where $\mathbf{Y} = [\mathbf{X}_1, \dots, \mathbf{X}_m, \mathbf{X}_{m+1}, \dots, \mathbf{X}_{\tilde{N}}]$ and $\mathbf{T}_i = [\mathbf{t}_{i1}, \dots, \mathbf{t}_{im}, \mathbf{t}_{im+1}, \dots, \mathbf{t}_{i\tilde{N}}]$, $1 \leq i \leq p$ are $p \times \tilde{N}$ random matrices satisfying the following three conditions:

- (i) $\mathbf{t}_{i1} = \dots = \mathbf{t}_{im}$;
- (ii) $\mathbf{T}_i \mathbf{1} = \mathbf{e}_i$, where $\mathbf{1} : \tilde{N} \times 1 = (1, \dots, 1)'$ and $\mathbf{e}_i : p \times 1 = (0, \dots, 0, 1, 0, \dots, 0)'$;
- (iii) $\mathbf{T} \mathbf{T}' = (d^2/\tilde{a}_m) \mathbf{I}_p \otimes \mathbf{S}_m^{-1}$, where $\mathbf{T} : p^2 \times \tilde{N} = [\mathbf{T}'_1, \dots, \mathbf{T}'_p]'$.

Then, define the confidence region $CR(\widetilde{\mathbf{X}}_{\tilde{N}})$ with $\widetilde{\mathbf{X}}_{\tilde{N}}$.

The conditional distribution of $\widetilde{\mathbf{X}}_{\tilde{N}}$ given \mathbf{S}_m is $N_p(\boldsymbol{\mu}, (Vd^2/\tilde{a}_m)\mathbf{I}_p)$. When $p = 1$, Stein [S45] showed that $\sqrt{\tilde{a}_m/(pd^2)}(\widetilde{\mathbf{X}}_{\tilde{N}} - \boldsymbol{\mu})$ has Student's t -distribution with $\nu = m - 1$ d.f. When $p = 2$, Chatterjee [C59] gave its distribution. When $p \geq 3$, the distribution of $\widetilde{\mathbf{X}}_{\tilde{N}}$ is complicated to handle exactly. In the case, a large sample approximation could be considered. Hyakutake and Siotani [HS87] obtained an asymptotic expansion for the distribution of $\sqrt{\tilde{a}_m/(pd^2)}(\widetilde{\mathbf{X}}_{\tilde{N}} - \boldsymbol{\mu})$ up to the order $O(\nu^{-2})$ by using the differential operator method. Then, its limiting distribution is $N_p(\mathbf{0}, \mathbf{I}_p)$.

Chatterjee [C60] showed that the above two-stage procedure also solves problem (P2) with coverage probability exactly equal to $1 - \alpha$. The stopping rule (4) is the least integer meeting the necessary and sufficient condition to choose \mathbf{T} satisfying three conditions (i)–(iii) simultaneously. (Several methods of generating \mathbf{T} were given by Hyakutake [H86] and Dudewicz and Taneja [DT87].) Chatterjee [C60] showed that procedure (P2)' is less efficient than (P2) in terms of both the sample size and the coverage probability. Furthermore, for any $(\boldsymbol{\mu}, \Sigma)$, it holds that $CR(\overline{\mathbf{X}}_N) \subset CR(\widetilde{\mathbf{X}}_{\tilde{N}})$ w.p.1 when placed at the same center. (Aoshima [A94a] gave assessments about both mean and variance of the sample size in procedure (P2) and compared them numerically with those given by Hyakutake and Siotani [HS89] in (P2)').) However, on the other hand, procedure (P2)' has an interesting

property that $U = \tilde{a}_m(pd^2)^{-1}||\widetilde{\mathbf{X}}_{\tilde{N}}||^2$ has the distribution depending only on $||\boldsymbol{\mu}||^2$ but completely free from nuisance parameters Σ . Chatterjee [C59] successfully showed that procedure (P2)' yields a test for $H_0 : \boldsymbol{\mu} = \mathbf{0}$ with power function completely free from Σ . (As for the distribution of U , an asymptotic expansion was studied by Mukaihata and Fujikoshi [MF93] together with its error bound.) Albers [Alb92] investigated to what extent this nice feature of having a power independent of certain aspects of the unknown underlying distribution can be generalized to the case of rank tests.

It has not been known whether a test given by procedure (P2) is uniformly more powerful than the test given by (P2)'. Chatterjee [C60] constructed a test which is uniformly more powerful and slightly more economical than (P2)', by replacing $m + p^2$ with $m + p^2 - 1$ in the stopping rule (4) and by replacing (d^2/\tilde{a}_m) with $tr(\mathbf{S}_m)/\tilde{N}$ in condition (iii) about \mathbf{T} . (Note that for \tilde{N} given by those replacements such a \mathbf{T} can always be found.) It is evident that procedure (P2)' with the above modification can also yield a confidence region slightly better than $CR(\widetilde{\mathbf{X}}_{\tilde{N}})$. However, as suggested in [C60], this slight improvement pays sacrifice that the associated power already depends on Σ . See also Chattejee [C91]. As for a purely sequential procedure, refer to Liu [Li97c] who studied a test problem when $p = 1$.

When some structure of Σ is supposed by prior information, several improvements of the stopping rule can be considered. For instance, let us consider the case that Σ has an intraclass correlation model such as $\Sigma = \sigma^2\{(1 - \rho)\mathbf{I}_p + \rho\mathbf{1}\mathbf{1}'\}$ where $\mathbf{1} = (1, \dots, 1)'$. Then, as for the maximum latent root of Σ , we have $\lambda = \max(\tau_1, \tau_2)$ where $\tau_1 = \sigma^2\{1 + (p - 1)\rho\}$ is simple while $\tau_2 = \sigma^2(1 - \rho)$ is $p - 1$ multiple. Note that $\hat{\tau}_{1m} = p^{-1}\mathbf{1}'\mathbf{S}_m\mathbf{1}$ and $\hat{\tau}_{2m} = (p - 1)^{-1}(tr(\mathbf{S}_m) - \hat{\tau}_{1m})$ are unbiased estimates of τ_1 and τ_2 respectively. A natural estimation of λ could be $\max(\hat{\tau}_1, \hat{\tau}_2)$. Hyakutake et al. [HTA95] considered an improvement of procedure (P2) by modifying the stopping rule (3) as

$$N = \max \left\{ m, \left[a_m \max(\hat{\tau}_{1m}, \hat{\tau}_{2m})/d^2 \right] + 1 \right\}. \quad (5)$$

Problem (P2) can be solved by choosing a_m as a constant such that $P\{U_1 + (p - 1)U_2 \leq a_m\} = 1 - \alpha$, where U_1 and U_2 are independent F -random variables with $(1, m - 1)$ d.f.s and $(p - 1, (p - 1)(m - 1))$ d.f.s, respectively. As expected, in that structure of Σ , procedure (P2) with (3) is improved by the modification with (5) in terms of the sample size. Takada and Hyakutake [TH97] considered a similar modification under a generalization of Σ that $\Sigma = \sigma_1 \mathbf{A}_1 + \dots + \sigma_q \mathbf{A}_q$, where \mathbf{A}_i is a known and symmetric matrix with $rank(\mathbf{A}_i) = r_i$ for each i and $\sum_{i=1}^q \mathbf{A}_i = \mathbf{I}_p$, $\sum_{i=1}^q r_i = p$ while σ_i 's are unknown and positive scalars. Under that general structure of Σ , Aoshima and Mukhopadhyay [AM99] gave asymptotic assessments up to the second order when $d \rightarrow 0$ about the average sample size and the coverage probability for the modified procedure (P2). As for a purely sequential procedure,

refer to Nagao [N96].

4. For several multivariate normal populations

Let us consider estimating the linear function of mean vectors coming from several multivariate normal populations. For each i ($1 \leq i \leq k$), let $\{\mathbf{X}_{ij}; j \geq 1\}$ be a sequence of i.i.d. random vectors with p -variate normal distribution $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ where $\boldsymbol{\Sigma}_i (> \mathbf{0})$. For (b_1, \dots, b_k) given along aim of inference, we define $\boldsymbol{\xi} = \sum_{i=1}^k b_i \boldsymbol{\mu}_i$. Let us consider the following two inference problems about $\boldsymbol{\xi}$.

(P3) For prespecified constant $\varepsilon > 0$, find $\mathbf{n} = (n_1, \dots, n_k)$, the sample sizes from each population, and an estimator $\boldsymbol{\delta}_n$ of $\boldsymbol{\xi}$ such that $E(\|\boldsymbol{\delta}_n - \boldsymbol{\xi}\|^2) \leq \varepsilon$ for any $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $1 \leq i \leq k$.

(P4) For value $d > 0$ of radius and confidence coefficient $1 - \alpha > 0$ both given beforehand, determine $\mathbf{n} = (n_1, \dots, n_k)$ such that the confidence region $CR(\boldsymbol{\delta}_n) = \{\boldsymbol{\xi} \mid \|\boldsymbol{\xi} - \boldsymbol{\delta}_n\| \leq d\}$ for $\boldsymbol{\xi}$ satisfies $P(\boldsymbol{\xi} \in CR(\boldsymbol{\delta}_n)) \geq 1 - \alpha$ for any $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $1 \leq i \leq k$.

Having recorded $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ of size n_i for each i ($1 \leq i \leq k$), we define the linear function of sample means by $\mathbf{T}_n = \sum_{i=1}^k b_i \bar{\mathbf{X}}_{in_i}$. As for (P3), we have that $E(\|\mathbf{T}_n - \boldsymbol{\xi}\|^2) = \sum_{i=1}^k b_i^2 \text{tr}(\boldsymbol{\Sigma}_i)/n_i$. When $\boldsymbol{\Sigma}_i$ ($1 \leq i \leq k$) is known, the sample sizes \mathbf{n} to solve (P3) would be determined as the smallest integer such that $n_i \geq |b_i| \sqrt{\text{tr}(\boldsymbol{\Sigma}_i)} \sum_{i=1}^k |b_i| \sqrt{\text{tr}(\boldsymbol{\Sigma}_i)}/\varepsilon$ for each i if one considers minimizing the total sample size $\sum_{i=1}^k n_i$. As for (P4), the sample sizes \mathbf{n} are determined as the smallest integer such that $n_i \geq a|b_i| \sqrt{\lambda_i} \sum_{i=1}^k |b_i| \sqrt{\lambda_i}/d^2$ for each i to construct $CR(\mathbf{T}_n)$. Here, a is the upper α point of the chi-squared distribution with p d.f., and λ_i denotes the maximum latent root of $\boldsymbol{\Sigma}_i$. However, if $\boldsymbol{\Sigma}_i$ ($1 \leq i \leq k$) is unknown, the sample sizes \mathbf{n} should be determined through estimates of $\boldsymbol{\Sigma}_i$, $1 \leq i \leq k$. As for (P3), it is solved as follows.

[Two-stage procedure (P3)]

(T1) First, take a pilot sample $\mathbf{X}_{i1}, \dots, \mathbf{X}_{im}$ of size m (≥ 4) and calculate $\mathbf{S}_{im} = (m-1)^{-1} \sum_{j=1}^m (\mathbf{X}_{ij} - \bar{\mathbf{X}}_{im})' (\mathbf{X}_{ij} - \bar{\mathbf{X}}_{im})$ as an estimate of $\boldsymbol{\Sigma}_i$ for each i . Define the stopping rule of each population by

$$N_i = \max \left\{ m, \left[c_m |b_i| \sqrt{\text{tr}(\mathbf{S}_{im})} \sum_{i=1}^k |b_i| \sqrt{\text{tr}(\mathbf{S}_{im})}/\varepsilon \right] + 1 \right\}, \quad (6)$$

where $c_m = (m-1)/(m-3)$ that is the same one as in (2).

(T2) Next, take an additional sample $\mathbf{X}_{im+1}, \dots, \mathbf{X}_{iN_i}$ of size $N_i - m$ for each i . Let $\mathbf{N} = (N_1, \dots, N_k)$. By combining the initial sample and the additional sample, calculate $\bar{\mathbf{X}}_{iN_i} = N_i^{-1} \sum_{j=1}^{N_i} \mathbf{X}_{ij}$ for each i . Then, estimate $\boldsymbol{\xi}$ by $\mathbf{T}_N = \sum_{i=1}^k b_i \bar{\mathbf{X}}_{iN_i}$.

Aoshima and Takada [AT02] showed that the above two-stage procedure solves problem (P3) and improves the predecessor research given in Ghosh et al. [GMS97, Chap. 6] in terms of the sample size.

As for (P4), it is solved as follows.

[Two-stage procedure (P4)]

(T1) First, take a pilot sample $\mathbf{X}_{i1}, \dots, \mathbf{X}_{im}$ of size $m (> p)$ and calculate the maximum latent root, ℓ_{im} , of \mathbf{S}_{im} for each i . Define the stopping rule of each population by

$$N_i = \max \left\{ m, \left[u_m |b_i| \sqrt{\ell_{im}} \sum_{i=1}^k |b_i| \sqrt{\ell_{im}} / d^2 \right] + 1 \right\}, \quad (7)$$

where $u_m > 0$ is determined by solving the equations

$$k F_{p,m-1} \left(\frac{u_m}{p} \right) - (k-1) F_{p,m} \left(\frac{u_m m}{p(m-1)} \right) = 1 - \alpha$$

when $p = 1, 2$, and

$$k \int_0^\infty F_p \left(\frac{u_m x}{m-1} \right) (1 - F_{m-1}(x))^{k-1} dF_{m-1}(x) = 1 - \alpha$$

when $p \geq 3$. Here, $F_{r,s}(\cdot)$ and $F_r(\cdot)$ denote the c.d.f.s of F -distribution with (r, s) d.f.s. and chi-squared distribution with r d.f., respectively.

(T2) Next, take an additional sample $\mathbf{X}_{im+1}, \dots, \mathbf{X}_{iN_i}$ of size $N_i - m$ for each i . Let $\mathbf{N} = (N_1, \dots, N_k)$. By combining the initial sample and the additional sample, calculate $\bar{\mathbf{X}}_{iN_i} = N_i^{-1} \sum_{j=1}^{N_i} \mathbf{X}_{ij}$ for each i . Then, define the confidence region $CR(\mathbf{T}_\mathbf{N})$ with $\mathbf{T}_\mathbf{N} = \sum_{i=1}^k b_i \bar{\mathbf{X}}_{iN_i}$.

Aoshima et al. [ATS02] showed that the above two-stage procedure solves problem (P4). Problem (P4) had been studied by Chapman [Ch50], Ghosh [Gh75], Mukhopadhyay and Liberman [ML89] and a lot of researchers, but an optimal solution had not been obtained even asymptotically. As for specific cases of (P4), when $p = 1$ and $k = 2$, Banerjee [B67] gave an asymptotic optimal solution and Schwabe [Sc93] gave an improvement of [B67] in terms of the sample size. Takada and Aoshima [TA96] developed Banerjee-Schwabe's result to a multivariate case but $\Sigma_i = \sigma_i^2 \mathbf{H}_i$ (\mathbf{H}_i is a known and positive definite matrix), $i = 1, 2$, and it was followed by Takada and Aoshima [TA97] for the case when $k \geq 3$. Under these references, [ATS02] reached to give an asymptotic optimal solution to problem (P4) in which Σ_i ($1 \leq i \leq k$) is completely unknown. As for second-order asymptotic properties in this context, refer to Aoshima and Mukhopadhyay

[AM02]. Under the structures for each i that $\Sigma_i = \sigma_{i1}\mathbf{A}_1 + \dots + \sigma_{iq}\mathbf{A}_q$ with known and symmetric \mathbf{A}_j ($1 \leq j \leq q$) such that $\text{rank}(\mathbf{A}_j) = r_j$, $\sum_{j=1}^q \mathbf{A}_j = \mathbf{I}_p$, $\sum_{j=1}^q r_j = p$ and with unknown and positive scalar σ_{ij} ($1 \leq j \leq q$), Aoshima [A98] and Hyakutake [H98] studied some modifications of the stopping rule (7).

When $k = 2$ and $(b_1, b_2) = (1, -1)$, problem (P4) is reduced to Behrens-Fisher problem. Then, as for testing $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, procedure (P2)' given in Section 3 could be considered naturally. Dudewicz and Bishop [DB79] proposed an application of procedure (P2)' to Behrens-Fisher problem, noting that the variance-covariance matrices of the generalized sample mean vectors given in procedure (P2)' are common and known for two populations if one conducts procedure (P2)' for each population independently but with a common size of $m (> p)$ for their pilot samples. Sample sizes for that testing are determined such that a test rule with significant level $\alpha > 0$ against $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ guarantees power β when $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\| = \delta_0$ for given $\delta_0 > 0$ and $\beta > 0$. Applications for constructing simultaneous confidence intervals are also available. See Hyakutake and Siotani [HS87], Siotani [Sio87], Siotani et al. [SHF85; Secs. 5.6.3, 6.4.3] and Aoshima [A94b] with those references.

When $p \geq 2$, we consider multiple comparisons experiments for correlated components (ξ_1, \dots, ξ_p) of $\boldsymbol{\xi} = \sum_{i=1}^k b_i \boldsymbol{\mu}_i$. When we suppose that there are several remedies to be compared with each other, say $k = 2$, and those effects are observed at p points of time series, the user would be typically interested in the direction and the magnitude of differences— which points are more significant differences, and by how much— with respect to p correlated time components (ξ_1, \dots, ξ_p) of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ ($b_1 = 1, b_2 = -1$). As for multiple comparisons methods, Tukey's (1953) method of all pairwise multiple comparisons (MCA), Hsu's (1984) method of multiple comparisons with the best (MCB), and Dunnett's (1955) method of multiple comparisons with a control (MCC) are well known and referred by many authors. (See Hochberg and Tamhane [HT87], Hirotsu [Hi92], Hsu [Hs96], and Nagata and Yoshida [NY97] as related textbooks.) Having recorded $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ from each i ($1 \leq i \leq k$), we have $\mathbf{T}_n = \sum_{i=1}^k b_i \bar{\mathbf{X}}_{in_i}$ with $\mathbf{n} = (n_1, \dots, n_k)$. Then, the components (ξ_1, \dots, ξ_p) are estimated by $\mathbf{T}_n = (T_{1n}, \dots, T_{pn})$. Tukey's MCA method, Hsu's MCB method and Dunnett's MCC method give the following simultaneous confidence intervals when d is specified (suitably narrow).

(MCA) For the $p(p - 1)/2$ differences of component effects,

$$SCI(\mathbf{T}_n) = \{\boldsymbol{\xi} \mid \xi_r - \xi_s \in [T_{rn} - T_{sn} - d, T_{rn} - T_{sn} + d], 1 \leq r < s \leq p\};$$

(MCB) For comparing each component with the best of the other components when a larger component effect is supposed to be better,

$$SCI(\mathbf{T}_n) = \{\boldsymbol{\xi} \mid \xi_r - \max_{s \neq r} \xi_s \in [-(T_{rn} - \max_{s \neq r} T_{sn} - d)^-, +(T_{rn} - \max_{s \neq r} T_{sn} + d)^+]\},$$

$$r = 1, \dots, p\},$$

where $+x^+ = \max\{0, x\}$ and $-x^- = \min\{0, x\}$;

(MCC) For comparing each component with a control component,

$$SCI(\mathbf{T}_n) = \{\boldsymbol{\xi} \mid \xi_r - \xi_p \in [T_{rn} - T_{pn} - d, T_{rn} - T_{pn} + d], r = 1, \dots, p-1\},$$

where the component p is supposed to be the control.

Then, we consider the following problem for each of (MCA), (MCB) and (MCC).

(P5) For value $d > 0$ and confidence coefficient $1 - \alpha > 0$ both given beforehand, determine $\mathbf{n} = (n_1, \dots, n_k)$ such that the simultaneous confidence intervals $SCI(\mathbf{T}_n)$ satisfies $P(\boldsymbol{\xi} \in SCI(\mathbf{T}_n)) \geq 1 - \alpha$ for any $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $1 \leq i \leq k$.

The coverage probability of simultaneous confidence intervals depends on the variance-covariance structure only through the variance of the pairwise differences of the component mean estimates. We assume that $\boldsymbol{\Sigma}_i = (\sigma_{(i)rs})$ has a spherical structure such as $\sigma_{(i)rr} + \sigma_{(i)ss} - 2\sigma_{(i)rs} = 2\tau_i^2$ ($1 \leq r < s \leq p$) depending only on the suffix i ($1 \leq i \leq k$). (See Hirotsu [Hi93, p.194] for instance.) A special case of the spherical model is the intraclass correlation model such as $\boldsymbol{\Sigma}_i = \sigma_i^2 \{(1 - \rho_i)\mathbf{I}_p + \rho_i\mathbf{1}\mathbf{1}'\}$ for some ρ_i , where $\mathbf{1} = (1, \dots, 1)'$. When $k = 1$ and the variance-covariance matrix has the spherical structure, Bhargava and Srivastava [BhS73], Chaturvedi et al. [CSS92], Hsu [Hs89], Nelson [Ne93], Nelson and Matejcik [NM95] and among others considered problem (P5) by using an appropriate procedure. However, the case when $k \geq 2$ had not been tackled fully considering interaction among populations.

When τ_i ($1 \leq i \leq k$) is known, the sample sizes \mathbf{n} to solve (P5) would be determined as the smallest integer such that $n_i \geq z^2 |b_i| \tau_i \sum_{i=1}^k |b_i| \tau_i / d^2$ for each i if one considers minimizing the total sample size $\sum_{i=1}^k n_i$. Here, for each of (MCA), (MCB) and (MCC), $z > 0$ is determined as a solution to a certain equation depending on (p, α) . (See Aoshima [A01] for the details.) When τ_i ($1 \leq i \leq k$) is unknown, the sample sizes \mathbf{n} should be determined through estimates of τ_i , $1 \leq i \leq k$. For each of (MCA), (MCB) and (MCC), problem (P5) is solved as follows.

[Two-stage procedure (P5)]

(T1) First, take a pilot sample $\mathbf{X}_{ij} = (X_{ij1}, \dots, X_{ijp})'$, $j = 1, \dots, m$, of size m (≥ 2) and calculate $S_{im}^2 = \nu^{-1} \sum_{r=1}^p \sum_{j=1}^m (X_{ijr} - \bar{X}_{ij.} - \bar{X}_{i.r} + \bar{X}_{i..})^2$ with $\nu = (p-1)(m-1)$ as an estimate of τ_i^2 for

each i . Here, $\bar{X}_{ij\cdot} = p^{-1} \sum_{r=1}^p X_{ijr}$, $\bar{X}_{i\cdot r} = m^{-1} \sum_{j=1}^m X_{ijr}$ and $\bar{X}_{i\cdot\cdot} = (pm)^{-1} \sum_{r=1}^p \sum_{j=1}^m X_{ijr}$. Define the stopping rule of each population by

$$N_i = \max \left\{ m, \left[t_m^2 |b_i| S_{im} \sum_{i=1}^k |b_i| S_{im} / d^2 \right] + 1 \right\}, \quad (8)$$

where $t_m > 0$ is determined as a solution to the following equation for each of (MCA), (MCB) and (MCC):

$$\begin{aligned} (\text{MCA}) \quad & kp \int_0^\infty \int_{-\infty}^\infty \left\{ \Phi(x) - \Phi(x - t_m \sqrt{y/\nu}) \right\}^{p-1} (1 - F_\nu(y))^{k-1} d\Phi(x) dF_\nu(y) \\ &= 1 - \alpha; \\ (\text{MCB}) \quad & k \int_0^\infty \int_{-\infty}^\infty \left\{ \Phi(x + t_m \sqrt{y/\nu}) \right\}^{p-1} (1 - F_\nu(y))^{k-1} d\Phi(x) dF_\nu(y) = 1 - \alpha; \\ (\text{MCC}) \quad & k \int_0^\infty \int_{-\infty}^\infty \left\{ \Phi(x + t_m \sqrt{y/\nu}) - \Phi(x - t_m \sqrt{y/\nu}) \right\}^{p-1} (1 - F_\nu(y))^{k-1} \\ & d\Phi(x) dF_\nu(y) = 1 - \alpha, \end{aligned}$$

where $\Phi(\cdot)$ denotes the standard normal distribution function and $F_r(\cdot)$ denotes the chi-squared distribution function with r d.f.

(T2) Next, take an additional sample $\mathbf{X}_{im+1}, \dots, \mathbf{X}_{iN_i}$ of size $N_i - m$ for each i . Let $\mathbf{N} = (N_1, \dots, N_k)$. By combining the initial sample and the additional sample, calculate $\bar{\mathbf{X}}_{iN_i} = N_i^{-1} \sum_{j=1}^{N_i} \mathbf{X}_{ij}$ for each i . Then, construct the simultaneous confidence intervals $SCI(\mathbf{T}_\mathbf{N})$ based on the components $(T_{1\mathbf{N}}, \dots, T_{p\mathbf{N}})$ of $\mathbf{T}_\mathbf{N} = \sum_{i=1}^k b_i \bar{\mathbf{X}}_{iN_i}$.

Aoshima [A01] showed that the above two-stage procedure solves problem (P5) and also discussed the case when Σ_i 's do not have any specific structure. Let t_{ma} , t_{mb} and t_{mc} denote the t_m value for each of (MCA), (MCB) and (MCC). Then, it holds that $t_{mb} < t_{mc} < t_{ma}$ and hence the sample should be required for each multiple comparisons method in such order. The reason why (MCB) requires fewest sample size is because fewest confidence intervals are required to be simultaneously correct. Note that (MCB) implies the inference of both the indifference-zone and the subset selection methodologies for ranking and selection of the best component with a specified probability of correct selection. See Hsu [Hs96] for its details. A special case of (MCB) in this context was given by Hyakutake [H00] under intraclass correlation models when $k = 2$. As for the second-order analysis related to procedure (P5), see Aoshima and Takada [AT00] and Aoshima and Miyajima [AMi01].

As for other related topics to apply two-stage procedures for, Aoshima and Mukhopadhyay [AM98] considered Scheffé-type simultaneous confidence intervals for a double linear combination

of $M = [\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k]$. Hyakutake [H92] and Aoshima, Aoki and Kai [AAK03] considered an application to select a most preferable component of Marshall-Olkin [MO67] multivariate exponential distribution in life testing. Aoshima and Chen [AC99] and Aoshima, Chen and Panchapakesan [ACP03] considered an application in voter preference problem when there is a nuisance cell in which Dirichlet integrals often used in inverse sampling technique do not work. We note that the two-stage procedure is simpler and faster than a sequential procedure in making a survey on voters in which some voters change their minds quickly after they see the news from TV and internet and it is essential to make the survey at one time point on as many voters as possible.

5. Beyond the conditions of two-stage procedure

It is natural to enquire whether departure from normality has any adverse effect on the performance of the two-stage procedure. However, robustness feature about the two-stage procedure had not been studied fully so far. When the conditions (a)-(b), given in Section 2, are not satisfied, the distribution of statistic is quite complicated and even an asymptotic expansion of the distribution is difficult to derive. Some robustness studies had been made by simulation for various types of departure from normality (see Ramkaran [Ra83] and its references) or had been done on some certain specific models (see Blumenthal and Govindarajulu [BG77] and Aoshima and Kano [AK97] for instance). To start this section, we shall theoretically attempt to expose the robustness of the two-stage procedure along the lines of Aoshima and Wakaki [AW01].

Let $\mathbf{X}, \{\mathbf{X}_i; i \geq 1\}$ be a sequence of i.i.d. random vectors with values in R^p . Let $\boldsymbol{\mu} = E(\mathbf{X})$, $\boldsymbol{\Sigma} = Cov(\mathbf{X}) (> \mathbf{0})$ and $\lambda \geq \lambda_2 \geq \dots \geq \lambda_p$ be the latent roots of $\boldsymbol{\Sigma}$. Let X_j denote the j -th element of \mathbf{X} and $\mu_{i_1 \dots i_r}$ be the moment of \mathbf{X} defined by $\mu_{i_1 \dots i_r} = E(X_{i_1} \cdots X_{i_r})$. Similarly the corresponding cumulant of \mathbf{X} is denoted by $\kappa_{i_1 \dots i_r}$. Let K_r be the r -th order tensor whose (i_1, \dots, i_r) -element is $\kappa_{i_1 \dots i_r}$. Having recorded $\mathbf{X}_1, \dots, \mathbf{X}_n$ of size n , $\boldsymbol{\mu}$ is estimated by the sample mean $\bar{\mathbf{X}}_n$. Let us consider the following inference problem about $\boldsymbol{\mu}$.

(P6) For value $d > 0$ and confidence coefficient $1 - \alpha > 0$ both given beforehand, determine the sample size n such that the ellipsoidal confidence region $ECR(\bar{\mathbf{X}}_n) = \{\boldsymbol{\mu} | n(\boldsymbol{\mu} - \bar{\mathbf{X}}_n)' \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{X}}_n) \leq c\}$ for some $c > 0$ satisfies $P(\boldsymbol{\mu} \in ECR(\bar{\mathbf{X}}_n)) \geq 1 - \alpha$ for any $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the maximum diameter $\leq 2d$.

When the population distribution is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we may use procedure (P2) given in Section 3 and define the region by $ECR(\bar{\mathbf{X}}_N) = \{\boldsymbol{\mu} | N(\boldsymbol{\mu} - \bar{\mathbf{X}}_N)' \mathbf{S}_m^{-1} (\boldsymbol{\mu} - \bar{\mathbf{X}}_N) \leq a_m\}$ where \mathbf{S}_m , a_m and N are

given in (T1) of (P2) and $\bar{\mathbf{X}}_N$ is given in (T2) of (P2). Then, Healy [He56] showed that procedure (P2) gives a solution to problem (P6) such as $P(\boldsymbol{\mu} \in ECR(\bar{\mathbf{X}}_N)) = 1 - \alpha$ for any $(\boldsymbol{\mu}, \Sigma)$. It is evident that the maximum diameter of $ECR(\bar{\mathbf{X}}_N) \leq 2d$ in view of the stopping rule (3). Note that the distribution of $T_N^2 = N(\bar{\mathbf{X}}_N - \boldsymbol{\mu})' \mathbf{S}_m^{-1} (\bar{\mathbf{X}}_N - \boldsymbol{\mu})$ is coincident with Hotelling's T^2 distribution and hence the constant c is given by a_m which is the upper α point of Hotelling's T^2 distribution.

It is interesting investigating into how robust the above solution is against departure from normality, that is whether the region $ECR(\bar{\mathbf{X}}_N)$ given by a_m guarantees its confidence coefficient the required $1 - \alpha$. It is difficult giving the distribution of T_N^2 without the assumption of normality in the form that is easy to handle theoretically. We note that the distribution of Hotelling's T^2 statistic under nonnormality is quite difficult to derive even in a single-stage sampling scheme and in which its asymptotic expansion form was given by Kano [Ka95] and Fujikoshi [Fu97]. As for the two-stage sampling scheme, a different approach from those former studies is required for derivation of an asymptotic expansion of the distribution. Aoshima and Wakaki [AW01] gave an asymptotic expansion for the distribution of T_N^2 under nonnormality when $m \rightarrow \infty$ and investigated into the robustness of the solution given by procedure (P2). We consider the following assumptions:

(A0) The Cramér's condition for the joint distribution of $(\mathbf{X}, \mathbf{XX}')$ holds;

(A1) The maximum latent root, λ , of Σ is simple;

(A2) $E(\|\mathbf{X}\|^{8+r}) < \infty$, $r > 0$;

(A3) $\lim_{m \rightarrow \infty} md^2 = c_0$ for some constant $c_0 \in (0, a\lambda)$, where a denotes the upper α point of the chi-squared distribution with p d.f.

Then, the distribution of T_N^2 is given in an asymptotic expansion form as follows:

$$P(T_N^2 \leq x) = F_p(x) + m^{-1} \sum_{j=0}^3 \beta_j F_{p+2j}(x) + o(m^{-1}),$$

where $F_r(\cdot)$ denotes a chi-squared distribution function with r d.f. Here, the coefficients $(\beta_0, \dots, \beta_3)$ are determined depending on p , $\rho = (md^2)^{-1}a\lambda$ and non-normality parameters such as $\kappa_3^{<1>} = \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p \kappa_{ijk}^2$, $\kappa_3^{<2>} = \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p \kappa_{iij} \kappa_{jkk}$, $\eta_{31}^{<2>} = \sum_{j=1}^p \kappa_{11j} \kappa_{j11}$, $\eta_{32}^{<2>} = \sum_{j=1}^p \sum_{k=1}^p \kappa_{11j}$, κ_{jkk} , $\xi_3^{<2>} = \sum_{i=1}^p \sum_{j=2}^p \lambda_j (\lambda - \lambda_j)^{-1} \kappa_{1ij}^2$, $\kappa_4^{<1>} = \sum_{i=1}^p \sum_{j=1}^p \kappa_{iijj}$ and $\eta_{41}^{<1>} = \sum_{j=1}^p \kappa_{11jj}$. (See Aoshima and Wakaki [AW01] for the details.) When the population distribution is symmetric, $\beta_3 = 0$. When the population distribution is a normal distribution, the above formula is coincident with an asymptotic expansion of Hotelling's T^2 distribution. The average of T_N^2 depends on both the third and fourth-order cumulants in the order of m^{-1} , while in a single-stage sampling scheme it depends only on the third-order cumulant. This reflects on various simulation studies conducted by

Blumenthal and Govindarajulu [BG77], Ramkaran [Ra83] and among others when $p = 1$. Putting $x = a_m$ in the above formula, one can asymptotically evaluate the coverage probability of the region $ECR(\bar{\mathbf{X}}_N)$ given by procedure (P2). We observe that if the population distribution is symmetric and longer-tailed when compared to the normal distribution, the coverage probability exceeds $1 - \alpha$ when $\rho < \rho^*$ for some ρ^* ($\in (1, 2)$) given depending on p . Further, if one applies Cornish-Fisher expansion to the distribution such as $P(T_N^2 \leq a + bm^{-1}) = 1 - \alpha + o(m^{-1})$, we could obtain the constant $a + bm^{-1}$ as an option to modify procedure (P2) with a_m against nonnormality. After taking this modification, we observe in the situation described above that the average sample size required in procedure (P2) becomes less than the one required in constructing the region $ECR(\bar{\mathbf{X}}_n) = \{\boldsymbol{\mu} | n(\boldsymbol{\mu} - \bar{\mathbf{X}}_n)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{X}}_n) \leq a\}$ with confidence $1 - \alpha + o(m^{-1})$ for known $\boldsymbol{\Sigma}$. It would be interesting considering an improvement of the estimate with higher order moments as seen in Uno and Isogai [UI00]. As for an improvement of the stopping rule, a monotone Bartlett-type correction for T^2 statistic might be considered under nonnormality by extending the techniques of Fujikoshi [Fu00] and among others.

Let us consider the estimation of parameters other than the mean. For instance, when considering a bounded risk problem about the variance of a normal distribution, the conditions required for the two-stage procedure are not satisfied about independence of estimates. Birnbaum and Healy [BH60] proposed a different two-stage procedure in which the ultimate estimator is defined by using only an additional sample. This procedure yields a bounded risk point estimation not only for the variance of a normal distribution but also for the parameters of a poisson distribution, a binomial distribution and a hypergeometric distribution and also for a scale parameter of the location-scale family of distributions: If followed by the use of Tchebychev's inequality, a fixed-width confidence interval of given confidence is naturally produced. Blum and Rosenblatt [BR69b] applied a similar technique to yielding a fixed-width confidence interval of given confidence for the moments of a distribution with increasing failure rate. However, the procedure given by [BH60] causes inefficiency necessarily, so it had been required to develop techniques so as to overcome this inconvenience. As for estimation for the variance of a normal distribution and for the parameter of a uniform distribution $U(0, \theta)$, Graybill and Connell [GC64ab] gave a technique to reduce the sample size by using an inherent inequality to those distributions instead of using Tchebychev's inequality. (That technique was applied by Takada [T86], Aoshima and Govindarajulu [AG02] and among others.) As for point estimation with bounded risk for a parameter of the scale family of distributions, Kubokawa [K89] proposed a technique to improve the ultimate estimator by combining with an initial sample. Further, Kubokawa [K90] applied this technique to point estimation with bounded risk for the generalized variance $|\boldsymbol{\Sigma}|$ of

$N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In addition, especially about fixed-width confidence intervals for the variance, Sproule [Sp74] applied an appropriate large sample theory of U -statistics to the sample variance in a purely sequential sampling scheme. See Ghosh et al. [GMS97] about related references.

In this article, several topics are omitted for brevity about especially other loss functions such as based on an asymmetric loss or based on a squared loss modified at the boundary of the parameter space. The loss function used in the sequential analysis does not have invariant property about scale transformations as seen in this article. This is because the sample size is determined by picking up information about the population distribution through estimation of the scale parameter. There is very little work on estimation for non-linear functions of means using two-stage procedures. Recently, Zheng et al. [ZSS98ab] considered a two-stage procedure for the estimation problem about product of means which arises in situations of determining area based on measurements of length and width in environmental applications. There are other problems such as sequential time series (Sriram [Srir87,01], Lee and Sriram [LS99], Galtchouk and Konev [GK01], Shiohama and Taniguchi [ST01] and among others), sequential change-point detection (Lai [La95], Siegmund and Venkatraman [SV95], Yakir [Y98] and among others) and sequential density estimation (Isogai [I93, 99], Martinsek [Ma92, 93], Xu and Martinsek [XM95], Honda [Hon98] and among others), when considering the precision of inference beforehand. However, when judging from character of the research fields, these problems should be handled with sequential sampling schemes rather than two-stage sampling schemes. Lastly, it would be interesting if a necessary and sufficient number of stages for sampling could be determined in some sense, and it might be applied to computer simulations and engineering in near future.

References

- [AkK96] Akahira, M. and Koike, K., On some properties of statistical sequential decision rules (in Japanese), *Sugaku* **48** (1996), 184–195.
- [Alb92] Albers, W., Asymptotic expansions for two-stage rank tests, *Ann. Inst. Statist. Math.*, **44** (1992), 335–356.
- [Alv77] Alvo, M., Bayesian sequential estimation, *Ann. Statist.*, **5** (1977), 955–968.
- [Alv78] Alvo, M., Sequential estimation of a truncation parameter, *J. Amer. Statist. Assoc.*, **73** (1978), 404–407.
- [A94a] Aoshima, M., Healy’s sample size of two-stage procedure in heteroscedastic simultaneous inference, *Commun. Statist.-Theory Meth.*, **23** (1994), 1297–1310.
- [A94b] Aoshima, M., In heteroscedastic simultaneous inference, the Heteroscedastic Method II is more economical than the original HM, *Amer. J. Math. Manage. Sci.*, **14** (1994), 167–196.
- [A98] Aoshima, M., A two-stage procedure for fixed-size confidence region when covariance matrices have some structures, *J. Japan Statist. Soc.*, **28** (1998), 59–67.
- [A00] Aoshima, M., Second-order properties of improved two-stage procedure for a multivariate normal distribution, *Commun. Statist.-Theory Meth.*, **29** (2000), 611–622.
- [A01] Aoshima, M., Sample size determination for multiple comparisons with components of a linear function of mean vectors, *Statistics: Reflections on the Past and Visions for the Future: An International Conference in Honor of Prof. C.R. Rao on the Occasion of His 80th Birthday*, R. Khattree and S.D. Peddada eds, *Commun. Statist.-Theory Meth.*, **30** (2001), 1773–1788.
- [AAK03] Aoshima, M., Aoki, M. and Kai, M., Two-stage procedures for selecting the best component of a multivariate exponential distribution, *Applications of Sequential Methodologies*, N. Mukhopadhyay, S. Datta and S. Chattopadhyay eds, Marcel Dekker, Inc., New York, in press (2003).
- [AC99] Aoshima, M. and Chen, P., A two-stage procedure for selecting the largest multinomial cell probability when nuisance cell is present, *Sequential Anal.*, **18** (1999), 143–155.
- [ACP03] Aoshima, M., Chen, P. and Panchapakesan, S., Sequential procedures for selecting the most probable multinomial cell when a nuisance cell is present, *Commun. Statist.-Theory Meth.*, **32** (2003), 893–905.
- [AG02] Aoshima, M. and Govindarajulu, Z., Fixed-width confidence interval for a lognormal mean, *International J. Mathematics and Mathematical Sciences*, **29** (2002), 143–154.
- [AK97] Aoshima, M. and Kano, Y., A note on robustness of two-stage procedure for a multivariate compounded normal distribution, *Sequential Anal.*, **16** (1997), 175–187.
- [AMi01] Aoshima, M. and Miyajima, H., Sample size reduction of two-stage procedure, *Kokyuroku 1224*, Research Institute for Mathematical Sciences (2001), 61–72.

- [AM98] Aoshima, M. and Mukhopadhyay, N., Fixed-width simultaneous confidence intervals for multinormal means in several intraclass correlation models, *J. Multiv. Anal.*, **66** (1998), 46–63.
- [AM99] Aoshima, M. and Mukhopadhyay, N., Second-order properties of a two-stage fixed-size confidence region when the covariance matrix has a structure, *Statistical Inference and Data Analysis—A Meeting in Honour of Prof. N. Sugiura*, *Commun. Statist.-Theory Meth.*, **28** (1999), 839–855.
- [AM02] Aoshima, M. and Mukhopadhyay, N., Two-stage estimation of a linear function of normal means with second-order approximations, *Sequential Anal.*, **21** (2002), 109–144.
- [AT00] Aoshima, M. and Takada, Y., Second-order properties of a two-stage procedure for comparing several treatments with a control, *J. Japan Statist. Soc.*, **30** (2000), 27–41.
- [AT02] Aoshima, M. and Takada, Y., Bounded risk point estimation of a linear function of k multinormal mean vectors when covariance matrices are unknown, *Advances on Theoretical and Methodological Aspects of Probability and Statistics*, N. Balakrishnan ed., Taylor & Francis, New York, (2002), 279–287.
- [ATS02] Aoshima, M., Takada, Y. and Srivastava, M.S., A two-stage procedure for estimating a linear function of k multinormal mean vectors when covariance matrices are unknown, *J. Statist. Plan. Inference*, **100** (2002), 109–119.
- [AW01] Aoshima, M. and Wakaki, H., Two-stage procedure under nonnormality, *Bull. International Statist. Inst.* 53rd Session, **1** (2001), 99–100.
- [ABG49] Arrow, K.J., Blackwell, D. and Girshick, M.A., Bayes and minimax solutions of sequential decision problems, *Econometrica*, **17** (1949), 213–244.
- [BaS56] Bahadur, R.R. and Savage, L.J., The nonexistence of certain statistical procedures in nonparametric problems, *Ann. Math. Statist.*, **27** (1956), 1115–1122.
- [B67] Banerjee, S., Confidence interval of preassigned length for the Behrens–Fisher problem, *Ann. Math. Statist.*, **38** (1967), 1175–1179.
- [BhS73] Bhargava, R.P. and Srivastava, M.S., On Tukey’s confidence intervals for the contrasts in the means of the intraclass correlation model, *J. Roy. Statist. Soc.*, **B 35** (1973), 147–152.
- [BY68] Bickel, P.J. and Yahav, J.A., Asymptotically optimal Bayes and minimax procedures in sequential estimation, *Ann. Math. Statist.*, **39** (1968), 442–456.
- [BH60] Birnbaum, A., and Healy, W.C., Jr., Estimates with prescribed variance based on two-stage sampling, *Ann. Math. Statist.*, **31** (1960), 662–676.
- [BR63] Blum, J.R. and Rosenblatt, J., On multistage estimation, *Ann. Math. Statist.*, **34** (1963), 1452–1458.
- [BR66] Blum, J.R. and Rosenblatt, J., On some statistical problems requiring purely sequential sampling schemes, *Ann. Inst. Statist. Math.*, **18** (1966), 351–355.
- [BR69a] Blum, J.R. and Rosenblatt, J., On fixed precision estimation in time series, *Ann. Math. Statist.*,

40 (1969), 1021–1032.

[BR69b] Blum, J.R. and Rosenblatt, J., Fixed precision estimation in the class of IFR distributions, *Ann. Inst. Statist. Math.*, **40** (1969), 211–213.

[BG77] Blumenthal, S. and Govindarajulu, Z, Robustness of Stein’s two-stage procedure for mixtures of normal populations, *J. Amer. Statist. Assoc.*, **72** (1977), 192–196.

[Cha50] Chapman, D.G., Some two sample tests, *Ann. Math. Statist.*, **21** (1950), 601–606.

[C59] Chatterjee, S.K., On an extension of Stein’s two-sample procedure to the multinormal problem, *Culcutta Statist. Assoc. Bull.*, **8** (1959), 121–148.

[C60] Chatterjee, S.K., Some further results on the multinormal extension of Stein’s two-sample procedure, *Culcutta Statist. Assoc. Bull.*, **9** (1960), 20–28.

[C91] Chatterjee, S.K., Two-stage and multistage procedures, *Handbook of Sequential Analysis*, B.K. Ghosh and P.K. Sen eds, Marcel Dekker, Inc., New York, (1991), 21–45.

[CSS92] Chaturvedi, A., Shukla, N.D. and Shukla, P.S., Fixed-width confidence intervals for contrasts in the means, *Ann. Inst. Statist. Math.*, **44** (1992), 157–167.

[CR65] Chow, Y.S. and Robbins, H., On the asymptotic theory of fixed-width sequential confidence intervals for the mean, *Ann. Math. Statist.*, **36** (1965), 457–462.

[CS84a] Cohen, A. and Sackrowitz, H.B., Bayes double sample estimation procedures, *Ann. Statist.*, **12** (1984), 1035–1049.

[CS84b] Cohen, A. and Sackrowitz, H.B., Results on double sample estimation for the binomial distribution, *Ann. Statist.*, **12** (1984), 1109–1116.

[D40] Dantzig, G.B., On the non-existence of tests of “Student’s” hypothesis having power functions independent of σ , *Ann. Math. Statist.*, **11** (1940), 186–192.

[DG00] Dmitrienko, A. and Govindarajulu, Z., Sequential confidence regions for maximum likelihood estimates, *Ann. Statist.*, **28** (2000), 1472–1501.

[DB79] Dudewicz, E.J. and Bishop, T.A., The heteroscedastic method, *Optim. Meth. Statist.*, J.S. Rustagi ed, Academic Press, New York, (1979), 183–203.

[DT87] Dudewicz, E.J. and Taneja, V.S., The Heteroscedastic Method: multivariate implementation, *Ann. Inst. Statist. Math.*, **39** (1987), 177–190.

[F66] Farrell, R.H., Bounded length confidence intervals for the p -point of a distribution function, III, *Ann. Math. Statist.*, **37** (1966), 586–592.

[Fu97] Fujikoshi, Y., An asymptotic expansion for the distribution of Hotelling’s T^2 -statistic under nonnormality, *J. Multiv. Anal.*, **61** (1997), 187–193.

[Fu00] Fujikoshi, Y., Transformations with improved chi-squared approximations, *J. Multiv. Anal.*, **72**

(2000), 249–263.

- [GK01] Galtchouk, L. and Konev, V., On sequential estimation of parameters in semimartingale regression models with continuous time parameter, *Ann. Statist.*, **29** (2001), 1508–1536.
- [Gh75] Ghosh, B.K., A two-stage procedure for the Behrens-Fisher problem, *J. Amer. Statist. Assoc.*, **70** (1975), 457–462.
- [GMS97] Ghosh, M., Mukhopadhyay, N. and Sen, P.K., *Sequential Estimation*, New York, Wiley, 1997.
- [GNS87] Ghosh, M., Nickerson, D.M. and Sen, P.K., Sequential shrinkage estimation, *Ann. Statist.*, **15** (1987), 817–829.
- [GS83] Ghosh, M. and Sen, P.K., On two-stage James–Stein estimators, *Sequential Anal.* **2** (1983), 359–367.
- [G58] Ghurye, S.G., Note on sufficient statistics and two-stage procedures, *Ann. Math. Statist.*, **29** (1958), 155–166.
- [GC64a] Graybill, F.A. and Connell, T.L., Sample size required for estimating the variance within d units of the true value, *Ann. Math. Statist.*, **35** (1964), 438–440.
- [GC64b] Graybill, F.A. and Connell, T.L., Sample size required to estimate the parameter in the uniform density within d units of the true value, *J. Amer. Statist. Assoc.*, **59** (1964), 550–556.
- [Ha81] Hall, P., Asymptotic theory of triple sampling for sequential estimation of a mean, *Ann. Statist.*, **9** (1981), 1229–1238.
- [Ha83] Hall, P., Sequential estimation saving sampling operations, *J. Roy. Statist. Soc., B* **45** (1983), 219–223.
- [He56] Healy, W.C., Jr., Two-sample procedures in simultaneous estimation, *Ann. Math. Statist.*, **27** (1956), 687–702.
- [Hi92] Hirotsu, C., *Analysis for Experimental Data: Beyond Analysis of Variance* (in Japanese), Kyoritsu, 1992.
- [Hi93] Hirotsu, C., Beyond analysis of variance techniques: Some applications in clinical trials, *Int. Statist. Rev.*, **61** (1993), 183–201.
- [HeS83] Hewett, J.E. and Spurrier, J.D., A survey of two stage tests of hypotheses: Theory and applications, *Commun. Statist.-Theory Meth.*, **12** (1983), 2307–2425.
- [HT87] Hochberg Y. and Tamhane, A.C., *Multiple Comparison Procedures*, John Wiley and Sons, Inc., New York, 1987.
- [Hol95] Holm, S., Confidence sets of fixed size with predetermined confidence level, *Commun. Statist. - Theory Meth.*, **24** (1995), 1521–1536.
- [Hol99] Holm, S., Confidence sets of fixed size for multinormal means, *J. Statist. Plan. Inference*, **82** (1999), 229–236.

- [Hon98] Honda, T., Sequential estimation of the marginal density function for a strongly mixing process, *Sequential Anal.*, **17** (1998), 239–251.
- [Hs89] Hsu, J.C., Sample size computation for designing multiple comparison experiments, *Comp. Statist. & Data Anal.*, **7** (1989), 79–91.
- [Hs96] Hsu, J.C., *Multiple Comparisons: Theory and methods*, Chapman & Hall, Inc., New York, 1996.
- [HC82] Hwang, J.T. and Casella, G., Minimax confidence sets for the mean of a multivariate normal distribution, *Ann. Statist.*, **10** (1982), 868–881.
- [HC84] Hwang, J.T. and Casella, G., Improved set estimators for a multivariate normal mean, *Statist. Decisions Suppl.* **1** (1984), 3–16.
- [H86] Hyakutake, H., A construction method of certain matrices required in the multivariate heteroscedastic method, *Ann. Inst. Statist. Math.*, **38** (1986), 523–528.
- [H92] Hyakutake, H., Selecting the better component of a bivariate exponential distribution, *Statist. Decisions*, **10** (1992), 153–162.
- [H98] Hyakutake, H., Fixed-size confidence region of the difference of two multinormal means when the covariance matrices have structure, *J. Japan Statist. Soc.*, **28** (1998), 175–180.
- [H00] Hyakutake, H., Selecting the component with maximal difference of two multinormal means, *Amer. J. Math. Manage. Sci.*, **20** (2000), 433–442.
- [HS87] Hyakutake, H. and Siotani, M., The multivariate heteroscedastic method: Distributions of statistics and an application, *Amer. J. Math. Manage. Sci.*, **7** (1987), 89–111.
- [HS89] Hyakutake, H. and Siotani, M., Mean and variance of sample size in multivariate heteroscedastic method, *Contributions to Probability and Statistics*, Springer-Verlag, New York, 1989, 227–240.
- [HTA95] Hyakutake, H., Takada, Y. and Aoshima, M., Fixed-size confidence regions for the multinormal mean in an intraclass correlation model, *Amer. J. Math. Manage. Sci.*, **15** (1995), 291–308.
- [I93] Isogai, E., Recursive estimation of non-parametric probability density function (in Japanese), *Sugaku* **45** (1993), 27–41.
- [I99] Isogai, E., Sequential estimation of a density and its derivatives with bounded MISE, *J. Statist. Plan. Inference*, **79** (1999), 59–78.
- [Ka95] Kano, Y., An asymptotic expansion of the distribution of Hotelling's T^2 -statistic under general condition, *Amer. J. Math. Manage. Sci.*, **15** (1995), 317–341.
- [Ko99] Koike, K., A lower bound for the Bayes risk in the sequential case, *Commun. Statist.-Theory Meth.*, **28** (1999), 857–871.
- [KOR64] Koopmans, L.H., Owen, D.B. and Rosenblatt, J.I., Confidence intervals for the coefficient of variation for the normal and log normal distributions, *Biometrika*, **51** (1964), 25–32.

- [K89] Kubokawa, T., Improving on two-stage estimators for scale families, *Metrika*, **36** (1989), 7–13.
- [K90] Kubokawa, T., Two-stage estimators with bounded risk in a growth curve model, *J. Japan Statist. Soc.*, **20** (1990), 77–87.
- [KS94] Kubokawa, T. and Saleh, A.K.Md.E., Two-stage point estimation with a shrinkage stopping rule, *Metrika*, **41** (1994), 293–306.
- [La95] Lai, T.L., Sequential change-point detection in quality control and dynamical systems, *J. Roy. Statist. Soc.*, **B 57** (1995), 613–658.
- [LS99] Lee, S. and Sriram, T.N., Sequential point estimation of parameters in a threshold AR(1) model, *Stochastic Process. Appl.*, **84** (1999), 343–355.
- [L51] Lehmann, E.L., *Notes on the Theory of Estimation*, Univ. of California Press, Berkeley, 1951.
- [Li97a] Liu, W., Improving the fully sequential sampling scheme of Anscombe-Chow-Robbins, *Ann. Statist.*, **25** (1997), 2164–2171.
- [Li97b] Liu, W., A k -stage sequential sampling procedure for estimation of normal mean, *J. Statist. Plan. Inference*, **65** (1997), 109–127.
- [Li97c] Liu, W., On some sample size formulae for controlling both size and power in clinical trials, *The Statistician*, **46** (1997), 239–251.
- [MO67] Marshall, A.W. and Olkin, I., A multivariate exponential distribution, *J. Amer. Statist. Assoc.*, **62** (1967), 30–44.
- [Ma87] Martinsek, A.T., Empirical Bayes methods in sequential estimation, *Sequential Anal.*, **6** (1987), 119–137.
- [Ma92] Martinsek, A.T., Using stopping rules to bound the mean integrated squared error in density estimation, *Ann. Statist.*, **20** (1992), 797–806.
- [Ma93] Martinsek, A.T., Fixed width confidence bands for density functions, *Sequential Anal.*, **12** (1993), 169–177.
- [MF93] Mukaihata, S. and Fujikoshi, Y., Error bounds for asymptotic expansions of some distributions in a multivariate two-stage procedure, *Hiroshima Math. J.*, **23** (1993), 565–575.
- [M82] Mukhopadhyay, N., Stein's two-stage procedure and exact consistency, *Scand. Actuarial J.*, (1982), 110–122.
- [M99] Mukhopadhyay, N., Second-order properties of a two-stage fixed-size confidence region for the mean vector of a multivariate normal distribution, *J. Multiv. Anal.*, **68** (1999), 250–263.
- [MD97] Mukhopadhyay, N. and Duggan, W.T., Can a two-stage procedure enjoy second-order properties ?, *Sankhya, Series A*, **59** (1997), 435–448.
- [MD99] Mukhopadhyay, N. and Duggan, W.T., On a two-stage procedure having second-order properties

- with applications, Ann. Inst. Statist. Math., **51** (1999), 621–636.
- [ML89] Mukhopadhyay, N. and Liberman, S., Sequential estimation of a linear function of mean vectors, Sequential Anal., **8** (1989), 381–395.
- [N96] Nagao, H., On fixed width confidence regions for multivariate normal mean when the covariance matrix has some structure, Sequential Anal., **15** (1996), 37–46.
- [N97a] Nagao, H., Asymptotically pointwise optimal rules for estimating the mean in general exponential distributions for squared loss, Sequential Anal., **16** (1997), 155–174.
- [N97b] Nagao, H., Asymptotically pointwise optimal rules of sequential estimation of mean vector when an information matrix has some structure in a multivariate normal population, Sequential Anal., **16** (1997), 363–374.
- [NY97] Nagata, Y. and Yoshida, M., Fundamentals of Statistical Multiple Comparisons (in Japanese), Scientist, 1997.
- [NS85] Natarajan, J. and Strawderman, W.E., Two-stage sequential estimation of a multivariate normal mean under quadratic loss, Ann. Statist., **13** (1985), 1509–1522.
- [Ne93] Nelson, B.L., Robust multiple comparisons under common random numbers, ACM Trans. Model. Comp. Simu., **3** (1993), 225–243.
- [NM95] Nelson, B.L. and Matejcik, F.J., Using common random numbers for indifference-zone selection and multiple comparisons in simulation, Manage. Sci., **41** (1995), 1935–1945.
- [R83] Ramkaran, The robustness of Stein’s two-stage procedure, Ann. Statist., **11** (1983), 1251–1256.
- [R73] Rao, C.R., Linear Statistical Inference and Its Application, 2nd ed., Wiley, New York, 1973.
- [Sc93] Schwabe, R., Some two-stage procedures for treating the Behrens–Fisher problem, Model-Oriented Data Analysis, W.G. Müller, H. P. Wynn and A. A. Zhigljavsky eds, Physica, Heidelberg (1993), 81–89.
- [Sh89] Shinozaki, N., Improved confidence sets for the mean of a multivariate normal distribution, Ann. Inst. Statist. Math., **41** (1989), 331–346.
- [ST01] Shiohama, T. and Taniguchi, M., Sequential estimation for a functional of the spectral density of a Gaussian stationary process, Ann. Inst. Statist. Math., **53** (2001), 142–158.
- [SV95] Siegmund, D. and Venkatraman, E.S., Using the generalized likelihood ratio statistics for sequential detection of a change-point, Ann. Statist., **23** (1995), 255–271.
- [Sin63] Singh, R., Existence of bounded length confidence intervals, Ann. Math. Statist., **34** (1963), 1474–1485.
- [Sio87] Siotani, M., Multivariate Behrens–Fisher problem by Heteroscedastic method, Advances in Multivariate Statistical Analysis, A.K. Gupta ed., (1987), 327–340.
- [SHF85] Siotani, M., Hayakawa, T. and Fujikoshi, Y., Modern Multivariate Statistical Analysis: A Graduate

Course and Handbook, American Sciences Press, Inc., Columbus, Ohio, 1985.

- [Sp74] Sproule, R.N., Asymptotic properties of U -statistics, *Trans. Amer. Math. Soc.*, **199** (1974), 55–64.
- [Srir87] Sriram, T.N., Sequential estimation of the mean of a first-order stationary autoregressive process, *Ann. Statist.*, **15** (1987), 1079–1090.
- [Srir01] Sriram, T.N., Fixed size confidence regions for parameters of threshold AR(1) models, *J. Statist. Plan. Inference*, **97** (2001), 293–304.
- [Sriv67] Srivastava, M.S., On fixed-width confidence bounds for regression parameters and the mean vector, *J. Roy. Statist. Soc., B* **29** (1967), 132–140.
- [S45] Stein, C., A two-sample test for a linear hypothesis whose power is independent of the variance, *Ann. Math. Statist.*, **16** (1945), 243–258.
- [T86] Takada, Y., Non-existence of fixed sample size procedures for scale families, *Sequential Anal.*, **5** (1986), 93–101.
- [T98a] Takada, Y., The nonexistence of procedures with bounded performance characteristics in certain parametric inference problems, *Ann. Inst. Statist. Math.*, **50** (1998), 325–335.
- [T98b] Takada, Y., Asymptotic improvement of the usual confidence set in a multivariate normal distribution with unknown variance, *J. Multiv. Anal.*, **64** (1998), 118–130.
- [TA96] Takada, Y. and Aoshima, M., Two-stage procedures for the difference of two multinormal means with covariance matrices different only by unknown scalar multipliers, *Commun. Statist.-Theory Meth.*, **25** (1996), 2371–2379.
- [TA97] Takada, Y. and Aoshima, M., Two-stage procedures for estimating a linear function of multinormal mean vectors, *Sequential Anal.*, **16** (1997), 353–362.
- [TH97] Takada, Y. and Hyakutake, H., Fixed-size confidence region by multivariate two-stage procedure when the covariance matrix has a structure, *J. Japan Statist. Soc.*, **27** (1997), 37–44.
- [UI00] Uno, C. and Isogai, E., Sequential estimation of a linear combination of means, *Statist. Probab. Lett.*, **47** (2000), 45–52.
- [W77] Woodroofe, M., Second-order approximations for sequential point and interval estimation, *Ann. Statist.*, **5** (1977), 984–995.
- [W81] Woodroofe, M., A.P.O. rules are asymptotically non deficient for estimation with squared error loss, *Z. Wahrschein. Verw. Geb.*, **58** (1981), 331–341.
- [XM95] Xu, Y. and Martinsek, A.T., Sequential confidence bands for densities, *Ann. Statist.*, **23** (1995), 2218–2240.
- [Y98] Yakir, B., On the average run length to false alarm in surveillance problems which possess an invariance structure, *Ann. Statist.*, **26** (1998), 1198–1214.

- [Z71] Zacks, S., The Theory of Statistical Inference, New York, Wiley, 1971.
- [ZSS98a] Zheng, S., Seila, A.F. and Sriram, T.N., Sequential fixed-width confidence interval for the product of two means, *Ann. Inst. Statist. Math.*, **50** (1998), 119–145.
- [ZSS98b] Zheng, S., Seila, A.F. and Sriram, T.N., Asymptotically risk efficient two stage procedure for estimating the product of k (≥ 2) means, *Statist. Decisions*, **16** (1998), 369–387.