Fixed Width Confidence Region for the Mean of a Multivariate Normal Distribution

Hisao Nagao

Osaka Prefecture University, Sakai, Osaka, Japan
E-mail: nagao@ms.osakafu-u.ac.jp

and

M. S. Srivastava

University of Toronto, Toronto, Canada
E-mail: srivasta@utstat.toronto.edu

Received November 30, 1998; published online December 12, 2001

Srivastava gave an asymptotically efficient and consistent sequential procedure to obtain a fixed-width confidence region for the mean vector of any \( p \)-dimensional random vector with finite second moments. For normally distributed random vectors, Srivastava and Bhargava showed that the specified coverage probability is attained independent of the width, the mean vector, and the covariance matrix by taking a finite number of observations over and above \( T \) prescribed by the sequential rule. However, the problem of showing that \( E(T - n_0) \) is bounded, where \( n_0 \) is the sample size required if the covariance matrix were known, has not been available. In this paper, we not only show that it is bounded but obtain sharper estimates of \( E(T) \) on the lines of Woodroofe. We also give an asymptotic expansion of the coverage probability. Similar results have recently been obtained by Nagao under the assumption that the covariance matrix \( S = \sum_{i=1}^{k} s_i A_i \) and \( \sum_{i=1}^{k} A_i = I \), where \( A_i \)'s are known matrices. We obtain the results of this paper under the sole assumption that the largest latent root of \( S \) is simple.


Keywords and phrases: spherical confidence region; asymptotic expansion; coverage probability; largest latent root of covariance matrix; theorem on implicit function; stopping variable; reverse martingale; contrasts of mean.

1. INTRODUCTION

Let \( X_1, X_2, \ldots, X_n, \ldots \) be iid \( p \)-dimensional random vectors which are normally distributed with unknown mean vector \( \mu \) and unknown positive definite covariance matrix \( \Sigma \). We denote it as \( N_p(\mu, \Sigma) \). We wish to find a
spherical confidence region of fixed diameter $2d$ and confidence coefficient $1 - \alpha$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ be the ordered latent roots of $\Sigma$ and $u$ be such that

$$\Pr(\chi^2_{(p)} \leq u) = 1 - \alpha,$$

where $\chi^2_{(p)}$ denotes a chi-square random variable with $p$ degrees of freedom. Then if $\Sigma$ were known, we could take a sample of size

$$n_0 \geq (u\lambda_1/d^2)$$ (1.1)

and obtain a confidence region $R_{n_0}$ of diameter $2d$ for $\mu$ as

$$R_{n_0} = \{ z: (z - \bar{X}_{n_0})' (z - \bar{X}_{n_0}) \leq d^2 \},$$

with the required confidence coefficient $1 - \alpha$, namely, $\Pr(\mu \in R_{n_0}) \geq 1 - \alpha$, where $\bar{X}_{n_0} = n_0^{-1} \sum_{n=1}^{n_0} X_n$. For unknown covariance matrix $\Sigma$, Srivastava (1967) proposed a sequential procedure in which $\Sigma$ is estimated at each stage and the sampling is stopped at

$$T = \inf \{ n \geq m | \hat{\lambda}_{1,n} \leq d^2 n/u_n \},$$ (1.2)

where $u_n \rightarrow u$, $m$ is a given integer number, and $\hat{\lambda}_{1,n}$ is the largest latent root of

$$A_n = \sum_{n=1}^{n} (X_n - \bar{X}_n)(X_n - \bar{X}_n)'/(n-1)$$ (1.3)

and

$$\bar{X}_n = n^{-1} \sum_{n=1}^{n} X_n.$$ (1.4)

The confidence region for $\mu$ is given by

$$R_T = \{ z: (z - \bar{X}_T)' (z - \bar{X}_T) \leq d^2 \}.$$ (1.5)

For $u_n \equiv u$, Srivastava and Bhargava (1979) showed that there exists a finite $k$ independent of $d$, $\mu$, and $\Sigma$ such that

$$\Pr(\mu \in R_{T+k}) \geq 1 - \alpha,$$ (1.6)

where $k$ may depend on $\alpha$. However, the problem of showing that $E(T - n_0)$ is bounded remained open.

In other words, the multivariate version of Theorem 10.3 of Woodrooffe (1982, p. 111) and Theorem 1, Eq. (10) of Simons (1968, p. 1948) have not
been available. Some so-called multivariate versions have been considered in the literature. For example, Mukhopadhyay and Al-Mousami (1986) considered the case when \( \Sigma = \sigma^2 H \), where \( H \) is a known \( p \times p \) symmetric positive definite matrix and Hyakutake et al. (1994) considered the case when \( \Sigma = \sigma^2 [(1 - \rho) I_p + \rho 1'1] \), where \( I_p \) is a \( p \times p \) identity matrix and \( 1' = (1, \ldots, 1) \); the intraclass correlation model. Recently, Nagao (1996) considered the case when \( \Sigma = \sum_{i=1}^{q} \sigma_i^2 A_i \), \( \sum_{i=1}^{q} A_i = I \), where \( A_i \) is a symmetric matrix of rank \( r_i \) such that \( \sum_{i=1}^{q} r_i = p \). Nagao’s model is the most general model in this category and includes the above two models as special cases of this model. All these models, however, have one thing in common, namely, they can all be diagonalized by a known orthogonal matrix; see Theorem 1.7.2 of Srivastava and Khatri (1979, p. 14). Thus, the observation vectors can be transformed by this known orthogonal matrix resulting in \( p \) independent components whose variances can be estimated by using the diagonal elements of the sample covariance matrix of the transformed observation vectors. These estimates are all independently distributed as chi-square with varying degrees of freedom depending on \( r_i \). For example, in the intraclass correlation model, the covariance matrix (after the transformation by known orthogonal matrix) becomes \( \text{diag}(\sigma^2(1 - \rho + pp), \sigma^2(1 - \rho), \ldots, \sigma^2(1 - \rho)) \), which can be estimated by the first diagonal element and the average of the remaining diagonal elements of the sample covariance matrix, respectively. Thus the stopping rule involves chi-square or the maximum of chi-square random variables. In fact the above so-called multivariate models are really many univariate models. The true multivariate case is quite different than the above cases. The stopping rule uses the largest latent root of the sample covariance matrix which is somewhat difficult to handle. In fact, even the asymptotic normality is not available for their case unless one assumes that the largest latent root of the population covariance matrix \( \Sigma \) is simple; see chapter 9 of Srivastava and Khatri (1979). We shall therefore assume that the largest latent root of \( \Sigma \) is simple as in the intraclass correlation model. Many numerical examples, see Srivastava and Carter (1983) chapter 10, show that this case does arise frequently in practice.

When \( p = 1 \), this problem has been considered many times in the statistical literature; see Stein (1945, 1949), Anscombe (1953), Chow and Robbins (1965), Starr (1966), Simons (1968), Woodroofe (1977), Srivastava and Bhargava (1979), and Nagao and Takada (1980). The last reference proposes new stopping rules for this problem and compares them with other rules. Also the second to last reference gives a confidence interval using \( t \)-statistic and width smaller than \( 2d \). Similar results on ellipsoidal confidence regions, except for asymptotic results given in Srivastava and Bhargava (1979), are not yet available.
The organization of the paper is as follows. In Section 2, we give an asymptotic expansion of the largest root of the sample covariance matrix $A_n$ defined in (1.3), the key result required to obtain the bound on average sample size and the coverage probability given in Section 3. In Section 4, we consider the problem of estimating a linear contrast under the quadratic loss function and when the sample cost is $c$ per unit of sample.

2. ASYMPTOTIC DISTRIBUTIONS OF THE LARGEST LATENT ROOT AND STOPPING RANDOM VARIABLE

We first obtain the asymptotic distributions of the largest latent root of the sample covariance matrix $A_n$ under the assumption that the largest latent root of the population covariance matrix $S$ is simple. Since the latent roots are invariant under orthogonal transformations, we shall assume without any loss of generality that the covariance matrix $S$ is a diagonal matrix $L$ with its diagonal elements $l_1 > l_2 > \cdots > l_p > 0$. Thus, it is assumed that $\lambda_1$ is simple. By Helmert’s transformation, we can write $A_{n+1} = \sum_{a=1}^n y_a y_a' / n$, where $y_a$’s are iid $N(0, L)$. At first, we show that $\hat{\lambda}_{1, n+1}$, the largest latent root of $A_{n+1}$, can be written in distribution as

$$\hat{\lambda}_{1, n+1} = \lambda_1 + (a_{11}^{(n)} - \lambda_1) + v_n = a_{11}^{(n)} + v_n,$$

where $a_{11}^{(n)}$ and $v_n$ is a slowly changing random variable, where $u_k$, $n \geq 1$, is said to be slowly changing if and only if $n^{-1} \max\{|u_1|, \ldots, |u_n|\} \to 0$ in probability as $n \to \infty$, and for every $\varepsilon > 0$, there is a $\delta > 0$ for which

$$\Pr\{ \max_{0 \leq k < n} |u_{n+k} - u_k| > \varepsilon \} < \varepsilon \quad \text{for all} \quad n \geq 1;$$

that is, it is also uniformly continuous in probability (ucip); see Woodroofe (1982, p. 41). We apply a theorem on implicit function. Let $f(A_{n+1}, \ell) = |A_{n+1} - \ell L| = 0$. Regarding $\hat{\lambda}_{1, n+1}$ as a function of $A_{n+1} = (a_{ij}^{(n)})$, we expand it around $A$. Letting $a_{ij}^{(n)} = a_{ij}$, we get

$$\frac{\partial f}{\partial a_{11}}|_{A_{n+1} = A} = \begin{vmatrix} 1 & 0 & \cdots & 0 \\ \lambda_2 - \lambda_1 & \ddots & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \lambda_p - \lambda_1 \end{vmatrix} = \prod_{k=2}^p (\lambda_k - \lambda_1)$$
and
\[\frac{\partial f}{\partial \ell} \bigg|_{A_{n+1}} = -\prod_{k=2}^{p} (\lambda_k - \lambda_1).\]
The other values \(\frac{\partial f}{\partial a_{ij}}\) on \(A\) are zero. We put \(f_{ij}(A_{n+1}, \ell) = (\partial/(\partial a_{ij})) f(A_{n+1}, \ell), \quad f_{ij,kl}(A_{n+1}, \ell) = (\partial^2/(\partial a_{ij} \partial a_{kl})) f(A_{n+1}, \ell), \quad f_{k}(A_{n+1}, \ell) = (\partial/\partial \ell) f(A_{n+1}, \ell)\) etc. Then
\[f_{ij}(A_{n+1}, \ell) + f_{ki}(A_{n+1}, \ell) \frac{\partial \ell}{\partial a_{ij}} = 0 \quad (2.3)\]
and
\[f_{ij,kl}(A_{n+1}, \ell) + f_{ij,l}(A_{n+1}, \ell) + f_{kl,i}(A_{n+1}, \ell) + f_{\ell,i}(A_{n+1}, \ell) \frac{\partial \ell}{\partial a_{ij}} + f_{\ell,k}(A_{n+1}, \ell) \frac{\partial^2 \ell}{\partial a_{ij} \partial a_{kl}} = 0. \quad (2.4)\]
Thus we have, as \(n \to \infty\), from (2.4)
\[\frac{\partial^2 \ell}{\partial a_{ij} \partial a_{kl}} \begin{cases} \frac{2}{\lambda_j - \lambda_i} & i = k = 1, \quad l = j \\ 0 & \text{otherwise.} \end{cases} \quad (2.5)\]
Therefore we have by (2.3) and (2.4)
\[\hat{\lambda}_{1,n+1} = \lambda_1 + (a_{11}^{(n)} - \lambda_1) + v_n \]
\[= a_{11}^{(n)} + v_n, \quad (2.6)\]
where
\[v_n = \sum_{i,j} \sum_{k<l} A_{ij,kl}(A^*)(a_{ij} - \lambda_i)(a_{kl} - \lambda_k \delta_{kl}) \quad (2.7)\]
and \(A_{ij,kl} = (\partial^2 \ell)/(2\partial a_{ij} \partial a_{kl}), \delta_{ij}\) is the Kronecker delta, and \(A^*\) is some point between \(A\) and \(A_{n+1}\). Since \(\sqrt{n}(A_{n+1} - A)\) has a limiting normal distribution, it follows that \(v_n\) is a slowly changing random variable. For asymptotic expansions, see Nagao (1970) and Sugiura (1973). Thus we get the following
Lemma 2.1. Let $\hat{\lambda}_{1,n+1}$ be the largest latent root of $A_{n+1} = (a_{ij}^{(n)})$ which is distributed as $W_p(\frac{1}{n}A, n)$, where $A = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$. $\hat{\lambda}_1 > \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p$; that is $\hat{\lambda}_1$ is a simple root. Then

$$\hat{\lambda}_{1,n+1} = a_{11}^{(n)} + v_n,$$

where $v_n$ is a slowly changing random variable defined above.

Next, we study the properties of the stopping variable $T_a$ defined by (2.9) with

$$u_n = u_\ell = u \left[ 1 + \frac{1}{n} \ell_0 + o(n^{-1}) \right], \quad \ell_0 > 0. \quad (2.8)$$

That is

$$T_a = \inf \left\{ n \geq m \mid n \geq \frac{u_\ell \hat{\lambda}_{1,n+1}}{d^2} \right\}. \quad (2.9)$$

Intuitively, the choice of $u_n$ in (2.8) is clear. For when the covariance matrix $\Sigma$ is known, $u_n$ is obtained from the chi-square distribution and when it is unknown, it may be obtained from an F-distribution with $p$ and $n-p$ degrees of freedom. Thus we have $u_n = \frac{p_n \ell_0}{n-p} F_{p,n-p}^p(1-\alpha)$, where $F_{p,n-p}^p(1-\alpha)$ is the lower 100$(1-\alpha)$% point of an F-distribution with $p$ and $n-p$ degrees of freedom. The $u_n$ so obtained should satisfy (2.8); see Srivastava and Bhargava (1979), Fisher (1925), and Fisher and Cornish (1960). Then, $\ell_0$ can be obtained with the help of F-distribution and chi-square. In fact, $u_n = u(1 + \frac{1}{n} \ell_0 + o(n^{-1}))$, with $\ell_0 = p + 1$, by Nagao (1973). The optimum value of $\ell_0$ is, however, difficult to obtain. We shall now introduce another stopping variable $N_a$ defined by

$$N_a = \inf \{ n \geq m-1 \mid Z_n > a \}, \quad (2.10)$$

where

$$Z_n = \frac{(n+1) \hat{\lambda}_1}{\hat{\lambda}_{1,n+1} \ell_{n+1}} \quad \text{and} \quad a \equiv n_0 = \frac{u_\lambda}{d^2}.$$ 

Then, $T_a = N_a + 1$. From (2.9) we can write $Z_n$ as

$$Z_n = S_n + \xi_n, \quad (2.11)$$
where \( S_n = \sum_{i=1}^{n} (2 - (1 / \lambda_i) y^{(i)}_\alpha) \) and \( y^{(i)}_\alpha \) is the first component of vector \( y_\alpha \) where \( y_\alpha \)'s are iid \( N_p(0, \Lambda) \), and

\[
\xi_n = -2(\ell_0 - 1) \frac{n}{\lambda_1} v_n + \frac{\ell_1}{\lambda_1^3} (n - (\ell_0 - 1 + b_n))(a^{(n)}_{11} - \lambda_1 + v_n)^2
\]

\[
+ \frac{(\ell_0 - 1 + b_n)}{\lambda_1}(a^{(n)}_{11} + v_n) = -2(\ell_0 - 1) + V_n
\]

(2.12)

and \( \lambda^*_1 \) is some point between \( \lambda_1 \) and \( \lambda_{1,n+1} \) and \( b_n = o(1) \). Here we prove that \( \xi_n \) is slowly changing. We find that \( \xi_n/n \to 0 \) a.s., since \( v_n \to 0 \) and \( a^{(n)}_{11} - \lambda_1 \to 0 \) a.s. Next we show that \( \xi_n \) is uniformly continuous in probability. Since \( |d_{ij,kl}(\Lambda*)| \leq M \) and \( \sqrt{n}(a^{(n)}_{ij} - \lambda_{dl}) \) are ucip and converges in law, \( v_n \) is also ucip. Similarly we show that \( (\lambda_1/\lambda^*_1)^2 \) is also ucip. For other terms, here we use the following simple lemma.

**Lemma 2.2.** Let \( U_n \) be a ucip and let it converge in law. If a real sequence \( a_n \) converges to zero, then \( a_n U_n \) is also a ucip.

**Proof.** \( \Pr(\max_{0 \leq k < n} |a_{n+k} U_{n+k} - a_n U_n| \geq \epsilon) = \Pr((\max_{0 \leq k < n} |a_{n+k} (U_{n+k} - U_n) + (a_{n+k} - a_n) U_n| \geq \epsilon) \leq \Pr(\max_{0 \leq k < n} a_{n+k} |U_{n+k} - U_n| \geq \epsilon/2) + \Pr(\epsilon' |U_n| \geq \epsilon/2) \leq \epsilon, \) for enough small \( \epsilon' > 0 \).

Thus \( \xi_n \) can be shown to be a ucip. Since \( m_n \) and \( \sqrt{n}(a^{(n)}_{11} - \lambda_1) \) converge in law, we have \( \xi_n / \sqrt{n} \to 0 \) in probability. By Woodroofe (1982), we have

**Lemma 2.3.** Let \( n_0 = u \lambda_1 / d^2 \), then we have

\[
\frac{N_n - n_0}{\sqrt{n_0}} \to N(0, 2).
\]

(2.13)

Next we have

**Lemma 2.4.**

\[
\begin{align*}
(1) & \quad \frac{N_n}{n_0} \to 1 \\
(2) & \quad \mathbb{E} \left( \sup_{a \geq 1} \left( \frac{N_n}{n_0} \right)^q \right) < \infty \quad \text{for all} \quad q > 0 \\
(3) & \quad \int_{N_n > 2a} N_n^2 \, dP \to 0 \quad \text{as} \quad a \to \infty.
\end{align*}
\]
We give an outline of the proof. The term (1) follows from $\xi_n/n \to 0$ a.s., since $v_n$ and $(a_{ij}^{(n)} - \lambda_i)$ converge to 0 a.s. For (2), we note that $(N_a/a)^r = ((T_a - 1)/a)^r \leq ((\ell_a p \operatorname{tr} A_a)/(p\lambda_i))^r$ and $(\operatorname{tr} A_a)/(p\lambda_i)$ is a reverse martingale, and hence the result follows. For (3), we need to show according to Woodroofe (1991) that

$$\sum_{n=1}^{\infty} n \Pr(\xi_n \leq -ne) < \infty \quad \text{for some } 0 < e < 1.$$  \hspace{1cm} (2.14)

We first note that

$$\Pr(\xi_n \leq -ne) \leq \Pr\left(\frac{1}{\lambda_i} |v_n| \geq e/3\right) + \Pr\left(\frac{\lambda_i}{\lambda_1} (n - (\ell_n - 1 + b_n)) \right. \times (a_{ij}^{(n)} - \lambda_i + v_n)^2 \geq e/3 \left. \right) + \Pr\left(\frac{(\ell_n - 1 + b_n)}{\lambda_i} (a_{ij}^{(n)} + |v_n|) \geq e/3\right).$$

By considering higher order moments of $(a_{ij}^{(n)} - \lambda_i, \delta_{ij})^2$, we can show that $\Pr((a_{ij}^{(n)} - \lambda_i, \delta_{ij})^2 \geq c \lambda) \leq (\text{const}) n^{-x}$ $(x \geq 3)$, and hence the first term is finite. The other terms can be proved by similar consideration. Thus we have (3).

**Lemma 2.5.**

$$\Pr(N_a = m-1) \sim \Pr\left(N_a \leq \frac{a}{2}\right) \sim ca^{-p(m-1)}.$$  \hspace{1cm} (2.15)

**Proof.** We have

$$\Pr(N_a = m-1) = \Pr\left(\frac{m\lambda_i}{\lambda_i, m\ell_m} > a\right) \leq \Pr\left(\frac{mp\lambda_i}{a\ell_m} \geq \operatorname{tr} A_m\right) \leq \Pr\left(\frac{m(m-1)p\lambda_i}{a\lambda_i, m\ell_m} > \chi^2_{[p(m-1)]}\right) \sim c'a^{-p(m-1)/2}$$

and similarly,

$$\Pr(N_a = m-1) = \Pr\left(\frac{m\lambda_i}{a\ell_m} > \lambda_i, m\right) \geq \Pr\left(\frac{m\lambda_i}{a\ell_m} > \operatorname{tr} A_m\right) \sim c'a^{-p(m-1)/2}.$$

Then, the proof follows as in Simons (1968), Srivastava (1973), Srivastava and Bhargava (1979), or Woodroofe (1982).
Lemma 2.6. If \( p(m-1) > 2 \), \( N_a^* = (N_a - n_0)/\sqrt{n_0} \) is uniformly integrable for \( a > 0 \).

Proof. At first we show that

\[
\int_{N_a \leq a/2} N_a^* \, dP + \int_{N_a > 2a} N_a^* \, dP \to 0. \tag{2.16}
\]

Since \( \int_{N_a \leq a/2} N_a^* \, dP \leq a^{-1}(m-1-a)^2 \Pr(N_a \leq a/2) \) and \( p(m-1) > 2 \), we get the first part of (2.16) from Lemma 2.5, the second part follows from Lemma 2.4(3). Next we consider the probability \( \Pr(N_a > a/2, N_a^* < -x) \).

When \( \sqrt{a}/2 \leq x \), the event is empty, so we consider \( 0 \leq x \leq \sqrt{a}/2 \). Thus we have \( ((k+1)\lambda_1)/(\hat{\lambda}_{1,k+1} \epsilon_{k+1}) > a \) for some \( k \in (a/2, a-\sqrt{a}/2) \). Since \( (k+1)/(a \epsilon_{k+1}) \leq 1/(a \epsilon_{k+1})(a-\sqrt{a}/2+1) = 1-\epsilon/a \) for some \( \epsilon > 0 \).

We have

\[
\Pr(N_a > a/2, N_a^* < -x) \leq \Pr \left( 1-\frac{a}{x} + \frac{\epsilon}{a} > \frac{\hat{\lambda}_{1,k+1}}{\lambda_1} \exists k \in (a/2, a-\sqrt{a}/2) \right)
\]

\[
\leq \Pr \left( \max_{k \leq a} \left| \frac{\hat{\lambda}_{1,k+1}}{\lambda_1} - 1 \right| \geq \frac{x}{2} + \epsilon' \right)
\]

\[
\leq \Pr \left( \max_{k \leq a} \left| \frac{d_{11}^{(k)}}{\lambda_1} - 1 \right| \geq \frac{x}{4} \right)
\]

\[
+ \Pr \left( \max_{k \leq a} \left| \frac{v_k}{\lambda_1} \right| \geq \frac{x}{4} \right).
\]

By martingale inequality, the last two formulas are less than \( cx^{-4} \). Next let us consider \( \Pr(N_a < 2a, N_a^* > x) \). When \( 0 \leq x \leq \sqrt{a} \), we have

\[
\Pr(N_a < 2a, N_a^* > x) \leq \Pr(N_a \geq a + \sqrt{a}x) \leq \Pr \left( \frac{(a+\sqrt{a}x)\hat{\lambda}_1}{\lambda_1} \leq a \right)
\]

\[
= \Pr \left( \left( 1 + \frac{x}{\sqrt{a}} \right) \frac{1}{\hat{\lambda}_{1,a+\sqrt{a}x}} \leq \frac{\hat{\lambda}_{1,a+\sqrt{a}x}}{\lambda_1} \right)
\]

\[
\leq \Pr \left( \frac{x}{\sqrt{a}} + o(a^{-1/2}) \leq \frac{\hat{\lambda}_{1,a+\sqrt{a}x}}{\lambda_1} - 1 \right).
\]

Since \( \hat{\lambda}_{1,n+1} = a_{11}^{(n)} + v_n \) in (2.6), Markov inequality yields \( \Pr(N_a < 2a, N_a^* > x) \leq cx^{-4} \). Thus \( N_a^* \) is uniformly integrable.
3. ASYMPTOTIC EXPANSION

In this section, we give asymptotic expansions of the mean of the time $T_a$ and $\Pr((\bar{X}_a - \mu) \leq d^2$). Since the random walk is given by $\sum_{i=1}^n (2 - (1/\lambda_i) y_{i,j}^{(l)})$, we have $E(2 - (1/\lambda_i) y_{i,j}^{(l)}) = 1$ and $\text{Var}(2 - (1/\lambda_i) y_{i,j}^{(l)}) = 2$.

Also by (2.5) as $n \to \infty$, we have $V_n \to -\sum_{i=2}^p (1/(\lambda_i - \lambda_1)) W_i^2 + \lambda_1^{-2} W^2 + (\ell_0 - 1)$, where $W_i$ are normally distributed random variables with mean 0 and variance $\lambda_1 \lambda_j$ and $W$ is a normally distributed random variable with mean 0 and variance $2\lambda_1^2$. Also we show that $\zeta_{a+k}$ for $0 \leq k \leq n$ uniformly integrable. We have

$$\Pr(\max_{0 \leq k \leq n} |\zeta_{a+k}| \geq y) \leq \Pr\left(\max_{0 \leq k \leq n} \frac{(n+k)}{\lambda_1} v_{n+k} \geq y/3\right)$$

$$+ \Pr\left(\max_{0 \leq k \leq n} \frac{\lambda_1}{\lambda_1} (n+k - (\ell_0 - 1 + b_{n+k}))(a_{i1}^{(n+k)} - \lambda_1 + v_{n+k})^2 \geq y/3\right)$$

$$+ \Pr\left(\max_{0 \leq k \leq n} \frac{(\ell - 1 + b_{n+k})}{\lambda_1} (a_{i1}^{(n+k)} - \lambda_1 + v_{n+k})^2 \geq y/3\right).$$

For example, by the martingale inequality, we can choose $\alpha$ large enough such that $\Pr(\max_{0 \leq k \leq n} (n+k)(a_{i1}^{(n+k)} - \lambda_1 \delta_i)^2 \geq cy) \leq \Pr(\max_{0 \leq k \leq n} (n+k)^2 (a_{i1}^{(n+k)} - \lambda_1 \delta_i)^2 \geq cy) \leq c y^{-\alpha}$. Similarly we can get the desired inequality by a similar consideration. The formula $\sum_{n=1}^\infty \Pr(\zeta_n \leq -nw) < \infty$ for some $0 < w < 1$ can be obtained by (2.14). Thus we have

**Theorem 3.1.** *Let $T_a$ be a stopping time defined by (2.9). If $\lambda_1$ is a simple root, then $a \to \infty$,*

$$E(T_a) = a + \rho + (\ell_0 - 2) + \sum_{j=2}^p \frac{\lambda_j}{\lambda_1} + o(1),$$

(3.1)

*where $\rho = \frac{1}{2} - \sum_{k=1}^\infty (S_k^-)^2$ and $S_k^- = \max(-S_k, 0)$.*

Note that

$$S_k^- = \max(-S_k, 0)$$

$$= \max\left(\frac{1}{\lambda_1} \sum_{s=1}^k y_s^{(l)} - 2k, 0\right)$$

$$= \max(\chi_k^2 - 2k, 0),$$
where $\chi^2_{[k]}$ is a chi-square random variable with $k$ degrees of freedom. Hence,

$$E(S^2) = E[(\chi^2_{[k]} - 2k) I(\chi^2_{[k]} > 2k)]$$

$$= k Pr(\chi^2_{[k+2]} > 2k) - 2k Pr(\chi^2_{[k]} > 2k).$$

Thus

$$p = 3 - \sum_{k=1}^{\infty} \left[ Pr(\chi^2_{[k+2]} > 2k) - 2 Pr(\chi^2_{[k]} > 2k) \right],$$

which can be evaluated.

We also note that $\sum_{i=2}^{p} \lambda_i / (\lambda_i - \lambda_1)$ vanishes when $p = 1$ and this summation is negative. Also when $p = 1$, this result reduces to the case in Woodroofe (1977).

Finally, we evaluate the probability $Pr((\bar{X}_a - \mu)' (\bar{X}_a - \mu) \leq d^2) \geq Pr(\chi^2_{[p]} \leq T_0 d^2 / \lambda_1) = E\psi_p(T_0 d^2 / \lambda_1) = E\psi_p(u((N_a + 1) / n_0))$, where $\psi_p(x) = Pr(\chi^2_{[p]} \leq x)$. By Lemma 2.3, $N_a / n_0 \to 1$. We give Taylor expansion $\psi_p(ux + h)$ about $x = 1$. After some calculation, we have

$$E\psi_p \left( u \frac{N_a + 1}{n_0} \right) = \psi_p(u) + u \frac{\psi'_p(u)}{n_0} - 1 + \frac{u^2}{2} E\psi''_p(\ast) \left( \frac{N_a}{n_0} - 1 \right) + o(a^{-1}),$$

where $\ast$ is some point between $u$ and $u((N_a + 1) / n_0)$. Since $\psi''_p(x) = \psi'_p(x) \{(p-2)x^{-1} - 1\}/2$, $\psi''_p(x)$ is bounded when $p - 2 \geq 2$ or $p - 2 = 0$. In these cases, by Lemma 2.6, $E\psi''_p(\ast) N_a^{-1} \to 2\psi''_p(u)$. When $p = 3$, we have $|\psi''_p(x)| \leq c_1 + c_2 x^{-3/2}$ for some constant numbers $c_1$ and $c_2$. When $N_a > a/2$, we have $\ast > u/2$. Thus $|\psi''_p(\ast)| \leq c_1 + c_2 (u/2)^{-1/2}$. Then on $N_a > a/2$, $\psi''_p(\ast) N_a^{-1/2}$ is uniformly integrable. The last step is to show that $\int_{N_a \leq a/2} \ast^{-1/2} N_a^{*} \, dP \to 0$ as $a \to \infty$. Since $N_a \leq a/2$, we have $\ast \geq N_a u / a$. Therefore

$$\int_{N_a \leq a/2} \ast^{-1/2} N_a^{*} \, dP \leq u^{-1/2} \int_{N_a \leq a/2} \left( \frac{a}{N_a} \right)^{1/2} N_a^{*} \, dP$$

$$\leq (mu)^{-1/2} a^{3/2} Pr(N_a \leq a/2).$$

When $(m - 1) p > 3$, the above converges to zero. Therefore we have
Theorem 3.2. If \((m-1)p > 3\) and \(\lambda_1\) is a simple root, we have

\[
\Pr((\bar{X}_T - \mu)' (\bar{X}_T - \mu) \leq d^2) \geq (1-\alpha) + \frac{\mu}{a} \psi_p'(u) \mathbb{E}(T_a - n_0) + \frac{\mu^2}{a} \psi_p''(u) + o(a^{-1}),
\]

(3.2)

where \(u\) is the upper 100\% point of the chi-square distribution with \(p\) degrees of freedom.

When \(p = 1\), we get Woodroofe’s (1977) result.

Thus \(\Pr(\mu \in R_T)\) could be less than \(1-\alpha\). On the other hand, it has been shown by Srivastava and Bhargava (1979) that there exists a \(k\) such that \(\Pr(\mu \in R_{T+k}) \geq 1-\alpha\). It would thus be desirable to take some additional observations than those dictated by the stopping rule. Starr (1966) has shown that the initial sample size plays a significant role.

4. SAMPLE SIZE FOR ESTIMATING CONTRASTS OF THE MEANS

In this section, we consider the problem of estimating all normalized linear combinations of the mean under the quadratic loss function and when the cost of sampling is \(c\) per unit sample, as defined in (4.1). The result of this section parallels that of Woodroofe’s chapter 10 as applied to the quadratic loss function. Other references in this area are Robbins (1959), Starr (1966), Starr and Woodroofe (1968, 1969), Woodroofe (1977), and Ghosh et al. (1997).

Thus, the problem with which we are concerned in this section is to find the sample size \(n\) such that

\[
\max_{|\alpha| = 1} \mathbb{E}((\alpha' \bar{X}_n - \alpha' \mu)^2) + cn
\]

is minimized, where \(c\) is the cost per unit sample. Then if \(\Sigma\) were known, the minimum sample size minimizing (4.1) would be given by \(n_0 = \sqrt{\lambda_1/c}\), where \(\lambda_1\) is a maximum latent root of \(\Sigma\). When \(\Sigma\) is unknown, we consider the following stopping time with \(\ell_n = 1 + (\ell_0/n) + o(n^{-1})\), \(\ell_0 > 0\),

\[
T^* = \inf\{n \geq m \mid n \geq \sqrt{\lambda_{1,e}\ell_0/c}\}.
\]

(4.2)

The reader is reminded that the \(\ell_0\) here has no connection with \(\ell_0\) used in the previous section. The \(\ell_0\) here should be chosen such that \(\mathbb{E}(\ell_n \hat{\lambda}_{1,e}) \approx \lambda_1\).
Similar to the previous section, we define

\[ N^*_a = \inf\{n \geq m - 1 \mid Z_n \geq a\}, \]

where \( Z_n = (n+1) \sqrt{\lambda_1/(\lambda_{n+1} + \ell_{a,n+1})} \) and \( \lambda = \sqrt{\lambda_1/c} \). Then we have \( T^* = N^*_a + 1 \) and \( Z_n = S_n + \hat{\epsilon}_n \), where \( S_n = \sum_{s=1}^{n} \left((3/2) - (1/2 \lambda_1) y_s^{(i)_s}\right) \) and

\[ \hat{\epsilon}_n = \frac{3}{2} - \frac{1}{2n \lambda_1} \sum_{s=1}^{n} y_s^{(i)_s} - \frac{(n+1)}{2 \lambda_1} v_n + \frac{3}{8} (n+1) b^{5/2} \left( \frac{1}{n \lambda_1} \sum_{s=1}^{n} y_s^{(i)_s} + v_n - 1 \right)^2 \]

with \( b \) between 1 and \( \hat{\ell}_{\lambda_1, n+1}/\lambda_1 \). Therefore we have

**Theorem 4.1.** When \( \lambda_1 \) is a simple root, the mean of \( T^* \) in (4.2) is given by

\[ E(T^*) = a + \rho - (\ell_0 + 2) + \sum_{j=2}^{\rho} \frac{\lambda_1}{\lambda_j - \lambda_1} + o(1), \]

where \( \rho = \frac{3}{\lambda_1} - \sum_{k=1}^{\infty} \mathbb{E}S_k^2 \) and \( S_k^2 = \max(-S_k, 0) \).

We note that \( \rho \) is given by

\[ \rho = \frac{3}{2} - \frac{1}{2} \sum_{k=1}^{\infty} \left\{ \Pr(\chi_{2k+2}^2 > 3k) - 3 \Pr(\chi_{2k}^2 > 3k) \right\}. \]

Next we consider the regret

\[ R = E \left( \frac{\lambda_1}{T^*_a + \epsilon T^*} - \frac{\lambda_1}{T^*_a + \epsilon n_0} \right) = cn_0 \left\{ E \left( \frac{n_0}{T^* + \epsilon n_0} \right) - 2 \right\}. \]

Let \( f(x) = x^{-1} + x \). Then the regret is written by

\[ R = cn_0 E \left( f \left( \frac{T^*}{n_0} \right) - f(1) \right). \]

As in Section 2, we set asymptotic results of \( T^* \) as \( T_0 \) as \( a \to \infty \). Then

\[ R = cn_0 E(b^{-3}(T^*_a/n_0 - 1)^2), \]

where \( b \) is some point between 1 and \( T^*_a/n_0 \). As \( a \to \infty \),

\[ n_0 \int_{T^*_a > n_0/2} b^{-3} \left( \frac{T^*_a}{n_0} - 1 \right)^2 \, dP = \int_{T^*_a > n_0/2} b^{-3} \left( \frac{T^*_a - n_0}{n_0} \right)^2 \, dP \to \frac{1}{2}. \]
Also for \( T^*/n_0 \leq \frac{1}{2} \), we evaluate
\[
0 \leq n_0 \int_{T^*/n_0 \leq 1/2} \left\{ \frac{n_0}{T^*} + \frac{T^* - 2}{n_0} \right\} dP \leq n_0 \int_{T^*/n_0 \leq 1/2} \left\{ \frac{n_0}{T^*} - \frac{3}{2} \right\} dP
\]
\[
\leq n_0 \int_{T^*/n_0 \leq 1/2} \frac{1}{T^*} dP \leq \frac{n_0}{m} \Pr\left( \frac{T^*}{n_0} \leq \frac{1}{2} \right) = o(1) \quad \text{if} \quad p(m-1) > 2.
\]
Thus we have the following theorem.

**Theorem 4.2.** If \( p(m-1) > 2 \) and \( \lambda_i \) is a simple root, the regret \( R \) is given by
\[
R = \frac{c}{2} + o(c).
\]

As a final problem, we consider the confidence interval of any linear combination of mean vector \( \mu \) for a fixed confidence coefficient \( 1 - \alpha \) and fixed width \( d > 0 \), that is, \( \Pr(||l'||X_n - l'\mu|| \leq d \text{ for all } ||l'|| = 1) \geq 1 - \alpha \). But, by Schwarz’s inequality, we have \( \Pr(||l'||X_n - l'\mu|| \leq d \text{ for all } ||l'|| = 1) = \Pr((X_n - \mu)'(X_n - \mu) \leq d^2) \). Thus this problem reduces to the one treated in Section 2.

**ACKNOWLEDGMENT**

A part of this paper was completed during Nagao’s visit to the University of Toronto. This research was partially supported by Natural Sciences and Engineering Research Council of Canada. The authors also thank the editor and the referee for their useful comments.

**REFERENCES**


