Jour. Ind. Soc. Ag. Statistics
47(1), 1995 : 1–11

Multi-Stage Point Estimation of the Mean Vector of a Multinormal Population

R. Karan Singh, Anita Singh and Ajit Chaturvedi*
Lucknow University, Lucknow, 226007
(Received : July, 1988)

SUMMARY

Three-stage and purely sequential procedures are developed for the minimum risk point estimation of the mean vector of a multinormal population. Asymptotic properties of the proposed estimation procedures are studied. A comparative study of the two estimation procedures is done and the three-stage procedure is found to be strongly competitive with the purely sequential procedure. Consideration is given to a family of loss functions.

Key Words : Mean vector, Multinormal population, Loss, Risk, Sequential and three-stage procedure, Risk-efficiency, Regret.

Introduction

Starr [8] developed a sequential procedure for the point estimation of the mean of univariate normal population (the variance being unknown) under a family of loss functions. He proved the asymptotic 'risk-efficiency' of the proposed procedure. Chaturvedi [3] obtained second-order approximations for the 'regret' associated with the sequential procedure of Starr [8].

Wald [11] mentioned that due to one-by-one sampling, the purely sequential procedures are complicated in nature to apply. However, this difficulty can be reduced by sampling in 'bulks'. In order to construct fixed-width confidence interval for the mean of a univariate normal population, Stein [10] developed a sampling scheme which required only two stages. However, Hall [4] pointed out that in many situations, Stein's two-stage procedure leads to considerable oversampling. We also observed that since it utilizes only first-stage sample for the estimation purpose, there is a loss of 'information' obtained from the sample.

From the theory discussed above, it is conlcuded that both the two-stage and purely sequential procedures suffer from certain drawbacks. Hall [4] deduced that if one more stage is appended to Stein's two-stage procedure, the resulting three-stage procedure becomes strongly competitive with the purely sequential procedure. Hall [4] utilized this three-stage procedure to handle the

* Ch. Charan Singh University, Meerut, 250 005.

problem of obtaining fixed-width confidence interval for a univariate normal mean. In fact, the three-stage procedure combines the advantages of both the two-stage and purely sequential procedures. The two-stage, three-stage and purely sequential procedures to construct fixed-size confidence region for the mean vector of a multinormal population are available in Singh and Chaturvedi [7].

The purpose of this note is two-fold. In section 2, the sequential procedures of Starr [8] amd Chaturvedi [3] are extended for the point estimation of the mean vector of a multinormal popultion. As compared to Starr [8], an alternative proof of the asymptotic 'risk-efficiency' is provided and second-order approximation for the 'regret' is obtained. In section 3 for the same estimation problem, a three-stage procedure is established. A comparative study of the two methods of estimation is done at the end of the note. The set-up of the problem is as follows.

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be a sequence of iid rv's from a p-variate normal population $N_p(\mu, \sigma^2 \Sigma)$. Here, μ is p×1 unknown mean vector, $\sigma \in (0, \infty)$ is an unknown scalar and Σ is a aknown p×p positive definite matrix. After having recorded a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ of size n, let the loss of estimating μ by

$$\overline{\mathbf{X}}_n = n^{-1} \sum_{i=1}^{n} x_i \qquad \text{be}$$

$$L(\mu, \overline{\mathbf{X}}_n) = A[(\overline{\mathbf{X}}_n - \mu), \Sigma^{-1} (\overline{\mathbf{X}}_n - \mu)]^{s/2} + C \log n \tag{1.1}$$

where A, s and C are known positive constants.

The risk corresponding to the loss function (1.1) is

$$R_n(C) = A(p, s)\, \sigma^s n^{-s/2} + C \log n, \tag{1.2}$$

where $A(p, s) = A\, E[X^2_p]^{s/2}$. The value $n = n_0$ which minimizes (1.2) is

$$n_o = [A(p, s)\,)s/2C)]^{2/s} \sigma^2, \tag{1.3}$$

and the corresponding minimum risk is

$$R_{n_o}(C) = 2s^{-1} C + C \log n_o. \tag{1.4}$$

But, σ being unknown, no fixed sample size procedure minimizes $R_n(C)$ sinultaneously for all values of σ. In such a situation, adopt estimation procedures determining sample sizes as rv's. The procedures and their properties are discussed in the next two sections

2. *The Purely Sequential Procedure*

Let, for $n \geq m(\geq 2)$, $\sigma_n^2 = [p(n-1)]^{-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)' \Sigma^{-1} (X_i - \overline{X}_n)$.
Then, the stopping time $N = N(C)$ is defined by

$$N = \inf [n \geq m : n \geq \{A(p,s)(s/2C)\}^{2/s} \sigma_n^2] . \tag{2.1}$$

Estimate μ by $\overline{X}_N$.

Since the events '$L(\mu, \overline{X}_N)$' and 'N' are stochasticaly independent [see Wang [12]] the risk associated with the sequential procedure (2.1) is

$$R_N(C) = 2s^{-1} C E[(n_o/N)^{s/2}] + C E(\log N) . \tag{2.2}$$

Following Starr [8] and Starr and Woodroofe [9] define the 'risk-efficiency' and 'regret' of the procedure (2.1). respectively by

$$R_e(C) = R_N(C)/R_{n_o}(C) \tag{2.3}$$

and

$$R_g(C) = R_N(C) - R_{n_o}(C) . \tag{2.4}$$

First establish two lemmas.

Lemma 1. Hayre [5] : Let $z_1, z_2, \ldots$ be iid chi-squared with one degree of freedom, and let

$$\eta_k = \inf_{n \geq k} [n^{-1}(z_1 + \ldots + z_n)] .$$

Then, for $t > 0, E(\eta_k^{-t}) < \infty$ if $k > 2t$.

Lemma 2 : N is well-defined, monotonically non-increasing in C, and

$$\lim_{C \to 0} N = \infty \text{ a.s.}, \tag{2.5}$$

$$\lim_{C \to 0} (N/n_o) = \text{a.s.}, \tag{2.6}$$

$$\lim_{C \to 0} E[(\log N)/(\log n_o)] = 1, \tag{2.7}$$

and, for $t > o$

$$\lim_{C \to 0} E[(n_o/N)^t] = 1, \text{ if } m > 1 + 2p^{-1} t . \tag{2.8}$$

Proof : Result (2.5) follows from the definition (2.1) of N.

Notice the inequality

$$A(p, s)\,(s/2C)^{2/s}\ \sigma_N^2 \le N \le A(P\,.\,s)\,(s/2C)^{2/s}\ \sigma_N^2 + (m-1)$$

or $$(\sigma_N^2/\sigma^2) \le (N/n_o) \le (\sigma_N^2/\sigma^2) + (m-1)/(n_o)\,.$$

which, on using (2.5) and the results $\lim_{N \to \infty} \sigma_N^2 = \sigma^2$ a.s. . ., $\lim_{C \to 0} n_o = \infty$ a.s., leads to (2.6).

To prove (2.7), let. for $0 < \epsilon < 1$, $\theta_1 = (1-\epsilon)\,n_o, \theta_2 = (1+\epsilon)\,n_o$. We have

$$\begin{aligned} E(\log N) &= \sum_{n=m}^{\infty} (\log n)\,P\,(N = n) \\ &\ge \sum_{n > \theta_1} (\log n)\,P\,(N = n) \\ &\ge (\log \theta_1)\ P\,(N > \theta_1) \\ &\ge [\,\log(1-\epsilon) + \log n_o]\ P(N/n_o > 1-\epsilon), \end{aligned}$$

which, on utilizing (2.6), gives

$$\lim_{C \to 0}\ \inf E\,[(\log N)/(\log n_o)] \ge 1. \qquad (2.9)$$

on the other hand,

$$\begin{aligned} E(\log n) &= \sum_{n=m}^{\theta_2} (\log n)\,P\,(N = n) + \sum_{n > \theta_2} (\log n\,)\,p\,(\,N = n) \\ &\le [\log(1+\epsilon) + \log n_o] + \sum_{n > \theta_2} n\,P\,(\,N = n). \end{aligned}$$

or, $$E\,[(\log N)/(\log n_0)] \le [1 + \{\log(1+\epsilon)/\log n_o\}] + (\log n_o)^{-1} . \sum_{n \ge \theta_2} (n+1)\,P\,(\,N = n+1\,).$$

Since ϵ is arbitrary, it follows that

$$\lim_{C \to 0} \sup E\,[(\log N)\,/\,(\log n_o)] \le 1. \qquad (2.10)$$

if it can be proved that, as $C \to 0$.

$$\sum_{n \geq \theta_2} (n+1)\, P\,(N = n+1) < K. \tag{2.11}$$

where K is free from C. To this end, using the fact that $\frac{p(n-1)\,\sigma_n^2}{\sigma^2} = V_n \sim X^2_{p(n-1)}$, we have

$$\sum_{n \geq \theta_2} (n+1)\, P\,(N = n+1 \leq \sum_{n \geq \theta_2} (n+1)\, P[n < \{A((p\,.\,s)\,.\,(s/2C)\}^{2/s}\, \sigma_n^2]$$

$$= \sum_{n \geq \theta_2} (n+1)\, P\,[V_n > p(n-1)\,.\,(n/n_o)].$$

Using exponential bounds [see Hogg and Craig [6], p.60] we get

$$\sum_{n \geq \theta_2} (n+1)\, P(N = n+1)$$

$$\leq \sum_{n \geq \theta_2} (n+1) \inf_{0<h<1/2} [\exp\{-hp(n-1)\,(n/n_o)\}\, E\,(e^{hV_n})]$$

$$\leq \sum_{n \geq \theta_2} (n+1) \inf_{0<h<1/2} [\exp\{-hp\,(n-1)\,(n/n_o)\}\ (1-2h)^{-p\,(n-1)/2}]\,.$$

The last inequality is also valid for the value $h_o \in (0, 1/2)$ of h, which minimizes the function

$$f(h) = \exp[-hp(n-1)\,(1+\in)]\,(1-2h)^{-p(n-1)/2},$$

i.e. $h_o = \left(\frac{1}{2}\right) [1-(1+\in)^{-1}]$, and substituting this value, we obtain

$$\sum_{n \geq \theta_2} (n+1)\, P(N = n+1) \leq \sum_{n \geq \theta_2} (n+1)\, [(1+\in) \exp\{1-(1+\in)\}]^{p(n-1)/2}$$

$$= \sum_{n \geq \theta_2} b_n \text{ (say).}$$

Since $\lim_{n \to \infty} b_n^{1/n} = [(1+\in) \exp\{1-(1+\in)\}]^{p/2} < 1$, (2.11) holds.

Result (2.7) now follows on combining (2.9) and (2.10).

By the definition of N,

$$N \geq [A(p, s)(s/2c)]^{2/s} \cdot \sigma_N^2$$

$$= n_o \; [p(N-1)]^{-1} \sum_{j=1}^{p(N-1)} Z_j,$$

with $Z_j \sim \chi^2_{(1)}$. Thus,

$$(n_o/N)^t \leq [\inf_N \{p(N-1)\}^{-1} \sum_{j=1}^{p(N-1)} z_j]^{-t}$$

$$= [\{p(m-1)\}^{-1} \sum_{j=1}^{p(m-1)} z_j]^{-t}$$

Applying Lemma 1, it is concluded that $E(n_o/N)^t < \infty$ if $p(m-1) > 2t$., i.e. $(n_o/N)^t$ is uniformly integrable for all $m > 1 + 2p^{-1} t$. Result (2.8) follows from (2.6) and the dominated convergvence theorem.

The following theorem establishes the asymptotic 'risk- efficiency' of the procedure (2.1).

Theorem 1. For all $m > 1 + p^{-1} s$,

$$\lim_{c \to 0} R_e(C) = 1.$$

Proof: From (1.4) and (2.2), substituting the values of $R_{n_o}(C)$ and $R_N(C)$ in (2.3), we obtain

$$R_e(C) = [(\log n_o)^{-1} + (2s^{-1})^{-1}]^{-1} \; [(\log n_o)^{-1} E\{n_o/N)^{s/2}\}$$

$$+ (2s^{-1})^{-1} \; E\left\{\frac{(\log N)}{(\log n_o)}\right\}]. \qquad (2.12)$$

The theorem is now in immediate consequence of (2.7) and (2.8).

Remark 1 : Letting p= 1 and $\Sigma = I = 1$, we get, $\lim_{C \to 0} R_e(C) = 1$ if $m > 1 + s$ and which is the result mentioned under corollary in Starr [8]

In the following theorem, we shall obtain second-order approximations for the 'regret' of the procedure (2.1).

Theorem 2 : For all $m > 1+2p^{-1}$, as $C \to 0$

$$R_g(C) = (C_s/2pn_o) + O(C^{2/s}).$$

Proof : Making substitutions from (1.4) and (2.2) in (2.4), we obtain

$$R_g(C) = CE \; f(N/n_o) - f(1), \tag{2.13}$$

where $f(x) = (2s^{-1})\, x^{-x/2} + \log X$. Denoting by K, any positive generic constant independent of C, we have from (2.13)

$$R_g(C)\, I\,(N \le n_o/2) \le KC\,[(n_o/m)^{s/2} + \{\log(n_o/2) - \log(n_o)\}]\, .\, P(N \le n_o/2)$$

$$\le KP\,(N \le n_o/2)\ . \tag{2.14}$$

Following the proof of Lemma 2 in Chaturvedi [2] , it can be shown that $P\,(N \le n_o/2) = O\,(C^{p(m-1)/s})$ as $C \to 0$. Hence, from (2.14),

$$R_g(C)\, I\,(N \le n_o/2) = O(C^{2/s}), \tag{2.15}$$

for all $m > 1+2p^{-1}$. Furthermore, using Taylor series expansion we obtain for $|W-1| \le |X-1|$

$$f(x) = f(1) + (x-1)\, f'(1) + \frac{(x-1)^2}{2!}\, f''(W).$$

Since

$$f'(x) = -x^{-(s/2+1)} + x^{-1}$$

and

$$f''(x) = (s/2+1)\, x^{-(s/2+2)} - x^{-2}$$

we obtain from (2.13) that , for $|W-1| \le |(N/n_o)-1|$.

$$R_g(C) = E[I_1 - I_2], \text{ say}, \tag{2.16}$$

where

$$I_1 = (C/2n_o^2)(s/2+1)\,[(N-n_o)^2\, W^{-(s/2+2)}] \tag{2.17}$$

and

$$I_2 = (C/2n_o^2)\,[(N-n_o)^2\, W^{-2}] \tag{2.18}$$

By the definition of 'W', on the event '$N > n_o/2$', $W \le 3/2$ and $W^{-1} \le 2$, that is, both the positive and negative powers of 'W' are bounded. Thus we conclude from (2.17) and (2.18) that

$$I_1 \leq KC^{(1+2/s)} \{(N-n_o)^2/n_o\} \tag{2.19}$$

and

$$I_2 \leq KC^{(1+2/s)} \{(N-n_o)^2/n_o\} \tag{2.20}$$

Form Theorem 2.4 of Woodroofe [13], $(N-n_o)^2/(n_o)$ is uniformly integrable for all $m > 1+2p^{-1}$. It now follows from (2.19) and (2.20) that for all $m > 1+2p^{-1}$, both I_1 and I_2 are unifromly integrable on the event '$N > n_o/2$'. It is easy to see that $W \xrightarrow{a.s.} 1$ as $C \to 0$ and from a result of Bhattacharya and Mallik [1], $(N-n_o)/(n_o)^{1/2} \xrightarrow{L} N\ (O,\ 2p^{-1})$ as $C \to 0$. Utilizing these results, we obtain from (2.16), (2.17) and (2.18) that, for all $m > 1+2p^{-1}$, as $C \to 0$,

$$\begin{aligned} R_g(C)\, I\,(N > n_o/2) &= (C/2n_o)(s/2+1)(2p^{-1}) - (C/2n_o)(2p^{-1}) \\ &= (C\,s/p\,n_o). \end{aligned} \tag{2.21}$$

The theorem now follows on combining (2.15) and (2.21).

Remark 2 : For $p = 1, \sum = I = 1, \lim_{C \to 0} R_g(C) = (Cs/2n_o) + O(C^{2/s})$,, which is the result obtained in Theorem 3 of Chaturvedi [3].

3. *The Three-Stage Procedure*

Choose and fix the number $r \in (0, 1)$. Start with a sample $\mathbf{X}_1, \ldots, \mathbf{X}_m$ of size $(m \geq 2)$. Denoting by $[q]^*$, the largest positive integer less than q, let us define

$$M = \max \{m, [r\, \{A(p, s)\,(s/2C)\}^{2/s}\ \sigma_m^2\,]^* + 1\}.$$

If $M > m$, extend the sample to $\mathbf{X}_1, \ldots, \mathbf{X}_m, \ldots, \mathbf{X}_M$. Let

$$N = \max \{M, [\, \{A(p, s)\,(s/2C)\}^{2/s}\ \sigma_M^2\,]^* + 1\} \tag{3.1}$$

and if $N > M$, take the difference to complete the sample $\mathbf{X}_1, \ldots, \mathbf{X}_M, \ldots, \mathbf{X}_N$. Estimate μ by $\overline{\mathbf{X}}_N$.

We first state a lemma, the proof of which can be obtained exactly along the lines of that of Theorem 1 in Hall [4]. We omit the details for brevity.

Lemma 3 : As $C \to 0$,

$$E(N) = n_o + (1/2 - r^{-1}) + 0(1),; \tag{3.2}$$

$$\text{Var}(N) = (2n_o/rp) + O(C^{-2/s}), \tag{3.3}$$

$$E\,|N - n_o|^3 = O(C^{-4/s}) \tag{3.4}$$

Corollary 1 : $\lim_{C \to 0} E(N/n_o) = 1.$

Proof : The proof follows from (3.2).

Lemma 4 :
$$\lim_{C \to 0} (N/n_o) = 1 \text{ a.s.}, \tag{3.5}$$

$$\lim_{C \to 0} E[(\log N)/(\log n_o)] = 1, \tag{3.6}$$

$$\lim_{C \to 0} E[(n_o/N)^t] = 1, \text{ if } m > 1 + 2p^{-1} t. \tag{3.7}$$

Proof : Result (3.5) is a direct consequence of the inequality

$$[A(p, s)(s/2C)]^{2/s}\, \sigma_M^2 \le N \le [A(p, s)(s/2C)]^{2/s} \cdot \sigma_M^2 + m$$

and the results $M \xrightarrow{\text{a.s.}} \infty$, $n_o \to \infty$ as $C \to 0$.

It is easy to see that $\lim_{C \to 0} [(\log N)/(\log n_o)] = 1$ a.s.

Thus, applying Fatou's lemma, we get

$$\lim_{C \to 0} \inf E[(\log N)/(\log n_o)] \ge 1. \tag{3.8}$$

Since $\log N$ is a concave function of N, using Jensen's inequality, we obtain

$$E[(\log N)/(\log n_o)] \le \log [E(N)]/(\log n_o),$$

which, on utilizing Corollary 1 gives

$$\lim_{C \to 0} \sup E[(\log N)/(\log n_o)] \le 1. \tag{3.9}$$

Result (3.6) now follows on combining (3.8) and (3.9).

From the definition of N,

$$N \ge M \ge r[A(p, s)(s/2C)]^{2/s}\, \sigma_m^2$$

$$= n_o [P(m-1)]^{-1} v_m,$$

with $V_m \sim \chi^2_{p(m-1)}$. Hence,

$$(n_o/N)^t \le [p(m-1)/r]^t \, V_m^{-t}.$$

We know that $E(V_m^{-t}) < \infty$ if $p(m-1) > 2t$, , i.e. $(n_o/N)^t$ is uniformlay integrable for all $m > 1 + 2p^{-1} t$. Result (3.7) now follows from (3.5) and the dominated convergence theorem.

It is to be noted that the risk associated with the three-stage procedure is of the same from as that given at (2.2). The only difference is that here N is determined by the rule (3.1).Thus, in what follows, as in section 2, we shall use the notations $R_e(C)$ and $R_g(C)$ to denote the 'risk-efficiency' and 'regret', respectively.

In the following theorem, we shall prove that the procedure (3.1) is asymptotically 'risk-efficient'.

Theorem 3 : For all $m > 1 + p^{-1} s$,

$$\lim_{C \to 0} R_e(C) = 1 .$$

Proof : Since the 'risk-efficiiiency' $R_e(C)$ of the procedure (3.1) can be put in the form (2.12), the theorem follows on applying the results (3.6) and (3.7).

In the next theorem, we shall study the asymptotic behaviour of 'regret' of the three-stage procedure (3.1).

Theorem 4 : For the procedure (3.1), as $C \to 0$,

$$R_g(C) = (Cs/2pm_o) + 0(C^{2/s}).$$

Proof : Utilizing Taylor's series expansion, we obtain

$$R_g(C) = 2s^{-1} C[E(n_o/N)^{s/2} - 1] + C[E(\log N) - (\log n_o)]$$

$$= (C/2n_o^2)(s/2) \, E[(N-n_o)^2] + \delta(C),$$

where the remainder term $\delta(C) = 0(n_o^{-3} E|N-n_o|^3)$. Now, applying Lemma 3, we get

$$R_g(C) = (CS/4n_o^2) \, [(2n_o/rp) + o(C^{2/s})] + 0(C^{2/s}),$$

and the theorem follows.

Remark 3 : Looking at the statements of Theorems 1 and 3, we note an interesting point that the initial sample sizes required to ensure the asymptotic 'risk-efficiency' remain same for both the purely sequential and three-stage procedure. Moreover, from Theorems 2 and 4 we arrive at the conclusion that letting 'r' closer to unity, we can make the 'regret' function of the three- stage procedure (3.1) closer to that of the purely sequential procedure (2.1). Thus, parallel to Hall [4], we conclude that in the point estimation case also, the three-stage procedure is strongly competitive with the purely sequential procedure.

ACKNOWLEDGEMENT

The authors are thankful to the referee for his comments.

REFERENCES

[1] Bhattacharya, and Mallik, A., 1973. Asymptotic normality of the stopping times of some sequential procedures. *Ann. Statist.* , **1**, 1203-1211.

[2] Chaturvedi, A, 1986. Further remarks on sequential point estimation of mean of a multinational population. *J. sequential Analysis*, **5**(3), 263-274.

[3] Chaturvedi, A., 1988. On sequential procedures for the point estimation of the mean of a nonnal population. *Ann. Inst. Statist. Math.*, **A40**, 769-783.

[4] Hall, P, 1981. Asymptotic theory of triple sampling for sequential estimation of a mean. *Ann. Statist.*, **9**, 1229-1238.

[5] Hayre, L.S., 1983. Sequential estimation of the differences between means of two nonnal populations. *Metrika*, **30**, 101-108.

[6] Hogg, R.V. and Craig, A.T., 1978. Introduction to Mathematical Statistics. The Macmillan Company, New York.

[7] Singh, R. Karan and Chaturvedi, A., 1988. Fixed size confidence region estimation of the mean vector. *Jour. Indian Soc. Ag. Statist*, **40**, 51-59.

[8] Starr, N., 1966. On the asymptotic efficiency of a sequential procedure for estimating the mean. *Ann. Math. Statist.*, **37**, 1173-1185.

[9] Starr, N. and Woodroofe, M., 1969. Remarks on sequential point estimation. *Proc. Nat. Acad. Sci.*, **63**, 285-288.

[10] Stein, C., 1945. A two sample test for a linear hypothesis whose power is independent of the variance. *Ann . Math. Statist.*, **16**, 243-258.

[11] Wald, A., 1947. Sequential analysis. John wiley and Jons Inc. , New York.

[12] Wang, Y.H., 1980. Sequential estimation of the mean of a multinormal population. *Jour. Amer. Statist. Assoc.*, **75**, 977-983.

[13] Woodroofe, M., 1977. Second order approximations for sequential point and interval estimation. *Ann. Statist.*, **5**, 984-995.