JOURNAL OF THE
CHUNGCHEONG MATHEMATICAL SOCIETY
Volume **26**, No. 1, 2013

SEQUENTIAL ESTIMATION OF THE MEAN VECTOR WITH BETA-PROTECTION IN THE MULTIVARIATE DISTRIBUTION

Sung Lai Kim*, Hae In Song**, Min Soo Kim***, and Yu Seon Jang****

Abstract. In the treatment of the sequential beta-protection procedure, we define the reasonable stopping time and investigate that for the stopping time Wijsman's requirements, coverage probability and beta-protection conditions, are satisfied in the estimation for the mean vector $\boldsymbol{\mu}$ by the sample from the multivariate normal distributed population with unknown mean vector $\boldsymbol{\mu}$ and a positive definite variance-covariance matrix $\boldsymbol{\Sigma}$.

1. Introduction

Most researchers use the statistical methods as the scientific analysis. The essential features of the hypothetic deductive view of scientific method are that a person observes or samples the natural world and uses all the information available to make an intuitive, logical guess, called an hypothesis, about how the system functions. The person has no way of knowing if their hypothesis is correct-it may or may not apply. Prediction made from the hypothesis are tested, either by further sampling or by doing experiments [7]. Sequential methods have been utilized in numerous disciplines, most notably in industrial process control and medical research involving clinical trials [1][3][4]. The sample size is controllable factor. In the test with the fixed width confidence interval, as the sample size increase, the significance level α is decrease, but the risk of Type 2 error, β, will be increase.

Received December 28, 2011; Accepted January 11, 2013.
2010 Mathematics Subject Classification: Primary 62H10, 62H12, 62H15.
Key words and phrases: sequential estimation, multivariate normal, coverage probability, beta-protection.
Correspondence should be addressed to Yu Seon Jang, ysjang@cnu.ac.kr.
This paper was supported by the research grant of Chungnam National University (Grant No. 2009-1668).

In the sequential test of the hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ based on a sample $X_1, X_2, \cdots, X_n$ from some population with the parameter $\theta \in \Theta \subset \mathbb{R}$, Wijsman [9] propose the following two requirements when n instead to some stopping time N; for any fixed α, β $(0 < \alpha, \beta < 1)$

$$P(\theta \in R_{c,n}) \geq 1 - \alpha \quad \text{for all } \theta \in \Theta \tag{1.1}$$

and

$$P\left\{\theta - \delta\left(I_{\{\theta > \theta_0\}}(\theta) - I_{\{\theta < \theta_0\}}(\theta)\right) \in R_{c,n}\right\} \leq \beta \quad \text{for all } \theta \in \Theta, \tag{1.2}$$

where $R_{c,n}$ is the confidence region, $I_A(\cdot)$ is the indicator function and δ is given imprecision function.

In the sampling from the normally distributed population with mean μ and the unknown variance σ^2, for given $0 < \alpha < 1$ and $d > 0$, taking the confidence interval $R_{c,n}$ for the mean μ with the width $2d$, the optimum fixed sample size is $n^* = (a^2 s_n^2)/d^2$, where $a = t_{\alpha/2,n-1}$ is defined by $P(|T| \leq t_{\alpha/2,n-1}) = 1 - \alpha$ for the Student's t-distribution random variable T with degree of freedom $n - 1$. The stopping time is given by

$$N = \min\left\{n \geq n_0 : n \geq a^2 s_n^2/d^2\right\}, \tag{1.3}$$

where s_n^2 is the sample variance which is an unbiased estimator of σ^2. In this time, for the significance level $\alpha \in (0, 1)$ the confidence interval for μ is $R_{c,n} = [\bar{X}_n - d, \bar{X}_n + d]$, where $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$ is the sample mean.

The precision of the confidence region $R_{c,n}$ will be controlled or partly controlled by requiring that with high probability the confidence set does not contain one or more specified parameter values different from the parameter μ of interest. Given an imprecision functions $\delta > 0$ and $0 < \beta < 1$, we have a probability

$$P\left(\mu - \delta \in R_{c,n}\right) \leq \beta. \tag{1.4}$$

This is called β-protection at $\mu - \delta$, and was first proposed as a measure of precision by Wijsman [5][9].

This method has been applied to a variety of problems, for the mean μ of a normal population with known variance, the μ/σ in the normal population [9][10], the mean of an exponential distribution, and the mean of a distribution in the presence of nuisance parameters [5] and sequential confidence sets with guaranteed coverage probability and beta-protection [2]. Specially, for one sided case $\mu > \mu_0$, in order to arrive at a sequential procedure which satisfies (1.1) and (1.2), as a confidence region $R_{c,n}$, replacing d by δ which is function of $\bar{X}_n$ in (1.3), Kim [5] showed that there is a fixed sample size confidence interval of the form $[\bar{X}_n - r\delta(\bar{X}_n), \infty)$ with $0 < r < 1$.

For $\mu < \mu_0$, in (1.4), if $\mu - \delta$ is replaced by $\mu + \delta$, then we have the beta-protection at $\mu + \delta$ such as following;

$$P\left(\mu + \delta(\mu) \in (-\infty, \bar{X}_n + r\delta(\bar{X})]\right) \leq \beta. \tag{1.5}$$

For the two sided situation this can be extended to the following;

$$P\left\{\mu - \delta(\mu)\left(I_{\{\mu>\mu_0\}} - I_{\{\mu<\mu_0\}}\right) \in \left[\bar{X}_n - r\delta, \bar{X}_n + r\delta\right]\right\} \leq \beta. \tag{1.6}$$

The sequential beta-protection procedure contains the Wijsman's requirements, (1.1) and (1.2), as well as the sequential estimation, such as confidence region for the parameter θ, given stopping time. In this paper we propose the new stopping time and we investigate the sequential beta-protection procedure in the sequential sampling from the multivariate normally distributed population.

2. Main theorem

Let $\{\mathbf{X}_k = (\mathbf{X}_{1k}, \mathbf{X}_{2k}, \cdots, \mathbf{X}_{pk}), k \geq 1\}$ be a sequence of multivariate normal random vectors with mean vector $\boldsymbol{\mu}$ and a positive definite variance-covariance matrix $\boldsymbol{\Sigma}$. For testing the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, a random sample of a fixed size n have recorded $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$. We write $\bar{\mathbf{X}}_n = n^{-1}\sum_{i=1}^n \bar{\mathbf{X}}_i$ and $\mathbf{S}_n = (n-1)^{-1}\sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}}_n)(\mathbf{X}_n - \bar{\mathbf{X}}_n)^t$ for $n \geq 2$. In order to find a confidence region $\mathbf{R}_{c,n}$ for the mean vector $\boldsymbol{\mu}$ such that (1.1) and (1.2) are achieved, given the positive real valued imprecision functions $\Delta_i : \mathbb{R}^p \to \mathbb{R}^+, 1 \leq i \leq p$, in the multivariate version, imprecision function $\Delta = (\Delta_1, \Delta_2, \cdots, \Delta_p)$ has to be satisfied several smoothness properties.

We begin with the following assumptions about the imprecision Δ:

(A1) Δ_i is continuously differentiable and has the bounded positive derivative for all $1 \leq i \leq p$

(A2) for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$, $\Delta_*(\mathbf{x})/\Delta_*(\mathbf{x}+\mathbf{y}) \to 1$ as $\mathbf{y} \to \mathbf{0}$ uniformly in $\mathbf{x}$, where $\Delta_* = \min\{\Delta_1, \Delta_2, \cdots, \Delta_p\}$ and $\mathbf{0} = (0, 0, \cdots, 0)$.

Let the length of the μ_i-axis of the region be bounded by $2d_i, i = 1, 2, \cdots, p$, given $\alpha \in (0,1)$, first we consider a confidence region $\mathbf{R}_{1,n} \subset$

$\mathbb{R}^p$ for mean vector $\boldsymbol{\mu}$ as following;

$$\mathbf{R}_{1,n} = \left\{\boldsymbol{\mu} \in \mathbb{R}^p : \mu_i - \Delta_i(\boldsymbol{\mu}) \geq \bar{X}_{in} - r\Delta_i(\bar{\mathbf{X}}_n),\ i = 1, \cdots, p\right\}, \tag{2.1}$$

where $\bar{X}_{in}$ and μ_i are i-th components of $\bar{\mathbf{X}}_n$ and $\boldsymbol{\mu}$ respectively.

Since the optimum fixed sample size n_p^* must be

$$n_p^* \geq a^2 s_{iin}/d_i^2, \text{ for all } 1 \leq i \leq p, \tag{2.2}$$

where $a > 0$ is to be chosen and s_{iin} is the (i,i) component of $\mathbf{S}_n$, replacing d_i by Δ_i, we propose the the sequential procedure in the estimation for the mean vector $\boldsymbol{\mu}$ from the initial sample $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_{n_0}$ of size $n_0(> p)$; stop according to the stopping time N_p defined by

$$N_p = \min\left\{n \geq n_0 : n \geq \frac{a^2\hat{\lambda}_{(p)}}{\Delta_*^2(\bar{\mathbf{X}}_n)}\right\}, \tag{2.3}$$

where $\hat{\lambda}_{(p)}$ is the largest eigenvalue of the sample variance covariance matrix $\mathbf{S}_n$ and $\Delta_* = \min\{\Delta_1, \Delta_2, \cdots, \Delta_n\}$. If $n < a^2\hat{\lambda}_{(p)}/\Delta_*^2(\bar{\mathbf{X}}_n)$, obtain the $(n+1)$ sampling and estimate $\boldsymbol{\mu}$ by $\bar{\mathbf{X}}_{n+1}$.

Let $\bar{\mathbf{Z}}_n = \mathbf{\Sigma}^{-1/2}\sum_{i=1}^n(\bar{\mathbf{X}}_i - \boldsymbol{\mu})$ and $\sigma_{(p)} = \max\{\sigma_1, \sigma_2, \cdots, \sigma_p\}$ for the (i,i) component σ_i^2 of the matrix $\mathbf{\Sigma}$. Then we have the following lemma.

LEMMA 2.1. *For any given $\varepsilon > 0$, if $\sigma_{(p)} \geq \varepsilon\Delta_*(\boldsymbol{\mu})$, then we have as $a \to \infty$*

(i) $N_p \to \infty$ *a.s*

(ii) $\hat{\lambda}_{(p)}/\sigma_{(p)}^2 \to 1$

(iii) $N_p^{1/2}\bar{\mathbf{Z}}_{N_p} \to \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ *in distribution and it is stochastically bounded*

(iv) $\Delta_*(\boldsymbol{\mu})/\Delta_*(\bar{\mathbf{Z}}_{N_p}) \to 1$ *a.s. uniformly in* $\boldsymbol{\mu}$

where $\mathbf{I}_p$ is the p-dimensional identity matrix.

Proof. Parts (i) and (ii) are immediate. For part (iii), from the multivariate version of Woodroofe's result we know that $\left\{n^{1/2}\bar{\mathbf{Z}}_n : n \geq 1\right\}$ is uniformly continuous in probability and $n^{1/2}\bar{\mathbf{Z}}_n \to \mathcal{N}(0, \mathbf{I}_p)$ by Anscombe's theorem. Since $N_p \to \infty$ as $a \to \infty$ if $\sigma_{(p)} \geq \varepsilon\Delta_*(\boldsymbol{\mu})$ for any given $\varepsilon > 0$, thus we have $N_p^{1/2}\bar{\mathbf{Z}}_{N_p} \to \mathcal{N}(0, \mathbf{I}_p)$. Now part (iv), write $\bar{\mathbf{X}}_n = \boldsymbol{\mu} + \mathbf{\Sigma}^{1/2}\bar{\mathbf{Z}}_n$. Since $\mathbf{\Sigma}^{1/2}\bar{\mathbf{Z}}_n \to 0$ *a.s.* as $n \to \infty$ by strong law of large number, it follows from (i) that $\mathbf{\Sigma}^{1/2}\bar{\mathbf{Z}}_{N_p} \to 0$ a.s. as $a \to \infty$. Then we have the result of (iv) by the assumption (A2). □

THEOREM 2.2. *For the stopping time N_p which was defined in* (2.3), *given* $0 < \alpha, \beta < 1$, *under the assumption in Lemma 2.1 we have that for all* $\boldsymbol{\mu}$

$$P\left(\boldsymbol{\mu} \in \mathbf{R}_{1,N_p}\right) \geq 1 - \alpha \tag{2.4}$$

and

$$P\left(\boldsymbol{\mu} - \Delta(\boldsymbol{\mu}) \in \mathbf{R}_{1,N_p}\right) \leq \beta. \tag{2.5}$$

Proof. Let $\bar{\mathbf{X}}_n = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\bar{\mathbf{Z}}_n$. From the assumption (A1) we have $\Delta(\bar{\mathbf{X}}_n) \approx \Delta(\boldsymbol{\mu}) + D\Sigma^{1/2}\bar{\mathbf{Z}}_n$ where $D = ((d_{ij}))_{p\times p}$ with $d_{ij} = (\partial\Delta_i/\partial x_j)(\boldsymbol{\xi}_n)$ in which $\boldsymbol{\xi}_n$ is on the line segment from $\boldsymbol{\mu}$ to $\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\bar{\mathbf{Z}}_n$. For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$, considering that $\mathbf{x} < \mathbf{y}$ means as $x_i < y_i$ for all $i(1 \leq i \leq p)$, where x_i and y_i are i-th components of $\mathbf{x}$ and $\mathbf{y}$ respectively, then for any $0 < r < 1$

$$\begin{aligned} P\left(\boldsymbol{\mu} \in \mathbf{R}_{1,n}\right) &= P\left\{\boldsymbol{\mu} \geq \bar{\mathbf{X}}_n - r\Delta(\bar{\mathbf{X}}_n)\right\} \\ &= P\left\{(I - rD)(\boldsymbol{\Sigma}^{1/2}/\sigma_{(p)})\bar{\mathbf{Z}}_n \leq r\Delta(\boldsymbol{\mu})/\sigma_{(p)}\right\} \end{aligned}$$

and

$$\begin{aligned} &1 - P\left(\boldsymbol{\mu} - \Delta(\boldsymbol{\mu}) \in \mathbf{R}_{1,n}\right) \\ &\quad = P\left\{\boldsymbol{\mu} - \Delta(\boldsymbol{\mu}) \leq \bar{\mathbf{X}}_n - r\Delta(\bar{\mathbf{X}}_n)\right\} \\ &\quad = P\left\{(I - rD)(\boldsymbol{\Sigma}^{1/2}/\sigma_{(p)})\bar{\mathbf{Z}}_n \geq -(1 - r)\Delta(\boldsymbol{\mu})/\sigma_{(p)}\right\}. \end{aligned}$$

Put

$$M = (I - rD)\Sigma^{1/2}\sigma_{(p)} = (m_{ij})_{p\times p}. \tag{2.6}$$

Then m_{ij} is bounded for all $1 \leq i, j \leq p$. Since $\bar{\mathbf{Z}}_n \to 0$ a.s. uniformly in $\boldsymbol{\mu}$ as $n \to \infty$ and M is bounded, for any given $\varepsilon > 0$, there exists $b_1 > 0$ and $b_2 > 0$ such that both

$$P_\theta(M\bar{\mathbf{Z}}_n \leq b_1\mathbf{1}, n = 1, 2, \cdots) > 1 - \alpha \tag{2.7}$$

and

$$P_\theta(M\bar{\mathbf{Z}}_n \geq -b_2\mathbf{1}, n = 1, 2, \cdots) > 1 - \beta, \tag{2.8}$$

where $1 = (1, 1, \cdots, 1)' \in \mathbb{R}^p$. Thus no matter what N_p is the stopping time we have

$$P_\theta(M\bar{\mathbf{Z}}_{N_p} \leq b_1\mathbf{1}) > 1 - \alpha \tag{2.9}$$

and

$$P_\theta(M\bar{\mathbf{Z}}_{N_p} \geq -b_2\mathbf{1}) > 1 - \beta. \tag{2.10}$$

From Lemma 2.1 we known that $MN_p^{1/2}\bar{\mathbf{Z}}_{N_p}$ is stochastically bounded as $a \to \infty$ if $\sigma_{(p)} \geq \varepsilon\Delta(\boldsymbol{\mu})$. Since $N_p^{1/2}\Delta(\boldsymbol{\mu})/\sigma_{(p)} \to \infty$ a.s. as $a \to \infty$ if $\sigma_{(p)} \geq \varepsilon\Delta_*(\boldsymbol{\mu})$, taking $b_1 = r/\varepsilon$ and $b_2 = (1-r)/\varepsilon$, the proof is complete. □

Now we consider the confidence region as following; for $0 < r < 1$

$$\mathbf{R}_{2,n} = \left\{\boldsymbol{\mu} \in \mathbb{R}^p : \mu_i + \Delta_i(\boldsymbol{\mu}) \leq \bar{X}_{in} + r\Delta_i(\bar{\mathbf{X}}_n),\ i = 1, \cdots, p\right\}, \tag{2.11}$$

where $\bar{X}_{in}$ and μ_i are i-th components of $\bar{\mathbf{X}}_n$ and $\boldsymbol{\mu}$ respectively.

An argument similar to that employed in the proof of Theorem 2.2 yields the multivariate version Wijsman's requirements.

Corollary 2.3. *Under the assumption in Theorem 2.2 we have that for all* $\boldsymbol{\mu}$

$$P\left(\boldsymbol{\mu} \in \mathbf{R}_{2,N_p}\right) \geq 1 - \alpha \tag{2.12}$$

and

$$P\left(\boldsymbol{\mu} + \Delta(\boldsymbol{\mu}) \in \mathbf{R}_{2,N_p}\right) \leq \beta. \tag{2.13}$$

Finally, letting $\mathbf{R}_{c,n} = \mathbf{R}_{1,n} \cap \mathbf{R}_{2,n}$, we have the confidence interval of the two sided

$$\mathbf{R}_{c,n} = \left\{\boldsymbol{\mu} \in \mathbb{R}^p : \left|\bar{X}_{in} - \mu_i\right| + \Delta_i(\boldsymbol{\mu}) \leq r\Delta_i(\bar{\mathbf{X}}_n)\ (1 \leq i \leq p)\right\}, \tag{2.14}$$

where $\bar{X}_{in}$ and μ_i are i-th components of $\bar{\mathbf{X}}$ and $\boldsymbol{\mu}$ respectively. In the choice of a, taking $t_{\alpha,n-1}$ by $t_{\alpha/2,n-1}$, the corollary follows immediately from Theorem 2.2 and Corollary 2.3.

Corollary 2.4. *Under the assumption in Theorem 2.2 we have that for all* $\boldsymbol{\mu}$

$$P\left(\boldsymbol{\mu} \in \mathbf{R}_{c,N_p}\right) \geq 1 - \alpha \tag{2.15}$$

and

$$P\left\{\boldsymbol{\mu} - \Delta(\boldsymbol{\mu})\left(I_{\mathbf{R}_{1,N_p}}(\boldsymbol{\mu}) - I_{\mathbf{R}_{2,N_p}}(\boldsymbol{\mu})\right) \in \mathbf{R}_{c,N_p}\right\} \leq \beta \tag{2.16}$$

where $I_A(\cdot)$ *is the indicator function.*

References

[1] D. Amengual, G. Fiorentini and E. Sentana, *Sequential estimation of shape parameters in multivariate dynamic models*, CEMFI Working Paper 1201, 2012.

[2] I. Fakhre-Zkkeri, Sequential confidence sets with guranteed coverage probability and beta-protection, *J. Multivariate Anal.* **33** (1990), 89-105.

[3] M. Ghosh, N. Mukhopadhyay and P. K. Sen, *Sequential Estimation*, John Wiley & Sons, 1997.

[4] B. K. Ghosh and P. K. Sen, *Handbook of Sequential Analysis*, Marcel Dekker, 1991.

[5] S. L. Kim, *Sequential confidence sets with β-protection in the presence of nuisance parameters*, Ph.D. thesis, University of Illinois at Urbana-Champaign, 1987.

[6] A. Liu and W. J. Hall, Unbiased estimation of secondary parameters following a sequential test, *Biometrika* **88** (2001), 895-900.

[7] S. Mckillup, *Statistics Explained: An introductory guide for life scientists*, Cambridge university press, 7th ed., 2011.

[8] N. Mukhopadhyay and S. Liberman, Sequential estimation of a linear function of mean vectors, *Sequential Anal.* **8** (1989), 381-394.

[9] R. A. Wijsman, Confidence sets based on sequential tests, *Common. Statist. Theor. Method A* **10** (1981), 2137-2147.

[10] R. A. Wijsman, Sequential confidence sets: estimation-oriented versus test oriented construction, *Statistical Decision Thoery and Related Topocs III* **2** (1982), 435-450.

[11] R. A. Wijsman, Sequential confidence intervals with beta protection in one-parameter families, *Lecture Notes–Monograph Series* **8** (1986), 96-107.

*
Department of Mathematics
Chungnam National University
Daejon 305-764, Republic of Korea
E-mail: slkim@cnu.ac.kr

**
Department of Mathematics
Chungnam National University
Daejeon 305-764, Republic of Korea
E-mail: ttlove17@naver.com

***
Department of Mathematics
Chungnam National University
Daejeon 305-764, Republic of Korea
E-mail: strikers@cnu.ac.kr

****
Department of Mathematics
Chungnam National University
Daejeon 305-764, Republic of Korea
E-mail: ysjang@cnu.ac.kr