Ann. Inst. Statist. Math.
Vol. 48, No. 3, 423–428 (1996)

THE MAXIMUM LIKELIHOOD ESTIMATORS IN A MULTIVARIATE NORMAL DISTRIBUTION WITH AR(1) COVARIANCE STRUCTURE FOR MONOTONE DATA

HIRONORI FUJISAWA

Department of Mathematics, Faculty of Science, Hiroshima University, Kagamiyama 1-3, Higashi-Hiroshima 739, Japan

(Received March 16, 1995; revised August 7, 1995)

Abstract. The maximum likelihood estimators are uniquely obtained in a multivariate normal distribution with AR(1) covariance structure for monotone data. The maximum likelihood estimator of mean is unbiased.

Key words and phrases: AR(1) covariance structure, conditional distribution, maximum likelihood estimator, missing data, monotone data, multivariate normal distribution.

1. Introduction

There are many studies about statistical inference based on missing data. Monotone data is a special type of missing data. It is expressed as the $i \times 1$ observation vector $\boldsymbol{x}_{ij}$ for $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$, especially for the case $k = 3$ as

$$\begin{array}{ccc} x_{11}^{(1)} \cdots x_{1n_1}^{(1)} & x_{21}^{(1)} \cdots x_{2n_2}^{(1)} & x_{31}^{(1)} \cdots x_{3n_3}^{(1)} \\ & x_{21}^{(2)} \cdots x_{2n_2}^{(2)} & x_{31}^{(2)} \cdots x_{3n_3}^{(2)} \\ & & x_{31}^{(3)} \cdots x_{3n_3}^{(3)} \end{array},$$

where $x_{ij}^{(l)}$ is the l-th component of $\boldsymbol{x}_{ij}$. This type of data appears in various situations.

Assume that monotone data are independently distributed as $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, in other words, $\boldsymbol{x}_{ij}$'s are mutually independent and $\boldsymbol{x}_{ij}$ is distributed as $\mathrm{N}_i(\boldsymbol{\mu}_i, \Sigma_i)$ where $\boldsymbol{\mu}_i$ is the first i components of $\boldsymbol{\mu}$ and Σ_i is the first $i \times i$ matrix of Σ. The maximum likelihood estimators of $\boldsymbol{\mu}$ and Σ were investigated by Anderson (1957), Bhargava (1975), Jinadasa and Tracy (1992) and Fujisawa (1995).

It is sometimes postulated that Σ has a structure given by $\Sigma = \sigma^2(\rho^{|i-j|})$, which is called AR(1) covariance structure. An example is the case when data is longitudinal. In this paper the maximum likelihood estimators of the parameters $\boldsymbol{\mu}$, σ^2 and ρ are obtained uniquely. The maximum likelihood estimator of $\boldsymbol{\mu}$ is unbiased.

The case $k = 2$ was considered by Dahiya and Korwar (1980). They proved that the likelihood equation for ρ has a unique root which always provides the maximum likelihood estimator of ρ. We extend this result to a general monotone data. The case $k = 2$ for general missing data was considered by Konishi and Shimizu (1994). They obtained the asymptotic normality of the maximum likelihood estimators and pointed out, based on a simulation study, that the likelihood equations had no unique solutions.

2. Maximum likelihood estimators

First we prepare some notations. For each l, let the subvectors of a vector $\boldsymbol{x}$ with the first l components and l-th component be denoted by

$$[\boldsymbol{x}]_l = (x_1 \cdots x_l)', \qquad [\boldsymbol{x}]^l = x_l,$$

and let the first $l \times l$ submatrix of a matrix A as $[A]_l$, and partion a $l \times l$ matrix A be denoted by

$$A = \begin{pmatrix} [A]_{11} & [A]_{12} \\ [A]_{21} & [A]_{22} \end{pmatrix}, \qquad [A]_{11} \text{ is the } (l-1) \times (l-1) \text{ matrix.}$$

Let the (m, n) component of a matrix A be denoted by $[A]_{(m,n)}$. Then $[\Sigma]_l = \Sigma_l$, $[\Sigma_l]_{11} = \Sigma_{l-1}$ and $[\Sigma_l]_{22} = [\Sigma]_{(ll)}$. Let $N_l = \sum_{i=l}^k n_i$,

$$\boldsymbol{m}_l = \frac{1}{N_l} \sum_{i=l}^{k} \sum_{j=1}^{n_i} [\boldsymbol{x}_{ij}]_l, \qquad Q_l = \frac{1}{N_l} \sum_{i=l}^{k} \sum_{j=1}^{n_i} ([\boldsymbol{x}_{ij}]_l - \boldsymbol{m}_l)([\boldsymbol{x}_{ij}]_l - \boldsymbol{m}_l)'.$$

Note that $\boldsymbol{m}_l$ and Q_l are natural estimators of $\boldsymbol{\mu}_l$ and Σ_l. Let $\Delta_{11} = \Sigma_1$, $\Delta_{l-1,l} = [\Sigma_l]_{11}^{-1}[\Sigma_l]_{12}$, $\Delta_{ll} = [\Sigma_l]_{22} - [\Sigma_l]_{21}[\Sigma_l]_{11}^{-1}[\Sigma_l]_{12}$, and let a symmetric positive definite matrix Δ defined by

$$\Delta = \begin{pmatrix} \Delta_{11} & \Delta_{12} & & & \\ & \Delta_{22} & \Delta_{23} & & \Delta_{k-1,k} \\ & & \Delta_{33} & & \\ & * & & \ddots & \\ & & & & \Delta_{k,k} \end{pmatrix}.$$

It may be noted that the correspondence between Δ and Σ is one-to-one and onto. Then we have for $l = 2, \ldots, k$,

$$\Delta_{11} = \sigma^2, \qquad \Delta_{ll} = \sigma^2(1 - \rho^2), \qquad \Delta_{l-1,l} = (0, \ldots, 0, \rho)'. \tag{2.1}$$

Let $f(\cdot)$ be a probability function and $f(\cdot \mid \cdot)$ be a conditional probability function. Noting that $[\boldsymbol{x}_{ij}]_l = ([\boldsymbol{x}_{ij}]_{l-1}'[\boldsymbol{x}_{ij}]^l)'$ for $l = 2, \ldots, i$, we can write the likelihood function of $\boldsymbol{x}_{ij}$'s as

$$\begin{aligned} \prod_{i=1}^{k} \prod_{j=1}^{n_i} f(\boldsymbol{x}_{ij}) &= \prod_{i=1}^{k} \prod_{j=1}^{n_i} f([\boldsymbol{x}_{ij}]_1) f([\boldsymbol{x}_{ij}]^2 \mid [\boldsymbol{x}_{ij}]_1) \cdots f([\boldsymbol{x}_{ij}]^i \mid [\boldsymbol{x}_{ij}]_{i-1}) \\ &= \prod_{i=1}^{k} \prod_{j=1}^{n_i} f([\boldsymbol{x}_{ij}]_1) \prod_{l=2}^{k} \prod_{i=l}^{k} \prod_{j=1}^{n_i} f([\boldsymbol{x}_{ij}]^l \mid [\boldsymbol{x}_{ij}]_{l-1}). \end{aligned}$$

The conditional distribution of $[\boldsymbol{x}_{ij}]^l$ given $[\boldsymbol{x}_{ij}]_{l-1}$ is normal with mean $\boldsymbol{\nu}_{ijl} = [\boldsymbol{\mu}]^l + \Delta'_{l-1,l}([\boldsymbol{x}_{ij}]_{l-1} - [\boldsymbol{\mu}]_{l-1})$ and variance Δ_{ll}. Thus we can obtain the log-likelihood $l(\boldsymbol{\mu}, \sigma^2, \rho)$, given by

$$\begin{aligned} l = \text{const} &- \frac{\tilde{N}_1}{2}\log\sigma^2 - \frac{\tilde{N}_2}{2}\log(1-\rho^2) - \frac{1}{2\sigma^2}N_1\tilde{\mu}_1^2 - \frac{1}{2\sigma^2(1-\rho^2)}\sum_{l=2}^{k} N_l\tilde{\mu}_l^2 \\ &- \frac{1}{2\sigma^2}\left\{N_1Q_1 + \frac{1}{1-\rho^2}(S^* + \rho^2 S - 2\rho S_{12})\right\}, \end{aligned}$$

where $\tilde{N}_1 = \sum_{l=1}^{k} N_l$, $\tilde{N}_2 = \sum_{l=2}^{k} N_l$, $\tilde{\mu}_1 = m_1 - \mu_1$, $\tilde{\mu}_l = [\boldsymbol{m}_l]^l - [\boldsymbol{\mu}]^l - \Delta'_{l-1,l}([\boldsymbol{m}_l]_{l-1} - [\boldsymbol{\mu}]_{l-1})$ for $l = 2, \dots, k$, and

$$S^* = \sum_{l=2}^{k} N_l[Q_l]_{(ll)}, \qquad S = \sum_{l=2}^{k} N_l[Q_l]_{(l-1,l-1)}, \qquad S_{12} = \sum_{l=2}^{k} N_l[Q_l]_{(l-1,l)}.$$

Since the likelihood l approaches 0 when some parameters approach the boundary of the regions, the maximum likelihood estimators $\hat{\boldsymbol{\mu}}$, $\hat{\sigma}^2$ and $\hat{\rho}$ are obtained as extremum values of the likelihood l. Simplifying the likelihood equations, it follows that the maximum likelihood estimators satisfy

$$[\hat{\boldsymbol{\mu}}]^1 = \boldsymbol{m}_1, \qquad [\hat{\boldsymbol{\mu}}]^l = [\boldsymbol{m}_l]^l + \hat{\rho}([\hat{\boldsymbol{\mu}}]^{l-1} - [\boldsymbol{m}_l]^{l-1}), \tag{2.2}$$

$$\hat{\rho} = S_{12}/\{N_1\hat{\sigma}^2 - N_1Q_1 + S\}, \tag{2.3}$$

$$\hat{\sigma}^2 = \left\{N_1Q_1 + \frac{1}{1-\hat{\rho}^2}(S^* + \hat{\rho}^2 S - 2\hat{\rho}S_{12})\right\}/\tilde{N}_1. \tag{2.4}$$

Eliminating $\hat{\sigma}^2$ from (2.3) and (2.4), we get a cubic equation in $\hat{\rho}$, given by

$$\begin{aligned} \tau(\hat{\rho}) = {} & \tilde{N}_2(N_1Q_1 - S)\hat{\rho}^3 + (\tilde{N}_2 - N_1)S_{12}\hat{\rho}^2 \\ & + \{N_1(S^* + S) - \tilde{N}_2(N_1Q_1 - S)\}\hat{\rho} - \tilde{N}_1S_{12} = 0. \end{aligned} \tag{2.5}$$

Dahiya and Korwar (1980) considered the case $k = 2$ and proved the uniqueness of the root of the equation $\tau(\rho) = 0$ on $(-1, 1)$. It may be noted that $N_1Q_1 - S > 0$ and $\tilde{N}_2 - N_1 < 0$ for the case $k = 2$. However, in this case the sign of the coefficients of the equation $\tau(\rho) = 0$ are unknown. We obtain the following theorem.

THEOREM 2.1. *Let $\hat{\boldsymbol{\mu}}$, $\hat{\sigma}^2$, $\hat{\rho}$ be the maximum likelihood estimators of the parameters $\boldsymbol{\mu}$, σ^2, ρ. Then $\hat{\rho}$ is* (i) *the unique root of $\tau(\rho) = 0$ on $(0, 1)$ when $S_{12} > 0$,* (ii) *zero when $S_{12} = 0$,* (iii) *the unique root of $\tau(\rho) = 0$ on $(-1, 0)$ when $S_{12} < 0$, and*

$$[\hat{\boldsymbol{\mu}}]^1 = \boldsymbol{m}_1, \qquad [\hat{\boldsymbol{\mu}}]^l = [\boldsymbol{m}_l]^l + \hat{\rho}([\hat{\boldsymbol{\mu}}]^{l-1} - [\boldsymbol{m}_l]^{l-1}),$$

$$\hat{\sigma}^2 = \left\{N_1Q_1 + \frac{1}{1-\hat{\rho}^2}(S^* + \hat{\rho}^2 S - 2\hat{\rho}S_{12})\right\}/\tilde{N}_1.$$

PROOF. If we prove that $\hat{\rho}$ is unique, then the proof is complete because of the formulas (2.2) and (2.4). When $S_{12} = 0$ we have $\hat{\rho} = 0$ from (2.3). The case $S_{12} > 0$ is now considered. The case $S_{12} < 0$ can be discussed in a similar way. The proof consists of three steps.

First we prove that there exists $\hat{\rho}$ on $(0, 1)$. The root of the equation $\tau(\rho) = 0$ is non-zero because $\tau(0) = -\tilde{N}_1 S_{12} \neq 0$. So, $\hat{\rho}$ is non-zero. For a fixed ρ, the maximum likelihood estimators $\hat{\boldsymbol{\mu}}(\rho)$ and $\hat{\sigma}^2(\rho)$ of $\boldsymbol{\mu}$ and σ^2 are obtained from (2.2) and (2.4). Let

$$\begin{aligned} v(\rho) &= l(\hat{\boldsymbol{\mu}}(\rho), \hat{\sigma}^2(\rho), \rho) \\ &= \text{const} - \frac{\tilde{N}_2}{2} \log(1 - \rho^2) \\ &\quad - \frac{\tilde{N}_1}{2} \log \left\{ N_1 Q_1 - S + \frac{1}{1 - \rho^2}(S^* + S - 2\rho S_{12}) \right\}. \end{aligned}$$

Then we can easily show that $v(\rho) > v(-\rho)$ for $\rho > 0$. So, there exists $\hat{\rho}$ on $(0, 1)$.

Now we show that the equation $\tau(\rho) = 0$ has the unique root on $(0, 1)$. We have

$$\begin{aligned} \tau(-1) &= -N_1(S + S^* + 2S_{12}) < 0, \\ \tau(0) &= -\tilde{N}_1 S_{12} < 0, \\ \tau(1) &= N_1(S + S^* - 2S_{12}) > 0. \end{aligned}$$

Hence, the equation $\tau(\rho) = 0$ has at least one root on $(0, 1)$. Let $\hat{\rho}_1$, $\hat{\rho}_2$, $\hat{\rho}_3$ be the non-zero roots of $\tau(\rho) = 0$. Then we have

$$\frac{\hat{\rho}_1 + \hat{\rho}_2 + \hat{\rho}_3}{\hat{\rho}_1 \hat{\rho}_2 \hat{\rho}_3} = \frac{(\tilde{N}_2 - N_1) S_{12}}{\tilde{N}_1 S_{12}} = \frac{\tilde{N}_2 - N_1}{\tilde{N}_2 + N_1} < 1.$$

Now let $0 < \hat{\rho}_i < 1$ for $i = 1, 2, 3$. Then we can obtain

$$3(\hat{\rho}_1 \hat{\rho}_2 \hat{\rho}_3)^{1/3} \leq \hat{\rho}_1 + \hat{\rho}_2 + \hat{\rho}_3 < \hat{\rho}_1 \hat{\rho}_2 \hat{\rho}_3.$$

So, we have $3 < (\hat{\rho}_1 \hat{\rho}_2 \hat{\rho}_3)^{2/3} < 1$. This is a contradiction. Now let $0 < \hat{\rho}_i < 1$ for $i = 1, 2$ without loss of generality. Then we have $N_1 Q_1 - S > 0$ and $0 < \hat{\rho}_3 < 1$ because $\tau(0) < 0$, $\tau(1) > 0$ and the equation $\tau(\rho) = 0$ is cubic. This is a contradiction. The result is proved.

Finally, we show that $\hat{\rho}$ is the unique root of the equation $\tau(\rho) = 0$ on $(0, 1)$. From the first step there exists $\hat{\rho}$, which is a root of the equation $\tau(\rho) = 0$, on $(0, 1)$. From the second step the equation $\tau(\rho) = 0$ has the unique root on $(0, 1)$. Therefore $\hat{\rho}$ is the unique root of $\tau(\rho) = 0$ on $(0, 1)$. □

The unbiasedness of the maximum likelihood estimator $\hat{\boldsymbol{\mu}}$ is obtained in the following theorem.

THEOREM 2.2. *The maximum likelihood estimator $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ is unbiased.*

PROOF. From Theorem 2.1 we have

$$\hat{\boldsymbol{\mu}} = \sum_{i=1}^{k} f_i \boldsymbol{d}_i \tag{2.6}$$

where $f_i = T_k \cdots T_i$, $T_k = I_k$, $T_j = (I_j\ \hat{\Delta}_{j,j+1})'$, $\hat{\Delta}_{j,j+1} = (0 \cdots 0\ \hat{\rho})'$ for $j = 1, \dots, k-1$, $\boldsymbol{d}_k = \boldsymbol{m}_k$ and $\boldsymbol{d}_i = \boldsymbol{m}_i - [\boldsymbol{m}_{i+1}]_i$ for $i = 1, \dots, k-1$. Because $\mathrm{E}(f_k \boldsymbol{d}_k) = \mathrm{E}(\boldsymbol{m}_k) = \boldsymbol{\mu}$, the proof is enough to show that $\mathrm{E}(\sum_{i=1}^{k-1} f_i \boldsymbol{d}_i) = \boldsymbol{0}$. Note that $\hat{\rho}$ is a function of Q_l's, and so is f_i for $i = 1, \dots, k-1$. Without loss of generality, let $\boldsymbol{\mu} = \boldsymbol{0}$ on the following discussions.

Let $\boldsymbol{m}_l^* = \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}/n_i$, $Q_i^* = \sum_{j=1}^{n_i} (\boldsymbol{x}_{ij} - \boldsymbol{m}_i^*)(\boldsymbol{x}_{ij} - \boldsymbol{m}_i^*)'/n_i$. We have

$$\boldsymbol{m}_l = \frac{1}{N_l} \sum_{i=l}^{k} n_i [\boldsymbol{m}_i^*]_l,$$

$$Q_l = \frac{1}{N_l} \sum_{i=l}^{k} n_i [Q_i^*]_l + \frac{1}{N_l} \sum_{i=l}^{k} n_i ([\boldsymbol{m}_i^*]_l - \boldsymbol{m}_l)([\boldsymbol{m}_i^*]_l - \boldsymbol{m}_l)'.$$

Note that Q_l is a function of Q_i^*'s and $[\boldsymbol{m}_1^*]^1 [\boldsymbol{m}_i^*]^l$'s because of

$$[\boldsymbol{m}_{i_1}^*]^{l_1} [\boldsymbol{m}_{i_2}^*]^{l_2} = [\boldsymbol{m}_1^*]^1 [\boldsymbol{m}_{i_1}^*]^{l_1} [\boldsymbol{m}_1^*]^1 [\boldsymbol{m}_{i_2}^*]^{l_2} / ([\boldsymbol{m}_1^*]^1)^2.$$

So is f_i because $\hat{\rho}$ is a function of Q_l's. We have

$$\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{k-1} f_i \boldsymbol{d}_i \right) &= \mathrm{E}\left(\mathrm{E}\left\{ \sum_{i=1}^{k-1} f_i \boldsymbol{d}_i \mid Q_i^*\text{'s and } [\boldsymbol{m}_1^*]^1 [\boldsymbol{m}_i^*]^l\text{'s} \right\} \right) \\
&= \mathrm{E}\left(\sum_{i=1}^{k-1} f_i \mathrm{E}\{ \boldsymbol{d}_i \mid Q_i^*\text{'s and } [\boldsymbol{m}_1^*]^1 [\boldsymbol{m}_i^*]^l\text{'s}\} \right) \\
&= \mathrm{E}\left(\sum_{i=1}^{k-1} f_i \mathrm{E}\{ \boldsymbol{d}_i \mid [\boldsymbol{m}_1^*]^1 [\boldsymbol{m}_i^*]^l\text{'s}\} \right).
\end{aligned}$$

Let $\boldsymbol{v} = ([\boldsymbol{m}_1^*]^1 [\boldsymbol{m}_2^*]^1 [\boldsymbol{m}_2^*]^2 [\boldsymbol{m}_3^*]^1 \cdots [\boldsymbol{m}_k^*]^k)' = (v_1 \cdots v_{k(k+1)/2})'$. Then $\boldsymbol{v}$ is normally distributed with mean zero. The set $V = \{\boldsymbol{v} \mid v_1 v_i = t_i, i = 1, \dots, k(k+1)/2\}$ is symmetric. In other words, if $\boldsymbol{v} \in V$, then $-\boldsymbol{v} \in V$. Therefore we have

$$\mathrm{E}\{\boldsymbol{v} \mid v_1 v_i = t_i, i = 1, \dots, k(k+1)/2\} = \boldsymbol{0}.$$

So,

$$\mathrm{E}\{\boldsymbol{d}_i \mid [\boldsymbol{m}_1^*]^1 [\boldsymbol{m}_i^*]^l\text{'s}\} = \boldsymbol{0}.$$

The proof is complete. □

Acknowledgements

I would like to thank Professor Y. Fujikoshi for his helps and referees for their comments and suggestions.

References

Anderson, T. W. (1957). Maximum likelihood estimators for a multivariate normal distribution when some observations are missing, *J. Amer. Statist. Assoc.*, **52**, 200–203.

Anderson, T. W. and Olkin, I. (1985). Maximum-likelihood estimation of the parameters of a multivariate normal distribution, *Linear Algebra Appl.*, **70**, 147–171.

Bhargava, R. P. (1975). Some one-sample hypothesis testing problems when there is a monotone sample from a multivariate normal population, *Ann. Inst. Statist. Math.*, **27**, 327–339.

Dahiya, R. C. and Korwar, R. M. (1980). Maximum likelihood estimators for a bivariate normal distribution with missing data, *Ann. Statist.*, **8**, 687–692.

Fujisawa, H. (1995). A note on the maximum likelihood estimators for multivariate normal distribution with monotone data, *Comm. Statist. Theory Methods*, **24**, 1377–1382.

Jinadasa, K. G. and Tracy, D. S. (1992). Maximum likelihood estimation for multivariate normal distribution with monotone sample, *Comm. Statist. Theory Methods*, **21**, 41–50.

Konishi, S. and Shimizu, K. (1994). Maximum likelihood estimation of an intraclass correlation in a bivariate normal distribution with missing observations, *Comm. Statist. Theory Methods*, **23**, 1593–1604.