International Journal of Statistics and Applied Mathematics 2021; 6(3): 16-19

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2021; 6(3): 16-19

www.mathsjournal.com
Received: 18-01-2021
Accepted: 10-03-2021

Jianqi Yu
College of Science, Guilin University of Technology, Guilin Guangxi, China

Xiang Wang
College of Science, Guilin University of Technology, Guilin Guangxi, China

Corresponding Author:
Jianqi Yu
College of Science, Guilin University of Technology, Guilin Guangxi, China

Maximum likelihood estimation for multivariate normal distribution with hierarchical missing data

Jianqi Yu and Xiang Wang

DOI: https://doi.org/10.22271/maths.2021.v6.i3a.681

Abstract
Closed forms are obtained for the maximum likelihood estimators (MLE) of the mean vector and the covariance matrix of a multivariate normal model with a hierarchical missing pattern. According to the missing pattern, the likelihood function is decomposed as product of several independent normal and conditional normal likelihood functions. The original parameters are transformed into a new set of parameters whose MLE are easy to derive. Since the MLE are invariant, the MLE of the original parameters are derived using the inverse transformation.

Keywords: Hierarchical missing data, maximum likelihood estimate, multivariate normal

Introduction

The problem of missing data is very common in practice, especially in public survey. For example, during data gathering and recording, when the experiment is involved a group of individuals over a period of time like in clinical trials or in a planned experiment where the variables that are expensive to measure are collected only from a subset of a sample, missing data arises.

There are several missing patterns considered in the literature, but the incomplete data with monotone pattern not only occur frequently in practice but also it allows the exact calculation of the maximum likelihood estimators (MLEs) and the likelihood ratio statistics and relevant distributions if multivariate normality is assumed. Jianqi Yu [1] defines hierarchical data missing pattern, which is a generalization of monotone data missing pattern. Anderson [2] gave a simple approach to derive the MLEs of bivariate normal data for a special case of monotone pattern. Kanda and Fujkoshi [3] studied some basic properties of the MLEs based on monotone data. Many authors developed asymptotic inferential procedures based on the likelihood ratio approach for multivariate normal distribution. We note, among many other papers, Bhargava [4], Morrison and Bhoj [5] and Naik [6].

This article derive the MLEs of the mean vector and the covariance matrix of a multivariate normal model with a hierarchical missing pattern. The causes for missing data could be various which will not be discussed in this article. However, to ignore the process that causes missing data, it is usually assumed that the data are missing at random (MAR). For an exposition of such issues, we refer to Little and Rubin [7] or Little [8]. Lu and Copas [9] pointed out that inference from the likelihood method ignoring the missing data mechanism is valid if and only if the missing data mechanism is MAR.

The hierarchical data missing pattern is like following data:

$$\begin{pmatrix} x_1,\cdots,x_n, & x_{n+1},\cdots,x_{n+m}, & x_{n+m+1},\cdots,x_{n+m+l}, & x_{n+m+l+1},\cdots,x_{n+m+l+k},\cdots, & x_{n+m+l+k+j} \\ y_1,\cdots,y_n, & y_{n+1},\cdots,y_{n+m} & & & \\ & & z_{n+m+1},\cdots,z_{n+m+l} & & \\ & & & u_{n+m+l+1},\cdots,u_{n+m+l+k}, & \\ v_1,\cdots,v_n & & & & \\ & w_{n+1},\cdots,w_{n+m} & & & \end{pmatrix} \quad (1)$$

Where the vector (x, y, z, u, v, w) represent the population with N = n + m + l + k + j observations, and the sub-index sets of the data satisfy following conditions:

1) The index set of the first row, i.e., (1, ., n, n + 1, ., n + m, n + m + 1, ., n + m + 1, n + m + 1 + 1., n), is the union of the index sets of all the other rows.
2) The index sets of two different rows are either disjoint, or inclusive.

It is easy to see that the monotone pattern is a special case of the hierarchical pattern.
The monotone pattern of missing data is like following data:

$$\begin{pmatrix} x_1,\cdots,x_n, & x_{n+1},\cdots,x_{n+m}, & x_{n+m+1},\cdots,x_{n+m+l}, & x_{n+m+l+1},\cdots,x_{n+m+l+k} \\ y_1,\cdots,y_n, & y_{n+1},\cdots,y_{n+m}, & y_{n+m+1},\cdots,y_{n+m+l} & \\ z_1,\cdots,z_n, & z_{n+1},\cdots,z_{n+m} & & \\ u_1,\cdots,u_n & & & \end{pmatrix} \qquad (2)$$

1. Maximum likelihood estimators

Assume that $(x, y, z, u, v, w) \sim N_p(\beta, \Sigma)$, x, y, z, u, v, w are p_i dimensional respectively, i = 1, 2, …, 6, $\sum p_i = p$.

We partition the parameters as follows:

$$\beta = (\beta_1,\ \beta_2,\ \beta_3,\ \beta_4,\ \beta_5,\ \beta_6)'$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} & \Sigma_{15} & \Sigma_{16} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{21} & \Sigma_{21} & \Sigma_{21} & \Sigma_{21} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} & \Sigma_{34} & \Sigma_{35} & \Sigma_{36} \\ \Sigma_{41} & \Sigma_{42} & \Sigma_{43} & \Sigma_{44} & \Sigma_{45} & \Sigma_{46} \\ \Sigma_{51} & \Sigma_{52} & \Sigma_{53} & \Sigma_{54} & \Sigma_{55} & \Sigma_{56} \\ \Sigma_{61} & \Sigma_{62} & \Sigma_{63} & \Sigma_{64} & \Sigma_{65} & \Sigma_{66} \end{pmatrix}$$

and partition the data as follows:

$$D_1 = \begin{pmatrix} x_1,\cdots,x_n \\ y_1,\cdots,y_n \\ \\ v_1,\cdots,v_n \end{pmatrix} = D_{xyv} \quad D_2 = \begin{pmatrix} x_{n+1},\cdots,x_{n+m} \\ y_{n+1},\cdots,y_{n+m} \\ \\ w_{n+1},\cdots,w_{n+m} \end{pmatrix} = D_{xyw} \quad D_3 = \begin{pmatrix} x_{n+m+1},\cdots,x_{n+m+l} \\ \\ z_{n+m+1},\cdots,z_{n+m+l} \\ \end{pmatrix} = D_{xz}$$

$$D_4 = \begin{pmatrix} x_{n+m+l+1},\cdots,x_{n+m+l+k}, & x_{n+m+l+k+1}\cdots, & x_{n+m+l+k+j} \\ & & \\ u_{n+m+l+1},\cdots,u_{n+m+l+k} & & \\ & & \end{pmatrix} \qquad (3)$$

Let

$$D_{xu} = \begin{pmatrix} x_{n+m+l+1},\cdots,x_{n+m+l+k} \\ u_{n+m+l+1},\cdots,u_{n+m+l+k} \end{pmatrix}, \quad D_{xy} = \begin{pmatrix} x_1,\cdots,x_n, x_{n+1},\cdots,x_{n+m} \\ y_1,\cdots,y_n, y_{n+1},\cdots,y_{n+m} \end{pmatrix}$$

Except D_4, we denote the sample mean vector and the sums of squares and products matrix based on D_i by $\overline{D_i}, S_i$. Let $\overline{D_i}^{\alpha}, S_i^{(\alpha,\gamma)}$ are $\alpha - th$ sub-vector and $(\alpha,\gamma)-th$ sub-matrix of $\overline{D_i}, S_i$ respectively.

Finally, denote the sample mean vector and the sums of squares and products matrix based on x by $\bar{x}, S_x$, and using similar notation for y, z, u, v, and w.

Consider the density function of data in (3). We note that the density can be written as the marginal density times the conditional density (we indicate the density of normal distribution by n (·) here), that is:

For D_1, $\prod_{i=1}^{n} n(x_i, y_i, v_i \mid \beta, \Sigma) = \prod_{i=1}^{n} n(x_i \mid \beta_1, \Sigma_{11}) n(y_i \mid \beta_{2.1} + B_{2.1}x_i, \Sigma_{2.1}) n(v_i \mid \beta_{5.21} + B_{5.12}(x_i, y_i), \Sigma_{5.21})$

Where $B_{2.1} = \Sigma_{21}\Sigma_{11}^{-1}, \beta_{2.1} = \beta_2 - B_{2.1}\beta_1, \Sigma_{2.1} = \Sigma_{22} - B_{2.1}\Sigma_{12}$

$$B_{5.21} = (\Sigma_{51}, \Sigma_{52})\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}^{-1}, \beta_{5.21} = \beta_5 - B_{5.21}\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}, \Sigma_{5.21} = \Sigma_{55} - B_{5.21}\begin{pmatrix}\Sigma_{15}\\ \Sigma_{25}\end{pmatrix} \quad (4)$$

For D_2 $\prod_{i=n+1}^{n+m} n(x_i, y_i, w_i \mid \beta, \Sigma) = \prod_{i=n+1}^{n+m} n(x_i \mid \beta_1, \Sigma_{11}) n(y_i \mid \beta_{2.1} + B_{2.1}x_i, \Sigma_{2.1}) n(w_i \mid \beta_{6.21} + B_{6.12}(x_i, y_i), \Sigma_{6.21})$

Where

$$B_{6.21} = (\Sigma_{61}, \Sigma_{62})\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}^{-1}, \beta_{6.21} = \beta_6 - B_{6.21}\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}, \Sigma_{6.21} = \Sigma_{66} - B_{6.21}\begin{pmatrix}\Sigma_{16}\\ \Sigma_{26}\end{pmatrix} \quad (5)$$

For D_3,

$$\prod_{i=n+m+1}^{n+m+l} n(x_i, z_i \mid \beta, \Sigma) = \prod_{i=n+m+1}^{n+m+l} n(x_i \mid \beta_1, \Sigma_{11}) n(z_i \mid \beta_{3.1} + B_{3.1}x_i, \Sigma_{3.1})$$

Where $B_{3.1} = \Sigma_{31}\Sigma_{11}^{-1}, \beta_{3.1} = \beta_3 - B_{3.1}\beta_1, \Sigma_{3.1} = \Sigma_{33} - B_{3.1}\Sigma_{13}$ (6)

For D_4, $\prod_{i=n+m+l+1}^{n+m+l+k} n(x_i,\ u_i \mid \beta, \Sigma) \prod_{i=n+m+l+k+1}^{n+m+l+k+j} n(x_i \mid \beta, \Sigma) = \prod_{i=n+m+l+1}^{n+m+l+k+j} n(x_i \mid \beta_1, \Sigma_{11}) \prod_{i=n+m+l+1}^{n+m+l+k} n(u_i \mid \beta_{4.1} + B_{4.1}x_i, \Sigma_{4.1})$

Where $B_{4.1} = \Sigma_{41}\Sigma_{11}^{-1}, \beta_{4.1} = \beta_4 - B_{4.1}\beta_1, \Sigma_{4.1} = \Sigma_{44} - B_{4.1}\Sigma_{14}$ (7)

The likelihood function is

$$L(x,y,z,u,v,w \mid \beta, \Sigma) = \prod_{i=1}^{n+m+l+k+j} n(x_i \mid \beta_1, \Sigma_{11}) \prod_{i=1}^{n+m} n(y_i \mid \beta_{2.1} + B_{2.1}x_i, \Sigma_{2.1}) \prod_{i=n+m+1}^{n+m+l} n(z_i \mid \beta_{3.1} + B_{3.1}x_i, \Sigma_{3.1})$$
$$\prod_{i=n+m+l+1}^{n+m+l+k} n(u_i \mid \beta_{4.1} + B_{4.1}x_i, \Sigma_{4.1}) \prod_{i=1}^{n} n(v_i \mid \beta_{5.21} + B_{5.12}(x_i, y_i), \Sigma_{5.21}) \prod_{i=n+1}^{n+m} n(w_i \mid \beta_{6.21} + B_{6.12}(x_i, y_i), \Sigma_{6.21}) \quad (8)$$

Consider MLE of the parameters in (8), it is obvious that

$$\hat{\beta}_1 = \bar{x}, \hat{\Sigma}_{11} = S_x \quad (9)$$

Next, Regress y on x in D_{xy}, we have

$$\hat{B}_{2.1} = S_{xy}^{(2,1)}(S_{xy}^{(1,1)})^{-1}, \hat{\beta}_{2.1} = \overline{D}_{xy}^{(2)} - \hat{B}_{2.1}\overline{D}_{xy}^{(1)}, \hat{\Sigma}_{2.1} = S_{xy}^{(2,2)} - \hat{B}_{2.1}S_{xy}^{(1,2)} \quad (10)$$

Similarly,

$$\hat{B}_{3.1} = S_{xz}^{(2,1)}(S_{xz}^{(1,1)})^{-1}, \hat{\beta}_{3.1} = \overline{D}_{xz}^{(2)} - \hat{B}_{3.1}\overline{D}_{xz}^{(1)}, \hat{\Sigma}_{3.1} = S_{xz}^{(2,2)} - \hat{B}_{3.1}S_{xz}^{(1,2)} \quad (11)$$

$$\hat{B}_{4.1} = S_{xu}^{(2,1)}(S_{xu}^{(1,1)})^{-1}, \hat{\beta}_{4.1} = \overline{D}_{xu}^{(2)} - \hat{B}_{4.1}\overline{D}_{xu}^{(1)}, \hat{\Sigma}_{4.1} = S_{xu}^{(2,2)} - \hat{B}_{4.1}S_{xu}^{(1,2)} \quad (12)$$

$$\hat{B}_{4.1} = S_{xu}^{(2,1)}(S_{xu}^{(1,1)})^{-1}, \hat{\beta}_{4.1} = \overline{D}_{xu}^{(2)} - \hat{B}_{4.1}\overline{D}_{xu}^{(1)}, \hat{\Sigma}_{4.1} = S_{xu}^{(2,2)} - \hat{B}_{4.1}S_{xu}^{(1,2)} \quad (13)$$

$$\hat{B}_{5.21} = (S_{xyv}^{(3,1)}, S_{xyv}^{(3,2)})\begin{pmatrix} S_{xyv}^{(1,1)} & S_{xyv}^{(1,2)} \\ S_{xyv}^{(2,1)} & S_{xyv}^{(2,2)} \end{pmatrix}^{-1}, \hat{\beta}_{5.21} = \overline{D}_{xyv}^{(3)} - \hat{B}_{5.21}\begin{pmatrix} \overline{D}_{xyv}^{(1)} \\ \overline{D}_{xyv}^{(2)} \end{pmatrix}, \hat{\Sigma}_{5.21} = S_{xyv}^{(3,3)} - \hat{B}_{5.21}\begin{pmatrix} S_{xyv}^{(1,3)} \\ S_{xyv}^{(2,3)} \end{pmatrix} \quad (14)$$

$$\hat{B}_{6.21} = (S_{xyw}^{(3,1)}, S_{xyw}^{(3,2)})\begin{pmatrix} S_{xyw}^{(1,1)} & S_{xyw}^{(1,2)} \\ S_{xyw}^{(2,1)} & S_{xyw}^{(2,2)} \end{pmatrix}^{-1}, \hat{\beta}_{6.21} = \overline{D}_{xyw}^{(3)} - \hat{B}_{6.21}\begin{pmatrix} \overline{D}_{xyw}^{(1)} \\ \overline{D}_{xyw}^{(2)} \end{pmatrix}, \hat{\Sigma}_{6.21} = S_{xyw}^{(3,3)} - \hat{B}_{6.21}\begin{pmatrix} S_{xyw}^{(1,3)} \\ S_{xyw}^{(2,3)} \end{pmatrix} \quad (15)$$

Solve the equations in (4), (5), (6), (7), we get the MLE of the original parameters:

$$\hat{\beta}_1 = \bar{x}, \hat{\Sigma}_{11} = S_x$$

$$\hat{\Sigma}_{21} = \hat{B}_{2.1}\hat{\Sigma}_{11}, \quad \hat{\beta}_2 = \hat{\beta}_{2.1} + \hat{B}_{2.1}\hat{\beta}_1, \hat{\Sigma}_{22} = \hat{B}_{2.1}(\hat{\Sigma}_{21})' + \hat{\Sigma}_{2.1}$$

$$\hat{\Sigma}_{31} = \hat{B}_{3.1}\hat{\Sigma}_{11}, \quad \hat{\beta}_3 = \hat{\beta}_{3.1} + \hat{B}_{3.1}\hat{\beta}_1, \hat{\Sigma}_{33} = \hat{B}_{3.1}(\hat{\Sigma}_{31})' + \hat{\Sigma}_{3.1}$$

$$\hat{\Sigma}_{41} = \hat{B}_{4.1}\hat{\Sigma}_{11}, \quad \hat{\beta}_4 = \hat{\beta}_{4.1} + \hat{B}_{4.1}\hat{\beta}_1, \hat{\Sigma}_{44} = \hat{B}_{4.1}(\hat{\Sigma}_{41})' + \hat{\Sigma}_{4.1}$$

$$(\hat{\Sigma}_{51}, \hat{\Sigma}_{52}) = \hat{B}_{5.21}\begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix}, \quad \hat{\beta}_5 = \hat{\beta}_{5.21} + \hat{B}_{5.1}\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix},, \hat{\Sigma}_{55} = \hat{B}_{5.21}(\hat{\Sigma}_{51}, \hat{\Sigma}_{52})' + \hat{\Sigma}_{5.21}$$

$$(\hat{\Sigma}_{61}, \hat{\Sigma}_{62}) = \hat{B}_{6.21}\begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix}, \quad \hat{\beta}_6 = \hat{\beta}_{6.21} + \hat{B}_{6.21}\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix},, \hat{\Sigma}_{66} = \hat{B}_{6.21}(\hat{\Sigma}_{61}, \hat{\Sigma}_{62})' + \hat{\Sigma}_{6.21}$$

It is worth to note that parameters such as Σ_{32} have no estimates, since there is no information of the correlation between z and x.

References

1. Jianqi Yu. Hierarchical Missing Data and Multivariate Behrens - Fisher Problem [J], Journal of Mathematics 2021. DOI:10.1155/2021/8837044.
2. Anderson TW. Maximum likelihood estimates for a multivariate normal distribution when some observations are missing [J]. Journal of American Statistical Association 1957(52).
3. Kanda T, Fujikoshi Y. Some basic properties of the MLEs for a multivariate normal distribution with monotone missing data [J]. Journal of Mathematics and Management Science 1998(18).
4. Bhargava BP. Multivariate tests of hypotheses with incomplete data [D]. Stanford CA: Stanford University 1962.
5. Morrison DF, Bhoj D. Power of the likelihood ratio test on the mean vector of the multivariate normal distribution with missing observations [J]. Biometrika 1973(60).
6. Naik UD. On testing equality of means of correlated variables with incomplete data [J]. Biometrika 1975(62).
7. Lu GB, Copas JB. Missing at random, likelihood ignitability and model completeness [J]. Annals of Statistics 2004(32).
8. Little RJA, Rubin DB. Statistical Analysis with Missing Data [M]. New York: Wiley 1987.
9. Little RJA. A test of missing completely at random for multivariate data with missing values [J]. Journal of American Statistical Association 1988(83).