Estimation for Multivariate Linear Mixed Models

I Nyoman Latra¹, Susanti Linuwih², Purhadi², and Suhartono²
1 Doctoral Candidate at Department of Statistics FMIPA-ITS Surabaya,
Email: i_nyoman_l@statistika.its.ac.id
2 Promotor and Co-promoters at Department of Statistics FMIPA-ITS Surabaya.

Abstract— This paper discusses about estimation of multivariate linear mixed model or multivariate component of variance model with equal number of replications. We focus on two estimation methods, namely Maximum Likelihood Estimation (MLE) and Restricted Maximum Likelihood Estimation (REML) methods. The results show that the parameter estimation of fixed effects yields unbiased estimators, whereas the estimation for random effects or variance components yields biased estimators. Moreover, assume that both likelihood and ln-likelihood functions hold some of regularity conditions, it can be proved that estimators as a solutions set of the likelihood equations satisfy strong consistency for large sample size, asymptotic normal and efficiency.

Index Term-- Linear Mixed Model, Multivariate Linear Model, Maximum Likelihood, Asymptotic Normal and Efficiency, Consistency.

I. INTRODUCTION

Linear mixed models or variance components models have been effectively and extensively used by statisticians for analyzing data when the response is univariate. Reference [12] discussed the latent variable model for mixed ordinal or discrete and continuous outcomes that was applied to birth defects data. Reference [16] showed that maximum likelihood estimation of variance components from twin data can be parameterized in the framework of linear mixed models. Specialized variance component estimation software that can handle pedigree data and user-defined covariance structures can be used to analyze multivariate data for simple and complex models with a large number of random effects. Reference [2] showed that Linear Mixed Models (LMM) could handle data where the observations were not independent or could be used for modeling data with correlated errors. There are some technical terms for predictor variables in linear mixed models, those are (i) random effects, i.e. the set of values of a categorical predictor variable that the values are selected not completely but as random sample of all possibility values (for example, the variable “product” has values representing only 5 of a possible 42 brands), (ii) hierarchical effects, i.e. predictor variables are measured at more than one level, and (iii) fixed effects, i.e. predictor variables which all possible category values (levels) are measured.

Otherwise, in this paragraph there are many papers discussed about linear model for multivariate cases. Reference [5] applied multivariate linear mixed model to Scholastic Aptitude Test and proposed Restricted Maximum Likelihood (REML) to estimate the parameters. Reference [6] used multivariate linear mixed model or multivariate variance components model with equal replication to predict the sum of the regression mean and the random effects of models. This prediction problem reduces to the estimation of the ratio of two covariance matrices. The estimation of the ratio matrix obtain the James-Stein type estimators based on the Bartlett's decomposition, the Stein type orthogonally equivariant estimators, and the Efron-Morris type estimators. Recently, Reference [10] applied linear mixed model in Statistics for Biology Systems to calculate both covariates and correlations between signals which followed non-stationary time series. They used the estimation algorithm based on Expectation-Maximization (EM) which involved dynamic programming for the segmentation step. Reference [11] discussed a joint model for multivariate mixed ordinal and continuous responses. The likelihood is found and modified Pearson residuals. The model is applied to medical data, obtained from an observational study on woman.

Most of previous papers about multivariate linear mixed models focused on estimation method and not yet discuss about the asymptotic properties of the estimators. By assumption that both likelihood and ln-likelihood functions hold for some of regularity conditions, then a solutions set of the likelihood equations satisfy strong consistency for large sample size, asymptotic normal and efficiency.

II. MULTIVARIATE LINEAR MIXED MODEL

In this section, we firstly discuss about the univariate linear mixed models and then continue to multivariate linear models without replications. Whereas discuss about the multivariate linear mixed models will be included in next section. Linear mixed models are statistical models for continuous outcome variables in which the residuals are normally distributed but may not be independent or have constant variance. A linear mixed model is a parametric linear model for clustered, longitudinal, or repeated-measures data. It may include both fixed-effect parameters associated with one or more continuous or categorical covariates and random effects that associated with one or more random factors. The fixed-effect parameters describe the relationships of the covariates to the dependent variable for an entire population, and the random effects are specific to clusters or subjects within a population [4], [9], [15], [17].

In matrix and vector, a linear mixed population model (univariate) is written as [4], [7], [9], [17]

$$ y = Xβ + Zδ + e $$

(1)

where $y$ is a $n \times 1$ vector of observations, $X$ is a $n \times (p+1)$ matrix of known covariates, $Z$ is a $n \times m$ known matrix, $β$ is a $(p + 1) \times 1$ vector of unknown regression coefficients which are usually known as the fixed effects, $δ$ is a $m \times 1$ vector of random effects, and $e$ is a $n \times 1$ vector of random errors. Both vectors $δ$ and $e$ are unobservable. The basic assumption of Eq.
y = Xβ + e’ or y ~ N(Xβ, Σ). (2)

Equation (2) is called marginal model or sometimes called as model of population mean.

For point estimation process, each of variance component of model can be notated by θ that include in vector θ = vech(Σ) = [θ₁, ..., θ_{p(p+1)/2}]’. It implies Gaussian mixed model has a joint probability density function as follows:

\[ f(y, β, θ) = \frac{1}{\sqrt{2π}^n |Σ|} \cdot \exp\left[-\frac{1}{2}(y - Xβ)^{T}Σ^{-1}(y - Xβ)/2\right] \]  

where \( n \) is dimension of vector \( y \) and \( \text{vech}(Σ) \) indicates a vector with all lower triangular elements of \( Σ \) have been stacked by column. A ln-likelihood function of (3) is

\[ \ell(β, θ, 0) = \ln L(y, β, θ) = -(n/2)ln(2π) - (1/2)ln|Σ| \]

\[-(1/2)(y - Xβ)^{T}Σ^{-1}(y - Xβ) \]  

(4)

Taking the partial derivatives of (4) with respect to \( β, θ_q \) and makes each equal to zero such that we have parameters estimation as follows:

\[ \hat{β} = (X^TΣ^{-1}X)^{-1}X^TΣ^{-1}y, \]  

\[ y^{T} \frac{∂Σ}{∂θ_q} \bigg|_{θ_q} \]  

\[ Py = \text{tr}(Σ^{-1} \frac{∂Σ}{∂θ_q} \bigg|_{θ_q}), q = 1, 2, ..., n(n+1)/2, \]  

(5)

where \( \hat{Σ} = Σ^{-1} - Σ^{-1}X(X^TΣ^{-1}X)^{-1}X^TΣ^{-1} \).

Equation (5) has a "closed-form". Whereas (6) does not have a "closed-form", therefore [4], [17] proposed three algorithms that could be used to calculate the parameter estimation \( θ_q \), i.e. EM (expectation-maximization), N-R (Newton-Raphson), and Fisher Scoring algorithm. Due to the MLE is consistent and asymptotically normal with covariance matrix asymptotic equal to inverse of Fisher Information matrix, so we need to obtain Fisher Information matrix components. In both univariate and multivariate linear mixed models, the regression coefficients and covariance matrix components will be placed it in to a vector \( ψ \), that is \( ψ = (β’, θ')' \). Thus, the Fisher Information matrix will have expressions as follows

\[ \text{Var}\left( \frac{∂ℓ}{∂θ_q} \right) = -E\left( \frac{∂^2ℓ}{∂ψ∂ψ'} \right), \]  

\[ \text{Var}\left( \frac{∂ℓ}{∂θ_q} \right) = -E\left( \frac{∂^2ℓ}{∂θ_q∂θ_q'} \right). \]  

by an assumption that second partial derivative of \( ψ \) exist. Mean value of second partial derivative respect to fixed components or variance components will be [4]

\[ E\left( \frac{∂^2ℓ}{∂θ_q∂θ_q'} \right) = -X^TΣ^{-1}X, \]  

\[ E\left( \frac{∂^2ℓ}{∂θ_q∂θ_q'} \right) = 0; 1 ≤ q ≤ n(n+1)/2, \]  

\[ E\left( \frac{∂^2ℓ}{∂θ_q∂θ_q'} \right) = -(1/2)κ\left\{ Σ^{-1} \frac{∂Σ}{∂θ_q} Σ^{-1} \frac{∂Σ}{∂θ_q'} \right\}, \]  

\[ 1 ≤ q, q' ≤ n(n+1)/2. \]  

Base on (5) and a result of Theorem 1 in [15], the ML Iterative Algorithm can be used to compute the MLE of the unknown parameters in model (3). By using iteration process, the estimator of variance component are found as elements of \( \text{vech}(Σ) = \hat{θ} \) such that estimator of covariance matrix can be written as,

\[ \hat{Σ} = Σ|_{θ_q}. \]  

(8)

that is generally bias. By substituting (8) to (5), it yields \( \text{Var}(\hat{β}) = (X^TΣ^{-1}X)^{-1} \) that is bias because (8) is bias. Bias in MLE can be eliminated by Restricted Maximum Likelihood Estimation (REMLE).

REML estimation uses a transformation \( y_1 = A^r y \), where \( \text{rank}(X) = p + 1 \). \( A \) is \( n × (n - p - 1) \) with \( \text{rank}(A) = n - p - 1 \), such that \( A^r X = 0 \). It is easy to show that that \( y_1 ∼ N(0, A^rΣA) \). Furthermore, the joint probability density function of \( y_1 \) is

\[ f_s(y_1) = \frac{1}{\sqrt{2π}^{(n-p-1)} |A^rΣA|} \cdot \exp\left[-(1/2)y_1^T(A^rΣA)^{-1}y_1\right]. \]  

(9)

The ln-likelihood of (9) can be written as

\[ \ell_s(0) = -(1/2)[ln|A^rΣA| - y_1^T(A^rΣA)^{-1}y_1]. \]  

(10)

Estimation process can be carried out by partially derivation of (10) respect to \( θ_q; q = 1, ..., n(n+1)/2 \) (because it does not contain components of \( β \)),

\[ \frac{∂ℓ_s}{∂θ_q} = (1/2)[y_1^TΣ^{-1}Py - tr(PΣ^{-1}Py - tr(PΣ^{-1}Py))]. \]  

(11)

where \( P \) in (11) is \( A(A^rΣA)^{-1}A^r \) which basically is equal to \( P \) that the estimator is \( \hat{P} \) in (6). Fisher Information matrix of (11) is

\[ \text{Var}\left( \frac{∂ℓ_s}{∂θ_q} \right) = -E\left( \frac{∂^2ℓ_s}{∂θ_q∂θ_q'} \right). \]  

(12)
By assuming that the second derivative of \( \Sigma \) in (10) is exist, then it can be proven that its components Fisher is

\[
E \left( \frac{\partial^2 \ell_i}{\partial \theta_i \partial \theta_j} \right) = \frac{1}{2} \text{tr} \left( \Sigma \frac{\partial \Sigma}{\partial \theta_i} \Sigma \frac{\partial \Sigma}{\partial \theta_j} \right);
\]

\[1 \leq i, j \leq n(n+1)/2 . \quad (13)\]

The estimator \( \hat{\Sigma} \) that be obtained from REMLE method is unbiased. However, the linear mixed models will yield \( \text{Var}(\hat{\beta}) \) that is bias, because \( \text{Var}(\hat{\beta}) \) is taken from (5) by doing a substitution of \( \Sigma \) by \( \hat{\Sigma} \). Bias from variance of fixed effect estimator that constitute the diagonal elements of \( \text{Var}(\hat{\beta}) \) is a downward bias both in MLE and REMLE method [17].

Multivariate linear model that does not depend to random factors was discussed by [1], [9]. Multivariate linear model without replications in the matrix form could be written as

\[
Y = XB + E \quad (14)
\]

where \( Y = [Y_1, ..., Y_s], E = [\epsilon_{ij}], X = [x_{ij}], \) and \( B = [\beta_{ij}]; i = 1, ..., n; j = 1, ..., s; a = 0, ..., p; \) intercepts \( x_{0} = 1 \) for all \( i \). For individual \( i \), (14) could be presented in the matrix and vector form as follows:

\[
y_i = B x_i + e_i \quad (15)
\]

where

\[
y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{is} \end{bmatrix}, \quad B = \begin{bmatrix} \beta_{i1} & \beta_{i2} & \cdots & \beta_{is} \\ \beta_{i1} & \beta_{i2} & \cdots & \beta_{is} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{ip} & \beta_{ip} & \cdots & \beta_{ip} \end{bmatrix},
\]

\[
x_i = \begin{bmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{ip} \end{bmatrix}, \quad \text{and} \ e_i = \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{is} \end{bmatrix}.
\]

The error vectors \( e_i \) in (15) are random vectors that be assumed follow multivariate normal distribution with zero mean vector in \( R^m \). The assumption of rows of \( E \) are independents because each of them is related to different observations, however the columns of \( E \) are still allowed to have a correlation. Thus,

\[
\text{Cov}(e_i, e_j) = \begin{cases} 0, & i \neq j \\ \Sigma, & \end{cases} \quad (16)
\]

for all \( i = 1, 2, ..., n \). The other assumption is matrix of variance errors in each observation \( \Sigma_i = \Sigma \) with size \( s \times s \) is same for all \( i \) and positive definite. Then, (16) can be written as [13]

\[
\text{Cov}(\text{vec}(E^T), \text{vec}(E^T)) = I_n \otimes \Sigma
\]

where \( \text{vec}(E^T) \) is a vector that indicates all elements of a matrix \( E \) have been stacked by rows and \( I_n \) is identity matrix with size \( n \times n \).

Moreover, estimation of \( B \) can be done by assuming that \( \Sigma \) is known. The likelihood function of (14) for \( n \) observations is

\[
L(\text{vec}(B^T), \text{vec}(\Sigma)) = 1/(2\pi)^{n/2} |\Sigma|^{-1} \exp\left(-Q_{slm}/2 \right) \quad (17)
\]

where \( Q_{slm} = \text{tr}(\Sigma^{-1}(Y' 
 (X \otimes I_p)\text{vec}(B^T))^T (X \otimes I_p)\text{vec}(B^T)) \)

Maximize (17) with constraint that \( \Sigma \) is known yield estimator

\[
\hat{B} = (X'X)^{-1}X'Y \quad \text{or} \quad \text{vec}(\hat{B}^T) = ((X'X)^{-1}X' \otimes I_p)\text{vec}(Y^T) . \quad (18)
\]

The estimation of \( \Sigma \) can be obtained by assuming that \( B \) is known. Then, maximize (17) and assume that \( B \) is known will yield variance estimator, i.e.

\[
\hat{\Sigma} = (1/2)Y' (I - X(X'X)^{-1}X')Y . \quad (19)
\]

Hence, estimator in (18) is an unbiased estimator with the variance matrix

\[
\text{Var}(\text{vec}(\hat{B}^T)) = (X'X)^{-1} \otimes \hat{\Sigma} . \quad (20)
\]

### III. PARAMETERS ESTIMATION for MULTIVARIATE LINEAR MIXED MODEL

The multivariate linear mixed model is an extension of the multivariate linear model. It means that the model can be constructed by adding a random component part to (14) and assuming that each elements of \( Y \) has a linear correlation with systematic part of the model. This model is called as the multivariate linear mixed model without replication. There are three types of the linear mixed model, i.e. cluster, longitudinal, and replication. This paper focuses on the linear mixed model without replication, i.e.

\[
Y = XB + ZD + E \quad (21)
\]

where \( X \) and \( Z \) are known as covariates matrices with size \( n \times (p+1) \) and \( n \times k \) respectively, \( B \) is a \( (p+1) \times s \) matrix of unknown regression coefficients of fixed effect, \( D \) is a \( k \times s \) matrix of specific coefficients of random effect, and \( E \) is a \( n \times s \) matrix of errors. It is also assumed that \( n \)-rows of \( E \) are independents and each row is \( N(0, \Sigma) \) and random effect \( D \) satisfies \( \text{vec}(D) \sim N(0, \Phi) \). Thus, the distribution of responses \( Y \) in model (21) can be written in matrix and vector form as
Let a matrix of variance $V=(Z \otimes I_j) \Phi (Z \otimes I_j)^T + (I_j \otimes \Sigma)$ with size $n \times n$, then variance components of model (21) are elements of $\Theta = vec(V) = [\theta_1, \ldots, \theta_{ntot+1/2}]^T$. The likelihood function of probability density function (22) is

$$L(\text{vec}(V^T), \text{vec}(V)) = 1/\sqrt{(2\pi)^n |V|} \exp(-Q_{\text{MLM}}/2) \quad (23)$$

where $Q_{\text{MLM}} = \{\text{vec}(Y^T) - (X \otimes I_j)\text{vec}(B^T)\}^T \cdot \text{vec}^{-1}(X \otimes I_j)\text{vec}(B^T)\}^T \cdot \text{vec}(Y^T) - (X \otimes I_j)\text{vec}(B^T)\}^T$.

By applying partial derivation of $ln$-likelihood function (23) with respect to $\text{vec}(B^T)$ and $\theta_j$, then make equal to zero and it will yield estimators $\text{vec}(B^T) = [(X \otimes I_j)^T \tilde{V}^{-1}(X \otimes I_j)]^{-1}(X \otimes I_j)^T \tilde{V}^{-1}\text{vec}(Y^T)$. (24)

and by using a nonlinear optimization method, with inequality constraints imposed on $\Theta$ so that positive definiteness requirements on the $\Phi$ and $\Sigma$ matrices are satisfied. There is no closed-form solution for the $\Theta$, so the estimate of $\Theta$ is obtained by the Fisher scoring algorithm or Newton-Raphson algorithm. After an iterative computational process we have $\tilde{V} = \text{vec}(\text{vec}(V^T)) = 1/\sqrt{(2\pi)^n |V|} \exp(-Q_{\text{MLM}}/2)$

Both components estimator of $\hat{\Phi}$ and $\hat{\Sigma}$ can be obtained too by applying iteration that are included in iteration process for computing $\tilde{V}$. The result of estimation in (24) is unbiased estimator with variance matrix

$$\text{Var}(\text{vec}(B^T)) = [(X \otimes I_j)^T \tilde{V}^{-1}(X \otimes I_j)]^{-1}.$$

**IV. ESTIMATOR PROPERTIES: CONSISTENCY, EFFICIENCY, and ASYMPTOTICALLY**

Let vectors of random samples in the form of observations $y_1, y_2, \ldots, y_n$ from a vector of random variable $y$ with distribution $F_y$ belonging to a family $\mathcal{F} = \{F_{\psi} : \psi \in \Theta\}$, where $\psi = [\theta, \ldots, \theta_j]^T$. For each $\psi \in \Theta$, a subset $N(\psi_y)$ of $\Theta$ is a neighborhood of $\psi_y$ if $N(\psi_y)$ is a superset of an open set $G$ containing $\psi_y$ [8]. Assume $\Theta \subset \mathbb{R}^k$ and there are three regularity conditions on $\mathcal{F}$, those are [3, 14]:

(R1). For each $\psi \in \Theta$, the derivatives $\frac{\partial^I(I(y, \psi))}{\partial \theta_j}$ exist, for all $y$ and $i, i^* = 1, 2, \ldots, k$.

(R2).

Proof:

The $L(y, \psi)$ is a likelihood function. Thus, it can be considered as a joint probability density function, so

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} L(y, \psi) dy = 1.$$

Taking the partial derivative with respect to vector $\psi$, then the results are as follows:

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \frac{\partial}{\partial \psi} L(y, \psi) dy = 0 \quad (28)$$

Equation (28) can be written as

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \frac{\partial}{\partial \psi} \ln L(y, \psi) L(y, \psi) dy = 0$$

or

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \frac{\partial^I L(y, \psi)}{\partial \psi} L(y, \psi) dy = 0 \quad (29)$$

For each $\psi \in \Theta$, there are exist functions $g(y), h(y)$, and $H(y)$ (poss. Writing (29) as an expectation, we have established
\[ E\left[ \frac{\partial l(y, \psi)}{\partial \psi} \right] = 0. \quad (30) \]

It means that the vector mean of random variable \( \frac{\partial l(y, \psi)}{\partial \psi} \) is 0. If (29) be partially differentiated again, it follows that

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \frac{\partial^2 (\ln L(y, \psi))}{\partial \psi \partial \psi'} \right) L(y, \psi) \, dy \, d\psi' + \frac{\partial (\ln L(y, \psi))}{\partial \psi} \frac{\partial}{\partial \psi} L(y, \psi) \, dy = 0.
\]

This is equivalent to

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \frac{\partial^2 (\ln L(y, \psi))}{\partial \psi \partial \psi'} \right) L(y, \psi) \, dy \, d\psi' + \int_{-\infty}^{\infty} \frac{\partial (\ln L(y, \psi))}{\partial \psi} \frac{\partial}{\partial \psi} L(y, \psi) \, dy = 0.
\]

This last equation can be written as an expectation, i.e.

\[
E\left[ \frac{\partial^2 l(y, \psi)}{\partial \psi \partial \psi'} \right] + E\left[ \frac{\partial l(y, \psi)}{\partial \psi} \right]^2 = 0
\]

By using (30) and definition of variance, the latter expression can be rewritten as

\[
E\left[ \frac{\partial^2 l(y, \psi)}{\partial \psi \partial \psi'} \right] + Var\left[ \frac{\partial l(y, \psi)}{\partial \psi} \right] = 0 \quad \text{or}
\]

\[
\text{I}(\psi) = Var\left[ \frac{\partial l(y, \psi)}{\partial \psi} \right] = -E\left[ \frac{\partial^2 l(y, \psi)}{\partial \psi \partial \psi'} \right]
\]

\[ = -E[\nabla^2 l(y, \psi)] \quad [3], \quad [14]. \]

\[ \square \]

**Theorem:**

Let \( n \) vectors observations \( y_1, y_2, \cdots, y_n \) be iid with distribution \( F_{\psi} \), for \( \psi \in \Theta \). Assume regularity conditions (R1), (R2), and (R3) hold on the family \( \mathcal{F} \). Then, by \( \hat{\psi}_n \) converges with probability 1 to \( \psi \), the likelihood equations admit a sequence of solutions \( \hat{\psi}_n \) satisfying

a). strong consistency:

\[ \hat{\psi}_n \xrightarrow{p} \psi, \quad m \to \infty; \quad (31) \]

b). asymptotic normality and efficiency

\[ \hat{\psi}_n \xrightarrow{d} \text{AN}(\psi, (1/m)I^{-1}(\psi)), \quad (32) \]

where \( \text{I}(\psi) = -E[\nabla^2 l(y, \psi)] \), \( \text{AN}(\cdot, \cdot) \) is a multivariate asymptotic normal distribution.

**Proof:**

Expanding the function \( \frac{\partial l(y, \psi)}{\partial \psi} \) into a Taylor series of second order about \( \psi_0 \) and evaluating it at \( \hat{\psi}_n \), as follows:

\[
\frac{\partial l(y, \hat{\psi}_n)}{\partial \psi} = \frac{\partial l(y, \psi_0)}{\partial \psi} + \left[ \nabla^2 l(y, \psi_0)(1/2)(\hat{\psi}_n - \psi_0)^T \right.
\]

\[ \otimes (\nabla l(y, \psi_0))^T \text{H}(y, \psi_0^*) \left( \hat{\psi}_n - \psi_0 \right) \quad (33) \]

where \( \| \psi_n - \psi_0 \| < \| \hat{\psi}_n - \psi_0 \| \), \( \text{H}(y, \psi_0^*) = (\nabla \psi)(\nabla \psi)^T \).

But \( \frac{\partial l(y, \hat{\psi}_n)}{\partial \psi} = 0 \), so Taylor series (33) can be written as

\[
\{-(1/m)\nabla^2 l(y, \psi_0)(1/2m)(\hat{\psi}_n - \psi_0)^T \otimes (\nabla l(y, \psi_0))^T \}
\]

\[ \cdot \text{H}(y, \psi_0^*)\nabla l(y, \psi_0) = (1/m)\nabla l(y, \psi_0) \quad (34) \]

Therefore, putting

\[ a_n = (1/\sqrt{m}) \sum_{i=1}^{m} \nabla l(y_i, \psi_0) \]

\[ \text{E}_{\psi_0}[a_n] = (1/\sqrt{m}) \sum_{i=1}^{m} \text{E}_{\psi_0}[\nabla l(y_i, \psi_0)] = 0 \]

\[ \text{Var}[a_n] = (1/m)^2 \text{I}(\psi_0) \]

\[ B_n = -(1/m) \sum_{i=1}^{m} \nabla^2 l(y_i, \psi_0) \]

\[ \text{E}_{\psi_0}[C_n] = -(1/2m) \text{E}_{\psi_0}[\text{H}(y_i, \psi_0^*)] \]

\[ = -(1/2m) \text{E}_{\psi_0}[\text{H}(y_i, \psi_0^*)] \]

Thus, (34) can be written as

\[ \{B_n + [(\hat{\psi}_n - \psi_0)^T \otimes (\nabla l(y, \psi_0))^T \} \nabla l(y, \psi_0) = a_n \quad (35) \]

By applying the Law of Large Numbers (LLN), it can be shown that

\[ a_n \xrightarrow{p} 0, \]

\[ B_n \xrightarrow{d} -\text{I}(\psi_0), \]

\[ C_n \xrightarrow{d} -(1/2m) \text{E}_{\psi_0}[\text{H}(y_i, \psi_0^*)]. \quad (36) \]

Next, it will be seen that \( \| \text{H}(y, \psi_0^*) \| \) is bounded in probability. Let \( c_0 \) be a constant and the \( \| \hat{\psi}_n - \psi_0 \| < c_0 \) implies

\[ \| \hat{\psi}_n - \psi_0 \| < c_0 \] such that,

\[ -(1/m)\text{H}(y, \psi_0^*) \leq (1/m)\sum_{i=1}^{m} \text{H}(y_i, \psi_0^*) \leq (1/m)\sum_{i=1}^{m} \text{H}(y_i, \psi_0^*) \]

(37)

By condition (R2) \( \| \text{E}_{\psi_0} \text{H}(y_i, \psi_0^*) \| \leq \infty \) and by applying the Law of Large Numbers (LLN), it can be shown that

\[ (1/m)\sum_{i=1}^{m} \text{H}(y_i, \psi_0^*) \xrightarrow{d} \text{E}_{\psi_0} \text{H}(y, \psi_0^*). \]

Let \( \varepsilon > 0 \) be given and we select \( 1 + \| \text{E}_{\psi_0} \text{H}(y_i, \psi_0^*) \| \). Choose \( M_1 \) and \( M_2 \) so that,

\[ m \geq M_1 \Rightarrow P[\| \hat{\psi}_n - \psi_0 \| < c_0] \geq 1 - \varepsilon/2 \quad (38) \]

\[ m \geq M_2 \Rightarrow P\left[ \| 1/(m)\sum_{i=1}^{m} \text{H}(y_i, \psi_0^*) - \text{E}_{\psi_0} \text{H}(y, \psi_0^*) \| < 1 \right] \geq 1 - \varepsilon/2. \quad (39) \]

It follows from (37), (38) and (39) that,
\[ m \geq \max \{ M_1, M_2 \} \Rightarrow \]
\[ P[1/(1/m)H(y, \psi_n) \leq 1 + \frac{1}{2} \|E_n[H(y, \psi_n)]\|] \geq 1 - \epsilon/2 \]
hence, \( \|1/(1/m)H(y, \psi_n)\| \) is bounded in probability.

Finally, to show that \( \hat{\psi}_n - \psi_0 \) converges with probability 1 to \( \psi_0 \), we have \( \|H(y, \psi_n)\| \) is bounded in probability and by Eq. (35) can be concluded that \( (\hat{\psi}_n - \psi_0) \overset{w}{\rightarrow} 0 \) for all \( \psi_0 \in \Theta \). The sequence of solutions \( \{\hat{\psi}_n\} \) satisfying strong consistency \( \hat{\psi}_m \overset{w}{\rightarrow} \psi \) as \( m \rightarrow \infty \) is proven.

Furthermore, by using the Central Limit Theorem (CLT), we have
\[ a_n \sim AN(0, I(\psi_n)) \] (40)
By (36):
\[ B_n \overset{d}{\rightarrow} I(\psi_n) \]
By (40):
\[ \sqrt{m}(\hat{\psi}_n - \psi_0) \overset{d}{\rightarrow} AN(0, I^{-1}(\psi_n)) \]
This expression can be written as
\[ \psi_n - \psi_0 \overset{d}{\rightarrow} AN(0, (1/m)I^{-1}(\psi)) \]
and
\[ \hat{\psi}_n \overset{d}{\rightarrow} AN(\psi, (1/m)I^{-1}(\psi)) \] [3], [14].

Now, let a sequence \( \{\hat{\psi}_n\} \) of solutions to the likelihood function (23) where
\[ \hat{\psi}_n = [\text{vec}(\hat{\Sigma}_n), (\text{vech} \hat{V}_n)] = [\hat{\eta}_1, \ldots, \hat{\eta}_{s(n,s+1)/2}] \]
By using the above Theorem, \( \hat{\psi}_n \) will be convergence with probability 1 to \( \psi \) or strong consistency \( \hat{\psi}_m \overset{p}{\rightarrow} \psi \) as \( m \rightarrow \infty \) and also efficiency and asymptotic normality distribution, i.e.
\[ \sqrt{m}(\hat{\psi}_n - \psi_0) \overset{d}{\rightarrow} AN(0, I^{-1}(\psi)) \]
\[ \psi_0 \overset{d}{\rightarrow} AN(\psi, n^{-1}I^{-1}(\psi)) \]

V. CONCLUSION

This paper has discussed about estimation of multivariate linear mixed model or multivariate component of variance model with equal number of replications. The results show that the parameter estimation of fixed effects yields unbiased estimators, whereas the estimation for random effects or variance components yields biased estimators. Moreover, it is assumed that both likelihood and \( L_1 \)-likelihood functions hold some of regularity conditions, it can be proved that estimators as a solution set of the likelihood equations satisfy strong consistency for large sample size, asymptotic normal and efficiency.

Based on the discussion at the previous section, it can be drawn some theoretical conclusions as follows:

1). The estimator of parameters by MLE method in the linear mixed linear model \( Y = XB + ZD + E \) with the variance \( V = (Z \otimes I_j) \Phi (Z \otimes I_j) \) is
\[ \text{vec} (\hat{B}^\top) = [\text{vec}(\hat{\Sigma}_n), (\text{vech} \hat{V}_n)] \]
\[ = (Z \otimes I_j) \Phi (Z \otimes I_j) \] and
\[ \text{Var} (\text{vec} (\hat{B}^\top)) = [(Z \otimes I_2) \Phi (Z \otimes I_j)]^{-1} \cdot (Z \otimes I_j) \Phi (Z \otimes I_j) \]

2). Let a sequence \( \{\hat{\psi}_n\} \), which \( \hat{\psi}_n = [\text{vec}(\hat{\Sigma}_n), (\text{vech} \hat{V}_n)] \) then by applying the proposed theorem it can be shown that \( \hat{\psi}_n \) is strong consistency, i.e.
\[ \hat{\psi}_n \overset{w}{\rightarrow} \psi \] and \( \hat{\psi}_n \overset{d}{\rightarrow} AN(\psi, n^{-1}I^{-1}(\psi)) \]

REFERENCES