Parameter Estimates in Random Intercept Mixed Effects Model for Repeated Measures

Yan SUN
School of Economics, Shanghai University of Finance and Economics, Shanghai 200433, P. R. China
E-mail: yansun02@hotmail.com

Gen Xiang CHAI
Department of Applied Mathematics, Tongji University, Shanghai 200092, P. R. China
E-mail: chaigx03@hotmail.com

Abstract In this article the following random intercept mixed effects model will be considered:

\[ y_{ij} = v_i + x_{ij}^T \beta + \epsilon_{ij}, \quad i = 1, \ldots, m; j = 1, 2, \ldots, n_i, \]

where \( \{v_i\} \) are i.i.d. random effects with mean \( \alpha \) and finite variance \( \sigma_v^2 \); \( \{\epsilon_{ij}\} \) are i.i.d. random errors with finite variance \( \sigma_\epsilon^2 \). Here we will estimate \( \alpha, \sigma_v^2, \sigma_\epsilon^2, \beta \) and study their large sample properties, such as strong consistency, strong convergence rates and asymptotic normality.

Keywords Repeated measures, Random effects, Convergence system, Strong convergence, Strong convergence rate

MR(2000) Subject Classification 62F10, 62F12

1 Introduction

“Repeated measurements” is a term used to describe data in which the response variable for each experimental unit is observed on multiple occasions and possibly under different experimental conditions. There are a lot of repeated measurements in surveys and experiments. For example, Laird, et al. [1] reported a subset of data for 537 children from Ohio, examined annually from age 7 to 10 for the presence or absence of respiratory infection in the survey of Air Pollution and Health.

The linear mixed effects model considers both the common and the difference among individuals. So there are a variety of disciplines which make use of it such as health and life sciences, social sciences and industrial designs.

In this article we will consider the following random intercept mixed effects model for repeated measurements:

\[ y_{ij} = v_i + x_{ij}^T \beta + \epsilon_{ij}, \quad i = 1, 2, \ldots, m; j = 1, 2, \ldots, n_i, \]

where \( y_{ij} \) represents a response variable and \( x_{ij} \in \mathbb{R}^p \) is a known design vector for observation \( j(i = 1, \ldots, m; j = 1, 2, \ldots, n_i) \) on subject \( i(i = 1, \ldots, m) \); \( \beta \) is a \( p \times 1 \) vector of population parameters; \( \{v_i\} \) are i.i.d. random effects with expectation \( \alpha \) and finite variance \( \sigma_v^2 \); the random errors \( \{\epsilon_{ij}\} \), independent of \( v_i \), are i.i.d. with zero mean and finite variance \( \sigma_\epsilon^2 \). Model (1) is also called the error components or variance components model (see Hsiao [2]), the primary objective is devoted to analysing the estimators of parameters \( \alpha, \beta, \sigma_v^2, \sigma_\epsilon^2 \). In the following we suppose that \( \{n_i\} \) is a bounded set and let \( n = \sum_{i=1}^m n_i \).
Arnold [3] studied the simpler balanced one-way mixed effects model: \( y_{ij} = \theta + \alpha_i + \epsilon_{ij}, i = 1, \ldots, k; j = 1, \ldots, n_i \), where \( \alpha_i \) and \( \epsilon_{ij} \) are independent, unobserved, random variables such that \( \epsilon_{ij} \) i.i.d. \( \sim N(0, \sigma^2) \); \( \epsilon_{ij} \) \( \sim N(0, \sigma^2) \), and used the method of complete sufficient statistics to establish the unbiased estimators of \( \sigma^2_\alpha \). Battese, Harter and Fuller [4] proposed model (1) in the context of estimating the mean acreage under a crop for countries (small areas) in Iowa, under the assumption of random effects \( v_i \) i.i.d. \( \sim N(0, \sigma^2_v) \), independent of the \( \epsilon_{ij} \), which are assumed to be i.i.d. \( \sim N(0, \sigma^2_\epsilon) \). Using the general theory of Henderson [5] for a linear mixed effects model, a two-stage estimator of the linear combination of the fixed effects \( \beta \) and the realized value of the random effect is obtained. Prasad and Rao [6] obtained the second order approximation to the mean square error (MSE) of the two-stage estimator and the estimator of the MSE approximation under normality. Furthermore they used a Monte Carlo study to summarize the efficiency of two-stage estimators and the accuracy of the proposed approximation to MSE. Normality of random effects is a routine assumption for the linear mixed effects model, but it may be unrealistic, obscuring important features of among-individual variation and may have a great influence upon the effect of some parameters’ estimators, especially the estimated random effect for each individual (see Lange and Ryan [7]).

Tao, et al. [8] used a predictive recursive method to provide a nonparametric smooth density estimate of the random effects under Gaussian errors, but it lacked of theoretic analysis. Qian and Chai [9] gave the explicit formulae of the estimator of the unknown parameters; the random effect density and obtained some statistical properties under the assumption of Gaussian errors.

Different from the above works, in this article we will establish the explicit formulae of unbiased estimators of \( \alpha, \beta, \sigma^2_\alpha, \sigma^2_\epsilon \) without assumption of distributions of \( v_i \) and \( \epsilon_{ij} \), and study the large sample properties more thoroughly such as the strong consistency of \( \alpha, \beta, \sigma^2_\alpha, \sigma^2_\epsilon \), the strong consistency rate and the approximate normality of some estimators.

2 Main Results

From model (1) we get \( \bar{y}_i = \bar{x}_i^\top \beta + v_i + \bar{\epsilon}_i \), where \( \bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \), \( \bar{\epsilon}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \epsilon_{ij} \). Thus model (1) can be changed to

\[
y_{ij} = \bar{y}_i = (x_{ij} - \bar{x}_i)^\top \beta + \epsilon_{ij} - \bar{\epsilon}_i. \tag{2}
\]

Let \( y_{ij}^* = y_{ij} - \bar{y}_i \), \( x_{ij}^* = x_{ij} - \bar{x}_i \), \( \gamma_{ij} = \epsilon_{ij} - \bar{\epsilon}_i \), \( Y^*_i = (y_{i1}^*, \ldots, y_{in_i}^*)^\top \), \( X^*_i = (x_{i1}^*, \ldots, x_{in_i}^*)^\top \), \( \gamma_i = (\gamma_{i1}, \ldots, \gamma_{in_i})^\top \), \( Y^* = (Y^*_1, \ldots, Y^*_m)^\top \), \( X^* = (X^*_1, \ldots, X^*_m)^\top \), \( \gamma = (\gamma^*_1, \ldots, \gamma^*_m) \). So (2) can be rewritten as

\[
y_{ij}^* = x_{ij}^\top \beta + \gamma_{ij}, \tag{3}
\]

where \( E\gamma_{ij}^* = (1 - \frac{1}{m_i})\sigma^2_\epsilon \), \( E\gamma_{ij}\gamma_{il} = -\frac{1}{m_i} \sigma^2_\epsilon, j \neq l \); \( E\gamma_{ij}\gamma_{lk} = 0, i \neq l \).

In the following we suppose \( \text{rk}(X^*) = p \). Hence the L.S. estimator of \( \beta \) is

\[
\hat{\beta}_n = (X^* X^*)^{-1} X^* Y^* = \left( \sum_{i=1}^{m} \sum_{j=1}^{n_i} x_{ij}^* x_{ij}^\top \right)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n_i} x_{ij}^* y_{ij}^*. \tag{4}
\]

Refer to Henderson [5], and we can obtain an estimator of \( \sigma^2_\epsilon \) as follows:

\[
\hat{\sigma}^2_{\epsilon n} = \frac{1}{n - m - p} \sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - x_{ij}^* \hat{\beta}_n)^2. \tag{5}
\]

Notice that \( (n - m - p)\hat{\sigma}^2_{\epsilon n} \) can also be expressed as

\[
(n - m - p)\hat{\sigma}^2_{\epsilon n} = Y^* (I_n - Q) Y^* = \gamma^\top (I_n - Q) \gamma,
\]

where \( I_n \) is an identity matrix of order \( n \), \( Q = X^*(X^* X^*)^{-1} X^* \), and it is obvious that \( Q \) is a projection matrix.

Let \( \bar{v}_i = v_i - \alpha \), \( \bar{x}_{ij} = (1, x_{ij}^\top) \), \( \theta = (\theta^\top) \), \( e_{ij} = \bar{v}_i + \bar{\epsilon}_i \). Then (1) can be represented as

\[
y_{ij} = \bar{x}_{ij}^\top \theta + e_{ij}, \quad i = 1, 2, \ldots, m; j = 1, 2, \ldots, n_i. \tag{6}
\]

where \( Ee_{ij} = 0, Ee_{ij}^2 = \sigma^2_\epsilon + \sigma^2_\alpha, Ee_{ik} e_{il} = \sigma^2_\epsilon, k \neq l \), \( Ee_{ij} e_{kl} = 0, i \neq k \).