Strong Consistency of MLE in Nonlinear Mixed-effects Models with large cluster size

Lei Nie and Min Yang
University of Maryland, Baltimore county
University of Nebraska Lincoln

Abstract

The search for conditions for the consistency of maximum likelihood estimators in nonlinear mixed effects models is difficult due to the fact that, in general, the likelihood can only be expressed as an integral over the random effects. For repeated measurements or clustered data, we focus on asymptotic theory for the maximum likelihood estimator for the case where the cluster sizes go to infinity, which is a minimum assumption required to validate most of the available methods of inference in nonlinear mixed-effects models. In particular, we establish sufficient conditions for the (strong) consistency of the maximum likelihood estimator of the fixed effects. Our results extend the results of Jennrich (1969) and Wu (1981) for nonlinear fixed-effects models to nonlinear mixed-effects models.

Running title: consistency of MLE.

Key Words and Phrases: Maximum likelihood estimator (MLE); Nonlinear models; Random effects; Strong consistency.

AMS subject Classification: Primary 62J05, 62F12.

1 Introduction

Nonlinear mixed models are increasingly used in studies of complex phenomena. Applications include biological growth studies and pharmacokinetics. For instance, tumor growth as a function of time is often nonlinear in the fixed effects parameters, as well as the random effects which would have to be included in order to accommodate the variations among subjects in the study. Davidian and Giltinan (1993) gives two examples. One is a study of the growth pattern differences between two soybean genotypes, while the other is a dose-response study for standard concentrations of a bioassay for the therapeutic protein relaxin. Other examples in pharmacokinetics are available in Mentré and Gomeli (1995), Pinheiro and Bates (1995), Roe (1997), Vonesh and Chinchilli (1997), Wolfinger and Lin (1997), and Vonesh, Wang, Nie, and Majumdar (2002).

Data for these experiments may be typically partitioned into n clusters, where the ith cluster consists of p_i observations,

$$y_i = (y_{i1}, ..., y_{ip_i})^T.$$

For instance, all observations on a subject would constitute a cluster. We consider the following model for the observations,

$$y_i = f_i(\theta, b_i) + \epsilon_i, \tag{1}$$

where $f_i(\theta, b_i) = (f_{i1}(\theta, b_i), \cdots, f_{ip_i}(\theta, b_i))^T$ is a vector of p_i (possibly nonlinear) functions, $\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{ip_i})^T$ is a vector of p_i random errors, and $b_i = (b_{i1}, \cdots, b_{iv})^T$ is a vector of random effects.

Although the problem can be framed in general terms, due to the complexity of the computations involved, we will focus on a particular setup that would cover many real world experiments where the measurements are on a continuous scale. Given b_i, we assume that the random errors $\epsilon_{i1}, \cdots, \epsilon_{ip_i}$ are distributed as i.i.d normal with mean 0 and variance σ^2, for $i = 1, \cdots, n$. Also, the random vectors $\epsilon_1, ..., \epsilon_n$ are independent. To complete the specification of the model, we assume that the unobservable random-effect vectors $b_1, ..., b_n$ are i.i.d. with cumulative distribution $F(b)$.

In this paper we investigate conditions for the consistency of the Maximum Likelihood Estimator (MLE) of θ in model (1) as $\min(p_i) \to \infty$ while n is fixed. Our results can be viewed as an extension of the results of Jennrich (1969) and Wu (1981) for nonlinear fixed-effects models to nonlinear mixed-effects models. Jennrich (1969) established the consistency of the least squares estimator for the nonlinear regression model,

$$y_i = f_i(\theta) + \varepsilon_i, \quad i = 1, ..., p, \tag{2}$$

where $f_i(\theta) = f(x_i, \theta)$, the $x_1, \ldots, x_p$ are known values of covariates, θ is an unknown (fixed) parameter, $f_i(\theta)$ is continuous in $\theta \in \Theta$, where $\Theta \subset R^q$ is compact and $\varepsilon_1, \ldots, \varepsilon_p$ are i.i.d. errors with mean zero and common variance $\sigma^2 > 0$. In general, $f_i(\theta)$ is nonlinear in θ. Jennrich (1969) showed that, under certain conditions, the least squares estimator, i.e., $\widehat{\theta}_p$ that minimizes $\sum_{i=1}^{p} (y_i - f_i(\theta))^2$, is strongly consistent, in other words, $\hat{\theta}_p \to \theta$ almost surely. The main condition imposed by Jennrich was,

$$p^{-1} D_p(\theta, \theta') \longrightarrow D(\theta, \theta'), \text{ uniformly, for a continuous } D(\theta, \theta'),$$

where $D_p(\theta, \theta') = \sum_{i=1}^{p} (f_i(\theta) - f_i(\theta'))^2$. This condition has been weakened by Wu (1981). We refer the reader to Wu (1981) for the exact set of conditions that he imposed, and also for further literature on this topic. Note that, if the random errors in model (2) are normally distributed, then the results of Jennrich (1969) and Wu (1981) establish the consistency of the MLE of θ. Wu (1981) provides one of the best strong consistency result for model (2). Bose and Sengupta (2003) is the another article which provides one of the best strong consistency result for model (2). Please refer to Sen and Sengupta (2003) for comparisons of these two excellent works. In this paper, we shall follow Wu (1981)'s assumptions and proofs. In the discussion section, we will comments on the usefulness of Bose and Sengupta (2003) to future work in this area.

The study of the consistency of the MLE of θ in model (1) is important for two reasons. The first one is obvious - in order to apply the MLE, we need to specify conditions under which its asymptotic properties hold. The second reason is more subtle. For models of the type (1) where $f_i(\theta, b_i)$ is nonlinear in b_i, the likelihood for θ is obtained only as an integral over the b_i's. Hence, computation of the MLE is generally difficult. While the latest version of SAS has a procedure to do this, it works well only when the dimension of b_i is no more than 2 due to computational complexity. As a result, θ is generally estimated using one of several alternative estimators proposed in the literature. For a general description of many such available estimators, the reader is referred to Davidian and Giltinan (1995a) and Vonesh and Chinchilli (1997). In essence, these estimators seek to approximate the MLE. Consistency of these estimators are generally established under two sets of conditions that allow (A) the proposed estimator to be a good approximation of the MLE, and (B) the MLE itself to be consistent, as in Vonesh (1996), Vonesh, Wang, Nie, and Majumdar (2002) and Nie (2002). It is the step (B) that is our concern here.

The consistency of the MLE may be studied under three different situations: (I) $n \to \infty$; (II) $\min_{1 \le i \le n}(p_i) \to \infty$; (III) $n \to \infty$ and $\min_{1 \le i \le n}(p_i) \to \infty$. The situation (I), $n \to \infty$ while p_i are bounded, has been considered in Nie (2004). While the situation (III) in which both n and p_i are allowed to grow, is technically the most difficult. This will be considered in the future. In this paper, we consider situation (II), i.e., n is finite, but $\min(p_i) \to \infty$. It should note that the focus and proof here are completely different than that in Nie (2004). Nie (2004) follows closely classical approach such as Cramer (1946), Wald (1949), Bradley and Gart (1962), and Hoadley (1971). The consistency as $n \to \infty$ is considered in all of these articles. In this paper, since n is fixed, we consider the consistency as $p \to \infty$.

Note that one approach to developing alternatives to the MLE is the use of the Laplace approximation to evaluate the integrated likelihood, as in Breslow and Clayton (1993), Vonesh (1996), Vonesh, Wang, Nie, and Majumdar (2002) and Nie (2002). It is interesting to note that the validity of the Laplace approximation depends on the assumption: $\min_{1 \le i \le n}(p_i) \to \infty$. Thus it is theoretically of interest to consider situation (II).

There are also real life experiments where n is small, but $p_i' s$ are large. For example, an animal experiment with n animals that receive a treatment at a specific time each day is restricted by the condition that to eliminate bias the same technician should administer the treatment to all animals at approximately the same time. This means that n cannot be very large, but p_i, the number of days, is not necessarily restricted. One such experiment, with two groups of $n = 5$ animals each treated at $p_i = 12$ time periods, is reported in Yamada et al. (2002). Another examples is the soybean growth model, (Davidian and Giltinan 1995b). 8 plots of lands were planted with soybean seeds. Each plant were randomly sampled 10 times. At each time, 6 plants were randomly selected and their weights are recorded as observations. In this example, $n = 8$ and $p = 60$. Other examples can be found in Lindstrom and Bates (1990) and (Nielsen, Ritz, and Streibig 2004).

The main technical difficulty in establishing the consistency of the MLE for situation n is finite, but $p_i \to \infty$ arises from the fact that we have to handle a likelihood function that is expressed as an integral over the b_i's. The proof in this paper provides a method to deal with the integrated likelihood and we expect this method will become a tool to overcome the difficulty due to the integrated likelihood.

If there are no random effects b_i in model (1) then no integration is involved. If $n = 1$, then the problem reduces to the problem considered by Jennrich (1969) and Wu (1981). Before we consider the case of a general (finite) n, we first extend the results of Jennrich (1969) and Wu (1981) to the nonlinear mixed-effects model (1) with $n = 1$. This model may be written as,

$$y_i = f_i(\theta, b) + \varepsilon_i, \quad i = 1, ..., p, \tag{3}$$

where $\theta \in \Theta \subset R^q$ is compact, b is a random variable with distribution function $F(b)$, $\varepsilon_1, \ldots, \varepsilon_p$ are $i.i.d$ $N(0,1)$ and they are independent of b. The function $f_i(\theta, b)$ is possibly nonlinear in θ and b. For $i = 1, \ldots, p$, $f_i(\theta, b) = f(\theta, b, x_i)$, where $x_1, \ldots, x_p$ are known values of covariates.

For this model we establish the strong consistency of the MLE of θ under certain conditions. In particular, we establish the Wald consistency of θ. Note that this is stronger than Cramer consistency (see Le Cam (1979), Jiang (1997)). Our result is proved separately for the cases: (a) the random effect b is discrete with finite support, (b) the random effect b is discrete with countable support and (c) the random effect b is continuous. For case (a), our main condition is similar to the main condition in Theorem 3 of Wu (1981). For cases (b) and (c), our main condition is similar to the main condition of Jennrich (1969) that we stated earlier. The consistency in the more general setup where the variance of ε_i is σ^2 will be briefly discussed in Remark 2 at the end of Section 2.

The results for model (3) are stated in Section 2. In Section 3 we apply the result to an example. In Section 4 we consider extensions to repeated measurements model (1). In Section 5 we conducted a simulation study. All proofs are relegated to the Appendix.

2 Consistency Results for Model (3)

In this section we investigate the consistency of the MLE of θ in model (3). We will do this separately for the case where the random effect b is discrete with finite support and the case where b is continuous or discrete with countable support. For each case a theorem is established that gives sufficient conditions for the consistency of the MLE.

Throughout we assume that the parameter space $\Theta \in \mathbb{R}^q$ is a compact set, and θ_0, the true value of θ, is an interior point of Θ. The support of b will be denoted by Ω^b, i.e. $F(\Omega^b) = 1$.

In the first theorem b is a discrete random variable with finite support, i.e. $\Omega^b = \{b_0, b_1, \ldots, b_K\}$.

Theorem 1. *Consider model (3). Let θ_0 denote the true value of the fixed-effects parameter θ. Suppose b has a discrete distribution with finite support: $P(b = b_j) = w_j$, $w_j \geq 0, j = 0, 1, \ldots, K, K < \infty$. $\sum_{j=0}^{K} w_j = 1$. Suppose the following conditions hold:*

(i) For any θ ($\theta \neq \theta_0$), b_j, b_l, there exist a neighborhood of θ, $O(\theta)$, such that:

(a) For some $c > 0$,

$$\limsup_{p\to\infty} \frac{\{\sum_{i=1}^{p} \sup_{\theta' \in O(\theta)} (f_i(\theta', b_j) - f_i(\theta_0, b_l))^2\}^{\frac{1+c}{2}}}{\inf_{\theta' \in O(\theta)} \sum_{i=1}^{p} (f_i(\theta', b_j) - f_i(\theta_0, b_l))^2} < \infty. \tag{4}$$

(b) $f_i(\theta, b_j)$ *is a Lipschitz function on* $O(\theta)$ *and*

$$\sup_{\theta', \theta'' \in O(\theta)} \frac{|f_i(\theta', b_j) - f_i(\theta'', b_j)|}{|\theta' - \theta''|} \leq M \sup_{\theta' \in O(\theta)} |f_i(\theta', b_j)|, \tag{5}$$

where M *is independent of* i*, but may depend on* $O(\theta)$*.*

(ii) For any θ $(\theta \neq \theta_0), b_j, b_l$,

$$\lim_{p \to \infty} \sum_{i=1}^{p} [f_i(\theta_0, b_j) - f_i(\theta, b_l)]^2 = +\infty. \tag{6}$$

Then $\hat{\theta}_p \to \theta_0$ *a.e.*

In the second theorem, b is a continuous random variable.

Theorem 2. *Consider model (3). Let* θ_0 *denote the true value of the fixed-effects parameter* θ*. Suppose* b *has a continuous distribution function* $F(b)$*. Suppose that for any given* $b_0 \in \Omega^b$*, and a fixed* $\theta \in \Theta$*,* $\theta \neq \theta_0$*, and* $\epsilon > 0$*, there exists a neighborhood of* θ*,* $O(\theta)$*, such that the following conditions hold:*

(i)The support of b *can be written as,* $\Omega^b = \cup_{l=1}^{K} B_l$*, where* K *is finite and* B_l *has either of the following two properties:*

(a)There exist b_l*, such that for* $b \in B_l$*,*

$$\sup_i \sup_{\theta' \in O(\theta)} \sup_{b \in B_l} |f_i(\theta', b) - f_i(\theta, b_l)| < \epsilon. \tag{7}$$

(b)For $b \in B_l$*, one of the following inequalities is satisfied for each* $i = 1, \ldots, p$*,*

$$\inf_i \inf_{\theta' \in O(\theta)} \inf_{b \in B_l} f_i(\theta', b) > f_i(\theta_0, b_0) + 2, \tag{8}$$

$$\sup_i \sup_{\theta' \in O(\theta)} \sup_{b \in B_l} f_i(\theta', b) < f_i(\theta_0, b_0) - 2. \tag{9}$$

(ii)There exists a neighborhood of b_0*,* $O(b_0)$*, such that,*

$$\sup_i \sup_{b \in O(b_0)} |f_i(\theta_0, b) - f_i(\theta_0, b_0)| < \epsilon. \tag{10}$$

(iii)There exists a nonnegative continuous function $D(\theta_0, \theta, b_0)$*, which is* 0 *if and only if* $\theta = \theta_0$*, such that,*

$$\inf_{b \in \Omega^b} \lim_{p \to +\infty} p^{-1} \sum_{i=1}^{p} [f_i(\theta_0, b_0) - f_i(\theta, b)]^2 \to D(\theta_0, \theta, b_0). \tag{11}$$

Then $\hat{\theta}_p \to \theta_0$ *a.e.*

The case of a discrete b with countable support, $\Omega^b = \{b_0, b_1, \ldots\}$, maybe treated along the same lines as a continuous b. This is stated (without proof) in the following Corollary.

Corollary 1. *Consider model (3). Let θ_0 denote the true value of the fixed-effects parameter θ. Suppose b has a discrete distribution with countable support. For a given b_0, a fixed $\theta \in \Theta$, $\theta \neq \theta_0$, and $\epsilon > 0$, suppose there exist a neighborhood of θ, $O(\theta)$, such that the following conditions hold:*

(i)The support set of b can be written as: $\Omega^b = \cup_{l=1}^{K} B_l$, $K < \infty$, where B_l is either a finite or an infinite set.

(a)If B_l is finite, then for each $b \in B_l$,

$$\sup_i \sup_{\theta' \in O(\theta)} |f_i(\theta', b) - f_i(\theta, b)| < \epsilon.$$

(b)If B_l is infinite, then either there exists a $b_l \in B_l$, such that,

$$\sup_i \sup_{\theta' \in O(\theta)} \sup_{b \in B_l} |f_i(\theta', b) - f_i(\theta, b_l)| < \epsilon,$$

or one of the following inequalities is satisfied,

$$\inf_i \inf_{\theta' \in O(\theta)} \inf_{b \in B_l} f_i(\theta', b) > f_i(\theta_0, b_0) + 2,$$

$$\sup_i \sup_{\theta' \in O(\theta)} \sup_{b \in B_l} f_i(\theta', b) < f_i(\theta_0, b_0) - 2.$$

(ii)There exists a nonnegative continuous function $D(\theta_0, \theta, b_0)$, which is 0 if and only if $\theta = \theta_0$, such that,

$$\inf_{b \in \Omega^b} \lim_{p \to +\infty} p^{-1} \sum_{i=1}^{p} [f_i(\theta_0, b_0) - f_i(\theta, b)]^2 \to D(\theta_0, \theta, b_0).$$

Then $\hat{\theta}_p \to \theta_0$ a.e.

Remark 1: Among the conditions of the theorem, (11) is the most demanding. This is akin to a strong identifiability condition. While this condition may not be necessary in general, there are examples where consistency fails when the condition is not satisfied. For instance, let

$$y_i = \theta + x_i b + \epsilon_i, \quad i = 1, \ldots, p,$$

where b is a random effect with distribution $N(0, 1)$, $\epsilon = (\epsilon_i, \ldots, \epsilon_p)'$ is $N_p(0, I)$, b and ϵ are independent, and $x_1, \ldots, x_p$ are known values of covariates. Clearly, conditions (7), (8), (9) and (10) hold. Condition (11) holds, unless $x_1 = \ldots = x_p$, thus our theorem can be applied whenever there are at least two distinct x_i's. In fact, it is easy to verify that the MLE of θ is not consistent for this special case $x_1 = \ldots = x_p$. Intuitively, θ and b are confounded.

The conditions of Theorem 2 can be simplified considerably when $f_i(\theta, b)$ is continuous in θ and b uniformly in i. This is stated in the following corollary.

Corollary 2. *Consider model (3). Let θ_0 denote the true value of the fixed-effect parameter θ. Suppose the random effect b has a continuous distribution function $F(b)$ and $f_i(\theta, b)$ is continuous in θ and b uniformly in i. For $b_0 \in \Omega^b$, for any $\theta \in \Theta$, $\theta \neq \theta_0$, and $\epsilon > 0$, suppose there exist a neighborhood of θ, $O'(\theta)$, such that the following conditions hold:*

(i)The support set of b can be written as, $\Omega^b = \cup_{l=1}^{K} B_l$, K is finite, and B_l is either compact or open. If B_l is open, it satisfies either condition (a) or condition (b) in Theorem 2.

(ii)There exists a nonnegative continuous function $D(\theta_0, \theta, b_0)$, which is 0 if and only if $\theta = \theta_0$, such that,

$$\inf_{b \in \Omega^b} \lim_{p \to +\infty} p^{-1} \sum_{i=1}^{p} [f_i(\theta_0, b_0) - f_i(\theta, b)]^2 \to D(\theta_0, \theta, b_0).$$

Then $\hat{\theta}_p \to \theta_0$ a.e.

It follows from Corollary 2 that when $f_i(\theta, b)$ is continuous in θ and b uniformly in i, we only need to check the tail of the distribution of b in order to verify whether condition (i) of Corollary 2 holds. For example, when $\Omega^b = (-\infty, +\infty)$, the condition is satisfied if we can show that $\lim_{b \to \infty} f_i(\theta, b)$ exists (may be ∞), for $i = 1, \ldots, p$.

Remark 2: For the case, $V(\varepsilon) = \sigma^2 I$, $\sigma^2 \neq 1$, we can transform $y_i \to y_i/\sigma$, $f_i(\theta, b) \to f_i(\theta, b)/\sigma$, $\epsilon_i \to \epsilon_i/\sigma$ and use the same approach. All conditions in the theorems and corollaries of this section hold as long as $\sigma^2 > 0$.

3 Examples

Jennrich (1969) studied the model $y_i = \theta_1 e^{\theta_2 x_i} + \epsilon_i$, where θ_1 and θ_2 are fixed effects, as an example. Wu (1981) also examined this model. We consider the same model with the exception that θ_1 will be taken to be a positive random effect. In effect, we consider the model,

$$y_i = e^{\theta x_i + b} + \epsilon_i,$$

where θ is a fixed effect which ranges over a compact set Θ and b is random effect with support R^1. Following Jennrich (1969), we assume $x_1, x_2, \ldots$ is a bounded sequence of real numbers whose sample distribution function G_n approaches a distribution function G which is not degenerate. We assume that $\epsilon_1, \epsilon_2, \ldots$ are i.i.d. $N(0, \sigma^2)$. It is easy to check that $e^{\theta x_i + b}$ is continuous in θ and b uniformly in i. Also $e^{\theta x_i + b} \to +\infty$ uniformly in i when $b \to +\infty$ and $e^{\theta x_i + b} \to 0$ uniformly in i when $b \to -\infty$, since $x_1, x_2, \ldots$ are bounded. Hence condition (i) in Corollary 2 is satisfied. Let

$$Q(\theta_0, b_0, \theta, b) = \lim_{n \to +\infty} n^{-1} \sum_{i=1}^{n} [e^{\theta x_i + b} - e^{\theta_0 x_i + b_0}]^2.$$

In order to verify condition (ii) in Corollary 2, we need to show that $\inf_b Q(\theta_0, b_0, \theta, b) > 0$.

Let $R^1 = (-\infty, -M) \cup [-M, M] \cup (M, \infty)$, where $M > 0$ is chosen such that for each i, $|e^{\theta x_i + b} - e^{\theta_0 x_i + b_0}| > 1$ for $b > M$ and $|e^{\theta x_i + b} - e^{\theta_0 x_i + b_0}| > \frac{1}{2} e^{\theta_0 x_i + b_0}$ for $b < -M$.

If $b \in (-\infty, -M) \cup (M, \infty)$, it is clear that $\inf_b Q(\theta_0, b_0, \theta, b) > 0$. Now suppose that $b \in [-M, M]$. Since $Q(\theta_0, b_0, \theta, b)$ is continuous in b, there exist a $b^* \in [-M, M]$, such that $Q(\theta_0, b_0, \theta, b^*) = \inf_{b \in [-M,M]} Q(\theta_0, b_0, \theta, b)$. Hence, it is sufficient to show that $Q(\theta_0, b_0, \theta, b^*) > 0$.

As in Jennrich (1969) we write,

$$Q(\theta_0, b_0, \theta, b^*) = \int [e^{\theta x + b^*} - e^{\theta_0 x + b_0}]^2 dG(x)$$

This expression is zero if and only if $\theta x + b^* = \theta_0 x + b_0$ on a set of x with G-measure one. Since $\theta \neq \theta_0$, this can happen only when G is degenerate. By the Corollary 2, the strong consistency of the MLE of θ follows.

It is not difficult to see that this result can be generalized to the model:

$$y_i = f(x_i\theta + z_i b) + \epsilon_i, \qquad i = 1, \ldots, p.$$

Under general conditions on f, x_i, and z_i, the MLE of the fixed effects θ will be strongly consistent. A necessary condition on f is that $f(\alpha)$ is either convergent or $f(\alpha) \to \pm\infty$ as $\alpha \to \infty$. Examples are the random coefficient models (see, for example, Fahrmeir and Tutz (1994)).

4 Consistency Results for Model (1)

In this section we consider the nonlinear mixed-effects model for clustered data, or the model (1) in the introduction. This model may be written as,

$$y_{ij} = f_{ij}(\theta, b_i) + \epsilon_{ij}, \quad i = 1, \ldots, n, j = 1, \ldots, p_i. \tag{12}$$

Here, $b_1, \ldots, b_n$ are random. The vector $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{ip_i})'$ is $N_{p_i}(0, \sigma^2 I)$, and $\epsilon_1, \ldots, \epsilon_n$ are independent. Moreover, for each $i = 1, \ldots, n$, the random effect b_i has the distribution function $F_i(b)$. Corollary 2, applied to model (12) for each $i = 1, \ldots, n$, gives the following result. Here, Ω_i^b denotes the support of $b_i, i = 1, \ldots, n$.

Corollary 3. *Consider model (12), for a finite n. Let θ_0 denote the true value of the fixed-effects parameter θ. For given $b_{i0} \in \Omega_i^b$ and any fixed $\theta \in \Theta$, $\theta \neq \theta_0$, and $\epsilon > 0$, there exist a neighborhood of θ, $O(\theta)$, such that the following conditions hold:*

(1) Suppose for each i, b_i is a continuous random variable and $f_{ij}(\theta, b_i)$ are continuous in b_i and θ uniformly in j. Furthermore they satisfy condition (i) in Corollary 2.

(2)There exists a nonnegative continuous function $D(\theta_0, \theta, b_{10}, \ldots, b_{n0})$, which is 0 if and only if $\theta_0 = \theta$, such that,

$$\inf_{b_i \in \Omega_i^b} \lim_{p \to +\infty} p^{-1} \sum_i^n \sum_{j=1}^{p_i} [f_{ij}(\theta_0, b_{i0}) - f_{ij}(\theta, b_i)]^2 \to D(\theta_0, \theta, b_{10}, \ldots, b_{n0}) > 0,$$

where, $p = \sum_{i=1}^n p_i$. Then the MLE of θ, $\hat{\theta}_p$ is strongly consistent, i.e., $\hat{\theta}_p \to \theta_0$ a.e.

Our first example is the growth curve model, see for example Lindstrom and Bates (1990) and Davidian and Giltinan (1993),

$$y_{ij} = \frac{b_i}{1+\exp(\beta_1 x_{ij} + \beta_2)} + \epsilon_{ij},\ i = 1, \cdots, n,\ j = 1, \cdots, p$$

where $b_i \sim N(\mu, \psi)$, and $\epsilon_{ij} \sim N(0, \sigma^2)$. It is easy to prove in a similar way as in Section 3 that β_1 and β_2 are consistent if $x_{ij}'s$ is a random variable with sample distribution function G_n approaches a distribution function G which is not degenerate. Our second example is the following pharmacokinetic model considered by Roe (1997) and Wolfinger and Lin (1997):

$$y_{ij} = \log[\frac{10k_a[\exp(-k_i t_{ij}) - \exp(-k_a t_{ij})]}{v_i(k_a - k_i)}] + e_{ij},$$

where y_{ij} is the concentration of the drug in the bloodstream of subject i at time t_{ij}. The parameter k_a is the absorbtion rate constant, k_i is the subject-specific elimination rate constant, and v_i is the volume of distribution. If we consider k_i and v_i as random effects, then the MLE of the fixed effect k_a can be shown to be strongly consistent by verifying the conditions of Corollary 3. This verification involves substantial computations that will not be shown here.

Please note that, although β_1 and β_2 can be consistently estimated, other parameters such as μ and ψ in the growth curve model may not be consistently estimated. We would like to explain this phenomena through a simple balanced one way ANOVA model,

$$y_{ij} = b_i + \epsilon_{ij},$$

$i = 1, \cdots, n$ and $j = 1, \cdots, p$ with $b_i's$ independent $N(\mu, \psi)$ and $\epsilon_{ij}'s$ independent $N(0, \sigma^2)$. It can be shown that,

$$\hat{\mu}_{MLE} = \bar{y}_{..}, \quad \text{var}(\hat{\mu}_{MLE}) = \frac{\sigma^2}{pn} + \frac{\sigma_b^2}{n} = O(n^{-1})$$

$$\hat{\psi}_{MLE} = \frac{(1-1/n)MST - MSE}{p}, \quad \text{var}(\hat{\psi}_{MLE}) = \frac{2\sigma^4}{n(p-1)p^2} + \frac{2(\sigma^2 + p\psi)^2}{(n-1)p^2} = O(n^{-1})$$

where MSE and MST are the mean square error and mean square of treatments, respectively, $\hat{\mu}_{MLE}$, $\hat{\sigma}^2_{MLE}$ and $\hat{\sigma}^2_{bMLE}$ are MLE's of parameters μ, σ^2 and σ_b^2. It can be seen,

$$\hat{\mu}_{MLE} - \mu_0 = O_p(n^{-1/2}),\ \hat{\psi}_{MLE} - \psi_0 = O_p(n^{-1/2}).$$

In other word, in this simple example neither MLE of μ and σ_b^2 is consistent.

5 A simulation study

We conducted a simulation study for two models: one is the model given in Section 3; the other one is the first example given in Section 4, with reparametrization.

Table 1: Simulation results for decay model

Parameter	True Value		Estimates		
			n=1, p=50	n=1, p=100	n=1, p=200
α_1	-2.5	mean	-2.22	-2.40	-2.43
		std	0.92	0.71	0.61
σ^2	1	mean	1.0	0.992	0.995
		std	0.21	0.14	0.10
ψ	1	mean	100.3	6.75	1.39
		std	1159	54.3	2.18

5.1 The simple exponential model

The model for y_{ij} is given by,

$$y_{ij} = exp(\theta x_{ij} + b_i) + \epsilon_{ij},\ i = 1, \cdots, n,\ j = 1, \cdots, p,$$

where $b_i \sim N(\beta_1, \psi)$, and $\epsilon_{ij} \sim N(0, \sigma^2)$. Let

$$\begin{aligned} x10 &= (-0.5, -0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.4, 0.5)^T, \\ x50 &= \{x10^T, x10^T, x10^T, x10^T, x10^T\}^T, \\ x100 &= \{x50^T, x50^T\}^T, \\ x200 &= \{x100^T, x100^T\}^T. \end{aligned}$$

We simulated three cases. For each case, 200 data sets were created and parameters are $\alpha = -2.5$, $\sigma^2 = 1$, and $\psi = 1$. These parameters values are similar to values in the first example in (Vonesh, Wang, Nie, and Majumdar 2002). Covariates $x_i = (x_{i1}, \cdots, x_{ip})$ are as follows:

1 n=1, p=50, $x_i = x50$;

2 n=1, p=100, $x_i = x100$;

3 n=1, p=200, $x_i = x200$;

The results are given on Table 1. According to Corollary 3, $\hat{\alpha}_{MLE}$, the maximum likelihood of α is consistent; on the other hand, as explained in the last paragraph in Section 4, ψ may not be consistent. We can see that the bias of $\hat{\alpha}$ and variation are small, and they become smaller as the sample size p increases. On the other hand, the estimation of ψ are not very good in general.

5.2 The Soybean growth model

Let y_{ij} denote the weight of the soybean plant at time x_{ij}, measured in weeks, $i = 1, \cdots, n$, $j = 1, \cdots, p$. The model is given by,

$$y_{ij} = \frac{b_i}{1 + \exp\{\beta_3(x_{ij} - \beta_2)\}} + \epsilon_{ij},\ i = 1, \cdots, n,\ j = 1, \cdots, p$$

Table 2: Simulation results for growth model

Parameter	True Value		Estimates		
			n=2,p=10	n=2, p=20	n=2, p=50
β_1	15	mean	15.00	15.00	14.99
		std	0.58	0.56	0.52
β_2	50	mean	50.03	50.02	49.99
		std	0.43	0.31	0.21
β_3	-0.1	mean	-0.1	-0.1	-0.1
		std	0.0030	0.0022	0.0014
σ^2	0.04	mean	0.036	0.039	0.039
		std	0.013	0.009	0.006
ψ	4	mean	0.285	0.276	0.287
		std	0.41	0.40	0.42

Note*: these values are all very close to 0.

where $b_i \sim N(\beta_1, \psi)$, and $\epsilon_{ij} \sim N(0, \sigma^2)$.

Let

$$
\begin{aligned}
x10 &= (14, 21, 28, 35, 42, 49, 56, 63, 70, 77)^T, \\
x50 &= \{x10^T, x10^T, x10^T, x10^T, x10^T\}^T, \\
x100 &= \{x50^T, x50^T\}^T, \\
x200 &= \{x100^T, x100^T\}^T.
\end{aligned}
$$

We simulated three cases. For each case, 200 data sets were created and parameters are $\beta_1 = 15$, $\beta_2 = 50$, $\beta_3 = -0.1$, $\sigma^2 = 0.04$, $\psi = 4$. These parameters were chosen in a similar way as they were in (Davidian and Giltinan 1993). Covariates $x_i = (x_{i1}, \cdots, x_{ip})$ are as follows:

1 n=2, p=10, $x_i = x10$;

2 n=2, p=20, $x_i = x20$;

3 n=2, p=50, $x_i = x50$;

The results are given on Table 2. According to Corollary 3, $\hat{\beta}_{2MLE}$ and $\hat{\beta}_{3MLE}$, the maximum likelihood of β_2 and β_3 are consistent and β_1 and ψ may not be consistent. From Table 2, the biases of the $\hat{\beta}_{2MLE}$ and $\hat{\beta}_{3MLE}$ are very small, and variances of $\hat{\beta}_{2MLE}$ and $\hat{\beta}_{3MLE}$ reduce quickly when the sample size p increase. On the other hand, the variance of $\hat{\beta}_1$ decrease very slow; the estimation of $\hat{\psi}$ is poor and this is expected. The simulation results roughly support the theorems developed in this paper.

6 Discussion

In this article, we extended the results by Wu (1981) to nonlinear mixed-effects models. Our assumptions are related to those assumptions in Wu (1981). However, it is interesting to prove the strong consistency following the approach in Bose and Sengupta (2003), which could possible reduce some assumptions assumed in this paper. Furthermore, Bose and Sengupta (2003) provides an possibility to establish strong consistency of MLE for both nonlinear mixed-effects models and generalized linear mixed-effects models, using the unified approach proposed in Bose and Sengupta (2003). As we know, generalized linear mixed-effects models are equally important as nonlinear mixed effects models, which have been used extensively in literature.

Since the likelihood function of the nonlinear mixed-effects model is presented through the integration, existing approaches cannot be applied directly to prove the consistency of the MLE. We break the likelihood function into a finite number of components, then consider the corresponding aymptotic properties of each individual component. Consequently, conditions in Theorem 1 and 2 are quite complicated. They are sufficient conditions for consistency of the MLEs, while they may not be the necessary conditions.

Theorem 1 considers the case when the support of distribution of random effect is finite. The likelihood function is already the finite summation of some likelihood functions. Therefore, conditions on $f_i(\theta, b)$ and proofs are similar to those of the nonlinear model (1.1), Wu (1981). The condition (i) in Theorem 1 is similar to the assumption A$'$, Page 506, Wu (1981). Precisely, (a) likes $(3.5)'$ and (b) likes $(3.6)'$. The condition (ii) is similar to another condition (other than A or A$'$) of Theorem 3 in Wu (1981).

Theorem 2 considers the case when the support of distribution of random effect is not finite. For example the random effect has the standard normal distribution, where the support of the random effect is $[-\infty, \infty]$. Therefore, conditions on $f_i(\theta, b)$ are different from those of the nonlinear models, Jennrich (1969) and Wu (1981). Precisely, we decompose the support of the random effect into a union of sets. e.g $[-\infty, \infty]$ can be decomposed into a union of some compact sets, $[-\infty, B]$, and $[C, +\infty]$. Condition (i)(a) is used to handle the compact sets and Condition (i)(b) is used to handle the open sets $[-\infty, B]$, and $[C, +\infty]$. Both conditions guarantee that we can get ride of the integration without losing much information. Condition (ii) is similar to Condition (i)(a). It also makes sure that we can get rid of the integration by approximation without losing much information. As it was stated in Corollary 2, Condition (ii) and Condition (i) (a) are implied by the uniform continuous conditions of $f_i(\theta, b)$, $i = 1, \cdots, p$. The condition (iii) is similar to but stronger than the corresponding condition of Theorem 3 in Wu (1981). Please refer to Remark 1, for further comments on this condition.

There are special features for the asymptotic normality of the MLE for this type of models. For one of these features, MLEs for some parameters may not be consistent. As a consequence, asymptotic normality may only be considered for some parameters. Another special feature is that classical tools are not directly applicable for establishing asymptotic normality for the models we considered. Precisely, the law of large number theory and the central limit theorem can not be used directly here since the first and second derivatives of the loglikelihood functions are integrated functions in our cases. Further, the conditions for establishing the asymptotic normality will be complicated and the proof are quite involved. We are currently working on it and wish to fully explore it in our future work.

Appendix

Lemma 1. *Let* $\{X_i\}$ *be a sequence of independent random variables, with* $\sum_{i=1}^p E(X_i) \to -\infty$, $\sum_{i=1}^p V(X_i) \to \infty$ *and* $\lim_{p\to\infty} \frac{[\sum_{i=1}^p V(X_i)]^{\frac{1}{2}+\delta}}{|\sum_{i=1}^p E(X_i)|} = 0$*, for some* $\delta > 0$*. Then for any constant* C*, we have:*

$$P(\lim_{p\to\infty} \sum_{i=1}^p X_i \leq C) = 1.$$

Proof. Since $\sum_{i=1}^p V(X_i) \to \infty$, following Chung (1974) page 126, we have,

$$P(\lim_{p\to\infty} \frac{\sum_{i=1}^p (X_i - E(X_i))}{[\sum_{i=1}^p V(X_i)]^{\frac{1}{2}+\delta}} = 0) = 1.$$

Observe that if

$$\lim_{p\to\infty} \frac{\sum_{i=1}^p (X_i - E(X_i))}{[\sum_{i=1}^p V(X_i)]^{\frac{1}{2}+\delta}} < 1,$$

then $\exists\ N$ such that for $p > N$,

$$\sum_{i=1}^p X_i < \sum_{i=1}^p E(X_i) + 2[\sum_{i=1}^p V(X_i)]^{\frac{1}{2}+\delta}.$$

It follows from the conditions of lemma that, given any $C < 0$, $\exists\ M$ such that $\forall\ p > M$,

$$\sum_{i=1}^p E(X_i) < 2C < 0,$$

$$\frac{[\sum_{i=1}^p V(X_i)]^{\frac{1}{2}+\delta}}{|\sum_{i=1}^p E(X_i)|} < \frac{1}{4}.$$

So we have:

$$\sum_{i=1}^p E(X_i) + 2[\sum_{i=1}^p V(X_i)]^{\frac{1}{2}+\delta} < \frac{1}{2}\sum_{i=1}^p E(X_i) < C.$$

It follows that, for $p > \max(N, M)$,

$$\sum_{i=1}^p X_i < C,$$

and

$$P(\lim_{p\to\infty}\sum_{i=1}^{p} X_i \le C) \ge P(\lim_{p\to\infty} \frac{\sum_{i=1}^{p}(X_i - E(X_i))}{[\sum_{i=1}^{p} V(X_i)]^{\frac{1}{2}+\delta}} = 0) = 1.$$

Thus,

$$P(\lim_{p\to\infty}\sum_{i=1}^{p} X_i \le C) = 1.$$

It also follows that this conclusion remains valid for any $C > 0$. □

Lemma 2. *For random variables $U_{jp} > 0, V_{jp} > 0$, if $P(\lim_{p\to+\infty} \frac{V_{jp}}{U_{jp}} \le \frac{1}{2}) = 1, j = 0, 1, \ldots, K$, then*

$$P(\lim_{p\to+\infty} \frac{\sum_{j=0}^{K} V_{jp}}{\sum_{j=0}^{K} U_{jp}} < 1) = 1.$$

Proof. For each $j = 0, 1, \ldots, K$, if $\lim_{p\to+\infty} \frac{V_{jp}}{U_{jp}} \le \frac{1}{2}$, then $\exists M_j$, such that $\forall\ p > M_j$, $\frac{V_{jp}}{U_{jp}} < 1$. If $p > Max\{M_0, M_1, \ldots, M_K\}$, then $\sum_{j=1}^{K} V_{jp} < \sum_{j=1}^{K} U_{jp}$. We note that $\sum_{j=1}^{K} U_{jp} > 0$. So,

$$P(\lim_{p\to+\infty} \frac{\sum_{j=0}^{K} V_{jp}}{\sum_{j=0}^{K} U_{jp}} < 1) \quad \ge \quad P(\cap_{j=0}^{K}\{\lim_{p\to+\infty} \frac{V_{jp}}{U_{jp}} \le \frac{1}{2}\}) = 1.$$

□

Lemma 3. *For any $b \in \Omega^b$, $\theta \ne \theta_0$, suppose there exist a neighborhood of θ, $O(\theta, b)$, which does not contain θ_0, such that:*

$$P_{y|b;\theta_0}(\lim_{p\to\infty} \frac{\sup_{\theta' \in O(\theta,b)} M_p(y;\theta')}{M_p(y;\theta_0)} < 1) = 1.$$

Then $\hat{\theta}_p \to \theta_0$ a.e.

Proof. For any $\epsilon > 0$, $\Theta_1 = \{\theta \in \Theta : |\theta - \theta_0| \ge \epsilon\}$ is a compact set. So there exist a finite number of points $\theta^{(1)}, \ldots, \theta^{(h)}$ in Θ_1, such that $\cup_{i=1}^{h} O(\theta^{(i)}, b) \supset \Theta_1$. Furthermore,

$$P_{y|b;\theta_0}(\lim_{p\to\infty} \frac{\sup_{\theta' \in O(\theta^{(i)},b)} M_p(y;\theta')}{M_p(y;\theta_0)} < 1) = 1,$$

and

$$P_{y|b;\theta_0}(\lim_{p\to\infty} \frac{\sup_{|\theta-\theta_0|\ge\epsilon} M_p(y;\theta)}{M_p(y;\theta_0)} < 1) \quad \ge \quad P_{y|b;\theta_0}(\cap_{i=1}^{h} \lim_{p\to\infty} \frac{\sup_{\theta' \in O(\theta^{(i)},b)} M_p(y;\theta')}{M_p(y;\theta_0)} < 1) = 1.$$

Since it is true for any $b \in \Omega^b$, we obtain,

$$P_{y;\theta_0}(\lim_{p\to\infty} \frac{\sup_{|\theta-\theta_0|\geq\epsilon} M_p(y;\theta)}{M_p(y;\theta_0)} < 1) = E_b P_{y|b;\theta_0}(\lim_{p\to\infty} \frac{\sup_{|\theta-\theta_0|\geq\epsilon} M_p(y;\theta)}{M_p(y;\theta_0)} < 1) = 1.$$

Suppose $\hat{\theta_p} \to \theta_0$ a.e. is not true. Then there exist $\delta > 0$, such that $P_{y;\theta_0}(\limsup_{p\to\infty} |\hat{\theta_p} - \theta_0| \geq \delta) > 0$, which implies that,

$$P_{y;\theta_0}(\limsup_{p\to\infty} \frac{\sup_{|\theta-\theta_0|\geq\delta} M_p(y;\theta)}{M_p(y;\theta_0)} \geq 1) > 0.$$

This contradicts (13), which completes the proof. □

Proof of Theorem 1. By the Lemma 3, it is sufficient to show that for any $\theta \neq \theta_0$, for any b_j, there exist a neighborhood of θ, $O(\theta)$, such that,

$$P_{y|b_j;\theta_0}(\lim_{p\to\infty} \frac{\sup_{\theta'\in O(\theta)} M_p(y;\theta')}{M_p(y;\theta_0)} < 1) = 1.$$

Without loss of generality, we establish this statement for $b_j = b_0$. Since b has discrete distribution with finite support, $M_p(y;\theta) = (2\pi)^{-p/2} \sum_{j=0}^K w_j \exp\{-\frac{1}{2}\sum_{i=1}^p (y_i - f_i(\theta, b_j))^2\}$, observe that,

$$\begin{aligned} M_p(y;\theta_0) &= (2\pi)^{-p/2} \sum_{j=0}^K w_j \exp\{-\frac{1}{2}\sum_{i=1}^p (y_i - f_i(\theta_0, b_j))^2\} \\ &\geq w_0 (2\pi)^{-p/2} \exp\{-\frac{1}{2}\sum_{i=1}^p (y_i - f_i(\theta_0, b_0))^2\} \\ &= (2\pi)^{-p/2} \sum_{j=0}^K w_0 w_j \exp\{-\frac{1}{2}\sum_{i=1}^p (y_i - f_i(\theta_0, b_0))^2\}. \end{aligned}$$

We note that,

$$\begin{aligned} & P_{y|b_0;\theta_0}(\lim_{p\to\infty} \frac{\sup_{\theta'\in O(\theta)} M_p(y;\theta')}{M_p(y;\theta_0)} < 1) \\ \geq\ & P_{y|b_0;\theta_0}(\lim_{p\to\infty} \frac{\sup_{\theta'\in O(\theta)} \sum_{j=0}^K w_j \exp\{-\frac{1}{2}\sum_{i=1}^p (y_i - f_i(\theta', b_j))^2\}}{\sum_{j=1}^K w_0 w_j \exp\{-\frac{1}{2}\sum_{i=1}^p (y_i - f_i(\theta_0, b_0))^2\}} < 1) \\ \geq\ & P_{y|b_0;\theta_0}(\lim_{p\to\infty} \frac{\sum_{j=0}^K w_j \sup_{\theta'\in O(\theta)} \exp\{-\frac{1}{2}\sum_{i=1}^p (y_i - f_i(\theta', b_j))^2\}}{\sum_{j=1}^K w_0 w_j \exp\{-\frac{1}{2}\sum_{i=1}^p (y_i - f_i(\theta_0, b_0))^2\}} < 1). \end{aligned}$$

Let,

$$A_j = \{y : \lim_{p\to\infty} \frac{\sup_{\theta'\in O(\theta)} w_j \exp\{-\frac{1}{2}\sum_{i=1}^p (y_i - f_i(\theta', b_j))^2\}}{w_0 w_j \exp\{-\frac{1}{2}\sum_{i=1}^p (y_i - f_i(\theta_0, b_0))^2\}} \leq \frac{1}{2}\}.$$

By Lemma 2, it is sufficient to show that $P(A_j) = 1$, for $j = 0, 1, \ldots, K$. Observe that,

$$
\begin{aligned}
& P_{y|b_0,\theta_0}(A_j) \\
= \quad & P_{y|\theta_0,b_0}(\lim_{p\to\infty} \sup_{\theta' \in O(\theta)} \exp\{-\frac{1}{2}\sum_{i=1}^{p}(y_i - f_i(\theta', b_j))^2 + \frac{1}{2}\sum_{i=1}^{p}(y_i - f_i(\theta_0, b_0))^2\} \leq \frac{w_0}{2}) \\
= \quad & P_{y|b_0,\theta_0}(\lim_{p\to\infty} \sup_{\theta' \in O(\theta)} -\frac{1}{2}\sum_{i=1}^{p}[(y_i - f_i(\theta', b_j))^2 - (y_i - f_i(\theta_0, b_0))^2] \leq \ln\frac{w_0}{2}) \\
= \quad & P_{y|b_0,\theta_0}(\lim_{p\to\infty} \inf_{\theta' \in O(\theta)} \sum_{i=1}^{p}[(f_i(\theta_0, b_0) - f_i(\theta', b_j))^2 + 2(y_i - f_i(\theta_0, b_0))(f_i(\theta_0, b_0) - f_i(\theta', b_j))] \geq -2\ln\frac{w_0}{2}).
\end{aligned}
$$

We note that,

$$
\begin{aligned}
& \inf_{\theta' \in O(\theta)} \sum_{i=1}^{p}[(f_i(\theta_0, b_0) - f_i(\theta', b_j))^2 + 2(y_i - f_i(\theta_0, b_0))(f_i(\theta_0, b_0) - f_i(\theta', b_j))] \\
\geq \quad & \inf_{\theta' \in O(\theta)} \sum_{i=1}^{p}(f_i(\theta_0, b_0) - f_i(\theta', b_j))^2 - \sup_{\theta' \in O(\theta)} |\sum_{i=1}^{p} 2(y_i - f_i(\theta_0, b_0))(f_i(\theta_0, b_0) - f_i(\theta', b_j))| \\
\geq \quad & \inf_{\theta' \in O(\theta)} \{\sum_{i=1}^{p}(f_i(\theta_0, b_0) - f_i(\theta', b_j))^2\}(1 - \frac{A(\theta, y)}{B(\theta, y)}),
\end{aligned}
$$

where

$$
\begin{aligned}
A(\theta, y) &= \sup_{\theta' \in O(\theta)} |\sum_{i=1}^{p} 2(y_i - f_i(\theta_0, b_0))(f_i(\theta_0, b_0) - f_i(\theta', b_j))|, \\
B(\theta, y) &= \inf_{\theta' \in O(\theta)} \sum_{i=1}^{p}(f_i(\theta_0, b_0) - f_i(\theta', b_j))^2.
\end{aligned}
$$

By (6), $\sum_{i=1}^{p} \sup_{\theta' \in O(\theta)}(f_i(\theta', b_j) - f_i(\theta_0, b_0))^2 \to \infty$ as $p \to \infty$, and by (4), $B(\theta, y) \to \infty$ as $p \to \infty$. Observe that,

$$
\frac{A(\theta, y)}{B(\theta, y)} = \frac{A(\theta, y)}{C(\theta, y)}\frac{C(\theta, y)}{B(\theta, y)},
$$

where

$$
C(\theta, y) = \{\sum_{i=1}^{p} \sup_{\theta' \in O(\theta)} (f_i(\theta_0, b_0) - f_i(\theta', b_j))^2\}^{\frac{1+c}{2}}
$$

It follows from (4) that, for the strong consistency of $\hat{\theta}_p$, it is sufficient to establish:

$$
\frac{A(\theta, y}{2C(\theta, y)} \to 0 \quad a.s.
$$

for some $c > 0$. Using condition (5), this follows from Corollary A from the Appendix of Wu (1981). □

Proof of Theorem 2. By Lemma 3, it is sufficient to show that for any $\theta \neq \theta_0$, $b_0 \in \Omega^b$, there exists a neighborhood of θ, $O(\theta)$, such that,

$$P_{y|b_0;\theta_0}(\lim_{p\to\infty} \frac{\sup_{\theta'\in O(\theta)} M_p(y;\theta')}{M_p(y;\theta_0)} < 1) = 1.$$

We notice that when θ_0, θ and b_0 are given, we can always find $0 < \epsilon_1 < \frac{1}{2}$ and $0 < \epsilon_2 < \frac{1}{2}$ satisfying the following inequality:

$$\epsilon_1^2 + 2\epsilon_1\sqrt{\frac{2}{\pi}} + \epsilon_2(2\sqrt{\frac{2}{\pi}} + 1) < \frac{1}{4}D(\theta_0, \theta, b_0). \tag{13}$$

Since ϵ_1 and ϵ_2 are determined by θ_0, θ and b_0 only, they can be treated as constants.

In condition (i), since K is a finite number, we can assume, without loss of generality, that the sets $\{B_l\}$ are disjoint, i.e., $\exists\ \{B'_l, l = 1, \ldots, K' < \infty\}$, such that $B'_l \cap B'_{l'} = \emptyset$, $\cup_{l=1}^{K'} B'_l = \Omega^b$, and $\{B'_l\}$ satisfies condition (i).

We note that:

$$\begin{aligned}
M_p(y;\theta) &= (2\pi)^{-n/2}\int_{\Omega^b} \exp(-\frac{1}{2}\sum_{i=1}^{p}[y_i - f_i(\theta, b)]^2)dF(b) \\
&= \sum_{l=1}^{K'}(2\pi)^{-n/2}\int_{B'_l} \exp(-\frac{1}{2}\sum_{i=1}^{p}[y_i - f_i(\theta, b)]^2)dF(b). \\
M_p(y;\theta_0) &= (2\pi)^{-n/2}\int_{\Omega^b} \exp(-\frac{1}{2}\sum_{i=1}^{p}[y_i - f_i(\theta_0, b)]^2)dF(b) \\
&\geq (2\pi)^{-n/2}\int_{O(b_0)} \exp(-\frac{1}{2}\sum_{i=1}^{p}[y_i - f_i(\theta_0, b)]^2)dF(b),
\end{aligned}$$

where $O(b_0)$ is defined in condition (ii), $O(b_0) \subset \Omega^b$.

Note that, as in the proof of Theorem 1,

$$\begin{aligned}
& P_{y|b_0;\theta_0}(\lim_{p\to\infty} \frac{\sup_{\theta'\in O(\theta)} M_p(y;\theta')}{M_p(y;\theta_0)} < 1) \\
\geq\ & P_{y|b_0;\theta_0}(\lim_{p\to\infty} \frac{\sum_{j=1}^{K'} \sup_{\theta'\in O(\theta)} \int_{B'_l} \exp(-\frac{1}{2}\sum_{i=1}^{p}[y_i - f_i(\theta', b)]^2)dF(b)}{\sum_{j=1}^{K'} F(B'_l)\int_{O(b_0)} \exp(-\frac{1}{2}\sum_{i=1}^{p}[y_i - f_i(\theta_0, b)]^2)dF(b)} < 1).
\end{aligned}$$

Define sets $A_1, A_2, \ldots, A_{K'}$ as:

$$A_l = \{\lim_{p\to+\infty} \frac{\sup_{\theta'\in O(\theta)} \int_{B'_l} exp(-\frac{1}{2}\sum_{i=1}^{p}[y_i - f_i(\theta', b)]^2)dF(b)}{F(B'_l)\int_{O(b_0)} exp(-\frac{1}{2}\sum_{i=1}^{p}[y_i - f_i(\theta_0, b)]^2)dF(b)} \leq \frac{1}{2}\}.$$

By Lemma 2, it suffices to show that,

$$P_{y|b_0;\theta_0}(A_l) = 1,\ l = 1, \ldots, K'.$$

We consider separately the numerator and denominator of the expression in the definition of A_l. First consider the denominator. By condition (10) with $\epsilon = \epsilon_1$, we obtain $\forall b \in O(b_0)$,

$$\begin{aligned}\sum_{i=1}^{p}[y_i - f_i(\theta_0, b)]^2 &= \sum_{i=1}^{p}[y_i - f_i(\theta_0, b_0) + f_i(\theta_0, b_0) - f_i(\theta_0, b)]^2 \\ &\leq \sum_{i=1}^{p}\{[y_i - f_i(\theta_0, b_0)]^2 + \epsilon_1^2 + 2\epsilon_1|y_i - f_i(\theta_0, b_0)|\}.\end{aligned}$$

So we have the following inequality:

$$\begin{aligned}&\int_{O(b_0)} \exp(-\frac{1}{2}\sum_{i=1}^{p}[y_i - f_i(\theta_0, b)]^2)dF(b) \\ \geq \quad & F(O(b_0))\exp\left(-\frac{1}{2}\sum_{i=1}^{p}\{[y_i - f_i(\theta_0, b_0)]^2 + \epsilon_1^2 + 2\epsilon_1|y_i - f_i(\theta_0, b_0)|\}\right). \qquad (14)\end{aligned}$$

Now, consider the numerator of the expression in the definition of A_l. Since B_l' satisfies condition (i), it satisfies either (a) or (b). We examine these cases separately.

Case 1. B_l' satisfies (a).

For $b \in B_l'$ and $\theta' \in O(\theta)$, we obtain from condition (7) with $\epsilon = \epsilon_2$,

$$\sum_{i=1}^{p}[y_i - f_i(\theta', b)]^2 > \sum_{i=1}^{p}\{[y_i - f_i(\theta, b_l)]^2 - 2\epsilon_2|y_i - f_i(\theta, b_l)|\}.$$

Then,

$$\begin{aligned}&\sup_{\theta' \in O(\theta)} \int_{B_l'} \exp(-\frac{1}{2}\sum_{i=1}^{p}[y_i - f_i(\theta', b)]^2)dF(b) \\ \leq \quad & F(B_l')\exp\left(-\frac{1}{2}\sum_{i=1}^{p}\{[y_i - f_i(\theta, b_l)]^2 - 2\epsilon_2|y_i - f_i(\theta, b_l)|\}\right). \qquad (15)\end{aligned}$$

By the definition of A_l, and the inequalities (14) and (15), we have,

$$\begin{aligned}P_{y|b_0;\theta_0}(A_l) \geq\; & P_{y|b_0;\theta_0}(\lim_{p\to+\infty}\sum_{i=1}^{p}\{[y_i - f_i(\theta_0, b_0)]^2 + \epsilon_1^2 + 2\epsilon_1|y_i - f_i(\theta_0, b_0)|\} - 2\ln(\frac{F(O(b_0))}{2}) \\ & -\sum_{i=1}^{p}\{[y_i - f_i(\theta, b_l)]^2 - 2\epsilon_2|y_i - f_i(\theta, b_l)|\} < 0) \\ =\; & P_{y|b_0;\theta_0}(\lim_{p\to+\infty}\sum_{i=1}^{p}\{[y_i - f_i(\theta_0, b_0)]^2 - [y_i - f_i(\theta, b_l)]^2 + 2\epsilon_2|y_i - f_i(\theta, b_l)| \\ & +2\epsilon_1|y_i - f_i(\theta_0, b_0)| + \epsilon_1^2\} < 2\ln(\frac{F(O(b_0))}{2}) \\ =\; & P_{y|b_0;\theta_0}\left(\lim_{p\to+\infty}\sum_{i=1}^{p}X_i < 2\ln(\frac{F(O(b_0))}{2})\right), \qquad (16)\end{aligned}$$

where,

$$
\begin{aligned}
X_i &= f_i^2(\theta_0, b_0) - f_i^2(\theta, b_l) - 2y_i[f_i(\theta_0, b_0) - f_i(\theta, b_l)] \\
&\quad +2\epsilon_2|y_i - f_i(\theta, b_l)| + 2\epsilon_1|y_i - f_i(\theta_0, b_0)| + \epsilon_1^2.
\end{aligned}
$$

We note that when the parameter is θ_0, given $b = b_0$, $y_i \sim N(f_i(\theta_0, b_0), 1)$. Hence, it can be verified that,

$$
E_{y|b_0;\theta_0}(\sum_{i=1}^{p} X_i) < -(1-\epsilon_2)\sum_{i=1}^{p}[f_i(\theta_0, b_0) - f_i(\theta, b_l)]^2 + \epsilon_2(2\sqrt{\frac{2}{\pi}}+1)p + p\epsilon_1^2 + 2\epsilon_1 p\sqrt{\frac{2}{\pi}}.
$$

Hence by (13),

$$
E_{y|b_0;\theta_0}(\sum_{i=1}^{p} X_i) < -\frac{1}{4}\sum_{i=1}^{p}[f_i(\theta_0, b_0) - f_i(\theta, b_l)]^2.
$$

Also,

$$
V_{y|b_0;\theta_0}(\sum_{i=1}^{p} X_i) \quad < \quad (8+16\epsilon_2^2)\sum_{i=1}^{p}[f_i(\theta_0, b_0) - f_i(\theta, b_l)]^2 + 16\epsilon_2^2 p + 16\epsilon_1^2 p
$$

By (11), $\lim_{p\to+\infty} E_{y|b_0;\theta_0}(\sum_{i=1}^{p} X_i) = -\infty$ and $\lim_{p\to+\infty} V_{y|b_0;\theta_0}(\sum_{i=1}^{p} X_i) = \infty$. Furthermore,

$$
\lim_{p\to+\infty} \frac{[V_{y|b_0;\theta_0}(\sum_{i=1}^{p} X_i)]^{\frac{3}{4}}}{|E_{y|b_0;\theta_0}(\sum_{i=1}^{p} X_i)|} = 0.
$$

Hence by Lemma 1, we have $P_{y|b_0;\theta_0}(A_l) = 1$ for all sets B_l' that satisfy (a).

Case 2. B_l' satisfies (b).

It follows from condition (b) that we can partition $\{1, \ldots, p\}$ into two sets G_1 and G_2, such that $b \in B_l'$ satisfies (8) if $i \in G_1$, and satisfies (9) if $i \in G_2$. So for $b \in B_l'$ and $\theta' \in O(\theta)$, we have the following inequalities:

$$
\begin{aligned}
\sum_{i\in G_1}[y_i - f_i(\theta', b)]^2 &> \sum_{i\in G_1}\{[y_i - f_i(\theta', b)]^2 I\{y_i \le f_i(\theta_0, b_0)+2\} + [y_i - f_i(\theta', b)]^2 I\{y_i > f_i(\theta_0, b_0)+2\}\} \\
&> \sum_{i\in G_1}[y_i - f_i(\theta_0, b_0) - 2]^2 I\{y_i \le f_i(\theta_0, b_0)+2\}. \\
\sum_{i\in G_2}[y_i - f_i(\theta', b)]^2 &> \sum_{i\in G_2}\{[y_i - f_i(\theta', b)]^2 I\{y_i < f_i(\theta_0, b_0)-2\} + [y_i - f_i(\theta', b)]^2 I\{y_i \ge f_i(\theta_0, b_0)-2\} \\
&> \sum_{i\in G_2}[y_i - f_i(\theta_0, b_0) + 2]^2 I\{y_i \ge f_i(\theta_0, b_0)-2\}.
\end{aligned}
$$

So we have

$$
\begin{aligned}
\sum_{i=1}^{p}[y_i - f_i(\theta', b)]^2 &= \sum_{i\in G_1}[y_i - f_i(\theta', b)]^2 + \sum_{i\in G_2}[y_i - f_i(\theta', b)]^2 \\
&> \sum_{i\in G_1}\big([y_i - f_i(\theta_0, b_0) - 2]^2 - [y_i - f_i(\theta_0, b_0) - 2]^2 I\{y_i > f_i(\theta_0, b_0) + 2\}\big) \\
&\quad + \sum_{i\in G_2}\big([y_i - f_i(\theta_0, b_0) + 2]^2 - [y_i - f_i(\theta_0, b_0) - 2]^2 I\{y_i < f_i(\theta_0, b_0) - 2\}\big) \\
&> \sum_{i=1}^{p}[y_i - f_i(\theta_0, b_0)]^2 - 4\sum_{i\in G_1}[y_i - f_i(\theta_0, b_0)] + 4\sum_{i\in G_2}[y_i - f_i(\theta_0, b_0)] + 4p \\
&\quad - \sum_{i\in G_1} X_{1i} - \sum_{i\in G_2} X_{2i}, \qquad (17)
\end{aligned}
$$

where

$$
\begin{aligned}
X_{1i} &= [y_i - f_i(\theta_0, b_0) - 2]^2 I\{y_i > f_i(\theta_0, b_0) + 2\}, \\
X_{2i} &= [y_i - f_i(\theta_0, b_0) + 2]^2 I\{y_i < f_i(\theta_0, b_0) - 2\}.
\end{aligned}
$$

Then

$$
\begin{aligned}
&\sup_{\theta'\in O(\theta)} \int_{B_l'} \exp(-\frac{1}{2}\sum_{i=1}^{p}[y_i - f_i(\theta', b)]^2)dF(b) \\
\le\ & F(B_l')\exp\{-\frac{1}{2}\sum_{i=1}^{p}[y_i - f_i(\theta_0, b_0)]^2 + 2\sum_{i\in G_1}[y_i - f_i(\theta_0, b_0)] - 2\sum_{i\in G_2}[y_i - f_i(\theta_0, b_0)] - 2p \\
& + \frac{1}{2}\sum_{i\in G_1} X_{1i} + \frac{1}{2}\sum_{i\in G_2} X_{2i}\}. \qquad (18)
\end{aligned}
$$

By the definition of A_l, and the inequalities (14) and (18), we have,

$$
\begin{aligned}
P_{y|b_0;\theta_0}(A_j) &\ge P_{y|\theta_0,b_0}(\lim_{p\to+\infty}\sum_{i=1}^{p}\{[y_i - f_i(\theta_0, b_0)]^2 + \epsilon_1^2 + 2\epsilon_1|y_i - f_i(\theta_0, b_0)|\} - 2\ln(\frac{F(O(b_0))}{2}) - \\
&\quad \sum_{i=1}^{p}[y_i - f_i(\theta_0, b_0)]^2 + 4\sum_{i\in G_1}[y_i - f_i(\theta_0, b_0)] - 4\sum_{i\in G_2}[y_i - f_i(\theta_0, b_0)] - 4p \\
&\quad + \sum_{i\in G_1} X_{1i} + \sum_{i\in G_2} X_{2i} < 0) \\
&= P_{y|b_0;\theta_0}(\lim_{p\to+\infty} X_{3i} < 2\ln(\frac{F(O(b_0))}{2})).
\end{aligned}
$$

Here

$$X_{3i} = 4\sum_{i\in G_1}[y_i - f_i(\theta_0, b_0)] - 4\sum_{i\in G_2}[y_i - f_i(\theta_0, b_0)] + 2\epsilon_1\sum_{i=1}^{p}|y_i - f_i(\theta_0, b_0)|$$
$$+\sum_{i\in G_1} X_{1i} + \sum_{i\in G_2} X_{2i} - 4p + p\epsilon_1^2.$$

We note that when the parameter is θ_0, given $b = b_0$, $y_i \sim N(f_i(\theta_0, b_0), 1)$. Hence, it can be verified that,

$$E_{y|b_0;\theta_0}X_{1i} < 1,\ E_{y|b_0;\theta_0}X_{2i} < 1,\ V_{y|b_0;\theta_0}X_{1i} < 2,\ V_{y|b_0;\theta_0}X_{2i} < 2.$$

So

$$E_{y|b_0;\theta_0}(\sum_{i=1}^{p} X_{3i}) < p\epsilon_1^2 + 2\epsilon_1\sqrt{\frac{2}{\pi}} - 3p,$$

and,

$$V_{y|b_0;\theta_0}(\sum_{i=1}^{p} X_{3i}) < 44p + 16\epsilon_1^2.$$

Since $\epsilon_1 < \frac{1}{2}$, $\lim_{p\to+\infty} E_{y|\theta_0,b_0}\sum_{i=1}^{p} X_{3i} = -\infty$ and $\lim_{p\to+\infty} V_{y|b_0;\theta_0}(\sum_{i=1}^{p} X_{3i}) = +\infty$. Furthermore,

$$\lim_{p\to+\infty}\frac{[V_{y|b_0;\theta_0}(\sum_{i=1}^{p} X_{3i})]^{\frac{3}{4}}}{|E_{y|b_0;\theta_0}(\sum_{i=1}^{p} X_{3i})|} = 0.$$

Hence by Lemma 1, we have $P_{y|b_0;\theta_0}(A_l) = 1$ for all sets B_l' that satisfy (b). The theorem now follows from Lemma 3. □

Proof of Corollary 2. Since $f_i(\theta, b)$ is continuous in b and θ uniformly in i, condition (10) in Theorem 2 is satisfied. Furthermore for any $\epsilon > 0$, for any compact set B_l, for each $b \in B_l$, there exist $O(b)$, a neighborhood of b, and $O(\theta_b)$, a neighborhood of θ, such that,

$$\sup_i \sup_{\theta'\in O(\theta_b)} \sup_{b'\in O(b)} |f_i(\theta', b') - f_i(\theta, b)| < \epsilon.$$

Since B_l is compact, we can find $O(b_{l_1}), \ldots, O(b_{l_{n_l}})$, such that $\cup_{j=1}^{n_l} O(b_{l_j}) \supset B_l$. So there exists sets $\{B_h'\}$ such that $B_h' = O(B_{l_j})$ for some l_j, and $\cup_{l=1}^{M} B_l' \supset \Omega^b$, for a finite M. It follows that for each B_l', there exist a neighborhood of θ, $O(\theta_l)$, such that B_l' satisfies one of (7), (8) and (9). Let $O(\theta) = \cap_{l=1}^{M} O(\theta_l)$, then each B_l' satisfies condition (a) or condition (b) in Theorem 2. Corollary 2 now follows from Theorem 2. □

Acknowledgement: The authors sincerely thank Professor D. Majumdar, who leads us to this interesting area. His consistent suggestions and comments greatly improved the quality of the paper. We also thank the editor and a referee for their very constructive comments and suggestions, which substantially improved the article.

References

Bose, A. and Sengupta, D. (2003). Strong consistency of minimum contrast estimators with applications. *Sankhya 65*, 440–463.

Bradley, R. A. and Gart, J. J. (1962). The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika 49*(1 and 2), 205–214.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association 88*(421), 9–25.

Chung, K. L. (1974). *A course in probability theory* (Second ed.). Academic Press, New York-London.

Cramer, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

Davidian, M. and Giltinan, D. M. (1993). Some general estimation methods for nonlinear mixed-effects models. *Journal of Biopharmaceutical Statistics 3*(1), 23–55.

Davidian, M. and Giltinan, D. M. (1995a). *Nonlinear Models for Repeated Measurements Data*. London: Chapman and Hall.

Davidian, M. and Giltinan, D. M. (1995b). *Nonlinear Models for Repeated Measurements Data*. London: Chapman and Hall.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate statistical modelling based on generalized linear models*. New York: Springer-Verlag.

Hoadley, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *The Annal of Mathematical Statistics 42*(6), 1977–1991.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist. 40*, 633–643.

Jiang, J. (1997). Wald consistency and the method of sieves in REML estimation. *Ann. Statist. 25*(4), 1781–1803.

Le Cam, L. (1979). A reduction theorem for certain sequential experiments. II. *Ann. Statist. 7*(4), 847–859.

Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics 46*, 673–687.

Mentré, F. and Gomeli, R. (1995). A two-step iterative algorithm for estimation in nonlinear mixed-effect models with an evaluation in population pharmacokinetics. *Journal of Biopharmaceutical Statistics 5*(2), 141–158.

Nie, L. (2002). *Laplace approximation in nonlinear mixed-effects models*. Ph. D. thesis, University of Illinois at Chicago.

Nie, L. (2004). Strong consistency of the maximum likelihood estimator in generalized linear and nonlinear mixed-effects models. *Metrika (accepted)*.

Nielsen, O. K., Ritz, C., and Streibig, J. C. (2004). Nonlinear mixed-model regression to analyze herbicide doseresponse relationships. *Weed Technology*.

Pinheiro, J. C. and Bates, D. M. (1995). Approximation to the log-likelihood function in the nonlinear mixed-effects models. *Journal of Computational and Graphical Statistics 4*, 12–35.

Roe, D. J. (1997). Comparison of population pharmacokienetic modeling methods using simulated data: Results from the population modeling workgroup. *Statistics in Medicine 16*, 1241–1262.

Vonesh, E. F. (1996). A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika 83*(2), 447–452.

Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker, Inc.

Vonesh, E. F., Wang, H., Nie, L., and Majumdar, D. (2002). Conditional second-order generalized estimating equations for generalized linear and nonlinear mixed-effects models. *Journal of the American Statistical Association 96*, 282–291.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *The Annal of Mathematical Statistics 20*, 595–601.

Wolfinger, R. D. and Lin, X. (1997). Two Taylor-series approximation methods for nonlinear mixed models. *Computatinal Statistics & Data Analysis 25*, 465–490.

Wu, C. F. (1981). Asymptotic theorem of nonlinear least squares estimation. *The Annal of Statistics 9*(5), 501–613.

Yamada, T., Goto, M., Punj, V., Zaborina, O., Chen, M. L., Kimbara, K., Majumdar, D., Cunningham, E., Gupta, T. K. D., and Chakrabarty, A. M. (2002). Bacterial redox protein azurin, tumor suppressor protein p53, and regression of cancer. *Proceedings of the National Academy of Sciences*, 14098–14103.