Working Paper 98-73
Statistics and Econometrics Series 32
September 1998

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax 34 - 91- 624.9849

ASYMPTOTIC PROPERTIES FOR A SIMULATED PSEUDO MAXIMUM LIKELIHOOD ESTIMATOR

Olivier Nuñez*

Abstract

We propose an estimator for parameters of nonlinear mixed effects model, obtained by maximization of a simulated pseudo likelihood. This simulated criterion is constructed from the likelihood of a Gaussian model whose means and variances are given by Monte Carlo approximations of means and variances of the true model. If the number of experimental units and the sample size of Monte Carlo simulations are respectively denoted by N and K, we obtained the strong consistency and asymptotic normality of the estimator when the ratio $N^{1/2}/K$ tends to zero.

Key Words

Nonlinear mixed-effects models; Simulation estimators; Asymptotic normality; Consistency.

*Department of Statistics y Econometrics, Universidad Carlos III de Madrid, e-mail: nunez@est-econ.uc3m.es.

Asymptotic Properties for a Simulated Pseudo Maximum Likelihood Estimator

Olivier Nuñez

SUMMARY

We propose an estimator for parameters of nonlinear mixed effects model, obtained by maximization of a simulated pseudo likelihood. This simulated criterion is constructed from the likelihood of a Gaussian model whose means and variances are given by Monte Carlo approximations of means and variances of the true model. If the number of experimental units and the sample size of Monte Carlo simulations are respectively denoted by N and K, we obtained the strong consistency and asymptotic normality of the estimator when the ratio $N^{\frac{1}{2}}/K$ tends to zero.

Some key words: Nonlinear mixed-effects models; Simulation estimators; Asymptotic normality; Consistency.

1. INTRODUCTION

We consider the following mixed effects model

$$Y_i = \eta_i(\Phi_i) + \Lambda_i(\Phi_i)\varepsilon_i, \ \ \varepsilon_i \sim \mathcal{N}(0, \sigma^2 I_{n_i}), \tag{1}$$

and

$$\Phi_i = A_i\alpha + B_i D^{\frac{1}{2}} E_i, \ \ E_i \sim \mathcal{N}(0, I_q),$$

where $i = 1...N$, N is the sample size.

The vector Y_i is a vector of n_i observations made on the i^{th} experimental unit.

The random vectors $(\varepsilon_i)_{i=1..N}$ and $(E_i)_{i=1..N}$ are unobserved and are assumed mutually independent. The matrices A_i and B_i are full rank matrices of deterministic explanatory variables. The vectorial functions $\eta_i(.)$ and the matrices $\Lambda_i(.)$ depend non-linearly on Φ_i and are assumed to be sufficiently regular. The vector α is a parameter vector of size r and D is a $q \times q$ covariance matrix.

We are interested in the estimation of the parameter vector $\theta \equiv (\alpha, \text{vec}(D), \sigma^2)$, where $\text{vec}(D)$ denotes the set of parameters of the matrix D.

Several different methods have been proposed for this estimation in the homoscedastic framework (i.e.: Λ_i depends on i only through its dimension). The methods suggested by Sheiner-Beal (1992) or Linstrom-Bates (1990) are based on a Laplacian approximation of the marginal log-likelihood of observations. These methods involve a linearization of the conditional model (1) with respect to the vector E_i about 0 or about a posterior mode, before integration.

The main problem which occurs when using these methods is related to asymptotic properties of estimators (when $N \to \infty$ and $\sup_{i \geq 1} n_i < \infty$) which are either not known or poor (for the non consistency see Ramos and Pantula (1995)).
These properties depend on the control, with respect to θ, of the remainder of the marginal log-likelihood expansion. Some authors (Vonesh 1996) imply that choosing "a good order" of expansion may improve these properties. But it seems difficult to provide suitable assumptions about the model curvature to control the remaining terms. Nevertheless, when the number of observations per experimental unit tends uniformly to infinity ($\inf_{i \geq 1} n_i \to \infty$), Vonesh (1996) gives a consistency result of estimators. But in this case, the usual assumption of independence between observations made on the same experimental unit (the Λ_i's assumed diagonal) is no more tenable.
As a consequence, it is difficult to make an inference (which is often the primary aim of the experiment) due to the lack of at least the asymptotic distribution of the estimator.

In this paper, we propose a simulated method of pseudo maximum likelihood, for which we give asymptotic properties. In section 2, the Simulated Pseudo Maximum Likelihood (SPML) method is presented. The main results are given in section 3, in which the strong consistency and asymptotic normality of the SPML estimator are demonstrated.

2. Simulated Pseudo Maximum Likelihood (SPML) Approach

The estimation of the parameter vector $\theta \equiv (\alpha, \mathrm{vec}(D), \sigma^2)$ with the SPML method, consists of constructing an objective function derived from a family of probability densities parametrized by the first moments of the observations. To this end, let us define for all $i \geq 1$,

$$\begin{aligned} \mu_i(\theta) &= \mathbb{E}_\theta (Y_i) \\ &= \mathbb{E}_\theta [\eta_i(\Phi_i)], \end{aligned}$$

where $\mu_i(\theta)$ depends on θ only through $(\alpha, \mathrm{vec}(D))$, and

$$\begin{aligned} V_i(\theta) &= \mathrm{var}_\theta (Y_i) \\ &= \mathrm{var}_\theta [\eta_i(\Phi_i)] + \sigma^2 \mathbb{E}_\theta \left[\Lambda_i^2(\Phi_i)\right]. \end{aligned}$$

These two first moments have generally no explicit form, and are respectively approximated using the following Monte-Carlo sums :

$$\bar{\mu}_{i,K}(\theta) = \frac{1}{K}\sum_{k=1}^{K}\eta_i(\Phi_i^k),$$

$$V_{i,K}(\theta) = \frac{1}{K-1}\sum_{k=1}^{K}\left[\eta_i(\Phi_i^k) - \bar{\mu}_{i,K}(\theta)\right]\left[\eta_i(\Phi_i^k) - \bar{\mu}_{i,K}(\theta)\right]' + \frac{\sigma^2}{K}\sum_{k=1}^{K}\Lambda_i^2(\Phi_i^k),$$

where $\Phi_i^k = A_i\alpha + B_i D^{\frac{1}{2}}\, E_i^k$, and E_i^k are drawn independently from $\mathcal{N}(0, I_q)$.

The objective function is constructed by considering that the Y_i's are respectively drawn from a gaussian distribution $\mathcal{N}_{n_i}\left(\mu_i(\theta), V_i(\theta)\right)$. This approach leads to consideration of the following objective function:

$$\bar{C}_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} C_i(\theta),$$

where

$$C_i(\theta) = \|Y_i - \mu_i(\theta)\|^2_{V_i^{-1}(\theta)} + \ln|V_i(\theta)|.$$

The two first simulated moments $\bar{\mu}_{i,K}(\theta)$ and $V_{i,K}(\theta)$ are used to evaluate each $C_i(\theta)$. Thus, the proposed simulated objective function is :

$$\bar{C}_N^K(\theta) \equiv \frac{1}{N}\sum_{i=1}^{N} C_i^K(\theta),$$

where

$$C_i^K(\theta) \equiv \left\|Y_i - \bar{\mu}_{i,K}(\theta)\right\|^2_{V_{i,K}^{-1}(\theta)} + \ln\left|V_{i,K}(\theta)\right|,$$

A similar idea was used by Gourieroux and Monfort (1991). These authors gave consistency results in the independent and identically distributed framework. This result cannot be extended readily to the non identically distributed case.

The estimator of θ which minimizes the criterion $\bar{C}_N^K(\theta)$ will be denoted by $\hat{\theta}_N^K$.

In this paper, the word asymptotic refers to the sample size N tending to infinity, and the numbers of observations per experimental unit being uniformly bounded (i.e. $\sup_{i\geq 1} n_i < \infty$).

3. Asymptotic Properties

Let us consider the following assumptions in order to ensure the strong consistency of the estimator $\hat{\theta}_N^K$.

(A1)

The true value of the parameter vector denoted by $\theta^* = (\alpha^*, vec(D^*), \sigma^{2*})$ is an interior point of a compact metric space $\Theta \subset \mathbb{R}^p$ (where p is the dimension of θ^*, $p = r + q(q+1)/2 + 1$).

(A2)

The size of observed vectors Y_i $(i \geq 1)$ are uniformly bounded :

$$\sup_{i \geq 1} n_i \; < \; \infty.$$

(A3)

For all $i \geq 1$, and for all θ in Θ, the four first moments of $C_i^K(\theta)$ are finite uniformly in K:

$$\sup_{K \geq 1} \mathbb{E}_\theta \left(C_i^K(\theta)^4 \right) \; < \; \infty.$$

(A4)

For all $i \geq 1$, the functions $\eta_i(.)$ and $\Lambda_i(.)$ are almost surely twice continuously differentiable and square integrable.

(A5)

the vector θ^* is second order identifiable :

$$\forall i \geq 1 \qquad \left.\begin{array}{l} \mu_i(\theta^*) = \mu_i(\theta) \\ V_i(\theta^*) = V_i(\theta) \end{array}\right\} \Longleftrightarrow \theta \; = \; \theta^*.$$

Assumption (A1) is not restrictive because we can generally include the parameter vector in a compact set. This assumption is one of the main argument in our framework to show results about the uniform convergence in Θ of the simulated criterion, this kind of convergence being very useful for establishing the strong consistency of $\hat{\theta}_N^K$ (see Andrews (1989)).
On the other hand, assumption (A3) can be quite restrictive because it requires the existence of the eight first moments of observations.

We start by giving a technical lemma which ensures that when N and K are sufficiently large, the simulated criterion $\bar{C}_N^K(\theta)$ is close to the expectation of $\bar{C}_N(\theta)$.

LEMMA. *Under assumptions (A1-A4), we have almost surely*

$$\bar{C}_N^K(\theta) - E_{\theta^*}(\bar{C}_N(\theta)) \xrightarrow[\substack{N \to \infty \\ K \to \infty}]{} 0, \text{ uniformly on } \Theta$$

This lemma shows the ability of $\bar{C}_N^K(\theta)$ to locate (at least for large N and K) the parameter vector θ^*, as can be seen by the study of $\mathbb{E}_{\theta^*}(\bar{C}_N(\theta))$.

$$\mathbb{E}_{\theta^*}\left(\bar{C}_N(\theta)\right) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\theta^*}\left(\|Y_i - \mu_i(\theta)\|^2_{V_i^{-1}(\theta)} + \ln|V_i(\theta)|\right)$$

is equal (up to an additive constant) to

$$-\frac{1}{2N}\int \ln\phi_N(y;\theta)\, dP_{\theta^*}^N(y),$$

where $P_{\theta^*}^N$ is the probability measure of the vector $Y = (Y_1, Y_2, .., Y_N)$ and $\phi_N(y;\theta)$ is the probability density of gaussian distribution

$$\bigotimes_{i=1}^{N}\mathcal{N}_{n_i}\left(\mu_i(\theta), V_i(\theta)\right).$$

But, this last expression is also equal to

$$-\frac{1}{2N}\int \ln\phi_N(y;\theta)\;\phi_N(y;\theta^*)\; dy,$$

because $\ln\phi_N(y;\theta)$ is a quadratic function of the vector Y, and the two first moments of $P_{\theta^*}^N$ and $\phi_N(y;\theta^*)$ coincide. Therefore, we deduce from Jensen's inequality that

$$\mathbb{E}_{\theta^*}\left(\bar{C}_N(\theta)\right) \geq \mathbb{E}_{\theta^*}\left(\bar{C}_N(\theta^*)\right), \tag{2}$$

and in view of assumption (A5), the equality holds if and only if $\theta = \theta^*$.

We conclude this discussion with the following proposition related to the strong consistency of $\hat{\theta}_N^K$.

PROPOSITION 1. *Under assumptions (A1-A5), the sequence* $\left(\hat{\theta}_N^K\right)_{K,N}$ *converges almost surely, when N and K tend to infinity, to the parameter vector* θ^*.

Proof. Let us consider the quantity

$$Q_N(\theta,\theta^*) \equiv \mathbb{E}_{\theta^*}\left(\bar{C}_N(\theta) - \bar{C}_N(\theta^*)\right),$$

and B an open ball of Θ centered in θ^*. Then, from (2) there exist $\varepsilon > 0$ such that the function $\theta \longmapsto Q_N(\theta,\theta^*)$ is undervalued by ε on $\Theta \setminus B$.
The demonstration is based on the following inclusion

$$\left\{\hat{\theta}_N^K \notin B\right\} \subset \left\{\inf_{\Theta\setminus B}\bar{C}_N^K(\theta) \leq \bar{C}_N^K(\theta^*)\right\}. \tag{3}$$

On the other hand, in view of the uniform convergence showed in the lemma, we have almost surely

$$\left[\inf_{\Theta\backslash B} \bar{C}_N^K(\theta) - \bar{C}_N^K(\theta^*)\right] - \inf_{\Theta\backslash B} Q_N(\theta,\theta^*) \tag{4}$$

and

$$\inf_{\Theta\backslash B} Q_N(\theta,\theta^*) > \varepsilon.$$

Therefore, from (3) and (4), we have

$$P\left(\limsup_{\substack{N\to\infty\\K\to\infty}} \left\{\hat{\theta}_N^K \not\in B\right\}\right) \leq P\left(\limsup_{\substack{N\to\infty\\K\to\infty}} \left\{\inf_{\Theta\backslash B} \bar{C}_N^K(\theta) - \bar{C}_N^K(\theta^*) \leq 0\right\}\right) = 0.$$

□

We can now study the asymptotic normality of the estimator $\hat{\theta}_N^K$.
The convergence in distribution of this estimator requires the additional following assumption.

(A6)

Let us consider the vector $\nabla_\theta M_i^K(\theta^*) = \nabla_\theta C_i^K(\theta^*) - \mathbb{E}_{\theta^*}\left(\nabla_\theta C_i^K(\theta^*)\right)$. There exists $\delta > 0$ such that, for N sufficiently large

$$\sum_{i=1}^N \mathbb{E}_{\theta^*}\left(\nabla_\theta M_i^K(\theta^*)^{2+\delta}\right) = o\left(\left[\sum_{i=1}^N \mathbb{E}_{\theta^*}\left(\nabla_\theta M_i^K(\theta^*)^2\right)\right]^\delta\right),$$

uniformly in K.
The operator ∇_θ denotes the gradient with respect to θ.

This condition is the so called Ljapunov's condition, which is generally easier to check than Lindeberg's in order to establish a central limit theorem. It is only required in the heteroskedastic case. This condition guarantees that the variances of individual terms $\nabla_\theta C_i^K(\theta)$ are small as compared to their sum (see for details Feller (1971)).

PROPOSITION 2. *Let us denote* $I(\theta^*) = \lim_{N\to\infty} \mathbb{E}_{\theta^*}\left(H_\theta \bar{C}_N(\theta^*)\right)$, *where* $H_\theta \bar{C}_N(\theta^*)$ *is the hessian of the function* $\bar{C}_N(\theta^*)$, *with respect to* θ, *and* $\Gamma(\theta^*) = \lim_{\substack{N\to\infty\\K\to\infty}} \frac{1}{N}\sum_{i=1}^N \text{var}_{\theta^*}\left(\nabla_\theta C_i^K(\theta^*)\right)$.
If the assumptions (A1-A6) hold, then when N and K tend to infinity, with $N^{\frac{1}{2}}/K \to 0$, *we have*

$$N^{\frac{1}{2}}\left(\hat{\theta}_N^K - \theta^*\right) \sim \mathcal{N}_p\left(0, I^{-1}(\theta^*)\Gamma(\theta^*)I^{-1}(\theta^*)\right).$$

Remark. By construction the hessian matrix $\mathbb{E}_{\theta^*}\left(H_\theta \bar{C}_N(\theta^*)\right)$ is (up to a constant) the Fisher's information matrix of the gaussian pseudo-model

$$\bigotimes_{i=1}^{N} \mathcal{N}_{n_i}\left(\mu_i(\theta), V_i(\theta)\right).$$

After further calculus, we derive that the $(l,m)^{th}$ element of the $(p+1)\times(p+1)$ matrix $\mathbb{E}_{\theta^*}\left(H_\theta \bar{C}_N(\theta^*)\right)$ is

$$\begin{aligned} &\mathbb{E}_{\theta^*}\left(\frac{\partial^2}{\partial\theta_l\partial\theta_m}\bar{C}_N(\theta^*)\right) \\ = &\frac{1}{N}\sum_{i=1}^{N}\left[2\left\{\frac{\partial}{\partial\theta_l}\mu_i(\theta^*)\right\}' V_i^{-1}(\theta^*)\frac{\partial}{\partial\theta_m}\mu_i(\theta^*) + \mathrm{tr}\left(V_i^{-1}(\theta^*)\frac{\partial}{\partial\theta_l}V_i(\theta^*)V_i^{-1}(\theta^*)\frac{\partial}{\partial\theta_m}V_i(\theta^*)\right)\right]. \end{aligned}$$

Finally, the expression of the variances $\mathrm{var}_{\theta^*}\left(\nabla_\theta C_i(\theta^*)\right)$ $(i = 1..N)$, obtained after tedious calculus, is presented in the appendix.

Appendix

Proof of the Lemma

Notice that

$$\begin{aligned} & \sup_\theta \left|\bar{C}_N^K(\theta) - \mathbb{E}_{\theta^*}\left(\bar{C}_N(\theta)\right)\right| \\ \leq\ & \sup_\theta \left|\bar{C}_N^K(\theta) - \mathbb{E}_{\theta^*}\left(\bar{C}_N^K(\theta)\right)\right| + \sup_\theta \left|\mathbb{E}_{\theta^*}\left(\bar{C}_N^K(\theta) - \bar{C}_N(\theta)\right)\right| \end{aligned}$$

then, it remains to demonstrate that the two terms of the right hand side converge to zero when N and K tend to infinity.
Let us start by showing that

$$\sup_N \sup_{\theta\in\Theta} \mathbb{E}_{\theta^*} \left|\bar{C}_N^K(\theta) - \bar{C}_N(\theta)\right| \xrightarrow[K\to\infty]{} 0.$$

For convenience, we adopt the following notations, the components of $\bar{\mu}_{i,K}(\theta)$ and the elements of matrix $V_{i,K}(\theta)$ are stored in the vector $\bar{\xi}_{i,K}(\theta)$.
Then, one can write

$$\bar{\xi}_{i,K}(\theta) = \frac{1}{K}\sum_{k=1}^{K} \xi_i^k(\theta),$$

where the vectors $(\xi_i^k(\theta))_{k=1..K}$ are uncorrelated and identically distributed.
The expectation $\mathbb{E}\xi_i^k(\theta)$ and the variance $\operatorname{var}(\xi_i^k(\theta))$ are respectively denoted by $m_i(\theta)$ and $W_i(\theta)$.
Finally, let us introduce the function $f_i(\xi) = \|Y_i - \mu\|^2_{V^{-1}} + \ln|V|$, where ξ denotes the components of μ and V.
According to the previous notations, we have for a given $\tau > 0$,

$$\begin{aligned} & \mathbb{E}_{\theta^*}\left|C_i^K(\theta) - C_i(\theta)\right| = \mathbb{E}_{\theta^*}\left|f_i(\bar{\xi}_{i,K}(\theta)) - f_i(m_i(\theta))\right| \\ =\ & \mathbb{E}_{\theta^*}\left(\left|f_i(\bar{\xi}_{i,K}(\theta)) - f_i(m_i(\theta))\right| 1\!\!1_{\{\|\bar{\xi}_{i,K}(\theta)-m_i(\theta)\|\geq\tau\}}\right) + \mathbb{E}_{\theta^*}\left(\left|f_i(\bar{\xi}_{i,K}(\theta)) - f_i(m_i(\theta))\right| 1\!\!1_{\{\|\bar{\xi}_{i,K}(\theta)-m_i(\theta)\|<\tau\}}\right) \end{aligned}$$

By continuity of f_i, the second term of the last expression is arbitrary small for τ sufficiently small, uniformly on Θ.
Using the Cauchy-Schwarz's inequality, we see that the first term is less than

$$\mathbb{E}^{\frac{1}{2}}_{\theta^*}\left(f_i(\bar{\xi}_{i,K}(\theta)) - f_i(m_i(\theta))\right)^2 P^{\frac{1}{2}}_{\theta^*}\left(\|\bar{\xi}_{i,K}(\theta) - m_i(\theta)\| \geq \tau\right),$$

and by Chebishev's inequality, the previous expression is bounded above by

$$\begin{aligned} & \mathbb{E}^{\frac{1}{2}}_{\theta^*}\left(f_i(\bar{\xi}_{i,K}(\theta)) - f_i(m_i(\theta))\right)^2 \frac{1}{\tau}\mathbb{E}^{\frac{1}{2}}_{\theta^*}\left(\|\bar{\xi}_{i,K}(\theta) - m_i(\theta)\|^2\right) \\ =\ & \mathbb{E}^{\frac{1}{2}}_{\theta^*}\left(C_i^K(\theta) - C_i(\theta)\right)^2 \frac{1}{\sqrt{K}\tau}\operatorname{tr}^{\frac{1}{2}}\left(W_i(\theta)\right). \end{aligned}$$

Now, $\mathrm{tr}\,(W_i(\theta))$ is uniformly bounded on Θ and this quantity is obviously an increasing function of n_i, but in view of (A2) we obtain that

$$\sup_\theta \sup_i \mathrm{tr}\,(W_i(\theta)) < \infty.$$

Finally, uniformly in i the expectation

$$\mathbb{E}_{\theta^*}\left(C_i^K(\theta) - C_i(\theta)\right)^2,$$

is from the same arguments used for $\mathrm{tr}\,(W_i(\theta))$, uniformly bounded on Θ. In addition, this expectation is from (A3) uniformly bounded for $K \geq 1$.
Thus, we can conclude that

$$\sup_N \mathbb{E}_{\theta^*}\left|\bar{C}_N^K(\theta) - \bar{C}_N(\theta)\right| \leq \sup_i \mathbb{E}_{\theta^*}\left|C_i^K(\theta) - C_i(\theta)\right| \xrightarrow[K\to\infty]{} 0,$$

uniformly on Θ.
It remains to show that almost surely,

$$\sup_K \sup_{\theta\in\Theta}\left|\bar{C}_N^K(\theta) - \mathbb{E}_{\theta^*}\left(\bar{C}_N(\theta)\right)\right| \xrightarrow[N\to\infty]{} 0,$$

Let's denote, for all $\theta \in \Theta$,

$$\bar{M}_N^K(\theta) = \frac{1}{N}\sum_{i=1}^N M_i^K(\theta),$$

with

$$M_i^K(\theta) = C_i^K(\theta) - \mathbb{E}_{\theta^*}\left(C_i^K(\theta)\right).$$

So, $\bar{M}_N^K(\theta)$ is defined as a sum of independent and centered variables.
In order to establish that if (A1-A4) hold then almost surely

$$\sup_{K\geq 1}\sup_\theta \bar{M}_N^K(\theta) \xrightarrow[N\to\infty]{} 0,$$

we proceed by contradiction .
If the previous convergence failed, there would exist a strictly increasing sequence $(N_r)_{r\geq 1}$ which would tend to infinity, and two sequences $(\theta_r)_r$ and $(K_r)_r$ both associated with $(N_r)_r$, such that

$$\bar{M}_{N_r}^{K_r}(\theta_r) \nrightarrow 0, \text{ when } r \to \infty. \tag{5}$$

From Borel-Canteli's lemma and Chebyshev's inequality, (5) cannot hold if there exist $\delta > 0$ such that

$$\sum_{r=1}^\infty \mathbb{E}_{\theta^*}\left|\bar{M}_{N_r}^{K_r}(\theta_r)\right|^\delta < \infty, \tag{6}$$

Now, notice that

$$\begin{aligned}
&\sum_{r=1}^{\infty} \mathbb{E}_{\theta^*} \bar{M}_{N_r}^{K_r}(\theta_r)^4 \\
&= \sum_{r=1}^{\infty} \frac{1}{N_r^4} \left[\sum_{i=1}^{N_r} \mathbb{E}_{\theta^*} M_i^{K_r}(\theta_r)^4 + 6 \sum_{i \neq j}^{N_r} \mathbb{E}_{\theta^*} \left(M_i^{K_r}(\theta_r)^2 \right) \mathbb{E}_{\theta^*} \left(M_j^{K_r}(\theta_r)^2 \right) \right] \\
&\leq \left(\sup_{\theta, K} \sup_i \mathbb{E}_{\theta^*} M_i^K(\theta)^4 \right) \sum_{r=1}^{\infty} \frac{1}{N_r^3} + 6 \left(\sup_{\theta, K} \sup_i \mathbb{E}_{\theta^*}^2 \left(M_i^K(\theta)^2 \right) \right) \sum_{r=1}^{\infty} \frac{N_r(N_r-1)}{N_r^4},
\end{aligned}$$

and,

$$\sum_{r=1}^{\infty} \frac{1}{N_r^3} \leq \sum_{r=1}^{\infty} \frac{N_r(N_r-1)}{N_r^4} < \infty$$

and the expressions in brackets are bounded in view of assumptions (A1-A4).
So, we have shown that (6) holds for $\delta = 4$, which contradicts (5).
This accomplishes the proof of the lemma.
□

Proof of the Proposition 2

This demonstration proceeds in three steps .

Step 1. We demonstrate that the normalized sum of variances

$$\frac{1}{N} \sum_{i=1}^{N} \mathrm{var}_{\theta^*} \left(\nabla_\theta C_i^K(\theta^*) \right)$$

converge when N tends to infinity and uniformly in K to a limit $\Gamma_K(\theta^*)$. Thus, $\frac{1}{N}\Gamma_K(\theta^*)$ is an asymptotic equivalent of $\mathrm{var}_{\theta^*} \left(\nabla_\theta \bar{C}_N^K(\theta^*) \right)$, when $N \longrightarrow \infty$.

Furthermore, we show that for K sufficiently large

$$\mathbb{E}_{\theta^*} \nabla_\theta \bar{C}_N^K(\theta^*) = O\left(\frac{1}{K}\right).$$

Step 2. We establish that for N and K sufficiently large

$$N^{\frac{1}{2}} \nabla_\theta \bar{C}_N^K(\theta^*) \sim \mathcal{N}_p \left(O\left(\frac{N^{\frac{1}{2}}}{K}\right), \Gamma_K(\theta^*) \right).$$

Step 3. We deduce from the uniform convergence obtained in lemma and in view of differentiability assumption (A4) that uniformly on Θ,

$$H_\theta \bar{C}_N^K(\theta) - E_{\theta^*} H_\theta \bar{C}_N(\theta) \xrightarrow[\substack{N \to \infty \\ K \to \infty}]{} 0, \text{ almost surely.} \tag{7}$$

Finally, we establish the asymptotic normality of $\hat{\theta}_N^K$, by considering the following Taylor expansion with Lagrange remainder.
We have, for all $1 \le l \le p$,

$$\begin{aligned} 0 &= N^{\frac{1}{2}} \frac{\partial}{\partial \theta_l} \bar{C}_N^K(\hat{\theta}_N^K) \\ &= N^{\frac{1}{2}} \frac{\partial}{\partial \theta_l} \bar{C}_N^K(\theta^*) + N^{\frac{1}{2}} \sum_{m=1}^{p} (\hat{\theta}_{N,m}^K - \theta_m^*) \int_0^1 \frac{\partial}{\partial \theta_l \partial \theta_m} \bar{C}_N^K \left(\theta^* + s(\hat{\theta}_N^K - \theta^*)\right) ds. \end{aligned}$$

Now from (7) and proposition 1, we have for all $1 \le l, m \le p$,

$$\int_0^1 \frac{\partial}{\partial \theta_l \partial \theta_m} \bar{C}_N^K \left(\theta^* + s(\hat{\theta}_N^K - \theta^*)\right) ds \; - \; \mathbb{E}_{\theta^*} \left(\frac{\partial}{\partial \theta_l \partial \theta_m} \bar{C}_N(\theta^*) \right) \xrightarrow[\substack{N \to \infty \\ K \to \infty}]{} 0, \text{ almost surely.}$$

Therefore from the previous steps we conclude that when N and K tend to infinity, with $N^{\frac{1}{2}}/K \to 0$,

$$N^{\frac{1}{2}} \left(\hat{\theta}_N^K - \theta^*\right) \sim \mathcal{N}_p \left(0, I^{-1}(\theta^*)\Gamma(\theta^*)I^{-1}(\theta^*)\right).$$

Proof of step 1. Using (A3) and (A4), we note that $\sup_i \mathrm{var}_{\theta^*} \left(\nabla_\theta C_i^K(\theta^*)\right)$ is bounded uniformly in K.
Since $\sum_{i=1}^{N} \mathrm{var}_{\theta^*} \left(\nabla_\theta C_i^K(\theta^*)\right) = O(N)$, the normalized sum

$$\frac{1}{N} \sum_{i=1}^{N} \mathrm{var}_{\theta^*} \left(\nabla_\theta C_i^K(\theta^*)\right)$$

converge when N tends to infinity, uniformly in K, and let us denote its limit by
$\Gamma_K(\theta^*) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathrm{var}_{\theta^*} \left(\nabla_\theta C_i^K(\theta^*)\right)$.

In an other way, using notations from the proof of the lemma, we have (when K tends to infinity)

$$\begin{aligned} & \mathbb{E}_{\theta^*} \left(\nabla_\theta \bar{C}_N^K(\theta^*) - \nabla_\theta \bar{C}_N(\theta^*)\right) \\ = \; & \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{\theta^*} \left[\nabla_\theta f_i(\bar{\xi}_{i,K}(\theta^*)) - \nabla_\theta f_i(m_i(\theta^*))\right] \\ = \; & \frac{1}{N} \sum_{i=1}^{N} \left(\nabla_{\xi_i} \nabla_\theta f_i(m_i(\theta^*))\right) \mathbb{E}\left(\bar{\xi}_{i,K}(\theta^*) - m_i(\theta^*)\right) + O\left(\|\bar{\xi}_{i,K}(\theta^*) - m_i(\theta^*)\|^2\right) \\ = \; & \frac{1}{N} \sum_{i=1}^{N} \frac{1}{K} O\left(\mathrm{tr}(W_i(\theta^*))\right) = \frac{1}{K} O\left(\mathrm{tr}(W_i(\theta^*))\right) = O\left(\frac{1}{K}\right). \end{aligned}$$

In view of Jensen's inequality (2) given in section 3, $\mathbb{E}_{\theta^*} \left(\nabla_\theta \bar{C}_N(\theta^*)\right) = 0$. Therefore, for K sufficiently large,

$$\mathbb{E}_{\theta^*} \left(\nabla_\theta \bar{C}_N^K(\theta^*)\right) = O\left(\frac{1}{K}\right).$$

Proof of step 2. Let us define the normalized sum

$$\nabla_\theta \bar{M}_N^K(\theta) = \frac{1}{N}\sum_{i=1}^{N} \nabla_\theta M_i^K(\theta).$$

Thus from (A3) and (A4), the vector $\nabla \bar{M}_N^K(\theta)$ is a sum of independent and centered vectors, which are square integrable uniformly in K.

Thus we can now, under assumption (A6), establish asymptotic normality (when N tends to infinity) of $\nabla_\theta \bar{M}_N^K(\theta)$. We obtain

$$\mathrm{var}_{\theta^*}\left(\nabla_\theta \bar{C}_N^K(\theta^*)\right)^{-\frac{1}{2}}\left(\nabla_\theta \bar{C}_N^K(\theta^*) - \mathbb{E}_{\theta^*}\left(\nabla_\theta \bar{C}_N^K(\theta^*)\right)\right)\xrightarrow[N\to\infty]{} \mathcal{N}_p(0, I) \tag{8}$$

uniformly in K.

Finally, we deduce from step 1, that for N and K sufficiently large

$$N^{\frac{1}{2}}\nabla_\theta \bar{C}_N^K(\theta^*) \sim \mathcal{N}_p\left(O\left(\frac{N^{\frac{1}{2}}}{K}\right), \Gamma_K(\theta^*)\right).$$

□

Expression of the variances $\mathrm{var}_{\theta^*}(\nabla_\theta C_i(\theta^*))$

The $(l,m)^{th}$ element of this $p \times p$ matrix is equal to

$$\begin{aligned}
&\mathbb{E}_{\theta^*}\left[\frac{\partial}{\partial\theta_l}C_i(\theta^*)\frac{\partial}{\partial\theta_m}C_i(\theta^*)\right] \\
= \;& 4\left[\frac{\partial}{\partial\theta_l}\mu_i(\theta^*)\right]' V_i^{-1}(\theta^*)\frac{\partial}{\partial\theta_m}\mu_i(\theta^*) \\
&+ \mathbb{E}_{\theta^*}\left(\mathrm{tr}\left(D_i^l(\theta^*)\sigma^2\Lambda_i^2(\Phi_i)\right)\|\eta_i(\Phi_i)-\mu_i(\theta^*)\|^2_{D_i^m(\theta^*)}\right) \\
&+ \mathbb{E}_{\theta^*}\left(\mathrm{tr}\left(D_i^m(\theta^*)\sigma^2\Lambda_i^2(\Phi_i)\right)\|\eta_i(\Phi_i)-\mu_i(\theta^*)\|^2_{D_i^l(\theta^*)}\right) \\
&+ \mathbb{E}_{\theta^*}\left(\|\eta_i(\Phi_i)-\mu_i(\theta^*)\|^2_{D_i^l(\theta^*)}\|\eta_i(\Phi_i)-\mu_i(\theta^*)\|^2_{D_i^m(\theta^*)}\right) - \mathrm{tr}\left(D_i^l(\theta^*)V_i(\theta^*)\right)\;\mathrm{tr}\left(D_i^m(\theta^*)V_i(\theta^*)\right) \\
&+ 2\mathbb{E}_{\theta^*}\left[\mathrm{tr}\left(D_i^l(\theta^*)\sigma^2\Lambda_i^2(\Phi_i)D_i^m(\theta^*)\sigma^2\Lambda_i^2(\Phi_i)\right)\right] + \mathbb{E}_{\theta^*}\left[\mathrm{tr}\left(D_i^l(\theta^*)\sigma^2\Lambda_i^2(\Phi_i)\right)\;\mathrm{tr}\left(D_i^m(\theta^*)\sigma^2\Lambda_i^2(\Phi_i)\right)\right] \\
&+ 2\left[\frac{\partial}{\partial\theta_l}\mu_i(\theta^*)\right]' V_i^{-1}(\theta^*)\mathbb{E}_{\theta^*}\left[(\eta_i(\Phi_i)-\mu_i(\theta^*))\left\{2\,\mathrm{tr}\left(D_i^m(\theta^*)\sigma^2\Lambda_i^2(\Phi_i)\right) + \|\eta_i(\Phi_i)-\mu_i(\theta^*)\|^2_{D_i^m(\theta^*)}\right\}\right] \\
&+ 2\left[\frac{\partial}{\partial\theta_m}\mu_i(\theta^*)\right]' V_i^{-1}(\theta^*)\mathbb{E}_{\theta^*}\left[(\eta_i(\Phi_i)-\mu_i(\theta^*))\left\{2\,\mathrm{tr}\left(D_i^l(\theta^*)\sigma^2\Lambda_i^2(\Phi_i)\right) + \|\eta_i(\Phi_i)-\mu_i(\theta^*)\|^2_{D_i^l(\theta^*)}\right\}\right],
\end{aligned}$$

where

$$D_i^l(\theta^*) = V_i^{-1}(\theta^*)\frac{\partial}{\partial\theta_l}V_i(\theta^*)V_i^{-1}(\theta^*).$$

References

AMARI, S.I.. (1982). Differential geometry of curved exponential families. Curvature and information loss. *Annals of Probability* **10**, 357-85.

ANDREWS, D. (1987). Consistency in nonlinear econometric models : a generic uniform law of large numbers. *Econometrica* **55**, 1465-71.

BEAL, S.L. & SHEINER, L.B. (1992). Nonmem user's guide. Nonlinear mixed effects models for repeated measures data. *University of California, San Francisco.*

DRAPER, N.R. & SMITH, H. (1981). Applied regression analysis (2nd ed.), Wiley, New York.

FELLER, W. (1971). An introduction to probability theory and its applications, vol.2. *Wiley, New York.*

GOURIEROUX, C. & MONFORT, A. (1991). Simulation based inference in models with heterogeneity. *Annales d'Economie et de Statistique* **20/21**, 69-107.

LINDSTOM, M.J. & BATES, D.M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**, 673-87.

PINHEIRO, J.C. & BATES, D.M. (1995). Approximation of the log-likelihood function in the nonlinear mixed-effects model. *Journal and Computational and Graphical Statistics.* **4**, 12-35.

RAMOS, R.Q. & PANTULA, S.G. (1995). Estimation of nonlinear random coefficient models. *Statistics and Probability Letters* **24**,49-56.

VONESH, E.F. (1996). A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika* **83**, 447-52.

WOLFINGER, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika* **80**, 791-95.