International Journal of Pure and Applied Mathematics

Volume 32 No. 4 2006, 435-443

STRONG CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATOR IN EXPONENTIAL FAMILY NONLINEAR MODELS

Tian Xia[1 §], Xue-Ren Wang[2], Yan Hu[3], Xue-Jun Jiang[4]

[1,2,4]Department of Statistics
Yunnan University
Kunming, 650091, P.R. CHINA
[1]e-mails: xiatian718@hotmail.com, xiatian403@yahoo.com.cn
[2]e-mail: wangxr_34@hotmail.com
[3]Department of Basic Studies
Vocational and Technical College
Anhui University
Hefei, 230011, P.R. CHINA

Abstract: This paper proposes the condition of eigenvalue, which is weaker than the corresponding condition in literature, and some other regularity conditions. On the basis of the proposed regularity conditions, we show the strong consistency of the maximum likelihood estimator (MLE) in exponential family nonlinear models (EFNM). Our results may be regarded as an essential improvement comparing to the relevant results in literature.

AMS Subject Classification: 62F12, 62J12
Key Words: exponential family nonlinear models, strong consistency, maximum likelihood estimator

1. Introduction

Suppose that the random variables $y_1, \dots, y_n$ are independent and each y_i has density:

$$p(y_i;\theta_i) = c(y_i)\,\exp\{y_i\theta_i - b(\theta_i)\},\ \ c \geq 0 \ \ \text{measurable}, i = 1, \dots, n, \qquad (1.1)$$

with respect to a σ-finite measure v, where b and c are known function, θ_i

Received: October 13, 2005

§Correspondence author

are the natural parameter. We assume Θ to be the natural parameter space. Then Θ is an interval in $\mathcal{R}$, and in the interior Θ° of Θ, all derivatives of $b(\theta)$ and all moments of y exist (we assume $\Theta^\circ \neq \emptyset$). In particular, we have $\mu_i = E(y_i) = \dot{b}(\theta_i), \mathrm{Var}(y_i) = \ddot{b}(\theta_i)$. In this paper, we use dots over the functions to denote the derivatives.

Suppose that there exists a link function G which is a known strictly monotone differentiable function such that

$$G(\mu_i) = f(x_i, \beta), \tag{1.2}$$

where $f(.;.)$ is a known function, x_i is a known q-vector defined in a subset $\mathcal{X}$ of R^q, β is the unknown p-vector parameter defined in a subset $\mathcal{B}$ of R^p. Then the model defined by (1.1)-(1.2) is called an exponential family nonlinear model (EFNM), which includes the generalized linear model (GLM) (see [4]) and the normal nonlinear model (see [1]) as their special cases.

For GLM, Fahrmeir and Kaufmann in [3], discussed strong consistency of the maximum likelihood estimator for the cases of natural link functions and nonnatural link functions, respectively. In the former case, they showed that strong consistency of the maximum likelihood estimator (MLE) under $\lambda_{min} F_n(\beta) \geq c(\lambda_{max} F_n(\beta_0))^\alpha$ $(\alpha > 1/2)$ and some other regularity conditions, where $F_n(\beta)$ is the Fisher information of the first n observations, $\lambda_{min}(\lambda_{max})$ denotes the smallest (largest) eigenvalue of a symmetric matrix. In the latter case, they also showed that strong consistency of MLE, but the condition of eigenvalue becomes $\lambda_{min} H_n(\beta) \geq c(\lambda_{max} F_n(\beta_0))^\alpha$ $(\alpha > 1/2)$, where $H_n(\beta)$ is the observed information matrix.

For EFNM, Wei in [6], pp. 195-210, investigated the existence, the strong consistency and the asymptotic normality of maximum likelihood estimator (MLE) under some regularity conditions. One of these conditions is Assumption C (b) of Wei (see [6], p. 57), i.e., there exists a positive definite and continuous matrix $L(\beta)$ such that

$$\frac{1}{n} F_n(\beta) \to L(\beta) \text{ uniformly in } \overline{N}(\delta) = \{\beta : ||\beta - \beta_0|| \leq \delta\} \ (\delta > 0),$$

which implies that $\lambda_{min} F_n(\beta) \geq cn$ for sufficiently large n, where $F_n(\beta)$ is the Fisher information matrix of the first n observations.

The above conditions of eigenvalue are too strong. In this paper, we weaken the eigenvalue conditions given by Fahrmeir and Kaufmann in [3], and Wei in [6], and add some other mild conditions to obtain the strong consistency of the maximum likelihood estimator in EFNM.

This paper is organized as follows. Section 2 introduces some regularity conditions and lemmas. In Section 3, we show the strong consistency of MLE in EFNM under some mild regularity conditions.

2. Conditions and Lemmas

From (1.2), we have $\mu_i = G^{-1}(f(x_i,\beta)) = G^{-1} \circ f(x_i,\beta)$, where $\circ$ denotes the product of two functions (as mappings). Let $\mu = G^{-1} \circ f, \mu_i \triangleq \mu_i(\beta) = \mu(x_i,\beta) = G^{-1} \circ f(x_i,\beta)$. Then the log-likelihood of a sample $y_1,\ldots,y_n$ can be expressed by:

$$\begin{aligned} &l_n(\beta) = \sum_{i=1}^{n} \{y_i\theta_i - b(\theta_i)\} + C, \\ &\theta_i = \dot{b}^{-1}(\mu_i), \quad \mu_i = \mu_i(\beta) = \mu(x_i,\beta), \quad i = 1,2,\ldots,n, \end{aligned} \tag{2.1}$$

where C does not depend on β.

It is easily seen from (2.1) that log-likelihood equation is given by

$$L_n(\beta) = \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} (\ddot{b}(\theta_i))^{-1} (y_i - \mu_i(\beta)) = 0. \tag{2.2}$$

Before formulating the assumptions, we introduce some notations. Let $\hat{\beta}_n$ denote the maximum likelihood estimator of β, which is the solution of the likelihood equation $L_n(\beta) = 0$; c represents an absolute positive constant which may take different values in each of its appearances, even in the same expression; β_0 is the true value of β. For a matrix $B = (b_{ij}) \in R^{p\times q}$, set $||B|| = (\sum_{i=1}^{p}\sum_{j=1}^{q} |b_{ij}|^2)^{1/2} = (\mathrm{tr}(B^TB))^{1/2}$. For notational simplicity, we would mostly drop the argument β_0 in E_{β_0}, P_{β_0}, etc., and use E, P, etc., instead.

To make inference for β we make the following assumptions.

Assumptions. (i) $\mathcal{X}$ is a compact subset in $\mathcal{R}^q$, and $\mathcal{B}$ is an open subset in $\mathcal{R}^p$,

(ii) $\mu(x,\beta)$, as a function of β, is differentiable up to the third order; The function $\mu(x,\beta)$ and all its derivatives are continuous in $\mathcal{X} \times \mathcal{B}$,

(iii) $D(\beta) = \partial\mu(\beta)/\partial\beta^T = (D_1,\ldots,D_n)^T$ is of full rank in column and $\inf_{\mathcal{X}} \det(\frac{\partial \mu}{\partial \beta}\frac{\partial \mu}{\partial \beta^T}) \neq 0$, where $\mu(\beta) = (\mu_1,\ldots,\mu_n)^T$,

(iv) $\sup_{i\geq 1} ||\frac{\partial \mu_i}{\partial \beta}|| < \infty$, and for sufficiently large n and some $\gamma \in (0,1]$, $\lambda_{min}(\sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta}\frac{\partial \mu_i}{\partial \beta^T}) \geq cn^\gamma$,

(v) $\mu_i = Ey_i = \mu(x_i, \beta)$ $(i = 1, 2, \dots)$, and for some $\alpha \geq 1/\gamma$, $\sup_{i\geq 1} E|y_i|^\alpha < \infty$,

(vi) $0 < \inf_{i\geq 1} \ddot{b}(\theta_i) \leq \sup_{i\geq 1} \ddot{b}(\theta_i) < \infty$.

For later use, we formulate two lemmas.

Lemma 1. (Bernstein Inequality, see Bennett, [2]) *Suppose that* $X_1, \dots, X_n$ *are independent random variables with zero expectation, and there exists a finite constant* b *such that* $|X_i| \leq b, 1 \leq i \leq n$. *Then for any* $\varepsilon > 0$,

$$P(\sum_{i=1}^{n} X_i \geq \varepsilon) \leq exp\{-\varepsilon^2/(2v + \frac{2b\varepsilon}{3})\}, \quad v = \sum_{i=1}^{n} EX_i^2.$$

Lemma 2. (see Ortega and Rheinboldt, [5], Corallary 6.3.4, p. 162-163) *Suppose that* C *is an open and bounded set in* R^n, $\overline{C}$ *and* ∂C *denote the closure and boundary of* C, *respectively. Suppose that* $F : \overline{C} \to R^n$ *is continuous, and satisfies* $(x - x^0)^T F(x) \leq 0$ *for some* $x^0 \in C$ *and for all* $x \in \partial C$, *then the equation* $F(x) = 0$ *has a solution in* $\overline{C}$.

3. Main Results

Theorem 1. *Suppose that Assumptions (i)-(vi) are satisfied. Then there exists a sequence* $\{\hat{\beta}_n\}$ *of estimates of* β_0 *such that with probability one for sufficiently large* n:

$$L_n(\hat{\beta}_n) = 0 \;\text{ and } \hat{\beta}_n \to \beta_0 \text{ a.s. (as } n \to \infty). \tag{3.1}$$

Proof. Taking ε such that

$$0 < 2\varepsilon\gamma < 1,\; t = \frac{1}{\gamma} + 2\varepsilon < \alpha,\; \delta = \gamma - \varepsilon\gamma^2 - \frac{1}{t} > 0,\; \rho_n = n^{-\delta} \to 0. \tag{3.2}$$

Let $S_{\rho_n} = \{\beta \in R^p : ||\beta - \beta_0|| \leq \rho_n\}$, and $\partial S_{\rho_n} = \{\beta \in R^p : ||\beta - \beta_0|| = \rho_n\}$. To prove (3.1), it is sufficient to show that with probability one for sufficiently large n

$$\sup_{\beta\in\partial S_{\rho_n}} \{(\beta - \beta_0)^T L_n(\beta)\} < 0. \tag{3.3}$$

Let $\eta = \beta - \beta_0, e_i = y_i - \mu(\beta_0)$. From Mean Value Theorem and Schwarz inequality, it follows that for $\beta \in \partial S_{\rho_n}$:

$$
\begin{aligned}
(\beta-\beta_0)^T L_n(\beta) &= \eta^T \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\theta_i))^{-1} e_i - \eta^T \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\beta_i))^{-1}(\frac{\partial \mu_i}{\partial \beta^T}|_{\beta=\beta^{i*}})\eta \\
&= \eta^T \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\theta_i))^{-1} e_i - \eta^T \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\theta_i))^{-1} \frac{\partial \mu_i}{\partial \beta^T}\eta \\
&+ \eta^T \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\theta_i))^{-1}[\frac{\partial \mu_i}{\partial \beta^T} - \frac{\partial \mu_i}{\partial \beta^T}|_{\beta=\beta^{i*}}]\eta \le J_n(\beta) - K_n(\beta) \\
&+ [K_n(\beta)]^{1/2}[\eta^T \sum_{i=1}^n (\frac{\partial \mu_i}{\partial \beta} - \frac{\partial \mu_i}{\partial \beta}|_{\beta=\beta^{i*}})(\ddot{b}(\theta_i))^{-1}(\frac{\partial \mu_i}{\partial \beta^T} - \frac{\partial \mu_i}{\partial \beta^T}|_{\beta=\beta^{i*}})\eta]^{1/2},
\end{aligned}
$$

where β^{i*} on the line segment between β and β_0,

$$
\begin{aligned}
J_n(\beta) &= \eta^T \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\theta_i))^{-1} e_i, \text{ and} \\
K_n(\beta) &= \eta^T \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\theta_i))^{-1} \frac{\partial \mu_i}{\partial \beta^T}\eta \ge c\eta^T \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta}\frac{\partial \mu_i}{\partial \beta^T}\eta.
\end{aligned} \tag{3.4}
$$

From Assumptions (i)-(iii), (vi) and (3.2), it follows that for all $\beta \in \partial S_{\rho_n}$ and $i \ge 1$

$$\lambda_{min}\{\frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\theta_i))^{-1}\frac{\partial \mu_i}{\partial \beta^T}\} \ge c > 0,$$

and that $||(\ddot{b}(\theta_i))^{-1/2}(\frac{\partial \mu_i}{\partial \beta} - \frac{\partial \mu_i}{\partial \beta}|_{\beta=\beta^{i*}})|| \to 0$ (as $n \to \infty$) uniformly for all $\beta \in \partial S_{\rho_n}$ and all $i \le n$. Therefore for any λ with $\lambda^T\lambda = 1$ and sufficiently large n, we have

$$
\begin{aligned}
&\lambda^T[\frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\theta_i))^{-1}\frac{\partial \mu_i}{\partial \beta^T} - (\frac{\partial \mu_i}{\partial \beta} - \frac{\partial \mu_i}{\partial \beta}|_{\beta=\beta^{i*}})(\ddot{b}(\theta_i))^{-1}(\frac{\partial \mu_i}{\partial \beta^T} - \frac{\partial \mu_i}{\partial \beta^T}|_{\beta=\beta^{i*}})]\lambda \\
&\quad \ge \lambda_{min}(\frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\theta_i))^{-1}\frac{\partial \mu_i}{\partial \beta^T}) - ||(\ddot{b}(\theta_i))^{-1/2}(\frac{\partial \mu_i}{\partial \beta} - \frac{\partial \mu_i}{\partial \beta}|_{\beta=\beta^{i*}})|| > c/2 > 0,
\end{aligned}
$$

implying for sufficiently large n,

$$(\beta-\beta_0)^T L_n(\beta) \le J_n(\beta) - G_n(\beta), \tag{3.5}$$

where $G_n(\beta) = c\eta^T \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta}\frac{\partial \mu_i}{\partial \beta^T}\eta$ (for some $c > 0$). By (3.5), it is sufficient to prove for sufficiently large n,

$$\sup_{\beta \in \partial S_{\rho_n}} \{J_n(\beta) - G_n(\beta)\} < 0. \tag{3.6}$$

Let $\bar{e}_i = e_i I(|e_i| \le i^{1/t}), e_i^* = \bar{e}_i - E\bar{e}_i$, where $I(.)$ is the indicator function of the relevant event. By Markov inequality, Assumptions (vi), (v) and (3.2), we have

$$|E\bar{e}_i| = |-E(e_i I(|e_i| > i^{1/t}))| \le i^{-(\alpha-1)/t} E|e_i|^\alpha, \tag{3.7}$$

$$\sum_{i=1}^{\infty} P(\bar{e}_i \neq e_i) \le \sup_{i\ge1} E|e_i|^\alpha \sum_{i=1}^{\infty} i^{-\alpha/t} < \infty,$$

by Borel-Cantelli Lemma, with probability one for sufficiently large n,

$$\bar{e}_n = e_n. \tag{3.8}$$

By (3.2) and Assumption (iv) as $n \to \infty$,

$$\inf_{\beta\in\partial S_{\rho n}} G_n(\beta) \ge cn^{\gamma-2\delta} \ge cn^{\gamma/(1+2\varepsilon\gamma)} \to \infty. \tag{3.9}$$

Let

$$J_n^*(\beta) = \sum_{i=1}^{n} \eta^T \frac{\partial\mu_i}{\partial\beta}(\ddot{b}(\theta_i))^{-1} e_i^*, \quad \bar{J}_n(\beta) = \sum_{i=1}^{n} \eta^T \frac{\partial\mu_i}{\partial\beta}(\ddot{b}(\theta_i))^{-1} \bar{e}_i. \tag{3.10}$$

By (3.8) and (3.9), to prove (3.6), it is sufficient to show that with probability one for sufficiently large n,

$$\sup_{\beta\in\partial S_{\rho n}} \{J_n^*(\beta) - \frac{G_n(\beta)}{3}\} < 0, \tag{3.11}$$

and

$$\sup_{\beta\in\partial S_{\rho n}} \{E\bar{J}_n(\beta) - \frac{G_n(\beta)}{3}\} < 0. \tag{3.12}$$

At first, we prove that (3.11) holds. By (3.2), we can divide ∂S_{ρ_n} into M parts, $U_1, U_2, \ldots, U_M$, such that the diameter of each part less than n^{-2}, and $M \le [(2n^2+1)^p]$. For any fixed $\beta_j \in U_j$, let $\eta_j = \beta_j - \beta_0, j = 1, \ldots, M$. From Assumptions (iv)-(vi), $|e_i^*| \le 2i^{1/t}$ and (3.2), it follows that for $j = 1, 2, \ldots, M, i = 1, 2, \ldots, n$.

$$|\eta^T \frac{\partial\mu_i}{\partial\beta}(\ddot{b}(\theta_i))^{-1} e_i^*| \le ||\eta^T|| || \frac{\partial\mu_i}{\partial\beta} || |(\ddot{b}(\theta_i))^{-1}| |e_i^*|$$
$$\le c||\eta^T|| i^{1/t} \le cn^{1/t-\delta}, \tag{3.13}$$

$$E(e_i^*)^2 \le \begin{cases} c, & \alpha \ge 2, \\ ci^{(2-\alpha)/t}, & 1 < \alpha < 2, \end{cases}$$

$$\begin{aligned}&\text{Var}(\eta^T \frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\theta_i))^{-1} e_i^*)\\ &\qquad \le c\eta^T \frac{\partial \mu_i}{\partial \beta}\frac{\partial \mu_i}{\partial \beta^T}\eta(i^{(2-\alpha)/t}+c) < c\eta^T \frac{\partial \mu_i}{\partial \beta}\frac{\partial \mu_i}{\partial \beta^T}\eta n^{1/t-\delta}, \quad (3.14)\end{aligned}$$

From Lemma 2.1, (3.5), (3.10), (3.13), (3.14), Assumptions (i), (ii), (vi), and (3.2), it follows that

$$\begin{aligned}P\{J_n^*(\beta_j) \ge G_n(\beta_j)/4\} &\le \exp\{-c\sum_{i=1}^n \eta_j^T \frac{\partial \mu_i}{\partial \beta}\frac{\partial \mu_i}{\partial \beta^T}\eta_j / n^{1/t-\delta}\}\\ &\le \exp\{-cn^{\gamma-\delta-1/t}\} = \exp\{-cn^{\varepsilon\gamma^2}\}, \quad (3.15)\end{aligned}$$

and therefore

$$\begin{aligned}&\sum_{i=1}^{\infty} P(\bigcup_{1\le j\le M}\{J_n^*(\beta_j) \ge G_n(\beta_j)/4\})\\ &\qquad \le \sum_{i=1}^{\infty}(2n^2+1)^p \exp\{-cn^{\varepsilon\gamma^2}\} < \infty. \quad (3.16)\end{aligned}$$

By Borel-Cantelli Lemma, for sufficiently large n,

$$J_n^*(\beta_j) \le \frac{G_n(\beta_j)}{4}, \text{ a.s. } j = 1, 2, \ldots, M. \quad (3.17)$$

For any given $\beta \in \partial S_{\rho_n}$, we can find $\beta_j \in U_j$ such that $||\beta - \beta_j|| \le n^{-2}$. Now

$$\begin{aligned}&|G_n(\beta) - G_n(\beta_j)| \le c||\eta - \eta_j|| \sum_{i=1}^n ||\frac{\partial \mu_i}{\partial \beta}\frac{\partial \mu_i}{\partial \beta^T}|| ||\eta||\\ &+ c||\eta_j|| \sum_{i=1}^n ||\frac{\partial \mu_i}{\partial \beta}\frac{\partial \mu_i}{\partial \beta^T}|| ||\eta - \eta_j|| + c||\eta_j|| \sum_{i=1}^n ||\frac{\partial \mu_i}{\partial \beta}\frac{\partial \mu_i}{\partial \beta^T} - \frac{\partial \mu_i}{\partial \beta}\frac{\partial \mu_i}{\partial \beta^T}|_{\beta=\beta_j}|| ||\eta_j||.\end{aligned}$$

Choose arbitrarily an element of the matrix

$$\frac{\partial \mu_i}{\partial \beta}\frac{\partial \mu_i}{\partial \beta^T} - \frac{\partial \mu_i}{\partial \beta}\frac{\partial \mu_i}{\partial \beta^T}|_{\beta=\beta_j},$$

which has the form

$$M = (g_i(\beta) - g_i(\beta_j)),$$

for some function g_i. Since $\beta_j \in U_j$ and $||\beta - \beta_j|| \leq n^{-2}$, from Assumptions (i), (ii) and the Mean Value Theorem, we have

$$|M| = |\frac{\partial g_i(\beta)}{\partial \beta}|_{\beta=\beta^{i**}}(\beta - \beta_j)| \leq ||\frac{\partial g_i(\beta)}{\partial \beta}|_{\beta=\beta^{i**}}||||\beta - \beta_j|| \leq cn^{-2},$$

where β^{i**} belongs to the line segment between β and β_j. Therefore, from $\eta = \beta - \beta_0$, $||\eta|| = \rho_n$, $||\eta - \eta_j|| \leq n^{-2}$ and Assumptions (i), (ii), we have

$$|G_n(\beta) - G_n(\beta_j)| \leq cn^{-2} \cdot n \cdot \rho_n + c\rho_n \cdot n \cdot n^{-2} + c\rho_n \cdot n \cdot n^{-2}\rho_n \leq c. \quad (3.18)$$

Similarly, we have

$$|J_n^*(\beta) - J_n^*(\beta_j)| \leq |(\eta - \eta_j)^T \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\theta_i))^{-1} e_i^*| + |\eta_j^T \sum_{i=1}^{n} [\frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\theta_i))^{-1} - (\frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\theta_i))^{-1})|_{\beta=\beta_j}] e_i^*|.$$

From $||\beta_j - \beta_0|| = \rho_n, ||\beta - \beta_j|| \leq n^{-2}, |e_i^*| \leq 2i^{1/t}$, and Assumptions (i), (ii), (vi), and the Mean Value Theorem, we have

$$|J_n^*(\beta) - J_n^*(\beta_j)| \leq c. \quad (3.19)$$

By (3.9) and (3.17)-(3.19), with probability one for sufficiently large n we have

$$\sup_{\beta \in \partial S_{\rho_n}} \{J_n^*(\beta) - \frac{G_n(\beta)}{3}\} < 0,$$

and hence (3.11) holds.

Secondly, we prove that (3.12) holds. By Assumptions (i), (ii), (vi), (3.2) and (3.7), for $\beta \in \partial S_{\rho_n}$, we have

$$\begin{aligned} |E\bar{J}_n(\beta)| &= |\eta^T \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta}(\ddot{b}(\theta_i))^{-1} E\bar{e}_i| \\ &\leq c||\eta|| \sum_{i=1}^{n} i^{-(\alpha-1)/t} \leq cn^{-\delta}(n^{1-(\alpha-1)/t} + \log n) \leq cn^{1/t-\delta}. \end{aligned} \quad (3.20)$$

From (3.2), it follows that $1/t - \delta < \frac{\gamma}{1+2\varepsilon\gamma}$. By (3.9) and (3.20), with probability one for sufficiently large n,

$$\sup_{\beta \in \partial S_{\rho_n}} \{E\bar{J}_n(\beta) - \frac{C_n(\beta)}{3}\} < 0,$$

and hence (3.12) holds.

From (3.11) and (3.12), it follows that (3.6) holds, implying (3.3) holds. From Lemma 2, it follows that with probability one for n sufficiently large, the log-likelihood equation (3.5) has a solution $\hat{\beta}_n \in S_{\rho_n}$, and since $\rho_n \to 0$ ($n \to \infty$),

$$\hat{\beta}_n \to \beta_0 \text{ a.s. } (n \to \infty),$$

which completes the proof of the theorem. □

Remark. Here the condition of eigenvalue is much weaker than those given by Fahrmeir and Kaufmann [3], and Wei [6]. Hence this result essentially improves the existing relevant results in literature.

Acknowledgments

This research is fully supported by grants from Natural Science Foundations of Yunnan University (2004A0002M) and NSFC (10561008), and Department of Education of Yunnan Provice (5Y0062A).

References

[1] D.M. Bates, D.G. Watts, *Nonlinear Regression Analysis and its Applications*, Wiley, New York (1988).

[2] G. Bennett, Probability inequality for sums of independent random variables, *JASA*, **57** (1962), 33-45.

[3] L. Fahrmer, H. Kaufmann, Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models, *Ann. Statist.*, **13** (1985), 342-368.

[4] J.A. Nelder, R.W.M. Wedderburn, Generalized linear models, *J. R. Statist. Soc. A*, **135** (1972), 370-384.

[5] J.M. Ortega, W.C. Rheinboldt, *Iterative Solution of Nonlinear Equation in Several Variables*, Academic Press, New York (1970).

[6] B.C. Wei, *Exponential Family Nonlinear Models*, Springer-Verlag, Singapore (1998), 195-209.