应用概率统计　第二十四卷
第六期　2008年12月

Chinese Journal of Applied Probability
and Statistics Vol.24 No.6 Dec. 2008

Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Exponential Family Nonlinear Models *

XIA TIAN
(*Department of Statistics, Yunnan University, Kunming, 650091*)
KONG FANCHAO
(*Department of Mathematics, Anhui University, Hefei, 230039*)

Abstract

This paper proposes some regularity conditions which weaken those given by Zhu & Wei (1997). On the basis of the proposed regularity conditions, the existence, the strong consistency and the asymptotic normality of maximum likelihood estimation (MLE) are proved in exponential family nonlinear models (EFNMs). Our results may be regarded as a further improvement of the work of Zhu & Wei (1997).

Keywords: Exponential family nonlinear models, consistency, asymptotic normality, maximum likelihood estimator.

AMS Subject Classification: 62F12, 62J02.

§1. Introduction

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

$$\begin{cases} p(y_i;\theta_i,\sigma^2) = \exp\Big\{\dfrac{y_i\theta_i - b(\theta_i)}{\sigma^2} + c(y_i;\sigma)\Big\}, \\ \theta_i = f(x_i,\beta), \qquad i = 1, 2, \cdots, n, \end{cases} \tag{1.1}$$

with respect to a σ-finite measure v, where b, c and f are known functions, x_i is a known q-vector defined in $\mathcal{X}$, β is an unknown p-vector parameter defined in $\mathcal{B}$, θ_i is a natural parameter, and σ is a scale parameter. Then models (1.1) are called the exponential family nonlinear models (EFNMs) which are natural extensions of generalized linear models (Nelder & Wedderburn, 1972) and normal nonlinear regression models. In the past two decades, a number of authors have been concerned about the inference of EFNMs,

*Project supported by Natural Science Foundation of China (10761011, 10561008), Mathematical Tianyuan Fund of Natural Science Foundation of China (10626048), Natural Science Foundation of Yunnan Province.

Received August 5, 2005. Revised September 28, 2007.

Cordeiro and Paula (1989) studied the improved likelihood ratio statistics; Cook and Tsai (1990) discussed the cubic approximations of confidence regions; Wei and Shi (1994) studied some diagnostics problems; Wei (1998) made a systematic study for the theory and the methods of EFNMs.

We first introduce some notation and basic properties of EFNMs to be used later. For all $\beta \in \mathcal{B}$ and a given $\delta > 0$, the neighborhoods of β with radius δ are denoted by $N(\delta) = \{\beta' : \|\beta' - \beta\| < \delta\}$, $\overline{N}(\delta) = \{\beta' : \|\beta' - \beta\| \le \delta\}$, $\partial N(\delta) = \{\beta' : \|\beta' - \beta\| = \delta\}$ separately. Let $\lambda_{\min} A(\lambda_{\max} A)$ denote the smallest (largest) eigenvalue of a symmetric matrix A; and let $A^{1/2}(A^{T/2})$ denote the left (right) square root of the positive definite matrix A, i.e. $A^{1/2}A^{T/2} = A$; denote $A^{-1/2} = (A^{1/2})^{-1}, A^{-T/2} = (A^{T/2})^{-1}$. Since the parameters of interest in EFNMs are β and the maximum likelihood estimators of β and σ^2 can be estimated separately, we may set $\sigma^2 = 1$ for simplicity. Then the distribution of the response y_i in (1.1) belongs to a natural exponential family, the natural parameter space Θ is an interval in $\mathcal{R}$, and $b(\theta)$ is an analytical convex function in Θ. In particular, we have $\mu_i = \mathsf{E}(y_i) = \dot{b}(\theta_i)$, $\mathsf{Var}\,(y_i) = \ddot{b}(\theta_i)$. In this paper, we use dots over the functions to denote the derivatives.

The log likelihood of a sample $y_1, \cdots, y_n$ is given by

$$l_n(\beta) = \sum_{i=1}^n \{y_i\theta_i - b(\theta_i)\} + C, \qquad \theta_i = \theta_i(\beta) = f(x_i, \beta), \quad i = 1, 2, \cdots, n, \tag{1.2}$$

where C does not depend on β.

It is easily seen from (1.2) that the score function, the observed information matrix and the Fisher information matrix for β can be respectively denoted by

$$S_n(\beta) = \dot{l}_n(\beta) = \sum_{i=1}^n \frac{\partial \theta_i}{\partial \beta}(y_i - \mu_i(\beta)) = D^T(\beta)e(\beta), \tag{1.3}$$

$$\begin{aligned} H_n(\beta) &= -\ddot{l}_n(\beta) = \sum_{i=1}^n \Big\{\frac{\partial \theta_i}{\partial \beta} \cdot \frac{\partial \theta_i}{\partial \beta^T} v_i(\beta) - \frac{\partial^2 \theta_i}{\partial \beta \partial \beta^T} e_i(\beta)\Big\} \\ &= D^T(\beta)V(\beta)D(\beta) - [e^T(\beta)][W(\beta)], \end{aligned}$$

$$J_n(\beta) = D^T(\beta)V(\beta)D(\beta), \qquad H_n(\beta) = J_n(\beta) - R_n(\beta), \tag{1.4}$$

where $R_n(\beta) = [e^T(\beta)][W(\beta)]$, $D(\beta) = \partial\theta(\beta)/\partial\beta^T$, $W(\beta) = \partial^2\theta/(\partial\beta\partial\beta^T)$, $\theta(\beta) = (\theta_1(\beta), \theta_2(\beta), \cdots, \theta_n(\beta))^T$, $e(\beta) = (e_1(\beta), e_2(\beta), \cdots, e_n(\beta))^T$, $e_i(\beta) = y_i - \mu_i(\beta)$, $\mu_i(\beta) = \dot{b}(\theta_i)$, $V(\beta) = \mathrm{diag}(v_1(\beta), v_2(\beta), \cdots, v_n(\beta))$, $v_i(\beta) = \ddot{b}(\theta_i)$, $\theta_i = f(x_i, \beta)$, and $[\cdot][\cdot]$ denotes array product (see Wei, 1989).

For EFNMs, Zhu & Wei (1997) discussed the existence, the strong consistence and asymptotic normality of MLE under some regularity conditions. One of those conditions is that there exist a neighborhood $N_0 \subset \mathcal{B}$ of β_0, constant $C > 0$ and a positive integer number m, such that $\lambda_{\min} J_n(\beta) \geq Cn$, $\beta \in N_0$, $n > m$ (Zhu & Wei, 1997, p.196). This condition is too strong, in this paper we weaken this condition to gain the corresponding results.

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

§2. Conditions and Lemmas

To make inference for β we make

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

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

(iii) $\theta_i = f(x_i, \beta) \in \Theta$, $i = 1, 2, \cdots$, for all $x_i \in \mathcal{X}$ and $\beta \in \mathcal{B}$,

(iv) β_0 is unknown true parameter of β and an interior point of $\mathcal{B}$,

(v) $D(\beta) = \partial\theta(\beta)/\partial\beta^T = (D_1, \cdots, D_n)^T$ is full rank, where $\theta(\beta) = (\theta_1, \cdots, \theta_n)^T$.

Remark 1 Condition (v) will ensure that the information matrix $J_n(\beta)$ is positive define for all $\beta \in \mathcal{B}$.

Assumption B (i) $\lambda_{\min} J_n(\beta_0) \to \infty$,

(ii) There is a neighborhood $N \subset \mathcal{B}$ of β_0 such that

$$\lambda_{\min} J_n(\beta) \geq c(\lambda_{\max} J_n(\beta_0)), \qquad \beta \in N,\ n \geq n_0,$$

with some constants $c > 0, n_0$,

(iii)

$$\lambda_{\max} \sum_{i=1}^{n} \frac{\partial^2 f(x_i, \beta)}{\partial\beta\partial\beta^T} \leq c_1 \lambda_{\max} J_n(\beta_0), \qquad n \geq n_0,$$

with some constants $c_1 > 0, n_0$.

Remark 2 Assumption B is similar to those of Fahrmeir & Kaufmann (1985, p.348, p.362).

Assumption C There exists a positive definite and continuous matrix $K(\beta)$ such that

$$\lambda_n^{-1} J_n(\beta) \to K(\beta), \qquad \text{uniformly in } \overline{N}(\delta),$$

here and in the sequel $\lambda_n = \lambda_{\max} J_n(\beta_0)$.

Remark 3 Assumption C is similar to that of Zhu & Wei (1997, p.197) and includes it (when $\lambda_n = n$).

Let $\widehat{\beta}_n$ denote the maximum likelihood estimator of β, which is the solution of the likelihood equation $\dot{l}_n(\beta) = 0$. For notational simplicity, we shall mostly drop the argument β_0 in $S_n(\beta_0)$, $H_n(\beta_0)$, E_{β_0}, P_{β_0} etc. and write S_n, H_n, E, P etc..

Lemma 2.1 Let $\{z_i\}$ be a sequence of independent random variables with $\mathsf{E}z_i = 0$ and $\mathsf{Var}\,(z_i) = \sigma_i^2 > 0$ and $A_n \to \infty$, $\limsup\limits_{n\to\infty} \Big(\sum\limits_{i=1}^n \sigma_i^2\Big)^{1/2+\delta}\Big/ A_n < +\infty$, for some $\delta > 0$, and $g(x,\beta)$ is a continuous function in $\mathcal{X} \times \mathcal{B}_1$, where $\mathcal{B}_1$ is a compact subset in $\mathcal{B}$. Then

$$\lim_{n\to\infty} \sum_{i=1}^n \frac{1}{A_n} z_i g(x_i,\beta) = 0, \quad \text{a.s.} \quad \text{uniformly in } \mathcal{B}_1. \tag{2.1}$$

In particular, if $\beta_n \to \beta_0$ in $\overline{N}(\delta)$, then

$$\frac{1}{A_n}\sum_{i=1}^n z_i g(x_i,\beta_n) \to 0 \qquad \text{(a.s.)}.$$

Proof For any given $\beta \in \mathcal{B}_1$, considering a point $\beta' \neq \beta$ in a neighborhood $N(\delta)$ of β, then

$$\begin{aligned} \Big|\frac{1}{A_n}\sum_{i=1}^n z_i g(x_i,\beta')\Big| &\le \Big|\frac{1}{A_n}\sum_{i=1}^n z_i g(x_i,\beta)\Big| + \Big|\frac{1}{A_n}\sum_{i=1}^n z_i[g(x_i,\beta') - g(x_i,\beta)]\Big| \\ &\le \Big|\frac{1}{A_n}\sum_{i=1}^n z_i g(x_i,\beta)\Big| + \sup_{\mathcal{X}} |g(x,\beta') - g(x,\beta)| \Big\{\frac{1}{A_n}\sum_{i=1}^n |z_i|\Big\}. \end{aligned} \tag{2.2}$$

Since $\mathsf{E}z_i g(x_i,\beta) = 0$, and

$$\begin{aligned} \frac{1}{A_n}\Big(\sum_{i=1}^n \mathsf{Var}\,(z_i g(x_i,\beta))\Big)^{1/2+\delta} &= \frac{1}{A_n}\Big(\sum_{i=1}^n g^2(x_i,\beta)\sigma_i^2\Big)^{1/2+\delta} \\ &\le \sup_{\mathcal{X}\times\mathcal{B}_1} g^2(x,\beta)) \frac{1}{A_n}\Big(\sum_{i=1}^n \sigma_i^2\Big)^{1/2+\delta} < +\infty, \end{aligned}$$

from Lemma 2 of Wu (1981), the first term of (2.2) tends to zero (a.s.). Since $g(x,\beta)$ is uniformly continuous in $\mathcal{X} \times \mathcal{B}_1$, we have $\sup\limits_{\mathcal{X}} |g(x,\beta') - g(x,\beta)| \to 0$, if $\beta' \to \beta$. On the other hand, we have

$$\frac{1}{A_n}\sum_{i=1}^n |z_i| = \frac{1}{A_n}\sum_{i=1}^n \{|z_i| - \mathsf{E}|z_i|\} + \frac{1}{A_n}\sum_{i=1}^n \mathsf{E}|z_i|.$$

Since

$$\frac{1}{A_n}\Big(\sum_{i=1}^n \mathsf{Var}\,(|z_i| - \mathsf{E}|z_i|)\Big)^{1/2+\delta} \le \frac{1}{A_n}\Big(\sum_{i=1}^n \mathsf{Var}\,(z_i)\Big)^{1/2+\delta} = \frac{1}{A_n}\Big(\sum_{i=1}^n \sigma_i^2\Big)^{1/2+\delta},$$

it follows from Lemma 2 of Wu (1981) that

$$\frac{1}{A_n}\sum_{i=1}^{n}\{|z_i|-\mathsf{E}|z_i|\}\to 0 \qquad \text{(a.s.)}.$$

Since $(1/A_n)\cdot\sum_{i=1}^{n}\mathsf{E}|z_i|$ is bounded, $(1/A_n)\cdot\sum_{i=1}^{n}|z_i|$ is a.s. bounded. Collecting together the above, we obtain the following conclusion.

For any given $\varepsilon>0$, there exist a neighborhood $N(\delta_1)$ of β with $0<\delta_1\le\delta$ and a random number n_1, such that

$$\mathsf{P}\Big\{\Big|\frac{1}{A_n}\sum_{i=1}^{n}z_i g(x_i,\beta')\Big|<\varepsilon, n>n_1\Big\}=1$$

for any $\beta'\in N(\delta_1)$. since $\mathcal{B}_1$ is compact, it follows from the finitely covered theorem that there exists $n_2>n_1$ such that

$$\mathsf{P}\Big\{\Big|\frac{1}{A_n}\sum_{i=1}^{n}z_i g(x_i,\beta)\Big|<\varepsilon, n>n_2\Big\}=1$$

for any $\beta\in\mathcal{B}_1$. Thus (2.1) is proved. □

Lemma 2.2 Suppose that Assumptions A and B, and the conditions of Lemma 2.1 hold. If there exists $\delta_0>0$ such that $\beta_n\to\beta_0$ in $\overline{N}(\delta_0)\subset\mathcal{B}_1$, then

$$\lambda_n^{-1}R_n(\beta_n)\to 0 \qquad \text{(a.s.)}. \tag{2.3}$$

Proof $R_n(\beta)$ may be written as

$$\begin{aligned} R_n(\beta) &= \sum_{i=1}^{n}\frac{\partial^2 f(x_i,\beta)}{\partial\beta\partial\beta^T}e_i(\beta_0)+\sum_{i=1}^{n}\frac{\partial^2 f(x_i,\beta)}{\partial\beta\partial\beta^T}(\mu_i(\beta_0)-\mu_i(\beta)) \\ &= B_n(\beta)+C_n(\beta), \end{aligned}$$

where

$$B_n(\beta)=\sum_{i=1}^{n}\frac{\partial^2 f(x_i,\beta)}{\partial\beta\partial\beta^T}e_i(\beta_0), \qquad C_n(\beta)=\sum_{i=1}^{n}\frac{\partial^2 f(x_i,\beta)}{\partial\beta\partial\beta^T}(\mu_i(\beta_0)-\mu_i(\beta)). \tag{2.4}$$

The component of $\lambda_n^{-1}B_n(\beta)$ at (a,b) may be written as

$$\{\lambda_n^{-1}B_n(\beta)\}_{a,b}=\lambda_n^{-1}\sum_{i=1}^{n}W_{iab}(\beta)e_i(\beta_0).$$

From Lemma 2.1, it follows that $\lambda_n^{-1}B_n(\beta_n)\to 0$ (a.s.).

For any λ with $\lambda^T\lambda=1$, we have

$$\begin{aligned} \lambda^T C_n(\beta)\lambda &\le \max_{i=1,2,\cdots,n}|\mu_i(\beta_0)-\mu_i(\beta)|\lambda^T\Big(\sum_{i=1}^{n}\frac{\partial^2 f(x_i,\beta)}{\partial\beta\partial\beta^T}\Big)\lambda \\ &\le c_1\lambda_n\max_{i=1,2,\cdots,n}|\mu_i(\beta_0)-\mu_i(\beta)|, \end{aligned}$$

with some constant c_1. Particularily, we have

$$\frac{\lambda^T C_n(\beta_n)\lambda}{\lambda_n} \le c_1 \max_{i=1,2,\cdots,n} |\mu_i(\beta_0) - \mu_i(\beta_n)|, \tag{2.5}$$

with some constant c_1. Since $\beta_n \to \beta_0$ $(n\to\infty)$, $\max\limits_{i=1,2,\cdots,n} |\mu_i(\beta_0)-\mu_i(\beta_n)|$ can be made arbitrarily small for sufficiently large n, and therefore it follows from (2.5) that $\lambda^T C_n(\beta_n)\lambda/\lambda_n$ converges to zero. Since λ with $\lambda^T\lambda = 1$ is arbitrary and pointwise convergence on the unit ball implies $\|C_n(\beta_n)\|/\lambda_n$ converges to zero, which leads to $\lambda_n^{-1} C_n(\beta_n) \to 0$.

Combining the above results yields $\lambda_n^{-1} R_n(\beta_n) \to 0$ (a.s.). □

Lemma 2.3 Under Assumptions A and B, the score function is asymptotically normal:

$$J_n^{-1/2} S_n \xrightarrow{\mathcal{L}} N(0, I_p),$$

where $\mathcal{L}$ denotes the convergence in distribution and I_p is a $p\times p$ unit matrix.

Proof To get lemma 2.3, it is enough to show that $Z_n = \lambda^T J_n^{-1/2} S_n \xrightarrow{\mathcal{L}} N(0,1)$ holds for any $\lambda^T\lambda = 1$. Z_n may be written as $Z_n = \sum\limits_{i=1}^n \alpha_{ni}\varepsilon_i$, where $\alpha_{ni} = \lambda^T J_n^{-1/2} v_i^{1/2} \partial\theta_i/\partial\beta$, $\varepsilon_i = v_i^{-1/2}(y_i - \mu_i(\beta_0))$. From Assumption A, it is easily seen that

$$\mathsf{E}(\alpha_{ni}\varepsilon_i) = 0, \qquad \mathsf{Var}\Big(\sum_{i=1}^n \alpha_{ni}\varepsilon_i\Big) = 1.$$

We shall show that the Lindeberg condition is satisfied (see Ferguson, 1996, p.27), i.e. for any $\rho > 0$

$$g_n(\rho) = A_n^{-2} \sum_{i=1}^n \mathsf{E}\{\alpha_{ni}^2 \varepsilon_i^2 I(|\alpha_{ni}\varepsilon_i|^2 > \rho^2 A_n^2)\} \to 0 \qquad (n\to\infty), \tag{2.6}$$

where $A_n^2 = \sum\limits_{i=1}^n \mathsf{Var}\,(\alpha_{ni}\varepsilon_i)$. Since $A_n^2 = 1$, inserting into (2.6),

$$\begin{aligned} g_n(\rho) &= \sum_{i=1}^n \mathsf{E}\{\alpha_{ni}^2\varepsilon_i^2 I(|\alpha_{ni}\varepsilon_i|^2 > \rho^2\} \\ &= \sum_{i=1}^n \alpha_{ni}^2 \mathsf{E}\Big\{\varepsilon_i^2 I\Big(\varepsilon_i^2 > \frac{\rho}{\alpha_{ni}^2}\Big)\Big\} \to 0 \qquad (n\to\infty). \end{aligned} \tag{2.7}$$

It follows from B (i) that $\lambda_{\max} J_n^{-1} \to 0$. Using the compactness of $\mathcal{X}$, we have

$$\max_{1\le i\le n} \alpha_{ni}^2 \to 0 \qquad (n\to\infty).$$

Let the distribution function of ε_i be F_X, define

$$h_c(x) = \int_{\{|z|>c\}} z^2 \mathrm{d}F_X.$$

We have, for any (large) $c > 0$,

$$\mathsf{E}\Big\{\varepsilon_i^2 I\Big(\varepsilon_i^2 > \frac{\rho^2}{\alpha_{ni}^2}\Big)\Big\} = \int_{\{z^2>\rho^2/\alpha_{ni}^2\}} z^2 \mathrm{d}F_X \le h_c(x_i), \qquad i = 1, 2, \cdots, n,\ n \ge n_2(c).$$

Combining the above results yields

$$g_n(\rho) \le \sum_{i=1}^n \alpha_{ni}^2 h_c(x_i).$$

By the Helly-Bray Lemma and the continuity properties of exponential function, the function $h_c(x)$ has the following properties: $h_c(x)$ is continuous in $\mathcal{X}$, $h_c(x) \to 0$ (as $c \to +\infty$) (pointwise for any $x \in \mathcal{X}$) and $h_c(x)$ is monotonously decreasing in c. Due to the above properties and the compactness of $\mathcal{X}$, $h_c(x) \to 0$ uniformly in $\mathcal{X}$ as $c \to +\infty$, i.e.

$$\sup_{\mathcal{X}} h_c(x) \to 0 \qquad (c \to +\infty). \tag{2.8}$$

By $\sum \alpha_{ni}^2 \le K < \infty$ with a constant K, we have

$$g_n(\rho) \le K \max_{x\in\mathcal{X}} h_c(x), \qquad n \ge n_2(c). \tag{2.9}$$

From (2.8) and (2.9), it is easily seen that $g_n(\rho) \to 0$. $\square$

§3. Main Results

Theorem 3.1 Under Assumptions A and B, there exist a sequence $\{\widehat{\beta}_n\}$ of random variables and a random number $n_0(Y)$, such that for any $n > n_0(Y)$,

i) $\mathsf{P}(S_n(\widehat{\beta}_n) = 0) \to 1$ (asymptotic existence),

ii) $\widehat{\beta_n} \to \beta_0$ a.s. (strong consistency).

Proof By (1.4), $\lambda_n^{-1}H_n(\beta) = \lambda_n^{-1}J_n(\beta) - \lambda_n^{-1}R_n(\beta)$, implying

$$\lambda^T \frac{H_n(\beta)}{\lambda_n}\lambda \ge \frac{1}{\lambda_n}\lambda_{\min} J_n(\beta) - \frac{\|R_n(\beta)\|}{\lambda_n} \qquad \text{(a.s.)}. \tag{3.1}$$

With analogous arguments as in the proof of Lemma 2.2 it can be shown that $\|R_n(\beta)\|/\lambda_n$ is arbitrarily small for sufficiently large n in a (sufficiently small) neighborhood $N(\delta_0)$ of β_0, and therefore there exists a random number $n_1(Y)$ such that for all $\beta \in N(\delta_0)$ and $n > n_1(Y)$

$$\frac{1}{\lambda_n}\lambda^T H_n(\beta)\lambda \ge \frac{c}{2}, \tag{3.2}$$

further yielding

$$\lambda^T H_n(\beta)\lambda \ge \frac{c}{2}\cdot\lambda_n \ge \frac{c}{2}M_1 > 0, \tag{3.3}$$

where M_1 is a positive constant. So $l_n(\beta)$ is concave in $N(\delta_0)$ a.s., and therefore we only need to show that for any δ $(0<\delta<\delta_0)$, there exists a random number $n_0(Y)$ such that for all $\beta \in \partial N(\delta)$ and $n > n_0(Y)$,

$$l_n(\beta) - l_n(\beta_0) < 0 \qquad \text{(a.s.)}. \tag{3.4}$$

This means that $\widehat{\beta}_n$ which maximizes $l_n(\beta)$ must be located in $N(\delta)$. Since $0<\delta<\delta_0$ and δ is arbitrary, (i) and (ii) of the theorem can be obtained.

To prove (3.4), letting $\lambda = (\beta - \beta_0)/\delta$, then the Taylor series expansion gives

$$l_n(\beta) - l_n(\beta_0) = \delta\lambda^T S_n - \frac{1}{2}\delta^2\lambda^T H_n(\beta_n^*)\lambda, \tag{3.5}$$

where $\beta_n^* = t_n\beta_0 + (1-t_n)\beta$ for some $0 \le t_n \le 1$. Then (3.4) is equivalent to

$$\frac{1}{\lambda_n}\lambda^T S_n < \frac{1}{2\lambda_n}\delta\lambda^T H_n(\beta_n^*)\lambda \qquad \text{a.s.}, \tag{3.6}$$

for all $\lambda^T\lambda = 1$, $n > n_0$. It follows from (1.3) that the a-th component of $\lambda_n^{-1}S_n = \lambda_n^{-1} \cdot \sum\limits_{i=1}^{n} e_i(\beta_0)\partial\theta_i/\partial\beta$ is

$$\frac{s_{na}}{\lambda_n} = \frac{1}{\lambda_n}\sum_{i=1}^{n}\frac{\partial\theta_i}{\partial\beta_a}(y_i - \mu_i(\beta_0)).$$

Under Assumptions A and B, it follows from Lemma 2.1 that $\lambda_n^{-1}s_{na} \to 0$ a.s., and hence $\lambda_n^{-1}\|S_n\| \to 0$ (a.s.). By the Cauchy-Schwarz inequality, for any $\lambda^T\lambda = 1$, we have

$$|\lambda^T S_n|^2 \le (\lambda^T\lambda)S_n^T S_n = |S_n\|^2,$$

and therefore

$$\lambda_n^{-1}\lambda^T S_n \to 0 \quad \text{a.s.} \qquad \text{for any } \lambda^T\lambda = 1. \tag{3.7}$$

On the other hand, from (3.2), there is δ $(0<\delta<\delta_0)$ such that

$$\lambda_n^{-1}\lambda^T H_n(\beta_n^*)\lambda > \frac{c}{2} > 0 \qquad \text{a.s.}, \tag{3.8}$$

for any $\beta \in N(\delta)$ and $n > n_1(Y)$. By (3.7), there exists a random number $n_0(Y) > n_1(Y)$ such that

$$\lambda_n^{-1}\lambda^T S_n < \frac{1}{2}\delta\cdot\frac{c}{2} \qquad \text{a.s.}, \tag{3.9}$$

for any $\beta \in N(\delta)$ and $n > n_0(Y)$. From (3.8) and (3.9), it is easily seen that (3.6) holds, and therefore (3.4) holds. (i) and (ii) of Theorem 3.1 are proved. □

Theorem 3.2 Suppose that Assumptions A, B (i), (iii) and C hold in model (1.1), then there exists a sequence $\{\widehat{\beta}_n\}$ of MLE with

i) $\widehat{\beta}_n \to \beta_0$ a.s.,

ii) $\lambda_n^{1/2}(\widehat{\beta}_n - \beta_0) \xrightarrow{\mathcal{L}} N(0, K^{-1}(\beta_0))$.

Proof By the continuity properties of $J_n(\beta) = D(\beta)^T V(\beta) D(\beta)$ and Assumption C, for any given ε $(0 < \varepsilon < (1/2)\cdot\lambda^T K(\beta_0)\lambda)$, there exist an integer number n_0 and a neighborhood $N(\delta_0)$ of β_0 such that

$$\begin{aligned} &\Big|\frac{1}{\lambda_n}\lambda^T J_n(\beta)\lambda - \lambda^T K(\beta_0)\lambda\Big| \\ \le\ &\Big|\frac{1}{\lambda_n}\lambda^T J_n(\beta)\lambda - \lambda^T K(\beta)\lambda\Big| + |\lambda^T K(\beta)\lambda - \lambda^T K(\beta_0)\lambda| < \varepsilon, \end{aligned} \tag{3.10}$$

for any $\beta \in N(\delta_0)$, $n > n_0(Y)$, $\lambda^T\lambda = 1$. Then we get

$$\lambda_n^{-1}\lambda^T J_n(\beta)\lambda > \lambda^T K(\beta_0)\lambda - \varepsilon > \frac{1}{2}\lambda^T K(\beta_0)\lambda > 0,$$

inducing (ii) of Assumption B. And therefore, we gain the (i) of the theorem from Theorem 3.1.

To prove part (ii) of the theorem, the Taylor expansion of S_n at $\widehat{\beta}_n$ can be used, which is

$$S_n = S_n(\widehat{\beta}_n) + H_n(\beta_n^*)(\beta_n^* - \beta_0) = H_n(\beta^*)(\widehat{\beta}_n - \beta_0),$$

where $\beta_n^* = t_n^*\beta_0 + (1 - t_n^*)\widehat{\beta}_n$ for some $0 \le t_n^* \le 1$. The above equation may be rewritten as

$$J_n^{-1/2} S_n = J_n^{-1/2} H_n(\beta_n^*) J_n^{-T/2} J_n^{T/2}(\widehat{\beta}_n - \beta_0),$$

where $J_n^{1/2}$ is a square root of the positive definite matrix J_n, i.e. $J_n = J_n^{1/2} J_n^{T/2}$, $J_n^{-1} = J_n^{-T/2} J_n^{-1/2}$. Then we have

$$\lambda_n^{1/2}(\widehat{\beta}_n - \beta_0) = (\lambda_n^{-1} J_n)^{-T/2} G_n^{-1}(\beta_n^*) J_n^{-1/2} S_n, \tag{3.11}$$

where $G_n(\beta_n^*) = J_n^{-1/2} H_n(\beta_n^*) J_n^{-T/2}$. On the other hand,

$$G_n(\beta_n^*) = (\lambda_n^{-1} J_n)^{-1/2}\{\lambda_n^{-1} J_n(\beta_n^*) - \lambda_n^{-1} R_n(\beta_n^*)\}(\lambda_n^{-1} J_n)^{-T/2}. \tag{3.12}$$

Since $\beta_n^* \to \beta_0$ a.s. $(n \to \infty)$, it follows from Lemma 2.2 that

$$\lambda_n^{-1} R_n(\beta_n^*) \to 0 \qquad \text{(a.s.)}. \tag{3.13}$$

From Assumption C, we get

$$(\lambda_n^{-1} J_n)^{-T/2} \to K^{-T/2}. \tag{3.14}$$

Since $\beta_n^* \to \beta_0$ a.s. $(n \to \infty)$, by analogous arguments as given in the proof of (3.10), it can be shown that

$$\lambda_n^{-1} J_n(\beta_n^*) \to K = K^{1/2} K^{T/2}. \tag{3.15}$$

Substituting (3.13)–(3.15) into (3.10) yields $G_n(\beta_n^*) \to I_p$ a.s.. Hence, it follows from (3.11) and Lemma 2.3 that $\lambda_n^{1/2}(\widehat{\beta}_n - \beta_0) \to N(0, K^{-1}(\beta_0))$. □

Theorem 3.3 Suppose that Assumptions A, B (i), (iii) and C hold in the models (1.1), then there exists a sequence $\{\widehat{\beta}_n\}$ satisfies

$$2\{l_n(\widehat{\beta}_n) - l_n(\beta_0)\} \xrightarrow{\mathcal{L}} \mathcal{X}^2(p), \tag{3.16}$$

where $\mathcal{X}^2(p)$ denotes $\mathcal{X}^2$-distribution with a degree of freedom, p.

Proof From (3.11), it follows that

$$(\widehat{\beta}_n - \beta_0) = J_n^{-T/2} G_n^{-1}(\beta_n^*) J_n^{-1/2} S_n, \tag{3.17}$$

where $G_n(\beta_n^*) = J_n^{-1/2} H_n(\beta_n^*) J_n^{-T/2}$. The Taylor expansion of $l_n(\beta)$ at $\widehat{\beta}_n$ gives

$$l_n(\beta_0) = l_n(\widehat{\beta}_n) + S_n^T(\widehat{\beta}_n)(\beta_0 - \widehat{\beta}_n) - \frac{1}{2}(\beta_0 - \widehat{\beta}_n)^T H_n(\beta_n^*)(\beta_0 - \widehat{\beta}_n),$$

where $\beta_n^* = t_n^* \beta_0 + (1 - t_n^*)\widehat{\beta}_n$, $0 \le t_n^* \le 1$. From $S_n(\widehat{\beta}_n) = 0$, we have

$$2\{l_n(\widehat{\beta}_n) - l_n(\beta_0)\} = (\beta_0 - \widehat{\beta}_n)^T H_n(\beta_n^*)(\beta_0 - \widehat{\beta}_n). \tag{3.18}$$

Since $\widehat{\beta}_n \to \beta_0$ $(n \to \infty)$, we have $\beta_n^* \to \beta_0$ $(n \to \infty)$. Substituting (3.17) into (3.18), we get

$$2\{l_n(\widehat{\beta}_n) - l_n(\beta_0)\} = \{J_n^{-1/2} S_n\}^T [G_n^{-1}(\beta_n^*)]^T \{J_n^{-1/2} S_n\}. \tag{3.19}$$

From $G_n^{-1}(\beta_n^*) \to I_p$ a.s., the continuity theorem, Lemma 2.3 and the definition of $\mathcal{X}^2$-distribution, it follows that (3.16) holds. □

Acknowledgement The authors are thankful to the referee for valuable comments for improving the paper.

References

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

[2] Cordeiro, G.M. and Paula, G.A., Improved likelihood ratio statistics for exponential family nonlinear models, *Biometrika*, **76**(1989), 93–100.

[3] Cook, R.D. and Tsai, C.L., Diagnostics for assessing the accuracy of normal approximations in exponential family nonlinear models, *J. Amer. Statist. Assoc.*, **85**(1990), 770–777.

[4] Wei, B.C. and Shi, J.Q., On statistical models in regression diagnostics, *Ann. Inst. Statist. Math.*, **46**(1994), 267–278.

[5] Wei, B.C., *Exponential Family Nonlinear Models*, Singapore: Springer-Verlag, 1998.

[6] Wei, B.C., *Modern Nonlinear Regression Analysis* (in China), Nanjing: SE University Press, 1989.

[7] Zhu, Z.Y. & Wei, B.C., Asymptotic properties of MLE in exponential family nonlinear models, *Sys. Sci. & Math. Sci.*, **10**(1997), 193–201.

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

[9] Wu, C.F., Asymptotic properties of nonlinear least squares estimators, *Ann. Statist.*, **9**(1981), 501–513.

[10] Ferguson, T.S., *A Course in Large Sample Theory*, London: Chapman & Hall, 1996.

指数族非线性模型最大似然估计的相合性和渐近正态性

夏 天 孔繁超

(云南大学统计系, 昆明, 650091) (安徽大学数学系, 合肥, 230039)

本文我们提出了一些正则条件, 这些条件减弱了Zhu and Wei (1997)文中的条件. 基于所提的正则条件, 我们证明了指数族非线性模型参数最大似然估计的相合性和渐近正态性. 我们的结果可被认为是Zhu and Wei (1997)工作的进一步改进.

关键词: 指数族非线性模型, 相合性, 渐近正态性, 最大似然估计.

学科分类号: O212.1.