Journal of Mathematical Research & Exposition
Feb., 2008, Vol. 28, No. 1, pp. 192–198
DOI:10.3770/j.issn:1000-341X.2008.01.024
Http://jmre.dlut.edu.cn

Strong Consistency of Maximum Quasi-Likelihood Estimator in Quasi-Likelihood Nonlinear Models

XIA Tian[1], *KONG Fan-chao*[2]
(1. Department of Statistics, Yunnan University, Yunnan 650091, China;
2. Department of Mathematics, Anhui University, Anhui 230039, China)
(E-mail: xiatian718@hotmail.com)

Abstract This paper proposes some regularity conditions. On the basis of the proposed regularity conditions, we show the strong consistency of maximum quasi-likelihood estimation (MQLE) in quasi-likelihood nonlinear models (QLNM). Our results may be regarded as a further generalization of the relevant results in Ref. [4].

Keywords maximum quasi-likelihood estimator; quasi-likelihood nonlinear models; strong consistency.

Document code A
MR(2000) Subject Classification 62F12; 62J02
Chinese Library Classification O212.1

1. Introduction

Since the generalized linear models (GLM) were proposed in the early 1970s[1], the researchers have tried to extend its scope of applications. At first, the distribution of the response variable is limited to belong to an exponential family. In 1974, Wedderburn[2] advanced the concept of quasi-likelihood, which requires only the correct specification of the expectation and variance function of the response variable. On the basis of this idea, we propose quasi-likelihood nonlinear models.

Suppose that the components of the response vector $Y = (y_1, \dots, y_n)^{\mathrm{T}}$ are independent with mean vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^{\mathrm{T}}$ and covariance matrix $\sigma^2 V(\boldsymbol{\mu})$, where $\mu_i = h(\boldsymbol{x}_i, \beta)$, $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \dots, x_{iq})^{\mathrm{T}}$ $(q < n)$ is a vector of explanatory variables for the ith observation, $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ $(p < n)$ is an unknown parameter vector to be estimated, $\boldsymbol{x}_i$ and β are defined in a subset $\mathcal{X}$ of R^q and a subset $\mathcal{B}$ of R^p, respectively, σ^2 may be unknown but does not depend on β, $V(\boldsymbol{\mu}) = \mathrm{diag}\{V_1(\mu_1), \dots, V_n(\mu_n)\}$ is a positive definite matrix of known functions and $h(\cdot,\cdot)$ is a known function. Then the log quasi-likelihood is defined as

$$Q(\beta, Y) = \sum_{i=1}^{n} \int_{y_i}^{\mu_i} \frac{y_i - t}{\sigma^2 V_i(t)} \mathrm{d}t, \quad \mu_i = h(\boldsymbol{x}_i, \beta) \triangleq \mu_i(\beta). \tag{1.1}$$

Received date: 2006-02-05; **Accepted date**: 2006-07-06
Foundation item: the Natural Science Foundation of Yunnan University (No. 2005Z007C); the Scientific Research Fund of Yunnan Provincial Education Department (No. 5Y0062A); Mathematical Tianyuan Fund of National Natural Science Foundation of China (No. 10626048); Special Foundation for Middle and Young Excellent Teachers of Yunnan University.

Model (1.1) is called quasi-likelihood nonlinear model (QLNM).

It is easy to see from (1.1) that log quasi-likelihood equation is

$$U_n(\beta)=\sum_{i=1}^{n}\frac{\partial\mu_i}{\partial\beta}(V_i(\mu_i))^{-1}(y_i-\mu_i(\beta))=0,\quad \mu_i(\beta)=h(\boldsymbol{x}_i,\beta). \tag{1.2}$$

When $\mu_i(\beta)=h(\boldsymbol{x}_i^{\mathrm{T}}\beta)$, and y_i are drawn independently from a one-parameter exponential family of distributions, with density

$$\exp\{\theta_i^{\mathrm{T}}y_i-b(\theta_i)\}\mathrm{d}\gamma(y_i),\quad i=1,\dots,n,$$

(1.2) can be rewritten as

$$\sum_{i=1}^{n}\boldsymbol{x}_i\frac{\partial h(t)}{\partial t}|_{t=\boldsymbol{x}_i^{\mathrm{T}}\beta}(\ddot{b}(\theta_i))^{-1}(y_i-\mu_i(\beta)),\quad \mu_i(\beta)=h(\boldsymbol{x}_i^{\mathrm{T}}\beta). \tag{1.3}$$

Equation (1.3) is just the well-known log-likelihood equation of exponential-distribution GLM (in the case when y_i is one dimensional random variable), see Ref. [3, p.882] and Ref. [4, p.1009). Hence, quasi-likelihood nonlinear models include generalized linear models[3,4] as a special case.

In recent years, a number of authors have been concerned about the strong consistency of maximum quasi-likelihood estimator (MQLE) of these models[3,4]. For example, for $q\times 1$ responses, general link function and bounded regressors, Yue Li & Chen Xi-ru[3] have showed that, if $\lambda_{min}(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}})\geq cn^{\alpha}$ for some $\alpha\in(3/4,1]$, $\sup_{i\geq1}E||y_i||^{7/3}<\infty$, and some regular conditions are satisfied, then with probability one for large n, the quasi-likelihood equation $U_n(\beta)=0$ has a solution $\hat{\beta}_n$ with $\hat{\beta}_n-\beta_0=O(n^{-(\alpha-1/2}(\mathrm{loglog}n)^{1/2})$. Yin Chang-ming & Zhao Lin-cheng[4] further weaken the condition of eigenvalue and obtain the strong consistency of MQLE of β, if $\lambda_{min}(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}})\geq cn^{\alpha}$ for some $\alpha\in(0,1]$, $\sup_{i\geq1}E||y_i||^{r}<\infty$ for some $r>1/\alpha$ and some other regular conditions are met. However, little work has been done on the strong consistency of the maximum quasi-likelihood estimator in QLNM until recently.

In this paper, we generalize the results of the strong consistency of Yin Chang-ming & Zhao Lin-cheng[4] to quasi-likelihood nonlinear models.

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

2. Conditions and lemmas

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

To make inference for β we make the following

Assumptions

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

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

(iii) $\boldsymbol{D}(\beta)=\partial\boldsymbol{\mu}(\beta)/\partial\beta^{\mathrm{T}}=(D_1,\ldots,D_n)^{\mathrm{T}}$ is of full rank , where $\boldsymbol{\mu}(\beta)=(\mu_1,\ldots,\mu_n)^{\mathrm{T}}$;

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

(v) $\mu_i=Ey_i=h(\boldsymbol{x}_i,\beta)$ $(i=1,2,\ldots)$, and for some $\alpha\geq 1/\gamma$, $\sup_{i\geq 1}E|y_i|^{\alpha}<\infty$;

(vi) $0<\inf_{i\geq 1}V_i(\mu_i)\leq\sup_{i\geq 1}V_i(\mu_i)<\infty$.

Lemma 1 (Bernstein inequality[5]) *Suppose that $X_1,\ldots,X_n$ 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 *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 $(x-x^0)^{\mathrm{T}}F(x)\leq 0$ for some $x^0\in C$ and for all $x\in\partial C$. Then the equation $F(x)=0$ has a solution in $\overline{C}$.*

For this lemma we refer to Ortega and Rheinboldt[6], p.162–163, Corallary 6.3.4.

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 n sufficiently large,*

$$U_n(\hat\beta_n)=0\ \text{ and } \hat\beta_n\to\beta_0 \text{ a.s. } (n\to\infty). \tag{3.1}$$

Proof Take 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 suffices to show that with probability one for n sufficiently large

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

Let $\eta=\beta-\beta_0$ and $e_i=y_i-h(\boldsymbol{x}_i,\beta_0)$. From the mean value theorem and the Schwarz inequality, it follows that for $\beta\in\partial S_{\rho_n}$,

$$\begin{aligned}&(\beta-\beta_0)^{\mathrm{T}}U_n(\beta)\\&=\eta^{\mathrm{T}}\sum_{i=1}^n\frac{\partial\mu_i}{\partial\beta}(V_i(\mu_i))^{-1}e_i-\eta^{\mathrm{T}}\sum_{i=1}^n\frac{\partial\mu_i}{\partial\beta}(V_i(\mu_i))^{-1}(\frac{\partial\mu_i}{\partial\beta^{\mathrm{T}}}|_{\beta=\beta^{i*}})\eta\end{aligned}$$

$$
\begin{aligned}
&= \eta^{\mathrm{T}} \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} (V_i(\mu_i))^{-1} e_i - \eta^{\mathrm{T}} \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} (V_i(\mu_i))^{-1} \frac{\partial \mu_i}{\partial \beta^{\mathrm{T}}} \eta + \\
&\quad \eta^{\mathrm{T}} \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} (V_i(\mu_i))^{-1} [\frac{\partial \mu_i}{\partial \beta^{\mathrm{T}}} - \frac{\partial \mu_i}{\partial \beta^{\mathrm{T}}} |_{\beta=\beta^{i*}}] \eta \\
&\leq F_n(\beta) - G_n(\beta) + [G_n(\beta)]^{1/2} [\eta^{\mathrm{T}} \sum_{i=1}^{n} (\frac{\partial \mu_i}{\partial \beta} - \frac{\partial \mu_i}{\partial \beta} |_{\beta=\beta^{i*}}) (V_i(\mu_i))^{-1} (\frac{\partial \mu_i}{\partial \beta^{\mathrm{T}}} - \frac{\partial \mu_i}{\partial \beta^{\mathrm{T}}} |_{\beta=\beta^{i*}}) \eta]^{1/2},
\end{aligned}
$$

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

$$
\begin{aligned}
F_n(\beta) &= \eta^{\mathrm{T}} \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} (V_i(\mu_i))^{-1} e_i, \\
G_n(\beta) &= \eta^{\mathrm{T}} \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} (V_i(\mu_i))^{-1} \frac{\partial \mu_i}{\partial \beta^{\mathrm{T}}} \eta \geq c \eta^{\mathrm{T}} \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} \frac{\partial \mu_i}{\partial \beta^{\mathrm{T}}} \eta, \qquad (3.4)
\end{aligned}
$$

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

$$\lambda_{\min} \{ \frac{\partial \mu_i}{\partial \beta} (V_i(\mu_i))^{-1} \frac{\partial \mu_i}{\partial \beta^{\mathrm{T}}} \} \geq c > 0,$$

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

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

which implies for n sufficiently large,

$$(\beta - \beta_0)^{\mathrm{T}} U_n(\beta) \leq F_n(\beta) - H_n(\beta), \qquad (3.5)$$

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

$$\sup_{\beta \in \partial S_{\rho_n}} \{F_n(\beta) - H_n(\beta)\} < 0. \qquad (3.6)$$

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

$$
\begin{aligned}
&|E\bar{e}_i| = |E(e_i I(|e_i| \leq i^{1/t}))| \leq i^{-(\alpha-1)/t} E|e_i|^{\alpha}, \qquad (3.7) \\
&\sum_{i=1}^{\infty} P(\bar{e}_i \neq e_i) \leq \sup_{i \geq 1} E|e_i|^{\alpha} \sum_{i=1}^{\infty} i^{-\alpha/t} < \infty.
\end{aligned}
$$

And by the Borel-Cantelli Lemma, for n sufficiently large,

$$\bar{e}_n = e_n. \qquad (3.8)$$

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

$$\inf_{\beta \in \partial S_{\rho_n}} H_n(\beta) \geq c n^{\gamma - 2\delta} \geq c n^{\gamma/(1+2\varepsilon\gamma)} \to \infty. \qquad (3.9)$$

Let

$$F_n^*(\beta) = \sum_{i=1}^n \eta^{\mathrm{T}} \frac{\partial \mu_i}{\partial \beta} (V_i(\mu_i))^{-1} e_i^*, \quad \overline{F}_n(\beta) = \sum_{i=1}^n \eta^{\mathrm{T}} \frac{\partial \mu_i}{\partial \beta} (V_i(\mu_i))^{-1} \bar{e}_i. \tag{3.10}$$

By (3.8) and (3.9), to prove (3.6), we need to show that with probability one for n sufficiently large,

$$\sup_{\beta \in \partial S_{\rho_n}} \{F_n^*(\beta) - \frac{H_n(\beta)}{3}\} < 0, \tag{3.11}$$

and

$$\sup_{\beta \in \partial S_{\rho_n}} \{E\overline{F}_n(\beta) - \frac{H_n(\beta)}{3}\} < 0. \tag{3.12}$$

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 is less than n^{-2}, and $M \leq [(2n^2+1)^p]$. Take any fixed $\beta_j \in U_j$ for $j = 1, 2, \ldots, M$ and let $\eta_j = \beta_j - \beta_0, j = 1, \ldots, M$. From Assumptions (iv)–(vi), $|e_i^*| \leq 2i^{1/t}$ and from (3.2), it follows that for $j = 1, 2, \ldots, M, i = 1, 2, \ldots, n$,

$$|\eta^{\mathrm{T}} \frac{\partial \mu_i}{\partial \beta} (V_i(\mu_i))^{-1} e_i^*| \leq ||\eta^{\mathrm{T}}|| || \frac{\partial \mu_i}{\partial \beta} || |(V_i(\mu_i))^{-1}| |e_i^*| \leq c||\eta^{\mathrm{T}}|| i^{1/t} \leq c n^{1/t-\delta}, \tag{3.13}$$

$$E(e_i^*)^2 \leq \begin{cases} c, & \alpha \geq 2, \\ c i^{(2-\alpha)/t}, & 1 < \alpha < 2, \end{cases}$$

$$\mathrm{Var}(\eta^{\mathrm{T}} \frac{\partial \mu_i}{\partial \beta} (V_i(\mu_i))^{-1} e_i^*) \leq c\eta^{\mathrm{T}} \frac{\partial \mu_i}{\partial \beta} \frac{\partial \mu_i}{\partial \beta^{\mathrm{T}}} \eta (i^{(2-\alpha)/t} + c) < c\eta^{\mathrm{T}} \frac{\partial \mu_i}{\partial \beta} \frac{\partial \mu_i}{\partial \beta^{\mathrm{T}}} \eta n^{1/t-\delta}. \tag{3.14}$$

From Lemma 1, (3.5), (3.10), (3.13), (3.14), and Assumption (iv) and (3.2), it follows that

$$\begin{aligned} P\{F_n^*(\beta_j) \geq H_n(\beta_j)/4\} &\leq \exp\{-c \sum_{i=1}^n \eta_j^{\mathrm{T}} \frac{\partial \mu_i}{\partial \beta} \frac{\partial \mu_i}{\partial \beta^{\mathrm{T}}} \eta_j / n^{1/t-\delta}\} \\ &\leq \exp\{-c n^{\gamma - \delta - 1/t}\} = \exp\{-c n^{\varepsilon \gamma^2}\}, \end{aligned} \tag{3.15}$$

and therefore

$$\sum_{i=1}^{\infty} P(\bigcup_{1 \leq j \leq M} \{F_n^*(\beta_j) \geq H_n(\beta_j)/4\}) \leq \sum_{i=1}^{\infty} (2n^2+1)^p \exp\{-n^{\varepsilon \gamma^2}\} < \infty. \tag{3.16}$$

By the Borel-Cantelli lemma, for n sufficiently large, we have

$$F_n^*(\beta_j) \leq H_n(\beta_j)/4, a.s.\ j = 1, 2, \ldots, M. \tag{3.17}$$

For any given $\beta \in \partial S_{\rho_n}$, we can find $\beta_j \in U_j$ such that $||\beta - \beta_j|| \leq n^{-2}$. Note that $\eta = \beta - \beta_0$, by Assumption (iv), $||\eta|| = \rho_n$ and $||\eta - \eta_j|| \leq n^{-2}$. So we have

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

$$\leq c|(\eta-\eta_j)^{\mathrm{T}}\sum_{i=1}^{n}\frac{\partial\mu_i}{\partial\beta}\frac{\partial\mu_i}{\partial\beta^{\mathrm{T}}}\eta|+c|\eta_j^{\mathrm{T}}\sum_{i=1}^{n}\frac{\partial\mu_i}{\partial\beta}\frac{\partial\mu_i}{\partial\beta^{\mathrm{T}}}(\eta-\eta_j)|+$$
$$c|\eta_j^{\mathrm{T}}\sum_{i=1}^{n}(\frac{\partial\mu_i}{\partial\beta}\frac{\partial\mu_i}{\partial\beta^{\mathrm{T}}}-\frac{\partial\mu_i}{\partial\beta}\frac{\partial\mu_i}{\partial\beta^{\mathrm{T}}}|_{\beta=\beta_j})\eta_j|.$$

Choose arbitrarily an element of the matrix

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

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**} is on 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

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

Similarly,

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

From $||\beta_j-\beta_0||=\rho_n, ||\beta-\beta_j||\leq n^{-2}, |e_i^*|\leq 2i^{1/t}$, and Assumption (iv), we have

$$|F_n^*(\beta)-F_n^*(\beta_j)|\leq c. \tag{3.19}$$

By (3.9) and (3.17)–(3.19), for n sufficiently large we have

$$\sup_{\beta\in\partial S_{\rho_n}}\{F_n^*(\beta)-\frac{H_n(\beta)}{3}\}<0.$$

Therefore (3.11) holds.

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

$$|E\overline{F}_n(\beta)|=|\eta^{\mathrm{T}}\sum_{i=1}^{n}\frac{\partial\mu_i}{\partial\beta}(V_i(\mu_i))^{-1}E\bar{e}_i|$$
$$\leq c||\eta||\sum_{i=1}^{n}i^{-(\alpha-1)/t}\leq cn^{-\delta}(n^{1-(\alpha-1)/t}+\log n)\leq cn^{1/t-\delta}. \tag{3.20}$$

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

$$\sup_{\beta\in\partial S_{\rho_n}}\{E\overline{F}_n(\beta)-\frac{H_n(\beta)}{3}\}<0,$$

and therefore (3.12) holds.

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

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

which completes the proof of the theorem.

References

[1] NELDER J A, WEDDERBURN R W M. *Generalized linear models* [J]. J. R. Statist. Soc. A., 1972, **135**: 370–384.
[2] WEDDERBURN R W M. *Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method* [J]. Biometrika, 1974, **61**: 439–447.
[3] YUE Li, CHEN Xi-ru. *Rate of strong consistency of quasi maximum likelihood estimate in generalized linear models* [J]. Sci. China Ser. A, 2004, **47**(6): 882–893.
[4] YIN Chang-ming, ZHAO Lin-cheng. *Strong consistency of maximum quasi-likelihood estimates in generalized linear models* [J]. Sci. China Ser. A, 2005, **48**(8): 1009–1014.
[5] BENNETT G. *Probability inequality for sums of independent random variables* [J]. JASA., 1962, **57**: 33–45.
[6] ORTEGA J M, RHEINBOLDT W C. *Iterative Solution of Nonlinear Equation in Several Variables* [M]. Academic Press, New York, 1970.