Electronic Journal of Statistics
Vol. 6 (2012) 1838–1846
ISSN: 1935-7524
DOI: 10.1214/12-EJS731

Maximum likelihood estimation in logistic regression models with a diverging number of covariates

Hua Liang*

Department of Biostatistics
and Computational Biology
University of Rochester
Rochester, NY 14642, USA
e-mail: hliang@bst.rochester.edu

and

Pang Du†

Department of Statistics
Virginia Tech
Blacksburg, VA 24061, USA
e-mail: pangdu@vt.edu

Abstract: Binary data with high-dimensional covariates have become more and more common in many disciplines. In this paper we consider the maximum likelihood estimation for logistic regression models with a diverging number of covariates. Under mild conditions we establish the asymptotic normality of the maximum likelihood estimate when the number of covariates p goes to infinity with the sample size n in the order of $p = o(n)$. This remarkably improves the existing results that can only allow p growing in an order of $o(n^\alpha)$ with $\alpha \in [1/5, 1/2]$ [12, 14]. A major innovation in our proof is the use of the injective function.

AMS 2000 subject classifications: Primary 62F12; secondary 62J12.
Keywords and phrases: High dimensional, asymptotic normality, injective function, "large n, diverging p", logistic regression.

Received March 2012.

1. Introduction

High dimensional logistic regression models have attracted much attention recently as binary data with a diverging number of covariates are becoming more and more common in many disciplines. Let y be a binary response variable and $\mathbf{x} = (x_1, \ldots, x_p)^\top$ be the vector of covariate whose relationship with y can be described as

$$\text{logit}\{P(y = 1|\mathbf{x})\} = \mathbf{x}^\top \boldsymbol{\beta}, \tag{1.1}$$

*Liang's research was partially supported by NSF grants DMS-1007167 and DMS-1207444.
†Corresponding author. Du's research was supported by NSF grant DMS-1007126.

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of unknown parameters. We are interested in the high-dimensional case when p diverges with the sample size n. Hence we may use p_n when needed to emphasize the dependence of p on n.

Logistic models are standard and powerful tools to describe the relationship between a binary response variable and a set of covariates. Estimation and inference based on the maximum likelihood estimation in logistic regression have been well studied in theory and widely used in practice [6, 9–11]. Recently, logistic regression models have been applied to analyze high-dimensional data where p may diverge with n [3, 4, 7, 13]. These papers primarily focus on developing various variable selection procedures. The success of these procedures relies on certain sparsity assumption that allows only a small number of covariates to have nonzero effects. Correspondingly, the asymptotic theories in these papers are devoted to the investigation of selection property like the well-known oracle property. In summary, the contribution of these work is the successful reduction of the possibly ultra-high dimension ($p >> n$) estimation problem to a problem with much lower dimensions ($p = o(n)$).

Despite these developments, theoretical properties of the maximum likelihood estimator (MLE) in a general setting of high dimensional $\boldsymbol{\beta}$ with $p = o(n)$ are not established yet. [3] offered some insights for a special case when all the components of the true $\boldsymbol{\beta}$ are nonzero and all the components of its MLE are not too close to zero. [14] considered generalized estimating equation analysis of clustered binary data with a diverging number of covariates. Particularly, she showed that the GEE estimator is consistent when the dimension diverges in the order of $o(n^{1/2})$ and its arbitrary linear combination is asymptotically normal when the dimension diverges in the order of $o(n^{1/3})$.

Our paper aims to fill in this important gap by showing the asymptotic normality of the MLE of high-dimensional $\boldsymbol{\beta}$ under mild conditions that only require $p/n \to 0$. A critical step in our theoretical derivation is an innovative use of the injective function.

The rest of the paper is organized as follows. We present the asymptotic normality result for high-dimensional logistic regression models in Section 2, and give the proof in Section 3. We conclude the paper with some discussions in Section 4.

2. MLE with diverging dimension

Let $\mathcal{F}(v) = \{1 + \exp(-v)\}^{-1}$ be the logistic distribution function. Then the log-likelihood function is

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \log\{\mathcal{F}(\mathbf{x}_i^\top \boldsymbol{\beta})\} + (1 - y_i) \log\{1 - \mathcal{F}(\mathbf{x}_i^\top \boldsymbol{\beta})\} \right]. \tag{2.1}$$

The maximum likelihood estimator $\widehat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ is the solution to

$$L_n(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{x}_i \{ y_i - \mathcal{F}(\mathbf{x}_i^\top \boldsymbol{\beta}) \} = \mathbf{0}. \tag{2.2}$$

Let $H(v)$ be the derivative of $\mathcal{F}(v)$. For any function η, denote $\eta_i(\boldsymbol{\beta}) = \eta(\mathbf{x}_i^\top \boldsymbol{\beta})$ and $\eta_i = \eta(\mathbf{x}_i^\top \boldsymbol{\beta}_0)$, for instance, $H_i(\boldsymbol{\beta}) = H(\mathbf{x}_i^\top \boldsymbol{\beta})$, $\mathcal{F}_i = \mathcal{F}_i(\boldsymbol{\beta}_0)$. Let $G_n(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{x}_i H_i(\boldsymbol{\beta}) \mathbf{x}_i^\top$ and $S_n = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top$.

When p is fixed, [2] studied asymptotic normality of the maximum likelihood estimator $\widehat{\boldsymbol{\beta}}_n$ under mild assumptions. [15] considered the same problem under weaker assumptions. With a diverging p_n, the problem becomes much more complicated and has not been investigated yet. Two questions need to be answered. First, are there any random variables such that (2.2) holds in probability? Second, if so, are these random variables still asymptotically normal and under what conditions? Our theorem below addresses these two questions. The tool we use to prove the existence is a local inverse function theorem developed by [1], who studied strong consistency of maximum quasi-likelihood estimators in generalized linear models, and the idea once used in [15].

In what follows, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the maximum and minimum eigenvalues for a matrix A respectively, and $A_{,j}$ and $A_{j,}$ the jth column and row of matrix A. $A_1 \geq A_2$ means $A_1 - A_2$ is semi-positive definite for two matrices A_1 and A_2. C will be a generic constant with different values in different places. Let $\|\cdot\|_2$ be the standard Euclidean norm on R^n.

The following conditions are imposed to obtain Theorem 1.

Assumption.

(A1) $p_n/n \to 0$

(A2) $\max_{i,j} |x_{ij}| < \infty$ and there exist two positive constants $c_{\min}$ and $C_{\max}$ such that $c_{\min} n \leq \lambda_{\min}(S_n) \leq \lambda_{\max}(S_n) \leq C_{\max} n$.

To the best of our knowledge, condition (A1) is the weakest assumption on the order of p_n comparing to the assumptions of $p_n = o(n^{1/2})$ or $o(n^{1/3})$ in the existing literature; see, e.g., [14] and the references therein. It might be very difficult to improve the order without any further assumptions such as sparsity. To bound the covariates, [14] requires $\sup_{i,j} |x_{ij}| = O(\sqrt{p_n})$. When p_n is a constant, this bound coincides with ours in condition (A2). When p_n diverges with n, our bound is a bit more restrictive. With our bound, one can always find a positive constant $c_{00} < 1/2$ such that

$$c_{00} \leq \max_{1 \leq i \leq n} \mathcal{F}(\mathbf{x}_i^\top \boldsymbol{\beta}_0)\{1 - \mathcal{F}(\mathbf{x}_i^\top \boldsymbol{\beta}_0)\} \leq 1 - c_{00}. \tag{2.3}$$

For example, the right-hand side $1 - c_{00}$ of (2.3) can always be replaced by $3/4$. Equation (2.3) indicates that, for any p_n-vector $\mathbf{v}$,

$$c_{00} \mathbf{v}^\top S_n \mathbf{v} \leq \mathbf{v}^\top \sum_{i=1}^n \mathbf{x}_i H_i(\boldsymbol{\beta}) \mathbf{x}_i^\top \mathbf{v} \leq (1 - c_{00}) \mathbf{v}^\top S_n \mathbf{v}.$$

The rest of condition (A2) bounds the eigenvalues of S_n. This is a stability assumption to ensure S_n/n is not ill-conditioned. This assumption is needed for asymptotic investigation of $\hat{\boldsymbol{\beta}}$ even in the designs with fixed number of covariates [1, 8]. Similar conditions are required in establishing asymptotic normality of

the maximum likelihood estimation for GLM with fixed number of covariates [see, for example, 1].

Theorem 1. *Suppose Assumptions (A1)-(A2) hold. Then there exist a sequence of random variables* $\widehat{\boldsymbol{\beta}}_n$ *such that*

$$P\{L_n(\widehat{\boldsymbol{\beta}}_n) = 0\} \to 1 \tag{2.4}$$

and

$$\mathbf{u}^\top G_n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(0,1), \tag{2.5}$$

where $\mathbf{u}$ *is an unit* p_n*-vector, and* $G_n = G_n(\boldsymbol{\beta}_0)$.

The first part indicates that with probability tending to 1, there exists a solution of the equation $L_n(\boldsymbol{\beta}) = 0$, while the second part ensures that this solution is asymptotically normal.

3. Proof of the main result

3.1. Technical Lemmas

We state or prove several preliminary lemmas first. In the following, $\|\cdot\|$ always refers to the l_2-norm $\|\cdot\|_2$.

Lemma 1. *[5] If* F *is continuously differentiable in a convex interval of* $\mathbb{R}$*, then*

$$F(t_2) - F(t_1) = (t_2 - t_1)\int_0^1 \frac{dF}{ds}|_{s=t_1+u(t_2-t_1)}du,$$

where $t_1, t_2 \in \mathbb{R}$. □

Lemma 2. *[1] Let* Υ *be a smooth injection from* $\mathbb{R}^{p_n}$ *to* $\mathbb{R}^{p_n}$ *with* $\Upsilon(\mathbf{x}_0) = \mathbf{y}_0$ *and* $\inf_{\|\mathbf{x}-\mathbf{x}_0\|=\delta}\|\Upsilon(\mathbf{x}) - \mathbf{y}_0\| \geq R$*. Then for any* $\mathbf{y}$ *with* $\|\mathbf{y} - \mathbf{y}_0\| \leq R$*, there is an* $\mathbf{x}$ *with* $\|\mathbf{x} - \mathbf{x}_0\| \leq \delta$ *such that* $\Upsilon(\mathbf{x}) = \mathbf{y}$. □

Lemma 3. *Under the conditions of Theorem 1, we have*

$$\sup_{\boldsymbol{\beta}\in N_n(\delta)} |\mathbf{u}^\top G_n^{-1/2} Q_n(\boldsymbol{\beta}) G_n^{-1/2}\mathbf{u} - 1| \to 0, \tag{3.6}$$

where $Q_n(\boldsymbol{\beta}) = \partial L_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^\top$ *and* $N_n(\delta) = \{\boldsymbol{\beta} : \|G_n^{1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\| \leq \delta\}$.

Proof. Let $\varepsilon_i = y_i - \mathcal{F}(\mathbf{x}_i^\top\boldsymbol{\beta}_0)$ and $\sigma_i^2 = \mathrm{var}(\varepsilon_i)$. A direct calculation yields

$$\mathbf{u}^\top G_n^{-1/2} Q_n(\boldsymbol{\beta}) G_n^{-1/2}\mathbf{u} - 1 = A_n(\boldsymbol{\beta}) - B_n - C_n(\boldsymbol{\beta}), \tag{3.7}$$

where $A_n(\boldsymbol{\beta}) = \mathbf{u}^\top G_n^{-1/2} G_n(\boldsymbol{\beta}) G_n^{-1/2}\mathbf{u} - 1$, $B_n = \sum_{i=1}^n \mathbf{u}^\top G_n^{-1/2}\mathbf{x}_i\mathbf{x}_i^\top G_n^{-1/2}\mathbf{u}\varepsilon_i$, $C_n(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{u}^\top\ G_n^{-1/2}\mathbf{x}_i\mathbf{x}_i^\top G_n^{-1/2}\mathbf{u}\{\mathcal{F}_i - \mathcal{F}_i(\boldsymbol{\beta})\}$. We will show that each of these three terms approaches to zero on $N_n(\delta)$.

Let $\mathbf{H} = \mathrm{diag}(H_1, \ldots, H_n)$ and recall $\mathbb{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top$. Then

$$\begin{aligned}
\max_{1\le i\le n} \|G_n^{-1/2}\mathbf{x}_i\|^2 &= \max_{1\le i\le n} \mathbf{x}_i^\top G_n^{-1}\mathbf{x}_i \\
&= \max_{1\le i\le n} \mathbf{x}_i^\top (\mathbb{X}^\top \mathbf{H}\mathbb{X}^\top)^{-1}\mathbf{x}_i \\
&\le \max_{1\le i\le n} \mathbf{x}_i^\top \lambda_{\min}^{-1}(\mathbf{H})(\mathbb{X}^\top\mathbb{X})^{-1}\mathbf{x}_i \\
&\le c_{00}^{-1}c_{\min}^{-1}n^{-1} \max_{1\le i\le n} \mathbf{x}_i^\top\mathbf{x}_i \\
&= O(n^{-1}p_n), \qquad (3.8)
\end{aligned}$$

and

$$\begin{aligned}
\max_{1\le i\le n} \|\mathbf{x}_i^\top(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\| &\le \max_{1\le i\le n} \|G_n^{-1/2}\mathbf{x}_i\|\|G_n^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\| \\
&= O(n^{-1/2}p_n^{1/2}\delta). \qquad (3.9)
\end{aligned}$$

In addition,

$$\begin{aligned}
\sum_{i=1}^n \|\mathbf{u}^\top G_n^{-1/2}\mathbf{x}_i\|^2 &= \sum_{i=1}^n \mathbf{u}^\top G_n^{-1/2}\mathbf{x}_i\mathbf{x}_i^\top G_n^{-1/2}\mathbf{u} \\
&= \mathbf{u}^\top G_n^{-1/2} S_n G_n^{-1/2}\mathbf{u} \\
&= \mathbf{u}^\top (\mathbb{X}\mathbf{H}\mathbb{X}^\top)^{-1/2} S_n (\mathbb{X}\mathbf{H}\mathbb{X}^\top)^{-1/2}\mathbf{u} \\
&\le c_{00}^{-1}\mathbf{u}^\top(\mathbb{X}\mathbb{X}^\top)^{-1/2}S_n(\mathbb{X}\mathbb{X}^\top)^{-1/2}\mathbf{u} = c_{00}^{-1}. \qquad (3.10)
\end{aligned}$$

Note that

$$\begin{aligned}
|A_n(\boldsymbol{\beta})| &= \mathbf{u}^\top G_n^{-1/2}\{G_n(\boldsymbol{\beta}) - G_n(\boldsymbol{\beta}_0)\}G_n^{-1/2}\mathbf{u} \\
&= \mathbf{u}^\top G_n^{-1/2} \sum_{i=1}^n \mathbf{x}_i H_i^{(1)}(\boldsymbol{\beta}^*)\mathbf{x}_i^\top(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\mathbf{x}_i^\top G_n^{-1/2}\mathbf{u} \\
&\le \mathbf{u}^\top G_n^{-1/2} \sum_{i=1}^n \mathbf{x}_i C_2 \|\mathbf{x}_i^\top(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|\mathbf{x}_i^\top G_n^{-1/2}\mathbf{u} \\
&\le C_2 \max_{1\le i\le n} \|\mathbf{x}_i^\top(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\| \sum_{i=1}^n \|\mathbf{u}^\top G_n^{-1/2}\mathbf{x}_i\|^2.
\end{aligned}$$

Then (3.9) and (3.10) indicate that $\sup_{\boldsymbol{\beta}\in N_n(\delta)} \|A_n(\boldsymbol{\beta})\| \to 0$.

To show $|B_n| \to 0$, it suffices to show $\mathrm{var}(B_n) \to 0$ as $n \to \infty$ because $E(B_n) = 0$. The latter is true since

$$\begin{aligned}
\mathrm{var}(B_n) &= \sum_{i=1}^n (\mathbf{u}^\top G_n^{-1/2}\mathbf{x}_i\mathbf{x}_i^\top G_n^{-1/2}\mathbf{u})^2\sigma_i^2 \\
&\le \sum_{i=1}^n \|\mathbf{u}^\top G_n^{-1/2}\mathbf{x}_i\|^2 \|\mathbf{x}_i^\top G_n^{-1/2}\mathbf{u}\|^2\sigma_i^2 \\
&\le \max_{1\le i\le n} \|\mathbf{u}^\top G_n^{-1/2}\mathbf{x}_i\|^2 \sum_{i=1}^n \|\mathbf{x}_i^\top G_n^{-1/2}\mathbf{u}\|^2\sigma_i^2 \to 0. \qquad (3.11)
\end{aligned}$$

Similar to the arguments for $A_n(\boldsymbol{\beta})$, we may prove

$$\sup_{\boldsymbol{\beta}\in N_n(\delta)} \|C_n(\boldsymbol{\beta})\| \le \max_{1\le i\le n} \|\mathbf{x}_i^\top(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\| \sum_{i=1}^n \|\mathbf{u}^\top G_n^{-1/2}\mathbf{x}_i\|^2 \to 0.$$

□

3.2. *Proof of Theorem 1*

The proof consists of three steps. We establish the asymptotic normality of $L_n(\boldsymbol{\beta}_0)$ in the first step, and prove (2.4) in the second step. In the third step, we justify that $\mathbf{u}^\top G_n^{1/2}(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)$ can be approximated as a combination of $L_n(\boldsymbol{\beta}_0)$, and thus complete the proof of the theorem.

Step 1. We will show

$$\mathbf{u}^\top G_n^{-1/2} L_n(\boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(0,1). \tag{3.12}$$

Let $\xi_i = \mathbf{u}^\top G_n^{-1/2}\mathbf{x}_i\varepsilon_i$. It is easy to verify that $E(\xi_i)=0$. It now suffices to prove that (Lindeberg's condition), for any $\zeta>0$, as $n\to\infty$,

$$g_n(\zeta) = \sum_{i=1}^n E\{|\xi_i|^2 I_{(|\xi_i|>\zeta)}\} \to 0. \tag{3.13}$$

Let $a_{ni} = \mathbf{u}^\top G_n^{-1/2}\mathbf{x}_i$. Similar to (3.10), we can show that $\max_{1\le i\le n}\|a_{ni}\|^2\to 0$. Also (3.10) showed that $\sum_{i=1}^n \|a_{ni}\|^2$ is bounded. Combining these with the Cauchy-Schwartz inequality and (2.3) ensures (3.13). The central limiting theorem then yields (3.12).

Step 2. By Lemma 1, we have

$$L_n(\boldsymbol{\beta}) - L_n(\boldsymbol{\beta}_0) = -Q_n^*(\boldsymbol{\beta})(\boldsymbol{\beta}-\boldsymbol{\beta}_0), \tag{3.14}$$

where $Q_n^*(\boldsymbol{\beta}) = \int_0^1 Q_n(\boldsymbol{\beta}_0 + s(\boldsymbol{\beta}-\boldsymbol{\beta}_0))ds$. Furthermore, it follows from Lemma 3 that

$$\sup_{\boldsymbol{\beta}\in N_n(\delta)} |\mathbf{u}^\top G_n^{-1/2} Q_n^*(\boldsymbol{\beta}) G_n^{-1/2}\mathbf{u} - 1| \to 0 \tag{3.15}$$

and

$$\sup_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\in N_n(\delta)} |\mathbf{u}^\top G_n^{-1/2} Q_n^*(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2) G_n^{-1/2}\mathbf{u} - 1| \to 0, \tag{3.16}$$

where $Q_n^*(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2) = \int_0^1 Q_n(\boldsymbol{\beta}_1 + s(\boldsymbol{\beta}_2-\boldsymbol{\beta}_1))ds$.

Next, we prove that for any $\zeta>0$ there is a $\delta>0$ such that when n is large enough

$$P\{\text{there is } \widehat{\boldsymbol{\beta}}_n \in N_n(\delta) \text{ such that } L_n(\widehat{\boldsymbol{\beta}}_n)=0\} > 1-\zeta \tag{3.17}$$

Write $\partial N_n(\delta) = \{\boldsymbol{\beta} : \|G_n^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\| = \delta\}$. Note that $\|G_n^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|/\delta = 1$ for $\boldsymbol{\beta} \in \partial N_n(\delta)$. By the Cauchy-Schwartz inequality, we have that for any $\delta > 0$,

$$\inf_{\boldsymbol{\beta}\in\partial N_n(\delta)} (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\top Q_n^*(\boldsymbol{\beta})^\top G_n^{-1} Q_n^*(\boldsymbol{\beta})(\boldsymbol{\beta}-\boldsymbol{\beta}_0) \geq \inf_{\boldsymbol{\beta}\in\partial N_n(\delta)} \delta^2 \{(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\top Q_n^*(\boldsymbol{\beta})^\top (\boldsymbol{\beta}-\boldsymbol{\beta}_0)/\delta^2\}^2. \quad (3.18)$$

It follows from (3.15) that, for any $\epsilon > 0$ and $\delta > 0$, there is a $c_0 \in (0,1)$ independent of δ, such that

$$P\left\{\inf_{\|\mathbf{e}\|=1, \boldsymbol{\beta}\in\partial N_n(\delta)} \mathbf{e}^\top G_n^{-1/2} Q_n^*(\boldsymbol{\beta})^\top G_n^{-1/2}\mathbf{e} \geq c_0\right\} > 1 - \frac{\epsilon}{4}. \quad (3.19)$$

(3.14), (3.18), and (3.19) indicate that, for any $\delta > 0$ such that

$$P\left\{\inf_{\boldsymbol{\beta}\in\partial N_n(\delta)} \|\mathbf{u}^\top G_n^{-1/2}\{L_n(\boldsymbol{\beta}) - L_n(\boldsymbol{\beta}_0)\}\| \geq c_0\delta\right\} > 1 - \frac{\epsilon}{4}. \quad (3.20)$$

Taking $\delta = (4/\epsilon)^{1/2}/c_0$ and using the Markov inequality and (3.12) yield

$$P\{\|\mathbf{u}^\top G_n^{1/2} L_n(\boldsymbol{\beta}_0)\| \leq c_0\delta\} \geq 1 - E\|\mathbf{u}^\top G_n^{-1/2} L_n(\boldsymbol{\beta}_0)\|^2/(c_0\delta)^2 \geq 1 - 1/(c_0\delta)^2 = 1 - \frac{\epsilon}{4}. \quad (3.21)$$

Write $E_n = \left\{\|\mathbf{u}^\top G_n^{-1/2} L_n(\boldsymbol{\beta}_0)\| \leq \inf_{\boldsymbol{\beta}\in\partial N_n(\delta)} \|\mathbf{u}^\top G_n^{-1/2}\{L_n(\boldsymbol{\beta}) - L_n(\boldsymbol{\beta}_0)\}\|\right\}$. (3.20) and (3.21) imply that

$$P(E_n) > 1 - \frac{\epsilon}{2}. \quad (3.22)$$

Write $E_n^* = \{\det\{Q_n^*(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\} \neq 0 \text{ for all } \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in N_n(\delta)\}$. Then (3.16) indicates that

$$P(E_n^*) > 1 - \frac{\epsilon}{2}. \quad (3.23)$$

Lemma 1 indicates that the map: $\boldsymbol{\beta} \to \mathbf{u}^\top G_n^{-1/2} L_n(\boldsymbol{\beta})$ is an injection for $\boldsymbol{\beta} \in N_n(\delta)$ on the set E_n^*. Using Lemma 2 we know that, on $E_n \cap E_n^*$, there is a $\widehat{\boldsymbol{\beta}}_n$ such that

$$\widehat{\boldsymbol{\beta}}_n \in N_n(\delta) \quad \text{and } L_n(\widehat{\boldsymbol{\beta}}_n) = 0 \quad (3.24)$$

(3.17) follows from (3.22)-(3.24). Then (2.4) holds.

Step 3. (2.4) means that $L_n(\boldsymbol{\beta}_0) = Q_n^*(\widehat{\boldsymbol{\beta}}_n)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$. As a result, we know that

$$G_n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \{G_n^{-1/2} Q_n^*(\widehat{\boldsymbol{\beta}}_n) G_n^{-1/2}\}^{-1} G_n^{-1/2} L_n(\boldsymbol{\beta}_0).$$

Furthermore, we can show by using (3.15) that

$$\left[\mathbf{u}^\top \{G_n^{-1/2} Q_n^*(\widehat{\boldsymbol{\beta}}_n) G_n^{-1/2}\}^{-1} - \mathbf{u}^\top\right] \mathbf{u} \to 0.$$

Thus, we have

$$\mathbf{u}^\top G_n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \mathbf{u}^\top G_n^{-1/2} L_n(\boldsymbol{\beta}_0) + o_p(1). \tag{3.25}$$

(2.5) therefore follows from (3.25), (3.17), and (3.12). We complete the proof of Theorem 1. □

4. Discussion

In a rather general setting, we have established the asymptotic normality for maximum likelihood estimators in logistic regression models with high-dimensional covariates. We believe that the procedure can be extended to other generalized linear models and similar theoretical results may be established with straightforward derivations. One potential complication for other generalized linear models is that the response y may not be bounded as in logistic regression models. Other possible extensions are to the Cox model, robust regression, and procedures based on quasi-likelihood functions. Further effort is needed to build up similar procedure and theoretical results under these settings.

References

[1] K. Chen, I. Hu, and Z. Ying. Strong consistency of maximum quasi-likelihood estimators in generalized linear models with fixed and adaptive designs. *Ann. Statist.*, 27:1155–1163, 1999. MR1740117

[2] L. Fahrmeir and H. Kaufmann. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.*, 13:342–368, 1985. MR0773172

[3] J. Fan and J. Lv. Non-Concave Penalized Likelihood with NP-Dimensionality. *IEEE Trans. on Inform. Theory*, 57:5467–5484, 2011. MR2849368

[4] J. Fan and R. Song. Sure independence screening in generalized linear models with NP-dimensionality. *Ann. Statist.*, 38:3567–3604, 2010. MR2766861

[5] H. Heuser. *Lehrbuch der Analysis. Teil 2.* B. G. Teubner, Stuttgart, 1981. MR0618121

[6] D. W. Hosmer and S. Lemeshow. *Applied Logistic Regression.* John Wiley & Sons, New York, 1989.

[7] J. Huang, S. Ma, and C.-H. Zhang. The iterated LASSO for high-dimensional logistic regression. *Technical report*, 2009.

[8] T. L. Lai and C. Z. Wei. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.*, 10:154–166, 1982. MR0642726

[9] J. K. Lindsey. *Applying Generalized Linear Models*. Springer Texts in Statistics. Springer, 1997.

[10] P. McCullagh and J. A. Nelder. *Generalized Linear Models*, volume 37 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London, 2 edition, 1989. MR0727836

[11] J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *J. R. Stat. Soc. Ser. A Statist. Soc.*, 135:370–384, 1972.

[12] S. Portnoy. Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.*, 16:356–366, 1988. MR0924876

[13] S. A. van de Geer. High-dimensional generalized linear models and the LASSO. *Ann. Statist.*, 36:614–645, 2008. MR2396809

[14] L. Wang. GEE analysis of clustered binary data with diverging number of covariates. *Ann. Statist.*, 39:389–417, 2011. MR2797851

[15] C. Yin, L. Zhao, and C. Wei. Asymptotic normality and strong consistency of maximum quasi-likelihood estimates in generalized linear models. *Sci. China Ser. A*, 49:145–157, 2006. MR2223705