

PROPRIETY OF POSTERIOR DISTRIBUTION FOR DICHOTOMOUS QUANTAL RESPONSE MODELS

MING-HUI CHEN AND QI-MAN SHAO

(Communicated by Wei Y. Loh)

ABSTRACT. In this article, we investigate the property of posterior distribution for dichotomous quantal response models using a uniform prior distribution on the regression parameters. Sufficient and necessary conditions for the propriety of the posterior distribution with a general link function are established. In addition, the sufficient conditions for the existence of the posterior moments and the posterior moment generating function are also obtained. Finally, the relationship between the propriety of posterior distribution and the existence of the maximum likelihood estimate is examined.

1. INTRODUCTION

Dichotomous quantal response models have been used in many medical and econometric studies to examine the relation between various covariates and a dichotomous outcome measure. Zellner and Rossi [13] analyzed quantal response models from the Bayesian point of view, using diffuse and informative prior distributions. When prior information is not readily available, a uniform prior distribution or some other noninformative priors such as the Jeffreys prior for the regression coefficients is often used. However, use of the uniform prior or non-informative prior typically leads to a challenging problem, that is, whether the resulting posterior distribution is proper.

Now, we introduce some notation, which will be used throughout the rest of the paper. Let $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ denote an $n \times 1$ vector of n independent dichotomous random variables. Assume that $y_i = 1$ with probability p_i and $y_i = 0$ with probability $1 - p_i$. In the dichotomous quantal response models, it is usually assumed that

$$(1.1) \quad p_i = F(\mathbf{x}_i' \boldsymbol{\beta}),$$

where $F(\cdot)$ denotes a cumulative distribution function, $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})'$ is a $k \times 1$ vector of covariates, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ is a $k \times 1$ vector of regression coefficients.

Received by the editors March 3, 1999.

2000 *Mathematics Subject Classification.* Primary 62F15, 62E15, 62J12.

Key words and phrases. Improper prior, logit model, log-log model, probit model, maximum likelihood estimate.

Research of the first author was partially supported by the National Science Foundation under Grant No. DMS-9702172, and of the second author by the National Science Foundation under Grant No. DMS-9802451.

In (1.1), x_{i1} may be 1, which corresponds to an intercept, and F^{-1} is called a link function. Three widely utilized functional forms for $F(\cdot)$ in (1.1) are

$$(1.2) \quad p_i = F(\mathbf{x}'_i \boldsymbol{\beta}) = \Phi(\mathbf{x}'_i \boldsymbol{\beta}),$$

the probit model,

$$(1.3) \quad p_i = F(\mathbf{x}'_i \boldsymbol{\beta}) = \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})},$$

the logit model, and

$$(1.4) \quad p_i = F(\mathbf{x}'_i \boldsymbol{\beta}) = \exp[-\exp(-\mathbf{x}'_i \boldsymbol{\beta})],$$

the log-log link model. In (1.2), $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. Of course, other functional forms for $F(\cdot)$, $0 < F(\cdot) < 1$, can be employed; see, for example, Agresti [1], McCullagh and Nelder [7], and Chen and Dey [3]. Using (1.1), the likelihood function, denoted by $L(\boldsymbol{\beta}|\mathbf{y})$, is

$$(1.5) \quad L(\boldsymbol{\beta}|\mathbf{y}) = \prod_{i=1}^n [F(\mathbf{x}'_i \boldsymbol{\beta})]^{y_i} [1 - F(\mathbf{x}'_i \boldsymbol{\beta})]^{1-y_i}.$$

In the context of Bayesian analysis, it is required to specify a prior distribution for $\boldsymbol{\beta}$, say, $\pi(\boldsymbol{\beta})$. Then, the posterior distribution of $\boldsymbol{\beta}$ is

$$(1.6) \quad \pi(\boldsymbol{\beta}|\mathbf{y}) \propto L(\boldsymbol{\beta}|\mathbf{y})\pi(\boldsymbol{\beta}) = \prod_{i=1}^n [F(\mathbf{x}'_i \boldsymbol{\beta})]^{y_i} [1 - F(\mathbf{x}'_i \boldsymbol{\beta})]^{1-y_i} \pi(\boldsymbol{\beta}).$$

Since an improper posterior makes Bayesian inference impossible, it is important to study the propriety of the posterior distribution. An investigation of the posterior propriety can avoid a poor experimental design, which may result in the parameters of interest not being identifiable without using a suitable proper prior. It is clear that the posterior $\pi(\boldsymbol{\beta}|\mathbf{y})$ is proper if and only if

$$(1.7) \quad \int_{R^k} \prod_{i=1}^n [F(\mathbf{x}'_i \boldsymbol{\beta})]^{y_i} [1 - F(\mathbf{x}'_i \boldsymbol{\beta})]^{1-y_i} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta} < \infty.$$

In this article, we are interested in the following problem. If a uniform prior distribution for $\boldsymbol{\beta}$, i.e., $\pi(\boldsymbol{\beta}) \propto 1$, is specified, what are the sufficient and necessary conditions for the propriety of $\pi(\boldsymbol{\beta}|\mathbf{y})$?

Zellner and Rossi [13] obtained some sufficient conditions for (1.7) when F^{-1} is logit, $x_{i1} = 1$, and $k = 2$. Without loss of generality, assume that $y_i = 0$ for $i = 1, 2, \dots, m$, and $x_{i2} > 0$ for $i = 1, 2, \dots, n$. Let the $x_{(i)2}$'s be the ordered x_{i2} 's, that is, $0 < x_{(1)2} < x_{(2)2} < \dots < x_{(n)2}$. Then, their sufficient conditions are (a) $m > 0$, (b) $n > m$, (c) $\sum_{i=1}^m x_{i2} > \sum_{i=1}^{m+1} x_{(i)2}$, and (d) $\sum_{i=1}^m x_{i2} < \sum_{i=0}^{m-2} x_{(n-i)2}$. These conditions, thereafter, are referred to as Zellner-Rossi sufficient conditions.

Other related works are Ibrahim and Laud [6] and Natarajan and McCulloch [8]. Ibrahim and Laud [6] established a sufficient condition for the propriety of the posterior distribution using the Jeffreys's prior for the generalized linear model, while Natarajan and McCulloch [6] provided necessary and sufficient conditions for the existence of the posterior distribution of the variance components in a class of mixed models for binomial responses. Their problems are different from ours.

In Section 2, we aim to establish precise necessary and sufficient conditions for the propriety (1.7) for a general link function F^{-1} . It is interesting to note that those conditions are almost F free. The remainder of this article is organized as

follows. The main results are presented in Section 2. In Section 3, we examine the relationship between the propriety of the posterior distribution and the existence of the maximum likelihood estimates. The proofs of all theorems are given in Section 4.

2. MAIN RESULTS

To obtain sufficient and necessary conditions for the propriety of the posterior distribution, let $z_i = 1$ if $y_i = 0$ and -1 if $y_i = 1$, \mathbf{X} be the $n \times k$ known design matrix, with rows \mathbf{x}'_i and define \mathbf{X}^* as the matrix with rows $z_i \mathbf{x}'_i$.

The main results are given as follows.

Theorem 2.1. *Assume the following conditions are satisfied:*

- (C1) *The design matrix \mathbf{X} is of full rank;*
- (C2) *There exists a **positive** vector $\mathbf{a} = (a_1, \dots, a_n)'$ $\in R^n$, i.e., each component $a_i > 0$, such that*

$$(2.1) \quad \mathbf{X}^{*'} \mathbf{a} = 0;$$

$$(C3) \quad \int_{-\infty}^{\infty} |u|^k dF(u) < \infty.$$

Then

$$(2.2) \quad \int_{R^k} L(\boldsymbol{\beta}|\mathbf{y}) d\boldsymbol{\beta} < \infty.$$

Theorem 2.2. *If $0 < F(0-) \leq F(0) < 1$, then (C1) and (C2) are necessary conditions for (2.2).*

The next result extends Theorem 2.1 to the existence of posterior moments and the posterior moment generating function.

Theorem 2.3. *Assume that (C1) and (C2) are satisfied. If $\int_{-\infty}^{\infty} |u|^{k+p} dF(u) < \infty$ for some $p \geq 0$, then*

$$(2.3) \quad \int_{R^k} \|\boldsymbol{\beta}\|^p L(\boldsymbol{\beta}|\mathbf{y}) d\boldsymbol{\beta} < \infty.$$

If $\int_{-\infty}^{\infty} e^{t_0|u|} dF(u) < \infty$ for some $t_0 > 0$, then

$$(2.4) \quad \int_{R^k} e^{\delta_0 \|\boldsymbol{\beta}\|} L(\boldsymbol{\beta}|\mathbf{y}) d\boldsymbol{\beta} < \infty,$$

for some $\delta_0 > 0$. Furthermore, if $\int_{-\infty}^{\infty} e^{t_0|u|^2} dF(u) < \infty$ for some $t_0 > 0$, then

$$(2.5) \quad \int_{R^k} e^{\delta_0 \|\boldsymbol{\beta}\|^2} L(\boldsymbol{\beta}|\mathbf{y}) d\boldsymbol{\beta} < \infty,$$

for some $\delta_0 > 0$, where $\|\cdot\|$ denotes the Euclidean norm.

For the three widely used probit, logit and log-log link models, it is easy to see that $0 < F(0-) = F(0) < 1$ and that $\int_{-\infty}^{\infty} e^{t_0|u|} dF(u) < \infty$ for some $t_0 > 0$. Hence, (C1) and (C2) are necessary and sufficient conditions for (2.2), (2.3) and (2.4).

The moment condition (C3) may not be satisfied for some other links, say the Cauchy link or t-link with less than k degrees of freedom. In general, however, it cannot be weakened.

Theorem 2.4. *Let $x_{i,1} = 1$ for $i = 1, 2, \dots, n$ and $0 < F(0) < 1$.*

- (i) *If $y_1 = 0$ and $y_i = 1$ for $i = 2, \dots, n$, then a necessary condition for (2.2) is $\int_0^\infty u^k dF(u) < \infty$.*
- (ii) *If $y_1 = 1$ and $y_i = 0$ for $i = 2, \dots, n$, then a necessary condition for (2.2) is $\int_{-\infty}^0 |u|^k dF(u) < \infty$.*

On the other hand, when the design matrix \mathbf{X}^* has some nice structure, it is possible to weaken the moment condition.

Theorem 2.5. *Define $\mathbf{X}_{l,m}^* = (z_j \mathbf{x}'_j, l < j \leq m)$ as the $(m - l) \times k$ matrix with rows $z_j \mathbf{x}'_j$, $l < j \leq m$. Assume that there exist $p \geq 1$, $0 = m_0 < m_1 < m_2 < \dots < m_p \leq n$ and positive vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p$ such that $\mathbf{X}_{m_{l-1}, m_l}^*$ is of full rank and*

$$(2.6) \quad \mathbf{X}_{m_{l-1}, m_l}^{*'} \mathbf{a}_l = \mathbf{0} \quad \text{for } l = 1, 2, \dots, p.$$

If

$$(C4) \quad \int_{-\infty}^\infty |u|^{k/p} dF(u) < \infty$$

is satisfied, then (2.2) holds.

The results in Theorems 2.4 and 2.5 suggest that the necessity of the moment condition depends on the structure of the design matrix and response variable \mathbf{y} . For some cases, a finite k^{th} moment of F is necessary, while for some other cases, even a finite first moment is sufficient.

It should be mentioned that verifying condition (C2) is a standard linear programming problem which can be done using commercially available software, for example CPLEX (CPLEX Optimization, 1992).

Remark 2.1. When $k = 2$, and $x_{i1} \neq 0$ for every $1 \leq i \leq n$, then (C1) and (C2) are equivalent to (C1) and

$$(2.7) \quad \begin{aligned} \min_{j: z_j x_{j1} < 0} (x_{j2}/x_{j1}) &< \max_{i: z_i x_{i1} > 0} (x_{i2}/x_{i1}) \quad \text{and} \\ \max_{j: z_j x_{j1} < 0} (x_{j2}/x_{j1}) &> \min_{i: z_i x_{i1} > 0} (x_{i2}/x_{i1}). \end{aligned}$$

Remark 2.2. The following example shows that if $F(0-) = 0$ or $F(0) = 1$, then (C2) is not a necessary condition for (2.2). Let $n = 5$, $k = 2$, $y_1 = y_2 = y_3 = 1$, $y_4 = y_5 = 0$, $\mathbf{x}_1 = \mathbf{x}_3 = (1, -1)'$, $\mathbf{x}_2 = \mathbf{x}_4 = (-1, 1)'$ and $\mathbf{x}_5 = (1, 1)'$. If $F(0-) = 0$ or $F(0) = 1$, then

$$\begin{aligned} L(\boldsymbol{\beta}|\mathbf{y}) &= F(\beta_1 - \beta_2)F(-\beta_1 + \beta_2)[1 - F(\beta_1 - \beta_2)][1 - F(-\beta_1 + \beta_2)]F(\beta_1 + \beta_2) \\ &= 0 \end{aligned}$$

for all $\beta_1 \neq \beta_2$ and hence

$$\int_{R^2} L(\boldsymbol{\beta}|\mathbf{y}) d\boldsymbol{\beta} = 0.$$

On the other hand, it is easy to see from Remark 2.1 that (C2) is not satisfied.

Remark 2.3. For $k = 2$ and $x_{i1} = 1$ for $i = 1, 2, \dots, n$, the Zellner-Rossi sufficient conditions are not necessary conditions for the propriety of the posterior distribution. First, we note that if Zellner-Rossi sufficient conditions hold, then conditions (C1) and (C2) automatically hold. Second, we use a simple illustration to show that conditions (C1) and (C2) do not guarantee that the Zellner-Rossi sufficient conditions are satisfied. Let $n = 3$, $x_{12} = 2$, $x_{22} = 1$, $x_{32} = 3$, $y_1 = 0$, $y_2 = y_3 = 1$. Then $m = 1$, $x_{(1)2} = 1$, $x_{(2)2} = 2$, and $x_{(3)2} = 3$. Thus, Zellner-Rossi sufficient condition (c) does not hold. However, it is easy to verify that our sufficient and necessary conditions (see (2.7) for (C2)) still hold. Therefore, the resulting posterior is proper. But, this result cannot be observed via the Zellner-Rossi conditions.

3. EXISTENCE OF THE MAXIMUM LIKELIHOOD ESTIMATE

In this section, we aim to examine the relationship between the propriety of the posterior distribution and the existence of the maximum likelihood estimate (MLE). The existence of the MLE was discussed by Haberman [5], Wedderburn [12] and Silvapulle [11] for the dichotomous quantal response models, Albert and Anderson [2] and Santner and Duffy [10] for multinomial logistic regression models. It was proved by Silvapulle [11] that if $-\log F$ and $\log(1 - F)$ are convex and the design matrix \mathbf{X} is of full rank (i.e., (C1) holds), then the MLE $\hat{\boldsymbol{\beta}}$ exists and the minimum set $\{\hat{\boldsymbol{\beta}}\}$ is bounded if and only if

$$(S1) \quad D_I \cap D_J \neq \emptyset \text{ or one of } D_I, D_J \text{ is } R^k, \text{ where } D_I = \{\sum_{i \in I} a_i \mathbf{x}_i : a_i > 0\}, \\ D_J = \{\sum_{j \in J} a_j \mathbf{x}_j : a_j > 0\}, I = \{i : y_i = 1\} \text{ and } J = \{j : y_j = 0\}.$$

However, some of Silvapulle's conditions such as the convexity on $-\log F$ or $\log(1 - F)$ are not necessary for the existence of the maximum likelihood estimate. A more general result is given in the following theorem, which confirms that (C1) and (C2) are sufficient for the existence of the maximum likelihood estimate for the general link F^{-1} .

Theorem 3.1. *Assume that F is continuous and that (C1) and (C2) are satisfied; then the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ exists. Furthermore, if $0 < F(x) < 1$ for every $x \in R^1$ and (C1) is satisfied, then (C2) is a necessary condition for the existence of the maximum likelihood estimate.*

Remark 3.1. The example below shows that (C1) is not a necessary condition for the existence of MLE. Let $\mathbf{x}_1 = \mathbf{x}_2 = (1, 1)'$ and $\mathbf{x}_3 = (-1, -1)$, $y_1 = y_2 = y_3 = 1$. Then

$$\sup_{\boldsymbol{\beta}} L(\boldsymbol{\beta}|\mathbf{y}) = \sup_{\boldsymbol{\beta}} F(\beta_1 + \beta_2)^2 F(-\beta_1 - \beta_2) = \sup_{\beta_1} L(\beta_1, 0|\mathbf{y}).$$

Clearly,

$$\lim_{|\beta_1| \rightarrow \infty} L(\beta_1, 0|\mathbf{y}) = 0.$$

Therefore, there exists $\hat{\beta}_1$ such that

$$\sup_{\beta_1} L(\beta_1, 0|\mathbf{y}) = L(\hat{\beta}_1, 0|\mathbf{y}).$$

This means that the MLE of $\boldsymbol{\beta}$ exists; however, it is easy to see that (C1) is not satisfied.

4. PROOFS OF THE THEOREMS

The following lemma plays a key role in the proofs of our theorems.

Lemma 4.1. *Assume that (C1) and (C2) are satisfied. Then there exists a constant K depending only on \mathbf{X}^* such that*

$$(4.1) \quad \|\boldsymbol{\beta}\| \leq K\|\mathbf{u}\|$$

whenever

$$(4.2) \quad \mathbf{X}^*\boldsymbol{\beta} \leq \mathbf{u}.$$

Proof. Let $\mathcal{E} = \{(\varepsilon_1, \dots, \varepsilon_k)' \in R^k : \varepsilon_i = \pm 1\}$. Since \mathbf{X}^* is of full rank, for every $\varepsilon \in \mathcal{E}$, there is a $\mathbf{b}_\varepsilon \in R^n$ such that

$$(4.3) \quad \mathbf{b}'_\varepsilon \mathbf{X}^* = \varepsilon'.$$

Let $\mathbf{a} = (a_1, \dots, a_n)' \in R^n$ be a positive vector satisfying (2.1). Put

$$\delta = \frac{\min_{1 \leq i \leq n} (a_i)}{2 \max_{\varepsilon \in \mathcal{E}} \|\mathbf{b}_\varepsilon\|}.$$

For $\varepsilon = \varepsilon_\beta = \text{sign}(\boldsymbol{\beta}') = (\text{sign}(\beta_1), \dots, \text{sign}(\beta_k))'$, we have $\delta > 0$ and $\mathbf{a} + \delta\mathbf{b}_\varepsilon > 0$. Hence, it follows from (4.2) that

$$\begin{aligned} (\mathbf{a} + \delta\mathbf{b}_\varepsilon)' \mathbf{u} &\geq (\mathbf{a} + \delta\mathbf{b}_\varepsilon)' \mathbf{X}^* \boldsymbol{\beta} \\ &= \delta \mathbf{b}'_\varepsilon \mathbf{X}^* \boldsymbol{\beta} = \delta \text{sign}(\boldsymbol{\beta}') \boldsymbol{\beta} \geq (\delta/k) \|\boldsymbol{\beta}\|, \end{aligned}$$

as desired. \square

Proof of Theorem 2.1. Observing that

$$1 - F(x) = \int_{-\infty}^{\infty} 1\{u > x\} dF(u)$$

and

$$F(x) = \int_{-\infty}^{\infty} 1\{u \leq x\} dF(u) = \int_{-\infty}^{\infty} 1\{u \geq -x\} d(-F(-u)),$$

we have

$$[F(\mathbf{x}'_i \boldsymbol{\beta})]^{y_i} [1 - F(\mathbf{x}'_i \boldsymbol{\beta})]^{1-y_i} \leq \int_{-\infty}^{\infty} 1\{u_i \geq z_i \mathbf{x}'_i \boldsymbol{\beta}\} d(z_i F(z_i u_i))$$

and

$$[F(\mathbf{x}'_i \boldsymbol{\beta})]^{y_i} [1 - F(\mathbf{x}'_i \boldsymbol{\beta})]^{1-y_i} \geq \int_{-\infty}^{\infty} 1\{u_i > z_i \mathbf{x}'_i \boldsymbol{\beta}\} d(z_i F(z_i u_i)).$$

By the Fubini theorem, we get

$$(4.4) \quad \begin{aligned} \int_{R^k} L(\boldsymbol{\beta}|\mathbf{y}) d\boldsymbol{\beta} &= \int_{R^n} \int_{R^k} 1\{u_i > z_i \mathbf{x}'_i \boldsymbol{\beta}, 1 \leq i \leq n\} d\boldsymbol{\beta} d\mathbf{F}(\mathbf{u}) \\ &= \int_{R^n} \int_{R^k} 1\{\mathbf{X}^* \boldsymbol{\beta} \leq \mathbf{u}\} d\boldsymbol{\beta} d\mathbf{F}(\mathbf{u}), \end{aligned}$$

where $d\mathbf{F}(\mathbf{u})$ stands for $d(z_1 F(z_1 u_1)) \cdots d(z_k F(z_k u_k))$.

It follows from Lemma 4.1 that

$$\int_{R^k} 1\{\mathbf{X}^* \boldsymbol{\beta} \leq \mathbf{u}\} d\boldsymbol{\beta} \leq K\|\mathbf{u}\|^k,$$

which yields (2.2) immediately, by (C3). \square

Proof of Theorem 2.2. Since $0 < F(0-) \leq F(0) < 1$, there exists a $\delta > 0$ such that

$$(4.5) \quad 0 < F(-\delta) \leq F(\delta) < 1.$$

If \mathbf{X} is not of full rank, then \mathbf{X}^* is not of full rank either. In this case, it is easy to see that the posterior distribution of $\boldsymbol{\beta}$ is not proper. Now assume that (C2) is not satisfied. Let

$$\mathcal{D} = \{\mathbf{X}^{*'} \mathbf{a} : \mathbf{a} > 0, \mathbf{a} \in R^n\}.$$

Then \mathcal{D} is a convex cone in R^k (see Section 2 of Rockafellar [9]). Since (C2) does not hold, by Corollary 11.7.3 of [9], there exists some non-zero vector $\mathbf{b} = (b_1, \dots, b_k)' \in R^k$ such that

$$\forall \mathbf{a} > 0, \mathbf{b}' \mathbf{X}^{*'} \mathbf{a} \leq 0$$

and hence

$$\forall \mathbf{a} \geq 0, \mathbf{b}' \mathbf{X}^{*'} \mathbf{a} \leq 0.$$

In particular, we have

$$(4.6) \quad z_i \mathbf{b}' \mathbf{x}_i \leq 0 \quad \text{for } i = 1, 2, \dots, n.$$

Without loss of generality, assume that $b_1 \neq 0$. Put $I = \{i : y_i = 0\}$ and $J = \{j : y_j = 1\}$. Letting $\boldsymbol{\beta} = s_1 \mathbf{b} + (0, s_2, \dots, s_k)'$ yields

$$\begin{aligned} \int_{R^k} L(\boldsymbol{\beta}|\mathbf{y}) d\boldsymbol{\beta} &= \int_{R^k} \prod_{j \in J} F(\mathbf{x}'_j \boldsymbol{\beta}) \prod_{i \in I} (1 - F(\mathbf{x}'_i \boldsymbol{\beta})) d\boldsymbol{\beta} \\ &= |b_1| \int_{R^k} \prod_{j \in J} F(s_1 \mathbf{x}'_j \mathbf{b} + \mathbf{x}'_j(0, s_2, \dots, s_k)) \\ &\quad \times \prod_{i \in I} (1 - F(s_1 \mathbf{x}'_i \mathbf{b} + \mathbf{x}'_i(0, s_2, \dots, s_k))) ds \\ &\geq |b_1| \int_{s_1 \geq 0, |s_i| \leq \eta, 2 \leq i \leq k} \prod_{j \in J} F(-k \|\mathbf{x}_j\| \eta) \\ &\quad \times \prod_{i \in I} (1 - F(k \|\mathbf{x}_j\| \eta)) ds \quad [\text{by (4.6)}] \\ &\geq |b_1| F(-\delta)^n (1 - F(\delta))^n \int_{s_1 \geq 0, |s_i| \leq \eta, 2 \leq i \leq k} ds \\ &= \infty \end{aligned}$$

provided $\eta > 0$ is chosen such that $\eta \max_{1 \leq i \leq n} \|\mathbf{x}_i\| \leq \delta/k$. □

Proof of Theorem 2.3. The proof follows along the same lines as that of Theorem 2.1. □

Proof of Theorem 2.4. Let $M = k \max_{1 \leq i \leq n} \sum_{l=1}^k |x_{il}|$. By (4.4), we have

$$\begin{aligned}
\int_{R^k} L(\boldsymbol{\beta}|\mathbf{y})d\boldsymbol{\beta} &= \int_{R^n} \int_{R^k} 1\{u_1 > \beta_1 + \sum_{l=2}^k x_{1l}\beta_l, -u_i < \beta_1 \\
&\quad + \sum_{l=2}^k x_{il}\beta_l, 2 \leq i \leq n\}d\boldsymbol{\beta}dF(\mathbf{u}) \\
&\geq \int_{R^{+n}} \int_{R^k} 1\{u_1 - \sum_{l=2}^k x_{1l}\beta_l > \beta_1 > -\min_{2 \leq i \leq n} \sum_{l=2}^k x_{il}\beta_l\}d\boldsymbol{\beta}dF(\mathbf{u}) \\
&\geq \int_{R^{+n}} \int_{0 < \beta_l < u_1/(2M)} d\boldsymbol{\beta}dF(\mathbf{u}) \\
&\geq \int_{R^{+n}} (u_1/(2M))^k dF(\mathbf{u}) \\
&\geq F(0)^{k-1} \int_0^\infty u^k dF(u).
\end{aligned}$$

This proves (i). Similarly, we have (ii). \square

Proof of Theorem 2.5. Let $\{\xi_i, 1 \leq i \leq n\}$ be independent random variables with the same distribution function F . From Lemma 4.1 it follows that there exists a constant K such that

$$(4.7) \quad \|\boldsymbol{\beta}\| \leq K^{1/k} \min_{1 \leq l \leq p} \max_{m_{l-1} < i \leq m_l} |u_i|$$

whenever $\mathbf{X}^* \boldsymbol{\beta} \leq \mathbf{u}$. Hence, in view of the proof of Theorem 2.1,

$$\begin{aligned}
\int_{R^k} L(\boldsymbol{\beta}|\mathbf{y})d\boldsymbol{\beta} &\leq K \int_{R^k} \min_{1 \leq l \leq p} \left(\max_{m_{l-1} < i \leq m_l} |u_i| \right)^k dF(\mathbf{u}) \\
&= K E \left(\min_{1 \leq l \leq p} \max_{m_{l-1} < i \leq m_l} |\xi_i|^k \right) \\
&\leq K E \left\{ \prod_{l=1}^p \left(\max_{m_{l-1} < i \leq m_l} |\xi_i|^k \right)^{1/p} \right\} \\
&= K \prod_{l=1}^p E \left(\max_{m_{l-1} < i \leq m_l} |\xi_i|^{k/p} \right) \\
&\leq K \prod_{l=1}^p \left(\sum_{m_{l-1} < i \leq m_l} E |\xi_i|^{k/p} \right) \\
&< \infty,
\end{aligned}$$

as desired. \square

Proof of Theorem 3.1. Sufficiency: When $L(\boldsymbol{\beta}|\mathbf{y}) \equiv 0$, the existence of the MLE of $\boldsymbol{\beta}$ is obvious. If $L(\boldsymbol{\beta}|\mathbf{y}) \not\equiv 0$, then there exists a $\boldsymbol{\beta}_0$ such that $L(\boldsymbol{\beta}_0|\mathbf{y}) > 0$. Let $M \geq 1$ such that

$$F(-M) + 1 - F(M) < L(\boldsymbol{\beta}_0|\mathbf{y}).$$

For β satisfying $\max_{1 \leq i \leq n} z_i \mathbf{x}'_i \beta > M$, i.e., $z_{i_0} \mathbf{x}'_{i_0} \beta > M$ for some i_0 , we have

$$\begin{aligned} L(\beta | \mathbf{y}) &\leq [F(\mathbf{x}'_{i_0} \beta)]^{y_{i_0}} [1 - F(\mathbf{x}'_{i_0} \beta)]^{1-y_{i_0}} \\ &= \begin{cases} F(\mathbf{x}'_{i_0} \beta) & \text{if } z_{i_0} = -1, \\ 1 - F(\mathbf{x}'_{i_0} \beta) & \text{if } z_{i_0} = 1, \end{cases} \\ &\leq \begin{cases} F(-M) & \text{if } z_{i_0} = -1, \mathbf{x}'_{i_0} \beta \leq -M, \\ 1 - F(M) & \text{if } z_{i_0} = 1, \mathbf{x}'_{i_0} \beta \geq M, \end{cases} \\ &\leq F(-M) + 1 - F(M) \text{ if } z_{i_0} \mathbf{x}'_{i_0} \beta \geq M, \\ &< L(\beta_0 | \mathbf{y}). \end{aligned}$$

When $\max_{1 \leq i \leq n} z_i \mathbf{x}'_i \beta \leq M$, by Lemma 4.1

$$\|\beta\| \leq D \text{ for some } 0 < D < \infty.$$

Therefore

$$\sup_{\beta} L(\beta | \mathbf{y}) = \sup_{\beta: \|\beta\| \leq D} L(\beta | \mathbf{y}).$$

Since $L(\beta | \mathbf{y})$ is a continuous function of β , there exists $\hat{\beta}$ such that

$$L(\hat{\beta} | \mathbf{y}) = \sup_{\beta: \|\beta\| \leq D} L(\beta | \mathbf{y})$$

and hence the MLE exists.

Necessity: Assume that the MLE of β exists, that is, there is a β^* such that

$$(4.8) \quad L(\beta^* | \mathbf{y}) = \sup_{\beta} L(\beta | \mathbf{y}).$$

Assume that (C2) is not satisfied. In view of the proof of Theorem 2.2, there exists a non-zero vector b such that (4.6) holds. Letting $I = \{i : y_i = 0\}$ and $J = \{j : y_j = 1\}$ yields

$$\begin{aligned} y &= \prod_{j \in J} F(\mathbf{x}'_j \beta^* + s \mathbf{x}'_j b) \prod_{i \in I} [1 - F(\mathbf{x}'_i \beta^* + s \mathbf{x}'_i b)] \\ &= \prod_{j \in J} F(\mathbf{x}'_j \beta^* - s z_j \mathbf{x}'_j b) \prod_{i \in I} [1 - F(\mathbf{x}'_i \beta^* + s z_i \mathbf{x}'_i b)]. \end{aligned}$$

Hence by (4.6), $L(\beta^* + s b)$ is an increasing function of s . This is a contradiction. This shows that (C2) is necessary for the existence of the MLE of β . \square

REFERENCES

- [1] Agresti, A. (1990), *Categorical Data Analysis*, New York: Wiley. MR **91k**:62055
- [2] Albert, A. and Anderson, J.A. (1984), On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. *Biometrika* **71**, 1-10. MR **85g**:62108
- [3] Chen, M.-H. and Dey, D.K. (1998). Bayesian Modeling of Correlated Binary Responses via Scale Mixture of Multivariate Normal Link Functions. *Sankhyā, Series A* **60**, 322-343.
- [4] CPLEX Optimization (1992), CPLEX, Incline Village, Nevada: CPLEX Optimization Inc.
- [5] Haberman, S.J. (1977), Maximum Likelihood Estimates in Exponential Response Models, *The Annals of Statistics*, **5**, 815-841. MR **58**:18861
- [6] Ibrahim, J.G. and Laud, P.W. (1991), On Bayesian Analysis of Generalized Linear Models Using Jeffreys's Prior, *Journal of the American Statistical Association*, **86**, 981-986. MR **93d**:62045
- [7] McCullagh, P. and Nelder, J.A. FRS (1989), *Generalized Linear Models*, Second Edition. London: Chapman & Hall. MR **85k**:62161 (first edition, 1983)

- [8] Natarajan, R. and McCulloch, C.E. (1995), A Note on the Existence of the Posterior Distribution for a Class of Mixed Models for Binomial responses, *Biometrika*, **82**, 639-643. MR **96h**:62039
- [9] Rockafellar, R. T. (1970), *Convex Analysis*. Princeton, N.J.: Princeton University Press. MR **43**:445
- [10] Santner, T.J. and Duffy, D.E. (1986), A note on A. Albert and J.A. Anderson's Conditions for the Existence of Maximum Likelihood Estimates in Logistic Regression Models. *Biometrika* **73**, 755-758. MR **88k**:62110
- [11] Silvapulle, M. J. (1981), On the Existence of Maximum Likelihood Estimators for the Binomial Response Models, *J. R. Statist. Soc., Ser. B* **43**, 310-313. MR **82k**:62138
- [12] Wedderburn, R.W.M. (1976), On the Existence and Uniqueness of the Maximum Likelihood Estimates for Certain Generalized Linear Models, *Biometrika*, **63**, 27-32. MR **53**:11859
- [13] Zellner, A., and Rossi, P. E. (1984), Bayesian Analysis of Dichotomous Quantal Response Models, *Journal of Econometrics*, **25**, 365-393. MR **85f**:62139

DEPARTMENT OF MATHEMATICAL SCIENCES, WORCESTER POLYTECHNIC INSTITUTE, 100 INSTITUTE ROAD, WORCESTER, MASSACHUSETTS 01609-2280

E-mail address: mhchen@wpi.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF OREGON, EUGENE, OREGON 97403

E-mail address: shao@math.uoregon.edu