Ann. Inst. Statist. Math.
Vol. 51, No. 4, 637–656 (1999)

EXISTENCE OF BAYESIAN ESTIMATES FOR THE POLYCHOTOMOUS QUANTAL RESPONSE MODELS

MING-HUI CHEN[1]* AND QI-MAN SHAO[2]**

[1] *Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280, U.S.A.*
[2] *Department of Mathematics, University of Oregon, Eugene, OR 97403-1222, U.S.A.*

(Received November 17, 1997; revised June 15, 1998)

Abstract. This paper investigates the existence of Bayesian estimates for polychotomous quantal response models using a uniform improper prior distribution on the regression parameters. Necessary and sufficient 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 obtained. It is also found that the propriety guarantees the existence of the maximum likelihood estimate.

Key words and phrases: Improper prior, logit model, log-log model, posterior distribution, probit model.

1. Introduction

Polychotomous quantal response models have been used in many medical and econometric studies to examine the relationship between various covariates and a polychotomous outcome measure. When prior information is not readily available, a uniform prior distribution or some other non-informative priors such as the Jeffreys prior for the regression coefficients is often used. Use of the uniform or non-informative improper priors typically leads to a challenging problem, that is, whether the resulting posterior distribution is proper. However, the posterior densities for the regression coefficients are always proper when proper non-informative priors such as "locally uniform" priors used by Box and Tiao ((1992), p. 23) are employed. While Box and Tiao introduce locally uniform proper priors, in practice they integrate from $-\infty$ to ∞ and thus certain integrals that would converge given a finite range of integration may not converge given a doubly or singly infinite

* Dr. Chen's research was partially supported by NSF grant No. DMS-9702172 and NIH grant 74015-01.

** Dr. Shao's research was partially supported by NSF grant No. DMS-9802451.

range of integration. A similar point can be made with respect to Zellner's maximal data information priors (Zellner (1997)) that in some cases are proper given a finite range of integration but not with an infinite range of integration. Also, the analysis in Ibrahim and Laud (1991) indicates that Jeffreys's "non-informative" prior is proper for certain generalized linear models and is not uniform. On the other hand, Jaynes' maxent prior (see, for example, Jaynes (1982)) given just that it be proper is a proper uniform prior defined over a finite range. Again, if one disregards the finite range, integrals may diverge.

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 identifiable without using a suitable proper prior.

When the response is dichotomous, Zellner and Rossi (1984) and Rossi (1996) examined the existence of Bayesian estimates for the logit model. More recently, Chen and Shao (1999*a*) established necessary and sufficient conditions for the propriety of the posterior distribution with a uniform improper prior on the regression coefficients for dichotomous response models. However, when the response is polychotomous, the literature on the study of propriety is almost nonexistent. This may be partially due to the technical difficulty of the problem. By extending the results of Chen and Shao (1999*a*), we are now able to study the existence of Bayesian estimates for polychotomous quantal response models with general link functions. However, the conditions on the existence of Bayesian estimates for the polychotomous response models are somehow quite different from the ones for the dichotomous response models, which may be mainly due to the fact that the dichotomous response models belong to the exponential family; but the polychotomous response models do not.

We first introduce some notation, which will be used throughout the rest of the paper. Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ denote an $n \times 1$ vector of n independent polychotomous random variables. Assume that y_i takes a value of l ($1 \le l \le L$, $L > 2$) with probability

$$p_{il} = F(\theta_l + \boldsymbol{x}_i'\boldsymbol{\beta}) - F(\theta_{l-1} + \boldsymbol{x}_i'\boldsymbol{\beta}), \tag{1.1}$$

for $l = 1, \ldots, L$, where $\theta_0 = -\infty$, $\theta_L = \infty$, $\theta_1 \le \theta_2 \le \cdots \le \theta_{L-1}$, $F(\cdot)$ denotes a cumulative distribution function (cdf), $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik})'$ is a $k \times 1$ vector of covariates, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{L-1})'$ is an $(L-1) \times 1$ vector of $L-1$ intercepts, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ is a $k \times 1$ vector of regression coefficients. In (1.1), F^{-1} is called a link function. Three widely utilized functional forms for $F(\cdot)$ in (1.1) are

$$F(u) = \Phi(u), \tag{1.2}$$

the probit model,

$$F(u) = \frac{\exp(u)}{1 + \exp(u)}, \tag{1.3}$$

the logit model, and

$$F(u) = \exp[-\exp(-u)], \tag{1.4}$$

the log-log link model. In (1.2), $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. Of course, other functional forms of $F(\cdot)$, $0 < F(\cdot) < 1$, can be employed; see, for example, Agresti (1990), McCullagh and Nelder (1989), and Chen and Dey (1998). Using (1.1), the likelihood function, denoted by $L(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \boldsymbol{y})$, is

$$L(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \boldsymbol{y}) = \prod_{i=1}^{n} [F(\theta_{y_i} + \boldsymbol{x}_i'\boldsymbol{\beta}) - F(\theta_{y_i - 1} + \boldsymbol{x}_i'\boldsymbol{\beta})]. \tag{1.5}$$

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

$$\begin{aligned} \pi(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \boldsymbol{y}) &\propto L(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \boldsymbol{y}) \pi(\boldsymbol{\beta}, \boldsymbol{\theta}) \\ &= \left\{ \prod_{i=1}^{n} [F(\theta_{y_i} + \boldsymbol{x}_i'\boldsymbol{\beta}) - F(\theta_{y_i - 1} + \boldsymbol{x}_i'\boldsymbol{\beta})] \right\} \pi(\boldsymbol{\beta}, \boldsymbol{\theta}). \end{aligned} \tag{1.6}$$

It is clear that the posterior $\pi(\boldsymbol{\beta}, \theta \mid \boldsymbol{y})$ is proper if and only if

$$\begin{aligned} \underset{\theta_1 \le \theta_2 \le \cdots \le \theta_{L-1}}{\int \cdots \int} \int_{R^k} \left\{ \prod_{i=1}^{n} [F(\theta_{y_i} + \boldsymbol{x}_i'\boldsymbol{\beta}) - F(\theta_{y_i - 1} + \boldsymbol{x}_i'\boldsymbol{\beta})] \right\} \pi(\boldsymbol{\beta}, \boldsymbol{\theta}) d\boldsymbol{\beta} d\boldsymbol{\theta} \\ < \infty, \end{aligned} \tag{1.7}$$

where R^k is the k dimensional Euclidean space. In this article, we are interested in the following problem. If a uniform improper prior distribution for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, i.e., $\pi(\boldsymbol{\beta}, \boldsymbol{\theta}) \propto 1$, is specified, what are the necessary and sufficient conditions for the propriety of $\pi(\boldsymbol{\beta}, \theta \mid \boldsymbol{y})$? In addition, we are also interested in the sufficient conditions for the existence of Bayesian estimates, such as the posterior means and variances of $\boldsymbol{\beta}$ and θ.

The remaining of the article is organized as follows. In Section 2, we introduce the mathematical formulation of Bayesian estimates and discuss the implementational issues for computing the posterior quantities of interest. The main results are presented in Section 3, in which we carefully describe the precise conditions on the existence of Bayesian estimates and examine the relationship between the propriety of posterior distribution and the existence of the maximum likelihood estimate. Two illustrative examples are give in Section 4. The proofs of all main results presented in Section 3 are left to Section 5.

2. Bayesian analysis of polychotomous quantal response models

As discussed in Zellner and Rossi (1984) and Rossi (1996), the posterior expectation of various functions of the model parameters must be calculated. For example, the posterior mean is the usual Bayesian estimate of model parameters. Let $c(y)$ be the normalizing constant of the posterior distribution $\pi(\boldsymbol{\beta}, \theta \mid \boldsymbol{y})$ with a uniform improper prior, that is,

$$c(y) = \underset{\theta_1 \le \theta_2 \le \cdots \le \theta_{L-1}}{\int \cdots \int} \int_{R^k} \left\{ \prod_{i=1}^{n} [F(\theta_{y_i} + \boldsymbol{x}_i'\boldsymbol{\beta}) - F(\theta_{y_i - 1} + \boldsymbol{x}_i'\boldsymbol{\beta})] \right\} d\boldsymbol{\beta} d\theta. \tag{2.1}$$

If $c(y) < \infty$, then the posterior $\pi(\boldsymbol{\beta}, \theta \mid \boldsymbol{y})$ is *proper* and

$$\pi(\boldsymbol{\beta}, \theta \mid \boldsymbol{y}) = \frac{1}{c(y)} \prod_{i=1}^{n} [F(\theta_{y_i} + \boldsymbol{x}_i'\boldsymbol{\beta}) - F(\theta_{y_i - 1} + \boldsymbol{x}_i'\boldsymbol{\beta})]. \tag{2.2}$$

Furthermore, letting $h(\boldsymbol{\beta}, \theta)$ be a real-valued function or a real-valued vector/ matrix of $\boldsymbol{\beta}$ and θ, the posterior expectation of $h(\boldsymbol{\beta}, \theta)$ is given by

$$E[h(\boldsymbol{\beta}, \theta) \mid y] = \int \cdots \int_{\theta_1 \le \theta_2 \le \cdots \le \theta_{L-1}} \int_{R^k} h(\boldsymbol{\beta}, \theta) \pi(\boldsymbol{\beta}, \theta \mid \boldsymbol{y}) d\boldsymbol{\beta} d\theta. \tag{2.3}$$

Here, the posterior expectation $E[h(\boldsymbol{\beta}, \theta) \mid y]$ is said to be existent iff (i) $\pi(\boldsymbol{\beta}, \theta \mid \boldsymbol{y})$ is proper, and (ii)

$$\int \cdots \int_{\theta_1 \le \theta_2 \le \cdots \le \theta_{L-1}} \int_{R^k} \|h(\boldsymbol{\beta}, \theta)\| \pi(\boldsymbol{\beta}, \theta \mid \boldsymbol{y}) d\boldsymbol{\beta} d\theta < \infty, \tag{2.4}$$

where $\|h(\boldsymbol{\beta}, \theta)\|$ denotes a usual norm of $h(\boldsymbol{\beta}, \theta)$, for example, when $h(\boldsymbol{\beta}, \theta)$ is a real-valued function, $\|h(\boldsymbol{\beta}, \theta)\|$ simply takes the absolute value of $h(\boldsymbol{\beta}, \theta)$. This convention will be used throughout the remainder of the article.

For Bayesian analyst's posterior-distribution problems, many posterior estimates can be expressed as $E[h(\boldsymbol{\beta}, \theta) \mid y]$ given by (2.3). For example, (2.3) reduces to (a) posterior mean of $\boldsymbol{\beta}$ when $h(\boldsymbol{\beta}, \theta) = \boldsymbol{\beta}$; (b) posterior covariance matrix of $\boldsymbol{\beta}$ if $h(\boldsymbol{\beta}, \theta) = (\boldsymbol{\beta} - E(\boldsymbol{\beta} \mid y))'(\boldsymbol{\beta} - E(\boldsymbol{\beta} \mid y))$, where $E(\boldsymbol{\beta} \mid y) = \int \cdots \int_{\theta_1 \le \theta_2 \le \cdots \le \theta_{L-1}} \int_{R^k} \boldsymbol{\beta} \pi(\boldsymbol{\beta}, \theta \mid \boldsymbol{y}) d\boldsymbol{\beta} d\theta$; (c) posterior probability of a set A if $h(\boldsymbol{\beta}, \theta) = 1_A(\boldsymbol{\beta}, \theta)$, where $1_A(\boldsymbol{\beta}, \theta) = 1$ if $(\boldsymbol{\beta}, \theta) \in A$ and 0 otherwise; (d) marginal posterior density for β_j evaluated at $\beta_j = \beta_j^*$ $(1 \le j \le k)$ when

$$h(\boldsymbol{\beta}, \theta) = w(\beta_j \mid \beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_k) \frac{\pi(\beta_1, \ldots, \beta_{j-1}, \beta_j^*, \beta_{j+1}, \ldots, \beta_k, \theta \mid \boldsymbol{y})}{\pi(\boldsymbol{\beta}, \theta \mid \boldsymbol{y})},$$

where $w(\beta_j \mid \beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_k)$ is a completely known conditional density (see Chen (1994)). Some other posterior estimators, including posterior quantiles, Bayesian credible intervals, highest probability density (HPD) intervals, cannot be expressed as (2.3); see Chen and Shao (1999*b*). Precise conditions providing for the existence of various Bayes posterior estimates will be investigated in the next section.

Finally, we note that for polychotomous quantal response models, almost all Bayesian estimates are analytically intractable. Fortunately, the recently developed Markov chain Monte Carlo (MCMC) sampling methods, which include the Gibbs sampler (for example, Geman and Geman (1984) and Gelfand and Smith (1990)), Metropolis-Hastings algorithms (for example, Metropolis *et al.* (1953), Hastings (1970) and Tierney (1994)), can be easily adopted for computing the posterior properties. Assume that $\{(\boldsymbol{\beta}_{(b)}, \theta_{(b)}), b = 1, 2, \ldots, B\}$ is a MCMC sample from $\pi(\boldsymbol{\beta}, \theta \mid y)$. Then, $E(h(\boldsymbol{\beta}, \theta) \mid y)$ can be approximated by

$$\hat{E}(h(\boldsymbol{\beta}, \theta) \mid y) = \frac{1}{B}\sum_{b=1}^{B} h(\boldsymbol{\beta}_{(b)}, \theta_{(b)}). \tag{2.5}$$

It can be shown that under certain regularity conditions such as *ergodicity*,

$$\hat{E}(h(\boldsymbol{\beta}, \theta) \mid y) \to E(h(\boldsymbol{\beta}, \theta) \mid y) \quad \text{a.s.}$$

as $B \to \infty$; see, for example, Tierney (1994). In particular, if $\prod_{i=1}^{n}[F(\theta_{y_i} + \boldsymbol{x}_i'\boldsymbol{\beta}) - F(\theta_{y_i-1} + \boldsymbol{x}_i'\boldsymbol{\beta})]$ is log-concave, which is true for all the three common links (probit, logit, log-log link), then we can use an adaptive rejection algorithm of Gilks and Wild (1992) to generate a sample from the posterior distribution $\pi(\boldsymbol{\beta}, \theta \mid y)$ without knowing the normalizing constant $c(y)$. In addition, if F is a scale mixture of normal link, we can use the efficient algorithms developed by Chen and Dey (1998).

3. Existence of the Bayesian estimates

To obtain necessary and sufficient conditions for the existence of Bayesian estimates described in Section 2, we first introduce some notation. Let

$$I = \{1 \le i \le n : y_i = L\}, \quad J = \{1 \le j \le n : y_j = 1\}, \quad \text{and}$$
$$T = \{1 \le i \le n : 1 < y_i < L\}.$$

That is, sets I, J and T divide the n observations $y_1, y_2, \ldots, y_n$ into three groups so that sets I and J include all polychotomous responses with values of L and 1 respectively, while set T contains all observations with values between 1 and L exclusively.

The main results are given as follows. We first present the necessary and sufficient conditions for the propriety of the posterior distribution given by (2.2).

THEOREM 3.1. *Assume that*

$$\int_{-\infty}^{\infty} |u|^{k+L-1} dF(u) < \infty. \tag{3.1}$$

If the following conditions are satisfied

(C1) *I and J are non-empty sets;*

(C2) $\forall\ \varepsilon_l = \pm 1$, $1 \le l \le k$, $\forall\ a_l \ge 0$, $b_r \ge 0$, *with* $\sum_{l=1}^{k} a_l + \sum_{r=2}^{L-1} b_r > 0$,

$$\min_{j \in J \cup T}\left(\sum_{r=2}^{y_j} b_r + \sum_{l=1}^{k} x_{jl}\varepsilon_l a_l\right) < \max_{i \in I \cup T}\left(\sum_{r=2}^{y_i - 1} b_r + \sum_{l=1}^{k} x_{il}\varepsilon_l a_l\right),$$

then the posterior is proper, that is,

$$\int \cdots \int_{\theta_1 \le \theta_2 \le \cdots \le \theta_{L-1}} \int_{R^k} L(\boldsymbol{\beta}, \theta \mid \boldsymbol{y}) d\boldsymbol{\beta} d\theta < \infty. \tag{3.2}$$

THEOREM 3.2. *If $F(b) - F(a) > 0$ for every $b > a$, then* (C1) *and* (C2) *are necessary conditions for* (3.2).

From Theorems 3.1 and 3.2, for a link function F^{-1}, if F satisfies (3.1) and $F(b) - F(a) > 0$ for every $b > a$, then (C1) and (C2) are the sufficient and necessary conditions for (3.2). As to condition (C1), if either I or J is empty, then there is a quasicomplete separation in the data points. In this case, obviously there is no information available to estimate either θ_1 or θ_{L-1}. Thus, (3.2) cannot hold. As to condition (C2), it is not difficult to show that (C2) implies that the $n \times k$ design matrix is of full column rank. Hence, the full rank condition on the design matrix is necessary for the propriety of the posterior but not sufficient. On the other hand, condition (C2) may not be easily verified. To this end, we present two sufficient conditions for (C2).

Let $\boldsymbol{X}$ be the $n \times k$ known design matrix with rows $\boldsymbol{x}_i'$. Define $\boldsymbol{x}_i^* = (1, \boldsymbol{x}_i')'$ for $i = 1, 2, \ldots, n$,

$$\boldsymbol{X}_1 = (\boldsymbol{x}_i^*, i \in I, -\boldsymbol{x}_j^*, j \in J \cup T)', \quad \text{and} \quad \boldsymbol{X}_2 = (\boldsymbol{x}_j^*, j \in J, -\boldsymbol{x}_i^*, i \in I \cup T)'.$$

Thus, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are the two reconstructed design matrices from $\boldsymbol{X}$. For example, X_1 is simply obtained by taking the i-th row to be $\boldsymbol{x}_i^*$ for $i \in I$ and the j-th row to be $-\boldsymbol{x}_j^*$ for $j \in J \cup T$.

PROPOSITION 3.1. *Assume that I and J are non-empty sets and that the design matrix $\boldsymbol{X}$ is of full column rank. Then* (C2) *holds if one of the following conditions is satisfied:*

(C3) *There exists a positive vector $\boldsymbol{a} = (a_1, \ldots, a_n)' \in R^n$, i.e., each component $a_i > 0$, such that*

$$\boldsymbol{a}'\boldsymbol{X}_1 = \boldsymbol{0}. \tag{3.3}$$

(C4) *There exists a positive vector $\boldsymbol{a} = (a_1, \ldots, a_n)' \in R^n$ such that*

$$\boldsymbol{a}'\boldsymbol{X}_2 = \boldsymbol{0}. \tag{3.4}$$

Thus, for verifying condition (C3) or (C4) it suffices to find a positive solution for (3.3) or (3.4). This is a standard linear programming problem which can be done using commercially available software, for example, CPLEX (CPLEX Optimization (1992)).

Next, we briefly discuss the conditions on link functions. For the three widely used probit, logit and log-log link models, it is easy to see that $F(b) - F(a) > 0$

for every $b > a$ and condition (3.1) holds. However, condition (3.1) may be too restrictive for some links other than probit, logit, and log-log link, such as t-links with degrees of freedom less than or equal to $k + L - 1$, stable distribution family links, and exponential power distribution family links (see Chen and Dey (1998)). At the cost of assuming more restrictive condition on $\boldsymbol{x}_i$ and y_i, we are able to weaken the assumption on the moment condition (3.1).

Let

$$T_{l,1} = \{(i,j) : i \in I, j \in J, x_{il} - x_{jl} > 0\},$$
$$T_{l,-1} = \{(i,j) : i \in I, j \in J, x_{il} - x_{jl} < 0\}.$$

For $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)$, where $\varepsilon_l = \pm 1$, let

$$T(\varepsilon) = \bigcap_{l=1}^{k} T_{l,\varepsilon_l}.$$

Thus, for a given ε, $T(\varepsilon)$ contains all pairs (i,j)'s of observations such that either $x_{il} - x_{jl} > 0$ or $x_{il} - x_{jl} < 0$ for all $1 \le l \le k$.

PROPOSITION 3.2. *Assume that $T(\varepsilon) \neq \emptyset$. Let $X_1, X_2, \ldots, X_n$ be independent random variables with the same cdf F. If*

$$\forall\, \varepsilon, \qquad E\left[\min_{(i,j)\in T(\varepsilon)} (|X_i| + |X_j|)^{k+L-1}\right] < \infty,$$

then (3.2) holds. In particular, if there are $I(\varepsilon)$ and $J(\varepsilon)$ such that $I(\varepsilon) \times J(\varepsilon) \subset T(\varepsilon)$ and

$$E\left[\min_{1\le i\le m^*} |X_i|^{k+L-1}\right] < \infty, \tag{3.5}$$

where $m^* = \min_\varepsilon \min(\text{card of } I(\varepsilon), \text{card of } J(\varepsilon))$, then (3.2) is true.

If $m^* > k + L - 1$, it is easy to verify that (3.5) holds even when F is the cdf of a Cauchy distribution. Therefore, the condition on F given in Proposition 3.2 is much weaker than the one given by (3.1). Of course, the restriction on the design matrix is much stronger. The condition $T(\varepsilon) \neq \emptyset$ requires that there must exist some common (i,j)'s such that either $x_{il} - x_{jl} > 0$ or $x_{il} - x_{jl} < 0$ for all l.

Now we consider the existence of Bayesian estimates. Given a proper posterior, the posterior median or the HPD intervals will exist. However, Theorem 3.1 cannot guarantee the existence of the posterior moments. Therefore, we present the sufficient conditions on the existence of the posterior moments and the posterior moment generating function in the next theorem.

THEOREM 3.3. *Assume that (C1) and (C2) are satisfied. If*

$$\int_{-\infty}^{\infty} |u|^{k+L-1+p} dF(u) < \infty \tag{3.6}$$

for some $p \geq 0$, *then*

$$\int \cdots \int_{\theta_1 \leq \theta_2 \leq \cdots \leq \theta_{L-1}} \int_{R^k} \|\zeta\|^p L(\boldsymbol{\beta}, \theta \mid \boldsymbol{y}) d\boldsymbol{\beta} d\theta < \infty, \tag{3.7}$$

where $\zeta = (\boldsymbol{\beta}', \theta')'$ *and the norm* $\|\zeta\| = (\sum_{j=1}^k \beta_j^2 + \sum_{l=1}^{L-1} \theta_l^2)^{1/2}$. *If*

$$\int_{-\infty}^{\infty} e^{t_0|u|} dF(u) < \infty \tag{3.8}$$

for some $t_0 > 0$, *then*

$$\int \cdots \int_{\theta_1 \leq \theta_2 \leq \cdots \leq \theta_{L-1}} \int_{R^k} e^{\delta_0 \|\zeta\|} L(\boldsymbol{\beta}, \theta \mid \boldsymbol{y}) d\boldsymbol{\beta} d\theta < \infty, \tag{3.9}$$

for some $\delta_0 > 0$.

For the three widely used probit, logit and log-log link functions, it is easy to see that conditions (3.6) and (3.8) holds. Hence, (C1) and (C2) are the sufficient conditions for (3.7) and (3.9). However, the moment generating condition (3.8) does not hold for all t-links.

Finally, we examine the relationship between the propriety and the existence of the maximum likelihood estimate. The next theorem confirms that (C1) and (C2) are the sufficient conditions for the existence of the maximum likelihood estimate.

THEOREM 3.4. *Assume that* F *is continuous and that* (C1) *and* (C2) *are satisfied, then the maximum likelihood estimate of* $(\boldsymbol{\beta}, \theta)$ *exists.*

When F is continuous and satisfies $F(b) - F(a) > 0$ for every $b > a$, it follows from Theorems 3.2 and 3.4 that if the posterior is proper, then the MLE exists. This result is useful since (i) the posterior is simply not proper if the MLE does not exist and (ii) the existence of the MLE is routinely checked by many widely used statistical softwares such as SAS and Splus. For example, in SAS the LOGISTIC procedure performs some checking to determine whether the input data have a configuration that leads to infinite parameter estimates (see Stokes *et al.* (1995) for detailed discussions). On the other hand, as proved by Wedderburn (1976), the MLE exists only for certain generalized linear models in the exponential family. Hence, Theorem 3.4 is of great importance in checking improperness of the posterior distribution, while Proposition 3.1 provides practically usable conditions to determine the propriety of a posterior density.

Table 1. The rating data.

Gender	female	female	female	male	male
Rating	poor	fair	good	poor	fair

4. Illustrative examples

In this section, we present two simple examples to illustrate how to verify the conditions stated in Section 3.

Example 4.1. At a private college, five students, including three females and two males, submitted course project reports in a given semester. The reports were graded and were rated as "poor", "fair", or "good" by their instructor. The rating results are given in Table 1.

This is a standard polychotomous response problem in which the response is the project grade while *gender* serves as a sole covariate. We code female as 0 and male as 1 and we also denote the response to be 1 for "poor", 2 for "fair", and 3 for "good". Then, the sample size $n = 5$, $k = 1$, $L = 3$, and the five observations are $x_1 = x_2 = x_3 = 0$, $x_4 = x_5 = 1$, $y_1 = 1$, $y_2 = 2$, $y_3 = 3$, $y_4 = 1$, and $y_5 = 2$. Using the notation given in Section 3, we have $I = \{3\}$, $J = \{1, 4\}$, $T = \{2, 5\}$, $\boldsymbol{X} = (0, 0, 0, 1, 1)'$, and

$$\boldsymbol{X}_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ -1 & 0 \\ -1 & 0 \\ -1 & -1 \end{pmatrix}.$$

It is easy to see that I and J are non-empty, $\boldsymbol{X}$ is of full rank, and $a = (2, 1, 1, 1, 1)'$ satisfies $a'\boldsymbol{X}_2 = 0$. Hence (C4) is satisfied, and it follows from Proposition 3.1 and Theorem 3.1 that the resulting posterior is proper. A small SAS program for this three-level ordinal response data set was also written. The LOGISTIC procedure returns the unique maximum likelihood estimates for all three commonly used links (probit, logit, and log-log link).

Example 4.2. We consider the same data given in Example 4.1 except the last male student dropped out from the instructor's grade sheet. Then, we have $x_1 = x_2 = x_3 = 0$, $x_4 = 1$, $y_1 = 1$, $y_2 = 2$, $y_3 = 3$, and $y_4 = 1$. Then $I = \{3\}$, $J = \{1, 4\}$ and $T = \{2\}$. For this case, we will show that the resulting posterior with a uniform improper prior is not proper.

Since $k = 1$, in Theorem 3.1 we take $\epsilon_1 = 1$, $a_1 = 1$ and $b_2 = 0$. Then, we have Table 2. From Table 2, it is easy to see that (C2) does not hold. Therefore, it directly follows from Theorems 3.1 and 3.2 that the resulting posterior is not proper. Similarly, we consulted with the LOGISTIC procedure in SAS for verifying the existence of the MLE, and the LOGISTIC procedure provides the following information:

Table 2. The computation of condition (C2).

$j \in J \cup T$	$\sum_{r=2}^{y_j} b_r + \sum_{l=1}^{k} x_{jl}\varepsilon_l a_l$	$i \in I \cup T$	$\sum_{r=2}^{y_i-1} b_r + \sum_{l=1}^{k} x_{il}\varepsilon_l a_l$
1	0	2	0
2	0	3	0
4	1		
	min = 0		max = 0

"There is possibly a quasicomplete separation in the sample points. The maximum likelihood estimate may not exist."

From the above two simple illustrative examples, we have learned that estimating the three unknown parameters for a three-level ordinal response model requires five data points.

5. The proofs of theorems

In this section, we provide the proofs for all the theorems and the propositions stated in Section 3.

PROOF OF THEOREM 3.1. It follows from the fact that

$$F(b) - F(a) = \int_{-\infty}^{\infty} 1\{a < u \le b\} dF(u),$$

where the indicator function $1\{a < u \le b\} = 1$ if $a < u \le b$ and 0 otherwise, and the Fubini theorem that

$$\begin{aligned} &\int_{\theta_1 \le \theta_2 \le \cdots \le \theta_{L-1}} \int_{R^k} L(\boldsymbol{\beta} \mid \boldsymbol{y}) d\boldsymbol{\beta} d\theta \\ &= \int_{R^n} \int_{\theta_1 \le \theta_2 \le \cdots \le \theta_{L-1}} \int_{R^k} \\ &\quad \cdot 1\{\theta_{y_i - 1} + \boldsymbol{x}_i'\boldsymbol{\beta} < u_i < \theta_{y_i} + \boldsymbol{x}_i'\boldsymbol{\beta}, 1 \le i \le n\} d\boldsymbol{\beta} d\theta d\boldsymbol{F}(\boldsymbol{u}), \end{aligned} \tag{5.1}$$

where $d\boldsymbol{F}(\boldsymbol{u})$ stands for $dF(u_1) \cdots dF(u_n)$. Put

$$h(\boldsymbol{u}) = \int_{\theta_1 \le \theta_2 \le \cdots \le \theta_{L-1}} \int_{R^k} 1\{\theta_{y_i - 1} + \boldsymbol{x}_i'\boldsymbol{\beta} < u_i < \theta_{y_i} + \boldsymbol{x}_i'\boldsymbol{\beta}, 1 \le i \le n\} d\boldsymbol{\beta} d\theta. \tag{5.2}$$

By (3.1), it suffices to show that

$$h(\boldsymbol{u}) \le A(\boldsymbol{x}, \boldsymbol{y}) \max_{1 \le i \le n} |u_i|^{k+L-1} \tag{5.3}$$

for some $A(\boldsymbol{x}, \boldsymbol{y}) < \infty$. To estimate $h(\boldsymbol{u})$, letting $t_1 = \theta_1$, $t_r = \theta_r - \theta_{r-1}$ for $r = 2, \ldots, L-1$ yields

$$h(\boldsymbol{u}) = \int_{t_r \geq 0, r=2,\ldots,L-1} \int_{R^k} \int_{-\infty}^{\infty} \cdot 1\left\{\sum_{r=1}^{y_i-1} t_r + \boldsymbol{x}_i'\boldsymbol{\beta} < u_i < \sum_{r=1}^{y_i} t_r + \boldsymbol{x}_i'\boldsymbol{\beta}, i \in T\right\} \cdot 1\left\{\sum_{r=1}^{y_i-1} t_r + \boldsymbol{x}_i'\boldsymbol{\beta} < u_i, i \in I\right\} \cdot 1\{u_i < t_1 + \boldsymbol{x}_i'\boldsymbol{\beta}, i \in J\}\, dt_1 d\boldsymbol{\beta} dt_2 \cdots dt_{L-1}, \tag{5.4}$$

where $1\{\sum_{r=1}^{y_i-1} t_r + \boldsymbol{x}_i'\boldsymbol{\beta} < u_i < \sum_{r=1}^{y_i} t_r + \boldsymbol{x}_i'\boldsymbol{\beta}, i \in T\}$ denotes 1 if T is empty. Note that

$$\begin{aligned} &1\left\{\sum_{r=1}^{y_i-1} t_r + \boldsymbol{x}_i'\boldsymbol{\beta} < u_i < \sum_{r=1}^{y_i} t_r + \boldsymbol{x}_i'\boldsymbol{\beta}, i \in T\right\} \\ &\quad \cdot 1\left\{\sum_{r=1}^{y_i-1} t_r + \boldsymbol{x}_i'\boldsymbol{\beta} < u_i, i \in I\right\} \cdot 1\{u_i < t_1 + \boldsymbol{x}_i'\boldsymbol{\beta}, i \in J\} \\ &= 1\left\{\max_{j \in J\cup T}\left(u_j - \sum_{r=2}^{y_j} t_r - \boldsymbol{x}_j'\boldsymbol{\beta}\right) < t_1 < \min_{i \in I\cup T}\left(u_i - \sum_{r=2}^{y_i-1} t_r - \boldsymbol{x}_i'\boldsymbol{\beta}\right)\right\} \\ &= 1\left\{-\min_{j \in J\cup T}\left(-u_j + \sum_{r=2}^{y_j} t_r + \boldsymbol{x}_j'\boldsymbol{\beta}\right) < t_1 < \min_{i \in I\cup T}\left(u_i - \sum_{r=2}^{y_i-1} t_r - \boldsymbol{x}_i'\boldsymbol{\beta}\right)\right\}. \end{aligned} \tag{5.5}$$

It follows from (5.4) and (5.5) that

$$\begin{aligned} h(\boldsymbol{u}) = \int_{t_r \geq 0, r=2,\ldots,L-1} \int_{R^k} &\left(\min_{j \in J\cup T}\left(-u_j + \sum_{r=2}^{y_j} t_r + \boldsymbol{x}_j'\boldsymbol{\beta}\right) + \min_{i \in I\cup T}\left(u_i - \sum_{r=2}^{y_i-1} t_r - \boldsymbol{x}_i'\boldsymbol{\beta}\right)\right) \\ &\cdot 1\left\{\min_{j \in J\cup T}\left(-u_j + \sum_{r=2}^{y_j} t_r + \boldsymbol{x}_j'\boldsymbol{\beta}\right) + \min_{i \in I\cup T}\left(u_i - \sum_{r=2}^{y_i-1} t_r - \boldsymbol{x}_i'\boldsymbol{\beta}\right) > 0\right\} d\boldsymbol{\beta} dt_2 \cdots dt_{L-1} \end{aligned} \tag{5.6}$$

$$= \sum_{\varepsilon_l=\pm1, 1\le l\le k} \int_{(R^+)^{k+L-2}} \left(\min_{j\in J\cup T} \left(-u_j + \sum_{r=2}^{y_j} t_r + \sum_{l=1}^{k} x_{jl}\varepsilon_l\beta_l \right) + \min_{i\in I\cup T} \left(u_i - \sum_{r=2}^{y_i-1} t_r - \sum_{l=1}^{k} x_{il}\varepsilon_l\beta_l \right) \right)$$
$$\cdot 1\left\{ \min_{j\in J\cup T} \left(-u_j + \sum_{r=2}^{y_j} t_r + \sum_{l=1}^{k} x_{jl}\varepsilon_l\beta_l \right) + \min_{i\in I\cup T} \left(u_i - \sum_{r=2}^{y_i-1} t_r - \sum_{l=1}^{k} x_{il}\varepsilon_l\beta_l \right) > 0 \right\}$$
$$\cdot d\beta_1 \cdots d\beta_k dt_2 \cdots dt_{L-1}.$$

Let

$$d = \inf_{\varepsilon_l=\pm1, 1\le l\le k, 0\le a_l\le 1, 0\le b_r\le 1, 2\le r\le L-1, a_{l_0}=1 \text{ for some } l_0, \text{ or } b_{r_0}=1 \text{ for some } r_0} \cdot \left(\max_{i\in I\cup T} \left(\sum_{r=2}^{y_i-1} b_r + \sum_{l=1}^{k} x_{il}\varepsilon_l a_l \right) - \min_{j\in J\cup T} \left(\sum_{r=2}^{y_j} b_r + \sum_{l=1}^{k} x_{jl}\varepsilon_l a_l \right) \right).$$

By (C2), we have $d > 0$. Denote $\max_{1\le l\le k, 2\le r\le L-1} \max(\beta_l, t_r)$ by M^*. We have

$$\begin{aligned} (5.7) \qquad 0 &\le \min_{j\in J\cup T} \left(-u_j + \sum_{r=2}^{y_j} t_r + \sum_{l=1}^{k} x_{jl}\varepsilon_l\beta_l \right) + \min_{i\in I\cup T} \left(u_i - \sum_{r=2}^{y_i-1} t_r - \sum_{l=1}^{k} x_{il}\varepsilon_l\beta_l \right) \\ &\le \min_{j\in J\cup T} \left(\sum_{r=2}^{y_j} t_r + \sum_{l=1}^{k} x_{jl}\varepsilon_l\beta_l \right) + \max_{j\in J\cup T} |u_j| \\ &\quad - \max_{i\in I\cup T} \left(\sum_{r=2}^{y_i-1} t_r + \sum_{l=1}^{k} x_{il}\varepsilon_l\beta_l \right) + \max_{i\in I\cup T} |u_i| \\ &\le 2\max_{1\le i\le n} |u_i| - M^* \left(\max_{i\in I\cup T} \left(\sum_{r=2}^{y_i-1} t_r/M^* + \sum_{l=1}^{k} x_{il}\varepsilon_l\beta_l/M^* \right) - \min_{j\in J\cup T} \left(\sum_{r=2}^{y_j} t_r/M^* + \sum_{l=1}^{k} x_{jl}\varepsilon_l\beta_l/M^* \right) \right) \\ &\le 2\max_{1\le i\le n} |u_i| - M^* d. \end{aligned}$$

Hence

$$(5.8) \qquad \max(\beta_l, t_r) = M^* \le 2\max_{1\le i\le n} |u_i|/d.$$

This proves (5.4), by (5.6) and (5.8). □

PROOF OF THEOREM 3.2. If $J = \emptyset$, by (5.2)

$$h(\boldsymbol{u}) = \int_{R^k} \int_{\theta_1 \le \theta_2 \le \cdots \le \theta_{L-1}} \cdot 1\{\theta_{y_i - 1} + \boldsymbol{x}_i'\boldsymbol{\beta} < u_i < \theta_{y_i} + \boldsymbol{x}_i'\boldsymbol{\beta}, i \in T \cup I\} d\theta_1 \cdots d\theta_{L-1} d\boldsymbol{\beta} = \infty.$$

Similarly, if $I = \emptyset$, we have also $h(\boldsymbol{u}) = \infty$. This proves (C1) necessary for (3.2).

Next we show that (C2) is necessary. If (C2) is not satisfied, then there exist sequences of ε_l, a_l and b_r such that

$$\min_{j \in J \cup T} \left(\sum_{r=2}^{y_j} b_r + \sum_{l=1}^{k} x_{jl}\varepsilon_l a_l \right) \ge \max_{i \in I \cup T} \left(\sum_{r=2}^{y_i - 1} b_r + \sum_{l=1}^{k} x_{il}\varepsilon_l a_l \right). \tag{5.9}$$

Noting that J is a non-empty set, (5.9) implies that $\sum_{l=1}^{k} a_l > 0$. Without loss of generality, assume $\varepsilon_l = 1$ for $l = 1, \ldots, k$ and $a_1 = 1$, that is,

$$\min_{j \in J \cup T} \left(\sum_{r=2}^{y_j} b_r + x_{j1} + \sum_{l=2}^{k} x_{jl} a_l \right) \ge \max_{i \in I \cup T} \left(\sum_{r=2}^{y_i - 1} b_r + x_{i1} + \sum_{l=2}^{k} x_{il} a_l \right). \tag{5.10}$$

Let $T_l = \{i : y_i = l\}$ for $l = 1, 2, \ldots, L$ and put

$$\mathcal{C} = \{\boldsymbol{u} : 0 \le -u_j + (l-1) \le 1/2 \text{ for } j \in T_l, 1 \le l \le L\}. \tag{5.11}$$

Since $F(b) - F(a) > 0$ for any $a < b$, it suffices to show that for any $\boldsymbol{u} \in \mathcal{C}$,

$$h(\boldsymbol{u}) = \infty. \tag{5.12}$$

From (5.6) it follows that

$$\begin{aligned} h(\boldsymbol{u}) \ge \int_{(R^+)^{k+L-2}} & \left(\min_{j \in J \cup T} \left(-u_j + \sum_{r=2}^{y_j} t_r + \boldsymbol{x}_j'\boldsymbol{\beta} \right) + \min_{i \in I \cup T} \left(u_i - \sum_{r=2}^{y_i - 1} t_r - \boldsymbol{x}_i'\boldsymbol{\beta} \right) \right) \\ & \cdot 1 \left\{ \min_{j \in J \cup T} \left(-u_j + \sum_{r=2}^{y_j} t_r + \boldsymbol{x}_j'\boldsymbol{\beta} \right) + \min_{i \in I \cup T} \left(u_i - \sum_{r=2}^{y_i - 1} t_r - \boldsymbol{x}_i'\boldsymbol{\beta} \right) > 0 \right\} d\boldsymbol{\beta} dt_2 \cdots dt_{L-1} \\ \ge \int_{t_r \ge 1, r=2,\ldots,L-1} \int_{(R^+)^k} & \left(\min_{j \in J \cup T} \left(-u_j + \sum_{r=2}^{y_j} t_r + \boldsymbol{x}_j'\boldsymbol{\beta} \right) \right. \end{aligned} \tag{5.13}$$

$$+\min_{i\in I\cup T}\left(u_i-\sum_{r=2}^{y_i-1}t_r-\boldsymbol{x}_i'\boldsymbol{\beta}\right)\Bigg)$$

$$\cdot 1\left\{\min_{j\in J\cup T}\left(-u_j+\sum_{r=2}^{y_j}t_r+\boldsymbol{x}_j'\boldsymbol{\beta}\right)\right.$$

$$\left.+\min_{i\in I\cup T}\left(u_i-\sum_{r=2}^{y_i-1}t_r-\boldsymbol{x}_i'\boldsymbol{\beta}\right)>0\right\}d\boldsymbol{\beta}dt_2\cdots dt_{L-1}$$

$$=\int_{(R^+)^{k+L-2}}\left(\min_{j\in J\cup T}\left(-u_j+(y_j-1)+\sum_{r=2}^{y_j}t_r+\boldsymbol{x}_j'\boldsymbol{\beta}\right)\right.$$

$$\left.+\min_{i\in I\cup T}\left(u_i-(y_i-2)-\sum_{r=2}^{y_i-1}t_r-\boldsymbol{x}_i'\boldsymbol{\beta}\right)\right)$$

$$\cdot 1\left\{\min_{j\in J\cup T}\left(-u_j+(y_j-1)+\sum_{r=2}^{y_j}t_r+\boldsymbol{x}_j'\boldsymbol{\beta}\right)\right.$$

$$\left.+\min_{i\in I\cup T}\left(u_i-(y_i-2)-\sum_{r=2}^{y_i-1}t_r-\boldsymbol{x}_i'\boldsymbol{\beta}\right)>0\right\}$$

$$\cdot d\boldsymbol{\beta}dt_2\cdots dt_{L-1}.$$

Noting that

$$\begin{aligned}
&\min_{j\in J\cup T}(-u_j+(y_j-1))+\min_{i\in I\cup T}(u_i-(y_i-2))\\
&\quad=\min_{1\le l\le L-1}\min_{j\in T_l}(-u_j+l-1)+\min_{2\le l\le L}\min_{i\in T_l}(u_i-(l-2))\\
&\quad=1+\min_{1\le l\le L-1}\min_{j\in T_l}(-u_j+l-1)-\max_{2\le l\le L}\max_{i\in T_l}(-u_i+(l-1))\ge 1/2
\end{aligned}$$

for $\boldsymbol{u}\in\mathcal{C}$, we have

$$\begin{aligned}
&\min_{j\in J\cup T}\left(-u_j+(y_j-1)+\sum_{r=2}^{y_j}t_r+\boldsymbol{x}_j'\boldsymbol{\beta}\right)+\min_{i\in I\cup T}\left(u_i-(y_i-2)-\sum_{r=2}^{y_i-1}t_r-\boldsymbol{x}_i'\boldsymbol{\beta}\right)\\
&\quad\ge\min_{j\in J\cup T}\left(\sum_{r=2}^{y_j}t_r+\boldsymbol{x}_j'\boldsymbol{\beta}\right)-\max_{i\in I\cup T}\left(\sum_{r=2}^{y_i-1}t_r+\boldsymbol{x}_i'\boldsymbol{\beta}\right)\\
&\qquad+\min_{j\in J\cup T}(-u_j+(y_j-1))+\min_{i\in I\cup T}(u_i-(y_i-2))\\
&\quad\ge 1/2+\min_{j\in J\cup T}\left(\sum_{r=2}^{y_j}t_r+\boldsymbol{x}_j'\boldsymbol{\beta}\right)-\max_{i\in I\cup T}\left(\sum_{r=2}^{y_i-1}t_r+\boldsymbol{x}_i'\boldsymbol{\beta}\right).
\end{aligned}$$

Thus, by (5.13)

$$h(\boldsymbol{u})\ge\int_{(R^+)^{k+L-2}}\left(1/2+\min_{j\in J\cup T}\left(\sum_{r=2}^{y_j}t_r+\boldsymbol{x}_j'\boldsymbol{\beta}\right)-\max_{i\in I\cup T}\left(\sum_{r=2}^{y_i-1}t_r+\boldsymbol{x}_i'\boldsymbol{\beta}\right)\right)$$

$$\cdot 1\left\{1/2+\min_{j\in J\cup T}\left(\sum_{r=2}^{y_j}t_r+\boldsymbol{x}_j'\boldsymbol{\beta}\right)-\max_{i\in I\cup T}\left(\sum_{r=2}^{y_i-1}t_r+\boldsymbol{x}_i'\boldsymbol{\beta}\right)>0\right\}$$
$$\cdot\, d\boldsymbol{\beta}dt_2\cdots dt_{L-1}.$$

Letting $\eta_1=\beta_1$, $\eta_i=\beta_i/\beta_1$ for $i=2,\ldots,k$, and $s_r=t_r/\beta_1$ for $r=2,\ldots,L-1$ yields

$$h(\boldsymbol{u})\geq\int_{(R^+)^{k+L-2}}\eta_1^{k+L-3}$$
$$\cdot\left[1/2+\eta_1\left\{\min_{j\in J\cup T}\left(\sum_{r=2}^{y_j}s_r+x_{j1}+\sum_{l=2}^{k}x_{jl}\eta_l\right)\right.\right.$$
$$\left.\left.-\max_{i\in I\cup T}\left(\sum_{r=2}^{y_i-1}s_r+x_{i1}+\sum_{l=2}^{k}x_{il}\eta_l\right)\right\}\right]$$
$$\cdot 1\left\{1/2+\eta_1\left\{\min_{j\in J\cup T}\left(\sum_{r=2}^{y_j}s_r+x_{j1}+\sum_{l=2}^{k}x_{jl}\eta_l\right)\right.\right.$$
$$\left.\left.-\max_{i\in I\cup T}\left(\sum_{r=2}^{y_i-1}s_r+x_{i1}+\sum_{l=2}^{k}x_{il}\eta_l\right)\right\}>0\right\}$$
$$\cdot\, d\boldsymbol{\eta}ds_2\cdots ds_{L-2}.$$

By (5.10),

$$\eta_1\left\{\min_{j\in J\cup T}\left(\sum_{r=2}^{y_j}s_r+x_{j1}+\sum_{l=2}^{k}x_{jl}\eta_l\right)-\max_{i\in I\cup T}\left(\sum_{r=2}^{y_i-1}s_r+x_{i1}+\sum_{l=2}^{k}x_{il}\eta_l\right)\right\}$$
$$\geq\eta_1\left\{\min_{j\in J\cup T}\left(\sum_{r=2}^{y_j}b_r+x_{j1}+\sum_{l=2}^{k}x_{jl}a_l\right)\right.$$
$$\left.-\max_{i\in I\cup T}\left(\sum_{r=2}^{y_i-1}b_r+x_{i1}+\sum_{l=2}^{k}x_{il}a_l\right)\right\}$$
$$+\eta_1\left\{\min_{j\in J\cup T}\left(\sum_{r=2}^{y_j}(s_r-b_r)+\sum_{l=2}^{k}x_{jl}(\eta_l-a_l)\right)\right.$$
$$\left.-\max_{i\in I\cup T}\left(\sum_{r=2}^{y_i-1}(s_r-b_r)+\sum_{l=2}^{k}x_{il}(\eta_l-a_l)\right)\right\}$$
$$\geq-\eta_1\left\{\max_{2\leq l\leq k}|\eta_l-a_l|+\max_{2\leq r\leq L-1}|s_r-b_r|\right\}B,$$

where $B=L+\max_{j\in I\cup J\cup T}\sum_{l=2}^{k}|x_{jl}|$. Hence,

$$h(\boldsymbol{u})\geq\int_{\eta_1>0,0<\eta_l-a_l<1/(8B\eta_1),2\leq l\leq k,0<s_r-b_r<1/(8B\eta_1),2\leq r\leq L-1}$$

$$\begin{aligned}
&\cdot \eta_1^{k+L-3}\left[1/2 - \eta_1 B\left(\max_{2\le l\le k}|\eta_l - a_l| + \max_{2\le r\le L-1}|s_r - b_r|\right)\right] \\
&\cdot 1\left\{1/2 - \eta_1 B\left(\max_{2\le l\le k}|\eta_l - a_l| + \max_{2\le r\le L-1}|s_r - b_r|\right) > 0\right\} \\
&\cdot d\boldsymbol{\eta} ds_2\cdots ds_{L-2} \\
\ge &\int_{\eta_1>0, 0<\eta_l - a_l<1/(8B\eta_1), 2\le l\le k, 0<s_r - b_r<1/(8B\eta_1), 2\le r\le L-1} \\
&\cdot (1/4)\eta_1^{k+L-3}\eta_1^{k+L-3} d\boldsymbol{\eta} ds_2\cdots ds_{L-2} \\
= &(1/4)\int_0^\infty \eta_1^{k+L-3}(1/(8B\eta_1))^{k+L-3} d\eta_1 \\
= &\infty,
\end{aligned}$$

as desired. □

PROOF OF PROPOSITION 3.1. We first show that (C3) implies (C2). Note that $\forall t_l$ and $b_r \ge 0$ satisfying $\sum_{l=1}^k |t_l| + \sum_{r=2}^{L-1} b_r > 0$,

$$\sum_{r=2}^{L-1} b_r + \max_{i\in I}\left(\sum_{l=1}^k x_{il}t_l\right) \le \max_{i\in I\cup T}\left(\sum_{r=2}^{y_i-1} b_r + \sum_{l=1}^k x_{il}t_l\right)$$

and

$$\min_{j\in J\cup T}\left(\sum_{r=2}^{y_j} b_r + \sum_{l=1}^k x_{jl}t_l\right) \le \sum_{r=2}^{L-1} b_r + \min_{j\in J\cup T}\left(\sum_{l=1}^k x_{jl}t_l\right).$$

It suffices to show that $\forall t_l$ with $\sum_{l=1}^k |t_l| > 0$

$$\min_{j\in J\cup T}\left(\sum_{l=1}^k x_{jl}t_l\right) < \max_{i\in I}\left(\sum_{l=1}^k x_{il}t_l\right). \tag{5.14}$$

Let $\boldsymbol{a} > \boldsymbol{0}$ satisfy $\boldsymbol{a}'\boldsymbol{X}_1 = \boldsymbol{0}$, namely

$$\sum_{j\in J\cup T} a_j = \sum_{i\in I} a_i$$

and

$$\sum_{j\in J\cup T} a_j x_{jl} = \sum_{i\in I} a_i x_{il} \quad \text{for} \quad l = 1,\ldots,k.$$

Without loss of generality, assume that

$$\sum_{j\in J\cup T} a_j = \sum_{i\in I} a_i = 1.$$

Hence,

$$\begin{aligned}
\min_{j\in J\cup T}\left(\sum_{l=1}^k x_{jl}t_l\right) &\le \sum_{j\in J\cup T} a_j \sum_{l=1}^k x_{jl}t_l \\
&= \sum_{l=1}^k \sum_{j\in J\cup T} a_j x_{jl}t_l = \sum_{l=1}^k \sum_{i\in I} a_i x_{il}t_l \le \max_{i\in I}\left(\sum_{l=1}^k x_{il}t_l\right).
\end{aligned}$$

Thus, we have

$$\min_{j\in J\cup T}\left(\sum_{l=1}^{k} x_{jl}t_l\right) \le \max_{i\in I}\left(\sum_{l=1}^{k} x_{il}t_l\right)$$

and the equality holds only if

$$\sum_{l=1}^{k} x_{il}t_l \equiv c \tag{5.15}$$

for some constant c and for all $1 \le i \le n$. Since the design matrix X is of full rank, (5.15) cannot hold for any t_l with $\sum_{l=1}^{k}|t_l| > 0$. This proves (5.14).

Similarly, one can prove that (C4) implies (C2). □

PROOF OF PROPOSITION 3.2. Note that for $t_r > 0$,

$$\begin{aligned}
&\min_{j\in J\cup T}\left(-u_j + \sum_{r=2}^{y_j} t_r + \boldsymbol{x}_j'\boldsymbol{\beta}\right) + \min_{i\in I\cup T}\left(u_i - \sum_{r=2}^{y_i-1} t_r - \boldsymbol{x}_i'\boldsymbol{\beta}\right)\\
&\quad = \min_{j\in J\cup T, i\in I\cup T}\left(-u_j + u_i + \sum_{r=2}^{y_j} t_r - \sum_{r=2}^{y_i-1} t_r - \sum_{l=1}^{k}(x_{il}-x_{jl})\varepsilon_l|\beta_l|\right)\\
&\quad \le \min_{j\in J, i\in I}\left(-u_j + u_i + \sum_{r=2}^{y_j} t_r - \sum_{r=2}^{y_i-1} t_r - \sum_{l=1}^{k}(x_{il}-x_{jl})\varepsilon_l|\beta_l|\right)\\
&\quad = \min_{j\in J, i\in I}\left(-u_j + u_i - \sum_{r=2}^{L-1} t_r - \sum_{l=1}^{k}(x_{il}-x_{jl})\varepsilon_l|\beta_l|\right)\\
&\quad \le -\sum_{r=2}^{L-1} t_r + \min_{j\in J, i\in I}\left(|u_j| + |u_i| - \sum_{l=1}^{k}(x_{il}-x_{jl})\varepsilon_l|\beta_l|\right)\\
&\quad \le -\sum_{r=2}^{L-1} t_r + \min_{(i,j)\in T(\varepsilon)}\left(|u_j| + |u_i| - \sum_{l=1}^{k}(x_{il}-x_{jl})\varepsilon_l|\beta_l|\right)\\
&\quad = -\sum_{r=2}^{L-1} t_r + \min_{(i,j)\in T(\varepsilon)}\left(|u_j| + |u_i| - \sum_{l=1}^{k}|x_{il}-x_{jl}||\beta_l|\right)\\
&\quad \le -\sum_{r=2}^{L-1} t_r + \min_{(i,j)\in T(\varepsilon)}\left(|u_j| + |u_i| - \sum_{l=1}^{k}\min_{(i,j)\in T(\varepsilon)}|x_{il}-x_{jl}||\beta_l|\right)\\
&\quad \le -\sum_{r=2}^{L-1} t_r + \min_{(i,j)\in T(\varepsilon)}(|u_j| + |u_i|) - \sum_{l=1}^{k}\min_{(i,j)\in T(\varepsilon)}|x_{il}-x_{jl}||\beta_l|\\
&\quad \le -\sum_{r=2}^{L-1} t_r + \min_{(i,j)\in T(\varepsilon)}(|u_j| + |u_i|) - B(\boldsymbol{x})\sum_{l=1}^{k}|\beta_l|.
\end{aligned}$$

Now (3.2) follows from (5.6). □

PROOF OF THEOREM 3.3. Similar to (5.1), we have

$$\int_{\theta_1\le\theta_2\le\cdots\le\theta_{L-1}}\int_{R^k}\|\zeta\|^p L(\boldsymbol{\beta}\mid \boldsymbol{y})d\boldsymbol{\beta}d\theta$$
$$=\int_{R^n}\int_{\theta_1\le\theta_2\le\cdots\le\theta_{L-1}}$$
$$\cdot\int_{R^k}\|\zeta\|^p 1\{\theta_{y_i-1}+\boldsymbol{x}_i'\boldsymbol{\beta}<u_i<\theta_{y_i}+\boldsymbol{x}_i'\boldsymbol{\beta}, 1\le i\le n\}d\boldsymbol{\beta}d\theta d\boldsymbol{F}(\boldsymbol{u}).$$

It follows from the proof of (5.3) that

$$\int_{\theta_1\le\theta_2\le\cdots\le\theta_{L-1}}\int_{R^k}\|\zeta\|^p 1\{\theta_{y_i-1}+\boldsymbol{x}_i'\boldsymbol{\beta}<u_i<\theta_{y_i}+\boldsymbol{x}_i'\boldsymbol{\beta}, 1\le i\le n\}d\boldsymbol{\beta}d\theta$$
$$\le A(\boldsymbol{x},\boldsymbol{y})\max_{1\le i\le n}|u_i|^{k+L-1+p}.$$

Therefore, (3.7) holds.

As to (3.9), we have

$$\int_{\theta_1\le\theta_2\le\cdots\le\theta_{L-1}}\int_{R^k}e^{\delta_0\|\zeta\|}L(\boldsymbol{\beta}\mid \boldsymbol{y})d\boldsymbol{\beta}d\theta$$
$$=\int_{R^n}\int_{\theta_1\le\theta_2\le\cdots\le\theta_{L-1}}$$
$$\cdot\int_{R^k}e^{\delta_0\|\zeta\|}1\{\theta_{y_i-1}+\boldsymbol{x}_i'\boldsymbol{\beta}<u_i<\theta_{y_i}+\boldsymbol{x}_i'\boldsymbol{\beta}, 1\le i\le n\}d\boldsymbol{\beta}d\theta d\boldsymbol{F}(\boldsymbol{u}).$$

Along the lines of the proof of (5.3), one can obtain that there exists $A(\boldsymbol{x},\boldsymbol{y})$ such that for any $\delta>0$

$$\int_{\theta_1\le\theta_2\le\cdots\le\theta_{L-1}}\int_{R^k}e^{\delta\|\zeta\|}1\{\theta_{y_i-1}+\boldsymbol{x}_i'\boldsymbol{\beta}<u_i<\theta_{y_i}+\boldsymbol{x}_i'\boldsymbol{\beta}, 1\le i\le n\}d\boldsymbol{\beta}d\theta$$
$$\le A(\boldsymbol{x},\boldsymbol{y})\max_{1\le i\le n}|u_i|^{k+L-1}\exp\left(\delta A(\boldsymbol{x},\boldsymbol{y})\max_{1\le i\le n}|u_i|\right).$$

Letting $\delta_0=t_0/(2A(\boldsymbol{x},\boldsymbol{y}))$ yields (3.9). □

PROOF OF THEOREM 3.4. Let $\boldsymbol{\zeta}=(\boldsymbol{\beta}',\theta')'$. When $L(\boldsymbol{\zeta}\mid\boldsymbol{y})\equiv 0$, the existence of MLE is obvious. If $L(\boldsymbol{\zeta}\mid\boldsymbol{y})\not\equiv 0$, then there exists a $\boldsymbol{\zeta}_0$ such that $L(\boldsymbol{\zeta}_0\mid\boldsymbol{y})>0$. Let $M\ge 1$ such that

$$F(-M)+1-F(M)<L(\boldsymbol{\zeta}_0\mid\boldsymbol{y}).$$

We next show that there exists D such that

$$\sup_{\boldsymbol{\zeta}:\|\boldsymbol{\zeta}\|>D}L(\boldsymbol{\zeta}\mid\boldsymbol{y})<L(\boldsymbol{\zeta}_0\mid\boldsymbol{y}), \tag{5.16}$$

where $\|\boldsymbol{\zeta}\| = \max_{1\le i\le k+L-1} |\zeta_i|$, which will imply that

$$\sup_{\boldsymbol{\zeta}} L(\boldsymbol{\zeta} \mid \boldsymbol{y}) = \sup_{\boldsymbol{\zeta}:\|\boldsymbol{\zeta}\|\le D} L(\boldsymbol{\zeta} \mid \boldsymbol{y}).$$

On the other hand, $L(\boldsymbol{\zeta} \mid \boldsymbol{y})$ is a continuous function of $\boldsymbol{\zeta}$, so there exists $\hat{\boldsymbol{\zeta}}$ such that

$$L(\hat{\boldsymbol{\zeta}} \mid \boldsymbol{y}) = \sup_{\boldsymbol{\zeta}:\|\boldsymbol{\zeta}\|\le D} L(\boldsymbol{\zeta} \mid \boldsymbol{y})$$

and hence the MLE exists.

Observe that if $\min_{1\le i\le n}\{\theta_{y_i} + \boldsymbol{x}_i'\boldsymbol{\beta}\} < -M$ or $\max_{1\le i\le n}\{\theta_{y_i-1} + \boldsymbol{x}_i'\boldsymbol{\beta}\} > M$, then

$$L(\boldsymbol{\zeta} \mid \boldsymbol{y}) \le \max(F(-M), 1 - F(M)) < L(\boldsymbol{\zeta}_0 \mid \boldsymbol{y}).$$

To prove (5.16), it suffices to show that

$$\min_{1\le i\le n} \{\theta_{y_i} + \boldsymbol{x}_i'\boldsymbol{\beta}\} \ge -M \quad \text{and} \quad \max_{1\le i\le n} \{\theta_{y_i-1} + \boldsymbol{x}_i'\boldsymbol{\beta}\} \le M \tag{5.17}$$

implies

$$\|\boldsymbol{\zeta}\| \le D \quad \text{for some} \quad D < \infty. \tag{5.18}$$

Let $t_1 = \theta_1$, $t_r = \theta_r - \theta_{r-1}$ for $r = 2, \ldots, L-1$. It follows from (5.17) that

$$-M - \min_{j\in J\cup T} \left(\boldsymbol{x}_j'\boldsymbol{\beta} + \sum_{l=2}^{y_j} t_l\right) \le \beta_1 \le M - \max_{i\in I\cup T} \left(\boldsymbol{x}_i'\boldsymbol{\beta} + \sum_{l=2}^{y_i-1} t_l\right) \tag{5.19}$$

and hence

$$0 \le 2M + \min_{j\in J\cup T} \left(\boldsymbol{x}_j'\boldsymbol{\beta} + \sum_{l=2}^{y_j} t_l\right) - \max_{i\in I\cup T} \left(\boldsymbol{x}_i'\boldsymbol{\beta} + \sum_{l=2}^{y_i-1} t_l\right). \tag{5.20}$$

From the proof of (5.7) and (5.8), it is ready to see that (5.20) implies

$$\max_{2\le l\le k, 1\le r\le L-1} \max(|\beta_l|, |t_r|) \le 2M/d,$$

which together with (5.19) gives (5.18). □

Acknowledgements

The authors wish to thank the editor, associate editor and a referee for several suggestions which have greatly improved the paper.

REFERENCES

Agresti, A. (1990). *Categorical Data Analysis*, Wiley, New York.

Box, G. E. P. and Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis*, Wiley, New York.

Chen, M.-H. (1994). Importance weighted marginal Bayesian posterior density estimation, *J. Amer. Statist. Assoc.*, **89**, 818–824.

Chen, M.-H. and Dey, D. K. (1998). Bayesian modeling of correlated binary responses via scale mixture of multivariate normal link functions, *Sankhyā, Ser. A*, **60**, 322–343.

Chen, M.-H. and Shao, Q.-M. (1999*a*). Propriety of posterior distribution for dichotomous quantal response models with general link functions, *Proceedings of the American Mathematical Society* (to appear).

Chen, M.-H. and Shao, Q.-M. (1999*b*). Monte Carlo estimation of Bayesian credible and HPD intervals, *J. Comput. Graph. Statist.*, **8**, 69–92.

CPLEX Optimization (1992). *CPLEX*, CPLEX Optimization Inc, Incline Village, Nevada.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities, *J. Amer. Statist. Assoc.*, **85**, 398–409.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling, *Applied Statistics*, **41**, 337–348.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**, 97–109.

Ibrahim, J. G. and Laud, P. W. (1991). On Bayesian analysis of generalized linear models using Jeffreys's prior, *J. Amer. Statist. Assoc.*, **86**, 981–986.

Jaynes, E. T. (1982). On the rationale of maximum-entropy methods, *Proceedings of IEEE*, **70**, 939–952.

McCullagh, P. and Nelder, J. A. FRS (1989). *Generalized Linear Models*, 2nd ed., Chapman & Hall, London.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machines, *Journal of Chemical Physics*, **21**, 1087–1092.

Rossi, P. (1996). Existence of Bayes estimators for the binomial logit model, *Bayesian Analysis in Statistics and Econometrics—Essays in Honor of Arnold Zellner* (eds. D. A. Berry, K. M. Chaloner and J. K. Geweke), 91–100, Wiley, New York.

Stokes, M. E., Davis, C. S. and Koch, G. G. (1995). *Categorical Data Analysis Using the SAS System*, SAS Institute Inc., Cary, North Carolina.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion), *Ann. Statist.*, **22**, 1701–1762.

Wedderburn, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models, *Biometrika*, **63**, 27–32.

Zellner, A. (1997). Past and recent results on maximal data information priors, *Bayesian Analysis in Econometrics and Statistics: The Zellner View and Papers*, 127–148, Edward Elgar, Cheltenham.

Zellner, A. and Rossi, P. E. (1984). Bayesian analysis of dichotomous quantal response models, *J. Econometrics*, **25**, 365–393.