Journal of Multivariate Analysis **71**, 277–296 (1999)
Article ID jmva.1999.1846, available online at http://www.idealibrary.com on IDEAL®

Properties of Prior and Posterior Distributions for Multivariate Categorical Response Data Models

Ming-Hui Chen

Worcester Polytechnic Institute

and

Qi-Man Shao

University of Oregon

Received June 26, 1998

In this article, we model multivariate categorical (binary and ordinal) response data using a very rich class of scale mixture of multivariate normal (SMMVN) link functions to accommodate heavy tailed distributions. We consider both noninformative as well as informative prior distributions for SMMVN-link models. The notation of informative prior elicitation is based on available similar historical studies. The main objectives of this article are (i) to derive theoretical properties of noninformative and informative priors as well as the resulting posteriors and (ii) to develop an efficient Markov chain Monte Carlo algorithm to sample from the resulting posterior distribution. A real data example from prostate cancer studies is used to illustrate the proposed methodologies.

AMS 1991 subject classifications: primary 62A15; secondary 62H05, 62E15.

Key words and phrases: Bayesian hierarchical model; Markov chain Monte Carlo; scale mixture of multivariate normal links.

1. INTRODUCTION

There is a growing interest in the statistical literature concerning the modeling and analysis of correlated binary or ordinal response data. Prentice (1988) provided a comprehensive review of various modeling strategies using generalized linear regression to analyze correlated binary response data with covariates associated at each binary response. Following Liang and Zeger (1986) and Zeger and Liang (1986), Prentice used the generalized estimating equation (GEE) approach to obtain consistent and asymptotically normal estimators of regression coefficients. Tan, Qu, and Kutner (1997) considered model diagnostics for correlated binary response data and used the GEE approach based on latent variables to derive the

0047-259X/99 $30.00

projection (hat) matrix, Cook's distance, and various residuals. The main disadvantages of a GEE approach are that (a) it heavily relies on asymptotic theory and therefore a large sample size is required, (b) it is difficult to work with the links other than few commonly used ones such as probit and logit, and (c) it excludes available historical information.

In a Bayesian framework, Albert and Chib (1993) used latent variables to analyze binary (and polychotomous) response data, but they did not include correlation among binary (or multinomial) variables. To analyze correlated or longitudinal binary response data, Chib and Greenberg (1998) used the multivariate probit (MVP) model and Dey and Chen (in press) used the MVP and multivariate *t*-link (MVT) models along with models proposed by Prentice (1988). Other approaches to the analysis of multivariate binary response data with logistic regressions are presented by Carey, Zeger and Diggle (1993) and Glonek and McCullagh (1995). More recently, Cowles, Carlin, and Connett (1996) considered multivariate tobit (MVT) models for longitudinal ordinal response data which include correlations among the latent variables. However, they considered only the three-level ordinal responses. Chen and Dey (1998, 1996) considered general scale mixture of multivariate normal (SMMVN) link functions for analyzing longitudinal binary or correlated ordinal response data; but the categorical responses they considered are either all binary or all ordinal. Furthermore, they used noninformative proper priors for model parameters.

However, categorical data obtained from surveys are often mixed. For example, some items in a questionnaire consist of two options (e.g., "true" or "false") while some other items have more than two options (e.g., "disagree," "neutral," and "agree"). Since the same individual answers all the items in a questionnaire, these mixed categorical responses are inherently correlated. Similarly, mixed categorical data are obtained from some medical studies. For example, in a prostate cancer study, some clinical responses such as pathological seminal vesicle invasion (PSVI) are binary and some other clinical responses such as pathological extracapsular extension (PECE) and pathological positive surgical margins (PPSM) are ordinal. Therefore, it is important to simultaneously model such correlated mixed categorical response data.

In this article, we incorporate the correlation across mixed categorical responses through the covariance matrix of the underlying latent variables, which follow from scale mixture of multivariate normal distributions. In addition, we consider both noninformative and informative prior distributions. As an improper posterior makes Bayesian inference impossible, it is important to study whether the resulting posterior distributions are proper when an improper prior is used. Recently, Chen and Shao (1998, 1999a) derived the precise necessary and sufficient conditions on the propriety of the posterior distribution with an improper uniform prior for the independent binomial model as well as for the independent ordinal response model. However, the correlated mixed categorical response problem is much more complicated and challenging than the ones considered by Chen and Shao (1998, 1999a). As one of the main objectives of this article, we investigate the theoretical properties regarding the propriety of priors as well as the resulting posterior distributions in details. In addition, we present efficient computational algorithms for computing posterior properties.

The rest of the article is organized as follows. In Section 2, we propose the SMMVN-link models for correlated mixed categorical response data. Section 3 is devoted to the study of the theoretical properties of the posterior distributions with an improper uniform prior as well as the informative prior. Here, the informative prior elicitation scheme is proposed based on available similar historical studies. In Section 4, we develop an efficient Markov chain Monte Carlo (MCMC) algorithm using the Multigrid Monte Carlo (MGMC) of Liu and Sabatti (1998) to sample from the resulting posterior distributions. A real data example from the two prostate cancer studies is presented for illustrating the proposed methods in Section 5. Section 6 provides the proofs of the theorems presented in Section 3. Finally, Section 7 gives brief concluding remarks.

2. THE MODELS

We first introduce some notations which will be used throughout the remainder of this article. Suppose that on the ith observation (or subject or individual), we observe a J-dimensional mixed categorical response $Y_i = (Y_{i1}, Y_{i2}, ..., Y_{iJ})$, where the first J_b components denote binary responses, and the last $J - J_b$ components correspond to ordinal responses. Assume that Y_{ij} takes a value between 1 and L_j, where $L_j \geqslant 2$. Also let $x_{ij} = (x_{ij1}, x_{ij2}, ..., x_{ijk_j})$ be the corresponding k_j-dimensional row regression vector for $i = 1, 2, ..., n$ and $j = 1, 2, ..., J$. (Notice that x_{ij1} may be 1, which corresponds to an intercept.) Denote $Y = (Y_1, Y_2, ..., Y_n)$ and assume that $Y_{i1}, Y_{i2}, ..., Y_{iJ}$, are dependent and $Y_1, Y_2, ..., Y_n$ are independent. Let $y_i = (y_{i1}, y_{i2}, ..., y_{iJ})$ and $y = (y_1, y_2, ..., y_n)$ be the observed data. We let $D = (n, y, x)$ denote the data from the current study. Also let $\beta_j = (\beta_{j1}, \beta_{j2}, ..., \beta_{jk_j})'$ be a k_j-dimensional column vector of regression coefficients and $\beta = (\beta'_1, \beta'_2, ..., \beta'_J)'$.

In order to set up our general scale mixture of multivariate normal (SMMVN) link models for correlated mixed categorical response data, we first introduce a J-dimensional (latent) random vector $w_i = (w_{i1}, w_{i2}, ..., w_{iJ})'$ such that

$$Y_{ij} = l, \qquad \text{if} \quad \gamma_{j,\,l-1} \leqslant w_{ij} < \gamma_{jl}, \tag{2.1}$$

for $l=1, 2, ..., L_j$, where

$$-\infty=\gamma_{j0}<\gamma_{j1}=0\leqslant \cdots \leqslant \gamma_{j,L_j-1}<\gamma_{jL_j}=\infty, \tag{2.2}$$

are the cutpoints, which divide the real line into L_j intervals. We take $\gamma_{j1}=0$ to ensure identifiability. Notice that in (2.1), for $j=1, 2, ..., J_b$, $L_j=2$ and therefore, there are no unknown cutpoints for the binary responses, while for $j=J_b+1, ..., J$, we allow ordinal components to have different levels L_j. Finally, we assume

$$w_i \mid \lambda_i \sim N(x_i\beta, \kappa(\lambda_i)\,\Sigma), \tag{2.3}$$

and

$$\lambda_i \sim \pi(\lambda_i), \tag{2.4}$$

where $\kappa(\lambda_i)$ is a positive function of a one-dimensional positive-valued scale mixing variable λ_i, $\pi(\lambda_i)$ is a mixing distribution which is either discrete or continuous, $\Sigma=(\sigma_{jj'})_{J\times J}$ is a positive definite covariance matrix, and

$$x_i=\mathrm{diag}(x_{i1}, ..., x_{iJ})=\begin{pmatrix} x_{i1} & 0 & \cdots & 0 \\ 0 & x_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{iJ} \end{pmatrix}$$

for $i=1, 2, ..., n$. To ensure identifiability of the model parameters, we assume that the covariance matrix Σ is a correlation matrix. We notice that Albert and Chib (1993) first introduced the strategy involving latent Gaussian random variables that essentially makes the analysis of mixed categorical response data possible.

It is easy to observe that the distribution of w_i determines the joint distribution of Y_i through (2.1) and the covariance matrix Σ captures the correlations among $Y_{i1}, Y_{i2}, ..., Y_{iJ}$. More specifically, the joint distribution of the mixed categorical response vector Y_i can be written as

$$\begin{aligned} &P(Y_{i1}=y_{i1}, Y_{i2}=y_{i2}, ..., Y_{iJ}=y_{iJ} \mid \beta, \Sigma, \gamma, \lambda_i, x_i) \\ &\quad=\int_{A_{i1}}\int_{A_{i2}}\cdots\int_{A_{iJ}} \frac{1}{(2\pi\kappa(\lambda_i))^{J/2}\,|\Sigma|^{1/2}} \\ &\qquad\cdot \exp\left\{-\frac{\kappa-1(\lambda_i)}{2}(w_i-x_i\beta)'\,\Sigma^{-1}(w_i-x_i\beta)\right\} dw_i, \end{aligned} \tag{2.5}$$

where $\gamma=(\gamma'_{J_b+1}, \gamma'_{J_b+2}, ..., \gamma'_J)'$, $\gamma_j=(\gamma_{j1}, \gamma_{j2}, ..., \gamma_{j,L_j-1})'$, and

$$A_{ij}=[\gamma_{j,l-1}, \gamma_{jl}) \qquad \text{if} \quad y_{ij}=l \quad (1\leqslant l\leqslant L_j) \tag{2.6}$$

for $j=1, 2, ..., J$. Thus, the likelihood function is given by

$$L(\beta, \Sigma, \gamma \mid D)=\prod_{i=1}^{n} \int_{0}^{\infty}\left[\int_{A_{i1}} \int_{A_{i2}} \cdots \int_{A_{iJ}} \frac{1}{(2\pi\kappa(\lambda_i))^{J/2}\,|\Sigma|^{1/2}}\right.$$
$$\left. \cdot \exp\left\{-\frac{\kappa^{-1}(\lambda_i)}{2}(w_i-x_i\beta)'\,\Sigma^{-1}(w_i-x_i\beta)\right\} dwi\right] \pi(\lambda_i)\, d\lambda_i\,. \tag{2.7}$$

The class of SMMVN links is quite rich, which includes multivariate probit (MVP), t-link (MVT), logit (MVL), symmetric stable distribution family links (MVS), symmetric exponential and power distribution family links (MVEP) models. In the interest of space, we will give a brief explanation for MVP, MVT, and MVL as follows. Detailed discussions for the other links can be found in Chen and Dey (1998).

Taking $\kappa(\lambda_i)=1$ and the mixing distribution $\pi(\{1\})=1$, the SMMVN-link reduces to the multivariate probit, i.e., MVP. Similar to the MVP, when we take $\kappa(\lambda_i)=1/\lambda_i$ and $\lambda_i \sim \mathcal{G}(v/2, v/2)$, i.e., $\pi(\lambda_i) \propto \lambda_i^{v/2-1} \exp\{(-v/2)\,\lambda_i\}$, the SMMVN-link gives a multivariate t-link (MVT) with v degrees of freedom. Note that the special case of MVT-link with $v=1$ is termed as a multivariate Cauchy (MVC) link, and another special case of MVT-link with $v\to\infty$ is the MVP. Logistic regression is widely used to fit binary response data (e.g., see Prentice, 1988). The multivariate logit is a special case of the SMMVN-link by taking $\kappa(\lambda_i)=4\lambda_i^2$ where λ_i follows an asymptotic Kolmogorov distribution with density $\pi(\lambda_i)=\pi_K(\lambda_i)=8\sum_{k=1}^{\infty}(-1)^{k+1}k^2\lambda_i \exp\{-2k^2\lambda_i^2\}$. The MVL models are attractive since the exchangeability on the correlation structure is not required, which is advantageous compared to the random effects type of logistic regression models, for example, stratified and mixture models as given in Prentice (1988).

It should be noticed that in the class of SMMVN links, the MVP and MVC links serve as the two extremes in light of the tail behavior, that is, the MVP has the lightest tail and the MVC link has the heaviest tail, while the others such as MVEP-link, MVL, and MVS-link have heavier tails than the MVP and lighter tails than the MVC.

3. THE PRIOR DISTRIBUTIONS

In this section, we propose novel classes of non-informative and informative priors for (β, Σ, γ), and discuss some theoretical properties of the proposed priors and resulting posteriors.

3.1. *Noninformative Priors*

We first consider an improper uniform prior for (β, Σ, γ) of the form

$$\pi(\beta, \Sigma, \gamma) \propto 1, \tag{3.1}$$

where γ is subject to the constraints given in (2.2), and $vec^*(\Sigma) = (\sigma_{12}, \sigma_{13}, ..., \sigma_{J-1,J})' \in V$, and the region V is a subset of the region $[-1, 1]^{J(J-1)/2}$ that leads to a proper correlation matrix. As mentioned by Chib and Greenberg (1998) and also shown by Rousseeuw and Molenberghs (1994), the region V forms a convex solid body in the hypercube $[-1, 1]^{J(J-1)/2}$ that leads to a proper correlation matrix. Since the resulting posterior inference is driven by the likelihood, we call the improper uniform prior given by (3.1) as a noninformative prior. In certain sense, the improper uniform prior plays a similar role as a *locally uniform* prior introduced by Box and Tiao (1992, p. 23).

Using (3.1), the posterior distribution of (β, Σ, γ) based on the observed data $D = (n, y, x)$ is given by

$$p(\beta, \Sigma, \gamma \mid D) \propto L(\beta, \Sigma, \gamma \mid D), \tag{3.2}$$

where $L(\beta, \Sigma, \gamma \mid D)$ is given by (2.7). We are led to the following theorem concerning the propriety of the posterior distribution in (3.2) using the improper uniform prior (3.1).

Before presenting Theorem 3.1, we introduce the following notations. For $1 \leqslant j \leqslant J_b$, let $z_{ij} = 1$ if $y_{ij} = 1$ and $z_{ij} = -1$ if $y_{ij} = 2$. Write

$$x_{ij}^* = z_{ij} x_{ij}, \qquad \text{and} \qquad x_i^* = \text{diag}(x_{i1}^*, ..., x_{iJ_b}^*).$$

Let $1\{A\}$ denote the indicator function such that $1\{A\} = 1$ if A is true, and $1\{A\} = 0$ if A is not true. For $j = J_b + 1, ..., J$, let

$$\begin{aligned}
&\tilde{x}_{ij} = -x_{ij} 1\{2 \leqslant y_{ij} \leqslant L_j\}, \qquad \hat{x}_{ij} = x_{ij} 1\{1 \leqslant y_{ij} \leqslant L_j - 1\},\\
&\tilde{c}_{ij} = (1\{3 \leqslant y_{ij}\}, ..., 1\{L_j \leqslant y_{ij}\}),\\
&\hat{c}_{ij} = -(1\{2 \leqslant y_{ij}\}, ..., 1\{L_j - 1 \leqslant y_{ij}\}) \, 1\{1 \leqslant y_{ij} \leqslant L_j - 1\},\\
&g_{ij} = \begin{pmatrix} \tilde{x}_{ij} \\ \hat{x}_{ij} \end{pmatrix}, \qquad h_{ij} = \begin{pmatrix} \tilde{c}_{ij} \\ \hat{c}_{ij} \end{pmatrix},\\
&g_i = \text{diag}(g_{i, J_b+1}, ..., g_{i,J}), \qquad h_i = \text{diag}(h_{i, J_b+1}, ..., h_{i,J}),\\
&X^* = \begin{pmatrix} x_1^* \\ \vdots \\ x_n^* \end{pmatrix}, \qquad G = \begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix}, \qquad \text{and} \quad H = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}.
\end{aligned}$$

THEOREM 3.1. *Assume that the following conditions are satisfied*:

(C1) X^* *and* (G, H) *are of full rank*,

(C2) *There exist positive vectors* a *and* b *such that*

$$a'X^* = 0, \qquad b'G = 0, \qquad \text{and} \qquad b'H \geqslant 0,$$

(C3) $\int_0^\infty \kappa^{q/2}(\lambda)\, \pi(\lambda)\, d\lambda < \infty$, *where* $q = \sum_{j=1}^J k_j + \sum_{j=J_b+1}^J (L_j - 2)$.

Then, the posterior $p(\beta, \Sigma, \gamma \mid D)$ *given in* (3.2) *is proper, that is*,

$$\int L(\beta, \Sigma, \gamma \mid D)\, d\beta\, d\Sigma\, d\gamma < \infty. \tag{3.3}$$

Conditions (C1) and (C2) are closely related to some geometric property of the design matrix X, G and H, so called *not full-dimensional* in Natarajan and McCulloch (1995). See also Chen and Shao (1998) for further discussion. As shown in Chen and Shao (1998), conditions (C1) and (C2) are necessary for (3.3) for independent binary regression models.

It is easy to see that the moment condition (C3) holds for the MVP and MVL models; but it may not be satisfied for other models such as the MVT models with degrees of freedom less than or equal to $q/2$. Next theorem shows that the moment condition can be weaken when the design matrix X^* has some nice structures.

THEOREM 3.2. *Define*

$$X^*_{l,m} = \begin{pmatrix} x_{l+1} \\ \vdots \\ x_m \end{pmatrix}, \qquad G_{l,m} = \begin{pmatrix} g_{l+1} \\ \vdots \\ g_m \end{pmatrix}, \qquad H_{l,m} = \begin{pmatrix} h_{l+1} \\ \vdots \\ h_m \end{pmatrix}.$$

Assume that there exist $p \geqslant 1$, $0 = m_0 < m_1 < m_2 < \cdots < m_p \leqslant n$ *and positive vectors* $a_1, a_2, ..., a_p$ *and* $b_1, b_2, ..., b_p$ *such that* $X^*_{m_{l-1}, m_l}$, $(G_{m_{l-1}, m_l}, H_{m_{l-1}, m_l})$ *are of full rank and*

$$a'_l X^*_{m_{l-1}, m_l} = 0, \qquad b'_l G_{m_{l-1}, m_l} = 0 \quad \text{and} \quad b'_l H_{m_{l-1}, m_l} \geqslant 0 \quad \text{for} \quad l = 1, ..., p. \tag{3.4}$$

If

(C4) $\int_0^\infty \kappa^{q/(2p)}(\lambda)\, \pi(\lambda)\, d\lambda < \infty$

is satisfied, then (3.3) *holds.*

Clearly, condition (C4) can be satisfied even for the MVC model as long as $2p > q$. Therefore, the results presented in Theorem 3.2 are less restrictive on the moment conditions, which is useful when one wishes to use a heavy-tailed link function.

The proofs of Theorems 3.1 and 3.2 are technical and thus left to Section 6.

3.2. *Informative Priors*

Now, we consider an informative prior. Our prior construction is based on the notion of the existence of a previous study that measures the same response variable and covariates as the current study. For ease of exposition, we assume only one previous study, as the extension to multiple previous studies is straightforward. To this end, let $D_0=(n_0, y_0, x_0)$ be the data from the historical study, where $y_0=(y_{01}, y_{02}, ..., y_{0n_0})$, $y_{0i}=(y_{0i1}, y_{0i2}, ..., y_{0iJ})$, the first J_b components of y_{0i} denote binary responses, and the last $J-J_b$ components of y_{0i} correspond to ordinal responses. Denote $w_{0i}=(w_{0i1}, ..., w_{0iJ})'$ to be the latent variable vector associated with the historical study.

Let $\pi_0(\beta, \Sigma, \gamma)$ be an *initial prior* distribution for (β, Σ, γ). Notice that $\pi_0(\beta, \Sigma, \gamma)$ may be an improper uniform prior. We wish to construct a prior distribution for (β, Σ, γ) based on the historical study. To this end, we propose an informative prior of the form

$$\pi(\beta \mid \Sigma, \gamma, a_0, D_0) \propto \pi^*(\beta \mid \Sigma, \gamma, a_0, D_0)\, \pi_0(\beta, \Sigma, \gamma), \tag{3.5}$$

where

$$\begin{aligned} \pi^*(\beta \mid \Sigma, \gamma, a_0, D_0) = \prod_{i=1}^{n_0} \int_0^\infty \int_{A_{0i1}} \cdots \int_{A_{0iJ}} & \frac{a_0^{J/2}\, |\Sigma|^{-1/2}}{(2\pi\kappa(\lambda_{0i}))^{J/2}} \\ & \cdot \exp\left\{-\frac{a_0 \kappa^{-1}(\lambda_{0i})}{2} (w_{0i}-x_{0i}\beta)'\, \Sigma^{-1}(w_{0i}-x_{0i}\beta)\right\} \\ & \times \pi(\lambda_{0i})\, dw_{0i}\, d\lambda_{0i}, \end{aligned} \tag{3.6}$$

$A_{0ij}=[\gamma_{j,l-1}, \gamma_{jl})$ if $y_{0ij}=l$ for $j=1, ..., J$, and the scale mixing distribution $\pi(\lambda_{0i})$ is given in (2.4).

In (3.5), a_0 can be interpreted as a scalar prior parameter that weights the prior data relative to the likelihood of the current study. It is reasonable to restrict the range of a_0 to be between 0 and 1, and thus we take $0 \leqslant a_0 \leqslant 1$. Notice that (3.6) has several appealing interpretations. Small values of a_0 give little prior weight to the historical control data relative to the likelihood of the current study whereas values of a_0 close to 1, for example, give roughly equal weight to the prior and the likelihood of the current study. Setting $a_0=1$, (3.5) corresponds to the usual Bayesian update of $\pi_0(\beta, \Sigma, \gamma)$ via Bayes theorem. That is, with $a_0=1$, (3.6) corresponds to the posterior distribution of (β, Σ, γ) from the previous study. When $a_0 \to 0$, then the prior does not depend on the historical data, and in this case, $\pi(\beta, \Sigma, \gamma \mid a_0, D_0)$ reduces to $\pi_0(\beta, \Sigma, \gamma)$. Therefore, the prior (3.5) can be viewed as a generalization of the usual Bayesian update of $\pi_0(\beta, \Sigma, \gamma)$. The parameter a_0 allows the investigator to control the influence of the historical data on the current study. Such control is important in cases where there is heterogeneity between the previous and current study, or when the sample sizes of the two studies are quite different. In practice, it is reasonable to take a noninformative prior for $\pi_0(\beta, \Sigma, \gamma)$, such as the one described in (3.1).

The prior specification is completed by specifying a prior distribution for a_0. We take a beta prior for a_0, and thus we propose a joint prior distribution for $(\beta, \Sigma, \gamma, a_0)$ of the form

$$\pi(\beta, \Sigma, \gamma, a_0 \mid D_0) \propto \pi^*(\beta \mid \Sigma, \gamma, a_0, D_0)\, \pi_0(\beta, \Sigma, \gamma)\, a_0^{\delta_0 - 1}(1 - a_0)^{\zeta_0 - 1}, \tag{3.7}$$

where (δ_0, ζ_0) are specified prior parameters. The prior in (3.7) does not have a closed form but it has several attractive theoretical properties. First, we note that if $\pi_0(\beta, \Sigma, \gamma)$ is proper, then (3.7) is guaranteed to be proper. Further, (3.7) can be proper even if $\pi_0(\beta, \Sigma, \gamma)$ is improper. The following theorem characterizes the propriety of (3.7) when $\pi_0(\beta, \Sigma, \gamma)$ is an improper uniform prior.

THEOREM 3.3. *In additional to conditions* (C1), (C2), *and* (C4), *assume that* $\delta_0 > q/(2p)$. *Then the joint prior given in* (3.7) *is proper, that is,*

$$\int \pi^*(\beta \mid \Sigma, \gamma, a_0, D_0)\, a_0^{\delta_0 - 1}(1 - a_0)^{\zeta_0 - 1}\, d\beta\, d\Sigma\, d\gamma\, da_0 < \infty. \tag{3.8}$$

Note that the prior specification (3.7) for (β, Σ, γ) is indeed the generalization of prior specification schemes proposed by Ibrahim, Ryan, and Chen (1998) and Chen, Ibrahim, and Yiannoutsos (1999) for analyzing univariate binary response data by using logistic regression, and Chen, Manatunga, and Williams (1998) for human twin data models.

4. COMPUTATIONAL DEVELOPMENT

In this section, we only present a Markov chain Monte Carlo (MCMC) algorithm to sample from the posterior distribution with the informative prior given by (3.7) as sampling from the posterior distribution with an improper uniform prior is much simpler.

The posterior distribution for our correlated mixed categorical data model is of the form,

$$p(\beta, \Sigma, \gamma, a_0 \mid D) \propto L(\beta, \Sigma, \gamma \mid D)\, \pi(\beta, \Sigma, \gamma, a_0 \mid D_0), \tag{4.1}$$

where the prior distribution $\pi(\beta, \Sigma, \gamma, a_0 \mid D_0)$ and the likelihood $L(\beta, \Sigma, \gamma \mid D)$ are given by (3.7) and (2.7) respectively. In (4.1), we assume that $\pi_0(\beta, \Sigma, \gamma) \propto 1$.

To sample β, Σ, γ, a_0 from (4.1), we introduce several auxiliary variables. These include the latent variables $w = (w_1', w_2', ..., w_n')'$ and the mixing variables $\lambda = (\lambda_1, \lambda_2, ..., \lambda_n)'$ for the current study and $w_0 = (w_{01}', ..., w_{0n_0}')'$ and $\lambda_0 = (\lambda_{01}, ..., \lambda_{0n_0})'$ for the historical study. In addition, we let $w_{(j)} = (w_{1j}, w_{2j}, ..., w_{nj})'$ and denote $w_{(-j)}$ to be w with $w_{(j)}$ deleted for $j = 1, 2, ..., J$. Also let $w_{0(j)} = (w_{01j}, w_{02j}, ..., w_{0n_0j})'$ and denote $w_{0(-j)}$ to be w_0 with $w_{0(j)}$ deleted for $j = 1, 2, ..., J$. To sample from the posterior distribution given in (4.1), we propose a two-step MCMC sampling algorithm, which consists of *Step* 1, Regular MCMC Sampling, and *Step* 2, the MGMC Adjustment.

In Step 1, we require sampling from the conditional distributions: (i) $[\beta \mid \Sigma, w, \lambda, w_0, \lambda_0, a_0, D, D_0]$; (ii) $[\Sigma \mid \beta, w, \lambda, w_0, \lambda_0, a_0, D, D_0]$; (iii) $[w_{(j)}, w_{0(j)}, \gamma_j \mid w_{(-j)}, w_{0(-j)}, \beta, \Sigma, \lambda, \lambda_0, a_0, D, D_0]$ for $j = 1, 2, ..., J$; (iv) $[a_0 \mid \beta, \Sigma, w_0, \lambda_0, D, D_0]$; and (v) $[\lambda, \lambda_0 \mid \beta, \Sigma, w, w_0, a_0, D, D_0]$. Now, we briefly discuss how to sample from each of the above conditional distributions. Let

$$B = a_0 \sum_{i=1}^{n_0} \kappa^{-1}(\lambda_{0i})\, x_{0i}' \Sigma^{-1} x_{0i} + \sum_{i=1}^{n} \kappa^{-1}(\lambda_i)\, x_i' \Sigma^{-1} x_i$$

and

$$\hat{\beta} = B^{-1} \left(a_0 \sum_{i=1}^{n_0} \kappa^{-1}(\lambda_{0i})\, x_{0i}' \Sigma^{-1} w_{0i} + \sum_{i=1}^{n} \kappa^{-1}(\lambda_i)\, x_i' \Sigma^{-1} w_i \right).$$

Then, $[\beta \mid \Sigma, w, \lambda, w_0\ \lambda_0, a_0, D, D_0]$ is $N(\hat{\beta}, B^{-1})$. Therefore, sampling β from its conditional distribution is straightforward. To sample the correlation matrix Σ from $[\Sigma \mid \beta, w, \lambda, w_0, \lambda_0, a_0, D, D_0]$, we use a Metropolized Hit-and-Run algorithm of Chen and Dey (1998), which is a generalization of the Metropolis algorithm of Chib and Greenberg (1998). The detail for generating Σ can be found in Chen and Dey (1998).

For (iii), we use a cycle of J Gibbs steps to generate γ_j, $w_{(j)}$, and $w_{0(j)}$ jointly from their conditional distributions for $j = 1, 2, ..., J$ in turn. For $j = 1, 2, ..., J_b$, we simply draw $w_{(j)}$ and $w_{0(j)}$ from their respective conditional posterior distributions since there are no unknown cutpoints γ_j for the binary responses. For $j = J_b + 1, J_b + 2, ..., J$, we first draw γ_j from

$[\gamma_j \mid \beta, \Sigma, w_{(-j)}, w_{0(-j)}, \lambda, \lambda_0, a_0, D]$, then draw $w_{(j)}$ from $[w_{(j)} \mid \gamma_j, \beta, \Sigma, w_{(-j)}, \lambda, D]$ and draw $w_{0(j)}$ from $[w_{0(j)} \mid \gamma_j, \beta, \Sigma, w_{0(-j)}, \lambda_0, a_0, D]$. We use the algorithm of Geweke (1991) to generate $w_{(j)}$ and $w_{0(j)}$ since their respective conditional posterior distributions are truncated multivariate normals over intervals defined by (2.6). Generating γ_j from its conditional posterior distribution is a challenging problem. Cowles (1996) proposed a Hastings scheme using a truncated normal proposal distribution and Nandram and Chen (1996) developed an improved algorithm based on a Dirichlet proposal distribution. Here, we adopt the rejection algorithm of Chen and Dey (1996) to generate γ_j using a transformation technique.

The conditional posterior density of $[a_0 \mid \beta, \Sigma, w_0, \lambda_0, D, D_0]$ is not log-concave in general. However, an efficient Metropolis algorithm developed by Chen *et al.* (1999) can be directly applied. The detailed description of this Metropolis algorithm can be found in Chen *et al.* (1999) and is omitted here for brevity. Finally, we briefly discuss how to generate mixing variables λ_i and λ_{0i}. The random generation for λ_i or λ_{0i} requires the known form of the mixing distribution $\pi(\lambda)$. For a MVP model, it does not require to generate λ_i or λ_{0i} since $\pi(\{\lambda_i = 1\}) = 1$ and $\pi(\{\lambda_{0i} = 1\}) = 1$. For a MVT model, $[\lambda_i \mid \beta, \Sigma, w_i, D]$ and $[\lambda_{0i} \mid \beta, \Sigma, w_{0i}, D_0]$ are gamma distributions, which are easy to sample from. For the MVL, MVS, and MVEP link models, Chen and Dey (1998) developed various efficient Metropolis algorithms. These algorithms can be directly applied to our mixed categorical data models, and thus we omit the details.

In the MGMC adjustment step, we propose to use the MGMC scheme of Liu and Sabatti (1998) J times. More specifically, for $j = 1, 2, \ldots, J_b$ we consider the transformation $g_j(\beta_j, w_{(j)}, w_{0(j)}) = (g_j \beta_j, g_j w_{(j)}, g_j w_{0(j)})$. Following from Liu and Sabatti (1998), it is easy to show that the Jacobian of this group transformation $J_{g_j} = g_j^{n+n_0+k_j}$, the Haar measure $H(dg_j) = dg_j/g_j$, and the distribution of g_j is

$$
\begin{aligned}
p(g_j) \propto g_j^{n+n_0+k_j-1} \exp\Bigg\{ -\frac{g_j^2}{2} \Bigg[\sum_{i=1}^{n} \kappa^{-1}(\lambda_i)(w_{ij} - x_{ij}\beta_j)^2 \\
+ a_0 \sum_{i=1}^{n_0} \kappa^{-1}(\lambda_{0i})(w_{0ij} - x_{0ij}\beta_j)^2 \Bigg] \\
- \frac{g_j}{2} \Bigg[2 \sum_{j' \neq j} \sigma_{jj'} \Bigg(\sum_{i=1}^{n} \kappa^{-1}(\lambda_i)(w_{ij} - x_{ij}\beta_j)(w_{ij'} - x_{ij'}\beta_{j'}) \\
+ a_0 \sum_{i=1}^{n_0} \kappa^{-1}(\lambda_{0i})(w_{0ij} - x_{0ij}\beta_j)(w_{0ij'} - x_{0ij'}\beta_{j'}) \Bigg) \Bigg] \Bigg\}. \qquad (4.2)
\end{aligned}
$$

For $j = J_b + 1, J_b + 2, \ldots, J$, since γ_j contains $L_j - 2$ parameters, the group transformation is taken to be $g_j(\beta_j, \gamma_j, w_{(j)}, w_{0(j)}) = (g_j\beta_j, g_j\gamma_j, g_jw_{(j)}, g_jw_{0(j)})$ and its corresponding Jacobian is $J_{g_j} = g_j^{n+n_0+k_j+L_j-2}$. The distribution of g_j is similar to (4.2) with $g_j^{n+n_0+k_j-1}$ being replaced by $g_j^{n+n_0+k_j+L_j-3}$. It can be shown that $p(g_j)$ is log-concave. Thus, we can use the adaptive rejection algorithm of Gilks and Wild (1992) to sample g_j from $p(g_j)$. After we obtain a draw g_j, we then adjust $(\beta_j, \gamma_j, w_{(j)}, w_{0(j)})$ by

$$\beta_j \leftarrow g_j\beta_j, \qquad \gamma_j \leftarrow g_j\gamma_j, \quad w_{(j)} \leftarrow g_j w_{(j)}, \quad \text{and} \quad w_{0(j)} \leftarrow g_j w_{0(j)}.$$

As shown in Chen and Liu (1999), the MGMC adjustment step can dramatically improve convergence of MCMC sampling. We shall employ this adjustment in our real prostate cancer data example.

5. PROSTATE CANCER STUDY DATA EXAMPLE

To illustrate the proposed methodologies, we use two data sets, called the PENN data and the MASS data, from two prostate cancer studies conducted at the Hospital of the University of Pennsylvania in Philadelphia and Brigham and Women's Hospital in Boston, respectively. The PENN data contain 713 patients and the same pathologist was involved for all patients from 1989 to 1995. The MASS data contain the information for a prospective study of 104 patients with prostate cancer and the treatment took place between August of 1995 and April of 1996. In these two studies, three clinical categorical response variables, Pathological Seminal Vesicle Invasion (PSVI), pathological extracapsular extension (PECE), and pathological positive surgical margins (PPSM), were observed and many preoperative staging system predictors were measured. For illustrative purposes, in this example we consider only three most important predictors, which are prostate specific antigen (PSA), clinical gleason score (GLEAS), and clinical stage (CLINS). PSVI is a binary (1-2) response, and PECE and PPSM are two ordinary (1-3) responses. In the prostate cancer study, it is important to predict the outcomes of PECE, PPSM, and PSVI in order to determine whether a prostate cancer patient needs to undergo the surgery. See Desjardin (1997) for a detailed discussion.

For patient i, we let Y_{1i}, Y_{2i}, and Y_{3i} denote PSVI, PECE, and PPSM and let x_{1i}, x_{2i}, and x_{3i} be PSA, GLEAS, and CLINS. Then, Y_{1i} is binary while both Y_{2i} and Y_{3i} are ordinal and each of them has three levels. Therefore, $J = 3$, $J_b = 1$, $L_1 = 2$, and $L_2 = L_3 = 3$. Since Y_{1i}, Y_{2i}, and Y_{3i} are observed from the same patient, they are naturally correlated. Furthermore, the MASS study was conducted recently while the PENN study was done earlier. Therefore, the MASS study naturally serves as a current study while the PENN study is a historical study. The sample size of the MASS data is 103, i.e., $n = 103$, which is relatively small. In order to perform a more accurate statistical analysis, it is important to include the available historical information, i.e., the PENN data, into the analysis.

Since the logit is the most dominated link used in medical research, we present the results of our analysis for the prostate cancer studies mainly based on the multivariate logit. However, the other links in the family of SMMVN-link functions are also considered. We implement the MCMC algorithms proposed in Section 4. To ease computational burden, we standardize all three covariates. We check the convergence of the MCMC algorithm using several diagnostic procedures recommended by Cowles and Carlin (1996) and after convergence, we find that the autocorrelations among the MCMC iterations are negligible with respect to their standard deviations at lag 10. The Metropolis algorithm for generating a_0 and the Metropolized Hit-and-Run algorithm for generating Σ work well, which result in the acceptance probabilities of 0.80 and 0.22 respectively for the MVL model.

Using 50,000 MCMC iterates, we compute all the posterior quantities of interest and the results are given in Tables I and II. In Tables I and II, the highest posterior density (HPD) intervals were computed using a Monte Carlo method of Chen and Shao (1999b). By comparing Table II to Table I, we can observe that (i) all posterior standard deviations in Table II are much greater than those in Table I; (ii) in Table II, PSA is the only significant predictor for all three categorical responses. These results indicate that (i) the inference based only on the MASS data may not be accurate and (ii) when the sample size of the current study is small, it is important to incorporate the available historical information into analysis.

TABLE I

Bayesian Estimates of the Regression Coefficients with a Uniform Prior on a_0

Response variable	Covariates	Posterior Mean	Posterior Std. dev.	95% HPD Intervals
PSVI	PSA	0.194	0.055	(0.090, 0.303)
	GLEAS	0.595	0.112	(0.378, 0.814)
	CLINS	0.438	0.171	(0.102, 0.773)
PECE	PSA	0.218	0.036	(0.149, 0.292)
	GLEAS	0.147	0.039	(0.072, 0.224)
	CLINS	0.246	0.058	(0.136, 0.361)
PPSM	PSA	0.895	0.214	(0.501, 1.323)
	GLEAS	0.458	0.230	(0.009, 0.910)
	CLINS	0.385	0.329	(−0.245, 1.048)

TABLE II

Bayesian Estimates of the Regression Coefficients without Incorporating Historical Information ($a_0 = 0$)

Response variable	Covariates	Posterior Mean	Posterior Std. dev.	95% HPD Intervals
PSVI	PSA	1.729	0.724	(0.407, 3.170)
	GLEAS	3.026	1.171	(0.778, 5.389)
	CLINS	−0.883	1.022	(−3.014, 1.021)
PECE	PSA	0.519	0.208	(0.129, 0.930)
	GLEAS	0.220	0.211	(−0.176, 0.653)
	CLINS	0.344	0.206	(−0.047, 0.772)
PPSM	PSA	0.426	0.189	(0.074, 0.805)
	GLEAS	0.371	0.210	(−0.021, 0.793)
	CLINS	−0.033	0.176	(−0.397, 0.301)

6. PROOFS OF THEOREMS

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

LEMMA 6.1. *Let $\theta = (\theta_1, ..., \theta_k)'$ and $\tau = (\tau_1, ..., \tau_l)'$, M be an $n \times k$ matrix and N be an $n \times l$ matrix, where $n > k + l$. Assume that (M, N) is of full rank and that there exists a positive vector a such that*

$$a'M = 0 \qquad \text{and} \qquad a'N \geqslant 0. \tag{6.1}$$

Then there exists a constant K depending only on (M, N) such that

$$\|\eta\| \leqslant K \|u\| \tag{6.2}$$

whenever

$$(M, N)\, \eta \leqslant u \qquad \text{and} \qquad \tau \geqslant 0, \tag{6.3}$$

where $\eta = (\theta', \tau')'$ and $\|\cdot\|$ denotes the Euclidean norm.

Proof. Let $\mathscr{E} = \{\varepsilon = (\varepsilon_1, ..., \varepsilon_{k+l})' \in R^{k+l}: \varepsilon_i = \pm 1\}$. Since (M, N) is of full rank, for every $\varepsilon \in \mathscr{E}$, there is a $b_\varepsilon \in R^n$ such that

$$b'_\varepsilon (M, N) = \varepsilon'. \tag{6.4}$$

Let $a = (a_1, ..., a_n)' \in R^n$ be the positive vector satisfying (0.0). Put

$$\delta = \frac{\min_{1 \leqslant i \leqslant n} (a_i)}{2 \max_{\varepsilon \in \mathscr{E}} \|b_\varepsilon\|}.$$

For $\varepsilon = \varepsilon_\eta = \text{sign}(\eta') = (\text{sign}(\eta_1), ..., \text{sign}(\eta_{k+l}))'$, we have $\delta > 0$ and $a + \delta b_\varepsilon > 0$. Hence, it follows from (6.1) and (6.3) that

$$\begin{aligned}(a+\delta b_\varepsilon)' u &\geqslant (a+\delta b_\varepsilon)' \,(M, N)\, \eta \\ &= a' M\theta + a' N\tau + \delta b'_\varepsilon (M, N)\, \eta \\ &\geqslant \delta b'_\varepsilon (M, N)\, \eta = \delta \,\text{sign}(\eta')\, \eta \\ &\geqslant (\delta/(k+l))\, \|\eta\|,\end{aligned}$$

as desired. ▮

Proof of Theorem 3.1. Let $\lambda_1, ..., \lambda_n$ be independent random variables with the common probability density function π. Let $\tilde{w}_i = (\tilde{w}_{i1}, ..., \tilde{w}_{iJ})'$ be independent random variables such that

$$\tilde{w}_i \mid \lambda_i \sim N(0, \kappa(\lambda_i)\, \Sigma),$$

that is, given λ_i, $\tilde{w}_i$ is normally distributed with mean zero and covariance matrix $\kappa(\lambda_i)\, \Sigma$. Put

$$A_i = A_{i,1} \times A_{i,2} \times \cdots \times A_{i,J}.$$

Thus, we can rewrite the likelihood function as

$$\begin{aligned}L(\beta, \Sigma, \gamma \mid D) &= E1\{(\tilde{w}_i + x_i\beta) \in A_i, 1 \leqslant i \leqslant n\} \\ &= E(1\{\tilde{w}_{ij} + x_{ij}\beta_j \in A_{ij}, 1 \leqslant j \leqslant J_b, 1 \leqslant i \leqslant n\} \\ &\qquad \times 1\{\tilde{w}_{ij} + x_{ij}\beta_j \in A_{ij}, J_b + 1 \leqslant j \leqslant J, 1 \leqslant i \leqslant n\}). \qquad (6.5)\end{aligned}$$

It is easy to see that

$$\begin{aligned}&\{\tilde{w}_{ij} + x_{ij}\beta_j \in A_{ij}, 1 \leqslant j \leqslant J_b, 1 \leqslant i \leqslant n\} \\ &\quad = \bigcap_{1 \leqslant j \leqslant J_b} (\{\tilde{w}_{ij} + x_{ij}\beta_j < 0,\ y_{ij} = 1, 1 \leqslant i \leqslant n\} \\ &\qquad \cup \{\tilde{w}_{ij} + x_{ij}\beta_j \geqslant 0,\ y_{ij} = 2, 1 \leqslant i \leqslant n\}) \\ &\quad \subset \{z_{ij} x_{ij} \beta_j \leqslant -z_{ij}\tilde{w}_{ij}, 1 \leqslant j \leqslant J_b, 1 \leqslant i \leqslant n\} \\ &\quad = \{X^*\beta \leqslant w^*\}, \qquad (6.6)\end{aligned}$$

where

$$w^* = \begin{pmatrix} w_1^* \\ \vdots \\ w_n^* \end{pmatrix}, \qquad \text{and} \qquad w_i^* = \begin{pmatrix} w_{i1}^* \\ \vdots \\ w_{iJ_b} \end{pmatrix}.$$

To deal with $\{\tilde{w}_{ij}+x_{ij}\beta_j \in A_{ij}, J_b+1 \leqslant j \leqslant J, 1 \leqslant i \leqslant n\}$, let

$$\gamma_j = \begin{pmatrix} \gamma_{j2} \\ \gamma_{j3}-\gamma_{j2} \\ \vdots \\ \gamma_{j,L_J-1}-\gamma_{j,L_J-2} \end{pmatrix}, \qquad \eta^{(o)} = \begin{pmatrix} \beta_{J_b+1} \\ \vdots \\ \beta_J \\ \gamma_{J_b+1} \\ \vdots \\ \gamma_J \end{pmatrix}, \quad w^{(o)} = \begin{pmatrix} w_1^{(o)} \\ \vdots \\ w_n^{(o)} \end{pmatrix},$$

$$w_i^{(o)} = (-\tilde{w}_{ij}1\{1 \leqslant y_{ij} \leqslant L_j-1\}, \tilde{w}_{ij}1\{2 \leqslant y_{ij} \leqslant L_j\}, J_b+1 \leqslant j \leqslant J)'.$$

Noting that $\gamma_{j0} = -\infty$, $\gamma_{j1}=0$ and $\gamma_{jL_j} = \infty$, we have

$$\begin{aligned}
&\{\tilde{w}_{ij}+x_{ij}\beta_j \in A_{ij}, J_b+1 \leqslant j \leqslant J, 1 \leqslant i \leqslant n\} \\
&\quad = \{\tilde{w}_{ij}+x_{ij}\beta_j < \gamma_{j,y_{ij}}, 1 \leqslant y_{ij} \leqslant L_j-1, 1 \leqslant i \leqslant n\} \\
&\quad \cap \{\tilde{w}_{ij}+x_{ij}\beta_j \geqslant \gamma_{j,y_{ij}-1}, 2 \leqslant y_{ij} \leqslant L_j, 1 \leqslant i \leqslant n\} \\
&\quad \subset \Bigg\{ -\sum_{l=2}^{L_j-1} (\gamma_{jl}-\gamma_{j,l-1})\, 1\{l \leqslant y_{ij}\} \\
&\qquad + x_{ij}\beta_j \leqslant -\tilde{w}_{ij}, 1 \leqslant y_{ij} \leqslant L_j-1, 1 \leqslant i \leqslant n \Bigg\} \\
&\quad \cap \Bigg\{ \sum_{l=2}^{L_j-1} (\gamma_{jl}-\gamma_{j,l-1})\, 1\{l+1 \leqslant y_{ij}\} \\
&\qquad - x_{ij}\beta_j < \tilde{w}_{ij}, 2 \leqslant y_{ij} \leqslant L_j, 1 \leqslant i \leqslant n \Bigg\} \\
&\quad = \{\hat{c}_{ij}\gamma_j + \hat{x}_{ij}\beta_j \leqslant -\tilde{w}_{ij}1\{1 \leqslant y_{ij} \leqslant L_j-1\}, 1 \leqslant i \leqslant n\} \\
&\quad \cap \{\tilde{c}_{ij}\gamma_j + \tilde{x}_{ij}\beta_j \leqslant \tilde{w}_{ij}1\{2 \leqslant y_{ij} \leqslant L_j\}, 1 \leqslant i \leqslant n\} \\
&\quad = \{(G, H)\,\eta^{(o)} \leqslant w^{(o)}\}.
\end{aligned}$$

Let

$$U = \operatorname{diag}(X^*, (G, H)), \qquad \eta = \begin{pmatrix} \beta^* \\ \eta^o \end{pmatrix}, \quad \bar{w} = \begin{pmatrix} w^* \\ w^o \end{pmatrix}.$$

Putting (6.5)–(6.7) together yields

$$L(\beta, \Sigma, \gamma \mid D) = E(1\{U\eta \leqslant \bar{w}\}). \tag{6.8}$$

Therefore, by (C1), (C2), and Lemma 6.1

$$\begin{aligned}
\int L(\beta, \Sigma, \gamma \mid D)\, d\beta\, d\Sigma\, d\gamma &= \int E(1\{U\eta \leqslant \bar{w}\})\, d\eta\, d\Sigma \\
&\leqslant \int E(1\{\|\beta\| \leqslant K\, \|\bar{w}\|\})\, d\eta\, d\Sigma \\
&\leqslant K \int E(\max_{1 \leqslant i \leqslant n} \|\tilde{w}_i\|)^q\, d\Sigma \\
&\leqslant K \int \sum_{i=1}^{n} \sum_{j=1}^{J} E\, |\tilde{w}_{ij}|^q\, d\Sigma \\
&\leqslant K \int \sum_{j=1}^{J} E\, |\kappa(\lambda)|^{q/2}\, d\Sigma \\
&< \infty
\end{aligned} \tag{6.9}$$

by (C3). ▮

Proof of Theorem 3.2. Following the proof of Theorem 3.1, we have

$$\begin{aligned}
&\int L(\beta, \Sigma, \gamma \mid D)\, d\beta\, d\Sigma\, d\gamma \\
&\quad \leqslant \int E(1\{\|\beta\| \leqslant K \min_{1 \leqslant l \leqslant p} \max_{m_{l-1} < i \leqslant m_l} \|\tilde{w}_i\|\})\, d\eta\, d\Sigma \\
&\quad \leqslant k \int E(\min_{1 \leqslant l \leqslant p} \max_{m_{l-1} < i \leqslant m_l} \|\tilde{w}_i\|)^q\, d\Sigma \\
&\quad \leqslant K \int E\left(\prod_{1 \leqslant l \leqslant p} \max_{m_{l-1} < i \leqslant m_l} \|\tilde{w}_i\|^{q/p}\right) d\Sigma \\
&\quad \leqslant K \int \prod_{1 \leqslant l \leqslant p} E(\max_{m_{l-1} < i \leqslant m_l} \|\tilde{w}_i\|^{q/p})\, d\Sigma \\
&\quad < \infty
\end{aligned} \tag{6.10}$$

by (C4). ▮

Proof of Theorem 3.3. Let $\lambda_1, ..., \lambda_n$ be independent random variables with the common probability density function π. Let $\tilde{w}_i = (\tilde{w}_{i1}, ..., \tilde{w}_{iJ})'$ be independent random variables such that

$$\tilde{w}_i \mid \lambda_i \sim N(0, (1/a_0)\, \kappa(\lambda_i)\, \Sigma).$$

Similar to (6.8),

$$\pi^*(\beta \mid \Sigma, \gamma, a_0, D_0) = E(1\{U\eta \leqslant \bar{w}\}).$$

Therefore, following the proof of (6.10)

$$\int \pi^*(\beta \mid \Sigma, \gamma, a_0, D_0)\, a_0^{\delta_0 - 1}(1 - a_0)^{\zeta_0 - 1}\, d\beta\, d\Sigma\, d\gamma\, da_0$$

$$\leqslant K \int \prod_{1 \leqslant l \leqslant p} E\Big(\max_{m_{l-1} < i \leqslant m_l} \|\tilde{w}_i\|^{q/p}\Big)\, a_0^{\delta_0 - 1}(1 - a_0)^{\zeta_0 - 1}\, d\Sigma\, da_0$$

$$\leqslant K \int_0^1 a_0^{-q/p} a_0^{\delta_0 - 1}(1 - a_0)^{\zeta_0 - 1}\, da_0 \tag{6.11}$$

$$< \infty. \quad \blacksquare \tag{6.12}$$

7. CONCLUDING REMARKS

In this article, we provide sufficient conditions for the propriety of the informative prior as well as the resulting posterior distributions. In addition, we develop an efficient MCMC sampling algorithm for the correlated mixed categorical response models. Our results are useful for the propriety study can avoid a poor experimental design, which may result in the parameters of interest not identifiable without using a strong informative prior.

In this article, we used a real data example from prostate cancer studies to illustrate proposed methodologies. From this example, we demonstrated that (i) the proposed MCMC sampling algorithm for simulating the posterior distribution works well; and (ii) incorporating prior information leads to improved interpretation of the results of a current study.

ACKNOWLEDGMENTS

The authors thank the editor, associate editor, and two referees for several suggestions which have greatly improved the paper. The authors also thank Dr. Anthony V. D'Amico of the Joint Center for Radiation Therapy at Harvard Medical School for providing the prostate cancer data sets. Dr. Chen's research was supported by the National Science Foundation under Grant DMS-9702172 and Dr. Shao's research was supported by the National Science Foundation under Grant DMS-9802451.

REFERENCES

1. J. H. Albert and S. Chib, Bayesian analysis of binary and polychotomous response data, *J. Amer. Statist. Assoc.* **88** (1993), 669–679.
2. J. H. Albert and S. Chib, Bayesian residual analysis for binary response regression models, *Biometrika* **82** (1995), 747–759.
3. G. E. P. Box and G. C. Tiao, "Bayesian Inference in Statistical Analysis," Wiley, New York, 1992.
4. V. C. Carey, S. L. Zeger, and P. J. Diggle, Modelling multivariate binary data with alternating logistic regressions, *Biometrika* **80** (1993), 517–526.
5. M.-H. Chen and D. K. Dey, Bayesian modeling of correlated binary responses via scale mixture of multivariate normal link functions, *Sankhyā Ser. A* **60** (1998), 322–343.
6. M.-H. Chen and D. K. Dey, "Bayesian Analysis and Computation for Correlated Ordinal Data Models," Technical Report No. 96-42, Department of Statistics, University of Connecticut, Storrs, CT, 1996.
7. M.-H. Chen, J. G. Ibrahim and C. Yiannoutsos, Prior elicitation, variable selection, and and Bayesian computation for logistic regression models, *J. Roy. Statist. Soc. Ser. B* **61** (1999), 223–242.
8. M.-H. Chen and C. Liu, Discussion on Liu and Sabatti's paper, *in* "Bayesian Statistics 6" (J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Eds.), pp. 402–405, Oxford Univ. Press, Oxford, 1999.
9. M.-H. Chen, A. K. Manatunga, and C. J. Williams, Heritability estimates from human twin data by incorporating historical prior information, *Biometrics* **54** (1998), 1348–1362.
10. M.-H. Chen and Q.-M. Shao, Existence of Bayes estimators for the polychotomous quantal response models, *Ann. Inst. Statist. Math.*, 1999a, in press.
11. M.-H. Chen and Q.-M. Shao, Monte Carlo estimation of Bayesian credible and HPD intervals, *J. Comput. Graph. Statist.* **8** (1999b), 69–92.
12. M.-H. Chen and Q.-M. Shao, "Propriety of Posterior Distribution for Dichotomous Quantal Response Models with General Link Functions," Technical Report No. MS-02-98-04, Department of Mathematical Sciences, Worcester Polytechnic Institute, 1998.
13. S. Chib and E. Greenberg, Bayesian analysis of multivariate probit models, *Biometrika* **85** (1998), 347–361.
14. M. K. Cowles, Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models, *Statist. Comput.* **6** (1996), 101–111.
15. M. K. Cowles and B. P. Carlin, Markov chain Monte Carlo convergence diagnostics: A comparative review, *J. Amer. Statist. Assoc.* **91** (1996), 883–904.
16. M. K. Cowles, B. P. Carlin, and J. E. Connett, Bayesian tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness, *J. Amer. Statist. Assoc.* **91** (1996), 86–98.
17. A. M. Desjardin, "Statistical Inference on Studies of Adenocarcinoma of the Prostate," Master Thesis, Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA, 1997.
18. D. K. Dey and M.-H. Chen, Bayesian model diagnostics for correlated binary data, *in* "Generalized Linear Models: A Bayesian Perspective" (D. K. Dey, S. K. Ghosh, and B. K. Mallick, Eds.), Dekker, New York, in press.
19. J. Geweke, Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints, *in* "Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface" (E. M. Keramidas, Eds.), pp. 571–578, Interface Foundation of North America, Inc., Fairfax Station, VA, 1991.
20. W. R. Gilks and P. Wild, Adaptive rejection sampling for Gibbs sampling, *Appl. Statist.* **41** (1992), 337–348.

21. G. F. V. Glonek and P. McCullagh, Multivariate logistic models, *J. Roy. Statist. Soc. Ser. B* **57** (1995), 533–546.
22. J. G. Ibrahim, L. M. Ryan, and M.-H. Chen, Use of historical controls to adjust for covariates in trend tests for binary data, *J. Amer. Statist. Assoc.* **93** (1998), 1282–1293.
23. K.-Y. Liang and S. L. Zeger, Longitudinal data analysis using generalized linear models, *Biometrika* **73** (1986), 13–22.
24. J. S. Liu and C. Sabatti, "Generalized Multigrid Monte Carlo for Bayesian computation," Technical Report, Department of Statistics, Stanford University.
25. B. Nandram and M.-H. Chen, Accelerating Gibbs sampler convergence in the generalized linear models via a reparameterization, *J. Statist. Comput. Simulation* **54** (1996), 129–144.
26. R. L. Prentice, Correlated binary regression with covariates specific to each binary observation, *Biometrics* **44** (1988), 1033–1048.
27. P. Rousseeuw and G. Molenberghs, The shape of correlation matrices, *Amer. Statist.* **48** (1994), 276–279.
28. M. Tan, Y. Qu, and M. H. Kutner, Model diagnostics for marginal regression analysis of correlated binary data, *Comm. Statist. Ser. B* **26** (1997), 539–558.
29. S. L. Zeger and K.-Y. Liang, Longitudinal data analysis for discrete and continuous outcomes, *Biometrics* **42** (1986), 121–130.