Journal of Statistical Planning and Inference 88 (2000) 99–115

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Noninformative priors for one-parameter item response models

Malay Ghosh[a,*], Atalanta Ghosh[b], Ming-Hui Chen[c], Alan Agresti[a]

[a] *Department of Statistics, University of Florida, Gainesville, FL 32611-8545, USA*

[b] *Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IA 46285-2233, USA*

[c] *Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA 01609-2280, USA*

Received 24 March 1999; received in revised form 14 October 1999; accepted 26 October 1999

Abstract

We present a unified Bayesian approach for the analysis of one-parameter item response models. A necessary and sufficient condition is given for the propriety of posteriors under improper priors with nonidentifiable likelihoods. Posterior distributions for item and subject parameters may be improper when the sum of the binary responses for an item or subject takes its minimum or maximum possible value. When the item parameters have a flat prior but the item totals do not fall at a boundary value, we prove the propriety of the Bayesian joint posterior under some sufficient conditions on the joint (proper) distribution of the subject parameters. The methods are implemented using Markov chain Monte Carlo and illustrated with an example from a cross-over study comparing three medical treatments.

MSC: primary 62F15; secondary 62P15

Keywords: Gibbs sampling; Improper prior; Propriety; Logit model; Markov chain Monte Carlo; Probit model; Rasch model

1. Introduction

Item response models were developed in educational testing to describe how the probability of a correct answer depends on the subject's ability and the question's difficulty. Specifically, the models assume that subject i has a parameter θ_i describing that subject's ability and question j has a parameter α_j such that its negative describes its level of difficulty ($i=1,\ldots,n,\ j=1,\ldots,k$). The response X_{ij} denotes the outcome for the ith subject on the jth question, where $X_{ij}=1$ for a correct response and 0 for an incorrect response.

* Corresponding author.

E-mail address: ghoshm@stat.ufl.edu (M. Ghosh).

PII: S0378-3758(99)00201-3

Let $p_{ij} = P(X_{ij} = 1)$. The simplest and most widely quoted model for p_{ij} is the *Rasch model* (Rasch, 1961), for which

$$p_{ij} = \exp(\theta_i + \alpha_j)[1 + \exp(\theta_i + \alpha_j)]^{-1}. \tag{1}$$

This is also referred to as the *one-parameter logistic model*, since as a function of θ_i it has the form of the distribution function of a one-parameter logistic distribution with location parameter $-\alpha_j$. A related popular model is the *probit model*, $p_{ij} = \Phi(\theta_i + \alpha_j)$, where Φ is the standard normal cumulative distribution function. These models are special cases of the generalized linear model

$$F^{-1}(p_{ij}) = \theta_i + \alpha_j, \tag{2}$$

where the link function F^{-1} is the inverse of an arbitrary continuous distribution function. A systematic development of item response theory from the classical point of view owes much to the pioneering work of Lord (e.g., Lord, 1953), Rasch (1961), and their colleagues. Among the many noteworthy contributions in the same vein are Andersen (1970) and Bock and Lieberman (1970).

Though originally introduced for psychometric applications, the item response form of model (2) is increasingly being used for a variety of applications, such as repeated measurements in biomedical studies. This article uses models of this form to analyze results of a cross-over study comparing three medical treatments on a binary response. In many applications, $\boldsymbol{\alpha}$ is the parameter of interest and $\boldsymbol{\theta}$ is treated as a nuisance parameter. For example, in educational testing interest usually focuses on inference for $\{-\alpha_j\}$, the difficulty values of the questions, rather than $\{\theta_i\}$, the subject-specific parameters. There could be other instances though, for example in longitudinal models with individual frailties, when one might be interested in $\{\theta_i\}$ themselves or in making predictions of p_{ij} for specific subjects or clusters. However, the Markov chain Monte Carlo numerical integration technique that we use for implementation of the Bayesian procedure is capable of producing the posteriors of $\{\theta_i\}$ also, if needed.

Throughout this article, we use the generic term “item parameters” to refer to components of $\boldsymbol{\alpha}$. In recent years, the literature has undergone dramatic expansion in the context of generalized linear mixed models (e.g., Breslow and Clayton, 1993), which treat $\boldsymbol{\theta}$ as a random effect. The parameter $\boldsymbol{\theta}$ captures heterogeneity in the individual subjects, and is very convenient for conditionally independent hierarchical modeling.

Our objective in this paper is to introduce a unified Bayesian approach for the analysis of item response models of form (2). Prior knowledge often suggests an asymmetric treatment of the parameters in the prior distributions. In many applications, for example in IQ tests, previous studies may suggest that $\{\theta_i\}$ are approximately normal with a certain standard deviation. On the other hand, there may be little, if any, information about the distribution of item parameters, suggesting that a flat or highly diffuse prior may be sensible for them. A primary question addressed in this paper is the extent to which it is possible to use an improper prior for $\boldsymbol{\alpha}$.

Section 2 discusses the selection of priors and the propriety of posteriors under different choices of improper priors. In particular, a simple necessary and sufficient condition is given for the propriety of the posteriors under improper priors with a nonidentifiable likelihood. As a corollary to this result, it is shown that for an arbitrary link function, the choice of flat priors for both $\boldsymbol{\theta}=(\theta_1,\dots,\theta_n)$ and $\boldsymbol{\alpha}=(\alpha_1,\dots,\alpha_k)$ leads to improper posteriors. Under the same choice of priors, the joint distribution of linear contrasts in $\{\alpha_j\}$ is also improper when $\sum_{i=1}^n x_{ij}=0$ or n for some j or $\sum_{j=1}^k x_{ij}=0$ or k for some i. Once these boundary values are excluded, it is shown that for a general class of link functions corresponding to cdf's with certain finite moments, Laplace's flat prior produces proper posteriors for the contrasts. This class of links includes, but is not limited to, the logit, probit, log–log, and certain other links which are inverses of Student's t cdf's. It is also shown that when the link function is the inverse of a Cauchy cdf, the posterior is improper under Laplace's prior. Moreover, if a proper prior is assigned to $\boldsymbol{\theta}$ but a flat prior is assigned to $\boldsymbol{\alpha}$, then exclusion of the two boundary values of $\sum_{i=1}^n x_{ij}$ for all $j=1,2,\dots,k$ (but not necessarily those of $\sum_{j=1}^k x_{ij}$) along with finiteness of certain prior moments of $\boldsymbol{\theta}$ yields a proper posterior for $(\boldsymbol{\theta},\boldsymbol{\alpha})$, and a fortiori for all contrasts, under weaker moment conditions for the link function. In particular, sufficient conditions are provided for a proper posterior of $(\boldsymbol{\theta},\boldsymbol{\alpha})$ under a multivariate t-prior for $\boldsymbol{\theta}$ and flat prior for $\boldsymbol{\alpha}$.

Section 3 discusses implementation of the Bayes procedures via Markov chain Monte Carlo (MCMC) integration techniques. A general result shows how to simplify the calculations when the inverse link F satisfies the increasing failure rate (IFR) property. Section 4 illustrates some of the proposed Bayesian methods for the cross-over study.

Bayesian methods previously proposed for item response models are link-specific. For the Rasch model and its two-parameter extension, relevant work includes Birnbaum (1969), Owen (1975), Swaminathan and Gifford (1982, 1985), Leonard and Novick (1985), Mislevy and Bock (1984), Kim et al. (1994), and Tsutakawa and various co-authors listed in the references (e.g., Tsutakawa and Lin (1986), Rigdon and Tsutakawa (1983, 1987), Tsutakawa (1984), Tsutakawa and Soltys (1988), Tsutakawa and Johnson (1990)). However, these methods are primarily approximate Bayes due to analytically intractable posteriors. Albert (1992) conducted a full Bayesian analysis for the two-parameter probit model, but his parameter augmentation technique applies only to the probit link. Natarajan and McCulloch (1995) have provided necessary and sufficient conditions for the propriety of posteriors for certain generalized linear models with the logit link, but their choice of priors is different from what we have considered. One of the main objectives of this paper is to present a unified Bayesian analysis of item response models that works for a variety of link functions.

2. Choice of priors and effect on posterior propriety

We assume the general one-parameter item response model (2), where $\{X_{ij}\}$ are independent. The likelihood function is

$$L(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{x})=\prod_{i=1}^{n}\prod_{j=1}^{k}[F^{x_{ij}}(\theta_i+\alpha_j)\bar{F}^{1-x_{ij}}(\theta_i+\alpha_j)], \tag{3}$$

where $\bar{F}=1-F$ and $\boldsymbol{x}=(x_{11},\dots,x_{1k},\dots,x_{n1},\dots,x_{nk})$.

2.1. Effects of flat priors

We first consider possible prior distributions $\pi(\boldsymbol{\theta},\boldsymbol{\alpha})$ for $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$, having a corresponding posterior distribution

$$\pi(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{x}) \propto L(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{x})\pi(\boldsymbol{\theta},\boldsymbol{\alpha}).$$

When the prior information is vague, one might consider noninformative priors. The simplest, due to Laplace, is $\pi(\boldsymbol{\theta},\boldsymbol{\alpha}) \propto 1$. However, this prior leads to improper posteriors for $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$, as a consequence of the following lemma. This lemma is possibly known to many Bayesians and seems to be implicit in Dawid (1979) and O'Hagan (1994, pp. 72, 158), but the following explicit formulation is worthwhile for this article.

Lemma 1. *Suppose the parameter vector is* (ϕ_1,ϕ_2), *where* ϕ_1 *and* ϕ_2 *may be vector valued. Suppose* $\boldsymbol{X}$ *has a nonidentifiable pdf* $f(\boldsymbol{x}|\phi_1,\phi_2) = f(\boldsymbol{x}|\phi_1)$. *For a prior* $\pi(\phi_1,\phi_2)$, *the posterior* $\pi(\phi_1,\phi_2|\boldsymbol{x})$ *is proper if and only if* $\pi(\phi_1|\boldsymbol{x})$ *and* $\pi(\phi_2|\phi_1)$ *are both proper.*

This lemma follows from the expression

$$\begin{aligned}\pi(\phi_1,\phi_2|\boldsymbol{x}) &\propto \boldsymbol{f}(\boldsymbol{x}|\phi_1,\phi_2)\pi(\phi_1,\phi_2)\\ &= \boldsymbol{f}(\boldsymbol{x}|\phi_1)\pi(\phi_1)\pi(\phi_2|\phi_1)\\ &\propto \pi(\phi_1|\boldsymbol{x})\pi(\phi_2|\phi_1).\end{aligned}$$

Remark 1. Dawid (1979) calls a parameter ϕ_2 nonidentifiable if $\pi(\phi_2|\phi_1,\boldsymbol{x})=\pi(\phi_2|\phi_1)$, that is observing data $\boldsymbol{x}$ does not increase the prior knowledge about ϕ_2 given ϕ_1. Noting that

$$\pi(\phi_2|\phi_1,\boldsymbol{x}) \propto f(\boldsymbol{x}|\phi_1,\phi_2)\pi(\phi_2|\phi_1)\pi(\phi_1),$$

ϕ_2 nonidentifiable if and only if $f(\boldsymbol{x}|\phi_1,\phi_2)$ is free of ϕ_2, that is $f(\boldsymbol{x}|\phi_1,\phi_2) = f(\boldsymbol{x}|\phi_1)$. Thus, Dawid's formal definition of nonidentifiability is equivalent to the lack of identifiability of the likelihood. The present lemma lays out conditions leading to the propriety or otherwise of posteriors in the presence of nonidentifiability.

Using this lemma, we show now that Laplace's prior leads necessarily to an improper posterior in the present context.

Theorem 1. *For the likelihood function* (3) *and the prior* $\pi(\boldsymbol{\theta},\boldsymbol{\alpha}) \propto 1$, *the posterior* $\pi(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{x})$ *is improper.*

Proof. Make the one-to-one linear transformation $\eta_i = \theta_i + \alpha_k$ $(i=1,\ldots,n)$, $\xi_j = \alpha_j - \alpha_k$ $(j=1,\ldots,k-1)$. Write $\boldsymbol{\eta} = (\eta_1,\ldots,\eta_n)$, $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_{k-1})$. Then $(\boldsymbol{\theta},\boldsymbol{\alpha})$ is one to one with $(\boldsymbol{\eta},\boldsymbol{\xi},\alpha_k)$ and the likelihood function given in (3) can be rewritten as

$$L(\boldsymbol{\eta},\boldsymbol{\xi},\alpha_k|\boldsymbol{x}) = \prod_{i=1}^{n}\prod_{j=1}^{k-1}[F^{x_{ij}}(\eta_i+\xi_j)\bar{F}^{1-x_{ij}}(\eta_i+\xi_j)]\prod_{i=1}^{n}[F^{x_{ik}}(\eta_i)\bar{F}^{1-x_{ik}}(\eta_i)]. \quad (4)$$

Since the Jacobian of the transformation from $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ to $(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k)$ is a constant free of any parameter, $\pi(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k) \propto 1$. This implies $\pi(\alpha_k | \boldsymbol{\eta}, \boldsymbol{\xi}) \propto 1$, which is improper. Now apply Lemma 1 with $\phi_2 = \alpha_k$ and $\phi_1 = (\boldsymbol{\eta}, \boldsymbol{\xi})$ to conclude that $\pi(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k | \boldsymbol{x})$ is improper. Hence,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \pi(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{x}) \, d\boldsymbol{\theta} \, d\boldsymbol{\alpha} \propto \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \pi(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k | \boldsymbol{x}) \, d\boldsymbol{\eta} \, d\boldsymbol{\xi} \, d\alpha_k = \infty.$$

The nonidentifiability of the likelihood becomes apparent with the reparameterization (4). A similar nonidentifiability occurs for more complex item response models as well. The result as stated in Lemma 1 should be of use in other similar contexts. Swaminathan and Gifford (1985, p. 353) suggested that using flat priors for both $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$, the Bayesian analysis is equivalent to the likelihood-based analysis. However, since the posterior is then improper, it is not possible to find any meaningful descriptive measures such as the posterior mean and posterior quantiles, although it is still possible to find a posterior mode that is equivalent to the ML estimate. The ML estimator, however, is inconsistent. (see, e.g., Baker, 1992).

Theorem 1 raises the question of whether the subset $(\boldsymbol{\eta}, \boldsymbol{\xi})$ of $(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k)$ has a proper posterior, since the nonidentifiability of the likelihood disappears when it is expressed only in terms of $\boldsymbol{\eta}$, and $\boldsymbol{\xi}$. However, for Laplace's prior the answer continues to be negative when either *for at least one* i, the $\{x_{ij},\ j=1,\ldots,k\}$ are all zeros or 1's, or *for at least one* j, $\{x_{ij},\ i=1,\ldots,n\}$ are all zeros or all 1's. To see this, consider likelihood (4) as $L(\boldsymbol{\eta}, \boldsymbol{\xi} | \boldsymbol{x})$. Also, let $t_i = \sum_{j=1}^{k} x_{ij}$ $(i = 1,\ldots,n)$ and $y_j = \sum_{i=1}^{n} x_{ij}$ $(j = 1,\ldots,k)$. Then, the following result holds.

Theorem 2. *Suppose that* $t_i = 0$ *or* k *for at least one* i *or that* $y_j = 0$ *or* n *for at least one* j. *Then, for the prior* $\pi(\boldsymbol{\eta}, \boldsymbol{\xi}) \propto 1$, $\pi(\boldsymbol{\eta}, \boldsymbol{\xi} | \boldsymbol{x})$ *is improper.*

Proof. Suppose $t_m = 0$, that is $x_{m1} = \cdots = x_{mk} = 0$. Now,

$$\int_{-\infty}^{\infty} \prod_{j=1}^{k-1} \bar{F}(\eta_m + \xi_j) \bar{F}(\eta_m) \, d\eta_m \geqslant \int_{-\infty}^{0} \prod_{j=1}^{k-1} \bar{F}(\xi_j) \bar{F}(0) \, d\eta_m = +\infty.$$

Similarly, if $t_m = k$, that is $x_{m1} = \cdots = x_{mk} = 1$, then

$$\int_{-\infty}^{\infty} \prod_{j=1}^{k-1} F(\eta_m + \xi_j) F(\eta_m) \, d\eta_m \geqslant \int_{0}^{\infty} \prod_{j=1}^{k-1} F(\xi_j) F(0) \, d\eta_m = +\infty.$$

Also, if $y_l = 0$, that is $x_{1l} = \cdots = x_{nl} = 0$, one gets

$$\int_{-\infty}^{\infty} \prod_{i=1}^{n} \bar{F}(\eta_i + \xi_l) \, d\xi_l \geqslant \int_{-\infty}^{0} \prod_{i=1}^{n} \bar{F}(\eta_i + \xi_l) \, d\xi_l \geqslant \prod_{i=1}^{n} \bar{F}(\eta_i) \int_{-\infty}^{0} d\xi_l = +\infty.$$

Finally, if $y_l = n$, that is $x_{1l} = \cdots = x_{nl} = 1$,

$$\int_{-\infty}^{\infty} \prod_{i=1}^{n} F(\eta_i + \xi_l) \, d\xi_l \geqslant \int_{0}^{\infty} \prod_{i=1}^{n} F(\eta_i + \xi_l) \, d\xi_l \geqslant \prod_{i=1}^{n} F(\eta_i) \int_{0}^{\infty} d\xi_l = +\infty.$$

As a consequence of the above findings, the joint posterior of $(\xi_1,\dots,\xi_{k-1})=(\alpha_1-\alpha_k,\dots,\alpha_{k-1}-\alpha_k)$ given $\boldsymbol{x}$ may be improper when $t_i=0$ or k for at least one i, or $y_j=0$ or n for at least one j. It may be noted here that the conditions $0<t_i<k$ for all i and $0<y_j<n$ for each j are precisely those needed for finiteness of the ML estimates (see Wedderburn, 1976, or Dellaportas and Smith, 1993). This result contrasts, however, with the conclusions of the usual one-way normal ANOVA model considered, for example, in Gelfand and Sahu (1999). To see this, suppose $Y_1,\dots,Y_k$ are independent $\mathrm{N}(\mu+\alpha_j,1)$, $(j=1,\dots,k)$. For Laplace's prior $\pi(\mu,\alpha_1,\dots,\alpha_k)\propto 1$, the joint posterior of $(\mu,\alpha_1,\dots,\alpha_k)$ is improper, as a consequence of Lemma 1. Yet if the likelihood is parameterized only in terms of $\boldsymbol{\zeta}=(\zeta_1,\dots,\zeta_k)$, where $\zeta_j=\mu+\alpha_j$, $(j=1,\dots,k)$, and Laplace's prior is used for these parameters, then the posterior of $\boldsymbol{\zeta}$ is a product of independent normals. This implies that the joint posterior of $(\alpha_1-\alpha_k,\dots,\alpha_{k-1}-\alpha_k)=(\zeta_1-\zeta_k,\dots,\zeta_{k-1}-\zeta_k)$ is also proper. Gelfand and Sahu (1999) have also discussed sufficient conditions ensuring the propriety of posteriors in genralized linear models. They have linked the propriety of the posterior with familiar notion of estimability.

A natural question to ask here is whether the posterior of $(\boldsymbol{\eta},\boldsymbol{\xi})$ necessarily becomes proper under the prior $\pi(\boldsymbol{\eta},\boldsymbol{\xi})\propto 1$ once the boundary values of the t_i's and the y_j's are excluded. The answer turns out to be negative unless some appropriate moments of F are finite, i.e., F satisfies a certain tail smoothness condition. To see this, let F be standard Cauchy, $n=k=2$, $x_{11}=1$, $x_{12}=0$, $x_{21}=0$, and $x_{22}=1$. Then, $t_1=t_2=y_1=y_2=1$ so that both boundary values 0 and 2 are excluded. The likelihood function $L(\eta_1,\eta_2,\xi_1|\boldsymbol{x})$ is then given by

$$L(\eta_1,\eta_2,\xi_1|\boldsymbol{x})=\left[\frac{1}{2}+\frac{1}{\pi}\tan^{-1}(\eta_1+\xi_1)\right]\left[\frac{1}{2}-\frac{1}{\pi}\tan^{-1}(\eta_1)\right]\times\left[\frac{1}{2}+\frac{1}{\pi}\tan^{-1}(\eta_2)\right]\left[\frac{1}{2}-\frac{1}{\pi}\tan^{-1}(\eta_2+\xi_1)\right].$$

We shall show $\int_{-\infty}^{\infty}L(\eta_1,\eta_2,\xi_1|\boldsymbol{x})\,\mathrm{d}\xi_1=\infty$. The first step towards this is the inequality

$$\int_{-\infty}^{\infty}L(\eta_1,\eta_2,\xi_1|\boldsymbol{x})\,\mathrm{d}\xi_1\geqslant\int_{0}^{\infty}L(\eta_1,\eta_2,\xi_1|\boldsymbol{x})\,\mathrm{d}\xi_1\geqslant g(\eta_1,\eta_2)\int_0^{\infty}\left[\frac{1}{2}-\frac{1}{\pi}\tan^{-1}(\eta_2+\xi_1)\right]\mathrm{d}\xi_1,$$

where

$$g(\eta_1,\eta_2)=\left[\frac{1}{2}+\frac{1}{\pi}\tan^{-1}(\eta_1)\right]\left[\frac{1}{2}-\frac{1}{\pi}\tan^{-1}(\eta_1)\right]\left[\frac{1}{2}+\frac{1}{\pi}\tan^{-1}(\eta_2)\right].$$

Hence, it suffices to show that

$$\mathrm{Int}\overset{\mathrm{def}}{=}\int_0^{\infty}\left[\frac{1}{2}-\frac{1}{\pi}\tan^{-1}(\eta_2+\xi_1)\right]\mathrm{d}\xi_1=\infty.$$

To this end, we first use the transformation $\tan^{-1}(\eta_2+\xi_2)=\frac{1}{2}(\pi-z)$. Then, "Int" simplifies to

$$\begin{aligned}
\text{Int} &= \int_0^{\pi-2\tan^{-1}(\eta_2)} \left(\frac{1}{2}z\right)\frac{1}{2}\sec^2((\pi-z)/2)\,\mathrm{d}z \\
&= \frac{1}{4}\int_0^{\pi-2\tan^{-1}(\eta_2)} \frac{z}{\cos^2((\pi-z)/2)}\,\mathrm{d}z \\
&= \frac{1}{4}\int_0^{\pi-2\tan^{-1}(\eta_2)} \frac{z}{\frac{1}{2}[1+\cos(\pi-z)]}\,\mathrm{d}z = \frac{1}{2}\int_0^{\pi-2\tan^{-1}(\eta_2)} \frac{z}{1-\cos z}\,\mathrm{d}z \\
&= \frac{1}{2}\int_0^{\pi-2\tan^{-1}(\eta_2)} \frac{z}{2\sin^2\left(\frac{1}{2}z\right)}\,\mathrm{d}z = \frac{1}{2}\int_0^{\pi-2\tan^{-1}(\eta_2)} \frac{(z/2)^2(4/z)}{2\sin^2\left(\frac{1}{2}z\right)}\,\mathrm{d}z \\
&\geqslant \int_0^{\pi-2\tan^{-1}(\eta_2)} \frac{\mathrm{d}z}{z} = \infty.
\end{aligned}$$

Thus, exclusion of the boundary values of $\{t_i\}$ and $\{y_j\}$, by itself, is not sufficient to guarantee the propriety of posteriors. We now prove a theorem which shows that under some sufficient moment conditions on F, the posterior is indeed proper.

Theorem 3. *Suppose* (i) $0<t_i<k$ *for all* i, (ii) $0<y_j<n$ *for all* j, *and* (iii) $\int_{-\infty}^{\infty}|z|^{n+k-1}\,\mathrm{d}F(z)<\infty$. *Then, under the prior* $\pi(\boldsymbol{\eta},\boldsymbol{\xi})\propto 1$, $\pi(\boldsymbol{\eta},\boldsymbol{\xi}|\boldsymbol{x})$ *is proper.*

Proof. Let $\boldsymbol{\beta}=(\eta_1,\eta_2,\dots,\eta_n,\xi_1,\xi_2,\dots,\xi_{k-1})^{\mathrm{T}}$. We write $u_{(i-1)k+j}=x_{ij}$ ($i=1,2,\dots,n$, $j=1,2,\dots,k$), $z_{(i-1)k+j}$ as the vector with 1 in places i and $n+j$, and zeros elsewhere for $j=1,2,\dots,k-1$, while z_{ik} is the vector with 1 in place i and zeros elsewhere. Each $z_{(i-1)k+j}$ is a vector with $n+k-1$ components. Let $\boldsymbol{Z}^{\mathrm{T}}=(z_1,\dots,z_k,\dots,z_{(n-1)k+1},\dots,z_{nk})$. Then it is easy to check that $F(\eta_i+\xi_j)=F(z_{(i-1)k+j}^{\mathrm{T}}\boldsymbol{\beta})$, and $\operatorname{rank}(\boldsymbol{Z})=n+k-1$. Write $w_{(i-1)k+j}=1-2x_{ij}$ and $\boldsymbol{Z}^{*\mathrm{T}}=(w_1z_1,\dots,w_kz_k,\dots,w_{(n-1)k+1}z_{(n-1)k+1},\dots,w_{nk}z_{nk})$. Since $1\leqslant t_i\leqslant k-1$ for all i and $1\leqslant y_j\leqslant n-1$ for all j, there exists $\boldsymbol{a}>\boldsymbol{0}$ such that $\boldsymbol{Z}^{*\mathrm{T}}\boldsymbol{a}=\boldsymbol{0}$.

Next, we show that there exists a constant K depending only on $\boldsymbol{Z}^*$ such that

$$||\boldsymbol{\beta}||\leqslant K||\boldsymbol{u}|| \tag{5}$$

whenever

$$\boldsymbol{Z}^*\boldsymbol{\beta}\leqslant\boldsymbol{u}, \tag{6}$$

where $\boldsymbol{u}=(u_1,u_2,\dots,u_{nk})^{\mathrm{T}}$, $||\boldsymbol{\beta}||=(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta})^{1/2}$, and $||\boldsymbol{u}||=(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u})^{1/2}$. Let

$$\epsilon=\operatorname{sign}(\boldsymbol{\beta})=(\operatorname{sign}(\eta_1),\dots,\operatorname{sign}(\eta_n),\operatorname{sign}(\xi_1),\dots,\operatorname{sign}(\xi_{k-1}))^{\mathrm{T}},$$

where the sign function takes the value 1 if its argument is nonnegative and is -1 otherwise. Since $\operatorname{rank}(\boldsymbol{Z}^*)=\operatorname{rank}(\boldsymbol{Z})=n+k-1$, there exists a $\boldsymbol{b}_\epsilon\in R^{nk}$ such that

$$\boldsymbol{b}_\epsilon^{\mathrm{T}}\boldsymbol{Z}^*=\epsilon^{\mathrm{T}}.$$

Let

$$\delta = \frac{\min_{1\leqslant l\leqslant nk}(a_l)}{2||\boldsymbol{b}_\epsilon||}.$$

Then $\delta > 0$ and $\boldsymbol{a} + \delta\boldsymbol{b}_\epsilon > \boldsymbol{0}$. Hence, by (6), one gets

$$(\boldsymbol{a} + \delta\boldsymbol{b}_\epsilon)^{\mathrm{T}}\boldsymbol{u} \geqslant (\boldsymbol{a} + \delta\boldsymbol{b}_\epsilon)^{\mathrm{T}}\boldsymbol{Z}^*\boldsymbol{\beta}$$
$$= \delta\boldsymbol{b}_\epsilon^{\mathrm{T}}\boldsymbol{Z}^*\boldsymbol{\beta} = \delta\,\mathrm{sign}(\boldsymbol{\beta})\boldsymbol{\beta} \geqslant \delta||\boldsymbol{\beta}||.$$

This proves (5).

Let $\mathrm{d}\boldsymbol{F}(\boldsymbol{u}) = \mathrm{d}(w_1F(u_1))\,\mathrm{d}(w_2F(u_2))\ldots\mathrm{d}(w_{nk}F(u_{nk}))$. It is easy to show that

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^{k-1}} L(\boldsymbol{\eta},\boldsymbol{\xi}|\boldsymbol{x})\,\mathrm{d}\boldsymbol{\eta}\,\mathrm{d}\boldsymbol{\xi} = \int_{\mathbb{R}^{n+k-1}} L(\boldsymbol{\beta}|\boldsymbol{x})\,\mathrm{d}\boldsymbol{\beta} = \int_{\mathbb{R}^{nk}}\int_{\mathbb{R}^{n+k-1}} 1_{[\boldsymbol{Z}^*\boldsymbol{\beta}\leqslant\boldsymbol{u}]}\,\mathrm{d}\boldsymbol{\beta}\,\mathrm{d}\boldsymbol{F}(\boldsymbol{u}),$$

where $1_{[\boldsymbol{Z}^*\boldsymbol{\beta}\leqslant\boldsymbol{u}]}$ is the usual indicator function. It follows from (5) and (6) that

$$\int_{\mathbb{R}^{n+k-1}} 1_{[\boldsymbol{Z}^*\boldsymbol{\beta}\leqslant\boldsymbol{u}]}\,\mathrm{d}\boldsymbol{\beta} \leqslant K||\boldsymbol{u}||^{n+k-1}.$$

Thus, assumption (iii) along with the above inequality directly yields the propriety of the posterior distribution.

Remark 2. Assumption (iii) of Theorem 3 automatically holds for any distribution with finite moments. Thus, the theorem applies for the logit, probit and log–log links, the corresponding distributions being logistic, normal and extreme valued. Also, the theorem is applicable to t-distributions with degrees of freedom exceeding $n + k - 1$.

2.2. Proper prior for subject parameters

As mentioned in the introduction, often it would be natural to assign a proper prior to $\boldsymbol{\theta}$ and a flat prior to $\boldsymbol{\alpha}$. We next investigate conditions under which such priors lead to a proper posterior for $(\boldsymbol{\theta},\boldsymbol{\alpha})$ and a fortiori for all independent linear contrasts in $\boldsymbol{\alpha}$. In this regard, it is straightforward to show that an invariance property holds for the posterior means of the elementary contrasts $\{\alpha_j - \alpha_m,\ 1\leqslant j \neq m\leqslant k\}$ for general link functions and the location-scale family of priors $g_{\mu,\sigma}(\boldsymbol{\theta}) = \sigma^{-n}g((\theta_1 - \mu)/\sigma,\ldots,(\theta_n - \mu)/\sigma)$ for $\boldsymbol{\theta}$ and the independent flat prior for $\boldsymbol{\alpha}$. The joint posterior of the contrasts $\{\alpha_j - \alpha_m\}$ (and hence their Bayes estimates) is location invariant, not depending on μ. Thus, for inference about the elementary contrasts for this prior, one can take $\mu = 0$ without loss of generality.

For this prior structure, to study conditions that lead to a proper posterior, we first need a slight modification of Theorem 3. Specifically, suppose $\boldsymbol{z}_i^{\mathrm{T}} = (\boldsymbol{z}_{i1}^{\mathrm{T}},\boldsymbol{z}_{i2}^{\mathrm{T}})$, where $\boldsymbol{z}_{i1}$ and $\boldsymbol{z}_{i2}$ are respectively of dimensions k_1 and k_2 $(k_1 + k_2 = k)$. Then, writing $\boldsymbol{\beta}^{\mathrm{T}} = (\boldsymbol{\beta}_1^{\mathrm{T}},\boldsymbol{\beta}_2^{\mathrm{T}})$, the likelihood function is given by

$$L(\boldsymbol{\beta}|\boldsymbol{u}) = \prod_{i=1}^{n} [F^{u_i}(\boldsymbol{z}_{i1}^{\mathrm{T}}\boldsymbol{\beta}_1 + \boldsymbol{z}_{i2}^{\mathrm{T}}\boldsymbol{\beta}_2)\bar{F}^{1-u_i}(\boldsymbol{z}_{i1}^{\mathrm{T}}\boldsymbol{\beta}_1 + \boldsymbol{z}_{i2}^{\mathrm{T}}\boldsymbol{\beta}_2)],$$

where each u_i is either 0 or 1. Consider the prior $\pi(\boldsymbol{\beta}) \propto g(\boldsymbol{\beta}_1)$, where $g(\boldsymbol{\beta}_1)$ is a proper prior. We want to find conditions under which the posterior $\pi(\boldsymbol{\beta}|\boldsymbol{u})$ is proper.

Let $\boldsymbol{Z}_2 = (z_{12}, \ldots, z_{n2})^{\mathrm{T}}$, $w_i = 1 - 2u_i$, $i = 1, \ldots, n$, $\boldsymbol{Z}_2^{*\mathrm{T}} = (w_1 z_{12}, \ldots, w_n z_{n2})$, and $\boldsymbol{Z}_1^{*\mathrm{T}} = (w_1 z_{11}, \ldots, w_n z_{n1})$. Following Theorem 3, we assume that

(a) $\boldsymbol{Z}_2$ has rank k_2;
(b) there exists $\boldsymbol{a} = (a_1, \ldots, a_n)^{\mathrm{T}}$, $a_i > 0$ for each i such that $\boldsymbol{Z}_2^{*\mathrm{T}} \boldsymbol{a} = \boldsymbol{0}$;
(c) $\int_{\mathbb{R}^n} ||\boldsymbol{u}||^{k_2} \,\mathrm{d}F(\boldsymbol{u}) < \infty$, where $\mathrm{d}F(\boldsymbol{u}) = \mathrm{d}(w_1 F(u_1), w_2 F(u_2), \ldots, w_n F(u_n))$;
(d) $\int_{\mathbb{R}^{k_1}} ||\boldsymbol{\beta}_1||^{k_2} g(\boldsymbol{\beta}_1)\,\mathrm{d}\boldsymbol{\beta}_1 < \infty$.

Then the following theorem holds.

Theorem 4. *Assume* (a)–(d). *Then*, $\pi(\boldsymbol{\beta}|\boldsymbol{u})$ *is proper.*

Remark 3. Before proving this theorem, we see some of its implications in the present context. Writing $\boldsymbol{\beta}_1 = \boldsymbol{\theta}$, and $\boldsymbol{\beta}_2 = \boldsymbol{\alpha}$, it follows that if $\boldsymbol{\theta}$ has a proper prior, but $\boldsymbol{\alpha}$ has a flat prior, assuming that $y_j = \sum_{i=1}^n x_{ij}$ does not attain its boundary values for each j, the joint posterior continues to be proper. The $t_i = \sum_{j=1}^k x_{ij}$ can take any values including its boundaries 0 or k. This is particularly useful when the number of items is small, for instance, in the example considered by us in Section 4, where $k = 3$ and $t_i = 0$ or 3 for 14 subjects.

Proof of Theorem 4. Following the proof of Theorem 3, write

$$L(\boldsymbol{\beta}|\boldsymbol{y}) = \int_{\mathbb{R}^n} 1_{[\boldsymbol{Z}_1^* \boldsymbol{\beta}_1 + \boldsymbol{Z}_2^* \boldsymbol{\beta}_1 \leqslant \boldsymbol{u}]} \,\mathrm{d}F(\boldsymbol{u}).$$

Now, using (5) and (6), we have

$$\begin{aligned}
&\int_{\mathbb{R}^{k_1}} \int_{\mathbb{R}^{k_2}} L(\boldsymbol{\beta}|\boldsymbol{y}) g(\boldsymbol{\beta}_1)\,\mathrm{d}\boldsymbol{\beta}_1\,\mathrm{d}\boldsymbol{\beta}_2 \\
&\quad = \int_{\mathbb{R}^n} \int_{\mathbb{R}^{k_1}} \left\{ \int_{\mathbb{R}^{k_2}} 1_{[\boldsymbol{Z}_1^* \boldsymbol{\beta}_1 + \boldsymbol{X}_2^* \boldsymbol{\beta}_2 \leqslant \boldsymbol{u}]} \,\mathrm{d}\boldsymbol{\beta}_2 \right\} g(\boldsymbol{\beta}_1)\,\mathrm{d}\boldsymbol{\beta}_1\,\mathrm{d}F(\boldsymbol{u}) \\
&\quad \leqslant K \int_{\mathbb{R}^n} \int_{\mathbb{R}^{k_1}} ||\boldsymbol{u} - \boldsymbol{Z}_1^* \boldsymbol{\beta}_1||^{k_2} g(\boldsymbol{\beta}_1)\,\mathrm{d}\boldsymbol{\beta}_1\,\mathrm{d}F(\boldsymbol{u}),
\end{aligned} \tag{7}$$

where K (> 0) is a generic constant. But

$$||\boldsymbol{u} - \boldsymbol{Z}_1^* \boldsymbol{\beta}_1||^{k_2} \leqslant 2^{k_2} [||\boldsymbol{u}||^{k_2} + ||\boldsymbol{Z}_1^* \boldsymbol{\beta}_1||^{k_2}].$$

Hence, it suffices to find conditions under which

$$I = \int_{\mathbb{R}^n} \int_{\mathbb{R}^{k_1}} [||\boldsymbol{u}||^{k_2} + ||\boldsymbol{X}_1^* \boldsymbol{\beta}_1||^{k_2}] g(\boldsymbol{\beta}_1)\,\mathrm{d}\boldsymbol{\beta}_1\,\mathrm{d}F(\boldsymbol{u}) < \infty.$$

But

$$I = \int_{\mathbb{R}^n} ||\boldsymbol{u}||^{k_2} \,\mathrm{d}F(\boldsymbol{u}) + \int_{R^{k_1}} ||\boldsymbol{Z}_1^* \boldsymbol{\beta}_1||^{k_2} g(\boldsymbol{\beta}_1)\,\mathrm{d}\boldsymbol{\beta}_1. \tag{8}$$

The first term on the right-hand side of (8) is finite because of (c). Then, writing $\lambda_1(\boldsymbol{Z}_1^{*\mathrm{T}}\boldsymbol{Z}_1^*)$ as the largest eigenvalue of $\boldsymbol{Z}_1^{*\mathrm{T}}\boldsymbol{Z}_1^*$, the second term on the right-hand side of (8) equals

$$\int_{\mathbb{R}^{k_1}} [\boldsymbol{\beta}_1^{\mathrm{T}}\boldsymbol{Z}_1^{*\mathrm{T}}\boldsymbol{Z}_1^*\boldsymbol{\beta}_1]^{k_2/2} g(\boldsymbol{\beta}_1)\,\mathrm{d}\boldsymbol{\beta}_1 \leqslant [\lambda_1(\boldsymbol{Z}_1^{*\mathrm{T}}\boldsymbol{Z}_1^*)]^{k_2/2} \int_{\mathbb{R}^{k_1}} ||\boldsymbol{\beta}_1||^{k_2} g(\boldsymbol{\beta}_1)\,\mathrm{d}\boldsymbol{\beta}_1$$

which is finite due to (d). This proves the theorem.

2.3. Hierarchical approach with normal subject prior

In practice, the most common way to apply the previous result would take $\{\theta_i\}$ to be iid $\mathrm{N}(0,\sigma^2)$. However, there would often be no obvious choice for σ. One could, alternatively, use a hierarchical Bayesian approach in which σ^{-2} has a gamma $(a/2, b/2)$ distribution, that is, *pdf* proportional to $(\sigma^2)^{-(1/2)b-1}\exp(-a/2\sigma^2)$, for fixed $a(>0)$ and $b(>0)$. This is equivalent to assuming the multivariate-t prior for $\boldsymbol{\theta}$,

$$\Pi_1(\boldsymbol{\theta}) \propto \left(a + \sum_{i=1}^{n} \theta_i^2\right)^{-(1/2)n-(1/2)b-1}. \tag{9}$$

It is generally agreed (e.g., Berger, 1985) that the t-priors are more robust than the normal prior. One can check directly that conditions (a) and (b) of the theorem hold. Once again, the finiteness of the kth moment of the link function ensures (c). Finally, in this case, condition (d) holds if

$$\int_{\mathbb{R}^n} ||\boldsymbol{\theta}||^k (a + ||\boldsymbol{\theta}||^2)^{-(1/2)n-(1/2)b-1}\,\mathrm{d}\boldsymbol{\theta} < \infty,$$

which holds if and only if $k - n - b - 2 + k - 1 < -1$, that is $n > 2k - b - 2$. This condition easily holds when the number of subjects is much larger than the number of items, which is almost always the case in practice. Also, it is interesting to notice the analogy of this condition with condition (c) in Theorem 1 of Hobert and Casella (1996).

3. Implementation of Bayes procedures

Consider first the general model (2) and the prior $\pi(\boldsymbol{\theta},\boldsymbol{\alpha}) \propto \prod_{i=1}^n g_1(\theta_i)\prod_{j=1}^k g_2(\alpha_j)$. Both g_1 and g_2 can be improper as long as the posterior $\pi(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{x})$ remains proper. However, due to nonconjugacy of the prior, the posterior is analytically intractable, and can be found only via numerical integration. Also, direct numerical integration seems infeasible because of the high dimensionality of the problem. Fortunately, the integration task has become easier due to the advent of Markov chain Monte Carlo techniques. We shall use, in particular, Gibbs sampling.

Gibbs sampling consists of finding the conditional distribution of every parameter given the remaining parameters and the data. In this case the full conditionals are given by

$$\pi(\theta_i|\theta_l\ (l \neq i), \boldsymbol{\alpha}, \boldsymbol{x}) \propto \prod_{j=1}^{k} [F^{x_{ij}}(\theta_i + \alpha_j)\bar{F}^{1-x_{ij}}(\theta_i + \alpha_j)]g_1(\theta_i), \tag{10}$$

$$\pi(\alpha_j|\alpha_m\ (m \neq j), \boldsymbol{\theta}, \boldsymbol{x}) \propto \prod_{i=1}^{n} [F^{x_{ij}}(\theta_i + \alpha_j)\bar{F}^{1-x_{ij}}(\theta_i + \alpha_j)]g_2(\alpha_j), \tag{11}$$

$i = 1,\dots,n$; $j = 1,\dots,k$. Note that the full conditional of θ_i does not involve the remaining $\{\theta_l,\ l \neq i\}$, and the full conditional of α_j does not involve the remaining $\{\alpha_m,\ m \neq j\}$). The full conditionals, however, are non-standard densities from which it is not possible to draw samples directly. The general procedure for generating samples in such cases is to use the Metropolis-Hastings accept–reject algorithm. If, however, F, $\bar{F}$, g_1 and g_2 are all log-concave, the full conditionals $\pi(\theta_i|\cdot)$ and $\pi(\alpha_j|\cdot)$ are all log-concave and one can then use the adaptive rejection sampling algorithm of Gilks and Wild (1992).

To see the log-concavity of $\pi(\theta_i|\cdot)$ and $\pi(\alpha_j|\cdot)$, simply write

$$\log \pi(\theta_i|\theta_l\ (l \neq i), \boldsymbol{\alpha}, \boldsymbol{x}) = \sum_{j=1}^{k} [x_{ij} \log F(\theta_i + \alpha_j) + (1 - x_{ij})\log \bar{F}(\theta_i + \alpha_j)] + \log g_1(\theta_i). \tag{12}$$

If F, $\bar{F}$ and g_1 are all log-concave then clearly $\pi(\theta_i|\theta_l\ (l \neq i), \boldsymbol{\alpha}, \boldsymbol{x})$ is log-concave. Similarly, if F, $\bar{F}$ and g_2 are log-concave, $\pi(\alpha_j|\alpha_m\ (m \neq j), \boldsymbol{\theta}, \boldsymbol{x})$ is log-concave. The log-concavity of F and $\bar{F}$ is ensured if F is an IFR df.

The full conditionals $\pi(\theta_i|\theta_l\ (l \neq i), \boldsymbol{\alpha}, \boldsymbol{x})$ $(i = 1,\dots,n)$ and $\pi(\alpha_j|\alpha_m\ (m \neq j), \boldsymbol{\theta}, \boldsymbol{x})$ $(j = 1,\dots,k)$ can all be proper, and yet the posterior $\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{x})$ can be improper (cf. Casella and George, 1992). For the Rasch model, if one assigns the flat prior $\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}) \propto 1$, then from Theorem 1, $\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{x})$ is improper. But, it can be shown after some algebra that the full conditionals $\pi(\theta_i|\theta_l\ (l \neq i), \boldsymbol{\alpha}, \boldsymbol{x})$, and $\pi(\alpha_j|\alpha_m,\ (m \neq j), \boldsymbol{\theta}, \boldsymbol{x})$ are all proper.

The multivariate t-prior given in (9) is not log-concave, so direct application of the algorithm discussed so far is not possible. However, recognizing the hierarchical structure of the t-prior, one can proceed essentially as before. To see this, rewrite the prior given in (9) as

$$[\theta|R = r] \sim \mathrm{N}(0, r^{-1}\boldsymbol{I}_n), \tag{13}$$

$$f(r) \propto \exp(-\tfrac{1}{2}ar)r^{(1/2)b-1}. \tag{14}$$

This parameter augmentation or introduction of R simplifies the MCMC technique (see e.g. Wakefield et al., 1994). One now has the full conditional

$$\pi(\theta_i|\theta_\ell\ (\ell \neq i), \boldsymbol{\alpha}, r, x) \propto \prod_{j=1}^{k} [F^{x_{ij}}(\theta_i + \alpha_j)\bar{F}^{1-x_{ij}}(\theta_i + \alpha_j)] \exp\left(-\frac{r}{2}\theta\right), \tag{15}$$

$$\pi(\alpha_j|\alpha_m\ (m \neq j), \boldsymbol{\theta}, r, \boldsymbol{x}) \propto \prod_{i=1}^{n} [F^{x_{ij}}(\theta_i + \alpha_j)\bar{F}^{1-x_{ij}}(\theta_i + \alpha_j)], \tag{16}$$

Table 1
Cell counts for cross-over study comparing treatments for relief of primary dysmenorrhea

Treatment B	Treatment C	Treatment A	
		Relief	No relief
Relief	Relief	8	45
Relief	Norelief	4	4
No relief	Relief	7	9
No relief	No relief	3	6

$$\pi(r|\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{x}) \propto \exp\left[-\frac{1}{2}r\left(a+\sum_{i=1}^{n}\theta_i^2\right)\right] r^{(1/2)(n+b)-1}. \tag{17}$$

Hence, F and $\bar{F}$ are both log-concave, so that the full conditionals given in (13) and (14) are also log-concave. Finally, the full conditional given in (15) is Gamma, and it is easy to generate samples from the same.

4. An example

We illustrate the Bayesian approaches to item response models using Table 1, previously analyzed by Jones and Kenward (1987) and Agresti (1993). The data result from a three-period cross-over trial designed to compare placebo (treatment A) with a low-dose analgesic (treatment B) and high-dose analgesic (treatment C) for relief of primary dysmenorrhea. At the end of each period, each subject rated the treatment as giving either some relief (1) or no relief (2). Let p_{ij} denote the probability of relief for subject i using treatment j ($j=A,B,C$). We estimate the treatment effects for the logit, probit, and log–log versions of model (2). Our interest focuses specifically on the posterior means and standard deviations of treatment differences $\{\alpha_j-\alpha_k\}$. Following the results and discussion at the end of Section 2, we use a flat prior for $\boldsymbol{\alpha}$ and a $\mathrm{N}(0,\sigma^2)$ prior for subject parameters $\boldsymbol{\theta}$ with inverse gamma prior for σ^2. By Theorem 4, the posterior is then proper, even though $t_i=0$ for 6 subjects and $t_i=k=3$ for 8 subjects.

Since the main interest is in estimating $\boldsymbol{\alpha}$, we regard $\boldsymbol{\theta}$ as a nuisance parameter, its presence in the model serving as a mechanism for inducing the dependence in the repeated measurements. Because of this we have presented the data in the form of Table 1 rather than listing the individual $\{X_{ij}\}$. Generally, the counts in the 2^k cross-classification of the repeated responses is sufficient for estimating $\boldsymbol{\alpha}$, as is seen by marginalizing (3) over $\boldsymbol{\theta}$. In this sense, the analysis presented here is analogous to repeated measurement methods for k-way correlated binary response data. However, even though we regard $\boldsymbol{\theta}$ primarily as a nuisance, it may be of interest to estimate the variability of $\boldsymbol{\theta}$ as a means of describing the degree of subject heterogeneity.

We calculated the posterior moments using the Gibbs sampler (Geman and Geman, 1984, Gelfand and Smith, 1990), using 1000 iterates for a burn-in and then taking

Table 2
Bayes treatment comparison estimates (standard errors in parentheses) for logit model with Table 1, using a variety of parameters (a, b) for t priors

Parameters for t Prior						
a	0.001	0.010	0.100	1.0	3.0	2.0
b	0.0	0.0	0.0	0.0	5.0	4.0
$\alpha_B - \alpha_A$	2.08	2.09	2.11	2.19	2.18	2.16
	(0.38)	(0.38)	(0.38)	(0.39)	(0.38)	(0.38)
$\alpha_C - \alpha_A$	2.62	2.64	2.66	2.75	2.74	2.72
	(0.42)	(0.43)	(0.43)	(0.43)	(0.41)	(0.41)
$\alpha_C - \alpha_B$	0.54	0.55	0.55	0.56	0.56	0.56
	(0.37)	(0.38)	(0.38)	(0.38)	(0.38)	(0.38)

every fifth iterate until generating 100,000 samples. Convergence of the Gibbs sampler was checked using the Gelman and Rubin (1992) algorithm. We estimate that the simulation error for the posterior estimates of $\{\alpha_j - \alpha_k\}$ is within 10% of the reported posterior standard error (e.g., within 0.04 when the reported standard error is 0.42).

We first consider the logit link. Table 2 displays the Bayes estimates of the treatment effects, with the standard deviations in parentheses. In order to use a relatively flat prior for σ^2, we chose (a, b) for its inverse gamma prior close to 0. As a check on the sensitivity of results to the particular selection of (a, b), Table 2 shows results for the variety of settings $(a, b) = (0.001, 0)$, $(0.01, 0)$, $(0.1, 0)$, $(1.0, 0)$. For comparison, we also obtained results for the very flat prior $(a, b) = (2.0001, 4.0001)$, for which $E(\sigma^2) = 1$ and $\mathrm{Var}(\sigma^2) = 20,000$, and for the more informative prior $(a, b) = (3, 5)$, for which $E(\sigma^2) = 1$ and $\mathrm{Var}(\sigma^2) = 2$. Results are relatively insensitive to the choice, and regardless of the prior we conclude that treatments B and C are substantially better than placebo, with only mild evidence that C is better than B.

The usual competitors of the Bayes estimates of $\boldsymbol{\alpha}$ are the maximum likelihood (ML) estimates for the mixed model treating $\boldsymbol{\theta}$ as a random effect. These are often referred to as marginal maximum likelihood estimators, since they are calculated by integrating the likelihood with respect to the assumed normal distribution of the random effect $\boldsymbol{\theta}$ and maximizing the resulting "marginal likelihood" of $\boldsymbol{x}$ with respect to $\boldsymbol{\alpha}$ and σ^2. The standard errors for estimates of the $\{\alpha_j - \alpha_k\}$ are obtained from the estimated inverse information matrix. Alternatively, for a fixed effects formulation with the logit link, one can use a conditional maximum likelihood approach, eliminating θ_i $(i = 1, \ldots, n)$ by conditioning on the sufficient statistics $T_i = \sum_{j=1}^{k} x_{ij}$, $i = 1, \ldots, n$. Andersen (1970) showed that the ML estimate of $\boldsymbol{\alpha}$ based on this likelihood is consistent.

Table 3 shows the marginal ML estimates and the conditional ML estimates. These are also the posterior modes for the Bayesian approach using a flat prior for $\boldsymbol{\alpha}$. Though the estimates differ somewhat from the Bayes estimates, the substantive conclusions are the same. For instance, with the conditional ML approach, we estimate the subject-specific odds of relief for the high-dose analgesic as $\exp(0.589) = 1.80$ times the odds of relief for the low-dose analgesic; this odds ratio estimate is $\exp(0.509) = 1.66$

Table 3
Marginal maximum likelihood and conditional maximum likelihood treatment comparisons for logit model with Table 1

Parameter	Conditional ML		Marginal ML	
	Estimate	Std. Error	Estimate	Std. Error
$\alpha_B - \alpha_A$	1.641	0.338	1.960	0.343
$\alpha_C - \alpha_A$	2.230	0.388	2.469	0.367
$\alpha_C - \alpha_B$	0.589	0.393	0.509	0.360

for the marginal ML approach and exp(0.54) = 1.72 for the hierarchical Bayes approach with $(a,b) = (0.001, 0.0)$. It is interesting to note that the marginal ML estimate of σ for the $\mathrm{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$ random effects distribution of $\boldsymbol{\theta}$ equals 0; that is, this is an unusual instance where the data do not provide evidence of within-subject dependence. The model with $\sigma = 0$, corresponding to subject homogeneity, is equivalent to one treating the three responses for a subject as if they came from three independent subjects.

These marginal ML estimates are similar to what one obtains with a Bayes approach using a degenerate normal prior distribution (mean = variance = 0) for each θ_i; in that case, the posterior means are 1.99 for $\alpha_B - \alpha_A$ (std. error = 0.35), 2.51 for $\alpha_C - \alpha_A$ (std. error = 0.37), and 0.52 for $\alpha_C - \alpha_B$ (std. error = 0.36). In fact, even with a relatively flat prior for σ, the posterior does not provide evidence of especially large variability in the θ_i. For instance, consider the t prior for $\boldsymbol{\theta}$ with $(a,b) = (0.001, 0)$. Then the posterior mean and standard deviation of θ_i equals (0.21, 0.55) for the 8 subjects who have relief with all three treatments, (0.04, 0.42) for the 56 subjects who have relief with two treatments, $(-0.13, 0.47)$ for the 16 subjects who have relief with only one treatment, and $(-0.29, 0.63)$ for the 6 subjects who have relief with none of the treatments. Hence with either the marginal ML or the Bayes approach a model is plausible that treats the subjects as homogeneous. Nonetheless, for model building we feel it safer to use the full model that permits the possibility of heterogeneity.

Table 4 shows the Bayes estimates for the probit and log–log links, using a variety of priors. The numerical values for treatment effects are not comparable to those in Tables 1 and 2, since the scales differ for these links. The probit link function corresponds to a cdf with a standard deviation of 1.0, whereas the logistic link function corresponds to a cdf with a standard deviation of $\pi/\sqrt{3} = 1.81$ and the log–log link function corresponds to a cdf with a standard deviation of $\pi/\sqrt{6} = 1.28$. It follows that when the different link functions provide decent fits, the estimates with the logit link should be on the order of 1.8 times those with probit link and on the order of $\sqrt{2} = 1.4$ times those with the log–log link. In fact, making this translation with the results in Table 4 provides comparable results to those in Table 2. In particular, the same substantive results occur as with the logit link, with the estimated effect being about 5 or 6 standard errors for the comparison of treatments B and C with A, and about 1.5 standard errors for the comparison of treatments C and B.

For Table 1, one can also consider generalizations of model (2) that take into account the sequence in which a subject receives the treatments. Other analyses of these

Table 4
Bayes treatment comparison estimates (standard errors in parentheses) for probit and log–log links with Table 1, using a variety of parameters (a, b) for t priors

Parameters for t Prior				
a	0.001	0.010	0.100	1.0
b	0.0	0.0	0.0	0.0
Probit link				
$\alpha_B - \alpha_A$	1.24	1.25	1.27	1.32
	(0.21)	(0.21)	(0.21)	(0.22)
$\alpha_C - \alpha_A$	1.55	1.56	1.58	1.65
	(0.22)	(0.22)	(0.22)	(0.23)
$\alpha_C - \alpha_B$	0.31	0.31	0.31	0.33
	(0.21)	(0.21)	(0.22)	(0.22)
Log–log link				
$\alpha_B - \alpha_A$	1.43	1.44	1.45	1.51
	(0.26)	(0.26)	(0.26)	(0.27)
$\alpha_C - \alpha_A$	1.89	1.90	1.91	1.98
	(0.29)	(0.29)	(0.30)	(0.30)
$\alpha_C - \alpha_B$	0.46	0.46	0.46	0.47
	(0.32)	(0.32)	(0.32)	(0.32)

data have considered potential treatment-by-sequence interaction, such as period or carry-over effects. Such analyses do not show any need for more complex modeling. See, for instance, Agresti (1993), who used a loglinear modeling approach that has treatment parameter estimates necessarily identical to the conditional ML estimates. We do not pursue these models here.

5. Concluding remarks

This article has introduced a unified Bayesian analysis of item response models. We have not considered many important questions, such as how to check model adequacy or how to choose the link function for the model. One way to choose the link uses the Bayes factors of one model relative to another to guide in model selection. Another way uses an adaptive link function, estimating the link from the data as in Mallick and Gelfand (1994).

The techniques used here have the potential of extension to more complex models, such as two-parameter item response models that contain “discrimination parameters” in addition to subject and item parameters. Also possible are extensions to multinomial responses, such as cumulative logit models for ordinal responses.

Acknowledgements

The authors thank Professors A. O’Hagan, A.P. Dawid and J.K. Ghosh for useful discussions leading to Lemma 1. Thanks are also due to two referees for their constructive comments. This research was partly supported by NSF grants SBR-9423996, SBR-9810968, DMS-9702172, and SBR-9730136, and NIH grant GM43824.

References

Agresti, A., 1993. Distribution-free fitting of logit models with random effects for repeated categorical responses. Statist. Med. 12, 1969–1987.

Albert, J.H., 1992. Bayesian estimation of normal ogive item response curve using Gibbs sampling. J. Educational Statist. 17, 251–269.

Andersen, E.B., 1970. Asymptotic properties of conditional maximum likelihood estimators. J. Roy. Statist. Soc. Ser. B 32, 283–301.

Baker, F.B., 1992. Item Response Theory. Marcel Dekker, New York.

Berger, J., 1985. Statistical Decision Theory and Bayesian Analysis, 2nd ed. Springer, New York.

Bock, R.D., Lieberman, M., 1970. Fitting a response model for *n* dichotomously scored items. Psychometrika 35, 179–197.

Birnbaum, A., 1969. Statistical theory for logistic mental test models with a prior distribution of ability. J. Math. Psychol. 6, 258–276.

Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. J. Amer. Statist. Assoc. 88, 9–25.

Casella, G., George, E.I., 1992. Explaining the Gibbs sampler. Amer. Statist. 46, 167–174.

Dawid, A.P., 1979. Conditional independence in statistical theory (with discussion). J. Roy. Statist. Soc. Ser. B 41, 1–31.

Dellaportas, P., Smith, A.F.M., 1993. Bayesian inference for generalised linear and proportional hazards model via Gibbs sampling. Appl. Statist. 42, 443–459.

Gelfand, A.E., Sahu, S.K., 1999. Identifiability, improper priors and Gibbs sampling for generalized linear models. J. Amer. Statist. Assoc. 94, 247–253.

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

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statist. Sci. 7, 457–511.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Trans. Pattn. Anal. Mach. Intel. 6, 721–741.

Gilks, W.R., Wild, P., 1992. Adaptive rejection sampling for Gibbs sampling. Appl. Statist. 41, 337–348.

Hobert, J.P., Casella, G., 1996. The effect of improper priors in Gibbs sampling in hierarchical linear mixed models. J. Amer. Statist. Assoc. 91, 1461–1474.

Jones, B., Kenward, M.G., 1987. Modeling binary data from a three-period cross-over trial. Statist. Med. 6, 555–564.

Kim, S., Cohen, A.S., Baker, F.B., Subkoviak, M.J., Leonard, T., 1994. An investigation of hierarchical Bayes procedures in item response theory. Psychometrika 59, 405–421.

Leonard, T., Novick, M.R., 1985. Bayesian inference and diagnostics for the three parameter logistic model. ONR Technical Report 85-5. The University of Iowa, CADA Research Group, Iowa City, IA.

Lord, F.M., 1953. The relation of test score to the trait underlying the test. Educational Psychol. Measurement 13, 517–548.

Mallick, B.K., Gelfand, A.E., 1994. Generalized linear models with unknown link functions. Biometrika 81, 237–245.

Mislevy, R.J., Bock, R.D., 1984. BILOG Maximum Likelihood Item Analysis and Test Scoring: Logistic Model. Scientific Software, Mooresville, ID.

Natarajan, R., McCulloch, C.E., 1995. A note on the existence of the posterior distribution for a class of mixed models for binomial responses. Biometrika 82, 639–643.

O'Hagan, A., 1994. Kendall's Advanced Theory of Statistics. Volume 2B: Bayesian Inference. Edward Arnold, New York.

Owen, R., 1975. A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. J. Amer. Statist. Assoc. 70, 351–356.

Rasch, G., 1961. On general laws and the meaning of measurement in psychology. Proceedings of the fourth Berkeley Symposium on Mathematical Statistics, Vol. 4. University of California Press, Berkeley, pp. 321–334.

Rigdon, S.E., Tsutakawa, R.K., 1983. Parameter estimation in latent trait models. Psychometrika 48, 567–574.

Rigdon, S.E., Tsutakawa, R.K., 1987. Estimation for the Rasch model when both ability and difficulty parameters are random. J. Educational Statist. 12, 76–86.

Swaminathan, H., Gifford, J.A., 1982. Bayesian estimation in the Rasch model. J. Educational Statist. 7, 175–192.

Swaminathan, H., Gifford, J.A., 1985. Bayesian estimation in the two-parameter logistic model. Psychometrika 50, 349–364.

Tsutakawa, R.K., 1984. Estimation of two-parameter logistic item response curves. J. Educational Statist. 9, 263–276.

Tsutakawa, R.K., Lin, H.Y., 1986. Bayesian estimation of item response curves. Psychometrika 51, 251–267.

Tsutakawa, R.K., Soltys, M.J., 1988. Approximation for Bayesian ability estimation. J. Educational Statist. 13, 117–130.

Tsutakawa, R.K., Johnson, J.C., 1990. The effect of uncertainty of item parameter estimation on ability estimates. Psychometrika 55, 371–390.

Wakefield, J.C., Smith, A.F.M., Racine-Poon, A., Gelfand, A.E., 1994. Bayesian analysis of linear and non-linear population models by using the Gibbs sampler. Appl. Statist. 43, 201–221.

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