The Canadian Journal of Statistics
Vol. 28, No. 4, 2000, Pages 715–730
La revue canadienne de statistique

Univariate and multirater ordinal cumulative link regression with covariate specific cutpoints

Hemant ISHWARAN

Key words and phrases: Bayesian hierarchical model; hybrid Monte Carlo; leapfrog algorithm; ordinal regression; random walk Metropolis–Hastings; staging analysis.

AMS 1991 subject classifications: Primary 62F15; secondary 62H99.

ABSTRACT

The author considers a reparameterized version of the Bayesian ordinal cumulative link regression model as a tool for exploring relationships between covariates and "cutpoint" parameters. The use of this parameterization allows one to fit models using the leapfrog hybrid Monte Carlo method, and to bypass latent variable data augmentation and the slow convergence of the cutpoints which it usually entails. The proposed Gibbs sampler is not model specific and can be easily modified to handle different link functions. The approach is illustrated by considering data from a pediatric radiology study.

RÉSUMÉ

L'auteur propose une nouvelle paramétrisation du modèle de régression ordinale bayésien à lien cumulatif dont il se sert pour explorer la relation entre des covariables et des "points de coupure." Cette reparamétrisation permet d'ajuster les modèles par une méthode de Monte-Carlo à saute-mouton modifiée, évitant ainsi le besoin d'augmentation de données de la variable latente et la lenteur de convergence des points de coupure qui en découle souvent. L'échantillonneur de Gibbs qui est proposé n'est pas spécifique au modèle et peut être adapté facilement à d'autres fonctions de lien. La méthode est illustrée au moyen d'une étude de radiologie pédiatrique.

1. INTRODUCTION

This paper discusses a class of Bayesian models suitable for analyzing ordinal categorical data that can be either univariate, in the case where there is a single ordinal response observed per individual in a study, or more generally for multivariate data, in the case where repeated measurements are collected for each individual. Multivariate ordinal data occur in many research settings, one of the most popular being the multirater experimental design, where a collection of "raters" each rate individuals in a study on the basis of an ordinal categorical scale. See Johnson & Albert (1998, Chapter 5) for related discussion. The models proposed here for the analysis of such data will be based on a cumulative link regression approach in which the cutpoints are reparameterized to allow covariates to be category specific. Cutpoint covariate specific models for univariate ordinal data have been studied in Fahrmeir & Tutz (1994, Section 3.3), who appear to be the first to have noticed the method of reparameterizing the ordinal regression model. This same parameterization has also been exploited by Albert & Chib (1998) in a Bayesian context for fitting univariate data with non-category-specific models. Also see Hedeker & Mermelstein (1998) and Hedeker, Mermelstein & Weeks (1999) for other approaches to modeling covariate-category specific models. Here, we consider the parameterization introduced by Fahrmeir & Tutz (1994, Chapter 3.3) and present a range of reparameterized Bayesian ordinal regression models, suitable for both univariate and multirater data, beginning with simpler versions that allow cutpoints to depend upon covariates uniformly in location, and uniformly in location and scale, and then moving to more general models that permit individual cutpoints to be covariate specific. The availability of models with differing levels of complexity is important in data analysis, allowing us to explore different scenarios for covariates to depend upon categories. For example, in the analysis of ordinal categorical data from diagnostic radiology studies,

such flexibility allows one to test for specific differences in the performance of diagnostic tools. This point will be further amplified in Section 4, where a staging analysis of a pediatric study will be based on models that test for differences in the way CT (computed tomography) and MR (magnetic resonance) interpret a specific stage of cancer.

A special case of the models to be considered is the Bayesian ordinal regression model presented in Albert & Chib (1993). However, our approach will bypass the use of latent variable data augmentation altogether (Albert & Chib 1993, 1998), with model fitting implemented instead through the use of a hybrid Monte Carlo (HMC) Gibbs sampling method. A key to using this method is the parameterization used for cutpoints, which, along with the selection of priors, ensures a key condition of a differentiable posterior, positive everywhere with respect to Lebesgue measure. The HMC method explored here is the "leapfrog" algorithm described by Duane, Kennedy, Pendleton & Roweth (1987), which has had a successful history of MCMC (Markov chain Monte Carlo) applications in statistical physics. See, for example, Duane & Kogut (1986), Duane, Kennedy, Pendleton & Roweth (1987) and Neal (1994). The leapfrog algorithm has also been recently applied successfully in statistics. See Neal (1996), Gustafson (1997) and Ishwaran (1999) for examples.

1.1. HMC versus data augmentation.

A special feature of the Albert & Chib (1993) data augmentation procedure is that it leads to a Gibbs sampler that utilizes exact updates for the location parameter in the probit ordinal regression model. Although this is a key attribute, and one of the reasons for the widespread popularity of this approach, a disadvantage arising from data augmentation is that it introduces as many nuisance latent variables as the number of observations (which can become quite large in multirater problems), and the simulation of a large number of latent variables can slow convergence of the Markov chain. However, an even more serious concern with the Albert & Chib (1993) approach is that the original method used for sampling the cutpoints is very inefficient, with high and often times non-vanishing autocorrelations. This problem has prompted some recent studies into methods for accelerating convergence for the cutpoints. For example, Cowles (1996) discusses the use of a multivariate Metropolis–Hastings step which updates cutpoints and latent variables simultaneously, while Nandram & Chen (1996) and Albert & Chib (1998) have used Metropolis–Hastings applied to reparameterized cutpoints with data augmentation.

The ordinal regression model in which cutpoints are covariate specific are even more challenging to fit, and thus conventional MCMC methods may be inappropriate in this setting. As a competitive approach, we have found that HMC applied within a non-data-augmentation scheme works quite well, generally promoting efficient mixing of the Markov chain. Another advantage with bypassing latent variable data augmentation is that our HMC Gibbs sampler is fairly generic and can easily be modified to accommodate fundamental changes to the regression model. For example, switching between a probit link to a logistic link regression model simply corresponds to changing the two lines in the program code that define the link function and its derivative. In contrast, switching between link functions with data augmentation requires fundamental changes to the Gibbs sampling algorithm (for example, it will no longer be possible to sample the location parameters exactly).

The layout of the paper is as follows. Section 2 presents the proposed class of models, while Section 3 discusses the selection of priors and the HMC Gibbs sampling method. Section 4 illustrates the method by considering data from a pediatric radiology study, and in Section 5, the method is applied to simulated data. A discussion and summary are given in Section 6.

2. CUMULATIVE LINK REGRESSION MODELS

In the description of our models, we will focus on the multiple repeated measurement problem, with the univariate (or single response problem) derived in an obvious way from this more general case. In the multirater setting, we observe ordinal response data $Y_{i,j}$, where $i = 1, \ldots, n$ indexes the n different individuals in our study and $j = 1, \ldots, J$ indexes the multiple readings

for an individual i from our J raters. The $Y_{i,j}$ are assumed to be ordinal categorical in nature, taking values from some ordered scale $\{r_1, \ldots, r_K\}$, where for notational convenience, we code $r_k = k$. In particular, this assumes that r_k is increasing in magnitude from $k = 1$ to $k = K$. Notationally, the multivariate nature of the data can be emphasized by writing $Y_i = (Y_{i,1}, \ldots, Y_{i,J})'$ for the vector of responses for individual i and $Y = (Y_1, \ldots, Y_n)'$. The case where not all raters will rate each individual i, i.e., when the dimension of Y_i varies with i, is easily handled with simple modifications. See Section 4 for illustration.

An important application giving rise to multirater data is the diagnostic radiology study, where, typically, the ordinal response $Y_{i,j}$ is the degree of suspicion elicited by a diagnostic reader j to patient i's diagnostic scan. In many instances, the study design may involve the use of several diagnostic tools, e.g., CT (computed tomography) or MR (magnetic resonance). In this case, $Y_{i,j}$ will represent the degree of suspicion elicited by reader j (who may be either a CT or MR reader) to a diagnostic scan from patient i (which can either be a CT or MR image). In the diagnostic study, the ordinal response $Y_{i,j}$ is recorded on a fixed ordinal scale that is typically encoded by a set of verbal descriptors enumerating the different categories of severity of disease. This type of data usually lends itself to an ROC (Receiver Operating Characteristic) analysis, where the cumulative ordinal regression model for $Y_{i,j}$ is typically based on patient clinical covariates, including the true disease status of the patient. See Hanley & McNeil (1982), Tosteson & Begg (1988), Thompson & Zucchini (1989), Gatsonis (1995), Obuchowski (1995) and Ishwaran & Gatsonis (2000) for related discussion. Another popular cumulative ordinal regression model for the analysis of diagnostic data is the predictive model for the true disease status. In this case, the ordinal response is the true disease status, while covariates will include the radiologists' degree of suspicion. In Section 4, an example of this nature will be considered in which the ordinal response is the patients' true stage of cancer recorded on the 4-point TNM scale.

2.1. General models.

In the cumulative regression model, along with the responses $Y_{i,j}$, there is also covariate information

$$(X_{i,j}, U_{i,j}, V_{i,j}, W_{i,j})$$

available for each individual i and for each repeated measurement j. We will use covariates $X_{i,j}$ for location parameters, $U_{i,j}$ for location random effects parameters, $V_{i,j}$ for scale parameters and $W_{i,j}$ for cutpoint parameters. Our ordinal regression model is based on a model for the cumulative distribution function (cdf) for $Y_{i,j}$, which is defined as

$$\mathrm{P}\{Y_{i,j} \le k \mid \eta_{i,j}(k)\} = F\{\eta_{i,j}(k)\}, \qquad k = 1, \ldots, K, \tag{1}$$

where F is a differentiable cdf with a density f that is positive everywhere with respect to Lebesgue measure. In (1)

$$\eta_{i,j}(k) = \mu_{i,j} + \theta_{i,j}(k),$$

where

$$\mu_{i,j} = X'_{i,j}\beta_0 + U'_{i,j}\beta_j \tag{2}$$

is a mixed effects location parameter, where β_0 is used to model fixed effects parameters and β_j are random effects parameters for raters. The expression for $\eta_{i,j}(k)$ also includes "cutpoint" parameters $\theta_{i,j}(k)$ that are defined by

$$\theta_{i,j}(k) = \exp(V'_{i,j}\alpha_0) \sum_{\ell=1}^{k} \exp(\gamma_\ell + W'_{i,j}\alpha_\ell), \tag{3}$$

where γ_ℓ are also parameters that we refer to as cutpoints and α_ℓ are covariate specific cutpoint parameters. However, we will interpret α_0 as a scale parameter. The reason for this distinction will be clarified in the following sections.

Note that the definition for the cutpoints $\theta_{i,j}(k)$ in (3) ensures that the probabilities defined by (1) are increasing in k. However, to ensure that the cdf for $Y_{i,j}$ is properly defined, we set

$$\gamma_K = +\infty, \qquad \alpha_K = 0$$

to guarantee that (1) is equal to one when $k = K$. The following section will also argue that constraining $\gamma_1 = -\infty$ is generally needed for proper model identification. However, other than these restrictions, the cutpoints γ_ℓ and α_ℓ are generally left unrestricted except for constraints used to test specific models, as we now discuss in the next few sections.

2.2. Location mixed effects models.

In the case where $\alpha_0 = 0$ and $\alpha_k = 0$ for $k = 1, \ldots, K$, the model defined by (1)–(3) reduces to

$$\mathrm{P}\{Y_{i,j} \le k \mid \eta_{i,j}(k)\} = F(X'_{i,j}\beta_0 + U'_{i,j}\beta_j + \theta^*_k), \qquad k = 1, \ldots, K,$$

where

$$\eta_{i,j}(k) = X'_{i,j}\beta_0 + U'_{i,j}\beta_j + \theta^*_k \qquad \text{and} \qquad \theta^*_k = \sum_{\ell=1}^{k} \exp(\gamma_\ell).$$

With the selection of a standard normal cdf for F, this model is similar to the probit ordinal regression model considered by Albert & Chib (1998) for studying univariate data, who also take advantage of reparameterizing the ordered cutpoints

$$\theta^*_1 \le \theta^*_2 \le \cdots \le \theta^*_K = +\infty$$

in terms of the unconstrained γ_ℓ parameters.

In the location model, the usual strategy to ensure identification is to set the first cutpoint θ^*_1 to zero and then designate a location parameter to act as an intercept term. See Albert & Chib (1993). This same effect is accomplished in our parameterization by setting $\gamma_1 = -\infty$. With this constraint, a particularly useful location model for multirater data is described by setting the random effects β_j to act as random intercept terms. More precisely,

$$\mathrm{P}\{Y_{i,j} \le k \mid \eta_{i,j}(k)\} = F(X'_{i,j}\beta_0 + \beta_j + \theta^*_k), \tag{4}$$

where $\gamma_1 = -\infty$ ($\theta^*_1 = 0$) to ensure identification and $\gamma_K = +\infty$ ($\theta^*_K = +\infty$) to ensure a properly defined cdf.

The model described above is especially useful in a diagnostic radiology study involving J different diagnostic readers who are believed to be interpreting diagnostic scans in roughly the same manner. In this scenario, β_j represents a reader random intercept term, with each parameter assumed to be derived from a common random effects population to emphasize the homogeneity between readers. The model can also be extended to accommodate the case when readers are heterogeneous, but form homogeneous groups. For example, in a study involving both CT and MR readers, we can group readers by modality with random intercept terms sampled from one of two random effects populations for CT and MR. More complex models for accommodating differences in readers are discussed in Section 2.4.

2.3. Location-scale models.

The effect of a covariate $X_{i,j}$ on the cdf for $Y_{i,j}$ in the location model (4) is restricted to translations of the distribution function F, which is a limitation with this type of model. Absorbing its value into each of the cutpoints, we see that the effect of a specific covariate $X_{i,j}$ is to uniformly translate the cutpoints by a constant. In particular, the location model with a logistic link function $F(x) = \{1 + \exp(-x)\}^{-1}$ produces what is commonly referred to as the proportional-odds model. The use of the location model in this case means that the log-odds ratio of the event

$\mathrm{P}\{Y_{i,j} \leq k \mid \eta_{i,j}(k)\}$ to the event $\mathrm{P}\{Y_{i,j} \leq k \mid \eta^*_{i,j}(k)\}$ equals $(X_{i,j} - X^*_{i,j})'\beta_0$, which is independent of the category k.

One method for extending this model is to permit the cutpoints to be location as well as scale transformed by covariates. This is the location-scale model, which is another special case of the model (1)–(3) and is similar to (4), but with α_0 left unconstrained:

$$\mathrm{P}\{Y_{i,j} \leq k \mid \eta_{i,j}(k)\} = F\left\{X'_{i,j}\beta_0 + \beta_j + \exp(V'_{i,j}\alpha_0)\theta^*_k\right\}. \tag{5}$$

In this case, the covariate $V_{i,j}$ uniformly scales each cutpoint by a constant. This scaling effect is why α_0 is referred to as a scale parameter.

2.4. Covariate specific cutpoint models.

The location-scale model is an improvement over the standard location model, but it is still limited because the effect of a covariate can only be uniform over the categories. For example, with a logistic link function, the log-odds ratio of $\mathrm{P}\{Y_{i,j} \leq k \mid \eta_{i,j}(k)\}$ to $\mathrm{P}\{Y_{i,j} \leq k \mid \eta^*_{i,j}(k)\}$ is

$$(X_{i,j} - X^*_{i,j})'\beta_0 + \left\{\exp(V'_{i,j}\alpha_0) - \exp(V^{*\prime}_{i,j}\alpha_0)\right\}\theta^*_k,$$

so that the effect of the scale covariate $V_{i,j}$ is uniform over each cutpoint θ^*_k.

The model (1)–(3), however, presents an approach for extending the location-scale model by allowing cutpoints to be covariate specific. This approach extends the models considered by Fahrmeir & Tutz (1994, Section 3.3) to the multirater problem. Setting $\alpha_0 = 0$, and including covariate specific cutpoint parameters α_k, we get

$$\mathrm{P}\{Y_{i,j} \leq k \mid \eta_{i,j}(k)\} = F\left\{X'_{i,j}\beta_0 + \beta_j + \sum_{\ell=1}^{k}\exp(\gamma_\ell + W'_{i,j}\alpha_\ell)\right\}, \tag{6}$$

where as before, $\gamma_1 = -\infty$ and $\alpha_1 = 0$ to ensure identification and $\gamma_K = +\infty$ and $\alpha_K = 0$ to guarantee a properly defined cdf. The model (6) is quite flexible and can be used to fit a number of different models. Section 4 illustrates an example of a diagnostic radiology study in which the use of a covariate specific cutpoint is used to test for a difference between diagnostic tests in interpreting a specific stage of disease. The model (6) can also be used in an analysis focusing on rater differences. By using covariates $W_{i,j}$ to index raters and α_ℓ as rater-category parameters, it can be used to test for differences in the way raters rate specific categories.

3. HYBRID MONTE CARLO GIBBS SAMPLING

In Section 4, we will investigate the use of each of the models, the location model, the location-scale model and the cutpoint covariate specific model, described by equations (4), (5) and (6), respectively. Here we present the hierarchy for our priors, followed by a description of the HMC method.

The priors for our parameters in (1)–(3) form a hierarchy that is based on normal priors for lower level parameters, and conjugate priors for hypermeans and hypervariances. As before, assume that $\gamma_K = +\infty$ and $\alpha_K = 0$ are fixed values. Other constraints, such as $\gamma_1 = -\infty$ and $\alpha_1 = 0$, can easily be accommodated within the following framework,

$$\begin{aligned}
(\beta_0 \mid B_0) &\sim \mathrm{N}(0, B_0 I) \\
(\beta_j \mid b, \sigma) &\overset{\text{i.i.d.}}{\sim} \mathrm{N}(b, \mathrm{diag}(\sigma)), \qquad j = 1, \ldots, J \\
(\alpha_0 \mid A_0) &\sim \mathrm{N}(0, A_0 I) \\
(\gamma_k \mid A) &\overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, AI) \\
(\alpha_k \mid A) &\overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, AI), \qquad k = 1, \ldots, K-1 \\
(b \mid B) &\sim \mathrm{N}(0, BI) \\
(\sigma_\ell^{-1} \mid s_1, s_2) &\overset{\text{i.i.d.}}{\sim} \mathrm{Gamma}(s_1, s_2), \qquad \ell = 1, \ldots, p.
\end{aligned} \tag{7}$$

Note that the model can also be easily extended to include more than one random effects population for β_j (as discussed earlier).

In (7) the notation diag(σ) is used to denote the diagonal matrix with diagonal elements $\sigma_1, \ldots, \sigma_p$, where p is the dimension for β_j. A noninformative inverse-gamma prior for σ_ℓ is used by selecting small values for the shape and scale parameters ($s_1 = s_2 = 0.0001$). Lower level parameters use flat priors through the use of large variances. The models studied here used $B_0 = A_0 = A = 1000$ and for the hypermean b, the value $B = 1000$ was used.

For notational convenience, let $\beta = (\beta_1, \ldots, \beta_J)'$, and with a constraint to the last cutpoint, set $\gamma = (\gamma_1, \ldots, \gamma_{K-1})'$, $\alpha = (\alpha_1, \ldots, \alpha_{K-1})'$. Also, define $\sigma = (\sigma_1, \ldots, \sigma_p)$. Sampling the posterior for $(\beta_0, \beta, \alpha_0, \gamma, \alpha, b, \sigma | Y)$ by the Gibbs sampler works by drawing values from the conditional distributions

$$(\beta_0, \beta \mid \alpha_0, \gamma, \alpha, b, \sigma, Y) \tag{8}$$

$$(\alpha_0 \mid \beta_0, \beta, \gamma, \alpha, Y) \tag{9}$$

$$(\gamma, \alpha \mid \beta_0, \beta, \alpha_0, Y) \tag{10}$$

$$(b, \sigma \mid \beta). \tag{11}$$

In implementing the sampler, the step (11) for b and σ is straightforward due to conjugacy. In particular,

$$(b \mid \beta, \sigma) \sim \mathrm{N}(B^* \mathrm{diag}(\sigma)^{-1}\beta^*, B^*),$$

where B^* is the diagonal matrix with elements $(J/\sigma_\ell + 1/B)^{-1}$ and $\beta^* = \sum \beta_j$. Furthermore,

$$(\sigma_\ell^{-1} | \beta, b) \overset{\text{ind}}{\sim} \mathrm{Gamma}\Big(s_1 + J/2, s_2 + \sum_{j=1}^{J} (\beta_{j,\ell} - b_l)^2/2\Big).$$

The updates (8)–(10) for the remaining lower level parameters are each completed using the leapfrog HMC method. A key criterion for implementing this approach requires that the density for the posterior be differentiable and strictly positive everywhere with respect to Lebesgue measure. To see why this condition is satisfied here, let $\eta_{i,j} = \eta_{i,j}(Y_{i,j})$ and define

$$\eta_{i,j}^- = \begin{cases} \eta_{i,j}(Y_{i,j} - 1) & \text{if } Y_{i,j} > 1, \\ -\infty & \text{when } Y_{i,j} = 1. \end{cases}$$

Then, the posterior density π for $(\beta_0, \beta, \alpha_0, \gamma, \alpha \mid b, \sigma, Y)$ is proportional to

$$\pi(\beta_0)\pi(\beta \mid b, \sigma)\pi(\alpha_0)\pi(\gamma)\pi(\alpha) \prod_{i=1}^{n} \prod_{j=1}^{J} \{F(\eta_{i,j}) - F(\eta_{i,j}^-)\}.$$

Note that the propriety of the posterior is easily established by observing that the product on the right-hand side is bounded by one. Establishing that the posterior is positive everywhere follows straightforwardly from our assumption that the density for F is positive, because $\eta_{i,j} > \eta_{i,j}^-$, and by our choice of normal priors for the parameters $(\beta_0, \beta, \alpha_0, \gamma, \alpha)$. It also follows that the posterior is differentiable by the choice of priors and our assumption that F has a density.

3.1. Leapfrog algorithm.

The HMC updates are implemented using the leapfrog method described by Duane *et al.* (1987). This method applies in general to a d-dimensional parameter ξ with a density π that is differentiable and positive everywhere with respect to Lebesgue measure. More theoretical details concerning the algorithm can be found in Duane, Kennedy, Pendleton & Roweth (1987),

Kennedy (1990) and Neal (1994; 1996, Chapter 3). The algorithmic details for generating an ergodic sample from π are as follows. If ξ_* is the current value for ξ, then the leapfrog algorithm generates a new value for ξ through the following steps:

1. Simulate a d-dimensional standard normal $Z_* \sim \mathrm{N}(0, I)$. Let

$$\xi_0 = \xi_* \qquad \text{and} \qquad Z_0 = Z_* + \delta\Delta(\xi_*)/2,$$

where Δ is the gradient of $\log \pi$.

2. For $\ell = 1, \ldots, L$, let

$$\begin{aligned} \xi_\ell &= \xi_{\ell-1} + \delta Z_{\ell-1}, \\ Z_\ell &= Z_{\ell-1} + \delta_\ell \Delta(\xi_\ell), \end{aligned}$$

where $\delta_\ell = \delta$ for $\ell < L$, otherwise $\delta_L = \delta/2$.

3. Move from the current value of ξ_* to the new candidate ξ_L with probability

$$\left[\frac{\pi(\xi_L)}{\pi(\xi_*)} \exp\left\{-\frac{1}{2}(Z_L' Z_L - Z_*' Z_*)\right\}\right] \bigwedge 1.$$

As discussed in Neal (1996, Chapter 3), and Ishwaran (1999), the strategy for selecting the number of leapfrog steps L and the step size δ used in the leapfrog algorithm is to choose values that encourage as large a trajectory as possible for ξ while leading to as high an acceptance rate as possible. The examples considered in this paper used $L = 100$ for the number of leapfrog steps, while δ was tuned in order to get approximately an 80%–90% acceptance rate. A good strategy for tuning δ is to start the Gibbs sampler off for a few iterations (100 steps were used here) using random walk Metropolis–Hastings. This brings us to a high probability region of the posterior where δ can then be tuned with minimal fuss. Starting the sampler off with HMC will usually require several adjustments to δ as the gradient rapidly changes for the first few iterations. Therefore, a better strategy is to wait several iterations before implementing HMC.

Creating the program code for the Metropolis step is straightforward and is easily adapted from the code used to run the leapfrog step. Once in place, it can be used to start the Gibbs sampler as discussed above, but it can also be used to create a mixed algorithm. In this approach, we can mix Metropolis with HMC in some preset proportion to reduce the amount of computation required in running HMC. Calculating the gradient Δ can sometimes be computationally expensive, especially with larger data sets, and with a mixture approach we can occasionally benefit from a large step due to HMC while reducing computation by using Metropolis–Hastings for the bulk of our iterations. We will investigate the use of a mixture algorithm in Sections 4 and 5.

3.2. Gradients.

To apply the leapfrog HMC algorithm, we must calculate the gradient vectors Δ for the model defined by (1)–(3) for each of the steps (8), (9) and (10). To define the necessary gradients, it will be convenient to introduce the following notation. Let $\theta_{i,j} = \theta_{i,j}(Y_{i,j})$ and define

$$\theta_{i,j}^- = \begin{cases} \theta_{i,j}(Y_{i,j} - 1) & \text{if} \quad Y_{i,j} > 1, \\ 0 & \text{when} \quad Y_{i,j} = 1. \end{cases}$$

In the following calculations, we will use the facts that $F(+\infty) = 1$ and $F(-\infty) = f(-\infty) = f(+\infty) = 0$.

TABLE 1: Mean values and standard deviations from the location model, the location-scale model and the model with covariate specific cutpoints applied to pediatric staging data. The last two rows (indicated by RMSE) are the root mean square error for CT and MR obtained by taking the root mean square difference between the observed predictive value and the estimated predictive value based on the value of imaging (see Table 2 for more details). Note that fixed effects parameters significant at the 5% level are indicated with a $*$.

		Model		
Type	Parameter	Location	Location-Scale	Cutpoint Specific
	age	$-0.39 \pm 0.14^*$	$0.42 \pm 0.19^*$	-0.03 ± 0.15
location	image=2	$-0.87 \pm 0.14^*$	$-0.98 \pm 0.19^*$	$-0.93 \pm 0.16^*$
β_0	image=3	$-1.14 \pm 0.13^*$	$-1.58 \pm 0.19^*$	$-1.44 \pm 0.15^*$
	image=4	$-2.60 \pm 0.16^*$	$-2.59 \pm 0.29^*$	$-2.66 \pm 0.19^*$
	CT reader 1	-0.38 ± 0.18	-0.52 ± 0.21	-0.36 ± 0.19
	CT reader 2	-0.07 ± 0.22	-0.11 ± 0.25	-0.01 ± 0.24
	CT reader 3	-0.38 ± 0.25	-0.52 ± 0.28	-0.37 ± 0.26
	CT reader 4	-0.35 ± 0.24	-0.45 ± 0.26	-0.29 ± 0.24
	CT reader 5	-0.43 ± 0.18	-0.61 ± 0.19	-0.46 ± 0.19
location	CT reader 6	-0.55 ± 0.16	-0.69 ± 0.19	-0.54 ± 0.17
random	CT reader 7	-0.44 ± 0.22	-0.56 ± 0.25	-0.44 ± 0.24
effects	CT reader 8	-0.06 ± 0.18	-0.17 ± 0.20	-0.07 ± 0.19
β_j	MR reader 1	0.25 ± 0.23	0.11 ± 0.22	0.20 ± 0.24
	MR reader 2	0.31 ± 0.19	0.12 ± 0.17	0.23 ± 0.18
	MR reader 3	0.03 ± 0.17	-0.22 ± 0.20	-0.06 ± 0.18
	MR reader 4	-0.16 ± 0.15	-0.41 ± 0.19	-0.21 ± 0.16
	MR reader 5	0.02 ± 0.20	-0.15 ± 0.21	-0.01 ± 0.21
	MR reader 6	0.14 ± 0.17	-0.08 ± 0.18	0.05 ± 0.17
	modality	-	0.11 ± 0.11	-
	age	-	$-1.38 \pm 0.24^*$	-
scale	image=2	-	0.18 ± 0.17	-
α_0	image=3	-	$0.49 \pm 0.15^*$	-
	image=4	-	0.06 ± 0.23	-
	modality	-	-	$0.37 \pm 0.17^*$
	age	-	-	$-2.57 \pm 0.46^*$
cutpoint 3	image=2	-	-	$1.38 \pm 0.46^*$
α_3	image=3	-	-	$2.12 \pm 0.43^*$
	image=4	-	-	$1.32 \pm 0.46^*$
RMSE	CT	0.06	0.04	0.03
RMSE	MR	0.11	0.09	0.07

TABLE 2: Observed and estimated predictive values for true stage conditioned on value of imaging from pediatric staging data. Top row corresponds to the observed value (in bold), while second, third and fourth rows are the mean values estimated from the location model, the location-scale model and the cutpoint specific regression model, respectively. *Key predictive values for CT and MR for stages 3 and 4 are emphasized to highlight the differences in the two modalities. Note that in all cases, predictive values sum to a probability of one across columns.

		True Stage			
Imaging		1	2	3	4
		0.389	**0.357**	**0.031**	**0.221**
Stage		0.298	0.324	0.181	0.195
by CT	1	0.347	0.263	0.147	0.241
		0.341	0.335	0.039	0.284
		0.049	**0.213**	**0.164**	**0.574***
Stage		0.086	0.216	0.209	0.486
by CT	2	0.097	0.199	0.190	0.514
		0.097	0.236	0.158	0.507
		0.029	**0.116**	**0.243***	**0.612***
Stage		0.053	0.168	0.191	0.587
by CT	3	0.035	0.164	0.216	0.586
		0.039	0.146	0.254	0.560
		0.011	**0.023**	**0.046**	**0.919**
Stage		0.001	0.013	0.036	0.949
by CT	4	0.002	0.013	0.031	0.953
		0.001	0.017	0.028	0.953
		0.288	**0.461**	**0.057**	**0.192**
Stage		0.447	0.313	0.134	0.104
by MR	1	0.449	0.303	0.125	0.122
		0.464	0.317	0.049	0.169
		0.255	**0.340**	**0.191**	**0.213***
Stage		0.196	0.300	0.208	0.296
by MR	2	0.183	0.326	0.213	0.277
		0.189	0.306	0.241	0.263
		0.150	**0.183**	**0.483***	**0.183***
Stage		0.129	0.256	0.215	0.398
by MR	3	0.069	0.323	0.279	0.329
		0.083	0.219	0.431	0.267
		0.012	**0.023**	**0.064**	**0.900**
Stage		0.004	0.034	0.072	0.889
by MR	4	0.006	0.033	0.067	0.894
		0.004	0.036	0.074	0.886

(a) From elementary calculus, it easily follows that for β_0, we have

$$\Delta(\beta_0) = \sum_{i,j} \frac{f(\eta_{i,j}) - f(\eta^-_{i,j})}{F(\eta_{i,j}) - F(\eta^-_{i,j})} X_{i,j} - \beta_0/B_0,$$

and similarly for β_j,

$$\Delta(\beta_j) = \sum_{i} \frac{f(\eta_{i,j}) - f(\eta^-_{i,j})}{F(\eta_{i,j}) - F(\eta^-_{i,j})} U_{i,j} - \text{diag}(\sigma)^{-1}(\beta_j - b), \qquad j = 1, \ldots, J.$$

The gradient vector formed by combining these $J + 1$ gradients allows us to complete (8).

(b) For (9), the required gradient is

$$\Delta(\alpha_0) = \sum_{i,j} \frac{f(\eta_{i,j})\theta_{i,j} - f(\eta^-_{i,j})\theta^-_{i,j}}{F(\eta_{i,j}) - F(\eta^-_{i,j})} V_{i,j} - \alpha_0/A_0,$$

where we are using the fact that $f(\infty) \times \infty = 0 \times \infty = 0$.

(c) For the cutpoints γ_k, we have

$$\Delta(\gamma_k) = \sum_{\{i,j:Y_{i,j} \geq k\}} \frac{f(\eta_{i,j}) - f(\eta^-_{i,j})\{Y_{i,j} > k\}}{F(\eta_{i,j}) - F(\eta^-_{i,j})} \exp\left(V'_{i,j}\alpha_0 + \gamma_k + W'_{i,j}\alpha_k\right) - \gamma_k/A.$$

Similarly, for the covariate specific cutpoints,

$$\Delta(\alpha_k) = \sum_{\{i,j:Y_{i,j} \geq k\}} \frac{f(\eta_{i,j}) - f(\eta^-_{i,j})\{Y_{i,j} > k\}}{F(\eta_{i,j}) - F(\eta^-_{i,j})} W_{i,j} \exp\left(V'_{i,j}\alpha_0 + \gamma_k + W'_{i,j}\alpha_k\right) - \alpha_k/A,$$

for $k = 2, \ldots, K - 1$. Combined together, these give the required gradient vector for step (10).

4. PEDIATRIC STAGING ANALYSIS

Our first example involves ordinal data collected from a multicenter, multireader diagnostic radiology study involving solid tumor cancer in 94 children of varying ages of up to 13 years. The ordinal data collected consists of the metastatic stage of cancer measured on the 4 point TNM scale, where the true stage of cancer was determined through pathological diagnosis, and in particular, it was based on bone biopsy for children who were stage 4, or for a small subset of these patients, through liver and lung biopsy.

Each of the $n = 94$ children underwent a series of CT and MR scans. These images were then circulated to the 6 different participating institutions, where they were interpreted by a total of 8 CT readers and 6 MR readers who each recorded their degree of suspicion of metastatic cancer on the 4-point TNM scale. Each patient's set of CT images was read by one CT reader per hospital, and therefore, not all CT images were read by all CT readers. The experimental design used had the effect of assigning each patient's set of CT images to a random subset of 6 readers from our pool of 8 CT radiologists. Because of missing values, there was a total of 688 observations: 358 resulting from CT and 330 from MR.

The emphasis in this study was on predicting the true stage of cancer for a patient. Therefore, the ordinal response in our case, $Y_{i,j}$, is the true stage of cancer as determined by pathology. Note that this is a fixed value for each patient i. Covariates included age, and the stage determined by the radiologist from the patient's image. An ROC analysis, using imaging for the response and true disease status as a covariate, is easily implemented using similar methods described here.

The data were fit using a probit link for the location model (4), the location-scale model (5) and the cutpoint covariate specific ordinal regression model (6) that were discussed in Section 2.

In the location model, in addition to the use of age and the value of imaging as location parameters, the model also included a random intercept term for the specific reader, where these parameters are assumed to arise from 2 different random effects populations representing CT and MR readers. With the location-scale model, we also used the same location parameters, but included scale parameters for age, imaging and the type of modality (either CT or MR). Note that modality cannot be introduced as a location term because of the nonidentifiability that would exist between it and the random intercept terms.

Neither the location model nor the location-scale model will perform very well in this example because of the fundamental difference between the way CT and MR perceive stage 4 cancer. Stage 4 cancer involves bone metastases which cannot be visualized by CT, and consequently CT cannot delineate clearly between stage 3 cancer and stage 4 cancer. Both the location model and location-scale model will fail to model this difference correctly, because in both models the effect of a covariate can only be uniform over the cutpoints. To properly model the effect of the stage-modality effect, a cutpoint specific model (6) was used by introducing a covariate specific cutpoint parameter α_3 which included age, modality and the value of imaging.

The posterior parameter estimates for each of the three models are given in Table 1. Note that in the parameterization used for covariates, the stage 1 value for imaging acts as a baseline value, and for modality, CT has been coded as $-1/2$ and MR as $1/2$. Age was measured in months and scaled by 1/100.

In all three models, we find that the value of imaging is significant as a location parameter, with parameter estimates decreasing in size as the value of imaging increases. This reflects a type of ordering, and in particular, it implies a decreasing probability of stage 1 cancer, and an increasing probability of stage 4 cancer, as the value observed for imaging increases. In the location model, the difference between CT and MR is expressed through the location random effects parameters β_j. In this case, we see that CT readers tend to have negative intercept terms, while MR readers tend to have positive terms. This has the effect of increasing the probability of stage 4 cancer for CT when the value of imaging is either stage 2 or stage 3. Although this leads to a correct adjustment for CT, it has the unfortunate effect of underestimating the probability of stage 3 cancer and overestimating the probability of stage 4 cancer when MR interprets an image as stage 3 (see Table 2.) The same problem also occurs in the location-scale model.

However, the cutpoint specific model corrects this problem, and it also leads to the lowest values for the RMSE (root mean square error) for both CT and MR (see the bottom of Table 1). The importance of including cutpoint specific parameters is also apparent when we consider the parameter estimates for α_3, which are all significant in this model. In particular, the significant effect due to modality indicates the anticipated difference between CT and MR for predicting stages 3 and 4. Interestingly, the parameter estimate for age was also significant, although there is no effect for age in location. Therefore, the large negative value for age seen in α_3 must imply that the impact of age is stage dependent, and in this case, indicates a higher probability of stage 4 cancer in older children. The parameter estimates for imaging in α_3 are also significant, indicating an image effect which is stage dependent. Indeed, because of the large RMSE observed for MR in the location model and the location-scale model (11% and 9% respectively), we suspect that this imaging effect is mostly due to MR. Thus, the effect of MR imaging will be different for predicting stages 1 and 2 then for predicting stages 3 and 4. Note that this also implies that the location and location-scale model will perform poorly even when applied only to the MR data.

Figure 1 shows that the cutpoint specific model appears to be mixing well, even with such a small number of burn-in iterations (1000) and without lagging sampled values. The autocorrelations for β_0 and α_3 in the figure are quite reasonable, although there is a high autocorrelation seen for one component of γ_k. The same mixing performance can be obtained through the use of a mixture algorithm as discussed in Section 3. To demonstrate this, a mixed algorithm was used based on mixing 90% random walk Metropolis–Hastings (RWMH) with 10% HMC. The RWMH step employed the same strategy for grouping parameters as the HMC Gibbs sampler

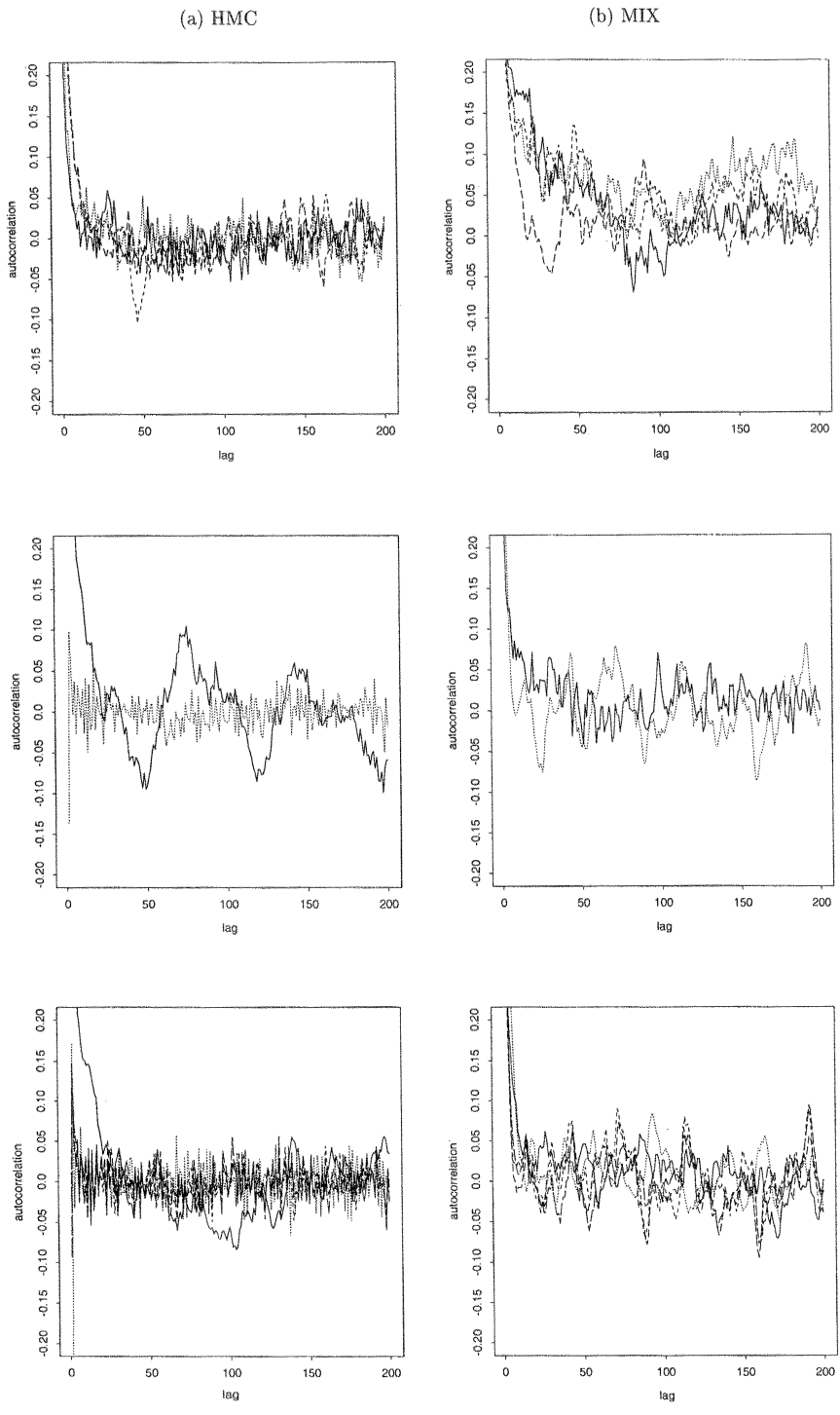

FIGURE 1: Autocorrelations for β_0 location parameters (top row), γ_k cutpoints (middle row) and covariate specific cutpoint α_3 (bottom row) obtained from the covariate specific ordinal regression model applied to the pediatric staging data. Left-hand side (a) is based on the HMC Gibbs sampler using 2,500 sampled values following a 1,000 iteration burn-in, while right-hand side (b) is a mixture of 10% HMC with 90% RWMH. Mixture algorithm used a 5,000 iteration burn-in followed by subsampling 2,500 values each 10th iteration. Both cases employed $L = 100$ leapfrog steps.

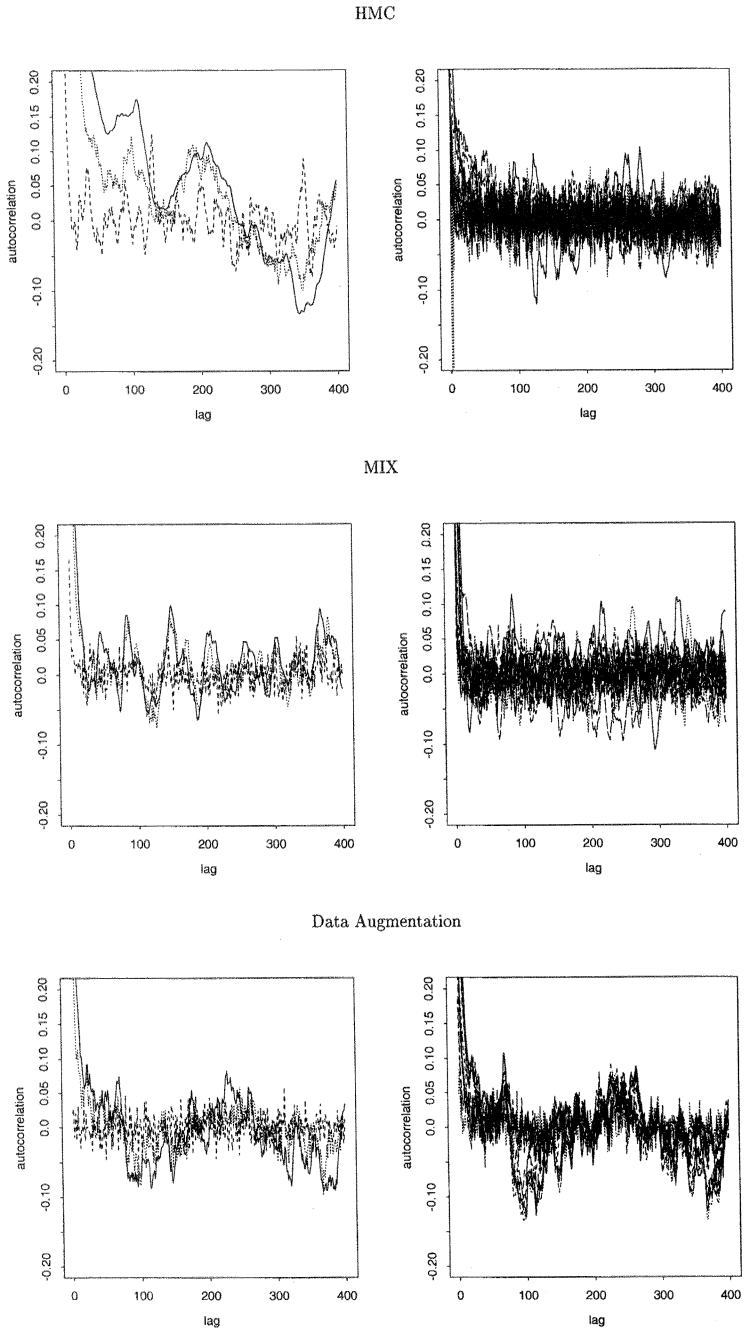

FIGURE 2: Autocorrelations for β_0 location parameters (left-hand side) and cutpoints (right-hand side) from our simulated example. The top row is the HMC Gibbs sampler based on 2,500 sampled values following a 1,000 iteration burn-in with $L = 100$ leapfrog steps. The middle row is obtained using a 90% mixture of RWMH and 10% HMC, while the bottom row is by data augmentation. The mixture algorithm was based on 2,500 sampled values subsampled each 10th iteration following an initial 5,000 iteration burn-in, while data augmentation used a 25,000 iteration burn-in followed by 2,500 sampled values subsampled each 100th iteration. Cutpoint autocorrelations are based on cutpoint values in the unconstrained γ parameterization for HMC and mixture algorithm and in the ordered θ^* parameterization for data augmentation.

TABLE 3: True values and posterior estimates from the simulated example. Mean values and standard deviations are based on HMC, mixture of 90% RWMH and 10% HMC (MIX) and data augmentation. Note that cutpoint values are given in the ordered θ^* parameterization.

Type	Parameter	True Value	HMC	MIX	Data Augmnt.
	$\beta_{0,1}$	-2.50	-2.76 ± 0.20	-2.80 ± 0.22	-2.91 ± 0.17
location	$\beta_{0,2}$	1.00	1.13 ± 0.07	1.13 ± 0.08	1.15 ± 0.06
	$\beta_{0,3}$	1.00	0.70 ± 0.15	0.71 ± 0.16	0.73 ± 0.16
	θ_1^*	0.00	-	-	-
	θ_2^*	0.25	0.12 ± 0.06	0.12 ± 0.07	0.15 ± 0.07
	θ_3^*	0.50	0.31 ± 0.10	0.32 ± 0.10	0.38 ± 0.11
	θ_4^*	0.75	0.59 ± 0.13	0.59 ± 0.14	0.67 ± 0.14
	θ_5^*	1.00	1.07 ± 0.17	1.09 ± 0.17	1.16 ± 0.17
	θ_6^*	1.25	1.20 ± 0.18	1.22 ± 0.18	1.32 ± 0.17
cutpoints	θ_7^*	1.50	1.32 ± 0.19	1.34 ± 0.19	1.46 ± 0.18
	θ_8^*	1.75	1.55 ± 0.20	1.57 ± 0.20	1.71 ± 0.18
	θ_9^*	3.00	3.09 ± 0.27	3.12 ± 0.28	3.21 ± 0.18
	θ_{10}^*	3.25	3.56 ± 0.28	3.59 ± 0.29	3.70 ± 0.22
	θ_{11}^*	3.50	3.79 ± 0.29	3.83 ± 0.31	3.96 ± 0.24
	θ_{12}^*	3.75	4.09 ± 0.30	4.13 ± 0.32	4.26 ± 0.25
	θ_{13}^*	4.00	4.27 ± 0.30	4.32 ± 0.32	4.47 ± 0.25
	θ_{14}^*	4.50	4.92 ± 0.33	4.97 ± 0.35	5.12 ± 0.25
	θ_{15}^*	$+\infty$	-	-	-

outlined in Section 3, but with updates for parameters using Metropolis–Hastings with a multivariate normal proposal. The number of iterations used was chosen to roughly equal the number of steps used in the fully HMC approach by counting each HMC iteration as L steps. Figure 1 shows that the mixture algorithm is performing well.

5. SIMULATIONS

In order to test the performance of the HMC method, $n = 200$ observations were simulated from a location model of the form (4) using a standard normal cdf for F. The simulation was based on a univariate model (one observation per i), with a 3-dimensional fixed effects parameter β_0, in which the first coordinate represented an intercept term. The covariates for the remaining two coordinates were simulated independently from a uniform distribution on $\{-3, -2, -1, 0, 1, 2, 3\}$ and a $\mathrm{N}(0, 0.5^2)$ distribution, respectively. The model included $K = 15$ preselected cutpoint parameters γ_k, where $\gamma_1 = -\infty$ and $\gamma_K = +\infty$.

In addition to using the HMC Gibbs sampler, a mixed algorithm of the type described in Section 4 was employed based on 10% HMC and 90% RWMH. The posterior was also sampled using the data augmentation procedure outlined in Albert & Chib (1993). This approach used the same priors as in the HMC Gibbs sampler, excepting the priors for cutpoints which employed a uniform prior over $\{0 = \theta_1^* \leq \theta_2^* \leq \cdots \leq \theta_{14}^* \leq 10\}$.

Parameter estimates are presented in Table 3 using the θ^* parameterization to facilitate comparisons with the Albert & Chib (1993) method. In comparing estimates, we find that all 3

procedures have done fairly well in recovering the true parameter values, although in each case the β_0 intercept term, corresponding to the first cutpoint, is slightly biased with a smaller value. This small bias also appears to affect the estimates for the first few θ^* cutpoints, which are also slightly smaller than the true values.

Looking at the autocorrelation plots in Figure 2, we find that both HMC and the mixture algorithm have mixed well for the cutpoints, but pure HMC has higher autocorrelations for the β_0 parameters, especially for the intercept term. Interestingly, this autocorrelation does not appear to effect the mixing for γ_k. For data augmentation, we find a non-vanishing autocorrelation for the cutpoints, and high autocorrelations for β_0, even when subsampling each 100th iteration (this means that 250,000 iterations were used after burn-in). As mentioned in the Introduction, working directly with ordered cutpoints generally leads to high autocorrelations, and is a well-known weakness of this approach (see, e.g., Cowles 1996 or Nandram & Chen 1996).

6. DISCUSSION AND SUMMARY

Conventional cumulative link ordinal regression models are limited by their inability to model covariates as category specific. However, by employing a different type of parameterization, this paper has presented a Bayesian class of models general enough to permit cutpoints to be covariate specific, but also flexible enough to fit more standard models, such as the mixed effects model and the location-scale ordinal regression model. Cutpoint specific models are especially relevant to analysis of diagnostic radiology studies, where the use of covariates that are category specific can be used to test for specific differences in the way diagnostic modalities operate. This point was illustrated in Section 4, where we considered data from a pediatric radiology study and tested the hypothesis that CT and MR operate differently in interpreting a specific stage of cancer. Although we did not explore models for testing for specific rater differences, the covariate specific model, as discussed in Section 2.4, can be used to test for differences in performance of raters for specific categories. In the analysis of radiology studies, this can be used as a tool for identifying the stages of a disease where radiologists are more likely to disagree in their diagnostic interpretation.

With greater model flexibility, there is also a greater challenge to model fitting. The strategy taken here was to fit models using the leapfrog HMC Gibbs sampler. This method leads to an efficiently mixing Markov chain and a Gibbs sampler that is more portable and generic in nature compared to those based on latent variable data augmentation. Although a weakness with the HMC method is the computational complexity required in evaluating gradients, the amount of computations required can be reduced by mixing the HMC algorithm with random walk Metropolis–Hastings. As we found in Sections 4 and 5, this method leads to a procedure that appears to perform as well as pure HMC.

A few important points should be followed when implementing HMC. To reduce correlations between parameters, and to encourage better mixing, it is a good strategy to center covariates to have mean zero and to rescale covariates to have variance one. Secondly, the number of leapfrog steps should be chosen fairly large (such as $L = 100$), and tuning for δ should lead to high acceptance rates in the range of 80%–90%. In the case where these methods fail to reduce autocorrelations, a final strategy might be to combine the steps (8)–(10) into one large multivariate HMC update.

ACKNOWLEDGEMENTS

The author is indebted to Constantine Gatsonis and George Papandonatos for helpful discussion and advice.

REFERENCES

J. H. Albert & S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88, 669–679.

J. H. Albert & S. Chib (1998). *Bayesian Methods for Cumulative, Sequential and Two-Step Ordinal Data Regression Models*. Technical Report.

M. K. Cowles (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing*, 6, 101–111.

S. Duane & J. B. Kogut (1986). The theory of hybrid stochastic algorithms. *Nuclear Physics B*, 275, 398–420.

S. Duane, A. D. Kennedy, B. J. Pendleton & D. Roweth (1987). Hybrid Monte Carlo. *Physics Letters B*, 195, 216–222.

L. Fahrmeir & G. Tutz (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag, New York.

C. A. Gatsonis (1995). Random-effects models for diagnostic accuracy data. *Academic Radiology*, 2 *Supplement*, S14–S21.

P. Gustafson (1997). Large hierarchical Bayesian analysis of multivariate survival data. *Biometrics*, 53, 230–242.

J. A. Hanley & B. J. McNeil (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29–36.

D. Hedeker & R. J. Mermelstein (1998). A multilevel thresholds of change model for analysis of stages of change data. *Multivariate Behavioral Research*, 33, 427–455.

D. Hedeker, R. J. Mermelstein & K. A. Weeks (1999). The thresholds of change model: an approach to analyzing stages of change data. *Annals of Behavioral Medicine*, 21, 61–70.

H. Ishwaran (1999). Applications of hybrid Monte Carlo to Bayesian generalized linear models: quasicomplete separation and neural networks. *Journal of Computational and Graphical Statistics*, 8, 779–799.

H. Ishwaran & C. A. Gatsonis (2000). A general class of hierarchical ordinal regression models with applications to correlated ROC analysis. *The Canadian Journal of Statistics*, 28, 731–750.

V. E. Johnson & J. H. Albert (1998). *Ordinal Data Modeling*. Springer-Verlag, New York.

A. D. Kennedy (1990). The theory of hybrid stochastic algorithms. In *Probabilistic methods in quantum field theory and quantum gravity* (P. H. Damgaard, H. Huffel & A. Rosenblum, eds.), Plenum Press, New York, pp. 209–223.

B. Nandram & M-H. Chen (1996). Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. *Journal of Statistical Computation and Simulation*, 54, 129–144.

R. M. Neal (1994). An improved acceptance procedure for the hybrid Monte Carlo algorithm. *Journal of Computational Physics*, 111, 194–203.

R. M. Neal (1996). *Bayesian Learning for Neural Networks*. Springer-Verlag, New York.

N. Obuchowski (1995). Multi-Reader multi-modality ROC studies: Hypothesis testing and sample size estimation using an ANOVA approach with dependent observations. *Academic Radiology*, 2 *Supplement*, S22–S29.

M. L. Thompson & W. Zucchini (1989). On the statistical analysis of ROC curves. *Statistics in Medicine*, 8, 1277–1290.

A. N. Tosteson & C. B. Begg (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making*, 8, 204–215.

Received 20 December 1999
Accepted 11 April 2000

Hemant ISHWARAN: ishwaran@bio.ri.ccf.org
Department of Biostatistcs and Epidemiology
Cleveland Clinic Foundation, Cleveland, OH 44195, USA