Statistics and Computing (1996) **6**, 101–111

Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models

MARY KATHRYN COWLES

Department of Biostatistics, Harvard School of Public Health, Boston, MA 02215, USA

Received October 1994 and accepted October 1995

The ordinal probit, univariate or multivariate, is a generalized linear model (GLM) structure that arises frequently in such disparate areas of statistical applications as medicine and econometrics. Despite the straightforwardness of its implementation using the Gibbs sampler, the ordinal probit may present challenges in obtaining satisfactory convergence.

We present a multivariate Hastings-within-Gibbs update step for generating latent data and bin boundary parameters jointly, instead of individually from their respective full conditionals. When the latent data are parameters of interest, this algorithm substantially improves Gibbs sampler convergence for large datasets. We also discuss Monte Carlo Markov chain (MCMC) implementation of cumulative logit (proportional odds) and cumulative complementary log-log (proportional hazards) models with latent data.

Keywords: Blocking, collapsing, data augmentation, Gibbs sampler, latent data

1. Introduction

Ordinal response variables are very common in biostatistical and econometric applications. For example, the outcome variable in a comparative trial of analgesics might be participants' self-report of change in pain status on a three-point scale consisting of 'improved', 'no change', and 'worse'. Generalized linear models with a cumulative link function are commonly used to analyse the relationship between an ordinal response variable and predictor variables, which may be continuous, nominal, or ordinal. Assume that for each subject i we observe a response variable W_i, which may take on any one of k ordered values labelled $1, 2, \dots, k$. Values of a set of predictor variables $\boldsymbol{x}_i$ are also observed. A cumulative link model (Agresti, 1990) for these data would be of the form

$$\Pr(W_i \leqslant j | \boldsymbol{x}, \boldsymbol{\beta}) = G(\gamma_j - \boldsymbol{x}_i^t \boldsymbol{\beta}),$$

where G is the cumulative distribution function of a continuous random variable having positive density over the entire real line; γ_j, $j = 0, 1, \dots k$, are ordered cutpoints dividing the real line into intervals; and $\boldsymbol{\beta}$ is a vector of coefficients of the predictors. If G is the logistic c.d.f., then the model is a *cumulative logit* or *proportional odds* model, while if G is the extreme value (minimum) c.d.f., the model is called the *cumulative complementary log-log* or *proportional hazards* model. We will first consider the model in which G is the normal c.d.f.—the *cumulative probit* or *ordinal probit* model. For each of the three link functions, maximum likelihood methods may be used to get point estimates and asymptotic standard errors of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, although the validity of the asymptotic standard errors is questionable for small sample sizes.

Albert and Chib (1993) present Bayesian implementations of the ordinal probit model using the Gibbs sampler. They point out that the ordinal probit may be visualized in terms of an unobservable, or 'latent' continuous variable y_i^* corresponding to each observed variable w_i. The value of each y_i^* falls into one of k contiguous bins on the real line demarcated by the cutpoints $\gamma_0, \gamma_1, \dots, \gamma_k$, and the observed values of the w_i's are determined by the relationship $w_i = j$ if $y_i^* \in (\gamma_{j-1}, \gamma_j]$. Then the assumption that $y_i^* \sim \mathrm{N}(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta}, 1)$ makes this latent variable model equivalent to the cumulative probit. The variance of the unobservable y_i^*'s is assumed to be 1 for consistency with the standard normal c.d.f. link function.

Applying the 'data augmentation' idea of Tanner and Wong (1987), Albert and Chib (1993) treat the unknown

0960-3174

$\boldsymbol{y}^*$ values as additional parameters to be simulated in the Gibbs sampler. Once values are obtained for $\boldsymbol{y}^*$, the problem of estimating $\boldsymbol{\beta}$ in the ordinal probit model simplifies to that of doing so in a standard normal linear model. If a flat prior is specified for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, then the full conditional distributions for $\boldsymbol{\beta}$ and $\boldsymbol{y}^*$, as laid out by Albert and Chib, are:

$$p(y_i^*|\boldsymbol{\beta}, \boldsymbol{\gamma}, w_i) = \mathrm{N}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}, 1) \tag{1a}$$

truncated to $(\gamma_{w_i-1}, \gamma_{w_i}]$, and

$$p(\boldsymbol{\beta}|\boldsymbol{w}, \boldsymbol{y}^*) = \mathrm{N}((\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}^*, (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}). \tag{1b}$$

In order that the domain of the y_i^*'s may be the entire real line, the extreme bin cutpoints, γ_0 and γ_k, must be fixed at $-\infty$ and $+\infty$ respectively. Carlin and Polson (1992) present a parametric approach to the remaining cutpoints that introduces additional assumptions into the ordinal probit model (although not into a binary probit). We pursue the more general approach of Albert and Chib. They note that, in order to make the parameters of the model identifiable, one additional cutpoint must be fixed; without loss of generality they fix γ_1 at 0. (An alternative to fixing γ_1 would be to omit the intercept from the model; however, we continue under the assumption that γ_1 is fixed.) Then the full conditional distribution for each variable γ_j is uniform:

$$\begin{aligned} &p(\gamma_j|\boldsymbol{w}, \boldsymbol{y}^*, \boldsymbol{\beta}, \{\gamma_1, l \neq j\} \\ &\quad = U[\max(\max\{y_i^* : w_i = j\}, \gamma_{j-1}), \\ &\qquad \min(\min\{y_i^* : w_i = j+1\}, \gamma_{j+1})]. \end{aligned} \tag{1c}$$

Albert and Chib used data augmentation in their implementations of the ordinal probit solely so that the full conditionals in the Gibbs sampler would be standard densities. However, in some problems, the values of the latent variables may be of interest. For example, Albert (1992) used latent data to estimate the polychoric correlation coefficient between two ordinal variables. Similarly, Cowles *et al.* (1996) used the values of a latent continuous variable underlying an ordinal response to estimate the correlations between the ordinal response and two continuous response variables. In the latter problem, the ordinal probit was just one component of a complex random-effects model with over 6000 parameters. In order to conserve computer resources, Cowles *et al.* (1996) sought a sampling algorithm that would provide good parameter estimates based on hundreds, rather than thousands or hundreds of thousands, of iterations. The present paper gives details of that algorithm.

In Section 2 we demonstrate that, for an ordinal probit with latent data, convergence of the Gibbs sampler using univariate full conditionals may be slow when the sample size is large. In Section 3 we propose a multivariate Hastings-within-Gibbs update step that substantially accelerates convergence for the three-bin problem, and in Section 4 we extend this method to an ordinal probit problem with more than three bins and to cumulative logit and cumulative complementary log-log models. Finally, in Section 5 we suggest areas for further work.

2. Convergence of the ordinal probit

Convergence of the Gibbs sampler implemented by simulating from the univariate full conditionals (1c), (1a), and (1b) in sequence appears to depend on how full the bins for the y_i^*'s are—that is, on the sample size. Convergence is very slow when the bins are full because the interval within which each γ_j must be generated from its full conditional (1c) is very narrow, so the cutpoint values can change very little between successive iterations. Until the cutpoints are in roughly the right places, the values of the y_i^*'s are distorted, so convergence of the β's is also retarded.

To simulate a simple example of a large-sample three-level ordinal probit, we generated $N = 2000$ data points from the model

$$y_i^* = \beta_0 + \beta_1 x_i + \epsilon_i$$
$$\epsilon_i = \mathrm{N}(0, 1)$$

with $\beta_0 = 1$ and $\beta_1 = -2$. We then calculated the 1/3 and 2/3 quantiles of the $\boldsymbol{y}^*$ and assigned a corresponding ordinal variable w_i to each y_i^* as follows:

$$w_i = \begin{cases} 1, & y_i^* \text{ in lowest tertile} \\ 2, & y_i^* \text{ in middle tertile} \\ 3, & y_i^* \text{ in highest tertile} \end{cases}$$

Using the w's and x's from the simulated data, we ran five parallel Gibbs sampler chains for 6000 iterations. Figure 1 shows resulting convergence plots for $\boldsymbol{\beta}$ and γ_2 (the only stochastic cutpoint in a three-bin ordinal probit). To assess Gibbs sampler convergence, for each parameter we computed Gelman and Rubin's (1992) 'shrink factor'—the factor by which variance in estimation is inflated due to stopping the chain after the number of iterations run instead of continuing sampling in the limit. Gelman and Rubin suggest running Gibbs sampler chains until the estimated shrink factors are less than about 1.1 for all parameters of interest. (For a comparative review of this and other convergence diagnostics, see Cowles and Carlin, 1996). The median and 97.5th percentiles of the shrink factors are shown above the plots. The plots suggest that, after approximately 3000 iterations, all chains for the β's are traversing the same sample space, and Gelman and Rubin's diagnostic perhaps implies that the last 3000 iterations may be used for estimation.

Since we were primarily interested in algorithms that would converge rapidly, necessitating fewer than 1000 iterations, we next examined the first 800 iterations of these chains, shown in Fig. 2. Here, both the graphical impression