Continuation-ratio Model for Categorical Data: A Gibbs Sampling Approach

Wan Kai, Pang *

Abstract—**In this paper we discuss the continuation-ratio model for ordinal data. This particular type of model is to model the probability of one particular category given the categories proceeding this one. It can be shown that estimation of the continuation-ratio model parameters can be done efficiently by using the techniques in fitting the binary data models. In this way, one does not have to estimate the cut-point parameters as in the cumulative probability models. A Bayesian approach with the use of Gibbs sampler is adopted in this paper. The adaptive rejection sampling method proposed by Gilks and Wild is used. The adaptive rejection sampling (ARS) algorithm is an efficient and direct method to sample from complicated log-concave densities often found in many Gibbs sampling scheme. We applied this model to analyse data obtained from experiments about quality of telephone connection conducted by British Telecom (BT) Laboratory. The final results are satisfactory.**

Keywords: Continuation-ratio Model; Ordered Categorical Data, Markov Chain Monte Carlo; Adaptive Rejection Sampling.

1 Introduction

A frequently used method of response in many scientific experiments is a three, four or five-point scale, graded, for example, subjectively from 'Excellent' to 'Bad'. In theory a preferable model for this kind of data would be one which takes into account the categorical nature of the response. The response is not continuous but rather discrete with only four/five possible values. More information is contained within the ordered structure of the categories. The categories are not separate, independent possibilities; they are strictly increasing (decreasing) from 'Bad' to 'Excellent'. Also relevant is the fact that these categories are not fixed but rather arbitrary cut-points of some underlying continuum. This underlying continuum is the unmeasurable subjective response.
As a result of these considerations, it is the probability of a response falling into a certain category (π_j) which is the focus of the modelling procedure. Associated with this are the cumulative probabilities (γ), the probabilities of the response falling in certain category or below it. Models for ordered categorical data had been studied by many researchers (either in a frequentist domain (McCullagh (1977, 1978), Stram *et. al* (1988), Jansen (1990), Agresti and Lang (1993)) or in a Bayesian approach (Albert and Chib (1993)). The popular models are proportional-odds model (cumulative logit model), cumulative probit model and the cumulative complementary log-log model. These models model the cumulative probability up to certain level of category and they are based on the generalized linear model settings with an appropriate link functions. Models for the cumulative probability of category j, $\gamma_j = \Pr(Y_j \leq j)$, can be founded in the monograph by Agresti (1996).
The model which is discussed in this article is the continuation-ratio model. This is defined as follows (Agresti (1996))

$$\text{logit}\left[P_i(\text{Cat} = j \mid \text{Cat} \geq j)\right] = \alpha_j - \mathbf{x}_i'\beta \qquad (1)$$

where α_j is the cut-point parameter for category j. This model is particularly flexible in the number of ways it may be rewritten to ease interpretation. Thus the following are all equivalently the continuation-ratio model:

$$\text{logit}\left(\frac{\pi_{ij}}{1-\gamma_{ij}+\pi_{ij}}\right) = \alpha_j - \mathbf{x}_i'\beta \qquad (2)$$

$$\log\left(\frac{\pi_{ij}}{1-\gamma_{ij}}\right) = \alpha_j - \mathbf{x}_i'\beta \qquad (3)$$

$$\log\left(\frac{\pi_{ij}}{\sum_{p=j+1}^{k} + \pi_{ip}}\right) = \alpha_j - \mathbf{x}_i'\beta \qquad (4)$$

Similarly there are the continuation-ratio complementary log-log and continuation-ratio probit models when the link function is the complementary log-log and probit respectively. The continuation-ratio link models do not have the same appeal to an underlying continuum as in the case with the cumulative link models. All continuation-ratio link models define strict stochastic ordering but they are not, in general, invariant to the collapsing of the contiguous categories.

*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, E-mail: mapangwk@inet.polyu.edu.hk. Acknowledgement: This research work is supported by the research committee of the Hong Kong Polytechnic University

2 Relationship between Continuation-ratio Model and Model for Binary Data

It is easy to show that the likelihood function for continuation-ratio link models can be split into $r-1$ independent binomial likelihood functions. To demonstrate, consider the case of an ordinal response having four categories (r =4). The likelihood L for the ijth observation is proportional to

$$L \quad \propto \quad \pi_{ij,1}^{y_{ij,1}} \times \pi_{ij,2}^{y_{ij,2}} \times \pi_{ij,3}^{y_{ij,3}} \times \pi_{ij,4}^{y_{ij,4}}, \qquad (5)$$

where $y_{ij,h} = 1$ if the ordinal response $y_{ij} = h$ and $y_{ij} = 0$ otherwise. Since $\sum_{h=1}^{4} \pi_{ij,h} = 1$, this is also the joint likelihood for one observation from each of the three independent binomial distributions,

(i) $\text{Bin}(1, \pi_{ij,1})$

(ii) $\text{Bin}\left(y_{ij,2} + y_{ij,3} + y_{ij,4}, \dfrac{\pi_{ij,2}}{\pi_{ij,2} + \pi_{ij,3} + \pi_{ij,4}}\right)$

(iii) $\text{Bin}\left(y_{ij,3} + y_{ij,4}, \dfrac{\pi_{ij,3}}{\pi_{ij,3} + \pi_{ij,4}}\right)$

A continuation-ratio link model is in this case

$$\begin{aligned} \text{Link}(\pi_{ij,1}) &= \alpha_1 - \mathbf{x}_i'\beta_{(1)} \\ \text{Link}\left(\frac{\pi_{ij,2}}{\pi_{ij,2} + \pi_{ij,3} + \pi_{ij,4}}\right) &= \alpha_2 - \mathbf{x}_i'\beta_{(2)} \\ \text{Link}\left(\frac{\pi_{ij,3}}{\pi_{ij,3} + \pi_{ij,4}}\right) &= \alpha_3 - \mathbf{x}_i'\beta_{(3)} \end{aligned}$$

So if the continuation-ratio link model is considered to have three levels then each level models one of the binomial probabilities in the expanded likelihood. This means that we can use the method for fitting logistic regression models for binary data to fit continuation-ratio link models at each level without having to estimate the cut-point parameter in each category level. Alternatively, we can use the expanded likelihood in (**??**) to estimate the cut-point parameters $\alpha_j,\ j = 1, \ldots, r-1$ and the regression parameter vector β for all categories. Methods for fitting binary data using Gibbs sampler can be found in Zeger and Karim (1991) and Pang (1999) and methods for fitting ordinal data can be found in Albert and Chib (1993). A brief description of Markov chain Monte Carlo techniques, Gibbs sampler method and the adaptive rejection sampling method (ARS) is introduced in the next section.

3 Markov Chain Monte Carlo Techniques

Markov Chain Monte Carlo (MCMC) methodology provides enormous scope for realistic statistical modelling and has become popular recently, for Bayesian computation in complex statistical models. MCMC is essentially Monte Carlo integration using Markov chains. Bayesian analysis requires integration over possibly high-dimensional probability distributions to make inferences about model parameters or to make predictions. However in the past Bayesian inference has been hampered by the problem of evaluating the expectation of the posterior densities by numerical integration. This problem becomes more acute when Bayesian statisticians have to solve high dimensional integrals. Complex numerical integration methods such as the Gaussian quadrature and Laplace approximation (Tierney and Kadane (1986), Shun and McCullagh (1995)) are utilized. Monte Carlo integration draws samples from the required distribution, and then forms sample averages to approximate expectations. The Markov Chain Monte Carlo approach draws these samples by running a suitably constructed Markov chain for a long time. There are many ways of constructing these chains, but most of them, including the Gibbs sampler (Geman and Geman, 1984), are special cases of the general framework of Metropolis *et al.* (1953) and Hastings (1970). Many MCMC algorithms are hybrids or generalizations of the simplest methods: the Gibbs sampler and the Metropolis-Hastings algorithm.

3.1 The Metropolis-Hastings Algorithm

Constructing such a Markov chain is not difficult. We first describe the Metropolis-Hastings algorithm. This algorithm is due to Hastings (1970), which is a generalization of the method first proposed by Metropolis *et al.* (1953).
Let $p(\theta)$ be the distribution of interest. Suppose at time t, θ_{t+1} is chosen by first sampling a candidate point η from a proposal distribution $q(. \mid \theta_t)$. The candidate η is the accepted with probability

$$\alpha(\theta, \eta) = \min\left(1, \frac{p(\eta)q(\theta \mid \eta)}{p(\theta)q(\eta \mid \theta)}\right)$$

If the candidate point is accepted, the next state becomes $\theta_{t+1} = \eta$. If it is rejected, the chain does not move. The proposal distribution can be any kind of continuous probability density and the stationary distribution of the chain will be $p(\theta)$.
The Metropolis algorithm considers only symmetric proposals, that is, $q(\theta \mid \eta) = q(\eta \mid \theta)$. It is often convenient to choose a proposal which generate each component of η conditionally independently, given θ_t. Therefore the acceptance probability of accepting a candidate point η is given by

$$\alpha(\theta, \eta) = \min\left(1, \frac{p(\eta)}{p(\theta)}\right)$$

ISBN: 978-988-98671-8-8

3.2 The Gibbs sampler

Many statistical applications of MCMC have used Gibbs sampler, which is easy to implement. Gelfand and Smith (1990) gave an overview, and suggested the approach for Bayesian computation. First, probability densities of unknown parameter θ of interests are denoted as $p(\theta) = F'(\theta)$, $F(\theta)$ is the cumulative distribution function (CDF) of θ. Therefore, in the sequel, joint, conditional and marginal densities appear as $p(\theta, \eta)$, $p(\theta|\eta)$, and $p(\eta)$. Now the Gibbs sampling algorithm is best described as follows: Let $\underline{\theta} = (\theta_1, \theta_2, ..., \theta_k)$ be a collection of random variables. Given an arbitrary initial values ${\theta_1}^{(0)}, ..., \theta^{(0)}$, we draw ${\theta_1}^{(1)}$ from conditional distribution $p(X_1 \mid {\theta_2}^{(0)}, ..., {\theta_k}^{(0)})$, then ${\theta_2}^{(1)}$ from $p(\theta_2 \mid {\theta_1}^{(1)}, {\theta_3}^{(0)}, ..., {\theta_k}^{(0)})$ and so on up to ${\theta_k}^{(1)}$ from $p(\theta_k \mid {\theta_1}^{(1)}, ..., {\theta_{k-1}}^{(1)})$ to complete one iteration of the scheme. This scheme is a Markov chain, with equilibrium distribution $p(\underline{\theta})$. After t such iterations we would arrive at $({\theta_1}^{(t)}, {\theta_2}^{(t)}, ..., {\theta_k}^{(t)})$. Thus, for t large enough, $\underline{\theta}^{(t)}$ can be viewed as a simulated observation from $p(\underline{\theta})$. Provided, we allow a suitable burn-in time, $\theta^{(t)}, \theta^{(t+1)}, \theta^{(t+2)}, \ldots$ can be thought of as a dependent sample from $p(\theta)$.

Similarly, suppose we wish to estimate the marginal distribution of a variable η which is a function $g(\theta_1, \theta_2, ..., \theta_k)$ of $\underline{\theta}$. Evaluating g at each of the $\underline{\theta}^{(t)}$ provides a sample of η. Marginal moments or tail areas are estimated by the corresponding sample quantities. Densities may be estimated using kernel density estimates.

3.3 Sampling Methods in Gibbs Sampler

The Gibbs sampler involves sampling from full conditional distributions. It is essential that sampling from full conditional distributions is highly efficient computationally. Rejection sampling and the ratio-of-uniforms are two techniques for sampling independently from a general density $p(\theta)$ where $p(\theta)$ is intractable analytically.
(1) Rejection sampling method:–Rejection sampling requires an envelope function G of $p(\theta)$ where $g(\theta) \geq p(\theta)$ for all θ. Samples are drawn from density proportional to g, and each sampled point θ is subjected to an acceptance/rejection test.
(2) Ratio-of-uniforms method:–Ratio-of-uniforms method is to introduce two variables U and V. Let D denote a region in $\{U, V\}$ space defined by $0 \leq U \leq \sqrt{p(V/U)}$. Sample a point U, V uniformly from D. This can be done by enveloping the entire region of D by a region Δ. U and V can then be generated by rejection sampling.
(3) Adaptive rejection sampling method:–Adaptive rejection sampling (ARS) method is proposed by Gilks and Wild (1992). In the rejection sampling and ratio-of-uniforms sampling methods, finding a tight envelope function g or a envelope region Δ is difficult. These can also be very time consuming in the sampling stage. However in many applications of Gibbs sampling, the full conditional densities $p(\theta)$ are often log-concave (that is $\frac{d^2 lnp(\theta)}{d\theta^2} < 0$). Gilks and Wild (1992) proposed the adaptive rejection sampling method to sample from a complicated full conditional density which satisfies the log-concavity condition. They showed that an envelope function for $\log p(\theta)$ can be constructed by drawing tangents to $\log p$ at each abscissae for a given set of abscissae. An envelope between any two adjacent abscissae is then constructed from the tangents at either end of that interval. Secants are drawn through $\log p$ at adjacent abscissae. The envelope is piece-wise exponential for which sampling is straightforward. We will use the adaptive rejection sampling method over the rejection sampling or ratio-of-uniforms method as the ARS is more direct and efficient in terms of sampling from the full conditional density. One may refer to the article by Gilks and Wild (1992) for more theoretical details about their method. Also one may refer to the book by Devroye (1986) or to the monograph by Gilks *et al.* (1996) for more details on rejection sampling method, ratio-of-uniforms method and the Metropolis-Hastings method.

4 Continuation-Ratio Model to Fit The Telephone Connection Quality Data

At the laboratories of British Telecom (BT) in Martlesham near Ipswich (UK), a series of experiments concerning the quality of telephone connections were conducted. One of their experiments is called conversation experiment. A conversation experiment consists of a number of pairs of subjects and each pair engages in a conversation over the telephone. The two subjects in each pair sit in two different cabinets; say cabinet A and B. In a conversation experiment a subject engages in conversation and then gives an opinion about the telephone connection. The duration of conversation is determined by the subjects. When the conversation is finished the subjects hang up and are prompted by the experiment controller to give an opinion of the transmission condition. The opinion is typically given on a five point scale graded from 'Bad' to 'Excellent'. This is an ordinal response scale. The subjects also give a binary responses to a question on difficulty in hearing over the connection.
The order in which a subject hears the transmission conditions is determined by an experimental design. This design can be set out as a two-way layout in which each row corresponds to a subject and each column corresponds to a period. In each period there is a particular level of transmission conditions. The model is linear in two factors; namely (i) rows (random effects), and (ii) transmission conditions.
For the experiments that we analyzed, one is called E198 experiment. This experiment has an limited duration in the conversation between the two subjects. There are 2

pairs of 16 subjects. Each subject (row) received 8 trials. In each trial one level of the 8 transmission conditions is set.
Here we propose to use the continuation-ratio model to model the E198 data set using the expanded likelihood with one common set of regression parameters. In the following we present the final results of our analysis. As in simulation, 10,000 random variates are generated for each parameter after 1000 burn-in values. The results of the final estimates are shown in Table 1.

Table 4. Results Continuation-ratio model of E198 experiment.

para.	mean	s. e.	s. e.[1]	$P_{0.025}$	$P_{0.975}$
Intercept	-8.9036	0.3304	0.0074	-9.4885	-8.3079
Cond 1	aliased	-	-	-	-
Cond 2	0.2835	0.4716	0.0106	-0.6140	1.1834
Cond 3	0.3856	0.4910	0.0208	-0.6078	-1.3223
Cond 4	7.3293	0.5168	0.0116	6.3213	8.3255
Cond 5	0.1846	0.5054	0.0417	-0.8195	1.1492
Cond 6	2.6563	0.4944	0.0301	1.6598	3.5756
Cond 7	6.4176	0.5043	0.0113	5.4333	7.3969
Cond 8	8.0449	0.5117	0.0205	7.0102	8.9981
Cut 1	aliased	-	-	-	-
Cut 2	3.3255	0.6902	0.0753	2.4298	5.9761
Cut 3	6.7925	0.6384	0.0114	5.4651	7.8235
Cut 4	9.3560	0.7098	0.0374	7.5261	10.7382

1: This is the standard error of the batching means.

As we can see from Table 4, the standard deviation of batching variances are very small relatively to overall standard deviation of the sample variances in each chain. The number of sample points in each batch is 500. There are 20 batches. This indicates that convergence is good in each of the Gibbs sampling scheme. The last two columns in each of the Tables show the usual 2.5% and 97.5% quantile values as shown in last three chapters.
We can in fact compare our results with those obtained by Lewis *et al.* (1993). They also fitted the continuation-ratio model to the E198 data. But they only included 16 subjects in the model and the covariate sets are quite different from the one that we used in here. In Table **, we presented partially their results. It is noted that the estimates for the cut-points and conditions, to certain extent, are similar to our results. Lewis *et al.* (1993) used GLIM software package to fit the continuation-ratio model.

Table 5: Results of E198 experiment by Lewis *et al.* (1993)*

parameter	estimate	s. e.
Intercept	-12.68	1.488
Cond 1	aliased	-
Cond 2	0.5848	0.7317
Cond 3	1.067	0.7392
Cond 4	8.924	1.043
Cond 5	0.9571	0.7349
Cond 6	4.427	0.7999
Cond 7	7.506	0.9487
Cond 8	10.24	1.125
Cut 1	aliased	
Cut 2	3.764	0.5271
Cut 3	6.992	0.7302
Cut 4	10.27	0.8831

*:This table is extracted from Table C.1 of Lewis *et al.* (1993).

5 Conclusions

In this paper, we presented the estimation results of continuation-ratio model for a large set of simulated data and a set of ordinal data from the telecommunication experiment. For the experiment data set (BT E198 data), we used the expanded likelihood to estimate the overall parameters. The other way of parameter estimation is to fit a model for each level and there are four separate models with no cut-points. However our experiences indicate that after converting the data into binary form. That is we let $Y_{ij} = 1$ if $Y_{ij} = h$ and $Y_{ij} = 0$ if $Y_{ij} > h$. The ARS estimation procedure failed to converge in each of the four cases due to the sparseness of $'1'$ in each subject. Similar results are encountered when we use the BUGS program for the same data set.

The interpretations of the continuation-ratio model is different from the model for strictly binary data. However the estimation technique is the same. The proposed continuation-ratio model is a model which, in general, models any response of an ordered categorical nature. This implies that the model would be suitable to cope with the many other ordered categorical scales used in historical experiments. The traditional British Telecom method of analysis for opinion score responses has been to perform an analysis-of-variance on the numerical scores assigned to the categories (0 to 4). One of the assumptions underlying the analysis-of-variance procedure is that the response variable follows a Normal distribution. The opinion score is constrained to one of five values, i.e., it is a discrete response rather than a continuous one. Approximating a discrete response with five values by a Normal curve is rather a crude approach. Also the scores attributed to different categories are arbitrary. The two considerations have the consequence of inefficient estimation of parameters by the standard analysis-of-variance approach.

Wolfe (1996) used continuation-ratio model to fit the BT experimental data. He found that the residual sum of squares of the continuation-ratio models are in general smaller that the residual sum of squares as minimised in the analysis-of-variance approach. Thus the continuation-ratio logit model fitted the data better than the analysis-of-variance approach. The proposed residual sum of squares (Wolfe, 1996) for the continuation-ratio logit model (RSS_{CR}) is calculated as follows:

$$RSS_{CR} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

with $\hat{y}_i$ given by

$$\hat{y}_i = \sum_{j=0}^{r-1} j\hat{\pi}_{ij} \tag{7}$$

The fitted value $\hat{y}_i$ is calculated by multiplying the fitted probabilities from the continuation-ratio logit model $\hat{\pi}_{ij}$ by the scores $j = 0$ to 4 as used in the ANOVA model for the data, giving a fitted mean score for the continuation-ratio logit model.

The Bayesian method of modelling the continuation-ratio logit model using Markov chain Monte Carlo (MCMC) technique provides a good alternative way to model ordinal data. In the Bayesian approach, one can identify immediately those parameters which have significant effects by conveniently inspecting whether zero value is contained in the $(1 - \alpha) \times 100\%$ probability intervals. 95% probability intervals are adopted here. Another advantage of Bayesian modelling is that we are able to compute the predicted probabilities of future events for given generated parameter values. When one uses Markov chain Monte Carlo (MCMC) technique for parameter estimation, one can also obtain the $(1 - \alpha) \times 100\%$ probability intervals for the predicted probabilities of future events. More information about model prediction is incorporated naturally in the Bayesian approach. The model parameters and future value of the observations are random variables in full probability model under discussion.

One major contribution of this paper is by fitting continuation-ratio model, we can estimate the cut-points parameters in the cumulative logit model for ordinal data. Algorithms proposed by Albert and Chib (1993), Cowles (1996) and Nandram and Chen (1996) to find the cut-points become unnecessary. Often these algorithms required more computational effort. However there is a drawback in continuation-ratio modelling. If there are not enough $'1'$s in the converted binary data set in any particular category (level), then the Gibbs sampler scheme will fail in the continuation-ratio logit model for that level. We have to resort to estimate parameters in the expanded model in (6) where we need to estimate each of the cut-points from the model (see Section 5). Otherwise modelling the continuation-ratio link models (logit, probit, complementary log-log) for ordinal data using Gibbs sampler method is an easy alternative over the usual cumulative probability link (logit, probit, complementary log-log) model.

References

[1] Agresti, J. H. (1996). *An Introduction to Categorical Data Analysis.* Wiley, New York.

[2] Albert, J. H. and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, **88**, 669-679.

[3] Agresti, J. H. and Lang J. B. (1993). A Proportional Odds Model with Subject-Specific Effects for Repeated Ordered Categorical Responses. *Biometrika*, **80**, 527-534.

[4] Albert, J. and Chib, S. (1995). Bayesian Residual Analysis for Binary Response Regression Models. *Biometrika*, **82**, 747-759.

[5] Cowles, M. K. (1996). Assessing and Accelerating Gibbs Sampler Convergence for Cumulative-Link Generalized Linear Models. *Statistics and Computing,* **6**, 101-111.

[6] Devroye L. (1986). *Non-Uniform Random Variate Generation.* Springer-Verlag, Berlin.

[7] Gelfand, A.E and Smith, A.F.M. (1990). Sampling Based Approach to Calculating Marginal Densities. *J. Am. Statist. Ass.*, **85**, 398-409.

[8] Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images . *IEEE Trans. Pattn. Anal. Math. Intell.*, **6**, 721-741.

[9] Gilks W. R., Best N. G. and Tan K. K. C. (1995). Adaptive Rejection Metropolis Sampling within Gibbs Sampling. *Appl. Statist.* **44**, 455-472.

[10] Gilks W. R. and Wild P. (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Appl. Statist.* **41**, 337-348.

[11] Gilks W. R. (1992). Derivative-free adaptive rejection sampling for Gibbs Sampling. In *Bayesian Statistics* **4**, (eds J. M. Bernardo, J. O. Berger, A.

P. Dawid, and A. F. Smith), pp 642-649. Oxford: Oxford University Press.

[12] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika* **57** 97-109.

[13] Jansen, J. (1990). On The Statistical Analysis of Ordinal Data when Extravariation is present. *Applied Statistics* **39**, 75-84.

[14] Lewis S. M., Firth D. and Russel K. G. (1993). Carry-over effects in speech-transmission experiments. *Progress Report* Department of Mathematics, University of Southampton.

[15] McCullagh, P. (1977). A logistic model for paired comparisons with ordered categorical data. *Biometrika*, **64**, 449-453.

[16] McCullagh, P. (1980). *Regression Models for Ordinal Data*. 2nd Edition. London, Chapman and Hall.

[17] McCullagh, P. (1978). A Class of Parametric Models for the Analysis of Square Contingency Tables with Ordered Categories. *Biometrika*, **65**, 413-418.

[18] McCullagh, P. (1980). Regression Models for Ordinal Data. *J. R. Statist. Soc.* **B, 42**, 109-142.

[19] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* 2nd Edition. London, Chapman and Hall.

[20] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machine *J. Chem. Phys.,* **21**, 1087-1091.

[21] Nandram, B. and Chen, M. H. (1996). Reparameterizing The Generalized Linear Model To Accelerate Gibbs Sampler Convergence. *J. Statist. Comput. Simul.*, **54**, 129-144.

[22] Pang W. K. (1999). Modelling Binary Data: A Gibbs Sampling Approach. *Journal of Statistical Computation and Simulation.*, **64**, 73-85.

[23] Pang W. K. (2000). Modelling Ordinal Categorical Data: A Gibbs Sampling Approach. *PhD Thesis.* Department of Marthematics, University of Southampton.

[24] Rubin, D. (1987). Comment on Paper by Tanner and Wong. *J. Am. Statist. Ass.*, **82**, 543-546.

[25] Shun Z. and McCullagh P. (1995). Laplace Approximation of High Dimensional Integrals. *J. R. Statist. Soc.* B Vol. **57**, 749-760.

[26] Stram, D. O., Wei, L. J. and Ware, J. H. (1988). Analysis of repeated ordered categorical outcomes with possibly missing observations and time dependent covariates. *J. Am. Statist. Ass.* **83**, 3631-637.

[27] Wolfe R. (1996). Models and Estimation for Repeated Ordinal Responses, with Application to Telecommunication Experiments. *PhD Thesis*, Department. of Mathematics, University of Southampton.

[28] Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Am. Statist. Ass.* **81**, 82-90.

[29] Zeger, S. L. and Karim, M. R. (1991). Generalized Linear Models with Random Effects: a Gibbs Sampling Approach. *J. AM. Statist. Ass*, **86**, 79-86.