A Note on the Existence of the Posterior Distribution for a Class of Mixed Models for Binomial Responses

BY RANJINI NATARAJAN

School of Operations Research and Industrial Engineering,

206 Engineering & Theory Center,

Cornell University, Ithaca, NY 14853

AND CHARLES E. MCCULLOCH

Biometrics Unit and Statistics Center,

337 Warren Hall,

Cornell University, Ithaca, NY 14853

BU-1270-MA

February 1995

Abstract

Necessary and sufficient conditions are given for the existence of the posterior distribution of the variance components in a class of mixed models for binomial responses. The implications of our results are illustrated through an example.

Some key words: Gibbs sampler; Improper Prior; Linear Programming; Logit; Mixed Model; Probit; Propriety; Variance Components.

1 Introduction

The question of the integrability of the posterior distribution arises when one imposes improper prior distributions on the parameters. Improper priors may be used for a variety of reasons in Bayesian analyses. In hierarchical models, one might impose improper prior distributions due to the absence of information on the hyperparameters at the lower levels of the hierarchy. In multi-parameter situations, elicitation of prior information and subsequent formulation into a distribution can be a difficult task. In such cases one might again consider analyses with improper priors to reflect vague information (Ibrahim and Laud, 1991). Improper priors may also be used in a frequentist context due to the equivalence of flat prior Bayes and maximum likelihood estimation.

Although a fair amount of work has focussed on studying the existence of maximum likelihood estimates for various models (Silvapulle (1981), Albert and Anderson (1984), Geyer and Thompson (1992)), very little has been done by way of verifying the existence of posterior distributions resulting from improper priors. We investigate conditions under which a class of improper priors on the variance components leads to proper posterior distributions for mixed models for binomial responses, specifically the logit-normal and probit-normal regression models. We are not concerned with the analytic tractability of the posterior, but rather its existence. Our conditions are very similar to those developed by Albert and Anderson (1984) on the existence of maximum likelihood estimates for the logit and probit models.

Our results have implications for the use of Monte Carlo Markov chain methods, such as the Gibbs sampler, to perform Bayesian analysis of these models. It is common in analyzing these models to impose improper priors on the parameters (Karim and Zeger, 1992). However, such priors do not necessarily lead to proper posterior distributions, even when they result in proper full conditional distributions. The use of the sampler in such situations can give seriously misleading results.

In Section 2 we formulate the model and state the main result. In Section 3 we illustrate the implications of our result through an example.

2 The Model

Let $w_1, ..., w_N$ be a set of N correlated binary observations. A flexible class of models can be generated by linking the mean of w_i to the fixed and random effects. More formally, conditional on the vector of random effects u, the (w_i) are independent with

$$E(w_i \mid u) \;=\; h(x_i \beta + z_i u), \text{ while } u \sim \mathcal{N}_q(0, \theta I), \qquad (1)$$

where $h(.)$ is a distribution function. We are particularly interested in $h(.)$ corresponding to the logistic and normal distributions, which lead to the logit-normal and probit-normal models respectively. The random effects u serve as a convenient way to specify the correlation between the w. They are also useful for prediction purposes (Harville and Mee, 1984).

2.1 Bayesian Hierarchy

We consider the following Bayesian hierarchical specification:

$$[w_i \mid u] \sim \texttt{Bernoulli}\ \{h(x_i\,\beta + z_i\,u)\}$$

$$[u \mid \theta] \sim N_q\,(0,\, \theta\, I)$$

$$[\theta \mid a] \propto \frac{1}{\theta^{a+1}}$$

where the square brackets [.] denote probability density or mass functions and a is a pre-specified constant characterizing the prior distribution of θ. Note that when $a = 0$ we have the classic non-informative prior on a normal variance (Box and Tiao, 1992 p58). Since our focus is on improper priors for the variance components, we assume β known. However, our results hold even if β is unknown, so long as we assign it a proper prior.

Let X be the $N \times p$ known design matrix, with rows x_i, and Z the $N \times q$ incidence matrix, with rows z_i. Define X^* as the matrix with rows $x_i^* = -x_i$ if $w_i = 1$, and $x_i^* = x_i$ if $w_i = 0$, and define Z^* similarly. The posterior distribution is given by:

$$[\theta \mid w_1, ..., w_N] = \frac{\int L\,(\beta,\, u \mid w_1, ..., w_N)\ [u \mid \theta]\,[\theta \mid a]\,du}{\int L\,(\beta,\, u \mid w_1, ..., w_N)\ [u \mid \theta]\,[\theta \mid a]\,du\,d\theta} \qquad (2)$$

where $L(\beta, u \mid w_1, ..., w_N) = \prod_{i=1}^{N} \{1 - h(x_i^* \beta + z_i^* u)\}$. It is clear that the posterior distribution of θ exists if and only if the integral in the denominator of (2) converges.

While using a data augmentation approach such as the Gibbs sampler to perform a Bayesian analysis of this model, it is typical to impose improper priors on the parameters. Karim and Zeger (1992) show that the full conditional specifications for logistic normal regression, using non-informative priors, are all proper distributions and relatively easy to generate from. Thus, implementation of the Gibbs sampler appears straightforward and computationally attractive. However, improper priors do not always lead to proper posterior distributions. We now state a theorem that guarantees the propriety of the posterior distribution of the variance components.

2.2 Existence Theorem

Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be the polyhedral cones defined by

$$\mathcal{C}_1 = \{\alpha : Z^* \alpha \leq 0\},$$

$$\mathcal{C}_2 = \{\alpha : (X^* \beta + Z^* \alpha) \leq 0\}.$$

Define conditions Π_1 and Π_2 as follows:

$$\Pi_1 \ : \ \text{dimension}(\mathcal{C}_1) < q,$$

$$\Pi_2 \ : \ \text{dimension}(\mathcal{C}_2) < q.$$

Our main result is as follows:

Theorem 1. For the model (1):

(i) The posterior distribution of θ exists only when Π_1 is satisfied and $-\frac{q}{2} < a < 0$.

(ii) When $h(.)$ is the logit or probit function, the posterior distribution of θ exists if Π_2 is satisfied and $-\frac{1}{2} < a < 0$.

The proof is given in the Appendix.

The conditions on the constant a stem from the contribution of the prior distribution to the posterior, while conditions Π_1 and Π_2 arise from the likelihood function. It is interesting to note that the classic non-informative prior on a normal variance, that is $a = 0$, does not lead to proper posterior distributions for this model.

Albert and Anderson (1984) developed conditions similar to Π_1 for the existence of maximum likelihood estimates of the fixed effects, for the logistic regression model. They proved that the maximum likelihood estimates exist if and only if there does not exist a non-zero α such that $X^* \alpha \leq 0$. Their interpretation of this condition has its roots in the regression and discrimination literature. For the purely fixed effects case this condition implies that the data set is overlapped. Although their condition appears to be simpler than Π_1 or Π_2, it is actually much more restrictive as illustrated in Section 3. It cannot be verified directly using a standard linear programming package, and needs to be reformulated in order to be solved. Santner and Duffy (1986) presented a mixed linear program to verify their condition.

We now discuss a method to verify conditions of the form Π_1 or Π_2. We show that our conditions reduce to checking the feasibility of a system of linear equations, which is a standard problem in the linear programming literature.

2.3 When is a polyhedral cone full-dimensional?

We say a cone in $\Re^n$ is full-dimensional if it has a non-empty interior. It is easy to see that the cone $\mathcal{C} = \{x \in \Re^n : A\,x \leq 0\}$ is full-dimensional if and only if the system of equations $A\,x < 0$ has a solution. By Farkas' lemma (1902), it follows that $\mathcal{C}$ is full-dimensional if and only if there does not exist a non-negative vector $y \in \Re^n \, (y \neq 0)$ such that $y\, A = 0$. Thus for verifying Π_1 or Π_2, it suffices to find such a y for $\mathcal{C}_1$ or $\mathcal{C}_2$. This is a standard linear programming problem which can be done using commercially available software, for example CPLEX.

3 Example

We consider the following mixed model with a single nested random effect:

$$E(w_{ij} \,|\, u_i) \;=\; h(\beta + u_i), \;\; j = 1, \ldots, k, \; i = 1, \ldots, q$$

$$\text{and} \quad u_i \;\sim\; N\,(0,\, \theta), \quad i = 1, \ldots, q$$

This model corresponds to k repeat observations being taken on each of the q levels of a single random effect. In this context the design matrix $X = 1_{q\,k}$, a $q\,k$ dimensional column vector of ones, and the incidence matrix is the direct product $Z = I_q \otimes 1_k$.

We first discuss condition Π_1. The cone $\mathcal{C}_1$ is $\{\alpha : (1 - 2\,w_{ij})\,\alpha_i \leq 0, \; i = 1,\ldots,q, \, j = 1,\ldots,k\}$. If for each i the outcomes are all successes or all failures, then $\mathcal{C}_1$ is the product of q half lines: $(-\infty, 0\,]$ for levels with only failures, and $[0, \infty)$ for levels with only successes. Then $\mathcal{C}_1$ is full-dimensional. However, if for some i, there are distinct indices j, j^* such that $w_{ij} = 1$, $w_{ij^*} = 0$, then $\alpha_i = 0$ for any $\alpha \in \mathcal{C}_1$, thereby decreasing the dimension of $\mathcal{C}_1$ by one. Thus, condition Π_1 requires that there be a success and a failure for at least one level of the random effect, to ensure the propriety of the posterior distribution of the variance component.

We now discuss the sufficient condition Π_2. We have $\mathcal{C}_2 = \{\alpha : (1 - 2\,w_{ij})\,(\beta + \alpha_i) \leq 0,\ \ i = 1, \ldots, q,\ j = 1, \ldots, k\}$. Again, $\mathcal{C}_2$ is full-dimensional if at each level of the random effect we have all successes or all failures; if there is at least one level i for which there is a success and a failure, then $\mathcal{C}_2$ is less than full-dimensional due to the binding constraint $\beta + \alpha_i = 0$. Thus, condition Π_2 states that it is sufficient to have a success and a failure for at least one level of the random effect to ensure proper posterior distributions.

For this simple one-way analysis of variance model, with logit or probit link, we have shown that, to ensure proper posterior distributions, it is both necessary and sufficient to have a success and a failure for at least one level of the random effect. This condition is less restrictive than the usual likelihood one, for the fixed effects case, which requires that, for the maximum likelihood estimates to exist, there must be at least one success and one failure for every level of the random effect (Albert and Anderson, 1984).

ACKNOWLEDGEMENTS: We wish to thank Professor Samorodnitsky, Cornell University, for his invaluable help and boundless patience. Thanks are also due to Dr. A. P. Dawid and a referee for several helpful comments.

Appendix

Denote the denominator in (2) by I. The posterior distribution $[\theta \mid w_1, ..., w_N]$ is proper if and only if I converges.

(i) Necessity: We first show that the condition $-\frac{q}{2} < a < 0$ is necessary by re-writing I as:

$$I \propto \int_\theta \int_u \prod_{i=1}^{N} \{1 - h(x_i^* \beta + z_i^* u)\} \, exp(-\frac{u' u}{2 \theta}) \, du \, \frac{d\theta}{\theta^{a + \frac{q}{2} + 1}} \quad (3)$$

$$= \int_\theta \int_v \prod_{i=1}^{N} \left\{1 - h(x_i^* \beta + \theta^{1/2} z_i^* v)\right\} \, exp(-\frac{v' v}{2}) \, dv \, \frac{d\theta}{\theta^{a + 1}} \quad (4)$$

where (4) follows from (3) on making the change of variable $v_i = u_i \, \theta^{-1/2}$, $i = 1, ..., q$. If $a > 0$ we see from (4) that the integrand diverges for θ in a neighbourhood of zero, whilst if

$a + \frac{q}{2} < 0$, we see from (3) that the integrand diverges for large θ.

We now prove that Π_1 is necessary. Suppose that Π_1 is not satisfied. Since the integrand in (4) is non-negative it is clear that

$$I \;\geq\; \int_\theta \int_{\mathcal{C}_1} \prod_{i=1}^{N} \{1 \,-\, h(x_i^* \,\beta)\} \; exp(-\frac{v'\,v}{2})\, dv \, \frac{d\theta}{\theta^{a+1}} \tag{5}$$

where we have also used the fact that $h(.)$ is monotone and θ is non-negative. The right hand side of (5) diverges due to the integral over θ.

(ii) Sufficiency: Integrating (3) over θ we have:

$$I \;\propto\; \int_u \prod_{i=1}^{N} \{1 - h(x_i^* \,\beta \,+\, z_i^* \,u)\} \; \frac{du}{(u'\,u)^{(a+\frac{q}{2})}} \;\; . \tag{6}$$

If Π_2 holds, then for every u there exists some index $j_u \in \{1, \ldots, N\}$ such that $x_{j_u}^* \,\beta \,+\, z_{j_u}^* \,u \,>\, 0$. Thus, we can bound I in the following way:

$$I \;\leq\; \sum_{j=1}^{N} \int_{\{u\,: x_j^*\,\beta \,+\, z_j^*\,u \,>\, 0\}} \left\{1 \,-\, h(x_j^* \,\beta \,+\, z_j^* \,u)\right\} \frac{du}{(u'\,u)^{(a+\frac{q}{2})}}$$

where we have also used the fact that $1 - h(.) \leq 1$ to retain only one term in the product in (6). It therefore suffices to inspect the convergence of

$$I_j \;=\; \int_{\{u\,: x_j^*\,\beta \,+\, z_j^*\,u \,>\, 0\}} \left\{1 \,-\, h(x_j^* \,\beta \,+\, z_j^* \,u)\right\} \frac{du}{(u'\,u)^{(a+\frac{q}{2})}} \;\; . \tag{7}$$

If $h(.)$ is the logit link, it can be seen readily that

$$I_j \;\le\; \int_{\{u \,: x_j^* \beta \,+\, z_j^* u \,>\, 0\}} exp\Big\{-(x_j^*\,\beta \,+\, z_j^*\,u)\Big\}\,\frac{du}{(u'\,u)^{(a\,+\,\frac{q}{2})}}\;.$$

We can assume without loss of generality that $z_j^* = (1,\, 0,\, \dots, 0)$ since this merely corresponds to the transformation $v_1 = z_j^*\, u$ and $v_k = u_k,\; k = 2, \dots, q$. Thus,

$$I_j \;\le\; \int_{\{u_1 \,: x_j^* \beta \,+\, u_1 \,>\, 0\}} \int_{u_2} \cdots \int_{u_q} exp\Big\{-(x_j^*\,\beta \,+\, u_1)\Big\}\,\frac{du}{(u'\,u)^{(a\,+\,\frac{q}{2})}}\;. \quad (8)$$

Make the change of variable $u_k = y_k\, u_1,\; k = 2, \dots, q$ in (8) to obtain:

$$I_j \;\le\; \int_{\{u_1 \,: x_j^* \beta \,+\, u_1 \,>\, 0\}} exp\Big\{-(x_j^*\,\beta \,+\, u_1)\Big\}\,\frac{du_1}{u_1^{2a+1}} \int_{y_2} \cdots \int_{y_q} \frac{dy_2 \dots dy_q}{(1 \,+\, y_2^2 \,+\, \dots \,+\, y_q^2)^{a\,+\,\frac{q}{2}}}\;.$$

The integral over u_1 converges since $a < 0$, while the integral over y can be transformed to

$$\int_0^\infty \frac{r^{q-2}\,dr}{(1 + r^2)^{a\,+\,\frac{q}{2}}}\;,$$

on using spherical co-ordinates. This integral converges if $a > -\frac{1}{2}$.

If $h(.)$ is the probit link, then the same proof applies after bounding $\{1 - h(.)\}$ by $exp\Big\{-\frac{1}{2}(.)^2\Big\}$ in (7) since it is easily shown that, for T standard normal,

$$P(T > \lambda) \le exp(-\tfrac{\lambda^2}{2}),\; \lambda > 0\,.$$

References

ALBERT, A. & ANDERSON, J. A. (1984) On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. *Biometrika* **71**, 1--10.

BOX, G. E. P. & TIAO, G. C. (1992) *Bayesian Inference in Statistical Analysis.* Canada: Wiley & Sons.

CPLEX, CPLEX OPTIMIZATION, INC. (1989 - 1994)

FARKAS, J. (1902) Theorie der einfachen Ungleichungen. *Journal für die reine und angewandte Mathematik* **124**: 1--27.

GEYER, C. J. AND THOMPSON, E. A. (1992) Constrained Monte Carlo Maximum Likelihood for Dependent Data. *Journal of the Royal Statistical Society, Series B* **54**, 657-699.

HARVILLE, D. A. & MEE, R. W. (1984) A Mixed-Model Procedure for Analyzing Ordered Categorical Data. *Biometrics* **40**, 393--408.

IBRAHIM, J. G. & LAUD, P. W (1991) A Bayesian Analysis of Generalized Linear Models using Jeffrey's Prior. *Journal of the American Statistical Association* **86**, 981--986.

KARIM, M. R. & ZEGER, S. L. (1992) Generalized Linear Models with Random Effects; Salamander Mating Revisited. *Biometrics* **48**, 631--644.

SANTNER, T. J. & DUFFY, D. E. (1986) A Note on A. Albert and J.

A. Anderson's Conditions for the Existence of Maximum Likelihood Estimates in Logistic Regression Models. *Biometrika* **73**, 755--758.

SILVAPULLE, M. J. (1981) On the Existence of Maximum Likelihood Estimates for Binomial Response Models. *Journal of the Royal Statistical Society, Series B* **43**, 310--313.