The Korean Communications
in Statistics Vol. 7, No. 3, 2000
pp. 817-827

# Bayesian Analysis for Random Effects Binomial Regression

**Dal Ho Kim[1], Eun Young Kim[2]**

## Abstract

In this paper, we investigate the Bayesian approach to random effect binomial regression models with improper prior due to the absence of information on parameter. We also propose a method of estimating the posterior moments and prediction and discuss some general methods for studying model assessment. The methodology is illustrated with Crowder's Seeds Data. Markov Chain Monte Carlo techniques are used to overcome the computational difficulties.

*Keywords* : Binomial regression, improper prior, model assesment, random effects

## 1. Introduction

In some regression problems, the response variable is categorical, often either success or failure. For such problem, the normal linear model is sure to be inappropriate, because normal errors do not correspond to a zero/one response. One important method that can be used in this situation is called binomial regression. It is one special case of the class of generalized linear models (GLMs) first proposed by Nelder and Wedderburn(1972). One useful extension involves models with random effects in the linear predictor. This is so-called generalized linear mixed models (GLMMs). GLMMs are known to be useful for accommodating the overdispersion often observed among outcomes; for modeling the dependence within clusters; for producing shrinkage estimates in multiparameter problems, such as the construction of maps of small area disease rates; and for smoothing of regression relationships.

As Nelder(1972) has pointed out in his discussion of the article by Lindley and Smith(1972) on Bayesian methods in regression, there is a strong connection between the random effects and Bayesian regression models. In Bayesian analysis, improper prior may be used for a variety of reasons. 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 multiparameter situations, elicitation of prior information and subsequent formulation into a

1) Assistant Professor, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea. E-mail : dalkim@bh.knu.ac.kr
2) Lecturer, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea. E-mail : yw20120723@hanmail.net

distribution can be a difficult task. In such cases one might again consider analysis with improper priors to reflect vague information (Ibrahim and Laud, 1991). When one imposes improper prior distributions on the parameter, the question of the properpriety of the posterior distribution arises.

Leonard(1972) discussed Bayesian hierarchical models for binomial data. Zellner and Rossi (1984) gave an overview of Bayesian methods for binomial regression models. Bedrick, Christensen and Johnson(1997) have provided a complete discussion of Bayesian inferences for binomial regression without the random effect. Breslow and Clayton(1993) gave an overview of frequentist analysis for GLMMs. Zeger and Karim(1991) discussed GLMMs in Bayesian perspective. But we are unable to verify that the prior considered by Zeger and Karim necessarily leads to a proper posterior.

In this paper, we investigate the Bayesian approach to binomial regression models with improper prior due to the absence of information on parameter. The methodology is illustrated with Crowder's Seeds Data. A concern in this data was the presence of extra-binomial variation. One way to explain this overdispersion is through random differences in the rate of germination between plates. So, this analysis is done for the random effects model. Extensions of generalized linear models to include random effects have been hampered by the need for numerical integration. In this paper, we adopt a Markov Chain Monte Carlo (MCMC) method to overcome the computational limitations.

The outline of the remaining sections is as follows. In Section 2, we introduce a random effects binomial regression models and find sufficient conditions for the propriety of posteriors under improper priors. Also we briefly discuss the application of the Gibbs sampler to our setting. In Section 3, we propose a method of estimating the posterior moments and prediction. In Section 4, we discuss some general methods for studying model assessment. This section also demonstrates how they can be applied to the proposed models. Finally, Section 5 provides the Bayes analysis of a real dataset, and compares it with other methods.

## 2. Random Effects Binomial Regression Models

Consider regression data $(n_i, y_i, \boldsymbol{x}_i)$, $i = 1, \ldots, N$, where the $y_i$'s are number of successes from independent binomial $n_i$ random variables and $\boldsymbol{x}_i$'s are known vectors of covariates, that is $y_i \sim$ Binomial$(n_i, p_i)$.

We write $\boldsymbol{y} = (y_1, \cdots, y_N)^T$ and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_N)^T$. Also, we write $\boldsymbol{X}^T = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N)$ and assume that rank $(\boldsymbol{X}) = p$.

Then the likelihood for the data is

$$L(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{N} \binom{n_i}{y_i} [F(\theta_i)]^{y_i} [1 - F(\theta_i)]^{n_i - y_i},$$

where the $\theta_i$'s are modelled as

$$\theta_i = F^{-1}(p_i) = \boldsymbol{x}_i^T \boldsymbol{\beta} + u_i ,$$

and $F(\cdot)$ corresponds to either the logistic, probit or complementary log-log models, that is

$$F(\theta_i) = \begin{cases} \dfrac{e^{\theta_i}}{1+e^{\theta_i}}, & \text{logistic} \\ \Phi(\theta_i), & \text{probit} \\ 1-\exp[-e^{\theta_i}], & \text{complementary } \log-\log . \end{cases}$$

The link function is $F^{-1}(p) = \theta$, where $F^{-1}(p) = \log[p/(1-p)]$, $\Phi^{-1}(p)$, and $\log[-\log(1-p)]$ for the three models, respectively. $\boldsymbol{\beta}(p\times 1)$ is the vector of unknown regression coefficients, and $u_i$ is the random effect associated with $i$th subject. It is assumed that the $u_i$ are i.i.d. $N(0, \sigma_u^2)$. Finally, it is assumed that $\boldsymbol{\beta}$ and $r_u = (\sigma_u^2)^{-1}$ are marginally independent with $\boldsymbol{\beta} \sim \text{uniform}(R^p)$ and $f(r_u) \propto \exp(-a/2\, r_u)\, r_u^{b/2-1}$, with $a > 0$ and $b > 0$. This is referred to as a $\text{Gamma}(a/2, b/2)$ prior.

Then the joint posterior density of $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $r_u$ given $\boldsymbol{y}$ under the link function $F^{-1}(\cdot)$ is given by

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\beta}, r_u \mid \boldsymbol{y}, F) \propto & \prod_{i=1}^{N} \binom{n_i}{y_i} [F(\theta_i)]^{y_i} [1-F(\theta_i)]^{n_i-y_i} \\ & \times \prod_{i=1}^{N} \left( r_u^{\frac{1}{2}} \exp\left[ -\frac{r_u}{2}(\theta_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2 \right] \right) \cdot e^{-\frac{a}{2} r_u} r_u^{\frac{b}{2}-1} . \end{aligned} \tag{2.1}$$

A necessary and sufficient condition for the propriety of the posterior in the binomial case is provided in Natarajan and McCulloch(1995) when $a = 0$, but $\boldsymbol{\beta}$ is known. The following theorem provides sufficient conditions under which the joint posterior of $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $r_u$ given $\boldsymbol{y}$ is proper.

**Theorem 1.** Let $\theta_i \in (\underline{\theta}_i, \overline{\theta}_i)$ $(i = 1, \cdots, N)$. Assume $a > 0$. Suppose

$$I_i = \int_{\underline{\theta}_i}^{\overline{\theta}_i} \binom{n_i}{y_i} [F(\theta_i)]^{y_i} [1-F(\theta_i)]^{n_i-y_i} \, d\theta_i < \infty \tag{2.2}$$

for at least one $i$. Also, let $s + b > p$, where $s$ is number of indices which $I_i < \infty$. Then the joint posterior of $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $r_u$ given $\boldsymbol{y}$ is proper.

**Proof.** Consider the joint posterior of $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $r_u$ given $\boldsymbol{y}$ is given by (2.1). First integrate out those $\theta_i$ for which $I_i$ given in (2.2) is infinite. Let $\boldsymbol{\theta}^*$ denote the set of $\theta_i$'s for which $I_i$ is finite. By assumption $\boldsymbol{\theta}^*$ has at least one element. Then, the joint posterior of $\boldsymbol{\theta}^*$, $\boldsymbol{\beta}$ and $r_u$ given $\boldsymbol{y}$ is

$$p(\boldsymbol{\theta}^*, \boldsymbol{\beta}, r_u | \boldsymbol{y}, F) \propto \prod_{i: T_i < \infty} \binom{n_i}{y_i} [F(\theta_i)]^{y_i} [1-F(\theta_i)]^{n_i - y_i} \times r_u^{\frac{s}{2}} \exp\left[ -\frac{r_u}{2} \sum_{i: T_i < \infty} (\theta_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2 \right] \cdot e^{-\frac{a}{2} r_u} r_u^{\frac{b}{2}-1}.$$

First integrating with respect to $\boldsymbol{\beta}$ and then integrating with respect to $r_u$, it follows that

$$p(\boldsymbol{\theta}^* | \boldsymbol{y}) \le K \prod_{i: T_i < \infty} \binom{n_i}{y_i} [F(\theta_i)]^{y_i} [1-F(\theta_i)]^{n_i - y_i},$$

where $K(>0)$ is a constant which does not depend on the $\theta_i$. From the assumption of the theorem, it follows now that $\int p(\boldsymbol{\theta}^* | \boldsymbol{y})\, d\boldsymbol{\theta}^* < \infty$.

The hierarchical Bayes (HB) method is implemented via the MCMC integration technique. This requires generation of samples from full conditionals for computing marginal posterior distributions of $\boldsymbol{\beta}$, $r_u$ and $\theta_i$'s respectively. The required full conditionals are

(i) $\boldsymbol{\beta} | \boldsymbol{\theta}, r_u, \boldsymbol{y} \sim N((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\theta},\ r_u^{-1}(\boldsymbol{X}'\boldsymbol{X})^{-1})$

(ii) $r_u | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{y} \sim \text{Gamma}\left(\frac{1}{2}\{a + \sum_{i=1}^{N} (\theta_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2\},\ \frac{1}{2}\{b+N\}\right)$ (2.3)

(iii) $\pi(\theta_i | \boldsymbol{\beta}, r_u, \boldsymbol{y}) \propto [F(\theta_i)]^{y_i} [1-F(\theta_i)]^{n_i - y_i} \cdot \exp\left[ -\frac{r_u}{2} (\theta_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2 \right]$.

It is easy to generate samples from the full conditionals given in (i) and (ii). However, (iii) is not a standard density from which one can generate samples easily. This difficulty is overcome by employing the Metropolis-Hastings algorithm (Chib and Greenberg, 1995). An alternative approach would be to use the adaptive rejection sampling (ARS) of Gilks and Wild (1992) since $\pi(\theta_i | \cdot)$ are all log-concave. But the latter is not pursued here.

## 3. Bayesian Inference

### 3.1 Estimation of Posterior Moments

If now $[\boldsymbol{\theta}, \boldsymbol{\beta}, r_u | \boldsymbol{y}]$ denotes the joint posterior density of $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $r_u$ given $\boldsymbol{y}$, the posterior pdf of $\boldsymbol{\beta}$ given $\boldsymbol{y}$, $[\boldsymbol{\beta} | \boldsymbol{y}]$ is given by

$$[\boldsymbol{\beta} | \boldsymbol{y}] = \int\int [\boldsymbol{\theta}, \boldsymbol{\beta}, r_u | \boldsymbol{y}]\, d\boldsymbol{\theta}\, dr_u = \int\int [\boldsymbol{\beta} | \boldsymbol{\theta}, r_u, \boldsymbol{y}] [\boldsymbol{\theta}, r_u | \boldsymbol{y}]\, d\boldsymbol{\theta}\, dr_u.$$

In order to compute the posterior pdf $[\boldsymbol{\beta} | \boldsymbol{y}]$, one typically first derives $[\boldsymbol{\beta} | \boldsymbol{\theta}, r_u, \boldsymbol{y}]$ and then $[\boldsymbol{\theta}, r_u | \boldsymbol{y}]$. The distribution of $\boldsymbol{\beta}$ given $\boldsymbol{\theta}$, $r_u$ and $\boldsymbol{y}$ is given in (2.3). The second density in the integrand of $[\boldsymbol{\beta}|\boldsymbol{y}]$, $[\boldsymbol{\theta}, r_u|\boldsymbol{y}]$ is

$$[\boldsymbol{\theta}, r_u|\boldsymbol{y}] = \int [\boldsymbol{\theta}, \boldsymbol{\beta}, r_u|\boldsymbol{y}]\, d\boldsymbol{\beta}$$

$$\propto \prod_{i=1}^{N} \left\{ \binom{n_i}{y_i} [F(\theta_i)]^{y_i} [1-F(\theta_i)]^{n_i - y_i} \right\} \cdot e^{-\frac{a}{2} r_u} r_u^{\frac{N+b-p}{2}-1}.$$

Hence, the posterior mean for the $\boldsymbol{\beta}$ is now obtained as

$$E[\boldsymbol{\beta}|\boldsymbol{y}] = E[E(\boldsymbol{\beta}|\boldsymbol{\theta}, r_u, \boldsymbol{y})|\boldsymbol{y}] = E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\theta}|\boldsymbol{y}], \tag{3.1}$$

and the covariance matrix for the $\boldsymbol{\beta}$ is

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{\beta}|\boldsymbol{y}) &= E[\mathrm{Cov}(\boldsymbol{\beta}|\boldsymbol{\theta}, r_u, \boldsymbol{y})|\boldsymbol{y}] + \mathrm{Cov}[E(\boldsymbol{\beta}|\boldsymbol{\theta}, r_u, \boldsymbol{y})|\boldsymbol{y}] \\ &= E[r_u^{-1}(\boldsymbol{X}'\boldsymbol{X})^{-1}|\boldsymbol{y}] \\ &\quad + E[\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\theta}\} \cdot \{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\theta}\}'|\boldsymbol{y}] \\ &\quad - E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\theta}|\boldsymbol{y}] \cdot E[\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\theta}\}'|\boldsymbol{y}]. \end{aligned} \tag{3.2}$$

To estimate the posterior moments using Gibbs sampling, we use the Rao-Blackwellized estimates as in Gelfand and Smith (1991). So, the posterior mean given in (3.1) is approximated by

$$E[\boldsymbol{\beta}|\boldsymbol{y}] \approx \frac{1}{L}\sum_{k=1}^{L} (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\theta}^{(k)}.$$

Also, using (3.2) the covariance matrix is approximated by

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{\beta}|\boldsymbol{y}) &\approx \frac{1}{L}\sum_{k=1}^{L}\left( (r_u^{(k)})^{-1}(\boldsymbol{X}'\boldsymbol{X})^{-1} \right) \\ &\quad + \frac{1}{L}\sum_{k=1}^{L}\left( \{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\theta}^{(k)}\} \cdot \{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\theta}^{(k)}\}' \right) \\ &\quad - \left( \frac{1}{L}\sum_{k=1}^{L}\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\theta}^{(k)}\} \right) \cdot \left( \frac{1}{L}\sum_{k=1}^{L}\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\theta}^{(k)}\}' \right). \end{aligned}$$

where $\{r_u^{(k)}, \boldsymbol{\theta}^{(k)}\}_{k=1}^{L}$ is Gibbs output from the full conditionals.

Similarly, the posterior pdf of $r_u$ given $\boldsymbol{y}$, $[r_u|\boldsymbol{y}]$ is given by

$$[r_u|\boldsymbol{y}] = \int\int [\boldsymbol{\theta}, \boldsymbol{\beta}, r_u|\boldsymbol{y}]\, d\boldsymbol{\theta}\, d\boldsymbol{\beta} = \int\int [r_u|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{y}] \cdot [\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{y}]\, d\boldsymbol{\theta}\, d\boldsymbol{\beta}.$$

The distribution of $r_u$ given $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\boldsymbol{y}$, $[r_u|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{y}]$ is also given in (2.3) and integrating $r_u$ out of $[\boldsymbol{\theta}, \boldsymbol{\beta}, r_u|\boldsymbol{y}]$ gives

$$[\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{y}] = \int [\boldsymbol{\theta}, \boldsymbol{\beta}, r_u|\boldsymbol{y}]\, dr_u$$

$$\propto \prod_{i=1}^{N} \left\{ \binom{n_i}{y_i} [F(\theta_i)]^{y_i} [1-F(\theta_i)]^{n_i - y_i} \right\} \cdot \left\{ a + \sum_{i=1}^{N} (\theta_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 \right\}^{-\frac{1}{2}(b+N)}.$$

Hence, the posterior mean for the $r_u$ satisfies the equation

$$E[r_u|\mathbf{y}] = E[E(r_u|\boldsymbol{\theta},\boldsymbol{\beta},\mathbf{y})|\mathbf{y}] = E\left[\frac{b+N}{a+\sum_{i=1}^{N}(\theta_i - \mathbf{x}_i'\boldsymbol{\beta})^2}\middle|\mathbf{y}\right],$$

and it is approximated by

$$E[r_u|\mathbf{y}] \approx \frac{1}{L}\sum_{k=1}^{L}\frac{b+N}{a+\sum_{i=1}^{N}(\theta_i^{(k)} - \mathbf{x}_i'\boldsymbol{\beta}^{(k)})^2}. \tag{3.3}$$

Also, we obtain the posterior variance for the $r_u$

$$\begin{aligned}
V(r_u|\mathbf{y}) &= E[V(r_u|\boldsymbol{\theta},\boldsymbol{\beta},\mathbf{y})|\mathbf{y}] + V[E(r_u|\boldsymbol{\theta},\boldsymbol{\beta},\mathbf{y})|\mathbf{y}]\\
&= E\left[\frac{2(b+N)}{\left\{a+\sum_{i=1}^{N}(\theta_i - \mathbf{x}_i'\boldsymbol{\beta})^2\right\}^2}\middle|\mathbf{y}\right] + E\left[\left\{\frac{b+N}{a+\sum_{i=1}^{N}(\theta_i - \mathbf{x}_i'\boldsymbol{\beta})^2}\right\}^2\middle|\mathbf{y}\right]\\
&\quad - \left(E\left[\frac{b+N}{a+\sum_{i=1}^{N}(\theta_i - \mathbf{x}_i'\boldsymbol{\beta})^2}\middle|\mathbf{y}\right]\right)^2\\
&\approx \frac{1}{L}\sum_{k=1}^{L}\frac{2(b+N)}{\left\{a+\sum_{i=1}^{N}(\theta_i^{(k)} - \mathbf{x}_i'\boldsymbol{\beta}^{(k)})^2\right\}^2} + \frac{1}{L}\sum_{k=1}^{L}\left(\frac{b+N}{a+\sum_{i=1}^{N}(\theta_i^{(k)} - \mathbf{x}_i'\boldsymbol{\beta}^{(k)})^2}\right)^2\\
&\quad - \left[\frac{1}{L}\sum_{k=1}^{L}\frac{b+N}{a+\sum_{i=1}^{N}(\theta_i^{(k)} - \mathbf{x}_i'\boldsymbol{\beta}^{(k)})^2}\right]^2.
\end{aligned} \tag{3.4}$$

As the case for $\boldsymbol{\beta}$, the approximations of (3.3) and (3.4) are by the Rao-Blackwellized estimates.

## 4.2 Prediction

We have a new case, not used to estimate parameters. We would like to predict the number of 'successes', $y_{N+1}$ in a sequence of $J$ i.i.d. Bernoulli trials, with covariate $\mathbf{x}_{N+1}$. The predictive probability of success in the $j$th new single trial, $y_{N+1,j}$ is

$$\begin{aligned}
p(y_{N+1,j} = 1|\mathbf{y}) &= \int p(y_{N+1,j} = 1|\theta_{N+1})\,p(\theta_{N+1}|\mathbf{y})\,d\theta_{N+1}\\
&= E[F(\theta_{N+1})|\mathbf{y}] \approx \frac{1}{L}\sum_{k=1}^{L}F(\theta_{N+1}^{(k)}) \qquad (j = 1,\dots,J),
\end{aligned}$$

where $\theta_{N+1}^{(k)} \sim N(\mathbf{x}'_{N+1}\boldsymbol{\beta}^{(k)}, (r_u^{(k)})^{-1})$, $k = 1,2,\dots,L$, and $\boldsymbol{\beta}^{(k)}$ and $r_u^{(k)}$ are outputs from the full conditionals using the Gibbs algorithm. Then the predicted number of successes, $y_{N+1}$ in $J$ trials is

$$y_{N+1} = J\cdot\left(\frac{1}{L}\sum_{k=1}^{L}F(\theta_{N+1}^{(k)})\right).$$

We can compare the predicted number of successes to the actual observed responses and get a measure of the predictive accuracy of our model.

## 4. Model Assessment

There are several Bayesian method available to studying model assessment. In this section, we briefly review some proposals for model adequacy and model comparison.

We first discuss one general method of checking model adequacy, which is known as posterior predictive assessment approach as advocated in Gelman, Meng and Stern(1996) and Gelman and Meng(1996). This approach is based on checking the fit of a model to data by simulating values of a discrepancy measure from posterior predictive distribution and comparing these sample with the corresponding measure for the observed data.

Let $\boldsymbol{y}_{\text{obs}}$ and $\boldsymbol{y}_{\text{new}}$ denote the vector of observed and generated data respectively and let $\boldsymbol{\theta}$ represent the vector of unknown parameters. The algorithm of this approach is summarized as follows.

( i ) Generate $\boldsymbol{\theta}^{(k)}(k=1,\ldots,L)$ from the posterior density of $\boldsymbol{\theta}$ given $\boldsymbol{y}_{\text{obs}}$ using the Gibbs output (for $k=1,\ldots,L$), (where $L$ is the total number of Gibbs iterations).

( ii ) Simulate $L$ hypothetical replicates of the data, say $\boldsymbol{y}^{(1)},\ldots,\boldsymbol{y}^{(L)}$, where $\boldsymbol{y}^{(k)}$ $(k=1,\ldots,L)$ is drawn from the sampling distribution of $\boldsymbol{y}$ given the simulated $\boldsymbol{\theta}^{(k)}$.

( iii ) Calculate $d(\boldsymbol{y}_{\text{obs}},\boldsymbol{\theta}^{(k)})$ and $d(\boldsymbol{y}^{(k)},\boldsymbol{\theta}^{(k)})$ for $k=1,\ldots,L$, where $d(\boldsymbol{y},\boldsymbol{\theta})$ is an appropriate discrepancy measure.

( iv ) Approximate $P\{d(\boldsymbol{y}_{\text{new}},\boldsymbol{\theta}^{(k)})\geq d(\boldsymbol{y}_{\text{obs}},\boldsymbol{\theta}^{(k)})\mid \boldsymbol{y}_{\text{obs}}\}$ by

$$L^{-1}\sum_{k=1}^{L} I\{d(\boldsymbol{y}^{(k)},\boldsymbol{\theta}^{(k)})\geq d(\boldsymbol{y}_{\text{obs}},\boldsymbol{\theta}^{(k)})\},$$

where $I(\cdot)$ is the usual indicator function.

If the model is adequate, then the probability should be near .5 ; conversely, for an inadequate model, this probability is close to 1 or 0. The discrepancy measure that we use for our proposed model is given by

$$d(\boldsymbol{y},\boldsymbol{\theta}) = \sum_{i=16}^{N}\frac{(y_i-n_i p_i^{(k)})^2}{n_i p_i^{(k)}(1-p_i^{(k)})},$$

where, $p_i^{(k)}=F(\theta_i^{(k)})$ is the generated value of $p_i$ from the $k$th iteration. Other choice of discrepancy measure have been provided by Gelfand and Ghosh(1998).

The second approach to model comparison is the posterior predictive divergence approach. Let $\boldsymbol{y}^{(k)}$ denote the simulates data vector from the posterior predictive distribution $f(\boldsymbol{y}_{\text{new}}\mid \boldsymbol{y}_{\text{obs}})$. We naturally choose a predictive loss to be the expected value of the average discordance between the simulated data and the observed data, which is

$E[\| y_{new} - y_{obs} \|^2 | y_{obs}]$, the divergence measure of Laud and Ibrahim (1995). This quantity is approximated by

$$L^{-1} \sum_{k=1}^{L} \| y_{new} - y_{obs} \|^2.$$

Clearly, there are other possible choice of the loss function beside the squared error loss. Gelfand and Ghosh(1998) proposed a number of loss function. In the next section, we apply these approaches for model checking.

## 5. Data Analysis

We illustrate in this section the Bayesian analysis for random effect binomial regression model with real dataset given in Crowder(1978). This example concerns the proportion of seeds that germinated on each of 21 plates arranged according to a $2 \times 2$ factorial layout by seed, *O. aegyptiaca* 75 and *O. aegyptiaca* 73 and type of root extract, bean and cucumber. The data are shown below, where $y_i$ and $n_i$ are the number of germinated and the total number of seeds on the $i$th plate, $i = 1, \ldots, N$. These data are also analyzed by, for example, Breslow and Clayton(1993).

**Table 1.** Crowder's Seeds Data

| seed *O. aegyptiaca* 75 | | | | | | seed *O. aegyptiaca* 73 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Bean | | | Cucumber | | | Bean | | | Cucumber | | |
| $y$ | $n$ | $y/n$ | $y$ | $n$ | $y/n$ | $y$ | $n$ | $y/n$ | $y$ | $n$ | $y/n$ |
| 10 | 39 | .26 | 5 | 6 | .83 | 8 | 16 | .50 | 3 | 12 | .25 |
| 23 | 62 | .37 | 53 | 74 | .72 | 10 | 30 | .33 | 22 | 41 | .54 |
| 23 | 81 | .28 | 55 | 72 | .76 | 8 | 28 | .29 | 15 | 30 | .50 |
| 26 | 51 | .51 | 32 | 51 | .63 | 23 | 45 | .51 | 32 | 51 | .63 |
| 17 | 39 | .44 | 46 | 79 | .58 | 0 | 4 | .00 | 3 | 7 | .43 |
| | | | | 13 | .77 | | | | | | |

Following to the posterior predictive assessment approach, we simulated values of the discrepancy measure and then approximated $E[\| y_{new} - y_{obs} \|^2 | y_{obs}]$. The estimated probability of 0.416 under the logit model strongly suggests the adequacy of the model whereas a value of 0.137 and 0.001 for the probit model and the complementary log-log model, respectively, indicates a lack of fit of the model.

In the posterior predictive divergence approach, the estimated predictive loss for the logit model, the probit model and the complementary log-log model are 384.62, 535.14 and 941.68, respectively. Clearly, it indicates the superiority of the logit model.

Thus, our appropriate choice is a random effects logistic model, allowing for over-dispersion.

If $p_i$ is the probability of germination on the $i$th plate, we assume

$$y_i|p_i \sim \text{Binomial}(n_i, p_i)$$
$$\theta_i = \text{logit}(p_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + u_i$$
$$u_i \sim \text{Normal}(0, \sigma_u^2),$$

where $x_{1i} = 0$ or 1 corresponding to *O. aegyptiaca* 75 or *O. aegyptiaca* 73 is cultivated in $i$-th plate ; $x_{2i} = 0$ or 1 corresponding to bean and cucumber. The seed type-root extract interaction was found significant in the frequentist logistic regression analysis. So, an interaction term $\beta_3 x_1 x_2$ is included in our analysis. Let $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$ represent the fixed effects associated with seed and root extract and $u_i$ represent random effects associated with each plate. The $u_i$'s are i.i.d. $N(0, \sigma_u^2)$. We assign a uniform $(R^4)$ prior for $\boldsymbol{\beta}$, and Gamma$(0.001, 0.001)$ prior for $r_u = (\sigma_u^2)^{-1}$.

Under the assumption of improper prior for regression parameters, we performed our production run of five parallel Gibbs chains for 2000 iterations each. We started the five chains at disparate points in the sample space and discarded the first 1000 iterations of each chain. In the Gibbs sampling algorithm, we employ the Metropolis-Hastings algorithm for (2.3)(iii).

After considering plots of the realized chains, sample autocorrelations, cross-correlations and the monitoring statistic of Gelman and Rubin(1992), we were satisfied with the convergence of our algorithm.

**Table 2.** Model Fits to the Crowder's Seeds Data

| | Method of Analysis | | | |
|---|---|---|---|---|
| | logistic regression | PQL | BUGS | improper prior |
| variable | $\hat{\beta} \pm SE$ | $\hat{\beta} \pm SE$ | $\hat{\beta} \pm SE$ | $\hat{\beta} \pm SE$ |
| constant( $\beta_0$ ) | -.558 ± .126 | -.542 ± .190 | -.542 ± .178 | -.552 ± .188 |
| seed( $\beta_1$ ) | .146 ± .223 | .077 ± .308 | .028 ± .340 | .065 ± .306 |
| extract( $\beta_2$ ) | 1.318 ± .117 | 1.339 ± .270 | 1.368 ± .253 | 1.345 ± .262 |
| interaction( $\beta_3$ ) | -.778 ± .306 | -.825 ± .430 | -.792 ± .426 | -.812 ± .422 |
| scale( $\sigma_u$ ) | --- | .313 ± .121 | .292 ± .152 | .275 ± .148 |

Table 2 shows the estimates of regression coefficients as well as scale in linear logistic regression models fitted to the 21 binomial proportion of seed germination. We may compare simple logistic regression analysis, penalized quasi-likelihood (PQL) analysis (Breslow and

Clayton, 1993), Bayesian analysis using BUGS output and our result. The classical logistic regression analysis has not considered at all random effect in modelling. PQL analysis is non-Bayesian and approximate inference. BUGS has carried out using diffused normal priors. From Table 2, we can see that all estimates of $\beta_i$'s and $\sigma_u$ are quite comparable to each other.

The covariance matrix for $\boldsymbol{\beta}$ is given by

$$\mathrm{Cov}(\boldsymbol{\beta}|\boldsymbol{y}) = \begin{pmatrix} 0.035 & -0.034 & -0.035 & 0.033 \\ -0.034 & 0.094 & 0.034 & -0.090 \\ -0.035 & 0.034 & 0.039 & -0.068 \\ 0.033 & -0.090 & -0.068 & 0.178 \end{pmatrix}.$$

To compare the predicted number of successes with the actual observed number of successes, we delete the 7th data point ( $n=74$, $y=53$, seed *O. aegyptiaca* 75, cucumber) that is selected at random. We compute for logit link the predictive probability of germination on the 7th plate based on of the remaining data. The predictive probability is 0.683. So, the predicted number of germination in 74 seeds is 51, which is quite close to actual value 53.

## References

[1] Bedrick, E. J., Christensen, R. and Johnson, W. (1997). Bayesian Binomial Regression: Predicting Survival at a Trauma Center, *The American Statistician*, **51**, 211-218.

[2] Breslow, N. E. and Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models, *Journal of the American Statistical Association*, **88**, 9-25.

[3] Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm, *The American Statistician*, **49**, 327-335.

[4] Crowder, M. J. (1978). Beta-binomial Anova for Proportions, *Applied Statistics*, **27**, 34- 37.

[5] Gelfand, A. E. and Ghosh, S. K. (1998). Model Choice: a posterior predictive loss approach, *Biometrics*, **85**, 1-11.

[6] Gelfand, A. E. and Smith, A. F. M. (1991). Gibbs Sampling for Marginal Posterior Expectations, *Communications in Statistics, Theory and Methods*, **20**, 1747-1766.

[7] Gelman, A., Meng, X.-L. and Stern, H. (1996). Posterior Predictive assessment of model fitness via realized discrepancies (with discussions), *Statistica Sinica*, **6**, 733-807[or 457-511].

[8] Gelman, A. and Meng, X.-L. (1996). Model Checking and Model Improvement. In *Marcov Chain Monte Carlo in Practice*, W. R. Gilks *at al.* (eds), 189-201. Chapman & Hall, London

[9] Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation (with

discussion), *Statistical Science*, **7**, 457-511.

[10] Ghosh, M., Natarajan, K., Stroud, T. W. F. and Carlin, B. P. (1998). Generalized Linear Models for Small-Area Estimation, *Journal of the American Statistical Association*, **93**, 273-282.

[11] Gilks, W. R. and Wild, P. (1992). Adaptive Rejection Sampling for Gibbs Sampling, *Applied Statistics*, **41**, 337-348.

[12] Ibrahim, J. G. and Laud, P. W. (1991). On Bayesian Analysis of Generalized Linear Models Using Jeffreys' Prior, *Journal of the American Statistical Association*, **86**, 981-986.

[13] Laud, P. and Ibrahim, J. (1995). Predictive model selection, *Journal of the Royal Statistical Society*, Ser. B, **57**, 247-262.

[14] Leonard, T. (1972). Bayesian Methods for Binomial Data, *Biometrika*, **59**, 581-589.

[15] Lindley, D. V. and Smith, A. F. M. (1972). Bayes Estimates for the Linear Model (with discussion), *Journal of the Royal Statistical Society*, Ser. B, **34**, 1-41.

[16] Natarajan, R. and McCulloch, C. E. (1995). A Note on the Existence of the Posterior Distribution for a Class of Mixed Models for Binomial Responses, *Biometrika*, **82**, 639-643.

[17] Nelder, J. A. (1972). Discussion on the Paper by Professor Lindley and Dr Smith, *Journal of the Royal Statistical Society*, Ser. B, **24**, 1-41.

[18] Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized Linear Models, *Journal of the Royal Statistical Society*, Ser. A, **135**, 370-384.

[19] Zeger, S. L. and Karim, M. R. (1991). Generalized Linear Models with Random Effects; A Gibbs Sampling Approach, *Journal of the American Statistical Association*, **86**, 79-86.

[20] Zellner, A. and Rossi, P. E. (1984). Bayesian Analysis of Dichotomous Quantal Response Models, *Journal of Econometrics*, **25**, 365-393.