

Fitting Bayesian Two-Stage Generalized Linear Models Using Random Samples via the SIR Algorithm

Bal gobin Nandram and Erik Barry Erhardt

Worcester Polytechnic Institute, Worcester, Massachusetts, USA

Abstract

Although it is common practice to fit a complex Bayesian model using Markov chain Monte Carlo (MCMC) methods, we provide an alternative sampling-based method to fit a two-stage hierarchical model in which there is conjugacy conditional on the parameters in the second stage. Using the sampling importance resampling (SIR) algorithm, our method subsamples independent samples from an approximate joint posterior density. This is an alternative to a Metropolis-Hastings (MH) algorithm normally used to draw samples from the joint posterior density. We also provide comparison with a Metropolis (MET) algorithm. We illustrate our method using a Poisson regression model which has much interest for the analysis of rare events from small areas. We also illustrate our method using a relatively new logistic regression model. We use four examples, three on Poisson regression and one on logistic regression, and a simulation study on the Poisson regression model to assess the performance of our method relative to the MH and the MET algorithms.

AMS (2000) subject classification. 62F15, 62J12.

Keywords and phrases. Collapsing, conditional conjugate, hierarchical Bayesian model, MCMC monitoring, Metropolis-Hastings algorithm, Rao-Blackwellized estimator, small areas.

1 Introduction

Bayesian generalized linear models are used routinely to incorporate covariates for discrete data analysis over small areas in many applications. These models are usually fit using Markov chain Monte Carlo (MCMC) methods. Generally Rao-Blackwellized estimators of the small area effects cannot be obtained because the conditional posterior densities of the area effects do not exist in simple forms. However, in a class of two-stage generalized linear models the conditional posterior densities of the area effects

are in simple forms, and Rao-Blackwellized estimators are possible. It is interesting that this class of models can be fit without using MCMC methods, and random samples via the sampling importance resampling (SIR) can be used. In this paper we explore the feasibility of this procedure.

The approach of using a conditional conjugate family of distributions is attractive. Our application is on the estimation of rates and proportions from small “areas”, a research activity that is of increasing interest (see Nandram et al., 2000 and Waller et al., 1997). In typical generalized linear models, the first two stages are not conjugate, and simple Rao-Blackwellized estimators of the first-stage parameters, usually the parameters of much interest, are not available (e.g., see Nandram et al. 2000 and Waller et al. 1997). A Rao-blackwellized estimator has smaller mean squared error than its empirical counterpart (Gelfand and Smith 1990). Also under conditional conjugacy one can collapse (Liu 1994) over the area effects to obtained a much reduced number of parameters for computation. In addition, most two-stage generalized linear models can be constructed with this structure. Also a method that does not use Markov chain Monte Carlo methods is desirable because it makes the methodology available to a wider range of practitioners. We propose a method for fitting such a model using a sampling-based method but without Markov chains, a method that subsamples independent samples using the SIR algorithm.

A standard sampling-based method for fitting complex Bayesian models is the Metropolis-Hastings sampler (see Metropolis et al., 1953, Hastings 1970 and Tierney 1994). Also see Chib and Greenberg (1995) for a pedagogical review of Metropolis-Hastings algorithm. Metropolis-type local move allows one to break a complex task into a series of manageable pieces. However, these local moves often lead to slow-converging algorithms that are easily trapped in a local mode, thereby providing an ineffective algorithm; see Liu et al. (2000) for appropriate comments about the Metropolis-Hastings sampler. With a pure Gibbs sampler, some of the conditional posterior densities might be too complex to work with. The adaptive rejection sampler (ARS) is powerful for log-concave densities, but does not apply to others (see Gilks and Wild 1992). Also all ARS algorithms use single-parameter updates, but multi-parameter updates (blocking) is usually more efficient. See also Gilks, Best and Tan (1995) who generalize the ARS to deal with non-log-concave full conditional densities by including a Metropolis-Hastings sampler; the best performance occurs when the conditional densities are near log-concavity. But the rejection rates can be very high, again providing an ineffective algorithm. It is usually very difficult to find efficient candidate generating densities, and monitoring and tuning of a Metropolis step can be

a daunting task. These remarks point to the simple idea that one should be careful when using the Metropolis-Hastings sampler.

Smith and Gelfand (1992) provide a sampling-resampling perspective of Bayes' theorem. They examine two resampling techniques whereby samples from one distribution may be modified to form samples from another distribution. Specifically both techniques can be used to modify posterior samples arising from one model specification to posterior samples arising from another. We are particularly interested in the SIR algorithm which is potentially useful to obtain a sample from a posterior distribution using a sample from another posterior distribution which approximates the original. We use the SIR algorithm to fit two-stage hierarchical models where, conditional on the parameters of distributions in the second stage, there is conjugacy.

We briefly describe the SIR algorithm. Let θ_i , $i = 1, \dots, M$, be a sample from $g(\theta)$, $q_i = f(\theta_i)/g(\theta_i)$ and $w_i = q_i / \sum_{i'=1}^M q_{i'}$. The SIR algorithm draws θ^* from the discrete distribution $\{(\theta_i, w_i), i = 1, \dots, M\}$ (i.e., a weighted bootstrap). Then θ^* is approximately distributed according to $f(\theta)$, with increasing accuracy as M increases (see Smith and Gelfand 1992 for a justification of this claim). It is worth noting that one does not need the normalization constant in $f(\theta)$. Smith and Gelfand (1992, p. 85) note that "The less f resembles g , the larger the sample size M will need to be in order that the distribution of θ^* well approximates f ." They use sampling without replacement, and they recommend that the resampling probabilities need to be at least 1 in 10. We use sampling with replacement, but note also that the larger the resampling probabilities, for sampling with replacement, the greater the dependence in the subsamples because there will be more repeats for many sampled values. Thus, we seek an approach where we first try to make $g(\theta)$ as close as possible to $f(\theta)$.

A hierarchical Bayesian Poisson regression model, called Poisson regression interactive multilevel modeling (PRIMM) introduced by Christiansen and Morris (1997), has been used to study mortality data and other rare events when there are occurrences from many areas. See Nandram (2000) for a review of Bayesian generalized linear models for small areas, which includes PRIMM. Christiansen and Morris (1997) have made some accurate analytical approximations, and it is important to note that these approximations avoid the use of Markov chain Monte Carlo methods. A sampling-based approach is desirable because it removes the analytical approximations associated with the method of Christiansen and Morris (1997). PRIMM is a special case of a two-stage hierarchical model in which the first two stages are conjugate conditional on the parameters of the distribution in the second stage.

Another example is a logistic regression model (i.e., an extension of a standard beta-binomial model). Here, like PRIMM, the parameters of the beta distribution hold the covariates, permitting conjugacy. Both the Poisson regression model and the logistic regression model are two useful examples of generalized linear models with conditional conjugacy on the area effects (i.e., the rates or proportions), and they fall within the same framework. Of course, other models can be constructed within the same spirit. We propose a method for fitting such models using a sampling-based method but without Markov chains, a method that subsamples independent samples using the SIR algorithm.

In a summary, we list our contributions in this paper.

- (a) We study a general class of conditional conjugate two-stage generalized linear models.
- (b) We use the SIR algorithm to fit these generalized linear models.
- (c) We use random samples to fit these models instead of Markov chain Monte Carlo methods.
- (d) We describe the Poisson and logistic regression models within this class, and show how to fit them using the SIR algorithm.
- (e) We compare our method with two Metropolis samplers, a Metropolis-Hastings sampler and a simple Metropolis sampler.
- (f) We use several real examples and a simulation study to compare our method with others.

We describe the general methodology in Section 2, and the methodology for Poisson and logistic regression in Section 3. In Section 4 we describe three examples on Poisson regression and one example on logistic regression to compare the Metropolis algorithms with our sampling-based method. Also in Section 4 a simulation study is used to compare our method and its competitors through the Poisson regression model. Section 5 has a discussion.

2 Methodology for Two-stage Hierarchical Models

In this section, using the SIR algorithm, we describe a general methodology to fit a two-stage hierarchical model with conjugacy conditional on the parameters of the second stage. If we can perform a sampling-based method using only independent samples, it will be much better than using

MCMC methods because monitoring convergence can be a daunting task. In the Metropolis-Hastings algorithm (Chib and Greenberg 1995) one needs to worry about how many iterates to “burn in”, autocorrelation and jumping probabilities (Gelman et al., 1996). Our strategy is to obtain a sample from an approximation to the posterior density and “tune” the algorithm that gives the approximate sample until it “conforms” to the true posterior density. Tuning of the SIR algorithm is relatively easy.

Let \underline{d} denote the data, $\underline{\theta}$ the first-stage parameters and $\underline{\beta}$ the parameters of the distribution of $\underline{\theta}$, and let $g(\underline{\beta} | \underline{d})$ denote the posterior distribution of $\underline{\beta}$, assuming propriety. Then, the posterior for $\underline{\theta}, \underline{\beta}$ is given by

$$\pi(\underline{\theta}, \underline{\beta} | \underline{d}) = f(\underline{\theta} | \underline{\beta}, \underline{d})g(\underline{\beta} | \underline{d}). \tag{1}$$

Then, if we can obtain samples from $g(\underline{\beta} | \underline{d})$, the composition method (Tanner 1993, p. 30) provides a natural way to get samples from (1). One can simply run a Metropolis algorithm to fit the posterior distribution $g(\underline{\beta} | \underline{d})$, and then in an output analysis for each sampled value $\underline{\beta}$, a sample can be drawn from $f(\underline{\theta} | \underline{\beta}, \underline{d})$; this is the method of collapsing in Markov chain Monte Carlo methods (see Liu 1994). With conjugacy it is easy to draw a sample from $f(\underline{\theta} | \underline{\beta}, \underline{d})$. In any problem of this kind, linear or generalized linear models, this approach can always be executed. Also, if one can draw samples from $g(\underline{\beta} | \underline{d})$, there will be considerable savings in computational time.

The question that arises is “How can a sample be drawn from $g(\underline{\beta} | \underline{d})$?” One can use a Metropolis algorithm although it is generally difficult to obtain good jumping probabilities. It is much easier to draw $f(\underline{\theta} | \underline{\beta}, \underline{d})$ and $h(\underline{\beta} | \underline{\theta}, \underline{d})$, say, using the Metropolis-Hastings algorithm (Chib and Greenberg 1995) with a Metropolis step for $h(\underline{\beta} | \underline{\theta}, \underline{d})$. However, one still needs to “tune” the Metropolis step to obtain good jumping probabilities. It is our experience that this latter procedure generally performs well using Markov chains, but the tuning can be difficult. By collapsing over $\underline{\theta}$, we can draw a sample from $g(\underline{\beta} | \underline{d})$ using a resampling method. Then samples for $f(\underline{\theta} | \underline{\beta}, \underline{d})$ can be obtained in an output analysis. Hence, we want to show how to obtain a sample from $g(\underline{\beta} | \underline{d})$ using the SIR algorithm.

Let $g_a(\underline{\beta} | \underline{d})$ be an approximation to $g(\underline{\beta} | \underline{d})$ such that one can draw independent samples from $g_a(\underline{\beta} | \underline{d})$ easily, and let $\underline{\beta}^{(1)}, \dots, \underline{\beta}^{(M)}$ denote a sample from $g_a(\underline{\beta} | \underline{d})$. Then with $q_i = g(\underline{\beta}^{(i)} | \underline{d})/g_a(\underline{\beta}^{(i)} | \underline{d})$ and $w_i = q_i / \sum_{i'=1}^M q_{i'}$, our version of the SIR algorithm draws a sample with replacement from the discrete distribution $\{(\underline{\beta}^{(i)}, w_i), i = 1, \dots, M\}$.

The closer w_i is to $1/M$, the closer the approximation $g_a(\underline{\beta} | \underline{d})$ will be to $g(\underline{\beta} | \underline{d})$. Thus, our key contribution is “How can $g_a(\underline{\beta} | \underline{d})$ be made as

close as possible to $g(\underline{\beta} | \underline{d})$?" We construct a multivariate normal density in which we incorporate symmetry and orthogonality (e.g., see Albert 1989). Typically, this can be easily done. For example, most of the components of $\underline{\beta}$ are regression coefficients with, say one nuisance parameter $\alpha > 0$, and α can be made approximately symmetric by a transformation (e.g., logarithm). Then, as a first approximation one can take $\underline{\beta} \sim N_p(\hat{\underline{\beta}}, \Delta)$ where $\hat{\underline{\beta}}$ is the posterior mode and $\Delta = (-H)^{-1}$ is the negative inverse Hessian matrix, evaluated at the posterior mode of $g(\underline{\beta} | \underline{d})$, assuming existence.

To make this modal-Hessian normal approximation close to $g(\underline{\beta} | \underline{d})$, we introduce tuning constants κ , with each component corresponding to a diagonal element of the covariance matrix Δ . That is, $\underline{\beta} | \underline{d} \sim N_p\{\hat{\underline{\beta}}, \text{diag}(\sqrt{\kappa_1}, \dots, \sqrt{\kappa_p}) \Delta \text{diag}(\sqrt{\kappa_1}, \dots, \sqrt{\kappa_p})\}$. This operation replaces each element, $\delta_{jj'}$, of Δ by $\sqrt{\kappa_j} \sqrt{\kappa_{j'}} \delta_{jj'}$. Then, $q_i = g(\underline{\beta}^{(i)} | \underline{d}) / g_{\kappa}(\underline{\beta}^{(i)} | \underline{d})$, $i = 1, \dots, M$ and $w_i = q_i / \sum_{i'=1}^M q_{i'}$. Thus, we draw samples from $g_{\kappa}(\underline{\beta} | \underline{d})$ and tune κ iteratively until $w_i \approx M^{-1}$, say, for all $i = 1, \dots, M$, where

$$g_{\kappa}(\underline{\beta} | \underline{d}) \propto \exp[-(\underline{\beta} - \hat{\underline{\beta}})' \times \{\text{diag}(\sqrt{\kappa_1}, \dots, \sqrt{\kappa_p}) \Delta \text{diag}(\sqrt{\kappa_1}, \dots, \sqrt{\kappa_p})\}^{-1} (\underline{\beta} - \hat{\underline{\beta}})]. \quad (2)$$

One advantage of (2) is that all directions in the multivariate density are considered asymmetrically. Then, (2) is a general proposal density and this tuning process forms our SIR algorithm. While one would like $w_i \approx M^{-1}$, one can settle for a variation such as $w_i \leq .05$.

3 Poisson and Logistic Regression Models

In this section we describe two Bayesian models: the Poisson-gamma hierarchical regression model and the binomial-beta hierarchical logistic regression model, and how to fit them using the SIR algorithm. We also consider two other algorithms, one based on the Metropolis-Hastings sampler and the other based on the simpler Metropolis sampler.

3.1. Our proposed method. Let θ_i denote the rate or proportion associated with the i^{th} unit (e.g., geographical area), $i = 1, \dots, \ell$. The observations consist of the number of incidents d_i and the population size n_i of the i^{th} unit, $i = 1, \dots, \ell$. To link d_i and n_i to the θ_i , in the Poisson-gamma model we assume that

$$d_i | \theta_i \stackrel{\text{ind}}{\sim} \text{Poisson}(n_i \theta_i), \quad i = 1, \dots, \ell, \quad (3)$$

and in the binomial-beta model we assume that

$$d_i | \theta_i \stackrel{\text{ind}}{\sim} \text{Binomial}(n_i, \theta_i), \quad i = 1, \dots, \ell. \quad (4)$$

Under these models the maximum likelihood estimator of θ_i is $r_i = d_i/n_i$, $i = 1, \dots, \ell$, the observed rate or proportion.

It is standard practice to estimate the θ_i by “borrowing strength” across the units. Thus, letting $\underline{x}_i = (1, x_{i,1}, \dots, x_{i,p-1})'$ denote an intercept and the vector of $(p - 1)$ covariates, in the Poisson-gamma model we assume that

$$\theta_i \mid \underline{\beta}, \tau \stackrel{ind}{\sim} \text{Gamma}(e^\tau, e^{\tau - \underline{x}'_i \underline{\beta}}) \tag{5}$$

and in the binomial-beta model we assume that

$$\begin{aligned} \theta_i \mid \underline{\beta}, \tau &\stackrel{ind}{\sim} \text{Beta}\{e^\tau \phi_i(\underline{\beta}), e^\tau (1 - \phi_i(\underline{\beta}))\} \\ \text{where } \phi_i(\underline{\beta}) &= e^{\underline{x}'_i \underline{\beta}} / (1 + e^{\underline{x}'_i \underline{\beta}}), i=1, \dots, \ell. \end{aligned} \tag{6}$$

In both regression models we have a priori

$$p(\underline{\beta}, \tau) \propto a_0 e^\tau (a_0 + e^\tau)^{-2}, -\infty < \tau < \infty, -\infty < \underline{\beta} < \infty \tag{7}$$

where $\alpha = e^\tau$. The model specified by (3), (5) and (7) was described by Christiansen and Morris (1997) who selected a_0 , and the model specified by (4), (6) and (7) is our new logistic regression model. See Christiansen and Morris (1997, Section 2.4) for a simple sufficient condition for posterior propriety in the Poisson regression model. As we have found that inference is not sensitive to the choice of a_0 , we choose $a_0 = 1$ (see Albert 1988 for the choice of $a_0 = 1$). For this exposition we will consider τ to be a nuisance parameter.

We make two remarks about these models. Both regression models incorporate the covariates at the area level similar to the Fay-Herriot model (Fay and Herriot, 1977). However, unlike the Fay-Herriot model, these models are not approximate (i.e., they do not rely on the central limit theorem). In the logistic regression model, it is required that $n_i > 1$ for some i , $i = 1, \dots, \ell$; otherwise τ is not identifiable.

Letting $\underline{\theta}$ denote the vector of rates or proportions, the joint posterior density for the θ_i is

$$\pi(\underline{\theta} \mid \underline{d}) = \int \int \left\{ \prod_{i=1}^{\ell} f(\theta_i \mid \underline{\beta}, \tau, d_i) \right\} g(\underline{\beta}, \tau \mid \underline{d}) d\underline{\beta} d\tau \tag{8}$$

where, in the Poisson-gamma model,

$$f(\theta_i \mid \underline{\beta}, \tau, d_i) = \frac{(n_i + e^{\tau - \underline{x}'_i \underline{\beta}})^{d_i + e^\tau} \theta_i^{d_i + e^\tau - 1} \exp\{-\theta_i(n_i + e^{\tau - \underline{x}'_i \underline{\beta}})\}}{\Gamma(d_i + e^\tau)} \tag{9}$$

and

$$g(\underline{\beta}, \tau \mid \underline{d}) \propto \frac{a_0 e^\tau}{(a_0 + e^\tau)^2} \times \prod_{i=1}^{\ell} \left[\frac{\Gamma(d_i + e^\tau)}{\Gamma(e^\tau)} \times \left\{ \frac{e^{\tau - x'_i \underline{\beta}}}{n_i + e^{\tau - x'_i \underline{\beta}}} \right\}^{e^\tau} \times \frac{1}{(n_i + e^{\tau - x'_i \underline{\beta}})^{d_i}} \right], \tag{10}$$

and in the binomial-beta model,

$$f(\theta_i \mid \underline{\beta}, \tau, d_i) = \frac{\theta_i^{d_i + e^\tau \phi_i(\underline{\beta}) - 1} (1 - \theta_i)^{n_i - d_i + e^\tau (1 - \phi_i(\underline{\beta})) - 1}}{B\{d_i + e^\tau \phi_i(\underline{\beta}), n_i - d_i + e^\tau (1 - \phi_i(\underline{\beta}))\}} \tag{11}$$

and

$$g(\underline{\beta}, \tau \mid \underline{d}) \propto \frac{a_0 e^\tau}{(a_0 + e^\tau)^2} \times \prod_{i=1}^{\ell} \left[\frac{B\{d_i + e^\tau \phi_i(\underline{\beta}), n_i - d_i + e^\tau (1 - \phi_i(\underline{\beta}))\}}{B\{e^\tau \phi_i(\underline{\beta}), e^\tau (1 - \phi_i(\underline{\beta}))\}} \right]. \tag{12}$$

Both models are attractive because of the conjugacy in which the conditional posterior density of θ_i is the simple gamma distribution in (9) or beta distribution in (11), permitting easy construction of Rao-Blackwellized estimators of the θ_i . In the standard generalized linear model in which the $\log(\theta_i)$ or $\log\{\theta_i/(1 - \theta_i)\}$ follow a normal linear model, it is not possible to obtain simple Rao-Blackwellized estimators of the θ_i ; only empirical estimators can be easily obtained. Also note that $g(\underline{\beta}, \tau \mid \underline{d})$ in (10) or (12) is not log-concave, prohibiting the use of the ARS.

Finally, we only need to show how to apply our algorithm to sample from $g(\underline{\beta}, \tau \mid \underline{d})$ in (10) or (12). In the Appendix A we describe the Hessian matrix, H , of $\log\{g(\underline{\beta}, \tau \mid \underline{d})\}$ evaluated at the posterior mode $(\hat{\underline{\beta}}, \hat{\tau})$ of $g(\underline{\beta}, \tau \mid \underline{d})$ which is obtained using the Nelder-Mead algorithm. Letting $\Sigma = (-H)^{-1}$, our proposal density is

$$\begin{pmatrix} \underline{\beta} \\ \tau \end{pmatrix} \mid \underline{d} \sim \text{Normal} \left\{ \begin{pmatrix} \underline{\beta} \\ \tau \end{pmatrix}, \text{diag}(\sqrt{\kappa_1}, \dots, \sqrt{\kappa_{p+1}}) \Sigma \text{diag}(\sqrt{\kappa_1}, \dots, \sqrt{\kappa_{p+1}}) \right\}, \tag{13}$$

which we denote by $g_{\kappa}(\underline{\beta}, \tau \mid \underline{d})$, a special case of the one in (2). Note that $\kappa_1, \dots, \kappa_p$ correspond to $\underline{\beta}$ and κ_{p+1} corresponds to τ . Then, the resampling probabilities for a random sample $(\underline{\beta}^{(1)}, \tau^{(1)}), \dots, (\underline{\beta}^{(M)}, \tau^{(M)})$ from (13) are

$$w_h(\underline{\kappa}) = \left\{ \frac{g(\underline{\beta}^{(h)}, \tau^{(h)} \mid \underline{d})}{g_{\underline{\kappa}}(\underline{\beta}^{(h)}, \tau^{(h)} \mid \underline{d})} \right\} / \sum_{h'=1}^M \left\{ \frac{g(\underline{\beta}^{(h')}, \tau^{(h')} \mid \underline{d})}{g_{\underline{\kappa}}(\underline{\beta}^{(h')}, \tau^{(h')} \mid \underline{d})} \right\}, \quad h = 1, \dots, M. \tag{14}$$

We tune the algorithm by varying κ until all the w_h are smaller than .05 and near M^{-1} . In practice, we start with all components of κ set at 2 with steps around .5 depending on how close we get. This is our SIR algorithm for regression models.

3.2 Other methods. We have compared our SIR algorithm with the Metropolis-Hastings (MH) algorithm and the simple Metropolis (MET) algorithm. We have also made comparison with the algorithm which simply uses the modal-Hessian normal approximation with no resampling (NR), instead of the correct posterior density (see Section 4). The NR method is in line with the method of Christiansen and Morris (1997) although they do not use a sampling-based method, and they select $a_0 \neq 1$.

First, we briefly describe the MH algorithm that Nandram et al. (2004) use to fit the Poisson regression model with the reparameterization (β, α) where $\alpha = e^\tau$. They draw β and α simultaneously from the joint conditional posterior density using a Metropolis step with an independence chain, and the θ_i from their conditional posterior densities. They obtain a proposal density for the Metropolis step by approximating $\pi(\beta, \alpha | \theta, d)$ by $\pi_a(\beta, \alpha | \theta, d) = g_1(\alpha | \beta, \theta, d)g_2(\beta | \theta, d)$ where $g_1(\alpha | \beta, \theta, d)$ is a gamma density and $g_2(\beta | \theta, d)$ is a multivariate normal density, both obtained from the modal-Hessian normal approximation of $\pi(\beta, \alpha | \theta, d)$. They ran the Metropolis-Hastings algorithm for 5500 iterations, with a “burn in” of 500 iterations, and picked every fifth iterate from the remaining 5000 to make the autocorrelations among the iterates negligible. The jumping probability of the Metropolis step is around .40 which they found by varying a tuning constant, a multiplier of the negative inverse Hessian matrix of $\log\{\pi(\beta, \alpha | \theta, d)\}$. This is the MH algorithm that we use for comparison with our method. For our examples we simply vary the tuning constant to maintain the jumping rate of .40.

Second, we describe the MET method based on the simple Metropolis algorithm. In terms of computational time, this is a fair comparison with our SIR algorithm to draw from the joint posterior density $g(\beta, \tau | d)$ in (12), not the conditional posterior density, using the simpler Metropolis sampler. Here we use the modal-Hessian normal approximation to construct a candidate generating density. We have

$$\begin{pmatrix} \beta \\ \tau \end{pmatrix} | d \overset{app}{\sim} \text{Normal} \left\{ \begin{pmatrix} \beta \\ \tau \end{pmatrix}, \begin{bmatrix} B_M & \nu_M \\ \nu'_M & t_M \end{bmatrix} \right\} \tag{15}$$

where $-kH^{-1} = \begin{bmatrix} B_M & \nu_M \\ \nu'_M & t_M \end{bmatrix}$, with k a tuning constant; see Appendix A

for the form of H . Thus, for the candidate generating density we take

$$\underline{\beta} \mid \underline{d} \stackrel{app}{\sim} \text{Normal}(\hat{\underline{\beta}}, B_M) \quad (16)$$

and independently

$$\alpha = e^\tau \mid \underline{d} \sim \text{Gamma}(\delta_0 \mu_0, \delta_0) \quad (17)$$

where $\mu_0 = e^{\hat{\tau} + t_M/2}$ and $\delta_0 = \{\mu_0(e^{t_M} - 1)\}^{-1}$. In this candidate generating density, setting $\underline{\nu}_M = \underline{0}$ provides a more efficient procedure. We have tried other forms (Chib and Greenberg 1995) of the candidate generating density, but this version appears to be the best for our application.

4 Examples and Simulation Study

We use four examples to compare the MH, SIR and MET algorithms under the regression models. Two examples, the pump and heart datasets, are taken from Christiansen and Morris (1997) and we reproduce the estimates from their non-sampling method (PRIMM). We also compare the computation time to run the MH, SIR and MET algorithms on our 666 MHz Alpha workstation. One example is given under the logistic regression model. For the Poisson regression model a small simulation study is used to compare these methods under identical repetitions.

4.1. Four examples with real data. The pump failure dataset contains the number of failures in a period (hours) of operation for 10 pumps with a single covariate that indicates whether the pump ran continuously or not (see Table 1 of Christiansen and Morris 1997). The heart dataset contains 30-day mortality rates at 15 U.S. heart transplant centers with a single covariate that reflects the expected number of deaths based on demographic and health risk variables for each patient (see Table 2 of Christiansen and Morris 1997). Our COPD dataset is a subset (National Center for Health Statistics 1998) that includes population census data for 1990 and mortalities due to chronic obstructive pulmonary disease (COPD) for the period 1988 through 1992. We use the subset of white males, age 65+, from the states of Washington, Oregon and California which contain 48 health service areas (HSA's). We include in our model the four covariates (Nandram et al. 2000), reflecting white male lung cancer rate per 1,000 population, square root of population density per 10,000, square root of elevation per 10,000 and annual rainfall per 100. Our osteoporosis dataset is a subset (National Center for Health Statistics 1994) of the Third National Health and Nutrition Examination Survey (NHANES III, 1988–92) that includes a small

sample providing the number of people with osteoporosis (measured by bone mineral density (BMD) for the sampled individuals, $BMD < 0.82$ indicates mild or severe). We include in our model the four binary covariates age (0=younger than 50, 1=50 years and older), race (0=white, 1=black), sex (0=male, 1=female) and income (0=less than 45000, 1=45000 and higher).

We have performed NR, MH, SIR and MET as described. For the pump dataset the SIR algorithm uses tuning constants $\kappa = [2.00, 2.00, 1.00]'$ which gives the largest resampling probability of 0.009, and the computation took 0.31 seconds, much less than 4.70 seconds for MH, but comparable to the 0.24 seconds for MET. For the heart dataset the SIR algorithm uses tuning constants $\kappa = [2.00, 2.00, 1.00]'$ which gives the largest resampling probability of 0.024, and the computation took 0.34 seconds, much less than 10.63 seconds for MH, but comparable to the 0.31 seconds for MET. For the COPD dataset the SIR algorithm used tuning constants $\kappa = [0.50, 0.50, 0.50, 1.50, 0.50, 0.50]'$ which gives the largest resampling probability of 0.054, and the computation took 2.28 seconds, much less than 39.62 seconds for MH, but comparable to the 1.64 seconds for MET. For the osteoporosis dataset the SIR algorithm uses tuning constants $\kappa = [2.00, 1.00, 2.00, 1.00, 1.00, 0.50]'$ which gives the largest resampling probability of 0.044, and the computation took 1.13 seconds, much less than 48.93 seconds for MH, but comparable to the 1.20 seconds for MET.

TABLE 1. PUMP DATA (CHRISTIANSEN AND MORRIS, 1997): COMPARISON OF THE POSTERIOR MEAN (PM), POSTERIOR STANDARD DEVIATION (PSD) AND PERCENTILES OF THE REGRESSION COEFFICIENTS FOR NR, MH AND SIR

		PM	PSD	Min	Percentiles					Max
					2.5	25	50	75	97.5	
NR	β_0	-0.545	0.374	-1.967	-1.269	-0.795	-0.534	-0.306	0.182	0.613
	β_1	-0.651	0.284	-1.860	-1.222	-0.840	-0.654	-0.454	-0.132	0.343
MH	β_0	-0.537	0.277	-2.111	-1.005	-0.696	-0.565	-0.419	0.160	0.653
	β_1	-0.658	0.211	-1.688	-1.131	-0.765	-0.654	-0.553	-0.181	0.536
SIR	β_0	-0.450	0.405	-1.627	-1.165	-0.736	-0.472	-0.242	0.490	1.094
	β_1	-0.632	0.283	-1.860	-1.222	-0.817	-0.632	-0.439	-0.104	-0.038
MET	β_0	-0.523	0.304	-1.375	-1.121	-0.722	-0.502	-0.343	0.072	0.326
	β_1	-0.655	0.283	-1.510	-1.423	-0.824	-0.648	-0.485	-0.010	0.043

In Tables 1, 3, 5 and 8 we present the posterior mean (PM), posterior standard deviation (PSD), and the five-number summaries for each of the regression coefficients. In general, there are differences for NR, MH, SIR and MET, but inferences about these parameters are similar for the pump and heart datasets. For the COPD dataset, except for β_2 , inference from MH, SIR and MET is the same. The PSD for SIR is larger making its 95%

credible interval wider (i.e., for β_2 compare SIR (-1.042, -0.010) with MH (-0.766, 0.027)). Also β_4 for NR is not the same as for MH, SIR and MET.

TABLE 2. PUMP DATA (CHRISTIANSEN AND MORRIS, 1997): COMPARISON OF THE POSTERIOR MEAN (PM), POSTERIOR STANDARD DEVIATION AND 95% CREDIBLE INTERVALS FOR THE FAILURE RATES FOR PRIMM, MH AND SIR

Pump	PRIMM		MH		
	PM	PSD	PM	PSD	CI
7	1.08	0.83	1.042	0.734	(0.113, 2.850)
8	1.08	0.83	1.017	0.739	(0.102, 3.016)
9	1.77	0.83	1.694	0.779	(0.548, 3.529)
5	0.67	0.35	0.678	0.320	(0.191, 1.430)
10	2.13	0.45	2.105	0.451	(1.340, 3.071)
2	0.12	0.09	0.133	0.091	(0.016, 0.363)
6	0.59	0.14	0.564	0.133	(0.333, 0.842)
3	0.09	0.04	0.093	0.038	(0.034, 0.184)
1	0.06	0.03	0.065	0.027	(0.024, 0.125)
4	0.12	0.03	0.119	0.030	(0.066, 0.185)

Pump	SIR			MET		
	PM	PSD	CI	PM	PSD	CI
7	1.071	0.814	(0.130, 3.278)	1.060	0.707	(0.150, 2.738)
8	1.108	0.856	(0.144, 3.238)	1.100	0.748	(0.168, 2.883)
9	1.767	0.871	(0.538, 3.904)	1.607	0.760	(0.531, 3.429)
5	0.670	0.339	(0.191, 1.503)	0.718	0.350	(0.206, 1.508)
10	2.078	0.446	(1.303, 3.024)	2.066	0.445	(1.284, 3.091)
2	0.135	0.097	(0.013, 0.379)	0.155	0.103	(0.018, 0.405)
6	0.572	0.134	(0.342, 0.859)	0.548	0.139	(0.303, 0.841)
3	0.093	0.039	(0.033, 0.179)	0.097	0.039	(0.037, 0.188)
1	0.064	0.026	(0.026, 0.118)	0.066	0.027	(0.026, 0.127)
4	0.122	0.031	(0.064, 0.190)	0.122	0.031	(0.072, 0.191)

NOTE: Christiansen and Morris (1997) did not present 95% credible intervals for the failure rates.

TABLE 3. HEART DATA (CHRISTIANSEN AND MORRIS, 1997): COMPARISON OF THE POSTERIOR MEAN (PM), POSTERIOR STANDARD DEVIATION (PSD) AND PERCENTILES OF THE REGRESSION COEFFICIENTS FOR NR, MH AND SIR

		PM	PSD	Percentiles						
				Min	2.5	25	50	75	97.5	Max
NR	β_0	-2.950	0.327	-4.280	-3.599	-3.165	-2.955	-2.743	-2.310	-1.879
	β_1	1.251	2.052	-5.062	-2.449	-0.154	1.226	2.615	5.381	8.494
MH	β_0	-2.978	0.237	-3.816	-3.447	-3.125	-2.970	-2.828	-2.508	-2.131
	β_1	1.179	1.463	-4.142	-1.708	0.295	1.108	2.068	4.077	6.518
SIR	β_0	-2.966	0.317	-4.214	-3.599	-3.185	-2.979	-2.766	-2.289	-1.810
	β_1	1.360	2.073	-4.447	-2.433	-0.030	1.322	2.758	5.422	8.494
MET	β_0	-2.949	0.259	-3.727	-3.467	-3.125	-2.945	-2.755	-2.499	-2.218
	β_1	0.953	1.568	-3.143	-1.936	-0.214	0.983	1.802	4.073	5.383

TABLE 4. HEART DATA (CHRISTIANSEN AND MORRIS, 1997): COMPARISON OF THE POSTERIOR MEAN (PM), POSTERIOR STANDARD DEVIATION AND 95% CREDIBLE INTERVALS FOR THE MORTALITY RATES FOR PRIMM, MH AND SIR

Center	PRIMM		MH		
	PM	PSD	PM	PSD	CI
1	0.117	0.057	0.110	0.053	(0.043, 0.243)
2	0.049	0.028	0.049	0.025	(0.013, 0.113)
3	0.083	0.038	0.076	0.032	(0.029, 0.150)
4	0.043	0.023	0.044	0.021	(0.012, 0.092)
5	0.047	0.027	0.045	0.022	(0.012, 0.094)
6	0.042	0.022	0.045	0.020	(0.011, 0.087)
7	0.051	0.026	0.054	0.023	(0.017, 0.107)
8	0.057	0.024	0.057	0.021	(0.022, 0.107)
9	0.079	0.028	0.077	0.025	(0.037, 0.139)
10	0.024	0.016	0.029	0.016	(0.004, 0.066)
11	0.083	0.029	0.077	0.026	(0.036, 0.138)
12	0.045	0.019	0.047	0.018	(0.017, 0.087)
13	0.042	0.018	0.045	0.017	(0.017, 0.085)
14	0.079	0.024	0.074	0.021	(0.040, 0.123)
15	0.067	0.022	0.067	0.021	(0.035, 0.114)

Center	SIR			MET		
	PM	PSD	CI	PM	PSD	CI
1	0.114	0.057	(0.038, 0.261)	0.081	0.032	(0.037, 0.165)
2	0.048	0.027	(0.011, 0.113)	0.052	0.020	(0.021, 0.094)
3	0.081	0.036	(0.030, 0.173)	0.069	0.023	(0.037, 0.128)
4	0.044	0.022	(0.010, 0.095)	0.052	0.017	(0.023, 0.088)
5	0.045	0.025	(0.010, 0.110)	0.048	0.024	(0.014, 0.102)
6	0.044	0.022	(0.010, 0.091)	0.055	0.016	(0.022, 0.085)
7	0.056	0.025	(0.015, 0.109)	0.063	0.023	(0.027, 0.116)
8	0.059	0.024	(0.021, 0.118)	0.061	0.017	(0.031, 0.100)
9	0.078	0.027	(0.037, 0.137)	0.069	0.019	(0.039, 0.114)
10	0.027	0.017	(0.002, 0.066)	0.046	0.015	(0.015, 0.072)
11	0.081	0.028	(0.037, 0.143)	0.065	0.021	(0.032, 0.118)
12	0.047	0.020	(0.016, 0.097)	0.052	0.014	(0.026, 0.081)
13	0.044	0.018	(0.015, 0.084)	0.052	0.014	(0.026, 0.081)
14	0.076	0.023	(0.042, 0.132)	0.065	0.016	(0.038, 0.107)
15	0.067	0.021	(0.036, 0.115)	0.065	0.017	(0.039, 0.104)

NOTE: Christiansen and Morris (1997) did not present 95% credible intervals for the mortality rates.

In Tables 2 and 4 we compare inference about the θ_i for PRIMM, MH, SIR and MET, and in Tables 6 and 9 we compare MH, SIR and MET. For the pump dataset for MH and MET the PSD's of the first three pumps are a bit smaller than for PRIMM and SIR. For the heart dataset the PRIMM, MH and SIR methods are very close. The MET method differs in the estimate for pumps 1, 3 and 10, and has generally smaller PSD's than the other methods. For the COPD dataset MH, SIR and MET are very similar; the

mortality rate for HSA 4 shows a little difference with a 95% credible interval (40.07, 54.06) for MH versus (40.97, 55.24) for SIR (see Table 6). For the osteoporosis dataset MH, SIR and MET are very similar. For cases 6, 8 and 14 the PSD differs slightly between methods. For each dataset we have also computed these quantities for NR; the PM's and PSD's are close to PRIMM for the pump and heart datasets and the 95% credible intervals are similar to those of MH and SIR.

TABLE 5. COPD DATA (NCHS, 1998): COMPARISON OF THE POSTERIOR MEAN (PM), POSTERIOR STANDARD DEVIATION (PSD) AND PERCENTILES OF THE REGRESSION COEFFICIENTS FOR NR, MH AND SIR

		PM	PSD	Percentiles						
				Min	2.5	25	50	75	97.5	Max
NR	β_0	-5.994	0.239	-6.727	-6.463	-6.162	-5.983	-5.832	-5.519	-5.287
	β_1	0.012	0.004	-0.001	0.003	0.009	0.012	0.015	0.021	0.026
	β_2	-0.535	0.267	-1.481	-1.051	-0.720	-0.533	-0.346	-0.057	0.163
	β_3	-0.001	0.002	-0.006	-0.004	-0.002	-0.001	0.001	0.003	0.005
	β_4	-0.402	0.266	-1.326	-0.921	-0.578	-0.413	-0.229	0.111	0.486
MH	β_0	-6.098	0.188	-6.723	-6.508	-6.213	-6.097	-5.979	-5.731	-5.374
	β_1	0.013	0.003	0.001	0.007	0.011	0.013	0.015	0.020	0.025
	β_2	-0.420	0.189	-1.091	-0.766	-0.543	-0.439	-0.308	0.027	0.485
	β_3	0.000	0.001	-0.004	-0.002	-0.001	0.000	0.001	0.003	0.006
	β_4	-0.363	0.109	-0.750	-0.588	-0.425	-0.370	-0.302	-0.136	0.110
SIR	β_0	-6.097	0.163	-6.550	-6.439	-6.207	-6.116	-5.979	-5.752	-5.596
	β_1	0.014	0.003	0.005	0.008	0.012	0.014	0.015	0.019	0.021
	β_2	-0.417	0.246	-1.162	-1.042	-0.567	-0.412	-0.249	-0.010	0.223
	β_3	0.000	0.001	-0.004	-0.003	-0.001	0.000	0.001	0.002	0.002
	β_4	-0.428	0.100	-0.734	-0.605	-0.515	-0.437	-0.333	-0.262	-0.057
MET	β_0	-6.160	0.151	-6.426	-6.426	-6.218	-6.171	-6.072	-5.830	-5.512
	β_1	0.014	0.003	0.002	0.009	0.013	0.014	0.015	0.019	0.020
	β_2	-0.340	0.164	-1.019	-0.669	-0.471	-0.292	-0.197	-0.042	0.079
	β_3	0.001	0.001	-0.003	-0.002	0.000	0.001	0.002	0.002	0.003
	β_4	-0.282	0.089	-0.651	-0.504	-0.331	-0.263	-0.204	-0.127	0.019

4.2. *Simulation study.* For the simulation study, we use the Poisson regression model and we started with the COPD dataset, and using MH, we obtained the posterior means of the hyperparameters which we denote by $\hat{\beta}^{(t)}$ and $\tau^{(t)}$ to represent the true values. We use the same values for the population sizes and the covariates (n_i, x_i) , $i = 1, \dots, 48$, as for the 48 HSA's in the COPD dataset.

TABLE 6. COPD DATA (NCHS, 1998): COMPARISON OF THE POSTERIOR MEAN (PM), POSTERIOR STANDARD DEVIATION AND 95% CREDIBLE INTERVALS FOR THE MORTALITY RATES FOR MH AND SIR

HSA	NR			MH		
	PM	PSD	CI	PM	PSD	CI
4	47.16	3.68	(40.27, 54.69)	46.88	3.58	(40.07, 54.06)
85	46.84	2.13	(42.77, 51.15)	46.93	2.10	(42.90, 51.13)
121	39.53	1.05	(37.61, 41.64)	39.61	1.13	(37.32, 41.79)
9	43.02	2.59	(38.21, 48.17)	43.18	2.60	(38.29, 48.92)
17	37.36	1.44	(34.67, 40.12)	37.30	1.45	(34.60, 40.27)
38	31.65	0.40	(30.86, 32.45)	31.70	0.43	(30.83, 32.55)
95	41.87	1.64	(38.55, 45.21)	41.84	1.67	(38.69, 45.15)
98	42.92	1.86	(39.29, 46.63)	42.85	1.81	(39.44, 46.44)
136	38.51	1.75	(35.07, 41.96)	38.62	1.69	(35.30, 41.75)
22	31.74	0.99	(29.90, 33.77)	31.76	1.01	(29.83, 33.79)
32	27.17	2.44	(22.40, 32.56)	27.97	2.56	(23.05, 32.86)
68	32.38	1.09	(30.34, 34.67)	32.40	1.05	(30.34, 34.30)
74	31.12	1.36	(28.53, 33.91)	31.21	1.36	(28.48, 33.87)
101	32.64	0.94	(30.78, 34.55)	32.68	0.97	(30.86, 34.64)
112	40.38	0.84	(38.66, 42.03)	40.32	0.85	(38.65, 41.99)
90	31.62	0.75	(30.11, 33.07)	31.64	0.72	(30.27, 32.99)
91	41.12	4.03	(33.53, 49.34)	41.33	4.03	(33.82, 49.06)
138	32.16	1.33	(29.48, 34.70)	32.11	1.43	(29.32, 34.85)
20	47.55	2.42	(42.80, 52.58)	47.36	2.51	(42.73, 52.31)

HSA	SIR			MET		
	PM	PSD	CI	PM	PSD	CI
4	47.90	3.48	(40.97, 55.24)	47.99	3.26	(41.72, 54.88)
85	46.68	2.12	(42.61, 50.90)	46.21	2.00	(42.21, 49.92)
121	39.54	1.16	(37.24, 41.73)	39.55	1.11	(37.32, 41.65)
9	43.38	2.54	(38.39, 48.44)	44.11	2.50	(39.31, 49.15)
17	37.35	1.42	(34.54, 40.20)	37.31	1.33	(34.88, 40.00)
38	31.66	0.40	(30.90, 32.45)	31.66	0.41	(30.86, 32.46)
95	41.92	1.59	(38.79, 45.20)	41.93	1.64	(38.75, 45.23)
98	42.93	1.83	(39.21, 46.57)	42.77	1.78	(39.21, 46.22)
136	38.81	1.68	(35.62, 42.03)	39.05	1.67	(35.98, 42.27)
22	31.83	1.02	(29.83, 33.84)	31.86	1.00	(29.95, 33.96)
32	27.82	2.31	(23.75, 32.71)	29.58	2.29	(25.05, 34.25)
68	32.45	1.07	(30.36, 34.66)	32.39	1.09	(30.39, 34.71)
74	31.16	1.32	(28.57, 33.81)	31.24	1.34	(28.87, 34.06)
101	32.65	0.93	(30.83, 34.44)	32.71	0.93	(30.86, 34.43)
112	40.37	0.83	(38.78, 41.96)	40.40	0.84	(38.79, 42.02)
90	31.64	0.75	(30.18, 33.07)	31.70	0.76	(30.22, 33.22)
91	40.78	3.79	(33.63, 48.36)	41.06	3.23	(34.92, 47.30)
138	32.25	1.46	(29.57, 35.25)	32.26	1.27	(29.94, 34.94)
20	47.33	2.48	(42.64, 52.13)	46.55	2.19	(42.35, 51.01)

NOTE: All values in this table are multiplied by 10^{-4} . We display the first 20 of the 48 HSA's (i.e., random selection).

TABLE 7. OSTEOPOROSIS DATA FOR WHITE FEMALES OLDER THAN 20 YEARS

case	n	d	$\hat{\theta}$	age	race	sex	inc
1	840	36	0.043	0	1	1	0
2	291	24	0.082	0	1	1	1
3	771	110	0.143	0	1	0	0
4	263	53	0.202	0	1	0	1
5	459	12	0.026	0	0	1	0
6	98	1	0.010	0	0	1	1
7	571	39	0.068	0	0	0	0
8	98	9	0.092	0	0	0	1
9	634	148	0.233	1	1	1	0
10	239	25	0.105	1	1	1	1
11	654	404	0.618	1	1	0	0
12	156	82	0.526	1	1	0	1
13	246	29	0.118	1	0	1	0
14	35	1	0.029	1	0	1	1
15	277	103	0.372	1	0	0	0
16	25	7	0.280	1	0	0	1

NOTE: Osteoporosis dataset is a small sample from the Third National Health and Nutrition Examination Survey (NHANES III, 1988–92) in which n is the population size, d is the number of people with osteoporosis, and $\hat{\theta} = d/n$ is the proportion of people with osteoporosis. The covariates are age (0=younger than 50, 1=50 years and older), race (0=white, 1=black), sex (0=male, 1=female) and inc for income (0=less than \$45000, 1=\$45000 and higher).

TABLE 8. OSTEOPOROSIS DATA (NHANES 3, 1988-94): COMPARISON OF THE POSTERIOR MEAN (PM), POSTERIOR STANDARD DEVIATION (PSD) AND PERCENTILES OF THE REGRESSION COEFFICIENTS FOR NR, MH AND SIR

				Percentiles						
		PM	PSD	Min	2.5	25	50	75	97.5	Max
NR	β_0	-2.197	0.525	-3.672	-3.234	-2.556	-2.212	-1.843	-1.124	-0.564
	β_1	1.683	0.556	-0.096	0.626	1.303	1.696	2.059	2.756	3.351
	β_2	0.834	0.341	-0.336	0.124	0.604	0.837	1.056	1.530	1.955
	β_3	-1.665	0.432	-2.976	-2.526	-1.947	-1.662	-1.373	-0.828	-0.338
	β_4	-0.284	0.194	-1.121	-0.664	-0.416	-0.280	-0.154	0.091	0.252
SIR	β_0	-2.243	0.258	-2.962	-2.682	-2.434	-2.241	-2.066	-1.733	-1.468
	β_1	1.716	0.241	1.018	1.309	1.529	1.721	1.855	2.220	2.460
	β_2	0.854	0.229	0.277	0.420	0.711	0.837	1.016	1.281	1.495
	β_3	-1.639	0.257	-2.244	-2.185	-1.787	-1.648	-1.429	-1.200	-0.975
	β_4	-0.311	0.192	-0.799	-0.799	-0.437	-0.264	-0.186	0.021	0.174
MH	β_0	-2.315	0.203	-2.942	-2.658	-2.467	-2.341	-2.181	-1.873	-1.610
	β_1	1.710	0.191	1.128	1.302	1.596	1.717	1.840	2.087	2.315
	β_2	0.909	0.215	0.309	0.517	0.794	0.888	1.026	1.301	1.564
	β_3	-1.724	0.244	-2.207	-2.161	-1.863	-1.697	-1.539	-1.276	-1.120
	β_4	-0.204	0.211	-0.770	-0.582	-0.362	-0.234	-0.047	0.133	0.325
MET	β_0	-2.297	0.279	-2.968	-2.868	-2.540	-2.300	-2.064	-1.802	-1.624
	β_1	1.695	0.232	1.134	1.238	1.534	1.657	1.864	2.140	2.314
	β_2	0.944	0.239	0.280	0.482	0.772	0.934	1.114	1.346	1.556
	β_3	-1.626	0.208	-2.516	-2.056	-1.757	-1.587	-1.508	-1.255	-1.001
	β_4	-0.327	0.257	-0.883	-0.831	-0.480	-0.285	-0.165	0.121	0.447

TABLE 9. OSTEOPOROSIS DATA (NHANES 3, 1988–92): COMPARISON OF NR, MH, SIR AND MET USING THE POSTERIOR MEAN (PM), POSTERIOR STANDARD DEVIATION AND 95% CREDIBLE INTERVALS FOR THE PROPORTION

case	NR			SIR		
	PM	PSD	CI	PM	PSD	CI
1	.044	.007	(.031, .060)	.044	.007	(.031, .060)
2	.074	.016	(.048, .107)	.074	.016	(.048, .107)
3	.148	.012	(.126, .173)	.148	.012	(.126, .173)
4	.193	.025	(.148, .243)	.193	.025	(.148, .243)
5	.027	.008	(.013, .044)	.027	.008	(.013, .044)
6	.016	.014	(.001, .054)	.016	.014	(.001, .054)
7	.073	.012	(.051, .097)	.073	.012	(.051, .097)
8	.089	.032	(.041, .159)	.089	.032	(.041, .159)
9	.231	.016	(.200, .261)	.231	.016	(.200, .261)
10	.119	.020	(.081, .162)	.119	.020	(.081, .162)
11	.614	.020	(.576, .654)	.614	.020	(.576, .654)
12	.520	.039	(.443, .595)	.520	.039	(.443, .595)
13	.114	.021	(.077, .158)	.114	.021	(.077, .158)
14	.069	.039	(.016, .164)	.069	.039	(.016, .164)
15	.373	.028	(.319, .432)	.373	.028	(.319, .432)
16	.305	.065	(.183, .436)	.305	.065	(.183, .436)

case	MH			MET		
	PM	PSD	CI	PM	PSD	CI
1	.043	.007	(.031, .058)	.043	.007	(.030, .057)
2	.073	.014	(.045, .103)	.073	.014	(.047, .101)
3	.147	.012	(.124, .172)	.149	.013	(.125, .173)
4	.194	.023	(.152, .241)	.192	.021	(.151, .235)
5	.025	.007	(.013, .040)	.026	.007	(.014, .041)
6	.012	.009	(.002, .035)	.012	.009	(.001, .035)
7	.071	.010	(.053, .093)	.071	.010	(.052, .094)
8	.086	.022	(.049, .132)	.083	.024	(.044, .136)
9	.230	.017	(.197, .262)	.232	.016	(.201, .264)
10	.119	.019	(.084, .161)	.120	.020	(.083, .159)
11	.613	.018	(.577, .648)	.614	.019	(.578, .651)
12	.526	.036	(.456, .601)	.520	.040	(.444, .593)
13	.112	.018	(.079, .149)	.113	.019	(.077, .150)
14	.061	.027	(.021, .125)	.062	.029	(.017, .127)
15	.367	.027	(.316, .421)	.366	.028	(.311, .422)
16	.305	.059	(.201, .427)	.284	.068	(.141, .416)

We generated 1,000 datasets in two steps. First, we drew the $\theta_i^{(t)}$ from $\theta_i^{(t)} \mid \tau^{(t)}, \beta^{(t)} \stackrel{ind}{\sim} \text{Gamma}(e^{\tau^{(t)}}, e^{\tau^{(t)} - x_i' \beta^{(t)}})$. Second, we drew the d_i from $d_i \mid \theta_i^{(t)} \stackrel{ind}{\sim} \text{Poisson}(n_i \theta_i^{(t)})$. Note that the true mortality rates $\theta_i^{(t)}$ change from one dataset to the other, but $(\beta^{(t)}, \tau^{(t)})$ is the same for all datasets. For a few datasets the Hessian matrices were singular, so we discarded them and generated replacements. Then, we used NR, MH, SIR and MET to

fit the 1,000 datasets in the same manner as for the COPD dataset (see the model in (3) and (5)), pretending that the true values of the hyperparameters are unknown. For each dataset we computed the 95% credible intervals for the $\theta_i^{(t)}$ and $\beta_0^{(t)}, \dots, \beta_4^{(t)}$, and we counted the number of credible intervals containing the true values $(\beta^{(t)}, \tau^{(t)})$ and $\theta_i^{(t)}$.

TABLE 10. SIMULATED DATA: COMPARISON OF THE AVERAGE OF THE END POINTS OF THE 95% CREDIBLE INTERVALS AND THE ESTIMATED PROBABILITY CONTENT (C) FOR β_0, \dots, β_4 FOR NH, MH, SIR AND MET

	NR		MH	
	CI	C	CI	C
β_0	(-6.605, -5.482)	97.5	(-6.511, -5.615)	95.4
β_1	(0.004, 0.022)	97.3	(0.005, 0.020)	94.1
β_2	(-1.150, 0.145)	97.9	(-0.961, 0.063)	96.5
β_3	(-0.004, 0.004)	97.2	(-0.003, 0.003)	96.2
β_4	(-0.813, 0.033)	98.8	(-0.646, -0.104)	95.8
	SIR		MET	
	CI	C	CI	C
β_0	(-6.492, -5.661)	93.0	(-6.488, -5.685)	91.8
β_1	(0.006, 0.020)	92.7	(0.006, 0.020)	90.8
β_2	(-0.925, 0.084)	95.6	(-0.870, 2.301)	92.2
β_3	(-0.003, 0.003)	94.0	(-0.003, 0.003)	92.8
β_4	(-0.615, -0.106)	92.8	(-0.604, -0.110)	90.4

NOTE: C is the percent intervals that contain the true value.

In Table 10 we present the average of the end points of the 1,000 95% credible intervals for $\beta_0^{(t)}, \dots, \beta_4^{(t)}$ and the proportion of intervals containing the true values. The 95% credible intervals are mostly similar for MH and SIR, but not so similar to NR and MET. While the coverage of the 95% credible intervals based on MET are below the nominal value, the 95% credible intervals based on NR are above the nominal value. SIR is closest to the nominal 95% coverage for β_2 and β_3 , and not too badly off for β_0 , β_1 and β_4 .

For the $\theta_i^{(t)}$, we first did exactly the same as for $\beta_0^{(t)}, \dots, \beta_4^{(t)}$ for the 48 HSA's. Then, we compute the ratios of the end points of the intervals relative to those obtained from the observed data (denoted by R_L and R_U) and the ratios of the probability contents relative to the nominal value of 95% (denoted by R_C). The five-number summaries of R_L , R_U and R_C over the 48 HSA's are presented in Table 11, and they show that the NR, MH and SIR methods are remarkably similar. MET shows some deviation from NR, MH and SIR: R_L is too small and R_U too large.

TABLE 11. SIMULATED DATA: COMPARISON OF THE FIVE-NUMBER SUMMARIES OF R_L , R_U AND R_C , CORRESPONDING TO THE MORTALITY RATES FOR NR, MH, SIR AND MET

		Min	25	50	75	Max
R_L	NR	0.831	0.939	0.982	1.056	1.310
	MH	0.833	0.947	0.986	1.056	1.290
	SIR	0.847	0.954	0.989	1.067	1.310
	MET	0.601	0.805	0.885	0.975	1.221
R_U	NR	0.877	0.957	1.004	1.058	1.252
	MH	0.879	0.960	1.000	1.063	1.229
	SIR	0.872	0.959	0.988	1.062	1.232
	MET	0.917	1.032	1.089	1.188	1.638
R_C	NR	0.991	1.001	1.006	1.011	1.031
	MH	0.986	1.001	1.007	1.012	1.034
	SIR	0.980	0.994	1.003	1.010	1.021
	MET	0.976	0.994	1.002	1.008	1.021

NOTE: R_L and R_U are the ratios of the average end points of the intervals relative to those obtained from the observed data and R_C is the ratio of the probability content relative to the nominal value of 95% for each of the 48 mortality rates.

5 Discussion

We have shown that it is possible to avoid the use of MCMC methods to fit two-stage generalized linear models when there is conditional conjugacy. We have exemplified our procedure using a hierarchical Poisson regression model which is normally fit using MCMC methods. We have also illustrated our method using a relatively new logistic regression model. Using the SIR algorithm we have shown that we can simply use random samples rather than Markov chains, and we have obtained Rao-Blackwellized estimators of the area effects. We have shown how to “tune” the normal approximation based on the mode-Hessian approximation through the SIR algorithm. In four examples and a simulation study, we have shown that NR (PRIMM), MH, SIR and MET give very similar inference for the rates and proportions and slightly less similar inference for the regression coefficients. Indeed, SIR is a good competitor to MH or MET.

The SIR algorithm is particularly useful in problems where the posterior density is extraordinarily complex. For example, suppose given data \underline{d} , the posterior density of $\underline{\theta}$ can be written as

$$\pi(\underline{\theta} \mid \underline{d}) = \pi_a(\underline{\theta} \mid \underline{d})R(\underline{\theta}; \underline{d}). \tag{18}$$

Suppose a sample can be drawn from $\pi_a(\underline{\theta} \mid \underline{d})$ but the computation of $R(\underline{\theta}; \underline{d})$ is too time-consuming. For example, $R(\underline{\theta}; \underline{d})$ might involve multi-dimensional integrals distinct from the normalization constant. Then, a

Metropolis-Hastings sampler may be prohibitive, and in this case $R(\theta; \underline{d})$ can be computed in an output analysis where the sampling path is already “thinned out”, so that $R(\theta; \underline{d})$ needs to be computed a fewer number of times. Here, $\pi_a(\theta | \underline{d})$ can be drawn using the Gibbs sampler or a non-Markovian (random) sampler. There are several examples in sample surveys (e.g., selection bias and nonresponse bias); see Nandram and Choi (2002), Nandram and Sedransk (2004) and Burgos and Nandram (2003).

Rubin (1987, p. 192) recommends using probability proportional to size (PPS) sampling for the SIR algorithm. His suggestion requires the population size N to be large relative to the sample size M , which he accommodates by taking $N/M \approx 20$. In our case, for a sample of $M = 1000$, his suggestion requires $N = 20000$. He also suggests choosing N/M adaptively, but this is still an open problem. See Cochran (1977, p. 250, 265) for a discussion of PPS sampling with or without replacement. Even with PPS sampling “without replacement” some units can enter the sample several times (especially true in our application) and the procedure, therefore, does not provide independent samples. Thus, for simplicity, in our work we have used PPS sampling with replacement in the SIR algorithm. Further research is needed to construct a more efficient SIR algorithm.

We believe that our approach would be preferred by many practitioners because it is conceptually simpler as its basis is independent samples not Markov chains, and the associated diagnostic procedures to monitor convergence disappear. Of course, for multi-dimensional problems one must rely on Markov Monte Carlo methods. For two-stage hierarchical Bayesian models with conditional conjugacy this is not necessary, and there is the additional benefit of Rao-Blackwellized estimators for area effects.

Appendix A. Hessian Matrix of $\log[\pi(\beta, \tau | \underline{d})]$ for the Gamma-Poisson Model

Letting $(\hat{\beta}, \hat{\tau})$ denote the posterior mode of $\pi(\beta, \tau | \underline{d})$ and $\psi(\cdot)$ and $\psi'(\cdot)$ denote the digamma and trigamma functions respectively, for $i = 1, \dots, \ell$, we define

$$\begin{aligned} w_i &= e^{\hat{\tau} - x'_i \hat{\beta}} (n_i + e^{\hat{\tau} - x'_i \hat{\beta}})^{-1}, \\ a_i &= \psi(d_i + e^{\hat{\tau}}) - \psi(e^{\hat{\tau}}) + \hat{\tau} + 1 - \ln(n_i + e^{\hat{\tau} - x'_i \hat{\beta}}) + e^{-x'_i \hat{\beta}} (d_i + e^{\hat{\tau}}) (n_i + e^{\hat{\tau} - x'_i \hat{\beta}})^{-1}, \\ b_i &= \psi'(d_i + e^{\hat{\tau}}) - \psi'(e^{\hat{\tau}}) + e^{-\hat{\tau}} + [-2n_i e^{-x'_i \hat{\beta}} + \{y_i - e^{\hat{\tau}}\} e^{2(\hat{\tau} - x'_i \hat{\beta})}] (n_i + e^{\hat{\tau} - x'_i \hat{\beta}})^{-2}. \end{aligned}$$

Then the Hessian matrix evaluated at the posterior mode is $H = \begin{bmatrix} B & \nu \\ \nu' & t \end{bmatrix}$,

where

$$\begin{aligned}
 B &= -\sum_{i=1}^{\ell} w_i(1-w_i)(d_i + e^{\hat{\tau}})x_i x'_i, \\
 \nu &= e^{\hat{\tau}} \sum_{i=1}^{\ell} [w_i\{1 + (1-w_i)(1 + d_i e^{-\hat{\tau}})\} - 1]x_i, \\
 t &= e^{-\hat{\tau}} \left\{ \sum_{i=1}^{\ell} a_i - 2(1 + e^{\hat{\tau}})^{-1} \right\} + e^{-2\hat{\tau}} \left\{ \sum_{i=1}^{\ell} b_i + 2(1 + e^{\hat{\tau}})^{-2} \right\}.
 \end{aligned}$$

Appendix B. Hessian Matrix of $\log[\pi(\beta, \tau | d)]$ for the Beta-Binomial Model

Let $p(\tau) = a_0 e^{\tau} / (a_0 + e^{\tau})^2$ and $\phi_i(\beta) = e^{x_i \beta} / (1 + e^{x_i \beta})$, $i = 1, \dots, \ell$. It follows that the first and second derivatives of $p(\tau)$ are $p'(\tau) = a_0 e^{\tau} (a_0 - e^{\tau}) / (a_0 + e^{\tau})^3$ and $p''(\tau) = a_0 e^{\tau} (a_0 - 4a_0 e^{\tau} + e^{2\tau}) / (a_0 + e^{\tau})^4$. The first and second derivatives of $\phi_i(\beta)$ are $\phi'_i(\beta) = x_i e^{x_i \beta} / (1 + e^{x_i \beta})^2$ and $\phi''_i(\beta) = x'_i x_i e^{x_i \beta} (1 - e^{x_i \beta}) / (1 + e^{x_i \beta})^3$.

Let $(\hat{\beta}, \hat{\tau})$ denote the posterior mode of $\pi(\beta, \tau | d)$ and for $i = 1, \dots, \ell$, let

$$\begin{aligned}
 d_{i1} &= \psi\{s_i + e^{\hat{\tau}} \phi_i(\hat{\beta})\} - \psi\{e^{\hat{\tau}} \phi_i(\hat{\beta})\}, \\
 d_{i2} &= \psi\{n_i - s_i + e^{\hat{\tau}} (1 - \phi_i(\hat{\beta}))\} - \psi\{e^{\hat{\tau}} (1 - \phi_i(\hat{\beta}))\}, \\
 d_{i3} &= \psi'\{s_i + e^{\hat{\tau}} \phi_i(\hat{\beta})\} - \psi'\{e^{\hat{\tau}} \phi_i(\hat{\beta})\}, \\
 d_{i4} &= \psi'\{n_i - s_i + e^{\hat{\tau}} (1 - \phi_i(\hat{\beta}))\} - \psi'\{e^{\hat{\tau}} (1 - \phi_i(\hat{\beta}))\}.
 \end{aligned}$$

Finally, the Hessian matrix evaluated at the posterior mode $(\hat{\beta}, \hat{\tau})$ is

$$H = \begin{bmatrix} B & \nu \\ \nu' & t \end{bmatrix} \text{ where}$$

$$\begin{aligned}
 B &= \sum_{i=1}^{\ell} e^{\hat{\tau}} \phi''_i(\hat{\beta}) \{d_{i1} - d_{i2}\} + (e^{\hat{\tau}} \phi'_i(\hat{\beta}))^2 \{d_{i3} - d_{i4}\}, \\
 t &= \sum_{i=1}^{\ell} e^{2\hat{\tau}} \left\{ (\phi_i(\hat{\beta}))^2 d_{i3} + (1 - \phi_i(\hat{\beta}))^2 d_{i4} - [\psi'\{n_i + e^{\hat{\tau}}\} - \psi'\{e^{\hat{\tau}}\}] \right\} \\
 &\quad + e^{\hat{\tau}} \left\{ \phi_i(\hat{\beta}) d_{i1} + (1 - \phi_i(\hat{\beta})) d_{i2} - [\psi\{n_i + e^{\hat{\tau}}\} - \psi\{e^{\hat{\tau}}\}] \right\} \\
 &\quad + \{p''(\hat{\tau})p(\hat{\tau}) - (p'(\hat{\tau}))^2\} / (p(\hat{\tau}))^2, \\
 \nu &= \sum_{i=1}^{\ell} e^{\hat{\tau}} \phi'_i(\hat{\beta}) \{d_{i1} - d_{i2}\} + e^{2\hat{\tau}} \phi'_i(\hat{\beta}) \left\{ \phi_i(\hat{\beta}) d_{i3} - (1 - \phi_i(\hat{\beta})) d_{i4} \right\}.
 \end{aligned}$$

References

- ALBERT, J. (1988). Bayesian estimation of Poisson means using a hierarchical log-linear model. In *Bayesian Statistics 3*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, eds., Oxford University Press, New York, 519-531.
- ALBERT, J. (1989). Nuisance parameters and the use of exploratory graphical methods in a Bayesian analysis, *Amer. Statist.*, **43**, 191-196.
- BURGOS, C.G. and NANDRAM, B. (2003). Bayesian predictive inference and selection bias, *J. Res. Science Engineering*, **1** 1-10.
- CHIB, S. and GREENBERG, E. (1995). Understanding the Metropolis-Hastings algorithm, *Amer. Statist.*, **49**, 327-335.
- COCHRAN, W.G. (1977). *Sampling Techniques*, 3rd edition, Wiley, New York.
- CHRISTIANSEN, C.L. and MORRIS, C.N. (1997). Hierarchical Poisson regression modeling, *J. Amer. Statist. Assoc.*, **92**, 618-631.
- FAY, R.E. and HERRIOT, R.A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data, *J. Amer. Statist. Assoc.*, **74**, 269-277.
- GELFAND, A.E. and SMITH, A.F.M. (1990). Sampling-based approaches to calculating marginal densities, *J. Amer. Statist. Assoc.*, **85**, 398-409.
- GELMAN, A., ROBERTS, G.O. and GILKS, W.R. (1996). Efficient Metropolis jumping rules. In *Bayesian Statistics 5*, J.M. Bernardo, J.O. Berger, A.P. Dawid, and A. F. M. Smith, eds., Oxford University Press, Oxford, 599-607.
- GILKS, W.R. and WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling, *Appl. Statist.*, **41**, 337-348.
- GILKS, W.R., BEST, N.G. and TAN, K.K.C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling, *Appl. Statist.*, **44**, 455-472.
- HASTINGS, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**, 97-109.
- LIU, J.S. (1994). The collapsed Gibbs sampler in Bayesian computations with application to a gene regulation problem, *J. Amer. Statist. Assoc.*, **89**, 958-966.
- LIU, J.S., LIANG, F. and WONG, W.H. (2000). The Multiple-Try Method and Local Optimization in Metropolis Sampling, *J. Amer. Statist. Assoc.*, **95**, 121-134.
- NANDRAM, B. (2000). Bayesian generalized linear models for inference about small areas. In *Generalized Linear Models: A Bayesian Perspective*, D.K. Dey, S.K. Ghosh, and B.K. Mallick, eds., Marcel Dekker, New York, 91-114.
- NANDRAM, B. and CHOI, J.W. (2002). Hierarchical Bayesian nonresponse models for binary data from small areas with uncertainty about ignorability. *J. Amer. Statist. Assoc.*, **97**, 381-388.
- NANDRAM, B., LIU, J. and CHOI, J.W. (2004). A comparison of the posterior choropleth maps for disease mapping, *J. Data Science*, **3**, (in press).
- NANDRAM, B. and SEDRANSK, J. (2004). A Bayesian predictive inference for finite population quantities under informative sampling, Technical Report, Worcester Polytechnic Institute.
- NANDRAM, B., SEDRANSK, J. and PICKLE, L.W. (2000). Bayesian analysis and mapping of mortality rates for chronic obstructive pulmonary disease, *J. Amer. Statist. Assoc.*, **95**, 1110-1118.

- NATIONAL CENTER FOR HEALTH STATISTICS (1998). *Health, United States*, DHHS Publication Number (PHS) 98-1232, National Center for Health Statistics, Hyattsville, MD.
- METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N., TELLER, A.H. and TELLER, E. (1953). Equations of state calculations by fast computing machines, *J. Chemical Physics*, **21**, 1087-1092.
- RUBIN, D. (1987). *Multiple Imputation for Nonresponse in Surveys*, Wiley, New York.
- SMITH, A.F.M. and GELFAND, A.E. (1992). Bayesian statistics without tears: a sampling-resampling perspective, *Amer. Statist.*, **46**, 84-88.
- TANNER, M.A. (1993). *Tools for Statistical Inference: Methods for Exploration of Posterior Distributions and Likelihood Functions*, Springer-Verlag, New York.
- TIERNEY, L. (1994). Markov chains for exploring posterior distributions, (with discussions), *Ann. Statist.*, **22**, 1701-1762.
- WALLER, L., CARLIN, B., XIA, H. and GELFAND, A.E. (1997). Hierarchical spatio-temporal mapping of disease rates, *J. Amer. Statist. Assoc.*, **92**, 607-617.

BALGOBIN NANDRAM AND ERIK BARRY ERHARDT

DEPARTMENT OF MATHEMATICAL SCIENCES

WORCESTER POLYTECHNIC INSTITUTE

100 INSTITUTE ROAD, WORCESTER

MA 01609-2280.

E-mail: balnan@wpi.edu

erike@wpi.edu

Paper received: April 2004; revised June 2004.