



Markov Chain Monte Carlo-Based Bayesian Analysis of Binary Response Regression, with Illustration in Dose-Response Assessment

Crispin M. Mutshinda

Department of Mathematics and Statistics (Gustaf Hällströmin katu 2b)

University of Helsinki

P.O. Box 68, FIN-00014, Helsinki, Finland

Tel: 3-58-9191-51439 E-mail: mutshinda@gmail.com

Abstract

This paper deals with the Bayesian analysis of binary response regression using Markov chain Monte Carlo (MCMC) methods, more specifically the Metropolis sampler, for posterior simulation. The methodology is illustrated with real-world data from a bioassay experiment. Inference about quantities of typical interest in the dose-response setting such as the lethal dose is discussed as well. MCMC are routinely implemented through popular Bayesian software such as Win-/Open-BUGS. However, these remain black boxes which provide no insight in the estimation procedure. This paper exemplifies that developing and implementing an MCMC sampler may, in many practical situations, be relatively straightforward. The R code for the Metropolis sampler is also provided in an appendix to the paper.

Keywords: Bayesian logistic regression, Generalized linear models, LD50, Metropolis sampler, Toxicological analysis

1. Introduction

Dose-response (D-R) analyses represent an important pillar upon which toxicological analyses regarding drugs and putative environmental contaminants are built. They allow identifying a range of intakes (doses) encompassing an appropriate benefit-to-risk ratio. Acute toxicity of drugs and various chemical compounds is typically assessed through bioassay (biological assay) experiments. These consist of measuring the effects of a substance on living organisms under controlled conditions. Typically, batches of animals are administered various doses of a drug or a chemical (e.g., an insecticide), and the outcomes are usually recorded on a dichotomous scale: dead/alive, tumor/no tumor, etc. The same type of data may arise from observational studies on environmental contamination involving humans, animals or plants, by relating a health defect outcome to the level of exposure to an alleged contaminant, using appropriate proxies for exposure. Such an experiment or observational study typically yields data of the form (x_i, n_i, y_i) where y_i denotes the number of individuals with positive outcome at dose x_i out of a total of n_i individuals.

Developing D-R models is central to determining “safe” and “hazardous” levels and dosages of drugs, potential pollutants and other substances to which humans, animals and plants are exposed. Conclusions of D-R analyses often serve as basis for public health regulations that limit exposure to hazardous substances (e.g., Mutshinda et al. 2008).

Statistical analyses of D-R data have mainly revolved around the maximum-likelihood method (e.g. Metz et al. 1982 amongst many others). However, the steady increase in popularity of the Bayesian methodology (Clark 2005) has prompted substantial interest in implementing it in the D-R and exposure risk assessment settings (e.g., French et al. 2002; Teunis et al. 2002; Englehardt 2004; Englehardt & Swartout. 2004, 2006; Mutshinda et al. 2008). Bayesian inference offers attractive options whereby available information such as results from previous studies or experts’ opinion in the form of a prior distribution can rationally be combined with information conveyed by empirical data and encoded in the likelihood function to boost the power of estimation. Owing to advances in sampling-based methods such as Markov chain Monte Carlo (MCMC) (Gilks et al. 1996) and the spread of related software in the vein of Win-/Open-BUGS (Spiegelhalter et al. 2003; Thomas et al. 2006), the computational burden that has long hampered the broad adoption of the Bayesian approach, or limited its application to the computationally convenient conjugate analysis should overall no longer be an issue in itself.

This paper discusses the Bayesian analysis of binary response regression using MCMC, more specifically the (random walk) Metropolis sampler for posterior simulation. The methodology is illustrated in a practical example using real-world data from a bioassay experiment. In spite of the relatively large body of literature on Bayesian dose-response assessment, the subject is worthy of additional attention, in particular as numerical methods such as the MCMC

implemented here are concerned. Indeed, the principles of these methods still need to be made broadly accessible to applied scientists. Notwithstanding the user-friendliness of available Bayesian software packages, these remain black boxes which provide no insight in the estimation process. This paper exemplifies that, in many practical cases, developing and implementing a MCMC sampler may be relatively straightforward once the underlying principles are well understood. In addition to a concise tutorial on generalized linear models under the umbrella of which falls binary response regression, this paper contributes to the literature by providing a handy explanation of the MCMC algorithm known as the Metropolis sampler, along with an illustration of its implementation in the increasingly popular R environment.

This paper is structured as follows. Section 2 presents the materials and methods. Section 3 deals with the analysis of the bioassay data, and section 4 is devoted to some conclusions.

2. Materials and methods

2.1 Description of data

We use real data from a bioassay experiment as presented Table 1. The data were reproduced (with permission) from Gelman et al. (2003), and are used here solely for illustrative purposes. The dichotomous outcomes are death/survival, and y_i denotes the number of animals with positive outcome at dose x_i (on the log-scale) out of n_i . Gelman et al. (2003) analyzed the same dataset with a Bayesian approach by approximating the posterior distribution in two ways: (i) using crude simulations, and (ii) using the normal approximation to the posterior. Here we use MCMC to directly simulate samples from the joint posterior of the model parameters. By using the same dataset, we have the opportunity to compare the results across different methods, including the method of maximum likelihood which is briefly carried out in section 2.2.2.

2.2 Methods

The probability of positive outcome is related to the dose level through a logit (logistic regression) model. Logistic regression falls into the category of generalized linear models (GLMs) which underlie most of the statistical analyses in applied research. We revisit GLMs in subsection 2.2.1 without any attempt to presenting an exhaustive discussion. Readers interested in more details are referred to appropriate literature which includes McCullough & Nelder (1989), Lindsey (1997), and Dobson (2002). In subsection 2.2.2 we fit the model to the bioassay data using the maximum likelihood method, and in subsection 2.2.3, we perform a MCMC-aided Bayesian analysis, simulating samples from the joint posterior of interest through a Metropolis sampler.

2.2.1 Generalized linear models: an overview

GLMs extend the ordinary linear regression (OLR) model to settings where assumptions of the latter are not met. This is so when: (i) the relation between, $\mu = E(Y)$ and X is not linear (in GLMs, it is rather a transformation of $\mu = E(Y)$ through a function $g(\cdot)$ called link function which linearly related to X as we discuss below), (ii) the variance is not constant across responses (OLR assumes constant variance across the responses), and (iii) the response variables are not Gaussian. This last assumption is nonetheless not critical. Indeed, in the OLR setting, $Y = X\beta + \varepsilon$, where Y is a (column) vector of n iid Gaussian responses, X is a design matrix with n rows and $(p+1)$ columns where p is the number of explanatory variables and the first column of X corresponds to the intercept. β is a (column) vector of the $(p+1)$ regression parameters (including the intercept), and ε is a (column) vector of the n residuals which are assumed to be normally distributed with mean zero and variance σ^2 i.e. $\varepsilon \sim N(0, \sigma^2 I_n)$. The ordinary least squares (OLS) estimator of β is given by $\tilde{\beta} = (X^T X)^{-1} X^T Y$, and its covariance matrix is $Var(\tilde{\beta}) = \{(X^T X)^{-1} X^T\} Var(Y) \{X(X^T X)^{-1}\} = \sigma^2 \{(X^T X)^{-1} X^T\} \{X(X^T X)^{-1}\}$ since $Var(Y) = \sigma^2 I_n$. Moreover, $\tilde{\beta}$ is Gaussian as a linear transformation of Y which is Gaussian by assumption in the OLR setting. Nevertheless, even if the responses, Y_i , are not normal, the OLS estimator $\tilde{\beta}$, which coincides with the MLE $\hat{\beta}$ is still asymptotically normal as long as the variance is constant, by virtue of the renowned central limit theorem.

A GLM consists of the following three components. First, a sequence of independent responses y_1, y_2, \dots, y_n assumed to be realizations of random variables Y_1, Y_2, \dots, Y_n which share the same form of parametric distribution from the exponential dispersion family. That is, the probability density function (pdf) of Y_i or the probability mass function (pmf) Y_i if is discrete can be written in the form

$$f(y_i) = \exp \{w_i [y_i \theta_i - b(\theta_i)] / \phi + c(y_i, \phi)\}, \quad (1)$$

where θ_i is the so-called natural (or canonical) parameter, ϕ is a scale parameter and the w_i 's are known weights which in the usual case are all equal to 1. Second, a systematic part or linear predictor $\eta_i = x_i^T \beta$, where

$X_i^T = (X_{i,1}, \dots, X_{i,p})$ is the i^{th} row of the design matrix $X_{n \times p}$, and β is a (column) vector of the $(p+1)$ regression parameters. The systematic part η_i is still linear (in the parameters!). And third, a monotonic and differentiable function $g(\cdot)$ called link function through which the systematic part $X_i^T \beta$ is related to the mean value $\mu_i = E(Y_i)$. More specifically $\theta_i(\mu_i) = g(\mu_i) = X_i^T \beta$ or equivalently, $\mu_i = g^{-1}(X_i^T \beta)$.

The exponential family includes a wide range of common distributions such as the normal, the binomial, the Poisson, and the gamma distributions. Below we derive the canonical forms of the three first distributions.

The normal pdf is $f(y_i) = 1/(2\pi\sigma^2)^{1/2} \exp(-(y_i - \mu_i)^2 / 2\sigma^2)$, $-\infty < y < \infty$, which can be written as $f(y_i) = \exp[(y_i \mu_i - \mu_i^2 / 2) / \sigma^2 - y_i^2 / 2\sigma^2 - (1/2) \log(2\pi\sigma^2)]$ which is in the canonical form (1) with $\theta_i = \mu_i$, $b(\theta_i) = \theta_i^2 / 2$, $\phi = \sigma^2$, $w_i = 1$, and $c(y_i, \phi) = -y_i^2 / 2\phi - (1/2) \log(2\pi\phi)$.

The binomial pmf is $f(y_i | \pi_i) = k \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i}$ where $k = n_i! / y_i! (n_i - y_i)!$ which can be written in the canonical form as $\exp\{y_i \log(\pi_i / [1 - \pi_i]) + n_i \log(1 - \pi_i) + \log(k)\}$, where $\theta_i = \log(\pi_i / [1 - \pi_i])$, i.e. $\pi_i = \exp(\theta_i) / [1 + \exp(\theta_i)]$, $b(\theta_i) = -n_i \log(1 - \pi_i) = n_i \log(1 + \exp(\theta_i))$, $\phi = 1$, $c(y_i, \phi) = \log(k)$, and $w_i = 1$.

The Poisson pmf is $f(y_i | \lambda_i) = e^{-\lambda_i} \lambda_i^{y_i} / y_i! = \exp(y_i \log \lambda_i - \lambda_i - \log y_i!)$ which is in the canonical form with $\theta_i(\lambda_i) = \log \lambda_i$, $b(\theta_i) = \exp(\theta_i)$, $\phi = 1$, $w_i = 1$, and $c(y_i, \phi) = -\log(y_i!)$.

It is worth emphasizing that for distributions belonging to the exponential family, the mean $\mu_i = E(Y_i)$, is related to the variance $\text{Var}(Y_i | \eta_i)$ through a special relationship, $V(\mu_i)$, called the variance function. Moreover, It can be shown that $\mu_i = E(Y_i) = \partial b(\theta_i) / \partial \theta_i$, and $\text{Var}(Y_i) = (\phi / w_i) \partial^2 b(\theta_i) / \partial \theta_i^2$. ϕ is usually a nuisance parameter. The variance function is defined as $V(\mu_i) = \partial^2 b(\theta_i(\mu_i)) / \partial \theta_i^2$. These relations can be easily derived in the following way (e.g. Lindsey 1997, p. 12 -13). For any likelihood function $L(\theta_i, \phi; Y_i)$, the quantity $U_i = \partial \log L(\theta_i, \phi; Y_i) / \partial \theta_i$ is commonly known as the score function, and in standard inference, it is known that $E(U_i) = 0$ and $\text{Var}(U_i) = E(-\partial U_i / \partial \theta_i)$. For distributions that belong to the exponential family, $U_i = w_i [Y_i - b'(\theta_i)] / \phi$. Hence, $E(U_i) = 0$ implies $E(Y_i) = b'(\theta_i)$ as required. Moreover, $\text{Var}(U_i) = (w_i / \phi)^2 \text{Var}(Y_i)$, and $E(\partial U_i / \partial \theta_i) = -(w_i / \phi) b''(\theta_i)$. The requirement $\text{Var}(U_i) = E(-\partial U_i / \partial \theta_i)$ yields $\text{Var}(Y_i | \eta_i) = b''(\theta_i) (\phi / w_i)$. Consequently, $V(\mu_i) = b''(\theta_i)$.

The function $b(\theta)$ is very important in GLMs as its first and second derivatives *with respect to the natural parameter* provide the mean and the variance functions, respectively. In the binomial case for instance, $E(Y_i) = b'(\theta_i) = n_i [\exp(\theta_i) / (1 + \exp(\theta_i))] = n_i \pi_i$, and $b''(\theta_i) = n_i [\exp(\theta_i) / (1 + \exp(\theta_i))^2]$ which is nothing but $n_i \pi_i (1 - \pi_i)$ in agreement with what we know about the binomial distribution. It can be easily checked that $V(\mu) = \mu$ for Poisson regression (with log link), $V(\mu) = \mu(1 - \mu)$ for logistic regression, $V(\mu) = \phi \mu^2$ for gamma (with inverse link $g(\mu) = 1 / \mu$), and $V(\mu) = \text{constant}$ for ordinary linear regression with identity as link function.

2.2.2 Maximum likelihood estimation of the binary response regression model, and application to the bioassay dose-response data

Let $f_\theta(X)$ be a statistical model (i.e. a family of distributions indexed by the unknown parameter θ) describing a process of interest, and let $X = X_1, \dots, X_n$ be independent and identically distributed (iid) random variables with density f_θ . The maximum likelihood method estimator (MLE) of θ is the value $\hat{\theta}$ which maximizes the likelihood $L(\theta; X) = p_\theta(X_1, \dots, X_n) = \prod_{i=1}^n f_\theta(X_i)$ of the observed data, which is the joint distribution of X evaluated viewed as a function of θ . Since the log function is monotonic and increasing, $L(\theta; x)$ and $l(\theta; X) = \log L(\theta; X)$ take their maximum at the same point, and it is often more convenient to work with the logarithm of the likelihood function $l(\theta; X) = \sum_{i=1}^n \log f_\theta(X_i)$ known as the log-likelihood function. The maximum likelihood estimator (MLE) $\hat{\theta}(X)$ is a random variable, since its value (the estimate) depends on a realization $x = (x_1, \dots, x_n)$ i.e. the data at hands. $\hat{\theta}(x)$ is then called the maximum likelihood estimate based on x . Maximum likelihood estimates have desirable large sample properties: as the sample size increases, they become unbiased,

efficient, and normally distributed with the “true” value as mean and asymptotic covariance matrix given by the inverse of the Fisher information matrix (the negative of the Hessian matrix of log-likelihood evaluated at the maximum likelihood estimate). Moreover, MLEs are re-parameterization invariant (see below), and the likelihood function can be used to construct confidence intervals and to test hypotheses about models and parameters. More details on maximum likelihood estimation can be found in many statistical books such as Casella & Berger (2001) or Garthwaite et al. (2002). Finding MLEs is an optimization problem which often requires iterative procedures.

In a dose-response setting, the outcomes of different trials are assumed to be (conditionally) independent given the level of dose. Hence, under controlled conditions, the dose-response relationship can be described by a logistic regression model. That is,

$$y_i | x_i \sim \text{Bin}(n_i, \pi_i), \text{ where } \eta_i = \text{logit}(\pi_i) = \beta_1 + \beta_2 x_i, \quad (i = 1, \dots, k) \quad (2)$$

where k is the number of dose levels, β_1 and β_2 are real-valued regression parameters. Consequently, the probability of positive outcome at dose level x_i is given by $\pi_i = \exp(\beta_1 + \beta_2 x_i) / [1 + \exp(\beta_1 + \beta_2 x_i)]$.

The regression parameters are often interpreted in terms of odds ratios. The log (odds) for the probability of positive response at dose level x is $\text{logit } \pi = \beta_1 + \beta_2 x$. If the dose increases by one unit from x to $(x + 1)$, the log (odds) becomes $\beta_1 + \beta_2 x + \beta_2$. By taking the exponential of the log (odds) to return to the odds scale, it turns out that the odds are multiplied by a factor of $\exp(\beta_2)$. Therefore, $\exp(\beta_2)$ is interpretable as an odds ratio; it is the multiplicative increase in the odds of positive response due to one unit increase (on the log-scale) in the level of dose. If the log is in base 2, each one unit increase corresponds to a doubling of the dose. A negative β_2 would suggest that increasing the dose does not increase the probability of having a positive response. In the case of a categorical explanatory variable with two categories $\exp(\beta_2)$ is just the odds ratio for one category compared with the other. Let $Z_i = 1 / [1 + \exp(\beta_1 + \beta_2 x_i)]$. Then the likelihood function (up to a constant of proportionality) is $L(\beta) \propto \prod_i (1 - Z_i)^{y_i} Z_i^{(n_i - y_i)}$, and the log-likelihood is $l(\beta) = \sum_i y_i \log(1 - Z_i) + (n_i - y_i) \log(Z_i)$.

The maximum likelihood estimator (MLE), $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)$ is obtained by (numerically) maximizing the (log)-likelihood with respect to each component of β . Presumably all standard software packages include functions for fitting GLMs. We used the *glm(.)* function built-in to the R statistical package (R script in appendix A) to carry out the ML model fitting to the bioassay data which yielded the ML estimate $\hat{\beta} = (0.847, 7.75)$. Elements of the asymptotic covariance matrix, C , of $\hat{\beta}$ were estimated to be $C_{1,1} = 1.04$, $C_{1,2} = 3.54$, and $C_{2,2} = 23.74$. The maximum likelihood estimate is pretty close to $(0.88, 7.93)$, the mean of Gelman et al.’s normal approximation to the posterior.

With mortality data, a typical quantity of interest is the so-called lethal dose which represents the level of dose resulting in the death of a specific proportion of experimental subjects (typically animals in a preclinical experiment). The cure effective dose is similarly defined as the dose resulting in the cure of a specific proportion of experimental subjects. Usually, the dose resulting in the death of half of the subjects, the so-called *LD50* henceforth denoted by φ is reported.

The formula $\varphi = -\beta_1 / \beta_2$, ($\beta_2 \neq 0$) is easily derived by solving $(1 + \exp(\beta_1 + \beta_2 * \varphi))^{-1} = 0.5$ for φ .

A handy feature of MLEs is their re-parameterization invariance property, by which the MLE of a function of parameters is just the same function of the appropriate MLE. It follows from this property that the MLE for φ is simply given by $\hat{\varphi} = -\hat{\beta}_1 / \hat{\beta}_2$, ($\hat{\beta}_2 \neq 0$). To derive a confidence interval for $\hat{\varphi}$, however, one needs to estimate its variance through what is known as the delta method (e.g., Casella & Berger 2002), which can be stated as follows. Let $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_d)$ be the maximum likelihood estimator of a d -dimensional parameter $\beta = (\beta_1, \dots, \beta_d)$ based on a sufficiently large-sample, with $\text{Var}(\hat{\beta}) = V$, and let $\varphi = g(\cdot)$ be a scalar and differentiable function of β . Then $\hat{\varphi} = g(\hat{\beta})$ is the MLE of φ . In addition, $\hat{\varphi}$ is asymptotically Gaussian with a distribution centred at the true value, φ , and approximate covariance variance matrix W given by $W = (\nabla \hat{\varphi})^T V (\nabla \hat{\varphi})$, where $\nabla \hat{\varphi} = (\partial \varphi / \partial \beta_1, \dots, \partial \varphi / \partial \beta_d)^T |_{\hat{\beta}}$ and T denotes matrix transposition. So, an expression for $\text{Var}(\hat{\varphi})$ can be derived via the delta method. Below we consider a MCMC-aided Bayesian approach.

2.2.3 Bayesian estimation of the binary response regression model via random walk Metropolis sampling

In the Bayesian paradigm, information conveyed by empirical data y through the likelihood function $p(y | \theta)$ is combined with prior information about the model parameter θ , encoded in a prior distribution $p(\theta)$, into a posterior

distribution $p(\theta|y)$ deriving from the joint distribution $p(y, \theta) = p(y|\theta)p(\theta)$ according to the celebrated Bayes' theorem as

$$p(\theta|y) = p(y|\theta)p(\theta) / p(y) \propto p(y|\theta)p(\theta) \tag{3}$$

where $p(y) = \int_{\Theta} p(\theta) p(y|\theta) d\theta$, and Θ the range of all possible values of θ .

The posterior distribution is the legitimate tool for Bayesian inference whose conclusions are essentially made in terms of probabilistic statements about plausible parameter values or outcomes of future observations.

The derivation of the posterior distribution may be difficult to impossible, owing to integration involved in the normalizing constant $p(y)$. Historically, Bayesian analyses were confined to trivial situations where the integral can be ignored. These situations involve the use of the so-called conjugate priors where the likelihood and the prior combine in a nice fashion so that the posterior can be determined without integration because the posterior and the prior have the same form. For example, a binomial likelihood coupled with a beta prior yield a beta posterior, a Poisson likelihood coupled with a gamma prior yield a gamma posterior. The advent of the computationally intensive methods known as Markov Chain Monte Carlo (MCMC) methods has made Bayesian analysis feasible for a wide class of models. This simulation-based approach greatly simplifies Bayesian analysis. The two most popular MCMC algorithms are the Gibbs sampler (e.g. Gelfand & Smith 1990) and the Metropolis-Hastings (M-H) algorithm (Hastings, 1970; Chib & Greenberg 1995)

For the purpose of our analysis, we first revisit M-H algorithm, which is an extension of the original Metropolis algorithm introduced by Metropolis et al. (1953). Like any other MCMC method, the principle of the M-H sampler is to generate (dependent) samples $\beta^{(1)}, \dots, \beta^{(N)}$ whose distribution converges to a target distribution $\pi(\beta)$. In Bayesian estimation of β , based on likelihood $p(y|\beta)$ and prior distribution $p(\beta)$, the target distribution is the posterior $\pi(\beta) = p(\beta|y)$. Starting from an arbitrary point in the parameter space, new states β^* are generated from a proposal (or instrumental) distribution $q(\beta^*|\beta^{(i)})$ giving the probability density of moving to state β^* from the current state $\beta^{(i)}$ of the chain at step i . Each proposal is either accepted (meaning that the chain moves to the new state) or rejected (i.e. the chain remains in the same state), depending on the value of a test ratio. A description of the M-H algorithm is given below.

1. Pick $\beta^{(1)}$ arbitrarily in the support of $\pi(\beta) = p(\beta|y)$
2. Draw β^* from $q(\beta^*|\beta^{(i)})$
3. Compute $r(\beta^{(i)}, \beta^*) = \pi(\beta^*) q(\beta^{(i)}|\beta^*) / \pi(\beta^{(i)}) q(\beta^*|\beta^{(i)})$

Generate $u \sim Uni(0,1)$

If $u < r(\beta^{(i)}, \beta^*)$ then set $\beta^{(i+1)} = \beta^*$

otherwise, set $\beta^{(i+1)} = \beta^{(i)}$

Repeat 2 -3 until "convergence"

The proposal density q can be anything, but a symmetric proposal, meaning that $q(\beta^{(i)}|\beta^*) = q(\beta^*|\beta^{(i)})$ yields the simpler form $r(\beta^{(i)}, \beta^*) = \pi(\beta^*) / \pi(\beta^{(i)})$ for the test ratio. A common symmetric choice is the random walk proposal which corresponds to the original Metropolis algorithm which is implemented here. Conditionally on the state, $\beta^{(k)}$, of the chain at step k , a random walk Metropolis sampler proposes a move to $\beta^* = \beta^{(k)} + \sigma \epsilon$ where $\epsilon \sim MVN(\mathbf{0}, \Sigma)$, and $\sigma > 0$ is a tuning constant usually selected in such a way that roughly 20%-40% of the proposals are accepted. The pre-convergence part of the MCMC called "burn-in" is usually discarded, and the remaining part might be thinned to say every j^{th} observation, for an integer $j > 1$ in order to reduce the correlation between consecutive draws. The thinning factor j is therefore the frequency with which successive parameter values are stored. After the burn-in and thinning of the MCMC, distributional summaries of the parameters of interest are produced from the stored samples.

3. Application to the bioassay data

We placed on the components of β independent and diffuse (with variance 1.0E+6) centered Gaussian priors. A random walk Metropolis was subsequently implemented, generating at step k proposals of the form $\beta^* = \beta^{(k)} + \sigma Az$,

where z is a random draw from the bivariate standard Gaussian, and A is the Cholesky factor of the ML asymptotic covariance matrix C i.e., means $C = A^T A$. Any symmetric and positive definite (SPD) matrix (all covariance matrices are indeed SPD) has a Cholesky decomposition (see e.g. Gentle, 1998 for more details).

We initially ran 5000 iterations of the MCMC for a couple of times in order to select a suitable tuning factor σ . The value $\sigma = 2$ was retained as it yielded a roughly 32% acceptance probability. After the tuning constant was set, we ran 5000 burn-in iterations of the MCMC followed by a further 20000 iterations, and thinned the MCMC to every 10th draw. A commented version of the R script implementing the sampler is provided in appendix B, where the variable *nacc* contains the number of accepted proposal and *acc_prob* contains the acceptance probability.

The MCMC reached the target distribution quite fast, and then proved to move freely over the parameter space. This good mixing is illustrated by the trace plots in Figure 1 which display 1000 posterior samples for each component of the parameter β .

The sensitivity of the results to the prior inputs was assessed by varying the range of the priors by order of magnitudes, while starting the MCMC from dispersed points in the parameter space, but the results remained broadly robust to these changes.

All our results are reported on the scale of the log dose. The posterior mean, $\bar{\beta}$ was (1.35, 12.37), which is relatively close to Gelman et al.'s approximate posterior (1.37, 11.93) obtained through simulations. The components of the posterior covariance matrix of $\hat{\beta}$ were estimated to be $\text{Var}(\hat{\beta}_1) = 1.12$,

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = 3.92, \text{ and } \text{Var}(\hat{\beta}_2) = 32.76.$$

Contour plots generally provide a better approach to displaying three-dimensional continuous data than the three-dimensional hidden-line plots. The un-normalized posterior was estimated at a 100 by 100 grid: $(\beta_1, \beta_2) \in [-5, 8] \times [-5, 35]$ and contour lines of the posterior surface were drawn at levels 0.922, 0.815, 0.701, 0.4, 0.1, 0.01, and 0.005 around the value at the mode. 1000 samples from the joint posterior of β_1 and β_2 were subsequently plotted on top of the contour plot as shown in Figure 2 (R code in appendix C). It can also be seen from this graph that the posterior mode is roughly equal to the ML estimate as expected under non-informative prior. However, the Bayesian approach allows for an effective handling of uncertainty through the posterior distribution. This is particularly important when it comes to predictions and risk assessment where a suitable accommodation of uncertainty is crucial.

As already pointed out, $\varphi = LD50$, as a function of the parameters, is defined by $\varphi = \varphi(\beta_1, \beta_2) = -\beta_1 / \beta_2, (\beta_2 \neq 0)$. A sample from the posterior distribution, $p(\varphi | y)$, of $LD50$ is straightforwardly obtained from the joint posterior of β , and an approximate $100 \times (1 - \alpha)\%$ credibility interval for φ is determined by the $\alpha/2$ and $(1 - \alpha/2)$ percentiles of the simulated $p(\varphi | y)$. For the bioassay data, the posterior mean and an approximate 95% credible interval for φ (on the log-scale) were respectively -0.101 and [-0.280, 0.091]. Figure 3 displays a histogram of the posterior distribution $p(\varphi | y)$.

4. Conclusions

In this paper we have discussed the Bayesian analysis of binary response regression using MCMC, more specifically the Metropolis sampler, for posterior simulation. We illustrated the methodology with real-world data from a bioassay experiment. The results were compared across different methods of posterior approximation and compared to their maximum likelihood counterparts. It is worth emphasizing that the bioassay dataset analyzed here was selected solely for illustration. There were only 5 replicates by level of dose, which resulted in low estimation precision as reflected by the large standard errors of the estimates.

In biological control, pest organisms may evolve over time with some characteristics that allow them to buffer the effect of say, pesticides or insecticides. Sequential replication over specific time-periods is necessary in order to track possible evolution of the dose-response relationship. This can be done by including random effects for different time periods. Combining link functions and exponential family variation with random effects leads to the cutting edge generalized linear mixed models (GLMMs) (e.g., Skrondal & Rabe-Hesketh 2004). GLMMs can easily be fitted with a Bayesian approach using MCMC.

References

- Casella, G., Berger, Roger L. (2002). *Statistical Inference*, 2nd edn. Duxbury Press, Belmont, CA.
- Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm, *The American Statistician* 49, 327-335.

- Clark, J. S. (2005). Why environmental scientists are becoming Bayesians *Ecology Letters*, Vol. 8: 2-14.
- Dobson, A. J. (2002). An introduction to generalized linear models, 2nd edn. Chapman & Hall/CRC, London
- Englehardt, J. (2004). Predictive Bayesian dose-response assessment for appraising absolute health risk from available information Human and Ecological Risk Assessment, The Association for Environmental Health and Sciences, 10(1), 69–78.
- Englehardt, J., and Swartout, J. (2004). Predictive population dose-response assessment for *Cryptosporidium parvum*: Infection end point. *Journal of Toxicology and Environmental Health Part A: Current Issues*, 67(8–10), 651–666.
- Englehardt, J., and Swartout, J. (2006). “Predictive Bayesian Microbial Dose-Response Assessment Based on Suggested Self-Organization in Primary Illness Response: *C. Parvum*,” *Risk Analysis*, vol. 26, no. 2, pp. 543-554.
- French, N. P., Kelly, L., Jones, R. & Clancy, D. (2002). Dose-response relationships for foot and mouth disease in cattle and sheep *Epidemiology and Infection* 128: 325-332 Cambridge University Press.
- Garthwaite, P. H., Joliffe, I. T. & Jones, B. (2002). Statistical inference 2nd Edition Oxford University Press, Oxford, UK.
- Gelfand, A., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities, *JASA* 85, 398-409.
- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2003). *Bayesian Data Analysis* Chapman & Hall/CRC.
- Gentle, J. E. (1998). *Numerical Linear Algebra for Applications in Statistics*. Springer-Verlag, Berlin.
- Gilks, W. R., Richardson, S. & Spiegelhalter, D. J., eds. (1996). *Markov Chain Monte Carlo in practice* Chapman and Hall, London, UK
- Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications *Biometrika*, 57(1):97-109.
- Lindsey, J. K. (1997). *Applying Generalized Linear Models* Springer-Verlag New York
- McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, New York.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N, Teller, A. H., & Teller, E. (1953). Equations of State Calculations by Fast Computing Machines *Journal of Chemical Physics* 21(6):1087-1092.
- Metz, C. E., Tokars, R. P., Kronman, H. B, and Griem, M. L. (1982). Maximum likelihood estimation of dose-response parameters for therapeutic operating characteristic (TOC) analysis of carcinoma of the nasopharynx. *Int J Radiat Oncol Biol Phys.* 8 (7):1185-92.
- Mutshinda, C.M., Antai, I., O'Hara R.B. (2008). A probabilistic approach to exposure risk assessment. *Stochastic Environmental Research & Risk Assessment* 22: 441-449
- Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.
- Spiegelhalter, D., Thomas, A., Best, N. & Lunn, D. (2003). WinBUGS version 1.4 User manual <http://www.mrc-bsu.cam.ac.uk/bug>
- Teunis, P., Chappell, C., & Okhuysen, P. (2002). *Cryptosporidium* dose response studies: Variation between hosts. *Risk Analysis*, 22(3), 475–485.
- Thomas, A., O'Hara, R. B., Ligges, U., and Sturtz, S. (2006). Making BUGS Open. *R News* 6: 12-17.

Table 1. The bioassay data reproduced (with permission) from Gelman et al. (2003).

x : Dose (log g/ml)	y : Number of Deaths	n : Number of Animals
-0.86	0	5
-0.30	1	5
-0.05	3	5
0.73	5	5

y_i denotes the number of animals with positive response at dose x_i out of n_i .

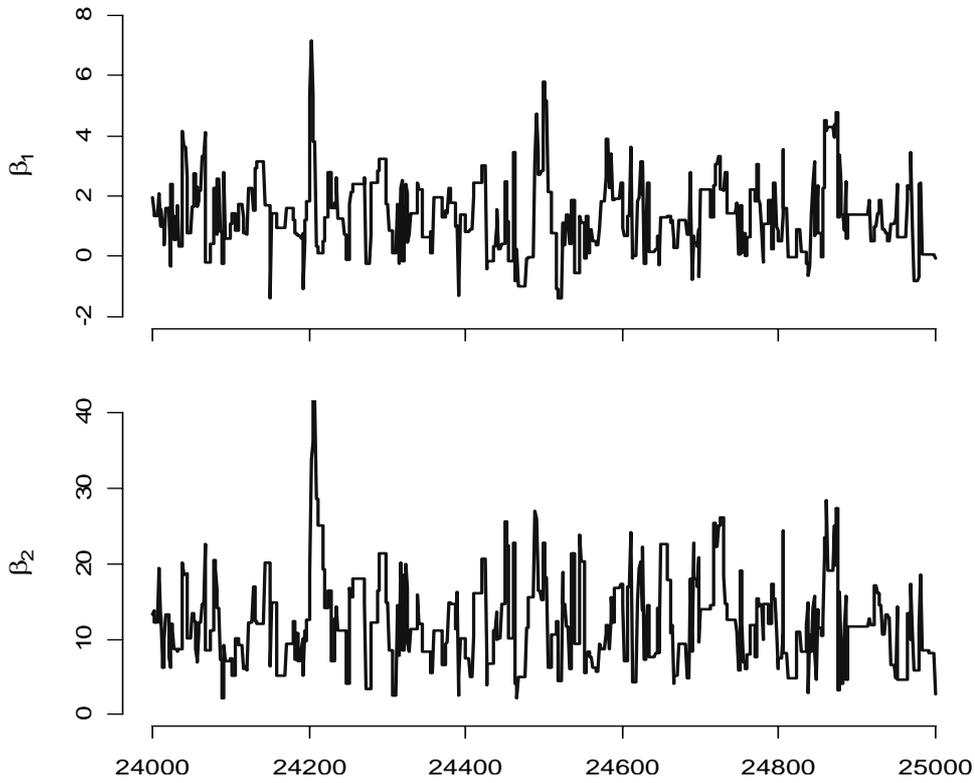


Figure 1. Trace plot of 1000 draws from the joint posterior of β_1 (top) and β_2 (bottom)

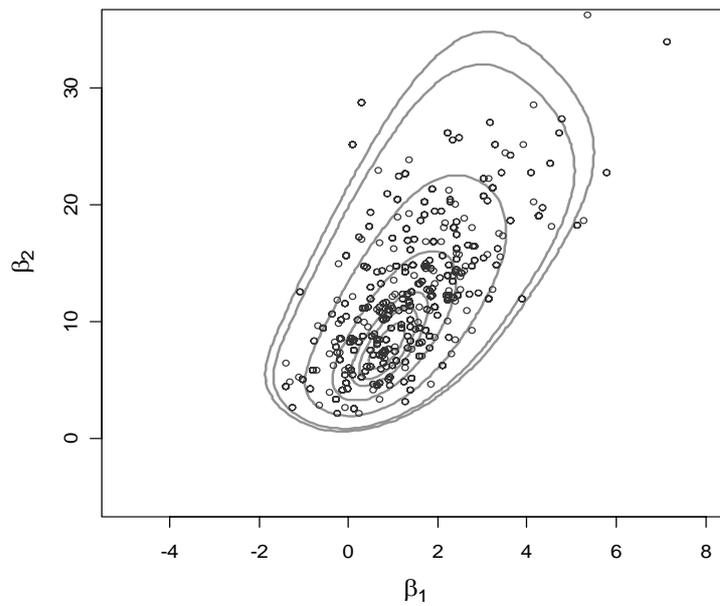


Figure 2. 1000 simulated samples from the joint the joint posterior of β_1 and β_2 plotted on top of a contour plot for the joint posterior density of the two parameters.

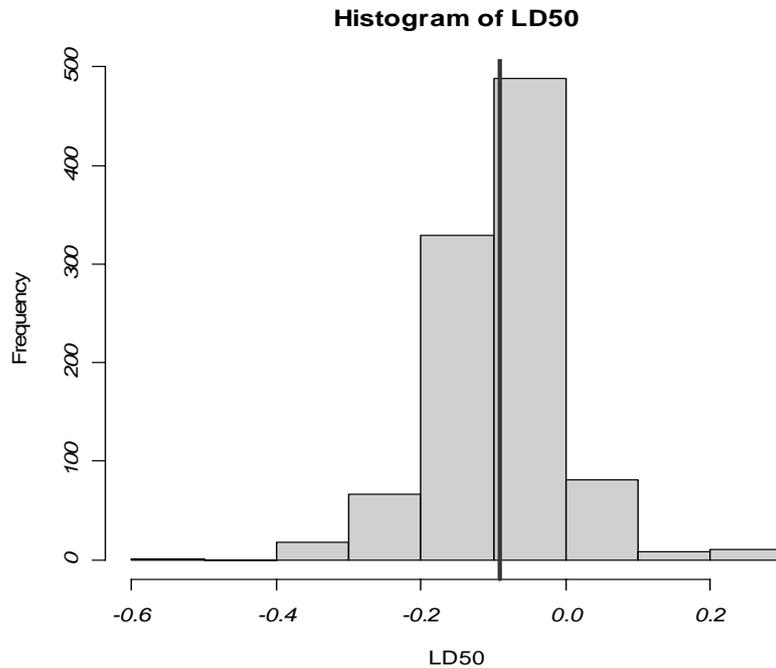


Figure 3. Histogram of simulated samples from the posterior of $LD50$. The solid vertical line is drawn at the posterior mean estimated to be -0.101 . A 95% credible interval for $LD50$ is $[-0.280, 0.091]$

Appendix A: R script for model fitting with the ML method

```
x<-c(-0.86, -0.3, -0.05, 0.73); n<-c(5, 5, 5, 5); y<-c(0, 1, 3, 5)
myfit<-glm(cbind(y, n-y)~x, family=binomial)
# Getting the covariance matrix of the parameter vector
C<-summary(myfit)$cov.unscaled
```

Appendix B: Annotated R script for the Metropolis sampler

```
###functions for loglik and (unnormalized) prior and posterior
loglik<-function(theta)
sum(y*(theta[1] + theta[2]*x - log(1 + exp(theta[1]+ theta[2]*x)))-(n-y)*log(1 + exp(theta[1] + theta[2]*x)))
logprior<-function(theta){
lp1<-dnorm(log=TRUE, theta[1], mean=0, sd=1000)
lp2<-dnorm(log=TRUE, theta[2], mean=0, sd=1000)
lp1+lp2}
## function for (unnormalized) log-posterior
logpost<-function(theta)
loglik(theta)+logprior(theta)
```

```
#### Data
x<- c(-0.86, -0.3, -0.05, 0.73); n<-c(5,5,5,5); y<-c(0,1,3,5)
## ML cov matrix, C, and its Cholesky factor A
C<-matrix(c(1.038,3.546,3.546, 23.744),nrow=2, byrow=T)
A<-t(chol(C))
### Setting the number B of MCMC iterations
B=25000
beta<-matrix(0,nrow=B, ncol=2)
beta[1,]<-c(0.5, 6)####initializing beta
sigma<-2.0 ## setting the tuning factor
nacc<-0 ##initializing number of accepted proposals
for (b in 2:B){
current<-beta[b-1,]
###Generating a proposal state
proposal<-current + sigma*A%%rnorm(2)
log_alpha<- logpost(proposal)-logpost(current)#log of the MH test ratio
alpha<-exp(log_alpha) ### the MH test ratio
u<-runif(1)
### accept or reject, that is the question
if(u<alpha){beta[b,]<-proposal; nacc=nacc+1}
  else{beta[b,]<-current}
}
## Acceptance rate
acc_prob<-nacc/B; print(acc_prob)
```

Appendix C: R script for displayed plots

```
## Plotting (1000) samples from the posterior (Figure 1)
par(mfrow=c(2,1))
par(oma=c(0,0,0,0)); par(mar=c(2,4,1,1))
t<-24000:25000
plot(t,beta[24000:25000,1],type="l", ylab=expression(beta[1]), axes=F,cex.lab=1.2, col="blue",lwd=2, font.lab=3,
xlab= " ", ylim=c(-2,8))
axis(2, lwd=1) ; axis(1, labels=NA)
par(mfg=c(2,1))
plot(t,beta[24000:25000,2],type="l",ylab=expression(beta[2]), axes=F, ylim=c(0,40), cex.lab=1.2, font.lab=3,
col="blue", lwd=2)
axis(2, lwd=1) ; axis(1, lwd=1)

### Contour plot (Figure 2)
par(mfcol=c(1,1))
### Setting a grid for contour plot
grid1 = seq(-5,8,len=100); grid2 = seq(-5, 35,len=100)
postg<-matrix(0,nrow=length(grid1),ncol=length(grid2))
```

```
for(i in seq(along=grid1)){
  for(j in seq(along=grid2)){
# Computing the (unnormalized posterior) over the grid
    postg[i,j]<-exp(logpost(c(grid1[i], grid2[j])))
  }
}
maxp<-max(postg)
lev<-c(0.922, 0.815, 0.701, 0.4, 0.1, 0.01, 0.005)*maxp
contour(grid1, grid2, postg, levels=lev, col=3, drawlabels=F, lwd=2, xlab=expression(beta[1]),ylab=expression(beta[2]),
cex.lab=1.3)
points(beta[24000:25000,1],beta[24000:25000,2], col=2)

## Histogram for LD50 (Figure 3)
LD50<- -beta[24000:25000,1]/ beta[24000:25000,2]
hist(LD50, bg="grey", col="lightblue", xlab="LD50",
font=3)
abline(v=mean(LD50),col=2,lwd=3)
-----
```