

**Bayeasian Logistic Regression with application to
Assessing Drug Safety Issues**

**A PROJECT
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA DULUTH
BY**

Xuran Yan

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN APPLIED AND COMPUTATIONAL
MATHEMATICS**

Kang James

July, 2014

© Xuran Yan 2014
ALL RIGHTS RESERVED

Acknowledgements

I would like take this opportunity to express my sincere gratitude to Professor Kang James. She is not only a great academic adviser, but also a great life mentor. I very appreciate her patience and support during the last two years. I would also like to thank Professor Barry James and Professor Xuan Li for serving as my committee member, who spent a lot of time reading my paper and gave a lot of useful advice to me. I would like to thank the whole Mathematics and Statistics department for giving me the opportunity to have such wonderful two years of study here.

*To my parents,
for supporting and sharing my dream.*

Abstract

When the outcome of a study only has two values, people usually apply logistic regression model to analyze the data. The estimation of the parameters in a logistic model involves the Newton-Raphson Method, which also generates the variance of the estimates at the same time. However, sometimes there is additional prior information we want to take into consideration, and then the Bayesian approach is to be employed. To get the posterior distribution of the parameters in the Bayesian approach, Gibbs Sampling, a special case of the MCMC algorithm is needed in the calculation. Applications of both approaches to drug safety assessment data will be given.

Contents

Acknowledgements	i
Abstract	iii
List of Tables	vi
1 Introduction	1
2 Logistic Regression	3
2.1 Introduction to logistic regression	3
2.2 Estimation of the parameters	5
2.2.1 Maximum likelihood estimation	5
2.3 Newton-Raphson Method	6
2.4 Odds Ratio	7
2.5 Confidence Interval for β_i	8
2.6 Pneumoconiosis Example	8
3 Bayesian Approach	9
3.1 Bayes' Theorem	9
3.2 Simple Example	10
3.3 Bayesian Logistic Regression	11
3.4 Monte Carlo Simulation	11
3.4.1 Metropolis-Hastings Algorithm	12
3.4.2 <i>Gibbs Sampling</i>	13
3.5 Pneumoconiosis Example	14

4	Assessing Drug Safety	16
4.1	Advantages of the Bayesian Approach	16
4.2	Data	16
4.3	Notation	19
4.4	Model	19
4.5	The Posterior Distribution	21
4.6	Results	21
5	Multivariate Bayesian Logistic Regression in Sparse Data Case	23
6	Conclusion and Future Work	26
7	Appendix	27

List of Tables

2.1	Table of pneumoconiosis data	8
4.1	Table of clinical data	18
4.2	Table of notations	19

Chapter 1

Introduction

Frequentist and Bayesian approaches are two classic methods, when one wants to learn the unknown quantities he is interested in by building models using the observable data. Frequentist method has been well developed for a long time, meanwhile, Bayesian statistics did not get well developed until the later 20th century, when the computers become powerful enough and easily accessible. We can calculate of posterior distribution of Bayesian method theoretically, however, the analytic computation was the difficult part to be done. Until the early 1990's, the discover of the method of drawing sample from the posterior distribution, which is know as MCMC algorithm, make the Bayesian approach widely applicable in practical problems. The Bayesian method has been used in diverse application areas.

In this paper, we focus on the problem of logistic regression. We study the theoretical foundation of both methods, then apply both methods to a drug safety assessment problem. Among statisticians, theoretically, there is not any conclusion about which approach is better, but for certain problems, we can find out which works better.

The first part of this paper covers the fundamental concepts of logistic regression, the type of data that requires the logistic model, estimation of the parameters, and statistical inference about the logistic model.

The second part covers the fundamental concepts of the Bayesian approach, posterior distribution, and the MCMC algorithms and Gibbs sampling method.

Meanwhile, a simple example of each method will be given in both parts.

The last part is the application of both methods to a drug safety assessment problem,

and the result will be discussed.

Chapter 2

Logistic Regression

2.1 Introduction to logistic regression

Regression analysis is a widely used method for investigating and modeling data. In the usual linear regression, we have

y is the response variable

x₁, x₂, x₃, ..., x_k are predictor variables,

The model is

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2),$$

where the β_j 's are the unknown parameters. Using the least square method, we can get the estimation of the parameters. However, in normal linear regression, there are two assumptions of the response variables namely: normality and equal variance.

Consider now, for example, that in the study of tumors, what people care about is whether the tumor is benign or malignant. In this case, the response variable y only has two possible outcomes, so it is a Bernoulli random variable with the following distribution:

y	Probability
1	$P(y = 1) = \pi$
0	$P(y = 0) = 1 - \pi$

Let $y = 1$ if the tumor is benign and $y = 0$ if the tumor is malignant. If we use the linear model:

$$y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon \quad \varepsilon \sim N(0, \sigma^2),$$

$$\text{where } \mathbf{x}' = [1, x_1, x_2, \dots, x_k], \boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \dots, \beta_k],$$

since

$$E(\varepsilon) = 0,$$

the expected value of y is

$$E(y) = 1(\pi) + 0(1 - \pi) = \mathbf{x}'\boldsymbol{\beta} = \pi.$$

Using the least square method, it is possible that the estimate will be outside the interval $[0,1]$. Moreover, for each observation, $var(y_i) = \pi_i(1 - \pi_i)$, which indicates unequal variance, unless $\pi_i = \pi, \forall i$. In this case, the logistic model is often used. This model assumes that $logit(\pi) = \mathbf{x}'\boldsymbol{\beta}$ instead of $\pi = \mathbf{x}'\boldsymbol{\beta}$. With this model, we see that

$$E(y) = \pi = \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}'\boldsymbol{\beta})} = \frac{1}{1 + \exp(-\mathbf{x}'\boldsymbol{\beta})}.$$

Let

$$\eta = \mathbf{x}'\boldsymbol{\beta}.$$

Then η is called the linear predictor, a linear combination of the unknown parameters. The link function shows the relationship between the linear predictor and the mean of the distribution function. Here it is the logit function:

$$\eta = \ln \frac{\pi}{1 - \pi},$$

and $\frac{\pi}{1 - \pi}$ is called odds.

In summary, logistic regression is used this way:

1. The response variable y_i 's are independent $B(\pi_i, 1), i = 1, 2, \dots, n$.
2. The assumption of the model:

$$\pi_i = \frac{\exp(\beta_0 + \sum x_{ij}\beta_j)}{1 + \exp(\beta_0 + \sum x_{ij}\beta_j)}, \text{ where } j = 1, 2, \dots, k.$$

x'_{ij} s are known constants.

3.

$$\eta_i = \text{logit } \pi_i = \sum_{j=1}^k x_{ij} \beta_j$$

is the link function.

2.2 Estimation of the parameters

2.2.1 Maximum likelihood estimation

Likelihood is the function of the parameters in a statistical model, given the data. It is the probability of getting the outcomes given the parameter. Suppose we have n observations; for each observation, the likelihood function for the Bernoulli distribution is[4]:

$$f_i(y_i) = (\pi_i)^{y_i} (1 - \pi_i)^{1-y_i}, \quad i = 1, 2, \dots, n.$$

Because the observations are independent, the likelihood function is

$$L(y_1, y_2, \dots, y_n, \boldsymbol{\beta}) = \prod_{i=1}^n f_i(y_i) = \prod_{i=1}^n (\pi_i)^{y_i} (1 - \pi_i)^{1-y_i}, \quad i = 1, 2, \dots, n.$$

Often, it is more convenient to use the log-likelihood, which is:

$$\ln L(y_1, y_2, \dots, y_n, \boldsymbol{\beta}) = \ln \prod_{i=1}^n f_i(y_i) = \sum_{i=1}^n [y_i \ln (\frac{\pi_i}{1 - \pi_i})] + \sum_{i=1}^n \ln (1 - \pi_i), \quad i = 1, 2, \dots, n.$$

Since

$$\pi_i = \frac{\exp(\mathbf{x}_i' \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i' \boldsymbol{\beta})},$$

we can get

$$1 - \pi_i = \frac{1}{1 + \exp(\mathbf{x}_i' \boldsymbol{\beta})}.$$

Then the likelihood and \log likelihood function can be written as [4]

$$L(y_1, y_2, \dots, y_n, \boldsymbol{\beta}) = \prod_{i=1}^n \left(\frac{\exp(\mathbf{x}_i' \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i' \boldsymbol{\beta})} \right)^{y_i} \left(\frac{1}{1 + \exp(\mathbf{x}_i' \boldsymbol{\beta})} \right)^{(1-y_i)}, \quad i = 1, 2, \dots, n,$$

and

$$\ln L(y_1, y_2, \dots, y_n, \boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{x}_i' \boldsymbol{\beta} y_i - \sum_{i=1}^n [1 + \exp(\mathbf{x}_i' \boldsymbol{\beta})].$$

The MLE is a solution of the equation

$$\frac{\partial \ln L(y_1, y_2, \dots, y_n, \boldsymbol{\beta})}{\partial \beta_s} = 0, \quad s = 1, 2, \dots, k.$$

In general, this equation is very complicated for the logistic model. We often use the *Newton – Raphson Method* to obtain a solution.

A very simple example is when $\eta_i = \text{logit}(\pi_i) = \beta_0 + \beta_1 x_i$.

Suppose

$$T_1 = \sum_i y_i, \quad T_2 = \sum_i x_i y_i.$$

Then

$$\begin{aligned} \ln L(y_1, y_2, \dots, y_n, \boldsymbol{\beta}) &= \beta_0 T_1 + \beta_1 T_2 - \sum_i (1 + \exp(\beta_0 + \beta_1 x_i)), \\ \frac{\partial \ln L(y_1, y_2, \dots, y_n, \boldsymbol{\beta})}{\partial \beta_0} &= T_1 - \sum_i \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}, \\ \frac{\partial \ln L(y_1, y_2, \dots, y_n, \boldsymbol{\beta})}{\partial \beta_1} &= T_2 - \sum_i \frac{\exp(\beta_0 + \beta_1 x_i) x_i}{1 + \exp(\beta_0 + \beta_1 x_i)}. \end{aligned}$$

We need to solve these two equations.

2.3 Newton-Raphson Method

For one variable case, $f(x) = 0$, we can solve the equation directly. Let

$$\begin{aligned} L(x) &= f(x_0) + (x - x_0) f'(x_0), \\ L_1(x) &= f(x_1) + (x - x_1) f'(x_1). \end{aligned}$$

Procedure:

1. Guess an initial solution x_0
2. Use x_1 as the second approximation, where x_1 is the solution of $L(x) = 0$

3. Use x_2 as the third approximation, where x_2 is the solution of $L_1(x) = 0$

4. Continue...Stop when x_i converges.

Since $L(x) = 0 = f(x_0) + (x - x_0)f'(x_0)$ x satisfies $x - x_0 = -\frac{f(x_0)}{f'(x_0)}$, so

$$\begin{aligned} x_1 &= x_0 - \frac{f(x_0)}{f'(x_0)}, \\ x_2 &= x_1 - \frac{f(x_1)}{f'(x_1)}, \\ &\vdots \\ x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)}. \end{aligned}$$

x_i will converge to the root of $L(x) = 0$. For k variable case, $f(x_1, x_2, \dots, x_k)$ we want to solve the equation:

$$L(x) = f(x_{01}, x_{02}, \dots, x_{0k}) + \sum_{l=1}^k (x_l - x_{0l}) \frac{\partial f}{\partial x_l} \Big|_{x_0}, \quad i = 1, 2, \dots, k.$$

So the calculation to get the MLE will involve the second derivatives of log-likelihood.

2.4 Odds Ratio

After we get the estimated parameters, the next step is to interpret them. Suppose

$$\hat{\eta}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i.$$

The difference of two $\hat{\eta}$'s one unit away from each other is:

$$\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) = \ln(odds_{x_i+1}) - \ln(odds_{x_i}) = \ln\left(\frac{odds_{x_i+1}}{odds_{x_i}}\right) = \hat{\beta}_1.$$

It is the log of the odds ratio, so we can get the odds ratio:

$$\widehat{O_R} = \frac{odds_{x_i+1}}{odds_{x_i}} = e^{\hat{\beta}_1}.$$

This is the estimated change of the probability of success, when there is a unit change in the predictor variable[4].

2.5 Confidence Interval for β_i

Using the theoretical result of the MLE, we can obtain the approximate confidence interval for β_i , $i = 0, 1, \dots, k$. Suppose $\hat{\beta}$ is the MLE of β . It is known that $\hat{\beta}$ is approximately normally distributed under regularity conditions. More precisely, $\hat{\beta} - \beta$ is approximately $N(0, i^{-1}(\beta))$. Where $i(\beta)$ is Fisher's information matrix:

$$i_{s,t}(\beta) = (-E \frac{\partial^2 \ln L(\beta)}{\partial \beta_s \partial \beta_t}) \quad s = 0, 1, \dots, k \quad t = 0, 1, \dots, k$$

An approximate $(1 - \alpha)100\%$ C.I. for β_s is

$$\hat{\beta}_s \pm Z_{\frac{\alpha}{2}} \sqrt{i^{-1}_{ss}(\hat{\beta})}$$

2.6 Pneumoconiosis Example

This example is from the journal *Biometrics*. It is a study about the relationship between the probability of coal miners experiencing Pneumoconiosis and the number of years they were exposed [4]. The data is given in Table2.1

Number of Years of Exposure	Number of Severe Pneumoconiosis Cases	Total Number of Miners	Proportion of Severe Cases
5.8	0	98	0
15.0	1	54	0.0185
21.5	3	43	0.0698
27.5	8	48	0.1667
33.2	9	51	0.1865
39.5	8	38	0.2105
46.5	10	28	0.3571
51.5	5	11	0.4545

Table 2.1: Table of pneumoconiosis data

Using the SAS program (code is in appendix), we can get the model as:

$$\hat{a} = \hat{\pi} = \frac{1}{1 + e^{4.7965 - 0.0935x}}.$$

Chapter 3

Bayesian Approach

3.1 Bayes' Theorem

To get to the Bayesian approach of logistic regression, first I will introduce Bayes' Theorem.

Theorem 1. (Bayes' Theorem) *For events A and B, given $P(B) \neq 0$.*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

The Bayesian Approach to logistic regression applies Bayes' Theorem to our model in this way. Suppose the density function of the parameters in our model is assigned, say $P(\beta)$, and the data distribution given the parameters, is known as $P(\mathbf{x}|\beta)$, which by definition is just the likelihood function of our parameter. Then the density of the parameters given the data will be

$$P(\beta|\mathbf{x}) = \frac{P(\beta)P(\mathbf{x}|\beta)}{P(\mathbf{x})}.$$

Here the distribution of our parameter, which will be chosen by us is called the *prior distribution*. The distribution of the parameters given the data, which we care about, is called the *posterior distribution*. The denominator in this formula can be calculated as

$$P(\mathbf{x}) = \int P(\mathbf{x}, \beta) d\beta = \int P(\beta) P(\mathbf{x}|\beta) d\beta,$$

where it is integrated over all possible value of the parameters. Since it does not depend on the parameter and the data is fixed, it is considered as a constant. Then we can get

$$P(\beta|\mathbf{x}) \propto P(\beta)P(\mathbf{x}|\beta).$$

Our task in the Bayesian approach is to find the parameters that maximize the probability $P(\beta|\mathbf{x})$, which is proportional to $\pi(\beta|\mathbf{x}) \equiv P(\beta)P(\mathbf{x}|\beta)$.

3.2 Simple Example

Here is a simple example about the Bayesian method. Suppose there is a kind of disease A that can be passed to the next generation and depends on a pair of genes, one a dominant gene, the other one a recessive gene. Suppose there is a woman, whose mother has one of each of these genes and is affected by this disease, but her father has both recessive genes and is not affected. The health state of the woman herself is unknown. Under this condition, when we think about the state of the woman, our prior distribution for β is: the probability that the woman is a carrier of this disease $P(\beta = 1) = 0.5$, and the probability that this women does not carry this disease is $P(\beta = 0) = 0.5$.

Then we get more information about the woman's two sons. Both of them are not affected by the disease. Suppose $y_i = 0, 1$ $i = 1, 2$ indicates if the son is affected or not. The likelihood of the unknown parameter is

$$P(y_1 = 0, y_2 = 0 | \beta = 1) = 0.5 \times 0.5 = 0.25$$

$$P(y_1 = 0, y_2 = 0 | \beta = 0) = 1 \times 1 = 1$$

The last step is to calculate the posterior distribution by combining the prior and likelihood using Bayes' rule: Let $y = (y_1, y_2)$, then the probability that the woman is a carrier of the the disease, given the data y , is:

$$\begin{aligned} P(\theta = 1|y) &= \frac{P(y|\theta = 1)P(\theta = 1)}{P(y|\theta = 1)P(\theta = 1) + P(y|\theta = 0)P(\theta = 0)} \\ &= \frac{(0.25)(0.5)}{(0.25)(0.5) + (1.0)(0.5)} = \frac{0.125}{0.625} = 0.20 \end{aligned}$$

As we can see, if the woman has both sons unaffected she is less likely to be a carrier of the disease[3].

3.3 Bayesian Logistic Regression

The likelihood function was mentioned in 2.2.1. Suppose we have a normal prior distribution, which is often used as a prior distribution for logistic regression, for all the parameters.

$$\beta_j \sim N(\mu_j, \sigma_j^2) \quad j = 1, 2, \dots, k$$

The posterior distribution is:

$$P(\beta|\mathbf{y}) = \prod_{i=1}^n (\pi_i)^{y_i} (1 - \pi_i)^{1-y_i} \times \prod_{j=0}^k \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{1}{2}\left(\frac{\beta_j - \mu_j}{\sigma_j}\right)^2\right].$$

Apparently, this expression does not have a closed form, and it is hard to get the marginal distribution of each coefficient by integration. For logistic regression, it is hard to get the exact numerical solution. In the statistics software, the most popular method used in estimate the parameters is the *Markov chain Monte Carlo* simulation, which would give us an approximate solution

3.4 Monte Carlo Simulation

For computational convenience, we can approximate the distribution of continuous variables by drawing samples from the distribution. When the samples are close enough so that no important properties are missing, we can approximate the target distribution properly. Here our target distribution is $P(\beta|\mathbf{x})$. The simple approximation is to use a bunch of equally spaced values $\beta_1, \beta_2, \dots, \beta_n$, and approximate the target distribution with density at β_i by the probability $P(\beta_i|\mathbf{x}) / \sum_{j=1}^n P(\beta_j|\mathbf{x})$

For example, the Markov chain method is one of the sampling methods, which draws samples sequentially to get the approximate distribution. Suppose our target distribution is $\pi(x)$. The next draw depends on the last value we have drawn, based on

a *transition distribution* $T(\beta^{t+1}|\beta^t)$, which is detailed balanced. Even though the samples are locally correlated, this sequence is still *ergodic*. By *ergodic*, it means the samples will go through every x in the proportion of $\pi(x)$ if we run the simulation long enough. Then we can compute the quantities of interest about the distribution with such an sample, and we will get the approximate distribution.

By detailed balance, it means the *transition distribution* satisfies

$$\pi(x_t)T(x_{t+1}|x_t) = \pi(x_{t+1})T(x_t|x_{t+1}),$$

,and detailed balance can guarantee the sequence is *ergodic*.

The reason can be simply explained by an single line prove as follow: Suppose we have drawn x_1 from our target distribution, the distribution of the next point x_2 we draw by using the transition distribution is:

$$\int \pi(x_1)T_t(x_2|x_1)dx_1 = \int \pi(x_2)T_t(x_1|x_2)dx_1 = \pi(x_2) \int T_t(x_1|x_2)dx_1 = \pi(x_2)$$

So the next step is to find such a *transition distribution* that satisfies the detailed balance[3].

3.4.1 Metropolis-Hastings Algorithm

We can use the *Metropolis–Hastings Algorithm* to draw the samples for *Markov chain* method, which would guarantee us to get an detailed balanced *transition distribution*. To implement this algorithm, the first step is to pick a “proposal distribution”, named $q(x_{n+1}|x_n)$ here. The proposal distribution can be any distribution as we want, but usually an normal distribution centered at x_n is used as the proposal distribution.

The algorithm is like this:

1. After we pick up a starting point x_1 , we draw a point x_{2c} from the proposal distribution around x_1 . x_{2c} will be the candidate of the next point x_2 we choose.
2. To decide if we accept x_{2c} as x_2 , we need to know the “acceptance probability”:

$$\alpha(x_1, x_{2c}) = \min\left\{1, \frac{\pi(x_{2c})q(x_1|x_{2c})}{\pi(x_1)q(x_{2c}|x_1)}\right\}$$

3. Choose x_2 to be x_{2c} with probability $\alpha(x_1, x_{2c})$, and to be x_1 with probability

$1 - \alpha(x_1, x_{2c})$. So the *transition probability* is

$$T_t(x_2|x_1) = q(x_2|x_1)\alpha(x_1, x_2) \quad (x_1 \neq x_2)$$

Here is the proof that using this algorithm, the *transition distribution* is always detailed balanced: First the "acceptance probability" is given as

$$\alpha(x_1, x_{2c}) = \min\left\{1, \frac{\pi(x_{2c})q(x_1|x_{2c})}{\pi(x_1)q(x_{2c}|x_1)}\right\}.$$

If we multiply both sides of the equation by the denominator, since the denominator is a positive constant, the *min* function still holds and it does not affect the equality. The equation becomes:

$$\pi(x_1)q(x_2|x_1)\alpha(x_1, x_2) = \min\{\pi(x_1)q(x_2|x_1), \pi(x_2)q(x_1|x_2)\}$$

The min function is symmetric, so if we exchange the position of the arguments, the equation becomes

$$\begin{aligned} \pi(x_1)q(x_2|x_1)\alpha(x_1, x_2) &= \min\{\pi(x_1)q(x_2|x_1), \pi(x_2)q(x_1|x_2)\} \\ &= \min\{\pi(x_2)q(x_1|x_2), \pi(x_1)q(x_2|x_1)\} \end{aligned}$$

By the same procedure, it is easy to prove:

$$\begin{aligned} \pi(x_1)q(x_2|x_1)\alpha(x_1, x_2) &= \min\{\pi(x_1)q(x_2|x_1), \pi(x_2)q(x_1|x_2)\} \\ &= \min\{\pi(x_2)q(x_1|x_2), \pi(x_1)q(x_2|x_1)\} \\ &= \pi(x_2)q(x_1|x_2)\alpha(x_2, x_1) \end{aligned}$$

Also we know that the *transition distribution* $T_t(x_2|x_1) = q(x_2|x_1)\alpha(x_1, x_2) \quad (x_1 \neq x_2)$. Thus $\pi(x_1)T_t(x_2|x_1) = \pi(x_2)T_t(x_1|x_2)$ which means it is *detailed balanced*. So, the sequence generated by the *Metropolis – Hastings Algorithm* is ergodic [3].

3.4.2 Gibbs Sampling

Gibbs Sampling is a special case of the *Metropolis – Hastings Algorithm* that works for the logistic regression model, which iterates one component of the parameter vector each time.

Definition 3.1. *Full conditional distribution is the normalized distribution obtained by sampling along one coordinate direction.*

It is written as $\pi(x_i|\mathbf{x}^-)$, where \mathbf{x}^- is the predictor vector with all component fixed except x_i .

Theorem 3.1. *A multivariate distribution is uniquely determined by all of its full conditional distributions.*

In the *Gibbs Sampling Algorithm*, we update one direction each time. So the “acceptance probability” for one direction is :

$$\alpha(x_{i,1}, x_{i,2c}|\mathbf{x}^-) = \min\left\{1, \frac{\pi(x_{i,2c}|\mathbf{x}^-)q(x_{i,1}|x_{i,2c}, \mathbf{x}^-)}{\pi(x_{i,1}|\mathbf{x}^-)q(x_{i,2c}|x_{i,1}, \mathbf{x}^-)}\right\}$$

Here the “proposal distribution” is set to be $q(x_2|x_1, \mathbf{x}^-) = \pi(x_2|\mathbf{x}^-)$. It is clear that by setting the proposal distribution like this, we can make the second argument in the *min* function be canceled to be 1. Therefore, no matter what we just accept the candidate. However we have to normalize the proposal distribution, so $\int \pi(x_i|\mathbf{x}^-)dx_i = 1$ along each direction every time.

Tens of thousands of iterations are generated for the posterior simulation. Nowadays there are many options of statistical software;, for example, SAS and R can do the iterations for us[3].

3.5 Pneumoconiosis Example

Here we get the Bayesian approach to the pneumoconiosis example. We can get the model as:

The SAS System

13:36 Tu

The GENMOD Procedure

Bayesian Analysis

Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
Intercept	5000	-4.8867	0.5827	-5.2666	-4.8630	-4.4807
year	5000	0.0955	0.0158	0.0845	0.0952	0.1059

$$\hat{y} = \hat{\pi} = \frac{1}{1 + e^{4.8867 - 0.0955x}}.$$

Here the prior we used was $N(0, 10^6)$, which is very non-informative. So we got the almost same result as in non-Bayesian approach. Which are:

$$\hat{y} = \hat{\pi} = \frac{1}{1 + e^{4.7965 - 0.0935x}},$$

$$\hat{y} = \hat{\pi} = \frac{1}{1 + e^{4.8867 - 0.0955x}}.$$

Chapter 4

Assessing Drug Safety

4.1 Advantages of the Bayesian Approach

Although the Bayesian approach requires more complex calculations, there are some advantages of that make it popular in applications. One important advantage is that, in practice, when we have multiple parameters, many of them are related in some way. For example, suppose θ_j is the survival probability of the patient in hospital j under a cardiac treatment. When we want to study the effectiveness of this cardiac treatment among a sample of hospitals, it is reasonable to regard the survival probability θ_j 's as related or connected to each other. We can show this connection by setting a prior distribution which makes the θ_j 's as a sample from a common population distribution. Since θ_j 's can not be observed directly, we can use a hierarchical model to connect them to the observable outcomes. The explicit explanation of the hierarchical model will be given in the next several section[3].

4.2 Data

The data we used in this paper is from Mehrotra and Heyse (2001) . 296 patients were randomly divided into two groups, the treatment group and control group. In the treatment group, the patients were toddlers who were given quadrivalent MMRV(measles, mumps, rubella, varicella) vaccine on day 0. On the other hand, the patients in the control group were toddlers only given MMR vaccine on day 0, then given V vaccine

on day 42. The data was collected between days 0-42 of the treatment group and days 42-84 of the control group. Our purpose is to study the clinical experimental data and flag if any adverse event (AE) is caused by the varicella vaccine.

There are 40 types of AEs we are curious about, like fever, cough and wheezing, for example, that are considered in the study. All the AE can be categorized into 8 body systems by biologists based on their biological ground. It is natural to consider the adverse drug effects that occur in the same body system as related. While the frequentist method focuses on type-I error, the Bayesian method focuses on the probability that the drug causes the AE's, taking consideration of all the information given. The Bayesian method allows different adverse drug effects to borrow strength from each other. That is one reason the Bayesian method is widely used in assessing drug safety issues. Here we give more details of this example. This example was studied in Berry & Berry(2004).

The collected data is given in Table4.1 [1]

			Treatment $N_T = 148$		Control $N_C = 132$		Fisher's exact p-value
b	j	Type of AE A_{bj}	Y_{bj}	Rate	X_{bj}	Rate	
1	1	Asthenia/fatigue	57	0.385	40	0.303	0.167
1	2	Fever	34	0.230	26	0.197	0.561
1	3	Infection,fungal	2	0.012	0	0.000	0.500
1	4	Infection,viral	3	0.020	1	0.008	0.625
1	5	Malaise	27	0.182	20	0.152	0.525
3	1	Anorexia	7	0.047	2	0.015	0.179
3	2	Candidiasis,oral	2	0.014	0	0.000	0.500
3	3	Constipation	2	0.014	0	0.000	0.500
3	4	Diarrhea	24	0.162	10	0.076	0.029
3	5	Gastroenteritis	3	0.020	1	0.008	0.625
3	6	Nausea	2	0.014	7	0.053	0.089
3	7	Vomiting	19	0.128	19	0.144	0.730
5	1	Lymphadenopathy	3	0.020	2	0.015	1.000

6	1	Dehydration	0	0.000	2	0.015	0.221
8	1	Crying	2	0.014	0	0.000	0.500
8	2	Insomnia	2	0.014	2	0.015	0.100
8	3	Irritability	75	0.507	43	0.326	0.003
9	1	Bronchitis	4	0.027	1	0.008	0.375
9	2	Congestion,nasal	4	0.027	2	0.015	0.375
9	3	Congestion.respiratory	1	0.007	2	0.015	0.603
9	4	Cough	13	0.088	8	0.061	0.497
9	5	Infection, upper respi- ratory	28	0.189	20	0.152	0.431
9	6	Laryngotracheobronchitis	2	0.014	1	0.008	1.000
9	7	Pharyngitis	13	0.088	8	0.061	0.497
9	8	Rhinorrhea	15	0.101	14	0.106	1.000
9	9	Sinusitis	3	0.020	1	0.008	0.625
9	10	Tonsillitis	2	0.014	1	0.008	1.000
9	11	Wheezing	3	0.020	1	0.008	0.625
10	1	Bite/Sting	4	0.027	0	0.000	0.125
10	2	Eczema	2	0.014	0	0.000	0.500
10	3	Pruritis	2	0.014	1	0.008	1.000
10	4	Rash	13	0.088	3	0.023	0.021
10	5	Rash,diaper	6	0.041	2	0.015	0.288
10	6	Rash,measles/rubella- like	8	0.054	1	0.008	0.039
10	7	Rash,varicella-like	4	0.027	2	0.015	0.687
10	8	Urticaria	0	0.000	2	0.015	0.221
10	9	Viral exanthema	1	0.007	2	0.015	0.603
11	1	Conjunctivitis	0	0.000	2	0.015	0.221
11	2	Otitis media	18	0.122	14	0.106	0.711
11	3	Otorrhea	2	0.014	1	0.008	1.000

Table 4.1: Table of clinical data

4.3 Notation

The notations we use here are given in Table4.2

B	Number of body systems	
b	Index for different body systems	$b = 1, 2, \dots, B$
k_b	Number of AEs in system b	
A_{bj}	The j th AE in system b	$j = 1, \dots, k_b$
N_C	Number of patients in control group	
N_T	Number of patients in treatment group	
X_{bj}	Number of patients experienced A_{bj} in control group	
Y_{bj}	Number of patients experienced A_{bj} in treatment group	
c_{bj}	Probability of experiencing A_{bj} in control group	
t_{bj}	Probability of experience A_{bj} in treatment group	

Table 4.2: Table of notations

4.4 Model

In this project , a tree-level Hierarchical model is used. The first level is different types of AEs; the second level is the different body systems containing these AEs; the third level is the body system as a whole. The logistic transformation is

$$\gamma_{bj} = \ln\left(\frac{c_{bj}}{1 - c_{bj}}\right) \quad \theta_{bj} = \ln\left(\frac{t_{bj}}{1 - t_{bj}}\right) - \gamma_{bj}.$$

Here the θ'_{bj} s are the log-odds ratios. If $c_{bj} = t_{bj}$, which means $\theta_{bj} = 0$, the patient in the treatment group and the control have the same probability to experience A_{bj} . Three levels of the model are:

1. The first level parameter distribution for γ is:

$$\gamma_{bj} \sim N(\mu_{\gamma b}, \sigma_{\gamma}^2) \quad b = 1, \dots, B \quad j = 1, \dots, k_b$$

and for θ is a mixture prior distribution:

$$\theta_{bj} \sim \pi_b I_{[0]} + (1 - \pi_b) N(\mu_{\theta b}, \tau_{\theta, b}^2)$$

2. Then for the hyperparameters we assign a prior distribution, the second level is:

$$\mu_{\gamma b} \sim N(\mu_{\gamma 0}, \tau_{\gamma 0}^2) \quad \text{and} \quad \sigma_{\gamma}^2 \sim IG(\alpha_{\sigma\gamma}, \beta_{\sigma\gamma})$$

and here π_b is the probability $\theta_{bj} = 0$. It is the same for all the AE in the same body system b .

The prior for π_b is:

$$\pi_b \sim Beta(\alpha_{\pi}, \beta_{\pi})$$

The prior for the hyperparameters in the normal part are:

$$\mu_{\theta b} \sim N(\mu_{\theta 0}, \tau_{\theta 0}^2) \quad \sigma_{\theta b}^2 \sim IG(\alpha_{\theta}, \beta_{\theta}).$$

3. The third level prior of the hyper-hyperparameters is:

$$\mu_{\gamma 0} \sim N(\mu_{\gamma 00}, \tau_{\gamma 00}^2) \quad \tau_{\gamma 0}^2 \sim IG(\alpha_{\tau\gamma}, \beta_{\tau\gamma})$$

To avoid the posterior π too concentrated at 1 or 0, α and β should be greater than 1. Their priors are the left truncated exponential distribution:

$$\alpha_{\pi} \sim \frac{\lambda_{\alpha} \exp(-\alpha \lambda_{\alpha})}{\exp(-\lambda_{\alpha})} I_{[\alpha > 1]} \quad \beta_{\pi} \sim \frac{\lambda_{\beta} \exp(-\beta \lambda_{\beta})}{\exp(-\lambda_{\beta})} I_{[\beta > 1]}.$$

If $\lambda_{\alpha} = \lambda_{\beta}$, α and β would be symmetric, and the priori probability that $\theta_{bj} = 0$ is 0.5. For the normal part:

$$\mu_{\theta 0} \sim N(\mu_{\theta 00}, \tau_{\theta 00}^2) \quad \tau_{\theta 0}^2 \sim IG(\alpha_{\theta 0}, \beta_{\theta 0}).$$

4. All the hyperparameters $\mu_{\gamma 00}, \tau_{\gamma 00}, \alpha_{\tau\gamma}, \beta_{\tau\gamma}, \alpha_{\sigma\gamma}, \beta_{\sigma\gamma}$ are fixed constants. We assign the value for the priors as: $\mu_{\theta 00} = 0, \tau_{\theta 00}^2 = 10, \alpha_{\theta} = 3, \beta_{\theta} = 1, \alpha_{\theta 0} = 3, \beta_{\theta 0} = 1, \alpha_{\tau\gamma} = \alpha_{\sigma\gamma} = 3, \beta_{\tau\gamma} = \beta_{\sigma\gamma} = 1, \lambda_{\alpha} = \lambda_{\beta} = 1$ [1].

4.5 The Posterior Distribution

After we set up the model, the next step is to use MCMC to simulate the posterior distribution. The posterior distribution is:

$$\begin{aligned}
& [\tilde{\theta}, \tilde{\gamma}, \tilde{\pi}, \tilde{\sigma}_\theta, \tilde{\mu}_\theta, \sigma_\gamma, \tilde{\mu}_\gamma, \alpha_\pi, \beta_\pi, \tau_{\gamma 0}, \mu_{\gamma 0}, \tau_{\theta 0}, \mu_{\theta 0} | \tilde{X}, \tilde{Y}] \\
& \propto \prod_{b=1}^B \prod_{j=1}^{k_b} \left[\frac{\exp(\theta_{bj} + \gamma_{bj})^{Y_{bj}}}{\{1 + \exp(\theta_{bj} + \gamma_{bj})\}^{N_T}} \frac{\exp(\gamma_{bj})^{X_{bj}}}{\{1 + \exp(\gamma_{bj})\}^{N_C}} \right] \\
& \times \prod_{b=1}^B \prod_{j=1}^{k_b} [\pi_b I_{[\theta_{bj}=0]} + (1 - \pi_b) \left(\frac{1}{\sigma_{\theta b}^2} \right) \times \exp\left\{-\frac{1}{2\sigma_{\theta b}^2} (\theta_{bj} - \mu_{\theta b})^2\right\}] \\
& \times \prod_{b=1}^B \left\{ \frac{\Gamma(\alpha_\pi + \beta_\pi)}{\Gamma(\alpha_\pi)\Gamma(\beta_\pi)} (\pi_b)^{\alpha_\pi-1} (1 - \pi_b)^{\beta_\pi-1} \times (\sigma_{\theta b}^2)^{-(\alpha_\theta+1)} \exp\left(-\frac{1}{\beta_\theta \sigma_{\theta b}^2}\right) \right\} \\
& \times [(\sigma_\gamma)^{-\sum_{b=1}^B k_b} \exp\left\{-\frac{1}{\sigma_\gamma^2} \sum_{b=1}^B \sum_{j=1}^{k_b} (\gamma_{bj} - \mu_{\gamma b})^2\right\}] \\
& \times [(\tau_{\gamma 0})^{-B} \exp\left\{-\frac{1}{2\tau_{\gamma 0}^2} \sum_{b=1}^B (\mu_{\gamma b} - \mu_{\gamma 0})^2\right\}] \\
& \times [(\tau_{\theta 0})^{-B} \exp\left\{-\frac{1}{2\tau_{\theta 0}^2} \sum_{b=1}^B (\mu_{\theta b} - \mu_{\theta 0})^2\right\}] \\
& \times \{ \exp(\alpha_\pi \lambda_\alpha) I_{[\alpha_\pi > 1]} \exp(-\beta_\pi \lambda_\beta) I_{[\beta_\pi > 1]} \} (\sigma_\gamma^2)^{\alpha_{\sigma\gamma}+1} \\
& \times \exp\left(-\frac{1}{\beta_{\sigma\gamma} \sigma_\gamma^2}\right) (\tau_{\sigma 0}^2)^{-(\alpha_{\tau\gamma}+1)} \exp\left(-\frac{1}{\beta_{\tau\gamma} \tau_{\gamma 0}^2}\right) \\
& \times (\tau_{\theta 0}^2)^{-(\alpha_{\tau\theta}+1)} \exp\left(-\frac{1}{\beta_{\tau\theta} \tau_{\theta 0}^2}\right) \\
& \times \exp\left\{-\frac{1}{2\tau_{\gamma 00}^2} (\mu_{\gamma 0} - \mu_{\gamma 00})^2\right\} \exp\left\{-\frac{1}{2\tau_{\theta 00}^2} (\mu_{\theta 0} - \mu_{\theta 00})^2\right\}
\end{aligned}$$

And we will use SAS software to help us solve it. The program codes given by Dr. Zink (2013) are included in the appendix[1].

4.6 Results

In the frequentist approach, only four AE's are to be significant, with small Fisher's exact test p-value. They are Diarrhea, Irritability, Rash and Rash,measles/rubella-like,

with p-values 0.029, 0.003, 0.021, and 0.039, respectively.

In the Bayesian approach, what we get is θ , the rate of a AE is more likely to occur in the treatment group than in the control group ($\theta > 0$), or not ($\theta = 0$). The result shows us that only the probability $\theta > 0$ for Irritability has a value greater than 0.5, which is 0.780. The probability $\theta > 0$ for Diarrhea, Rash and Rash,measles/rubella-like are all much smaller, which are 0.231, 0.190, 0.126, respectively. It was because diarrhea was in body system 3 with 6 other AE's besides itself, Rash and Rash,measles/rubella-like was in body system 10 with 7 other AE's besides themselves. The Bayesian setting take the body system into consideration, which makes the results of AEs in the same body system shrink towards each other. Since all the other AE's in system 3 and 10 are not significant, the probability $\theta > 0$ of these three AE's just shrink towards 0. In fact, the probability $\theta > 0$ for Irritability also shrinks, however, there are only 2 other AE's in body system 8, so it is still relatively big.

Chapter 5

Multivariate Bayesian Logistic Regression in Sparse Data Case

The paper of Dr. DuMouchel (2012) talked about another situation where the Bayesian approach is necessary. When it comes to clinical study designed for efficacy more than safety investigation, the adverse drug events data(safety issue) are often sparse. The response variables are dichotomous, with 1 meaning the patient experienced this issue, and 0 meaning he did not. There will be a lot of 0's in our collected data. In this case, there are two methods of dealing with this kind of problem. One is by analyzing the safety issues separately; however, for some event, the small count may lead to non significant result. The other one is to treat all the events as one event, and define it as present when any of these issues appear; however, the significant issues may be submerged into other issue, and then we could know little about the original problem. Using Multivariate Bayesian logistic regression, we can make a compromise between these two methods.

The example in Dumouchel's paper contains data from a pool of 8 studies; 10 safety issues that are medically related were analyzed. The adverse events are like Anuria, Dry mouth, Hyperkalaemia, etc. The predictor variables are also dichotomous or categorical, like age, gender and medical history. The main predictor in our problem is the study arm, which has two values "Treatment" and "Comparator".

The subjects(patients) with the same location in the predictor categories were grouped

together. Suppose there were m such groups, n_i subjects in group i , and N_{ik} denotes the number of subjects in group i ($i = 1, 2, \dots, m$) who experienced issue k .

Different issues were assigned a single regression model. Suppose there were J predictors besides the treatment arm, and g_j is the number of categories in predictor J . Let $G = \sum g_j$. For each group, the probability of experiencing issue k is P_{ik} , the model is:

$$P_{ik} = 1/[1 + \exp(-Z_{ik})],$$

$$Z_{ik} = \alpha_0 k + \sum_{1 \leq g \leq G} X_{ig} \alpha_{gk} + T_i (\beta_{0k} + \sum_{1 \leq g \leq G} X_{ig} \beta_{gk}),$$

where X_{ig} is the dummy variable for the J covariates and T_i is the dummy variable that indicates the treatment status of the group. In this model, our priors for the parameters are

$$\begin{aligned} \alpha_{gk} | A_g &\sim N(A_g, \sigma_A^2), & k = 1, \dots, K; g = 1, \dots, G, \\ \beta_{0k} | B_0 &\sim N(B_0, \sigma_0^2), & k = 1, \dots, K \\ \beta_{gk} | B_g &\sim N(B_g, \sigma_B^2), & k = 1, \dots, K; g = 1, \dots, G, \\ B_g &\sim N(0, \tau^2), & g = 1, \dots, G, \end{aligned}$$

with constraints $\sum_g \alpha_{gk} = 0, \sum_g \beta_{gk} = 0$. The prior distribution of α_0, A_g, B_0 is assumed to be uniform within $(-\infty, +\infty)$. If the four standard deviations ($\sigma_A, \sigma_0, \sigma_B, \tau$) are fixed, the posterior joint distribution of our parameters will be

$$\begin{aligned} L &= p(A_1, \dots, A_G, B_0, \dots, B_G, \{\alpha\}\{\beta\}\{N_{ik}\}) \\ &= p(\{N_{ik}\} | \{A\}\{B\}\{\alpha\}\{\beta\}) \cdot p(A_1, \dots, A_G, B_0, \dots, B_G, \{\alpha\}\{\beta\}) \\ &= p(A_1, \dots, A_G, B_0, \dots, B_G) \cdot \prod_k (\alpha_{0k}, \dots, \alpha_{Gk}, \beta_{0k}, \dots, \beta_{Gk} | \{A\}\{B\}) \cdot p(\{N_{ik}\} | \{A\}\{B\}\{\alpha\}\{\beta\}) \\ &= C \cdot \prod_{ik} (P_{ik})^{N_{ik}} (1 - P_{ik})^{n_i - N_{ik}} \cdot \prod_{g>0, k} \frac{1}{\sigma_A} \exp\left(-\frac{(\alpha_{gk} - A_g)^2}{2\sigma_A^2}\right) \cdot \prod_{g>0, k} \frac{1}{\sigma_B} \exp\left(-\frac{(\beta_{gk} - B_g)^2}{2\sigma_B^2}\right) \\ &\quad \cdot \prod_k \frac{1}{\sigma_0} \exp\left(-\frac{(\beta_{0k} - B_0)^2}{2\sigma_0^2}\right) \cdot \prod_{g>0} \frac{1}{\tau} \exp\left(-\frac{B_g^2}{2\tau^2}\right). \end{aligned}$$

The log posterior joint distribution, up to a constant, would be

$$\begin{aligned}
\ln L = & \sum_i \sum_k [N_{ik} \ln(P_{ik}) + (n_i - N_{ik}) \ln(1 - P_{ik})] \\
& - [\sum_{g>0} \sum_k \frac{(\alpha_{gk} - A_g)^2}{\sigma_A^2} + (G - J) K \ln(\sigma_A^2)] \\
& - [\sum_{g>0} \sum_k \frac{(\beta_{gk} - B_g)^2}{\sigma_B^2} + (G - J) K \ln(\sigma_B^2)] \\
& - [\frac{\sum (\beta_{0k} - B)^2}{\sigma_0^2} + K \ln(\sigma_0^2) - [\frac{\sum_{g>0} B_g^2}{\tau^2} + (G - J) \ln(\tau^2)]
\end{aligned}$$

Using this hierarchical model, applying the Newton-Raphson method we can get the estimates of the parameters which maximize the posterior. The choice of the standard deviations $(\sigma_A, \sigma_0, \sigma_B, \tau)$ involves the response surface method, which will not be discussed here [2].

Chapter 6

Conclusion and Future Work

The Bayesian method has some advantages, when we are dealing with assessing drug safety data.

1. It can help us to get a result when standard logistic regression does not work.
2. It can take consideration of all the information we have.

However, what prior information should be taken into consideration depends on the choices of the researcher. So one should be very careful when it comes to choose the prior parameters. In our example in Chapter 4, the priors were chosen to be very vague. Different priors may cause a big change in the model. The assignment of each AE in body system makes a big difference. Future work could be to try to assign Irritability into a body system with more AE's, and see what the result would be.

Chapter 7

Appendix

1. Codes for Section 2.6[4]:

```
data rr;
input year cases total portion;
datalines;
5.8 0 98 0
15.0 1 54 0.0185
21.5 3 43 0.0698
27.5 8 48 0.1667
33.5 9 51 0.1765
39.5 8 38 0.2105
46.0 10 28 0.3571
51.5 5 11 0.4545
;
run;
proc genmod data=rr;
model cases/total = year /dist = bin
link = logit type1 type3
lrci;
run;
2. Codes for Chapter 4[6]:
libname out "||Your Directory Name Here||";
```

```

data ae;
format term $28.;
input e s term $ N_T N_C ;
datalines;
1 1 Astenia/fatigue 57 40
2 1 Fever 34 26
3 1 Infection-fungal 2 0
4 1 Infection-viral 3 1
5 1 Malaise 27 20
6 2 Anorexia 7 2
7 2 Cendisiasis-oral 2 0
8 2 Constipation 2 0
9 2 Diarrhea 24 10
10 2 Gastroenteritis 3 1
11 2 Nausea 2 7
12 2 Vomiting 19 19
13 3 Lymphadenopathy 3 2
14 4 Dehydration 0 2
15 5 Crying 2 0
16 5 Insomnia 2 2
17 5 Irritability 75 43
18 6 Bronchitis 4 1
19 6 Congestion-nasal 4 2
20 6 Congestion-respiratory 1 2
21 6 Cough 13 8
22 6 Infection-upper-respiratory 28 20
23 6 Laryngotracheobronchitis 2 1
24 6 Pharyngitis 13 8
25 6 Rhinorrhea 15 14
26 6 Sinusitis 3 1
27 6 Tonsillitis 2 1
28 6 Wheezing 3 1

```

```

29 7 Bite/sting 4 0
30 7 Eczema 2 0
31 7 Pruritis 2 1
32 7 Rash 13 3
33 7 Rash-diaper 6 2
34 7 Rash-measles/rubella-like 8 1
35 7 Rash-varicella-like 4 2
36 7 Urticaria 0 2
37 7 Viral-exanthema 1 2
38 8 Conjunctivitis 0 2
39 8 Otitis-media 18 14
40 8 Otorrhea 2 1
;
run;
data ae(drop = N_T N_C);
set ae;
count = N_T; trt = 1; ntc = 148; output;
count = N_C; trt = 0; ntc = 132; output;
run;
%macro MCMC(suffix = , seed = );
%let mu_G00 = 0; %let mu_T00 = 0;
%let tau2_G00 = 10; %let tau2_T00 = 10;
%let A_G = 3; %let A_T = 3;
%let A_G00 = 3; %let A_T00 = 3;
%let B_G = 1; %let B_T = 1;
%let B_G00 = 1; %let B_T00 = 1;
%let L_A = 0.1; %let L_B = 0.1;
proc mcmc data = ae seed = &seed nmc = 20000 dic maxtune = 500
monitor = (_parms_ T_1-T_40) outpost = outpost&suffix nbi = 2000;
*** Treatment Effects for the 40 Adverse Events ***;
array T_[40] T_1 - T_40;
parms mu_G0;

```

```

parms tau2_G0;
parms mu_T0;
parms tau2_T0;
parms A_P;
parms B_P;
*** Hyperpriors for Gammas ***;
prior mu_G0  normal(&mu_G00, prec = %sysevalf(1/&tau2_G00));
prior tau2_G0  gamma(&A_G00, iscale = &B_G00);
*** Hyperpriors for Thetas ***;
prior mu_T0  normal(&mu_T00, prec = %sysevalf(1/&tau2_T00));
prior tau2_T0  gamma(&A_T00, iscale = &B_T00);
*** Hyperpriors for Pis ***;
prior A_P  expon(iscale = &L_A, lower = 1);
prior B_P  expon(iscale = &L_B, lower = 1);
random mu_Ga  normal(mu_G0, prec = tau2_G0) subject = s monitor = (mu_Ga);
random tau_Ga  gamma(&A_G, iscale = &B_G) subject = s monitor = (tau_Ga);
random G  normal(mu_Ga, prec = tau_Ga) subject = e monitor = (G);
random mu-Ta  normal(mu_T0, prec = tau2_T0) subject = s monitor = (mu-Ta);
random tau-Ta  gamma(&A_T, iscale = &B_T) subject = s monitor = (tau-Ta);
random T1  normal(mu-Ta, prec = tau-Ta) subject = e monitor = (T1);
random P  beta(A_P, B_P) subject = s monitor = (P);
random Bi  binary(P) subject = e monitor = (Bi);
T_[e] = (1 - Bi) * T1;
lp = G + T_[e] * trt;
ptc = logistic(lp);
model count  binomial(n = ntc, p = ptc);
run;
%mend MCMC;
*** Generate 3 Independent Chains ***;
%MCMC(suffix = 1, seed = 500);
%MCMC(suffix = 2, seed = 203);
%MCMC(suffix = 3, seed = 140);

```

```
data out.events(keep = chain iteration T_1-T_40);  
set outpost1(in = ina) outpost2(in = inb) outpost3(in = inc);  
if ina then chain = 1;  
else if inb then chain = 2;  
else if inc then chain = 3;  
run;
```

References

- [1] Berry, Scott M., and Donald A. Berry. "Accounting for Multiplicities in Assessing Drug Safety: A Three-Level Hierarchical Mixture Model." *Biometrics* 60.2 (2004): 418-26. Web. 21 July 2014.
- [2] Dumouchel, William. "Multivariate Bayesian Logistic Regression for Analysis of Clinical Study Safety Issues." *Statistical Science* 27.3 (2012): 319-39. JSTOR. Web. 21 July 2014.
- [3] Gelman, Andrew. *Bayesian Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC, 2004. Print.
- [4] Montgomery, Douglas C., and Elizabeth A. Peck. *Introduction to Linear Regression Analysis*. New York: J. Wiley, 1992. Print.
- [5] Mehrotra, Devan V. "Use of the False Discovery Rate for Evaluating Clinical Safety Data." *Stat Methods Med Res* 13 (2004): 227-38. Print.
- [6] Zink, Richard C. "Assessing Drug Safety with Bayesian Hierarchical Modeling Using PROC MCMC and JMP." *Pharma and Health Care* (2013): n. pag. Web. 21 July 2014.