Generalized Linear Models Using Markov Chain Monte Carlo

Abdul-ganiy Babatunde USMAN (abdul.usman@aims.ac.rw)
Supervised by: Doctor Viani DJEUNDJE BIATAT
University of Edinburgh, United Kingdom

June 2017

AN ESSAY PRESENTED TO AIMS RWANDA IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE IN MATHEMATICAL SCIENCES

DECLARATION

This work was carried out at AIMS Rwanda in partial fulfilment of the requirements for a Master of Science Degree.

I hereby declare that except where due acknowledgement is made, this work has never been presented wholly or in part for the award of a degree at AIMS Rwanda or any other University.

Student: Abdul-ganiy Babatunde Usman 

Supervisor: DJEUNDJE BIATAT Viani A 

ACKNOWLEDGEMENTS

Firstly, I appreciate the Supreme Being, the Master of the whole universe, for the successful completion of this program.

I can't find the right words to appreciate my parents Mr. and Mrs. Odeniyi, you are the best parents one can have. I also recognize the support of my siblings, thank you all.

I would like to thank my supervisor Dr. Viani DJEUNDJE BIATAT for his friendly support and guidance and my tutor Savannah Nuwagaba for her advise and assistance.

My acknowledgement also goes to the awesome AIMS family, staff and students. You all have greatly impacted in me.

DEDICATION

This project work is dedicated to Almighty ALLAH, the creator and fashioner of all existing being, to whom I am forever indebted to serve.

Abstract

In statistics, there are two approaches to model fitting. This study applied both approaches, the Classical and Bayesian (using Markov chain Monte Carlo), in fitting Generalized Linear Models, GLM. We considered fitting the GLM using 3 distributions from the exponential family of distribution; the Normal, Binomial and Poisson. Results showed that both approaches used gave quite similar model parameters' values. The Mean squared error, Akaike Information Criteria and Bayesian Information Criteria were used as tools to select the best model distribution. Based on the mortality data used, the Binomial and Poisson distributions are most appropriate as the model distributions as they both had the least MSE.

Contents

1. Introduction

Decision making is a major part of human daily activities and plays a vital role in science. The ability to predict and project into the future has made science impacted greatly. However in many circumstances, predictions come with challenges of uncertainty which scientists have tried to understand and include in their prediction settings (Pielke, 2003). One major prediction method is the use of classical linear models and least squares method. This began with the work of Gauss and Legendre who applied the methods to astronomical data (Stigler and Stephen, 1981). The data they observed were usually measurements of continuous quantities such as the positions and magnitudes of the planetary objects. And the variability in the observations obtained was largely the effect of measurement errors. The Normal, or Gaussian, distribution was viewed as a mathematical construct developed to describe the properties of such errors. Later in the nineteenth century the same distribution was used to describe the variation between individuals in a biological population in respect of a character such as height, and this led numerous applications of linear models in biology. Gauss introduced the Normal distribution of errors as a device for describing variability. He showed, however, that many of the important properties of linear model least-squares estimates depend not only on Normality but on the assumptions of constant variance (McCullagh and Nelder, 1989).

The classical linear model assumes that the (continuous, normally distributed with constant variance) response variable is a linear function of (set of) independent variables. These independent variables can be continuous, categorical or combination of both (Dunteman and Ho, 2005). In practice however, the response variable can have different forms including categorical, counts, binary and non-normal continuous. Models from such problems result in the conditional mean being a non-linear function of the independent variables (Dunteman and Ho, 2005). The classical linear modelling framework is not flexible enough to handle these issues. A solution is found in the so-called Generalized Linear Models.

Generalized Linear Models allow the modelling of data from a wider class of distributions. Such distribution affects the assumptions we make regarding variances (Olsson, 2002). It is important to mention that the linear model framework as well as its generalization assumes the model parameters as fixed and unknown constants to be estimated. This view is known as the *classical* (or *frequentist*) view. A contrary view considers the model parameters as uncertainties described by some (prior) probability distribution. This view is the *Bayesian*'s (Kery, 2010).

In Bayesian modelling, inference about the model parameters requires integrating over possibly high-dimensional probability distributions. This is quite tedious and in most cases, not feasible numerically (Gilks et al., 1995). A general intervention is evaluating the integrand at a randomly generated set of points, a technique known as Monte Carlo integration. This method is based on (pseudo-) random sampling from probability distributions (Cowles, 2013). Gamerman (2006) defined Markov chain as a type of stochastic process that deals with characterization of sequences of random variables. The stochastic processes are independent in their states; present, past and future. In other words, Markov chains are random variables that are generated sequentially over time. Since Monte Carlo technique allows sampling from complex probability density functions, Markov Chain Monte Carlo (MCMC) is essentially the integration of Monte Carlo using Markov

chains. MCMC applications in statistics are largely oriented towards Bayesian modelling and inference.

It is important to mention that Bayesian approach to modelling comes with different many assets. In particular, complex statistical models which are intractable using classical methods of fitting, can be fitted using Bayesian computational method (Link et al., 2002). Often-times, there are assumptions around some parameters in conducted studies/experiments. As a result, one need to build model that will accommodate the prior knowledge about such parameters. Bayesian modelling allows incorporating external knowledge into an analysis and this may greatly increase the precision of the estimates (McCARTHY and MASTERS, 2005). Some parameters may only be estimable through precisely this combination of information. Classical inference is unbiased only for large sample sizes. In contrast, Bayesian inference is exact for any sample size (Kery, 2010).

Despite the advantages of Bayesian approach in modelling over its classical counterpart, there is some resistance to the Bayesian methods as it is perceived to, unconsciously, inject information into an analysis via the priors (Kery, 2010). Also, practical implementation of Bayesian analysis typically involves custom-written code in high level programming languages. Therefore, solid knowledge in statistics and computing is required for Bayesian analyses.

This essay, essentially, investigates the generalized linear models analysis results using the classical and Bayesian approach (using MCMC). In Chapter 2, we introduced the linear models and extend it to its generalization, generalized linear models. In Chapter 3, we discussed the Bayesian approach with Markov chain Monte Carlo. In Chapter 4, we presented the data used as well as methods and results of analysis conducted. Chapter 5 contains the study inferences and future interventions.

2. Basic concepts and definitions

In this chapter, we introduce the foundations of (classical) linear models which is the pre-requisite to understanding generalized linear models. But first, it is important to know the root of every analysis: data.

2.1 Data collection

Any statistical analysis is as good as the data on which it is based. This makes data collection an essential part of an analysis. Montgomery et al. (2012) has classified methods for collecting data into three categories: retrospective study, observational study and experimental design.

- A retrospective study is employing the use of previously collected data for a present study. This will definitely minimize the cost of study. However, such historical data often involves problems of missing observations, reliability and quality.
- An observational study as the name imply, simply observes the process within which our interest lies in collecting information. This study requires time to obtain significant number of observations.
- A designed experiment requires one to conduct planned and well-defined strategy, called the experimental design, to collect data. This data collection method is the best but involves high cost.

Problem arises from analysis when poor scheme is employed in data collection. This consequently leads to result misinterpretation.

2.1.1 Data types. In general, data can take two forms: Numerical and Categorical.

1 Numerical data are measured quantitatively, they are presented not in terms of description but with numbers. Numerical data can be divided into two classes: discrete and continuous.
   - Discrete data represents numbers that can be listed out. They are counts represented by elements in the natural numbers, $\mathbb{N}$.
   - Continuous data are measurements whose possible values cannot be counted and can only be described using intervals on the real number line.

2 Categorical data are characteristics or qualities that can not be measured quantitatively. They can, however, be represented with elements, objects or sometimes numbers that have no mathematical meaning.

Random variable is a variable whose possible values are numerical outcomes of a random phenomenon (or experiment). Random variable, denoted by X or Y, can either be *discrete* or

continuous. If discrete, its values are countable distinct numbers, for example, the number of students in a class. Continuous random variable is one which takes an infinite number of possible values.They are usually measurements. Examples include height and weight (Easton and McColl, 1997).

2.2 Classical Linear model

Linear models, also known as regression models, are used to study how a numerical variable measured on a particular scale depends on one or more predictor (or explanatory) variables. The predictors themselves may be quantitative or qualitative. More specifically, regression analysis is a statistical method employed in investigating and modelling the relationships between variables. A linear regression model in a simple form is given by:

$$y = \beta_0 + \beta_1 x + \epsilon. \tag{2.2.1}$$

y is called the response variable, x the predictor or explanatory variable, β's are the model parameters and ϵ stochastic error. Equation (2.2.1) is often referred to as a simple linear regression model because it involves only one predictor.

A classical regression model can however have more than one predictor and thus written in a general form:

$$Y_i = \mathbf{x}_i^T \beta. \tag{2.2.2}$$

where $Y_1, \cdots, Y_N$ are observations of the response variable, and $\mathbf{x}_i$ are the observations of the covariates or predictors. This model is written compactly as

$$\mathbf{y} = X\beta + \mathbf{e}. \tag{2.2.3}$$

where

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}, \qquad X = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \qquad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

, and the $e_i's$ are independently and identically distributed random variables with $e_i \sim N(0, \sigma^2)$. A common method of estimating the parameters, β's, is the *least squares estimation* method. The method is about estimating model parameters by minimizing the squared errors (or discrepancies) between the observed data.

An important quality of the least squares estimators, $\boldsymbol{\beta}$, for the regression model in (2.2.3) above, according to the Gauss-Markov theorem, is that they are unbiased with minimum variance when compared with all other unbiased estimators that are linear combination of $\mathbf{y}$(Montgomery et al., 2012)

2.3 Generalized Linear Model

Generalized linear models, GLM, extend linear regression to response types other than Gaussian. GLM are used to model response variables from a large class of distributions known as the exponential family of distributions. Unlike classical linear regressions, GLMs do not assume constant variance. Additionally, they are used when the model involves non-linear functions of the mean. Essentially, a GLM consists of three components: the random component, the systematic component and the link function.

- Random component: an assumption often made in the classical linear model is that the response set Y_i are independently normally distributed with constant variance. We can relax this assumption to permit the distribution to be any that belongs to the exponential family of distributions. For such distribution with parameter η, called the canonical parameter, we have
$$\eta = E(Y) = \mu$$
.
- The systematic component specifies the linear combination of the predictor variables (covariates), $\mathbf{x}_i, i = 1, \cdots, p$, in the model.
$$X\beta = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}, \qquad i = i, \cdots, n$$
.
- Link function is basically the connection (link) between the random and systematic component. It describes how the mean response, $E(Y) = \mu$, is linked to the linear predictors.
$$\eta = g(\mu) = X\beta$$
.

2.4 Exponential family

Let Y be a random variable with probability distribution $f(y, \theta, \phi)$, the distribution of Y belongs to the exponential family if it can be written in the form:

$$f(y, \theta, \phi) = exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}. \tag{2.4.1}$$

for some specific functions $a(\cdot), b(\cdot)$ and $c(\cdot)$. θ is the canonical or location parameter, ϕ is the scale or dispersion parameter (Faraway and James, 2016).

Not all statistical distribution can be written in the form (2.4.1) above, for example, the student-t distribution. Some common probability distributions that belong to the exponential family are the normal, poisson and binomial distributions.

2.4.1 Properties of exponential family distributions (Dobson and Barnett, 2008).

i

$$\int f(y,\theta,\phi)dy = 1. \tag{2.4.2}$$

This is a general property for any continuous statistical distribution. In the discrete case, the integral is replaced by sum (over the range of y). Also, we recall from the definition of expectation of a random variable, Y, that;

$$E(Y) = \mu = \int yf(y,\theta,\phi)dy. \tag{2.4.3}$$

and the expectation that it differs from its mean is

$$E(Y-\mu) = var(y) = \int (y-\mu)f(y,\theta,\phi)dy. \tag{2.4.4}$$

ii $E(Y) = \frac{db(\theta)}{d\theta} = b'(\theta)$.

Proof. First note that; $\frac{d}{d\theta}\int f(y,\theta,\phi)dy = 0$, reversing the order of the differentiation and integration we obtain the same result :

$$\int \frac{df(y,\theta,\phi)}{d\theta}dy = 0. \tag{2.4.5}$$

same for,

$$\int \frac{d^2f(y,\theta,\phi)}{d\theta^2}dy = 0. \tag{2.4.6}$$

Now from (2.4.1),

$$\begin{aligned} f(y,\theta,\phi) &= exp\left\{\frac{y\theta - b(\theta)]}{a(\phi)} + c(y,\phi)\right\}, \\ \frac{df(y,\theta,\phi)}{d\theta} &= \frac{y-b'(\theta)}{a(\phi)}f(y,\theta,\phi), \\ &= \frac{1}{a(\phi)}[yf(y,\theta,\phi) - b'(\theta)f(y,\theta,\phi)], \end{aligned}$$

integrating with respect to y

$$\begin{aligned} \int \frac{df(y,\theta,\phi)}{d\theta}dy, &= \int \frac{1}{a(\phi)}[yf(y,\theta,\phi) - b'(\theta)f(y,\theta,\phi)]dy \\ &= \frac{1}{a(\phi)}[E(Y) - b'(\theta)], \end{aligned}$$

$$\frac{1}{a(\phi)}[E(Y) - b'(\theta)] = 0, \qquad \text{from (2.4.5)}$$
$$E(Y) = b'(\theta).$$

207 □

208 • $Var(Y) = \frac{d^2 b(\theta)}{d\theta^2} = b''(\theta)a(\phi)$.

Proof. Using the same argument from the mean:

$$\frac{d^2 f(y,\theta,\phi)}{d^2\theta} = \left(\frac{y - b'(\theta)}{a(\phi)}\right)^2 f(y,\theta,\phi) - \frac{b''(\theta)}{a(\phi)} f(y,\theta,\phi),$$

But $E(Y) = b'(\theta)$,

$$\frac{d^2 f(y,\theta,\phi)}{d^2\theta} = \left(\frac{y - E(Y)}{a(\phi)}\right)^2 f(y,\theta,\phi) - \frac{b''(\theta)}{a(\phi)} f(y,\theta,\phi),$$

integrating with respect to y,

$$\begin{aligned}\int \frac{d^2 f(y,\theta,\phi)}{d^2\theta} dy &= \int \left(\frac{y - E(Y)}{a(\phi)}\right)^2 f(y,\theta,\phi) dy - \int \frac{b''(\theta)}{a(\phi)} f(y,\theta,\phi) dy, \\ &= \frac{1}{[a(\phi)]^2} \int (y - E(Y))^2 f(y,\theta,\phi) dy - \frac{b''(\theta)}{a(\phi)} \int f(y,\theta,\phi) dy, \\ &= \frac{1}{[a(\phi)]^2} var(Y) - \frac{b''(\theta)}{a(\phi)},\end{aligned}$$

$$\frac{1}{[a(\phi)]^2} var(Y) - \frac{b''(\theta)}{a(\phi)} = 0, \qquad \text{from (2.4.6)}$$

And so

$$Var(Y) = b''(\theta)a(\phi).$$

209 □

2.4.2 Some distributions from exponential family.

1 The Normal distribution is best appropriate to model continuous response variable. The distribution for any random variable $Y, -\infty < Y < \infty$, is given thus

$$f(y,\theta,\phi) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}.$$

And this can be written in its exponential form as

$$f(y,\theta,\phi) = exp\left\{-\frac{1}{2}ln(2\pi\sigma^2) - \frac{(y^2+\mu^2-2y\mu)}{2\sigma^2}\right\},$$
$$f(y,\theta,\phi) = exp\left\{\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}ln(2\pi\sigma^2)\right\}.$$

With, $\theta = \mu$, $b(\theta) = \frac{\mu^2}{2}$, $a(\phi) = \sigma^2$ and $c(y,\phi) = -\frac{y^2}{2\sigma^2} - \frac{1}{2}ln(2\pi\sigma^2)$.

$$E(Y) = b'(\theta) = \mu, \qquad Var(Y) = b''(\theta)a(\phi) = \sigma^2.$$

2 The Binomial distribution is used for observations involving more than one binary outcome for each combination of the predictors. In this case, the response becomes some number of one type and some number of the other type (Faraway and James, 2016).

$$\begin{aligned} f(y,\theta,\phi) &= \binom{n}{k}\pi^y(1-\pi)^{n-y}, \\ &= exp\left\{ln\binom{n}{k} + yln\pi + (n-y)ln(1-\pi)\right\}, \\ &= exp\left\{yln\left(\frac{\pi}{1-\pi}\right) + nln(1-\pi) + ln\binom{n}{k}\right\}. \end{aligned}$$

For the binomial distribution,

$\theta = ln\left(\frac{\pi}{1-\pi}\right)$, $b(\theta) = nln(1-\pi)$, $a(\phi) = 1$ and $c(y) = ln\binom{n}{k}$.

It is therefore trivial to show that

$$E(Y) = n\pi, \qquad Var(Y) = n\pi(1-\pi).$$

3 The Poisson distribution is often used for count events that occur randomly over time at a particular rate (Agresti, 2015). When the response is given as an unbounded count, we can explain this in terms of the given predictors by using the poisson model.

If Y is Poisson with mean $\lambda > 0$, then:

$$f(y,\theta,\phi) = \frac{\lambda^y e^{-\lambda}}{y!}, \qquad y = 0,1,2,\cdots.$$

And in the exponential form

$$f(y,\theta,\phi) = exp\{yln\lambda - \lambda - ln(y!)\}.$$

We then have that, $\theta = ln\lambda$, $b(\theta) = \lambda$, $a(\phi) = 1$ and $c(y) = -ln(y!)$.

$$E(Y) = \lambda, \qquad Var(Y) = \lambda.$$

This is a property of the Poisson distribution, the mean is equal to the variance.

We summarize thus

Distribution	Canonical link function	
Normal	Identity	μ
Binomial	Logit	$log\left[\frac{\pi}{1-\pi}\right]$
Poisson	log	$log\lambda$
Gamma	Reciprocal	$\frac{1}{\mu}$
Inverse Gaussian	Squared reciprocal	$\frac{1}{\mu^2}$

Table 2.1: Canonical links for GLMs

2.5 Estimation of GLMs

One method to estimate parameters in GLM is to use the method of maximum likelihood. Maximum likelihood estimation is based on the notion that estimated parameters should be those that maximize the value of the sampled data distribution. This implies that maximum likelihood estimation finds the parameter values that most likely generated the sample observations (Dunteman and Ho, 2005).

Consider independent random variables, Y_i. To estimate the parameters $\boldsymbol{\beta}$, which are related to the Y_i through the link function $g(\mu_i)$, we use the log-likelihood function of (2.4.1).

Likelihood function is the probability density of the data given the parameters. It is regarded as a function of the parameters (Filippi, 2015). In our case

$$L(\boldsymbol{\beta}, \mathbf{y}) = \prod_{i=1}^{n} f(y_i, \theta, \phi). \tag{2.5.1}$$

The log of (2.5.1) is given by

$$l = logL(\boldsymbol{\beta}, \mathbf{y}) = log \prod_{i=1}^{n} f(y_i, \theta, \phi),$$

$$l = \sum_{i}^{n} \left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right). \tag{2.5.2}$$

We then maximize by differentiating (2.5.2) w.r.t the parameters $\boldsymbol{\beta}$. This can be done using the chain rule:

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \frac{dl}{d\theta_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{d\eta_i}{d\boldsymbol{\beta}}.$$

Note that:

1. $\frac{dl}{d\theta_i} = \frac{1}{a(\phi)}\left[y_i - \frac{db(\theta_i)}{d\theta_i}\right] = \frac{1}{a(\phi)}(y_i - \mu_i)$,
2. $\frac{d\theta_i}{d\mu_i} = \frac{1}{b''(\theta)} = \frac{a(\phi)}{Var(y_i)}$,
3. $\frac{d\eta_i}{d\boldsymbol{\beta}} = \mathbf{x_i}$.

These together yields

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)\mathbf{x}_i}{Var(y_i)}\frac{d\mu_i}{d\eta_i}.$$

where $\eta_i = g(\mu_i) = \mathbf{X}\boldsymbol{\beta}$. The maximum likelihood estimator of each β_j can be found by using Iteratively Re-weighted Least Square (IRLS) procedure (Charnes et al., 1976; Montgomery et al., 2012). It can be shown that the solution takes the form

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{z}.$$

where $z_i = \hat{\eta}_i + (y_i - \hat{\mu}_i)\frac{d\eta_i}{d\mu_i}$.

The diagonal elements of the matrix $\mathbf{V}$ are the variances of z_i's.

3. Bayesian Models and Markov Chain Monte Carlo

3.1 Bayesian method

Statistical inference from models is based on the (unknown) model parameters, say θ_i, and the observed data, Y_i. Unlike the classical statistical theory, Bayesian approach assumes that all unknown parameters, including the regression coefficients, are random. This randomness is taken into account via the so-called prior distribution (Ntzoufras, 2009). A prior distribution is the belief we have about the model parameter. After having seen the data, there is need to update the belief on the unknown parameters. The probability distribution assigned to the parameters on the basis of the prior distribution and the observed data is called the posterior distribution, and this is calculated using the Bayes' theorem.

Bayes' formula for the posterior distribution of a parameter θ is

$$P(\theta|\mathbf{y}) = \frac{P(\mathbf{y}|\theta)P(\theta)}{\int P(\mathbf{y}|\theta)P(\theta)d\theta}. \tag{3.1.1}$$

3.1.1 Bayesian Models. The mathematical description of Bayesian model as explained by Liang et al. (2010) is given thus

i specifying a sample distribution for the observed variable Y,

$$Y \sim f(Y|\theta),$$

ii specifying the prior distribution before seeing the data,

$$\theta \sim f(\theta),$$

iii Finding the Posterior distribution of the parameter after seeing the data,

$$f(\theta|Y) = \frac{f(\mathbf{y}|\theta)f(\theta)}{f(\mathbf{y})}, \qquad f(\mathbf{y}|\theta) = \prod_i^n f(y_i|\theta). \tag{3.1.2}$$

Since $f(y)$ is not dependent on θ. Equation (3.1.2) can be written as

$$f(\theta|Y) \quad \propto f(\mathbf{y}|\theta)f(\theta). \tag{3.1.3}$$

In other words,

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

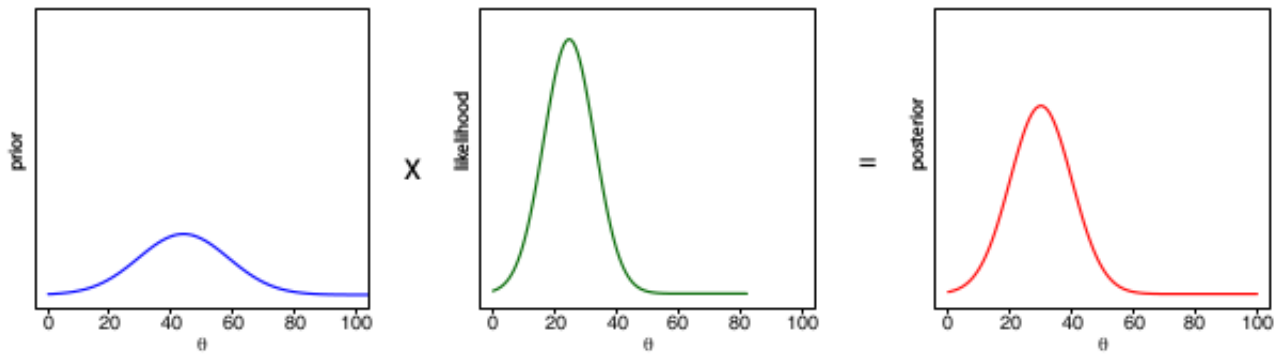


Figure 3.1: Typical Bayes plot (Filippi, 2015)

Figure 3.1 shows that the prior distribution is only moderately informative. But with the likelihood, calculated from the data, it will give a highly informative posterior distribution.

3.2 Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) is a methodology that allows to construct models by drawing a long chain samples from the required distribution and form sample averages to approximate the parameters of interest (Gilks et al., 1995). It is a technique that is well used in Bayesian statistical modelling.

3.2.1 Monte Carlo. Suppose θ is a vector of unknown parameters and $\mathbf{y}$ is the observed data, then the posterior mean of a function $g(\theta)$ is given as

$$E[g(\theta)] = \int g(\theta)p(\theta|y)d\theta. \tag{3.2.1}$$

where $p(\theta|y)$ is the posterior distribution.

The property described above implies that when we have a large sample from $p(\theta|y)$ we can estimate (approximate) the expected value of $g(\theta)$ more accurately. If

$$\theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(n)}$$

are n samples drawn from $p(\theta|y)$ such that this samples are independent and identically distributed. The quantity

$$g_n(\theta) = \frac{1}{n}\sum g(\theta^{(n)}),$$

by law of large numbers, will converge to $E[g(\theta)]$. This is known as Monte Carlo (Ramuada, 2016)

3.2.2 How does the MCMC work?. The MCMC works such that samples

$$\theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(n)}$$

are drawn from the posterior distribution $p(\theta|y)$. That is,

$$\theta^{(i)} \sim p(\theta|y), \text{ for } i = 1, 2, 3, \cdots, n.$$

Then $\{\theta^{(i)}\}_{i=1}^{n}$ is a bag of samples which is used as a proxy for $p(\theta|y)$, the posterior distribution, (SAS-Publishing, 2009). This implies that MCMC produces a sequence (a chain) of random variables $\{\theta^{(i)}\}_{i=1}^{n}$ such that $p(\theta^{(i)}) \approx p(\theta|y)$ (SAS-Publishing, 2009; Ramuada, 2016).

The most prominent MCMC algorithms are the Metropolis-Hastings algorithm and the Gibbs sampler (Kroese et al., 2011).

Metropolis-Hasting algorithm (Ramuada, 2016) simulates samples from a probability distribution by making use of the full joint density function and (independent) proposal distributions. In practice, it works as follows:

1. Initialize $\theta^{(0)} \sim q(\theta)$
2. For iteration $i = 1, 2, \cdots$
   - i Draw θ^* from $q(\theta^{(i)}|\theta^{(i-1)})$, called the *proposal or candidate* distribution.
   - ii compute the acceptance probability

$$\alpha(\theta^*|\theta^{(i-1)}) = min\left(1, \frac{q(\theta^{(i-1)}|\theta^*)\pi(\theta^*)}{q(\theta^*|\theta^{(i-1)})\pi(\theta^{(i-1)})}\right)$$

   where $\pi(\theta^*)$ is the probability that the next state of the chain is θ^* and $\pi(\theta^{(i-1)})$ is the probability that the next state of the chain is $\theta^{(i-1)}$
   - iii sample $u \sim U(0, 1)$
     - if $u \leq \alpha(\theta^*|\theta^{(i-1)})$, we set current $\theta^{(i)}$ to be equal to the draw value θ^*, that is, $\theta^{(i)} = \theta^*$.
     - else, we set $\theta^{(i)} = \theta^{(i-1)}$.

This algorithm is repeated sufficiently (n times) for step 2 until we obtain a number of samples. This will result in a Markov chain $\{\theta^{(i)}\}_{i=1}^{n}$ to be used for posterior inference.

Gibbs sampler (Gamerman, 2006) Gibbs sampling technique basically generates posterior samples by sweeping through each variable and samples from its conditional distribution with the remaining variables fixed to their current values. The algorithm can be summarised as follows:

1. Set initial values $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \cdots, \theta_d^{(0)})$
2. For iteration $i = 1, 2, 3, \cdots$, obtain a new value $\theta^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \cdots, \theta_d^{(i)})$ from $\theta^{(i-1)}$ through successive generation of the values

$$\begin{aligned}
\theta_1^{(i)} &\sim p(\theta_1|\theta_2^{(i-1)}, \cdots, \theta_d^{(i-1)}) \\
\theta_2^{(i)} &\sim p(\theta_2|\theta_1^{(i)}, \cdots, \theta_d^{(i-1)}) \\
&\vdots \\
\theta_d^{(i)} &\sim p(\theta_d|\theta_1^{(i)}, \cdots, \theta_{d-1}^{(i)})
\end{aligned}$$

The iteration in step 2 continues until convergence. When the convergence is reached the resulting value $\theta^{(i)}$ is a draw from the stationary (posterior) distribution of interest (Gamerman, 2006).

3.2.3 Convergence of the Markov Chain. The simulated draws are used to make inferences about the posterior distribution. But a serious issue which we face is deciding if Markov Chains have converged to the stationary distribution. A virtual analysis test called the Trace plot can be used. It is a plot of the samples against the number of iterations. According to SAS-Publishing (2009) if the distribution of points is not changing with time as the Markov Chain progresses, then we say that the chain has reached convergence.

3.2.4 Implementing MCMC. Many software packages have been developed overtime to run MCMC algorithms. The prominent one is the BUGS language, Bayesian analysis using Gibbs Sampling, which includes WINBUGS and OPENBUGS (Lunn et al., 2009). Other similar MCMC packages are STAN and JAGS. The JAGS language is essentially a clone of BUGS and it has been used for the analysis contained in this work.

3.3 Modelling in JAGS

JAGS, an acronym for Just Another Gibbs Sampler, is a program designed for inference on Bayesian models using Markov Chain Monte Carlo (MCMC) simulation (Plummer, 2015). It essentially generates samples from the posterior distribution of the model parameters. This is achieved in five basic steps: definition of the model, compilation, initialization, adaptation and monitoring.

3.3.1 Model definition. The model is defined in a text file using BUGS language dialect. The model definition is in two parts contained inside a block delimited by curly brackets and preceded by the keyword *model* or *function*. The first part is the likelihood definition and the second part is the (parameters') prior distribution specification. A typical example is contained in the appendix.

3.3.2 Compilation.

3.3.3 Initialization. Running a model requires initialization of model parameters. This can be done by:

- the user setting chosen initial values for the model parameters,
- randomly generating values for each chain or
- Samplers object in the model automatically choosing initials for all parameters and update them from one iteration to the next.

3.3.4 Adaptation. In practice, as the number of iterations tends to a finite limit MCMC sampler converges to the target distribution (i.e., parameters). By convention, MCMC output is divided into: an initial "burn-in" period, which is discarded, and the remainder of the run, in which the output is considered to converge (sufficiently close) to the target distribution. Samples from the second part are used to create approximate summary statistics of the target distribution.

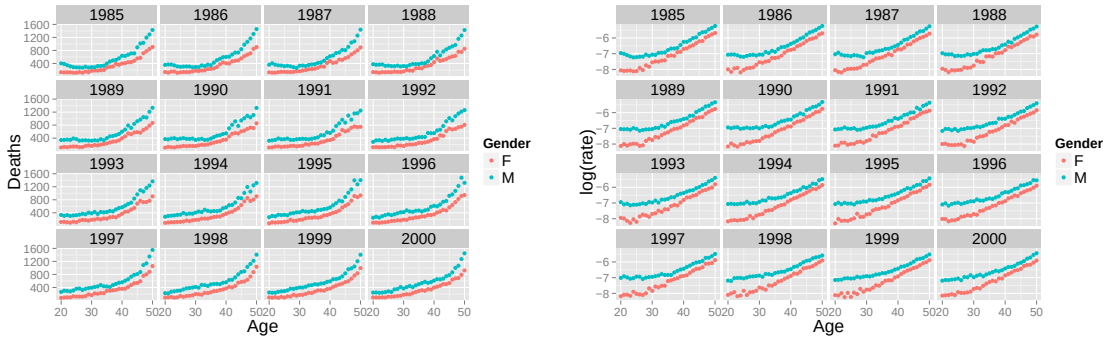
322 **3.3.5 Monitoring.** Monitoring in JAGS involves object that records sampled values. The simplest
323 monitor is the trace monitor which stores the sampled values of a defined parameter (or node)
324 at each iteration.

4. Data, Methods and Analysis

4.1 Data Presentation and description

For illustration, we used aggregated mortality data from males and females of ages from 20 to 50 and calendar years 1985 to 2000. The dataset was supplied by my supervisor and contains the following columns: Exposure, Deaths, Age, Gender, Year. Exposure is the number of people exposed to the risk of dying. Deaths is the numbers of death counts. Age refers to the age of individuals at death. All these informations were collected over 16 years period.

Figure 4.1a below shows a graphical description of the data. It can be observed that the number of death (for both males and females) from the exposed individuals increases with age. The log transformation of the death rate, $\frac{Deaths}{Exposure}$, in figure 4.1b somewhat shows a linear trend between the death rate and age.



(a) Original data

(b) Transformed data

Figure 4.1: Data description

4.2 Model specification and implementation

GLM was fitted following the classical and Bayesian view of statistical modelling. The classical model involve using the **glm** function in **R** while the MCMC approach will be used for the Bayesian modelling. Three distribution types will be considered:

1 **Normal distribution**: with appropriate transformation of the response variable, death rate. The model was fitted using the normal distribution with the identity link function.

2 **Binomial distribution**: supposing the data a result from a Bernoulli experiment involving success and failure repeated in a number of times. Success in this case is the death count and Exposure represent the number of repetitive trial of the experiment. In the generalized linear model, we specify a binomial distribution for the response with it's canonical logit link.

3 **Poisson distribution**: we took the response variable as count of (death) event. Since the number of deaths, for each observation, is recorded from different number of exposed individuals,we assume that the mean is proportional to the exposure.

And so,

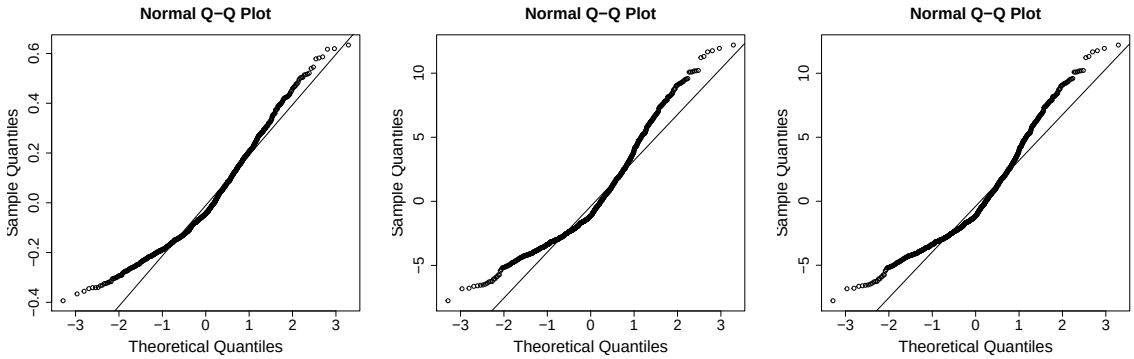
$$Deaths \sim Poisson(Exposed \times \lambda).$$

The canonical link of the poisson distribution is the log;

$$\eta = g(\theta) = \mathbf{X}\boldsymbol{\beta},$$
$$g(\theta) = log\left(\frac{Deaths}{Exposed}\right),$$
$$log(Deaths) = log(Exposed) + \mathbf{X}\boldsymbol{\beta}.$$

The term $log(Exposed)$ is called the offset.

4.2.1 Covariates. At first, we considered, in the model, all the predictors: Age, Gender and Year (scaled by 2000). A quantile-quantile plot was used to assess the plausibility of the covariates and the model distribution as shown below.



(a) Normal qqplot (b) Binomial qqplot (c) Poisson qqplot

Figure 4.2: Quantile-Quantile plot of the distributions considered

The figures 4.2a, 4.2b and 4.2c above clearly indicate that the models considered do not fit well. This was improved by adding (to the model) a quadratic Age term and an interaction term between Age and Gender.

4.2.2 Markov Chain Monte Carlo implementation.

- The packages **R2jags** and **mcmcplots** in **R** was used to perform analysis and plotting of MCMC output,
- We simulated 100000 draws for four parallel chains. The first 1000 part of the simulations were discarded - burn-in, to minimize the influence of the Markov Chain initial state.
- Trace plot was used to access the convergence of the chain to the stationary distribution.

4.2.3 Model comparison. Since three different distributions were considered in the model fit, the need was to select the best family of distribution appropriate for our data. For the classical model fitting, the criteria for comparison among the distributions used was the Mean Squared Error, MSE. The Akaike Information Criterion, AIC (Akaike, 1978), and Schwart criterion, BIC (Schwarz, 1978) was used to compare between the Binomial and Poisson models, since they both have the same response form.

In statistics, the MSE tells how close the model fit to the data. The smaller the mean squared error, the better the model fit. The AIC and BIC are criteria for model selection among a finite set of models; the model with the lowest AIC and/or BIC is preferred.

Instead of the AIC and BIC in the classical approach, the Deviance Information Criterion, DIC (van der Linde, 2005; Spiegelhalter et al., 2002), serves as the selection criterion for models fitted using the Bayesian MCMC approach. The deviance information criterion (DIC) is a generalization of the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). It is particularly useful in Bayesian model selection problems where the posterior distributions of the models have been obtained by Markov chain Monte Carlo (MCMC) simulation. However, it suffices to infer using only the outcome from the classical comparison.

4.3 Result and Analysis

4.3.1 Results of the Classical approach.

- **Normal distribution**

Listing 4.1: Normal model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	−7.87930	0.046362	−169.95	0.00e+00
Age	−0.05229	0.002692	−19.42	2.05e−71
GenderM	1.50273	0.021911	68.58	0.00e+00
I(Age^2)	0.00190	0.000038	50.10	3.03e−273
Year_f	0.00489	0.000588	8.31	3.16e−16
Age:GenderM	−0.02460	0.000607	−40.56	2.14e−212

- **Binomial distribution**

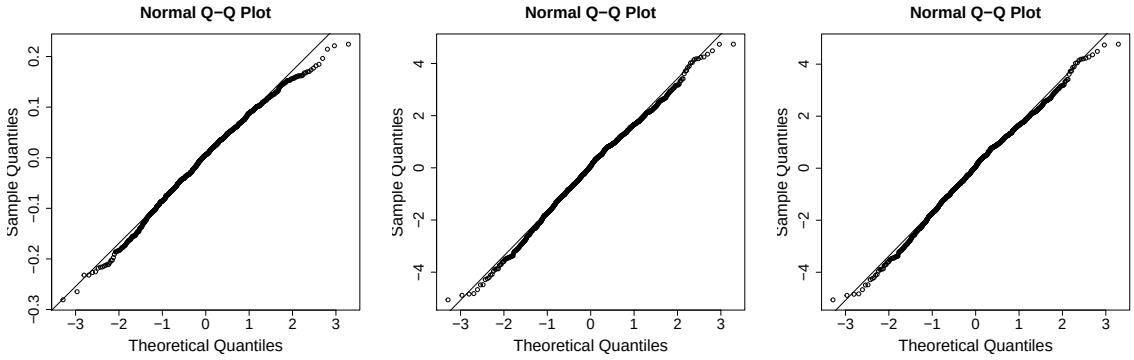
Listing 4.2: Binomial model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−7.82579	3.14e−02	−249.3	0.00e+00
Age	−0.05337	1.66e−03	−32.2	4.15e−227
GenderM	1.36038	1.51e−02	89.9	0.00e+00
I(Age^2)	0.00188	2.19e−05	85.8	0.00e+00
Year_f	0.00825	3.25e−04	25.4	3.37e−142
Age:GenderM	−0.02065	3.69e−04	−56.0	0.00e+00

- **Poisson distribution**

Listing 4.3: Poisson model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−7.83145	3.14e−02	−249.6	0.00e+00
Age	−0.05299	1.66e−03	−32.0	2.83e−224
GenderM	1.36070	1.51e−02	90.0	0.00e+00
I(Age^2)	0.00187	2.19e−05	85.6	0.00e+00
Year_f	0.00824	3.25e−04	25.4	5.86e−142
Age:GenderM	−0.02068	3.68e−04	−56.1	0.00e+00



(a) Normal qqplot (b) Binomial qqplot (c) Poisson qqplot

Figure 4.3: Quantile-Quantile plot of the distributions considered

From the output listings 4.1, 4.2 and 4.3 we can observe quite similar values for the parameters in each model. These parameter values are very much significant in all considerations. The Quantile-Quantile plots in figure 4.3 dipict a good model fit.

Distribution	Criterion		
	MSE	AIC	BIC
Normal	2.07e-08		
Binomial	1.52e-08	10590	10619.41
Poisson	1.52e-08	10586	10615.78

Table 4.1: Result summary for the classical approach

Table 4.1 above shows the values of each model selection criterion. The models with the Binomial and Poisson distribution give the least MSE. Using the AIC and BIC values, the Possion model seems better than the Binomial model. This implies that the Poisson distribution is fits best to model our data.

4.3.2 Results of the MCMC approach.

- **Normal distribution Posterior summary**

Listing 4.4: Normal mcmc model

	Mean	SD	Naive SE	Time-series SE
Age	−2.78e−02	2.94e−03	4.64e−05	4.56e−05
Age_GenderM	−2.46e−02	6.48e−04	1.02e−05	1.02e−05
Age_squared	1.90e−03	4.05e−05	6.41e−07	6.40e−07
GenderM	1.50e+00	2.35e−02	3.71e−04	3.71e−04
Intercept	−9.38e+00	5.73e−02	9.05e−04	8.82e−04
Year	4.89e−03	6.19e−04	9.79e−06	9.50e−06
deviance	−2.06e+03	1.79e+01	2.83e−01	2.83e−01

- **Binomial distribution**

Listing 4.5: Binomial mcmc model

	Mean	SD	Naive SE	Time-series SE
Age	−3.28e−02	1.69e−03	2.66e−05	1.18e−04
Age_GenderM	−2.07e−02	3.57e−04	5.64e−06	2.57e−05
Age_squared	1.88e−03	2.05e−05	3.25e−07	1.26e−06
GenderM	1.36e+00	1.49e−02	2.35e−04	1.24e−03
Intercept	−9.19e+00	3.86e−02	6.10e−04	3.40e−03
Year	8.26e−03	3.38e−04	5.35e−06	2.23e−05
deviance	1.06e+04	3.75e+00	5.94e−02	1.57e−01

- **Poisson distribution**

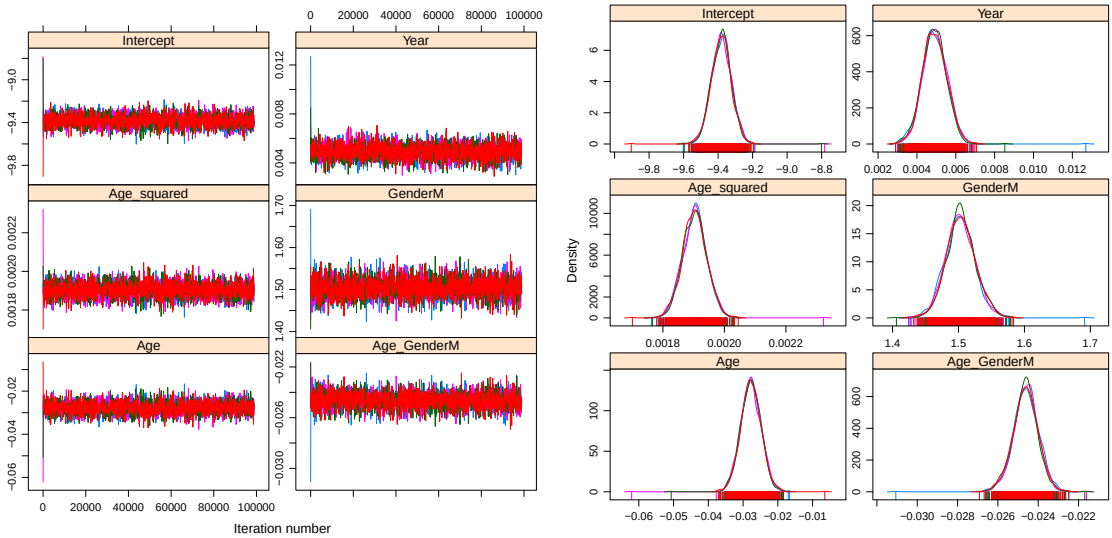
Listing 4.6: Poisson mcmc model

	Mean	SD	Naive SE	Time-series SE
Age	−3.24e−02	1.81e−03	2.86e−05	2.86e−05
Age_GenderM	−2.07e−02	3.67e−04	5.80e−06	5.80e−06
Age_squared	1.87e−03	2.22e−05	3.51e−07	3.60e−07
GenderM	1.36e+00	1.52e−02	2.40e−04	2.45e−04
Intercept	−9.19e+00	4.09e−02	6.47e−04	6.47e−04
Year	8.28e−03	3.27e−04	5.17e−06	5.17e−06
deviance	1.06e+04	5.01e+00	7.92e−02	7.93e−02

The output listings 4.4, 4.5 and 4.6 above show the summary results of the generalized linear model using MCMC. Carefully considering each parameter values in the table, it is approximately similar to those obtain from the classical approach.

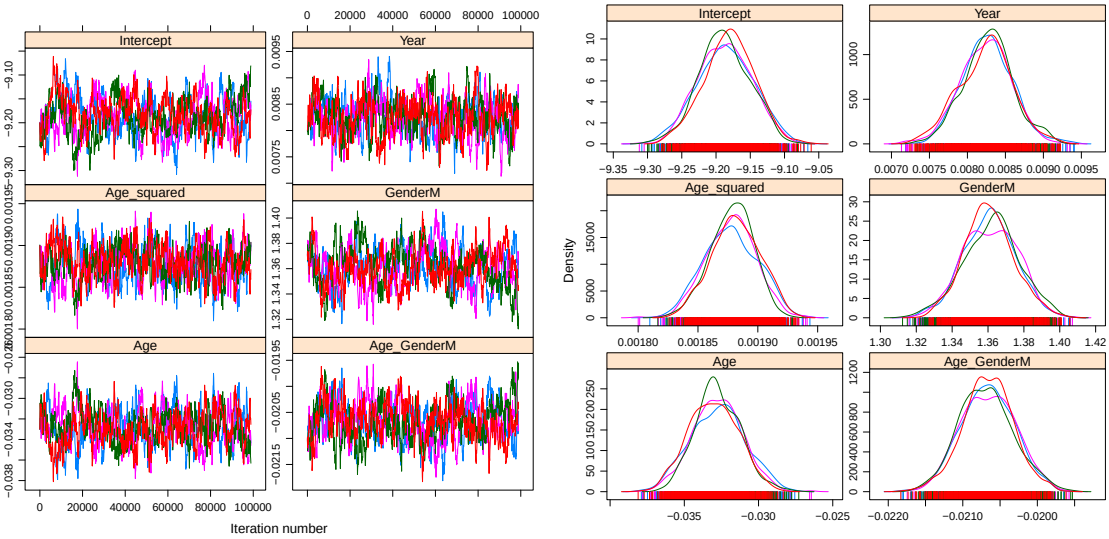
Figures 4.4, 4.5 and 4.6 below depict the trace plots (on the left side) and density plots (on the right side) of the model parameters. The trace plots shows the values the parameter took during the runtime of the MCMC chains. The density plots, basically, are the (smoothened) histogram of the values in the trace-plots, i.e., the distribution of the values of the parameter in the chains.

Figures 4.4a, 4.5a and 4.6a display "perfect" trace plots. Note that the center of the chains appear to be around the parameter value, with very small fluctuations. This indicates that the chains are mixing well and could have reached the stationary distribution.



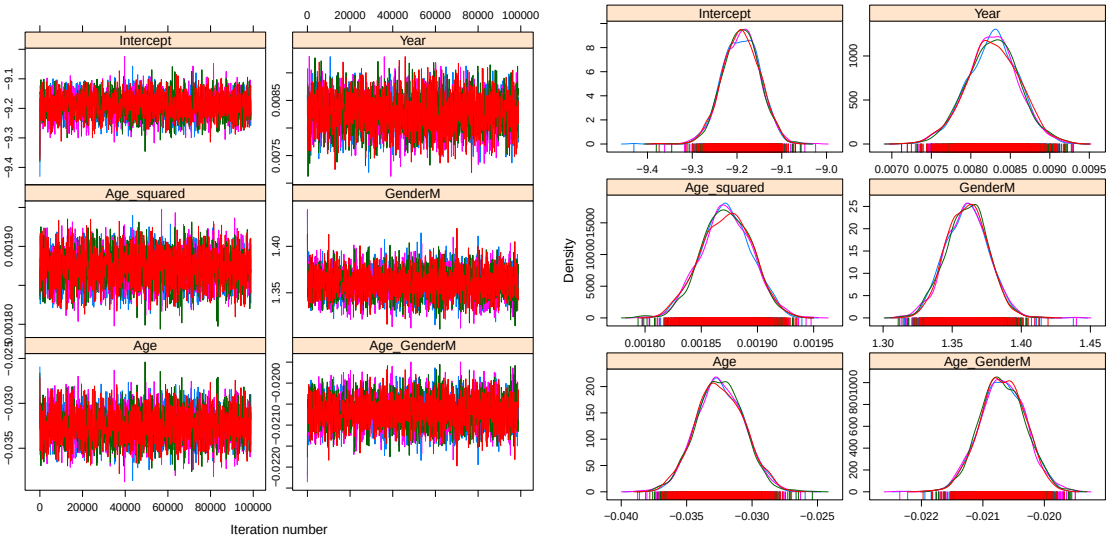
(a) trace plot (b) density plot

Figure 4.4: MCMC plots of the Normal distribution



(a) Trace plot (b) Density plot

Figure 4.5: MCMC plots of the Binomial distribution



(a) Trace plot (b) Density plot

Figure 4.6: MCMC plots of the Poisson distribution

5. Conclusion

The aim of this work was to investigate Generalized Linear Models (GLM) with application to mortality data. We started by considering the classical approach and then the Bayesian approach using MCMC. In each case, the Normal, Binomial and Poisson distributions were considered and models were fitted using aggregate mortality data.

Results showed that model parameters' values are approximately the same for the classical and Bayesian approaches. Also, the best model fit, based on the criteria considered is the Poisson distribution model. The Binomial model follows as a better fit to the data.

Implementing MCMC allows specifying some initial values for the model parameters. In this study, parameters values are initialized to zero except for the Binomial case. Future study might include investigation of the influence of the initials in model fitting. We might as well consider approach based criteria, whether to model our data using the classical or Bayesian. This might produce parameter values with good precision and helps greatly in statistical predictions.

Reference

A. Agresti. *Foundations of Linear and Generalized Linear Models*. Wiley Series in Probability and Statistics. Wiley, 1 edition, 2015. ISBN 1118730038,9781118730034. URL http://gen.lib.rus.ec/book/index.php?md5=EDE8D9717B9942ABC676AEAD386381E4.

H. Akaike. A bayesian analysis of the minimum aic procedure. *Annals of the Institute of Statistical Mathematics*, 30, 1978. doi: 10.1007/bf02480194. URL http://gen.lib.rus.ec/scimag/index.php?s=10.1007/bf02480194.

A. Charnes, E. L. Frome, and P. L. Yu. The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. *Journal of the American Statistical Association*, 71, 03 1976. doi: 10.2307/2285762. URL http://gen.lib.rus.ec/scimag/index.php?s=10.2307/2285762.

M. K. Cowles. *Applied Bayesian Statistics: With R and OpenBUGS Examples*. Springer Texts in Statistics 98. Springer-Verlag New York, 1 edition, 2013. ISBN 978-1-4614-5695-7,978-1-4614-5696-4. URL http://gen.lib.rus.ec/book/index.php?md5=4CF7C5AC3922437BB63D7137AB1B3785.

A. J. Dobson and A. Barnett. *An Introduction to Generalized Linear Models*. Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC, 3 edition, 2008. ISBN 1584889500,9781584889502. URL http://gen.lib.rus.ec/book/index.php?md5=472B57FA461867F6CFB4334BFED60010.

G. H. Dunteman and Dr. Moon-Ho R. Ho. *An Introduction to Generalized Linear Models (Quantitative Applications in the Social Sciences)*. Quantitative Applications in the Social Sciences 145. Sage Publications, Inc, 1 edition, 2005. ISBN 0761920846,9780761920847. URL http://gen.lib.rus.ec/book/index.php?md5=B22E2AE0EF4EAD514407DC5DFB07FCFD.

V. J. Easton and J. H. McColl. Statistics glossary v1.1, 1997.

Faraway and J. James. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Texts in statistical science. CRC Press, second edition edition, 2016. ISBN 978-1-4987-2098-4,103-103-108-1,1498720986. URL http://gen.lib.rus.ec/book/index.php?md5=9A3B5AA44BE0F6AEAC4319E9624180E7.

S. Filippi. Hierarchical models - statistical methods, 2015. Unpublished manuscript.

D. Gamerman. *Markov chain Monte Carlo : stochastic simulation for Bayesian inference*. Texts in statistical science. Chapman and Hall, 2nd ed edition, 2006. ISBN 1584885874,9781584885870. URL http://gen.lib.rus.ec/book/index.php?md5=84B08D7DB8E96EFFB1EFDA564161C2BB.

W.R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC Interdisciplinary Statistics Series. Chapman and Hall, 1 edition, 1995. ISBN 0412055511,9780412055515. URL http://gen.lib.rus.ec/book/index.php?md5=4D623F3301BA8E3F27F8F32454FC80E8.

M. Kery. *Introduction to WinBUGS for Ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses*. 1 edition, 2010. ISBN 0123786053,9780123786050. URL http://gen.lib.rus.ec/book/index.php?md5=E53AB3BEAE8CDC157A19C683E89DD59A.

D. P. Kroese, T. Taimre, and Z. I. Botev. *Handbook of Monte Carlo Methods (Wiley Series in Probability and Statistics)*. Wiley Series in Probability and Statistics 706. Wiley, 1 edition, 2011. ISBN 0470177934,9780470177938. URL http://gen.lib.rus.ec/book/index.php?md5=3818E025500D51FAEA6F269706322A46.

F. Liang, C. Liu, and R. Carroll. *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples (Wiley Series in Computational Statistics)*. Wiley, 1 edition, 2010. ISBN 0470748265,9780470748268,9780470669730,047066973X. URL http://gen.lib.rus.ec/book/index.php?md5=D79E8C161A8E6C296AF4D6A227E49B88.

W. A. Link, E. Cam, J. D. Nichols, and E. G. Cooch. Of bugs and birds: Markov chain monte carlo for hierarchical modeling in wildlife research. *Journal of Wildlife Management*, 66, 04 2002. doi: 10.2307/3803160. URL http://gen.lib.rus.ec/scimag/index.php?s=10.2307/3803160.

D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The bugs project: Evolution, critique and future directions. *Statistics in Medicine*, 28, 2009. doi: 10.1002/sim.3680. URL http://gen.lib.rus.ec/scimag/index.php?s=10.1002/sim.3680.

M. A. McCARTHY and PIP MASTERS. Profiting from prior information in bayesian analyses of ecological data. *Journal of Applied Ecology*, 42, 2005. doi: 10.1111/j.1365-2664.2005.01101.x. URL http://gen.lib.rus.ec/scimag/index.php?s=10.1111/j.1365-2664.2005.01101.x.

P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Monographs on Statistics and Applied Probability. Springer US, 1989. ISBN 978-0-412-31760-6,978-1-4899-3242-6. URL http://gen.lib.rus.ec/book/index.php?md5=A204C384673F55C1BB7485B74A657FA1.

D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to Linear Regression Analysis*. Wiley, 5 edition, 2012. ISBN 0470542810,9780470542811. URL http://gen.lib.rus.ec/book/index.php?md5=801CF5D2E242A49529EC50A61542FD54.

I. Ntzoufras. *Bayesian Modeling Using WinBUGS (Wiley Series in Computational Statistics)*. 1 edition, 2009. ISBN 047014114X,9780470141144,9780470434543. URL http://gen.lib.rus.ec/book/index.php?md5=3882F2874AB29401B16FE9DF7D6E886D.

U. Olsson. *Generalized Linear Models: An Applied Approach*. Studentlitteratur AB, 2002. ISBN 9144041551,9789144041551. URL http://gen.lib.rus.ec/book/index.php?md5=C9B880B63251C496D5820A4EB75ED67F.

R. A. Pielke. The role of models in prediction for decision. 2003.

M. Plummer. Jags version 4.0.0 user manual, 2015.

V. C. Ramuada. Bayesian estimation of the garch(1,1) stochastic volatility model, 2016.

SAS-Publishing. *SAS STAT 9.2 User's Guide*. Second edition edition, 2009. ISBN 1607642476,9781607642473. URL http://gen.lib.rus.ec/book/index.php?md5=21146D880205B7E6DBD0F38CCD786E6B.

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6, 03 1978. doi: 10.2307/2958889. URL http://gen.lib.rus.ec/scimag/index.php?s=10.2307/2958889.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 2002. doi: 10.1111/1467-9868.00353. URL http://gen.lib.rus.ec/scimag/index.php?s=10.1111/1467-9868.00353.

Stigler and M. Stephen. Gauss and the invention of least squares. *The Annals of Statistics*, 9, 05 1981. doi: 10.1214/aos/1176345451. URL http://gen.lib.rus.ec/scimag/index.php?s=10.1214/aos/1176345451.

A. van der Linde. Dic in variable selection. *Statistica Neerlandica*, 59(1):45–56, 2005. ISSN 1467-9574. doi: 10.1111/j.1467-9574.2005.00278.x. URL http://dx.doi.org/10.1111/j.1467-9574.2005.00278.x.