Derivation of BIWLS versus regular MH to compare convergence rate

Chase Joyner

March 28, 2017

We assume the binary model

$$Y_i \sim \text{Bern}(p_i)$$

for $i = 1, ..., n$. The logit link is used and so we have the relation

$$\log \frac{p_i}{1-p_i} = \mathbf{x}_i' \boldsymbol{\beta}.$$

First notice that there is a one-to-one correspondence between p_i and $\boldsymbol{\beta}$ and so for the sake of notation, we will write p_i throughout for the likelihood term. Also, we can write the data distribution as

$$\begin{aligned} f(y_i|\theta_i) &= p_i^{y_i}(1-p_i)^{1-y_i} = (1-p_i)\left(\frac{p_i}{1-p_i}\right)^{y_i} = \exp\left\{ y_i \log \frac{p_i}{1-p_i} + \log(1-p_i) \right\} \\ &= \exp\left\{ y_i \log \frac{p_i}{1-p_i} - \log \frac{1}{1-p_i} \right\} \end{aligned}$$

and thus this is a member of the exponential family with

$$\theta_i = \log \frac{p_i}{1-p_i}, \qquad b(\theta_i) = \log\left(1 + e^{\theta_i}\right).$$

Now, we assume that $\boldsymbol{\beta} \sim N(\mathbf{a}, \mathbf{R})$ and so the posterior distribution is

$$f(\boldsymbol{\beta}|\mathbf{y}) \propto \exp\left\{ -\frac{1}{2}(\boldsymbol{\beta}-\mathbf{a})'\mathbf{R}^{-1}(\boldsymbol{\beta}-\mathbf{a}) + \sum_{i=1}^{n} \left(y_i \log \frac{p_i}{1-p_i} - \log \frac{1}{1-p_i} \right) \right\}.$$

We will consider the Metropolis-Hastings algorithm, but with two different normal proposal distributions. The first is a naive approach where we use a normal distribution centered around the previous value of $\boldsymbol{\beta}$ with a tuning covariance matrix as the proposal distribution and the second is the BIWLS method. We derive the necessary pieces for the BIWLS method below. First, notice that $\mu_i = p_i$ for all i and so

$$g'(p_i) = \frac{d}{dp_i} \log \frac{p_i}{1 - p_i} = \frac{1}{p_i(1 - p_i)}.$$

Therefore, the transformed observations has components

$$\widetilde{y}_i(\boldsymbol{\beta}) = \mathbf{x}_i'\boldsymbol{\beta} + \frac{y_i - p_i}{p_i(1 - p_i)}.$$

Lastly, the second derivative of $b(\theta_i)$ wrt θ_i is

$$b''(\theta_i) = \frac{e^{\theta_i}}{(1 + e^{\theta_i})^2} = p_i(1 - p_i)$$

and therefore the inverse of the diagonal weight matrix has entries

$$W_{ii}^{-1}(\boldsymbol{\beta}) = p_i(1 - p_i) \cdot \frac{1}{\big(p_i(1 - p_i)\big)^2} = \frac{1}{p_i(1 - p_i)}.$$

Thus, the diagonal weight matrix has entries given by

$$W_{ii}(\boldsymbol{\beta}) = p_i(1 - p_i).$$

Now, we use the normal distribution with $\mathbf{m}$ and $\mathbf{C}$ given in the Gamerman paper.