MDA Assignment No. 2

Aditya Kunar(5074274)

December 18 2019

Likelihood

Consider the Poisson regression model which is a basic model for count-data. So we assume data $\{(x_i, y_i)\}_{i=1}^n$, where $y_i \in \{0, 1, 2, ...\}$ and $x_i \in \mathbb{R}^p$. The model is given by $y_i \sim Pois(\mu_i)$, where $log\mu_i = x_i^T\theta$ for an unknown parameter vector $\theta \in \mathbb{R}^p$. Hence, for $k \in \{0, 1, 2, ...\}$ we have $\mathbb{P}(Y_i = k) = \frac{e^{-\mu_i}\mu_i^k}{(k!)}$.

(1) Give the loglikelihood, assuming all y_i are independent.

The likelihood is defined as follows -:

$L(\theta|Y) = \mathbb{P}(Y|\theta) = \prod_{i=1}^n \mathbb{P}(y_i|\theta)$

$\mathbb{P}(Y|\theta) = \prod_{i=1}^n \frac{e^{-\mu_i}\mu_i^{y_i}}{(y_i!)}$

$log\mathbb{P}(Y|\theta) = log\prod_{i=1}^n \frac{e^{-\mu_i}\mu_i^{y_i}}{(y_i!)}$

$= \sum_{i=1}^n log\frac{e^{-\mu_i}\mu_i^{y_i}}{(y_i!)}$

$= \sum_{i=1}^n y_i log\mu_i - \mu_i - log(y_i!)$

Gradient

(2) Derive an expression for the gradient and hessian of the loglikelihood

The gradient is defined as follows -:

$LogL(\theta|Y) = \sum_{i=1}^n y_i log\mu_i - \mu_i - log(y_i!)$

As $log\mu_i = x_i^T\theta$

$\therefore LogL(\theta|Y) = \sum_{i=1}^n y_i x_i^T\theta - e^{x_i^T\theta} - log(y_i!)$

$\nabla LogL(\theta|Y) = \frac{\partial}{\partial\theta}\sum_{i=1}^n y_i x_i^T\theta - e^{x_i^T\theta} - log(y_i!)$

$= \sum_{i=1}^{n} y_i x_i - x_i e^{x_i^T \theta}$

Hessian

The hessian is defined as follows -:

$\frac{\partial^2 logL(\theta|Y)}{\partial\theta\partial\theta^T} = \frac{\partial}{\partial\theta^T} \sum_{i=1}^{n} y_i x_i - x_i e^{x_i^T \theta}$

$= \sum_{i=1}^{n} -x_i x_i^T e^{x_i^T \theta}$

Laplace Approximation

(3) In the following we consider the dataset dataexercise2.csv. We take a Bayesian point of view, where we assume $y_i|\theta \overset{ind}{\sim} Pois(\mu_i)$, where $log\mu_i = x_i^T\theta$, $\theta \sim N(0, \tilde{\sigma}^2)$. Assume the prior standard deviation is given by $\tilde{\sigma} = 5$. Implement a Newton algorithm for computing the Laplace approximation to the posterior distribution. Report mean and covariance matrix of the approximation.

The mean and covariance for the Laplace Approximation are as follows -:

$$mean = \begin{bmatrix} 1.12649475 & 0.42890301 & 0.01506197 & -0.05408826 \end{bmatrix}$$

$$Covariance = \begin{bmatrix} 0.0313 & -0.0081 & 0.0011 & -0.0013 \\ -0.0081 & 0.0030 & -0.0003 & 0.0006 \\ 0.0011 & -0.0003 & 0.01480 & -0.0014 \\ -0.0013 & 0.0006 & -0.0014 & 0.0117 \end{bmatrix}$$

M-H Algorithm

(4) Implement a random-walk Metropolis Hastings algorithm to sample from the posterior. Take proposals of the form $\theta^\circ := \theta + \sigma_{proposal} N(0, I_p)$. Tune $\sigma_{proposal}$ to achieve an acceptance rate of about $25\% - 50\%$. Make a plot of the iterates where you plot $\theta 2$ versus $\theta 1$, with colour indicating the iteration number. Report the Monte-Carlo estimate of the posterior mean (where you "throw away" burnin samples, i.e. initial samples where the chain has not reached its stationary region).

The tuned value of $\sigma_{proposal} = 0.067$
Monte-Carlo estimate of the posterior mean: $\begin{bmatrix} 1.07563637 & 0.43882114 & 0.01468655 & -0.0550479 \end{bmatrix}$

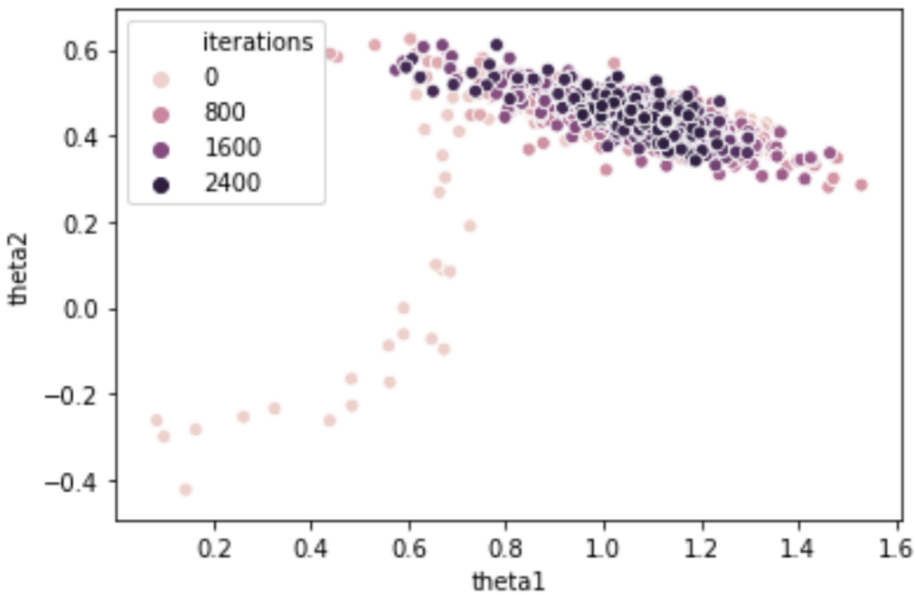

Figure 1: Scatter Plot of $\theta 2$ vs. $\theta 1$ ($\sigma_{proposal} = 0.067$).

Gibbs Sampling

(5) The results may be sensitive to the choice of $\tilde{\sigma}$. For that reason we add an extra layer to the hierarchical model in the following way:
$y_i|\theta \overset{ind}{\sim} Pois(\mu_i)$, where $log\mu_i = x_i^T\theta$,
$\theta|\tilde{\sigma} \sim N(0, \tilde{\sigma}^2 I_p)$,
$\tilde{\sigma}^2 \sim IG(\alpha, \beta)$
Here $IG(\alpha, \beta)$ denotes the inverse Gamma distribution with parameters α and β (its density function is given in exercise 3.12 in RG). Take $\alpha = \beta = 0.1$. Implement a Gibbs sampler that iteratively samples from the full conditionals of θ and $\tilde{\sigma}$. Include a derivation for the update-step for $\tilde{\sigma}^2$ in your report. Also include a traceplot of the posterior samples of $\tilde{\sigma}^2$ (a traceplot is a plot of iterate value versus iterate number).

The derivation for the update step for $\tilde{\sigma}^2$ is as follows-:

$\mathbb{P}(\tilde{\sigma}^2|Y, \theta) \propto \mathbb{P}(\theta|\tilde{\sigma}^2)\mathbb{P}(\tilde{\sigma}^2|\alpha, \beta)$

And

$\theta|\tilde{\sigma} \sim N(0, \tilde{\sigma}^2 I_p)$,

$\tilde{\sigma}^2 \sim IG(\alpha, \beta)$.

$IG(\alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\beta}{x}}$

$\therefore \mathbb{P}(\tilde{\sigma}^2|Y, \theta) \propto \left((2\pi)^{-\frac{k}{2}} \det(\tilde{\sigma}^2 I_p)^{-\frac{1}{2}} e^{(-\frac{1}{2}\theta\Sigma^{-1}\theta^T)}\right)\left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}(\tilde{\sigma}^2)^{-\alpha-1} e^{-(\frac{\beta}{\tilde{\sigma}^2})}\right)$

$\mathbb{P}(\tilde{\sigma}^2|Y, \theta) \propto \tilde{\sigma}^{-p}(\tilde{\sigma}^2)^{-\alpha-1} e^{(-\frac{1}{\tilde{\sigma}^2}(\frac{\theta\theta^T}{2}+\beta))}$

$\mathbb{P}(\tilde{\sigma}^2|Y, \theta) \propto (\tilde{\sigma}^2)^{-(\alpha+\frac{p}{2})-1} e^{(-\frac{1}{\tilde{\sigma}^2}(\frac{\theta\theta^T}{2}+\beta))}$

Therefore we see that this results in a new inv-gamma distribution with the following parameters -:

$\mathbb{P}(\tilde{\sigma}^2|Y, \theta) \sim IG(\alpha^{\circ}, \beta^{\circ})$ where $\alpha^{\circ} = \alpha + \frac{p}{2}$ and $\beta^{\circ} = \beta + \frac{\theta\theta^T}{2}$

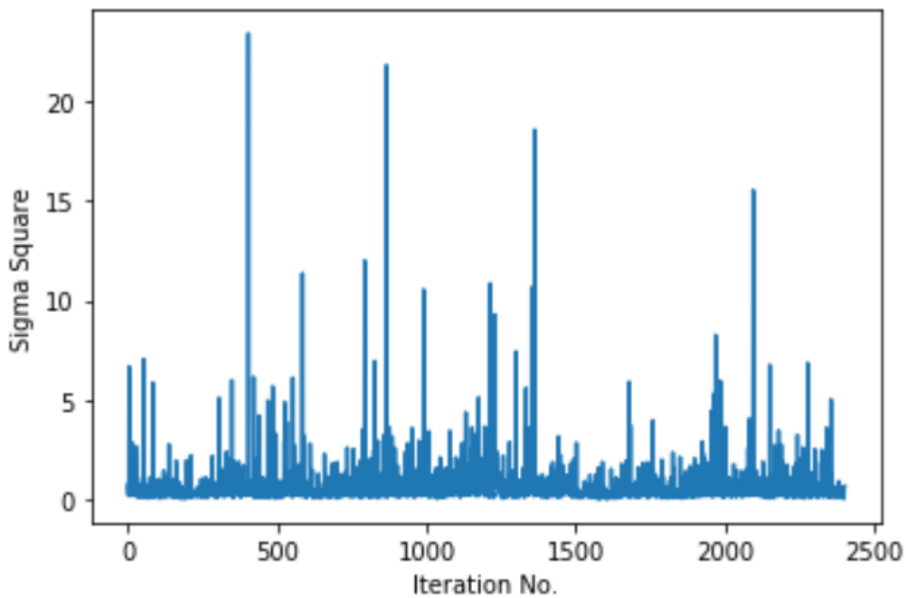

Figure 2: Trace Plot of $\tilde{\sigma}^2$