Homework 9

Problem 1

Consider the Bernoulli regression model,

$$P(y_i = 1) = p_i = \frac{e^{x_i \beta}}{1 + e^{x_i \beta}}, \quad y_i \in \{0, 1\}, \quad i = 1, ..., n,$$

with β a one dimensional unknown parameter. The log–likelihood function is given by

$$L(\beta) = \beta \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \log\left(1 + e^{x_i \beta}\right).$$

(1) (10 pts) By finding $dL/d\beta$ and $d^2L/d\beta^2$, show that the Newton–Raphson algorithm for finding the maximum likelihood estimator $\widehat{\beta}$ is given by

$$\beta^{(t+1)} = \beta^{(t)} + \frac{\sum_{i=1}^{n} x_i (y_i - p_i^{(t)})}{\sum_{i=1}^{n} x_i^2 p_i^{(t)} (1 - p_i^{(t)})},$$

where

$$p_i^{(t)} = \frac{e^{x_i \beta^{(t)}}}{1 + e^{x_i \beta^{(t)}}}.$$

(2) (10 pts) If $n = 10$ and W is the $n \times n$ diagonal matrix with ith element $\widehat{p}_i(1 - \widehat{p}_i)$, where $\widehat{p}_i$ is p_i estimated at $\widehat{\beta}$, and assuming that approximately

$$\widehat{\beta} = N\left(\beta, (X'WX)^{-1}\right),$$

where X is the $n \times 1$ vector of predictor variables ($x_i = i/10$), find the approximate variance of $\widehat{\beta}$ when its observed value is $\widehat{\beta} = -0.34$.

(3) (10 pts) What is the value of the test statistic for testing the hypothesis $\beta = 0$.

(4) (10 pts) What is the outcome of the test if the level of significance is chosen to be 0.1.

(5) (10 pts) Write down an expresson for the deviance of the model and what is the approximate distribution of it if the model is correct.

Problem 2

The Bernoulli regression model, as given in Problem 1, is analyzed using a Bayesian approach and the prior for β, i.e. $\pi(\beta)$, is chosen to be normal with mean 0 and variance σ^2.

(1) (10 pts) Write down the posterior density (proportional to) for β in terms of the (x_i, y_i).

(2) (10 pts) A way to sample from a density directly is available if the logarithm of the density is concave. Show that the log of the posterior density is concave.

(3) (10 pts) If the prior is taken to be improper, i.e. $\pi(\beta) = 1$, what is the relationship between the mode of the posterior and the MLE estimator.

(4) (10 pts) With this improper prior, the posterior is sampled using a Metropolis algorithm with proposal density $N(\beta^* \mid \beta_0, v^2)$, where β_0 is the current value of β in the chain. If β^* has been sampled from the proposal, what is the probability that $\beta_1 = \beta^\star$; i.e. accept $\beta^\star$ as the next value of the chain.

(5) (10 pts) Explain the problem with running a chain with v too small and v too big.