Homework 9

Rafael Espinosa Castañeda

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

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

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

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} \frac{e^{x_i\beta} x_i}{1+e^{x_i\beta}}$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} p_i^{(t)} x_i$$

$$\boxed{\frac{\partial L}{\partial \beta} = \sum_{i=1}^{N} x_i (y_i - p_i^{(t)})}$$

$$\frac{\partial^2 L}{\partial \beta^2} = -\sum_{i=1}^{N} \frac{\left[e^{x_i\beta} x_i^2 (1+e^{x_i\beta}) - e^{2x_i\beta} x_i^2\right.}{(1+e^{x_i\beta})^2}$$

$$= -\sum_{i=1}^{N} \frac{e^{x_i\beta} x_i^2}{(1+e^{x_i\beta})} - \frac{e^{2x_i\beta} x_i^2}{(1+e^{x_i\beta})^2}$$

$$\frac{\partial^2 L}{\partial \beta^2} = -\sum_{i=1}^{N} p_i(t) x_i^2 - p_i(t)^2 x_i^2$$

$$\boxed{\frac{\partial^2 L}{\partial \beta^2} = -\sum_{i=1}^{N} p_i(t)\, x_j^2 (1 - p_i(t))}$$

Hence,

$$\beta^{(t+1)} = \beta^{(t)} - \frac{\partial L}{\partial \beta} \Big/ \frac{\partial^2 L}{\partial \beta^2} = \beta^{(t)} + \frac{\sum_{i=1}^{N} x_i (y_i - p_i^{(t)})}{\sum_{i=1}^{N} p_i(t)\, x_j^2 (1 - p_i(t))}$$

(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 $\widehat{\beta} = -0.34$.

code:

```
x<-0
W<-matrix(0,nrow=10,ncol=10)
p_i<-0
beta<- -0.34
for (i in 1:10)
  {
  x[i]<-i/10
  p_i<- exp(x[i]*beta)/(1+exp(x[i]*beta))
  W[i,i]<- p_i
  }
 print(x)
 variance<-solve(t(x)%*%W%*%x)

 print(variance)
```

Variance = 0.5989221

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

We use a z-statistic

$$z = \frac{\hat{\beta} - \beta}{\sqrt{Var(\hat{\beta})}} = \frac{-0.34 - 0}{\sqrt{0.5989221}}$$

$$z = -0.43933292$$

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

Given that the p-value = 0.6604203

And p-value > α we accept the null hypothesis.

Hence, we have sufficient evidence to believe that $\beta = 0$.

(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.

The deviance will be given by

$$D(\hat{\beta}) = -2\,\ell(\hat{\beta})$$

$$= -2\left[\hat{\beta}\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} \log\left(1+e^{x_i\hat{\beta}}\right)\right]$$

If the model is correct it will follow aproximately a χ^2 distribution.

Problem 2

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).

$$\Pi_n \propto \Pi(\text{data}|\beta)\,\Pi(\beta)$$

$\propto$ proportional

where $\Pi(\beta) = N(0, \sigma^2)$

$$\Pi(\beta) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-y_i^2/2\sigma^2}$$

and

$\Pi(\text{data}|\beta) =$ Likelihood function

Let's denote as $l =$ Likelihood function

$\Rightarrow$

$$\underbrace{L(\beta)}_{\text{Log-likelihood}} = \ln(l) \Rightarrow$$

$$l = e^{L(\beta)}$$

where $L(\beta) = \beta \sum x_i y_i - \sum \log(1+e^{x_i\beta})$

$$\Rightarrow l = e^{\beta \sum x_i y_i - \sum \log(1+e^{x_i\beta})}$$

Hence, the posterior is given by

$$\pi_n(\beta) = \frac{1}{\sqrt{2\pi}\sigma} e^{\sum y_i^2/2\sigma^2} e^{\beta \sum x_i y_i - \sum \log(1+e^{x_i\beta})}$$

(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.

$$\ln(\pi_n(\beta)) =$$

$$\ln\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{\sum y_i^2}{2\sigma^2} + \beta \sum x_i y_i$$

$$- \sum_{i=1}^{N} \ln\left(1 + e^{x_i \beta}\right)$$

So,

$$\frac{\partial}{\partial \beta} \ln(\pi_n(\beta)) = \sum x_i y_i - \sum_{i=1}^{N} \frac{e^{x_i\beta} x_i}{1 + e^{x_i \beta}}$$

$$\frac{\partial^2 \ln(\pi_n(\beta))}{\partial \beta^2} =$$

$$= - \sum_{i=1}^{N} \frac{(e^{x_i\beta} x_i^2)(1 + e^{x_i\beta}) + e^{2x_i\beta} x_i^2}{(1 + e^{x_i\beta})^2}$$

$$= -\left[\underbrace{\sum_{i=1}^{N} \frac{e^{x_i\beta} x_i^2}{(1+e^{x_i\beta})} + \frac{e^{2x_i\beta}}{(1+e^{x_i\beta})^2} x_i^2}_{a} \right]$$

$a > 0$, we have that all terms in a are positive. Hence the sum of all them are positive. By the minus sign outside the sum, we obtain a negative second derivative

$\frac{\partial^2}{\partial \beta^2} \ln(\Pi_n(\beta)) < 0$. Hence, the function 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.

IF $\pi(\beta) = 1$

$$\pi_n(\beta) = e^{\beta \sum x_i y_i - \sum \log(1 + e^{x_i \beta})}$$

$\Rightarrow$

$$\frac{\partial \pi_n(\beta)}{\partial \beta} = e^{\beta \sum x_i y_i - \sum \log(1+e^{x_i\beta})} \left(\sum_{i=1}^{n} x_i y_i - \frac{e^{x_i\beta} x_i}{1+e^{x_i\beta}} \right) = 0$$

$$\Rightarrow \boxed{\sum x_i y_i = \sum \left(e^{x_i\beta} x_i / 1 + e^{x_i\beta} \right)}$$

At the same time the MLE estimator is the $\hat{\beta}$ that extremizes the log-likelihood function, which will satisfy

$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow$$

$$\frac{\partial L}{\partial \beta} = \frac{\partial}{\partial \beta}\left(\beta \sum x_i y_i - \sum \log\left(1 + e^{x_i\beta}\right) \right)$$

$$= \sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} \frac{e^{x_i \beta} x_i}{1+e^{x_i \beta}} = 0 \Rightarrow$$

$$\boxed{\sum x_i y_i = \sum_{i=1}^{N} \frac{e^{x_i \beta} x_i}{1+e^{x_i \beta}}}$$

So, the mode and the MLE will satisfy the same equation and will be the same value.

(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.

$$\alpha = \min\left\{ 1, \frac{\pi_n(\beta^*)}{\pi_n(\beta^0)} \right\}$$

where

$$\pi_n(\beta^*) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\sum y_i^2/2\sigma^2} e^{\beta^* \sum x_i y_i - \sum \log(1+e^{x_i \beta^*})}$$

$$\pi_n(\beta^0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\sum y_i^2/2\sigma^2} e^{\beta^0 \sum x_i y_i - \sum \log(1+e^{x_i \beta^0})}$$

5) if v too large: The proposal will be too far from the current value and the posterior value could be small. So, the chain will not move much

if v too small: the chain can move easily as the posterior values are similar. However, the chain will not move far and the correlation between samples will be high.