Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

LØSNINGSFORSLAG
EXAM IN TMA4315 GENERALIZED LINEAR MODELS
Monday 30 November 2009
Time: 09:00–13:00

Oppgave 1 N ($N > 100$) independent observations $Y_1, ..., Y_N$ have normal distributions with the same variance σ^2.

$$EY_1 = \mu_1 = \beta_1 + \beta_2,$$
$$EY_2 = EY_3 = ... = EY_N = \mu = \beta_1,$$

where β_1 and β_2 are parameters of interest.

a) What is the design matrix X?

Solution. X is a $N \times 2$ matrix of the form

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ & ... \\ 1 & 0 \end{bmatrix}$$

b) Using the simple rule of thumb $h_{ii} > 2p/N$ show that the first observation (and only it) is highly influential (here $H = [h_{ij}]$ is the hat matrix, and p is the number of parameters, $p = 2$).

Solution.

$$X^TX = \begin{bmatrix} N & 1 \\ 1 & 1 \end{bmatrix}, \ (X^TX)^{-1} = \frac{1}{N-1}\begin{bmatrix} 1 & -1 \\ -1 & N \end{bmatrix}$$

and therefore the hat matrix is

$$H = X(X^TX)^{-1}X^T = \begin{bmatrix} 1 & 0 & ... & 0 \\ 0 & \frac{1}{N-1} & ... & \frac{1}{N-1} \\ . & . & . & . \\ 0 & \frac{1}{N-1} & ... & \frac{1}{N-1} \end{bmatrix}.$$

Thus

$$h_{11} > \frac{2p}{N}$$

and

$$h_{ii} < \frac{2p}{N}, \quad i = 2, ..., N$$

(here $p = 2$ and $N > 100$).

c) Find the maximum likelihood estimator of $\beta = (\beta_1, \beta_2)^T$.

Solution. It is easy to see that

$$(X^T X)^{-1} X^T = \frac{1}{N-1} \begin{bmatrix} 0 & 1 & 1 & \dots & 1 \\ N-1 & -1 & -1 & \dots & -1 \end{bmatrix},$$

therefore MLE is

$$b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = (X^T X)^{-1} X^T Y = \begin{bmatrix} \frac{1}{N-1} \sum_{i=2}^{N} Y_i \\ Y_1 - \frac{1}{N-1} \sum_{i=2}^{N} Y_i \end{bmatrix}.$$

Oppgave 2 X_1 and X_2 are two independent random variables having the same distribution with the probability density function

$$f_X(x; \theta) = \theta x^{\theta - 1} I_{(0,1)}(x), \ \theta > 0.$$

Let $Y = \max\{X_1, X_2\}$.

a) Does the distribution of Y belong to the exponential family?

Solution.

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0, \\ x^\theta & \text{for } 0 \le x \le 1, \\ 1 & \text{for } x > 1, \end{cases}$$

therefore

$$F_Y(y) = [F_X(y)]^2 = \begin{cases} 0 & \text{for } y < 0, \\ y^{2\theta} & \text{for } 0 \le y \le 1, \\ 1 & \text{for } y > 1, \end{cases}$$

$$f_Y(y) = 2\theta y^{2\theta - 1} I_{(0,1)}(y) = e^{2\theta \ln y + \ln \theta + \ln 2 - \ln y} \text{ for } 0 < x < 1.$$

The distribution of Y belongs to the exponential family:

$$a(y) = \ln y, \ b(\theta) = 2\theta, \ c(\theta) = \ln \theta, \ d(y) = \ln 2 - \ln y.$$

b) Show that

$$E \ln Y = -\frac{1}{2\theta}.$$

Solution.

$$E \ln Y = Ea(Y) = -\frac{c'(\theta)}{b'(\theta)} = -\frac{1}{2\theta}.$$

Oppgave 3 Consider the following GLM: $Y_1, ..., Y_N$ are independent, $N = 2n$; $Y_i \sim$binomial(n_i, π_i), i=1,...,N (n_i are known); $\beta = (\beta_1, \beta_2)^T$ are parameters of interest;

$$\ln \frac{\pi_i}{1-\pi_i} = x_i^T \beta,$$

where the design matrix has form

$$X = \begin{bmatrix} a_1 & & 0 \\ & \cdots & \\ a_n & & 0 \\ 0 & & c_1 \\ & \cdots & \\ 0 & & c_n \end{bmatrix}.$$

a) Find the score vector (in terms of Y_i, n_i, π_i, a_i, c_i) and show that b_1 does not depend of $Y_{n+1}, ..., Y_N$ while b_2 does not depend of $Y_1, ..., Y_n$, where $b = (b_1, b_2)^T$ is the maximum likelihood estimator of $\beta = (\beta_1, \beta_2)^T$.

Solution. The log-likelihood function is

$$l(\beta; Y) = \sum_{i=1}^{N} \left(Y_i \ln \frac{\pi_i}{1-\pi_i} + n_i \ln(1-\pi_i) + \ln \begin{pmatrix} n_i \\ Y_i \end{pmatrix} \right) =$$

$$= \sum_{i=1}^{N} \left(Y_i x_i^T \beta - n_i \ln(1 + e^{x_i^T \beta}) + \ln \begin{pmatrix} n_i \\ Y_i \end{pmatrix} \right).$$

U_1 and U_2 are obtained by differentiation of $l(\beta; Y)$ with respect to β_1 and β_2 respectively. Taking into account that

$$\pi_i = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}}$$

we obtain

$$U = (U_1, U_2)^T = \left(\sum_{i=1}^{n} (Y_i - n_i \pi_i) a_i, \sum_{i=n+1}^{N} (Y_i - n_i \pi_i) c_{i-n} \right)^T.$$

Since $\pi_1, ..., \pi_n$ depend only of β_1 and $\pi_{n+1}, ..., \pi_N$ depend only of β_2, so do U_1 and U_2 respectively. Equations $U_1(\beta_1) = 0$, $U_2(\beta_2) = 0$ are solved separately.

b) Find the information matrix and show that the method of scoring for b has the form

$$b_1^{(m)} = b_1^{(m-1)} + \frac{\sum_{i=1}^{n}(Y_i - n_i\pi_i^{(m-1)})a_i}{\sum_{i=1}^{n} n_i\pi_i^{(m-1)}(1-\pi_i^{(m-1)})a_i^2},$$

$$b_2^{(m)} = b_2^{(m-1)} + \frac{\sum_{i=n+1}^{N}(Y_i - n_i\pi_i^{(m-1)})c_{i-n}}{\sum_{i=n+1}^{N} n_i\pi_i^{(m-1)}(1-\pi_i^{(m-1)})c_{i-n}^2}.$$

Solution. The information matrix is

$$I = \begin{bmatrix} EU_1^2 & EU_1U_2 \\ EU_1U_2 & EU_2^2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} n_i\pi_i(1-\pi_i)a_i^2 & 0 \\ 0 & \sum_{i=n+1}^{N} n_i\pi_i(1-\pi_i)c_i^2 \end{bmatrix}.$$

Thus

$$I^{-1}U = \begin{bmatrix} \frac{\sum_{i=1}^{n}(Y_i - n_i\pi_i)a_i}{\sum_{i=1}^{n} n_i\pi_i(1-\pi_i)a_i^2}, & \frac{\sum_{i=n+1}^{N}(Y_i - n_i\pi_i)c_{i-n}}{\sum_{i=n+1}^{N} n_i\pi_i(1-\pi_i)c_{i-n}^2} \end{bmatrix}^T,$$

and the result follows.