Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

LØSNINGSFORSLAG
EXAM IN TMA4315 GENERALIZED LINEAR MODELS
Monday 15 December 2008
Time: 09:00–13:00

Oppgave 1

Suppose that there are two categorical explanatory variables, sex (male or female) and handedness (right- or left-handed). Suppose that people, coming to some place, say, a shoping center, are investigated: their sex is registered and they are asked about being left- or righthanded. Let probabilities that a person, coming to the center, is MR, ML, FR, and FL (MR means male, right-handed etc.) are θ_{11}, θ_{12}, θ_{21}, and θ_{22} respectively. Denote Y_{11}, Y_{12}, Y_{21}, and Y_{22} the number of MR, ML, FR, and FL

(i) among first 1000 people,

(ii) coming during the day.

Suppose that people come independently of each other, and that the total number of people, coming during the day, has the Poisson distribution with parameter λ.

a) Find the distribution of $\mathbf{Y} = [Y_{11}, Y_{12}, Y_{21}, Y_{22}]$ in case (i) and in case (ii).

Solution. (i) Multinomial with parameters (1000, $\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}$).

(ii)

$$\mathrm{P}(Y_{11} = y_{11}, Y_{12} = y_{12}, Y_{21} = y_{21}, Y_{22} = y_{22}) = \prod_{j=1}^{2}\prod_{i=1}^{2} \frac{e^{-\lambda\theta_{ij}}(\lambda\theta_{ij})^{y_{ij}}}{y_{ij}!}, \quad y_{ij} = 0, 1, 2, ...$$

b) Suppose that design (i) is used, and results of the investigation are presented in the contingency table below. The question of interest is whether there is an association between sex and handedness. Examine this by testing the null hypothesis that the variables are independent.

	right-handed	left-handed
male	430	90
female	440	40

Solution. Denote

$$Y_{1\cdot} = Y_{11} + Y_{12},\ Y_{\cdot 1} = Y_{11} + Y_{21},\ Y_{2\cdot} = Y_{21} + Y_{22},\ Y_{\cdot 2} = Y_{12} + Y_{22},$$

$$n = 1000,$$

$$\hat{\theta}_{i\cdot} = \frac{Y_{i\cdot}}{n},\ i = 1, 2;\ \hat{\theta}_{\cdot j} = \frac{Y_{\cdot j}}{n},\ j = 1, 2.$$

Then under null

$$X^2 = \sum_{j=1}^{2} \sum_{i=1}^{2} \frac{(Y_{ij} - n\hat{\theta}_{i\cdot}\hat{\theta}_{\cdot j})^2}{n\hat{\theta}_{i\cdot}\hat{\theta}_{\cdot j}}$$

has chi-squre distribution with 1 degree of fredom.

In our case $\hat{\theta}_{1\cdot} = 0.52$, $\hat{\theta}_{\cdot 1} = 0.87$, $\hat{\theta}_{2\cdot} = 0.48$, $\hat{\theta}_{\cdot 2} = 0.13$, and $X^2 = 17.774$. H_0 should be rejected, because even for significance level $\alpha = 0.005$, the critical value is 7.88.

Oppgave 2

Consider the following model (binomial model with one continuous explanatory variable): $Y_1, ..., Y_N$ are independent;

$$Y_i \sim \mathit{binomial}(n_i, \pi_i),\ \pi_i = F(x_i\beta),\ i = 1, ..., N,$$

where x_i and β are scalars, and F is a differentiable, strictly monotone cumulative distribution function. Denote the corresponding density by f.

a) Show that the score statistic is

$$U = \sum_{i=1}^{N} (Y_i - n_i F(x_i\beta)) \frac{f(x_i\beta)}{F(x_i\beta)(1 - F(x_i\beta))} x_i.$$

Solution. This expression for U is obtained by differentiating the log-likelihood function, which is

$$l = \sum_{i=1}^{N} \left[Y_i \ln F(x_i\beta) + (n_i - Y_i) \ln(1 - F(x_i\beta)) + \ln \binom{n_i}{Y_i} \right].$$

In the rest of the exercise suppose that F is the logistic distribution function:

$$F(x) = \frac{e^x}{1+e^x}, \quad -\infty < x < \infty.$$

b) Show that in this case the expression for U can be simplified. (Hint: find $f/F(1-F)$.)

Solution. For the logistic distribution

$$\frac{f(x)}{F(x)(1-F(x))} = 1$$

and therefore

$$U = \sum_{i=1}^{N}(Y_i - n_i F(x_i\beta))x_i = \sum_{i=1}^{N}\left(Y_i - \frac{n_i e^{x_i\beta}}{1+e^{x_i\beta}}\right)x_i.$$

c) Find the information $I = \mathrm{E}U^2$.

Solution.

$$I = \sum_{i=1}^{N} x_i^2 VarY_i = \sum_{i=1}^{N} x_i^2 n_i\pi_i(1-\pi_i)$$

d) Show that the method of scoring for the maximum likelihood estimator has the form

$$b^{(m)} = b^{(m-1)} + \frac{\sum_{i=1}^{N}(Y_i - n_i\pi_i^{(m-1)})x_i}{\sum_{j=1}^{N} n_j\pi_j^{(m-1)}(1-\pi_j^{(m-1)})x_j^2},$$

where

$$\pi_i^{(m-1)} = \frac{e^{x_i b^{(m-1)}}}{1+e^{x_i b^{(m-1)}}}.$$

Solution. This follows from b), c) and that the method of scoring in one parameter case has the form

$$b^{(m)} = b^{(m-1)} + U^{(m-1)}/I^{(m-1)}.$$

Oppgave 3 Let the cumulative distribution function of Y be

$$F(y) = e^{-e^{-(y-\theta)}}, \quad -\infty < y < \infty; \quad -\infty < \theta < \infty.$$

Show that the distribution of Y belongs to the exponential family. Find Ee^{-Y}.

Solution. The density is

$$f(y) = e^{-(y-\theta)-e^{-(y-\theta)}} = e^{-e^{-y}e^{\theta}+\theta-y} = e^{a(y)b(\theta)+c(\theta)+d(y)},$$

where

$$a(y) = e^{-y},\ b(\theta) = -e^{\theta},\ c(\theta) = \theta,\ d(y) = -y.$$

Then

$$Ee^{-Y} = Ea(Y) = -\frac{c'(\theta)}{b'(\theta)} = e^{-\theta}.$$