Tentative solutions to TMA4315 GLM, Exam 07.12.2017

Problem 1: Poisson regression

$$f(y) = \exp\left\{ \frac{y\theta - b(\theta)}{\phi} \cdot w + c(y, w, \phi) \right\}$$

exp. fam.

Poisson: $f(y) = \frac{\lambda^y}{y!} e^{-\lambda}$, $y = 0, 1, 2 \ldots$
$\lambda > 0$

a)

$$f(y) = \exp\left(y \ln \lambda - \ln(y!) - \lambda \right)$$

$$= \exp\left(\frac{y \ln \lambda - \lambda}{1} \cdot 1 - \ln(y!) \right)$$

i) $\underline{\underline{\theta = \ln \lambda}}$ $\quad \underline{\underline{\phi = 1}}$ $\quad \underline{\underline{w = 1}}$ $\quad c(y, w=1, \phi=1) = \underline{\underline{\ln(y!)}}$

$\underline{b(\theta) = \lambda = \underline{\exp(\theta)}}$

ii) $E(Y) = b'(\theta)$ and $Var(Y) = b''(\theta) \cdot \frac{\phi}{w}$

iii) $E(Y) = \frac{d}{d\theta} \exp(\theta) = \exp(\theta) = \underline{\underline{\lambda}}$

$Var(Y_i) = \frac{d^2}{d\theta^2} \exp(\theta) = \exp(\theta) = \underline{\underline{\lambda}}$

iv) Canonical link: $\theta = \eta$ but as a function of $\mu = E(Y)$

$$\eta = \theta = \ln \lambda = \ln(\mu) \qquad \underline{g(\mu) = \ln(\mu)}$$

The log link is the canonical link.

b) Poisson regression with log link:

1) $Y_i \sim \text{Poisson}(\lambda_i)$ ← yes, is exp. fam and (Y_i, x_i) independent

2) $\eta_i = x_i^T \beta$ where β is $p \times 1$ ✓ yes is one-to-one and twice differentiable

3) $\eta_i = \underset{\uparrow \text{link}}{g(\mu_i)} = \ln \mu_i$ and $h(\eta_i) = \underset{\uparrow \text{responsfunction}}{\exp(\eta_i)} = \mu_i$

In addition we assume that design matrix $\underline{X} = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix}$ has full rank.

$$L(\beta) = \prod_{i=1}^{n} f_i(y_i; \beta) = \prod_{i=1}^{n} \frac{\lambda_i^{y_i}}{y_i!} e^{-\lambda_i}$$

$$l(\beta) = \sum_{i=1}^{n} \underbrace{\ln f_i(y_i; \beta)}_{l_i(\beta)} = \underline{\underline{\sum_{i=1}^{n} \underbrace{\left(y_i \ln \lambda_i - \lambda_i - \ln(y_i!)\right)}_{l_i(\beta)}}}$$

↑ can stop here. = full score

$\ln \lambda_i = x_i^T \beta$

$\lambda_i = \exp(x_i^T \beta)$

$$\underline{\underline{l(\beta) = \sum_{i=1}^{n} y_i x_i^T \beta - \sum_{i=1}^{n} \exp(x_i^T \beta) - \sum_{i=1}^{n} \ln(y_i!)}} \leftarrow \text{or stop here}$$

$$\underset{\substack{\uparrow \\ \text{score function}}}{s(\beta)} = \frac{\partial l}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{i=1}^{n} l_i(\beta) = \sum_{i=1}^{n} \frac{\partial}{\partial \beta} l_i(\beta) = \sum_{i=1}^{n} s_i(\beta)$$

$$s_i(\beta) = \frac{\partial}{\partial\beta} l_i(\beta) = \frac{\partial}{\partial\beta}\left[y_i x_i^T\beta - \exp(x_i^T\beta) + C\right] \circledast$$

$$= \left(y_i x_i - \exp(x_i^T\beta)\cdot x_i\right)$$

$$\overset{p\times 1}{s(\beta)} = \sum_{i=1}^{n}\left[y_i - \exp(x_i^T\beta)\right]x_i = \sum_{i=1}^{n}(y_i - \mu_i)\overset{p\times 1}{x_i}$$

$s(\beta)$ is a $p\times 1$ column vector

$\circledast$ Remember the rules for derivatives wrt a vector:

$$\frac{\partial}{\partial\beta}\left(x_i^T\beta\right) = x_i$$

since $x_i^T\beta = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$ and

$$\frac{\partial}{\partial\beta} x_i^T\beta = \begin{bmatrix} \frac{\partial x_i^T\beta}{\partial\beta_0} \\ \frac{\partial x_i^T\beta}{\partial\beta_1} \\ \vdots \\ \frac{\partial x_i^T\beta}{\partial\beta_k} \end{bmatrix} = \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{ik} \end{bmatrix} = x_i$$

which leeds to $\frac{\partial}{\partial\beta}\left(y_i x_i^T\beta\right) = y_i x_i$

and $\frac{\partial}{\partial\beta}\left[\exp(x_i^T\beta)\right] = \exp(x_i^T\beta)\cdot\frac{\partial x_i^T\beta}{\partial\beta}$

$$= \exp(x_i^T\beta)\cdot x_i$$

Then, the expected Fisher information matrix.

Two possible ways further: $\underline{F(\beta) = E(H(\beta)) \text{ or } F(\beta) = Cov(S(\beta))}$

$$H(\beta) = -\frac{\partial^2}{\partial\beta\partial\beta^T} \ell(\beta) = -\frac{\partial}{\partial\beta^T}\left(S(\beta)\right) = -\frac{\partial}{\partial\beta^T}\left(\sum_{i=1}^{n} S_i(\beta)\right)$$

$$= -\sum_{i=1}^{n} \frac{\partial}{\partial\beta^T} S_i(\beta) = \sum_{i=1}^{n} H_i(\beta)$$

$$H_i(\beta) = -\frac{\partial}{\partial\beta^T} S_i(\beta) = -\frac{\partial}{\partial\beta^T}\left[(y_i - \exp(x_i^T\beta))\cdot x_i\right]$$

$$\circledast = -0 + \exp(x_i^T\beta)\cdot x_i \cdot x_i^T = \underbrace{\exp(x_i^T\beta)}_{\mu_i}\cdot x_i x_i^T$$

so $$\underset{p\times p}{H(\beta)} = \sum_{i=1}^{n} \mu_i x_i x_i^T = \sum_{i=1}^{n} \exp(x_i^T\beta)\cdot \underset{p\times 1}{x_i}\underset{1\times p}{x_i^T}$$

$$F(\beta) = E(H(\beta)) = E\left[\sum_{i=1}^{n} \exp(x_i^T\beta)\cdot x_i x_i^T\right]$$

$$= \underline{\underline{\sum_{i=1}^{n} \exp(x_i^T\beta)\cdot x_i x_i^T}} \qquad \underline{\underline{\text{which is } p\times p}}$$

$\circledast$ $$\frac{\partial}{\partial\beta^T}\left[\exp(x_i^T\beta)\cdot x_i\right] = x_i\cdot\frac{\partial}{\partial\beta^T}\left[\exp(x_i^T\beta)\right]$$

$$= x_i \cdot \exp(x_i^T\beta)\cdot\frac{\partial}{\partial\beta^T}\left[x_i^T\beta\right] = \exp(x_i^T\beta)\cdot x_i \cdot x_i^T$$

Since $$\left(\frac{\partial x_i^T\beta}{\partial\beta_0}\ \frac{\partial x_i^T\beta}{\partial\beta_1}\ \frac{\partial x_i^T\beta}{\partial\beta_2} \cdots \frac{\partial x_i^T\beta}{\partial\beta_k}\right) = \left(1\ x_{i1}\ x_{i2} \cdots x_{ik}\right) = x_i^T$$

Alternative way: $F(\beta) = Cov(s(\beta)) = E(s(\beta)\, s(\beta)^T)$

since

$$E(s(\beta)) = E\left(\sum_{i=1}^{n} (Y_i - \mu_i)\, x_i\right) = 0 \quad \text{and also}$$

$$E(s_i(\beta)) = E((Y_i - \mu_i)\, x_i) = 0$$

Since the Y_i's are independent

$$F(\beta) = E(s(\beta)\, s(\beta)^T) = \sum_{i=1}^{n} E(s_i(\beta)\, s_i(\beta)^T)$$

$$= \sum_{i=1}^{n} E\left((Y_i - \mu_i)\, x_i\, x_i^T (Y_i - \mu_i)\right) = \sum_{i=1}^{n} x_i x_i^T\, \underbrace{E[(Y_i - \mu_i)^2]}_{Var(Y_i) \,=\, \mu_i}$$

$$= \underline{\underline{\sum_{i=1}^{n} \mu_i\, x_i x_i^T}}$$

To find the mle $\hat{\beta}$ we solve $s(\hat{\beta}) = 0$ using the Fisher scoring algorithm

$$\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} + \underbrace{F(\hat{\beta}^{(t)})^{-1}}_{\left[\sum_{i=1}^{n} \mu_i x_i x_i^T\right]^{-1}}\ \underbrace{s(\hat{\beta}^{(t)})}_{\left[\sum_{i=1}^{n} (y_i - \mu_i)\, x\right]}$$

There is <u>no</u> closed form solution here.

c) Interpret R-print-out from Galapagos island data set

i) $\hat{\beta} \approx N(\beta, F^{-1}(\hat{\beta}))$

where $F(\hat{\beta})$ is the expected Fisher information matrix inserted the param. estimate $\hat{\beta}$.

[The Fisher information matrix was in b) found to be
$F(\beta) = \sum_{i=1}^{n} \mu_i x_i x_i^T$ and $\mu_i = \exp(\eta_i) = \exp(x_i^T \beta)$ so
$$\underbrace{F(\hat{\beta})}_{6\times 6 \text{ matrix}} = \sum_{i=1}^{n} \exp(x_i^T \hat{\beta})\, x_i x_i^T$$
]

ii) β_2 = regression coefficient for elevation (m)

$\hat{\beta}_2 = 3.54 \cdot 10^{-3}$ (under Estimate in print-out)

iii) Since $\hat{\mu}_i = \exp(x_i^T \hat{\beta}) = \exp(\hat{\beta}_0) \exp(\hat{\beta}_1)^{x_{i1}} \cdot \exp(\hat{\beta}_2)^{x_{i2}} \cdots \exp(\hat{\beta}_5)^{x_{i5}}$

If we consider two islands A and B with the same values for (x_1, x_3, x_4, x_5), but where the elevation of island A is 1m higher than island B, then the predicted number of species on island B is a factor $\exp(\hat{\beta}_2) = 1.003547$ higher than on island A.

[10 m higher $\exp(\hat{\beta}_2 \cdot 10) = 1.03604$
100 $\exp(\hat{\beta}_2 \cdot 100) = 1.42$]

iii) $\hat{SD}(\hat{\beta}_2) = 8.741 \cdot 10^{-5}$

iv) 95% CI for β_2:

$\hat{\beta}_2 \approx N(\beta_2, \hat{SD}(\hat{\beta}_2))$

$\hat{\beta} \pm 1.96 \cdot \hat{SD}(\hat{\beta}_2) = 3.541 \cdot 10^{-3} \pm 1.96 \cdot 8.741 \cdot 10^{-5}$

$= [0.003369, 0.003712]$

Problem 2: Titanic

i) 1) $Y_i \sim \text{bin}(n_i, \pi_i)$, (x_i, Y_i) independent pairs

2) $\eta_i = x_i^T \beta$ with $p = 6$ param. to estimate

here X has full rank

3) Logit link: $\eta_i = g(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$

one-to-one and twice differentiable

ii) $\hat{\beta}_{crew} = -\hat{\beta}_{class1} - \hat{\beta}_{class2} - \hat{\beta}_{class3}$

$= -0.91338 + 0.10471 + 0.86438 = \underline{\underline{0.0557}}$

iii) $H_0: \beta_{class1} = \beta_{class2} = \beta_{class3} = 0$
vs H_1: at leest one different from 0

Cen be written as $C\beta = d$

$H_0: C\beta = d$ vs. $H_1: C\beta \neq d$

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \qquad d = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Test statistic is then

$$X^2 = (C\hat{\beta} - d)^T [C F^{-1}(\hat{\beta}) C^T]^{-1} (C\hat{\beta} - d) \approx \chi^2_3$$

and reject H_0 when $P(\chi^2_3 > x^2) \leq 0.05$.

Alternatively, use the likelihood ratio test where
A: large model is the model with Class1, Class2, Class3 as covariates
B: small without Class as covariate

$\left(-2\ln(l_B - l_A) = 180.9 \qquad P(\chi^2_3 > 180.9) = 0 \right)$

iv) AgeAdult $\hat{\beta} = -1.06154$

Compare an Adult to a Child, where these two people are of the same Sex and Class (1,2,3).
(Female, Male)

$$\frac{P(\text{survive} \mid \text{Adult})}{P(\text{not survive} \mid \text{Adult})} = \frac{P(\text{survive} \mid \text{Child})}{P(\text{not survive} \mid \text{Child})} \cdot e^{\beta_{AgeAdult}}$$

$e^{\beta_{AgeAdult}}$: $\underbrace{\exp(-1.06154)}_{0.35}$

odds of survival for adult $= 0.35 \cdot$ odds of survival for Child

v) Deviance test

$$D = -2\Big(\underbrace{\ell(\text{candidate})}_{\substack{\text{loglikelihood of}\\ \text{our model}\\ \hat{\mu}_i = x_i^T \hat{\beta}}} - \underbrace{\ell(\text{saturated})}_{\substack{\text{loglikelihood of}\\ \text{model with Sex * Class * Age}}}\Big)$$

all 14 groups get $\tilde{\mu}_j = \bar{y}_j$ in group j

$D = 112.57$

Asymptotically: $D \approx \chi^2_8$

number of groups $14 - 6 = 8$ number of estimated pars.

$$P(\chi^2_8 > 112.57) = 0$$

H_0: our model is good
H_1: not so

reject H_0, this is not a good model compared to the saturated model.

Next step: try to include interaction effects?

<u>Problem 3:</u> Random intercept linear mixed effects model

Example: Effect of $\underbrace{\text{life style change}}_{\text{(an intervention)}}$ on $\underbrace{\text{fitness}}_{}$.

(as measured by VO_2max oxygen uptake)

Have n_i measurements for each of m persons.

Here n_i could be two = before and after the intervention.

X_i could include $\underbrace{\text{age, sex}}_{\text{not change}}$, $\underbrace{\text{weight etc and}}_{\text{might change}}$ time

before → time ← after

Model assumptions:

$$\underbrace{Y_i}_{n_i\times 1} = \underbrace{X_i}_{n_i\times p}\underbrace{\beta}_{p\times 1} + \underbrace{1}_{n_i\times 1}\underbrace{\gamma_{0i}}_{1\times 1} + \underbrace{\varepsilon_i}_{n_i\times 1}$$

ε_i and γ_{0i} are independent

$$\varepsilon_i \sim N(0, \sigma^2 I) \quad \text{and} \quad \gamma_{0i} \sim N(0, \tau_0^2)$$

$(x_1, Y_1), (x_2, Y_2), (x_3, Y_3), \ldots, (x_m, Y_m)$ are independent pairs.

Conditional model

$$Y_i \mid \gamma_{0i} \sim N(X_i\beta + 1\gamma_{0i}, \sigma^2)$$

Combined with

$$\gamma_{0i} \sim N(0, \tau_0^2)$$

Marginal model

$$Y_i = X_i\beta + \overbrace{(1\gamma_{0i} + \varepsilon_i)}^{\varepsilon_i^*} = X_i\beta + \varepsilon_i^*$$

$$\text{and} \quad \varepsilon_i^* \sim N(0, 1\tau_0 1^T + \sigma^2 I)$$

$$Y_i \sim N(X_i\beta, \underbrace{\tau_0 1 1^T + \sigma^2 I}_{V_i})$$

Intraclass correlation:

$$Cov(Y_{ij}, Y_{ik}) = (\tau_0 11^T + \sigma^2 I)_{j,k} = \overbrace{\begin{bmatrix} \tau_0^2+\sigma^2 & \tau_0 & & \\ \tau_0 & \tau_0^2+\sigma^2 & \ddots & \\ \tau_0 & & & \tau_0+\sigma \end{bmatrix}}^{V_i}_{\{i,j,k\}} = \tau_0^2$$

$$Corr(Y_{ij}, Y_{ik}) = \frac{Cov(Y_{ij}, Y_{ik})}{\sqrt{\underbrace{Var(Y_{ij})}_{\tau_0^2+\sigma^2} \cdot \underbrace{Var(Y_{ik})}_{\tau_0^2+\sigma^2}}} = \frac{\tau_0^2}{\tau_0^2+\sigma^2}$$

If we select two obs from the same cluster (or person), this is the correlation between these two obs. that is <u>provided</u> in the model.

Parameter estimation is performed based on the (marginal) likelihood, and maximum likelihood assuming σ^2, τ_0 known gives

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

where $V = Cov(Y)$ and based on cluster i, so alternatively

$$\hat{\beta} = \left(\sum_{i=1}^{m} X_i^T V_i^{-1} X_i\right)^{-1} \sum_{i=1}^{m} X_i^T V_i^{-1} Y_i$$

where in our case for the random intercept model

$$V_i = \tau_0^2 11^T + \sigma^2 I$$

$\theta = (\tau_0^2, \sigma^2)$ are estimated based on REML.

For known V then

$$\hat{\beta} \sim N\left(\beta, (X^T V^{-1} X)^{-1} = \left(\sum_{i=1}^{m} X_i^T V_i^{-1} X_i\right)^{-1}\right)$$

but with estimated $\hat{V}_i$'s then

$$\hat{\beta} = \left[\sum_{i=1}^{m} X_i^T V_i^{-1} X_i\right]^{-1} \sum_{i=1}^{m} X_i^T V_i^{-1} Y_i$$

$$\hat{\beta} \approx N\left(\beta, \left[\sum_{i=1}^{m} X_i^T V_i^{-1} X_i\right]^{-1}\right)$$

approx

Inference on β is performed using this approximation.

Inference on θ can be performed assuming

$$\theta \approx N(\theta, \hat{Cov}(\hat{\theta}))$$

where θ is estimated using REML. However, for hypotheses on the boundary of the param. space this approx. is not very good and other solutions are found