UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in: STK3100/STK4100 — Introduction to Generalized Linear Models
SKETCH OF SOLUTION

Day of examination: Monday 6th December 2021

Examination hours: 15.00 – 19.00

This problem set consists of 5 pages.

Appendices: Formulas in STK3100/4100

Permitted aids: Approved calculator

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Problem 1

a

We have

$$\frac{\partial \log f(Y;\theta,\phi)}{\partial \theta} = \left(Y - b'(\theta)\right)/a(\phi),$$

and hence combined with the hint

$$E\left[\left(Y - b'(\theta)\right)/a(\phi)\right] = 0 \quad \Rightarrow \quad E[Y] = b'(\theta).$$

Furthermore we have

$$\frac{\partial^2 \log f(Y;\theta,\phi)}{\partial \theta^2} = -b''(\theta)/a(\phi),$$

which combined with the result $E[Y] = b'(\theta)$ shown above and the hint gives

$$\begin{aligned}
-E\left[-b''(\theta)/a(\phi)\right] &= E\left[\left(\left(Y - b'(\theta)\right)/a(\phi)\right)^2\right] \\
&\Downarrow \\
b''(\theta)/a(\phi) &= E\left[\left(Y - b'(\theta)\right)^2\right]/a(\phi)^2 \\
&\Downarrow \\
E\left[\left(Y - E[Y]\right)^2\right] &= \mathrm{Var}[Y] = b''(\theta)a(\phi)
\end{aligned}$$

b

We can write

$$P(Y = y) = \exp\left\{\log P(Y = y)\right\} = \exp\left\{y \log \mu - \log y! - \mu\right\}$$

which is identical to (1) with $\theta = \log \mu$, $b(\theta) = \mu = \exp(\theta)$, $a(\phi) = 1$ and $c(y,\phi) = -\log y!$.

(Continued on page 2.)

c

From the results proven in **a** we get

$$E[Y] = b'(\theta) = \exp(\theta) = \mu$$
$$\mathrm{Var}[Y] = b''(\theta)a(\phi) = \exp(\theta) = \mu$$

d

The log-likelihood function is

$$L(\beta_0, \beta_1) = \log\left\{\prod_{i=1}^{n} \exp\{Y_i \log \mu_i - \log Y_i! - \mu_i\}\right\} = \sum_{i=1}^{n}\{Y_i \log \mu_i - \log Y_i! - \mu_i\}$$
$$= \sum_{i=1}^{n}\{Y_i\eta_i - e^{\eta_i} - \log Y_i!\} \qquad (1)$$

From (1) we get for $j = 0, 1$, with $x_{i0} = 1$ and $x_{i1} = x_i$,

$$\frac{\partial L(\beta_0, \beta_1)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial\{Y_i\eta_i - e^{\eta_i} - \log Y_i!\}}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^{n}(Y_i - e^{\eta_i})\frac{\partial \eta_i}{\partial \beta_j}$$
$$= \sum_{i=1}^{n}(Y_i - \mu_i)x_{ij} = 0$$

which implies $\sum_{i=1}^{n}(Y_i - \mu_i) = 0$ and $\sum_{i=1}^{n}(Y_i - \mu_i)x_i = 0$.

It is also ok to find $\hat{\beta}_0$ and $\hat{\beta}_1$ from the likelihood equations for a GLM given in the appendix.

Problem 2

a

The model behind the fit reported in **a** is a logistic regression model. For $i = 1, \ldots, n$ ($n = 2702$), let Y_i denote the response variable, where $Y_i = 1$ if there was objection against patent i, and $Y_i = o$ if there was no objection against patent i, and $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}, x_{i6}, x_{i7})$ denote the vector of explanatory variables, where $x_{i1} = 1$ for the intercept and

- x_{i2} = grant year
- x_{i3} = number of citations for the patent
- $x_{i4} = 1$ if US twin patent exists, otherwise $x_{i4} = 0$
- $x_{i5} = 1$ if patent holder is from the US, otherwise $x_{i5} = 0$
- $x_{i6} = 1$ if patent holder is from Germany, Switzerland or Great Britain, otherwise $x_{i6} = 0$
- x_{i7} = number of designated countries for the patent

(Continued on page 3.)

Let $\pi_i = P(Y_i = 1)$. The model assumes that $Y_1, \dots, Y_n$ are independent and, with $\boldsymbol{\beta} = (\beta_1, \dots, \beta_7)^T$, that

$$\text{logit}(\pi_i) = \frac{\pi_i}{1-\pi_i} = \eta_i = \mathbf{x}_i\boldsymbol{\beta} = \sum_{j=1}^{7} \beta_j x_{ij},$$

which is equivalent to

$$\pi_i = \frac{\exp\left(\sum_{j=1}^{7} \beta_j x_{ij}\right)}{1+\exp\left(\sum_{j=1}^{7} \beta_j x_{ij}\right)}$$

The estimate $\hat{\beta}_5$ for `patus` can be interpreted by considering two different patents i and k, that have the same values of all the explanatory variables except `patus`, i.e. $x_{ij} = x_{kj}$ for $j = 1, 2, 3, 4, 6, 7$ and $x_{i5} \neq x_{k5}$. In words they share grant year, number of citations for the patent, either both patent holders have a US twin patent or not, neither of the patent holders are from Germany, Switzerland or Great Britain, and they have the same number of designated countries, but the holder of patent i is from the US, while the holder of patent k is not. Then the odds ratio for the two patents is is

$$\frac{P(Y_i = 1)/\left[1 - P(Y_i = 1)\right]}{P(Y_k = 1)/\left[1 - P(Y_k = 1)\right]} = \frac{\pi_i/\left[1-\pi_i\right]}{\pi_k/\left[1-\pi_k\right]}$$

$$= \frac{\frac{\exp\left(\sum_{j=1}^{7} \beta_j x_{ij}\right)}{1+\exp\left(\sum_{j=1}^{7} \beta_j x_{ij}\right)} \Big/ \left[1 - \frac{\exp\left(\sum_{j=1}^{7} \beta_j x_{ij}\right)}{1+\exp\left(\sum_{j=1}^{7} \beta_j x_{ij}\right)}\right]}{\frac{\exp\left(\sum_{j=1}^{7} \beta_j x_{kj}\right)}{1+\exp\left(\sum_{j=1}^{7} \beta_j x_{kj}\right)} \Big/ \left[1 - \frac{\exp\left(\sum_{j=1}^{7} \beta_j x_{kj}\right)}{1+\exp\left(\sum_{j=1}^{7} \beta_j x_{kj}\right)}\right]} = \frac{\exp\left(\sum_{j=1}^{7} \beta_j x_{ij}\right)}{\exp\left(\sum_{j=1}^{7} \beta_j x_{kj}\right)} = e^{\beta_5}$$

Hence the estimated odds ratio for the two patents is $e^{\hat{\beta}_5} = e^{-0.43685} = 0.646$, which means that $\hat{\beta}_5$ tells us that we estimate that the odds of experiencing opposition against a patent is reduced by 35.4% when the patent holder is from the US.

b

The missing numbers are filled in in the table below. They are found in the following way

- Df for Model 2: The difference in residual degrees of freedom between Model 1 and Model 2, hence $2695 - 2694 = 1$
- Deviance for Model 2: The difference between the deviance of Model 1 and Model 2, hence $2996.8 - 2981.7 = 15.1$
- Resid. Df for Model 3: The residual degrees of freedom, which is the difference between the number of observations (2702) and the number of parameters in Model 3. The number of parameters is 1 (intercept)+ 6 (1 for each of the main effects) + 1 (year:patus) + 1 (patus:ncountry)= 9, hence Resid. Df for Model 3 is $2702 - 9 = 2693$
- Resid. Dev for Model 4: The deviance for the fit of Model 4, which can be found by subtracting the difference between the deviance of Model 3 and Model 4 (2.5899) from the deviance for the fot of Model 3 (2979.2), hence $2979.2 - 2.5899 = 2976.61$

(Continued on page 4.)

```
     > anova(fit1,fit2,fit3,fit4,test="LRT")
Analysis of Deviance Table

Model 1: opp ~ year + ncit + ustwin + patus + patgsgr + ncountry
Model 2: opp ~ year + ncit + ustwin + patus + patgsgr + ncountry + year:patus
Model 3: opp ~ year + ncit + ustwin + patus + patgsgr + ncountry + year:patus +
    patus:ncountry
Model 4: opp ~ year + ncit + ustwin + patus + patgsgr + ncountry + year:patus +
    patus:ncountry + patgsgr:ncountry
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      2695     2996.8
2      2694     2981.7  1     15.1 0.0001011 ***
3      2693     2979.2  1   2.5449 0.1106478
4      2692     2976.61 1   2.5899 0.1075493
```

c

The AIC for a model with p parameters and maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ is defined as $-2\left(L(\hat{\boldsymbol{\beta}}) - p\right)$. Minimising this over a set of models identifies which of these models that has a distribution closest to the true distribution.

The likelihood ratio test can only be used for nested models, while comparing AIC values does not have this restriction. In this example the probit model achieves the smallest AIC value, which is indicates that it is the best choice of link function here. However, the differences are very small, in particular between probit and logit.

Problem 3

a

The log-likelihood function is

$$L(\boldsymbol{\beta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log \mid \mathbf{V} \mid -\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Hence, using the hint and that both a scalar and $\mathbf{V}^{-1}$ are symmetric,

$$\begin{aligned}
\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= -\frac{1}{2}\frac{\partial(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\
&= -\frac{1}{2}\frac{\partial(\mathbf{y}^T\mathbf{V}^{-1}\mathbf{y} - \mathbf{y}^T\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} + \ \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\
&= -\frac{1}{2}\frac{\partial(\mathbf{y}^T\mathbf{V}^{-1}\mathbf{y} - 2\mathbf{y}^T\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\
&= (\mathbf{y}^T\mathbf{V}^{-1}\mathbf{X})^T - \frac{1}{2}\left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} + \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^T\right)\boldsymbol{\beta} \\
&= \mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} - \mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta}
\end{aligned}$$

which is $= 0$ for

$$\boldsymbol{\beta} = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$$

(Continued on page 5.)

b

Since both $\mathbf{Y}|\mathbf{u}$ and $\mathbf{u}$ are normally distributed, the joint distribution of $\mathbf{Y}$ and $\mathbf{u}$ is also normal. We have $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$, $E[\mathbf{u}] = \mathbf{0}$, $\text{Var}[\mathbf{Y}] = \mathbf{V}$ and $\text{Var}[\mathbf{u}] = \boldsymbol{\Sigma}_\mathbf{u}$. Furthermore

$$\text{Cov}(\mathbf{Y}, \mathbf{u}) = \text{Cov}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \mathbf{u}) = \text{Cov}(\mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \mathbf{u}) = \mathbf{Z}\text{Cov}(\mathbf{u}, \mathbf{u}) = \mathbf{Z}\text{Var}(\mathbf{u}) = \mathbf{Z}\boldsymbol{\Sigma}_\mathbf{u}$$

which means that $\text{Cov}(\mathbf{Y}, \mathbf{u}) = (\mathbf{Z}\boldsymbol{\Sigma}_\mathbf{u})^T = \boldsymbol{\Sigma}_\mathbf{u}\mathbf{Z}^T$. Hence

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{u} \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V} & \mathbf{Z}\boldsymbol{\Sigma}_\mathbf{u} \\ \boldsymbol{\Sigma}_\mathbf{u}\mathbf{Z}^T & \boldsymbol{\Sigma}_\mathbf{u} \end{pmatrix}\right]$$

Using the formula for conditional expectation for the multivariate normal distribution given in the appendix, we get

$$E[\boldsymbol{u}|\mathbf{Y} = \mathbf{y}] = \mathbf{0} + (\mathbf{Z}\boldsymbol{\Sigma}_\mathbf{u})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\Sigma}_\mathbf{u}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \qquad (2)$$

Since $\mathbf{V}$ and $\boldsymbol{\Sigma}_\mathbf{u}$, are assumed known, a prediction of the random effects can be achieved by replacing $\boldsymbol{\beta}$ by $\tilde{\boldsymbol{\beta}}$ in (2)

$$\hat{\mathbf{u}} = \boldsymbol{\Sigma}_\mathbf{u}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})$$

c

The marginal expected value of Y_{ij} is

$$\begin{aligned}\mu_{ij} = E[Y_{ij}] = E[E[Y_{ij}|\mathbf{u}_i]] = \int E[Y_{ij}|\mathbf{u}_i]f(\mathbf{u}_i)d\mathbf{u}_i = \int (\mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{u}_i)f(\mathbf{u}_i)d\mathbf{u}_i \\ = \mathbf{x}_{ij}\boldsymbol{\beta}\int f(\mathbf{u}_i)d\mathbf{u}_i + \mathbf{z}_{ij}\int \mathbf{u}_i f(\mathbf{u}_i)d\mathbf{u}_i = \mathbf{x}_{ij}\boldsymbol{\beta}\end{aligned}$$

since $\int f(\mathbf{u}_i)d\mathbf{u}_i = 1$ and $\int \mathbf{u}_i f(\mathbf{u}_i)d\mathbf{u}_i = E[\mathbf{u}_i] = \mathbf{0}$. Hence, as for the GLMM, the link function for the implied marginal model is also the identity link. This is not a general result for all lunk functions.

END