UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in:	STK3100/STK4100 — Introduction to Generalized Linear Models
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 distribution with probability mass function (pmf) or probability density function (pdf)

$$f(y;\theta,\phi) = \exp\left\{(\theta y - b(\theta))/a(\phi) + c(y,\phi)\right\} \tag{1}$$

where $a(\cdot)$, $b(\cdot)$ and $c(\cdot,\cdot)$ are functions, and θ and ϕ parameters, belongs to the *exponential dispersion family* of distributions.

a

Show that if the random variable Y has a distribution from the exponential dispersion family with parameters θ and ϕ, then $E[Y] = b'(\theta)$ and $\text{Var}[Y] = b''(\theta)a(\phi)$.

Hint: Use that we have from general likelihood results that

$$E\left[\frac{\partial \log f(Y;\theta,\phi)}{\partial\theta}\right] = 0$$
$$-E\left[\frac{\partial^2 \log f(Y;\theta,\phi)}{\partial\theta^2}\right] = E\left[\left(\frac{\partial \log f(Y;\theta,\phi)}{\partial\theta}\right)^2\right].$$

and in addition that $\text{Var}[Y] = E\left[(Y - E[Y])^2\right]$.

In the following we assume that the random variable Y is Poisson distributed with pmf

$$P(Y=y) = \frac{\mu^y}{y!}\exp(-\mu), \quad y = 0,1,2,\dots \tag{2}$$

b

Show that the Poisson distribution belongs to the exponential dispersion family of distributions, that is show that (2) can be written in the form of (1). Determine θ, $a(\phi)$, $b(\theta)$ and $c(y,\phi)$.

(Continued on page 2.)

c

Show that $E[Y] = \text{Var}[Y] = \mu$.

Assume now that $Y_1, Y_2, \ldots, Y_n$ are independent and that Y_i is Poisson distributed with mean μ_i, $i = 1, \ldots, n$. Given the values x_i, $i = 1, \ldots, n$, of an explanatory variable, we assume a generalized linear model (GLM) for $Y_1, Y_2, \ldots, Y_n$, with the linear predictor $\eta_i = \beta_0 + \beta_1 x_i$ linked to the mean through the canonical link function $g(\mu_i) = \log(\mu_i) = \eta_i$.

d

Derive an expression for the log-likelihood function $L(\beta_0, \beta_1)$, and show that the maximum likelihood estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are the solutions to the following equations

$$\sum_{i=1}^{n}(Y_i - \mu_i) = 0 \quad \text{and} \quad \sum_{i=1}^{n}(Y_i - \mu_i)x_i = 0$$

Problem 2

In this problem we will consider data collected for a study analysing the probability of objections agains patents granted by the European patent office. The data set concerns 2702 patents from the sector of semiconductor/computer industry. On each of the 2702 patents the following variables are recorded

- `opp`: Patent opposition (1=yes; 0=no)
- `year`: Grant year
- `ncit`: Number of citations for the patent
- `ustwin`: US twin patent exists (1=yes; 0=no)
- `patus`: Patent holder from the US (1=yes; 0=no)
- `patgsgr`: Patent holder from Germany, Switzerland or Great Britain (1=yes; 0=no)
- `ncountry`: Number of designated countries for the patent

We will investigate how the probability of having an objection against a patent depends on grant year, number of citations for the patent, whether a US twin patent exists, whether the patent holder is from the US, whether the patent holder is from Germany, Switzerland or Great Britain, and the number of designated countries for the patent.

a

On the next page you find output from R from a fit with only main effects. Describe the model behind this fit, including necessary assumptions. Give an interpretation of the estimate for `patus`.

(Continued on page 3.)

```
 Call:
glm(formula = opp ~ year + ncit + ustwin + patus + patgsgr +
    ncountry, family = binomial(link = logit), data = patents)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7594  -0.8181  -0.6328   1.1397   2.2151

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) 185.25738   21.85850   8.475  < 2e-16 ***
year         -0.09373    0.01098  -8.537  < 2e-16 ***
ncit          0.12128    0.02214   5.477 4.33e-08 ***
ustwin       -0.41085    0.09959  -4.126 3.70e-05 ***
patus        -0.43685    0.10998  -3.972 7.12e-05 ***
patgsgr       0.18723    0.11691   1.602    0.109
ncountry      0.10302    0.01481   6.957 3.48e-12 ***
---
Null deviance: 3214.5  on 2701  degrees of freedom
Residual deviance: 2996.8  on 2695  degrees of freedom
AIC: 3010.8
```

b

We have also fitted models with interactions. A summary of the fits can be seen in the R output of an analysis of variance table below. Some of the entries in the table are missing and have been replaced by question marks. Give the missing numbers, along with an explanation of how you found them. Which of the models fit the data best, and why?

```
     > 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  ?       ?    0.0001011 ***
3        ?      2979.2  1  2.5449   0.1106478
4      2692        ?    1  2.5899   0.1075493
```

(Continued on page 4.)

c

The table below shows the AIC values for a binomial GLM-model with three different link functions, and the same linear predictor as for the best model in b. Explain what AIC is. Argue why it is sensible to use this as a criteria to compare the different models arising from the different link-functions in this situation, instead of a test such as the likelihood ratio test. Based on these values, which link function should be chosen?

Link function	AIC
logit	2997.724
probit	2996.969
cloglog	3000.778

Problem 3

Let $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{id})^T$, $i = 1, \ldots, n$, be independent vectors that correspond to observations from each of n groups. Given model matrices $\mathbf{X}_i$ (of dimension $d \times p$, with vector $\mathbf{x}_{ij}$ for observation j from group i in row j) and $\mathbf{Z}_i$ (of dimension $d \times q$ with vector $\mathbf{z}_{ij}$ for observation j from group i in row j) , we assume a normal linear mixed model

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, n$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects, $\mathbf{u}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_u)$ is a $q \times 1$ vector of random effects, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{id})^T \sim N(\mathbf{0}, \mathbf{R})$ is independent of $\mathbf{u}_i$.

Now with

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_n \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

$$\boldsymbol{\Sigma}_\mathbf{u} = \begin{pmatrix} \boldsymbol{\Sigma}_u & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_u & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_u \end{pmatrix}, \quad \mathbf{R}_\varepsilon = \begin{pmatrix} \mathbf{R} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R} \end{pmatrix}$$

we may write $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$. Furthermore, we have that

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$$

where $\mathbf{V} = \mathbf{Z}\boldsymbol{\Sigma}_\mathbf{u}\mathbf{Z}^T + \mathbf{R}_\varepsilon$. We assume throughout this problem that that $\mathbf{V}$ is known.

a

Given observed data $\mathbf{Y} = \mathbf{y}$, show that the maximum likelihood estimate of $\boldsymbol{\beta}$ is

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

Hint: Use the matrix derivative results

$$\partial(\boldsymbol{a}^T\gamma)/\partial\gamma = \boldsymbol{a} \quad \text{and} \quad \partial(\gamma^T\boldsymbol{A}\gamma)/\partial\gamma = (\boldsymbol{A} + \boldsymbol{A}^T)\gamma$$

(Continued on page 5.)

b

Show that

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

and that

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

Given $\mathbf{Y} = \mathbf{y}$ and $\tilde{\boldsymbol{\beta}}$, what is a sensible prediction for the random effects?

c

With $\mathbf{Y}_i$, $\mathbf{x}_{ij}$, $\mathbf{z}_{ij}$, $\boldsymbol{\beta}$ and $\mathbf{u}_i$ defined as before, the *generalized* linear mixed model (GLMM) for $\mathbf{Y}_i$ has the form

$$g\left(E[Y_{ij}|\mathbf{u}_i]\right) = \mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{u}_i, \quad i = 1, \ldots, n, \quad j = 1, \ldots, d.$$

Still assuming that g is the identity link function, find the marginal expected value $\mu_{ij} = E[Y_{ij}]$. Comment on the link function for this marginal model implied by the GLMM. Is this relationship a general result for all link functions?

END

STK3100/STK4100 - Introduction to Generalized Linear Models

Additional information

- In **Problem 3**, $\mathbf{\Sigma_u}$ and $\mathbf{R}_\varepsilon$ can be assumed known.
- Hint for **Problem 3 c**: $E[Y] = E[E[Y|X]]$

APPENDIX: Formulas in STK3100/4100

1) Linear models and least squares

a) Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ be a vector of random variables with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^{\mathrm{T}}$ and covariance matrix $\boldsymbol{V} = E\{(\boldsymbol{Y} - \boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\mu})^{\mathrm{T}}\}$. We consider the linear model $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$, where the model matrix $\boldsymbol{X}$ is a $n \times p$ matrix, and assume that $\boldsymbol{V} = \sigma^2 \boldsymbol{I}$. If we observe $\boldsymbol{Y} = \boldsymbol{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$, then the least squares estimate $\widehat{\boldsymbol{\beta}}$ and the fitted values $\widehat{\boldsymbol{\mu}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ are obtained by minimizing $\|\boldsymbol{y} - \boldsymbol{\mu}\|^2 = (\boldsymbol{y} - \boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{\mu})$.

b) Let $C(\boldsymbol{X})$ denote the model space, i.e. the subspace of $\mathbb{R}^n$ that is spanned by the columns of $\boldsymbol{X}$, and let $\boldsymbol{P_X}$ denote the projection matrix onto $C(\boldsymbol{X})$. Then $\widehat{\boldsymbol{\mu}} = \boldsymbol{P_X}\boldsymbol{y}$. The projection matrix is symmetric and idempotent (i.e. $\boldsymbol{P}_{\boldsymbol{X}}^2 = \boldsymbol{P_X}$), and $\mathrm{rank}(\boldsymbol{P_X}) = \mathrm{trace}(\boldsymbol{P_X})$.

c) The projection matrix $\boldsymbol{P_X}$ is unique, i.e. it depends only on the subspace $C(\boldsymbol{X})$ and not on the choice of basis vectors for the subspace. If $\boldsymbol{X}$ has full rank, we have $\boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}$.

d) For a random vector $\boldsymbol{Y}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{V}$ and a fixed matrix $\boldsymbol{A}$, we have $E(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Y}) = \mathrm{trace}(\boldsymbol{A}\boldsymbol{V}) + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\mu}$.

2) Multivariate normal distribution and normal linear models

a) $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{V}$, written $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{V})$, if its joint pdf is given by

$$f(\boldsymbol{y}; \boldsymbol{\mu}, \boldsymbol{V}) = (2\pi)^{-n/2}|\boldsymbol{V}|^{-1/2} \exp\{-(1/2)(\boldsymbol{y} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\}$$

b) Suppose $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{V})$ is partitioned as

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} \quad \text{with} \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{pmatrix}$$

then

$$\boldsymbol{Y}_1 | \boldsymbol{Y}_2 = \boldsymbol{y}_2 \sim N\left(\boldsymbol{\mu}_1 + \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2), \boldsymbol{V}_{11} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{21}\right)$$

c) [Cochran's theorem] Assume that $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \sigma^2\boldsymbol{I})$ and that $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_k$ are projection matrices with $\sum_{i=1}^{k} \boldsymbol{P}_i = \boldsymbol{I}$. Then $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}_i\boldsymbol{Y}$ are independent for $i = 1, \ldots k$, and $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}_i\boldsymbol{Y}/\sigma^2$ has a non-central chi-squared distribution with non-centrality parameter $\lambda_i = \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{P}_i\boldsymbol{\mu}/\sigma^2$ and degrees of freedom equal to the rank of $\boldsymbol{P}_i$.

3) Generalized linear models (GLMs)

a) A random variable Y_i has a distribution in the exponential dispersion family if its pmf/pdf may be written

$$f(y_i; \theta_i, \phi) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i, \phi)\},$$

where θ_i is the natural parameter and ϕ is the dispersion parameter. We have $E(Y_i) = b'(\theta_i)$ and $\mathrm{var}(Y_i) = b''(\theta_i)a(\phi)$.

b) For a GLM we have that $Y_1, \ldots Y_n$ are independent with pmf/pdf from the exponential dispersion family. The linear predictors $\eta_1, \ldots, \eta_n$ are given by $\eta_i = \sum_{j=1}^{p} x_{ij}\beta_j = \boldsymbol{x}_i\boldsymbol{\beta}$, and the expected values $\mu_i = E(Y_i)$ satisfy $g(\mu_i) = \eta_i$ for a strictly increasing and differentiable link function g. For the canonical link function $g(\mu_i) = (b')^{-1}(\mu_i)$ we have $\theta_i = \eta_i$.

c) The likelihood equations for a GLM are given by

$$\sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{\mathrm{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0 \quad \text{for} \quad j = 1, \ldots, p.$$

d) Let $\widehat{\boldsymbol{\beta}}$ be the maximum likelihood (ML) estimator for a GLM. Then

$$\widehat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X})^{-1}\right), \quad \text{approximately}$$

where $\boldsymbol{X}$ is the model matrix and $\boldsymbol{W}$ is the diagonal matrix with elements $w_i = (\partial\mu_i/\partial\eta_i)^2/\mathrm{var}(Y_i)$.

e) Consider a GLM with $a(\phi) = \phi/\omega_i$. Let $\widehat{\mu}_i = b'(\widehat{\theta}_i)$ be the ML estimate of μ_i under the actual model, and let $y_i = b'(\tilde{\theta}_i)$ be the ML estimate of μ_i under the saturated model. Then

$$-2\log\left(\frac{\text{max likelihood for actual model}}{\text{max likelihood for saturated model}}\right) = D(\boldsymbol{y}; \widehat{\boldsymbol{\mu}})/\phi$$

where

$$D(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n} \omega_i \left[y_i\left(\tilde{\theta}_i - \widehat{\theta}_i\right) - b(\tilde{\theta}_i) + b(\widehat{\theta}_i)\right]$$

is the deviance.

4) Normal and generalized linear mixed models

a) We assume that $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{id})^{\mathrm{T}}$ for $i = 1, \ldots n$ are independent vectors that correspond to d observations from each of n clusters. A normal linear mixed effects model is given by

$$Y_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{u}_i + \epsilon_{ij},$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects, $\boldsymbol{u}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_u)$ is a $q \times 1$ vector of random effects, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots \epsilon_{id})^{\mathrm{T}} \sim N(\boldsymbol{0}, \boldsymbol{R})$ is independent of $\boldsymbol{u}_i$. Often one will have $\boldsymbol{R} = \sigma^2\boldsymbol{I}$.

b) For a generalized linear mixed model we assume that the conditional pmf/pdf of Y_{ij} given $\boldsymbol{u}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_u)$ is in the exponential dispersion family, and that for a link function g we have

$$g\left[E(Y_{ij} \mid \boldsymbol{u}_i)\right] = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{u}_i.$$