UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in:	STK3100/STK4100 — Introduction to Generalized Linear Models SKETCH OF SOLUTION
Day of examination:	Thursday 8th December 2022

This problem set consists of 5 pages.

Appendices:

Permitted aids:

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

Problem 1

a

We can write (1) as

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

which is identical to (2) with $\theta = \log\mu$, $b(\theta) = \mu = \exp(\theta)$, $a(\phi) = 1$ and $c(y,\phi) = -\log y!$.

b

The definition of a generalized linear model (GLM) for $Y_i \sim \text{Poisson}(\mu_i)$ and covariates $x_{ij}, i = 1,\dots n, j = 1,\dots,p$, with $x_{i1} = 1$ to represent the intercept, using the canonical link function

- $Y_1,\dots,Y_N$ are independent with $Y_i \sim \text{Poisson}(\mu_i)$ (i.e. pmf of the form (**??**))
- For each Y_i with covariates $x_{ij}, i = 1,\dots n, j = 1,\dots,p$, we have the linear predictor $\eta_i = \sum_{j=1}^{p}\beta_j x_{ij} = \boldsymbol{x}_i\boldsymbol{\beta}$, for $\boldsymbol{x}_i = (x_{i1},\dots,x_{ip})$ and $\boldsymbol{\beta} = (\beta_1,\dots,\beta_p)^T$
- For each Y_i, the mean $\mu_i = E(Y_i)$ is linked to the linear predictor by the link function $g(\mu_i) = \eta_i$. Here we are using the canonical link function $g(\mu_i) = \theta_i = \log\mu$.

c

The log-likelihood function is

$$L(\boldsymbol{\mu};\mathbf{y}) = \log\left\{\prod_{i=1}^{n}\exp\{y_i\log\mu_i - \mu_i - \log y_i!\}\right\} = \sum_{i=1}^{n}\{y_i\log\mu_i - \mu_i - \log y_i!\} \tag{1}$$

(Continued on page 2.)

The saturated model is the most general model a separate parameter μ_i for each subject i, hence without any restrictions on μ_i. For this model the maximum likelihood fit $\tilde{\mu}_i = y_i$, because $\frac{\partial L(\boldsymbol{\mu};\mathbf{y})}{\partial \mu_i} = \frac{y_i}{\mu_i} - 1$ is $= 0$ for $\tilde{\mu}_i = y_i$. It is useful as a baseline for assessing the quality of fit for other models, as it is the model out of all possible models that achieves the maximum achievable value of the log likelihood $L(\mathbf{y};\mathbf{y})$. It is however not a very useful model for estimation of the underlying truth nor for prediction of new observations, e.g. it overfits the data and does not smooth. The maximum log likelihood for the saturated model is

$$L(\mathbf{y};\mathbf{y}) = \sum_{i=1}^{n} \{y_i \log y_i - y_i - \log y_i!\}$$

d

The deviance $D(\mathbf{y};\hat{\boldsymbol{\mu}})$ for a Poisson GLM is (because $a\phi = \phi = 1$)

$$\begin{aligned} D(\mathbf{y};\hat{\boldsymbol{\mu}}) &= -2\left[L(\hat{\boldsymbol{\mu}};\mathbf{y}) - L(\mathbf{y};\mathbf{y})\right] = -2\left[\sum_{i=1}^{n}\{y_i \log \hat{\mu}_i - \hat{\mu}_i - \log y_i!\} - \sum_{i=1}^{n}\{y_i \log y_i - y_i - \log y_i!\}\right] \\ &= 2\sum_{i=1}^{n}\left\{y_i \log \frac{y_i}{\hat{\mu}_i} - y_i + \hat{\mu}_i\right\} \end{aligned}$$

This deviance may be used as a test statistic to compare nested Poisson GLMs by considering the difference in the deviance for the smaller model and the bigger model. To explain this, assume that the Poisson GLM model M_1 with p_1 parameters holds, and that M_0 with $p_0 < p_1$ parameters is nested in model M_1 (i.e. has the same distribution and link function as M_1, here Poisson with log-link, and the same linear predictor as M_1, but with $p_1 - p_0$ of the β_j's of M_1 set to zero). Let $\hat{\boldsymbol{\mu}}_0$ and $\hat{\boldsymbol{\mu}}_1$ be the maximum likelihood estimates of μ under M_0 and M_1 respectively. For testing the null hypothesis that M_0 holds, we have the likelihood ratio statistics

$$\begin{aligned} G^2(M_0 \mid M_1) &= -2\left(L(\hat{\boldsymbol{\mu}}_0;\mathbf{y}) - L(\hat{\boldsymbol{\mu}}_1;\mathbf{y})\right) = -2\left(L(\hat{\boldsymbol{\mu}}_0;\mathbf{y}) - L(\mathbf{y};\mathbf{y})\right) - \left(-2\left(L(\hat{\boldsymbol{\mu}}_1;\mathbf{y}) - L(\mathbf{y};\mathbf{y})\right)\right) \\ &= D(\mathbf{y};\hat{\boldsymbol{\mu}}_0) - D(\mathbf{y};\hat{\boldsymbol{\mu}}_1) \end{aligned}$$

which is approximately $\mathcal{X}^2_{p_1-p_0}$ distributed if M_0 holds.

Problem 2

In this problem we assume that Y comes from a negative binomial distribution. Here we let the pmf of the negative binomial distribution take the form

$$p(y;\mu,k) = \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)}\left(\frac{\mu}{\mu+k}\right)^{y}\left(\frac{k}{\mu+k}\right)^{k} \;;\; y = 0,1,2,\ldots$$

We will assume that $k > 0$ is a given constant, and consider the random variable $Y^* = Y/k$. Then $P(Y^* = y^*) = P(Y = ky^*)$ for $y^* = 0, \frac{1}{k}, \frac{2}{k}, \ldots$, so Y^* has pmf

$$p^*(y^*;\mu,k) = \frac{\Gamma(ky^*+k)}{\Gamma(k)\Gamma(ky^*+1)}\left(\frac{\mu}{\mu+k}\right)^{ky^*}\left(\frac{k}{\mu+k}\right)^{k} \;;\; y^* = 0, \frac{1}{k}, \frac{2}{k}\ldots \tag{2}$$

(Continued on page 3.)

a

We can write (2) as

$$\begin{aligned} p^*(y^*;\mu,k) &= \exp\log p^*(y^*;\mu,k) \\ &= \exp\left\{\log\Gamma(y+k) - \log\Gamma(k) - \log\Gamma(ky^*+1) + ky^*\log\left(\frac{\mu}{\mu+k}\right) + k\log\left(\frac{k}{\mu+k}\right)\right\} \\ &= \exp\left\{\frac{y^*\log\left(\frac{\mu}{\mu+k}\right) + \log\left(\frac{k}{\mu+k}\right)}{\frac{1}{k}} + c(y,\phi)\right\} = \exp\left\{\left(\theta y^* - b(\theta)\right)/a(\phi) + c(y^*,\phi)\right\} \end{aligned}$$

where $\theta = \log\left(\frac{\mu}{\mu+k}\right)$, $b(\theta) = -\log\left(\frac{k}{\mu+k}\right) = -\log\left(1-\frac{\mu}{\mu+k}\right) = -\log\left(1-e^\theta\right)$ and $a(\phi) = 1/k$ (and $c(y^*,\phi) = \log\Gamma(y+k) - \log\Gamma(k) - \log\Gamma(ky^*+1)$).

b

Mean of Y^*

$$E[Y^*] = b'(\theta) = \frac{e^\theta}{1-e^\theta} = \frac{\frac{\mu}{\mu+k}}{1-\frac{\mu}{\mu+k}} = \frac{\mu}{\mu+k-\mu} = \frac{\mu}{k}$$

Variance of Y^*

$$\mathrm{Var}[Y^*] = b''(\theta)\cdot a(\phi) = \frac{e^\theta(1-e^\theta) - e^\theta(-e^\theta)}{(1-e^\theta)^2}\cdot\frac{1}{k} = \frac{e^\theta}{k(1-e^\theta)^2} = \frac{\frac{\mu}{\mu+k}}{k(1-\frac{\mu}{\mu+k})^2} = \frac{\mu(\mu+k)}{k^3}$$

which gives us

$$E[Y] = E[kY^*] = kE[Y^*] = k\frac{\mu}{k} = \mu$$

and

$$\mathrm{Var}[Y] = \mathrm{Var}[kY^*] = k^2\mathrm{Var}[Y^*] = k^2\frac{\mu(\mu+k)}{k^3} = \mu + \frac{\mu^2}{k}$$

c

If Y is negative-binomially distributed we have $\mathrm{Var}[Y]/E[Y] = 1+\mu/k$, and if Y is Poisson distributed we have $\mathrm{Var}[Y]/E[Y] = 1$. Hence if $k < \infty$, or $\gamma = 1/k > 0$, the variance is larger than the mean for the negative binomial distribution, while they are equal for the Poisson distribution. When you have count data for which it is reasonable to assume equal mean and variance, the Poisson is a good model, while if the variance is assumed to be larger than the mean, there is overdispersion and the negative binomial is a better model than the Poisson.

Problem 3

a

The model behind the analysis is a Poisson GLM with log link. Hence

- The response variables $Y_1, \ldots, Y_N$ are assumed to be independent with $Y_i \sim \text{Poisson}(\mu_i)$ (i.e. pmf of the form (**??**))

(Continued on page 4.)

- For each Y_i with covariates $x_{ij}, i = 1, \dots n, j = 1, \dots, 5$, we have the linear predictor $\eta_i = \sum_{j=1}^{5} \beta_j x_{ij} = \boldsymbol{x}_i \boldsymbol{\beta}$, for $\boldsymbol{x}_i = (x_{i1}, \dots, x_{i5})$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_5)^T$. We have here
  - $x_{i1} = 1, \forall i$ to represent the intercept
  - $x_{i2} = 1$ if `procedure` of patient i is CABG, and $x_{i2} = 0$ if PTCA
  - $x_{i3} = 1$ if patient i is male, and $x_{i3} = 0$ if female
  - $x_{i4} = 1$ if patient i was admitted as an emergency, and $x_{i4} = 0$ if the procedure was pre-planned
  - $x_{i5} = 1$ if the age of patient i is over 75, and $x_{i5} = 0$ if it is less than or equal to 75
- For each Y_i, the mean $\mu_i = E(Y_i)$ is linked to the linear predictor by the link function $g(\mu_i) = \eta_i$. Here we are using the canonical link function $g(\mu_i) = \theta_i = \log \mu$, and hence $\mu_i = E(Y_i) = \exp\left\{\sum_{j=1}^{5} \beta_j x_{ij}\right\}$

b

The p-value of the likelihood ratio test described in Problem 1d for comparing the model with interaction (M_1) to the model without interactions (M_0) is reported to be $< 2.2e - 16$, which is very small, which means that there is reason to reject M_0 and conclude that model M_1 with interactions is to be preferred. It can be commented that this is supported by the fact that the AIC is lower for M_1 than M_0.

To describe the effects of `procedure` and `admit` on the estimated mean length of stay, we consider a female patient ($x_{i3} = 0$) of age less than or equal to 75 ($x_{i5} = 0$). The estimated mean length of stay for the different combinations of `procedure` and `admit` are then

- `procedure` of patient i is PTCA ($x_{i2} = 0$) and the procedure was pre-planned ($x_{i4} = 0$):
$$\exp\left\{\sum_{j=1}^{5} \hat{\beta}_j x_{ij}\right\} = \exp\left\{\hat{\beta}_1\right\} = \exp\{1.23851\} \approx 3.5$$
- `procedure` of patient i is CABG ($x_{i2} = 1$) and the procedure was pre-planned ($x_{i4} = 0$):
$$\exp\left\{\sum_{j=1}^{5} \hat{\beta}_j x_{ij}\right\} = \exp\left\{\hat{\beta}_1 + \hat{\beta}_2\right\} = \exp\{1.23851 + 1.24765\} \approx 12$$
- `procedure` of patient i is PTCA ($x_{i2} = 0$) and the procedure was an emergency ($x_{i4} = 1$):
$$\exp\left\{\sum_{j=1}^{5} \hat{\beta}_j x_{ij}\right\} = \exp\left\{\hat{\beta}_1 + \hat{\beta}_4\right\} = \exp\{1.23851 + 0.61606\} \approx 6.4$$

(Continued on page 5.)

- `procedure` of patient i is CABG ($x_{i2} = 1$) and the procedure was an emergency ($x_{i4} = 1$):

$$\exp\left\{\sum_{j=1}^{5}\hat{\beta}_j x_{ij}\right\} = \exp\left\{\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_4 + \hat{\beta}_6\right\} = \exp\{1.23851 + 1.24765 + 0.61606 - 0.39658\} \approx 15$$

c

The p-value of the test is (close to) 0 and hence the null hypothesis of $\gamma = 0$ should be rejected. This implies that there is overdispersion and the negative binomial is to be preferred over the Poisson. This is supported by the fact that the AIC value for the fit with the negative binomial (19857) is lower than the one for the Poisson (22184).

Problem 4

a

The estimated β_j's are the same, which is not surprising, since they are estimated by the exact same equations. Their estimated standard errors are however different, smaller for the Poisson GLM. This is because the ones for the Poisson are based on an inadequate assumption for the variance function (that it is equal to the mean), while the quasi-likelihood allows for this to be multiplied by ϕ. Hence the estimated standard errors of the estimated β_j's of the quasi-likelihood approach are more robust and reliable.

b

The likelihood equations for a GLM with variance function $\text{var}(Y_i) = v^*(\mu_i)$ are

$$\sum_{i=1}^{n}\frac{(y_i - \mu_i)x_{ij}}{v^*(\mu_i)}\frac{\partial\mu_i}{\partial\eta_i} = 0, \qquad j = 1, \ldots, p$$

The only dependence on the assumed distributon is here through how the variance function $v^*(\mu_i)$ depends on the mean μ_i.

We can allow for over-dispersion and instead assume $\text{var}(Y_i) = v(\mu_i) = \phi v^*(\mu_i)$. Then we have the quasi-likelihood equations

$$\sum_{i=1}^{n}\frac{(y_i - \mu_i)x_{ij}}{\phi v^*(\mu_i)}\frac{\partial\mu_i}{\partial\eta_i} = 0, \qquad j = 1, \ldots, p$$

that are not dependent on any particular distribution, only that $\text{var}(Y_i) = v(\mu_i) = \phi v^*(\mu_i)$ and $g(\mu_i) = g(E[Y_i]) = \eta_i = \boldsymbol{x}_i\boldsymbol{\beta}$ for a link function g. ϕ can be estimated

These equations give estimates of $\boldsymbol{\beta}$ that are (as mentioned in a)) the same as obtained fro the ML estimates for the original GLM (because ϕ cancels from the quasi-likelihood equations), and they are approximately normally distributed with mean $\boldsymbol{\beta}$ and covariance matrix equal to the covariance matrix obtained fro the ML estimates for the original GLM multiplied by ϕ, which can be estimated.