Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4315 Generalized linear models, Solution, December 2019

Problem 1

a) The pmf is

$$\begin{aligned} f_Y(y) &= \exp(y\ln(1-p) - \ln(1-p) + \ln p) \\ &= \exp(y\theta - \theta + \ln(1-e^\theta)) \end{aligned} \tag{1}$$

so it belongs to the expoential family. The canonical parameter is $\theta = \ln(1-p)$ and $b(\theta) = \theta - \ln(1-e^\theta)$.

Via general formulas, we recover the usual expressions for the mean and variance,

$$EY = b'(\theta) = 1 + \frac{e^\theta}{1-e^\theta} = \frac{1}{1-e^\theta} = \frac{1}{p} \tag{2}$$

and

$$\operatorname{Var} Y = \frac{\phi}{w_i} b''(\theta) = \frac{e^\theta}{(1-e^\theta)^2} = \frac{1-p}{p^2}, \tag{3}$$

of the geometric distribution.

b) Using the canonical link function, the log likelihood is

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n y_i\eta_i - b(\eta_i), \tag{4}$$

where $\eta_i = \mathbf{x}_i^T\boldsymbol{\beta}$. Differentiation with respect to $\boldsymbol{\beta}$ leads to

$$\mathbf{s}(\boldsymbol{\beta}) = \sum_{i=1}^n (y_i - \mu_i)\frac{\partial \eta_i}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n (y_i - \frac{1}{p_i})\mathbf{x}_i, \tag{5}$$

and

$$\mathbf{F}(\boldsymbol{\beta}) = \operatorname{Var}(\mathbf{s}(\boldsymbol{\beta})) = \sum_{i=1}^n \frac{1-p_i}{p_i^2}\mathbf{x}_i\mathbf{x}_i^T. \tag{6}$$

In one or several dimensions dimension, Newton's methods search for the root of the equation $f(x) = 0$ by setting a linear approximation of $f(x)$ around our current estimate x_j of the solution equal to zero and solving. This is repeated until convergence.

For optimizations problems (such as finding the solution of $\mathbf{s}(\boldsymbol{\beta}) = \mathbf{0}$) the algorithm takes the form

$$\boldsymbol{\beta}_{j+1} = \boldsymbol{\beta}_j + \mathbf{H}(\boldsymbol{\beta}_j)^{-1}\mathbf{s}(\boldsymbol{\beta}) \tag{7}$$

where $\mathbf{H}(\boldsymbol{\beta}_j) = -\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}$. Fisher scoring differs form this in that $\mathbf{H}(\boldsymbol{\beta})$ is replaced by the expected Fisher information $\mathbf{F}(\boldsymbol{\beta}) = E\mathbf{H}(\boldsymbol{\beta})$ which has the advantage that it is always positive definite and invertible at the interior of the parameter space.

We have the constraint $0 < p_i = 1 - e^{\eta_i} < 1$ which translates to $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} < 0$ for all observations $i = 1, 2, \ldots, n$. Thus, if the model includes an intercept β_0, the parameters need to satisfy the constraint ~~$\beta_0 < \min_{i=1}^n (\beta_1 x_{i1} + \ldots \beta_k x_{ik})$~~ $\beta_0 < \min_{i=1}^n (-\beta_1 x_{i1} - \cdots - \beta_k x_{ik})$.

Thus, an initial parameter vector $\boldsymbol{\beta}_0 = (0, 0, \ldots, 0)^T$ would not work.

Based on asymptotic normality of MLEs, approximate the standard errors of the parameters estimates can be found by taking the square root of the diagonal elements of the inverse of the expected Fisher information at the MLE of $\boldsymbol{\beta}$.

Problem 2

a) The model assumes independent $y_i \sim \text{Poisson}(\lambda_i)$ where $\ln \lambda_i = \beta_0 + \beta_1 \text{altitude}_i + \ln \text{area}_i$ for $i = 1, 2, \ldots, n$.

Keeping everything else constant, the effect of a 100 meter increase in altitude is to change the expected number of plants inside a sampling square by a factor of $\exp(-0.000586 \cdot 100) = 0.94$, that is, the expected number of plants is reduced by about 6%.

b) If the individuals plants are points in a spatial Poisson process within each sampling square, then the total plant count inside each square will follow a Poisson distribution.

Using $\ln \text{area}_i$ as an offset means that

$$E(y_i) = e^{\beta_0 + \beta_1 \text{altitude}_i + \ln \text{area}_i} = e^{\beta_0 + \beta_1 \text{altitude}_i} \text{area}_i \tag{8}$$

such that the resonable a priori assumption of direct proportionality between $E(y_i)$ and area_i has been built into the model.

c) Under the null hypothesis of no overdispersion ($\phi = 1$ ~~$\phi = 0$~~), the deviance D is chi-square with $n - p = 100 - 2 = 98$ degrees of freedom. Large values of D indicate overdispersion so we reject if $D > \chi^2_{0.95,98} = 122.1$. Given the observed value of $D = 235.4$ we thus reject H_0 and conclude that there is overdispersion in the data.

This can be caused by missing covariates, an incorrect link function, or positive covariance between the number of plants in disjoint subareas in the spatial Poisson process.

The dispersion parameter can be estimated by $\hat{\phi} = D/(n-p) = 235.4/98 = 2.40$. For the quasilikelihood model, the corresponding Wald Z statistic for the effect of altitude changes to

$$T = \frac{\hat{\beta}_1}{\sqrt{\hat{\phi}SE(\hat{\beta}_1)}} = \frac{-0.000586}{\sqrt{2.41} \cdot 0.000415} = -0.909. \tag{9}$$

Under H_0, this statistic is approximately student-T with $n-p$ degrees of freedom so we reject if $|T| > t_{0.025,98} = 1.98$. Hence, we cannot rejct $H_0 : \beta_1 = 0$.

d) Conditional in the random effects $\gamma_{01}, \ldots, \gamma_{0m}$ associated with each year $i = 1, 2, \ldots, m$, the model assumes independent $y_{ij}|\gamma_{0i} \sim \text{Poisson}(\lambda_{ij})$ where $\ln \lambda_{ij} = \beta_0 + \beta_1 \text{altitude}_{ij} + \gamma_{0i} + \ln \text{area}_{ij}$ for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n_i$. The random effects $\gamma_{01}, \ldots, \gamma_{0m}$ are iid $N(0, \tau_0^2)$.

The MLE of τ_0^2 is $\hat{\tau}_0^2 = 0.44$.

For $\gamma_{0i} = 0$ and at sea level in a 50 square meter sampling square, an estimate of the conditional expected response becomes

$$E(\widehat{y_{ij}|\gamma_{0i}} = 0) = e^{-2.77 - 0.00125 \cdot 0 + \log(50)} = 3.13 \tag{10}$$

The expected value in a randomly chosen year becomes

$$\begin{aligned} Ey_{ij} &= EE(y_{ij}|\gamma_{0i}) \\ &= Ee^{\beta_0 + \beta_1 \text{altitude}_{ij} + \gamma_{0i} + \log \text{area}_{ij}} \\ &= e^{\beta_0 + \beta_1 \text{altitude}_{ij} + \log \text{area}_{ij}} E(e^{\gamma_{0i}}) \\ &= e^{\beta_0 + \beta_1 \text{altitude}_{ij} + \log \text{area}_{ij}} e^{\tau_0^2/2} \end{aligned} \tag{11}$$

since $e^{\gamma_{0i}} \sim \text{lognormal}(0, \hat{\tau}_0^2)$. An estimate of this unconditional expectation is therefore

$$\widehat{Ey_{ij}} = E(e^{-2.77 - 0.00125 \cdot 0 + \log(50) + .44/2}) = 3.90 \tag{12}$$

e) We want to test the null hypothesis $H_0 : \tau_0^2 = 0$ against the alternative $H_1 : \tau_0^2 > 0$.

The expected value of all components of the score vector $\mathbf{s}(\boldsymbol{\beta}, \boldsymbol{\theta})$ are zero when evaluated at the true parameter values. Thus, under the above H_0 in particular, since the score vector asymptotically has a normal (symmetric) distribution, there is a 50% chance that the MLE of τ_0 under H_1 falls on the boundary of the parameter space such that the corresponding LRT statistic takes a value of 0.

The overall asymptotic distribution of the LRT statistics can thus be seen as a 50-50% mixture of two chi-squares with 0 (a point mass at zero) and 1 degree of freedom respectively.

We want to find a critical value c, such that the probability of rejection equals the nominal level of $\alpha = 0.005$, that is,

$$\alpha = P(\text{LRT} > c) = \underbrace{\frac{1}{2}P(\chi_0^2 > c)}_{=0} + \frac{1}{2}P(\chi_1^2 > c). \tag{13}$$

The critical value c is therefore the upper 2α quantile of the chi-square distribution with 1 degree of freedom, $c = 6.63$.

Given the observed values of $\text{LRT} = 2(-177 - (-225)) = 96$ we can thus reject H_0 in favour of the model including the random intercepts.

f) If we only observe $z_i = 1$ if $y_i \geq 1$ and $z_i = 0$ otherwise, we get the glm where $z_i \sim \text{bin}(1, p_i)$ and

$$p_i = P(y_i \geq 1) = 1 - P(y_i = 0) = 1 - e^{-\lambda_i} = 1 - e^{-e^{\eta_i}} \tag{14}$$

such that

$$\text{cloglog}\, p_i = \ln(-\ln(1 - p_i)) = \eta_i \tag{15}$$

and η_i equals the linear predictor in point d.

Problem 3

a) Using the law of total probability and the assumption of independence between clusters, the likelihood can be written

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{i=1}^{m} \int f(y_i|\boldsymbol{\beta}, \boldsymbol{\gamma}_i) f(\boldsymbol{\gamma}_i|\boldsymbol{\theta}) d\boldsymbol{\gamma}_i. \tag{16}$$

This can be computed via a Laplace approximation or via numerical integration.

For GLMMs, the restricted likelihood can be defined as

$$L_R(\boldsymbol{\theta}) = \int L(\boldsymbol{\beta}, \boldsymbol{\theta}) d\boldsymbol{\beta}, \tag{17}$$

that is, we also integrate out $\boldsymbol{\beta}$ in addition to $\boldsymbol{\gamma}$. This can be done by doing the above Laplace approximation instead with respect to both $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$. As for LMMs, the REML estimators of $\boldsymbol{\theta}$ are often less biased. The estimates can be interpreted as the maximum a posteriori estimates of $\boldsymbol{\theta}$ if a uniform improper prior were used on $\boldsymbol{\beta}$ so in a sense we are dealing with $\boldsymbol{\beta}$ (which are nuisance parameters from the point of view of estimating $\boldsymbol{\theta}$) in a similar way to how nuisance parameters are dealt with in Bayesian inference.

b) It follows that

$$\mathbf{w} = \mathbf{A}^T\mathbf{y} = \mathbf{A}^T(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = \mathbf{A}^T\boldsymbol{\epsilon}. \tag{18}$$

Thus, $\mathbf{w}$ is multinormal with zero mean vector and variance matrix $\mathbf{A}^T\sigma^2\mathbf{I}_n\mathbf{A} = \sigma^2\mathbf{I}_{n-p}$ The restricted likelihood is therefore

$$L(\sigma^2) = (2\pi)^{-(n-p)/2}|\sigma^2\mathbf{I}_{n-p}|^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}\mathbf{w}^T\mathbf{w}\right\} \tag{19}$$

Since $|\sigma^2\mathbf{I}_{n-p}|^{-1/2} = (\sigma^2)^{-(n-p)/2}$ and $\mathbf{w}^T\mathbf{w} = (\mathbf{A}^T\mathbf{y})^T\mathbf{A}^T\mathbf{y} = \mathbf{y}^T\mathbf{A}\mathbf{A}^T\mathbf{y} = \mathbf{y}^T(\mathbf{I}_n - \mathbf{H})\mathbf{y} = ((\mathbf{I}_n - \mathbf{H})\mathbf{y})^T(\mathbf{I}_n - \mathbf{H})\mathbf{y} = (\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}}) = SSE$, the log likelihood simplifies to

$$l(\sigma^2) = -\frac{n-p}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}SSE. \tag{20}$$

Maximising this we find the REML estimator of σ^2,

$$\hat{\sigma}^2 = \frac{1}{n-p}SSE, \tag{21}$$

which equals the usual unbiased estimator of σ^2 for LMs.