AMS 274 – Generalized Linear Models (Fall 2016)

Homework 4 (due Tuesday November 29)

1. The table below reports results from a developmental toxicity study involving ordinal categorical outcomes. This study administered diethylene glycol dimethyl ether (an industrial solvent used in the manufacture of protective coatings) to pregnant mice. Each mouse was exposed to one of five concentration levels for ten days early in the pregnancy (with concentration 0 corresponding to controls). Two days later, the uterine contents of the pregnant mice were examined for defects. One of three (ordered) outcomes ("Dead", "Malformation", "Normal") was recorded for each fetus.

Concentration (mg/kg per day) (x_i)	Response: Dead (y_{i1})	Response: Malformation (y_{i2})	Response: Normal (y_{i3})	Total number of subjects (m_i)
0	15	1	281	297
62.5	17	0	225	242
125	22	7	283	312
250	38	59	202	299
500	144	132	9	285

Build a multinomial regression model for these data using continuation-ratio logits for the response probabilities $\pi_j(x)$, $j = 1, 2, 3$, as a function of concentration level, x. Specifically, consider the following model

$$L_1^{(\mathrm{cr})} = \log\left(\frac{\pi_1}{\pi_2 + \pi_3}\right) = \alpha_1 + \beta_1 x; \quad L_2^{(\mathrm{cr})} = \log\left(\frac{\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x$$

for the multinomial response probabilities $\pi_j \equiv \pi_j(x)$, $j = 1, 2, 3$.

(a) Show that the model, involving the multinomial likelihood for the data $= \{(y_{i1}, y_{i2}, y_{i3}, x_i) : i = 1, ..., 5\}$, can be fitted by fitting separately two Binomial GLMs. Provide details for your argument, including the specific form of the Binomial GLMs.

(b) Use the result from part (a) to obtain the MLE estimates and corresponding standard errors for parameters $(\alpha_1, \alpha_2, \beta_1, \beta_2)$. Plot the estimated response curves $\hat{\pi}_j(x)$, for $j = 1, 2, 3$, and discuss the results.

(c) Develop and implement a Bayesian version of the model above. Discuss your prior choice, and provide details for the posterior simulation method. Provide point and interval estimates for the response curves $\pi_j(x)$, for $j = 1, 2, 3$.

2. Consider the "alligator food choice" data example, the full version of which is discussed in Section 7.1 of Agresti (2002), *Categorical Data Analysis*, Second Edition. Here, consider the subset of the data reported in Table 7.16 (page 304) of the above book. This data set involves observations on the primary food choice for $n = 63$ alligators caught in Lake George, Florida. The nominal response variable is the primary food type (in volume) found in each alligator's stomach, with three categories: "fish", "invertebrate", and "other". The invertebrates were mainly apple snails, aquatic insects, and crayfish. The "other" category included amphibian, mammal, bird, reptile, and plant material. Also available for each alligator is covariate information on its length (in meters) and gender.

   (a) Focus first on length as the single covariate to explain the response probabilities for the "fish", "invertebrate" and "other" food choice categories. Develop a Bayesian multinomial regression model, using the baseline-category logits formulation with "fish" as the baseline category, to estimate (with point and interval estimates) the response probabilities as a function of length. (Note that in this data example, we have $m_i = 1$, for $i = 1, ..., n$.) Discuss your prior choice and approach to MCMC posterior simulation.

   (b) Extend the model from part (a) to describe the effects of both length and gender on food choice. Based on your proposed model, provide point and interval estimates for the length-dependent response probabilities for male and female alligators.

3. Consider the inverse Gaussian distribution with density function

$$f(y \mid \mu, \phi) = (2\pi\phi y^3)^{-1/2} \exp\left\{-\frac{(y-\mu)^2}{2\phi\mu^2 y}\right\}, \qquad y > 0;\;\; \mu > 0, \phi > 0.$$

   Denote the inverse Gaussian distribution with parameters μ and ϕ by $IG(\mu, \phi)$.

   (a) Show that the inverse Gaussian distribution is a member of the exponential dispersion family. Show that μ is the mean of the distribution and obtain the variance function.

   (b) Consider a GLM with random component defined by the inverse Gaussian distribution. That is, assume that y_i are realizations of independent random variables Y_i with $IG(\mu_i, \phi)$ distributions, for $i = 1$,...,n. Here, $g(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}$, where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ $(p < n)$ is the vector of regression coefficients, and $\boldsymbol{x}_i = (x_{i1}, ..., x_{ip})^T$ is the covariate vector for the ith response, $i = 1$,...,n. Define the full model so that the y_i are realizations of independent $IG(\mu_i, \phi)$ distributed random variables Y_i, with a distinct μ_i for each y_i. Obtain the scaled deviance for the comparison of the full model with the inverse Gaussian GLM.