STA 2201 S 2011 Assignment 4

This assignment is about generalized linear models. Lecture material should be all you need, but you might want to look at a textbook treatment too. If so, take a look at a classic, McCullagh and Nelder's Generalized Linear Models. There is a link from the course home page as well as from this document. You could read just pages 21-32 and 40-43 (book's numbering, not the page numbering of the pdf). Or, you can read the whole book if you feel like it.

1. The random variable Y belongs to the exponential family of distributions if its support does not depend upon any unknown parameters and its density or probability mass function takes the form

$$f(y|\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}.$$

   We will assume ϕ is known, so we have a one-parameter exponential family. One thing to remember about the distributions in the exponential family is that it is okay to differentiate under the integral sign (or sum).

   (a) Denoting $\log f(Y|\theta,\phi)$ by $\ell(\theta)$, show $E\left(\frac{\partial \ell}{\partial\theta}\right) = 0$.

   (b) Show $E\left(\frac{\partial^2 \ell}{\partial\theta^2}\right) + E\left(\frac{\partial \ell}{\partial\theta}\right)^2 = 0$.

   (c) Show $E(Y) = \frac{\partial b(\theta)}{\partial\theta}$.

   (d) Show $Var(Y) = a(\phi)\,\frac{\partial^2 b(\theta)}{\partial\theta^2}$.

2. In repeated independent trials, suppose an event occurs with probability p. Let Y represent the number of trails required for the event to happen for the first time. Then Y has a geometric distribution with probability mass function

$$f(y|p) = p(1-p)^{y-1} \text{ for } y = 1, 2, \ldots$$

   (a) Write $f(y|p)$ in the form of the exponential family. In terms of p, what is the natural parameter θ?

   (b) What is the function $b(\theta)$? Express the answer in terms of θ, not p.

   (c) What is $\mu = E(Y)$ in terms of θ?

   (d) What is the variance function $V(\mu)$?

   (e) What is the natural link function $\eta = g(\mu)$?

   (f) Write the likelihood function in terms of the linear predictor $\eta_i = \mathbf{x}_i'\boldsymbol{\beta}$ for $i = 1, \ldots, n$. What is the sufficient statistic?

   (g) I believe that in this case the "natural" link function may be natural and lead to a nice sufficient statistic, but it is very unpleasant and inconvenient. Do you agree that we must have $\eta_i = \mathbf{x}_i'\boldsymbol{\beta} < 0$? This imposes complicated restrictions on the β parameters, and those restrictions depend on the x_i values. Ugh!

   (h) So we will adopt another link function. How about an upside down logit: $\eta = \log\frac{1-p}{p} = \log(\mu - 1)$. Following the recipe for iteratively re-weighted least squares with this link,

i. What is $\widehat{\mu}_i$ in terms of $\widehat{\eta}_i$?

ii. What is z_i?

iii. What is w_i?

(i) The file georeg.data has data for a geometric regression with 3 independent variables. There is also a link from the course home page. Estimate the parameters with a generalized linear model, using the upside down logit link. Your final answer is a set of 4 numbers. Please include a printout showing the R input and output. Just so you will know whether your answers are reasonable, I'll tell you that the data were simulated using the following true parameter values: $\beta_0 = 1, \beta_1 = 1, \beta_2 = -1, \beta_3 = 0$. My answers are reasonably close to these.