Homework 8

Each part of the problems 5 points
Due in class or by email before 9am on Thursday April 7

1. Present each of the following distributions in the exponential family form. Identify the relevant components necessary for use in a GLM: (1) the canonical parameter $\theta$, (2) the dispersion parameter $\phi$, (3) $a(\cdot)$, $b(\cdot)$, $c(\cdot)$, (4) the variance function, (5) the canonical link, $\mu(\theta)$, (6) the deviance and (7) the scaled deviance. Show your work.

(a) Normal distribution
(b) Binomial distribution
(c) Poisson distribution
(d) Gamma distribution $G(\nu, \mu/\nu)$ with density function

$$f(y) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^\nu y^{\nu-1} e^{-y\nu/\mu}, \ y > 0$$

(e) Inverse Gaussian distribution $IG(\mu, \lambda)$ with density function

$$f(y) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp\left\{ -\frac{\lambda(y-\mu)^2}{2\mu^2y} \right\}, \ y, \mu, \lambda > 0$$

2. [Methods qualifying exam, August 2010: use paper and pencil.] Data are generated for the exponential distribution with density $f(y) = \lambda \exp(-\lambda y)$, where $\lambda, y > 0$. The distribution is a member of the exponential family which takes the general form

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

(a) Identify the specific form of $\theta$, $\phi$, $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ for the exponential distribution.
(b) What’s the canonical link and variance function for a generalized linear model (GLM) with a response following the exponential distribution?
(c) Identify a practical difficulty that may arise when using the canonical link in this instance.
(d) When comparing nested models in this case, should an $F$ or $\chi^2$ test be used? Explain.
(e) Express the deviance in terms of the responses $y_i$ and the fitted values $\hat{\mu}_i$. 

1
3. [Methods qualifying exam, January 2008: use paper and pencil.] The Conway-Maxwell-Poisson distribution has the probability function

\[ Y(Y = y) = \frac{\lambda^y}{(y!)^\nu} \frac{1}{Z(\lambda, \nu)}, \quad y = 0, 1, 2, \ldots \]

where

\[ Z(\lambda, \nu) = \sum_{i=1}^{\infty} \frac{\lambda^i}{(i!)^\nu} \]

(a) Place this distribution in an exponential family form with respect to both parameters, and identify all the relevant components.

(b) Explain why this distribution can be used to model overdispersion for count data.

4. [Methods qualifying exam, August 2007: use paper and pencil.] Consider a study intended to investigate race discrimination in calling fouls by referees in National Basketball Association (NBA). \(n\) black referees and \(n\) white referees were randomly selected (\(n > 0\)). For each referee, \(K\) foul calls were randomly selected to count how many calls were given to black players and how many calls were given to white players. Therefore, we have the dataset \(\{(Y_i, X_i), \ i = 1, \ldots, n\}\), where \(Y_i\) is the number of foul calls given to black players by \(i\)th referee, and \(X_i\) indicates whether the \(i\)th referee is black.

(a) Please construct a generalized linear model to serve the study goal:

i. Specify \(\theta_i, \phi, b(\theta_i), a\) and \(c(y_i)\) which defines the exponential family of distributions of \(Y_i\)

\[ P(Y_i = y_i|X_i) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}; \]

ii. State the canonical link function;

iii. State the variance function.

(b) Using the canonical link function, suppose we have \(\hat{\theta}_i = 1.2 - 0.5X_i\) estimated from the data. Please interpret the result.

(c) Suppose \(K = 1\), i.e., only one foul call by each referee was recorded. With this dataset, is it possible for us to build up a model to predict the referee’s race for a foul call on a black player? Please explain.