Generalized Linear Models

Homework Assignment 1: Spring 2003

1. Suppose $Y \sim Ga(\mu, \phi)$. Note that Y has an exponential distribution when the $\phi = 1$, and that the density of Y/μ does not depend on the value of μ.

   (a) Evaluate the normal approximation to the distribution of Y by plotting the density of $(Y - \mu)/\sqrt{\phi}\mu$ for various values of ϕ.

   (b) Derive the moment generating function of $X = \log Y$ and hence find formulas for $E(X)$ and $\mathrm{var}(X)$.

   (c) Plot the density of $(X - E(X))/\sqrt{\mathrm{var}(X)}$ for the values of the dispersion parameter considered in part (a). Hence evaluate the normal approximation to the distribution of X.

   (d) Plot the density of the signed (and scaled) likelihood root statistic

   $$r = \mathrm{sign}(Y - \mu)\sqrt{\frac{2}{\phi}\left\{-\log\frac{Y}{\mu} + \frac{Y - \mu}{\mu}\right\}}$$

   for the values of the dispersion parameter considered in part (a). Hence evaluate the normal approximation to the distribution of r.

2. Let $X_1, \ldots, X_n$ be an i.i.d. sample from the negative binomial distribution, $Nb(\lambda, p)$, and let $\theta = \log p$

   (a) Find the m.l.e. for θ when λ is known.

   (b) Derive the m.l. estimating equation for λ when θ is known.

   (c) Describe an algorithm for finding the joint m.l.e. $(\hat{\theta}, \hat{\lambda})$.

   (d) The number of species of fish in 70 lakes around the world are given in the dataset "Fish counts" available from the class website. Test the goodness-of-fit of the Poisson distribution to these counts.

   (e) Write a computer program to fit a negative binomial distribution to the counts. Email your program to me in a form that I can run. The output from the program should include the results of each iteration.

3. This problem concerns the Inverse Gaussian distribution. Let Φ denote the standard normal CDF and consider the function

$$F(y) = \begin{cases} 0 & y \le 0 \\ \Phi\left(\sqrt{\frac{\lambda}{y}}\left(-1+\frac{y}{\mu}\right)\right) + e^{2\lambda/\mu}\Phi\left(-\sqrt{\frac{\lambda}{y}}\left(1+\frac{y}{\mu}\right)\right) & y > 0\,. \end{cases}$$

   (a) Show that

$$F'(y) = \begin{cases} 0 & y \le 0 \\ \left(\frac{\lambda}{2\pi y^3}\right)^{1/2} \exp\left\{-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\right\} & y > 0\,. \end{cases}$$

   Hence show that F is a continuous CDF with corresponding density $f(y) = F'(y)$.

   (b) Show that the density can be written in exponential dispersion form. Identify the function, $b(\theta)$, the relationship between the canonical and mean parameters, and the variance function for this model.

4. Write a computer program to fit the logistic regression model

$$\operatorname{logit} \pi = \beta_0 + \beta_1 x$$

   to the space shuttle data available from the class website using the IRLS algorithm described in class. In this model x denotes the temperature at launch time and $\pi = P(ndo > 0)$, where ndo is the number of damaged o-rings.

   (a) Email your program to me in a form that I can run. The output from the program should include the results of each iteration and the Wald and LR statistics for testing $H_o : \beta_1 = 0$.

   (b) Explain the criterion you used to determine convergence.

   (c) Plot the fitted value of π as a function of temperature and mark the estimated value for $x = 32^o$ (the temperature the morning of the Challenger disaster).