AMS 274 - Generalized Linear Models (Winter 2008)

Homework 2 (due Tuesday 29 January)

1. Consider the special case of the Cauchy distribution, $C(\theta, 1)$, with scale parameter $\sigma = 1$, and density function

$$f(y; \theta) = \frac{1}{\pi(1 + (y - \theta)^2)}, \quad y \in \mathbb{R}, \theta \in \mathbb{R}.$$ 

Here $\theta$ is a location parameter, in fact, the median of the distribution.

(a) Let $y = (y_1, \ldots, y_n)$ be a random sample from the $C(\theta, 1)$ distribution. Develop the Newton-Raphson method and the method of scoring to approximate the maximum likelihood estimate of $\theta$ based on the sample $y$.

(Hint: For the method of scoring, you can use the result $\int_0^\infty (1 - x^2)/((1 + x^2)^3)dx = \pi/8$.)

(b) Consider a sample, assumed to arise from the $C(\theta, 1)$ distribution, with $n = 9$ and $y = (-0.774, 0.597, 7.575, 0.397, 0.865, 0.318, 0.125, 0.961, 1.039)$. Apply both methods from (a) to estimate $\theta$. To check your results, try a few different starting values and also plot the likelihood function for $\theta$.

(c) Now consider a sample (again, assumed to arise from the $C(\theta, 1)$ distribution) with $n = 3$ and $y = (0, 5, 9)$. Apply again the methods from (a) to estimate $\theta$, using three different starting values, $\theta^0 = -1$, $\theta^0 = 4.67$, $\theta^0 = 10$. Comment on the results (as in (b), plot the likelihood function for $\theta$).

2. The data in the table below show the number of cases of AIDS in Australia by date of diagnosis for successive 3-months periods from 1984 to 1988.

<table>
<thead>
<tr>
<th>Year</th>
<th>Quarter 1</th>
<th>Quarter 2</th>
<th>Quarter 3</th>
<th>Quarter 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>1</td>
<td>6</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>1985</td>
<td>27</td>
<td>39</td>
<td>31</td>
<td>30</td>
</tr>
<tr>
<td>1986</td>
<td>43</td>
<td>51</td>
<td>63</td>
<td>70</td>
</tr>
<tr>
<td>1987</td>
<td>88</td>
<td>97</td>
<td>91</td>
<td>104</td>
</tr>
<tr>
<td>1988</td>
<td>110</td>
<td>113</td>
<td>149</td>
<td>159</td>
</tr>
</tbody>
</table>

Let $x_i = \log i$, where $i$ denotes the time period for $i = 1, \ldots, 20$. Consider a GLM for this data set based on a Poisson response distribution with mean $\mu$, systematic component $\beta_1 + \beta_2 x_i$, and logarithmic link function $g(\mu) = \log(\mu)$.

(a) Fit this GLM to the data working from first principles, that is, derive the expressions that are needed for the scoring method, and implement the algorithm to obtain the maximum likelihood estimates for $\beta_1$ and $\beta_2$.

(b) Use function “glm” in R to verify your results from part (a).
3. Consider the data set from:
on the incidence of faults in the manufacturing of rolls of fabric. The first column contains the
length of each roll (the covariate with values $x_i$), and the second contains the number of faults
(the response with means $\mu_i$).
(a) Use R to fit a Poisson regression model, with a logarithmic link,

$$\log(\mu_i) = \beta_1 + \beta_2 x_i$$

(1)

to explain the number of faults in terms of length of roll.
(b) Fit the regression model for the response means in (1) using the quasi-likelihood estimation
method, which allows for a dispersion parameter in the response variance function. (Use the
quasipoisson “family” in R.) Discuss the results.
(c) Derive point estimates and asymptotic interval estimates for the linear predictor, $\eta_0 = \beta_1 + \\
\beta_2 x_0$, at a new value $x_0$ for length of roll, under the standard (likelihood) estimation method
from part (a) and also under the quasi-likelihood estimation method from part (b). Evaluate
the point and interval estimates at $x_0 = 500$ and $x_0 = 995$.
(Under both cases, use the asymptotic bivariate normality of the vector $(\hat{\beta}_1, \hat{\beta}_2)$ to obtain the
asymptotic distribution of $\hat{\eta}_0 = \hat{\beta}_1 + \hat{\beta}_2 x_0$.)