DEPARTMENT OF STATISTICS
Paper STATS 765 Special Topic in Regression 2006

Assignment 6. Due Friday October 20.

Instructions: hand in a solution to all three problems.

1. Suppose we have a model with a continuous response y and two factors A and B, each having 3 levels. Each factor level combination is measured twice. Carefully explain how R constructs the model matrix (i.e. the matrix of dummy variables) for each of the models

   (a) `y~A*B`;

   (b) `y~A+B`.

   In each case, illustrate your answer with a numerical example. You may find the R function `model.matrix` useful. [*10 marks*]

2. Draw a contour plot of the logistic regression log-likelihood for the CHD data. Mark the position of the maximum on the plot. [*10 marks*]

3. Suppose we have a contingency table with m cells, and we want to fit a model to the table.

   (a) The "multinomial" model models the cell counts $y_1, \ldots, y_m$ as a multinomial distribution, with parameters $n = \sum_{i=1}^m y_i$ and $\pi_i,\ i =, \ldots, m$. In terms of these parameters, the log-likelihood for this model is

   $$l_M = \sum_{i=1}^{m} y_i \log(\pi_i).$$

   Suppose we parameterise the probabilities π_i with $m-1$ parameters $\alpha_2, \ldots, \alpha_m$, by setting

   $$\begin{aligned} \pi_1 &= \frac{1}{1 + \sum_{i=2}^m e^{\alpha_i}}, \\ \pi_i &= \frac{e^{\alpha_i}}{1 + \sum_{i=2}^m e^{\alpha_i}}, i = 2, \ldots, m. \end{aligned}$$

   Show that, in terms of the new parameters, the log-likelihood is

   $$l_M(\alpha_2, \ldots, \alpha_m) = \sum_{i=2}^{m} y_i \alpha_i - n \log\left(1 + \sum_{i=2}^{m} e^{\alpha_i}\right).$$

   [*4 marks*]

(b) Calculate the derivative of l_M and hence show that the MLE's of $\alpha_2, \ldots, \alpha_m$ are the solution to the equations

$$y_i = n \frac{e^{\alpha_i}}{(1 + \sum_{i=2}^m e^{\alpha_i})}, i = 2, \ldots, m. \tag{1}$$

[*4 marks*] Hint: Recall that the derivatives are zero at the maximum.

(c) The "Poisson" model models the cell counts $y_1, \ldots, y_m$ as independent Poisson distributions, with means $\mu_i, \ i = 1, \ldots, m$. In terms of these parameters, the log-likelihood for this model is

$$l_P = \sum_{i=1}^m \{y_i \log(\mu_i) - \mu_i\}.$$

Suppose we re-parameterise the means with m parameters $\mu, \ \alpha_2, \ldots, \alpha_m$, by setting

$$\begin{aligned} \mu_1 &= \exp(\mu), \\ \mu_i &= \exp(\mu + \alpha_i), \ i = 2, \ldots, m. \end{aligned}$$

Show that, in terms of the new parameters, the log-likelihood can be written

$$l_P(\mu, \alpha_2, \ldots, \alpha_m) = n\mu - e^\mu + \sum_{i=2}^m (y_i \alpha_i - e^{\mu + \alpha_i}).$$

[*4 marks*]

(d) Calculate the derivative of l_P and hence show that the MLE's of $\alpha_2, \ldots, \alpha_m$ for the Poisson model are the same as those for the multinomial model.[*4 marks*]

(e) Use the "death by falling" example to verify numerically that the coefficients calculated in R using the Poisson model are the same as the solutions to (1). [*4 marks*]