HW 4 for Stat 7249 - Spring 2009

Due March 19

Reading in text for this assignment

- Chapter 8

Datasets

- posted on class web page

1. Show that the complementary log-log model discussed in class (i.e., the Proportional hazards model) is equivalent to the 'continuation-ratio' model given by

$$g\{\pi_j(x)/(1-\gamma_{j-1}(x))\} = \alpha_j - \beta^T x. \tag{1}$$

   if $g(\cdot)$ is the complementary log-log link. Also, express α_j in terms of the cutpoints $\theta_1, \ldots, \theta_{k-1}$ appearing in the PH model.

2. Consider the proportional odds model,

$$\text{logit}\gamma_j(x_i) = \theta_j - \beta x_i$$

   with x and β both scalars. Denote by $\hat{\theta}_j$ and $\hat{\pi}_j$ the fitted parameters and probabilities under the hypothesis that $\beta = 0$. Show that the derivative of the log likelihood with respect to β at $\beta = 0$, $\theta_j = \hat{\theta}_j$, is given by

$$T = \sum R_{ij} x_i s_j$$

   where $R_{ij} = Y_{ij} - m_i \hat{\pi}_j$ is the residual under independence and $s_j = \hat{\gamma}_j + \hat{\gamma}_{j-1}$. [Note: This is the score test of the hypothesis that $\beta = 0$ in the PO model].

3. Consider the adjacent categories logit model,

$$L_j = log(\pi_j(x)/\pi_{j+1}(x)) = \alpha_j + \beta^T x. \tag{2}$$

   L_j is equal to the logit of what probability? How is this model related to the baseline-category logit model for ordinal data? Equate the parameter of L_j to the parameters in the baseline category logit model. Based on your answer, can we use software to fit the baseline category logit model to fit the adjacent category logit model? Explain.

4. Show that the multinomial distribution for sample size n and parameters $\pi_j, j = 1, \ldots, k$ is in the $(k-1)-$parameter exponential family with the baseline category logits as natural parameters.

5. Consider the following dataset (mental.dat) on mental impairment, which was discretized into four ordered categories (well, mild symptom formation, moderate symptom formation, impaired). It was of interest to determine how this impairment was related to the covariates socioeconomic status (high or low) and a life events index (composite measure of both the number and severity of important life events). Consider both the proportional odds and proportional hazards models. Find the best fitting model (in terms of covariates and the link function). Can we use the deviance to assess goodness of fit here? Explain. Compute the predicted probability of moderate symptom formation for a high SES individual with a life events index values of 5. Also, compute a 95% confidence interval for this probability. Think carefully about the best way to construct this interval for 'optimal' accuracy. Another way to assess the fit of these models is to fit a more complex model which has the model of interest nested within it and then do a LR test. Suppose we fit the following model

$$logit\{\gamma_j(x)\} = \alpha_j - \beta_j^T x. \tag{3}$$

Will there be any complications in fitting this model? Explain. Finally, using your result from Problem 3, fit the adjacent categories logit model to this data. How does this fit relative to the PO and PH models? Give evidence.

6. Consider the following dataset on the primary food choice of alligators in Florida (alligator.dat). The response, primary food choice, was broken into five categories, fish, invertebrate, reptile, bird, other. It was of interest to determine how their primary food choice was related to the size of the alligator (big or small), the gender, and the lake of residence (hancock, oklawaha, trafford, george). Fit baseline category logit models to this data. Which model fits best? Does your best fitting model provide a good fit? Give pertinent evidence. Derive and compute a confidence interval for the difference in probabilities of primary food choice bird vs. primary food choice reptile for a small, female alligator in lake hancock. Think carefully about how to compute this interval.