MA 576 – Qualifying Exam

Spring 2013

1. An exponentially *tilted* density f based on a base density f_0 and with natural parameter θ is given by

$$f(y;\theta) = \frac{\exp(\theta y) f_0(y)}{M(\theta)},$$

where $M(\theta)$ is the moment generating function of f_0. Take f_0 to be the hyperbolic cosecant density

$$f_0(y) = \frac{y}{2\sinh(\pi y/2)}, \quad -\infty < y < \infty,$$

with $\log M(\theta) = -2\log\cos\theta$ (for $|\theta| < \pi/2$.)

   (a) Verify that the distribution f generated by f_0 belongs to an exponential family. What are the dispersion parameter and the cumulant function $b(\theta)$ of f?

   (b) Find the mean and variance of this distribution. Express the variance as a function of the mean.

   (c) What is the canonical link for a GLM from this distribution?

   (d) Write down an expression for the deviance associated with this distribution.

2. The following dataset contains tumor responses of male and female patients receiving two treatment regimes for lung cancer (Holtbrugger and Schumacher, 1991.)

Treatment	Sex	Progressive disease	No change	Partial remission	Complete remission
Sequential	Male	28	45	29	26
	Female	4	12	5	2
Alternating	Male	41	44	20	20
	Female	12	7	3	1

Since the responses in the table are ordinal with $K = 4$ categories, a proportional odds model seems adequate. The predictors $\mathbf{x}$ are factors for sex with levels "male" and "female" and for treatment with levels "sequential" and "alternating". Under this formulation, the cumulative probabilities $\gamma_j = \mathbb{P}(Y \le j \,|\, \mathbf{x})$ are given by

$$\log\left(\frac{\gamma_j}{1-\gamma_j}\right) = \theta_j - \mathbf{x}^\top\boldsymbol{\beta} = \theta_j - \beta_m - \beta_s, \qquad j = 1, \ldots, K-1,$$

where β_m is the contrast for males and β_s is the contrast for the "sequential" treatment.

   (a) Here is a summary output from fitting the proportional odds model above. How would you interpret the coefficients θ_j, β_m and β_s? In particular, which treatment seems to give better results? Explain (*Hint*: use a latent variable representation.)

```
Call:
polr(formula = response ~ sex + treatment, data = tumor)

Coefficients:
                     Value Std. Error t value
sexmale             0.5414     0.2872   1.885
treatmentsequential 0.5807     0.2121   2.737

Intercepts:
                                    Value   Std. Error t value
progressivedisease|nochange        -0.1960  0.2893    -0.6774
nochange|partialremission           1.3713  0.3000     4.5706
partialremission|completeremission  2.4221  0.3224     7.5119
```

(b) Give point estimates for the probability of response being "progressive disease" for the four combinations between sex and treatment, namely, for males undertaking sequential and alternating treatment and for females under sequential and alternating treatment.

(c) Compute a 95% confidence interval for β_m using a Wald statistic. Based on this interval, do you see evidence of a difference in responses based on sex?

(d) The following analysis of deviance table compares three models with different predictors: (1) sex only; (2) sex and treatment; and (3) sex, treatment, and their interaction. Based on this output: Is there evidence for a difference in tumor responses due to treatments? How would you assess the adequacy of a simpler model without an interaction effect between sex and treatment? To answer these questions formally conduct hypothesis tests; state your test statistic, its distribution under the null, and your conclusion at which significance level.

```
Likelihood ratio tests of ordinal regression models

Response: response
            Model Resid. df Resid. Dev   Test   Df LR stat.    Pr(Chi)
1             sex       295   796.6268
2 sex + treatment       294   789.0566 1 vs 2    1 7.570185 0.00593417
3 sex * treatment       293   788.0098 2 vs 3    1 1.046787 0.30624834
```

(e) How would the model change if the response categories were reversed, that is, if "progressive disease" were recorded as "complete remission" and vice-versa, and if "no change" were recorded as "partial remission" and vice-versa? List the parameter estimates for θ_j, β_m and β_s with the reversed responses; would your tests and conclusions above change? Explain.