Project 2 of ST 755, Spring 2008

Due: Thursday, 3/20/2008

1. Consider the following special linear mixed model

$$Y_{ij} = \beta_0 + b_i + e_{ij} \quad i = 1, ..., m, \quad j = 1, 2, ..., n,$$

where $b_i \sim \mathrm{N}(0, \sigma_b^2)$ and $e_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_e^2)$, independent of b_i. Do the following:

(a) Show that

$$\Phi(\theta) = (X^T V^{-1} X)^{-1} = (n\sigma_b^2 + \sigma_e^2)/(mn).$$

(b) Find the asymptotic variance matrix for the MLEs of (σ_b^2, σ_e^2). Note that you can write

$$(\sigma_b^2 11^T + \sigma_e^2 I)^{-1} = \sigma_e^{-2}(I - r11^T),$$

where $r = \sigma_b^2/(n\sigma_b^2 + \sigma_e^2)$.

(c) Using Satterthwaite approach, find the denominator degrees of freedom for testing the hypothesis $H_0 : \beta_0 = c$.

2. Consider a special case of Poisson regression. Suppose there are m subjects in the study sample with n observations per subject. The response y_{ij} is assumed to be a realization from the Poisson distribution with mean μ_{ij} that satisfies

$$\log(\mu_{ij}) = \alpha_0 + x_j \alpha_1 \quad i = 1, ..., m, \quad j = 1, 2, ..., n.$$

Although the data from the same individuals are possibly correlated, we can still obtain consistent estimates of α_0 and α_1 by ignoring the correlation (this approach is often called GEE, see Diggle, Liang and Zeger, for example). Denote by $\hat{\alpha}_0$, $\hat{\alpha}_1$ the estimates of α_0 and α_1 by assuming the data are independent. Show that $\hat{\alpha}_0$, $\hat{\alpha}_1$ satisfy the following equation

$$\sum_{j=1}^{n} (\bar{y}_{\cdot j} - e^{\alpha_0 + x_j \alpha_1}) = 0,$$
$$\sum_{j=1}^{n} x_j (\bar{y}_{\cdot j} - e^{\alpha_0 + x_j \alpha_1}) = 0.$$

3. Now consider a GLMM for the same data given in the previous problem. Conditional on $b_i \sim \mathrm{N}(0, \theta)$, y_{ij} are assumed to be independent observations from the Poisson distribution with mean μ_{ij}^b that satisfies

$$\log(\mu_{ij}) = \beta_0 + x_j\beta_1 + b_i, \quad i = 1, ..., m, \quad j = 1, 2, ..., n.$$

Let us consider the PQL estimates $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{b}_i$ of β_0, β_1 and b_i for any estimate of θ. Let $\hat{\beta}_0^* = \hat{\beta}_0 + \log(\frac{1}{m}\sum_{i=1}^m e^{\hat{b}_i})$. Show that $(\hat{\beta}_0^*, \beta_1)$ satisfy the same equation as $(\hat{\alpha}_0, \hat{\alpha}_1)$ and hence show that $\hat{\beta}_1 = \hat{\alpha}_1$.

REMARK: The above result is very important. We know that PQL estimates are generally biased and GEE estimates are consistent. The above result indicates that in the special case we considered here, the PQL estimate of the regression coefficient is **consistent**!