Statistics 600 Problem Set 6

Due Friday, December 10th at midnight

1. Suppose we have a $n \times p$ design matrix X and coefficient vector $\beta \in \mathcal{R}^p$. Let $\eta = X\beta$ be the linear predictor and set $\mu_i = 1/(1 + \exp(-\eta_i))$. Let $v_i = \lambda\mu_i(1 - \mu_i)$ where $\lambda \in \mathcal{R}$ and let $y_i \in [0, 1]$ follow a beta distribution with mean μ_i and variance v_i. The Wikipedia page on the beta distribution has a sidebar that contains the moment information you will need to do this. (i) Show that any value $\lambda < 1$ can be achieved, (ii) use logistic regression to model the conditional distribution of y given X – explore the bias of $\hat{\beta}$ and assess the validity of the GLM standard errors; (iii) use simulation to assess whether the Pearson χ^2 estimate of the scale parameters allows us to correctly identify the variance function; (iv) make scatterplots of the squared and absolute residuals against the fitted values, including a loess curve on the plot to further aid in evaluating the variance structure.

2. Let X be a $n \times p$ design matrix and let $\beta \in \mathcal{R}^p$ be a coefficient vector. Let $\eta = X\beta$ be the linear predictor and let $\mu = \exp(\eta)$ be the conditional means. We wish to simulate quasi-Poisson data that has mean μ and (over-dispersed) variance $\lambda\mu$ where $\lambda > 1$. Moreover, we wish these data to be correlated within given groups. To do this, use a copula approach as follows. First, for group i of size n_i, sample z_i from a multivariate normal distribution with exchangeable correlation and unit variances. Second, transform z_i to u_i having marginal uniform distributions using the standard normal CDF. Third, transform u_i using a Gamma quantile function with the proper mean and variance to produce the responses y for group i. (i) Use simulation to understand the correlation coefficients among the y values within a common group. (ii) Use Poisson GEE to model the data, assess the bias of $\hat{\beta}$ and assess the validity of the GEE standard errors. (iii) Assess the accuracy of the GEE estimate of the exchangeable correlation coefficient (the ICC). (iv) Assess the accuracy of the Pearson χ^2 estimate of the scale parameter.

3. Let X be a $n \times p$ design matrix, and let $\beta \in \mathcal{R}^p$ be a coefficient vector. Let $\eta = X\beta$ be the linear predictor and let $\mu = \exp(\eta)$ be the conditional mean function. Let the responses y_i be independent with $E[y_i] = \mu_i$ and $\mathrm{var}[y_i] = 1$. Choose β such that the probability

of observing $y_i \leq 0$ is extremely small. Analyze the data two different ways: first use OLS to regress $\log(y)$ on X. Second, use a Gaussian GLM with a (non-canonical) log link function. Assess the bias and precision of $\hat{\beta}$ in each case, and compare the validity of the standard errors in the two regression approaches.