Chapter 7

Likelihood Models for Repeated Binary Data

In Section 6.1 we showed the close correspondence between exponential family likelihood theory and quasi-likelihood for generalized linear models in the univariate setting. In the multivariate case, it is not possible to completely generalize the GLM theory to likelihoods in a way which is entirely satisfactory for non-normal data. Here we will deal with some approaches to constructing likelihood models for repeated binary data. To fix ideas, we will first consider a single sample, no covariates, with $n_i = n$. We also drop the subscript i for much of the discussion for simplicity.

With measured multivariate responses, likelihood based analyses are invariably based on the Multivariate Normal (MVN) distribution, although how the mean and variance are parameterized may differ for different models. The MVN has many attractive features:

1. In the general case, the distribution is indexed by n parameters for the mean μ and $n(n+1)/2$ parameters for the variance-covariance matrix Σ.

2. Any subset of the n-vector Y, say Y_s, also has a multivariate normal with μ_s and Σ_s being the corresponding subsets of (μ, Σ). This property is known as reproducibility.

3. The MVN ensures consistency of ML estimates of μ and Σ, *even if* MVN does not hold. In particular, if $E(Y) = X\beta$ and $V(Y) = \Sigma(\alpha)$, then the ML estimates of (β, α) will be consistent even if Y is *not* MVN.