Longitudinal data in a Generalized Linear Model

Two approaches to handling dependence in Generalized Linear Models:

**Marginal models:** based on the *consequences* of dependence on estimating model parameters.

- Target of inference is the *population*.

**Random effect models:** based on the *sources* of dependence.

- Target of inference is the *subject*. 
• Notation:

\[ Y_{i,j} = j^{th} \text{ measurement on } i^{th} \text{ subject,} \]
made on occasion \( t_{i,j}, \; j = 1, 2, \ldots, n_i, \; i = 1, 2, \ldots, N. \)

• As a vector:

\[
Y_i = \begin{pmatrix}
Y_{i,1} \\
Y_{i,2} \\
\vdots \\
Y_{i,n_i}
\end{pmatrix}
\]
Covariates:

- Associated with $Y_{i,j}$: covariates $X_{i,j,k}, k = 1, 2, \ldots, p$

$$X_{i,j} = \begin{pmatrix} X_{i,j,1} \\ X_{i,j,2} \\ \vdots \\ X_{i,j,p} \end{pmatrix}$$

- Matrix of covariates ($n_i \times p$):

$$X_i = \begin{pmatrix} X'_{i,1} \\ X'_{i,2} \\ \vdots \\ X'_{i,n_i} \end{pmatrix}, i = 1, 2, \ldots, N.$$
Specification of a Marginal Model

- Link function:
  \[ g(\mu_{i,j}) = \eta_{i,j} = \beta_1 X_{i,j,1} + \beta_2 X_{i,j,2} + \cdots + \beta_p X_{i,j,p}, j = 1, 2 \ldots, n_i. \]
  or
  \[ g(\mu_{i,j}) = X'_{i,j} \beta. \]

- Variance function:
  \[ \text{Var}(Y_{i,j}) = \phi v(\mu_{i,j}). \]

- Within-subject association.
Within subject association

- Assume:

\[
\text{Cov}(Y_i) = V_i = A_i^{\frac{1}{2}} \text{Corr}(Y_i) A_i^{\frac{1}{2}}
\]

where

\[
A_i = \phi \begin{pmatrix}
  v(\mu_{i,1}) & 0 & \ldots & 0 \\
  0 & v(\mu_{i,2}) & \ldots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \ldots & v(\mu_{i,n_i})
\end{pmatrix}.
\]

- Corr\((Y_i)\) may be unstructured or patterned.

- No actual distribution for \(Y_i\) is specified.
Estimating Equations

- If responses were independent from the corresponding distribution, ML equations are:

\[
\sum_{i=1}^{N} D_i A_i^{-1} (y_i - \mu_i) = 0
\]

where \( D_i \) is the gradient matrix \( \partial \mu_i/\partial \beta \).

- By analogy with GLS, with dependent responses, use generalized equations

\[
\sum_{i=1}^{N} D_i' V_i^{-1} (y_i - \mu_i) = 0
\]
Numerical Solution:

- $\mu_i$ is a (non-linear) function of $\beta$;

- $V_i$ depends on $\mu_i$, and hence $\beta$, and also the response correlations.

Alternating iteration:

- hold $V_i$ fixed, and solve for $\beta$;

- estimate $V_i$ from current standardized residuals

\[ e_{i,j} = \frac{Y_{i,j} - \hat{\mu}_{i,j}}{\sqrt{v(\hat{\mu}_{i,j})}}. \]
Properties:

• $\hat{\beta}$ is consistent for $\beta$.

• $\hat{\beta}$ is asymptotically normal with mean $\beta$ and

$$\text{Cov} (\hat{\beta}) = B^{-1}MB^{-1},$$

where

$$B = \sum_{i=1}^{N} D'_i \mathbf{V}_i^{-1} D_i,$$

$$M = \sum_{i=1}^{N} D'_i \mathbf{V}_i^{-1} \text{Cov} (Y_i) \mathbf{V}_i^{-1} D_i,$$
Variance specification:

- If $V_i$ is modeled correctly, i.e. $\text{Cov}(Y_i) = V_i$, then $B = M$ and
  \[ \text{Cov}(\hat{\beta}) = B^{-1}, \]
  the *model-based* estimator.

- Otherwise, estimate $B$ and $M$ separately, and use the “sandwich” or *empirical* estimator.
The Canonical Link:

- Above approach works for any link function;

- In general,

$$D_i = \frac{\partial \mu_i}{\partial \beta} = \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta} = \frac{\partial \mu_i}{\partial \eta_i} X_i;$$

- For the canonical link,

$$\frac{\partial \mu_i}{\partial \eta_i} = A_i;$$
• Generalized estimating equations become
\[ \sum_{i=1}^{N} X'_i A_i V_i^{-1} (y_i - \mu_i) = 0; \]

• Under independence, \( V_i = A_i \), so equations simplify:
\[ \sum_{i=1}^{N} X'_i (y_i - \mu_i) = 0; \]

• With dependence,
\[ A_i V_i^{-1} = \left( A_i^{\frac{1}{2}} \text{Corr}(Y_i) A_i^{-\frac{1}{2}} \right)^{-1}; \]
no corresponding simplification.
Paraphrasing McCullagh and Nelder (writing about independence case):

- Canonical link leads to desirable statistical properties, particularly in small samples.

- In general, no *a priori* reason why systematic effects should be additive on the scale of the canonical link.

- Canonical link is convenient, but convenience alone does not preempt quality of fit.

- Canonical link is often sensible on scientific grounds.