GEE for Mean and Association Parameters (GEE2)

GEE2

Previously: Under GEE1, we treat the correlation as (lower than) a nuisance. **But what if we are interested in the correlation?**

Idea: Jointly model marginal mean and association (correlation) using two estimating equations.

Note: We here concentrate on modeling multivariate binary data.

- One estimating equation for the mean.
- One estimating equation for the odds ratio (correlation).

Likelihood Parameterizations for Multivariate Binary Data

$$\mathbf{Y}_i = (Y_{i1}, \cdots, Y_{n_i}) : \text{Binary outcome vector for subject } i$$

Goals:

1. What is the likelihood and parameterization for $\mathbf{Y}_i$: $L(\mathbf{Y}_i, \boldsymbol{\theta}_i)$?
2. Next few slides: Four different parameterizations
3. Result: full likelihood inference for multivariate binary data is hard!

Note: for now, we suppress subscript i.

Bishop's Parameterization

$$
\begin{aligned}
\Pr(\mathbf{Y} = \mathbf{y}) &= c(\boldsymbol{\theta}) exp\{\sum_{j=1}^{n} \theta_j y_j + \sum_{k_1<k_2} \theta_{k_1,k_2} y_{k_1} y_{k_2} \\
&\quad + \cdots + \theta_{1,2,\cdots,n} y_1 \cdots y_n\}
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\theta} &= (\theta_1, \cdots, \theta_n, \theta_{1,2}, \cdots, \theta_{n-1,n}, \cdots \theta_{1,2\cdots n}) \\
&= (2^n - 1) \text{ vector of canonical parameters}
\end{aligned}
$$

Basically, we are modeling every possible combination.

Interpretation of $\boldsymbol{\theta}$

$$
\begin{aligned}
\theta_j &= \textit{logit}\,\Pr[Y_j = 1 | Y_k = 0, k \neq j] \\
\\
\theta_{jk} &= \log OR(Y_j, Y_k | Y_l = 0, l \neq j, k) \\
&= \textit{logit}\,\Pr[Y_j = 1 | Y_k = 1, Y_l = 0, l \neq j, k] \\
&\quad - \textit{logit}\,\Pr[Y_j = 1 | Y_k = 0, Y_l = 0, l \neq j, k] \\
\\
\theta_{123} &= \log OR[Y_1, Y_2 | Y_3 = 1, Y_l = 0, l > 3] \\
&\quad - \log OR[Y_1, Y_2 | Y_3 = 0, Y_l = 0, l > 3]
\end{aligned}
$$

Challenges of Modeling $\boldsymbol{\theta}$

Directly modeling $\boldsymbol{\theta}$ is not desirable

- Interpretations of θ_j and θ_{jk} are in terms of the conditional probability or the conditional odds ratio given particular values for all the other variables, e.g. $\theta_{jk} = \log$ OR describing the association between Y_j, Y_k given all the other responses are fixed.

- The interpretation of $\boldsymbol{\theta}$ depends on the number of observations n in a cluster.

Special Case: Quadratic exponential model:

$$\Pr(\mathbf{Y} = \mathbf{y}) = c(\boldsymbol{\theta}) exp\{\sum_{j=1}^{n} \theta_j y_j + \sum_{k_1 < k_2} \theta_{k_1 k_2} y_{k_1} y_{k_2}\}$$

Reference:
Zhao & Prentice (1990)
Prentice & Zhao (1991)

Question: How to parameterize $\Pr(\mathbf{Y} = \mathbf{y})$ using the marginal mean?

Fitzmaurice, Laird and Rotnitzky's Parameterization (1)

From earlier model:

$$
\begin{aligned}
\Pr(\mathbf{Y} = \mathbf{y}) &= c(\boldsymbol{\theta}) exp\{\mathbf{y}^T \boldsymbol{\theta_1} + \mathbf{w}^T \boldsymbol{\theta_2}\} \\
\boldsymbol{\theta_1} &= (\theta_1, \cdots, \theta_n)^T \\
\boldsymbol{\theta_2} &= (\theta_{1,2}, \cdots, \theta_{n-1,n}, \cdots \theta_{1,2\cdots n})^T \\
\mathbf{y} &= (y_1, \cdots, y_n)^T \\
\mathbf{w} &= (y_1 y_2, \cdots, y_{n-1} y_n, \cdots, y_1 y_2 \cdots y_n)^T
\end{aligned}
$$

As $\boldsymbol{\theta}_1$ does not have an attractive interpretation, we reparameterize $\boldsymbol{\theta_1}$ using the marginal mean of $\boldsymbol{Y}$:

$$
\begin{aligned}
\mu_j &= \Pr[Y_j = 1] \\
\boldsymbol{\mu} &= (\mu_1, \cdots \mu_n)^T = \boldsymbol{\mu}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})
\end{aligned}
$$

Fitzmaurice, Laird and Rotnitzky's Parameterization (2)

Then

$$Pr[\mathbf{Y} = \mathbf{y}; \boldsymbol{\theta_1}, \boldsymbol{\theta_2}] = Pr[\mathbf{Y} = \mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\theta_2}]$$

Model:

$$logit(\mu_{ij}) = \mathbf{X}_{ij}^T \beta$$

GEE for β (under this parameterization):

$$\sum \mathbf{D_i^T V_i^{-1}(Y_i} - \boldsymbol{\mu_i}) = \mathbf{0}$$

is the score equation for β given $\mathbf{V}_i$=cov($\mathbf{Y}_i$) is correctly specified.

$\Longrightarrow$

$\hat{\beta}$ is the MLE when $\mathbf{V}_i$ is correct,e.g., when $\boldsymbol{\theta_2}$ is known.

Drawbacks

The prior model and estimating equations addresses limitations with $\boldsymbol{\theta}_1$: forming model based on $\boldsymbol{\mu}$.

$\boldsymbol{\theta}_2$ is still a problem!

- $\boldsymbol{\theta}_2 =$ conditional log OR [conditioning on other $\mathbf{Y}$] and depends on the cluster size n.
- Again, difficult to interpret under this parameterization.

Question: How to parameterize $\Pr[\mathbf{Y} = \mathbf{y}]$ in terms of the marginal means and the marginal odds ratios?

Bahadur (1961) Parameterization

Let

$$
\begin{aligned}
R_j &= \frac{Y_j - \mu_j}{\sqrt{\mu_j(1-\mu_j)}} \\
\rho_{jk} &= corr(Y_j, Y_k) = E(R_j R_k) \\
\rho_{jkl} &= E(R_j R_k R_l)
\end{aligned}
$$

Then

$$
P[\mathbf{Y} = \mathbf{y}] = \prod_{j=1}^{n} \mu_j^{y_j} (1-\mu_j)^{(1-y_j)} (1 + \sum_{j<k} \rho_{jk} r_j r_k + \cdots + \rho_{1\ldots n} r_1 \cdots r_n)
$$

Advantage: $P[\mathbf{Y} = \mathbf{y}]$ is parameterized in terms of marginal means, correlations and higher order moments.

Problems with Bahadur Parameterization

Disadvantage:

The correlation ρ_{jk} is constrained in complicated ways by the μ_j.

Hence if $logit(\mu_j) = \mathbf{X}_j^T\beta$, then it is not correct to assume ρ_{jk} is independent of $\mathbf{X}_j$.

Liang, Zeger and Qaqish (1992)'s Parameterization

Idea: To parameterize $P[\mathbf{Y} = \mathbf{y}]$ in terms of marginal means, marginal odds ratios and contrasts of ORs.

Likelihood:

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\xi})$$

where

$$\begin{aligned} \mu_j &= P[Y_j = 1] \\ \nu_{jk} &= \frac{P[Y_j = 1, Y_k = 1]P[Y_j = 0, Y_k = 0]}{P[Y_j = 1, Y_k = 0]P[Y_j = 0, Y_k = 1]} = \ln OR[Y_j, Y_k] \end{aligned}$$

Note that μ_j is the marginal mean of Y_j and ν_{jk} is the marginal odds ratio between Y_j and Y_k.

Liang, Zeger and Qaqish (1992)'s Parameterization (2)

$$
\begin{aligned}
\nu_{jk} &= \frac{P[Y_j = 1, Y_k = 1]/P[Y_j = 0, Y_k = 1]}{P[Y_j = 1, Y_k = 0]/P[Y_j = 0, Y_k = 0]} \\
\xi_{jkl} &= \ln OR[Y_j, Y_k | Y_l = 1] - \ln OR[Y_j, Y_k | Y_l = 0] \\
\xi_{j_1 \ldots j_n} &= \sum_{Y_{j3}, \ldots Y_{jn} = 0,1} (-1)^{b(y)} \ln OR[Y_{j1}, Y_{j2} | Y_{j3}, \ldots Y_{jn}]
\end{aligned}
$$

Unfortunately, a full MLE for multivariate binary data is complicated.

Question: How can we jointly model mean and association (OR)?

GEE for Mean and Association (Finally!)

Model:

$$
\begin{aligned}
g(\mu_{ij}) &= \mathbf{X}_{ij}^T \boldsymbol{\beta} \\
ln(\nu_{ijk}) &= \mathbf{Z}_{ijk}^T \boldsymbol{\alpha}
\end{aligned}
$$

where

$$
\begin{aligned}
\mu_{ij} &= E(Y_{ij}) = Pr[Y_{ij} = 1] \\
\nu_{ijk} &= OR(Y_{ij}, Y_{ik})
\end{aligned}
$$

e.g. $ln(\nu_{ijk}) = \alpha_0 + \alpha_1 |t_{ij} - t_{ik}|^{-1}$, no constraint on α's.

GEE2

$$\begin{aligned}
\mathbf{U}_{\beta}(\beta, \boldsymbol{\alpha}) &= \sum_{i=1}^{m} \frac{\partial \boldsymbol{\mu}_i^T}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0 \\
\mathbf{U}_{\alpha}(\beta, \boldsymbol{\alpha}) &= \sum_{i=1}^{m} \frac{\partial \boldsymbol{\eta}_i^T}{\partial \boldsymbol{\alpha}} \mathbf{H}_i^{-1}(\mathbf{W}_i - \boldsymbol{\eta}_i) = 0
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{V_i} &= \text{working covariance of } \mathbf{Y_i} \\
\mathbf{W}_i &= (R_{i1}R_{i2}, R_{i2}R_{i3}, \cdots, R_{i,n_i-1}R_{i,n_i})^T \\
\boldsymbol{\eta}_i &= E(\mathbf{W}_i) \\
\mathbf{H}_i &= \text{working covariance of } \mathbf{W_i}
\end{aligned}$$