

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

$\mathbf{Y}_i = (Y_{i1}, \dots, Y_{n_i})$: 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

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

where

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

Basically, we are modeling every possible moment.

Interpretation of θ

$$\theta_j = \text{logit Pr}[Y_j = 1 | Y_k = 0, k \neq j]$$

$$\begin{aligned}\theta_{jk} &= \log OR(Y_j, Y_k | Y_l = 0, l \neq j, k) \\ &= \text{logit Pr}[Y_j = 1 | Y_k = 1, Y_l = 0, l \neq j, k] \\ &\quad - \text{logit Pr}[Y_j = 1 | Y_k = 0, Y_l = 0, l \neq j, k]\end{aligned}$$

$$\begin{aligned}\theta_{123} &= \log OR[Y_1, Y_2 | Y_3 = 1, Y_l = 0, l > 3] \\ &\quad - \log OR[Y_1, Y_2 | Y_3 = 0, Y_l = 0, l > 3]\end{aligned}$$

Challenges of Modeling θ

Directly modeling θ 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. $\theta_{jk} = \log \text{OR}$ describing the association between Y_j , Y_k given all the other responses are fixed.
- The interpretation of θ depends on the number of observations n in a cluster.

Special Case: Quadratic exponential model:

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

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:

$$\begin{aligned}\Pr(\mathbf{Y} = \mathbf{y}) &= c(\boldsymbol{\theta}) \exp\{\mathbf{y}^T \boldsymbol{\theta}_1 + \mathbf{w}^T \boldsymbol{\theta}_2\} \\ \boldsymbol{\theta}_1 &= (\theta_1, \dots, \theta_n)^T \\ \boldsymbol{\theta}_2 &= (\theta_{1,2}, \dots, \theta_{n-1,n}, \dots, \theta_{1,2 \dots n})^T \\ \mathbf{y} &= (y_1, \dots, y_n)^T \\ \mathbf{w} &= (y_1 y_2, \dots, y_{n-1} y_n, \dots, y_1 y_2 \dots y_n)^T\end{aligned}$$

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

$$\begin{aligned}\mu_j &= \Pr[Y_j = 1] \\ \boldsymbol{\mu} &= (\mu_1, \dots, \mu_n)^T = \boldsymbol{\mu}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\end{aligned}$$

Fitzmaurice, Laird and Rotnitzky's Parameterization (2)

Then

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

Model:

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

GEE for β (under this parameterization):

$$\sum \mathbf{D}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mu_i) = \mathbf{0}$$

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



$\hat{\beta}$ is the MLE when \mathbf{V}_i is correct, e.g., when θ_2 is known.

Drawbacks

The prior model and estimating equations addresses limitations with θ_1 : forming model based on μ .

θ_2 is still a problem!

- θ_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

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

Then

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

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 $\text{logit}(\mu_j) = \mathbf{X}_j^T \boldsymbol{\beta}$, 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:

$$L(\theta) = L(\mu, \nu, \xi)$$

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)

$$\begin{aligned}\nu_{jk} &= \frac{P[Y_j = 1, Y_k = 1]/P[Y_j = 0, Y_k = 1]}{P[Y_j = 1, Y_k = 0]/P[Y_j = 0, Y_k = 0]} \\ \xi_{jkl} &= \ln OR[Y_j, Y_k | Y_l = 1] - \ln OR[Y_j, Y_k | Y_l = 0] \\ \xi_{j_1 \dots j_n} &= \sum_{Y_{j_3}, \dots, Y_{j_n} = 0, 1} (-1)^{b(y)} \ln OR[Y_{j_1}, Y_{j_2} | Y_{j_3}, \dots, Y_{j_n}]\end{aligned}$$

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

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

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

where

\mathbf{V}_i = working covariance of \mathbf{Y}_i

$\mathbf{W}_i = (R_{i1}R_{i2}, R_{i2}R_{i3}, \dots, R_{i,n_i-1}R_{i,n_i})^T$

$\boldsymbol{\eta}_i = E(\mathbf{W}_i)$

\mathbf{H}_i = working covariance of \mathbf{W}_i