Marginal Model for Categorical Data

Outline

- Marginal Model
- GEE1
- Augmented GEE1 (GEE1.5)
- GEE2

Marginal Model

- For longitudinal data, the marginal model separates the modeling of the response on explanatory variables and the modeling of within-subject correlation. The former is modeled through the marginal mean $\mathrm{E}(Y_{ij})$ while the latter is modeled through the covariance structure $\mathrm{Cov}(Y_{ij}, Y_{ik})$.
- By "marginal mean", we mean the average response over the population. It is also called a **population average model** (as opposed to **subject specific model**).
- A marginal model has the following components:
  1. *Mean model*: the marginal mean depends on covariates via a link function

$$\begin{aligned} \mathrm{E}(Y_{ij} \mid \boldsymbol{X}_{ij}) &= \mu_{ij} \\ g(\mu_{ij}) &= \boldsymbol{X}_{ij}^T \boldsymbol{\beta} \end{aligned}$$

  2. *Correlation model* (nuisance):

$$\begin{aligned} \mathrm{Var}(Y_{ij} \mid X_i) &= v_{ij} = \phi v(\mu_{ij}) \\ \mathrm{Cor}(Y_{ij}, Y_{jk} \mid X_i) &= \rho_{ijk} \\ \mathrm{Cov}(\mathbf{Y}_i \mid X_i) &= V_i(\phi, \boldsymbol{\alpha}) = C_i^{1/2} R_i C_i^{1/2} \end{aligned}$$

  where R_i is the correlation matrix and $C_i = \mathrm{diag}(v_{ij})$ is a diagonal matrix of variances. The parameter $\boldsymbol{\alpha}$ characterizes the correlation and ϕ is a scale parameter for variances.
- Further assumptions are needed to specify a complete probability model, which may be difficult for categorical data.
- Without a likelihood function, estimation and valid inference are achieved by constructing an unbiased estimating function.

Example: Indonesian Children's Health Study

- Consider the effect of vitamin A deficiency (Xerophthalmia, X) on respiratory infection (RI, Y). Let i indicate the child and j the visit. The marginal mean model is:

$$\text{logit}(\mu_{ij}) = \log \frac{\Pr(Y_{ij} = 1)}{\Pr(Y_{ij} = 0)} = \beta_0 + \beta_1 \mathrm{I}_{X_{ij}=1}.$$

  The variance model is:

$$\begin{aligned} \text{Var}(Y_{ij}) &= \mu_{ij}(1 - \mu_{ij}), \\ \text{Cor}(Y_{ij}, Y_{ik}) &= \alpha. \end{aligned}$$

- The parameter of interest is β_1,

$$\exp(\beta_1) = \frac{\text{Odds of RI among vitamin A deficient children}}{\text{Odds of RI among non-deficient children}}.$$

  – When the prevalence of RI is low, the odds ratio (OR) is approximately the same as relative risk (RR).
  – The risk may be different for different children with the same covariates, so the parameter is a population average (assuming random sample).

- The correlation between two binary variables Y_1 and Y_2 has a constrained range that depends on μ_1 and μ_2. So it might be desirable to model the correlation differently, for example, using the odds ratio.

GEE1

Estimating Function

- When $(\phi, \boldsymbol{\alpha})$ are known, then the estimator $\hat{\boldsymbol{\beta}}$ is defined by the estimating equation:

$$\mathbf{0} = \sum_{i=1}^{m} \boldsymbol{U}_i(\boldsymbol{\beta}) = \sum_{i=1}^{m} \mathsf{D}_i^T \mathsf{V}_i^{-1} \left\{ \boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}) \right\},$$

where

$$\mathsf{D}_i(\boldsymbol{\beta}) = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}, D_i(j,k) = \frac{\partial \mu_{ij}}{\partial \beta_k},$$
$$\mathsf{V}_i(\boldsymbol{\beta}, \phi, \boldsymbol{\alpha}) = \mathsf{C}_i^{1/2} \mathsf{R}_i(\boldsymbol{\alpha}) \mathsf{C}_i^{1/2}.$$

- For example, for logistic model with one covariate:

$$\mu_{ij} = \frac{\exp(\beta_0 + \beta_1 x_{ij})}{1 + \exp(\beta_0 + \beta_1 x_{ij})}$$

$$D_i(j) = \left(\frac{\partial \mu_{ij}}{\partial \beta_0}, \frac{\partial \mu_{ij}}{\partial \beta_1} \right)$$

$$\mathsf{C}_i = \begin{pmatrix} \mu_{i0}(1-\mu_{i0}) & 0 & \cdots & 0 \\ 0 & \mu_{i1}(1-\mu_{i1}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_{in}(1-\mu_{in}) \end{pmatrix}$$

$$\mathsf{R}_i = \begin{pmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \cdots & \alpha & 1 \end{pmatrix}$$

GEE1 — Variance of $\hat{\boldsymbol{\beta}}$

- The solution $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically normal.
- If the correlation model is correct, then, the *model-based* estimate for the variance of $\hat{\boldsymbol{\beta}}$ is $\hat{V}(\hat{\boldsymbol{\beta}}) = A^{-1}$, where

$$A = \sum_{i=1}^{m} D_i^T(\hat{\boldsymbol{\beta}}) V_i^{-1}(\hat{\boldsymbol{\beta}}, \phi, \boldsymbol{\alpha}) D_i(\hat{\boldsymbol{\beta}}).$$

- If the correlation model is *not* correct, then we can use the *empirical* variance estimate:

$$\tilde{V}(\hat{\boldsymbol{\beta}}) = A^{-1} B A^{-1},$$

$$B = \sum_{i=1}^{m} \boldsymbol{U}_i \boldsymbol{U}_i^T$$

$$= \sum_{i=1}^{m} D_i^T(\hat{\boldsymbol{\beta}}) V_i^{-1}(\hat{\boldsymbol{\beta}}, \phi, \boldsymbol{\alpha}) \hat{\text{Cov}}(\boldsymbol{Y}_i) V_i^{-1}(\hat{\boldsymbol{\beta}}, \phi, \boldsymbol{\alpha}) D_i(\hat{\boldsymbol{\beta}})$$

where

$$\hat{\text{Cov}}(\boldsymbol{Y}_i) = (\boldsymbol{Y}_i - \boldsymbol{\mu}_i)(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)^T.$$

- $\hat{\text{Cov}}(\boldsymbol{Y}_i)$ is a poor estimator for $\text{Cov}(\boldsymbol{Y}_i)$. However, we do not need a good estimator for each $\text{Cov}(\boldsymbol{Y}_i)$. With sufficient independent replication (m large), the average covariance can be well estimated (consistency).
- What if $(\phi, \boldsymbol{\alpha})$ are unknown? How can we estimate them and what is the impact on the estimation of $\boldsymbol{\beta}$?
- Liang and Zeger (1986) proposed to use moment estimators for the unknown parameters (GEE1).

GEE1 — Estimating $\boldsymbol{\alpha}$

- Let $N = \sum_{i=1}^{m} n_i$.
- Recall that $\mathrm{Var}(Y_{ij} \mid \boldsymbol{X}_i) = \phi\nu(\mu_{ij})$ where ν is a known function.
- The scale parameter ϕ, if exists, can be estimated by

$$\hat{\phi} = \frac{1}{N-p} \sum_{i=1}^{m} \sum_{j=1}^{n_j} \frac{(Y_{ij} - \hat{\mu}_{ij})^2}{\hat{\nu}_{ij}(\hat{\mu}_{ij})},$$

  where p is the dimension of $\boldsymbol{\beta}$.

  – Binomial: $\hat{\nu}_{ij} = \hat{\mu}_{ij}(1 - \hat{\mu}_{ij})$.

  – Poisson: $\hat{\nu}_{ij} = \hat{\mu}_{ij}$.

- Define the residuals:

$$r_{ij} = \frac{Y_{ij} - \hat{\mu}_{ij}(\hat{\boldsymbol{\beta}})}{\hat{V}_{ij}^{1/2}},$$

  where

$$\hat{V}_{ij} = \hat{\mathrm{Var}}(Y_{ij}) = \hat{\phi}\hat{\nu}_{ij}.$$

- The correlation parameter $\boldsymbol{\alpha}$ can be estimated as simple functions of r_{ij}.

  1. Unstructured correlation:

$$\hat{R}(j,k) = \frac{1}{m-p} \sum_{i=1}^{m} r_{ij} r_{ik}.$$

  2. Exchangeable correlation:

$$\hat{\alpha} = \frac{1}{\sum_i n_i(n_i - 1) - p} \sum_{i=1}^{m} \sum_{j \neq k} r_{ij} r_{ik}.$$

GEE1 — Estimation

An iterative algorithm is used to find $(\hat{\boldsymbol{\beta}}, \hat{\phi}, \hat{\boldsymbol{\alpha}})$:

1. Starting with an estimate of $\boldsymbol{\beta}$, i.e., assuming independence.
2. Given $\hat{\boldsymbol{\beta}}^{(j)}$, calculate method-of-moments estimates for ϕ and $\boldsymbol{\alpha}$.
3. Given estimates for ϕ and $\boldsymbol{\alpha}$, solve the estimating equation using Fisher's scoring algorithm:

$$\hat{\boldsymbol{\beta}}^{(j+1)} = \hat{\boldsymbol{\beta}}^{(j)} + \left(\sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right)^{-1} \sum_{i=1}^{m} D_i^T V_i^{-1} \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i\}.$$

4. Iterate the above two steps until convergence is achieved.

Working Correlation

- The model chosen for $R_i(\boldsymbol{\alpha})$ is called the "working correlation" since it needs not be the true correlation to obtain a valid point estimate $\hat{\boldsymbol{\beta}}$ (consistent and asymptotically normal).
- If $R_i(\boldsymbol{\alpha})$ is the correct correlation, then the model-based estimates of the standard errors for $\hat{\boldsymbol{\beta}}$ can used. Otherwise we use the empirical estimates of standard errors.
- Replace $\boldsymbol{\alpha}$ (ϕ) with a (any) consistent estimator does not affect the large sample properties of $\hat{\boldsymbol{\beta}}$. The asymptotic variance of $\hat{\boldsymbol{\beta}}$ would be the same as if $\boldsymbol{\alpha}$ is known. (Liang and Zeger, 1986).
- Is it worthwhile to model the correlation at all? Why not simply use the "working independence" model?
  - A good model that closely approximate $\text{Cov}(\boldsymbol{Y}_i)$ can improve efficiency, sometimes greatly, over the working independence model.
  - For between-subject covariates (e.g. gender) and moderate correlation, the loss of efficiency is not very large.

Hypothesis Testing

- **Wald Test**:
  - $H_0 : \beta_j = 0$

$$\frac{\hat{\beta}_j}{\hat{\text{s.e.}}} \sim \mathcal{N}(0, 1).$$

  - Write $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$

$$\hat{\boldsymbol{\beta}}_1^T \mathsf{V}_1^{-1} \hat{\boldsymbol{\beta}}_1 \sim \chi_r^2,$$

    where r is the dimension of $\boldsymbol{\beta}_1$ and V_1 is the estimated variance matrix corresponding to $\hat{\boldsymbol{\beta}}_1$.

- **Score Test**: $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$.

$$T_s = \frac{1}{m} \boldsymbol{U}_1(\mathbf{0}, \hat{\boldsymbol{\beta}}_2)^T \Sigma_1^{-1} \boldsymbol{U}_1(\mathbf{0}, \hat{\boldsymbol{\beta}}_2) \sim \chi_r^2.$$

GEE1 — What about $\boldsymbol{\alpha}$?

- Recall that we use simple MoM estimators to estimate $(\phi, \boldsymbol{\alpha})$ in GEE1. The scale parameter ϕ is often considered to a nuisance and it does not affect the estimates of $\boldsymbol{\beta}$. But what about $\boldsymbol{\alpha}$?
  1. Shouldn't we consider the parameter as $(\boldsymbol{\beta}, \boldsymbol{\alpha})$?
  2. Can't we improve upon the estimation of $\boldsymbol{\alpha}$?
  3. Would "better" estimation of $\boldsymbol{\alpha}$ help us to "better" estimate $\boldsymbol{\beta}$?
- Shouldn't we consider the parameter as $(\boldsymbol{\beta}, \boldsymbol{\alpha})$?

  Answer: It depends.
  - Is $\boldsymbol{\alpha}$ a nuisance? If the covariance structure is of secondary interest (often the case) then GEE1 is usually fine. However, if the covariance matrix *is* of primary interest then GEE1 is not ideal.
  - Are you willing to sacrifice some model robustness in order to let $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ be the target parameter? Note that in GEE1, the estimate $\hat{\boldsymbol{\beta}}$ is consistent even if the model for $\boldsymbol{\alpha}$ is wrong. Other approaches that treat $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ on equal ground may not have this property.
- Can't we improve upon the estimation of $\boldsymbol{\alpha}$?

  Answer: Yes!
  - *Model*: We can adopt a more flexible class of covariance models.
  - *Model*: We can adopt alternative association (dependence) models that are more suitable for categorical data.
  - *Estimator*: We can use estimators that are more efficient in estimating $\boldsymbol{\alpha}$ but do not sacrifice the robustness of $\hat{\boldsymbol{\beta}}$ (GEE1.5, ALR).

  - *Estimator*: We can create estimators that are targeted at $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ jointly and are efficient for both (GEE2, likelihood methods).

- Would "better" estimation of $\boldsymbol{\alpha}$ help us to "better" estimate $\boldsymbol{\beta}$?

  Answer: It depends.

  - *Model* for $\boldsymbol{\alpha}$ is important for the efficiency of $\hat{\boldsymbol{\beta}}$.
  - *Estimator* choice may not be important (given a decent model).

GEE1.5

Prentice (1988)'s Augmented GEE1 for Binary Responses

- **Idea**: Joint estimation of parameters in both the marginal response probabilities and the pairwise correlations.
- GEE1 uses an estimating function U_1 based on the centered first moments $(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)$ for the estimation of $\boldsymbol{\beta}$. We can add a second estimation function based on the centered *second* moments:
$$\frac{(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})}{[\mu_{ij}(1 - \mu_{ij})\mu_{ik}(1 - \mu_{ik})]^{1/2}} - \sigma_{ijk}$$
to estimate $\boldsymbol{\alpha}$.
- Now, the joint estimating functions are:
$$\begin{aligned}
\boldsymbol{U}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_{i=1}^{m} D_i^T(\boldsymbol{\beta}) V_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\} \\
\boldsymbol{U}_2(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_{i=1}^{m} E_i^T(\boldsymbol{\beta}, \boldsymbol{\alpha}) W_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \{\boldsymbol{S}_i - \boldsymbol{\sigma}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})\} \\
\boldsymbol{S}_i &= (R_{i1}R_{i2}, R_{i1}R_{i3}, R_{i1}R_{in_i}, \cdots, R_{in_i-1}R_{in_i})^T \\
R_{ij} &= (Y_{ij} - \mu_{ij})/[\mu_{ij}(1 - \mu_{ij})]^{1/2} \\
\boldsymbol{\sigma}_i &= \mathrm{E}(\boldsymbol{S}_i) \\
E_i &= \frac{\partial \boldsymbol{\sigma}_i}{\partial \boldsymbol{\alpha}} \\
W_i &= \mathrm{diag}(\mathrm{Var}(R_{i1}, R_{i2}), \cdots, \mathrm{Var}(R_{in_i-1}R_{in_i}))
\end{aligned}$$
- Note that the above W_i is an $n_i(n_i - 1)/2$-dimensional working independent variance matrix for $\boldsymbol{S}_i$. Other working matrices than the independent one could readily be substituted. To model $\mathrm{Cov}(\boldsymbol{S}_i)$ properly, we need specify models for higher moments (with more parameters), which is typically difficult.

- Iterative method can be used for solving the joint estimating equations

$$\mathbf{0} = \boldsymbol{U}_1(\boldsymbol{\beta}, \boldsymbol{\alpha})$$
$$\mathbf{0} = \boldsymbol{U}_2(\boldsymbol{\beta}, \boldsymbol{\alpha})$$

  Given $(\hat{\boldsymbol{\beta}}^{(r)}, \hat{\boldsymbol{\alpha}}^{(r)})$:

  1. Fixed $\hat{\boldsymbol{\alpha}}^{(r)}$, solve $\mathbf{0} = \boldsymbol{U}_1(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}^{(r)})$ to get $\hat{\boldsymbol{\beta}}^{(r+1)}$.
  2. Fixed $\hat{\boldsymbol{\beta}}^{(r+1)}$, solve $\mathbf{0} = \boldsymbol{U}_2(\hat{\boldsymbol{\beta}}^{(r+1)}, \boldsymbol{\alpha})$ to get $\hat{\boldsymbol{\alpha}}^{(r+1)}$.

- $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$ is consistent and asymptotically normal under correct model specification. Similar to GEE1, $\hat{\boldsymbol{\beta}}$ is consistent even if the model for $\boldsymbol{\alpha}$ is misspecified.

A Note on Modeling Correlation of Binary Responses

- Correlation for binary data are constrained by their means.

  Let $\mathrm{E}(Y_1) = \mu_1$, $\mathrm{E}(Y_2) = \mu_2$, $\rho_{12} = \mathrm{Cor}(Y_1, Y_2)$, and $\pi_{12} = \mathrm{E}(Y_1 Y_2)$, then $\pi_{12} < \min(\mu_1, \mu_2)$, and hence,

$$\rho_{12}^2 \leq \min\left\{\frac{\mu_1(1-\mu_2)}{\mu_2(1-\mu_1)}, \frac{\mu_2(1-\mu_1)}{\mu_1(1-\mu_2)}\right\},$$

  because the pairwise correlation

$$\rho_{12} = \frac{\pi_{12} - \mu_1\mu_2}{[\mu_1(1-\mu_1)\mu_2(1-\mu_2)]^{1/2}},$$

  and π_{12} is constrained to satisfy

$$\max(0, \mu_1 + \mu_2 - 1) < \pi_{12} < \min(\mu_1, \mu_2).$$

- Modeling odds ratios:

$$\begin{aligned}\Psi_{ijk} &= \frac{\Pr(Y_{ij}=1, Y_{ik}=1)\Pr(Y_{ij}=0, Y_{ik}=0)}{\Pr(Y_{ij}=1, Y_{ik}=0)\Pr(Y_{ij}=0, Y_{ik}=1)}, \\ &= \frac{\Pr(Y_{ij}=1 \mid Y_{ik}=1)/\Pr(Y_{ij}=0 \mid Y_{ik}=1)}{\Pr(Y_{ij}=1 \mid Y_{ik}=0)/\Pr(Y_{ij}=0 \mid Y_{ik}=0)}.\end{aligned}$$

  - Invariant to marginal specification of μ_1 and μ_2.
  - The log odds ratios $\log\Psi \in (-\infty, \infty)$, are symmetric about 0 and not constrained by the marginal means.
  - Interpretation: $\Psi_{ijk} = 1$ or $\log\Psi_{ijk} = 0$ implies (Y_{ij}, Y_{ik}) are uncorrelated.

- The odds ratio, Ψ_{ijk} and the marginal means μ_{ij}, μ_{ik} determine the $\pi_{ijk} = \mathrm{E}(Y_{ij}Y_{ik})$, the correlation ρ_{ijk}, and variance $\boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$.

$$\begin{aligned}\Psi_{ijk} &= \frac{\pi_{ijk}(1-\mu_{ij}-\mu_{ik}+\pi_{ijk})}{(\mu_{ij}-\pi_{ijk})(\mu_{ik}-\pi_{ijk})}, \text{hence,} \\ \pi_{ijk} &= \frac{A - [A^2 - 4(\Psi_{ijk}-1)\Psi_{ijk}\mu_{ij}\mu_{ik}]^{1/2}}{2(\Psi_{ijk}-1)} \\ A &= 1 - (\mu_{ij}+\mu_{ik})(1-\Psi_{ijk})\end{aligned}$$

Using Marginal Odds Ratios to Model Association

- The joint models (Lipsitz et al, 1991) are:

$$\textbf{Mean model} \qquad \text{logit}(\mu_{ij}) = \boldsymbol{X}_{ij}^T\boldsymbol{\beta}$$
$$\textbf{Correlation model} \qquad \log(\Psi_{ijk}) = \boldsymbol{Z}_{ijk}^T\boldsymbol{\alpha}$$

- **Alternating logistic regression (ALR)** (Carey et al, 1993) Let $\gamma_{ijk} = \log\Psi_{ijk} = \boldsymbol{Z}_{ijk}^T\boldsymbol{\alpha}$, note that the pairwise conditional expectations:

$$\text{logit}\,\text{E}(Y_{ij} \mid Y_{ik}, \boldsymbol{X}_i) = \gamma_{ijk}Y_{ik} + \Delta_{ijk}$$
$$\Delta_{ijk} = \log\left(\frac{\mu_{ij} - \pi_{ijk}}{1 - \mu_{ij} - \mu_{ik} + \pi_{ijk}}\right).$$

- Suppose that all the odds ratios are the same, $\gamma_{ijk} = \gamma$, then an estimate for γ could be obtained by a logistic regression of y_{ij} on y_{ik}, $1 \le j < k \le n_i, i = 1, \cdots, m$, using Δ_{ijk} as an offset.
- More generally, $\log(\Psi_{ijk}) = \boldsymbol{Z}_{ijk}^T\boldsymbol{\alpha}$, where $\boldsymbol{Z}_{ijk}$ is a set of covariates characterizing the log-odds ratio between observations j and k. For example, in family data, Z_{ijk} would encode the type of family relationship for Y_{ij} and Y_{ik}: husband-wife, parent-sib, or sib-sib. We then estimate the vector $\boldsymbol{\alpha}$ by a logistic regression of y_{ij} on the product $\boldsymbol{Z}_{ijk}^T\boldsymbol{\alpha}$ with the same offset Δ_{ijk}.
- Note that the offset depends on both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ so that iteration is required. We alternate the following two steps until convergence.
  1. Given the current values of $(\boldsymbol{\beta}^{(r)}, \boldsymbol{\alpha}^{(r)})$, calculate $\hat{V}^{(r)}$ and solve the $\boldsymbol{U}_1(\boldsymbol{\beta}, \boldsymbol{\alpha})$ for an updated $\hat{\boldsymbol{\beta}}^{(r+1)}$.
  2. Given $\hat{\boldsymbol{\beta}}^{(r+1)}$ and $\hat{\boldsymbol{\alpha}}^{(r)}$, evaluate the offset and perform the offset logistic regression of Y_{ij} on $Z_{ijk}Y_{ik}$ with a total of $\sum_{i=1}^m n_i(n_i - 1)/2$ observations to obtain $\hat{\boldsymbol{\alpha}}^{(r+1)}$.

- **ALR estimating equations**. Formally the ALR uses the same model as in Lipsitz et al. (1991) but uses this estimating equation for $\boldsymbol{\alpha}$:

$$\boldsymbol{U}_{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{m} \mathsf{F}_i^T(\boldsymbol{\beta}, \boldsymbol{\alpha}) \tilde{\mathsf{W}}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \boldsymbol{T}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}),$$

where

$$\begin{aligned} \boldsymbol{T}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \text{vec}(Y_{ij} - \zeta_{ijk}) \\ \zeta_{ijk} &= \text{E}(Y_{ij} \mid Y_{ik}) \\ \tilde{\mathsf{W}}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \text{diag}(\text{Var}(Y_{ij} \mid Y_{ik})) \\ &= \text{diag}(\zeta_{ijk}(1 - \zeta_{ijk})) \\ \mathsf{F}_i &= \frac{\partial \boldsymbol{\zeta}_i}{\partial \boldsymbol{\alpha}} \end{aligned}$$

  - The ALR $\boldsymbol{\alpha}$ is more efficient that the model of Lipsitz et al. (1991).
  - The efficiency is comparable to GEE2 but more computationally efficient for large clusters (does not require the inverse of large matrices).
  - The ALR $(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha})$ are consistent and asymptotically normal. Sandwich variance estimates.

- **When the scale parameter ϕ is important** (over-dispersion, heteroscedasticity), Yan and Fine (2004) proposed to use a third model for the scale parameter:

$$g_3(\phi_{ij}) = \boldsymbol{T}_{3i}^T \boldsymbol{\gamma},$$

where g_3 is the link function, e.g. log. The estimating equation is

$$\boldsymbol{U}_\phi = \sum_{i=1}^{m} \mathsf{D}_{3i}^T \mathsf{V}_{3i}(\boldsymbol{s}_i - \boldsymbol{\phi}_i) = \mathbf{0},$$

where

$$s_{ij} = \frac{(Y_{ij} - \mu_{ij})^2}{v_{ij}}, D_{3i} = \frac{\partial \phi_i}{\partial \gamma}.$$

The method is implemented in R package `geepack`.

GEE2

- Prentice and Zhao (1991) considered $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$ as the parameter and the optimal estimating function for $\boldsymbol{\delta}$.
- Paired models:

$$g_1(\mu_{ij}) = \boldsymbol{X}_{ij}^T \boldsymbol{\beta}$$
$$g_2(\sigma_{ijk}) = \boldsymbol{Z}_{ijk}^T \boldsymbol{\alpha}$$

  where $\sigma_{ijk} = \text{Cov}(Y_{ij}, Y_{ik})$.
- Optimal estimating equations for $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$:

$$\begin{aligned}
\boldsymbol{U}(\boldsymbol{\delta}) &= \sum_{i=1}^{m} \mathsf{D}_i^T(\boldsymbol{\delta}) \mathsf{V}_i^T(\boldsymbol{\delta}) \boldsymbol{T}_i(\boldsymbol{\delta}) \\
&= \begin{pmatrix} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} & \frac{\partial \boldsymbol{\sigma}_i}{\partial \boldsymbol{\beta}} \\ 0 & \frac{\partial \boldsymbol{\sigma}_i}{\partial \boldsymbol{\alpha}} \end{pmatrix} \begin{pmatrix} \mathsf{V}_i(1,1) & \mathsf{V}_i(1,2) \\ \mathsf{V}_i(1,2)^T & \mathsf{V}_i(2,2) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{Y}_i - \boldsymbol{\mu}_i \\ \boldsymbol{S}_i - \boldsymbol{\sigma}_i \end{pmatrix} \\
\mathsf{V}_i(1,1) &= \text{Cov}(\boldsymbol{Y}_i) \\
\mathsf{V}_i(1,2) &= \text{Cov}(\boldsymbol{Y}_i, \boldsymbol{S}_i) \\
\mathsf{V}_i(2,2) &= \text{Cov}(\boldsymbol{S}_i) \\
S_{ijk} &= (Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})
\end{aligned}$$

- First and second moment models are not enough to obtain $\mathsf{V}_i(1,2)$ and $\mathsf{V}_i(2,2)$.
- In GEE2, a working 3rd/4th moment model is used.
- Note that maximum likelihood method can be used if we specify all moments (so that the full likelihood can be determined).
- GEE2 equations can be derived as the score equations for a Quadratic Exponential Family (QEF) model:

$$\ell_i = \boldsymbol{\theta}_{1i}^T \boldsymbol{Y}_i + \boldsymbol{\theta}_{2i}^T \boldsymbol{S}_i + \delta_i + c_i(\boldsymbol{Y}_i).$$

- Working (*ad hoc*) 3rd/4th moment models (Prentice and Zhao, 1991)
  - Independence working models

$$V_i(1,2) = 0$$
$$V_i(2,2) = \text{diagonal matrix}$$

  - Gaussian working models

$$V_i(1,2) = 0$$
$$V_i(2,2) : \mathrm{Cov}(S_{ijk}, S_{ilm}) = \sigma_{ijl}\sigma_{ikm} + \sigma_{ijm}\sigma_{ikl}$$

- Liang et al. (1992) considered GEE2 model for binary data using odds ratios.
- Estimation can be done via Fisher scoring and sandwich variance estimator is used to protect against the 3rd/4th moment model misspecification.
- Consistency of both $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$ depends on the **correct modeling of both mean and covariance**.
- The matrix V_i has dimension $M_i \times M_i$ where

$$M_i = n_i + n_i(n_i - 1)/2,$$

  and its inverse is required!
- In Liang et al (1992), solutions of higher order polynomial equations are needed!
- The efficiency gained of GEE2 comparing with GEE1.5 depends on the correct specification of the 3rd/4th moments.
- Conclusion: GEE2 may not be worthwhile after all. If we want to specify higher order moments, why not use the likelihood?

Further Reading

- Chapters 7.1, 7.5, and 8 of DHLZ.

References

- Cavey V, Zeger SL, and Diggle P (1993). Modelling multivariate binary data with alternating logistic regression. *Biometrika* **80**:517-26.
- Ling KY and Zeger SL (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**:13-22.
- Liang KY, Zeger SL, and Qaquish B (1992). Multivariate regression analysse for categorical data (with discussion). *JRSSB* **54**:3-40.
- Lipsitz SR, Laird NM, and Harrington DP (1991). Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika* **78**:153-60.
- Prentice RL (1988). Correlated binary regression with covariates specific o each binary observations. *Biometrics* **44**:1033-48.
- Prentice RL and Zhao LP (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* **47**:825-39.
- Yan J and Fine J (2004). Estimating equations for association structures. *Statistics in Medicine* **23**:859-74.