5. Marginal Models

- For longitudinal data the marginal model separates the modeling of the between-subject and within-subject covariate effects. The former is modeled through the marginal mean $E(Y_{ij})$ while the latter is modeled through the covariance structure $cov(Y_{ij}, Y_{ik})$.

- This marginal modeling approach is all what we can do if we only have cross-sectional data (one observation per subject).

- For a random sample, it is also called a <u>population average</u> model (as opposed to <u>subject specific</u> model).

- Specifically, a marginal model has the following components:

  1. <u>Mean model</u>: the marginal mean depends on

covariates via a link function

$$E(Y_{ij}|X_{ij}) = \mu_{ij},$$
$$g(\mu_{ij}) = X_{ij}^T \beta.$$

2. Correlation model (nuisance)

$$var(Y_{ij}|X_i) = v_{ij} = \phi v(\mu_{ij}),$$
$$cov(Y_{ij}, Y_{ik}|X_i) = \rho_{ijk},$$
$$cov(Y_i|X_i) = V_i(\phi, \alpha) = C_i^{1/2} R_i C_i^{1/2}$$

where R_i is the correlation matrix and $C_i = diag(v_{ij})$ is a diagonal matrix of variances. The parameter α characterizes the correlation and ϕ is a scale parameter for variances.

- Further assumptions are needed to specify a complete probability model which may be different 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

$$\text{logit}\mu_{it} = \log \frac{P(Y_{ij} = 1)}{P(Y_{ij} = 0)} = \beta_0 + \beta_1 I_{x_{ij}=1}.$$

- The variance model can be written as

$$var(Y_{ij}) = \mu_{ij}(1 - \mu_{ij}),$$
$$corr(Y_{ij}, Y_{ik}) = \alpha.$$

- 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 (more about this later).

GEE1-Estimating β

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

$$0 = \sum_{i=1}^{m} U_i(\beta) = \sum_{i=1}^{m} D_i^T V_i^{-1} \{Y_i - \mu_i(\beta)\}$$

where

$$D_i(\beta) = \frac{\partial \mu_i}{\partial \beta}, \quad D_i(j,k) = \frac{\partial \mu_{ij}}{\partial \beta_k},$$

$$V_i(\beta, \phi, \alpha) = C_i^{1/2} R_i(\alpha) C_i^{1/2}.$$

- For logistic model with one covariate:

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

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

$$C_i = \begin{pmatrix} \mu_{i1}(1-\mu_{i1}) & 0 & \cdots & 0 \\ 0 & \mu_{i2}(1-\mu_{i2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_{in}(1-\mu_{in}) \end{pmatrix},$$

$$R_i = \begin{pmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \cdots & 1 \end{pmatrix}.$$

GEE1-Variance

The solution $\hat{\beta}$ is consistent and asymptotically normal. If the correlation model is correct, then the model-based esimate for the variance of $\hat{\beta}$ is

$$\hat{V}(\hat{\beta}) = \hat{A}^{-1}.$$

where $\hat{A} = \sum_{i=1}^{m} D_i^T(\hat{\beta}) V_i^{-1}(\hat{\beta}, \phi, \alpha) D_i(\hat{\beta})$.
If the correlation model is not correct, then we can use

the <u>empirical variance estimate</u>:

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

where

$$B = \sum_{i=1}^{m} U_i U_i^T = \sum_{i=1}^{m} D_i^T(\hat{\beta}) V_i^{-1}(\hat{\beta}, \phi, \alpha) c\hat{o}v(Y_i) V_i^{-1}(\hat{\beta}, \phi, \alpha) D_i(\hat{\beta}),$$

$$c\hat{o}v(Y_i) = (Y_i - \mu_i)(Y_i - \mu_i)^T.$$

Note that $c\hat{o}v(Y_i)$ is a poor estimator for $cov(Y_i)$. However we do not need a good estimator for each $cov(Y_i)$. With sufficient independent replication, the average covariance can be well estimated (consistency).

What if (ϕ, α) are <u>unknown</u>? (How can we estimate them and what is the impact on the estimation of β?) Liang and Zeger (1986) proposed to use simple method-of-moment estimators based on the residuals (GEE1).

GEE1-Estimating α

Let $N = \sum_{i=1}^{m} n_i$. Recall that $var(Y_{ij}|X_i) = \phi v(\mu_{ij})$

where v is a known function. The scale parameter ϕ (if exists) can be estimated by

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

where p is the dimension of β.

- Binomial: $\hat{v}_{ij} = \hat{\mu}_{ij}(1 - \hat{\mu}_{ij})$,

- Poisson: $\hat{v}_{ij} = \hat{\mu}_{ij}$.

Define the residuals

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

where $\hat{V}_{ij}^{1/2} = v\hat{a}r(Y_{ij}) = \hat{\phi}\hat{v}_{ij}$.

The correlation parameter α can be estimated as simple functions of r_{ij}.

- Unstructured correlation

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

- Exchangeable correlation

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

GEE1-Estimation

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

1. Start with an estimate of β. i.e., assuming independence.

2. Given $\hat{\beta}^{(j)}$, calculate method-of-moments estimates for ϕ and α.

3. Given estimates for ϕ and α, solve the estimating

equation using Fisher scoring algorithm:

$$\hat{\beta}^{(j+1)} = \hat{\beta}^{(j)} + \left(\sum_{i=1}^{m} D_i^T V_i^{-1} D_i \right)^{-1} \sum_{i=1}^{m} D_i^T V_i^{-1} (Y_i - \mu_i).$$

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

- Working Correlation
  - The model chosen for $R_i(\alpha)$ is called the "working correlation" since it needs not be the true correlation to obtain a valid point estimate $\hat{\beta}$ (consistent and asymptotically normal).
  - If $R_i(\alpha)$ is the correct correlation, then the model-based estimates of the standard errors for $\hat{\beta}$ can be used from A^{-1}. Otherwise, we use the empirical estimates of standard errors ($A^{-1}BA^{-1}$).
  - Replacing $\alpha(\phi)$ with a (any) consistent estimator does not affect the large sample properties of $\hat{\beta}$. The asymptotic variance of $\hat{\beta}$ would be the same as if α is known (Liang and Zeger, 1986).

– Is it worthwhile to model the correlation at all? Why not simply use the "working independence" model?

1. A good model that closely approximate $cov(Y_i)$ can improve efficiency, sometimes greatly, over the working independence model.
2. For between-subject covariates and moderate correlation, the loss of efficiency is not very large.

- Hypothesis Testing

  – Wald Test

1. $H_0 : \beta_j = 0$, $\frac{\hat{\beta}_j}{se} \sim N(0, 1)$.
2. Write $\beta = (\beta_1, \beta_2)$, $H_0 : \beta_1 = 0$.

$$\hat{\beta}_1^T V_1^{-1} \hat{\beta}_1 \sim \chi_r^2$$

where r is the dimension of β_1 and V_1 is the estimated variance matrix corresponding to $\hat{\beta}_1$.

– Score Test

$$H_0 : \beta_1 = 0.$$
$$T_s = \frac{1}{m} U_1(0, \hat{\beta}_2)^T \Sigma_1^{-1} U_1(0, \hat{\beta}_2) \sim \chi_r^2.$$

- Caveat
  - MCAR assumed
  - Time-dependent covariates and modeling of covariance.

Examples

GEE1 - What about α?

Recall that we use simple method-of-moments to estimate (ϕ, α) in GEE1. The scale parameter ϕ is often considered to be a nuisance and it does not affect the estimates of β. But what about α?

1. Should not we consider the parameter (β, α)?

2. Cannot we improve upon the estimation of α?

3. Would "better" estimation of α help us to "better" estimate of β?

- Should not we consider the parameter as (β, α)? Answer: It depends.
  - Is α 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 (β, α) be the target parameter? Note that in GEE1, the estimate $\hat{\beta}$ is consistent

even if the model for α is wrong. Other approaches that treat β and α on equal ground may not have this property.

- Cannot we improve upon the estimation of α?
  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 α but do not sacrifice the robustness of $\hat{\beta}$ (GEE1.5, ALR).
  - Estimator: We can create estimators that are targeted at (β, α) jointly and are efficient for both (GEE2, likelihood methods).

- Would "better" estimation of α help us to "better" estimate of β?
  Answer: It depends.
  - Model for α is important for the efficiency of $\hat{\beta}$.

– Estimator choice may not be important (given a decent model).

Augmented GEE1 (GEE1.5)

- For binary responses (Prentice, 1988), GEE1 uses an estimating function U_1 based on the centered first moment $(Y_i - \mu_i)$ for the estimation of β. We can add a second estimating function based on the centered second moments:

$$(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik}) - \sigma_{ijk}$$

to estimate α.

$$U_1(\beta, \alpha) = \sum_{i=1}^{m} D_i^T(\beta) V_i^{-1}(\beta, \alpha) \left\{ Y_i - \mu_i(\beta) \right\},$$

$$U_2(\beta, \alpha) = \sum_{i=1}^{m} E_i^T(\beta, \alpha) W_i^{-1}(\beta, \alpha) \left\{ S_i - \sigma_i(\beta, \alpha) \right\},$$

where

$$
\begin{aligned}
S_i &= (Y_i - \mu_i) \otimes (Y_i - \mu_i) \\
&\quad = vec\left[(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})\right], \\
\sigma_i &= E(S_i) = vec(\sigma ijk), \\
E_i &= \frac{\partial \sigma_i}{\partial \alpha}, \\
W_i &\approx cov(S_i).
\end{aligned}
$$

To model $cov(S_i)$ properly, we need specify models for higher moments (with more parameters), which is typically difficult. In practice, we can use a simple working correlation matrix (e.g., working independence) and empirical variance for $\hat{\alpha}$.

- Paired models:

$$
\text{Mean model: logit}\mu_{ij} = x_{ij}^T \beta,
$$

$$
\text{Conditional model: } g_2(\rho_{ijk}) = z_{ijk}^T \alpha.
$$

– g_2 is a second link function which can be, for example, Fisher's Z transformation

$$g_2(\rho_{ijk}) = \log\left(\frac{1+\rho_{ijk}}{1-\rho_{ijk}}\right) \in (-\infty, \infty).$$

– Allows a flexible class of models for dependence.
– Note that for binary responses, no variance model is needed:

$$var(Y_{ij}) = \mu_{ij}(1-\mu_{ij}).$$

– The correlation "design" matrix Z_i is $n_i(n_i - 1)/2 \times q$ matrix (q is the length of α).

- Paired Estimating Equations:

$$0 = U_1(\beta, \alpha),$$
$$0 = U_2(\beta, \alpha).$$

The two equations are solved iteratively. Given $(\hat{\beta}^{(k)}, \hat{\alpha}^{(k)})$,

1. Fixed $\hat{\alpha}^{(k)}$, solve $0 = U_1(\beta, \hat{\alpha}^{(k)})$.

2. Fixed $\hat{\beta}^{(k+1)}$, solve $0 = U_2(\hat{\beta}^{(k+1)}, \alpha)$.

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

- Model for correlation for binary responses.

  – Correlation for binary data are constrained by their means. Suppose $E(Y_1) = \mu_1$ and $E(Y_2) = \mu_2$, and $\pi_{12} = E(Y_1 Y_2)$, then

$$
\begin{aligned}
&\pi_{12} \leq \min(\mu_1, \mu_2), \\
&\rho_{12}^2 \leq \min\left\{\frac{\mu_1(1-\mu_2)}{\mu_2(1-\mu_1)}, \frac{\mu_2(1-\mu_1)}{\mu_1(1-\mu_2)}\right\}.
\end{aligned}
$$

  For example, $E(Y_1) = 0.3$ and $E(Y_2) = 0.1$, then $\rho^2 \leq 0.260$.
  Therefore, when there is restriction in correlation, Fisher Z-transformation method dose not work.

– Modeling odds ratios

$$\Psi_{ijk} = \frac{P(Y_{ij}=1, Y_{ik}=1)P(Y_{ij}=0, Y_{ik}=0)}{P(Y_{ij}=1, Y_{ik}=0)P(Y_{ij}=0, Y_{ik}=1)}$$
$$= \frac{P(Y_{ij}=1|Y_{ik}=1)/P(Y_{ij}=0|Y_{ik}=1)}{P(Y_{ij}=1|Y_{ik}=0)/P(Y_{ij}=0|Y_{ik}=0)}.$$

<u>Note</u>

1. The log odds ratios $\log\Psi_{ijk} \in (-\infty, \infty)$ are symmetric about 0 and not constrained by the marginal means.
2. Interpretation: $\Psi = 1$ or $\log\Psi = 0$ implies (Y_1, Y_2) are uncorrelated.
3. Invariant to marginal specification of μ_1 and μ_2 (applicable in case control studies).

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

$$\Psi_{ijk} = \frac{\pi_{ijk}(1 - \mu_{ij} - \mu_{ik} + \pi_{ijk})}{(\mu_{ij} - \pi_{ijk})(\mu_{ik} - \pi_{ijk})},$$

$$\pi_{ijk} = \frac{A - \left[A^2 - 4(\Psi_{ijk} - 1)\Psi_{ijk}\mu_{ij}\mu_{ik}\right]^{1/2}}{2(\Psi_{ijk} - 1)},$$

$$A = 1 - (\mu_{ij} + \mu_{ik})(1 - \Psi_{ijk}).$$

GEE1.5 - Odds Ratios

- Paired models for odds ratios (Lipsitz et al., 1991)

  Mean model: $\text{logit}\mu_{ij} = X_{ij}^T\beta$,

  Correlation model: $\log \Psi_{ijk} = Z_{ijk}^T\alpha$.

- Alternating logistic regression (ALR) (Carey et al., 1993).
  Let $\gamma_{ijk} = \log \Psi_{ijk} = Z_{ijk}^T\alpha$. We consider the pairwise conditional expectations:

  $$\text{logit}E(Y_{ij}|Y_{ik}, X_i) = \Delta_{ijk} + \gamma_{ijk}Y_{ik}$$

where $\Delta_{ijk} = \log\left(\frac{\mu_{ij}-\pi_{ijk}}{1-\mu_{ij}-\mu_{ik}+\pi_{ijk}}\right)$.
An estimator for α could be obtained by alternating the following two steps until convergence.

1. A logistic regression of Y_{ij} on $X_{ij} => \hat{\beta}$
2. A logistic regression of Y_{ij} on $Z_{ijk}^T Y_{ik}$ with offset Δ_{ijk} (a fixed constant in the regression model) $=> \hat{\alpha}$.
   Note that the offset depends on both α and β and it is calculated using current estimates.

Formally the ALR uses the same model as in Lipsitz et al. (1991) but uses this estimating equation for α

$$U_\alpha(\beta,\alpha) = \sum_{i=1}^{m} F_i^T(\beta,\alpha)\tilde{W}_i^{-1}(\beta,\alpha)T_i(\beta,\alpha),$$

where

$$T_i(\beta,\alpha) = vec(Y_{ij} - \zeta_{ijk}),$$

$$\zeta_{ijk} = E(Y_{ij}|Y_{ik}),$$

$$\tilde{W}_i^{-1}(\beta,\alpha) = diag(var(Y_{ij}|Y_{ik})) = diag\left(\zeta_{ijk}(1-\zeta_{ijk})\right),$$

$$F_i = \frac{\partial\zeta_i}{\partial\alpha}.$$

- The ALR α is more efficient than 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$(\hat{\beta}, \hat{\alpha})$ 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 estimation equation for the scale parameter:

$$g_3(\phi_{ij}) = T_{3i}^T \gamma$$

and

$$U_\phi = \sum_{i=1}^{m} D_{3i}^T V_{3i} (S_i - \phi_i) = 0$$

where $S_{ij} = \frac{(Y_{ij} - \mu_{ij})^2}{v_{ij}}$.

The method is implemented in R package geepack.

Example

GEE2 - Joint Estimating Equations

- Prentice and Zhao (1991) considered $\delta = (\beta, \alpha)$ as the parameter and the optimal estimating function for δ.

- Paired models:

$$g_1(\mu_{ij}) = X_{ij}^T \beta,$$
$$g_2(\sigma_{ijk}) = Z_{ijk}^T \alpha,$$

where $\sigma_{ijk} = cov(Y_{ij}, Y_{ik})$.

- Optimal estimating equations for $\delta = (\beta, \alpha)$:

$$\begin{aligned} U(\delta) &= \sum_{i=1}^{m} D_i^T(\delta) V_i^{-1}(\delta) T_i(\delta) \\ &= \sum_{i=1}^{m} \begin{pmatrix} \frac{\partial \mu_i}{\partial \beta} & \frac{\partial \sigma_i}{\partial \beta} \\ 0 & \frac{\partial \sigma_i}{\partial \alpha} \end{pmatrix}^T \begin{pmatrix} V_i(1,1) & V_i(1,2) \\ V_i(1,2)^T & V_i(2,2) \end{pmatrix}^{-1} \begin{pmatrix} Y_i - \mu_i \\ S_i - \sigma_i \end{pmatrix} \end{aligned}$$

where $V_i(1,1) = cov(Y_i)$, $V_i(1,2) = cov(Y_i, S_i)$, $V_i(2,2) = cov(S_i)$, and $S_{ijk} = (Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})$.

- Note

  – First and second moment models are not enough to obtain $V_i(1,2)$ and $V_i(2,2)$.
  – Maximum likelihood method can be used if we specify all moments.
  – In GEE2, a working 3rd/4th moment model is used.
  – GEE2 equations can be derived as the score equations for a quadratic exponential family (QEF) model:

$$l_i = \theta_{1i}^T Y_i + \theta_{2i}^T S_i + \delta_i + c_i(Y_i).$$

- 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

$$V_i(1,2) = 0,$$
$$V_i(2,2): \ cov(s_{ijk}, s_{ilm}) = \sigma_{ijl}\sigma_{ikm} + \sigma_{ijm}\sigma_{ikl}.$$

- Liang et al. (1992) considered GEE2 model for binary data using odds ratios. One working 3rd/4th moment model is to fix marginal 3-way and 4-way log odds ratio contrasts at 0.

- 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 $\hat{\beta}$ and $\hat{\alpha}$ depends on the correct modeling of both mean and variance.

- 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), solution 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?

References

- Carey V, Zeger SL, and Diggle P (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika.* **80**: 517-526.
- Liang KY and Zeger SL (1986). Longitudinal data analysis using generalized linear models. *Biometrika.* **73**: 13-22.
- Liang KY, Zeger SL, and Qaqish B (1992). Multivariate regression analyses for categorical data (with discussion). *Journal of the Royal Statistical Society, Series B. 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**: 153-160.
- Prentice RL (1988). Correlated binary regression with covariates specific to each binary observations. *Biometrics.* **44**: 1033-1048.
- Prentice RL and Zhao LP (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics.* **47**: 825-839.
- Yan J and Fine J (2004). Estimating equations for association structures. *Statistics in Medicine* **23**: 859-874.