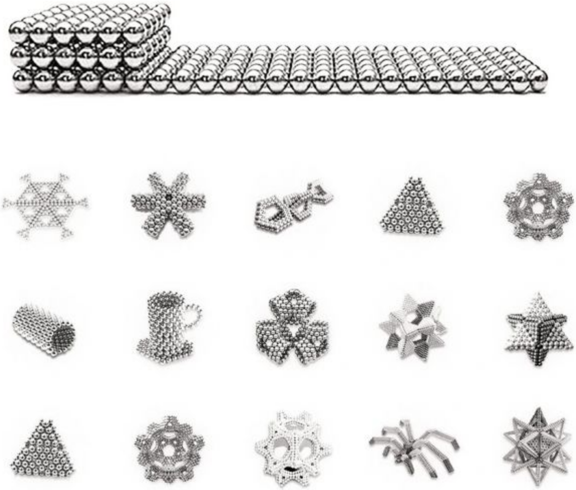
Lecture 17. Marginal Models

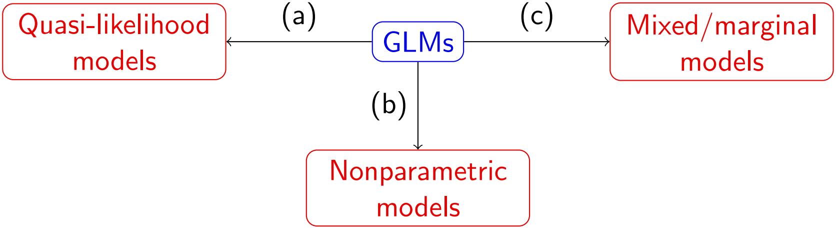
Nan Ye

School of Mathematics and Physics
University of Queensland

Competition Prize



Recall: Extending GLMs



(a) Relax assumption on the random component.

(b) Relax assumption on the systematic component.

(c) Relax assumption on the data (independence).

Recall: Correlated Data

So far...

- We have been working under the assumption that the responses are independent given the covariates.
- This assumption does not hold for many problems.

Examples of correlated responses

- Measurements on clusters of subjects
    - e.g. measurements on patients from the same hospital may be correlated because they are attended by the same set of nurses and doctors, and they are likely to share demographic or socio-economic features.
- Repeated measurements on same subject

Recall: mixed models

A generalized linear mixed model has the following structure

$$\mathbb{E}(Y_{ij} \mid \mathbf{x}_{ij}, \mathbf{z}_{ij}, \alpha_i) = h(\mathbf{x}_{ij}^\top \beta + \mathbf{z}_{ij}^\top \alpha_i),$$
$$Y_{ij} \mid \mathbf{x}_{ij}, \mathbf{z}_{ij}, \alpha_i \sim \text{ an exponential family distribution},$$
$$\alpha_j \overset{ind}{\sim} N(0, \Sigma_A).$$

What if we are policy makers responsible for controlling what $\mathbf{x}_{ij}$ should be for the whole population?

If we are policy makers...

- We will be mainly concerned with how Y_{ij}'s depend on $\mathbf{x}_{ij}$ for the population.
- However, we need to take correlation between the responses into account when building such a model.
- This leads to the class of marginal models.

This Lecture

Marginal models for correlated data

- Model structure and examples
- Comparison with mixed models
- More on specifying association structures
- Parameter estimation
- Example

Structure of Marginal Models

- Recall: A GLM has the following structure

$$\begin{aligned} \text{(systematic)} \quad & \mathbb{E}(Y \mid \mathbf{x}) = h(\beta^\top \mathbf{x}), \\ \text{(random)} \quad & Y \mid \mathbf{x} \text{ follows an exponential family distribution.} \end{aligned}$$

- A marginal model has the following structure

$$\begin{aligned} \text{(systematic)} \quad & \mu_{ij} = \mathbb{E}(Y_{ij} \mid \mathbf{x}_{ij}) = h(\beta^\top \mathbf{x}_{ij}), \\ \text{(random)} \quad & \text{var}(Y_{ij} \mid \mathbf{x}_{ij}) = \phi V(\mu_{ij}) \\ & \text{association}(Y_{ij}, Y_{ij'}) = C(\mu_{ij}, \mu_{ij'}, \gamma), \end{aligned}$$

  where C is a function, and γ is called an association parameter.
- β is often estimated using the generalized estimating equation.

Remarks

- The first two components of a marginal model have the same form as a quasi-likelihood model.
- The association component usually measures the correlation for continuous random variables, but in general this just needs to be some numerical measures reflecting how responses correlate with each other.
- We assume that responses from different clusters are uncorrelated.
- The generalized estimating equation is a generalization of quasi-score equation (more on this later).

Example 1. Continuous longitudinal data

- Assume we have K subjects, and we measure a continuous response of subject i at time $t_{i1}, \ldots, t_{in_i}$.
- A marginal model for such data may have the following structure

$$\mu_{ij} = \mathbb{E}(Y_{ij} \mid \mathbf{x}_{ij}) = \beta^\top \mathbf{x}_{ij},$$
$$\text{var}(Y_{ij} \mid \mathbf{x}_{ij}) = \phi,$$
$$\text{corr}(Y_{ij}, Y_{ij'}) = \gamma^{t_{ij} - t_{ij'}},$$

  where $0 \leq \gamma \leq 1$.
- That is, we assume an identity link and a constant variance for each Y_{ij}, and we assume that Y_{ij} and $Y_{ij'}$ are less correlated if they are further apart in time.

Example 2. Clustered count data

- Assume we have K subjects, and there are n_i measurements on a count response for subject i.
- A marginal model for such data may have the following structure

$$\mu_{ij} = \mathbb{E}(Y_{ij} \mid \mathbf{x}_{ij}) = \exp(\beta^\top \mathbf{x}_{ij}),$$
$$\text{var}(Y_{ij} \mid \mathbf{x}_{ij}) = \phi\mu_{ij}$$
$$\text{corr}(Y_{ij}, Y_{ij'}) = \gamma_{jj'},$$

  where $-1 \leq \gamma_{jj'} \leq 1$.
- That is, we assume a log-link, a linear variance, and a completely unstructured association.

Example 3. Clustered binary data

- Assume we have K subjects, and there are n_i measurements on a binary response for subject i.
- A marginal model for such data may have the following structure

$$\begin{aligned} \mu_{ij} = \mathbb{E}(Y_{ij} \mid \mathbf{x}_{ij}) &= \text{logistic}(\beta^\top \mathbf{x}_{ij}), \\ \text{var}(Y_{ij} \mid \mathbf{x}_{ij}) &= \phi\mu_{ij}(1 - \mu_{ij}) \\ \ln \text{OR}(Y_{ij}, Y_{ij'}) &= \gamma_{jj'}, \end{aligned}$$

  where the odds ratio is defined as $\text{OR}(Y_{ij}, Y_{ij'}) = \frac{P(Y_{ij}=1, Y_{ij'}=1)P(Y_{ij}=0, Y_{ij'}=1)}{P(Y_{ij}=1, Y_{ij'}=0)P(Y_{ij}=0, Y_{ij'}=0)}$, and $-\infty \leq \gamma_{jj'} \leq \infty$.
- That is, we assume a log-link, a linear variance, and a completely unstructured association.

Marginal models vs. mixed models

- Mixed models are subject-specific models
  - They are used to answer questions about what we can do for individuals or clusters.
  - Correlation is modelled by modifying the systematic component of GLMs by adding a subject-specific (or cluster-specific) term to the linear predictor.
- Marginal model are population-averaged models
  - They are used to answer questions about what we can do for the population.
  - Correlation is modelled by introducing a component to capture within-cluster correlation.

Some Association Structures

Independence correlation

- This is used when there is no correlation within the clusters.
- The correlation matrix is the identity matrix.
- Example correlation matrix for Y_{11}, Y_{12}, Y_{13}

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Exchangeable correlation

- This is used when the within cluster correlation is constant.
- Example correlation matrix for Y_{11}, Y_{12}, Y_{13}

$$\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

Autoregressive correlation

- This is used when the within cluster responses are assumed to be a time series, or more precisely, an autoregressive process.
- In an autoregressive process of order m (AR(m)), the next response is a linear combination of previous m responses and a white noise.
- For AR(1), if Y_{11}, Y_{12}, Y_{13} are responses taken at three consecutive time steps, then the correlation matrix has the form

$$\begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix},$$

that is, $\text{corr}(Y_{1j}, Y_{1j'}) = \rho^{|j-j'|}$.

Parameter Estimation

Revisiting parameter estimation for quasi-models

- Recall: for GLM and quasi-likelihood model, an estimate of β is obtained by solving

$$\sum_i \frac{y_i - \mu_i}{g'(\mu_i) V_i} \mathbf{x}_i = 0,$$

  where $\mu_i(\beta) = \mathbb{E}(Y_i \mid \mathbf{x}_i, \beta)$, and $V_i = \text{var}(Y_i \mid \mathbf{x}_i, \beta)$.
- We can show that $\nabla \mu_i(\beta) = \mathbf{x}_i / g'(\mu_i)$, thus the above equation can be written as

$$\sum_i \frac{y_i - \mu_i}{V_i} \nabla \mu_i(\beta) = 0.$$

- We can further write the equation in matrix notation as

$$\left(\frac{\partial \boldsymbol{\mu}}{\partial \beta}\right)^{\top} \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = 0,$$

  where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$, $\frac{\partial \boldsymbol{\mu}}{\partial \beta}$ is the Jacobian of $\boldsymbol{\mu}$ ($\nabla \mu_i(\beta)$ is the i-th row of the Jacobian), and $\mathbf{V} = \text{diag}(V_1, \ldots, V_n)$ is the covariance matrix.
- This can be applied to marginal models by simply replacing $\mathbf{V}$ using the covariance matrix of marginal models!

Generalized Estimating Equation (GEE)

- The GEE for a marginal model is

$$\sum_i \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \beta} \right)^\top \mathbf{V}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = 0,$$

  where $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{in_i})$, $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})$, and $\mathbf{V}_i = \mathrm{diag}(V_{i1}, \ldots, V_{in_i})^{1/2} \, \mathrm{corr}(\mathbf{y}_i) \, \mathrm{diag}(V_{i1}, \ldots, V_{in_i})^{1/2}$ are the vector of means, the vector of responses, and the covariance matrix for cluster i respectively.

- The GEE can be solved by iterating between a modified Fisher scoring algorithm for solving β given $\mathbf{V}_i$'s, and estimating $\mathbf{V}_i$'s using the residuals for given β.

Properties

- When the correlation matrix is chosen to be an identity matrix, the GEE reduces to the quasi-score equation for the corresponding quasi-model (marginal model with the association component removed).
- If the data actually satisfies $\mathbb{E}(Y_{ij} \mid \mathbf{x}_{ij}) = \mathbf{x}_{ij}^\top \beta^*$ for some β^*, then under certain regularity conditions, the estimate $\hat{\beta}$ given K clusters is asymptotically normally distributed with mean β^* and covariance $\mathbf{V}^*/K$, where $\mathbf{V}^*$ depends on both the true covariance and the assumed covariance.
- The GEE approach usually yields similar β as the corresponding quasi-model, but is able to adjust the standard errors using the empirical covariance so as to provide more accurate estimates for the standard errors.

Example

Data

```
> library(gee)
> dim(warpbreaks)
[1] 54  3
> head(warpbreaks)
  breaks wool tension
1     26    A       L
2     30    A       L
3     54    A       L
4     25    A       L
5     70    A       L
6     52    A       L
```

- The data records the number of warp breaks per loom (a fixed length of yarn).
- Three levels of tension (L, M, H) and two different types of wool (A and B).
- We want to see how tension affects the number of breaks.

Quasi-Poisson regression

```
> fit.po = glm(breaks ~ tension, data=warpbreaks,
    family=quasipoisson)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.59426    0.08382  42.881  < 2e-16 ***
tensionM    -0.32132    0.12928  -2.485 0.016260 *
tensionH    -0.51849    0.13721  -3.779 0.000414 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Marginal Poisson model with independence correlation

```
> fit.po.ex = gee(breaks ~ tension, id=wool, data=warpbreaks,
    family=poisson, corstr="independence")
Coefficients:
              Estimate Naive S.E.   Naive z Robust S.E.  Robust z
(Intercept)  3.5942635 0.08382008 42.880698  0.15869419 22.648992
tensionM    -0.3213204 0.12928272 -2.485409  0.22270597 -1.442801
tensionH    -0.5184885 0.13720625 -3.778898  0.06441329 -8.049403

Estimated Scale Parameter:  4.601903
Number of Iterations:  1
```

- We use independence correlation structure, and thus the marginal model should be the same as quasi-Poisson.
- Indeed, we recover the same estimates for the coefficients, and the naive S.E. and naive z are the same as those given by the quasi-model.
- The robust S.E. and robust z from gee are quite different though.

Marginal Poisson model with exchangeable correlation

```
> fit.po.ex = gee(breaks ~ tension, id=wool, data=warpbreaks,
    family=poisson, corstr="exchangeable")
Coefficients:
              Estimate Naive S.E.   Naive z Robust S.E.  Robust z
(Intercept)  3.5942635 0.09055356 39.692126  0.15869419 22.648992
tensionM    -0.3213204 0.12808197 -2.508709  0.22270597 -1.442801
tensionH    -0.5184885 0.13619100 -3.807069  0.06441329 -8.049403

Estimated Scale Parameter:  4.601903
Number of Iterations:  1
```

- It may be more realistic to assume the presence of within cluster correlation.
- With an exchangeable correlation structure, we get different naive S.E. values but the same parameter estimates and robust S.E. values.

Marginal Poisson model with AR(1) correlation

```
> fit.po.arm = gee(breaks ~ tension, id=wool, data=warpbreaks,
    family=poisson, corstr="AR-M", Mv=1)
> summary(fit.po.arm)
Coefficients:
              Estimate Naive S.E.   Naive z Robust S.E.  Robust z
(Intercept)  3.5949833 0.08689454 41.371797  0.15918021 22.584360
tensionM    -0.3245601 0.13381637 -2.425414  0.22496947 -1.442685
tensionH    -0.5178782 0.14221001 -3.641644  0.06567149 -7.885892

Estimated Scale Parameter:  4.601424
Number of Iterations:  2
```

- If the measurements are taken in consecutive time steps within the same group, an autoregressive correlation structure can be more appropriate.
- We get slightly different parameter estimates and robust S.E. values.

- We get similar parameter estimates using different correlation structures.
- Naive S.E. values change quite significantly in some cases, but robust S.E. values are quite similar using different correlation structures.

What You Need to Know

Marginal models for correlated data

- Marginal model extends quasi-model with a correlation structure.
- Marginal models are population-averaged, but mixed models are subject-specific.
- Some common association structures.
- Parameter estimation using GEE and robust standard errors.