Stat 579: Generalized Linear Models and Extensions

Linear Mixed Models for Longitudinal Data

Yan Lu

April, 2018, week 15

THE UNIVERSITY OF NEW MEXICO

Data structure

		t1	t2	$\cdots$	tn_i
Experimental	1st subject	y_{11}	y_{12}		y_{1n_1}
	2nd subject	y_{21}	y_{22}		y_{2n_2}
	$\vdots$				
	m_1 th subject	$y_{m_1 1}$	$y_{m_1 2}$		$y_{m_1 n_{m_1}}$
Control	1st subject	$y_{m_1+1,1}$	$y_{m_1+1,2}$		$y_{m_1+1,n_{m_1+1}}$
	2nd subject	$y_{m_1+2,1}$	$y_{m_1+2,2}$		$y_{m_1+2,n_{m_1+2}}$
	$\vdots$				
	m th subject	y_{m1}	y_{m2}		y_{mn_m}

Linear mixed model

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \tag{1}$$

where

$$\boldsymbol{\beta}_g = \begin{bmatrix} \beta_{g0} \\ \beta_{g1} \end{bmatrix}, \boldsymbol{\beta}_B = \begin{bmatrix} \beta_{B0} \\ \beta_{B1} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_g \\ \boldsymbol{\beta}_B \end{bmatrix} = \begin{bmatrix} \beta_{g0} \\ \beta_{g1} \\ \beta_{B0} \\ \beta_{B1} \end{bmatrix}, \mathbf{b}_i = \begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix}$$

$$\mathbf{b}_i \sim N(0, \mathbf{g}) \text{ indep } \boldsymbol{\epsilon}_i \sim N(0, \mathbf{R}_i)$$

Therefore,

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{g}\mathbf{Z}_i' + \mathbf{R}_i)$$

$$\mathbf{Y}_i = \begin{cases} \mathbf{X}_i\boldsymbol{\beta}_g + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i & \text{child i is a girl} \\ \mathbf{X}_i\boldsymbol{\beta}_B + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i & \text{child i is a boy} \end{cases}$$

where

$$\mathbf{Z}_i = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{bmatrix}, \mathbf{X}_i = \begin{bmatrix} \delta_i & \delta_i t_{i1} & (1-\delta_i) & (1-\delta_i)t_{i1} \\ \delta_i & \delta_i t_{i2} & (1-\delta_i) & (1-\delta_i)t_{i2} \\ \vdots & \vdots & \vdots & \vdots \\ \delta_i & \delta_i t_{in_i} & (1-\delta_i) & (1-\delta_i)t_{in_i} \end{bmatrix}$$

with

$$\delta_i = \begin{cases} 1 & \text{girls} \\ 0 & \text{boys} \end{cases}$$

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \epsilon_i$$

- The parameters of the model are β and those in $\mathbf{g}$ and $\mathbf{R}_i$.
- The $\mathbf{b}_i$ are unknown random variables, while $\mathbf{Z}_i$ is fixed and known.
- The likelihood function is based on the marginal distribution

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\beta, \mathbf{Z}_i\mathbf{g}\mathbf{Z}_i' + \mathbf{R}_i)$$

where responses $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_m$ for different individuals are assumed independent.
——same form as the longitudinal models

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\beta, \mathbf{\Sigma}_i)$$

——where here $\mathbf{\Sigma}_i = \mathbf{Z}_i\mathbf{g}\mathbf{Z}_i' + \mathbf{R}_i$

- Thus standard ML and REML can be used for inference on $\boldsymbol{\beta}$, whereas AIC/BIC may be used for selections of appropriate covariance structures.

We can also state model (1) as

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \epsilon \qquad (2)$$

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_m \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_m \end{bmatrix} \beta + \begin{bmatrix} \mathbf{Z}_1 & 0 & 0 & \cdots & 0 \\ 0 & \mathbf{Z}_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \mathbf{Z}_m \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{bmatrix}$$

where $E(\epsilon) = \mathbf{0}$, $E(\mathbf{b}) = \mathbf{0}$ and

$$var(\epsilon) = \begin{bmatrix} \mathbf{R}_1 & 0 & 0 & 0 \\ 0 & \mathbf{R}_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \mathbf{R}_m \end{bmatrix} = \mathbf{R}$$

and

$$var(\mathbf{b}) = \begin{bmatrix} \mathbf{g} & 0 & 0 & 0 \\ 0 & \mathbf{g} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \mathbf{g} \end{bmatrix} = \mathbf{G}$$

- Note that $\mathbf{G}$ and $\mathbf{g}$ are different.
- $\mathbf{R}$ and $\mathbf{G}$ are block diagonal which reflects assumptions that $\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_m$ are mutually independent, so are $\epsilon_1, \epsilon_2, \cdots, \epsilon_m$.
- As before, we also assume the $\mathbf{b}_i$s and ϵ_is are independent

Concisely, the linear mixed model for longitudinal data can be written as

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \epsilon$$

with

$$\mathbf{b} \sim N(\mathbf{0}, \mathbf{G}) \text{ indep of } \epsilon \sim N(\mathbf{0}, \mathbf{R}).$$

- $$\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{V}),$$

——where $var(\mathbf{Y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$
——-$\mathbf{V}$ is block-diagonal with ith diagonal block $var(\mathbf{Y}_i) = \mathbf{Z}_i\mathbf{g}\mathbf{Z}_i' + \mathbf{R}_i$

Model (1) includes many settings for special case

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \epsilon_i$$

- the standard linear regression model, which excludes $\mathbf{Z}_i\mathbf{b}_i$ and assumes that $\mathbf{R}_i = \sigma^2\mathbf{I}_{n_i}$ independent errors with constant variance (same is true for standard ANOVA)
- longitudinal model from week 13 which also excludes $\mathbf{Z}_i\mathbf{b}_i$
- the random coefficient model where $\mathbf{X}_i\beta$ was $\mathbf{Z}_i\beta$
- the split plot, randomized block and one-way random effects model

when we assume $E(\mathbf{b}_i) = 0$, we typically are thinking of $\mathbf{b}_i$ as subject specific deviations, which makes sense when the effects in $\mathbf{Z}_i$ are contained in $\mathbf{X}_i$.

One-way random effects model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, i = 1, 2, \cdots, m, j = 1, \cdots, n_i \tag{3}$$

$$\mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{pmatrix}, \boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix}, \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix}, \mathbf{X}_i = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{n_i \times 1}$$

- $\boldsymbol{\alpha} \sim N(\mathbf{0}, \sigma_\alpha^2 \mathbf{I}_m)$
  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$
  $\boldsymbol{\alpha}$ and $\boldsymbol{\epsilon}$ independent
- $\mathbf{Y}_i = \mathbf{X}_i \mu + \mathbf{Z}_i \alpha_i + \boldsymbol{\epsilon}_i$, where $\mathbf{Z}_i = \mathbf{X}_i$

$$\beta = \mu, \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_m \end{pmatrix}_{mk \times 1}, \mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{pmatrix},$$

$$\mathbf{Z} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \end{pmatrix}_{mk \times m}$$

One way random effects model (3) can be written as

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$$

where

- $\mu = \beta, \boldsymbol{\alpha} = \mathbf{b}$, $\mathbf{J}_{n_i}$ is an $n_i \times 1$ vector of 1s, $\mathbf{g} = \sigma_\alpha^2$ and $\mathbf{R}_i = \sigma_e^2 \mathbf{I}_{n_i}$, $\mathbf{I}_{n_i}$ is an $n_i \times 1$ identity matrix
- which implies $\mathbf{Z}_i \mathbf{g} \mathbf{Z}_i' + \mathbf{R}_i = \sigma_\alpha^2 \mathbf{J}_{n_i} + \sigma_e^2 \mathbf{I}_{n_i}$

Best linear unbiased prediction (BLUPS)

The expected response for subject i conditional on $\mathbf{b}_i$ is

$$E(\mathbf{Y}_i|\mathbf{b}_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$$

Unconditional expectation, averaging over all subjects is

$$\begin{aligned} E(\mathbf{Y}_i) &= E\left\{E(\mathbf{Y}_i|\mathbf{b}_i)\right\} \\ &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_iE(\mathbf{b}_i) = \mathbf{X}_i\boldsymbol{\beta} \end{aligned}$$

Given the ML or REML estiator of $\boldsymbol{\beta}$, say $\hat{\boldsymbol{\beta}}$, we estimate the unconditional (or pop averaged) mean with $\mathbf{X}_i\hat{\boldsymbol{\beta}}$

How do we estimate (actually predict) the subject specific $E(\mathbf{Y}_i|\mathbf{b}_i) = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i$

- It is reasonable to think $\mathbf{X}_i\hat{\beta} + \mathbf{Z}_i\hat{\mathbf{b}}_i$ for a suitable chosen $\hat{\mathbf{b}}_i$.
- Statistical theory suggests that the best prediction for the random variable $\mathbf{b}_i$ based on $\mathbf{Y}_i$ is $E(\mathbf{b}_i|\mathbf{Y}_i)$.
  —— Under normality, this can be shown to equal

$$E(\mathbf{b}_i|\mathbf{Y}_i) = \mathbf{g}\mathbf{Z}_i'\mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\beta)$$

——This follows because $\mathbf{Y}_i|\mathbf{b}_i \sim N(\mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i, \mathbf{R}_i)$ and $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{g})$ implies $(\mathbf{Y}_i, \mathbf{b}_i)$ has a multivariate normal distribution. It is well known that $\mathbf{b}_i|\mathbf{Y}_i$ is normal with

$$E(\mathbf{b}_i|\mathbf{Y}_i) = E(\mathbf{b}_i) + cov(\mathbf{b}_i, \mathbf{Y}_i)var(\mathbf{Y}_i)^{-1}(\mathbf{Y}_i - E(\mathbf{Y}_i))$$

$$= \mathbf{0} + \mathbf{g}\mathbf{Z}_i'\mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\beta)$$

Here $\mathbf{V}_i = var(\mathbf{Y}_i)$ and

$$\begin{aligned} cov(\mathbf{b}_i, \mathbf{Y}_i) &= cov(\mathbf{b}_i, \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i) \\ &= cov(\mathbf{b}_i, \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i) \\ &= cov(\mathbf{b}_i, \mathbf{Z}_i\mathbf{b}_i) + cov(\mathbf{b}_i, \boldsymbol{\epsilon}_i) \\ &= cov(\mathbf{b}_i, \mathbf{b}_i)\mathbf{Z}_i' \\ &= \mathbf{g}\mathbf{Z}_i' \end{aligned}$$

So result follows.

Best Prediction

$$E(\mathbf{b}_i|\mathbf{Y}_i) = \mathbf{g}\mathbf{Z}_i'\mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{X}\beta)$$

Replace unknown parameters $\mathbf{g}, \mathbf{V}_i$ and β by MLES or REMLs gives

$$\hat{\mathbf{b}}_i = \hat{\mathbf{g}}\mathbf{Z}_i'\hat{\mathbf{V}}_i^{-1}(\mathbf{Y}_i - \mathbf{X}\hat{\beta})$$

which is often called the empirical estimated BLUP (best unbiased predictor).

Thus the subject specific trajectory $\mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i$ is predicted via

$$\begin{aligned} \mathbf{X}_i\hat{\beta} + \mathbf{Z}_i\hat{\mathbf{b}}_i &= \mathbf{X}_i\hat{\beta} + \mathbf{Z}_i\hat{\mathbf{g}}\mathbf{Z}_i'\hat{\mathbf{V}}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta}) \\ &= (\mathbf{I} - \mathbf{Z}_i\hat{\mathbf{g}}\mathbf{Z}_i'\hat{\mathbf{V}}_i^{-1})\mathbf{X}_i\hat{\beta} + \mathbf{Z}_i\hat{\mathbf{g}}\mathbf{Z}_i'\hat{\mathbf{V}}_i^{-1}\mathbf{Y}_i \\ &= (\mathbf{I} - \mathbf{W}_i)\mathbf{X}_i\hat{\beta} + \mathbf{W}_i\mathbf{Y}_i \end{aligned}$$

If you consider $\mathbf{W}_i$ a "weight", then our base prediction is a weighted average of an individual's response Y_i and the estimated mean response $\mathbf{X}_i\hat{\beta}$, averaged over all individuals with same design matrix $\mathbf{X}_i$ as subject i.
Note that prediction in the one-way random effects ANOVA we give before are special cases of the above results.

Tests on variance component

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}$$

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix}\right)$$

Suppose we wish to test

$$H_0 : \sigma_1^2 = 0 \text{ i.e.,} b_{1i} = 0 \text{ for all } i.$$

In this case β_1 is the slope of each individual's regression line. This test can be done informally using AIC/BIC (i.e., compare null model to alternative model with arbitrary σ_1^2), or using LR test

$$LR = -2\log(\hat{L}_{red}/\hat{L}_{full})$$

where $\hat{L}_{red}$ and $\hat{L}_{full}$ are the maximized log-likelihood for the reduced and full models.

Boundry correction

H_0 specifies that σ_1^2 lies on the boundary of possible value for σ_1^2 (i.e., $\sigma_1^2 \geq 0$).

- The standard theory for LR tests doesn't hold.
- One can show under H_0,

$$LR \sim 0.5\chi_1^2 + 0.5\chi_2^2$$

  i.e., LR statistic has a null distribution that is a 50:50 mixture of χ_1^2 and χ_2^2 distribution (rather than usual) χ_1^2.
  —— Result mixture holds since we are comparing 2 nested models, one with 1 random effect b_{0i}, the other with two b_{0i} and b_{1i}.
- More generally if we are comparing nested models with q and $q+1$ correlated random-effects (which still differ by 1 random effect), then $LR \sim 0.5\chi_q^2 + 0.5\chi_{q+1}^2$

Time dependent covariance

- Mixed models allow for a variety of complications with longitudinal data, for example
  ——dropout/missing
  ——covariates in fixed effects can either be static (constant over time) such as baseline measurements or time dependent such as time.
- IN some cases, the treatment (or group) one is assigned changes with time. For example, you may be switched to placebo group if you are having all adverse reactions to a drug, or switched to a higher dose of a drug if given dose is not effective. Some care is needed to ensure proper inferences with these complications.

Generalized estimating equations (GEEs) Liang and Zeger (1986)

- Apply to non-normal data, shortage of joint distribution such as multivariate probability distribution
- GEE approach bypasses the distribution problem by focusing on modeling the mean function and correlation function only.
- An estimating equation is proposed for estimating the regression effects and then appropriate statistical theory is applied to generate standard errors and procedures for inference.

Standard longitudinal model for normal responses.

The standard model for longitudinal data can be written as

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

with $\epsilon \sim N(\mathbf{0}, \mathbf{\Sigma})$.

- $$\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{\Sigma}),$$

  ——where $\mathbf{\Sigma}$ is block-diagonal with ith diagonal block $\mathbf{\Sigma}_i$
  ——usually write $\mathbf{\Sigma} = \sigma^2\mathbf{R}$, $var(y_{ij}) = \sigma^2$, $cov(\mathbf{y}_i) = \mathbf{\Sigma}_i$, $\mathbf{\Sigma}_i = \sigma^2\mathbf{R}_i$
- Equivalently, $\mathbf{y}_i \sim N(\mathbf{X}_i\beta, \sigma^2\mathbf{R}_i)$

- $$\frac{\partial L}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2}\mathbf{X}'\mathbf{R}'(\mathbf{y}-\boldsymbol{\mu}) = \frac{1}{\sigma^2}\sum_{j=1}^{k}\mathbf{X}_j'\mathbf{R}_j^{-1}(\mathbf{y}_j-\boldsymbol{\mu}_j) = \mathbf{0} \quad (4)$$

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ &= \sum_{j=1}^{k}(\mathbf{X}_j'\mathbf{R}_j^{-1}\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{R}_j^{-1}\mathbf{y}_j \quad (5)\end{aligned}$$

- MLEs of σ^2 and $\mathbf{R}$ can be computed independently of $\boldsymbol{\beta}$, plugging these into (5) gives the overall MLE of $\boldsymbol{\beta}$.

Generalized linear models

$y_i \overset{indep}{\sim} Bin(n_i, \mu_i)$
A glm has the following form with some link function $g(\cdot)$
$g(\mu_i) = \eta_i = \mathbf{x}_i'\boldsymbol{\beta}$

$$g(\mu_i) = \begin{cases} \log\left(\dfrac{\mu_i}{1-\mu_i}\right), & \text{logit link: logistic regression} \\ \Phi^{-1}(\mu_i), & \text{Probit link: probit regression} \\ \log\{-\log(1-\mu_i)\} & \text{Complementary log-log link} \end{cases}$$

$$\mu_i = g^{-1}(\eta_i) = \begin{cases} \exp(\eta_i)/(1+\exp(\eta_i)), & \text{logit link} \\ \Phi(\eta_i), & \text{Probit link} \\ 1-\exp\{-\exp(\eta_i)\} & \text{Complementary log-log} \end{cases}$$

General longitudinal model

In a general case, $\mathbf{y} \sim N(\mathbf{X}\beta, \mathbf{\Sigma})$
$\hat{\beta} = (\mathbf{X}'\hat{\mathbf{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{\Sigma}}^{-1}\mathbf{y}$ is the MLE of β and $\hat{\mathbf{\Sigma}}$ is the MLE of $\mathbf{\Sigma}$.

- Key to generalizing glms to longitudinal responses is to modify equation (4) to accommodate $g(\mu_{ij}) = \mathbf{X}'_{ij}\beta$ instead of $\mu_{ij} = \mathbf{X}'_{ij}\beta$ for some link function $g(\cdot)$.
- If the responses within an individual were independent and satisfied an EF model, this would be a glm and the score function for β based solely on $\mathbf{y}_i$ would be

$$\frac{\partial l_i}{\partial \beta} = \mathbf{X}'_i \mathbf{w}_i \mathbf{\Delta}_i (\mathbf{y}_i - \boldsymbol{\mu}_i)$$

$$\mathbf{\Delta}_i = \text{diag}(g'(\mu_{ij}))$$

$$w_{ii} = \text{diag}\frac{w_{ij}}{\phi V(\mu_{ij}) g'(\mu_i)^2}$$

w_{ij}, weight for jth obsn in $\mathbf{y}_i$, i.e., y_{ij}

- Assuming responses between individuals are independent, the score function based on the sample is

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{k} \frac{\partial L_j}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{k} \mathbf{X}_j' \mathbf{w}_j \boldsymbol{\Delta}_j (\mathbf{y}_j - \boldsymbol{\mu}_j)$$

- Recall the score function under normality,

$$0 = \frac{1}{\sigma^2} \sum_{j=1}^{k} \mathbf{X}_j' \mathbf{R}_j^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_j)$$

  $\mathbf{R}_j = \mathbf{I}_{n_j}$, corresponds to the case where responses within an individual are independent.

- Idea behind GEEs for glms is to take the score function when responses within an individual are independent. $\mathbf{0} = \sum_{j=1}^{k} \mathbf{X}_j' \mathbf{w}_j \boldsymbol{\Delta}_j (\mathbf{y}_j - \boldsymbol{\mu}_j)$ and replace $\mathbf{w}_j \boldsymbol{\Delta}_j$ by a matrix that captures the within individual correlation structure, but reduces to $\mathbf{w}_j \boldsymbol{\Delta}_j$ if responses within an individual are independent.

SAS

$$
\begin{aligned}
\mathbf{w}_i \boldsymbol{\Delta}_i &= \operatorname{diag}\left(\frac{1}{\phi} \cdot \frac{w_{ij}}{V(\mu_{ij}) g'(\mu_i)^2} \cdot g'(\mu_{ij})\right) \\
&= \operatorname{diag}\left(\frac{w_{ij}}{\phi V(\mu_{ij})} \cdot \frac{1}{g'(\mu_{ij})}\right) \\
&= \operatorname{diag}\left(\frac{1}{g'(\mu_{ij})}\right) \operatorname{diag}\left(\frac{w_{ij}}{\phi V(\mu_{ij})}\right) \\
&= \boldsymbol{\Delta}_i^{-1} \mathbf{V}_i^{-1}
\end{aligned}
$$

Recall for a glm that

$$var(y_{ij}) = \frac{\phi V(\mu_{ij})}{w_{ij}}$$

ϕ is the dispersion parameter, so if responses within an individual were independent, then $var(\mathbf{y}_i) = \mathbf{V}_i$ and the score function for $\boldsymbol{\beta}$ for a glm when responses within an individual are independent reduces to

$$\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{\beta}} &= \sum_{j=1}^{k} \mathbf{X}_j' \mathbf{w}_j \boldsymbol{\Delta}_j (\mathbf{y}_j - \boldsymbol{\mu}_j) \\
&= \sum_{j=1}^{k} \mathbf{X}_j' \boldsymbol{\Delta}_j^{-1} \mathbf{V}_j^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_j) \\
&= \sum_{j=1}^{k} \mathbf{D}_j' \mathbf{V}_j^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_j) = \mathbf{0}
\end{aligned}$$

$\mathbf{D}_j$ depends on $\mathbf{X}_j$ and link function, but not the variance function.

GEE extension

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{k} \mathbf{D}_j' \mathbf{V}_j^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_j) = \mathbf{0} \tag{6}$$

GEE extension replaces $\mathbf{V}_i$ as defined above by a covariance matrix for $\mathbf{y}_i$ that reflects the correlation structure for a longitudinal response vector.

$$Var(\mathbf{y}_i) = \phi \mathbf{A}_i^{\frac{1}{2}} \mathbf{W}_i^{*-\frac{1}{2}} \mathbf{R}_i(\alpha) \mathbf{W}_i^{*-\frac{1}{2}} \mathbf{A}_i^{\frac{1}{2}}$$

where $\mathbf{A}_i = \text{diag}(V(\boldsymbol{\mu}_{ij}))$, $w_i^* = \text{diag}(\mathbf{w}_{ij})$, so that

$\mathbf{A}_i^{0.5} = \text{diag}\left(\sqrt{V(\boldsymbol{\mu}_{ij})}\right)$

$(w_i^*)^{-0.5} = \text{diag}\left(\frac{1}{\sqrt{\mathbf{w}_{ij}}}\right)$

$A_i^{0.5}(w_i^*)^{-0.5} = \text{diag}\left(\sqrt{\frac{V(\mu_{ij})}{\mathbf{w}_{ij}}}\right)$

$\mathbf{R}_i(\alpha) = n_i * n_i$ correlation matrix for $\mathbf{y}_i$ which depends on a set $\boldsymbol{\alpha}$ of correlation parameters. If responses within an individual are independent, $\mathbf{R}_i(\boldsymbol{\alpha}) = \mathbf{I}_{n_i}$ and thus

$$
\begin{aligned}
V_i &= \phi A_i^{0.5} w_i^{*\,-0.5} w_i^{*\,-0.5} A_i^{0.5} \\
&= \text{diag}\left(\frac{\phi V(\mu_{ij})}{w_{ij}}\right)
\end{aligned}
$$

as before.

Several comments

Consider equation (6),

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{k} \mathbf{D}_j' \mathbf{V}_j^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_j) = \mathbf{0}$$

$\mathbf{D}_j = \mathbf{D}_j(\boldsymbol{\beta}), \mathbf{V}_j = \mathbf{V}_j(\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi), \boldsymbol{\mu}_j = \boldsymbol{\mu}_j(\boldsymbol{\beta})$

- for normal response and $\mathbf{V}_i = \sigma^2 \mathbf{R}_i(\boldsymbol{\alpha})$, the GEE of $\boldsymbol{\beta}$ is MLE when σ^2 and $\boldsymbol{\alpha}$ are estimated by ML (and REML if σ^2 and $\boldsymbol{\alpha}$ are estimated by REML)
- GEE estimators are MLEs when $\mathbf{R}_i(\boldsymbol{\alpha}) = \mathbf{I}_{n_i}$ and responses are from an EF distribution
- Scale parameter ϕ is optional (not required for some models)

Choice of $\mathbf{R}_i(\alpha)$

- Independent $\mathbf{R}_i(\boldsymbol{\alpha}) = \mathbf{I}_{n_i}$
- Exchangeable

$$\mathbf{R}_i(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & \alpha & \alpha & \alpha \\ \alpha & 1 & \alpha & \alpha \\ \alpha & \alpha & \ddots & \alpha \\ \cdots & \cdots & \cdots & 1 \end{bmatrix}$$

- Unstructured, $corr(y_{ij}, y_{ik}) = \alpha_{jk}$
- AR(1), $corr(y_{ij}, y_{ik}) = \alpha^{|k-j|}$

Estimators

Need to estimate $\boldsymbol{\alpha}$ to compute the GEE, using exchangeable $\mathbf{R}_i(\boldsymbol{\alpha})$ as an example.

$\mathbf{y}_i = (y_{i1}, y_{i2}, \cdots, y_{in_i})'$, $corr(\mathbf{y}_i) = \mathbf{R}_i(\boldsymbol{\alpha})$, $\mathbf{R}_i(\boldsymbol{\alpha})$ is exchangeable.

- scaled Pearson residuals

$$\Gamma_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{var(y_{ij})}} = \frac{y_{ij} - \mu_{ij}}{\sqrt{\phi V(\mu_{ij})/w_{ij}}}$$

- $\alpha = \rho_{jk} = corr(y_{ij}, y_{ik}) = E(\Gamma_{ij}, \Gamma_{ik})$, $j \neq k$
  ——if the parameters in Γ_{ij} were known, then each pair Γ_{ij}, Γ_{ik} would be an unbiased estimator of α.
  —— There are $n_i(n_i - 1)/2$ such pairs for individual i. Adding up all pairs over all subjects and divided by the total number of pairs $N^* = 0.5\sum_{i=1}^m n_i(n_i - 1)$ gives an estimate of α
- $\hat{\alpha} = \frac{1}{N^*} \sum_{i=1}^{m} \sum_{j,k=1}^{n_i} \Gamma_{ij}\Gamma_{ik}$

- SAS unscaled residuals

$$e_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{V(\mu_{ij})/w_{ij}}}$$

- $\Gamma_{ij} = e_{ij}/\sqrt{\phi}$, so

$$\hat{\alpha} = \frac{1}{N^*\phi} \sum_{i=1}^{m} \sum_{j<k} e_{ij} e_{ik} \tag{7}$$

Note that

  - $\hat{\alpha}$ depends on ϕ and $\boldsymbol{\beta}$ (through μ_{ij}), $\hat{\alpha}(\phi, \boldsymbol{\beta})$
  - $E(e_{ij}^2) = var(e_{ij}) = \phi$, $var(\Gamma_{ij}) = 1$ so $var(e_{ij}) = \phi$

$$\hat{\phi} = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n_i} e_{ij}^2, N = \sum_{i=1}^{m} n_i \tag{8}$$

e_{ij} depends on $\boldsymbol{\beta}$, $\hat{\phi}$ is a function of $\boldsymbol{\beta}$, say $\hat{\phi}(\boldsymbol{\beta})$.

Fitting algorithm

1. Get an initial estimate of $\boldsymbol{\beta}$, say $\hat{\boldsymbol{\beta}}_0$ by solving equation (6), assuming independence structure (i.e., $\mathbf{R}_i(\boldsymbol{\alpha}) = \mathbf{I}_{n_i}$) doesn't require estimates of ϕ or $\boldsymbol{\alpha}$.
2. Compute $\hat{\mu}_{ij} = \mu_{ij}(\hat{\boldsymbol{\beta}}_0)$ and $\hat{e}_{ij} = \dfrac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{V(\hat{\mu}_{ij})/w_{ij}}}$ and plug into (8) to get $\hat{\phi}_0 = \hat{\phi}(\hat{\beta}_i)$ and into (7) with $\hat{\phi}_0$ to get $\hat{\alpha}_0 = \hat{\alpha}(\hat{\phi}_0, \hat{\beta}_0)$
3. Compute $V_i = V_i(\hat{\alpha}_0, \hat{\beta}_0, \hat{\phi}_0)$ and $\mathbf{D}_i = \mathbf{D}_i(\hat{\boldsymbol{\beta}}_0)$ and plug into (6). Solve GEE score function (6) for $\hat{\boldsymbol{\beta}}$, say $\hat{\boldsymbol{\beta}}^{(1)}$
4. Repeat 2 and 3 with $\hat{\boldsymbol{\beta}}^{(1)}$, iterate until (hopefully) convergence
5. Final estimates $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}$ and $\hat{\phi}$

Same idea applied to other correlation structures.

The working correlation structure

It is difficult to propose a suitable correlation structure $\mathbf{R}_i(\alpha)$ for non-normal longitudinal data. GEE approach has some good properties to deal with the problem

- β is consistently estimated even if $\mathbf{R}_i(\alpha)$ is uncorrectly specified
- Standard errors and inference procedure are available to account for $\mathbf{R}_i(\alpha)$ possibly being misspecified
- In standard ML based inference for normally distributed repeated measures, the ML estimate of β is also consistent even $\mathbf{R}_i(\alpha)$ is misspecified, but the ML-based standard errors and tests are not robust to misspecification of $\mathbf{R}_i(\alpha)$. This makes GEE-based inference attractive even for normal response.
- $\mathbf{R}_i(\alpha)$ is not necessarily taken seriously, it is called "working correlation structure". But estimators with less variability will be produced using $\mathbf{R}_i(\alpha)$ that mimic the "true" correlation.

Distribution of GEE estimators $\hat{\beta}$

Under suitable regularity conditions, $\hat{\beta}$ is approximately normally distributed

$$\hat{\beta} \sim N(\beta, \mathbf{\Sigma})$$

given and estimator $\hat{\mathbf{\Sigma}}$ of $\mathbf{\Sigma}$, we can construct CI, tests and inferences.