STAT 574 Linear and Nonlinear Mixed Models

Lecture 6: Nonlinear Marginal Model

Chencheng Cai

Washington State University

Marginal v.s. Conditional Model

Suppose b_i is a random effect.

- **Conditional Model**: we specify $p(Y_{ij} \mid b_i, \boldsymbol{\theta})$
- **Marginal Model**: we specify $\mathbb{E}(Y_{ij} \mid \boldsymbol{\theta})$ and $\mathrm{Var}(Y_{ij} \mid \boldsymbol{\theta})$

Is LME a marginal or conditional model?

Marginal v.s. Conditional Model

Suppose b_i is a random effect.

- **Conditional Model**: we specify $p(Y_{ij} \mid b_i, \boldsymbol{\theta})$
- **Marginal Model**: we specify $\mathbb{E}(Y_{ij} \mid \boldsymbol{\theta})$ and $\mathrm{Var}(Y_{ij} \mid \boldsymbol{\theta})$

Is LME a marginal or conditional model? BOTH!

Marginal v.s. Conditional Model

When we consider nonlinear models, the two models are usually different.

- **Conditional Model** involves an arbitrary conditional distribution for $Y_{ij} \mid b_i, \boldsymbol{\theta}$ — may include interaction terms between b_i and $\boldsymbol{\theta}$.
- **Marginal Model** assumes the following form:

$$Y_{ij} = \mathbb{E}[Y_{ij} \mid \boldsymbol{\theta}] + \eta_{ij},$$

with η_{ij} as the deviation of Y_{ij} from its mean, which follows a specified covariance structure.

Nonlinear Marginal Model

We generalize the linear mixed effects model to a nonlinear mixed model in which random effects **enter the model in a linear fashion**. *This type of mixed model will be called* **marginal**.

Nonlinear Marginal Model:

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\theta}) + \boldsymbol{\eta}_i,$$

with $\boldsymbol{f}_i$ a nonlinear function and $\mathrm{Cov}(\boldsymbol{\eta}_i) = \boldsymbol{V}_i(\boldsymbol{\theta})$.

Why the above model is a marginal model?

Nonlinear Marginal Model

We generalize the linear mixed effects model to a nonlinear mixed model in which random effects **enter the model in a linear fashion**. *This type of mixed model will be called* **marginal**.

Nonlinear Marginal Model:

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\theta}) + \boldsymbol{\eta}_i,$$

with $\boldsymbol{f}_i$ a nonlinear function and $\mathrm{Cov}(\boldsymbol{\eta}_i) = \boldsymbol{V}_i(\boldsymbol{\theta})$.

Why the above model is a marginal model?

- $\boldsymbol{f}_i(\boldsymbol{\theta})$ specifies the marginal mean.
- $\boldsymbol{V}_i(\boldsymbol{\theta})$ specifies the marginal covariance.

Sometimes, we further specify $\boldsymbol{\eta}_i$ to obtain the desired covariance structure.

Fixed Matrix of Random Effects

Consider the following nonlinear model with additive random-effects

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\beta}) + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i,$$

with

$$\mathbb{E}(\boldsymbol{b}_i) = 0, \quad \mathbb{E}(\boldsymbol{\epsilon}_i) = 0, \quad \mathrm{Cov}(\boldsymbol{b}_i) = \sigma^2\boldsymbol{D}, \quad \mathrm{Cov}(\boldsymbol{\epsilon}_i) = \sigma^2\boldsymbol{I}.$$

Fixed Matrix of Random Effects

Consider the following nonlinear model with additive random-effects

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\beta}) + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i,$$

with

$$\mathbb{E}(\boldsymbol{b}_i) = 0, \quad \mathbb{E}(\boldsymbol{\epsilon}_i) = 0, \quad \mathrm{Cov}(\boldsymbol{b}_i) = \sigma^2 \boldsymbol{D}, \quad \mathrm{Cov}(\boldsymbol{\epsilon}_i) = \sigma^2 \boldsymbol{I}.$$

- Is this a conditional or marginal model?

Fixed Matrix of Random Effects

Consider the following nonlinear model with additive random-effects

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\beta}) + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i,$$

with

$$\mathbb{E}(\boldsymbol{b}_i) = 0, \quad \mathbb{E}(\boldsymbol{\epsilon}_i) = 0, \quad \mathrm{Cov}(\boldsymbol{b}_i) = \sigma^2 \boldsymbol{D}, \quad \mathrm{Cov}(\boldsymbol{\epsilon}_i) = \sigma^2 \boldsymbol{I}.$$

- Is this a conditional or marginal model?
  - Marginal, because it specifies the mean and variance of $\boldsymbol{y}_i$.
  - Not conditional, because the true distribution $\boldsymbol{y}_i \mid \boldsymbol{b}_i$ may be beyond linear.

Fixed Matrix of Random Effects

Consider the following nonlinear model with additive random-effects

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\beta}) + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i,$$

with

$$\mathbb{E}(\boldsymbol{b}_i) = 0, \quad \mathbb{E}(\boldsymbol{\epsilon}_i) = 0, \quad \mathrm{Cov}(\boldsymbol{b}_i) = \sigma^2 \boldsymbol{D}, \quad \mathrm{Cov}(\boldsymbol{\epsilon}_i) = \sigma^2 \boldsymbol{I}.$$

- Is this a conditional or marginal model?
  - Marginal, because it specifies the mean and variance of $\boldsymbol{y}_i$.
  - Not conditional, because the true distribution $\boldsymbol{y}_i \mid \boldsymbol{b}_i$ may be beyond linear.
- Do we need other conditions?

Fixed Matrix of Random Effects

Consider the following nonlinear model with additive random-effects

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\beta}) + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i,$$

with

$$\mathbb{E}(\boldsymbol{b}_i) = 0, \quad \mathbb{E}(\boldsymbol{\epsilon}_i) = 0, \quad \mathrm{Cov}(\boldsymbol{b}_i) = \sigma^2 \boldsymbol{D}, \quad \mathrm{Cov}(\boldsymbol{\epsilon}_i) = \sigma^2 \boldsymbol{I}.$$

- Is this a conditional or marginal model?
  - Marginal, because it specifies the mean and variance of $\boldsymbol{y}_i$.
  - Not conditional, because the true distribution $\boldsymbol{y}_i \mid \boldsymbol{b}_i$ may be beyond linear.
- Do we need other conditions?
  - $\boldsymbol{f}_i(\boldsymbol{\beta})$'s have to be identifiable. That is, $\boldsymbol{f}_i(\boldsymbol{\beta}_1) = \boldsymbol{f}_i(\boldsymbol{\beta}_2)$ for all i implies $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.
  - It is called **mean identifiability**.

Fixed Matrix of Random Effects — Estimate β

Nonlinear Least Squares (NLS):

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \|\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta})\|^2$$

The corresponding estimation equations:

$$\sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{f}_i}{\partial \boldsymbol{\beta}} \right)^T (\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta})) = \boldsymbol{0}$$

The minimizer $\hat{\beta}^{(NLS)}$ (an M-estimator) is consistent because

$$\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{f}_i}{\partial \boldsymbol{\beta}} \right)^T (\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta})) = \boldsymbol{0}.$$

Fixed Matrix of Random Effects — Estimate $\boldsymbol{\beta}$

A more efficient way (assuming $\boldsymbol{D}$ is known) is (weighted NLS)

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} [\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta})]^T (\boldsymbol{I} + \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^T)^{-1} [\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta})]$$

The corresponding estimation equations are

$$\sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{f}_i}{\partial \boldsymbol{\beta}} \right)^T (\boldsymbol{I} + \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^T)^{-1} [\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta})] = \boldsymbol{0}$$

Fixed Matrix of Random Effects — MLE

If we assume $\boldsymbol{b}_i$'s and $\boldsymbol{\epsilon}_i$'s are normal, that is

$$\boldsymbol{y}_i \sim \mathcal{N}(\boldsymbol{f}_i(\boldsymbol{\beta}), \sigma^2 \boldsymbol{V}_i)$$

The log-likelihood function is

$$\ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{D}) = -\frac{1}{2}\left\{ N_t \log \sigma^2 + \sum_{i=1}^{N} \left[\log |\boldsymbol{V}_i| + \sigma^{-2}(\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}))^T \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}))\right] \right\}$$

Fixed Matrix of Random Effects — Distribution-free Estimation

- Pooled Variance estimator
  - Minimization:

$$S_{min} = \min_{\boldsymbol{\beta}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_N} \sum_{i=1}^{N} \|\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}) - \boldsymbol{Z}_i \boldsymbol{\gamma}_i\|^2$$

  - Estimation of variance: $\hat{\sigma}^2 = S_{min}/d.f.$.
- Method of Moments
  - Calculate $\hat{\boldsymbol{b}}_i = (\boldsymbol{Z}_i^T \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i^T (\boldsymbol{y}_i - \boldsymbol{f}_i(\hat{\boldsymbol{\beta}}^{(NLS)}))$.
  - Estimate

$$\widehat{\sigma^2 \boldsymbol{D}} = \frac{1}{N} \sum_{i=1}^{N} \hat{\boldsymbol{b}}_i \hat{\boldsymbol{b}}_i^T - \hat{\sigma}^2 \sum_{i=1}^{N} (\boldsymbol{Z}_i^T \boldsymbol{Z}_i)^{-1}$$

- Variance Least Squares (VLS)
  - Get $\hat{\boldsymbol{e}}_i = \boldsymbol{y}_i - \boldsymbol{f}_i(\hat{\boldsymbol{\beta}}^{(NLS)})$.
  - minimize

$$\min_{\sigma^2, \boldsymbol{D}} \sum_{i=1}^{N} \|\hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_i^T - \sigma^2 \boldsymbol{V}_i\|_F^2$$

Fixed Matrix of Random Effects — Asymptotic Results

Nonlinear marginal models are similar to LME models:

- NLS: consistency and asymptotic normality
- weighted NLS: consistency and asymptotic normality
- Use MM or VLS estimators for $\boldsymbol{D}$ in weighted NLS converges in distribution to the one using true $\boldsymbol{D}$.

If both $\boldsymbol{b}_i$ and $\boldsymbol{\epsilon}_i$ are normal,

- The MLEs for $\boldsymbol{\beta}$ and variance parameters are independent.
- Use of MM or VLS in weighted NLS leads to $\boldsymbol{\beta}$ estimates asymptotically equivalent to the MLE.

Varied Matrix of Random Effects

A generalization is that the design matrix of the random effect is also related to $\boldsymbol{\beta}$.

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\beta}) + \boldsymbol{Z}_i(\boldsymbol{\beta})\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, N$$

Because the dependence of $\boldsymbol{Z}_i$ on $\boldsymbol{\beta}$,

- Weighted NLS has asymptotic normality.
- Weighted NLS is not efficient.
- Need to consider MLE.

Varied Matrix of Random Effects — MLE

Log-likelihood function:

$$\ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{D}) = -\frac{1}{2}\left\{N \log \sigma^2 + \sum_{i=1}^{N}\left[\log |\boldsymbol{V}_i(\boldsymbol{\beta})| + \frac{1}{\sigma^2}\boldsymbol{e}_i^T(\boldsymbol{\beta})\boldsymbol{V}_i^{-1}(\boldsymbol{\beta})\boldsymbol{e}_i(\boldsymbol{\beta})\right]\right\},$$

where

$$\boldsymbol{e}_i(\boldsymbol{\beta}) = \boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}), \quad \boldsymbol{V}_i(\boldsymbol{\beta}) = \boldsymbol{I} + \boldsymbol{Z}_i(\boldsymbol{\beta})\boldsymbol{D}\boldsymbol{Z}_i^T(\boldsymbol{\beta}).$$

- Difficulty: both $\boldsymbol{V}_i$ and $\boldsymbol{e}_i$ are $\boldsymbol{\beta}$-dependent — more complicated gradient.

Varied Matrix of Random Effects — An iterative algorithm

Notice that, if $\boldsymbol{V}_i$'s are given, the MLE of $\boldsymbol{\beta}$ is the weighted NLS solution:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \boldsymbol{e}_i^T(\boldsymbol{\beta})\boldsymbol{V}_i^{-1}\boldsymbol{e}_i(\boldsymbol{\beta})$$

An easy estimate for $\boldsymbol{V}_i$ is $\boldsymbol{V}_i(\hat{\boldsymbol{\beta}})$ for some estimated $\boldsymbol{\beta}$. Therefore, we have the following iterative algorithm. **Iteratively reweighted least squares (IRLS)**

1. Start from $\hat{\boldsymbol{\beta}}^{(0)}$ (estimated from NLS with $\boldsymbol{D} = \boldsymbol{0}$)
2. For iteration 1 to K,
   - update $\hat{\boldsymbol{D}}$ and $\hat{\sigma}^2$ by maximizing conditional likelihood.
   - update

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \boldsymbol{e}_i^T(\boldsymbol{\beta})\boldsymbol{V}_i^{-1}(\hat{\boldsymbol{\beta}}^{(k)})\boldsymbol{e}_i(\boldsymbol{\beta})$$

Varied Matrix of Random Effects — An iterative algorithm

From the update equation

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \boldsymbol{e}_i^T(\boldsymbol{\beta})\boldsymbol{V}_i^{-1}(\hat{\boldsymbol{\beta}}^{(k)})\boldsymbol{e}_i(\boldsymbol{\beta})$$

we know

$$\sum_{i=1}^{N} \boldsymbol{F}_i^T(\hat{\boldsymbol{\beta}}^{(k+1)})\boldsymbol{V}_i^{-1}(\hat{\boldsymbol{\beta}}^{(k)})\boldsymbol{e}_i(\hat{\boldsymbol{\beta}}^{(k+1)}) = \mathbf{0},$$

where $\boldsymbol{F}_i$ is the gradient of $\boldsymbol{f}_i$.

- At convergence, the solution $\hat{\boldsymbol{\beta}}$ is a fixed point of the first equation.
- The converged value $\hat{\boldsymbol{\beta}} = \lim_{k\to\infty} \hat{\boldsymbol{\beta}}^{(k)}$ satisfy

$$\sum_{i=1}^{N} \boldsymbol{F}_i^T(\hat{\boldsymbol{\beta}})\boldsymbol{V}_i^{-1}(\hat{\boldsymbol{\beta}})\boldsymbol{e}_i(\hat{\boldsymbol{\beta}}) = \mathbf{0}$$

Varied Matrix of Random Effects — GEE

From our previous arguments, the iterative algorithm is to find the solution to

$$\sum_{i=1}^{N} \boldsymbol{F}_i^T(\boldsymbol{\beta}) \left(\boldsymbol{I} + \boldsymbol{Z}_i(\boldsymbol{\beta})\boldsymbol{D}\boldsymbol{Z}_i^T(\boldsymbol{\beta})\right)^{-1} (\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta})) = \mathbf{0}$$

The above equations are called **generalized estimating equations (GEE)**.

- The estimator is the solution to the GEE.
- The GEE are NOT gradients of the log-likelihood function.
- Need to have an estimate $\hat{\boldsymbol{D}}$, which is usually obtained by
  - distribution-free variance estimator
  - solving the score equation of the log-likelihood function.

Varied Matrix of Random Effects — GEE Properties

- GEE is consistent and asymptotically normally distributed.
  - Because GEE is a Z-estimator (solution to $\boldsymbol{\Psi}(\boldsymbol{\theta}) = \mathbf{0}$)
  - Consistency under consistent estimator $\hat{\boldsymbol{D}}$ is ensured by the generalized Slutsky's theorem.
  - Asymptotic normality is ensured by the LLN and delta method.
- GEE is less efficient than MLE.

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{\mathrm{IRLS}}) = \sigma^2 \left(\sum_{i=1}^{N} \boldsymbol{F}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{F}_i \right)^{-1}$$

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{\mathrm{ML}}) = \sigma^2 \left(\sum_{i=1}^{N} \boldsymbol{F}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{F}_i + \sigma^2 \sum_{i=1}^{N} \boldsymbol{T}_i \right)^{-1}$$

for some positive semi-definite matrix $\boldsymbol{T}_i$'s.

Varied Matrix of Random Effects — GEE Properties

- GEE and MLE have similar efficiency when
  - σ^2 is small.
  - Or $\boldsymbol{T}_i$'s are small — when $\boldsymbol{Z}_i(\boldsymbol{\beta})$ is quite linear in $\boldsymbol{\beta}$.
- When the distribution is misspecified, both IRLS and MLE give consistent estimators.
- When the covariance matrix is misspecified, only IRLS gives a consistent estimator and asymptotic normality.

Three Types of Nonlinear Marginal Models

- Type I nonlinear marginal model:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\beta}) + \boldsymbol{\eta}, \quad \mathrm{Cov}(\boldsymbol{\eta}) = \boldsymbol{V}(\boldsymbol{\gamma})$$

- Type II nonlinear marginal model:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\beta}) + \boldsymbol{\eta}, \quad \mathrm{Cov}(\boldsymbol{\eta}) = \boldsymbol{V}(\boldsymbol{\beta}, \boldsymbol{\gamma})$$

- Type III nonlinear marginal model:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\theta}) + \boldsymbol{\eta}, \quad \mathrm{Cov}(\boldsymbol{\eta}) = \boldsymbol{V}(\boldsymbol{\theta})$$

- identifiability: same mean and variance imply same parameters.
- mean-identifiability: same mean implies same parameters.

Total Generalized Estimating Equations

- For Type I models, IRLS is the same as MLE.
- For Type II models, IRLS is easy to implement but less efficient than MLE.
- For Type III models, IRLS does not work because the mean and variance share the same parameter.

Total Generalized Estimating Equations

Assume the following Type III model:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\theta}) + \boldsymbol{\eta}, \quad \mathrm{Cov}(\boldsymbol{\eta}) = \boldsymbol{V}(\boldsymbol{\theta})$$

Under normal assumption, we have the log-likelihood function:

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\left\{\log|\boldsymbol{V}| + (\boldsymbol{y} - \boldsymbol{f})^T \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{f})\right\}$$

The estimating equation for MLE is

$$\boldsymbol{F}^T \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{f}) + \frac{1}{2}\boldsymbol{G}^T \left[\left(\boldsymbol{V}^T(\boldsymbol{y} - \boldsymbol{f})\right) \otimes \left(\boldsymbol{V}^T(\boldsymbol{y} - \boldsymbol{f})\right) - \mathrm{vec}(\boldsymbol{V}^{-1})\right] = \boldsymbol{0},$$

where $\boldsymbol{F} = \partial \boldsymbol{f}/\partial\boldsymbol{\theta}$ and $\boldsymbol{G} = \partial\mathrm{vec}(\boldsymbol{V})/\partial\boldsymbol{\theta}$.

Total Generalized Estimating Equations

For general cases (NOT normally distributed),
we consider the following **total GEE (TGEE)**:

$$\boldsymbol{F}^T \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{f}) + \nu \boldsymbol{G}^T \left[\left(\boldsymbol{V}^T (\boldsymbol{y} - \boldsymbol{f}) \right) \otimes \left(\boldsymbol{V}^T (\boldsymbol{y} - \boldsymbol{f}) \right) - \operatorname{vec}(\boldsymbol{V}^{-1}) \right] = \boldsymbol{0}$$

for some $\nu > 0$.

Total Generalized Estimating Equations

- $\hat{\boldsymbol{\theta}}_{TGEE}$ is consistent and asymptotically normal for any $\nu > 0$.
- Let $\boldsymbol{\xi} = \boldsymbol{V}^{-1/2}(\boldsymbol{y} - \boldsymbol{f})$ and $\mathrm{Var}(\xi_i) = \kappa - 1$. ($\kappa$ the kurtosis)
- If $\mathbb{E}[\xi_i^3] = 0$,

$$\mathrm{Cov}\left(\hat{\boldsymbol{\theta}}_{TGEE}\right) = (\boldsymbol{P} + \nu\boldsymbol{Q})^{-1}\left(\boldsymbol{P} + \nu^2(\kappa - 1)\boldsymbol{Q}\right)(\boldsymbol{P} + \nu\boldsymbol{Q})^{-1},$$

  where

$$\boldsymbol{P} = \boldsymbol{F}^T\boldsymbol{V}^{-1}\boldsymbol{F}, \quad \boldsymbol{Q} = \boldsymbol{G}^T(\boldsymbol{V}^{-1} \otimes \boldsymbol{V}^{-1})\boldsymbol{G}.$$

- Minimum variance attained at $\nu = 1/(\kappa - 1)$. (See Table 6.3 in textbook)
- For MLE under normal assumption,

$$\mathrm{Cov}\left(\hat{\boldsymbol{\theta}}_{ML}\right) = \left(\boldsymbol{P} + \frac{1}{2}\boldsymbol{Q}\right)^{-1}$$

Total Generalized Estimating Equations

For mixed-effect models

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\theta}) + \boldsymbol{\eta}_i, \quad \mathrm{Cov}(\boldsymbol{\eta}_i) = \boldsymbol{V}_i(\boldsymbol{\theta}),$$

the TGEE is the solution to

$$\sum_{i=1}^{N} \left\{ \boldsymbol{F}_i^T \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{f}_i) + \nu \boldsymbol{G}_i^T \left[\left(\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{f}_i) \right) \otimes \left(\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{f}_i) \right) - \mathrm{vec}(\boldsymbol{V}_i^{-1}) \right] \right\} = \boldsymbol{0}$$

The components in the variance are

$$\boldsymbol{P} = \sum_{i=1}^{N} \boldsymbol{F}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{F}_i, \quad \boldsymbol{Q} = \sum_{i=1}^{N} \boldsymbol{G}_i^T (\boldsymbol{V}_i^{-1} \otimes \boldsymbol{V}_i^{-1}) \boldsymbol{G}_i.$$

Expected Newton-Raphson Algorithm for Total GEE

- To solve the TGEE in the form of $\boldsymbol{\Psi}(\boldsymbol{\theta}) = \mathbf{0}$, the Newton-Raphson suggests

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - \left(\frac{\partial \boldsymbol{\Psi}}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}^{(k)}} \right)^{-1} \boldsymbol{\Psi}(\boldsymbol{\theta}^{(k)})$$

- The expected Newton-Raphson (ENR) algorithm replaces the Jacobian with its expectation and set

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \left(\boldsymbol{P}(\boldsymbol{\theta}^{(k)}) + \nu \boldsymbol{Q}(\boldsymbol{\theta}^{(k)}) \right)^{-1} \boldsymbol{\Psi}(\boldsymbol{\theta}^{(k)})$$

Example: LME model

Consider the LME model:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i,$$

with $\mathrm{Cov}(\boldsymbol{b}_i) = \sigma^2\boldsymbol{D}$ and $\mathrm{Cov}(\boldsymbol{\epsilon}_i) = \sigma^2\boldsymbol{I}$.

- The parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \mathrm{vec}(\boldsymbol{D})^T)$.
- Using notations for the marginal models, we have

$$\boldsymbol{F}_i = [\boldsymbol{X}_i, \boldsymbol{0}, \boldsymbol{0}], \quad \boldsymbol{G}_i = [\boldsymbol{0}, \mathrm{vec}(\boldsymbol{I}), \boldsymbol{Z}_i \otimes \boldsymbol{Z}_i].$$

- The TGEEs are:

$$\sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) = \boldsymbol{0}$$

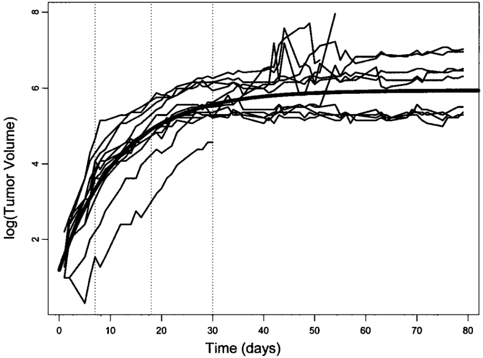
$$\sum_{i=1}^{N} (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^T \boldsymbol{V}_i^{-2}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) - \mathrm{tr}(\boldsymbol{V}_i^{-1}) = 0$$

$$\sum_{i=1}^{N} \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^T \boldsymbol{V}_i^{-1}\boldsymbol{Z}_i - \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1}\boldsymbol{Z}_i = \boldsymbol{0}$$

Example — Log-Gompertz Curve

Consider a Gompertz Growth Curve model:

$$Y(t) = Ae^{-e^{b-ct}}$$



Example — Log-Gompertz Curve

By re-parametrization, we have the following model

$$y = \beta_1 - \beta_2 e^{-\beta_3 t} + \epsilon.$$

By considering random intercept and adding indices, we have

$$y_{ij} = (\beta_1 + b_i) - \beta_2 e^{-\beta_3 t_{ij}} + \epsilon_{ij}$$

with $\mathbb{E}[\epsilon_{ij}] = \mathbb{E}[b_i] = 0$ and $\mathrm{Var}(\epsilon_{ij}) = \sigma^2$, $\mathrm{Var}(b_i) = \sigma^2 d$.

Example — Log-Gompertz Curve

$$y_{ij} = (\beta_1 + b_i) - \beta_2 e^{-\beta_3 t_{ij}} + \epsilon_{ij}$$

As a nonlinear marginal model, we have

- $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$.
- $\boldsymbol{f}_i = \beta_1 - \beta_2 e^{-\beta_3 \boldsymbol{t}_i}$.
- $\mathrm{Cov}(\boldsymbol{y}_i) = \sigma^2 \boldsymbol{I} + \sigma^2 d \mathbf{1}\mathbf{1}^T$.

Example — Log-Gompertz Curve — NLS

```
data = read.csv("./Data/MixedModels/Chapter06/TUMspher.txt")

fit.nls = nls(lntumvol ~ a1-a2*exp(-a3*day), data=data,
              start=list(a1=5, a2=1, a3=0.1))

print(fit.nls)
```

```
Nonlinear regression model
    model: lntumvol ~ a1 - a2 * exp(-a3 * day)
        data: data
            a1      a2      a3
        6.03797 4.83443 0.08211
    residual sum-of-squares: 259.8

    Number of iterations to convergence: 5
    Achieved convergence tolerance: 7.034e-07
```

Example — Log-Gompertz Curve — NLME

```
library(nlme)
fit.nlme = nlme(lntumvol ~ a1-a2*exp(-a3*day),
                fixed = a1+a2+a3~1,
                random = a1~1|id,
                data=data,
                start=c(5,1,0.1))
print(fit.nlme)
```

```
Nonlinear mixed-effects model fit by maximum likelihood
    Model: lntumvol ~ a1 - a2 * exp(-a3 * day)
    Data: data
    Log-likelihood: -328.624
    Fixed: a1 + a2 + a3 ~ 1
            a1          a2          a3
    5.93781796 4.74846672 0.08549033

Random effects:
Formula: a1 ~ 1 | id
            a1  Residual
StdDev: 0.6293778 0.4023551

Number of Observations: 590
Number of Groups: 12
```

Example — Log-Gompertz Curve — TGEE

- Since $\boldsymbol{f}_i = \beta_1 - \beta_2 e^{-\beta_3 \boldsymbol{t}_i}$, we have

$$\boldsymbol{F}_{i,\beta} = (\mathbf{1}, -e^{-\beta_3 \boldsymbol{t}_i}, \beta_2 \boldsymbol{t}_i \odot e^{-\beta_3 \boldsymbol{t}_i})$$

- Since $\boldsymbol{V}_i = \sigma^2(\boldsymbol{I} + d\mathbf{1}\mathbf{1}^T)$,

$$\boldsymbol{G}_{i,\beta} = \mathbf{0}, \quad \boldsymbol{G}_{i,\sigma^2} = \mathrm{vec}(\boldsymbol{I}) + d\mathbf{1}, \quad \boldsymbol{G}_{i,d} = \sigma^2\mathbf{1}.$$

Example — Log-Gompertz Curve — TGEE

- Since $\boldsymbol{f}_i = \beta_1 - \beta_2 e^{-\beta_3 \boldsymbol{t}_i}$, we have

$$\boldsymbol{F}_{i,\beta} = (\mathbf{1}, -e^{-\beta_3 \boldsymbol{t}_i}, \beta_2 \boldsymbol{t}_i \odot e^{-\beta_3 \boldsymbol{t}_i})$$

- Since $\boldsymbol{V}_i = \sigma^2(\boldsymbol{I} + d\mathbf{1}\mathbf{1}^T)$,

$$\boldsymbol{G}_{i,\beta} = \mathbf{0}, \quad \boldsymbol{G}_{i,\sigma^2} = \mathrm{vec}(\boldsymbol{I}) + d\mathbf{1}, \quad \boldsymbol{G}_{i,d} = \sigma^2 \mathbf{1}.$$

Example — Log-Gompertz Curve — TGEE

```
step <- function(a1, a2, a3, sigma, d, nu){
P = 0
Q = 0
TGEE = rep(0, 5)
for(i in 1:12){
     sub = data[data$id==i,]
     ni = nrow(sub)
     Vi = sigma^2 * (diag(ni) + d)
     Vi_inv = (diag(ni) - 1/(ni+1/d))/sigma^2
     eps = sub$lntumvol - a1 + a2 * exp(-a3 * sub$day)
     Fi = cbind(1, -exp(-a3 * sub$day), a2*sub$day*exp(-a3*sub$day))
     P = P + t(Fi)%*%Vi_inv%*%Fi
     Gi_sig = c((diag(ni) + d))
     Gi = cbind(Gi_sig, sigma^2)
     Q = Q + t(Gi)%*%kronecker(Vi_inv, Vi_inv)%*%Gi
     TGEE[1:3] = TGEE[1:3] + t(Fi)%*%Vi_inv%*%eps
     veps = Vi_inv%*%eps
     TGEE[4:5] = TGEE[4:5] + nu* t(Gi)%*%(kronecker(veps, veps)-c(Vi_inv))}
return(solve(bdiag(P, nu*Q),TGEE))
}
```

Example — Log-Gompertz Curve — TGEE

```r
theta = c(5, 1, 0.1, 1, 1)
while(T){
    delta = step(theta[1], theta[2], theta[3], theta[4], theta[5],
        0.5)
    if(norm(delta) <= 0.0001) break
    else theta = theta + delta
}

print(theta)
```

The TGEE estimators are:

$$\hat{\beta}_1 = 5.9378,\ \hat{\beta}_2 = 4.7485,\ \hat{\beta}_3 = 0.0855,\ \hat{\sigma}^2 = 0.1618,\ \hat{d} = 2.4468$$