STAT 574 Linear and Nonlinear Mixed Models

Lecture 7: Generalized Linear Mixed Models

Chencheng Cai

Washington State University

Regression Models for Binary Data

- If the outcome $\boldsymbol{y}$ is binary, we often model that with

$$\mathbb{P}[y_i = 1] = \mu(\boldsymbol{\beta}^T \boldsymbol{x}_i)$$

- $\boldsymbol{x}_i$: covariate vector for unit i.
- y_i is Bernoulli with parameter $\mu(\boldsymbol{\beta}^T \boldsymbol{x}_i)$.
- μ is the **inverse link** function.
- Equivalently,

$$\eta(\mathbb{P}[y_i = 1]) = \boldsymbol{\beta}^T \boldsymbol{x}_i,$$

  where η is the **link** function.
- Mean and variance:

$$\mathbb{E}[y_i] = \mu(\boldsymbol{\beta}^T \boldsymbol{x}_i), \quad \mathrm{Var}(y_i) = \mu(\boldsymbol{\beta}^T \boldsymbol{x}_i)(1 - \mu(\boldsymbol{\beta}^T \boldsymbol{x}_i)).$$

Regression Models for Binary Data

Properties for the inverse link function μ.

1. The function $\mu(s)$ is defined for all $s \in (-\infty, \infty)$.
2. $0 < \mu(s) < 1$, $\lim_{x \to -\infty} \mu(s) = 0$, and $\lim_{x \to \infty} \mu(s) = 1$
3. $d\mu(s)/ds = \mu' > 0$.
4. $d^2 \log \mu(s)/ds^2 < 0$
5. Symmetry: $\mu(s) = 1 - \mu(-s)$.

Interpretations:

- (1) and (2): $\mu(s)$ is a valid parameter for Bernoulli distribution.
- (3): strictly monotonic probability
- (4): concave log-likelihood function
- (5): symmetric under the transformation $y_i \to 1 - y_i$.

Regression Models for Binary Data

Some choices for the inverse link function μ:

- Logistic regression:

$$\mu(s) = \frac{e^s}{1 + e^s}$$

- Probit regression:

$$\mu(s) = \Phi(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s} e^{-t^2/2} dt$$

- Log-log regression:

$$\mu(s) = 1 - e^{-e^s}$$

Regression Models for Binary Data

Some choices for the inverse link function μ:

- Logistic regression: popular in epidemiology, biomedicine, and machine learning.

$$\frac{\mathbb{P}[y_i = 1 \mid \boldsymbol{x}_i]}{\mathbb{P}[y_i = 0 \mid \boldsymbol{x}_i]} = e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}$$

- Probit regression: popular in engineering and econometric studies. Binary response as from a Gaussian outcome with cut-off

$$z_i = \boldsymbol{\beta}^T \boldsymbol{x}_i + \epsilon_i, \quad y_i = \mathbf{1}\{z_i > c\}$$

- Log-log regression: related to Poisson distribution.

$$z_i \sim \text{Poisson}(e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}), \quad y_i = \mathbf{1}\{z_i \geq 1\}$$

Regression Models for Binary Data

Logistic and probit functions are similar:

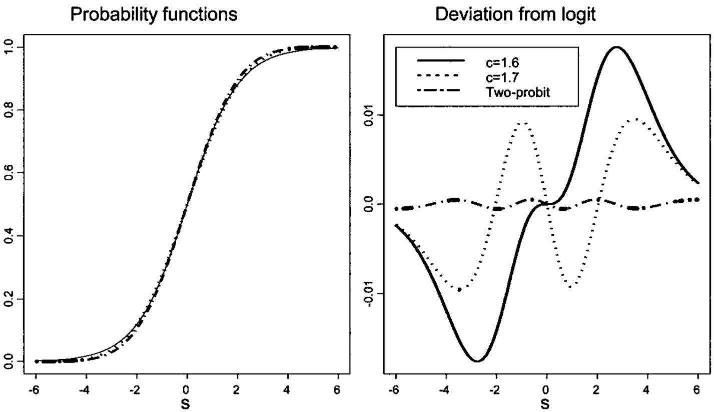

FIGURE 7.2. Logit approximation by probit(s). One-probit approximation (7.9) uses $c = 1.6$ and $c = 1.7$. The latter c gives an absolute error of approximation of less than 0.01. The two-probit approximation (7.11) yields an absolute error of 0.000526.

Regression Models for Binary Data

The model:

$$\mathbb{P}[y_i = 1] = \mu(\boldsymbol{\beta}^T \boldsymbol{x}_i)$$

The likelihood function is

$$L(\boldsymbol{\beta}) = \prod_{i:y_i=1} \mu(\boldsymbol{\beta}^T \boldsymbol{x}_i) \prod_{i:y_i=0} [1 - \mu(\boldsymbol{\beta}^T \boldsymbol{x}_i)]$$

The log-likelihood is

$$\begin{aligned} \ell(\boldsymbol{\beta}) &= \sum_{i:y_i=1} \log \mu(\boldsymbol{\beta}^T \boldsymbol{x}_i) + \sum_{i:y_i=0} \log(1 - \mu(\boldsymbol{\beta}^T \boldsymbol{x}_i)) \\ &= \sum_{i=1}^{n} \left[y_i \log \mu(\boldsymbol{\beta}^T \boldsymbol{x}_i) + (1 - y_i) \log(1 - \mu(\boldsymbol{\beta}^T \boldsymbol{x}_i)) \right] \end{aligned}$$

Regression Models for Binary Data

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \log \mu(\boldsymbol{\beta}^T \boldsymbol{x}_i) + (1 - y_i) \log(1 - \mu(\boldsymbol{\beta}^T \boldsymbol{x}_i)) \right]$$

- $\ell(\boldsymbol{\beta}) \leq 0$. (proof?)
- score equations:

$$\left(\frac{\partial \ell}{\partial \boldsymbol{\beta}} \right)^T = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\mu_i (1 - \mu_i)} \mu_i' \boldsymbol{x}_i = 0$$

- information matrix:

$$\mathcal{I} = -\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}^2} \right] = \sum_{i=1}^{n} \frac{(\mu_i')^2}{\mu_i (1 - \mu_i)} \boldsymbol{x}_i \boldsymbol{x}_i^T$$

Regression Models for Binary Data

Algorithm to find the MLE:

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \hat{\boldsymbol{\beta}}^{(k)} + \lambda_k \boldsymbol{H}_s^{-1} \left(\frac{\partial \ell}{\partial \boldsymbol{\beta}} \right)^T$$

- Newton-Raphson: $\boldsymbol{H}_s$ is the negative Hessian.
- Fisher-Scoring: $\boldsymbol{H}_s$ is the information matrix.
- step size λ_s can be determined by the unit-step algorithm (see Chapter 7.1.5 in textbook)

Binary model with subject-specific intercept

Now we assume a mixed-effect style model such that the intercept is group-specific.

$$\mathbb{P}[y_{ij} = 1] = \mu(a_i + \boldsymbol{\beta}^T \boldsymbol{x}_{ij})$$

We have two interpretations:

1. Fixed effects model. a_i's are fixed unknown parameters.
   - Regular generalized linear regression with (a lot of) dummy variables.
   - Curse of dimensionality.
2. Random effects model. $a_i = \alpha + u_i$ is random with $\mathbb{E}[u_i] = 0$ and $\mathrm{Var}[u_i] = \sigma^2$.
   - More flexible, less number of parameters.
   - Need more complicated estimation methods.

Binary model with subject-specific intercept

Now we consider

$$\mathbb{P}[y_{ij} = 1] = \mu(a_i + \boldsymbol{\beta}^T \boldsymbol{x}_{ij}), \quad a_i \sim \mathcal{N}(\alpha, \sigma^2)$$

if σ^2 is known, the log-likelihood function is

$$\ell(\alpha, \boldsymbol{\beta}) = -\frac{N \log(2\pi\sigma^2)}{2} + \sum_{i=1}^{N} \ell_i(\alpha, \boldsymbol{\beta})$$

with

$$\ell_i(\alpha, \boldsymbol{\beta}) = \log \int_{-\infty}^{\infty} \exp\left\{\tilde{\ell}_i(\alpha, \boldsymbol{\beta}) - \frac{(a_i - \alpha)^2}{2\sigma^2} da_i\right\}$$

and

$$\tilde{\ell}_i(\alpha, \boldsymbol{\beta}) = \sum_{j=1}^{n_j} \left[y_{ij}\mu(a_i + \boldsymbol{\beta}^T \boldsymbol{x}_i) + (1 - y_{ij})(1 - \mu(a_i + \boldsymbol{\beta}^T \boldsymbol{x}_i))\right]$$

Binary model with subject-specific intercept

- When $\sigma^2 \to \infty$, the random effects model turns into a fixed effects model.
- MLE does not exist if there exists $(a, \boldsymbol{\beta})$ such that

$$a + \boldsymbol{\beta}^T x_{ij} < 0 \text{ when } y_{ij} = 0$$
$$a + \boldsymbol{\beta}^T x_{ij} > 0 \text{ when } y_{ij} = 1$$

Logistic regression with random intercept

Consider a logistic regression with random intercept.

$$\mathbb{P}[y_{ij} = 1 \mid a_i] = \frac{e^{a_i + \boldsymbol{\beta}^T \boldsymbol{x}_{ij}}}{1 + e^{a_i + \boldsymbol{\beta}^T \boldsymbol{x}_{ij}}}$$

with

$$a_i = \alpha + b_i \sim \mathcal{N}(\alpha, \sigma^2)$$

Or, equivalently, by adding the constant variable 1 to $\boldsymbol{x}_{ij}$, we have

$$\mathbb{P}[y_{ij} = 1 \mid b_i] = \frac{e^{b_i + \boldsymbol{\beta}^T \boldsymbol{x}_{ij}}}{1 + e^{b_i + \boldsymbol{\beta}^T \boldsymbol{x}_{ij}}}$$

with

$$b_i \sim \mathcal{N}(0, \sigma^2)$$

Logistic regression with random intercept

$$\mathbb{P}[y_{ij} = 1 \mid b_i] = \frac{e^{b_i + \boldsymbol{\beta}^T \boldsymbol{x}_{ij}}}{1 + e^{b_i + \boldsymbol{\beta}^T \boldsymbol{x}_{ij}}}, \quad b_i \sim \mathcal{N}(0, \sigma^2)$$

Therefore, the marginal probability of $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})$ is

$$\begin{aligned}
\mathbb{P}[\boldsymbol{y}_i] &= \int \mathbb{P}[\boldsymbol{y}_i \mid b_i] \; dP(b_i) = \int \prod_{j=1}^{n_j} \mathbb{P}[y_{ij} \mid b_i] \; dP(b_i) \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{b_i^2}{2\sigma^2}} \prod_{j=1}^{n_j} \frac{e^{y_{ij}(b_i + \boldsymbol{\beta}^T \boldsymbol{x}_{ij})}}{1 + e^{b_i + \boldsymbol{\beta}^T \boldsymbol{x}_{ij}}} \; db_i \\
&= \frac{e^{\boldsymbol{\beta}^T (\sum_j \boldsymbol{y}_{ij} \boldsymbol{x}_{ij})}}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \frac{e^{-\frac{b_i^2}{2\sigma^2} + b_i \sum_j y_{ij}}}{\prod_j (1 + e^{b_i + \boldsymbol{\beta}^T \boldsymbol{x}_{ij}})} \; db_i
\end{aligned}$$

Logistic regression with random intercept — MLE

Then we have the log-likelihood function:

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{N}{2}\log(2\pi\sigma^2) + \boldsymbol{\beta}^T \boldsymbol{r} + \sum_{i=1}^{N} \log \int_{-\infty}^{\infty} e^{h_i(\boldsymbol{\beta}; u)} du$$

and

$$h_i(\boldsymbol{\beta}; u) = -\frac{u^2}{2\sigma^2} + k_i u - \sum_{j=1}^{n_j} \log(1 + e^{u + \boldsymbol{\beta}^T \boldsymbol{x}_{ij}})$$

with

$$k_i = \sum_{j=1}^{n_j} y_{ij}, \quad \text{and} \quad \boldsymbol{r} = \sum_{i=1}^{N} \sum_{j=1}^{n_j} y_{ij} \boldsymbol{x}_{ij}$$

- In order to find the MLE, one needs to approximate the integrals involved in $\ell(\boldsymbol{\beta}, \sigma^2)$ and its partial derivatives. (See textbook for integral approximation methods)

Logistic regression with random intercept — Conditional MLE

Instead of considering MLE, we consider the following conditional probability:

$$\mathbb{P}\left[\boldsymbol{y}_i \middle| \sum_{i=1}^{n_j} y_{ij} = k_i\right] = \frac{\mathbb{P}[\boldsymbol{y}_i]}{\mathbb{P}\left[\sum_{i=1}^{n_j} y_{ij} = k_i\right]} = \frac{\prod_j \frac{e^{y_{ij}(b_i+\boldsymbol{\beta}^T \boldsymbol{x}_{ij})}}{1+e^{b_i+\boldsymbol{\beta}^T \boldsymbol{x}_{ij}}}}{\sum_{\boldsymbol{z}_i \in \mathcal{S}_{n_j,k_i}} \prod_j \frac{e^{z_{ij}(b_i+\boldsymbol{\beta}^T \boldsymbol{x}_{ij})}}{1+e^{b_i+\boldsymbol{\beta}^T \boldsymbol{x}_{ij}}}}$$

$$= \frac{e^{b_i k_i + \boldsymbol{\beta}^T \sum_j y_{ij} \boldsymbol{x}_{ij}}}{\sum_{\boldsymbol{z}_i \in \mathcal{S}_{n_j,k_i}} e^{b_i k_i + \boldsymbol{\beta}^T \sum_j z_{ij} \boldsymbol{x}_{ij}}} = \frac{e^{\boldsymbol{\beta}^T \sum_j y_{ij} \boldsymbol{x}_{ij}}}{\sum_{\boldsymbol{z}_i \in \mathcal{S}_{n_j,k_i}} e^{\boldsymbol{\beta}^T \sum_j z_{ij} \boldsymbol{x}_{ij}}}$$

The conditional probability does not depend on b_i. Therefore, we consider the conditional MLE as

$$\hat{\boldsymbol{\beta}}_{CML} = \arg\max_{\boldsymbol{\beta}} \; \boldsymbol{\beta}^T \left(\sum_{i,j} y_{ij} \boldsymbol{x}_{ij} \right) - \sum_{i=1}^{N} \log \left(\sum_{\boldsymbol{z}: \sum_j z_j = \sum_j y_{ij}} e^{\boldsymbol{\beta}^T \sum_j z_{ij} \boldsymbol{x}_{ij}} \right)$$

Logistic regression with random intercept — Conditional MLE

- Conditional MLE estimates $\boldsymbol{\beta}$ and bypass the random intercepts.
- Conditional MLE is tractable without integral.
- $\mathcal{S}_{n_j,k_i}$ contains all binary vectors of length n_j that sum up to k_i.
- Complex computation for large n_i.

Logistic regression with random intercept — Fixed Sample Approximation

The integral involved in the log-likelihood function:

$$\frac{1}{\sigma}\int_{-\infty}^{\infty} e^{h_i(\boldsymbol{\beta};u)}du = \frac{1}{\sigma}\int_{-\infty}^{\infty} e^{\ell_i(\boldsymbol{\beta};u)}e^{-\frac{u^2}{2\sigma^2}}du = \int_{-\infty}^{\infty} e^{\ell_i(\boldsymbol{\beta};u/\sigma)}e^{-u^2/2}du$$

- Fixed Sample Approximation:
  - Draw a weighted sample (u_s, w_s) from $\mathcal{N}(0,1)$.
  - Approximate the integral by

$$\int_{-\infty}^{\infty} e^{\ell_i(\boldsymbol{\beta};u/\sigma)}e^{-u^2/2}du \approx \sqrt{2\pi}\sum_{s=1}^{S} w_s e^{\ell_i(\boldsymbol{\beta};u_s/\sigma)}$$

Logistic regression with random intercept — Quadratic Approximation

$$\int_{-\infty}^{\infty} e^{\ell_i(\boldsymbol{\beta}; u/\sigma)} e^{-u^2/2} du \quad \text{with} \quad \ell_i(\boldsymbol{\beta}, u) = k_i u - \sum_{j=1}^{n_j} \log(1 + e^{u + \boldsymbol{\beta}^T \boldsymbol{x}_{ij}})$$

- Quadratic Approximation:
  - We find the following quadratic approximation to the second term in ℓ_i:

$$\sum_{j=1}^{n_j} \log(1 + e^{u + \boldsymbol{\beta}^T \boldsymbol{x}_{ij}}) \approx C_{0j} + C_{1j} u + \frac{1}{2} C_{2j} u^2$$

    with

$$C_{0j} = \sum_{j=1}^{n_j} \log(1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_{ij}}), \quad C_{1j} = \sum_{j=1}^{n_j} \frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}_{ij}}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_{ij}}}, \quad C_{2j} = \sum_{j=1}^{n_j} \frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}_{ij}}}{\left(1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_{ij}}\right)^2}$$

  - The integral is now tractable in the form of $\int e^{C_{0j} + (C_{1j} + k_i) u - (1 - C_{2j}) u^2/2} du$.

Logistic regression with random intercept — Laplace Approximation

We consider the integral

$$\int e^{h(x)} dx$$

If we approximate $h(x)$ at its maximum by

$$h(x) \approx h_{max} + \frac{1}{2}(x - x_{max})^2 \left(- \frac{d^2 h}{dx^2} \Big|_{x = x_{max}} \right)$$

(why no first-order term?)
Then we have

$$\int_{-\infty}^{\infty} e^{h(x)} dx \approx \sqrt{2\pi} h_{max} \sqrt{ - \frac{d^2 h}{dx^2} \Big|_{x = x_{max}} }$$

- Similar idea as the quadratic approximation — but taken at the maximum.

Logistic regression with random intercept — Summary

- MLE is difficult to maximize because of the integral.
- Approximate the integral in a tractable way:
  - Fixed Sample Approximation — replace integral by weighted sum.
  - Quadratic Approximation — approximate exponent by quadratic function.
  - Laplace Approximation — Taylor expansion of the exponent at the maximum.
- Conditional MLE: bypass the integral by considering a conditional likelihood.
- Choice of models:
  - Small n_i: conditional MLE.
  - Large n_i ($\gg N$): fixed-effect models with dummy variables.
  - Large N or moderate n_i: random-effect models.

Mixed Models with Multiple Random Effects

Now we consider the following model:

$$\mathbb{P}[y_{ij} = 1] = \mu(\boldsymbol{\beta}^T \boldsymbol{x}_{ij} + \boldsymbol{b}_i^T \boldsymbol{z}_{ij})$$

with $\boldsymbol{b}_i \sim \mathcal{N}(0, \boldsymbol{D}_*)$.

Why $\boldsymbol{D}_*$ not $\sigma^2 \boldsymbol{D}$?

We will focus on the logistic regression.

Logistic Regression with Multiple Random Effects

$$\mathbb{P}[y_{ij} = 1] = \frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}_{ij} + \boldsymbol{b}_i^T \boldsymbol{z}_{ij}}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_{ij} + \boldsymbol{b}_i^T \boldsymbol{z}_{ij}}}$$

with $\boldsymbol{b}_i \sim \mathcal{N}(0, \boldsymbol{D}_*)$.
The log-likelihood function: (let $\boldsymbol{D}_- = \boldsymbol{D}_*^{-1}$)

$$\ell(\boldsymbol{\beta}, \boldsymbol{D}_-) = \frac{N}{2} \log |\boldsymbol{D}_-| + \boldsymbol{\beta}^T \boldsymbol{r} + \sum_{i=1}^{N} \log \int_{\mathbb{R}^k} e^{h_i(\boldsymbol{\beta}; \boldsymbol{u})} d\boldsymbol{u} + \text{const}$$

with

$$h_i(\boldsymbol{\beta}; \boldsymbol{u}) = \boldsymbol{k}_i^T \boldsymbol{u} - \frac{1}{2} \boldsymbol{u}^T \boldsymbol{D}_- \boldsymbol{u} - \sum_{j=1}^{n_i} \log(1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_{ij} + \boldsymbol{b}_i^T \boldsymbol{z}_{ij}})$$

and $\boldsymbol{k}_i = \sum_{j=1}^{n_i} y_{ij} \boldsymbol{z}_{ij}$ and $\boldsymbol{r} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} y_{ij} \boldsymbol{x}_{ij}$.
Solution: Laplace Approximation + Penalized Quasi-likelihood.

Generalized Linear Mixed Models via Exponential Families

- Exponential Family: (in natural parameters)

$$f(y;\theta) = e^{\theta y - b(\theta) - c(y)}$$

- GLMM via exponential family:

$$y_{ij} \mid \boldsymbol{b}_i \sim f(y; \boldsymbol{\beta}^T \boldsymbol{x}_{ij} + \boldsymbol{b}_i^T \boldsymbol{z}_{ij}), \quad \boldsymbol{b}_i \sim \mathcal{N}(0, \boldsymbol{D}_*)$$

- For example, in logistic regression,

$$f(y;\theta) = e^{\theta y - \log(1+e^{\theta})}$$

Generalized Linear Mixed Models via Exponential Families

- The log-likelihood function:

$$\ell(\boldsymbol{\beta}, \boldsymbol{D}_*) = -\frac{N}{2} \log |\boldsymbol{D}_*| + \sum_{i=1}^{N} \log \int e^{\ell_i(\boldsymbol{\beta}, \boldsymbol{u}) - \boldsymbol{u}^T \boldsymbol{D}_*^{-1} \boldsymbol{u}/2} d\boldsymbol{u}$$

with

$$\ell_i(\boldsymbol{\beta}, \boldsymbol{u}) = \sum_{j=1}^{n_i} \left[(\boldsymbol{\beta}^T \boldsymbol{x}_{ij} + \boldsymbol{u}^T \boldsymbol{z}_{ij}) y_{ij} + b(\boldsymbol{\beta}^T \boldsymbol{x}_{ij} + \boldsymbol{u}^T \boldsymbol{z}_{ij}) \right]$$

- Solution: Laplace Approximation + Penalized Quasi-likelihood

Penalized Quasi-likelihood

- Let $\ell_{LA}(\boldsymbol{\beta}, \boldsymbol{D}_*; \boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_N^*)$ be the Laplace approximation of the log-likelihood function at maximums $\boldsymbol{u}_i, i = 1, \ldots, N$.
- Let the penalized quasi-likelihood function be

$$\ell_{PQL}(\boldsymbol{\beta}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_N; \boldsymbol{D}_*) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\beta}, \boldsymbol{u}) - \frac{1}{2} \sum_{i=1}^{N} \boldsymbol{u}_i^T \boldsymbol{D}_*^{-1} \boldsymbol{u}_i$$

- The LA + PQL algorithm:
  - Initialized the estimators.
  - Update $\boldsymbol{u}_i^*, i = 1, \ldots, N$ according to ℓ_{PQL}
  - Update $\boldsymbol{D}_*$ according to ℓ_{LA}.
  - Update $\boldsymbol{\beta}$ according to ℓ_{PQL}.
  - Repeat until convergence.

Marginal Model

Now we consider a marginal model:

$$\mathbb{E}[\boldsymbol{y}_i] = \mu(\boldsymbol{X}_i\boldsymbol{\beta}), \quad \text{Cov}(\boldsymbol{y}_i) = \boldsymbol{V}_i$$

- Often, we assume certain structure for $\boldsymbol{V}_i$. For example:

$$\boldsymbol{V}_i = \boldsymbol{D}_i^{1/2}\boldsymbol{R}\boldsymbol{D}_i^{1/2},$$

  where $\boldsymbol{D}_i$ is diagonal with elements $\mu(1-\mu)$, and $\boldsymbol{R}$ is assumed to be an exchangeable **correlation** matrix.
- To solve the marginal model, we can use GEE:

$$\sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{\mu}_i' \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \boldsymbol{0}$$

Example

We consider the example of Problem 6 of Section 7.5 in the textbook. The dataset looks like

```r
source("./Data/MixedModels/Chapter07/psdat.r")
head(psdat)
```

	hrr	visitin	visitout	visittot	black	age10	female	daysfu
1	1	3	0	3	0	8.3	0	2
2	1	23	4	27	0	8.0	1	933
3	1	21	24	45	1	7.9	0	301
4	1	13	5	18	0	6.7	0	411
5	1	39	48	87	0	7.7	0	1325
6	1	0	0	0	1	8.1	0	12

Example — GLMM with PQL

```r
library(MASS)
glmmPQL(fixed=visittot~black+female+age10+age10^2,
        random=~1|hrr,
        data=psdat,
        family=poisson)
```

```
Linear mixed-effects model fit by maximum likelihood
    Data: psdat
    Log-likelihood: NA
Fixed: visittot ~ black + female + age10 + age10^2
(Intercept)       black      female       age10
4.99719386 -0.02701705 -0.02863541 -0.14080444

Random effects:
Formula: ~1 | hrr
        (Intercept) Residual
StdDev:   0.1973408 5.820526
```

Example — GLMM with MLE

```r
library(lme4)
glmer(visittot~black+female+age10+age10^2 + (1|hrr),
      data=psdat,
      family=poisson,
      nAGQ=1)
```

```
Generalized linear mixed model fit by maximum likelihood   (Laplace Approximation) ['glmerMod']
Family: poisson  ( log )
Formula: visittot ~ black + female + age10 + age10^2 + (1 | hrr)
Data: psdat
    AIC       BIC    logLik  deviance  df.resid
342622.9  342658.9 -171306.5  342612.9      9861
Random effects:
Groups Name        Std.Dev.
hrr    (Intercept) 0.2967
Number of obs: 9866, groups:  hrr, 306
Fixed Effects:
(Intercept)        black       female        age10
    4.92558     -0.04742     -0.02831     -0.13820
```

Example — GLMM with MLE

- The `nAGQ` specifies how to approximate the integral in the log-likelihood function.
- `nAGQ`: parameter for Adaptive Gauss-Hermite Quadrature.
- `nAGQ=1`: Laplace Approximation
- Higher `nAGQ`: better approximation but slower computation.

Example — GEE

```r
library(gee)
gee(visittot~black+female+age10+age10^2,
     id=hrr,
     data=psdat,
     family=poisson,
     corstr="exchangeable")
```

```
GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
 Link:                      Logarithm
 Variance to Mean Relation: Poisson
 Correlation Structure:     Exchangeable
...
Coefficients:
(Intercept)       black      female       age10
 5.02944411 -0.03652173 -0.02930477 -0.14431704

Estimated Scale Parameter:  41.755
Number of Iterations:  4
```

Example — `family` argument

- The `family` argument specifies the error structure and link function of the generalized linear model.
- The formula is `dist(link)`, e.g. `binomial(link='logit')` is the logistic model.
- Common families:
  - Binomial/Bernoulli: `binomial`
    link: `'logit'`, `'probit'`, `'log'`, `'cloglog'`
  - Poisson: `poisson`
    link: `'log'`, `'identity'`, `'sqrt'`
  - Gaussian: `gaussian`
    link: `'identity'`, `'log'`, `'inverse'`