Lecture 10 | 05.05.2025

Generalized linear models
with random effects

GLM extensions for the longitudinal data

- **Marginal models**
  - primary interest is given to the conditional mean structure
  - separate model for the mean and the correlated observations
- **Random effects models**
  - one equation used to account for both—the mean and the correlation
  - mostly used when subject specific inference is of some interest
- **Transition models**
  - primary interest again with respect to the mean structure
  - the correlation structure due to historical observations within the model

All three categories of the regression models for correlated (repeated) observations lead to the same model (with the same interpretation) for the Gaussian type of the data but for the discrete data different models can produce different interpretations (due to non-linearity involved)

GLM with random effects

- **Mean structure**

$$\mu_{ij} = E[Y_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{w}_i] = \psi'(\theta_{ij})$$

- **Variance structure**

$$v_{ij} = Var[Y_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{w}_i] = \psi''(\theta_{ij})\phi$$

where we assume the exponential family for the conditional distribution of $Y_{ij}|(\boldsymbol{X}_{ij}, \boldsymbol{w}_i)$ with $f_{(Y|\boldsymbol{X}_{ij},\boldsymbol{w}_i)}(y) = \exp\{[y\theta_{ij} - \psi(\theta_{ij})]/\phi + c(y,\phi)\}$, where $g(\mu_{ij}) = \boldsymbol{X}_{ij}^\top\beta + \boldsymbol{Z}_{ij}^\top\boldsymbol{w}_i$, $\psi'(\theta_{ij}) = \mu_{ij}$, and $v_{ij} = v(\mu_{ij})\phi$
(link function g; normalization/cumulant function ψ; variance function v)

- **Covariance structure**
Determined/induced by random effects $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N$ which are independent and have some common underlying distribution and the subject specific responses $Y_{i1}, \ldots, Y_{in_i}$ are, conditionally on $\boldsymbol{w}_i$, independent

Generalized linear mixed models (GLMM)

- for a subject specific response vector $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in_i})^\top$ and explanatory vectors of covariates $\mathbb{X}_i = (\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{in_i})^\top \in \mathbb{R}^{n_1 \times p}$ it is assumed that responses Y_{ij} for $j = 1, \dots, n_i$, are, given the random effects $\boldsymbol{w}_i \in \mathbb{R}^r$, independent with the density function,

$$f(y|\boldsymbol{X}_{ij}, \boldsymbol{w}_i) = \exp\{\phi^{-1}[y\theta_{ij} - \psi(\theta_{ij})] + c(y, \phi)\}, \quad \text{for } y \in \mathbb{R}$$

  where, in addition, it holds that

  - canonical parameter $\theta_{ij} = \boldsymbol{X}_{ij}^\top \beta + \boldsymbol{Z}_{ij}^\top \boldsymbol{w}_i$
  - random effects follow the Gaussian distribution $\boldsymbol{w}_i \sim N_r(\boldsymbol{0}, \mathbb{G})$
  - the mean-variance relationship is induced by the full joint distribution

- thus, for the whole subject's specific vector $\boldsymbol{Y}_i$ given $\mathbb{X}_i$ and the random effects in $\boldsymbol{w}_i$ we obtain the conditional density of $\boldsymbol{Y}_i|(\mathbb{X}_i, \boldsymbol{w}_i)$

$$f_i(\boldsymbol{y}|\mathbb{X}_i, \boldsymbol{w}_i) = \prod_{j=1}^{n_i} f(y_j|\boldsymbol{X}_{ij}, \boldsymbol{w}_i), \quad \text{for } \boldsymbol{y} = (y_1, \dots, y_{n_i})^\top \in \mathbb{R}^{n_i}$$

Note that both densities still depend on $\beta \in \mathbb{R}^p$ and $\phi > 0$ thus, we have a fully specified parametric model—the notation, for brevity, does not reflect this dependence

From hierarchical to marginal model

- Note, that the model specification is given in terms of a hierarchical model (two hierarchical model equations and subject specific interpretation)

$$Y_{ij}|(\boldsymbol{X}_{ij}, \boldsymbol{w}_i) \sim f(y|\boldsymbol{X}_{ij}, \boldsymbol{w}_i) \quad \text{and} \quad \boldsymbol{w}_i \sim N_r(\boldsymbol{0}, \mathbb{G})$$

  and, analogously, also

$$\boldsymbol{Y}_i|(\mathbb{X}_i, \boldsymbol{w}_i) \sim f_i(\boldsymbol{y}|\mathbb{X}_i, \boldsymbol{w}_i) \quad \text{and} \quad \boldsymbol{w}_i \sim N_r(\boldsymbol{0}, \mathbb{G})$$

  (note that $f \equiv f(y|\boldsymbol{X}_{ij}, \boldsymbol{w}_i): \ \mathbb{R} \to \mathbb{R}$ while $f_i \equiv f_i(\boldsymbol{y}|\mathbb{X}_i, \boldsymbol{w}_i): \ \mathbb{R}^{n_i} \to \mathbb{R}$)

- The marginal model $f(\boldsymbol{y}|\mathbb{X}_i)$ is obtained by integrating over $\boldsymbol{w}_i \sim N_r(\boldsymbol{0}, \mathbb{G})$ which gives the marginal (conditional) density (population interpretation)

$$f_i(\boldsymbol{y}|\mathbb{X}_i) = \int_{\mathbb{R}^r} f_i(\boldsymbol{y}|\mathbb{X}_i, \boldsymbol{w})\mathrm{d}\Phi(\boldsymbol{w}) = \int_{\mathbb{R}^r} \prod_{j=1}^{n_i} f(y_j|\boldsymbol{X}_{ij}, \boldsymbol{w})f_\Phi(\boldsymbol{w})\mathrm{d}\boldsymbol{w}$$

  where $\Phi(\cdot)$ is the distribution function of the Gaussian $N_r(\boldsymbol{0}, \mathbb{G})$ distribution and f_Φ is the corresponding density function

GLMM Likelihood and log-likelihood

- the marginal distribution of $\boldsymbol{Y}_i$ (conditionally on $\mathbb{X}_i$) depends on unknown (mean structure) parameters $\boldsymbol{\beta} \in \mathbb{R}^p$, the over-dispersion parameter $\phi > 0$, and a (positive-definite) variance-covariance matrix $\mathbb{G} \in \mathbb{R}^{r \times r}$
- thus, given the longitudinal observations in $\mathcal{D} = \{(\boldsymbol{Y}_i, \mathbb{X}_i);\ i = 1, \dots, N\}$ the full likelihood for $\boldsymbol{\beta}, \phi$, and $\mathbb{G}$ can be expressed as

$$
\begin{aligned}
L(\boldsymbol{\beta}, \mathbb{G}, \phi, \mathcal{D}) &= \prod_{i=1}^{N} f_i(\boldsymbol{Y}_i | \mathbb{X}_i) \\
&= \prod_{i=1}^{N} \int_{\mathbb{R}^r} \prod_{j=1}^{n_i} f(y_j | \boldsymbol{X}_{ij}, \boldsymbol{w}) f_{\Phi}(\boldsymbol{w}) \mathrm{d}\boldsymbol{w}
\end{aligned}
$$

(first equality due to independent subjects, second equality due to conditional independence given $\boldsymbol{w}_i$)

- Under the normal linear model, the integral(s) can be worked out analytically but, in general, various (numerical) approximations are needed to obtain the final solution (likelihood)
- Subject specific interpretation can be carried out in terms of the random effects predictions based on the posterior distribution $f(\boldsymbol{w}_i | \boldsymbol{Y}_i, \mathbb{X}_i)$

Computational issues and approximations

- The marginal and joint likelihoods involve integrating out the random effects where the integrals are typically not available in closed form (especially for non-Gaussian outcomes)

- **Most crucial problems**
  - High-dimensional integrals
    (dimensionality of the random effects, $r \in \mathbb{N}$)
  - Non-convexity of the likelihood
    (sensitive to starting values)
  - Identifiability issues
    (fixed & random effects, variance & overdispersion)
  - Slow convergence of iterative algorithms
    (resp., the algorithms may get stuck in local maxima)

- **Three types of numerical approximations are typically used**
  - approximation of the integrand
  - approximation of the data
  - approximation of the integral

I. Laplace approximation of the integrand

- the integral(s) in the likelihood $L(\beta, \mathbb{G}, \phi, \mathcal{D})$ can be equivalently rewritten in a form $\int e^{Q(\boldsymbol{w})} \mathrm{d}\boldsymbol{w}$ and the second order Taylor expansion can be applied to approximate the quantity $Q(\boldsymbol{w})$
- The idea is to approximate an untractable non-Gaussian integral by by a Gaussian (normal) integral, which can be (typically) evaluated analytically
- Taylor expansion around the mode $\overline{\boldsymbol{w}} \in \mathbb{R}^r$ (i.e., the point where $Q(\boldsymbol{w})$ is maximized) as

$$Q(\boldsymbol{w}) \approx Q(\overline{\boldsymbol{w}}) + \frac{1}{2}(\boldsymbol{w} - \overline{\boldsymbol{w}})^\top Q''(\overline{\boldsymbol{w}})(\boldsymbol{w} - \overline{\boldsymbol{w}})$$

- provides reasonably good approximation in case of many repeated measurements per each subject and faster than numerical integration
- the accuracy of the approximation substantially decreases when the dimensionality increases and when the likelihood is flat

II. Approximation of the data

- The GLMM model $g(\mu_{ij}) = g(E[Y_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{w}_i]) = \boldsymbol{X}_{ij}^\top \beta + \boldsymbol{Z}_{ij}^\top \boldsymbol{w}_i$ is rewritten (approximated) in a (linear) form

$$Y_{ij} = g^{-1}(\boldsymbol{X}_{ij}^\top \beta + \boldsymbol{Z}_{ij}^\top \boldsymbol{w}_i) + \varepsilon_{ij} = \mu_{ij} + \varepsilon_{ij}$$

  where $\varepsilon_{ij} \sim (0, \phi v(\mu_{ij}))$ and the Taylor expansion is applied to μ_{ij}

- First/second order Taylor expansion
  - around current values of $\widehat{\beta}$ and $\widehat{\boldsymbol{w}}_i$ Penalized quasi-likelihood (PQL)
    (random effects are treated as penalty terms and standard LMM techniques are applied)
  - around current values of $\widehat{\beta}$ and $\widehat{\boldsymbol{w}}_i = \boldsymbol{0}$ Marginal quasi-likelihood (MQL)
    (linearization of the fixed effects only but often underestimating the variance components)
- MQL only performs relatively well if the random effects variance is small
- Both, PQL and MQL perform bad for binary data with only few observations per subject
- For increasing number of measurements within each subject MQL remains biased while PQL is consistent (MQL not that much used today)

III. Approximation of the integral

- The likelihood contribution of each subject is

$$f_i(\boldsymbol{Y}_i|\mathbb{X}_i) = \int_{\mathbb{R}^r} \prod_{j=1}^{n_i} f(Y_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{w}) f_\Phi(\boldsymbol{w}) \mathrm{d}\boldsymbol{w} = \int f(\boldsymbol{w}) f_\Phi(\boldsymbol{w}) \mathrm{d}\boldsymbol{w}$$

  where $f(\boldsymbol{w}) = f_i(\boldsymbol{Y}_i|\mathbb{X}_i, \boldsymbol{w})$, taken as the function of $\boldsymbol{w} \in \mathbb{R}^r$
- Gaussian quadrature method is used to approximate the last integral above in a way

$$\int f(\boldsymbol{w}) f_\Phi(\boldsymbol{w}) \mathrm{d}\boldsymbol{w} \approx \sum_{q=1}^{Q} \omega_q f(\boldsymbol{w}_q)$$

  for some weights $\omega_q > 0$ and a sequence of nodes $\boldsymbol{w}_q \in \mathbb{R}^r$, $q = 1, \ldots, Q$
- The Gaussian quadrature uses nodes and weights that are fixed and the adaptive Gaussian quadrature adjusts the nodes and the weights to adapt to the support of $f(\boldsymbol{w}) f_\Phi(\boldsymbol{w})$ (but it is also more time consuming) (addaptive Gaussian quadrature with $Q = 1$ is equivalent with Laplace)

Parameter interpretation

- The parameter vector $\boldsymbol{\beta} \in \mathbb{R}^p$ is interpreted differently in GEE model (marginal interpretation) and GLMM model (conditional interpretation)
- In general, the marginal model is not of the same parametric form as the conditional average in the GLMM formulation
- For logistic mixed regression model in particular (with a normally distributed random intercept) it can be shown that the marginal model can be well approximated by another logistic (marginal) model, however, with the parameters $\widehat{\boldsymbol{\beta}}^{\text{GLMM}} = \sqrt{c^2\sigma^2 + 1}\widehat{\boldsymbol{\beta}}^{\text{M}}$ where σ^2 is the variance of the random intercept and $c \approx 16\sqrt{3}/(15\pi)$
- In practice, the marginal interpretation can be derived from the GLMM output by integrating out the random effects part of the model which can be done numerically, for instance, by the Gaussian quadrature or sampling methods

Summary

- GLM models with random effects are particularly suitable in situations where subject specific inference is of interest
- Underlying theory and the estimation is based on maximum likelihood and its properties (quasilikelihood and marginal likelihood)
- Different computational approaches and algorithms are used to obtain the solution—the estimates of the fixed effect parameters
- Marginal form of the GLM model with random effects has, generally, different interpretation of the parameters than the marginal model
- The right choice of the model always depends on the question of interest, the underlying data, and the sole purpose of the model