Generalized Linear Mixed Models with Random Effects

- The logistic regression model with random intercept

  Example: 2×2 crossover trial

  Example: Indonesian Children Health Study

- The Poisson regression model with random intercept

  Example: seizure data

Random Effects GLM

Tha basic idea: there is natural heterogeneity among subjects.

- Systematic part

$$g(E[Y_{ij} \mid U_i]) = \beta_0 + \beta_1 x_{ij} + U_i$$

- Random Part

$$\begin{aligned} Y_{ij} \mid U_i &\sim \text{GLM} \\ &\quad\ \text{Normal} \\ &\quad\ \text{Bernoulli} \\ &\quad\ \text{Poisson} \\ U_i &\sim N(0, G) \end{aligned}$$

Example: 2×2 crossover trial

In random effects models, the regression coefficients measure the more direct influence of explanatory variables on the responses for heterogeneous individuals. For example

$$\begin{aligned} \text{logit} P(Y_{ij} \mid U_i) &= \beta_0^* + U_i + \beta_1^* x_{ij} \\ x_{ij} &= \begin{cases} 1 & \text{treat. A} \\ 0 & \text{plac. B} \end{cases} \end{aligned}$$

This model states that:

1. each person has their own probability of positive response under a placebo (B)

$$P(Y_{ij} = 1 \mid U_i, x_{ij} = 0) = \frac{\exp(\beta_0^* + U_i)}{1 + \exp(\beta_0^* + U_i)}$$

2. a person's odds of a normal response are multiplied by $\exp(\beta_1^*)$ when taking the drug A, regardless of the initial risk

$$\begin{aligned} \frac{P(Y_{ij}=1|U_i,x_{ij}=0)}{P(Y_{ij}=0|U_i,x_{ij}=0)} &= \exp(\beta_0^* + U_i) \\ \frac{P(Y_{ij}=1|U_i,x_{ij}=1)}{P(Y_{ij}=0|U_i,x_{ij}=1)} &= \exp(\beta_0^* + \beta_1^* + U_i) \\ \frac{P(Y_{ij}=1|U_i,x_{ij}=1)}{P(Y_{ij}=0|U_i,x_{ij}=1)} &= \frac{P(Y_{ij}=1|U_i,x_{ij}=0)}{P(Y_{ij}=0|U_i,x_{ij}=0)} \exp(\beta_1^*) \end{aligned}$$

In Logistic Models

1. $| \boldsymbol{\beta} | \leq | \boldsymbol{\beta}^* |$
2. $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ if and only if $VarU_i = 0$
3. $\hat{\boldsymbol{\beta}}/se(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}^*/se(\hat{\boldsymbol{\beta}}^*)$
4. Marginal Model estimates smaller than random effects estimates
5. Tests of hypotheses approximately the same

Correspondence between regression parameters in random effects and marginal models

- let $\boldsymbol{\beta}$ the vector of regression coefficients under a marginal model
- let $\boldsymbol{\beta}^*$ the vector of regression coefficients under a random effects model
- G is the variance of the random effects

$$\boldsymbol{\beta} \simeq (0.346G^2 + 1)^{-1/2}\boldsymbol{\beta}^*$$

Estimation of Generalized Linear Mixed Models

- $f(Y_{ij} \mid U_i)$ in the exponential family
- $Y_{i1}, \ldots, Y_{in_i} \mid U_i$ are independent
- $U_i \sim f(U_i, G)$
- **Maximum Likelihood**

  U_i is a set of unobserved variables which we integrate out of the likelihood

Maximum Likelihood estimation of G and β

1. We will treat $U_i \sim N(0, G)$ *we can learn about one individual's coefficients by understanding the variability in coefficients across the population*
2. if G_i is small $\rightarrow$ *rely on population average coefficients to estimate those for an individual*

   we weight the cross-sectional information more heavily and we borrow strength across subjects
3. if G_i is large $\rightarrow$ *rely more heavily on the data from each individual in estimating their own coefficients*

   we weight the longitudinal information more heavily since comparisons within a subject are likely to be more precise than comparisons among subjects.

Maximum likelihood estimation

"averaged-away" the random effects

$$\underbrace{L(\boldsymbol{\beta}, G, \boldsymbol{Y})}_{\text{what we see}} = \prod_{i=1}^{m} \int \prod_{j=1}^{n_i} \underbrace{Pr(Y_{ij} \mid U_i, \boldsymbol{\beta})}_{\text{what we think exists}} \quad f(U_i, G) dU_i$$

- Used all the data
- EM algorithm
- Numerical integration

Maximum Likelihood

Example: 2×2 crossover trial

$$\begin{aligned} \text{logit} P(Y_{ij} = 1 \mid U_i) &= \beta_0 + U_i + \beta_1 x_{ij1} + \beta_2 x_{ij2} \\ U_i &\sim N(0, G) \end{aligned}$$

- $x_{ij1} = 1$ if active drug (A) or 0 if placebo (B)
- $x_{ij2} = 1$ if period 2 or 0 if period 1
- $\sqrt{\hat{G}} = 4.9$: 95% of the subjects would fall between $(-2 \times 4.9, 2 \times 4.9)$ logit units of the overall mean.
- This range on the logit scale translates into probabilities between 0 and 1, i.e. some people have little chance and others have very high chance of a normal reading in the placebo and the treatment group.
- Assuming a constant treatment effect for all persons, the odds of a normal response for a subject are estimated to be $\exp(\hat{\beta}_1) = \exp(1.9) = 6.7$ times higher on the active drug than on the placebo.

In Summary

Example: 2×2 crossover trial

- Marginal Model $\hat{\beta}_1 = 0.57(0.23),\ \exp(0.57) \simeq 2$
- Random effects Model (maximum likelihood) $\hat{\beta}_1 = 1.9(0.91),\ \exp(1.$ $7,\ \hat{G} = 4.9$
- The smaller value from the marginal analysis is consistent with the theoretical inequality: $(0.346\hat{G} + 1)^{1/2} = 3.1$

Indonesian Study - Maximum Likelihood approach

With a random effect model, here we can address the question of how an individual child's risk for respiratory infection will change if their vitamin A status were to change.

- we assume that each child have a distinct intercept which represents their propensity to infection
- we have accounted for correlation by including random intercepts $U_i \sim N(0, G)$
- $\sqrt{\hat{G}} = 0.72 \Rightarrow$ considerable heterogeneity among children
- Among children with linear predictor equal to the intercept -2.2 (average age, height, female, vitamin A sufficient), about 95% would have a probability of infection between 0.03 and 0.31:

$$\begin{aligned} P(Y_{ij} = 1 \mid U_i = -2 \times 0.72) &= \tfrac{\exp(-2.2-2\times 0.72)}{1+\exp(-2.2-2\times 0.72)} = 0.025 \\ P(Y_{ij} = 1 \mid U_i = +2 \times 0.72) &= \tfrac{\exp(-2.2+2\times 0.72)}{1+\exp(-2.2+2\times 0.72)} = 0.31 \end{aligned}$$

- relative odds of infection associated with vitamin A deficiency are $\exp(0.54) = 1.7$

- the longitudinal age effect on the risk of RI in Model 2, can be explained by the seasonal trend in model 3
- Because of the small heterogeneity the estimates of the coefficients obtained under a random effects models are similar to the marginal model coefficients
- the ratio of the RE coefficients and marginal coefficients are close to $(0.346\hat{G}+1)^{1/2}$

Poisson model with random intercept: Epileptic seizure example

$$\begin{aligned}\log E(Y_{ij} \mid \gamma_i) &= \gamma_i + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \\ &+ \beta_3 x_{ij1} x_{ij2} + \log(t_{ij}) \\ j &= 0, 1, \ldots, 4 \\ i &= 1, \ldots, 59 \\ \gamma_i &= \beta_0 + U_i\end{aligned}$$

where

$$x_{ij1} = \begin{cases} 1 & \text{if the } ith \text{ subject is assigned to the prog. group} \\ 0 & \text{if the } ith \text{ subject is assigned to the pla. group} \end{cases}$$

$$x_{ij2} = \begin{cases} 1 & \text{if } j = 1, 2, 3, 4 \\ 0 & \text{if } j = 0. \end{cases}$$

Seizure Example

Group	Visit	Expected # seizures
Placebo $(x_1 = 0)$	0 $(x_2 = 0)$	$8\exp(\gamma_i)$
Placebo $(x_1 = 0)$	1,2,3,4 $(x_2 = 1)$	$2\exp(\gamma_i + \beta_2)$
Progabide $(x_1 = 1)$	0 $(x_2 = 0)$	$8\exp(\gamma_i + \beta_1)$
Progabide $(x_1 = 1)$	1,2,3,4 $(x_2 = 1)$	$2\exp(\gamma_i + \beta_1 + \beta_2 + \beta_3)$

- $t_{i0} = 8$
- $t_{i1} = t_{i2} = t_{i3} = t_{i4} = 2$

- $\exp(\gamma_i)$ is the expected baseline seizure count for the ith subject, $i = 1, \ldots, 59$

$$E[Y_{ij} \mid U_i, x_1 = 0, x_2 = 0]/8 = \exp(\gamma_i)$$

- β_2 represents the log ratio of the seizure rates post versus pre-randomization for the placebo group *(same for each subject)*:

$$\frac{(E[Y_{ij}|U_i, x_1=0, x_2=1]/2)}{(E[Y_{ij}|U_i, x_1=0, x_2=0]/8)} = \frac{\exp(\gamma_i+\beta_2)}{\exp(\gamma_i)} = \exp(\beta_2)$$

- $\beta_2 + \beta_3$ represents the log ratio of the seizure rates post versus pre-randomization for the treatment group *(same for each subject)*:

$$\frac{(E[Y_{ij}|U_i, x_1=1, x_2=1]/2)}{(E[Y_{ij}|U_i, x_1=1, x_2=0]/2)} = \frac{\exp(\gamma_i+\beta_1+\beta_2+\beta_3)}{\exp(\gamma_i+\beta_1)} = \exp(\beta_2+\beta_3)$$

- $\exp(\beta_3)$ represents the ratio of seizure rates post versus pre-randomization for the treatment group divided by seizure rates post versus pre-randomization for the control group.
- a negative value of β_3 indicates that a relatively larger fraction of the total seizures in the treatment group occurred before rather

than after randomization as compared to the placebo group. (i.e. the treatment is effective)

- $\hat{\beta}_3 = -0.10(0.065)$ (modest evidence that the progabide is effective)
- the model doesn't fit well
- extend the model by including a random slope U_{i2}

Poisson-Gaussian random effects models: Epileptic seizure

$$\begin{aligned}\log E(Y_{ij} \mid U_{i1}, U_{i2}) &= \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \\ &+ \beta_3 x_{ij1} x_{ij2} + U_{i1} + U_{i2} x_{ij2} + \\ &+ \log(t_{ij}) \\ (U_{i1}, U_{i2}) &\sim N\left((0,0), \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}\right)\end{aligned}$$

- here we are assuming that there might be heterogeneity among subjects in the ratio of the expected seizure counts before and after the randomization.
- the degree of heterogeneity can be measured by G_{22}
- maximum likelihood estimation
- ratio of seizure counts in the placebo post-to-pre treatment is subject specific

$$\begin{aligned}\frac{(E[Y_{ij}|U_{i1},U_{i2},x_1=0,x_2=1]/2)}{(E[Y_{ij}|U_{i1},U_{i2},x_1=0,x_2=0]/2)} &= \frac{\exp(\beta_0+U_{i1}+\beta_2+U_{i2})}{\exp(\beta_0+U_{i1})} \\ &= \exp(\beta_2 + U_{i2})\end{aligned}$$

- ratio of seizure counts in the progabide group post-to-pre treatment is subject specific

$$\begin{aligned}\frac{(E[Y_{ij}|U_{i1},U_{i2},x_1=1,x_2=1]/2)}{(E[Y_{ij}|U_{i1},U_{i2},x_1=1,x_2=0]/2)} &= \frac{\exp(\beta_0+U_{i1}+\beta_1+\beta_2+U_{i2}+\beta_3)}{\exp(\beta_0+U_{i1}+\beta_1)} = \\ &= \exp(\beta_2 + U_{i2} + \beta_3)\end{aligned}$$

Results

- The estimate of G_{22} is statistical significant, therefore the data give support of between subject variability in the the ratio of the expected seizure counts before and after the randomization.
- $\exp(\hat{\beta}_2) = \exp(0.002) = 1.002$ subjects in the placebo group with $U_{i2} = 0$ have expected seizure rates after the treatment which are estimated to be roughly the same as before treatment.
- Subjects in the progabide group with $U_{i2} = 0$, the seizure rates are reduced after the treatment by about 27% $(1-\exp(\hat{\beta}_2+\hat{\beta}_3) = (1 - \exp(0.002 - 0.31) = 0.27)$
- the treatment seems to have a modest effect: $\hat{\beta}_3 = -0.31,\ (0.15)$
- without patient 207 the evidence of the progabide is stronger $\hat{\beta}_3 = -0.34(0.15)$