Models with Random Effects

Generalized Linear Mixed Models

Demetris Athienitis

Introduction

- Conditional models allow for subject specific terms in the model and all interpretations are conditional on subject. Correlation will be accounted for by the subject specific effects.
- Assume that data are clustered (in some way)
  - Same observations on one subject form a cluster
  - Individuals within a family are a cluster

  If we are referring to a specific cluster i, we will use $(Y_{i1}, \ldots, Y_{iT})$
- Assume that there is correlation within outcomes in a cluster, but not across clusters

Introduction

Random effect models introduce subject or cluster-specific effects. These effects are assumed to follow some probability distribution in the population, typically a normal distribution.

Randomness in the random effects induces correlation within a cluster.

GLMM

A Generalized Linear Mixed Model (GLMM) with a random effects is able to account for having multiple responses per subject (or "cluster") by putting a subject term in model.

Let Y_{it} = binary response by subject i at time t.

$$\text{logit}\left[P(Y_{it}=1)\right] = \alpha_i + \beta x_{it}, \quad t = 1, \ldots, T$$

α_i varies by subject so that a heterogeneous population implies a highly variable $\{\alpha_i\}$. Treating α_i as fixed is not possible because this model would yield at least n parameters, yielding an over-parameterized model, so the solution is to treat it as random, i.e. $\alpha_i \overset{\text{ind.}}{\sim} N(\alpha, \sigma_u^2)$ or equivalently

$$\alpha_i = \alpha + u_i, \quad u_i \sim N(0, \sigma_u^2)$$

GLMM

- Magnitude of σ_u^2 controls the amount of variability across subjects and how much correlation exists within a cluster. Larger σ_u^2 values lead to higher correlation within clusters
- Assuming a common distribution “borrows information" across subjects and cluster-specific effects are shrunk towards an overall mean
- Parameters α and β are *fixed effects* and $\{u_i\}$ are *random effects*. Fixed effects are estimated, as well as σ_u^2, and predictions of $\{u_i\}$ can use the mean 0 or randomly generate from $N(0, \sigma_u^2)$

$$\text{logit}\left[P(Y_{it} = 1)\right] = \alpha + u_i + \beta x_{it}, \quad t = 1, \ldots, T$$

GLMM

$Y_{i1}, Y_{i2}, \ldots, Y_{iT}$ are conditionally independent given u_i, but marginally dependent. That is, responses within subject more alike than between subjects.

E.g. regular regression model, i.e. identity link and normal random component,

$$Y_{it} = u_i + \alpha + \beta_1 x_{it1} + \epsilon_{it}$$

- $\epsilon_{it} \overset{\text{ind.}}{\sim} N(0, \sigma^2)$ represents random error.
- $u_i \overset{\text{ind.}}{\sim} N(0, \sigma_u^2)$ induce correlation within a cluster.
- $\{\epsilon_{it}\}$ and $\{u_i\}$ are independent.

GLMM

The covariance between two data points in the same cluster, noting that ϵ's are independent to each other and to the u_i

$$\begin{aligned}
\text{Cov}(Y_{it}, Y_{it'}) &= \text{Cov}(u_i + \alpha + \beta_1 x_{it1} + \epsilon_{it}, u_i + \alpha + \beta_1 x_{it'1} + \epsilon_{it'}) \\
&= \text{Cov}(u_i + \epsilon_{it}, u_i + \epsilon_{it'}) \\
&= \text{Cov}(u_i, u_i) && \epsilon \text{ are ind.} \\
&= V(u_i) = \sigma_u^2
\end{aligned}$$

The random intercept induces positive correlation, and the magnitude is governed by σ_u^2

Example (Depression)

Response on mental depression (normal, abnormal) measured three times (after 1, 2, and 4 weeks of treatment) with two drug treatments (standard, new) and two severity of initial diagnosis groups (mild, severe).
Is the rate of improvement better with the new drug?

	Time	Response Pattern							
	0	A	A	A	A	N	N	N	N
	1	A	A	N	N	A	A	N	N
	2	A	N	A	N	A	N	A	N
Severity	Drug								
Mild	Std	6	15	4	14	3	9	13	16
	New	0	9	2	22	0	6	0	31
Severe	Std	28	27	15	9	9	8	2	2
	New	6	32	5	31	2	5	2	7

Example (continued)

- Y_t = response of randomly selected subject at time t (1 = normal, 0 = abnormal)
- s = severity of initial diagnosis (1 = severe, 0 = mild)
- d = drug (1 = new, 0 = std)
- t = time (0, 1, 2), which is log2(weeks of trt)

$$\log\left[\frac{P(Y_t = 1)}{P(Y_t = 0)}\right] = u_i + \alpha + \beta_1 s + \beta_2 d + \beta_3 t + \beta_4 (dt)$$

Example (continued)

```
> dep.lme=glmer((response=="normal")~severity+drug*time+
+    (1|subject), data=depression, family=binomial)
> summary(dep.lme)
     AIC      BIC   logLik deviance df.resid
  1173.9   1203.5   -581.0   1161.9     1014

Random effects:
 Groups  Name        Variance Std.Dev.
 subject (Intercept) 0.003231 0.05684
Number of obs: 1020, groups:  subject, 340

Fixed effects:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.02797    0.16406  -0.170    0.865
severitysevere -1.31488    0.15261  -8.616  < 2e-16 ***
drug1          -0.05967    0.22239  -0.268    0.788
time            0.48274    0.11566   4.174 3.00e-05 ***
drug1:time      1.01817    0.19150   5.317 1.06e-07 ***
```

Example (continued)

GLMM and GEE estimates and se's for fixed effects are nearly identical:

	GLMM		GEE	
	Est	SE	Est	SE
alpha	-0.03	0.16	-0.03	0.17
beta.1	-1.31	0.15	-1.31	0.15
beta.2	-0.06	0.22	-0.06	0.23
beta.3	0.48	0.11	0.48	0.12
beta.4	1.02	0.19	1.02	0.19

- $\hat{\rho} = -0.003 \approx 0$ in GEE with exchangeable working correlation
- $\hat{\sigma} = 0.057 \approx 0$ in GLMM. According to model, 95% of all individuals will have u_i between $\pm 1.96\sigma \approx \pm 0.11$. But $e^{\pm 0.11} \rightarrow (0.89, 1.12)$, so effect of u_i on odds is estimated to be small for most subjects

Example (continued)

```
Correlation of Fixed Effects:
             (Intr) svrtys drug1  time
severitysvr -0.389
drug1       -0.614 -0.005
time        -0.673 -0.123  0.524
drug1:time   0.462 -0.121 -0.742 -0.562
```

We learned

- Mixed models contain fixed and random effects
- GLMM are subject specific, i.e. conditional by adding random intercept
- Saw an example where a standard GLM is sufficient but could not start with that assumption