Linear Mixed Models

$$Y = X\beta + Zb + \epsilon$$

where X and Z are specified design matrices, β is a vector of fixed effect coefficients, b and ϵ are random, mean zero, Gaussian if needed.

Usually think of b being constant over subjects, the ϵ as independent between subjects, possibly correlated within subjects. Let ω denote free parameters in the variance specification.

Likelihood

We observe n r.v.'s Y. Once the error structure is fully specified, and $\text{cov}(b, \epsilon) = 0$,

$$\begin{aligned} Y &\sim N(X\beta, V(\omega)) \\ V(\omega) &= \text{var}(\epsilon) + Z\text{var}(b)Z^T \end{aligned}$$

so minus twice the log-likelihood is

$$(Y - X\beta)^T V^{-1}(\omega)(Y - X\beta) + \log|V(\omega)|$$

Thus, given ω we find the MLE of β by generalized least squares.

One-way layout

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \qquad i = 1, ..., n_i$$

If we treat the $\tau_i \sim N(0, \sigma_b^2)$, we have a special case. The log-likelihood depends on β through the group means $m_i = \overline{y_{i\cdot}}$. Now

$$m_i \sim N(\mu, \sigma_b^2 + \sigma^2/n_i)$$

which suggests that we take a weighted mean of m_i with weights inversely proportional to $\text{var}(m_i)$. This is MVUE and is in fact the MLE of μ (using the special structure of V).

What if the variances are unknown? For a balanced layout the estimator does not depend on them. In general it depends on σ_b^2/σ_2.

We can find the MLEs of σ_b^2 and σ^2, but even in the balanced case they are not the traditional ones: they have no adjustment for fitting means.

REML

Restricted / residual / reduced maximum likelihood: a method of estimation in LMEs.

Suppose that we can find some linear combinations AY whose distribution does not depend on β. In fact we can find up to $n - p$ linearly independent such. One choice is any $n - p$ of the least-squares residuals of the regression of Y on X.

In REML we treat AY as the data and use maximum-likelihood estimation of ω (the parameters in V).

The REML estimates do not depend on the choice of A, so this procedure is not as arbitrary as it sounds. Indeed, the REML estimates minimize

$$\begin{aligned} &(Y - X\beta)^T V^{-1}(\omega)(Y - X\beta) + \log|V(\omega)| \\ &\quad + \log|X^T V^{-1}(\omega)X| \end{aligned}$$

Clearly the REML estimator of β is still GLS, plugging in the REML estimate of ω: *slightly* simpler to compute than MLEs.

Another perspective

The REML fit criterion is the marginal likelihood, integrating β out with a vague prior.

Relationship to classical ideas

In balanced designs REML gives the classical moment estimates of variance components (constrained to be non-negative).

Consider a paired comparison: REML will give the paired t–test analysis, ML will get the variance consistently low (by a factor of a half).

Drawbacks

No equivalents of likelihood-ratio tests (REMLs on models with different fixed effects are not comparable).

May be able to use Wald-like tests of extra parameters, but relevant asymptotic theory is hard to find.

Usual to quote GLS-based variances $X^T V^{-1}(\hat{\omega})X$ for $\hat{\beta}$ in both ML and REML procedures.

BLUPs

Best linear unbiased predictions. In an LME it is not clear what fitted values and hence residuals are. Our best prediction for subject i is *not* given by the mean relationship. We need to specify just what is common with an example we have already seen.

BLUPs replace the random effects b by their conditional means $\hat{b}$ given the data, and then make predictions using those values,

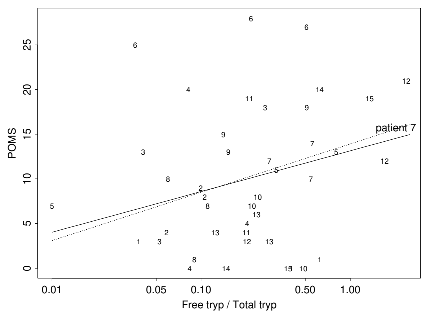
$$\hat{Y} = X\hat{\beta} + Z\hat{b}$$

Since everything is Gaussian, these are linear functions of the data, and as everything is linear, they are unbiased. They have minimum variance amongst such estimators.

Obviously if we have a new subject, $\hat{b} = 0$, and similarly in multilevel models. Therefore find several (in general) fitted values and several residuals.

Effects of Free Trytophan

James McGuire measured mood (POMS score) and abundance of free trytophan in the blood for 15 post-operative patients.



Classical model is

$$y_{ij} = \mu + \alpha_i + \beta x_{ij} + \epsilon_{ij}, \qquad \epsilon \sim N(0, \sigma^2)$$

a parallel line for each patient.

LME is

$$y_{ij} = \mu + \eta_i + \beta x_{ij} + \epsilon_{ij}, \qquad \eta \sim N(0, \sigma^2_\eta)$$

We could also consider a random effect for slope. This is hopeless in the classical case. The LME becomes LME is

$$y_{ij} = \mu + \eta_i + (\beta + \zeta) x_{ij} + \epsilon_{ij}, \qquad \zeta \sim N(0, \sigma^2_\zeta)$$

where either we allow η and ζ to be correlated or we centre x carefully. The estimate of β and its estimated s.e. are almost unchanged. The BLUPs for each patient are very different from the classical fits.

Non-linear Mixed-Effect Models

$$Y_{ij} = f(x_{ij}; \beta, \eta_i) + \epsilon_{ij}$$

will be general enough for our discussion. What we usually assume is that

$$Y_{ij} = f(x_{ij}; \beta + \eta_i) + \epsilon_{ij}$$

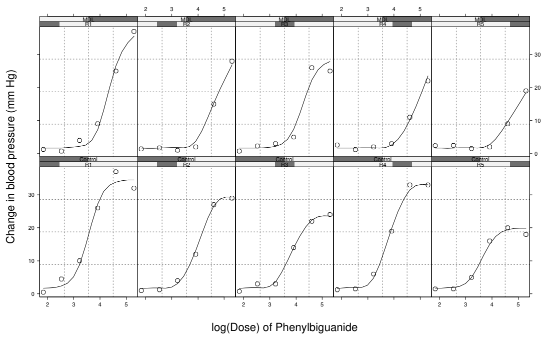
where some components of η_i may always be zero. (Only α and θ have random effects in the next example.)

What is the likelihood? Only rarely can we integrate over (η_i). So 'MLEs' of NLMEs are based on approximations.

Blood Pressure in Rabbits

Five rabbits were studied on two occasions, after treatment with saline (control) and after treatment with the 5-HT_3 antagonist MDL 72222. After each treatment ascending doses of phenylbiguanide (PBG) were injected intravenously at 10 minute intervals and the responses of mean blood pressure measured. The goal was to test whether the cardiogenic chemoreflex elicited by PBG depends on the activation of 5-HT_3 receptors.

The response is the *change* in blood pressure relative to the start of the experiment.



$$f(x;\alpha,\beta,\lambda,\theta) = \alpha + \frac{\beta-\alpha}{1+\exp[(x-\lambda)/\theta]}$$

Fitting NLMEs

1. Fit a non-linear regression to each subject, and treat the parameter values as the data at subject level. If there is within-subject correlation, pool estimates of correlation parameters across subjects.
2. Use a Taylor-series expansion about the mean effects. This gives an LME which we can fit. Repeatedly expand about the fixed effects, that is write

$$Y_{ij} = f(x_{ij};\hat\beta^0,0) + X(\beta-\hat\beta^0) + Z\eta_i + \epsilon_{ij}$$

3. Use a Taylor-series expansion about estimates of (η_i):

$$Y_{ij} = f(x_{ij};\hat\beta^0,\hat\eta^0) + X(\beta-\hat\beta^0) + Z(\eta_i-\hat\eta^0) + \epsilon_{ij}$$

Lindstrom–Bates fit by simultaneously minimizing over (β,η_i); this effectively uses the BLUPs in the local linearization.

Inference in NLMEs

A problem! We have no likelihood to compare, and the `nlme` software appears to quote the likelihood of the final linearization.

We can use the estimated variance of the parameters and Wald–like tests.

Rabbits

Note that there are three strata of variation:

1. Animals
2. Occasions within animals
3. Measurements on the animal/occasion combination.

and the effect of interest, the treatment, varies in the second stratum.

We start by fitting separate models for each treatment:

```
Control:
  Log-likelihood: -66.502
  Fixed: list(A ~ 1, B ~ 1, ld50 ~ 1, th ~ 1)
      A      B   ld50      th
 28.332 1.5134 3.7744 0.28957

Random effects:
 Formula: list(A ~ 1, ld50 ~ 1)
 Structure: General positive-definite
          StdDev  Corr
       A 5.76889 A
    ld50 0.17953 0.112
Residual 1.36735

Treatment:
  Log-likelihood: -65.422
  Fixed: list(A ~ 1, B ~ 1, ld50 ~ 1, th ~ 1)
      A      B   ld50      th
 27.521 1.7839 4.5257 0.24236

Random effects:
 Formula: list(A ~ 1, ld50 ~ 1)
 Structure: General positive-definite
         StdDev   Corr
       A 5.36549 A
    ld50 0.18999 -0.594
Residual 1.44172
```

Now a combined model

```
R.nlme1 <-
  nlme(BPchange ~ Fpl(Dose, A, B, ld50, th),
    fixed = list(A ~ Treatment,
                 B ~ Treatment,
                 ld50 ~ Treatment,
                 th ~ Treatment),
    random =  A + ld50 ~ 1 | Animal/Run,
    data = Rabbit, ...)
```

```
Random effects:
 Formula: list(A ~ 1, ld50 ~ 1)
 Level: Animal
 Structure: General positive-definite

                 StdDev   Corr
   A.(Intercept) 4.6063 A.(Int
ld50.(Intercept) 0.0626 -0.166

 Formula: list(A ~ 1, ld50 ~ 1)
 Level: Run %in% Animal
 Structure: General positive-definite
                 StdDev   Corr
   A.(Intercept) 3.2489 A.(Int
ld50.(Intercept) 0.1707 -0.348
        Residual 1.4113
```

```
Fixed effects:
                   Value Std.Error t-value p-value
   A.(Intercept)  28.326    2.7802  10.188  <.0001
     A.Treatment  -0.727    2.5184  -0.288  0.7744
   B.(Intercept)   1.525    0.5155   2.958  0.0050
     B.Treatment   0.261    0.6460   0.405  0.6877
ld50.(Intercept)   3.778    0.0955  39.579  <.0001
  ld50.Treatment   0.747    0.1286   5.809  <.0001
  th.(Intercept)   0.290    0.0323   8.957  <.0001
    th.Treatment  -0.047    0.0459  -1.020  0.3135
```

This suggests that the only difference by treatment is to shift the mean curve along (λ varies by treatment).

```
                  Value Std.Error t-value p-value
               A 28.170    2.4909  11.309  <.0001
               B  1.667    0.3069   5.433  <.0001
ld50.(Intercept)  3.779    0.0921  41.036  <.0001
  ld50.Treatment  0.759    0.1217   6.233  <.0001
              th  0.271    0.0226  11.964  <.0001
```

Bayesian Analysis

In the linear case, both β and b_i are regarded as random variables, and we have a hierarchical linear model. Relatively little to say: Bayesians just need to find the posterior distributions of the quantities of interest (still β and perhaps ω).

For a long time the issues were computational, but the re-discovery of MCMC and the Gibbs sampler has made even mainstream Bayesians realize that there are relatively simple ways to do this. (Given ω, everything is joint Gaussian, so empirical Bayes procedures were popular.)

The only issues involve ω. For simple specifications (a few variances or general covariance matrices) with conjugate priors finding the full conditionals is straightforward. For parameters such as the correlation in an AR(1) process, need to use numerical simulation techniques, using a profile of the joint density.

In the non-linear case the issues are a little harder. Consider

$$Y_{ij} = f(x_{ij}; \beta + \eta_i) + \epsilon_{ij}$$

and let $\beta_i = \beta + \eta$. Consider the r.v.'s $(Y, \beta, \beta_i, \omega)$. We can easily simulate

$$\begin{array}{rl} Y \mid \beta, (\beta_i), \omega & \text{the model} \\ \beta \mid Y, (\beta_i), \omega & \text{normal} \end{array}$$

For

$$\begin{array}{l} \omega \mid Y, \beta, (\beta_i) \\ \beta_i \mid Y, \beta, (\beta_j, j \neq i), \omega \end{array}$$

we have a joint density and can use numerical simulation procedures.

Generalized Linear Mixed Models

Suppose we have a binomial or Poisson response. We can apply the same ideas, with linear predictor

$$\eta = X\beta + Zb$$

and distribution of Y_i depending on η_i through the link function.

Note that unless we have a Gaussian GLM with identity link, the marginal distribution of Y_i is not binomial, Poisson etc; the (Y_i) are always dependent (and usually positively correlated in clusters).

This is known as a *subject-specific* model. The alternative is a *marginal* or *population-averaged* model where the marginal distribution of the Y_i is binomial, Poisson, etc, but they are correlated in clusters.

Logistic GLMM

Simplest case, a random-intercept model:

$$Y_{ij} \sim \text{bin}(n_{ij}, p_{ij}), \quad \operatorname{logit} p_{ij} = b_i + (X\beta)_{ij}$$

Here i labels the cluster.

Methods:

- Conditional analysis, conditional on $\sum_j y_{ij}$, which eliminates the random intercept.
- Approximate MLEs based on Laplace expansion.
- Approximate MLEs based on numerical integration (and need to estimate the variance of b_i).
- Bayesian analysis by Gibbs sampler.

Note that this is one case where the marginal distributions may be approximately binomial (Bernoulli) but correlated, and with a different regression on the covariates.

Marginal Models

Suppose we have several observations Y_{ij} on each cluster i. We allow the mean μ_{ij} of Y_{ij} to depend on η_{ij} for a linear predictor $\eta = X\beta$, the variance of Y_{ij} to depend on its mean (and possibly a dispersion parameter ϕ). Observations on different clusters are independent, but $(Y_{i\cdot})$ are dependent, with a correlation matrix depending on parameters ω.

Apart from the dependence, this is how we model a GLM.

Identity link

Suppose $\mu_{ij} = \eta_{ij}$, and we fit β by GLS with weight matrix W,

$$\hat\beta_W = (X^T W X)^{-1} X^T W Y$$

Then asymptotically $\hat\beta_W$ is unbiased and normal with variance matrix

$$\Sigma_W = \left[(X^T W X)^{-1} X^T W\right] \text{var}(Y) \left[W X (X^T W X)^{-1}\right]$$

- We may be able to estimate var(Y) some other way (REML from a saturated model?)
- All we lose by not having the correct weights W is efficiency.

General link

Still use GLS, ignore the dependence of var$(Y_{i\cdot})$ on β:

$$\sum_{\text{clusters } i} \frac{\partial \mu_{ij}}{\partial \beta} \, \text{var}(Y_{i\cdot})^{-1} \left[Y_{i\cdot} - \mu_{i\cdot}\right] = 0$$

These are the GLM score equations, except for the correlations, which need to estimate simultaneously.

This approach (including equations for α) is known as GEE, *Generalized Estimating Equations*. It has asymptotic theory that shows consistency, asymptotic normality with estimable variance matrix.