Mixed Models

Bret Larget

Departments of Botany and of Statistics
University of Wisconsin—Madison

March 15, 2007

The Big Picture

• Mixed models contain a combination of fixed and random effects.
• It is important to properly identify which variables should be modeled as random and which as fixed.
• In the case of completely balanced designs, ANOVA tables can be used for inference with $F$ tests, but care must be taken to compute p-values properly.
• Blocking variables are a special type of categorical variable.
• Blocks can be random or fixed depending on the context.
Parameter Estimation

- Recall the Antigua Corn data from last time.
- The model had three parameters, a mean harvest $\hat{\mu} = 4.2917$, a site variance $\hat{\sigma}_a^2 = 2.3677$, and an individual variance $\hat{\sigma}_e^2 = 0.5775$.
- The function `mcmcsamp` provides a means to estimate uncertainty in parameter estimates in mixed models by producing a sample of the parameter values from a Bayesian posterior distribution.
- The 0.025 and 0.975 quantiles of the sample are estimates of a 95% credible region which you can interpret as a confidence interval.
- The function estimates logarithms of the variance components, so we need to take exponentials to put them back on scale.
- Credible regions need not be symmetric — for variance components they usually are not.

MCMC Sample Example

```r
> library(DAAG)
> library(lme4)
> corn1.lmer = lmer(harvwt ~ 1 + (1 | site), data = ant111b)
> set.seed(3425)
> corn1.samp = mcmcsamp(corn1.lmer, n = 1000)
> apply(corn1.samp, 2, function(x) quantile(x, probs = c(0.025, 0.975)))
   (Intercept) log(sigma^2) log(site.(In))
2.5%  2.875464 -1.05292374 -0.01928132
97.5%  5.679904  0.06247114  2.37010691
> apply(corn1.samp, 2, function(x) quantile(exp(x), probs = c(0.025, 0.975)))
   (Intercept) log(sigma^2) log(site.(In))
2.5%  17.73366  0.3489163  0.9809044
97.5% 292.92262  1.0644637 10.6986503
```

- `set.seed` is unnecessary, but makes the sample repeatable.
- The result of `mcmcsamp` is a matrix with 1000 rows and three columns.
- The `apply` function applies to the columns (2nd argument 2) of `corn1.samp` (1st argument) the function (3rd argument) which find quantiles of either $x$ or $exp(x)$ in this case.
Credible Intervals

- The estimated intercept is $\hat{\mu} = 4.29$, $P\{2.88 < \mu < 5.68\} = 0.95$.
- The estimated site variance is $\hat{\sigma}_a^2 = 2.37$, $P\{0.98 < \sigma_a^2 < 10.7\} = 0.95$.
- The estimated individual variance is $\hat{\sigma}_e^2 = 0.58$, $P\{0.35 < \sigma_e^2 < 1.06\} = 0.95$.

Soil Moisture

- In an experiment, four different irrigation methods are compared.
- There are 16 different plots in the experiment, considered to be a sample from a larger population of plots.
- Each irrigation treatment is used on four plots, assigned at random.
- Each plot is subsampled — three soil moisture measurements are taken within each plot.
- There is one fixed effect — *irrigation treatment*.
- There are two sources of variation — *plot-to-plot* variability and *within-plot* variability.
Model

- A model is

\[ y_{ijk} = \mu + \alpha_i + b_{ij} + e_{ijk} \]

where:
- \( \mu \) is an overall mean,
- \( \alpha_i \) is the effect of treatment \( i \) where \( \sum_i \alpha_i = 0 \), for \( i = 1, \ldots, 4 \),
- \( b_{ij} \sim iid \, N(0, \sigma_b^2) \) is the effect of the \( j \)th plot (nested within the \( i \)th treatment), \( j = 1, \ldots, 4 \),
- and \( e_{ijk} \sim iid \, N(0, \sigma_e^2) \) is individual random error, \( k = 1, \ldots, 3 \).

Plot of Data

- Notice that the variation within treatments is substantial, but the treatment distributions appear to be different.
- Variation within treatments looks to be reasonably symmetric without outliers.
Examples

Soil Moisture

Fitting a Random Effects Model

```r
> library(lme4)
> soil.lmer = lmer(moisture ~ irrigation + (1 | irrigation:plot))
> summary(soil.lmer)
```

Linear mixed-effects model fit by REML
Formula: moisture ~ irrigation + (1 | irrigation:plot)

AIC   BIC logLik MLdeviance REMLdeviance
83.5 92.86 -36.75 72.22 73.5

Random effects:
Groups   Name         Variance Std.Dev.
         irrigation:plot (Intercept) 0.44020  0.66347
         Residual                0.12813  0.35795
Number of obs: 48, groups: irrigation:plot, 16

Fixed effects:
   Estimate Std. Error t value
(Intercept) 12.0750     0.3475   34.75
irrigationB  0.7417     0.4914    1.51
irrigationC -0.9750     0.4914   -1.98
irrigationD  1.0583     0.4914    2.15

Correlation of Fixed Effects:
   (Intr) irrigB irrigC
irrigationB -0.707
irrigationC -0.707  0.500
irrigationD -0.707  0.500  0.500

irrigation is the fixed effect.

irrigation:plot is an interaction between irrigation and plot, which is how we specify nesting here.

There is no main plot effect, so plot does not appear without irrigation in the model.

The term (1 | irrigation:plot) means that there is a random effect for each plot and this effect is nested within the intercept (the whole model).

There are two sources of random variation, one for plot and one for individual measurements.

Residual Plot

```r
> plot(fitted(soil.lmer), residuals(soil.lmer))
> abline(h = 0)
```

There are no bad patterns in the residual plot.
**Continue Analysis in R**

- Draw a nesting diagram for the variables in the model.
- Use R to show alternative analysis using fixed effects.
- Draw connection to the correct ANOVA table calculation.

---

**Nesting**

- A categorical variable $A$ is *nested* in a categorical variable $B$ if all of the observations within a single level of $A$ are also within a single level of $B$.
- Irrigation method is nested within the grand mean.
- Plot is nested within irrigation method.
- Individual error is nested within each other variable.
For balanced designs, an appropriate $F$-test uses for its denominator the MS from the closest nested random effect.

The correct p-value to test irrigation is only 0.0064, not $10^{-14}$.

```r
> soil.lm = lm(moisture ~ irrigation + irrigation:plot)
> anova(soil.lm)

Analysis of Variance Table

Response: moisture
df Sum Sq Mean Sq F value   Pr(>F)
irrigation  3 29.4073 9.8024 76.507 1.097e-14 ***
irrigation:plot 12 17.3858 1.4488 11.308 2.342e-08 ***
Residuals  32  4.1000 0.1281
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(soil.lmer)

Analysis of Variance Table

Df Sum Sq Mean Sq
irrigation  3 2.60086 0.86695
---

> c(9.8024/1.4488, 0.86695/0.1281)

> 1 - pf(6.76, 3, 12)
[1] 0.006385453
```