Mixed-effects modeling

Colin Wilson

Phonetics Seminar, May 21, 2007

Main sources

- Baayen (2004) Statistics in Psycholinguistics: A critique of some gold standards. *Mental Lexicon Working Papers I*.
- Baayen (to appear) *Analyzing Linguistic Data: A practical approach to statistics*. Cambridge University Press.
- Baayen, Davidson and Bates (2006) Mixed-effects modeling with crossed random effects for subjects and items. Submitted.

Problem: crossed random effects

In many types of linguistic experiments (artificial grammar learning, lexical decision, phonetic production and perception, . . .) participants and materials are **random** and **crossed** effects.

(Clark. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *JVLVB* 12, 335-359)

Random effects

- A **random effect** is a factor whose levels in the experiment are nonexhaustively sampled from a larger population of interest.
- The analysis must take into account the fact that we wish to generalize our results beyond the particular sample to the population.*

*This does not imply that the results necessarily or plausibly hold of every member of the population. To take a typical case, the mean μ of a normal distrib. is a population parameter, but $\Pr(x = \mu) = 0$.

Random effects (Baayen et al. 2006)

‘the interest of most studies is not about experimental effects present only in the individuals who participated in the experiment, but rather in effects present in speakers everywhere’ (2)

‘most materials in a single experiment do not exhaust all possible syllables, words, or sentences that could be found in a given language’ (2)

‘Any naturalistic stimulus which is a member of a population of stimuli which has not been exhaustively sampled should be considered a random variable for the purposes of an experiment’ (31).

Random effects (Baayen et al. 2006)

‘The current practice of psychophysiologists and neuorimaging researchers typically ignores the issue of whether linguistic materials should be modeled with fixed or random effect models’ (30).

‘Individual subjects and items may have intercepts and slopes that diverge considerably from the population means’ (24).

‘we know that no two brains are the same’ (25)

Crossed effects

- Two effects are **crossed** in an experiment if every level of one effect co-occurs with every level of the other effect.
- Counterbalancing, while important, does not eliminate the crossing of participants and materials — typically, it *requires* crossing.

  "Each subject saw each item in exactly one condition. The number of items in each condition was the same for each subject, and each item occurred in each condition the same number of times across subjects."

Problem: crossed random effects

In many types of linguistic experiments (artificial grammar learning, lexical decision, phonetic production and perception, . . .) participants and materials are **random** and **crossed** effects.

(Clark. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *JVLVB* 12, 335-359)

Standard solution (Baayen et al. 2006)

'Clarks' oft-cited paper presented a technical solution to this modeling problem, based on statistical theory and computational methods available at the time (e.g., Winer, 1971). This solution involved computing a quasi-F statistic which, in the simplest-to-use form, could be approximated by the use of a combined minimum-F statistic derived from separate participants (F1) and items (F2) analyses' (2).

Critique of the standard (Baayen et al. 2006)

- Deficient statistical power (see §4 and Baayen 2004: §3.2.3 for simulation results)
- Inability to handle missing data
- Different methods for handling continuous and categorical data (see Baayen 2004: §4)
- ‘unprincipled methods of modeling heteroskedasticity and non-spherical error variance’ (3)

Critique of the standard (Baayen 2004)

- Cost of dichotomization

  'It is widely believed that [dichotimization] is the most powerful means of ascertaining the independent effect of variables such as frequency of occurrence that are correlated with many other potentially relevant predictors.

  Unfortunately, this belief is incorrect' (2).

  (references: Cohen 1983, Harrell 2001)

Also: Arbitrarity of cutoff points for low and high conditions (e.g., what counts as 'hard' or 'frequent'?).

Critique of the standard (Baayen 2004)

- Cost of prior averaging

  'The by-subject and by-item analyses that are currently the norm in psycholinguistic studies also bring along systematic data loss. It is widely believed that these averaging techniques are the best that current statistics has to offer' (8).

[This seems to ignore one benefit of averaging, namely reduction of the error of the estimates.]

Critique of the standard (Baayen 2004)

- Summary

  'Factorial designs are commonly used where regression is more appropriate. Dichotomization and factorization of numerical predictors, although widely practised, lead to a loss of power and should be avoided. Psycholinguists are generally very reluctant to include covariates in their analyses, even though including relevant covariates is part and parcel of statistical common sense' (37).

The future is now (Baayen 2004)

‘as anyone following statistical developments outside the field of psycholinguistics (for instance, in *Psychological Methods* or in *Behavioral Research Methods, Instruments and Computers*, or in Venables & Ripley, 2003) will have realized, current statistics has a lot more to offer, both in power and in the insight provided into the quantitative structure of the data’ (38).

The future is now (Baayen et al. 2006)

‘In the 30+ years since [Clark 1973], statistical techniques have expanded the space of possible solutions to this problem, but these techniques have not yet been applied widely in the field of language and memory studies’ (2)

‘we introduce a very recent development in computational statistics, namely, the possibility to include subject and items as crossed random effects, as opposed to hierarchical or multilevel models in which random effects must be assumed to be nested’ (2).
(see Bates & Penheiro, 1998; Pinheiro & Bates 2000)

Proposal: Mixed-effect models

- Modern mixed-effect modeling
  - allows fixed and random effects to be combined (i.e., ‘mixed’)
  - allows random effects to be crossed (a result of the ‘recent developments’)
  - allows covariates to be included in the model (e.g., trial number)
  - is a form of regression (and so does not require dichotomization or aggregation)
  - generalizes to categorical responses

Hypothetical example (Baayen et al. 2006)

Three participants (s1, s2, s3) responded to three items (w1, w2, w3) in a primed lexical decision task under both short and long SOA [stimulus onset asynchrony]

- Two random effects (crossed)
  - participants
  - items
- One fixed effect (crossed with rand. effects)
  - SOA (short vs. long)

Hypothetical data

Subject	Item	SOA	RT
s1	w1	long	466
s1	w2	long	520
s1	w3	long	502
s1	w1	short	475
s1	w2	short	494
s1	w3	short	490
s2	w1	long	516

...

Mixed-effect formula

$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \epsilon$, $\epsilon \sim N(0, \sigma^2\mathbf{I})$, $\mathbf{b} \sim N(0, \sigma^2\mathbf{\Sigma})$, ϵ and $\mathbf{b}$ are independent random variables

where

$\mathbf{y}$	is the vector of dependent values (RTs)
$\mathbf{X}$	is the fixed-effect design matrix
β	is the vector of fixed-effect coefficients
$\mathbf{Z}$	is the random-effect design matrix
$\mathbf{b}$	is the vector of adjustments for subjects, items
ϵ	is the vector of residual errors (subjects $\times$ items

Mixed effect formula (one subject-item pair)

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\beta + \mathbf{Z}_{ij}\mathbf{b}_{ij} + \epsilon_{ij}$$

where i is a subject and j is an item.

Suppose that for all i and j in the population:

$$\mathbf{X}_{ij} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \beta = \begin{pmatrix} 522.2 \\ -19.10 \end{pmatrix} \begin{matrix} intercept \\ SOAshort \end{matrix}$$

and suppose that for subject 1 and item 1:

$$\mathbf{Z}_{11} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \mathbf{b}_{11} = \begin{pmatrix} -26.2 \\ 11.0 \\ -28.3 \end{pmatrix} \begin{matrix} s1\ intercept \\ s1\ \&\ SOAshort \\ w1\ intercept \end{matrix}$$

Mixed effect formula (one subject-item pair)

$$\mathbf{y}_{11} = \mathbf{X}_{11}\beta + \mathbf{Z}_{11}\mathbf{b}_{11} + \epsilon_{11}$$

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 522.2 \\ -19.10 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -26.2 \\ 11.0 \\ -28.3 \end{pmatrix} + \epsilon_{11}$$

$$= \begin{pmatrix} 522.2 \\ 503.1 \end{pmatrix} + \begin{pmatrix} -54.5 \\ -43.5 \end{pmatrix} + \epsilon_{11}$$

$$= \begin{pmatrix} 467.7 \\ 459.6 \end{pmatrix} + \epsilon_{11}$$

by comparison with RT data, $\epsilon_{11} \approx \begin{pmatrix} -2 \\ 15 \end{pmatrix}$

Mixed-model analysis

- The goal of statistical analysis is to provide estimates of the population parameters and measures of the reliability of the estimates.
- The analysis allows us to test whether items contribute to responses independently (assumed by crossing) or only via interaction with subjects (the nested alternative).
- Despite the inclusion of both random effects, the analysis does not estimate a separate free parameter for each subject and item (or combination) (see Baayen et al. 2006:7).

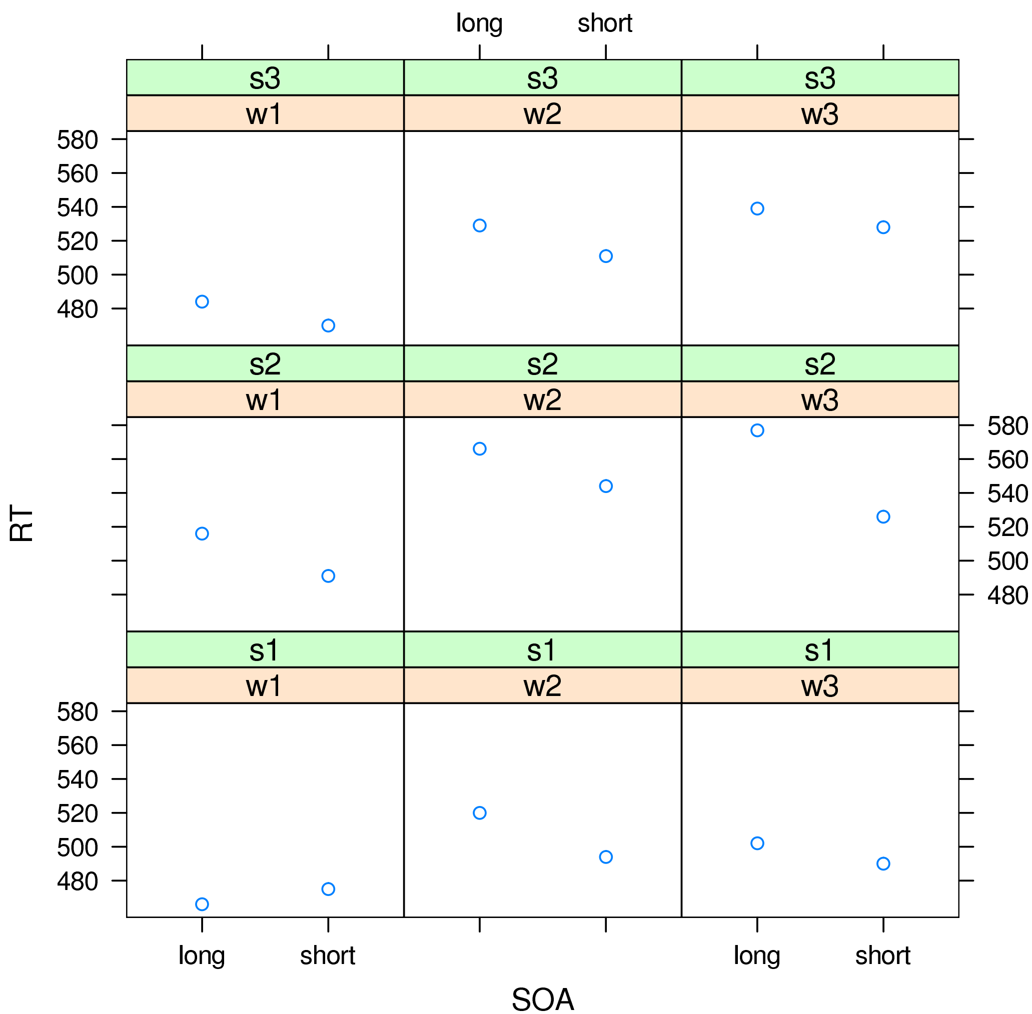
R (http://www.r-project.org/)

‘To our knowledge, the only software currently available for fitting mixed-effects models with crossed random effects is the lme4 package (Bates, 2005; Bates & Sarkar, 2005) in R, an open-source language and environment for statistical computing (R development core team, 2005) . . . . In statistical computing, R is the leading platform for research and development, which explains why mixed-effects models with crossed random effects are not (yet) available in commercial software packages’ (8).

Analysis in R with lme4

```
library(grid) # plotting
library(lme4) # analysis
d <- read.table(".../BaayenDavidsonBates.data"
    + header=T)
xyplot(RT ˜ SOA | Item + Subject, data = d)
fit1 <- lmer(RT ˜ SOA + (1 | Item) +
        + (1 | Subject), data = d, method="ML"
```

Graph of data



Result of fitting

```
Linear mixed-effects model fit by maximum like
Formula: RT ~ SOA + (1 | Item) + (1 | Subject)
   Data: d
   AIC   BIC logLik MLdeviance REMLdeviance
 162.3 165.9 -77.16      154.3        141.5
Random effects:
 Groups   Name        Variance Std.Dev.
 Item     (Intercept) 473.18   21.753
 Subject  (Intercept) 401.40   20.035
 Residual             127.01   11.270
number of obs: 18, groups: Item, 3; Subject, 3
```

Fixed effects:

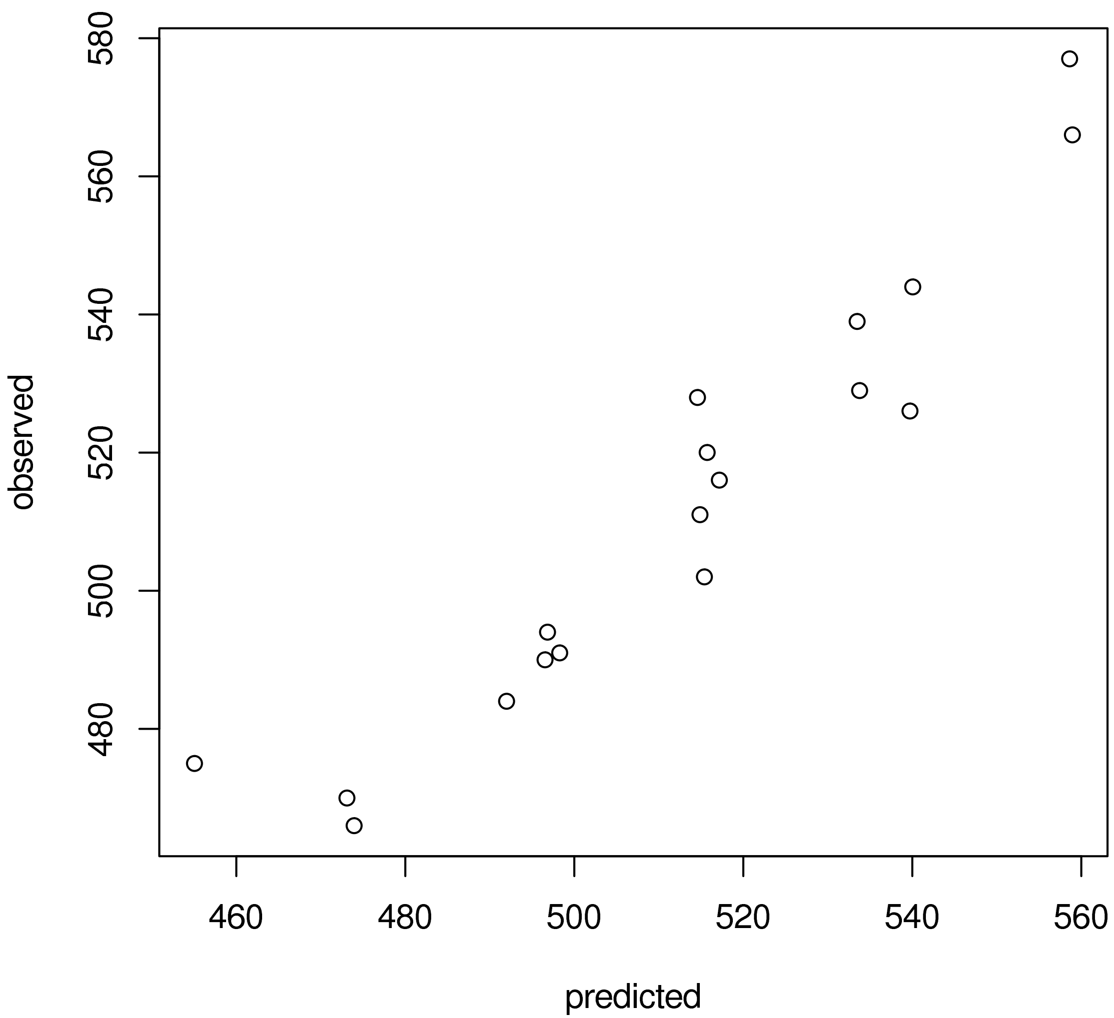
Result of fitting (cont.)

```
Fixed effects:
            Estimate Std. Error t value
(Intercept)  522.111     17.483  29.865
SOAshort     -18.889      5.313  -3.555

Correlation of Fixed Effects:
         (Intr)
SOAshort -0.152
```

[Note that the estimates of the fixed effects are quite close to the hypothetical parameter values a few slides back.]

Regression plot ($r^2 = .90$)



In-class excercise

How would we obtain the *predicted* response for subject 1 on item 1 (both long and short SOA) given the fitted model above?

[This is analogous to the calculation we did with the hypothetical population estimates. It's *just* a matter of pulling the right numbers out of R . . . .]

Assessing the model fit

- Note that t-values are given without ps, because determination of degrees of freedom is difficult (Bates, R-News). If $t > 2$, should be significant if the data set is ‘large’. Alternatively, use MCMC sampling methods (Baayen et al. 2006:11ff., Baayen to appear).
- The log likelihood (logLik) of the data gives a measure of how well the fitted model matches the data (MLdeviance = -2*logLik).

Assessing the model fit

- ANOVA can be used to compare different models (Raudenbusch & Bryk 2002: 60-61) based on log likelihood / deviance.

```
fit2 <- lmer(RT ~ SOA +
        + (1 + Item | Subject),
        + data = d, method="ML")
anova(fit1, fit2) # Chi-Sq(4) = 3.0176, p<.6
# improvement in log likelihood
# (-77.16 vs. -75.56) is not
# sufficient to justify nesting
```

Summary

- The problem addressed was estimating population parameters in the case of crossed random factors, typical in psycholinguistics.
- Modern mixed-effect modeling allows crossed random factors, is based on well-known statistical objectives (e.g., maximum likelihood), and provides alternatives to standard statistical tests.

What was the crucial advance?

- Classic statistical methods allow key values to be calculate *exactly*, in closed-form (e.g., partitioning the variance and computing F).
- Modern optimization methods such as Expectation Maximization and sampling techniques allow quantities that cannot be computed in closed form to be obtained or approximated to within any desired precision.
- These methods allow models with crossed random factors, for which no closed-form solution exists, to be fitted to the data.

Modeling timecourse (Baayen et al. 2006)

‘An important new possibility offered by mixed-effect modeling is to bring effects that unfold during the course of an experiment into account. There are several kinds of longitudinal effects. First, there are effects of learning or fatigue. . . . . Second, in chronometric paradigms, the response to a target trial is heavily influenced by how the preceding trials were processed. In lexical decision, for instance, the reaction time to the preceding word in the experiment is one of the best predictors for the target latency . . . . Third, qualitative properties of preceding trials should be brought under statistical control’ (25-26).