15th class: Generalized linear mixed models

Monday, November 26, 2007
2:22 PM

We'll do a little bit of the Baayen et al paper, but not a lot (Chris finds parts of it confusing, too, don't worry).

Baayen, R.H., Davidson, D.J. and Bates, D.M. (submitted). [Mixed-effects modeling with crossed random effects for subjects and items](). MS, 2007.

Why used mixed effect models?

a. Language isn't a fixed effect (Clark 1973, especially the beginning before the heavy stats, though Clark does a lot less of the heavy stats now). The early research didn't include the items at all. You can prove some results, but they are only valid for the exact items you tested. But unless you do more, you have no evidence that the result extends to general language use.
    i. Clark's nice example is about people who want to study whether nouns or verbs are read more slowly. We'll suppose that reading times vary around 500 ms. Each of two researchers tests 10 nouns and 10 verbs. You can well imagine how to researchers could both get results that for their experimental items that were contradictory.
    ii. What is a random effect variable? Everything where you haven't seen the entire population--any variable where there's some complete population. If you've seen every variable, then a fixed effect is okay. Otherwise, you've just sampled values and need to use random effects.
    iii. Corollary: People don't actually pull things out by random, they have sampling biases. These can invalidate your results. Joan talks about how if you look at a lot of acceptability judgments people put in papers is really just a sampling bias. To make the difference more extreme, you pull in animacy, pronominality, etc, which exaggerate the size of the effect.

b. You can't ignore random effects (as is traditional in psycholinguistics).

c. You can't make random effects into fixed effects. Random effects can constrain the amount that individuals can vary, but fixed effects can suck up all your explanation (everything varies just by the state the person's from and nothing else!=a sample of 100, 2 from each state, with state incorrectly treated as a fixed effect).

d. (See Jaeger on categorical variables and mixed effect models.)

Baayen et al summary

a. Baayen et al paper is just saying that even if you just want a plain linear model, you still want to use generalized linear mixed models.
    i. Summary from CM: "People who know tons more about stats think this is better to use than ANOVA." (The Clark argument you should understand, the Florian argument you should understand, but everything in Baayen you can just trust.)
    ii. GLMM's can handle missing data.
    iii. ANOVA's assume that all random effects are necessary, but GLMM's allow you to test if they're necessary.
    iv. SOA is, btw, the time between two things. Stimulus Onset Asynchrony.
    v. Baayen's text book has lots of examples but no real explanation behind what's going on. This paper has all explanation but not many examples. What about in between?
    vi. There's a standard linear model and for the random effects: the way the models are constructed is that when you're fitting a GLMM, the only thing you're fitting for random effects is how much variance there is. You aren't fitting a parameter for all 50 subjects, just how much variance there is. (The BLUPs tell you about the individual adjustments, but that's not in the model.) This is parsimonious because it says only "how much spread am I getting for these individuals".
    vii. T-tests look at how much distance there is between means compared to how many data points support each mean and how much spread there is. These turn into z-tests. If you have a small amount of data points, it's complicated to

Exploring data

Subset is a useful command in R.

CM is ambivalent about how useful mosaic plots are.

Xtabs are helpful for getting started. Xtabs take a model formula. If you want to stick in a complex expression, you have to quote it by putting it inside a capital I:

- Xtabs(~ RealizationOfRecipient + I(Verb=="give"))

In making a GLMM, you end up doing (1|Speaker) + (1|Verb) to model the random effects based on the intercept (the 1). You could also put some slope variables in here, like length of recipient could depend on the speaker ID. In practice, this overparameterizes the model and is unnecessary. In practice, people avoid this.

print(dativ.glmm1, corr=F) helps keep the output shorter. You probably won't bother at your console.

Laplace is the default algorithm and the one that's generally recommended. It shows you that you want the log likelihood to be a large negative number, but you want the deviance to be small (it shows you how much different you are from a model that has a parameter for every data point).

CM doesn't fully understand "Estimated scale (compared to 1)". Here the value is .77, which suggests that we're underdispersed compared to the assumptions of normal distribution, but it's still big enough we don't really have to worry.

The best model doesn't attribute anything to the speaker (variance: 5.000e-10), but it puts a lot on the verb (variance: 4.211e+00). CM takes the speaker variance as evidence that he doesn't have to prove it's statistically insignificant. You don't need to worry about subject as a random effect. You can leave it out. Once you do that, estimating models is 20x's faster!

But not having verb as a random effect would be a big mistake.

Note that this is a logistic regression model. The same lmer() command can do both logistic or linear. For logistic you say "family="binomial"". If you don't say that you get a linear mixed model.

(In the design package, you use ols for a plain linear model, lrm for a logistic model.)

Note that lmer() requires at least one random effect. If you don't have any random effects, you move to lrm() in the design package. If you chop out verb as a random effect, you start seeing, for example, a bunch of things being significant that weren't before. You want to get that random effect for verb because you otherwise have to count on a bunch of other factors to capture a part of what the verb random effect could do on its own.

get all the degrees of freedom. But in practice, don't worry about it, you have hundreds of observations so just use a z-test and don't sweat it.

viii. How good is the fit and how much are you modeling? The bottom corner of the middle model is about deleting the random effect from the model. It's parsimonious because it has the least model complexity (the random effect is just gone, the middle model). As you attribute more of what's going on to the random effect, then your model is scored as more and more complex. On the other hand, by attributing stuff to random effects of participants, your ability to get the best coverage (far right). The picture on the left is the joint optimization of both. I want the highest likelihood combining both scores. You make some use of random effects but "not too much".

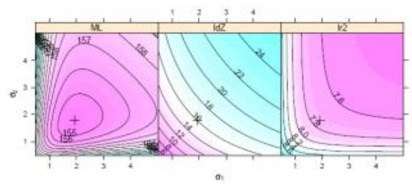

Figure 2: Contours of the profiled deviance as a function of the relative standard deviations of the item random effects and the subject random effects. The leftmost

ix. ANOVA models are really complicated: every different experimental design has a different ANOVA design. GLMM's are a lot more flexible and easy to deal with.

x. They want to show that using these mixed effect models is a ton better than using ANOVAs. The ANOVAs and quasi-F measures don't deliver the sig levels they claim or they have less power than GLMM.

   1) With stat hypothesis tests ("gender is a sig effect in people's response" vs. "no it isn't"). Your test could say:

	Really is an effect	Really isn't an effect
Test is sig	Hooray. You should be here 95% of the time.	Type I error. Conclude sig when you shouldn't have. You should be here about 5% of the time. (Maybe the tests assume normal distribution and you don't have it?)
Test isn't sig	Type II error. You've lost an opportunity. Not so bad, though you can't write your paper. This is modeled by the power of the tests. You want to make sure your test is powerful enough but maintains Type I error standards.	