GLMMs and related

18 Nov 2024

Intercept random effects

$$
\begin{aligned}
y_{ij} &= \beta_0 + \beta_1 x_{ij} + \epsilon_{0,ij} + \epsilon_{1,j} \\
&= (\beta_0 + \epsilon_{1,j}) + \beta_1 x_{ij} + \epsilon_{1,j} \\
\epsilon_{0,ij} &\sim \text{Normal}(0, \sigma_0^2) \\
\epsilon_{1,j} &\sim \text{Normal}(0, \sigma_1^2)
\end{aligned}
$$

- Could have multiple, nested levels of random effects (genotype within population within region ...), or *crossed* REs
- formula: `y ~ 1 + x + (1 | g)`

Random-slopes model

$$
\begin{aligned}
y_{ij} &= \beta_0 + \beta_1 x_{ij} + \epsilon_{0,ij} + \epsilon_{1,j} + \epsilon_{2,j} x_{ij} \\
&= (\beta_0 + \epsilon_{1,j}) + (\beta_1 + \epsilon_{2,j}) x_{ij} + \epsilon_{0,ij} \\
\epsilon_{0,ij} &\sim \text{Normal}(0, \sigma_0^2) \\
\{\epsilon_{1,j}, \epsilon_{2,j}\} &\sim \text{MVN}(0, \Sigma)
\end{aligned}
$$

- variation in the *effect* of a treatment or covariate across groups
- estimate the correlation between the intercept and slope
- formula: `y ~ 1 + x + (1 + x | g)`

General definition

$$\underbrace{Y_i}_{\text{response}} \sim \overbrace{\text{Distr}}^{\substack{\text{conditional}\\ \text{distribution}}} (\underbrace{g^{-1}(\eta_i)}_{\substack{\text{inverse}\\ \text{link}\\ \text{function}}}, \underbrace{\phi}_{\substack{\text{scale}\\ \text{parameter}}})$$

$$\underbrace{\eta}_{\substack{\text{linear}\\ \text{predictor}}} = \underbrace{X\beta}_{\substack{\text{fixed}\\ \text{effects}}} + \underbrace{Zb}_{\substack{\text{random}\\ \text{effects}}}$$

$$\underbrace{b}_{\substack{\text{conditional}\\ \text{modes}}} \sim \text{MVN}(0, \underbrace{\Sigma(\theta)}_{\substack{\text{covariance}\\ \text{matrix}}})$$

- the structure of Z and Σ reflect one or more underlying categorical *grouping variables* (*clusters*, *blocks*, subjects, etc. etc.) or combinations thereof

What are random effects?

A method for ...

- accounting for correlations among observations within clusters
- compromising between
  complete pooling (no among-cluster variance)
  and *fixed effects* (large among-cluster variance)
- handling levels selected at random from a larger population
- sharing information among levels (*shrinkage estimation*)
- estimating variability among clusters
- allowing predictions for unmeasured clusters

Random-effect myths

- clusters must always be sampled at random
- a complete sample cannot be treated as a random effect
- random effects are always a *nuisance variable*
- nothing can be said about the predictions of a random effect
- you should always use a random effect no matter how few levels you have

Why use random effects? (inferential/philosophical)

When you:

- **do** want to
  - quantify variation among groups
  - make predictions about unobserved groups
- have (randomly) sampled clusters from a larger population
- have clusters that are **exchangeable**
- **don't** want to
  - test hypotheses about differences between particular clusters

Why use random effects? (practical) (Crawley 2002; Gelman 2005)

- want to combine information across groups
- have variation in information per cluster (number of samples or noisiness);
- have a categorical predictor that is a nuisance variable (i.e., it is not of direct interest, but should be controlled for).
- have more than 5-6 groups, or regularizing/using priors (otherwise, use fixed)

Avoiding MM

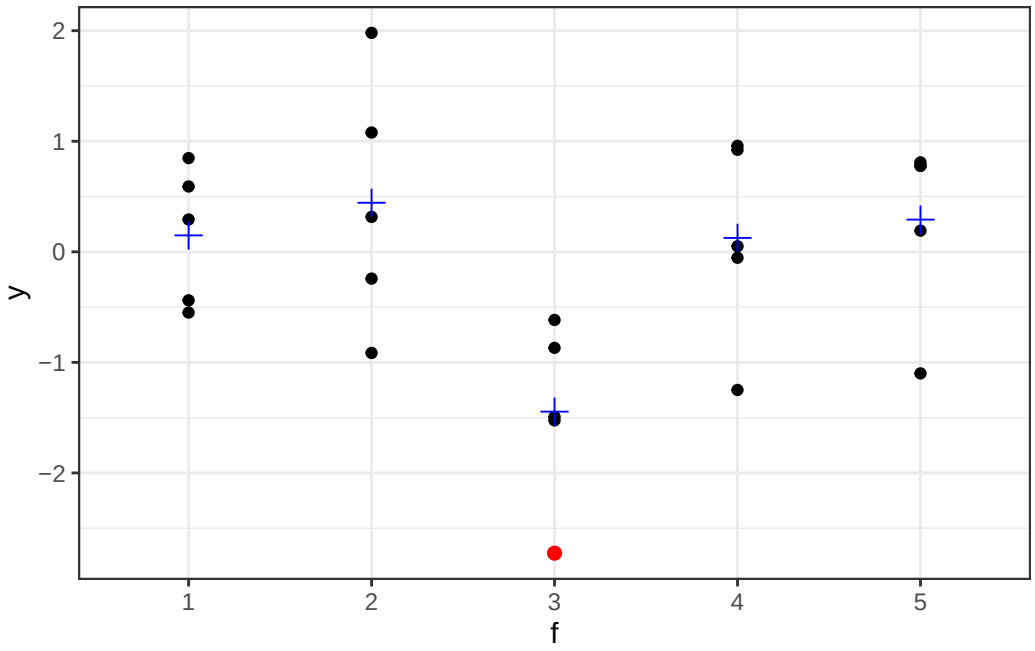
- for *nested* designs: compute cluster means (Murtaugh 2007)
- use fixed effects (or *two-stage models*) when there are
  - many samples per cluster
  - few clusters

Maximum likelihood estimation

- Best fit is a compromise between two components
  (consistency of data with fixed effects and conditional modes; consistency of random effect with RE distribution)
- $\underbrace{L(\beta,\theta)}_{\substack{\text{marginal}\\ \text{likelihood}}} = \int \underbrace{L(\mathbf{y}|\beta,b)}_{\substack{\text{conditional}\\ \text{likelihood}}} \cdot L(\mathbf{b}|\Sigma(\theta))\, d\mathbf{b}$

...

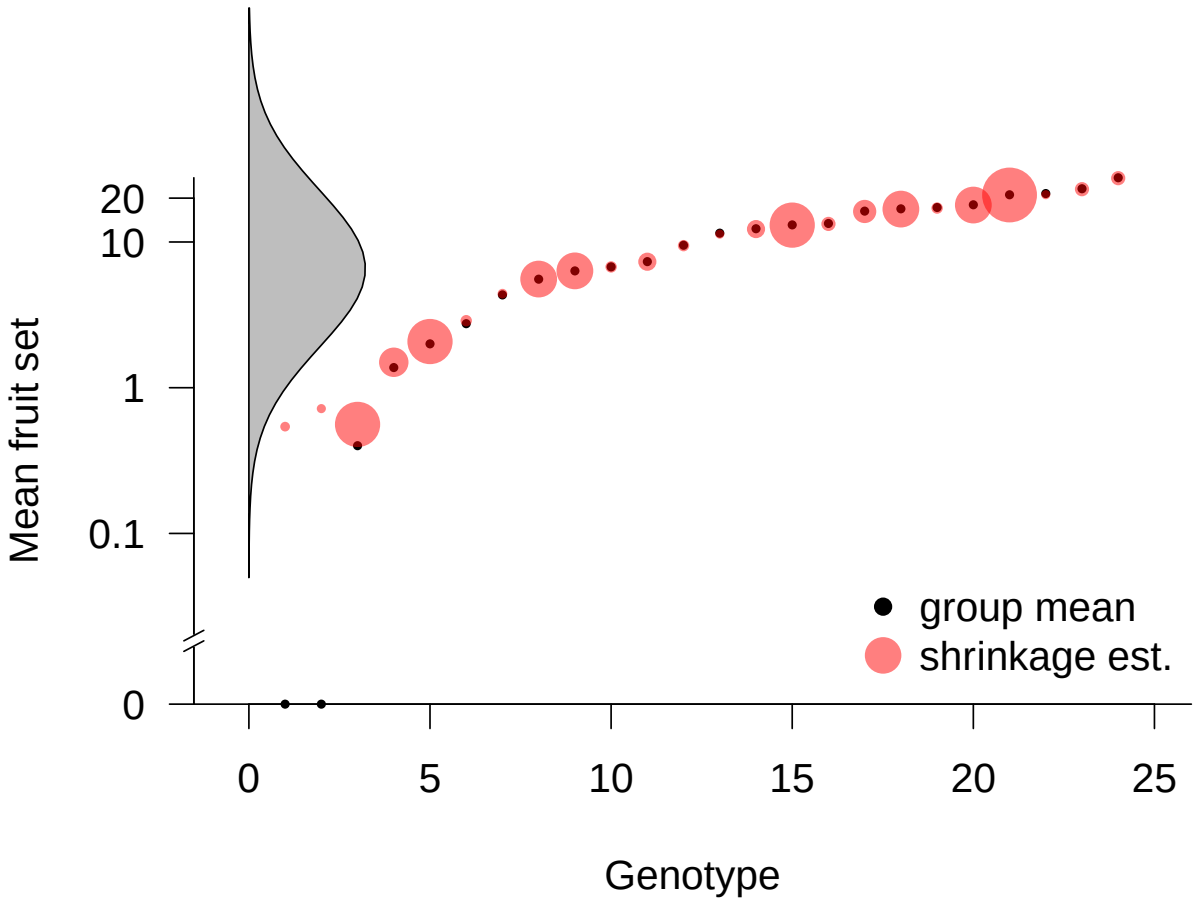
```
set.seed(101)
dd <- data.frame(f=gl(5,5))
dd$y <- simulate(~1+(1|f),newdata=dd,
                 family=gaussian,seed=101,
                 newparams=list(theta=1,beta=0,sigma=1))[[1]]
ggplot(dd,aes(x=f,y=y))+geom_point()+
    stat_summary(fun=mean,geom="point",size=3,colour="blue",
                 pch=3)+
     geom_point(data=subset(dd,y<(-2)),colour="red",size=2)+
         theme_update(panel.grid.major=element_blank(),
                      panel.grid.minor=element_blank())
```



Shrinkage: *Arabidopsis* example

```
load("../data/Banta.RData")
z<- subset(dat.tf,amd=="clipped" & nutrient=="1")
m1 <- glm(total.fruits~gen-1,data=z,family="poisson")
m2 <- glmer(total.fruits~1+(1|gen),data=z,family="poisson")
```

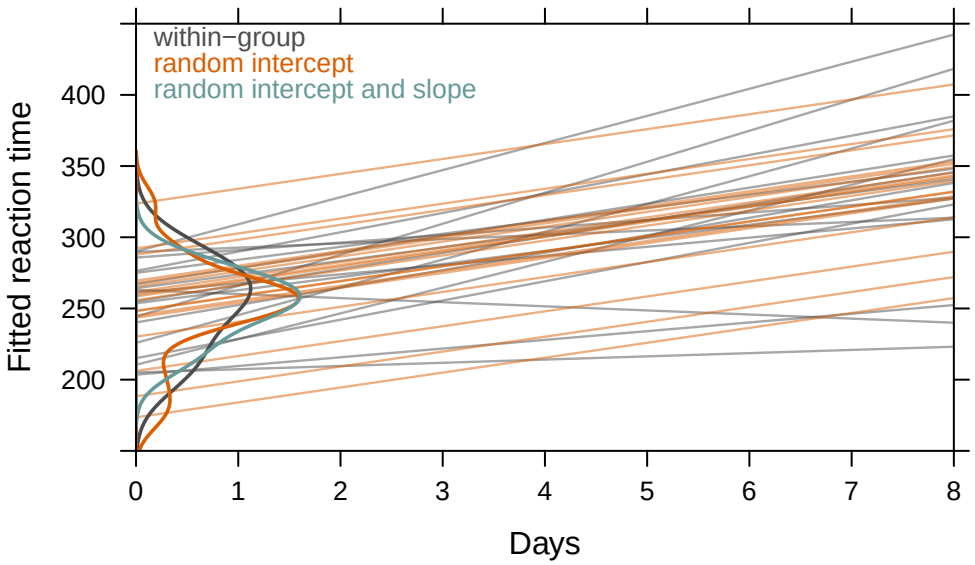
```
tt <- table(z$gen)
rr <- unlist(ranef(m2)$gen)[order(coef(m1))]+fixef(m2)
m1s <- sort(coef(m1))
m1s[1:2] <- rep(-5,2)
gsd <- attr(VarCorr(m2)$gen,"stddev")
gm <- fixef(m2)
nseq <- seq(-3,6,length.out=50)
sizefun <- function(x,smin=0.5,smax=3,pow=2) {
    smin+(smax-smin)*((x-min(x))/diff(range(x)))^pow
}
nv <- dnorm(nseq,mean=gm,sd=gsd)
##
op <- par(las=1,cex=1.5,bty="l")
plot(exp(m1s),xlab="Genotype",ylab="Mean fruit set",
     axes=FALSE,xlim=c(-0.5,25),log="y",yaxs="i",xpd=NA,
     pch=16,cex=0.5)
axis(side=1)
axis(side=2,at=c(exp(-5),0.1,1,10,20),
     labels=c(0,0.1,1,10,20),cex=0.8)
##     ylim=c(-3,5))
polygon(c(rep(0,50),nv*10),exp(c(rev(nseq),nseq)),col="gray",xpd=NA)
n <- tt[order(coef(m1))]
points(exp(rr),pch=16,col=adjustcolor("red",alpha=0.5),
       cex=sizefun(n),xpd=NA)
## text(seq_along(rr),rr,n,pos=3,xpd=NA,cex=0.6)
box()
plotrix::axis.break(axis=2,breakpos=exp(-4))
legend("bottomright",
       c("group mean","shrinkage est."),
       pch=16,pt.cex=c(1,2),
       col=c("black",adjustcolor("red",alpha=0.5)),
       bty="n")
```




```
par(op)
```

Shrinkage in a random-slopes model

From Christophe Lalanne, see here:



Estimation

- we need to compute an integral
- in *linear* mixed models the integral goes away (replaced by fancy linear algebra)
- deterministic
  - various approximate integrals (Breslow 2004):
    penalized quasi-likelihood, Laplace, Gauss-Hermite quadrature, ... (Biswas 2015);
  - more care needed for large variance, small clusters (e.g. binary data)
  - flexibility and speed vs. accuracy
- stochastic (Monte Carlo): frequentist and Bayesian (Booth and Hobert 1999; Sung and Geyer 2007; Ponciano et al. 2009). MCMC, importance sampling
  - (much) slower but flexible and accurate

Model specification

Model formula

- specify as `(t|g)`; `t` is the *varying term* and `g` is the *grouping factor*
- for intercepts `(1|g)` [**scalar random effects**], just the indicator matrix
- for more complex models (random slopes), take the *Khatri-Rao* product of the model matrix of the term with the indicator matrix of `g`

- concatenate multiple random effects terms into a single `Z` matrix
- all *varying terms* within a term can be correlated
- random effect *blocks* are independent (block diagonal)
- RE *terms* are independent (block diagonal)

Complexities

- how many/which grouping variables?
- **crossed** or **nested** ?
- what terms vary within each group?
- e.g. psychology experiments: only one grouping variable (subject), but terms can be complicated (priming $\times$ stimulus)
- e.g. teaching evaluations: students and professors are crossed random effects

What is the maximal model?

- Which effects vary *within* which groups?
- If effects don't vary within groups, then we *can't* estimate among-group variation in the effect
  - convenient, but less powerful
- e.g. female rats exposed to different doses of radiation, multiple pups per mother, multiple measurements per pup (labeled by time). Maximal model ... ?
- Maximal model is often impractical/unfeasible
  - *Culcita* (coral-reef) example: randomized-block design, so each treatment (none/crabs/shrimp/both) is repeated in every block; thus `(treat|block)` is maximal
  - CBPP data: each herd is measured in every period, so in principle we could use `(period|herd)`, not just `(1|herd)`

Singular fits

- variances equal to zero, or non-positive-definite correlation matrices
- too little data (== too little signal)
- simple case: 1-way ANOVA example
- can be (very) non-obvious in larger models
- `rePCA()`

Convergence problems

- indication of *some* kind of numerical issues
- scale/center variables
- simplify model?
- try different packages/optimizers: `allFit()`

Simplifying models

- Lots of disagreement on how to do this
- Barr et al. (2013) ("keep it maximal"); simplify until non-singular
- Bates, Kliegl, et al. (2015), Matuschek et al. (2017): stepwise reduction

Simplification strategies

- drop varying terms
- drop correlations between terms (center first!)
- reduce complexity from "general positive-definite": **compound symmetric** models, etc.

Inference

Wald tests/CIs

- need to know the "denominator degrees of freedom"
- "between-within"/"containment"
- Satterthwaite approximation
- Kenward-Roger correction (Stroup 2014); `pbkrtest` package

Likelihood ratio tests/profiling

Nonparametric bootstrap

- Bootstrapping: slow, but gold standard for frequentist models
- Need to respect structure when resampling
  - Residual bootstrapping for LMMs
  - Nested resampling where possible
- `lmeresampler` package

Parametric bootstrap

- works for any model (including crossed-RE GLMMs)
- fit null model to data
- simulate "data" from null model
- fit null and working model, compute likelihood difference
- repeat to estimate null distribution

- assumes model correctly specified
- `bootMer()`, `pbkrtest` package

How do we estimate this?

- can use EM algorithm (e.g. see here, or the lmm package)
- Or by linear algebra. For LMMs, we do a more complicated version of *data augmentation*.
- given a value for the random-effects variance, we can calculate the log-likelihood in one step (see
- large, sparse matrix computation
- has to be done *repeatedly*
- most efficient if we analyze the matrix and permute to optimize structure (Bates, Mächler, et al. 2015)
- then we need to do some kind of search over the space of variances
- derivatives are available in particular special cases

constructing the covariance matrix

- what's the best way to parameterize a positive-(semi)definite matrix? (Pinheiro and Bates 1996)
- Cholesky decomposition
  - scaled or unscaled?
  - Cholesky or log-Cholesky scale?
- separating correlation and SD vectors: glmmTMB:

$$\Sigma = D^{-1/2} L L^\top D^{-1/2}, \quad D = \text{diag}(L L^\top)$$

Zero-inflation models

- discrete (finite) mixture model; *structural* and *sampling* zeros
- e.g. for Z-I Poisson

$$\text{Prob}(0) = p_Z + (1 - p_Z)\exp(-\lambda)$$
$$\text{Prob}(x) = (1 - p_Z) \cdot \frac{\lambda^x \exp(-\lambda)}{x!}, \qquad x > 0$$

Key references

- Bates, Mächler, et al. (2015)
- Bolker (2015)
- GLMM FAQ
- mixed models task view

References

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. "Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal." *Journal of Memory and Language* 68 (3): 255–78. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. 2015. "Parsimonious Mixed Models." *arXiv:1506.04967* [*Stat*], June. https://arxiv.org/abs/1506.04967.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using Lme4." *Journal of Statistical Software* 67 (October): 1–48. https://doi.org/10.18637/jss.v067.i01.

Biswas, Keya. 2015. "Performances of Different Estimation Methods for Generalized Linear Mixed Models." Master's thesis, McMaster University. https://macsphere.mcmaster.ca/bitstream/11375/17272/2/M.Sc_Thesis_final_Keya_Biswas.pdf.

Bolker, Benjamin M. 2015. "Generalized Linear Mixed Models." In *Ecological Statistics: Contemporary Theory and Application*, edited by Gordon A. Fox, Simoneta Negrete-Yankelevich, and Vinicio J. Sosa. Oxford University Press.

Booth, James G., and James P. Hobert. 1999. "Maximizing Generalized Linear Mixed Model Likelihoods with an Automated Monte Carlo EM Algorithm." *Journal of the Royal Statistical Society. Series B* 61 (1): 265–85. https://doi.org/10.1111/1467-9868.00176.

Breslow, N. E. 2004. "Whither PQL?" In *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*, edited by Danyu Y. Lin and P. J. Heagerty, 1–22. Springer.

Crawley, Michael J. 2002. *Statistical Computing: An Introduction to Data Analysis Using S-PLUS*. John Wiley & Sons.

Gelman, Andrew. 2005. "Analysis of Variance: Why It Is More Important Than Ever." *Annals of Statistics* 33 (1): 1–53. https://doi.org/doi:10.1214/009053604000001048.

Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, Harald Baayen, and Douglas Bates. 2017. "Balancing Type I Error and Power in Linear Mixed Models." *Journal of Memory and Language* 94 (June): 305–15. https://doi.org/10.1016/j.jml.2017.01.001.

Murtaugh, Paul A. 2007. "Simplicity and Complexity in Ecological Data Analysis." *Ecology* 88 (1): 56–62. http://www.esajournals.org/doi/abs/10.1890/0012-9658%282007%2988%5B56%3ASACIED%5D2.0.CO%3B2.

Pinheiro, José C., and Douglas M. Bates. 1996. "Unconstrained Parametrizations for Variance-Covariance Matrices." *Statistics and Computing* 6 (3): 289–96. https://doi.org/10.1007/BF00140873.

Ponciano, José Miguel, Mark L. Taper, Brian Dennis, and Subhash R. Lele. 2009. "Hierarchical Models in Ecology: Confidence Intervals, Hypothesis Testing, and Model Selection Using Data Cloning." *Ecology* 90 (2): 356–62. http://www.jstor.org/stable/27650990.

Stroup, Walter W. 2014. "Rethinking the Analysis of Non-Normal Data in Plant and Soil Science." *Agronomy Journal* 106: 1–17. https://doi.org/10.2134/agronj2013.0342.

Sung, Yun Ju, and Charles J. Geyer. 2007. "Monte Carlo Likelihood Inference for Missing Data Models." *The Annals of Statistics* 35 (3): 990–1011. https://doi.org/10.1214/009053606000001389.