The methods of VetResearchLMM

Muhammad Yaseen

2026-11-10

Table of contents

1 Abstract **1**

2 The model **2**

3 Estimation **2**
3.1 Likelihoods 2
3.2 Maximisation 2
3.3 Method of moments 3

4 Fixed effects **3**

5 Denominator degrees of freedom **3**

6 Inference spaces and prediction **4**

7 Expected mean squares **4**

8 Fit statistics **4**

9 Notation **5**

10 References **6**

1 Abstract

This vignette states the methods that `vlm_fit()` and the functions that use a fit implement: the linear mixed model, restricted and full maximum likelihood, the generalised least-squares solution in the overparameterised coding, estimability, Type 3 and Type 1 tests, the five methods for the denominator degrees of freedom, the broad and narrow inference spaces, best linear unbiased prediction, expected mean squares, covariance structures and fit statistics. The notation follows the book: scalars upper case, vectors lower-case bold, matrices upper-case bold.

2 The model

With N observations,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u}+\mathbf{e}, \qquad \mathbf{u} \sim N(\mathbf{0}, \mathbf{G}), \quad \mathbf{e} \sim N(\mathbf{0}, \mathbf{R}), \quad \mathrm{Var}(\mathbf{y}) = \mathbf{V}(\boldsymbol{\theta}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^\top+\mathbf{R},$$

where $\boldsymbol{\theta}$ holds the covariance parameters. $\mathbf{X}$ is overparameterised, as in SAS: one column for every level (or observed cell) of a classification effect, so that its rank P is smaller than its number of columns.

Random effects. A term of `random = ~ Region + Region:Drug` contributes $\sigma_k^2 \mathbf{Z}_k \mathbf{Z}_k^\top$ to $\mathbf{V}$, with $\mathbf{Z}_k$ the indicators of its levels. `vlm_random(effects = ~ 1 + time, subject = ~ animal, type = "un")` gives every subject s a vector $\mathbf{b}_s \sim N(\mathbf{0}, \mathbf{G}_0)$ of coefficients; `type = "un"` leaves $\mathbf{G}_0$ unstructured, with parameters $UN(1,1), UN(2,1), UN(2,2)$, and `type = "vc"` makes it diagonal.

Residual structures within a subject (`vlm_repeated()`), with t_i the time and r_i the rank of the time of observation i:

$$\mathrm{CS}: \sigma_1+\sigma^2\delta_{ij}, \qquad \mathrm{AR}(1): \sigma^2\rho^{|r_i-r_j|}, \qquad \mathrm{SP(POW)}: \sigma^2\rho^{|t_i-t_j|}, \qquad \mathrm{UN}: \sigma_{t_i t_j}.$$

3 Estimation

3.1 Likelihoods

With $\mathbf{r} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$ the generalised least-squares residuals, the restricted (REML) and full (ML) log-likelihoods are

$$\ell_R(\boldsymbol{\theta}) = -\tfrac{1}{2}\left\{(N-P)\log 2\pi + \log|\mathbf{V}| + \log|\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X}| + \mathbf{r}^\top\mathbf{V}^{-1}\mathbf{r}\right\}, \qquad \ell(\boldsymbol{\theta}) = -\tfrac{1}{2}\left\{N\log 2\pi + \log|\mathbf{V}| + \mathbf{r}^\top\mathbf{V}^{-}\right.$$

where $\mathbf{X}$ is restricted to its P non-aliased columns (Patterson and Thompson 1971; Harville 1977). REML accounts for the degrees of freedom used by the fixed effects; ML underestimates the variances (Table 2.4.2.1 of the book).

3.2 Maximisation

With $\mathbf{V}_i = \partial\mathbf{V}/\partial\theta_i$ and $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}^\top\mathbf{V}^{-1}$ (for ML, $\mathbf{P} = \mathbf{V}^{-1}$), the score and the average information (Gilmour, Thompson and Cullis 1995) are

$$\frac{\partial\ell_R}{\partial\theta_i} = -\tfrac{1}{2}\,\mathrm{tr}(\mathbf{P}\mathbf{V}_i) + \tfrac{1}{2}\mathbf{y}^\top\mathbf{P}\mathbf{V}_i\mathbf{P}\mathbf{y}, \qquad \mathcal{I}_{ij} = \tfrac{1}{2}\mathbf{y}^\top\mathbf{P}\mathbf{V}_i\mathbf{P}\mathbf{V}_j\mathbf{P}\mathbf{y}.$$

Newton steps $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}+\mathcal{I}^{-1}\nabla\ell_R$ are halved until the likelihood increases; variances stay non-negative (parameters on the bound leave the step), as in SAS. The standard errors of $\widehat{\boldsymbol{\theta}}$ come from the observed information, the Hessian of $-\ell_R$, computed from the analytic score.

3.3 Method of moments

`method = "Type3"` equates the Type 3 mean squares of the random effects and the residual to their expectations (next section but one) and solves for the variance components; negative solutions are set to zero.

4 Fixed effects

The generalised least-squares solution is

$$\widehat{\boldsymbol{\beta}} = \mathbf{C}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{y}, \qquad \mathbf{C} = (\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^-,$$

with the generalised inverse that sets aliased columns (in model order) to zero: the G2 inverse of the SAS sweep operator, which gives the last level of each effect the value 0. A combination $\mathbf{l}^\top\boldsymbol{\beta}$ is *estimable* when $\mathbf{l}$ lies in the row space of $\mathbf{X}$; then $\mathbf{l}^\top\widehat{\boldsymbol{\beta}}$ and $\mathrm{Var}(\mathbf{l}^\top\widehat{\boldsymbol{\beta}}) = \mathbf{l}^\top\mathbf{C}\mathbf{l}$ do not depend on the generalised inverse. Non-estimable combinations are returned as `NA`.

Least-squares means use $\mathbf{l}$ with 1 for the intercept and the matching level, equal weights over the levels of the other factors of each effect, and the means of the covariates (or the values of `at`).

Type 3 tests use the sum-to-zero (Sigma-restricted) coding of every effect, in which effects nested in another sum to zero within it: the test of an effect is the Wald test that its coefficients are zero. This is the SAS Type 3 hypothesis whenever every cell is observed. **Type 1 tests** are sequential: after whitening with $\mathbf{V}^{-1/2}$, the rows of the triangular factor of the QR decomposition of $\mathbf{X}$ that belong to the effect.

For $\mathbf{L}$ of rank Q,

$$F = \frac{(\mathbf{L}\widehat{\boldsymbol{\beta}})^\top(\mathbf{L}\mathbf{C}\mathbf{L}^\top)^{-1}(\mathbf{L}\widehat{\boldsymbol{\beta}})}{Q}.$$

5 Denominator degrees of freedom

- **Containment**: the rank contribution of the random effects that contain the fixed effect (all its variables, covariates included), the smallest one if several do; $N - \mathrm{rank}[\mathbf{X}\ \mathbf{Z}]$ if none does.
- **Residual**: $N - \mathrm{rank}[\mathbf{X}\ \mathbf{Z}]$.
- **Between-within**: for effects constant within subjects, the number of subjects minus the rank of those effects; for the others, $N - \mathrm{rank}[\mathbf{X}\ \mathbf{Z}_s]$ with $\mathbf{Z}_s$ the subject indicators.
- **Satterthwaite** (Satterthwaite 1946; Giesbrecht and Burns 1985): with $\mathbf{A}$ the asymptotic covariance of $\hat{\boldsymbol{\theta}}$ and $g_i = \partial(\mathbf{l}^\top\mathbf{C}\mathbf{l})/\partial\theta_i = -\mathbf{l}^\top\mathbf{C}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{X}\mathbf{C}\mathbf{l}$,

$$\nu = \frac{2(\mathbf{l}^\top\mathbf{C}\mathbf{l})^2}{\mathbf{g}^\top\mathbf{A}\mathbf{g}}.$$

For $Q > 1$ rows, $\mathbf{LCL}^\top$ is decomposed into Q independent rows with ν_m each, and $\nu = 2E/(E-Q)$ with $E = \sum_m \nu_m/(\nu_m - 2)$ (Fai and Cornelius 1996).

- **Kenward-Roger** (Kenward and Roger 1997): the covariance of $\widehat{\boldsymbol{\beta}}$ is inflated for the uncertainty in $\hat{\boldsymbol{\theta}}$,

$$\boldsymbol{\Phi}_A = \mathbf{C} + 2\mathbf{C}\Big\{\sum_{i,j} W_{ij}(\mathbf{Q}_{ij} - \mathbf{P}_i\mathbf{C}\mathbf{P}_j - \tfrac{1}{4}\mathbf{R}_{ij})\Big\}\mathbf{C},$$

with $\mathbf{P}_i = -\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{X}$, $\mathbf{Q}_{ij} = \mathbf{X}^\top\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j\mathbf{V}^{-1}\mathbf{X}$, $\mathbf{R}_{ij} = \mathbf{X}^\top\mathbf{V}^{-1}(\partial^2\mathbf{V}/\partial\theta_i\partial\theta_j)\mathbf{V}^{-1}\mathbf{X}$ and $\mathbf{W}$ the inverse expected information; the F statistic is scaled and its degrees of freedom matched to the first two moments.

6 Inference spaces and prediction

Henderson's mixed model equations give the best linear unbiased predictions $\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}^\top\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})$ and the covariance of the errors

$$\operatorname{Var}\begin{pmatrix}\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\mathbf{u}} - \mathbf{u}\end{pmatrix} = \begin{pmatrix}\mathbf{C} & -\mathbf{C}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{Z}\mathbf{G} \\ \cdot & \mathbf{G} - \mathbf{G}\mathbf{Z}^\top\mathbf{P}\mathbf{Z}\mathbf{G}\end{pmatrix}.$$

An estimate $\mathbf{l}^\top\boldsymbol{\beta}$ is in the **broad** inference space (valid for the population of regions, herds, …); adding $\mathbf{k}^\top\mathbf{u}$ (argument `random` of `vlm_estimate()`) restricts it to the observed levels, the **narrow** space (Example 2.5.4.1 of the book; McLean, Sanders and Stroup 1991).

7 Expected mean squares

A sum of squares is a quadratic form $\mathbf{y}^\top\mathbf{A}\mathbf{y}$ of rank DF. For random effects of design $\mathbf{Z}_k$,

$$\mathrm{E}(MS) = \sigma^2 + \sum_k \frac{\operatorname{tr}(\mathbf{Z}_k^\top\mathbf{A}\mathbf{Z}_k)}{DF}\sigma_k^2 + Q(\text{fixed}),$$

which with one observation deleted from Example 1.2.4 gives the coefficients 9.7143 and 9.7222 of Table 1.3.2.3. `vlm_ems()` builds the error term of each test as the combination of mean squares with the same expectation under the null hypothesis, with Satterthwaite degrees of freedom.

8 Fit statistics

With d covariance parameters (plus P for ML), $N^* = N - P$ for REML (N for ML) and s subjects:

$$AIC = -2\ell + 2d, \qquad AICC = -2\ell + \frac{2dN^*}{N^* - d - 1}, \qquad BIC = -2\ell + d\log s,$$

smaller is better. SAS version 6, which the book used, printed the "larger is better" forms $\ell - d$ and $\ell - \frac{d}{2}\log N^*$; both are given. The null model likelihood ratio statistic compares the model with the one with independent errors only.

9 Notation

```
nt <-
  vlm_notation(
      format = c("table", "path", "tex")[1]
    )
knitr::kable(nt[, c("text", "meaning", "argument")])
```

text	meaning	argument
N	number of observations	data
P	rank of the fixed-effects design matrix	fixed
y	response (observations)	fixed
X	design matrix of the fixed effects (overparameterised: one column per level)	fixed
beta	fixed effects	fixed
Z	design matrix of the random effects	random
u	random effects	random
e	residual errors	
G	covariance matrix of the random effects	random
R	covariance matrix of the residuals	repeated
V	covariance matrix of the observations: Z G Z' + R	
theta	covariance parameters (variance components, covariances, correlations)	
V_i	derivative of V with respect to the i-th covariance parameter	
P	V^-1 - V^-1 X (X' V^-1 X)^- X' V^-1	
C	(X' V^-1 X)^-: covariance matrix of the fixed-effects solution	
l	restricted and full log-likelihood	method
L	coefficients of an estimate or contrast of fixed effects	L
k	coefficients of random effects in a narrow-space estimate	random
nu	denominator degrees of freedom	ddfm
A	asymptotic covariance matrix of the covariance-parameter estimates	
Phi_A	Kenward-Roger adjusted covariance matrix of the fixed effects	ddfm
sigma2	residual variance	
sigma2_k	variance of the k-th random effect	random
rho	correlation of the first-order autoregressive or spatial power structure	repeated

text	meaning	argument
EMS	expected mean square	type
d	number of covariance parameters	
s	number of subjects	repeated

10 References

Fai, A. H.-T. and Cornelius, P. L. (1996). Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal of Statistical Computation and Simulation*, 54, 363-378.

Giesbrecht, F. G. and Burns, J. C. (1985). Two-stage analysis based on a mixed model. *Biometrics*, 41, 477-486.

Gilmour, A. R., Thompson, R. and Cullis, B. R. (1995). Average information REML. *Biometrics*, 51, 1440-1450.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation. *Journal of the American Statistical Association*, 72, 320-338.

Kenward, M. G. and Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53, 983-997.

McLean, R. A., Sanders, W. L. and Stroup, W. W. (1991). A unified approach to mixed linear models. *The American Statistician*, 45, 54-64.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58, 545-554.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2, 110-114.