merDeriv: Derivative Computations for Generalized Linear Mixed Effects Models with Application

2021 Psychoco

Ting Wang

merDeriv

Compute case-wise and cluster-wise derivative for mixed effects models with respect to fixed effects parameter, random effect (co)variances, and residual variance (Wang & Merkle, 2018; Wang, Graves, Rosseel, & Merkle, 2020).

Outline

Motivation

Within R ecosystem:

- *sandwich*: Robust Covariance Matrix Estimators
- *strucchange*: Testing, Monitoring, and Dating Structural Changes
- *nonnest2*: Tests of Non-Nested Models
- *partykit*: A Toolkit for Recursive Partytioning

Motivation

- All utilize casewise partial first derivatives (scores) and second derivatives (Fisher information matrix/Hessian) of the log-likelihood.
- Not available for *lme4* models (LMM and GLMM).
- Utilize *lme4* output to compute these quantities, so that all packages mentioned above can be used on LMM and GLMM.

LMM

- Computation (analytical)
- Application

LMM

$$\boldsymbol{y}|\boldsymbol{b} \sim N(\boldsymbol{X\beta} + \boldsymbol{Zb}, \boldsymbol{R}) \quad (1)$$

$$\boldsymbol{b} \sim N(\boldsymbol{0}, \boldsymbol{G}) \quad (2)$$

$$\boldsymbol{R} = \sigma_r^2 \boldsymbol{I}_n, \quad (3)$$

- The marginal distribution of the LMM is

$$\boldsymbol{y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{V}) \tag{4}$$

where

$$\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^\top + \sigma_r^2 \boldsymbol{I}_n \tag{5}$$

- The marginal log-likelihood can be expressed as

$$\ell(\boldsymbol{\sigma}^2, \boldsymbol{\beta}; \boldsymbol{y}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{V}|) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \tag{6}$$

Scores for $\boldsymbol{\beta}$

$$\frac{\partial \ell(\boldsymbol{\sigma}^2, \boldsymbol{\beta}; \boldsymbol{y})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}). \tag{7}$$

$$s(\boldsymbol{\beta}; \boldsymbol{y}) = \left\{\boldsymbol{X}^\top \boldsymbol{V}^{-1}\right\}^T \circ (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}). \tag{8}$$

Scores for $\boldsymbol{\sigma}^2$

$$\frac{\partial \ell(\boldsymbol{\sigma}^2, \boldsymbol{\beta}; \boldsymbol{y})}{\partial \sigma_k^2} = -\frac{1}{2}\mathrm{tr}\left[\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \sigma_k^2}\right] + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1}\left(\frac{\partial \boldsymbol{V}}{\partial \sigma_k^2}\right)\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \quad (9)$$

$$s(\sigma_k^2; \boldsymbol{y}) = -\frac{1}{2}\mathrm{diag}\left[\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \sigma_k^2}\right] + \left\{\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1}\left(\frac{\partial \boldsymbol{V}}{\partial \sigma_k^2}\right)\boldsymbol{V}^{-1}\right\}^T \circ (\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \quad (10)$$

Fisher information matrix

$$\boldsymbol{A} = \left[\begin{array}{c|c} -E\left(\frac{\partial^2 \ell(\boldsymbol{\sigma}^2,\boldsymbol{\beta};\boldsymbol{y})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right) & -E\left(\frac{\partial^2 \ell(\boldsymbol{\sigma}^2,\boldsymbol{\beta};\boldsymbol{y})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\sigma}^2}\right) \\ \hline -E\left(\frac{\partial^2 \ell(\boldsymbol{\sigma}^2,\boldsymbol{\beta};\boldsymbol{y})}{\partial\boldsymbol{\sigma}^2\partial\boldsymbol{\beta}}\right) & -E\left(\frac{\partial^2 \ell(\boldsymbol{\sigma}^2,\boldsymbol{\beta};\boldsymbol{y})}{\partial\boldsymbol{\sigma}^2\partial\boldsymbol{\sigma}^2}\right) \end{array}\right]$$

Fisher information matrix

$$\boldsymbol{A} = \left[\begin{array}{c|c} \boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X} & \mathbf{0} \\ \hline \mathbf{0} & \left(\frac{1}{2}\right) \operatorname{tr}\left[\boldsymbol{V}^{-1} \left(\frac{\partial \boldsymbol{V}}{\partial \sigma_{k1}^2}\right) \boldsymbol{V}^{-1} \left(\frac{\partial \boldsymbol{V}}{\partial \sigma_{k2}^2}\right)\right] \end{array}\right]$$

Application: Huber-White sandwich estimator

$$\boldsymbol{V}(\hat{\boldsymbol{\xi}}) = (\boldsymbol{A})^{-1}\boldsymbol{B}(\boldsymbol{A})^{-1}, \tag{11}$$

- $\boldsymbol{A} = -E\left(\ell''(\hat{\boldsymbol{\xi}}; \boldsymbol{y})\right)$
- $\boldsymbol{B} = \sum_{j=1}^{J}\left[\sum_{i\in c_j} s_i(y_i|\boldsymbol{\xi})\right]^{\top}\left[\sum_{i\in c_j} s_i(y_i|\boldsymbol{\xi})\right].$

Square roots of the diagonal elements of $\boldsymbol{V}$ are the “robust standard errors” (Zeileis, 2006).

```
R> library("lme4")
R> library("merDeriv")
R> lme4fit <- lmer(Reaction ~ Days + (Days|Subject),
+                  sleepstudy, REML = FALSE)
```

- casewise score

```
R> score1 <- estfun.lmerMod(lme4fit, level = 1)
R> dim(score1)
[1] 180   6
```

- clusterwise score

```
R> score2 <- estfun.lmerMod(lme4fit, level = 2)
R> dim(score2)
[1] 18  6
```

full = TRUE; not available in *lme4*

```
R> vcov.lmerMod(lme4fit, level = 2, full = TRUE)

6 x 6 Matrix of class "dgeMatrix"
     (Intercept)  Days cov_Subject.(Intercept) cov_Subject.Days.(Intercept)
[1,]       43.99 -1.37                     0.0                            0
[2,]       -1.37  2.26                     0.0                            0
[3,]        0.00  0.00                 70359.6                        -2282
[4,]        0.00  0.00                 -2282.4                         1838
[5,]        0.00  0.00                    92.6                         -115
[6,]        0.00  0.00                 -2058.1                          325
     cov_Subject.Days residual
[1,]              0.0      0.0
[2,]              0.0      0.0
[3,]             92.6  -2058.1
[4,]           -115.3    325.0
[5,]            184.2    -72.2
[6,]            -72.2   5957.7
```

```
R> library("sandwich")
R> sandwich(lme4fit, bread. = bread.lmerMod(lme4fit, full = TRUE),
+           meat. = meat(lme4fit, level = 2, full = TRUE))

6 x 6 Matrix of class "dgeMatrix"
     (Intercept)    Days cov_Subject.(Intercept) cov_Subject.Days.(Intercept)
[1,]       43.99  -1.370                  -523.4                      -20.768
[2,]       -1.37   2.257                   -56.1                        0.185
[3,]     -523.40 -56.094                 45232.1                     1055.380
[4,]      -20.77   0.185                  1055.4                     1862.988
[5,]       -5.92  -1.977                   427.4                      -89.284
[6,]      149.15  78.709                -27398.6                     1214.371
     cov_Subject.Days residual
[1,]            -5.92    149.2
[2,]            -1.98     78.7
[3,]           427.39 -27398.6
[4,]           -89.28   1214.4
[5,]           137.89   -492.6
[6,]          -492.56  43229.0
```

Score-based tests background

- Score-based tests: Utilize deviations in the model scores, i.e., the first derivatives of the model's log likelihood function.
- Scores are individual terms of the gradient. They tell us how well a particular parameter describes a particular individual.
- Used to detect parameter instability and related issues (Zeileis, Leisch, Hornik, & Kleiber, 2002; Zeileis & Hornik, 2007; Merkle & Zeileis, 2013; Merkle, Fan, & Zeileis, 2014; Wang, Merkle, Anguera, & Turner, 2020).

Fit the model

- 7185 U.S. high-school students from 160 schools completed a math achievement test, with the students' socioeconomic status (cses) as a level 1 covariate.
- It is plausible that the relationship between cses and math achievement differs across schools with different meanses (level 2 covariate)
- Heterogeneity in random effect or residual variance parameters would result in incorrect significance test in fixed effects' coefficients.
- Score-based test provides a simple, systematic way to detect heterogeneity in typical LMM.

Fit the model

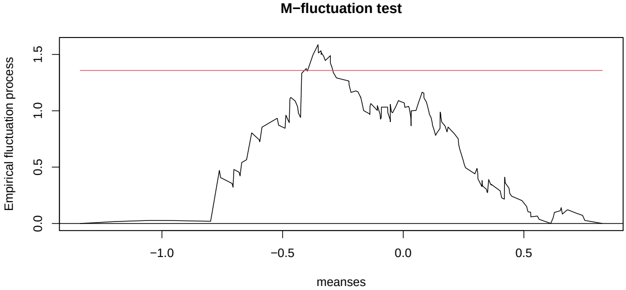
```
R> library("mlmRev")
R> library("lme4")
R> m1 <- lmer(mAch ~ cses + (cses | school), data = Hsb82, REML = FALSE)
```

Test parameters of interest

```
R> library("strucchange")
R> library("merDeriv")
R> dm <- sctest(m1, order.by = unique(orderHsb82$meanses), parm = 5,
+                functional = "DM", plot = FALSE)
R> dm$p.value
[1] 0.013
```

Application: Statistical test

Figure 1: Fluctuation process of variance of random intercept variance across values of meanses



GLMM

- Computation (numerical)
- Application

GLMM

$$E(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\Lambda_\theta}) = \boldsymbol{\mu}|\boldsymbol{\Lambda_\theta}, \boldsymbol{u} \tag{12}$$

$$\boldsymbol{\mu} = g^{-1}(\boldsymbol{\eta}|\boldsymbol{\Lambda_\theta}, \boldsymbol{u}) \tag{13}$$

$$\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Zb} \tag{14}$$

$$\boldsymbol{b} = \boldsymbol{\Lambda_\theta u} \tag{15}$$

$$\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{I}) \tag{16}$$

$$\boldsymbol{G} = \boldsymbol{\Lambda_\theta \Lambda_\theta^T} \tag{17}$$

The marginal log likelihood can be expressed as:

$$\ell = \log \int f_{\boldsymbol{y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{u}) f_{\boldsymbol{u}}(\boldsymbol{u}) d\boldsymbol{u}. \tag{18}$$

Scores for $\boldsymbol{\beta}$

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \frac{\int \frac{\partial \log f_{\boldsymbol{y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{u})}{\partial \boldsymbol{\beta}} f_{\boldsymbol{y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{u}) f_{\boldsymbol{u}}(\boldsymbol{u})\, d\boldsymbol{u}}{f_{\boldsymbol{y}}(\boldsymbol{y})}, \tag{19}$$

where $f_{\boldsymbol{y}}(\boldsymbol{y}) = \int f_{\boldsymbol{y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{u}) f_{\boldsymbol{u}}(\boldsymbol{u})\, d\boldsymbol{u}$.

$f_{\boldsymbol{y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{u})$ is GLM.

$$\frac{\partial \ell}{\partial \boldsymbol{\Lambda}_{\boldsymbol{\theta}}} = \frac{\int \frac{\partial \log f_{\boldsymbol{y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{u})}{\partial \boldsymbol{\Lambda}_{\boldsymbol{\theta}}} f_{\boldsymbol{y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{u}) f_{\boldsymbol{u}}(\boldsymbol{u})\, d\boldsymbol{u}}{f_{\boldsymbol{y}}(\boldsymbol{y})}, \tag{20}$$

where $\frac{\partial \log f_{\boldsymbol{y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{u})}{\partial \boldsymbol{\Lambda}_{\boldsymbol{\theta}}}$ equals to $\boldsymbol{u}^T \frac{\partial \boldsymbol{\Lambda}_{\boldsymbol{\theta}}}{\partial \theta} \boldsymbol{Z}^T (\boldsymbol{y} - \boldsymbol{\mu})$.

Reparameterization

- Chain rule from Cholesky factor to variance.
- Chain rule from variance to standard deviation.
- *ranpar* argument in *estfun* or *vcov* in *merDeriv* to specify the parameters of interest: *sd, var, theta*.

Quadrature

- All the derivatives above involve integrals that marginalize over the model random effects $\boldsymbol{u}$.
- Simplified version of multivariate adaptive Gauss-Hermite quadrature
- Simplifications are based on the fact that the “adaptive” step can be replaced by the posterior modes and variances of random effects from *lme4* (Merkle, Furr, & Rabe-Hesketh, 2019; Wang, Graves, et al., 2020)

Hessian

- *lme4* computes Hessian for *glmer* in *optinfo* (finite difference approach)
- The *merDeriv* package provides a convenient function to access this Hessian

Vuong's test

- SPISA in the R package *psychotree* (Strobl, Kopf, & Zeileis, 2015): 1075 Bavarian university students who took a general knowledge quiz (45 items).
- Explanatory IRT: covariates such as age, gender, semester of university enrollment, and elite university status.
- mod1 includes age and gender as covariates, while mod2 uses current semester of university enrollment and whether the university has been granted "elite" status or not.
- Comparison for non-nested models (Merkle, You, & Preacher, 2016).

Vuong's test

```
R> mod1 <- glmer(response ~ -1 + item + agecent + gender + (1|pnum),
+                data = longdat, family = binomial,
+                control = glmerControl(optimizer='bobyqa',
+                                       optCtrl=list(maxfun=10000)))
R> mod2 <- glmer(response ~ -1 + item + semester + elite + (1|pnum),
+                data = longdat, family = binomial,
+                control = glmerControl(optimizer='bobyqa',
+                                       optCtrl=list(maxfun=10000)))
```

Vuong's test

```
R> vcg <- function(obj) vcov(obj, full = TRUE)
R> vuongtest(mod1, mod2, ll1 = llcont.glmerMod, ll2 = llcont.glmerMod,
+            score1 = estfun.glmerMod, score2 = estfun.glmerMod,
+            vc1 = vcg, vc2 = vcg)

Model 1
 Class: glmerMod
 Call: glmer(formula = response ~ -1 + item + agecent + gender + (1 | ...

Model 2
 Class: glmerMod
 Call: glmer(formula = response ~ -1 + item + semester + elite + (1 | ...

Variance test
  H0: Model 1 and Model 2 are indistinguishable
  H1: Model 1 and Model 2 are distinguishable
    w2 = 0.033,   p = 6.34e-07

Non-nested likelihood ratio test
  H0: Model fits are equal for the focal population
  H1A: Model 1 fits better than Model 2
    z = -0.356,   p = 0.639
  H1B: Model 2 fits better than Model 1
    z = -0.356,   p = 0.3611
```

Future Developments

- GLMM tree: can be immediately extended.
- Heterogeneity in GLMM: is potentially more problematic than in LMM, because it will impact fixed effects' estimates.
- LMM/GLMM with cross/nested random effects: need to find a way to "decorrelate" correlations in rows of score matrix.

Thanks

- *install.packages("merDeriv")*

- Wang, T., & Merkle, E. C. (2018). merDeriv: Derivative computations for linear mixed effects models with application to robust standard errors. Journal of Statistical Software, Code Snippets, 87(1), 1-16.
- Wang, T., Graves, B., Rosseel, Y., & Merkle, E. C. (2020). Computation and application of generalized linear mixed model derivatives using lme4. arXiv preprint arXiv:2011.10414.

Acknowledgments

Edgar C. Merkle

Benjamin Graves

Yves Rosseel

NSF-1460719

References

Merkle, E. C., Fan, J., & Zeileis, A. (2014). Testing for measurement invariance with respect to an ordinal variable. *Psychometrika*, *79*, 569–584.

Merkle, E. C., Furr, D., & Rabe-Hesketh, S. (2019). Bayesian comparison of latent variable models: Conditional versus marginal likelihoods. *Psychometrika*, *84*, 802–829.

Merkle, E. C., You, D., & Preacher, K. J. (2016). Testing nonnested structural equation models. *Psychological Methods*, *21*(2), 151–163.

Merkle, E. C., & Zeileis, A. (2013). Tests of measurement invariance without subgroups: A generalization of classical methods. *Psychometrika*, *78*, 59–82.

Strobl, C., Kopf, J., & Zeileis, A. (2015). Rasch trees: A new method for detecting differential item functioning in the Rasch model. *Psychometrika*, *80*(2), 289–316. doi: 10.1007/s11336-013-9388-3

Wang, T., Graves, B., Rosseel, Y., & Merkle, E. C. (2020). *Computation and application of generalized linear mixed model derivatives using lme4.*

Wang, T., & Merkle, E. C. (2018). merDeriv: Derivative computations for linear mixed effects models with application to robust standard errors. *Journal of Statistical Software*, *87*(1), 1–16. doi: 10.18637/jss.v087.c01

Wang, T., Merkle, E. C., Anguera, J. A., & Turner, B. M. (2020). Score-based tests for detecting heterogeneity in linear mixed models. *Behavior Research Methods*.

Zeileis, A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, *16*(9), 1–16. doi: 10.18637/jss.v016.i09

Zeileis, A., & Hornik, K. (2007). Generalized M-fluctuation tests for parameter instability. *Statistica Neerlandica*, *61*, 488–508.

Zeileis, A., Leisch, F., Hornik, K., & Kleiber, C. (2002). strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software*, *7*(2), 1–38. Retrieved from `http://www.jstatsoft.org/v07/i02/`