Translating lme4 models to sommer

Giovanny Covarrubias-Pazaran

2022-04-16

The sommer package was developed to provide R users a powerful and reliable multivariate mixed model solver. The package is focused on problems of the type \( p > n \) (more effects to estimate than observations) and its core algorithm is coded in C++ using the Armadillo library. This package allows the user to fit mixed models with the advantage of specifying the variance-covariance structure for the random effects, and specifying heterogeneous variances, and obtaining other parameters such as BLUPs, BLUEs, residuals, fitted values, variances for fixed and random effects, etc.

The purpose of this vignette is to show how to translate the syntax formula from \texttt{lme4} models to \texttt{sommer} models. Feel free to remove the comment marks from the \texttt{lme4} code so you can compare the results.

1) Random slopes with same intercept
2) Random slopes and random intercepts (without correlation)
3) Random slopes and random intercepts (with correlation)
4) Random slopes with a different intercept
5) Other models not available in \texttt{lme4}

1) Random slopes

This is the simplest model people use when a random effect is desired and the levels of the random effect are considered to have the same intercept.

```r
# install.packages("lme4")
# library(lme4)
library(sommer)
data(DT_sleepstudy)
DT <- DT_sleepstudy

###########
## lme4
###########
fm1 <- lmer(Reaction ~ Days + (1 | Subject), data=DT)
summary(fm1)  # or vc <- VarCorr(fm1); print(vc, comp=c("Variance"))

# Random effects:
# Groups Name Variance Std.Dev.
# Subject (Intercept) 1378.2 37.12
# Residual 960.5 30.99
# Number of obs: 180, groups: Subject, 18

###########
## sommer
###########
fm2 <- mmer(Reaction ~ Days, random= ~ Subject, data=DT, tolparinv = 1e-6, verbose = FALSE)
summary(fm2)$varcomp
```
2) Random slopes and random intercepts (without correlation)

This is the a model where you assume that the random effect has different intercepts based on the levels of another variable. In addition the || in \texttt{lme4} assumes that slopes and intercepts have no correlation.

\begin{verbatim}
# fm1 <- lmer(Reaction ~ Days + (Days || Subject), data=DT)
# summary(fm1) # or vc <- VarCorr(fm1); print(vc,comp=c("Variance"))

# Groups   Name Variance Std.Dev.
# Subject (Intercept) 627.57  25.051
# Subject.1 Days     35.86   5.988
# Residual                      653.58  25.565
# Number of obs: 180, groups: Subject, 18

# fm2 <- mmer(Reaction ~ Days, 
# random = ~ Subject + vs(Days, Subject), 
# data=DT, tolparinv = 1e-6, verbose = FALSE)
# summary(fm2)$varcomp

# Groups   Name Variance Std.Dev. Corr
# Subject (Intercept) 627.54087 283.52939 2.213319
# Days:Subject.Reaction-Reaction 35.86008 14.53187 2.467686
# units.Reaction-Reaction  653.58305 76.72711 8.518281
# Number of obs: 180, groups: Subject, 18
\end{verbatim}

Notice that Days is a numerical (not factor) variable.

3) Random slopes and random intercepts (with correlation)

This is the a model where you assume that the random effect has different intercepts based on the levels of another variable. In addition a single | in \texttt{lme4} assumes that slopes and intercepts have a correlation to be estimated.

\begin{verbatim}
# fm1 <- lmer(Reaction ~ Days + (Days | Subject), data=DT)
# summary(fm1) # or vc <- VarCorr(fm1); print(vc,comp=c("Variance"))

# Groups   Name Variance Std.Dev. Corr
# Subject (Intercept) 612.10   24.741
# Days                 35.07    5.922   0.07
# Residual             654.94   25.592
# Number of obs: 180, groups: Subject, 18

# fm2 <- mmer(Reaction ~ Days, 
# random = ~ Subject + vs(Days, Subject), 
# data=DT, tolparinv = 1e-6, verbose = FALSE)
# summary(fm2)$varcomp

# Groups   Name Variance Std.Dev. Corr
# Subject (Intercept) 612.10   24.741
# Days                 35.07    5.922   0.07
# Residual             654.94   25.592
# Number of obs: 180, groups: Subject, 18
\end{verbatim}
### no equivalence in sommer to find the correlation between the 2 vc
### this is the most similar which is equivalent to (intercept // slope)

```r
fm2 <- mmer(Reaction ~ Days,
    random = ~ Subject + vs(Days, Subject),
    data = DT, tolparinv = 1e-6, verbose = FALSE)
summary(fm2)
```

#### o) Random slopes with a different intercept

This is the a model where you assume that the random effect has different intercepts based on the levels of another variable but there’s not a main effect. The 0 in the intercept in lme4 assumes that random slopes interact with an intercept but without a main effect.

```r
library(orthopolynom)
## diagonal model
fm2 <- mmer(Reaction ~ Days,
    random = ~ vs(ds(Daysf), Subject),
    data = DT, tolparinv = 1e-6, verbose = FALSE)
summary(fm2)
```

### 4) Other models available in sommer but not in lme4

One of the strengths of sommer is the availability of other variance covariance structures. In this section we show 4 models available in sommer that are not available in lme4 and might be useful.

```r
library(orthopolynom)
## diagonal model
fm2 <- mmer(Reaction ~ Days,
    random = ~ vs(ds(Daysf), Subject),
    data = DT, tolparinv = 1e-6, verbose = FALSE)
summary(fm2)
```
### 0: Subject.Reaction-Reaction 139.5473 399.5095 0.3492967 Positive
### 1: Subject.Reaction-Reaction 196.8544 411.8262 0.4780037 Positive
### 2: Subject.Reaction-Reaction 0.0000 365.3178 0.0000000 Positive
### 3: Subject.Reaction-Reaction 556.0773 501.2665 1.1093445 Positive
### 4: Subject.Reaction-Reaction 855.2104 581.8190 1.4698910 Positive
### 5: Subject.Reaction-Reaction 1699.4269 820.4561 2.0713197 Positive
### 6: Subject.Reaction-Reaction 2910.8975 1175.7872 2.4757011 Positive
### 7: Subject.Reaction-Reaction 1539.6201 779.1437 1.9760413 Positive
### 8: Subject.Reaction-Reaction 2597.5337 1089.4522 2.3842568 Positive
### 9: Subject.Reaction-Reaction 3472.7108 1351.5702 2.5693899 Positive
### units.Reaction-Reaction 879.6958 247.4680 3.5547862 Positive

### unstructured model

```r
fm2 <- mmer(Reaction ~ Days, 
随机 = ~ vs(us(Daysf), Subject), 
data = DT, tolparinv = 1e-6, verbose = FALSE)

summary(fm2)$varcomp
```

<table>
<thead>
<tr>
<th>VarComp</th>
<th>VarCompSE</th>
<th>Zratio</th>
<th>Constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: Subject.Reaction-Reaction</td>
<td>402.6286</td>
<td>572.0867</td>
<td>0.7037894</td>
</tr>
<tr>
<td>1: 0: Subject.Reaction-Reaction</td>
<td>1022.5098</td>
<td>393.6922</td>
<td>2.5972314</td>
</tr>
<tr>
<td>1: 1: Subject.Reaction-Reaction</td>
<td>417.6460</td>
<td>521.3722</td>
<td>0.8010515</td>
</tr>
<tr>
<td>2: 0: Subject.Reaction-Reaction</td>
<td>540.3746</td>
<td>287.1704</td>
<td>1.8817210</td>
</tr>
<tr>
<td>2: 1: Subject.Reaction-Reaction</td>
<td>828.5156</td>
<td>325.7576</td>
<td>2.5433499</td>
</tr>
<tr>
<td>2: 2: Subject.Reaction-Reaction</td>
<td>0.0000</td>
<td>509.8962</td>
<td>0.0000000</td>
</tr>
<tr>
<td>3: 0: Subject.Reaction-Reaction</td>
<td>798.3750</td>
<td>397.0884</td>
<td>2.0105726</td>
</tr>
<tr>
<td>3: 1: Subject.Reaction-Reaction</td>
<td>1137.3863</td>
<td>443.9056</td>
<td>2.562256</td>
</tr>
<tr>
<td>3: 2: Subject.Reaction-Reaction</td>
<td>1057.0708</td>
<td>385.9026</td>
<td>2.7392162</td>
</tr>
<tr>
<td>3: 3: Subject.Reaction-Reaction</td>
<td>760.2469</td>
<td>436.7463</td>
<td>1.7407060</td>
</tr>
<tr>
<td>4: 0: Subject.Reaction-Reaction</td>
<td>757.8909</td>
<td>411.2464</td>
<td>1.8429119</td>
</tr>
<tr>
<td>4: 1: Subject.Reaction-Reaction</td>
<td>1039.6332</td>
<td>447.5192</td>
<td>2.3232148</td>
</tr>
<tr>
<td>4: 2: Subject.Reaction-Reaction</td>
<td>911.1369</td>
<td>377.9651</td>
<td>2.4106377</td>
</tr>
<tr>
<td>4: 3: Subject.Reaction-Reaction</td>
<td>1590.6778</td>
<td>566.5376</td>
<td>2.8077180</td>
</tr>
<tr>
<td>4: 4: Subject.Reaction-Reaction</td>
<td>957.1797</td>
<td>364.0599</td>
<td>2.6291817</td>
</tr>
<tr>
<td>5: 0: Subject.Reaction-Reaction</td>
<td>932.5247</td>
<td>516.7169</td>
<td>1.8047110</td>
</tr>
<tr>
<td>5: 1: Subject.Reaction-Reaction</td>
<td>1179.5219</td>
<td>547.9498</td>
<td>2.1526095</td>
</tr>
<tr>
<td>5: 2: Subject.Reaction-Reaction</td>
<td>859.1635</td>
<td>440.5250</td>
<td>1.9503173</td>
</tr>
<tr>
<td>5: 3: Subject.Reaction-Reaction</td>
<td>1672.9989</td>
<td>664.0846</td>
<td>2.5192556</td>
</tr>
<tr>
<td>5: 4: Subject.Reaction-Reaction</td>
<td>2003.0167</td>
<td>738.6399</td>
<td>2.7117633</td>
</tr>
<tr>
<td>6: 0: Subject.Reaction-Reaction</td>
<td>2067.9299</td>
<td>553.3254</td>
<td>3.7372765</td>
</tr>
<tr>
<td>6: 1: Subject.Reaction-Reaction</td>
<td>666.1077</td>
<td>565.7589</td>
<td>1.1773702</td>
</tr>
<tr>
<td>6: 2: Subject.Reaction-Reaction</td>
<td>850.9395</td>
<td>583.6190</td>
<td>1.4580394</td>
</tr>
<tr>
<td>6: 3: Subject.Reaction-Reaction</td>
<td>916.2375</td>
<td>504.0273</td>
<td>1.8178333</td>
</tr>
<tr>
<td>6: 4: Subject.Reaction-Reaction</td>
<td>1785.8432</td>
<td>750.7274</td>
<td>2.3788171</td>
</tr>
<tr>
<td>6: 5: Subject.Reaction-Reaction</td>
<td>2077.5064</td>
<td>822.0777</td>
<td>2.5271412</td>
</tr>
<tr>
<td>6: 6: Subject.Reaction-Reaction</td>
<td>2603.2823</td>
<td>1035.1406</td>
<td>2.5149070</td>
</tr>
<tr>
<td>7: 0: Subject.Reaction-Reaction</td>
<td>3123.2005</td>
<td>1049.0352</td>
<td>2.9772123</td>
</tr>
<tr>
<td>7: 1: Subject.Reaction-Reaction</td>
<td>932.8190</td>
<td>490.4744</td>
<td>1.9018709</td>
</tr>
<tr>
<td>7: 2: Subject.Reaction-Reaction</td>
<td>927.3416</td>
<td>492.7764</td>
<td>1.8818709</td>
</tr>
<tr>
<td>7: 3: Subject.Reaction-Reaction</td>
<td>924.0709</td>
<td>426.2387</td>
<td>2.1694602</td>
</tr>
<tr>
<td>7: 4: Subject.Reaction-Reaction</td>
<td>1282.8637</td>
<td>583.3415</td>
<td>2.1991642</td>
</tr>
<tr>
<td>7: 5: Subject.Reaction-Reaction</td>
<td>1549.9053</td>
<td>643.7083</td>
<td>2.4077575</td>
</tr>
<tr>
<td>7: 6: Subject.Reaction-Reaction</td>
<td>2306.0261</td>
<td>951.5128</td>
<td>2.4235367</td>
</tr>
<tr>
<td>7: 7: Subject.Reaction-Reaction</td>
<td>1669.8274</td>
<td>612.0081</td>
<td>2.7284398</td>
</tr>
</tbody>
</table>
## random regression (legendre polynomials)

```r
fm2 <- 
mer(Reaction ~ Days,
    random = ~ vs(leg(Days,1), Subject),
    data=DT, tolparinv = 1e-6, verbose = FALSE)
summary(fm2)$varcomp
```

<table>
<thead>
<tr>
<th></th>
<th>VarComp</th>
<th>VarCompSE</th>
<th>Zratio</th>
<th>Constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td>leg0:Subject.Reaction-Reaction</td>
<td>2817.4048</td>
<td>1011.23903</td>
<td>2.786092</td>
<td>Positive</td>
</tr>
<tr>
<td>leg1:Subject.Reaction-Reaction</td>
<td>473.4608</td>
<td>199.53635</td>
<td>2.372805</td>
<td>Positive</td>
</tr>
<tr>
<td>units.Reaction-Reaction</td>
<td>654.9433</td>
<td>77.18822</td>
<td>8.485016</td>
<td>Positive</td>
</tr>
</tbody>
</table>

## unstructured random regression (legendre)

```r
fm2 <- 
mer(Reaction ~ Days,
    random = ~ vs(us(leg(Days,1)), Subject),
    data=DT, tolparinv = 1e-6, verbose = FALSE)
summary(fm2)$varcomp
```

<table>
<thead>
<tr>
<th></th>
<th>VarComp</th>
<th>VarCompSE</th>
<th>Zratio</th>
<th>Constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td>leg0:Subject.Reaction-Reaction</td>
<td>2817.4056</td>
<td>1011.24156</td>
<td>2.786086</td>
<td>Positive</td>
</tr>
<tr>
<td>leg1:leg0:Subject.Reaction-Reaction</td>
<td>869.9590</td>
<td>381.02481</td>
<td>2.283208</td>
<td>Unconstr</td>
</tr>
<tr>
<td>leg1:Subject.Reaction-Reaction</td>
<td>473.4608</td>
<td>199.53612</td>
<td>2.372807</td>
<td>Positive</td>
</tr>
<tr>
<td>units.Reaction-Reaction</td>
<td>654.9428</td>
<td>77.18763</td>
<td>8.485075</td>
<td>Positive</td>
</tr>
</tbody>
</table>

**Final remarks**

Keep in mind that `sommer` uses the direct inversion (DI) algorithm which can be very slow for large datasets. The package is focused on problems of the type \( p > n \) (more random effect levels than observations) and models with dense covariance structures. For example, for an experiment with dense covariance structures with low-replication (i.e. 2000 records from 1000 individuals replicated twice with a covariance structure of 1000x1000), `sommer` will be faster than MME-based software. Also for genomic problems with large number of random effect levels, i.e. 300 individuals (\( n \)) with 100,000 genetic markers (\( p \)). For highly replicated trials with small number of individuals and covariance structures or \( n > p \) (i.e. 2000 records from 200 individuals replicated 10 times with covariance structure of 200x200), `asreml` or other MME-based algorithms will be much faster and we recommend you to opt for those software programs.
Literature


Covarrubias-Pazaran G. 2018. Software update: Moving the R package sommer to multivariate mixed models for genome-assisted prediction. doi: https://doi.org/10.1101/354639


