Applied Statistical Methods - Solution 10

Peter von Rohr

2022-05-16

Problem 1: Sire Model

Use the following dataset to predict breeding values using a sire model. The dataset is available from

```
## https://charlotte-ngs.github.io/asmss2022/data/asm_ped_sim_data.csv
```

Hints

- The variance component σ_s^2 of the sire effect can be assumed to be 2.25.
- The variance component σ_e^2 of the random resiudals is 36.
- Sex is modelled as a fixed effect.
- The sire pedigree relationship can be computed using the `pedigreemm` package.

Solution

- Read the data

```
s_ex10_p01_data_path <- "https://charlotte-ngs.github.io/asmss2022/data/asm_ped_sim_data.csv"
tbl_ex10_p01 <- readr::read_csv(s_ex10_p01_data_path)
```

```
## Rows: 8 Columns: 5

## -- Column specification --------------------------------------------------------------------------
## Delimiter: ","
## chr (1): SEX
## dbl (4): ID, SIRE, DAM, P

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

- Compute the inverse sire relationship matrix

```
vec_sire <- unique(tbl_ex10_p01$SIRE)
vec_sire <- vec_sire[!is.na(vec_sire)]
n_nr_sire <- length(vec_sire)
ped_sire <- pedigreemm::pedigree(sire = c(NA, NA, 2),
                                 dam = rep(NA, n_nr_sire),
                                 label = as.character(vec_sire))
mat_A_inv_sire <- as.matrix(pedigreemm::getAInv(ped = ped_sire))
mat_A_inv_sire
```

```
##   1         2          8
## 1 1  0.0000000  0.0000000
## 2 0  1.3333333 -0.6666667
## 8 0 -0.6666667  1.3333333
```

- Setup the mixed model equations. The mixed model equations are

$$\left[\begin{array}{cc} X^TX & X^TZ \\ Z^TX & Z^TZ + \lambda_s * A_s^{-1} \end{array}\right] \left[\begin{array}{c} \hat{b} \\ \hat{s} \end{array}\right] = \left[\begin{array}{c} X^Ty \\ Z^Ty \end{array}\right]$$

where $\lambda_s = \sigma_e^2/\sigma_s^2$.

The components of the mixed model equations are shown in the following table

Component	Description
X	Given in the data
Z	Given in the data
y	Given in the data
λ_s	Given by variance components
A_s^{-1}	Computed above

The matrix X

```
mat_X <- model.matrix(lm(P ~ 0 + SEX, data = tbl_ex10_p01))
attr(mat_X, "assign") <- NULL
attr(mat_X, "contrasts") <- NULL
colnames(mat_X) <- NULL
mat_X
```

```
##   [,1] [,2]
## 1    1    0
## 2    1    0
## 3    0    1
## 4    0    1
## 5    0    1
## 6    0    1
## 7    0    1
## 8    1    0
```

The matrix Z

```
mat_Z <- model.matrix(lm(P ~ 0 + as.factor(SIRE), data = tbl_ex10_p01))
attr(mat_Z, "assign") <- NULL
attr(mat_Z, "contrasts") <- NULL
colnames(mat_Z) <- NULL
mat_Z
```

```
##   [,1] [,2] [,3]
## 1    1    0    0
## 2    0    1    0
## 3    1    0    0
## 4    0    1    0
## 5    1    0    0
## 6    0    0    1
## 7    1    0    0
## 8    0    0    1
```

The vector y

```
vec_y <- tbl_ex10_p01$P
vec_y
```

```
## [1] 16.7 13.9 26.0  4.3 18.8  5.2  6.6 27.5
```

The mixed model equations are

```
mat_xtx <- crossprod(mat_X)
mat_xtz <- crossprod(mat_X, mat_Z)
mat_ztx <- t(mat_xtz)
lambda_s <- sigma_e2 / sigma_s2
mat_ztz_a_inv_lambda <- crossprod(mat_Z) + lambda_s * mat_A_inv_sire
mat_coef <- rbind(cbind(mat_xtx, mat_xtz), cbind(mat_ztx, mat_ztz_a_inv_lambda))
mat_xty <- crossprod(mat_X, vec_y)
mat_zty <- crossprod(mat_Z, vec_y)
mat_rhs <- rbind(mat_xty, mat_zty)
```

- Solve mixed model equations. The solution is obtained by

```
mat_sol_sire <- solve(mat_coef, mat_rhs)
mat_sol_sire
```

```
##         [,1]
##   19.4721453
##   11.9901384
## 1  0.6328720
## 2 -0.6878893
## 8 -0.2614187
```

The solution for the fixed effects are

```
mat_sol_sire[1:2,]
```

```
## 
## 19.47215 11.99014
```

The predicted breeding values are

```
mat_sol_sire[3:nrow(mat_sol_sire),]
```

```
##          1          2          8 
##  0.6328720 -0.6878893 -0.2614187
```

Problem 2: Animal Model

Use the same dataset as in Problem 1 to predict breeding values, but use an animal model instead of a sire model. The dataset is available from

```
## https://charlotte-ngs.github.io/asmss2022/data/asm_ped_sim_data.csv
```

Hints

- The variance component σ_u^2 of the breeding value can be assumed to be 9.
- The variance component σ_e^2 of the random resiudals is 36.
- Sex is modelled as a fixed effect.
- The numerator relationship matrix can be computed using the `pedigreemm` package.

Solution

- Read the data

```
s_ex10_p02_data_path <- "https://charlotte-ngs.github.io/asmss2022/data/asm_ped_sim_data.csv"
tbl_ex10_p02 <- readr::read_csv(s_ex10_p02_data_path)
```

```
## Rows: 8 Columns: 5
## -- Column specification --------------------------------------------------------------------------
## Delimiter: ","
## chr (1): SEX
## dbl (4): ID, SIRE, DAM, P
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

- Compute the inverse sire relationship matrix

```
ped <- pedigreemm::pedigree(sire = c(rep(NA, 4), tbl_ex10_p02$SIRE),
                            dam   = c(rep(NA, 4), tbl_ex10_p02$DAM),
                            label = as.character(c(1:4, tbl_ex10_p02$ID)))
mat_A_inv <- as.matrix(pedigreemm::getAInv(ped = ped))
mat_A_inv
```

```
##       1    2    3    4  5  6  7  8  9 10 11 12
## 1   3.0  0.0  0.5  0.5 -1  1 -1  0 -1  0 -1  0
## 2   0.0  2.0  0.5  0.5  0 -1  0 -1  0  0  0  0
## 3   0.5  0.5  2.0  0.0  0  0 -1 -1  0  0  0  0
## 4   0.5  0.5  0.0  2.0 -1 -1  0  0  0  0  0  0
## 5  -1.0  0.0  0.0 -1.0  3  0  0  1  0 -1  0 -1
## 6   1.0 -1.0  0.0 -1.0  0  3  0  0 -1  0 -1  0
## 7  -1.0  0.0 -1.0  0.0  0  0  2  0  0  0  0  0
## 8   0.0 -1.0 -1.0  0.0  1  0  0  3  0 -1  0 -1
## 9  -1.0  0.0  0.0  0.0  0 -1  0  0  2  0  0  0
## 10  0.0  0.0  0.0  0.0 -1  0  0 -1  0  2  0  0
## 11 -1.0  0.0  0.0  0.0  0 -1  0  0  0  0  2  0
## 12  0.0  0.0  0.0  0.0 -1  0  0 -1  0  0  0  2
```

- Setup the mixed model equations. The mixed model equations are

$$\left[\begin{array}{cc} X^TX & X^TZ \\ Z^TX & Z^TZ + \lambda * A^{-1} \end{array}\right] \left[\begin{array}{c} \hat{b} \\ \hat{u} \end{array}\right] = \left[\begin{array}{c} X^Ty \\ Z^Ty \end{array}\right]$$

where $\lambda = \sigma_e^2/\sigma_u^2$.

The components of the mixed model equations are shown in the following table

Component	Description
X	Given in the data
Z	Given in the data
y	Given in the data
λ	Given by variance components
A^{-1}	Computed above

The matrix X

```
mat_X <- model.matrix(lm(P ~ 0 + SEX, data = tbl_ex10_p02))
attr(mat_X, "assign") <- NULL
attr(mat_X, "contrasts") <- NULL
colnames(mat_X) <- NULL
mat_X
```

```
##   [,1] [,2]
## 1    1    0
## 2    1    0
## 3    0    1
## 4    0    1
## 5    0    1
## 6    0    1
## 7    0    1
## 8    1    0
```

The matrix Z

```
# model matrix from data
mat_Z <- model.matrix(lm(P ~ 0 + as.factor(ID), data = tbl_ex10_p02))
attr(mat_Z, "assign") <- NULL
attr(mat_Z, "contrasts") <- NULL
colnames(mat_Z) <- NULL
# add founders
mat_Z <- cbind(matrix(0, nrow = nrow(mat_Z), ncol = 4), mat_Z)
mat_Z
```

```
##   [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
## 1    0    0    0    0    1    0    0    0    0     0     0     0
## 2    0    0    0    0    0    1    0    0    0     0     0     0
## 3    0    0    0    0    0    0    1    0    0     0     0     0
## 4    0    0    0    0    0    0    0    1    0     0     0     0
## 5    0    0    0    0    0    0    0    0    1     0     0     0
## 6    0    0    0    0    0    0    0    0    0     1     0     0
## 7    0    0    0    0    0    0    0    0    0     0     1     0
## 8    0    0    0    0    0    0    0    0    0     0     0     1
```

The vector y

```
vec_y <- tbl_ex10_p02$P
vec_y
```

```
## [1] 16.7 13.9 26.0  4.3 18.8  5.2  6.6 27.5
```

The mixed model equations are

```
mat_xtx <- crossprod(mat_X)
mat_xtz <- crossprod(mat_X, mat_Z)
mat_ztx <- t(mat_xtz)
lambda <- sigma_e2 / sigma_u2
mat_ztz_a_inv_lambda <- crossprod(mat_Z) + lambda * mat_A_inv
mat_coef <- rbind(cbind(mat_xtx, mat_xtz), cbind(mat_ztx, mat_ztz_a_inv_lambda))
mat_xty <- crossprod(mat_X, vec_y)
mat_zty <- crossprod(mat_Z, vec_y)
mat_rhs <- rbind(mat_xty, mat_zty)
```

- Solve mixed model equations. The solution is obtained by

```
mat_sol <- solve(mat_coef, mat_rhs)
mat_sol
```

```
##              [,1]
##    19.7175571343
##    12.1523850711
```

```
## 1   1.2950766779
## 2  -1.2250000000
## 3   0.6784481962
## 4  -0.7485248741
## 5  -0.0007843862
## 6  -1.4612270230
## 7   2.4157460473
## 8  -1.0238113159
## 9   0.6647792832
## 10 -1.2278630978
## 11 -0.6907762724
## 12  0.4093400063
```

The solution for the fixed effects are

```
mat_sol[1:2,]
```

```
##
## 19.71756 12.15239
```

The predicted breeding values are

```
mat_sol[3:nrow(mat_sol),]
```

```
##             1             2             3             4             5             6             7
##  1.2950766779 -1.2250000000  0.6784481962 -0.7485248741 -0.0007843862 -1.4612270230  2.4157460473 -1
##             9            10            11            12
##  0.6647792832 -1.2278630978 -0.6907762724  0.4093400063
```

Problem 3: Model Comparison

Compare the order of the predicted breeding values for the sires from the sire model and from the animal model.

Solution

- Sire model

```
order(mat_sol_sire[3:nrow(mat_sol_sire)], decreasing = TRUE)
```

```
## [1] 1 3 2
```

- Animal model

```
order(mat_sol[3:nrow(mat_sol)], decreasing = TRUE)
```

```
##  [1]  7  1  3  9 12  5 11  4  8  2 10  6
```

The order of the sires is the same under both models