Employing asremlPlus, in conjunction with asreml, to calculate and use information criteria

Chris Brien
19 June, 2024

This vignette illustrates the facilities in asremlPlus (Brien, 2024), in conjunction with asreml (Butler et al., 2023), for calculating and using information. Here, asremlPlus and asreml are packages for the R Statistical Computing environment (R Core Team, 2024).

It is divided into the following main sections:

1. Set up the maximal model for this experiment
2. Obtaining information criteria for separate models
3. Obtaining information criteria for a prescribed sequence of model changes
4. Using information criteria to decide model changes

1. Set up the maximal model for this experiment

```r
library(knitr)
opts_chunk$set("tidy" = FALSE, comment = NA)
suppressMessages(library(asreml, quietly=TRUE))
```

```r
## Offline License checked out Wed Jun 19 15:46:12 2024
packageVersion("asreml")
```

```r
# [1] '4.2.0.332'

suppressMessages(library(asremlPlus))
packageVersion("asremlPlus")
```

```r
# [1] '4.4.34'

options(width = 100)
```

Get data available in asremlPlus

The data are from a 1976 spring wheat experiment and are taken from Gilmour et al. (1995). An analysis is presented in the asreml manual by Butler et al. (2023, Section 7.6), although they suggest that it is a barley experiment.
Fit the maximal model

In the following a model is fitted that has the terms that would be included for a balanced lattice. In addition, a term **WithinColPairs** has been included to allow for extraneous variation arising between pairs of adjacent lanes. Also, separable ar1 residual autocorrelation has been included. This model represents the maximal anticipated model,

```
max.asr <- asreml(yield ~ WithinColPairs + Variety,
                   random = ~ Rep/(Row + Column) + units,
                   residual = ~ ar1(Row):ar1(Column),
                   data=Wheat.dat)
```

<table>
<thead>
<tr>
<th>LogLik</th>
<th>Sigma2</th>
<th>DF</th>
<th>wall</th>
</tr>
</thead>
<tbody>
<tr>
<td>-717.4149</td>
<td>9206.931</td>
<td>124</td>
<td>15:46:13</td>
</tr>
<tr>
<td>-694.8752</td>
<td>26492.99</td>
<td>124</td>
<td>15:46:13</td>
</tr>
<tr>
<td>-694.1600</td>
<td>33101.80</td>
<td>124</td>
<td>15:46:13</td>
</tr>
<tr>
<td>-692.0020</td>
<td>36912.26</td>
<td>124</td>
<td>15:46:13</td>
</tr>
<tr>
<td>-691.7892</td>
<td>46701.51</td>
<td>124</td>
<td>15:46:13</td>
</tr>
<tr>
<td>-691.7749</td>
<td>47698.26</td>
<td>124</td>
<td>15:46:13</td>
</tr>
<tr>
<td>-691.7711</td>
<td>47041.85</td>
<td>124</td>
<td>15:46:13</td>
</tr>
</tbody>
</table>

Warning in asreml(yield ~ WithinColPairs + Variety, random = ~Rep/(Row + : Some components changed by more than 1% on the last iteration

The warning from `asreml` is probably due to a bound term.

Initialize a testing sequence by loading the current fit into an `asrtests` object

```
max.asrt <- as.asrtests(max.asr, NULL, NULL)
```

Check for and remove any boundary terms

```
max.asrt <- rmboundary(max.asrt)
summary(max.asrt$asreml.obj)$varcomp
```

<table>
<thead>
<tr>
<th>component</th>
<th>std.error</th>
<th>z.ratio</th>
<th>bound</th>
<th>%ch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rep:Row</td>
<td>4.293282e+03</td>
<td>3.199458e+03</td>
<td>1.3418779</td>
<td>P 0.0</td>
</tr>
<tr>
<td>Rep:Column</td>
<td>1.575689e+02</td>
<td>1.480357e+03</td>
<td>0.1064398</td>
<td>P 0.7</td>
</tr>
<tr>
<td>units</td>
<td>5.742689e+03</td>
<td>1.652457e+03</td>
<td>3.4752438</td>
<td>P 0.0</td>
</tr>
<tr>
<td>Row:Column!R</td>
<td>4.706787e+04</td>
<td>2.515832e+04</td>
<td>1.8708669</td>
<td>P 0.0</td>
</tr>
<tr>
<td>Row:Column!Row!cor</td>
<td>7.920301e-01</td>
<td>1.014691e-01</td>
<td>7.8056280</td>
<td>U 0.0</td>
</tr>
<tr>
<td>Row:Column!Column!cor</td>
<td>8.799559e-01</td>
<td>7.370402e-02</td>
<td>11.9390486</td>
<td>U 0.0</td>
</tr>
</tbody>
</table>
### Sequence of model investigations

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

<table>
<thead>
<tr>
<th>terms</th>
<th>DF</th>
<th>denDF</th>
<th>p</th>
<th>AIC</th>
<th>BIC</th>
<th>action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rep</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Boundary</td>
</tr>
</tbody>
</table>

Rep has been removed because it has been constrained to zero. Following the recommendation of Littel et al. (2006, p. 150), the bound on all variance components is set to unconstrained (U) using `setvariances.asreml` so as to avoid bias in the estimate of the residual variance. Alternatively, one could move Rep to the fixed model.

#### Unbind Rep, Row and Column components and reload into an asrtests object

```r
max.asr <- setvarianceterms(max.asr$call,
                            terms = c("Rep", "Rep:Row", "Rep:Column"),
                            bounds = "U")
```

```
ASReml Version 4.2 19/06/2024 15:46:14
LogLik Sigma2 DF wall
1 -724.1213 23034.14 124 15:46:14
2 -717.4149 9206.931 124 15:46:14 (2 restrained)
3 -694.8752 26492.99 124 15:46:14 (2 restrained)
4 -693.9744 33129.65 124 15:46:14 (1 restrained)
5 -692.8856 39662.12 124 15:46:14
6 -691.4276 53103.83 124 15:46:14
7 -691.2387 48092.17 124 15:46:14
8 -691.1808 47278.94 124 15:46:14
9 -691.1710 46850.98 124 15:46:14
10 -691.1700 46690.46 124 15:46:14
```

Warning in asreml(fixed = yield ~ WithinColPairs + Variety, random = ~Rep/(Row + : Some components changed by more than 1% on the last iteration

```r
max.asrt <- as.asrtests(max.asr, NULL, NULL)
max.asrt <- rmboundary(max.asrt)
summary(max.asrt$asreml.obj)$varcomp
```

<table>
<thead>
<tr>
<th>component</th>
<th>std.error</th>
<th>z.ratio</th>
<th>bound</th>
<th>%ch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rep</td>
<td>-2458.3485841 1.197491e+03</td>
<td>-2.0529167</td>
<td>U</td>
<td>0.0</td>
</tr>
<tr>
<td>Rep:Row</td>
<td>5008.7151486 3.401335e+03</td>
<td>1.4725732</td>
<td>U</td>
<td>0.0</td>
</tr>
<tr>
<td>Rep:Column</td>
<td>916.4641198 1.699576e+03</td>
<td>0.5392309</td>
<td>P</td>
<td>0.0</td>
</tr>
<tr>
<td>units</td>
<td>5959.0220817 1.609649e+03</td>
<td>3.7020634</td>
<td>P</td>
<td>0.0</td>
</tr>
<tr>
<td>Row:Column!R</td>
<td>46637.6303429 2.724392e+04</td>
<td>1.7118545</td>
<td>P</td>
<td>0.0</td>
</tr>
<tr>
<td>Row:Column!Row!cor</td>
<td>0.8150590 1.000281e-01</td>
<td>8.1483012</td>
<td>U</td>
<td>0.0</td>
</tr>
<tr>
<td>Row:Column!Column!cor</td>
<td>0.8856824 7.492514e-02</td>
<td>11.8208968</td>
<td>U</td>
<td>0.0</td>
</tr>
</tbody>
</table>
print(max.asrt, which = "testsummary")

### Sequence of model investigations

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

[1] terms DF denDF p AIC BIC action
<0 rows> (or 0-length row.names)

Now the Rep component estimate is negative.
The test.summary output shows that no changes have been made to the model loaded using as.asrtests. The pseudo-anova table shows that Varieties are highly significant ($p < 0.001$)

2. Obtaining information criteria for separate models

The method infoCriteria has two methods for calculating information criteria. One, infoCriteria.asreml, is a method for asreml objects and the other, infoCriteria.list, if for 'listobjects, the components of the list being asreml' objects.

**Single models**

Firstly, infoCriteria is called with the default IClikelihood, which is REML. Then it is called with IClikelihood set to full (Verbyla, 2019).

```r
infoCriteria(max.asr)
```

<table>
<thead>
<tr>
<th>fixedDF</th>
<th>varDF</th>
<th>NBound</th>
<th>AIC</th>
<th>BIC</th>
<th>loglik</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>7</td>
<td>1396.34</td>
<td>1416.082</td>
<td>-691.17</td>
</tr>
</tbody>
</table>

```r
infoCriteria(max.asr, IClikelihood = "full")
```

ASReml Version 4.2 19/06/2024 15:46:15

<table>
<thead>
<tr>
<th>LogLik</th>
<th>Sigma2</th>
<th>DF</th>
<th>wall</th>
</tr>
</thead>
<tbody>
<tr>
<td>-691.1700</td>
<td>46627.05</td>
<td>124</td>
<td>15:46:15</td>
</tr>
</tbody>
</table>

Warning in asreml(fixed = yield ~ WithinColPairs + Variety, random = -Rep/(Row + : Log-likelihood not converged

<table>
<thead>
<tr>
<th>fixedDF</th>
<th>varDF</th>
<th>NBound</th>
<th>AIC</th>
<th>BIC</th>
<th>loglik</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26</td>
<td>7</td>
<td>1647.191</td>
<td>1746.542</td>
<td>-790.5957</td>
</tr>
</tbody>
</table>
A list of models

Now, a second model, from which the `withinColPairs` term has been omitted, is fitted; to be consistent, the variance components are unconstrained using `setvariances.asreml`. Then the `asreml` objects for this model and the maximal model are combined into a `list` and a `data.frame` produced that includes their information criteria.

```r
m1.asr <- asreml(yield ~ Variety,
                 random = ~ Rep/(Row + Column) + units,
                 residual = ~ ar1(Row):ar1(Column),
                 data=Wheat.dat)
```

```
ASReml Version 4.2 19/06/2024 15:46:15
LogLik Sigma2 DF wall
1  -727.7742  22898.99 125 15:46:15
2  -721.0966  9190.303 125 15:46:15 ( 2 restrained)
3  -698.3135  26671.76 125 15:46:15 ( 2 restrained)
4  -697.5170  32677.28 125 15:46:15 ( 1 restrained)
5  -695.4192  36662.76 125 15:46:15 ( 1 restrained)
6  -695.2077  46263.96 125 15:46:15 ( 2 restrained)
7  -695.1975  46156.63 125 15:46:15
8  -695.1906  46630.21 125 15:46:15
```

Warning in asreml(yield ~ Variety, random = ~Rep/(Row + Column) + units, : Some components changed by more than 1% on the last iteration

```r
m1.asr <- setvariances(m1.asr$call,
                        terms = c("Rep", "Rep:Row", "Rep:Column"),
                        bounds = "U")
```

```
ASReml Version 4.2 19/06/2024 15:46:15
LogLik Sigma2 DF wall
1  -727.7742  22898.99 125 15:46:15
2  -721.0966  9190.303 125 15:46:15 ( 2 restrained)
3  -698.3135  26671.76 125 15:46:15 ( 2 restrained)
4  -697.5170  32677.28 125 15:46:15 ( 1 restrained)
5  -697.3331  32689.33 125 15:46:15 ( 1 restrained)
6  -697.0164  39975.97 125 15:46:15
7  -695.0695  54825.30 125 15:46:15
8  -694.7571  47637.20 125 15:46:15
9  -694.6436  46775.41 125 15:46:15
10 -694.6181  45940.69 125 15:46:15
```

Warning in asreml(fixed = yield ~ Variety, random = ~Rep/(Row + Column) + : Some components changed by more than 1% on the last iteration

```r
mods <- list(max = max.asr, m1 = m1.asr)
ic <- infoCriteria(mods, ICLikelihood = "full")
print(ic)
```

```
fixedDF varDF NBound    AIC    BIC    loglik
max   26   7     0 1647.191 1746.542 -790.5957
m1    25   7     0 1645.318 1741.658 -790.6588
```
3. Obtaining information criteria for a prescribed sequence of model changes

The use of `changeTerms.asrtests` is demonstrated for a sequence of models, starting with the maximal model.

Drop the term for within Column pairs (a post hoc factor)

```r
current.asrt <- as.asrtests(max.asrt$asreml.obj, NULL, NULL,
                         label = "Maximal model", IClikelihood = "full")
```

> Warning in asreml(fixed = yield ~ WithinColPairs + Variety, random = ~Rep/(Row + : Log-likelihood not converged

```r
current.asrt <- changeTerms(current.asrt, dropFixed = "WithinColPairs",
                          label = "Drop withinColPairs", IClikelihood = "full")
```

> Warning in asreml(fixed = yield ~ Variety, random = ~Rep + units + Rep:Row + : Some components changed by more than 1% on the last iteration

> Warning in asreml(fixed = yield ~ Variety, random = ~Rep + units + Rep:Row + : Some components changed by more than 1% on the last iteration

```r
print(current.asrt, which = "testsummary", omit.columns = "p")
```

### Sequence of model investigations

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

<table>
<thead>
<tr>
<th>terms</th>
<th>DF</th>
<th>denDF</th>
<th>AIC</th>
<th>BIC</th>
<th>action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximal model</td>
<td>26</td>
<td>7</td>
<td>1647.191</td>
<td>1746.542</td>
<td>Starting model</td>
</tr>
<tr>
<td>Drop withinColPairs</td>
<td>25</td>
<td>7</td>
<td>1645.325</td>
<td>1741.666</td>
<td>Changed fixed</td>
</tr>
</tbody>
</table>

So the same values of the information criteria have been obtained as when `infoCriteria.list` was used on a list containing the `asreml` objects for the two models. The differences is that here there is ultimately only one fitted model, the model stored in the `asreml` object in the `asrtests` object named `current.asrt`: this is the model with `withinColPairs` omitted.

Note this use of the `omit.columns` argument from `print.test.summary` to omit the irrelevant column `p` from the `test.summary`.

Drop nugget term

```r
current.asrt <- changeTerms(current.asrt, dropRandom = "units",
                          label = "Drop units", IClikelihood = "full")
```
Warning in asreml(fixed = yield ~ Variety, random = ~Rep + Rep:Row + Rep:Column, : Some components changed by more than 1% on the last iteration
Warning in asreml(fixed = yield ~ Variety, random = ~Rep + Rep:Row + Rep:Column, : Some components changed by more than 1% on the last iteration

Check Row autocorrelation

```r
current.asrt <- changeTerms(current.asrt, newResidual = "Row:ar1(Column)",
                           label="Row autocorrelation", IClelihood = "full")
print(current.asrt, which = "testsummary", omit.columns = "p")
```

#### Sequence of model investigations

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

<table>
<thead>
<tr>
<th>terms</th>
<th>DF</th>
<th>denDF</th>
<th>AIC</th>
<th>BIC</th>
<th>action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximal model</td>
<td>26</td>
<td>7</td>
<td>1647.191</td>
<td>1746.542</td>
<td>Starting model</td>
</tr>
<tr>
<td>Drop withinColPairs</td>
<td>25</td>
<td>7</td>
<td>1645.325</td>
<td>1741.666</td>
<td>Changed fixed</td>
</tr>
<tr>
<td>Drop units 25</td>
<td>25</td>
<td>6</td>
<td>1650.126</td>
<td>1743.456</td>
<td>Changed random</td>
</tr>
<tr>
<td>Row autocorrelation 25</td>
<td>25</td>
<td>5</td>
<td>1660.882</td>
<td>1751.201</td>
<td>Changed residual</td>
</tr>
</tbody>
</table>

### Sequence of model investigations

4. Using information criteria to decide model changes

This section illustrates the use of `changeModelOnIC.asrtests` to decide between consecutive models in a sequence of models. The default information criterion to use for this is the \textbf{AIC}. However, \textbf{which.IC} can be used to specify the use of the \textbf{BIC} or both. Here we use the \textbf{AIC} and the full likelihood.

Check the term for within Column pairs (a post hoc factor)

As before, we start with the maximal model, in which the variance components have been unconstrained and look to decide whether of not to drop the withinColPairs term.

```r
current.asrt <- as.asrtests(max.asrt$asreml.obj, NULL, NULL,
                           label = "Maximal model", IClelihood = "full")
```

Warning in asreml(fixed = yield ~ WithinColPairs + Variety, random = ~Rep/(Row + : Log-likelihood not converged

```r
current.asrt <- iterate(current.asrt)
current.asrt <- changeModelOnIC(current.asrt, dropFixed = "WithinColPairs",
                                 label = "withinColPairs",
                                 IClelihood = "full", which.IC = "AIC",
                                 allow.unconverged = FALSE)
```

Warning in asreml(fixed = yield ~ Variety, random = ~Rep + units + Rep:Row + : Some components changed by more than 1% on the last iteration
Warning in asreml(fixed = yield ~ Variety, random = ~Rep + units + Rep:Row + : Some components changed by more than 1% on the last iteration

\texttt{print(current.asrt, which = "testsummary", omit.columns = "p")}

#### Sequence of model investigations

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

<table>
<thead>
<tr>
<th>terms</th>
<th>DF</th>
<th>denDF</th>
<th>AIC</th>
<th>BIC</th>
<th>action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximal model</td>
<td>26</td>
<td>7</td>
<td>1647.191452</td>
<td>1746.542417</td>
<td>Starting model</td>
</tr>
<tr>
<td>withinColPairs</td>
<td>-1</td>
<td>0</td>
<td>-1.866103</td>
<td>-4.876738</td>
<td>Swapped</td>
</tr>
</tbody>
</table>

Given the warning about a lack of convergence, we use \texttt{iterate.asrtests} to perform additional iterations of the fitting process. It seems that it was successful.

It can be seen from the \texttt{test.summary} that the term has been swapped out and this has the effect of reducing the number of fixed parameters by one and makes no change to the variance parameters.

Check the nugget term

\texttt{current.asrt <- changeModelOnIC(current.asrt, dropRandom = "units", label = "units", IClikelihood = "full", allow.unconverged = FALSE)}

Warning in asreml(fixed = yield ~ Variety, random = ~Rep + Rep:Row + Rep:Column, : Some components changed by more than 1% on the last iteration
Warning in asreml(fixed = yield ~ Variety, random = ~Rep + Rep:Row + Rep:Column, : Some components changed by more than 1% on the last iteration

Check Row autocorrelation

\texttt{current.asrt <- changeModelOnIC(current.asrt, newResidual = "Row:ar1(Column)", label="Row autocorrelation", IClikelihood = "full", allow.unconverged = FALSE)}

Warning in asreml(fixed = yield ~ Variety, random = ~Rep + units + Rep:Row + : Log-likelihood not converged
Warning in asreml(fixed = yield ~ Variety, random = ~Rep + units + Rep:Row + : Some components changed by more than 1% on the last iteration
Warning in asreml(fixed = yield ~ Variety, random = ~Rep + units + Rep:Row + : Log-likelihood not converged
Warning in asreml(fixed = yield ~ Variety, random = ~Rep + units + Rep:Row + : Some components changed by more than 1% on the last iteration
Warning in newfit.asreml(asreml.obj, fixed. = fix.form, random. = ran.form, :
Check Column autocorrelation (depends on whether Row autocorrelation retained)

```r
last.action <- current.asrt$test.summary$terms == "Row autocorrelation"
if (grepl("Unswapped", last.action, fixed = TRUE) |
    grepl("Unchanged", last.action, fixed = TRUE))
current.asrt <- changeModelOnIC(current.asrt, newResidual = "ar1(Row):Column",
    label="Col autocorrelation", IClikelihood = "full",
    allow.unconverged = FALSE)
else
current.asrt <- testresidual(current.asrt, newResidual = "Row:Column",
    label="Col autocorrelation", IClikelihood = "full",
    allow.unconverged = FALSE)
```

Warning in infoCriteria.asreml(asreml.obj, IClikelihood = ic.lik, bound.exclusions = bound.exclusions):
Row:Column!Row!cor

Warning in rmboundary.asrtests(as.asrtests(asreml.obj, wald.tab, test.summary, : In analysing yield, est
Row:Column!Row!cor

Warning in infoCriteria.asreml(new.asrtests.obj$asreml.obj, IClikelihood = ic.lik, : The following bound
Row:Column!Row!cor

Output the results

```r
print(current.asrt, which = "test", omit.columns = "p")
```

```r
### Sequence of model investigations

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

<table>
<thead>
<tr>
<th>terms</th>
<th>DF</th>
<th>denDF</th>
<th>AIC</th>
<th>BIC</th>
<th>action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximal model</td>
<td>26</td>
<td>7</td>
<td>1.647191e+03</td>
<td>1.746542e+03</td>
<td>Starting model</td>
</tr>
<tr>
<td>withinColPairs</td>
<td>-1</td>
<td>0</td>
<td>-1.866103e+00</td>
<td>-4.876738e+00</td>
<td>Swapped</td>
</tr>
<tr>
<td>units</td>
<td>0</td>
<td>-1</td>
<td>4.801053e+00</td>
<td>1.790418e+00</td>
<td>Unswapped</td>
</tr>
<tr>
<td>Row autocorrelation</td>
<td>0</td>
<td>0</td>
<td>-7.342295e+03</td>
<td>-7.342295e-03</td>
<td>Unchanged - new unconverged</td>
</tr>
<tr>
<td>Col autocorrelation</td>
<td>0</td>
<td>-2</td>
<td>1.947985e+01</td>
<td>1.345858e+01</td>
<td>Unswapped</td>
</tr>
</tbody>
</table>

summary(current.asrt$asreml.obj)$varcomp

<table>
<thead>
<tr>
<th>component</th>
<th>std.error</th>
<th>z.ratio</th>
<th>bound</th>
<th>%ch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rep</td>
<td>-2391.894799</td>
<td>1.194671e+03</td>
<td>-2.002136</td>
<td>U</td>
</tr>
<tr>
<td>Rep:Row</td>
<td>5035.482349</td>
<td>3.406065e+03</td>
<td>1.478387</td>
<td>U</td>
</tr>
</tbody>
</table>
The test.summary shows us that the model without the autocorrelation failed to converge and so no change was made to the model. It, and the messages from checking the Column autocorrelation, also show us that the omission of the Column autocorrelation resulted in the Row autocorrelation becoming bound. That is, dropping the Column autocorrelation resulted in the dropping of two variance parameters.

The function printFormulae.asreml is used to display the fitted model.

```r
printFormulae(current.asrt$asreml.obj)
```

#### Formulae from asreml object

```r
fixed: yield ~ Variety
random: ~ Rep + units + Rep:Row + Rep:Column
residual: ~ ar1(Rep):ar1(Column)
```

References


