Does the "AI" in AI–REML stand for Artificial Intelligence?

S. D. Kachman
Department of Biometry
University of Nebraska–Lincoln

What will I talk about?

- Introduction
- Mixed Models
- REML
- Alternative Estimators
- Computational Shortcuts
- AI-REML

Does the “AI” in AI–REML stand for Artificial Intelligence?

- Short example

- Summary

Introduction

- Some of the many challenges facing Animal Breeders
  - Analysis of large messy data sets
  - Increasingly complex models
  - Computationally intensive procedures

- Larger and faster computers
  - If you wait long enough
  - There will always be a larger problem

- Computing strategies

– Most bang for the cycle?
– Several methods of obtaining REML estimates
– Sparse matrix techniques
– Reduced animal model

Mixed Models

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$$

$$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{G} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{R} \end{pmatrix}\right]$$

Variance Components

$$\begin{pmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{pmatrix}$$

- $\boldsymbol{G}$ and $\boldsymbol{R}$ need to be estimated
- Function of a set of variance components $\boldsymbol{\sigma}$

REML

- Restricted/Residual Maximum Likelihood

- Find the set of parameters which maximize the likelihood of the residuals

Residuals $\boldsymbol{K}$'$\boldsymbol{y}$

$$L(\boldsymbol{K}'y;\boldsymbol{\sigma}) = -\frac{1}{2}[\ln|\boldsymbol{V}| + \ln|\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}| + \boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}]$$

where

$$\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}$$

Alternative Estimators

- DF-REML
  - Evaluate the log-likelihood at many different places until a maximum is found
- EM-REML
  - If we had $\boldsymbol{u}$ and $\boldsymbol{e}$ it would be easy
  - Take the expected value of "$\boldsymbol{u}$" and "$\boldsymbol{e}$" given $\boldsymbol{y}$ and $\boldsymbol{\sigma}$ and maximize
- Derivative based methods

Derivative based methods

$$L(\boldsymbol{K}'y;\boldsymbol{\sigma}) = -\frac{1}{2}[\ln|\boldsymbol{V}| + \ln|\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}| + \boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}]$$

Score function

$$\frac{\partial L}{\partial \sigma_i} = -\frac{1}{2}[\mathrm{tr}(\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_i}) - \boldsymbol{y}'\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_i}\boldsymbol{P}\boldsymbol{y}]$$

Set this equal to zero and solve.

$$\frac{\partial \boldsymbol{V}}{\partial \sigma_i} = \boldsymbol{Z}\frac{\partial \boldsymbol{G}}{\partial \sigma_i}\boldsymbol{Z}' + \frac{\partial \boldsymbol{R}}{\partial \sigma_i}$$

$$\begin{aligned}\frac{\partial \boldsymbol{V}}{\partial \sigma_i}\boldsymbol{P}\boldsymbol{y} &= \boldsymbol{Z}\frac{\partial \boldsymbol{G}}{\partial \sigma_i}\boldsymbol{G}^{-1}\widehat{\boldsymbol{u}} \\ &\quad + \frac{\partial \boldsymbol{R}}{\partial \sigma_i}\boldsymbol{R}^{-1}\widehat{\boldsymbol{e}}\end{aligned}$$

$$\begin{aligned}\boldsymbol{y}'\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_i}\boldsymbol{P}\boldsymbol{y} &= \widehat{\boldsymbol{u}}'\boldsymbol{G}^{-1}\frac{\partial \boldsymbol{G}}{\partial \sigma_i}\boldsymbol{G}^{-1}\widehat{\boldsymbol{u}} \\ &\quad + \widehat{\boldsymbol{e}}'\boldsymbol{R}^{-1}\frac{\partial \boldsymbol{R}}{\partial \sigma_i}\boldsymbol{R}^{-1}\widehat{\boldsymbol{e}}\end{aligned}$$

$$
\begin{aligned}
\mathrm{tr}(\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_i}) &= \mathrm{tr}(\boldsymbol{P}\frac{\partial \boldsymbol{R}}{\partial \sigma_i} + \boldsymbol{P}\boldsymbol{Z}\frac{\partial \boldsymbol{G}}{\partial \sigma_i}\boldsymbol{Z}') \\
&= \mathrm{tr}([\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\,\mathrm{var}(\widehat{\boldsymbol{e}} - \boldsymbol{e})\boldsymbol{R}^{-1}]\frac{\partial \boldsymbol{R}}{\partial \sigma_i}) \\
&\quad + \mathrm{tr}([\boldsymbol{G}^{-1} \\
&\qquad - \boldsymbol{G}^{-1}\,\mathrm{var}(\widehat{\boldsymbol{u}} - \boldsymbol{u})\boldsymbol{G}^{-1}]\frac{\partial \boldsymbol{G}}{\partial \sigma_i})
\end{aligned}
$$

Information matrix

- Newton Raphson

$$-\left(\frac{\partial^2 L}{\partial \boldsymbol{\sigma} \partial \boldsymbol{\sigma}'}\right)(\Delta \boldsymbol{\sigma}) = \frac{\partial L}{\partial \boldsymbol{\sigma}}$$

- Fisher's method of Scoring

$$-\mathrm{E}\left(\frac{\partial^2 L}{\partial \boldsymbol{\sigma} \partial \boldsymbol{\sigma}'}\right)(\Delta \boldsymbol{\sigma}) = \frac{\partial L}{\partial \boldsymbol{\sigma}}$$

- Assuming the second partials of $\boldsymbol{V}$ with respect to the variance components are zero.

$$\frac{\partial^2 L}{\partial \boldsymbol{\sigma} \partial \boldsymbol{\sigma}'} = -\frac{1}{2}[-\operatorname{tr}(\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_i}\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_j}) + 2\boldsymbol{y}'\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_i}\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_j}\boldsymbol{P}\boldsymbol{y}]$$

$$\mathrm{E}\left[\frac{\partial^2 L}{\partial \boldsymbol{\sigma} \partial \boldsymbol{\sigma}'}\right] = -\frac{1}{2}[\operatorname{tr}(\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_i}\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_j})]$$

- Quadratic form in $\widehat{\boldsymbol{u}}$ and $\widehat{\boldsymbol{e}}$

$$\boldsymbol{f}_i = \boldsymbol{Z}\frac{\partial \boldsymbol{G}}{\partial \sigma_i}\boldsymbol{G}^{-1}\widehat{\boldsymbol{u}} + \frac{\partial \boldsymbol{R}}{\partial \sigma_i}\boldsymbol{R}^{-1}\widehat{\boldsymbol{e}}$$

- Also involves

$$\mathrm{tr}(\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_i}\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_j})$$

- Occurs in NR with a $\frac{1}{2}$ and in FS with a $-\frac{1}{2}$

AI-REML

- Newton Raphson tends to be quicker
- Newton Raphson may go in the wrong direction
  - Information matrix does not have to be positive semi-definite
- Fisher's method of scoring will at least go uphill
  - may also go up the other side
- Both involve a complicated trace

Does the "AI" in AI–REML stand for Artificial Intelligence?

- If we average them the complicated trace drops out and we are left with a number of quadratic forms to evaluate

Average information

$$-\frac{1}{2}\boldsymbol{y}'\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_i}\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_j}\boldsymbol{P}\boldsymbol{y}$$

- Which after a little simplification reduces to

$$
\begin{aligned}
\boldsymbol{I}_A &= \boldsymbol{F}'\boldsymbol{P}\boldsymbol{F} \\
\boldsymbol{F} &= \boldsymbol{Z}\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{\sigma}}\boldsymbol{G}^{-1}\widehat{\boldsymbol{u}} + \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{\sigma}}\boldsymbol{R}^{-1}\widehat{\boldsymbol{e}} \\
\boldsymbol{I}_A &= \boldsymbol{F}'\boldsymbol{R}^{-1}\boldsymbol{F} - \begin{pmatrix} \widehat{\boldsymbol{\beta}}_F \\ \widehat{\boldsymbol{u}}_F \end{pmatrix}' \begin{pmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{F} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{F} \end{pmatrix}
\end{aligned}
$$

- $\boldsymbol{I}_A$ is a positive semi-definite matrix

Algorithm

$$(\boldsymbol{F}'\boldsymbol{P}\boldsymbol{F})\Delta\sigma = \boldsymbol{F}'\boldsymbol{P}\boldsymbol{y} - \{\mathrm{tr}(\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma_i})\}$$

- A number of quadratic forms
- Solve the mixed model equations once for the original data, and once for each of the variance components
- Done correctly the subsequent solves are not very expensive computationally

Small Example

- Crimson Clover Example
- Treatment design was a 2^3 Factorial
- Seven replicates each consisting of four plants
- Plants were divided in half with a different treatment applied to each half
- Dependent variable was hard seed percentage

$$y_{rtp} = rep_r + \tau_t + plant_p(rep)_r + e_{rtp}$$

$$\begin{pmatrix} \boldsymbol{rep} \\ \boldsymbol{plant} \\ \boldsymbol{e} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{I}_7\sigma_r^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{28}\sigma_p^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I}_{56}\sigma_e^2 \end{pmatrix}\right]$$

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_r^2 & \sigma_p^2 & \sigma_e^2 \end{pmatrix}$$

Program

```
sol=lhsi*rhs;
u1=sol(p+1:p+q1,:);
u2=sol(p+q1+1:p+q1+q2,:);
e=y-w*sol;

f0=e/sig0;
f1=z1*u1/sig1;
f2=z2*u2/sig2;
F=<f0 f1 f2>;


tr1=sum(diag(lhsi((p+1):(p+q1),..
```

```
  (p+1):(p+q1))));
tr2=sum(diag(lhsi((p+q1+1):..
   (p+q1+q2),(p+q1+1):(p+q1+q2))));
ss1=q1/sig1-(tr1+u1'*u1)/(sig1*sig1);
ss2=q2/sig2-(tr2+u2'*u2)/(sig2*sig2);
ss0=((n-(p+q1+q2))+..
  (tr1/sig1+tr2/sig2))/sig0-..
e'*e/(sig0*sig0);
ss=-.5*<ss0;ss1;ss2>;

Flhs=w'*f/sig0;
fsol=lhsi*Flhs;
Inf=F'*F*(1/sig0)-fsol'*flhs;
```

```
Inf=.5*Inf;

ascv=inv(inf)

delt=ascv*ss
var=<sig0;sig1;sig2>+delt
```

Results

- Starting with 1, 1, and 1

- Converged after 12 iterations

$$\widehat{\sigma}_r^2 = 47 \pm 63$$
$$\widehat{\sigma}_p^2 = 184 \pm 74$$
$$\widehat{\sigma}_e^2 = 89 \pm 27$$

Summary

- The "AI" in AI-REML stands for average information
- Uses the average of the observed and expected information
- Computationally less expensive/iteration than either NR or FS
- Appears to converge fairly quickly
- Unlike NR it will always head in a "reasonable" direction
- Generalizes easily to the case when $\boldsymbol{V} \neq \sum \boldsymbol{V}_i \sigma_i$

Does the "AI" in AI–REML stand for Artificial Intelligence?