Covariance Matrices

LRS

CGIL

July-Aug 2012

Balanced AOV-Fisher

Balanced Data

Every smallest subclass of the model is filled and has the same number of observations.

- All factors in the model were random, except for μ.
- Unbiased estimators, simple to calculate, high chance of being negative.

Henderson, 1953

- Published Methods 1, 2, and 3.
- Method 1, all random factors, unbalanced data
- Method 2, some fixed factors, no interactions with random factors
- Method 3, Fitting Constants Method, full model and submodels
- All methods unbiased, negative estimates possible, easy to calculate for those times using desk calculators.

Hartley and Rao, 1967

- Published Maximum Likelihood Method
- Biased estimates, forced to be positive
- Iterative
- More accurate than unbiased methods

1971

- Restricted Maximum Likelihood presented; Less biased than ML; Iterative; Accurate; Patterson and Thompson.
- Minimum Variance Quadratic Unbiased Estimation presented by C. R. Rao; if estimates stay positive then same as REML

Gianola et al, 1980's

- Introduced Bayesian methodology; Gibbs Sampling; MCMC methods
- Biased estimates (positive), logical approach

Today, 2012

- Everyone uses either REML or Bayesian Methods, or both.
- Will only present these two methods.

Intro

Let **y** be a random vector variable of length n, then the *variance-covariance* matrix of **y** is:

$$\begin{aligned} Var(\mathbf{y}) &= E(\mathbf{y}\mathbf{y}') - E(\mathbf{y})E(\mathbf{y}') \\ &= \begin{pmatrix} \sigma^2_{y_1} & \sigma_{y_1 y_2} & \cdots & \sigma_{y_1 y_n} \\ \sigma_{y_1 y_2} & \sigma^2_{y_2} & \cdots & \sigma_{y_2 y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{y_1 y_n} & \sigma_{y_2 y_n} & \cdots & \sigma^2_{y_n} \end{pmatrix} \\ &= \mathbf{V} \end{aligned}$$

A variance-covariance (VCV) matrix is square, symmetric and should always be positive definite, i.e. all of the eigenvalues must be positive.

Quadratic Forms

A general quadratic form is

$$\mathbf{y}'\mathbf{Q}\mathbf{y}$$

Usually $\mathbf{Q}$ is a symmetric matrix, but not necessarily positive definite.
Examples of $\mathbf{Q}$

- $\mathbf{Q} = \mathbf{I}$, then $\mathbf{y}'\mathbf{Q}\mathbf{y} = \mathbf{y}'\mathbf{y}$ which is a total sum of squares of the elements in $\mathbf{y}$.
- $\mathbf{Q} = \mathbf{J}(1/n)$, then $\mathbf{y}'\mathbf{Q}\mathbf{y} = \mathbf{y}'\mathbf{J}\mathbf{y}(1/n)$ where n is the length of $\mathbf{y}$.

$$\mathbf{y}'\mathbf{J}\mathbf{y} = (\mathbf{y}'\mathbf{1})(\mathbf{1}'\mathbf{y})$$

- $\mathbf{Q} = \left(\mathbf{I} - \mathbf{J}(1/n)\right)/(n-1)$, then $\mathbf{y}'\mathbf{Q}\mathbf{y}$ gives the variance of the elements in $\mathbf{y}$, σ_y^2.

$Var(\mathbf{y})$

The (co)variance matrix of **y** is

$$\begin{aligned}\mathbf{V} &= \sum_{i=1}^{s} \mathbf{Z}_i \mathbf{G}_i \mathbf{Z}_i' \sigma_i^2 + \mathbf{R}\sigma_0^2 \\ &= \mathbf{ZGZ}' + \mathbf{R}.\end{aligned}$$

The inverse of **V** is

$$\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}$$

Proof is in notes.

Useful Results

- $$| \mathbf{A}k | = k^m | \mathbf{A} |$$
- For general square matrices, **M** and **U**, of the same order then
$$| \mathbf{MU} | = | \mathbf{M} | \, | \mathbf{U} |$$

Useful Results II

- $\mathbf{A}$ and $\mathbf{D}$ are square and non-singular

$$\begin{vmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = \mid \mathbf{A} \mid \mid \mathbf{D} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \mid = \mid \mathbf{D} \mid \mid \mathbf{A} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C} \mid$$

If $\mathbf{A} = \mathbf{I}$ and $\mathbf{D} = \mathbf{I}$, and $\mid \mathbf{I} \mid = 1$,

$$\begin{aligned} \mid \mathbf{I} + \mathbf{C}\mathbf{B} \mid &= \mid \mathbf{I} + \mathbf{B}\mathbf{C} \mid \\ &= \mid \mathbf{I} + \mathbf{B}'\mathbf{C}' \mid \\ &= \mid \mathbf{I} + \mathbf{C}'\mathbf{B}' \mid . \end{aligned}$$

Useful Results III

$$
\begin{aligned}
|\,\mathbf{V}\,| &= |\,\mathbf{R}+\mathbf{ZGZ}'\,| \\
&= |\,\mathbf{R}(\mathbf{I}+\mathbf{R}^{-1}\mathbf{ZGZ}')\,| \\
&= |\,\mathbf{R}\,|\,|\,\mathbf{I}+\mathbf{R}^{-1}\mathbf{ZGZ}'\,| \\
&= |\,\mathbf{R}\,|\,|\,\mathbf{I}+\mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZG}\,| \\
&= |\,\mathbf{R}\,|\,|\,(\mathbf{G}^{-1}+\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z})\mathbf{G}\,| \\
&= |\,\mathbf{R}\,|\,|\,\mathbf{G}^{-1}+\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\,|\,|\,\mathbf{G}\,|\,.
\end{aligned}
$$

Useful Results

The mixed model coefficient matrix of Henderson is

$$\mathbf{C} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix}$$

then the determinant of $\mathbf{C}$ is

$$\begin{aligned} \mid \mathbf{C} \mid &= \mid \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} \mid \\ &\quad \times \mid \mathbf{G}^{-1} + \mathbf{Z}'(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{R}^{-1})\mathbf{Z} \mid \\ &= \mid \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \mid \\ &\quad \times \mid \mathbf{X}'(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1})\mathbf{X} \mid \end{aligned}$$

Useful Results V

$\mathbf{S} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{R}^{-1}$ then

$$\begin{aligned}
\mid \mathbf{C} \mid &= \mid \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} \mid \;\; \mid \mathbf{G}^{-1} + \mathbf{Z}'\mathbf{S}\mathbf{Z} \mid \\
&= \mid \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \mid \mid \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \mid .
\end{aligned}$$

Derivatives

$$\frac{\partial \mathbf{V}^{-1}}{\partial \sigma_i^2} = -\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{V}^{-1}$$

$$\frac{\partial \ln \mid \mathbf{V} \mid}{\partial \sigma_i^2} = tr\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\right)$$

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

$$\frac{\partial \mathbf{P}}{\partial \sigma_i^2} = -\mathbf{P}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{P}$$

Derivatives

$$\frac{\partial \mathbf{V}}{\partial \sigma_i^2} = \mathbf{Z}_i \mathbf{G}_i \mathbf{Z}_i'$$

$$\frac{\partial \ln \mid \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \mid}{\partial \sigma_i^2} = \mathrm{tr}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{V}^{-1}\mathbf{X}$$

Forcing PD

```
# Compute eigenvalues and eigenvectors
mkPD = function(A){
  D = eigen(A)
  V = D$values;     U = D$vectors
  nn = length(V[V<0]);      N = nrow(A)
kpos = N-nn;     ff = V[kpos]
if( nn > 0){
   for(i in 1:nn){
   ff = ff*0.8;    V[kpos+i] = ff
   }
  }
 B = U %*% diag(V) %*% t(U)
 return(B)
}
```

Example

$$\mathbf{A} = \begin{pmatrix} 100 & 80 & 20 & 6 \\ 80 & 50 & 10 & 2 \\ 20 & 10 & 6 & 1 \\ 6 & 2 & 1 & 1 \end{pmatrix}$$

The eigenvalues are

$$\begin{pmatrix} 162.1627196 & 4.1339019 & 0.9171925 & -10.213814 \end{pmatrix}$$

Change the last one to be 0.733754, then reconstruct original matrix,

$$\mathbf{A}^* = \mathbf{UDU}' = \begin{pmatrix} 103.87 & 75.18 & 18.26 & 4.94 \\ 75.18 & 56.00 & 12.16 & 3.32 \\ 18.26 & 12.16 & 6.78 & 1.47 \\ 4.94 & 3.32 & 1.47 & 1.29 \end{pmatrix}$$

where **U** are the eigenvectors.

REML

The multivariate normal distribution likelihood function is

$$L(\mathbf{y}) = (2\pi)^{-.5N} \mid \mathbf{V} \mid^{-.5} \exp(-.5(\mathbf{y} - \mathbf{Xb})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Xb}))$$

The log of the likelihood, L_1 is

$$L_1 = -0.5[N \ln(2\pi) + \ln \mid \mathbf{V} \mid + (\mathbf{y} - \mathbf{Xb})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Xb})]$$

The term $N \ln(2\pi)$ is a constant

REML

$$|\mathbf{V}| = |\mathbf{R}|\,|\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}|\,|\mathbf{G}|,$$

and therefore,

$$\ln|\mathbf{V}| = \ln|\mathbf{R}| + \ln|\mathbf{G}| + \ln|\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}|\,.$$

If $\mathbf{R} = \mathbf{I}\sigma_0^2$, then

$$\begin{aligned} \ln|\mathbf{R}| &= \ln|\mathbf{I}\sigma_0^2| \\ &= \ln(\sigma_0^2)^N\,|\mathbf{I}| \\ &= N\ln\sigma_0^2(1) \end{aligned}$$

REML

If $\mathbf{G} = \sum^{+} \mathbf{I}\sigma_i^2$, where $i = 1$ to s, then

$$\begin{aligned} \ln \mid \mathbf{G} \mid &= \sum_{i=1}^{s} \ln \mid \mathbf{I}\sigma_i^2 \mid \\ &= \sum_{i=1}^{s} q_i \ln \sigma_i^2 \end{aligned}$$

In animal models one of the $\mathbf{G}_i$ is equal to $\mathbf{A}\sigma_i^2$. In that case,

$$\ln \mid \mathbf{A}\sigma_i^2 \mid = \ln(\sigma_i^2)^{q_i} \mid \mathbf{A} \mid$$

which is

$$\ln \mid \mathbf{A}\sigma_i^2 \mid = q_i \ln \sigma_i^2 \mid \mathbf{A} \mid = \ q_i \ln \sigma_i^2 + \ln \mid \mathbf{A} \mid$$

REML

$$\mathbf{C} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix}$$

and

$$\mid \mathbf{C} \mid = \mid \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \mid \mid \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \mid$$

so that

$$\ln \mid \mathbf{C} \mid = \ln \mid \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \mid + \ln \mid \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \mid$$

REML

REML

Restricted (or Residual) maximum likelihood was suggested by Thompson (1962), and described formally by Patterson and Thompson (1971)

- $\mathbf{y}$ has a multivariate normal distribution
- Translation invariant, $\mathbf{QX} = \mathbf{0}$
- Estimates within the allowable parameter space(i.e. zero to plus infinity)
- REML is asymptotically unbiased procedure
- Several computational variants

Derivation

Begin with residual contrasts,

$$\mathbf{K}'\mathbf{y}$$

where $\mathbf{K}'\mathbf{X} = 0$, and $\mathbf{K}'$ has rank equal to $N - r(\mathbf{X})$

$$L(\mathbf{K}'\mathbf{y}) = (2\pi)^{-.5(N-r(\mathbf{X}))} \mid \mathbf{K}'\mathbf{V}\mathbf{K} \mid^{-.5} \exp(-.5(\mathbf{K}'\mathbf{y})'(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}(\mathbf{K}'\mathbf{y}))$$

The natural log of the likelihood function is

$$L_3 = -.5(N - r(\mathbf{X}))\ln(2\pi) - .5\ln \mid \mathbf{K}'\mathbf{V}\mathbf{K} \mid -.5\mathbf{y}'\mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}'\mathbf{y}$$

$$\ln \mid \mathbf{K}'\mathbf{V}\mathbf{K} \mid \;=\; \ln \mid \mathbf{V} \mid + \ln \mid \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \mid$$

and

Derivation II

$$\mathbf{y}'\mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}'\mathbf{y} = \mathbf{y}'\mathbf{P}\mathbf{y} = (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$$

for any $\mathbf{K}'$ such that $\mathbf{K}'\mathbf{X} = \mathbf{0}$.

$$L_4 = -.5 \ln \mid \mathbf{V} \mid -.5 \ln \mid \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \mid -.5(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$$

Computational Variants

- **Derivative Free approach**, search technique to find the parameters that maximize the log likelihood function.
- **First Derivatives and EM**, the first derivatives of L_4 set to zero in order to maximize the likelihood function. Solutions need to be obtained by iteration because the resulting equations are non linear.
- **Second Derivatives**, gradient methods used to find the parameters that make the first derivatives equal to zero. Newton-Raphson (involves the observed information matrix) and Fishers Method of Scoring (involves the expected information matrix) have been used.
- **Average Information**, averages the observed and expected information matrices.

DF REML

Various alternative forms of L_4 can be derived.

$$\ln \mid \mathbf{V} \mid = \ln \mid \mathbf{R} \mid + \ln \mid \mathbf{G} \mid + \ln \mid \mathbf{G}^{-1} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \mid$$

and that

$$\ln \mid \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \mid = \ln \mid \mathbf{C} \mid - \ln \mid \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \mid$$

combining these results gives

$$L_4 = -.5 \ln \mid \mathbf{R} \mid -.5 \ln \mid \mathbf{G} \mid -.5 \ln \mid \mathbf{C} \mid -.5 \mathbf{y}'\mathbf{P}\mathbf{y}$$

DF REML

$$
\begin{aligned}
\ln \mid \mathbf{R} \mid &= \ln \mid \mathbf{I}\sigma_0^2 \mid \\
&= N \ln \sigma_0^2 \\
\ln \mid \mathbf{G} \mid &= \sum_{i=1}^{s} q_i \ln \sigma_i^2 \\
\ln \mid \mathbf{C} \mid &= \ln \mid \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} \mid + \ln \mid \mathbf{Z}'\mathbf{S}\mathbf{Z} + \mathbf{G}^{-1} \mid \\
\ln \mid \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} \mid &= \ln \mid \mathbf{X}'\mathbf{X}\sigma_0^{-2} \mid \\
&= \ln(\sigma_0^{-2})^{r(\mathbf{X})} \mid \mathbf{X}'\mathbf{X} \mid = \ln \mid \mathbf{X}'\mathbf{X} \mid - r(\mathbf{X}) \ln \sigma_0^2 \\
\mathbf{S} &= \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{R}^{-1} \\
\mathbf{Z}'\mathbf{S}\mathbf{Z} + \mathbf{G}^{-1} &= \sigma_0^{-2}\mathbf{Z}'\mathbf{M}\mathbf{Z} + \mathbf{G}^{-1} \\
&= \sigma_0^{-2}(\mathbf{Z}'\mathbf{M}\mathbf{Z} + \mathbf{G}^{-1}\sigma_0^2)
\end{aligned}
$$

DF REML

$$\ln \mid \mathbf{C} \mid = \ln \mid \mathbf{X}'\mathbf{X} \mid - r(\mathbf{X}) \ln \sigma_0^2 - q \ln \sigma_0^2 + \ln \mid \mathbf{Z}'\mathbf{M}\mathbf{Z} + \mathbf{G}^{-1}\sigma_0^2 \mid$$

$$\begin{aligned} L_4 &= -.5(N - r(\mathbf{X}) - q) \ln \sigma_0^2 - .5 \sum_{i=1}^{s} q_i \ln \sigma_i^2 \\ &\quad -.5 \ln \mid \mathbf{C}^{\star} \mid -.5\mathbf{y}'\mathbf{P}\mathbf{y} \end{aligned}$$

DF REML

$$\mathbf{C}^{\star} = \begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{G}^{-1}\sigma_0^2 \end{pmatrix}$$

$$\begin{aligned} q_i \ln \sigma_i^2 &= q_i \ln \sigma_0^2/\alpha_i \\ &= q_i(\ln \sigma_0^2 - \ln \alpha_i) \end{aligned}$$

$$L_4 = -.5[(N - r(\mathbf{X})) \ln \sigma_0^2 - \sum_{i=1}^{s} q_i \ln \alpha_i + \ln \mid \mathbf{C}^{\star} \mid + \mathbf{y}'\mathbf{P}\mathbf{y}]$$

$$\mathbf{y}'\mathbf{P}\mathbf{y} = \mathbf{y}'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}\hat{\mathbf{u}})/\sigma_0^2$$

DF REML

- Pick several sets of possible parameter values.
- Calculate L_4 for each set (negative values).
- Choose new sets of parameters but closer to the set that had the largest L_4 in the first group.
- Continue narrowing the differences among the sets until they are almost all the same.
- To check convergence, start over with a completely different first group, and re-do.

EM REML, First Derivatives

$$
\begin{aligned}
\frac{\partial L_4}{\partial \sigma_i^2} &= -.5tr\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2} - .5tr(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{V}^{-1}\mathbf{X} \\
&\quad +.5(\mathbf{y}-\mathbf{X}\hat{\mathbf{b}})'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\mathbf{b}}) \\
&= -.5tr\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i' + .5\mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y} \quad \mathrm{OR} \\
&= -.5tr\mathbf{P} + .5\mathbf{y}'\mathbf{P}\mathbf{P}\mathbf{y}
\end{aligned}
$$

EM REML

$$tr\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i' = q_i/\sigma_i^2 - tr\mathbf{C}_{ii}\sigma_0^2/\sigma_i^4$$

and

$$tr\mathbf{P} = (N - r(\mathbf{X}))\sigma_0^2 - \sum_{i=1}^{s} \hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i/\sigma_i^2$$

$$\begin{aligned}
\mathbf{P}\mathbf{y} &= \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \\
\mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y} &= \hat{\mathbf{u}}_i'\mathbf{G}_i^{-2}\hat{\mathbf{u}}_i \\
\hat{\sigma}_i^2 &= (\hat{\mathbf{u}}_i'\mathbf{G}_i^{-1}\hat{\mathbf{u}}_i + tr\mathbf{G}_i^{-1}\mathbf{C}_{ii}\sigma_0^2)/q_i \\
\hat{\sigma}_0^2 &= \mathbf{y}'\mathbf{P}\mathbf{y}/(N - r(\mathbf{X}))
\end{aligned}$$

EM REML

- Slow convergence, many iterations needed.
- May not converge at all, to zero or infinity.
- $tr\mathbf{C}_{ii}$ may not be possible.

AI REML

The second derivatives give a matrix of quantities. The elements of the *observed information* matrix (Gilmour et al. 1995) are

$$
\begin{aligned}
-\frac{\partial^2 L_4}{\partial\sigma_i^2\partial\sigma_0^2} &= 0.5\mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y}/\sigma_0^4 \\
-\frac{\partial^2 L_4}{\partial\sigma_i^2\partial\sigma_j^2} &= 0.5tr(\mathbf{P}\mathbf{Z}_i\mathbf{Z}_j') - 0.5tr(\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{Z}_j\mathbf{Z}_j') \\
&\quad +\mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{Z}_j\mathbf{Z}_j'\mathbf{P}\mathbf{y}/\sigma_0^2 - 0.5\mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{Z}_j'\mathbf{P}\mathbf{y}/\sigma_0^2 \\
-\frac{\partial^2 L_4}{\partial\sigma_0^2\partial\sigma_0^2} &= \mathbf{y}'\mathbf{P}\mathbf{y}/\sigma_0^6 - 0.5(N - r(\mathbf{X}))/\sigma_0^4
\end{aligned}
$$

AI REML

the *expected information* matrix (Gilmour et al. 1995) are

$$
\begin{aligned}
E[-\frac{\partial^2 L_4}{\partial\sigma_i^2 \partial\sigma_0^2}] &= 0.5tr(\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i')/\sigma_0^2 \\
E[-\frac{\partial^2 L_4}{\partial\sigma_i^2 \partial\sigma_j^2}] &= 0.5tr(\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{Z}_j\mathbf{Z}_j') \\
E[-\frac{\partial^2 L_4}{\partial\sigma_0^2 \partial\sigma_0^2}] &= 0.5(N - r(\mathbf{X}))/\sigma_0^4
\end{aligned}
$$

AI REML

Average Information implies the average of *observed* and *expected* information matrices.

$$
\begin{aligned}
I[\sigma_i^2, \sigma_0^2] &= 0.5\mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y}/\sigma_0^4 \\
I[\sigma_i^2, \sigma_j^2] &= \mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{Z}_j\mathbf{Z}_j'\mathbf{P}\mathbf{y}/\sigma_0^2 \\
I[\sigma_0^2, \sigma_0^2] &= 0.5\mathbf{y}'\mathbf{P}\mathbf{y}/\sigma_0^6
\end{aligned}
$$

Use ASREML package, expensive, memory hog.

Summary

- Likelihood methods may not converge.
- Likelihood methods may take a long time.
- Likelihood estimates have good accuracy.
- Accuracy of estimates of variance components depends upon
  - Number of data points
  - Number of levels of random factors
  - True variances
  - Distribution of random elements over fixed effects levels