Estimation of Variance Components in Animal Breeding

Day 1	Distributions Building Blocks Unbiased Estimation
Day 2	Likelihood Methods ML, REML
Day 3	Bayesian Estimation Data Preparation Simple Animal Models
Day 4	Software VCE, DMU, MTDFREML
Day 5	Other Models

Why do we need variances?

1. To estimate heritability.
2. To use in BLUP.
3. To develop index weights.
4. To determine breeding strategies.
5. To predict future responses.
6. To design experiments.

What are variances?

1. Positive quantities, greater than 0.
2. Quadratic forms of $\mathbf{y}$, $\mathbf{y}'\mathbf{Q}\mathbf{y}$.
3. Measure of dispersion around a mean.

Random variable, y

Has a distribution function, $p(y)$

$$\begin{aligned} E(y) &= \mu \\ Var(y) &= E(y^2) - E(y)E(y) \\ &= \sigma_y^2 \end{aligned}$$

Discrete Random Variables

y	$p(y)$
0	.125
1	.250
2	.250
3	.375

$$
\begin{aligned}
Pr(Y = 0 \text{ or } Y = 2) &= p(0) + p(2) \;=\; .375 \\
F(y) &= Pr(Y \leq y) \\
F(0) &= p(0) \\
F(1) &= p(0) + p(1) \\
F(2) &= p(0) + p(1) + p(2) \\
F(3) &= F(2) + p(3) \;=\; 1
\end{aligned}
$$

y	$p(y)$
0	.125
1	.250
2	.250
3	.375

$$\begin{aligned} E(y) &= \Sigma\, y\ p(y) \\ &= 0(.125) + 1(.25) + 2(.25) + 3(.375) \\ &= 1.875 \end{aligned}$$

$$\begin{aligned} E(y^2) &= \Sigma\, y^2\ p(y) \\ &= 0(.125) + 1(.25) + 4(.25) + 9(.375) \\ &= 4.625 \end{aligned}$$

$$\begin{aligned} Var(y) &= E(y^2) - [E(y)]^2 \\ &= 4.625 - (1.875)^2 \\ &= 1.109375 \end{aligned}$$

Types of discrete distributions

Binomial	Success or Failure, (0-1)
Poisson	Number of embryos, 0-20, mean 3-4.

Matrix formulation

$\mathbf{y}_{Nx1} =$ vector of random variables

$E(\mathbf{y}) = \mu_{Nx1}$

$E(c\,\mathbf{y}) = c\,\mu$

$E(\mathbf{C}\mathbf{y}) = \mathbf{C}\mu$

$E(\mathbf{y}\mathbf{y}') =$ a matrix of order N

$Var(\mathbf{y}) = E(\mathbf{y}\mathbf{y}') - E(\mathbf{y})E(\mathbf{y}')$

$Var(\mathbf{C}\mathbf{y}) = \mathbf{C}[Var(\mathbf{y})]\mathbf{C}'$

Continuous Distributions

$$E(y) = \int y \; p(y) \partial y$$

$$E(y^2) = \int y^2 \partial y$$

$$Var(y) = E(y^2) - [E(y)]^2$$

Types of Continuous

Uniform - interval a to b

Underlies all random number generators.

Multiply with carry algorithms.

$$x_{n-1} = 123456$$

$$\begin{aligned} x_n &= 672 * [x_{n-1} - (\frac{x_{n-1}}{1000}) * 1000] \ + \ \frac{x_{n-1}}{1000} \\ &= 672 * [456] \ + \ 123 \\ &= 306555 \end{aligned}$$

$$\begin{aligned} x_{n+1} &= 672 * [555] \ + \ 306 \\ &= 373266. \end{aligned}$$

Repeating Cycle

`0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 ...`

`6 9 3 4 4 2 1 5 7 8 6 9 3 4 4 2 1 ...`

Desired: A long cycle

Most: 2^{32}, 2^{80}

Ours: 2^{250}

Types of continuous

Normal

$$p(y) = \frac{1}{(2\pi)^{.5}\sigma} \exp^{-.5(y-\mu)^2/\sigma}$$

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

Random Normal Deviates - utilizes Uniform variates

Types of continuous

- Beta
- Gamma
- t-
- F-
- Chi-square, from a Gamma. If

  $$\mathbf{y} \sim N(0, I)$$

  then

  $$\mathbf{y}'\mathbf{y} \sim \chi^2_N.$$

  If

  $$\mathbf{y} \sim N(\mu, V)$$

  then

  $$\mathbf{y}'\mathbf{y} \sim \chi^{2'}_{N,\lambda},$$

  where

  $$\lambda = .5\mu'\mathbf{V}\mu.$$

Wishart Distribution

- Like a multivariate Chi-square
- A matrix rather than scalar
- Uses random Chi-square and Normal variates

Building Blocks

1. $Var(\mathbf{y}) = \mathbf{V} = \mathbf{ZGZ'} + \mathbf{R}$

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

3. $\mid \mathbf{A}k \mid = k^m \mid \mathbf{A} \mid$

4. $\mid \mathbf{MU} \mid = \mid \mathbf{M} \mid\mid \mathbf{U} \mid$

5. $$\begin{vmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{Q} & \mathbf{D} \end{vmatrix} = |\, \mathbf{A} \,||\, \mathbf{D} + \mathbf{Q}\mathbf{A}^{-1}\mathbf{B} \,|$$

$$= |\, \mathbf{D} \,||\, \mathbf{A} + \mathbf{B}\mathbf{D}^{-1}\mathbf{Q} \,|$$

if $\mathbf{A} = \mathbf{I},\ \mathbf{D} = \mathbf{I},$

$$= |\, \mathbf{I} + \mathbf{Q}\mathbf{B} \,|$$

$$= |\, \mathbf{I} + \mathbf{B}\mathbf{Q} \,|$$

$$= |\, \mathbf{I} + \mathbf{B}'\mathbf{Q}' \,|$$

$$= |\, \mathbf{I} + \mathbf{Q}'\mathbf{B}' \,|$$

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

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

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

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

$$\mathbf{PX} = \mathbf{0}.$$

9. Blocks 9 to 13 are derivatives, needed tomorrow.

History of Methods

B.C. Balanced data designs, AOV

A.D. Henderson's 1, 2, 3

ML, Hartley and Rao

REML, Patterson and Thompson

MIVQUE, C. R. Rao

Bayesian, Gianola, Harville

Circa 1960's - Simple Models

$$\begin{pmatrix} 29 \\ 58 \\ 13 \\ 44 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu + \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \mathbf{e}$$

OR

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e}.$$

$$\mathbf{V}_0 = \mathbf{I},$$

$$\mathbf{V}_1 = \mathbf{Z}_1\mathbf{Z}_1' = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix},$$

$$\mathbf{V}_2 = \mathbf{Z}_2\mathbf{Z}_2' = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix},$$

Quadratic Forms - Method

$$\mathbf{Q}_1 = \begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{pmatrix},$$

$$\mathbf{y}'\mathbf{Q}_1 = \begin{pmatrix} -44 & 74 & -76 & 46 \end{pmatrix}$$

$$\mathbf{y}'\mathbf{Q}_1\mathbf{y} = 4052.$$

$$\mathbf{Q}_2 = \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix},$$

$$\mathbf{y}'\mathbf{Q}_2 = \begin{pmatrix} -28 & 88 & -92 & 32 \end{pmatrix}$$

$$\mathbf{y}'\mathbf{Q}_2\mathbf{y} = 4504.$$

$$\mathbf{Q}_3 = \begin{pmatrix} 4 & -2 & -1 & -1 \\ -2 & 4 & -1 & -1 \\ -1 & -1 & 4 & -2 \\ -1 & -1 & -2 & 4 \end{pmatrix},$$

$$\mathbf{y}'\mathbf{Q}_3 = \begin{pmatrix} -57 & 117 & -123 & 63 \end{pmatrix}$$

$$\mathbf{y}'\mathbf{Q}_3\mathbf{y} = 6306.$$

Expectations

$$\begin{aligned} E(\mathbf{y}'\mathbf{Q}\mathbf{y}) &= \mu\mathbf{1}'\mathbf{Q}\mathbf{1}\mu + tr\mathbf{Q}\mathbf{V}_0\sigma_0^2 \\ &\quad + tr\mathbf{Q}\mathbf{V}_1\sigma_1^2 + tr\mathbf{Q}\mathbf{V}_2\sigma_2^2 \end{aligned}$$

$$\begin{aligned} tr\mathbf{Q}_1\mathbf{V}_0 &= 8, \\ tr\mathbf{Q}_2\mathbf{V}_0 &= 12, \\ tr\mathbf{Q}_3\mathbf{V}_0 &= 16. \end{aligned}$$

Expectation of $\mathbf{y}'\mathbf{Q}_1\mathbf{y}$

$$\begin{aligned}
tr\mathbf{Q}_1\mathbf{V}_0 &= 8, \\
tr\mathbf{Q}_1\mathbf{V}_1 &= 4, \\
tr\mathbf{Q}_1\mathbf{V}_2 &= 8, \\
E(\mathbf{y}'\mathbf{Q}_1\mathbf{y}) &= 8\sigma_0^2 + 4\sigma_1^2 + 8\sigma_2^2
\end{aligned}$$

Collectively

$$E\begin{pmatrix} \mathbf{y}'\mathbf{Q}_1\mathbf{y} \\ \mathbf{y}'\mathbf{Q}_2\mathbf{y} \\ \mathbf{y}'\mathbf{Q}_3\mathbf{y} \end{pmatrix} = \begin{pmatrix} 8 & 4 & 8 \\ 12 & 8 & 8 \\ 16 & 12 & 8 \end{pmatrix} \begin{pmatrix} \sigma_0^2 \\ \sigma_1^2 \\ \sigma_2^2 \end{pmatrix}.$$

Equate and Solve

$$\begin{pmatrix} \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \end{pmatrix} = \begin{pmatrix} 8 & 4 & 8 \\ 12 & 8 & 8 \\ 16 & 12 & 8 \end{pmatrix}^{-1} \begin{pmatrix} 4051 \\ 4504 \\ 6306 \end{pmatrix} = \begin{pmatrix} 1 \\ 112 \\ 449.5 \end{pmatrix}.$$

Sampling Variances

$$Var(\mathbf{y}'\mathbf{Q}\mathbf{y}) = 2\ tr\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{V}$$

$$\begin{aligned} = \ & 2\ tr\mathbf{Q}\mathbf{V}_0\mathbf{Q}\mathbf{V}_0\sigma_0^2\sigma_0^2 + 4\ tr\mathbf{Q}\mathbf{V}_0\mathbf{Q}\mathbf{V}_1\sigma_0^2\sigma_1^2 \\ & +4\ tr\mathbf{Q}\mathbf{V}_0\mathbf{Q}\mathbf{V}_2\sigma_0^2\sigma_2^2 + 2\ tr\mathbf{Q}\mathbf{V}_1\mathbf{Q}\mathbf{V}_1\sigma_1^2\sigma_1^2 \\ & +4\ tr\mathbf{Q}\mathbf{V}_1\mathbf{Q}\mathbf{V}_2\sigma_1^2\sigma_2^2 + 2\ tr\mathbf{Q}\mathbf{V}_2\mathbf{Q}\mathbf{V}_2\sigma_2^2\sigma_2^2 \end{aligned}$$

$$Cov(\mathbf{y}'\mathbf{Q}_i\mathbf{y}, \mathbf{y}'\mathbf{Q}_j\mathbf{y}) = 2\ tr\mathbf{Q}_i\mathbf{V}\mathbf{Q}_j\mathbf{V}$$

Sampling Variances

Depend upon

- $\mathbf{Q}$'s, i.e. method of estimation.
- $\mathbf{V}_i$, i.e. design of the data.
- σ_i^2, i.e. unknown true variances.

Due to the size of $\mathbf{Q}$'s and $\mathbf{V}$'s, $tr\mathbf{QVQV}$ is generally impossible to compute.

Estimation of Variance Components in Animal Breeding

✓	Day 1	Distributions
✓		Building Blocks
✓		Unbiased Estimation
	Day 2	Likelihood Methods
		ML, REML, Bayesian
	Day 3	Data Preparation
		Simple Animal Models
	Day 4	Software
		VCE, DMU, MTDFREML
	Day 5	Other Models

Likelihood Methods

- More than unbiasedness
- Within the parameter space
- Smaller sampling variances

Normal Distribution

Log Likelihood

$$\begin{aligned} L(\mathbf{y}) &= (2\pi)^{-.5N} \mid \mathbf{V} \mid^{-.5} \\ &\quad \exp(-.5(\mathbf{y} - \mathbf{X}\mathbf{b})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b})). \end{aligned}$$

Taking logs gives

$$\begin{aligned} L_1 &= -0.5[N \ln(2\pi) + \ln \mid \mathbf{V} \mid \\ &\quad +(\mathbf{y} - \mathbf{X}\mathbf{b})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b})]. \end{aligned}$$

Now make use of Building Blocks,

$$\begin{aligned} \mid \mathbf{V} \mid &= \mid \mathbf{R} \mid \mid \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \mid \mid \mathbf{G} \mid, \\ \ln \mid \mathbf{V} \mid &= \ln \mid \mathbf{R} \mid + \ln \mid \mathbf{G} \mid \\ &\quad + \ln \mid \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \mid . \end{aligned}$$

Parts of Log Likelihood

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

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

Similarly, if $\mathbf{G} = {\scriptstyle\Sigma}^{+} \mathbf{I}\sigma_i^2$, 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

$$\mathbf{G}_i = \mathbf{A}\sigma_i^2$$

$$\begin{aligned} \ln \mid \mathbf{G}_i \mid &= \ln \mid \mathbf{A}\sigma_i^2 \mid \\ &= \ln[(\sigma_i^2)^{q_i} \mid \mathbf{A} \mid] \\ &= q_i \ln \sigma_i^2 + \ln \mid \mathbf{A} \mid . \end{aligned}$$

Another Part of Log Likelihood

Recall that

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

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

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

Many ways to manipulate the log likelihood function.

Maximum Likelihood

$$L_2 = -0.5[\ln \mid \mathbf{V} \mid + (\mathbf{y}-\mathbf{Xb})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{Xb})]$$

Use Building Blocks 9-13, derivatives.

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

and

$$tr[\mathbf{V}^{-1}\mathbf{V}_i] = (\mathbf{y}-\mathbf{X}\hat{\mathbf{b}})'\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\mathbf{b}}).$$

Let

$$\mathbf{T} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1},$$

then

$$\hat{\sigma}_i^2 = (\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i + tr\mathbf{T}_{ii}\hat{\sigma}_0^2)/q_i$$

for $i = 1, 2, \ldots, s$, and for $i = 0$ gives

$$\hat{\sigma}_0^2 = (\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \hat{\mathbf{u}}'\mathbf{Z}'\mathbf{y})/\ N.$$

Simple Example

$$\begin{pmatrix} 50 & 0 & 5 & 15 & 30 \\ 0 & 40 & 5 & 15 & 20 \\ 5 & 5 & 10+\alpha & 0 & 0 \\ 15 & 15 & 0 & 30+\alpha & 0 \\ 30 & 20 & 0 & 0 & 50+\alpha \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ A_1 \\ A_2 \\ A_3 \end{pmatrix} = \begin{pmatrix} 3200 \\ 2380 \\ 580 \\ 1860 \\ 3140 \end{pmatrix}.$$

$$\mathbf{y}'\mathbf{y} = 356,000$$

$$\alpha = 25$$

Example

$$\begin{pmatrix} \hat{F}_1 \\ \hat{F}_2 \\ \hat{A}_1 \\ \hat{A}_2 \\ \hat{A}_3 \end{pmatrix} = \begin{pmatrix} 63.6138 \\ 59.1774 \\ -0.9702 \\ 0.3297 \\ 0.6405 \end{pmatrix},$$

$$\mathbf{T} = \begin{pmatrix} 0.02857 & 0 & 0 \\ 0 & 0.01818 & 0 \\ 0 & 0 & 0.01333 \end{pmatrix},$$

$$\begin{aligned} \hat{\beta}'\mathbf{W}'\mathbf{y} &= 346,468.07 \\ \hat{\sigma}_0^2 &= (\mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{W}'\mathbf{y})/N \\ &= 105.91034, \\ \hat{\mathbf{u}}'\hat{\mathbf{u}} &= 1.46016, \\ \hat{\sigma}_a^2 &= (\hat{\mathbf{u}}'\hat{\mathbf{u}} + \hat{\sigma}_0^2 tr\mathbf{T})/q \\ &= (1.46016 + 105.91034(0.0600866))/3 \\ &= 2.6079837. \end{aligned}$$

Must iterate to convergence!! Not guaranteed.

EM Algorithm
Expectation
Maximization

- Conditional expected values
- Maximize likelihood
- Equivalent to previous example

Restricted ML
Residual ML

- ML uses N in the denominator for $\hat{\sigma}_0^2$,
- ML uses $tr\mathbf{T}$ in formula for $\hat{\sigma}_i^2$,
- ML does not account for $r(\mathbf{X})$.
- Too much bias in ML.
- Patterson and Thompson (1971).
- Log Likelihood for error contrasts rather than for y.

REML

Four different computing approaches.

1. Derivative Free
2. EM algorithm
3. MIVQUE
4. Average Information

Error Contrasts

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

There are an infinite number of $\mathbf{K}'$ that could be used. We only need to know that one can be formed.

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

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

- Ignore $-.5(N - r(\mathbf{X}))\ln(2\pi)$.
- Searle (1979) showed

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

and

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

- New log likelihood is

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

DF REML

- Smith and Graser (1986).
- Search the parameter space in an efficient manner to find the values of unknown variances that maximize L_4.
- Global versus local maxima.

Strategy

- Choose several sets of possible parameter values of σ_i^2 for all i.
- Compute the value of L_4 for each set.
- From the set that gives the maximum, choose some new sets around this set.
- Re-compute the value of L_4 for each new set.
- Continue refining the values in the sets until the differences between sets are trivial.

Computation of L_4

Various alternative forms of L_4 have been derived for particular models. Utilizing previous results, one form is

$$\begin{aligned} L_4 &= -.5\ln\mid\mathbf{R}\mid -.5\ln\mid\mathbf{G}\mid \\ &\quad -.5\ln\mid\mathbf{C}\mid -.5\mathbf{y}'\mathbf{P}\mathbf{y}. \end{aligned}$$

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

and

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

The difficult parts are $\ln\mid\mathbf{C}^\star\mid$ and $\mathbf{y}'\mathbf{P}\mathbf{y}$.

Gaussian Elimination

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

$$\begin{pmatrix} 50 & 10 & 3200 \\ 10 & 40 & 2380 \\ 3200 & 2380 & 356000 \end{pmatrix}$$

Pivot 1 = 50.

$$\begin{pmatrix} 50 & 10 & 3200 \\ 0 & 38 & 1740 \\ 0 & 1740 & 151200 \end{pmatrix},$$

Pivot 2 = 38.

$$\begin{pmatrix} 50 & 10 & 3200 \\ 0 & 38 & 1740 \\ 0 & 0 & 71526.3158 \end{pmatrix}.$$

The determinant of $\mathbf{C}^\star$ is the product of the pivots and $\ln \mid \mathbf{C}^\star \mid$ is the sum of the logs of the pivots.

$$\begin{aligned} \mid \mathbf{C}^\star \mid &= 50 * 38 = 1900, \\ \ln \mid \mathbf{C}^\star \mid &= \ln(1900) = 7.5496 \\ &= \ln(50) + \ln(38) \\ &= 3.91202 + 3.63759 \\ &= 7.5496. \end{aligned}$$

And

$$\mathbf{y}'\mathbf{P}\mathbf{y} = 71526.3158.$$

Example, Page 40

- Two random factors, A and B.
- Fix $\alpha_B = 10$, set of values for α_A

α_A	L_4
5	-251.4442
10	-251.1504
20	-250.9822
30	-250.9274
40	-250.9019

$$\alpha_A = 40$$

$$\begin{aligned}
\ln \mid \mathbf{C}^\star \mid &= 32.052454, \\
\sigma_0^2 &= 96.399728, \\
\ln \sigma_0^2 &= 4.5685034, \\
\mathbf{y}'\mathbf{P}\mathbf{y} &= 8483.176/\sigma_0^2 \;=\; 88, \\
q_A \ln \alpha_A &= 11.0666385, \\
q_B \ln \alpha_B &= 9.2103404, \\
(N - r(\mathbf{X})) &= 88,
\end{aligned}$$

$$\begin{aligned}
L_4 \;=\; & -0.5[(N - r(\mathbf{X})) \ln \sigma_0^2 && \text{-0.5 [88(4.5685)} \\
& -q_A \ln \alpha_A && \text{-11.0666} \\
& -q_B \ln \alpha_B && \text{-9.2103} \\
& + \ln \mid \mathbf{C}^\star \mid && \text{32.0525} \\
& + \mathbf{y}'\mathbf{P}\mathbf{y}/\sigma_0^2] && \text{88]} \\
= & && \text{-250.9019.}
\end{aligned}$$

$$\mathbf{Q} = \begin{pmatrix} 1 & 5 & 25 \\ 1 & 10 & 100 \\ 1 & 20 & 400 \\ 1 & 30 & 900 \\ 1 & 40 & 1600 \end{pmatrix} \text{ and } \mathbf{Y} = \begin{pmatrix} -251.4442 \\ -251.1504 \\ -250.9822 \\ -250.9274 \\ -250.9019 \end{pmatrix}$$

$$\hat{\beta} = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{Y} = \begin{pmatrix} -251.6016 \\ .0448877 \\ -.000698 \end{pmatrix}.$$

$$L_4 = \; -251.6016 + .04489\alpha_A - .000698\alpha_A^2.$$

$$\alpha_A = .04489/(2(.000698)) = 32.1546.$$

α_B	L_4
2	-250.2722
3	-250.1954
4	-250.2379
5	-250.3295
6	-250.4419
7	-250.5624
8	-250.6843
9	-250.8042
10	-250.9204

$$\alpha_B = 1.2625.$$

Try new values of α_A with $\alpha_B = 1.2625$, etc.

The Simplex Method (Nelder and Mead, 1965), MTDFREML

EM Algorithm

Take derivatives of L_4 equate to null matrices or zero.

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

Note the differences from ML.

Simple Example from ML.

EM REML is iterative, must continue to convergence.

$$\hat{\sigma}_a^2 = (\hat{\mathbf{a}}'\mathbf{A}^{-1}\hat{\mathbf{a}} + tr\mathbf{C}\mathbf{A}^{-1}\sigma_0^2)/q.$$

Example

$$
\begin{aligned}
\hat{\mathbf{u}}' &= (-.9702 \quad .3297 \quad .6405\ ),\\
\hat{\mathbf{u}}'\hat{\mathbf{u}} &= 1.4602,\\
\mathbf{C}_{uu} &= \begin{pmatrix} .03076 & .00418 & .00506 \\ .00418 & .02615 & .00967 \\ .00506 & .00967 & .02526 \end{pmatrix},\\
tr\mathbf{C}_{uu} &= .0821769,\\
\hat{\sigma}_0^2 &= (\mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{W}'\mathbf{y})/(N - r(\mathbf{X})),\\
&= 108.3174,\\
\hat{\sigma}_u^2 &= (\hat{\mathbf{u}}'\hat{\mathbf{u}} + tr\mathbf{C}_{uu}\hat{\sigma}_0^2)/q,\\
&= (1.4602 + (.0821769)108.3174)/3,\\
&= 3.4538.
\end{aligned}
$$

Average Information

Gradient method

$$\theta^{(t+1)} = \theta^{(t)} + \mathbf{M}^{(t)}\mathbf{d}^{(t)},$$

$\mathbf{M}^{(t)}$, matrix of second derivatives, inverted,

$\mathbf{d}^{(t)}$, vector of first derivatives.

First Derivatives

$$\mathbf{d} = \begin{pmatrix} 0.5(\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i/\sigma_i^4 - q_i/\sigma_i^2 + tr\mathbf{C}_{ii}\sigma_0^2/\sigma_i^4) \\ \vdots \\ 0.5([\mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{W}'\mathbf{y}]/\sigma_0^2 - (N - r(\mathbf{X}))\sigma_0^2) \end{pmatrix}$$

These quantities tend to zero at convergence.

Second Derivatives
Observed Information Matrix

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

and

$$
-\frac{\partial^2 L_4}{\partial\sigma_0^2\partial\sigma_0^2} = \mathbf{y}'\mathbf{P}\mathbf{y}/\sigma_0^6 - 0.5(N - r(\mathbf{X}))/\sigma_0^4.
$$

Second Derivatives
Expected Information Matrix

$$E[-\frac{\partial^2 L_4}{\partial \sigma_i^2 \partial \sigma_0^2}] \;=\; 0.5tr(\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i')/\sigma_0^2,$$

$$E[-\frac{\partial^2 L_4}{\partial \sigma_i^2 \partial \sigma_j^2}] \;=\; 0.5tr(\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{Z}_j\mathbf{Z}_j'),$$

and

$$E[-\frac{\partial^2 L_4}{\partial \sigma_0^2 \partial \sigma_0^2}] \;=\; 0.5(N - r(\mathbf{X}))/\sigma_0^4.$$

Average Information Matrix
(Observed + Expected)/2

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

and

$$I[\sigma_0^2, \sigma_0^2] = 0.5\mathbf{y}'\mathbf{P}\mathbf{y}/\sigma_0^6.$$

$$\mathbf{M}^{(t)} = I[\sigma, \sigma]^{-1}.$$

Estimation of Variance Components in Animal Breeding

√	Day 1	Distributions
√		Building Blocks
√		Unbiased Estimation
√	Day 2	Likelihood Methods
√		ML, REML
	Day 3	Bayesian Estimation
		Data Preparation
		Simple Animal Models
	Day 4	Software
		VCE, DMU, MTDFREML
	Day 5	Other Models

Bayesian Methods

- Every variable is random and has an underlying distribution function.
- Joint posterior distribution
- Conditional posterior distributions
- Gibbs sampling, random samples from conditional posterior distributions, eventually converge to joint posterior distribution.

Joint Posterior Distribution

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}.$$

θ includes $\mathbf{b}$, $\mathbf{a}$, σ_a^2, and σ_e^2.

$$p(\theta \mid \mathbf{y}) = \frac{p(\theta)p(\mathbf{y} \mid \theta)}{p(\mathbf{y})}$$

$$\begin{aligned} p(\mathbf{y} \mid \theta) &\propto (\sigma_e^2)^{(-N/2)} \\ &\quad \exp[-(\mathbf{y} - \mathbf{Xb} - \mathbf{Za})' \\ &\quad (\mathbf{y} - \mathbf{Xb} - \mathbf{Za})/2\sigma_e^2] \\ p(\mathbf{b}) &\propto \text{constant}, \end{aligned}$$

$$p(\mathbf{a}) \propto (\sigma_a^2)^{(-q/2)} \exp\left[-\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}/2\sigma_a^2\right],$$

$$\begin{aligned} p(\sigma_a^2 \mid v_a, S_a^2) &\propto (\sigma_a^2)^{-(\frac{v_a}{2}+1)} \exp\left(-\frac{v_a}{2}\frac{S_a^2}{\sigma_a^2}\right), \\ p(\sigma_e^2 \mid v_e, S_e^2) &\propto (\sigma_e^2)^{-(\frac{v_e}{2}+1)} \exp\left(-\frac{v_e}{2}\frac{S_e^2}{\sigma_e^2}\right). \end{aligned}$$

Joint Posterior Distribution

$$p(\mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2 \mid \mathbf{y})$$

$$\propto$$

$$p(\mathbf{b})p(\mathbf{a} \mid \sigma_a^2)p(\sigma_a^2)p(\sigma_e^2)p(\mathbf{y} \mid \mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2)$$

$$\begin{aligned}\propto\ & (\sigma_e^2)^{-(\frac{N+v_e}{2}+1)} \\ & \exp[-\frac{1}{2\sigma_e^2}((\mathbf{y} - \mathbf{Xb} - \mathbf{Za})' \\ & (\mathbf{y} - \mathbf{Xb} - \mathbf{Za}) + v_e S_e^2)] \\ & (\sigma_a^2)^{-(\frac{q+v_a}{2}+1)} \\ & \exp\left[-\frac{1}{2\sigma_a^2}(\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} + v_a S_a^2)\right].\end{aligned}$$

Conditional Posterior Distributions

$$\begin{aligned}
\mathbf{W} &= (\mathbf{X} \;\; \mathbf{Z}), \\
\beta' &= (\mathbf{b}' \;\; \mathbf{a}'), \\
\Sigma &= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-1}k \end{pmatrix}, \\
\mathbf{C} &= \mathbf{W}'\mathbf{W} + \Sigma \\
\mathbf{C}\hat{\beta} &= \mathbf{W}'\mathbf{y}
\end{aligned}$$

$$\beta' = (\beta_i \;\; \beta'_{-i}),$$

$$\begin{aligned}
\mathbf{W}' &= (\mathbf{W}_i \;\; \mathbf{W}_{-i})' \\
\mathbf{C} &= \begin{pmatrix} C_{i,i} & \mathbf{C}_{i,-i} \\ \mathbf{C}_{-i,i} & \mathbf{C}_{-i,-i} \end{pmatrix}.
\end{aligned}$$

$$\begin{aligned}
\beta_i \mid \beta_{-i}, \sigma_a^2, \sigma_e^2, \mathbf{y} &\sim N(\hat{\beta}_i, C_{i,i}^{-1}\sigma_e^2) \\
C_{i,i}\hat{\beta}_i &= (\mathbf{W}_i'\mathbf{y} - \mathbf{C}_{i,-i}\beta_{-i}).
\end{aligned}$$

$$\begin{aligned}
b_i \mid \mathbf{b}_{-i}, \mathbf{a}, \sigma_a^2, \sigma_e^2, \mathbf{y} &\sim N(\hat{b}_i, C_{i,i}^{-1}\sigma_e^2), \\
C_{i,i} &= \mathbf{x}_i'\mathbf{x}_i.
\end{aligned}$$

$$\begin{aligned}
a_i \mid \mathbf{b}, \mathbf{a}_{-i}, \sigma_a^2, \sigma_e^2, \mathbf{y} &\sim N(\hat{a}_i, C_{i,i}^{-1}\sigma_e^2), \\
C_{i,i} &= (\mathbf{z}_i'\mathbf{z}_i + A^{i,i}k),
\end{aligned}$$

$$k = \sigma_e^2/\sigma_a^2.$$

$$
\begin{aligned}
\sigma_a^2 \mid \mathbf{b}, \mathbf{a}, \sigma_e^2, \mathbf{y} &\sim \tilde{v}_a \tilde{S}_a^2 \chi_{\tilde{v}_a}^{-2} \\
\tilde{v}_a &= q + v_a, \\
\tilde{S}_a^2 &= (\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} + v_a S_a^2)/\tilde{v}_a,
\end{aligned}
$$

$$
\begin{aligned}
\sigma_e^2 \mid \mathbf{b}, \mathbf{a}, \sigma_a^2, \mathbf{y} &\sim \tilde{v}_e \tilde{S}_e^2 \chi_{\tilde{v}_e}^{-2} \\
\tilde{v}_e &= N + v_e, \\
\tilde{S}_e^2 &= (\mathbf{e}'\mathbf{e} + v_e S_e^2)/\tilde{v}_e, \\
\mathbf{e} &= \mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{a}.
\end{aligned}
$$

Gibbs Sampling

$$\begin{pmatrix} 5 & 1 & 1 & 1 & 1 & 1 \\ 1 & 29 & 7 & -7 & -14 & 0 \\ 1 & 7 & 30 & -14 & 8 & -16 \\ 1 & -7 & -14 & 36 & -14 & 0 \\ 1 & -14 & 8 & -14 & 37 & -16 \\ 1 & 0 & -16 & 0 & -16 & 33 \end{pmatrix} \begin{pmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} = \begin{pmatrix} 238.2 \\ 38.5 \\ 48.9 \\ 64.3 \\ 50.5 \\ 36.0 \end{pmatrix},$$

$$\begin{aligned} \hat{\mu} &= (238.2 - a_1 - a_2 - a_3 - a_4 - a_5)/5 \\ &= 47.64 \\ \mu &= \hat{\mu} + RND * (\sigma_e^2/5)^{.5} \\ &= 47.64 + (-1.21) * (4.32) \\ &= 42.41 \end{aligned}$$

Gibbs Sampling

$$\begin{pmatrix} 5 & 1 & 1 & 1 & 1 & 1 \\ 1 & 29 & 7 & -7 & -14 & 0 \\ 1 & 7 & 30 & -14 & 8 & -16 \\ 1 & -7 & -14 & 36 & -14 & 0 \\ 1 & -14 & 8 & -14 & 37 & -16 \\ 1 & 0 & -16 & 0 & -16 & 33 \end{pmatrix} \begin{pmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} = \begin{pmatrix} 238.2 \\ 38.5 \\ 48.9 \\ 64.3 \\ 50.5 \\ 36.0 \end{pmatrix},$$

$$\begin{aligned} \hat{a}_1 &= (38.5 - \mu - 7a_2 + 7a_3 + 14a_4)/29 \\ &= -.1349 \\ a_1 &= \hat{a}_1 + RND * (\sigma_e^2/29)^{.5} \\ &= -.1349 + (1.138)(1.794) \\ &= 1.9067 \end{aligned}$$

$$\begin{pmatrix} 5 & 1 & 1 & 1 & 1 & 1 \\ 1 & 29 & 7 & -7 & -14 & 0 \\ 1 & 7 & 30 & -14 & 8 & -16 \\ 1 & -7 & -14 & 36 & -14 & 0 \\ 1 & -14 & 8 & -14 & 37 & -16 \\ 1 & 0 & -16 & 0 & -16 & 33 \end{pmatrix} \begin{pmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} = \begin{pmatrix} 238.2 \\ 38.5 \\ 48.9 \\ 64.3 \\ 50.5 \\ 36.0 \end{pmatrix},$$

$$\begin{aligned} e_1 &= 38.5 - 42.41 - 1.9067 = -5.8167 \\ e_2 &= 48.9 - 42.41 + .2203 = 6.7103 \\ e_3 &= 64.3 - 42.41 + .8879 = 22.7779 \\ e_4 &= 50.5 - 42.41 + 1.3006 = 9.3906 \\ e_5 &= 36.0 - 42.41 + 2.0200 = -4.3900 \\ \mathbf{e}'\mathbf{e} &= 705.1503 \end{aligned}$$

$$\begin{aligned} \sigma_e^2 &= (\mathbf{e}'\mathbf{e} + v_e S_e^2)/CHI(15) \\ &= (705.1503 + (10)(93.3333))/17.1321 \\ &= 95.6382. \end{aligned}$$

Additive Genetic Variance

$$\mathbf{A}^{-1} = \mathbf{T}^{-1}\mathbf{D}^{-2}\mathbf{T}'^{-1},$$

$$\begin{aligned}\mathbf{T}'^{-1}\hat{\mathbf{a}} &= \hat{\mathbf{m}} \\ &= (\hat{a}_i - 0.5(\hat{a}_s + \hat{a}_d)),\end{aligned}$$

$$\begin{aligned}\hat{\mathbf{a}}'\mathbf{A}^{-1}\hat{\mathbf{a}} &= \hat{\mathbf{a}}'\mathbf{T}^{-1}\mathbf{D}^{-2}\mathbf{T}'^{-1}\hat{\mathbf{a}} \\ &= \hat{\mathbf{m}}'\mathbf{D}^{-2}\hat{\mathbf{m}} \\ &= \sum_{i=1}^{q} \hat{m}_i^2 d^{ii}, \\ \mathbf{a}'\mathbf{A}^{-1}\mathbf{a} &= 19.85586.\end{aligned}$$

$$\begin{aligned}\sigma_a^2 &= (\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} + v_a S_a^2)/CHI(15) \\ &= (19.85586 + (10)(6.66667))/10.7341 \\ &= 8.0605.\end{aligned}$$

Gibbs Samples

- A 'burn-in' period is necessary.
- Samples from burn-in not used.
- Run two chains beginning with the same random seed, but different starting values.
- When samples converge, burn-in is over.
- Do post analyses:
  - Determine number of independent samples,
  - Average and SD of independent samples gives estimate and its SE, mode, distribution.
- One long chain with at least 100 independent samples??

Data Preparation

- Takes longer than analysis.
- Data file.
- Pedigree file.

Pedigree Files

Animal	Sire	Dam	Birth Year
348097	351604	351342	1981
349876	352515	350873	1981
350010	348097	349876	1981
350873	0	0	1980
351604	350121	0	1980
352012	352229	350873	1982
353118	350873	351604	1981

- IDs appear as 'sire' or 'dam' but not as 'animal'.
- 'Sire' and 'Dam' IDs sometimes larger than progeny ID.
- IDs appear as both 'sire' and 'dam'.
- IDs do not follow birth year ordering.
- Some 'animals' have missing parentage.

Sequencing Pedigrees

Animal	Sire	Dam	1st	2nd	3rd	Code
A	D	E	1			
B	F	G	1			
C	A	B	1			
D	–	–	1			
E	–	–	1			
F	–	–	1			
G	–	–	1			

Sequencing Pedigrees

Animal	Sire	Dam	1st	2nd	3rd	Code
A	D	E	1	2		1
B	F	G	1	2		2
C	A	B	1	1		0
D	–	–	1	2		1
E	–	–	1	2		2
F	–	–	1	2		1
G	–	–	1	2		2

Sequencing Pedigrees

Animal	Sire	Dam	1st	2nd	3rd	Code
A	D	E	1	2	2	1
B	F	G	1	2	2	2
C	A	B	1	1	1	0
D	–	–	1	2	3	1
E	–	–	1	2	3	2
F	–	–	1	2	3	1
G	–	–	1	2	3	2

Sequencing Pedigrees

Animal	Sire	Dam	1st	2nd	3rd	Code
A	D	E	1	2	2	1
B	F	G	1	2	2	2
C	A	B	1	1	1	0
D	-	-	1	2	3	1
E	-	-	1	2	3	2
F	-	-	1	2	3	1
G	-	-	1	2	3	2

Gen	Animal	Sire	Dam	Code
3	D	-	-	1
3	F	-	-	1
3	E	-	-	2
3	G	-	-	2
2	A	D	E	1
2	B	F	G	2
1	C	A	B	0

Renumber Animals

Gen	Animal	Sire	Dam	Code
3	1(D)	-	-	1
3	2(F)	-	-	1
3	3(E)	-	-	2
3	4(G)	-	-	2
2	5(A)	1(D)	3(E)	1
2	6(B)	2(F)	4(G)	2
1	7(C)	5(A)	6(B)	0

Progeny numbers are always greater than their parent numbers.

```
pedf03.f - Assign generation numbers
sort output

pedf04.f - Re-number animals
sort output
inbrd.f
```

Inbreeding Calculations

1) Initialize pedigree memory structure
   `call fxini(nam)  nam=number of animals`

2) Add an individual, sire,dam = int
   `call fxadd(sire,dam,fi) fi=r*4,`

3) Get a relationship between i and j
   `call fxrel(i,j,r)          r=real*4`

4) Get within family (Mendelian) genetic
   `call fxdd(sire,dam,d)`

5) Get sire of animal i, or dam
   `call fxsire(i,isire)`
   `call fxdam(i,idam)`

6) Free up memory
   `call fxfree`

Software Packages

- Require data file separate from pedigree file.
- Require specific formats for data and pedigree.
- Operate via a 'control', 'parameter', or 'driver' file.

Data File

Can be re-numbered using sorted output from `pedf03.f`.

- Animal ID
- All Integer Variables
- All Real Variables
- No characters
- Blank space between all variables

```
Animal Herd Year Month Age Milk
2468    98  1988   10   34 7632.

    Fat  Protein  Wt
    329.  275.    524.
```

PAMST - LRS

Programs for Animal Models and Single Traits

pamst1.f - reads control file and prepares files

pamst2.f - performs Gibbs sampling or BLUP

```
Read(9,*,end=99)ia, (ix(j),j=1,mfix),
  (ty(k),k=1,mrel)
```

Control File

Order of input is critical for `pamst1.f` and most other software.

```
$ Give name of analysis and a date
Test Analysis of Test Data,July 21, 2004
$
$ Name of pedigree file
PRACN.d
$
$ Name of data file
test1.d
$
$ Number of integer, real values
    4  4
$ Names of integer and real variables, in
Herd
Year
Month
Age
Milk
```

```
Fat
Protein
Weight
$ Type of Analysis - Gibbs or BLUP
Gibbs
$
$ Random number seed, number of samples
15743298  50000
$
$ Trait to be analyzed / minimum value
Milk
500.
$
$ Number of fixed factors in model
 3
Herd
Year
Age
$ Number of covariates
   1
Weight
```

```
$
$ Number of random factors
$ / names / df, ratio, var
  2
  additive
   5. 1.5  125666.
  residual
   5. 1.0  188500.
  $
  $
```

PAMST

pamst01.f

- Creates a coded pedigree file, needs to be sorted.
- Creates a new data file, reduced for particular model.
- Creates a PARM2.f file - has necessary numbers for assigning memory allocation in pamst02.f

Post Analyses

1. SAS can be used to plot the sample results for each component estimated. This is useful to determine the length of burn-in needed.
2. autoc.f is another useful program, to determine the number of independent samples. Number of samples, length of burn-in and names of the components must be specified. Average and SD of sample values over ALL samples post burn-in are calculated.
3. Use SAS to plot distribution of sample values.

Estimation of Variance Components in Animal Breeding

✓	Day 1	Distributions
✓		Building Blocks
✓		Unbiased Estimation
✓	Day 2	Likelihood Methods
✓		ML, REML
✓	Day 3	Bayesian Estimation
✓		Data Preparation
✓		Simple Animal Models
	Day 4	Software
		VCE, DMU, MTDFREML
		Models
	Day 5	Other Models

Random Numbers

pamst02.f used random number generators for the normal and Chi-square distributions.

```
  Integer iseed,ndf,n,ncat,ix()
  Real*8 u,p,chi, alpha,beta
Initialize generators:
  call firan(iseed)

Uniform: call fguni1(u)
Normal:  call fgnor3(p)
Chi-sq:  call fgchi1(ndf,chi)
Multinom: call fgmul1(n,p,ncat,ix)
Gamma:   call fggam1(alpha,beta,r)
Exponent: call fgexp1(alpha,p)
Beta:    call fgbet1(alpha,beta,p)
```

Software Packages

- DMU - Denmark, Just Jensen, Per Madsen
- VCE - Eildert Groeneveld
- MTDFREML - Boldman
- ASREML - Arthur Gilmour
- Karin Meyer, Ignacy Misztal

DMU Package

- DMU1 - must always be used. Reads the control file, takes care of pedigrees and data. Pedigree information can be specified in different ways.
- DMU4 - Solves MME for EBVs.
  1. JCG - Jacobi Conjugate Gradient
  2. JSI - Jacobi Semi-Iteration
  3. SOR - Successive Overrelaxation
  4. SSOR - Symmetric SOR Conjugate Gradient
  5. SSOR - Semi-Iteration
  6. RCG - Reduced System Conjugate Gradient
  7. RSI - Reduced System Semi-Iteration
  8. FSPAK - speed, memory optimized
- DMUAI - AI or EM REML.
- DMUGib - Bayesian method via Gibbs sampling.

DMU1

```
$COMMENT
Example driver file
Single trait animal model
Trait is T1
Fixed factors are   RC, RAS
Random factors are  HRC, ID
```

Up to 10 lines of comments. These appear in the output.

```
$ANALYSE 1 1 0 0
```

Task	1 = REML 11 = BLUP
Method task 1	1 = AI REML 2 = EM REML, Thompson 3 = EM REML, Mantysaari 4 = AI (step halving)
Method task 11	1 = JCG, JSI, etc.
Scaling	0 = no, 1 = yes
Printout	0 = standard, 1, 2

```
$DATA  ASCII (6,8,-99) /u/name/dairy.d
```

ASCII or Binary

6 = Number of integer variables

8 = Number of real variables

-99 = indicates missing values

File pathname

```
$VARIABLE
 ID RC RAS HRC HS DAM ROUND
 LP1 LP2 LP3 LP4 LP5 T1 T2
```

First 6 are integer variables.

Last 8 are real variables.

Use any names.

This command is optional.

```
$MODEL
1
0
7 0 4 2 3 4 1
2 1 2
0
0
```

1. Number of traits, 1.
2. Intended for future releases, one 0 for each trait.

3. Model

```
$VARIABLE
 ID RC RAS HRC HS DAM ROUND
  1  2  3   4   5  6    1
 LP1 LP2 LP3 LP4 LP5 T1 T2
  2   3   4   5   6   7  8
```

7 0 4 2 3 4 1

7 trait is $T1$

0 no weighting of observations

4 number of fixed and random factors

2 RC

3 RAS

4 HRC

1 Animal ID

The model is

$$T1 = RC + RAS + HRC + ID + e$$

4. Random.

   `2 1 2`

   First number is the number of random factors, assumed to be the last two factors in the model, i.e. HRC and ID. HRC has structure 1, and ID has structure 2.

5. Regressions. In this case 0.
6. Number of covariances among residual effects. In this case 0.

```
$VAR_STR 2 PED 1 ASCII /u/name/ped.d
```

Random factors are assumed to have the assumed structure, $\mathbf{I}\sigma^2$ unless otherwise indicated. This command allows the user to indicate which factor should use the additive genetic relationship matrix. Here the factor with structure 2. The 1 after PED indicates how the inverse should be constructed, there are six possibilities.

1. = Sires and dams, inbred situation.
2. = Sires and dams, non-inbred situation.
3. = Sires and MGS, inbred.
4. = Sires and MGS, non-inbred.
5. = Not used.
6. = Same as 2 with phantom groups.

Then the ASCII or BINARY version of the pedigree file is given with its pathname.

DMU - Driver File

There are other commands to specify starting values for the variances and covariances, to indicate optional input for DMU4, or to indicate optional input for DMUAI.

Example Driver Files in the notes, pages 78-80.

Comments on DMU

- Driver files are easy.
- How does it get starting values?
- Should always use inbred situation?
- Model specification is tricky.
- Maternal effects models do not always work correctly.
- Programs are being upgraded continually.
- When programs fail, not easy to figure out.
- Help from Denmark is tricky.

VCE

Control files are called *parameter files* or *pfiles*. VCE is keyword oriented.

VCE can perform AI REML, EM REML, or Bayesian methods via Gibbs Sampling.

VCE does not distinguish between capital or small letters.

The documentation states that VCE can handle dominance genetic effects and random regression models.

The keywords are

```
COMMENT        DATA        MODEL
COVARIANCE     SYSTEM      OUTPUT
END
```

```
COMMENT  job = labproblem
  The job name will appear on
  all output.
  Unlimited number of lines.
  Describe the analysis.
```

The words "data" or "Model" or "covariance" cannot be used in the Comment section, because these are keywords.

DATA keyword

Three types of data sets.

```
DATA
 datfile  = 'dairy.d'
 format = '(4f12.0,8f8.0)'
 dep = AFS NRF NRC CTFS
 indep = Animal YS YSH MF SS AP AF PE
 group_by = MF
 header = 0
 crossbreeding = .false.
 ;
```

The data must be in a different sequence for VCE compared to DMU and PAMST.

Format must be all 'real' even if some factors are integer.

`Group by` is to estimate different residual variances, in this case, for each level of `MF`.

`header` indicates number of lines to skip in data file.

```
DATA
 pedfile = 'ped.d'
 format = '(4i10)'
 link = animal
 dominance = fanimal
 indep = litter
 header = 0
 ;
```

Pedigree file assumes first three variables are animal, sire, dam. If there are more variables then these are specifed in `indep`.

```
DATA
 ranfile = 'rfile.d'
 format = '(f8.0)'
 link = YSH
 indep =
 header = 0
 ;
```

A `ranfile` identifies levels of heterogeneous variances.

```
MODEL
  AFS = YS YSH ANIMAL;
  NRF NRC = YS MF AF YSH SS ANIMAL;
  CTFS = YS AP YSH ANIMAL PE;
```

A variable should only appear once in the equation.

`p2(age)` covariates of age.

`pa2(age)` covariates of (age-mean).

`[1, p2(age)]year` covariates nested within `year`.

`[p5(dim)]animal` and `[plg5(dim)]animal`.

```
COVARIANCE
  ANIMAL ;
  YSH ;
  SS: NRF NRC;
  PE: CTFS;
```

```
SYSTEM
 method = 'GI'
 mc_seed = 38742209
 burn_max = 5000
 burn_stop = .0001
 restart = .true.
 inbreeding = .true.
 missing_value = -99
 non_zero = 1000000
```

`GI` is Gibbs sampling, `AG` is AI REML, `AE` is EM REML.

```
OUTPUT
  covfile = 'name' format='( )' next=1;
  inbreeding = 'name' format='( )' next=0
  gibbs_log = 'name';
```

Other options are available. `covfile` will contain the sample values (after burn-in) for each parameter.

```
END
```

Comments on VCE

- Easy to specify models, flexibility.
- Groeneveld may not always answer requests for help, but usually very helpful when he does respond.
- Lengthy manual.
- Need to know the keywords.
- PEST programs.

MTDFREML

- Latest documentation was April 1995.
- Interactive on a PC.
- Over 30 questions to be answered.
- Only DF REML is used.
- The Simplex method for finding the maximum of the likelihood is used.
- The package consists of `MTDFNRM`, `MTDFPREP`, and `MTDFRUN`.
- User supplied data and pedigree information should be readable in free format.

MTDFNRM

Prepares the inverse of the additive genetic relationship matrix. The following information is needed.

1. Animal model(0) or Sire-MGS model(1)
2. Largest animal ID in pedigree file, 99999999
3. Smallest animal ID in pedigree file, 0
4. Name of pedigree file, ped.d
5. Original IDs and inbreeding values wanted, yes
6. Number of integer fields in file
7. Which one is Animal ID
8. Which one is Sire ID
9. Which one is Dam ID
10. Number of genetic groups

Files from MTDFNRM

- MTDF11 - list of animals, coded, original.
- MTDF13 - coded ID, inbreeding coefficients.
- MTDF44 - inverse elements, $\ln \mid \mathbf{A} \mid$.
- MTDF56 - summary information.

MTDFPREP

Data files as in PAMST and DMU. `PARAM.DAT` file must be provided.

`MAXTRT`	`Number of traits`
`MAXINTR`	`Number of integer variables`
`MAXR8`	`Number of real variables`
`MAXANIM`	`Number of animals (with records, ancestors and groups)`
`MAXCOV`	`Number of covariates per trait`
`MAXNFR`	`Number of regression coefficients per trait`
`MAXFIX`	`Number of fixed effects per trait`
`MAXNFL`	`Number of levels for a fixed effect`
`MAXCONS`	`Number of constraints`
`MAXRAN`	`Number of uncorrelated`

	random effects per trait
MAXNRL	Number of levels for uncorrelated random effects
MAXINV	Order of submatrix to be invert
MAXORDS	Used by SPARSPAK
MAXNZE	Used by FSPAK
NHASH	Used by FSPAK for Hashing

MTDFPREP

See Page 89 of notes.

6 lines of comments

Input of information is much more friendly with DMU or VCE.

MTDFRUN

- Type of Analysis, new(0) or old(0)
- Options, estimate variances, solve MME, sampling variances.
- Another series of questions, Page 90.
- Number of Simplex Rounds, greater than number of parameters.

Comments on MTDFREML

- Handles maternal effects models appropriately.
- Very cumbersome input of information.
- Manual and programs out of date.
- Only DF REML can be run.
- Many files created.
- Restart capabilities.

Maternal Effects Models

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{m} + \mathbf{Z}_3\mathbf{p} + \mathbf{e},$$

$$Var\begin{pmatrix} \mathbf{a} \\ \mathbf{m} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{A}\sigma_a^2 & \mathbf{A}\sigma_{am} & \mathbf{0} & \mathbf{0} \\ \mathbf{A}\sigma_{am} & \mathbf{A}\sigma_m^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_p^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2 \end{pmatrix},$$

$$\left(\begin{array}{c} \mathbf{a} \\ \\ \mathbf{m} \end{array} \middle| \mathbf{A}, \mathbf{G} \right) \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \; \mathbf{G} \otimes \mathbf{A} \right),$$

where

$$\mathbf{G} = \begin{pmatrix} \sigma_a^2 & \sigma_{am} \\ \sigma_{am} & \sigma_m^2 \end{pmatrix},$$

and

$$\mathbf{p} \mid \mathbf{I}, \sigma_p^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_p^2),$$

and

$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2).$$

Simulation of Data

$$\mathbf{G} = \begin{pmatrix} \sigma_a^2 & \sigma_{am} \\ \sigma_{am} & \sigma_m^2 \end{pmatrix} = \begin{pmatrix} 49 & -7 \\ -7 & 26 \end{pmatrix}.$$

$$\begin{aligned} \mathbf{G} &= \mathbf{LL}' \\ \mathbf{L} &= \begin{pmatrix} 7 & 0 \\ -1 & 5 \end{pmatrix}. \end{aligned}$$

Let $\sigma_p^2 = 9$ and $\sigma_e^2 = 81.$

Three Animals

A and B are parents of C.

$$\begin{aligned}\begin{pmatrix} a_A \\ m_A \end{pmatrix} &= \mathbf{Lw} \\ &= \begin{pmatrix} 7 & 0 \\ -1 & 5 \end{pmatrix}\begin{pmatrix} 2.533 \\ -.299 \end{pmatrix} \\ &= \begin{pmatrix} 17.731 \\ -4.028 \end{pmatrix}.\end{aligned}$$

$$\begin{aligned}\begin{pmatrix} a_B \\ m_B \end{pmatrix} &= \begin{pmatrix} 7 & 0 \\ -1 & 5 \end{pmatrix}\begin{pmatrix} -1.141 \\ .235 \end{pmatrix} \\ &= \begin{pmatrix} -7.987 \\ 2.316 \end{pmatrix}.\end{aligned}$$

$$\begin{aligned}\begin{pmatrix} a_C \\ m_C \end{pmatrix} &= \frac{1}{2}\begin{pmatrix} a_A + a_B \\ m_A + m_B \end{pmatrix} + (b_{ii})^{.5}\mathbf{L}\mathbf{w} \\ &= \frac{1}{2}\begin{pmatrix} 17.731 - 7.987 \\ -4.028 + 2.316 \end{pmatrix} + (\frac{1}{2})^{.5}\mathbf{L}\begin{pmatrix} .275 \\ .402 \end{pmatrix} \\ &= \begin{pmatrix} 6.233 \\ .371 \end{pmatrix}.\end{aligned}$$

Record for C

$$\begin{pmatrix} a_C \\ m_C \end{pmatrix} = \begin{pmatrix} 6.233 \\ .371 \end{pmatrix},$$

$$\begin{pmatrix} a_B \\ m_B \end{pmatrix} = \begin{pmatrix} -7.987 \\ 2.316 \end{pmatrix}.$$

$$p_B = \sigma_p * RND = -4.491.$$

Then

$$\begin{aligned} y &= \mu + a_C + m_B + p_B + \sigma_e * RND \\ &= 140 + 6.233 + 2.316 + (-4.491) + (9)(1.074) \\ &= 153.724. \end{aligned}$$

MME

$$\left(\begin{array}{cccc} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_1 & \mathbf{X}'\mathbf{Z}_2 & \mathbf{X}'\mathbf{Z}_3 \\ \mathbf{Z}_1'\mathbf{X} & \mathbf{Z}_1'\mathbf{Z}_1 + \mathbf{A}^{-1}k_{11} & \mathbf{Z}_1'\mathbf{Z}_2 + \mathbf{A}^{-1}k_{12} & \mathbf{Z}_1'\mathbf{Z}_3 \\ \mathbf{Z}_2'\mathbf{X} & \mathbf{Z}_2'\mathbf{Z}_1 + \mathbf{A}^{-1}k_{12} & \mathbf{Z}_2'\mathbf{Z}_2 + \mathbf{A}^{-1}k_{22} & \mathbf{Z}_2'\mathbf{Z}_3 \\ \mathbf{Z}_3'\mathbf{X} & \mathbf{Z}_3'\mathbf{Z}_1 & \mathbf{Z}_3'\mathbf{Z}_2 & \mathbf{Z}_3'\mathbf{Z}_3 + \mathbf{I}k \end{array}\right.$$

$$\begin{aligned} \begin{pmatrix} k_{11} & k_{12} \\ k_{12} & k_{22} \end{pmatrix} &= \begin{pmatrix} \sigma_a^2 & \sigma_{am} \\ \sigma_{am} & \sigma_m^2 \end{pmatrix}^{-1} \sigma_e^2, \\ &= \begin{pmatrix} 49 & -7 \\ -7 & 26 \end{pmatrix}^{-1} (81), \\ &= \begin{pmatrix} 1.7192 & .4628 \\ .4628 & 3.2400 \end{pmatrix}. \end{aligned}$$

Finally, $k_{33} = \sigma_e^2/\sigma_p^2 = 81/9 = 9.$

Example

Animal	Sire	Dam	CG	Weight
5	1	3	1	156
6	2	3	1	124
7	1	4	1	135
8	2	4	2	163
9	1	3	2	149
10	2	4	2	138

EM-REML

$$\hat{\sigma}_e^2 = (\mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{W}'\mathbf{y})/(N - r(\mathbf{X})).$$

$$q\begin{pmatrix} \hat{\sigma}_a^2 & \hat{\sigma}_{am} \\ \hat{\sigma}_{am} & \hat{\sigma}_m^2 \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{a}}'\mathbf{A}^{-1}\hat{\mathbf{a}} & \hat{\mathbf{a}}'\mathbf{A}^{-1}\hat{\mathbf{m}} \\ \hat{\mathbf{m}}'\mathbf{A}^{-1}\hat{\mathbf{a}} & \hat{\mathbf{m}}'\mathbf{A}^{-1}\hat{\mathbf{m}} \end{pmatrix} + \hat{\sigma}_e^2\begin{pmatrix} tr(\mathbf{A}^{-1}\mathbf{C}_{11}) & tr(\mathbf{A}^{-1}\mathbf{C}_{12}) \\ tr(\mathbf{A}^{-1}\mathbf{C}_{21}) & tr(\mathbf{A}^{-1}\mathbf{C}_{22}) \end{pmatrix}.$$

$$\hat{\sigma}_p^2 = (\hat{\mathbf{p}}'\hat{\mathbf{p}} + \hat{\sigma}_e^2 tr\mathbf{C}_{33})/2.$$

Bayesian Estimation

$$\mathbf{G}_i = \begin{pmatrix} \mathbf{a}_i'\mathbf{A}^{-1}\mathbf{a}_i & \mathbf{a}_i'\mathbf{A}^{-1}\mathbf{m}_i \\ \mathbf{m}_i'\mathbf{A}^{-1}\mathbf{a}_i & \mathbf{m}_i'\mathbf{A}^{-1}\mathbf{m}_i \end{pmatrix}.$$

1. Invert the i^{th} sample matrix, $\mathbf{G}_i$.
2. Compute the Cholesky decomposition of this inverse,
$$\mathbf{T} = Chol(\mathbf{G}_i^{-1}),$$
where $\mathbf{T}$ is a lower triangular matrix.
3. Generate a new sample for $\mathbf{G}_{i+1}^{-1}$ from a Wishart distribution based on q degrees of freedom, where q is the number of animals, in this case $q = 10$.
4. Invert the previous matrix to give $\mathbf{G}_{i+1}$.

$$\sigma_e^2 = \hat{\mathbf{e}}'\hat{\mathbf{e}}/\chi_N^2,$$

$$\sigma_p^2 = \hat{\mathbf{p}}'\hat{\mathbf{p}}/\chi_p^2.$$

Maternal Effects Models

- Must have good data structure.
- Research herds versus field data.
- i.e. Females must have a birthweight, and must have progeny with birthweights.
- Most field data are not appropriate.

Estimation of Variance Components in Animal Breeding

✓	Day 1	Distributions
✓		Building Blocks
✓		Unbiased Estimation
✓	Day 2	Likelihood Methods
✓		ML, REML
✓	Day 3	Bayesian Estimation
✓		Data Preparation
✓		Simple Animal Models
✓	Day 4	Software
✓		VCE, DMU, MTDFREML
✓		Maternal Effects
	Day 5	Random Regression
		Multiple Traits
		Non-Additive Genetic Effects

Random Regression Models

Animal	Days on Test					
	10	20	30	40	50	60
1	42	53	60	72	83	94
2	30	50	58	68	76	85
3	38	44	51	60	70	77
SD	1.6	3.7	3.9	5.0	5.3	5.6

Longitudinal Data

Infinitely Dimensional

Six Traits or One?

$$\mathbf{G} = \begin{pmatrix} 2.5 & 4.9 & 4.6 & 4.6 & 4.3 & 4.0 \\ 4.9 & 13.5 & 12.1 & 12.3 & 11.9 & 10.7 \\ 4.6 & 12.1 & 15.2 & 14.5 & 14.6 & 12.5 \\ 4.6 & 12.3 & 14.5 & 20.0 & 19.0 & 16.9 \\ 4.3 & 11.9 & 14.6 & 19.0 & 25.0 & 20.3 \\ 4.0 & 10.7 & 12.5 & 16.9 & 20.3 & 30.0 \end{pmatrix}.$$

$$\mathbf{R} = \begin{pmatrix} 3.8 & 7.4 & 6.9 & 6.8 & 6.4 & 6.0 \\ 7.4 & 20.3 & 18.2 & 18.4 & 17.9 & 16.1 \\ 6.9 & 18.2 & 22.8 & 21.8 & 21.9 & 18.8 \\ 6.8 & 18.4 & 21.8 & 30.0 & 28.5 & 25.4 \\ 6.4 & 17.9 & 21.9 & 28.5 & 37.5 & 30.5 \\ 6.0 & 16.1 & 18.8 & 25.4 & 30.5 & 45.0 \end{pmatrix}.$$

MT MME Solutions

Days on Test	Mean	Animal 1	Animal 2	Animal 3
10	36.67	2.10	-2.61	0.51
20	49.00	1.57	0.45	-2.02
30	56.33	1.48	0.64	-2.12
40	66.67	2.21	0.39	-2.60
50	76.33	2.72	-0.24	-2.48
60	85.33	3.48	-0.16	-3.32

Covariance Functions

Animal 1		Animal 2		Animal 3	
Age	Wt	Age	Wt	Age	Wt
28	42	32	30	35	38
38	53	42	50	45	44
48	60	52	58	55	51
58	72	62	68	65	60
68	83	72	76	75	70
78	94	82	85	85	77

Orthogonal Polynomials

$$\begin{aligned} P_0(x) &= 1, \quad \text{and} \\ P_1(x) &= x, \end{aligned}$$

$$P_{n+1}(x) = \frac{1}{n+1}\left((2n+1)xP_n(x) - nP_{n-1}(x)\right).$$

These quantities are "normalized" using

$$\phi_n(x) = \left(\frac{2n+1}{2}\right)^{.5} P_n(x).$$

Legendre Polynomials

$$
\begin{aligned}
\phi_0(x) &= \left(\frac{1}{2}\right)^{.5} P_0(x) = .7071 \\
\phi_1(x) &= \left(\frac{3}{2}\right)^{.5} P_1(x) \\
&= 1.2247x \\
P_2(x) &= \frac{1}{2}(3xP_1(x) - 1P_0(x)) \\
\phi_2(x) &= \left(\frac{5}{2}\right)^{.5} (\frac{3}{2}x^2 - \frac{1}{2}) \\
&= -.7906 + 2.3717x^2,
\end{aligned}
$$

$$
\Lambda' = \left(\begin{array}{rrrrrr}
.7071 & 0 & 0 & 0 & 0 & \\
0 & 1.2247 & 0 & 0 & 0 & \\
-.7906 & 0 & 2.3717 & 0 & 0 & \\
0 & -2.8062 & 0 & 4.6771 & 0 & \\
.7955 & 0 & -7.9550 & 0 & 9.2808 & \\
0 & 4.3973 & 0 & -20.5206 & 0 & 1
\end{array} \right.
$$

Standardized Time

x in Legendre polynomials must range from -1 to +1.

$$q_\ell = -1 + 2\left(\frac{t_\ell - t_{min}}{t_{max} - t_{min}}\right).$$

Days on Test	Age	Standardized Value
10	31	-1.000
20	41	-.600
30	51	-.200
40	61	.200
50	71	.600
60	81	1.000

Covariance Function

$$\begin{aligned}\mathbf{G} &= \Phi\mathbf{H}\Phi' \\ &= \mathbf{M}(\Lambda\mathbf{H}\Lambda')\mathbf{M}' \\ &= \mathbf{MTM}'.\end{aligned}$$

$$\mathbf{M} = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -.600 & .360 & -.216 & .130 & -.078 \\ 1 & -.200 & .040 & -.008 & .002 & -.000 \\ 1 & .200 & .040 & .008 & .002 & .000 \\ 1 & .600 & .360 & .216 & .130 & .078 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

$$\begin{aligned}\Phi &= \mathbf{M}\Lambda, \\ &= \begin{pmatrix} .7071 & -1.2247 & 1.5811 & -1.8708 & 2.1213 & -2.3 \\ .7071 & -.7348 & .0632 & .6735 & -.8655 & .3 \\ .7071 & -.2449 & -.6957 & .5238 & .4921 & -.7 \\ .7071 & .2449 & -.6957 & -.5238 & .4921 & .7 \\ .7071 & .7348 & .0632 & -.6735 & -.8655 & -.3 \\ .7071 & 1.2247 & 1.5811 & 1.8708 & 2.1213 & 2.3 \end{pmatrix}\end{aligned}$$

$$\begin{aligned}\mathbf{G} &= \Phi\mathbf{H}\Phi' \\ &= \mathbf{M}(\Lambda\mathbf{H}\Lambda')\mathbf{M}' \\ &= \mathbf{MTM}'.\end{aligned}$$

$$\begin{aligned}\mathbf{H} &= \Phi^{-1}\mathbf{G}\Phi^{-T}, \\ &= \begin{pmatrix} 27.69 & 5.29 & -1.95 & 0.05 & -1.17 & 0.52 \\ 5.29 & 4.99 & 0.42 & -0.25 & -0.30 & -0.75 \\ -1.95 & 0.42 & 1.51 & 0.20 & -0.33 & -0.07 \\ 0.05 & -0.25 & 0.20 & 1.19 & 0.06 & -0.71 \\ -1.17 & -0.30 & -0.33 & 0.06 & 0.58 & 0.15 \\ 0.52 & -0.75 & -0.07 & -0.71 & 0.15 & 1.12 \end{pmatrix},\end{aligned}$$

$$
\begin{aligned}
\mathbf{T} &= \mathbf{M}^{-1}\mathbf{G}\mathbf{M}^{-T} \\
&= \left(\begin{array}{rrrrrr}
16.44 & 6.48 & -5.93 & -11.49 & -0.93 & \\
6.48 & 49.87 & -2.05 & -155.34 & 1.44 & \\
-5.93 & -2.05 & 57.71 & 28.62 & -50.06 & - \\
-11.49 & -155.34 & 28.62 & 635.49 & -26.91 & - \\
-0.93 & 1.44 & -50.06 & -26.91 & 49.80 & \\
10.02 & 111.23 & -25.73 & -486.90 & 26.49 &
\end{array}\right.
\end{aligned}
$$

See Page 106.

Reduced orders of fit.

Model

$$y_{ijkn:t} = F_i + g(t)_j + r(a, x, m1)_k + r(pe, x, m2)_k + e_{ij}$$

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{p} + \mathbf{e},$$

$$Var\begin{pmatrix}\mathbf{a}\\ \mathbf{p}\\ \mathbf{e}\end{pmatrix} = \begin{pmatrix}\mathbf{A}\otimes\mathbf{G} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\otimes\mathbf{P} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{R}\end{pmatrix},$$

Example, Page 112

Solutions

$$\hat{\mathbf{a}} = \begin{pmatrix} -1.747298 & .124789 & -.001223 \\ 5.774393 & -.553689 & .005612 \\ -2.899020 & .475908 & -.004998 \\ -4.926784 & .159792 & -.001347 \\ -2.002508 & .301390 & -.003149 \\ 3.285314 & -.297302 & .002997 \\ 1.692846 & -.215472 & .002232 \\ -2.975451 & .211306 & -.002080 \end{pmatrix}.$$

The Legendre polynomials for ages 24, 36, and 48 mo are given in the rows of the following matrix $\mathbf{L}$,

$$\mathbf{L} = \begin{pmatrix} .7071 & -.8328 & .3061 \\ .7071 & -.3429 & -.6046 \\ .7071 & .2449 & -.6957 \end{pmatrix}.$$

Animal	EBV(24)	EBV(36)	EBV(48)
1	-1.34	-1.28	-1.20
2	4.55	4.27	3.94
8	-2.28	-2.18	-2.05

EM REML

$$\hat{\mathbf{a}}'\mathbf{A}^{-1}\hat{\mathbf{a}} = \begin{pmatrix} 66.679755 & -4.925545 & .048641 \\ -4.925545 & .557022 & -.005720 \\ .048641 & -.005720 & .000059 \end{pmatrix}.$$

$$tr_3((\mathbf{A}^{-1}\otimes\mathbf{I}_3)\mathbf{C}_{aa}) = \begin{pmatrix} 68.347139 & -2.689821 & .02 \\ -2.689821 & 1.284185 & -.01 \\ .021540 & -.012359 & .00 \end{pmatrix}$$

$$\begin{aligned} \hat{\mathbf{G}} &= (\hat{\mathbf{a}}'\mathbf{A}^{-1}\hat{\mathbf{a}} + tr_3((\mathbf{A}^{-1}\otimes\mathbf{I}_3)\mathbf{C}_{aa})\hat{\sigma}_e^2)/8, \\ &= \begin{pmatrix} 157.476730 & -6.485209 & .053082 \\ -6.485209 & 2.871876 & -.027685 \\ .053082 & -.027685 & .002714 \end{pmatrix}. \end{aligned}$$

$$\hat{\mathbf{P}} = (\hat{\mathbf{p}}'\hat{\mathbf{p}} + tr_3(\mathbf{C}_{pp})\hat{\sigma}_e^2)/4.$$

Bayesian Estimation

A change in handling animal solutions in MME.

$$(\mathbf{Z}_i'\mathbf{Z}_i + a^{ii}\mathbf{G}_k^{-1}\sigma_{e_k}^2) =$$

$$\mathbf{Q}_{i_k} = \begin{pmatrix} 1.7675 & -.2017 & 1.0346 \\ -.2017 & 19.7604 & 173.6637 \\ 1.0346 & 173.6637 & 17839.79 \end{pmatrix},$$

$$\hat{\mathbf{a}}_i = \begin{pmatrix} -1.7473 \\ .1248 \\ -.0012 \end{pmatrix},$$

Current sample value of σ_e^2 was 17.4570. To generate a new set of sample additive genetic effects for animal i then

1. Invert $\mathbf{Q}_{i_k}$ and multiply by the current sample value of σ_e^2 which gives

$$\mathbf{Q}_{i_k}^{-1}\sigma_e^2 = \begin{pmatrix} 9.8906 & .115912 & -.001720 \\ .115912 & .967441 & -.009424 \\ -.001720 & -.009424 & .001070 \end{pmatrix}.$$

2. Apply a Cholesky decomposition to the previous matrix,

$$Chol(\mathbf{Q}_{i_k}^{-1}\sigma_e^2) = \begin{pmatrix} 3.144938 & 0 & 0 \\ .036857 & .982895 & 0 \\ -.000541 & -.009568 & .031282 \end{pmatrix}.$$

3. Generate a vector of three random normal deviates, suppose they are

$$\mathbf{f}' = \begin{pmatrix} 1.0673 & -.5892 & -.9814 \end{pmatrix}.$$

4. The new sample values for animal i additive genetic random regression coefficients is then

$$\begin{aligned} \mathbf{a}_i &= \hat{\mathbf{a}}_i + \mathbf{Lf}, \\ &= \begin{pmatrix} -1.7473 \\ .1248 \\ -.0012 \end{pmatrix} + \left(\begin{matrix} 3.144938 & 0 & 0 \\ .036857 & .982895 & 0 \\ -.000541 & -.009568 & .031282 \end{matrix} \right. \\ &= \begin{pmatrix} 1.6093 \\ -.4150 \\ -.0269 \end{pmatrix}. \end{aligned}$$

$$
\begin{aligned}
\mathbf{c}'\mathbf{c} &= 4.11, \\
\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} &= \begin{pmatrix} 64.71 & -4.15 & .0532 \\ -4.15 & .51 & -.0049 \\ .0532 & -.0049 & .000054 \end{pmatrix}, \\
&= \mathbf{G}_k, \\
\mathbf{p}'\mathbf{p} &= \begin{pmatrix} 37.22 & -1.25 & .0247 \\ -1.25 & .0873 & -.0011 \\ .0247 & -.0011 & .000022 \end{pmatrix}, \\
&= \mathbf{P}_k, \\
\mathbf{e} &= \mathbf{y} - \mathbf{W}\beta, \\
\mathbf{e}'\mathbf{e} &= 74.8851. \\
\sigma_v^2 &= \mathbf{c}'\mathbf{c}/\chi_4^2, \\
\sigma_e^2 &= \mathbf{e}'\mathbf{e}/\chi_{12}^2, \\
\mathbf{T} &= Chol(\mathbf{G}_k^{-1}), \\
\mathbf{G}_{k+1}^{-1} &= \text{Wishart}(\text{T}, 8), \\
\mathbf{S} &= Chol(\mathbf{P}_k^{-1}), \\
\mathbf{P}_{k+1}^{-1} &= \text{Wishart}(\text{S}, 4).
\end{aligned}
$$

Multiple Trait Models

1. Low Heritability Traits.
2. Big difference between genetic and residual correlations.
3. Missing traits.
4. Culling or selection bias.

Animal	Sire	Dam	B-level	C-level	Trait 1	Trait 2
1	0	0	1	1	2.3	
2	0	0	1	2	2.6	
3	0	0	1	3	9.8	53
4	0	0	1	1	4.7	4
5	0	0	1	2	5.5	63
6	1	3	2	3	2.5	
7	1	4	2	2	8.4	35
8	1	5	2	3	8.2	41
9	2	3	2	1	9.0	27
10	2	4	2	1	7.8	32
11	2	5	2	2	2.8	
12	6	10	2	3	7.4	67

Computing Aspects

- Assume the same model for both traits. If the model for trait 1 is

  $$y_{1ij} = B_{1i} + a_{1j} + e_{1ij},$$

  and the model for trait 2 is

  $$y_{2ij} = C_{2i} + a_{2j} + e_{2ij},$$

  then the assumed model could be

  $$y_{tijkl} = B_{tj} + C_{tk} + a_{tl} + e_{tijkl}.$$

  During each round of iteration, the solutions for C_{tk} for trait 1 have to be fixed to 0, and the solutions for B_{tj} for trait 2 have to be fixed to 0.

- Assume that all animals are observed for all traits. Missing observations are assigned to a subclass in which it is the only observation. For example, animal 1 had trait 2 missing, and therefore, it could be assigned to level 4 of factor C. Similarly, animal 2 also was missing trait two, and it could be assigned to level 5 of factor C, and so on. The value of the missing observation used in the analysis does not matter because the observation is 'absorbed' out. However, a 0 is safer to use. Proof is on page 125.

Estimation

- MTDFREML, VCE, DMU all allow multiple trait models, but do not utilize the computing tricks just shown.
- Many different ways to manipulate the log likelihood for DF REML algorithms.
- REML and Bayesian methods are similar to that for the random regression model. Block updating of solutions to MME is necessary.
- Computing time goes up with an increase in number of traits.

- Number of likelihoods to be evaluated in DF REML goes up with the number of parameters to be estimated.
- Number of Gibbs samples, length of burn-in increase.
- Must check that resulting matrices are positive definite.
- Traits with genetic correlations close to 1 or -1 may cause convergence problems.

Non-Additive Genetic Models

There are many types of possible interactions among loci in the genome. A simple model might be

$$y_{ij} = \mu + a_i + d_i + (ad)_i + p_i + e_{ij},$$

where

a_i is the additive genetic effect, the effect of a single allele at a locus, (all such possible effects in the genome),

d_i is the dominance genetic effect, interaction between alleles at one locus, (all possible such interactions in the genome),

$(ad)_i$ is the additive by dominance genetic effect, interaction between two alleles at one locus with one allele at another locus, (all possible such interactions in the genome),

p_i is a permanent environmental effect,

y_{ij}, μ, e_{ij} are obvious, hopefully.

VCV Matrices

$$Var\begin{pmatrix}\mathbf{a}\\ \mathbf{d}\\ \mathbf{ad}\\ \mathbf{p}\\ \mathbf{e}\end{pmatrix}=\begin{pmatrix}\mathbf{A}\sigma^2_{10} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}\sigma^2_{01} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{A\#D}\sigma^2_{11} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma^2_p & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma^2_e\end{pmatrix}.$$

Genomic Relationship Matrix

	1A	1B	2A	2B	3A	3B
1A	1	0	0	0	.5	0
1B	0	1	0	0	.5	0
2A	0	0	1	0	0	.5
2B	0	0	0	1	0	.5
3A	.5	.5	0	0	1	0
3B	0	0	.5	.5	0	1

	XA	XB
WA	.5	.25
WB	0	.5

$$a_{WX} = 0.5 * (.5 + .5 + .25 + 0) = .625,$$

$$d_{WX} = .5(.5) + 0(.25) = .25.$$

MME

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}+\mathbf{A}^{-1}k_{10} & \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z}+\mathbf{D}^{-1}k_{01} & \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z}+(\mathbf{A}\#\mathbf{D})^{-1}k_{11} & \mathbf{Z}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z}+\mathbf{I}k_p \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{d}} \\ \widehat{\mathbf{ad}} \\ \hat{\mathbf{p}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{pmatrix},$$

Note that

$$\begin{aligned} \mathbf{A}^{-1}k_{10}\hat{\mathbf{a}} - \mathbf{D}^{-1}k_{01}\hat{\mathbf{d}} &= \mathbf{0} \\ \mathbf{A}^{-1}k_{10}\hat{\mathbf{a}} - (\mathbf{A}\#\mathbf{D})^{-1}k_{11}\widehat{\mathbf{ad}} &= \mathbf{0} \\ \mathbf{A}^{-1}k_{10}\hat{\mathbf{a}} - \mathbf{I}^{-1}k_p\hat{\mathbf{p}} &= \mathbf{0} \end{aligned}$$

Re-arranging terms, then

$$\hat{\mathbf{d}} = \mathbf{D}\mathbf{A}^{-1}(k_{10}/k_{01})\hat{\mathbf{a}}$$

$$\begin{aligned} \hat{\mathbf{ad}} &= (\mathbf{A}\#\mathbf{D})\mathbf{A}^{-1}(k_{10}/k_{11})\hat{\mathbf{a}} \\ \hat{\mathbf{p}} &= \mathbf{A}^{-1}(k_{10}/k_p)\hat{\mathbf{a}} \end{aligned}$$

1. Using the current sample values for $\mathbf{d}$, $\mathbf{ad}$, and $\mathbf{p}$, adjust the observations and construct the animal model MME for the adjusted $\mathbf{y}$.

$$\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{Z}(\mathbf{d} + \mathbf{ad} + \mathbf{p}),$$

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}k_{10} \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ \mathbf{a} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\tilde{\mathbf{y}} \\ \mathbf{Z}'\tilde{\mathbf{y}} \end{pmatrix}.$$

2. Go through the MME equations and compute new sample values for $\mathbf{b}$ and $\mathbf{a}$ as usual, i.e.,

$$\begin{aligned} \hat{b}_i &= [\mathbf{x}_i'(\tilde{\mathbf{y}} - X_{-i}\mathbf{b}_{-i} - Z_i\mathbf{a})]/\mathbf{x}_i'\mathbf{x}_i, \\ b_i &= \hat{b}_i + RND * (\sigma_e^2/\mathbf{x}_i'\mathbf{x}_i)^{.5}, \end{aligned}$$

and

$$\begin{aligned} \hat{a}_i &= [\mathbf{z}_i'(\tilde{\mathbf{y}} - X_i\mathbf{b} - Z_{-i}\mathbf{a}_{-i}) - \mathbf{A}_{-i}^{-1}k_{10}\mathbf{a}_{-i}]/(\mathbf{z}_i' \\ a_i &= \hat{a}_i + RND * (\sigma_e^2/(\mathbf{z}_i'\mathbf{z}_i + a^{ii}k_{10}))^{.5}. \end{aligned}$$

3. New samples of the epistatic genetic effects are just functions of the samples values for $\mathbf{a}$, namely,

$$\mathbf{d} = \mathbf{D}\mathbf{A}^{-1}(k_{10}/k_{01})\mathbf{a}$$

$$
\begin{aligned}
\mathbf{ad} &= (\mathbf{A}\#\mathbf{D})\mathbf{A}^{-1}(k_{10}/k_{11})\mathbf{a} \\
\mathbf{p} &= \mathbf{A}^{-1}(k_{10}/k_p)\mathbf{a}.
\end{aligned}
$$

4. Let

$$
\begin{aligned}
\mathbf{w}_{01} &= (\mathbf{D}^{-1}\hat{\mathbf{d}}) \;=\; \mathbf{A}^{-1}(k_{10}/k_{01})\hat{\mathbf{a}} \\
\mathbf{w}_{11} &= ((\mathbf{A}\#\mathbf{D})^{-1}\widehat{\mathbf{ad}}) \;=\; \mathbf{A}^{-1}(k_{10}/k_{11})\hat{\mathbf{a}} \\
\mathbf{w}_p &= (\mathbf{I}\hat{\mathbf{p}}) \;=\; \mathbf{A}^{-1}(k_{10}/k_p)\hat{\mathbf{a}}.
\end{aligned}
$$

5. The necessary quadratic forms, are then

$$
\begin{aligned}
&\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} \\
\mathbf{e} &= \mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}(\mathbf{a} + \mathbf{d} + \mathbf{ad} + \mathbf{p}), \\
&\mathbf{e}'\mathbf{e} \\
\mathbf{d}'\mathbf{w}_{01} &= \mathbf{d}'\mathbf{D}^{-1}\mathbf{d}, \\
\mathbf{ad}'\mathbf{w}_{11} &= \mathbf{ad}'(\mathbf{A}\#\mathbf{D})^{-1}\mathbf{ad}, \\
\mathbf{p}'\mathbf{w}_p &= \mathbf{p}'\mathbf{p},
\end{aligned}
$$

6. New samples for variances are

$$
\begin{aligned}
\sigma^2_{10} &= \mathbf{a}'\mathbf{A}^{-1}\mathbf{a}/CHI(q) \\
\sigma^2_{01} &= \mathbf{d}'\mathbf{w}_{01}/CHI(q),
\end{aligned}
$$

$$
\begin{aligned}
\sigma_{11}^2 &= \mathbf{ad}'\mathbf{w}_{11}/CHI(q), \\
\sigma_p^2 &= \mathbf{p}'\mathbf{w}_p/CHI(N_p), \\
\sigma_e^2 &= \mathbf{e}'\mathbf{e}/CHI(N).
\end{aligned}
$$

Estimation of Variance Components in Animal Breeding

√	Day 1	Distributions
√		Building Blocks
√		Unbiased Estimation
√	Day 2	Likelihood Methods
√		ML, REML
√	Day 3	Bayesian Estimation
√		Data Preparation
√		Simple Animal Models
√	Day 4	Software
√		VCE, DMU, MTDFREML
√		Maternal Effects
√	Day 5	Random Regression
√		Multiple Traits
√		Non-Additive Genetic Effects

THE END