Estimation of Variance Components in Animal Breeding

L. R. Schaeffer

July 19-23, 2004
Iowa State University
Short Course

Contents

Chapter 1

Distributions

1.1 Random Variables

A *random variable* is a real-valued function which exists within the domain of a defined sample space. A random variable is conventionally designated by a capital letter, say Y, and the value of Y is denoted by a small letter, say y. The sample space is the range of values that y can be assigned. The small letter y is used in these notes for Y and y.

Random variables can be either discrete or continuous. A discrete random variable can assume only a finite number of distinct values, such as zero or one, for example. A continuous random variable can assume any value within the range of the sample space.

Random variables usually follow a distribution function or probability density which can be described mathematically in most cases. The distribution could have a number of parameters associated with it, such as a mean and variance, and all parameters will generally be designated collectively as a vector $\boldsymbol{\theta}$. The goal of statistical analysis is usually to estimate elements of $\boldsymbol{\theta}$ from the observed random variables, y.

If y represents a random variable from some distribution, then the expectation of y is denoted by

$$E(y) = \mu$$

where $E()$ means *expected value*. The expected value of y depends on its distribution and range of allowable values. The expected value is known as the mean, or the first moment of the distribution. Also of importance is the variance of y that could be expected with that distribution. The variance of a scalar random variable, y, is defined as

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

and is commonly represented as σ_y^2. Variances are known as the second moment of the distribution. Variances are always greater than zero.

1.2 Discrete Random Variables

In the general discrete case, the probability that Y takes the value y, is defined as the sum of the probabilities of all sample points that are assigned the value y. That is,

$$P(Y = y) = p(y).$$

The *probability distribution* of Y lists the probabilities for each value of y. Suppose Y can take on four values with the following probabilities:

y	$p(y)$
0	1/8
1	1/4
2	1/4
3	3/8

Any other values of y are assumed to have zero probability, and the sum of all probabilities is 1, as required of a valid distribution function.

Various kinds of probabilities may be calculated from this table.

$$\begin{aligned} Pr(Y = 0) &= p(0) = \frac{1}{8} \\ Pr(Y = 0 \text{ or } Y = 2) &= p(0) + p(2) = \frac{3}{8} \\ Pr(Y = 1 \text{ or } Y = 2) &= p(1) + p(2) = \frac{1}{2} \end{aligned}$$

The *cumulative distribution function* is

$$F(y) = Pr(Y \leq y),$$

for example,

$$\begin{aligned} F(0) &= p(0), \\ F(1) &= p(0) + p(1) = \frac{3}{8}, \\ F(2) &= p(0) + p(1) + p(2) = \frac{5}{8}, \\ F(3) &= p(0) + p(1) + p(2) + p(3) = 1. \end{aligned}$$

Also, if $y_1 \leq y_2$, then $F(y_1) \leq F(y_2)$. Finally,

$$\begin{aligned} Pr(y_1 < Y \leq y_2) &= Pr(Y \leq y_2) - Pr(Y \leq y_1) \\ &= F(y_2) - F(y_1). \end{aligned}$$

The expected value of a discrete random variable is defined as

$$E(y) = \sum_y y\ p(y).$$

For the example above,

$$E(y) = (0(1/8) + 1(1/4) + 2(1/4) + 3(3/8)) = 1.875.$$

Similarly, the expected value of a function of Y, say $g(Y)$ is given by

$$E(g(y)) = \sum_y g(y)\ p(y).$$

Suppose $g(y) = y^2$, then

$$E(y^2) = (0(1/8) + 1(1/4) + 4(1/4) + 9(3/8)) = 4.625.$$

The variance of discrete random variable Y is

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

For the example,

$$\begin{aligned} Var(y) &= (-1.875)^2(1/8) + (-.875)^2(1/4) + (.125)^2(1/4) + (1.125)^2(3/8) \\ &= 4.625 - (1.875)^2 \\ &= 1.109375 \end{aligned}$$

1.2.1 Binomial Distribution

A common discrete distribution is the *binomial* distribution. A binomial event can take on only two possible outcomes, success or failure, zero or one, heads or tails, diseased or not diseased, and so on. The probability of one outcome is q and the probability of the other outcome is $1-q$. Trials, or a succession of binomial events, are assumed to be independent. The random variable Y is the number of successes. The probability distribution is given by

$$p(y) = \left(\begin{array}{c} n \\ y \end{array} \right) q^y (1-q)^{n-y},$$

for $y = 0,\ 1,\ 2,\ ..., n$ and $0 \leq\ q\ \leq 1$. The number of trials is n. The expected value and variance of the binomial distribution are

$$\begin{aligned} E(y) &= n\ q \\ Var(y) &= n\ q\ (1-q). \end{aligned}$$

1.2.2 Poisson Distribution

A Poisson probability distribution provides a good model for the probability distribution of the number Y of rare events that occur in a given space, time, volume, or any other dimension, and λ is the average value of Y. In dairy cattle breeding, for example, the number of quality embryos produced by a cow during superovulation can range from 0 to 20 (or more), but the average might be only 3 or 4. The Poisson probability distribution is given by

$$p(y) = \frac{\lambda^y}{y!} \exp -\lambda,$$

for $y = 0,\ 1,\ 2,\ \ldots$ and $\lambda > 0$. Also,

$$\begin{aligned} E(y) &= \lambda \\ Var(y) &= \lambda. \end{aligned}$$

The mean and the variance are equal.

1.3 General Matrix Results

Extending results from scalar random variables to vectors of random variables, also called a random vector variable, then the following general results apply. Vectors are denoted by boldfaced small letters.

1.3.1 Expectations

Let $\mathbf{y}_1$ and $\mathbf{y}_2$ be two random vector variables, then for $i = 1$ or 2, then

$$E(\mathbf{y}_i) = \boldsymbol{\mu}_i = \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{in} \end{pmatrix},$$

for a vector of length n. If c is a scalar constant, then

$$E(c\mathbf{y}_i) = c\boldsymbol{\mu}_i.$$

Similarly, if $\mathbf{C}$ is a matrix of constants, then

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

Finally,

$$\begin{aligned} E(\mathbf{y}_1 + \mathbf{y}_2) &= E(\mathbf{y}_1) + E(\mathbf{y}_2) \\ &= \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2. \end{aligned}$$

1.3.2 Variance-Covariance Matrices

Let $\mathbf{y}$ be a random vector variable of length n, then the *variance-covariance* matrix of $\mathbf{y}$ is as follows:

$$\begin{aligned} Var(\mathbf{y}) &= E(\mathbf{y}\mathbf{y}') - E(\mathbf{y})E(\mathbf{y}') \\ &= \begin{pmatrix} \sigma^2_{y_1} & \sigma_{y_1y_2} & \cdots & \sigma_{y_1y_n} \\ \sigma_{y_1y_2} & \sigma^2_{y_2} & \cdots & \sigma_{y_2y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{y_1y_n} & \sigma_{y_2y_n} & \cdots & \sigma^2_{y_n} \end{pmatrix} \\ &= \mathbf{V} \end{aligned}$$

A variance-covariance (VCV) matrix of a random vector contains variances on the diagonals and covariances in the off-diagonals. A VCV matrix is square, symmetric and should always be positive definite or positive semi-definite, i.e. all of the eigenvalues must be positive. Another name for VCV matrix is *dispersion* matrix or (co)variance matrix.

Let $\mathbf{C}$ be a matrix of constants conformable for multiplication with the vector $\mathbf{y}$, then

$$\begin{aligned} Var(\mathbf{C}\mathbf{y}) &= E(\mathbf{C}\mathbf{y}\mathbf{y}'\mathbf{C}') - E(\mathbf{C}\mathbf{y})E(\mathbf{y}'\mathbf{C}') \\ &= \mathbf{C}E(\mathbf{y}\mathbf{y}')\mathbf{C}' - \mathbf{C}E(\mathbf{y})E(\mathbf{y}')\mathbf{C}' \\ &= \mathbf{C}\left(E(\mathbf{y}\mathbf{y}') - E(\mathbf{y})E(\mathbf{y}')\right)\mathbf{C}' \\ &= \mathbf{C}Var(\mathbf{y})\mathbf{C}' = \mathbf{C}\mathbf{V}\mathbf{C}'. \end{aligned}$$

If there are two sets of functions of $\mathbf{y}$, say $\mathbf{C}_1\mathbf{y}$ and $\mathbf{C}_2\mathbf{y}$, then

$$Cov(\mathbf{C}_1\mathbf{y}, \mathbf{C}_2\mathbf{y}) = \mathbf{C}_1\mathbf{V}\mathbf{C}_2'.$$

Similarly, if $\mathbf{y}$ and $\mathbf{z}$ represent two different random vectors, possibly of different orders, and if the (co)variance matrix between these two vectors is $\mathbf{W}$, then

$$Cov(\mathbf{C}_1\mathbf{y}, \mathbf{C}_2\mathbf{z}) = \mathbf{C}_1\mathbf{W}\mathbf{C}_2'.$$

1.4 Continuous Distributions

Consider measuring the amount of milk given by a dairy cow at a particular milking. Even if a machine of perfect accuracy was used, the amount of milk would be a unique point on a continuum of possible values, such as 32.35769842.... kg of milk. As such it is mathematically impossible to assign a nonzero probability to all of the infinite possible points in the continuum. Thus, a different method of describing a probability distribution of a continuous random variable must be used. The sum of the probabilities (if they

could be assigned) through the continuum is still assumed to sum to 1. The *cumulative distribution function* of a random variable is

$$F(y) = P(Y \leq y),$$

for $-\infty < y < \infty$. As y approaches $-\infty$, then $F(y)$ approaches 0. As y approaches ∞, then $F(y)$ approaches 1. Thus, $F(y)$ is said to be a nondecreasing function of y. If $a < b$, then $F(a) < F(b)$.

If $F(y)$ is the cumulative distribution function of Y, then the *probability density function* of Y is given by

$$p(y) = \frac{\partial F(y)}{\partial y} = F'(y),$$

wherever the derivative exists. Always for $p(y)$ being a probability density function,

$$\int_{-\infty}^{\infty} p(y)\, \partial y = 1.$$

Conversely,

$$F(y) = \int_{-\infty}^{y} p(t)\, \partial t.$$

The expected value of a continuous random variable y is

$$E(y) = \int_{-\infty}^{\infty} y\, p(y)\, \partial y$$

provided that the integral exists. If $g(y)$ is a function of y, then

$$E(g(y)) = \int_{-\infty}^{\infty} g(y)\, p(y)\, \partial y$$

provided that the integral exists. Finally,

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

1.4.1 Uniform Distribution

The Uniform Distribution is one of the basic distributions in statistics. The primary application is in the generation of random numbers from other distributions.

$$\begin{aligned} y &\sim U(a, b) \\ p(y) &= U(y \mid a, b) \end{aligned}$$

where b is greater than a. In a uniform distribution, every value between a and b has an equal probability of existing, i.e. $p(y) = 1/(b - a)$. Usually $b = 1$ and $a = 0$ so that

the observed y are between 0 and 1, representing probabilities. The expected value and variance of this distribution are

$$\begin{aligned}
E(y) &= \int_a^b y\ p(y)\partial y \\
&= \int_a^b y/(b-a)\partial y \\
&= \frac{1}{2}y^2/(b-a)\ |_a^b \\
&= \frac{1}{2}b^2/(b-a)\ -\ \frac{1}{2}a^2/(b-a) \\
&= \frac{1}{2}(b^2-a^2)/(b-a) \\
&= (a+b)/2, \\
E(y^2) &= \int_a^b y^2\ p(y)\partial y \\
&= \int_a^b y^2/(b-a)\partial y \\
&= \frac{1}{3}y^3/(b-a)\ |_a^b \\
&= \frac{1}{3}(b^3-a^3)/(b-a) \\
&= \frac{1}{3}[(a+b)^2-ab].
\end{aligned}$$

Then

$$\begin{aligned}
Var(y) &= E(y^2)-E(y)^2 \\
&= (4[(a+b)^2-ab]-3(a+b)^2)/12 \\
&= (b-a)^2/12.
\end{aligned}$$

Uniform Random Number Generators

George Marsaglia of Florida State University developed a uniform random number generator that has passed 18 different tests, and which has a long cycle time (i.e. the number of calls to the subroutine before the sequence of random numbers begins to repeat itself). The strategy utilized is a multiply-with-carry scheme. To give an idea of how this scheme works, start with $x_{n-1} = 123456$, which is called the seed. The next number would be generated by

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

This would be followed by

$$x_n = 672 * 555 + 306 = 373266.$$

The number 672 is carefully chosen, and there is a process for doing this. The cycle time, or period, of the above generator is 336,000. Thus, after 336,000 operations, the numbers begin again in the same sequence. This is an example of how multiply-with-carry works.

Marsaglia's proposed generator makes use of 8 numbers at a time rather than just the previous random number, as well as two sequences of numbers rather than just one. The period for this generator is 2^{250} which is a very, very large number. The two sequences are

$$\begin{aligned} x_n &= 12013x_{n-8} + 1066x_{n-7} + 1215x_{n-6} + 1492x_{n-5} \\ &\quad +1776x_{n-4} + 1812x_{n-3} + 1860x_{n-2} + 1941x_{n-1} + \\ &\quad \text{carry mod } 2^{16}, \end{aligned}$$

and

$$\begin{aligned} x_n &= 9272x_{n-8} + 7777x_{n-7} + 6666x_{n-6} + 5555x_{n-5} \\ &\quad +4444x_{n-4} + 3333x_{n-3} + 2222x_{n-2} + 1111x_{n-1} + \\ &\quad \text{carry mod } 2^{16}. \end{aligned}$$

The coefficients were carefully chosen. Each sequence provides a 16 bit integer number, and therefore, combining the two (by concatenation) gives a 32-bit random integer.

Example Usage

The uniform distribution random number generator can be used to simulate a discrete random variable (such as that given earlier - Section 1.2). The steps of a simple program to assign y a value based upon the probabilities given earlier, would be as follows:

```
call uniform(p)
y = 0
if(p.gt.0.125)y=1
if(p.gt.0.375)y=2
if(p.gt.0.625)y=3
```

1.4.2 Normal Distribution

The normal distributiou has been the most commonly assumed distribution in animal breeding and statistical genetics. The properties of the distribution are well known; computations for estimators are relatively easy; and normal distributions suffice for a majority of situations.

A scalar random variable y has a normal probability distribution if and only if

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

for $-\infty < x < +\infty$, where σ^2 is the variance of y and μ is the expected value of y.

For a random vector variable, $\mathbf{y}$, the multivariate normal density function is

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

denoted as $\mathbf{y} \sim N(\boldsymbol{\mu}, \mathbf{V})$ where $\mathbf{V}$ is the variance-covariance matrix of $\mathbf{y}$. Note that the determinant of $\mathbf{V}$ must be positive, otherwise the density function is undefined.

Random Normal Deviates

There have been several ways to generate random variables from a normal distribution. Let r_1 and r_2 be values from a uniform distribution generator, between 0 and 1, then two random normal deviates can be computed as

$$y_1 = (-2\ln r_1)^{.5} \cos 2\pi r_2,$$

and

$$y_2 = (-2\ln r_1)^{.5} \sin 2\pi r_2.$$

The quality of the results depend on the quality of the uniform distribution generator.

Another algorithm again uses two values from a uniform distribution generator, say r_1 and r_2, then

$$\begin{aligned} v_1 &= r_1 + r_1 - 1.0, \quad \text{and} \\ v_2 &= r_2 + r_2 - 1.0, \\ q &= v_1 * v_1 + v_2 * v_2, \end{aligned}$$

if q is less than 1 and greater than 0, then

$$\begin{aligned} f &= (-2\ln q/q)^{.5} \\ y_1 &= v_1 * f, \\ y_2 &= v_2 * f. \end{aligned}$$

1.4.3 Beta Distribution

Beta distributions may be assumed for random variables such as heritability, probability, or gene frequency all of which are limited to a value between zero and 1.

$$p(y \mid a, b) = Cy^{a-1}(1-y)^{b-1},$$

for $y \in [0, 1]$, where the constant of integration is

$$C = \Gamma(a+b)/(\Gamma(a)\Gamma(b)),$$

and a and b are greater than 0. The gamma function, $\Gamma(\cdot)$ is

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} \exp^{-y} \partial y,$$

for $\alpha > 0$. When $\alpha = 1$, then $\Gamma(1) = 1$. For α being an integer and greater than 1, then

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) = (\alpha - 1)!$$

The expectation of a beta variable is

$$E(y \mid a, b) = C \int_0^1 y \left[y^{a-1}(1-y)^{b-1} \right] \partial y,$$

which simplifies to

$$E(y \mid a, b) = \frac{a}{a+b}.$$

The variance of a beta variable is

$$Var(y \mid a, b) = \frac{ab}{(a+b)^2(a+b+1)}.$$

1.4.4 Gamma Distribution

The Gamma Distribution has the following form;

$$p(y \mid a, b) = Cy^{a-1} \exp^{-by},$$

for y being greater than 0. Variance components are always supposed to be greater than 0, thus a gamma distribution may be appropriate. The constant of integration is

$$C = \frac{b^a}{\Gamma(a)},$$

and both a and b are greater than zero. The mean of the distribution is

$$E(y \mid a, b) = \frac{a}{b},$$

and the variance of the distribution is

$$Var(y \mid a, b) = \frac{a}{b^2}.$$

The coefficient of variation (standard deviation divided by the mean) is equal to $a^{-.5}$.

The gamma distribution is the "parent" distribution for two special cases. When $a = 1$, then the gamma distribution becomes the Exponential Distribution, i.e.,

$$p(y \mid b) = b \exp^{-by},$$

for y greater than zero. The other case is when $a = v/2$ and $b = 0.5$, where v is commonly known as the degrees of freedom and is greater than zero. Then

$$p(y \mid v) = Cy^{(v/2)-1} \exp^{-y/2},$$

and

$$C = \frac{(1/2)^{v/2}}{\Gamma(v/2)}.$$

This is the Central Chi-square distribution.

1.4.5 Chi-Square Distribution

In the estimation of variance components, quadratic forms of $\mathbf{y}$ are needed. If $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I})$, then $\mathbf{y}'\mathbf{y} \sim \chi^2_n$, where χ^2_n is a *central chi-square* distribution with n degrees of freedom and n is the length of the random vector variable $\mathbf{y}$. The mean of the central chi-square distribution is n, and the variance is $2n$. The probability distribution function of $s = \mathbf{y}'\mathbf{y}$ is

$$p(s \mid n) = (s)^{(n/2)-1} \exp -0.5s/[2^{0.5n}\Gamma(0.5n)],$$

for $s > 0$.

If $\mathbf{y} \sim N(\boldsymbol{\mu}, \mathbf{I})$, then $\mathbf{y}'\mathbf{y} \sim \chi^2_{n,\lambda}$ where λ is the noncentrality parameter which is equal to $.5\boldsymbol{\mu}'\boldsymbol{\mu}$. The mean of a noncentral chi-square distribution is $n + 2\lambda$ and the variance is $2n + 8\lambda$.

If $\mathbf{y} \sim N(\boldsymbol{\mu}, \mathbf{V})$, then $\mathbf{y}'\mathbf{Q}\mathbf{y}$ has a noncentral chi-square distribution only if $\mathbf{QV}$ is idempotent, i.e. $\mathbf{QVQV} = \mathbf{QV}$. The noncentrality parameter is $\lambda = .5\boldsymbol{\mu}'\mathbf{QVQ}\boldsymbol{\mu}$ and the mean and variance of the distribution are $tr(\mathbf{QV}) + 2\lambda$ and $2tr(\mathbf{QV}) + 8\lambda$, respectively.

If there are two quadratic forms of $\mathbf{y}$, say $\mathbf{y}'\mathbf{Q}\mathbf{y}$ and $\mathbf{y}'\mathbf{P}\mathbf{y}$, and both quadratic forms have chi-square distributions, then the two quadratic forms are independent if $\mathbf{QVP} = \mathbf{0}$. Independence of quadratic forms is necessary for the construction of valid tests of hypotheses. This property is not required for estimation of variances and covariances.

Random Chi-Square Generator

One way to generate a central Chi-square variate with n degrees of freedom is to generate a vector of length n of random normal deviates, then sum the squares of these deviates. Thus, a very good random normal deviate generator would be necessary. Also, the computing time to generate one Chi-square variate would depend on n, the number of random normal deviates to be generated.

In order to save time for large n, a Chi-square variate is generated using a random gamma distribution variate, which requires two uniform variates, r_1 and r_2, then

```
 1    call uniform(r1)
      call uniform(r2)
      V1 = r1 + r1 - 1
      V2 = r2 + r2 - 1
      f = V1*V1 + V2*V2
      if ( f.gt.1.0)go to 1
      if (V1.eq.0.0)then
         c=0.0
        else
         c=V2/V1
        endif
      anm = n/2 - 1
      s = dsqrt(anm+anm+1.0)
      y = s*c + anm
      if(y.le.0.0)go to 1
      temp = anm*dlog(y/anm)-s*c
      if(dabs(temp).gt.85)then
          go to 1
        else
          e = (1.0+c*c)*dexp(temp)
          call uniform(r3)
          if(r3.gt.e)go to 1
        endif
c   at this point  y  is the gamma variate
      return
```

The Chi-square variate with n degrees of freedom is then derived as

```
If(n is even) then
   call gamma(iseed,n,gam)
   chi = gam + gam
 else (n is odd)
   m = n-1
   call gamma(iseed,m,gam)
   chi = gam + gam
   call normal(znorm)
   chi = chi + znorm*znorm
 endif
```

Chi-square variates are needed in Gibbs sampling to obtain new sample values of a variance component.

1.4.6 The Wishart Distribution

The Wishart Distribution is akin to a multivariate Chi-square Distribution. An entire matrix is envisioned of which the diagonals have a Chi-square distribution, and the off-diagonals have a built-in correlation structure. The resulting matrix is positive definite.

1. To generate a matrix having a Wishart distribution, let $\mathbf{U}$ equal a matrix with q rows and m columns (and $q > m$), then $\mathbf{W} = \mathbf{U}'\mathbf{U}$ is an m by m positive definite matrix.

2. Perform a Cholesky decomposition of $\mathbf{W}$, so that

$$\mathbf{W} = \mathbf{T}\mathbf{T}',$$

   and $\mathbf{T}$ is a lower triangular matrix.

3. Fill an m by m matrix, $\mathbf{Z}$, with random normal deviates, and fill a vector, $\mathbf{V}$, of length m with the square roots of random Chi-square variates, such that the first element has $q - 1$ degrees of freedom, the second has $q - 2$ df, and so on, and the last element has $q - m$ df.

4. Now form a matrix $\mathbf{B}$ such that the diagonals are

$$B_{i,i} = v_i * v_i,$$

   for $i = 1$ and for $i > 1$

$$B_{i,i} = v_i * v_i + \sum_{j=1}^{i-1} z_{i,j} * z_{i,j}.$$

   The offdiagonals of $\mathbf{B}$ are

$$B_{i,j} = B_{j,i} = z_{i,j} * v_i,$$

   for $j = 2$ to m and $i = 1$, and for $i > 1$

$$B_{i,j} = B_{j,i} = z_{i,j} * v_i + \sum_{k=1}^{i-1} z_{k,i} * z_{k,j},$$

   for $j = 2$ to m and $i = 2, (j - 1)$.

5. Finally, calculate $\mathbf{S} = \mathbf{T}\mathbf{B}\mathbf{T}'$.

The matrix $\mathbf{S}$ is a random Wishart matrix based upon the relationships in $\mathbf{W}$. Note that the off-diagonals in $\mathbf{B}$ should be close to zero, on average.

1.4.7 The t-Distribution

The t-distribution is based on the ratio of two independent random variables. The first random variable follows a univariate normal distribution, and the second random variable follows a central chi-square distribution. Let $y \sim N(0,1)$ and $s \sim \chi^2_n$ with y and s being independent, then

$$\frac{y}{(s/n)^{.5}} \sim t_n.$$

The mean of a t-distribution is the mean of the y variable, and the variance is $n/(n-2)$, and n is the degrees of freedom of the distribution. As n becomes larger, the t-distribution becomes very similar to the normal distribution in shape. A t-distribution with small degrees of freedom would allow more observations to occur in the tails of the distribution, and therefore would look like a squashed normal distribution.

1.4.8 The F-distribution

A common distribution used in the testing of hypotheses is the F-distribution. A central F-distribution is based on the ratio of two independent central chi-square variables. Let $s \sim \chi^2_n$ and $v \sim \chi^2_m$ with s and v being independent, then

$$\frac{(s/n)}{(v/m)} \sim F_{n,m}.$$

The mean of the F-distribution is $m/(m-2)$ and the variance is

$$\frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}.$$

Tables of F-values have been constructed for various probability levels as criteria to test if the numerator chi-square variable has a noncentral chi-square distribution. If the calculated F-value is greater than the value in the tables, then s is implied to have a noncentral chi-square distribution, otherwise s has a central chi-square distribution. The square of a t-distribution variable gives a variable that has an F-distribution with 1 and n degrees of freedom.

Noncentral F-distributions exist depending on whether the numerator or denominator variables have noncentral chi-square distributions. Tables for noncentral F-distributions generally do not exist because of the difficulty in predicting the noncentrality parameters. However, using random chi-square generators it is possible to numerically calculate an expected noncentral F value for specific situations. When both the numerator and denominator chi-square variables are from noncentral distributions, then their ratio follows a doubly noncentral F-distribution.

1.5 Quadratic Forms

A quadratic form is a weighted sum of squares of elements of a random vector variable. The general form is $\mathbf{y}'\mathbf{Q}\mathbf{y}$, where $\mathbf{y}$ is a random vector variable, and $\mathbf{Q}$ is a regulator matrix. The regulator matrix can take on various forms and values depending on the situation. Usually $\mathbf{Q}$ is a symmetric matrix, but not necessarily positive definite. Examples of different $\mathbf{Q}$ matrices are as follows:

1. $\mathbf{Q} = \mathbf{I}$, then $\mathbf{y}'\mathbf{Q}\mathbf{y} = \mathbf{y}'\mathbf{y}$ which is a total sum of squares of the elements in $\mathbf{y}$.
2. $\mathbf{Q} = \mathbf{J}(1/n)$, then $\mathbf{y}'\mathbf{Q}\mathbf{y} = \mathbf{y}'\mathbf{J}\mathbf{y}(1/n)$ where n is the length of $\mathbf{y}$. Note that $\mathbf{J} = \mathbf{1}\mathbf{1}'$, so that $\mathbf{y}'\mathbf{J}\mathbf{y} = (\mathbf{y}'\mathbf{1})(\mathbf{1}'\mathbf{y})$ and $(\mathbf{1}'\mathbf{y})$ is the sum of the elements in $\mathbf{y}$.
3. $\mathbf{Q} = (\mathbf{I} - \mathbf{J}(1/n))/(n-1)$, then $\mathbf{y}'\mathbf{Q}\mathbf{y}$ gives the variance of the elements in $\mathbf{y}$, σ_y^2.

The expected value of a quadratic form is

$$E(\mathbf{y}'\mathbf{Q}\mathbf{y}) = E(tr(\mathbf{y}'\mathbf{Q}\mathbf{y})) = E(tr(\mathbf{Q}\mathbf{y}\mathbf{y}')) = tr(\mathbf{Q}E(\mathbf{y}\mathbf{y}')).$$

However,

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

so that

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

then

$$E(\mathbf{y}'\mathbf{Q}\mathbf{y}) = tr(\mathbf{Q}(Var(\mathbf{y}) + E(\mathbf{y})E(\mathbf{y}'))).$$

If we let $Var(\mathbf{y}) = \mathbf{V}$ and $E(\mathbf{y}) = \mu$, then

$$\begin{aligned} E(\mathbf{y}'\mathbf{Q}\mathbf{y}) &= tr(\mathbf{Q}(\mathbf{V} + \mu\mu')) \\ &= tr(\mathbf{Q}\mathbf{V}) + tr(\mathbf{Q}\mu\mu') \\ &= tr(\mathbf{Q}\mathbf{V}) + \mu'\mathbf{Q}\mu. \end{aligned}$$

The expectation of a quadratic form does not depend on the distribution of $\mathbf{y}$. However, the variance of a quadratic form requires that $\mathbf{y}$ follows a multivariate normal distribution. Without showing the derivation, the variance of a quadratic form, assuming $\mathbf{y}$ has a multivariate normal distribution, is

$$Var(\mathbf{y}'\mathbf{Q}\mathbf{y}) = 2tr(\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{V}) + 4\mu'\mathbf{Q}\mathbf{V}\mathbf{Q}\mu.$$

The quadratic form, $\mathbf{y}'\mathbf{Q}\mathbf{y}$, has a chi-square distribution if

$$tr(\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{V}) = tr(\mathbf{Q}\mathbf{V}), \text{ and } \mu'\mathbf{Q}\mathbf{V}\mathbf{Q}\mu = \mu'\mathbf{Q}\mu,$$

or the single condition that $\mathbf{Q}\mathbf{V}$ is idempotent. Then if

$$m = tr(\mathbf{Q}\mathbf{V}) \text{ and } \lambda = .5\mu'\mathbf{Q}\mu,$$

the expected value of $\mathbf{y}'\mathbf{Q}\mathbf{y}$ is $m + 2\lambda$ and the variance is $2m + 8\lambda$, which are the usual results for a noncentral chi-square variable.

The covariance between two quadratic forms, say $\mathbf{y}'\mathbf{Q}\mathbf{y}$ and $\mathbf{y}'\mathbf{P}\mathbf{y}$, is

$$Cov(\mathbf{y}'\mathbf{Q}\mathbf{y}, \mathbf{y}'\mathbf{P}\mathbf{y}) = 2tr(\mathbf{Q}\mathbf{V}\mathbf{P}\mathbf{V}) + 4\boldsymbol{\mu}'\mathbf{Q}\mathbf{V}\mathbf{P}\boldsymbol{\mu}.$$

The covariance is zero if $\mathbf{Q}\mathbf{V}\mathbf{P} = \mathbf{0}$, then the two quadratic forms are said to be independent.

Chapter 2

Building Blocks

2.1 Basic Blocks

In order to derive Maximum Likelihood and Residual Maximum Likelihood, a number of results about derivatives of determinants and other quantities needs to be reviewed. These will be called **Building Blocks** or **BB** for short, because they are not theorems or conjectures.

BB-1. The (co)variance matrix of $\mathbf{y}$ is

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

Usually, each $\mathbf{G}_i$ is assumed to be $\mathbf{I}$ for most random factors, but for animal models $\mathbf{G}_i$ might be equal to $\mathbf{A}$, the additive genetic relationship matrix. Thus, $\mathbf{G}_i$ does not always have to be diagonal, and will not be an identity in animal model analyses.

BB-2. The inverse of $\mathbf{V}$ is

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

To prove, show that $\mathbf{V}\mathbf{V}^{-1} = \mathbf{I}$. Let $\mathbf{T} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}$, then

$$\begin{aligned} \mathbf{V}\mathbf{V}^{-1} &= (\mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R})[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}\mathbf{T}^{-1}\mathbf{Z}'\mathbf{R}^{-1}] \\ &= \mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{R}^{-1} - \mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{T}^{-1}\mathbf{Z}'\mathbf{R}^{-1} \\ &\quad + \mathbf{I} - \mathbf{Z}\mathbf{T}^{-1}\mathbf{Z}'\mathbf{R}^{-1} \end{aligned}$$

$$\begin{aligned}
&= \mathbf{I} + [\mathbf{ZGT} - \mathbf{ZGZ'R^{-1}Z} - \mathbf{Z}](\mathbf{T^{-1}Z'R^{-1}}) \\
&= \mathbf{I} + [\mathbf{ZG}(\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}}) - \mathbf{ZGZ'R^{-1}Z} - \mathbf{Z}](\mathbf{T^{-1}Z'R^{-1}}) \\
&= \mathbf{I} + [\mathbf{ZGZ'R^{-1}Z} + \mathbf{Z} - \mathbf{ZGZ'R^{-1}Z} - \mathbf{Z}](\mathbf{T^{-1}Z'R^{-1}}) \\
&= \mathbf{I} + [\mathbf{0}](\mathbf{T^{-1}Z'R^{-1}}) \\
&= \mathbf{I}.
\end{aligned}$$

BB-3. If k is a scalar constant and $\mathbf{A}$ is any square matrix of order m, then

$$|\, \mathbf{A}k \,| = k^m \,|\, \mathbf{A} \,|.$$

BB-4. For general square matrices, say $\mathbf{M}$ and $\mathbf{U}$, of the same order then

$$|\, \mathbf{MU} \,| = |\, \mathbf{M} \,|\, |\, \mathbf{U} \,|.$$

BB-5. For the general matrix below with $\mathbf{A}$ and $\mathbf{D}$ being square and non-singular (i.e. the inverse of each exists), then

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

Then if $\mathbf{A} = \mathbf{I}$ and $\mathbf{D} = \mathbf{I}$, then $|\, \mathbf{I} \,| = 1$, so that

$$\begin{aligned}
|\, \mathbf{I} + \mathbf{QB} \,| &= |\, \mathbf{I} + \mathbf{BQ} \,| \\
&= |\, \mathbf{I} + \mathbf{B'Q'} \,| \\
&= |\, \mathbf{I} + \mathbf{Q'B'} \,|.
\end{aligned}$$

BB-6. Using the results in (**BB-4**) and (**BB-5**), then

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

BB-7. The mixed model coefficient matrix of Henderson can be denoted by

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

then the determinant of $\mathbf{C}$ can be derived as

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

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

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

BB-8. A projection matrix, $\mathbf{P}$, is defined as

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

Properties of $\mathbf{P}$:

$$\begin{aligned} \mathbf{P}\mathbf{X} &= \mathbf{0}, \\ \mathbf{P}\mathbf{y} &= \mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\mathbf{b}}), \quad \text{where} \\ \hat{\mathbf{b}} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \end{aligned}$$

Therefore,

$$\mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{G}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y} = (\mathbf{y}-\mathbf{X}\hat{\mathbf{b}})'\mathbf{V}^{-1}\mathbf{Z}_i\mathbf{G}_i\mathbf{Z}_i'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\mathbf{b}}).$$

BB-9. Derivative of $\mathbf{V}^{-1}$ is

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

BB-10. Derivative of $\ln |\,\mathbf{V}\,|$ is

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

BB-11. Derivative of $\mathbf{P}$ is

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

BB-12. Derivative of $\mathbf{V}$ is

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

BB-13. Derivative of $\ln \mid \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \mid$ is

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

2.2 Basic Model

The following simple model will be assumed. The general linear model is described as

$$\begin{aligned} \mathbf{y} &= \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}, \\ \text{where } E(\mathbf{y}) &= \mathbf{Xb}, \\ E(\mathbf{u}) &= \mathbf{0}, \\ \text{and } E(\mathbf{e}) &= \mathbf{0}. \end{aligned}$$

Often $\mathbf{u}$ is partitioned into s factors as

$$\mathbf{u}' = (\mathbf{u}_1' \;\; \mathbf{u}_2' \;\; \ldots \;\; \mathbf{u}_s').$$

The (co)variance matrices are defined as

$$\mathbf{G} = Var(\mathbf{u}) = Var\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_s \end{pmatrix} = \begin{pmatrix} \mathbf{G}_1\sigma_1^2 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2\sigma_2^2 & \ldots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{G}_s\sigma_s^2 \end{pmatrix}$$

and

$$\mathbf{R} = Var(\mathbf{e}) = \mathbf{I}\sigma_0^2.$$

Then

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

and if $\mathbf{Z}$ is partitioned corresponding to $\mathbf{u}$, as

$$\begin{aligned} \mathbf{Z} &= [\ \mathbf{Z}_1\ \ \mathbf{Z}_2\ \ldots\ \mathbf{Z}_s\ ], \text{ then} \\ \mathbf{ZGZ}' &= \sum_{i=1}^{s} \mathbf{Z}_i\mathbf{G}_i\mathbf{Z}_i'\sigma_i^2. \\ \text{Let } \mathbf{V}_i &= \mathbf{Z}_i\mathbf{G}_i\mathbf{Z}_i' \text{ and} \\ \mathbf{V}_0 &= \mathbf{I}, \text{ then} \\ \mathbf{V} &= \sum_{i=o}^{s} \mathbf{V}_i\sigma_i^2. \end{aligned}$$

Covariances between random factors in the model will be allowed later, such as in a maternal effects model where there are covariances between direct and maternal effects, or random regression models with covariances between the random regression coefficients, or multiple trait models.

2.3 Mixed Model Equations

Henderson's mixed model equations (MME) are written as

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_2 & \ldots & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_s \\ \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_1 + \mathbf{G}_1^{-1}\sigma_1^{-2} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_2 & \ldots & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_s \\ \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_2 + \mathbf{G}_2^{-1}\sigma_2^{-2} & \ldots & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_s \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{Z}_s'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_s'\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{Z}_s'\mathbf{R}^{-1}\mathbf{Z}_2 & \ldots & \mathbf{Z}_s'\mathbf{R}^{-1}\mathbf{Z}_s + \mathbf{G}_s^{-1}\sigma_s^{-2} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \\ \vdots \\ \hat{\mathbf{u}}_s \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{y} \\ \vdots \\ \mathbf{Z}_s'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

Quadratic forms from solutions to these equations are

$$\hat{\mathbf{u}}_i'\mathbf{G}_i\hat{\mathbf{u}}_i,$$

where

$$\hat{\mathbf{u}} = \mathbf{G}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y},$$

and for the residual variance,

$$\hat{\mathbf{e}}'\hat{\mathbf{e}},$$

where

$$\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}\hat{\mathbf{u}}.$$

2.4 Unbiased Estimation of Variances

Unbiased estimation is no longer practiced in animal breeding, but the history of the development from unbiased to likelihood based methods is important to understand. The best way to describe unbiased methods of estimation is to give a small example with only three observations. Assume that all $\mathbf{G}_i$ are equal to $\mathbf{I}$ for this example, so that $\mathbf{Z}_i\mathbf{G}_i\mathbf{Z}_i'$ simplifies to $\mathbf{Z}_i\mathbf{Z}_i'$. Let

$$\begin{aligned}\mathbf{X} &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \ \mathbf{Z}_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \\ \mathbf{Z}_2 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \text{ and } \mathbf{y} = \begin{pmatrix} 29 \\ 53 \\ 44 \end{pmatrix},\end{aligned}$$

Then

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

and

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

and $\mathbf{V}_0 = \mathbf{I}$.

In this example, there are 3 unknown variances to be estimated, and consequently, at least three quadratic forms are needed in order to estimate the variances. The $\mathbf{Q}$-matrices are the 'weights' of the observations in the quadratic forms. These matrices differ depending on the method of estimation that is chosen. Below are three arbitrary $\mathbf{Q}$-matrices that were chosen such that $\mathbf{Q}_k\mathbf{X} = \mathbf{0}$. They do not necessarily correspond to any known method of estimation, but are for illustration of the calculations. Let

$$\begin{aligned}\mathbf{Q}_1 &= \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}, \\ \mathbf{Q}_2 &= \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 2 \end{pmatrix}, \\ \text{and } \mathbf{Q}_3 &= \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.\end{aligned}$$

The numeric values of the quadratic forms are

$$\mathbf{y}'\mathbf{Q}_1\mathbf{y} = 657,$$

$$\begin{aligned} \mathbf{y}'\mathbf{Q}_2\mathbf{y} &= 306, \\ \text{and } \mathbf{y}'\mathbf{Q}_3\mathbf{y} &= 882. \end{aligned}$$

For example,

$$\mathbf{y}'\mathbf{Q}_1\mathbf{y} = \begin{pmatrix} 29 & 53 & 44 \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 29 \\ 53 \\ 44 \end{pmatrix} = 657.$$

The expectations of the quadratic forms are

$$\begin{aligned} E(\mathbf{y}'\mathbf{Q}_1\mathbf{y}) &= tr\mathbf{Q}_1\mathbf{V}_0\sigma_0^2 + tr\mathbf{Q}_1\mathbf{V}_1\sigma_1^2 + tr\mathbf{Q}_1\mathbf{V}_2\sigma_2^2 \\ &= 4\sigma_0^2 + 2\sigma_1^2 + 2\sigma_2^2 \\ E(\mathbf{y}'\mathbf{Q}_2\mathbf{y}) &= 4\sigma_0^2 + 4\sigma_1^2 + 2\sigma_2^2, \\ E(\mathbf{y}'\mathbf{Q}_3\mathbf{y}) &= 6\sigma_0^2 + 4\sigma_1^2 + 4\sigma_2^2. \end{aligned}$$

Now equate the values of the quadratic forms to their corresponding expectations, which gives a system of equations to be solved, such as $\mathbf{F}\sigma = \mathbf{w}$. In this case, the equations would be

$$\begin{pmatrix} 4 & 2 & 2 \\ 4 & 4 & 2 \\ 6 & 4 & 4 \end{pmatrix} \begin{pmatrix} \sigma_0^2 \\ \sigma_1^2 \\ \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 657. \\ 306. \\ 882. \end{pmatrix},$$

which gives the solution as $\hat{\sigma} = \mathbf{F}^{-1}\mathbf{w}$, or

$$\begin{pmatrix} \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \end{pmatrix} = \begin{pmatrix} 216.0 \\ -175.5 \\ 72.0 \end{pmatrix}.$$

Note that one of the estimates is negative, which is not appropriate for a variance component. The estimate is said to be out of the parameter space.

2.5 Variances of Quadratic Forms

The variance of a quadratic form is given by

$$Var(\mathbf{y}'\mathbf{Q}\mathbf{y}) = 2tr\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{V} + 4\mathbf{b}'\mathbf{X}'\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{X}\mathbf{b}.$$

Only translation invariant quadratic forms are typically considered in variance component estimation, that means $\mathbf{b}'\mathbf{X}'\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{X}\mathbf{b} = 0$. Thus, only $2tr\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{V}$ needs to be calculated. Remember that $\mathbf{V}$ can be written as the sum of $s+1$ matrices, $\mathbf{V}_i\sigma_i^2$, then

$$\begin{aligned} tr\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{V} &= tr\mathbf{Q}\sum_{i=o}^{s}\mathbf{V}_i\sigma_i^2\ \mathbf{Q}\sum_{j=o}^{s}\mathbf{V}_j\sigma_j^2 \\ &= \sum_{i=o}^{s}\sum_{j=o}^{s} tr\ \mathbf{Q}\mathbf{V}_i\mathbf{Q}\mathbf{V}_j\ \sigma_i^2\ \sigma_j^2 \end{aligned}$$

For example, if $s = 2$, then

$$\begin{aligned} tr\mathbf{QVQV} &= tr\mathbf{QV}_0\mathbf{QV}_0\sigma_0^4 + 2tr\mathbf{QV}_0\mathbf{QV}_1\sigma_0^2\sigma_1^2 \\ &\quad + 2tr\mathbf{QV}_0\mathbf{QV}_2\sigma_0^2\sigma_2^2 + tr\mathbf{QV}_1\mathbf{QV}_1\sigma_1^4 \\ &\quad + 2tr\mathbf{QV}_1\mathbf{QV}_2\sigma_1^2\sigma_2^2 + tr\mathbf{QV}_2\mathbf{QV}_2\sigma_2^4. \end{aligned}$$

The exact sampling variances require the true, unknown components of variance. The magnitude of the sampling variances depends on

1. The true magnitude of the individual components,
2. The matrix $\mathbf{Q}$ which depends on the method of estimation and the model, and
3. The structure and amount of the data through $\mathbf{X}$ and $\mathbf{Z}$.

Normally, the variance-covariance matrix of the estimates, commonly known as the sampling variances of the estimates, were never actually computed during the days of unbiased methods due to their computational complexity. However, with today's computers their calculation can still be very challenging and usually impossible. For small examples, the calculations can be easily demonstrated. In this case,

$$Var(\mathbf{F}^{-1}\mathbf{w}) = \mathbf{F}^{-1}Var(\mathbf{w})\mathbf{F}^{-1'},$$

a function of the variance-covariance matrix of the quadratic forms.

Using the small example of the previous section, the $Var(\mathbf{w})$ is a 3x3 matrix. The (1,1) element is the variance of $\mathbf{y}'\mathbf{Q}_1\mathbf{y}$ which is

$$\begin{aligned} Var(\mathbf{y}'\mathbf{Q}_1\mathbf{y}) &= 2tr\mathbf{Q}_1\mathbf{V}\mathbf{Q}_1\mathbf{V} \\ &= 2tr\mathbf{Q}_1\mathbf{V}_0\mathbf{Q}_1\mathbf{V}_0\sigma_0^4 + 4tr\mathbf{Q}_1\mathbf{V}_0\mathbf{Q}_1\mathbf{V}_1\sigma_0^2\sigma_1^2 \\ &\quad +4tr\mathbf{Q}_1\mathbf{V}_0\mathbf{Q}_1\mathbf{V}_2\sigma_0^2\sigma_2^2 + 2tr\mathbf{Q}_1\mathbf{V}_1\mathbf{Q}_1\mathbf{V}_1\sigma_1^4 \\ &\quad +4tr\mathbf{Q}_1\mathbf{V}_1\mathbf{Q}_1\mathbf{V}_2\sigma_1^2\sigma_2^2 + 2tr\mathbf{Q}_1\mathbf{V}_2\mathbf{Q}_1\mathbf{V}_2\sigma_2^4 \\ &= 20\sigma_0^4 + 16\sigma_0^2\sigma_1^2 + 16\sigma_0^2\sigma_2^2 + 8\sigma_1^4 + 0\sigma_1^2\sigma_2^2 + 8\sigma_2^4 \end{aligned}$$

The (1,2) element is the covariance between the first and second quadratic forms,

$$Cov(\mathbf{y}'\mathbf{Q}_1\mathbf{y}, \mathbf{y}'\mathbf{Q}_2\mathbf{y}) = 2tr\mathbf{Q}_1\mathbf{V}\mathbf{Q}_2\mathbf{V},$$

and similarly for the other terms. All of the results are summarized in the table below.

Forms	σ_0^4	$\sigma_0^2\sigma_1^2$	$\sigma_0^2\sigma_2^2$	σ_1^4	$\sigma_1^2\sigma_2^2$	σ_2^4
$Var(w_1)$	20	16	16	8	0	8
$Cov(w_1, w_2)$	14	24	8	16	0	8
$Cov(w_1, w_3)$	24	24	24	16	0	16
$Var(w_2)$	20	48	16	32	16	8
$Cov(w_2, w_3)$	24	48	24	32	16	16
$Var(w_3)$	36	48	48	32	16	32

To get numeric values for these variances, the true components need to be known. Assume that the true values are $\sigma_0^2 = 250$, $\sigma_1^2 = 10$, and $\sigma_2^2 = 80$, then the variance of w_1 is

$$\begin{aligned} Var(w_1) &= 20(250)^2 + 16(250)(10) + 16(250)(80) \\ &\quad +8(10)^2 + 0(10)(80) + 8(80)^2 \\ &= 1,662,000. \end{aligned}$$

The complete variance- covariance matrix of the quadratic forms is

$$Var\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 1,662,000 & 1,147,800 & 2,144,000 \\ 1,147,800 & 1,757,200 & 2,218,400 \\ 2,144,000 & 2,218,400 & 3,550,800 \end{pmatrix}.$$

The variance-covariance matrix of the estimated variances (assuming the above true values) would be

$$\begin{aligned} Var(\hat{\sigma}) &= \mathbf{F}^{-1}Var(\mathbf{w})\mathbf{F}^{-1'} \\ &= \begin{pmatrix} 405,700 & -275,700 & -240,700 \\ -275,700 & 280,900 & 141,950 \\ -240,700 & 141,950 & 293,500 \end{pmatrix} = \mathbf{C}. \end{aligned}$$

2.6 Variance of A Ratio of Variance Estimates

Often estimates of ratios of functions of the variances are needed for animal breeding work, such as heritabilities, repeatabilities, and variance ratios. Let such a ratio be denoted as a/c where

$$a = \hat{\sigma_2^2} = (0 \;\; 0 \;\; 1)\hat{\sigma} = 72.$$

and

$$c = \hat{\sigma_0^2} + \hat{\sigma_1^2} + \hat{\sigma_2^2} = (1 \;\; 1 \;\; 1)\hat{\sigma} = 288.$$

(NOTE: the negative estimate for $\hat{\sigma}_1^2$ was set to zero before calculating c.

From Osborne and Patterson (1952) and Rao (1968) an approximation to the variance of a ratio is given by

$$Var(a/c) = (c^2Var(a) + a^2Var(c) - 2ac\ Cov(a,c))/c^4.$$

Now note that

$$\begin{aligned} Var(a) &= (0 \;\; 0 \;\; 1)\mathbf{C}(0 \;\; 0 \;\; 1)' = 293,500, \\ Var(c) &= (1 \;\; 1 \;\; 1)\mathbf{C}(1 \;\; 1 \;\; 1)' = 231,200, \\ Cov(a,c,) &= (0 \;\; 0 \;\; 1)\mathbf{C}(1 \;\; 1 \;\; 1)' = 194,750. \end{aligned}$$

Then

$$\begin{aligned} Var(a/c) &= [(288)^2(293,500) + (72)^2(231,200) \\ &\quad -2(72)(288)(194,750)]/(288)^4 \\ &= 2.53876 \end{aligned}$$

This result is very large, but could be expected from only 3 observations. Thus, $(a/c) = .25$ with a standard deviation of 1.5933.

Another approximation method assumes that the denominator has been estimated accurately, so that it is considered to be a constant, such as the estimate of σ_e^2. Then,

$$Var(a/c) \cong Var(a)/c^2.$$

For the example problem, this gives

$$Var(a/c) \cong 293,500/(288)^2 = 3.53853,$$

which is slightly larger than the previous approximation. The second approximation would not be suitable for a ratio of the residual variance to the variance of one of the other components. Suppose $a = \hat{\sigma}_0^2 = 216$, and $c = \hat{\sigma}_2^2 = 72$, then $(a/c) = 3.0$, and

$$\begin{aligned} Var(a/c) &= [(72)^2(405,700) + (216)^2(293,500) \\ &\quad -2(72)(216)(-240,700)]/(72)^4 \\ &= 866.3966, \end{aligned}$$

with the first method, and

$$Var(a/c) = 405,700/(72)^2 = 78.26,$$

with the second method. The first method is probably more realistic in this situation, but both are very large.

Chapter 3

Likelihood Methods

3.1 The Likelihood Function

The normal distribution likelihood function is commonly assumed in animal breeding, and is used in REML and related methods of estimation. Use will be made of the building blocks of the previous chapter. 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, say 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 that does not involve any of the unknown variances or effects in the model, and therefore, it is commonly omitted during maximization computations. Maximizing the log likelihood maximizes the original likelihood function.

In the previous chapter,

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

and therefore,

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

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

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

Except, that 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

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

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

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

3.2 Maximum Likelihood

Hartley and Rao (1967) described the maximum likelihood approach for the estimation of variance components. Let L_2 be equivalent to L_1 except for the constant involving π.

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

The derivatives of L_2 with respect to $\mathbf{b}$ and to σ_i^2 for $i = 0, 1, \ldots s$ are

$$\frac{\partial L_2}{\partial \mathbf{b}} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{b} - \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

and

$$\begin{aligned} \frac{\partial L_2}{\partial \sigma_i^2} &= -.5 \; tr[\mathbf{V}^{-1}(\partial \mathbf{V}/\partial \sigma_i^2)] \\ &\quad + .5(\mathbf{y} - \mathbf{X}\mathbf{b})'\mathbf{V}^{-1}(\partial \mathbf{V}/\partial \sigma_i^2)\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b}) \\ &= -.5 \; tr[\mathbf{V}^{-1}\mathbf{V}_i] + .5(\mathbf{y} - \mathbf{X}\mathbf{b})'\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b}) \end{aligned}$$

Equating the derivatives to zero gives

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

and

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

Recall that

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

where $\mathbf{P}$ is the projection matrix, and that $\mathbf{V}_i = \mathbf{Z}_i\mathbf{Z}_i'$, then

$$tr[\mathbf{V}^{-1}\mathbf{Z}_i\mathbf{Z}_i'] = \mathbf{y}'\mathbf{P}\mathbf{V}_i\mathbf{P}\mathbf{y}.$$

In usual mixed model theory, the solution vector for a random factor may be written as

$$\hat{\mathbf{u}}_i = \mathbf{G}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y},$$

so that

$$\begin{aligned}\mathbf{y}'\mathbf{P}\mathbf{V}_i\mathbf{P}\mathbf{y} &= \mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{G}_i\mathbf{G}_i^{-2}\mathbf{G}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y} \\ &= \hat{\mathbf{u}}_i'\mathbf{G}_i^{-2}\hat{\mathbf{u}}_i \\ &= \hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i/\sigma_i^4.\end{aligned}$$

Also,

$$tr[\mathbf{V}^{-1}\mathbf{V}_i] = tr[(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1})\mathbf{Z}_i\mathbf{Z}_i'].$$

Let

$$\begin{aligned}\mathbf{T} &= (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}, \\ \mathbf{R} &= \mathbf{I}\sigma_0^2, \\ \text{and } \mathbf{G} &= \overset{+}{\sum}\mathbf{I}\sigma_i^2,\end{aligned}$$

then

$$tr[\mathbf{V}^{-1}\mathbf{V}_i] = tr(\mathbf{Z}_i'\mathbf{Z}_i)\sigma_0^{-2} - tr(\mathbf{Z}_i'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{Z}_i)\sigma_0^{-4}.$$

If $\mathbf{T}$ can be partitioned into submatrices for each random factor, then

$$\mathbf{T}\sigma_0^{-2}(\mathbf{Z}'\mathbf{Z} + \overset{+}{\sum}\mathbf{I}\alpha_i) = \mathbf{I},$$

and

$$\begin{aligned}\mathbf{T}\mathbf{Z}'\mathbf{Z}\sigma_0^{-2} &= \mathbf{I} - \mathbf{T}(\overset{+}{\sum}\mathbf{I}\sigma_i^{-2}), \\ \mathbf{T}\mathbf{Z}'\mathbf{Z}_i\sigma_0^{-2} &= \mathbf{I} - \mathbf{T}_{ii}\sigma_i^{-2},\end{aligned}$$

which yields

$$tr(\mathbf{Z}_i'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{Z}_i)\sigma_0^{-4} = tr(\mathbf{Z}_i'\mathbf{Z}_i)\sigma_0^{-2} - tr(\mathbf{I} - \mathbf{T}_{ii}\sigma_i^{-2})\sigma_i^{-2}.$$

Finally,

$$\begin{aligned}tr[\mathbf{V}^{-1}\mathbf{V}_i] &= tr(\mathbf{I} - \mathbf{T}_{ii}\sigma_i^{-2})\sigma_i^{-2} \\ &= tr\mathbf{I}\sigma_i^{-2} - tr\mathbf{T}_{ii}\sigma_i^{-4} \\ &= q_i\sigma_i^{-2} - tr\mathbf{T}_{ii}\sigma_i^{-4}.\end{aligned}$$

Combining results gives

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

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

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

3.2.1 The EM Algorithm

EM stands for Expectation Maximization. From Searle, Casella, and McCulloch (1992) the following explanation is given. The procedure alternates between calculating conditional expected values and maximizing simplified likelihoods. The actual data $\mathbf{y}$ are called the incomplete data in the EM algorithm, and the complete data are considered to be $\mathbf{y}$ and the unobservable random effects, $\mathbf{u}_i$. If the realized values of the unobservable random effects were known, then their variance would be the average of their squared values, i.e.,

$$\hat{\sigma}_i^2 = \mathbf{u}_i'\mathbf{u}_i/q_i.$$

However, in real life the realized values of the random effects are unknown.

The steps of the EM algorithm are as follows:

Step 0. Decide on starting values for the variances and set $m = 0$.

Step 1.(E-step) Calculate the conditional expectation of the sufficient statistics, conditional on the incomplete data.

$$\begin{aligned} E(\mathbf{u}_i'\mathbf{u}_i \mid \mathbf{y}) &= \sigma_i^{4(m)}\mathbf{y}'\mathbf{P}^{(m)}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}^{(m)}\mathbf{y} \\ &\quad + tr(\sigma_i^{2(m)}\mathbf{I} - \sigma_i^{4(m)}\mathbf{Z}_i'(\mathbf{V}^{(m)})^{-1}\mathbf{Z}_i) \\ &= \hat{t}_i^{(m)} \end{aligned}$$

Step 2.(M-step) Maximize the likelihood of the complete data,

$$\sigma_i^{2(m+1)} = \hat{t}_i^{(m)}/q_i, \quad i = 0, 1, 2, \ldots, s.$$

Step 3. If convergence is reached, set $\hat{\boldsymbol{\sigma}} = \boldsymbol{\sigma}^{(m+1)}$, otherwise increase m by one and return to Step 1.

This is equivalent to constructing and solving the mixed model equations with a given set of variances, $\boldsymbol{\sigma}^{(m)}$, and then

$$\begin{aligned} \sigma_0^{2(m+1)} &= (\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \hat{\mathbf{u}}'\mathbf{Z}'\mathbf{y})/N, \\ \text{and } \sigma_i^{2(m+1)} &= (\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i + \sigma_0^{2(m+1)}tr\mathbf{T}_{ii})/q_i. \end{aligned}$$

3.3 Restricted Maximum Likelihood

Restricted (or Residual) maximum likelihood (REML), was first suggested by Thompson (1962), and was described formally by Patterson and Thompson (1971). The procedure requires that y have a multivariate normal distribution. The method is translation invariant. The maximum likelihood approach antomatically keeps the estimator within the allowable parameter space(i.e. zero to plus infinity), and therefore, REML is a biased procedure. REML was proposed as an improvement to ML in order to account for the degrees of freedom lost in estimating fixed effects.

The likelihood function used in REML is that for a set of error contrasts (i.e. residuals) that are assumed to have a multivariate normal distribution. The multivariate normal distribution likelihood function for the residual contrasts, $\mathbf{K}'\mathbf{y}$, where $\mathbf{K}'\mathbf{X} = 0$, and $\mathbf{K}'$ has rank equal to $N - r(\mathbf{X})$, is

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

The natural log of the likelihood function is

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

Notice that $-.5(N - r(\mathbf{X}))\ln(2\pi)$ is a constant that does not depend on the unknown variance components or factors in the model, and therefore, can be ignored to give L_4. Searle (1979) showed that

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

and

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

for any $\mathbf{K}'$ such that $\mathbf{K}'\mathbf{X} = \mathbf{0}$. Hence, L_4 can be written as

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

REML can be calculated a number of different ways.

1. **Derivative Free approach** is a search technique to find the parameters that maximize the log likelihood function. Two techniques will be described here.
2. **First Derivatives and EM** is where the first derivatives of the log likelihood are determined and set to zero in order to maximize the likelihood function. Solutions need to be obtained by iteration because the resulting equations are non linear.
3. **Second Derivatives** are generally more computationally demanding. Gradient methods are 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. Lately, the "average information" algorithm (averages the observed and expected information matrices) has been used to reduce the computational time.

All of the approaches attempt to maximize the log likelihood function of the error contrasts. To illustrate the methods, consider a single trait model with three factors (F, A, B), of which A and B are random factors. There were a total of 90 observations, and the total sum of squares was 356,000. The least sqnares equations for this small example are shown below.

$$\begin{pmatrix} 50 & 0 & 5 & 15 & 30 & 5 & 10 & 20 & 15 \\ 0 & 40 & 5 & 15 & 20 & 5 & 10 & 20 & 5 \\ 5 & 5 & 10 & 0 & 0 & 2 & 3 & 4 & 1 \\ 15 & 15 & 0 & 30 & 0 & 5 & 7 & 11 & 7 \\ 30 & 20 & 0 & 0 & 50 & 3 & 10 & 25 & 12 \\ 5 & 5 & 2 & 5 & 3 & 10 & 0 & 0 & 0 \\ 10 & 10 & 3 & 7 & 10 & 0 & 20 & 0 & 0 \\ 20 & 20 & 4 & 11 & 25 & 0 & 0 & 40 & 0 \\ 15 & 5 & 1 & 7 & 12 & 0 & 0 & 0 & 20 \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ A_1 \\ A_2 \\ A_3 \\ B_1 \\ B_2 \\ B_3 \\ B_4 \end{pmatrix} = \begin{pmatrix} 3200 \\ 2380 \\ 580 \\ 1860 \\ 3140 \\ 700 \\ 1320 \\ 2400 \\ 1160 \end{pmatrix}.$$

3.3.1 Derivative Free REML

Derivative Free REML was proposed by Smith and Graser(1986) and Graser, Smith and Tier(1987) and has been expanded upon by Meyer (1987,91) who has developed a set of programs for computing estimates of variance components for a whole range of univariate and multivariate models. The description given below is a very simplified version of the method for basic understanding of the technique.

Imagine an s dimensional array containing the values of the likelihood function for every possible set of values of the ratios of the components to the residual variance. The technique is to search this array and find the set of ratios for which thc likelihood function is maximized. There is more than one way to conduct this search. Care must be taken to find the 'global' maximum rather than one of possibly many 'local' maxima. At the same time thc number of likelihood evaluations to be computed must also be minimized.

Various alternative forms of L_4 can be derived. Note that

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

and that

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

and that combining these results gives

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

Now note that

$$\ln \mid \mathbf{R} \mid \; = \; \ln \mid \mathbf{I}\sigma_0^2 \mid$$

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

Then

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

and finally, the log-likelihood fuuction becomes

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

where

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

Note that

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

so that

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

The quantity $\mathbf{y}'\mathbf{P}\mathbf{y}$ is $\mathbf{y}'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}\hat{\mathbf{u}})/\sigma_0^2$. The computations are achieved by constructing the following matrix,

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

then by Gaussian elimination of one row at a time, the sum of the log of the non-zero pivots (using the same ordering for each evaluation of the likelihood) gives log $\mid \mathbf{C}^{\star} \mid$ and $\mathbf{y}'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}\hat{\mathbf{u}})$. Gaussian elimination, using sparse matrix techniques, requires less computing time than inverting the coefficient matrix of the mixed model equations. The ordering of factors within the equations could be critical to the computational process and some experimentation may be necessary to determine the best ordering. The likelihood

function can be evaluated without the calculation of solutions to the mixed model equations, without inverting the coefficient matrix of the mixed model equations, and without computing any of the σ_i^2. The formulations for more general models and multiple trait models are more complex, but follow the same ideas.

Searching the array of likelihood values for various values of α_i can be done in several different ways. One method is to fix the values of all but one of the s α_i, and then evaluate L_2 for four or more different values of the α_i that were not fixed. Then one can use a quadratic regression analysis to determine the value of that one ratio which maximizes L_2 given that the other ratios are fixed. This is repeated for each of the s ratios, and the process is repeated until a maximum likelihood is obtained. The calculations are demonstrated in the example that follows.

Begin by fixing the value of $\alpha_B = 10$ and letting the value of α_A take on the values of (5, 10, 20, 30, 40). Using L_4 to evaluate the likelihood, then the results were as follows:

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

For example, the likelihood value for $\alpha_A = 40$, would be

$$L_4 = -\frac{1}{2}[(N - r(\mathbf{X}))\ln\sigma_0^2 - q_A \ln\alpha_A - q_B \ln\alpha_B + \ln \mid \mathbf{C}^* \mid + \mathbf{y}'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}\hat{\mathbf{u}})/\sigma_0^2]$$

where

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

then

$$\begin{aligned}
L_4 &= -0.5[88(4.5685) - 11.0666 - 9.2103 + 32.0525 + (8483.176/96.3997)] \\
&= -250.9019.
\end{aligned}$$

To find the value of α_A that maximizes L_4 for $\alpha_B = 10$, let

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

then

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

From this a prediction equation for L_4 can be written as

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

This equation can be differentiated with respect to α_A and then equated to zero to find the value of the ratio that maximizes the prediction equation. This gives

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

Now keep $\alpha_A = 32.1546$ and try a number of values of α_B from 2 to 10, which give the following results.

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

Applying the quadratic regression to these points gives

$$\alpha_B = 1.2625.$$

The next step would be to fix $\alpha_B = 1.2625$ and to try new values for α_A, such as 25 to 40 by units of 1. The range of values becomes finer and finer. To insure that one has found the global maximum, the entire process could be started with vastly different starting values for the ratios, such as $\alpha_B = 50$ and let values for α_A be 40, 50, 60, and 70. The more components there are to estimate, the more evaluations of the likelihood

that are going to be needed, and the more probable that convergence might be to a local maximum rather than to the global maximum.

Please refer to the literature for specification of the log likelihood function for particular models and situations. Also, refer to work by Boldman and Van Vleck (1993) which found a simplification of Meyer's algorithms which reduced computational time by several orders of magnitude. Even so, DFREML has been applied to fairly small data sets and can take considerable time to find estimates for these. The available software may not be able to handle particular models, and so the user should be aware of these possible problems.

The Simplex Method

The Simplex Method (Nelder and Mead, 1965) is a procedure for finding the minimum of a function (i.e. the minimum of $-2L_4$ or the maximum of L_4) with respect to the unknown variances and covariances. The best way to describe the method is using the example data from the previous sections. Begin by constructing a set of 'points' for which L_4 is to be evaluated. A 'point' is a vector of values for the unknowns (α_A, α_B), for example,

$$\theta_1 = \begin{pmatrix} 12.1 & 3.8 \end{pmatrix},$$

then form two more points by changing one unknown at a time. Let the three points be as shown in the following table.

No.	α_A	α_B
1	12.1	3.8
2	13.1	3.8
3	12.1	4.3

Now calculate L_4 for each point and arrange from largest to lowest value.

No.	α_A	α_B	L_4
2	13.1	3.8	-250.3047
1	12.1	3.8	-250.3197
3	12.1	4.3	-250.3662

The idea now is to find another point to replace the last one(lowest L_4). This is done by a process called *reflection*. Compute the mean of all points excluding the one with the lowest L_4.

$$\theta_m = \begin{pmatrix} 12.6 & 3.8 \end{pmatrix},$$

then the reflection step is

$$\theta_4 = \theta_m + r * (\theta_m - \theta_{last}),$$

where r is recommended by Nelder and Mead (1965) to be 1, giving

$$\theta_4 = \begin{pmatrix} 13.1 & 3.3 \end{pmatrix}.$$

The corresponding L_4 for this point was -250.2722. Compared to those in the table it has the largest value, and therefore, is a better point than the other three.

No.	α_A	α_B	L_4
4	13.1	3.3	-250.2722
2	13.1	3.8	-250.3047
1	12.1	3.8	-250.3197

Given this success, the Simplex method calls for an *expansion* step, i.e. to make a bigger change. Thus,

$$\theta_5 = \theta_m + E * (\theta_4 - \theta_m),$$

where E is suggested to be equal to 2. Hence

$$\theta_5 = \begin{pmatrix} 13.6 & 2.8 \end{pmatrix}.$$

Then $L_4 = -250.2546$, and the expanded point is better yet. Now drop θ_1 from the table and put θ_5 at the top.

No.	α_A	α_B	L_4
5	13.6	2.8	-250.2546
4	13.1	3.3	-250.2722
2	13.1	3.8	-250.3047

This completes one iteration. Begin the next iteration by computing the mean of all points excluding the point with the lowest L_4.

$$\theta_m = \begin{pmatrix} 13.35 & 3.05 \end{pmatrix}.$$

Another reflection step gives

$$\begin{aligned} \theta_6 &= \theta_m + r * (\theta_m - \theta_{last}), \\ &= \begin{pmatrix} 13.6 & 2.3 \end{pmatrix}. \end{aligned}$$

However, this gives $L_4 = -250.2761$, which is between θ_2 and θ_4, and can push out θ_2 from the table.

No.	α_A	α_B	L_4
5	13.6	2.8	-250.2546
4	13.1	3.3	-250.2722
6	13.6	2.3	-250.2761

Instead of an expansion step, a *contraction* step is needed because θ_6 did not give a greater L_4 than the first two. Thus,

$$\theta_7 = \theta_m + c * (\theta_6 - \theta_m),$$

where $c = 0.5$ is recommended. Hence,

$$\theta_7 = \begin{pmatrix} 13.475 & 3.05 \end{pmatrix}.$$

Then $L_4 = -250.2586$ is better than that given by θ_4, but not by θ_5, thus the new table becomes as follows:

No.	α_A	α_B	L_4
5	13.6	2.8	-250.2546
7	13.475	3.05	-250.2586
4	13.1	3.3	-250.2722

The following steps were taken in the next iteration.

1. The mean of the top two L_4 is

$$\theta_m = \begin{pmatrix} 13.5375 & 2.925 \end{pmatrix}.$$

2. A reflection step gives

$$\begin{aligned} \theta_8 &= \theta_m + r * (\theta_m - \theta_{last}), \\ &= \begin{pmatrix} 13.975 & 2.55 \end{pmatrix}, \end{aligned}$$

   which gave $L_4 = -250.2563$, which is better than θ_7.

3. Add θ_8 to the table and drop θ_4.

No.	α_A	α_B	L_4
5	13.6	2.8	-250.2546
8	13.975	2.55	-250.2563
7	13.475	3.05	-250.2586

4. Because L_4 for θ_8 was not larger than L_4 for θ_5 or smaller than L_4 for θ_7, then no expansion or contraction step is necessary. Begin the next iteration.

The Simplex method continues in this manner until all point entries in the table are equal. The constants recommended by Nelder and Mead (1965) for reflection, expansion, and contraction could be adjusted for a particular data set. This method may converge to a local maximum, and so different starting values are needed to see if it converges to the same point. The Simplex method does not work well with a large number of parameters to be estimated.

3.3.2 First Derivatives and EM Algorithm

To derive formulas for estimating the variance components take the derivatives of L_4 with respect to the unknown components.

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

Combine the two terms involving the traces and note that

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

then

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

for $i = 1, \ldots, s$ or

$$= -.5tr\mathbf{P} + .5\mathbf{y}'\mathbf{P}\mathbf{P}\mathbf{y}$$

for $i = 0$ for the residual component. Using $\mathbf{P}$ and the fact that

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

then

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

and

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

The other terms, $\mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y}$ and $\mathbf{y}'\mathbf{P}\mathbf{P}\mathbf{y}$, were simplified by Henderson (1973) to show that they could be calculated from the Mixed Model Equations. Note that Henderson (1973) showed

$$\begin{aligned}\mathbf{P}\mathbf{y} &= \mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\mathbf{b}}), \\ \hat{\mathbf{b}} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \\ \hat{\mathbf{u}}_i &= \mathbf{G}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y},\end{aligned}$$

then

$$\begin{aligned}\mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y} &= \mathbf{y}'\mathbf{P}\mathbf{Z}_i[\mathbf{G}_i\mathbf{G}_i^{-1}\mathbf{G}_i^{-1}\mathbf{G}_i]\mathbf{Z}_i'\mathbf{P}\mathbf{y} \\ &= (\mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{G}_i)\mathbf{G}_i^{-2}(\mathbf{G}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y}) \\ &= \hat{\mathbf{u}}_i'\mathbf{G}_i^{-2}\hat{\mathbf{u}}_i\end{aligned}$$

which when $\mathbf{G}_i = \mathbf{I}\sigma_i^2$ gives

$$\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i/\sigma_i^4.$$

Similarly for the residual component, Henderson showed that

$$\mathbf{y}'\mathbf{P}\mathbf{P}\mathbf{y} = [\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \sum_{i=1}^{s}(\hat{\mathbf{u}}_i'\mathbf{Z}_i'\mathbf{y} + \hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i\alpha_i)]/\sigma_0^2,$$

where $\alpha_i = \sigma_0^2/\sigma_i^2$.

Equate the derivatives to zero incorporating the above simplifications and obtain

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

As with ML, solutions using the EM algorithm must be computed iteratively. Convergence is usually very slow, if it occurs, and the process may also diverge.

Notice the differences between REML and ML. The denominator for $\hat{\sigma}_0^2$ is $N - r(\mathbf{X})$ rather than N, and in $\hat{\sigma}_i^2$ is $tr\mathbf{C}_{ii}$ rather than $tr\mathbf{T}_{ii}$. The quadratic forms, however, are identical in REML and ML. Accounting for the degrees of freedom to estimate $\mathbf{b}$ has resulted in the REML algorithm.

A major computing problem with the EM algorithm is the calculation of $tr\mathbf{C}_{ii}$, which is the corresponding inverse elements of the mixed model equations for the i^{th} random factor. With most applications in animal breeding, the order of the mixed model equations are too large to be inverted, and solutions to the equations are obtained by Gauss-Seidel iterations. However, there have been several attempts to approximate $tr\mathbf{C}_{ii}$, but these have not been totally suitable.

To demonstrate the EM algorithm let $\alpha_A = 10$ and $\alpha_B = 5$ be the starting values of the ratios for factors A and B, respectively. There were $N = 90$ total observations, and $r(\mathbf{X}) = 2$. The solution vector is

$$\begin{pmatrix} F_1 \\ F_2 \\ A_1 \\ A_2 \\ A_3 \\ B_1 \\ B_2 \\ B_3 \\ B_4 \end{pmatrix} = \begin{pmatrix} 64.6313 \\ 59.4225 \\ -2.1363 \\ .4955 \\ 1.6368 \\ 5.1064 \\ 2.6402 \\ -2.6433 \\ -5.1034 \end{pmatrix}.$$

Then

$$\mathbf{y}'(\mathbf{X}\hat{\mathbf{b}} + \mathbf{Z}\hat{\mathbf{u}}) = 347,871.2661$$

and from the inverse of the coefficient matrix,

$$tr\mathbf{C}_{AA} = .16493, \text{ and } tr\mathbf{C}_{BB} = .3309886$$

which give rise to the following estimates,

$$\begin{aligned}\hat{\sigma}_0^2 &= (356,000 - 347,871.2661)/88 \\ &= 92.371976, \\ \hat{\sigma}_A^2 &= (7.4925463 + .16493(92.371976))/3 \\ &= 7.575855, \\ \hat{\sigma}_B^2 &= (66.0771576 + .3309886(92.371976))/4 \\ &= 24.16280774.\end{aligned}$$

New ratios are formed as

$$\alpha_A = 92.371976/7.575855 = 12.192944,$$

and

$$\alpha_B = 92.371976/24.16280774 = 3.822899$$

and these are used to form the mixed model equations again, new solutions and traces are calculated, and so on, until the estimated ratios and the prior values of the ratios are equal. The estimates converge to

$$\begin{aligned}\hat{\sigma}_0^2 &= 91.8639, \\ \hat{\sigma}_A^2 &= 2.5692, \\ \hat{\sigma}_B^2 &= 30.5190.\end{aligned}$$

or

$$\alpha_A = 35.7558, \quad \text{and} \quad \alpha_B = 3.0101.$$

3.3.3 Second Derivatives, Average Information

Second derivatives of the log likelihood lead to the expectations of the quadratic forms. One technique, MIVQUE (Minimum Variance Quadratic Unbiased Estimation) equates the quadratic forms to their expectations. The estimates are unbiased and if all variances remain positive, then convergence will be to the REML estimates. However, due to a shortage of data or an inappropriate model, the estimates derived in this manner can be negative. Computing the expectations of the quadratic forms requires the inverse of the mixed model equations coefficient matrix, and then products and crossproducts of various parts of the inverse.

A gradient method using first and second derivatives can be used (Hofer, 1998). The gradient, $\mathbf{d}$ (the vector of first derivatives of the log likelihood), is used to determine the direction towards the parameters that give the maximum of the log likelihood, such that

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

where $\mathbf{d}^{(t)}$ are the first derivatives evaluated at $\theta = \theta^{(t)}$, and $\mathbf{M}^{(t)}$ in the Newton-Raphson(NR) algorithm is the observed information matrix, and in the Fisher Method of Scoring(FS) it is the expected information matrix.

The first derivatives are as follows (from earlier in these notes):

$$\frac{\partial L_4}{\partial \sigma_i^2} = -.5tr\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i' + .5\mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y} = 0$$

for $i = 1, \ldots, s$ or

$$\frac{\partial L_4}{\partial \sigma_0^2} = -.5tr\mathbf{P} + .5\mathbf{y}'\mathbf{P}\mathbf{P}\mathbf{y} = 0$$

for the residual component. Then from earlier results,

$$\begin{aligned} tr\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i' &= q_i/\sigma_i^2 - tr\mathbf{C}_{ii}\sigma_0^2/\sigma_i^4, \\ \mathbf{y}'\mathbf{P}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}\mathbf{y} &= \hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i/\sigma_i^4 \end{aligned}$$

which combined give

$$0.5(\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i/\sigma_i^4 - q_i/\sigma_i^2 + tr\mathbf{C}_{ii}\sigma_0^2/\sigma_i^4) = 0,$$

for $i = 1, \ldots, s$, and

$$\begin{aligned} tr\mathbf{P} &= (N - r(\mathbf{X}))\sigma_0^2 - \sum_{i=1}^{s} \hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i/\sigma_i^2 \\ \mathbf{y}'\mathbf{P}\mathbf{P}\mathbf{y} &= [\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \sum_{i=1}^{s}(\hat{\mathbf{u}}_i'\mathbf{Z}_i'\mathbf{y} + \hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i\alpha_i)]/\sigma_0^2 \end{aligned}$$

which combined give

$$0.5([\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \sum_{i=1}^{s}(\hat{\mathbf{u}}_i'\mathbf{Z}_i'\mathbf{y} + \hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i\alpha_i)]/\sigma_0^2 - (N - r(\mathbf{X}))\sigma_0^2 + \sum_{i=1}^{s}\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i/\sigma_i^2) = 0,$$

which simplifies to

$$0.5([\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \sum_{i=1}^{s}\hat{\mathbf{u}}_i'\mathbf{Z}_i'\mathbf{y}]/\sigma_0^2 - (N - r(\mathbf{X}))\sigma_0^2) = 0.$$

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

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

and

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

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

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

As the name *Average Information* implies, average the *observed* and *expected* information matrices to give the following matrix of elements.

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

The first derivatives form the vector, $\mathbf{d}^{(t)}$, and

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

The rest of this method is computational detail to simplify the requirements for inverse elements and solutions to MME. The calculations can not be illustrated very easily for the example data because the **y**-vector is not available.

3.3.4 Animal Models

The model commonly applied to estimation of variance components in livestock genetics since 1989 has been an animal model. The animal model assumes a large, random mating population, an infinite number of loci each with a small and equal effect on the trait, only additive genetic effects, and all relationships among animals are known and tracible to an unselected base population (somewhere in the past). Animals may bave more than one record each. The equation of the model is

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

where **a** is the vector of animal additive genetic effects (one per animal), and **p** is a vector of permanent environmental (p.e.) effects associated with each animal.

$$\begin{aligned} E(\mathbf{y}) &= \mathbf{X}\mathbf{b}, \\ Var\begin{pmatrix} \mathbf{a} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} &= \begin{pmatrix} \mathbf{A}\sigma_a^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_p^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2 \end{pmatrix}. \end{aligned}$$

The matrix $\mathbf{A}$ is called the numerator relationship matrix. Wright defined relationships among animals as correlations, but $\mathbf{A}$ is essentially relationships defined as covariances (the numerators of the correlation coefficients). Also, these only represent the additive genetic relationships between animals.

The MME for this model are

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}+\mathbf{A}^{-1}k_a & \mathbf{Z}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z}+\mathbf{I}k_p \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{p}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{pmatrix}.$$

Note that k_a is the ratio of residual to additive genetic variances, and k_p is the ratio of residual to permanent environmental variances. Also, in MME the inverse of $\mathbf{A}$ is required.

The EM-REML procedure gives

$$\begin{aligned} \hat{\sigma}_e^2 &= (\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \hat{\mathbf{a}}'\mathbf{Z}'\mathbf{y} - \hat{\mathbf{p}}'\mathbf{Z}'\mathbf{y})/(N - r(\mathbf{X})), \\ \hat{\sigma}_a^2 &= (\hat{\mathbf{a}}'\mathbf{A}^{-1}\hat{\mathbf{a}} + tr\mathbf{A}^{-1}\mathbf{C}_{aa}\hat{\sigma}_e^2)/n, \\ \hat{\sigma}_p^2 &= (\hat{\mathbf{p}}'\hat{\mathbf{p}} + tr\mathbf{C}_{pp}\hat{\sigma}_e^2)/n, \end{aligned}$$

where n is the total number of animals, N is the total number of records, and $\mathbf{C}_{aa}$ are the inverse elements of the MME for the animal additive genetic effects, and $\mathbf{C}_{pp}$ are the inverse elements of the MME for the animal permanent environmental effects. An example of this model will be given in later notes.

Quadratic Forms in an Animal Model

A necessary quadratic form in an animal model is $\hat{\mathbf{a}}'\mathbf{A}^{-1}\hat{\mathbf{a}}$, and this can be computed very easily. Note that the inverse of $\mathbf{A}$ may be written as

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

where $\mathbf{T}^{-1}$ is an upper triangular matrix, and diagonal matrix $\mathbf{D}^{-2}$ has elements equal to 1, 2, or 4/3 in noninbred situations, and values greater than 2 in inbred situations. In Henderson (1975), this inverse was shown to be composed of just three numbers, i.e. 0, 1's on the diagonals, and -.5 corresponding to the parents of an animal. For example,

$$\mathbf{T}'^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -.5 & -.5 & 1 & 0 \\ -.5 & 0 & -.5 & 1 \end{pmatrix}.$$

Then

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

for the i^{th} animal, and $\hat{a}_s$ and $\hat{a}_d$ are the sire and dam estimated breeding values, respectively. Consequently,

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

where d^{ii} are the diagonal elements of $\mathbf{D}^{-2}$, and q is the number of animals.

Chapter 4

Bayesian Methods

These notes are based on a course given by Daniel Sorensen in 1998 at Armidale, NSW, Australia. Subsequently, a book has been published in 2002 by Sorensen and Gianola entitled "Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics". A Bayesian approach to estimation problems, in general, seems intuitively appealing to animal breeders. Every element of a model is a random variable derived from a distribution function. A fixed factor becomes a random variable with possibly a uniform distribution going from a lower limit to an upper limit. A component of variance is a random variable having a Gamma or Chi-square distribution with x degrees of freedom. In addition, the researcher may have information from previous experiments that strongly indicate the value that a variance component may have, and the Bayes approach allows the *apriori* information to be included in the analysis.

The Bayesian process is to specify distributions for each random variable of the model. These are combined to form the joint posterior distribution. Finding estimators via differentiation of the joint posterior distribution may be difficult to achieve. Gibbs Sampling is a tool for deriving estimates of parameters from the joint posterior distribution without the differentiations. By determining conditional marginal distributions for each random variable of the model, then generating random samples from these distributions eventually converge to random samples from the joint posterior distribution. Computationally, any program that calculates solutions to Henderson's mixed model equations can be modified to implement Gibbs Sampling. Very good random number generators and a substantial amount of computer time are needed for large data sets in animal breeding to apply Gibbs Sampling.

4.1 The Joint Posterior Distribution

Begin with the simple single trait animal model. That is,

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

The Bayesian approach is to derive the joint posterior distribution by application of Bayes theorem. If θ is a vector of random variables and $\mathbf{y}$ is the data vector, then

$$\begin{aligned} p(\theta, \mathbf{y}) &= p(\theta)\ p(\mathbf{y} \mid \theta) \\ &= p(\mathbf{y})\ p(\theta \mid \mathbf{y}) \end{aligned}$$

Re-arranging gives

$$\begin{aligned} p(\theta \mid \mathbf{y}) &= \frac{p(\theta) p(\mathbf{y} \mid \theta)}{p(\mathbf{y})} \\ &= (\text{prior for } \theta)\frac{p(\mathbf{y} \mid \theta)}{p(\mathbf{y})} \\ &= \text{posterior probability function of } \theta \end{aligned}$$

In terms of the simple animal model, θ includes $\mathbf{b}$, $\mathbf{a}$, σ_a^2, and σ_e^2. The conditional distribution of $\mathbf{y}$ given θ is

$$\mathbf{y} \mid \mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2 \sim N(\mathbf{Xb} + \mathbf{Za}, \mathbf{I}\sigma_e^2),$$

and

$$p(\mathbf{y} \mid \mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2) \propto (\sigma_e^2)^{(-N/2)} \exp\left[-(\mathbf{y} - \mathbf{Xb} - \mathbf{Za})'(\mathbf{y} - \mathbf{Xb} - \mathbf{Za})/2\sigma_e^2\right].$$

Prior distributions need to be assigned to the components in θ, and these need to be multiplied together and times the conditional distribution of $\mathbf{y}$ given θ. For the fixed effects vector, $\mathbf{b}$, there is little prior knowledge about the values that elements in that vector might have. This is represented by assuming

$$p(\mathbf{b}) \propto \text{constant}.$$

For $\mathbf{a}$, the vector of additive genetic values, quantitative genetics theory suggests that they follow a normal distribution, i.e.

$$\mathbf{a} \mid \mathbf{A}, \sigma_a^2 \sim N(\mathbf{0}, \mathbf{A}\sigma_a^2)$$

and

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

where q is the length of $\mathbf{a}$. A natural estimator of σ_a^2 is $\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}/q$, call it S_a^2, where

$$S_a^2 \sim \chi_q^2 \sigma_a^2/q.$$

Multiply both sides by q and divid by χ^2_q to give

$$\sigma^2_a \sim qS^2_a/\chi^2_q$$

which is a scaled, inverted Chi-square distribution, written as

$$p(\sigma^2_a \mid v_a, S^2_a) \propto (\sigma^2_a)^{-(\frac{v_a}{2}+1)} \exp\left(-\frac{v_a}{2}\frac{S^2_a}{\sigma^2_a}\right),$$

where v_a and S^2_a are hyperparameters with S^2_a being a prior guess about the value of σ^2_a and v_a being the degrees of belief in that prior value. Usually q is much larger than v_a and therefore, the data provide nearly all of the information about σ^2_a. Similarly, for the residual variance,

$$p(\sigma^2_e \mid v_e, S^2_e) \quad \propto \quad (\sigma^2_e)^{-(\frac{v_e}{2}+1)} \exp\left(-\frac{v_e}{2}\frac{S^2_e}{\sigma^2_e}\right).$$

Now form the joint posterior distribution as

$$p(\mathbf{b}, \mathbf{a}, \sigma^2_a, \sigma^2_e \mid \mathbf{y}) \propto p(\mathbf{b})p(\mathbf{a} \mid \sigma^2_a)p(\sigma^2_a)p(\sigma^2_e)p(\mathbf{y} \mid \mathbf{b}, \mathbf{a}, \sigma^2_a, \sigma^2_e)$$

which can be written as

$$\propto (\sigma^2_e)^{-(\frac{N+v_e}{2}+1)} \exp\left[-\frac{1}{2\sigma^2_e}((\mathbf{y}-\mathbf{Xb}-\mathbf{Za})'(\mathbf{y}-\mathbf{Xb}-\mathbf{Za}) + v_e S^2_e)\right]$$

$$(\sigma^2_a)^{-(\frac{q+v_a}{2}+1)} \exp\left[-\frac{1}{2\sigma^2_a}(\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} + v_a S^2_a)\right].$$

4.2 Fully Conditional Posterior Distributions

In order to implement Gibbs sampling, all of the fully conditional posterior distributions (one for each component of θ) need to be derived from the above joint posterior distribution. The conditional posterior distribntion is derived from the joint posterior distribution by picking out the parts that involve the unknown parameter in question. Let

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

A new notation is introduced, let

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

where β_i is a scalar representing just one element of the vector β, and β_{-i} is a vector representing all of the other elements except β_i. Similarly, $\mathbf{C}$ and $\mathbf{W}$ can be partitioned in the same manner as

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

In general terms, the conditional posterior distribution of β is a normal distribution,

$$\beta_i \mid \beta_{-i}, \sigma_a^2, \sigma_e^2, \mathbf{y} \sim N(\hat{\beta}_i, C_{i,i}^{-1}\sigma_e^2)$$

where

$$C_{i,i}\hat{\beta}_i = (\mathbf{W}_i'\mathbf{y} - \mathbf{C}_{i,-i}\beta_{-i}).$$

Then

$$b_i \mid \mathbf{b}_{-i}, \mathbf{a}, \sigma_a^2, \sigma_e^2, \mathbf{y} \sim N(\hat{b}_i, C_{i,i}^{-1}\sigma_e^2),$$

for

$$C_{i,i} = \mathbf{x}_i'\mathbf{x}_i.$$

Also,

$$a_i \mid \mathbf{b}, \mathbf{a}_{-i}, \sigma_a^2, \sigma_e^2, \mathbf{y} \sim N(\hat{a}_i, C_{i,i}^{-1}\sigma_e^2),$$

where $C_{i,i} = (\mathbf{z}_i'\mathbf{z}_i + A^{i,i}k)$, for $k = \sigma_e^2/\sigma_a^2$.

The conditional posterior distributions for the variances are inverted Chi-square distributions,

$$\sigma_a^2 \mid \mathbf{b}, \mathbf{a}, \sigma_e^2, \mathbf{y} \sim \tilde{v}_a\tilde{S}_a^2\chi_{\tilde{v}_a}^{-2}$$

for $\tilde{v}_a = q + v_a$, and $\tilde{S}_a^2 = (\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} + v_aS_a^2)/\tilde{v}_a$, and

$$\sigma_e^2 \mid \mathbf{b}, \mathbf{a}, \sigma_a^2, \mathbf{y} \sim \tilde{v}_e\tilde{S}_e^2\chi_{\tilde{v}_e}^{-2}$$

for $\tilde{v}_e = N + v_e$, and $\tilde{S}_e^2 = (\mathbf{e}'\mathbf{e} + v_eS_e^2)/\tilde{v}_e$, and $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathrm{b} - \mathbf{Za}$.

4.3 Computational Scheme

Gibbs sampling is much like Gauss-Seidel iteration. When a new solution is calculated in the Mixed Model Equations for a level of a fixed or random factor, a random amount is added to the solution based upon its conditional posterior distribution variance before proceeding to the next level of that factor or the next factor. After all equations have been processed, new values of the variances are calculated and a new variance ratio is

determined prior to beginning the next round. The following MME for five animals will be used to illustrate the Gibbs sampling scheme:

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

where $k = \sigma_e^2/\sigma_a^2 = 14$, and

$$\mathbf{A}^{-1} = \frac{1}{14} \begin{pmatrix} 28 & 7 & -7 & -14 & 0 \\ 7 & 29 & -14 & 8 & -16 \\ -7 & -14 & 35 & -14 & 0 \\ -14 & 8 & -14 & 36 & -16 \\ 0 & -16 & 0 & -16 & 32 \end{pmatrix}.$$

The starting values for $\beta = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$, and for $v_a = v_e = 10$, and $S_e^2 = 93\frac{1}{3}$ and $S_a^2 = 6\frac{2}{3}$, so that $k = 14$. Let RND represent a random normal deviate from a random normal deviate generator, and let $CHI(idf)$ represent a random Chi-square variate from a random Chi-Square variate generator with idf degrees of freedom. To begin, let $\sigma_e^2 = S_e^2$ and $\sigma_a^2 = S_a^2$. Below are descriptions of calculations in the first two rounds.

4.3.1 Round 1

Process each factor in the model, one equation at a time.

Overall mean

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

Animal 1

$$\hat{a}_1 = (38.5 - \mu - 7a_2 + 7a_3 + 14a_4)/29$$

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

Animal 2

$$
\begin{aligned}
\hat{a}_2 &= (48.9 - \mu - 7a_1 + 14a_3 - 8a_4 + 16a_5)/30 \\
&= -6.8591/30 = -.2286 \\
a_2 &= \hat{a}_2 + RND * (\sigma_e^2/30)^{.5} \\
&= -.2286 + (.0047)(1.7638) \\
&= -.2203
\end{aligned}
$$

Animal 3

$$
\begin{aligned}
\hat{a}_3 &= (64.3 - \mu + 7a_1 + 14a_2 + 14a_4)/36 \\
&= .8931 \\
a_3 &= \hat{a}_3 + RND * (\sigma_e^2/36)^{.5} \\
&= .8931 + (-1.1061)(1.6102) \\
&= -.8879
\end{aligned}
$$

Animal 4

$$
\begin{aligned}
\hat{a}_4 &= (50.5 - \mu + 14a_1 - 8a_2 + 14a_3 + 16a_5)/37 \\
&= .6518 \\
a_4 &= \hat{a}_4 + RND * (\sigma_e^2/37)^{.5} \\
&= .6518 + (-1.2293)(1.5882) \\
&= -1.3006
\end{aligned}
$$

Animal 5

$$\begin{aligned}
\hat{a}_5 &= (36.0 - \mu + 16a_2 + 16a_4)/33 \\
&= -.9316 \\
a_5 &= \hat{a}_5 + RND * (\sigma_e^2/33)^{.5} \\
&= -.9316 + (-.6472)(1.6817) \\
&= -2.0200
\end{aligned}$$

Residual Variance

Now calculate the residuals and their sum of squares in order to obtain a new residual variance.

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

A new sample value of the residual variance is

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

Additive Genetic Variance

The additive genetic variance requires calculation of $\mathbf{a'A^{-1}a}$ using the a-values obtained above, which gives

$$\mathbf{a'A^{-1}a} = 19.85586.$$

Then

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

A new sample value of the variance ratio becomes

$$k = 95.6382/8.0605 = 11.8650.$$

4.3.2 Round 2

Round 2 begins by re-forming the MME using the new variance ratio. The equations change to

$$\begin{pmatrix} 5 & 1 & 1 & 1 & 1 & 1 \\ 1 & 24.73 & 5.93 & -5.93 & -11.86 & 0 \\ 1 & 5.93 & 25.58 & -11.86 & 6.78 & -13.56 \\ 1 & -5.93 & -11.86 & 30.66 & -11.86 & 0 \\ 1 & -11.86 & 6.78 & -11.86 & 31.51 & -13.56 \\ 1 & 0 & -13.56 & 0 & -13.56 & 28.12 \end{pmatrix} \begin{pmatrix} \mu \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} = \begin{pmatrix} 238.2 \\ 38.5 \\ 48.9 \\ 64.3 \\ 50.5 \\ 36.0 \end{pmatrix}.$$

The process is repeated using the last values of μ and $\mathbf{a}$ and σ_e^2.

$$\begin{aligned}
\hat{\mu} &= (238.2 - a_1 - a_2 - a_3 - a_4 - a_5)/5 \\
&= 48.14 \\
\mu &= \hat{\mu} + RND * (\sigma_e^2/5)^{.5} \\
&= 48.14 + (.7465) * (4.3735) \\
&= 51.41 \\
\hat{a}_1 &= (38.5 - \mu - 5.93a_2 + 5.93a_3 + 11.86_4)/24.73 \\
&= -1.3059 \\
a_1 &= \hat{a}_1 + RND * (\sigma_e^2/24.73)^{.5} \\
&= -1.3059 + (-.0478)(1.9665) \\
&= -1.3999 \\
\hat{a}_2 &= (48.9 - \mu - 5.93a_1 + 11.86a_3 - 6.78a_4 + 13.56a_5)/25.58 \\
&= -.9113 \\
a_2 &= \hat{a}_2 + RND * (\sigma_e^2/25.58)^{.5} \\
&= -.9113 + (.8386)(1.9336) \\
&= .7102 \\
\hat{a}_3 &= -2.41355/30.66 \\
&= -.0787 \\
a_3 &= \hat{a}_3 + RND * (\sigma_e^2/30.66)^{.5} \\
&= -.0787 + (-1.8414)(1.7662) \\
&= -3.3309 \\
\hat{a}_4 &= -89.2236/31.51 = -2.8316 \\
a_4 &= -2.8316 + (-1.2549)(1.7422) \\
&= -5.0179
\end{aligned}$$

$$\begin{aligned}\hat{a}_5 &= -73.8224/28.12 = -2.6253\\ a_5 &= -2.6253 + (.8184)(1.8442)\\ &= -1.1160\end{aligned}$$

The residuals and their sum of squares are

$$\begin{aligned}e_1 &= 38.5 - 51.41 + 1.3999 = -11.5101\\ e_2 &= 48.9 - 51.41 - .7102 = -3.2202\\ e_3 &= 64.3 - 51.41 + 3.3309 = 16.2209\\ e_4 &= 50.5 - 51.41 + 5.0179 = 4.1079\\ e_5 &= 36.0 - 51.41 + 1.1160 = -14.2940\\ \mathbf{e}'\mathbf{e} &= 627.1630\end{aligned}$$

The new sample value of the residual variance is

$$\begin{aligned}\sigma_e^2 &= (\mathbf{e}'\mathbf{e} + v_e S_e^2)/CHI(15)\\ &= (627.1630 + (10)(93.3333))/20.4957\\ &= 76.1377.\end{aligned}$$

The new sample value of the additive genetic variance is

$$\begin{aligned}\sigma_a^2 &= (\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} + v_a S_a^2)/CHI(15)\\ &= (36.8306 + (10)(6.66667))/16.6012\\ &= 6.2343.\end{aligned}$$

The new variance ratio becomes

$$k = 76.1377/6.2343 = 12.2127.$$

Continue taking samples for thousands of rounds.

4.3.3 Burn-In Periods and Estimates

The samples do not immediately represent samples from the joint posterior distribution. Generally, this takes anywhere from 100 to 10,000 samples depending on the model. This period is known as the *burn-in period.* Samples from the burn-in period are discarded. The length of the burn-in period (i.e. number of samples) is usually judged by visually inspecting a plot of sample values across rounds.

A less subjective approach to determine convergence to the joint posterior distribution is to run two chains at the same time, both beginning with the same random number

seed. However, the starting values (in variances) for each chain are usually greatly different, e.g. one set is greatly above the expected outcome and the other set is greatly below the expected outcome. When the two chains essentially become one chain, i.e. the squared difference between variance estimates is less than a specified value (like 10^{-5}), then convergence to the joint posterior distribution has occurred. All previous samples are considered to be part of the burn-in period and are discarded.

After burn-in, each round of Gibbs sampling is dependent on the results of the previous round. Depending on the total number of observations and parameters, one round may be positively correlated with the next twenty to three hundred rounds. The user can determine the effective number of samples by calculating lag correlations, i.e. the correlation of estimates between rounds, between every other round, between every third round, etc. Suppose a total of 12,000 samples (after removing the burn-in rounds) gave an effective number of samples equal to 500. This implies that samples that are 24 rounds apart should be uncorrelated.

An overall estimate of a parameter can be obtained by averaging all of the 12,000 samples (after the burn-in). However, to derive a confidence interval or to plot the distribution of the samples or to calculate the standard deviation of the sample values, the variance of the 500 independent samples should be used.

The final estimates are therefore, an average of the sample estimates. Some research has shown that the mode of the estimates might be a better estimate, which indicates that the distribution of sample estimates is skewed. One could report both the mean and mode of the samples, however, the mode should be based on the independent samples only.

4.3.4 Influence of the Priors

In the small example, $v_a = v_e = 10$ whereas N was only 5. Thus, the prior values of the variances received more weight than information coming from the data. This is probably appropriate for this small example, but if N were 5,000,000, then the influence of the priors would be next to nothing. The amount of influence of the priors is not directly determined by the ratio of v_i to N. In the small example, even though $v_a/(N + v_a) = \frac{2}{3}$, the influence of S_a^2 could be greater than $\frac{2}{3}$, (Schenkel, 1998)

4.3.5 Long Chain or Many Chains?

Early papers on MCMC (Monte Carlo Markov Chain) methods recommended running many chains of samples and then averaging the final values from each chain. This was to insure independence of the samples. Another philosophy recommends one single long chain. For animal breeding applications this could mean 100,000 samples or more. If a month is needed to run 50,000 samples, then maybe three chains of 50,000 would be

preferable. If only an hour is needed for 50,000 samples, then 1,000,000 samples would not be difficult to run.

Two chains that utilize the same sequence of random numbers, but which use different starting variances, are recommended for determining the burn-in period, after which enough samples need to be run to generate a sufficient number of independent samples for obtaining standard deviations of the samples. A sufficient number of independent samples may be 100 or more depending on the amount of time needed to generate samples.

Chapter 5

Data Readiness

Prior to any analysis, data and pedigree files need to be prepared as input to a general program. General programs expect the pedigree to be arranged in a particular manner and the data file might also have to be re-arranged and sorted. Data preparation and validation are probably the most time consuming stages of any analysis.

5.1 Pedigree Files

The requisite information in a pedigree file is the animal ID, the sire ID, the dam ID, and possibly the year or date of birth. Additional information might be if the animal was a clone of another animal, or was the result of embryo transfer or embryo splitting. The surrogate dam should not be recorded as the biological dam of such animals, but should be in the pedigree file to account for maternal effects provided by the surrogate dam. Below is an example of a pedigree file that will be used in these notes.

Animal	Sire	Dam	Birth Year
348097	351604	351342	1981
349876	352515	350873	1981
350010	348097	349876	1981
350873	0	0	1980
351011	353118	351342	1982
351342	0	0	1980
351604	350121	0	1980
352012	352229	350873	1982
352229	352515	351342	1981
352515	0	0	1980
353013	353118	350010	1982
353118	350873	351604	1981

```
354018  353013  352012  1983
354317  348097  351011  1983
354516  351604  351342  1983
354715  352515  351011  1983
354914  351604  349876  1982
```

There are 17 animals in the list sorted by animal registration number. The following features can be identified about this file.

1. The registration numbers cannot be used to order parents before their progeny. The sire ID of the first animal is greater than the animal's ID, which would not have occurred if animals were registered at birth in a consecutive manner.
2. Some of the pedigrees have missing information on sire and/or dam.
3. Animal 350121 appears as a sire, but does not have its own record in the file. Thus, there should be 18 animals in the file.
4. Animals 350873 and 351604 appear as both sire and dam in the file.

The requirements for a pedigree file are

1. All animals should be numbered consecutively, and
2. Parent ID numbers should be smaller than the smallest progeny ID.

An additional requirement might be that an animal ID can appear only as a sire or a dam, but not both. Sometimes sire and dam IDs may be entered in the wrong fields, and if they have several progeny then they could appear as both a sire and a dam. In most cases this is not a problem, but if the genetic model includes maternal genetic effects, then sire IDs should obviously not be in the file as a dam ID.

Finally, all animals that appear as a sire ID or dam ID should be in the pedigree file in the animal ID column including their sire and dam IDs, if known. They should be assigned unknown parents and the earliest birth year.

5.1.1 Achieving Chronological Order

The first step is to assign generation numbers to each animal, and this is done in an iterative manner until the generation numbers no longer need altering. Every animal begins with a generation number equal to one. Parents are given generation numbers that are at least one greater than the generation number of their offspring. Below are the

animals and the generation numbers assigned by iteration. Five iterations were needed with this small pedigree to correctly assign generation numbers. A code of 0 is given to animals without progeny, 1 if the animal appears as a sire, 2 if the animal appears as a dam, and 3 if the animal appears as both a sire and a dam. The animals that appear as both sire and dam can be corrected later, if necessary.

Animal	Sire	Dam	1st	2nd	3rd	4th	5th	Code
348097	351604	351342	1	2	3	4	4	1
349876	352515	350873	1	2	3	4	4	2
350010	348097	349876	1	2	3	3	3	2
350121	0	0	1	2	4	5	6	1
350873	0	0	1	3	3	4	5	3
351011	353118	351342	1	2	2	2	2	2
351342	0	0	1	3	4	4	5	2
351604	350121	0	1	3	3	4	5	3
352012	352229	350873	1	2	2	2	2	2
352229	352515	351342	1	2	3	3	3	1
352515	0	0	1	3	4	4	5	1
353013	353118	350010	1	2	2	2	2	1
353118	350873	351604	1	2	3	3	3	1
354018	353013	352012	1	1	1	1	1	0
354317	348097	351011	1	1	1	1	1	0
354516	351604	351342	1	1	1	1	1	0
354715	352515	351011	1	1	1	1	1	0
354914	351604	349876	1	1	1	1	1	0

An output file with the generation numbers from the 5th iteration can be written, and this file must be sorted by the generation numbers to give the following file.

Gen. No.	Animal	Sire	Dam	Code	Birth Year
6	350121	0	0	1	1980
5	350873	0	0	3	1980
5	351342	0	0	2	1980
5	352515	0	0	1	1980
5	351604	350121	0	3	1980
4	349876	352515	350873	2	1981
4	348097	351604	351342	1	1981
3	352229	352515	351342	1	1981
3	353118	350873	351604	1	1981
3	350010	348097	349876	2	1981
2	353013	353118	350010	1	1982
2	352012	352229	350873	2	1982
2	351011	353118	351342	2	1982
1	354914	351694	349876	0	1982

```
1     354715  352515  351011    0     1983
1     354317  348097  351011    0     1983
1     354516  351604  351342    0     1983
1     354018  353013  352012    0     1983
```

The second step is to renumber the animals consecutively from 1 to 18, and to replace the original registration numbers with the consecutive numbers for animals, sires, and dams. Parents will have smaller numbers than their progeny. The code for indicating sires or dams have been dropped. The sorted output file is as follows;

Animal	BirthYr	Sire	Dam	Original ID
1	1980	0	0	350121
2	1980	0	0	350873
3	1980	0	0	351342
4	1980	0	0	352515
5	1980	1	0	351604
6	1981	4	2	349876
7	1981	5	3	348097
8	1981	4	3	352229
9	1981	2	5	353118
10	1981	7	6	350010
11	1982	9	10	353013
12	1982	8	2	352012
13	1982	9	3	351011
14	1982	5	6	354914
15	1983	4	13	354715
16	1983	7	13	354317
17	1983	5	3	354516
18	1983	11	12	354018

5.1.2 Inbreeding Calculations

The next step is to compute the inbreeding coefficients for each animal according to the Meuwissen and Luo (1997) algorithm. These are needed to compute the diagonals of $\mathbf{D}^{-2}$ in $\mathbf{A}^{-1} = \mathbf{T}^{-1}\mathbf{D}^{-2}\mathbf{T}'^{-1}$, which are, in turn, needed to construct elements of $\mathbf{A}^{-1}$ following Henderson's rules (1975). The diagonals of $\mathbf{D}^{-2}$ for the above animals were equal to 1 for animals 1 to 4, 1.3333 for animal 5, 2.13333 for animal 18, and 2 for all other animals (6 to 17).

Programs that prepare the pedigree file in the above manner are available.

pedf01.f Assigns generation numbers to each animal and fills in the pedigree file. Animal IDs are assumed to be numeric only.

msort Sort routine to sort animals from high to low generation numbers, so that parents appear before their progeny.

pedf02.f Re-numbers animals consecutively, including sire and dam IDs.

msort Sort the output in consecutive number order.

inbrd.f Compute inbreeding coefficients and diagonals of $\mathbf{D}^{-2}$ for use in other programs.

5.2 Data File

The software packages available to estimate variances and covariances generally require data to be prepared in specific formats. Any factor which might be included in a model should be created and present in the data file. For example, Age-of-Dam and Sex-of-Calf might be in the data. If the model will include an interaction between Age-of-Dam and Sex-of-Calf, then a code for the interaction effect, Age-X-Sex, should be included in the data file too. Most of the software packages do not manipulate items in the data to form interaction effects, except SAS.

Unfortunately, software packages differ in the exact requirements for data files. The following generic format will be recommended. Data come as integer values and real values (with decimal places). Some, but not all programs require the y variables to be real values, even if they are integer in nature. Every variable in the data should be separated from the next by at least one blank space. No character data are allowed in these data files. A record in the data file should be arranged as follows:

1. **Animal ID**. The animal ID should be the same as that in the pedigree file. Thus, if the animals in the pedigree file have been re-numbered consecutively, then the animal IDs in the data should be numbered in the same manner. Consequently, the data file should not include any animals that did not appear in the pedigree file. However, the pedigree file may contain many more animals than those that have records.

2. **Integer Values**. The next variables in the record should be any number of integer values. These indicate levels of factors that may be in the model, such as days in milk, age, herd ID, dam ID (consecutive numbers as in pedigree file), month, diet, year, etc.

3. **Real Values**. All covariates and observed traits are included in the real values, with or without a decimal in the number. If the model will be a random regression model, then the appropriate covariates need to be created and stored in the data file. If the observations will be weighted by the inverse of a residual variance, then the weights should be included in the data file.

The specific formats for each software package will be given later, but the above arrangement should be generally suitable.

5.3 Control Files

Each software package requires at least one *control file* that does the following (not necessarily in this order):

1. Identifies the pedigree and data files and the path to find them.
2. Names the variables that are integer and that are real, and a format (sometimes).
3. Model specification. How many traits; what are the fixed factors, covariates, and random factors of the model; weighting factors, if any.
4. Type of analysis. REML EM, DF REML, AI REML, or Bayesian methodology. Not all methods are available in each software package.
5. Starting values for parameters, degrees of belief.
6. Other optional choices, dependent on the software package. These might include choice of output features, choice of level of optimization, choice of Jacobi or Gauss-Seidel iteration strategies, choice of sparse matrix packages, degree of convergence desired, etc.

Each software package has a different name for this *control file*, e.g. parameter file, driver file. Some control files are very easy to set up, and some control files are very tedious. However, a program can be written to convert pedigree and data files (in the above generic formats) into the appropriate format for a specific software package, but none has been written.

Each software package creates a number of different files for its own use. These files are generally written in binary format to save space and to speed up their reading during the analysis phase of the program. Such files are not readable by the user in most cases. Some of these files are important, for example, if a Gibbs sampling chain is interrupted and then must be re-started (without starting from scratch). Some of the files will be output from the analysis, and these may require further processing for interpretation.

Chapter 6

Software Packages

Three software packages are described in this chapter. This does not mean that these packages are recommended or are the best available. The packages are 1) VCE by Groeneveld; 2) DMU by Jensen and Madsen; and 3) MTDFREML by Boldman and others. The package ASREML by Gilmour is not described because this software has been commercialized and is very expensive per copy. Also, the three packages cover all of the REML methods including AI REML. VCE is free. MTDFREML only uses DF REML. Users should try to learn how to use each package, as one package may not be suitable for all models and analyses. Because of the general nature of these programs to handle many different models and methods, the memory requirements can be substantial, thereby limiting either the amount of data that can be accommodated or the model being applied. If the model has to be compromised, then the user may wish to use another software package. The more traits there are and the more parameters to be estimated, the more memory and computing time that will be needed to get estimates. Finally, be cautious with results obtained from any of these programs. Just because results were obtained, the program may not have handled the model correctly, or the way you intended. Correct use of any package is the user's responsibility. If possible, use two software packages or two methodologies and see if the results agree.

6.1 DMU-Jensen and Madsen

DMU was written by Per Madsen and Just Jensen at the Danish Institute of Agricultural Sciences (DIAS), the Research Centre Foulum. The programs are in Fortran 90 or 95. The DMU package consists of modules.

DMU1 This module must always be used. This program reads the control file, and does preliminary massaging of the pedigree and data files for use in the other modules. Pedigree information can be specified in different ways.

DMU4 This module is for solving the mixed model equations in order to obtain estimated breeding values. Standard errors of prediction can be obtained. Several methods of solving the equations can be used. They are

1. Jacobi Conjugate Gradient (JCG),
2. Jacobi Semi-Iteration (JSI),
3. Successive Overrelaxation (SOR),
4. Symmetric SOR Conjugate Gradient,
5. Symmetric SOR Semi-Iteration,
6. Reduced system Conjugate Gradient,
7. Reduced system Semi-Iteration,
8. FSPAK, speed optimized or memory optimized.

DMUAI Module for estimation of covariance components using AI REML or EM REML.

DMUGib Module for estimation of covariance components via Gibbs Sampling and Bayesian methodology.

There is complete documentation available for this package from the source. The best way to explain the program is to describe the creation of the control files for different kinds of models. If something does not work from these notes, then please consult the documentation.

6.1.1 Control File

The control file in DMU is called a *driver file*. The driver file is organized into sections and contains *keywords*, some of which are mandatory.

Single Trait Animal Model

```
$COMMENT
Example driver file for a SINGLE TRAIT, Animal Model
Trait for analysis is T1
Fixed factors are     RC RAS
Random Factors are    HRC ID

$ANALYSE 1 1 0 0

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

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

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

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

Keywords begin with $ and have a specific syntax for what follows it.

$COMMENT Up to 10 lines may follow this keyword. All lines are repeated on the output files to identify the type of analysis. The user can input anything they want to identify the analysis.

$ANALYSE is followed by four numbers.

- The first number specifies the *task* and is either a 1 or 11. A 1 indicates that REML estimation is to be conducted using DMUAI. An 11 indicates that DMU4 is to be used.
- The second number indicates the *method* to be used. If the first number was a 1, then the method can be 1 - AI REML; 2 - EM REML based on Robin Thompson's algorithm; 3 - EM REML based on Esa Mantysaari's algorithm; and 4 - AI REML using step halving if an update goes outside the parameter space. If the first number was 11, then the method can be one of 10 possible methods.

- The third number refers to scaling the data to a residual variance of unity. A 1 indicates yes to do scaling, and a 0 is for no scaling.
- The fourth number refers to the amount of printout to be generated. A 0 stands for standard output. Values of 1 or 2 ask for more output, but the volume of output could be very large.

Thus, in this example, variances are to be estimated using AI REML, no scaling of the data, and only standard output.

item[$DATA] keyword is followed by the format of the data, either ASCII or BINARY, followed by (number of integer values, number of real values, value below which real values are assumed missing), followed by the file name and the full pathway to the file. In this example, the file is in ASCII format, the first 6 numbers are integers and the next 8 are real values. Any real values below -99 are considered to be missing. The location and name of the data file is given.

$VARIABLE keyword is used to give names to the integer and real values in a record. A name may be up to 8 characters in length. Thus, `ID RC RAS HRC HS DAM` are the six integer values and the following 8 are real values. This keyword is not mandatory, but it helps in understanding the next section of the driver file.

$MODEL keyword is for describing the traits and models for each trait.

- **TRAITS**. The first line indicates the number of traits in this analysis. In this case, only 1 trait.
- **ABSORB**. Intended for future releases of DMU. For now there should be one 0 for each trait to be analyzed, each on a separate line. In this example there is only one trait and therefore only one line with 0 in it.
- **MODEL**.

  7 the seventh real value is y,
  0 no weighting of observations,
  4 number of class variables (fixed and random),
  2 second integer value (RC - Round-classifier subclasses),
  3 third int value (RAS - Round-age-season),
  4 fourth int value (HRC - Herd-round-classifier),
  1 first int value (animal ID).

  The model equation could be written as

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

- **RANDOM**. Indicates how many factors in the model are random factors, and a numbering of those factors. Factors with the same number have a correlation structure between them, while factors with different numbers are independent of each other. The example says there are 2 random factors (assumed to be the last 2 specified in the MODEL, i.e. HRC and ID), and that these are independent, HRC has structure 1 and ID has structure 2, to be defined in a later keyword.

- **REGRES**. For indicating number of covariates in a model with one line per trait. In this example, no covariates are in the model.
- **NOCOV**. Number of covariances among residual effects. In this example, no covariances exist.

$VAR_STR is used to specify the covariance structures of the random variables. If this keyword is omitted, then the assumed variance structure of all random variables is assumed to be $\mathbf{I}\sigma_i^2$. However, in this example, an animal model is being employed, therefore, the structure is $\mathbf{A}\sigma_a^2$. The `ID` variable was coded as number 2 in the **RANDOM** section of the previous keyword, and thus the first 2 in this keyword command. This is followed by `PED` to indicate the relationship matrix is to he used. There are six options for the type of relationship matrix to be constructed.

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 but with phantom groups.

This code is followed by `ASCII` or `BINARY` and the full path and name of the pedigree file.

Others . There are three more keywords which can be used to 1) specify prior or starting values of the variances and covariances; 2) indicate optional input to DMU4; and 3) indicate optional input to DMUAI. These options expect filenames for the files where the additional information is stored. There are fixed formats for the additional information too. These are for fine-tuning an analysis, or possibly for saving information in case of a system malfunction, so that the analysis may be re-started with the same or different information.

Multiple Traits, Maternal Effects

The following driver file is for running two traits, the second trait has a maternal genetic effect.

```
$COMMENT
Example driver file for a MULTIPLE TRAIT, Animal Model
Two traits and including maternal effects for second trait
Traits for analysis are   T1 T2
Fixed factors are          RC RAS
Random Factors for T1     HRC ID
Random factors for T2     HRC DAM ID

$ANALYSE 1 1 0 0

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

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

$MODEL
2
0
0
7 0 4 2 3 4 1
8 0 5 2 3 4 6 1
2 1 2
3 1 2 2
0
0
0

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

This driver file is very similar to that in the previous section. The only differences occur in the $MODEL statements. The model for the first trait is given by 7 0 4 2 3 4 1 which is similar to the first example, and the model for the second trait is 8 0 5 2 3 4 6 1 , where the 6 indicates the DAM ID of the animal. The first model has 2 random factors, coded as 1 and 2 for HRC and ID, respectively. The second model has 3 random factors namely, HRC, DAM, and ID, coded as 1, 2, and 2, respectively. Thus HRC have a covariance between traits, ID has a covariance between traits, and ID and DAM have a covariance between them within trait 2. The structure of the covariances for those coded as 2 is the

additive genetic relationship matrix given in the $VAR_STR keyword.

Single trait model with covariates

```
$COMMENT
Example driver file for a SINGLE TRAIT, Animal Model
With a fixed covariate nested within a fixed factor
Trait for analysis is T1
Fixed factors are     RC
Random Factors are    HRC ID
Covariate is          AGE nested within RC

$ANALYSE 1 1 0 0

$DATA  ASCII (6,9,-99) /u/name/test/dairy.d

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

$MODEL
1
0
7 0 3 2 4 1
2 1 2
1 2(1)
0
# (1) refers to factor 1 of the model (RC)
# and () refer to nested

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

A new variable, AGE, was added to the data, which was the second real variable. Note that comments may be added to the driver file. Any lines starting with # and blank lines are ignored.

Random regression model

```
$COMMENT
Example driver file for a SINGLE TRAIT, Animal Model
Using longitudinal data, random regression model
Trait for analysis is T1
Legendre polynomials are LP1 LP2 LP3 LP4 LP5
Fixed factors are     RC RAS
Random Factors are    HRC ID

$ANALYSE 1 1 0 0

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

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

$MODEL
1
0
7 0 4 2 3 4 1
2 1 2
3 2(4) 3(4) 4(4)
0
0

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

Thus, LP1,LP2, and LP3 are nested within levels of factor 4 of the model which is the animal ID. Because the animal ID factor is designated to have the additive relationship matrix covariance structure, then that structure is also applied to the covariates nested within the animal ID, as well as covariances between the covariates.

6.1.2 Comments on DMU

The driver files are not very complicated to construct, but the user may get confused by the number codes, and the manual will likely need to be utilized every time a new driver file is made, in order to remember what the codes represent. The driver file should have been made to be more explicit in itself. However, if DMU is used often, then the user may memorize the options and coding procedures so that they become second nature.

6.2 VCE - Groeneveld

The control files for VCE are called *parameter files* or *pfiles*. pfiles are organized into sections which must be in a certain order. VCE is keyword oriented, and therefore, the user must not use these keywords to describe data or other variables. There is an online manual available to assist in understanding the use of VCE. VCE is capable of performing AI REML coded as (AG) in methods, Monte Carlo EM REML, coded as (AE), and Bayesian methods via Gibbs Sampling, coded as (GI). VCE is capable of handling dominance genetic effects and random regression models.

6.2.1 pfile

The parameter file for VCE is made up of sections identified by the following keywords in this order.

COMMENT	Optional
DATA	Mandatory
MODEL	Mandatory
COVARIANCE	Mandatory
SYSTEM	Optional
OUTPUT	Optional
END	Optional

Each section will be described separately, and then some examples will be given for particular models. VCE does not distinguish between capital or small letters. Thus, `ANIMAL` and `animal` are the same.

6.2.2 COMMENT

```
COMMENT  job = iowastate
 The job name will appear on every page of the printout.
 An unlimited number of lines can follow.
 This section is for describing the analysis that the
 user is running.
```

Care must be taken not to use another keyword in the comments section.

6.2.3 DATA

There are three types of data sets. An example DATA section follows.

```
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      = variable
    header        = 0
    crossbreeding = .false.
    :
 pedfile       = 'ped.d'
    format        = '(4i10)'
    link          = ANIMAL
    dominance     = fANIMAL
    indep         =
    header        = 0
    ;
 ranfile       = 'rfile.d'
    format        = '(  )'
    link          =
    indep         =
    header        = 0
    :
```

- A `format` is not needed. If omitted, then the default option is free format. That is, a space must be inserted between each variable. If a format is given for the data file, a typical FORTRAN format is used, and all variables must be read with the 'F' indicator even if the variables are integer.
- Note that the records in the data file must be arranged so that the observations or traits appear before any of the classification variables that may be used in the models. Or a format has to be given that uses the 'T' specifier for specific columns in which a variable is located to order the variables. Thus, `dep` refers to traits or dependent variables, and `indep` refers to independent variables.
- The `header` indicates the number of records that should be skipped in the data file before records on animals begin. Usually the header is equal to 0, but if the data file is an Excel file, for example with headings on the columns, then the headings must be skipped. This statement allows the user to skip a certain number of records.
- The `crossbreeding` keyword is used to indicate if crossbred animals appear in the data.

- The `group_by` keyword is to identify a variable in the data such that the residual variance could be different for each level of that variable, such as by herd-year-seasons, for example.
- Note the semicolon after each file description.
- There can be several data files. If the first data file is sorted by `YSH`, for example, then subsequent data files should be sorted in the same sequence.
- The default option for the pedigree file assumes that the file contains animal, sire, and dam. If other information is on the file, such as breed codes, litter codes, or birthdates, then `indep` can be used to specify these variables.
- The `link` keyword is usually equated to `animal` in order to have additive genetic effects in the model.
- Dominance genetic relationships will be calculated if `dominance = fanimal`.
- The `ranfile` is a file that identifies levels of heterogeneous variances for a random factor. An example will be given later.

6.2.4 MODEL

A typical simple specification for four traits, each with a different model is shown below.

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

If `NRF` and `NRC` have the same model, then this could be written as

```
NRF NRC = YS MF AF YSH SS ANIMAL;
```

- The `ANIMAL` effect should not be the first effect in the model equation.
- The `ANIMAL` effect should appear in the model before the `maternal` effect, if included in the data.
- Note that `PE` is the same as `ANIMAL`, because a variable should only appear once in the equation.

- Covariates are designated as `p2(age)`, for example, which means to include age and age-squared.
- To deviate covariates from their means then use `pa2(age)`. This is generally a good idea to avoid rounding problems.
- If the covariates are nested within year for example, then `[1, p2(age)]year` would be used to include an intercept, age, and age-squared within years.
- For a random regression model, this could be specified by `[p5(dim)]animal`, which is a polynomial of dim, (days in milk), to the power 5 nested within animals.
- If the user wanted to use Legendre polynomials, then `[plg5(dim)]animal`.

6.2.5 COVARIANCE

This section is for describing the structure of covariance matrices of the random variables.

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

The `ANIMAL` covariances are applied to all traits, as with `YSH`. Because the pedigree file was given in the `DATA` section and linked to `ANIMAL`, then the additive genetic relationship matrix will be employed. For `YSH`, the assumption is that the covariance structure is diagonal for a single trait and between traits. `SS` and `PE` are applied only to two traits each. There are options for specifying heterogeneous variances. The `residual` covariance structure, if not stated, is assumed to be diagonal within and between traits.

Starting values for parameters can be determined by the program. Covariances are initially assumed to be close to zero. However, the user may specify starting values in a file, or values from a previous analysis may be used as input. Please consult the manual for details.

6.2.6 SYSTEM

This section describes the analysis that the user wishes to conduct, and how to proceed. There are keywords for iterative methods, such as for Gibbs sampling. There are keywords for constructing the additive genetic relationship matrix, and there are keywords for indicating missing data or observations that the user might want to skip. Some examples, but not all possibilities are shown below.

```
SYSTEM
  method        = 'GI'
  mc_seed       = 56437281
  burn_max      = 5000
  burn_stop     = .0001
  restart       = .true.

  inbreeding    = .true.

  missing_value = -99
  non_zero      = 1000000
```

non_zero is the number of nonzero elements in the MME. The other variables should be obvious. The burn-in period would be limited to 5000 samples or to the level of 0.0001 difference between the two chains. With Gibbs sampling, VCE uses two chains, with different starting parameters, but the same random number sequence. When the agreement between the two chains reaches the minimum desired, then burn-in has been achieved and only samples from this point are used as samples from the joint posterior distribution.

6.2.7 OUTPUT

This section is for creating the desired output files.

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

Thus, the covfile will contain the estimated parameters from each Gibbs sample, after burn-in. Inbreeding coefficients of each animal can be saved. There are other options for getting exactly what you want or need. Please consult the manual.

6.2.8 END

The END statement terminates input to the program. Any lines that come after this are totally ignored. Thus, the user may put additional comments about the analysis after this statement.

The manual for VCE Version 5.1 from December 2003 contains a lot of information, but some of the explanations are not very clear. Thus, the manual can be difficult to

follow and to get what is really needed. Assistance from the authors is not always readily at hand. However, VCE seems to be a very useful program if you understand the options and features. There is slightly more flexibility in modeling than in DMU. No comparisons have been made as to speed or memory requirements.

6.3 MTDFREML

These notes are based on the latest documentation which was April 1995. The authors were K. Boldman, L. Kriese, L. D. Van Vleck, C. P. Van Tassell, and S. D. Kachman. The programs were originally designed to run on personal computers interactively. The user would answer questions interactively with the program. If one of the answers was unacceptable or in error, then the program would abort and the user would need to begin again. With over 30 questions to be answered, this method of providing information to the program was slow, tedious, and subject to errors. The user can now create a file that has the answers to those questions. The MTDFREML package uses the Simplex method to locate the parameters that maximize the log likelihood function. The package also utilizes SPARSPAK routines for sparse matrix manipulations. The package consists of several programs. MTDFNRM is used to prepare the additive genetic relationship matrix inverse and recode animals. MTDFPREP uses output from MTDFNRM to recode the data file, recode levels of fixed and random factors and to create new data files for the next program, including information on the dimensions of vectors and arrays. MTDFRUN is the program that searches for the maximum of the log likelihood function.

MTDFNRM prepares 4 files; MTDFPREP prepares another 6 files, and MTDFRUN creates 10 files, of which 4 are necessary to continue from the last search. All pedigree and data input files from the user are assumed to be readable in free format.

6.3.1 MTDFNRM

This program computes the inverse of the additive genetic numerator relationship matrix, $\mathbf{A}^{-1}$. The information required is given in the table below.

1	Animal model(0) or Sire-MGS model(1)	0
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	6
7	Which one is Animal ID	1
8	Which one is Sire ID	2
9	Which one is Dam ID	3
10	Number of genetic groups	0

The following data files are created.

MTDF11	Number of animals, list of animals by original ID with recoded ID, in ascending order.
MTDF13	Recoded pedigree, original pedigree, and inbreeding coefficients.
MTDF44	Value of 0.5 ln $\mid \mathbf{A} \mid$, followed by non zero inverse elements in binary format.
MTDF56	Summary information from the run.

6.3.2 MTDFPREP

The data file must be prepared as for DMU with integer variables first followed by real variables. A PARAM.DAT file must be prepared and included during the compilation of the program. The file indicates the maximum values possible for this particular analysis. The variables that can be changed in that file are as follows:

MAXTRT	Number of traits
MAXINTR	Number of integer variables
MAXR8	Number of real variables
MAXANIM	Number of animals(with records and 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 inverted
MAXORDS	Used by SPARSPAK
MAXNZE	Used by FSPAK
NHASH	Used by FSPAK for Hashing function

The following information is needed, in this order.

1			Name of data file
2			Up to 6 lines, description of analysis, ended by in column 1 after last comment line
3			Number of integer variables
4			Number of real variables
5			Number of traits
6			For each trait the following must be known
6	a		Name of trait
6	b		Position of trait in list of real variables
6	c		Missing value designation
6	d		Number of covariates in model for this trait
6	e		For each covariate with this trait
6	e	1	Name of covariate
6	e	2	Position of covariate in real variables
6	e	3	Type of covariate, linear, quadratic,
6	f		Number of fixed effects in model for this trait
6	g		For each fixed effect
6	g	1	Name of fixed effect
6	g	2	Position of fixed effect
6	g	3	Write levels to MTDF66? 1=yes
6	h		Position of animal ID in list of integers
6	i		Number of animals in relationship matrix
6	j		Is there a maternal effect? 1 = yes
6	j	1	If yes, Name of effect, i.e. maternal
6	j	2	If yes, Position of Maternal ID
6	k		Number of uncorrelated random effects
6	m		For each random effect
6	m	1	Name of random effect
6	m	2	Position of random effect in integer list
6	m	3	Write levels to MTDF66? 1=yes
7			Save labels for covariates and fixed effects? 1=yes
8			Save labels for random effects? 1=yes

Six data files are created by this program that are used as input to the analysis program. The same file names are generated each time the program is run. Therefore, information from the previous run will be lost or overwritten. Re-name the files if you think they might be used again at a later time before running this program.

6.3.3 MTDFRUN

Several options are possible.

TYPE OF ANALYSIS . Enter 0 if this is a completely new run. Enter 1 if this is a

continuation of a previous run.

OPTIONS .

1. Estimate variance components.
2. Solve MME only.
3. Calculate sampling variances only.
4. Solve MME, then sampling variances.

Beginning a New Analysis

Because SPARSPAK is used in MTDFRUN, the MME must be full rank. Only the input needed for OPTION 1, estimation of variance components, will be presented here. There are different inputs to the other OPTIONS. The program will ask the following questions.

1		Up to 6 lines to describe the analysis with in the first column after the last comment line.
2		Is this a continuation? yes=1
3		OPTION number = 1
4		Number of constraints, put 0
5		Does file MTDF58 exist? yes=1
6		Input of prior values for animal (and maternal) genetic effects according to model
7		Verification of the priors.
8		Number of parameters to hold constant.
8	a	Position of parameters to hold constant.
9		Input of prior values for other random factors.
10		Verification of the priors.
11		Number of parameters to hold constant.
11	a	Position of parameters to hold constant.
12		Input of priors for residual covariances.
13		Verification of the priors.
14		Number of parameters to hold constant.
14	a	Position of parameters to hold constant.
15		Write solutions for covariates and fixed effects? 1=yes, 2=no.
16		If yes, Merge labels with solutions? 1=yes, 2=no.
17		Write animal (and maternal) solutions? 1=yes, 2=no.
18		Write solutions to other random effects? 1=yes, 2=no.
19		If yes, Merge labels with solutions? 1=yes, 2=no.
20		Convergence criterion, 1.e-6 or 1.e-8.
21		Number of Simplex rounds.

The number of Simplex rounds should be at least one greater than the number of parameters to be estimated. Depending on the number of parameters and complexity of

the model, the number of rounds required could be 1000 or more. If the user enters a 1 for number of Simplex rounds, then the program will give a timing per round, from which the user can determine the number of rounds to perform. However, if the method reaches convergence before the specified number of rounds, then the program will terminate.

If the user wants to restart, then the questions are the same except that the input of the priors is skipped (items 4 through 14). The manual is filled with examples for different kinds of models which are useful to understand the program.

Chapter 7

Maternal Effects Models

7.1 Introduction

In some species of livestock, such as beef cattle, sheep or swine, the female provides an environment for its offspring to survive and grow. Females vary in their ability to provide a good environment for their offspring, and this variability has a genetic basis. The offspring inherit directly an ability to grow (or survive) from both parents, and environmentally do better or poorer depending on their dam's maternal ability. Maternal ability is a genetic trait and is transmitted, as usual, from both parents, but maternal ability is only expressed by females when they have a calf (i.e. much like milk yield in dairy cows).

A model to account for maternal ability is

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

where $\mathbf{y}$ is the growth trait of a young animal, $\mathbf{b}$ is a vector of fixed factors influencing growth, such as contemporary group, sex of the offspring, or age of dam, $\mathbf{a}$ is a vector of random additive genetic effects (i.e. direct genetic effects) of the animals, $\mathbf{m}$ is a vector of random maternal genetic (dam) effects, and $\mathbf{p}$, in this model, is a vector of maternal permanent environmental effects (because dams may have more than one offspring in the data).

The expectations of the random vectors, $\mathbf{a}$, $\mathbf{m}$, $\mathbf{p}$, and $\mathbf{e}$ are all null vectors in a model without selection, and the variance-covariance structure is

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

where σ_a^2 is the additive genetic variance, σ_m^2 is the maternal genetic variance, σ_{am} is the additive genetic by maternal genetic covariance, and σ_p^2 is the maternal permanent

environmental variance. Also,

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

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

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

In this model, a female animal, i, could have its own growth record for estimating $\hat{a}_i$. The same female could later have offspring of its own for estimating $\hat{m}_i$ and $\hat{p}_i$, and the offspring would also contribute towards $\hat{a}_i$. The maternal effects model can be more complicated if, for example, embryo transfer is practiced. Recipient dams would have maternal effects, but would not have direct genetic effects on that calf, see Schaeffer and Kennedy (1989).

7.2 Simulation of Records

To better understand this model, think about how records might be sampled in reality. For example, let

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

Any positive definite matrix can be partitioned into the product of a matrix times its transpose (i.e. Cholesky decomposition), or

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

Let $\sigma_p^2 = 9$ and $\sigma_e^2 = 81$. Both the additive genetic and maternal genetic effects need to be sampled simultaneously because these effects are genetically correlated.

Consider three animals, A, B, and C, where C is an offspring of sire A and dam B. First, sample additive genetic values for the parents, and then generate the additive genetic effects of their progeny, animal C. For animal A, generate a vector of two random

normal deviates from a distribution with mean zero and variance unity, which will be pre-multiplied by $\mathbf{L}$. Animals A and B are base population animals unrelated to each other. Let the vector of random normal deviates be $\mathbf{w}' = (2.533 \quad -.299)$, then for animal A

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

Similarly for animal B, generate another vector of random normal deviates, say $\mathbf{w}' = (-1.141 \quad .235)$, then

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

The additive genetic value of animal C take the average of the parents' true breeding values and add a random Mendelian sampling term. Generate another vector or random normal deviates, $\mathbf{w}' = (.275 \quad .402)$, then

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

where b_{ii} comes from factoring the additive genetic relationship matrix as $\mathbf{A} = \mathbf{TBT}'$, for $\mathbf{T}$ being lower triangular and $\mathbf{B}$ being a diagonal matrix. Hence b_{ii} is a diagonal element of $\mathbf{B}$. For non-inbred animals with both parents known then $b_{ii} = 0.5$, otherwise it can be less than 0.5, depending on the amount of inbreeding.

Maternal permanent environmental effects must be generated for each dam with progeny. In this case only animal B is a dam. Multiply a new random normal deviate by $\sigma_p = 3$, suppose the result is -4.491. An record for animal C is created by following the model equation,

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

where μ was arbitrarily set to 140, and RND refers to a new random normal deviate. The record on animal C consists of the direct genetic effect of animal C plus the maternal

genetic effect of the dam (B) plus the maternal permanent environmental effect of the dam (B) plus a residual. Records in real life are probably measured to the nearest whole unit, so $y = 154$ for animal C.

7.3 Mixed Model Equations

To illustrate the mixed model equations, assume the data as given in the table below.

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

where CG stands for contemporary group, the only fixed effect in this example. Assume that the appropriate variauce parameters were those used in the simulation in the previous section. Based on the matrix formulation of the model, the MME are

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_1 & \mathbf{X}'\mathbf{Z}_2 & \mathbf{X}'\mathbf{Z}_3 \\ \mathbf{Z}_1'\mathbf{X} & \mathbf{Z}_1'\mathbf{Z}_1 + \mathbf{A}^{-1}k_{11} & \mathbf{Z}_1'\mathbf{Z}_2 + \mathbf{A}^{-1}k_{12} & \mathbf{Z}_1'\mathbf{Z}_3 \\ \mathbf{Z}_2'\mathbf{X} & \mathbf{Z}_2'\mathbf{Z}_1 + \mathbf{A}^{-1}k_{12} & \mathbf{Z}_2'\mathbf{Z}_2 + \mathbf{A}^{-1}k_{22} & \mathbf{Z}_2'\mathbf{Z}_3 \\ \mathbf{Z}_3'\mathbf{X} & \mathbf{Z}_3'\mathbf{Z}_1 & \mathbf{Z}_3'\mathbf{Z}_2 & \mathbf{Z}_3'\mathbf{Z}_3 + \mathbf{I}k_{33} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{m}} \\ \hat{\mathbf{p}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}_1'\mathbf{y} \\ \mathbf{Z}_2'\mathbf{y} \\ \mathbf{Z}_3'\mathbf{y} \end{pmatrix},$$

where

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

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

The matrices for the example data are

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{X}'\mathbf{y} = \begin{pmatrix} 415 \\ 450 \end{pmatrix},$$

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

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

$$\mathbf{Z}_3 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{Z}_3'\mathbf{y} = \begin{pmatrix} 429 \\ 436 \end{pmatrix}.$$

The other two right hand side matrices can be easily obtained from $\mathbf{y}$ and $\mathbf{Z}_3'\mathbf{y}$. The order of the MME is 24. The inverse of the relationship matrix is

$$\mathbf{A}^{-1} = \frac{1}{2}\begin{pmatrix} 5 & 0 & 2 & 1 & -2 & 0 & -2 & 0 & -2 & 0 \\ 0 & 5 & 1 & 2 & 0 & -2 & 0 & -2 & 0 & -2 \\ 2 & 1 & 5 & 0 & -2 & -2 & 0 & 0 & -2 & 0 \\ 1 & 2 & 0 & 5 & 0 & 0 & -2 & -2 & 0 & -2 \\ -2 & 0 & -2 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & -2 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & -2 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & -2 & 0 & -2 & 0 & 0 & 0 & 4 & 0 & 0 \\ -2 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & -2 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 4 \end{pmatrix}.$$

The solutions to the MME are

$$\hat{\mathbf{b}} = \begin{pmatrix} 137.8469 \\ 150.4864 \end{pmatrix}, \quad \hat{\mathbf{p}} = \begin{pmatrix} .0658 \\ -.0658 \end{pmatrix},$$

$$\hat{\mathbf{a}} = \begin{pmatrix} 2.3295 \\ -2.3295 \\ .1280 \\ -.1280 \\ 5.1055 \\ -4.1143 \\ .2375 \\ 2.0161 \\ .5447 \\ -3.7896 \end{pmatrix}, \text{ and } \hat{\mathbf{m}} = \begin{pmatrix} -.3328 \\ .3328 \\ .1646 \\ -.1646 \\ -.6379 \\ .6792 \\ -.1254 \\ -.3795 \\ .0136 \\ .4499 \end{pmatrix}.$$

7.4 Estimation of Covariances

Let the inverse of the coefficient matrix of the MME be represented as

$$\begin{pmatrix} \mathbf{C}_{xx} & \mathbf{C}_{x1} & \mathbf{C}_{x2} & \mathbf{C}_{x3} \\ \mathbf{C}_{1x} & \mathbf{C}_{11} & \mathbf{C}_{12} & \mathbf{C}_{13} \\ \mathbf{C}_{2x} & \mathbf{C}_{21} & \mathbf{C}_{22} & \mathbf{C}_{23} \\ \mathbf{C}_{3x} & \mathbf{C}_{31} & \mathbf{C}_{32} & \mathbf{C}_{33} \end{pmatrix}.$$

The quadratic forms required for REML or Bayesian estimation are

$$\begin{aligned} \begin{pmatrix} \hat{\mathbf{a}}' \\ \hat{\mathbf{m}}' \end{pmatrix} \mathbf{A}^{-1} \begin{pmatrix} \hat{\mathbf{a}} & \hat{\mathbf{m}} \end{pmatrix} &= \begin{pmatrix} \hat{\mathbf{a}}'\mathbf{A}^{-1}\hat{\mathbf{a}} & \hat{\mathbf{a}}'\mathbf{A}^{-1}\hat{\mathbf{m}} \\ \hat{\mathbf{m}}'\mathbf{A}^{-1}\hat{\mathbf{a}} & \hat{\mathbf{m}}'\mathbf{A}^{-1}\hat{\mathbf{m}} \end{pmatrix}, \\ &= \begin{pmatrix} 95.7075 & -13.6257 \\ -13.6257 & 2.0067 \end{pmatrix}, \\ \hat{\mathbf{p}}'\hat{\mathbf{p}} &= 0.008668, \\ \hat{\mathbf{e}}'\hat{\mathbf{e}} &= 463.56943. \end{aligned}$$

7.4.1 EM-REML

For EM-REML estimation,

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

where $\mathbf{W} = (\mathbf{X}\ \mathbf{Z}_1\ \mathbf{Z}_2\ \mathbf{Z}_3)$ and $\beta' = (\mathbf{b}'\ \mathbf{a}'\ \mathbf{m}'\ \mathbf{p}')$. In the example, $N = 6$, $r(\mathbf{X}) = 2$, and

$$\hat{\sigma}_e = (125,751\ -\ 125,128.93)/4\ =\ 155.51863.$$

The genetic components are given by

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

For the example, these are

$$\begin{aligned} \begin{pmatrix} \hat{\mathbf{a}}'\mathbf{A}^{-1}\hat{\mathbf{a}} & \hat{\mathbf{a}}'\mathbf{A}^{-1}\hat{\mathbf{m}} \\ \hat{\mathbf{m}}'\mathbf{A}^{-1}\hat{\mathbf{a}} & \hat{\mathbf{m}}'\mathbf{A}^{-1}\hat{\mathbf{m}} \end{pmatrix} &= \begin{pmatrix} 95.7075 & -13.6257 \\ -13.6257 & 2.0067 \end{pmatrix}, \\ \begin{pmatrix} tr(\mathbf{A}^{-1}\mathbf{C}_{11}) & tr(\mathbf{A}^{-1}\mathbf{C}_{12}) \\ tr(\mathbf{A}^{-1}\mathbf{C}_{21}) & tr(\mathbf{A}^{-1}\mathbf{C}_{22}) \end{pmatrix} &= \begin{pmatrix} 5.4036 & -0.8391 \\ -0.8391 & 3.1199 \end{pmatrix}, \\ \begin{pmatrix} \hat{\sigma}_a^2 & \hat{\sigma}_{am} \\ \hat{\sigma}_{am} & \hat{\sigma}_m^2 \end{pmatrix} &= \begin{pmatrix} 93.6068 & -14.4124 \\ -14.4124 & 48.7213 \end{pmatrix}. \end{aligned}$$

Finally, the estimate of the maternal permanent environmental component would be

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

where $\hat{\mathbf{p}}'\hat{\mathbf{p}} = 0.0086681$, $tr\mathbf{C}_{33} = 0.2097853$, and $\hat{\sigma}_p^2 = 3.2634$. The new estimates should be used to reconstruct the MME to get new solutions and new estimates of variances and covariances, and this is repeated until convergence is achieved.

7.4.2 Bayesian Estimation

For Bayesian estimation using Gibbs sampling, the MME would be solved and new samples generated for $\beta' = (\mathbf{b}'\ \mathbf{a}'\ \mathbf{m}'\ \mathbf{p}')$ as usual. For the genetics components, the necessary quadratic forms for the i^{th} sample are given by

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

This matrix follows an inverted Wishart distribution. To sample a new $\mathbf{G}$,

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 $\mathbf{G}_{i+1}$.

For the residual and maternal permanent environmental effects, a new sample value for the residual variance is given by

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

and a new sample value for the maternal permanent environmental variance is given by

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

where p is the number of dams with progeny. Many samples may need to be drawn. The $\hat{e}$ and $\hat{p}$ are not estimates in the usual sense, but represent the current sample values in the Gibbs sampling chain.

7.5 Warnings

The influence of the correlation between direct and maternal effects on the relationship between solutions for direct and maternal effects can be a matter of concern. If the genetic correlation between direct and maternal true breeding values is negative (-0.1), for example, and if an animal has a high, positive direct EBV based on its own growth record, then the maternal EBV could be very negative due to the correlation alone. Thus, if few of the animals with growth records have progeny, then the relationship between direct and maternal EBVs will be strongly negative (like -0.8)(reflecting the assumed negative

correlation amongst true breeding values). However, if the data are complete and animals have both their own records and those of several progeny, then the correlation between direct and maternal EBVs should more closely follow the assumed genetic correlation.

The data structure can affect the correct estimation of the genetic correlation between direct and maternal effects. Estimates of this correlation in beef cattle have ranged from -0.5 to +0.5, and this mostly reflects the differences in quality (completeness) of data used. In experimental station herds with several generations of animals and fairly complete data on pedigrees, the estimates of the genetic correlation have tended to be zero or slightly positive between direct and maternal effects. On the other hand, in field data with almost no ties between growth of calves with performance of offspring as a dam, the estimates of the genetic correlation have tended to be negative. To determine if your data are complete, create a file that has an animal's own record plus the average growth record of its dam. If you have 3 million records, but only 100 dam-offspring pairs, then the reliability of the estimated correlation between direct and maternal effects will be low. One can also look at the number of female progeny of each sire that have their own progeny as a percentage of all female progeny. If that percentage is low (i.e. less than 20%), then the reliability of the estimated genetic correlation could be low. With poor data structure, the possibility of a strong negative genetic correlation is very likely if the estimation process is started with a negative genetic correlation.

Chapter 8

Random Regression Models

8.1 Introduction

All livestock grow and perform over their lifetime. Traits that are measured at various times during that life are known as *longitudinal* data. Examples are body weights, body lengths, milk production, feed intake, fat deposition, and egg production. On a biological basis there could be different genes that turn on or turn off as an animal ages causing changes in physiology and performance. Also, an animal's age can be recorded in years, months, weeks, days, hours, minutes, or seconds, so that, in effect, there could be a continuum or continuous range of points in time when an animal could be observed for a trait. These traits have also been called *infinitely dimensional* traits.

Take body weight on gilts over 60 days on test as an example.

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

The differences among the three animals increase with days on test as the gilts become heavier. As the mean weight increases, so also the standard deviation of weights increases. The weights over time could be modeled as a mean plus covariates of days on test and days on test squared. Depending on the species and trait, perhaps a cubic or spline function would fit the data better. The point is that the means can be fit by a linear model with a certain number of parameters.

8.2 Multiple Trait Approach

The data presented in the previous table have typically been analyzed such that the weights at each day on test are different traits. If t is the day on test, i.e. 10, 20, 30, 40, 50, or 60, then a model for any one of the weights could be

$$\mathbf{y}_t = \mathbf{X}\mathbf{b}_t + \mathbf{a}_t + \mathbf{e}_t,$$

which is just a simple, single record, animal model. Analyses are usually done so that the genetic and residual variances and covariances are estimated among the six weights. Suppose that an estimate of the genetic variances and covariances was

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

Let the residual covariance matrix be

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

Assuming a model with only an intercept, and that the three animals are unrelated, then

$$(\mathbf{X} \;\; \mathbf{Z}) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \otimes \mathbf{I}_6,$$

where the identity is of order 6 and $\otimes$ is the direct product operator. The observations would be ordered by days on test within animals, i.e.,

$$\mathbf{y}' = \begin{pmatrix} 42 & 53 & 60 & 72 & 83 & 94 & \cdots & 60 & 70 & 77 \end{pmatrix}.$$

The resulting MME would be of order 24 by 24, and the solutions would be as follows.

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

Animal 1 clearly grew faster than the other two animals and its superiority grew larger with time. Animals 2 and 3 switched rankings after the first 10 days, and Animal 3 was clearly the slower growing animal. The estimates for the mean give an average growth curve for the 3 animals.

A multiple trait approach may be appropriate here because every animal was weighed on exactly the same number of days on test throughout the trial. However, suppose the animals were of different ages at the start of test, and suppose that instead of days on test, the ages for each weight were given. Assume at start of test that Animal 1 was 18 days old, Animal 2 was 22, and Animal 3 was 25. The multiple trait model could include a factor (classification or covariable) to account for different starting ages. The differences observed at any point in time could be due to the ages of the animals rather than just on the number of days on test. The analysis shown above would have an implied assumption that all animals began the test at the same age.

8.3 Covariance Functions and Orthogonal Polynomials

Let the example data be as shown below, allowing for the different ages at each test. Note that the ages range from 28 days to 85 days, and that none of the animals were ever weighed at exactly the same age.

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

Kirkpatrick et al.(1991) proposed the use of covariance functions for longitudinal data of this kind. A covariance function (CF) is a way to model the variances and covariances of a longitudinal trait. Orthogonal polynomials are used in this model and the user must decide the order of fit that is best. Legendre polynomials are the easiest to apply. They were first published around 1797.

To calculate Legendre polynomials, first define

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

then, in general, the $n+1$ polynomial is described by the following recursive equation:

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

These quantities are "normalized" using

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

This gives the following series,

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

and so on. The first six can be put into a matrix, Λ, as

$$\Lambda' = \begin{pmatrix}
.7071 & 0 & 0 & 0 & 0 & 0 \\
0 & 1.2247 & 0 & 0 & 0 & 0 \\
-.7906 & 0 & 2.3717 & 0 & 0 & 0 \\
0 & -2.8062 & 0 & 4.6771 & 0 & 0 \\
.7955 & 0 & -7.9550 & 0 & 9.2808 & 0 \\
0 & 4.3973 & 0 & -20.5206 & 0 & 18.4685
\end{pmatrix}.$$

Now define another matrix, **M**, as a matrix containing the polynomials of standardized time values. Legendre polynomials are defined within the range of values from -1 to +1. Thus, ages or time periods have to be standardized (converted) to the interval between -1 to +1. The formula is

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

Let the minimum starting age for pigs on test be 15 days and the maximum starting age be 28 days, then the maximum age at end of test was 88 days. Thus, $t_{min} = 25 = (15+10)$ and $t_{max} = 88 = (28 + 60)$.

The matrix **G** was based on weights taken on pigs that were all 21 days of age at start of test. The table below shows the ages and standardized time values for the six weigh dates.

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

Therefore,

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

This gives

$$\begin{aligned} \Phi &= \mathbf{M}\Lambda, \\ &= \begin{pmatrix} .7071 & -1.2247 & 1.5811 & -1.8708 & 2.1213 & -2.3452 \\ .7071 & -.7348 & .0632 & .6735 & -.8655 & .3580 \\ .7071 & -.2449 & -.6957 & .5238 & .4921 & -.7212 \\ .7071 & .2449 & -.6957 & -.5238 & .4921 & .7212 \\ .7071 & .7348 & .0632 & -.6735 & -.8655 & -.3580 \\ .7071 & 1.2247 & 1.5811 & 1.8708 & 2.1213 & 2.3452 \end{pmatrix}. \end{aligned}$$

which can be used to specify the elements of $\mathbf{G}$ as

$$\begin{aligned} \mathbf{G} &= \Phi\mathbf{H}\Phi' \\ &= \mathbf{M}(\Lambda\mathbf{H}\Lambda')\mathbf{M}' \\ &= \mathbf{M}\mathbf{T}\mathbf{M}'. \end{aligned}$$

Note that Φ, $\mathbf{M}$, and Λ are matrices defined by the Legendre polynomial functions and by the standardized time values and do not depend on the data or values in the matrix $\mathbf{G}$. Therefore, it is possible to estimate either $\mathbf{H}$ or $\mathbf{T}$,

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

and

$$\mathbf{T} = \mathbf{M}^{-1}\mathbf{G}\mathbf{M}^{-T}$$

$$= \begin{pmatrix} 16.44 & 6.48 & -5.93 & -11.49 & -0.93 & 10.02 \\ 6.48 & 49.87 & -2.05 & -155.34 & 1.44 & 111.23 \\ -5.93 & -2.05 & 57.71 & 28.62 & -50.06 & -25.73 \\ -11.49 & -155.34 & 28.62 & 635.49 & -26.91 & -486.90 \\ -0.93 & 1.44 & -50.06 & -26.91 & 49.80 & 26.49 \\ 10.02 & 111.23 & -25.73 & -486.90 & 26.49 & 382.79 \end{pmatrix}.$$

Why orthogonal polynomials? Convert **T** and **H** to correlation matrices.

$$\mathbf{T}_{cor} = \begin{pmatrix} 1.00 & .23 & -.19 & -.11 & -.03 & .13 \\ .23 & 1.00 & -.04 & -.87 & .03 & .81 \\ -.19 & -.04 & 1.00 & .15 & -.93 & -.17 \\ -.11 & -.87 & .15 & 1.00 & -.15 & -.99 \\ -.03 & .03 & -.93 & -.15 & 1.00 & .19 \\ .13 & .81 & -.17 & -.99 & .19 & 1.00 \end{pmatrix},$$

and

$$\mathbf{H}_{cor} = \begin{pmatrix} 1.00 & .45 & -.30 & .01 & -.29 & .09 \\ .45 & 1.00 & .15 & -.10 & -.17 & -.32 \\ -.30 & .15 & 1.00 & .15 & -.36 & -.05 \\ .01 & -.10 & .15 & 1.00 & .07 & -.62 \\ -.29 & -.17 & -.36 & .07 & 1.00 & .19 \\ .09 & -.32 & -.05 & -.62 & .19 & 1.00 \end{pmatrix}.$$

The largest absolute correlation in **T** was .99, while the largest absolute correlation in **H** was only .62. Orthogonal polynomials tend to reduce the correlations between estimated regression coefficients. This is advantageous when trying to estimate **H** by REML or Bayesian methods, because the estimates would converge faster to the maximum or appropriate posterior distribution than trying to estimate **T**. The matrix **T** actually had four correlations greater than 0.80 in absolute value, while **H** had none. There are other kinds of orthogonal polynomials, but Legendre polynomials are probably the easiest to calculate and utilize.

H can be used to calculate the covariance between any two days on test between 10 and 60 days. To compute the covariance between days 25 and 55, calculate the Legendre polynomial covariates as in calculating a row of Φ. The standardized time values for days 25 and 55 are -0.4 and 0.8, respectively. The Legendre polynomials (stored in **L** are

$$\mathbf{L} = \begin{pmatrix} .7071 & -.4899 & -.4111 & .8232 & -.2397 & -.6347 \\ .7071 & .9798 & .7273 & .1497 & -.4943 & -.9370 \end{pmatrix}.$$

Then the variances and covariance for those two ages are

$$\mathbf{LHL}' = \begin{pmatrix} 14.4226 & 13.7370 \\ 13.7370 & 28.9395 \end{pmatrix}.$$

Thus, the genetic correlation between days 25 and 55 is 0.67. The same calculations could be repeated for the residual variance-covariance matrix. Let

$$\mathbf{S} = \Phi^{-1}\mathbf{R}\Phi^{-T},$$

$$= \begin{pmatrix} 41.57 & 7.94 & -2.91 & 0.11 & -1.76 & 0.76 \\ 7.94 & 7.45 & 0.62 & -0.41 & -0.44 & -1.07 \\ -2.91 & 0.62 & 2.29 & 0.31 & -0.52 & -0.12 \\ 0.11 & -0.41 & 0.31 & 1.76 & 0.08 & -1.04 \\ -1.76 & -0.44 & -0.52 & 0.08 & 0.88 & 0.24 \\ 0.76 & -1.07 & -0.12 & -1.04 & 0.24 & 1.64 \end{pmatrix},$$

then the residual variances and covariances for days 25 and 55 would be

$$\mathbf{LSL}' = \begin{pmatrix} 21.6645 & 20.6166 \\ 20.6166 & 43.3442 \end{pmatrix}.$$

8.3.1 Reduced Orders of Fit

Although the order of $\mathbf{G}$ in the previous example was six and polynomials of standardized ages to the fifth power were used to derive the covariance functions, perhaps only squared or cubed powers are needed to adequately describe the elements of $\mathbf{G}$. That is, find Φ^* such that it is rectangular and $\mathbf{H}^*$ has a smaller order, $m < k$, but still

$$\mathbf{G} = \Phi^* \mathbf{H}^* \Phi^{*'}.$$

To determine $\mathbf{H}^*$, first pre-multiply $\mathbf{G}$ by $\Phi^{*'}$ and post-multiply that by Φ^* as

$$\begin{aligned} \Phi^{*'}\mathbf{G}\Phi^* &= \Phi^{*'}(\Phi^*\mathbf{H}^*\Phi^{*'})\Phi^* \\ &= (\Phi^{*'}\Phi^*)\mathbf{H}^*(\Phi^{*'}\Phi^*). \end{aligned}$$

Now pre- and post- multiply by the inverse of $(\Phi^{*'}\Phi^*) = \mathbf{P}$ to determine $\mathbf{H}^*$,

$$\mathbf{H}^* = \mathbf{P}^{-1}\Phi^{*'}\mathbf{G}\Phi^*\mathbf{P}^{-1}.$$

To illustrate, let $m = 3$, then

$$\Phi^* = \begin{pmatrix} .7071 & -1.2247 & 1.5811 \\ .7071 & -.7348 & .0632 \\ .7071 & -.2449 & -.6957 \\ .7071 & .2449 & -.6957 \\ .7071 & .7348 & .0632 \\ .7071 & 1.2247 & 1.5811 \end{pmatrix},$$

and

$$\begin{aligned} \Phi^{*'}\Phi^* &= \begin{pmatrix} 3.0000 & 0.0000 & 1.3415 \\ 0.0000 & 4.1997 & 0.0000 \\ 1.3415 & 0.0000 & 5.9758 \end{pmatrix}, \\ (\Phi^{*'}\Phi^*)^{-1} &= \begin{pmatrix} .3705 & .0000 & -.0832 \\ .0000 & .2381 & .0000 \\ -.0832 & .0000 & .1860 \end{pmatrix}. \end{aligned}$$

Also,

$$\Phi^{*'}\mathbf{G}\Phi^{*} = \begin{pmatrix} 220.2958 & 78.0080 & 61.4449 \\ 78.0080 & 67.5670 & 44.9707 \\ 61.4449 & 44.9707 & 50.5819 \end{pmatrix}.$$

The matrix $\mathbf{H}^{*}$ is then

$$\begin{pmatrix} 26.8082 & 5.9919 & -2.9122 \\ 5.9919 & 3.8309 & .4468 \\ -2.9122 & .4468 & 1.3730 \end{pmatrix}.$$

What order of reduced fit is sufficient to explain the variances and covariances in **G**? Kirkpatrick et al.(1990) suggested looking at the eigenvalues of the matrix **H** from a full rank fit. Below are the values. The sum of all the eigenvalues was , and also shown is the percentage of that total.

H	
Eigenvalue	Percentage
29.0357	.7831
4.2922	.1158
1.8161	.0490
1.3558	.0366
.5445	.0147
.0355	.0010

The majority of change in elements in **G** is explained by a constant, and by a linear increment. Both suggest that a quadratic function of the polynomials is probably sufficient. Is there a way to statistically test the reduced orders of fit to determine which is sufficient? A goodness of fit statistic is $\hat{\mathbf{e}}'\hat{\mathbf{e}}$ where

$$\hat{\mathbf{e}} = \mathbf{g} - \hat{\mathbf{g}}$$

and $\mathbf{g}$ is a vector of the half-stored elements of the matrix **G**, i.e.,

$$\mathbf{g}' = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{16} & g_{22} & \cdots & g_{66} \end{pmatrix}.$$

A half-stored matrix of order k has $k(k+1)/2$ elements. For $k = 6$ there are 21 values. Likewise, $\hat{\mathbf{g}}$ is a vector of half stored elements of the matrix $\Phi^{*}\mathbf{H}^{*}\Phi^{*'}$. Although this matrix also has 21 values, because **M** has only $m < k$ columns, the number of independent values is $m(m+1)/2$. For $m = 3$ this number is 6.

The test statistic, $\hat{\mathbf{e}}'\hat{\mathbf{e}}$, has a Chi-square distribution with $k(k+1)/2 - m(m+1)/2$ degrees of freedom. In the example with $m = 3$,

$$\Phi^{*}\mathbf{H}^{*}\Phi^{*'} = \begin{pmatrix} 3.9622 & 4.7467 & 5.2006 & 5.3239 & 5.1165 & 4.5786 \\ 4.7467 & 8.9493 & 11.4058 & 12.1162 & 11.0804 & 8.2986 \\ 5.2006 & 11.4058 & 15.2402 & 16.7038 & 15.7966 & 12.5186 \\ 5.3239 & 12.1162 & 16.7038 & 19.0868 & 19.2650 & 17.2386 \\ 5.1165 & 11.0804 & 15.7966 & 19.2650 & 21.4857 & 22.4586 \\ 4.5786 & 8.2986 & 12.5186 & 17.2386 & 22.4586 & 28.1786 \end{pmatrix},$$

and the residuals (differences from the original **G**) are

$$\begin{pmatrix} -1.4622 & .1533 & -.6006 & -.7239 & -.8165 & -.5786 \\ .1533 & 4.5507 & .6942 & .1838 & .8196 & 2.4014 \\ -.6006 & .6942 & -.0402 & -2.2038 & -1.1966 & -.0186 \\ -.7239 & .1838 & -2.2038 & .9132 & -.2650 & -.3386 \\ -.8165 & .8196 & -1.1966 & -.2650 & 3.5143 & -2.1586 \\ -.5786 & 2.4014 & -.0186 & -.3386 & -2.1586 & 1.8214 \end{pmatrix},$$

so that the goodness of fit statistic is

$$\hat{\mathbf{e}}'\hat{\mathbf{e}} = 59.3476,$$

with 21-6=15 degrees of freedom.

Is a fit of order 3 poorer than a fit of order 5? An F-statistic is possible by taking the difference in the goodness of fit statistics, divided by an estimate of the residual variance. The residual variance is estimated from a fit of order $k-1$ or in this case of order 5. The goodness of fit statistic for order 5 was 7.2139 with 21-15=6 degrees of freedom. Hence the residual variance is

$$\sigma^2 = 7.2139/6 \;=\; 1.2023.$$

The F-statistic to test if a fit of order 3 is different from a fit of order 5 is

$$\begin{aligned} F &= \frac{(\hat{\mathbf{e}}'\hat{\mathbf{e}}_{m=3} - \hat{\mathbf{e}}'\hat{\mathbf{e}}_{m=5})/(15-6)}{\sigma^2} \\ &= \frac{(59.3476 - 7.2139)/9}{1.2023} \\ &= 5.7926/1.2023 \;=\; 4.8180, \end{aligned}$$

with (9,6) degrees of freedom. The table F-value at the ($P = .05$) level is 4.10. Thus, the difference is significant, and a fit of order 5 is better than a fit of order 3.

8.4 Basic Structure of RRM

Random regression models have a basic structure that is similar in most applications. A simplified RRM for a single trait can be written as

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

where

$y_{ijkn:t}$ is the n^{th} observation on the k^{th} animal at time t belonging to the i^{th} fixed factor and the j^{th} group;

F_i is a fixed effect that is independent of the time scale for the observations, such as a cage effect, a location effect or a herd-test date effect;

$g(t)_j$ is a function or functions that account for the phenotypic trajectory of the average observations across all animals belonging to the j^{th} group;

$r(a, x, m1)_k = \sum_{\ell=0}^{m1} a_{k\ell} x_{ijk:\ell}$ is the notation adopted for a random regression function. In this case, a denotes the additive genetic effects of the k^{th} animal, x is the vector of time covariates, and $m1$ is the order of the regression function. So that $x_{ijk:\ell}$ are the covariables related to time t, and $a_{k\ell}$ are the animal additive genetic regression coefficients to be estimated;

$r(pe, x, m2)_k = \sum_{\ell=0}^{m2} p_{k\ell} x_{ijk:\ell}$ is a similar random regression function for the permanent environmental (pe) effects of the k^{th} animal; and

$e_{ijkn:t}$ is a random residual effect with mean null and with possibly different variances for each t or functions of t.

The function, $g(t)_j$, can be either linear or nonlinear in t. Such a function is necessary in a RRM to account for the phenotypic relationship between y and the time covariables (or other types of covariables that could be used in a RRM). In a test day model, $g(t)_j$ accounts for different lactation curve shapes for groups of animals defined by years of birth, parity number, and age and season of calving within parities, for example. With growth data, $g(t)_j$ accounts for the growth curve of males or females of breed X or breed Y from young or old dams.

If the shape of the phenotypic relationship is not known or is nonlinear, then $g(t)_j$ could be a set of classification variables. Classification variables take up more degrees of freedom and require a large number of observations per level, but they do not force the user to explicitly define the shape of the trajectory. A mathematical function, on the other hand, does not use many degrees of freedom and gives a smooth trajectory over time regardless of the number of observations. The choice of classification variables or mathematical function is up to the researcher. If data are very numerous, and the mathematical function fits the data well, then either approach will generally lead to the same results. The phenotypic relationships, $g(t)_j$, are important to a RRM analysis and deserve care and effort in their correct specification.

The random regressions are intended to model the deviations around the phenotypic trajectories. The pattern of variation may be very different in shape or appearance from the phenotypic relationships, and may be more simple than $g(t)_j$. Orthogonal polynomials of standardized units of time have been recommended as covariables (Kirkpatrick et al., 1990). Orthogonal polynomials have computational advantages. The primary general advantage is the reduced correlations among the estimated coefficients. A standardized unit of time, w, ranges from -1 to +1, and is derived as

$$w = \frac{2 * (t - t_{min})}{(t_{max} - t_{min})} - 1,$$

where t_{min} is the earliest date (or the youngest age) and t_{max} is the latest date (or oldest age) represented in the data. The order of the orthogonal polynomials would be m_1 and m_2, i.e. the highest power of polynomial. Note that m_1 and m_2 do not need to be equal, but often (for simplicity of computing) they are chosen to be the same. Meyer(2000) and Pool et al. (2000), for example, compared many RRM models with different orders of orthogonal polynomials for the genetic and *pe* effects. Several types of orthogonal polynomials are available, but Legendre polynomials have been utilized (Kirkpatrick et al., 1990).

The residual variance should not be assumed to be constant from t_{min} to t_{max}. The residual effect is also known as a temporary environmental effect. Changes in residual variance might be predictable depending on the trajectory of the phenotypic data. For example, if RRM were being applied to growth data, weights may increase linearly with age, and the variance of weights may increase quadratically with age. Thus, the residual variance would be expected to increase in a similar manner as the phenotypic variance. Residual variances can be fit with a function of t, or assumed to have an autoregressive structure, or can be grouped into intervals having equal variance within the intervals. Research in this area is needed.

In matrix notation the RRM is

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

where $\mathbf{b}$ contains F_i and $g(t)_j$ effects, $\mathbf{a}$ contains $m_1 + 1$ additive genetic regression coefficients for each animal, $\mathbf{p}$ contains $m_2 + 1$ permanent environmental regression coefficients for each animal with data, and $\mathbf{e}$ contains the temporary environmental effects. Also,

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

where $\mathbf{G}$ is the variance-covariance matrix of the additive genetic random regression coefficients of order m_1+1; $\mathbf{P}$ is the variance-covariance matrix of the permanent environmental random regression coefficients of order $m_2 + 1$; and $\mathbf{R}$ is a diagonal matrix of temporary environmental variances which could vary depending on t, or $\mathbf{R}$ could be block diagonal with an autocorrelation structure for each animal's records. The mixed model equations (MME) are represented as

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_2 \\ \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_1 + \mathbf{A}^{-1}\otimes\mathbf{G}^{-1} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_2 \\ \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_2 + \mathbf{I}\otimes\mathbf{P}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{p}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}.$$

Assumptions about the distributions of $\mathbf{y}$ and other random variables are not necessary to derive best linear unbiased predictors (BLUP)(Goldberger, 1962; Henderson, 1984) or the

MME, but when **y** is normally distributed then BLUP is also BLP if the model is correct and variances and covariances are known. In order to estimate the elements of **G**, **P**, and **R** via Bayesian methods or restricted maximum likelihood, then normality of the random variables must be assumed (See for example Jamrozik and Schaeffer, 1997).

8.4.1 Example Data Analysis By RRM

Below are the data structure and pedigrees of four dairy cows. Given is the age at which they were observed for a trait during four visits to one herd.

			Age; Obs. at Visit			
Cow	Sire	Dam	Visit 1	Visit 2	Visit 3	Visit 4
1	7	5	22;224	34;236	47;239	
2	7	6	30;244	42;247	55;241	66;244
3	8	5	28;224	40;242		
4	8	1		20;220	33;234	44;228

The model equation might be

$$\begin{aligned} y_{jik:t} &= V_j + b_0 + b_1(A) + b_2(A)^2 \\ &\quad +(a_{i0}z_0 + a_{i1}z_1 + a_{i2}z_2) \\ &\quad +(p_{i0}z_0 + p_{i1}z_1 + p_{i2}z_2) + e_{jik:t} \end{aligned}$$

where

V_j is a random contemporary group effect which is assumed to follow a normal distribution with mean 0 and variance, $\sigma_c^2 = 4$.

b_0, b_1, and b_2 are fixed regression coefficients on (A) = age and age squared which describes the general relationship between age and the observations,

a_{i0}, a_{i1}, and a_{i2} are random regression coefficients for animal i additive genetic effects, assumed to follow a multivariate normal distribution with mean vector null and variance-covariance matrix, **G**,

p_{i0}, p_{i1}, and p_{i2} are random regression coefficients for animal i permanent environmental effects, assumed to follow a multivariate normal distribution with mean vector null and variance-covariance matrix, **P**,

z_0, z_1, and z_2 are the Legendre polynomials based on standardized ages and derived as indicated earlier. The minimum age was set at 18 and the maximum age was set at 68 for calculating the Legendre polynomials.

and e_{jik} is a temporary residual error term assumed to follow a normal distribution with mean 0 and variance, $\sigma_e^2 = 9$. In this example, the residual variance is assumed to be constant across ages.

The model in matrix notation is

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Wv} + \mathbf{Za} + \mathbf{Zp} + \mathbf{e},$$

where

$$\mathbf{X} = \begin{pmatrix} 1 & 22 & 484 \\ 1 & 30 & 900 \\ 1 & 28 & 784 \\ 1 & 34 & 1156 \\ 1 & 42 & 1764 \\ 1 & 40 & 1600 \\ 1 & 20 & 400 \\ 1 & 47 & 2209 \\ 1 & 55 & 3025 \\ 1 & 33 & 1089 \\ 1 & 66 & 4356 \\ 1 & 44 & 1936 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 224 \\ 244 \\ 224 \\ 236 \\ 247 \\ 242 \\ 220 \\ 239 \\ 241 \\ 234 \\ 244 \\ 228 \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and

$$\mathbf{Z} = \begin{pmatrix} .7071 & -1.0288 & .8829 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .7071 & -.6369 & -.1493 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .7071 & -.7348 & .0632 & 0 & 0 & 0 \\ .7071 & -.4409 & -.4832 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .7071 & -.0490 & -.7868 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .7071 & -.1470 & -.7564 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .7071 & -1.1268 & 1.2168 \\ .7071 & .1960 & -.7299 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .7071 & .5879 & -.2441 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .7071 & -.4899 & -.4111 \\ 0 & 0 & 0 & .7071 & 1.1268 & 1.2168 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .7071 & .0490 & -.7868 \end{pmatrix}.$$

In order to reduce rounding errors the covariates of age for the fixed regressions can be forced to have a mean of approximately zero by subtracting 38 from all ages and 1642 from all ages squared. Then

$$\mathbf{X} = \begin{pmatrix} 1 & -16 & -1158 \\ 1 & -8 & -742 \\ 1 & -10 & -858 \\ 1 & -4 & -486 \\ 1 & 4 & 122 \\ 1 & 2 & -42 \\ 1 & -18 & -1242 \\ 1 & 9 & 567 \\ 1 & 17 & 1383 \\ 1 & -5 & -553 \\ 1 & 28 & 2714 \\ 1 & 6 & 294 \end{pmatrix}.$$

The mixed model equations that need to be constructed to provide estimated breeding values are as follows;

$$\begin{pmatrix} \mathbf{X'X} & \mathbf{X'W} & \mathbf{X'Z} & \mathbf{0} & \mathbf{X'Z} \\ \mathbf{W'X} & \mathbf{W'W}+\mathbf{I}k_1 & \mathbf{W'Z} & \mathbf{0} & \mathbf{W'Z} \\ \mathbf{Z'X} & \mathbf{Z'W} & \mathbf{Z'Z}+\mathbf{A}^{nn}\otimes\mathbf{G}^{-1}\sigma_e^2 & \mathbf{A}^{nb}\otimes\mathbf{G}^{-1}\sigma_e^2 & \mathbf{Z'Z} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}^{bn}\otimes\mathbf{G}^{-1}\sigma_e^2 & \mathbf{A}^{bb}\otimes\mathbf{G}^{-1}\sigma_e^2 & \mathbf{0} \\ \mathbf{Z'X} & \mathbf{Z'W} & \mathbf{Z'Z} & \mathbf{0} & \mathbf{Z'Z}+\mathbf{I}\otimes\mathbf{P}^{-1}\sigma_e^2 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{c}} \\ \hat{\mathbf{a}}_n \\ \hat{\mathbf{a}}_b \\ \hat{\mathbf{p}} \end{pmatrix} = \begin{pmatrix} \mathbf{X'y} \\ \mathbf{W'y} \\ \mathbf{Z'y} \\ \mathbf{0} \\ \mathbf{Z'y} \end{pmatrix}.$$

where $k_1 = \sigma_e^2/\sigma_v^2$.

The entire MME can not be presented (order 43), but parts of the MME are given below.

$$\mathbf{W'W} = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix},$$

$$\mathbf{W'X} = \begin{pmatrix} 3 & -34 & -2758 \\ 4 & -16 & -1648 \\ 3 & 21 & 1397 \\ 2 & 34 & 3008 \end{pmatrix},$$

$$\mathbf{X'X} = \begin{pmatrix} 12 & 5 & -1 \\ 5 & 1995 & 166,883 \\ -1 & 166,883 & 14,415,319 \end{pmatrix},$$

$\mathbf{Z'Z}$ is composed of the following four blocks of order 3, for the four animals with records;

$$\text{Animal 1} \quad \begin{pmatrix} 1.5 & -.9006 & -.2335 \\ -.9006 & 1.2912 & -.8383 \\ -.2335 & -.8383 & 1.5457 \end{pmatrix},$$

$$\text{Animal 2} \quad \begin{pmatrix} 2 & .7275 & .0259 \\ .7275 & 2.0233 & 1.3612 \\ .0259 & 1.3612 & 2.1815 \end{pmatrix},$$

$$\text{Animal 3} \quad \begin{pmatrix} 1 & -.6235 & -.4902 \\ -.6235 & .5615 & .0648 \\ -.4902 & .0648 & .5761 \end{pmatrix},$$

$$\text{Animal 4} \quad \begin{pmatrix} 1.5 & -1.1085 & .0134 \\ -1.1085 & 1.5121 & -1.2082 \\ .0134 & -1.2082 & 2.2687 \end{pmatrix}.$$

and $\mathbf{Z}'\mathbf{X}$ is

$$\mathbf{Z}'\mathbf{X} = \begin{pmatrix} 2.1213 & -7.7781 & -761.5467 \\ -1.2737 & 19.9884 & 1516.7598 \\ -.3302 & -18.7627 & -1201.416 \\ 2.8284 & 28.9911 & 2458.5867 \\ 1.0288 & 46.4439 & 4337.8027 \\ .0366 & 27.9679 & 2979.5959 \\ 1.4142 & -5.6568 & -636.3900 \\ -.8818 & 7.0540 & 636.6324 \\ -.6932 & -2.1448 & -22.4568 \\ 2.1213 & -12.0207 & -1061.3570 \\ -1.5677 & 23.0259 & 1684.8063 \\ .0189 & -24.5677 & -1515.2470 \end{pmatrix}.$$

The right hand sides of the MME are

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} 2823 \\ 2070 \\ 68,064 \end{pmatrix},$$

$$\mathbf{W}'\mathbf{y} = \begin{pmatrix} 692 \\ 945 \\ 714 \\ 472 \end{pmatrix},$$

and

$$\mathbf{Z}'\mathbf{y} = \begin{pmatrix} 494.2629 \\ -287.6596 \\ -90.7117 \\ 690.1296 \\ 249.1165 \\ 7.3023 \\ 329.5086 \\ -200.1692 \\ -168.8920 \\ 482.2422 \\ -351.3606 \\ -7.8918 \end{pmatrix}.$$

The assumed variance-covariance matrices of the additive and permanent environmental effects need to be known for BLUP. Let

$$\mathbf{G} = \begin{pmatrix} 94.0000 & -3.8500 & .03098 \\ -3.8500 & 1.5000 & -.0144 \\ .03098 & -.0144 & .0014 \end{pmatrix},$$

$$\mathbf{P} = \begin{pmatrix} 63.0000 & -2.1263 & .0447 \\ -2.1263 & .5058 & -.00486 \\ .0447 & -.00486 & .0005 \end{pmatrix},$$

and $\sigma_v^2 = 4$, and $\sigma_e^2 = 9$. The solutions to MME are

$$\hat{\mathbf{b}}' = \begin{pmatrix} 234.4349 & 1.5957 & -.01600 \end{pmatrix},$$

$$\hat{\mathbf{c}}' = \begin{pmatrix} -.8213 & 1.5179 & .0770 & -.7736 \end{pmatrix}.$$

Let the solutions for the animal additive genetic random regression coefficients be presented as follows, where each row represents the coefficients for one animal (i.e. for the intercept, linear, and quadratic regressions).

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

Similarly, the solutions for the animal permanent environmental random regression coefficients can be given as

$$\hat{\mathbf{p}} = \begin{pmatrix} -.370066 & .059735 & -.000696 \\ 4.308127 & -.250192 & .004092 \\ -.424394 & .145076 & -.001497 \\ -3.513555 & .045355 & -.001899 \end{pmatrix}.$$

The problem is to rank the animals for selection purposes. If animals are ranked on the basis of a_0, (the intercepts) then animal 2 would be the highest (if that was desirable). If ranked on the basis of a_1 (the linear regression coefficient), then animal 3 would be the highest, and if ranked on the basis of a_2 (the quadratic coefficient), then animal 2 would be the highest. To properly rank the animals, an EBV at different ages should be calculated, and then these could be combined with appropriate economic weights. Suppose EBVs for 24, 36, and 48 mo of age are of particular interest, and the economic weights for these ages might be 2, 1, and .5, respectively. A Total Economic Value can be calculated as

$$\text{TEV} = 2 * \text{EBV}(24) + 1 * \text{EBV}(36) + .5 * \text{EBV}(48).$$

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

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

The results are shown in the following table.

Animal	EBV(24)	EBV(36)	EBV(48)	TEV
1	-1.34	-1.28	-1.20	-4.56
2	4.55	4.27	3.94	15.33
3	-2.45	-2.21	-1.93	-8.07
4	-3.62	-3.54	-3.44	-12.49
5	-1.67	-1.52	-1.34	-5.52
6	2.57	2.42	2.25	8.69
7	1.38	1.27	1.14	4.60
8	-2.28	-2.18	-2.05	-7.76

The animal with the highest TEV was animal 2. All animals ranked rather similarly at each age on their EBVs. Rankings of animals could change with age. Thus, the pattern of growth could be changed to one that is desirable.

8.5 Estimation of Parameters

8.5.1 EM-REML

Estimation of the common residual variance is

$$\hat{\sigma}_e^2 = (\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} - \hat{\mathbf{c}}'\mathbf{W}'\mathbf{y} - \hat{\mathbf{a}}_n'\mathbf{M}'\mathbf{y} - \hat{\mathbf{p}}'\mathbf{M}'\mathbf{y})/(N - r(\mathbf{X})),$$

where

$$\begin{aligned} \mathbf{y}'\mathbf{y} &= 665,035, \\ \hat{\beta}'\mathbf{W}'\mathbf{y} &= 664877.89, \\ N - r(\mathbf{X}) &= 12 - 3 = 9, \\ \hat{\sigma}_e^2 &= 17.4570. \end{aligned}$$

Let $\mathbf{C}$ represent the inverse of the MME coefficient matrix, and let $\mathbf{C}_{aa}$ and $\mathbf{C}_{pp}$ represent the submatrices of $\mathbf{C}$ corresponding to animal additive genetic effects and animal permanent environmental effects, respectively. Note that $\mathbf{C}_{aa}$ is of order 24, and $\mathbf{C}_{pp}$ has order 12. Because the coefficients are ordered within animals, then both $\mathbf{C}_{aa}$ and $\mathbf{C}_{pp}$ can be partitioned into submatrices each of order 3 by animals. The submatrix $\mathbf{C}_{cc}$ represents the inverse elements for the visit effects.

The variance due to visit effects is estimated as usual in EM-REML as

$$\hat{\sigma}_v^2 = (\hat{\mathbf{c}}'\hat{\mathbf{c}} + tr(\mathbf{C}_{cc})\hat{\sigma}_e^2)/4,$$

where

$$\hat{\mathbf{c}}'\hat{\mathbf{c}} = 3.5828,$$

$$\begin{aligned} tr(\mathbf{C}_{cc}) &= 1.2890, \\ \hat{\sigma}_v^2 &= (3.5828 + (1.2890)17.4570)/4, \\ &= 6.5214. \end{aligned}$$

To estimate $\mathbf{G}$, first calculate $\hat{\mathbf{a}}'\mathbf{A}^{-1}\hat{\mathbf{a}}$ using the $\hat{\mathbf{a}}$ given in the previous section, which gives

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

The tricky part is calculating the appropriate traces. They are

$$tr_3((\mathbf{A}^{-1} \otimes \mathbf{I}_3)\mathbf{C}_{aa}) = \begin{pmatrix} 68.347139 & -2.689821 & .021540 \\ -2.689821 & 1.284185 & -.012359 \\ .021540 & -.012359 & .001240 \end{pmatrix},$$

where the trace is the sum of the 3 by 3 diagonal blocks of the matrix product indicated. Theses are combined to estimate $\mathbf{G}$ as

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

Similarly, the estimate of $\mathbf{P}$ is also a 3 by 3 matrix.

$$\begin{aligned} \hat{\mathbf{p}}'\hat{\mathbf{p}} &= \begin{pmatrix} 31.222084 & -1.320890 & .025192 \\ -1.320890 & .089268 & -.001369 \\ .025192 & -.001369 & .000023 \end{pmatrix}, \\ tr_3(\mathbf{C}_{pp}) &= \begin{pmatrix} 18.467546 & -.590303 & .012875 \\ -.590303 & .209410 & -.001880 \\ .012875 & -.001880 & .000217 \end{pmatrix}, \\ \hat{\mathbf{P}} &= (\hat{\mathbf{p}}'\hat{\mathbf{p}} + tr_3(\mathbf{C}_{pp})\hat{\sigma}_e^2)/4, \\ &= \begin{pmatrix} 88.402374 & -2.906448 & .062487 \\ -2.906448 & .936231 & -.008547 \\ .062487 & -.008547 & .000952 \end{pmatrix}. \end{aligned}$$

The new parameter estimates would be used to re-construct the MME and to solve them again. Then repeat the estimation of parameters and continue iterating these steps until convergence has been achieved, which will be when the global maximum has been attained.

8.5.2 Bayesian Estimation

Using Gibbs sampling, as with other models, the MME would be solved and new samples generated for the fixed regressions, visit effects, animal additive genetic and animal permanent environmental effects by processing one equation at a time. Better performance can be achieved by using blocks of 3 by 3 matrices for the 3 random regression coefficients for animal additive genetic and animal permanent environmental effects. To illustrate, suppose the diagonal block for animal i additive genetic effects from the MME was

$$(\mathbf{Z}_i'\mathbf{Z}_i + a^{ii}\mathbf{G}_k^{-1}\sigma^2_{e_k}) = \mathbf{Q}_{i_k} = \begin{pmatrix} 1.7675 & -.2017 & 1.0346 \\ -.2017 & 19.7604 & 173.6637 \\ 1.0346 & 173.6637 & 17839.79 \end{pmatrix},$$

and that the current 'solution' for animal i was $\hat{a}_i$,

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

and the current sample value of σ^2_e 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^2_e which gives

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

2. Apply a Cholesky decomposition to the previous matrix,

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

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

$$\begin{aligned} \mathbf{a}_i &= \hat{\mathbf{a}}_i + \mathbf{L}\mathbf{f}, \\ &= \begin{pmatrix} -1.7473 \\ .1248 \\ -.0012 \end{pmatrix} + \begin{pmatrix} 3.144938 & 0 & 0 \\ .036857 & .982895 & 0 \\ -.000541 & -.009568 & .031282 \end{pmatrix} \begin{pmatrix} 1.0673 \\ -.5892 \\ -.9814 \end{pmatrix}, \\ &= \begin{pmatrix} 1.6093 \\ -.4150 \\ -.0269 \end{pmatrix}. \end{aligned}$$

The animal permanent environmental regression coefficients would be handled in the same way as the animal additive genetic coefficients. After new sample values for all effects in the MME have been obtained, then the following quadratic forms are calculated, for example,

$$\begin{aligned}
\mathbf{c}'\mathbf{c} &= 4.11, \\
\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} &= \begin{pmatrix} 64.71 & -4.15 & .0532 \\ -4.15 & .51 & -.0049 \\ .0532 & -.0049 & .000054 \end{pmatrix}, \\
&= \mathbf{G}_k, \\
\mathbf{p}'\mathbf{p} &= \begin{pmatrix} 37.22 & -1.25 & .0247 \\ -1.25 & .0873 & -.0011 \\ .0247 & -.0011 & .000022 \end{pmatrix}, \\
&= \mathbf{P}_k, \\
\mathbf{e} &= \mathbf{y} - \mathbf{W}\beta, \\
\mathbf{e}'\mathbf{e} &= 74.8851.
\end{aligned}$$

The forms $\mathbf{c}'\mathbf{c}$ aud $\mathbf{e}'\mathbf{e}$ follow inverted Chi-square distributions while the other two follow inverted Wishart distributions. New sample values for the variances would be given by the following:

$$\begin{aligned}
\sigma_v^2 &= \mathbf{c}'\mathbf{c}/\chi_4^2, \\
\sigma_e^2 &= \mathbf{e}'\mathbf{e}/\chi_{12}^2, \\
\mathbf{T} &= Chol(\mathbf{G}_k^{-1}), \\
\mathbf{G}_{k+1}^{-1} &= \text{Wishart}(\text{T}, 8), \\
\mathbf{S} &= Chol(\mathbf{P}_k^{-1}), \\
\mathbf{P}_{k+1}^{-1} &= \text{Wishart}(\text{S}, 4),
\end{aligned}$$

and these give new sample values for $\mathbf{G}$ and $\mathbf{P}$.

Chapter 9

Multiple Trait Models

9.1 Introduction

Animals are commonly observed for more than one trait because many traits affect overall profitability. A multiple trait (MT) model is one in which two or more traits are analyzed simultaneously in order to take advantage of genetic and environmental correlations between traits.

Reasons for using a multiple trait analysis are as follows.

1. **Low Heritability Traits.** MT models are useful for traits where the difference between genetic and residnal correlations are large (e.g. greater than .5 difference) or where one trait has a much higher heritability than the other trait. Traits with low heritability tend to gain in accuracy when analyzed with high heritability traits, although all traits benefit to some degree from the simultaneous analysis.

2. **Culling.** Traits that occur at different times in the life of the animal, such that culling of animals may occur between measurements, are suitable for MT analyses. Consequently, animals which have observations later in life tend to be selected based on their performance for earlier traits. Thus, analysis of later life traits by themselves could suffer from the effects of culling bias, and the resulting EBV could lead to errors in selecting future parents. An MT analysis that includes all observations on an animal upon which culling decisions have been based, has been shown to account for the selection that has taken place, and therefore gives unbiased estimates of breeding values for all traits.

3. **Missing Traits.** Some traits may be difficult to record on animals, and therefore, relatively few animals are actually measured for a trait. However, because this trait is genetically correlated to the more easily recorded traits, then a multiple trait analysis could help the accuracy of the less recorded trait. An example in dairy

cattle could be a newly introduced conformation trait for which only the more recent animals have been scored.

For cases where heritabilities of traits are similar in magnitude, or where both genetic and residual correlations are relatively the same, or where every animal is measured for all traits, the benefits of a MT analysis will be almost unnoticeable. However, if culling bias exists, then an MT analysis should be performed even if the heritabilities and correlations are similar among traits.

The accuracy of the assumed genetic and residual correlations are critical to the success of a multiple trait analysis. If the parameter estimates are greatly different from the unknown true values, then an MT analysis could do as much harm as it might do good.

Lastly, multiple trait analyses can be very costly and time consuming to execute. MT programs are more complicated than single trait programs, more memory and disk storage are usually needed, and verification of results might be more complicated. These have to be balanced against the benefits of an MT analysis. If culling bias is the main concern, then an MT model must be used regardless of the costs or no analysis should be done at all, except for the traits not affected by culling bias.

9.2 Models

Consider two traits with a single observation per trait on animals. A model should be specified separately for each trait. Usually, the same model is assumed for each trait, and this can greatly simplify the computational aspects, but such an assumption may be unrealistic in many situations.

Let the model equation for trait 1 be

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

where B_{1i} is a fixed effect with p_B levels, a_{1j} is a random, animal additive genetic effect for trait 1, and e_{1ij} is a random residual environmental effect for trait 1. The model equation for trait 2 might be

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

where C_{2i} is a fixed effect (different from B_{1i} for trait 1) with p_C levels, a_{2j} is a random, animal additive genetic effect for trait 2, and e_{2ij} is a random residual environmental effect for trait 2.

For example, y_{1ij} could be a trait like birthweight, so that B_{1i} could identify animals born in the same season. Trait 2 could be yearling weights and C_{2i} could identify contemporary groups of animals of the same sex, same herd, and same rearing unit within herd.

Because the two traits will be analyzed simultaneously, the variances and covariances need to be specified for the traits together. For example, the additive genetic variance-covariance (VCV) matrix could be written as

$$\mathbf{G} = \begin{pmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 15 \end{pmatrix},$$

and the residual environmental VCV matrix as

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{pmatrix} = \begin{pmatrix} 10 & 5 \\ 5 & 100 \end{pmatrix}.$$

The genetic and residual correlations are, respectively,

$$\begin{aligned} \rho_g &= 2/(15)^{.5} = .516, \\ \rho_r &= 5/(1000)^{.5} = .158 \end{aligned}$$

with

$$h_1^2 = \frac{1}{11} = .0909,$$

and

$$h_2^2 = \frac{15}{115} = .1304.$$

For all data, then

$$Var \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A}g_{11} & \mathbf{A}g_{12} \\ \mathbf{A}g_{12} & \mathbf{A}g_{22} \end{pmatrix}.$$

The structure of the residual VCV matrix over all observations can be written several ways depending on whether allowance is made for missing observations on either trait for some animals. If all animals were observed for both traits, then

$$Var \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I}r_{11} & \mathbf{I}r_{12} \\ \mathbf{I}r_{12} & \mathbf{I}r_{22} \end{pmatrix}.$$

9.3 Example Data

The following data were available following the previous models. Note that some of the trait 2 observations are missing, and therefore, possible culling bias is a reason to use a multiple trait analysis.

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

9.3.1 The MME

Organize the data by traits within animals. The residual covariance matrix is given by $\mathbf{R}$ below.

$$\mathbf{R} = \begin{pmatrix} 10 & 5 \\ 5 & 100 \end{pmatrix}.$$

Two computing algorithms will be employed in order to simplify the multiple trait analysis.

1. Assume that there is a common model for all traits. In this case the common model is

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

   There are 2 levels of factor B, three levels of factor C, and 12 animals. Previously, factor B was only associated with trait 1. Therefore, during the analysis (by iteration on data, for example), the solutions for the B factor for trait 2 must always be kept at zero. Similarly, the C factor was only for trait 2, so that the solutions for the C factor pertaining to trait 1 must always be kept at zero.

2. Assume that all animals are observed for all traits. Therefore, $\mathbf{R}$ is the same for all animals and missing observations can be left at 0. However, in order to handle the missing observations appropriately, the missing observation has to be assigned to a fixed factor, such that this observation is the only observation in that level of the factor. For example, for Animal 1 in the table, trait 2 is missing. Therefore, this observation is assigned to level 4 of factor C, C_{24}. Animals which have trait 2 are assigned to the appropriate level of factor C. Similarly, Animal 2 is also missing trait 2 and is assigned to level 5 of factor C. The other missing trait 2 observations belong to Animals 6 and 11, which are assigned to levels 6 and 7 of factor C, respectively.

Because trait 1 is present for all animals, all trait 1 observations are assigned to the appropriate levels of factor B. If trait 1 was missing for an animal, then a new level of factor B would be created for that animal. By putting an animal into a level of a factor by itself automatically takes care of the observation being missing. This algorithm is due to Bruce Tier (AGBU-1998-PhD Thesis).

To prove that this works, take $\mathbf{R}^{-1}$ and do a Gaussian elimination (i.e. absorption) of the row and column corresponding to the missing trait, say trait 2,

$$\begin{pmatrix} r^{11} & r^{12} \\ r^{12} & r^{22} \end{pmatrix} - \begin{pmatrix} r^{12} \\ r^{22} \end{pmatrix} (r^{22})^{-1} \begin{pmatrix} r^{12} & r^{22} \end{pmatrix},$$

$$= \begin{pmatrix} r^{11} & r^{12} \\ r^{12} & r^{22} \end{pmatrix} - \begin{pmatrix} r^{12}(r^{22})^{-1}r^{12} & r^{12} \\ r^{12} & r^{22} \end{pmatrix},$$

$$= \begin{pmatrix} r^{11} - r^{12}(r^{22})^{-1}r^{12} & 0 \\ 0 & 0 \end{pmatrix}.$$

Recall that for a matrix of order 2 that

$$\begin{aligned} r^{11} &= r_{22}/\mid \mathbf{R} \mid, \\ r^{12} &= -r_{12}/\mid \mathbf{R} \mid, \\ r^{22} &= r_{11}/\mid \mathbf{R} \mid, \\ \mid \mathbf{R} \mid &= (r_{11}r_{22} - r_{12}r_{12}) \end{aligned}$$

then

$$\begin{aligned} r^{11} - r^{12}(r^{22})^{-1}r^{12} &= (r_{22} - r_{12}(r_{11})^{-1}r_{12})/\mid \mathbf{R} \mid \\ &= r_{11}(r_{22} - r_{12}(r_{11})^{-1}r_{12}/r_{11}(r_{11}r_{22} - r_{12}r_{12}) \\ &= (r_{11})^{-1} \end{aligned}$$

which is exactly the weight applied to records on animals with only trait 1 observed. This proof can be extended to any number of traits recorded and any number missing, by partitioning $\mathbf{R}$ into

$$\begin{pmatrix} \mathbf{R}_{oo} & \mathbf{R}_{om} \\ \mathbf{R}_{mo} & \mathbf{R}_{mm} \end{pmatrix},$$

where the subscript o refers to traits that were observed and m refers to traits that were missing on an animal. Then it can be easily shown that

$$\mathbf{R}_{oo}^{-1} = \mathbf{R}^{oo} - \mathbf{R}^{om}(\mathbf{R}^{mm})^{-1}\mathbf{R}^{mo}.$$

Iteration on Data Scheme

The following will demonstrate the iteration on data technique to solve MME for the multiple traits, using the example data. Assume that the iterated solutions are currently

at the following values.

$$\begin{pmatrix} B_{11} \\ B_{12} \\ C_{21} \\ C_{22} \\ C_{23} \\ C_{24} \\ C_{25} \\ C_{26} \\ C_{27} \end{pmatrix} = \begin{pmatrix} 5.0209 \\ 6.5592 \\ 20.0882 \\ 49.0575 \\ 51.9553 \\ 2.8590 \\ 0.1322 \\ 2.1189 \\ 1.1463 \end{pmatrix},$$

and the animal additive genetic current solutions are in the following table.

Animal	Sire	Dam	Trait 1	Trait 2
1	0	0	-.3573	-1.6772
2	0	0	-.0730	1.0418
3	0	0	.4105	1.1707
4	0	0	-.0449	-1.4923
5	0	0	.0646	.9570
6	1	3	-.1033	-.1410
7	1	4	-.1975	-2.2983
8	1	5	-.1410	-.9633
9	2	3	.3079	1.6227
10	2	4	.1426	1.1273
11	2	5	-.1830	.6418
12	6	10	.1554	1.5089

The MME are ordered by B-factor equations, then C-factor, then animal additive genetic.

Factor B. Go through the data, one animal at a time. For animal 1, for example, take the observations and subtract the solutions for all other factors, except factor B, i.e.,

$$\begin{pmatrix} 2.3 - a_{11} \\ 0 - C_{24} - a_{21} \end{pmatrix} = \begin{pmatrix} 2.30 + 0.3573 \\ 0.00 - 2.8590 + 1.6772 \end{pmatrix} = \begin{pmatrix} 2.6573 \\ -1.1818 \end{pmatrix}.$$

Premultiply by $\mathbf{R}^{-1}$ and accumulate into the right hand sides for factor B, level 1.

$$\mathbf{R}^{-1} \begin{pmatrix} 2.6573 \\ -1.1818 \end{pmatrix} = \begin{pmatrix} 0.2786 \\ -0.0257 \end{pmatrix}.$$

The accumulated right-hand-sides for level B1 over the first five animals is then

$$RHS = \begin{pmatrix} 2.5748 \\ -0.1697 \end{pmatrix},$$

and a new solution for level 1 of factor B for trait one is obtained by dividing the right hand side, 2.5748, by the diagonal which is $5r^{11} = 0.5128$, giving 5.0209. Because factor B is not in the model for trait 2, then the solution for the second trait is ignored, and B_{21} should be set to 0.

Similarly for level B2, after accumulating the deviations on animals 6 through 12,

$$RHS = \begin{pmatrix} 4.7092 \\ -0.1945 \end{pmatrix},$$

and the new level 2 solution for factor B is

$$4.7092/(7r^{11}) = 6.5592,$$

and the solution for trait 2, B_{22}, must be made 0.

Factor C. Go through the data a second time. Subtract solutions for all other factors except those for factor C. Accumulate the deviations in the appropriate right-hand-sides for the three levels of C. For animal 1,

$$\begin{pmatrix} 2.3 - B_{11} - a_{11} \\ 0 - a_{21} \end{pmatrix} = \begin{pmatrix} 2.30 - 5.0209 + 0.3573 \\ 0.00 + 1.6772 \end{pmatrix} = \begin{pmatrix} -2.3636 \\ 1.6772 \end{pmatrix}.$$

Premultiply by $\mathbf{R}^{-1}$ to give

$$\mathbf{R}^{-1} \begin{pmatrix} -2.3636 \\ 1.6772 \end{pmatrix} = \begin{pmatrix} -0.2510 \\ 0.0293 \end{pmatrix}.$$

The 0.0293 is accumulated in the RHS for level 4 of factor C because in this case the trait 2 observation was missing. Because this is the only observation in level 4, the new solution would be

$$C_{24} = 0.0293/r^{22} = 2.8590.$$

For animal 2,

$$\begin{pmatrix} 2.6 - B_{11} - a_{12} \\ 0 - a_{22} \end{pmatrix} = \begin{pmatrix} 2.6 - 5.0209 + 0.0730 \\ 0.0 - 1.0418 \end{pmatrix} = \begin{pmatrix} -2.3479 \\ -1.0418 \end{pmatrix}.$$

Premultiply by $\mathbf{R}^{-1}$ to give

$$\mathbf{R}^{-1} \begin{pmatrix} -2.3479 \\ -1.0418 \end{pmatrix} = \begin{pmatrix} -0.2355 \\ 0.0013556 \end{pmatrix}.$$

The 0.0013556 is accumulated in the RHS for level 5 of factor C because in this case the trait 2 observation was missing. Because this is the only observation in level 5, the new solution would be

$$C_{25} = 0.0013556/r^{22} = 0.132171.$$

For animal 3, where both traits were observed, for example,

$$\begin{pmatrix} 9.8 - B_{11} - a_{13} \\ 53 - a_{23} \end{pmatrix} = \begin{pmatrix} 9.80 - 5.0209 - 0.4105 \\ 53.00 - 1.1707 \end{pmatrix} = \begin{pmatrix} 4.3686 \\ 51.8293 \end{pmatrix}.$$

Animal 3 belonged to level 3 of factor C. Pre-multiply by $\mathbf{R}^{-1}$ to obtain

$$\mathbf{R}^{-1} \begin{pmatrix} 4.3686 \\ 51.8293 \end{pmatrix} = \begin{pmatrix} 0.1823 \\ 0.5092 \end{pmatrix}.$$

The value 0.5092 is added to the right hand sides for level 3 of Factor C.

After processing all 12 animals, the right-hand-sides for each level of factor C are

$$\begin{aligned} RHS_{21} &= 0.6181, \\ RHS_{22} &= 1.0063, \\ RHS_{23} &= 1.5986, \\ RHS_{24} &= 0.0293, \\ RHS_{25} &= 0.0013556, \\ RHS_{26} &= 0.0217, \\ RHS_{27} &= 0.0118. \end{aligned}$$

The new solutions are

$$\begin{aligned} C_{21} &= RHS_{21}/3r^{22} = 20.0883, \\ C_{22} &= RHS_{22}/2r^{22} = 49.0575, \\ C_{23} &= RHS_{23}/3r^{22} = 51.9553, \\ C_{24} &= RHS_{24}/r^{22} = 2.8590, \\ C_{25} &= RHS_{25}/r^{22} = 0.132171, \\ C_{26} &= RHS_{26}/r^{22} = 2.1189, \\ C_{27} &= RHS_{27}/r^{22} = 1.1463. \end{aligned}$$

Animal Additive Genetic. The inverse of the additive genetic relationship matrix for the twelve animals is

$$\mathbf{A}^{-1} = \frac{1}{2} \begin{pmatrix} 5 & 0 & 1 & 1 & 1 & -2 & -2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 5 & 1 & 1 & 1 & 0 & 0 & 0 & -2 & -2 & -2 & 0 \\ 1 & 1 & 4 & 0 & 0 & -2 & 0 & 0 & -2 & 0 & 0 & 0 \\ 1 & 1 & 0 & 4 & 0 & 0 & -2 & 0 & 0 & -2 & 0 & 0 \\ 1 & 1 & 0 & 0 & 4 & 0 & 0 & -2 & 0 & 0 & -2 & 0 \\ -2 & 0 & -2 & 0 & 0 & 5 & 0 & 0 & 0 & 1 & 0 & -2 \\ -2 & 0 & 0 & -2 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 & -2 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\ 0 & -2 & -2 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & -2 & 0 & -2 & 0 & 1 & 0 & 0 & 0 & 5 & 0 & -2 \\ 0 & -2 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & -2 & 0 & 4 \end{pmatrix}.$$

To get a new solution for Animal 1, for example, the record and its adjustments are

$$\begin{pmatrix} 2.3 - B_{11} \\ 0 - C_{24} \end{pmatrix} = \begin{pmatrix} 2.30 - 5.0209 \\ 0.00 - 2.8590 \end{pmatrix} = \begin{pmatrix} -2.7209 \\ -2.8590 \end{pmatrix}.$$

Premultiply this by $\mathbf{R}^{-1}$, giving

$$RHS = \begin{pmatrix} -0.2644 \\ -0.01537 \end{pmatrix}.$$

Now adjustments need to be made due to the relationship matrix. For Animal 1 these are

$$\begin{aligned} ADJ &= \frac{1}{2}(-a_3 - a_4 - a_5 + 2a_6 + 2a_7 + 2a_8), \\ &= \begin{pmatrix} -0.6569 \\ -3.7203 \end{pmatrix}. \end{aligned}$$

ADJ must be pre-multiplied by $\mathbf{G}^{-1}$ and added to RHS. The end result must be pre-multiplied by the inverse of the diagonal block for Animal 1, which is $\mathbf{R}^{-1} + 2.5\mathbf{G}^{-1}$.

$$\begin{aligned} \mathbf{G}^{-1}ADJ &= \frac{1}{11}\begin{pmatrix} 15 & -2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} -0.6569 \\ -3.7203 \end{pmatrix}, \\ &= \begin{pmatrix} -0.219355 \\ -0.218773 \end{pmatrix}, \\ RHS + \mathbf{G}^{-1}ADJ &= \begin{pmatrix} -0.483755 \\ -0.234143 \end{pmatrix}. \end{aligned}$$

The solutions for Animal 1 are then

$$(\mathbf{R}^{-1} + 2.5\mathbf{G}^{-1})^{-1}(RHS + \mathbf{G}^{-1}ADJ) = \begin{pmatrix} -0.3573 \\ -1.6772 \end{pmatrix}.$$

The same process is followed for each animal, until new solutions have been computed for each animal.

The results shown above were converged (to 4 decimal places at least), and therefore, the new solutions were identical to the old ones. However, this process is repeated until convergence is achieved, as it is here.

9.4 Estimation of Covariance Matrices

Derivative free REML is one option for estimating variances and covariances in a multi-trait situation, such as with MTDFREML. MTDFREML uses the simplex algorithm of

finding the global maximum of the likelihood function. Suppose there are two traits and only additive and residual covariance matrices. Then there are a total of six parameters to be estimated. MTDFREML would begin by evaluating seven likelihoods, and then do more to find the maximum. If the number of traits is increased to 3, then there are 12 parameters to be estimated and 13 initial likelihoods to evaluate. The numbers quickly increase as the number of traits increases. Also, if there is another random factor in the models, then there is another increase in parameters. Soon, the number of likelihood evaluations becomes too large to permit analysis.

The EM algorithm is not suitable due to the requirement for the traces of inverse elements that are needed. Inverses of the MME are much more of a problem than the log of the determinant of the MME coefficient matrix, and therefore, EM-REML would not be practical.

The Bayesian approach via Gibbs Sampling is very feasible, but the number of necessary samples might be too large or take too much time to complete. This section is intended to describe the Gibbs sampling approach to the multiple trait problem. Use will be made of the Iteration on Data algorithm given in the previous section.

9.4.1 Sampling Solutions to MME

For a t-trait analysis, any solution vector for a fixed effect factor in the MME is a $t \times 1$ vector, and can be represented as

$$\hat{\beta}_i = (\mathbf{X}_i'\mathbf{R}^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{R}^{-1}(\mathrm{y}_i - \mathbf{W}_{-i}\beta_{-i}),$$

then a new sample vector is generated by

$$\beta_i = \hat{\beta}_i + \mathbf{L}\mathbf{v},$$

where

$$\mathbf{L}\mathbf{L}' = (\mathbf{X}_i'\mathbf{R}^{-1}\mathbf{X}_i)^{-1},$$

and $\mathbf{v}$ is a $t \times 1$ vector of random normal deviates.

For factor B and C effects in the example, because factor B is associated with trait 1 only, and factor C is associated with trait 2 only, the sampling procedure for these effects are scalar. Thus, the new B_{11} was 5.0209, and the diagonal of the MME corresponding to that equation was $5r^{11}$. Therefore, the variance of the estimate would be the inverse of this or 1/0.5128. A new sample value would be

$$B_{11} = 5.0209 + RND * 1.95,$$

so that if $RND = -0.22$, then the new $B_{11} = 4.5919$. Similarly, for C_{21} the estimate was 20.0883 and the diagonal element of the MME for C_{21} was $3r^{22} = .0307692$, and a new sample value would be

$$C_{21} = 20.0883 + RND * 32.5,$$

and if $RND = 0.43$, then $C_{21} = 34.0633$.

For the additive genetic effects of animals, using animal 1 as an example, the solution vector was

$$\mathbf{a}_1 = \begin{pmatrix} -0.3573 \\ -1.6772 \end{pmatrix},$$

and the diagonal block of the MME for this animal was

$$\mathbf{R}^{-1} + 2.5 * \mathbf{G}^{-1} = \begin{pmatrix} 3.5117 & -0.4597 \\ -0.4597 & 0.2375 \end{pmatrix}.$$

The diagonal block must be inverted and then decomposed via the Cholesky decomposition, i.e.

$$\begin{pmatrix} 0.3814 & 0.7381 \\ 0.7381 & 5.6383 \end{pmatrix} = \begin{pmatrix} 0.6176 & 0 \\ 1.1951 & 2.0518 \end{pmatrix} \begin{pmatrix} 0.6176 & 1.1951 \\ 0 & 2.0518 \end{pmatrix}.$$

Generate a t by 1 vector of random normal deviates, say,

$$\mathbf{v} = \begin{pmatrix} -1.6794 \\ 0.5536 \end{pmatrix},$$

then the new sample values for the animal additive genetic effects is

$$\mathbf{a}_1 = \begin{pmatrix} -0.3573 \\ -1.6772 \end{pmatrix} + \begin{pmatrix} 0.6176 & 0 \\ 1.1951 & 2.0518 \end{pmatrix} \begin{pmatrix} -1.6794 \\ 0.5536 \end{pmatrix} = \begin{pmatrix} -1.0371 \\ -0.8711 \end{pmatrix}.$$

9.4.2 Sampling New Covariance Matrices

If $\mathbf{a}_i$ is the $q \times 1$ vector of animal solutions for trait i, then form

$$\mathbf{U} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_t \end{pmatrix},$$

followed by

$$\mathbf{S}_a = (\mathbf{U}'\mathbf{A}^{-1}\mathbf{U} + \nu_a \mathbf{G}_a).$$

If $\nu_a = 0$, then

$$\mathbf{S}_a = \mathbf{U}'\mathbf{A}^{-1}\mathbf{U} = \begin{pmatrix} 0.5623 & 2.1843 \\ 2.1843 & 16.6941 \end{pmatrix},$$

with $q = 12$ degrees of freedom. Invert this matrix and apply a Cholesky decomposition, giving

$$\begin{pmatrix} 1.9019 & 0 \\ -0.2488 & 0.2447 \end{pmatrix}.$$

This matrix is supplied to a Wishart distribution random generator along with q as the degrees of freedom, and from this a new $\mathbf{G}^{-1}$ is obtained.

The residual effects for each animal are obtained in a similar manner as in the iteration on data. Subtract the sample values of each of the effects of the model (for each trait) from the observations, $e_{ti} = y_{ti} - \mathbf{w}_i'\beta_t$. The residuals, for example, could be

$$\mathbf{E} = \begin{pmatrix} -2.3636 & -1.1818 \\ -2.3480 & -1.1740 \\ 4.3686 & -0.1260 \\ -0.2760 & -14.5960 \\ 0.4145 & 12.9856 \\ -3.9559 & -1.9779 \\ 2.0383 & -11.7592 \\ 1.7818 & -9.9920 \\ 2.1329 & 5.2890 \\ 1.0982 & 10.7845 \\ -3.5762 & -1.7881 \\ 0.6854 & 13.5358 \end{pmatrix}.$$

Note that within traits the residuals sum to zero. Also, that there are BLUP estimates of the residuals for the missing trait 2 observations.

Once the residuals are calculated for all animals, then calculate

$$\mathbf{E}'\mathbf{E} = \begin{pmatrix} 72.4247 & 19.2601 \\ 19.2601 & 957.1828 \end{pmatrix}.$$

Let

$$\mathbf{S}_e = (\mathbf{E}'\mathbf{E} + \nu_e \mathbf{R}_e),$$

which is then inverted and a Cholesky decomposition is applied to the inverse, i.e.

$$\mathbf{L}_e \mathbf{L}_e' = \mathbf{S}_e^{-1},$$

where $\mathbf{L}_e$ is supplied to a Wishart distribution random number generator to give a new sample matrix for the inverse of the residual variances and covariances, $\mathbf{R}^{-1}$.

The sampling process must be repeated until it converges to be samples from the posterior distribution. This may take longer than for scalar estimation of variances.

Chapter 10

Non-Additive Genetic Models

10.1 Introduction

In most animal breeding applications, only additive genetic effects are considered in the evaluation of animals. An infinitesimal animal model is assumed, where animals have been randomly mating. Application of non-additive genetic models has been limited because of difficulties in 1) computing dominance genetic relationships among animals in large populations; 2) computing the inverse of the dominance genetic covariance matrix (and any other epistatic covariance matrices that could be created); and 3) constructing and solving Henderson's MME which increase in size equal to the number of animal equations for each non-additive effect included in the model. These notes describe methods that avoid the problems in 2) and 3) above. Dominance genetic relationships can be calculated using a genomic (or gametic) relationship matrix which provides a number that can be used, but there is a limit to the number of pairs of animals for which a dominance relationship could be computed.

10.2 The Model

If non-additive genetic effects are included in an animal model, then the assumption of random mating is still required. Otherwise non-zero covariances can arise between additive and dominance genetic effects. Thus, the model in these notes is based on approximations (as is any model). Consider a simple animal model with additive, dominance, and additive by dominance genetic effects, and repeated observations per animal, i.e.,

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

where μ is the overall mean, a_i is the additive genetic effect of animal i, d_i is the dominance genetic effect of animal i, $(ad)_i$ is the additive by dominance genetic effect of animal i, p_i

is the permanent environmental effect for an animal with records, and e_i is the residual effect. Also,

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

10.3 Example Data

Below are data and pedigrees for 16 animals. Only animals 11 to 16 have observations.

Animal	Sire	Dam	1	2	3	4
1	0	0				
2	0	0				
3	0	0				
4	0	0				
5	1	2				
6	3	4				
7	3	2				
8	1	4				
9	5	8				
10	7	6				
11	5	6	69	66	56	
12	5	6	29	6		
13	5	6	1			
14	7	8	44	103	85	70
15	7	8	35	77	68	
16	9	10	19	20	33	24

Assume initially that

$$\begin{aligned} \sigma^2_{10} &= 324, \quad \sigma^2_{01} = 169, \\ \sigma^2_{11} &= 49, \quad \sigma^2_p = 144, \\ \sigma^2_e &= 400. \end{aligned}$$

10.3.1 Genetic Relationships

The first step is to compute the additive and dominance genetic relationship matrices for these 16 animals. To do this construct the genomic relationship matrix, which will be of

order 32 (2 times the number of animals). The entire matrix is not given, but only for the animals with records.

Part of the Genomic Relationship Matrix
for the Example Data (All numbers times 16).

	11A	11B	12A	12B	13A	13B	14A	14B	15A	15B	16A	16B
11A	16	0	8	0	8	0	2	2	2	2	5	1
11B	0	16	0	8	0	8	2	2	2	2	1	5
12A	8	0	16	0	8	0	2	2	2	2	5	1
12B	0	8	0	16	0	8	2	2	2	2	1	5
13A	8	0	8	0	16	0	2	2	2	2	5	1
13B	0	8	0	8	0	16	2	2	2	2	1	5
14A	2	2	2	2	2	2	16	0	8	0	1	5
14B	2	2	2	2	2	2	0	16	0	8	5	1
15A	2	2	2	2	2	2	8	0	16	0	1	5
15B	2	2	2	2	2	2	0	8	0	16	5	1
16A	5	1	5	1	5	1	1	5	1	5	16	1
16B	1	5	1	5	1	5	5	1	5	1	1	16

Recall that the additive genetic relationship between two individuals is calculated by adding the four numbers in a block and dividing by 2. Thus, the additive genetic relationship between animals 11 and 16 is

$$a_{11,16} = 0.5 * (\frac{5}{16} + \frac{1}{16} + \frac{1}{16} + \frac{5}{16}) = \frac{3}{8}.$$

The dominance genetic relationship is given by multiplying together the opposite corners of a block and adding the two results together. Hence for animals 11 and 16,

$$d_{11,16} = (\frac{5}{16} * \frac{5}{16}) + (\frac{1}{16} * \frac{1}{16}) = \frac{25}{256} + \frac{1}{256} = \frac{26}{256} = \frac{13}{128}.$$

The complete **A** and **D** for all 16 animals follow.

$$\mathbf{A} = \frac{1}{16}\begin{pmatrix}
16 & 0 & 0 & 0 & 8 & 0 & 0 & 8 & 8 & 0 & 4 & 4 & 4 & 4 & 4 & 4 \\
0 & 16 & 0 & 0 & 8 & 0 & 8 & 0 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\
0 & 0 & 16 & 0 & 0 & 8 & 8 & 0 & 0 & 8 & 4 & 4 & 4 & 4 & 4 & 4 \\
0 & 0 & 0 & 16 & 0 & 8 & 0 & 8 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\
8 & 8 & 0 & 0 & 16 & 0 & 4 & 4 & 10 & 2 & 8 & 8 & 8 & 4 & 4 & 6 \\
0 & 0 & 8 & 8 & 0 & 16 & 4 & 4 & 2 & 10 & 8 & 8 & 8 & 4 & 4 & 6 \\
0 & 8 & 8 & 0 & 4 & 4 & 16 & 0 & 2 & 10 & 4 & 4 & 4 & 8 & 8 & 6 \\
8 & 0 & 0 & 8 & 4 & 4 & 0 & 16 & 10 & 2 & 4 & 4 & 4 & 8 & 8 & 6 \\
8 & 4 & 0 & 4 & 10 & 2 & 2 & 10 & 18 & 2 & 6 & 6 & 6 & 6 & 6 & 10 \\
0 & 4 & 8 & 4 & 2 & 10 & 10 & 2 & 2 & 18 & 6 & 6 & 6 & 6 & 6 & 10 \\
4 & 4 & 4 & 4 & 8 & 8 & 4 & 4 & 6 & 6 & 16 & 8 & 8 & 4 & 4 & 6 \\
4 & 4 & 4 & 4 & 8 & 8 & 4 & 4 & 6 & 6 & 8 & 16 & 8 & 4 & 4 & 6 \\
4 & 4 & 4 & 4 & 8 & 8 & 4 & 4 & 6 & 6 & 8 & 8 & 16 & 4 & 4 & 6 \\
4 & 4 & 4 & 4 & 4 & 4 & 8 & 8 & 6 & 6 & 4 & 4 & 4 & 16 & 8 & 6 \\
4 & 4 & 4 & 4 & 4 & 4 & 8 & 8 & 6 & 6 & 4 & 4 & 4 & 8 & 16 & 6 \\
4 & 4 & 4 & 4 & 6 & 6 & 6 & 6 & 10 & 10 & 6 & 6 & 6 & 6 & 6 & 17
\end{pmatrix},$$

and

$$\mathbf{D} = \frac{1}{256}\begin{pmatrix}
256 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 32 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 256 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 \\
0 & 0 & 256 & 0 & 0 & 0 & 0 & 0 & 0 & 32 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 256 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 \\
0 & 0 & 0 & 0 & 256 & 0 & 0 & 0 & 32 & 0 & 0 & 0 & 0 & 16 & 16 & 12 \\
0 & 0 & 0 & 0 & 0 & 256 & 0 & 0 & 0 & 32 & 0 & 0 & 0 & 16 & 16 & 12 \\
0 & 0 & 0 & 0 & 0 & 0 & 256 & 0 & 0 & 32 & 16 & 16 & 16 & 0 & 0 & 12 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 256 & 32 & 0 & 16 & 16 & 16 & 0 & 0 & 12 \\
32 & 0 & 0 & 0 & 32 & 0 & 0 & 32 & 260 & 4 & 16 & 16 & 16 & 16 & 16 & 18 \\
0 & 0 & 32 & 0 & 0 & 32 & 32 & 0 & 4 & 260 & 16 & 16 & 16 & 16 & 16 & 18 \\
0 & 0 & 0 & 0 & 0 & 0 & 16 & 16 & 16 & 16 & 256 & 64 & 64 & 8 & 8 & 26 \\
0 & 0 & 0 & 0 & 0 & 0 & 16 & 16 & 16 & 16 & 64 & 256 & 64 & 8 & 8 & 26 \\
0 & 0 & 0 & 0 & 0 & 0 & 16 & 16 & 16 & 16 & 64 & 64 & 256 & 8 & 8 & 26 \\
0 & 0 & 0 & 0 & 16 & 16 & 0 & 0 & 16 & 16 & 8 & 8 & 8 & 256 & 64 & 26 \\
0 & 0 & 0 & 0 & 16 & 16 & 0 & 0 & 16 & 16 & 8 & 8 & 8 & 64 & 256 & 26 \\
0 & 8 & 0 & 8 & 12 & 12 & 12 & 12 & 18 & 18 & 26 & 26 & 26 & 26 & 26 & 257
\end{pmatrix}.$$

From these the matrix $\mathbf{A}\#\mathbf{D}$ can be derived, as

$$\mathbf{A}\#\mathbf{D} = \frac{1}{4096}\begin{pmatrix}
4096 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 4096 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 4096 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 4096 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 4096 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 4096 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 4096 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
256 & 0 & 0 & 0 & 320 & 0 & 0 \\
0 & 0 & 256 & 0 & 0 & 320 & 320 \\
0 & 0 & 0 & 0 & 0 & 0 & 64 \\
0 & 0 & 0 & 0 & 0 & 0 & 64 \\
0 & 0 & 0 & 0 & 0 & 0 & 64 \\
0 & 0 & 0 & 0 & 64 & 64 & 0 \\
0 & 0 & 0 & 0 & 64 & 64 & 0 \\
0 & 32 & 0 & 32 & 72 & 72 & 72
\end{pmatrix}$$

$$\begin{pmatrix}
0 & 256 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 32 \\
0 & 0 & 256 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 32 \\
0 & 320 & 0 & 0 & 0 & 0 & 64 & 64 & 72 \\
0 & 0 & 320 & 0 & 0 & 0 & 64 & 64 & 72 \\
0 & 0 & 320 & 64 & 64 & 64 & 0 & 0 & 72 \\
4096 & 320 & 0 & 64 & 64 & 64 & 0 & 0 & 72 \\
320 & 4680 & 8 & 96 & 96 & 96 & 96 & 96 & 180 \\
0 & 8 & 4680 & 96 & 96 & 96 & 96 & 96 & 180 \\
64 & 96 & 96 & 4096 & 512 & 512 & 32 & 32 & 156 \\
64 & 96 & 96 & 512 & 4096 & 512 & 32 & 32 & 156 \\
64 & 96 & 96 & 512 & 512 & 4096 & 32 & 32 & 156 \\
0 & 96 & 96 & 32 & 32 & 32 & 4096 & 512 & 156 \\
0 & 96 & 96 & 32 & 32 & 32 & 512 & 4096 & 156 \\
72 & 180 & 180 & 156 & 156 & 156 & 156 & 156 & 4369
\end{pmatrix}.$$

Once $\mathbf{A}$ and $\mathbf{D}$ have been obtained then any epistatic genetic component covariance matrix can be obtained. For example, for the additive by additive by dominance component would be the result of $\mathbf{A}\#\mathbf{A}\#\mathbf{D}$.

10.3.2 HMME

The appropriate MME needed for the analysis as described by Henderson (1984) would be

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

where $k_{10} = 400/324$, $k_{01} = 400/169$, $k_{11} = 400/49$, and $k_p = 400/144$. Thus, the order is 55 for these 16 animals, with only 17 observations. Note that in the above equations, the inverses of $\mathbf{A}$, $\mathbf{D}$, and $(\mathbf{A}\#\mathbf{D})$ are necessary. Only the inverse of $\mathbf{A}$ can be calculated easily.

For the example data,

$$\mathbf{X'y} = (805),$$

and

$$\mathbf{Z'y} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 191 \\ 35 \\ 1 \\ 302 \\ 180 \\ 96 \end{pmatrix}.$$

The total sum of squares was 52,005.

The solutions are

$$\hat{\mathbf{a}} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -3.77 \\ -3.77 \\ 3.77 \\ 3.77 \\ -2.40 \\ -2.40 \\ 2.98 \\ -8.76 \\ -10.67 \\ 11.00 \\ 6.48 \\ -6.59 \end{pmatrix}, \quad \hat{\mathbf{d}} = \begin{pmatrix} 0 \\ -0.16 \\ 0 \\ -0.16 \\ 0.41 \\ 0.41 \\ -0.57 \\ -0.57 \\ -0.04 \\ -0.04 \\ 3.75 \\ -5.43 \\ -6.92 \\ 7.57 \\ 4.04 \\ -4.51 \end{pmatrix}, \quad \hat{\mathbf{ad}} = \begin{pmatrix} 0 \\ -0.01 \\ 0 \\ -0.01 \\ 0.02 \\ 0.02 \\ -0.05 \\ -0.05 \\ -0.03 \\ -0.03 \\ 1.56 \\ -1.55 \\ -2.05 \\ 2.22 \\ 1.03 \\ -1.49 \end{pmatrix}, \quad \hat{\mathbf{p}} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 6.00 \\ -4.43 \\ -6.13 \\ 6.42 \\ 2.41 \\ -4.26 \end{pmatrix},$$

and $\hat{\mu} = 43.82$. The total genetic merit of an animal can be estimated by adding together the solutions for the additive, dominance, and additive by dominance genetic values,

$$\hat{\mathbf{g}} = \begin{pmatrix} 0 \\ -0.17 \\ 0 \\ -0.17 \\ -3.33 \\ -3.33 \\ 3.15 \\ 3.15 \\ -2.47 \\ -2.47 \\ 8.29 \\ -15.73 \\ -19.65 \\ 20.79 \\ 11.54 \\ -12.60 \end{pmatrix} = (\hat{\mathbf{a}} + \hat{\mathbf{d}} + \hat{\mathbf{ad}}).$$

On the practical side, the solutions for the individual dominance and additive by dominance solutions should be used in breeding programs, but how? Dominance effects arise due to particular sire-dam matings, and thus, dominance genetic values could be used to determine which matings were better. However, additive by dominance genetic solutions may be less useful. Perhaps the main point is that if non-additive genetic effects are significant, then they should be removed through the model to obtain more accurate estimates of the additive genetic effects, assuming that these have a much larger effect than the non-additive genetic effects.

10.4 Estimation of Variances

Take the MME as shown earlier, i.e.

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

Now subtract the equation for dominance genetic effects from the equation for additive genetic effects, and similarly for the additive by dominance and permanent environmental effects, giving

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

Re-arranging terms, then

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

The only inverse that is needed is for $\mathbf{A}$, and the equations to solve are only as large as the usual animal model MME.

The following Gibbs sampling scheme is proposed for this model. (This is a pseudo Gibbs sampling scheme and may not be appropriate. It does not sample the epistatic genetic effects, assuming that these are merely functions of the additive genetic effects. The solutions for these effects could also be sampled, but the variances of these effects would need to be approximated in some way.)

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

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

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

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

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

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

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

4. Quadratic forms for the variances are needed. Note from the fact that the epistatic genetic effects are functions of $\mathbf{a}$, then so too are the quadratic forms. Let

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

The necessary quadratic forms, are then

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

The computation of these quadratic forms is not very complicated and do not require the inverses of $\mathbf{D}$ or $\mathbf{A}\#\mathbf{D}$.

5. New sample values of the variances are given by dividing the quadratic forms by a random Chi-square variate with appropriate degrees of freedom. If q is the number of animals in the relationship matrices, N_p is the number of animals with records, N is the number of records, and $CHI(\nu)$ is a random Chi-square variate with ν degrees of freedom, then the new samples are

$$\begin{aligned}\sigma_{10}^2 &= \mathbf{a}'\mathbf{A}^{-1}\mathbf{a}/CHI(q)\\ \sigma_{01}^2 &= \mathbf{d}'\mathbf{w}_{01}/CHI(q),\\ \sigma_{11}^2 &= \mathbf{ad}'\mathbf{w}_{11}/CHI(q),\\ \sigma_p^2 &= \mathbf{p}'\mathbf{w}_p/CHI(N_p),\\ \sigma_e^2 &= \mathbf{e}'\mathbf{e}/CHI(N).\end{aligned}$$

6. Form the new variance ratios and begin the process again.

The only step missing to make this a valid Gibbs sampling scheme is the sampling of new values for $\mathbf{d}$, $\mathbf{ad}$, and $\mathbf{p}$. For $\mathbf{p}$ one could use the diagonals of $\mathbf{Z}'\mathbf{Z} + \mathbf{I}k_p$ for the variance of $\hat{\mathbf{p}}$. The problem would be in getting the variances for $\mathbf{d}$ and $\mathbf{ad}$ which would require the inverses of $\mathbf{D}$ and $\mathbf{A}\#\mathbf{D}$, respectively, which are being avoided. These could be approximated by the diagonals of $\mathbf{Z}'\mathbf{Z}+\mathbf{I}k_{01}$ and $\mathbf{Z}'\mathbf{Z}+\mathbf{I}k_{11}$. This might be better than not obtaining any new samples of $\mathbf{d}$ and $\mathbf{ad}$, but comparisons to Gibbs sampling with the full MME using the inverses of $\mathbf{D}$ and $\mathbf{A}\#\mathbf{D}$ need to be made.