SYLLABUS
MIXED MODELS IN QUANTITATIVE GENETICS
SIS, Seattle, 19 – 21 July 2017

INSTRUCTORS:

Guilherme Rosa, Department of Animal Sciences, University of Wisconsin, Madison
grosa@wisc.edu

Bruce Walsh, Department of Ecology & Evolutionary Biology, University of Arizona
jbwalsh@u.arizona.edu

LW = Lynch & Walsh: Genetics and Analysis of Quantitative Traits (book)

WL = Walsh & Lynch: Evolution and Selection of Quantitative Traits (website)
http://nitro.biosci.arizona.edu/zbook/NewVolume_2/newvol2.html

LECTURE SCHEDULE

Wednesday, 19 July

2:00 3:30 pm 1) Introduction to matrix algebra and calculus (Walsh)
Background reading: LW, Chapter 8
Additional reading: LW Appendix 3; WL Appendices 4,5

3:30 3:50 pm Break

3:50 5:20 pm 2) The General Linear Model (Walsh)
Background reading: LW Chapter 8
Additional reading: LW Appendices 3, 4; WL Appendices 2, 3

Thursday, 20 July

8:30 10:00 am 3) Overview and Derivation of the mixed model (Rosa)
Additional reading: LW Chapters 26, 27

10:00 10:20 am Break

10:20 12:00 4) Application: BLUP breeding values (Rosa)

12:00 2:00 pm Lunch

2:00 3:30 pm 5) Application: Genomic selection (Rosa)

3:30 3:50 pm Break

3:50 5:00 pm 6) Application: QTL/association mapping (Walsh)
Additional reading: LW Chapters 14, 16

Friday, 21 July

8:30 10:00 am 7) Application: BLUP maternal genetic (Rosa)

10:00 10:20 am Break

10:20 12:00 8) Application: Associative effects (Walsh)
Additional reading: WL Chapter 18

12:00 2:00 pm Lunch

2:00 3:30 pm 9) More on Mixed Models: estimation of variance components, Gibbs sampling (Rosa)

3:30 3:50 pm Break

3:50 5:00 pm 10) Summary and wrap-up (Walsh and Rosa)

ADDITIONAL BOOKS ON MIXED MODELS IN QUANTITATIVE GENETICS

Bernardo, R. (2010) *Breeding for Quantitative Traits in Plants*, 2nd Edition, Stemma Press.

Christou, P., Savin, R., Costa-Pierce, B. A., Misztal, I. and Whitelaw, C. B. A. (eds.) 2013. *Sustainable Food Production*, Vol 1. Springer.

Gondro, C., van der Werf, J and Hayes, H. (eds.) 2013. *Genome-wide Association Studies and Genomic Prediction*. Humana Press.

Henderson, C. R. 1984. *Applications of Linear Models in Animal Breeding*. University of Guelph Press.

Mrode, R. A. 2014. *Linear Models for the Prediction of Animal Breeding Values*, 3rd Edition, CAB International.

Searle, S. R., Casella, G. and McCulloch, C. E. (2006) *Variance Components*. Willey.

Sorensen, D. and Gianola, D. (2002) *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics*. Springer.

VanVleck, L. D. (1993) *Selection Index and Introduction to Mixed Model Methods*. CRC Press.

Lecture 1: Intro/refresher in Matrix Algebra

Bruce Walsh lecture notes
Introduction to Mixed Models
SISG, Seattle
19 – 21 July 2017

Topics

- Definitions, dimensionality, addition, subtraction
- Matrix multiplication
- Inverses, solving systems of equations
- Quadratic products and covariances
- The multivariate normal distribution
- Eigenstructure
- Basic matrix calculations in R
- The Singular Value Decompositon (SVD)

Matrices: An array of elements

Vectors: A matrix with either one row or one column.

Usually written in bold lowercase, e.g. **a, b, c**

$$\mathbf{a} = \begin{pmatrix} 12 \\ 13 \\ 47 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} 2 & 0 & 5 & 21 \end{pmatrix}$$

Column vector (3 x 1)

Row vector (1 x 4)

Dimensionality of a matrix: r x c (rows x columns)
think of Railroad Car

General Matrices

Usually written in bold uppercase, e.g. **A, C, D**

$$\mathbf{C} = \begin{pmatrix} 3 & 1 & 2 \\ 2 & 5 & 4 \\ 1 & 1 & 2 \end{pmatrix} \quad \mathbf{D} = \begin{pmatrix} 0 & 1 \\ 3 & 4 \\ 2 & 9 \end{pmatrix}$$

(3 x 3)

Square matrix

(3 x 2)

Dimensionality of a matrix: r x c (rows x columns)
think of Railroad Car

A matrix is defined by a list of its elements.
B has ij-th element B_{ij} -- the element in row i and column j

Addition and Subtraction of Matrices

If two matrices have the same dimension (both are r x c), then matrix addition and subtraction simply follows by adding (or subtracting) on an element by element basis

Matrix addition: $(A+B)_{ij} = A_{ij} + B_{ij}$

Matrix subtraction: $(A-B)_{ij} = A_{ij} - B_{ij}$

Examples:

$$\mathbf{A} = \begin{pmatrix} 3 & 0 \\ 1 & 2 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{pmatrix} 4 & 2 \\ 3 & 3 \end{pmatrix} \quad \text{and} \quad \text{D = A-B} = \begin{pmatrix} 2 & -2 \\ -1 & 1 \end{pmatrix}$$

Partitioned Matrices

It will often prove useful to divide (or partition) the elements of a matrix into a matrix whose elements are itself matrices.

$$\mathbf{C} = \begin{pmatrix} 3 & 1 & 2 \\ 2 & 5 & 4 \\ 1 & 1 & 2 \end{pmatrix} = \left(\begin{array}{c|cc} 3 & 1 & 2 \\ \hline 2 & 5 & 4 \\ 1 & 1 & 2 \end{array}\right) = \begin{pmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{d} & \mathbf{B} \end{pmatrix}$$

$$\mathbf{a} = (3), \quad \mathbf{b} = (1 \quad 2), \quad \mathbf{d} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 5 & 4 \\ 1 & 2 \end{pmatrix}$$

One useful partition is to write the matrix as either a row vector of column vectors or a column vector of row vectors

$$\mathbf{C} = \begin{pmatrix} 3 & 1 & 2 \\ 2 & 5 & 4 \\ 1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{pmatrix}$$

A column vector whose elements are row vectors

$$\mathbf{r}_1 = (3 \quad 1 \quad 2)$$
$$\mathbf{r}_2 = (2 \quad 5 \quad 4)$$
$$\mathbf{r}_3 = (1 \quad 1 \quad 2)$$

$$\mathbf{C} = \begin{pmatrix} 3 & 1 & 2 \\ 2 & 5 & 4 \\ 1 & 1 & 2 \end{pmatrix} = (\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{c}_3)$$

A row vector whose elements are column vectors

$$\mathbf{c}_1 = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}, \quad \mathbf{c}_2 = \begin{pmatrix} 1 \\ 5 \\ 1 \end{pmatrix}, \quad \mathbf{c}_3 = \begin{pmatrix} 2 \\ 4 \\ 2 \end{pmatrix}$$

Towards Matrix Multiplication: dot products

The dot (or inner) product of two vectors (both of length n) is defined as follows:

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i$$

Example:

$$\mathbf{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \quad \text{and} \quad \mathbf{b} = (4 \quad 5 \quad 7 \quad 9)$$

a ·b = 1*4 + 2*5 + 3*7 + 4*9 = 60

Matrices are compact ways to write systems of equations

$$
\begin{aligned}
5x_1 + 6x_2 + 4x_3 &= 6 \\
7x_1 - 3x_2 + 5x_3 &= -9 \\
-x_1 - x_2 + 6x_3 &= 12
\end{aligned}
$$

$$
\begin{pmatrix} 5 & 6 & 4 \\ 7 & -3 & 5 \\ -1 & -1 & 6 \end{pmatrix}
\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}
=
\begin{pmatrix} 6 \\ -9 \\ 12 \end{pmatrix}
$$

$$
\mathbf{Ax} = \mathbf{c}, \qquad \text{or} \quad \mathbf{x} = \mathbf{A}^{-1}\mathbf{c}
$$

The least-squares solution for the linear model

$$y = \mu + \beta_1 z_1 + \cdots \beta_n z_n$$

yields the following system of equations for the β_i

$$
\begin{aligned}
\sigma(y, z_1) &= \beta_1\sigma^2(z_1) + \beta_2\sigma(z_1, z_2) + \cdots + \beta_n\sigma(z_1, z_n) \\
\sigma(y, z_2) &= \beta_1\sigma(z_1, z_2) + \beta_2\sigma^2(z_2) + \cdots + \beta_n\sigma(z_2, z_n) \\
&\vdots \quad\quad \vdots \quad\quad \vdots \quad\quad \ddots \quad\quad \vdots \\
\sigma(y, z_n) &= \beta_1\sigma(z_1, z_n) + \beta_2\sigma(z_2, z_n) + \cdots + \beta_n\sigma^2(z_n)
\end{aligned}
$$

This can be more compactly written in matrix form as

$$
\begin{pmatrix}
\sigma^2(z_1) & \sigma(z_1, z_2) & \dots & \sigma(z_1, z_n) \\
\sigma(z_1, z_2) & \sigma^2(z_2) & \dots & \sigma(z_2, z_n) \\
\vdots & \vdots & \ddots & \vdots \\
\sigma(z_1, z_n) & \sigma(z_2, z_n) & \dots & \sigma^2(z_n)
\end{pmatrix}
\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}
=
\begin{pmatrix} \sigma(y, z_1) \\ \sigma(y, z_2) \\ \vdots \\ \sigma(y, z_n) \end{pmatrix}
$$

X^TX $\quad\quad$ β $\quad\quad$ X^Ty

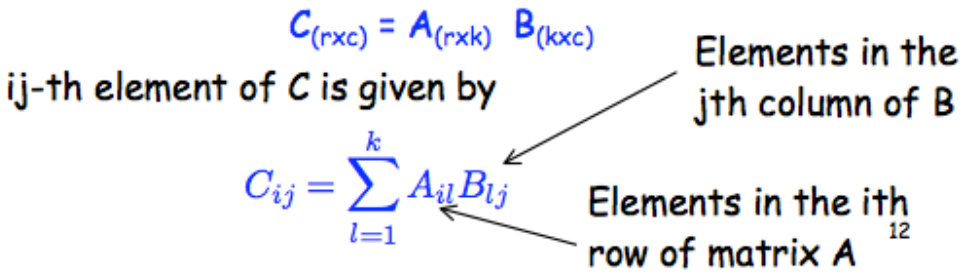
or, $\beta = (X^TX)^{-1} X^Ty$

Matrix Multiplication:

The order in which matrices are multiplied affects the matrix product, e.g. $AB \neq BA$

For the product of two matrices to exist, the matrices must conform. For AB, the number of columns of A must equal the number of rows of B.

The matrix C = AB has the same number of rows as A and the same number of columns as B.



Outer indices given dimensions of resulting matrix, with r rows (A) and c columns (B)

$$C_{(rxc)} = A_{(rxk)}\ B_{(kxc)}$$

Inner indices must match columns of A = rows of B

Example: Is the product ABCD defined? If so, what is its dimensionality? Suppose

$$A_{3x5}\ B_{5x9}\ C_{9x6}\ D_{6x23}$$

Yes, defined, as inner indices match. Result is a 3 x 23 matrix (3 rows, 23 columns)

More formally, consider the product L = MN

Express the matrix M as a column vector of row vectors

$$\mathbf{M} = \begin{pmatrix} \mathbf{m_1} \\ \mathbf{m_2} \\ \vdots \\ \mathbf{m_r} \end{pmatrix} \quad \text{where} \quad \mathbf{m_i} = (M_{i1} \quad M_{i2} \quad \cdots \quad M_{ic})$$

Likewise express N as a row vector of column vectors

$$\mathbf{N} = (\mathbf{n_1} \quad \mathbf{n_2} \quad \cdots \quad \mathbf{n_b}) \quad \text{where} \quad \mathbf{n_j} = \begin{pmatrix} N_{1j} \\ N_{2j} \\ \vdots \\ N_{cj} \end{pmatrix}$$

The ij-th element of L is the inner product of M's row i with N's column j

$$\mathbf{L} = \begin{pmatrix} \mathbf{m_1 \cdot n_1} & \mathbf{m_1 \cdot n_2} & \cdots & \mathbf{m_1 \cdot n_b} \\ \mathbf{m_2 \cdot n_1} & \mathbf{m_2 \cdot n_2} & \cdots & \mathbf{m_2 \cdot n_b} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{m_r \cdot n_1} & \mathbf{m_r \cdot n_2} & \cdots & \mathbf{m_r \cdot n_b} \end{pmatrix}$$

Example

$$\mathbf{AB} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix}$$

Likewise

$$\mathbf{BA} = \begin{pmatrix} ae+cf & eb+df \\ ga+ch & gd+dh \end{pmatrix}$$

ORDER of multiplication matters! Indeed, consider $C_{3x5}\, D_{5x5}$ which gives a 3 x 5 matrix, versus $D_{5x5}\, C_{3x5}$, which is not defined.

Matrix multiplication in R

```
> A<-matrix(c(1,2,3,4),nrow=2)
> B<-matrix(c(4,5,6,7),nrow=2)
> A
     [,1] [,2]
[1,]    1    3
[2,]    2    4
> B
     [,1] [,2]
[1,]    4    6
[2,]    5    7
> A %*% B
     [,1] [,2]
[1,]   19   27
[2,]   28   40
```

R fills in the matrix from the list c by filling in as columns, here with 2 rows (nrow=2)

Entering A or B displays what was entered (always a good thing to check)

The command %*% is the R code for the multiplication of two matrices

On your own: What is the matrix resulting from BA? What is A if nrow=1 or nrow=4 is used?

The Transpose of a Matrix

The transpose of a matrix exchanges the rows and columns, $A^T_{ij} = A_{ji}$

Useful identities

$(AB)^T = B^T A^T$

$(ABC)^T = C^T B^T A^T$

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

Inner product = $a^T b = a^T_{(1 \times n)}\, b_{(n \times 1)}$

Indices match, matrices conform

Dimension of resulting product is 1 X 1 (i.e. a scalar)

$$(a_1 \quad \cdots \quad a_n) \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \mathbf{a}^T\mathbf{b} = \sum_{i=1}^{n} a_i b_i$$

Note that $b^T a = (b^T a)^T = a^T b$

Outer product = $ab^T = a_{(n \times 1)} b^T_{(1 \times n)}$

Resulting product is an n x n matrix

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \begin{pmatrix} b_1 & b_2 & \cdots & b_n \end{pmatrix}$$

$$= \begin{pmatrix} a_1b_1 & a_1b_2 & \dots & a_1b_n \\ a_2b_1 & a_2b_2 & \dots & a_2b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_nb_1 & a_nb_2 & \dots & a_nb_{bn} \end{pmatrix}$$

R code for transposition

```
> t(A)
     [,1] [,2]
[1,]    1    2
[2,]    3    4
```

t(A) = transpose of A

```
> a<-matrix(c(1,2,3),nrow=3)
> a
     [,1]
[1,]    1
[2,]    2
[3,]    3
> t(a) %*% a
     [,1]
[1,]   14
> a %*% t(a)
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    2    4    6
[3,]    3    6    9
```

Enter the column vector a

Compute inner product $\mathbf{a}^T\mathbf{a}$

Compute outer product $\mathbf{a}\mathbf{a}^T$

Solving equations

- The identity matrix I
  - Serves the same role as 1 in scalar algebra, e.g., a*1=1*a =a, with AI=IA= A
- The inverse matrix A^{-1} (IF it exists)
  - Defined by $A A^{-1} = I$, $A^{-1}A = I$
  - Serves the same role as scalar division
    - To solve ax = c, multiply both sides by (1/a) to give:
    - (1/a)*ax = (1/a)c or (1/a)*a*x = 1*x = x,
    - Hence x = (1/a)c
    - To solve Ax = c, $A^{-1}Ax = A^{-1} c$
    - Or $A^{-1}Ax = Ix = x = A^{-1} c$

The Identity Matrix, I

The identity matrix serves the role of the number 1 in matrix multiplication: AI =A, IA = A

I is a square diagonal matrix, with all diagonal elements being one, all off-diagonal elements zero.

$$I_{ij} = \begin{array}{l} 1 \text{ for } i = j \\ 0 \text{ otherwise} \end{array}$$

$$\mathbf{I}_{3x3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The Identity Matrix in R

diag(k), where k is an integer, return the k x k I matix

```
> I<-diag(4)
> I
     [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0    1    0    0
[3,]    0    0    1    0
[4,]    0    0    0    1
> I2 <-diag(2)
> I2
     [,1] [,2]
[1,]    1    0
[2,]    0    1
```


The Inverse Matrix, A^{-1}

For a square matrix A, define its Inverse A^{-1}, as the matrix satisfying

$$\mathbf{A^{-1}A = AA^{-1} = I}$$

For $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ $\mathbf{A}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

If this quantity (the determinant) is zero, the inverse does not exist.

If det(A) is not zero, A^{-1} exists and A is said to be non-singular. If det(A) = 0, A is singular, and no *unique* inverse exists (generalized inverses do)

Generalized inverses, and their uses in solving systems of equations, are discussed in Appendix 3 of Lynch & Walsh

A^- is the typical notation to denote the G-inverse of a matrix

When a G-inverse is used, *provided* the system is consistent, then some of the variables have a family of solutions (e.g., $x_1 = 2$, but $x_2 + x_3 = 6$)

Inversion in R

solve(A) computes A^{-1}

det(A) computes determinant of A

Using A entered earlier

```
> A
     [,1] [,2]
[1,]    1    3
[2,]    2    4
```

Compute A^{-1}

```
> solve(A)
     [,1] [,2]
[1,]   -2  1.5
[2,]    1 -0.5
```

Showing that $A^{-1} A = I$

```
> solve(A) %*% A
     [,1]          [,2]
[1,]    1 -8.881784e-16
[2,]    0  1.000000e+00
```

Computing determinant of A

```
> det(A)
[1] -2
```

Homework

Put the following system of equations in matrix form, and solve using R

$$
\begin{aligned}
3x_1 + 4x_2 + 4x_3 + 6x_4 &= -10 \\
9x_1 + 2x_2 - x_3 - 6x_4 &= 20 \\
x_1 + x_2 + x_3 - 10x_4 &= 2 \\
2x_1 + 9x_2 + 2x_3 - x_4 &= -10
\end{aligned}
$$

Example: solve the OLS for β in $y = \alpha + \beta_1 z_1 + \beta_2 z_2 + e$

$$\beta = \mathbf{V}^{-1}\,\mathbf{c} \qquad \mathbf{c} = \begin{pmatrix} \sigma(y, z_1) \\ \sigma(y, z_2) \end{pmatrix} \qquad \mathbf{V} = \begin{pmatrix} \sigma^2(z_1) & \sigma(z_1, z_2) \\ \sigma(z_1, z_2) & \sigma^2(z_2) \end{pmatrix}$$

It is more compact to use $\sigma(z_1, z_2) = \rho_{12}\,\sigma(z_1)\sigma(z_2)$

$$\mathbf{V}^{-1} = \frac{1}{\sigma^2(z_1)\sigma^2(z_2)\,(1-\rho_{12}^2)} \begin{pmatrix} \sigma^2(z_2) & -\sigma(z_1, z_2) \\ -\sigma(z_1, z_2) & \sigma^2(z_1) \end{pmatrix}$$

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \frac{1}{\sigma^2(z_1)\sigma^2(z_2)\,(1-\rho_{12}^2)} \begin{pmatrix} \sigma^2(z_2) & -\sigma(z_1, z_2) \\ -\sigma(z_1, z_2) & \sigma^2(z_1) \end{pmatrix} \begin{pmatrix} \sigma(y, z_1) \\ \sigma(y, z_2) \end{pmatrix}$$

$$\beta_1 = \frac{1}{1-\rho_{12}^2}\left[\frac{\sigma(y,z_1)}{\sigma^2(z_1)} - \rho_{12}\frac{\sigma(y,z_2)}{\sigma(z_1)\sigma(z_2)}\right]$$

$$\beta_2 = \frac{1}{1-\rho_{12}^2}\left[\frac{\sigma(y,z_2)}{\sigma^2(z_2)} - \rho_{12}\frac{\sigma(y,z_1)}{\sigma(z_1)\sigma(z_2)}\right]$$

If $\rho_{12} = 0$, these reduce to the two univariate slopes,

$$\beta_1 = \frac{\sigma(y,z_1)}{\sigma^2(z_1)} \quad \text{and} \quad \beta_2 = \frac{\sigma(y,z_2)}{\sigma^2(z_2)}$$

Likewise, if $\rho_{12} = 1$, this reduces to a univariate regression,

Useful identities

$(A^T)^{-1} = (A^{-1})^T$

$(AB)^{-1} = B^{-1} A^{-1}$

For a diagonal matrix **D**, then det (**D**), which is also denoted by |**D**|, = product of the diagonal elements

Also, the determinant of any square matrix **A**, det(**A**), is simply the product of the eigenvalues λ of **A**, which statisfy

$$\mathbf{Ae} = \lambda\mathbf{e}$$

If A is n x n, solutions to λ are an n-degree polynomial. **e** is the eigenvector associated with λ. If any of the roots to the equation are zero, A^{-1} is not defined. In this case, for some linear combination **b**, we have **Ab** = 0.

Variance-Covariance matrix

- A very important square matrix is the variance-covariance matrix **V** associated with a vector **x** of random variables.
- $V_{ij} = Cov(x_i,x_j)$, so that the i-th diagonal element of **V** is the variance of x_i, and off-diagonal elements are covariances
- **V** is a symmetric, square matrix

The trace

The trace, tr(**A**) or trace(**A**), of a square matrix A is simply the sum of its diagonal elements

The importance of the trace is that it equals the sum of the eigenvalues of **A**, $tr(\mathbf{A}) = \Sigma \lambda_i$

For a covariance matrix **V**, tr(V) measures the total amount of variation in the variables

λ_i / tr(**V**) is the fraction of the total variation in x contained in the linear combination $\mathbf{e}_i^T\mathbf{x}$, where $\mathbf{e}_i$, the i-th principal component of **V** is also the i-th eigenvector of **V** ($\mathbf{V}\mathbf{e}_i = \lambda_i \mathbf{e}_i$)

Eigenstructure in R

`eigen(A)` returns the eigenvalues and vectors of A

```
> V<-matrix(c(10,-5,10,-5,20,0,10,0,30),nrow=3)
> V
     [,1] [,2] [,3]
[1,]   10   -5   10
[2,]   -5   20    0
[3,]   10    0   30
> eigen(V)
$values
[1] 34.410103 21.117310  4.472587

$vectors
           [,1]       [,2]       [,3]
[1,]  0.3996151  0.2117936  0.8918807
[2,] -0.1386580 -0.9477830  0.2871955
[3,]  0.9061356 -0.2384340 -0.3493816
```

PC 1

Trace = 60

PC 1 accounts for 34.4/60 = 57% of all the variation

$0.400^* x_1 - 0.139^* x_2 + 0.906^* x_3$

Quadratic and Bilinear Forms

Quadratic product: for $A_{n \times n}$ and $x_{n \times 1}$

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$$ Scalar (1 x 1)

Bilinear Form (generalization of quadratic product)
for $A_{m \times n}$, $a_{n \times 1}$, $b_{m \times 1}$ their bilinear form is $b^T_{1 \times m} A_{m \times n} a_{n \times 1}$

$$\mathbf{b}^T\mathbf{A}\mathbf{a} = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}b_i a_j$$

Note that $b^T A\, a = a^T A^T b$

Covariance Matrices for Transformed Variables

What is the variance of the linear combination, $c_1x_1 + c_2x_2 + \ldots + c_nx_n$? (note this is a scalar)

$$\begin{aligned}\sigma^2(\mathbf{c}^T\mathbf{x}) &= \sigma^2\left(\sum_{i=1}^{n} c_i x_i\right) = \sigma\left(\sum_{i=1}^{n} c_i x_i, \sum_{j=1}^{n} c_j x_j\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sigma(c_i x_i, c_j x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \sigma(x_i, x_j) \\ &= \mathbf{c}^T\mathbf{V}\mathbf{c}\end{aligned}$$

Likewise, the covariance between two linear combinations can be expressed as a bilinear form,

$$\sigma(\mathbf{a}^T\mathbf{x}, \mathbf{b}^T\mathbf{x}) = \mathbf{a}^T\mathbf{V}\,\mathbf{b}$$

Example: Suppose the variances of x_1, x_2, and x_3 are 10, 20, and 30. x_1 and x_2 have a covariance of -5, x_1 and x_3 of 10, while x_2 and x_3 are uncorrelated.

What are the variances of the indices $y_1 = x_1-2x_2+5x_3$ and $y_2 = 6x_2-4x_3$?

$$\mathbf{V} = \begin{pmatrix} 10 & -5 & 10 \\ -5 & 20 & 0 \\ 10 & 0 & 30 \end{pmatrix}, \quad \mathbf{c}_1 = \begin{pmatrix} 1 \\ -2 \\ 5 \end{pmatrix}, \quad \mathbf{c}_2 = \begin{pmatrix} 0 \\ 6 \\ -4 \end{pmatrix}$$

$\text{Var}(y_1) = \text{Var}(c_1^T x) = c_1^T \text{Var}(x)\, c_1 = 960$

$\text{Var}(y_2) = \text{Var}(c_2^T x) = c_2^T \text{Var}(x)\, c_2 = 1200$

$\text{Cov}(y_1,y_2) = \text{Cov}(c_1^T x, c_2^T x) = c_1^T \text{Var}(x)\, c_2 = -910$

Homework: use R to compute the above values

The Multivariate Normal Distribution (MVN)

Consider the pdf for n independent normal random variables, the ith of which has mean μ_i and variance σ^2_i

$$p(\mathbf{x}) = \prod_{i=1}^{n} (2\pi)^{-1/2} \sigma_i^{-1} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right)$$

$$= (2\pi)^{-n/2} \left(\prod_{i=1}^{n} \sigma_i\right)^{-1} \exp\left(-\sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right)$$

This can be expressed more compactly in matrix form

Define the covariance matrix V for the vector x of the n normal random variable by

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_n^2 \end{pmatrix} \qquad |\mathbf{V}| = \prod_{i=1}^{n} \sigma_i^2$$

Define the mean vector μ by gives

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$$

$$\sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2} = (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

Hence in matrix from the MVN pdf becomes

$$p(\mathbf{x}) = (2\pi)^{-n/2} |\mathbf{V}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]$$

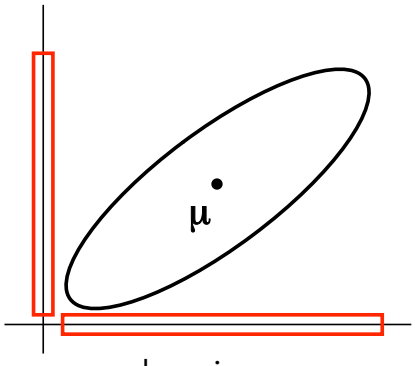
Notice this holds for any vector μ and symmetric positive -definite matrix V, as | V | > 0.

The multivariate normal

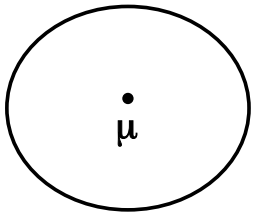
- Just as a univariate normal is defined by its mean and spread, a multivariate normal is defined by its mean vector $\boldsymbol{\mu}$ (also called the centroid) and variance -covariance matrix $\mathbf{V}$

Vector of means $\boldsymbol{\mu}$ determines location

Spread (geometry) about $\boldsymbol{\mu}$ determined by $\mathbf{V}$



x_1, x_2 equal variances, positively correlated

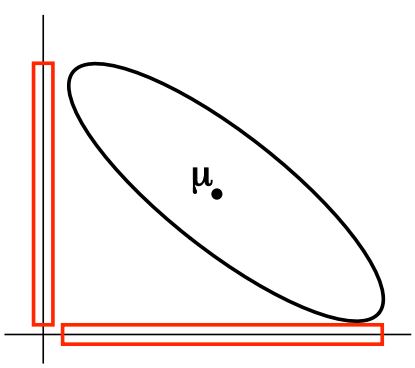


x_1, x_2 equal variances, uncorrelated

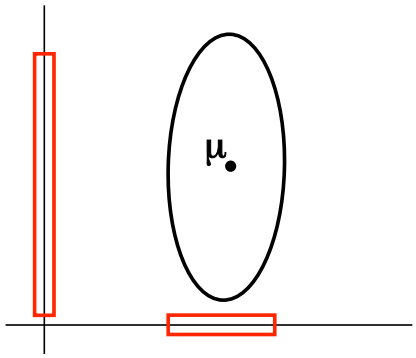
Eigenstructure (the eigenvectors and their corresponding eigenvalues) determines the geometry of $\mathbf{V}$.

Vector of means μ determines location

Spread (geometry) about μ determined by $\mathbf{V}$



x_1, x_2 equal variances, negatively correlated

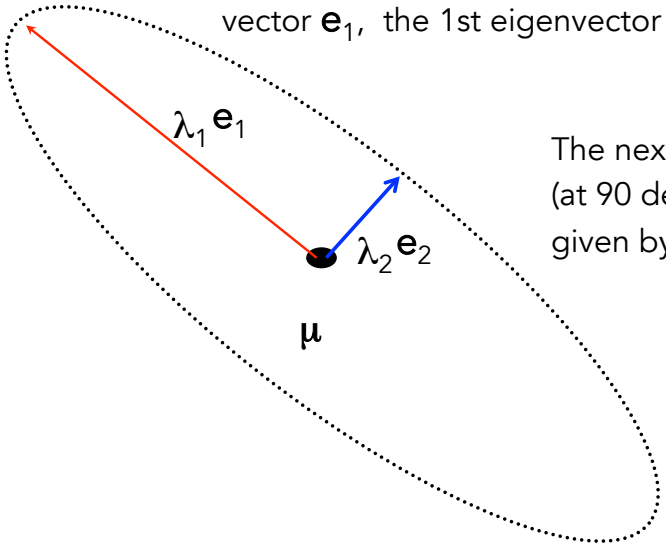


$Var(x_1) < Var(x_2)$, uncorrelated

Positive tilt = positive correlations
Negative tilt = negative correlation
No tilt = uncorrelated

Eigenstructure of V

The direction of the largest axis of variation is given by the unit-length vector $\mathbf{e}_1$, the 1st eigenvector of V.



The next largest axis of orthogonal (at 90 degrees from) $\mathbf{e}_1$, is given by $\mathbf{e}_2$, the 2nd eigenvector

Principal components

- The principal components (or PCs) of a covariance matrix define the axes of variation.
  - PC1 is the direction (linear combination $\mathbf{c}^T\mathbf{x}$) that explains the most variation.
  - PC2 is the next largest direction (at 90degree from PC1), and so on
- PC_i = ith eigenvector of V
- Fraction of variation accounted for by PCi = λ_i / trace(V)
- If V has a few large eigenvalues, most of the variation is distributed along a few linear combinations (axis of variation)
- The singular value decomposition is the generalization of this idea to nonsquare matrices

Properties of the MVN - I

1) If **x** is MVN, any subset of the variables in **x** is also MVN

2) If **x** is MVN, any linear combination of the elements of **x** is also MVN. If **x** ~ MVN(μ,V)

$$\text{for} \quad \mathbf{y} = \mathbf{x} + \mathbf{a}, \qquad \mathbf{y} \text{ is } \mathrm{MVN}_n(\boldsymbol{\mu} + \mathbf{a}, \mathbf{V})$$

$$\text{for} \quad y = \mathbf{a}^T\mathbf{x} = \sum_{k=1}^{n} a_i x_i, \qquad y \text{ is } \mathrm{N}(\mathbf{a}^T\boldsymbol{\mu}, \mathbf{a}^T\mathbf{V}\,\mathbf{a})$$

$$\text{for} \quad \mathbf{y} = \mathbf{A}\mathbf{x}, \qquad \mathbf{y} \text{ is } \mathrm{MVN}_m\left(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}^T\mathbf{V}\mathbf{A}\right)$$

Properties of the MVN - II

3) Conditional distributions are also MVN. Partition x into two components, x_1 (m dimensional column vector) and x_2 (n-m dimensional column vector)

$$\mathbf{x} = \begin{pmatrix} \mathbf{x_1} \\ \mathbf{x_2} \end{pmatrix} \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix} \quad \text{and} \quad \mathbf{V} = \begin{pmatrix} \mathbf{V_{x_1 x_1}} & \mathbf{V_{x_1 x_2}} \\ \mathbf{V^T_{x_1 x_2}} & \mathbf{V_{x_2 x_2}} \end{pmatrix}$$

x_1 | x_2 is MVN with m-dimensional mean vector

$$\boldsymbol{\mu}_{\mathbf{x_1}|\mathbf{x_2}} = \boldsymbol{\mu_1} + \mathbf{V_{x_1 x_2}} \mathbf{V^{-1}_{x_2 x_2}} (\mathbf{x_2} - \boldsymbol{\mu_2})$$

and m x m covariance matrix

$$\mathbf{V}_{\mathbf{x_1}|\mathbf{x_2}} = \mathbf{V_{x_1 x_1}} - \mathbf{V_{x_1 x_2}} \mathbf{V^{-1}_{x_2 x_2}} \mathbf{V^T_{x_1 x_2}}$$

Properties of the MVN - III

4) If x is MVN, the regression of any subset of x on another subset is linear and homoscedastic

$$\mathbf{x_1} = \boldsymbol{\mu}_{\mathbf{x_1}|\mathbf{x_2}} + \mathbf{e}$$
$$= \boldsymbol{\mu_1} + \mathbf{V_{x_1 x_2}} \mathbf{V^{-1}_{x_2 x_2}} (\mathbf{x_2} - \boldsymbol{\mu_2}) + \mathbf{e}$$

Where e is MVN with mean vector 0 and variance-covariance matrix $\mathbf{V}_{\mathbf{x_1}|\mathbf{x_2}}$

$$\boldsymbol{\mu_1} + \mathbf{V_{x_1 x_2}} \mathbf{V}^{-1}_{\mathbf{x_2 x_2}} (\mathbf{x_2} - \boldsymbol{\mu_2}) + \mathbf{e}$$

The regression is linear because it is a linear function of x_2

The regression is homoscedastic because the variance-covariance matrix for e does not depend on the value of the x's

$$\mathbf{V_{x_1|x_2}} = \mathbf{V_{x_1 x_1}} - \mathbf{V_{x_1 x_2}} \mathbf{V}^{-1}_{\mathbf{x_2 x_2}} \mathbf{V}^{T}_{\mathbf{x_1 x_2}}$$

All these matrices are constant, and hence the same for any value of x

Example: Regression of Offspring value on Parental values

Assume the vector of offspring value and the values of both its parents is MVN. Then from the correlations among (outbred) relatives,

$$\begin{pmatrix} z_o \\ z_s \\ z_d \end{pmatrix} \sim \text{MVN} \left[\begin{pmatrix} \mu_o \\ \mu_s \\ \mu_d \end{pmatrix}, \sigma_z^2 \begin{pmatrix} 1 & h^2/2 & h^2/2 \\ h^2/2 & 1 & 0 \\ h^2/2 & 0 & 1 \end{pmatrix} \right]$$

Let $\mathbf{x_1} = (z_o), \quad \mathbf{x_2} = \begin{pmatrix} z_s \\ z_d \end{pmatrix}$

$$\mathbf{V_{x_1,x_1}} = \sigma_z^2, \quad \mathbf{V_{x_1,x_2}} = \frac{h^2 \sigma_z^2}{2} (1 \quad 1), \quad \mathbf{V_{x_2,x_2}} = \sigma_z^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$= \boldsymbol{\mu_1} + \mathbf{V_{x_1 x_2}} \mathbf{V}^{-1}_{\mathbf{x_2 x_2}} (\mathbf{x_2} - \boldsymbol{\mu_2}) + \mathbf{e}$$

Regression of Offspring value on Parental values (cont.)

$$= \boldsymbol{\mu_1} + \mathbf{V_{x_1 x_2}} \mathbf{V^{-1}_{x_2 x_2}} (\mathbf{x_2} - \boldsymbol{\mu_2}) + \mathbf{e}$$

$$\mathbf{V}_{\mathbf{x}_1,\mathbf{x}_1} = \sigma_z^2, \quad \mathbf{V}_{\mathbf{x}_1,\mathbf{x}_2} = \frac{h^2\sigma_z^2}{2}(1 \quad 1), \quad \mathbf{V}_{\mathbf{x}_2,\mathbf{x}_2} = \sigma_z^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Hence,

$$\begin{aligned} z_o &= \mu_o + \frac{h^2\sigma_z^2}{2}(1 \quad 1)\,\sigma_z^{-2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_s - \mu_s \\ z_d - \mu_d \end{pmatrix} + e \\ &= \mu_o + \frac{h^2}{2}(z_s - \mu_s) + \frac{h^2}{2}(z_d - \mu_d) + e \end{aligned}$$

Where e is normal with mean zero and variance

$$\mathbf{V_{x_1|x_2}} = \mathbf{V_{x_1 x_1}} - \mathbf{V_{x_1 x_2}} \mathbf{V^{-1}_{x_2 x_2}} \mathbf{V^{T}_{x_1 x_2}}$$

$$\begin{aligned} \sigma_e^2 &= \sigma_z^2 - \frac{h^2\sigma_z^2}{2}(1 \quad 1)\,\sigma_z^{-2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \frac{h^2\sigma_z^2}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \sigma_z^2 \left(1 - \frac{h^4}{2}\right) \end{aligned}$$

Hence, the regression of offspring trait value given the trait values of its parents is

$z_o = \mu_o + h^2/2(z_s - \mu_s) + h^2/2(z_d - \mu_d) + e$

where the residual e is normal with mean zero and $Var(e) = \sigma_z^2(1-h^4/2)$

Similar logic gives the regression of offspring breeding value on parental breeding value as

$A_o = \mu_o + (A_s - \mu_s)/2 + (A_d - \mu_d)/2 + e$
$= A_s/2 + A_d/2 + e$

where the residual e is normal with mean zero and $Var(e) = \sigma_A^2/2$

The Singular-Value Decomposition (SVD)

An $n \times p$ matrix $\mathbf{A}$ can always be decomposed as the product of three matrices: an $n \times p$ diagonal matrix $\boldsymbol{\Lambda}$ and two unitary matrices, $\mathbf{U}$ which is $n \times n$ and $\mathbf{V}$ which is $p \times p$. The resulting **singular value decomposition** (**SVD**) of $\mathbf{A}$ is given by

$$\mathbf{A}_{n\times p} = \mathbf{U}_{n\times n}\boldsymbol{\Lambda}_{n\times p}\mathbf{V}^T_{p\times p} \tag{39.16a}$$

We have indicated the dimensionality of each matrix to allow the reader to verify that each matrix multiplication conforms. The diagonal elements $\lambda_1, \cdots, \lambda_s$ of $\boldsymbol{\Lambda}$ correspond to the **singular values** of $\mathbf{A}$ and are ordered by decreasing magnitude. Returning to the unitary matrices $\mathbf{U}$ and $\mathbf{V}$, we can write each as a row vector of column vectors,

$$\mathbf{U} = (\mathbf{u}_1, \cdots, \mathbf{u}_i, \cdots \mathbf{u}_n), \qquad \mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_i, \cdots \mathbf{v}_p) \tag{39.16b}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are n and p-dimensional column vectors (often called the **left** and **right singular vectors**, respectively). Since both $\mathbf{U}$ and $\mathbf{V}$ are unitary, by definition (Appendix 4) each column vector has length one and are mutually orthogonal (i.e., if $i \neq j$, $\mathbf{u}_i\mathbf{u}_j^T = \mathbf{v}_i\mathbf{v}_j^T = 0$). Since $\boldsymbol{\Lambda}$ is diagonal, it immediately follows from matrix multiplication that we can write any element in $\mathbf{A}$ as

$$A_{ij} = \sum_{k=1}^{s} \lambda_k \, u_{ik} \, v_{kj} \tag{39.16c}$$

where λ_k is the kth singular value and $s \leq \min(p, n)$ is the number of non-zero singular values.

The importance of the singular value decomposition in the analysis of G×E arises from the **Eckart-Young theorem** (1938), which relates the best approximation of a matrix by some lower-rank (say k) matrix with the SVD. Define as our measure of goodness of fit between a matrix $\mathbf{A}$ and a lower rank approximation $\widehat{\mathbf{A}}$ as the sum of squared differences over all elements,

$$\sum_{ij}(A_{ij} - \hat{A}_{ij})^2$$

Eckart and Young show that the best fitting approximation $\widehat{\mathbf{A}}$ of rank $m < s$ is given from the first m terms of the singular value decomposition (the **rank-m SVD**),

$$\hat{A}_{ij} = \sum_{k=1}^{m} \lambda_k \, u_{ik} \, v_{kj} \tag{39.17a}$$

For example, the best rank-2 approximation for the G×E interaction is given by

$$GE_{ij} \simeq \lambda_1 \, u_{i1} \, v_{j1} + \lambda_2 \, u_{i2} \, v_{j2} \tag{39.17b}$$

where λ_i is the ith singular value of the **GE** matrix, $\mathbf{u}$ and $\mathbf{v}$ are the associated singular vectors (see Example 39.3). The fraction of total variation of a matrix accounted for by taking the first m terms in its SVD is

$$\sum_{k=1}^{m} \lambda_k^2 / \sum_{ij} A_{ij}^2 = \frac{\lambda_1^2 + \cdots + \lambda_m^2}{\lambda_1^2 + \cdots + \lambda_s^2}$$

A data set for soybeans grown in New York (Gauch 1992) gives the GE matrix as

$$\mathbf{GE} = \begin{pmatrix} 57 & 176 & -233 \\ -36 & -196 & 233 \\ -45 & -324 & 369 \\ -66 & 178 & -112 \\ 89 & 165 & -254 \end{pmatrix}$$

Where GE_{ij} = value for Genotype i in envir. j

In **R**, the compact SVD (Equation 39.16d) of a matrix X is given by **svd(X)**, returning the SVD of $\mathbf{GE}$ as

$$\begin{pmatrix} 0.40 & 0.21 & 0.18 \\ -0.41 & 0.00 & 0.91 \\ -0.66 & 0.12 & -0.30 \\ 0.26 & -0.83 & 0.11 \\ 0.41 & 0.50 & 0.19 \end{pmatrix} \begin{pmatrix} 746.10 & 0 & 0 \\ 0 & 131.36 & 0 \\ 0 & 0 & 0.53 \end{pmatrix} \begin{pmatrix} 0.12 & 0.64 & -0.76 \\ 0.81 & -0.51 & -0.30 \\ 0.58 & 0.58 & 0.58 \end{pmatrix}$$

The first singular value accounts for $746.10^2/(743.26^2 + 131.36^2 + 0.53^2) = 97.0\%$ of the total variation of $\mathbf{GE}$, while the second singular value accounts for 3.0%, so that together they account for essentially all of the total variation. The rank-1 SVD approximation of $\mathbf{GE}$ is given by setting all of the diagonal elements of $\boldsymbol{\Lambda}$ except the first entry to zero,

$$\mathbf{GE}_1 = \begin{pmatrix} 0.40 & 0.21 & 0.18 \\ -0.41 & 0.00 & 0.91 \\ -0.66 & 0.12 & -0.30 \\ 0.26 & -0.83 & 0.11 \\ 0.41 & 0.50 & 0.19 \end{pmatrix} \begin{pmatrix} 746.10 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0.12 & 0.64 & -0.76 \\ 0.81 & -0.51 & -0.30 \\ 0.58 & 0.58 & 0.58 \end{pmatrix}$$

Similarly, the rank-2 SVD is given by setting all but the first two singular values to zero,

$$\mathbf{GE}_2 = \begin{pmatrix} 0.40 & 0.21 & 0.18 \\ -0.41 & 0.00 & 0.91 \\ -0.66 & 0.12 & -0.30 \\ 0.26 & -0.83 & 0.11 \\ 0.41 & 0.50 & 0.19 \end{pmatrix} \begin{pmatrix} 746.10 & 0 & 0 \\ 0 & 131.36 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0.12 & 0.64 & -0.76 \\ 0.81 & -0.51 & -0.30 \\ 0.58 & 0.58 & 0.58 \end{pmatrix}$$

For example, the rank-1 SVD approximation for GE_{32} is $g_{31}\lambda_1 e_{12}$ = 746.10*(-0.66)*0.64 = -315

While the rank-2 SVD approximation is $g_{31}\lambda_2 e_{12} + g_{32}\lambda_2 e_{22}$ = 746.10*(-0.66)*0.64 + 131.36* 0.12*(-0.51) = -323

Actual value is -324

Generally, the rank-2 SVD approximation for GE_{ij} is $g_{i1}\lambda_1 e_{1j} + g_{i2}\lambda_2 e_{2j}$

Additional R matrix commands

Operator or Function	Description
A * B	Element-wise multiplication
A %*% B	Matrix multiplication
A %o% B	Outer product. **AB'**
crossprod(A,B) **crossprod(A)**	**A'B** and **A'A** respectively.
t(A)	Transpose
diag(x)	Creates diagonal matrix with elements of **x** in the principal diagonal
diag(A)	Returns a vector containing the elements of the principal diagonal
diag(k)	If k is a scalar, this creates a k x k identity matrix. Go figure.
solve(A, b)	Returns vector **x** in the equation **b = Ax** (i.e., $A^{-1}b$)
solve(A)	Inverse of **A** where A is a square matrix.
ginv(A)	Moore-Penrose Generalized Inverse of **A**. ginv(A) requires loading the **MASS** package.
y<-eigen(A)	**y$val** are the eigenvalues of **A** **y$vec** are the eigenvectors of **A**
y<-svd(A)	Single value decomposition of **A**. **y$d** = vector containing the singular values of **A** **y$u** = matrix with columns contain the left singular vectors of **A** **y$v** = matrix with columns contain the right singular vectors of **A**

Additional R matrix commands (cont)

R <- chol(A)	Choleski factorization of **A**. Returns the upper triangular factor, such that **R'R = A**.
y <- qr(A)	QR decomposition of **A**. **y$qr** has an upper triangle that contains the decomposition and a lower triangle that contains information on the Q decomposition. **y$rank** is the rank of A. **y$qraux** a vector which contains additional information on Q. **y$pivot** contains information on the pivoting strategy used.
cbind(A,B,...)	Combine matrices(vectors) horizontally. Returns a matrix.
rbind(A,B,...)	Combine matrices(vectors) vertically. Returns a matrix.
rowMeans(A)	Returns vector of row means.
rowSums(A)	Returns vector of row sums.
colMeans(A)	Returns vector of column means.
colSums(A)	Returns vector of coumn means.

Additional references

- Lynch & Walsh Chapter 8 (intro to matrices)
- Online notes:
  - Appendix 4 (Matrix geometry)
  - Appendix 5 (Matrix derivatives)

Lecture 2:
Linear and Mixed Models

Bruce Walsh lecture notes
Introduction to Mixed Models
SISG, Seattle
19 – 21 July 2017

Quick Review of the Major Points

The general linear model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

- $\mathbf{y}$ = vector of observed dependent values
- $\mathbf{X}$ = Design matrix: observations of the variables in the assumed linear model
- $\boldsymbol{\beta}$ = vector of unknown parameters to estimate
- $\mathbf{e}$ = vector of residuals (deviation from model fit), $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$

$y = X\beta + e$

Solution to β depends on the ***covariance structure*** (= covariance matrix) of the vector **e** of residuals

Ordinary least squares (OLS)

- OLS: $\mathbf{e} \sim MVN(\mathbf{0}, \sigma^2 \mathbf{I})$
- Residuals are homoscedastic and uncorrelated, so that we can write the cov matrix of **e** as $Cov(e) = \sigma^2 I$
- the OLS estimate, $OLS(\beta) = (X^T X)^{-1} X^T y$

Generalized least squares (GLS)

- GLS: $e \sim MVN(\mathbf{0}, \mathbf{V})$
- Residuals are heteroscedastic and/or dependent,
- $GLS(\beta) = (X^T V^{-1} X)^{-1} V^{-1} X^T y$

BLUE

- Both the OLS and GLS solutions are also called the Best Linear Unbiased Estimator (or BLUE for short)
- Whether the OLS or GLS form is used depends on the assumed covariance structure for the residuals
  - Special case of $Var(\mathbf{e}) = \sigma_e^2 \mathbf{I}$ -- OLS
  - All others, i.e., $Var(\mathbf{e}) = \mathbf{R}$ -- GLS

Linear Models

One tries to explain a dependent variable y as a linear function of a number of independent (or predictor) variables.

A multiple regression is a typical linear model,

$$y = \mu + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_x + e$$

Here e is the residual, or deviation between the true value observed and the value predicted by the linear model.

The (partial) regression coefficients are interpreted as follows: a unit change in x_i while holding all other variables constant results in a change of β_i in y

Linear Models

As with a univariate regression (y = a + bx + e), the model parameters are typically chosen by least squares, wherein they are chosen to minimize the sum of squared residuals, $\Sigma\ e_i^2$

This unweighted sum of squared residuals assumes an OLS error structure, so all residuals are equally weighted (homoscedastic) and uncorrelated

If the residuals differ in variances and/or some are correlated (GLS conditions), then we need to minimize the weighted sum $\mathbf{e}^T\mathbf{V}^{-1}\mathbf{e}$, which removes correlations and gives all residuals equal variance.

Linear Models in Matrix Form

Suppose we have 3 variables in a multiple regression, with four (y,x) vectors of observations.

$$y_i = \mu + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i$$

In matrix form, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ 1 & x_{31} & x_{32} & x_{33} \\ 1 & x_{41} & x_{42} & x_{43} \end{pmatrix} \quad \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix}$$

The design matrix X. Details of both the experimental design and the observed values of the predictor variables all reside solely in X

Rank of the design matrix

- With n observations and p unknowns, X is an n x p matrix, so that X^TX is p x p
- Thus, at most X can provide unique estimates for up to p < n parameters
- The rank of X is the number of independent rows of X. If X is of full rank, then rank = p
- A parameter is said to be estimable if we can provide a unique estimate of it. If the rank of X is k < p, then exactly k parameters are estimable (some as linear combinations, e.g. β_1-3β_3 = 4)
- if $\det(X^TX) = 0$, then X is not of full rank
- Number of nonzero eigenvalues of X^TX gives the rank of X.

Experimental design and X

- The structure of X determines not only which parameters are estimable, but also the expected sample variances, as $Var(\beta) = k (X^TX)^{-1}$
- Experimental design determines the structure of X before an experiment (of course, missing data almost always means the final X is different form the proposed X)
- Different criteria used for an optimal design. Let $V = (X^TX)^{-1}$. The idea is to chose a design for X given the constraints of the experiment that:
  - A-optimality: minimizes tr(V)
  - D-optimality: minimizes det(V)
  - E-optimality: minimizes leading eigenvalue of V

Ordinary Least Squares (OLS)

When the covariance structure of the residuals has a certain form, we solve for the vector β using OLS

If residuals follow a MVN distribution, OLS = ML solution

If the residuals are homoscedastic and uncorrelated, $\sigma^2(e_i) = \sigma_e^2$, $\sigma(e_i,e_j) = 0$. Hence, each residual is equally weighted,

Sum of squared residuals can be written as

$$\sum_{i=1}^{n} \widehat{e}_i^2 = \widehat{\mathbf{e}}^T\widehat{\mathbf{e}} = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$$

Predicted value of the y's

Ordinary Least Squares (OLS)

$$\sum_{i=1}^{n} \widehat{e}_i^2 = \widehat{\mathbf{e}}^T \widehat{\mathbf{e}} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Taking (matrix) derivatives shows this is minimized by

$$\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

This is the OLS estimate of the vector β

The variance-covariance estimate for the sample estimates is

$$\mathbf{V}_{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\sigma_e^2$$

The ij-th element gives the covariance between the estimates of β_i and β_j.

Sample Variances/Covariances

The residual variance can be estimated as

$$\widehat{\sigma_e^2} = \frac{1}{n - \operatorname{rank}(X)} \sum_{i=1}^{n} \widehat{e}_i^2$$

The estimated residual variance can be substituted into

$$\mathbf{V}_{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\sigma_e^2$$

To give an approximation for the sampling variance and covariances of our estimates.

Confidence intervals follow since the vector of estimates ~ MVN($\boldsymbol{\beta}$, $\mathbf{V}_{\boldsymbol{\beta}}$)

Example: Regression Through the Origin

$$y_i = \beta x_i + e_i$$

Here $\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ $\boldsymbol{\beta} = (\beta)$

$$\mathbf{X}^T\mathbf{X} = \sum_{i=1}^{n} x_i^2 \qquad \mathbf{X}^T\mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

$$\beta = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} = \frac{\sum x_i y_i}{\sum x_i^2} \qquad \sigma^2(b) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\sigma_e^2 = \frac{\sigma_e^2}{\sum x_i^2}$$

$$\sigma^2(\beta) = \frac{1}{n-1}\frac{\sum(y_i - \beta x_i)^2}{\sum x_i^2} \qquad \sigma_e^2 = \frac{1}{n-1}\sum(y_i - \beta x_i)^2$$

Polynomial Regressions

GLM can easily handle any function of the observed predictor variables, provided the parameters to estimate are still linear, e.g. $Y = \alpha + \beta_1 f(x) + \beta_2 g(x) + \cdots + e$

Quadratic regression:

$$y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + e_i$$

$$\boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix}$$

Interaction Effects

Interaction terms (e.g. sex x age) are handled similarly

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + e_i$$

$$\boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{11}x_{12} \\ 1 & x_{21} & x_{22} & x_{21}x_{22} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n1}x_{n2} \end{pmatrix}$$

With x_1 held constant, a unit change in x_2 changes y by $\beta_2 + \beta_3 x_1$ (i.e., the slope in x_2 depends on the current value of x_1)

Likewise, a unit change in x_1 changes y by $\beta_1 + \beta_3 x_2$

The GLM lets you build your own model!

- Suppose you want a quadratic regression forced through the origin where the slope of the quadratic term can vary over the sexes (pollen vs. seed parents)
- $Y_i = \beta_1 x_i + \beta_2 x_i^2 + \beta_3 s_i x_i^2$
- s_i is an indicator (0/1) variable for the sex (0 = male, 1 = female).
  - Male slope = β_2,
  - Female slope = $\beta_2 + \beta_3$

Generalized Least Squares (GLS)

Suppose the residuals no longer have the same variance (i.e., display heteroscedasticity). Clearly we do not wish to minimize the *unweighted* sum of squared residuals, because those residuals with smaller variance should receive more weight.

Likewise in the event the residuals are correlated, we also wish to take this into account (i.e., perform a suitable transformation to remove the correlations) before minimizing the sum of squares.

Either of the above settings leads to a GLS solution in place of an OLS solution.

In the GLS setting, the covariance matrix for the vector e of residuals is written as R where $R_{ij} = \sigma(e_i, e_j)$

The linear model becomes $y = X\beta + e$, $cov(e) = R$

The GLS solution for β is

$$\mathbf{b} = \left(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{R}^{-1}\mathbf{y}$$

The variance-covariance of the estimated model parameters is given by

$$\mathbf{V_b} = \left(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\sigma_e^2$$

Model diagnostics

- It's all about the residuals
- Plot the residuals
  - Quick and easy screen for outliers
  - Plot y or yhat on e
- Test for normality among estimated residuals
  - Q-Q plot
  - Wilk-Shapiro test
  - If non-normal, try transformations, such as log

OLS, GLS summary

	OLS	GLS
Assumed distribution of residuals	$\mathbf{e} \sim (\mathbf{0}, \sigma_e^2 \mathbf{I})$	$\mathbf{e} \sim (\mathbf{0}, \mathbf{V})$
Least-squares estimator of $\boldsymbol{\beta}$	$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$	$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$
$\mathrm{Var}(\widehat{\boldsymbol{\beta}})$	$(\mathbf{X}^T\mathbf{X})^{-1}\sigma_e^2$	$(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$
Predicted values, $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$	$\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$	$\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$
$\mathrm{Var}(\widehat{\mathbf{y}})$	$\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma_e^2$	$\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T$

Fixed vs. Random Effects

In linear models are are trying to accomplish two goals: estimation the values of model parameters and estimate any appropriate variances.

For example, in the simplest regression model, $y = \alpha + \beta x + e$, we estimate the values for α and β and also the variance of e. We, of course, can also estimate the $e_i = y_i - (\alpha + \beta x_i)$

Note that α/β are ***fixed constants*** are we trying to estimate (fixed factors or fixed effects), while the e_i values are drawn from some probability distribution (typically Normal with mean 0, variance σ^2_e). The e_i are random effects.

This distinction between fixed and random effects is extremely important in terms of how we analyzed a model. If a parameter is a fixed constant we wish to estimate, it is a fixed effect. If a parameter is drawn from some probability distribution and we are trying to make inferences on either the distribution and/or specific realizations from this distribution, it is a random effect.

We generally speak of estimating fixed factors (BLUE) and predicting random effects (BLUP -- best linear unbiased Predictor)

"Mixed" models (MM) contain both fixed and random factors

$$y = Xb + Zu + e, \quad u \sim MVN(0,R), \; e \sim MVN(0,\sigma^2_e I)$$

Key: need to specify covariance structures for MM

Random effects models

- It is often useful to treat certain effects as random, as opposed to fixed
  - Suppose we have k effects. If we treat these as fixed, we lose k degrees of freedom
  - If we assume each of the k realizations are drawn from a normal with mean zero and unknown variance, only one degree of freedom lost --- that for estimating the variance
    - We can then predict the values of the k realizations

Environmental effects

- Consider yield data measured over several years in a series of plots.
- Standard to treat year-to-year variation at a specific site as being random effects
- Often the plot effects (mean value over years) are also treated as random.
- For example, consider plants group in growing region i, location j within that region, and year (season) k for that location-region effect
  - $E = R_i + L_{ij} + e_{ijk}$
  - Typically R can be a fixed effect, while L and e are random effects, $L_{ik} \sim N(0,\sigma^2_L)$ and $e_{ikj} \sim N(0,\sigma^2_e)$

Random models

- With a random model, one is assuming that all "levels" of a factor are not observed. Rather, some subset of values are drawn from some underlying distribution
  - For example, year to year variation in rainfall at a location. Each year is a random sample from the long-term distribution of rainfall values
  - Typically, assume a functional form for this underlying distribution (e.g., normal with mean 0) and then use observations to estimate the distribution parameters (here, the variance)

Random models (cont)

- Key feature:
  - Only one degree of freedom used (estimate of the variance)
  - Using the fixed effects and the estimated underlying distribution parameters, one then predicts the actual realizations of the individual values (i.e., the year effects)
  - Assumption: the covariance structure among the individual realizations of the realized effects. If only a variance is assume, this implies they are independent. If they are assumed to be correlated, this structure must be estimated.

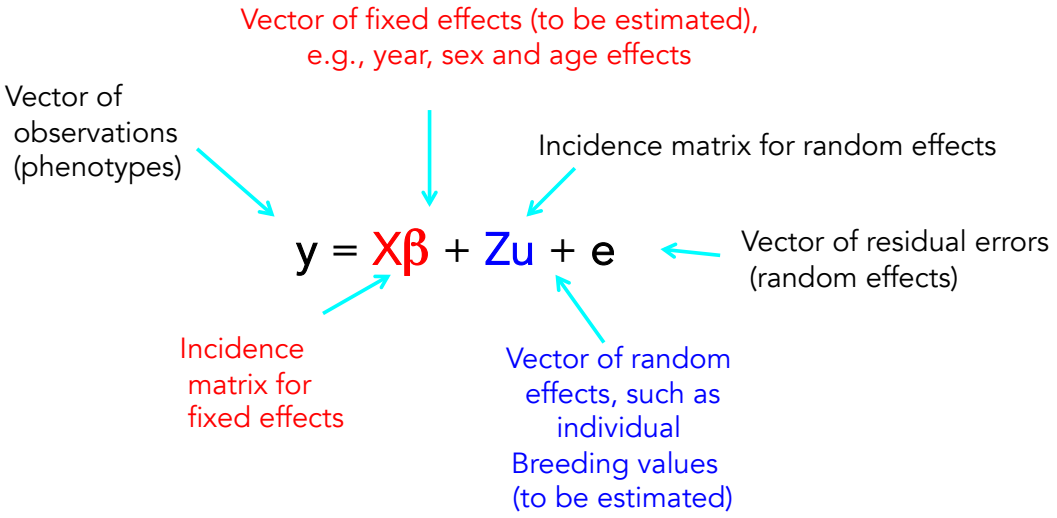
Random models

- Let's go back to treating yearly effects as random
- If assume these are uncorrelated, only use one degree of freedom, but makes assumptions about covariance structure
  - Standard: Uncorrelated
  - Option: some sort of autocorrelation process, say with a yearly decay of r (must also be estimated)
- Conversely, could all be treated as fixed, but would use k degrees of freedom for k years, but no assumptions on their relationships (covariance structure)

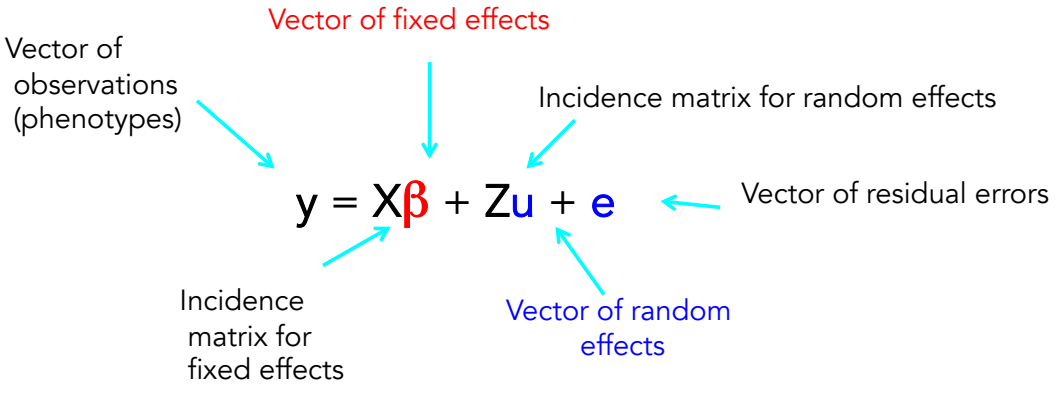
Identifiability

- Recall that a fixed effect is said to be estimable if we can obtain a unique estimate for it (either because X is of full rank or when using a generalized inverse it returns a unique estimate)
  - Lack of estimable arises because the experiment design confounds effects
- The analogous term for random models is identifiability
  - The variance components have unique estimates

The general mixed model



The general mixed model



Observe **y, X, Z.**

Estimate fixed effects **β**

Estimate random effects **u, e**

Means & Variances for $y = X\beta + Zu + e$

Means: $E(u) = E(e) = 0$, $E(y) = X\beta$

Variances:

Let R be the covariance matrix for the residuals. We typically assume $R = \sigma^2_e * I$

Let G be the covariance matrix for the vector **u** of random effects

The covariance matrix for y becomes

$V = ZGZ^T + R$

Hence, $y \sim MVN(X\beta, V)$

Mean $X\beta$ due to fixed effects
Variance V due to random effects

Estimating fixed Effects & Predicting Random Effects

For a mixed model, we observe **y**, X, and Z

β, u, R, and G are generally unknown

Two complementary estimation issues

(i) Estimation of **β** and **u**

$$\hat{\beta} = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$$ Estimation of fixed effects

BLUE = Best Linear Unbiased Estimator

$$\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)$$ Prediction of random effects

BLUP = Best Linear Unbiased Predictor

Recall $V = ZGZ^T + R$

Different statistical models

- GLM = general linear model
  - OLS ordinary least squares: e ~ MVN(0,cI)
  - GLS generalized least squares: e ~ MVN(0,R)
- Mixed models
  - Both fixed and random effects (beyond the residual)
- Mixture models
  - A weighted mixture of distributions
- Generalized linear models
  - Nonlinear functions, non-normality

Mixture models

- Under a mixture model, an observation potentially comes from one of several different distributions, so that the density function is $\pi_1\phi_1 + \pi_2\phi_2 + \pi_3\phi_3$
  - The mixture proportions π_i sum to one
  - The ϕ_i represent different distribution, e.g., normal with mean μ_i and variance σ^2
- Mixture models come up in QTL mapping -- an individual could have QTL genotype QQ, Qq, or qq
  - See Lynch & Walsh Chapter 13
- They also come up in codon models of evolution, were a site may be neutral, deleterious, or advantageous, each with a different distribution of selection coefficients
  - See Walsh & Lynch (volume 2A website), Chapters 10,11

Generalized linear models

The **Generalized Linear Model** (note the **ized** ending) takes this a step further by assuming for some monotonic function g, that

$$E[y_i] = g\left(\mu + \sum_{k=1}^{n} \beta_k x_{ik}\right) \tag{2}$$

In particular, taking the inverse g^{-1} of the function g returns a linear model, with

$$g^{-1}(E[y_i]) = \mu + \sum_{k=1}^{n} \beta_k x_{ik} \tag{3}$$

The function f with the property that expresses the expected value of the response variable as a linear function of the predictor variables, i.e.,

$$f(E[y_i]) = \mu + \sum_{k=1}^{n} \beta_k x_{ik}$$

is called the **link function** of the particular generalized linear model.

Typically assume non-normal distribution for residuals, e.g., Poisson, binomial, gamma, etc

Lecture 3
Overview and Derivation of the Mixed Model

Guilherme J. M. Rosa

University of Wisconsin-Madison

Introduction to Quantitative Genetics
SISG, Seattle
19 – 21 July 2017

OUTLINE

- General Linear Model (fixed effects)
- Maximum Likelihood Estimation
- Linear Mixed Model
- BLUE and BLUP

General Linear Model

(Fixed Effects Model)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- $\mathbf{y}$: responses
- $\mathbf{X}$: design/incidence matrix (known)
- $\boldsymbol{\beta}$: overall mean + fixed effects parameters
- $\boldsymbol{\varepsilon}$: residuals

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, I_n\sigma^2) \rightarrow \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

⇨ Fixed effect: levels included in the study represent all levels about which inference is to be made. Fixed effects models: models containing only fixed effects

Example 1

Experiment to compare growth performance of pigs under two experimental groups (Control and Treatment), with three replications each.

Control	Treatment
53	61
46	66
58	57

Model:

$$y_{ij} = \mu + \delta_i + e_{ij}$$

- y_{ij}: weight gain of pig j of group i
- μ: constant; general mean
- δ_i: effect of group i
- e_{ij}: residual term

4

Matrix Notation

Control	Treatment
53	61
46	66
58	57

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} 53 \\ 46 \\ 58 \\ 61 \\ 66 \\ 57 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \delta_1 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{21} \\ e_{23} \end{bmatrix}$$

5

Alternative Parameterizations

⇨ Equivalent models with different parameterizations

For example, if the average weight gain in each group is expressed as $\mu_i = \mu + \delta_i$, the model becomes:

$$\begin{bmatrix} 53 \\ 46 \\ 58 \\ 61 \\ 66 \\ 57 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{21} \\ e_{23} \end{bmatrix}$$

Alternatively, the model can be expressed in terms of the average weight gain of the Control (μ_1) and the difference on weight gain between the two groups ($\tau = \mu_2 - \mu_1$):

$$\begin{bmatrix} 53 \\ 46 \\ 58 \\ 61 \\ 66 \\ 57 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \tau \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{21} \\ e_{23} \end{bmatrix}$$

Example 2

Flowering time (days, log scale) of *Brassica napus* according to genotype in specific locus, such as a candidate gene

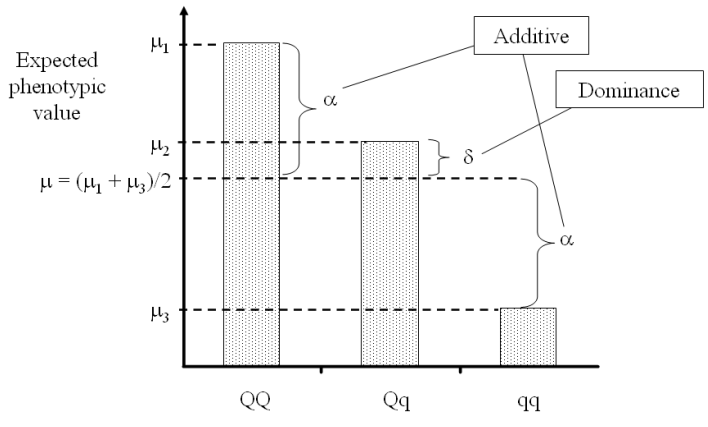
Genotype		
qq	Qq	QQ
3.4	2.9	3.1
3.7	2.5	2.6
3.2		

Model: $y_{ij} = \mu_i + e_{ij}$

- y_{ij}: flowering time of replication j (j = 1,…, n_i) of genotype i (i = qq, Qq and QQ)
- μ_i: expected flowering time of plants of genotype i
- e_{ij}: residual (environment and polygenic effects)

7

⇨ The expected phenotypic values μ_i, however, can be expressed as a function of the additive and dominant effects



Expected phenotypic value according to the genotype on a specific locus.

The model can be written then as:

$$y_{ij} = \mu + x_{ij}\alpha + (1 - |x_{ij}|)\delta + e_{ij}$$

- μ: constant (mid-point flowering time between homozygous genotypes)
- x_{ij}: indicator variable (genotype), coded as -1, 0 and 1 for genotypes qq, Qq and QQ
- α and β: additive and dominance effects

In matrix notation:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 3.4 \\ 3.7 \\ 3.2 \\ 2.9 \\ 2.5 \\ 3.1 \\ 2.6 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & -1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha \\ \delta \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{31} \\ e_{32} \end{bmatrix}$$

Least-Squares Estimation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim (\mathbf{0}, \mathbf{I}_n\sigma^2) \quad \rightarrow \quad \varepsilon_i \overset{iid}{\sim} (0, \sigma^2)$$

An estimate ($\hat{\boldsymbol{\beta}}$) of the vector β can be obtained by the method of least-squares, which aims to minimize the residual sum of squares, given (in matrix notation) by:

$$RSS = \sum_{i=1}^{n} (\hat{\varepsilon}_i)^2 = \hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

Taking the derivatives and equating to zero, it can be shown that the least-squares estimator of β is:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

➲ It is shown that $E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$ and $Var[\hat{\boldsymbol{\beta}}] = (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2$

10

More on the LS Methodology

The estimator $\hat{\boldsymbol{\beta}}_{OLS} = \hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ is called ordinary least squares (OLS) estimator, and it is indicated only in situations with homoscedastic and uncorrelated residuals

If the residual variance is heterogeneous (i.e., $\mathrm{Var}(\boldsymbol{\varepsilon}_i) = \sigma_i^2 = w_i\sigma^2$), the residual variance matrix can be expressed as $\mathrm{Var}(\boldsymbol{\varepsilon}) = \mathbf{W}\sigma^2$, where **W** is a diagonal matrix with the elements w_i, a better estimator of β is given by: $\hat{\boldsymbol{\beta}}_{WLS} = (\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}\mathbf{y}$

which is generally referred to as weighted least squares (WLS) estimator.

Furthermore, in situations with a general residual variance-covariance matrix **V**, including correlated residuals, a generalized least squares (GLS) estimator $\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$ is obtained by minimizing the generalized sum of squares, given by: $GSS = \boldsymbol{\varepsilon}^T\mathbf{V}^{-1}\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$

Maximum Likelihood Estimation

Likelihood Function: any function of the model parameters that is proportional to the density function of the data

Hence, to use a likelihood-based approach for estimating model parameters, some extra assumptions must be made regarding the distribution of the data

In the case of the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, if the residuals are assumed normally distributed with mean vector zero and variance-covariance matrix **V**, i.e. $\boldsymbol{\varepsilon} \sim \mathrm{MVN}(\mathbf{0}, \mathbf{V})$, the response vector **y** is also normally distributed, with expectation $\mathrm{E}[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$ and variance $\mathrm{Var}[\mathbf{y}] = \mathbf{V}$

Maximum Likelihood Estimation

The distribution of **y** has a density function given by:

$$p(\mathbf{y} \mid \boldsymbol{\beta}, V) = (2\pi)^{-n/2} \mid \mathbf{V} \mid^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{T} \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$$

so that the likelihood and the log-likelihood functions can be expressed respectively as:

$$L(\boldsymbol{\beta}, V) \propto \mid \mathbf{V} \mid^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{T} \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$$

and

$$l(\boldsymbol{\beta}, V) = \log[L(\boldsymbol{\beta}, V)] \propto -\frac{1}{2}\log \mid \mathbf{V} \mid -\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{T} \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Maximum Likelihood Estimation

Assuming **V** known, the likelihood equations for **β** are given by taking the first derivatives of l(β,**V**) with respect to **β** and equating it to zero:

$$\frac{\partial l(\boldsymbol{\beta}, V)}{\partial \boldsymbol{\beta}} \equiv \frac{\partial}{\partial \boldsymbol{\beta}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{T} \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0$$

from which the following system of equations is obtained:

$$\mathbf{X}^{T}\mathbf{V}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^{T}\mathbf{V}^{-1}\mathbf{y}$$

The maximum likelihood estimator (MLE) for **β** is given then by:

$$\mathrm{MLE}(\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}} = (\mathbf{X}^{T}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{T}\mathbf{V}^{-1}\mathbf{y}$$

Maximum Likelihood Estimation

If the inverse of $\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}$ does not exist, a generalized inverse $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^-$ can be used to obtain a solution for the system of likelihood equations:

$$\boldsymbol{\beta}^0 = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^- \mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$$

Note: Under normality the MLE coincides with the GLS estimator discussed previously. Similarly, in situations in which the matrix **V** is diagonal, or when **V** can be represented as $\mathbf{V} = \mathbf{I}_n\sigma^2$, the MLE coincides with the WLS and the OLS estimators, respectively

Maximum Likelihood Estimation

The expectation and the variance-covariance matrix of the MLE are given by:

$$E[\hat{\boldsymbol{\beta}}] = E[(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}] = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}E[\mathbf{y}] = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}$$

$$\begin{aligned}\mathrm{Var}[\hat{\boldsymbol{\beta}}] &= \mathrm{Var}[(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}] = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathrm{Var}[\mathbf{y}]\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\end{aligned}$$

As $\hat{\boldsymbol{\beta}}$ is a linear combination of the response vector **y**, we have that $\hat{\boldsymbol{\beta}} \sim \mathrm{MVN}(\boldsymbol{\beta}, (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1})$, from which confidence intervals (regions) and hypothesis testing regarding any (set of) element(s) of $\boldsymbol{\beta}$ can be easily obtained

The estimation of variance and covariance parameters will be discussed later

Maximum Likelihood Estimation

⇨ Note: In the case of the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, with $\boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}, \mathbf{I}\sigma^2)$, it can be shown that:

$$\begin{cases} \hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \quad \rightarrow \quad \hat{\boldsymbol{\beta}} \sim \mathrm{N}(\boldsymbol{\beta}, (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\sigma^2) \\ \hat{\sigma}^2 = \dfrac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \dfrac{1}{n} \| \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \|^2 \end{cases}$$

$$\hat{\sigma}^2 \sim \sigma^2 \frac{\chi^2_{(n-k)}}{n} \qquad \left(\mathrm{E}[\hat{\sigma}^2] = \frac{n-k}{n}\sigma^2 \right)$$

$$\tilde{\sigma}^2 = s^2 = \frac{n}{n-k}\hat{\sigma}^2 = \frac{1}{n-k} \| \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \|^2 \quad \rightarrow \quad \tilde{\sigma}^2 \sim \sigma^2 \frac{\chi^2_{(n-k)}}{n-k}$$

Two-stage Analysis of Longitudinal Data

Step 1

Supposed a series of longitudinal data (e.g., repeated measurements on time) on n individuals. Let y_{ij} represent the observation j ($j = 1,2,\ldots,n_i$) on individual i ($i = 1,2,\ldots,n$), and the following quadratic regression of measurements on time (z_{ij}) for each individual:

$$y_{ij} = \beta_{0i} + \beta_{1i} z_{ij} + \beta_{2i} z_{ij}^2 + \varepsilon_{ij}$$

where β_{0i}, β_{1i} and β_{2i} are subject-specific regression parameters, and ε_{ij} are residual terms, assumed normally distributed with mean zero and variance σ_ε^2

In matrix notation such subject-specific regressions can be expressed as:

$$\mathbf{y}_i = \mathbf{Z}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i \qquad (1)$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})^T$, $\boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i}, \beta_{2i})^T$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in_i})^T \sim N(\mathbf{0}, \mathbf{I}\sigma^2_{\varepsilon})$ and

$$\mathbf{Z}_i = \begin{bmatrix} 1 & z_{i1} & z_{i1}^2 \\ 1 & z_{i2} & z_{i2}^2 \\ \vdots & \vdots & \vdots \\ 1 & z_{in_i} & z_{in_i}^2 \end{bmatrix}$$

19

Under these specifications, it is shown that the least-squares estimate of β_i is:

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{Z}_i^T\mathbf{Z}_i)^{-1}\mathbf{Z}_i^T\mathbf{y}_i$$

Note that this is also the maximum likelihood estimate of β_i

Such estimates can be viewed as summary statistics for the longitudinal data, the same way one could use area under the curve (AUC), or peak (maximum value of y_{ij}), or mean response.

20

Two-stage Analysis of Longitudinal Data
Step 2

Supposed now we are interested on the effect of some other variables (such as gender, treatment, year, etc.) on the values of β_i

Such effects could be studied using a model as:

$$\hat{\boldsymbol{\beta}}_i = \mathbf{W}_i\boldsymbol{\beta} + \mathbf{u}_i$$

where $\mathbf{u}_i \sim N(\mathbf{0},\mathbf{D})$, which is an approximation for the model:

$$\boldsymbol{\beta}_i = \mathbf{W}_i\boldsymbol{\beta} + \mathbf{u}_i \qquad (2)$$

21

Single-stage Analysis of Longitudinal Data

The two step-analysis described here can be merged into a single stage approach by substituting (2) in (1):

$$\mathbf{y}_i = \mathbf{Z}_i[\mathbf{W}_i\boldsymbol{\beta} + \mathbf{u}_i] + \boldsymbol{\varepsilon}_i$$

which can be expressed as:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$$

where $\mathbf{X}_i = \mathbf{Z}_i\mathbf{W}_i$. By concatenating observations from multiple individuals, we have the following mixed model:

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

22

Mixed Effects Models

Frequently, linear models contain factors whose levels represent a random sample of a population of all possible factor levels

Models containing both fixed and random effects are called mixed effects models

Linear mixed effects models have been widely used in analysis of data where responses are clustered around some random effects, such that there is a natural dependence between observations in the same cluster

For example, consider repeated measurements taken on each subject in longitudinal data, or observations taken on members of the same family in a genetic study

Linear Mixed Effects Model

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

where:

- **y**: response vector; observations
- **β**: vector of fixed effects
- **u**: vector of random effects; **u** ~ N(**0**, ***G***)
- **X** and **Z**: (known) incidence matrices
- **e**: residual vector; **e** ~ N(**0**, **Σ**)

Linear Mixed Effects Model

Generally, it is assumed that **u** and **e** are independent from each other, such that:

$$\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}\end{bmatrix}\right)$$

Inferences regarding mixed effects models refer to the estimation of fixed effects, the prediction of random effects, and the estimation of variance and covariance components, which are briefly discussed next

Estimation of Fixed Effects

Let $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} = \mathbf{Z}\mathbf{u} + \mathbf{e}$

$$\begin{cases} \text{E}[\boldsymbol{\varepsilon}] = \text{E}[\mathbf{Z}\mathbf{u} + \mathbf{e}] = \mathbf{Z}\text{E}[\mathbf{u}] + \text{E}[\mathbf{e}] = \mathbf{0} \\ \text{Var}[\boldsymbol{\varepsilon}] = \text{Var}[\mathbf{Z}\mathbf{u} + \mathbf{e}] = \mathbf{Z}\text{Var}[\mathbf{u}]\mathbf{Z}^{\text{T}} + \text{Var}[\mathbf{e}] = \mathbf{Z}\mathbf{G}\mathbf{Z}^{\text{T}} + \boldsymbol{\Sigma} \end{cases}$$

such that $\mathbf{y} \sim \text{MVN}(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, where $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^{\text{T}} + \boldsymbol{\Sigma}$

Under these circumstances, the MLE for β is:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\text{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\text{T}}\mathbf{V}^{-1}\mathbf{y} \sim \text{MVN}(\boldsymbol{\beta}, (\mathbf{X}^{\text{T}}\mathbf{V}^{-1}\mathbf{X})^{-1})$$

Estimation of Fixed Effects

As **G** and **Σ** are generally unknown, an estimate of **V** is used instead such that the estimator becomes:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}^{-1}\mathbf{y}$$

The variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is now approximated by $(\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}$

Note: $(\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}$ is biased downwards as a consequence of ignoring the variability introduced by working with estimates of (co)variance components instead of their true (unknown) parameter values

Estimation of Fixed Effects

Approximated confidence regions and test statistics for estimable functions of the type $\mathbf{K}^{\mathrm{T}}\boldsymbol{\beta}$ can be obtained by using the result:

$$\frac{(\mathbf{K}^{\mathrm{T}}\boldsymbol{\beta}^{0})^{\mathrm{T}}(\mathbf{K}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{K})^{-1}(\mathbf{K}^{\mathrm{T}}\boldsymbol{\beta}^{0})}{\mathrm{rank}(\mathbf{K})} \approx \mathrm{F}_{[\varphi_{\mathrm{N}},\varphi_{\mathrm{D}}]}$$

where $\mathrm{F}_{[\varphi_{\mathrm{N}},\varphi_{\mathrm{D}}]}$ refers to an F-distribution with $\varphi_{\mathrm{N}} = \mathrm{rank}(\mathbf{K})$ degrees of freedom for the numerator, and φ_{D} degrees of freedom for the denominator, which is generally calculated from the data using, for example, the Satterthwaite's approach

Estimation (Prediction) of Random Effects

In addition to the estimation of fixed effects, very often in genetics interest is also on prediction of random effects.

In linear (Gaussian) models such predictions are given by the conditional expectation of **u** given the data, i.e. $\mathrm{E}[\mathbf{u} \mid \mathbf{y}]$

Given the model specifications, the joint distribution of **y** and **u** is:

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{u}\end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix}\mathbf{X\beta}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{V} & \mathbf{ZG}\\ \mathbf{GZ}^{\mathrm{T}} & \mathbf{G}\end{bmatrix}\right)$$

Estimation (Prediction) of Random Effects

From the properties of multivariate normal distribution, we have that:

$$\begin{aligned}\mathrm{E}[\mathbf{u} \mid \mathbf{y}] &= \mathrm{E}[\mathbf{u}] + \mathrm{Cov}[\mathbf{u}, \mathbf{y}^{\mathrm{T}}]\mathrm{Var}^{-1}[\mathbf{y}](\mathbf{y} - \mathrm{E}[\mathbf{y}])\\ &= \mathbf{GZ}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X\beta}) = \mathbf{GZ}^{\mathrm{T}}(\mathbf{ZGZ}^{\mathrm{T}} + \mathbf{\Sigma})^{-1}(\mathbf{y} - \mathbf{X\beta})\end{aligned}$$

The fixed effects **β** are typically replaced by their estimates, so that predictions are made based on the following expression:

$$\hat{\mathbf{u}} = \mathbf{GZ}^{\mathrm{T}}(\mathbf{ZGZ}^{\mathrm{T}} + \mathbf{\Sigma})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

Mixed Model Equations

The solutions $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ discussed before require $\mathbf{V}^{-1}$

As **V** can be of huge dimensions, especially in animal breeding applications, its inverse is generally computationally demanding if not unfeasible.

However, Henderson (1950) presented the mixed model equations (MME) to estimate **β** and **u** simultaneously, without the need for computing $\mathbf{V}^{-1}$

The MME were derived by maximizing (for **β** and **u**) the joint density of **y** and **u**, expressed as:

$$p(\mathbf{y},\mathbf{u}\,|\,\boldsymbol{\beta},\mathbf{G},\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2}|\mathbf{G}|^{-1/2} \times \exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})-\frac{1}{2}\mathbf{u}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{u}\right\}$$

Mixed Model Equations

The logarithm of this function is:

$$\begin{aligned}\ell = \log[p(\mathbf{y},\mathbf{u}\,|\,\boldsymbol{\beta},\mathbf{G},\boldsymbol{\Sigma})] \propto |\boldsymbol{\Sigma}| + |\mathbf{G}| + (\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u}) + \mathbf{u}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{u} \\ = |\boldsymbol{\Sigma}| + |\mathbf{G}| + \mathbf{y}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y} - 2\mathbf{y}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} - 2\mathbf{y}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z}\mathbf{u} \\ + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} + 2\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z}\mathbf{u} + \mathbf{u}^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z}\mathbf{u} + \mathbf{u}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{u}\end{aligned}$$

The derivatives of ℓ regarding **β** and **u** are:

$$\begin{bmatrix}\frac{\partial \ell}{\partial \boldsymbol{\beta}} \\ \frac{\partial \ell}{\partial \mathbf{u}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y} - \mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z}\hat{\mathbf{u}} \\ \mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y} - \mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z}\hat{\mathbf{u}} - \mathbf{G}^{-1}\hat{\mathbf{u}}\end{bmatrix}$$

Mixed Model Equations

Equating them to zero gives the following system:

$$\begin{bmatrix} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}\hat{\mathbf{u}} \\ \mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}\hat{\mathbf{u}} + \mathbf{G}^{-1}\hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} \\ \mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{y} \end{bmatrix}$$

which can be expressed as:

$$\begin{bmatrix} \mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X} & \mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z} \\ \mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X} & \mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y} \\ \mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y} \end{bmatrix}$$

known as the mixed model equations (MME)

BLUE and BLUP

Using the second part of the MME, we have that:

$$\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + (\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\hat{\mathbf{u}} = \mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

so that:

$$\hat{\mathbf{u}} = (\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

It can be shown that this expression is equivalent to:

$$\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}^{\mathrm{T}}(\mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}} + \boldsymbol{\Sigma})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

and, more importantly, that $\hat{\mathbf{u}}$ is the best linear unbiased predictor (BLUP) of $\mathbf{u}$

BLUE and BLUP

Using this result into the first part of the MME, we have that:

$$\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z}\hat{\mathbf{u}} = \mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

$$\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

$$\hat{\boldsymbol{\beta}} = \{\mathbf{X}^{\mathrm{T}}[\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}]\mathbf{X}\}^{-1}\mathbf{X}^{\mathrm{T}}[\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}]\mathbf{y}$$

Similarly, it is shown that this expression is equivalent to $\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{y}$, which is the best linear unbiased estimator (BLUE) of β.

BLUE and BLUP

It is important to note that $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ require knowledge of ***G*** and **Σ**. These matrices, however, are rarely known. This is a problem without an exact solution using classical methods.

The practical approach is to replace ***G*** and **Σ** by their estimates ($\hat{\mathbf{G}}$ and $\hat{\boldsymbol{\Sigma}}$) into the MME:

$$\begin{bmatrix} \mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X} & \mathbf{Z}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y} \\ \mathbf{Z}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y} \end{bmatrix}$$

Estimation of Variance Components

BLUE and BLUP require knowledge of $\mathbf{G}$ and $\mathbf{\Sigma}$

These matrices, however, are rarely known and must be estimated

Variance and covariance components estimation:

- Analysis of Variance (ANOVA)
- Maximum likelihood
- Restricted maximum likelihood (REML)
- Bayesian approach (to be discussed later)

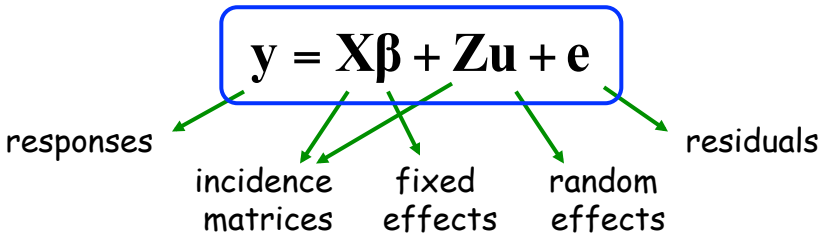
Lecture 4
BLUP Breeding Values

Guilherme J. M. Rosa
University of Wisconsin-Madison

Introduction to Quantitative Genetics
SISG, Seattle
19 – 21 July 2017

Linear Mixed Effects Model

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



$$\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}\end{bmatrix}\right)$$

Estimation of Fixed Effects

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

with $\boldsymbol{\varepsilon} = \mathbf{Z}\mathbf{u} + \mathbf{e}$, such that $\mathrm{Var}[\boldsymbol{\varepsilon}] = \mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}} + \boldsymbol{\Sigma}$

➔ MLE for β :

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{y} \sim \mathrm{MVN}(\boldsymbol{\beta}, (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1})$$

where $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}} + \boldsymbol{\Sigma}$

Prediction of Random Effects

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{u}\end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix}\mathbf{X}\boldsymbol{\beta}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{V} & \mathbf{Z}\mathbf{G}\\ \mathbf{G}\mathbf{Z}^{\mathrm{T}} & \mathbf{G}\end{bmatrix}\right)$$

$$\begin{aligned}\mathrm{E}[\mathbf{u}\,|\,\mathbf{y}] &= \mathrm{E}[\mathbf{u}] + \mathrm{Cov}[\mathbf{u}, \mathbf{y}^{\mathrm{T}}]\mathrm{Var}^{-1}[\mathbf{y}](\mathbf{y} - \mathrm{E}[\mathbf{y}])\\ &= \mathbf{G}\mathbf{Z}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{G}\mathbf{Z}^{\mathrm{T}}(\mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}} + \boldsymbol{\Sigma})^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\end{aligned}$$

Replacing β by its estimate:

$$\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}^{\mathrm{T}}(\mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}} + \boldsymbol{\Sigma})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

Mixed Model Equations

$$\begin{bmatrix} \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X} & \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{X} & \mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z}+\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} \\ \mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} \end{bmatrix}$$

BLUP and BLUE:

$$\begin{cases} \hat{\mathbf{u}} = (\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z}+\mathbf{G}^{-1})^{-1}\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) \\ \hat{\boldsymbol{\beta}} = \{\mathbf{X}^T[\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z}+\mathbf{G}^{-1})^{-1}\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}]\mathbf{X}\}^{-1} \\ \qquad \times \mathbf{X}^T[\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z}+\mathbf{G}^{-1})^{-1}\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}]\mathbf{y} \end{cases}$$

Mixed Models in Animal and Plant Breeding

Animal/plant breeding programs are based on the principle that phenotypic observations on related individuals can provide information about their underlying genotypic values

The additive component of genetic variation is the primary determinant of the degree to which offspring resemble their parents, and therefore this is usually the component of interest in artificial selection programs

Mixed Models in Animal and Plant Breeding

Many statistical methods for analysis of genetic data are specific (or more appropriate) for phenotypic measurements obtained from planned experimental designs and with balanced data sets

While such situations may be possible within laboratory or greenhouse experimental settings, data from natural populations and agricultural species are generally highly unbalanced and fragmented by numerous kinds of relationships

Animal Model

Culling of data to accommodate conventional statistical techniques (e.g. ANOVA) may introduce bias and/or lead to a substantial loss of information

The mixed model methodology allows efficient estimation of genetic parameters (such as variance components and heritability) and breeding values while accommodating extended pedigrees, unequal family sizes, overlapping generations, sex-limited traits, assortative mating, and natural or artificial selection

To illustrate such application of mixed models in breeding programs, we consider here the so-called Animal Model in situations with a single trait and a single observation (including missing values) per individual

Animal Model

The animal model can be described as:

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

- $\mathbf{y}$ is an ($n \times 1$) vector of observations (phenotypic scores)
- β is a ($p \times 1$) vector of fixed effects (e.g. herd-year-season effects)
- $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$ is a ($q \times 1$) vector of breeding values (relative to all individuals with record or in the pedigree file, such that q is in general bigger than n)
- $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}_n\sigma_e^2)$ represents residual effects, where σ_e^2 is the residual variance

The Matrix *A*

The matrix ***G*** describing the covariances among the random effects (here the breeding values) follows from standard results for the covariances between relatives

It is seen that the additive genetic covariance between two relatives i and i' is given by $2\theta_{ii'}\sigma_a^2$, where $\theta_{ii'}$ is the coefficient of coancestry between individuals i and i', and σ_a^2 is the additive genetic variance in the base population

Hence, under the animal model, $\mathbf{G} = \mathbf{A}\sigma_a^2$, where ***A*** is the additive genetic (or numerator) relationship matrix, having elements given by $a_{ii'} = 2\theta_{ii'}$

The Matrix *A*

For each animal i in the pedigree (i = 1, 2,…,n), going from older to younger animals, compute a_{ii} and a_{ij} (j = 1, 2,…,i-1) as follows:

If both parents (s and d) of animal i are known:

$a_{ij} = a_{ji} = (a_{js} + a_{jd})/2$ and $a_{ii} = 1 + a_{sd}/2$

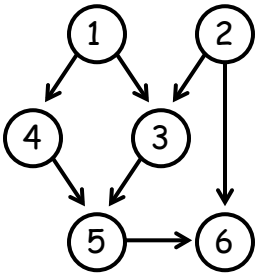
If only one parent (e.g. d) of animal i is known:

$a_{ij} = a_{ji} = a_{jd}/2$ and $a_{ii} = 1$

If parents unknown:

$a_{ij} = a_{ji} = 0$ and $a_{ii} = 1$

Example



Animal	Sire	Dam
1	-	-
2	-	-
3	1	2
4	1	-
5	4	3
6	5	2

$$A = \begin{bmatrix} 1 & 0 & .5 & .5 & .5 & .25 \\ 0 & 1 & .5 & 0 & .25 & .625 \\ .5 & .5 & 1 & .25 & .625 & .563 \\ .5 & 0 & .25 & 1 & .625 & .313 \\ .5 & .25 & .625 & .625 & 1.125 & .688 \\ .25 & .625 & .563 & .313 & .688 & 1.125 \end{bmatrix}$$

pedigree matrix A

Animal Model

In general, in animal/plant breeding interest is on prediction of breeding values (for selection of superior individuals), and on estimation of variance components and functions thereof, such as heritability

The fixed effects are, in some sense, nuisance factors with no central interest in terms of inferences, but which need to be taken into account (i.e., they need to be corrected for when inferring breeding values)

Animal Model

Since under the animal model $\mathbf{G}^{-1} = \mathbf{A}^{-1}\sigma_a^{-2}$ and $\mathbf{R}^{-1} = \mathbf{I}_n\sigma_e^{-2}$, the mixed model equations can be expressed as:

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z}+\lambda\mathbf{A}^{-1} \end{bmatrix}\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{bmatrix}$$

where $\lambda = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-h^2}{h^2}$, such that:

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z}+\lambda\mathbf{A}^{-1} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{bmatrix}$$

Conditional on the variance components ratio λ, the BLUP of the breeding values are given then by:

$$\hat{\mathbf{u}} = (\mathbf{Z}^T\mathbf{Z} + \lambda\mathbf{A}^{-1})^{-1}\mathbf{Z}^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

These are generally referred to as Estimated Breeding Values (EBV)

Alternatively, some breeders associations express their results as Predicted Transmitting Abilities (PTA) (or Estimated Transmitting Abilities (ETA) or Expected Progeny Difference (EPD)), which are equal to half the EBV, representing the portion of an animal's breeding values that is passed to its offspring

The amount of information contained in an animal's genetic evaluation depends on the availability of its own record, as well as how many (and how close) relatives it has with phenotypic information

As a measure of amount of information in livestock genetic evaluations, EBVs are typically reported with its associated accuracies

Accuracy of predictions is defined as the correlation between true and estimated breeding values, i.e., $r_i = \rho(\hat{u}_i, u_i)$

Instead of accuracy, some livestock species genetic evaluations use reliability, which is the squared correlation of accuracy (r_i^2)

Prediction Accuracy

The calculation of $\rho(\hat{u}_i, u_i)$ requires the diagonal elements of the inverse of the MME coefficient matrix, represented as:

$$\mathbf{C} = \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \lambda\mathbf{A}^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C}^{\beta\beta} & \mathbf{C}^{\beta u} \\ \mathbf{C}^{u\beta} & \mathbf{C}^{uu} \end{bmatrix}$$

It is shown that the prediction error variance of EBV $\hat{u}_i$ is given by:

$$PEV = Var(\hat{u}_i - u_i) = c_i^{uu}\sigma_e^2$$

where c_i^{uu} is the i-th diagonal element of $\mathbf{C}^{uu}$, relative to animal i.

Prediction Accuracy

The PEV can be interpreted as the fraction of additive genetic variance not accounted for by the prediction

Therefore, PEV can be expressed also as:

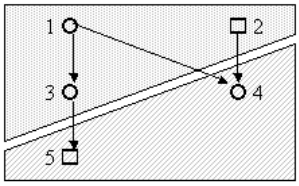
$$PEV = (1 - r_i^2)\sigma_a^2$$

such that $c_i^{uu}\sigma_e^2 = (1 - r_i^2)\sigma_a^2$, from which the reliability is obtained as:

$$r_i^2 = 1 - c_i^{uu}\sigma_e^2 / \sigma_a^2 = 1 - \lambda c_i^{uu}$$

Animal Model

herd 1



herd 2

Animal	Sire	Dam	Herd	Observation
1	–	–	h1	310
2	–	–	h1	–
3	–	1	h1	270
4	2	1	h2	350
5	–	3	h2	–

$$\begin{bmatrix}310\\270\\350\end{bmatrix} = \begin{bmatrix}1 & 0\\1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}h_1\\h_2\end{bmatrix} + \begin{bmatrix}1 & 0 & 0 & 0 & 0\\0 & 0 & 1 & 0 & 0\\0 & 0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}u_1\\u_2\\u_3\\u_4\\u_5\end{bmatrix} + \begin{bmatrix}e_1\\e_3\\e_4\end{bmatrix}$$

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

Animal Model

Breeding values: $\mathbf{u} \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2)$, with

$$\mathbf{A} = \begin{bmatrix}1 & 0 & 0.5 & 0.5 & 0.25\\0 & 1 & 0 & 0.5 & 0\\0.5 & 0 & 1 & 0.25 & 0.5\\0.5 & 0.5 & 0.25 & 1 & 0.125\\0.25 & 0 & 0.5 & 0.125 & 1\end{bmatrix}$$

$$\begin{bmatrix}\hat{\boldsymbol{\beta}}\\\hat{\mathbf{u}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z}\\\mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \lambda\mathbf{A}^{-1}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}^T\mathbf{y}\\\mathbf{Z}^T\mathbf{y}\end{bmatrix}$$

$$\lambda = \frac{\sigma_e^2}{\sigma_u^2} = \frac{1-h^2}{h^2}$$

R Code

animal model
toy example

```
y<-matrix(c(310,270,350),nrow=3)
X<-matrix(c(1,1,0,0,0,1),nrow=3)
Z<-matrix(c(1,0,0,0,0,0,0,1,0,0,0,0,0,1,0),nrow=3, byrow = TRUE)
A<-matrix(c(1,0,0.5,0.5,0.25,
            0,1,0,0.5,0,
            0.5,0,1,0.25,0.5,
            0.5,0.5,0.25,1,0.125,
            0.25,0,0.5,0.125,1),nrow=5)

h2<-1/3 # heritability
a=(1-h2)/h2

# crossproducts
XX<-crossprod(X,X)
XZ<-t(X) %*% Z
ZX<-t(Z) %*% X
ZZ<-crossprod(Z,Z)+a*solve(A)

# mixed model equations
# coefficient matrix and right hand side
C<-rbind(cbind(XX,XZ),cbind(ZX,ZZ))
rhs<-rbind(t(X) %*% y,t(Z) %*% y)

#solution
theta.hat <- solve(C) %*% rhs
```

$$h^2 = \frac{1}{3} \rightarrow \alpha = 2 \Rightarrow \begin{cases} \hat{h}_1 = 290 \\ \hat{h}_2 = 348 \\ \hat{u}_1 = \ 4.0 \\ \hat{u}_2 = \ 0.0 \\ \hat{u}_3 = -4.0 \\ \hat{u}_4 = \ 2.0 \\ \hat{u}_5 = -2.0 \end{cases}$$

Animal Model

The animal model can be extended to model multiple (correlated) traits, multiple random effects (such as maternal effects and common environmental effects), repeated records (e.g. test day models), and so on

Example (Mrode 1996, pp74-76): Weaning weight (kg) of piglets, progeny of three sows mated to two boars:

Piglet	Sire	Dam	Sex	Weight
6	1	2	1	90
7	1	2	2	70
8	1	2	2	65
9	3	4	2	98
10	3	4	1	106
11	3	4	2	60
12	3	4	2	80
13	1	5	1	100
14	1	5	2	85
15	1	5	1	68

A linear model with the (fixed) effect of sex, and the (random) effects of common environment (related to each litter) and breeding values can be expressed as **X**:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{Wc} + \mathbf{e}$$

Weight — Sex — Breeding values — Common environment — Residual

Assuming that $\sigma^2_u = 20$, $\sigma^2_c = 15$ and $\sigma^2_e = 65$, the MME are as follows:

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} & \mathbf{X}^T\mathbf{W} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \mathbf{A}^{-1}\lambda_1 & \mathbf{Z}^T\mathbf{W} \\ \mathbf{W}^T\mathbf{X} & \mathbf{W}^T\mathbf{Z} & \mathbf{W}^T\mathbf{W} + \mathbf{I}\lambda_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{c}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \\ \mathbf{W}^T\mathbf{y} \end{bmatrix}$$

where $\lambda_1 = \frac{\sigma^2_e}{\sigma^2_u} = 3.25$ and $\lambda_2 = \frac{\sigma^2_e}{\sigma^2_c} = 4.\dot{3}$

The BLUEs and BLUPs (inverting the numerator relationship matrix) are:

Effects	Solutions
Sex	
1	91.493
2	75.764
Animals	
1	-1.441
2	-1.175
3	1.441
4	1.441
5	-0.266
6	-1.098
7	-1.667
8	-2.334
9	3.925
10	2.895
11	-1.141
12	1.525
13	0.448
14	0.545
15	-3.819
Environ.	
2	-1.762
4	2.161
5	-0.399

Mrode example

Lecture 5
Genomic Selection

Guilherme J. M. Rosa
University of Wisconsin-Madison

Introduction to Quantitative Genetics
SISG, Seattle
19 – 21 July 2017

OUTLINE

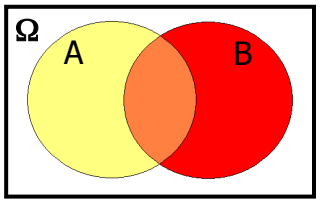
- Bayesian Analysis and MCMC
- Marker Assisted Selection
- Genomic Selection
- Models & Techniques

Bayesian Data Analysis

Inferences using probability models for quantities we observe and for quantities about which we wish to learn

Explicit use of probability for quantifying uncertainty in inferences based on statistical data analysis

Conditional Probability (Bayes' Rule)



$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B \mid A)}{P(B)}$$

Bayesian Inference

- y: observed data; $y \sim p(y|\theta)$
- θ: parameters (all unobserved quantities)

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

$$p(\theta \mid y) \propto p(\theta)p(y \mid \theta)$$

posterior distribution

prior distribution

sampling distribution

Prior Distributions

Informative and Noninformative

Proper and Improper

Conjugate and Nonconjugate

Jeffreys' Prior

Maximum Entropy

Reference Prior

Example 1: Binomial Distribution

Data: $y_1, y_2, \ldots, y_n \overset{iid}{\sim} \text{Bin}(n_i, \theta)$, θ = Prob(y = 1)

Sampling model: $$p(\mathbf{y} \mid \theta) = \prod_{i=1}^{n} p(y_i \mid \theta) = \prod_{i=1}^{n} \binom{1}{y_i} \theta^{y_i} (1-\theta)^{1-y_i}$$

$$\propto \theta^{\sum y_i} (1-\theta)^{n-\sum y_i}$$

Prior: $$p(\theta) = \text{Beta}(a, b) \propto \theta^{a-1} (1-\theta)^{b-1}$$

Posterior: $$p(\theta \mid \mathbf{y}) \propto \theta^{a+\sum y_i - 1} (1-\theta)^{n+b-\sum y_i - 1}$$

$$\boxed{\theta \mid \mathbf{y} \sim \text{Beta}\left(a + \sum y_i,\ n + b - \sum y_i\right)}$$

Example 1: Binomial Distribution

$$\boxed{\theta \mid \mathbf{y} \sim \text{Beta}\left(a + \sum y_i,\ n + b - \sum y_i\right)}$$

Features of the posterior distribution:

Posterior mean: $$E[\theta \mid \mathbf{y}] = \frac{a + \sum y_i}{n + a + b}$$

Posterior mode: $$\text{Mode}[\theta \mid \mathbf{y}] = \frac{a + \sum y_i - 1}{n + a + b - 2}$$

Posterior variance: $$\text{Var}[\theta \mid \mathbf{y}] = \frac{\left(a + \sum y_i\right)\left(n + b - \sum y_i\right)}{(n + a + b)^2 (n + a + b + 1)}$$

percentis, HPD, etc.

Example 1: Binomial Distribution

Setting, for example a = 1 and b = 1:

Prior: $p(\theta) = \text{Uniform}(0,1)$

Posterior: $p(\theta \mid \mathbf{y}) \propto \theta^{\sum y_i - 1}(1-\theta)^{n-\sum y_i}$

$$\boxed{\theta \mid \mathbf{y} \sim \text{Beta}\left(1+\sum y_i,\ n+1-\sum y_i\right)}$$

Note that in this case the posterior mode coincides with the maximum likelihood estimate of θ:

$$\text{Mode}[\theta \mid \mathbf{y}] = \frac{1}{n}\sum y_i$$

Example 2: Normal Distribution

Data: $y_1, y_2, \ldots, y_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, with known σ^2

Sampling model: $p(y_i \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i-\mu)^2\right\}$

$$\begin{aligned} p(\mathbf{y} \mid \mu, \sigma^2) &= \prod_{i=1}^{n} p(y_i \mid \mu, \sigma^2) \\ &\propto \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right\} \\ &\propto \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}n(\bar{y}-\mu)^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\bar{y})^2\right\} \end{aligned}$$

Example 2: Normal Distribution

Prior (Conjugate): $\mu \sim N(\phi, \tau^2)$

$$p(\mu) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(\mu-\phi)^2\right\}$$

Joint posterior:

$$p(\mu \mid y) \propto p(y \mid \mu, \sigma^2) \times p(\mu)$$

$$\propto \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} n(\bar{y}-\mu)^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\bar{y})^2\right\}$$

$$\times \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(\mu-\phi)^2\right\}$$

Joint posterior (cont'ed):

$$p(\mu \mid y) \propto \exp\left\{-\frac{1}{2\sigma^2} n(\bar{y}-\mu)^2\right\} \exp\left\{-\frac{1}{2\tau^2}(\mu-\phi)^2\right\}$$

$$\propto \exp\left\{-\frac{n(\bar{y}-\mu)^2}{2\sigma^2} - \frac{(\mu-\phi)^2}{2\tau^2}\right\}$$

$$\propto \exp\left\{-\frac{\mu^2}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right) + \mu\left(\frac{n\bar{y}}{\sigma^2}+\frac{\phi}{\tau^2}\right) - \frac{1}{2\tau}\left(\frac{n\bar{y}^2}{\sigma^2}+\frac{\phi^2}{\tau^2}\right)\right\}$$

$$= \exp\left\{-\frac{1}{2\sigma_n^2}(\mu+\mu_n)^2\right\}$$

where $\frac{1}{\sigma_n^2} = \left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)$ and $\frac{\mu_n}{\sigma_n^2} = \left(\frac{n\bar{y}}{\sigma^2}+\frac{\phi}{\tau^2}\right)$

Hence: $\mu \mid y \sim N\left(\frac{n\tau^2}{n\tau^2+\sigma^2}\bar{y} + \frac{\sigma^2}{n\tau^2+\sigma^2}\phi,\ \left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}\right)$

Multi Parameter Models

$$y \sim p(y \mid \theta_1, \theta_2, \ldots, \theta_p)$$

$$p(\theta_1, \theta_2, \ldots, \theta_p \mid y) \sim p(\theta_1, \theta_2, \ldots, \theta_p) p(y \mid \theta_1, \theta_2, \ldots, \theta_p)$$

Marginal Posterior Distributions

$$p(\theta_1 \mid y) \propto \int_{\theta \neq \theta_1} p(\theta_1, \theta_2, \ldots, \theta_p \mid y) d\theta_{\theta \neq \theta_1}$$

Marginal Posterior Distributions

Marginalization (i.e. integrals) in multi-dimensional models can be cumbersome and some times do not have analytical form

An alternative in this regard: Monte Carlo methods

Monte Carlo integration consists of sampling from the posterior distribution, and then using such sampled values to calculate features of interest on the (joint or marginal) posterior distribution

There are many algorithms that can be used to sample from a distribution; some are based on Markov chains, among which the Gibbs sampling is probably the most popular

Gibbs Sampling

$$\theta = (\theta_1, \theta_2, \ldots, \theta_r) \rightarrow p(\theta_i \mid \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_r)$$

$$\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_r^{(0)})$$

$$\theta_1^{(1)} \mid \theta_2^{(0)}, \theta_3^{(0)}, \ldots, \theta_r^{(0)}$$

$$\theta_2^{(1)} \mid \theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_r^{(0)}$$

$$\vdots$$

$$\theta_r^{(1)} \mid \theta_2^{(1)}, \theta_3^{(1)}, \ldots, \theta_{r-1}^{(1)}$$

Burn-in & Convergence

Tinning interval & Lag correlations

Sample size & Monte Carlo error

Monte Carlo Approximations

After convergence, each sampled vector is a sample from the joint posterior distribution, and so each sampled element (scalar) is a sample from the respective marginal posterior distribution

For each parameter (e.g., θ_1) we'll have then a series of values:

$$\theta_1^{(1)}, \theta_1^{(2)}, \theta_1^{(3)}, \ldots, \theta_1^{(N)}$$

from which features of its distribution (e.g., posterior mean) can be approximated, for example:

$$E[\theta_1 \mid y] \cong \frac{1}{N} \sum_{j=1}^{N} \theta_1^{(j)}$$

Monte Carlo Approximations

Other often interesting features used to represent a marginal posterior distribution are: posterior variance (or standard deviation), posterior mode or median, percentiles, highest posterior density (HPD), etc.

Very useful property: If one is interested on the distribution of a function of the model parameters, samples from such a distribution can be obtained simply by applying that specific function to the sampled values of those parameters

For example, the posterior mean of the heritability can be obtained as:

$$E[h^2 \mid \mathbf{y}] \cong \frac{1}{N}\sum_{j=1}^{N} \frac{\sigma_u^{(j)}}{\sigma_u^{(j)} + \sigma_\varepsilon^{(j)}}$$

Marker Assisted Selection

MAS: Use of genetic markers to improve the efficiency of genetic selection

Basic idea behind of MAS:

- Most traits of economic importance are controlled by a fairly large number of genes
- Some of these genes, however, with larger effect
- Following the pattern of inheritance of such genes might assist in selection

MAS Could Help Improve

Low heritability traits

Phenotypes that can be measured on one sex only

Characteristics that are not measurable before sexual maturity

Traits that are difficult to measured or require sacrifice

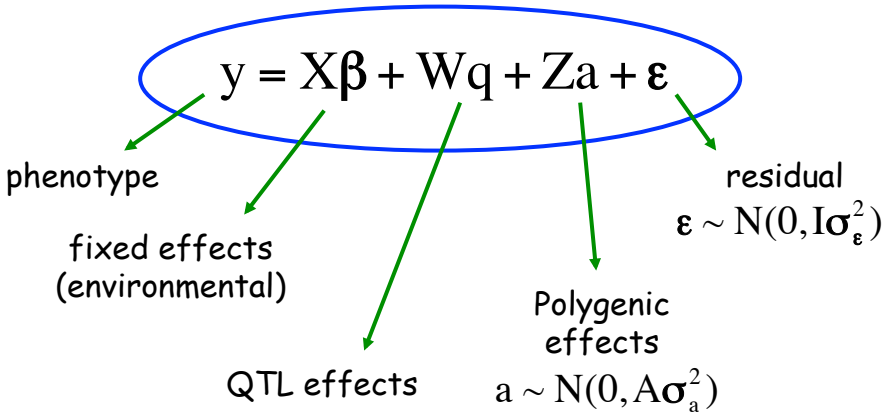
Efficiency of MAS

Size (effect) of QTL

Frequency of favorable allele

Recombination rate between marker(s) and QTL

Modeling Effects at The QTL Genotype



Modeling Effects at the QTL Genotype

QTL-genotype as a fixed effect: Regression of phenotypes using QTL genotype probabilities from segregation analysis (Kinghorn et al. 1993, Meuwissen and Goddard 1997)

QTL-genotype as a random effect: QTL effect is modeled as the sum of the two gametic effects (Fernando and Grossman 1989)

$$y = X\beta + Wv + Za + \varepsilon \;, \quad Var\begin{pmatrix} v \\ a \\ \varepsilon \end{pmatrix} = \begin{pmatrix} G_v\sigma_v^2 & 0 & 0 \\ 0 & A\sigma_a^2 & 0 \\ 0 & 0 & I\sigma_\varepsilon^2 \end{pmatrix}$$

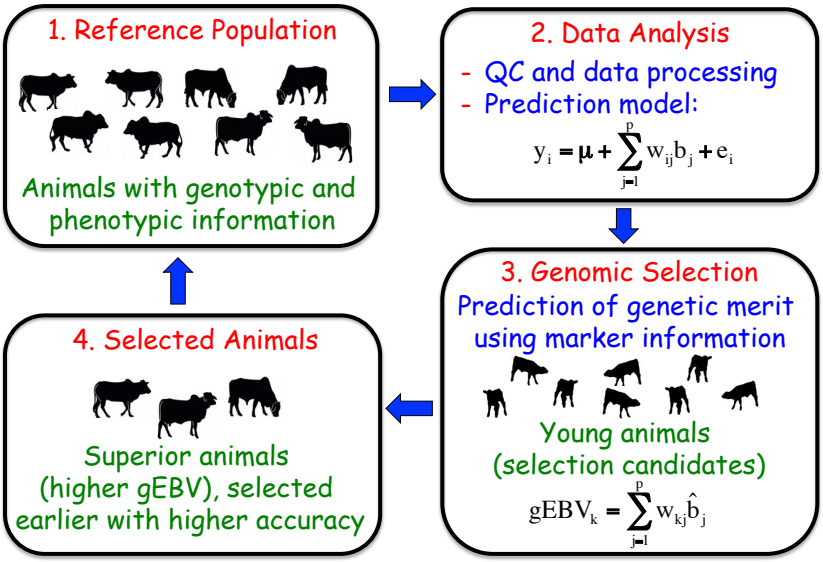
Gametic relationship matrix

Genomic Selection
(Genome-wide Marker Assisted Selection)

As most quantitative traits are influenced by many genes, tracking a small number of them using molecular markers will explain only a small fraction of the total genetic variance

GWMAS, on the other hand, makes use of a very dense set of markers covering the entire genome, which potentially explain all genetic variance

Genomic Selection



Genomic Selection

(Meuwissen et al., 2001)

$$y_i = \mu + x_{i1}g_1 + x_{i2}g_2 + \ldots + x_{ip}g_p + e_i$$

Marker genotypes — Genetic effects

Genomic EBV: $GEBV = x_{i1}\hat{g}_1 + x_{i2}\hat{g}_2 + \ldots + x_{ip}\hat{g}_p = \sum_{j=1}^{p} x_{ij}\hat{g}_j$

- ‘big p small n paradigm’
- Dimension reduction techniques (e.g. SVD and PLS), and stepwise strategies
- Alternatively, ridge regression, random effects models, and hierarchical modeling

Least Squares

Two-step Procedure:

- Test each marker (chromosome segment) for presence of QTL and select those with significant effects
- Fit selected markers simultaneously using multiple regression
- Predict breeding values using fitted regression (similar to LD- MAS approach with multiple markers)

Problems:

- Over estimation of markers effects due to first-step (selection)
- Do not capture all QTL

BLUP

$$\mathbf{y} = \mathbf{1}\mu + \sum_{j=1}^{p} \mathbf{X}_j g_j + \mathbf{e}$$

$$g_j \sim N(0, \sigma_0^2)$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{X} \\ \mathbf{X}'\mathbf{1} & \mathbf{X}'\mathbf{X} + \mathbf{I}\gamma \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}'\mathbf{y} \\ \mathbf{X}'\mathbf{y} \end{bmatrix}$$

$$\gamma = \sigma_e^2 / \sigma_0^2$$

How to choose σ_0^2 ?

- Arbitrary; but σ_0^2 controls amount of shrinkage
- Alternative: set $\sigma_0^2 = \sigma_u^2 / p$, where σ_u^2 is an estimate (prior) of total additive genetic variance

Bayes A

$$\mathbf{y} = \mathbf{1}\mu + \sum_{j=1}^{p} \mathbf{X}_j \mathbf{g}_j + \mathbf{e} \quad \longrightarrow \quad \mathbf{y} \mid \mu, \mathbf{g}_j, \sigma_e^2 \sim N(\mathbf{1}\mu + \sum_{j=1}^{p} \mathbf{X}_j \mathbf{g}_j, \mathbf{I}\sigma_e^2)$$

Prior distributions:

$$\begin{cases} g_j \mid \sigma_j^2 \sim N(0, \sigma_j^2) \\ \sigma_j^2 \sim \chi^{-2}(\nu, S) \\ \quad \text{(scaled inverted chi-square distribution with scale parameter S and } \nu \text{ degrees of freedom)} \\ \sigma_e^2 \sim \chi^{-2}(-2, 0) \end{cases}$$

Bayes B

$$\mathbf{y} = \mathbf{1}\mu + \sum_{j=1}^{p} \mathbf{X}_j \mathbf{g}_j + \mathbf{e} \quad \longrightarrow \quad \mathbf{y} \mid \mu, \mathbf{g}_j, \sigma_e^2 \sim N(\mathbf{1}\mu + \sum_{j=1}^{p} \mathbf{X}_j \mathbf{g}_j, \mathbf{I}\sigma_e^2)$$

Prior distributions:

$$\begin{cases} \begin{cases} g_j = 0 & \text{with probability } \pi \\ g_j \mid \sigma_j^2 \sim N(0, \sigma_j^2) & \text{with probability } (1 - \pi) \end{cases} \\ \sigma_j^2 \sim \chi^{-2}(\nu, S) \\ \sigma_e^2 \sim \chi^{-2}(-2, 0) \end{cases}$$

SIMULATION STUDY

Genome: 1000 cM with markers every 1 cM

Markers surrounding each 1 cM region combined into haplotypes

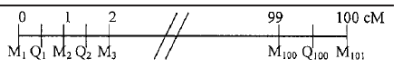
LD between marker and QTLs due to finite population size (N_e = 100)

Training sample: single generation with 2,000 animals

Test sample: prediction of breeding values of their progeny based on marker genotypes

SIMULATION STUDY

The parameters of the simulated genetic model

Map per chromosome[a]	
Number of chromosomes is the total number of morgans	10
Mutation rate of QTL	2.5×10^{-5}
Distribution of additive mutational effects	Gamma(1.66; 0.4)
Dominance of QTL effects	0
Mutation rate of marker loci	2.5×10^{-3}
Population structure	
Generations 1–1000	Ideal[b], $N = 100$
Generation 1001	Ideal[b], $N = 200$
Generation 1002	20 half-sib families, $N = 2000$
Generation 1003 and later	Ideal[b], $N = 2000$
Marker genotyping	Generations 1001 and later
Phenotypic recording	Generations 1001 and 1002

[a] M, marker position; Q, QTL position.

[b] Ideal denotes a population structure where the effective size equals the actual population size. This structure is simulated by giving every male (female) in generation $t - 1$ an equal probability of becoming the sire (dam) of animal i in generation t, which implies no selection and random mating of males and females.

SIMULATION STUDY

Comparing estimated *vs.* true breeding values in generation 1003

	$r_{TBV,EBV}$ + SE	$b_{TBV,EBV}$ + SE
LS	0.318 ± 0.018	0.285 ± 0.024
BLUP	0.732 ± 0.030	0.896 ± 0.045
BayesA	0.798	0.827
BayesB	0.848 + 0.012	0.946 + 0.018

Mean of five replicated simulations, except for BayesA which is based on one replicate. LS, least squares; BLUP, best linear unbiased prediction; BayesA, Bayesian method with inverse chi-square prior distribution; BayesB, Bayesian method where the prior density of having zero QTL effects was increased; $r_{TBV,EBV}$, correlation between estimated and true breeding values (equals accuracy of selection); $b_{TBV,EBV}$, regression of true on estimated breeding value.

SIMULATION STUDY

Correlations between true and estimated breeding values when the number of phenotypic records is varied

	No. of phenotypic records		
	500	1000	2200
LS	0.124	0.204	0.318
BLUP	0.579	0.659	0.732
BayesB	0.708	0.787	0.848

Correlations between true and estimated breeding values when the density of the marker map is varied and effective population size is 100

	Marker spacing (cM)		
	1	2	4
LS	0.318	0.354	0.363
BLUP	0.732	0.708	0.668
BayesB	0.848	0.810	0.737

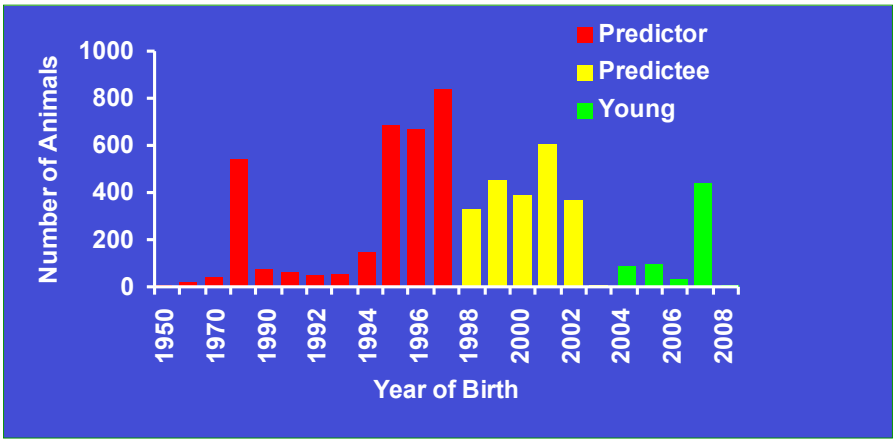
SIMULATION STUDY

The correlation between estimated and true breeding values in generations 1003–1008, where the estimated breeding values are obtained from the BayesB marker estimates in generations 1001 and 1002

Generation	$r_{TBV;EBV}$
1003	0.848
1004	0.804
1005	0.768
1006	0.758
1007	0.734
1008	0.718

The generations 1004–1008 are obtained in the same way as 1003 from their parental generations.

Application with Real Data



(VanRaden et al., 2008)

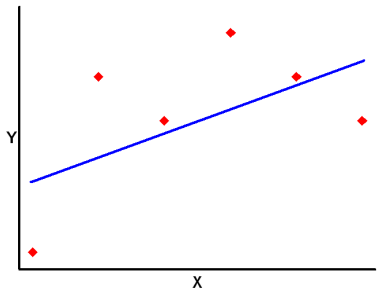
Table 2. Coefficients of determination ($R^2 \times 100$) for 2008 daughter deviations with 2003 predictions

Trait	Traditional parent average	Genomic prediction			Gain from nonlinear genomic prediction compared with parent average
		Linear	Nonlinear	Difference[1]	
Net merit	11	28	28	0	17
Milk yield	28	47	49	2	21
Fat yield	15	42	44	2	29
Protein yield	27	47	47	0	20
Fat percentage	25	55	63	8	38
Protein percentage	28	51	58	7	30
Productive life	17	26	27	1	10
SCS	23	37	38	1	15
Daughter pregnancy rate	20	30	29	−1	9
Sire calving ease	17	21	22	1	5
Daughter calving ease	14	22	22	0	8
Final score	23	35	36	1	13
Stature	27	49	50	1	23
Strength	16	33	34	1	18
Body depth	17	36	37	1	20
Dairy form	9	29	28	−1	19
Foot angle	13	23	21	−2	8
Rear legs (side view)	10	27	27	0	17
Rear legs (rear view)	11	21	19	−2	8
Rump angle	20	44	43	−1	23
Rump width	19	38	36	−2	17
Fore udder	17	39	40	1	23
Rear udder height	20	35	36	1	16
Udder depth	18	47	46	−1	28
Udder cleft	18	30	30	0	12
Front teat placement	22	41	42	1	20
Teat length	12	35	34	−1	22
All	19	36	37	1	18

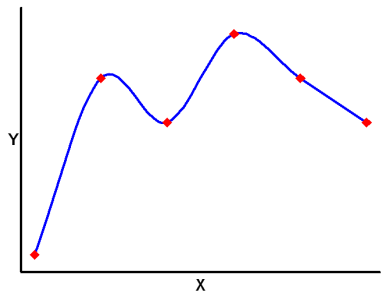
[1]Nonlinear minus linear genomic prediction.

Model Selection

⇨ Goodness-of-fit vs. Model Complexity
(Bias-variance tradeoff)



Over-reduction



Over-fit

Model Selection

⇨ Goodness-of-fit

- likelihood ratio approach (LRT; nested models)

$$\mathrm{LRT} = -2\ln\left(\frac{L_1}{L_2}\right) \sim \chi^2_{(p_1-p_2)}$$

⇨ Model complexity

- number of free parameters, p (effective number)

Linear (regularized) fitting: $\hat{\mathbf{y}} = \mathbf{S}\mathbf{y} \longrightarrow p = \mathrm{trace}(\mathbf{S})$

Model Selection

⇨ Balancing goodness-of-fit and complexity

- Akaike information criterion (AIC):

$$\mathrm{AIC} = 2p - \ln(L)$$

- Bayesian information criterion (BIC):
  (or Schwarz Criterion)

$$\mathrm{BIC} = p\ln(n) - 2\ln(L)$$

☞ If $e_i \overset{\text{iid}}{\sim} N(0, \sigma_e^2)$ then:

$$\mathrm{AIC} = 2p + n\ln\left(\frac{\mathrm{RSS}}{n}\right) \quad \text{and} \quad \mathrm{BIC} = \frac{1}{\sigma_e^2}\mathrm{RSS} + p\ln(L)$$

Ridge Regression

$$\hat{\boldsymbol{\beta}}^{\text{ridge}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}$$

$\lambda \geq 0$ (complexity parameter)

$$\text{or, equivalently: } \hat{\boldsymbol{\beta}}^{\text{ridge}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2,$$

$$\text{subject to: } \sum_{j=1}^{p} \beta_j^2 \leq s$$

Ridge Regression

$$\begin{cases} \hat{\beta}_0 = \bar{y} = \sum y_i / N \\ \text{after centering } y_i \text{ and } x_i\text{'s (i.e., } y_i - \bar{y} \text{ and } x_i - \bar{x}) \end{cases}$$

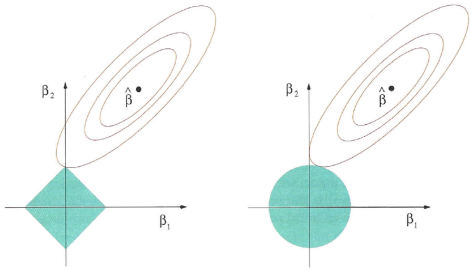
$$\text{RSS}(\lambda) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}'\boldsymbol{\beta}$$

$$\hat{\boldsymbol{\beta}}^{\text{ridge}} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}$$

LASSO

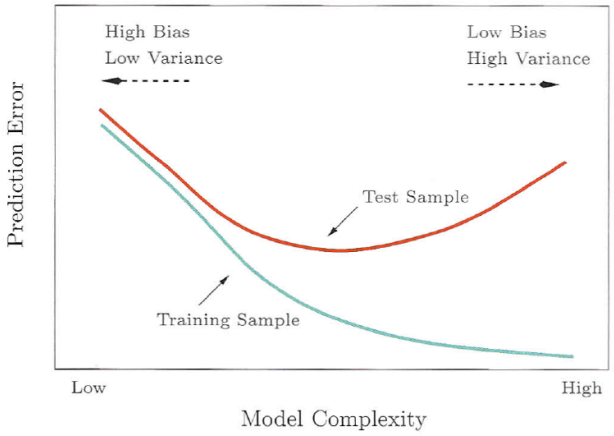
$$\hat{\boldsymbol{\beta}}^{\text{lasso}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2, \text{ subject to: } \sum_{j=1}^{p} |\beta_j| \le t$$

- Estimation picture for the LASSO (left) and Ridge Regression (right)



The solid blue areas are the constraint regions $|\beta_1| + |\beta_2| \le t$ (lasso) and $\beta_1^2 + \beta_2^2 \le t^2$ (ridge regression), while the red ellipses are the contours of the least squares error function.

Predictive Ability

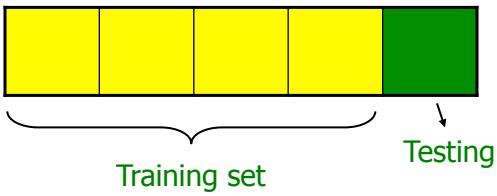


(Hastie et al 2009)

Behavior of test sample and training sample error as the model complexity is varied

Cross-validation

⇨ *K*-FOLD



$$\begin{cases} \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \\ \hat{\boldsymbol{\beta}} : \text{estimate of } \boldsymbol{\beta} \end{cases} \Rightarrow \begin{cases} \text{PMSE} = \dfrac{1}{m}\sum_i (y_i - \hat{y}_i)^2 \\ \hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \end{cases}$$

⇨ LEAVE-ONE-OUT ("*n*-FOLD")

Bayesian Alternative

$$\mathbf{y} = \mathbf{1}\mu + \sum_{j=1}^{p} \mathbf{X}_j \mathbf{g}_j + \mathbf{e} \;\rightarrow\; \mathbf{y} \mid \mu, \mathbf{g}_j, \sigma_e^2 \sim N(\mathbf{1}\mu + \sum_{j=1}^{p} \mathbf{X}_j \mathbf{g}_j, \mathbf{I}\sigma_e^2)$$

BRR: $g_j \mid \sigma_0^2 \sim N(0, \sigma_0^2)$

Bayes A: $g_j \mid \sigma_j^2 \sim N(0, \sigma_j^2), \;\; \sigma_j^2 \sim \chi^{-2}(\nu, S)$

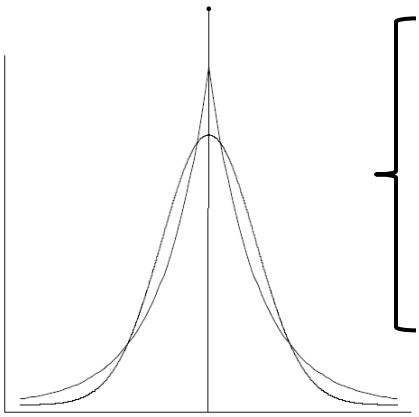
Bayes B,C: $g_j \mid k, \sigma_j^2 \sim \pi \times N(0, k\sigma_j^2) + (1-\pi) \times N(0, \sigma_j^2)$

BLasso: $g_j \mid \sigma_j^2 \sim N(0, \sigma_j^2), \;\; \sigma_j^2 \sim \text{Exponential}(\lambda)$

BX: $g_j \mid \sigma_j^2 \sim N(0, \sigma_j^2), \;\; \sigma_j^2 \sim X$

Normal/Independent Distributions

$$p(g_j) = \int_{\sigma_j^2} p(g_j \mid \sigma_j^2) p(\sigma_j^2) d\sigma_j^2$$



- BRR: Normal
- Bayes A: Student-t
- Bayes B,C: Mixtures
- BLasso: Double exponential

GBLUP

Regression with genetic effects with normal distribution with common variance

$$\mathbf{y} = \mathbf{1}\mu + \sum_{j=1}^{p} \mathbf{X}_j \mathbf{g}_j + \mathbf{e} \quad \text{, with:} \quad g_j \mid \sigma_g^2 \sim N(0, \sigma_g^2)$$

Equivalent Model

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{a} + \mathbf{e} \quad \text{, with:} \quad \mathbf{a} \mid \sigma_a^2 \sim N(\mathbf{0}, \mathbf{G}\sigma_a^2)$$

⇨ ***G*** is the genomic relationship matrix:

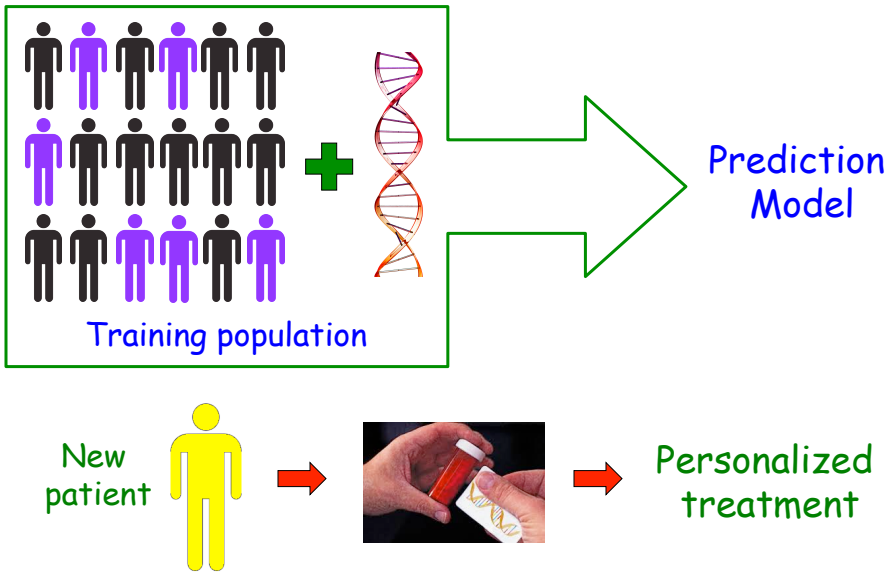
$$\mathbf{G} = \left(2 \sum_{j=1}^{p} p_j (1 - p_j) \right)^{-1} (\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})'$$

ssGBLUP

Single-step GBLUP: Single mixed model with all animals (genotyped and non-genotyped) included, with matrix ***A*** replaced by **H**

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

Preventive and Personalized Medicine



A Comprehensive Genetic Approach for Improving Prediction of Skin Cancer Risk in Humans

Ana I. Vazquez,*,[1] Gustavo de los Campos,* Yann C. Klimentidis,* Guilherme J. M. Rosa,[†] Daniel Gianola,[†] Nengjun Yi,* and David B. Allison*
*Section on Statistical Genetics, Department of Biostatistics, University of Alabama, Birmingham, Alabama 35294, and [†]Department of Animal Sciences, University of Wisconsin, Madison, Wisconsin 53705

Genetics, Vol. 192, 1493–1502 December 2012

- 5,132 subjects from Framingham Heart Study
- Phenotypes measured from 1948 until death
- Genotypes: Affymetrix 500K SNPs

Three generations of participants.

Photo: http://www.framinghamheartstudy.org/

Models

1. No-SNP: standard covariables
2. Covariates + familial relationships
3. Covariates + SNPs (PC or Bayesian LASSO)

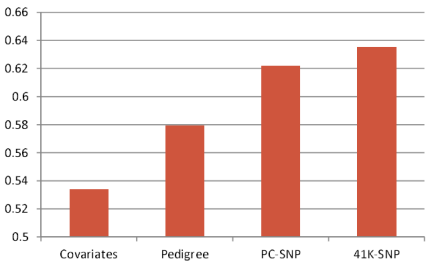
Probit B-LASSO $$p(\mathbf{y}|\boldsymbol{\beta},\mathbf{u}) = \prod_{i=1}^{5132}\left\{[\Phi(\eta_i)]^{y_i}[1-\Phi(\eta_i)]^{1-y_i}\right\}$$

$$\eta_i = \beta_0 + \sum_{j=1}^{p_1} x_{1ij}\beta_{1j} + \sum_{j=1}^{p_2} x_{2ij}\beta_{2j} \quad \text{or} \quad \eta_i = \beta_0 + \sum_{j=1}^{p_1} x_{1ij}\beta_{1j} + u_i$$

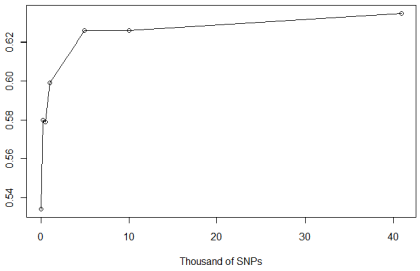
$$\begin{aligned} p(\beta_0,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\mathbf{u},\boldsymbol{\tau}^2,\lambda) \propto & \left[\prod_{j=1}^{p} N\left(\beta_{2j}\middle|0,\tau_j^2\right)\right] \\ & \times \left[\prod_{j=1}^{p} \mathrm{Exp}\left(\tau_j^2\middle|\lambda^2\right)\right] \times G(\lambda^2|\alpha_1,\alpha_2) \\ & \times N(\mathbf{u}|\mathbf{0},\mathbf{A}\sigma_u^2)\times\chi^{-2}(\sigma_u^2|S,df), \end{aligned}$$

Results (ROC, Area Under the Curve)

Comparison of Models



Models with increasing number of SNPs



Lecture 6
QTL and Association Mapping with Mixed Models

Bruce Walsh lecture notes
Introduction to Mixed Models
SISG, Seattle
19 – 21 July 2017

QTL & Association mapping

- We would like to know both the genomic locations (map positions) and effects (either genotypic means or variances) for genes underlying quantitative trait variation
- QTL mapping
  - Using linkage information on a set of known relatives
- Association mapping
  - Using very fine scale LD to map genes in a set of random individuals from a population

Outline

- Basics of QTL mapping
  - Line crosses
    - typically fixed effects models
  - Outbred populations
    - Random effects family models
    - General pedigree methods
- High parameter models
  - Shrinkage approaches for detecting epistasis
- Association mapping

Inbred Line Cross QTL mapping

- Most powerful design
  - Cross two fully inbred lines, look at marker-trait segregation in the F_2 (or other, such as F_n) generations
  - P1: MMQQ, P2:mmqq
  - All F_1 same genotype/phase: MQ/mq
  - Hence, in the F1, all parents have the same genotype
  - At most only two alleles, each with freq 1/2
  - Idea: Does the mean trait value of (say) MM individuals differ from (say) mm
    - Different marker genotypes have different mean trait values

Expected Marker Means

The expected trait mean for marker genotype M_j is just

$$\mu_{M_j} = \sum_{k=1}^{N} \mu_{Q_k} \Pr(Q_k \mid M_j)$$

For example, if QQ = 2a, Qq = a(1+k), qq = 0, then in the F2 of an MMQQ/mmqq cross,

$$(\mu_{MM} - \mu_{mm})/2 = a(1-2c)$$

- If the trait mean is significantly different for the genotypes at a marker locus, it is linked to a QTL
- A small MM-mm difference could be (i) a tightly-linked QTL of small effect or (ii) loose linkage to a large QTL

Linear Models for QTL Detection

The use of differences in the mean trait value for different marker genotypes to detect a QTL and estimate its effects is a use of linear models.

One-way ANOVA.

Value of trait in kth individual of marker genotype type i

$$z_{ik} = \mu + b_i + e_{ik}$$

Effect of marker genotype i on trait value

$$z_{ik} = \mu + b_i + e_{ik}$$

Detection: a QTL is linked to the marker if at least one of the b_i is significantly different from zero

Estimation: (QTL effect and position): This requires relating the b_i to the QTL effects and map position

Detecting epistasis

One major advantage of linear models is their flexibility. To test for epistasis between two QTLs, use ANOVA with an interaction term

$$z = \mu + a_i + b_k + d_{ik} + e$$

Effect from marker genotype at first marker set (can be > 1 loci)

Effect from marker genotype at second marker set

Interaction between marker genotypes i in 1st marker set and k in 2nd marker set

Detecting epistasis

$$z = \mu + a_i + b_k + d_{ik} + e$$

- At least one of the a_i significantly different from 0
  ---- QTL linked to first marker set
- At least one of the b_k significantly different from 0
  ---- QTL linked to second marker set
- At least one of the d_{ik} significantly different from 0
  ---- interactions between QTL in sets 1 and two

Problem: Huge number of potential interaction terms (order m^2, where m = number of markers)

Model selection

- With (say) 300 markers, we have (potentially) 300 single-marker terms and 300*299/2 = 44,850 epistatic terms
  - Hence, a model with up to p= 45,150 possible parameters
  - 2^p possible submodels = $10^{13,600}$ ouch!
- The issue of Model selection becomes very important.
- How do we find the best model?
  - Stepwise regression approaches
    - Forward selection (add terms one at a time)
    - Backwards selection (delete terms one at a time)
  - Try all models, assess best fit
  - Mixed-model approaches (Stochastic Search Variable Selection, or SSVS)

Model Selection

Model Selection: Use some criteria to chose among a number of candidate models. Weight goodness-of-fit (L, value of the likelihood at the MLEs) vs. number of estimated parameters (k)

AIC = Akaike's information criterion
AIC = 2k - 2 Ln(L)

BIC = Bayesian information criterion (Schwarz criterion)
BIC = k*ln(n)/n - 2 Ln(L)/n
BIC penalizes free parameters more strongly than AIC

Other measures. For these (and AIVC, BIC) smaller score indicates better model fit

Model averaging

Model averaging: Generate a composite model by weighting (averaging) the various models, using AIC, BIC, or other

Idea: Perhaps no "best" model, but several models all extremely close. Better to report this "distribution" rather than the best one

One approach is to average the coefficients on the "best-fitting" models using some scheme to return a composite model

Supersaturated Models

A problem with many QTL approaches is that there are far more parameters (p) to estimate than there are independent samples (n). Case in point: epistasis

Such supersaturated models arise commonly in Genomics. How do we deal with them?

One approach is to have all parameters included, but some are shrunk back (regressed) towards zero by assigning them a very small posterior variance

Shrinkage estimators

Shrinkage estimates: Rather than adding interaction terms one at a time, a shrinkage method starts with all interactions included, and then shrinks most back to zero.

Under a Bayesian analysis, any effect is *random*. One can assume the effect for (say) interaction *ij* is drawn from a normal with mean zero and variance σ^2_{ij}

Further, the interaction-specific variances are themselves random variables drawn from a hyperparameter distribution, such as an inverse chi-square.

One then estimates the hyperparameters and uses these to predict the variances, with effects with small variances shrinking back to zero, and effects with large variances remaining in the model.

What is a "QTL"

- A detected "QTL" in a mapping experiment is a region of a chromosome detected by linkage.
- Usually large (typically 10-40 cM)
- When further examined, most "large" QTLs turn out to be a linked collection of locations with increasingly smaller effects
- The more one localizes, the more subregions that are found, and the smaller the effect in each subregion
- This is called fractionation

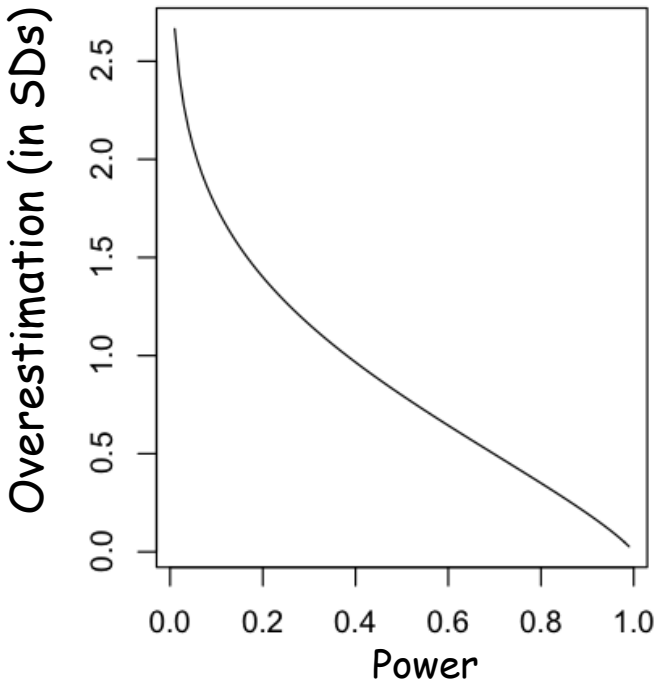
Limitations of QTL mapping

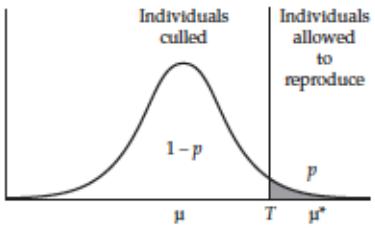
- Poor resolution (~20 cM or greater in most designs with sample sizes in low to mid 100's)
  - Detected "QTLs" are thus large chromosomal regions
- Fine mapping requires either
  - Further crosses (recombinations) involving regions of interest (i.e., RILs, NILs)
  - Enormous sample sizes
    - If marker-QTL distance is 0.5cM, require sample sizes in excess of 3400 to have a 95% chance of 10 (or more) recombination events in sample
    - 10 recombination events allows one to separate effects that differ by ~ 0.6 SD

Limitations of QTL mapping (cont)

- "Major" QTLs typically fractionate
  - QTLs of large effect (accounting for > 10% of the variance) are routinely discovered.
  - However, a large QTL peak in an initial experiment generally becomes a series of smaller and smaller peaks upon subsequent fine-mapping.
- The Beavis effect:
  - When power for detection is low, marker-trait associations declared to be statistically significant significantly overestimate their true effects.
  - This effect can be very large (order of magnitude) when power is low.

Beavis effect is akin to a selection intensity





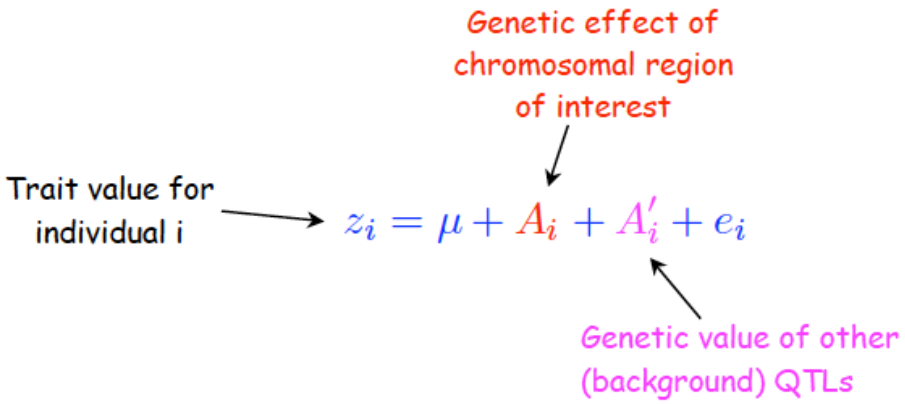
$$\bar{i} = \frac{S}{\sigma} = \frac{\varphi(x_{[1-p]})}{p}$$

Outbred populations

- When we move from the simple framework of an inbred line cross QTL design to a set of parents from an outbred population, complications arise as the parents don't all have the same genotypes
  - Differences in linkage phase
  - Many uninformative as to linkage (varies over makers)
  - Possibility of multiple alleles
- Result: express marker effects in terms of the variance in trait value it explains, rather than in terms of mean marker effects

General Pedigree Methods

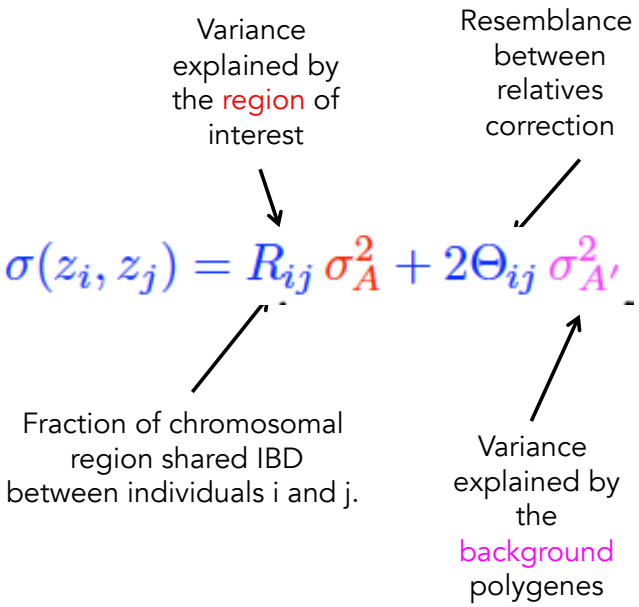
Random effects (hence, variance component) method for detecting QTLs in general pedigrees



The model is rerun for each marker

$$z_i = \mu + A_i + A_i' + e_i$$

The covariance between individuals i and j is thus



Assume z is MVN, giving the covariance matrix as

$$\mathbf{V} = \mathbf{R}\,\sigma_A^2 + \mathbf{A}\,\sigma_{A'}^2 + \mathbf{I}\,\sigma_e^2$$

Here

$$\mathbf{R}_{ij} = \begin{cases} 1 & \text{for } i = j \\ \widehat{R}_{ij} & \text{for } i \neq j \end{cases}, \qquad \mathbf{A}_{ij} = \begin{cases} 1 & \text{for } i = j \\ 2\Theta_{ij} & \text{for } i \neq j \end{cases}$$

Estimated from marker data

Estimated from the pedigree

The resulting likelihood function is

$$\ell(\mathbf{z}\,|\,\mu, \sigma_A^2, \sigma_{A'}^2, \sigma_e^2) = \frac{1}{\sqrt{(2\pi)^n\,|\mathbf{V}|}} \exp\left[-\frac{1}{2}(\mathbf{z}-\mu)^T\,\mathbf{V}^{-1}\,(\mathbf{z}-\mu)\right]$$

A significant σ_A^2 indicates a linked QTL.

Association & LD mapping

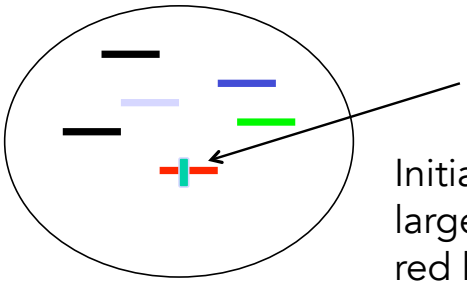
Mapping major genes (LD mapping) vs. trying to Map QTLs (Association mapping)

Idea: Collect random sample of individuals, contrast trait means over marker genotypes

If a dense enough marker map, likely population level linkage disequilibrium (LD) between closely-linked genes

Fine-mapping genes

Suppose an allele causing an effect on the trait arose as a single mutation in a closed population



New mutation arises on red chromosome

Initially, the new mutation is largely associated with the red haplotype

Hence, markers that define the red haplotype are likely to be associated (i.e. in LD) with the mutant allele

Background: Association mapping

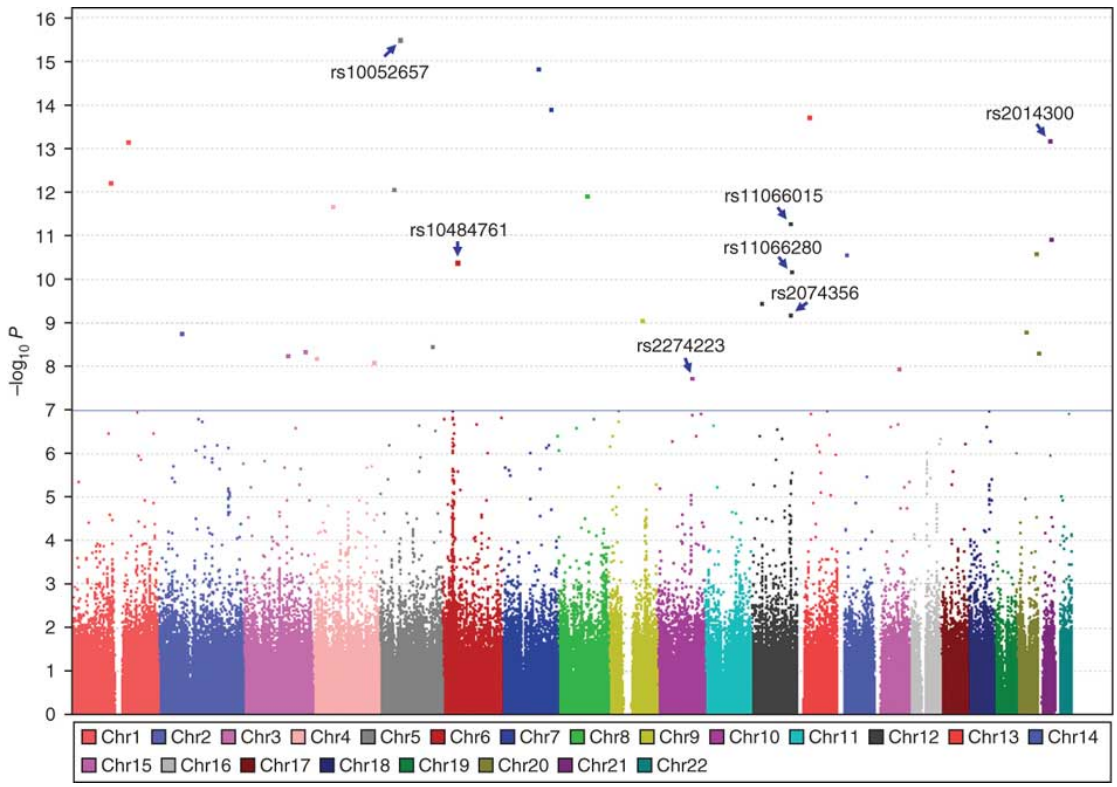
- If one has a very large number of SNPs, then new mutations (such as those that influence a trait) will be in LD with very close SNPs for hundreds to thousands of generations, generating a marker-trait association.
  - Association mapping looks over all sets of SNPs for trait -SNP associations. GWAS = genome-wide association studies.
  - This is also the basis for genomic selection
- Main point from extensive human association studies
  - Almost all QTLs have very small effects
  - Marker-trait associations do not fully recapture all of the additive variance in the trait (due to incomplete LD)
  - This has been called the "missing heritability problem" by human geneticists, but not really a problem at all (more shortly).

Association mapping

- Marker-trait associations within a population of unrelated individuals
- Very high marker density (~ 100s of markers/cM) required
  - Marker density no less than the average track length of linkage disequilibrium (LD)
- Relies on very slow breakdown of initial LD generated by a new mutation near a marker to generate marker-trait associations
  - LD decays very quickly unless very tight linkage
  - Hence, resolution on the scale of LD in the population(s) being studied (1 ~ 40 kB)
- Widely used since mid 1990's. Mainstay of human genetics, strong inroads in breeding, evolutionary genetics
- Power a function of the genetic variance of a QTL, not its mean effects

Manhattan plots

- The results for a Genome-wide Association study (or GWAS) are typically displayed using a Manhattan plot.
  - At each SNP, -ln(p), the negative log of the p value for a significant marker-trait association is plotted. Values above a threshold indicate significant effects
  - Threshold set by Bonferroni-style multiple comparisons correction
  - With n markers, an overall false-positive rate of p requires each marker be tested using p/n.
  - With n = 10^6 SNPs, p must exceed $0.01/10^6$ or 10^{-8} to have a control of 1% of a false-positive



Population Stratification

When population being sampled actually consists of several distinct subpopulations we have lumped together, marker alleles may provide information as to which group an individual belongs. If there are other risk factors in a group, this can create a false association btw marker and trait

Example. The Gm marker was thought (for biological reasons) to be an excellent candidate gene for diabetes in the high-risk population of Pima Indians in the American Southwest. Initially a very strong association was observed:

Gm^+	Total	% with diabetes
Present	293	8%
Absent	4,627	29%

Gm^+	Total	% with diabetes
Present	293	8%
Absent	4,627	29%

Problem: freq(Gm^+) in Caucasians (lower-risk diabetes Population) is 67%, Gm^+ rare in full-blooded Pima

The association was re-examined in a population of Pima that were 7/8th (or more) full heritage:

Gm^+	Total	% with diabetes
Present	17	59%
Absent	1,764	60%

Linkage vs. Association

The distinction between linkage and association is subtle, yet critical

Marker allele M is associated with the trait if

Cov(M,y) ≠ 0

While such associations can arise via linkage, they can also arise via population structure.

Thus, association DOES NOT imply linkage, and linkage is not sufficient for association

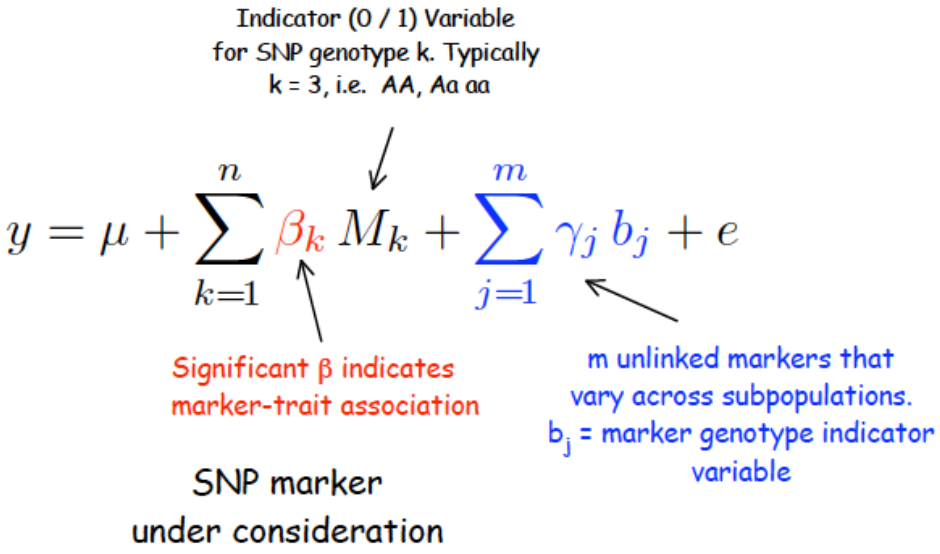
Accounting for population structure

- Three classes of approaches proposed
  - 1) Attempts to correct for common pop structure signal (regression/PC methods)
  - 2) Attempts to first assign individuals into subpopulations and then perform association mapping in each set (Structure)
  - 3) Mixed models that use all of the marker information (Tassle, EMMA, many others)
    - These can also account for cryptic relatedness in the data set, which also causes false-positives.

Regression Approaches

One approach to control for structure is simply to include a number of markers, outside of the SNP of interest, chosen because they are expected to vary over any subpopulations

How might you choose these in a sample? Try those markers (read STRs) that show the largest departure from Hardy-Weinberg, as this is expected in markers that vary the most over subpopulations.



Variations on this theme (eigenstrat) --- use all of the marker information to extract a set of significant PCs, which are then included in the model as cofactors

Structured Association Mapping

Pritchard and Rosenberg (1999) proposed Structured Association Mapping, wherein one assumes k subpopulations (each in Hardy-Weinberg).

Given a large number of markers, one then attempts to assign individuals to groups using an MCMC Bayesian classifier

Once individuals assigned to groups, association mapping without any correction can occur in each group.

Mixed-model approaches

- Mixed models use marker data to
  - Account for population structure
  - Account for cryptic relatedness
- Three general approaches:
  - Treat a single SNP as fixed
    - TASSLE, EMMA
  - Treat a single SNP as random
    - General pedigree method
  - Fit all of the SNPs at once as random
    - GBLUP

Structure plus Kinship Methods

Association mapping in plants offer occurs by first taking a large collection of lines, some closely related, others more distantly related. Thus, in addition to this collection being a series of subpopulations (derivatives from a number of founding lines), there can also be additional structure within each subpopulation (groups of more closely related lines within any particular lineage).

$$Y = X\beta + Sa + Qv + Zu + e$$

Fixed effects in blue, random effects in red

This is a mixed-model approach. The program TASSEL runs this model.

Q-K method

$$Y = X\beta + Sa + Qv + Zu + e$$

β = vector of fixed effects

a = SNP effects (fits SNPs one at a time)

v = vector of subpopulation effects (STRUCTURE)
Q_{ij} = Prob(individual i in group j). Determined from STRUCTURE output

u = shared polygenic effects due to kinship.
Cov(u) = var(A)*A, where the relationship matrix A estimated from marker data matrix K, also called a GRM – a genomic relationship matrix

Which markers to include in K?

- Best approach is to leave out the marker being tested (and any in LD with it) when construction the genomic relationship matrix
  - LOCO approach – leave out one chromosome (which the tested marker is linked to)
- Best approach seems to be to use most of the markers
- Other mixed-model approaches along these lines

Treat Single SNP as random: General Pedigree method

$$\mathbf{V} = \mathbf{R}\,\sigma_A^2 + \mathbf{A}\,\sigma_{A'}^2 + \mathbf{I}\,\sigma_e^2$$

Here

$$\mathbf{R}_{ij} = \begin{cases} 1 & \text{for } i = j \\ \widehat{R}_{ij} & \text{for } i \neq j \end{cases}, \qquad \mathbf{A}_{ij} = \begin{cases} 1 & \text{for } i = j \\ 2\Theta_{ij} & \text{for } i \neq j \end{cases}$$

Estimated from marker data

Estimated from the pedigree

The resulting likelihood function is

$$\ell(\mathbf{z}\,|\,\mu, \sigma_A^2, \sigma_{A'}^2, \sigma_e^2) = \frac{1}{\sqrt{(2\pi)^n\,|\mathbf{V}|}} \exp\left[-\frac{1}{2}(\mathbf{z}-\mu)^T\,\mathbf{V}^{-1}\,(\mathbf{z}-\mu)\right]$$

A significant σ_A^2 indicates a linked QTL.

GBLUP

- The Q-K method tests SNPs one at a time, treating them as fixed effects
- The general pedigree method (slides 24-26) also tests one marker at a time, treating them as random effects
- Genomic selection can be though of as estimating all of the SNP effects at once and hence can also be used for GWAS

BLUP, GBLUP, and GWAS

- Pedigree information gives EXPECTED value of shared sites (i.e., ½ for full-sibs)
  - A matrix in BLUP
  - The actual realization of the fraction of shared genes for a particular pair of relatives can be rather different, due to sampling variance in segregation of alleles
  - GRM (or K or marker matrix M)
  - Hence "identical" relatives can differ significantly in faction of shared regions
  - Dense marker information can account for this

The general setting

- Suppose we have n measured individuals (the n x 1 vector **y** of trait values)
- The n x n relationship matrix **A** gives the relatedness among the sampled individuals, where the elements of **A** are obtained from the pedigree of measured individuals
- We may also have p (>> n) SNPs per individual, where the n x p marker information matrix **M** contains the marker data, where M_{ij} = score for SNP j (i.e., 0 for 00, 1 for 10, 2 for 11) in individual i.

Covariance structure of random effects

- A critical element specifying the mixed model is the covariance structure (matrix) of the vector **u** of random effects
- Standard form is that Cov(**u**) = variance component * matrix of known constants
  - This is the case for pedigree data, where **u** is typically the vector of breeding values, and the pedigree defines a relationship matrix **A**, with Cov(**u**) = Var(A) * **A**, the additive variance times the relationship matrix
  - With marker data, the covariance of random effects are functions of the marker information matrix **M**.
    - If **u** is the vector of p marker effects, then Cov(**u**) = Var(m) * $\mathbf{M}^T\mathbf{M}$, the marker variance times the covariance structure of the markers.

$$Y = X\beta + Zu + e$$

Pedigree-based BV estimation: (BLUP)
u_{nx1} = vector of BVs, Cov(u) = Var(A) A_{nxn}

Marker-based BV estimation: (GBLUP)
u_{nx1} = vector of BVs, Cov(u) = Var(m) M^TM (n x n)

GWAS: u_{px1} = vector of marker effects,
Cov(u) = Var(m) MM^T (p x p)

Genomic selection: predicted vector of breeding values from marker effects, $GBV_{nx1} = M_{nxp}u_{px1}$.
Note that Cov(GBV) = Var(m) M^TM (n x n)

Lots of variations of these general ideas by adding additional assumptions on covariance structure.

GWAS Model diagnostics

The "Genomic Control" parameter λ

Devlin and Roeder (1999). Basic idea is that association tests (marker presence/absence vs. trait presence/absence) is typically done with a standard 2 x 2 χ^2 test.

When population structure is present, the test statistic now follows a scaled χ^2, so that if S is the test statistic, then $S/\lambda \sim \chi^2_1$ (so $S \sim \lambda\chi^2_1$) . Hence, population structure should inflate all of the tests (on average) by a common amount λ.

Hence, if we have suitably corrected for population structure, the estimated inflation factor λ among tests should be ~ 1.

A robust estimator for λ is offered from the medium (50% value) of the test statistics, so that for m tests

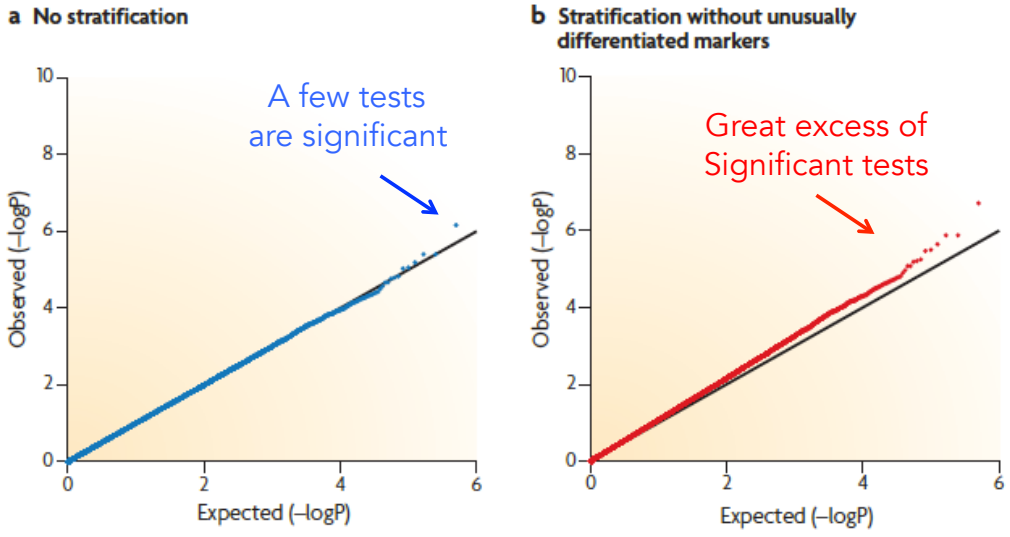
$$\widehat{\lambda} = \frac{\text{medium}\,(S_1, \cdot\cdot; S_m)}{0.456}$$

Genomic control λ as a diagnostic tool

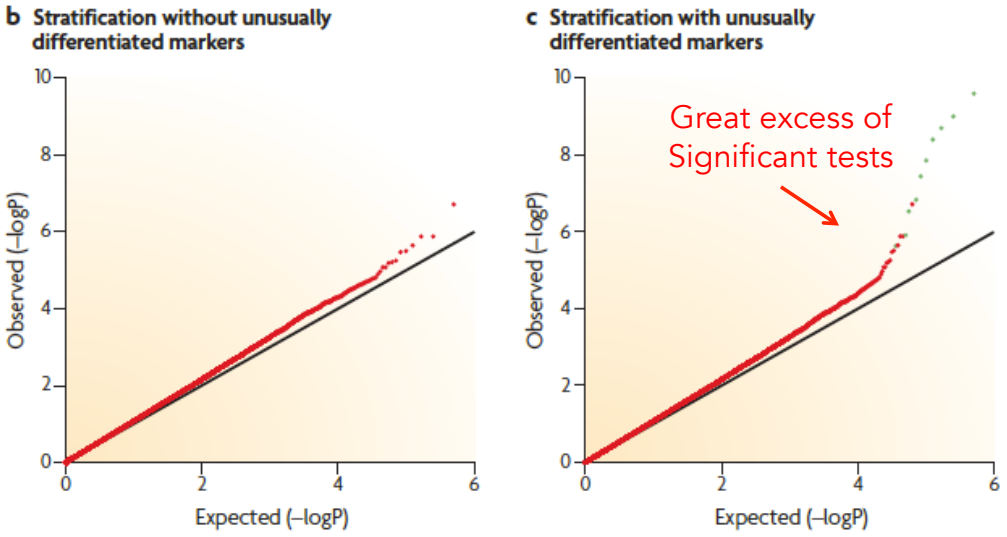
- Presence of population structure will inflate the λ parameter
- A value above 1 is considered evidence of additional structure in the data
  - Could be population structure, cryptic relatedness, or both
  - A lambda value less that 1.05 is generally considered benign
- One issue is that if the true polygenic model holds (lots of sites of small effect), then a significant fraction will have inflated p values, and hence an inflated λ value.
- Hence, often one computes the λ following attempts to remove population structure. If the resulting value is below 1.05, suggestion that structure has been largely removed.

P – P plots

- Another powerful diagnostic tool is the p-p plot.
- If all tests are drawn from the null, then the distribution of p values should be uniform.
  - There should be a slight excess of tests with very low p indicating true positives
- This gives a straight line of a log-log plot of observed (seen) and expected (uniform) p values with a slight rise near small values
  - If the fraction of true positives is high (i.e., many sites influence the trait), this also bends the p-p plot



Price et al. 2010 Nat Rev Gene 11: 459



As with using λ, one should construct p-p following some approach to correct for structure & relatedness to see if they look unusual.

Association mapping (power)

Q/q is the polymorphic site contributing to trait variation, M/m alleles (at a SNP) used as a marker

Let p be the frequency of M, and assume that Q only resides on the M background (complete disequilibrium)

Haloptype	Frequency	effect
QM	rp	a
qM	(1-r)p	0
qm	1-p	0

Haloptype	Frequency	effect
QM	rp	a
qM	(1-r)p	0
qm	1-p	0

Effect of m = 0

Effect of M = ar

Genetic variation associated with Q = $2(rp)(1-rp)a^2$
~ $2rpa^2$ when Q rare. Hence, little power if Q rare

Genetic variation associated with marker M is
$2p(1-p)(ar)^2 \sim 2pa^2r^2$

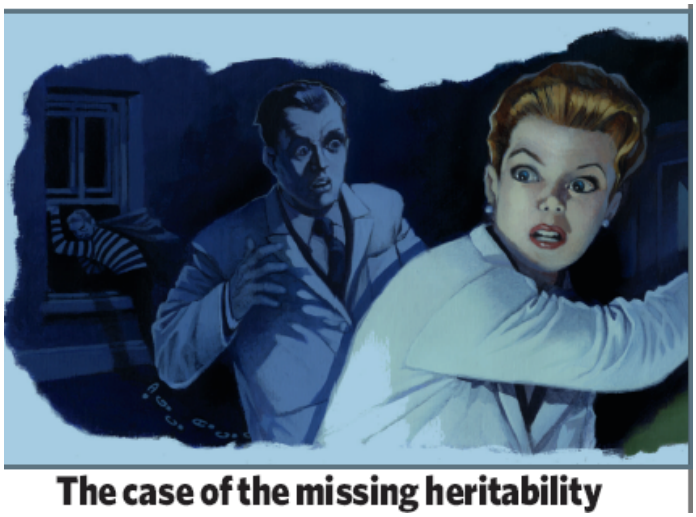
Ratio of marker/true effect variance is ~ r

Hence, if Q rare within the A class, even less power, as M only captures a fraction of the associated QTL.

Common variants

- Association mapping is only powerful for common variants
  - freq(Q) moderate
  - freq (r) of Q within M haplotypes modest to large
- Large effect alleles (a large) can leave small signals.
- The fraction of the actual variance accounted for by the markers is no greater than ~ ave(r), the average frequency of Q within a haplotype class
- Hence, don't expect to capture all of Var(A) with markers, esp. when QTL alleles are rare but markers are common (e.g. common SNPs, p > 0.05)
- Low power to detect G x G, G x E interactions

"How wonderful that we have met with a paradox. Now we have some hope of making progress" -- Neils Bohr



Infamous figure from *Nature* on the angst of human geneticists over the finding that all of their discovered SNPs still accounted for only a fraction of relative-based heritability estimates of human disease.

- "There is something simultaneously remarkable and encouraging about the fact that a centuries-old method requiring no more than a ruler, a pencil and (I suppose) a slide rule out performed, by an order of magnitude, the fruits of the genomic revolution"
- --Ben Sheldon (2013)

The "missing heritability" paradox

- A number of GWAS workers noted that the sum of their significant marker variances was much less (typically 10%) than the additive variance estimated from biometrical methods
- The "missing heritability" problem was birthed from this observation.
- Not a paradox at all
  - Low power means small effect (i.e. variance) sites are unlikely to be called as significant, esp. given the high stringency associated with control of false positives over tens of thousands of tests
  - Further, even if all markers are detected, only a fraction ~ r (the frequency of the causative site within a marker haplotype class) of the underlying variance is accounted for.

Lecture 7
Multi-Trait Models, Binary and Count Traits

Guilherme J. M. Rosa

University of Wisconsin-Madison

Introduction to Quantitative Genetics
SISG, Seattle
19 – 21 July 2017

OUTLINE

- Multiple-trait Model
- Repeatability Model
- Maternal Effects
- Generalized Linear Models

Animal Model

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

responses — incidence matrices — fixed effects — breeding values — residuals

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{A}\sigma_a^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_e^2 \end{bmatrix} \right)$$

Mixed Model Equations

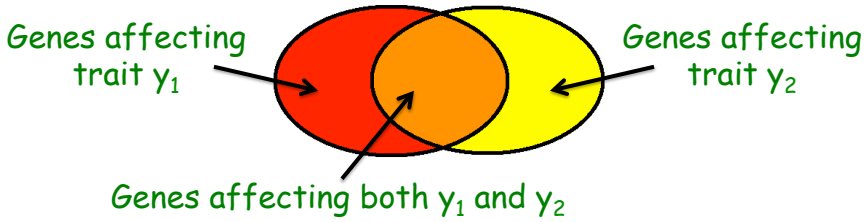
$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \lambda\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{bmatrix}$$

$$\lambda = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1 - h^2}{h^2}$$

BLUP: $\hat{\mathbf{u}} = (\mathbf{Z}^T\mathbf{Z} + \lambda\mathbf{A}^{-1})^{-1}\mathbf{Z}^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$

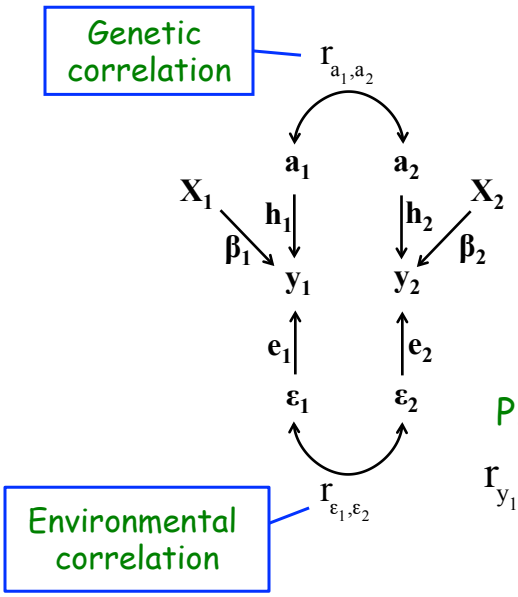
Genetic Correlation

Schematic representation of pleiotropy



- Pleiotropic genes affect both y_1 and y_2 resulting in a genetic correlation between the two traits
- In addition to pleiotropy, genetic correlations can be caused also by linkage disequilibrium (LD) between genes affecting the different traits. LD however is a 'temporary' cause of genetic correlation as recombination can breakdown LD over the generations

Multiple (Correlated) Traits



$$\begin{cases} h_j = \sqrt{h_j^2} \\ e_j = 1 - h_j \end{cases}$$

Phenotypic correlation:

$$r_{y_1,y_2} = h_1 h_2 r_{a_1,a_2} + e_1 e_2 r_{\varepsilon_1,\varepsilon_2}$$

Multiple (Correlated) Traits

The animal model can be extended for the joint analysis of multiple traits

Let the model for each of k traits be:

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{Z}_j\mathbf{a}_j + \boldsymbol{\varepsilon}_j$$

where j is an index to indicate the trait (j = 1, 2,…,k).

For the joint analysis of the k trait, the model becomes:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon}$$

with design matrices given by:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_k \end{bmatrix} \qquad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_k \end{bmatrix}$$

Multiple (Correlated) Traits

In this case it is assumed that:

$$\text{Var}\begin{bmatrix} \mathbf{a} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{G} \otimes \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \otimes \mathbf{I} \end{bmatrix}$$

where **G** and **Σ** are the genetic and residual variance-covariance matrices, given by:

$$\mathbf{G} = \begin{bmatrix} \sigma^2_{a_1} & \sigma_{a_1a_2} & \cdots & \sigma_{a_1a_k} \\ \sigma_{a_1a_2} & \sigma^2_{a_2} & \cdots & \sigma_{a_2a_2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{a_1a_k} & \sigma_{a_2a_k} & \cdots & \sigma^2_{a_k} \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2_{\boldsymbol{\varepsilon}_1} & \sigma_{\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_2} & \cdots & \sigma_{\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_k} \\ \sigma_{\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_2} & \sigma^2_{\boldsymbol{\varepsilon}_2} & \cdots & \sigma_{\boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}_2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_k} & \sigma_{\boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}_2} & \cdots & \sigma^2_{\boldsymbol{\varepsilon}_k} \end{bmatrix}$$

Note: $\otimes$ represents the direct (Kronecker) product

Multiple (Correlated) Traits

The MME for multi-trait analyses are of the same form as before, i.e.:

$$\begin{bmatrix} \mathbf{X}'(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I})\mathbf{X} & \mathbf{X}'(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I})\mathbf{Z} \\ \mathbf{Z}'(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I})\mathbf{X} & \mathbf{Z}'(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I})\mathbf{Z}+\mathbf{G}^{-1}\otimes\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I})\mathbf{y} \\ \mathbf{Z}'(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I})\mathbf{y} \end{bmatrix}$$

from which the BLUEs and BLUPs of **β** and **a** can be obtained.

Multiple (Correlated) Traits

The dimensionality of multi-trait MME, however, can become a hurdle for solving it when more than two or three traits are considered

An alternative for the analysis of multiple traits is to use a canonical transformation of the traits, which consists of transforming the vectors of correlated traits into a new vector of uncorrelated variables

In such case, each transformed variable can be analyzed independently using standard single trait models, and subsequently the estimated breeding values are transformed back to the original scale of measurement

Repeatability Model




Repeatability Model

For the analysis of repeated measurements, environmental effects can be partitioned into permanent and temporary effects

In this case, the mixed model, usually called 'repeatability model', can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{p} + \boldsymbol{\varepsilon}$$

where $\mathbf{p} \sim N(\mathbf{0}, \mathbf{I}\sigma_p^2)$ is the vector of permanent environmental effects, with each level pertaining to a common effect to all observations of each animal

Repeatability Model

It is often assumed that **a**, **p**, and **ε**, which are independent from each other

Under these assumptions, the MME becomes:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{W} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}+\lambda_a\mathbf{A}^{-1} & \mathbf{Z}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{Z} & \mathbf{W}'\mathbf{W}+\lambda_p\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}$$

with $\lambda_a = \sigma^2_{\boldsymbol{\varepsilon}}/\sigma^2_a$ and $\lambda_p = \sigma^2_{\boldsymbol{\varepsilon}}/\sigma^2_p$

Repeatability Model

An important definition related to repeated measurements refers to repeatability (r), which is given by the intraclass correlation, i.e., the ratio of the within-individual (or between repeated measurements) to the phenotypic variances:

$$r = \frac{\sigma^2_a + \sigma^2_p}{\sigma^2_y} = \frac{\sigma^2_a + \sigma^2_p}{\sigma^2_a + \sigma^2_p + \sigma^2_{\varepsilon}}$$

The repeatability coefficient measures the correlation between records on the same animal, and so it is useful for example in the estimation of producing ability and an animal

Maternal Effects

Maternal Effects

There are some traits of interest in livestock, such as weaning weight in beef cattle, in which progeny performance is affected by the dam's ability to affect the calf's environment, such as in the form of nourishment through her milk production, the quantity and quality of which is in part genetically determined

In such cases, dams contribute to the performance of their progeny not only through the genes passed to the progeny (the "direct genetic effects") but also through their ability to provide a suitable environment (the "indirect genetic effects")

Maternal Effects

Maternally influenced traits can be analyzed by using a model as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{K}\mathbf{m} + \mathbf{W}\mathbf{p} + \boldsymbol{\varepsilon}$$

where **m** is a vector of random maternal genetic effects, and **p** is a vector of random maternal permanent environmental effects

It is assumed that $\mathbf{m} \sim N(\mathbf{0}, \mathbf{A}\sigma_m^2)$ and $\mathbf{p} \sim N(\mathbf{0}, \mathbf{I}\sigma_p^2)$, and quite often a covariance structure between direct and maternal additive genetic effects is considered, assumed equal to $\mathbf{A}\sigma_{a,m}$

Computing Strategies

Solving the MME does not necessary require the inversion of the coefficient matrix **C**

More computationally convenient alternatives for solving high dimensional systems of linear equations include methods based on iteration on the MME, such as the Jacobi or Gauss-Seidel iteration, and the "iteration on the data" strategy, which is commonly used methodology in national genetic evaluations involving millions of records

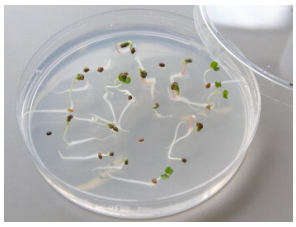
Generalized Linear Mixed Models

The models discussed so far assumed a Gaussian (normal) distribution of the phenotypic traits

Often however phenotypic traits are expressed a a binary (e.g., pregancy in dairy cattle, or germination in seeds) or count variable (e.g., litter size in swine, or fruits in trees)

In such cases the linear (Gaussian) model is not appropriate, and a generalized linear model (GLM) approach is necessary

Generalized Linear Mixed Models



Generalized Linear Mixed Models

GLM can actually model outcomes (response variables) generated from any distribution from the exponential family, which includes the normal, binomial, Poisson and gamma distributions, among others

The GLM consists of three elements:

1. Probability distribution from the exponential family.
2. Linear predictor $\eta = X\beta$
3. Link function g such that $E(Y) = \mu = g^{-1}(\eta)$.

Generalized Linear Mixed Models

Notice that the Gaussian model is a specific case of the GLM, with the normal distribution and an identity link function

In the case of Generalized Linear Mixed Models, including the applications in animal/plant breeding, the model is defined as:

1. Probability distribution from the exponential family.
2. Linear predictor $\eta = X\beta + Zu$
3. Link function g such that $E(Y|u) = \mu = g^{-1}(\eta)$

GLMM in R

GLMM can be implemented in R using the package lme4

lme4, however, assumes independence between levels of random effects, and as such it is not suitable for many animal/ plant breeding applications

pedigreemm is an R package that uses lme4 with a Cholesky decomposition strategy to overcome this problem

pedigreemm

An R package for fitting generalized linear mixed models in animal breeding

$$g\left(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{U}}\right) = \mathbf{Z}\mathbf{u} + \mathbf{X}\boldsymbol{\beta}$$

$$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{U}} = E\left[\mathbf{Y}\middle|\mathbf{U} = \mathbf{u}\right] \qquad \mathbf{u} \sim N\left(\mathbf{0}, \mathbf{A}\sigma_u^2\right)$$

$$\mathbf{u}^* = \mathbf{L}^{-1}\mathbf{u} \longrightarrow g\left(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{U}}\right) = \mathbf{Z}\mathbf{L}\left(\mathbf{L}^{-1}\mathbf{u}\right) + \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}^*\mathbf{u}^* + \mathbf{X}\boldsymbol{\beta}$$

$$\mathbf{A} = \mathbf{L}\mathbf{L}' \qquad \mathbf{u}^* \sim N\left(\mathbf{0}, \mathbf{I}\sigma_u^2\right)$$

(Harville and Callanan 1989)

Technical note: An R package for fitting generalized linear mixed models in animal breeding[1]

A. I. Vazquez,*[2] D. M. Bates,† G. J. M. Rosa,* D. Gianola,*‡ and K. A. Weigel*

*Department of Dairy Science, †Department of Statistics, and ‡Department of Animal Sciences, University of Wisconsin, Madison 53706

J. Anim. Sci. 2010. 88:497–504
doi:10.2527/jas.2009-1952

Data Set 1. Milk production records of 3,397 lactations from first- through fifth-parity Holsteins were available. These records were from 1,359 cows, daughters of 38 sires in 57 herds. Records are in the *milk* data set in the *pedigreemm* package. The data were downloaded from the USDA site (http://www.aipl.arsusda.gov/). All lactation records represent cows with at least 100 d in milk, with an average of 347 d. Milk yield ranged from 4,065 to 19,345 kg estimated for 305 d, averaging 11,636 kg. There were 1,314, 1,006, 640, 334, and 103 records for first-, second-, third-, fourth-, and fifth-lactation animals, respectively. A 5-generation pedigree of the cows with a total of 6,547 animals was used in the analysis (http://www.aipl.arsusda.gov/). The pedigree information is available in the *pedCows* and *pedCowsR* pedigree objects also included in the package; the second one is a lighter pedigree (with 70% of the information on *pedCows*). The milk production data used in the first 2 examples are described below.

pedigreemm example

Lecture 08:
Associate effects models, kin/group selection, inclusive fitness

Bruce Walsh lecture notes
Introduction to Mixed Models
SISG, Seattle
19 – 21 July 2017

Associative effects models

- A very powerful recent development in quantitative genetics (although the idea dates back to Griffin's work in the 1960s) is the notion of direct vs. associative (or social, or indirect genetic) effects
- This idea unifies kin and group selection, offers models for the evolution of social (group-level) traits, and shows why selection can often fail
- The basic idea is that the phenotype of a target individual is a function of some intrinsic direct value and also the phenotypes of those individuals with which it interacts.

Direct & Associative effects

- Consider egg production from chickens raised in cages. Production is a function of both a chicken's own genetics and the environment (her other cage-mates)
  - Direct effects = intrinsic egg production
  - Associative effects = competitive ability
- Suppose our focal individual (i) interacts with n-1 others in a group

$$z_i = P_{d,i} + \sum_{j \neq i}^{n} P_{j,s}$$

Direct and associative effects can be antagonistic

- Consider a plant with a trait that allows it to more efficiently garner resources
- This gives it a high direct effect but a negative associative effect --- it reduces the trait values in those individuals with which it interacts
- Thus, the best performing single plants can have very low average plot performance

Example 20.1. This point was made in a classic paper by Weibe et al (1976), who examined yield in mixed- versus single-genotype plots of barley. They observed that genotypes which yielded well in mixed stands had poorer yield in pure stands, while those genotypes that did poorly in mixed stands had the highest yield in pure stands. In our framework, we could imagine that lines which do well in mixed stands have both high direct effects and high negative associative effects, suppressing the phenotypes of their neighbors. When grown in a pure strand, the high negative associative effects suppress plot yield. Conversely, lines that perform poorly in mixed strands might have low direct effects but high positive associative effects, so that the phenotypes of their neighbors are enhanced (or atleast not hindered). When grown as a pure strand, these high positive associative effects more than compensate for the low direct effects, increasing yield.

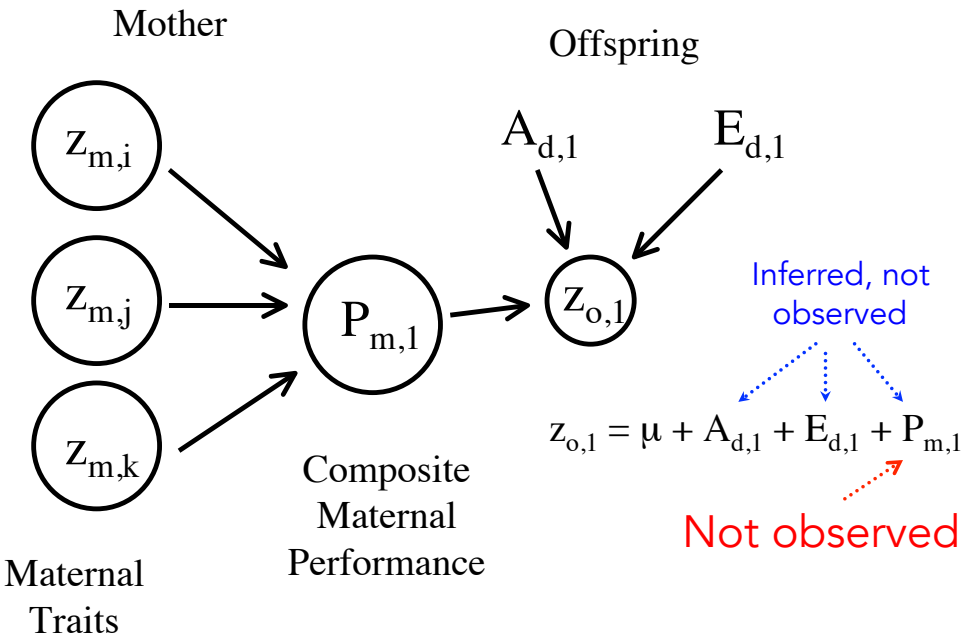
Roots of associative-effects models trace to maternal effects

- Maternal effects are a classic example of associative effects (maternal performance).
- Two different approaches to model maternal effects
  - Falconer model: an observed trait value (e.g., litter size) influences offspring. Trait-based
  - Willham model: Maternal performance is a latent (unobserved) variable, and hence we don't need to specify it. Variance-component based. We focus on these models here.

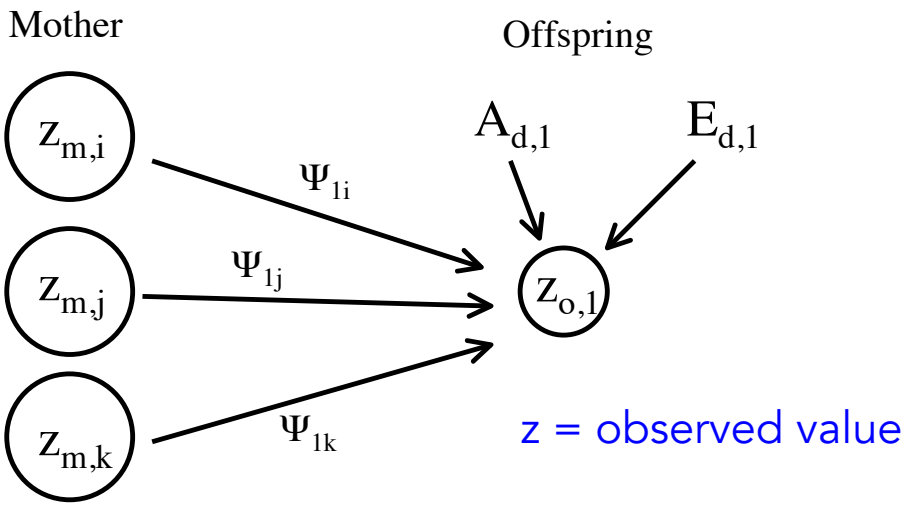
Trait-based vs. variance-component models

- Trait based:
  - Trait values of associative effects in group members are observed
- Variance-component models
  - A composite latent (unmeasured) variable for associative effects is created

Variance components



Trait-based models



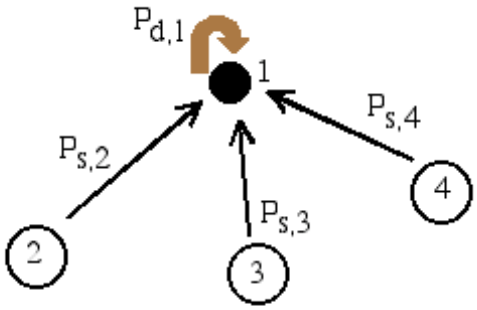
$$z_{o,1} = \mu + A_{d,1} + E_{d,1} + \Psi_{1,i}\, z_{m,i} + \Psi_{1,j}\, z_{m,j} + \Psi_{1,k}\, z_{m,k}$$

Decomposition

- Consider the phenotype of a focal individual
- Sum of a direct effect and an associative effect
- Both of these can have a breeding value and an environment (residual) deviation
- The breeding values of the direct & associative effects can be correlated
- This is a multiple-traits problem

$$z_i = P_{d,i} + \sum_{j \neq i}^{n} P_{j,s}$$

- i's phenotype z_i is the sum of its direct effect ($P_{d,i}$) plus the sum of the associative (or social) effects ($P_{s,j}$) from its n-1 group members



$z_1 = P_{d,1} + P_{s,2} + P_{s,3} + P_{s,4}$

Breeding values for direct (A_d) and associative (A_s) effects

- Can express the phenotype of i in terms of its direct breeding value ($A_{d,i}$) and the associative breeding values ($A_{s,j}$) of its group mates

$$z_i = \mu + (A_{d_i} + E_{d_i}) + \sum_{j \neq i} (A_{s_j} + E_{s_j})$$

$$z_i = \mu + A_{d_i} + \sum_{j \neq i} A_{s_j} + e_i, \qquad e_i = E_{d_i} + \sum_{j \neq i} E_{s_j}$$

Total response

The trait mean equals the mean of the direct effects plus the means of the associative effects,

$$\mu_z = \mu_{A_d} + (n-1)\mu_{A_s}$$

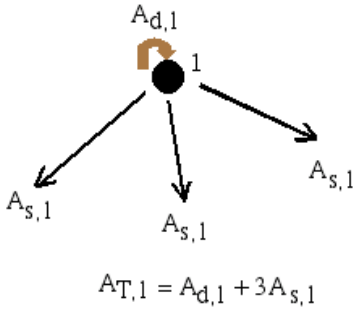
Total response is the sum of the response R_d in the direct breeding values plus the sum of the responses R_s in the associative effects breeding values,

$$R_z = R_d + (n-1)R_s$$

Total breeding value

The key to predicting response is the total breeding value of an individual, where

$$A_{T,i} = A_{d,i} + (n-1)A_{s,i}$$



Note that part ($A_{s,i}$) of the total breeding value of i never appears in its phenotype. Must either use informative from relatives or the group to estimate it.

h^2 and τ^2

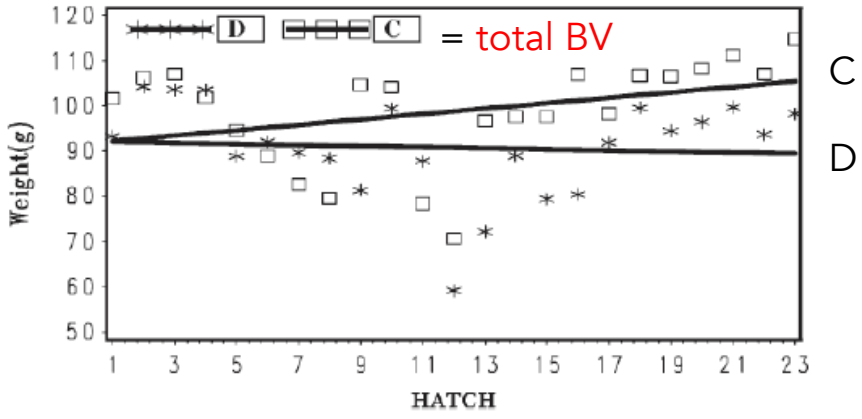
- τ^2, the analog for h^2, is the ratio of the total breeding value to the individual phenotypic variance
  - $\tau^2 = Var(A_T)/Var(z)$
- Note that, unlike h^2, τ^2 can exceed one,
- Why? A potentially large fraction of A_T never appears in z, and hence Var(z)
  - $Var(A_T) = Var(A_d) + (n-1)Var(A_s)$
  - $\tau^2 = Var(A_d)/Var(z) + (n-1)Var(A_s)/Var(z)$
  - $= h^2 + (n-1)Var(A_s)/Var(z)$

BLUP estimation

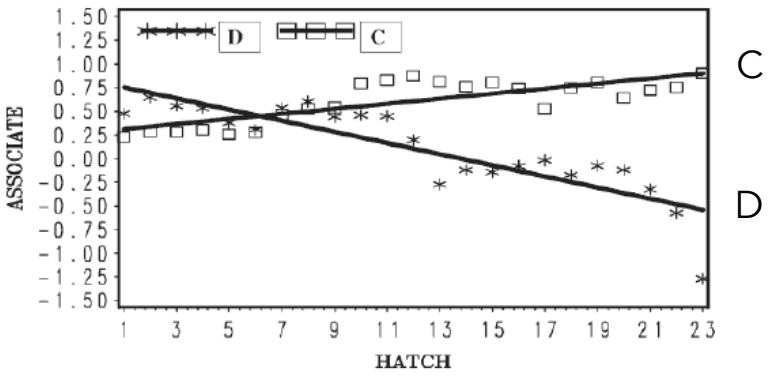
- While the total breeding value cannot be estimated directly from an individual's phenotype, using an appropriate mixed model, we can obtain
  - BLUPs of Direct breeding values (A_d)
  - BLUPs of Associative (or social) BVs (A_s)
  - REML estimates of $\sigma^2(A_d)$, $\sigma^2(A_s)$, and the direct-associate effects covariance $\sigma(A_d, A_s)$

This works: Muir's result

- Bill Muir (Purdue University) selection on six-week weight in Japanese quail over 23 generations using two different schemes
  - BLUP selection on estimated direct BV (D)
    - Denoted by D-BLUP
  - BLUP selection on estimated total BV
    - Denoted by C-BLUP



Weighted increased under selection using total BV (C), decreased under selection using direct BV (D).



C

D

Under BLUP selection on direct BV (D), significant decline in the mean social value, which over-rode the positive response in the direct value

Under BLUP selection of total BV (C), *both* increase

The mixed model

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_d\,\mathbf{a}_d + \mathbf{Z}_s\,\mathbf{a}_s + \mathbf{e}$$

Example: Individuals 1-4 and 5-8 are half sibs from unrelated families

$$\mathbf{A} = \begin{pmatrix} 1 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 & 0 \\ 0.25 & 1 & 0.25 & 0.25 & 0 & 0 & 0 & 0 \\ 0.25 & 0.25 & 1 & 0.25 & 0 & 0 & 0 & 0 \\ 0.25 & 0.25 & 0.25 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0.25 & 0.25 & 0.25 \\ 0 & 0 & 0 & 0 & 0.25 & 1 & 0.25 & 0.25 \\ 0 & 0 & 0 & 0 & 0.25 & 0.25 & 1 & 0.25 \\ 0 & 0 & 0 & 0 & 0.25 & 0.25 & 0.25 & 1 \end{pmatrix}$$

Filling out Z_s

- Suppose group one contains individuals 1, 2, 5, 6. The resulting values for these individuals become
  - $z_1 = m + A_{d1} + A_{s2} + A_{s5} + A_{s6} + e$
  - $z_2 = m + A_{d2} + A_{s1} + A_{s5} + A_{s6} + e$
  - $z_5 = m + A_{d5} + A_{s1} + A_{s2} + A_{s6} + e$
  - $z_6 = m + A_{d6} + A_{s1} + A_{s2} + A_{s5} + e$
  - The result Z_d and Z_s incident matrices become

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_d\,\mathbf{a}_d + \mathbf{Z}_s\,\mathbf{a}_s + \mathbf{e}$$

$$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \\ z_7 \\ z_8 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{a}_d = \begin{pmatrix} A_{d,1} \\ A_{d,2} \\ A_{d,3} \\ A_{d,4} \\ A_{d,5} \\ A_{d,6} \\ A_{d,7} \\ A_{d,8} \end{pmatrix}, \quad \mathbf{Z}_d = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} = \mathbf{I}_8$$

Group one contains individuals 1,2,5,6; while group two contains 3,4,7,8.

$$\mathbf{Z}_s = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{a}_s = \begin{pmatrix} A_{s,1} \\ A_{s,2} \\ A_{s,3} \\ A_{s,4} \\ A_{s,5} \\ A_{s,6} \\ A_{s,7} \\ A_{s,8} \end{pmatrix}$$

Lots of hidden variation to exploit

- Bergsma et al. (2008) examined four traits in 14,000 pigs grown in pens of 6-12 animals.
- Heritability for these traits was estimated in a model without social effects,

	Growth	Back fat	Muscle	Intake
$\sigma^2(A)$	2,583	2.83	7.94	41,275
h^2	0.37	0.36	0.25	0.41

Next, a model was fit allowing for heritable social effects, $\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_d\mathbf{a}_d + \mathbf{Z}_s\mathbf{a}_s + \mathbf{Z}_c\mathbf{c} + \mathbf{e}$, which gave estimates of

	Growth	Back fat	Muscle	Intake
$\sigma^2(A_d)$	1,522	2.75	6.68	16,950
h^2_d	0.21	0.35	0.21	0.17
$\sigma^2(A_s)$	51	0.01	0.03	596
$\sigma^2(A_T)$	5,208	3.19	10.35	68,687
τ^2	0.71	0.41	0.32	0.70

Here $h^2_d = \sigma^2(A_d)/\sigma^2(z)$, while $\tau^2 = \sigma^2(A_T)/\sigma^2(z)$. h^2_d measures the response potential under phenotypic selection, while $\tau^2 \geq h^2_d$ measures the total genetic potential for improvement under specialized selection designs.

	Growth	Back fat	Muscle	Intake
$\sigma^2(A)$	2,583	2.83	7.94	41,275
h^2	0.37	0.36	0.25	0.41

Hence, for growth and food intake, lots of additional genetic variation for trait response lies “hidden” in associative effects.

Consequences

- How can we exploit this variation in breeding?
- What are the consequences for evolutionary biologists?
- Need to consider selection response
  - Has both a direct and associative effects component

$$z_i = \mu + (A_{d_i} + E_{d_i}) + \sum_{j \neq i} (A_{s_j} + E_{s_j}) \tag{20.1b}$$

We can write this compactly as

$$z_i = \mu + A_{d_i} + \sum_{j \neq i} A_{s_j} + e_i, \quad \text{where} \quad e_i = E_{d_i} + \sum_{j \neq i} E_{s_j} \tag{20.1c}$$

Since the environmental values have expected value zero, the mean phenotypic value in the group is just

$$\mu_z = \mu_{A_d} + (n-1)\mu_{A_s} \tag{20.1d}$$

Further, the change in the mean trait value within a group following selection is

$$\Delta\mu_z = \Delta\mu_{A_d} + (n-1)\Delta\mu_{A_s} \tag{20.1e}$$

total response

Direct response

Associative response

Example 20.2. Consider a trait in a group of four (unrelated) individuals, where (for illustrative purposes) we assume no environmental values so that $P_d = A_d$ and $P_s = A_s$. The population mean is 20, and the four group members have the following breeding values for direct, associative, and total effects:

Individual	A_d	A_s	A_T	$\sum_{j \neq i} A_{s_j}$	z
1	9	-4	-3	4	33
2	5	-1	2	1	26
3	-6	2	0	-2	12
4	-8	3	1	-3	9

Since $n = 4$, $A_T = A_d + 3\,A_s$. The sum $\sum_{j \neq i} A_{s_j}$ represents the contribution of the associative effects of the other three individuals to i's value. For example, for individual 1, the contributions from individuals 2 through 4 is -1 + 2 + 3 = 4. From Equation 20.1c, the phenotypic value we would observe is

$$z_i = 20 + A_{d_i} + \sum_{j \neq i} A_{s_j}$$

Individual one has the largest direct effect (9) and the largest observed trait value (33). This individual also has the most unfavorable associative value (-4), and the smallest total breeding value (-3). Conversely, it has the largest contribution (4) to its trait value from the associative effects of the other group members. Its high trait value is due to this combination of a high direct effect and a high contribution from the associative effects of the other group members. Its unfavorable associative effects do not appear in its own phenotype, but rather are expressed in the trait values of the other group members. As a result, its own phenotypic value is a poor predictor of A_T.

Individual	A_d	A_s	A_T	$\sum_{j \neq i} A_{s_j}$	z
1	9	-4	-3	4	33
2	5	-1	2	1	26
3	-6	2	0	-2	12
4	-8	3	1	-3	9

If the next generation is formed by crossing the two individuals (1 and 2) with the largest trait values, the expected offspring mean is 20 + (-3+2)/2 = 19.5, the mean plus their average total breeding values. Although the two largest individuals were chosen, the population mean *decreases*. Conversely, crossing the two smallest individuals gives an expected offspring mean of 20 + (0+1)/2 = 20.5, increasing the mean. While the two smallest individuals have the smallest direct effects, they also have the most favorable associative effects, and hence give a more favorable response. The greatest expected response occurs by crossing the two individuals (2 and 4) with the largest total breeding values, for an expected mean of 20 + (2+1)/2 = 21.5.

Response: It's about covariances

- Selection response is a function of the covariance between our unit u of selection and the total breeding value, $\sigma(A_T, u)$
  - $R = i * \sigma(A_T, u) / \sigma(u)$ (generalized breeder's Eq.)
- The "unit" could be a
  - single individual (individual selection)
  - The group mean (group selection)
  - Some index of these
- Members of a group can be
  - Unrelated
  - Related (kin selection)
- All these considerations influence $\sigma(A_T, u)$

The covariance between an individual's phenotype and total breeding value is

$$\sigma(z_i, A_{T_i}) = \sigma\left(\mu + A_{d_i} + \sum_{j \neq i} A_{s_j} + e_i, A_{d_i} + (n-1)A_{s_i}\right)$$

General expression

$$= \sigma\left(A_{d_i}, A_{d_i} + (n-1)A_{s_i}\right) + \sum_{j \neq i} \sigma\left(A_{s_j}, A_{d_i} + (n-1)A_{s_i}\right) \quad (20.4a)$$

For now, we assume unrelated group members, in which case the covariances in the summation are all zero, giving

Group members unrelated (r = 0)

$$\sigma(z, A_T) = \sigma^2(A_d) + (n-1)\sigma(A_d, A_s) \quad (20.4b)$$

If the direct and associative effects are uncorrelated, this reduces to our standard result that the covariance between an individual's phenotype and breeding value is just the additive genetic variance (in this case, of direct effects). By contrast, the variance of the total breeding value becomes

$$\sigma^2(A_T) = \sigma^2\left[A_d + (n-1)A_s\right]$$
$$= \sigma^2(A_d) + 2(n-1)\sigma(A_d, A_s) + (n-1)^2\sigma^2(A_s) \quad (20.4c)$$
$$= \sigma(z, A_T) + (n-1)\left[2\sigma(A_d, A_s) + (n-1)\sigma^2(A_s)\right] \quad (20.4d)$$

Group members unrelated (r = 0)

Now consider the phenotypic variance,

$$\sigma_z^2 = \sigma^2\left(P_{d_i} + \sum_{j \neq i} P_{s_j}\right). \tag{20.5a}$$

Assuming (for now) that the group members are unrelated, so that $\sigma(P_{d_i}, P_{s_j}) = 0$. For a group of size n Equation 20.5a reduces to

$$\begin{aligned}
\sigma_z^2 &= \sigma^2(P_d) + (n-1)\sigma^2(P_s) && (20.5b)\\
&= \sigma^2(A_d) + (n-1)\sigma^2(A_s) + \sigma^2(E_d) + (n-1)\sigma^2(E_s) && (20.5c)\\
&= \sigma^2(A_d) + (n-1)\sigma^2(A_s) + \sigma^2(e) && (20.5d)
\end{aligned}$$

where e is given by Equation 20.1c. With the phenotypic variance in hand, we can define the heritability of the direct and associative effects as

$$h_d^2 = \frac{\sigma^2(A_d)}{\sigma_z^2}, \qquad \text{and} \qquad h_s^2 = \frac{\sigma^2(A_s)}{\sigma_z^2} \tag{20.6a}$$

Example 20.3. Consider a trait in a group of 10 unrelated individuals, with $\sigma^2_{P_d} = 10$, $\sigma^2_{P_s} = 1$, and both direct and associative effects have modest heritabilities measured on the scale of the effect themselves ($h^{'2}_d = 0.4, h^{'2}_s = 0.3$). To simplify matters, assume $\sigma(A_d, A_s) = 0$. Applying Equation 20.5b, the resulting phenotypic variance is

$$\sigma_z^2 = \sigma^2_{P_d} + 9 \cdot \sigma^2_{P_s} = 10 + 9 \cdot 1 = 19$$

From Equation 20.4c, the variance in total breeding value becomes

$$\sigma^2_{A_T} = \sigma^2_{A_d} + 9^2 \cdot \sigma^2_{A_s} = h^{'2}_d\, \sigma^2_{P_d} + 9^2 \cdot h^{'2}_s\, \sigma^2_{P_s} = 4 + 81 \cdot 0.3 = 28.3,$$

giving T^2 = 28.3 / 18 = 1.57.

A real world example of large potential differences in h_d^2 versus T^2 is survival days in chickens (Bijma et al. 2007b). Ignoring associative effects gives a heritability h_d^2 of 6.7%, while when using a mixed model that incorporates associative effects (detailed later in the chapter), the estimate of T^2 was 20%, a threefold increase. Hence, under the conditions in the study, roughly two-thirds of the heritable variation in the trait arises from interactions between individuals and is thus hidden from standard analyses which ignore these. As discussed below, this component is only fully accessible under individual selection if the group includes relatives.

One of the key results when associative effects are present is that individual selection can result in a reversed response, while group selection always results in a positive response (although it may be far from optimal). These points were clearly made by Griffing (1967) for the simple case of two interacting, and unrelated, individuals within each group. For selection on individual phenotype, the response becomes

$$R = \frac{\overline{i}}{\sigma_z}\left[\sigma^2(A_d) + \sigma(A_d, A_s)\right] \tag{20.11a}$$

A negative covariance between direct and associative effects reduces the efficiency of selection, and if sufficiently negative, gives a reversed response. This loss of efficiency occurs because the only information an individual's phenotype contains about their breeding value for associative effects is that provided by the covariance between direct and associative breeding values (which can be negative). Conversely, if we select based on the mean of a group, we are selecting on both direct and associative effects to improve trait value. For the case of $n = 2$, Griffing obtained the expected response as

$$R = \frac{\overline{i}}{2\sigma(\overline{z})}\left[\sigma^2(A_d) + 2\sigma(A_d, A_s) + \sigma^2(A_s)\right] = \frac{\overline{i}}{2\sigma(\overline{z})}\,\sigma^2(A_T) \tag{20.11b}$$

While group selection always give a non-negative response, if the associative effects are weak, this approach is very inefficient relative to individual selection. For example, in the absence of associative effects, $\sigma^2(\overline{z}) = \sigma^2(z)/2$, and Equation 20.11b reduces to $\overline{i}\,\sigma(A_d)/[\sqrt{2}\,\sigma(z)]$, or $1/\sqrt{2} = 0.701$ of the response under individual selection.

Covariances with related group members

$$\sigma(z, A_T) = r\sigma^2(A_T) + (1 - r)\left[\sigma^2(A_d) + (n-1)\sigma(A_d, A_s)\right]$$

$$\begin{aligned}\sigma^2(z) &= \sigma^2(A_d) + \sigma^2(E_d) + (n-1)\left[\sigma^2(A_s) + \sigma^2(E_s)\right] \\ &\quad + (n-1)r\left[2\sigma(A_s, A_d) + (n-2)\sigma^2(A_d)\right] \\ &= \sigma^2(z \mid r = 0) + (n-1)r\left[2\sigma(A_s, A_d) + (n-2)\sigma^2(A_d)\right]\end{aligned}$$

Group members related (r > 0)

The response to selection is simply the change in the mean total breeding value, which (from Chapter 10) is the within-generation change in the phenotypic mean after selection (the selection differential S) times the slope of the regression of A_T on phenotype z,

$$R = \frac{\sigma(z, A_T)}{\sigma_z^2}\,S = \frac{\sigma(z, A_T)}{\sigma_z}\,\overline{i} \tag{20.14}$$

with the second formulation following from the standard identity that $S = \sigma_z\,\overline{i}$ (Equation 10.6a). For $n = 2$ and $r = 0$, we recover Griffin's result (Equation 20.11a).

Example 20.4. Muir (2005) estimated variance components for six-week body weight in Japanese quail (*Coturnix coturnix japonica*) housed in groups of $n = 16$ per cage. REML estimates of the genetic variances were $\sigma^2(A_d) = 33.7$ and $\sigma^2(A_s) = 2.87$, while $\sigma(A_d, A_s) = -5.5$. Under these values, the predicted response to individual selection in a group of 16 unrelated individuals is

$$R = \frac{\overline{i}}{\sigma_z}\left[\sigma^2(A_d) + (n-1)\sigma(A_d, A_s)\right] = \frac{\overline{i}}{\sigma_z}\left[33.7 + 15 \cdot (-5.5)\right] = -48.8\,\frac{\overline{i}}{\sigma_z}$$

The strong negative covariance between direct and social (competitive) effects results in an expected reversed response if directional selection is used, as the positive gain from improvement of direct effects is swamped by the negative effects from the correlated response in social values.

The presence of relatives within the group results in some fraction of $\sigma^2(A_s)$ being incorporated into the response under individual selection. Suppose the group of 16 consists of two half-sib families. In this case, the average relationship is 0.125, and from Equation 20.12d the resulting covariance between phenotype and total breeding values becomes

$$\begin{aligned}\sigma(z, A_T) &= \sigma(z, A_T \mid r = 0) + (n-1)r\left[\sigma(A_s, A_d) + (n-1)\sigma^2(A_s)\right] \\ &= -48.4 + 15 \cdot 0.125\,(-5.5 + 15 \cdot 2.87) = 21.6\end{aligned}$$

Simply by using groups of relatives (as opposed to groups of unrelated individuals) allows individual selection to give an expected positive response.

Individual selection: Direct vs. Associate response

Here unit of selection u = z, the phenotype of an individual

$$R_z = R_d + (n-1)R_s, \quad \text{where} \quad R_d = \frac{\sigma(A_d, z)}{\sigma_z}\,\overline{i} \quad \text{and} \quad R_s = \frac{\sigma(A_s, z)}{\sigma_z}\,\overline{i} \qquad (20.15a)$$

Here

$$\sigma(A_d, z) = \sigma\left(A_d, A_d + \sum_{i \neq j} A_{s,i} + e\right) = \sigma^2(A_d) + r(n-1)\sigma(A_d, A_s) \qquad (20.15b)$$

while

$$\sigma(A_s, z) = \sigma\left(A_s, A_d + \sum_{i \neq j} A_{s,i} + e\right) = \sigma(A_d, A_s) + r(n-1)\sigma^2(A_s) \qquad (20.15c)$$

Unless (i) A_s, A_d correlated OR (ii) group members are relatives, value of z provides information on A_d, but NOT on its A_s value

Example 20.6. Consider the response in a family of half-sibs from Example 20.5, where the expected total response was 15.39 $\bar{\imath}$. What were the contributions from the direct and social response? For the values used in that example,

$$\sigma(A_d, z) = \sigma^2(A_d) + r(n-1)\sigma(A_d, A_s) = 500 + 0.25 \cdot 5 \cdot (-39.5) = 450.63$$

and

$$\sigma(A_s, z) = \sigma(A_d, A_s) + r(n-1)\sigma^2(A_s) = -39.5 + 0.24 \cdot 5 \cdot 50 = 23.0$$

Recalling (for half-sibs) that $\sigma_z^2 = 1350.6$, Equation 20.15 gives the two components of response as

$$R_d = \frac{450.63}{\sqrt{1350.6}}\,\bar{\imath} = 12.26\,\bar{\imath}, \quad \text{and} \quad R_s = \frac{23}{\sqrt{1350.6}}\,\bar{\imath} = 0.63\,\bar{\imath}$$

Hence, 80% (12.26/ 15.39) of the total response was due to response in direct effects, while 20% was from the response in social effects (5·0.63 / 15.39). Under individual selection on half-sib families, both the mean direct and mean social values improved. By contrast, if group members are unrelated, then (Example 20.5) $\sigma_z^2 = 1150$, while

$$\sigma(A_d, z) = \sigma^2(A_d) = 500, \quad \sigma(A_s, z) = \sigma(A_d, A_s) = -39.5$$

giving responses of

$$R_d = \frac{500}{\sqrt{1150}}\,\bar{\imath} = 14.744\,\bar{\imath}, \quad \text{and} \quad R_s = \frac{-39.5}{\sqrt{1150}}\,\bar{\imath} = -1.165\,\bar{\imath}$$

While the total response in this case was positive, the large direct response (14.74) was significantly offset by a decrease in the mean social environment ($5 \cdot [-1.16] = -5.83$), giving the total response as (14.74 - 5.82) $\bar{\imath}$ = 8.92$\bar{\imath}$. The lack of relatedness implies no direct selection involving $\sigma^2(A_s)$, and hence the social breeding values only change through their correlation with the direct values, which in this example was negative.

Maternal effects

$$z_i = P_{d_i} + P_{m,j} \tag{20.16a}$$

In the absence of inbreeding, $r = 1/2$ for this group (mother-offspring) with $n = 2$. From Equation 20.12c, the covariance between phenotype and total breeding value ($A_T = A_d + A_m$),

$$\sigma(z, A_T) = \sigma^2(A_d) + (3/2)\sigma(A_d, A_m) + (1/2)\sigma^2(A_m), \tag{20.16b}$$

while Equation 20.13a gives the phenotypic variance as

$$\sigma^2(z) = \sigma^2(A_d) + \sigma(A_d, A_s) + \sigma^2(A_s) + \sigma_e^2 \tag{20.16c}$$

giving the resulting response to selection as

total $$R = \frac{\sigma(z, A_T)}{\sigma_z}\bar{\imath} = \frac{\sigma^2(A_d) + (3/2)\sigma(A_d, A_m) + (1/2)\sigma^2(A_m)}{\sqrt{\sigma^2(A_d) + \sigma(A_d, A_s) + \sigma^2(A_s) + \sigma_e^2}}\bar{\imath} \tag{20.16d}$$

$$R_d = \frac{\sigma(A_d, z)}{\sigma_z}\bar{\imath} = \frac{\sigma^2(A_d) + (1/2)\sigma(A_d, A_m)}{\sigma_z}\bar{\imath}$$ Direct response

$$R_m = \frac{\sigma(A_m, z)}{\sigma_z}\bar{\imath} = \frac{\sigma(A_d, A_m) + (1/2)\sigma^2_{A_m}}{\sigma_z}\bar{\imath}$$ Maternal response

Group selection

Unit of selection u = group mean

$$\sigma\left(A_{T_i}, \sum_{j=1}^{n} z_j\right) = \sigma\left(A_{T_i}, \sum_{j=1}^{n}(A_{T_j} + e_j)\right) = \sum_{j=1}^{n}\sigma\left(A_{T_i}, A_{T_j}\right) = \sigma^2(A_T)\sum_{j=1} r_{ij}$$
$$= \sigma^2(A_T)\left(1 + \sum_{j\neq i} r_{ij}\right) \tag{20.19a}$$

If group members are unrelated, then

$$\sigma\left(A_{T_i}, \sum_{j=1}^{n} z_j\right) = \sigma^2(A_T) \tag{20.19b}$$

which implies $\sigma(A_{T_i}, \bar{z}) = \sigma^2(A_T)/n$. Hence, group selection acts on the total breeding value of an individual, rather than on only part of this as is the case with individual selection (e.g., Equation 20.12e). The associative effects contribution to the total breeding value does not influence the phenotype of the focal individual, but does influence the phenotype of other group members. Group selection directly targets these effects. If all members have the same degree of relationship r,

$$\sigma\left(A_{T_i}, \sum_{j=1}^{n} z_j\right) = \sigma^2(A_T)\left[1 + (n-1)r\right] \tag{20.19c}$$

Key: group mean always correlated with A_T

Group selection -- role of relatives

$$\sigma\left(A_{T_i}, \overline{z}\right) = \frac{1}{n}\sigma^2(A_T)\left[1 + (n-1)r\right] = \sigma^2(A_T)\left(r + \frac{1-r}{n}\right)$$

Group of size n, with r = average relatedness among group members

Note that zbar directly correlated with A_T. Correlation increases if members are related (r > 0)

Response under group selection

$$R = \frac{\sigma(A_T, \overline{z})}{\sigma^2(\overline{z})} S = \frac{\sigma^2(A_T) r_n}{\sigma^2(A_T) r_n + \sigma_e^2 \rho_n} S$$
$$= \frac{\sigma(A_T, \overline{z})}{\sigma(\overline{z})} \overline{\imath} = \frac{\sigma^2(A_T) r_n}{\sqrt{\sigma^2(A_T) r_n + \sigma_e^2 \rho_n}} \overline{\imath}$$

$$r_n = r + \frac{1-r}{n} \quad \textbf{and} \quad \rho_n = \rho + \frac{1-\rho}{n}$$

r = genetic correlation

ρ = environmental correlation among group members

Example 20.7. Consider group selection using Muir's quail data from Example 20.4. Here $\sigma^2(A_d) = 33.7, \sigma^2(A_s) = 2.87, \sigma(A_d, A_s) = -5.5, n = 16$. Muir estimated the residual variance as $\sigma_e^2 = 69.0$, while Muir's model assumed $\rho = 0$, giving $\rho_n = 1/n$ and hence $\sigma_e^2 \rho_n = 69.0/16 = 4.32$. Applying Equation 20.4 gives the total additive variance as

$$\begin{aligned}\sigma^2(A_T) &= \sigma^2(A_d) + 2(n-1)\sigma(A_d, A_s) + (n-1)^2\sigma^2(A_s) \\ &= 33.7 + 30 \cdot (-5.5) + 30^2 \cdot 2.87 = 2451.7,\end{aligned}$$

while Equation 20.26b gives the response as

$$R = \frac{\sigma^2(A_T) r_n}{\sqrt{\sigma^2(A_T) r_n + \sigma_e^2 \rho_n}} \bar{\imath} = \frac{2451.7 \cdot r_n}{\sqrt{2451.7 \cdot r_n + 4.32}} \bar{\imath}$$

For groups of unrelated individuals, $r = 0$ and $r_n = 0.0625$, and the response becomes R= 12.2 $\bar{\imath}$. For half- and full-sibs, r_n = 0.297 and 0.531, with responses of 26.9 $\bar{\imath}$ and 36.0 $\bar{\imath}$, a two- and three-hold increase relative to a group of unrelated individuals.

Group + kin selection

Unit of selection
u = I is an index

$$I_i = z_i + g \sum_{j \neq i} z_j$$

This index can also be written as

$$I_i = (1-g) z_i + g \sum_{j=1}^{n} z_j = (1-g) z_i + g\, n\, \overline{z},$$

$$R = \frac{\sigma(I, A_T)}{\sigma(I)} \bar{\imath}_I$$

g = group selection
r = kin selection

$$\sigma(A_T, I) = [g + r + (n-2)gr]\, \sigma^2(A_T) + (1-g)(1-r)\left[\sigma^2(A_d) + (n-1)\sigma(A_s, A_d)\right]$$

g & r have symmetric roles

Key: Use group + relatives to maximize Cov(u, A_T)

Consequences: Evolution of fitness

Examining the expected change in mean fitness is straightfoward. Using previous results, we simply take the trait being followed as individual fitness ($z = W$). From Equation 20.1c, the fitness of individual i becomes

$$W_i = \mu + A_{d_i} + \sum_{j \neq i} A_{s_j} + e_i \tag{20.47a}$$

A_d is the direct breeding value of fitness, while A_s is the social breeding value (how a focal individual influences the fitness of others in its group). As above, A_{s_i} does not contribute to W_i, while A_{s_j} for $j \neq i$ does. Likewise, as before the total breeding value for fitness of an individual is simply

$$AT_i = A_{d_i} + (n-1)A_{s_i} \tag{20.47b}$$

with variance

$$\sigma^2(A_T) = \sigma^2(A_d) + 2(n-1)\sigma(A_d, A_s) + (n-1)^2\sigma^2(A_s) \tag{20.47c}$$

The first term is the classical additive genetic variance in fitness in the absence of associative effects. When interactions are present, there is the potential for substantially more heritable variation in fitness. Indeed, the total genetic variance in fitness has the potential to exceed the actual variance in individual fitness ($\sigma^2(A_T) > \sigma^2_W$), as much of the variation is hidden in interactions with others, which do not appear in one's individual fitness.

Mean fitness can decrease when associative effects are strong

Applying Equation 20.12c gives the response in terms of the variance components as

$$R_W = \frac{1}{\overline{W}}\left[\sigma^2(A_d) + (n-1)(1+r)\sigma(A_d, A_s) + r(n-1)^2\sigma^2(A_s)\right] \tag{20.48c}$$

Just as we have seen for other traits, when $r = 0$, the possibility of a reversed response occurs if the breeding value for direct and social effects on fitness are sufficiently negatively correlated. Hence, under rather realistic conditions, individual selection can result in a *decrease* (and a potentially rather significant one at that) in mean fitness.

If the BVs of direct and associative effects on fitness are sufficiently negatively-correlated, can get a reversed response -- fitness goes down

Ironically, even though a negative response can occur in the presence of associative effects, there is actually more total variance potentially available when they are present, as $\sigma^2(A_T) \geq \sigma^2(A_d)$. However, only a fraction of this may be accessible to *individual* selection, and this fraction (being a covariance rather than a variance) can be negative. The key for exploiting the available variance is either selection among groups and/or the presence of relatives in one's group of interacting individuals.

To see this, note from Equation 20.12e that we can express Equation 20.48c as

$$R_W = \frac{1}{\overline{W}}\left(r\sigma^2(A_T) + (1-r)\left[\sigma^2(A_d) + (n-1)\sigma(A_d, A_s)\right]\right) \tag{20.48d}$$

The term in square brackets represents the response in a group of non-relatives. When interactions occur among kin ($r > 0$), then for sufficiently close relatives, the response becomes positive (mean fitness increases) even if it is negative when $r = 0$. At the extreme, when $r = 1$ (all interactions are among clones), the response in mean fitness is simply $\sigma^2(A_T)/\overline{W}$ and all of the heritable variance in fitness is utilized. Conversely, when interactions occur among unrelated individuals, only a fraction of this genetic potential is exploited. This observation lead Bijma (2010a) to suggest that when heritable fitness interactions are present, the key to evolutionary success is interacting with relatives. The reason for this is clear from our previous discussions. With interactions among unrelated individuals, one's phenotype (here fitness) provides very little information about their social breeding value. With interacting kin, the breeding values of the kin's social effects influences your fitness, and these are positively correlated (via kinship) with your own breeding value for social effects.

47

Direct and social effects responses

Finally, we can decompose the total response in fitness into response from changes in the mean of the direct effects and response from changes in the mean of the social effects. Equation 20.15a gives

$$R_W = R_{W,d} + (n-1)R_{W,s} \tag{20.49a}$$

Recalling Equation 20.48a, Equations 20.15b,c give these response components as

$$R_{W,d} = \frac{\sigma^2(A_d) + r(n-1)\sigma(A_d, A_s)}{\overline{W}} \tag{20.49b}$$

and

$$R_{W,s} = \frac{\sigma(A_d, A_s) + r(n-1)\sigma^2(A_s)}{\overline{W}} \tag{20.49c}$$

Altruistic traits: An example of a reversed response

Example 20.16. Haldane (1932) coined the term **altruistic trait** to denote a behavior (or trait) that harms an individual, but benefits others. The classic example are alarm calls — others in a group are warned (increasingly their fitness), but at some expense to the individual making the call (a direct effect decreasing fitness). Note that the increase in an altrustic trait is an example of a reversed response, as the trait lowers the fitness of the individual that bears it. What are the conditions for such traits to spread? In terms of our fitness model with associative effects (Equation 20.47a), we can rephrase this as the conditions for the mean value of A_s to increase, which are given by Equation 20.49c. From the definition of altruism, $\sigma(A_d, A_s) < 0$, as performing an altruistic act decreases your direct fitness while increasing the fitness of those in your group. Equation 20.49c shows that a necessary (but not sufficient!) condition for altruism to evolve under individual selection is $r > 0$, i.e., individuals interact in groups of relatives.

As pointed out by Bijma and Wade (2008), we can view $\sigma(A_d, A_s)$ as the **cost** $(-c)$ for an altruistic act towards others in your group. Conversely, the altruistic contribution to you from others in your group is $(n-1)\sigma^2(A_s) \geq 0$, which we denote as the **benefit** b. With these definitions, from Equation 20.49c the condition for altruism to evolve under individual selection is just

$$-c + rb > 0, \quad \text{or} \quad r > b/c$$

This is the classic **Hamilton's rule** (Hamilton 1963; 1964a,b).

Inclusive Fitness

As Equation 20.47a illustrates, when heritable interactions are present, the fitness of an individual depends on both their own genes as well as the genes in others. Hamilton (1964a,b) suggested that the focus should shift from individual fitness to what he called **inclusive fitness** — that component of fitness influenced only by the alleles carried by the focal individual. Hamilton argued that individuals strive to increase their inclusive, as opposed to individual, fitness (also see Michod and Abugov 1980, Grafen 2006). Formally, the inclusive fitness of an individual is context-specific, and is defined as individual fitness minus any contribution to that fitness from the group environment plus the effect of that individual on the fitness of others, weighted by relatedness. While sounding rather abstract, when placed in an associative effect framework, this definition is quite clear.

From Equation 20.47a, for individual i, A_{d_i} is the heritable component of individual fitness W_i remaining when the social contributions from others have been removed. The focal individual's social breeding value A_{s_i} does not influence their own fitness, but the social effects of other group members do, with the (heritable) contribution to individual i's fitness from individual j being A_{s_j}. The correlation between the breeding value A_{s_i} carried by i and the contribution to i's fitness from j is their relatedness r_{ij}, so that $r_{ij}A_{s_i}$ is the predicted value of A_{s_j} given A_{s_i}. Putting these together gives the heritable component (i.e., breeding value) of i's inclusive fitness as

$$A_{incl,i} = A_{d_i} + A_{s_i}\sum_{j\neq i}^{n} r_{ij} = A_{d_i} + r(n-1)A_{s_i} \tag{20.51a}$$

where the last equality makes our standard assumption that all group members are equally related (which is easily relaxed). The resulting variance in the breeding value for inclusive fitness becomes

$$\sigma^2(A_{incl}) = \sigma^2(A_d) + 2r(n-1)\sigma(A_d, A_d) + r^2(n-1)^2\sigma^2(A_s) \tag{20.51b}$$

In the absence of heritable associative effects ($\sigma^2(A_s) = 0$) this simply reduces to the additive variance in direct fitness. Importantly, note that the heritable component of inclusive fitness is not the same as the total breeding value A_T for fitness, as

$$A_{T_i} = A_{incl,i} + (1-r)(n-1)A_{s_i} \tag{20.51c}$$

Just as Equation 20.49 decomposed the total response into components from direct and associative effects, we can similarly decompose the change in mean individual fitness into change in mean inclusive fitness plus the residual response. From Equation 20.51c,

$$R_W = R_{W,incl} + (1-r)(n-1)R_{W,s} \tag{20.52a}$$

so that total response in fitness is the change in inclusive fitness plus any response in the residual of the mean social value (after the effects of group relatives are absorbed into inclusive fitness). From Equation 20.48b, the response in the mean inclusive fitness is given by

$$R_{W,incl} = \frac{1}{\overline{W}}\sigma(W, A_{incl}) \tag{20.52b}$$

where

$$\begin{aligned}\sigma(W, A_{incl}) &= \sigma\left(\mu + A_{d_i} + \sum_{j\neq i} A_{s_j} + e_i, A_{d_i} + r(n-1)A_{s_i}\right) \\ &= \sigma^2(A_d) + 2r(n-1)\sigma(A_d, A_d) + r^2(n-1)^2\sigma^2(A_s)\end{aligned} \tag{20.52c}$$

The last line follows by evaluating the covariance in a similar fashion as done throughout this chapter. Note by comparison with Equation 20.51b, that this is simply $\sigma^2(A_{incl})$, yielding

$$R_{W,incl} = \frac{\sigma^2(A_{incl})}{\overline{W}} \tag{20.52d}$$

Hence (under our simple model), the response in mean inclusive fitness is proportional to the additive variance in inclusive fitness, so that mean inclusive fitness is non-decreasing. Why, then, can the mean of *individual* fitness decline despite the continual increase in mean inclusive fitness? The reason is an even faster decline in the mean (residual) social value. Recalling Equation 20.49c, Equation 20.51a becomes

$$R_W - R_{W,incl} = \frac{(1-r)(n-1)}{\overline{W}}\left(\sigma(A_d, A_s) + r(n-1)\sigma^2(A_s)\right) \tag{20.52e}$$

Hence, if the covariance between direct and associative effects is sufficiently negative, any increase in inclusive fitness is more than countered by the decline in the mean social environment. Note that increasingly the relatedness of group members decreases the residual response between mean individual and inclusive fitness, which in turn increases the chances that individual mean fitness increases.

Key: mean inclusive fitness (unlike individual fitness) is non-decreasing

Lecture 9
Estimation of Variance Components

Guilherme J. M. Rosa

University of Wisconsin-Madison

Introduction to Quantitative Genetics
SISG, Seattle
19 – 21 July 2017

Estimation of Variance Components
ANOVA Estimation

Consider the data set below, related to observations of half-sib families of k unrelated sires. The following model can be used to represent these data:

Sire			
1	2	…	k
y_{11}	y_{21}	…	y_{k1}
y_{12}	y_{22}	…	y_{k2}
⋮	⋮		⋮
y_{1n_1}	y_{2n_2}	…	y_{kn_k}

$$y_{ij} = \mu + s_i + e_{ij}$$

where y_{ij} represents the phenotypic trait observation of progeny j (j = 1, 2,…, n_i) in family i, μ is a mean, s_i is an effect common to all animals having sire i, and e_{ij} is a residual term

Estimation of Variance Components

ANOVA Estimation

The sire effect s_i is equivalent to the transmitting ability (which is equal to one-half additive genetic value) of sire i, as one-half of its genes are (randomly) transmitted to each of its n_i progeny.

The residual terms e_{ij} refer to additional genetics effects (such as the effect of dams) and environmental components.

It is assumed that $s_i \overset{ind}{\sim} (0, \sigma_s^2)$ and $e_{ij} \overset{ind}{\sim} (0, \sigma_e^2)$

From the model settings discussed before we have that

$$E[y_{ij}] = \mu \quad \text{and} \quad Var[y_{ij}] = \sigma_s^2 + \sigma_e^2$$

The overall sample mean is given by $\bar{y}_{\bullet\bullet} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} = \frac{1}{N}\sum_{i=1}^{k} y_{i\bullet}$

where $N = \sum_{i=1}^{k} n_i$, and $\bar{y}_{i\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$ are sire-specific means.

The ANOVA approach consists of an orthogonal decomposition of the total sum of squares (TSS) into between classes (or, in our case, sires) and within classes (or residual) components. The corrected (in terms of the general mean) TSS is given by: $TSS = \sum_{i=1}^{k}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{\bullet\bullet})^2$

By adding and subtracting $\bar{y}_{i\bullet}$ within the parentheses, the TSS can be expressed as:

$$\begin{aligned} TSS &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}[(y_{ij}-\bar{y}_{i\bullet})+(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})]^2 \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\bullet})^2+\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2+2\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\bullet})(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet}) \end{aligned}$$

It is seen that the last part of this expression is equal to zero, so that TSS can be written as two components:

$$SSS=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2 \quad \text{and} \quad RSS=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\bullet})^2$$

which are the sire and the residual sum of squares, respectively. The SSS term measures the variation of each progeny family around the overall mean, while the RSS term measures the extra variation related to each observation around its sire average

It can be shown that the expectation of these sums of squares terms are:

$$E[SSS]=\left(N-\frac{1}{N}\sum_{i=1}^{n_i}n_i^2\right)\sigma_s^2+(k-1)\sigma_e^2 \quad \text{and} \quad E[RSS]=(N-k)\sigma_e^2$$

so that the ANOVA estimators of the sire and residual variance components are given by:

$$\hat{\sigma}_s^2=\left(N-\frac{1}{N}\sum_{i=1}^{n_i}n_i^2\right)^{-1}[SSS-(k-1)\hat{\sigma}_e^2] \quad \text{and} \quad \hat{\sigma}_e^2=\frac{1}{(N-k)}RSS$$

In the specific case of balanced data, i.e. the same progeny size for all sires, $n_i = n = N/k$ and the ANOVA estimators become:

$$\hat{\sigma}_s^2=\frac{1}{n}\left[\frac{1}{(k-1)}SSS-\hat{\sigma}_e^2\right] \quad \text{and} \quad \hat{\sigma}_e^2=\frac{1}{k(n-1)}RSS$$

Estimation of Variance Components

ANOVA approach works well for simple models (such as a one-way structure) or balanced data (such as data from designed experiments with no missing data), but they are not indicated for more complex models and data structures

Other proposed methods: expected mean squares approach of Henderson (1953), and the minimum norm quadratic unbiased estimation (Rao 1971a, 1971b), among others.

However, maximum likelihood based methods are currently the most popular, especially the restricted (or residual) maximum likelihood (REML) approach, which attempts to correct for the well-known bias in the classical maximum likelihood (ML) estimation of variance components. These two methods are briefly described next.

Estimation of Variance Components
Maximum Likelihood (ML) Estimator

Maximum likelihood estimates of the variance components can be obtained by maximizing the log-likelihood $L(\boldsymbol{\beta}, \mathbf{G}, \boldsymbol{\Sigma})$ with respect to each element of $\mathbf{G}$ and $\boldsymbol{\Sigma}$, after replacing $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$

Alternatively, $\mathbf{G}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\beta}$ can be estimated simultaneously by maximizing their joint log-likelihood with respect to the variance components and the fixed effects.

As a simple example of maximum likelihood estimation of variance components, consider the balanced case (i.e., constant progeny sizes) half-sib families data set discussed previously, and the linear model:

$$y_{ij} = \mu + s_i + e_{ij}$$

with the same definitions as before, but with the additional assumption of normality of both the sire and the residual effects, i.e.:

$$s_i \overset{ind}{\sim} N(0, \sigma_s^2) \quad \text{and} \quad e_{ij} \overset{ind}{\sim} N(0, \sigma_e^2)$$

In matrix notation, this model can be expressed as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_k \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n \\ \mathbf{1}_n \\ \vdots \\ \mathbf{1}_n \end{bmatrix} \mu + \begin{bmatrix} \mathbf{1}_n & \mathbf{0}_n & \cdots & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{1}_n & \cdots & \mathbf{0}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_n & \mathbf{0}_n & \cdots & \mathbf{1}_n \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_k \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_k \end{bmatrix}$$

where $\mathbf{y}_i = [y_{i1}\ y_{i2} \cdots y_{ik}]^T$ represents the vector of observations of progeny i (i.e., relative to sire i); $\mathbf{1}_n$ and $\mathbf{0}_n$ represent n-dimensional column vectors of 1's and 0's, respectively; and $\mathbf{e}_i = [e_{i1}, e_{i2}, \ldots, e_{ik}]^T$ is the vector of residuals associated with progeny i

The vector of observations $\mathbf{y} = [\mathbf{y}_1^T \ \mathbf{y}_2^T \ \dots \ \mathbf{y}_k^T]^T$ has then a multivariate normal distr. with mean vector $\boldsymbol{\mu} = \mathbf{1}_N \mu$ and variance-covariance matrix given by $\mathbf{I}_s \otimes (\mathbf{1}_n \sigma_s^2 \mathbf{1}_n^T) + \mathbf{I}_N \sigma_e^2$, and its density function (from which the likelihood function obtained) can be written as:

$$p(\mathbf{y} \mid \mu, \sigma_s^2, \sigma_e^2) = \frac{1}{(2\pi)^{N/2} \mid \mathbf{I}_s \otimes \mathbf{J}_n \sigma_s^2 + \mathbf{I}_N \sigma_e^2 \mid^{1/2}}$$

$$\times \exp\left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{1}_N \mu)^T (\mathbf{J}_n \sigma_s^2 + \mathbf{I}_n \sigma_e^2)^{-1} (\mathbf{y} - \mathbf{1}_N \mu) \right\}$$

$$= (2\pi)^{-\frac{N}{2}} (\sigma_e^2)^{-\frac{(N-k)}{2}} (\sigma_e^2 + n\sigma_s^2)^{-\frac{k}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{1}_N \mu)^T \left[\mathbf{I}_s \otimes \mathbf{J}_n \left(\frac{1}{n} \left(\frac{1}{\sigma_e^2 + n\sigma_s^2} - \frac{1}{\sigma_e^2} \right) \right) \right] (\mathbf{y} - \mathbf{1}_N \mu) \right\}$$

where $\mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n^T$ is an (n × n) matrix of 1's, and ⊗ is the Kronecker product

The log-likelihood function can be written then as:

$$l(\mu, \sigma_s^2, \sigma_e^2) \propto -\frac{(N-k)}{2} \log(\sigma_e^2) - \frac{k}{2} \log(\sigma_e^2 + n\sigma_s^2) - \frac{1}{2\sigma_e^2} \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i\bullet})^2 - \frac{1}{2} \sum_{i=1}^{k} \frac{n(\bar{y}_{i\bullet} - \mu)^2}{\sigma_e^2 + n\sigma_s^2}$$

By taking the derivatives and setting them to 0, the following solutions are obtained:

$$\hat{\mu} = \bar{y}_{\bullet\bullet} \ , \ \hat{\sigma}_e^2 = \frac{1}{k(n-1)} \mathrm{RSS} \quad \text{and} \quad \hat{\sigma}_s^2 = \frac{1}{n} \left[\frac{\mathrm{SSS}}{k} - \hat{\sigma}_e^2 \right]$$

from which ML estimates of the variance components are obtained, except if $\hat{\sigma}_s^2 < 0$, in which case the estimate is set to zero

ML estimates of variance components are biased downwards as they do not take into account the degrees of freedom used for estimating the fixed effects

Estimation of Variance Components

Residual Maximum Likelihood (REML) Estimator

Restricted (or residual) maximum likelihood approach (REML): corrects the bias associated with ML estimates by taking into account the degrees of freedom used for estimating the fixed effects

REML maximizes the likelihood function of a set of error contrasts $\mathbf{d} = \mathbf{L}^T\mathbf{y}$, where $\mathbf{L}$ is a [n x (n – p)] full-rank matrix with columns orthogonal to the columns of the incidence matrix $\mathbf{X}$

The vector $\mathbf{d}$ follows a multivariate normal distribution with null mean vector and variance-covariance matrix $\mathbf{L}^T\mathbf{V}\mathbf{L} = \mathbf{L}^T(\mathbf{Z}\mathbf{G}\mathbf{Z}^T + \boldsymbol{\Sigma})\mathbf{L}$. Note that the distribution of $\mathbf{d}$ does not depend on $\boldsymbol{\beta}$.

The residual likelihood function for the variance components is then:

$$L(\mathbf{G},\boldsymbol{\Sigma}\,|\,\mathbf{y}) = (2\pi)^{-(n-p)/2}\,|\,\mathbf{L}^T\mathbf{V}\mathbf{L}\,|^{-1/2}\,\exp\left\{-\frac{1}{2}\mathbf{d}^T(\mathbf{L}^T\mathbf{V}\mathbf{L})^{-1}\mathbf{d}\right\}$$

Another approach for obtaining the residual likelihood function for the variance components is by integrating the fixed effects out of the 'full' likelihood function, i.e.:

$$L(\mathbf{G},\boldsymbol{\Sigma}\,|\,\mathbf{y}) = \int L(\boldsymbol{\beta},\mathbf{G},\boldsymbol{\Sigma}\,|\,\mathbf{y})\,d\boldsymbol{\beta}$$

as illustrated in the following example.

Recall the balanced half-sib families data set, and its associated likelihood function:

$$L(\mu,\sigma_s^2,\sigma_e^2) = (2\pi)^{-\frac{N}{2}}(\sigma_e^2)^{-\frac{(N-k)}{2}}(\sigma_e^2+n\sigma_s^2)^{-\frac{k}{2}} \times \exp\left\{-\frac{1}{2\sigma_e^2}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2 - \frac{1}{2}\sum_{i=1}^{k}\frac{n(\bar{y}_{i\bullet}-\mu)^2}{\sigma_e^2+n\sigma_s^2}\right\}$$

Its residual likelihood is then:

$$L(\sigma_s^2,\sigma_e^2) = \int L(\mu,\sigma_s^2,\sigma_e^2)d\mu$$
$$= (2\pi)^{-\frac{N}{2}}(\sigma_e^2)^{-\frac{(N-k)}{2}}(\sigma_e^2+n\sigma_s^2)^{-\frac{k}{2}} \times \exp\left\{-\frac{1}{2\sigma_e^2}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2\right\}\int \exp\left\{-\frac{1}{2}\sum_{i=1}^{k}\frac{n(\bar{y}_{i\bullet}-\mu)^2}{\sigma_e^2+n\sigma_s^2}\right\}d\mu$$

which is equal to:

$$L(\sigma_s^2,\sigma_e^2) = (2\pi)^{-\frac{N}{2}}(\sigma_e^2)^{-\frac{(N-k)}{2}}\lambda^{-\frac{k}{2}} \times \exp\left\{-\frac{1}{2\sigma_e^2}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2\right\}\exp\left\{-\frac{n}{2\lambda}\sum_{i=1}^{k}(\bar{y}_{i\bullet}-\mu)^2\right\}\sqrt{2\pi\frac{\lambda}{kn}}$$

where $\lambda = \sigma_e^2 + n\sigma_s^2$.

By taking the derivatives with respect to λ and σ_e^2, and by using the invariance property of maximum likelihood estimators, the following solutions are obtained:

$$\hat{\sigma}_e^2 = \frac{1}{k(n-1)} RSS \quad \text{and} \quad \hat{\sigma}_s^2 = \frac{1}{n}\left[\frac{1}{(k-1)} SSS - \hat{\sigma}_e^2\right]$$

which are the REML estimates of the variance components, except if $\hat{\sigma}_s^2 < 0$, i.e. if

$$SSS < \frac{(k-1)}{k(n-1)} RSS$$

Explicit forms of ML and REML estimators are often not available for more complex mixed effects models

ML and REML estimates are then generally obtained by iterative approaches such as the expectation-maximization (EM) algorithm and Newton-Raphson-based procedures

Bayesian MCMC Methods

Table A.1 Continuous distributions

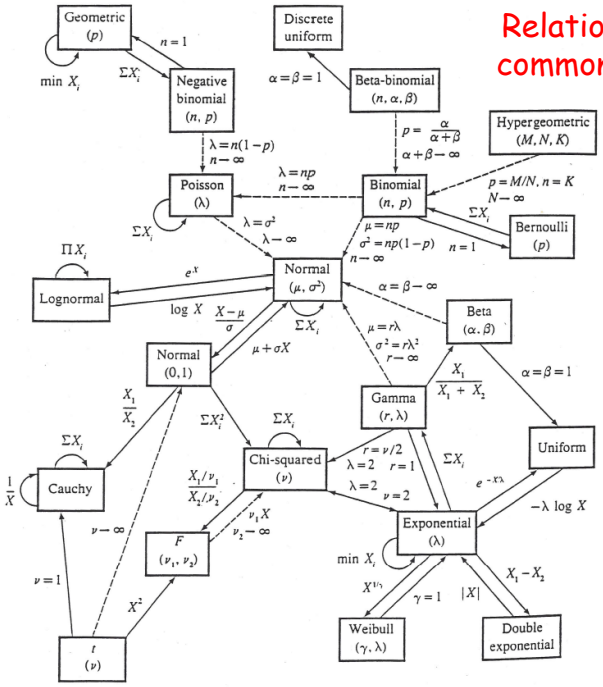
Distribution	Notation	Parameters	Density function	Mean, variance, and mode
Uniform	$\theta \sim \text{U}(a,b)$ $p(\theta) = \text{U}(\theta\|a,b)$	boundaries a,b with $b > a$	$p(\theta) = \frac{1}{b-a},\ \theta \in [a,b]$	$\text{E}(\theta) = \frac{a+b}{2}$, $\text{var}(\theta) = \frac{(b-a)^2}{12}$ no mode
Normal	$\theta \sim \text{N}(\mu,\sigma^2)$ $p(\theta) = \text{N}(\theta\|\mu,\sigma^2)$	location μ scale $\sigma > 0$	$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(\theta-\mu)^2\right)$	$\text{E}(\theta) = \mu$, $\text{var}(\theta) = \sigma^2$ $\text{mode}(\theta) = \mu$
Multivariate normal	$\theta \sim \text{N}(\mu,\Sigma)$ $p(\theta) = \text{N}(\theta\|\mu,\Sigma)$ (implicit dimension d)	symmetric, pos. def., $d \times d$ cov. matrix Σ	$p(\theta) = (2\pi)^{-d/2}\|\Sigma\|^{-1/2}$ $\times \exp\left(-\frac{1}{2}(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)\right)$	$\text{E}(\theta) = \mu$, $\text{var}(\theta) = \Sigma$ $\text{mode}(\theta) = \mu$
Gamma	$\theta \sim \text{Gamma}(\alpha,\beta)$ $p(\theta) = \text{Gamma}(\theta\|\alpha,\beta)$	shape $\alpha > 0$ inverse scale $\beta > 0$	$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta},\ \theta > 0$	$\text{E}(\theta) = \frac{\alpha}{\beta}$ $\text{var}(\theta) = \frac{\alpha}{\beta^2}$ $\text{mode}(\theta) = \frac{\alpha-1}{\beta}$, for $\alpha \geq 1$
Inverse-gamma	$\theta \sim \text{Inv-gamma}(\alpha,\beta)$ $p(\theta) = \text{Inv-gamma}(\theta\|\alpha,\beta)$	shape $\alpha > 0$ scale $\beta > 0$	$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{-(\alpha+1)}e^{-\beta/\theta},\ \theta > 0$	$\text{E}(\theta) = \frac{\beta}{\alpha-1}$, for $\alpha > 1$ $\text{var}(\theta) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \alpha > 2$ $\text{mode}(\theta) = \frac{\beta}{\alpha+1}$
Chi-square	$\theta \sim \chi^2_\nu$ $p(\theta) = \chi^2_\nu(\theta)$	deg. of freedom $\nu > 0$	$p(\theta) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)}\theta^{\nu/2-1}e^{-\theta/2},\ \theta > 0$ same as $\text{Gamma}(\alpha = \frac{\nu}{2}, \beta = \frac{1}{2})$	$\text{E}(\theta) = \nu$, $\text{var}(\theta) = 2\nu$ $\text{mode}(\theta) = \nu - 2$, for $\nu \geq 2$
Inverse-chi-square	$\theta \sim \text{Inv-}\chi^2_\nu$ $p(\theta) = \text{Inv-}\chi^2_\nu(\theta)$	deg. of freedom $\nu > 0$	$p(\theta) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)}\theta^{-(\nu/2+1)}e^{-1/(2\theta)},\ \theta > 0$ same as $\text{Inv-gamma}(\alpha = \frac{\nu}{2}, \beta = \frac{1}{2})$	$\text{E}(\theta) = \frac{1}{\nu-2}$, for $\nu > 2$ $\text{var}(\theta) = \frac{2}{(\nu-2)^2(\nu-4)}, \nu > 4$ $\text{mode}(\theta) = \frac{1}{\nu+2}$
Scaled inverse-chi-square	$\theta \sim \text{Inv-}\chi^2(\nu,s^2)$ $p(\theta) = \text{Inv-}\chi^2(\theta\|\nu,s^2)$	deg. of freedom $\nu > 0$ scale $s > 0$	$p(\theta) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)}s^\nu\theta^{-(\nu/2+1)}e^{-\nu s^2/(2\theta)},\ \theta > 0$ same as $\text{Inv-gamma}(\alpha = \frac{\nu}{2}, \beta = \frac{\nu}{2}s^2)$	$\text{E}(\theta) = \frac{\nu}{\nu-2}s^2$ $\text{var}(\theta) = \frac{2\nu^2}{(\nu-2)^2(\nu-4)}s^4$ $\text{mode}(\theta) = \frac{\nu}{\nu+2}s^2$
Exponential	$\theta \sim \text{Expon}(\beta)$ $p(\theta) = \text{Expon}(\theta\|\beta)$	inverse scale $\beta > 0$	$p(\theta) = \beta e^{-\beta\theta},\ \theta > 0$ same as $\text{Gamma}(\alpha = 1, \beta)$	$\text{E}(\theta) = \frac{1}{\beta}$, $\text{var}(\theta) = \frac{1}{\beta^2}$ $\text{mode}(\theta) = 0$
Wishart	$W \sim \text{Wishart}_\nu(S)$ $p(W) = \text{Wishart}_\nu(W\|S)$ (implicit dimension $k \times k$)	deg. of freedom ν symmetric, pos. def. $k \times k$ scale matrix S	$p(W) = \left(2^{\nu k/2}\pi^{k(k-1)/4}\prod_{i=1}^k \Gamma\left(\frac{\nu+1-i}{2}\right)\right)^{-1}$ $\times\|S\|^{-\nu/2}\|W\|^{(\nu-k-1)/2}$ $\times \exp\left(-\frac{1}{2}\text{tr}(S^{-1}W)\right)$, W pos. def.	$\text{E}(W) = \nu S$
Inverse-Wishart	$W \sim \text{Inv-Wishart}_\nu(S^{-1})$ $p(W) = \text{Inv-Wishart}_\nu(W\|S^{-1})$ (implicit dimension $k \times k$)	deg. of freedom ν symmetric, pos. def. $k \times k$ scale matrix S	$p(W) = \left(2^{\nu k/2}\pi^{k(k-1)/4}\prod_{i=1}^k \Gamma\left(\frac{\nu+1-i}{2}\right)\right)^{-1}$ $\times\|S\|^{\nu/2}\|W\|^{-(\nu+k+1)/2}$ $\times \exp\left(-\frac{1}{2}\text{tr}(SW^{-1})\right)$, W pos. def.	$\text{E}(W) = (\nu-k-1)^{-1}S$

Table A.1 Continuous distributions *continued*

Distribution	Notation	Parameters	Density function	Mean, variance, and mode
Student-t	$\theta \sim t_\nu(\mu, \sigma^2)$ $p(\theta) = t_\nu(\theta\|\mu, \sigma^2)$ t_ν is short for $t_\nu(0,1)$	deg. of freedom $\nu > 0$ location μ scale $\sigma > 0$	$p(\theta) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi}\sigma}(1 + \frac{1}{\nu}(\frac{\theta-\mu}{\sigma})^2)^{-(\nu+1)/2}$	$\mathrm{E}(\theta) = \mu$, for $\nu > 1$ $\mathrm{var}(\theta) = \frac{\nu}{\nu-2}\sigma^2$, for $\nu > 2$ $\mathrm{mode}(\theta) = \mu$
Multivariate Student-t	$\theta \sim t_\nu(\mu, \Sigma)$ $p(\theta) = t_\nu(\theta\|\mu, \Sigma)$ (implicit dimension d)	deg. of freedom $\nu > 0$ location $\mu = (\mu_1, \ldots, \mu_d)$ symmetric, pos. def. $d \times d$ scale matrix Σ	$p(\theta) = \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}}\|\Sigma\|^{-1/2}$ $\times(1 + \frac{1}{\nu}(\theta-\mu)^T\Sigma^{-1}(\theta-\mu))^{-(\nu+d)/2}$	$\mathrm{E}(\theta) = \mu$, for $\nu > 1$ $\mathrm{var}(\theta) = \frac{\nu}{\nu-2}\Sigma$, for $\nu > 2$ $\mathrm{mode}(\theta) = \mu$
Beta	$\theta \sim \mathrm{Beta}(\alpha, \beta)$ $p(\theta) = \mathrm{Beta}(\theta\|\alpha, \beta)$	'prior sample sizes' $\alpha > 0, \beta > 0$	$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$ $\theta \in [0,1]$	$\mathrm{E}(\theta) = \frac{\alpha}{\alpha+\beta}$ $\mathrm{var}(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ $\mathrm{mode}(\theta) = \frac{\alpha-1}{\alpha+\beta-2}$
Dirichlet	$\theta \sim \mathrm{Dirichlet}(\alpha_1, \ldots, \alpha_k)$ $p(\theta) = \mathrm{Dirichlet}(\theta\|\alpha_1, \ldots, \alpha_k)$	'prior sample sizes' $\alpha_j > 0; \alpha_0 \equiv \sum_{j=1}^k \alpha_j$	$p(\theta) = \frac{\Gamma(\alpha_1+\cdots+\alpha_k)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)}\theta_1^{\alpha_1-1}\ldots\theta_k^{\alpha_k-1}$ $\theta_1, \ldots, \theta_k \geq 0; \sum_{j=1}^k \theta_j = 1$	$\mathrm{E}(\theta_j) = \frac{\alpha_j}{\alpha_0}$ $\mathrm{var}(\theta_j) = \frac{\alpha_j(\alpha_0-\alpha_j)}{\alpha_0^2(\alpha_0+1)}$ $\mathrm{cov}(\theta_i, \theta_j) = -\frac{\alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)}$ $\mathrm{mode}(\theta_j) = \frac{\alpha_j-1}{\alpha_0-k}$

Table A.2 Discrete distributions

Distribution	Notation	Parameters	Density function	Mean, variance, and mode
Poisson	$\theta \sim \mathrm{Poisson}(\lambda)$ $p(\theta) = \mathrm{Poisson}(\theta\|\lambda)$	'rate' $\lambda > 0$	$p(\theta) = \frac{1}{\theta!}\lambda^\theta \exp(-\lambda)$ $\theta = 0, 1, 2, \ldots$	$\mathrm{E}(\theta) = \lambda$, $\mathrm{var}(\theta) = \lambda$ $\mathrm{mode}(\theta) = \lfloor\lambda\rfloor$
Binomial	$\theta \sim \mathrm{Bin}(n, p)$ $p(\theta) = \mathrm{Bin}(\theta\|n, p)$	'sample size' n (pos. integer) 'probability' $p \in [0,1]$	$p(\theta) = \binom{n}{\theta}p^\theta(1-p)^{n-\theta}$ $\theta = 0, 1, 2, \ldots, n$	$\mathrm{E}(\theta) = np$ $\mathrm{var}(\theta) = np(1-p)$ $\mathrm{mode}(\theta) = \lfloor(n+1)p\rfloor$
Multinomial	$\theta \sim \mathrm{Multin}(n; p_1, \ldots, p_k)$ $p(\theta) = \mathrm{Multin}(\theta\|n; p_1, \ldots, p_k)$	'sample size' n (pos. integer) 'probabilities' $p_j \in [0,1]$: $\sum_{j=1}^k p_j = 1$	$p(\theta) = \binom{n}{\theta_1\ \theta_2\cdots\theta_k}p_1^{\theta_1}\cdots p_k^{\theta_k}$ $\theta_j = 0, 1, 2, \ldots, n; \sum_{j=1}^k \theta_j = n$	$\mathrm{E}(\theta_j) = np_j$ $\mathrm{var}(\theta_j) = np_j(1-p_j)$ $\mathrm{cov}(\theta_i, \theta_j) = -np_ip_j$
Negative binomial	$\theta \sim \mathrm{Neg\text{-}bin}(\alpha, \beta)$ $p(\theta) = \mathrm{Neg\text{-}bin}(\theta\|\alpha, \beta)$	shape $\alpha > 0$ inverse scale $\beta > 0$	$p(\theta) = \binom{\theta+\alpha-1}{\alpha-1}\left(\frac{\beta}{\beta+1}\right)^\alpha\left(\frac{1}{\beta+1}\right)^\theta$ $\theta = 0, 1, 2, \ldots$	$\mathrm{E}(\theta) = \frac{\alpha}{\beta}$ $\mathrm{var}(\theta) = \frac{\alpha}{\beta^2}(\beta+1)$
Beta-binomial	$\theta \sim \mathrm{Beta\text{-}bin}(n, \alpha, \beta)$ $p(\theta) = \mathrm{Beta\text{-}bin}(\theta\|n, \alpha, \beta)$	'sample size' n (pos. integer) 'prior sample sizes' $\alpha > 0, \beta > 0$	$p(\theta) = \frac{\Gamma(n+1)}{\Gamma(\theta+1)\Gamma(n-\theta+1)}\frac{\Gamma(\alpha+\theta)\Gamma(n+b-\theta)}{\Gamma(a+b+n)}$ $\times\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}, \quad \theta = 0, 1, 2, \ldots, n$	$\mathrm{E}(\theta) = n\frac{\alpha}{\alpha+\beta}$ $\mathrm{var}(\theta) = n\frac{\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$



Relationships among common distributions

Solid lines: transformations and special cases

Dashed lines: limits

(Leemis, 1986)

Bayesian Inference

y: observed data; $y \sim p(y|\theta)$

θ: parameters (all unobserved quantities)

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

$$p(\theta \mid y) \propto p(\theta)p(y \mid \theta)$$

posterior distribution

prior distribution

sampling distribution

Multi Parameter Models

$$\mathbf{y} \sim p(\mathbf{y} \mid \theta_1, \theta_2, \ldots, \theta_p)$$

$$p(\theta_1, \theta_2, \ldots, \theta_p \mid \mathbf{y}) \propto p(\theta_1, \theta_2, \ldots, \theta_p)p(\mathbf{y} \mid \theta_1, \theta_2, \ldots, \theta_p)$$

Marginal Posterior Distributions

$$p(\theta_1 \mid \mathbf{y}) \propto \int_{\theta \neq \theta_1} p(\theta_1, \theta_2, \ldots, \theta_p \mid \mathbf{y}) d\theta_{\theta \neq \theta_1}$$

Example: Normal Distribution

Data: $y_1, y_2, \ldots, y_n \overset{iid}{\sim} N(\mu, \sigma^2)$

Sampling model: $p(\mathbf{y} \mid \mu, \sigma^2) = \prod_{i=1}^{n} p(y_i \mid \mu, \sigma^2)$

$$\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right\}$$

Prior (Jeffreys'): $p(\mu, \sigma^2) = (\sigma^2)^{-1}$

Joint posterior:

$$p(\mu, \sigma^2 \mid \mathbf{y}) \propto (\sigma^2)^{-(n+2)/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right\}$$

$$= (\sigma^2)^{-(n+2)/2} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y} - \mu)^2\right]\right\}$$

Example: Normal Distribution

Marginal posterior of μ:

$$p(\mu \mid \mathbf{y}) = \int_0^\infty p(\mu, \sigma^2 \mid \mathbf{y}) d\sigma^2$$

$$\propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2}\right]^{-\frac{n}{2}} \sim t_{n-1}(\bar{y}, s^2/n)$$

Marginal posterior of σ^2:

$$p(\sigma^2 \mid \mathbf{y}) = \int_{-\infty}^{\infty} p(\mu, \sigma^2 \mid y) d\mu$$

$$\propto (\sigma^2)^{-(n+2)/2} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\} \sim \text{Inv} - \chi^2(n-1, s^2)$$

Linear Mixed Models

Data: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$, with $\mathbf{u} \mid \sigma_u^2 \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2)$

Sampling model: $p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{u}, \sigma_\varepsilon^2) \propto (\sigma_\varepsilon^2)^{-n/2}$

$$\times \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})\right\}$$

Prior distribution: $p(\boldsymbol{\beta}, \sigma_u^2, \sigma_\varepsilon^2) = p(\boldsymbol{\beta})p(\sigma_u^2)p(\sigma_\varepsilon^2)$

(Note: independence was assumed a priori)

Joint posterior distribution:

$$p(\boldsymbol{\beta}, \mathbf{u}, \sigma_u^2, \sigma_\varepsilon^2 \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{u}, \sigma_\varepsilon^2)p(\mathbf{u} \mid \sigma_u^2)p(\boldsymbol{\beta}, \sigma_u^2, \sigma_\varepsilon^2)$$

$$\propto (\sigma_\varepsilon^2)^{-n/2}(\sigma_u^2)^{-q/2}\, p(\boldsymbol{\beta}, \sigma_u^2, \sigma_\varepsilon^2)$$

$$\times \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_\varepsilon^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \frac{1}{\sigma_u^2}\mathbf{u}^T\mathbf{A}^{-1}\mathbf{u}\right]\right\}$$

Marginal Posterior Distributions

Fixed effects vector:

$$p(\boldsymbol{\beta} \mid \mathbf{y}) = \int_{\mathbf{u}} \int_{\sigma_u^2} \int_{\sigma_\varepsilon^2} p(\boldsymbol{\beta}, \mathbf{u}, \sigma_u^2, \sigma_\varepsilon^2 \mid \mathbf{y}) d\sigma_\varepsilon^2 d\sigma_u^2 d\mathbf{u}$$

Note that integrating over a vector (e.g., vector **u**) implies integrating over each element in that vector, i.e.

$$p(\boldsymbol{\beta} \mid \mathbf{y}) = \int_{u_1} \int_{u_2} \cdots \int_{u_q} \int_{\sigma_u^2} \int_{\sigma_\varepsilon^2} p(\boldsymbol{\beta}, \mathbf{u}, \sigma_u^2, \sigma_\varepsilon^2 \mid \mathbf{y}) d\sigma_\varepsilon^2 d\sigma_u^2 du_q \ldots du_2 du_1$$

Single element of $\boldsymbol{\beta}$ (e.g. β_1):

$$p(\beta_1 \mid \mathbf{y}) = \int_{\beta_2} \int_{\beta_3} \cdots \int_{\beta_p} p(\boldsymbol{\beta} \mid \mathbf{y}) d\beta_p \ldots d\beta_3 d\beta_2$$

Marginal Posterior Distributions

Random effects vector:

$$p(\mathbf{u} \mid \mathbf{y}) = \int_{\boldsymbol{\beta}} \int_{\sigma_u^2} \int_{\sigma_\varepsilon^2} p(\boldsymbol{\beta}, \mathbf{u}, \sigma_u^2, \sigma_\varepsilon^2 \mid \mathbf{y}) d\sigma_\varepsilon^2 d\sigma_u^2 d\boldsymbol{\beta}$$

Variance components:

$$p(\sigma_u^2 \mid \mathbf{y}) = \int_{\boldsymbol{\beta}} \int_{\mathbf{u}} \int_{\sigma_\varepsilon^2} p(\boldsymbol{\beta}, \mathbf{u}, \sigma_u^2, \sigma_\varepsilon^2 \mid \mathbf{y}) d\sigma_\varepsilon^2 d\mathbf{u} d\boldsymbol{\beta}$$

$$p(\sigma_\varepsilon^2 \mid \mathbf{y}) = \int_{\boldsymbol{\beta}} \int_{\mathbf{u}} \int_{\sigma_u^2} p(\boldsymbol{\beta}, \mathbf{u}, \sigma_u^2, \sigma_\varepsilon^2 \mid \mathbf{y}) d\sigma_u^2 d\mathbf{u} d\boldsymbol{\beta}$$

Marginal Posterior Distributions

Marginalization (i.e. integrals) in multi-dimensional models can be cumbersome and some times do not have analytical form

An alternative in this regard: Monte Carlo methods

Monte Carlo integration consists of sampling from the posterior distribution, and then using such sampled values to calculate features of interest on the (joint or marginal) posterior distribution

There are many algorithms that can be used to sample from a distribution; some are based on Markov chains, among which the Gibbs sampling is probably the most popular

Monte Carlo Methods

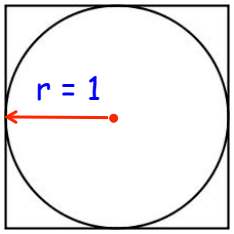
Any method which solves a problem by generating a series of random numbers and counting the incidences that obey specific property(ies)

The method is useful for obtaining numerical solutions to problems which are too complicated to solve analytically

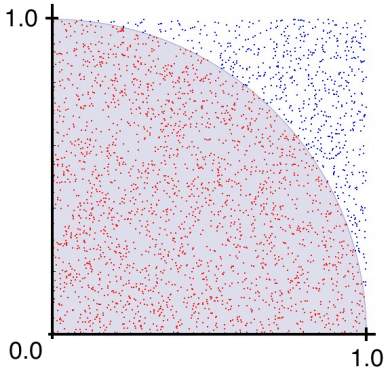
The most common application of the Monte Carlo method is Monte Carlo integration

Monte Carlo Methods

Example: approximating the number π using a circle inscribed in a square



Area of circle = πr^2
Area of square = $4 r^2$



$x^2 + y^2 = r^2$

Monte Carlo Methods

Example: approximating the number π using a circle inscribed in a square

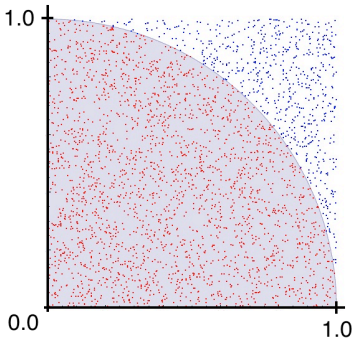
Sample x from Uniform(0,1)

Sample y from Uniform(0,1)

Check if point (x,y) is within the circle, i.e. $y^2 < 1 - x^2$

Repeat the process N times and count how many points (m) fall within the circle

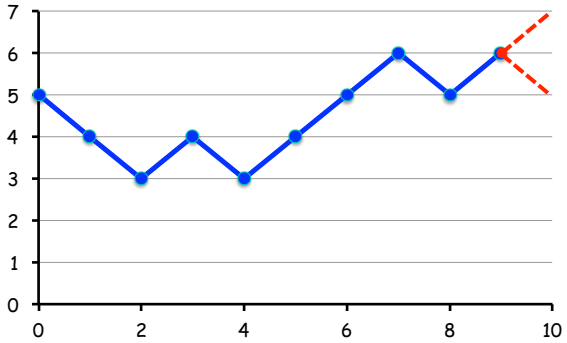
The ratio 4 x m/N is a Monte Carlo approximation for π



$x^2 + y^2 = r^2$

Markov Process

Markov process is a stochastic process that satisfies the Markov property (the memoryless property), i.e., predictions for the future of the process can be made based solely on its present state



Markov Process

Example: Suppose that weather on any given day can be classified into two states: sunny (S) or rainy (R)

Suppose also that, based on past experience, we know that:

Pr(Next day is S | Given today is R) = 0.50 and

Pr(Next day is S | Given today is S) = 0.90

Then, a transition matrix representing the probabilities of the weather moving from one state to another state can be expressed as:

$$P = \begin{array}{c} \begin{matrix} S & R \end{matrix} \\ \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} \end{array} \begin{matrix} S \\ R \end{matrix}$$

Markov Process

If the weather is sunny today (time 0), what is the chance that it will be sunny tomorrow (time 1) as well?

$$Pr(S_1 \mid S_0) = 0.90$$

What about two days from today?

$$Pr(S_2 \mid S_0) = Pr(S_2 \mid S_1) \times Pr(S_1 \mid S_0) + Pr(S_2 \mid R_1) \times Pr(R_1 \mid S_0)$$

$$= 0.9 \times 0.9 + 0.1 \times 0.5 = 0.86$$

Using the same approach to forecast weather on n-th day will approach the following 'equilibrium' probabilities as n increases:

$$Pr(S_n) = 0.833 \text{ and } Pr(R_n) = 0.167$$

Gibbs Sampling

$$\theta = (\theta_1, \theta_2, \ldots, \theta_r) \rightarrow p(\theta_i \mid \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_r)$$

$$\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_r^{(0)})$$

$$\theta_1^{(1)} \mid \theta_2^{(0)}, \theta_3^{(0)}, \ldots, \theta_r^{(0)}$$

$$\theta_2^{(1)} \mid \theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_r^{(0)}$$

$$\vdots$$

$$\theta_r^{(1)} \mid \theta_2^{(1)}, \theta_3^{(1)}, \ldots, \theta_{r-1}^{(1)}$$

Burn-in & Convergence

Tinning interval & Lag correlations

Sample size & Monte Carlo error

Monte Carlo Approximations

After convergence, each sampled vector is a sample from the joint posterior distribution, and so each sampled element (scalar) is a sample from the respective marginal posterior distribution

For each parameter (e.g., θ_1) we'll have then a series of values:

$$\theta_1^{(1)}, \theta_1^{(2)}, \theta_1^{(3)}, \ldots, \theta_1^{(N)}$$

from which features of its distribution (e.g., posterior mean) can be approximated, for example:

$$E[\theta_1 \mid y] \cong \frac{1}{N} \sum_{j=1}^{N} \theta_1^{(j)}$$

Monte Carlo Approximations

Other often interesting features used to represent a marginal posterior distribution are: posterior variance (or standard deviation), posterior mode or median, percentiles, highest posterior density (HPD), etc.

Very useful property: If one is interested on the distribution of a function of the model parameters, samples from such a distribution can be obtained simply by applying that specific function to the sampled values of those parameters

For example, the posterior mean of the heritability can be obtained as:

$$E[h^2 \mid \mathbf{y}] \cong \frac{1}{N}\sum_{j=1}^{N} \frac{\sigma_u^{(j)}}{\sigma_u^{(j)} + \sigma_\varepsilon^{(j)}}$$

Example: Linear Model

Data: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad \mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$

Sampling model: $p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma_e^2) \propto (\sigma_e^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$

Prior distribution: $p(\boldsymbol{\beta}, \sigma_e^2) = p(\boldsymbol{\beta})p(\sigma_e^2) \propto (\sigma_e^2)^{-1}$

Joint posterior distribution:

$$p(\boldsymbol{\beta}, \sigma_e^2 \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma_e^2)p(\boldsymbol{\beta}, \sigma_e^2)$$

$$\propto (\sigma_e^2)^{-(n+2)/2} \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$$

Example: Linear Model

Conditional distribution of location parameters:

$$p(\boldsymbol{\beta} \mid \sigma_e^2, \mathbf{y}) \propto \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}$$

Recall $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ and note that $\mathbf{y}-\mathbf{X}\boldsymbol{\beta} = \mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{X}\hat{\boldsymbol{\beta}}-\mathbf{X}\boldsymbol{\beta}$

such that:

$$(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = \left[(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})-(\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\hat{\boldsymbol{\beta}})\right]^T\left[(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})-(\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\hat{\boldsymbol{\beta}})\right]$$

$$= \Big[\underbrace{(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})}_{\text{independent of } \boldsymbol{\beta}} - \underbrace{2(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\hat{\boldsymbol{\beta}})}_{\text{equal to zero}} + (\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\hat{\boldsymbol{\beta}})\Big]$$

Example: Linear Model

Conditional distribution of location parameters:

Hence: $$p(\boldsymbol{\beta} \mid \sigma_e^2, \mathbf{y}) \propto \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\hat{\boldsymbol{\beta}})\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma_e^2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T\mathbf{X}^T\mathbf{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right\}$$

and so: $$\boldsymbol{\beta} \mid \sigma_e^2, \mathbf{y} \sim N\left(\hat{\boldsymbol{\beta}},\ (\mathbf{X}^T\mathbf{X})^{-1}\sigma_e^2\right)$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$

Example: Linear Model

Conditional distribution of residual variance:

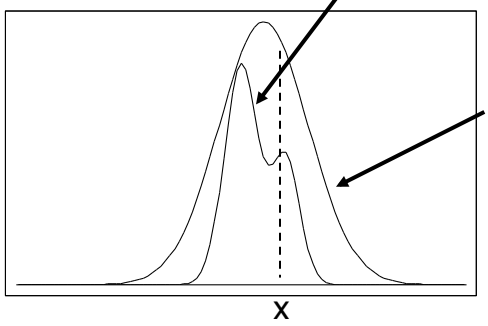
$$p(\sigma_e^2 \mid \boldsymbol{\beta}, \mathbf{y}) \propto (\sigma_e^2)^{-(n+2)/2} \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}$$

Hence: $\sigma_e^2 \mid \boldsymbol{\beta}, \mathbf{y} \sim \text{Inv-gamma}\left(\frac{n}{2}, \frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right)$

Bayes linear regression

Rejection Sampling

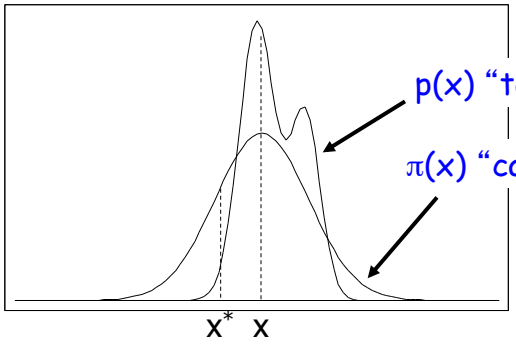
$K f(x) \geq p(x), \ \forall x$ p(x) "target distribution"



① Sample x from f(x)

② Decision: Probability of accepting x: $\alpha = \frac{p(x)}{Kf(x)}$

Metropolis-Hastings Algorithm



① x: current value; sample x^* from $\pi(x)$, e.g. $\pi(x) \sim N(x, \tau^2)$

② The chain moves from x to x^* with probability:

$$\alpha = \min\left[1, \frac{p(x^*)\pi(x)}{p(x)\pi(x^*)}\right]$$

Otherwise the chain remains at the current value