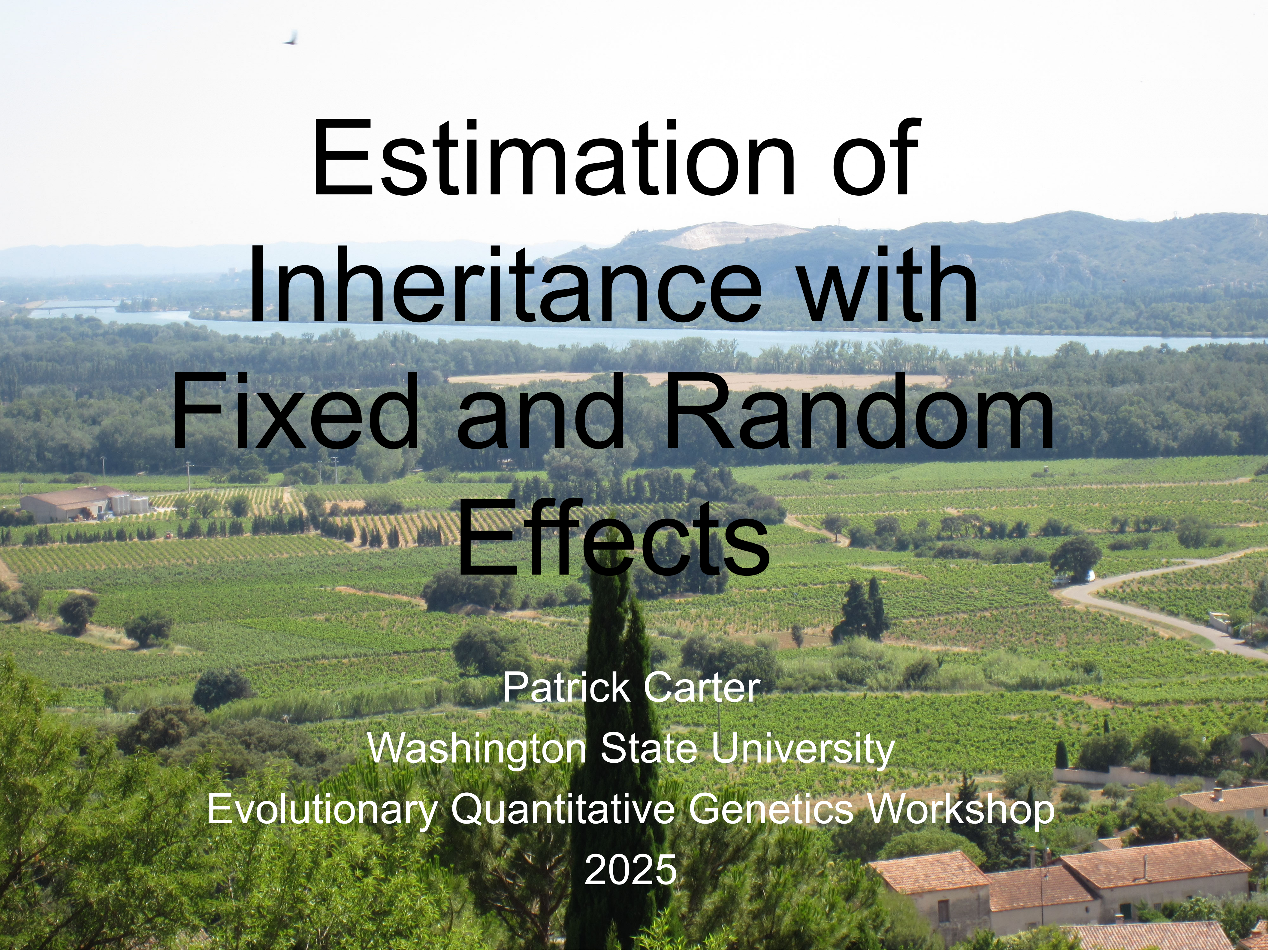


Estimation of Inheritance with Fixed and Random Effects

Patrick Carter
Washington State University
Evolutionary Quantitative Genetics Workshop
2025

Estimating Variance Components

- Many ways to do it
  - ANOVA
  - Regression (developed for estimating h^2!)
- For large unbalanced data sets with complex pedigrees: Linear Mixed Models
- Fixed
  - Experimental/population specific effects
- Random
  - Variables randomly sampled, e.g., individuals randomly sampled from a population

The “Animal Model”

- $\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{e}$
  - $\mathbf{y}$ = (nx1) vector of phenotypes (n = sample size)
  - $\boldsymbol{\beta}$ = (px1) vector of fixed effect regression coefficients (p = number of fixed effects) (FIXED)
  - $\mathbf{X}$ = (nxp) design matrix relating $\mathbf{y}$ to $\boldsymbol{\beta}$ (FIXED)
  - $\mathbf{a}$ = (qx1) vector of additive effects (q = number of individuals in the pedigree) (RANDOM)
  - $\mathbf{Z}$ = (nxq) design matrix relating $\mathbf{y}$ to $\mathbf{a}$ (RANDOM)
  - $\mathbf{e}$ = (nx1) vector of errors
- Solve for $\boldsymbol{\beta}$ and $\mathbf{a}$
- Assume
  - $Var(\mathbf{e}) = \mathbf{I}\sigma_e^2$ (errors are independent)
  - $Var(\mathbf{a}) = \mathbf{A}\sigma_a^2$ (var of $\mathbf{a}$ depends on relationship matrix $\mathbf{A}$)
  - $cov(\mathbf{a},\mathbf{e}) = 0$
  - $\alpha = \sigma_e^2/\sigma_a^2$ is known (!!!!) (need a starting point)

Henderson's Mixed Model Equation

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+A^{-1}\alpha \end{bmatrix} \begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

y = (nx1) vector of phenotypic measures = KNOWN

β = (px1) vector of fixed effect regression coefficients (FIXED) = UNKNOWN

X = (nxp) design matrix relating **y** to **β** (FIXED) = KNOWN

a = (qx1) vector of additive effects (q = number of individuals in the pedigree) (RANDOM) = UNKNOWN

Z = (nxq) design matrix relating **y** to **a** (RANDOM) = KNOWN

A relationship matrix = KNOWN = (q x q) matrix where q = number of individuals in the pedigree

$\alpha = \sigma_e^2 / \sigma_a^2$ = ESTIMATED

Henderson's Mixed Model Equation

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix} \begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

so

$$\begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Ex: Data and System of Linear Equations

id	Sire	Dam	Herd	Ptype
1	-	-	1	78
2	-	-	2	83
3	-	-	2	70
4	2	1	1	86
5	2	3	2	77

$78 = herd_1 + animal_1 + error_1$

$83 = herd_2 + animal_2 + error_2$

$70 = herd_2 + animal_3 + error_3$

$86 = herd_1 + animal_4 + error_4$

$77 = herd_2 + animal_5 + error_5$

Ex: Data and Knowns

id	Sire	Dam	Herd	Ptype
1	-	-	1	78
2	-	-	2	83
3	-	-	2	70
4	2	1	1	86
5	2	3	2	77

$$y = \begin{bmatrix} 78 \\ 83 \\ 70 \\ 86 \\ 77 \end{bmatrix} \quad X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Note: $Z = I$

Ex: Data and Unknowns

id	Sire	Dam	Herd	Ptype
1	-	-	1	78
2	-	-	2	83
3	-	-	2	70
4	2	1	1	86
5	2	3	2	77

$$a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} \qquad \beta = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$

Assume $h^2 = .33$, so $\alpha = \sigma_e^2 / \sigma_a^2 = 2.0$

Example: Data and “A” Matrix

id	Sire	Dam	Herd	Ptype
1	-	-	1	78
2	-	-	2	83
3	-	-	2	70
4	2	1	1	86
5	2	3	2	77

$$
A = \begin{array}{c|ccccc}
 & 1 & 2 & 3 & 4 & 5 \\
\hline
1 & 1 & & & & \\
2 & 0 & 1 & & & \\
3 & 0 & 0 & 1 & & \\
4 & \frac{1}{2} & \frac{1}{2} & 0 & 1 & \\
5 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & 1
\end{array}
$$

Henderson's Mixed Model Equation

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+A^{-1}\alpha \end{bmatrix} \begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

so

$$\begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+A^{-1}\alpha \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Henderson's Mixed Model Equation

X'X =

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Henderson's Mixed Model Equation

$$\begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Henderson's Mixed Model Equation

$X'Z =$

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

Henderson’s Mixed Model Equation

$$\begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+A^{-1}\alpha \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Henderson's Mixed Model Equation

$Z'X = IX =$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Henderson's Mixed Model Equation

$$\begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+A^{-1}\alpha \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Henderson's Mixed Model Equation

$Z'Z + A^{-1}\alpha \;=$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} + 2.0 * \begin{bmatrix} 1.5 & 0.5 & 0 & -1 & 0 \\ 0.5 & 2 & 0.5 & -1 & -1 \\ 0 & 0.5 & 1.5 & 0 & -1 \\ -1 & -1 & 0 & 2 & 0 \\ 0 & -1 & -1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 1 & 0 & -2 & 0 \\ 1 & 5 & 1 & -2 & -2 \\ 0 & 1 & 4 & 0 & -2 \\ -2 & -2 & 0 & 5 & 0 \\ 0 & -2 & -2 & 0 & 5 \end{bmatrix}$$

Henderson's Mixed Model Equation

$$\begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+A^{-1}\alpha \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Henderson's Mixed Model Equation

$X'y =$

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 78 \\ 83 \\ 70 \\ 86 \\ 77 \end{bmatrix} = \begin{bmatrix} 164 \\ 230 \end{bmatrix}$$

Henderson's Mixed Model Equation

$$\begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Henderson's Mixed Model Equation

$$Z'y = Iy = \begin{bmatrix} 78 \\ 83 \\ 70 \\ 86 \\ 77 \end{bmatrix}$$

Ex: Solving the Mixed Model Equation

$$\begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

$$\begin{bmatrix} h_1 \\ h_2 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 4 & 1 & 0 & -2 & 0 \\ 1 & 0 & 1 & 5 & 1 & -2 & -2 \\ 0 & 1 & 0 & 1 & 4 & 0 & -2 \\ 0 & 1 & -2 & -2 & 0 & 5 & 0 \\ 0 & 1 & 0 & -2 & -2 & 0 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 164 \\ 230 \\ 78 \\ 83 \\ 70 \\ 86 \\ 77 \end{bmatrix}$$

Ex: Solutions

$h_1 = 79.17$

$h_2 = 77.77$

$a_1 = 0.53$

$a_2 = 2.13$

$a_3 = -2.66$

$a_4 = 2.71$

$a_5 = -0.37$

Additional Effects

- More fixed effects: Just increases the length of Beta = very easy
- More random effects: each one adds a new term consisting of a vector of random effects times a design matrix
  - Example: add maternal environmental effects

“Animal Model” with Maternal Effects

- **y** = **Xβ** + **Za** + **Wm** + **e**
  y = (nx1) vector of phenotypic measures
  β = (px1) vector of fixed effect regression coefficients (FIXED)
  X = (nxp) design matrix relating **y** to **β** (FIXED)
  a = (qx1) vector of additive effects (q = number of individuals in the pedigree) (RANDOM)
  Z = (nxq) design matrix relating **y** to **a** (RANDOM)
  m = (fx1) vector of maternal effects (f = number of dams in pedigree) (RANDOM)
  W = (nxf) design matrix relating **y** to **m** (RANDOM)
  e = (nx1) vector of errors
- Solve for **β** and **a** and **m**
- Assume
  Var(**e**) = $\mathbf{I}\sigma_e^2$ (errors are independent)
  Var(**a**) = $\mathbf{A}\sigma_a^2$ (var of **a** depends on relationship matrix **A**)
  Var(**m**) = $\mathbf{I}\,\sigma_m^2$ (maternal effects are independent)
  cov(**a**,**m**) and cov (**e**,**m**) are both 0 (**a**, **m**, **e** independent of each other)
  σ_e^2/σ_a^2 and σ_e^2/σ_m^2 are known (need starting points)

Henderson's Mixed Model Equation

$$\begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + A^{-1}\alpha_a & Z'W \\ W'X & W'Z & W'W + I\alpha_m \end{bmatrix} = \begin{bmatrix} \beta \\ a \\ m \end{bmatrix} \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix}$$

y = (nx1) vector of phenotypic measures = KNOWN

β = (px1) vector of fixed effect regression coefficients (FIXED) = UNKNOWN

X = (nxp) design matrix relating **y** to **β** (FIXED) = KNOWN

a = (qx1) vector of additive effects (q = number of individuals in the pedigree) (RANDOM) = UNKNOWN

Z = (nxq) design matrix relating **y** to **a** (RANDOM) = KNOWN

A relationship matrix = KNOWN

m = (fx1) vector of maternal effects (f = number of dams in pedigree) (RANDOM)

W = (nxf) design matrix relating **y** to **m** (RANDOM)

I = identity matrix

$\alpha_a = \sigma_e^2/\sigma_a^2$ = ESTIMATED; $\alpha_m = \sigma_e^2/\sigma_m^2$ = ESTIMATED

Likelihood Approach

- Likelihood = $\Pr(y|\mu, \sigma^2)$ = conditional probability of the data (y) given the parameters:
  $\mu = \mathbf{X\beta}$
  $\sigma^2 = \mathbf{Z(A\sigma_a^2)Z' + I\sigma_e^2}$
  genetic variance + error variance
- Iterations:
  $(\sigma_e^2/\sigma_a^2)_{sv} \rightarrow \mathbf{a_i} \rightarrow (\sigma_e^2/\sigma_a^2)_i$
  $(\sigma_e^2/\sigma_a^2)_i \rightarrow \mathbf{a_j} \rightarrow (\sigma_e^2/\sigma_a^2)_j$
  $(\sigma_e^2/\sigma_a^2)_j \rightarrow \mathbf{a_k} \rightarrow (\sigma_e^2/\sigma_a^2)_k$
- Rinse & Repeat until parameters stabilize

Bayesian Approach

- $\Pr(\mu, \sigma^2|y)$ = conditional probability of the parameters given the data
  - Yay! More intuitive than Likelihood
- Except that it is proportional to:

$\Pr(y|\mu, \sigma^2)\Pr(\mu, \sigma^2)$

(Likelihood)(Prior belief in possible parameter values)

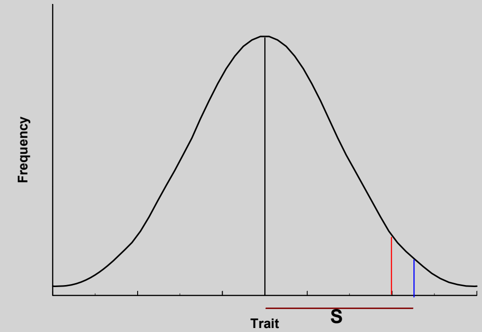
- Produces a posterior distribution = probabilistic distribution associating each value of a parameter to a probability
- Need iterative process to solve: MCMC

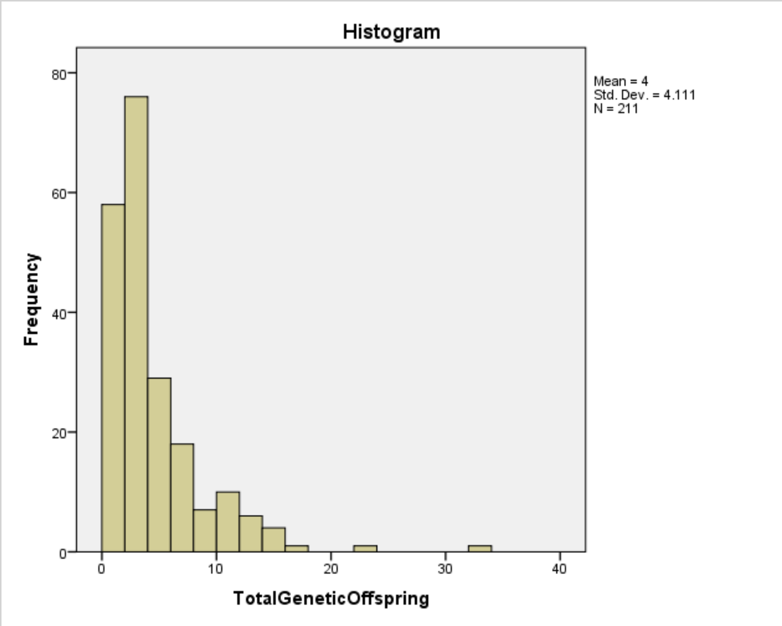
Markov Chain Monte Carlo

- Algorithm based on the proposal of a new value for a parameter, as a function of the value of the other parameters, at each iterative step.
- Saving the value of the parameter at each iteration, we ultimately get a series of values, which is the posterior distribution of interest, i.e., a posterior probability distribution for each parameter.

Non-Gaussian Traits

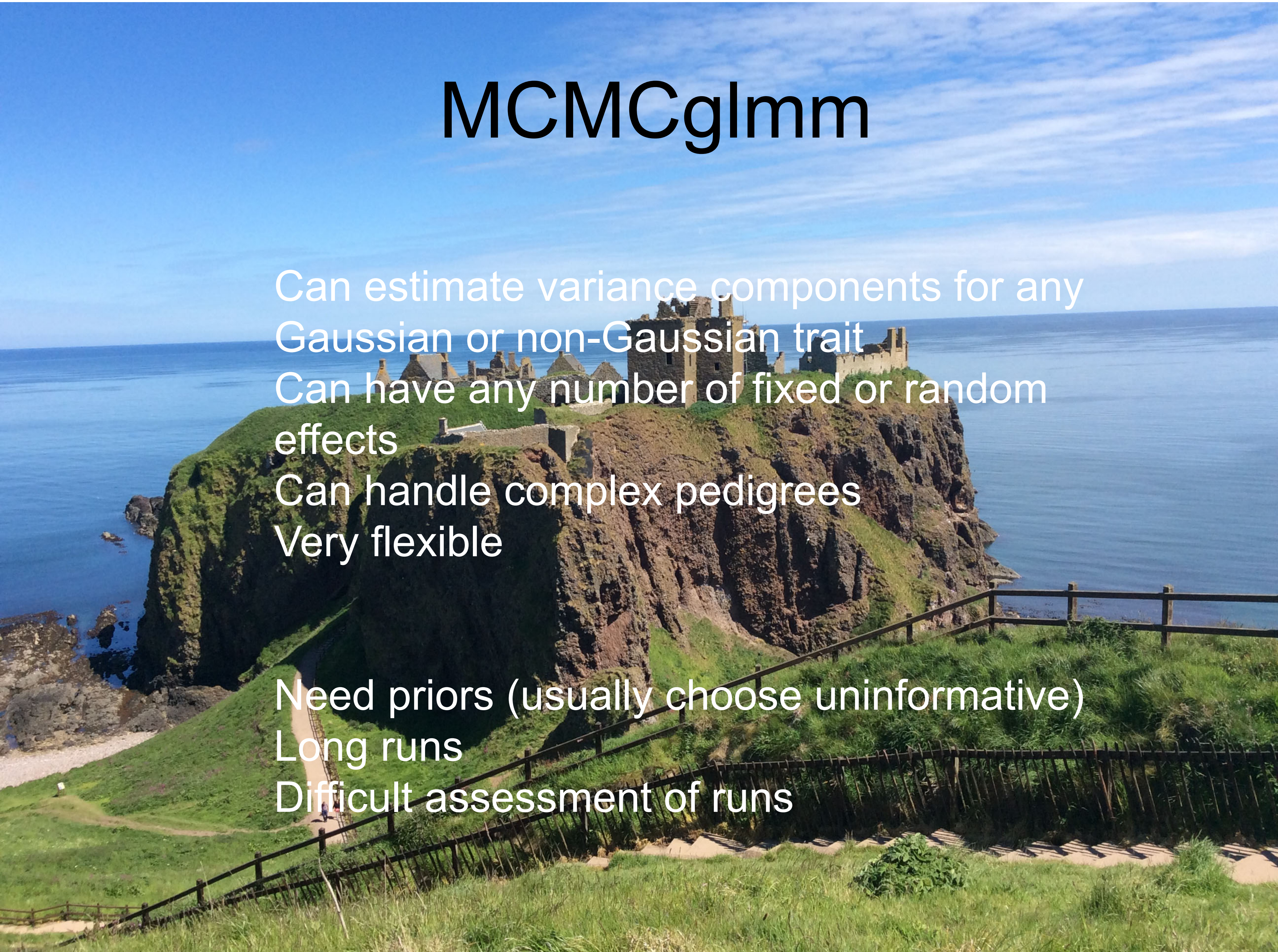
- For a continuous trait, if family relationships in the population are known, and assuming many loci of small effect, the phenotypic variance can be partitioned:
  - $V_P = V_G + V_E$
- But what if the trait of interest is not continuous or is non-Gaussian? How can we estimate variance components?





Generalized Linear Mixed Model (GLMM)

- Combination of Generalized Linear Models and Linear Mixed Models
- Can handle non-Gaussian traits
  - Requires use of a link function (e.g., probability density function for binary outcomes) otherwise still…
- An “Animal Model”
    - $\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{e}$
- Requires Bayesian approaches



MCMCglmm

Can estimate variance components for any Gaussian or non-Gaussian trait
Can have any number of fixed or random effects
Can handle complex pedigrees
Very flexible

Need priors (usually choose uninformative)
Long runs
Difficult assessment of runs