

Course roadmap

AlphaSimR

Day 1: Simulation of breeding programmes **BASICS**

Day 2: ... Quantitative genetics

Day 3: ... Estimation with linear mixed models

Day 4: ... Spatial variation & GxE interactions

Day 5: ... Ancestral recombination graphs

Day 3 agenda – Estimation with linear mixed models

- 09:00-10:30 Introduction to genetic evaluation (simple & pedigree)
- 10:30-11:00 Refreshments break
- 11:00-12:30 Practicals
- 12:30-13:30 Lunch break
- 13:30-15:00 ... using genomic data
- 15:00-15:30 Refreshments break
- 15:30-17:00 Practicals

Preliminaries

- Focus on concepts
- Will make large jumps at a time
- Ask questions at any time

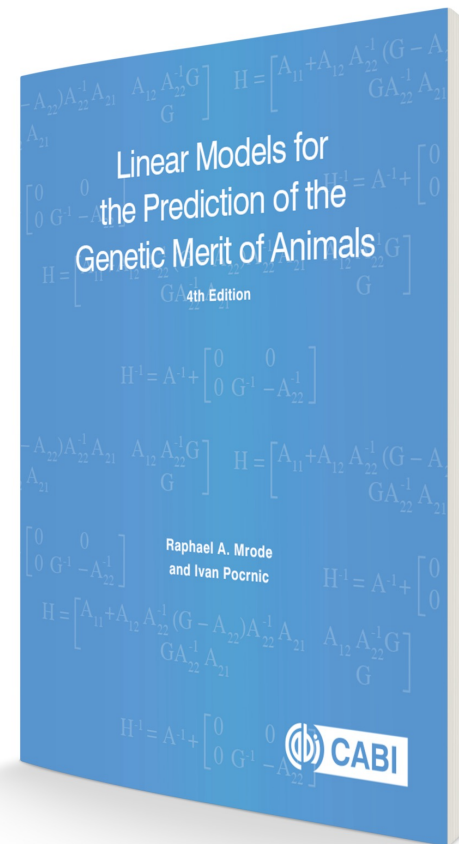
Linear Models for the Prediction of the Genetic Merit of Animals

CABI Biotechnology Series

September 2023 | 412pp

Raphael A Mrode
Ivan Pocrnic

Robin Thompson
Gregor Gorjanc



See chapter 2!



THE UNIVERSITY
of EDINBURGH



Introduction to genetic evaluation (intuition & simple cases)

Gregor Gorjanc, Chris Gaynor, Jon Bancic, Daniel Tolhurst

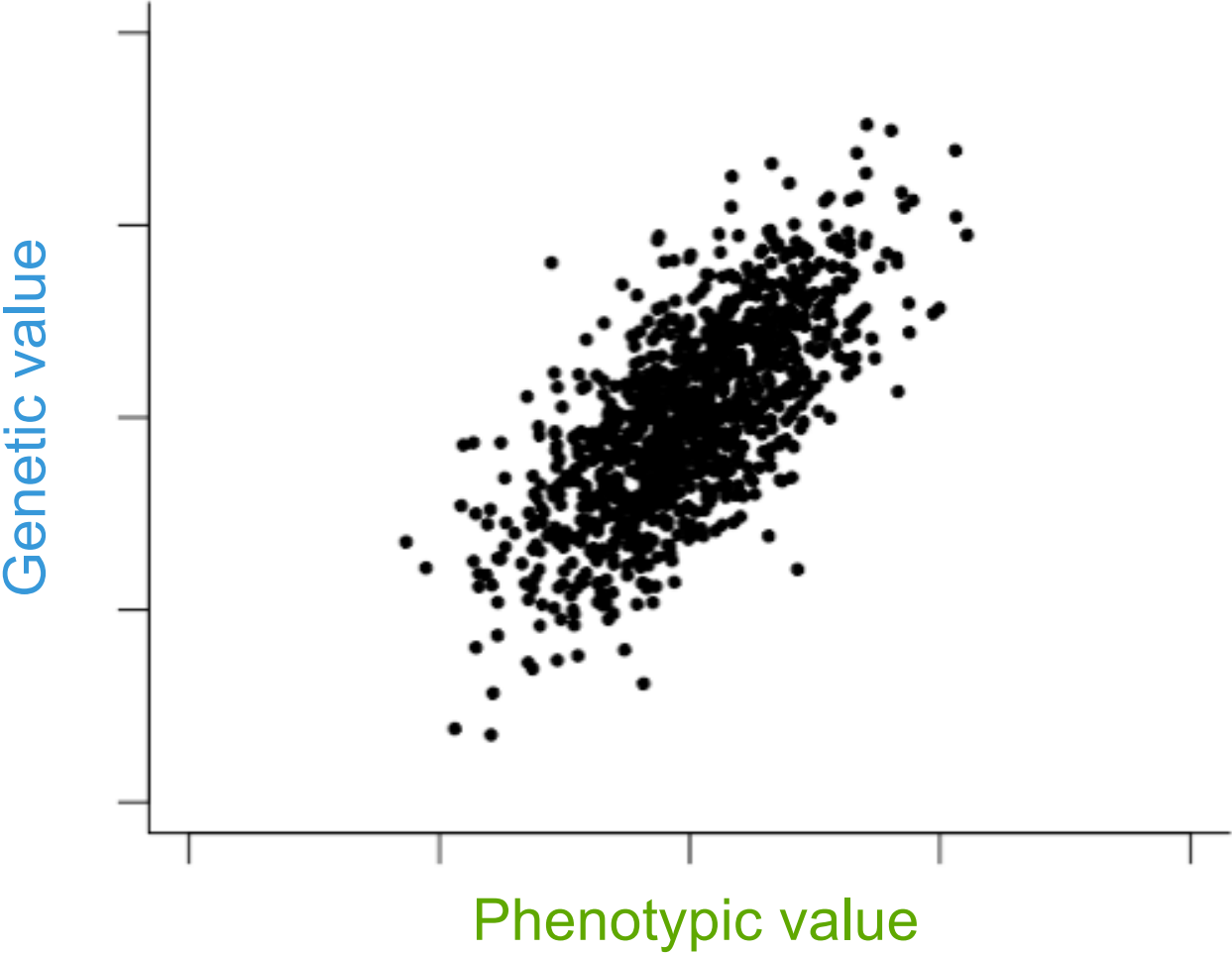
UNE, Armidale
2024-02-07



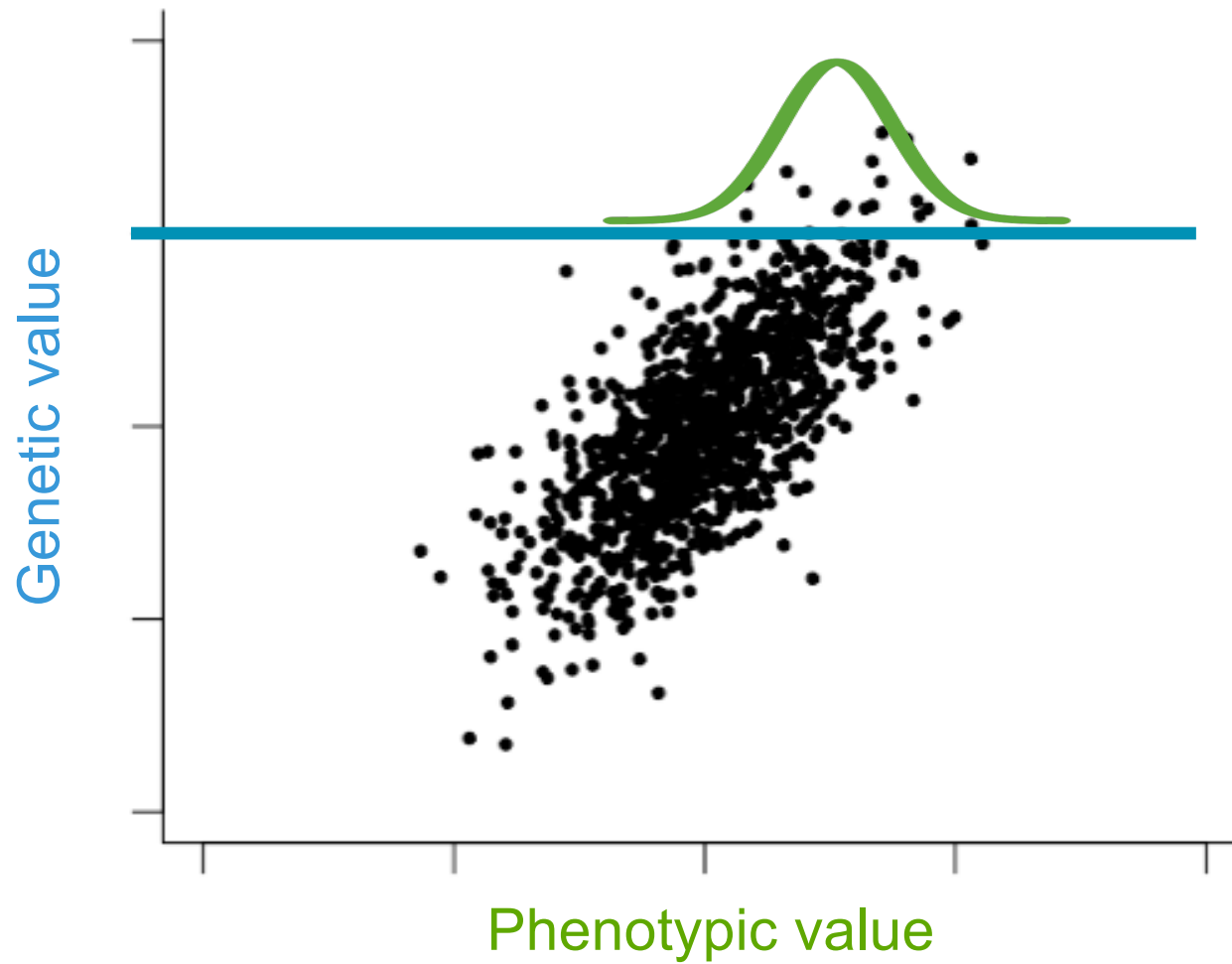
Learning objectives

- Understand forward and backward/inverse views of working with models
- Familiarise yourself with simple cases of inferring breeding values from the data
- Practice inference of breeding values for simple cases

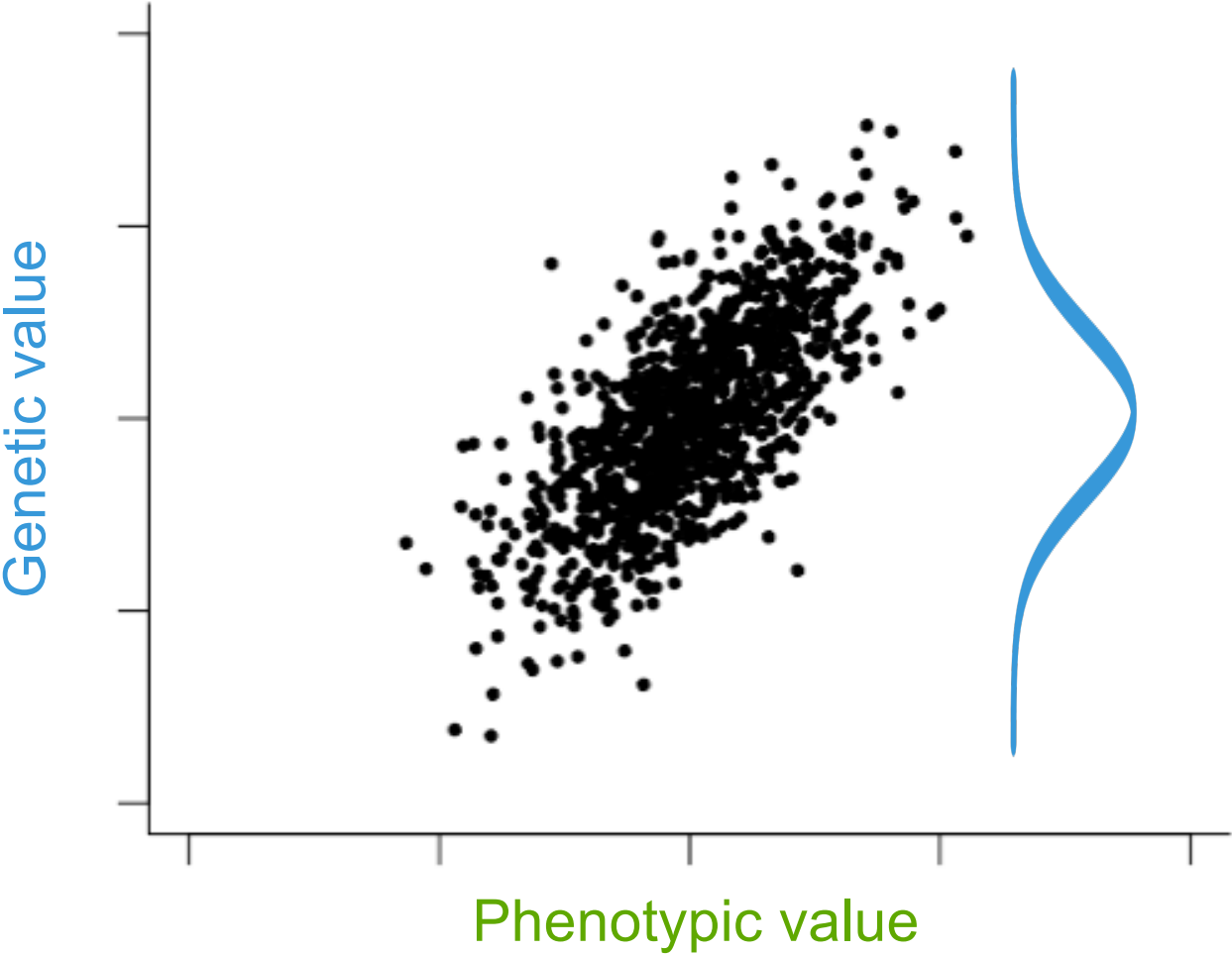
Phenotypic and genetic values



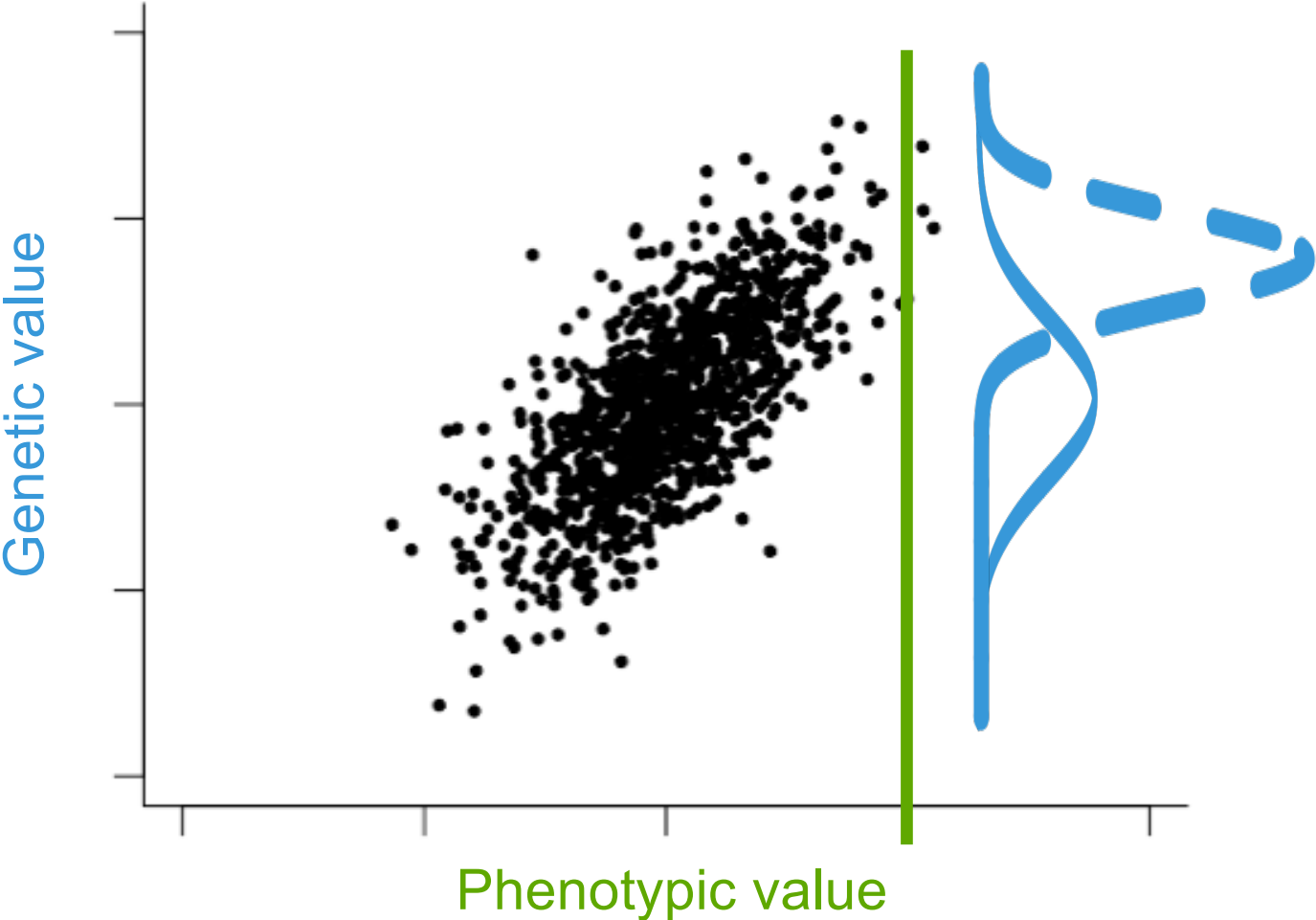
Forward (DATA GENERATION) view



Backward/Inverse (INFERENCE) view - no info.



Backward/Inverse (INFERENCE) view – pheno info.



Forward and Backward/Inverse views

- Forward (DATA GENERATION) view
 - parameters \rightarrow data
 - the true process (unknown!!!)
 - a simulation model (assumptions!!!)
 - a data analysis model (assumptions!!!)
- Backward/Inverse (INFERENCE) view
 - data \rightarrow parameters
 - inferring parameters of the data analysis model
- Don't confuse with Backward/Forward-in-time simulation!
(that is all DATA GENERATION)

Estimand, estimator, estimate!

- Estimand
→ The quantity that we want to infer from the data
- Estimator
→ The inference procedure based on some assumptions (an assumed model)
- Estimate
→ The value obtained from the inference procedure applied to your data



ESTIMAND



ESTIMATE

ESTIMATOR

Ingredients	Directions
150g unsalted butter	1. Heat oven to 160C.
150g chocolate pieces	Grease 1 liter glass baking pan. Line a 450g loaf tin with baking paper.
150g all-purpose flour	2. Melt butter and chocolate in a saucepan over low heat.
1/2 tsp baking powder	
1/2 tsp baking soda	
200g brown sugar	
2 large eggs	



Richard McElreath 
@rlmcelreath

Questions?!

Fisher's 1918 model

- Observe: Phenotypic value, y_i
- Estimand: Genotypic/Genetic value, g_i (or its components)
- Assume:

Phenotype = Genotype + Environment

$$y_i = \mu + g_i + e_i$$

Genotype = Additive + Dominance + Epistasis

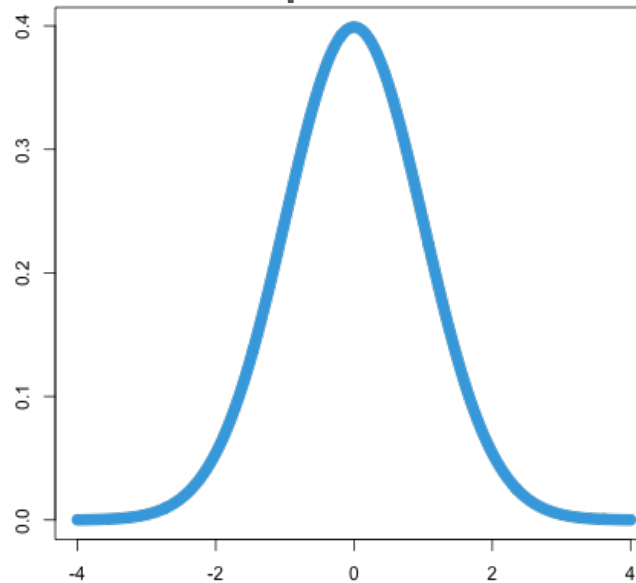
$$g_i = a_i + d_i + x_i$$

$$y_i \approx \mu + a_i + e_i$$

Case with no information

- Assume: $a_i \sim N(0, \sigma_a^2)$
- Having no information about an individual, what can we say about its breeding value?
- Let's follow the above assumptions!

a_i

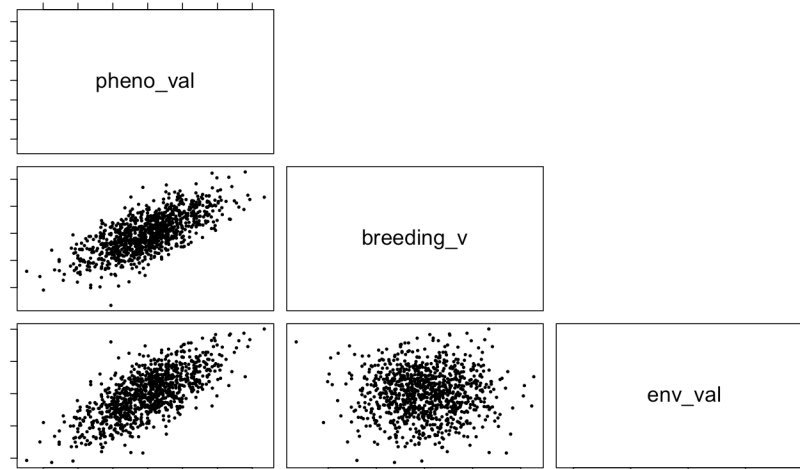
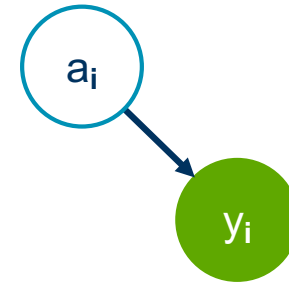


Case with own phenotype

- Assume:

$$y_i \approx \mu + a_i + e_i$$

$$\begin{pmatrix} y_i \\ a_i \\ e_i \end{pmatrix} \sim N \left(\begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & & \\ \sigma_a^2 & \sigma_a^2 & \\ \sigma_e^2 & 0 & \sigma_e^2 \end{pmatrix} \text{sym.} \right)$$

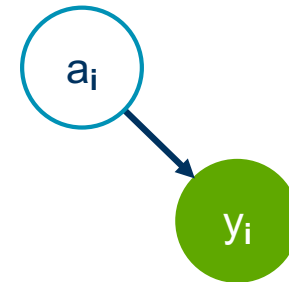


Case with own phenotype

- Assume:

$$y_i \approx \mu + a_i + e_i$$

$$\begin{pmatrix} y_i \\ a_i \\ e_i \end{pmatrix} \sim N \left(\begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & & \\ \sigma_a^2 & \sigma_a^2 & \\ \sigma_e^2 & 0 & \sigma_e^2 \end{pmatrix} \right) \begin{matrix} \\ \\ \text{sym.} \end{matrix}$$

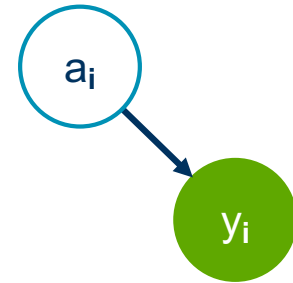


- Observing phenotypic value of an individual, what can we say about its breeding value?
- Given the above assumptions and observed phenotype we can calculate conditional distribution $p(a_i|y_i)$

Case with own phenotype

- Assume: $y_i \approx \mu + a_i + e_i$

$$\begin{pmatrix} y_i \\ a_i \\ e_i \end{pmatrix} \sim N \left(\begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & & \\ \sigma_a^2 & \sigma_a^2 & \\ \sigma_e^2 & 0 & \sigma_e^2 \end{pmatrix} \text{sym.} \right)$$



- Estimator (summarising the conditional distribution):

$$E(a_i | y_i) = \hat{a}_i = E(a_i) + \text{Cov}(a_i, y_i) \text{Var}(y_i)^{-1} (y_i - E(y_i))$$

$$\text{Var}(a_i | y_i) = \text{Var}(a_i) - \text{Cov}(a_i, y_i) \text{Var}(y_i)^{-1} \text{Cov}(y_i, a_i)$$

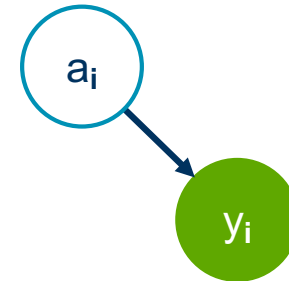
$$\text{Cor}(a_i, y_i)^2 = 1 - \text{Var}(a_i | y_i) / \text{Var}(a_i)$$

Case with own phenotype

- Assume:

$$y_i \approx \mu + a_i + e_i$$

$$\begin{pmatrix} y_i \\ a_i \\ e_i \end{pmatrix} \sim N \left(\begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & & \\ \sigma_a^2 & \sigma_a^2 & \\ \sigma_e^2 & 0 & \sigma_e^2 \end{pmatrix} \right) \begin{matrix} \\ \\ \text{sym.} \end{matrix}$$



- Estimator (summarising the conditional distribution):

$$E(a_i | y_i) = \sigma_a^2 \sigma_y^{-2} (y_i - \mu) = h^2 (y_i - \mu)$$

$$\text{Var}(a_i | y_i) = \sigma_a^2 - \sigma_a^2 \sigma_y^{-2} \sigma_a^2 = \sigma_a^2 (1 - h^2)$$

$$\text{Cor}(a_i, y_i) = h \quad \text{Cor}(a_i, y_i)^2 = h^2$$

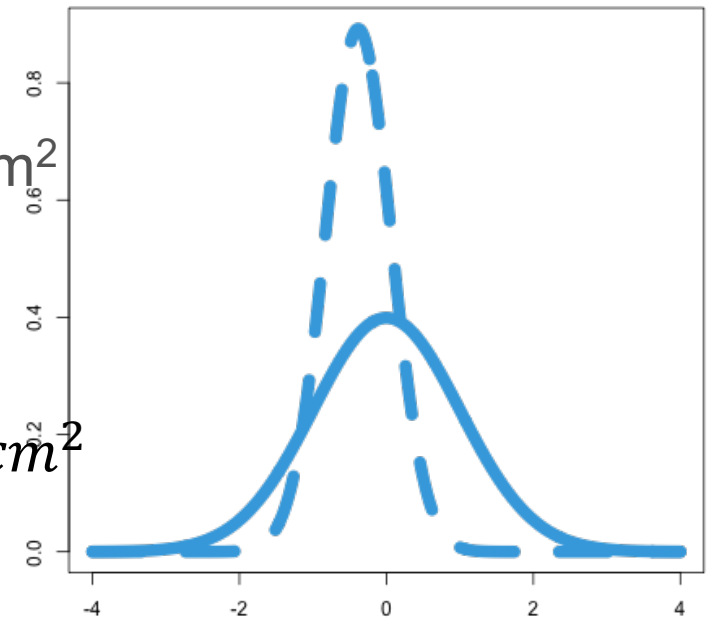
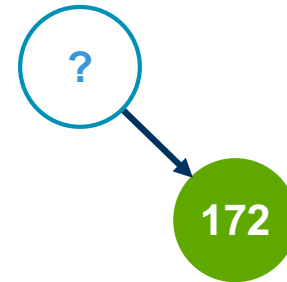
Case with own phenotype - example

- Observe: GG's height is 172 cm
- Assume (from literature):
 - mean height in males is ~175 cm
 - heritability for height is ~0.8
 - phenotypic standard deviation is ~7 cm
 - genetic variance is $7^2 \cdot 0.8 = 39 \text{ cm}^2 = 6.3^2 \text{ cm}^2$
- Estimator & Estimate:

$$E(a_i | y_i = 172) = 0.8(172 - 175) = -2.4 \text{ cm}$$

$$\text{Var}(a_i | y_i) = 7^2 \cdot 0.8(1 - 0.8) = 7.84 \text{ cm}^2 = 2.8^2 \text{ cm}^2$$

$$\text{Cor}(a_i, y_i) = \sim 0.9, \text{Cor}(a_i, y_i)^2 = 0.8$$



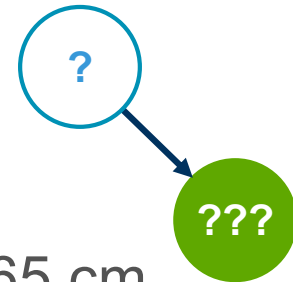
Case with own phenotype - example

- Observe: YOUR height is ??? cm
- Assume (from literature):
 - mean height in males is ~175 cm / females is ~165 cm
 - heritability for height is ~0.8
 - phenotypic standard deviation is ~7 cm
 - genetic variance is $7^2 \cdot 0.8 = 39 \text{ cm}^2 = 6.3^2 \text{ cm}^2$
- Estimator & Estimate:

$$E(a_i | y_i = ???) = 0.8(??? - ???) = -??? \text{ cm}$$

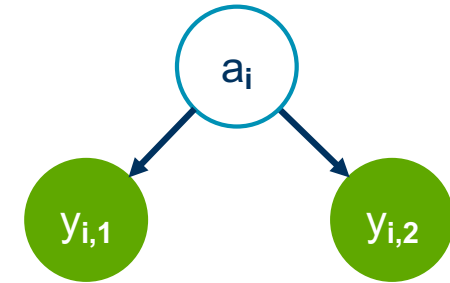
$$\text{Var}(a_i | y_i) = 7^2 \cdot 0.8(1 - 0.8) = 7.84 \text{ cm}^2 = 2.8^2 \text{ cm}^2$$

$$\text{Cor}(a_i, y_i) = \sim 0.9, \text{Cor}(a_i, y_i)^2 = 0.8$$



Case with repeated own phenotypes

- Assume: $y_{i,j} \approx \mu + a_i + p_i + e_{i,j}$



- Estimator:

$$E(a_i | \bar{y}_i) = \hat{a}_i = k(\bar{y}_i - \mu)$$

$$\text{Cor}(a_i, \bar{y}_i)^2 = k$$

$$k = \frac{nh^2}{1+(n-1)(h^2+p^2)}$$

$$p^2 = \sigma_p^2 \sigma_y^{-2}$$

Case with repeated own phenotypes

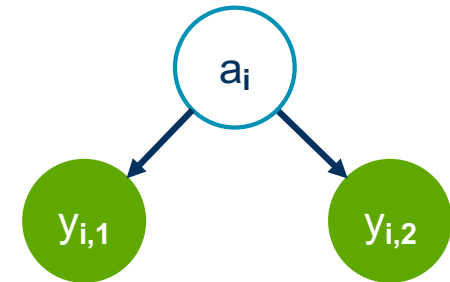


Table 1.1. Percentage increase in accuracy of prediction with repeated records compared with single records at a heritability of 0.35.

<i>t</i> values	Number of records				
	2	4	6	8	10
0.2	29	58	73	83	89
0.5	15	26	31	33	35
0.8	5	8	10	10	10

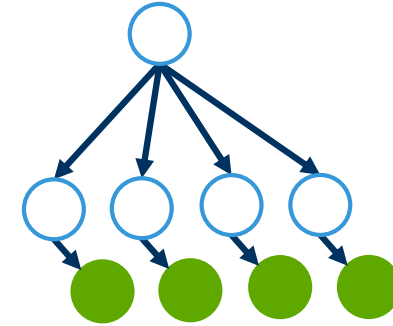
$$t = h^2 + p^2$$

See chapter 2 in
Mrode et al. (2023)

Case with progeny phenotypes

- Assume:

$$y_{i,j} \approx \mu + \frac{1}{2}a_i + e_{i,j}$$



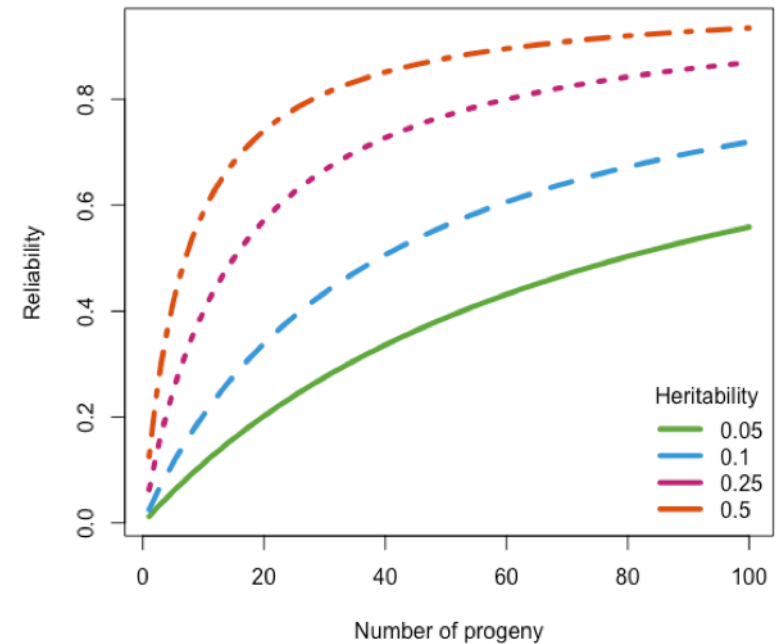
- Estimator:

$$E(a_i | \bar{y}_i) = \hat{a}_i = 2k(\bar{y}_i - \mu)$$

$$Cor(a_i, \bar{y}_i)^2 = k$$

$$k = \frac{n}{n + (4 - h^2)/h^2}$$

See chapter 2 in
Mrode et al. (2023)



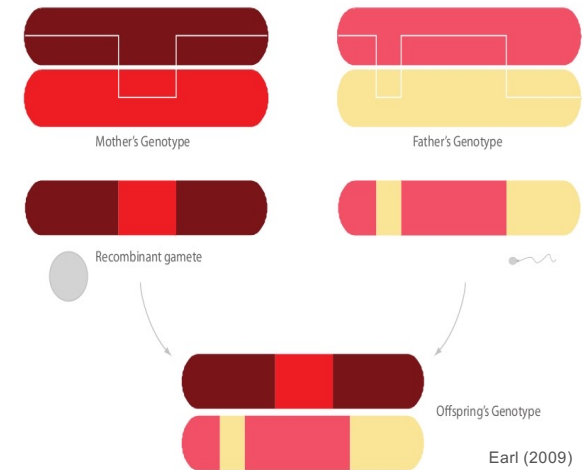
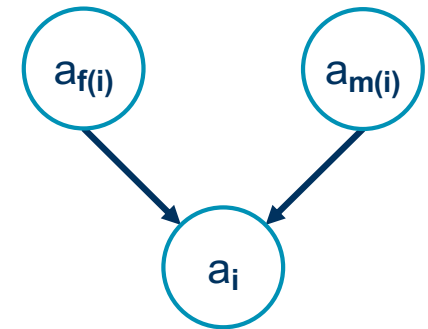
Questions?!

Case with progeny prediction (a pedigree detour)

- Assume:

$$a_i = a_{i,1} + a_{i,2}$$

- Variance between breeding values?
- Key rules:
 - $\text{Var}(x \pm y) = \text{Var}(x) + \text{Var}(y) \pm 2\text{Cov}(x, y)$
 - $\text{Var}(ax) = a^2 \text{Var}(x)$



Case with progeny prediction (a pedigree detour)

- Assume:

$$a_i = a_{i,1} + a_{i,2}$$

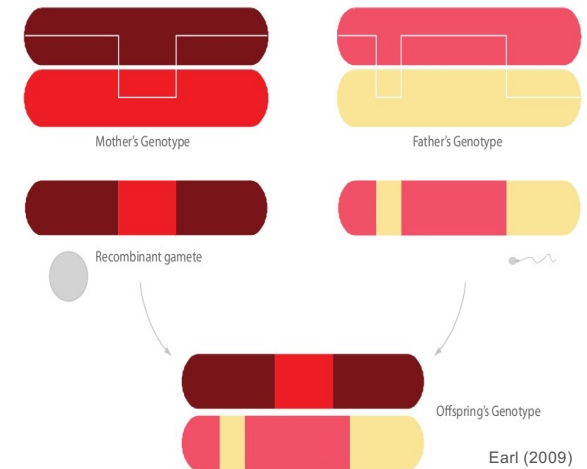
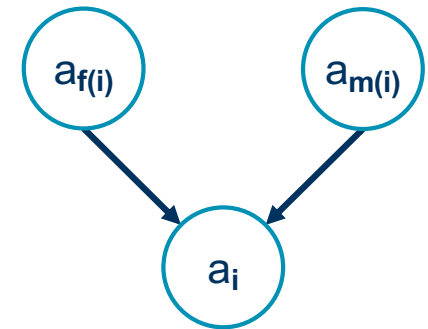
$$\begin{aligned} \text{Var}(a_i) &= \text{Var}(a_{i,1} + a_{i,2}) = \frac{1}{2}\sigma_a^2 + \frac{1}{2}\sigma_a^2 + F_i\sigma_a^2 \\ &= (1 + F_i)\sigma_a^2 \end{aligned}$$

$$a_i \sim N(0, (1 + F_i)\sigma_a^2)$$

- Some cases:

– Non-inbred individuals $a_i \sim N(0, \sigma_a^2)$

– Fully-inbred individuals $a_i \sim N(0, 2\sigma_a^2)$



Case with progeny prediction (a pedigree detour)

- Assume:

$$a_i = a_{i,1} + a_{i,2}$$

$$\begin{aligned} \text{Var}(a_i) &= \text{Var}(a_{i,1} + a_{i,2}) = \frac{1}{2}\sigma_a^2 + \frac{1}{2}\sigma_a^2 + F_i\sigma_a^2 \\ &= (1 + F_i)\sigma_a^2 \end{aligned}$$

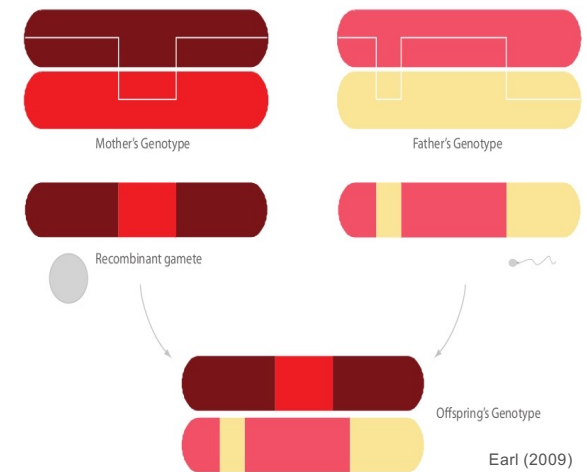
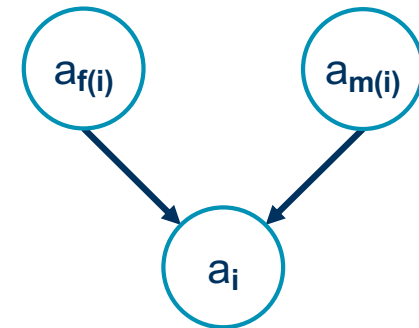
$$a_i \sim N(0, (1 + F_i)\sigma_a^2)$$

$$a_{i,1} = \frac{1}{2}a_{f(i),1} + \frac{1}{2}a_{f(i),2} + r_{i,1}$$

$$a_{i,2} = \frac{1}{2}a_{m(i),1} + \frac{1}{2}a_{m(i),2} + r_{i,2}$$

$$a_i = \frac{1}{2}a_{f(i)} + \frac{1}{2}a_{m(i)} + r_i$$

$$a_i | a_{f(i)}, a_{m(i)} \sim N\left(\frac{1}{2}a_{f(i)} + \frac{1}{2}a_{m(i)}, k\sigma_a^2\right)$$



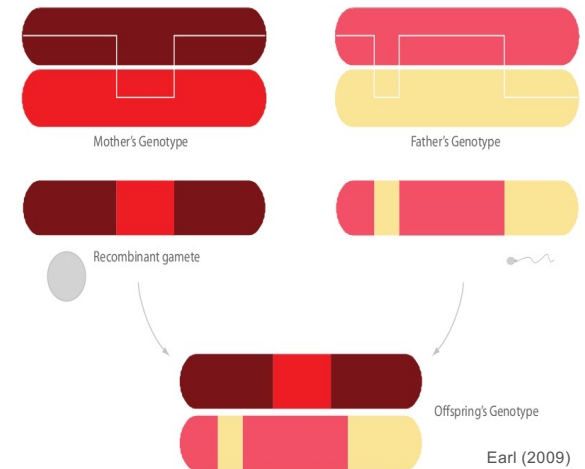
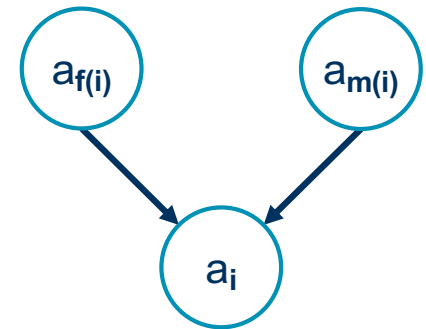
Case with progeny prediction (a pedigree detour)

- Assume:

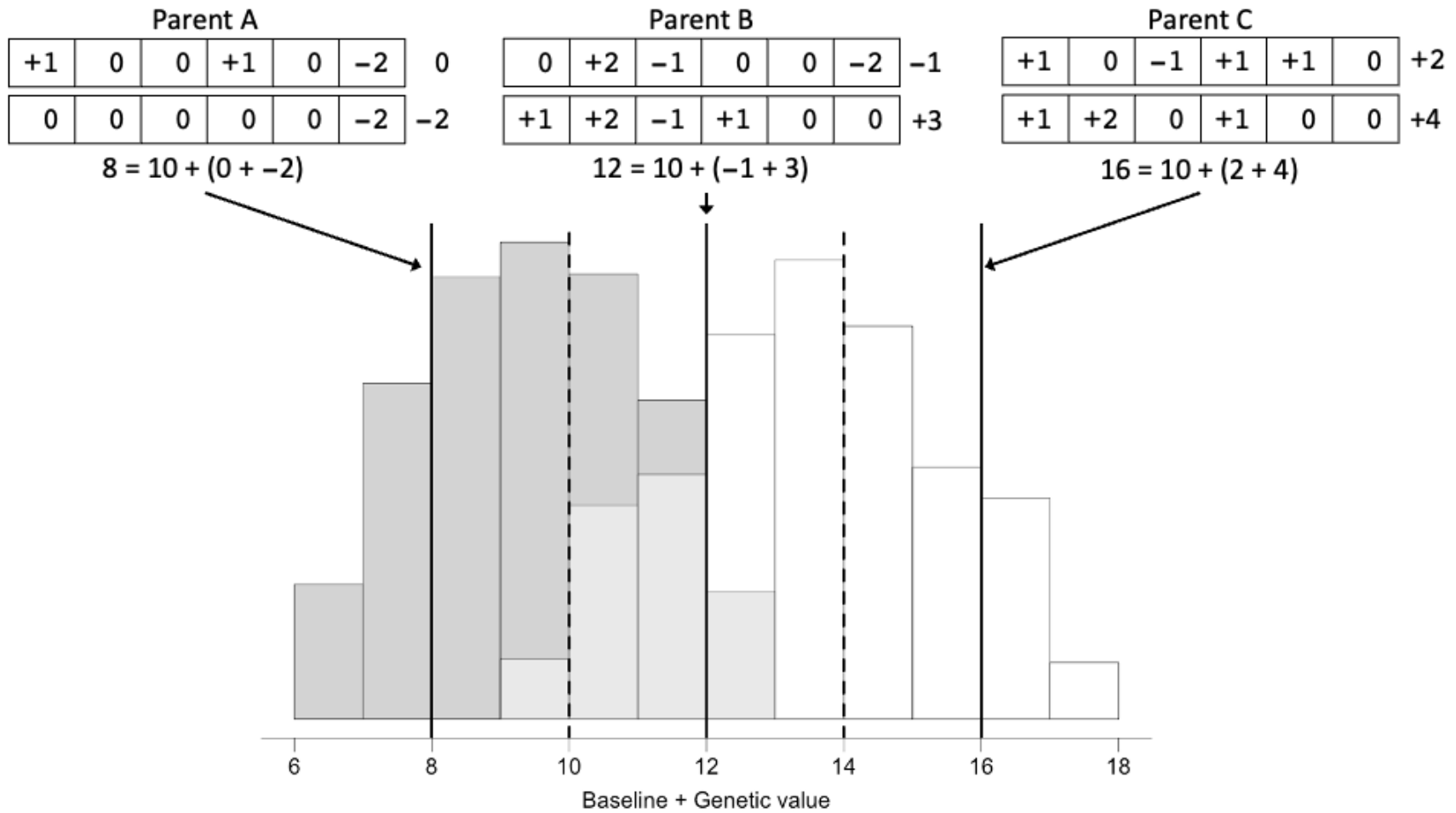
$$a_i = \frac{1}{2}a_{f(i)} + \frac{1}{2}a_{m(i)} + r_i$$

$$a_i | a_{f(i)}, a_{m(i)} \sim N\left(\frac{1}{2}a_{f(i)} + \frac{1}{2}a_{m(i)}, k\sigma_a^2\right)$$

- Parent average (expectation) $\frac{1}{2}a_{f(i)} + \frac{1}{2}a_{m(i)}$
- Mendelian sampling r_i (deviation from the expectation due to recombination, segregation, and an odd mutation)



Between and within family genetic variation



Case with progeny prediction (a pedigree detour)

- Assume:

$$a_i = \frac{1}{2}a_{f(i)} + \frac{1}{2}a_{m(i)} + r_i$$

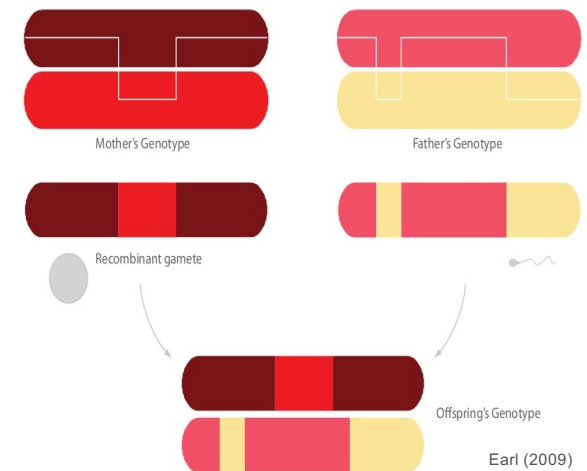
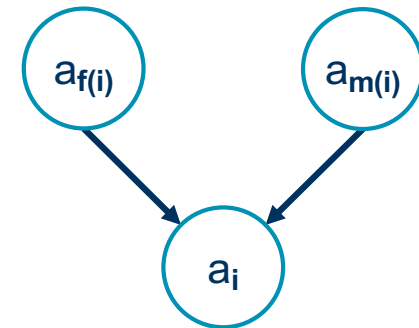
$$a_i | a_{f(i)}, a_{m(i)} \sim N\left(\frac{1}{2}a_{f(i)} + \frac{1}{2}a_{m(i)}, k\sigma_a^2\right)$$

$$r_i \sim N(0, k\sigma_a^2)$$

$$k = \frac{1}{2} - \frac{1}{4}(F_{f(i)} + F_{m(i)})$$

- Some cases:

- Non-inbred parents $k = \frac{1}{2}$
- Fully-inbred parents $k = 0$



Case with progeny prediction

- Assume:

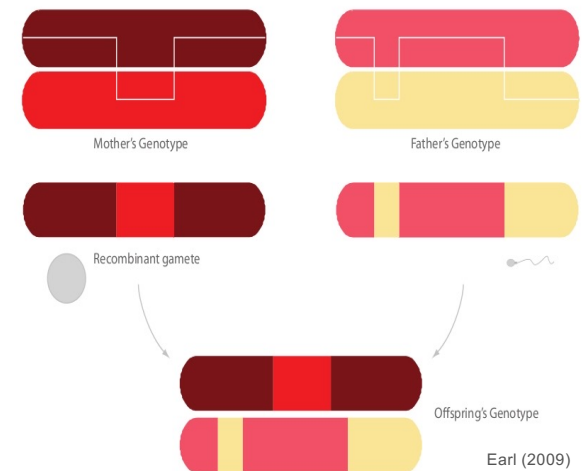
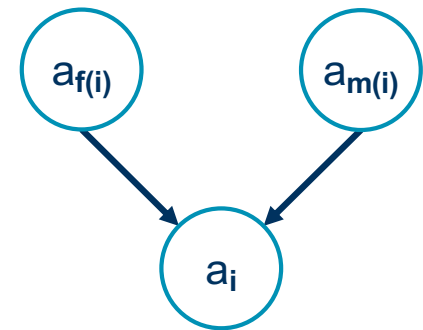
$$a_i = \frac{1}{2}a_{f(i)} + \frac{1}{2}a_{m(i)} + r_i$$

$$a_i | a_{f(i)}, a_{m(i)} \sim N\left(\frac{1}{2}a_{f(i)} + \frac{1}{2}a_{m(i)}, k\sigma_a^2\right)$$

- Estimator/predictor:

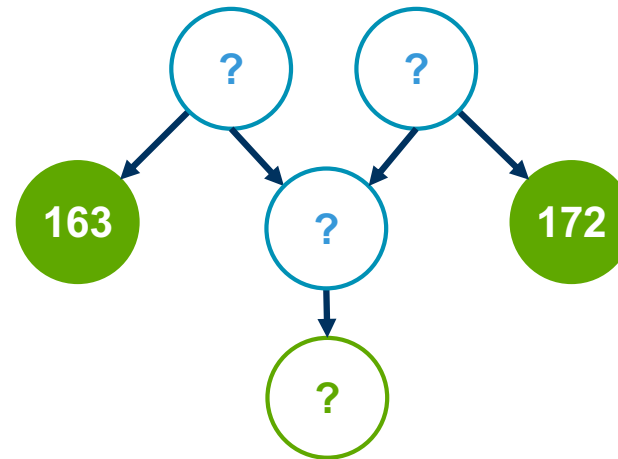
$$E(a_i | \hat{a}_{f(i)}, \hat{a}_{m(i)}) = \hat{a}_i = \frac{1}{2}\hat{a}_{f(i)} + \frac{1}{2}\hat{a}_{m(i)}$$

$$Cor(a_i, \frac{1}{2}\hat{a}_{f(i)} + \frac{1}{2}\hat{a}_{m(i)})^2 = \frac{1}{4}(R_{f(i)}^2 + R_{m(i)}^2)$$



Case with progeny prediction - example

- Assume:
 - average height
 - males ~175 cm
 - females ~165 cm
 - heritability ~0.8
- Calculate



$$E(a_{f(i)} | y_{f(i)} = 172) = 0.8(172 - 175) = -2.4cm \quad R^2 = 0.80$$

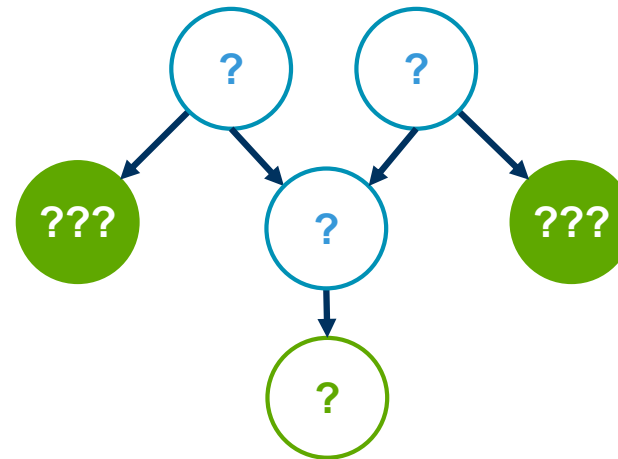
$$E(a_{m(i)} | y_{m(i)} = 163) = 0.8(163 - 165) = -1.6cm \quad R^2 = 0.80$$

$$E(a_i | \hat{a}_{f(i)}, \hat{a}_{m(i)}) = \frac{1}{2}(-2.4 - 1.6) = -2cm \quad R^2 = 0.40$$

$$E(y_i | \hat{a}_i) = 165 - 2 = 163cm \quad R^2 = 0.32$$

Case with progeny prediction - exercise

- Assume:
 - average height
 - males ~175 cm
 - females ~165 cm
 - heritability ~0.8
- Calculate



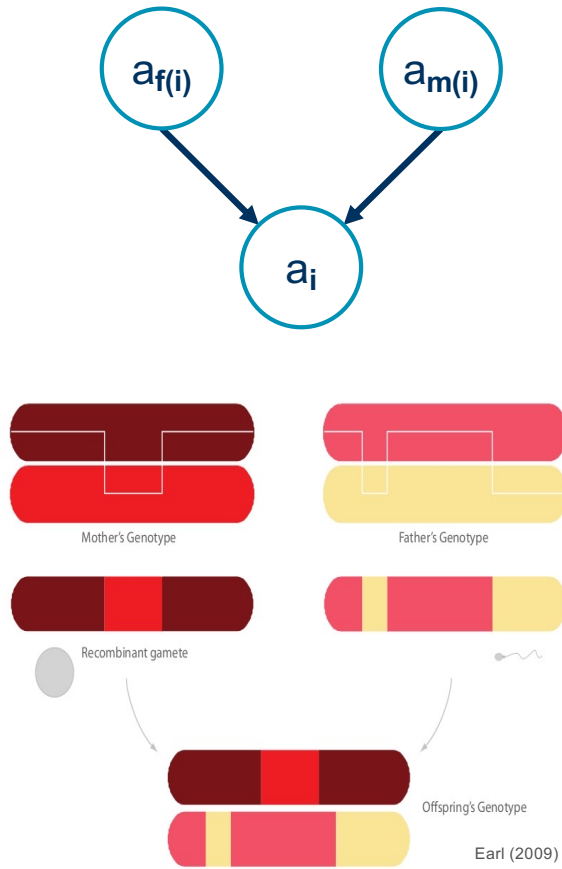
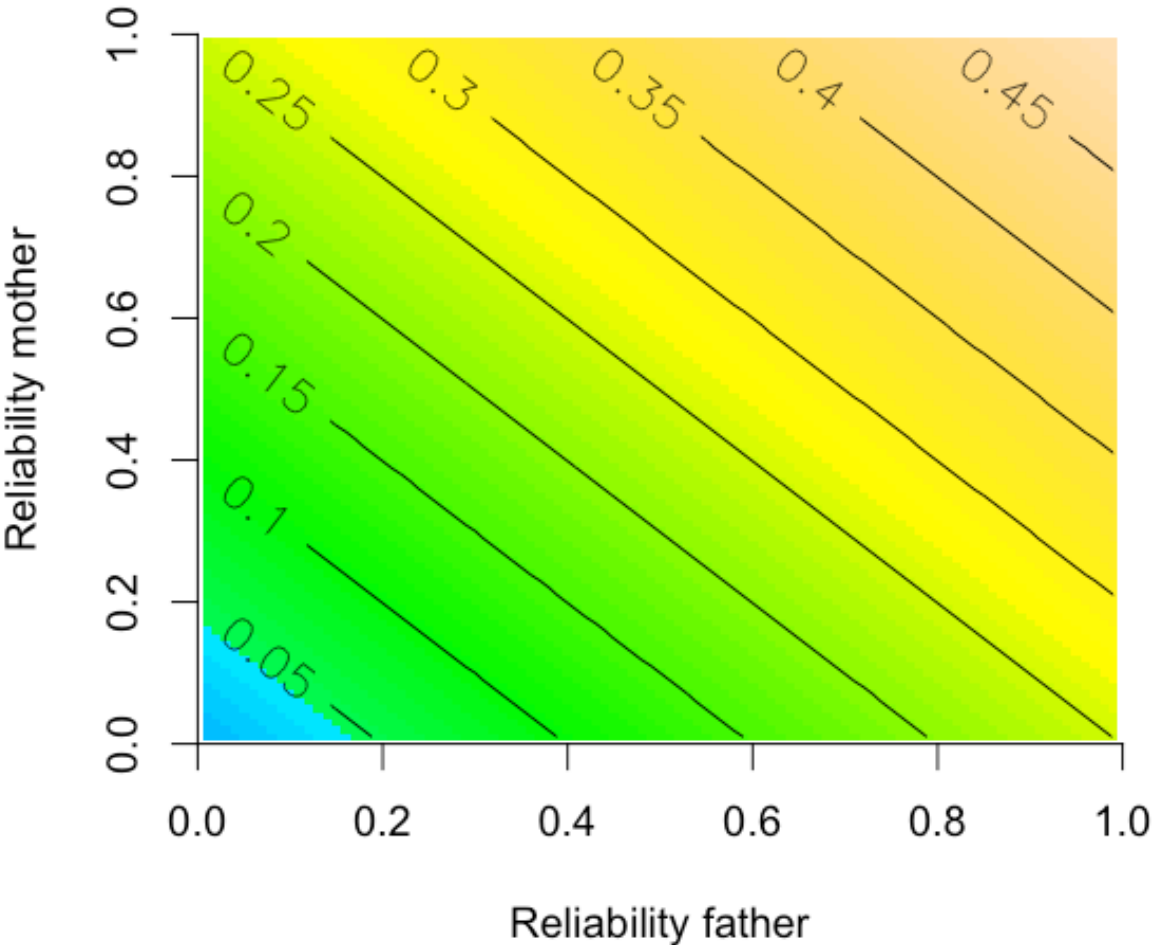
$$E(a_{f(i)} | y_{f(i)} = \text{?}) = R^2 = 0.80$$

$$E(a_{m(i)} | y_{m(i)} = \text{?}) = R^2 = 0.80$$

$$E(a_i | \hat{a}_{f(i)}, \hat{a}_{m(i)}) = R^2 = 0.40$$

$$E(y_i | \hat{a}_i) = R^2 = 0.32$$

Case with progeny prediction



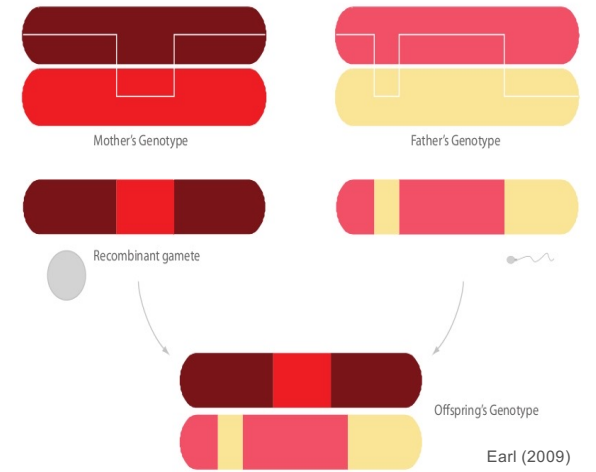
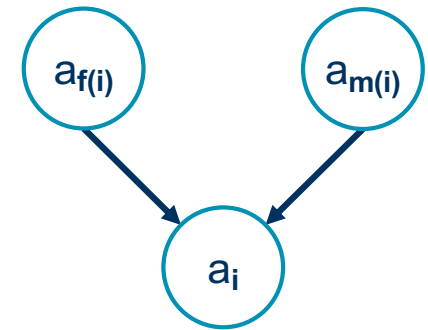
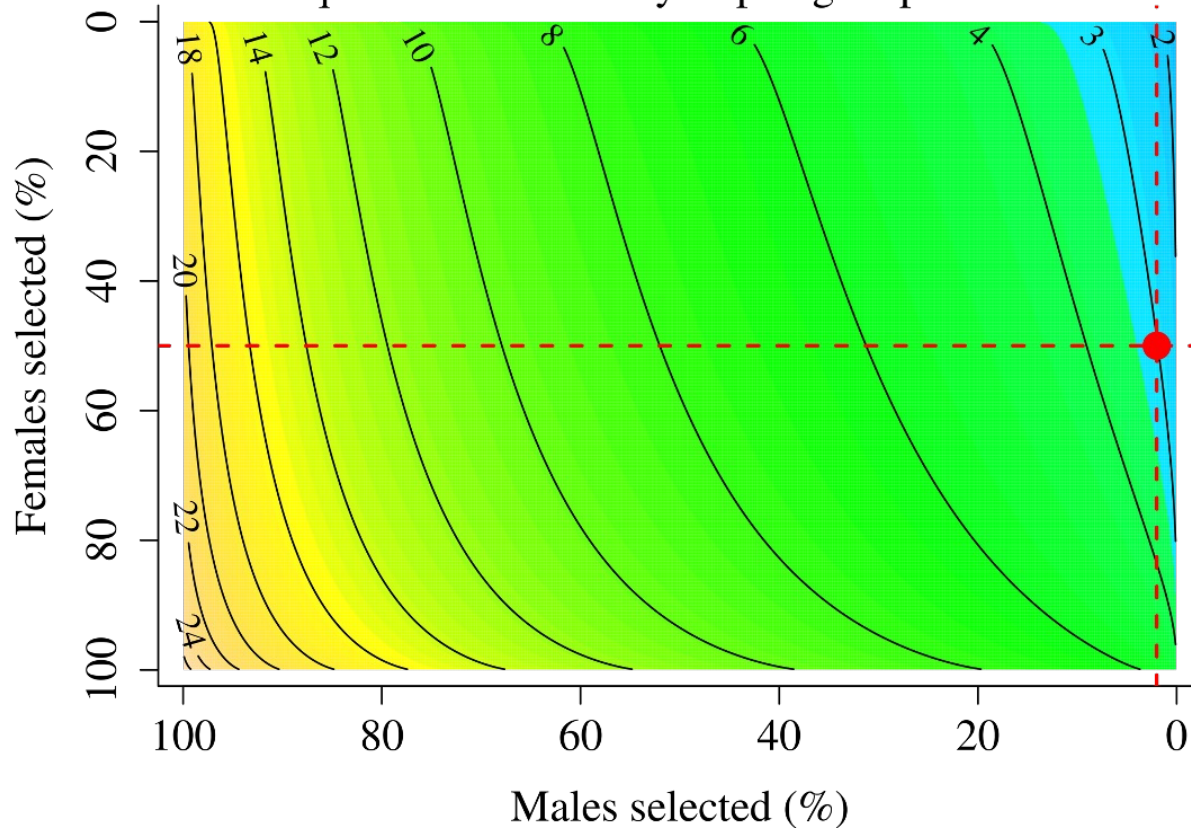
Careful with selection based on progeny predictions!

- Low reliability (halves with every generation)
→ very quickly ~random selection
- Parent average captures only between-family genetic variance (=half of genetic variance)
→ selection on it erodes diversity quickly

Case with progeny prediction with selection

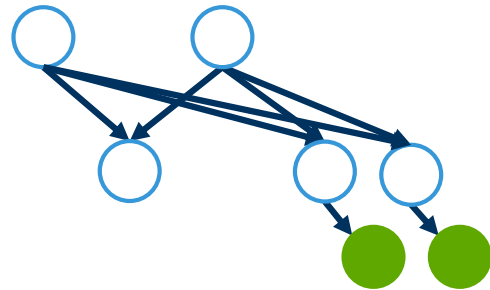
Gorjanc et al. (2015) GSE 47:65

Equilibrium reliability of pedigree prediction

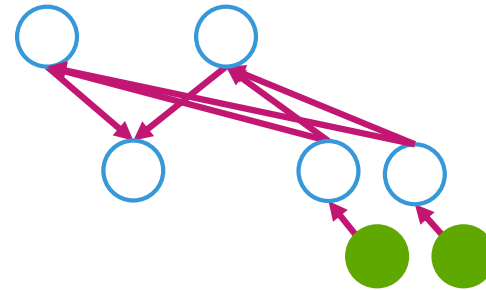


Case with sib phenotypes

Gene flow



Information flow



Sib test = Progeny test of parents &
Progeny prediction

Without genomic data sibs don't give direct information
about other sibs

Learning objectives

- Understand forward and backward/inverse views of working with models
- Familiarise yourself with simple cases of inferring breeding values from the data
- Practice inference of breeding values for simple cases

Questions?!



THE UNIVERSITY
of EDINBURGH



Introduction to genetic evaluation (intuition & simple cases)

Gregor Gorjanc, Chris Gaynor, Jon Bancic, Daniel Tolhurst

UNE, Armidale

2024-02-07

