

# THE INFINITESIMAL WITH DOMINANCE

*CIRM, February 2020*

## Recap of the additive model

$$\text{Trait value} = \underbrace{\text{genetic}}_Z + \underbrace{\text{non-genetic}}_E$$

I am going to ignore the environmental component  $E$ .

Parental trait values  $z_1, z_2$ . Genetic component:

$$Z = \frac{z_1 + z_2}{2} + R, \quad R \sim \mathcal{N}(0, V_0)$$

Deterministic shared component, normally distributed residual

In a large outcrossing population,  $V_0 = \text{constant}$ , otherwise decreases in proportion to relatedness, for parents  $i[1], i[2]$ ,  
 $V_0 = \sigma_A^2(1 - F_{i[1],i[2]})$ .

$$\sigma_A^2 = \frac{2}{M} \sum_{l=1}^M \mathbb{E}[\eta_l(\hat{\chi}_l)^2]$$

# The infinitesimal model

Let

1.  $\mathcal{P}^{(t)}$  denote the *pedigree* relationships between all individuals up to and including generation  $t$ ;
2.  $Z^{(t)}$  denote the *traits* of all individuals in the pedigree up to and including the  $t$ th generation.

The distribution of trait values in generation  $t$ , conditional on knowing  $\mathcal{P}^{(t)}$  and  $Z^{(t-1)}$ , is multivariate normal.

**THIS IS A STATEMENT ABOUT DISTRIBUTION WITHIN FAMILIES, NOT ACROSS THE WHOLE POPULATION**

The pedigree can be quite general. It can, for example, capture population structure.

## Why does it work in the additive case?

For simplicity, consider haploid model. In Amandine's notation,

$$Z^j = \bar{z}_0 + \sum_{l=1}^M \frac{1}{\sqrt{M}} \eta_l(\chi_l^j).$$

Simplest imaginable model: suppose  $\eta_l$  are i.i.d. with  $\eta_l = \pm 1$  with equal probability,  $\bar{z}_0 = 0$ .

Because many different combinations of alleles can lead to the same allelic state, 'typically', knowing the trait value of an individual tells us very little about the allelic state at any given locus.

## Put more mathematically

$$\begin{aligned}\mathbb{P}[\eta_1 = 1 | Z = k/\sqrt{M}] &= \frac{\mathbb{P}\left[\sum_{l=1}^M \eta_l = k \mid \eta_1 = 1\right]}{\mathbb{P}\left[\sum_{l=1}^M \eta_l = k\right]} \mathbb{P}[\eta_1 = 1] \\ &= \frac{\mathbb{P}\left[\sum_{l=2}^M \eta_l = (k-1)\right]}{\mathbb{P}\left[\sum_{l=1}^M \eta_l = k\right]} \mathbb{P}[\eta_1 = 1] \\ &= \frac{\frac{1}{2^{M-1}} \binom{M-1}{(M+k-2)/2}}{\frac{1}{2^M} \binom{M}{(M+k)/2}} \mathbb{P}[\eta_1 = 1] \\ &= \left(1 + \frac{k}{M}\right) \mathbb{P}[\eta_1 = 1].\end{aligned}$$

## Toy example continued

If scaled allelic effects are i.i.d. Bernoulli,

$$\mathbb{P} \left[ \eta_1 = 1 \mid Z = \frac{k}{\sqrt{M}} \right] = \left( 1 + \frac{k}{M} \right) \mathbb{P} [\eta_1 = 1].$$

For a 'typical' trait value,  $k/M = \mathcal{O}(1/\sqrt{M})$ .

For extreme values ( $k = \pm M$ ), the trait gives complete information about the allelic effect at each locus.

For 'typical'  $k$ , the distribution of  $\eta_1$  is almost unchanged because there are so many different configurations of allelic effects that correspond to the same trait value.

# Dominance?

Trait value:

$$Z = \bar{z}_0 + \frac{1}{\sqrt{M}} \sum_{l=1}^M (\eta_l(\chi_l^1) + \eta_l(\chi_l^2) + \phi_l(\chi_l^1, \chi_l^2)).$$

For an individual in the ancestral population,  $\hat{\chi}_l^1, \hat{\chi}_l^2$ , two independent draws from  $\hat{\nu}_l$ .

Conventions:  $\mathbb{E}[\eta_l(\hat{\chi}_l)] = 0$ ,  $\mathbb{E}[\phi_l(\hat{\chi}_l^1, \hat{\chi}_l^2)] = 0$ ,

for any value  $x'$  of the allelic state at locus  $l$ ,  
 $\mathbb{E}[\phi_l(\hat{\chi}_l, x')] = 0 = \mathbb{E}[\phi_l(x', \hat{\chi}_l)]$ .

## Why does this not result in a loss in generality?

Replace  $\phi_l(\hat{\chi}_l^1, \hat{\chi}_l^2)$  by

$$\phi_l(\hat{\chi}_l^1, \hat{\chi}_l^2) - \mathbb{E}[\phi_l(\hat{\chi}_l^1, \hat{\chi}_l^2) | \hat{\chi}_l^1] - \mathbb{E}[\phi_l(\hat{\chi}_l^1, \hat{\chi}_l^2) | \hat{\chi}_l^2] + \mathbb{E}[\phi_l(\hat{\chi}_l^1, \hat{\chi}_l^2)].$$

Second and third terms (functions of  $\hat{\chi}_l^1, \hat{\chi}_l^2$ ) can be subsumed into  $\eta_l(\hat{\chi}_l)$ .

For any value  $x'$  of the allelic state at locus  $l$ :

$$\mathbb{E}[\phi_l(\hat{\chi}_l, x')] = \int \phi_l(x, x') \hat{\nu}_l(dx) = 0 = \mathbb{E}[\phi_l(x', \hat{\chi}_l)].$$

With this modification,  $\mathbb{E}[\phi_l(\hat{\chi}_l^1, \hat{\chi}_l^2)] = 0$ . Moreover, absorbing the mean into  $\bar{z}_0$ , we may assume that  $\mathbb{E}[\eta_l(\hat{\chi}_l)] = 0$ .



# Decomposing the trait value

Write  $i[1]$ ,  $i[2]$  for parents of individual  $i$

Decompose trait value into two parts,

- ▶ component shared by all offspring of  $Z^{i[1]}$  and  $Z^{i[2]}$ ;
- ▶ component which is independent for each individual within the family.

$$\underbrace{\text{trait value}}_{Z^i} = \underbrace{\text{shared}}_{A^i + D^i} + \underbrace{\text{residual}}_{R^i + S^i}$$

$A^i + R^i$  additive component;  $D^i + S^i$  dominance deviation.

Residuals  $R^i$ ,  $S^i$  determined by Mendelian inheritance

# Inheritance

Label chromosomes in individual  $i$ , 1 and 2, according to whether inherited from  $i[1]$  or  $i[2]$ .

$X_l^i, Y_l^i$ , independent  $\text{Ber}(1/2)$  r.v.'s

- ▶  $X_l^i = 1$  if type at locus  $l$  on chromosome 1 in individual  $i$ , inherited from chromosome 1 in  $i[1]$ ;
- ▶  $Y_l^i = 1$  if type at locus  $l$  on chromosome 2 in individual  $i$  inherited from chromosome 1 in  $i[2]$ .

$$Z^i = \bar{z}_0 + A^i + D^i + R^i + S^i.$$

Mirrors our approach in additive case:  $A^i + D^i$  centre on the means of  $X_l^i, Y_l^i$ ; residuals  $R^i + S^i$  capture deviation of  $X_l^i, Y_l^i$  from their means, i.e. the randomness of Mendelian inheritance.

## Terms shared by all descendants of $i[1]$ , $i[2]$

$$A^i = \frac{1}{2\sqrt{M}} \sum_{l=1}^M \left( \eta(\chi_l^{i[1],1}) + \eta(\chi_l^{i[1],2}) + \eta(\chi_l^{i[2],1}) + \eta(\chi_l^{i[2],2}) \right)$$

and

$$D^i = \frac{1}{4\sqrt{M}} \sum_{l=1}^M \left\{ \begin{aligned} &\phi_l(\chi_l^{i[1],1}, \chi_l^{i[2],1}) + \phi_l(\chi_l^{i[1],1}, \chi_l^{i[2],2}) \\ &+ \phi_l(\chi_l^{i[1],2}, \chi_l^{i[2],1}) + \phi_l(\chi_l^{i[1],2}, \chi_l^{i[2],2}) \end{aligned} \right\}.$$

Even if we condition on  $Z^{i[1]}$ ,  $Z^{i[2]}$ , with nontrivial dominance, these terms are *random*

# Residuals

$$R^i = \frac{1}{\sqrt{M}} \sum_{l=1}^M \left\{ (X_l^i - \frac{1}{2}) \eta_l(\chi_l^{i[1],1}) + (\frac{1}{2} - X_l^i) \eta_l(\chi_l^{i[1],2}) \right. \\ \left. + (Y_l^i - \frac{1}{2}) \eta_l(\chi_l^{i[2],1}) + (\frac{1}{2} - Y_l^i) \eta_l(\chi_l^{i[2],2}) \right\},$$

$$S^i = \frac{1}{\sqrt{M}} \sum_{l=1}^M \left\{ (X_l^i Y_l^i - \frac{1}{4}) \phi_l(\chi_l^{i[1],1}, \chi_l^{i[2],1}) \right. \\ \left. + (X_l^i (1 - Y_l^i) - \frac{1}{4}) \phi_l(\chi_l^{i[1],1}, \chi_l^{i[2],2}) \right. \\ \left. + ((1 - X_l^i) Y_l^i - \frac{1}{4}) \phi_l(\chi_l^{i[1],2}, \chi_l^{i[2],1}) \right. \\ \left. + ((1 - X_l^i) (1 - Y_l^i) - \frac{1}{4}) \phi_l(\chi_l^{i[1],2}, \chi_l^{i[2],2}) \right\}.$$

## When will everything be well defined?

By construction,  $\mathbb{E}[R^i + S^i] = 0$ , and  $\mathbb{E}[\eta_l(\hat{\chi}_l)] = 0 \implies \mathbb{E}[A^i] = 0$ .

$$D^i = \frac{1}{4\sqrt{M}} \sum_{l=1}^M \left\{ \phi_l(\chi_l^{i[1],1}, \chi_l^{i[2],1}) + \phi_l(\chi_l^{i[1],1}, \chi_l^{i[2],2}) \right. \\ \left. + \phi_l(\chi_l^{i[1],2}, \chi_l^{i[2],1}) + \phi_l(\chi_l^{i[1],2}, \chi_l^{i[2],2}) \right\}.$$

$\mathbb{E}[\phi_l(\hat{\chi}_l^1, \hat{\chi}_l^2)] = 0$ , but what if parents are related?

Define *inbreeding depression*

$$\iota = \frac{1}{\sqrt{M}} \sum_{l=1}^M \mathbb{E}[\phi_l(\hat{\chi}_l, \hat{\chi}_l)]$$

$$\mathbb{E}[D^i] = \iota F_{i[1],i[2]}.$$

Need this to be finite

... and then it gets really messy

To calculate the variance of  $D^i$ , we need to know quantities like

$$\mathbb{E} \left[ \left\{ \phi_l(\chi_l^{i[1],1}, \chi_l^{i[2],1}) + \phi_l(\chi_l^{i[1],1}, \chi_l^{i[2],2}) \right. \right. \\ \left. \left. + \phi_l(\chi_l^{i[1],2}, \chi_l^{i[2],1}) + \phi_l(\chi_l^{i[1],2}, \chi_l^{i[2],2}) \right\}^2 \right]$$

and this will depend upon 4-way identities.

# Identities required in expressions for variances

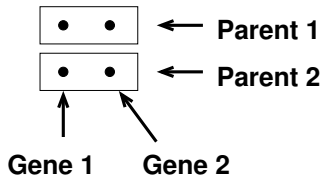
$$F_{1122} = P \left[ \begin{array}{cc} \bullet & \bullet \\ \bullet & \bullet \end{array} \right]$$

$$\tilde{F}_{1122} = P \left[ \begin{array}{cc} \bullet & \bullet \\ \bullet & \bullet \end{array} \right]$$

$$F_{112} = \frac{1}{2} P \left[ \begin{array}{cc} \bullet & \bullet \\ \bullet & \bullet \end{array} \right] \text{ or } \left[ \begin{array}{cc} \bullet & \bullet \\ \bullet & \bullet \end{array} \right]$$

$$F_{122} = \frac{1}{2} P \left[ \begin{array}{cc} \bullet & \bullet \\ \bullet & \bullet \end{array} \right] \text{ or } \left[ \begin{array}{cc} \bullet & \bullet \\ \bullet & \bullet \end{array} \right]$$

$$\tilde{F}_{1212} = \frac{1}{2} P \left[ \begin{array}{cc} \bullet & \bullet \\ \bullet & \bullet \end{array} \right] \text{ or } \left[ \begin{array}{cc} \bullet & \bullet \\ \bullet & \bullet \end{array} \right]$$



## QG coefficients required to describe variance

Additive variance  $\sigma_A^2 = \frac{2}{M} \sum_{l=1}^M \mathbb{E}[\eta_l(\hat{\chi}_l)^2]$

Dominance variance  $\sigma_D^2 = \frac{1}{M} \sum_{l=1}^M \mathbb{E}[\phi_l(\hat{\chi}_l^1, \hat{\chi}_l^2)^2]$

Inbreeding depression  $\iota = \frac{1}{\sqrt{M}} \sum_{l=1}^M \mathbb{E}[\phi_l(\hat{\chi}_l, \hat{\chi}_l)]$

Sum of squared locus-specific inbreeding depressions  $\iota^* = \frac{1}{M} \sum_{l=1}^M \mathbb{E}[\phi_l(\hat{\chi}_l, \hat{\chi}_l)]^2$

Variance of dominance effects in inbred individuals  $\sigma_{DI}^2 = \frac{1}{M} \sum_{l=1}^M (\mathbb{E}[\phi_l(\hat{\chi}_l, \hat{\chi}_l)^2] - \mathbb{E}[\phi_l(\hat{\chi}_l, \hat{\chi}_l)]^2)$

Covariance of additive and dominance effects in inbred individuals  $\sigma_{ADI} = \frac{2}{M} \sum_{l=1}^M \mathbb{E}[\eta_l(\hat{\chi}_l)\phi_l(\hat{\chi}_l, \hat{\chi}_l)]$



But it can be done . . .

And, since the actual trait value depends only on the two alleles carried at each locus,  $\text{Var}(Z^i)$  only depends on pairwise identities:

$$\begin{aligned}\text{Var}(Z^i) = & \sigma_A^2(1 + F_{ii}) + \sigma_D^2(1 - F_{ii}) + (\sigma_{DI}^2 + \iota^*)F_{ii} \\ & + 2F_{ii}\sigma_{ADI} - F_{ii}^2\iota^*.\end{aligned}$$

(We have assumed that the ancestral population is in linkage equilibrium, or there is another term.)

## Conditioning on trait values of parents

$$Z = \bar{z}_0 + \frac{1}{\sqrt{M}} \sum_{l=1}^M (\eta_l(\chi_l^1) + \eta_l(\chi_l^2) + \phi_l(\chi_l^1, \chi_l^2)).$$

Suppose  $\eta_l, \phi_l$  uniformly bounded, and inbreeding depression  $\iota = \frac{1}{\sqrt{M}} \sum_{l=1}^M \mathbb{E}[\phi_l(\hat{\chi}_l, \hat{\chi}_l)] < \infty$ .

Before conditioning, normality from a CLT.

Knowing the trait values of the parents, tells us little about allelic states at any pair of loci.

Mendelian inheritance independent at unlinked loci, guarantees normality residuals after conditioning much as in additive case.

For  $A^i + D^i$  have to work harder; use a variant of Stein's method of exchangeable pairs.

# Main result

- ▶ Trait values across pedigree well approximated by multivariate normal;
- ▶ Trait can be decomposed into two (normally distributed) parts,  $(A^i + D^i)$  (shared by all individuals in family) and  $(R^i + S^i)$  (picked independently for each individual in the family);
- ▶ Conditioning on the parental trait values does not affect the variance of the trait values among offspring, and it shifts mean trait values in a predictable way.

The effect of conditioning can be calculated in the usual way:

$$\begin{bmatrix} x_A \\ x_B \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix} \right).$$

Then

$$x_A | x_B \sim \mathcal{N} \left( \mu_A + \Sigma_{AB} \Sigma_{BB}^{-1} (x_B - \mu_B), \Sigma_{AA} - \Sigma_{AB} \Sigma_{BB}^{-1} \Sigma_{BA} \right).$$

## Key conditions for the approximation

- ▶ Pedigree not too inbred. ( $\Sigma_t$  not too small);
- ▶ Parental trait values not too extreme. ( $\Delta_t$  not too big);
- ▶ inbreeding depression,  $\iota F_{12}$  not too big.

*When we say that the infinitesimal with dominance holds, we mean that the trait values across the pedigree are a multivariate normal; the distribution of trait values across the whole population may be far from normal.*

*Knowing trait values of the parents shifts the mean in a predictable way; the variance is independent of the parental trait values, and is determined by the pedigree and variance components in the ancestral population.*

BACK TO NICK