**Breeding Value Prediction**

Jeffrey Endelman

Agron/Hort 812, Spring 2020

**Average Effects for a Single Locus**

Genotypic value can be partitioned into additive and dominance effects defined by regression equations (Fisher 1918). Let denote which alleles are present in a diploid zygote , for a single locus with *m* alleles. The genotypic value of genotype is denoted , and its frequency at panmictic equilibrium is the product of the allele frequencies: . Using notation from factorial experiments, the regression of the genotypic value on allele dosage is

|  |  |
| --- | --- |
|  | [1] |

where *i* is the **additive effect** of allele *i* and is the population mean. The residual in Eq. 1 is the **dominance effect** for allele pair *ij*. Solving by ordinary least squares leads to the following linear constraint on the additive effects,

|  |  |
| --- | --- |
|  | [2] |

and the solution is the average genotypic value of all individuals with allele *i*, relative to the population mean:

|  |  |
| --- | --- |
|  | [3] |

From Eq. 3, the mean of the additive effects is zero:

|  |  |
| --- | --- |
|  | [4] |

And therefore the mean of the dominance effects is also zero.

At panmictic equilibrium, the covariance between additive and dominance effects is zero, which leads to the following decomposition of variance:

|  |  |
| --- | --- |
| Additive variance =  Dominance variance = | [5] |

The breeding value (BV) of an individual, which is defined as twice the mean genotypic value of its progeny relative to the population mean, equals the sum of its additive effects:

**Bi-allelic case**

When there are only two alleles (*B*/*b*), with the frequency of *B* denoted by *p*, Eq. 1 is a system of three equations:

and Eq. 2 becomes . Defining the **allele substitution effect** as leads to the following expressions:

**Additive Relationship Model**

Previously, we considered a population with i.i.d. genotypic effects. By decomposing the genotypic effect into additive and non-additive effects, we can now model the covariance between genotypes.

|  |  |  |
| --- | --- | --- |
| i.i.d. model |  | [6] |
| Additive Relationship model |  | [7] |

The additive value for genotype is the sum of its additive effects, and the additive relationship (*A*) between genotypes is defined based on Eq. 7: it is the covariance between the additive values, divided by the additive variance:

The vector **r** contains the non-additive values for the genotypes, which are assumed to be i.i.d. For a single locus, the non-additive value is the dominance effect, while for multi-locus traits, epistatic effects are also present. Although there are covariance models for non-additive effects, for now we will assume they are i.i.d.

The covariance between the additive values can be expanded to 4 terms involving the expectation of the product of additive effects:

In the classical theory, this expectation refers to a random process of sampling individuals from a population with a known pedigree relationship. When the two genes are **identical by descent** (IBD)from a common ancestor, the expectation equals (See Eq. 5). The probability that one gene from genotype and one gene from are IBD is defined as **kinship** , also known as the coefficient of coancestry. It follows that

In other words, the additive relationship is twice the kinship. When generalized to polyploids, one finds that additive relationship equals the ploidy times kinship.

**Calculating Additive Relationship** (see also Bernardo 2.9.1–2.9.2)

Kinship coefficients can be calculated from pedigree records using a recursive approach. First, the individuals need to be sorted from oldest to youngest; in other words, ancestors must precede their descendants. The founders are individuals without recorded ancestors, for which assumptions must be made about their degree of inbreeding and their relationship. Typically, they are assumed to be unrelated.

Changing notation, rather than identifying each genotype by its two genes, a single index will be used that takes on values from 1 to *n* (the number of individuals). The kinship between two individuals can be expressed as the average kinship between the older individual (*i*) and the parents of the younger individual (*j*):

|  |  |
| --- | --- |
| for | [8] |

In Eq. 8, and refer to the parents of . The kinship between an individual and itself (i.e., the diagonal elements of the kinship matrix **K**) is calculated assuming the genes are sampled **with replacement**. This is closely related to the inbreeding coefficient *F*, which is the probability that the two genes (sampling **without replacement**) are IBD:

|  |  |
| --- | --- |
|  | [9] |

The second equality in Eq. 9 follows because the inbreeding coefficient equals the kinship between the parents. Using the result *A* = 2*K* derived earlier, recursive formulas for the additive relationship follow:

|  |  |
| --- | --- |
| for | [10] |

Eq. 10 assumes that all individuals are represented in the pedigree. However, in plant breeding, inbred lines are often derived by selfing following sexual hybridization, and the intermediate generations are not explicitly represented in the pedigree. The recursive formula for the additive relationship between two different individuals still holds in this case, but the diagonal elements of the additive relationship matrix need to be modified. For Sz individuals, which have undergone *z* generations of selfing relative to the F1 = S0 generation,

|  |  |
| --- | --- |
|  | [11] |

Example

Arrows drawn from parent to offspring



Assumptions

A11 = A22 = A33 = 1 (founders are outbred)

A12 = A13 = A23 = 0 (founders are unrelated)

Parent-Offspring Relationship:

A14 = ½ (A11 + A12) = ½

Full-Sib Relationship:

A45 = ½ (A41 + A42) = ½ (½ + ½) = ½

Half-Sib Relationship:

A56 = ½ (A52 + A53) = ½ (½ + 0) = ¼

Diagonal Elements:

A44 = 1 + ½ A12 = 1