2.4.6 Summary of Values

The following table summarizes all genotypic values all breeding values and the dominance deviations.

Genotyp	genotypic value	Breeding Value	Dominance Deviation
G_iG_j	V_{ij}	BV_{ij}	D_{ij}
G_1G_1	a	$2q\alpha$	$-2q^2d$
G_1G_2	d	$(q-p)\alpha$	$2pqd$
G_2G_2	$-a$	$-2p\alpha$	$-2p^2d$

The formulas in the above shown table assume that G_1 is the favorable allele with frequency $f(G_1) = p$. The allele frequency of G_2 is $f(G_2) = q$. Since we have a bi-allelic locus $p + q = 1$.

Based on the definition of dominance deviation, the genotypic values V_{ij} can be decomposed into the components population mean (μ), breeding value (BV_{ij}) and dominance deviation (D_{ij}) according to equation (2.16).

$$V_{ij} = \mu + BV_{ij} + D_{ij} \tag{2.16}$$

Taking expected values on both sides of equation (2.16) and knowing that the population mean μ was defined as the expected value of the genotypic values in the population, i.e. $E[V] = \mu$, it follows that the expected values of both the breeding values and the dominance deviations must be 0. More formally, we have

$$\begin{aligned} E[V] &= E[\mu + BV + D] \\ &= E[\mu] + E[BV] + E[D] \\ &= \mu \end{aligned} \tag{2.17}$$

From the last line in equation (2.17), it follows that $E[BV] = E[D] = 0$. This also shows that both breeding values and dominance deviations are defined as deviation from a given mean.

2.5 Variances

The population mean μ and derived from that the breeding values were defined as expected values. Their main purpose is to assess the state of a given population with respect to a certain genetic locus and its effect on a phenotypic trait of interest. One of our primary goals in livestock breeding is to improve the populations at the genetic level through the means of selection and mating. Selection of potential parents that produce offspring that are closer to our breeding goals is only possible, if the selection candidates show a certain level of variation in the traits that we are interested in. In populations where there is no variation which means that all individuals are exactly at the same level, it is not possible to select potential parents for the next generation.

In statistics the measure that is most often used to assess variation in a certain population is called **variance**. For any given discrete random variable X the variance is defined as the second central moment of X which is computed as shown in equation (2.18).

$$Var[X] = \sum_{x_i \in \mathcal{X}} (x_i - \mu_X)^2 * f(x_i) \tag{2.18}$$

where	$\mathcal{X}$:	set of all possible x-values
	$f(x_i)$	probability that x assumes the value of x_i
	μ_X	expected value $E[X]$ of X

In this section we will be focusing on separating the obtained variances into different components according to their causative sources. Applying the definition of variance given in equation (2.18) to the genotypic values V_{ij}, we obtain the following expression.

$$\begin{aligned}\sigma_G^2 = Var[V] = &(V_{11} - \mu)^2 * f(G_1G_1) \\ &+ (V_{12} - \mu)^2 * f(G_1G_2) \\ &+ (V_{22} - \mu)^2 * f(G_2G_2)\end{aligned} \tag{2.19}$$

where $\mu = (p-q)a + 2pqd$ the population mean.

Based on the decomposition of the genotypic value V_{ij} given in (2.16), the difference between V_{ij} and μ can be written as the sum of the breeding value and the dominance deviation. Then σ_G^2 can be written as

$$\begin{aligned}\sigma_G^2 = Var[V] = &(BV_{11} + D_{11})^2 * f(G_1G_1) \\ &+ (BV_{12} + D_{12})^2 * f(G_1G_2) \\ &+ (BV_{22} + D_{22})^2 * f(G_2G_2)\end{aligned} \tag{2.20}$$

Inserting the expressions for the breeding values BV_{ij} and for the dominance deviation D_{ij} found earlier and simplifying the equation leads to the result in (2.21). A more detailed derivation of σ_G^2 is given in the appendix (2.8) of this chapter.

$$\begin{aligned}\sigma_G^2 &= 2pq\alpha^2 + (2pqd)^2 \\ &= \sigma_A^2 + \sigma_D^2\end{aligned} \tag{2.21}$$

The formula in equation (2.21) shows that σ_G^2 consists of two components. The first component σ_A^2 is called the **genetic additive variance** and the second component σ_D^2 is termed **dominance variance**. As shown in equation (2.25) σ_A^2 corresponds to the variance of the breeding values. Because we have already seen that the breeding values are additive in the number of favorable alleles, σ_A^2 is called genetic additive variance. Because σ_D^2 corresponds to the variance of the dominance deviation effects (see equation (2.27)) it is called dominance variance.

2.6 Extension To More Loci

When only a single locus is considered, the genotypic values (V_{ij}) can be decomposed according to equation (2.16) into population mean, breeding value and dominance deviation. When a genotype refers to more than one locus, the genotypic value may contain an additional deviation caused by non-additive combination effects.

2.6.1 Epistatic Interaction

Let V_A be the genotypic value of locus A and V_B denote the genotypic value of a second locus B, then the total aggregate genotypic value V attributed to both loci A and B can be written as

$$V = V_A + V_B + I_{AB} \tag{2.22}$$

where I_{AB} is the deviation from additive combination of these genotypic values. When computing the population mean earlier in this chapter, we assumed that I was zero for all combinations of genotypes. If I is not zero for any combination of genes at different loci, those genes are said to **interact** with each other or to exhibit **epistasis**. The deviation I is called interaction deviation or epistatic deviation. If I is zero, the genes are called to act additively between loci. Hence *additive action* may mean different things. When referring to one locus, it means absence of dominance. When referring to different loci, it means absence of epistasis.

Interaction between loci may occur between pairs or between higher numbers of different loci. The complex nature of higher order interactions, i.e., interactions between higher number of loci does not need to concern us. Because in the aggregate genotypic value V, interaction deviations of all sorts are treated together in an overall interaction deviation I. This leads to the following generalized form of decomposing the overall aggregate genotype V for the case of multiple loci affecting a certain trait of interest.

$$V = \mu + A + D + I \tag{2.23}$$

where A is the sum of the breeding values attributable to the separate loci and D is the sum of all dominance deviations. For our purposes in livestock breeding where we want to assess the genetic potential of a selection candidate to be a parent of offspring forming the next generation, the **breeding value** is the most important quantity. The breeding value is of primary importance because a given parent passes a random sample of its alleles to its offspring. We have seen in section 2.4.4 that breeding values are additive in the number of favorable alleles. Hence the more favorable alleles a given parent passes to its offspring the higher the breeding value of this parent.

On the other hand, the dominance deviation measures the effect of a certain genotype occurring in an individual and the interaction deviation estimates the effects of combining different genotypes at different loci in the genome. But because parents do not pass complete genotypes nor do they pass stretches of DNA with several loci, but only a random collection of its alleles, it is really the breeding value that is of primary importance in assessing the genetic potential of a given selection candidate.

2.6.2 Interaction Variance

If genotypes at different loci show epistatic interaction effects as described in section 2.6.1, the interactions give rise to an additional variance component called V_I, which is the variance of interaction deviations. This new variance component V_I can be further decomposed into sub-components. The first level of sub-components is according to the number of loci that are considered. Two-way interactions involve two loci, three-way interactions consider three loci and in general n-way interactions arise from n different loci. The next level of subdivision is according to whether they include additive effects, dominance deviations or both.

In general it can be said that for practical purposes, interaction effects explain only a very small amount of the overall variation. As already mentioned in section @ref(#epistatic-interaction) for livestock breeding, we are mostly interested in the additive effects. This is also true when looking at the variance components. Hence dominance variance and variances of interaction deviations are not used very often in practical livestock breeding application.

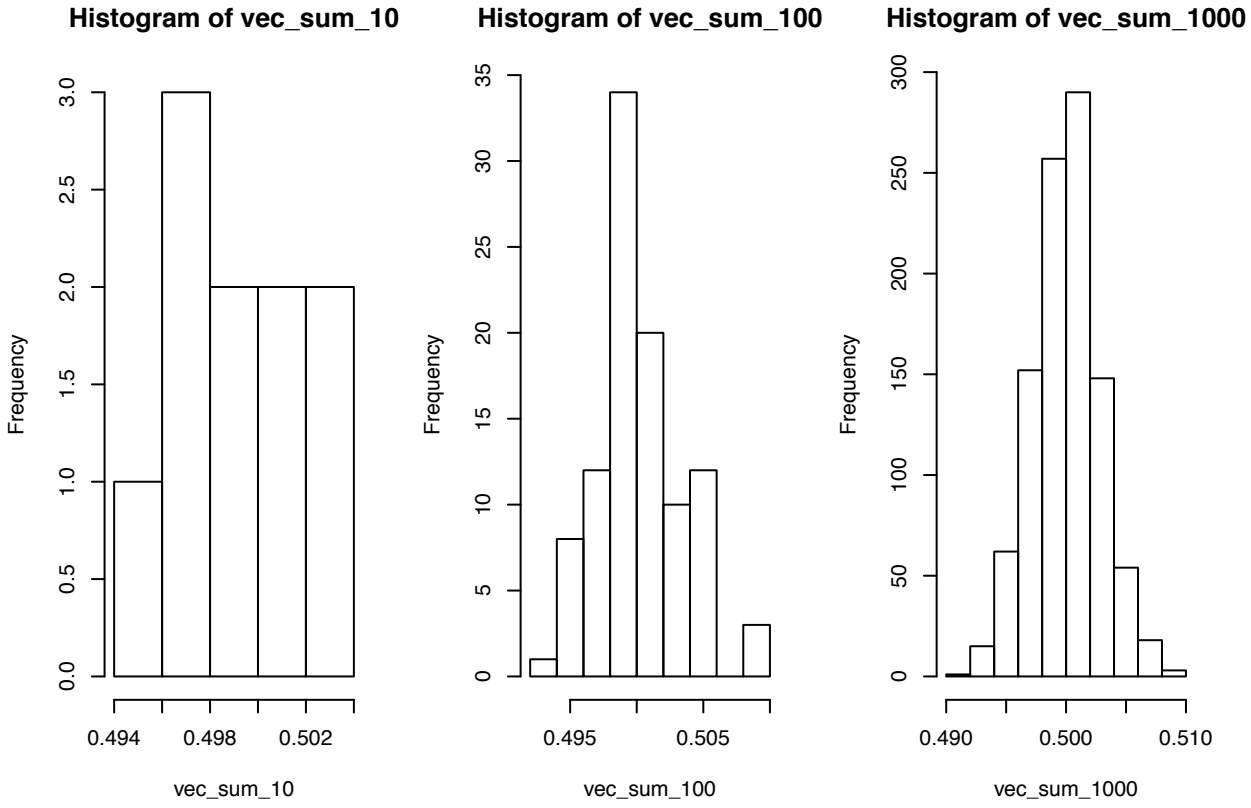

Figure 2.3: Distribution of Sums of Different Numbers of Components

2.7 Genetic Models

In this chapter, we have seen how to model the genetic basis of a quantitative trait when a single locus affects the trait of interest. We call this a single-locus model. When several loci have an effect on a certain trait, then we talk about a **polygenic model**. Letting the number of loci affecting a certain phenotype tend to infinity, the resulting model is called **infinitesimal model**.

From a statistical point of view, the breeding values in an infinitesimal model are considered as a random effect with a known distribution. Due to the central limit theorem, this distribution is assumed to be a normal distribution. The central limit theorem says that the distribution of any sum of a large number of very small effects converges to a normal distribution. For our case where a given trait of interest is thought to be influenced by a large number of genetic loci each having a small effect, the sum of the breeding values of all loci together can be approximated by a normal distribution. Figure (2.3) shows the distribution for a sum of 10, 100 and 1000 components each. The histograms show a better approximation to the normal distribution the larger the number of components considered in the sum.

2.7.1 Model Usage In Routine Evaluations

Traditional prediction of breeding values before the introduction of genomic selection is based on the infinitesimal model. When genomic selection was introduced which takes into account the information from a large number of loci, genomic breeding values are estimated using a polygenic model.

2.8 Appendix: Derivations

This section shows how the genetic variance in equation (2.21) is computed.

$$
\begin{aligned}
\sigma_G^2 &= (BV_{11} + D_{11})^2 * p^2 \\
&+ \ (BV_{12} + D_{12})^2 * 2pq \\
&+ \ (BV_{22} + D_{22})^2 * q^2 \\
&= \left(2q\alpha - 2q^2 d\right)^2 * p^2 \\
&+ \ ((q-p)\alpha + 2pqd)^2 * 2pq \\
&+ \ \left(-2p\alpha - 2p^2 d\right)^2 * q^2 \\
&= \left(4q^2\alpha^2 - 8q^3 d\alpha + 4q^4 d^2\right) * p^2 \\
&+ \ \left(q^2\alpha^2 - 2pq\alpha^2 + p^2\alpha^2 - 4(q-p)pqd\alpha + 4p^2q^2d^2\right) * 2pq \\
&+ \ \left(4p^2\alpha^2 + 8p^3 d\alpha + 4p^4\alpha^2\right) * q^2 \\
&= 4p^2q^2\alpha^2 - 8p^2q^3 d\alpha + 4p^2q^4 d^2 \\
&+ \ 2pq^3\alpha^2 - 4p^2q^2\alpha^2 + 2p^3q\alpha^2 \\
&- \ 8p^3q^2 d\alpha + 8p^2q^3 d\alpha + 8p^3q^3 d^2 \\
&+ \ 4p^2q^2\alpha^2 + 8p^3q^2 d\alpha + 4p^4q^2 d^2 \\
&= 4p^2q^2\alpha^2 + 4p^2q^4 d^2 \\
&+ \ 2pq^3\alpha^2 + 2p^3q\alpha^2 \\
&+ \ 8p^3q^3 d^2 \\
&+ \ 4p^4q^2 d^2 \\
&= 2pq\alpha^2\left(p^2 + 2pq + q^2\right) \\
&+ \ (2pqd)^2\left(p^2 + 2pq + q^2\right) \\
&= 2pq\alpha^2 + (2pqd)^2 \\
&= \sigma_A^2 + \sigma_D^2
\end{aligned}
\tag{2.24}
$$

From the last two lines of (2.24) it follows that $\sigma_A^2 = 2pq\alpha^2$ and $\sigma_D^2 = (2pqd)^2$. It can be shown that σ_A^2 corresponds to the squared breeding values times the associated genotype frequencies. Because the expected values of the breeding values is zero, σ_A^2 is equivalent to the variance of the breeding values.

$$
\begin{aligned}
\sigma_A^2 &= Var\left[BV\right] = (BV_{11} - E\left[BV\right])^2 * f(G_1G_1) + (BV_{12} - E\left[BV\right])^2 * f(G_1G_2) + (BV_{22} - E\left[BV\right])^2 * f(G_2G_2) \\
&= BV_{11}^2 * f(G_1G_1) + BV_{12}^2 * f(G_1G_2) + BV_{22}^2 * f(G_2G_2) \\
&= (2q\alpha)^2 * p^2 + ((q-p)\alpha)^2 * 2pq + (-2p\alpha)^2 * q^2 \\
&= 4p^2q^2\alpha^2 + \left(q^2\alpha^2 - 2pq\alpha^2 + p^2\alpha^2\right) * 2pq + 4p^2q^2\alpha^2 \\
&= 8p^2q^2\alpha^2 + 2pq^3\alpha^2 - 4p^2q^2\alpha^2 + 2p^3q\alpha^2 \\
&= 4p^2q^2\alpha^2 + 2pq^3\alpha^2 + 2p^3q\alpha^2 \\
&= 2pq\alpha^2\left(2pq + q^2 + p^2\right) \\
&= 2pq\alpha^2
\end{aligned}
\tag{2.25}
$$

In the above derivation in (2.25) of the variance of the breeding values, we were using the fact that the expected value $E\left[BV\right] = 0$. This can be shown more formally as follows

$$\begin{aligned} E[BV] &= BV_{11} * f(G_1G_1) + BV_{12} * f(G_1G_2) + BV_{22} * f(G_2G_2) \\ &= 2q\alpha * p^2 + (q-p)\alpha * 2pq + (-2p\alpha) * q^2 \\ &= 2p^2q\alpha + 2pq^2\alpha - 2p^2q\alpha - 2pq^2\alpha \\ &= 0 \end{aligned} \tag{2.26}$$

Similarly to (2.25) we can show that σ_D^2 corresponds to the squared dominance deviations times the frequencies of the corresponding genotypes. That is the reason why σ_D^2 is called dominance variance.

$$\begin{aligned} \sigma_D^2 &= D_{11}^2 * f(G_1G_1) + D_{12}^2 * f(G_1G_2) + D_{22}^2 * f(G_2G_2) \\ &= (-2q^2d)^2 * p^2 + (2pqd)^2 * 2pq + (-2p^2d)^2 * q^2 \\ &= 4p^2q^4d^2 + 8p^3q^3d^2 + 4p^4q^2d^2 \\ &= 4p^2q^2d^2\left(q^2 + 2pq + p^2\right) \\ &= 4p^2q^2d^2 \end{aligned} \tag{2.27}$$