

## Chapter 6

# Variance and Inbreeding

Recalling from chapter 5 the variance ( $var(u_i)$ ) of a breeding value  $u_i$  of animal  $i$  is given by

$$var(u_i) = (1 + F_i)\sigma_u^2 \quad (6.1)$$

where  $F_i$  is the inbreeding coefficient of animal  $i$  and  $\sigma_u^2$  corresponds to the additive genetic variance. At first sight this seems difficult to understand why the inbreeding coefficient increases the variance of a breeding value. This chapter aims at explaining the relationship between inbreeding and the genetic variance. The material presented here is based on chapters 3 and 14 of (Falconer and Mackay, 1996).

### 6.1 Inbreeding

Inbreeding means mating related individuals. The degree of relationship between individuals in a population depends on the size of the population. In a population of bisexual organisms every individual has  $2^t$  ancestors when looking  $t$  generations back. Already for small  $t$  the number of required individuals in a population becomes quite large in order to provide separate unrelated ancestors. As a consequence of that any pair of individuals must be related to some degree. Furthermore, pairs mating at random are expected to be more related in smaller populations compared to when individuals mate at random in large populations. Therefore properties of small populations can be treated as consequences of inbreeding.

An important consequence of two individuals having a common ancestor is that they may both carry replicates of one of the alleles present in the common ancestor. If the two individuals mate, they may pass on these replicates to their offspring.

### 6.1.1 Inbreeding in Idealized Population

The coefficient of inbreeding is deduced assuming an idealized population. Starting with the base population and its progeny forming generation 1 the development of the inbreeding coefficient is computed. What is meant by the term **idealized population** is shown in Figure 6.1.

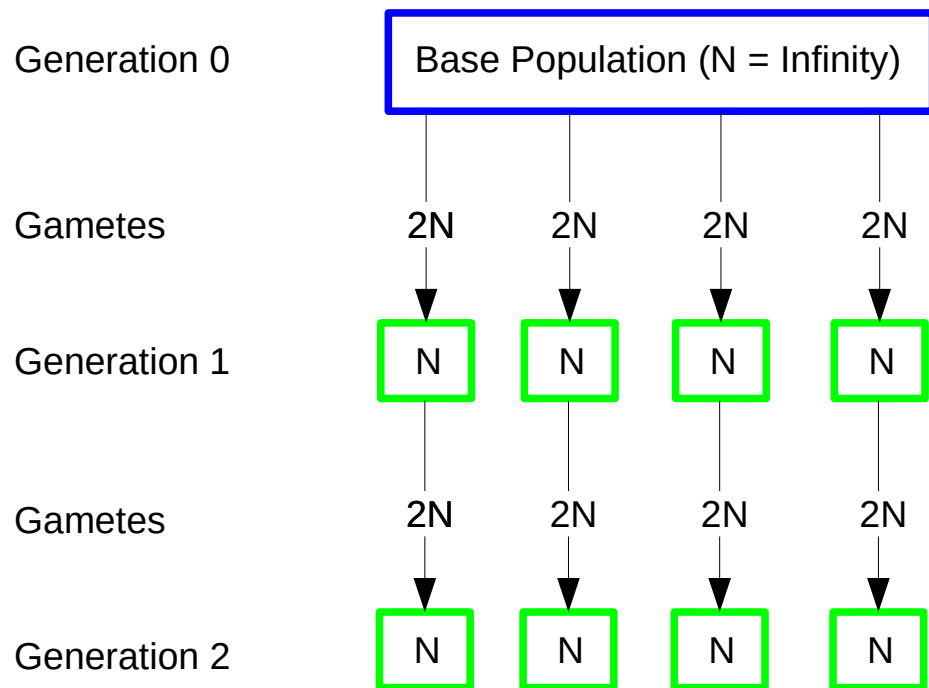


Figure 6.1: Subdivision of a single large population into a number of subpopulations or lines

The computation of the inbreeding coefficient may be visualized by the following situation. Let us assume a hermaphrodite marine organism, capable of self-fertilization shedding eggs and sperm into the sea. There are  $N$  individuals, each shedding equal numbers of gametes which mate at random. At a given locus, all the alleles in the base population have to be regarded as non-identical. For that single locus, among the gametes shed by the base population there

are  $2N$  different sorts in equal number. What is the probability that a pair of gametes taken at random carry identical alleles? This probability corresponds to the **inbreeding coefficient** ( $F$ ). Any gamete has a chance of  $1/(2N)$  to mate with a gamete carrying the same allele. Hence the inbreeding coefficient ( $F_1$ ) in generation 1 corresponds to

$$F_1 = \frac{1}{2N} \quad (6.2)$$

In generation 2 there are two ways in which identical homozygotes can arise, first from new replication of alleles and second from previous replications. The probability of newly replicated alleles coming together in a new zygote is again  $1/(2N)$ . The remaining proportion  $1 - (1/(2N))$  of zygotes carries alleles that are not identical, but may have been identical from the previous generation. The total probability of identical zygotes in generation 2 is

$$F_2 = \frac{1}{2N} + (1 - \frac{1}{2N}) * F_1 \quad (6.3)$$

The same argument leading to equation (6.3) applies to any subsequent generations. We can therefore write the more general statement

$$F_t = \frac{1}{2N} + (1 - \frac{1}{2N}) * F_{t-1} \quad (6.4)$$

Thus the inbreeding coefficient given in (6.4) consists of two parts: first an increment ( $1/(2N)$ ) attributable to new inbreeding and a remainder that is caused by inbreeding of previous generations. The increment ( $1/(2N)$ ) is assigned to a new variable  $\Delta F$ , so that

$$\Delta F = \frac{1}{2N} \quad (6.5)$$

With that equation (6.4) can be re-written as

$$F_t = \Delta F + (1 - \Delta F) * F_{t-1} \quad (6.6)$$

Solving (6.6) for  $\Delta F$  results in

$$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}} \quad (6.7)$$

The measure of the **rate of inbreeding** given in equation (6.7) provides a convenient way of generalising the concept of inbreeding beyond the simplifications of the idealized population. This generalization provides a means of

comparing inbreeding effects of different breeding systems. When expressing inbreeding in terms of  $\Delta F$ , equation (6.4) is valid for any breeding system and is not restricted to the idealized population where  $\Delta F$  is set to  $1/(2N)$ . So far, we have just related the inbreeding coefficient in one generation to the previous generation. It remains to express the inbreeding coefficient in terms of a set of properties of the base population. This is simplified by defining the symbol  $P$  as the complement of the inbreeding coefficient  $F$ , hence

$$P = 1 - F \quad (6.8)$$

The quantity symbolized by  $P$  is known as the **panmictic index**. Using (6.8) and inserting it into (6.7) leads to

$$\frac{P_t}{P_{t-1}} = 1 - \Delta F \quad (6.9)$$

Hence the rate at which  $P$  increases from one generation to the next is reduced to a constant  $1 - \Delta F$ . Going back  $t$  generations to the base population leads to

$$P_t = (1 - \Delta F)^t * P_0 \quad (6.10)$$

In the base population, we assumed no inbreeding, hence  $F_0 = 0$  and  $P_0 = 1$ . Using the result of (6.10) to compute  $F_t$  leads to

$$F_t = 1 - (1 - \Delta F)^t \quad (6.11)$$

### 6.1.2 Variance of Gene Frequency

According to the Hardy-Weinberg Equilibrium, gene frequencies are constant over generations. But this is only true, if the base population is not divided into subpopulations or lines. If the base population is split into separate lines as shown in Figure 6.1, the gene frequencies in the single lines start to show variation. The amount of the variation is quantified by the variance of the gene frequencies.

The variance ( $\sigma_{\Delta q}^2$ ) of the change of gene frequency in one generation is first of all the variance of a binomial random variable and can be expressed in terms of the rate of inbreeding, as shown below.

$$\sigma_{\Delta q}^2 = \frac{p_0 q_0}{2N} = p_0 q_0 \Delta F \quad (6.12)$$

An equivalent way of writing (6.13) is in terms of the inbreeding coefficient ( $F_1$ ) and the variance ( $\sigma_q^2$ ) of gene frequencies after one generation. It follows that

the relationship is the same after any number of generations, so that after  $t$  generations

$$\sigma_q^2 = p_0 q_0 F_t \quad (6.13)$$

### 6.1.3 Genotype Frequencies

The genotype frequencies in the population as a whole (across all generations) can be deduced from the knowledge of the variance of gene frequencies. If an allele has frequency  $q$  in a given line, homozygotes of that allele have frequency  $q^2$  in that line. The frequency of the homozygotes in the complete population over all lines will be the mean value of  $q^2$  across all lines. The mean frequency of homozygotes is written as  $\bar{q}^2$ . The value of  $\bar{q}^2$  is obtained by the knowledge of the variance of gene frequencies. In general the variance of a series of observations is obtained by

From the general formula of obtaining the variance of a set of observations corresponding to

$$\sigma_q^2 = (\bar{q}^2) - \bar{q}^2 \quad (6.14)$$

the mean frequency of homozygotes  $\bar{q}^2$  is obtained as

$$\bar{q}^2 = \sigma_q^2 + \bar{q}^2 \quad (6.15)$$

where  $\bar{q}$  is the mean gene frequency which is the same as the original gene frequency  $q_0$ . Thus in the complete population, the frequency of the homozygotes of a particular allele increases and is always in excess of the original frequency by an amount equal to the variance of the gene frequency among the lines. In a two-allele system, the same applies to the other allele. The genotypic frequencies for a locus with two alleles can then be summarized as shown in Table 6.1.

The genotype frequencies shown in Table 6.1 are no longer in Hardy-Weinberg equilibrium. This change in genotype frequencies is the result of a mechanism which is called the **dispersive process**. The dispersive process is active as soon

Table 6.1: Genotype Frequencies in Population as a Whole

Genotype	Frequency
$A_1 A_1$	$p_0^2 + \sigma_q^2$
$A_1 A_2$	$2p_0 q_0 - 2\sigma_q^2$
$A_2 A_2$	$q_0^2 + \sigma_q^2$

as the idealized base population is subdivided into single lines. The increase of the frequency of the homozygous genotypes is the source of a phenomenon called **inbreeding depression**. This depression refers to the reduced fitness of individuals in populations with increasing levels of inbreeding.

Combining the formulas (6.13) and (6.15) and furthermore dropping the subscript  $t$  in  $F_t$  leads to

$$\bar{q}^2 = \bar{q}_0^2 + \sigma_q^2 = \bar{q}_0^2 + p_0q_0F \quad (6.16)$$

Based on (6.16) Table 6.1 with the genotype frequencies can be re-written as shown in Table 6.2 where genotype frequencies are now expressed in terms of the inbreeding coefficient  $F$ .

## 6.2 Changes of Mean Value

So far, we have explained the consequences of inbreeding on the genotype frequencies. In this section, we have a look at how inbreeding affects the mean values of metric characters. The most important consequence of inbreeding is the reduction of the mean phenotypic value of characters connected to reproduction and fitness. This phenomenon is known as **inbreeding depression**. In saying that a certain trait shows inbreeding depression, we refer to the average change of mean value in a number of lines. The separate lines are commonly found to differ to a greater or lesser extent in the change they show, as indeed, we should expect in consequence of random drift of gene frequencies. The change of mean value that we discuss now refers to changes of the mean value of a number of lines derived from a common base population.

Consider a population, subdivided into a number of lines, with a coefficient of inbreeding  $F$ . Table 6.3 shows the genotype frequencies derived earlier, the values of the single genotypes and the contribution to the population mean for each genotype. The allele frequencies  $\bar{p}$  and  $\bar{q}$  correspond to the frequencies in the whole population.

Summing over the last column in Table 6.3 results in the population mean for the given trait.

Table 6.2: Genotype Frequencies for a bi-allelic locus, expressed in terms of inbreeding coefficient  $F$

Genotype	Original Frequencies	Changes due to inbreeding
$A_1A_1$	$p_0^2$	$+p_0q_0F$
$A_1A_2$	$2p_0q_0$	$-2p_0q_0F$
$A_2A_2$	$q_0^2$	$+p_0q_0F$

Table 6.3: Derivation of Inbreeding Depression

Genotype	Frequency	Value	Product
$A_1A_1$	$\bar{p}^2 + \bar{p}\bar{q}F$	$a$	$(\bar{p}^2 + \bar{p}\bar{q}F)a$
$A_1A_2$	$2\bar{p}\bar{q} - 2\bar{p}\bar{q}F$	$d$	$(2\bar{p}\bar{q} - 2\bar{p}\bar{q}F)d$
$A_2A_2$	$\bar{q}^2 + \bar{p}\bar{q}F$	$-a$	$-(\bar{q}^2 + \bar{p}\bar{q}F)a$

$$\begin{aligned}
M_F &= (\bar{p}^2 + \bar{p}\bar{q}F)a + (2\bar{p}\bar{q} - 2\bar{p}\bar{q}F)d - (\bar{q}^2 + \bar{p}\bar{q}F)a \\
&= a(\bar{p} - \bar{q}) + 2d\bar{p}\bar{q} - 2d\bar{p}\bar{q}F \\
&= a(\bar{p} - \bar{q}) + 2d\bar{p}\bar{q}(1 - F) \\
&= M_0 - 2d\bar{p}\bar{q}F
\end{aligned} \tag{6.17}$$

where  $M_0$  is the population mean before inbreeding. The change of mean resulting from inbreeding is  $2d\bar{p}\bar{q}F$ .

## 6.3 Changes of Variance

The effect of inbreeding on the genetic variance becomes apparent when again imagining the change of gene frequencies in different lines that emerge from a homogeneous base population. Within the different lines, the gene frequencies change to the dispersive process of random drift. This makes that over time some allele frequencies will tend towards 0 and frequencies of other alleles will tend towards 1. This tendency towards the extremes is different in the different lines. As a result in the populations, the similarity within lines increases, but between the lines the differences get bigger.

The subdivision of a population into lines introduces an additional component of variance, the between-line variance component. This means that the total genetic variance is re-distributed into the within-line component and the between line component.

### 6.3.1 Redistribution of Genetic Variance

For reasons of simplicity, we are currently just looking at genetic loci with purely additive effects. That means the dominance term  $d$  for such additive loci is 0. Strictly speaking, the results shown here apply only to traits with no non-additive variance. But still, these results serve as a good approximation to the effect of inbreeding on the genetic variance.

Consider first a single locus. When there is not dominance the genotypic variance in the base population is given by

$$V_G = V_A = 2p_0q_0a^2 \quad (6.18)$$

The variance within one given line is  $V_G = 2pqa^2$ , where  $p$  and  $q$  are the allele frequencies in that given line. The mean genetic variance ( $V_{\bar{G}}$ ) within the lines is

$$V_{\bar{G}} = 2(\bar{p}\bar{q})a^2 \quad (6.19)$$

where  $(\bar{p}\bar{q})$  is the mean value of  $pq$  over all lines. The term  $2(\bar{p}\bar{q})$  is the overall frequency of heterozygotes in the whole population, which, by Table 6.2, is equal to  $2p_0q_0(1 - F)$  where  $F$  is the inbreeding coefficient. Therefore

$$\begin{aligned} V_{\bar{G}} &= 2(\bar{p}\bar{q})a^2 \\ &= 2p_0q_0(1 - F)a^2 \\ &= V_G(1 - F) \end{aligned} \quad (6.20)$$

This remains true when summing the variances over all loci that affect a given trait. The within-line variance corresponds to the original variance times  $(1 - F)$ . As  $F$  approaches 1, the within-line variance tends toward 0.

Now consider the between-line variance. This is the variance of the true means of lines, and would be estimated from an analysis of variance as the between-line component. For a single locus with no dominance, the mean genotypic value of a line with allele frequencies  $p$  and  $q$  is obtained as

$$M = a(p - q) = a(1 - 2q) \quad (6.21)$$

Now we have to find the variance of  $M$ . The term  $a$  is a constant, meaning that it is the same in all the lines. Hence the only random term in  $M$  is the allele frequency  $q$ . Therefore

$$var(M) = \sigma_M^2 = 4a^2\sigma_q^2 = 4a^2p_0q_0F \quad (6.22)$$

Comparing the results of (6.22) and (6.20) shows that  $\sigma_M^2 = 2FV_G$ . Putting the two components together leads to the total genetic variance as shown in Table 6.4.

From the last row of Table 6.4, we can see that the total additive genetic variance in a population with inbreeding corresponds to  $(1 + F)V_G$  which is exactly what we wanted to show at the beginning of this chapter in equation (6.1).



Table 6.4: Partitioning of the genetic additive variance in a population with lines and a given inbreeding coefficient  $F$

Source	Variance
Between lines	$2FV_G$
Within lines	$(1 - F)V_G$
Total	$(1 + F)V_G$