5 THE GENETICS OF POPULATIONS

The fundamental idea of population genetics is that of a 'gene frequency'. Consider a population of diploid organisms, and suppose that at a particular locus there are two, and only two alleles, A and a, present in the population. Then there are three possible genotypes present in the population, AA, Aa and aa. In principle it would be possible to count these genotypes and find their proportions, $P\,AA : Q\,Aa : R\,aa$, where $P+Q+R = 1$. (In practice, it is usually difficult to distinguish the heterozygote Aa from one of the homozygotes AA or aa.) P, Q and R are then the genotype frequencies. From them we can calculate the gene frequencies $p\,A : q\,a$ as follows:

$$p = P + \tfrac{1}{2}Q; \quad q = \tfrac{1}{2}Q + R. \tag{5.1}$$

In so doing, we have 'counted' the A and a genes in the population, by allowing two A genes in an AA homozygote, two a genes in an aa homozygote, and one A and one a gene in an Aa heterozygote.

A. The Hardy–Weinberg ratio

The relations (5.1) hold for all 'autosomal' (i.e. not sex-linked) loci in a diploid, whatever the mating system. The relations enable us to calculate p and q if we know P, Q and R. Thus if a population consists of 60 % AA, 10 % Aa, and 30 % aa, then $p = 0{\cdot}65$ and $q = 0{\cdot}35$. But the reverse is not true; we cannot calculate P, Q and R merely from a knowledge of p and q.

It is however possible to find P, Q and R from p and q if we assume 'random mating'; that is, if we assume that the probability that an individual will mate with an AA, Aa or aa partner is independent of the genotype of the individual. Thus, if the frequences of genes A and a are p and q respectively, the probability that a child will inherit gene A from its father is p. The probability that a child will

inherit gene A from its mother is likewise p, and, if mating is random, is independent of whether the child also inherited gene A from its father.

Hence the probability that a child inherits gene A from both parents is p^2, and this is equal to P, the frequency of the AA genotype in the population.

This argument can be extended in a tabular form:

gene from father	gene from mother	genotype	frequency
A	A	AA	p^2
A	a	Aa	pq
a	A	Aa	pq
a	a	aa	q^2

Hence the 'Hardy–Weinberg' ratio, which states that if in a diploid population two allelic genes are present in the frequencies $p\,A : q\,a$, then random mating will give rise to zygotes with genotypes in the proportions $p^2\,AA : 2pq\,Aa : q^2\,aa$. This ratio is reached in a single generation of random mating, whatever the genotype frequencies in the parental population.

The Hardy–Weinberg ratio is widely assumed to be true in population genetics. The assumption is justified only if mating is random for the genotypes concerned. How are we to decide whether mating is random? In general we cannot. But if we can identify all three genotypes in a sample of a population, we can count them and see whether they agree with the Hardy–Weinberg ratio. If they do, this confirms that mating is random. For example, mating has been shown to be near enough random in this way for blood groups in man, and for black, ginger and tortoise-shell in London's cats. On the other hand, it is known that tall people tend to marry one another, and likewise short people, so mating is not random for genes affecting height. Finally, if we count genotypes in an adult population and find that they depart significantly from the Hardy–Weinberg ratio, this does not prove that mating is non-random; the discrepancy could equally well be caused by differential mortality.

B. Selection

Suppose that in a random-mating population there are two allelic genes A and a, A being dominant to a, and that the probabilities of survival from fertilised egg to breeding adult are:

$$\begin{aligned} &\text{for } AA \text{ and } Aa, && S \\ &\text{and for } aa, && S(1-k). \end{aligned}$$

Thus if k is positive, aa has a lower 'fitness' than AA or Aa. It is assumed that the fertilities of the three genotypes are the same. What will happen to such a population, and how rapidly will it happen? For simplicity, we will assume that generations are separate. The procedure is then to calculate the gene frequency in one generation in terms of the gene frequency at the same stage of the preceding one.

Let the frequency of A in the adults of the nth generation be p_n. Then with random mating, zygotes of the $(n+1)$th generation will be formed with the frequencies.

$$p_n^2\, AA : 2p_n(1-p_n)\, Aa : (1-p_n)^2\, aa.$$

The adults of the $(n+1)$th generation will then be in the relative proportions:

$$Sp_n^2\, AA : 2Sp_n(1-p_n)\, Aa : S(1-p_n)^2(1-k)\, aa. \qquad (5.2)$$

And p_{n+1}, the frequency of A genes in the adults of the $(n+1)$th generation, could then be found by 'counting' the A genes as a fraction of all genes in (5.2). However, this procedure leads to a rather clumsy expression for p_{n+1} in terms of p_n. It turns out that we get a neater expression if we work with $u_n = p_n/(1-p_n)$; i.e. the ratio of A genes to a genes in the nth generation.

Thus if we divide each term in (5.2) by $(1-p_n)^2$, we see that the adults in the $(n+1)$th generation are formed in the proportions

$$Su_n^2\, AA : 2Su_n\, Aa : S(1-k)\, aa,$$

and hence
$$u_{n+1} = \frac{2Su_n^2 + 2Su_n}{2Su_n + 2S(1-k)} = \frac{u_n(u_n+1)}{u_n+1-k}. \qquad (5.3)$$

This is a recurrence relation which enables us to calculate the

frequency of A genes in the $(n+1)$th generation in terms of the frequency in the nth generation. Knowing the initial frequency of gene A, and the 'selective disadvantage' k of aa, we could calculate the frequency of gene A in any subsequent generation by numerical iteration.

It would be convenient to find an analytical solution of (5.3), so that we could find for example u_{100} in terms of u_0 and k without the labour of 100 iterative steps. Unfortunately, this is not in general possible. However, it is possible to solve the important case when k

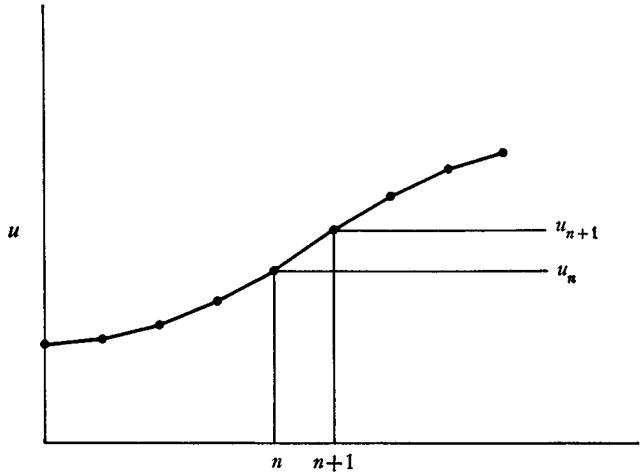

Fig. 21. The solution of a recurrence relation which could safely be replaced by a differential equation.

is small (say 0·01 or less). This we do by turning (5.3) into a differential equation—a procedure which is only justified when the change in u_n in one generation is not greatly different from the change in the preceding and in the following generation. Thus fig. 21 shows a possible graph of u_n against n which satisfies this condition. The graph consisting of a series of straight segments can safely be replaced by a continuously curving one whose slope at generation n is given by

$$\frac{u_{n+1}-u_n}{(n+1)-n} = u_{n+1}-u_n;$$

i.e. by a graph whose equation is

$$\frac{du_n}{dn} = u_{n+1} - u_n.$$

We will therefore replace equation (5.3) by

$$\frac{du_n}{dn} = u_{n+1} - u_n = \frac{ku_n}{u_n + 1 - k}. \tag{5.4}$$

Before attempting to solve (5.4), we note that if k is small, $u_{n+1} - u_n$ does not change rapidly with time, so it is safe to replace the recurrence relation by a differential equation. Also, if k is small, (5.4) can be replaced by

$$\frac{du_n}{dn} = \frac{ku_n}{u_n + 1}. \tag{5.5}$$

This is a differential equation with variables separate, so that

$$\int (1 + 1/u_n)\, du_n = k \int dn = kn.$$

$$\therefore \quad kn = [u_n + \ln u_n]_0^n$$

$$= u_n - u_0 + \ln \frac{u_n}{u_0}. \tag{5.6}$$

Thus suppose we have a recessive gene with a selective disadvantage of 1 % ($k = 0{\cdot}01$) and an initial frequency of 99·9 %. Then $u_0 = \dfrac{0{\cdot}001}{0{\cdot}999} \simeq 0{\cdot}001$, and hence

$$kn = u_n - 0{\cdot}001 + \ln \frac{u_n}{0{\cdot}001}. \tag{5.7}$$

From (5.7) we can calculate the number of generations taken for a given change in gene frequency, as shown in table 1.

Table 1

p_n	u_n	$\dfrac{u_n}{0{\cdot}001}$	$\ln \dfrac{u_n}{0{\cdot}001}$	kn	n
0·001	0·001	1	0	0	0
0·01	0·01	10	2·303	2·312	231
0·1	0·111	111	4·71	4·82	482
0·5	1·0	1,000	6·91	7·91	791
0·9	9	9,000	9·11	18·11	1811
0·99	99	99,000	11·52	110·52	11052
0·999	999	999,000	13·83	1012·83	101283

Thus with a 1 % advantage, an initially rare dominant gene will increase in frequency from 0·1 % to 10 % in 482 generations, from 10 to 90 % in 1329 generations, and will take almost one hundred thousand generations to increase from 90 to 99·9 %. The slowness with which the recessive is finally eliminated is due to the fact that a rare recessive is almost always present in heterozygotes, and so is not exposed to selection; for the same reason a rare but advantageous recessive gene increases in frequency very slowly.

c. Selection when all three genotypes have different fitnesses

If the fitnesses of the three genotypes have the relation

$$AA < Aa < aa,$$

then A will be replaced by a, as in the case when $AA = Aa < aa$, but at a greater rate, particularly when a is rare. But a novel type of behaviour arises when $AA < Aa > aa$; it will now be shown that an equilibrium exists with both A and a present in the population.

Let p be the frequency of gene A in the adult breeding population in the nth generation, and let $q = 1-p$ be the frequency of a. Let the relative fitnesses of the three genotypes AA, Aa and aa be $1-K:1:1-k$.

Then if N zygotes are formed by random mating, we have:

genotype	number of zygotes	number of surviving adults in generation $(n+1)$
AA	Np^2	$Np^2(1-K)$
Aa	$2Npq$	$2Npq$
aa	Nq^2	$Nq^2(1-k)$
Total	N	$N(1-Kp^2-kq^2)$.

Then the number of A genes in the adult population is

$$\begin{aligned} 2Np^2(1-K)+2Npq &= 2Np(p-pK+q) \\ &= 2Np(1-pK). \end{aligned}$$

Hence if p' is the frequency of gene A in the $n+1$th generation,

$$p' = \frac{2Np(1-pK)}{2N(1-Kp^2-kq^2)}. \qquad (5.8)$$

Now if the population is in equilibrium, p does not change from generation to generation; i.e. $p = p'$, and so at equilibrium either

$$p = 0 \quad \text{or} \quad 1 - Kp^2 - kq^2 = 1 - pK.$$

Substituting $q = 1 - p$, and collecting terms in p^2 and p, this becomes

$$(K+k)p^2 - (K+2k)p + k = 0,$$

$$\therefore \quad (p-1)[(K+k)p - k] = 0,$$

or at equilibrium $p = 0$ or 1 or $\dfrac{k}{K+k}$. (5.9)

The equilibria at $p = 0$ and 1 are trivial, but that at $p = k/(K+k)$ is of great interest. An equilibrium is only meaningful if p lies between 0 and 1, and this requires that K and k have the same sign. Thus an equilibrium exists if the heterozygote is the fittest of the three genotypes, (K and k positive), or if the heterozygote is the least fit (K and k negative).

Are these equilibria stable? We can get a preliminary answer by asking how the gene frequency will change when A is rare, and when a is rare. Consider the case when K and k are positive. When A is rare, most A genes occur in heterozygotes (i.e. Aa is much commoner than AA). Then since k is positive, Aa is fitter than aa, and since AA is too rare to influence the result, A will increase in frequency. Similarly, when a is rare it will increase in frequency if K is positive.

It follows that when K and k are positive (i.e. the heterozygote is the fittest of the three genotypes) the equilibrium is stable; similarly, when K and k are negative, the equilibrium is unstable.

But this argument does not tell us whether, when K and k are positive, p will oscillate about its equilibrium value.

To settle this question we investigate small departures from the equilibrium.

In the nth generation, let $p = k/(K+k) + \delta$, where δ is a small departure from the equilibrium, and let $p' = k/(K+k) + \delta'$.

These values can be substituted in equation (5.8). After some algebraic manipulation, the resulting equation reduces to

$$\frac{k}{K+k} + \delta' = \left[\frac{k}{K+k} + \delta\right]\left[1 - K\delta \Big/ \left(1 - \frac{Kk}{K+k}\right)\right].$$

This can be further simplified if we remember that since δ is small, terms in δ^2 can be ignored. Hence

$$\delta' = \frac{K+k-2Kk}{K+k-Kk}\delta. \tag{5.10}$$

Let $\frac{K+k-2Kk}{K+k-Kk} = R$. Then if R lies between 0 and 1, the equilibrium is stable and non-oscillatory. In fact, when the heterozygote is the fittest of the three genotypes these conditions are satisfied. In this case, K and k are positive, but both must lie between 0 and 1, since if for example $K > 1$, then $1-K$, the fitness of AA, would be negative, and negative fitnesses are meaningless. If K and k lie between 0 and 1, it is easy to verify that R also lies between 0 and 1.

If K and k are both negative, $R > 1$, and the equilibrium is unstable. If K and k have different signs, it has already been shown that no equilibrium exists.

D. The balance between selection and mutation

What will be the frequency of a gene which reduces fitness, but which is continuously reappearing by mutation?

This problem will first be solved for a deleterious dominant gene A, such that if the fitness of the 'normal' recessive homozygote aa is 1, the fitness of AA and Aa is $1-K$. K is then a number between 0 and 1.

The 'mutation rate' from a to A will be taken as μ. This has the following meaning. Every gene present in a zygote has arisen by a series of replications from a gene present in the zygote from which one of its parents developed. The mutation rate μ is the probability that a mutation has taken place in this time interval of one generation.

Knowing μ and K, we want to find p_E, the equilibrium frequency of gene A. Before proceeding in detail, it will help to outline the method to be adopted:

(i) We assume that in one generation of zygotes the frequency of A is p and of a is q, where $p+q = 1$.

(ii) We then calculate successively the effects on p of selection and of mutation, and hence find the value p' of the frequency of A in the next generation of zygotes, in terms of μ and K.

(iii) We then argue that at equilibrium there is no change in gene frequency, and hence that $p = p' = p_E$.

In the initial population of zygotes, if mating is random, the genotype frequencies are $p^2\,AA : 2pq\,Aa : q^2\,aa$. Thus if we start with a population of N zygotes and hence of $2N$ genes, we have after selection:

genotype	number after selection $\times 1/N$	gene	number after selection $\times 1/N$
AA	$p^2(1-K)$	A	$(2p^2+2pq)(1-K) = 2p(1-K)$
Aa	$2pq(1-K)$	a	$2q^2+2pq(1-K) = 2q-2pqK$
aa	q^2	Total	$2-2Kp(1+q)$
Total	$1-K(p^2+2pq)$		$=2-2K(p^2+2pq)$

Note that the totals check; the number of genes is twice the number of individuals.

Mutation does not alter the total number of genes, but increases the number of A genes by $\mu(2q-2pqK)$, and decreases the number of a genes by the same amount. Hence after selection and mutation the frequency of A is

$$p' = \frac{2p(1-K)+2\mu q(1-pK)}{2-2K(p^2+2pq)}$$

$$= \frac{p-pK+\mu q-\mu pqK}{1-Kp(p+2q)}. \qquad (5.11)$$

This expression can be greatly simplified if we make the assumption that p is very small. The assumption is justified because mutation rates are small (of the order of 10^{-5} or less) and hence genes which lower fitness can be maintained only at very low frequency by mutation.

If p is small, $1-Kp(p+2q) \simeq 1$, and μpqK in the numerator is small compared to μq. Hence (5.11) becomes

$$p' = p-pK+\mu q,$$

and at equilibrium

$$p_E = p_E - Kp_E + \mu(1-p_E),$$

or

$$p_E = \frac{\mu(1-p_E)}{K} \simeq \frac{\mu}{K}. \qquad (5.12)$$

This simple result could have been reached more quickly if it

had been assumed from the outset that p is small. Note that the assumption is justified provided that $\mu \ll K$.

Equation (5.12) sometimes enables us to determine mutation rates in human populations. Suppose we know F, the frequency at birth of an abnormality determined by a dominant gene A. Then provided that there is no difference in foetal mortality between normal and abnormal individuals, and that mating is random,

$$F = p^2 + 2pq \quad \text{or, if } p \text{ is small} \quad F \simeq 2p$$

and hence

$$\mu = pK = \frac{FK}{2}. \tag{5.13}$$

Now if the fitness of affected individuals is very low—either through sterility or because they die before reaching reproductive age—then $K \simeq 1$, and the mutation rate μ equals half the frequency of abnormal births.

In the case of a rare and harmful recessive gene a, with frequency p, if the relative fitnesses of AA, Aa and aa are $1:1:1-k$, it can be shown that at equilibrium

$$\mu = kp^2. \tag{5.14}$$

If we again assume random mating and no foetal mortality, the frequency of abnormal births is p^2. However, we cannot use equation (5.14) to estimate mutation rates from known frequencies of abnormal births, for two reasons:

(i) Equation (5.14) assumes random mating. But in human populations cousin marriages, and other marriages between relatives, occur frequently enough to have a large effect on the frequency of individuals homozygous for rare genes (see p. 86, example 6). This could perhaps be allowed for, but:

(ii) It is impossible to be sure that AA and Aa have the same fitness. If Aa is only slightly fitter than AA, this would keep gene a at a frequency considerably higher than could be maintained by mutation alone.

E. Inbreeding

If two cousins marry, they may at any locus transmit to a child a pair of genes both of which are derived by successive replications from the same individual gene in one of the grandparents they have

in common. In other words, two DNA molecules, each of which is a direct 'copy' of the same DNA molecule in an ancestor, may come together in a child.

In what follows, we will assume that two genes which are copies of the same gene in a recent ancestor are identical; that is, we will ignore the small probability that a mutation has occurred in the recent past. We will also assume that two genes at the same locus which are not copies of the same gene in a recent ancestor have a probability P_0 of being identical; in general, we do not know the value of P_0.

We will now define the 'coefficient of inbreeding' and the 'coefficient of parentage'.

I, the coefficient of inbreeding, is a property of an individual. It is the probability that, at any autosomal locus, the two genes present in that individual are identical.

R, the coefficient of parentage, is a property of two individuals. It is the probability that, at any autosomal locus, if one gene is drawn at random from each individual, those two genes will be identical. (For those already familiar with population genetics R corresponds to Malecot's '*coefficient de parente*' when $P_0 = 0$; I have used definitions of I and R which are unorthodox but which seem to me simpler to understand.)

It follows from these definitions that, if two individuals A and B have an offspring C, then

$$R_{AB} = I_C. \tag{5.15}$$

We will calculate first the value of I for individual C in the pedigree shown in fig. 22. C is in fact the offspring of a mating between half-sibs. Considering the two genes at any autosomal locus in C:

$$I_C = P_s + (1 - P_s) P_0,$$

where P_s is the probability that both genes in C are copies of the same gene in male G.

Consider the gene which C inherits from A: there is a probability of $\frac{1}{2}$ that this gene was inherited from G; if so, there is a probability of $\frac{1}{2}$ that G transmitted an identical gene to B; and if so, there is a probability of $\frac{1}{2}$ that B transmitted an identical gene to C.

Hence $P_s = (\frac{1}{2})^3$,

and $I_C = \frac{1}{8} + \frac{7}{8}P_0$.

The coefficient of parentage between half-sibs is likewise $\frac{1}{8} + \frac{7}{8}P_0$.

The argument whereby P_s was calculated can be generalised. It amounted to saying that a gene was certainly transmitted from A to C, and that for each of the steps $A-G$, $G-B$ and $B-C$, there

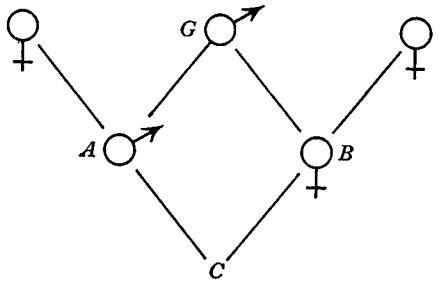

Fig. 22. Pedigree showing a mating between half sibs.

is a probability of $\frac{1}{2}$ that an identical gene was transmitted. The general proposition is then as follows:

If an individual C has parents A and B who have a common ancestor G, and if the 'loop' $C-A-X-Y-G-Z-B-C$ has n steps connecting parent and offspring, then

$$I_C = R_{AB} = (\tfrac{1}{2})^{n-1} + [1 - (\tfrac{1}{2})^{n-1}]P_0. \qquad (5.16)$$

What if there are two common ancestors, as in the pedigree in fig. 23, showing the mating of a bother and sister?

There are three mutually exclusive possibilities:

(i) the gene pair in C are copies of the same gene in G_1, with probability $(\frac{1}{2})^3 = \frac{1}{8}$;

(ii) the gene pair in C are copies of the same gene in G_2, with probability $\frac{1}{8}$;

(iii) the gene pair in C are not copies of a single gene in a recent ancestor, with probability $1 - \frac{1}{8} - \frac{1}{8} = \frac{3}{4}$.

Hence $I_C = R_{AB} = \frac{1}{8} + \frac{1}{8} + \frac{3}{4}P_0 = \frac{1}{4} + \frac{3}{4}P_0$.

We are now ready to tackle the problem of the rate of approach to homozygosity in a brother-sister mated line (fig. 24). Suppose that a male and female are selected from a large population to be the original parents (generation 0) of such an inbred line, and that at N loci the 4 genes present (2 in each parent) are all different.

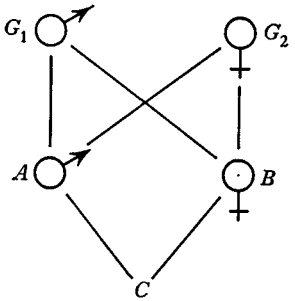

Fig. 23. Pedigree showing a mating between full sibs.

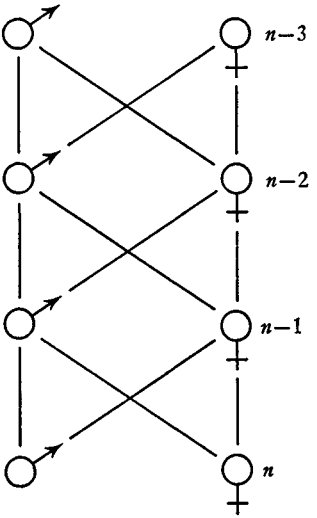

Fig. 24. Pedigree showing successive generations of brother-sister mating.

Then if for these loci I_n and R_n are the coefficients of inbreeding and parentage respectively in the nth generation, we have, remembering (5.15)

$$I_o = 0 \quad \text{and} \quad R_o = I_1 = 0. \tag{5.17}$$

(If we supposed that the original parents were homozygous at all loci, but homozygous for different alleles at N loci, then for these loci our initial conditions would be $I_o = 1$ and $R_o = I_1 = 0$.)

We want to find I_n, the proportion of N originally segregating loci for which an individual in the nth generation is homozygous; since $I_n = R_{n-1}$, this will also give a measure of the genetic similarity between members of the population.

Consider a locus in an individual in the nth generation. There are three mutually exclusive possibilities:

(i) the two genes are copies of the same gene in the same grandparent, with probability $\frac{1}{8}+\frac{1}{8}=\frac{1}{4}$; if so, the probability that they are identical is 1;

(ii) the two genes are copies of different genes in the same grandparent, with probability $\frac{1}{4}$; if so, the probability that they are identical is I_{n-2};

(iii) the two genes are copies of genes from different grandparents, with probability $\frac{1}{2}$; if so, the probability that they are identical is $R_{n-2}=I_{n-1}$.

Hence
$$I_n=\tfrac{1}{4}+\tfrac{1}{4}I_{n-2}+\tfrac{1}{2}I_{n-1},$$

or
$$4I_n=1+I_{n-2}+2I_{n-1}. \tag{5.18}$$

Before finding an analytical solution of (5.18), its behaviour can be investigated numerically. Thus from (5.17) $I_o=I_1=0$, and hence

$$I_2=\tfrac{1}{4}(1+0+0)=0{\cdot}25$$
$$I_3=\tfrac{1}{4}(1+0+0{\cdot}5)=0{\cdot}375$$
$$I_4=\tfrac{1}{4}(1+0{\cdot}25+0{\cdot}75)=0{\cdot}5$$
$$I_5=\tfrac{1}{4}(1+0{\cdot}375+1)=0{\cdot}594$$
$$I_6=\tfrac{1}{4}(1+0{\cdot}5+1{\cdot}188)=0{\cdot}672 \quad \text{and so on.}$$

Thus after 6 generations of brother-sister mating, approximately two thirds of the initially segregating loci would be homozygous.

In seeking an analytical solution of (5.18), we notice that it closely resembles the equation solved in appendix 4, to which the solution had the form $x=A\lambda_1^n+B\lambda_2^n$. This will not quite do in the present case, because of the constant term. However, a solution of the form

$$I_n=1+A\lambda_1^n+B_2^n$$

will work, because when this is substituted in (5.18), the constant terms cancel out, and we are left with the requirement that λ_1 and λ_2 satisfy the equation
$$4\lambda^2-2\lambda-1=0,$$

or
$$\lambda_1=+0{\cdot}808, \quad \lambda_2=-0{\cdot}308.$$

A and B can then be chosen to fit the initial conditions.

Thus if $I_0 = 0, \quad A + B = -1,$

and if $I_1 = 0, \quad 0{\cdot}808A - 0{\cdot}308B = -1,$

and hence $A = 1{\cdot}172 \quad \text{and} \quad B = 0{\cdot}172.$

Hence $$I_n = 1 - 1{\cdot}172 \times 0{\cdot}808^n + 0{\cdot}172(-0{\cdot}308)^n, \qquad (5.19)$$

and substituting various values of n:

generations	I_n	
6	0·672	as before
10	0·861	
50	$1 - 2{\cdot}75 \times 10^{-5}$.	

Examples

1 A *Drosophila* population cage is started by introducing 300 flies homozygous for the gene 'vestigial', 100 wild-type homozygotes, and 200 heterozygotes, each class consisting of equal number of males and females. Assuming random mating, write down the frequencies of different types of mating, and hence the proportions in which the three genotypes occur in the next generation. Check that these numbers agree with the Hardy–Weinberg ratio. What assumptions, other than random mating, have you made?

2 The ABO blood groups in man are determined by a system of 3 alleles, A, B and O. Genotypes AA and AO are group A, BB and BO are group B, AB is group AB, and OO is group O. The frequencies of the blood groups in England are 32·1 % A, 22·4 % B, 7·1 % AB and 38·4 % O. Are these proportions consistent with the assumption of random mating?

3 Homozygotes for a recessive gene r have a 2 % greater chance of survival than either R/R or R/r. The initial frequency of r in a random mating population is 1 per thousand. How many generations will elapse before the frequency of r reaches 50 %?

4 There is a gene S in man such that S/S individuals die soon after birth of anaemia. However, in areas where some $+/+$ individuals die of malaria as children, $S/+$ individuals never do. In an area of Africa, 10 % of adults are $S/+$. If this situation is

stable, what proportion of $+/+$ individuals die of malaria in childhood? (The facts, concerning sickle-cell anaemia, have been simplified for the sake of the example.)

5 Derive equation (5.14).

6 What is the coefficient of parentage of first cousins? A harmful recessive gene is present in a population with a frequency of 1/200. What is the frequency of homozygotes among the children of (*a*) unrelated parents, (*b*) first cousins?

7 An hermaphroditic organism reproduces by self-fertilisation. If an ancestral individual has a coefficient of inbreeding of I_0, what will be the coefficient of inbreeding n generations later?