from: H. Tuckwell: Elementary Applications of Probability Theory, publ. Chapman & Hall

8 Population genetics and Markov chains

8.1 GENES AND THEIR FREQUENCIES IN POPULATIONS

Chromosomes, which contain hereditary material, are located in the nuclei of the cells of living organisms. In *diploid* organisms, the chromosomes occur in pairs. For example, the nuclei of human cells contain 23 pairs of chromosomes, whereas those of dogs have 39 pairs. Sections of chromosomes that determine, by means of very complex chemical reactions, certain properties of an organism are called **genes**.

Genes may be in different forms at a given location, or **locus**, on a chromosome, and the different forms are called **alleles**. For example, in a diploid organism with two alleles labelled A_1, A_2 we may have the so called **genotypes** A_1A_1, A_1A_2, A_2A_1 or A_2A_2 as illustrated in Fig. 8.1. It is usually assumed that A_1A_2 and A_2A_1 give rise to the same properties. Individuals with both genes the same (A_1A_1 or A_2A_2) are called **homozygous**, whereas individuals with different genes (A_1A_2 or A_2A_1) are called **heterozygous**. Note that nearly all the cells of an organism have the same chromosomal structure.

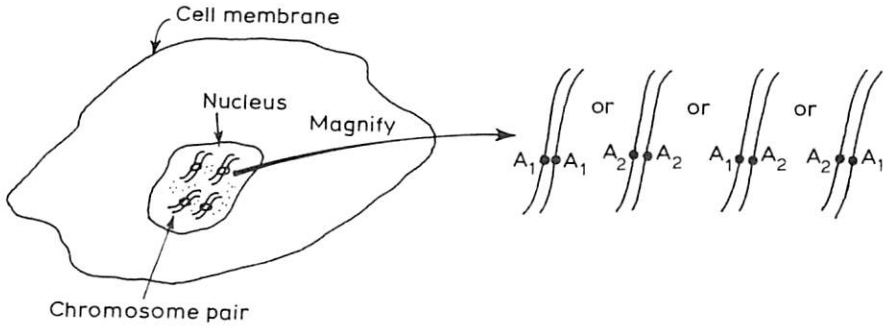

Figure 8.1 Schematic representation of a cell, its nucleus and chromosome pairs. On the right of the figure a chromosome pair is magnified to show possible genotypes at a particular locus.

In some reproductive processes (e.g. human) the chromosome pairs of the offspring contain one chromosome from each parent. Population genetics concerns itself with the numbers of genes of various types in populations, usually with a view to studying their changes from generation to generation.

Frequencies

Consider a population of N diploid individuals. At a particular locus there are a total of $2N$ genes. Let there be N_1 individuals of type A_1A_1, N_2 of type A_1A_2 or A_2A_1 and N_3 of type A_2A_2. The **genotype frequencies** are the fractions of the numbers of genotypes of each kind. We define

$$f = N_1/N$$
$$g = N_2/N$$
$$h = N_3/N,$$

and observe that

$$f + g + h = 1.$$

If we count the numbers of genes of each kind we see there are

$$p = 2N_1 + N_2$$

of type A_1, and

$$q = N_2 + 2N_3$$

of type A_2. We define the **gene (or allele) frequencies** as the fractions of the numbers of genes of each kind:

$$x = p/2N$$
$$y = q/2N.$$

We see that the relations

$$x = \frac{2N_1 + N_2}{2N} = f + g/2 \tag{8.1}$$

$$y = \frac{N_2 + 2N_3}{2N} = g/2 + h \tag{8.2}$$

$$x + y = 1,$$

must hold.

It should be noted that populations with different genotype frequencies may have the same gene frequencies. To illustrate consider a population of 20 individuals of which 10 are A_1A_1 and 10 are A_2A_2. Then the genotype

frequencies are

$$f = 1/2, \qquad g = 0, \qquad h = 1/2,$$

while the gene frequencies are

$$x = y = 1/2.$$

If on the other hand there are $5\ A_1A_1$, $10\ A_1A_2$ and $5\ A_2A_2$, the gene frequencies are the same, but the genotype frequencies are now

$$f = 1/4, \qquad g = 1/2, \qquad h = 1/4.$$

The factors affecting the evolution of gene frequencies

We will see in a simplified picture that in 'infinite populations' of individuals which mate randomly, the gene frequencies remain constant. This is contained in the **Hardy–Weinberg principle** which we will prove in the next section. In finite populations, however, there is so called **random drift** which leads eventually to the elimination of the heterozygous genotypes. The model we use to study this phenomenon is a Markov chain and we will discuss the general introductory theory of such processes. We will then see how the properties of the Markov chain are altered when the genes themselves may change from one form to another (**mutation**). The remaining forces of evolution, namely the selective advantages of some genes over others (**selection**) and the influx or efflux of individuals (**migration**) will not be discussed here. The reader may consult Crow and Kimura (1970) or Ewens (1979) for in-depth mathematical treatments. For a fascinating account of the basic biology and biochemistry the work of Watson (1970) will be a delight to read.

8.2 THE HARDY–WEINBERG PRINCIPLE

If a population is infinite we may interpret the genotype and gene frequencies as *bona fide* (actual) probabilities. We observe such a population as it evolves in time and let the genotype frequencies at generation n be f_n, g_n and h_n and the corresponding gene frequencies be x_n and y_n, where $n = 0, 1, 2, \ldots$. We then have the following.

Theorem 8.1 (Hardy–Weinberg principle) In an infinite, randomly mating diploid population, for genes at a single locus with two alleles, the gene frequencies do not change from generation to generation. Further, no matter what the initial genotype frequencies, the genotype frequencies at the first ($n = 1$) and subsequent generations are fixed and determined only by the initial gene frequencies.

Table 8.1 Possible matings, their probabilities and the conditional probabilities of various offspring genotypes ($A_1A_2 = A_1A_2$ or A_2A_1).

	Mating		$\Pr(M_i)$	$\Pr(A_1A_1 \mid M_i)$	$\Pr(A_1A_2 \mid M_i)$	$\Pr(A_2A_2 \mid M_i)$
	Male	Female				
M_1	A_1A_1	A_1A_1	f_0^2	1	0	0
M_2	A_1A_1	A_1A_2	f_0g_0	$\frac{1}{2}$	$\frac{1}{2}$	0
M_3	A_1A_1	A_2A_2	f_0h_0	0	1	0
M_4	A_1A_2	A_1A_1	f_0g_0	$\frac{1}{2}$	$\frac{1}{2}$	0
M_5	A_1A_2	A_1A_2	g_0^2	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
M_6	A_1A_2	A_2A_2	g_0h_0	0	$\frac{1}{2}$	$\frac{1}{2}$
M_7	A_2A_2	A_1A_1	f_0h_0	0	1	0
M_8	A_2A_2	A_1A_2	g_0h_0	0	$\frac{1}{2}$	$\frac{1}{2}$
M_9	A_2A_2	A_2A_2	h_0^2	0	0	1

In symbols this becomes

(a) $x_n = x_0, \quad y_n = y_0, \qquad n = 1, 2, 3, \ldots$

(b) $f_n = f_1, \quad g_n = g_1, \quad h_n = h_1, \qquad n = 2, 3, \ldots$

(c) f_1, g_1 **and** h_1 **depend only upon** x_0.

Proof In the population under consideration the various possible matings are shown in Table 8.1. Since both male and female parents may be any of three genotypes there are 9 possible combinations, denoted by M_i, $i = 1, 2, \ldots, 9$.

The probabilities of occurrence of the different matings are found as in the following example. The probability that an offspring of the first generation has an A_1A_1 male parent is f_0, and this is also the probability it has an A_1A_1 female parent. Hence the probability that a mating is of type M_1 is f_0^2. Similarly the remaining entries in the third column of Table 8.1 are found.

Now, given that the mating is of type M_1, the conditional probability of an A_1A_1 offspring is 1 etc. The matings M_1–M_9 are mutually exclusive and their union is the whole sample space. Hence by the law of total probability, the probability that an offspring in the first generation is A_1A_1 is

$$\begin{aligned} f_1 &= \Pr\{\text{member of first generation is } A_1A_1\} \\ &= \sum_{i=1}^{9} \Pr(\text{member of first generation is } A_1A_1 \mid M_i) \Pr(M_i) \\ &= f_0^2 + f_0g_0 + g_0^2/4 \\ &= (f_0 + g_0/2)^2 \end{aligned}$$

Using (8.1) we get

$$f_1 = x_0^2. \tag{8.3}$$

Similarly,

$$g_1 = \Pr(\text{member of first generation is } A_1A_2)$$

$$= f_0 g_0 + f_0 h_0 + g_0 h_0 + g_0^2/2$$
$$= 2(f_0 + g_0/2)(g_0/2 + h_0),$$

so that, from (8.1) and (8.2),

$$g_1 = 2x_0 y_0.$$

Also,

$$h_1 = \Pr(\text{member of first generation is } A_2A_2)$$
$$= g_0^2/4 + g_0 h_0 + h_0^2$$
$$= (g_0/2 + h_0)^2$$
$$= y_0^2.$$

Thus we have established that the genotype frequencies in any generation are completely determined by the gene frequencies in the previous generation, regardless of the genotype frequencies in the previous generation (part (c)).

Furthermore, the frequency of A_1 in the first generation is, from (8.1),

$$x_1 = f_1 + g_1/2$$
$$= x_0^2 + x_0 y_0$$
$$= x_0(x_0 + y_0)$$
$$= x_0.$$

Hence $x_1 = x_0$ and $y_1 = y_0$. Hence the gene frequencies in any generation must be the same as those in the preceding generation (part (a)). Part (b) follows because f_2 ($= x_1^2$ from (8.3)) is determined by x_1 and $x_1 = x_0$ and so on, for $f_3, f_4, \ldots$. This completes the proof of the above form of the Hardy–Weinberg principle.

Because the gene frequencies never change and because the genotype frequencies are constant from the first generation onwards, the population is said to be in **equilibrium** or **Hardy–Weinberg equilibrium** for the gene under consideration.

An example from human genetics

The following data obtained in an actual experimental study (see Strickberger, 1968, Chapter 30) lend support to the existence of Hardy–Weinberg equilibria in nature.

When human red blood cells are injected into the bloodstreams of rabbits, an immune reaction occurs (production of antibodies) in the rabbits. However, the blood from various humans leads to different reactions and the three different human genotypes MM, MN and NN can be classified. In a group of 104 North American Indians there were found to be 61 MM, 36 MN and 7 NN individuals.

Let x be the frequency of the M allele and let $y = 1 - x$ be the frequency of the N allele. If the gene under consideration is in Hardy–Weinberg equilibrium, the genotype frequencies for MM, MN and NN should be x^2, $2xy$ and y^2 respectively. We will calculate the gene frequencies and see if the genotype frequencies are as predicted by the Hardy–Weinberg formula.

From the data we find

$$x = (122 + 36)/208 = .7596$$

and $y = .2404$. Under the hypothesis of a Hardy–Weinberg equilibrium, the expected numbers of MM, MN and NN are obtained by multiplying 104 by the genotype frequencies x^2, $2xy$ and y^2. This gives 60.01, 37.98 and 6.009, respectively.

The value of the chi-squared statistic is

$$\chi^2 = \frac{(60.01 - 61)^2}{60.01} + \frac{(37.98 - 36)^2}{37.98} + \frac{(6.009 - 7)^2}{6.006}$$
$$= 0.283.$$

There are three terms in the sum but one degree of freedom is lost because we estimated x from the data, and another degree of freedom is lost because the numbers of genotypes must add to 104. Thus there is one degree of freedom for chi-squared and from tables we find $\Pr\{\chi_1^2 > 3.842\} = .05$. The observed value of chi-squared is safely less than the critical value at the .05 level of significance, lending strong support to the existence of a Hardy–Weinberg equilibrium.

8.3 RANDOM MATING IN FINITE POPULATIONS: A MARKOV CHAIN MODEL

In the previous section an infinite randomly mating diploid population was considered. The frequencies, or probabilities of occurrence of two alleles at a single locus were found to be constant.

We now wish to study the behaviour of gene frequencies in a finite population of N diploid individuals. Again we concentrate on a single locus with genotypes A_1A_1, A_1A_2 and A_2A_2. The total number of genes is fixed at $2N$ in all generations, it being assumed that the total population size is constant in time.

Notation

We introduce the following notation:

X_n = the number of A_1-genes in the nth generation, $n = 0, 1, 2, \ldots$

Thus there are $2N - X_n$ genes of type A_2 in generation n.

Random mating assumption

Randomness enters the model as follows. The $2N$ genes of any generation are chosen randomly from those in the previous generation in $2N$ Bernoulli trials in which the probability of a given gene (A_1 or A_2) is equal to its frequency in the previous generation.

Thus the number X_n of A_1-genes in generation n is a random variable and the whole sequence $X = \{X_0, X_1, X_2, \ldots\}$ is a **discrete-time random process.** Since the possible values of the X_n consist of the discrete set $\{0, 1, 2, \ldots, 2N\}$, X has a **discrete state space.** The process X is a **Markov chain.**

Transition probabilities

Suppose we are given that $X_n = j$. We ask, conditioned on this event, what is the probability that $X_{n+1} = k$. Since by the above random mating assumption, X_{n+1} is a binomial random variable with parameters $2N$ and $j/2N$, we have

$$\boxed{\Pr\{X_{n+1} = k | X_n = j\} = \binom{2N}{k}\left(\frac{j}{2N}\right)^k\left(1 - \frac{j}{2N}\right)^{2N-k}}, \qquad (8.4)$$

$$j, k = 0, 1, 2, \ldots, 2N.$$

This set of $(2N + 1)^2$ quantities is called the one-step transition probabilities. They can be arranged as a matrix $\mathbf{P}$ with elements

$$p_{jk} = \Pr\{X_{n+1} = k | X_n = j\}.$$

Before investigating the properties and behaviour of this genetical random process we give a brief introduction to the general theory of Markov chains.

8.4 GENERAL DESCRIPTION OF MARKOV CHAINS

Let $X \doteq \{X_n, n = 0, 1, 2, \ldots\}$ be a discrete-time random process with a discrete state space $\mathcal{S}$ whose elements are $s_1, s_2, \ldots$. We have seen that X is a Markov chain if for any $n \geqslant 0$, the probability that X_{n+1} takes on any value $s_k \in \mathcal{S}$ is conditional only on the value of X_n (and possibly n) but does not depend on the values of $X_{n-1}, X_{n-2}, \ldots$. This leads to the introduction of the one-time-step transition probabilities

$$p_{jk}(n) = \Pr\{X_{n+1} = s_k | X_n = s_j\}; \qquad j, k = 1, 2, \ldots, \qquad n = 0, 1, 2, \ldots \qquad (8.5)$$

We have allowed here for the possibility that the transition probabilities may depend on n. When they do not, they are called **stationary** and the process is referred to as a **temporally homogeneous** Markov chain. When they do depend on n, the term **nonhomogeneous** Markov chain is used. All the Markov chains we will consider later are temporally homogeneous.

Since X_0 is a random variable, which we refer to as the **initial value**, we introduce its probability distribution

$$p_j(0) = \Pr\{X_0 = s_j\}, \qquad j = 1, 2, \ldots \qquad (8.6)$$

We will now prove the following.

Theorem 8.2 The set of one-time-step transition probabilities (8.5) and the distribution of X_0 given by (8.6) completely determine the joint distribution of $\{X_0, X_1, \ldots, X_n\}$ for any $n \geqslant 1$.

Proof We will first prove this for $n = 1$ and $n = 2$.

$n = 1$.

We have, for any j, k, by definition of conditional probability,

$$\Pr\{X_1 = s_k | X_0 = s_j\} = \frac{\Pr\{X_0 = s_j, X_1 = s_k\}}{\Pr\{X_0 = s_j\}}.$$

On rearranging this,

$$\begin{aligned}\Pr\{X_0 = s_j, X_1 = s_k\} &= \Pr\{X_0 = s_j\}\Pr\{X_1 = s_k | X_0 = s_j\} \\ &= p_j(0)p_{jk}(1). \end{aligned} \qquad (8.7)$$

$n = 2$.

Again by definition of conditional probability

$$\Pr\{X_2 = s_l | X_1 = s_k, X_0 = s_j\} = \frac{\Pr\{X_0 = s_j, X_1 = s_k, X_2 = s_l\}}{\Pr\{X_1 = s_k, X_0 = s_j\}},$$

so

$$\begin{aligned}&\Pr\{X_0 = s_j, X_1 = s_k, X_2 = s_l\} \\ &\quad = \Pr\{X_1 = s_k, X_0 = s_j\}\Pr\{X_2 = s_l | X_1 = s_k, X_0 = s_j\}.\end{aligned}$$

But $\Pr\{X_2 = s_l | X_1 = s_k, X_0 = s_j\} = \Pr\{X_2 = s_l | X_1 = s_k\}$ by the **Markov property** and so, using (8.7) as well we get

$$\Pr\{X_0 = s_j, X_1 = s_k, X_2 = s_l\} = p_j(0)p_{jk}(1)p_{kl}(2).$$

This is generalized easily to $n > 2$ (see Exercise 2).

8.5 TEMPORALLY HOMOGENEOUS MARKOV CHAINS

If a Markov chain is temporally homogeneous and there are M possible states (i.e. possible values of X), then

$$p_{jk} = \Pr\{X_{n+1} = s_k | X_n = s_j\}; \qquad j, k = 1, 2, \ldots, M, \qquad (8.8)$$

regardless of the value of n.

Definition The matrix P whose elements are given by (8.8) is called the transition matrix of the Markov chain.

Properties of P

Writing out the array **P** we have

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \cdot & \cdot & \cdots & \cdot \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{bmatrix}$$

It is seen that **P** has M rows and M columns. Every element of **P** satisfies the **non-negativity condition**

$$p_{jk} \geqslant 0. \tag{8.9}$$

Also, the sum of the elements in each row of **P** is unity. That is,

$$\sum_{k=1}^{M} p_{jk} = 1, \qquad j = 1, \ldots, M. \tag{8.10}$$

A square matrix whose elements satisfy (8.9) and (8.10) is called a **stochastic matrix.**

The probability distribution of X_n

The M quantities

$$p_j(0) = \Pr\{X_0 = s_j\}$$

can be arranged as the components of a row vector:

$$\mathbf{p}(0) = [p_1(0)\, p_2(0) \cdots p_M(0)]$$

Similarly, for X_n, $n \geqslant 1$, let

$$p_j(n) = \Pr\{X_n = s_j\},$$

and

$$\mathbf{p}(n) = [p_1(n)\, p_2(n) \cdots p_M(n)].$$

We now prove the following.

Theorem 8.3 The probability distribution of X_n, $n \geqslant 1$, is given in terms of that of X_0 by

$$\boxed{\mathbf{p}(n) = \mathbf{p}(0)\mathbf{P}^n}, \tag{8.11}$$

where P is the transition matrix of the Markov chain.

Proof We proceed by induction, first showing that (8.11) is true for $n = 1$.

If the value of X_0 is s_k, the value of X_1 will be s_j only if a transition is made from s_k to s_j. The events '$X_0 = s_k, k = 1, 2, \ldots, M$' are mutually exclusive and one of them must occur. Hence, by the law of total probability

$$\Pr\{X_1 = s_j\} = \sum_{k=1}^{M} \Pr\{X_0 = s_k\} \Pr\{X_1 = s_j | X_0 = s_k\},$$

or

$$p_j(1) = \sum_{k=1}^{M} p_k(0) p_{kj}, \qquad j = 1, 2, \ldots, M. \tag{8.12}$$

Recall now that if **A** is an $m \times n$ matrix with element a_{ij} in its ith row and jth column, and if **B** is an $n \times p$ matrix with general element b_{ij}, then the $m \times p$ product matrix $\mathbf{C} = \mathbf{AB}$ has general element

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}; \qquad i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, p.$$

From (8.12),

$$\mathbf{p}(1) = \mathbf{p}(0)\mathbf{P}.$$

Assume now the truth of (8.11), for some $n > 1$. Clearly

$$\Pr\{X_{n+1} = s_j\} = \sum_{k=1}^{M} \Pr\{X_n = s_k\} \Pr\{X_{n+1} = s_j | X_n = s_k\},$$

or,

$$p_j(n+1) = \sum_{k=1}^{M} p_k(n) p_{kj}.$$

In terms of vectors and matrices this becomes

$$\begin{aligned} \mathbf{p}(n+1) &= \mathbf{p}(n)\mathbf{P} \\ &= \mathbf{p}(0)\mathbf{P}^n\mathbf{P}, \end{aligned}$$

because we have assumed (8.11) is true. Since $\mathbf{P}^n\mathbf{P} = \mathbf{P}^{n+1}$, we find

$$\mathbf{p}(n+1) = \mathbf{p}(0)\mathbf{P}^{n+1}.$$

This completes the inductive proof as it follows that (8.11) is true for all $n \geqslant 1$.

The matrix $\mathbf{P}^n$ also has M rows and M columns. Its elements, denoted by $p_{jk}^{(n)}$, are called the ***n*-step transition probabilities** since they give the probabilities of transitions from s_j to s_k in n time steps. It is left as Exercise 3 to prove the **Chapman–Kolmogorov forward equations**

$$p_{ik}^{(m+n)} = \sum_{k=1}^{M} p_{ij}^{(m)} p_{jk}^{(n)}.$$

8.6 RANDOM GENETIC DRIFT

We now return to study the Markov chain of Section 8.3 in which X_n is the number of genes of type A_1 in a randomly mating population of size N. The state space $\mathcal{S}$ contains $2N+1$ elements which are just the integers $0, 1, 2, \ldots, 2N$. The elements of the transition matrix are given by (8.4):

$$p_{jk} = \binom{2N}{k}\left(\frac{j}{2N}\right)^k\left(1-\frac{j}{2N}\right)^{2N-k}; \qquad j,k = 0,1,\ldots,2N. \tag{8.13}$$

Thus $\mathbf{P}$ has $2N+1$ rows and $2N+1$ columns, and in Exercise 8.4 it is shown that $\mathbf{P}$ is stochastic. For $N=1$ the transition matrix is

$$\mathbf{P} = \tfrac{1}{4}\begin{bmatrix} 4 & 0 & 0 \\ 1 & 2 & 1 \\ 0 & 0 & 4 \end{bmatrix}.$$

When $N=2$ we find

$$\mathbf{P} = \tfrac{1}{256}\begin{bmatrix} 256 & 0 & 0 & 0 & 0 \\ 81 & 108 & 54 & 12 & 1 \\ 16 & 64 & 96 & 64 & 16 \\ 1 & 12 & 54 & 108 & 81 \\ 0 & 0 & 0 & 0 & 256 \end{bmatrix} \tag{8.14}$$

Recall now that individuals with A_1A_1 or A_2A_2 are called homozygous whereas those with A_1A_2 are called heterozygous. We will see, first heuristically by a numerical example with $N=2$, that a finite population of individuals which mate randomly according to our assumptions, evolves to a state in which there are no heterozygous individuals. Note that for a population of size N consisting of only homozygous individuals, the number of A_1 alleles is either 0 (corresponding to all A_2A_2) or $2N$ (all A_1A_1).

We choose a probability distribution for X_0 so that the probability that the population is homozygous is zero:

$$\mathbf{p}(0) = [0 \quad \tfrac{1}{4} \quad \tfrac{1}{2} \quad \tfrac{1}{4} \quad 0].$$

We now compute $\mathbf{p}(1) = \mathbf{p}(0)\mathbf{P}$ by matrix multiplication to find the probability distribution of X_1. This gives

$$\mathbf{p}(1) = [0.1113 \quad 0.2422 \quad 0.2930 \quad 0.2422 \quad 0.1113]$$

Similarly, the distribution of X_2 is given by $\mathbf{p}(2) = \mathbf{p}(1)\mathbf{P} = \mathbf{p}(0)\mathbf{P}^2$:

$$\mathbf{p}(2) = [0.2072 \quad 0.1868 \quad 0.2121 \quad 0.1868 \quad 0.2072]$$

The probability distributions of the number of A_1-alleles in the next four

generations are found to be as follows:

$$\mathbf{p}(3) = [0.2803 \quad 0.1406 \quad 0.1583 \quad 0.1406 \quad 0.2803]$$
$$\mathbf{p}(4) = [0.3352 \quad 0.1055 \quad 0.1187 \quad 0.1055 \quad 0.3352]$$
$$\mathbf{p}(5) = [0.3764 \quad 0.0791 \quad 0.0890 \quad 0.0791 \quad 0.3764]$$
$$\mathbf{p}(6) = [0.4073 \quad 0.0593 \quad 0.0667 \quad 0.0593 \quad 0.4073].$$

Figure 8.2 shows sketches of the distributions of $X_0, X_1, \ldots, X_6$.

It can be seen that by the third generation ($n=3$) there is more probability

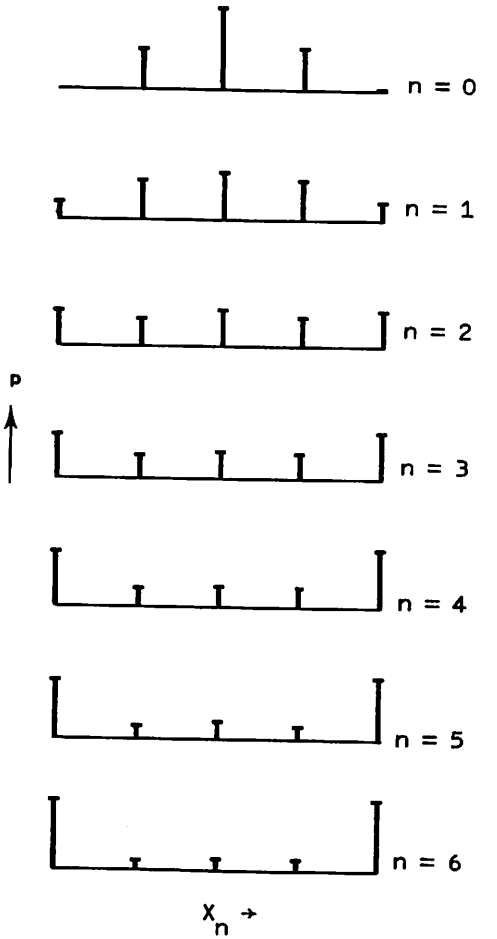

Figure 8.2 Evolution of the probability distribution of the number of A_1 genes in the text example.

mass concentrated at the homozygous states than in the heterozygous states. This contrasts with the situation in the initial population in which the probability of homozygous states was zero. By the sixth generation the probability that the population is homozygous has grown to 0.8146. Eventually, there is zero chance that the population is heterozygous, even if it started as heterozygous with probability one.

This tendency for a population to become homozygous is referred to as random genetic drift or just random drift. It was first studied theoretically by the pioneering population geneticists R.A. Fisher and Sewall Wright. This phenomenon is in direct contrast with the Hardy–Weinberg prediction of constant gene frequencies in infinite populations. It is purely due to the random sampling of gametes (egg and sperm cells) from a parent generation to form the individuals of the next generation. However, it only occurs in finite populations and the smaller the population (N), the faster is the approach to homozygosity; or, as population geneticists say, the faster do the genes become fixed in the population (all A_1A_1 or all A_2A_2).

Provided the assumptions which led to this theoretical prediction are fulfilled, we expect in small populations that after a few generations there is a large chance of having all homozygous individuals. In large populations the drift will proceed more slowly. The fact that nearly all Chinese have the same black hair colouring, the same brown colour eyes, etc., probably means that the population has been around a very long time and a state of homozygosity has been reached for the genes controlling these physical characteristics.

8.7 MARKOV CHAINS WITH ABSORBING STATES

Let $\{X_n, n = 0, 1, 2, \ldots\}$ be a temporally homogeneous Markov chain with state space $\mathscr{S}$ containing elements $s_1, s_2, \ldots$. Suppose it is possible to get from state s_j to state s_k in a finite time; that is, $p_{jk}^{(n)} > 0$ for some n. Then we say that state s_k is **accessible** from state s_j, or s_k can be reached from s_j. If s_j is also accessible from state s_k we say that states s_j and s_k **communicate**. A state may, of course, communicate with itself.

However, some states may act as traps so that once entered they cannot be left, as for example in the random walk of Section 7.3. If s_j is such a state, then $p_{jj} = 1$ and s_j is called **absorbing**. The jth row of the transition matrix will then consist of all zeros except for the 1 in column j. There may be just one or several absorbing states.

Absorption is certain

We make the following assumptions concerning the states of a temporally homogeneous Markov chain.

Assumptions

(i) The state space $\mathscr{S} = \{s_1, s_2, \ldots, s_M\}$ contains a finite number of elements.
(ii) The states in the set $\mathscr{A} = \{s_1, s_2, \ldots, s_A\}$, where $A \geqslant 1$, are absorbing. That is, there is at least one absorbing state.
(iii) At least one of the absorbing states is accessible from any member of the set $\mathscr{B} = \{s_{A+1}, \ldots, s_M\}$ of non-absorbing states.

We now prove the following.

Theorem 8.4 Under the above assumptions,

$$\Pr\{X_n \in \mathscr{A}\} \underset{n\to\infty}{\longrightarrow} 1.$$

That is, absorption of X in one or other of the absorbing states is certain.

Proof If $X_0 \in \mathscr{A}$ there is nothing to prove, since X is already absorbed. Therefore, let $X_0 \in \mathscr{B}$. By assumption there is at least one state in $\mathscr{A}$ which is accessible from any state in $\mathscr{B}$. Hence, there is a state $s_k \in \mathscr{A}$ which is accessible from $s_j \in \mathscr{B}$, and so we may define $n_{jk} < \infty$ as the smallest number n such that $p_{jk}^{(n)} > 0$.

For a given state s_j let n_j be the largest of the collection of n_{jk} as k varies and let n' be the largest of the n_j as j varies. After n' time steps, no matter what the initial state of the process, there is a probability $p > 0$ that the process is in an absorbing state. Hence

$$\Pr\{X_{n'} \in \mathscr{B}\} = 1 - p$$

and $0 < (1 - p) < 1$. It follows by temporal homogeneity and the Markov property that $\Pr\{X_{2n'} \in \mathscr{B}\} \leqslant (1 - p)^2$ and, in general,

$$\Pr\{X_{kn'} \in \mathscr{B}\} \leqslant (1 - p)^k, \qquad k = 1, 2, \ldots$$

Since as $k \to \infty$, $(1 - p)^k \to 0$, we see that $\Pr\{X_n \in \mathscr{B}\} \to 0$ as $n \to \infty$. This proves that the process must eventually end up in one of the absorbing states.

Theorem 8.4 and the above proof are based on Theorem 3.1.1 of Kemeny and Snell (1960).

Example 1

For the Markov chain of Section 8.3 in which X_n is the number of genes of type A_1 in generation n, the values 0 and $2N$ are absorbing since $p_{00} = 1$ and $p_{2N,2N} = 1$. The assumptions of Theorem 8.4 are fulfilled and it follows immediately that absorption in one or the other of the absorbing states must eventually occur. That is,

$$\Pr\{X_n = 0 \cup X_n = 2N\} \underset{n\to\infty}{\longrightarrow} 1.$$

Example 2

Consider the simple random walk of Section 7.3 where X_n is the 'position of the particle' or a 'gambler's fortune' at epoch n, with absorbing barriers at 0 and c. The elements of the transition matrix of this temporally homogeneous Markov chain are, for $j = 1, 2, \ldots, c-1$,

$$p_{jk} = \Pr\{X_{n+1} = k | X_n = j\} = \begin{cases} p, & \text{if } k = j+1 \\ q, & \text{if } k = j-1 \\ 0, & \text{otherwise,} \end{cases}$$

whereas

$$\begin{cases} p_{00} = 1, \\ p_{0k} = 0, & k = 1, \ldots, c, \\ p_{c,k} = 0, & k = 0, \ldots, c-1, \\ p_{cc} = 1. \end{cases}$$

Thus $\mathbf{P}$ has $c+1$ rows and $c+1$ columns and has the form

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ q & 0 & p & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & q & 0 & p & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & q & 0 & p & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & q & 0 & p \\ 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 1 \end{bmatrix}$$

It is intuitively clear that the absorbing states are accessible from any of the non-absorbing states, $1, 2, \ldots, c-1$. By Theorem 8.4 absorption at 0 or c is certain as $n \to \infty$, a fact that we proved by a different method in Section 7.5.

8.8 ABSORPTION PROBABILITIES

Given a temporally homogeneous Markov chain which satisfies assumptions (i)–(iii) of the previous section, we have seen that the process must terminate in one of the absorbing states. If there is more than one absorbing state we may wish to know the chances of absorption in the individual absorbing states. For example, in the Markov chain model which displays random genetic drift, we would like to know the probability that the population ends up having all individuals of genotype A_1A_1 as opposed to all A_2A_2. We thus require the **absorption probabilities** for the various absorbing states. In this section we show how to calculate these probabilities as functions of the initial value of the process.

If states $s_1, \ldots, s_A$ are absorbing and there are M states altogether, the

transition matrix can be put in the form

$$\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \\ \cdot \\ \cdot \\ A \\ A+1 \\ \cdot \\ \cdot \\ M \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & \cdot & \cdot & A & A+1 & \cdot & \cdot & M \end{array} \\ \left[\begin{array}{cccc:ccc} 1 & 0 & \cdot & \cdot & 0 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & \cdot & \cdot & 0 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 1 & 0 & \cdot & \cdot & 0 \\ \hdashline p_{A+1,1} & \cdot & \cdot & \cdot & p_{A+1,A} & p_{A+1,A+1} & \cdot & \cdot & p_{A+1,M} \\ \cdot & & & & \cdot & \cdot & & & \cdot \\ \cdot & & & & \cdot & \cdot & & & \cdot \\ p_{M,1} & \cdot & \cdot & \cdot & p_{M,A} & p_{M,A+1} & \cdot & \cdot & p_{M,M} \end{array}\right] \end{array} \tag{8.15}$$

Introducing the $(M-A) \times (M-A)$ submatrix

$$Q = \begin{bmatrix} p_{A+1,A+1} & \cdot & \cdot & \cdot & p_{A+1,M} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ p_{M,A+1} & \cdot & \cdot & \cdot & p_{M,M} \end{bmatrix}$$

and the $(M-A) \times A$ submatrix

$$\mathbf{R} = \begin{bmatrix} p_{A+1,1} & \cdot & \cdot & \cdot & p_{A+1,A} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ p_{M,1} & \cdot & \cdot & \cdot & p_{M,A} \end{bmatrix} \tag{8.16}$$

the matrix $\mathbf{P}$ can be partitioned as

$$\mathbf{P} = \left[\begin{array}{c:c} \mathbf{I} & \mathbf{0} \\ \hdashline \mathbf{R} & \mathbf{Q} \end{array}\right]$$

where $\mathbf{I}$ is an $A \times A$ identity matrix and $\mathbf{0}$ is an $A \times (M-A)$ zero matrix. The elements of $\mathbf{Q}$ are the one-step transition probabilities among the non-absorbing states, and the elements of $\mathbf{R}$ are the one-step transition probabilities from non-absorbing to absorbing states.

We now define the matrix $\mathbf{\Pi}$ whose elements are the required absorption probabilities:

$$\pi_{jk} = \Pr\{\text{process is absorbed in state } s_k \in \mathscr{A} | \text{starts in } s_j \in \mathscr{B}\} \tag{8.17}$$

It is seen that $\mathbf{\Pi}$ has $(M-A)$ rows and A columns. We introduce the matrix

$$\boxed{\mathbf{\Phi} = (\mathbf{I} - \mathbf{Q})^{-1}},$$

which is called the **fundamental matrix** of the Markov chain, where here $\mathbf{I}$ is an identity matrix with the same number of rows and columns as $\mathbf{Q}$. In terms of $\mathbf{\Phi}$ and the matrix $\mathbf{R}$ defined by (8.16) we have the following result.

Theorem 8.5 The matrix whose elements are the absorption probabilities (8.17) is given by

$$\boxed{\mathbf{\Pi} = \mathbf{\Phi R}}$$

Proof From the state $s_j \in \mathscr{B}$ the process goes at the first time-step to state $s_i \in \mathscr{S}$ with probability p_{ji}. Allowing for these possible first transitions we have

$$\pi_{jk} = \sum_{i=1}^{M} p_{ji} \Pr\{\text{process is absorbed in state } s_k | \text{starts in } s_i\}. \tag{8.18}$$

Allowing for the contingencies

$$\Pr\{\text{process starts in state } s_i \text{ and is absorbed in state } s_k\} = \begin{cases} 1, & s_i = s_k, \\ 0, & s_i \in \mathscr{A}, \quad s_i \neq s_k, \\ \pi_{ik}, & s_i \in \mathscr{B}, \quad i = A+1, \ldots, M, \end{cases}$$

equation (8.18) becomes

$$\pi_{jk} = p_{jk} + \sum_{i=A+1}^{M} p_{ji}\pi_{ik}, \qquad j = A+1, \ldots, M; \quad k = 1, \ldots, A. \tag{8.19}$$

But $p_{jk}, j = A+1, \ldots, M; k = 1, \ldots, A$ are the elements of $\mathbf{R}$, whereas p_{ji}, $j = A+1, \ldots, M; i = A+1, \ldots, M$ are the elements of $\mathbf{Q}$. Hence, in matrix notation, (8.19) becomes

$$\mathbf{\Pi} = \mathbf{R} + \mathbf{Q\Pi}$$

Rearranging and being careful to preserve the order of matrix multiplication,

$$(\mathbf{I} - \mathbf{Q})\mathbf{\Pi} = \mathbf{R}.$$

Premultiplying both sides with the inverse of $(\mathbf{I} - \mathbf{Q})$ gives

$$\mathbf{\Pi} = (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{R},$$

which proves the theorem, since $\mathbf{\Phi} = (\mathbf{I} - \mathbf{Q})^{-1}$.

Example 1

Consider the Markov chain model for the numbers of A_1-genes in a (self-fertilizing) population with $N = 1$. The possible values are 0, 1, 2. The matrix of probabilities of transitions among the non-absorbing states consists of a single entry,

$$\mathbf{Q} = [p_{11}] = [\tfrac{1}{2}].$$

The matrix of probabilities of transitions from non-absorbing to absorbing states is

$$\mathbf{R} = [p_{10}\, p_{12}] = [\tfrac{1}{4}\, \tfrac{1}{4}].$$

Then

$$\mathbf{I} - \mathbf{Q} = [\tfrac{1}{2}]$$

and

$$\mathbf{\Phi} = (\mathbf{I} - \mathbf{Q})^{-1} = [2]. \tag{8.20}$$

Thus, from Theorem 8.5, the absorption probabilities are

$$\mathbf{\Pi} = [\pi_{10}\, \pi_{12}] = \mathbf{\Phi R} = [\tfrac{1}{2}\, \tfrac{1}{2}].$$

Example 2

Let a simple random walk be restricted by absorbing barriers at 0 and $c = 3$. The transition matrix is

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

The matrix $\mathbf{Q}$ is given by

$$\mathbf{Q} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 0 & p \\ q & 0 \end{bmatrix},$$

and

$$\mathbf{R} = \begin{bmatrix} p_{10} & p_{13} \\ p_{20} & p_{23} \end{bmatrix} = \begin{bmatrix} q & 0 \\ 0 & p \end{bmatrix}$$

Then

$$(\mathbf{I} - \mathbf{Q}) = \begin{bmatrix} 1 & -p \\ -q & 1 \end{bmatrix}$$

Recall that the inverse of a general 2×2 matrix

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

is

$$\mathbf{A}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \qquad ad - bc \neq 0,$$

as can be checked by showing $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$.

Hence the fundamental matrix for this Markov chain is

$$\mathbf{\Phi} = \frac{1}{1-pq}\begin{bmatrix} 1 & p \\ q & 1 \end{bmatrix} \tag{8.21}$$

The probabilities of absorption into states 0 and 3 are, by Theorem 8.5,

$$\begin{aligned} \mathbf{\Pi} &= \begin{bmatrix} \pi_{10} & \pi_{13} \\ \pi_{20} & \pi_{23} \end{bmatrix} = \mathbf{\Phi R} \\ &= \frac{1}{1-pq}\begin{bmatrix} 1 & p \\ q & 1 \end{bmatrix}\begin{bmatrix} q & 0 \\ 0 & p \end{bmatrix} \\ &= \frac{1}{1-pq}\begin{bmatrix} q & p^2 \\ q^2 & p \end{bmatrix} \end{aligned}$$

In the exercises it is confirmed that the probability of absorption (P_a) at zero for an initial value a, as given by formula (7.17) with $c = 3$, agrees with the values of π_{10} and π_{20}. The row sums of $\mathbf{\Pi}$ are unity, since absorption into one or the other absorbing states is certain. This is also confirmed in Exercise 7.

Example 3

Consider the Markov chain model for the number of A_1 genes but now let the population size be $N = 2$. The state space consists of $0, 1, 2, 3, 4$ and the transition matrix is given by (8.14).

The matrices $\mathbf{Q}$ and $\mathbf{R}$ are

$$\begin{aligned} \mathbf{Q} &= \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \\ &= \tfrac{1}{256}\begin{bmatrix} 108 & 54 & 12 \\ 64 & 96 & 64 \\ 12 & 54 & 108 \end{bmatrix}, \end{aligned}$$

$$\begin{aligned} \mathbf{R} &= \begin{bmatrix} p_{10} & p_{14} \\ p_{20} & p_{24} \\ p_{30} & p_{34} \end{bmatrix} \\ &= \tfrac{1}{256}\begin{bmatrix} 81 & 1 \\ 16 & 16 \\ 1 & 81 \end{bmatrix}. \end{aligned}$$

Thus

$$(\mathbf{I} - \mathbf{Q}) = \tfrac{1}{256}\begin{bmatrix} 148 & -54 & -12 \\ -64 & 160 & -64 \\ -12 & -54 & 148 \end{bmatrix}.$$

To invert this matrix by hand to find $\mathbf{\Phi}$ is too messy. However, it will be readily verified (see exercises) that the solutions of the equations

$$(\mathbf{I} - \mathbf{Q})\mathbf{\Pi} = \mathbf{R},$$

with

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{10} & \pi_{14} \\ \pi_{20} & \pi_{24} \\ \pi_{30} & \pi_{34} \end{bmatrix}$$

are

$$\mathbf{\Pi} = \tfrac{1}{4}\begin{bmatrix} 3 & 1 \\ 2 & 2 \\ 1 & 3 \end{bmatrix}.$$

In fact the general result with a population of size N is

$$\pi_{k,0} = 1 - \frac{k}{2N}, \qquad \pi_{k,2N} = \frac{k}{2N}, \qquad k = 1, 2, \ldots, 2N-1, \tag{8.22}$$

as will also be seen in Exercises 8 and 9.

8.9 THE MEAN TIME TO ABSORPTION

For the Markov chain with transition probabilities given by (8.15) we would like to have as much information as possible concerning the number of time units required to reach an absorbing state from a non-absorbing state. This length of time is of course a random variable which we call the **time to absorption**.

In the population genetics example, the time to absorption of the Markov chain is the time it takes for the heterozygotes to disappear completely from the population. In the random walk with absorbing barriers, the time to absorption is, in the gambling context, the duration of the game or the time required for one player to go broke. In this section we obtain formulas for the mean of the time to absorption.

We define the following two **random variables.**

Definition Let N_{jk} be the number of times the non-absorbing state s_k is occupied until absorption takes place when the Markov chain starts in the non-absorbing state s_j, The collection of N_{jk} forms the $(M-A) \times (M-A)$ matrix N.

Definition Let T_j be the total number of time units until absorption when the Markov chain starts in the non-absorbing state s_j.

The random variable T_j is the time to absorption from state s_j. The collection of T_j, with $j = A+1, \ldots, M$, forms the $1 \times (M-A)$ row-vector of

absorption times for various initial states:

$$\mathbf{T} = [T_{A+1} T_{A+2} \cdots T_M].$$

Since the time to absorption is the total number of times that all the non-absorbing states are occupied, the following relation holds between T_j and the N_{jk}:

$$T_j = \sum_{k=A+1}^{M} N_{jk}. \tag{8.23}$$

The following result gives the expectation of T_j as the sum of the elements in the jth row of the fundamental matrix Φ.

Theorem 8.6 The mean time to absorption from state s_j is

$$E(T_j) = \sum_{k=A+1}^{M} \phi_{jk}, \qquad j = A+1, \ldots, M, \tag{8.24}$$

where ϕ_{jk} is the (j,k)-element of the fundamental matrix Φ.

The equations (8.24) may be written in matrix-vector notation as

$$E(\mathbf{T}) = \Phi\xi$$

where ξ is the $(M-A) \times 1$ column vector

$$\xi = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}.$$

Proof The sketch in Fig. 8.3 depicts an initial state s_j and the possible states s_i after the first transition. States $1, \ldots, A$ are absorbing and are lumped together. We will calculate $E(N_{jk})$, there being two separate cases to consider.

Case (i): $k \neq j$.

If the first transition is to an absorbing state, then $N_{jk} = 0$. Hence

$$N_{jk} = 0 \qquad \text{with probability } \sum_{i=1}^{A} p_{ji}.$$

If the first transition is to a non-absorbing state s_i then the total number of times that state s_k is occupied is N_{ik}. Hence

$$N_{jk} = N_{ik} \qquad \text{with probability } p_{ji}, \qquad i = A+1, \ldots, M.$$

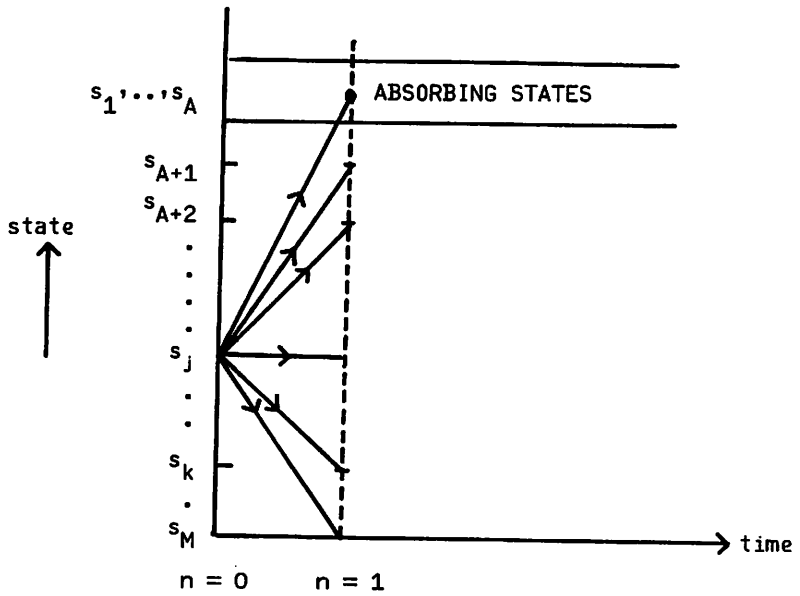

Figure 8.3 Possible transitions from the initial state s_j.

By the law of total probability applied to expectations (see Chapter 1), we must have

$$E(N_{jk}) = \sum_{i=1}^{M} \Pr\{\text{1st transition is from } s_j \text{ to } s_i\} \times E(N_{jk} | \text{1st transition is from } s_j \text{ to } s_i).$$

The absorbing states contribute zero, so

$$E(N_{jk}) = 0 \times \sum_{i=1}^{A} p_{ji} + \sum_{i=A+1}^{M} p_{ji} E(N_{jk} | \text{1st transition is from } s_j \text{ to } s_i).$$

But we have seen that

$$E(N_{jk} | \text{1st transition is from } s_j \text{ to } s_i) = E(N_{ik}).$$

Hence, for $k \neq j$,

$$E(N_{jk}) = \sum_{i=A+1}^{M} p_{ji} E(N_{ik}). \tag{8.25}$$

Case (ii): $k = j$.

We have $N_{jj} = 1$ if absorption occurs on the first transition, so

$$N_{jj} = 1 \qquad \text{with probability } \sum_{i=1}^{A} p_{ji}.$$

If the first transition is to a non-absorbing state s_i, then

$$N_{jj} = 1 + N_{ij} \qquad \text{with probability } p_{ji}.$$

Thus,

$$\begin{aligned} E(N_{jj}) &= \sum_{i=1}^{M} \Pr\{\text{1st transition is from } s_j \text{ to } s_i\} \\ &\quad \times E(N_{jj} | \text{1st transition is from } s_j \text{ to } s_i) \\ &= 1 \times \sum_{i=1}^{A} p_{ji} + \sum_{i=A+1}^{M} p_{ji}(1 + E(N_{ij})) \\ &= 1 + \sum_{i=A+1}^{M} p_{ji} E(N_{ij}). \end{aligned} \tag{8.26}$$

Introducing the symbol (Kronecker's delta)

$$\delta_{jk} = \begin{cases} 1, & j = k, \\ 0, & j \neq k, \end{cases}$$

equations (8.25) and (8.26) may be summarized as

$$E(N_{jk}) = \delta_{jk} + \sum_{i=A+1}^{M} p_{ji} E(N_{ik}).$$

In matrix form,

$$E(\mathbf{N}) = \mathbf{I} + \mathbf{Q}E(\mathbf{N}),$$

or

$$(\mathbf{I} - \mathbf{Q})E(\mathbf{N}) = \mathbf{I}.$$

Hence

$$E(\mathbf{N}) = \mathbf{\Phi},$$

by definition of $\mathbf{\Phi} = (\mathbf{I} - \mathbf{Q})^{-1}$. Combining this with (8.23),

$$E(T_j) = \sum_{k=A+1}^{M} E(N_{jk}) = \sum_{k=A+1}^{M} \phi_{jk},$$

so the value of $E(T_j)$ is the sum of the elements in the jth row of $\mathbf{\Phi}$ as required.

Example 1

For the Markov chain model exhibiting random drift, when $N = 1$, there is only one non-absorbing state, corresponding to an A_1A_2 individual. The states are s_0, s_1 and s_2. The fundamental matrix is (see (8.20))

$$\mathbf{\Phi} = [2].$$

The expected time to absorption into either state s_0 or s_2 is, by Theorem 8.6, the sum of the elements in the row of the fundamental matrix corresponding to the initial state. Hence the mean time to fixation (A_1A_1 or A_2A_2) is

$$E(T_1) = 2.$$

That is, the mean fixation time is 2 generations.

Example 2

For the random walk with absorbing barriers at 0 and $c = 3$, the fundamental matrix is (see (8.21))

$$\mathbf{\Phi} = \frac{1}{1 - pq}\begin{bmatrix} 1 & p \\ q & 1 \end{bmatrix}.$$

The expected value of T_1, the time to absorption when $X_0 = 1$, is the sum of the elements in the first row of $\mathbf{\Phi}$:

$$E(T_1) = \frac{1 + p}{1 - pq}.$$

Similarly, the expected time to absorption when $X_0 = 2$ is

$$E(T_2) = \frac{1 + q}{1 - pq}.$$

In Exercise 10 it is verified that these results agree with those derived previously (equation (7.23)).

8.10 MUTATION

Genes can be changed by certain stimuli, such as radiation. Sometimes in the 'natural' course of events, chemical accidents may occur which change one allele to another. Such alteration of genetic material is called mutation.

In this section we modify the Markov chain model of Section 8.3 to allow for the possibility of mutation of A_1 alleles to A_2 alleles and vice versa. We will see that this modification drastically alters the properties of the Markov chain.

Suppose that each A_1 allele mutates to A_2 with probability α_1 per generation and that each A_2 allele mutates to A_1 with probability α_2 per generation. By considering the ways in which a gene may be of type A_1 we see that

$$\begin{aligned} &\Pr\{\text{a gene is } A_1 \text{ after mutation}\} \\ &\quad = \Pr\{\text{gene is } A_1 \text{ before mutation}\} \\ &\qquad \times \Pr\{\text{gene does not mutate from } A_1 \text{ to } A_2\} \\ &\qquad + \Pr\{\text{gene is } A_2 \text{ before mutation}\} \\ &\qquad \times \Pr\{\text{gene mutates from } A_2 \text{ to } A_1\}. \end{aligned}$$

Thus, if there were j genes of type A_1 in the parent population before mutation, the probability p_j of choosing an A_1 gene when we form the next generation is

$$p_j = \frac{j}{2N}(1-\alpha_1) + \left(1 - \frac{j}{2N}\right)\alpha_2. \tag{8.27}$$

Let X_n be the number of A_1 alleles at generation n. Suppose in fact that there were j genes of type A_1 in generation n. We choose $2N$ genes to form generation $n+1$ with probability p_j of an A_1 and probability $1-p_j$ of an A_2 at each trial. Then $\{X_n, n=0,1,2,\ldots\}$ is a temporally homogeneous Markov chain with one-step transition probabilities

$$\begin{aligned} p_{jk} &= \Pr\{X_{n+1}=k \mid X_n = j\} \\ &= \binom{2N}{k}(p_j)^k(1-p_j)^{2N-k}; \qquad j,k = 0,1,2,\ldots,2N, \end{aligned} \tag{8.28}$$

with p_j given by (8.27).

Example

Let $N=1$, $\alpha_1 = \alpha_2 = \frac{1}{4}$. Then, substituting in (8.27),

$$p_j = \frac{j}{4} + \frac{1}{4}.$$

This gives $p_0 = \frac{1}{4}, p_1 = \frac{1}{2}, p_2 = \frac{3}{4}$. The elements of **P** are, from (8.28), for $k = 0,1,2$,

$$p_{0k} = \binom{2}{k}(\tfrac{1}{4})^k(\tfrac{3}{4})^{2-k}$$

$$p_{1k} = \binom{2}{k}(\tfrac{1}{2})^2$$

$$p_{2k} = \binom{2}{k}(\tfrac{3}{4})^k(\tfrac{1}{4})^{2-k}.$$

Evaluating these we obtain

$$\mathbf{P} = \tfrac{1}{16}\begin{bmatrix} 9 & 6 & 1 \\ 4 & 8 & 4 \\ 1 & 6 & 9 \end{bmatrix}.$$

There are no absorbing states, as there are no ones on the principal diagonal. All the elements of **P** are non-zero, so at any time step a transition is possible from any state to any other. We will see that in contrast to the case where there is no mutation, an equilibrium probability distribution is eventually achieved.

8.11 STATIONARY DISTRIBUTIONS

Let **P** be the transition matrix of a temporally homogeneous Markov chain $\{X_n, n=0,1,2,\ldots\}$. Suppose there exists a **probability vector** $\hat{\mathbf{p}}$ (i.e., a row vector with non-negative components whose sum is unity) satisfying

$$\hat{\mathbf{p}}\mathbf{P} = \hat{\mathbf{p}}. \tag{8.29}$$

Now let the probability distribution of X_n be given by $\mathbf{p}(n)$ and suppose the process starts with

$$\mathbf{p}(0) = \hat{\mathbf{p}}.$$

Then we must have

$$\mathbf{p}(1) = \mathbf{p}(0)\mathbf{P} = \hat{\mathbf{p}}\mathbf{P} = \hat{\mathbf{p}}$$
$$\mathbf{p}(2) = \mathbf{p}(1)\mathbf{P} = \hat{\mathbf{p}}\mathbf{P} = \hat{\mathbf{p}},$$

and it can be seen that

$$\mathbf{p}(n) = \hat{\mathbf{p}},$$

for all $n = 0,1,2,\ldots$

We find that if the initial probability distribution is given by $\hat{\mathbf{p}}$, the probability distribution of the process at each time step is the same. We call such a probability distribution **stationary** or **time-invariant**. The random variables $X_0, X_1, X_2, \ldots$ are thus identically distributed if X_0 has the distribution $\hat{\mathbf{p}}$.

Definition Let P be the transition matrix of a temporally homogeneous Markov chain. If there exists a probability vector $\hat{\mathbf{p}}$ such that $\hat{\mathbf{p}} = \hat{\mathbf{p}}\mathbf{P}$, then $\hat{\mathbf{p}}$ is called a stationary distribution for the Markov chain.

Note on terminology

A vector **x** is said to be a (left) **eigenvector** of the matrix **A** if **xA** is a scalar (real or complex) multiple of **x**. That is,

$$\mathbf{x}\mathbf{A} = \lambda\mathbf{x},$$

where λ is a scalar called the corresponding **eigenvalue**. According to (8.29) $\hat{\mathbf{p}}\mathbf{P} = 1\hat{\mathbf{p}}$. Hence a stationary distribution is an eigenvector of **P** with eigenvalue 1. Any non-zero multiple of an eigenvector is also an eigenvector with the same eigenvalue. To fix $\hat{\mathbf{p}}$ we insist that its components sum to unity.

Example

Let the transition matrix of a temporally homogeneous Markov chain be

$$\mathbf{P} = \begin{bmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix}. \tag{8.30}$$

An eigenvector $\mathbf{x} = [x_1 x_2]$ of $\mathbf{P}$ with eigenvalue 1 must satisfy

$$[x_1 \quad x_2]\begin{bmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix} = [x_1 \quad x_2].$$

Thus

$$0.4x_1 + 0.2x_2 = x_1$$
$$0.6x_1 + 0.8x_2 = x_2.$$

From the first (or second) of these equations we find

$$x_2 = 3x_1.$$

Hence any multiple of the row vector

$$\mathbf{x} = [1 \quad 3]$$

is an eigenvector with eigenvalue 1. To obtain a probability vector we must divide by the sum of the components. Thus a stationary probability vector for this Markov chain is

$$\hat{\mathbf{p}} = [0.25 \quad 0.75] \doteq [\hat{p}_1 \quad \hat{p}_2].$$

8.12 APPROACH TO A STATIONARY DISTRIBUTION AS $n \to \infty$

Consider again the Markov chain with $\mathbf{P}$ given by (8.30). Computing successive powers of $\mathbf{P}$ we find

$$\mathbf{P}^2 = \begin{bmatrix} 0.28 & 0.72 \\ 0.24 & 0.76 \end{bmatrix}$$

$$\mathbf{P}^3 = \begin{bmatrix} 0.256 & 0.744 \\ 0.248 & 0.752 \end{bmatrix}$$

$$\mathbf{P}^4 = \begin{bmatrix} 0.2512 & 0.7488 \\ 0.2496 & 0.7504 \end{bmatrix}$$

It would seem, and we will see that it is true, that as n increases $\mathbf{P}^n$ is approaching, element by element, the matrix

$$\hat{\mathbf{P}} = \begin{bmatrix} 0.25 & 0.75 \\ 0.25 & 0.75 \end{bmatrix}.$$

That is,

$$\lim_{n\to\infty} \mathbf{P}^n = \hat{\mathbf{P}}. \tag{8.31}$$

Note that each row of $\hat{\mathbf{P}}$ is the same as the stationary probability vector $\hat{\mathbf{p}}$, so

$$\hat{\mathbf{P}} = \begin{bmatrix} \hat{p}_1 & \hat{p}_2 \\ \hat{p}_1 & \hat{p}_2 \end{bmatrix}.$$

In terms of matrix elements,

$$\lim_{n\to\infty} p_{jk}^{(n)} = \hat{p}_k,$$

regardless of the value of j.

Let us see what happens, if (8.31) is true, to the probability distribution of X_n as $n \to \infty$, for an arbitrary initial distribution

$$\mathbf{p}(0) = [p_1(0) \quad p_2(0)].$$

Since

$$\mathbf{p}(n) = \mathbf{p}(0)\mathbf{P}^n,$$

we have

$$\begin{aligned} \lim_{n\to\infty} \mathbf{p}(n) &= \mathbf{p}(0)\hat{\mathbf{P}} \\ &= [p_1(0) \quad p_2(0)]\begin{bmatrix} \hat{p}_1 & \hat{p}_2 \\ \hat{p}_1 & \hat{p}_2 \end{bmatrix} \\ &= [\hat{p}_1(p_1(0) + p_2(0)) \quad \hat{p}_2(p_1(0) + p_2(0))] \\ &= [\hat{p}_1 \quad \hat{p}_2], \end{aligned}$$

since the components of $\mathbf{p}(0)$ must add to one. Thus

$$\mathbf{p}(n) \xrightarrow[n\to\infty]{} \hat{\mathbf{p}}$$

for an arbitrary initial probability distribution. Under these circumstances we say that the distribution of X_n approaches a **steady-state distribution** which coincides with the stationary probability vector $\hat{\mathbf{p}}$.

The question arises as to what conditions guarantee the approach to a steady-state distribution. Before stating the main result we make the following definition.

Definition A Markov chain is regular if there is a finite positive integer m such that after m time-steps, every state has a non-zero chance of being occupied, no matter what the initial state.

Notation

If every element a_{jk} of a matrix $\mathbf{A}$ satisfies the inequality

$$a_{jk} > 0$$

then we write

$$\mathbf{A} > \mathbf{0}.$$

Thus, for a regular Markov chain with transition matrix $\mathbf{P}$, there exists an $m > 0$ such that

$$\mathbf{P}^m > \mathbf{0}.$$

In Exercise 16 it is proved that

$$\mathbf{P}^m > \mathbf{0} \Rightarrow \mathbf{P}^{m+k} > \mathbf{0}, \qquad k = 1, 2, \ldots$$

For regular Markov chains we have the following result concerning steady-state distributions.

Theorem 8.7 Let $X = \{X_0, X_1, \ldots\}$ be a regular temporally homogeneous Markov chain with a finite number M of states and transition matrix P. Then,

(i) Regardless of the value of $j = 1, 2, \ldots, M$,

$$\lim_{n \to \infty} p_{jk}^{(n)} = \hat{p}_k, \qquad k = 1, 2, \ldots, M.$$

or equivalently,

(ii) $$\lim_{n \to \infty} \mathbf{P}^n = \hat{\mathbf{P}},$$

where $\hat{\mathbf{P}}$ is a matrix whose rows are identical and equal to the probability vector

$$\hat{\mathbf{p}} = [\hat{p}_1 \hat{p}_2 \cdots \hat{p}_M].$$

(iii) No matter what the probability distribution p(0) of X_0, the probability distribution of X_n approaches $\hat{\mathbf{p}}$ as $n \to \infty$:

$$\mathbf{p}(n) \xrightarrow[n \to \infty]{} \hat{\mathbf{p}}$$

(iv) $\hat{\mathbf{p}}$ is the unique solution of

$$\hat{\mathbf{p}}\mathbf{P} = \hat{\mathbf{p}}$$

satisfying $\hat{\mathbf{p}} > \mathbf{0}$ and $\sum_k \hat{p}_k = 1$.

For a proof see Kemeny and Snell (1960) or Feller (1968). Note that in the terminology of Feller a regular Markov chain is irreducible, aperiodic and has only ergodic states. The terminology for Markov chains is confusing as different authors use the same word with different meanings as well as several different words for the same thing. It seemed best to avoid these altogether in an introductory treatment. A matrix A satisfying $\mathbf{A}^m > \mathbf{0}$ for some positive integer m is called **primitive**. The theory of such matrices is well developed, including the useful Perron–Frobenius theorems. See, for example, Seneta (1983).

Example 1

For the population genetics example of the previous section

$$\mathbf{P} = \tfrac{1}{16}\begin{bmatrix} 9 & 6 & 1 \\ 4 & 8 & 4 \\ 1 & 6 & 9 \end{bmatrix}.$$

Since $\mathbf{P} > \mathbf{0}$ we see from Theorem 8.7 that a steady-state probability distribution will be approached as $n \to \infty$. To obtain the stationary distribution we must find a left eigenvector of P with eigenvalue 1 whose components add to unity.

Any eigenvector $\mathbf{x} = [x_1 \quad x_2 \quad x_3]$ with eigenvalue 1 must satisfy $\mathbf{xP} = \mathbf{x}$. Hence

$$[x_1 \quad x_2 \quad x_3]\begin{bmatrix} 9 & 6 & 1 \\ 4 & 8 & 4 \\ 1 & 6 & 9 \end{bmatrix} = 16[x_1 \quad x_2 \quad x_3].$$

This yields three equations, of which only two are needed. The first two equations are

$$9x_1 + 4x_2 + x_3 = 16x_1$$
$$6x_1 + 8x_2 + 6x_3 = 16x_2,$$

or

$$7x_1 - 4x_2 - x_3 = 0$$
$$-6x_1 + 8x_2 - 6x_3 = 0.$$

Since one of the components of x is arbitrary we may set $x_3 = 1$ and solve

$$7x_1 - 4x_2 = 1$$
$$-6x_1 + 8x_2 = 6.$$

This yields $x_1 = 1, x_2 = 3/2$ and $x_3 = 1$ so any left eigenvector of P with eigenvalue 1 is a non-zero multiple of

$$\mathbf{x} = [1 \quad 3/2 \quad 1].$$

The sum of the components of x is 7/2 so dividing x by 7/2 we obtain the required stationary probability vector

$$\hat{\mathbf{p}} = [2/7 \quad 3/7 \quad 2/7].$$

For any initial probability vector p(0), the probability distribution of X_n approaches $\hat{\mathbf{p}}$. In particular, even if the population starts with say, all A_1A_1, so

$$\mathbf{p}(0) = [0 \quad 0 \quad 1],$$

there is probability 3/7 that the population will eventually be heterozygous. Compare this behaviour with random drift.

Example 2

This example, from Ash (1970), shows that a stationary distribution may exist but this does not imply that a steady-state is approached as $n \to \infty$.

Consider a Markov chain with two states and

$$\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

so that transitions are only possible from one state to the other. Solving

$$\mathbf{xP} = \mathbf{x}$$

or,

$$[x_1 \quad x_2]\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} = [x_1 \quad x_2]$$

gives $x_2 = x_1$. Hence

$$\hat{\mathbf{p}} = [\tfrac{1}{2} \quad \tfrac{1}{2}],$$

is a stationary probability vector. However, as $n \to \infty$, $\mathbf{P}^n$ does not approach a constant matrix because

$$\mathbf{P}^n = \begin{cases}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}, & n = 1,3,5,\ldots\\ \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, & n = 2,4,6,\ldots\end{cases}$$

The conditions of Theorem 8.7 are violated, this not being a regular Markov chain. It is seen that state 2 can only be entered from state 1 and vice versa on time steps $1,3,5,\ldots$. Such a Markov chain is called **periodic** or **cyclic** with period 2. For a discussion of such Markov chains see Feller (1968).

REFERENCES

Ash, R.B. (1970). *Basic Probability Theory.* Wiley, New York.
Bailey, N.T.J. (1964). *The Elements of Stochastic Processes.* Wiley, New York.
Crow, J.F. and Kimura, M. (1970). *An Introduction to Population Genetics Theory.* Harper and Row, New York.
Ewens, W.J. (1979). *Mathematical Population Genetics.* Springer-Verlag, New York.
Feller, W. (1968). *An Introduction to Probability Theory and its Applications.* Wiley, New York.
Isaacson, D.L. and Madsen, R.W. (1976). *Markov Chains Theory and Applications.* Wiley, New York.
Kemeny, J.G. and Snell, J.L. (1960). *Finite Markov Chains.* Van Nostrand, Princeton, N.J.
Seneta, E. (1983). *Nonnegative Matrices and Markov Chains*, Springer-Verlag, New York.
Strickberger, M.W. (1968). *Genetics.* Macmillan, New York.
Watson, J.D. (1970). *Molecular Biology of the Gene*, Benjamin, New York.

Additional works including advanced and specialized treatments not referenced in the text

Bartholomew, D.J. (1967). *Stochastic Models for Social Processes.* Wiley, London.
Chung, K.L. (1967). *Markov chains with Stationary Transition Probabilities.* Springer-Verlag, New York.

Cox, D.R. and Miller, H.D. (1965). *The Theory of Stochastic Processes.* Wiley, New York.
Karlin, S. and Taylor, H.M. (1975). *A First Course in Stochastic Processes.* Academic Press, New York.
Kemeny, J.G., Snell, J.L. and Knaff, A.W. (1966). *Denumerable Markov Chains.* Van Nostrand, Princeton, N.J.
Parzen, E. (1962). *Stochastic Processes.* Holden-Day, San Francisco.

EXERCISES

1. A gene is present in human populations which has two alleles A_1 and A_2. If a group initially has 40 A_1A_1, 30 A_1A_2 or A_2A_1 and 30 A_2A_2 individuals, what will the equilibrium (HW) genotype frequencies be?
2. Complete the proof of Theorem 8.2; that

$$\Pr(X_0 = s_{j_0}, X_1 = s_{j_1}, \ldots, X_n = s_{j_n}) = P_{j_0}(0)p_{j_0j_1}(1)\cdots p_{j_{n-1}j_n}(n)$$

for $n \geqslant 1$. (*Hint*: Use mathematical induction.)
3. Establish the **Chapman–Kolmogorov equations**

$$p_{jk}^{(m+n)} = \sum_{i=1}^{M} p_{ji}^{(m)} p_{ik}^{(n)}.$$

(*Hint*: Use matrix multiplication.)
4. Show that the matrix with elements given by (8.13) is stochastic.
5. Any stochastic matrix defines a temporally homogeneous Markov chain. Which of the following matrices are stochastic?

(a) $\begin{bmatrix}1/4 & 3/4\\ 1 & 0\end{bmatrix}$ (b) $\begin{bmatrix}1/2 & 1/2 & 0\\ 1/2 & 1/4 & 0\\ 1 & 0 & 0\end{bmatrix}$

(c) $\begin{bmatrix}0 & 1 & 0 & 0\\ 1/3 & 1/3 & 0 & 1/3\\ 1/2 & 1/4 & 1/8 & 1/8\\ 1/2 & 0 & 1/4 & 1/4\end{bmatrix}$

6. For the Markov chain for random mating with no mutation, the transition matrix when $N = 1$ is

$$\mathbf{P} = \begin{bmatrix}1 & 0 & 0\\ 1/4 & 1/2 & 1/4\\ 0 & 0 & 1\end{bmatrix}$$

If X_0 has the distribution $\mathbf{p}(0) = [0 \quad \tfrac{1}{2} \quad \tfrac{1}{2}]$, find the probability

distributions of X_1, X_2, X_3 and X_4. Plot these distributions and observe the phenomenon of random drift.

7. The matrix $\mathbf{\Pi}$ of absorption probabilities for the simple random walk with absorbing barriers at 0 and 3 was found to be

$$\mathbf{\Pi} = \frac{1}{1-pq}\begin{bmatrix} q & p^2 \\ q^2 & p \end{bmatrix}.$$

Verify that
(a) the row sums of $\mathbf{\Pi}$ are unity
(b) the probabilities of absorption at 0 agree with those given by (7.17).

8. For the genetic Markov chain (Section 8.3) with a population of N diploid individuals, find the matrices $\mathbf{Q}$ and $\mathbf{R}$. Verify that the matrix $\mathbf{\Pi}$ of absorption probabilities

$$\mathbf{\Pi} = \begin{bmatrix} 1-1/2N & 1/2N \\ 1-2/2N & 2/2N \\ \vdots & \vdots \\ 1-k/2N & k/2N \\ \vdots & \vdots \\ 1/2N & 1-1/2N \end{bmatrix}$$

satisfies

$$(\mathbf{I}-\mathbf{Q})\mathbf{\Pi} = \mathbf{R}.$$

9. Prove that the Markov chain $\{X_n\}$ for random genetic drift defined in Section 8.3 is a martingale. (cf. Exercise 14 of Chapter 7.) Use the optional stopping theorem to deduce immediately that the probabilities of fixation are given by (8.22).
10. For the simple random walk with absorbing barriers at 0 and 3, verify that the formulas

$$E(T_1) = \frac{1+p}{1-pq}, \quad E(T_2) = \frac{1+q}{1-pq},$$

for the expected times to absorption from $X_0 = 1$, $X_0 = 2$, respectively, agree with those given by (7.23).

11. The following problem is based upon one in Kemeny and Snell (1960). In each year of a three-year degree course, a university student has probability p of not returning the following year, probability q of having to repeat the year and probability r of passing ($p+q+r=1$). The states are: dropped out (s_1), graduated (s_2), is a third-year student (s_3), is a second-year student (s_4), and is a first-year student (s_5). Find the transition matrix $\mathbf{P}$ and the matrices $\mathbf{Q}$ and $\mathbf{R}$. (Note that this is a random walk with absorbing barriers.)

12. For the Markov chain of Exercise 11, solve the equations $(\mathbf{I}-\mathbf{Q})\mathbf{\Phi} = \mathbf{I}$ to obtain the fundamental matrix $\mathbf{\Phi} = (\mathbf{I}-\mathbf{Q})^{-1}$.
13. For the Markov chain of Exercise 11, find a student's chances of graduating if he is in years 1, 2 and 3.
14. For the Markov chain of Exercise 11, find the average number of years a first-year, second-year and third-year student will remain in university.
15. The following example is based on an application discussed in Isaacson and Madsen (1976). Farms are divided into four categories: very small (s_1), very large (s_2), large (s_3) and small (s_4). Farms grow or shrink as land is bought or sold. It is assumed that once a farm is very small or very large it stays as such. Small and large farms increase in size each year into the next category with probability $\frac{1}{2}$, remain the same size with probability $\frac{1}{4}$ and decrease in size to the category below with probability $\frac{1}{4}$. Find the transition matrix and the expected time for a small farm to become either very small or very large.
16. Prove that $\mathbf{P}^m > \mathbf{0} \Rightarrow \mathbf{P}^{m+k} > \mathbf{0}$, $k = 1, 2, \ldots$
17. The following learning model, due to Bush and Mosteller, is discussed in Bailey (1964). In a learning experiment let s_1 be a correct response and s_2 an incorrect response. The response at any trial depends only on the result of the previous trial and the transition matrix is

$$\mathbf{P} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}, \qquad 0 < p, q < 1.$$

Let X_n be the response on trial n, $n = 0, 1, 2, \ldots$
(a) Find the stationary probability vector $\hat{\mathbf{p}}$.
(b) Will the probability distribution of X_n approach $\hat{\mathbf{p}}$ as $n \to \infty$?
(c) Find the matrix $\hat{\mathbf{P}}$.
(d) Prove, using induction or otherwise, that

$$\mathbf{P}^n = \frac{1}{p+q}\begin{bmatrix} q & p \\ q & p \end{bmatrix} + \frac{(1-p-q)^n}{p+q}\begin{bmatrix} p & -p \\ -q & q \end{bmatrix}.$$

Hence verify your result (c).
(e) If the initial response is correct so $\mathbf{p}(0) = [1 \quad 0]$, what is the probability of a correct response on trial n?

18. For a simple random walk assume there are reflecting barriers at 0 and 3. That is, when the particle gets to 0 or 3 it goes on the next step to states 1 or 2 (respectively) with probability one. Thus the transition matrix is

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

(a) Without doing any calculations, is this a regular Markov chain? Why?
(b) If the answer to (a) is yes, compute the equilibrium probability distribution $\hat{\mathbf{p}}$.
(c) If $X_0 = 3$, what is the eventual probability that the position of the particle is 3?

19. In the Markov chain model of random mating with mutation in a population of size N, find $\mathbf{P}$ if $\alpha_1 = \alpha_2 = \alpha \neq 0$. Given an arbitrary initial probability distribution $\mathbf{p}(0)$, find $\mathbf{p}(1)$ and deduce that the stationary distribution is attained in one generation.
20. What will happen in the Markov chain model of random mating with mutation if $\alpha_1 \neq 0$ but $\alpha_2 = 0$?

9

Population growth I: birth and death processes

9.1 INTRODUCTION

It is clearly desirable that governments and some businesses be able to predict future human population numbers. Not only are the total numbers of male and female individuals of interest but also the numbers in certain categories such as age groups. The subject which deals with population numbers and movements is called **demography**.

Some of the data of concern to human demographers is obtained from our filling out census forms. The type of data is exemplified by that in Tables 9.1 and 9.2. In Table 9.1 is given the total population of Australia at various times since 1881 and Table 9.2 contains some data on births and deaths and their rates. Notice the drastic fall in the birth rate in the last few decades compared with an almost steady death rate.

Table 9.1

Time	Population of Australia (thousands)*
3 April 1881	2250.2
5 April 1891	3177.8
31 March 1901	3773.8
3 April 1911	4455.0
4 April 1921	5435.7
30 June 1933	6629.8
30 June 1947	7579.4
30 June 1954	8986.5
30 June 1961	10548.3
30 June 1966	11599.5
30 June 1971	12937.2
30 June 1979	14417.2
30 June 1981	14923.3

*Obtained from Cameron (1982) and Australian Bureau of Statistics.