RESPONSE   TO   SELECTION

IN

FINITE   POPULATIONS

by

PREM   NARAIN

B.Sc. (LUCKNOW)

B.Sc.(Hons)(LUCKNOW)

M.Sc. (LUCKNOW)

Thesis presented for the Degree of
Doctor of Philosophy of the
University of Edinburgh in the
Faculty of Science.

Institute of Animal Genetics,

EDINBURGH   May, 1969

CONTENTS

Contents (continued)

Contents (continued)

SUMMARY

1.   A theoretical investigation was made on (a) the limits and duration of response to selection and (b) the properties of lethal mutants and their expectation of life in small populations. The problems were formulated mathematically in terms of transition matrices and in some cases in terms of diffusion approximations. In a few cases explicit formulae were obtained by expanding matrices and solving differential equations. In most of the cases, however, numerical solutions were obtained by evaluating matrix functions on the computer.

2.   In the case of selection at a di- and tri- allelic locus, the transition probabilities between the segregating states of a population of gametes of constant size 2N, determine a Q-matrix. The average of the total number of times the population spends in the different segregating states on the way to fixation, corresponding to various initial states are determined by a fundamental matrix $(\underset{\sim}{I} - \underset{\sim}{Q})^{-1}$ where $\underset{\sim}{I}$ is a unit matrix. The expected change in the gene frequency of a particular allele in the limit and mean and variance of the number of generations until fixation of this allele are found to depend on the fundamental matrix. The frequency distribution of gene frequency at equilibrium between mutation and selection at a diallelic locus is also found to depend on the elements of the first row of the fundamental matrix.

3.   With binomial transition probabilities and selection, the expansion of matrix formulae gave explicit expressions for the expected change in the gene frequency in the limit, and for the half-life of the selection process. For large N and small values of selective coefficient (s), such that Ns is constant, the expression for the expected change in the gene frequency in the limit was found to be the same as the diffusion approximation formula of the chance of fixation expanded in powers of Ns. In the case of artificial selection based on individual measurements the total advance by selection for additive genes, is $2NI(1 + C_1 I + C_2 I^2)$, where I is the gain in the first generation, $C_1 = \frac{2}{3}N \mu_{3g}/\sigma_g^4$ and $C_2 = -\frac{2}{3}N^2(1 + n^2 V^*/\sigma_g^4)/n\sigma_g^2$. Here σ_g^2 and μ_{3g} are respectively the variance and the third moment (about the origin) of the breeding values. V^* refers to the variance between the contributions of the n different loci to the additive genetic variance of the character. The half-life of the selection process was found to be a function of Ns only. When $Ns = 1$ and initial gene frequency is half, the half-life is about $1.30N$ but for initial gene frequency tending to 0 and 1, the half-lives are $1.58N$ and $1.03N$ respectively. It was further proved that the moments of the distribution of time until fixation of a gene are independent of terms in s.

4.   The diffusion approximation for the variance of time until fixation, for a selectively neutral gene and initial gene frequency tending to zero, is found to be $16N_e^2(\frac{\pi^2}{3} - 3)$ where N_e is the effective size of the population. With the mean time as $4N_e$, this amounts to a coefficient of variation of about 54 per cent.

5.   Numerical results on the mean and the coefficient of variation of time until fixation of a gene, at a diallelic locus, obtained by the use of transition matrices suggested that these are functions of Ns only. Selection decreases the mean and the coefficient of variation of time until fixation of a gene with additive effects. With rise in the initial gene frequency, the mean decreases but the coefficient of variation increases. Selection decreases the mean time until fixation of a recessive gene also. For a dominant gene, however, the mean time increases when Ns is small, attains a maximum and then decreases. A gene takes less time, on an average, to reach fixation when it is recessive than when it is dominant. When there is no selection and the two alleles are equally represented in the initial population, the mean time until homozygosity is found to be maximum but the coefficient of variation of time until homozygosity happens to be minimum. With selection the maximum, in the case of the mean and the minimum, in the case of the coefficient of variation, are found to occur at initial gene frequencies less than half. As N becomes large and initial gene frequency tends to zero, the mean and the coefficient of variation of time until homozygosity approaches the mean and the coefficient of variation of time until loss respectively.

6.   Numerical results on chance of fixation of a particular allele A_1 at a tri-allelic locus $(A_1-A_2-A_3)$ obtained by the use of the transition matrices suggested that it is/on $N\bar{s}$ and N^2V_m only dependent where $\bar{s}$ is the average superiority of A_1 over A_2 and A_3 lumped together and V_m is the variance in superiority, (multiple allelic variance). The chance of fixation increases as $N\bar{s}$ increases in almost

the same manner as in the case of two alleles at a locus. It, however, decreases as N^2V_m increases. The mean time until fixation of A_1 decreases as $N\bar{s}$ increases provided $N^2V_m \leq 1$. At higher values of N^2V_m it increases with increase in $N\bar{s}$, attains a maximum and then decreases. The effect of increasing N^2V_m on the mean time is to decrease it provided $N\bar{s} \leq 1$. At higher values of $N\bar{s}$, it increases with increase in N^2V_m, attains a maximum and then decreases. The mean time until homozygosity is found to be maximum when there is no selection and the three alleles are equally represented in the initial population. This maximum is found to be greater than the corresponding maximum in the two allele case.

7. The mean and variance of the distribution of frequencies of lethal mutants at equivalent loci in a finite population depends on the mutation rate per locus, the population size and the selective advantage/disadvantage of the heterozygotes carrying lethals. Assuming a mutation rate of 10^{-5} , the mean and variance declines as the heterozygotes declines in their fitness from 1.05 to 0.95. At any given value of the selective coefficient of the heterozygote in this range the mean is larger for a population size of 50 than that of 10, but the variance behaves in this manner only if the values of selective coefficient between 1.05 to 0.00 are considered. In other cases it is smaller for $N = 50$ than for $N = 10$. The mean heterozygosity behaves in a similar manner as the mean.

8.   Assuming 500 loci on lethal bearing chromosomes on which lethal mutations can occur so that the possibility of the chance of allelism due to recurrent mutation is practically nil and a mutation rate of 10^{-6} per locus, the chance of allelism due to identity of genes by common descent have been studied. It is found that it increases as the heterozygotes decrease in fitness at a particular population size. It is larger for the smaller population size and for very small population size it may not depend on the heterozygote's fitness. The allelism of lethal genes declines with time. The decline is sharper for a smaller population size. Heterotic lethals do not show as rapid a decline as the completely or partially recessive lethals. The initial rate of decline is found to be about 4 to 5 per cent for completely recessive lethals. This rate agrees well with that predicted by the formula of rate of decline derived with the help of the distribution of lethal genes based on diffusion approximation. For a population of size N and a chromosomal mutation rate of U, the formula for the initial rate of decline is found to be

$$2 \pi NU \left(\frac{\pi}{8N}\right)^{\frac{1}{2}} / [1 - \left(\frac{\pi}{8N}\right)^{\frac{1}{2}} + 2\pi NU].$$

9.   The lethal mutants present in a population at any given time have an average age since they first appeared in the population. This is equivalent to the average time which these mutants would spend before disappearing from the population, in other words, their average expectation of life. It depends on the fitness of the heterozygote

and is very high for heterotic lethals but declines sharply as the heterozygote tends to be neutral and then disadvantageous. The expectation of life of a new lethal mutant at its initial occurrence is always smaller than its expectation of life at any subsequent time.

CHAPTER I

GENERAL INTRODUCTION

Selection changes the genetic make up of a population by changing
its gene frequencies. In an infinite population it leads ultimately to
the fixation of the favoured allele unless there is heterozygote
advantage. But in a finite population, the gene frequency undergoes a
random change also from generation to generation. This results in a
distribution of gene frequencies which can be regarded either as the
distribution of frequencies at equivalent loci in one population or as
the distribution of frequencies at a single locus replicated in many
equivalent populations (Wright 1931). As time proceeds this dis-
tribution gets broadened with irreversible fixation (or loss) of gene
leading to a state of steady decay when the distribution curve attains
a constant form. The height of the curve then decreases at a constant
rate and becomes zero in the limit. The expected frequency of the
favoured allele in the limit is the same as the chance of its fixation
(Kimura 1957, 1962). This chance of fixation can be regarded either
as the proportion of equivalent loci which would be expected to be
fixed in the limit in any line or as the proportion of replicate
selected lines in which an individual gene would be expected to be
fixed in the limit. The limit of response to selection in a finite
population is measured by the difference between the chance of fixation

of the favoured gene and its initial frequency. The determination of
the chance of fixation of a gene is therefore the basic problem in
the study of limits of response to selection in a finite population.

For a single locus with two alleles Kimura (1957) gave a formula
for the chance of fixation. His method was based on Kolmogorov's
backward diffusion equation. Robertson (1960), Ewens (1963), Allan and
Robertson (1964), Hill and Robertson (1966), Hill and Robertson (1968)
and others used methods based on transition matrices and determined the
chance of fixation numerically with the help of a computer. In the
present study a general theory of the transition matrix approach has
been developed which can give the chance of fixation as well as the
expected change in the gene frequency by a given time in any genetic
situation. The formulae are, however, in terms of the matrices. For the
case of single locus with two alleles and binomial transition
probabilities the matrix formulae have been expanded, giving a formula
for the chance of fixation expressed as a series. Under the conditions
in which diffusion approximation holds, this formula reduces to Kimura's
formula expressed as a series. For the case of single locus with three
alleles, the chance of fixation of a gene is not known. Using
trinomial transition probabilities this has been determined on a
computer and presented in this study. Robertson (personal communication)
has, however, expanded the matrix formula of the chance of fixation for
a multi-allelic locus and given an approximate series formula. His
formula has been compared with the exact computer results for a

tri-allelic locus.

Since the limit of response to selection is attained asymptotically, the time to attain the limit is expressed in terms of half-life i.e. the time by which the expected gene frequency gets half-way to the limit (Robertson 1960). In the present study, a more accurate formula for half-life than known earlier has been developed.

The fixation or loss of genes occur after a variable number of generations. Hence the distribution of time to fixation of a particular allele disregarding the cases in which it is lost can describe the nature of the life of a gene until it is fixed. In particular, the mean and variance of this distribution are of interest. The general theory of the transition matrix approach provide with matrix formulae for the determination of the mean and variance of time until fixation of a gene in any genetic situation. For the case of a single locus with two alleles, computer results have been presented for the mean and the coefficient of variation of time until fixation of a gene. Kimura and Ohta (1968) have recently given formulae for the mean time until fixation from the diffusion approach. Their results for a selectively neutral single mutant introduced in the population have been compared with the computer results. They have, however, not derived formula for the variance of the time until fixation. This has been done in the present investigation. For the case of single locus with three alleles, the mean time until fixation of a gene is not known. Computer results have, therefore, been presented for this case.

The distribution of time until homozygosity i.e. either fixation or loss of the gene describes the nature of the life of a gene until it is either fixed or lost. When the chance of fixation of a gene is very near to unity, the two distributions may coincide. It is therefore of some interest to compare the mean and variance of time until fixation of a gene with those until homozygosity. For the case of single locus with two alleles this comparison has been presented. Ewens (1963) gave computer results as well as diffusion approximations for the mean time until homozygosity but not the variance of time until homozygosity. For the case of single locus with three alleles, the mean time until homozygosity is not known. Computer results for the same have, therefore, been presented.

Selection with random fluctuation of gene frequencies leads to fixation or loss of genes in a population. But reversible mutation or migration tends to restore the intermediate gene frequencies. This results in a stable distribution of gene frequencies. In any given case, a knowledge of the mean and the variance of small changes in gene frequency gives this stationary distribution with the help of a general formula given by Wright (1938, 1945). This formula is based on Kolmogorov's forward diffusion equation. However, for the case of balance between mutation and selection with random drift, it has been proved that the stable distribution can also be given by the means of the total number of times the population spends in the different transient states on the way to fixation from an initial state in which

the mutant individual is represented once only. These means are given by the elements of the first row of a function of the transition matrix. This approach can therefore be thought of as the transition matrix approach for the determination of the stable distribution and their properties. Ewens (1964), however, described the distribution based on the mean time spent in an interval before absorption as a pseudo-transient distribution.

If we consider a deleterious gene in a large population, mutation opposed by moderately severe selection would tend to keep the gene at a low equilibrium frequency determined by the mutation rate, selective coefficient and the degree of dominance of the gene. But if we take small samples from such a population with dominance lacking and the same degree of severity of selection is considered, the mean frequency of the gene rises to the equilibrium value expected on the basis of reversible mutations alone. If unfavourable mutation is much more frequent than the reverse, this may lead to approximate fixation of the gene. In the case of a recessive lethal the effect of this bottle-neck due to reduction in the population size is, however, a considerable decrease in the mean frequency of lethal gene. The distribution of lethal gene frequencies is given by Wright (1937) whereas the distribution of lethal chromosomes is derived by Nei (1968). In the present study the properties of the lethal gene frequency distribution has been studied from the transition matrix approach and compared with the properties known from the diffusion approach. In the case of

heterotic and partially recessive lethals, however, the present approach gives results which were not known previously.

In the case of lethal genes an important problem is the determination of the chance of allelism. This can be regarded as the proportion of crosses between lethal bearing chromosomes which produce lethal zygotes. Allelic rates have been found useful in discriminating, from experimental data on <u>Drosophila</u> populations, whether recessive lethals extracted from natural populations have a deleterious effect as heterozygotes or not (Crow and Temin 1964). Dobzhansky and Wright (1941), and Wright, Dobzhansky and Hovanitz (1942) gave a formula for the determination of allelic rates. Nei (1968) modified this formula. Here Nei's formula has further been modified to deal with an infinite number of loci on lethal bearing chromosomes. Allelic rates have been determined using transition matrix approach, in the case of heterotic, completely recessive and partially recessive lethals.

Another interesting problem, in the case of lethal genes, is the decline of allelism of recessive lethals extracted from natural populations at different times (Wallace 1966). Wallace suggested a functional relationship between the allelic rate and the time interval between the sampling of the first and the second set of lethals. Prout (1967) proved this relationship to be approximately correct by using recurrence relations between the gene frequencies in the successive generations subject to systematic and random pressures. Using transition matrices an alternative way of predicting the decline of

allelism with time has been suggested by Robertson (personal communication). On this basis, the decline of allelism with time has been studied in the present investigation. The initial rate of decline has also been worked out with the help of Wright's equilibrium distribution formula.

The expectations of life of a lethal mutant at its initial occurrence as well as at any given time when it is represented in the population more than once have further been investigated in this study. These do not appear to have been studied previously.

The plan to be followed in the succeeding chapters is this: a general theory of the transition matrix approach for the study of (a) response to selection in a finite population and (b) the gene frequency distribution at equilibrium between selection and mutation is given in Chapter II. For a single locus with two alleles and binomial transition probabilities, matrix formulae are expanded in Chapter III. Diffusion approximation for the variance of time until fixation has also been discussed in this chapter. This is followed by Chapter IV dealing with numerical results on the mean and the coefficient of variation of time until fixation using binomial transition probabilities. The numerical results on selection with trinomial transition probabilities are given in Chapter V. In the last Chapter, the properties of lethal gene distribution, allelism, and the expectation of life of a lethal mutant are presented.

CHAPTER II

THEORY OF TRANSITION MATRIX APPROACH

In this section the basic theory of the transition matrix
approach is developed from the first principles. First only selection
in a finite population is considered. Matrix formulae are developed
for the fixation probabilities, the expected changes in the gene
frequency and the moments of the distribution of time to fixation for a
single locus with two alleles and three alleles. Next mutation at a
low rate is introduced in the case of single locus with two alleles to
compensate for the homozygosity due to random elimination. The resul-
ting stable distribution is expressed as the elements of a certain row
of the fundamental matrix, and the properties of the distribution are
developed.

2.1   SELECTION AND RANDOM DRIFT

2.11   DIALLELIC LOCUS

Consider a finite population of gametes of constant size $2N$ and
a single locus with two alleles A_1 and A_2. Such a population can
assume $(2N + 1)$ states E_o, E_1, ... E_{2N}, the ith state E_i
representing the state of i A_1 genes and $(2N - i)$ A_2 genes. The
states E_o and E_{2N} represent respectively the state of A_2 and A_1

genes entirely and therefore once the population assumes these states, it gets fixed for either A_2 or A_1 allele. In Markov chain terminology, such states are known as absorbing states. On the other hand, any other state E_i, $i = 1,2, \ldots (2N-1)$, represents a mixture of A_1 and A_2 genes and therefore once the population is in this state, it has a possibility of going out of this state to any other state including the absorbing ones. In other words, this means E_i state represents a state segregating for A_1 and A_2 genes with proportions $q_i = i/(2N)$ and $(1 - q_i)$ respectively. These are known as transient states. Suppose P_{ij} represents the conditional probability that there are j A_1 genes out of $2N$ genes after one generation, given that there were i A_1 genes out of $2N$ genes in the previous generation. Since there are $(2N + 1)$ possibilities in the previous as well as in this generation, we have $(2N + 1)\times(2N + 1)$ P_{ij}'s which can be conveniently represented by a matrix P as given below. Also, if all the genes are either of A_1 or A_2 types, P_{oo} and $P_{2N,2N}$ will each be one but P_{oj} and $P_{2N,j}$ will each be zero, where j can have any value between 1 and $(2N - 1)$. So

$$(2.1) \quad \underset{\sim}{P} = \begin{bmatrix} 1 & 0 & . \; . & 0 & 0 \\ P_{10} & P_{11} & . \; . & P_{1,(2N-1)} & P_{1,2N} \\ \cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot \\ P_{(2N-1),0} & P_{(2N-1),1} & . \; . & P_{(2N-1),(2N-1)} & P_{(2N-1),2N} \\ 0 & 0 & . \; . & 0 & 1 \end{bmatrix}$$

Suppose we consider transitions between the transient states only, then the transition probability matrix can be represented by $\underset{\sim}{Q}$ as given below.

$$(2.2) \qquad \underset{\sim}{Q} = \begin{bmatrix} P_{11} & \cdots\cdots & P_{1,(2N-1)} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ P_{(2N-1),1} & \cdots\cdots & P_{(2N-1),(2N-1)} \end{bmatrix}$$

If $\underline{P}'_o$, $\underline{P}'_{2N}$ and $\underline{0}'$ denote the row vectors as given below

$$(2.3) \qquad \underline{P}'_o = [P_{10}, P_{20}, \cdots P_{(2N-1),0}]$$

$$(2.4) \qquad \underline{P}'_{2N} = [P_{1,2N}, P_{2,2N}, \cdots P_{(2N-1),2N}]$$

$$(2.5) \qquad \underline{0}' = [0, 0, \cdots 0]$$

then $\underset{\sim}{P}$ can be written in a partitioned form as

$$(2.6) \qquad P = \begin{bmatrix} 1 & 0 & \underline{0}' \\ 0 & 1 & \underline{0}' \\ \underline{P}_o & \underline{P}_{2N} & \underset{\sim}{Q} \end{bmatrix}$$

Now suppose we consider the transition probabilities after t-generations and denote them by $P_{ij}^{(t)}$ with the corresponding $\underset{\sim}{Q}$ matrix as $\underset{\sim}{Q}(t)$. Then we know from the theory of finite Markov chains

(Kemeny and Snell 1960) that

(2.7) $\underset{\sim}{Q}(t) = \underset{\sim}{Q}^t$

Now consider the matrix sum

(2.8) $\underset{\sim}{T}(t) = \underset{\sim}{I} + \underset{\sim}{Q} + \underset{\sim}{Q}^2 + \ldots + \underset{\sim}{Q}^t$

where $\underset{\sim}{I}$ is a unit matrix with ones as the diagonal elements and zeros elsewhere. The elements in the ith row of $\underset{\sim}{T}(t)$ give the expected total number of times the population spends in the different transient states by the tth generation, having started from the state E_i. Suppose $\underset{\sim}{U}(t)$ and $\underset{\sim}{R}(t)$ denote the column vectors of the fixation probabilities and the expected changes in the gene frequency of A_1 respectively by the tth generation, whereas $\underline{P}_{2N}$ and $\underline{\triangle q}$ denote the column vectors of the fixation probabilities and the expected changes in the gene frequency of A_1 respectively in one step. Now fixation by the tth generation means that starting from E_1, the population assumes the different transient states in the first $(t - 1)$ steps and then achieves fixation in one step from the assumed transient states. This means that the fixation probability is the sum of the expected total number of times the population spends in the different transient state by the $(t - 1)^{th}$ generation, multiplied by the corresponding probability of fixation in one step. That is,

(2.9) $\underline{U}(t) = \underset{\sim}{T}(t - 1)\ \underline{P}_{2N}$

Similar considerations show that

(2.10) $\underline{R}(t) = \underset{\sim}{T}(t - 1)\ \underline{\triangle q}$

Since the $\underset{\sim}{P}$ matrix is a Markov matrix with elements as probabilities, it follows from the matrix theory (Faddeeva, 1958) that the roots of $\underset{\sim}{Q}$ are all positive and less than unity and therefore the inverse of $(\underset{\sim}{I} - \underset{\sim}{Q})$ exists i.e. $|\underset{\sim}{I} - \underset{\sim}{Q}| \neq 0$. Also we have, then

$$(2.11) \quad \underset{\sim}{T}(t - 1) = \underset{\sim}{I} + \underset{\sim}{Q} + \ldots \ldots + \underset{\sim}{Q}^{t-1}$$
$$= (\underset{\sim}{I} - \underset{\sim}{Q}^t)(\underset{\sim}{I} - \underset{\sim}{Q})^{-1}$$

Hence, we have

$$(2.12) \quad \underset{\sim}{U}(t) = (\underset{\sim}{I} - \underset{\sim}{Q}^t)(\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \underline{P_{2N}}$$

$$(2.13) \quad \underline{R}(t) = (\underset{\sim}{I} - \underset{\sim}{Q}^t)(\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \underline{\Delta q}$$

Moreover, as $t \to \infty$, $\underset{\sim}{Q}^t \to \underset{\sim}{0}$, a null matrix with all the elements as zeros, so that if $\underline{U}$ and $\underline{R}$ denote the vectors of the eventual fixation probabilities and the expected changes in the gene frequency of A_1 in the limit (i.e. the selection limit) respectively, these are given by

$$(2.14) \quad \underline{U} = (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \underline{P_{2N}}$$

$$(2.15) \quad \underline{R} = (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \underline{\Delta q}$$

$\underline{U}(t)$ and $\underline{R}(t)$ can now, alternatively, be expressed as

$$(2.16) \quad \underline{U}(t) = (\underset{\sim}{I} - \underset{\sim}{Q}^t) \underline{U}$$

$$(2.17) \quad \underline{R}(t) = (\underset{\sim}{I} - \underset{\sim}{Q}^t) \underline{R}$$

The time which a population, with a given initial gene frequency, takes to get fixed for this gene is a random variable on the hypothesis that fixation for this gene takes place with certainty. It is, therefore,

important to investigate the average number of generations taken for the fixation of the desirable allele disregarding the cases in which it is lost. Since

$$(2.18) \quad \lim_{t \to \infty} T(t) = I + Q + Q^2 + \ldots \ldots$$

$$= (I - Q)^{-1}$$

$$= T$$

it follows that the elements in the ith row of T are the averages of the total number of times the population spends in the different transient states on the way to eventual fixation from an initial state i. This is called the fundamental matrix of the absorbing Markov chain. The row sums of this matrix give the average time to absorption in the absorbing states from various initial states. Here it corresponds to the average time to homozygosity. Expressed in vector notations, this is given by

$$(2.19) \quad m = T e$$

where e is a column vector with all the elements as unity. As the proportion of times that a population goes from a particular state to the fixation of the allele A_1 is given by the elements of the vector U, the vector M given by

$$(2.20) \quad M = (I - Q)^{-1} U$$

$$= (I - Q)^{-2} P_{2N}$$

gives the mean total number of steps needed for the fixation of the allele A_1. The mean time until fixation of A_1 is therefore given by the ratio of the elements of vectors $\underline{M}$ and $\underline{U}$ respectively. If we expand $(\underset{\sim}{I} - \underset{\sim}{Q})^{-2}$ in (2.20) we obtain

$$(2.21) \quad \underline{M} = (\underset{\sim}{I} + 2\underset{\sim}{Q} + 3\underset{\sim}{Q}^2 + \ldots \ldots) \underline{P_{2N}}$$

This shows that we can obtain the various moments of the distribution of time until fixation of the particular allele disregarding the cases in which it is lost, by generalizing this formula. For instance, the second moment is given by the ratio of the elements of vectors $\underline{V}$ and $\underline{U}$ respectively where $\underline{V}$ is given by

$$(2.22) \quad \underline{V} = (\underset{\sim}{I} + 2^2\underset{\sim}{Q} + 3^2\underset{\sim}{Q}^2 + \ldots \ldots) \underline{P_{2N}}$$
$$= [2(\underset{\sim}{I} - \underset{\sim}{Q})^{-3} - (\underset{\sim}{I} - \underset{\sim}{Q})^{-2}] \underline{P_{2N}}$$

The variance can then be obtained by subtracting the square of the mean from it.

The matrix formulae for the moments of the distribution of time until loss of A_1 disregarding the cases in which it is fixed, can be obtained by substituting $\underline{P_o}$ in place of $\underline{P_{2N}}$ in $\underline{U}$, $\underline{M}$, and $\underline{V}$.

An alternative method of deriving the matrix formulae for $\underline{U}$, $\underline{R}$, $\underline{M}$ and $\underline{V}$ is given in the Appendix.

2.12 TRI-ALLELIC LOCUS

Consider a finite population of gametes of constant size $2N$ and a single locus with three alleles A_1, A_2 and A_3. Such a population can assume $(N + 1)(2N + 1)$ states. If the population consists of i_1 A_1 genes, i_2 A_2 genes and $(2N - i_1 - i_2) A_3$ genes, the population is said to be in state E_{i_1, i_2} with $i_1 + i_2$ less than or equal to $2N$. These states can be grouped into three classes depending upon the kind of genes present in the population:

class I : $\quad E_{0,0}$ $\qquad A_1, A_2$ lost i.e. A_3 is fixed

$\qquad\qquad E_{0,2N}$ $\qquad A_1, A_3$ lost i.e. A_2 is fixed

$\qquad\qquad E_{2N,0}$ $\qquad A_2, A_3$ lost i.e. A_1 is fixed

Thus there are three absorbing states

class II : $\quad E_{i_1, i_2}$: $i_1, i_2 > 0$ such $i_1 + i_2 = 2N$, i.e. A_3 is lost

$\qquad\qquad E_{i_1,0}$: $i_1 \neq 0$ or $2N$ $\qquad\qquad$ i.e. A_2 is lost

$\qquad\qquad E_{0,i_2}$: $i_2 \neq 0$ or $2N$ $\qquad\qquad$ i.e. A_1 is lost

Thus there are $3(2N - 1)$ states such that exactly one of the three genes is lost. In these cases the population behaves as described in the previous section

class III : E_{i_1, i_2} : $i_1 \neq 0$, $i_2 \neq 0$ and $i_1 + i_2 < 2N$

Thus there are $(N - 1)(2N - 1)$ states such that all the three genes are present.

These states can be geometrically represented by triangular coordinates as in Figure 1. The states in class I are represented by the vertices of the triangle. The states in class II are represented by the interior points of the sides of the triangle. The states in class III are the interior points of the triangle. The totality of transient states is the sum of states in class II and class III i.e. $(N + 2)(2N - 1)$. There is, however, an important distinction between the transient character of the states in class II and class III. If the population is in one of the three categories of class II states it cannot go to the transient states of the other two categories of this class as well as to states of class III and to one of the absorbing states. For instance if the transient states of the type $E_{i_1, 0}$ are considered the transitions are not possible to states of type E_{i_1, i_2}, E_{0, i_2} of class II, E_{i_1, i_2} of class III and $E_{0, 2N}$. This simplifies the structure of the transition matrix considerably as will be shown below. But if a state belongs to class III, the transitions are possible to all the states of class I, II and III.

The transition probabilities are of the type $P_{(i_1 \, i_2)(j_1 \, j_2)}$ representing the conditional probability that there are j_1 A_1 genes, j_2 A_2 genes out of $2N$ genes after one generation given that there
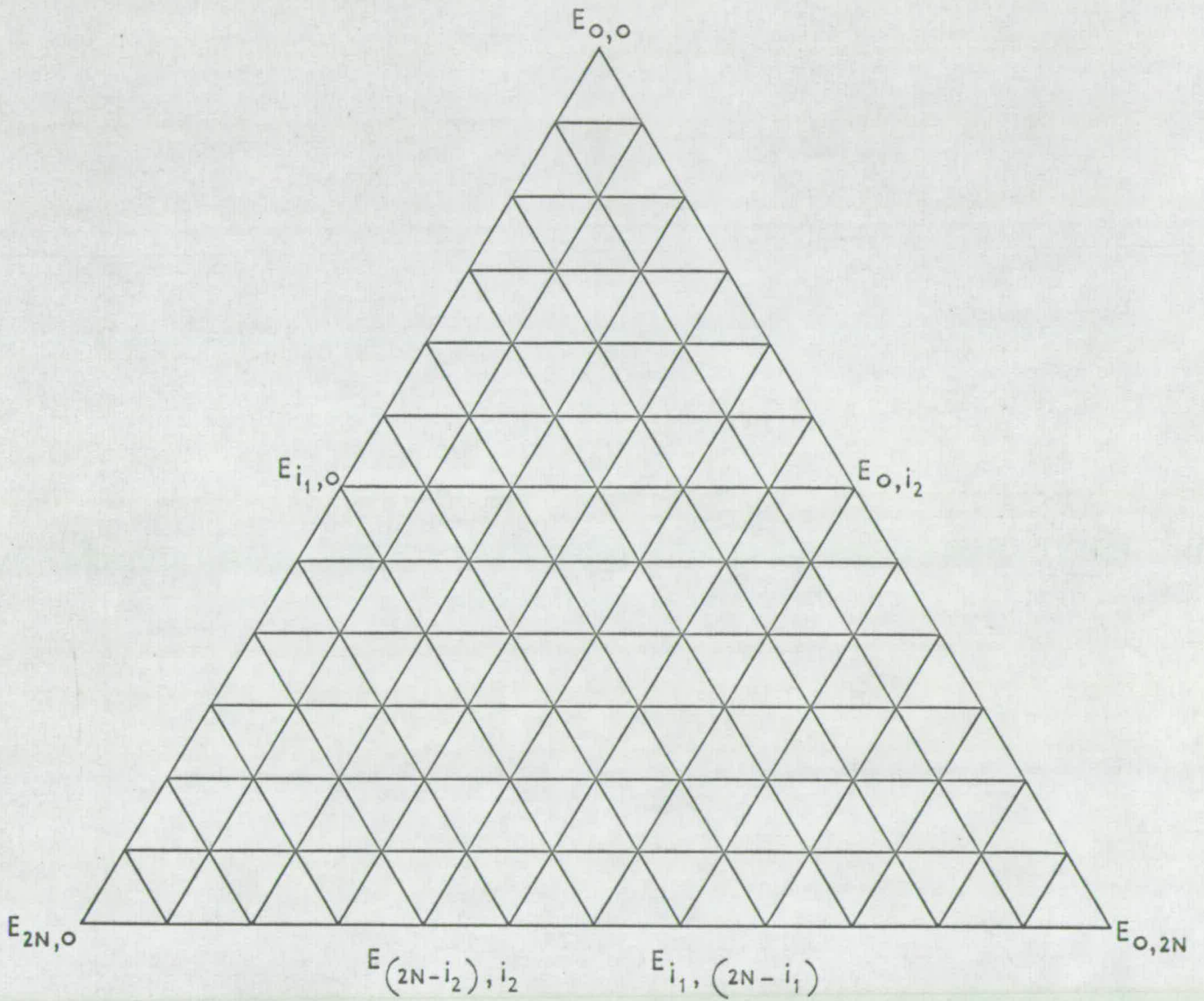
Fig. 1   A geometric representation of the state
space at a tri-allelic locus.

were i_1 A_1 genes and i_2 A_2 genes out of $2N$ genes in the previous generation. Since there are in all $(N + 1)(2N + 1)$ states, there transition probabilities are arranged in a $\underset{\sim}{P}$ of order $(N + 1)(2N + 1) \times (N + 1)(2N + 1)$ whereas since there $(N + 2)(2N - 1)$ transient states, the $Q-$ matrix will be of the order $(N + 2)(2N - 1) \times (N + 2)(2N - 1)$. If we denote by $\underline{P'}_{o,o},$ $\underline{P'}_{o,2N}$ and $\underline{P'}_{2N,o}$ row vectors of order $(N + 2)(2N - 1)$ as representing the one-step transitions for fixation of A_3, A_2 and A_1 genes respectively, the partitioned form of $\underset{\sim}{P}$ is given by

$$(2.23) \quad \underset{\sim}{P} = \begin{bmatrix} 1 & 0 & 0 & \underline{0}' \\ 0 & 1 & 0 & \underline{0}' \\ 0 & 0 & 1 & \underline{0}' \\ \underline{P}_{o,o} & \underline{P}_{o,2N} & \underline{P}_{2N,o} & \underset{\sim}{Q} \end{bmatrix}$$

where $\underline{0}'$ is a null row vector of order $(N + 2)(2N - 1)$. The $Q-$matrix can further be partitioned as

$$(2.24) \quad \underset{\sim}{Q} = \begin{bmatrix} \underset{\sim}{Q}_{12} & \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{Q}_{13} & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{Q}_{23} & \underset{\sim}{0} \\ \underset{\sim}{D}_{12} & \underset{\sim}{D}_{13} & \underset{\sim}{D}_{23} & \underset{\sim}{Q}^* \end{bmatrix}$$

where $\underset{\sim}{Q}_{12}$ represents the transition probabilities between states

E_{i_1,i_2} of class II i.e. a two-allele situation $A_1 - A_2$.
Similarly $\underset{\sim}{Q}_{13}$ is for $A_1 - A_3$ and $\underset{\sim}{Q}_{23}$ is for $A_2 - A_3$ situations
respectively. $\underset{\sim}{Q}^*$ represents the $(N-1)(2N-1) \times (N-1)(2N-1)$
matrix for transitions between states E_{i_1,i_2} of class III. $\underset{\sim}{D}_{12}$
represents the transition from states of class III to E_{i_1,i_2} type
states of class II, $\underset{\sim}{D}_{12}$ represents the transitions from states of
class III to $E_{i_1,0}$ type states of class II and $\underset{\sim}{D}_{23}$ represents the
transitions from states of class III to E_{0,i_2} type states of class
II.

In view of the well known matrix operations on partitioned
matrices, the fundamental matrix $\underset{\sim}{T}$ introduced in the previous section
and equal to $(\underset{\sim}{I} - \underset{\sim}{Q})^{-1}$ is now given by

$$(2.25) \quad \underset{\sim}{T} = \begin{bmatrix} \underset{\sim}{T}_{12} & \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{T}_{13} & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{T}_{23} & \underset{\sim}{0} \\ \underset{\sim}{T}^* \underset{\sim}{D}_{12} \underset{\sim}{T}_{12} & \underset{\sim}{T}^* \underset{\sim}{D}_{13} \underset{\sim}{T}_{13} & \underset{\sim}{T}^* \underset{\sim}{D}_{23} \underset{\sim}{T}_{23} & \underset{\sim}{T}^* \end{bmatrix}$$

where

$$(2.26) \quad \underset{\sim}{T}_{12} = (\underset{\sim}{I} - \underset{\sim}{Q}_{12})^{-1}$$

$$(2.27) \quad \underset{\sim}{T}_{13} = (\underset{\sim}{I} - \underset{\sim}{Q}_{13})^{-1}$$

$$(2.28) \quad \underset{\sim}{T}_{23} = (\underset{\sim}{I} - \underset{\sim}{Q}_{23})^{-1}$$

$$(2.29) \quad \underset{\sim}{T}^* = (\underset{\sim}{I} - \underset{\sim}{Q}^*)^{-1}$$

The vectors of the probability of fixation and the expected change in the gene frequency of A_1 in the limit are, therefore, given by

$$(2.30) \quad \underline{U}_1 = \underset{\sim}{T} \, \underline{P}_{2N,o}$$

$$(2.31) \quad \underline{R}_1 = \underset{\sim}{T} \, \underline{\Delta q}_1$$

Now $\underline{P}_{2N,o}$ is a vector of transition probabilities for transitions from the transient states to the absorbing state representing the vertex $E_{2N,o}$ in the triangle in Fig. 1 in one step. It can be represented by a partitioned row vector as

$$(2.32) \quad \underline{P'}_{2N,o} = [\, {}_2\underline{P'}_{2N}, \quad {}_3\underline{P'}_{2N}, \quad \underline{0'}, \quad \underline{P^*}'_{2N} \,]$$

where ${}_2\underline{P'}_{2N}$ represents the row vector of one-step transitions for the fixation of A_1 starting from the cases in which A_3 is absent. ${}_3\underline{P'}_{2N}$ is similarly the vector for the case when initially A_2 is absent, and $\underline{P^*}'_{2N}$ is for the case when initially all the three genes are present. Obviously if initially A_1 is absent, there is no question of fixation of A_1. Hence there is a null row vector also in $\underline{P'}_{2N,o}$. Now with this partitioning the fixation probability vector $\underline{U}_1$ consists of four component vectors $\underline{U}_{12}$, $\underline{U}_{13}$, $\underline{U}_{23}$ and $\underline{U}^*$ given by

$$(2.33) \quad \underline{U}_{12} = \underset{\sim}{T}_{12} \, {}_2\underline{P}_{2N}$$

$$(2.34) \quad \underline{U}_{13} = \underset{\sim}{T}_{13} \, {}_3\underline{P}_{2N}$$

$(2.35) \quad \underline{U}_{23} = \underline{0}$

$(2.36) \quad \underline{U}^* = \underset{\sim}{T}^* (\underset{\sim}{D}_{12} \underline{U}_{12} + \underset{\sim}{D}_{13} \underline{U}_{13} + \underline{P}^*_{2N})$

Here $\underline{U}_{12}$ and $\underline{U}_{13}$ are the usual fixation probability vectors for two allele $A_1 - A_2$ and $A_1 - A_3$ situations as discussed in the previous section. $\underline{U}^*$ is the fixation probability vector for A_1 when initially all the three genes exist or in other words the initial state is somewhere in the interior of the triangle in Fig. 1. For fixation in the vertex $E_{2N,o}$ from such a state, there are three possibilities. One possibility is that it can go to one side of the triangle representing E_{i_1,i_2} with $i_1 + i_2 = 2N$ and then to $E_{2N,o}$. The second possibility is that it can go to the side of the triangle representing $E_{i_1,o}$ and then to $E_{2N,o}$. Third possibility is that it can straight go to $E_{2N,o}$. The matrix formula (2.36) derived above shows that when the initial state is somewhere in the interior of the triangle, the fixation probability is the sum of the expected total number of times the population spends in the interior states multiplied by the sum of the corresponding one-step probabilities of fixation via the three different paths enumerated above.

Similar considerations show that if we break down the vector of the expected changes in the frequency of A_1 in one step into components $_2\underline{\Delta q}_1$, $_3\underline{\Delta q}_1$, $\underline{0}$, $\underline{\Delta}^* q_1$ as

$(2.37) \quad \underline{\Delta}' q_1 = [\, _2\underline{\Delta}' q_1, \quad _3\underline{\Delta}' q_1, \quad \underline{0}', \quad \underline{\Delta}^{*'} q_1 \,]$

and the selection limit vector into components $\underline{R}_{12}$, $\underline{R}_{13}$, $\underline{R}_{23}$ and $\underline{R}^*$, we get

$$(2.38) \qquad \underline{R}_{12} = \underset{\sim}{T}_{12} \; _2\underline{\Delta q}_1$$

$$(2.39) \qquad \underline{R}_{13} = \underset{\sim}{T}_{13} \; _3\underline{\Delta q}_1$$

$$(2.40) \qquad \underline{R}_{23} = \underline{0}$$

$$(2.41) \qquad \underline{R}^* = \underset{\sim}{T}^* \, (\underset{\sim}{D}_{12} \, \underline{R}_{12} + \underset{\sim}{D}_{13} \, \underline{R}_{13} + \underline{\Delta}^* q_1)$$

This gives a matrix formula for the evaluation of the expected limit of response to selection in a single locus three allele situation.

We have derived above the matrix formulae for the fixation probability and the expected change in the gene frequency for the allele A_1. Similar results hold for the alleles A_2 and A_3.

For the mean time until homozygosity, we can write, in vector notations

$$(2.42) \qquad \underline{m} = \underset{\sim}{T} \, \underline{e}$$

If $\underline{m}$ consists of four components $\underline{m}_1$, $\underline{m}_2$, $\underline{m}_3$ and $\underline{m}^*$ corresponding to the four initial situations described earlier, we get

$$(2.43) \qquad \underline{m}_1 = \underset{\sim}{T}_1 \, \underline{e}$$

$$(2.44) \qquad \underline{m}_2 = \underset{\sim}{T}_2 \, \underline{e}$$

$$(2.45) \qquad \underline{m}_3 = \underset{\sim}{T}_3 \, \underline{e}$$

$$(2.46) \qquad \underline{m}^* = \underset{\sim}{T}^* \, [\underset{\sim}{D}_{12} \, \underline{m}_1 + \underset{\sim}{D}_{13} \, \underline{m}_2 + \underset{\sim}{D}_{23} \, \underline{m}_3 + \underline{e}]$$

For the second moment of time until homozygosity, we have the relation

$$(2.47) \quad \underline{v} = (2 \underset{\sim}{T} - \underset{\sim}{I}) \underline{m}$$

If $\underline{v}$ is broken down into four components $\underline{v}_1$, $\underline{v}_2$, $\underline{v}_3$ and $\underline{v}^*$ we get

$$(2.48) \quad \underline{v}_1 = (2 \underset{\sim}{T}_1 - \underset{\sim}{I}) \underline{m}_1$$

$$(2.49) \quad \underline{v}_2 = (2 \underset{\sim}{T}_2 - \underset{\sim}{I}) \underline{m}_2$$

$$(2.50) \quad \underline{v}_3 = (2 \underset{\sim}{T}_3 - \underset{\sim}{I}) \underline{m}_3$$

$$(2.51) \quad \underline{v}^* = \underset{\sim}{T}^* [\underset{\sim}{D}_{12} \underline{v}_1 + \underset{\sim}{D}_{13} \underline{v}_2 + \underset{\sim}{D}_{23} \underline{v}_3 + 2 \underline{m}^* - \underline{e}]$$

In order to evaluate the mean time until fixation of the A_1 allele disregarding the cases in which it is lost, we have to evaluate

$$(2.52) \quad \underline{M}_1 = \underset{\sim}{T} \underline{U}_1$$

As usual if we let

$$(2.53) \quad \underline{M}'_1 = [M'_{12}, \ M'_{13}, \ M'_{23}, \ \underline{M}^{*'}]$$

we get

$$(2.54) \quad \underline{M}_{12} = \underset{\sim}{T}_{12} \ \underline{U}_{12}$$

$$(2.55) \quad \underline{M}_{13} = \underset{\sim}{T}_{13} \ \underline{U}_{13}$$

$$(2.56) \quad \underline{M}_{23} = \underline{0}$$

(2.57) $\underset{\sim}{M}* = \underset{\sim}{T}* (\underset{\sim 12}{D} \underline{M}_{12} + \underset{\sim 13}{D} \underline{M}_{13} + \underline{U}*)$

The required mean time will be given by the ratios of the elements of
the vectors $\underline{M}_{12}$ and $\underline{U}_{12}$, $\underline{M}_{13}$ and $\underline{U}_{13}$ and $\underline{M}*$ and $\underline{U}*$ respectively.

For the second moment of the distribution of time until fixation
of the A_1 allele disregarding the cases in which it is lost, we have the
relation

(2.58) $\underline{V}_1 = (2 \underset{\sim}{T} - \underset{\sim}{I}) \underline{M}_1$

If $\underline{V}_1$ consists of $\underline{V}_{12}$, $\underline{V}_{13}$, $\underline{V}_{23}$ and $\underline{V}*$ we get

(2.59) $\underline{V}_{12} = (2 \underset{\sim 12}{T} - \underset{\sim}{I}) \underline{M}_{12}$

(2.60) $\underline{V}_{13} = (2 \underset{\sim 13}{T} - \underset{\sim}{I}) \underline{M}_{13}$

(2.61) $\underline{V}_{23} = \underline{0}$

(2.62) $\underline{V}* = \underset{\sim}{T}* (\underset{\sim 12}{D} \underline{V}_{12} + \underset{\sim 13}{D} \underline{V}_{13} + 2\underline{M}* - \underline{U}*)$

The required second moment will be given by the ratios of the elements
of the vectors $\underline{V}_{12}$ and $\underline{U}_{12}$, $\underline{V}_{13}$ and $\underline{U}_{13}$ and $\underline{V}*$ and $\underline{U}*$
respectively.

Similarly the moments of the distribution of time until fixation of
either A_2 or A_3 disregarding the others can be derived.

2.2   MUTATION, SELECTION AND RANDOM DRIFT

Suppose we consider S loci at each of which there is
initially one mutant A_1 and $(2N - 1)$ normal A_2 alleles. This
means that the population is initially in E_1 state. Each locus will
eventually become homozygous in the absence of further mutation.
Suppose S is so chosen that on an average one locus becomes homo-
zygous per generation. We can, then introduce a new locus having one
mutant and $(2N - 1)$ normals each generation to balance this loss.
This would lead to an equilibrium state with mean number S_{1j} of loci
having j mutants, where j can take values 1, 2, ... $(2N - 1)$
and $S = \sum_{j=1}^{(2N-1)} S_{1j}$. It follows from the ergodic property of irreducible
recurrent Markov chains that the S_{1j}'s can be obtained by pre-
multiplying the fundamental matrix $\underset{\sim}{T}$ by a row vector $\underline{f'}_1 = (1,0,0 \ldots 0)$
containing unity in the first place and zeros elsewhere and therefore

$$(2.63) \quad S = \underline{f'}_1 \, (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \, \underline{e}$$

This shows that the mean number of loci having j mutants in
equilibrium is equivalent to the mean number of generations which the
population spends in the state E_j on the way to fixation or loss if
there had been no balancing effect by mutation. Now if the mutation rate
from A_2 to A_1 is u (reverse mutation being negligible), there would
be on an average 2N u new loci each having one mutant and the rest
normal each generation. Hence there would be required 2N u S loci

initially instead of S to determine the equilibrium state and the mean number of loci having j mutants in equilibrium would be $2N\,u\,S_{1j}$. The mean and variance of q, the frequency of A_1 in the equilibrium state are given by

$$(2.64) \quad E(q) = 2N\,u \sum_{j=1}^{2N-1} S_{1j}\,(j/2N)$$

$$(2.65) \quad E[q - E(q)]^2 = 2N\,u \sum_{j=1}^{2N-1} S_{1j}\,(j/2N)^2 - [E(q)]^2$$

The mean frequency of heterozygotes in the equilibrium state can be obtained from

$$(2.66) \quad E[2q(1-q)] = 4N\,u \sum_{j=1}^{2N-1} S_{1j}\,(j/2N)\,(1 - j/2N)$$

This quantity is proportional to the contribution of the locus towards the genetic variance of the stable population and shows that it is proportional to the population size if $N\,u$ is small.

CHAPTER III

EXPANSION OF MATRIX FORMULAE AND
DIFFUSION APPROXIMATIONS

The theory developed in Chapter II can be applied to specific genetic models. In this chapter an analytical treatment of selection with random drift and diallelic locus is presented. In the first part, matrix formulae developed in 2.11 are expanded taking into account the binomial transition probabilities. The algebraic expressions for selection limit, half-life and moments of the distribution of time to fixation are discussed. In the second part, the diffusion approximation for the variance of time until fixation is derived.

3.1 EXPANSION WITH BINOMIAL TRANSITION PROBABILITY

If no selective forces are operating and the transition probabilities are of the binomial type, the situation is what is often known as Wright's model (Wright 1931). Here starting with a given frequency $q_i = i/(2N)$ of allele A_1 (with $1 - q_i$ = frequency of A_2) in lines of constant breeding size of N individuals, we can consider the second generation as derived from the first by the sampling of groups of $2N$ haploid sets, the gene frequency in the different groups being distributed binomially with mean $2Nq_i$ and index $2N$. The next generation is then the repetition of this process, each line giving rise to a group of lines

whose gene frequencies are binomially distributed about the mean of the parent line. Now let A_1 genes have a selective advantage of $(1 + s/2)$ over A_2 genes with selective advantage unity, so that the relative number of offsprings have expectations proportional to $(1 + s/2)$ and 1 respectively where s is small. That is, we assume that a large number of offsprings are produced but exactly 2N of these survive. Assuming further that the selection operates before sampling, the gene frequency in the different groups are distributed binomially with mean $2Nq_i'$ and index 2N, where q_i' is the gene frequency of A_1 after selection and is given by

$$(3.1) \quad q_i' = q_i + \delta q_i$$

$$= q_i + \frac{s}{2} q_i(1 - q_i)/(1 + q_i s/2)$$

δq_i being the change in the mean frequency of A_1 due to selection in one step when its frequency in the previous generation is q_i. Thus with haploid selection and binomial sampling, we can regard the number of A_1 genes in any generation as a Markovian variate with transition probability P_{ij} given by

$$(3.2) \quad P_{ij} = \binom{2N}{j}(q_i')^j (1 - q_i')^{\overline{2N - j}}$$

These probabilities, therefore, determine the P- and Q-matrices introduced in 2.11. We can now expand the transition probabilities in

powers of 's' for a given N and utilise the properties of the transition probability matrices with no selection, referred to here-after, as P_o and Q_o. The properties of these matrices, in terms of the eigen-roots and vectors, are discussed in Feller (1951) and Robertson (1952).

Let δq_i be approximated by $\frac{s}{2} q_i (1 - q_i) (1 - \frac{s}{2} q_i)$. Then P_{ij} can be expressed approximately upto terms in s^2 as

$$(3.3) \quad P_{ij} = P_{ij}(o) [1 + s\, a_{ij} + s^2 b_{ij}]$$

where

$$(3.4) \quad P_{ij}(o) = \binom{2N}{j} q_i^j (1 - q_i)^{2N - j}$$

is the transition probability with no selection and a_{ij} and b_{ij} are given by

$$(3.5) \quad a_{ij} = N(q_j - q_i)$$

$$(3.6) \quad b_{ij} = \frac{N^2}{2} [q_j^2 + (1 + \frac{1}{2N})q_i^2 - 2q_i q_j - \frac{1}{2N} q_j] ,$$

q_j being $j/(2N)$.

In terms of Q, we can write, upto terms involving s^2,

$$(3.7) \quad Q = Q_o + s\, Q_o' + s^2 Q_o''$$

where Q_o' and Q_o'' are $(2N - 1) \times (2N - 1)$ matrices with i, j

elements respectively as

(3.8) $\quad a_{ij}\, P_{ij}(o)$

and

(3.9) $\quad b_{ij}\, P_{ij}(o)$

Now we have proved in 2.11 that the vectors $\underline{R}$ and $\underline{R}(t)$ are obtained by operating certain functions of $\underset{\sim}{Q}_o$ on to the column vector of the change in the mean gene frequency in a single generation. This column vector is expressible in terms of the right hand eigen-vectors of $\underset{\sim}{Q}_o$. Hence the expanded form of $\underline{R}$ and $\underline{R}(t)$ can be determined if we know the operations of certain functions of $\underset{\sim}{Q}_o$ on to these vectors. It has been found that we need nine operators and the first three vectors of $\underset{\sim}{Q}_o$. The vectors are

(3.10) $\quad \underline{X}_1 = [X_{11},\ X_{12},\ \ldots\ldots\ X_{1,(2N-1)}],\quad X_{1i} = q_i(1-q_i)$

(3.11) $\quad \underline{X}_2 = [X_{21},\ X_{22},\ \ldots\ldots\ X_{2,(2N-1)}],\quad X_{2i} = q_i(1-q_i)(1-2q_i)$

(3.12) $\quad \underline{X}_3 = [X_{31},\ X_{32},\ \ldots\ldots\ X_{3,(2N-1)}],$

$$X_{3i} = q_i(1-q_i)\left[\frac{2N-1}{10N-6} - q_i(1-q_i)\right]$$

These vectors correspond to the three eigen - roots of $\underset{\sim}{Q}_o$ given by

(3.13) $\quad \lambda_1 = (1 - 1/2N)$

(3.14) $\quad \lambda_2 = (1 - 1/2N)(1 - 2/2N)$

(3.15) $\quad \lambda_3 = (1 - 1/2N)(1 - 2/2N)(1 - 3/2N)$

respectively. The results of performing these operations are given in Table 1, where only the 18 operations out of the 27 possible are given, which are actually needed in the expansion of $\underline{R}$ and $\underline{R}(t)$.

3.11 EXPANSION FOR THE EXPECTED SELECTION LIMIT

In Chapter II, the selection limit vector has been expressed as the operation of $(\underset{\sim}{I} - \underset{\sim}{Q})^{-1}$ on to $\underline{\triangle} q$, the vector of expected change in the gene frequency in one step. Each of these, when expressed in powers of s, upto terms involving s^2, are given by

$$(3.16) \quad (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} = (\underset{\sim}{I} - \underset{\sim}{Q_o})^{-1} + s(\underset{\sim}{I} - \underset{\sim}{Q_o})^{-1} \underset{\sim}{Q_o}'(\underset{\sim}{I} - \underset{\sim}{Q_o})^{-1}$$
$$+ s^2 [(\underset{\sim}{I} - \underset{\sim}{Q_o})^{-1} \underset{\sim}{Q_o}''(\underset{\sim}{I} - \underset{\sim}{Q_o})^{-1}$$
$$+ (\underset{\sim}{I} - \underset{\sim}{Q_o})^{-1} \underset{\sim}{Q_o}'(\underset{\sim}{I} - \underset{\sim}{Q_o})^{-1} \underset{\sim}{Q_o}'(\underset{\sim}{I} - \underset{\sim}{Q_o})^{-1}]$$

$$(3.17) \quad \underline{\triangle} q = \frac{s}{2} (1 - \frac{s}{4}) \underline{X_1} + \frac{s^2}{8} \underline{X_2}$$

The item by item operation of the terms in the expansion of the matrix on to those in the expansion of the vector has been performed with the help of Table 1. The result, expressed as a linear function of the three vectors $\underline{X_1}$, $\underline{X_2}$ and $\underline{X_3}$ is given by

$$(3.18) \quad \underline{R} = a_{10} \underline{X_1} + a_{20} \underline{X_2} + a_{30} \underline{X_3}$$

TABLE 1   MATRIX OPERATORS AND VECTORS

	X_1	X_2	X_3
Q_0	$\lambda_1 X_1$	$\lambda_2 X_2$	$\lambda_3 X_3$
T_0	$2N X_1$	$\dfrac{2N^2}{3N-1} X_2$	$\dfrac{4N^3}{12N^2 - 11N + 3} X_3$
Q_0^t	$\lambda_1^t X_1$	$\lambda_2^t X_2$	$\lambda_3^t X_3$
Q_0'	$\tfrac{1}{2}\lambda_1 X_2$	$\tfrac{1}{2}\lambda_2\left(6 X_3 - \dfrac{N}{5N-3} X_1\right)$	---
$T_0 Q_0' T_0$	$\dfrac{2N^3}{3N-1}\lambda_1 X_2$	$\dfrac{2N^3}{3N-1}\lambda_2\left(\dfrac{12N^2}{12N^2-11N+3} X_3 - \dfrac{N}{5N-3} X_1\right)$	---
$(Q_0^t)'$	$\tfrac{1}{2}\lambda_1 W_1 X_2$	$\tfrac{1}{2}\lambda_2\left(6 W_3 X_3 - \dfrac{N}{5N-3} W_1 X_1\right)$	---

TABLE 1   CONTINUED

	$\underline{X}_1$	$\underline{X}_2$	$\underline{X}_3$
$\underset{\sim}{Q}_0''$	$\frac{1}{8} \lambda_1 (6 \underline{X}_3 - \frac{N}{5N-3} \underline{X}_1 - \underline{X}_2)$	--	--
$\underset{\sim}{T}_0 \underset{\sim}{Q}_0'' \underset{\sim}{T}_0$	$\frac{1}{2} N^2 \lambda_1 (\frac{12N^2}{12N^2 - 11N + 3} \underline{X}_3 - \frac{N}{5N-3} \underline{X}_1 - \frac{N}{3N-1} \underline{X}_2)$	--	--
$(\underset{\sim}{Q}_0^t)''$	$\frac{1}{8} \lambda_1 [6(\frac{\lambda_2}{\lambda_2 - \lambda_3} W_1 - \frac{\lambda_2}{\lambda_1 - \lambda_2} W_3 + \frac{\lambda_2^2 - \lambda_1 \lambda_3}{(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)} W_2) \underline{X}_3$	--	--
	$- W_1 \underline{X}_2 - \frac{N}{5N-3} (t \lambda_1^{t-1} \frac{(\lambda_1 + \lambda_2)}{(\lambda_1 - \lambda_2)} - \frac{2\lambda_2}{(\lambda_1 - \lambda_2)} W_1) \underline{X}_1]$		

N.B.

$$T_0 = (\underset{\sim}{I} - \underset{\sim}{Q}_0)^{-1} \qquad W_2 = (\frac{\lambda_1^t - \lambda_3^t}{\lambda_1 - \lambda_3})$$

$$W_1 = (\frac{\lambda_1^t - \lambda_2^t}{\lambda_1 - \lambda_2}) \qquad W_3 = (\frac{\lambda_2^t - \lambda_3^t}{\lambda_2 - \lambda_3})$$

where

$$(3.19) \quad a_{10} = Ns \left[1 - (s/4) - \left(\frac{N-1}{5N-3}\right) \left(\frac{8N^3 - 7N + 2}{3N - 1}\right) (s^2/8) \right]$$

$$(3.20) \quad a_{20} = N^2 s^2 \left[\frac{4N - 1}{4(3N - 1)} (1 - s/2) \right]$$

$$(3.21) \quad a_{30} = N^3 s^3 \left[\frac{3(4N - 1)(N - 1)(2N - 1)}{2(3N - 1)(12N^2 - 11N + 3)} + \frac{3N - 1}{12N^2 - 11N + 3} \right]$$

Starting with an initial gene frequency q, the expected selection limit or the chance of its eventual fixation $u(q)$ minus the initial frequency is then given by

$$(3.22) \quad u(q) - q = a_{10} q(1 - q) + a_{20} q(1 - q)(1 - 2q)$$

$$+ a_{30} q(1 - q) \left[\frac{2N - 1}{10N - 6} - q(1 - q) \right]$$

If N becomes very large and s becomes very small such that Ns is kept constant, we get

$$(3.23) \quad u(q) - q = Ns(1 - N^2 s^2/15) q(1 - q) + \frac{1}{3} N^2 s^2 q(1 - q)(1 - 2q)$$

$$+ \frac{1}{3} N^3 s^3 q(1 - q) \left[\frac{1}{5} - q(1 - q) \right]$$

$$= Ns\, q(1 - q) + \frac{1}{3} N^2 s^2 q(1 - q)(1 - 2q) - \frac{1}{3} N^3 s^3 q^2 (1 - q)^2$$

This expression is exactly the same, upto terms involving $N^3 s^3$, as the expanded form of the formula of chance of fixation obtained by diffusion approximation and given in Kimura (1964).

The ratio of the expected response in the limit to the initial response of $\frac{s}{2} q(1 - q)$ is given by

$$(3.24) \quad \frac{u(q) - q}{\frac{s}{2} q(1 - q)} = 2N \left[1 + \frac{1}{3} Ns(1 - 2q) - \frac{1}{3} N^2 s^2 q(1 - q) \right]$$

This shows that for $q = \frac{1}{2}$, this ratio is $2N(1 - \frac{1}{12} N^2 s^2)$ but for q tending to 0 and 1, this tends respectively to $2N(1 + \frac{1}{3} Ns)$ and $2N(1 - \frac{1}{3} Ns)$.

If we wish to project the response, expressed above in terms of the gene frequency, on to the underlying quantitative trait, in the case of artificial selection, we have to assume the absence of natural selection and use the following relation between the selective advantage s and the intensity of selection $\bar{i}$, imposed artificially, as given in Robertson (1960),

$$(3.25) \quad s = \bar{i} \, a/\sigma$$

where 'a' refers to the difference of units on the metric scale between the mean of the two homozygotes for the additive locus under consideration and σ is the phenotypic standard deviation. It may be noted that the haploid model is a fairly good approximation for the diploid situation with additive gene action, the diploid genetic variance merely being double that of the haploid genetic variance.

Summing the contributions of the individual loci and using the relation between s and $\bar{i}$, the expected change in the limit in the character (L) under consideration is given by

$$(3.26) \quad L = \Sigma a \left[u(q) - q\right]$$

$$= \frac{N\bar{i}}{\sigma} \Sigma a^2 q(1 - q) + \frac{1}{3} \left(\frac{N\bar{i}}{\sigma}\right)^2 \Sigma a^3 q(1 - q)(1 - 2q)$$

$$- \frac{1}{3} \left(\frac{N\bar{i}}{\sigma}\right)^3 \Sigma a^4 q^2(1 - q)^2$$

If σ_g and μ_{3g} denote the standard deviation and the third moment (about the mean) of the breeding values (Falconer 1960), we have

$$(3.27) \quad \sigma_g^2 = \tfrac{1}{2} \Sigma a^2 q(1 - q)$$

$$\mu_{3g} = \tfrac{1}{4} \Sigma a^3 q(1 - q)(1 - 2q)$$

$$(3.28) \quad nV^* + \frac{1}{n} \sigma_g^4 = \tfrac{1}{4} \Sigma a^4 q^2(1 - q)^2$$

where V^* refers to the variance between the contributions of the n different loci to the additive genetic variance of the character. Hence L is expressed as

$$(3.29) \quad L = 2N \left[\bar{I} h \sigma_g + \frac{2}{3} N \frac{\mu_{3g}}{\sigma_g^4} (\bar{I} h \sigma_g)^2 \right.$$

$$\left. - \frac{2}{3} \frac{N^2}{n\sigma_g^2} \left(1 + \frac{n^2 V^*}{\sigma_g^4}\right) (\bar{I} h \sigma_g)^3 \right]$$

where $h^2 = \sigma_g^2/\sigma^2$, the heritability of the trait. Since the initial change in the character is given by

$$(3.30) \quad I = \Sigma a \delta q = \bar{I} h \sigma_g$$

the limit L can be finally expressed as

$$(3.31) \quad L = 2NI(1 + C_1 I + C_2 I^2)$$

where

$$(3.32) \quad C_1 = \frac{2}{3} N \frac{\mu_{3g}}{\sigma_g^4}$$

$$(3.33) \quad C_2 = -\frac{2}{3} \frac{N^2}{n\sigma_g^2} \left(1 + \frac{n^2 V^*}{\sigma_g^4}\right)$$

3.12 EXPANSION FOR THE EXPECTED CHANGE IN THE GENE FREQUENCY BY THE tth GENERATION

The vector of the expected change in the gene frequency of A_1 by the tth generation has been expressed in 2.11 as the operation of the matrix $(I - Q^t)$ on to R, the vector of the expected selection limit. The former, when expressed in powers of s upto terms involving s^2, is given by

$$(3.34) \quad (\underset{\sim}{I} - \underset{\sim}{Q}^t) = (\underset{\sim}{I} - \underset{\sim}{Q}_o^t) - s(\underset{\sim}{Q}^t)' - s^2(\underset{\sim}{Q}^t)''$$

whereas the latter has already been expressed as the linear function of $\underline{X}_1$, $\underline{X}_2$ and $\underline{X}_3$ by (3.18). The item by item operation of terms in $(\underset{\sim}{I} - \underset{\sim}{Q}^t)$ on to those in R has been performed with the help of Table 1. The result expressed as a linear function of $\underline{X}_1$, $\underline{X}_2$ and $\underline{X}_3$ is given by

$$(3.35) \quad \underline{R}(t) = (a_{10} + a_{11} \lambda_1^t + a_{12} \lambda_2^t) \underline{X}_1$$
$$+ (a_{20} + a_{21} \lambda_1^t + a_{22} \lambda_2^t) \underline{X}_2$$
$$+ (a_{30} + a_{31} \lambda_1^t + a_{32} \lambda_2^t + a_{33} \lambda_3^t) \underline{X}_3$$

where

$$(3.36) \quad a_{11} = -Ns \left[1 - \left(\frac{s}{4}\right) - \frac{(N-1)(2N^2 + N - 2) + N(2N-1)t}{5N - 3} \left(\frac{s^2}{8}\right) \right]$$

$$(3.37) \quad a_{12} = (Ns)^3 \left[\frac{(N-1)(2N-1)}{8(5N-3)(3N-1)} \right]$$

$$(3.38) \quad a_{21} = -(Ns)^2 \left[\frac{1}{2}(1 - s/2) \right]$$

$$(3.39) \quad a_{22} = (Ns)^2 \left[\frac{(2N-1)}{4(3N-1)}(1 - s/2) \right]$$

$$(3.40) \quad a_{31} = -(Ns)^3 \left[\frac{12N^2 - 9N + 1}{4N(5N-3)} \right]$$

$$(3.41) \quad a_{32} = (Ns)^3 \left[\frac{(2N-1)}{2(3N-1)} \right]$$

$$(3.42) \quad a_{33} = -(Ns)^3 \left[\frac{48N^5 - 4N^4 - 73N^3 + 74N^2 - 23N + 3}{4N(3N-1)(5N-3)(12N^2 - 11N + 3)} \right]$$

For initial gene frequency q, the expected gene frequency by the tth generation, written as $E(q_t)$, is then given by

$$(3.43) \quad E(q_t) = q + (a_{10} + a_{11} \lambda_1^t + a_{12} \lambda_2^t) q(1-q)$$

$$+ (a_{20} + a_{21} \lambda_1^t + a_{22} \lambda_2^t) q(1-q)(1 - 2q)$$

$$+ (a_{30} + a_{31} \lambda_1^t + a_{32} \lambda_2^t + a_{33} \lambda_3^t)$$

$$\cdot q(1-q) \left[\frac{2N-1}{10N - 6} - q(1-q) \right]$$

When N becomes very large and s very small such that Ns is held constant, we get

$$(3.44) \quad E(q_t) = q + Ns[1 - (1 - \frac{3}{100} N^2s^2) \exp(-t/2N) - \frac{N^2s^2}{10} \exp(-2t/2N)$$

$$+ \frac{N^2s^2}{12} \exp(-3t/2N)$$

$$- \frac{N^2s^2}{75} \exp(-6t/2N)] \, q(1 - q)$$

$$+ \frac{1}{3} N^2s^2 [1 - \frac{3}{2} \exp(-t/2N) + \frac{1}{2} \exp(-3t/2N)]$$

$$. \, q(1 - q)(1 - 2q)$$

$$- \frac{1}{3} N^3s^3 [1 - \frac{9}{5} \exp(-t/2N) + \exp(-3t/2N) - \frac{1}{5} \exp(-6t/2N)]$$

$$. \, q^2(1 - q)^2$$

3.13 HALF - LIFE

The half-life is the time t_h (in terms of the number of generations) by which the expected change in the gene frequency of A_1 is half that expected in the limit. Hence we have to solve for t in

the following equation

$$(3.45) \quad E(q_t) - q = (1/2) [u(q) - q]$$

after substituting $u(q)$ and $E(q_t)$ from (3.23) and (3.44) respectively. This gives an equation of the sixth order in $x = \exp(-t/2N)$

$$(3.46) \quad Ax^6 + Bx^3 + Cx^2 + Dx + E = 0$$

where

$$(3.47) \quad A = -\frac{1}{15} (Ns)^2 [\frac{1}{5} - q(1 - q)]$$

$$(3.48) \quad B = \frac{1}{6} Ns (1 - 2q) + \frac{1}{3} (Ns)^2 [\frac{1}{4} - q(1 - q)]$$

$$(3.49) \quad C = -\frac{1}{10} (Ns)^2$$

$$(3.50) \quad D = -1 - \frac{1}{2} Ns(1 - 2q) + \frac{3}{5} (Ns)^2 [\frac{1}{20} + q(1 - q)]$$

$$(3.51) \quad E = \frac{1}{2} + \frac{1}{6} Ns (1 - 2q) - \frac{1}{6} (Ns)^2 q(1 - q)$$

Since $x_o = \frac{1}{2}$ is an approximate solution (Robertson 1960), an improved value of x can be obtained by applying Newton - Raphson method to the equation.

This is given by

$$(3.52) \quad x = x_o - \frac{Ax_o^6 + Bx_o^3 + Cx_o^2 + Dx_o + E}{6Ax_o^5 + 3Bx_o^2 + 2Cx_o + D}$$

$$= \frac{1}{2} \left[\frac{1 + \frac{1}{4} Ns(1 - 2q) + \frac{5}{96}(Ns)^2 (\frac{1}{5} - q(1-q)) - \frac{1}{8}N^2s^2 q(1-q)}{1 + \frac{3}{8} Ns(1 - 2q) + \frac{1}{20} N^2s^2 (\frac{1}{5} - q(1-q)) - \frac{5}{16}N^2s^2 q(1-q)} \right]$$

Since $t_h = -2N \log_e x$, it is approximately given by

$$(3.53) \quad t_h = [1.4 + \frac{1}{4} Ns(1 - 2q) + 1.12 N^2s^2 (0.16 - q(1-q))]N$$

For $Ns = 1$ and $q = \frac{1}{2}$, t_h is found to be 1.30N. For q tending to 0 or to 1, we have to make use of $t_h = -2N \log_e x$ as such, without further approximating. That is

$$(3.54) \quad t_h = -2N [\log_e \frac{1}{2} + \log_e [1 + \frac{1}{4} Ns(1 - 2q) + \frac{5}{96} N^2s^2 (\frac{1}{5} - q(1-q))$$

$$- \frac{1}{8} N^2s^2 q(1 - q)]$$

$$- \log_e [1 + \frac{3}{8} Ns(1 - 2q) + \frac{1}{20} N^2s^2 (\frac{1}{5} - q(1-q))$$

$$- \frac{5}{16} N^2s^2 q(1 - q)] \quad]$$

For $Ns = 1$ and q tending to 0, the half-life is found to be 1.58N whereas for q tending to 1, it is found to be 1.03N. The half-lives

thus obtained for $Ns = 1$ and $q = \frac{1}{2}$, $q \to 0$ and $q \to 1$ are very close to those obtained by matrix iteration method and given in Hill and Robertson (1966) (see Fig. 11 in this paper). The half-life expressions discussed above pertain to the haploid selection process when stochastic variation in the gene frequency due to finite size of the population is taken into account. It has, however, a counterpart in the deterministic situation as well when the population size is assumed infinite. In this case the gene frequency in the tth generation is given by

$$(3.55) \quad q_t = (1 + \frac{s}{2})^t q \bigg/ \left(1 + [(1 + \frac{s}{2})^t - 1] q \right)$$

The number of generations required to change the gene frequency by $(1 - q)/2$ is

$$(3.56) \quad t = \frac{\text{Log}_e (1 + 1/2q)}{\text{Log}_e (1 + s/2)}$$

For $q = \frac{1}{2}$ and positive and small s, this can be approximated by

$$(3.59) \quad t = 2 \text{ Log}_e 2/s$$

When $s = 0.002$, about 700 generations are required to achieve the required change in gene frequency and there is no finite limit for it as $s \to o$. In contrast to this, the half-life t_h, apparently, approaches the value $1.4N$ as s tends to be small.

3.14 EXPANSION FOR THE MEAN AND VARIANCE OF TIME UNTIL FIXATION

In chapter II we have developed matrix formulae for the mean and variance of the time until fixation of A_1 disregarding the cases in which it is lost. However, in order to see the effect of the expansion of $\underset{\sim}{Q}$ in powers of s on the mean and the variance of the distribution of time to fixation, the probability generating function of the number of steps needed for fixation of A_1 disregarding the cases in which it is lost, as introduced in the Appendix, is considered.

The matrix analogue of the probability generating function $\underline{\Pi}(z)$ given by

$$(3.58) \quad \underline{\Pi}(z) = z \, \underset{\sim}{D}_U^{-1} (\underset{\sim}{I} - z\underset{\sim}{Q})^{-1} (\underset{\sim}{I} - \underset{\sim}{Q}) \, \underline{U}$$

can be expressed in terms of the eigen-roots $\lambda_i(s)$ and spectral matrices

$$(3.59) \quad \underset{\sim}{H}_i(s) = \underline{X}_i(s) \, \underline{Y}_i(s)$$

$i = 1, 2, \ldots \ldots (2N - 1)$ of $\underset{\sim}{Q}$, where $\underline{X}_i(s)$ and $\underline{Y}_i(s)$ are respectively the right and left eigen-vectors corresponding to the root $\lambda_i(s)$ and s in the paranthesis indicates that all are functions of s. The resulting expression is given by

$$(3.60) \quad \underline{\Pi}(z) = z\left[\sum_{i=1}^{2N-1} (1 - z \, \lambda_i(s))^{-1} (1 - \lambda_i(s)) \underset{\sim}{D}_U^{-1} \underset{\sim}{H}_i(s) \right] \underline{U}$$

Expanding (3.58) in powers of s and neglecting terms involving s^2 and higher powers of s, we get

$$(3.61) \quad \underset{\sim}{\Pi}(z) = z\underset{\sim x_o}{D^{-1}} (\underset{\sim}{I} - z\underset{\sim}{Q_o})^{-1} (\underset{\sim}{I} - \underset{\sim}{Q_o}) \underline{X_o} + s\left[\frac{d\underset{\sim}{\Pi}(z)}{ds}\right]_{s=o}$$

where

$$(3.62) \quad \underline{X_o} = [X_{o1}, X_{o2}, \dots X_{o,(2N-1)}], \quad X_{oi} = q_i$$

$$(3.63) \quad \underset{\sim x_o}{D} = \text{diag}\,[q_1, q_2 \dots, q_{(2N-1)}]$$

$$(3.64) \quad \left[\frac{d\underset{\sim}{\Pi}(z)}{ds}\right]_{s=o} = 2Nz[(\underset{\sim}{I} - z\underset{\sim}{Q_o})^{-1} + \underset{\sim x_o}{D^{-1}}(\underset{\sim}{I} - z\underset{\sim}{Q_o})^{-1}$$

$$\cdot \left\{ z(\underset{\sim}{Q_o} \underset{\sim x_o}{D} - \underset{\sim x_o}{D}\, \underset{\sim}{Q_o}) (\underset{\sim}{I} - z\underset{\sim}{Q_o})^{-1} - \underset{\sim x_o}{D} \right\}]$$

$$\cdot (\underset{\sim}{I} - \underset{\sim}{Q_o}) \underline{X_o}$$

Expanding (3.61) in terms of roots and vectors of $\underset{\sim}{Q_o}$, it is found that the coefficient of s vanishes, giving

$$(3.65) \quad \underset{\sim}{\Pi}(z) = z \sum_{i=1}^{2N-1} [(1 - \lambda_i)/(1 - z\lambda_i)]\, \underset{\sim x_o}{D^{-1}} \underset{\sim i}{H}\, \underset{\sim x_o}{D^{-1}}\, \underline{e}$$

This shows that to the order s, the probability generating function is the same as that for no selection case. It is a considerable simplification since it shows that all the moments of the distribution of time to fixation are independent of terms in s.

Differentiating $\Pi (z)$ once with respect to z and setting $z = 1$, we get the vector for the mean time until fixation as

$$(3.66) \quad \underline{\mu} = \sum_{i=1}^{2N-1} k_i (1 - \lambda_i)^{-1} \underset{\sim}{D}_{x_o}^{-1} \underline{X_i}$$

where

$$(3.67) \quad k_i = \sum_{j=1}^{2N-1} Y_{ij} q_j$$

Y_{ij} being the jth element of the ith left-eigen vector $\underline{Y_i}$ of $\underset{\sim}{Q}_o$.

Differentiating $\Pi (z)$ twice with respect to z and setting $z = 1$, we get the expected value of $t(t - 1)$ and adding the expected value of t we get the second moment (about the origin) vector $\underline{\mu_2}$ as

$$(3.68) \quad \underline{\mu_2} = \sum_{i=1}^{2N-1} k_i (1 + \lambda_i) (1 - \lambda_i)^{-2} \underset{\sim}{D}_{x_o}^{-1} \underline{X_i}$$

These relations show that explicit expressions for the moments of the distribution of time to fixation are expressible only in terms of all the roots and eigen-vectors of $\underset{\sim}{Q}_o$, the evaluation of the latter requiring the use of computer. Hence the mean and the variance of time to fixation have been studied by evaluating the matrix formulae on the computer in Chapter IV.

3.2   DIFFUSION APPROXIMATION FOR THE VARIANCE

OF TIME UNTIL FIXATION

The diffusion approximation for the mean time until fixation of
a particular allele has recently been given by Kimura and Ohta (1968).
For a neutral gene and initial gene frequency tending to zero, the mean
time is shown to be $4N_e$ where N_e is the variance effective number
which may differ from the actual population number N if the mating
is not random or if the distribution of the number of offspring does
not follow a Poisson distribution. The variance effective number of a
population is defined as the size of an idealized population that would
have the same amount of random gene frequency drift as the population
under consideration (Kimura and Crow, 1963). The diffusion approxi-
mation for the variance of the time until fixation has, however, not
been discussed by Kimura and Ohta (1968).

Let us consider a mutant allele A_2 with frequency q (so that
the frequency of the normal allele A_1 is $1 - q$) in a diploid
population of N individuals with variance effective number N_e. Let
$u(q, t)$ be the probability that A_2 gets fixed by the tth generation
starting with frequency q at $t = o$. Let

$$(3.69) \quad T_1(q) = \int_0^\infty t \, \frac{\partial u(q,t)}{\partial t} \, dt$$

$$(3.70) \quad S_1(q) = \int_0^\infty t^2 \frac{\partial u(q,t)}{\partial t} \, dt$$

Then

$$(3.71) \quad M_1(q) = T_1(q) / u(q)$$

$$(3.72) \quad V_1(q) = S_1(q) / u(q)$$

represent respectively the average and the second moment about the origin of the length of time until the mutant allele becomes fixed in the population excluding the cases in which it is lost from it. Here $u(q)$ is the probability of ultimate fixation such that

$$(3.73) \quad u(q) = \lim_{t \to \infty} u(q, t)$$

If $\alpha(q)$ and $\beta(q)$ represent the mean and the variance of the rate of change in the frequency of A_2 per generation, then following Kimura (1962), $u(q, t)$ satisfies the Kolmogorov backward equation

$$(3.74) \quad \frac{\partial u(q,t)}{\partial t} = \tfrac{1}{2} \beta(q) \frac{\partial^2 u(q,t)}{\partial q^2} + \alpha(q) \frac{\partial u(q,t)}{\partial q}$$

Following the technique of Kimura and Ohta (1968), the set of

differential equations for $T_1(q)$ and $S_1(q)$ are respectively given by

$$(3.75) \quad \tfrac{1}{2}\beta(q)\,\frac{d^2 T_1(q)}{dq^2} + \alpha(q)\,\frac{d\,T_1(q)}{dq} + u(q) = 0$$

$$(3.76) \quad \tfrac{1}{2}\beta(q)\,\frac{d^2 S_1(q)}{dq^2} + \alpha(q)\,\frac{d\,S_1(q)}{dq} + 2T_1(q) = 0$$

The former differential equation has been derived and solved by Kimura and Ohta (1968) with boundary conditions

$$(3.77) \quad \lim_{q \to 0} T_1(q) = K_1\,u(q), \quad K_1 \text{ being a finite quantity}$$

$$(3.78) \quad T_1(1) = 0$$

It is shown by them that

$$(3.79) \quad K_1 = 4N_e$$

$$(3.80) \quad M_1(q) = -4N_e\,\frac{(1-q)}{q}\,\log_e(1-q)$$

For the second moment (about the origin) of the length of time until fixation of A_2, we proceed to transform (3.76) into a differential equation for $V_1(q)$ by differentiating $S_1(q) = V_1(q)\cdot u(q)$ twice and substituting in (3.76).

This gives

(3.81) $\frac{1}{2} \beta (q) \dfrac{d^2 V_1 (q)}{dq^2} + \left[a (q) + \dfrac{\beta(q) \; G (q)}{u (q)} \right] \dfrac{d V_1 (q)}{dq} + 2M_1 (q) = 0$

where

(3.82) $G (q) = \dfrac{d \, u (q)}{dq}$

The boundary conditions to be imposed are

(3.83) $\underset{q \to 0}{\text{Lim.}} \; V_1 (q) = K_2$, K_2 being a finite quantity

(3.84) $V_1 (1) = 0$

In the case of random drift alone, we have

(3.85) $a (q) = 0$

(3.86) $\beta (q) = q(1 - q) \, / \, 2N_e$

(3.87) $u (q) = q$

(3.88) $G (q) = 1$

The differential equation, then reduces to

$$(3.89) \quad \frac{d^2 V_1(q)}{dq^2} + \frac{2}{q} \frac{d V_1(q)}{dq} + \frac{8N_e M_1(q)}{q(1-q)} = 0$$

The solution of this differential equation after substituting for $M_1(q)$ is given by

$$(3.90) \quad V_1(q) = B - \frac{A}{q} - 32N_e^2 \left[\left(1 - \frac{1}{q} - \log_e q\right) \log_e(1-q) - F(q) \right]$$

where A and B are constants of integration and $F(q)$ is given by

$$(3.91) \quad F(q) = \int \frac{\log q}{1-q} \, dq$$

Using the boundary conditions, we get

$$(3.92) \quad B = A + 32N_e^2 F(1)$$

$$(3.93) \quad K_2 = A - 32N_e^2 \left[1 + \int_o^1 \frac{\log q}{1-q} \, dq\right]$$

Thus $V_1(q)$ is given by

$$(3.94) \quad V_1(q) = 32N_e^2 \left[-\int_q^1 \frac{\text{Log}_e\, q}{1-q}\, dq - \left(1 - \frac{1}{q} - \text{Log}_e\, q\right) \text{Log}_e(1-q) \right]$$

$$+ \left(1 - \frac{1}{q}\right)\left[K_2 + 32N_e^2 \left(\int_0^1 \frac{\text{Log}_e\, q}{1-q}\, dq + 1\right)\right]$$

Hence if

$$(3.95) \quad K_2 = -32N_e^2 \left(\int_0^1 \frac{\text{Log}_e\, q}{1-q}\, dq + 1\right)$$

we get

$$(3.96) \quad V_1(q) = 32N_e^2 \left[-\left(1 - \frac{1}{q}\right)\text{Log}_e(1-q) + \text{Log}_e\, q\, \text{Log}_e(1-q)\right.$$

$$\left. -\int_q^1 \frac{\text{Log}\, q}{1-q}\, dq\right]$$

Now

$$(3.97) \quad f(q) = F(q) - F(1)$$

$$= -\int_1^q \frac{\text{Log}_e\, q}{q-1}\, dq$$

is a dilogarithm whose series expansion is given by

$$(3.98) \quad f(q) = \sum_{k=1}^{\infty} (-)^k (q-1)^k / k^2$$

and there is a functional relationship between $f(q)$ and $f(1-q)$ given by

$$(3.99) \quad f(q) + f(1-q) = - \text{Log}_e q \, \text{Log}_e (1-q) + \frac{\pi^2}{6} \quad (1 \geqslant q \geqslant 0)$$

(for these results on Dilogarithm see Abramowitz and Stegun 1965)

Hence $V_1(q)$ reduces to

$$(3.100) \quad V_1(q) = 32 N_e^2 [(\frac{1-q}{q}) \text{Log}_e (1-q) + \frac{\pi^2}{6} - \sum_{k=1}^{\infty} q^k / k^2]$$

Also

$$(3.101) \quad \lim_{q \to 0} V_1(q) = K_2$$

$$= 32 N_e^2 (\frac{\pi^2}{6} - 1)$$

The variance of time until fixation for $q \to 0$ is then given by

$$(3.102) \quad V = K_2 - K_1^2$$

$$= 16 N_e^2 (\frac{\pi^2}{3} - 3)$$

This gives a coefficient of variation of about 53.8%.

CHAPTER   IV

SELECTION WITH BINOMIAL TRANSITION PROBABILITIES

In this chapter numerical results on selection at a diallelic
locus with binomial transition probabilities are presented. The ratios
of the response to selection in the limit to the initial response,
analytically treated in Chapter III in the case of additive genes, are
obtained exactly (using matrix formula) for additive, recessive and
dominant genes. For additive genes results are presented for the mean
and the coefficient of variation of time to fixation. A comparison
between the times to fixation, loss and homozygosity is also made in
this case. For recessive and dominant genes however, only the mean
times to fixation are presented. All numerical values were obtained with
the help of the KDF9 computer of Edinburgh University.

4.1   ADDITIVE GENES

The matrix formulae (2.15), (2.19), (2.20) and (2.22) discussed in
Chapter II and analogous formulae for time to loss and homozygosity
were used for obtaining computer results presented in this section.

4.11   RATIO OF SELECTION LIMIT TO INITIAL RESPONSE

It is shown in Robertson (1960), that the chance of fixation of a
gene and hence the selection limit are dependent on Ns only so that
these can be evaluated for a given N at a series of values of s.

This would give results for a series of values of Ns applicable to values of N other than the one used if s is suitably adjusted. Hence the ratios have been worked out for N = 8 and Ns = 0.031 to Ns = 8.000. Further, the formulae (3.24) and (3.31) show that these ratios can be expressed in terms of the population size. The factor (ratio / N) has, therefore, been presented in Table 2 for initial gene frequencies 0.0625, 0.5000 and 0.9375. It is apparent that for q = 0.5000 and 0.9375 it decreases as Ns increases. For q = 0.0625, it increases as Ns increases till Ns = 2 after which it starts decreasing. When Ns is as small as 0.031, the value of this factor is greater than 2 when q is less than $\frac{1}{2}$ but less than 2 when q is greater than $\frac{1}{2}$. At q = $\frac{1}{2}$, it is equal to 2. When Ns is as high as 8.00, it is about 1.7, 0.5 and 0.4 for q $< \frac{1}{2}$, = $\frac{1}{2}$ and $> \frac{1}{2}$ respectively. It always decreases as gene frequency increases at a particular value of Ns.

4.12 MEAN AND VARIABILITY OF TIME TO FIXATION

The evaluation of matrix formulae on the computer depends on the population size (N) and the selective coefficient of the gene (s). But it is known (Robertson, 1960 and Hill and Robertson, 1966) that, under the conditions in which diffusion approximation holds, the time scale of the selection process is proportional to N and therefore if

TABLE 2: Values of (ratio / N) as calculated by the
transition matrix method for different values of Ns
(additive genes)

	INITIAL FREQUENCY (q)		
Ns	0.0625	0.5000	0.9375
0.031	2.016	2.000	1.984
0.063	2.031	1.999	1.969
0.125	2.062	1.997	1.937
0.250	2.121	1.990	1.873
0.500	2.233	1.927	1.743
1.000	2.418	1.854	1.492
2.000	2.611	1.535	1.080
4.000	2.490	0.970	0.648
8.000	1.692	0.500	0.383

the time is measured in units of N, the pattern of the selection process is determined by the parameter Ns at a given initial gene frequency. The mean time has, therefore, been expressed in units of N. In order to see whether the mean time and the coefficient of variation for fixation of a gene, disregarding the cases in which it is lost, is a function of Ns, a comparison of these for a few population sizes at a particular value of Ns is shown in Tables 3 and 4 respectively. There is found to be a fair degree of stability in these quantities due to variations in N at a fixed value of Ns. The dependence of the mean time and the coefficient of variation on Ns has, therefore, been graphically shown in Figures 2 and 3 respectively for initial gene frequencies 0.0312, 0.5000 and 0.9687. Both the mean and the coefficient of variation decreases as Ns increases. For a fixed Ns, however, the mean time is highest at low initial gene frequency and lowest at high initial gene frequency whereas the coefficient of variation is highest at high gene frequency and lowest at low gene frequency. When $Ns = 1$ and the initial gene frequency is 0.5, the mean time is about $2N$ with a percent coefficient of variation of about 70, but for low initial gene frequency of 0.0312, the mean rises to about $3N$ with a percent coefficient of variation of about 50. It is interesting to observe that genic selection shortens the fixation time but increases the rate of steady decay (Kimura, 1957).

TABLE   3   The average number of generations / N until
                  fixation of a gene with selective value s
                  as calculated by the transition matrix
                  method for different population sizes (N).

Ns	N	Initial frequency		
		0.25	0.50	0.75
16	16	0.46	0.36	0.27
8	16	0.74	0.56	0.40
	8	0.82	0.61	0.43
4	8	1.29	0.94	0.63
	4	1.40	1.02	0.67
2	4	2.03	1.51	0.95
	2	2.07	1.55	1.01

TABLE 4   The coefficient of variation (%) of the
number of generations until fixation of a
gene with selective value s as calculated
by the transition matrix method for dif-
ferent population sizes (N).

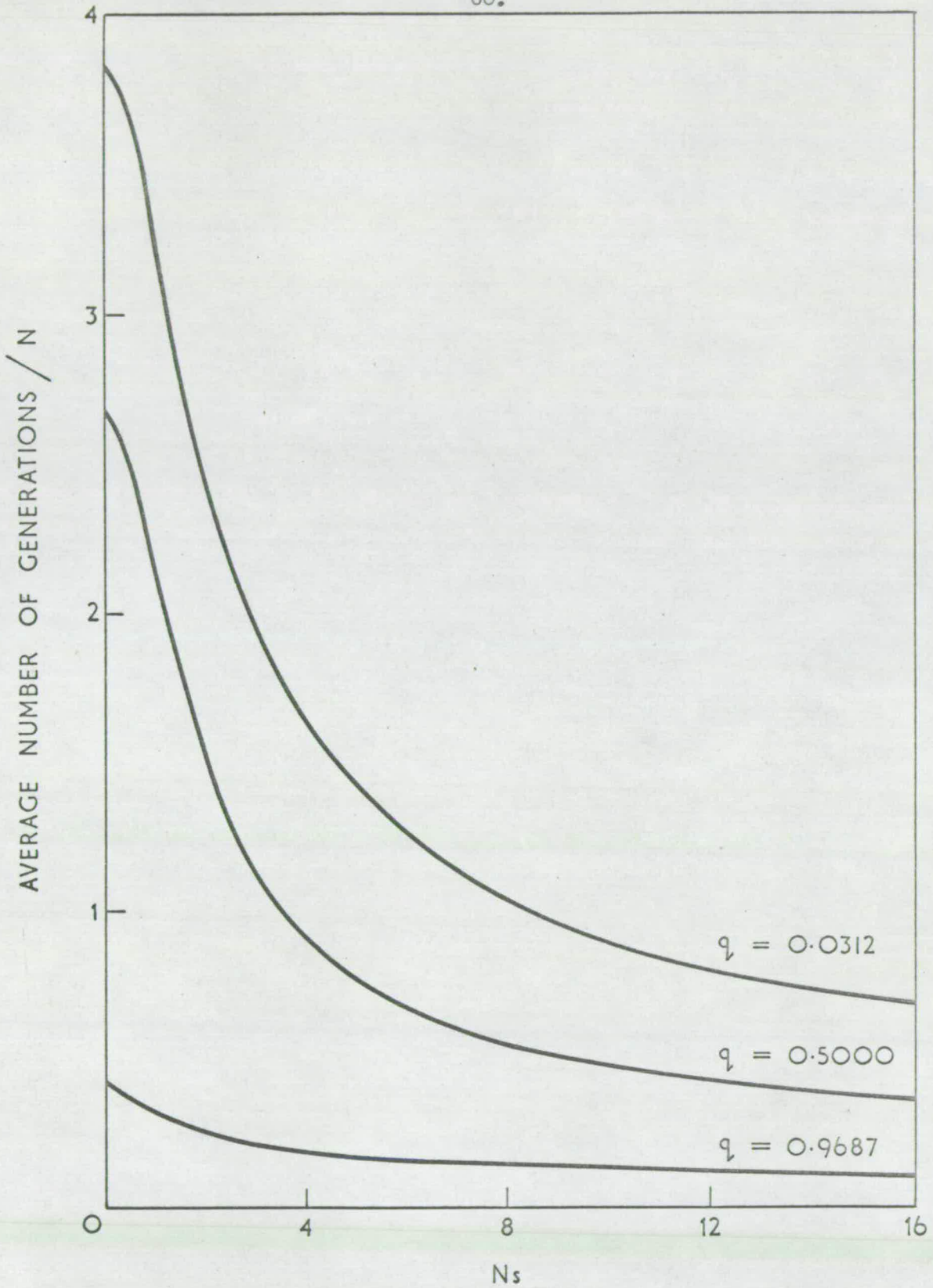| Ns	N	Initial frequency		
		0.25	0.50	0.75
16	16	28.36	34.25	43.68
8	16	33.04	39.23	50.23
	8	35.33	42.20	54.32
4	8	42.59	50.44	64.26
	4	45.24	54.62	69.40
2	4	53.12	65.06	84.26
	2	54.70	68.25	82.66

Fig. 2   Average number of generations/N until fixation of a
          gene acting additively. The curves are drawn for
          different initial gene frequencies.
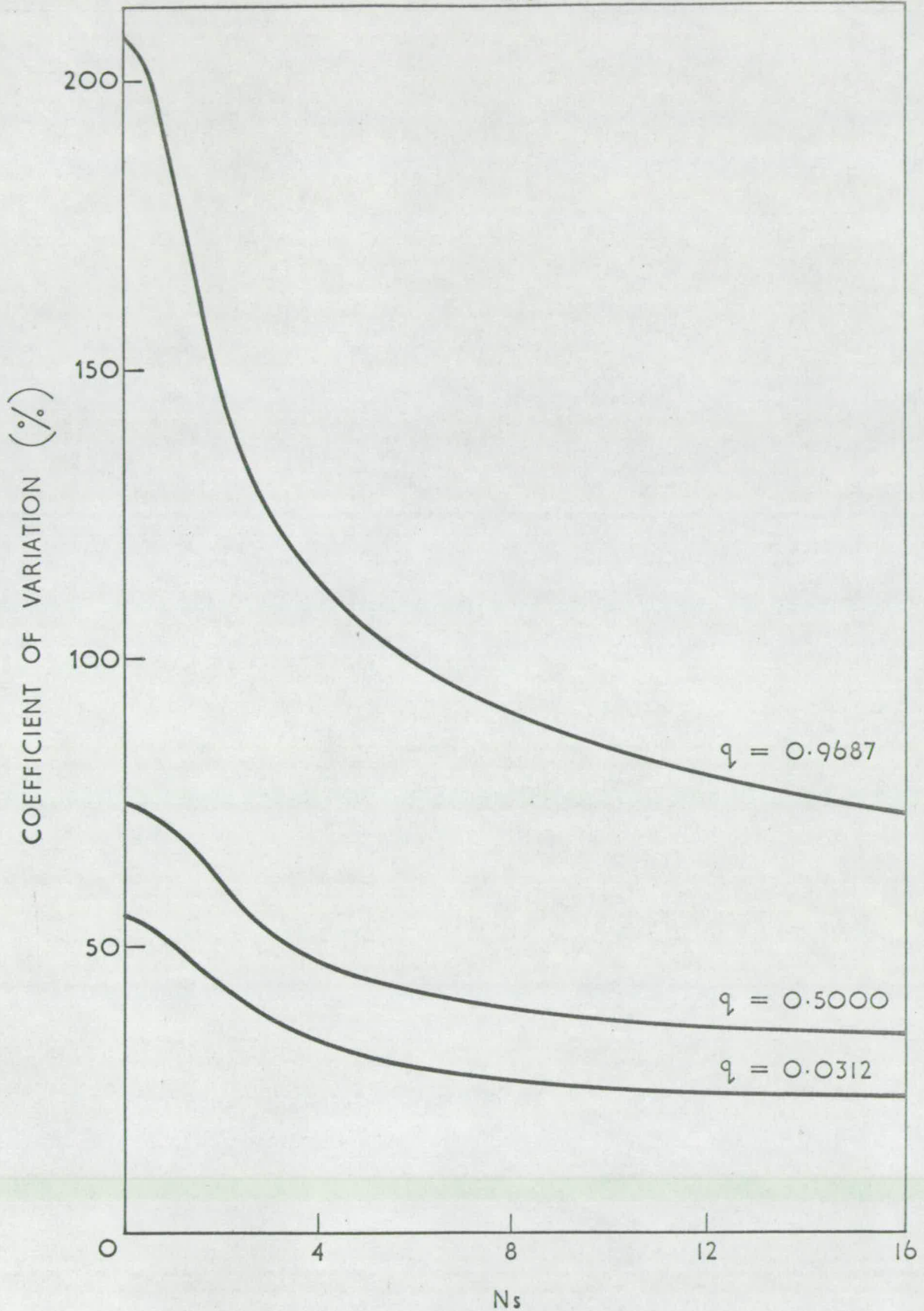
Fig. 3   Coefficient of variation (%) of number of generations
          until fixation of a gene acting additively. The curves
          are drawn for different initial gene frequency.

When there is no selection, the mean increases linearly with N and the coefficient of variation is practically steady. However, in the case of mean time, the slopes and in the case of coefficient of variation, the average values, vary with initial gene frequency. These variations are shown respectively in Figures 4 and 5. The slope increases but the average percent coefficient of variation declines as the gene frequency decreases from 1 to 0. Extrapolating from these graphs it is found that the limiting value of the slope and the average percent coefficient of variation are about 4 and 55 respectively as gene frequency approaches zero, as against the diffusion approximation results of 4 and about 54 respectively.

4.13 COMPARISONS OF TIMES TO FIXATION, LOSS AND HOMOZYGOSITY

It is interesting to compare the mean and the variability of the time to fixation of a particular allele with those of the time to loss and to homozygosity. As far as diffusion approximation to the mean times are concerned, the results of Ewens (1964) and Kimura and Ohta (1968), provide the following expressions for a neutral gene with initial frequency q.
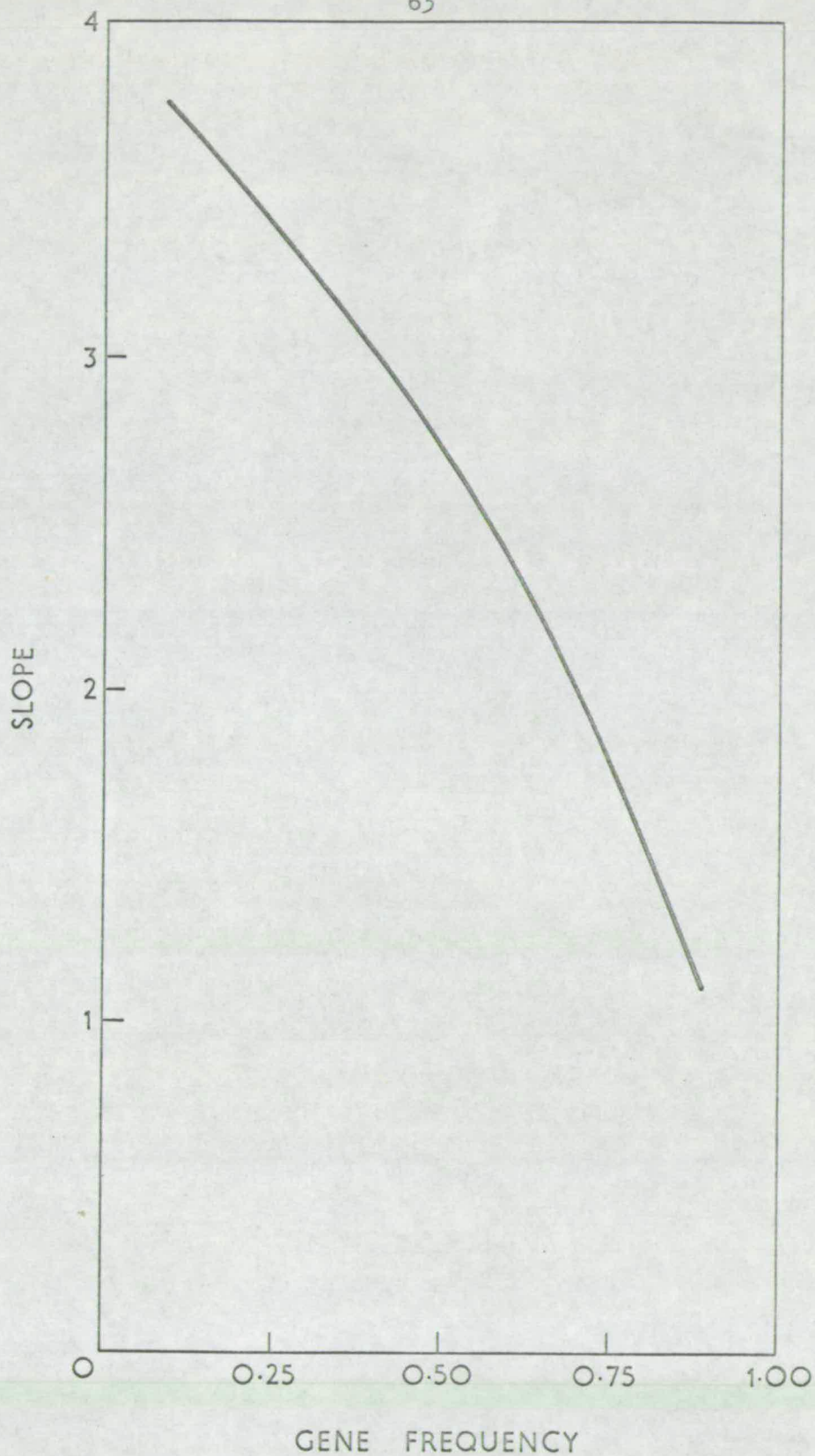
Fig. 4.   Slope of average number of generations
until fixation on population size at
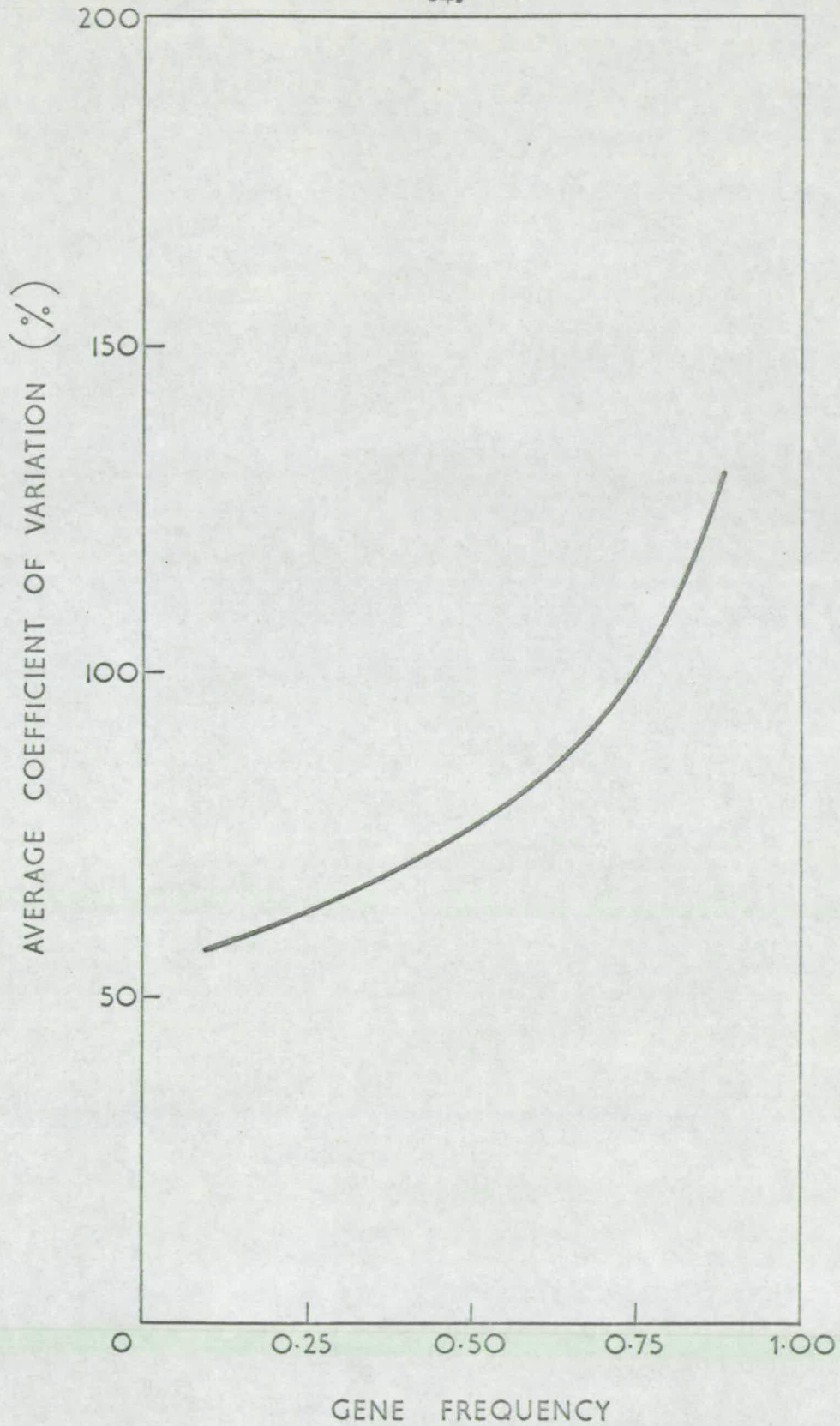different values of initial gene
frequency.

Fig. 5. Coefficient of variation (%) of number
of generations until fixation averaged
over population sizes between 2 to 16
at different values of initial gene
frequency.

Mean time until fixation $\qquad - 4N_e \dfrac{1-q}{q} \, Log_e \, (1-q)$ Kimura & Ohta
(1968)

Mean time until loss $\qquad - 4N_e \dfrac{q}{1-q} \, Log_e \, q$ Kimura & Ohta
(1968)

Mean time until homozygosity $\qquad - 4N_e [q \, Log_e \, q + (1-q) \, Log_e \, (1-q)]$

Ewens (1964)

In these formulae N_e is the variance effective number which may
differ from the actual population number N if the mating is not
random or if the distribution of the number of offsprings does not
follow a Poisson distribution. According to Crow and Morton (1955), the
formula connecting N and N_e is given by

$$(4.1) \quad N_e = \frac{2N}{1 - F' + (1 + F') \, V_k/\mu_k}$$

where F' is Wright's coefficient of inbreeding used as a measure

of the departure from random mating zygotic proportions among the parents, μ_k and V_k are the mean and the variance of the number of surviving offspring per parent. When $q = 1/(2N)$ and N is very large, the mean time until fixation is close to $4N_e$ while that until loss is $\frac{2N_e}{N} \text{Log}_e 2N$ so that the mean time until homozygosity should approach the value $\frac{2N_e}{N} \text{Log}_e 2N$. These comparisons, on the basis of transition matrix methods, are presented in Figures 6 and 7 for $Ns = o$ and 2 respectively. The dotted lines refer to time until fixation and until loss (marked respectively 1 and 2 in the figures) whereas solid line (marked 3 in the figures) refers to time until homozygosity. When $Ns = 0$ and initial frequency of gene is 0.5, all the three curves give the same value of about 2.55N as against the diffusion approximation of $-4N_e \text{Log}_e \frac{1}{2} = 2.8N_e$. The diffusion approximation, therefore, overestimates the mean time. Comparison with the half-life of $1.4N_e$, shows that the half-life is attained much earlier than the mean time. For initial gene frequencies greater than or less than 0.5, the three curves differ and as expected the curve 3 always lies between the curves 1 and 2. For $q < 0.5$ the mean time until homozygosity is more near the mean time until loss than that until fixation so that when $q = 1/(2N)$, the mean time until loss is 0.79N as against that until homozygosity
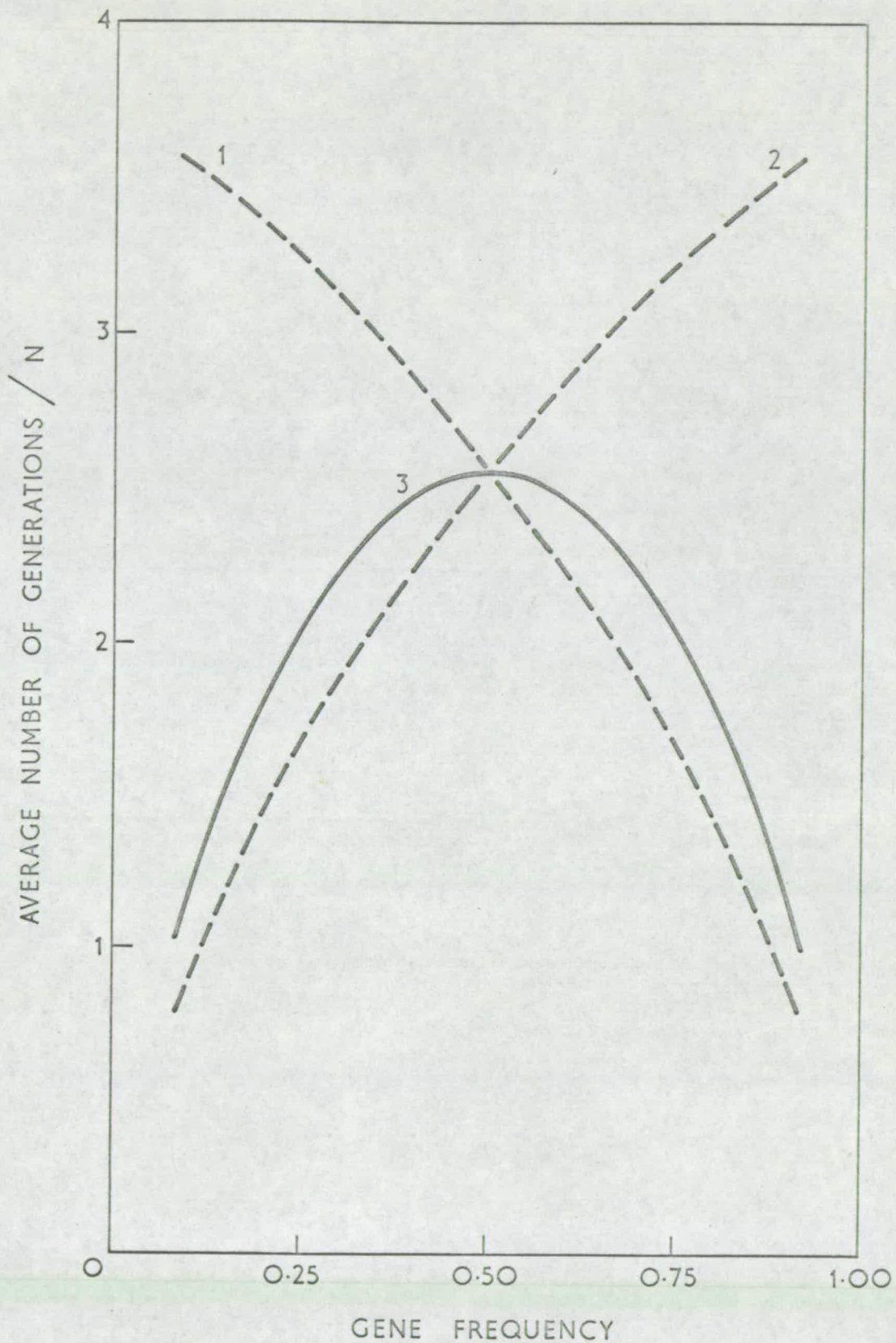
Fig. 6.   Average number of generations/N until fixation
(1), loss (2) and homozygosity (3) at different
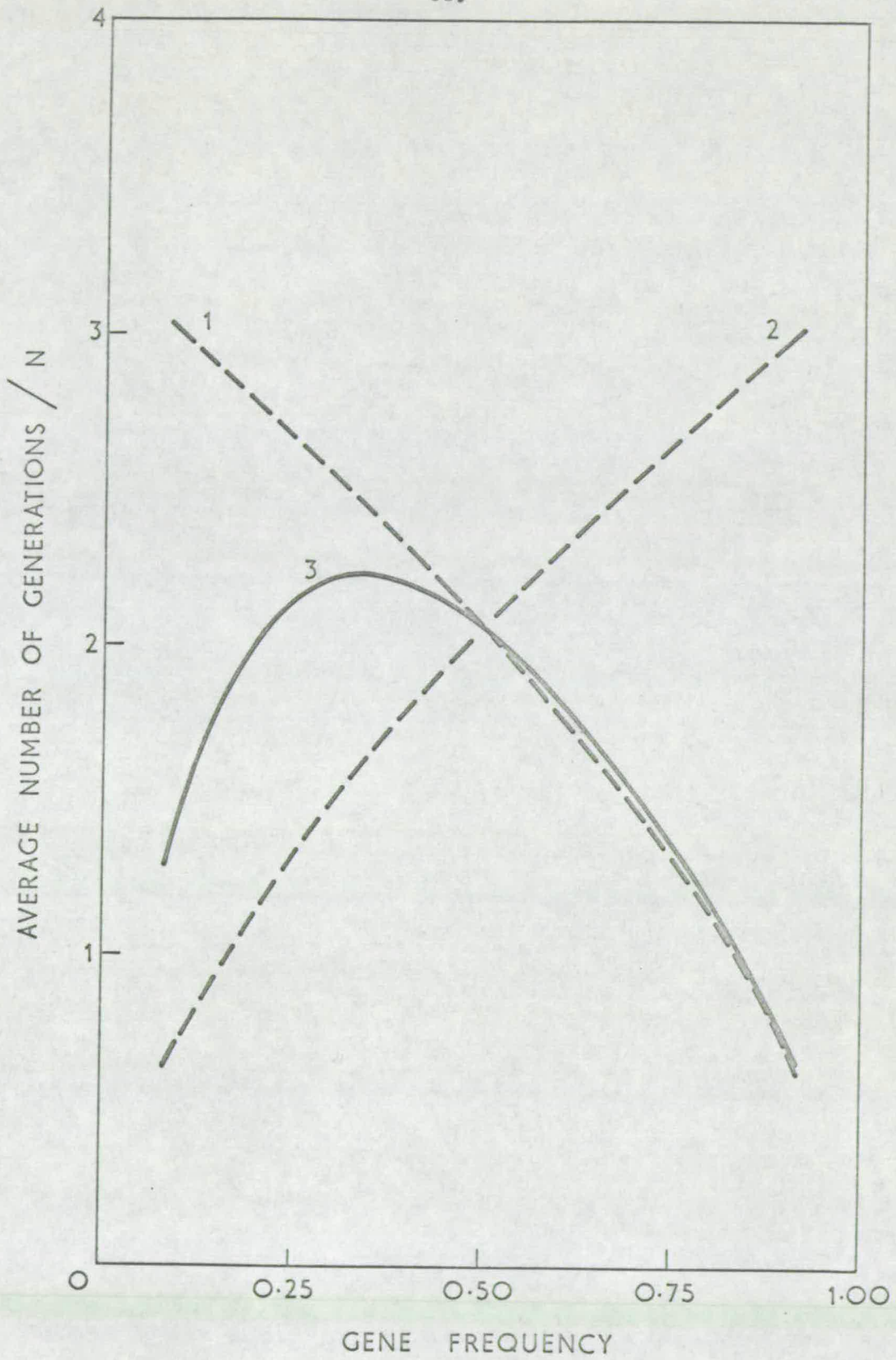values of initial gene frequency. (Ns = 0).

Fig. 7. Average number of generations/N until fixation (1), loss (2) and homozygosity (3) at different values of initial gene frequency. (Ns = 2).

of 1.02N. The difference is expected to decrease as the computer
results from higher values of N are compared. In the limit, these
should, therefore, be the same as expected from the diffusion
approximation. For $q > 0.5$, the mean time until homozygosity gets
nearer to the mean time until fixation. When $Ns = 2$, the mean time
until homozygosity is found to be practically the same as the mean
time until fixation unless the gene frequency is well below 0.5. It
is interesting to observe that, while with no selection, the maximum
mean time until homozygosity occurs at $q = 0.5$, with selection
this maximum shifts and occurs at gene frequency less than 0.5.
This shift has further been found to increase as selection becomes
more intense.

The comparisons of the coefficients of variation are shown in
Figures 8 and 9 for $Ns = 0$ and 2 respectively. As before dotted
lines refer to time until fixation (1) and loss (2) whereas solid line
(3) refers to homozygosity. When $Ns = 0$ and initial gene frequency is
0.5, all the three curves give the same value of about 76%. For
$q < 0.5$, the coefficient of variation for time until homozygosity is
more near the mean time until loss than that until fixation. For
$q > 0.5$, it gets nearer to that until fixation. When $Ns = 2$, the
coefficients of variation for homozygosity and fixation are practically
the same unless q is well below 0.5. It is interesting to note
that the coefficient of variation of time until homozygosity is minimum
at $q = 0.5$ with no selection but this minimum shifts to $q < 0.5$
with selection.

Fig. 8. Coefficient of variation (%) of number of
generations until fixation (1), loss (2) and
homozygosity (3) at different values of initial
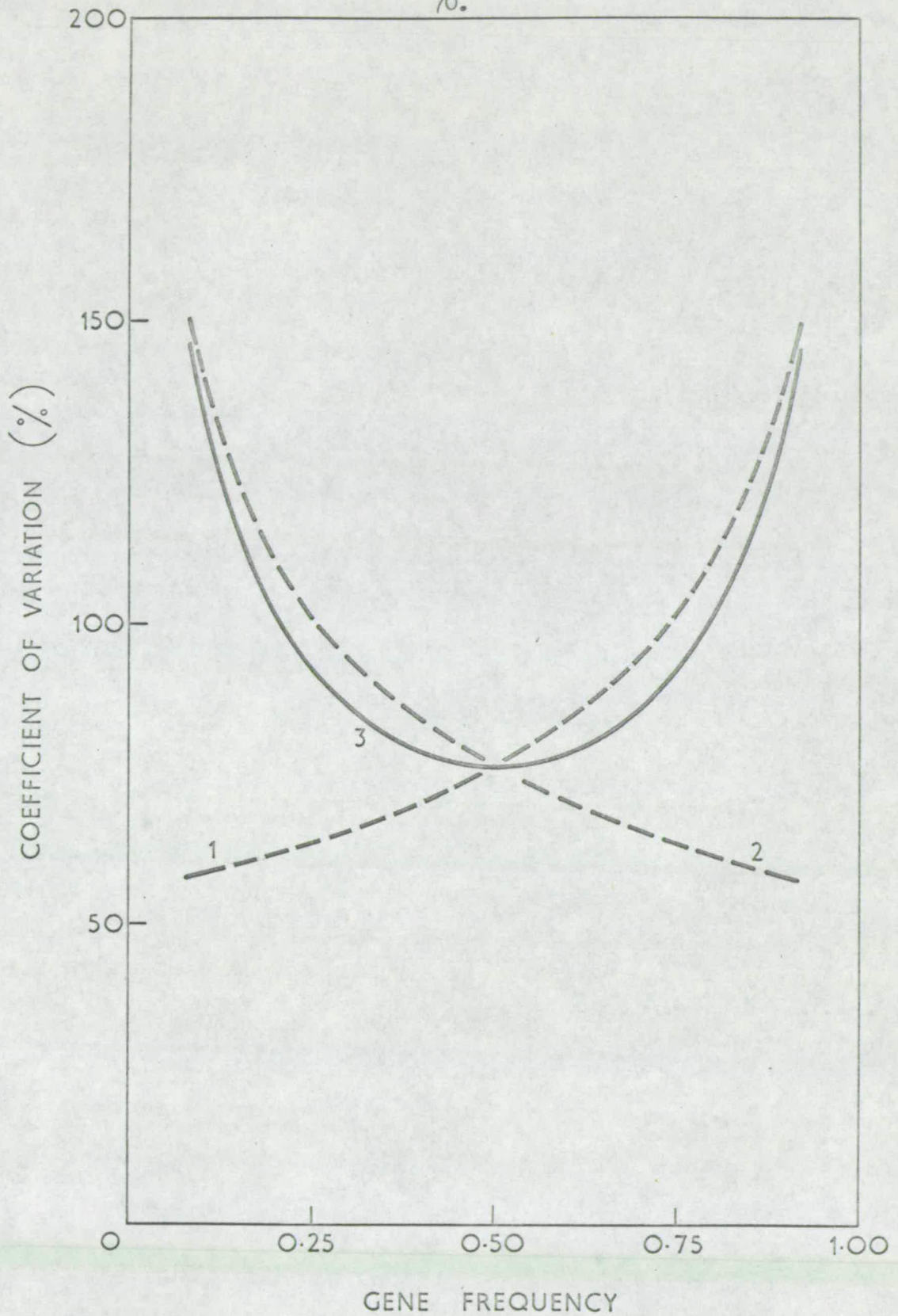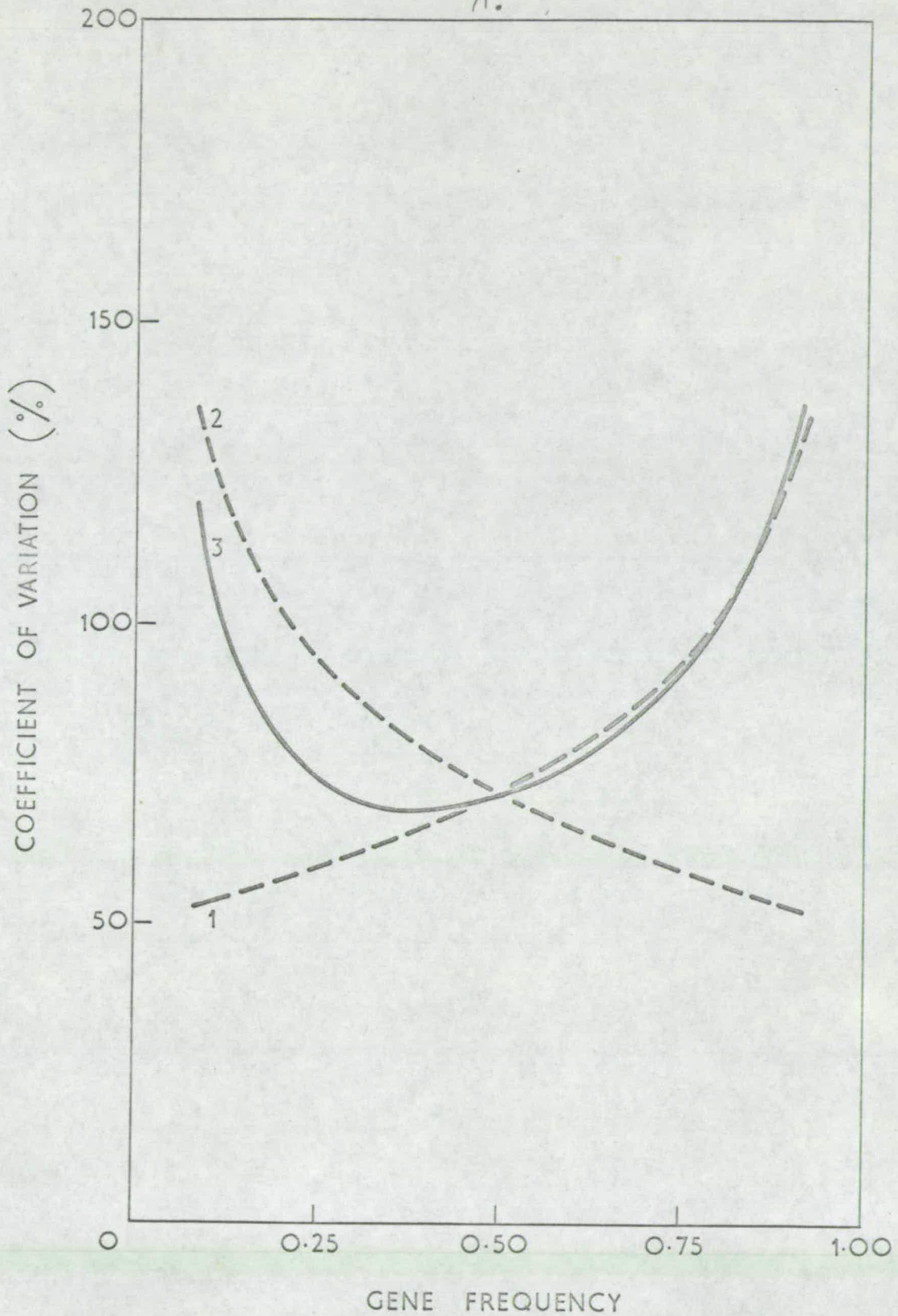gene frequency. (Ns = 0).

Fig. 9. Coefficient of variation (%) of number of
generations until fixation (1), loss (2) and
homozygosity (3) at different values of initial
gene frequency. (Ns = 2).

4.2   RECESSIVE AND DOMINANT GENES

The matrix formulae used in the case of additive genes were also used for obtaining results in respect of recessive and dominant genes with the difference that the change in the gene frequency in one-step was fed in as

$$(4.2) \quad \delta q_1 = \frac{\left(1 + \frac{s}{2}\right) q_1^2 + q_1 (1 - q_1) \left(1 + \frac{h}{2}\right)}{1 - \frac{s}{2} (1 - 2q_1) + h \, q_1 (1 - q_1)}$$

where $h = -s$ for recessive genes and $h = +s$ for dominant genes.

4.21   RATIO OF SELECTION LIMIT TO INITIAL RESPONSE

The factors (ratio / N) are given in Tables 5 and 6 respectively for completely recessive and completely dominant genes. In the case of recessive genes, it is found that for $q = 0.5$ and 0.9375 the factor decreases as Ns increases. For $q = 0.0625$ it increases as Ns increases till $Ns = 0.25$ after which it starts decreasing. When Ns is as small as 0.031, its value is 11.14 for $q = 0.0625$. This is in contrast with the additive genes where it is only 2.016. For $q = 0.5000$ and $q = 0.9375$, however, the factor is less than their corresponding values in the additive case. When Ns is as high as 8.00 it is about 5.10, 0.38 and 0.20 for $q < \frac{1}{2}$, $= \frac{1}{2}$ and $> \frac{1}{2}$ respectively. In the case of dominant genes, however, it is found that

TABLE 5   Values of (ratio / N) as calculated by the
            transition matrix method for different values of
            Ns (recessive genes).

Ns	INITIAL FREQUENCY (q)		
	0.0625	0.5000	0.9375
0.031	11.141	1.995	1.374
0.063	11.150	1.989	1.356
0.125	11.165	1.977	1.322
0.250	11.177	1.950	1.254
0.500	11.142	1.884	1.125
1.000	10.874	1.719	0.900
2.000	9.920	1.361	0.595
4.000	7.913	0.829	0.338
8.000	5.101	0.375	0.196

TABLE   6   Values of (ratio/N) as calculated by the
               transition matrix method for different
               values of Ns (dominant genes)

Ns	INITIAL FREQUENCY (q)		
	0.0625	0.5000	0.9375
0.031	1.409	2.005	11.118
0.063	1.426	2.010	11.105
0.125	1.461	2.018	11.073
0.250	1.532	2.030	10.994
0.500	1.670	2.039	10.765
1.000	1.918	1.989	10.066
2.000	2.194	1.700	8.178
4.000	2.014	1.104	5.270
8.000	1.184	0.625	3.191

for q = 0.9375 the factor decreases as Ns increases but for
q = 0.0625 and 0.5000 it increases initially and then decreases after-
wards with the increase in Ns values. When q = 0.5, it increases
from 2.005 (when Ns = 0.031) to 2.039 (when Ns = 0.500) and
decreases thereafter. For q = 0.0625, however, it increases from
1.409 (when Ns = 0.031) to 2.194 (when Ns = 2.000) and decreases
thereafter. When Ns = 0.031 the factor increases as q increases from
0.0625 to 0.9375. This trend continues upto Ns = 1.00 but when
Ns is equal to 2 or greater then 2, it is lowest at q = 0.5000
and highest at q = 0.9375.

4.22   MEAN TIME TO FIXATION

The behaviours of the mean time until fixation of a recessive and
dominant gene with variation in Ns are shown in Figures 10 and 11 for
initial gene frequencies 0.0625, 0.5000 and 0.9375. As found, in the
case of additive genes, the time to fixation decreases as Ns increases,
as far as a recessive gene is concerned but for a dominant gene, it
increases initially for small values of Ns and then decreases. For a
gene with low frequency the maximum occurs at a value of $Ns < 1$
whereas for gene frequency as 0.5 it occurs at Ns = 1 and for high
gene frequency it occurs at $Ns > 1$. For a rare dominant gene, therefore,
the limiting value of the mean time until fixation may well be above
$4N_e$. For gene frequency equal to half (Ns = 1) a dominant gene would
take about 2.77N generations to fixation. On the other hand a rare
recessive gene at Ns = 1 takes about 3.18N generations. It would
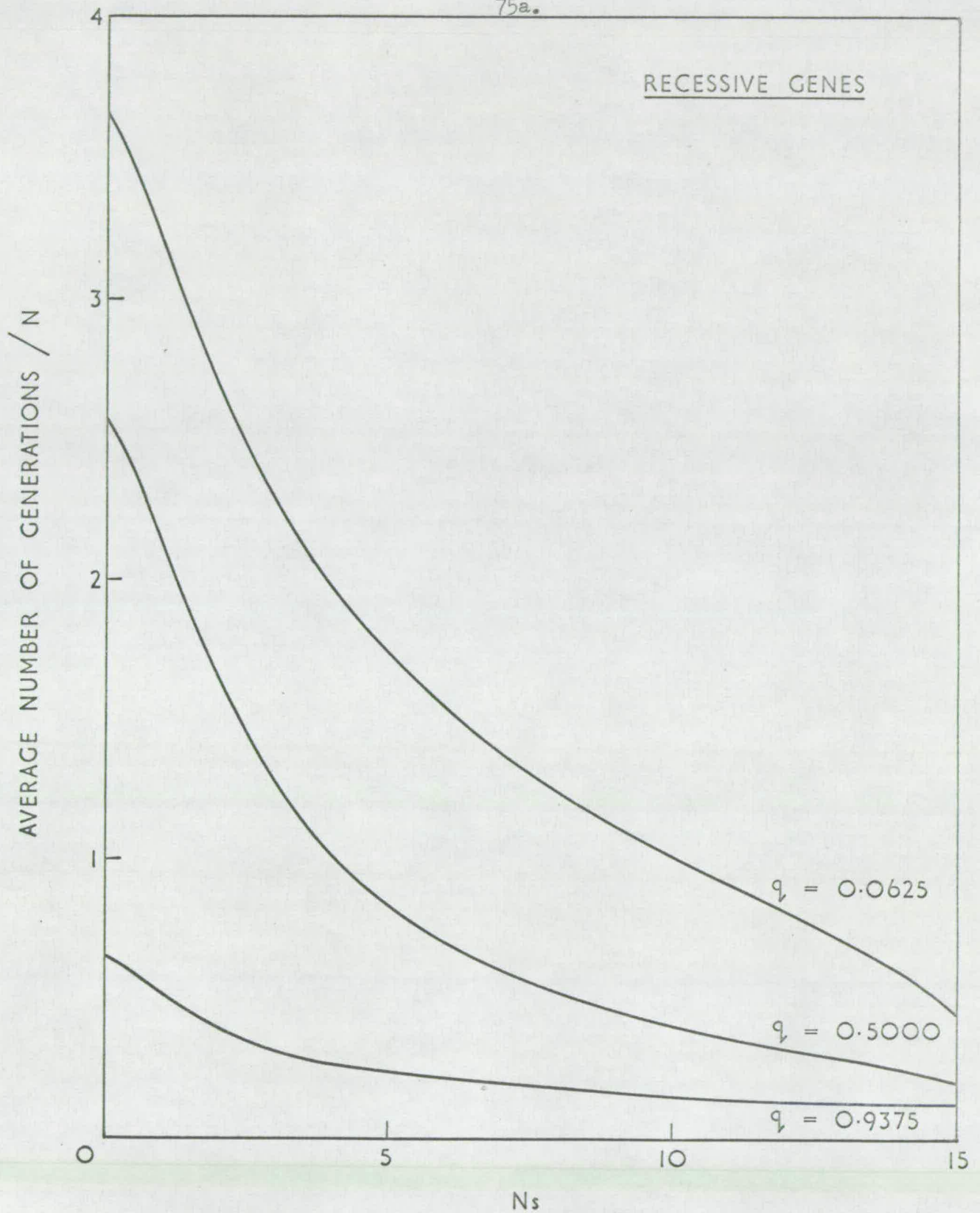
Fig. 10   Average number of generations/N until fixation of a recessive
gene.   The curves are drawn for different initial recessive
frequencies.
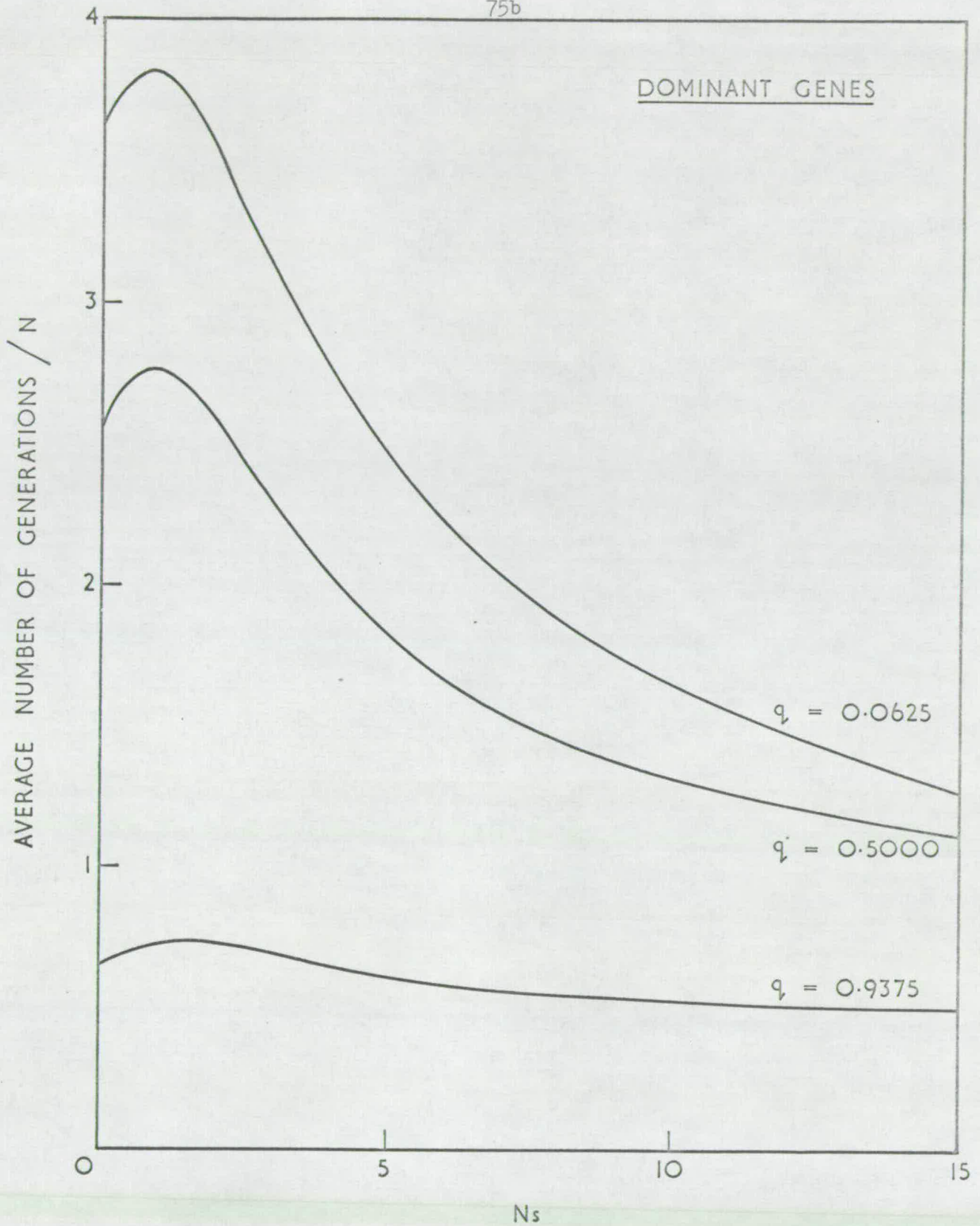
Fig. 11   Average number of generations/N until fixation of a dominant
gene. The curves are drawn for different initial dominant
frequencies.

take 2.14N generations if it is as frequent as its alternative allele. Thus a gene takes less time to reach fixation when it is recessive than when it is dominant. It is interesting to compare this result with those of Kimura (1957) on the behaviour of the final rate of decay for recessive and dominant genes. Selection towards dominants decreases the final rate of decay whereas selection against dominants increases it.

CHAPTER V

SELECTION WITH TRINOMIAL TRANSITION PROBABILITIES

The theory developed in Chapter II for selection with random drift in the case of single locus with three alleles can be applied to specific genetic models. In this section numerical results on the chance of fixation of a particular allele A_1, the average number of generations taken until its fixation and the average time to homozygosity are presented when the transition probabilities are of the trinomial type and the relative selective advantages of A_1, A_2 and A_3 alleles are respectively $(1 + s_1/2)$, $(1 + s_2/2)$ and 1.

Suppose first there are no selective forces operating. Then starting with a given pair of frequencies $q_{1i} = i_1/(2N)$, $q_{2i} = i_2/(2N)$ of alleles A_1 and A_2 respectively (with $1 - q_{1i} - q_{2i}$ = frequency of A_3), in lines of constant breeding size of N individuals, we can consider the second generation as derived from the first by the sampling of groups of 2N haploid sets, the gene frequencies in the different groups being distributed trinomially with pairs of means $2Nq_{1i}$ and $2Nq_{2i}$ and index 2N. The next generation is then the repetition of this process, each line giving rise to a group of lines whose gene frequencies are trinomially distributed about the pairs of means of the parent line. Now let A_1 genes have a selective advantage of $(1 + s_1/2)$, A_2 genes have a selective advantage of $(1 + s_2/2)$ and A_3 genes have a selective advantage of unity where s_1 and s_2

are small and positive. This means that while the selective advantage of A_1 over A_3 is $(1 + s_1/2)$ that of A_1 over A_2 is $(1 + (s_1 - s_2)/2)$ approximately. Assuming that selection operates before sampling of gametes, the gene frequencies of A_1 and A_2 in the different groups are distributed trinomially with means $2Nq'_{1i}$ and $2Nq'_{2i}$ and index $2N$, where q'_{1i} and q'_{2i} are the gene frequencies of A_1 and A_2 after selection and are given by

$$(5.1) \quad q'_{1i} = q_{1i} + \delta q_{1i}$$

$$= q_{1i} + \frac{\frac{s_1}{2} q_{1i}(1 - q_{1i}) - \frac{s_2}{2} q_{1i} q_{2i}}{(1 + q_{1i} s_1/2 + q_{2i} s_2/2)}$$

and

$$(5.2) \quad q'_{2i} = q_{2i} + \delta q_{2i}$$

$$= q_{2i} + \frac{\frac{s_2}{2} q_{2i}(1 - q_{2i}) - \frac{s_1}{2} q_{1i} q_{2i}}{(1 + q_{1i} s_1/2 + q_{2i} s_2/2)}$$

respectively. Here δq_{1i} and δq_{2i} are the changes in the mean frequencies of A_1 and A_2 respectively due to selection in one step when their frequencies in the previous generations are respectively $i_1/(2N)$ and $i_2/(2N)$. Thus with haploid selection and trinomial

sampling, the conditional probability that there are j_1 A_1 genes and j_2 A_2 genes out of $2N$ genes after one generation, given that there were i_1 A_1 genes and i_2 A_2 genes out of $2N$ in the previous generation, $P_{(i_1 i_2)(j_1 j_2)}$ are given by

$$(5.3) \quad P_{(i_1 i_2)(j_1 j_2)} = \frac{(2N)!}{j_1! \ j_2! \ (2N - j_1 - j_2)!} (q'_{1i})^{j_1} (q'_{2i})^{j_2}$$

$$\cdot (1 - q'_{1i} - q'_{2i})^{(2N - j_1 - j_2)}$$

These probabilities, therefore, determine the $\underset{\sim}{P}$, $\underset{\sim}{Q}$ and the components of $\underset{\sim}{Q}$ introduced in section 2 for the case of single locus with three alleles. With the help of the matrix formulae (2.33), (2.34) and (2.36) the chance of fixation of A_1 can be determined and in addition using (2.54), (2.55) and (2.57) gives the average time until its fixation. For average time until homozygosity matrix formulae (2.43), (2.44), (2.45) and (2.46) can be used. However, since the fixation of either A_2 or A_3 reduces the problem to that of a locus with two alleles, it is sufficient to consider only those situations in which initially neither of the A_2 and A_3 are absent. This means we have to use only the relations (2.36), (2.57) and (2.46). These have been evaluated numerically on the KDF9 computer of Edinburgh University. The capacity of the computer did not permit the use of a population size greater than six. Hence all the results presented relate to $N = 6$. The computer

programme developed for this purpose made regular use of the KDF9 matrix pack.

5.1 CHANCE OF FIXATION OF A_1

In order to present results in a meaningful way it is desirable to examine the dependence of the change in the frequency of A_1 on the selective coefficients of the other alleles and the population size. First we consider an infinite population and determine the parameters which govern the changes in the gene frequency of A_1 initially and in subsequent generations. Thereafter we examine how these parameters can be combined with the population size when a finite population is considered.

Consider three alleles A_1, A_2 and A_3 with selection coefficients $(1 + s_1/2)$, $(1 + s_2/2)$ and 1 with frequencies q_1, q_2 and q_3 respectively. The superiority of A_1 over A_2 is $(s_1 - s_2)/2$ whereas that of A_1 over A_3 is $s_1/2$. This gives an average superiority $\bar{s}$ and the variance in superiority V_m of A_1 over A_2 and A_3 considered together as a group, given by

$$(5.4) \quad \bar{s} = \tfrac{1}{2}(s_1 - rs_2)$$

$$(5.5) \quad V_m = \tfrac{1}{4} r (1 - r) s_2^2$$

where
$$(5.6) \quad r = q_2 / (q_2 + q_3).$$

The variance in the selective advantage of the three alleles can be partitioned as

$$(5.7) \quad V_s = q_1 (1 - q_1) \bar{s}^2 + (q_2 + q_3) V_m$$

$$= \tfrac{1}{2} \sigma_a^2 + \tfrac{1}{2} \sigma_m^2$$

where σ_a^2 is the contribution of the locus to the additive genetic variance of the genotypes when the three allele system is collapsed into two allele system and σ_m^2 is the component of total genetic variance due to the multiple allelic effect. That is the total additive genetic variance can be partitioned into two components, one reflecting the contribution of the locus with two effective alleles A_1 and $\bar{A}_2$ where $\bar{A}_2$ refers to the group of A_2 and A_3 and the other emphasising the contribution due to the distinction made between the two alleles A_2 and A_3 on the basis of their different selective advantages. This idea of distinguishing multiple alleles on the basis of their quantitative effects was first introduced by Narain (1965) in connection with the description of gene action in continuous variation when multiple alleles are taken into account. A similar partitioning was suggested for metric traits for describing the components of genetic variance. Thus V_m can be generalized and defined as multiple allelic variance in the selective advantages of the genotypes. If q_1' and q_1'' are the gene frequency of A_1 after the

first and the second generation of selection respectively then, approximately

$$(5.8) \quad q_1' = q_1 + \bar{s} \, q_1 \, (1 - q_1)$$

neglecting terms involving powers of s_2 greater than one.

$$(5.9) \quad q_1'' = q_1' + \bar{s}' \, q_1' \, (1 - q_1')$$
$$= q_1' + (\bar{s} - V_m) \, q_1 (1 - q_1) + \bar{s}^2 \, q_1 \, (1 - q_1) \, (1 - 2q_1)$$

neglecting terms involving powers of s_2 greater than two and where

$$(5.10) \quad \bar{s}' = \tfrac{1}{2} \, (s_1 - r' \, s_2)$$

$$(5.11) \quad r' = r[1 + (1 - r) \, (1 - r \, \frac{s_2}{2}) \, \frac{s_2}{2}]$$

$$(5.12) \quad q_1'(1 - q_1') = q_1(1 - q_1) \, [1 + \bar{s} \, (1 - 2q_1) - \bar{s}^2 \, q_1 \, (1 - q_1)]$$

From (5.8) and (5.9) it is clear that, in an infinite population, the change in the frequency of A_1 depends on $\bar{s}$ and V_m whereas in the two allele case it depends only on the superiority in selective advantage of one allele over the other. In the present case, when $V_m = o$, the change in the gene frequency is simply $\bar{s} \, q_1 \, (1 - q_1)$ if

the term involving $\bar{s}^2$ is neglected and this, in fact, corresponds to a two allele situation where A_2 and A_3 are indistinguishable is iso-alleles. The allele A_1 increases in frequency and eventually gets fixed. But when $\bar{s} = 0$, we get

$$(5.13) \quad q_1' = q_1$$

$$(5.14) \quad q_1'' = q_1 - V_m\, q_1\, (1 - q_1)$$

This shows that in the first generation, the frequency of A_1 does not change and in the second generation it decreases; the decrease being $-V_m\, q_1\, (1 - q_1)$. Under these conditions, it would eventually be eliminated from the population. When neither of $\bar{s}$ and V_m are zero, then initially the allele would increase in frequency but in the subsequent generations it would increase and get eventually fixed when the values of $\bar{s}$, V_m and q_1 make $(q_1'' - q_1')$ as positive. For q_1 equal to or less than half, the condition for this to happen is $\bar{s} > V_m$.

When the population is finite we write $\phi(q_1, q_2, x_1, x_2; t)$ for the probability density that the frequencies of A_1 and A_2 become x_1 and x_2 at the tth generation given that their frequencies are q_1 and q_2 at $t = o$. Using the diffusion model (Kimura, 1964) we can write down the process of change with time as

$$(5.15) \quad \frac{\partial \phi}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial x_1^2} [\frac{x_1 (1 - x_1)}{2N} \phi] + \frac{1}{2} \frac{\partial^2}{\partial x_2^2} [\frac{x_2 (1 - x_2)}{2N} \phi]$$

$$- \frac{\partial^2}{\partial x_1 \partial x_2} [\frac{x_1 x_2}{2N} \phi] - \frac{\partial}{\partial x_1} [(\frac{s_1}{2} x_1 (1 - x_1) - \frac{s_2}{2} x_1 x_2) \phi]$$

$$- \frac{\partial}{\partial x_2} [(\frac{s_2}{2} x_2 (1 - x_2) - \frac{s_1}{2} x_1 x_2) \phi]$$

This may be rearranged as

$$(5.16) \quad \frac{\partial \phi}{\partial (t/N)} = \frac{1}{4} \frac{\partial^2}{\partial x_1^2} [x_1 (1 - x_1) \phi] + \frac{1}{4} \frac{\partial^2}{\partial x_2^2} [x_2 (1 - x_2) \phi]$$

$$- \frac{1}{2} \frac{\partial^2}{\partial x_1 \partial x_2} [x_1 x_2 \phi]$$

$$- \frac{1}{2} Ns_1 [\frac{\partial}{\partial x_1} (x_1 (1 - x_1) \phi) - x_1 \frac{\partial}{\partial x_2} (x_2 \phi)]$$

$$- \frac{1}{2} Ns_2 [\frac{\partial}{\partial x_2} (x_2 (1 - x_2) \phi) - x_2 \frac{\partial}{\partial x_1} (x_1 \phi)]$$

This shows that, under the conditions in which the diffusion approximation
holds, the pattern of the selection process is entirely governed by
the parameters Ns_1 and Ns_2 on a time scale (t/N) starting from
a given initial configuration (q_1, q_2, q_3) of the gene frequencies of
the three alleles. Since the change in the frequency of A_1 in an

infinite population, is shown to be dependent on $\bar{s}$ and V_m, the parameters Ns_1 and Ns_2 can be transformed to $N\bar{s}$ and N^2V_m. In order to see whether the chance of fixation of A_1 is a function of $N\bar{s}$ and N^2V_m only, a comparison of these for population sizes $N = 3$ and 6 at particular values of $N\bar{s}$ and N^2V_m is shown in Table 7.

TABLE 7   The chance of fixation of A_1 as calculated by the
            transition matrix method for different population
            sizes (N)

$N\bar{s}$	N^2V_m	N	Initial frequencies (A_1, A_2, A_3)
			(1/3, 1/3, 1/3)
0.5000	0.0625	3	0.42033
		6	0.42299

The two values of the chance of fixation of A_1 are quite close, thus indicating that the population size N can appear in combination with $\bar{s}$ and V_m at particular initial gene frequencies of A_1, A_2 and A_3,

The dependence of the chance of fixation of A_1 on $N\bar{s}$ and N^2V_m are shown in Figures 12 and 13. In Figure 12, it is shown against $N\bar{s}$ for N^2V_m = 0, 1, 4 and 9 when initial frequencies (A_1, A_2, A_3) are (1/3, 1/3, 1/3). In Figure 13, it is shown against N^2V_m for initial frequencies (A_1, A_2, A_3) = (1/6, 5/12, 5/12), (1/2, 1/4, 1/4) and (5/6, 1/12, 1/12) when $N\bar{s}$ = 0. In the latter case, the three initial configuration of the gene frequencies have the same value of $r = 1/2$.

In Figure 12, the curve for N^2V_m = 0 gives the same value of the chance of fixation of A_1 as provided by the formula for two alleles given by Kimura (1957) as

$$(5.17) \quad u(q_1) = \frac{1 - \exp(-2Ns_1 q_1)}{1 - \exp(-2Ns_1)}$$

When N^2V_m is greater than zero, the chance of fixation of A_1, at a particular value of $N\bar{s}$, is reduced. It, however, varies with $N\bar{s}$, at a particular value of N^2V_m, in almost the same fashion as when N^2V_m = 0. But the characteristic feature, in this case, is that it becomes less than the initial frequency of the gene when $N\bar{s}$ is less than a certain value depending upon the particular value of N^2V_m considered. Thus when $N\bar{s}$ = 1 the chance of fixation is reduced from about 52.5% to about 30% when N^2V_m is increased from 0 to 1 and the
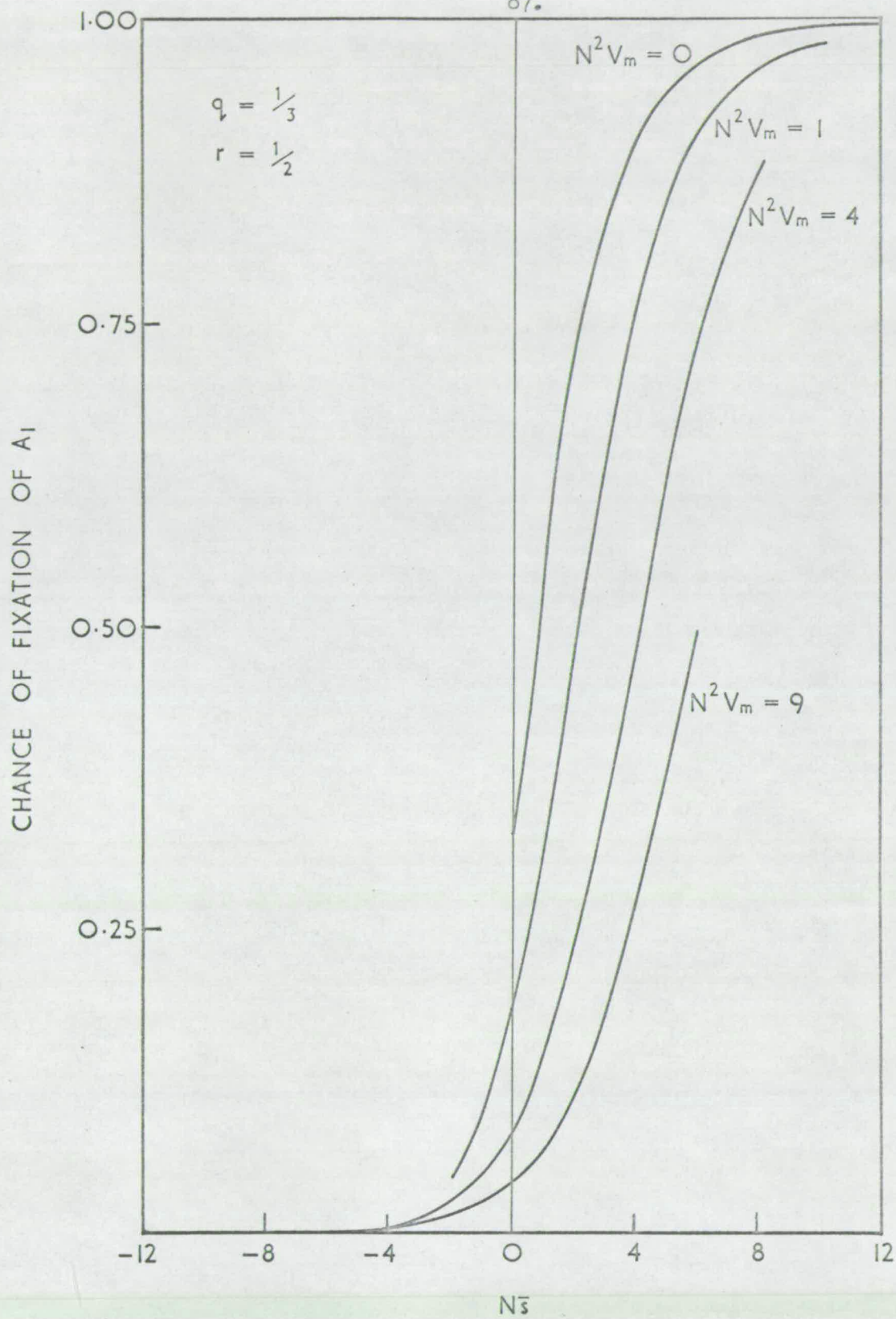
Fig. 12   The chance of fixation of allele A_1 at a tri-
allelic locus $(q_1 = 1/3, r = \frac{1}{2})$. The curves
are drawn for different values of $N^2 V_m$.
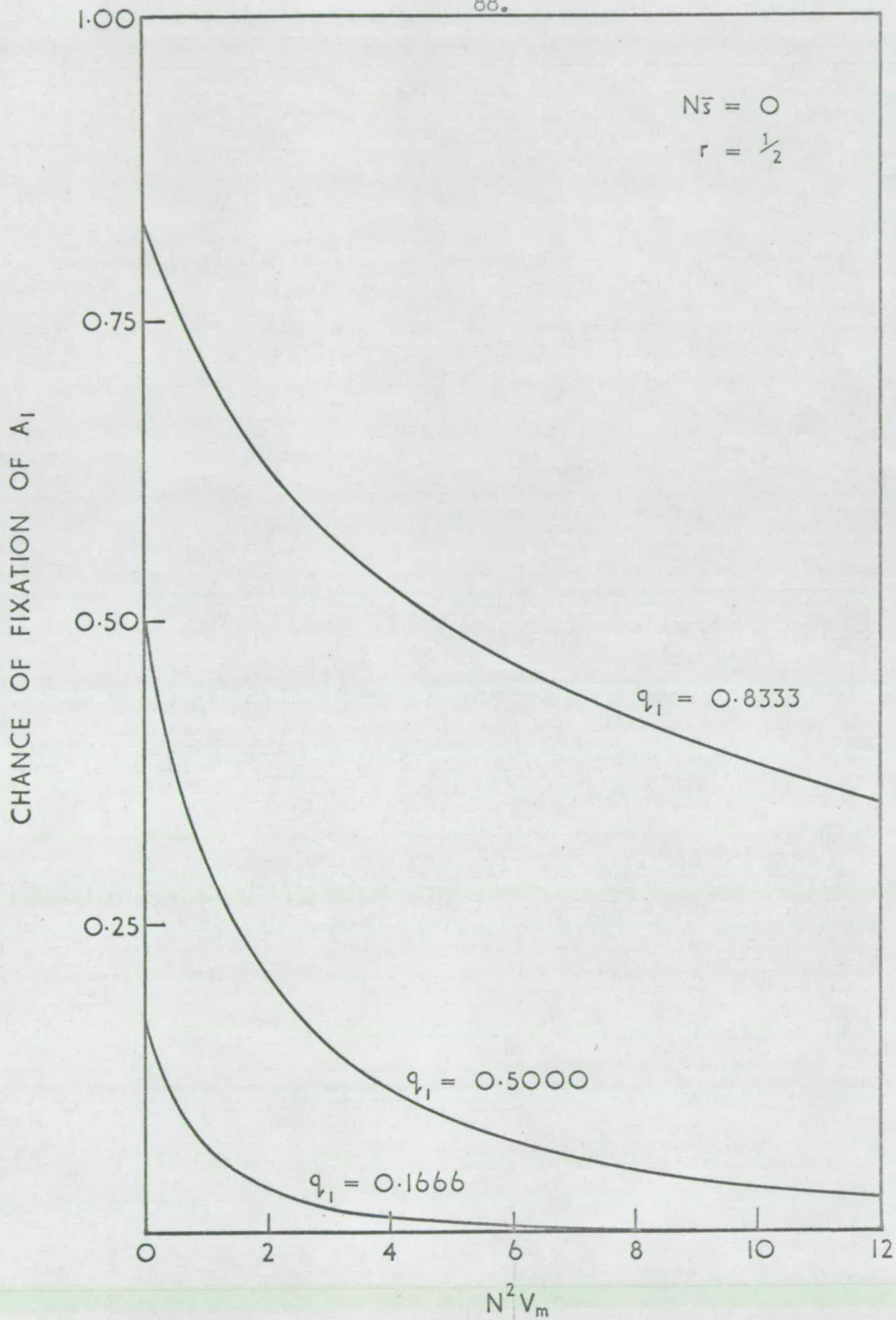
Fig. 13   The chance of fixation of allele A_1 at a
          tri-allelic locus ($N\bar{s} = 0$, $r = \frac{1}{2}$).[1] The curves
          are drawn for different initial frequencies of
          A_1.

chance of fixation becomes less than the initial gene frequency. In order to have the same chance of fixation at $N^2V_m = 1$ as that at $N\bar{s} = 1$ with $N^2V_m = 0$, $N\bar{s}$ is to be increased from 1 to about 2.2. The effect of the multiple allelic variance in reducing the probability of fixation of a gene can be compensated by a suitable increase in the average selective advantage of the gene. At very high values of N^2V_m, the chance of fixation could be practically zero unless there is some average selective advantage of the gene to compensate for this. It can be seen from the graph that when $N^2V_m = 9$, a neutral gene would have as small a chance as 4% but could have the same chance as its own initial frequency if it were of as high a selective advantage as $N\bar{s} = 4.5$.

Figure 13 clearly brings out the effect of multiple allelic variance on the chance of fixation of A_1 when it is neutral, on an average, with respect to A_2 and A_3 with variance V_m. In an infinite population, A_1 would ultimately be lost due to the effect of V_m but in a finite population, it could get fixed by chance. This chance fixation would increase as the population decreases in size. For small populations, random drift would predominate over its being selected against and the allele would get fixed but for large populations, selective forces acting against it would overcome random drift and the allele would have very little chance of getting fixed. Since the selective forces acting against A_1 are inherent in V_m, the chance of its fixation would equal its initial frequency when $V_m = 0$.

The total possible change in the frequency of A_1 in the limit is $- q_1$ whereas in a finite population the expected change in the limit is $u(q_1) - q_1$. This expected change would be negative. For a gene with frequency as half, $N^2 V_m = 1$ can bring down its expected frequency in the limit to about 0.30. It is apparent from figure 13 that a rare gene $(q_1 = 0.16667)$ can quickly be eliminated by multiple allelic variance but a frequent gene $(q_1 = 0.8333)$ can still have some chance of getting fixed in spite of the enormous effect of multiple alleles to knock it out.

The results on the chance of fixation of A_1 presented above can be compared with a general formula for the chance of fixation of an allele in a multiple allelic situation suggested by Robertson (personal communication). With respect to k alleles A_1, A_2, . . . A_k with respective initial frequencies q_1, q_2, . . . q_k and selective coefficients S_1, S_2, . . . S_k such that

$$(5.18) \quad \bar{S} = \sum_n q_n S_n$$

$$(5.19) \quad V_S = \sum_n q_n S_n^2 - \bar{S}^2$$

$$(5.20) \quad (\delta q)_n = q_n (S_n - \bar{S})$$

the chance of fixation u_n of the nth allele A_n is given

by

$$(5.21) \quad u_n = q_n + 2N (\delta q)_n + \frac{(2N)^2}{3} q_n [(s_n - \bar{s})^2 - V_S]$$

$$- \frac{(2N)^3}{3} (\delta q)_n V_S + - - - - -$$

For the case of three alleles the chance of fixation of allele A_1 with frequency q_1, which is neutral on an average, with respect to A_2 and A_3 with a variance V_m between them, this formula reduces to

$$(5.22) \quad u_1 (q_1) = q_1 - \frac{(2N)^2}{3} q_1 (1 - q_1) V_m$$

A comparison of $(u_1 (q_1) - q_1)$ calculated from this formula with that obtained from the computer is shown in Table 8 when $q_1 = q_2 = q_3 = \frac{1}{3}$. The agreement between the two results is quite close.

5.2   MEAN TIME UNTIL FIXATION OF ALLELE A_1

When no selective forces are operating, the mean time until fixation of A_1 is found to be independent of the frequency of alleles

TABLE 8   Comparative values for the expected change
in the limit in the frequency of allele A_1
having a mean superiority of $\bar{s} = o$ with
variance V_m in a population of size $N = 6$
for $q_1 = q_2 = q_3 = \frac{1}{3}$.

$N^2 V_m$	Computer	Formula	Difference
0.0000	0.00000	0.00000	0.00000
0.0144	-0.00376	-0.00427	0.00051
0.0576	-0.01470	-0.01707	0.00237
0.1296	-0.03183	-0.03840	0.00657
0.2304	-0.05369	-0.06827	0.01458
0.3600	-0.07863	-0.10667	0.02804

A_2 or A_3 and is the same as that given in Chapter IV for two alleles with additive gene action. At particular values of $Ns_1 = 2$ and $Ns_2 = 1$, the mean times are presented in Table 9 corresponding to the possible initial configurations of the three alleles in a population of size $N = 6$, when all the three alleles are present. For a fixed value of q_2, the mean times can be read vertically down for various initial frequencies of A_1. It decreases as the gene frequency increases. For instance, when A_2 allele is initially present with a frequency 0.5, the time to fixation of A_1 with frequency 1/12 is 3.26N whereas when its frequency is 5/12 it is 2.59N. However, the effect of the frequency of A_2 allele on the time to fixation of A_1 allele is the reverse. For a given frequency of A_1 allele, it takes longer for its fixation if the frequency of A_2 allele is frequent than when it is rare.

The effect of $N\bar{s}$ and N^2V_m on the time to fixation of A_1 is shown in Table 10, for the case when the initial frequency of A_1 allele is 1/3. When $N^2V_m = o$, the mean time decreases with increase in $N\bar{s}$. So is the case with $N^2V_m = 1$ but at higher values of N^2V_m it increases first with increase in $N\bar{s}$ and decreases thereafter. When $N\bar{s} = o$, the effect of multiple allelic variance is to decrease the mean time. So is the case at $N\bar{s} = 1$ but at higher values of $N\bar{s}$, the mean time increases first with increase in N^2V_m but decreases thereafter. It is interesting to note that the behaviour of the mean time over variations in $N\bar{s}$ for a fixed N^2V_m is almost the

TABLE 9   Average number of generations / N until fixation of A_1 (N = 6)

$Ns_1 = 2$, $Ns_2 = 1$.

$12q_2$ / $12q_1$	1	2	3	4	5	6	7	8	9	10
1	3.07	3.11	3.15	3.19	3.22	3.26	3.29	3.33	3.36	3.39
2	2.90	2.94	2.98	3.02	3.06	3.10	3.14	3.17	3.21	
3	2.72	2.76	2.81	2.85	2.90	2.94	2.98	3.02		
4	2.53	2.58	2.63	2.68	2.72	2.77	2.81			
5	2.33	2.39	2.44	2.49	2.54	2.59				
6	2.12	2.18	2.24	2.30	2.36					
7	1.90	1.97	2.03	2.10						
8	1.66	1.74	1.81							
9	1.40	1.48								
10	1.10									

TABLE 10   Average No. of generations until

fixation of A_1 $(N = 6)$ with

initial frequency $1/3$.

N^2V_m / $N\bar{s}$	0	1	4	9
0.0	3.00	2.81	2.45	2.11
1.0	2.83	2.75	2.51	2.23
2.0	2.48	2.55	2.54	2.40
3.0	2.12	2.28	2.44	2.48
4.0	1.84	2.00	2.24	2.42
5.0	1.61	1.76	2.01	2.26
6.0	1.44	1.57	1.79	2.06
7.0	1.31	1.42	1.61	1.86
8.0	1.20	1.30	1.46	1.68

same as that of the mean time over variations in N^2V_m for a fixed $N\bar{s}$.

5.3 MEAN TIME UNTIL HOMOZYGOSITY

The results of the mean time until homozygosity are presented in Table 11 and 12. In Table 11, the three alleles are neutral and the variation of the mean time until homozygosity over the different initial gene configurations is shown. At a fixed value of q_1, the mean time increases first with increase in the frequency of A_2; attains a maximum and decreases thereafter. Since the situation is symmetrical, the same is true for variations over the frequency of A_1 for a fixed q_2. Also the distribution is symmetrical so that the same mean time is obtained for frequencies less than or greater than the mid-point of the range. The highest mean time is $3N$ and occurs when $q_1 = q_2 = \frac{1}{3}$ whereas the lowest is $1.66N$ and occurs when either $q_1 = q_2 = \frac{1}{12}$ or $q_1 = \frac{1}{12}$, $q_2 = \frac{10}{12}$ or $q_1 = \frac{10}{12}$, $q_2 = \frac{1}{12}$. In Table 12, the variation in the mean time until homozygosity is shown over the different values of Ns_1 and Ns_2 when all the three alleles are at equal frequencies. There is a perfect symmetry as far as Ns_1 and Ns_2 are concerned. For a fixed Ns_2 (unless it is 0), the mean time first increases, attains a maximum and then decreases with increase in Ns_1. The maximum shifts to the right side as Ns_2 increases. For $Ns_2 = 0$, the maximum is $3N$ and occurs at $Ns_1 = 0$, the time for the neutral case. This can be compared with the maximum mean time until homozygosity of $2.55N$

TABLE 11   Average number of generations / N until homozygosity   (N = 6)
Three neutral alleles at a locus.

$12q_2$ $12q_1$	1	2	3	4	5	6	7	8	9	10
1	1.66	2.11	2.43	2.63	2.73	2.73	2.63	2.43	2.11	1.66
2	2.11	2.46	2.70	2.84	2.89	2.84	2.70	2.46	2.11	
3	2.43	2.70	2.88	2.96	2.96	2.88	2.70	2.43		
4	2.63	2.84	2.96	3.00	2.96	2.84	2.63			
5	2.73	2.89	2.96	2.96	2.89	2.73				
6	2.73	2.84	2.88	2.84	2.73					
7	2.63	2.70	2.70	2.63						
8	2.43	2.46	2.43							
9	2.11	2.11								
10	1.66									

Table 12   Average number of generations / N until homozygosity (N = 6)

$$q_1 = \frac{1}{3}, \qquad q_2 = \frac{1}{3}, \qquad q_3 = \frac{1}{3}$$

Ns_2 \ Ns_1	0	1	2	3	4	5	6	7	8
0	3.00	2.87	2.55	2.20	1.90	1.66	1.48	1.33	1.22
1	2.87	2.90	2.71	2.41	2.10	1.83	1.62	1.45	1.31
2	2.55	2.71	2.73	2.57	2.31	2.03	1.79	1.60	1.44
3	2.20	2.41	2.57	2.61	2.48	2.26	2.01	1.79	1.60
4	1.90	2.10	2.31	2.48	2.53	2.44	2.24	2.01	1.80
5	1.66	1.83	2.03	2.26	2.44	2.50	2.42	2.24	2.03
6	1.48	1.62	1.79	2.01	2.24	2.42	2.48	2.41	2.25
7	1.33	1.45	1.60	1.79	2.01	2.24	2.41	2.47	2.41
8	1.22	1.31	1.44	1.60	1.80	2.03	2.25	2.41	2.47

obtained in the two allele case in Chapter IV. It shows that the maximum time to homozygosity is increased with the increase in the number of alleles and that it occurs when all the alleles are equally represented in the initial population.

CHAPTER VI

LETHAL GENES AND ALLELISM IN FINITE POPULATIONS

In Chapter II, it has been shown that the properties of the
stationary distribution of gene frequencies realised due to the
balance between the mutation occurring at a low rate in one direction
(reverse mutation being negligible) and selection with random drift
can be studied with the help of the element of the first row of
fundamental matrix. In this Chapter, this theory is applied to study
the population dynamics of completely lethal genes in finite popu-
lations. It will further be shown how this theory can predict allelic
rates and the decline of allelism with time. The determination of
allelic rates is of great importance in experimental investigations with
natural populations to decide between whether recessive lethals have
deleterious effect as heterozygotes. This problem has been discussed at
length by Crow and Temin (1964) and Wallace (1966). If the heterozygotes
carrying lethals have reduced fitness, then the mutation from normal to
lethal allele is more than sufficient to balance their elimination by
homozygosity and the population suffers from a mutational load. On the
other hand, if the heterozygotes have enhanced fitness, then the mutation
is not sufficient to balance the elimination of lethal genes, and the
load is said to be segregational (balanced). Most of the earlier
experimental investigations on natural populationswith _Drosophila_ like
those of Dobzhansky and Wright (1941), Wright, Dobzhansky and Hovanitz
(1942) and Crow and Temin (1964) give evidence in support of the

deleterious effect of lethal genes in the heterozygous condition but Wallace (1966) obtains results which show an overdominant effect of lethal genes. Nei (1968), however, argues that Wallace's conclusion could be due to the small effective size of the population sampled and suggests that the lethal genes are, on the average, slightly deleterious in the heterozygous condition. The decline of allelism of recessive lethals with time is another interesting problem invoked by Wallace (1966). His conjecture, on the functional relationship between the allelic rate and the time interval between the sampling of the first and the second set of lethals has been proved as approximately correct by Prout (1967). An alternative method, however, based on the transition matrix approach is developed in this section on the basis of suggestions made by Robertson (personal communication).

6.1   DISTRIBUTION OF LETHAL GENES

Consider a randomly mating population of N adult individuals with i heterozygotes and $(N - i)$ homozygotes. This would mean that the gene frequency of lethal a is $\frac{1}{2}q_i = i/(2N)$ and that of normal A is $(1 - \frac{1}{2}q_i)$. Let the fitnesses of the three possible genotypes AA, Aa and aa be 1, $1 - h$ and 0 respectively. Then after selection, the frequency of heterozygotes will be

$$(6.1) \quad q_i' \quad = \quad \frac{q_i \ (1 - h)}{1 + \frac{1}{2}q_i \ (1 - 2h)}$$

whereas that of homozygotes will be $(1 - q_i')$. On sampling N adult individuals from such a population will produce j heterozygotes and $(N - j)$ homozygotes with transition probability

$$(6.2) \quad P_{ij} \quad = \quad \binom{N}{j} (q_i')^j \ (1 - q_i')^{(N - j)}$$

Here i goes from 0 to N and j also goes from 0 to N. The Markov Chain specified by such probabilities has 0 as the only absorbing state. These probabilities determine the Q matrix where i and j each go from 1 to N. If the mutation rate from A to a is u per generation, the reverse mutation being negligible, then as shown in Chapter II the stable distribution of the lethal gene frequency is given by $2Nu \ S_{1j}$ where S_{1j} is the jth element of the first row of $(I - Q)^{-1}$. The mean and the variance of the frequency of a and the mean frequency of heterozygotes in the stable state are given by

$$(6.3) \quad E(q) \quad = \quad 2Nu \sum_{j=1}^{N} S_{1j} \ (j/(2N))$$

$$(6.4) \quad E[q - E(q)]^2 = 2Nu \sum_{j=1}^{N} S_{1j} (j/(2N))^2 - [E(q)]^2$$

$$(6.5) \quad E[2q(1-q)] = 4Nu \sum_{j=1}^{N} S_{1j} (j/(2N)) (1 - j/(2N))$$

These quantities have been evaluated on the computer for population size $N = 50$ and 10. The results are shown in figures 14, 15 and 16. A negative value of h means heterotic lethals whereas a positive value means partially recessive lethals. $h = 0$ corresponds to completely recessive lethals. A mutation rate of 10^{-5} has been assumed throughout.

6.11 MEAN GENE FREQUENCY

From figure 14, it is found that the mean gene frequency declines almost linearly for small population sizes such as 10 as we pass from heterotic lethals through completely recessive lethals to partially recessive lethals. But for large population sizes such as 50, the decline is non-linear.

The mean gene frequency for completely recessive lethals is known from Wright's equilibrium distribution formula (Wright, 1937). The distribution for small q is given by
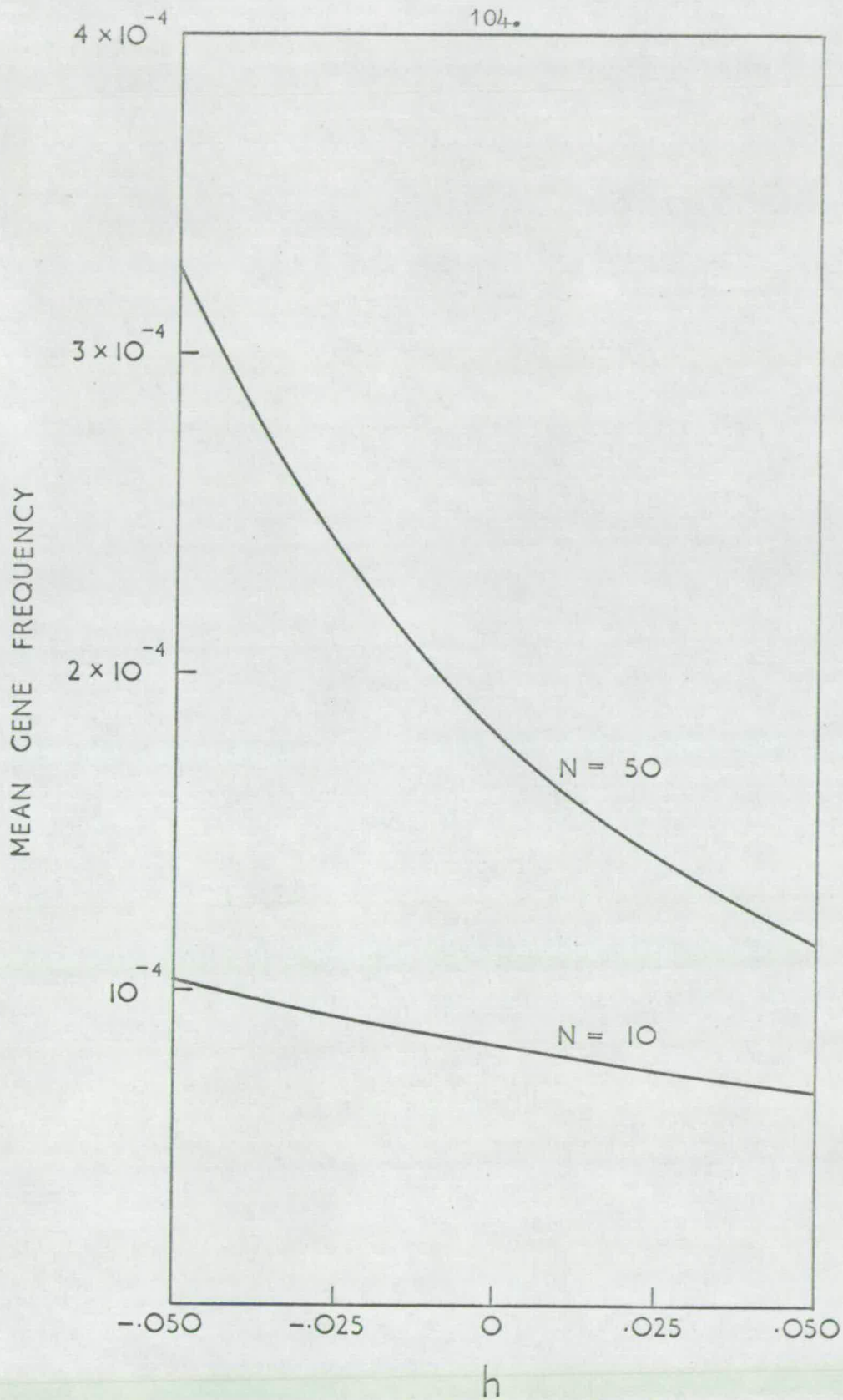
Fig. 14   Mean of the distribution of
            frequencies of a lethal gene
            (u = 0.00001). The curves are drawn
            for different population sizes.

$$(6.6) \quad \phi(q) = \text{const. exp} \left(- 2Nq^2\right) q^{(4Nu - 1)}$$

Substituting $t = q^2$ and applying the condition

$$(6.7) \quad \int_0^1 \phi(q) \, dq = 1$$

gives

$$(6.8) \quad \text{const.} = \frac{2}{\Gamma(2Nu)} \, (2N)^{2Nu}$$

The mean and variance of the gene frequency are given by

$$(6.9) \quad \bar{q} = \Gamma(2Nu + \tfrac{1}{2}) \, / \, (2N)^{\tfrac{1}{2}} \, \Gamma(2Nu)$$

$$(6.10) \quad \sigma_q^2 = u - \bar{q}^2$$

For $2Nu > 1$, $\bar{q}$ is close to $u^{\tfrac{1}{2}}$ i.e. the equilibrium value expected in an infinite population. For $2Nu < 1$, however,

(6.11) $\quad \Gamma(2Nu) = \dfrac{1}{2Nu}$

(6.12) $\quad \Gamma(2Nu + \frac{1}{2}) = \pi^{\frac{1}{2}}$

approximately, so that

(6.13) $\quad \bar{q} = u(2 \pi N)^{\frac{1}{2}}$

The conditions imposed in the matrix approach require that $2Nu < 1$ and therefore the mean gene frequency obtained by this approach should be compared with $u(2 \pi N)^{\frac{1}{2}}$. This comparison is shown in Table 13.

Table 13   Comparison of mean lethal frequency as
calculated from Wright's formula and
transition matrix method for different
population sizes $(h = 0,\ u = 10^{-5})$.

N	Wright's formula	Computer results
50	17.72×10^{-5}	18.18×10^{-5}
10	7.93×10^{-5}	8.39×10^{-5}

The agreement between the two approaches is quite close, Wright's formula giving a lower value than the computer results.

For heterotic lethals, the equilibrium gene frequency by Wright's formula is known in only special cases. The equilibrium frequency q expected in an infinite population due to the superiority of hetero- zygotes only is, however, $k/(1 + k)$ where $k = -h$ which is approximately $-h$ for small values of h. The matrix approach, therefore, gives the result that in the range $q = 0$ to $q = .05$, the mean gene frequency decreases linearly as q approaches 0 if $N = 10$ but decreases very sharply if $N = 50$. When $k = +.05$, the mean gene frequency is 3.28×10^{-4} for $N = 50$ but only 1.4×10^{-4} for $N = 10$. This shows that, in order that the equilibrium frequency of 0.05 is actually realized the population size has to be very very large. The gene frequency increases as the population size increases.

6.12 VARIANCE OF GENE FREQUENCY

The variances of gene frequency for population size $N = 10$ and 50 are shown in figure 15. The variance decreases as the reduction in the fitness of the heterozygotes rises from -0.05 through 0 to $+0.05$ for either of the two population sizes. But unlike the curves for the mean gene frequency, the curves here intersect. In most cases of the heterotic situation, the variance is more for the larger population size but for completely recessive lethals and partially recessive lethals, the variance is smaller for the larger population size. For $h = 0$ and $N = 50$, the variance is found to be 1.08×10^{-5}
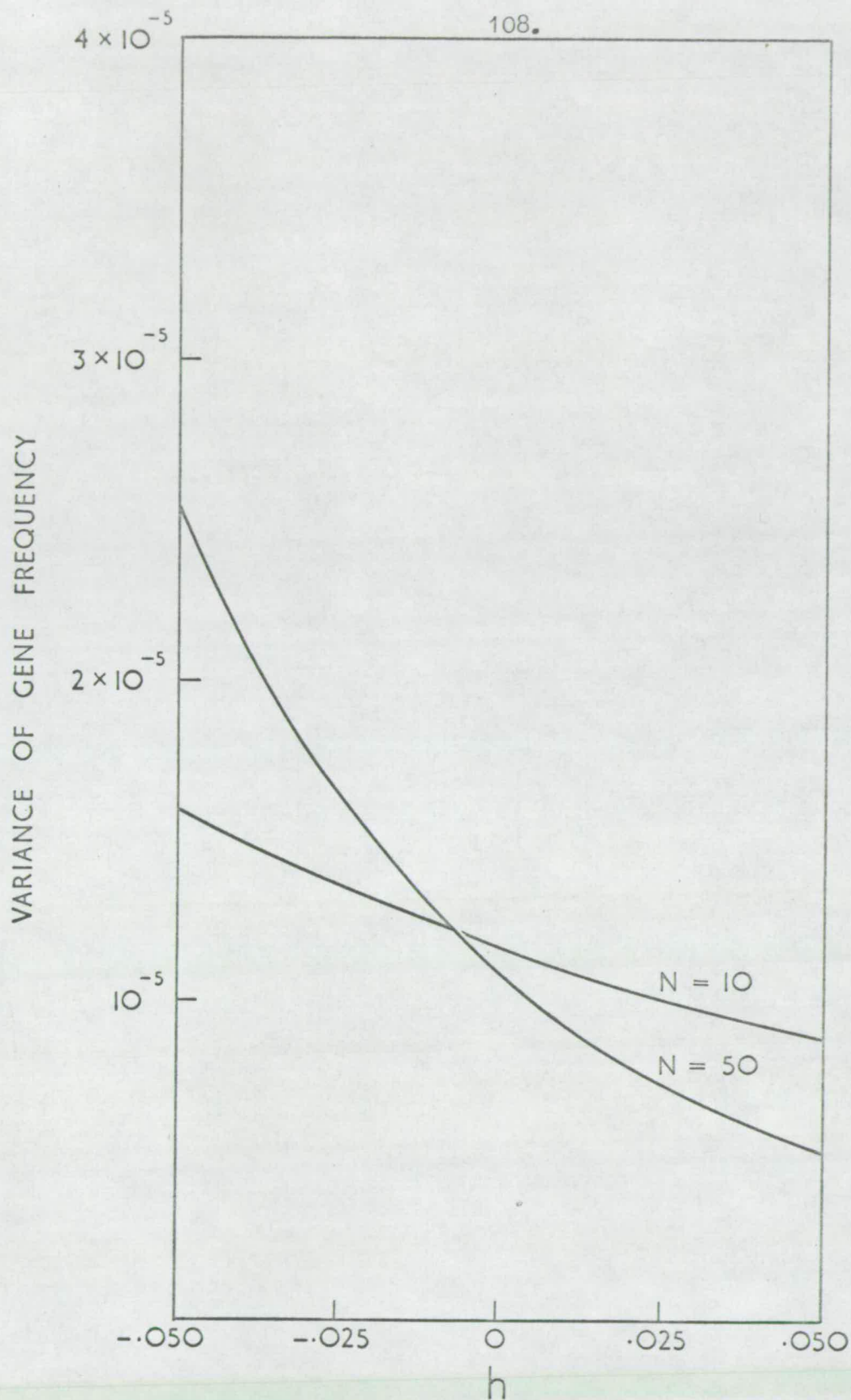
Fig. 15   Variance of the distribution of
          frequencies of a lethal gene (u = 0.00001).
          The curves are drawn for different pop-
          ulation sizes.

which compares very well with the value of 0.99686×10^{-5} obtained from Wright's variance formula given by (6.10).

6.13 MEAN HETEROZYGOSITY AT EQUILIBRIUM

The curves for the mean heterozygosity for $N = 50$ and $N = 10$ as presented in figure 16 show almost the same pattern as that of the mean gene frequency. For overdominant genes, this similarity between the mean gene frequency and average heterozygosity was also noted by Robertson (1962). For $h = 0$ and lethal gene, his formula for heterozygosity reduces to $2u(2\pi N)^{\frac{1}{2}}$ which is twice the mean gene frequency. It can be verified from figures 14 and 16 that for $h = 0$, the mean heterozygosity is almost double that of the mean gene frequency. As noted in Chapter II, the mean heterozygosity is proportional to the contribution of the locus towards the genetic variance of the stable population. If Nu is small as would usually be the case with $u = 10^{-5}$ unless N is very large, the genetic variance would increase as the size of the population is increased.

6.14 PROPORTION OF POPULATIONS SEGREGATING FOR A GIVEN LETHAL

Following Wright (1931) the proportion $F(o)$ of populations from which lethals are absent is such that the frequency $2Nu\,F(o)$ of recurrence in a population equals the frequency $(\frac{1}{2})\,F(\frac{1}{2}N)$ of loss where

110.

MEAN HETEROZYGOSITY AT EQUILIBRIUM

100×10^{-5}

75×10^{-5}

50×10^{-5}

25×10^{-5}

N = 50

N = 10
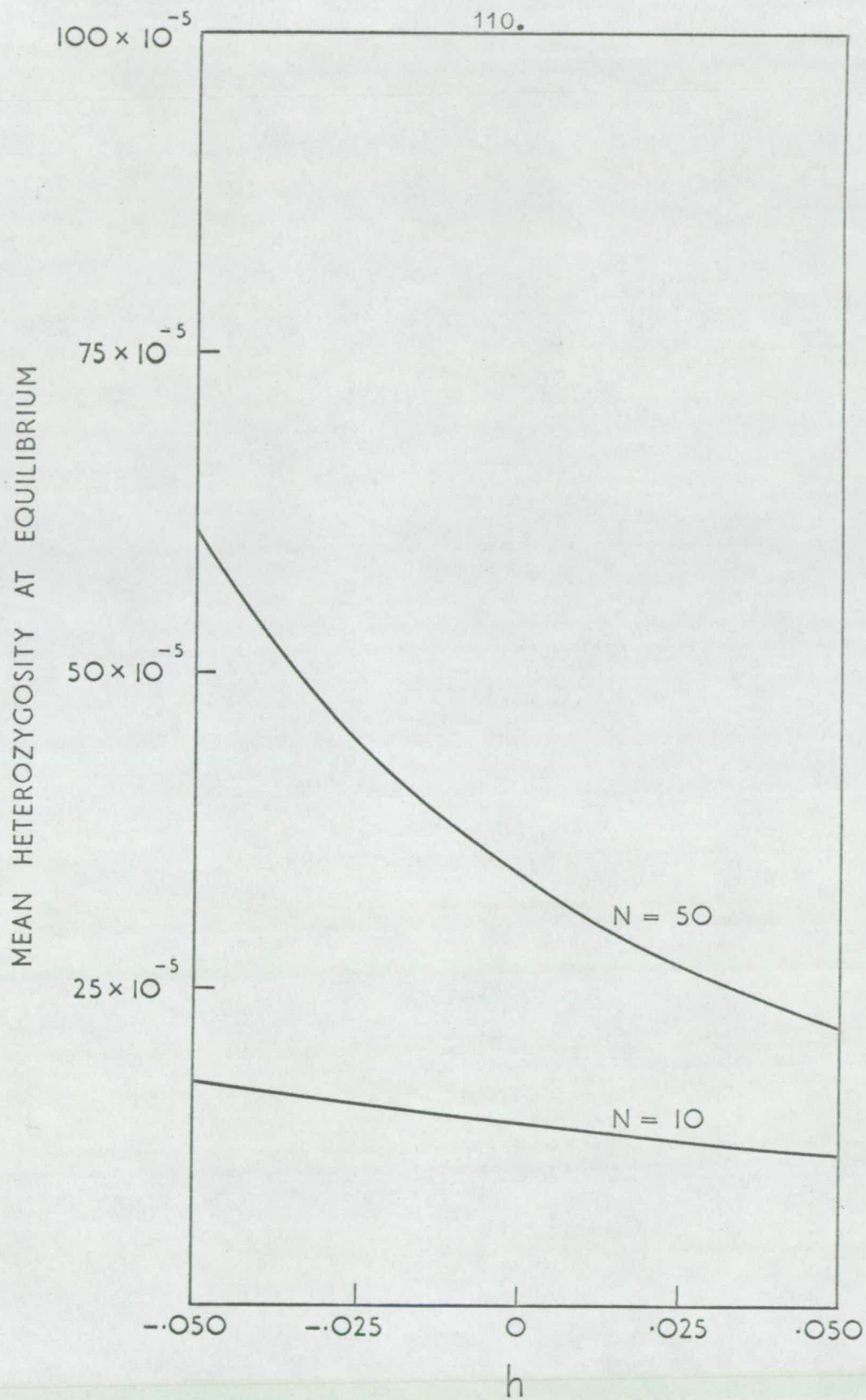
−·050   −·025   0   ·025   ·050

h

Fig. 16 Mean heterozygosity at equilibrium
(u = 0.00001). The curves are drawn
for different population sizes.

111.

$$(6.14) \quad F(q) = \frac{1}{2N} \phi(q)$$

and $\phi(q)$ is given by (6.6). This gives

$$(6.15) \quad F(o) = (2N)^{-2Nu}$$
$$= 1 - 2Nu \log_e 2N \quad \text{approximately}$$

In the matrix approach, however, if $f(o)$ represents the mean number of loci having no mutant at the equilibrium state when, on an average one locus becomes homozygous per generation, in the absence of any mutation, then

$$(6.16) \quad 2Nu \, f(o) + 2Nu \, S = 1$$

where

$$(6.17) \quad S = \sum_j S_{1j}$$

This gives the value of $F(o)$ as

$$(6.18) \quad F(o) = 2Nu \, f(o)$$
$$= 1 - 2Nu \, S$$

In other words, the proportion of populations segregating for lethals is $2Nu\,S$. For a completely recessive lethal with $N = 50$, the proportion of populations from which lethals are absent is, using (6.18), found to be 0.994360 as against the value of 0.995395 obtained by Wright's approach [using (6.15)].

6.2   CHANCE OF ALLELISM

In order to study the allelism in a given sample of lethal bearing chromosomes, all the possible heterozygotes between these lethal bearing chromosomes are made. If the same lethal gene is carried on two or more chromosomes in the sample, then a lethal zygote will be formed everytime these chromosomes meet. If a lethal gene is represented only once in the sample, then the chromosome carrying it will form viable heterozygotes with all other lethal chromosomes in the sample. By making all possible crosses among strains with lethal chromosomes, the number of times various lethal genes are represented as well as the total frequency of lethal zygotes formed can be determined. The chance of allelism of lethal chromosomes is then regarded as the proportion of crosses between lethal bearing chromosomes which produce lethal zygotes. When two lethal chromosomes prove to be allelic, the lethal genes they carry are either identical by descent or are different mutations which by chance have affected a finite number of equivalent loci. The total chance of allelism is the sum of the chances of allelism due to these

two causes. If the population size is infinite, the allelism is due to recurrent mutations only whereas if the number of equivalent loci is infinite the allelism is due to identity by descent only.

Apart from the allelism of lethal bearing chromosomes i_c, there is also the allelism of random pairs of lethal genes i. Knowing i_c from experimental data, i can be evaluated by the formula (Nei 1968)

$$(6.19) \quad i = \frac{-\text{Log}_e \ (1 - i_c \ q^{*2})}{[\text{Log}_e \ (1 - q^*)]^2}$$

where q^* is the frequency of lethal chromosomes. In what follows, however, we discuss i, the chance of allelism of random pairs of lethal genes with the help of the properties of lethal gene frequency distribution discussed in the previous section.

If lethal mutations occur at n out of a large number of loci and the proportion of lethals at a locus to that of all lethals is p, then the chance that a second lethal occurs at the same locus as the first is p^2. Averaging over the n loci gives the total probability as $\sum_n p^2$ that the two lethals taken at random from the population will be allelic. Since $\sum_n p = 1$ and p is proportional to the frequency q of lethal gene, the chance of allelism is $(\sum q^2) / (\sum q)^2$. Following Nei (1968) we take the expectations of the numerator and the denominator and obtain $n \ E(q^2)$ and $n\sigma_q^2 + n^2 \ \bar{q}^2$ respectively. This gives the chance of allelism i as

$$(6.20) \quad i = \frac{x}{(1 - \frac{1}{n}) + x}$$

where

$$(6.21) \quad x = \frac{E(q^2)}{n\bar{q}^2}$$

$$= \frac{(\sigma_q^2 + \bar{q}^2)}{n\bar{q}^2}$$

is the formula for the chance of allelism given in Dobzhansky and Wright (1941), and Wright, Dobzhansky and Hovanitz (1942). In the present case it is assumed that the mutation rates are the same for all loci so that there is no contribution to the variance of the gene frequency due to differences in the mutation rates. When the population is infinite, the variance in the gene frequency due to random drift vanishes and i reduces to $\frac{1}{n}$. But when an infinite number of equivalent loci are affected by recurrent mutation, $\frac{1}{n}$ tends to zero and i reduces to i_F given by

$$(6.22) \quad i_F = \frac{x}{1 + x}$$

This is the chance of allelism due to finite size of the population only and has been investigated in this section by the transition matrix approach. Using (6.3) and (6.4) for $E(q)$ and $E(q^2)$ we get the value of x as

$$(6.23) \quad x = \frac{2Nu \sum_{j} S_{1j} \left(\frac{j}{2N}\right)^2}{(2Nu)^2 n \left[\sum S_{1j} \frac{j}{2N}\right]^2}$$

$$= \frac{\sum_{j} S_{1j} \left(\frac{j}{2N}\right)^2}{2NU \left[\sum S_{1j} \frac{j}{2N}\right]^2}$$

where

$$(6.24) \quad U = nu$$

is the mutation rate for lethal chromosomes. It refers to a very large number of loci with very small mutation rate per locus such that the product of the two is constant. Taking $n = 500$ and $u = 10^{-5}$ so that the mutation rate for lethal chromosome is 0.005, the computer results on the allelic rates are shown in figure 17. It is apparent that the slope of the curve is more when $N = 50$ than when $N = 10$. This shows that for very small population sizes, the
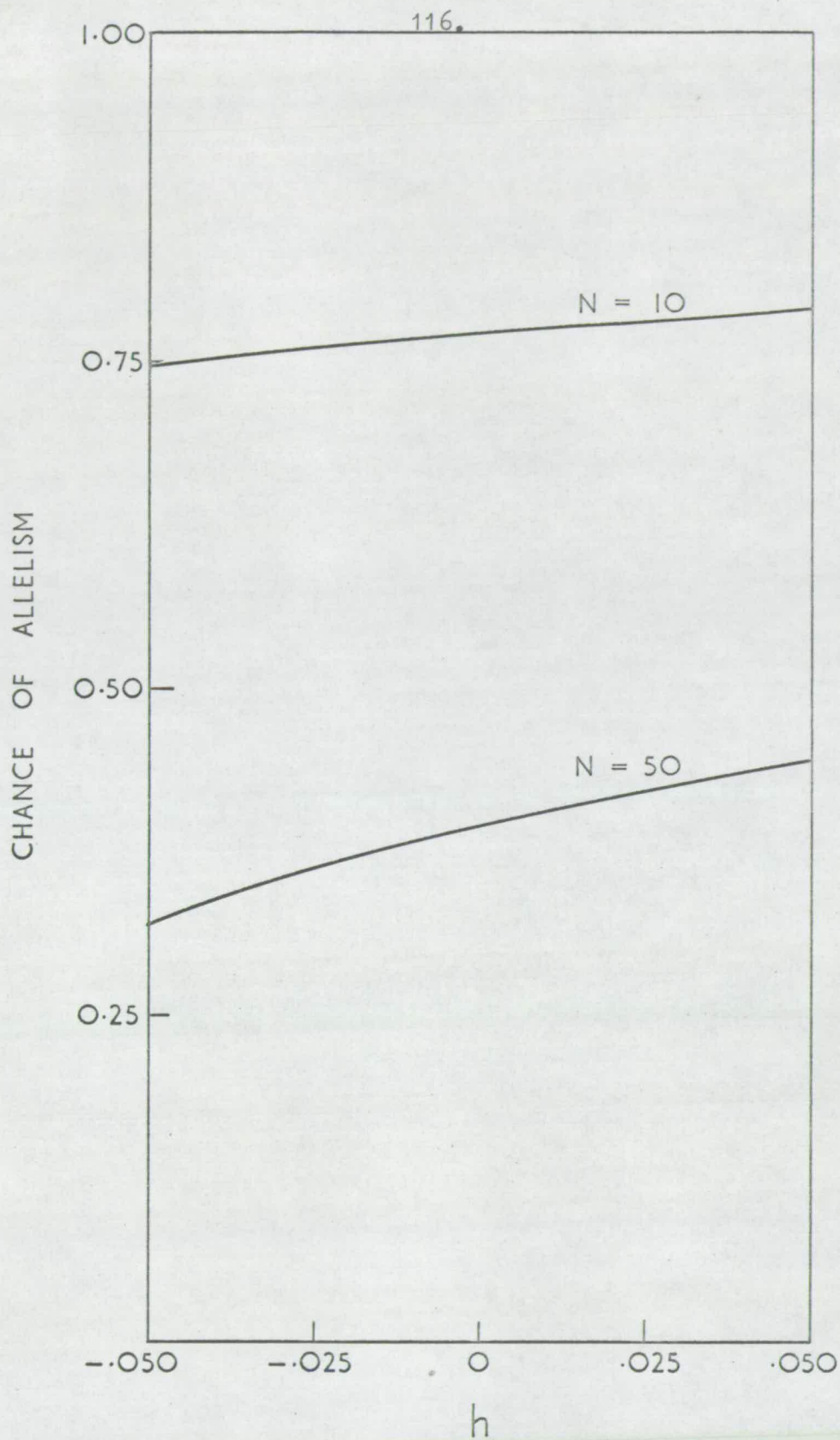
Fig. 17   The chance of allelism of lethal
genes (U = 0.005). The curves are
drawn for different population sizes.

allelic rates may practically be independent of the nature of lethal
genes. But for appreciable population size, the allelic rates are
lower for heterotic lethals than for partially recessive lethals.
For very large population sizes, these rates are solely determined
by the nature of the loci and can be very small for heterotic loci.
As the population size decreases the rates increase. Similar results
were obtained by Nei (1968), using the properties of Wright's stable
distribution of gene frequencies. For completely recessive lethals
($h = 0$) and $N = 10$, the chance of allelism is 77% as against Nei's
result of 76.2% whereas for $N = 50$, the values obtained by the
matrix approach and Wright's formula are respectively 0.40 and 0.39.

6.3 DECLINE OF ALLELISM WITH TIME

Wallace (1966) presented an analysis of the allelism of recessive
lethal genes in certain populations of Drosophila melanogaster. His
main concern was the determination of the allelic rates of lethals
taken from a natural population at collection sites of varying distance
apart. He argued that the allelism between two sets of lethals from a
given distance apart can be related to the allelism of two sets of
lethals taken from the same point in the population but separated by a
given interval in time. The formula suggested by him for the allelic
rate i_{T_t} and the time interval (t) between the sampling of the

first and the second set of lethals have the relationship

$$(6.25) \quad i_{T_t} = i_n + i_F (1 - K)^t$$

where

i_{T_t} = the allelism between lethal genes t generations apart.

i_F = the allelism component of i_{T_t} within a generation produced by identity by descent in a finite population.

i_n = the allelism within a generation of the same set of lethal loci if they were in a population of infinite size.

K = a constant related to the rate of turnover of lethals within the population.

Prout (1967) demonstrated with algebraic details that this functional relationship is approximately correct.

If the lethal gene frequency at time t is denoted by q_t and if we assume that the mean gene frequency $\bar{q}$ does not change appreciably

during the time interval t, then the probability of choosing one and the same lethal from among lethals of generation 0 and from among lethals of generation t is $\Sigma qq_t/(n\bar{q})^2$. The expectation of this ratio can again, be expressed as

$$(6.26) \quad i_{T_t} = \frac{y_t}{1 + y_t}$$

where

$$(6.27) \quad y_t = \frac{nE(qq_t)}{(n\bar{q})^2}$$

We already know that $\bar{q} = \overset{2Nu}{\underset{\wedge\,j}{\Sigma}} S_{1j} \left(\frac{j}{2N}\right)$. The expected value of qq_t has further to be evaluated by the matrix approach. The stable gene frequency distribution is given by the elements $2Nu\, S_{1j}$, $j = 1, 2, \ldots \ldots N$, representing the mean number of loci having j mutants. The weighted/gene frequency vector at time 0 is then given by

$$(6.28) \quad q = 2Nu\, [S_{11} \tfrac{1}{2N}, \ldots \ldots S_{1j} \tfrac{j}{2N}, \ldots \ldots S_{1N} \tfrac{1}{2}]$$

For a given number of j mutants occurring initially, the conditional mean gene frequency column vector at time t is given by

$$(6.29) \quad q_t = \begin{bmatrix} \sum_k P_{1k}(t) \dfrac{k}{2N} \\ \\ \\ \sum_k P_{jk}(t) \dfrac{k}{2N} \\ \\ \\ \sum_k P_{Nk}(t) \dfrac{k}{2N} \end{bmatrix}$$

where $P_{jk}(t)$ are the elements of the matrix $\underset{\sim}{Q}^t$. The expectation of qq_t is then given by

$$(6.30) \quad E(qq_t) = 2Nu \sum_j \left[\left(S_{1j} \dfrac{j}{2N} \right) \left(\sum_k P_{jk} \dfrac{j}{2N} \right) \right]$$

Hence y_t is given by

$$(6.31) \quad y_t = \frac{2Nu \; n \; \sum\limits_{j} [(S_{1j} \frac{j}{2N}) (\sum\limits_{k} P_{jk}^{(t)} \frac{k}{2N})]}{(2Nu)^2 \; n^2 \; (\sum\limits_{j} S_{1j} \frac{j}{2N})^2}$$

$$= \frac{\sum\limits_{j} [(S_{1j} \frac{j}{2N}) (\sum\limits_{k} P_{jk}^{(t)} \frac{k}{2N})]}{2NU \; (\sum\limits_{j} S_{1j} \frac{j}{2N})^2}$$

Using this formula the computer results were obtained for i_{T_t} for two population sizes $N = 10$ and $N = 50$ and $h = -.05$, 0 and $+.05$ taking U to be 0.005. In each case t was varied from 0 to 28. The decline of allelism with time, thus obtained, is shown in figures 18, 19 and 20 respectively for $h = 0$, $+ 0.05$ and $- 0.05$. In each case the decline is sharp for $N = 10$ as compared to $N = 50$ and for about first twenty generations the allelism is higher for the smaller population size. The difference declines with time and beyond about twenty generations the allelism for $N = 10$ is lower than that for $N = 50$. These graphs can be compared with the sketch of hypo- thetical data given in Prout (1967). The graphs for $N = 50$ show the same pattern of exponential decline but for $N = 10$ the pattern of decline is non-exponential in the first several generations. From figure 20, it appears that for heterotic lethals $(h = - 0.05)$ the decline of allelism with time is not so rapid. This would be so
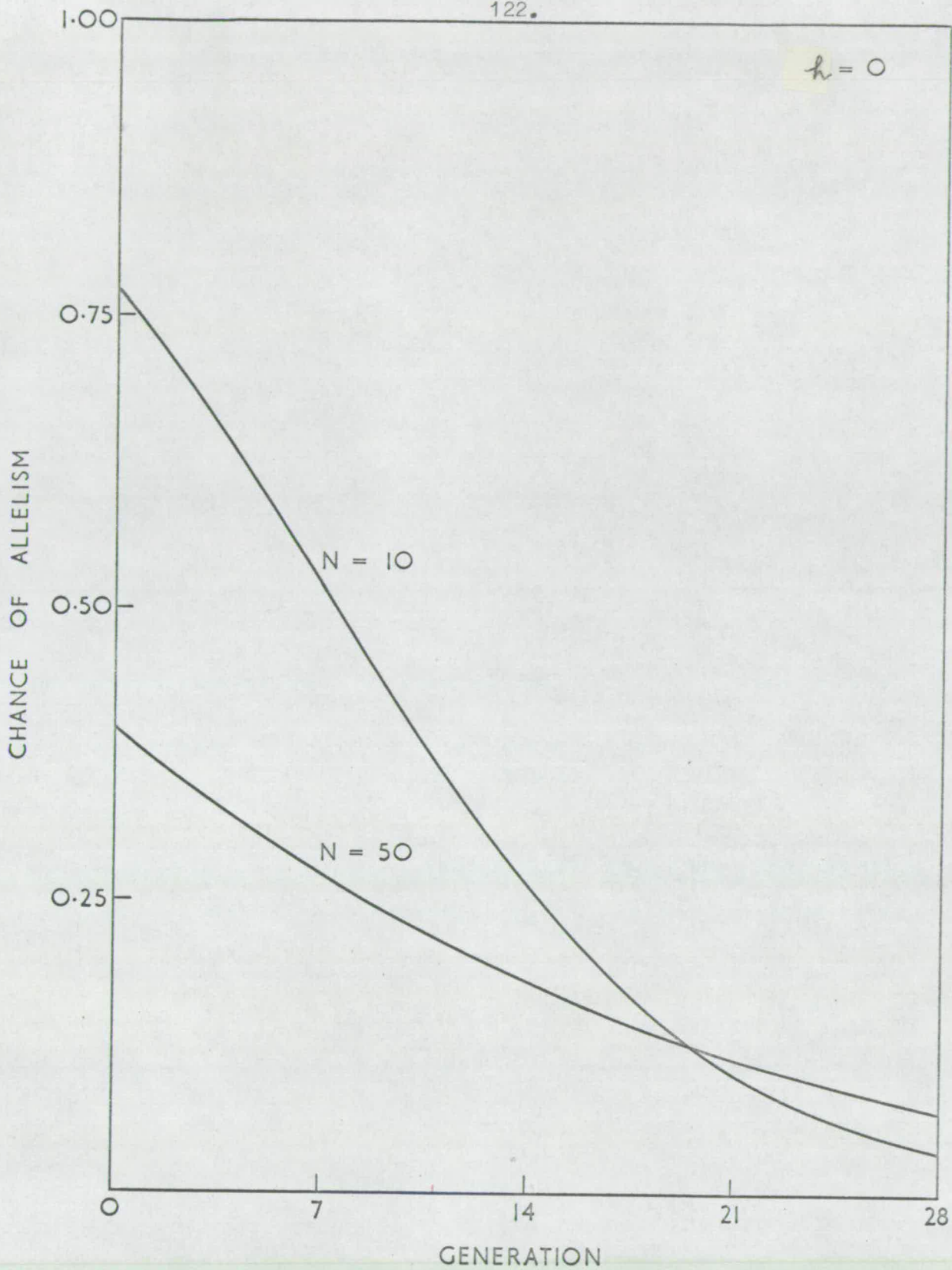
Fig. 18   Allelism of completely recessive lethal genes
           (U = 0.005).   The curves are drawn for different pop-
           ulation sizes.
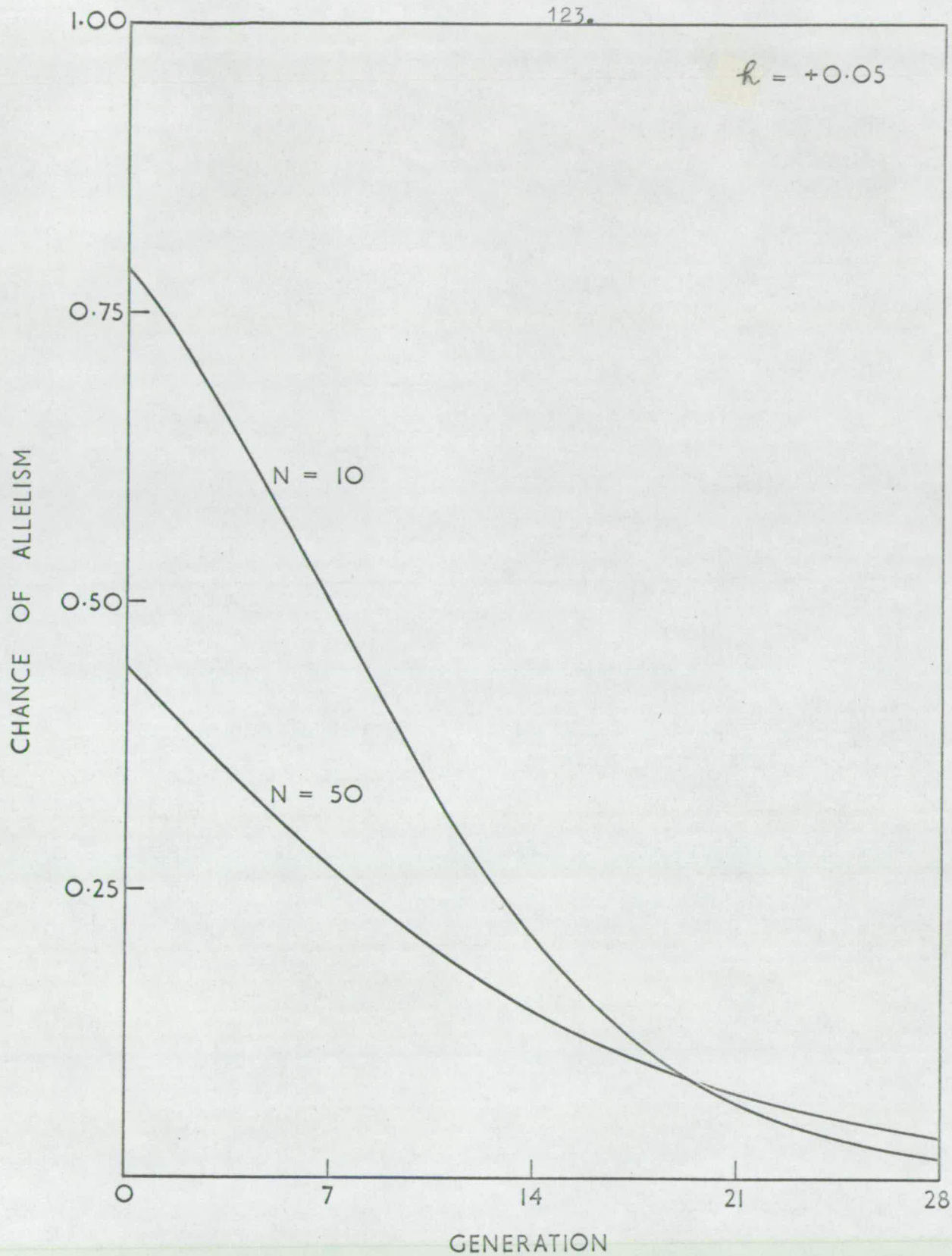
Fig. 19   Allelism of partially recessive lethal genes (U = 0.005).
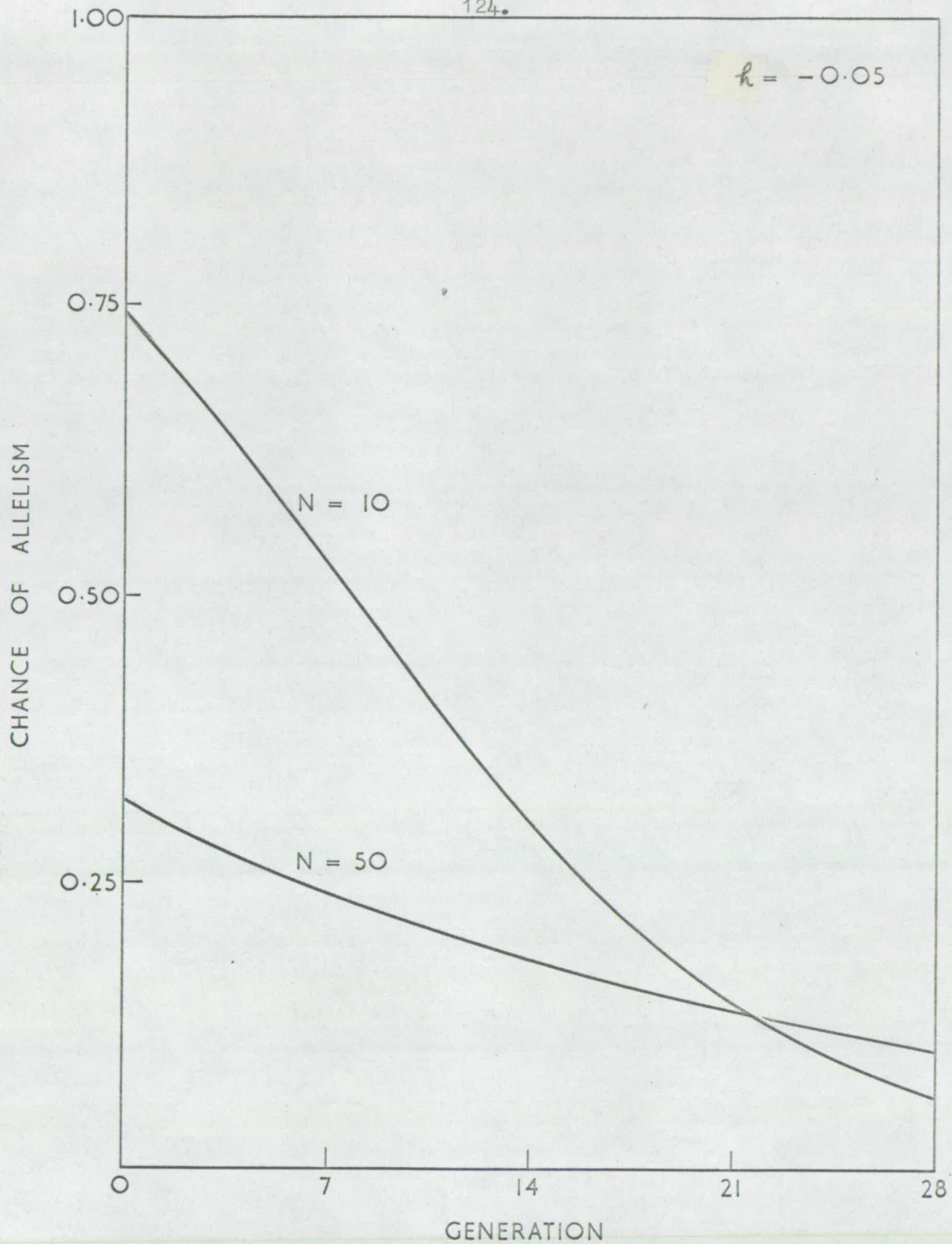The curves are drawn for different population sizes.

Fig. 20   Allelism of heterotic lethals (U = 0.005).   The curves
are drawn for different population sizes.

because heterotic lethals are likely to persist longer in the population. The limiting value of the allelic rate (i_T) is likely to be $\frac{1}{500} = 0.002$ since the chromosomal mutation rate is based on 500 loci. It is likely to occur well beyond 28 generations.

For completely recessive lethals $(h = 0)$ an expression of initial rate of decline can, however, be obtained using the equilibrium distribution of lethal genes given by (6.6). Let q_1 be the gene frequency of lethals in the next generation and let δq be the change in the gene frequency in one generation, then δq is approximately $- q^2$ so that

$$(6.32) \quad E(qq_1) = E[q(q + q)] = E(q^2) - E(q^3)$$

From (6.10), $E(q^2)$ is equal to u, $E(q^3)$ is given by

$$(6.33) \quad E(q^3) = \int_0^1 q^3 \phi(q)\, dq$$

$$= \frac{2(2N)^{2Nu}}{\Gamma(2Nu)} \int_0^1 q^3 \exp(-2Nq^2) q^{(4Nu - 1)}\, dq$$

$$= \frac{(2N)^{-\frac{3}{2}}}{\Gamma(2Nu)} \int_0^\infty t^{(2Nu + \frac{1}{2})} \exp(-t)\, dt$$

$$= \frac{(2N)^{-\frac{3}{2}} \, \Gamma(2Nu + \frac{3}{2})}{\Gamma(2Nu)}$$

$$= (2N)^{-\frac{3}{2}} (2Nu + \tfrac{1}{2}) \, \frac{\Gamma(2Nu + \tfrac{1}{2})}{\Gamma(2Nu)}$$

Using (6.11) and (6.12) we get, approximately

$$(6.34) \quad E(q^3) = u(4Nu + 1) \, (\frac{\pi}{8N})^{\frac{1}{2}}$$

$$\doteq u \, (\frac{\pi}{8N})^{\frac{1}{2}}$$

if u is so small that the term involving u^2 can be neglected.
This gives

$$(6.35) \quad y_1 = \frac{n \, E(qq_1)}{(n\bar{q})^2}$$

$$= \frac{1 - (\frac{\pi}{8N})^{\frac{1}{2}}}{2\pi Nu}$$

and

$$(6.36) \quad y_o = \frac{n \, E(q^2)}{(n\bar{q})^2}$$

$$= \frac{1}{2\pi NU}$$

Hence i_{T_1} and i_{T_o} are given by

$$(6.37) \quad i_{T_1} = \frac{[1 - (\frac{\pi}{8N})^{\frac{1}{2}}]}{[1 - (\frac{\pi}{8N})^{\frac{1}{2}} + 2\pi NU]}$$

$$(6.38) \quad i_{T_o} = \frac{1}{1 + 2\pi NU}$$

Both i_{T_o} and i_{T_1} show that the allelic rates will decrease if the chromosomal mutation rate increases. The initial rate of decline $r*$ is then given by

$$(6.39) \quad r^* = 1 - \frac{{}_i T_1}{{}_i T_0}$$

$$= \frac{2\pi NU \left(\frac{\pi}{8N}\right)^{\frac{1}{2}}}{\left[1 - \left(\frac{\pi}{8N}\right)^{\frac{1}{2}} + 2\pi NU\right]}$$

A comparison of the initial rates of decline as predicted from (6.39) and as obtained numerically (figure 18) for completley recessive lethals is shown in Table 14.

TABLE 14   Comparison of initial rate of decline as predicted from Wright's distribution formula and as obtained by the transition matrix method for different population sizes (h = 0, U = 0.005)

N	Predicted	Computer	Difference
50	0.05606	0.04917	0.00689
10	0.05576	0.04047	0.01529

The difference between the two sets of results is less marked when N = 50 than when N = 10.

6.4 EXPECTATION OF LIFE AND AVERAGE AGE OF A LETHAL MUTANT

6.41 EXPECTATION OF LIFE OF A LETHAL MUTANT

Consider first the case when there is initially one lethal mutant a and $(2N - 1)$ normal alleles A in a population of size $2N$. If there is no further mutation from normal to lethal alleles, the lethal mutant will ultimately be lost from the population. As shown in Chapter II, the mean total number of generations required for the elimination of the lethal is given by

$$(6.40) \qquad S_1^* = \sum_j S_{1j}$$

where S_{1j} represents the mean number of generations which the population spends in the state E_j characterized by j mutants and $(2N - j)$ normals on the way to homozygosity. S_1^* can be thought of as the expectation of life of the new mutant at its initial occurrence.

If the lethal mutant a occurs i times and normal allele A occurs $(2N - i)$ times initially in a population of size $2N$, the mean total number of generations required for the elimination of the i lethals is given by

$$(6.41) \quad S_i^* = \sum_j S_{ij}$$

$$= \underline{f'_i} (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \underline{e}$$

where S_{ij} represents the mean number of generations which the population spends in the state E_j before becoming homozygous and $\underline{f'_i}$ is a row vector with unity in the ith place and zeros elsewhere.

For population size as $N = 8$, the computer results for the average number of generations until loss of lethals for various initial gene frequencies of mutants and $h = -0.5$, 0 and $+0.5$ are shown in Table 15. It may be noted that the initial gene frequency of mutant can at most be $\frac{1}{2}$ corresponding to a population containing only heterozygotes.

TABLE 15   Average number of generations until loss of lethals in a population of size $N = 8$.

Initial frequency	h = - 0.5	h = 0	h = +0.5
0.0625	19.15	4.15	1.77
0.1250	26.54	6.02	2.35
0.1875	30.02	7.21	2.80
0.2500	31.87	8.02	3.15
0.3125	32.96	8.60	3.44
0.3750	33.64	9.03	3.68
0.4375	34.10	9.36	3.89
0.5000	34.43	9.63	4.08

It is apparent from Table 15 that heterotic lethals take considerably longer time, on an average, to disappear from the population than completely or partially recessive lethals. For instance, when the initial gene frequency of the mutant is 0.25, heterotic lethals would persist for about 32 generations as against about 8 generations required for completely recessive lethals and about 3 generations for partially recessive lethals. The effect of the initial gene frequency is also to increase the average time until loss. The expectation of life of a single mutant, when the population size is $N = 8$, corresponds to the initial gene frequency of $(\frac{1}{2}N) = 0.0625$ and is about 19, 4 and 2 generations for heterotic, completely recessive and partially recessive lethals.

Consider now the case when the population is in equilibrium due to the opposing processes of selection and mutation so that there is no change in the gene frequency. The opposition between these centripetal processes as a group and the scattering effects of random processes determine a frequency distribution of lethal genes. The frequency of classes segregating for lethals is given by $2Nu\,S_{1j}$ for $j = 1, 2, \ldots N$ whereas that of its absence from the population is $F(0)$ given by (6.18). If we consider the frequency distribution of the lethals present at any given time disregarding the cases when it is absent, the distribution is obtained by taking $\dfrac{F(q)}{1 - F(0)}$. This is, therefore, given by

$$(6.42) \quad S_{1j}{}^{*} = \frac{S_{1j}}{S_1{}^{*}} \quad \text{for } j = 1, 2, \ldots \ldots N$$

and is such that

$$(6.43) \quad \sum_j S_{1j}^* = 1$$

Now in such an equilibrium population, the proportion of equivalent loci at which the lethal is represented with a frequency $j/(2N)$ is S_{1j}^*. From this class of loci, the mean time of elimination of all mutants is S_j^*. Summing over all possible classes of loci, the expectation of life of the lethals present at any given time is, therefore

$$(6.44) \quad L = \sum_j S_{1j}^* S_j^*$$

The expectation of life of a lethal tells us how long a mutant lethal is expected to last, on an average.

The expectation of life of a new lethal mutant at its initial occurrence (S_1^*) and of any lethal (L) were obtained on the computer for $N = 50$ and for various values of h. The results are shown in figure 21. It is apparent from the graph that heterotic lethals have expectations of life higher than completely recessive lethals whereas the latter type of lethals have higher expectations than partially recessive lethals. A completely recessive lethal $(h = 0)$ takes, on an average,
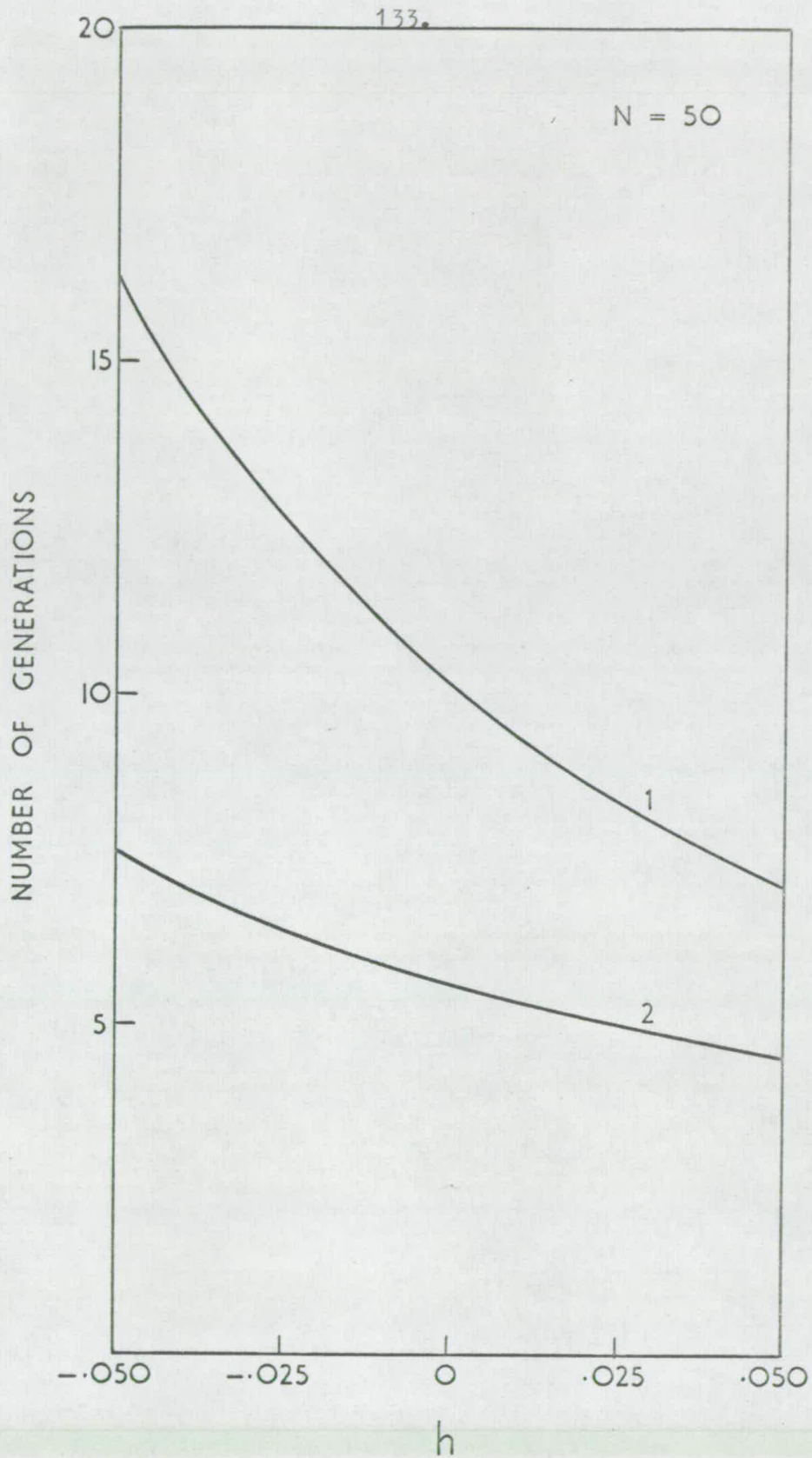
Fig. 21   The expectation of life of a lethal
           mutant (N=50). (1) refers to expec-
           tation at any given time, whereas
           (2) refers to expectation at the
           initial occurrence of the mutant.

about 6 generations to disappear from the population since its
initial occurrence. But in a random sample of completely recessive
lethals present at any given time, a lethal takes, on an average,
about 10 generations before disappearing. Another point worth noting
is that the decline in the expectation of life from heterotic conditions
through complete recessiveness, to partially recessiveness is more sharp
in L than in S_1^*. The former is, however, always higher than the
latter. The ratios (L/S_1^*) for various values of h are shown in
Table 16.

TABLE 16 (L/S_1^*) for various values of h $(N = 50)$.

h	(L/S_1^*)
-0.05	2.13
-0.04	2.06
-0.03	1.99
-0.02	1.93
-0.01	1.87
0.00	1.81
0.01	1.76
0.02	1.71
0.03	1.66
0.04	1.62
0.05	1.58

The ratio varies between 2.13 to 1.58. For a completely recessive
lethal it is 1.81.

6.42   AVERAGE AGE OF A LETHAL MUTANT

The realisation of a stable distribution of lethal genes is based
on regarding the processes of gene mutation and selection due to
lethality as occurring continually through time in a population of
finite size. A mutation to lethal gene, when occurs at any time, may
either disappear after its initial occurrence or may exist for several
generations before disappearing. Suppose a random point is chosen on
the time axis and a cross section of the then existing population is
taken. The lethals present at this point of time, might have existed
for different lengths of time since their initial occurrences. The mean
length of time for which they have stayed in the population up to the
given instant of time provides with the average age of a lethal mutant.
Further the lethals present at this moment of time may have different
lengths of time till they disappear. The mean length of time till they
disappear provide with the expectation of life of a lethal mutant.
We may also argue in terms of the proportion of equivalent loci at which
a given lethal mutant survives for different numbers of generations.
Suppose we start initially with a single mutant in a population of size
$2N$ genes and suppose l_t represents the number of equivalent loci at
which the lethal survives for t generations where $t = 0, 1, 2, \ldots$
with $l_o = 1$. Then the relating frequency of loci at which the lethal
survives for t generations is $l_t / \Sigma l_t$. The average age of the lethal

$\bar{A}$ is, therefore,

$$(6.45) \quad \bar{A} = \frac{\Sigma \, t l_t}{\Sigma l_t}$$

A given mutant at time t occurs in l_t loci but at time $(t + 1)$ it occurs in l_{t+1} loci, at time $(t + 2)$ in l_{t+2} loci and so on. Hence its expectation of life at time t is given by

$$(6.46) \quad L_t = \frac{l_{t+1} + l_{t+2} + \ldots \ldots}{l_t}$$

Summing over time, the average expectation of life is given by

$$(6.47) \quad L = \frac{\Sigma \, l_t \, L_t}{\Sigma l_t}$$

But

$$
\begin{aligned}
(6.48) \quad \Sigma \, l_t \, L_t &= \Sigma (l_{t+1} + l_{t+2} + \ldots \ldots) \\
&= l_1 + l_2 + l_3 + \ldots \ldots \\
&\quad + l_2 + l_3 + \ldots \ldots \\
&= \Sigma \, t l_t
\end{aligned}
$$

Hence

$$(6.49) \quad L = \bar{A}$$

It is thus shown that the average age of a lethal is the same as the average expectation of life. From the matrix approach this can be proved as follows.

It has already been shown that the elements S_{1j}^{*} of the stable distribution of the existing lethals are provided by the fundamental matrix $(\underset{\sim}{I} - \underset{\sim}{Q})^{-1}$, normalized to unity. If this is expanded, we get

$$(6.50) \quad \frac{1}{S_1^{*}} (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} = \frac{I + Q + Q^2 + \ldots\ldots}{S_1^{*}}$$

As shown in Chapter II S_{1j}^{*} can be obtained from (6.45) by premultiplying by $\underline{f_1}' = (1\ 0\ 0\ \ldots\ 0)$. This means that the terms in the above series when premultiplied by $\underline{f_1}'$ represent successively the probability distribution of the existing mutant individuals after one generation, two generations, and so on. If, therefore, we consider the matrix sum

$$(6.51) \quad \frac{1}{S_1^*} (\underset{\sim}{I} - \underset{\sim}{Q})^{-2} = \frac{\underset{\sim}{I} + 2\underset{\sim}{Q} + 3\underset{\sim}{Q}^2 + \ldots \ldots}{S_1^*}$$

premultiply it by f_1' and add the elements of the resulting vector, we obtain the average age of a lethal in the stable distribution of lethal genes. That is

$$(6.52) \quad \bar{A} = \frac{1}{S_1^*} f_1' (\underset{\sim}{I} - \underset{\sim}{Q})^{-2} \underset{\sim}{e}$$

$$= \frac{1}{S_1^*} f_1' (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \underset{\sim}{e}$$

$$= \frac{1}{S_1^*} \sum_j S_{1j} S_j^*$$

$$= \sum_j S_{1j}^* S_j^*$$

$$= L$$

Hence $\bar{A} = L$.

ACKNOWLEDGEMENTS

REFERENCES

ALLAN, J.S. and ROBERTSON, A. 1964. The effect of an initial
reverse selection upon total selection response. Genet.
Res. $\underline{5}$: 68-79.

CROW, J.F. and MORTON, N.E. 1955. Measurement of gene frequency
drift in small populations. Evolution, $\underline{9}$: 202-214.

CROW, J.F. and TEMIN, R.G. 1964. Evidence for the partial dominance
of recessive lethal genes in natural populations of
Drosophila. Am. Naturalist $\underline{98}$: 21-33.

DOBZHANSKY, Th. and WRIGHT, S. 1941. Genetics of natural populations.
V. Relations between mutation rate and accumulation of
lethals in populations of Drosophila pseudoobscura.
Genetics $\underline{26}$: 23-52.

EWENS, W.J. 1963. Numerical results and diffusion approximations in a
genetic process. Biometrika $\underline{50}$: 241-249.

EWENS, W.J. 1964. The pseudo-transient distribution and its uses in
genetics. J. Appl. Prob. $\underline{1}$: 141-156.

FADDEEVA, V.N. 1958. Computational Methods of Linear Algebra. Dover,
New York.

FALCONER, D.S. 1960. Introduction to Quantitative Genetics. Oliver and
Boyd, Edinburgh.

FELLER, W. 1951. Diffusion processes in genetics. Proc. 2nd Berkeley Symp. Mathematical Statistics and Probability, pp. 227-246. University of California Press.

HILL, W.G. and ROBERTSON, A. 1966. The effect of linkage on limits to artificial selection. Genet. Res., Camb. 8: 269-294.

HILL, W.G. and ROBERTSON, A. 1968. The effects of inbreeding at loci with heterozygote advantage. Genetics 60: 615-628.

KEMENY, J.G. and SNELL, J.L. 1960. Finite Markov Chains. Princeton, New York.

KIMURA, M. 1957. Some problems of stochastic processes in genetics. Ann. Math. Statist. 28: 882-901.

KIMURA, M. 1962. On the probability of fixation of mutant genes in a population. Genetics 47: 713-719.

KIMURA, M. 1964. Diffusion Models in Population Genetics. Methuen's Review Series in Applied Probability Vol. 2.

KIMURA, M. and CROW, J.F. 1963. The measurement of effective population number. Evolution 17: 279-288.

KIMURA, M. and OHTA, T. 1968. The average number of generations until fixation of a mutant gene in a finite population. Genetics. In press.

NARAIN, P. 1965. The description of gene action and interaction with multiple alleles in continuous variation. Genetics 52: 43-53.

NEI, M. 1968. The frequency distribution of lethal chromosomes in finite populations. Proc. Nat. Acad. Sci. 60: 517-524.

PROUT, T. 1967. Theory of the allelism between Drosophila lethals collected at different times. Genetics 56: 659-666.

ROBERTSON, A. 1952. The effect of inbreeding on the variation due to recessive genes. Genetics 37: 189-207.

ROBERTSON, A. 1960. A theory of limits in artificial selection. Proc. Roy. Soc. London B 153: 234-249.

ROBERTSON, A. 1962. Selection for heterozygotes in small populations. Genetics 47: 1291-1300.

STEGUN, I.A. 1965. Miscellaneous Functions. Mathematical functions with formulae, graphs and mathematical tables. Ed. Milton Abramowitz and Irene A. Stegun, Dover publications, Inc; New York. pp. 1004-1005.

WALLACE, B. 1966. Distance and the allelism of lethals in a tropical population of Drosophila melanogaster. Am. Naturalist 100: 565-578.

WRIGHT, S. 1931. Evolution in Mendelian populations. Genetics 16: 97-159.

WRIGHT, S. 1937. The distribution of gene frequencies in populations. Proc. Nat. Acad. Sci. 23: 307-320.

WRIGHT, S. 1938. The distribution of gene frequencies under irreversible mutation. Proc. Nat. Acad. Sci. 24: 253-259.

WRIGHT, S. 1945. The differential equation of the distribution of
gene frequencies. Proc. Nat. Acad. Sci. 31: 382-389.

WRIGHT, S., DOBZHANSKY, Th. and HOVANITZ, W. 1942. Genetics of
natural populations. VII. The allelism of lethals in
the third chromosome of Drosophila pseudoobscura.
Genetics 27: 363-395.

APPENDIX

DERIVATION OF MATRIX FORMULAE

(a) <u>Probability of fixation and expected change in the gene frequency</u>

Consider a finite absorbing Markov chain with $(2N + 1)$ states
E_o, E_1, E_2, - - - - - - - E_{2N-1}, E_{2N}, the ith state E_i representing
the state of i A_1 genes and $(2N - i)$ A_2 genes in a population of
$2N$ genes. In other words, a variate X can assume the values
$X_i = i/(2N)$, for $i = 0, 1, - - - - - - 2N$. The states E_o and E_{2N}
are called absorbing states whereas E_i, $i = 1, 2, - - - (2N - 1)$ are
called transient states. Let $P_{ij}(t_1, t_2)$ be the conditional probability
that the system is in state E_j at time t_1, given that it was in state
E_i at time t_2 i.e. it represents the probability of transition from
E_i to E_j after a time $(t_1 - t_2)$. Mathematically, this means

Probability $[X = X_j$ at $t_1 / X = X_i$ at $t_2]$; $t_1 > t_2$

Let the process be homogeneous in time i.e. $P_{ij}(t_1, t_2)$ depends only on
the difference $(t_1 - t_2)$ and not on t_1 and t_2. We can write this
probability as $P_{ij}(t)$, representing the probability that the system is
in state E_j at time $t + \zeta > t$, given that it was in state E_i

at time τ for $\tau \geqslant 0$. This is known as t-step transition probability, one step transition probability being written as P_{ij}. A transition from i to j after t-steps means a transition from i to k in one step and then from k to j in $(t - 1)$-steps, the probability of simultaneous realization of these events being $P_{ik} P_{kj}^{(t-1)}$ for $k = 0, 1, - - - - - -, 2N$. Hence we have, the Chapman-Kolmogorov equation (Feller 1951)

$$(A.1) \quad P_{ij}^{(t)} = \sum_{k=0}^{2N} P_{ik} P_{kj}^{(t-1)}$$

The one-step transition probability matrix has been represented as $\underset{\sim}{P}$ by (2.1). Let the t-step transition probability matrix be represented by $\underset{\sim}{P}(t)$. Then the matrix equation corresponding to (A.1) is given by

$$(A.2) \quad \underset{\sim}{P}(t) = \underset{\sim}{P} \, \underset{\sim}{P}(t - 1)$$

$$= \underset{\sim}{P}^{t-1} \, \underset{\sim}{P}(1)$$

$$= \underset{\sim}{P}^{t}$$

The elements of $\underset{\sim}{P}(t)$ satisfy the conditions

(A.3) $P_{ij}(t) \geq 0$ for all i, j

(A.4) $\sum_{j=0}^{2N} P_{ij}(t) = 1$ for all i

Since $\underset{\sim}{P}$ has been partitioned as in (2.6), the partitioning of $\underset{\sim}{P}(t)$ will be given by

(A.5) $\underset{\sim}{P}(t) = \begin{bmatrix} 1 & 0 & \underline{0}' \\ 0 & 1 & \underline{0}' \\ \underline{P}_o(t) & \underline{P}_{2N}(t) & \underset{\sim}{Q}^t \end{bmatrix}$

where

(A.6) $\underline{P}_o(t) = (\underset{\sim}{I} - \underset{\sim}{Q}^t)(\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \underline{P}_o$

(A.7) $\underline{P}_{2N}(t) = (\underset{\sim}{I} - \underset{\sim}{Q}^t)(\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \underline{P}_{2N}$

We shall now show that vectors $\underline{P}_{2N}(t)$ and $\underline{P}_o(t)$ respectively give the probabilities of fixation of A_1 and A_2

Let $u_i(t)$ be the probability that at time t, and not sooner, the population with initial gene frequency of A_1 as $i/(2N)$ becomes fixed

for A_1 and let $U_i^{(t)}$ be the probability that it has become fixed for A_1 by the tth generation. In other words, we have

$$(A.8) \quad U_i^{(t)} = \sum_1^t u_i^{(t)}$$

Since fixation at time t in one generation means that transition from initial state to the absorbing state takes place in one step, we have

$$(A.9) \quad U_i^{(1)} = u_i^{(1)} = P_{i,\,2N}$$

Now fixation at time t can take place in $(2N-1)$ mutually exclusive ways, the kth way being that the initial gene frequency becomes $k/(2N)$ in the first step and then fixation takes place in $(t-1)$ steps. The probability of simultaneous realization of these two independent events is $P_{ik}\,u_k^{(t-1)}$. Hence

$$(A10) \quad u_i^{(t)} = \sum_{k=1}^{2N-1} P_{ik}\,u_k^{(t-1)}$$

If we denote by $\underline{u}(t)$ and $\underline{U}(t)$ the column vectors of $u_i^{(t)}$ and

$U_i(t)$ respectively, we can write these relations in matrix notations as

$$(A.11) \quad \underline{u}(t) = \underset{\sim}{Q} \, \underline{u}\,(t-1)$$

$$= \; - \; - \; - \; - \; -$$

$$= \underset{\sim}{Q}^{t-1} \, \underline{u}\,(1)$$

$$= \underset{\sim}{Q}^{t-1} \, \underline{P_{2N}}$$

Then we have

$$(A.12) \quad \underline{U}(t) = (\underset{\sim}{I} + \underset{\sim}{Q} + \underset{\sim}{Q}^2 + - \; - \; - + \underset{\sim}{Q}^{t-1}) \; \underline{P_{2N}}$$

$$= (\underset{\sim}{I} - \underset{\sim}{Q}^{t}) \, (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \; \underline{P_{2N}}$$

$$= \underline{P_{2N}}(t)$$

Similarly if $\underline{l}(t)$ and $\underline{L}(t)$ denote the corresponding vector of fixation probabilities of the other gene A_2, we have

$$(A.13) \quad \underset{\sim}{L}(t) = (\underset{\sim}{I} - \underset{\sim}{Q}^t)(\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \underset{\sim}{P}_o$$

$$= \underset{\sim}{P}_o(t)$$

If $\underset{\sim}{W}(t)$ denotes the vector of probabilities that a population with initial frequency of A_1 gene as $i/(2N)$ is still segregating in t^{th} generation, we have

$$(A.14) \quad \underset{\sim}{W}(t) = \underset{\sim}{Q}^t \underset{\sim}{e}$$

We, thus see that the fixation probability vectors $\underset{\sim}{U}(t)$ and $\underset{\sim}{L}(t)$ can be obtained in two ways either by powering the $P-$ matrix as in (A.5) or as matrix functions of $Q-$ matrix as proved in (A.12) and (A.13). When $t \to \infty$, we get

$$(A.15) \quad \underset{\sim}{U} = \underset{\sim}{U}(\infty) = (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \underset{\sim}{P}_{2N}$$

$$(A.16) \quad \underset{\sim}{L} = \underset{\sim}{L}(\infty) = (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \underset{\sim}{P}_o$$

$$(A.17) \quad \underset{\sim}{W} = \underset{\sim}{W}(\infty) = \underset{\sim}{0}$$

We thus see that (A.15) provide with an alternative way of proving the matrix formula (2.14).

Now we give an alternative way of proving (2.15) as below:

Let the expected frequency of A_1 by the tth generation be denoted by $q_i^{(t)}$ when the initial population has its frequency as $q_i^{(o)} = i/(2N)$, and the expected response by the tth generation be $R_i^{(t)} = q_i^{(t)} - q_i^{(o)}$. In vector notations, we can put the response as $\underline{R}(t) = \underline{q}(t) - \underline{q}(o)$. The expected gene frequency by the tth generation can be obtained by finding the mean of the variate $X_j = j/(2N)$ for the distribution given by the ith row of $\underset{\sim}{P}(t)$ i.e.

$$(A.18) \quad q_i^{(t)} = \sum_{j=o}^{2N} P_{ij}^{(t)} X_j$$

$$= \sum_{j=1}^{2N-1} P_{ij}^{(t)} X_j + P_{i,\,2N}^{(t)}$$

In matrix notations, this means

$$(A.19) \quad \underline{q}(t) = \underset{\sim}{Q}^t \underline{q}(o) + \underline{U}(t)$$

If $\triangle q_i$ represents the initial change in the mean gene frequency, we have

$$(A.20) \quad q_i^{(1)} = \Delta q_i + q_i^{(0)}$$

$$= \sum_{j=1}^{2N-1} P_{ij} X_j + P_{i, 2N}$$

In matrix notation, this becomes

$$(A.21) \quad \underline{q}(1) = \Delta \underline{q} + \underline{q}(o)$$

$$= \underset{\sim}{Q} \, \underline{q}(o) + \underline{P_{2N}}$$

so that

$$(A.22) \quad (\underset{\sim}{I} - \underset{\sim}{Q}) \, \underline{q}(o) = \underline{P_{2N}} - \Delta \underline{q}$$

giving

$$(A.23) \quad (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \, \underline{P_{2N}} - \underline{q}(o) = (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \Delta \underline{q}$$

$$(A.24) \quad \underset{\sim}{R}(t) \quad = \quad \underline{q}(t) - \underline{q}(o)$$

$$= \quad \underset{\sim}{Q}^t \, \underline{q}(o) + \underline{U}(t) - \underline{q}(o)$$

$$= \quad -(\underset{\sim}{I} - \underset{\sim}{Q}^t) \, \underline{q}(o) + (\underset{\sim}{I} - \underset{\sim}{Q}^t)(\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \, \underline{P_{2N}}$$

$$= \quad (\underset{\sim}{I} - \underset{\sim}{Q}^t) \, [(\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \, \underline{P_{2N}} - \underline{q}(o)]$$

$$= \oint (\underset{\sim}{I} - \underset{\sim}{Q}^t)(\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \triangle q$$

Letting $t \rightarrow \infty$, this gives the expected selection limit vector $\underline{R}$ as

$$(A.25) \quad \underline{R} \quad = \quad \underline{U} - \underline{q}(o)$$

$$= \quad (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \triangle q$$

which is the same as (2.15)

(b) <u>Probability generating function of time until fixation of A_1.</u>

In order to study the distribution of time until fixation of allele A_1 disregarding the cases in which it is lost, we have to consider conditional transition matrix $\underset{\sim}{Q}^c$ giving transition probabilities P_{ij}^c relative to the hypothesis that the population ends up with the fixation of A_1. Following Kemeny and Snell (1960), we can define P_{ij}^c as

$$(A.26) \quad P_{ij}^c = \frac{P_{ij} U_j}{U_i}$$

with

$$(A.27) \quad U_{2N} = 1$$

Let

$$(A.28) \quad \underset{\sim}{D}_U = \text{diag} (U_1, U_2, - - - - - - U_{(2N-1)})$$

Then

$$(A.29) \quad \underset{\sim}{Q}^c = \underset{\sim}{D}_U^{-1} \underset{\sim}{Q} \underset{\sim}{D}_U$$

$$(A.30) \quad (\underset{\sim}{I} - \underset{\sim}{Q}^c)^{-1} = \underset{\sim}{D}_U^{-1} (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \underset{\sim}{D}_U$$

In order to derive the probability generating function for the time until fixation of A_1 relative to the hypothesis that the population ends up with the fixation of A_1, **are** make use of the probability generating function of time until homozygosity given in Watterson (1961). Let T_i be the time taken to first reach fixation or loss of A_1, given the initial population with i A_1 genes and $(2N-i)$ A_2 genes and let $S_i^{(t)}$ be the probability that $T_i = t$. Then

$$(A.31) \quad S_i^{(t)} = P_{io} + P_{i, 2N}$$

and the probability generating function $\pi_i(z)$ is given by

$$(A.32) \quad \pi_i(z) = \sum_{t=0}^{\infty} z^t \cdot S_i(t)$$

$$= \sum_{t=1}^{\infty} z^t S_i(t)$$

$$= z S_i(1) + \sum_{t=2}^{\infty} z^t S_i(t)$$

$$= z S_i(1) + \sum_{t=2}^{\infty} z^t \sum_{k=1}^{2N-1} P_{ik} S_k(t-1)$$

$$= z S_i(1) + z \sum_k P_{ik} \sum_{t=2}^{\infty} z^{t-1} S_k(t-1)$$

$$= z(P_{io} + P_{i,\,2N}) + z \sum_k P_{ik} \pi_k(z)$$

This can be expressed in matrix notations as

$$(A.33) \quad (I - z Q) \pi(z) = z(I - Q) e$$

where z is still a scalar and $\pi(z)$ is the vector of probability generating functions. Hence

$$(A.34) \quad \pi(z) = z(I - zQ)^{-1}(I - Q) e$$

In the conditional case Q and its functions are to be replaced by Q^c

and its appropriate functions. This gives

$$(A.35) \quad \underset{\sim}{\Pi}^{c}(z) \; = \; z \, (\underset{\sim}{I} - z\underset{\sim}{Q}^{c})^{-1} \, (\underset{\sim}{I} - \underset{\sim}{Q}^{c})$$

$$= \; z \, \underset{\sim U}{D}^{-1} \, (\underset{\sim}{I} - z\underset{\sim}{Q})^{-1} \, \underset{\sim U}{D} \, \underset{\sim U}{D}^{-1} \, (\underset{\sim}{I} - \underset{\sim}{Q}) \, \underset{\sim U}{D}$$

$$= \; z \, \underset{\sim U}{D}^{-1} \, (\underset{\sim}{I} - z\underset{\sim}{Q})^{-1} \, (\underset{\sim}{I} - \underset{\sim}{Q}) \, \underline{U}$$

This has been used in (3.58).

For mean time until fixation of A_1, we differentiate the probability generating function once and put $z = 1$. This gives

$$(A.36) \quad (d/dz) \, \underset{\sim}{\Pi}^{c}(z) \Big|_{z=1} \; = \; (d/dz) \, \underset{\sim U}{D}^{-1} \, (z^{-1}\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \, (\underset{\sim}{I} - \underset{\sim}{Q})\underline{U} \Big|_{z=1}$$

$$= \; \underset{\sim U}{D}^{-1} \, z^{-2} \, (z^{-1} \, \underset{\sim}{I} - \underset{\sim}{Q})^{-2} \, (\underset{\sim}{I} - \underset{\sim}{Q}) \, \underline{U} \Big|_{z=1}$$

$$= \; \underset{\sim U}{D}^{-1} \, (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \, \underline{U}$$

This can be compared with (2.20). The second factorial moment i.e. $E[t\,(t-1)]$ is given by

$$(A.37) \quad (d^2/dz^2) \, \underline{\Pi}^0(z)\Big|_{z=1} = \underset{\sim}{D}_U^{-1} (-2) \, z^{-3} (z^{-1} \underset{\sim}{I} - \underset{\sim}{Q})^{-2} (\underset{\sim}{I} - \underset{\sim}{Q})\underline{U}\Big|_{z=1}$$

$$+ \underset{\sim}{D}_U^{-1} (2) \, z^{-4} (z^{-1} \underset{\sim}{I} - \underset{\sim}{Q})^{-3} (\underset{\sim}{I} - \underset{\sim}{Q})\underline{U}\Big|_{z=1}$$

$$= 2\underset{\sim}{D}_U^{-1} [(\underset{\sim}{I} - \underset{\sim}{Q})^{-2} - (\underset{\sim}{I} - \underset{\sim}{Q})^{-1}] \, \underline{U}$$

$$(A.38) \quad \therefore \; E(t^2) = E[t\,(t-1)] + E(t)$$

$$= 2 \underset{\sim}{D}_U^{-1} [(\underset{\sim}{I} - \underset{\sim}{Q})^{-2} - (\underset{\sim}{I} - \underset{\sim}{Q})^{-1}]\underline{U} + \underset{\sim}{D}_U^{-1} (\underset{\sim}{I} - \underset{\sim}{Q})^{-1} \, \underline{U}$$

$$= \underset{\sim}{D}_U^{-1} [2 (\underset{\sim}{I} - \underset{\sim}{Q})^{-2} - (\underset{\sim}{I} - \underset{\sim}{Q})^{-1}] \, \underline{U}$$

which can be compared by (2.22).

REFERENCES

Feller, W. 1950. <u>An Introduction to Probability Theory and its Applications</u>.
New York: John Wiley and sons.

Kemeny, J.G. and Snell, J.L. 1960. <u>Finite Markov chains</u>. Princeton,
New York.

Watterson, G. 1961. Markov chains with absorbing states (a genetic
example). Ann. Math. Statist. <u>32</u>, 716-729.