

THE CONDITIONAL MARKOV CHAIN IN A GENETIC CONTEXT

PREM NARAIN

Institute of Agricultural Research Statistics, Library Avenue
New Delhi 110 012

Population genetics deals with a study of changes which the gene pool of a Mendelian population may undergo when it is exposed to systematic forces such as selection, mutation and migration. When the population is of limited size, the sample of genes transmitted to the next generation can deviate randomly from the true genetic composition of the parental generation and these random changes can accumulate over several generations. In other words, the change in the gene frequency over time due to systematic as well as random forces is a stochastic process. Usually the behaviour of the gene frequency in a generation depends only on its value in the immediately preceding generation so that the process is Markovian in structure. It can be studied either approximately as a diffusion process in which gene frequency as well as generations are treated as continuous or strictly as a finite Markov chain in which gene frequency is a discrete random variable and generations are discrete. In a series of investigations [Narain (1969), Narain and Robertson (1969), Robertson and Narain (1971), Narain (1971a)], it was shown how the process can be treated as a finite Markov chain and how the use of a transition probability matrix can be helpful in a genetic context. In particular, it was shown how to calculate the probability of fixation of a gene as well as the first two moments of the distribution of time taken for its fixation, disregarding the cases in which it is lost. The calculation of the first two moments of time until fixation of a particular allele was also attempted by the diffusion approach [Narain (1970), Narain (1974)]. The last investigation as well as that of Ewens (1973) have demonstrated that invoking a conditional process facilitates the calculation of the moments of the distribution of time taken for the fixation of a gene. This aspect is intimately connected with the concept of average number of generations required to attain limits of genetic improvement due to artificial selection which was first introduced in Narain (1969) and later elaborated in Narain (1971b). Although the diffusion approach to the conditional process is completely documented in Ewens (1973) and Narain (1974), the transition matrix approach is only briefly indicated in the former reference. The purpose of this paper is therefore to describe the conditional Markov chain and demonstrate its application in a genetic context relating to response to selection in finite populations. In addition, the theory has been applied to study the effect of linkage on the mean and variance of time until fixation of a gamete in populations practising self-fertilization.

CONDITIONAL MARKOV CHAIN

Assuming no mutation, consider a finite population of gametes of size $2N$ (corresponding to a population of diploid individuals of size N) and a single locus with two alleles A and a . Such a population can assume $(2N+1)$ states $E_0, E_1, \dots, E_{2N-1}, E_{2N}$, the state E_i representing the state of i A genes and $(2N-i)$ a genes. The frequency of A , denoted by x_i for the population in state E_i , can then take values $i/2N$, $i = 0, 1, \dots, (2N-1), 2N$. When $x_i = 0$ or 1 for $i = 0$ and $2N$ respectively, the population is said to be fixed for A or a respectively. But when $x_i \neq 0$ or 1 , the population is said to be segregating for A and a alleles. Such a genetic situation corresponds to a finite absorbing Markov Chain with two absorbing states E_0 and E_{2N} and $(2N-1)$ transient states $E_1, E_2, \dots, E_{2N-1}$. A detailed description of this chain, in such a context, is given in Narain (1971a). If P_{ij} represents one step transition probability for the system to move from E_i to E_j , the transition probability matrix \underline{P} of order $(2N+1) \times (2N+1)$ takes the form

$$\underline{P} = \begin{bmatrix} 1 & \underline{0} & 0 \\ \underline{P}_0 & \underline{Q} & \underline{P}_{2N} \\ 0 & \underline{0} & 1 \end{bmatrix} \quad (1)$$

where \underline{Q} is of order $(2N-1) \times (2N-1)$, giving the one-step transition probabilities amongst the transient states only, \underline{P}_0 and \underline{P}_{2N} are column-vectors of order $(2N-1) \times 1$ representing the one-step transition probabilities from a transient state to E_0 and E_{2N} respectively. The vectors of the eventual probabilities of fixation of A and a , denoted by \underline{U} and \underline{L} are respectively given by

$$\underline{U} = (\underline{I} - \underline{Q})^{-1} \underline{P}_{2N} \quad (2)$$

$$\underline{L} = (\underline{I} - \underline{Q})^{-1} \underline{P}_0 \quad (3)$$

Consider now a finite absorbing Markov Chain conditional to the eventual absorption in E_{2N} . We then have only one absorbing state E_{2N} and $(2N-1)$ transient states E_1, \dots, E_{2N-1} from which absorption is only possible in E_{2N} . Let $P_{ij}^{(C1)}$ be the one-step transition probability for the system to move from E_i to E_j relative to the event of ultimate absorption in E_{2N} . Denoting by U_i , the i -th element of vector \underline{U} , the eventual probability of fixation of A when initially the population was in state E_i and following Kemeny and Snell (1960), we can define

$$P_{ij}^{(C1)} = P_{ij} U_j / U_i \quad (4)$$

with $U_{2N} = 1$. We then have the conditional one step transition probability matrix $\underline{P}^{(C1)}$, of order $2N \times 2N$ given by

$$\underline{P}^{(C1)} = \begin{bmatrix} \underline{Q}^{(C1)} & \underline{P}_{2N}^{(C1)} \\ \underline{0} & 1 \end{bmatrix} \quad (5)$$

where $\underline{Q}^{(C1)}$ is of order $(2N-1) \times (2N-1)$, giving the one-step transition probabilities amongst the transient states only, conditional to fixation in E_{2N} and $\underline{P}_{2N}^{(C1)}$ is the column vector of order $(2N-1) \times 1$ representing the one-step transition probability from a transient state to E_{2N} relative to the eventual absorption in E_{2N} . The corresponding t -step transition matrix is given by

$$\underline{P}^{(C1)}(t) = \begin{bmatrix} \underline{Q}^{(C1)}(t) & \underline{P}_{2N}^{(C1)}(t) \\ \underline{0}' & 1 \end{bmatrix} \quad (6)$$

and the use of Chapman-Kolmogorov [Feller (1951)] for the Conditional Markov Chain gives

$$\underline{P}^{(C1)}(t) = \left[\underline{P}^{(C1)} \right]^t \quad (7)$$

so that

$$\underline{Q}^{(C1)}(t) = \left[\underline{Q}^{(C1)} \right]^t \quad (8)$$

$$\underline{P}_{2N}^{(C1)}(t) = \left[\underline{I} - \underline{Q}^{(C1)}(t) \right] \left[\underline{I} - \underline{Q}^{(C1)} \right]^{-1} \underline{P}_{2N}^{(C1)} \quad (9)$$

Following Narain (1971a), the column vector $\underline{U}^{(C1)}(t)$ of the probability of fixation of A by the t -th generation relative to the eventual fixation for A is obtained as

$$\underline{U}^{(C1)}(t) = \underline{P}_{2N}^{(C1)}(t) \quad (10)$$

As t tends to infinity, $(\underline{Q}^{(C1)})^t$ tends to zero so that the vector of the eventual probability of fixation for A, for the conditional process, is given by

$$\underline{U}^{(C1)} = \left[\underline{I} - \underline{Q}^{(C1)} \right]^{-1} \underline{P}_{2N}^{(C1)} \quad (11)$$

Writing $\underline{D}_1 = \text{diag}(U_1, U_2, \dots, U_{2N-1})$ of order $(2N-1) \times (2N-1)$, we find

$$\underline{Q}^{(C1)} = \underline{D}_1^{-1} \underline{Q} \underline{D}_1 \quad (12)$$

$$\left[\underline{Q}^{(C1)} \right]^t = \underline{D}_1^{-1} \underline{Q}^t \underline{D}_1 \quad (13)$$

$$\left[\underline{I} - \underline{Q}^{(C1)} \right]^{-1} = \underline{D}_1^{-1} (\underline{I} - \underline{Q})^{-1} \underline{D}_1 \quad (14)$$

$$\underline{P}_{2N}^{(C1)} = \underline{D}_1^{-1} \underline{P}_{2N} \quad (15)$$

It, therefore, follows that

$$\underline{U}^{(C1)} = \underline{D}_1^{-1} \underline{U} = \underline{e} \quad (16)$$

a column vector of unities, as expected due to the conditioning of the process. Further, we get

$$\underline{U}^{(C1)}(t) = \left[\underline{I} - \underline{D}_1^{-1} \underline{Q}^t \underline{D}_1 \right] \underline{e} \quad (17)$$

for working out the probabilities of fixation of A by the t-th generation.

Similarly, if we consider a finite absorbing Markov Chain conditional to the eventual absorption in E_0 , we have again one absorbing state and $(2N-1)$ transient states. Defining the corresponding one-step transition probability $P_{ij}^{(C0)} = P_{ij} L_j / L_i$ with $L_0 = 1$ and proceeding in the same way as above, we get

$$\underline{P}^{(C0)}(t) = \left[\begin{array}{cc} 1 & \underline{0}^t \\ \underline{P}^{(C0)}(t) & \underline{Q}^{(C0)}(t) \end{array} \right] \quad (19)$$

where, writing $\underline{D}_0 = \text{diag}(L_1, L_2, \dots, L_{2N-1})$,

$$\underline{Q}^{(C0)}(t) = \underline{D}_0^{-1} \underline{Q}^t \underline{D}_0 \quad (20)$$

$$\begin{aligned} \text{and } \underline{P}_0^{(C0)}(t) &= \left[\underline{I} - \underline{D}_0^{-1} \underline{Q}^t \underline{D}_0 \right] \underline{e} \\ &= \underline{L}^{(C0)}(t) , \end{aligned} \quad (21)$$

giving the probabilities of fixation of a by the t-th generation in the conditional process with $\underline{L}^{(C0)} = \underline{e}$ as usual.

CONDITIONAL EXPECTED RESPONSE DUE TO ARTIFICIAL SELECTION

The random change in gene frequency due to finite population size has important applications in animal breeding as shown by Robertson (1960). The probability of fixation of the desirable allele can be converted into the expected response in the character under selection at the limit by making use of the relation between the selective advantage of a gene with its effect on the metric character under selection given first by Haldane (1931). Under the assumption of independent segregation of several loci affecting the character, the expected response at the limit, expressed in relation to the initial genetic standard deviation, is a function of Nih (the product of population size, intensity of selection and the square-root of heritability) and the initial frequency, p of the desirable allele, assumed equal at all loci. Narain (1971a) showed that this expected limit of response to selection, expressed in terms of the vector of changes in the frequency of desirable allele and denoted by $E(\underline{R})$ is given by

$$E(\underline{R}) = (\underline{I} - \underline{Q})^{-1} E(\underline{\delta p}) \quad (22)$$

where $E(\underline{\delta p})$ is the vector of initial expected responses. Also, the the vector of

the expected response by the t -th generation, $E[\underline{R}(t)]$ was shown to be equal to

$$E[\underline{R}(t)] = (\underline{I} - \underline{Q}^t) E(\underline{R}) \quad (23)$$

Invoking a conditional process of selection along the same lines as in the previous section, we get the vector of conditional expected response due to selection by the t -th generation as

$$E\left[\underline{R}^{(C)}(t)\right] = \left[\underline{I} - (\underline{Q}^{(C1)})^t\right] E(\underline{R}^{(C)}) \quad (24)$$

relative to the eventual fixation of A regarded as a desirable allele. The expression for $E(\underline{R}^{(C)})$, however, becomes, as expected

$$\begin{aligned} E(\underline{R}^{(C)}) &= (\underline{I} - \underline{Q}^{(C1)})^{-1} E(\delta p) \\ &= (\underline{I} - \underline{Q}^{(C1)})^{-1} \underline{P}_{2N}^{(C1)} - \underline{p}(0) \end{aligned} \quad (25)$$

where $\underline{p}(0)$ is the vector of the frequency of desirable allele in the initial population. We then have,

$$E\left[\underline{R}^{(C)}(t)\right] = \left[\underline{I} - \underline{D}_1^{-1} \underline{Q}^t \underline{D}_1\right] \left[\underline{e} - \underline{p}(0)\right] \quad (26)$$

PROBABILITY GENERATING FUNCTION OF THE DISTRIBUTION OF TIME UNTIL FIXATION OF A PARTICULAR ALLELE

Let T_i be the time taken to first reach fixation of A, given that the initial population contains i A genes and $(2N-i)$ a genes relative to the hypothesis of eventual absorption in E_{2N} . Let $S_i^{(t)}$ be the probability that $T_i = t$. Then clearly,

$$S_i^{(1)} = P_{i,2N}^{(C1)} \quad (27)$$

The probability generating function $\pi_i^{(C1)}(z)$, in this case, can then be expressed as

$$\begin{aligned} \pi_i^{(C1)}(z) &= z S_i^{(1)} + \sum_{t=2}^{\infty} z^t S_i^{(t)} \\ &= z P_{i,2N}^{(C1)} + z \sum_{t=2}^{\infty} z^{t-1} \sum_{i=1}^{2N-1} P_{ik}^{(C1)} S_k^{(t-1)} \\ &= z P_{i,2N}^{(C1)} + z \sum_{i=1}^{2N-1} P_{ik}^{(C1)} \pi_k^{(C1)}(z) \end{aligned} \quad (28)$$

In matrix notations, we can write it as

$$\underline{\pi}^{(C1)}(z) = z(\underline{I} - z\underline{Q}^{(C1)})^{-1} (\underline{I} - \underline{Q}^{(C1)}) \underline{e} \quad (29)$$

where z is still a scalar and $\underline{\pi}^{(C1)}(z)$ is the vector of probability generating functions conditional to fixation of A. Using the relationship between functions of $\underline{Q}^{(C1)}$ and \underline{Q} given in Section 2, we get

$$\underline{\pi}^{(C1)}(z) = z \underline{D}_1^{-1} (\underline{I} - z\underline{Q})^{-1} (\underline{I} - \underline{Q}) \underline{U} \quad (30)$$

The vector of the first moments of the distributions of time until fixation of A is obtained by differentiating $\underline{\pi}^{(C1)}(z)$ once and putting $z = 1$. This gives

$$\begin{aligned} E(\underline{T}_1) &= \left| (d/dz) \underline{\pi}^{(C1)}(z) \right|_{z=1} = \left| (d/dz) \underline{D}_1^{-1} (z^{-1} \underline{I} - \underline{Q})^{-1} (\underline{I} - \underline{Q}) \underline{U} \right|_{z=1} \\ &= \left| \underline{D}_1^{-1} z^{-2} (z^{-1} \underline{I} - \underline{Q})^{-2} (\underline{I} - \underline{Q}) \underline{U} \right|_{z=1} \\ &= \underline{D}_1^{-1} (\underline{I} - \underline{Q})^{-1} \underline{U} \end{aligned} \quad (31)$$

It is easy to see that this is also equivalent to $(\underline{I} - \underline{Q}^{(C1)})^{-1} \underline{e}$. The vector of the second factorial moment is given by

$$\begin{aligned} E(\underline{T}_1^2) - E(\underline{T}_1)^2 &= \left| (d^2/dz^2) \underline{\pi}^{(C1)}(z) \right|_{z=1} \\ &= 2 \underline{D}_1^{-1} [(\underline{I} - \underline{Q})^{-2} - (\underline{I} - \underline{Q})^{-1}] \underline{U} \end{aligned} \quad (32)$$

Using (31) and (32), the vector of the second moment about origin is obtained as

$$E(\underline{T}_1^2) = \underline{D}_1^{-1} [2(\underline{I} - \underline{Q})^{-2} - (\underline{I} - \underline{Q})^{-1}] \underline{U} \quad (33)$$

With the help of the elements of the vectors given by (31) and (33), one can obtain the variances of the time until fixation of A.

In a similar manner, we obtain the vector of the probability generating functions $\underline{\pi}^{(C0)}(z)$, of the distributions of time until fixation of a. This is given by

$$\begin{aligned} \underline{\pi}^{(C0)}(z) &= z(\underline{I} - z\underline{Q}^{(C0)})^{-1} (\underline{I} - \underline{Q}^{(C0)}) \underline{e} \\ &= z \underline{D}_0^{-1} (\underline{I} - z\underline{Q})^{-1} (\underline{I} - \underline{Q}) \underline{L} \end{aligned} \quad (34)$$

The vectors of the first and second moments in this case are given by

$$E(\underline{T}_0) = \underline{D}_0^{-1} (\underline{I} - \underline{Q})^{-1} \underline{L} \quad (35)$$

$$E(\underline{T}_0^2) = \underline{D}_0^{-1} [2(\underline{I} - \underline{Q})^{-2} - (\underline{I} - \underline{Q})^{-1}] \underline{L} \quad (36)$$

From which one can get the corresponding variances of the time until fixation of a.

EIGEN-ROOTS AND EIGEN-VECTORS OF THE CONDITIONAL MARKOV CHAIN WITH BINOMIAL TRANSITION PROBABILITIES

It is apparent from the above matrix derivations that for applying this theory, the element of $\underline{P}^{(C1)}$, $\underline{P}^{(C0)}$, \underline{D}_1 and \underline{D}_0 are required to be known. Analytically, this involves working out the eigen-roots and eigen-vectors of $\underline{P}^{(C1)}$ and $\underline{P}^{(C0)}$. Since the conditional transition matrices $\underline{P}^{(C1)}$ and $\underline{P}^{(C0)}$ depend on the conditional transition matrix \underline{P} and since \underline{D}_1 and \underline{D}_0 are shown to be certain functions of \underline{P} [Narain (1971a)], the problem boils down to a study of \underline{P} or its derivative \underline{Q} . For specifying the elements of \underline{P} , we consider, as an example, the binomial transition probabilities. This case is commonly known as Wright's model [Wright (1931)]. It assumes absence of selective forces and considers only random drift based on binomial sampling with a constant population size N . The eigen-roots and vectors of \underline{P} in such a case are also known [Feller (1951)]. Extension of such a model so as to involve selection in the context of limits of response to selection has been extensively studied by Narain and Robertson (1969). However, it is still of interest to study the eigen-roots and eigen-vectors of the conditional transition matrices $\underline{P}^{(C1)}$ and $\underline{P}^{(C0)}$. For this purpose we follow the approach given in Feller (1951).

With binomial sampling and no selection, we have

$$P_{ij} = \binom{2N}{j} p_1^j (1-p_1)^{2N-j} \quad \begin{array}{l} i = 0, 1, \dots, 2N \\ j = 0, 1, \dots, 2N \end{array} \quad (37)$$

$$P_{ij}^{(C1)} = \binom{2N}{j} p_1^j (1-p_1)^{2N-j} (p_j/p_1) \quad \begin{array}{l} i = 1, 2, \dots, 2N \\ j = 1, 2, \dots, 2N \end{array} \quad (38)$$

where $p_i = i/2N$.

The eigen-roots of $\underline{P}^{(C1)}$ are obtained by solving the characteristic equation

$$|\underline{P}^{(C1)} - \lambda \underline{I}| = 0 \quad (39)$$

It is found that the roots are given by

$$\lambda_r^{(C)} = (1-r/2N) \binom{2N}{r} r! / (2N)^r, \quad r = 0, 1, \dots, (2N-1) \quad (40)$$

For $r = 1, 2, \dots, (2N-1)$, the roots are the same as that of $\underline{Q}^{(C1)}$ and similarly as that of $\underline{Q}^{(C0)}$, which in view of (12), are the same as that of \underline{Q} viz.

$$\lambda_r = \binom{2N}{r} r! / (2N)^r, \quad r = 1, \dots, 2N-1 \quad (41)$$

Writing $j_{(v)} = j(j-1)\dots(j-v+1)$, we get

$$\sum_{j=1}^{2N} P_{ij}^{(C1)} j_{(v)} = [(2N)_{(v+1)} p_1^v + v(2N)_{(v)} p_1^{v-1}] / 2N \quad (42)$$

This shows that, taking $v=1$, the expected value of the gene frequency, with pure random drift, will not be simply p_i but instead given by

$$\sum_{j=1}^{2N} p_{ij}^{(C1)} (j/2N) = p_i + (1-p_i)/2N \quad (43)$$

so that

$$E(\delta p_i) = (1-p_i)/2N \quad (44)$$

Similarly, with $v=2$, we get

$$\sum_{j=1}^{2N} p_{ij}^{(C1)} (j/2N)^2 - (1/2N) \sum_{j=1}^{2N} p_{ij}^{(C1)} (j/2N) = (1-1/2N)(1-2/2N)p_i^2 + (2/2N)(1-1/2N)p_i \quad (45)$$

This shows that, using (43), the variance of the change in the gene frequency, with pure random drift, will not be simply $p_i(1-p_i)/2N$ but, instead, given by

$$V(\delta p_i) = (1/2N)(1-1/2N)p_i(1-p_i) \quad (46)$$

From (44) and (46) it is evident that for a population so large that $(1/2N)^2$ is negligible, the mean and variance of the change in gene frequency due to random drift, in the conditional process, are $(1-p_i)/2N$ and $p_i(1-p_i)/2N$ as against 0 and $p_i(1-p_i)/2N$ respectively in the unconditional case. This is exactly what we get from the diffusion approach for the pure random drift case as shown in Narain (1974). It is however interesting to note that for the exact process, conditioning the process increases the mean but decreases the variance.

Corresponding to each characteristic root $\lambda_r^{(C)}$ given by (40), the system of linear equations

$$\sum_{j=1}^{2N} p_{ij}^{(C1)} x_{jr} = \lambda_r^{(C)} x_{ir}, \quad i=1,2,\dots,2N \quad (47)$$

admits a non-trivial solution $\underline{x}_r = (x_{1r} \dots x_{(2N)r})$ known as the right-hand eigenvector for $\lambda_r^{(C)}$. It is, therefore, always possible to find constants a_0, a_1, \dots, a_r (not all of them zero) such that

$$x_{jr} = \sum_{v=0}^r a_v j(v) \quad (48)$$

is a solution of (47). Substituting (48) in (47) and using (42), we get

$$\sum_{v=0}^r a_v [(2N-v)P_{i+v}] \frac{(2N)_v P_i^{v-1}}{2N} = \lambda_r^{(C)} \sum_{v=0}^r a_v i_{(v)} \quad i=1,2,\dots,2N \quad (49)$$

Since coefficients of a_v on the both sides of (49) are polynomials of degree v in i , it is possible to write

$$P_i^v = \sum_{s=0}^v C_{s,v} i_{(s)} \quad (50)$$

where $C_{s,v}$'s are independent of i . Substituting (50) in (49) and equating the coefficients of $i_{(t)}$ for $t=0,1,\dots,r$, we get,

$$\lambda_r^{(C)} a_t = a_t (2N)_{(t+1)} C_{tt} / (2N) + \sum_{v=t+1}^r a_v [(2N-v)C_{tv} + vC_{t,(v-1)}] (2N)_v / (2N), \quad (t=0,1,\dots,r) \quad (51)$$

If we take $t=r$ in (51) and use (40) as well as (50), it is found that (51) is satisfied for $v=r$ and arbitrary a_r . We can then put $a_r=1$ in (51) for $t=(r-1)$ giving a_{r-1} . This procedure allows us to calculate a_{r-2}, \dots, a_1 and a_0 in succession, giving thereby the j -th element of the right-hand eigen-vector corresponding to $\lambda_r^{(C)}$ given by (48).

Using the above procedure of obtaining the eigen-vectors for $r=1,2$ and 3 , we find that for the three eigen-roots,

$$\begin{aligned} \lambda_1^{(C)} &= (1-1/2N) \\ \lambda_2^{(C)} &= (1-1/2N)(1-2/2N) \\ \lambda_3^{(C)} &= (1-1/2N)(1-2/2N)(1-3/2N) \end{aligned} \quad (52)$$

the vectors are, respectively, given by

$$\begin{aligned} x_{j1} &= (1-p_j) \\ x_{j2} &= (1-p_j)(1-2p_j) \\ x_{j3} &= (1-p_j)[(2N-1)/(10N-6) - p_j(1-p_j)] \end{aligned} \quad (53)$$

with $j=1,2,\dots,2N$. These results can be compared with the corresponding results of the unconditional case detailed in Narain and Robertson (1969). Although the roots are the same, the elements of the vectors are now $(1/p_j)$ of those in the unconditional case. Alternatively, since $\lambda_r^{(C)}$ is an eigenvalue of $\underline{\underline{Q}}^{(C1)}$ with the associated right eigen-vector, \underline{x}_r , we have

$$\underline{\underline{Q}}^{(C1)} \underline{x}_r = \lambda_r^{(C)} \underline{x}_r, \quad r = 1, 2, \dots, 2N-1$$

In view of (12), we get

$$\underline{\underline{Q}} \underline{\underline{D}}_1 \underline{x}_r = \lambda_r^{(C)} \underline{\underline{D}}_1 \underline{x}_r, \quad r = 1, 2, \dots, 2N-1.$$

This shows that $\lambda_r^{(C)}$ is also an eigenvalue of $\underline{\underline{Q}}$ with the associated right-eigenvector $\underline{z}_r = \underline{\underline{D}}_1 \underline{x}_r$, so that by definition of $\underline{\underline{D}}_1$, the j -th element of \underline{x}_r is $2N/j$ times as large as the j -th element of \underline{z}_r i.e. those corresponding to the unconditional case.

EFFECT OF LINKAGE ON THE MEAN AND VARIANCE OF TIME UNTIL FIXATION OF A GAMETE IN SELFED POPULATIONS

The effect of linkage on the probability of fixation of a gamete in populations practising self-fertilization was studied in Narain (1971a) which can be consulted for details. The case of self-fertilization corresponds to the situation when $N=1$. The population is sub-divided into lines from each of which two gametes are chosen to form one mature individual only. With two linked loci each with two alleles A-a and B-b respectively with recombination probability r , with $s = 1-r$, and assuming no mutation, there are 10 states of the system corresponding to 10 types of lines out of which four homozygous ones represent absorbing states and the remaining six are transient states. Amongst the transient states, the two corresponding to two double heterozygotes, AB/ab (coupling) and Ab/aB (repulsion) are important from the point of view of linkage, the remaining four involving single heterozygotes only. Taking the P-matrix of the process and the \underline{U}_{AB} vector of the probabilities of fixation of gamete AB from Narain (1971a), the $\underline{\underline{P}}^{(C11)}$ matrix for the process, conditional to absorption in AB/AB, has the form

$$\underline{\underline{P}}^{(C11)} = \begin{bmatrix} \underline{\underline{Q}}^{(C11)} & \underline{P}_{AB}^{(C)} \\ \underline{0}^r & I \end{bmatrix} \quad (54)$$

where

$$\underline{P}_{AB}^{(C)} = (1/2, 1/2, 0, 0, s^2(1+2r)/2, r(1+2r)/4) \quad (55)$$

$$\underline{\underline{Q}}^{(C11)} = \begin{bmatrix} 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ rs(1+2r)/2 & rs(1+2r)/2 & 0 & 0 & s^2/2 & r^3 \\ s(1+2r)/4 & s(1+2r)/4 & 0 & 0 & r/4 & s^2/2 \end{bmatrix} \quad (56)$$

The ordering of the states being AB/Ab, AB/aB, Ab/ab, aB/ab, AB/ab and Ab/aB. With the help of the results given in Narain (1971a) and using (31) & (33), the vectors of mean as well as second moment about origin of time until fixation of AB

are respectively given by

$$E(\underline{T}_{AB}^1) = [2, 2, 0, 0, \alpha_{AB}^{(c)}, \alpha_{AB}^{(r)}] \quad (57)$$

$$E(\underline{T}_{AB}^2) = [6, 6, 0, 0, \beta_{AB}^{(c)}, \beta_{AB}^{(r)}] \quad (58)$$

where $\alpha_{AB}^{(c)}$, $\beta_{AB}^{(c)}$ corresponding to the situation where the population is initially in the coupling phase, are given by

$$\alpha_{AB}^{(c)} = (1+2r)(1+4rs)/(1+2rs) + (1-2r)/(1+2r) \quad (59)$$

$$\beta_{AB}^{(c)} = (1+2r)(3+26rs+24r^2s^2)/(1+2rs)^2 + (1-2r)(3-2r)/(1+2r)^2 \quad (60)$$

and $\alpha_{AB}^{(r)}$, $\beta_{AB}^{(r)}$ corresponding to the situation when the population is initially in the repulsion phase, are given by

$$\alpha_{AB}^{(r)} = (1+2r)(1+4rs)/2r(1+2rs) - (1-2r)/2r(1+2r) \quad (61)$$

$$\beta_{AB}^{(r)} = (1+2r)(3+26rs+24r^2s^2)/2r(1+2rs)^2 - (1-2r)(3-2r)/2r(1+2r)^2 \quad (62)$$

In a similar manner, we get the corresponding vectors for moments of time until fixation of Ab, aB and ab. It is found that when the population is initially in the coupling phase, the means and second moments about origin of time until fixation of Ab as well as aB are the same as that given by (61) and (62) respectively whereas when the initial population is in repulsion phase, these are correspondingly given by (59) and (60). These results for the time until fixation of ab are exactly the same as that until fixation of AB given by (59) to (62). In each case, the variance of time until fixation is calculated by subtracting the square of a from β .

It is interesting to note that mean and variance of time until homozygosity can further be obtained by multiplying the mean and variance of time until fixation of a gamete by the corresponding probability of fixation and adding over the four possible cases. For the situation when the initial population is in the coupling phase, these are given by

$$E(T) = U_{AB}^{(c)} \alpha_{AB}^{(c)} + U_{Ab}^{(c)} \alpha_{Ab}^{(c)} + U_{aB}^{(c)} \alpha_{aB}^{(c)} + U_{ab}^{(c)} \alpha_{ab}^{(c)} = 2(1+4rs)/(1+2rs) \quad (63)$$

$$\begin{aligned} \text{Var}(T) &= U_{AB}^{(c)} \beta_{AB}^{(c)} + U_{Ab}^{(c)} \beta_{Ab}^{(c)} + U_{aB}^{(c)} \beta_{aB}^{(c)} + U_{ab}^{(c)} \beta_{ab}^{(c)} - [E(T)]^2 \\ &= 2(1+10rs-8r^2s^2)/(1+2rs)^2 \end{aligned} \quad (64)$$

where $U_{AB}^{(c)} = U_{ab}^{(c)} = 1/2(1+2r)$ and $U_{Ab}^{(c)} = U_{aB}^{(c)} = r/(1+2r)$ [Narain (1971a)]. Because of symmetry, (63) and (64) hold for the repulsion phase also. The values of $E(T)$ and $\text{Var}(T)$ obtained here are exactly the same as those obtained by Puri (1968) who obtained them directly without working out the time until fixation of a particular gamete.

The effect of the recombination fraction r on the mean and the standard deviation of time until fixation was numerically studied with the help of expressions (59) to (62). The results for the case when the initial population is in coupling phase are presented in Table 1. For the case when the initial population is in repulsion phase, the results are obtainable from the Table by interchanging either A and a or B and b.

Table 1: Mean and standard deviation of time (number of generations) until fixation of a gamete for the initial population with heterozygotes in coupling phase.

r	AB or ab		Ab or aB	
	Mean	s. d.	Mean	s. d.
0.0000	2.0000	1.4142	4.0000	2.0000
0.0625	2.0208	1.4337	3.7244	1.8622
0.1250	2.0743	1.4753	3.4973	1.7682
0.1875	2.1506	1.5213	3.3103	1.7078
0.2500	2.2424	1.5635	3.1516	1.6709
0.3125	2.3441	1.5959	3.0120	1.6486
0.3750	2.4513	1.6187	2.8873	1.6382
0.4375	2.5602	1.6306	2.7734	1.6354
0.5000	2.6666	1.6329	2.6666	1.6329

It is found that when the initial population is in coupling phase, the effect of linkage is to decrease the average and standard deviation of the number of generations until fixation for a coupled gamete (AB or ab) but to increase the same for a repulsed gamete (Ab or aB). It may be noted that when we consider independently segregating loci ($r = 0.50$) and fixation of coupled gametes with initial population in coupling phase (or of repulsed gametes with initial population in repulsion phase), the average time to fixation is about 1.33 times that for completely linked loci whereas the chance of fixation is half of its value for the completely linked case. As expected, with independent segregation, the average time to fixation of corresponding repulsed (or coupled) gametes, the chance of fixation being the same viz. 0.25 in all the four cases. But for every tight linkage (r approaching zero) average time to fixation of a repulsed gamete with initial population in coupling phase (or of a coupled gamete with initial population in repulsion phase) tends to a limiting value of 4 with chance of its fixation becoming very very small. As regards variability in time to fixation, a characteristic feature, true for all situations, is that a larger mean is accompanied by a larger standard deviation.

SUMMARY

A theory of the stochastic change in the frequency of a gene in finite populations conditional to its eventual fixation has been developed employing a Condi-

ditional Markov Chain. The probability generating function of the distribution of time until fixation of a particular allele as well as the eigen-roots and eigen-vectors of the conditional process with binomial transition probabilities have been studied. The theory has been applied to investigate the effect of linkage on the mean and standard deviation of time until fixation of a gamete in populations practising self-fertilization. It has been found that linkage decreases or increases the average and standard deviation of time to fixation of a coupled gamete according as the initial population consists of a coupling or repulsion heterozygote respectively.

ACKNOWLEDGEMENTS

The author is grateful to Professor E. Pollak, Iowa State University, U.S.A. as well as the referee for helpful suggestions which have considerably improved the presentation.

REFERENCES

- EWENS, W.G., 1973. Conditional diffusion processes in population genetics. *Theor. Pop. Bio.* 4, 21-30.
- FELLER, W., 1951. Diffusion processes in genetics. *Proc. 2nd Berkeley Symp. Math. Statist. and Prob.*, pp.227-246, Univ. of California Press, Berkeley, C.A.
- HALDANE, J.B.S., 1931. A mathematical theory of natural and artificial selection. VII. Selection intensity as a function of mortality rate. *Proc. Camb. Phil. Soc.* 27, 131-136.
- KEMENY, J.G. and SNELL, J.L., 1960. Finite Markov Chains. Princeton, New York.
- NARAIN, P., 1969. Response to selection in finite populations. Ph.D. thesis, Edinburgh University.
- NARAIN, P., 1970. A note on the diffusion approximation for the variance of the number of generations until fixation of a neutral mutant gene. *Genet. Res.*, Camb. 15, 251-255.
- NARAIN, P., 1971a. The use of transition probability matrices in studies on limits of response to selection. *J. Indian Soc. Agri. Statist.* 23, 48-60.
- NARAIN, P., 1971b. Average number of generations required to attain limits of genetic improvement. *SABRAO Newsletter*. 3, 135-142.
- NARAIN, P., 1974. The conditioned diffusion equation and its uses in population genetics. *J. Royal Statist. Soc. London B* 36, 258-266.
- NARAIN, P. and ROBERTSON, A., 1969. Limits and duration of response to selection in finite populations: the use of transition probability matrices. *Indian J. Heredity*, 1, 79-97.

- PURI, P.S., 1968. A study through a Markov Chain of a population undergoing certain mating systems in the presence of linkage. Contributions in Statist. and Agri. Sciences. Vol. presented to Dr. V.G. Panse at his 62nd birthday: 127-141.
- ROBERTSON, A., 1960. A theory of limits in artificial selection. Proc. Roy. Soc. London B 153, 234-249.
- ROBERTSON, A. and NARAIN, P., 1971. Survival of recessive lethals in finite populations. Theor. Pop. Bio. 2, 24-50.
- WRIGHT, S., 1931. Evolution in Mendelian populations. Genetics, 16, 97-159.