

THE USE OF TRANSITION PROBABILITY MATRICES IN STUDIES ON LIMITS OF RESPONSE TO SELECTION

PREM NARAIN

Institute of Agricultural Research Statistic, (I. C. A. R.), New Delhi-12

Introduction

A fundamental problem in population genetics is to describe the changes in the frequency of a gene over time due to systematic forces like selection, mutation and migration. When the size of the given population is very large and the individuals of the population mate at random, the change in the gene frequency is deterministic and can be easily studied by simple algebraic principles. In such cases the population ends up either with the fixation of a favoured gene or with a polymorphic state due to the balancing of various forces. However, when the population is small, consisting of a finite number of individuals, the gene frequency is also subject to fluctuations over time due to random forces created by the random sampling of gametes. The change in the gene frequency over time is then a stochastic process and can only be studied with the help of the mathematical and statistical techniques used for studying a stochastic process. If the rate of change in the gene frequency per generation is very small, the process is approximated by a continuous stochastic process with gene frequency as a random variable lying between 0 and 1. Moreover, in most of the cases encountered, the behaviour of the gene frequency in a generation depends only on its value in the preceding generation so that the process is Markovian in structure. A Markov process, continuous in gene frequency as well as in time parameter can best be studied with the help of differential equations introduced by Kolmogorov (1931). Making use of these techniques, Wright (1931) and Kimura (1957) respectively gave the concepts of distribution of gene frequency and probability of fixation of a gene. These concepts have proved fundamentals in the theory of population genetics applied to finite populations. In particular, Robertson (1960) made use of these concepts and developed a theory of limits of response to artificial selection with useful applications in animal breeding. When the changes in the gene frequency over time are not small, these cannot be described by a continuous stochastic process. The gene frequency is now a discrete random variable changing by steps between 0 and 1 depending on the population size. The changes in the gene frequency are therefore described exactly by a discrete Markov process with discrete time parameter. Such a process is often referred to as a finite Markov chain and can be studied with the help of transition probability matrices as shown by Narain (1969). Besides translating the concepts of distribution of gene frequency and fixation probability into the transition matrix approach, Narain (1969)

introduced the concept of average time taken for the fixation of a gene. A general theory of limits and duration of response to selection in finite populations was developed subsequently by Narain and Robertson (1969). In this paper an application of the theory to the case of selection at a locus with two alleles, assuming a constant population size, was also studied in detail. Later on, the effect of dominance and recessive nature of the gene on the limits and duration of response was investigated by Narain (1971). In this paper the estimates of genetic change per generation were also obtained with the help of the limits and time taken to achieve them. The transition matrix approach was further used to study the survival of recessive lethals in small populations by Narain (1969) and later by Robertson and Narain (1971). In the later paper, *Monté Carlo* methods were also used to investigate the effect of linkage on the survival of recessive lethals in small populations.

Apart from the above applications of the transition matrix approach to specific genetic problems, a rigorous theory of this approach has not been discussed so far. In this paper, therefore, the basic theory of this approach has been described briefly with special reference to the problem of genetic selection in finite populations. In addition, the theory has been applied to study the effect of linkage on the probability of fixation of a gamete in populations practising self-fertilization.

2. Transition Probability Matrix

Consider a finite population of gametes of size $2N$ 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} , the i^{th} state E_i representing the state of i A genes and $(2N-i)$ a genes. The gene frequency of A , denoted by x_i for the population in the state E_i can then take values, $x_i = i/2N$, $i=0, 1, \dots, 2N$. The states E_0 and E_{2N} represent the states of a and A genes entirely and therefore once the population assumes these states, it gets fixed for either a or A alleles, i.e., the gene frequencies are $x_0=0$ and $x_{2N}=1$ respectively. On the other hand, any state E_i , $i=1, 2, \dots, (2N-1)$, represents a state segregating for A and a genes and therefore once the population is in such a state, it has a possibility of moving from this state to any other state. Then the gene frequencies of A and a , in this population, are x_i and $(1-x_i)$ respectively. Thus a population of gametes of size $2N$ with two alleles A and a corresponds to a random walk with E_0 and E_{2N} as absorbing states and E_i , $i=1, 2, \dots, (2N-1)$ as transient states.

Let $P_{ij}(t_1, t_2)$ be the conditional probability that the population is in state E_j at time t_1 , given that it was in state E_i at time t_2 (less than t_1) i.e. it represents the probability of transition from E_i to E_j after a time (t_1-t_2) . Mathematically, this means

$$P_{ij}(t_1, t_2) = P(x \in E_j \text{ at } t_1 / x \in E_i \text{ at } t_2), t_1 > t_2 \quad (1)$$

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 then denote this probability by $P_{ij}^{(t)}$, representing the probability that the population is in state E_j at time $t + \tau$, greater than t , given that it was in state E_i at time τ for $\tau \geq 0$. This is known as t -step transition probability from E_i to E_j so that one-step transition probability can be denoted by P_{ij} . Varying i and j from 0 to $2N$ in steps of 1, we get $(2N+1) \times (2N+1)$ transition probabilities which can conveniently be represented by matrices \mathbf{P} and $\mathbf{P}(t)$ for one and t -step transitions respectively. If all the genes are either A or a then for all t , $P_{0,0}^{(t)}$ and $P_{2N,2N}^{(t)}$ will each be one and $P_{0,i}^{(t)}$, and $P_{2N,j}^{(t)}$ will each be zero for $j=1, 2, \dots, (2N-1)$. Suppose we denote the matrix of transition probabilities associated with transient states by \mathbf{Q} and $\mathbf{Q}(t)$ respectively for one and t -step transitions. Further, suppose \underline{P}_0 and $\underline{P}_0(t)$ represent, respectively for one and t -step transitions, column vectors for transitions from a transient state to E_0 . Similarly \underline{P}_{2N} and $\underline{P}_{2N}(t)$ represent the corresponding column vectors for transitions from a transient state to E_{2N} . Then \mathbf{P} and $\mathbf{P}(t)$ can be written as

$$\mathbf{P} = \begin{bmatrix} 1 & \underline{O}' & 0 \\ \underline{P}_0 & \mathbf{Q} & \underline{P}_{2N} \\ 0 & \underline{O}' & 1 \end{bmatrix} \tag{2}$$

$$\mathbf{P}(t) = \begin{bmatrix} 1 & \underline{O}' & 0 \\ \underline{P}_0(t) & \mathbf{Q}(t) & \underline{P}_{2N}(t) \\ 0 & \underline{O}' & 1 \end{bmatrix} \tag{3}$$

The elements of $\mathbf{P}(t)$ satisfy the condition

$$P_{ij}^{(t)} \geq 0 \text{ for all } i, j \tag{4}$$

$$\sum_{j=0}^{2N} P_{ij}^{(t)} = 1 \text{ for all } i \tag{5}$$

Since a transition from E_i to E_j after t steps means a transition from E_i to E_k in one step and then from E_k to E_j in $(t-1)$ steps, the probabilities of simultaneous realization of these events are

$$P_{ik} P_{kj}^{(t-1)} \text{ for } k=0, 1, 2, 3, \dots, 2N$$

Hence we have, the Chapman-Kolmogorov equation (Feller 1951)

$$P_{ij}^{(t)} = \sum_{k=0}^{2N} P_{ik} P_{kj}^{(t-1)} \tag{6}$$

The corresponding matrix equation is obtained as

$$\begin{aligned} P(t) &= P P(t-1) \\ &= P^2 P(t-2) \\ &= \dots\dots\dots \\ &= P^t \end{aligned} \quad (7)$$

Powering the P -matrix and assuming that inverse of $(I-Q)$ exists we get the following relations :

$$Q(t) = Q^t \quad (8)$$

$$\underline{P}_O(t) = (I - Q^t)(I - Q)^{-1} \underline{P}_O \quad (9)$$

$$\underline{P}_{2N}(t) = (I - Q^t)(I - Q)^{-1} \underline{P}_{2N} \quad (10)$$

3. Probability of Fixation of a Gene

Let $u_i(t)$ be the probability that at time t , and not sooner, the population with initial gene frequency of A as $i/2N$ becomes fixed for A and let $U_i(t)$ be the probability that it has become fixed for A by t^{th} generation. Then the probabilities of fixation of A with initial gene frequency, $x_i = i/2N$ is given by

$$U_i(t) = \sum_{r=1}^t u_i(r) \quad (11)$$

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

$$U_i(1) = u_i(1) = P_{i, 2N} \quad (12)$$

Now fixation at time t can take place in $(2N-1)$ mutually exclusive ways, the k^{th} 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

$$u_i(t) = \sum_{k=1}^{2N-1} P_{ik} u_k(t-1) \quad (13)$$

If we denote by $\underline{U}(t)$ and $\underline{u}(t)$ the column vectors of $U_i(t)$ and $u_i(t)$ respectively for $i=1, \dots, (2N-1)$, we can write these relations as

$$\begin{aligned}
 \underline{u}(t) &= \underline{Q}u(t-1) \\
 &= \underline{Q}^2u(t-2) \\
 &= \dots\dots\dots \\
 &= \underline{Q}^{t-1}u(1) \\
 &= \underline{Q}^{t-1}\underline{P}_{2N}
 \end{aligned} \tag{14}$$

Then

$$\begin{aligned}
 \underline{U}(t) &= \sum_{r=1}^t \underline{u}(r) \\
 &= \sum_{r=1}^t \underline{Q}^{r-1}\underline{P}_{2N} \\
 &= (\underline{I}-\underline{Q}^t)(\underline{I}-\underline{Q})^{-1}\underline{P}_{2N}
 \end{aligned} \tag{15}$$

The expression for $\underline{U}(t)$ is the same as $\underline{P}_{2N}(t)$ given by (10) showing thereby that the fixation probability by t^{th} generation can alternatively be obtained by powering the transition matrix \underline{P} , t times.

Similarly, if $\underline{l}(t)$ and $\underline{L}(t)$ denote the corresponding vectors for the fixation of gene a , we have

$$\begin{aligned}
 \underline{L}(t) &= \sum_{r=1}^t \underline{l}(r) \\
 &= \sum_{r=1}^t \underline{Q}^{r-1}\underline{P}_O \\
 &= (\underline{I}-\underline{Q}^t)(\underline{I}-\underline{Q})^{-1}\underline{P}_O
 \end{aligned} \tag{16}$$

which is the same as $\underline{P}_O(t)$ in view of (9). Also if $\underline{W}(t)$ denotes the vector of probabilities that a population with gene frequency x_i of A is still segregating for it by the t^{th} generation, we have

$$\underline{W}(t) = \underline{Q}^t \underline{e} \tag{17}$$

where \underline{e} is a column vector of unities.

As t tends to infinity, we get

$$\underline{U} = \underline{U}(\infty) = (\underline{I}-\underline{Q})^{-1}\underline{P}_{2N} \tag{18}$$

$$\underline{L} = \underline{L}(\infty) = (\underline{I}-\underline{Q})^{-1}\underline{P}_O \tag{19}$$

$$\underline{W} = \underline{W}(\infty) = \underline{O} \tag{20}$$

These relations show that ultimately the population is going to be fixed either for A or for a with fixation probabilities given by (18) and (19) respectively.

We can now express (15) and (16) as

$$\underline{U}(t) = (\mathbf{I} - \mathbf{Q}^t) \underline{U} \quad (21)$$

$$\underline{L}(t) = (\mathbf{I} - \mathbf{Q}^t) \underline{L} \quad (22)$$

4. Expected Change in the Frequency of a Gene

Let the expected frequency of A by the t^{th} generation be denoted by $E[q_i(t)]$ when the initial population had its frequency as $q_i(0) = i/2N$. The expected response in the gene frequency by the t^{th} generation is then

$$E[R_i(t)] = E[q_i(t)] - q_i(0) \quad (23)$$

Let this be represented in vector notation by

$$E[\underline{R}(t)] = E[\underline{q}(t)] - \underline{q}(0) \quad (24)$$

The expected gene frequency by the t^{th} generation, $E[q_i(t)]$ can be obtained by finding the mean of the variate $x_j = j/2N$ for the distribution given by the i^{th} row of $\mathbf{P}(t)$ i.e.

$$\begin{aligned} E[q_i(t)] &= \sum_{j=0}^{2N} P_{ij}(t) x_j \\ &= \sum_{j=1}^{2N-1} P_{ij}(t) x_j + P_{i,2N}(t) \end{aligned} \quad (25)$$

In matrix notation, this means

$$\begin{aligned} E[\underline{q}(t)] &= \mathbf{Q}^t \underline{q}(0) + \underline{P}_{2N}(t) \\ &= \mathbf{Q}^t \underline{q}(0) + \underline{U}(t) \end{aligned} \quad (26)$$

in view of (10) and (15).

If Δq_i denotes the response in the mean gene frequency due to the first generation of selection i.e. the initial response, we have

$$\begin{aligned} q_i(0) + E(\Delta q_i) &= E[q_i(1)] \\ &= \sum_{j=1}^{2N-1} P_{ij} x_j + P_{i,2N} \end{aligned} \quad (27)$$

In matrix notations, we have

$$\begin{aligned} \underline{q}(0) + E(\Delta \underline{q}) &= E[\underline{q}(1)] \\ &= \mathbf{Q} \underline{q}(0) + \underline{P}_{2N} \end{aligned} \quad (28)$$

where $E(\Delta \underline{q})$ is the vector of initial expected responses.

This can be manipulated to give

$$(\mathbf{I} - \mathbf{Q})^{-1} \underline{P}_{2N} - \underline{q}(0) = (\mathbf{I} - \mathbf{Q})^{-1} E(\Delta \underline{q}) \quad (29)$$

Then

$$\begin{aligned}
 E[\underline{R}(t)] &= \mathbf{Q}^t \underline{q}(0) + \underline{U}(t) - \underline{q}(0) \\
 &= -(\mathbf{I} - \mathbf{Q}^t) \underline{q}(0) + (\mathbf{I} - \mathbf{Q}^t)(\mathbf{I} - \mathbf{Q})^{-1} \underline{P}_{2N} \\
 &= (\mathbf{I} - \mathbf{Q}^t)[(\mathbf{I} - \mathbf{Q})^{-1} \underline{P}_{2N} - \underline{q}(0)] \\
 &= (\mathbf{I} - \mathbf{Q}^t)(\mathbf{I} - \mathbf{Q})^{-1} E(\underline{\Delta q})
 \end{aligned} \tag{30}$$

in view of (15) and (29). This shows that the expected response vector by the t^{th} generation is similar to probability fixation vector given by (15) with the difference that \underline{P}_{2N} in (15) is replaced by $E(\underline{\Delta q})$ in (30). It is also interesting to note that $E(\underline{\Delta q})$ and \underline{P}_{2N} give respectively the initial expected responses and initial step fixation probabilities.

As t tends to infinity, we get

$$E(\underline{R}) = E[\underline{R}(\infty)] = (\mathbf{I} - \mathbf{Q})^{-1} E(\underline{\Delta q}) \tag{31}$$

$$E[\underline{R}(t)] = (\mathbf{I} - \mathbf{Q}^t) E(\underline{R}) \tag{32}$$

which are similar to (18) and (21) respectively. In view of (29) and (18), we also have

$$E(\underline{R}) = \underline{U} - \underline{q}(0) \tag{33}$$

This shows that the expected limit of response to selection can, otherwise, be obtained by subtracting the initial gene frequency from the eventual probability of fixation. But it must be noted that the expected response by the t^{th} generation cannot, similarly, be obtained *i. e.* $E[\underline{R}(t)]$ is not equal to $\underline{U}(t) - \underline{q}(0)$ in view of (30). When there is no selection, $E(\underline{\Delta q}) = \underline{0}$ and (31) and (33) give the expected result, $\underline{U} = \underline{q}(0)$ *i. e.* there is no ultimate response, making the fixation probability of a gene equal to its initial gene frequency.

5. Calculation of the probability of fixation and the expected change in the gene frequency

It is apparent from the above derivations that the probability of fixation of a gene and the expected change in the gene frequency by a given number of generations as well as in the limit can be obtained by performing matrix operations on the transition matrix \mathbf{Q} . However, before these operations, the transition probability P_{ij} is to be specified with the help of the knowledge of the genetic situation involved. Thereafter, matrix functions are to be evaluated either numerically or by analytical methods, for a given population size. For instance, Narain (1969) and Narain and Robertson (1969) used binomial transition probabilities for describing the random sampling of gametes. In some cases they evaluated matrix functions numerically on the computer for a given population size. In some cases, however, the results were also obtained with the help of analytical techniques by using the eigen-roots and eigen-vectors of \mathbf{Q} . In this section we outline the analytical approach which is applied to a genetic problem studied in the next section,

Let the eigen-roots of \mathbf{Q} , a $k \times k$ matrix, obtained by solving the characteristic equation

$$|\mathbf{Q} - \lambda \mathbf{I}| = 0 \quad (34)$$

be given by $\lambda_1, \lambda_2, \dots, \lambda_k$. Also, let \underline{x}_i and \underline{y}_i be the right and left column vectors respectively corresponding to the root λ_i for $i=1, 2, \dots, k$. The spectral set of matrices for \mathbf{Q} are therefore

$$\mathbf{H}_i = \underline{x}_i \underline{y}_i', \quad i=1, 2, \dots, k. \quad (35)$$

Any function of \mathbf{Q} , $f(\mathbf{Q})$, can then be expressed as

$$\begin{aligned} f(\mathbf{Q}) &= \sum_{i=1}^k f(\lambda_i) \mathbf{H}_i \\ &= \sum_{i=1}^k f(\lambda_i) \underline{x}_i \underline{y}_i' \end{aligned} \quad (36)$$

The formulae for calculating the probability of fixation and the expected change in the gene frequency are then given by

$$\underline{U} = \sum_{i=1}^k (1-\lambda_i)^{-1} \underline{x}_i \underline{y}_i' P_{2N} \quad (37)$$

$$\underline{U}(t) = \sum_{i=1}^k (1-\lambda_i^t) \underline{x}_i \underline{y}_i' \underline{U} \quad (38)$$

$$E(\underline{R}) = \sum_{i=1}^k (1-\lambda_i)^{-1} \underline{x}_i \underline{y}_i' E(\underline{\Delta q}) \quad (39)$$

$$E[\underline{R}(t)] = \sum_{i=1}^k (1-\lambda_i^t) \underline{x}_i \underline{y}_i' E(\underline{R}) \quad (40)$$

6. Effect of linkage on the probability of fixation of a gamete in selfed populations

The case of selfed populations corresponds, in the above discussion, to the situation when $N=1$. The population gets subdivided into lines from each of which two gametes are chosen to form one mature individual only. If we consider a single locus with two alleles A and a , there would be three types of lines. Two of these would have only homozygotes AA and aa respectively and each would occur with frequency $1/4$. The third type would have only heterozygotes Aa and would occur with frequency $1/2$. It is obvious that the first two types correspond to the absorbing states with frequency of A as 1 and 0 respectively whereas the third type

corresponds to the transient state with frequency of A as $1/2$. The matrix \mathbf{P} , then assumes the form

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} \quad (41)$$

In this case the matrix \mathbf{Q} is a scalar quantity equal to $1/2$. It is easy to see that (18) and (21) give, for this case, the following results :

$$\text{Probability of fixation of } A \text{ in the limit} = \frac{1}{2} \quad (42)$$

$$\text{Probability of fixation of } A \text{ by the } t^{\text{th}} \text{ generation} = \frac{1}{2}(1 - \frac{1}{2}^t) \quad (43)$$

With two loci each with two alleles $A-a$ and $B-b$ respectively, there would be four gametes AB, Ab, aB, ab . Since two gametes per line are chosen, there would be ten types of lines. Four of these would have only homozygotes respectively as $AB/AB, Ab/Ab, aB/aB$ and ab/ab . These would represent absorbing states. Another four types of lines would contain only single heterozygotes respectively as $AB/Ab, AB/aB, Ab/ab, aB/ab$. The corresponding states would be transient. The remaining two types of lines would have only double heterozygotes of the types AB/ab (coupling) and Ab/aB (repulsion) respectively and would also correspond to transient states. The linkage between genes at the two loci would affect the contributions of double heterozygotes only. If we denote the probability of crossing-over between two loci by r , with $s=1-r$, the 10 states of the system can be classified according to the distribution amongst the four types of gametes as below :

	AB	Ab	aB	ab
E_1	1	0	0	0
E_2	0	1	0	0
E_3	0	0	1	0
E_4	0	0	0	1
E_5	$1/2$	$1/2$	0	0
E_6	$1/2$	0	$1/2$	0
E_7	0	$1/2$	0	$1/2$
E_8	0	0	$1/2$	$1/2$
E_9	$s/2$	$r/2$	$r/2$	$s/2$
E_{10}	$r/2$	$s/2$	$s/2$	$r/2$

Journal of the Indian Society of Agricultural Statistics

Vol. XXIII]

SILVER JUBILEE NUMBER

DECEMBER 1971

[No. 2

CONTENTS

1. Requirements of statistical data for agricultural planning <i>by J. S. Sarma</i>	...	1
2. The statistician, the computer and experimentation <i>by D. J. Finney</i>	...	7
3. Invariant tests of discriminant coefficients in classification problems <i>by N. Giri</i>	...	28
4. Efficiency of T-statistics for testing two normal samples <i>by P. V. K. Iyer, B. N. Singh and M. S. Holla</i>	...	36
5. The use of transition probability matrices in studies on limits of response to selection <i>by Prem Narain</i>	...	48
6. On a model for rank analysis in fractional paired comparisons <i>by S. C. Rai</i>	...	61
7. Some thoughts on foundations of survey sampling <i>by J. N. K. Rao</i>	...	69
8. Use of transformation in sampling <i>by S. Mohanty and M. N. Das</i>	...	83
9. The Canadian water-fowl mail surveys—design, results and their reliability <i>by A. R. Sen</i>	...	88
10. Relative efficiencies of some alternative procedures in two-stage sampling on successive occasions <i>by O. P. Kathuria and D. Singh</i>	...	101
11. Determination of optimum strata boundaries <i>by Ravindra Singh</i>	...	115
12. On the comparison of certain classes of balanced 28 fractional designs of resolution V with respect to the trace criterion <i>by J. N. Srivastava and D. V. Chopra</i>	...	124
13. Construction and analysis of designs $qx 3^2$ in six blocks <i>by P. R. Sreenath</i>	...	132
14. Linking of bio-assay contrasts and factorial contrasts <i>by A. C. Kulshreshtha</i>	...	136
15. The dynamics of populations under partial inbreeding <i>by Gauri L. Ghai and Oscar Kempthorne</i>	...	142
16. Asian Statistical Institute <i>by K. R. Nair and M. N. Murty</i>	...	152
17. Note on the transformation of variables in simultaneous equations systems <i>by Gerhard Tintner and Gopal Kadekodi</i>	...	163
18. Price of milk and feed cost as factors influencing increased milk production <i>by K. C. Raut and T. Jacob</i>	...	174
19. A note on an approximation of the F distribution <i>by C. G. Khatri</i>	...	178
20. Assessment of mean-depth of precipitation for a catchment <i>by G. A. Kulkarni and G. D. Khataavkar</i>	181
21. Agricultural Census in India <i>by I. J. Naidu</i>	186
22. List of members of the Society	...	193
23. Hindi Supplement		i-xiii

JOURNAL of the Indian Society of Agricultural Statistics

Vol. XXIII]

JUNE, 1971

[No. 1

CONTENTS

Obituary (Prof. D.R. Gadgil)

1. The story of green revolution in India ... 1
By C. Subramaniam
2. Some rethinking on statistical training and research ... 11
in India
by P.K. Bose
3. Missing and mixed up observations in analysis ... 25
of covariance
by D. Raghavarao & Har Vishan Lal
4. Comparisons of estimates in two-stage sampling ... 31
on successive occasions
by O.P. Kathuria and D. Singh
5. Ranking in fractional triad comparisons ... 52
by S.C. Rai
6. On construction of a series of balanced incomplete ... 62
block designs
by Basant Lal
7. Intercensal estimates of livestock numbers ... 67
by V.N. Amble & M.S. Avadhani
8. Symposium on model sampling experiments ... 84
9. Abstract of papers presented at the Twenty-fourth ... 91
Annual Conference of the Society
10. Review of the activities of the Society for the ... 116
year 1969-70
11. Twenty-fourth Annual Conference ... 1 0
12. Hindi supplement ... *i-iv*

The P -matrix, then, takes the form :

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & \underline{0'} \\ 0 & 1 & 0 & 0 & \underline{0'} \\ 0 & 0 & 1 & 0 & \underline{0'} \\ 0 & 0 & 0 & 1 & \underline{0'} \\ \underline{P_{AB}} & \underline{P_{Ab}} & \underline{P_{aB}} & \underline{P_{ab}} & \underline{Q} \end{bmatrix} \quad (44)$$

In this matrix, $\underline{0'}$ is a 1×6 row vector of zeros. $\underline{P_{AB}}$, $\underline{P_{Ab}}$, $\underline{P_{aB}}$, and $\underline{P_{ab}}$, are 6×1 column vectors given by

$$\underline{P'_{AB}} = (1/4, 1/4, 0, 0, s^2/4, r^2/4) \quad (45)$$

$$\underline{P'_{Ab}} = (1/4, 0, 1/4, 0, r^2/4, s^2/4) \quad (46)$$

$$\underline{P'_{aB}} = (0, 1/4, 0, 1/4, r^2/4, s^2/4) \quad (47)$$

$$\underline{P'_{ab}} = (0, 0, 1/4, 1/4, s^2/4, r^2/4) \quad (48)$$

Further, Q is a 6×6 matrix given by

$$Q = \begin{bmatrix} 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 0 \\ rs/2 & rs/2 & rs/2 & rs/2 & s^2/2 & r^2/2 \\ rs/2 & rs/2 & rs/2 & rs/2 & r^2/2 & s^2/2 \end{bmatrix} \quad (49)$$

The Q -matrix is found to yield the 6 roots $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ and λ_6 given by

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 1/2 \quad (50)$$

$$\lambda_5 = (1 - 2rs)/2 \quad (51)$$

$$\lambda_6 = (1 - 2r)/2 \quad (52)$$

The matrix X of the right-hand column vectors x_i corresponding to λ_i for $i=1, 2, 3, 4, 5, 6$, is derived from the results given by Puri (1968) and is given by

$$X = (\underline{x}_1, \underline{x}_2, \underline{x}_3, \underline{x}_4, \underline{x}_5, \underline{x}_6) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1/2 & 1/2 & 1/2 & 1/2 & 1/\sqrt{2} & 1/\sqrt{2} \\ 1/2 & 1/2 & 1/2 & 1/2 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \quad (53)$$

Similarly, the matrix Y of the left-hand row vectors y'_i corresponding to the six roots is given by

$$Y = (y'_1, y'_2, y'_3, y'_4, y'_5, y'_6) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ -1/\sqrt{2} & -1/\sqrt{2} & -1/\sqrt{2} & -1/\sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \quad (54)$$

With the help of vectors x_i and y'_i , the spectral set of the matrices H_i , $i=1, 2, \dots, 6$ given by (35) are obtained. These matrices, together with the six-roots, provide with the elements of the matrix $(I-Q)^{-1}$ since

$$(I-Q)^{-1} = \sum_{i=1}^6 (1-\lambda_i)^{-1} H_i = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 1-w & 1-w & 1-w & 1-w & w+v & w-v \\ 1-w & 1-w & 1-w & 1-w & w-v & w+v \end{bmatrix} \quad (55)$$

where $w=1/(1+2rs)$ and $v=1/(1+2r)$.

If the vectors of probabilities of fixation of gametes AB , Ab , aB and ab are denoted respectively by $\underline{U_{AB}}$, $\underline{U_{Ab}}$, $\underline{U_{aB}}$ and $\underline{U_{ab}}$, then, in view of (37) and (45) to (48), we get

$$\underline{U_{AB}} = (1/2 \quad 1/2 \quad 0 \quad 0 \quad v/2 \quad rv)' \quad (56)$$

$$\underline{U_{Ab}} = (1/2 \quad 0 \quad 1/2 \quad 0 \quad rv \quad v/2)' \quad (57)$$

$$\underline{U_{aB}} = (0 \quad 1/2 \quad 0 \quad 1/2 \quad rv \quad v/2)' \quad (58)$$

$$\underline{U_{ab}} = (0 \quad 0 \quad 1/2 \quad 1/2 \quad v/2 \quad rv)' \quad (59)$$

It is confirmed from the above results that the linkage can only have its effect when the population is initially either in state E_9 (coupling phase) or in state E_{10} (repulsion phase). In each of these two situations, the probabilities of fixation of the four types of gametes depend on the recombination fraction r , as shown by the last two elements of each of the vectors given by (56)–(59). As expected, the probability of fixation of gametes AB is the same as that of ab . Similarly, the probability of fixation of Ab is the same as that of aB . The effect of linkage is to increase the probability of fixation of a coupled gamete (AB or ab) if the initial population consists of a double heterozygote in coupling phase. On the other hand with repulsion phase the probability is reduced. For repulsion gametes (Ab or aB), the probability is increased when the initial population consists of a double heterozygote in repulsion phase but decreased with coupling phase of linkage.

7. Summary

A rigorous theory of the transition matrix approach for studying the change in the frequency of a gene in finite populations is developed. The probability of fixation of a gene and the expected change in the gene frequency by a given number of generations as well as in the limit are expressed as functions of the transition probability matrix. The analytical as well as the numerical procedures for the calculation of these quantities are outlined. The theory is applied to study the effect of linkage on the probability of fixation of a gamete in populations practising self-fertilization. It is found that linkage increases or decreases the probability of fixation of a coupled gamete according as the initial population consists of a coupling or repulsion heterozygote respectively.

REFERENCES

- Feller, W. (1951). Diffusion processes in genetics. *Proc. 2nd Berkeley Sym. Mathematical Statistics and Probability*. pp. 227-246. University of California Press.
- Kimura, M. (1957). Some problems of stochastic processes in genetics. *Ann. Math. Statist.* 28 : 882-901.
- Kolmogorov, A. (1931) Über die analytischen Methoden in der Wahrscheinlich keitsrechnung. *Math. Ann.* 104 : 415-458.
- Narain, P. (1969). Response to selection in finite populations. *Ph.D. thesis*, Edinburgh University.
- Narain, P. (1971). Average number of generations required to attain limits of genetic improvement. *S.A.B.R.A.O. Newsletter*, 3(2) : 135-142.
- Narain, P. and Robertson, A. (1969). Limits and duration of response to selection in finite populations : the use of transition probability matrices. *The Indian Jour. of Heredity*. 1 : 79-97.
- Puri, Prem S. (1968). A study through a Markov chain of a population undergoing certain mating systems in the presence of linkage. *Contributions in Statistics 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. *Theo. Population Biol.* 2 : 24-50.
- Takacs, L. (1960). *Stochastic Processes : Problems and Solutions*. Methuen, London. Wiley, New York.
- Wright, S. (1931). Evolutions in Mendelian populations. *Genetics*, 16 : 97-159.