MARKOV CHAINS WITH ABSORBING STATES: A GENETIC EXAMPLE[1]

By G. A. Watterson

Virginia Polytechnic Institute

1. Summary and introduction. If a finite Markov chain (discrete time, discrete states) has a number of absorbing states, one of these will eventually be reached. In this paper are given theoretical formulae for the probability distribution, its generating function and moments of the time taken to first reach an absorbing state, and these formulae are applied to an example taken from genetics.

While first passage time problems and their solutions are known for a wide variety of Markov chain processes (e.g., [2], [7], [4]), the theory seems not to have been used in population genetics. Suppose a genetic population consists of a constant number of individuals and the state of the population is defined by the numbers of the various genotypes existing at a given time. Then if mutation is absent, all individuals will eventually become of the same genotype because of random influences such as births, deaths, mating, selection, chromosome breakages and recombinations. The population behavior may in some circumstances be approximated by a Markov chain with absorbing states.

In Section 2, two alternative approaches are given for the theoretical determination of absorption time properties, using well known techniques. In Section 3, the consequences of the theoretical results are investigated for a particular population model introduced by Moran [9], [10], and explicit expressions for the distribution of the gene fixation time are obtained in terms of Chebyshev's orthogonal polynomials. The derivation requires finding the pre- and post-eigenvectors of the matrix of transition probabilities, and an incidental by-product is the proof of certain identities for the orthogonal polynomials.

The material presented in Section 2 and Section 3 is obtained by exact methods. In Section 4, the Fokker-Planck diffusion equation is used to obtain approximate results, and these are compared with those of the exact theory to ascertain the accuracy of the diffusion approximation.

2. Markov chains with absorbing states.

(a) *Arbitrary initial state.* Consider a Markov chain with variable $x(t)$, which at time t ($t = 0, 1, 2, \cdots$) can be in any of the states $0, 1, 2, \cdots, M$. Let

$$P_{ij} = \Pr\{x(\tau) = j \mid x(\tau - 1) = i\} \tag{1}$$

be the unit-time transition probabilities, and write

$$P_{ij}^{(t)} = \Pr\{x(t + \tau) = j \mid x(\tau) = i\}, \quad t, \tau = 0, 1, 2, \cdots. \tag{2}$$

Then, if $\mathbf{P}$ is the matrix of elements P_{ij}, the elements of $\mathbf{P}^t$ are the t-step transition probabilities (2).

Received September 7, 1960; revised February 13, 1961.

[1] The major portion of this work was completed while the author was a research scholar at the Australian National University.

We assume that the states 0 and M are absorbing, and that the states 1, 2, $\cdots$, $M - 1$ are transient. The following theory could be adapted to the case with more (or fewer) absorbing states, but the application to genetics makes the specific case important. Therefore we have

$$(3)\quad \begin{aligned} &P_{00} = P_{MM} = 1, \\ &P_{01} = P_{02} = \cdots = P_{0M} = P_{M0} = P_{M1} = \cdots = P_{M\,M-1} = 0. \end{aligned}$$

Let T_i be the time taken for the chain to first reach one or other absorbing state, given the initial state $x(0) = i$, and write $S_i^{(t)}$ for the probability that $T_i = t$. Clearly

$$(4)\quad S_i^{(t)} = P_{i0}^{(t)} + P_{iM}^{(t)} - P_{i0}^{(t-1)} - P_{iM}^{(t-1)}, \qquad t = 1, 2, \cdots,$$

but in particular,

$$(5)\quad S_0^{(0)} = S_M^{(0)} = 1, \qquad S_0^{(t)} = S_M^{(t)} = 0, \qquad t \geqq 1,$$

and for $i \neq 0$ or M,

$$(6)\quad S_i^{(0)} = 0, \qquad S_i^{(1)} = P_{i0} + P_{iM}.$$

If we write $\mathbf{S}^{(t)}$ as the column vector whose transpose is

$$\mathbf{S}^{(t)\prime} = (S_0^{(t)}, S_1^{(t)}, \cdots, S_M^{(t)}),$$

and, in particular, from (5) and (6),

$$(7)\quad \begin{aligned} \mathbf{S}^{(0)\prime} &= (1, 0, 0, \cdots, 0, 1), \\ \mathbf{S}^{(1)\prime} &= (0, P_{10} + P_{1M}, P_{20} + P_{2M}, \cdots, P_{M-1\,0} + P_{M-1\,M}, 0), \end{aligned}$$

then (4) becomes

$$(8)\quad \mathbf{S}^{(t)} = (\mathbf{P}^t - \mathbf{P}^{t-1})\mathbf{S}^{(0)} = \mathbf{P}^{t-1}\mathbf{S}^{(1)}, \qquad t = 1, 2, 3, \cdots.$$

The calculation of absorption probabilities by (8) will generally be a difficult task unless the eigenvalues and eigenvectors of $\mathbf{P}$ are known. If they are, however, we can proceed as follows. Let λ_j be the jth eigenvalue of $\mathbf{P}$, and $\mathbf{K}_j$ the corresponding post-eigenvector. Then

$$\mathbf{PK}_j = \mathbf{K}_j\lambda_j, \qquad j = 0, 1, \cdots, M;$$

that is $\mathbf{PK} = \mathbf{K\Lambda}$ where $\mathbf{K} = (\mathbf{K}_0, \mathbf{K}_1, \cdots, \mathbf{K}_M)$, $\mathbf{\Lambda} = (\delta_{ij}\lambda_j)$. While the columns of $\mathbf{K}$ are the post-eigenvectors, the rows of $\mathbf{K}^{-1}$ are the pre-eigenvectors, and we have $\mathbf{P} = \mathbf{K\Lambda K}^{-1}$ or, more generally,

$$(9)\quad \mathbf{P}^{t-1} = \mathbf{K\Lambda}^{t-1}\mathbf{K}^{-1}, \qquad t = 1, 2, \cdots,$$

where $\mathbf{\Lambda}^{t-1}$ is a diagonal matrix with elements $\lambda_j^{t-1}(j = 0, 1, \cdots, M)$ in the diagonal. Thus substituting (9) into (8) gives

$$(10)\quad \mathbf{S}^{(t)} = \mathbf{K\Lambda}^{t-1}\mathbf{K}^{-1}\mathbf{S}^{(1)}, \qquad t = 1, 2, \cdots.$$

At least in theory (10) gives the distributions we seek.

(b) *Transient initial state.* An alternative approach can be made by assuming that $x(0) = i$ is not absorbing. Write $\mathbf{P}_\triangle$[2] for the matrix obtained by ignoring the first and last rows and columns of $\mathbf{P}$; $\mathbf{P}_\triangle$ is not stochastic since some non-zero elements have been removed from the stochastic matrix $\mathbf{P}$. Further, with the notation

$$\mathbf{S}_\triangle^{(t)'} = (S_1^{(t)}, S_2^{(t)}, \cdots, S_{M-1}^{(t)}),$$

we find from (6) that

$$\begin{aligned} \mathbf{S}_\triangle^{(0)'} &= (0, 0, \cdots, 0), \\ \mathbf{S}_\triangle^{(1)'} &= (P_{10} + P_{1M}, P_{20} + P_{2M}, \cdots, P_{M-1\,0} + P_{M-1\,M}), \end{aligned} \tag{11}$$

that is, $\mathbf{S}_\triangle^{(1)} = (\mathbf{I}_\triangle - \mathbf{P}_\triangle)\mathbf{1}_\triangle$, where $\mathbf{I}_\triangle$ is the unit matrix, $\mathbf{1}_\triangle' = (1, 1, 1, \cdots, 1)$, both of order $M - 1$.

Corresponding to (8) we then have

$$\mathbf{S}_\triangle^{(t)} = \mathbf{P}_\triangle^{t-1}\mathbf{S}_\triangle^{(1)} = \mathbf{P}_\triangle^{t-1}(\mathbf{I}_\triangle - \mathbf{P}_\triangle)\mathbf{1}_\triangle. \tag{12}$$

From this, an equation analogous to (10) could be written down, but we will not require it in the sequel.

For the probability generating function, we write

$$\mathbf{G}_\triangle(z) = \sum_{t=0}^{\infty} z^t \mathbf{S}_\triangle^{(t)} = \sum_{t=1}^{\infty} z^t \mathbf{S}_\triangle^{(t)},$$

which by (12) is

$$\begin{aligned} \mathbf{G}_\triangle(z) &= z(\mathbf{I}_\triangle - \mathbf{P}_\triangle)\mathbf{1}_\triangle + z\sum_{t=2}^{\infty} z^{t-1}\mathbf{S}_\triangle(t) \\ &= z(\mathbf{I}_\triangle - \mathbf{P}_\triangle)\mathbf{1}_\triangle + z\sum_{t=2}^{\infty} (z\mathbf{P}_\triangle)^{t-1}(\mathbf{I}_\triangle - \mathbf{P}_\triangle)\mathbf{1}_\triangle \\ &= z(\mathbf{I}_\triangle - \mathbf{P}_\triangle)\mathbf{1}_\triangle + z^2\mathbf{P}_\triangle(\mathbf{I}_\triangle - z\mathbf{P}_\triangle)^{-1}(\mathbf{I}_\triangle - \mathbf{P}_\triangle)\mathbf{1}_\triangle \\ &= (z^{-1}\mathbf{I}_\triangle - \mathbf{P}_\triangle)^{-1}(\mathbf{I}_\triangle - \mathbf{P}_\triangle)\mathbf{1}_\triangle. \end{aligned} \tag{13}$$

Although (13) does not involve knowledge of the eigenvectors of $\mathbf{P}$, the resolvent $(z^{-1}\mathbf{I}_\triangle - \mathbf{P}_\triangle)^{-1}$ must be known for this to be a useful alternative. In the example of Section 3, we shall meet a case where the resolvent is known at $z = 1$, and this is sufficient to calculate moments.

In [5], Karlin and McGregor have discussed the problem of random walks with "ignored" absorbing states. The transition matrix for the transient states, our $\mathbf{P}_\triangle$ above, is assumed to have the Jacobi form with

$$P_{ij} = 0 \text{ for } |i - j| > 1. \tag{14}$$

[2] In what follows, the subscript $\triangle$ will distinguish $(M-1)$-order matrices and vectors, got by deleting the elements for states 0 and M, from the corresponding $(M+1)$-order quantities.

Our genetics example in Section 3 is of this type, but does not appear to yield to their methods.

3. Application to a genetic population model.

(a) *The model.* We consider a population model in which there are M individuals, each being one or other of two haploid genotypes. The birth-death model postulates that at each unit of time, one individual is chosen at random to die, and is replaced by a new individual whose genotype is determined at random from those existing before the death. Thus the number of individuals of a given genotype—the state of the population—can take any of the values 0, 1, 2, $\cdots$, M, and can change by at most unity during one birth-death event. This model was introduced by Moran [9], and further discussed by him, [10]. Actually, Moran considered as well the more general case when gene mutation was allowed. Here, as mutation is assumed absent, there is no source for new genes, and once all individuals are of the same genotype the population state remains unchanged thereafter.

The transition probabilities are (see [9] with a new notation)

$$
\begin{aligned}
P_{ij} &= 0, &\text{if } |i - j| > 1,\\
P_{ii-1} &= i(M - i)M^{-2} &\\
P_{ii} &= 1 - 2i(M - i)M^{-2} &\\
P_{ii+1} &= i(M - i)M^{-2}. &
\end{aligned}
\tag{15}
$$

The states 0 and M are absorbing, those in-between are transient.

(b) *Known results.* Hannan, in an appendix to [9], has proved the following theorem, expressed here in our notation.

THEOREM 1. *Transforming the matrix* $\mathbf{P}$ *of* (15) *by the matrix* $\mathbf{R}$, *where* $\mathbf{R}$ *has the typical element* $R_{ij} = \binom{i}{j}$ *and* $\mathbf{R}^{-1}$ *has the typical element* $(-1)^{i+j}\binom{i}{j}$, $i, j = 0, 1, \cdots, M$, *then* $\mathbf{R}^{-1}\mathbf{PR}$ *has non-zero terms only in the leading and first super diagonals. The ith row is*

$$(0, 0, \cdots, 0, 1 - i(i-1)M^{-2}, -i(M-i)M^{-2}, 0, 0, \cdots, 0), \tag{16}$$

the quantity $1 - i(i-1)M^{-2}$ *in the diagonal position is the ith eigenvalue of* $\mathbf{P}$.

Moran [10] stated the following results, again expressed here in our notation.

THEOREM 2. *If* $\mathbf{K} = (\mathbf{K}_0, \mathbf{K}_1, \cdots, \mathbf{K}_M)$ *is a matrix of eigenvectors* $\mathbf{K}_j' = (K_{0j}, K_{1j}, \cdots, K_{Mj})$ *such that*

$$\mathbf{K}^{-1}\mathbf{PK} = \mathbf{\Lambda} = (\delta_{ij}[1 - i(i-1)M^{-2}]),$$

then, apart from normalizing constants,

(i) $K_{i0} = 1$, $K_{i1} = i$, $K_{i2} = i(M - i)$, $K_{i3} = i(M - i)(M - 2i)$, $K_{i4} = i(M - i)(M^2 - 5Mi + 5i^2 + 1)$, $i = 0, 1, \cdots, M$.

(ii) *The jth element of the ith prevector,* K^{ij} *say, is proportional to* $j^{-1}(M - j)^{-1}K_{ji}$, $\quad j = 1, 2, \cdots, M - 1, \quad i = 0, 1, \cdots, M$.

(iii) *The generating function* $\sum_{j=0}^{M} K^{ij} z^j$ *satisfies a particular case of Heun's differential equation.*

(c) *New results.* In Theorem 2, Moran has given some of the early eigenvectors of the matrix **P**. Theorem 1, however, can be used to obtain an explicit expression for all the post-eigenvectors. Suppose that **W** is a matrix such that

$$\mathbf{R}^{-1}\mathbf{P}\mathbf{R}\mathbf{W} = \mathbf{W}\mathbf{\Lambda}.$$

Then **RW** has columns which are the required post-eigenvectors. By using (16) and the known eigenvalues, it is seen that the elements of **W** must satisfy the difference equation

$$i(M - i)W_{i+1\, j} = [j(j-1) - i(i-1)]W_{ij}\,.$$

One solution is

$$W_{00} = 1, \qquad W_{i0} = 0, \qquad\qquad i = 1, 2, \cdots, M,$$

and for $j \geqq 1$,

$$W_{0j} = 0, \qquad W_{1j} = 1, \qquad W_{ij} = \prod_{k=1}^{i-1} [j(j-1) - k(k-1)]k^{-1}(M-k)^{-1},$$
$$i = 2, 3, \cdots, M.$$

Hence we have

THEOREM 3. *The post-eigenvectors* $\mathbf{K}_j$ *of* **P** *have elements* K_{ij} *proportional to*

$$\sum_{k=0}^{M} R_{ik} W_{kj} = \sum_{k=0}^{i} \binom{i}{k} W_{kj}\,.$$

This result would be sufficient to obtain explicit expressions for the probability of first absorption at time t. However, the resulting expressions are rather complicated; luckily a different approach leads to tractible algebra. Consider the Chebyshev orthogonal polynomials defined by

$$\xi_j(x) = c_j\, \Delta^j \binom{x}{j}\binom{x-M}{j}, \qquad\qquad j = 0, 1, \cdots, M-1, \tag{17}$$

where

$$\Delta f(x) = f(x+1) - f(x),$$
$$\Delta^j f(x) = \sum_{k=0}^{j} (-1)^k \binom{j}{k} f(x + j - k),$$

and

$$c_j = j!(2j+1)^{\frac{1}{2}}[M(M^2 - 1^2)(M^2 - 2^2) \cdots (M^2 - j^2)]^{-\frac{1}{2}}.$$

Then $\xi_j(x)$ is a polynomial of degree j in x, and the set is orthogonal in the sense that

$$\sum_{k=0}^{M-1} \xi_i(k)\xi_j(k) = \delta_{ij}\,, \quad i, j = 0, 1, 2, \cdots, M-1, \tag{18}$$

see [1], p. 223. In what follows it will be convenient to use the conventions

$$\xi_j(-1) = 0, \qquad j = 0, 1, \cdots, M-1 \tag{19}$$

and to introduce the new function

$$\xi_{-1}(k) = \begin{cases} 0 & \text{if} \quad k = 0, 1, \cdots, M-1, \\ 1 & \text{if} \quad k = -1. \end{cases} \tag{20}$$

From (18), (19) and (20), it is clear that the augmented set $\{\xi_j(k)\}$ $j = -1, 0, 1, \cdots, M-1$ is orthogonal over $k = -1, 0, 1, \cdots, M-1$.

The non-trivial values of $\xi_j(k)$ have been tabulated in [11] for $M = 3(1)52$, $j = 1(1)6$, and references are given there to more extensive tabulations; for each j, the values of $\xi_j(k)$ are multiplied by the smallest constant to make the tabulated entries integers.

The functions $\xi_j(x)$ satisfy the difference equation

$$(x+2)(x-M+2)\Delta^2\xi_j(x) + [2x - M + 3 - j(j+1)]\Delta\xi_j(x) - j(j+1)\xi_j(x) = 0,$$

see [1], p. 223, and this may be written

$$j(j+1)\xi_j(x+1) = \Delta[(x+1)(x-M+1)\Delta\xi_j(x)].$$

Summing over the integers, and renaming the variables, we get

$$(j-1)j\sum_{k=0}^{i} \xi_{j-1}(k-1) = i(M-i)[\xi_{j-1}(i-1) - \xi_{j-1}(i)], \qquad i, j = 0, 1, \cdots, M, \tag{21}$$

where the conventions (19) and (20) have been used.

We are now in a position to prove

THEOREM 4. *For the matrix* $\mathbf{P}$ *defined in* (15),

(i) *The eigenvalues are*

$$\lambda_j = 1 - j(j-1)M^{-2}, \qquad j = 0, 1, \cdots, M. \tag{22}$$

(ii) *The post-eigenvectors are the columns* $\mathbf{K}_j$ *of the matrix*

$$\mathbf{K} = (\mathbf{K}_0, \mathbf{K}_1, \cdots, \mathbf{K}_M) = \mathbf{C}\mathbf{\Xi}, \tag{23}$$

where

$$\mathbf{C} = \begin{bmatrix} 1 & & & & & \\ 1 & 1 & & & 0 & \\ 1 & 1 & 1 & & & \\ 1 & 1 & 1 & 1 & & \\ \cdot & & & & \cdot & \\ \cdot & & & & & \cdot \\ \cdot & & & & & & \cdot \\ 1 & 1 & 1 & 1 & \cdots & 1 \end{bmatrix},$$

and $\mathbf{\Xi}$ *has* $\xi_{j-1}(i-1)$ *in the* (i,j) *position,* $i, j = 0, 1, \cdots, M$.

(iii) *The pre-eigenvectors are the rows of the matrix*

$$\mathbf{K}^{-1} = \mathbf{\Xi}'\mathbf{C}^{-1}, \tag{24}$$

where $\mathbf{\Xi}'$ *is the transpose of* $\mathbf{\Xi}$, *and*

$$\mathbf{C}^{-1} = \begin{bmatrix} 1 & & & & & & \\ -1 & 1 & & & & & \\ & -1 & 1 & & 0 & & \\ & & -1 & 1 & & & \\ & & & \cdot & \cdot & & \\ & 0 & & & \cdot & \cdot & \\ & & & & & \cdot & \cdot \\ & & & & & -1 & 1 \end{bmatrix}.$$

Proof: Parts (ii) and (iii) of the theorem are either true or false together, because $\mathbf{\Xi}$ is an orthogonal matrix, $\mathbf{C}^{-1}$ has the stated form, and the inverse of the post-eigenvector matrix gives the pre-vectors. It will therefore be sufficient to prove that (i) and (ii) are correct, and this is done by proving

$$\mathbf{PK} = \mathbf{K\Lambda}$$

for the particular definitions used here. Write g_{ij} and h_{ij} for the typical elements of the left- and right-hand sides respectively; then we have to show that $g_{ij} = h_{ij}$ for $i, j = 0, 1, \cdots, M$.

Multiplying out $\mathbf{PK} = \mathbf{PC\Xi}$ we find

$$g_{ij} = \sum_{k=0}^{M} [\xi_{j-1}(k-1) \sum_{n=k}^{M} P_{in}],$$

and with the substitution for P_{in} from (15) we get

$$g_{ij} = \sum_{k=0}^{i-1} \xi_{j-1}(k-1) + \left[1 - \frac{i}{M}\left(1 - \frac{i}{M}\right)\right] \xi_{j-1}(i-1) + \frac{i}{M}\left(1 - \frac{i}{M}\right) \xi_{j-1}(i).$$

Again, multiplying out $\mathbf{K\Lambda} = \mathbf{C\Xi\Lambda}$ we have

$$h_{ij} = \lambda_j \sum_{k=0}^{i} \xi_{j-1}(k-1) = \left[1 - \frac{j(j-1)}{M^2}\right] \sum_{k=0}^{i} \xi_{j-1}(k-1).$$

The equality of g_{ij} and h_{ij} follows from (21), and holds for all relevant i, j. This completes the proof of the theorem.

Corollary.

$$K_{ij} = i(M-i)(j-1)^{-1}j^{-1}[\xi_{j-1}(i-1) - \xi_{j-1}(i)]. \tag{25}$$

This follows immediately from (23), which gives $K_{ij} = \sum_{k=0}^{i} \xi_{j-1}(k-1)$, and (21).

The theorem is not completely new. It restates the eigenvalues given already in Theorem 1. The first five eigenvectors in (23) agree, except for multiplicative constants, with those of Theorem 2(i). Also the relationship between pre- and post-eigenvectors is found to be $K^{ij} = (i-1)ij^{-1}(M-j)^{-1}K_{ji}$, using (24) and (25), and this verifies Theorem 2(ii).

With these preliminaries, we take up the problem of absorption time for the population.

THEOREM 5. *The probability of first absorption at time t, given an initial state i, is*

$$S_i^{(t)} = 2i(M-i)M^{-2}\sum_{j=1}^{[\frac{1}{2}M]}\{[1-2j(2j-1)M^{-2}]^{t-1}\xi_{2j-1}(0) \cdot[\xi_{2j-1}(i-1)-\xi_{2j-1}(i)]\}, \quad t=1,2,3,\cdots, \quad i=0,1,2,\cdots,M \tag{26}$$

where $[\frac{1}{2}M]$ *is the integral part of* $\frac{1}{2}M$.

PROOF. From (10), (23), (24) we have

$$\mathbf{S}^{(t)} = \mathbf{K\Lambda}^{t-1}\mathbf{K}^{-1}\mathbf{S}^{(1)} = \mathbf{C\Xi\Lambda}^{t-1}\mathbf{\Xi}'\mathbf{C}^{-1}\mathbf{S}^{(1)}, \quad t=1,2,3,\cdots,$$

where, by (7), (15),

$$\mathbf{S}^{(1)\prime} = (0,(M-1)M^{-2},0,0,\cdots,0,(M-1)M^{-2},0).$$

Multiplying out the matrices involved, we have, for the ith element,

$$S_i^{(t)} = (M-1)M^{-2}\sum_{j=0}^{M}\{[1-j(j-1)M^{-2}]^{t-1}[\xi_{j-1}(0) - \xi_{j-1}(1)+\xi_{j-1}(M-2)-\xi_{j-1}(M-1)] \cdot i(M-i)(j-1)^{-1}j^{-1}[\xi_{j-1}(i-1)-\xi_{j-1}(i)]\}, \quad t=1,2,3,\cdots, \quad i=0,1,2,\cdots,M. \tag{27}$$

Because

$$\xi_{j-1}(0) = \pm\xi_{j-1}(M-1), \qquad \xi_{j-1}(1) = \pm\xi_{j-1}(M-2), \tag{28}$$

depending on whether j is odd or even, only terms with j even need be included in the summation. Hence we replace j by $2j$, and add to the term with $j = [\frac{1}{2}M]$, the integral part of $\frac{1}{2}M$. From (21), (28), we have that

$$\begin{aligned}(M-1)M^{-2}&[\xi_{2j-1}(0)-\xi_{2j-1}(1)]\\ &= (M-1)M^{-2}[\xi_{2j-1}(M-2)-\xi_{2j-1}(M-1)]\\ &= 2j(2j-1)M^{-2}\xi_{2j-1}(0).\end{aligned} \tag{29}$$

Substituting for (28), (29), into (27) gives the expression (26) of the theorem.

While (26) seems to be the simplest form for the absorption probability at general time t, for t reasonably small, a direct evaluation of (10) could be used. For example, with $t = 1$ we know

$$S_i^{(1)} = (\delta_{i1}+\delta_{i\,M-1})(M-1)M^{-2},$$

and hence we have the

COROLLARY.

$$(\delta_{i1} + \delta_{iM-1})(M - 1)M^{-2} \equiv 2i(M - i)M^{-2} \sum_{j=1}^{[\frac{1}{2}M]} \cdot\{\xi_{2j-1}(0)[\xi_{2j-1}(i-1) - \xi_{2j-1}(i)]\} \quad i = 0, 1, \cdots, M.$$

This, and other identities for the orthogonal polynomials can be obtained from

$$\mathbf{C\Xi\Lambda}^t\mathbf{\Xi}'\mathbf{C}^{-1} \equiv \mathbf{P}^t, \qquad t = 0, 1, 2, \cdots.$$

We note that (26) agrees with (5), in so much as the right hand side is zero when $i = 0$ or $i = M$, for all $t \geqq 1$.

THEOREM 6. *The probability generating function for the first absorption time distribution is*

$$G_i(z) = \delta_{i0} + \delta_{iM} + 2zi(M - i)M^{-2} \sum_{j=1}^{[\frac{1}{2}M]} \cdot\{[1 - 2zj(2j-1)M^{-2}]^{-1}\xi_{2j-1}(0)[\xi_{2j-1}(i-1) - \xi_{2j-1}(i)]\}. \tag{30}$$

PROOF. By definition,

$$G_i(z) = \sum_{t=0}^{\infty} S_i^{(t)} z^t = \delta_{i0} + \delta_{iM} + z\sum_{t=1}^{\infty} S_i^{(t)} z^{t-1}.$$

Substituting for the $S_i^{(t)}$ from (26), and summing the geometric series involved, gives (30).

Because $G_i(z)$ is a probability generating function, we must have $G_i(1) = 1$ for all i, and hence

COROLLARY.

$$2i(M - i)M^{-2} \sum_{j=1}^{[\frac{1}{2}M]} \{[1 - 2j(2j-1)M^{-2}]^{-1}\xi_{2j-1}(0)[\xi_{2j-1}(i-1) - \xi_{2j-1}(i)]\} \equiv 1 - \delta_{i0} - \delta_{iM}, \quad i = 0, 1, \cdots, M.$$

The generating function in Theorem 6 must be consistent with (13), although this is not obvious. One can obtain all the moments of the absorption time T_i by suitable differentiation of $G_i(z)$. Thus

THEOREM 7.

$$E(T_i) = (d/dz)G_i(z)\mid_{z=1} = 2M^2 i(M - i) \sum_{j=1}^{[\frac{1}{2}M]} \{[2j(2j-1)]^{-2}\xi_{2j-1}(0)[\xi_{2j-1}(i-1) - \xi_{2j-1}(i)]\}. \tag{31}$$

$$\text{Var}(T_i) = 4M^4 i(M - 1) \sum_{j=1}^{[\frac{1}{2}M]} \cdot\{[2j(2j-1)]^{-3}\xi_{2j-1}(0)[\xi_{2j-1}(i-1) - \xi_{2j-1}(i)]\} - E(T_i) - [E(T_i)]^2. \tag{32}$$

Proof. The expression (31) for the expected value of T_i is immediate from (30). For the variance, we have

$$\operatorname{Var}(T_i) = E[T_i(T_i - 1)] + E(T_i) - [E(T_i)]^2$$
$$= (d^2/dz^2)G_i(z)\,|_{z=1} + E(T_i) - [E(T_i)]^2.$$

But

$$(d^2/dz^2)G_i(z)\,|_{z=1} = 4M^4 i(M-i) \sum_{j=1}^{[\frac{1}{2}M]} \{[2j(2j-1)]^{-3}[1 - 2j(2j-1)M^{-2}]$$
$$\cdot \xi_{2j-1}(0)[\xi_{2j-1}(i-1) - \xi_{2j-1}(i)]\},$$
$$= 4M^4 i(M-i) \sum_{j=1}^{[\frac{1}{2}M]} \{[2j(2j-1)]^{-3}\xi_{2j-1}(0)$$
$$\cdot[\xi_{2j-1}(i-1) - \xi_{2j-1}(i)]\} - 2E(T_i),$$

by (31). Hence, we obtain (32).

While the above discussion is sufficient to solve all problems of interest, the expressions obtained are not simple to use in practice, even assuming that M is sufficiently small for the values of $\xi_j(k)$ to be available in tables. We shall now show how the moments of T_i can be obtained in terms of elementary functions by using approach Section 2(b). Here we assume that the initial state i is not absorbing and consider the truncated matrix $\mathbf{P}_\Delta$.

Differentiating (13) with respect to z, and evaluating the result at $z = 1$ gives

$$(d/dz)\mathbf{G}_\Delta(z)\,|_{z=1} = [z^{-2}(z^{-1}\mathbf{I}_\Delta - \mathbf{P}_\Delta)^{-2}]_{z=1}(\mathbf{I}_\Delta - \mathbf{P}_\Delta)\mathbf{1}_\Delta$$
$$= (\mathbf{I}_\Delta - \mathbf{P}_\Delta)^{-1}\mathbf{1}_\Delta. \tag{33}$$

This equation was given in [2], p. 378, ex. 17, and in [6], p. 51, with different notations and derivations. In [6], $(\mathbf{I}_\Delta - \mathbf{P}_\Delta)^{-1}$ was called the "fundamental matrix." It was used in [6], p. 177 in the genetics problem of a family tree with non-random mating, whereas here we are concerned with the entire population.

Higher moments can be obtained similarly. For example, for the second factorial moment we have

$$(d^2/dz^2)\mathbf{G}_\Delta(z)\,|_{z=1}$$
$$= [-2z^{-3}(z^{-1}\mathbf{I}_\Delta - \mathbf{P}_\Delta)^{-2} + 2z^{-4}(z^{-1}\mathbf{I}_\Delta - \mathbf{P}_\Delta)^{-3}]_{z=1}(\mathbf{I}_\Delta - \mathbf{P}_\Delta)\mathbf{1}_\Delta$$
$$= -2(\mathbf{I}_\Delta - \mathbf{P}_\Delta)^{-1}\mathbf{1}_\Delta + 2(\mathbf{I}_\Delta - \mathbf{P}_\Delta)^{-2}\mathbf{1}_\Delta \tag{34}$$
$$= 2[(\mathbf{I}_\Delta - \mathbf{P}_\Delta)^{-1} - \mathbf{I}_\Delta](d/dz)\mathbf{G}_\Delta(z)\,|_{z=1}.$$

This formula was given in [6], p. 51, and a genetics example worked in [6], p. 177. For the particular population model (15), we have

Theorem 8. *The first two moments of the absorption time, given an initial state i, are*

$$E(T_i) = (M-i)\sum_{j=1}^{i}(1-jM^{-1})^{-1} + i\sum_{j=1}^{M-i-1}(1-jM^{-1})^{-1}, \tag{35}$$

$$\operatorname{Var}(T_i) = \left\{2M(M-i)\sum_{k=1}^{i} + 2Mi\sum_{k=1}^{M-i-1}\right\}$$

$$\left\{\sum_{j=1}^{k}(1-jM^{-1})^{-1} + k(M-k)^{-1}\sum_{j=1}^{M-k-1}(1-jM^{-1})^{-1}\right\} \tag{36}$$

$$- E(T_i) - [E(T_i)]^2,$$

where the terms in braces are to be multiplied symbolically to give four double sums.

PROOF. Omitting the first and last rows and columns of **P** defined in (15), we see that $(\mathbf{I}_\Delta - \mathbf{P}_\Delta)$ can be written as the product

$$M^{-2}\begin{bmatrix} 1(M-1) & & & & & \\ & 2(M-2) & & & & \\ & & 3(M-3) & & 0 & \\ & & & \ddots & & \\ & 0 & & & (M-2)2 & \\ & & & & & (M-1)1 \end{bmatrix}$$

$$\cdot\begin{bmatrix} 2 & -1 & & & & 0 \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & 0 & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix}$$

The inverse is therefore the product of two symmetric matrices $(\mathbf{I} - \mathbf{P})^{-1} =$

$$M\begin{bmatrix} M-1 & M-2 & M-3 & \cdots & 2 & 1 \\ M-2 & 2(M-2) & 2(M-3) & \cdots & 4 & 2 \\ \vdots & & & & & \\ M-i & 2(M-i) & 3(M-i)\cdots i(M-i)\cdots & & 2i & i \\ \vdots & & & & & \\ 1 & 2 & 3 & \cdots & M-2 & M-1 \end{bmatrix}$$

$$\cdot\begin{bmatrix} 1^{-1}(M-1)^{-1} & & & \\ & 2^{-1}(M-2)^{-1} & & 0 \\ & & \ddots & \\ & 0 & & (M-1)^{-1}1^{-1} \end{bmatrix}$$

Substituting into (33) gives (35) for the ith element, and into (34) gives the ith element

$$\left.\frac{d^2G_i(z)}{dz^2}\right|_{z=1} = \left\{2M(M-i)\sum_{k=1}^{i} + 2Mi\sum_{k=1}^{M-i-1}\right\}$$
$$\cdot \left\{\sum_{j=1}^{k}(1-jM^{-1})^{-1} + k(M-k)^{-1}\sum_{j=1}^{M-k-1}(1-jM^{-1})^{-1}\right\} - 2E(T_i),$$

from which (36) follows.

COROLLARY. Equating (35) and (36) with (31) and (32), respectively, results in two identities for the orthogonal polynomials.

If M is small, (35) and (36) appear to be preferable to (31) and (32), but in actual populations M is large and approximate procedures are required. These are discussed below.

4. Approximations and the diffusion equation. When the population size M is large, the Markov chain can be approximated by a diffusion process continuous in space and time. We make the time scale transformation $u = M^{-2}t$ and the state transformation $y(u) = M^{-1}x(M^2u)$. Since $x = 0, 1, 2, \cdots, M$, we have $y = 0, M^{-1}, 2M^{-1}, \cdots, 1$ and letting $M \to \infty$ but keeping u fixed, it can be shown that the distribution of $y(u)$ approaches a distribution function which has jumps at $y = 0$ and $y = 1$ but is differentiable in the open interval $(0, 1)$, see [12], [13] for similar examples. In other words, the discrete variable $y(u)$ has an approximately continuous distribution within $(0, 1)$ for sufficiently large M. Write the derivative of this distribution as $f(y, u)$; then it may be shown that

$$\frac{\partial f(y, u)}{\partial u} = \frac{\partial^2 y(1-y)f(y, u)}{\partial y^2} \quad \text{for} \quad 0 < y < 1$$

and apart from the accumulations of probability at $y = 0$ and $y = 1$, $f(y, u)$ behaves as an approximate density for $y(u)$. This equation is a special case of the "Fokker-Planck diffusion equation," and requires for its unique solution a specification of the initial function $f(y, 0)$.

Further, following [8], or [12], the probability that the diffusing state is absorbed at $y = 1$ at or before time u is given by the backward equation solution

$$\frac{\partial G(p, u)}{\partial u} = p(1-p)\frac{\partial^2 G(p, u)}{\partial p^2},$$

where $p = y(0) = M^{-1}x(0) = iM^{-1}$, and the boundary conditions $G(0, u) = 0$, $G(1, u) = 1$, must be satisfied for all $u > 0$. The solution of this equation is (see [8], eqn (5.3) with a different notation)

$$G(p, u) = p + \sum_{j=1}^{\infty}(2j+1)p(1-p)(-1)^jF(1-j, j+2, 2, p)e^{-j(j+1)u} \tag{37}$$

From (37) we see that the probability of ultimate absorption at $y = 1 (x = M)$ is $\lim_{u\to\infty} G(p, u) = p$. This result happens to be exactly correct for the discrete Markov process, for the ultimate value of $x(t)$ can only be 0 or M, and it is easily checked from (15) that $E(x(t))$ remains constant throughout time, and therefore $E(x(\infty)) = i \equiv pM$. This is therefore one aspect of the model's behaviour that the diffusion approximation predicts exactly.

Also, from (37), we can find the probability of absorption in either state at or before u by symmetry. It is

$$\begin{aligned} &G(p, u) + G(1-p, u) \\ (38) \qquad &= 1 + \sum_{j=1}^{\infty} (2j+1)p(1-p)(-1)^j[F(1-j, j+2, 2, p) \\ &\qquad + F(1-j, j+2, 2, 1-p)]e^{-j(j+1)u}, \end{aligned}$$

and hence the probability of absorption at exactly time t (on the old scale) is approximately

$$\begin{aligned} S_i^{(t)} &\doteqdot G(p, M^{-2}t) + G(1-p, M^{-2}t) - G(p, M^{-2}(t-1)) \\ &\qquad - G(1-p, M^{-2}(t-1)) \\ (39) \qquad &= \sum_{j=1}^{\infty} (2j+1)p(1-p)(-1)^j[F(1-j, j+2, 2, p) \\ &\qquad + F(1-j, j+2, 2, 1-p)]e^{-j(j+1)tM^{-2}}[1 - e^{j(j+1)M^{-2}}] \end{aligned}$$

where $p = iM^{-1}$. An exact formula for $S_i^{(t)}$ was given in (26) but comparisons for the accuracy of (39) as an approximation seem hopeless. In any case (39) seems no easier to compute than (26).

The moments of the first absorption time T_i can be calculated with considerable difficulty from the approximate distribution (39), but for the mean a simpler procedure is the following. Write $U(p)$ as the expected value of the time—measured on the u-scale—for one or other absorbing state to be first reached. Then Feller [3] states that $U(p)$ is the solution of an ordinary differential equation which reduces to

$$p(1-p)[d^2U(p)/dp^2] = -1$$

in our case, with the boundary conditions $U(0) = U(1) = 0$. The solution is

$$U(p) = \log[p^{-p}(1-p)^{-(1-p)}],$$

or measured on the t scale and with $p = iM^{-1}$,

$$(40) \qquad E(T_i) \doteqdot M^2 \log[p^{-p}(1-p)^{-(1-p)}].$$

Comparing this with the exact result found in (35), we see that the diffusion approximation is equivalent to replacing the Riemann summation in (35) by integration, and hence the approximation should be good for all but very small values of M.

Writing $V(p)$ as the second moment of the distribution (38), that is

$$V(p) = \int_0^{\infty} u^2\, d[G(p, u) + G(1-p, u)],$$

one can show that $V(p)$ is the solution of the equation

$$p(1-p)\frac{d^2V(p)}{dp^2} = -2U(p),$$

where $U(p)$ is as above, and $V(0) = V(1) = 0$. The solution is

$$V(p) = \frac{1}{3}\pi^2 - 2\sum_{j=1}^{\infty}\frac{p^{j+1} + (1-p)^{j+1}}{j^2} - 2\log[p^{-p}(1-p)^{-(1-p)}].$$

Making the mean correction and reverting to the t-time scale, we have for the variance of the first absorption time,

$$\text{Var}(T_i) \doteqdot M^4V(p) - [E(T_i)]^2, \qquad p = iM^{-1}.$$

It may be verified that this approximation is got if, in the exact formula (36), the summations are replaced by integrations, and terms of order less than M^4 are ignored. Thus for both the mean and the variance, the diffusion approximation should be adequate for all but very small M.

Other aspects of the model's behaviour have been given in [9], [10].

Acknowledgment. The author thanks Mr. J. E. Moyal for suggesting the use of the eigenvector expansion (9), and Professor E. J. Hannan for discussion of his derivation of the eigenvalues leading to Theorem 1.

REFERENCES

[1] Harry Bateman, *Higher Transcendal Functions* (compiled by Erdélyi, *et al.*), Vol. 2, McGraw-Hill, New York, 1953.

[2] William Feller, *An Introduction to Probability Theory and its Applications*, Vol. 1, 2nd ed., Wiley, New York, 1957.

[3] William Feller, "Diffusion processes in one dimension," *Trans. Amer. Math. Soc.*, Vol. 77 (1954), pp. 1–31.

[4] Samuel Karlin and James McGregor, "Linear growth, birth and death processes," *J. Math. Mech.*, Vol. 7 (1958), pp. 643–662.

[5] Samuel Karlin and James McGregor, "Random walks," *Illinois J. Math.*, Vol. 3 (1959), pp. 66–81.

[6] John G. Kemeny and J. Laurie Snell, *Finite Markov Chains*, D. Van Nostrand, Princeton, 1959.

[7] D. G. Kendall, "Les processus stochastiques de croissance en biologie," *Ann. Inst. Henri Poincaré*, Vol. 13 (1952), pp. 43–108.

[8] Motoo Kimura, "Some problems of stochastic processes in genetics," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 882–901.

[9] P. A. P. Moran, "Random processes in genetics," *Proc. Camb. Phil. Soc.*, Vol. 54 (1958), pp. 60–71.

[10] P. A. P. Moran, "The rate of approach to homozygosity," *Ann. Hum. Genet.*, Vol. 23 (1958), pp. 1–5.

[11] E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for Statisticians*, Vol. 1, Cambridge University Press, Cambridge, 1956.

[12] G. A. Watterson, "On the relationship between a discrete and a continuous diffusion process, with applications to genetic populations. I. Theoretical aspects," submitted to *Ann. Math. Stat.*

[13] G. A. Watterson, "On the relationship between a discrete and a continuous diffusion process, with application to genetic populations. II. Some finite population models," in preparation.