

A Note on Finite Homogeneous Continuous-Time Markov Chains

S. TAVARÉ¹

Department of Probability and Statistics, The University, Sheffield, England

Summary

A simple and intuitive method of deriving some properties of finite homogeneous continuous-time Markov chains is given. These chains have absorbing barriers, and arise in the study of population genetics. The method involves the representation of such a process in terms of a discrete Markov chain, and a series of waiting times, which reconstruct the original time scale. An example from population genetics is given.

1. Introduction

Finite, homogeneous, continuous-time Markov chains (for convenience referred to as FCMC's) are often used as models for mechanisms arising in physical sciences; such models arise frequently in population genetics. FCMC's with absorbing states arise naturally as models for the evolution of the number of a particular allele at a single locus in a population of fixed size m , say, at which two alleles are possible (*cf.* Moran 1962). For such models, the questions of importance to geneticists include ultimate fixation probabilities, the distribution and moments of the absorption times, the distribution of sojourn times at particular gene frequencies and, more recently, the age of alleles problem (*cf.* Watterson 1977). Explicit solutions to such problems are often hard to find, and one usually resorts to diffusion approximations or computes numerical solutions. Computationally convenient formulas for corresponding discrete-time models are well known (Kemeny and Snell 1960). This note gives the corresponding results for FCMC's, and provides simple, intuitively appealing proofs. The proofs, which are given in the appendix, use the discrete-time results, and the representation of a FCMC in terms of a related discrete-time Markov chain and a sequence of waiting times which recreate the original time scale.

The methods given here can be applied to any FCMC with at least one absorbing state, to the conditioned chains developed for use in genetics by Ewens (1973), and can be modified to accommodate processes in which the waiting times do not have exponential distributions (*i.e.*, semi-Markov processes).

2. Notation and Results

For simplicity, suppose that the FCMC $\{X(t), t \geq 0\}$ has state space $B = \{0, 1, \dots, m\}$, with absorbing barriers at 0 and m . Such a process is uniquely determined by its in-

¹Present Address: Department of Mathematics, University of Utah, Salt Lake City, Utah 84112, U.S.A.

Key Words: Markov chains; Absorption times; Genetic models.

finitesimal generator (q_{ij}) , where q_{ij} is the transition rate from state i to state j . The elements q_{ij} satisfy, for each i ,

$$\sum_{j \neq i} q_{ij} = -q_{ii} = q_i, \quad 0 \leq q_i < \infty. \quad (1)$$

For convenience, write the generator in the form

$$(q_{ij}) = \begin{pmatrix} 0 & \mathbf{0}' & 0 \\ \mathbf{q}_0 & \mathbf{Q} & \mathbf{q}_m \\ 0 & \mathbf{0}' & 0 \end{pmatrix} \quad (2)$$

The matrix \mathbf{Q} gives transition rates among the transient states $T = \{1, 2, \dots, m-1\}$, while $\mathbf{q}_j = (q_{1j}, q_{2j}, \dots, q_{m-1,j})'$ gives the transition rates into absorbing state j . If $k \in T$, then $q_k > 0$, since k is not absorbing, and the waiting time W_k in state k has an exponential distribution with mean q_k^{-1} . The associated discrete-time jump chain $\{J_n, n \geq 0\}$ describes the sequence of states through which X travels on its way to absorption. It has transition matrix $P = (p_{ij})$ determined by

$$p_{00} = p_{mm} = 1; \quad p_{ii} = 0; \quad p_{ij} = q_{ij}q_i^{-1}, \quad i \in T, j \in B, j \neq i. \quad (3)$$

The behaviour of X is determined by the jump chain with $J_0 = X(0)$, and the waiting times $\{W_1, W_2, \dots, W_{m-1}\}$. Starting from $X(0) = i$ the process waits a length of time W_i , and then jumps to state j with probability p_{ij} . If $j \in T$, the process waits a length of time W_j in state j , and then jumps to state k with probability p_{jk} . This is continued until a jump to an absorbing state is made and the process stops.

For $i, j \in T$, let S_{ij} be the time that X spends in state j before absorption occurs, starting from $X(0) = i$. Let $n_{ij} = ES_{ij}$ be the mean of the sojourn time S_{ij} , and let \mathbf{N} be a square matrix with elements n_{ij} . The time to reach an absorbing state from $X(0) = i$ is denoted by S_i ; clearly, $S_i = \sum S_{ij}$. The mean absorption time ES_i is denoted by n_i , and the variance by v_i . Finally, let π_{ik} be the probability that X reaches absorbing state k , starting from i , and let π_k be a vector with elements π_{ik} , $1 \leq i \leq m-1$.

We will use the notation $(*)$ to denote the corresponding quantities of the jump chain $J = \{J_n, n \geq 0\}$. For example S_{ij}^* is the time that J spends in state j , given $J_0 = i$, before absorption occurs, and n_{ij}^* is the mean sojourn time, ES_{ij}^* . Finally, let \mathbf{e}_i be a vector with a 1 in the i th place, and 0 elsewhere, and let \mathbf{e} be a vector of 1's.

We now state a theorem which gives the continuous-time analogues of some classical results from the theory of discrete-time Markov chains. The proof, which uses the representation of X in terms of the waiting times and jump chain J , is given in the appendix.

Theorem. Let $i, j \in T$. Then:

- (i) \mathbf{Q} is nonsingular, and $\mathbf{N} = -\mathbf{Q}^{-1}$.
- (ii) $n_{ij} = n_{ij}^* q_j^{-1}$.
- (iii) $P(S_{ij} = 0) = 1 - (n_{ij}n_{jj}^{-1})$; $P(0 < S_{ij} \leq t) = (n_{ij}n_{jj}^{-1})[1 - \exp(-tn_{jj}^{-1})]$, $t > 0$.
- (iv) $\pi_k = \mathbf{N}\mathbf{q}_k$, $k = 0, m$.
- (v) For $\theta_1, \dots, \theta_{m-1} > 0$, the joint distribution of $(S_{i1}, \dots, S_{i,m-1})$ has a Laplace transform given by

$$E\left(\prod_{j=1}^{m-1} \exp(-\theta_j S_{ij})\right) = \mathbf{e}_i'(\mathbf{Q} - \mathbf{\Theta})^{-1}\mathbf{Q}\mathbf{e},$$

where $\mathbf{\Theta} = \text{diag}\{\theta_1, \dots, \theta_{m-1}\}$.

(vi) The Laplace transform of S_i is given by $E[\exp(-\theta S_i)] = \mathbf{e}'_i(\mathbf{Q} - \theta \mathbf{I})^{-1} \mathbf{Q} \mathbf{e}$, $\theta > 0$, and hence S_i has a matrix exponential distribution with density

$$f_i(t) = -\mathbf{e}'_i \mathbf{Q} \exp(\mathbf{Q}t) \mathbf{e}, \quad t \geq 0.$$

(vii) $n_i = \mathbf{e}'_i \mathbf{N} \mathbf{e}$, $v_i = 2\mathbf{e}'_i \mathbf{N}^2 \mathbf{e} - n_i^2$, $i \in T$.

3. Comments and Example

Application of the results given here depends crucially on the determination of \mathbf{N} . The formula in (i) is computationally convenient, since \mathbf{Q} is usually strongly diagonal, and thus easy to invert numerically. Part (iii) agrees with the results of Nagylaki (1974), who considered Markov processes with continuous state-space. The probability mass at the origin accounts for the fact that j may never be visited from i before absorption. Equation (iv) is again useful computationally. The Laplace transform in (v) is of limited use in evaluating distributions, but a useful identity for computing moments. The density given in (vi) was derived by a different method by Tan (1976) in the special case of birth-death processes. Both these formulas are natural generalizations of the usual exponential distributions.

There are some cases in which \mathbf{N} can be determined explicitly. One particular case is the birth-death process, where we can make use of known formulas for the jumping chain, which is a random walk (see, e.g., Ewens 1964). A special case is given below.

Example: A genetic model of Moran (1962). (See also Karlin and McGregor 1962 and Tan 1976.) In this model $X(t)$ is the number of A_1 alleles at time t in a haploid population with two alleles, A_1 and A_2 , and fixed population size m . We assume there are no mutation forces acting. The model is then well known to be a birth-death process, with transition rates determined by

$$q_{i,i+1} = i(m-i)\lambda_2/m \quad \text{and} \quad q_{i,i-1} = i(m-i)\lambda_1/m$$

where $\lambda_1, \lambda_2 > 0$. In the case of no selection, $\lambda_1 = \lambda_2$. When $\lambda_1 \neq \lambda_2$, let $\gamma = \lambda_2/\lambda_1$. Then

$$n_{ij} = \frac{m}{j(m-j)(\lambda_2 - \lambda_1)(\gamma^m - 1)} \begin{cases} (\gamma^j - 1)(\gamma^{m-i} - 1); & 1 \leq j \leq i, \\ (\gamma^j - 1)(\gamma^{m-i} - \gamma^{j-i}); & i \leq j \leq m-1, \end{cases}$$

and if $\lambda_1 = \lambda_2 = \lambda$, then

$$n_{ij} = (m-i)/\lambda(m-j), \quad 1 \leq j \leq i; = i/\lambda j, \quad i \leq j \leq m-1.$$

Acknowledgments

I would like to thank the editor for comments which have clarified the presentation of the results.

Résumé

On donne une méthode simple et intuitive pour déduire certaines propriétés de chaînes de Markov finies homogènes à temps continu. Ces chaînes sont à barrières absorbantes, et apparaissent dans L'étude de la génétique des populations. Le méthode consiste à représenter un tel processus en termes de chaîne de Markov discrète et d'une suite de temps d'attente qui reconstruisent l'échelle de temps originale. On donne un exemple de génétique des populations.

References

Barnett, V. (1964). Joint distribution of occupation totals for a simple random walk. *Journal of the Australian Mathematical Society* 4, 518-528.
 Chung, K. L. (1967). *Markov Chains with Stationary Transition Probabilities*. Springer Verlag, Berlin.
 Ewens, W. J. (1964). The pseudo-transient distribution and its use in genetics. *Journal of Applied Probability* 1, 141-156.
 Ewens, W. J. (1973). Conditional diffusion processes in population genetics. *Theoretical Population Biology* 4, 21-30.
 Karlin, S. and McGregor, J. (1962). On a genetics model of Moran. *Proceedings of the Cambridge Philosophical Society* 58, 299-311.
 Kemeny, J. G. and Snell, J. L. (1960). *Finite Markov Chains*. Van Nostrand, New York.
 Moran, P. A. P. (1962). *The Statistical Processes of Evolutionary Theory*. Clarendon Press, Oxford.
 Nagylaki, T. (1974). The moments of stochastic integrals and the distribution of sojourn times. *Proceedings of the National Academy of Sciences* 71, 746-749.
 Tan, W. Y. (1976). On the absorption probabilities and absorption times of finite homogeneous birth-death processes. *Biometrics* 32, 745-752.
 Watterson, G. A. (1977). Reversibility and the age of an allele. II. Two allele models with selection and mutation. *Theoretical Population Biology* 12, 179-196.

Received June 1978; Revised March 1979

Appendix

First, let $D = \text{diag} \{q_1, \dots, q_{m-1}\}$, and write

$$P = \begin{bmatrix} 1 & \mathbf{0}' & 0 \\ \mathbf{p}_0 & Q^* & \mathbf{p}_m \\ 0 & \mathbf{0}' & 1 \end{bmatrix}.$$

From (2) and (3), $\mathbf{p}_k = D^{-1}\mathbf{q}_k$ when $k = 0$ or m , and $Q = -D(I - Q^*)$. From Kemeny and Shell (1960, p. 47), $(I - Q^*)$ is nonsingular, and $N^* = (I - Q^*)^{-1}$. It follows that Q is nonsingular. Next, note that $\{X(t)\}$ and $\{J_n\}$ execute the same transitions in the same order. Clearly, we have

$$S_{ij} = \sum_{\ell=1}^{S_{ij}^*} Y_{ij}^{\ell}, \tag{4}$$

where, using the strong Markov property, the $\{Y_{ij}^{\ell}\}$ are i.i.d. with the same distribution as W_j . Hence $ES_{ij} = n_{ij} = ES_{ij}^* EY_1^{\ell} = n_{ij}^* q_j^{-1}$, and (ii) follows. Then $N = N^*D^{-1} = (I - Q^*)^{-1}D^{-1} = [D(I - Q^*)]^{-1} = -Q^{-1}$, which completes the proof of (i). To prove (iv), notice that the absorption probabilities of X and J are identical. Hence $\pi_k = \pi_k^* = N^*\mathbf{p}_k = N^*D^{-1}\mathbf{q}_k = N\mathbf{q}_k$, $k = 0$ or m . To prove (iii), we use a results of Kemeny and Snell (1960, p. 62). Let $|z| < 1$. Then they show that

$$E(z^{S_{ij}^*}) = (n_{ij}^*/n_{jj}^*)z[n_{jj}^* - (n_{jj}^* - 1)z]^{-1} + 1 - (n_{ij}^*/n_{jj}^*). \tag{5}$$

Hence for $\theta > 0$, we have

$$\begin{aligned} E[\exp(-\theta S_{ij})] &= \left(E \exp\left(-\theta \sum_{\ell=1}^{S_{ij}^*} Y_{ij}^{\ell}\right) \right) = E\{E[\exp(-\theta \sum Y_{ij}^{\ell}) | S_{ij}^*]\} \\ &= E[q_j/(\theta + q_j)]^{S_{ij}^*} = (n_{ij}/n_{jj})[n_{jj}^{-1}/(\theta + n_{jj}^{-1})] + 1 - (n_{ij}/n_{jj}). \end{aligned}$$

The last expression is the Laplace transform of the distribution given (iii).

The proof of (vi) follows essentially the same lines, but the starting point is not (5), but instead uses the matrix generating from

$$E(z_1^{S_{11}^*} z_2^{S_{12}^*} \dots z_m^{S_{m-1}^*}) = \mathbf{e}' (Z^{-1} - Q^*)^{-1} (I - Q^*) \mathbf{e},$$

where $Z = \text{diag} \{z_1, \dots, z_{m-1}\}$. See, for example, Barnett (1964). By conditioning on $(S_{11}, \dots, S_{im-1}^*)$ the result of (v) follows. Point (vi) is an immediate consequence of (v), obtained by setting $\theta_1 = \theta_2 = \dots = \theta_{m-1} = \theta$. The last formula follows from (vi) by differentiation with respect to θ . Chung (1967) provides a more general proof of (iii) which does not use the characterization used here.