Reprinted from
ADVANCES IN APPLIED PROBABILITY Vol. 8, No. 4, December 1976
SYMPOSIUM ON MATHEMATICAL GENETICS
Liverpool, 5–6 April 1976

Sojourn times for conditioned Markov chains in genetics

S. TAVARÉ, *University of Sheffield*

We consider a Markov chain $\{X_n, n \geqq 0\}$ with state space $\{0, 1, \cdots, M\}$, with absorbing barriers at 0 and M. Following the notation of Kemeny and Snell (1960), we let T_{ij} be the number of times the chain is in state j, given $X_0 = i$, before absorption at 0 or M. We write the one-step transition matrix of the chain in the form

$$P = (p_{ij}) = \begin{pmatrix} I & 0 \\ R & Q \end{pmatrix}$$

Then $N = (n_{ij}) = (ET_{ij}) = (I - Q)^{-1}$.

Let π_i be the probability of absorption at 0, given $X_0 = i$, and consider the chain $\{X_n^*\}$ conditioned to be absorbed at 0. For the conditioned chain, we have a one-step transition matrix P^*, and

$$N^* = (n^*_{ij}) = (ET^*_{ij}) = D^{-1}ND$$

where $D = \text{diag}\{\pi_1, \cdots, \pi_{M-1}\}$.

The class of chains we consider will be assumed to have the following properties:

(A) (i) $\{X_n\}$ is a martingale;

(ii) $p_{ij} = p_{M-i,M-j}$.

Two chains satisfying (A) are the pure-drift genetic model of Wright, for which

(1) $$p_{ij} = \binom{M}{J}\left(\frac{i}{M}\right)^j\left(1-\frac{i}{M}\right)^{M-j}; \qquad i,j = 0,1,\cdots,M$$

and Moran, for which

(2) $$\begin{aligned} p_{ii-1} &= p_{ii+1} = i(M-i)/M^2 = p_i \\ p_{ii} &= 1-2p_i \\ p_{ij} &= 0, \quad |i-j| \geqq 2. \end{aligned}$$

For details of the Wright and Moran models see Moran (1962). We can also modify these transition probabilities slightly by setting

(3) $$p_i = i(M-i)/MC$$

where C is a function of M chosen to ensure $0 < p_i \leqq \frac{1}{2}$, $\forall i$, M. Condition (ii) of (A) is a consequence of the 'exchangeability' of this type of genetic model. See Cannings (1974).

Ewens (1973) has considered diffusion approximations to this type of chain. After appropriate scaling of the time and state spaces of $\{X_n\}$, we obtain a process $X(t)$ diffusing over $[0,1]$, with drift and diffusion coefficients given by

(B) $$\begin{aligned} \alpha(x) &= 0, \\ \beta(x) &= x(1-x), \text{ respectively.} \end{aligned}$$

He shows that for processes satisfying (B) which are conditioned to hit 0, the mean time spent in any interval (a,b) is given by

$$t^*(a,b) = \int_b^a t^*(x)dx,$$

where

(4) $$\begin{aligned} t^*(x) &= 2, \quad 0 \leqq x \leqq p \\ &= 2p(1-x)/x(1-p), \quad p \leqq x \leqq 1. \end{aligned}$$

Translated back to the underlying conditioned Markov chains, we see that (4) is equivalent to requiring that N^* satisfy

(C) $$n^*_{ij} = C \quad \text{for} \quad 1 \leqq j \leqq i \leqq M-1.$$

We now want to know which models satisfying (A), whose diffusion coefficients (after suitable scaling) satisfy (B), also satisfy (C). It turns out that $\{X_n\}$ must be a symmetric random walk, whose transition probabilities satisfy (3). It is, of course, possible to construct transition matrices P^* which also satisfy (C). In this case, they must be skip-free downwards, but we cannot in general find absorption probabilities, and so reconstruct a P which is the transition matrix of a chain with two absorbing barriers. This gives a reason for Assumption (A). In particular, the Wright model (1) does not satisfy (C). Some discussion of the errors involved in the diffusion approximation in this case is given by Pollak and Arnold (1975). Further, if we let $\phi^*_{ij}(s)$ be the probability generating function of T^*_{ij}, and suppose $i \geqq j$, then $\phi^*_{ij}(s)$ is independent of i and j iff P satisfies (3). In this case, all higher moments of T^*_{ij} will be independent of i and j also.

Our diffusion approximation processes have turned out to be limits (in an exact sense) of Moran-type random walks satisfying (3). This is reasonable—nicely-behaved diffusions usually arise as limits of absorbing chains which have only local changes of state. The errors in using this diffusion approximation for Wright's model may be attributed to the early behaviour of the chain, where there is a higher probability of hitting an absorbing barrier. So the application of diffusion approximation methods to processes like Wright's model are not really appropriate. However, we can approximately reproduce Wright's model by looking at Moran's model at time points $0, [M/2], M, \cdots$ instead of $0, 1, 2, \cdots$. The transition matrix of this new chain is given by $P^{[M/2]}$, where P is as in (2). For large values of M, the eigenvalues of Wright's model and those of $P^{[M/2]}$ are close (see e.g. Cannings (1974)) and this provides an explanation of why the diffusion approximation appears insensitive to the type of Markov chain to which it is applied.

References

CANNINGS, C. (1974) The latent roots of certain Markov chains arising in genetics: A new approach. I. Haploid models. *Adv. Appl. Prob.* **6**, 260–290.

EWENS, W. J. (1973) Conditional diffusion processes in population genetics. *Theor. Pop. Biol.* **4**, 21–30.

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.

POLLAK, E. AND ARNOLD, B. C. (1975) On sojourn times at particular gene frequencies. *Gen. Res., Camb.* **25**, 89–94.