LOOKING FORWARDS AND BACKWARDS IN THE MULTI-ALLELIC NEUTRAL CANNINGS POPULATION MODEL

M. MÖHLE,* *Eberhard Karls Universität Tübingen*

Abstract

We look forwards and backwards in the multi-allelic neutral exchangeable Cannings model with fixed population size and nonoverlapping generations. The Markov chain X is studied which describes the allelic composition of the population forward in time. A duality relation (inversion formula) between the transition matrix of X and an appropriate backward matrix is discussed. The probabilities of the backward matrix are explicitly expressed in terms of the offspring distribution, complementing the work of Gladstien (1978). The results are applied to fundamental multi-allelic Cannings models, among them the Moran model, the Wright–Fisher model, the Kimura model, and the Karlin and McGregor model. As a side effect, number theoretical sieve formulae occur in these examples.

Keywords: Cannings model; duality; exchangeability; generalized Stirling number; inversion formula; multi-allele model; neutrality; principle of inclusion and exclusion; Silvester's sieve formula

2010 Mathematics Subject Classification: Primary 60J10; 92D10
Secondary 60K35; 92D25

1. Introduction

Cannings [1], [2] introduced haploid discrete population models with constant population size $N \in \mathbb{N} := \{1, 2, \ldots\}$ and nonoverlapping generations $r \in \mathbb{N}_0 := \{0, 1, 2, \ldots\}$. In each generation the individuals are randomly labeled from 1 to N. The ith individual of the rth generation produces a random number, $\nu_i^{(r)}$, of offspring, where $\nu_1^{(r)} + \cdots + \nu_N^{(r)} = N$. Cannings assumed that, for each fixed r, the random variables $\nu_1^{(r)}, \ldots, \nu_N^{(r)}$ are exchangeable and that the model is time homogeneous in the sense that the random vectors $\nu^{(r)} := (\nu_1^{(r)}, \ldots, \nu_N^{(r)})$, $r \in \mathbb{N}_0$, are independent and identically distributed. For convenience, define $\nu := \nu^{(0)}$ and $\nu_i := \nu_i^{(0)}$ for $i \in \{1, \ldots, N\}$. The most celebrated examples are the Wright–Fisher model [14], in which ν has a symmetric multinomial distribution, and the Moran model, in which ν is a random permutation of $(0, 1, 1, \ldots, 1, 1, 2)$. Many other discrete population models in the literature, in particular most of the examples collected by Gladstien [4]–[7], are (or can at least be viewed as) Cannings models. For example, as explained in Section 5, the Kimura model (see [7, p. 636] or [10]) can be viewed as a Cannings model with symmetric multi-hypergeometric joint offspring distribution (see (11)). Another example studied in more detail in Section 5 is the Karlin and McGregor conditional branching process model (see [7, p. 636] or [9]). These classical examples demonstrate the importance of the class of Cannings models. In this paper we are interested in the multi-allelic version of the Cannings model, which is less intensively

studied in the literature. Gladstien [7, pp. 638–640] considered multi-type models; however, he did not provide explicit formulae for the multi-allelic Cannings model. The basic reproduction model is the Cannings model described above. In addition, in the multi-allelic version it is assumed that each individual has one of $K \in \mathbb{N}$ possible types and that each offspring inherits the type of its parent (neutrality, no mutation, no selection). For a nonneutral discrete Moran model, we refer the reader to [8]. The paper is organized as follows. In Section 2 we briefly focus on the Markov chain which describes the allelic composition of the population forward in time. In Section 3 we look backward in time and derive explicit expressions for the so-called backward matrix of the multi-allelic Cannings model. In Section 4 the duality relation in the spirit of Gladstien [7] is discussed in more detail. For fundamental examples (Moran model, Wright–Fisher model, Kimura model, Karlin and McGregor model, uniform model) detailed explicit formulae for the multi-allelic forward and backward matrices are derived in Section 5.

2. Forward structure

Let $X_k(r)$ denote the number of descendants of type $k \in \{1, \ldots, K\}$ in generation $r \in \mathbb{N}_0$, and set $X(r) := (X_1(r), \ldots, X_K(r))$. It is easily seen that $X := (X(r))_{r \in \mathbb{N}_0}$ is a time-homogeneous Markov chain with state space

$$E_{N,K} := \{i = (i_1, \ldots, i_K) \in \mathbb{N}_0^K : i_1 + \cdots + i_K = N\}.$$

Note that $|E_{N,K}| = \binom{N+K-1}{K-1}$. The chain X moves from the state $i = (i_1, \ldots, i_K) \in E_{N,K}$ to the state $j = (j_1, \ldots, j_K) \in E_{N,K}$ with transition probability

$$\pi_{ij} := P(X(r+1) = j \mid X(r) = i) = P(C = j), \tag{1}$$

where $C := (C_1, \ldots, C_K)$ with $C_k := \sum_{s=s_{k-1}+1}^{s_k} \nu_s$ for $k \in \{1, \ldots, K\}$, $s_0 := 0$, and $s_k := i_1 + \cdots + i_k$ for $k \in \{1, \ldots, K\}$. Note that C depends on i and that $C_1 + \cdots + C_K = \nu_1 + \cdots + \nu_N = N$. In general, the transition probability π_{ij} depends on the population size N and on the number of types K.

3. Backward structure

In the following the space $S_{N,K} := \{i = (i_1, \ldots, i_K) \in \mathbb{N}_0^K : i_1 + \cdots + i_K \leq N\}$ will play an important role. From $S_{N,K} = \bigcup_{m=0}^{N} E_{m,K}$, it follows that

$$|S_{N,K}| = \sum_{m=0}^{N} |E_{m,K}| = \sum_{m=0}^{N} \binom{m+K-1}{K-1} = \binom{N+K}{K}.$$

Note that $|E_{N,K+1}| = |S_{N,K}|$, a fact which we will come back to in Section 4. For most purposes, the order of the elements of $S_{N,K}$ is unimportant; however, it is convenient to think of the elements ordered lexicographically, for example, the 10 elements of $S_{3,2}$ in the order $(0, 0)$, $(0, 1)$, $(0, 2)$, $(0, 3)$, $(1, 0)$, $(1, 1)$, $(1, 2)$, $(2, 0)$, $(2, 1)$, and $(3, 0)$. In the brief Section 2 the Cannings model was studied forward in time. It is also reasonable to look from some generation $r \in \mathbb{N}$ backward in time. More precisely, fix some $i = (i_1, \ldots, i_K) \in S_{N,K}$ and take in some generation $r \in \mathbb{N}$ a sample of $|i| := i_1 + \cdots + i_K$ individuals and suppose that i_k of these individuals are of type k, $k \in \{1, \ldots, K\}$. For $j = (j_1, \ldots, j_K) \in S_{N,K}$, let A_{ij} denote the event that the i_1 individuals of type 1 have exactly j_1 parents, the i_2 individuals of type 2 have exactly j_2 parents, $\ldots$, and the i_K individuals of type K have exactly j_K parents. Note

that, for fixed $i \in S_{N,K}$, the events A_{ij}, $j \in S_{N,K}$, are disjoint. The following proposition provides an explicit formula for the probability $p_{ij} := P(A_{ij})$.

Proposition 1. *For* $i = (i_1, \ldots, i_K)$, $j = (j_1, \ldots, j_K) \in S_{N,K}$,

$$p_{ij} := P(A_{ij}) = \frac{(N - |i|)!\, i_1! \cdots i_K!}{(N - |j|)!\, j_1! \cdots j_K!} \sum_m E\left(\binom{\nu_1}{m_1} \cdots \binom{\nu_{|j|}}{m_{|j|}}\right), \qquad (2)$$

where

$$|i| := i_1 + \cdots + i_K, \qquad |j| := j_1 + \cdots + j_K,$$

and the sum extends over all $m = (m_1, \ldots, m_{|j|}) \in \mathbb{N}^{|j|}$ *satisfying* $m_1 + \cdots + m_{j_1} = i_1$, $m_{j_1+1} + \cdots + m_{j_1+j_2} = i_2, \ldots, m_{j_1+\cdots+j_{K-1}+1} + \cdots + m_{|j|} = i_K$. *In particular,* $p_{ij} = 0$ *if* $j_k > i_k$ *for some* $k \in \{1, \ldots, K\}$. *Thus, thinking of* $S_{N,K}$ *as being lexicographically ordered, the matrix* $P := (p_{ij})_{i,j \in S_{N,K}}$ *is left lower triangular. The eigenvalues of* P *are hence* $p_{ii} = E(\nu_1 \cdots \nu_{|i|})$, $i \in S_{N,K}$.

Remarks. 1. Comparing the definition of the matrix P with the definition of Gladstien's matrix G in [7, pp. 638–639], it is clear that $P = G^\top$, the transpose of G. If $K = 1$ then (2) reduces to (see, for example, Gladstien [7, p. 637])

$$p_{ij} = \frac{\binom{N}{j}}{\binom{N}{i}} \sum_{\substack{m_1, \ldots, m_j \in \mathbb{N} \\ m_1 + \cdots + m_j = i}} E\left(\binom{\nu_1}{m_1} \cdots \binom{\nu_j}{m_j}\right), \qquad i, j \in \{0, \ldots, N\}.$$

Proposition 1 extends this formula to the multi-allelic case. Note that Gladstien did not provide explicit expressions for the entries of G for the multi-allelic Cannings model ($K > 1$).

2. Since, for each fixed $i \in S_{N,K}$, the events A_{ij}, $j \in S_{N,K}$, are disjoint, the inequality

$$\sum_{j \in S_{N,K}} p_{ij} = P\left(\bigcup_{j \in S_{N,K}} A_{ij}\right) \le 1$$

holds for all $i \in S_{N,K}$. Thus, the matrix P is substochastic. For $K = 1$, the matrix P is well known to be stochastic, whereas, for $K > 1$, the matrix P is in general not stochastic. This fact can be explained as follows. Let $M_{i_k} \subseteq \{1, \ldots, N\}$ denote the random set of all parents of the i_k individuals of type k, $k \in \{1, \ldots, K\}$. Since individuals of different type cannot have the same parent, it follows that $\bigcup_{j \in S_{N,K}} A_{ij} = \bigcap_{1 \le k < l \le K} \{M_{i_k} \cap M_{i_l} = \varnothing\}$, an event which, for $K > 1$, in general does not coincide with the full space Ω. In fact, for particular models, the matrix P may even have rows with only zero entries (see, for example, the simple model presented at the beginning of Section 5).

Proof of Proposition 1. Fix $i, j \in S_{N,K}$ and $r \in \mathbb{N}$. Obviously,

$$p_{ij} = \sum_k P(A_{ij} \mid \nu^{(r-1)} = k)\, P(\nu^{(r-1)} = k),$$

where the sum extends over all $k = (k_1, \ldots, k_N) \in \mathbb{N}_0^N$ with $k_1 + \cdots + k_N = N$ and $P(\nu^{(r-1)} = k) > 0$. The computation of the conditional probability $P(A_{ij} \mid \nu^{(r-1)} = k)$ is equivalent to the following 'putting balls into boxes' problem. Suppose that there are N balls

given, i_k of them of type k, $k \in \{1, \ldots, K\}$. The balls are labeled from 1 to N such that the balls of type 1 have labels $1, \ldots, i_1$, the balls of type 2 have labels $i_1 + 1, \ldots, i_1 + i_2$, and so on. Note that the last $N - |i| = N - (i_1 + \cdots + i_K)$ balls are not so important and, hence, not assigned any type. We call these $N - |i|$ balls the neutral balls. Moreover, there are N empty boxes given, k_l of color l, $l \in \{1, \ldots, N\}$. The boxes are labeled from 1 to N such that the boxes $1, \ldots, k_1$ have color 1, the boxes $k_1 + 1, \ldots, k_1 + k_2$ have color 2, and so on. Each box has space for exactly one ball. The N balls are thrown randomly on to the N boxes such that, after the experiment, each box contains exactly one ball. There are obviously $N!$ outcomes of this experiment. We are interested in the probability of the event that, after the experiment, boxes of the same color only contain balls of the same type (and possibly some further neutral balls) and that, for each $k \in \{1, \ldots, K\}$, the number of colors of those boxes containing balls of type k is equal to j_k. Let L_k denote the number of colors of the boxes which contain a ball of type k. Let m_{kl} denote the number of balls of type k which belong to boxes of color l. The event we are interested in corresponds to the constrains that $|L_k| = j_k$ for all $k \in \{1, \ldots, K\}$ and that $m_{kl} \in \mathbb{N}$ for all $k \in \{1, \ldots, K\}$ and $l \in L_k$ with $\sum_{l \in L_k} m_{kl} = i_k$ for all $k \in \{1, \ldots, K\}$. There are $\binom{k_l}{m_{kl}}$ possibilities to select among the k_l boxes of color l those m_{kl} boxes which will contain a ball of type k. There are $i_k!$ possibilities to distribute the i_k balls of type k among the $\sum_{l \in L_k} m_{kl} = i_k$ selected boxes. For the remaining $N - |i|$ balls, there are $(N - |i|)!$ possibilities to distribute them among the remaining $N - |i|$ free boxes. Therefore,

$$P(A_{ij} \mid \nu^{(r-1)} = k) = \frac{1}{N!} \sum_{L_1, \ldots, L_K} \sum_{m_{kl}} (N - |i|)! \prod_{k=1}^{K} \left(i_k! \prod_{l \in L_k} \binom{k_l}{m_{kl}} \right).$$

Here the first sum extends over all disjoint sets $L_1, \ldots, L_K \subseteq \{1, \ldots, N\}$ satisfying $|L_k| = j_k$ for all $k \in \{1, \ldots, K\}$ and the second sum extends over all $m_{kl} \in \mathbb{N}$, $k \in \{1, \ldots, K\}, l \in L_k$, with $\sum_{l \in L_k} m_{kl} = i_k$ for all $k \in \{1, \ldots, K\}$. Thus,

$$p_{ij} = \frac{(N - |i|)! \, i_1! \cdots i_K!}{N!} \sum_{L_1, \ldots, L_K} \sum_{m_{kl}} E\left(\prod_{k=1}^{K} \prod_{l \in L_k} \binom{\nu_l}{m_{kl}} \right).$$

Since the random variables $\nu_1, \ldots, \nu_N$ are exchangeable, the last expectation does not depend on the particular choice of the subsets $L_1, \ldots, L_K$. Thus, we can choose $L_1 := \{1, \ldots, j_1\}$, $L_2 := \{j_1 + 1, \ldots, j_1 + j_2\}$, and so on. Since there are exactly $N!/(j_1! \cdots j_K! (N - |j|)!)$ choices of subsets $L_1, \ldots, L_K$, it follows that

$$p_{ij} = \frac{(N - |i|)! \, i_1! \cdots i_K!}{(N - |j|)! \, j_1! \cdots j_K!} \sum_{m_{kl}} E\left(\prod_{k=1}^{K} \prod_{l=j_1 + \cdots + j_{k-1} + 1}^{j_1 + \cdots + j_k} \binom{\nu_l}{m_{kl}} \right).$$

The index substitution $m_l := m_{kl}$ completes the proof.

The sum on the right-hand side of (2) has $|j|$ summation variables. There is the following alternative formula for p_{ij} which involves only K summation variables.

Proposition 2. *For* $i = (i_1, \ldots, i_K)$, $j = (j_1, \ldots, j_K) \in S_{N,K}$,

$$p_{ij} = \frac{(N - |i|)! \, i_1! \cdots i_K!}{(N - |j|)! \, j_1! \cdots j_K!} \sum_{l_1=0}^{j_1} \cdots \sum_{l_K=0}^{j_K} \left(\prod_{k=1}^{K} (-1)^{j_k - l_k} \binom{j_k}{l_k} \right) E\left(\prod_{k=1}^{K} \binom{D_k}{i_k} \right), \quad (3)$$

where $|i| := i_1 + \cdots + i_K$, $|j| := j_1 + \cdots + j_K$, *and* $D_k := v_{r_{k-1}+1} + \cdots + v_{r_k}$ *for* $k \in \{1, \ldots, K\}$ *with* $r_0 := 0$ *and* $r_k := l_1 + \cdots + l_k$ *for* $k \in \{1, \ldots, K\}$.

Remark. If $K = 1$ then (3) reduces to (see [12, p. 766, Equation (7)])

$$p_{ij} = \frac{\binom{N}{j}}{\binom{N}{i}} \sum_{l=0}^{j} (-1)^{j-l} \binom{j}{l} \mathrm{E}\left(\binom{v_1 + \cdots + v_l}{i}\right), \qquad i, j \in \{0, \ldots, N\}.$$

Proof of Proposition 2. Fix $i, j \in S_{N,K}$. Define $s_0 := 0$ and $s_k := j_1 + \cdots + j_k$ for $k \in \{1, \ldots, K\}$. Note that $s_K = |j|$. Let A denote the set of all $m = (m_1, \ldots, m_{|j|}) \in \mathbb{N}_0^{|j|}$ satisfying $m_1 + \cdots + m_{s_1} = i_1$, $m_{s_1+1} + \cdots + m_{s_2} = i_2, \ldots, m_{s_{K-1}+1} + \cdots + m_{|j|} = i_K$. For $l \in \{1, \ldots, |j|\}$, define the subset $A_l \subset A$ via $A_l := \{m \in A \mid m_l = 0\}$ and denote $\overline{A}_l := A \setminus A_l$. For $m \in A$, define $Z_m := \binom{v_1}{m_1} \cdots \binom{v_{|j|}}{m_{|j|}}$. By (2) we have to manipulate the sum

$$s := \sum_{m \in \overline{A}_1 \cap \cdots \cap \overline{A}_{|j|}} \mathrm{E}(Z_m) = \sum_{m \in A} \mathrm{E}(Z_m) - \sum_{m \in A_1 \cup \cdots \cup A_{|j|}} \mathrm{E}(Z_m).$$

By the principle of inclusion and exclusion (Silvester's sieve formula), it follows that

$$s = \sum_{m \in A} \mathrm{E}(Z_m) - \sum_{n=1}^{|j|} (-1)^{n-1} \sum_{\substack{L \subseteq \{1, \ldots, |j|\} \\ |L|=n}} \sum_{m \in \bigcap_{l \in L} A_l} \mathrm{E}(Z_m)$$

$$= \sum_{n=0}^{|j|} (-1)^n \sum_{L_1, \ldots, L_K} \sum_{m \in \bigcap_{k=1}^K \bigcap_{l \in L_k} A_l} \mathrm{E}(Z_m),$$

where the sum $\sum_{L_1, \ldots, L_K}$ extends over all $L_1 \subseteq \{1, \ldots, s_1\}$, $L_2 \subseteq \{s_1 + 1, \ldots, s_2\}, \ldots,$ $L_K \subseteq \{s_{K-1} + 1, \ldots, s_K\}$ satisfying $\sum_{k=1}^K |L_k| = n$. From the exchangeability of the variables $v_1, \ldots, v_N$, it follows that, for $m \in \bigcap_{k=1}^K \bigcap_{l \in L_k} A_l$, the mean $\mathrm{E}(Z_m)$ depends only via $(n_1, \ldots, n_K) := (|L_1|, \ldots, |L_K|)$ on $L_1, \ldots, L_K$. Thus, we are allowed to compute $\mathrm{E}(Z_m)$ for the particular choice of subsets $L_1 := \{s_1 - n_1 + 1, \ldots, s_1\}$, $L_2 := \{s_2 - n_2 + 1, \ldots, s_2\}$, and so on, and multiply with the number $\binom{j_1}{n_1} \cdots \binom{j_K}{n_K}$ of all possible subsets $L_1, \ldots, L_K$ and obtain

$$s = \sum_{n=0}^{|j|} (-1)^n$$

$$\times \sum_{\substack{n_1, \ldots, n_K \in \mathbb{N}_0 \\ n_1 + \cdots + n_K = n}} \binom{j_1}{n_1} \cdots \binom{j_K}{n_K} \mathrm{E}\left(\prod_{k=1}^K \sum_{\substack{m_{s_{k-1}+1}, \ldots, m_{s_k-n_k} \in \mathbb{N}_0 \\ m_{s_{k-1}+1} + \cdots + m_{s_k-n_k} = i_k}} \binom{v_{s_{k-1}+1}}{m_{s_{k-1}+1}} \cdots \binom{v_{s_k-n_k}}{m_{s_k-n_k}}\right)$$

$$= \sum_{n=0}^{|j|} (-1)^n \sum_{\substack{n_1, \ldots, n_K \in \mathbb{N}_0 \\ n_1 + \cdots + n_K = n}} \binom{j_1}{n_1} \cdots \binom{j_K}{n_K} \mathrm{E}\left(\prod_{k=1}^K \binom{v_{s_{k-1}+1} + \cdots + v_{s_k-n_k}}{i_k}\right).$$

The index transformations $l_k := j_k - n_k$ and afterwards $l := |j| - n$ yield

$$
\begin{aligned}
s &= \sum_{n=0}^{|j|} (-1)^n \sum_{\substack{l_1,\dots,l_K \in \mathbb{N}_0 \\ l_1+\cdots+l_K=|j|-n}} \binom{j_1}{l_1} \cdots \binom{j_K}{l_K} \mathrm{E}\left(\prod_{k=1}^{K} \binom{\nu_{s_{k-1}+1} + \cdots + \nu_{s_{k-1}+l_k}}{i_k} \right) \\
&= \sum_{l=0}^{|j|} (-1)^{|j|-l} \sum_{\substack{l_1,\dots,l_K \in \mathbb{N}_0 \\ l_1+\cdots+l_K=l}} \binom{j_1}{l_1} \cdots \binom{j_K}{l_K} \mathrm{E}\left(\prod_{k=1}^{K} \binom{\nu_{s_{k-1}+1} + \cdots + \nu_{s_{k-1}+l_k}}{i_k} \right) \\
&= \sum_{l_1=0}^{j_1} \cdots \sum_{l_K=0}^{j_K} \left(\prod_{k=1}^{K} (-1)^{j_k-l_k} \binom{j_k}{l_k} \right) \mathrm{E}\left(\prod_{k=1}^{K} \binom{\nu_{s_{k-1}+1} + \cdots + \nu_{s_{k-1}+l_k}}{i_k} \right).
\end{aligned}
$$

By exchangeability we can replace $\nu_{s_{k-1}+1} + \cdots + \nu_{s_{k-1}+l_k}$ by D_k, which completes the proof.

Remark. The structure of the matrix $P = (p_{ij})_{i,j \in S_{N,K}}$ becomes clearer if the states of $S_{N,K}$ are ordered (differently from the lexicographical order) as follows. A state $i \in S_{N,K}$ is called absorbing if $p_{ij} = 0$ for all $j \in S_{N,K} \setminus \{i\}$. Note that, since P is in general only substochastic, this does not necessarily imply that $p_{ii} = 1$. It is readily checked (by using Proposition 1 or Proposition 2) that at least all the states $i = (i_1, \dots, i_K) \in \{0,1\}^K$ are absorbing. These are 2^K states. Let i be one of these 2^K absorbing states. To i is attached the set of all states $j = (j_1, \dots, j_K) \in S_{N,K}$ satisfying $j_k = 0$ if $i_k = 0$ and $j_k \geq i_k$ if $i_k = 1$, $k \in \{1, \dots, K\}$. These are the only states j from which the absorbing state i is possibly reachable. Note that, if exactly $m \in \{0, \dots, K\}$ entries of i are equal to 1 then there are $\binom{N}{m}$ such states j. Suppose that the states of $S_{N,K}$ are ordered with respect to the order induced by the 2^K absorbing states and the set of states attached to each of these absorbing states, for example, the 10 states of $S_{3,2}$ in the order $(0,0), (0,1), (0,2), (0,3), (1,0), (2,0), (3,0), (1,1), (1,2),$ and $(2,1)$. With respect to this order, P is a block diagonal matrix containing 2^K square blocks, where, for each $m \in \{0, \dots, K\}$, there are exactly $\binom{K}{m}$ square blocks of the same dimension $\binom{N}{m}$, in agreement with the formula $\sum_{m=0}^{K} \binom{K}{m}\binom{N}{m} = (K+N)!/(N!\,K!) = |S_{N,K}|$. Moreover, if the states inside each block are ordered lexicographically then each block itself is a left lower triangular matrix.

4. Duality

Fix $N, K \in \mathbb{N}$. We identify each element $(i_1, \dots, i_{K+1}) \in E_{N,K+1}$ with the element $(i_1, \dots, i_K) \in S_{N,K}$. In particular, $|E_{N,K+1}| = |S_{N,K}|$ and we can write the transition matrix of the forward process with $K+1$ types in the form $\Pi = (\pi_{ij})_{i,j \in S_{N,K}}$. Let $H = (h_{ij})_{i,j \in S_{N,K}}$ denote the matrix with entries

$$
h_{ij} := \frac{(i_1)_{j_1} \cdots (i_K)_{j_K}}{(N)_{|j|}}, \qquad i, j \in S_{N,K}, \tag{4}
$$

where $|j| := j_1 + \cdots + j_K$ for $j \in S_{N,K}$, $(x)_0 := 1$, and $(x)_j := x(x-1)\cdots(x-j+1)$ for $x \in \mathbb{R}$ and $k \in \mathbb{N}$. Assuming that the states of $S_{N,K}$ are ordered lexicographically, H is a left lower triangular regular matrix with determinant

$$
\det H = \prod_{i \in S_{N,K}} h_{ii} = \prod_{i \in S_{N,K}} \frac{i_1! \cdots i_K!}{(N)_{|i|}} > 0.
$$

The following result is from Gladstien [7, Lemma 4]. We provide an alternative proof based on the explicit formulae for π_{ij}, (1), and for p_{ij}, (3).

Proposition 3. (Duality.) *The transition matrix Π of the forward chain X (with $K + 1$ types) and the transpose $P^\top$ of the backward matrix P are similar with respect to H, i.e. $\Pi H = H P^\top$.*

Remark. For $K = 1$, the matrix P is stochastic and Proposition 3 is in fact a duality relation between two Markov chains (the forward chain X and a backward chain Y with transition matrix P) in the sense of Liggett [11]. We also refer the reader to [12] in this context. For $K > 1$, the matrix P is in general only substochastic, such that a backward chain Y cannot be defined without adding a kind of 'coffin state'.

Proof of Proposition 3. Fix $i = (i_1, \ldots, i_K)$, $j = (j_1, \ldots, j_K) \in S_{N,K}$, and define the random vector $C = (C_1, \ldots, C_K)$ as in Section 2. Then, by (1) and (4),

$$(\Pi H)_{ij} = \sum_{l \in S_{N,K}} \pi_{il} h_{lj} = \sum_{l \in S_{N,K}} \mathrm{P}(C = l) \frac{(l_1)_{j_1} \cdots (l_K)_{j_K}}{(N)_{|j|}} = \frac{\mathrm{E}((C_1)_{j_1} \cdots (C_K)_{j_K})}{(N)_{|j|}}.$$

On the other hand, by (3),

$$
\begin{aligned}
(H P^\top)_{ij} &= \sum_{l \in S_{N,K}} h_{il} p_{jl} \\
&= \sum_{l \in S_{N,K}} \frac{(i_1)_{l_1} \cdots (i_K)_{l_K}}{(N)_{|l|}} \frac{(N - |j|)! \, j_1! \cdots j_K!}{(N - |l|)! \, l_1! \cdots l_K!} \\
&\quad \times \sum_{m_1=0}^{l_1} \cdots \sum_{m_K=0}^{l_K} \left(\prod_{k=1}^{K} (-1)^{l_k - m_k} \binom{l_k}{m_k} \right) \mathrm{E}\left(\prod_{k=1}^{K} \binom{D_k}{j_k} \right) \\
&= \frac{1}{(N)_j} \sum_{m_1=0}^{i_1} \cdots \sum_{m_K=0}^{i_K} \mathrm{E}\left(\prod_{k=1}^{K} (D_k)_{j_k} \right) \sum_{l \in S_{N,K}} \prod_{k=1}^{K} (-1)^{l_k - m_k} \binom{i_k}{l_k} \binom{l_k}{m_k}. \quad (5)
\end{aligned}
$$

The last sum simplifies to

$$
\begin{aligned}
\sum_{l \in S_{N,K}} (-1)^{l_k - m_k} \binom{i_k}{l_k} \binom{l_k}{m_k} &= \prod_{k=1}^{K} \sum_{l_k=0}^{i_k} (-1)^{l_k - m_k} \binom{i_k}{l_k} \binom{l_k}{m_k} \\
&= \prod_{k=1}^{K} \binom{i_k}{m_k} \sum_{l_k=m_k}^{i_k} (-1)^{l_k - m_k} \binom{i_k - m_k}{i_k - l_k} \\
&= \prod_{k=1}^{K} \binom{i_k}{m_k} \sum_{s_k=0}^{i_k - m_k} (-1)^{i_k - m_k - s_k} \binom{i_k - m_k}{s_k} \\
&= \prod_{k=1}^{K} \binom{i_k}{m_k} (1 - 1)^{i_k - m_k} \\
&= \delta_{im},
\end{aligned}
$$

where δ_{im} denotes the Kronecker symbol. Thus, only the multi-index $m = i$ provides a nonzero

contribution to (5). But, for $m = i$, the random vector D coincides with C and it follows from (5) that

$$(HP^\top)_{ij} = \frac{E((C_1)_{j_1} \cdots (C_K)_{j_K})}{(N)_{|j|}}.$$

Thus, $\Pi H = HP^\top$ and the proof is complete.

In some cases it might be useful to derive the backward matrix P directly from the forward matrix Π via $P^\top = H^{-1}\Pi H$. This approach is of course only advisable if the inverse H^{-1} of the matrix H is known. Of course, H, as well as H^{-1}, is left lower triangular. The following proposition provides an explicit formula for the entries of the inverse of the matrix H with entries (4).

Proposition 4. *Suppose that H has entries (4). Then, for $i = (i_1, \ldots, i_K)$, $j = (j_1, \ldots, j_K) \in S_{N,K}$,*

$$(H^{-1})_{ij} = (-1)^{|i|-|j|} \frac{(N)_{|i|}}{i_1! \cdots i_K!} \binom{i_1}{j_1} \cdots \binom{i_K}{j_K}, \tag{6}$$

where $|i| := i_1 + \cdots + i_K$ and $|j| := j_1 + \cdots + j_K$.

Proof. Let b_{ij} denote the right-hand side of (6), and set $B := (b_{ij})_{i,j \in S_{N,K}}$. For $i, j \in S_{N,K}$, it follows that

$$
\begin{aligned}
(HB)_{ij} &= \sum_{l \in S_{N,K}} h_{il} b_{lj} \\
&= \sum_{l \in S_{N,K}} \frac{(i_1)_{l_1} \cdots (i_K)_{l_K}}{(N)_{|l|}} (-1)^{|l|-|j|} \frac{(N)_{|l|}}{l_1! \cdots l_K!} \binom{l_1}{j_1} \cdots \binom{l_K}{j_K} \\
&= \sum_{l_1=j_1}^{i_1} \cdots \sum_{l_K=j_K}^{i_K} \prod_{k=1}^{K} (-1)^{l_k-j_k} \binom{i_k}{j_k} \binom{i_k - j_k}{l_k - j_k} \\
&= \prod_{k=1}^{K} \binom{i_k}{j_k} \sum_{l_k=j_k}^{i_k} (-1)^{l_k-j_k} \binom{i_k - j_k}{l_k - j_k} \\
&= \prod_{k=1}^{K} \binom{i_k}{j_k} (1 - 1)^{i_k-j_k} \\
&= \delta_{ij}.
\end{aligned}
$$

Thus, $HB = E$ (identity matrix). Similarly, it follows that $BH = E$. Thus, $B = H^{-1}$.

Remark. Since, by Proposition 4, the inverse H^{-1} is known explicitly, the duality relation $\Pi H = HP^\top$ can also be viewed via $\Pi = HP^\top H^{-1}$ and $P^\top = H^{-1}\Pi H$ as a number theoretical inversion formula between Π and $P^\top$. We will see in the following section that, for typical examples, this inversion formula involves (generalized) Stirling numbers of the second kind.

Remark. It might be useful to replace the matrix H with entries (4) by a more general matrix H with entries

$$h_{ij} = \frac{\binom{i_1}{j_1} \cdots \binom{i_K}{j_K}}{w(j)}, \qquad i, j \in S_{N,K}, \tag{7}$$

where the $w(j)$, $j \in S_{N,K}$, are some given nonzero weights. Note that (4) corresponds to the particular weights $w(j) := (N)_{|j|}/(j_1! \cdots j_K!)$, $j \in S_{N,K}$. Adapting the proof of Proposition 4 to this more general matrix H shows that the inverse H^{-1} exists and has entries

$$(H^{-1})_{ij} = (-1)^{|i|-|j|} w(i) \binom{i_1}{j_1} \cdots \binom{i_1}{j_K}, \qquad i, j \in S_{N,K}.$$

Similar as in the proof of Proposition 3, it follows that

$$(\Pi H)_{ij} = \frac{1}{w(j)} \mathrm{E}\left(\binom{C_1}{j_1} \cdots \binom{C_K}{j_K} \right).$$

Again, there exists a matrix $P = (p_{ij})_{i,j \in S_{N,K}}$ which satisfies the duality relation $\Pi H = H P^\top$. The entries of P are

$$
\begin{aligned}
p_{ij} &= (H^{-1} \Pi H)_{ji} \\
&= \sum_{l \in S_{N,K}} (H^{-1})_{jl} (\Pi H)_{li} \\
&= \sum_{l \in S_{N,K}} (-1)^{|j|-|l|} w(j) \binom{j_1}{l_1} \cdots \binom{j_K}{l_K} \frac{1}{w(i)} \mathrm{E}\left(\binom{D_1}{i_1} \cdots \binom{D_K}{i_K} \right) \\
&= \frac{w(j)}{w(i)} \sum_{l_1=0}^{j_1} \cdots \sum_{l_K=0}^{j_K} \left(\prod_{k=1}^{K} (-1)^{j_k-l_k} \binom{j_k}{l_k} \right) \mathrm{E}\left(\prod_{k=1}^{K} \binom{D_k}{i_k} \right),
\end{aligned}
$$

in agreement with Proposition 2 for the particular weights $w(j) = (N)_{|j|}/(j_1! \cdots j_K!)$, $j \in S_{N,K}$. Note that, the diagonal entries p_{ii}, $i \in S_{N,K}$, do not depend on the weights $w(j)$, $j \in S_{N,K}$. Thus, the diagonal entries are always of the form $p_{ii} = \mathrm{E}(\nu_1 \cdots \nu_{|i|})$ no matter how the weights are chosen. Suppose now that $K > 1$ and that $\mathrm{P}(\nu_1 = 1) < 1$. Then, for the particular absorbing state $i = (1, 1, 0, \ldots, 0)$, we obtain

$$\sum_{j \in S_{N,K}} p_{ij} = p_{ii} = \mathrm{E}(\nu_1 \nu_2) < 1.$$

Therefore, for $K > 1$, the matrix P is not stochastic no matter how the weights are chosen, except for the trivial model in which each individual produces exactly one offspring.

5. Examples

In this section the multi-allelic versions of six fundamental examples of Cannings models (a simple model, Moran model, Wright–Fisher model, Kimura model, Karlin and McGregor model, and uniform model) are studied. Explicit formulae for the forward transition matrix Π, the backward matrix P, and their eigenvalues are derived. Note that all the results derived in the previous sections, in particular the duality result (Proposition 3), apply to all these examples.

5.1. A simple model

We start with a somewhat nonnatural but simple Cannings model in which it is assumed that the offspring vector $\nu = (\nu_1, \ldots, \nu_N)$ is a random permutation of $(0, \ldots, 0, N)$, i.e. one randomly chosen individual has N offspring, whereas all other $N - 1$ individuals do not have

any offspring. It is easily checked that the forward transition matrix $\Pi = (\pi_{ij})_{i,j \in E_{N,K}}$ has entries

$$\pi_{ij} = \begin{cases} \dfrac{i_k}{N} & \text{if } j_k = N \text{ for some } k \in \{1, \ldots, K\}, \\ 0 & \text{otherwise.} \end{cases}$$

For $k \in \{1, \ldots, K\}$, let e_k denote the kth unit vector in $\mathbb{R}^K$. Straightforward considerations show that the matrix $P = (p_{ij})_{i,j \in S_{N,K}}$ has entries

$$p_{ij} = \begin{cases} 1 & \text{if } (i, j) = (0, 0), \\ 1 & \text{if } i = i_k e_k \text{ and } j = e_k \text{ for some } k \in \{1, \ldots, K\} \text{ and some } i_k \in \{1, \ldots, N\}, \\ 0 & \text{otherwise,} \end{cases}$$

in agreement with Proposition 1. In particular, for $K > 1$, the matrix P has several rows with only zero entries, namely the ith row has only zero entries whenever the vector $i = (i_1, \ldots, i_K)$ has at least two nonzero components. The matrix P is hence highly substochastic.

5.2. Moran model

We consider first the following Moran model (see [7, p. 636]) with overlapping generations and constant population size $N \in \mathbb{N} \setminus \{1\}$. At each time $r \in \mathbb{N}_0$ one randomly chosen individual contributes one offspring to the next generation $r + 1$. Afterwards one individual—not the same individual who produced the offspring—is randomly selected and removed from the population. We are interested in the multi-allelic version of the model, so each individual has one of K possible types and each child inherits the type of its parent. Fortunately, this model can be interpreted as a model with nonoverlapping generations by identifying individuals who are still alive in the next generation as being children in this next generation. More precisely, the model can be seen as a Cannings model, where the offspring vector $\nu = (\nu_1, \ldots, \nu_N)$ is a random permutation of $(0, 1, 1, \ldots, 1, 1, 2)$. Let e_k denote the kth unit vector in $\mathbb{R}^K$, $k \in \{1, \ldots, K\}$. It is readily seen from the definition of the model and can alternatively be derived from (1) that the multi-allelic forward Markov chain X moves from the state $i \in E_{N,K}$ to the state $j \in E_{N,K}$ with transition probability

$$\pi_{ij} = \begin{cases} \displaystyle\sum_{k=1}^{K} \dfrac{i_k(i_k - 1)}{N(N-1)} & \text{if } j = i, \\ \dfrac{i_k i_l}{N(N-1)} & \text{if } j = i + e_k - e_l \text{ for some } k, l \in \{1, \ldots, K\} \text{ with } k \neq l, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

We now turn to the backward matrix $P = (p_{ij})_{i,j \in S_{N,K}}$. By (2), for $i \in S_{N,K}$,

$$\begin{aligned} p_{ii} &= \mathrm{E}(\nu_1 \cdots \nu_{|i|}) \\ &= \mathrm{P}(\nu_1 = \cdots = \nu_{|i|} = 1) + 2\,\mathrm{P}(\nu_1 \cdots \nu_{|i|} = 2) \\ &= \frac{(N - |i|)(N - |i| - 1)}{N(N-1)} + 2\frac{|i|(N - |i|)}{N(N-1)} \\ &= 1 - \frac{|i|(|i| - 1)}{N(N-1)}. \end{aligned}$$

Suppose now that $i, j \in S_{N,K}$ with $j = i - e_k$ for some $k \in \{1, \ldots, K\}$. Note that $|j| = |i| - 1$ and that $j_k = i_k - 1$. Then, (2) yields

$$
\begin{aligned}
p_{ij} &= \frac{(N-|i|)! \, i_k!}{(N-|j|)! \, j_k!} \sum_{s=j_1+\cdots+j_{k-1}+1}^{j_1+\cdots+j_k} \mathrm{E}\left(\nu_1 \cdots \nu_{s-1} \binom{\nu_s}{2} \nu_{s+1} \cdots \nu_{|j|} \right) \\
&= \frac{i_k}{N-|j|} \sum_{s=j_1+\cdots+j_{k-1}+1}^{j_1+\cdots+j_k} \mathrm{P}(\nu_1 = 1, \ldots, \nu_{s-1} = 1, \, \nu_s = 2, \, \nu_{s+1} = 1, \ldots, \nu_{|j|} = 1) \\
&= \frac{i_k}{N-|j|} j_k \frac{N-|j|}{N(N-1)} \\
&= \frac{i_k(i_k-1)}{N(N-1)}.
\end{aligned}
$$

All other entries p_{ij} are equal to 0, so in summary, for $i, j \in S_{N,K}$,

$$
p_{ij} = \begin{cases}
1 - \dfrac{|i|(|i|-1)}{N(N-1)} & \text{if } j = i, \\[2mm]
\dfrac{i_k(i_k-1)}{N(N-1)} & \text{if } j = i - e_k \text{ for some } k \in \{1, \ldots, K\}, \\[2mm]
0 & \text{otherwise.}
\end{cases}
$$

In particular, the Moran model has effective population size $N_{\mathrm{e}} := 1/p_{21} = N(N-1)/2$. Note that

$$
\sum_{j \in S_{N,K}} p_{ij} = 1 - \frac{1}{N(N-1)} \sum_{\substack{k,l=1 \\ k \neq l}}^{K} i_k i_l.
$$

Thus, for $K > 1$, the matrix P is not stochastic but substochastic. The eigenvalues of P are

$$
p_{ii} = 1 - \frac{|i|(|i|-1)}{N(N-1)}, \qquad i \in S_{N,K}.
$$

By the duality result of Section 4, these are also the eigenvalues of the transition matrix of the forward chain X with $K+1$ types. The classical Moran model (see [13] and [7, p. 635]) slightly differs from the Moran model described above, since, in the classical version, the individual who is removed from the population is randomly selected from the complete population and is hence possibly the same individual who produced the offspring. Fortunately, the classical Moran model can also be interpreted as a Cannings model as follows. In each generation, with probability $1 - 1/N$, the population evolves according to the Moran model described above and, with the complementary probability $1/N$, the population evolves according to the trivial Cannings model in which each individual produces exactly one offspring. In other words, the classical Moran model is a convex mixture of the Moran model described above and the trivial Cannings model. Note that the corresponding offspring vector $\nu = (\nu_1, \ldots, \nu_N)$ of the classical Moran model has distribution $\mathrm{P}(\nu = k) = 1/N$ if $k = (k_1, \ldots, k_N) = (1, \ldots, 1)$ and $\mathrm{P}(\nu = k) = (1 - 1/N)(N-2)!/N! = 1/N^2$ if $k = (k_1, \ldots, k_N)$ with $k_i = 2$, $k_j = 0$, and $k_l = 1$ for all $l \in \{1, \ldots, N\} \setminus \{i, j\}$ for some $i, j \in \{1, \ldots, N\}$ with $i \neq j$. It is easily seen that, for the classical multi-allelic Moran model, the forward chain X moves from $i \in E_{N,K}$ to

$j \in E_{N,K}$ with transition probability

$$
\tilde{\pi}_{ij} = \begin{cases}
\dfrac{\sum_{k=1}^{K} i_k^2}{N^2} & \text{if } j = i, \\[2mm]
\dfrac{i_k i_l}{N^2} & \text{if } j = i + e_k - e_l \text{ for some } k, l \in \{1, \ldots, K\} \text{ with } k \neq l, \\[2mm]
0 & \text{otherwise.}
\end{cases}
$$

Note that $\tilde{\pi}_{ij} = (1 - 1/N)\pi_{ij} + (1/N)\delta_{ij}$, where π_{ij} is defined via (8) and δ_{ij} denotes the Kronecker symbol, in agreement with the interpretation of the classical Moran model as a convex mixture. Similarly, the entries $\tilde{p}_{ij}$, $i, j \in S_{N,K}$, of the backward matrix $\tilde{P}$ of the multi-allelic classical Moran model are

$$
\tilde{p}_{ij} = \left(1 - \frac{1}{N}\right)p_{ij} + \frac{\delta_{ij}}{N}
$$

$$
= \begin{cases}
1 - \dfrac{|i|(|i| - 1)}{N^2} & \text{if } j = i, \\[2mm]
\dfrac{i_k(i_k - 1)}{N^2} & \text{if } j = i - e_k \text{ for some } k \in \{1, \ldots, K\}, \\[2mm]
0 & \text{otherwise.}
\end{cases}
$$

Note that the classical Moran model has effective population size $\tilde{N}_e := 1/\tilde{p}_{21} = N^2/2$. By the duality result of Section 4, the transition matrix of the forward chain X (with $K + 1$ types) has eigenvalues $\tilde{p}_{ii} = 1 - |i|(|i| - 1)/N^2$, $i \in S_{N,K}$.

5.3. Wright–Fisher model

In the Wright–Fisher model [14] the offspring vector $\nu = (\nu_1, \ldots, \nu_N)$ is assumed to have a symmetric multinomial distribution. Note that ν has probability generating function $E(z_1^{\nu_1} \cdots z_N^{\nu_N}) = ((z_1 + \cdots + z_N)/N)^N$, $z_1, \ldots, z_N \in \mathbb{C}$. The random vector $C = (C_1, \ldots, C_K)$ introduced in Section 2 has again a multinomial distribution with parameters N and $i_1/N, \ldots, i_K/N$, and, by (1), the forward chain X has multinomial transition probabilities

$$
\pi_{ij} = \frac{N!}{j_1! \cdots j_K!} \left(\frac{i_1}{N}\right)^{j_1} \cdots \left(\frac{i_K}{N}\right)^{j_K}, \qquad i, j \in E_{N,K}.
$$

Moreover,

$$
E\left(\binom{\nu_1}{m_1} \cdots \binom{\nu_{|j|}}{m_{|j|}}\right) = \binom{N}{|i|} \frac{(|i|)!}{m_1! \cdots m_{|j|}!} \left(\frac{1}{N}\right)^{|i|}
$$

and it follows from (2) that the backward matrix P has entries

$$
p_{ij} = \frac{(N - |i|)! \, i_1! \cdots i_K!}{(N - |j|)! \, j_1! \cdots j_K!} \sum_m \binom{N}{|i|} \frac{(|i|)!}{m_1! \cdots m_{|j|}!} \left(\frac{1}{N}\right)^{|i|}
$$

$$
= \frac{N!}{(N - |j|)!} \left(\frac{1}{N}\right)^{|i|} \prod_{k=1}^{K} \left(\frac{i_k!}{j_k!} \sum_{\substack{m_{s_{k-1}+1}, \ldots, m_{s_k} \in \mathbb{N} \\ m_{s_{k-1}+1} + \cdots + m_{s_k} = i_k}} \frac{1}{m_{s_{k-1}+1}! \cdots m_{s_k}!}\right)
$$

$$
= (N)_{|j|} N^{-|i|} \prod_{k=1}^{K} S(i_k, j_k), \qquad i, j \in S_{N,K}, \tag{9}
$$

where (see also the remark after Lemma 1 in Appendix A)

$$S(m, n) = \frac{m!}{n!} \sum_{\substack{m_1,\ldots,m_n \in \mathbb{N} \\ m_1+\cdots+m_n=m}} \frac{1}{m_1! \cdots m_n!} = \frac{1}{n!} \sum_{l=0}^{n} (-1)^{n-l} \binom{n}{l} l^m, \qquad m, n \in \mathbb{N}_0, \quad (10)$$

are the Stirling numbers of the second kind. Note that $S(m, 0) = \delta_{m0}$, $m \in \mathbb{N}_0$. In particular, P has eigenvalues $p_{ii} = (N)_{|i|} N^{-|i|}$, $i \in S_{N,K}$. The effective population size $N_e := 1/p_{21} = 1/(1 - p_{22}) = N$ coincides with the total population size.

5.4. Kimura model

The Kimura model with parameter $c \in \mathbb{N}$ (see [7, p. 636] or [10]) can be viewed as a Cannings model with symmetric multi-hypergeometric joint offspring distribution

$$P(\nu = k) = \frac{\binom{c}{k_1} \cdots \binom{c}{k_N}}{\binom{cN}{N}} \qquad (11)$$

for $k = (k_1, \ldots, k_N) \in \mathbb{N}_0^N$ with $k_1 + \cdots + k_N = N$. Kimura's model behaves quite similar to the Wright–Fisher model, except for the fact that the number of offspring per individual is uniformly bounded by c, since $P(\nu_i \leq c) = 1$ for all $i \in \{1, \ldots, N\}$. For $c \to \infty$, the distribution of ν weakly converges to the symmetric multinomial distribution. Thus, in the limit $c \to \infty$ the Kimura model coincides with the Wright–Fisher model. The random vector $C = (C_1, \ldots, C_K)$ introduced in Section 2 has a multi-hypergeometric distribution with parameters N, cN, and $ci_1, \ldots, ci_K$. From (1), it follows that the forward chain X has transition probabilities (see also [9])

$$\pi_{ij} = P(C = j) = \frac{\binom{ci_1}{j_1} \cdots \binom{ci_K}{j_K}}{\binom{cN}{N}}, \qquad i, j \in E_{N,K}.$$

Since the expectation under the sum $\sum_m$ on the right-hand side in (2) is given by (note that $m_1 + \cdots + m_{|j|} = |i|$)

$$E\left(\binom{\nu_1}{m_1} \cdots \binom{\nu_{|j|}}{m_{|j|}}\right) = \frac{\binom{N}{|i|}}{\binom{cN}{|i|}} \binom{c}{m_1} \cdots \binom{c}{m_{|j|}},$$

and, similarly, the random vector $D = (D_1, \ldots, D_K)$ in Proposition 2 satisfies

$$E\left(\binom{D_1}{i_1} \cdots \binom{D_K}{i_K}\right) = \frac{\binom{N}{|i|}}{\binom{cN}{|i|}} \binom{cl_1}{i_1} \cdots \binom{cl_K}{i_K},$$

it follows from (2) and (3) that

$$p_{ij} = \frac{(N)_{|j|}}{(cN)_{|i|}} \prod_{k=1}^{K} S_c(i_k, j_k), \qquad i, j \in S_{N,K}, \quad (12)$$

where, for $m, n \in \mathbb{N}_0$,

$$S_c(m, n) := \frac{m!}{n!} \sum_{\substack{m_1,\ldots,m_n \in \mathbb{N} \\ m_1+\cdots+m_n=m}} \binom{c}{m_1} \cdots \binom{c}{m_n} = \frac{m!}{n!} \sum_{l=0}^{j} (-1)^{n-l} \binom{n}{l} \binom{cl}{m} \quad (13)$$

is some sort of generalized Stirling number of the second kind. Note that $S_c(m, n) \sim c^m S(m, n)$ as $c \to \infty$, where $S(m, n)$ is the usual Stirling number of the second kind defined in (10). Thus, as $c \to \infty$, the Kimura p_{ij} given via (12) converges to the Wright–Fisher p_{ij} given via (9) as expected. From $S_c(m, m) = c^m$, it follows that the backward matrix P has eigenvalues $p_{ii} = c^{|i|}(N)_{|i|}/(cN)_{|i|}$, $i \in S_{N,K}$, extending the result of Gladstien [7, p. 636] and Karlin and McGregor [9] for $K = 1$. Note that $S_c(2, 1) = c(c - 1)$. Hence, $p_{21} = (c - 1)/(cN - 1)$ and, for $c \neq 1$, the Kimura model has effective population size $N_e := 1/p_{21} = (cN - 1)/(c - 1)$.

5.5. Karlin and McGregor model

Suppose that a sequence $X_1, X_2, \ldots$ of independent and identically distributed nonnegative integer-valued random variables is given. Let f denote the probability generating function (PGF) of X_1. For any PGF g and $n \in \mathbb{N}_0$, we use the standard notation g^n for the nth power of g ($g^0 = 1$) and the notation $g^{(n)}$ for the nth derivative of g. Moreover, $[x^n]g(x)$ denotes the coefficient in front of x^n in the Taylor expansion of g around 0. The conditional branching process model of Karlin and McGregor [9] is a Cannings model with joint offspring distribution

$$P(\nu = k) = \frac{P(X_1 = k_1) \cdots P(X_N = k_N)}{P(X_1 + \cdots + X_N = N)} = \frac{[x^{k_1}]f(x) \cdots [x^{k_N}]f(x)}{[x^N]f^N(x)} \tag{14}$$

for $k = (k_1, \ldots, k_N) \in \mathbb{N}_0^N$ with $k_1 + \cdots + k_N = N$, a fact already known in 1971 by Felsenstein [3, p. 399], although at that time the Cannings models had not been introduced yet. Felsenstein also mentioned that $\nu = (\nu_1, \ldots, \nu_N)$ has PGF

$$E(z_1^{\nu_1} \cdots z_N^{\nu_N}) = \frac{[x^N](f(xz_1) \cdots f(xz_N))}{[x^N]f^N(x)}, \qquad |z_1|, \ldots, |z_N| \leq 1. \tag{15}$$

If X_1 is Poisson distributed with some parameter $\alpha > 0$ then (independent of α) the model reduces to the Wright–Fisher model. If X_1 has a binomial distribution with PGF $f(x) = (px + 1 - p)^c$ for some parameters $c \in \mathbb{N}$ and $p \in (0, 1)$, then (independent of p) the model reduces to the Kimura model with parameter c. Thus, the Karlin and McGregor model is a generalization of the Wright–Fisher model and the Kimura model. In order to compute the entries of the forward transition matrix, fix $i, j \in E_{N,K}$ and, as in Section 2, define $s_0 := 0$ and $s_k := i_1 + \cdots + i_k$ for $k \in \{1, \ldots, K\}$ and put $Y_k := \sum_{s=s_{k-1}+1}^{s_k} X_s$ for $k \in \{1, \ldots, K\}$. Note that Y_k has PGF f^{i_k}. From (1), it follows that the forward transition matrix has entries

$$\pi_{ij} = \frac{P(Y_1 = j_1) \cdots P(Y_K = j_K)}{P(X_1 + \cdots + X_N = N)} = \frac{[x^{j_1}]f^{i_1}(x) \cdots [x^{j_K}]f^{i_K}(x)}{[x^N]f^N(x)}, \qquad i, j \in E_{N,K}. \tag{16}$$

For $K = 2$, we obtain

$$\pi_{ij} = \frac{[x^{j_1}]f^{i_1}(x)[x^{j_2}]f^{i_2}(x)}{[x^N]f^N(x)},$$

a formula which is slightly simpler than the equivalent expression at the top of page 637 of [7], since only one variable x is involved. Applying the 'derivative operator'

$$\frac{\partial^{m_1}}{\partial^{m_1} z_1} \cdots \frac{\partial^{m_N}}{\partial^{m_N} z_N}$$

to (15), and noting that it is allowed to interchange the 'coefficient operator $[x^N]$' with this 'derivative operator', it follows after choosing $z_1 = \cdots = z_N = 1$ that $\nu = (\nu_1, \ldots, \nu_N)$ has

joint descending factorial moments

$$\mathrm{E}((v_1)_{m_1} \cdots (v_N)_{m_N}) = \frac{[x^{N-(m_1+\cdots+m_N)}](f^{(m_1)}(x) \cdots f^{(m_N)}(x))}{[x^N]f^N(x)}, \qquad m_1, \ldots, m_N \in \mathbb{N}_0.$$

Substituting this expression into (2) shows that the backward matrix P has entries

$$p_{ij} = \frac{(N-|i|)! \, i_1! \cdots i_K!}{(N-|j|)! \, j_1! \cdots j_K!} \frac{1}{[x^N]f^N(x)} \sum_m \frac{[x^{N-|i|}](f^{N-|j|}(x) f^{(m_1)}(x) \cdots f^{(m_{|j|})}(x))}{m_1! \cdots m_{|j|}!},$$

$i, j \in S_{N,K}$, where the multi-index $m = (m_1, \ldots, m_{|j|})$ runs as explained in Proposition 1. This expression further simplifies to

$$p_{ij} = \frac{(N-|i|)!}{(N-|j|)!} \frac{[x^{N-|i|}](f^{N-|j|}(x) \prod_{k=1}^K S_{f,x}(i_k, j_k))}{[x^N]f^N(x)}, \tag{17}$$

where, for $m, n \in \mathbb{N}_0$,

$$S_{f,x}(m, n) := \frac{m!}{n!} \sum_{\substack{m_1, \ldots, m_n \in \mathbb{N} \\ m_1 + \cdots + m_n = m}} \frac{f^{(m_1)}(x)}{m_1!} \cdots \frac{f^{(m_n)}(x)}{m_n!} \tag{18}$$

is again a kind of generalized Stirling number of the second kind, with the convention that $S_{f,x}(m, 0) = \delta_{m0}$ for $m \in \mathbb{N}_0$. For example, if X_1 is Poisson distributed with PGF $f(x) = e^{\alpha(x-1)}$ for some parameter $\alpha > 0$, then $S_{f,x}(m, n) = \alpha^m e^{\alpha n(x-1)} S(m, n)$, where $S(m, n)$ denotes the usual Stirling number of the second kind defined in (10). If X_1 has a binomial distribution with PGF $f(x) = (px + 1 - p)^c$ for some parameters $c \in \mathbb{N}$ and $p \in (0, 1)$, then $S_{f,x}(m, n) = (px + 1 - p)^{nc-m} S_c(m, n)$, with $S_c(m, n)$ defined in (13). Since $S_{f,x}(m, m) = (f'(x))^m$, it follows that P has eigenvalues

$$p_{ii} = \frac{[x^{N-|i|}](f^{N-|i|}(x)(f'(x))^{|i|})}{[x^N]f^N(x)}, \qquad i \in S_{N,K},$$

generalizing the known results for $K = 1$ of [3, p. 400], [7, p. 637], and [9] to the multi-allelic case. From $S_{f,x}(2, 1) = f''(x)$ and (17), it follows that the Karlin and McGregor model has effective population size

$$N_e := \frac{1}{p_{21}} = (N-1) \frac{[x^N]f^N(x)}{[x^{N-2}](f^{N-1}(x)f''(x))}.$$

An alternative expression for p_{ij} is obtained from (3) as follows. Let $l_1, \ldots, l_K \in \mathbb{N}_0$ with $|l| := l_1 + \cdots + l_K \leq N$. Choosing in (15) the first l_1 variables $z_1, \ldots, z_{l_1}$ all equal to some given u_1, the next l_2 variables $z_{l_1+1}, \ldots, z_{l_1+l_2}$ all equal to some given $u_2, \ldots$, the next l_K variables $z_{l_1+\cdots+l_{K-1}+1}, \ldots, z_{|l|}$ all equal to some u_K, and the last $N - |l|$ variables $z_{|l|+1}, \ldots, z_N$ all equal to 1, it follows that the random vector $D = (D_1, \ldots, D_K)$ in Proposition 2 has PGF

$$\mathrm{E}(u_1^{D_1} \cdots u_K^{D_K}) = \frac{[x^N](f^{N-|l|}(x) f^{l_1}(xu_1) \cdots f^{l_K}(xu_K))}{[x^N]f^N(x)}, \qquad |u_1|, \ldots, |u_K| \leq 1.$$

The same method used before for the vector $v = (v_1, \ldots, v_N)$ yields the fact that D has joint factorial moments

$$\mathrm{E}((D_1)_{i_1} \cdots (D_K)_{i_K}) = \frac{[x^{N-|i|}](f^{N-|l|}(x)(f^{l_1})^{(i_1)}(x) \cdots (f^{l_K})^{(i_K)}(x))}{[x^N]f^N(x)}.$$

Substituting this expression into (3) yields (17) with

$$S_{f,x}(m,n) = \frac{1}{n!} \sum_{l=0}^{n} (-1)^{n-l} \binom{n}{l} f^{n-l}(x)(f^l)^{(m)}(x), \qquad m, n \in \mathbb{N}_0,$$

which indeed coincides with (18) by Lemma 1 given in Appendix A.

5.6. The uniform model and extensions

We finally mention a further particular Karlin and McGregor model. Suppose that X_1 has a geometric distribution with PGF $f(x) = p/(1-qx)$ for some parameter $p \in (0,1)$, where $q := 1 - p$. Then $X_1 + \cdots + X_N$ has a negative binomial distribution with parameters N and p, and it follows from (14) that $P(\nu = k) = 1/\binom{2N-1}{N}$ for all $k = (k_1, \ldots, k_N) \in \mathbb{N}_0^N$ with $k_1 + \cdots + k_N = N$. Thus, ν is uniformly distributed. Let us call this particular model the uniform Cannings model. Using (16), the transition probabilities of the forward chain turn out to be of the form

$$\pi_{ij} = \frac{\binom{i_1+j_1-1}{j_1} \cdots \binom{i_K+j_K-1}{j_K}}{\binom{2N-1}{N}}, \qquad i, j \in E_{N,K}.$$

Since f has derivatives $f^{(m)}(x) = m!\, pq^m/(1-qx)^{m+1}$, $m \in \mathbb{N}_0$, the generalized Stirling numbers in (18) are given by

$$S_{f,x}(m,n) = \frac{m!}{n!} \binom{m-1}{n-1} \frac{p^n q^m}{(1-qx)^{m+n}}, \qquad m, n \in \mathbb{N}_0$$

(with the convention $\binom{m-1}{-1} := \delta_{m0}$ for $m \in \mathbb{N}_0$), and it can be readily derived from (17) that

$$p_{ij} = \frac{N!\,(N-1)!}{(N+|i|-1)!\,(N-|j|)!} \prod_{k=1}^{K} \frac{i_k!}{j_k!} \binom{i_k-1}{j_k-1}, \qquad i, j \in S_{N,K}.$$

In particular, P has eigenvalues $p_{ii} = \binom{2N-1}{N-|i|}/\binom{2N-1}{N}$, $i \in S_{N,K}$. The effective population size is $N_e := 1/p_{21} = (N+1)/2$. The uniform model can be extended to a more general Karlin and McGregor model by replacing the geometric distribution of X_1 by a negative binomial distribution with PGF $f(x) = (p/(1-qx))^\alpha$, $\alpha > 0$, $p \in (0,1)$, $q := 1 - p$. Then $X_1 + \cdots + X_N$ has a negative binomial distribution with parameters αN and p, and it follows from (14) that

$$P(\nu = k) = \frac{\binom{\alpha+k_1-1}{k_1} \cdots \binom{\alpha+k_N-1}{k_N}}{\binom{\alpha N+N-1}{N}} = \frac{N!}{k_1! \cdots k_N!} \frac{[\alpha]_{k_1} \cdots [\alpha]_{k_N}}{[\alpha N]_N}$$

for all $k = (k_1, \ldots, k_N) \in \mathbb{N}_0^N$ with $k_1 + \cdots + k_N = N$, where $[x]_0 := 1$ and $[x]_k := x(x+1) \cdots (x+k-1)$ for $x \in \mathbb{R}$ and $k \in \mathbb{N}$. For $\alpha = 1$, we are back in the uniform model and, for $\alpha \to \infty$, this model converges to the Wright–Fisher model. Using (16), the transition probabilities of the forward chain turn out to be of the form

$$\pi_{ij} = \frac{\binom{\alpha i_1+j_1-1}{j_1} \cdots \binom{\alpha i_K+j_K-1}{j_K}}{\binom{\alpha N+N-1}{N}} = \frac{N!}{j_1! \cdots j_K!} \frac{[\alpha i_1]_{j_1} \cdots [\alpha i_K]_{j_K}}{[\alpha N]_N}, \qquad i, j \in E_{N,K}.$$

Since f has derivatives

$$f^{(m)}(x) = \frac{[\alpha]_m p^\alpha q^m}{(1-qx)^{\alpha+m}} = \frac{m! \binom{\alpha+m-1}{m} p^\alpha q^m}{(1-qx)^{\alpha+m}}, \qquad m \in \mathbb{N}_0,$$

the generalized Stirling numbers in (18) are given by

$$S_{f,x}(m,n) = \frac{p^{\alpha n} q^m}{(1-qx)^{\alpha n+m}} s_\alpha(m,n), \qquad m, n \in \mathbb{N}_0,$$

where

$$s_\alpha(m,n) := \frac{m!}{n!} \sum_{\substack{m_1,\dots,m_n \in \mathbb{N} \\ m_1+\cdots+m_n=m}} \binom{\alpha+m_1-1}{m_1} \cdots \binom{\alpha+m_n-1}{m_n}$$

$$= \frac{m!}{n!} \sum_{l=0}^n (-1)^{n-l} \binom{n}{l} \binom{\alpha l + m - 1}{m},$$

and it can be readily derived from (17) that

$$p_{ij} = \frac{(N)_{|j|}}{(\alpha N + |i| - 1)_{|i|}} \prod_{k=1}^K s_\alpha(i_k, j_k), \qquad i, j \in S_{N,K}.$$

In particular, P has eigenvalues $p_{ii} = \alpha^{|i|} \binom{\alpha N+N-1}{N-|i|} / \binom{\alpha N+N-1}{N}$, $i \in S_{N,K}$. From $p_{22} = \alpha(N-1)/(\alpha N+1)$, it follows that this model has effective population size $N_{\mathrm{e}} := 1/p_{21} = 1/(1-p_{22}) = (\alpha N+1)/(\alpha+1)$.

Appendix A

In this appendix a particular combinatorial identity is verified, which is a consequence of the principle of inclusion and exclusion and of Leibniz's derivative rule for products of functions. The corresponding formula (19) below is probably known from the combinatorics literature; however, we have not been able to find an appropriate reference.

Lemma 1. *Let $U \subseteq \mathbb{R}$ be an open set, and let $f : U \to \mathbb{R}$ be a function which is m times differentiable in a point $x \in U$ for some given $m \in \mathbb{N}_0$. Then, for all $n \in \mathbb{N}_0$,*

$$\sum_{\substack{m_1,\dots,m_n \in \mathbb{N} \\ m_1+\cdots+m_n=m}} \frac{m!}{m_1! \cdots m_n!} f^{(m_1)}(x) \cdots f^{(m_n)}(x)$$

$$= \sum_{l=0}^n (-1)^{n-l} \binom{n}{l} f^{n-l}(x) (f^l)^{(m)}(x), \tag{19}$$

with the convention that the left-hand side is equal to δ_{m0} for $n = 0$.

Remark. For $f(x) = \mathrm{e}^x$, (19) yields the two representations in (10) for the Stirling numbers of the second kind. For $f(x) = x^c$ for some fixed $c \in \mathbb{N}$, (19) yields the two representations in (13) for the numbers $S_c(m,n)$. For $f(x) = 1/x$, (19) leads to the combinatorial identity $\binom{m-1}{n-1} = \sum_{l=0}^n (-1)^{n-l} \binom{n}{l} \binom{l+m-1}{m}$, $m, n \in \mathbb{N}$.

Proof of Lemma 1. It is easily checked that (19) holds for $m = 0$ or $n = 0$. Suppose now that $m, n \in \mathbb{N}$. Define a finite signed measure μ on

$$\Omega = \Omega(m, n) := \{\omega = (m_1, \ldots, m_n) \in \mathbb{N}_0^n \colon m_1 + \cdots + m_n = m\}$$

via

$$\mu(\omega) := \frac{m!}{m_1! \cdots m_n!} f^{(m_1)}(x) \cdots f^{(m_n)}(x), \qquad \omega = (m_1, \ldots, m_n) \in \Omega.$$

Note that, by the Leibniz rule, μ has total mass $\mu(\Omega) = (f^n)^{(m)}(x) \in \mathbb{R}$. For $i \in \{1, \ldots, n\}$, let $A_i \subset \Omega$ denote the subset of all $\omega = (m_1, \ldots, m_n) \in \Omega$ satisfying $m_i = 0$. Let L denote the left-hand side of (19). From the definition of μ, it follows that

$$L = \mu(A_1^c \cap \cdots \cap A_n^c) = \mu(\Omega) - \mu(A_1 \cup \cdots \cup A_n),$$

where A_i^c denotes the complement of A_i, $i \in \{1, \ldots, n\}$. By the principle of inclusion and exclusion, which is applicable for finite signed measures, we obtain

$$L = \mu(\Omega) - \sum_{j=1}^n (-1)^{j-1} \sum_{1 \le i_1 < \cdots < i_j \le n} \mu(A_{i_1} \cap \cdots \cap A_{i_j}). \qquad (20)$$

By the definition of the signed measure μ,

$$\mu(A_{i_1} \cap \cdots \cap A_{i_j}) = \left(\prod_{\substack{k=1 \\ k \in \{i_1, \ldots, i_j\}}}^n \frac{f^{(0)}(x)}{0!} \right) \left(m! \prod_{\substack{k=1 \\ k \notin \{i_1, \ldots, j_j\}}}^n \frac{f^{(m_k)}(x)}{m_k!} \right) = f^j(x)(f^{n-j})^{(m)}(x),$$

where the last equation follows again from the Leibniz rule. Thus, (20) reduces to

$$L = \sum_{j=0}^n (-1)^j \binom{n}{j} f^j(x)(f^{n-j})^{(m)}(x).$$

The index transformation $l = n - j$ yields the result.

Acknowledgement

References

[1] 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.

[2] CANNINGS, C. (1975). The latent roots of certain Markov chains arising in genetics: a new approach. II. Further haploid models. *Adv. Appl. Prob.* **7,** 264–282.

[3] FELSENSTEIN, J. (1971). The rate of loss of multiple alleles in finite haploid populations. *Theoret. Pop. Biol.* **2,** 391–403.

[4] GLADSTIEN, K. (1976). Loss of alleles in a haploid population with varying environment. *Theoret. Pop. Biol.* **10,** 383–394.

[5] GLADSTIEN, K. (1977). Haploid populations subject to varying environment: the characteristic values and the rate of loss of alleles. *SIAM J. Appl. Math.* **32,** 778–783.

[6] GLADSTIEN, K. (1977). Subdivided populations: the characteristic values and rate of loss of alleles. *J. Appl. Prob.* **14,** 241–248.

[7] GLADSTIEN, K. (1978). The characteristic values and vectors for a class of stochastic matrices arising in genetics. *SIAM J. Appl. Math.* **34,** 630–642.

[8] HUILLET, T. AND MÖHLE, M. (2009). Duality and asymptotics for a class of nonneutral discrete Moran models. *J. Appl. Prob.* **46,** 866–893.

[9] KARLIN, S. AND McGREGOR, J. (1965). Direct product branching processes and related induced Markoff chains. I. Calculations of rates of approach to homozygosity. In *Proc. Internat. Res. Sem.*, Springer, New York, pp. 111–145.

[10] KIMURA, M. (1957). Some problems of stochastic processes in genetics. *Ann. Math. Statist.* **28,** 882–901.

[11] LIGGETT, T. M. (1985). *Interacting Particle Systems*. Springer, New York.

[12] MÖHLE, M. (1999). The concept of duality and applications to Markov processes arising in neutral population genetics models. *Bernoulli* **5,** 761–777.

[13] MORAN, P. A. P. (1958). Random processes in genetics. *Proc. Camb. Phil. Soc.* **54,** 60–71.

[14] WRIGHT, S. (1931). Evolution in Mendelian populations. *Genetics* **16,** 97–159.