The concept of duality and applications to Markov processes arising in neutral population genetics models

M A R T I N   M Ö H L E

Johannes Gutenberg-Universität Mainz, Fachbereich Mathematik, Saarstraße 21, 55099 Mainz, Germany. E-mail:moehle@mathematik.uni-mainz.de

One possible and widely used definition of the duality of Markov processes employs functions H relating one process to the other in a certain way. For given processes X and Y the space U of all such functions H, called the duality space of X and Y, is studied in this paper. The algebraic structure of U is closely related to the eigenvalues and eigenvectors of the transition matrices of X and Y. Often as for example in physics (interacting particle systems) and in biology (population genetics models) dual processes arise naturally by looking forwards and backwards in time. In particular, time-reversible Markov processes are self-dual. In this paper, results on the duality space are presented for classes of haploid and two-sex population models. For example $\dim U = N + 3$ for the classical haploid Wright–Fisher model with fixed population size N.

Keywords: centralizer; duality space; generalized Jordan matrix; haploid population models; interacting random processes; Markov chains with discrete parameter; Moran model; population dynamics; spectral form; two-sex population models; Wright–Fisher model

1. Introduction

The concept of duality is a powerful tool in the stochastic theory of interacting particle systems (Liggett 1985). Often dual processes arise when some system is considered forwards and backwards in time. In particular, time-reversible Markov processes are self-dual. Our aim in this paper is to link the theory of duality to finite population genetics models. Duality has been used by Donnelly and Kurtz (1996a,b) as well as by Krone and Neuhauser (1997a,b). These workers have concentrated on particle system models and genetics models arising from the limiting diffusion which is appropriate for large populations. In this paper a more general concept of duality is used and the results apply to finite population models.

Suppose that $X = (X_t)_{t \in T}$ and $Y = (Y_t)_{t \in T}$ are Markov processes with state spaces $(E_1, \mathscr{F}_1)$ and $(E_2, \mathscr{F}_2)$ respectively. Let $B(E_1 \times E_2)$ denote the set of all bounded measurable functions on $E_1 \times E_2$. The following definition of duality is due to Liggett (1985). Similar defining equations for duality have been used by Sudbury and Lloyd (1995).

Definition 1.1. *The process X is said to be dual to Y with respect to $H \in B(E_1 \times E_2)$ if*

$$\mathrm{E}^x H(X_t, y) = \mathrm{E}^y H(x, Y_t) \tag{1}$$

for all $x \in E_1$, $y \in E_2$ and $t \in T$, where E^x denotes the expectation given that the process X starts in $X_0 = x$ and E^y denotes the expectation given that the process Y starts in $Y_0 = y$.

Let $\{R_t\}_{t \in T}$ and $\{S_t\}_{t \in T}$ denote the semigroups of X and Y respectively, i.e. $R_t f(x) = \mathrm{E}^x f(X_t)$ and $S_t g(y) = \mathrm{E}^y g(Y_t)$ for all functions f on E_1 and g on E_2 such that the expectations exist. Each function $H \in B(E_1 \times E_2)$ induces a family of functions $H(\cdot, y) \in B(E_1)$, $y \in E_2$ and a family of functions $H(x, \cdot) \in B(E_2)$, $x \in E_1$. Hence the semigroups induce operators on $B(E_1 \times E_2)$ which are denoted again by R_t and S_t for simplicity and are defined by $R_t H(x, y) := \mathrm{E}^x H(X_t, y)$ and $S_t H(x, y) := \mathrm{E}^y H(x, Y_t)$. Equation (1) becomes then simply

$$R_t H = S_t H \quad \forall t \in T. \tag{2}$$

For given processes X and Y the set

$$U = U(X, Y) := \{H \in B(E_1 \times E_2) | X \text{ is dual to } Y \text{ with respect to } H\} \tag{3}$$

of all functions H satisfying (1) is of special interest in this paper. If X is dual to Y with respect to H, then Y is dual to X with respect to F given by $F(y, x) := H(x, y)$. Hence, $U(X, Y)$ is isomorphic to $U(Y, X)$. Obviously U is a linear subspace of $B(E_1 \times E_2)$. Call U the *duality space* of X and Y. One major interest is to characterize U, e.g. to find a basis of U (in terms of the processes X and Y) or expressions for the dimension of U. The space U might be very large, e.g. $U = B(E_1 \times E_2)$ for the simple case when X and Y are both constant.

If the state spaces E_1 and E_2 are countable, then each $H \in B(E_1 \times E_2)$ corresponds to a bounded matrix with entries $H_{ij} = H(i, j)$, $i \in E_1$, $j \in E_2$. Hence, for Markov processes X and Y with countable state spaces, (1) is equivalent to

$$\sum_{k \in E_1} \pi_{ik}(t) H_{kj} = \sum_{l \in E_2} H_{il} p_{jl}(t) \quad \forall i \in E_1, j \in E_2, t \in T,$$

where $\pi_{ik}(t) := P(X_t = k | X_0 = i)$ and $p_{jl}(t) := P(Y_t = l | Y_0 = j)$ are the transition probabilities of X and Y respectively. It is convenient to write this in matrix notation

$$\Pi_t H = H P_t' \quad \forall t \in T, \tag{4}$$

where $\Pi_t := (\pi_{ik}(t))_{i,k \in E_1}$ and $P_t := (p_{jl}(t))_{j,l \in E_2}$ are the transition matrices of the processes X and Y respectively and P_t' denotes the transpose of P_t. For time-homogeneous Markov processes with $T = \mathbb{N}_0 := \{0, 1, 2, \ldots\}$, (4) is equivalent to $\Pi H = H P'$, where Π and P are the corresponding one-step transition matrices. Note that, in this case, $\Pi H \in U$ and $H P' \in U$, if $H \in U$. Equation (4) is a "countable state-space version" of the semigroup representation (2) and can be considered as a more algebraic definition of the duality, but the more measure-theoretic definitions (1) and (2) are certainly more general. Nevertheless, as our interest here is focused on Markov processes with countable or even finite state spaces, (4) links the concept of duality directly to the theory of (linear) algebra

which turns out to be quite helpful. Two simple examples of duality are presented now. The first is an example with infinite countable state space, and the second is a corresponding finite state-space version.

Examples. Fix $p \in (0, 1)$ and define $q := 1 - p$.

(1) Let $X = (X_t)_{t \in \mathbb{N}_0}$ be a random walk on $E := \mathbb{N}_0$ with transition probabilities $\pi_{00} := 1$ (absorption at 0) and $\pi_{i,i+1} = 1 - \pi_{i,i-1} := p$ for all $i \in \mathbb{N}$. Further let $Y = (Y_t)_{t \in \mathbb{N}_0}$ be a random walk on $\mathbb{N}_0$ with transition probabilities $p_{00} := p$, $p_{01} := q$ (reflection at 0 with probability q) and $p_{i,i+1} = 1 - p_{i,i-1} := q$ for all $i \in \mathbb{N}$. Then for each $H: \mathbb{N}_0 \times \mathbb{N}_0 \to \mathbb{R}$ it follows that

$$(\Pi H)_{ij} = \sum_{k=0}^{\infty} \pi_{ik} H_{kj} = \begin{cases} H_{0j} & \text{if } i = 0, \\ q H_{i-1,j} + p H_{i+1,j} & \text{if } i > 0, \end{cases}$$

and

$$(HP')_{ij} = \sum_{k=0}^{\infty} H_{ik} p_{jk} = \begin{cases} p H_{i0} + q H_{i1} & \text{if } j = 0, \\ p H_{i,j-1} + q H_{i,j+1} & \text{if } j > 0. \end{cases}$$

Thus $\Pi H = HP'$, i.e. $H \in U$ for example for the function $H: E^2 \to \mathbb{R}$ given by

$$H(i, j) := \begin{cases} 1 & \text{if } i \leqslant j, \\ 0 & \text{otherwise,} \end{cases} \qquad H^{-1}(i, j) = \begin{cases} 1 & \text{if } i = j, \\ -1 & \text{if } i = j - 1, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The duality space U is generated by the matrices $\Pi^n H$, $n \in \mathbb{N}_0$.

(2) It is now obvious how to construct an example with finite state space $E := \{0, \ldots, N\}$. Let X be a random walk on E with transition probabilities $\pi_{00} := 1$ (absorption at 0), $\pi_{i,i+1} = 1 - \pi_{i,i-1} := p$ for all $i \in \{1, \ldots, N-1\}$, $\pi_{N,N-1} := q$ and $\pi_{NN} := p$ (reflection at N with probability q). Further let $Y = (Y_t)_{t \in \mathbb{N}_0}$ be a random walk on E with transition probabilities $p_{00} := p$, $p_{01} := q$ (reflection at 0 with probability q), $p_{i,i+1} = 1 - p_{i,i-1} := q$ for all $i \in \{1, \ldots, N-1\}$ and $p_{NN} := 1$ (absorption at N). In other words, Y looks like X viewed in reverse, i.e. $p_{ij} = \pi_{N-i,N-j}$. Then again $\Pi H = HP'$ for the function $H: E^2 \to \mathbb{R}$ defined as in (5). The duality space U is generated by the matrices $\Pi^n H$, $n \in \{0, \ldots, N\}$. Especially $\dim U = N + 1$.

Remark. A simple example with state space $E = [0, \infty)$ (Brownian motion with absorption at 0 and Brownian motion with reflection at 0) has been described by Liggett (1985). The analogy to the above random walk example with countable state space is obvious.

2. Haploid population models

We consider first the haploid population models with non-overlapping generations and fixed population size $N \in \mathbb{N}$ introduced by Cannings (1974; 1975).

2.1. The forward process

Looking forward in time this leads to the so-called descendant process or forward process $(X_n^{(i)})_{n\in\mathbb{N}_0}$, where $X_n^{(i)}$ denotes (by definition) the number of descendants in generation n of the individuals $1, \ldots, i$ of generation 0. The transition probabilities are given by

$$\pi_{jk} := P(X_{n+1}^{(i)} = k | X_n^{(i)} = j) = P(\nu_1 + \cdots + \nu_j = k) \tag{6}$$

(Cannings 1974), where ν_i denotes the number of offspring of individual i. Note that

$$\mathrm{E}(X_{n+1}^{(i)} | X_n^{(i)} = j) = \sum_k k P(\nu_1 + \cdots + \nu_j = k) = \mathrm{E}(\nu_1 + \cdots + \nu_j) = j,$$

i.e. $(X_n^{(i)})_{n\in\mathbb{N}_0}$ is a martingale which converges almost surely to a random variable $X_\infty^{(i)}$ as n tends to infinity such that $(X_n^{(i)})_{n\in\mathbb{N}_0\cup\{\infty\}}$ is still a martingale. Let $\pi_{ij}^{(n)} := P(X_n^{(i)} = j)$ denote the n-step transition probabilities of the forward process.

Lemma 2.1. *The following conditions are equivalent.*

(a) *There exists a constant $n \in \mathbb{N}$ such that $\inf_{0\leqslant i\leqslant N}(\pi_{i0}^{(n)} + \pi_{iN}^{(n)}) > 0$.*
(b) *$P(X_\infty^{(i)} \in \{0, N\}) = 1$ for all $i \in \{0, \ldots, N\}$.*
(c) *$\lim_{n\to\infty}\inf_{0\leqslant i\leqslant N}(\pi_{i0}^{(n)} + \pi_{iN}^{(n)}) = 1$.*
(d) *The states $1, \ldots, N-1$ are transient.*

Remark. The next lemma shows that these conditions are satisfied except for the trivial model $\nu_1 \equiv 1$.

Proof. "(a) $\Rightarrow$ (b)": from (a) it follows that $\alpha := 1 - \inf_{0\leqslant i\leqslant N}(\pi_{i0}^{(n)} + \pi_{iN}^{(n)}) < 1$. The homogeneity of the forward process $(X_n^{(i)})_{n\in\mathbb{N}_0}$ yields

$$P(0 < X_{nm}^{(i)} < N) = \sum_{j=1}^{N-1} P(0 < X_{nm}^{(i)} < N | X_{n(m-1)}^{(i)} = j) P(X_{n(m-1)}^{(i)} = j)$$

$$= \sum_{j=1}^{N-1} P(0 < X_n^{(j)} < N) P(X_{n(m-1)}^{(i)} = j)$$

$$= \sum_{j=1}^{N-1} (1 - \pi_{j0}^{(n)} - \pi_{jN}^{(n)}) P(X_{n(m-1)}^{(i)} = j)$$

$$\leqslant \alpha P(0 < X_{n(m-1)}^{(i)} < N)$$

for all $m \in \mathbb{N}$. By induction on m it follows that $P(0 < X_{nm}^{(i)} < N) \leqslant \alpha^m$. Note that $X_{nm}^{(i)}$ converges almost surely and hence also in distribution to $X_\infty^{(i)}$ as m tends to infinity. Thus $P(0 < X_\infty^{(i)} < N) \leqslant \lim_{m\to\infty}\alpha^m = 0$ and therefore $P(X_\infty^{(i)} \in \{0, N\}) = 1$.

"(b) $\Rightarrow$ (c)": obviously

$$\lim_{n\to\infty}(\pi_{i0}^{(n)} + \pi_{iN}^{(n)}) = \lim_{n\to\infty}(P(X_n^{(i)} = 0) + P(X_n^{(i)} = N))$$

$$= P(X_\infty^{(i)} = 0) + P(X_\infty^{(i)} = N)$$

$$\overset{(b)}{=} 1$$

for all $i \in \{0, \ldots, N\}$. As the state space $\{0, \ldots, N\}$ is finite, (c) is obtained.

"(c) $\Rightarrow$ (a)": this is trivial.

"(b) $\Leftrightarrow$ (d)": obviously

$$P(X_n^{(i)} \in \{0, N\} \text{ finally}) = P\left(\bigcup_{n\in\mathbb{N}_0}\{X_n^{(i)} \in \{0, N\}\}\right)$$

$$= \lim_{n\to\infty} P(X_n^{(i)} \in \{0, N\})$$

$$= P(X_\infty^{(i)} \in \{0, N\})$$

and hence

$$P(0 < X_n^{(i)} < N \text{ infinitely often}) = P(0 < X_\infty^{(i)} < N).$$

If all the states $1, \ldots, N-1$ are transient, then by definition the left-hand side of the above equation is equal to zero. Hence the right-hand side of the equation has to be equal to zero and (b) is established. Assume now that (b) is satisfied. Then for each $j \in \{1, \ldots, N-1\}$

$$P(X_n^{(i)} = j \text{ infinitely often}) \leq P(0 < X_n^{(i)} < N \text{ infinitely often})$$

$$= P(0 < X_\infty^{(i)} < N) \overset{(b)}{=} 0,$$

i.e. j is transient. $\qquad\square$

Lemma 2.2. *If $P(\nu_1 = 1) < 1$, then the conditions of Lemma 2.1 are satisfied, i.e. the states $1, \ldots, N-1$ are transient, the states 0 and N are absorbing and $X_\infty^{(i)}$ takes only the two values 0 and N with probability $q_i := P(X_n^{(i)} = 0 \text{ finally})$ and $1 - q_i$ respectively.*

Proof. For all $i \in \{1, \ldots, N-1\}$ it follows that

$$\sum_{j=i+1}^{N} \pi_{ij} = P(\nu_1 + \cdots + \nu_i > i) > 0,$$

i.e. from a state $i \in \{1, \ldots, N-1\}$ some state $j > i$ is reachable with positive probability.

As the state space is finite there exists a constant $n \in \mathbb{N}$ such that $\pi_{iN}^{(n)} > 0$. Hence condition
(a) of Lemma 2.1 is satisfied. The rest of the lemma follows immediately.   □

Remark. If $P(\nu_1 = 1) < 1$, then $i = \mathrm{E}(X_0^{(i)}) = \mathrm{E}(X_\infty^{(i)}) = N(1 - q_i)$, i.e. for all haploid models
(except the trivial one) the extinction probability is given by $q_i = 1 - i/N$.

2.2. The backward process

For $n, r \in \mathbb{N}_0$ with $n \leqslant r$ let $R_n^{(r)}$ denote the number of ancestors of all the individuals of
generation r in generation $r - n$, i.e. n generations backwards in time. This leads to the so-
called backward process $(R_n^{(r)})_{n \in \{0,\ldots,r\}}$ with transition probabilities (Cannings 1974, Theorem
11; Gladstien 1978, Examples)

$$p_{ij} := P(R_{n+1}^{(r)} = j \,|\, R_n^{(r)} = i)$$

$$= \left\{ \binom{N}{j} \bigg/ \binom{N}{i} \right\} \sum_{\substack{m_1,\ldots,m_j \in N \\ m_1 + \cdots + m_j = i}} \mathrm{E}\left(\binom{\nu_1}{m_1} \cdots \binom{\nu_j}{m_j} \right),$$

where $i, j \in \{0, \ldots, N\}$. An application of the principle of inclusion and exclusion
(also known as the sieve formula of Sylvester) leads to the simpler but alternating
summation

$$p_{ij} = \left\{ \binom{N}{j} \bigg/ \binom{N}{i} \right\} \sum_{l=0}^{j} (-1)^{j-l} \binom{j}{l} \mathrm{E}\left(\binom{\nu_1 + \cdots + \nu_l}{i} \right). \qquad (7)$$

3. Duality in haploid population models

Let $\Pi = (\pi_{ij})_{i,j \in \{0,\ldots,N\}}$ denote the transition matrix of the forward process and let
$P = (p_{ij})_{i,j \in \{0,\ldots,N\}}$ denote the transition matrix of the backward process.

Theorem 3.1. *There exists a non-singular symmetric left upper triangular matrix H such that*
$\Pi H = H P'$, *i.e. the forward process is dual to the backward process with respect to H.*

Remark. The matrix $H = (H_{ij})_{i,j \in \{0,\ldots,N\}}$ is given by

$$H_{ij} := \binom{N-i}{j} \bigg/ \binom{N}{j} = \binom{N-j}{i} \bigg/ \binom{N}{i} = \prod_{k=0}^{i-1} \frac{N-j-k}{N-k} = \prod_{k=0}^{j-1} \frac{N-i-k}{N-k}. \qquad (8)$$

These entries have the following probabilistic interpretation. Put i black and $N - i$ white
balls in a box. Choose j balls randomly without replacement. Then H_{ij} is the probability that

all the j chosen balls are white (hypergeometric distribution). The inverse of H is a symmetric right lower triangular matrix with entries

$$(H^{-1})_{ij} = (-1)^{i+j-N} \binom{i}{N-j} \binom{N}{i} = (-1)^{i+j-N} \binom{j}{N-i} \binom{N}{j}.$$

Proof. Obviously H is a symmetric left upper triangular matrix with

$$\det H = (-1)^{N(N+1)/2} \prod_{i=0}^{N} H_{i,N-i} = (-1)^{N(N+1)/2} \prod_{i=0}^{N} \left[\binom{N}{i}^{-1} \right] \neq 0.$$

Two proofs are presented now. The first is a straightforward calculation based on (6) and (7) for the transition probabilities. The second is based on a more general probabilistic method which turns out in Section 4 to work also for other (two-sex) classes of models.

Alternative 1. Fix $i, k \in \{0, \ldots, N\}$ and let $C_i := \nu_1 + \cdots + \nu_i$ denote the number of children of the individuals $1, \ldots, i$ of generation 0. Obviously

$$(\Pi H)_{ik} = \sum_{j=0}^{N} \pi_{ij} H_{jk}$$

$$= \sum_{j=0}^{N-k} P(C_i = j) \binom{N-j}{k} \Big/ \binom{N}{k} \tag{9}$$

$$= \mathrm{E}\left(\binom{N-C_i}{k} \right) \Big/ \binom{N}{k}.$$

On the other hand it follows from (7) that

$$(HP')_{ik} = \sum_{j=0}^{N} H_{ij} p_{kj}$$

$$= \sum_{j=0}^{N-i} \left\{ \binom{N-i}{j} \Big/ \binom{N}{j} \right\} \left\{ \binom{N}{j} \Big/ \binom{N}{k} \right\} \sum_{l=0}^{j} (-1)^{j-l} \binom{j}{l} \mathrm{E}\left(\binom{C_l}{k} \right)$$

$$= \binom{N}{k}^{-1} \sum_{l=0}^{N-i} \mathrm{E}\left(\binom{C_l}{k} \right) \sum_{j=l}^{N-i} (-1)^{j-l} \binom{N-i}{j} \binom{j}{l}$$

$$= \binom{N}{k}^{-1} \sum_{l=0}^{N-i} \mathrm{E}\left(\binom{C_l}{k} \right) \binom{N-i}{l} \sum_{j=l}^{N-i} (-1)^{j-l} \binom{N-i-l}{j-l}$$

$$= \binom{N}{k}^{-1} \sum_{l=0}^{N-i} \mathrm{E}\left(\binom{C_l}{k}\right)\binom{N-i}{l}(1-1)^{N-i-l}$$

$$= \binom{N}{k}^{-1} \mathrm{E}\left(\binom{C_{N-i}}{k}\right)$$

$$= \left\{1 \Big/ \binom{N}{k}\right\} \mathrm{E}\left(\binom{N-C_i}{k}\right) \Big/ \binom{N}{k},$$

as $C_{N-i} = \nu_1 + \cdots + \nu_{N-i} = N - (\nu_{N-i+1} + \cdots + \nu_N) \overset{d}{=} N - (\nu_1 + \cdots + \nu_i) = N - C_i$.
Hence $\Pi H = HP'$.

Alternative 2. Fix $i, k \in \{0, \ldots, N\}$ and $n, r \in \mathbb{N}_0$ such that $n \le r$. Consider the event E, that there exist in generation n exactly k ancestors, i.e. $R^{(r)}_{r-n} = k$, but none of these k ancestral individuals is descended from one of the individuals $1, \ldots, i$ of generation 0. In order to prove the theorem the probability of E is now calculated in two different ways. Obviously

$$E = \bigcup_{j=0}^{N} \{R^{(r)}_{r-n} = k \text{ and } X^{(i)}_n = j, \text{ but none of these } j \text{ descendants belongs to the } k \text{ ancestors}\}$$

and from the exchangeability of the model it follows that

$$P(E) = \sum_{j=0}^{N} P(R^{(r)}_{r-n} = k)P(X^{(i)}_n = j)\binom{N-k}{j}\Big/\binom{N}{j}$$

$$= P(R^{(r)}_{r-n} = k) \sum_{j=0}^{N} \pi^{(n)}_{ij} H_{jk}.$$

On the other hand

$$E = \bigcup_{j=0}^{N} \{R^{(r)}_{r-n} = k \text{ and } R^{(r)}_r = j,$$

$$\text{but none of these } j \text{ ancestors belongs to the individuals } 1, \ldots, i\}$$

and therefore (under the assumptions of the model)

$$P(E) = \sum_{j=0}^{N} \left\{\binom{N-i}{j}\Big/\binom{N}{j}\right\} P(R^{(r)}_r = j, R^{(r)}_{r-n} = k)$$

$$= \sum_{j=0}^{N} H_{ij} P(R^{(r)}_r = j, R^{(r)}_{r-n} = k).$$

Division by $P(R_{r-n}^{(r)} = k)$ yields $\sum_{j=0}^{N} \pi_{ij}^{(n)} H_{jk} = \sum_{j=0}^{N} H_{ij} p_{kj}^{(n)}$, i.e. $\Pi^n H = H(P^n)'$. $\qquad\square$

Remarks. The matrix H given in (8) is not the only one which satisfies the equation $\Pi H = HP'$. Cannings (1974) and Gladstien (1976, 1977, 1978) showed that $\Pi H = HP'$ for the matrix H with entries

$$H_{ij} := \binom{i}{j} \Big/ \binom{N}{j} = \prod_{k=0}^{j-1} \frac{i-k}{N-k}, \qquad (H^{-1})_{ij} = (-1)^{i-j} \binom{i}{j} \binom{N}{i}. \qquad (10)$$

Their proof is a simple modification of that given for Theorem 3.1. For haploid models the matrix (10) is in some sense more natural to consider. The probablistic interpretation is obvious, but we shall point out later that the matrix in (8) is useful for two-sex models as well, while the matrix (10) will not satisfy $\Pi H = HP'$ for those models any longer.

Whenever a non-singular matrix H exists such that $\Pi H = HP'$, then the eigenvalues of Π are identical with those of P. As for the haploid models considered here, P is triangular; these eigenvalues are given by $\lambda_k = p_{kk}$, $k \in \{0, \ldots, N\}$. This is the method used by Cannings and by Gladstien to calculate the eigenvalues (and eigenvectors) of Π for these models.

In any case there exists a non-singular matrix X such that $X\Pi = JX$, where the matrix J is the so-called *spectral form* or *generalized Jordan form* of Π (Gabriel 1996, p. 79). From Theorem 3.1 it follows that P' is similar to the same spectral form J, i.e. there exists a non-singular matrix Y such that $P'Y = YJ$. Now, if $\Pi H = HP'$, then $JXHY = X\Pi HY = XHP'Y = XHYJ$, i.e. $XHY \in C(J)$, where $C(J) := \{H | JH = HJ\}$ denotes the so-called *centralizer* of J. Thus the duality space $U = \{H | \Pi H = HP'\}$ is via $H \mapsto XHY$ isomorphic to the centralizer of J. As the structure of the spectral form J is quite simple, this provides a general method to characterize the duality space. The following theorem deals with the special case when $J = D$ is diagonal.

Theorem 3.2. *If Π (or equivalently P) is diagonalizable, then the duality space U is isomorphic to $V_0 \times \cdots \times V_N$, i.e.*

$$U \cong V_0 \times \cdots \times V_N,$$

where the $V_k := \{x \in \mathbb{R}^{N+1} | \Pi x = \lambda_k x\}$, $k \in \{0, \ldots, N\}$ are the right eigenspaces of Π. In particular, $\dim U = \sum_{k=0}^{N} \dim V_k \geq N + 1$.

Remark. By Theorem 3.1 there exists a non-singular matrix H such that $\Pi H = HP'$. If x is a right eigenvector of P' for λ_k, then $\Pi Hx = HP'x = H\lambda_k x = \lambda_k Hx$, i.e. Hx is then a right eigenvector of Π for λ_k. Hence $W_k := \{x \in \mathbb{R}^{N+1} | P'x = \lambda_k x\}$ is via $x \mapsto Hx$ isomorphic to V_k and $U \cong \times_{k=0}^{N} W_k$ under the conditions of Theorem 3.2.

Proof. As P (and hence P') is diagonalizable, there exists a non-singular matrix Y such that $P'Y = YD$, where D is diagonal with entries $d_{kk} = \lambda_k$. Obviously the duality space $U = \{H | \Pi H = HP'\}$, is via $\alpha(H) := HY$ isomorphic to the space $V := \{H | \Pi H = HD\}$. Now consider the map $\beta \colon V \to \times_{k=0}^{N} V_k$ defined by $\beta(H) = (v_0, \ldots, v_N)$, where v_k denotes

the kth column of H. Note that $H \in V$, i.e. $\Pi H = HD$ if and only if $\Pi v_k = \lambda_k v_k$, i.e. $v_k \in V_k$ for all $k \in \{0, \dots, N\}$. Thus β is an isomorphism and hence $U \overset{\alpha}{\cong} V \overset{\beta}{\cong} \times_{k=0}^{N} V_k$. $\qquad\square$

Example. For the trivial model $\nu_1 \equiv 1$ it follows that $\Pi = P = I$, $V_k = \mathbb{R}^{N+1}$ and hence $\dim U = \sum_{k=0}^{N} \dim V_k = \sum_{k=0}^{N} (N+1) = (N+1)^2$, which follows also directly from $\Pi H = H = HP'$ for all H.

From now on we assume that $P(\nu_1 = 1) < 1$.

Corollary 3.3. *If the non-unit eigenvalues are distinct, then Π is diagonalizable and $\dim U = N + 3$.*

Proof. Let $1 = \lambda_0 = \lambda_1 > \lambda_2 > \cdots > \lambda_N$ denote the eigenvalues and V_k, $k \in \{0, \dots, N\}$, the corresponding eigenspaces of Π. The eigenspace $V_0 (= V_1)$ is generated by the two vectors $v_0 := (1, \dots, 1)'$ and $v_1 := (q_0, \dots, q_N)'$, where the $q_i := P(X_n^{(i)} = 0$ finally), $i \in \{0, \dots, N\}$ are the extinction probabilities. Note that $q_N = 0$ and hence v_0 and v_1 are linear independent. For $k \in \{2, \dots, N\}$ choose an eigenvector v_k for λ_k. If X denotes the matrix with columns v_k, then it follows that $\Pi X = XD$, i.e. Π is diagonalizable. The previous theorem yields $\dim U = \sum_{k=0}^{N} \dim V_k = 2 + 2 + \sum_{k=2}^{N} 1 = N + 3$. $\qquad\square$

Examples. For the haploid Moran model the eigenvalues are given by $\lambda_k = p_{kk} = 1 - k(k-1)/N^2$. For the haploid Wright–Fisher model the eigenvalues are given by $\lambda_k = p_{kk} = (N)_k N^{-k}$, where $(N)_0 := 1$ and $(N)_k := N(N-1) \cdots (N-k+1)$ for all $k \in \mathbb{N}$. Hence the non-unit eigenvalues are distinct and $\dim U = N + 3$ for both models.

From Theorem 3.2 it seems to follow that finding a basis of U is (at least) as difficult as finding the right eigenvectors of Π or P'. Some results for the eigenvectors have been given by Gladstien (1978). The following theorem gives detailed information about the structure of the duality space U.

Theorem 3.4. *If the non-unit eigenvalues are distinct, then the $N + 3$ matrices H_1, H_2, H_3 and $\Pi^n H$, $n \in \{0, \dots, N-1\}$, built a basis of U, where H is given as in (8) or (10), $(H_1)_{ij} := \delta_{0j}$, $(H_2)_{ij} := q_i \delta_{0j}$, $(H_3)_{ij} := 1 - \delta_{0j}$ and δ_{ij} denotes the Kronecker symbol.*

Proof. Obviously H_1, H_2, $H_3 \in U$. Further $H \in U$ and hence $\Pi^n H \in U$ for all $n \in \mathbb{N}_0$. By Corollary 3.3 it is sufficient to verify that the $N + 3$ matrices H_1, H_2, H_3 and $\Pi^n H$, $n \in \{0, \dots, N-1\}$ are linear independent. This is satisfied if the N matrices $\Pi^n H$, $n \in \{0, \dots, N-1\}$ are linear independent. Assume that this is not the case. Then there exist constants a_n, $n \in \{0, \dots, N-1\}$ such that $\sum_{n=0}^{N-1} a_n \Pi^n H = 0$. As H is non-singular it follows that $g(\Pi) := \sum_{n=0}^{N-1} a_n \Pi^n = 0$. Hence the minimal polynomial m_Π of Π satisfies

$$\operatorname{grad} m_\Pi \leqslant \operatorname{grad} g = N - 1,$$

but Π is diagonalizable and the non-unit eigenvalues $\lambda_2, \ldots, \lambda_N$ are distinct. Hence

$$m_\Pi = \prod_{k=1}^{N}(x - \lambda_k)$$

and grad $m_\Pi = N$. This is obviously a contradiction. $\qquad\square$

In any case the matrices (8) and (10) belong to U and these two are linear independent as the first has a left upper triangular structure and the second has a left lower triangular structure. The entries of both matrices have a probabilistic interpretation via an urn model, where balls are sampled without replacement. One might think about a similar urn model where the balls are sampled with replacement. The corresponding matrices are then given by

$$H_{ij} := \left(\frac{i}{N}\right)^{j} \tag{11}$$

and

$$H_{ij} := \left(1 - \frac{i}{N}\right)^{j}. \tag{12}$$

Both matrices are "Vandermonde" matrices and hence non-singular. Unfortunately for most models these two matrices do not belong to U. A nice exception is the Wright–Fisher model.

Proposition 3.5. *For the Wright–Fisher model the matrices* (11) *and* (12) *belong to U.*

Proof. The proof is given for the matrix (12) here. For the other matrix the proof is similar. Fix $i, k \in \{0, \ldots, N\}$. Obviously

$$(\Pi H)_{ik} = \sum_{j=0}^{N} \pi_{ij} H_{jk}$$

$$= \sum_{j=0}^{N} P(X_1^{(i)} = j)\left(1 - \frac{j}{N}\right)^{k}$$

$$= \mathrm{E}\left(\left(1 - \frac{X_1^{(i)}}{N}\right)^{k}\right)$$

$$= N^{-k}\mathrm{E}((N - X_1^{(i)})^{k})$$

$$= N^{-k}\sum_{j=0}^{N} \mathrm{E}((N - X_1^{(i)})_j)S(k, j),$$

where $S(k, j)$ denotes the Stirling numbers of the second kind. The variable $N - X_1^{(i)}$ counts the number of individuals in generation 1, which are not descended from one of the individuals $1, \ldots, i$ of generation 0. In the Wright–Fisher model this variable is binomially distributed with the parameters N and $1 - i/N$. The factorial moments are therefore given by $E((N - X_1^{(i)})_j) = (N)_j(1 - i/N)^j$. Thus it follows that

$$(\Pi H)_{ik} = \sum_{j=0}^{N} \left(1 - \frac{i}{N}\right)^j S(k, j)(N)_j N^{-k} = \sum_{j=0}^{N} H_{ij} p_{kj} = (HP')_{ik},$$

where the formula $p_{ij} = S(i, j)(N)_j N^{-i}$ for the transition probabilities of the backward process has been used. Thus $\Pi H = HP'$, i.e. $H \in U$. $\square$

Remark. Another proof of the above proposition is based on the following observation. The matrix H given in (8) belongs to U. Hence also $\Pi H \in U$, but from (9) it follows that

$$(\Pi H)_{ik} = E\left(\left(\begin{array}{c} N - X_1^{(i)} \\ k \end{array}\right)\right) \bigg/ \left(\begin{array}{c} N \\ k \end{array}\right) = \left(1 - \frac{i}{N}\right)^k,$$

as $N - X_1^{(i)}$ is binomially distributed with the parameters N and $1 - i/N$.

4. Duality in two-sex population models

Consider a two-sex population model with non-overlapping generations $n \in \mathbb{N}_0$. Assume a fixed number of N pairs of individuals consisting of a male and a female. The N pairs of a generation n produce N daughters and N sons altogether, and these $2N$ children form the N pairs of the next generation $n + 1$ at random. Let $\nu_i^{(n)}$ and $\mu_i^{(n)}$ denote the number of daughters and sons of the pair i in generation n. The offspring vectors $(\nu_1^{(n)}, \mu_1^{(n)}, \ldots, \nu_N^{(n)}, \mu_N^{(n)})$ are requested to be independent and identically distributed for different generations and to satisfy $\sum_{i=1}^{N} \nu_i^{(n)} = N = \sum_{i=1}^{N} \mu_i^{(n)}$. Write ν_i for $\nu_i^{(0)}$ for convenience.

4.1. The forward process

Fix $i \in \{0, \ldots, N\}$ and choose i pairs from the generation 0. A pair is called a descendant pair if at least one individual is descended from one of these i pairs of generation 0. Let $X_n^{(i)}$ denote the number of descendant pairs in generation $n \in \mathbb{N}_0$. The transition probabilities of the forward process $(X_n^{(i)})_{n \in \mathbb{N}_0}$ are given by

$$\pi_{jk} := P(X_{n+1}^{(i)} = k | X_n^{(i)} = j)$$

$$= E\left(\binom{C_j}{C_j + D_j - k} \binom{N - C_j}{k - C_j} \Big/ \binom{N}{D_j} \right),$$

where $C_j := \nu_1 + \cdots + \nu_j$ and $D_j := \mu_1 + \cdots + \mu_j$ denote the number of daughters and the number of sons respectively of the pairs $1, \ldots, i$ of generation 0. Unfortunately the forward process is not a martingale any more (as is the case for the haploid models) but from the structure of the transition probabilities it follows that $(X_n^{(i)})_{n \in \mathbb{N}_0}$ is at least a submartingale. Thus $X_n^{(i)}$ converges again almost surely to some random variable $X_\infty^{(i)}$ as the time n tends to infinity. Lemma 2.1 is still valid for the two-sex models. It is not difficult to verify that there exists a constant $n \in \mathbb{N}$ such that $\pi_{iN}^{(n)} > 0$ for all $i \in \{0, \ldots, N\}$. Hence all the states $1, \ldots, N-1$ are transient and $X_\infty^{(i)}$ takes the two values 0 and N with probability q_i and $1 - q_i$ respectively, where $q_i := P(X_n^{(i)} = 0 \text{ finally})$ denotes the extinction probability.

4.2. The backward process

The backward process $(R_n^{(r)})_{n \in \{0, \ldots, r\}}$ has transition probabilities

$$p_{ij} := P(R_{n+1}^{(r)} = j | R_n^{(r)} = i) = \left\{ \binom{N}{j} \Big/ \binom{N}{i}^2 \right\} \sum_{l=0}^{j} (-1)^{j-l} \binom{j}{l} E\left(\binom{C_l}{i} \binom{D_l}{i} \right),$$

$i, j \in \{0, \ldots, N\}$ (Möhle 1994, Lemma 5.3).

Lemma 4.1. *The forward process is dual to the backward process with respect to H given in* (8).

Proof. For all $i, k \in \{0, \ldots, N\}$

$$(\Pi H)_{ik} = \sum_{j=0}^{N} \pi_{ij} H_{jk}$$

$$= \sum_{j=0}^{N} E\left(\binom{C_i}{C_i + D_i - j} \binom{N - C_i}{j - C_i} \Big/ \binom{N}{D_i} \right) \left\{ \binom{N - j}{k} \Big/ \binom{N}{k} \right\}$$

$$= E\left(\left\{ \binom{N - C_i}{k} \binom{N - D_i}{k} \Big/ \binom{N}{k}^2 \right\} \right.$$

$$\left. \times \sum_{j=0}^{N} \left\{ \binom{C_i}{C_i + D_i - j} \binom{N - k - C_i}{j - C_i} \Big/ \binom{N - k}{D_i} \right\} \right)$$

$$= \mathrm{E}\left(\binom{N-C_i}{k}\binom{N-D_i}{k}\right)\bigg/\binom{N}{k}^2.$$

A similar calculation to that in the first proof of Lemma 3.1 shows that the last expression is equal to $(HP')_{ik}$. □

Remarks. The above proof corresponds to the first alternative given for the haploid models in Lemma 3.1. The second proof of Lemma 3.1 works also for the two-sex models, if pairs (instead of individuals) are considered, but note that in general $\Pi H \neq HP'$ for the matrix (10) because of the definition of a "descendant pair". Both proofs break down for the matrix (10).

Theorem 3.2. and Corollary 3.3 are still valid for the two-sex models and they can in principle be used to characterize the duality space for two-sex models. Unfortunately for the two-sex models the transition matrix P is usually not triangular and hence more or less nothing is known about the eigenvalues. For a large class of two-sex population models, e.g. for the two-sex Wright–Fisher model, the non-unit eigenvalues seem to be all non-negative and distinct, but the author was not able to verify this conjecture analytically.

Finally, a modification of Proposition 3.5 for the two-sex Wright–Fisher model is presented.

Proposition 4.2. *For the two-sex Wright–Fisher model the matrix $H = (H_{ij})_{i,j\in\{0,\dots,N\}}$ with entries $H_{ij} = (1 - i/N)^{2j}$ belongs to U.*

Proof. This is a straightforward modification of the corresponding proof for the haploid Wright–Fisher model. Note that for the two-sex Wright–Fisher model $N - X_1^{(i)}$ is binomially distributed with the parameters N and $(1 - i/N)^2$ and that the backward process has transition probabilities $p_{ij} = S(2i, j)(N)_j N^{-2i}$. □

5. An application of the duality: stationary distribution

The duality is a powerful property for analysing the corresponding processes. As an example the behaviour of the processes for $n \to \infty$ is analysed here. It turns out that the ancestral process has a stationary distribution and that there is a one-to-one correspondence between the stationary distribution and the extinction probabilities of the forward process. The choice $k := N$ in the duality equation

$$\sum_{j=0}^{N} \pi_{ij}^{(n)} H_{jk} = \sum_{j=0}^{N} H_{ij} p_{kj}^{(n)}$$

leads to

$$P(X_n^{(i)} = 0) = \mathrm{E}\left(\binom{N-R_n}{i}\right)\bigg/\binom{N}{i}. \tag{13}$$

The *iN*th component of the matrix $(P^n)' = H^{-1}\Pi^n H$ is given by

$$P(R_n = i) = p_{Ni}^{(n)} = \sum_{k,l=0}^{N} (H^{-1})_{ik} \pi_{kl}^{(n)} H_{lN}$$

$$= \sum_{k=0}^{N} (H^{-1})_{ik} \pi_{k0}^{(n)}$$

$$= \binom{N}{i} \sum_{k=N-i}^{N} (-1)^{i+k-N} \binom{i}{N-k} P(X_n^{(k)} = 0) \qquad (14)$$

$$= \binom{N}{i} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} P(X_n^{(N-j)} = 0).$$

By definition the limit $q_i := \lim_{n\to\infty} P(X_n^{(i)} = 0)$ exists for each $i \in \{0, \dots, N\}$. Hence by (14) also the limit $p_i := \lim_{n\to\infty} P(R_n = i)$ exists for each $i \in \{0, \dots, N\}$. Further

$$\sum_{i=0}^{N} p_i = \sum_{i=0}^{N} \lim_{n\to\infty} P(R_n = i) = \lim_{n\to\infty} \sum_{i=0}^{N} P(R_n = i) = 1.$$

Thus there exists a random variable $\mathscr{R} = \mathscr{R}_N$ such that

$$P(\mathscr{R} = i) = p_i = \lim_{n\to\infty} P(R_n = i) \quad \forall i \in \{0, \dots, N\}.$$

Obviously the distribution of $\mathscr{R}$ is a stationary distribution of the backward process. Letting $n \to \infty$ in (13) and (14) leads to

$$q_i = \mathrm{E}\left(\binom{N-\mathscr{R}}{i}\right) \Big/ \binom{N}{i} = \sum_{j=0}^{N-i} \left\{ \binom{N-j}{i} \Big/ \binom{N}{i} \right\} p_j \qquad (15)$$

and

$$p_i = P(\mathscr{R} = i) = \binom{N}{i} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} q_{N-j}, \qquad (16)$$

where $i \in \{0, \dots, N\}$. This is a one-to-one correspondence between the extinction probabilities q_i of the forward process $(X_n^{(i)})_{n\in\mathbb{N}_0}$, $i \in \{0, \dots, N\}$ and the stationary distribution $p := (p_0, \dots, p_N)$ of the backward process.

Remark. For all haploid models (except the trivial one) the extinction probabilities are given by $q_i = 1 - i/N$. From (16) it follows that $p_1 = P(\mathscr{R} = 1) = 1$. This means that, after going sufficiently far backwards in time, one eventually reaches the most recent common ancestor of the population. This is also well known from the so-called coalescent theory (Kingman 1982a,b,c; Tavaré 1984). For the two-sex models the behaviour is quite different, as pairs instead of individuals are considered. The forward process $(X_n^{(i)})_{n\in\mathbb{N}_0}$ is then only a

submartingale and hence in general only the inequality $q_i \leqslant 1 - i/N$ is satisfied. In fact for many models as for example for the two-sex Wright–Fisher model the strict inequality $q_i < 1 - i/N$ holds for $i \in \{1, \ldots, N-1\}$. For the two-sex Wright–Fisher model it is further known (Möhle 1994, Theorem 4.5) that $\lim_{N \to \infty} q_i = x_0^i$, where $x_0 \approx 0.2032$ is the smallest fixed point of the probability generating function $s \mapsto e^{-2(1-s)}$ of a Poisson random variable with parameter 2 in the interval $[0, 1]$. The stationary distribution of the backward process is asymptotically normal (Möhle 1994, Theorem 9.2), i.e.

$$\frac{\mathscr{R}_N - \mathrm{E}(\mathscr{R}_N)}{\{\mathrm{var}(\mathscr{R}_N)\}^{1/2}} \xrightarrow[N \to \infty]{\mathrm{d}} N(0, 1) \tag{17}$$

or equivalently

$$\frac{\mathscr{R}_N - N(1 - x_0)}{N^{1/2}} \xrightarrow[N \to \infty]{\mathrm{d}} N\left(0, \frac{x_0(1 - x_0)}{1 + 2x_0}\right). \tag{18}$$

Nevertheless, the coalescent-theory is still applicable for two-sex models if genes instead of pairs or individuals are considered (Möhle 1998).

Acknowledgements

References

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

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

Donnelly, P. and Kurtz, T.G. (1996a) A countable representation of the Fleming–Viot measure-valued diffusion. *Ann. Probab.*, **24**, 698–742.

Donnelly, P. and Kurtz, T.G. (1996b) The asymptotic behavior of an urn model arising in population genetics. *Stochastic Processes Applic.*, **64**, 1–16.

Gabriel, P. (1996) *Matrizen, Geometrie, Lineare Algebra*. Basel: Birkhäuser.

Gladstien, K. (1976) Loss of alleles in a haploid population with varying environment. *Theor. Popul. Biol.*, **10**, 383–394.

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.

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

Kingman, J.F.C. (1982a) On the genealogy of large populations. *J. Appl. Probab. A.*, **19**, 27–43.

Kingman, J.F.C. (1982b) Exchangeability and the evolution of large populations. In: G. Koch and F.

Spizzichino (eds), *Exchangeability in Probability and Statistics*, pp. 97–112. Amsterdam: North–Holland.

Kingman, J.F.C. (1982c) The coalescent. *Stochastic Processes Applic*., **13**, 235–248.

Krone, S.M. and Neuhauser, C. (1997a) The genealogy of samples in models with selection. *Genetics*, **145**, 519–534.

Krone, S.M. and Neuhauser, C. (1997b) Ancestral processes with selection. *Theor. Popul. Biol*., **51**, 210–237.

Liggett, T.M. (1985) *Interacting Particle Systems*. Berlin: Springer-Verlag.

Möhle, M. (1994) Forward and backward processes in bisexual models with fixed population sizes. *J. Appl. Probab.* **31**, 309–332.

Möhle, M. (1998) Coalescent results for two-sexual population models. *Adv. Appl. Probab*, **30**, 513–520.

Sudbury, A. and Lloyd, P. (1995) Quantum operators in classical probability theory: II. The concept of duality in interacting particle systems. *Ann. Probab*., **23**, 1816–1830.

Tavaré, S. (1984) Line-of-descent and genealogical processes, and their applications in population genetics models. *Theor. Popul. Biol*., **26**, 119–164.