Research Article

A Duality Approach to the Genealogies of Discrete Nonneutral Wright-Fisher Models

Thierry E. Huillet

Laboratoire de Physique Théorique et Modélisation, Université de Cergy-Pontoise, CNRS-UMR 8089,
2 Avenue Adolphe Chauvin, 95302 Cergy-Pontoise Cedex, France

Correspondence should be addressed to Thierry E. Huillet, thierry.huillet@u-cergy.fr

Received 3 September 2008; Revised 5 November 2008; Accepted 6 November 2008

Recommended by Rongling Wu

Discrete ancestral problems arising in population genetics are investigated. In the neutral case, the
duality concept has been proved of particular interest in the understanding of backward in time
ancestral process from the forward in time branching population dynamics. We show that duality
formulae still are of great use when considering discrete nonneutral Wright-Fisher models. This
concerns a large class of nonneutral models with completely monotone (CM) bias probabilities.
We show that most classical bias probabilities used in the genetics literature fall within this CM
class or are amenable to it through some “reciprocal mechanism” which we define. Next, using
elementary algebra on CM functions, some suggested novel evolutionary mechanisms of potential
interest are introduced and discussed.

Copyright © 2009 Thierry E. Huillet. This is an open access article distributed under the Creative
Commons Attribution License, which permits unrestricted use, distribution, and reproduction in
any medium, provided the original work is properly cited.

1. Introduction

Forward evolution of large populations in genetics has a long history, starting in the 1920s; it
is closely attached to the names of R. A. Fisher and S. Wright; see Nagylaki [1] for historical
commentaries and on the role played by the French geneticist G. Malécot, starting shortly
before the second world war. The book of Ewens [2] is an excellent modern presentation of
the current mathematical theory. Coalescent theory is the corresponding backward problem
obtained while running the forward evolution processes backward in time. It was discovered
independently by several researchers in the 1980s, but definitive formalization is commonly
attributed to Kingman [3]. Major contributions to the development of coalescent theory were
made (among others) by P. Donnelly, R. Griffiths, R. Hudson, F. Tajima, and S. Tavaré (see
the course of Tavaré in Saint-Flour [4] for a review). It included incorporating variations
and Sagitov [6], independently, introduced coalescent processes with multiple collisions
of ancestral lineages. Shortly later, the full class of exchangeable coalescent processes with
simultaneous multiple mergers of ancestral lineages was discovered by Möhle and Sagitov [7], and Schweinsberg [8]. All these recent developments and improvements concern chiefly the discrete neutral case and its various scaling limits in continuous time and/or space. As was shown by Möhle [9, 10], neutral forward and backward theories learn much from one another by using a concept of duality introduced by Liggett [11]. Backward theory in the presence of mutations in the forward process is well understood, as it requires the study of a marked Kingman’s tree (see Tavaré [4], for a review). In the works of Neuhauser and Krone [12], there is also some use of the duality concept in an attempt to understand the genealogies of a Wright-Fisher diffusion (as a limit of a discrete Wright-Fisher model) presenting a selection mechanism; this led these authors to the idea of the ancestral selection graph extending Kingman’s coalescent tree of the neutral theory; see also Huillet [13] for related objectives in the context of Wright-Fisher diffusions with and without drifts. There is, therefore, some evidence that the concept of duality could help one understand the backward theory even in nonneutral situations when various evolutionary forces are the causes of deviation to neutrality (see Crow and Kimura [14], Maruyama [15], Gillespie [16], and Ewens [2], for a discussion on various models of utmost interest in population genetics).

In this note, we focus on discrete nonneutral Wright-Fisher (WF) models and on the conditions on the bias probabilities under which forward branching population dynamics is directly amenable to a dual discrete ancestral coalescent. We emphasize that duality formulae still are of great use when considering discrete nonneutral Wright-Fisher models, at least for specific deviation forces to neutrality. It is shown that it concerns a large class of nonneutral models involving completely monotone bias probabilities. Several classical examples are supplied in the light of complete monotonicity. In the process leading us to focus on these peculiar bias models, some unsuspected evolutionary mechanisms of potential interest are introduced and discussed, as suggested by elementary algebra on completely monotone functions. We emphasize that the relevance of these novel bias mechanisms in Biology seems to deserve additional work and confrontation with real-world problems is urged for to pinpoint their biological significance.

We will finally briefly outline the content of this manuscript. Section 2 is designed to fix the background and ideas. We introduce some basic facts about the discrete-time forward (Section 2.2) and backward processes (Section 2.3) arising from exchangeable reproduction laws (Section 2.1). In Section 2.4, we introduce a concept of duality and briefly recall its relevance to the study of the neutral case problem. The basic question we address in subsequent sections is whether this notion of duality still makes sense in nonneutral situations. We start supplying important nonneutral examples in Section 3. In Section 4, we show that duality does indeed make sense in the framework of discrete nonneutral Wright-Fisher models, but only for the class of completely-monotone state-dependent transition frequencies. In Section 5, we show that most nonneutrality mechanisms used in the literature fall within this class, or are amenable to it via some “reciprocal transformation,” starting with elementary mechanisms and ending up with more complex ones. In Section 6, we show that duality can be used in nonneutral situations to compute the extinction probabilities (invariant measure) of the dual backward ancestral process if one knows the invariant measure (resp., extinction probabilities) of the forward branching process.

### 2. Discrete-Time Neutral Coalescent

In this section, to fix the background and notations, we review some well-known facts from the cited literature.
2.1. Exchangeable Neutral Population Models: Reproduction Laws (the Cannings Model [17])

Consider a population with nonoverlapping generations \( r \in \mathbb{Z} \). Assume the population size is constant, say \( n \) (\( n \) individuals (or genes)) over generations. Assume the random reproduction law at generation 0 is \( \nu_0 \) satisfying

\[
\sum_{m=1}^{n} \nu_{m,n} = n. \tag{2.1}
\]

Here, \( \nu_{m,n} \) is the number of offspring of gene \( m \). We avoid the trivial case: \( \nu_{m,n} = 1, m = 1, \ldots, n \). One iterates the reproduction over generations, while imposing the following additional assumptions:

(i) exchangeability: \( (\nu_{1,n}, \ldots, \nu_{n,n}) \overset{d}{=} (\nu_{\sigma(1),n}, \ldots, \nu_{\sigma(n),n}) \), for all permutations \( \sigma \in S_n \);

(ii) time-homogeneity: reproduction laws are independent and identically distributed (iid) at each generation \( r \in \mathbb{Z} \).

This model, therefore, consists of a conservative-conditioned branching Galton-Watson process in \( [n]^2 \), where \( [n] := \{0, 1, \ldots, n\} \) (see Karlin-McGregor [18]).

Famous reproduction laws are as follows.

Example 2.1. The multinomial Dirichlet model: \( \nu_n \overset{d}{=} \text{Multin.Dirichlet}(n; \theta) \), where \( \theta > 0 \) is a disorder parameter. With \( k_n := (k_1, \ldots, k_n) \), \( \nu_n \) admits the following joint exchangeable distribution on the simplex \( |k_n| := \sum_{m=1}^{n} k_m = n \):

\[
P_{\theta}(\nu_n = k_n) = \frac{n!}{[\theta]_n} \prod_{m=1}^{n} \frac{[\theta]_{k_m}}{k_m!},
\]

where \( [\theta]_k = \theta(\theta+1) \cdots (\theta+k-1) \) is the rising factorial of \( \theta \). This distribution can be obtained by conditioning \( n \) independent mean 1 Pólya distributed random variables \( \xi_n = (\xi_1, \ldots, \xi_n) \) on summing to \( n \), that is to say, \( \nu_n \overset{d}{=} (\xi_n \mid |\xi_n| = n) \), where

\[
P_{\theta}(\xi_1 = k) = \frac{[\theta]_k}{k!} (1 + \theta)^{-k} \left( \frac{\theta}{1 + \theta} \right)^{\theta}, \quad k \in \mathbb{N}.
\]

When \( \theta \uparrow \infty \), this distribution reduces to the Wright-Fisher model for which \( \nu_n \overset{d}{=} \text{Multin}(n; 1/n, \ldots, 1/n) \). Indeed, \( \nu_n \) admits the following joint exchangeable multinomial distribution on the simplex \( |k_n| = n \):

\[
P_{\infty}(\nu_n = k_n) = \frac{n! n^{-n}}{\prod_{m=1}^{n} k_m!}.
\]


This distribution can be obtained by conditioning \( n \) independent mean 1 Poisson distributed random variables \( \xi_n = (\xi_1, \ldots, \xi_n) \) on summing to \( n \) : \( \nu_n \overset{d}{=} (\xi_n \mid |\xi_n| = n) \). When \( n \) is large, using Stirling formula, \( n! \sim \sqrt{2\pi n} n^{n+1/2} e^{-n} \), it follows that \( \nu_n \overset{d}{\to} \xi_\infty \) with joint finite-dimensional law: \( \mathbb{P}(\xi_n = k_n) = \prod_{m=1}^{n} (e^{-1}/k_m!) = (e^{-n}/\prod_{m=1}^{n} k_m!) \) on \( \mathbb{N}^n \). Thanks to the product form of all finite-dimensional laws of \( \xi_\infty \), we get an asymptotic independence property of \( \nu_n \).

**Example 2.2.** In the Moran model, \( \nu_n \overset{d}{\to} \) random permutation of \((2, 0, 1, \ldots, 1)\). In such a model, only one new gene per generation may come to life, at the expense of the simultaneous disappearance of some other genes.

### 2.2. Forward in Time Branching Process

Take a subsample of size \( m \) from \([n] := \{0, 1, \ldots, n\}\) at generation 0. Let

\[
N_r(m) = \# \text{offspring at generation } r \in \mathbb{N}_+, \text{ forward in time.}
\]

(2.5)

This sibship process is a discrete-time homogeneous Markov chain, with transition probability

\[
\mathbb{P}(N_{r+1}(m) = k' \mid N_r(m) = k) = \mathbb{P}(\nu_{1,n} + \cdots + \nu_{k,n} = k').
\]

(2.6)

It is a martingale, with state space \([0, \ldots, n]\), initial state \( m \), absorbing states \([0, n]\), and transient states \([1, \ldots, n - 1]\). The first hitting time of boundaries \([0, n]\), which is \( \tau(m) = \tau_0(m) \land \tau_n(m) \), is finite with probability 1 and has finite mean. Omitting reference to any specific initial condition \( m \), the process \((N_r; r \in \mathbb{N})\) has the transition matrix \( \Pi_n \) with entries \( \Pi_n(k, k') = \mathbb{P}(\nu_{1,n} + \cdots + \nu_{k,n} = k') \) given by (2.6). We have \( \Pi_n(0, k') = \delta_{0,k'} \) and \( \Pi_n(n, k') = \delta_{n,k'} \), and \( \Pi_n \) is not irreducible. However, \( \Pi_n \) is aperiodic and (apart from absorbing states) cannot be broken down into noncommunicating subsets; as a result it is diagonalizable, with eigenvalues \(|\lambda_0| \geq |\lambda_1| \geq \cdots \geq |\lambda_n|\) and \( 1 = \lambda_0 = \lambda_1 > |\lambda_2| \).

**Example 2.3** (Dirichlet binomial). With \( U \) a \((0, 1)\)-valued random variable with density \( \beta(k\theta, (n - k)\theta) \)

\[
\mathbb{P}(\nu_{1,n} + \cdots + \nu_{k,n} = k') = \binom{n}{k'} \frac{[k\theta]_k [(n-k)\theta]_{n-k}}{[n\theta]_n} = \mathbb{E} \left[ \binom{n}{k'} U^{k'} (1-U)^{n-k'} \right],
\]

(2.7)

which is a beta mixture of the binomial distribution \( \text{Bin}(n, u) \).

**Example 2.4.** The Wright-Fisher model has a \( \text{Bin}(n, k/n) \) transition matrix

\[
\mathbb{P}(N_{r+1}(m) = k' \mid N_r(m) = k) = \binom{n}{k'} \left( \frac{k}{n} \right)^k \left( 1 - \frac{k}{n} \right)^{n-k'}.
\]

(2.8)
Remark 2.5 (statistical symmetry). Due to exchangeability of the reproduction law, neutral models are symmetric in the following sense. The transition probabilities of \( \bar{N}_r(m) := n - N_r(m) \) are equal to the transition probabilities of \( N_r(m) \).

### 2.3. Backward in Time Process (Neutral Coalescent)

The coalescent backward process can be defined as follows. Take a subsample of size \( m \) from \([n]\) at generation 0. Identify two individuals from \([m]\) at each step if they share a common ancestor one generation backward in time. This defines an equivalence relation between 2 genes from the set \([m]\). Define the induced ancestral backward process

\[ \mathcal{A}_r(m) \in \mathcal{E}_m = \{\text{equivalence classes (partitions) of } [m]\}, \quad r \in \mathbb{N}, \text{ backward in time}. \quad (2.9) \]

The ancestral process is a discrete-time Markov chain with transition probability:

\[ P(\mathcal{A}_{r+1}(m) = \alpha | \mathcal{A}_r(m) = \beta) = P_{\beta,\alpha}^{\alpha}; \quad \text{with } (\alpha, \beta) \in \mathcal{E}_m, \alpha \subseteq \beta, \quad (2.10) \]

where, with \( a = |\alpha| = \text{number of equivalence classes of } \alpha, b = |\beta| = \text{number of equivalence classes of } \beta, \ b_a := (b_1, \ldots, b_a) = \text{clusters sizes of } \beta, \text{ and } (m)_a := m(m-1)\cdots(m-a+1) \) a falling factorial

\[ P_{\beta,\alpha}^{\alpha} = P_{\nu,n}^{(n)}(b_a) = \frac{(n)_a}{(n)_b} \text{E} \left( \prod_{l=1}^{a} \left( \nu_{l,n} \right)_{b_l} \right) \quad (2.11) \]

is the probability of a \( b_a \)-merger. This is the probability that \( b \) randomly chosen individuals out of \( n \) have \( a \leq b \) distinct parents, \( c \) merging classes and cluster sizes \( b_1 \geq \cdots \geq b_c \geq 2, \ b_{c+1} = \cdots = b_a = 1 \).

- If \( c = 1 \) a unique multiple collision occurs of order \( b_1 \geq 2 \).
- If \( b_1 = 2 \) a simple binary collision occurs involving only two clusters.
- If \( c > 1 \), simultaneous multiple collisions of orders \( b_1 \geq \cdots \geq b_c \geq 2 \) occur.

Thus, the jump’s height of a transition \( b \rightarrow a \) is \( b - a = \sum_{i=1}^{c} (b_i - 1) \), corresponding to a partition of integer \( b - a \) into \( c \) summands, each \( \geq 1 \).

The chain’s state-space is \{equivalence relations on (partitions of) \([1, \ldots, m]\)\}; it has dimension \( B_m := \sum_{l=0}^{m} S_{m,l} \) (a Bell number), where \( S_{m,l} \) are the second-kind Stirling numbers.

The chain has initial state \( \mathcal{A}_0 = \{(1, \ldots, m)\} \), and terminal absorbing state \( \{\emptyset\} \).

### Examples

From the Dirichlet Example 2.3, we get, \( P_{\nu,n}^{(n)}(b_a) = (n)_a / [n\theta]_b \prod_{m=1}^{a} [\theta]_{b_m} \).

From the WF Example 2.4: In this case, \( P_{\nu,n}^{(n)}(b_a) = (n)_a / n^b \) is the uniform distribution on \{\( b_a : b_1 + \cdots + b_a = b \)\}. 

The ancestral count process

Let

\[ A_r(m) = \text{# ancestors at generation } r \in \mathbb{N}, \text{ backward-in-time.} \]  

(2.12)

The backward ancestral count process is a discrete-time Markov chain with transition probabilities [17, 19]

\[
\mathbb{P}(A_{r+1}(m) = a \mid A_r(m) = b) = P^{(n)}_{b,a} \\
\quad := \frac{b!}{a!} \sum_{b_1, \ldots, b_n \in \mathbb{N}} \frac{P^{(n)}_{b,a} (b_1 \cdots b_n)}{b_1! \cdots b_n!},
\]

(2.13)

This Markov chain has state-space \( \{0, \ldots, m\} \), initial state \( m \), absorbing states \( \{0,1\} \). The process \( (A_r; r \in \mathbb{N}) \) has the transition matrix \( P_n \) with entries \( P_n(b,a) = P^{(n)}_{b,a} \) given by (2.13).

Note, by inclusion-exclusion principle, the alternative alternating expression of \( P^{(n)}_{b,a} \):

\[
P^{(n)}_{b,a} := \binom{n}{b} \sum_{m=0}^{a} (-1)^{a-m} \binom{a}{m} \mathbb{E} \left( \left( \nu_{1,n} + \cdots + \nu_{m,n} \right) \right).
\]

(2.14)

2.4. Duality (Neutral Case)

We start with a definition of the duality concept which is relevant to our purposes.

**Definition 2.6** (see [11]). Two Markov processes \( (X^1_t, X^2_t; \ t \geq 0) \), with state-spaces \( (\mathcal{E}_1, \mathcal{E}_2) \), are said to be dual with respect to some real-valued function \( \Phi \) on the product space \( \mathcal{E}_1 \times \mathcal{E}_2 \) if \( \forall x_1 \in \mathcal{E}_1, \ \forall x_2 \in \mathcal{E}_2, \ \forall t \geq 0 \):

\[
\mathbb{E}_{x_1} \Phi(X^1_t, x_2) = \mathbb{E}_{x_2} \Phi(x_1, X^2_t).
\]

(2.15)

We then recall basic examples of dual processes from the neutral and exchangeable population models (Möhle [10]). The neutral forward and backward processes \( (N_r, A_r; r \in \mathbb{N}) \) introduced in the two preceding subsections are dual with respect to the hypergeometric sampling without replacement kernels

\[
\begin{align*}
(i) & \quad \Phi^1_n(m,k) = \binom{m}{k}, \\
(ii) & \quad \Phi^2_n(m,k) = \binom{n-m}{k} \quad \text{on } \{0, \ldots, n\}^2.
\end{align*}
\]

(2.16)
Namely, (i) reads

\[
\mathbb{E}_{m} \left[ \begin{array}{c} \binom{N_r}{k} \\ \binom{n}{k} \end{array} \right] = \mathbb{E}_k \left[ \begin{array}{c} \binom{m}{A_r} \\ \binom{n}{A_r} \end{array} \right] \\
= \mathbb{E}_k \left[ \begin{array}{c} \binom{n-A_r}{m-m} \\ \binom{n}{m-m} \end{array} \right].
\] (2.17)

Call type A individuals the descendants of the \( m \) first chosen individuals (allele \( A \)) in the study of the forward process; type \( a \) individuals are the remaining ones (allele \( a \)). The left-hand-side of the above equality is an expression of the probability that \( k \)-samples (without replacement) from population of size \( N_r \) at time \( r \) are all of type \( A \), given \( N_0 = m \). If these \( k \)-samples are all descendants of \( A_r \) ancestors at time \( -r \), this probability must be equal to the probability that a \( (n - m) \)-sample from population of size \( A_r \) at time \( -r \) are all of type \( a \). This is the meaning of the right-hand-side.

Moreover, (ii) reads

\[
\mathbb{E}_{m} \left[ \begin{array}{c} \binom{n-N_r}{k} \\ \binom{n}{k} \end{array} \right] = \mathbb{E}_k \left[ \begin{array}{c} \binom{n-m}{A_r} \\ \binom{n}{A_r} \end{array} \right] \\
= \mathbb{E}_k \left[ \begin{array}{c} \binom{n-A_r}{m} \\ \binom{n}{m} \end{array} \right].
\] (2.18)

The left-hand-side is the probability that a \( k \)-sample (without replacement) from population of size \( N_r \) at time \( r \) are all of type \( a \), given \( N_0 = m \). If these \( k \)-samples are all descendants of \( A_r \) ancestors at time \( -r \), this probability must be equal to the probability that \( m \)-samples from population of size \( A_r \) at time \( -r \) are themselves all of type \( a \).

With \( P_n^\prime \) the transpose of \( P_n \), a one-step \((r = 1)\) version of these formulae is

\[
\begin{align*}
(i) \quad & \Pi_n \Phi_n^1 = \Phi_n^1 P_n^\prime, \\
(ii) \quad & \Pi_n \Phi_n^2 = \Phi_n^2 P_n^\prime,
\end{align*}
\] (2.19)

where \( (\Phi_n^1, \Phi_n^2) \) are \( n \times n \) matrices with entries \( \Phi_n^1(m, k) \) and \( \Phi_n^2(m, k) \), respectively, and \((\Pi_n, P_n)\) the transition matrices of forward and backward processes. Note that the matrix \( \Phi_n^2 \) is symmetric and left-upper triangular. The matrices \( \Phi_n^1 \) and \( \Phi_n^2 \) are both invertible, with respective entries

\[
\begin{align*}
[\Phi_n^1]^{-1}(i, j) &= (-1)^{i-j} \binom{n}{i} \binom{n}{j}, \\
[\Phi_n^2]^{-1}(i, j) &= (-1)^{i+j-n} \binom{n}{i} \binom{n}{j},
\end{align*}
\] (2.20)
The matrix $[\Phi_n^1]^{-1}$ is left-lower triangular, while $[\Phi_n^2]^{-1}$ is symmetric right-lower triangular. Thus,

\begin{align}
(i) \quad [\Phi_n^1]^{-1} \Pi_n \Phi_n^1 &= P_n', \\
(ii) \quad [\Phi_n^2]^{-1} \Pi_n \Phi_n^2 &= P_n'.
\end{align}

In any case, being similar matrices, $\Pi_n$ and $P_n'$ (or $P_n$) both share the same eigenvalues. If $R_n$ diagonalizing $\Pi_n$ is known so that $R_n^{-1} \Pi_n R_n = \Lambda_n := \text{diag}(\lambda_0, \ldots, \lambda_n)$, the diagonal matrix of the eigenvalues of $\Pi_n$, then, with $\Phi_n = \Phi_n^1$ or $\Phi_n^2$, $\tilde{R}_n := \Phi_n^{-1} R_n$ diagonalizes $P_n'$ and is obtained for free (and conversely). $R_n$ is the matrix whose columns are the right-eigenvectors of $\Pi_n$ and $\tilde{R}_n$ is the matrix whose columns (rows) are the right eigenvectors (left-eigenvectors) of $P_n'$ (of $P_n$). Similarly, if $L_n$ is the matrix whose rows are the left eigenvectors of $\Pi_n$, $\tilde{L}_n := L_n \Phi_n$ is the matrix whose rows (columns) are the left eigenvectors (right eigenvectors) of $P_n'$ (of $P_n$). With $l_k'$ the $k$th row of $L_n$ and $r_k$ the $k$th column of $R_n$, the spectral decomposition of $\Pi_n$ is

$$\Pi_n' = \sum_{k=0}^{n} \frac{\lambda_k}{r_k'} \frac{l_k l_k'}{r_k'^2}, \quad r \in \mathbb{N},$$

whereas, with $\tilde{l}_k$ the $k$th column of $\tilde{L}_n$ and $\tilde{r}_k'$ the $k$th row of $\tilde{R}_n$, the one of $P_n$ reads

$$P_n' = \sum_{k=0}^{n} \frac{\lambda_k}{\tilde{l}_k r_k'} \frac{\tilde{l}_k \tilde{r}_k'}{\tilde{r}_k'^2} = \sum_{k=0}^{n} \frac{\lambda_k}{\tilde{r}_k'} (\Phi_n^{-1} r_k) l_k', \quad r \in \mathbb{N}.$$ 

In Möhle [10], a direct combinatorial proof of the duality result can be found (in the general exchangeable or neutral case); it was obtained by directly checking the consistency of (2.6), (2.13), and (2.16).

The duality formulae allow one to deduce the probabilistic structure of one process from the one of the other. The question we address now is does duality still make sense in nonneutral situations? We will see that it does in discrete nonneutral Wright-Fisher models, but only for some class of state-dependent transition frequencies.

### 3. Beyond Neutrality (Symmetry Breaking)

Discrete forward nonneutral models (with nonnull drifts) can be obtained by substituting

$$k \rightarrow np\left(\frac{k}{n}\right) \quad \text{in} \quad \mathbb{P}\left(v_{1,n} + \cdots + v_{k,n} = k'\right),$$

(3.1)
where

\[ p(x) : x \in (0, 1) \rightarrow (0, 1) \] is continuous, increasing, with \( p(0) = 0, \ p(1) = 1. \] (3.2)

\( p(x) \) is the state-dependent Bernoulli bias probability different from identity \( x \) (as in neutral case).

When particularized to the WF model, this leads to the biased transition probabilities

\[ \mathbb{P}(N_{r+1}(m) = k' \mid N_r(m) = k) = \left( \frac{\lfloor k \rfloor}{n} \right) p \left( \frac{k'}{n} \right) \left( 1 - p \left( \frac{k}{n} \right) \right)^{n-k'} \] (3.3)

In this binomial \( n \)-sampling with replacement model, a type \( A \) individual is drawn with probability \( p(k/n) \) which is different from the uniform distribution \( k/n \) due to bias effects.

From this, we conclude (a symmetry breaking property) the following: the transition probabilities of \( \overline{N}_r(m) := n - N_r(m), \ r \in \mathbb{N} \) are

\[ \text{Bin} \left( n, 1 - p \left( 1 - \frac{k}{n} \right) \right) \neq \text{Bin} \left( n, p \left( \frac{k}{n} \right) \right), \] (3.4)

and so, no longer coincide with the ones of \( (N_r(m); \ r \in \mathbb{N}) \). The process \( N_r(m), \ r \in \mathbb{N} \) no longer is a martingale. Rather, if \( x \rightarrow p(x) \) is concave (convex), \( N_r(m), \ r \in \mathbb{N} \) is a submartingale (supermartingale) because \( \mathbb{E}(N_{r+1}(m) \mid N_r(m)) = np(N_r(m)/n) \geq N_r(m) \) (resp., \( \leq N_r(m) \)).

In the binomial neutral Wright-Fisher transition probabilities, we replaced the success probability \( k/n \) by a more general function \( p(k/n) \). However, this replacement leaves open the question what model is in the background and what quantity the process \( (N_r, \ r \in \mathbb{N}) \) really counts? A concrete model in terms of offspring variables must be provided instead. To address this question, we emphasize that the reproduction law corresponding to the biased binomial model is multinomial and asymmetric, namely, \( \nu_n \overset{d}{\sim} \text{Multin}(n; \pi_n) \), where \( \pi_n := (\pi_1, \ldots, \pi_n) \) and \( \pi_{m,n} = p(m/n) - p((m-1)/n), \ m = 1, \ldots, n \). We note that under our hypothesis,

\[ \sum_{m=1}^{n} \pi_{m,n} = p(1) - p(0) = 1. \] (3.5)

Due to its asymmetry, the law of the biased \( \nu_n \) no longer is exchangeable.

We now recall some well-known bias examples arising in population genetics.

Example 3.1 (homographic model, selection). Assume

\[ p(x) = \frac{(1 + s)x}{1 + sx}, \] (3.6)

where \( s > -1 \) is a selection parameter. This model arises when gene \( A \) (resp., \( a \)), with frequency \( x \) (resp., \( 1 - x \)), has fitness \( 1 + s \) (resp., 1). The case \( s > 0 \) arises when gene of type \( A \) is selectively advantageous, whereas it is disadvantageous when \( s \in (-1, 0) \).
Example 3.2 (selection with dominance). Assume

$$p(x) = \frac{(1+s)x^2 + (1+sh)x(1-x)}{1+sx^2 + 2shx(1-x)}.$$  \hfill (3.7)

In this model, genotype $AA$ (resp., $Aa$ and $aa$), with frequency $x^2$ (resp., $2x(1-x)$ and $(1-x)^2$) has fitness $1+s$ (resp., $1+sh$ and 1). $h$ is a measure of the degree of dominance of heterozygote $Aa$. We impose $s > -1$ and $sh > -1$. Note that the latter quantity can be put into the canonical form of deviation to neutrality:

$$p(x) = x + sx(1-x) - \frac{h-x(2h-1)}{1+sx^2 + 2shx(1-x)},$$  \hfill (3.8)

where the ratio appearing in the right-hand-side is the ratio of the difference of marginal fitnesses of $A$ and $a$ to their mean fitness. The case $h = 1/2$ corresponds to balancing selection with $p(x) = x + (s/2)(x(1-x)/(1+sx))$.

Example 3.3 (quadratic model). With $c \in [-1,1]$, a curvature parameter, one may choose

$$p(x) = x(1+c-cx).$$  \hfill (3.9)

If $c = 1$, $p(x) = x(2-x) = 1 - (1-x)^2$, this bias appears in a discrete 2-sex population model [9, 10]. We will give below an interpretation of this quadratic model when $c \in (0,1]$ in terms of a joint one-way mutations and neutrality effects model.

We can relax the assumption $p(0) = 0$, $p(1) = 1$ by assuming $0 \leq p(0) \leq p(1) \leq 1$, $p(1) - p(0) \in [0,1)$.

Example 3.4 (affine model). Take, for example,

$$p(x) = (1-\mu_2)x + \mu_1(1-x),$$  \hfill (3.10)

where $(\mu_1, \mu_2)$ are mutation probabilities, satisfying $\mu_1 \leq 1-\mu_2$. It corresponds to the mutation scheme $\mu_1^A \mu_2^A$. To avoid discussions of intermediate cases, we will assume that $p(0) = \mu_1 > 0$ and $p(1) < 1$ ($\mu_2 > 0$). In this case, the matrix $\Pi_n$ is irreducible and even primitive and all states of this Markov chain are now recurrent. We have $P(N_{r+1} > 0 \mid N_r = 0) = 1 - (1 - p(0))^n > 0$ and $P(N_{r+1} < n \mid N_r = n) = 1 - (1 - p(0))^n > 0$ and the boundaries $\{0\}$ and $\{n\}$ no longer are strictly absorbing as there is a positive reflection probability inside the domain $\{0,1,\ldots,n\}$.

For reasons to appear now, we will be only interested in functions $q$ such that $q(x) := 1-p(x)$ is a completely monotone function (CM) on $(0,1)$, that is, satisfying

$$(-1)^j q^{(j)}(x) \geq 0, \quad \forall x \in (0,1),$$  \hfill (3.11)
for all order-$l$ derivatives $q^{(l)}$ of $q$, $l \geq 0$. If $p(x)$ is such that $q$ is CM, we will call it an admissible bias mechanism.

4. Nonneutral Wright-Fisher Models and Duality

Preliminaries 4.1. Let $v_n := (v(0), v(1), \ldots, v(n))$ be a $(n+1)$-vector of $[0,1]$-valued numbers. Define the backward difference operator $\nabla$ acting on $v_n$ by: $\nabla v(m) = v(m) - v(m-1)$, $m = 1, \ldots, n$. We have $\nabla^0 v(m) = v(m)$, $\nabla^2 v(m) = v(m) - 2v(m-1) + v(m-2)$, and so forth, and, starting from the endpoint $v(n)$

$$\nabla^j v(m) \big|_{m=n} = \sum_{i=0}^{j} (-1)^{j-i} \binom{j}{i} v(n-i), \quad j = 0, \ldots, n. \quad (4.1)$$

Let $u$ be some continuous function $[0,1] \rightarrow [0,1]$. Consider the $(n+1)$-vector $u_n := (u(0/n), \ldots, u(m/n), \ldots, u(n/n))$, sampling $u$ at points $m/n$. The function $u$ is said to be $\nabla$-completely monotonic if $(-1)^j \nabla^j u(m/n) \big|_{m=n} \geq 0$, for all $j = 0, \ldots, n$ and all $n \geq 0$. Let $(u^1, u^2)$ be two continuous functions on $[0,1]$. Let $u = u^1 \cdot u^2$. With $u_n$ the pointwise product of $u^1_n$ and $u^2_n$, assuming both functions $(u^1, u^2)$ to be $\nabla$-completely monotonic, so will be $u$, by the Leibniz rule. In particular, if $u$ is $\nabla$-completely monotonic, so will be its integral powers $u^i$, $i \in \mathbb{N}$. Our main result is as follows.

**Theorem 4.2.** Consider a nonneutral WF forward model $(N_r; r \in \mathbb{N})$ on $\{0, \ldots, n\}$, with continuous, nondecreasing bias $p(x)$, satisfying

$$0 \leq p(0) \leq p(1) \leq 1, \quad p(1) - p(0) \in [0,1]. \quad (4.2)$$

This process has forward transition matrix

$$\Pi_n(k,k') = \mathbb{P}(v_{1,n} + \cdots + v_{k,n} = k') = \binom{n}{k'} p \left( \frac{k}{n} \right)^k \left( 1 - p \left( \frac{k}{n} \right) \right)^{n-k} \quad (4.3)$$

There exists a Markov chain $(A_r; r \in \mathbb{N})$ on $\{0, \ldots, n\}$ such that $(N_r, A_r; r \in \mathbb{N})$ are dual with respect to $\Phi_n^2(m, k) = \binom{n}{k} \binom{m}{k} / \binom{m}{k}$ if and only if $x \rightarrow q(x) = 1 - p(x)$ is completely monotone on $(0,1)$. In this case, the transition probability matrix of $(A_r; r \in \mathbb{N})$ is

$$P_n(i,j) = \binom{n}{j} \sum_{l=0}^{j} (-1)^{j-l} \binom{j}{l} q \left( 1 - \frac{l}{n} \right) \geq 0, \quad (4.4)$$

where $P_n$ is a stochastic matrix if and only if $p(0) = 0$; else, if $p(0) > 0$, the matrix $P_n$ is substochastic.
Proof. Developing \([\Phi_n^2]^{-1}\Pi_n\Phi_n^2 = P_n^r\), we easily obtain

\[
P_n^r(j, i) = P_n(i, j) = \binom{n}{j} \sum_{l=0}^{j} (-1)^{j-l} \binom{j}{l} \left[ 1 - p\left(\frac{n-l}{n}\right) \right]^i
= \binom{n}{j} (-1)^j \nabla\left(q\left(\frac{m}{n}\right)^i\right)\bigg|_{m=n}.
\]

This entry is nonnegative if and only if \((-1)^j \nabla\left(q\left(\frac{m}{n}\right)^i\right)\bigg|_{m=n} \geq 0\), for all \(i, j = 0, \ldots, n\). But due to the above argument on \(\nabla\)-complete monotonicity of integral powers, this will be the case if and only if \((-1)^j \nabla\left(q\left(\frac{m}{n}\right)^i\right)\bigg|_{m=n} \geq 0\), for all \(j = 0, \ldots, n\). As this must be true for arbitrary value of population size \(n\), function \(q\) has to be \(\nabla\)-completely monotonic. Adapting now the arguments of Theorem 2 developed in [20, page 223], for absolutely monotone functions on \((0, 1)\), this will be the case if and only if \(x \rightarrow q(x) := 1 - p(x)\) is a completely monotone function on \((0, 1)\) in the sense that

\[
(-1)^l q^{(l)}(x) \geq 0, \; \forall x \in (0, 1), \; l \in \mathbb{N}.
\]

Next, since \((I - \nabla)u(m) = u(m - 1)\) is a simple back-shift,

\[
\sum_{j=0}^{n} P_n^r(j, i) = \sum_{j=0}^{n} P_n(i, j)
= (I - \nabla)^n \left(q\left(\frac{m}{n}\right)^i\right)\bigg|_{m=n}
= q(0)^i,
\]

and if \(q\) is CM, \(P_n\) is a stochastic matrix if and only if \(q(0) = 1\); else, if \(q(0) < 1\), the matrix \(P_n\) is substochastic.

We note that the first column of the matrix \(P_n\) is \(P_n(i, 0) = q(1)^i\) whereas its first line is \(P_n(0, j) = \delta_{0,j}\), expressing, as required, that the state 0 of \((A_r; r \in \mathbb{N})\) is absorbing.

\[
\square
\]

5. Examples

We show here that most of the simplest nonneutrality mechanisms used in the literature fall within the class which we would like to draw the attention on, or are amenable to it via some “reciprocal transformation” which we define. Elementary algebraic manipulations on CM functions allow to exhibit a vast class of unsuspected mechanisms. Note that in some cases, their biological relevance remains to be elucidated. The results presented in this section seem to be new. They serve as an illustration of our theorem.
5.1. Elementary Mechanisms

Assume first that \( p(x) \equiv x \) corresponding to the simple neutral case. Then \( q(x) = 1 - x \) is completely monotone on \((0,1)\). With \( S_{ij} \) the second kind Stirling numbers, we get a lower left triangular stochastic transition matrix

\[
P_n(i,j) = \begin{cases} 
(n/j) \sum_{l=0}^{j} (-1)^{j-l} \left( \binom{j}{l} \left( \frac{1}{n} \right) \right)^{l} = (n)_j \cdot n^{-i} \cdot S_{i,j}, & j \leq i, \\
0, & \text{else.}
\end{cases}
\]  

(5.1)

The diagonal terms (eigenvalues) are all distinct with \( P_n(i,i) = (n)_i \cdot n^{-i} \). The matrix \( P_n \) is stochastic. Due to triangularity, ancestral process is a pure death Markov process which may be viewed as a discrete coalescence tree.

From Example 3.4 (mutation). Assume that (3.10) holds \( p(x) = (1 - \mu_2)x + \mu_1(1 - x) \) where \((\mu_1, \mu_2)\) are mutation probabilities. Then, with \( \kappa := 1 - (\mu_1 + \mu_2) \), \( q(x) = 1 - \mu_1 - xx \) is completely monotone on \((0,1)\) if and only \( \mu_1 \leq 1 - \mu_2 \ (\kappa \geq 0) \). In this case, \( P_n \) is again lower left triangular (a pure death process). We have

\[
P_n(i,j) = \begin{cases} 
\left( \binom{n}{j} \right) \sum_{l=0}^{j} (-1)^{j-l} \left( \binom{j}{l} \left( \mu_2 + \frac{\kappa}{n} \right) \right)^{l} =: (n)_j \cdot n^{-i} \cdot S^\mu_{i,j} \left( \frac{\kappa}{n} \right), & j \leq i, \\
0, & \text{else,}
\end{cases}
\]  

(5.2)

in terms of generalized Stirling numbers \( S^\mu_{i,j}(\kappa/n) \). We have \( P_n(i,i) = (n)_i \cdot n^{-i} \) and the spectrum of \( P_n \) is real. When \( \mu_1 > 0 \), this matrix is substochastic with \( \sum_{j=0}^{n} P_n(i,j) = (1 - \mu_1)^i \).

A particular case deals with one-way mutations \((\mu_1 + \mu_2 > 0, \mu_1 \cdot \mu_2 = 0)\).

If \( \mu_2 = 0 \), \( P_n(i,j) = (1 - \mu_1)^i \cdot (n)_j \cdot n^{-i} \cdot S_{i,j}, j \leq i, = 0 \), else. Further, \( \sum_{j=0}^{n} P_n(i,j) = (1 - \mu_1)^i < 1 \).

If \( \mu_1 = 0 \), \( P_n(i,j) = (n)_j \cdot n^{-i} \cdot S^\mu_{i,j}(1 - \mu_2/n), j \leq i, = 0 \), else. The corresponding matrix \( P_n \) is stochastic.

From Example 3.3 (quadratic). Assume that \( p(x) = x(1 + c - cx) \), as in (3.9). Then \( q(x) = (1 - x)(1 - cx) \) which is completely monotone if and only if \( c \in [0,1] \). The case \( c = 0 \) is the neutral case, whereas \( c = 1 \) appears in a 2-sex model of Möhle. In this quadratic case, since \( \nabla^i(q(m/n)^i) = 0 \) if \( j > 2i \), then \( P_n(i,j) = 0 \) if \( j > 2i \) and so \( P_n \) is a Hessenberg-like matrix.

Note that \( \sum_{j=0}^{n} P_n(i,j) = q(0)^i = 1 \).

From the selection Example 3.1, when (3.6) holds

\[
p(x) = \frac{(1+s)x}{1+sx},
\]  

(5.3)

where \( q(x) = 1 - p(x) = (1-x)/(1+sx) \) is CM whenever selection parameter \( s > 0 \). The induced matrix \( P_n \) is stochastic. It is no longer lower left triangular so that the ancestral no longer is a pure death process, being rather a birth and death process. The induced
coalescence pattern no longer is a discrete tree, but rather a graph (a discrete version of the ancestral selection graph of Neuhauser-Krone [12]).

From Example 3.2 (selection with dominance). The corresponding mechanism (3.7) with parameters \((s, h)\) satisfying \(s > -1\) and \(sh > -1\) is CM if and only if \(s \geq s_h := (1-2h)/h^2 > 0\) and \(h \in (0,1/2)\). The case \(h \in (0,1/2)\) corresponds to directional selection, where genotype \(AA\) has highest fitness compared to \(aa\)'s and the heterozygote class \(Aa\) has intermediate fitness compared to both homozygote classes. In this situation, marginal fitness of \(A\) exceeds the one of \(a\) and selective sweep is expected. When \(h \in (0,1/2)\), allele \(A\) is dominant to \(a\), whereas when \(h \in (1/2,1)\), allele \(A\) is recessive to \(a\) (a stabilizing effect slowing down the sweep). Critical value \(h = 1/2\) is a case of pure genic balancing selection.

**Example 5.1.** Consider the mechanism \(p(x) = x^\gamma\) for some \(\gamma > 0\). The function \(q(x) = 1 - p(x)\) is CM if and only if \(\gamma \in (0,1)\). Although this model seems quite appealing, we could find no reference to it in the specialized mathematical genetics literature.

**Example 5.2** (reciprocal mechanism). If \(p(x)\) is not admissible in that \(q\) is not CM, it can be that \(\bar{p}(x) := 1 - p(1-x)\) is itself admissible. As observed before, if \(N_r(m)\) has transition probabilities given by \(\text{Bin}(n, p(k/n))\), \(\bar{p}(x)\) arises in the transition probabilities of \(\bar{N}_r(m) := n - N_r(m)\). Indeed, such transitions are \(\text{Bin}(n, \bar{p}(k/n))\) distributed.

If \(p(x)\) is the selection mechanism of Example 3.1, (3.6), with \(s \in (-1,0)\) (not admissible), \(\bar{p}(x) = (1+\beta)x/(1+\beta x)\) is itself an admissible selection mechanism because it has reciprocal selection parameter \(\bar{s} = -s/(1+s) > 0\). If \(p(x)\) is the mechanism of Example 3.2, namely, (3.7), with parameters \((s,h)\), then \(\bar{p}(x)\) is itself a selection with dominance mechanism with reciprocal parameters \(\bar{s} = -s/(1+s)\) and \(\bar{h} = 1-h\). Assuming that \((s \leq s_h < 0, h \in (1/2,1))\), \(p(x)\) is not admissible whereas \(\bar{p}(x)\) is because \(\bar{s} \geq \bar{s}_h > 0\) and \(\bar{h} \in (0,1/2)\). Similarly, when \(\gamma \in (0,1)\), the mechanism \(p(x) = 1 - (1-x)^\gamma\) is not admissible but, from Example 5.1, \(\bar{p}(x) := 1 - p(1-x) = x^\gamma\) is as follows.

### 5.2. Bias Mechanisms with Mutational Effects

Let \(p_M(x) = (1-\mu_2)x + \mu_1 (1-x)\) be the mutational bias mechanism (with \(k = 1-(\mu_1 + \mu_2) \geq 0\)). Let \(p(x)\) be a bias mechanism such that \(q(x)\) is CM with \(p(1) - p(0) = 1\). Then,

\[
\bar{p}_M(x) = p_M(p(x))
\]

is such that \(\bar{q}_M(x) := 1 - \bar{p}_M(x)\) is CM. It is, therefore, admissible and adds mutational effects to the primary mechanism \(p(x)\). For example,

\[
\bar{p}_M(x) = \frac{\mu_1 + x((1+s)(1-\mu_2) - \mu_1)}{1+sx}
\]

is a mechanism of selection combined with mutational effects. We have \(\bar{p}_M(0) = \mu_1, \bar{p}_M(1) = 1-\mu_2\). The mechanisms \(\bar{p}_M(x)\) obtained in this way all share the specificity \(\bar{p}_M(1) - \bar{p}_M(0) =: \kappa < 1\).

Note that, except for the mutational affine mechanism, it is not true in general that whenever \(p^1(x)\) and \(p^2(x)\) are two admissible bias mechanisms, then \(p^1(p^2(x))\) is admissible.
5.3. Joint Bias Effects and Compound Bias

Let $p^1(x)$ and $p^2(x)$ be two admissible bias in that $q^1(x) := 1 - p^1(x)$ and $q^2(x) := 1 - p^2(x)$ are both completely monotone. Then

$$q(x) = q^1(x)q^2(x) \text{ is CM.}$$

(5.6)

Thus, with $x_1 \circ x_2 := x_1 + x_2 - x_1x_2$, the probabilistic product in $[0,1]$ is

$$(p^1(x), p^2(x)) \rightarrow p(x) = p^1(x) \circ p^2(x).$$

(5.7)

Whenever a WF model is considered with bias $p(x) = p^1(x) \circ p^2(x)$ obtained from two distinct bias $p^1(x)$ and $p^2(x)$, we call it a WF model with joint bias effect.

Example 5.3 (joint selection and mutational effects). Let $p^1(x) = p_M(x)$ and $p^2(x) = (1 + s)x/(1 + sx)$. We get

$$q(x) = \frac{(1 - \mu_1 - \kappa x)(1 - x)}{1 + sx},$$

$$p(x) = \frac{\mu_1 + x(s + 1 - \mu_1 + \kappa) - \kappa x²}{1 + sx},$$

(5.8)

with $p(0) = \mu_1$, $p(1) = 1$. This mechanism differs from the traditional mechanism of selection combined with mutational effects.

Example 5.4 (joint mutation and neutral effects). Let $p^1(x) = (1 - \mu_2)x + \mu_1(1 - x)$ and $p^2(x) = x$. We get

$$q(x) = (1 - x)(1 - \mu_1 - \kappa x),$$

$$p(x) = \mu_1 + x(1 - \mu_1 + \kappa(1 - x)),$$

(5.9)

with $p(0) = \mu_1$, $p(1) = 1$. When $\mu_1 = 0$ (one-way mutations), we recover the quadratic mechanism with curvature parameter $c = 1 - \mu_2$. This finding justifies some interest into the quadratic mechanisms with $c \neq 1$.

With $j = 1, 2$, the reproduction law of each elementary effect is $\nu_n^j \sim \text{Multin}(n; \pi_n^j)$, where $\pi_{m,n}^j = p^j(m/n) - p^j((m - 1)/n)$, $m = 1, \ldots, n$. Then, $\nu_n \sim \text{Multin}(n; \pi_n)$, $\pi_{m,n} = p(m/n) - p(m - 1)/n)$, $m = 1, \ldots, n$, where $\pi_n := \pi_n^1 \odot \pi_n^2$ is easily obtained componentwise by

$$\pi_{m,n} = \pi_{m,n}^1 \sum_{l=1}^m \pi_{l,n}^2 + \pi_{m,n}^2 \sum_{l=1}^m \pi_{l,n}^1, \quad m = 1, \ldots, n.$$ 

(5.10)

We let $\nu_n := \nu_n^1 \odot \nu_n^2 \sim \text{Multin}(n; \pi_n^1 \odot \pi_n^2)$. It is the reproduction law of a WF model obtained jointly from the two bias $p^1(x)$ and $p^2(x)$. 
Let \( \phi(x) : (0, 1) \to (0, 1) \) be an absolutely monotone function satisfying \( \phi^{(l)}(x) \geq 0 \) for all \( l \)th derivatives \( \phi^{(l)} \) of \( \phi \), for all \( x \in (0, 1) \). Such functions are well known to be probability generating functions (pgfs) of \( \mathbb{N} \)-valued random variables, say \( N \), that is to say, \( \phi(x) = \mathbb{E}[x^N] \). Clearly, if \( q \) is CM on \( (0, 1) \), then so is \( q_\phi(x) := \phi(q(x)) \). Thus, \( p_\phi(x) := 1 - \phi(1 - p(x)) \) is an admissible bias mechanism in that \( q_\phi(x) := 1 - p_\phi(x) \) is CM. We call it a compound bias.

**Example 5.5.** The general mechanism with mutational effects is in this class. Indeed,

\[
\tilde{q}_M(x) = 1 - p_M(x) \\
= 1 - p_M(p(x)) \\
= 1 - p_M(1 - q(x)),
\]

and so \( \phi(x) = 1 - p_M(1 - x) = 1 - (1 - \mu_2)(1 - x) - \mu_1 x = \mu_2 + \kappa x \) which is absolutely monotone as soon as \( \kappa = 1 - (\mu_1 + \mu_2) \geq 0 \).

**Example 5.6.** With \( \theta > 0 \), taking \( \phi(x) = e^{-\theta(1-x)} \) or \( (e^{\theta x} - 1)/(e^\theta - 1) \), the pgf of a Poisson (or shifted-Poisson) random variable, \( p_\phi(x) = 1 - \phi(1 - p(x)) \) is admissible if \( p(x) \) is. Note that if \( q \) is of the form \( q_\phi \), where \( \phi \) is the pgf of the Poisson random variable, then \( q_\phi(x)^a \) is admissible, for all \( \alpha > 0 \), a property reminiscent of infinite divisibility for pgfs. Taking \( \phi(x) = (1 - \pi)/(1 - \pi x) \) or \( x(1 - \pi)/(1 - \pi x), \pi \in (0, 1) \), the pgf of a geometric (or shifted-geometric) random variable, \( p_\phi(x) = sp(x)/(1 + sp(x)) \) or \( (s+1)p(x)/(1+sp(x)) \) is admissible if \( p(x) \) is (with \( s = \pi/(1 - \pi) > 0 \)). In the external latter mechanism, one recognizes the one in (3.6) occurring in the model with selection of Example 3.1.

**Example 5.7.** Let \( p(x) = x^\gamma \) with \( \gamma \in (0, 1) \) as in Example 5.1. Then, \( p_\phi(x) = 1 - q_\phi(x) \), where \( q_\phi(x) = e^{-\theta(1-q(x))} = e^{-\theta x^\gamma}, \theta > 0 \), is admissible. Note that \( p_\phi(x) \sim \theta x^\gamma \). The reciprocal function \( \tilde{p}_\phi(x) = q_\phi(1 - x) = e^{-\theta(1-x)^\gamma} \) also interprets as an absolutely monotone discrete-stable pgf (see Steutel and van Harn [21]). It is not admissible.

Proceeding in this way, one can produce a wealth of admissible bias probabilities \( p_\phi \), the signification of which in population genetics remaining though to be pinpointed, in each specific case study.

### 6. Limit Laws

Consider a WF model \( (N_r; r \in \mathbb{N}) \) on \( \{0, \ldots, n\} \) with forward transition matrix

\[
\Pi_k(k, k') = \binom{n}{k'} p^k \mathbin{k'} \left( 1 - p \frac{k}{n} \right)^{n-k'}, \tag{6.1}
\]

with admissible bias \( p(x) \). Define \( (A_r; r \in \mathbb{N}) \) as the dual Markov chain on \( \{0, \ldots, n\} \) with transition probability

\[
P_n(i, j) = \binom{n}{j} \sum_{l=0}^{j} (-1)^{j-l} \binom{j}{l} q \left( 1 - \frac{l}{n} \right)^i. \tag{6.2}
\]
Then, \((N_r, A_r; r \in \mathbb{N})\) are dual with respect to \(\Phi_n(m, k) := \Phi_n^r(m, k) = \binom{n-m}{k}/\binom{n}{k}\), to wit

\[
\mathbb{E}_m \left[ \binom{n-N_r}{k} \right] = \mathbb{E}_k \left[ \binom{n-m}{A_r} / \binom{m}{A_r} \right] = \mathbb{E}_k \left[ \binom{n-A_r}{n} / \binom{n}{m} \right].
\]  

(6.3)

We will distinguish two cases.

Case 1. Assume first that

\[ N_r \rightarrow^d N_\infty \text{ as } r \uparrow \infty, \text{ independently of } N_0 = m \geq 1. \]  

(6.4)

Let \(\pi_\infty(i) = \mathbb{P}(N_\infty = i)\) and \(\pi_\infty = (\pi_\infty(0), \ldots, \pi_\infty(n))'\). The line vector \(\pi_\infty'\) is the left eigenvector of \(\Pi_n\) associated to the eigenvalue 1 : \(\pi_\infty' = \pi_\infty' \Pi_n\). It is the (unique) invariant probability measure (stationary distribution) of \((N_r; r \in \mathbb{N})\).

If this stationary distribution exists, then, using duality formula, necessarily, \(A_r \rightarrow 0\) as \(r \uparrow \infty\) with probability \(\mathbb{P}_k(A_\infty = 0) =: \rho_\infty(k) < 1\). The numbers \(\rho_\infty(k)\) are the extinction probabilities of the dual process started at \(k\). As is well known, \(\rho_\infty = (\rho_\infty(0), \ldots, \rho_\infty(n))'\) is the unique solution to \((I - P_n)\rho_\infty = 0\) with \(\rho_\infty(0) = 1\).

Remark 6.1. Typical situations where \((N_r; r \in \mathbb{N})\) has an invariant measure is when mutational effects are present, and more generally when the bias mechanism satisfies \(p(0) > 0\) and \(p(1) < 1\). In this situation, the forward stochastic transition matrix \(\Pi_n\) has an algebraically simple dominant eigenvalue 1. By Perron-Frobenius theorem

\[
\lim_{r \uparrow \infty} \Pi_n^r = 1 \pi_\infty'.
\]  

(6.5)

where \(1' = (1, \ldots, 1)\). The invariant probability measure can be approximated by subsequent iterates of \(\Pi_n\), the convergence being exponentially fast, with rate governed by the second largest eigenvalue. Of course, detailed balance (stating that \(\pi_\infty \Pi_n(k, k') = \pi_\infty \Pi_n(k', k)\)) does not hold here and the forward chain in equilibrium is not time-reversible.

In these recurrent cases, the dual ancestral process \(A_r\) started at \(k\) gets extinct with probability \(\rho_\infty(k)\). The numbers \(1 - \rho_\infty(k)\) are the probabilities that it gets killed before getting extinct; in other words, \(1 - \rho_\infty(k)\) are the probabilities that \(A_r\) first hits an extra coffin state, say \(\{0\}\), before hitting \(\{0\}\).

In terms of moments, by the duality formula, we conclude that

\[
\left( \begin{array}{c} n \\ k \end{array} \right)^{-1} \mathbb{E} \left[ \binom{n-N_\infty}{k} \right] = \rho_\infty(k) = \mathbb{P}_k(A_\infty = 0),
\]  

(6.6)
relating $k$-factorial moments of $n - N_\infty$ to the extinction probabilities of $A_r$ given $A_0 = k$. We also have

$$\sum_{k=0}^{n} \sigma^k \mathbb{E}
\left[\binom{n - N_\infty}{k}\right] = \mathbb{E}[(1 + \sigma)^{n - N_\infty}]
= \sum_{k=0}^{n} \binom{n}{k} \rho_\infty(k) \sigma^k,$$

and so the probability generating function of $N_\infty$ can be expressed as $(u \in [0, 1])$:}

$$\mathbb{E}[u^{N_\infty}] = \sum_{k=0}^{n} \binom{n}{k} \rho_\infty(k) u^{n-k}(1-u)^k,$$

in terms of the Bernstein-Bézier polynomial of $(\rho_\infty(n-k); k = 0, \ldots, n)$.

Let $\rho_\infty = (\rho_\infty(0), \ldots, \rho_\infty(n))^\prime$. The vector $\rho_\infty$ is the right eigenvector of $P_n$ associated to the eigenvalue 1: $\rho_\infty = P_n \rho_\infty$. In this case, the matrix $P_n$ is sub-stochastic and the extinction probability of $(A_r; r \in \mathbb{N})$ given $A_0 = k$ is less than one. Thanks to duality, we have

$$\Pi_n \Phi_n = \Phi_n P_n^\prime$$

where the matrix $\Phi_n$ is symmetric whereas the matrix $\Phi_n^{-1}$ is symmetric right lower triangular, with

$$\Phi_n(m, k) = \binom{n-m}{k} \binom{n-k}{m},$$

$$\Phi_n^{-1}(i, j) = (-1)^{i+j-n} \binom{i}{n-j} \binom{j}{n-i},$$

Thus,

$$\Pi_n \Pi_n \Phi_n = \Phi_\infty \Phi_n = \Phi_\infty \Phi_n P_n^\prime,$$

showing that $\rho_\infty$ and $\pi_\infty$ are related through

$$\rho_\infty = \Phi_n \pi_\infty \quad \text{or} \quad \pi_\infty = \Phi_n^{-1} \rho_\infty.$$

The knowledge of the invariant measure $\pi_\infty$ of the forward process allows one to compute the extinction probabilities $\rho_\infty$ of the dual backward ancestral process (and conversely).
Example

Consider the discrete WF model with mutations of Example 3.4. In this case, \( N_r \xrightarrow{d} N_\infty \) as \( r \uparrow \infty \), regardless of \( N_0 = m \) and \( (N_r; r \in \mathbb{N}) \) has an invariant measure which is difficult to compute. Looking at the backward process, the matrix \( P_n \) so that boundaries \( \{i, j\}; j \leq i \), \( i = 0, \ldots, n \). We, therefore, get the following alternating expression for the invariant measure:

\[
\pi_\infty(i) = \left( \begin{array}{c} n \\ i \end{array} \right) \sum_{j=0}^{i} (-1)^{i-j} \left( \begin{array}{c} i \\ j \end{array} \right) \rho_\infty(n-j).
\] (6.13)

Concerning moments, for instance, we have \( \rho_\infty(1) = \mu_2/(\mu_1 + \mu_2) \) so that \( \mathbb{E}[N_\infty] = n\mu_1/(\mu_1 + \mu_2) \); from (5.2) we also have

\[
\rho_\infty(2) = \frac{\mu_2[n\mu_2(1 + \kappa) + \kappa^2]}{[1 - \kappa](n - (n - 1)\kappa^2)]} = \frac{1}{n(n-1)} \left( n(n-1) - n \frac{2n-1}{\mu_1 + \mu_2} + \mathbb{E}[N_\infty^2] \right),
\] (6.14)

allowing to compute \( \mathbb{E}[N_\infty^2] \) and then the variance of \( N_\infty \). We get

\[
\sigma^2(N_\infty) = \frac{n^2\mu_1\mu_2}{(\mu_1 + \mu_2)^2(2n(\mu_1 + \mu_2) + 1)} + o(n) \sim \frac{\mu_1\mu_2}{n^{1/2}} \frac{n}{2(\mu_1 + \mu_2)^3} n,
\] (6.15)

suggesting (when \( \mu_1\mu_2 > 0 \)) a central limit theorem for \( N_\infty \) as \( n \) grows large:

\[
\frac{1}{\sqrt{n}} \left( N_\infty - n \frac{\mu_1}{\mu_1 + \mu_2} \right) \xrightarrow{d} A \left( 0, \frac{\mu_1\mu_2}{2(\mu_1 + \mu_2)^3} \right).
\] (6.16)

Case 2. Conversely, assume now that given \( N_0 = m \)

\[
N_r \xrightarrow{d} 0 \quad \text{as} \quad r \uparrow \infty, \quad \text{with probability} \quad \mathbb{P}_m(N_\infty = 0) =: \rho_\infty(m),
\] (6.17)

so that boundaries \( \{0, n\} \) are absorbing. Then, the ancestral process \( (A_r; r \in \mathbb{N}) \) possesses an invariant distribution, in that

\[
A_r \xrightarrow{d} A_\infty \quad \text{as} \quad r \uparrow \infty, \quad \text{independently of} \quad A_0 = k \in [n].
\] (6.18)

In terms of moments, the duality formula means that

\[
\left( \begin{array}{c} n \\ m \end{array} \right)^{-1} \mathbb{E} \left[ \left( \begin{array}{c} n - A_\infty \\ m \end{array} \right) \right] = \rho_\infty(m) = \mathbb{P}_m(N_\infty = 0),
\] (6.19)
relating $m$-factorial moments of $n - A_\infty$ to the extinction probabilities of $N_r$ given $N_0 = m$. Stated differently, the probability generating function of $A_\infty$ is $(u \in [0,1])$

$$
\mathbb{E}[u^{A_\infty}] = \sum_{m=0}^{\infty} \binom{n}{m} \rho_\infty(m) u^{n-m}(1-u)^m.
$$

Let $\pi_\infty(i) = \mathbb{P}(A_\infty = i)$, with $\pi'_\infty = \pi'_\infty P_n$. Then, using duality, $\rho_\infty$ is the right eigenvector of $\Pi_n$ associated to the eigenvalue 1: $\rho_\infty = \Pi_n \rho_\infty$. Thus, $\rho_\infty$ and $\pi_\infty$ are related through

$$
\rho_\infty = \Phi_n \pi_\infty \quad \text{or} \quad \pi_\infty = \Phi_n^{-1} \rho_\infty.
$$

The knowledge of the extinction probabilities $\rho_\infty$ of the forward process allows one to compute the invariant measure $\pi_\infty$ of the dual backward ancestral process (and conversely).

**Examples**

Typical situations where boundaries $\{0,n\}$ are absorbing to $(N_r; r \in \mathbb{N})$ occur when $p(0) = 0$ and $p(1) = 1$. The simplest case is the neutral case, but the nonneutral selection and selection with dominance mechanisms or the quadratic mechanism (Examples 3.1, 3.2 and 3.3) are also in this class. For instance,

(i) in the neutral case, $\rho_\infty(m) = 1 - m/n$. Thus, $\pi_\infty(i) = \binom{n}{i} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} (j/n) = \delta_{i,1}$ and $A_r \xrightarrow{d} 1$ as $r \uparrow \infty$, the degenerate state reached when the most recent common ancestor (MRCA) is attained,

(ii) nondegenerate solutions of $A_\infty$ are obtained when considering bias mechanisms with $p(0) = 0$ and $p(1) = 1$,

(iii) consider any biased WF model with $p(0) = 0$ and $p(1) = 1$ for which $p(x) \sim \lambda x$, $\lambda > 1$. Then, due to large sample asymptotic independence,

$$
\nu_n \xrightarrow{d} \xi_\infty
$$

where $\xi_\infty$ is an iid sequence with $\xi_1 \sim \text{Poisson}(\lambda)$ (as it can easily be checked by the Poisson limit to the binomial distribution). In this case, the limiting extinction probability of $(N_r; r \in \mathbb{N})$ given $N_0 = m$ is $\lim_{n \to \infty} \rho_\infty(m) = \rho^m$, $m = 1, 2, \ldots$, where $0 < \rho < 1$ is the smallest solution to the fixed point equation

$$
x = e^{-\lambda(1-x)},
$$

where $\rho$ is the singleton extinction probability of a supercritical Galton-Watson process with offspring distribution Poisson($\lambda$). More precisely, proceeding as in M"ohle [9, Theorem 4.5], we have

$$
n(\rho^m - \rho_\infty(m)) = \rho^m \left( \frac{1 - \rho}{1 + \lambda \rho} m^2 + \frac{\lambda (1 - \rho) \rho}{1 - (\lambda \rho)^2} m \right).
$$
showing that the convergence of \( \rho_n(m) \) to \( \rho^m \) is of order \( n^{-1} \). As a result, we get the asymptotic normality

\[
\frac{1}{\sqrt{n}} (A_\infty - n(1 - \rho)) \overset{d}{\to} \mathcal{N}(0, \rho(1 - \rho)/1 + \lambda \rho).
\]  

(6.25)

Intuitively, \((1/n)\mathbb{E}[n - A_\infty] = \rho(1) = P_1(N_\infty = 0) \to \rho\), showing that \(\mathbb{E}[A_\infty] \sim n(1 - \rho)\) and

\[
\frac{1}{n(n - 1)} \mathbb{E}[(n - A_\infty)(n - 1 - A_\infty)] = \rho(2) = P_2(N_\infty = 0) \to \rho^2.
\]  

(6.26)

showing that \(\sigma^2(A_\infty) \sim n\rho(1 - \rho)/(1 + \lambda \rho)\).

For quadratic Example 3.3, \(p(x) = x(1 + c - cx)\), with \(c \in [0,1]\), \(\lambda = 1 + c > 1\) as soon as \(c > 0\). When \(c \in (0,1]\), we thus always have asymptotic normality. For Example 3.1, with selection, \(p(x) = (1 + s)x/(1 + sx)\), with \(s > -1 : p(x) \sim (1 + s)x\) and so \(\lambda = 1 + s\).

We have asymptotic normality only when \(s > 0\), that is, when the fitness is advantageous (corresponding as required to complete monotonicity of corresponding \(q = 1 - p\)).

Note that this asymptotic behavior does not hold for the Lipshitz continuous admissible mechanism \(p(x) = x^4\) of Example 5.1 (or more generally for mechanisms satisfying \(p(x) \sim \theta x^\gamma, \theta > 0\) as in compound bias Example 5.7) with \(\gamma \in (0,1)\) because its behavior near 0 is not linear. This puzzling class of models seems to deserve a special study as deviation to normality is expected. We postpone it to a future work.

7. Concluding Remarks

In this note, we focused on discrete nonneutral Wright-Fisher models and on the conditions on the bias probabilities under which the forward branching dynamics is amenable to a dual discrete ancestral coalescent. It was shown that it concerns a large class of nonneutral models involving completely monotone bias probabilities. Several examples were supplied, some standard, some less classical. The Wright-Fisher model with forward binomial transition matrix is a particular instance of the Dirichlet model with Dirichlet-binomial transition matrix. Following the same lines, using the representation of the Dirichlet binomial distribution as a beta mixture of the binomial distribution, it would be interesting to exhibit that the corresponding conditions on the bias mechanism were the starting point to be a forward Dirichlet branching process. Also of particular interest in this respect would be the discrete nonneutral Moran models whose forward transition matrices are simpler because of their tridiagonal Jacobi structure. We hope to be able to consider shortly these cases (and maybe others) in a future work.

References


