HAL
open science

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

Thierry Huillet

To cite this version:

Thierry Huillet. A Duality Approach to the Genealogies of Discrete Non-Neutral Wright-Fisher Models. Journal of Probability and Statistics, 2009, Article ID 714701, pp.1-22. 10.1155/2009/714701 . hal-00337394

HAL Id: hal-00337394

https://hal.science/hal-00337394v1

Submitted on 6 Nov 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A DUALITY APPROACH TO THE GENEALOGIES OF DISCRETE NON-NEUTRAL WRIGHT-FISHER MODELS

THIERRY E. HUILLET

Abstract. Discrete ancestral problems arising in population genetics are investigated. In the neutral case, the duality concept has 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 non-neutral Wright-Fisher models. This concerns a large class of non-neutral 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,

Running title: Duality and Discrete Non-Neutral Wright-Fisher Models.

Keywords: Wright-Fisher Models; Markov chains; Duality; Mutational and evolutionary processes; Population dynamics; Phylogeny.

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 T. Nagylaki ('1999) for historical commentaries and on the role played by the French geneticist G. Malécot, starting shortly before the second world war. The book of W. Ewens ('2004) 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 J. Kingman ('1982). 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 '2004 for a review). It included incorporating variations in population size, mutation, recombination, selection... In ('1999), J. Pitman and S. Sagitov, 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 M. Möhle and S. Sagitov ('2001) and J. Schweinsberg ('2000). All these recent developments and improvements concern chiefly the discrete neutral case and their various scaling limits in continuous time and/or space. As was shown by Möhle ('1994 and '1999), neutral forward and backward theories learn much from one another by using a concept of duality

introduced by T. Liggett ('1985). 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, ('2004) for a review). In the works of C. Neuhauser and S. Krone ('1997), 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 T. Huillet ('2007) 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 non-neutral situations when various evolutionary forces are the causes of deviation to neutrality (see J. Crow and M. Kimura, ('1970), T. Maruyama ('1977), J. Gillepsie ('1991) and W. Ewens ('2004), for a discussion on various models of utmost interest in population genetics).

In this Note, we focus on discrete non-neutral Wright-Fisher (say 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 non-neutral Wright-Fisher models, at least for specific deviation forces to neutrality. It is shown that it concerns a large class of non-neutral 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 shall 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 (subsection 2.2) and backward processes (subsection 2.3) arising from exchangeable reproduction laws (subsection 2.1). In subsection 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 non-neutral situations. We start supplying important non-neutral examples in section 3. In section 4, we show that duality does indeed make sense in the framework of discrete non-neutral Wright-Fisher models, but only for the class of completely-monotone state-dependent transition frequencies. In section 5, we show that most non-neutrality 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 non-neutral situations to compute the extinction probabilities (invariant measure) of the dual backward ancestral process if one knows the invariant measure (respectively, 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: '1974). Consider a population with non-overlapping 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 $\boldsymbol{\nu}_n := (\nu_{1,n}, ..., \nu_{n,n})$, satisfying:

$$\sum_{m=1}^{n} \nu_{m,n} = n.$$

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

- Exchangeability: $(\nu_{1,n}, ..., \nu_{n,n}) \overset{d}{=} \left(\nu_{\sigma(1),n}, ..., \nu_{\sigma(n),n}\right)$, for all permutations $\sigma \in \mathcal{S}_n$.

- 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 $[n]^{\mathbb{Z}}$, where $[n] := \{0, 1, ..., n\}$ (see Karlin-McGregor, '1964).

Famous reproduction laws are:

Example 2.1.1 The multinomial Dirichlet model: $\boldsymbol{\nu}_n \overset{d}{\sim}$ Multin-Dirichlet$(n; \theta)$, where $\theta > 0$ is a disorder parameter. With $\mathbf{k}_n := (k_1, ..., k_n)$, $\boldsymbol{\nu}_n$ admits the following joint exchangeable distribution on the simplex $|\mathbf{k}_n| := \sum_{m=1}^{n} k_m = n$:

$$\mathbb{P}_\theta\left(\boldsymbol{\nu}_n = \mathbf{k}_n\right) = \frac{n!}{[n\theta]_n} \prod_{m=1}^{n} \frac{[\theta]_{k_m}}{k_m!},$$

where $[\theta]_k = \theta(\theta+1)...(\theta+k-1)$ is the rising factorial of θ. This distribution can be obtained by conditioning n independent mean 1 Pòlya distributed random variables $\boldsymbol{\xi}_n = (\xi_1, ..., \xi_n)$ on summing to n, that is to say: $\boldsymbol{\nu}_n \overset{d}{=} (\boldsymbol{\xi}_n \mid |\boldsymbol{\xi}_n| = n)$, where

$$\mathbb{P}_\theta\left(\xi_1 = k\right) = \frac{[\theta]_k}{k!}(1+\theta)^{-k}\left(\theta/(1+\theta)\right)^{\theta}, \; k \in \mathbb{N}.$$

When $\theta \uparrow \infty$, this distribution reduces to the Wright-Fisher model for which $\boldsymbol{\nu}_n \overset{d}{\sim}$ Multin$(n; 1/n, ..., 1/n)$. Indeed, $\boldsymbol{\nu}_n$ admits the following joint exchangeable multinomial distribution on the simplex $|\mathbf{k}_n| = n$:

$$\mathbb{P}_\infty\left(\boldsymbol{\nu}_n = \mathbf{k}_n\right) = \frac{n! \cdot n^{-n}}{\prod_{m=1}^{n} k_m!}.$$

This distribution can be obtained by conditioning n independent mean 1 Poisson distributed random variables $\boldsymbol{\xi}_n = (\xi_1, ..., \xi_n)$ on summing to n: $\boldsymbol{\nu}_n \overset{d}{=} (\boldsymbol{\xi}_n \mid |\boldsymbol{\xi}_n| = n)$. When n is large, using Stirling formula, $n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n}$; it follows that $\boldsymbol{\nu}_n \underset{n\uparrow\infty}{\overset{d}{\rightarrow}} \boldsymbol{\xi}_\infty$ with joint finite-dimensional law: $\mathbb{P}\left(\boldsymbol{\xi}_n = \mathbf{k}_n\right) = \prod_{m=1}^{n} \frac{e^{-1}}{k_m!} = \frac{e^{-n}}{\prod_{m=1}^{n} k_m!}$

on $\mathbb{N}^n$. Thanks to the product form of all finite-dimensional laws of $\boldsymbol{\xi}_\infty$, we get an asymptotic independence property of $\boldsymbol{\nu}_n$.

Example 2.1.2 In the Moran model, $\boldsymbol{\nu}_n \overset{d}{\sim}$ random permutation of $(2,0,1,...,1)$: in such a model, only one new gene per generation may come to life, at the expense of the simultaneous disappearance of some other gene.

2.2. **Forward in time branching process.** Take a sub-sample of size m from $[n] := \{0,1,...,n\}$ at generation 0. Let

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

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

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

It is a martingale, with state-space $\{0,...,n\}$, initial state m, absorbing states $\{0,n\}$ and transient states $\{1,...,n-1\}$. The first hitting time of boundaries $\{0,n\}$, which is: $\tau(m) = \tau_{\{0\}}(m) \wedge \tau_{\{n\}}(m)$ 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 Π_n with entries $\Pi_n(k,k') = \mathbb{P}(\nu_{1,n} + ... + \nu_{k,n} = k')$ given by (1). We have $\Pi_n(0,k') = \delta_{0,k'}$ and $\Pi_n(n,k') = \delta_{n,k'}$ and Π_n is not irreducible. However, Π_n is aperiodic and (apart from absorbing states) cannot be broken down into non-communicating subsets; as a result it is diagonalizable, with eigenvalues $|\lambda_0| \geq |\lambda_1| \geq ... \geq |\lambda_n|$ and $1 = \lambda_0 = \lambda_1 > |\lambda_2|$.

Example 2.2.1 (Dirichlet binomial): With U a $(0,1)-$valued random variable with density beta$(k\theta, (n-k)\theta)$

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

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

Example 2.2.2 The Wright-Fisher model has a Bin$(n, k/n)$ transition matrix:

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

Remark (statistical symmetry): Due to exchangeability of the reproduction law, neutral models are symmetric in the following sense: The transition probabilities of $\overline{N}_r(m) := n - N_r(m)$ 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 sub-sample 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:

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

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

$$\mathbb{P}\left(\mathcal{A}_{r+1}(m)=\alpha \mid \mathcal{A}_r(m)=\beta\right)=P_{\beta;\alpha};\text{ with }(\alpha,\beta)\in\mathcal{E}_m,\ \alpha\subseteq\beta$$

where, with $a=|\alpha|$ = number of equivalence classes of α, $b=|\beta|$ = number of equivalence classes of β, $\mathbf{b}_a:=(b_1,...,b_a)$ clusters sizes of β and $(m)_a:=m(m-1)...(m-a+1)$ a falling factorial,

$$P_{\beta;\alpha}=P_{b;a}^{(n)}(\mathbf{b}_a)=\frac{(n)_a}{(n)_b}\mathbb{E}\left(\prod_{l=1}^{a}(\nu_{l,n})_{b_l}\right)$$

is the probability of a $\mathbf{b}_a-$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 $b_1\geq...\geq b_c\geq 2$, $b_{c+1}=...=b_a=1$.

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 $b_1\geq...\geq b_c\geq 2$ occur.

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

The chain's state-space is: {equivalence relations on (partitions of) $\{1,...,m\}$}; 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 $\mathcal{A}_0=\{(1),...,(m)\}$, and terminal absorbing state $\{(1,...,m)\}$.

Examples:

From the Dirichlet example 2.2.1, we get: $P_{b;a}^{(n)}(\mathbf{b}_a)=\frac{(n)_a}{[n\theta]_b}\prod_{m=1}^{a}[\theta]_{b_m}$.

From the WF example 2.2.2: In this case, $P_{b;a}^{(n)}(\mathbf{b}_a)=\frac{(n)_a}{n^b}$ is the uniform distribution on $\{\mathbf{b}_a:b_1+...+b_a=b\}$.

The ancestral Count Process: Let

$$A_r(m)\ =\#\text{ ancestors at generation }r\in\mathbb{N},\text{ backward-in-time.}$$
$$\text{Then: }A_r(m)=\#\text{ blocks of }\mathcal{A}_r(m).$$

The backward ancestral count process is a discrete-time Markov chain with transition probabilities (Cannings, '1974 and Gladstien '1978):

$$\mathbb{P}\left(A_{r+1}(m)=a\mid A_r(m)=b\right)=P_{b,a}^{(n)}:=\frac{b!}{a!}\sum_{\substack{b_1,...,b_a\in\mathbb{N}_+\\ b_1+...+b_a=b}}\frac{P_{b;a}^{(n)}(\mathbf{b}_a)}{b_1!...b_a!}. \tag{2}$$

$$=\frac{\binom{n}{a}}{\binom{n}{b}}\sum_{\substack{b_1,...,b_a\in\mathbb{N}_+\\ b_1+...+b_a=b}}\mathbb{E}\left(\prod_{l=1}^{a}\binom{\nu_{l,n}}{b_l}\right).$$

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

$P_{b,a}^{(n)}$ given by (2). Note, by inclusion-exclusion principle, the alternative alternating expression of $P_{b,a}^{(n)}$:

$$P_{b,a}^{(n)} := \frac{\binom{n}{a}}{\binom{n}{b}} \sum_{m=0}^{a} (-1)^{a-m} \binom{a}{m} \mathbb{E}\left(\binom{\nu_{1,n} + ... + \nu_{m,n}}{b} \right).$$

2.4. **Duality (neutral case).** We start with a definition of the duality concept which is relevant to our purposes.

Definition [Liggett, '1985]: *Two Markov processes* $\left(X_t^1, X_t^2; t \geq 0\right)$, *with state-spaces* $(\mathcal{E}_1, \mathcal{E}_2)$, *are said to be dual with respect to some real-valued function* Φ *on the product space* $\mathcal{E}_1 \times \mathcal{E}_2$ *if* $\forall x_1 \in \mathcal{E}_1$, $\forall x_2 \in \mathcal{E}_2$, $\forall t \geq 0$:

$$\mathbb{E}_{x_1} \Phi\left(X_t^1, x_2\right) = \mathbb{E}_{x_2} \Phi\left(x_1, X_t^2\right). \tag{3}$$

We then recall basic examples of dual processes from the neutral and exchangeable population models (Möhle, '1997): 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:

$$(i\ )\ \Phi_n^1(m,k) = \binom{m}{k} / \binom{n}{k} \text{ and} \tag{4}$$

$$(ii)\ \Phi_n^2(m,k) = \binom{n-m}{k} / \binom{n}{k} \text{ on } \{0, ..., n\}^2.$$

Namely (i) reads:

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

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 a $k-$sample (without replacement) from population of size N_r at time r are all of type A, given $N_0 = m$. If this $k-$sample 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.

And (ii) reads:

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

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 this $k-$sample are all descendants of A_r ancestors at time $-r$, this probability must be equal to the probability that a $m-$sample from population of size A_r at time $-r$ are themselves all of type a.

With P_n' the transpose of P_n, a one-step $(r = 1)$ version of these formulae is:

$$(i)\ \Pi_n \Phi_n^1 = \Phi_n^1 P_n' \text{ and } (ii)\ \Pi_n \Phi_n^2 = \Phi_n^2 P_n'$$

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

$$\left[\Phi_n^1\right]^{-1}(i,j) = (-1)^{i-j}\binom{i}{j}\binom{n}{i}$$

and

$$\left[\Phi_n^2\right]^{-1}(i,j) = (-1)^{i+j-n}\binom{i}{n-j}\binom{n}{i} = (-1)^{i+j-n}\binom{j}{n-i}\binom{n}{j}.$$

The matrix $\left[\Phi_n^1\right]^{-1}$ is left-lower triangular, while $\left[\Phi_n^2\right]^{-1}$ is symmetric right-lower triangular. Thus,

$$(i)\ \ \left[\Phi_n^1\right]^{-1}\Pi_n\Phi_n^1 = P_n' \text{ and } (ii)\ \ \left[\Phi_n^2\right]^{-1}\Pi_n\Phi_n^2 = P_n'.$$

In any case, being similar matrices, Π_n and P_n' (or P_n) both share the same eigenvalues. If R_n diagonalizing Π_n is known so that $R_n^{-1}\Pi_n R_n = \Lambda_n := \operatorname{diag}(\lambda_0, ..., \lambda_n)$, the diagonal matrix of the eigenvalues of Π_n, then, with $\Phi_n = \Phi_n^1$ or Φ_n^2, $\widetilde{R}_n := \Phi_n^{-1}R_n$ diagonalizes P_n' and is obtained for free (and conversely). R_n is the matrix whose columns are the right-eigenvectors of Π_n and $\widetilde{R}_n$ 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 Π_n, $\widetilde{L}_n := L_n\Phi_n$ 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 Π_n is:

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

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

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

In Möhle '1999, 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 (1), (2) and (4).

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 non-neutral situations? We shall see that it does in discrete non-neutral Wright-Fisher models, but only for some class of state-dependent transition frequencies.

3. Beyond neutrality (symmetry breaking)

Discrete forward non-neutral models (with non-null drifts) can be obtained by substituting

$$k \to np\left(\frac{k}{n}\right) \text{ in } \mathbb{P}\left(\nu_{1,n} + ... + \nu_{k,n} = k'\right),$$

where:

$$p(x) : x \in (0,1) \to (0,1) \text{ is continuous, increasing, with } p(0) = 0,\ p(1) = 1.$$

$p(x)$ 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:

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

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

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

$$\text{Bin}(n, 1 - p(1 - k/n)) \neq \text{Bin}(n, p(k/n)),$$

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 $x \to p(x)$ is concave (convex), $N_r(m)$, $r \in \mathbb{N}$ is a submartingale (supermartingale), because: $\mathbb{E}(N_{r+1}(m) \mid N_r(m)) = np(N_r(m)/n) \geq N_r(m)$ (respectively $\leq N_r(m)$).

In the binomial neutral Wright-Fisher transition probabilities, we replaced the success probability $\frac{k}{n}$ by a more general function $p\left(\frac{k}{n}\right)$. 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: $\boldsymbol{\nu}_n \overset{d}{\sim} \text{Multin}(n; \boldsymbol{\pi}_n)$, where $\boldsymbol{\pi}_n := (\pi_{1,n}, ..., \pi_{n,n})$ and: $\pi_{m,n} = p\left(\frac{m}{n}\right) - p\left(\frac{m-1}{n}\right)$, $m = 1, ..., n$. We note that under our hypothesis,

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

Due to its asymmetry, the law of the biased $\boldsymbol{\nu}_n$ no longer is exchangeable.

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

Example 3.1 (homographic model, selection). Assume

$$p(x) = (1+s)\,x/(1+sx), \tag{5}$$

where $s > -1$ is a selection parameter. This model arises when gene A (respectively a), with frequency x (respectively $1-x$), has fitness $1+s$ (respectively 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

$$p(x) = \frac{(1+s)\,x^2 + (1+sh)\,x(1-x)}{1 + sx^2 + 2shx(1-x)}. \tag{6}$$

In this model, genotype AA (respectively Aa and aa), with frequency x^2 (respectively $2x(1-x)$ and $(1-x)^2$) has fitness $1+s$ (respectively $1+sh$ 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:

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

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: $p(x) = x + \frac{s}{2}\frac{x(1-x)}{1+sx}$.

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

$$p(x) = x(1 + c - cx). \tag{7}$$

If $c = 1$, $p(x) = x(2-x) = 1-(1-x)^2$: this bias appears in a discrete 2-sex population model (Möhle, '1994, '1998)). We shall 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 $0 \le p(0) \le p(1) \le 1$, $p(1) - p(0) \in [0,1)$.

Example 3.4 (affine model) Take for example

$$p(x) = (1-\mu_2)x + \mu_1(1-x), \tag{8}$$

where (μ_1, μ_2) are mutation probabilities, satisfying $\mu_1 \le 1-\mu_2$. It corresponds to the mutation scheme: $a \underset{\mu_2}{\overset{\mu_1}{\rightleftarrows}} A$. 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 Π_n is irreducible and even primitive and all states of this Markov chain are now recurrent. We have $\mathbb{P}(N_{r+1} > 0 \mid N_r = 0) = 1-(1-p(0))^n > 0$ and $\mathbb{P}(N_{r+1} < n \mid N_r = n) = 1 - p(1)^n > 0$ and the boundaries $\{0\}$ and $\{n\}$ no longer are strictly absorbing as there is a positive reflection probability inside the domain $\{0,1,...,n\}$.

For reasons to appear now, we shall 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:

$$(-1)^l q^{(l)}(x) \ge 0, \text{ for all } x \in (0,1),$$

for all order-l derivatives $q^{(l)}$ of q, $l \ge 0$. If $p(x)$ is such that q is CM, we shall call it an admissible bias mechanism.

4. Non-neutral Wright-Fisher models and duality

Preliminaries: Let $\mathbf{v}_n := (v(0), v(1), ..., v(n))$ be a $(n+1)-$vector of $[0,1]-$valued numbers. Define the backward difference operator ∇ acting on $\mathbf{v}_n$ by: $\nabla v(m) =$

$v(m) - v(m-1)$, $m = 1, ..., n$. We have $\nabla^0 v(m) = v(m)$, $\nabla^2 v(m) = v(m) - 2v(m-1) + v(m-2)$, etc..., and, starting from the endpoint $v(n)$

$$\nabla^j v(m) \mid_{m=n} = \sum_{l=0}^{j} (-1)^{j-l} \binom{j}{l} v(n-l), \; j = 0, ..., n.$$

Let u be some continuous function: $[0,1] \to [0,1]$. Consider the $(n+1)-$vector $\mathbf{u}_n := \left(u\left(\frac{0}{n}\right), ..., u\left(\frac{m}{n}\right), ..., u\left(\frac{n}{n}\right)\right)$, sampling u at points m/n. The function u is said to be $\nabla-$completely monotonic if $(-1)^j \nabla^j u\left(\frac{m}{n}\right) \mid_{m=n} \geq 0$, for all $j = 0, ..., n$ and all $n \geq 0$. Let $\left(u^1, u^2\right)$ be two continuous functions on $[0,1]$. Let $u = u^1 \cdot u^2$. With $\mathbf{u}_n$ the point-wise product of $\mathbf{u}_n^1$ and $\mathbf{u}_n^2$, assuming both functions $\left(u^1, u^2\right)$ 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:

Theorem: *Consider a non-neutral WF forward model $(N_r; r \in \mathbb{N})$ on $\{0, ..., n\}$, with continuous, non-decreasing bias $p(x)$, satisfying:*

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

This process has forward transition matrix:

$$\Pi_n(k, k') = \mathbb{P}(\nu_{1,n} + ... + \nu_{k,n} = k') = \binom{n}{k'} p\left(\frac{k}{n}\right)^{k'} \left(1 - p\left(\frac{k}{n}\right)\right)^{n-k'}.$$

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

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

P_n is a stochastic matrix if and only if $p(0) = 0$; else, if $p(0) > 0$, the matrix P_n is sub-stochastic.

Proof: Developing $\left[\Phi_n^2\right]^{-1} \Pi_n \Phi_n^2 = P_n'$, we easily obtain:

$$\begin{aligned} P_n^{'}(j,i) &= P_n(i,j) = \binom{n}{j} \sum_{l=0}^{j} (-1)^{j-l} \binom{j}{l} \left[1 - p\left(\frac{n-l}{n}\right)\right]^i \\ &= \binom{n}{j} (-1)^j \nabla^j \left(q\left(\frac{m}{n}\right)^i\right) \mid_{m=n} \end{aligned}$$

This entry is non-negative if and only if $(-1)^j \nabla^j \left(q\left(\frac{m}{n}\right)^i\right) \mid_{m=n} \geq 0$, for all $i, j = 0, ..., n$. But, due to the above argument on $\nabla-$complete monotonicity of integral powers, this will be the case if and only if $(-1)^j \nabla^j \left(q\left(\frac{m}{n}\right)\right) \mid_{m=n} \geq 0$, for all $j = 0, ..., n$. 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 Feller '1971, page 223, for absolutely monotone functions on $(0,1)$,

this will be the case if and only if $x \to q(x) := 1 - p(x)$ is a completely monotone function on $(0,1)$ in the sense that:

$$(-1)^l q^{(l)}(x) \geq 0, \text{ for all } x \in (0,1), \, l \in \mathbb{N}.$$

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

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

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 sub-stochastic.

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. □

5. Examples

We show here that most of the simplest non-neutrality 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 allows 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 $p(x) = x$ corresponding to the simple neutral case. Then $q(x) = 1 - x$ is completely monotone on $(0,1)$. With $S_{i,j}$ the second kind Stirling numbers, we get a lower left triangular stochastic transition matrix

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

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 (8) holds: $p(x) = (1 - \mu_2) x + \mu_1 (1 - x)$ where (μ_1, μ_2) are mutation probabilities. Then, with $\kappa := 1 - (\mu_1 + \mu_2)$, $q(x) = 1 - \mu_1 - \kappa x$ 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

$$P_n(i,j) = \binom{n}{j} \sum_{l=0}^{j} (-1)^{j-l} \binom{j}{l} \left(\mu_2 + \kappa \frac{l}{n}\right)^i =: (n)_j \cdot n^{-i} \cdot S_{i,j}^{\mu_2}(\kappa/n), \; j \leq i \tag{9}$$

$$P_n(i,j) = 0, \text{ else,}$$

in terms of generalized Stirling numbers $S_{i,j}^{\mu_2}(\kappa/n)$. We have $P_n(i,i) = (n)_i \left(\frac{\kappa}{n}\right)^i$ and the spectrum of P_n is real. When $\mu_1 > 0$, this matrix is sub-stochastic 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 $\mu_1 = 0$, $P_n(i,j) = (n)_j \cdot n^{-i} \cdot S_{i,j}^{\mu_2}((1-\mu_2)/n)$, $j \leq i$, $= 0$, else. The corresponding matrix P_n is stochastic.

From example 3.3 (quadratic). Assume $p(x) = x(1+c-cx)$, as in (7). 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 $\nabla^j \left(q\left(\frac{m}{n}\right)^i\right) = 0$ 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 (5) holds

$$p(x) = (1+s)\,x/(1+sx),$$

$q(x) = 1 - p(x) = (1-x)/(1+sx)$ 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, 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 '1997).

From example 3.2 (selection with dominance). The corresponding mechanism (6) with parameters (s,h) satisfying $s > -1$ and $sh > -1$ is CM if and only if $s > 0$ and $h \in (0,1/2)$. The case $h \in (0,1)$ 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 5.1.1 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 5.1.2 (Reciprocal mechanism) If $p(x)$ is not admissible in that q is not CM, it can be that $\overline{p}(x) := 1 - p(1-x)$ is itself admissible. As observed before, if $N_r(m)$ has transition probabilities given by $\mathrm{Bin}(n, p(k/n))$, $\overline{p}(x)$ arises in the transition probabilities of $\overline{N}_r(m) := n - N_r(m)$. Indeed, such transitions are $\mathrm{Bin}(n, \overline{p}(k/n))$ distributed.

If $p(x)$ is the selection mechanism of example 3.1, (5), with $s \in (-1,0)$ (not admissible), $\overline{p}(x) = (1+\overline{s})x/(1+\overline{s}x)$ is itself an admissible selection mechanism because it has reciprocal selection parameter $\overline{s} = -s/(1+s) > 0$. If $p(x)$ is the mechanism of example 3.2, namely (6), with parameters (s,h), then $\overline{p}(x)$ is itself a selection with dominance mechanism with reciprocal parameters $\overline{s} = -s/(1+s)$ and $\overline{h} = 1-h$. Assuming $(s<0, h \in (1/2,1))$, $p(x)$ is not admissible whereas $\overline{p}(x)$ is because $\overline{s} > 0$ and $\overline{h} \in (0,1/2)$. Similarly, when $\gamma \in (0,1)$, the mechanism $p(x) = 1-(1-x)^{\gamma}$ is not admissible but, from example 5.1.1, $\overline{p}(x) := 1-p(1-x) = x^{\gamma}$ is.

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 $\kappa = 1-(\mu_1+\mu_2) \geq 0$). Let $p(x)$ be a bias mechanism such that $q(x)$ is CM with $p(1)-p(0)=1$. Then

$$\widetilde{p}_M(x) = p_M(p(x))$$

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

$$\widetilde{p}_M(x) = \frac{\mu_1 + x((1+s)(1-\mu_2)-\mu_1)}{1+sx}$$

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

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 $p^1\left(p^2(x)\right)$ 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

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

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

$$\left(p^1(x), p^2(x)\right) \to p(x) = p^1(x) \circ p^2(x).$$

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

$$q(x) = \frac{(1-\mu_1-\kappa x)(1-x)}{1+sx} \text{ and } p(x) = \frac{\mu_1 + x(s+1-\mu_1+\kappa) - \kappa x^2}{1+sx},$$

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

Example 5.3.2 (Joint mutation and neutral effects). Let $p^1(x) = (1-\mu_2)x + \mu_1(1-x)$ and $p^2(x) = x$. We get

$$q(x) = (1-x)(1-\mu_1-\kappa x) \text{ and } p(x) = \mu_1 + x(1-\mu_1+\kappa(1-x)),$$

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 $\boldsymbol{\nu}_n^j \overset{d}{\sim} \text{Multin}\left(n; \boldsymbol{\pi}_n^j\right)$, where $\pi_{m,n}^j = p^j\left(\frac{m}{n}\right) - p^j\left(\frac{m-1}{n}\right)$, $m = 1, ..., n$. Then, $\boldsymbol{\nu}_n \overset{d}{\sim} \text{Multin}(n; \boldsymbol{\pi}_n)$, $\pi_{m,n} = p\left(\frac{m}{n}\right) - p\left(\frac{m-1}{n}\right)$, $m = 1, ..., n$, where $\boldsymbol{\pi}_n := \boldsymbol{\pi}_n^1 \odot \boldsymbol{\pi}_n^2$ is easily obtained componentwise by:

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

We let: $\boldsymbol{\nu}_n := \boldsymbol{\nu}_n^1 \odot \boldsymbol{\nu}_n^2 \overset{d}{\sim} \text{Multin}\left(n; \boldsymbol{\pi}_n^1 \odot \boldsymbol{\pi}_n^2\right)$. 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 ϕ, 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: $\phi(x) = \mathbb{E}\left[x^N\right]$. 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 5.3.3 The general mechanism with mutational effects is in this class. Indeed,

$$\widetilde{q}_M(x) = 1 - \widetilde{p}_M(x) = 1 - p_M(p(x)) = 1 - p_M(1 - q(x))$$

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 5.3.4 With $\theta > 0$, taking $\phi(x) = e^{-\theta(1-x)}$ or $\left(e^{\theta x} - 1\right) / \left(e^{\theta} - 1\right)$, 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_ϕ where ϕ is the pgf of a Poisson random variable, then $q_\phi(x)^\alpha$ 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 (5) occurring in the model with selection of example 3.1.

Example 5.3.5 Let $p(x) = x^\gamma$ with $\gamma \in (0,1)$ as in example 5.1.1. 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 $p_\phi(x) \underset{x \downarrow 0}{\sim} \theta x^\gamma$. The reciprocal function $\overline{p}_\phi(x) = q_\phi(1 - x) = e^{-\theta(1-x)^\gamma}$ also interprets as an absolutely monotone discrete-stable pgf (see Steutel, van Harn, '1979). It is not admissible.

Proceeding in this way, one can produce a wealth of admissible bias probabilities p_ϕ, 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 $\{0, ..., n\}$ with forward transition matrix:

$$\Pi_n(k, k') = \binom{n}{k'} p\left(\frac{k}{n}\right)^{k'} \left(1 - p\left(\frac{k}{n}\right)\right)^{n-k'},$$

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

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

Then, $(N_r, A_r; r \in \mathbb{N})$ are dual with respect to $\Phi_n(m, k) := \Phi_n^2(m, k) = \binom{n-m}{k} / \binom{n}{k}$, to wit:

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

We shall distinguish two cases.

Case 1: Assume first that

$$N_r \xrightarrow{d} N_\infty \text{ as } r \uparrow \infty, \text{ independently of } N_0 = m \geq 1.$$

Let $\pi_\infty(i) = \mathbb{P}(N_\infty = i)$ and $\boldsymbol{\pi}_\infty = (\pi_\infty(0), ..., \pi_\infty(n))'$. The line vector $\boldsymbol{\pi}'_\infty$ is the left eigenvector of Π_n associated to the eigenvalue $1 : \boldsymbol{\pi}'_\infty = \boldsymbol{\pi}'_\infty \Pi_n$. 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, $A_r \to 0$ 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, $\boldsymbol{\rho}_\infty = (\rho_\infty(0), ..., \rho_\infty(n))'$ is the unique solution to $(I - P_n)\boldsymbol{\rho}_\infty = 0$ with $\rho_\infty(0) = 1$.

Remark: 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 Π_n has an algebraically simple dominant eigenvalue 1. By Perron-Frobenius theorem:

$$\lim_{r \uparrow \infty} \Pi_n^r = \mathbf{1}\boldsymbol{\pi}'_\infty,$$

where $\mathbf{1}' = (1, ..., 1)$. The invariant probability measure can be approximated by subsequent iterates of Π_n, the convergence being exponentially fast, with rate governed by the second largest eigenvalue. Of course, detailed balance (stating that $\pi_k \Pi_n(k, k') = \pi_{k'} \Pi_n(k', k)$) 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 $\{\partial\}$, before hitting $\{0\}$. □

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

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

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

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

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

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

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

Let $\boldsymbol{\rho}_\infty=(\rho_\infty(0),...,\rho_\infty(n))'$. The vector $\boldsymbol{\rho}_\infty$ is the right eigenvector of P_n associated to the eigenvalue $1:\boldsymbol{\rho}_\infty=P_n\boldsymbol{\rho}_\infty$. 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:

$$\Pi_n\Phi_n=\Phi_nP_n'.$$

where the matrix Φ_n is symmetric whereas the matrix Φ_n^{-1} is symmetric right-lower triangular, with:

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

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

Thus,

$$\boldsymbol{\pi}_\infty'\Pi_n\Phi_n=\boldsymbol{\pi}_\infty'\Phi_n=\boldsymbol{\pi}_\infty'\Phi_nP_n',$$

showing that $\boldsymbol{\rho}_\infty$ and $\boldsymbol{\pi}_\infty$ are related through:

$$\boldsymbol{\rho}_\infty=\Phi_n\boldsymbol{\pi}_\infty\text{ or }\boldsymbol{\pi}_\infty=\Phi_n^{-1}\boldsymbol{\rho}_\infty.$$

The knowledge of the invariant measure $\boldsymbol{\pi}_\infty$ of the forward process allows one to compute the extinction probabilities $\boldsymbol{\rho}_\infty$ of the dual backward ancestral process (and conversely).

Example: Consider the discrete WF model with mutations of example 3.4. In this case, $N_r\overset{d}{\to}N_\infty$ 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 is sub-stochastic (if $\mu_1>0$) and lower-left triangular. Due to triangularity, the right eigenvector $\boldsymbol{\rho}_\infty$ of P_n can easily be computed explicitly in terms of $(P_n(i,j);\ j\leq i),i=0,...,n$. We therefore get the following alternating expression for the invariant measure:

$$\pi_\infty(i)=\binom{n}{i}\sum_{j=0}^{i}(-1)^{i-j}\binom{i}{j}\rho_\infty(n-j).$$

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 (9) we also have:

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

allowing to compute $\mathbb{E}\left[N_\infty^2\right]$ and then the variance of N_∞. We get:

$$\sigma^2(N_\infty) = \frac{n^2\mu_1\mu_2}{(\mu_1+\mu_2)^2(2n(\mu_1+\mu_2)+1)} + o(n) \underset{n\uparrow\infty}{\sim} \frac{\mu_1\mu_2}{2(\mu_1+\mu_2)^3}n,$$

suggesting (when $\mu_1\mu_2 > 0$) a Central Limit Theorem for N_∞ as n grows large:

$$\frac{1}{\sqrt{n}}\left(N_\infty - n\frac{\mu_1}{\mu_1+\mu_2}\right) \underset{n\uparrow\infty}{\overset{d}{\to}} \mathcal{N}\left(0, \frac{\mu_1\mu_2}{2(\mu_1+\mu_2)^3}\right).$$

Case 2. Conversely, assume now that given $N_0 = m$

$$N_r \overset{d}{\to} 0 \text{ as } r\uparrow\infty, \text{ with probability } \mathbb{P}_m(N_\infty = 0) =: \rho_\infty(m),$$

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

$$A_r \overset{d}{\to} A_\infty \text{ as } r\uparrow\infty, \text{ independently of } A_0 = k\in[n].$$

In terms of moments, the duality formula means that:

$$\binom{n}{m}^{-1}\mathbb{E}\left[\binom{n-A_\infty}{m}\right] = \rho_\infty(m) = \mathbb{P}_m(N_\infty = 0),$$

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_∞ is ($u\in[0,1]$):

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

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

$$\boldsymbol{\rho}_\infty = \Phi_n\boldsymbol{\pi}_\infty \text{ or } \boldsymbol{\pi}_\infty = \Phi_n^{-1}\boldsymbol{\rho}_\infty.$$

The knowledge of the extinction probabilities $\boldsymbol{\rho}_\infty$ of the forward process allows one to compute the invariant measure $\boldsymbol{\pi}_\infty$ 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 non-neutral 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, $\pi_\infty(i) = \binom{n}{i}\sum_{j=0}^{i}(-1)^{i-j}\binom{i}{j}\frac{j}{n} = \delta_{i,1}$ and $A_r \xrightarrow{d} 1$ as $r \uparrow \infty$, the degenerate state reached when the most recent common ancestor (MRCA) is attained.

(ii) Non-degenerate solutions of A_∞ 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 $p(x) \underset{x\uparrow 0}{\sim} \lambda x$, $\lambda > 1$. Then, due to large sample asymptotic independence:

$$\boldsymbol{\nu}_n \xrightarrow{d} \boldsymbol{\xi}_\infty,$$

where $\boldsymbol{\xi}_\infty$ is an iid sequence with $\xi_1 \overset{d}{\sim}$ Poisson(λ) (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 $\lim_{n\uparrow\infty}\rho_\infty(m) = \rho^m$, $m = 1, 2, ...$, where $0 < \rho < 1$ is the smallest solution to the fixed point equation

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

ρ is the singleton extinction probability of a super-critical Galton-Watson process with offspring distribution Poisson(λ). More precisely, proceeding as in Möhle '1994, Theorem 4.5, we have

$$n\left(\rho^m - \rho_\infty(m)\right) = \rho^m\left(\frac{1-\rho}{1+\lambda\rho}m^2 + \frac{\lambda(1-\rho)\rho}{1-(\lambda\rho)^2}m\right),$$

showing that the convergence of $\rho_\infty(m)$ to ρ^m is of order n^{-1}. As a result, we get the asymptotic normality:

$$\frac{1}{\sqrt{n}}\left(A_\infty - n(1-\rho)\right) \underset{n\uparrow\infty}{\xrightarrow{d}} \mathcal{N}\left(0, \frac{\rho(1-\rho)}{1+\lambda\rho}\right).$$

Intuitively, $\frac{1}{n}\mathbb{E}[n - A_\infty] = \rho_\infty(1) = \mathbb{P}_1(N_\infty = 0) \to \rho$, showing that $\mathbb{E}[A_\infty] \underset{n\uparrow\infty}{\sim} n(1-\rho)$ and

$$\frac{1}{n(n-1)}\mathbb{E}\left[(n - A_\infty)(n - 1 - A_\infty)\right] = \rho_\infty(2) = \mathbb{P}_2(N_\infty = 0) \to \rho^2,$$

showing that $\sigma^2(A_\infty) \underset{n\uparrow\infty}{\sim} n\rho(1-\rho)/(1+\lambda\rho)$.

For the 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 the example 3.1 with selection, $p(x) = (1+s)x/(1+sx)$, with $s > -1$: $p(x) \underset{x\uparrow 0}{\sim} (1+s)x$ and so $\lambda = 1 + s$. We have asymptotic normality only when $s > 0$, i.e. 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 $p(x) = x^\gamma$ of example 5.1.1 (or more generally for mechanisms satisfying $p(x) \underset{x\downarrow 0}{\sim} \theta x^\gamma$, $\theta > 0$ as in the compound bias example 5.3.5) 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 non-neutral 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 non-neutral 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 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 non-neutral 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

[1] Cannings, C. The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. Advances in Appl. Probability 6 (1974), 260–290.

[2] Crow, J.; F.; Kimura, M. *An introduction to population genetics theory.* Harper & Row, Publishers, New York-London (1970).

[3] Ewens, W. J. *Mathematical population genetics. I. Theoretical introduction.* Second edition. Interdisciplinary Applied Mathematics, 27. Springer-Verlag, New York, (2004).

[4] Feller, W. *An introduction to probability theory and its applications,* Vol. **2**. John Wiley and Sons, Second Edition, New York (1971).

[5] Gillespie, J. H. *The Causes of Molecular Evolution.* New York and Oxford: Oxford University Press, (1991).

[6] Gladstien, K. The characteristic values and vectors for a class of stochastic matrices arising in genetics. SIAM J. Appl. Math. 34 (1978), no. 4, 630–642.

[7] Huillet, T. On Wright-Fisher diffusion and its relatives. J. Stat. Mech., Th. and Exp. P11006, vol. 11, (2007).

[8] Karlin, S.; McGregor, J. Direct product branching processes and related Markov chains. Proc. Nat. Acad. Sci. U.S.A., 51, 1964, 598–602.

[9] Kingman, J.F.C. The coalescent. Stochastic Process. Appl., 13, 235-248, (1982).

[10] Liggett, T. M. *Interacting particle systems.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 276. Springer-Verlag, New York, (1985).

[11] Maruyama, T. *Stochastic problems in population genetics.* Lecture Notes in Biomathematics, 17. Springer-Verlag, Berlin-New York, (1977).

[12] Möhle, M. Forward and backward processes in bisexual models with fixed population sizes. J. Appl. Probab. 31 (1994), no. 2, 309–332.

[13] Möhle, M. The concept of duality and applications to Markov processes arising in neutral population genetics models. Bernoulli 5 (1999), no. 5, 761–777.

[14] Möhle, M.; Sagitov, S. A classification of coalescent processes for haploid exchangeable population models. Ann. Probab. 29 (2001), no. 4, 1547–1562.

[15] Nagylaki, T. Anecdotal, historical and critical commentaries on Genetics. Gustave Malécot and the transition from classical to modern population genetics. Edited by James F. Crow and William F. Dove, Genetics, 122, 253-268, (1989).

[16] Neuhauser, C.; Krone, S. M. The genealogy of samples in models with selection. Genetics, Vol 145, 519-534, (1997).

[17] Pitman, J. Coalescents with multiple collisions. Ann. Probab. 27 (1999), no. 4, 1870–1902.

[18] Sagitov, S. The general coalescent with asynchronous mergers of ancestral lines. J. Appl. Probab. 36 (1999), no. 4, 1116–1125.

[19] Schweinsberg, J. Coalescents with simultaneous multiple collisions. Electron. J. Probab. 5 (2000), Paper no. 12, 50 pp.
[20] Steutel, F. W.; van Harn, K. Discrete analogues of self-decomposability and stability. Ann. Probab. 7 (1979), no. 5, 893–899.
[21] Tavaré, S. Ancestral inference in population genetics. Lectures on probability theory and statistics, Saint-Flour 2001, Lecture Notes in Math., 1837, (1-188) Springer, (2004).

Laboratoire de Physique Théorique et Modélisation, CNRS-UMR 8089 et Université de Cergy-Pontoise, 2 Avenue Adolphe Chauvin, 95302, Cergy-Pontoise, FRANCE, E-mail: Thierry.Huillet@u-cergy.fr