

# SOME MARKOV CHAINS ARISING IN POPULATION GENETICS

THIERRY E. HUILLET

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

## 1. GW BRANCHING PROCESS (REMINDER)

1.1. **The pgf approach.** Consider a discrete-time Bienyamé-Galton-Watson branching process (Harris (1963)) whose reproduction law is given by the (sub-)probability law  $\mathbf{P}(\nu = k) =: \pi(k)$ ,  $k \geq 0$  for the number  $\nu$  of offspring per capita. We assume  $\pi(0) > 0$  so that the process can go extinct. We let  $\phi(z) = \mathbf{E}(z^\nu) = \sum_{k \geq 0} \pi(k) z^k$  be the probability generating function (pgf) of  $\nu$  and we assume  $\phi(1) \leq 1$ .

With  $N_r(1)$  the number of individuals alive at generation  $r$  given  $N_0 = 1$ , we have (with convention  $\sum_{l=1}^0 \nu_l = 0$ )

$$N_{r+1}(1) = \sum_{l=1}^{N_r(1)} \nu_l, \quad r \geq 0$$

and so,

$$\Phi_{r+1}(z) = \phi(\Phi_r(z)), \quad \Phi_0(z) = z$$

with

$$(1) \quad \Phi_r(z) := \mathbf{E}\left(z^{N_r(1)}\right) := \mathbf{E}\left(z^{N_r} \mid N_0 = 1\right) = \phi^{\circ r}(z),$$

where  $\phi^{\circ r}(z)$  is the  $r$ -th composition of  $\phi(z)$  with itself, <sup>1</sup>. Similarly, if  $N_r(m)$  is the number of individuals alive at generation  $r$  given there are  $N_0 = m$  independent founders, we clearly get

$$(2) \quad \mathbf{E}\left(z^{N_r(m)}\right) := \mathbf{E}\left(z^{N_r} \mid N_0 = m\right) = \phi^{\circ r}(z)^m.$$

We shall also let

$$\tau_{m,m'} = \inf(r \geq 1 : N_r = m' \mid N_0 = m),$$

the first hitting time of state  $m' \neq m$  given  $N_0 = m \neq 0$ . In particular (time to extinction),

$$\tau_{m,0} = \inf(r \geq 1 : N_r(m) = 0).$$

---

<sup>1</sup>Throughout this work, a pgf will therefore be a function  $\phi$  which is absolutely monotone on  $(0, 1)$  with all nonnegative derivatives of any order there, obeying  $\phi(1) \leq 1$ .

- If  $\phi(1) < 1$ , there is a positive probability  $1 - \phi(1) =: \pi(\infty)$  that  $\nu = \infty$  (explosion is made possible even at the first branching step): following Sagitov and Lindo (2015), we shall speak of an explosive or non-regular process.

- If  $\phi(1) = 1$  (regular case), depending on  $\mu := \mathbf{E}(\nu) \leq 1$  (i.e. the (sub-)critical case) or  $\mu > 1$  (supercritical case): the process  $N_r(1)$  goes extinct with probability 1 or goes extinct with probability  $\rho < 1$  where  $\rho$  is the smallest fixed point solution in  $[0, 1]$  to  $\phi(\rho) = \rho$ , respectively. Results from  $\rho = \Phi_\infty(0) = \mathbf{P}(N_\infty(1) = 0)$ .

Specifically,

- If  $\mu := \mathbf{E}(\nu) = \phi'(1) < 1$ ,  $N_r(1)$  goes extinct wp 1: subcritical case.  $\mathbf{E}(N_r(1)) = \mu^r \rightarrow 0$  (extinction). If  $\mathbf{E}(\nu \log \nu) < \infty$ ,

$$\mu^{-r} \mathbf{P}(\tau_{1,0} > r) \rightarrow C > 0$$

(geometric tails, fast extinction).

- If  $\mu > 1$ ,  $N_r(1)$  goes extinct wp  $\rho$  and diverges to  $\infty$ , wp  $\bar{\rho} := 1 - \rho$ : supercritical case.  $\mathbf{E}(N_r(1)) = \mu^r \rightarrow \infty$  (Malthus growth). Time to extinction:

$$\phi'(\rho)^{-r} [\mathbf{P}(\tau_{1,0} > r) - \bar{\rho}] \rightarrow C > 0, \quad \phi'(\rho) \in (0, 1).$$

Gives the geometric rate of approach to  $\bar{\rho}$ .

- If  $\phi'(1) = 1$ ,  $N_r(1)$  goes extinct wp 1: critical case. But it will take a long time to do so! If  $\mathbf{E}(\nu^2) < \infty$ :

$$r \mathbf{P}(\tau_{1,0} > r) \rightarrow C > 0$$

(Pareto tails, slow extinction).

In the critical case ( $N_r(m)$  martingale), for each  $r \geq 0$ ,  $\mathbf{E}(N_r(m)) = m$  and the initial population size is conserved on average.

In the supercritical case, the distribution of the time to extinction  $\tau_{1,0}$  is given by

$$\mathbf{P}(\tau_{1,0} \leq r) = \mathbf{P}(N_r(1) = 0) = \phi^{\circ r}(0),$$

and the process explodes with probability  $\bar{\rho} := 1 - \rho$ , but not in finite time: only state  $\{0\}$  is absorbing. Clearly also, if there are  $m$  independent founders instead of simply 1,

$$\mathbf{P}(\tau_{m,0} \leq r) = \mathbf{P}(N_r(m) = 0) = \phi^{\circ r}(0)^m.$$

- If  $\phi(1) < 1$  (explosive case),  $\mu := \mathbf{E}(\nu) = \infty$  because there is a positive probability  $1 - \phi(1)$  that  $\nu = \infty$ . Notice that

$$\phi'(1) = \mathbf{E}(\nu \cdot \mathbf{1}_{\{\nu < \infty\}}) = \sum_{k \geq 1} k \pi(k),$$

if this quantity exists (is finite). If  $\phi(1) < 1$ , state  $\{\infty\}$  should be added to the state-space  $\mathbb{N}_0 = \{0, 1, \dots\}$  of  $N_r(m)$  and then both states are  $\{0, \infty\}$  are absorbing.

In this supercritical case,  $\rho < 1$  always, and both the time to extinction  $\tau_{1,0}$  and the time to explosion  $\tau_{1,\infty}$  of  $N_r(1)$  are finite with positive probability, now with

$$(3) \quad \begin{cases} \mathbf{P}(\tau_{1,0} \leq r) = \mathbf{P}(N_r(1) = 0) = \phi^{\circ r}(0) \xrightarrow{r \rightarrow \infty} \rho = \mathbf{P}(\tau_{1,0} < \infty). \\ \mathbf{P}(\tau_{1,\infty} > r) = \mathbf{P}(N_r(1) < \infty) = \phi^{\circ r}(1) \xrightarrow{r \rightarrow \infty} \rho = \mathbf{P}(\tau_{1,\infty} = \infty). \end{cases}$$

Thus  $\rho$  and  $\bar{\rho}$  are now also the probabilities that  $\tau_{1,0} < \infty$  and  $\tau_{1,\infty} < \infty$ , respectively. We thus have

$$(4) \quad \begin{cases} \mathbf{P}(r < \tau_{1,0} < \infty) = \rho - \phi^{\circ r}(0), \\ \mathbf{P}(r < \tau_{1,\infty} < \infty) = \bar{\rho} - (1 - \phi^{\circ r}(1)) = \phi^{\circ r}(1) - \rho, \text{ and,} \\ \mathbf{P}(r < \tau_1 < \infty) = \mathbf{P}(0 < N_r(1) < \infty) = \phi^{\circ r}(1) - \phi^{\circ r}(0), \end{cases}$$

where we defined the global absorption time  $\tau_1 := \tau_{1,0} \wedge \tau_{1,\infty}$ . Clearly also, with  $\tau_m := \tau_{m,0} \wedge \tau_{m,\infty}$

$$(5) \quad \begin{cases} \mathbf{P}(r < \tau_{m,0} < \infty) = \rho^m - \phi^{\circ r}(0)^m, \\ \mathbf{P}(r < \tau_{m,\infty} < \infty) = (1 - \rho^m) - (1 - \phi^{\circ r}(1)^m) = \phi^{\circ r}(1)^m - \rho^m, \text{ and,} \\ \mathbf{P}(r < \tau_m < \infty) = \mathbf{P}(0 < N_r(m) < \infty) = \phi^{\circ r}(1)^m - \phi^{\circ r}(0)^m. \end{cases}$$

**1.2. The (stochastic) transition matrix approach.** A Bienaymé-Galton-Watson process is a time-homogeneous Markov chain with denumerable state-space  $\mathbb{N}_0 := \{0, 1, \dots\}$ . Its stochastic one-step transition matrix is  $P$ , with entries

$$P(m, m') = [z^{m'}] \phi(z)^m = \mathbf{P}(N_1(m) = m')$$

(with  $[z^{m'}] \phi(z)^m$  meaning the  $z^{m'}$ -coefficient of the pgf  $\phi(z)^m$ ). Initial condition  $\mathbf{P}(N_0(m) = m') = \delta_{m,m'}$ . Note  $P(0, m') = \delta_{m',0}$  (0 absorbing).

When there is explosion and in the supercritical cases, an interesting problem arises when conditioning  $N_r$  either on extinction or on explosion. This may be understood as follows:

The harmonic column vector  $\varphi$ , solution to  $P\varphi = \varphi$ , is given by its coordinates  $\varphi(m) = \rho^m$ ,  $m \geq 0$ , because  $\sum_{m' \geq 0} P(m, m') \rho^{m'} = \phi(\rho)^m = \rho^m$ .

Consequently, with  $\tau_m = \tau_{m,0} \wedge \tau_{m,\infty}$ ,  $\varphi(N_{r \wedge \tau_m}(m)) = \rho^{N_{r \wedge \tau_m}(m)}$  is a martingale because  $\mathbf{E}(\varphi(N_{r \wedge \tau_m}(m))) = \phi^{\circ r}(\rho)^m = \rho^m = \varphi(m) = \varphi(N_0(m))$ . Hence

$$(6) \quad \begin{aligned} \mathbf{E}(\varphi(N_{r \wedge \tau_m}(m))) &= \varphi(m) = \varphi(0) \mathbf{P}(\tau_{m,0} < \tau_{m,\infty}) + \varphi(\infty) \mathbf{P}(\tau_{m,\infty} < \tau_{m,0}) \\ &= \rho^m = \mathbf{P}(\tau_{m,0} < \tau_{m,\infty}) + 0 \end{aligned}$$

Letting  $D_\varphi := \text{diag}(\varphi(0), \varphi(1), \dots)$ , introduce the stochastic matrix  $P_\varphi$  given by a Doob transform (Norris (1998) and Rogers and Williams (1994), p. 327):

$$P_\varphi := D_\varphi^{-1} P D_\varphi$$

or  $P_\varphi(m, m') = \varphi(m)^{-1} P(m, m') \varphi(m') = P(m, m') \rho^{m'-m}$ ,  $m, m' \geq 0$ .

Then  $P_\varphi$  is the transition matrix of  $N_r$  conditioned on almost sure extinction.

Equivalently, when conditioning  $N_r$  on almost sure extinction, one is led to a regular subcritical BGW process with new branching mechanism  $\phi_0(z) = \phi(\rho z)/\rho$ , satisfying  $\phi_0(1) = 1$  and  $\phi'_0(1) = \phi'(\rho) < 1$ . Indeed,  $\phi_0(z) = \sum_{m' \geq 0} P_\varphi(1, m') z^{m'}$ . Upon iterating, we get  $\phi_0^{\circ r}(z) = \phi^{\circ r}(\rho z)/\rho$ .

Similarly, when conditioning  $N_r$  on almost sure explosion, one is led to an explosive supercritical BGW process with new Harris-Sevastyanov branching mechanism  $\phi_\infty(z) = [\phi(\rho + \bar{\rho}z) - \rho]/\bar{\rho}$ , satisfying  $\phi_\infty(0) = 0$  and  $\phi_\infty(1) = (\phi(1) - \rho)/\bar{\rho} < 1$ . Upon iterating, we have  $\phi_\infty^{\circ r}(z) = [\phi^{\circ r}(\rho + \bar{\rho}z) - \rho]/\bar{\rho}$ .

The second largest eigenvalue of  $P$  is  $\gamma = \phi'(\rho) < 1$ . The corresponding eigenvector  $\mathbf{u}$  obeys  $P\mathbf{u} = \gamma\mathbf{u}$  with  $u(m) = m\rho^{m-1}$ ,  $m \geq 1$ , because  $\sum_{m' \geq 1} P(m, m') m' \rho^{m'-1} = \phi'(\rho) m \phi(\rho)^{m-1} = \gamma m \rho^{m-1}$ . Conditioning  $N_r$  on never hitting  $\{0, \infty\}$  in the remote future is given by the  $Q$ -process with stochastic transition matrix  $Q = \gamma^{-1} D_{\mathbf{u}}^{-1} P D_{\mathbf{u}}$  or  $Q(m, m') = \gamma^{-1} u(m)^{-1} P(m, m') u(m') = \gamma^{-1} \rho^{m'-m} m^{-1} P(m, m') m'$ ,  $m, m' \geq 1$  (see Lambert (2010) and Sagitov and Lindo (2015), Section 6 in the  $\theta$ -special case).

There are classes of discrete branching processes for which the pgf  $\phi^{\circ r}(z)$  of  $N_r(1)$  is exactly computable, thereby making the above computations concrete and somehow explicit.

**1.3. GW with immigration.** GW very unstable (transient)! Immigration stabilizes it (recurrence very likely).

With  $(\xi_1, \xi_2, \dots)$  iid  $\mathbb{N}$ -valued sequence representing the adjunction of a random number of immigrants at each generation (possibly none if  $\xi = 0$ ), let

$$N_{r+1}(1) = \sum_{l=1}^{N_r(1)} \nu_l + \xi_{r+1}, \quad r \geq 0$$

If  $\mathbf{E}(\nu) < 1$ , subcritical GW process goes extinct wp 1 and is reinvigorated by the arrival of immigrants. Competition between death and birth! With  $\phi_\xi(z) = \mathbf{E}(z^\xi)$  the pgf of  $\xi$ , assuming  $N_0(1) = m$ , we get the recurrence for  $\Phi_r(z) = \mathbf{E}(z^{N_r(1)})$

$$\Phi_{r+1}(z) = \phi_\xi(z) \Phi_r(\phi(z)), \quad \Phi_0(z) = z^m.$$

Iterating,

$$\begin{aligned} \Phi_r(z) &= \phi^{\circ r}(z)^m \prod_{r'=0}^{r-1} \phi_\xi(\phi^{\circ r'}(z)), \quad r \geq 1. \\ P^r(m, m') &= \left[ z^{m'} \right] \phi^{\circ r}(z)^m \prod_{r'=0}^{r-1} \phi_\xi(\phi^{\circ r'}(z)) \quad (r \text{ steps}) \\ P(m, m') &= \left[ z^{m'} \right] \phi(z)^m \phi_\xi(z) \\ &= (\pi^{*m} * \mu)(m') \quad (\text{one-step}) \end{aligned}$$

if  $\mu_m = \mathbf{P}(\xi_1 = m)$ . Existence of invariant probability measure:  $\Phi_\infty(z)$  is a proper pgf iff  $\mathbf{E}(\log_+ \xi) < \infty$ .

## 2. BRANCHING WITH CONSTANT POPULATION SIZE. 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. Gillespie ('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.

### 3. DISCRETE-TIME NEUTRAL COALESCENT

In this Section, to fix the background and notations, we review some well-known facts from the cited literature.

**3.1. 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  $\nu_n := (\nu_{1,n}, \dots, \nu_{n,n})$ , 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:  $\nu_{m,n} = 1$ ,  $m = 1, \dots, n$ . One iterates the reproduction over generations, while imposing the following additional assumptions:

- Exchangeability:  $(\nu_{1,n}, \dots, \nu_{n,n}) \stackrel{d}{=} (\nu_{\sigma(1),n}, \dots, \nu_{\sigma(n),n})$ , 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  $[n] := \{0, 1, \dots, n\}$  (see Karlin-McGregor, ’1964).

Famous reproduction laws are:

*Example 2.1.1* The multinomial Dirichlet model:  $\nu_n \stackrel{d}{\sim} \text{Multin-Dirichlet}(n; \theta)$ , where  $\theta > 0$  is a disorder parameter. With  $\mathbf{k}_n := (k_1, \dots, k_n)$ ,  $\nu_n$  admits the

following joint exchangeable distribution on the simplex  $|\mathbf{k}_n| := \sum_{m=1}^n k_m = n$ :

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

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

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

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

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

This distribution can be obtained by conditioning  $n$  independent mean 1 Poisson distributed random variables  $\boldsymbol{\xi}_n = (\xi_1, \dots, \xi_n)$  on summing to  $n$ :  $\boldsymbol{\nu}_n \stackrel{d}{=} (\boldsymbol{\xi}_n \mid |\boldsymbol{\xi}_n| = n)$ .

$$\mathbb{P}_\theta(\xi_1 = k) = \frac{[\theta]_k}{k!} (1+\theta)^{-k} (\theta/(1+\theta))^\theta \underset{\theta \rightarrow \infty}{\sim} \frac{e^{-1}}{k!}, \quad \text{mean 1 Poisson.}$$

When  $n$  is large, using Stirling formula,  $n! \sim \sqrt{2\pi n} n^{n+1/2} e^{-n}$ ; it follows that  $\boldsymbol{\nu}_n \xrightarrow[n \uparrow \infty]{d} \boldsymbol{\xi}_\infty$  with joint finite-dimensional law:  $\mathbb{P}(\boldsymbol{\xi}_n = \mathbf{k}_n) = \prod_{m=1}^n \frac{e^{-1}}{k_m!} = \frac{e^{-n}}{\prod_{m=1}^n k_m!}$  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,  $\boldsymbol{\nu}_n \stackrel{d}{\sim}$  random permutation of  $(2, 0, 1, \dots, 1)$ : 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.

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

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

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

$$(7) \quad \mathbb{P}(N_{r+1}(m) = k' \mid N_r(m) = k) = \mathbb{P}(\nu_{1,n} + \dots + \nu_{k,n} = k').$$

It is a martingale, with state-space  $\{0, \dots, n\}$ , initial state  $m$ , absorbing states  $\{0, n\}$  and transient states  $\{1, \dots, n-1\}$ . 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  $\Pi_n$  with entries  $\Pi_n(k, k') = \mathbb{P}(\nu_{1,n} + \dots + \nu_{k,n} = k')$  given by (7). 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 non-communicating subsets; as a result it is diagonalizable, with eigenvalues



### 3.3. 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:

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

backward-in-time.

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

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

where, with  $a = |\alpha|$  = number of equivalence classes of  $\alpha$ ,  $b = |\beta|$  = number of equivalence classes of  $\beta$ ,  $\mathbf{b}_a := (b_1, \dots, b_a)$  clusters sizes of  $\beta$  and  $(m)_a := m(m-1)\dots(m-a+1)$  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  $b_1 \geq \dots \geq b_c \geq 2, b_{c+1} = \dots = b_a = 1$ .

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  $b_1 \geq \dots \geq b_c \geq 2$  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)  $\{1, \dots, m\}$ }; 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  $\mathcal{A}_0 = \{(1), \dots, (m)\}$ , and terminal absorbing state  $\{(1, \dots, m)\}$ .

Note that if  $a = b - 1$ , obviously,

$$c_n := P_{b;b-1}^{(n)}(\mathbf{b}_{b-1}) = \frac{1}{(n)_2} \sum_{l=1}^n \mathbf{E}((\nu_{l,n})_2) \sim 2n^{-1} \mathbf{E}((\nu_{1,n})_2) \text{ as } n \rightarrow \infty$$

This is the '*coalescence*' probability that 2 randomly chosen individuals (without replacement) out of  $n$  share a common ancestor one generation before.

*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  $\{\mathbf{b}_a : b_1 + \dots + b_a = b\}$ .

In particular,

$$\mathbf{E}((\nu_{1,n})_k) = (n)_k n^{-k} \leq 1$$

leading to

$$c_n := P_{b;b-1}^{(n)}(\mathbf{b}_{b-1}) = \frac{\binom{n}{b-1}}{n^b} \sim \frac{1}{n} \text{ as } n \rightarrow \infty$$

From the Moran example 2.2.3.

$$c_n = 1/\binom{n}{2} \sim \frac{2}{n^2} \text{ as } n \rightarrow \infty$$

We shall see the role of this  $c_n$  to derive scaling limits of these processes in the large population limit  $n \rightarrow \infty$ .

*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):

$$(11) \quad \mathbb{P}(A_{r+1}(m) = a \mid A_r(m) = b) = P_{b,a}^{(n)} := \frac{b!}{a!} \sum_{\substack{b_1, \dots, b_a \in \mathbb{N}_+ \\ b_1 + \dots + b_a = b}} \frac{P_{b;a}^{(n)}(\mathbf{b}_a)}{b_1! \dots b_a!} \\ = \frac{\binom{n}{a}}{\binom{n}{b}} \sum_{\substack{b_1, \dots, b_a \in \mathbb{N}_+ \\ b_1 + \dots + b_a = b}} \mathbb{E} \left( \prod_{l=1}^a \binom{\nu_{l,n}}{b_l} \right).$$

This Markov chain has state-space  $\{0, \dots, m\}$ , 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 (11). Note, by inclusion-exclusion principle, the alternative alternating expression of  $P_{b,a}^{(n)}$ :

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

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

**Definition** [Liggett, '1985]: *Two Markov processes  $(X_t^1, X_t^2; t \geq 0)$ , 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  $\forall x_1 \in \mathcal{E}_1, \forall x_2 \in \mathcal{E}_2, \forall t \geq 0$ :*

$$(12) \quad \mathbb{E}_{x_1} \Phi(X_t^1, x_2) = \mathbb{E}_{x_2} \Phi(x_1, X_t^2).$$

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:

$$(13) \quad (i) \quad \Phi_n^1(m, k) = \binom{m}{k} / \binom{n}{k} \text{ and}$$

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

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:

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

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

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

and

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

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

$$(i) \quad [\Phi_n^1]^{-1} \Pi_n \Phi_n^1 = P'_n \text{ and } (ii) \quad [\Phi_n^2]^{-1} \Pi_n \Phi_n^2 = P'_n.$$

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  $R_n^{-1} \Pi_n R_n = \Lambda_n := \text{diag}(\lambda_0, \dots, \lambda_n)$ , 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:

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

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:

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

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 (7), (11) and (13).

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.

#### 4. BEYOND NEUTRALITY (SYMMETRY BREAKING)

*Reference:*

*Thierry Huillet, Martin Mohle. Duality and Asymptotics for a class of non-neutral discrete Moran models. J. App. Prob. 46, 866-893, 2009.*

*Thierry Huillet, Servet Martinez. Discrete evolutionary genetics. Multiplicative fitnesses and the mutation-fitness balance. Applied Mathematics, Scientific Research Publishing, 2011, Vol 2 (no 1), pp. 11-22. ff10.4236/am.2011.21002ff. fhal-00526859*

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

$$k \rightarrow np \binom{k}{n} \text{ in } \mathbb{P}(\nu_{1,n} + \dots + \nu_{k,n} = k'),$$

where:

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

$p(x)$  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  $\bar{N}_r(m) := n - N_r(m)$ ,  $r \in \mathbb{N}$  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  $x \rightarrow p(x)$  is concave (convex),  $N_r(m)$ ,  $r \in \mathbb{N}$  is a submartingale (supermartingale), because:  $\mathbb{E}(N_{r+1}(m) \mid N_r(m)) = np(N_r(m)/n) \geq N_r(m)$  (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:  $\nu_n \stackrel{d}{\sim} \text{Multin}(n; \pi_n)$ , where  $\pi_n := (\pi_{1,n}, \dots, \pi_{n,n})$  and:  $\pi_{m,n} = p \left( \frac{m}{n} \right) - p \left( \frac{m-1}{n} \right)$ ,  $m = 1, \dots, n$ . 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  $\nu_n$  no longer is exchangeable.

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

*Example 3.1* (homographic model, selection, haploidy). Assume, with  $\mathbf{x} = (x, 1 - x)'$  and, with  $\mathbf{w} = (1 + s, 1)'$ ,  $W = \mathbf{w}\mathbf{w}'$ , a (symmetric) fitness matrix

$$\mathbf{x}^+ = \frac{D_{\mathbf{x}} W \mathbf{x}}{\mathbf{x}' W \mathbf{x}} = \frac{D_{\mathbf{x}} \mathbf{w}}{\mathbf{x}' \mathbf{w}}$$

mapping the simplex to the simplex. First allele  $A$  marginal dynamics is

$$(14) \quad x^+ = p(x) = (1 + s)x / (1 + sx) = x + \frac{sx(1 - x)}{1 + sx},$$

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, diploidy). Assume

$$(15) \quad p(x) = \frac{(1 + s)x^2 + (1 + sh)x(1 - x)}{1 + sx^2 + 2shx(1 - x)}.$$

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  $1$ ): Asymmetric fitness matrix is

$$W = \begin{bmatrix} 1+s & 1+sh \\ 1+sh & 1 \end{bmatrix}$$

where  $h$  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:

$$(16) \quad p(x) = x(1+c-cx).$$

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

*Example 3.4* (mutations, affine model)

Assume  $M = [\mu_{k,l}]_{K \times K}$  a mutation matrix  $k \xrightarrow{\mu_{k,l}} l$  ( $K$  types)

$$\mathbf{x}^+ = \mathcal{M}\mathbf{x}$$

where  $\mathcal{M} = I - D_{M\mathbf{1}} + M'$ , such that  $\mathbf{1}'\mathcal{M} = \mathbf{1}'$  ( $\mathcal{M}'$  stochastic). Increment  $\mathbf{x}^+ - \mathbf{x} = (M' - D_{M\mathbf{1}})\mathbf{x}$ .

Ex:  $K = 2$ . With  $\mathbf{x} = (x, 1-x)'$ ,  $x$  frequency of allele  $A$

$$(17) \quad x^+ = p(x) = (1-\mu_2)x + \mu_1(1-x)$$

$$(18) \quad = x + \mu_1 - (\mu_1 + \mu_2)x,$$

where  $(\mu_1, \mu_2)$  are mutation probabilities ( $\mu_1 = \mu_{1,2}, \mu_2 = \mu_{2,1}$ ), satisfying  $\mu_1 \leq 1 - \mu_2$ . It corresponds to the mutation scheme:

$$\begin{array}{c} \bar{\mu}_1 \\ \circlearrowleft a \xrightarrow{\mu_1} A \xrightarrow{\bar{\mu}_2} \circlearrowright \end{array}$$

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  $\mathbb{P}(N_{r+1} > 0 \mid N_r = 0) = 1 - (1-p(0))^n > 0$  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  $\{0, 1, \dots, n\}$ .

Irreversible mutations:  $\mu_1 = 0$ : loss of allele  $A$

$$x^+ = x - \mu_2 x < x.$$

*Example 3.5* (emigration,  $\epsilon > 0$  small)

$$(19) \quad x^\dagger = p(x) = (x - \epsilon)_+ = x - x \wedge \epsilon.$$

translating the systematic removal of a fixed frequency  $\epsilon$  (if available) from the population with frequency  $x$ , as a result of emigration.

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:

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

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

## 5. NON-NEUTRAL WRIGHT-FISHER MODELS AND DUALITY

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

$$\nabla^j v(m) |_{m=n} = \sum_{l=0}^j (-1)^{j-l} \binom{j}{l} v(n-l), \quad j = 0, \dots, n.$$

Let  $u$  be some continuous function:  $[0, 1] \rightarrow [0, 1]$ . Consider the  $(n+1)$ -vector  $\mathbf{u}_n := (u(\frac{0}{n}), \dots, u(\frac{m}{n}), \dots, u(\frac{n}{n}))$ , sampling  $u$  at points  $m/n$ . The function  $u$  is said to be  $\nabla$ -completely monotonic if  $(-1)^j \nabla^j u(\frac{m}{n}) |_{m=n} \geq 0$ , for all  $j = 0, \dots, n$  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  $\mathbf{u}_n$  the point-wise product of  $\mathbf{u}_n^1$  and  $\mathbf{u}_n^2$ , 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:

**Theorem:** Consider a non-neutral WF forward model  $(N_r; r \in \mathbb{N})$  on  $\{0, \dots, n\}$ , 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:

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

There exists a Markov chain  $(A_r; r \in \mathbb{N})$  on  $\{0, \dots, n\}$  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:  $x \rightarrow q(x) = 1 - p(x)$  is completely monotone on  $(0, 1)$ . In this case, the transition probability matrix of  $(A_r; r \in \mathbb{N})$  is:

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

$P_n$  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  $[\Phi_n^2]^{-1} \Pi_n \Phi_n^2 = P'_n$ , we easily obtain:

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

This entry is non-negative if and only if  $(-1)^j \nabla^j \left( q\left(\frac{m}{n}\right)^i \right) \Big|_{m=n} \geq 0$ , for all  $i, j = 0, \dots, n$ . But, due to the above argument on  $\nabla$ -complete monotonicity of integral powers, this will be the case if and only if  $(-1)^j \nabla^j \left( q\left(\frac{m}{n}\right) \right) \Big|_{m=n} \geq 0$ , for all  $j = 0, \dots, n$ . 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  $x \rightarrow q(x) := 1 - p(x)$  is a completely monotone function on  $(0, 1)$  in the sense that:

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

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

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

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.  $\square$

## 6. 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.

**6.1. 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

$$P_n(i, j) = \binom{n}{j} \sum_{l=0}^j (-1)^{j-l} \binom{j}{l} \left(\frac{l}{n}\right)^i = (n)_j \cdot n^{-i} \cdot S_{i,j}, \quad j \leq i$$

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

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

$$(20) \quad P_n(i, j) = \binom{n}{j} \sum_{l=0}^j (-1)^{j-l} \binom{j}{l} \left(\mu_2 + \kappa \frac{l}{n}\right)^i =: (n)_j \cdot n^{-i} \cdot S_{i,j}^{\mu_2}(\kappa/n), \quad j \leq i$$

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

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 (16). 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 (14) 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 (15) with parameters  $(s, h)$  satisfying  $s > -1$  and  $sh > -1$  is CM if and only if  $s \geq \frac{1-2h}{h^2} > 0$  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  $\bar{p}(x) := 1 - p(1 - x)$  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, (14), with  $s \in (-1, 0)$  (not admissible),  $\bar{p}(x) = (1 + \bar{s})x / (1 + \bar{s}x)$  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 (15), 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  $(s \leq \frac{1-2h}{h^2} < 0, h \in (1/2, 1))$ ,  $p(x)$  is not admissible whereas  $\bar{p}(x)$  is because  $\bar{s} \geq \frac{1-2\bar{h}}{h} > 0$  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 5.1.1,  $\bar{p}(x) := 1 - p(1 - x) = x^\gamma$  is.

**6.2. 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

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

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

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

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

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.

**6.3. 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  $[0, 1]$

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

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  $\nu_n^j \stackrel{d}{\sim} \text{Multin}(n; \pi_n^j)$ , where  $\pi_{m,n}^j = p^j\left(\frac{m}{n}\right) - p^j\left(\frac{m-1}{n}\right)$ ,  $m = 1, \dots, n$ . Then,  $\nu_n \stackrel{d}{\sim} \text{Multin}(n; \pi_n)$ ,  $\pi_{m,n} = p\left(\frac{m}{n}\right) - p\left(\frac{m-1}{n}\right)$ ,  $m = 1, \dots, n$ , where  $\pi_n := \pi_n^1 \circ \pi_n^2$  is easily obtained component-wise by:

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

We let:  $\nu_n := \nu_n^1 \circ \nu_n^2 \stackrel{d}{\sim} \text{Multin}(n; \pi_n^1 \circ \pi_n^2)$ . It is the reproduction law of a WF model obtained jointly from the two bias  $p^1(x)$  and  $p^2(x)$ .

Let  $\phi(x) : (0, 1) \rightarrow (0, 1)$  be an absolutely monotone function satisfying:  $\phi^{(l)}(x) \geq 0$  for all  $l$ -th derivatives  $\phi^{(l)}$  of  $\phi$ , 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 5.3.3* The general mechanism with mutational effects is in this class. Indeed,

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

and so  $\phi(x) = 1 - p_M(1 - x) = 1 - (1 - \mu_2)(1 - x) - \mu_1x = \mu_2 + \kappa x$  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  $(e^{\theta x} - 1)/(e^\theta - 1)$ , 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 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 (14) 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  $\bar{p}_\phi(x) = q_\phi(1 - x) = e^{-\theta(1-x)^\gamma}$  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  $p_\phi$ , the signification of which in Population Genetics remaining though to be pinpointed, in each specific case study.

## 7. LIMIT LAWS

Consider a WF model  $(N_r; r \in \mathbb{N})$  on  $\{0, \dots, n\}$  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  $\{0, \dots, n\}$  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  $\boldsymbol{\pi}_\infty = (\pi_\infty(0), \dots, \pi_\infty(n))'$ . The line vector  $\boldsymbol{\pi}'_\infty$  is the left eigenvector of  $\Pi_n$  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,  $A_r \rightarrow 0$  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,  $\boldsymbol{\rho}_\infty = (\rho_\infty(0), \dots, \rho_\infty(n))'$  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  $\Pi_n$  has an algebraically simple dominant eigenvalue 1. By Perron-Frobenius theorem:

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

where  $\mathbf{1}' = (1, \dots, 1)$ . 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  $\pi_k \Pi_n(k, k') = \pi_{k'} \Pi_n(k', k)$ ) 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  $\{0\}$ .  $\square$

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  $N_\infty$  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  $(\rho_\infty(n - k); k = 0, \dots, n)$ .

Let  $\boldsymbol{\rho}_\infty = (\rho_\infty(0), \dots, \rho_\infty(n))'$ . 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:

$$\Pi_n \Phi_n = \Phi_n P'_n.$$

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

$$\begin{aligned} \Phi_n(m, k) &= \binom{n - m}{k} / \binom{n}{k} = \binom{n - k}{m} / \binom{n}{m} \\ \Phi_n^{-1}(i, j) &= (-1)^{i+j-n} \binom{i}{n - j} \binom{n}{i} = (-1)^{i+j-n} \binom{j}{n - i} \binom{n}{j}. \end{aligned}$$

Thus,

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

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,  $N_r \xrightarrow{d} N_\infty$  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  $P_n$  is sub-stochastic (if  $\mu_1 > 0$ ) and lower-left triangular. Due to triangularity, the right eigenvector  $\rho_\infty$  of  $P_n$  can easily be computed explicitly in terms of  $(P_n(i, j); j \leq i), i = 0, \dots, n$ . 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 (20) we also have:

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

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

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

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

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

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

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

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

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

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  $A_\infty$  is ( $u \in [0, 1]$ ):

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

Let  $\pi_\infty(i) = \mathbb{P}(A_\infty = i)$ , with  $\pi'_\infty = \pi'_\infty P_n$ . 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:

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

*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 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  $A_\infty$  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  $p(x) \underset{x \uparrow 0}{\sim} \lambda x$ ,  $\lambda > 1$ . Then, due to large sample asymptotic independence:

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

where  $\xi_\infty$  is an iid sequence with  $\xi_1 \stackrel{d}{\sim} \text{Poisson}(\lambda)$  (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  $\lim_{n \uparrow \infty} \rho_\infty(m) = \rho^m$ ,  $m = 1, 2, \dots$ , where  $0 < \rho < 1$  is the smallest solution to the fixed point equation

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

$\rho$  is the singleton extinction probability of a super-critical Galton-Watson process with offspring distribution  $\text{Poisson}(\lambda)$ . More precisely, proceeding as in Möhle '1994, 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  $\rho_\infty(m)$  to  $\rho^m$  is of order  $n^{-1}$ . As a result, we get the asymptotic normality:

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

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

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

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 Lipschitz-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.

## 8. 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).

## IEGMUND DUALITY WITH APPLICATIONS TO THE NEUTRAL MORAN MODEL CONDITIONED ON NEVER BEING ABSORBED

### 9. INTRODUCTION AND OUTLINE OF THE RESULTS

Consider a population constituted of  $N$  individuals (alleles) of two types. Suppose at all time  $n$  there are  $Z_n$  individuals of type 1 and  $N - Z_n$  individuals of type 2. Elementary population genetics aims at elucidating the fate of the type 1 allele frequency  $Z_n$ , when various driving ‘forces’ are at stake in the gene pool. One major driving source we shall stick to in this manuscript is the genetic drift whose nature is exclusively random. In this context, we shall first consider the classical *neutral* Moran model with two alleles (see [8] or [21]). This process is a well-known birth and death chain on the state-space  $\{0, \dots, N\}$ . The process  $Z_n$  either becomes extinct or it fixes; it has two natural boundaries  $\{0, N\}$  which are both absorbing. We will next introduce an ergodic version of  $Z_n$  obtained by conditioning  $Z_n$  to never hit the boundaries, making use of a Doob transform. We shall call the resulting process, say  $X_n$ , the recurrent Moran process. Its state-space is now  $\{1, \dots, N - 1\}$ . The recurrent Moran process  $X_n$  will be the main concern of this paper. Its main features are first derived in Section 2.

Then we run into the rich duality idea between Markov chains. Dual processes occur in many applications; for typical examples and use of dual processes, we refer the reader to the mathematics and physics literature on interacting particle systems (see for example [7], [32], [27], [31], [18], [4] and [22]). Other examples occur in the population genetics context (see for example [10], [11], [29], [19] and [20]) and essentially go back to Siegmund similar early concerns for stochastically monotone processes (see [30] and [5]).

In this Note, we will focus on the Siegmund duality and on its use in our genetics problem. We will show that the Siegmund dual of the recurrent Moran process  $X_n$  exists and is a substochastic birth and death chain. It loses mass at state  $\{1\}$  and is absorbed in state  $\{N - 1\}$ . We shall call this process  $\hat{X}_n$ . By adding an extra coffin state, say  $\{0\}$ , we will consider an enlarged proper Markov chain  $\hat{X}_n^+$ , now with the two absorbing states  $\{0\}$  and  $\{N - 1\}$ . While conditioning the enlarged process  $\hat{X}_n^+$  to exit in its natural absorbing state  $\{N - 1\}$ , we shall construct another proper stochastic process  $\tilde{X}_n$ , but now with this unique absorbing state. This new process  $\tilde{X}_n$  turns out to be intertwined with the original recurrent Moran process  $X_n$ . It is again a birth and death chain.

The intertwining construction shows that the original positive recurrent birth and death chain  $X_n$  may also be viewed as the output (through a stochastic link) of a dual hidden Markov chain  $\tilde{X}_n$  which is absorbed in a single state. Its peculiarity is that  $X_n$  is a Markov output of  $\tilde{X}_n$  which is itself Markov. As a result of this intertwining construction, the time needed for  $\tilde{X}_n$  to first hit its absorbing state is related to the time to reach stationarity for the recurrent Moran process  $X_n$  under study. This is when both intertwined processes  $X_n$  and  $\tilde{X}_n$  start in the same state  $\{1\}$ . Because we deal here with a birth and death chain  $X_n$  with known spectral characteristics, the estimation of the hitting time for  $\tilde{X}_n$  is somehow made easier

(see [24] and [16]). These constructions are described in Section 3; they are in the spirit of [1] and [7] and actually serve as an additional illustration of the theory.

Using spectral information on the intertwined chain  $\tilde{X}_n$ , we will show in Section 4 that there is no abrupt relaxation to equilibrium for  $X_n$  (no cutoff phenomenon<sup>1</sup> in the sense of Diaconis, [6]). More precisely, we will show that the expected value of the time  $\tilde{\tau}_{1,N-1}$  for  $\tilde{X}_n$  to first hit its absorbing state  $\{N-1\}$ , starting from state  $\{1\}$ , grows like  $N^2$ , just like its standard deviation. We will identify the limit law of  $\tilde{\tau}_{1,N-1}$  (normalized by its mean value) as a quadratically weighted sum of exponentially distributed random variables. The random time  $\tilde{\tau}_{1,N-1}$  being a strong stationarity time, the ergodic chain  $X_n$  enters gradually into its invariant measure after about  $N^2$  steps. By the end of this Section 4, we add some considerations on the scaling limit attached to the recurrent Moran model  $X_n$ . These follow from the observation that the shifted recurrent Moran model is a Moran model with mutations.

In Section 5, other transition times of the chain  $X_n$  are briefly investigated. It is shown that the first return time to state  $\{1\}$  grows like  $N^2$  whereas the expected time for  $X_n$  to move from one end  $\{1\}$  to the other end  $\{N-1\}$  of the state-space grows like  $N^3$ .

**Remark<sup>1</sup>:** Similar considerations for the *nonneutral* ergodic Moran model including mutations (described in [23]) can be found in [12] and [11]. It includes the Ehrenfest urn model and the Bernoulli-Laplace urn model (see [15], [9]), both being simple heat exchange models of statistical physics. In this context, the conclusions are radically different: the Moran model with mutations exhibits a cutoff phenomenon. This ergodic chain (of a different nature from the  $X_n$  under study here) enters abruptly into its invariant measure after about  $\frac{N}{2\mu} \log N$  steps, where  $\mu$  is the total mutation pressure.

## 10. THE NEUTRAL MORAN MODEL CONDITIONED ON NON-ABSORPTION

Consider a population constituted of  $N$  individuals (alleles) of two types. Suppose at time  $n$  there are  $Z_n = x$  individuals of type 1 and  $N-x$  individuals of type 2. We are first interested in the random Moran neutral evolution of type 1 individuals, namely of  $Z_n$ ,  $n \geq 0$ . It can be described as follows: Pick at random 2 distinct individuals from the whole 2-types population at time  $n$ . Assume that one is bound to die, the other one is bound to survive in the next generation and produce an additional offspring so that, at step  $n+1$ , the total population size remains constant. From this model, given  $Z_n = x$ , the updated number of type 1 individuals is

$$(21) \quad Z_{n+1} = x + 1 \text{ with probability } \alpha_x = \left(1 - \frac{x}{N}\right) \frac{x}{N}$$

$$(22) \quad Z_{n+1} = x - 1 \text{ with probability } \beta_x = \frac{x}{N} \left(1 - \frac{x}{N}\right)$$

$$(23) \quad Z_{n+1} = x \text{ with probability } \rho_x = \left(\frac{x}{N}\right)^2 + \left(1 - \frac{x}{N}\right)^2$$

The neutral Moran process  $Z_n$  is thus a transient birth and death Markov chain. It is a cornerstone model of population genetics (the so-called random genetic drift model). See [8], p. 104, [21] or [9] for example. Its stochastic transition matrix reads (the empty entries are 0).

$$\Pi = \begin{bmatrix} 1 & 0 & & & & \\ \beta_1 & \rho_1 & \alpha_1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & \beta_{N-1} & \rho_{N-1} & \alpha_{N-1} & \\ & & & 0 & & 1 \end{bmatrix}.$$

The eigenvalues of  $\Pi$  are easily seen to be  $\theta_k = 1 - \frac{k(k-1)}{N^2}$ ,  $k = 0, \dots, N$  (see [8], p. 106). Note that  $\theta_0 = \theta_1 = 1$ , translating the fact that the two states  $\{0, N\}$  are absorbing for  $Z_n$ . We shall let  $\tau_{x,0}$ ,  $\tau_{x,N}$  be the first random times at which  $Z_n$  hits  $\{0\}$  and  $\{N\}$  respectively, starting from  $Z_0 = x$ . We shall also let  $\tau_x = \tau_{x,0} \wedge \tau_{x,N}$ , with  $\tau_x < \infty$  almost surely. Let

$$\bar{\Pi} = \begin{bmatrix} \rho_1 & \alpha_1 & & & \\ \beta_2 & \rho_2 & \alpha_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \beta_{N-2} & \rho_{N-2} & \alpha_{N-2} \\ & & & \beta_{N-1} & \rho_{N-1} \end{bmatrix}$$

be obtained from  $\Pi$  after removing its first and last rows and columns. This matrix of size  $N - 1$  is substochastic in the sense that it loses mass at states  $\{1, N - 1\}$ .

Let  $\bar{\pi}'$  be the transpose row-vector of some column-vector  $\bar{\pi}$ . Let  $\bar{\pi}'$  and  $\bar{\psi}$  be defined as the left and right (row and column) eigenvectors associated to the spectral radius (largest eigenvalue)  $\theta_2$  of  $\bar{\Pi}$ , namely:

$$(24) \quad \bar{\pi}' \bar{\Pi} = \theta_2 \bar{\pi}' \quad \text{and} \quad \bar{\Pi} \bar{\psi} = \theta_2 \bar{\psi}.$$

We have  $\theta_2 = 1 - 2/N^2$  and  $\bar{\psi} = (\bar{\psi}_x, x = 1, \dots, N - 1)$ , where  $\bar{\psi}_x = \frac{x}{N} (1 - \frac{x}{N})$ . Note also that  $\bar{\pi} = (\bar{\pi}_x, x = 1, \dots, N - 1)$ , where  $\bar{\pi}_x = \frac{1}{N-1}$ . Regardless of the starting point  $x$ , we therefore have

$$\mathbb{P}_x(Z_n = z \mid \tau_x > n) \rightarrow \bar{\pi}_z = \frac{1}{N-1} \quad \text{as } n \rightarrow \infty,$$

and  $\bar{\pi}$  has an interpretation in terms of the uniform Yaglom limiting measure of  $Z_n$  on  $\{1, \dots, N - 1\}$ .

Let  $P$  be the birth and death stochastic matrix defined from  $\bar{\Pi}$  by applying the Doob transform

$$(25) \quad P = \theta_2^{-1} D_{\bar{\psi}}^{-1} \bar{\Pi} D_{\bar{\psi}},$$

where  $D_{\bar{\psi}} \equiv \text{diag}(\bar{\psi}_1, \dots, \bar{\psi}_{N-1})$ . We have

$$P = \begin{bmatrix} r_1 & p_1 & & & \\ q_2 & r_2 & p_2 & & \\ & \ddots & \ddots & \ddots & \\ & & q_{N-2} & r_{N-2} & p_{N-2} \\ & & & q_{N-1} & r_{N-1} \end{bmatrix}$$

where

$$(26) \quad p_x = \theta_2^{-1} \alpha_x \frac{\bar{\psi}_{x+1}}{\bar{\psi}_x} = \theta_2^{-1} \frac{x+1}{N} \left(1 - \frac{x+1}{N}\right)$$

$$(27) \quad q_x = \theta_2^{-1} \beta_x \frac{\bar{\psi}_{x+1}}{\bar{\psi}_x} = \theta_2^{-1} \frac{x-1}{N} \left(1 - \frac{x-1}{N}\right)$$

$$(28) \quad r_x = \theta_2^{-1} \rho_x = \theta_2^{-1} \left[ \left(\frac{x}{N}\right)^2 + \left(1 - \frac{x}{N}\right)^2 \right]$$

The invariant measure  $\boldsymbol{\pi} = (\pi_x, x = 1, \dots, N-1)$  of the corresponding random walk (say  $X_n$ ) governed by  $P$  is given by  $\pi_x = \pi_1 \prod_{y=1}^{x-1} \frac{p_y}{q_{y+1}}$ .  $\pi_x$  can also be expressed in terms of the normalized Hadamard product of  $\bar{\boldsymbol{\pi}}$  and  $\bar{\boldsymbol{\psi}}$ :

$$(29) \quad \pi_x = \frac{\bar{\pi}_x \bar{\psi}_x}{\sum_{x=1}^{N-1} \bar{\pi}_x \bar{\psi}_x} = \frac{6x(N-x)}{N(N-1)(N+1)}.$$

This invariant probability measure is a special incarnation of the bivariate generalized hypergeometric distribution (a discrete version of the beta(2, 2) distribution supported by the unit interval). Note that the recurrent Moran model governed by  $P$  is reversible as a result of  $P$  obeying detailed balance which is true for all birth and death chains.

Clearly, the eigenvalues of  $P$  are

$$t_k = \theta_2^{-1} \left(1 - \frac{k(k+1)}{N^2}\right), \quad k = 1, \dots, N-1$$

which may be recast as

$$(30) \quad t_k = 1 - \frac{(k-1)(k+2)}{N^2-2}, \quad k = 1, \dots, N-1.$$

Note that  $t_1 = 1$ ,  $t_2 = 1 - \frac{4}{N^2-2} < 1$  and the spectral gap of  $P$  is  $1 - t_2 = \frac{4}{N^2-2}$ . We have  $t_{N-1} = \frac{N}{N^2-2} > 0$  so that all eigenvalues of  $P$  are  $> 0$ .

The random walk  $X_n$ , governed by  $P$ , corresponds to the random neutral Moran process  $Z_n$  conditioned to never become extinct nor fixed in the remote future<sup>2</sup>. See [18] for a continuous-time version of this construction. It is again a birth and death Markov chain but with the modified transition probabilities (26, 28). Because it is a positive recurrent ergodic Markov chain, we shall call the process  $X_n$  governed by  $P$  the recurrent neutral Moran process (or shortly the *recurrent* Moran model).

Note that if  $X_0 \stackrel{d}{\sim} \boldsymbol{\pi}$  (meaning that the distribution of  $X_0$  is  $\boldsymbol{\pi}$ ), then  $X_n \stackrel{d}{\sim} \boldsymbol{\pi}$  for all  $n \geq 1$ : the distribution of  $X_n$  remains unchanged at all times.

We can check from (26, 27) that for all  $N \geq 3$  and for all  $x \in \{1, \dots, N-2\}$

$$(31) \quad p_x + q_{x+1} \leq 1$$

which means that the random walk  $X_n$  is stochastically monotone in the sense that for all  $n \geq 0$  and  $y \geq 1$ , the function  $x \rightarrow \mathbb{P}_x(X_n \leq y)$  is a non-increasing function of  $x$ . This fact may also be viewed as a direct consequence of  $P$  having all its eigenvalues positive. This property will prove essential in the following statement that a Siegmund dual to  $X_n$  exists, [30]. See [5] for similar considerations for birth and

death chains on the half-line.

**Remarks<sup>2</sup>:**

(i) The original idea of conditioning a Markov chain to never become extinct in the remote future seems to be due to Spitzer (unpublished) in the sixties and then Lamperti-Ney, [26]; see Athreya-Ney [2], p 56 – 59. It was first introduced in the context of the celebrated Bienyamé-Galton-Watson branching process.

(ii) Non-neutral Moran model. Consider the 2-allele Moran model with bias mechanism  $p$ . Let

$$(32) \quad p(u) : u \in [0, 1] \rightarrow [0, 1], \text{ with } 0 \leq p(0) \text{ and } p(1) \leq 1$$

be continuous and  $\bar{p}(u) := 1 - p(u)$ .

The Moran model  $X_n$  is characterized by the following transition probabilities, given the walker is in state  $x \in \{0, \dots, N\}$  :

$$(33) \quad q_x = \frac{x}{N} \bar{p}\left(\frac{x}{N}\right), \quad r_x = \frac{x}{N} p\left(\frac{x}{N}\right) + \left(1 - \frac{x}{N}\right) \bar{p}\left(\frac{x}{N}\right), \quad p_x = \left(1 - \frac{x}{N}\right) p\left(\frac{x}{N}\right).$$

For a Moran model,  $X_n$  is the amount of, say allele or type 1 individuals, in a population of size  $N$  and the reproduction law is obtained while choosing two individuals at random among  $N$ , one dying and the other one passing to the next generation while giving birth to an additional individual so as to keep constant the full population size  $N$  over the generations. Given then  $X_n = x$ , the transition  $X_n \rightarrow X_n + 1$  ( $X_n \rightarrow X_n - 1$ ) occurs with probability  $p_x$  ( $q_x$ ) translating the fact that the individual bound to die is chosen among type 2 (type 1) individuals with probability  $1 - \frac{x}{N}$  ( $\frac{x}{N}$ ) and the one which splits into two new born is chosen among the other type, with probability  $p\left(\frac{x}{N}\right)$  (respectively  $\bar{p}\left(\frac{x}{N}\right)$ ), after deforming the neutral frequency  $\frac{x}{N}$  by the bias mechanism  $p$ .

## 11. THE SIEGMUND DUAL TO THE RECURRENT MORAN MODEL

In this Section, we shall illustrate the power of the duality/intertwining relationship by considering the simplest Siegmund dual of a birth and death chain, with the recurrent neutral Moran birth and death example in mind. This will prove useful to address the question of computing the strong stationary time distribution that helps quantifying the ‘distance’ to equilibrium of the original positive recurrent birth and death process.

**Definition 1.** [27]: *Two discrete-time Markov processes  $(X_n, \hat{X}_n; n \geq 0)$ , with state-spaces  $(\mathcal{X}, \mathcal{Y})$ , possibly with substochastic transition kernels, are said to be dual with respect to some non-singular duality kernel  $H \geq 0$  on the product space  $\mathcal{X} \times \mathcal{Y}$  if  $\forall x \in \mathcal{X}, \forall y \in \mathcal{Y}, \forall n \in \mathbb{N}$  :*

$$(34) \quad \mathbb{E}_x H(X_n, y) = \mathbb{E}_y H(x, \hat{X}_n).$$

When the state-spaces  $(\mathcal{X}, \mathcal{Y}) = \{1, \dots, N - 1\}^2$  are finite and identical, the duality kernel is a square-matrix and the transition matrix of the dual process  $\hat{X}_n$ , say  $\hat{P}$ , is obtained from the one  $P$  of the original process  $X_n$  by:

$$(35) \quad \hat{P}' = H^{-1} P H,$$





in terms of the conditional probability

$$(41) \quad \Lambda(\tilde{x}, x) = \mathbb{P}\left(X_n = x \mid \tilde{X}_n = \tilde{x}\right), \text{ for all } n \geq 1.$$

(ii) The link matrix  $\Lambda$  satisfies

$$(42) \quad \Lambda(N-1, x) = \pi_x, \quad x = 0, \dots, N.$$

(iii)  $\tilde{\pi}'_0 = \pi'_0 = \mathbf{e}'_0 \equiv (1, 0, \dots, 0)$  are admissible initial distributions for the chains  $\tilde{X}_n$  and  $X_n$ , satisfying

$$(43) \quad \pi'_0 = \tilde{\pi}'_0 \Lambda.$$

**Proof:** (i) Using (39) and (35)

$$\tilde{P} = D_{\pi^c}^{-1} \hat{P} D_{\pi^c} = D_{\pi^c}^{-1} H' P H'^{-1} D_{\pi^c}.$$

The random walk  $X_n$  being reversible (detailed balance holds),  $P' = D_{\pi} P D_{\pi}^{-1}$ . Thus:

$$\tilde{P} = (D_{\pi^c}^{-1} H' D_{\pi}) P (D_{\pi}^{-1} H'^{-1} D_{\pi^c}) = \Lambda P \Lambda^{-1}$$

where  $\Lambda = D_{\pi^c}^{-1} H' D_{\pi}$ . The entries of  $\Lambda$  are  $\Lambda(\tilde{x}, x) = \frac{\pi_x}{\pi_{\tilde{x}}} \mathbf{1}(x \leq \tilde{x})$ , satisfying  $\sum_{x \leq \tilde{x}} \frac{\pi_x}{\pi_{\tilde{x}}} = 1$ . Thus  $\Lambda$  is stochastic as its row sums all sum to 1.

Suppose  $X_0 \stackrel{d}{\sim} \pi_0$ ,  $\tilde{X}_0 \stackrel{d}{\sim} \tilde{\pi}_0$  and assume the initial distributions  $\pi_0$  and  $\tilde{\pi}_0$  are  $\Lambda$ -compatible, that is  $\pi'_0 = \tilde{\pi}'_0 \Lambda$ . Using (40),  $\tilde{P}^n \Lambda = \Lambda P^n$  for all  $n \geq 1$ . This means  $\pi'_n = \tilde{\pi}'_n \Lambda$ , for all  $n \geq 1$ , where  $\pi_n(\cdot) = \mathbb{P}_{\pi_0}(X_n = \cdot)$ ,  $\tilde{\pi}_n(\cdot) = \mathbb{P}_{\tilde{\pi}_0}(\tilde{X}_n = \cdot)$  are the laws at time  $n$  of  $X_n$  and  $\tilde{X}_n$ . Thus  $\Lambda(\tilde{x}, x) = \mathbb{P}(X_n = x \mid \tilde{X}_n = \tilde{x})$ , for all  $n \geq 1$ .

(ii) The last row of  $\Lambda$  is given by  $\Lambda(N-1, x) = \pi_x$  so that once  $\tilde{X}_n$  hits state  $\{N-1\}$ , the law of  $X_n$  is  $\pi$ .

(iii) The first row of  $\Lambda$  is  $(1, 0, \dots, 0)$  so that  $\mathbf{e}'_0 = \mathbf{e}'_0 \Lambda$ . Thus both chains  $X_n$  and  $\tilde{X}_n$  may consistently start in state  $\{1\}$ .  $\triangle$

So  $\tilde{P}$  (as the algebraic composition of the Siegmund dual of  $P$  with a Doob-transform) can be obtained from  $P$  through a stochastic link  $\Lambda$ .

**Intertwining and strong stationary time.** The intertwining construction shows that the original positive recurrent birth and death chain  $X_n$  with transition matrix  $P$  may also be viewed as the output (through the stochastic link  $\Lambda$ ) of a dual hidden Markov chain  $\tilde{X}_n$  governed by  $\tilde{P} = \Lambda P \Lambda^{-1}$  and absorbed in the single state  $\{N-1\}$ . This is a setup reminiscent of filtering theory with  $\tilde{X}_n$  the hidden process and  $X_n$  the observable. The peculiarity of the intertwining construction is that  $X_n$  is a Markov output which is itself Markov. We emphasize that there is so far no clear ‘genetical’ interpretation of the process governed by  $\tilde{P}$ . Because the starting point process governed by  $P$  may be viewed as a Markovian output of the  $\tilde{P}$ -process, this suggests that the latter process should have a very basic (still missing) meaning.

What we only get is that once  $\tilde{X}_n$  hits its absorbing state  $\{N-1\}$ , the random walk  $X_n$  is distributed like  $\pi$ , provided both  $X_n$  and  $\tilde{X}_n$  were both started in  $\{1\}$ . This

suggests that, given  $X_0 = \tilde{X}_0 = 1$ , the time needed for  $X_n$  to enter its invariant probability measure should be related to the entrance time of the hidden process  $\tilde{X}_n$  into its absorbing state  $\{N - 1\}$ .

## 12. TIME TO REACH STATIONARITY FOR THE RECURRENT MORAN MODEL

We shall let

$$(44) \quad \tilde{\tau}_{1,N-1} = \inf \left( n : \tilde{X}_n = N - 1 \mid \tilde{X}_0 = 1 \right)$$

be the first hitting time of  $\{N - 1\}$  of  $\tilde{X}_n$ , starting from the state  $\{1\}$ . The random time  $\tilde{\tau}_{1,N-1}$  gives some information on the speed of convergence of the law of the original process  $X_n$  to its invariant measure. It is a strong stationary time in the sense of Diaconis and Fill, [7]. The facts (40, 41, 42, 43) indeed guarantee that  $\tilde{\tau}_{1,N-1}$  is a strong stationary time of  $X_n$  in the sense that  $X_{\tilde{\tau}_{1,N-1}} \stackrel{d}{\sim} \pi$  and is independent of  $\tilde{\tau}_{1,N-1}$  (see [7] Theorems 2.4 and 2.17 or [16] Theorem 2.1). Equivalently (see [1], Prop. 3.2), it holds that:

$$(45) \quad \text{sep}(\pi_{n,1}, \pi) \leq \mathbb{P}(\tilde{\tau}_{1,N-1} > n) \leq \mathbb{E}(\tilde{\tau}_{1,N-1})/n$$

where  $\pi_{n,1}(\cdot) = \mathbb{P}_1(X_n = \cdot)$  is the law of  $X_n$  started in  $\{1\}$ ,  $\pi$  its invariant measure. In (45), the separation discrepancy is defined by:

$$\text{sep}(\pi_{n,1}, \pi) \equiv \sup_y [1 - \pi_n(y) / \pi_y].$$

It satisfies  $\text{sep}(\pi_{n,1}, \pi) \geq \|\pi_{n,1} - \pi\|_{TV}$  where  $\|\pi_{n,1} - \pi\|_{TV} = \frac{1}{2} \sum_y |\pi_{n,1}(y) - \pi_y|$  is the total variation distance between  $\pi_{n,1}$  and  $\pi$ .

Furthermore, from (40, 43), there is a unique ‘witness’ state  $d$  for  $X_n$ , here with  $d = N - 1$ , such that either

$$\mathbb{P}_1(\tilde{X}_n = N - 1) \equiv \tilde{\pi}_{n,1}(N - 1) = 0 \Rightarrow \pi_{n,1}(d) = 0$$

or

$$\tilde{\pi}_{n,1}(N - 1) > 0 \Rightarrow \pi_{n,1}(d) = \tilde{\pi}_{n,1}(N - 1) \pi_d > 0,$$

showing that this random time is stochastically the smallest since the first inequality in (45) turns out to be an equality (see Remark 2.39 of [7] and Proposition 13 of [22]).

In our context of the birth and death chain  $\tilde{X}_n$  absorbed in  $\{N - 1\}$ , the probability generating function of  $\tilde{\tau}_{1,N-1} \geq N - 2$  is explicitly given by, [24], [16]:

$$(46) \quad \mathbb{E}(z^{\tilde{\tau}_{1,N-1}}) = \prod_{k=2}^{N-1} \frac{(1 - t_k)z}{1 - t_k z}, \quad z \in [0, 1],$$

where  $0 < t_k < 1$ ,  $k = 2, \dots, N - 1$  are the  $N - 2$  distinct eigenvalues of both  $\tilde{P}$  and  $P$ , avoiding  $t_1 = 1$  described in (30). Because the eigenvalues  $t_k$  are positive, then  $\tilde{\tau}_{1,N-1} \stackrel{d}{=} \sum_{k=2}^{N-1} \tau_k$  where the  $\tau_k$ s are independent with  $\tau_k \stackrel{d}{\sim} \text{geom}(1 - t_k)$ , the geometric distribution with success parameter  $1 - t_k$  on the half-line  $\{1, 2, \dots\}$ . The formula (66) also reads

$$\mathbb{E}(z^{\tilde{\tau}_{1,N-1}}) = \sum_{l=2}^{N-1} A_l \frac{z(1 - t_l)}{1 - t_l z},$$

where

$$A_l = \prod_{k \neq l} \frac{1 - t_k}{t_l - t_k}.$$

Therefore

$$(47) \quad \mathbb{P}(\tilde{\tau}_{1,N-1} > n) = \sum_{l=2}^{N-1} \prod_{k \neq l} \frac{1 - t_k}{t_l - t_k} t_l^n, \quad n \geq N - 1.$$

Thus,  $t_2^{-n} \mathbb{P}(\tilde{\tau}_{1,N-1} > n) \rightarrow_{n \uparrow \infty} \prod_{k=3}^{N-1} \frac{1-t_k}{t_2-t_k}$  and  $\tilde{\tau}_{1,N-1}$  has geometric tails with exponent  $t_2$ . We also have:

$$(48) \quad \mu_N \equiv \mathbb{E}(\tilde{\tau}_{1,N-1}) = \sum_{k=2}^{N-1} (1 - t_k)^{-1} \quad \text{and}$$

$$(49) \quad \sigma_N^2 \equiv \sigma^2(\tilde{\tau}_{0,N}) = \sum_{k=2}^{N-1} t_k (1 - t_k)^{-2} = \sum_{k=2}^{N-1} (1 - t_k)^{-2} - \sum_{k=2}^{N-1} (1 - t_k)^{-1}.$$

Note that since  $t_2$  is the dominant eigenvalue

$$(50) \quad \sigma^2(\tilde{\tau}_{1,N-1}) \leq \frac{\mathbb{E}(\tilde{\tau}_{1,N-1})}{1 - t_2}.$$

We can summarize these results (in the spirit of [1], [7] and [16]) as follows:

**Proposition 2.** *The Siegmund dual for the recurrent Moran birth and death chain  $X_n$  exists. There exists a Markov chain  $\tilde{X}_n$ , intertwined with  $X_n$ , with  $\{N - 1\}$  as an absorbing state and fully described in Proposition 1. The random time  $\tilde{\tau}_{1,N-1}$  is a fastest strong stationary time for  $X_n$  whose law is characterized either by (66) or (47) involving the spectrum (given in (30)), of either  $P$  or  $\tilde{P}$ , the transition matrices governing the two processes.*

**Conditions for a cutoff phenomenon.** Because the eigenvalues  $t_k$  of  $P$  are known explicitly from (30), it is possible to compute  $\mu_N = \mathbb{E}(\tilde{\tau}_{1,N-1})$  and  $\sigma_N^2 = \sigma^2(\tilde{\tau}_{1,N-1})$  and decide whether or not

$$(51) \quad \mathbb{E}(\tilde{\tau}_{1,N-1}) \rightarrow \infty \quad \text{and} \quad \sigma^2 \left( \frac{\tilde{\tau}_{1,N-1}}{\mathbb{E}(\tilde{\tau}_{1,N-1})} \right) \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

meaning  $\sigma_N/\mu_N \rightarrow 0$  as  $N \rightarrow \infty$ .

Would these conditions both hold, then  $\frac{\tilde{\tau}_{1,N-1}}{\mathbb{E}(\tilde{\tau}_{1,N-1})} \rightarrow 1$  in probability and  $\lfloor \mu_N/2 \rfloor$  would be expected to be a cutoff time for  $X_n$  started in  $\{1\}$  in the sense ([6]) that: With  $n_N(x) = \lfloor (\mu_N + x\sigma_N) \rfloor$ , then

$$\|\pi_{n_N(x),1} - \pi\|_{TV} \xrightarrow{N \uparrow \infty} c(x)$$

where  $c(x) \rightarrow_{x \uparrow \infty} 0$  and  $c(x) \rightarrow_{x \uparrow -\infty} 1$ . The chain would then exhibit a sharp cutoff: if  $n < n_N(x)$ ,  $\|\pi_{n,1} - \pi\|$  would keep close to its maximal value 1, whereas if  $n > n_N(x)$  it would suddenly drop to a smaller value.

When there is cutoff, the expected mixing time (measured in terms of  $\text{sep}(\pi_{n,1}, \pi)$ ) is  $\mu_N$  whereas the spectral gap is  $1 - t_2$ , the product of the 2 of which tends

to  $\infty$ . Recalling  $\sigma^2(\tilde{\tau}_{1,N-1}) \leq \frac{\mu_N}{1-t_2}$ , then  $\sigma^2(\tilde{\tau}_{1,N-1}/\mu_N) = \mu_N^{-2}\sigma^2(\tilde{\tau}_{1,N-1}) \leq 1/((1-t_2)\mu_N)$  and the condition

$$(1-t_2)\mu_N \rightarrow \infty$$

is a sufficient condition for  $\sigma^2(\tilde{\tau}_{1,N-1}/\mu_N) \rightarrow 0$ . If this condition holds, the contribution of  $\sum_{k=3}^{N-1} (1-t_k)^{-1}$  to  $\mu_N$  dominates the lead term  $(1-t_2)^{-1}$  (see [8] for recent developments and precisions).

For some birth and death models, it may happen that  $\mu_N \rightarrow \infty$  and  $\sigma_N/\mu_N \rightarrow 0$  as  $N \rightarrow \infty$ . In such cases, there is no cutoff and this is precisely what happens for the recurrent Moran model under concern.

**Computing the mean and variance of  $\tilde{\tau}_{1,N-1}$  for the recurrent Moran model.** For the case of the recurrent Moran model, using (30):  $1-t_k = \frac{k(k+1)-2}{N^2-2}$  and the large  $N$  values of  $\mu_N$  and  $\sigma_N^2$  can be estimated. Using indeed the integral approximation

$$\mu_N \sim (N-2)(N^2-2) \int_0^1 \frac{dx}{(2+x(N-2))(3+x(N-2))-2}$$

(where we used the change of variables  $\frac{k-2}{N-2} = x$  and  $\mu_N \sim \alpha_N \Leftrightarrow \mu_N/\alpha_N \rightarrow 1$  as  $N \rightarrow \infty$ ), with  $c_1 \equiv \frac{2 \log 2}{3}$ , we easily get

$$\mu_N \sim N \int_0^1 \frac{dx}{(x+1/N)(x+4/N)} = c_1 N^2.$$

In this situation, the contribution of  $\sum_{k=3}^{N-1} (1-t_k)^{-1}$  to  $\mu_N$  contribute equally likely with the lead term  $(1-t_2)^{-1}$ . Note also that for this model,  $(1-t_2)\mu_N \sim 4c_1 \rightarrow \infty$  as  $N \rightarrow \infty$  suggesting that there is no cutoff.

Using similar techniques indeed, with  $c_2 \equiv \frac{5}{36} + \frac{4 \log 2}{27}$ , we get

$$\sigma_N^2 \sim N \int_0^1 \frac{dx}{(x+1/N)^2(x+4/N)^2} \sim c_2 N^4$$

showing that  $\sigma^2\left(\frac{\tilde{\tau}_{1,N-1}}{\mathbb{E}(\tilde{\tau}_{1,N-1})}\right) \sim \frac{c_2}{c_1^2} \rightarrow 0$  as  $N \rightarrow \infty$ .

Thus  $\frac{\tilde{\tau}_{1,N-1}}{\mathbb{E}(\tilde{\tau}_{1,N-1})} \rightarrow 1$  in probability and there is no cutoff phenomenon. In fact, in the recurrent Moran case, we have

**Proposition 3.**

$$\frac{\tilde{\tau}_{1,N-1}}{\mathbb{E}(\tilde{\tau}_{1,N-1})} \xrightarrow{d} T \text{ as } N \rightarrow \infty$$

where the limit  $T$  is random with

$$T \stackrel{d}{=} \sum_{l \geq 1} \frac{1}{l(l+3)} E_l(c_1)$$

where the  $E_l(c_1)$  are independent and identically distributed (iid) random variables, with common distribution  $\mathbb{P}(E_1(c_1) > x) = e^{-c_1 x}$ .

**Proof:** Indeed, from the large  $N$  approximation  $t_k/(1-t_k) \sim \frac{N^2}{k(k+1)-2}$  and making use of (66), for all  $t \in \mathbb{R}$ , we get

$$\begin{aligned} \mathbb{E} \left( e^{it \frac{\tilde{\tau}_{1,N-1}}{\mu_N}} \right) &= e^{it \frac{N-2}{\mu_N}} \prod_{k=2}^{N-1} \frac{(1-t_k) e^{\frac{it}{\mu_N}}}{1-t_k e^{\frac{it}{\mu_N}}} \sim e^{\frac{it}{c_1 N}} \prod_{k=2}^{N-1} \frac{1}{1 - \frac{t_k}{1-t_k} \frac{it}{c_1 N^2}} \\ &\sim \prod_{l=1}^{N-2} \frac{1}{1 - \frac{it}{c_1 l(l+3)}} \rightarrow \Phi(t), \end{aligned}$$

where  $\Phi(t) = \prod_{l=1}^{\infty} \frac{1}{1 - \frac{it/c_1}{l(l+3)}}$ .  $\Phi(t)$  is the characteristic function of the random variable

$$T = \sum_{l \geq 1} \frac{1}{l(l+3)} E_l(c_1),$$

where the  $E_l(c_1)$  are iid random variables, with common distribution  $\mathbb{P}(E_1(c_1) > x) = e^{-c_1 x}$ . The limiting  $T$  is thus a weighted sum of iid exponentially-distributed random variables with quadratic weights. Such random variables are known to have all their integral moments finite (see [3]). For instance

$$\mathbb{E}(T) = \frac{1}{c_1} \sum_{l \geq 1} \frac{1}{l(l+3)} = \frac{1}{3c_1} \left( \sum_{l \geq 1} \frac{1}{l} - \sum_{l \geq 4} \frac{1}{l} \right) = \frac{11}{18c_1}. \quad \Delta$$

We conclude from (45) that as  $N$  is large

$$\text{sep}(\boldsymbol{\pi}_{n,1}, \boldsymbol{\pi}) = \mathbb{P}(\tilde{\tau}_{1,N-1} > n) \sim \bar{F}_T \left( \frac{n}{c_1 N^2} \right)$$

where  $\bar{F}_T(x) = \mathbb{P}(T > x)$  is the complementary probability distribution of  $T > 0$ . This shows that, for the recurrent neutral Moran model, the transition to stationarity occurs gradually after about  $c_1 N^2$  steps. There is no abrupt entrance of this particular chain into the invariant measure (no cutoff phenomenon).

**A plausible scaling limit.** Let  $Y_n = X_n - 1$  be a shifted version of the recurrent model with state-space  $\{0, \dots, N-2\}$ . First, we observe that  $Y_n$  is a Moran model with mutations (see [8] p. 107). Indeed, from (26, 28), the transition probabilities of  $Y_n$  read:  $p_y = \theta_2^{-1} \frac{y+2}{N} \left(1 - \frac{y+2}{N}\right)$ ,  $q_y = \theta_2^{-1} \frac{y}{N} \left(1 - \frac{y}{N}\right)$ ,  $r_y = 1 - (p_y + q_y)$ . Recalling  $\theta_2 = 1 - \frac{2}{N^2-2}$ , these may be recast as

$$\begin{aligned} p_y &= \left(1 - \frac{y}{N-2}\right) \pi \left(\frac{y}{N-2}\right) \\ q_y &= \frac{y}{N-2} \left(1 - \pi \left(\frac{y}{N-2}\right)\right) \\ r_y &= \frac{y}{N-2} \pi \left(\frac{y}{N-2}\right) + \left(1 - \frac{y}{N-2}\right) \left(1 - \pi \left(\frac{y}{N-2}\right)\right), \end{aligned}$$

where

$$\pi(u) = \mu_1(1-u) + (1-\mu_2)u$$

is an affine mutation mapping from  $[0, 1]$  to  $[0, 1]$ , with mutation probabilities from type 2 to 1 and from type 1 to 2

$$\mu_1 = \frac{2(N-2)}{N^2-2} \sim \frac{2}{N} \quad \text{and} \quad \mu_2 = \frac{2(N-1)}{N^2-2} \sim \frac{2}{N}.$$

Let now  $\Delta > 0$  be a small constant. Let  $t = n\Delta$  be a discretization of time and consider the scaled process  $U_t = Y_{n\Delta}/(N-2) \in [0, 1]$ . Given  $U_t = u$ , assume

$$\begin{aligned} U_{t+\Delta} &= u \pm \frac{1}{N} \text{ with probability } \Delta \cdot p_{(N-2)u} \text{ (respectively } \Delta \cdot q_{(N-2)u}) \\ &= u \text{ with probability } 1 - \Delta \cdot r_{(N-2)u}. \end{aligned}$$

Then  $\mathbb{E}_u(U_{t+\Delta} - u) = \frac{\Delta}{N-2} (p_{(N-2)u} - q_{(N-2)u}) = \frac{\Delta}{N-2} (\pi(u) - u)$ , with  $\pi(u) - u \sim \frac{1}{N} (2 - 4u)$  for large  $N$ . Thus

$$\frac{1}{\Delta} \mathbb{E}_u(U_{t+\Delta} - u) \sim \frac{1}{N^2} (2 - 4u).$$

Similarly,  $\mathbb{E}_u[(U_{t+\Delta} - u)^2] = \frac{\Delta}{(N-2)^2} (1 - r_{(N-2)u})$  leading to the dominant order in  $N$  to

$$\frac{1}{\Delta} \mathbb{E}_u[(U_{t+\Delta} - u)^2] \sim \frac{2}{N^2} u(1 - u).$$

Following ([21], Chapter 2), this shows (by passing to the limit in  $\Delta$ ), that  $U_t \in [0, 1]$  obeys the Wright-Fisher-like diffusion approximation equation driven by Brownian motion  $(B_t)$

$$dU_t = \frac{2}{N^2} (1 - 2U_t) dt + \frac{1}{N} \sqrt{2U_t(1 - U_t)} dB_t.$$

Performing the time substitution  $\tau = N^{-2}t$  and defining  $V_\tau = U_{N^2\tau}$ ,  $V_\tau$  obeys the diffusion equation

$$dV_\tau = 2(1 - 2V_\tau) d\tau + \sqrt{2V_\tau(1 - V_\tau)} dB_\tau$$

with a unit time  $\tau$  corresponding to  $N^2$  generations in the discrete model. The above considerations on the time to reach stationarity of  $X_n$  suggests that  $V_\tau$  enters gradually into its invariant measure (which is  $6v(1-v)dv$ ) by time  $O(1)$ .

### 13. OTHER TRANSITION TIMES FOR THE RECURRENT MORAN MODEL

Consider the ergodic Moran birth and death Markov chain  $X_n$  on the state-space  $\{1, \dots, N-1\}$ . It is well-known that the mean return time to state  $\{1\}$  is  $\mathbb{E}(\tau_{1,1}) = \frac{1}{\pi_1}$ . Thus, using (29)

$$\mathbb{E}(\tau_{1,1}) \sim \frac{N^2}{6} \text{ as } N \rightarrow \infty.$$

We wish now to estimate the time it takes for  $X_n$  to move from one end of the state-space to the other, that is from state  $\{1\}$  to  $\{N-1\}$ . Let  $\tau_{1,N-1}$  be this random time. We shall prove the following estimation of its mean value:

**Proposition 4.**

$$\mathbb{E}(\tau_{1,N-1}) \sim \frac{N^3}{6} \text{ as } N \rightarrow \infty.$$

**Proof:** Let indeed  $\tau_{x,x+1}$  be the random time to first hit the state  $\{x+1\}$  starting from the state  $\{x\}$ . Depending on whether the move starting in  $x$  is up, down or no move,  $\tau_{x,x+1}$  is either 1 or  $1 + \tau'_{x,x+1}$  (with  $\tau'_{x,x+1}$  a statistical copy of  $\tau_{x,x+1}$ ) or  $1 + \tau_{x-1,x} + \tau'_{x,x+1}$ . If we let  $\mu_x$  be the mean value of  $\tau_{x,x+1}$ , we thus get

$\mu_x = p_x + r_x(1 + \mu_x) + q_x(1 + \mu_{x-1} + \mu_x)$ ,  $x \geq 2$ . This leads to the recurrence ( $\mu_1 = 1/p_1$ ):

$$\mu_x = \frac{q_x}{p_x} \mu_{x-1} + \frac{1}{p_x}, \quad x \geq 2.$$

This recurrence can be solved to give

$$\mu_x = \frac{1}{p_x \pi_x} \sum_{y=1}^x \pi_y.$$

Thus

$$\mathbb{E}(\tau_{1,N-1}) = \sum_{x=1}^{N-2} \mu_x = \sum_{x=1}^{N-2} \frac{1}{p_x \pi_x} \sum_{y=1}^x \pi_y.$$

Looking at this sum formula, one expects that its leading term is  $\frac{1}{p_{N-2} \pi_{N-2}} \sum_{y=1}^{N-2} \pi_y$  because this is where  $\sum_{y=1}^x \pi_y$  is the largest and  $\pi_x$  the smallest. Looking at the expressions of  $p_x$  and  $\pi_x$  at  $x = N - 2$  in (26, 29), this crude estimation would lead to  $\mathbb{E}(\tau_{1,N-1}) \sim N^3/12$  but this is not quite true because the other terms contribute equally likely.

In fact, the computations can be done explicitly. From (29), we have

$$\sum_{y=1}^x \pi_y = \frac{x(x+1)(3N - (2x+1))}{N(N-1)(N+1)}$$

and so

$$\begin{aligned} \mathbb{E}(\tau_{1,N-1}) &= \frac{\theta_2 N^2}{6} \sum_{x=1}^{N-2} \frac{3N - (2x+1)}{(N-x)(N-x-1)} \\ &= \frac{\theta_2 N^2}{6} \sum_{y=1}^{N-2} \frac{N+1+2y}{y(y+1)} = \frac{\theta_2 N^2}{6} \left( \sum_{y=1}^{N-2} \frac{N+1}{y} - \sum_{y=1}^{N-2} \frac{N-1}{y+1} \right) \\ &= \frac{\theta_2 N^2}{6} [(N+1)H_{N-2} - (N-1)(H_{N-1} - 1)] \end{aligned}$$

where  $H_N$  is the  $N$ -harmonic number. Thus

$$\begin{aligned} \mathbb{E}(\tau_{1,N-1}) &= \frac{\theta_2 N^2}{6} \left[ 2H_{N-1} - \frac{N+1}{N-1} + N - 1 \right] \\ &\sim \frac{N^3}{6} \text{ as } N \rightarrow \infty. \quad \triangle \end{aligned}$$

#### REFERENCES

- [1] Aldous, D.; Diaconis, P. Strong uniform times and finite random walks. Adv. in Appl. Math. 8, no. 1, 69–97, (1987).
- [2] Athreya, K. B.; Ney, P. E. *Branching processes*. Die Grundlehren der mathematischen Wissenschaften, Band 196. Springer-Verlag, New York-Heidelberg, 1972. xi+287 pp.
- [3] Biane, P.; Pitman, J.; Yor, M. Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions. Bull. Amer. Math. Soc. (N.S.) 38, (2001), no. 4, 435–465.
- [4] Chen, J. Extremality of invariant measures and ergodicity of stochastic systems. J. Phys. A 35 (2002), no. 2, 229–238.
- [5] Dette, H.; Fill, J. A.; Pitman, J.; Studden, W. J. Wall and Siegmund duality relations for birth and death chains with reflecting barrier. Dedicated to Murray Rosenblatt. J. Theoret. Probab. 10, (1997), no. 2, 349–374.

- [6] Diaconis, P. The cutoff phenomenon in finite Markov chains. Proc. Natl. Acad. Sci. USA 36, (1996), 1659–1664.
- [7] Diaconis, P.; Fill, J. A. Strong stationary times via a new form of duality. Ann. Probab. 18, no. 4, 1483–1522, (1990).
- [8] Diaconis, P.; Saloff-Coste, L. Separation cut-offs for birth and death chains. Ann. Appl. Probab. 16, (2006), no. 4, 2098–2122.
- [9] Diaconis, P.; Shahshahani, M. Time to reach stationarity in the Bernoulli-Laplace diffusion model. SIAM J. Math. Anal. 18 (1987), no. 1, 208–218.
- [10] Donnelly, P. *Dual processes in population genetics*. Stochastic spatial processes (Heidelberg, 1984), 94–105, Lecture Notes in Math., 1212, Springer, Berlin, 1986.
- [11] Donnelly, P.; Rodrigues, E. R. Convergence to stationarity in the Moran model. J. Appl. Probab. 37, (2000), no. 3, 705–717.
- [12] Dunham, B. Fluctuation theory for Moran’s genetics model. J. Math. Anal. Appl. 210, (1997), no. 2, 777–789.
- [13] Dynkin, E. B. *Markov processes*. Vols. I, II. Translated with the authorization and assistance of the author by J. Fabius, V. Greenberg, A. Maitra, G. Majone. Die Grundlehren der Mathematischen Wissenschaften, Bände 121, 122 Academic Press Inc., Publishers, New York; Springer-Verlag, Berlin-Göttingen-Heidelberg 1965.
- [14] Ewens, W. J. *Mathematical population genetics. I. Theoretical introduction*. Second edition. Interdisciplinary Applied Mathematics, 27. Springer-Verlag, New York, 2004.
- [15] Feller, W. *An introduction to probability theory and its applications*, Vol. 1. John Wiley and Sons, Third Edition, New York, 1968.
- [16] Fill, J. A. The passage time distribution for a birth-and-death chain: Strong stationary duality gives a first stochastic proof. Journal of Theoretical Probability, Volume 22, Number 3, 543–557, (2009).
- [17] Gillespie, J. H. *The Causes of Molecular Evolution*. New York and Oxford: Oxford University Press, 1991.
- [18] Griffeath, D. Frank Spitzer’s pioneering work on interacting particle systems. Ann. Probab. 21 (1993), no. 2, 608–621.
- [19] Huillet, T. A Duality Approach to the Genealogies of Discrete Non-Neutral Wright-Fisher Models. Journal of Probability and Statistics, vol. 2009, Article ID 714701, 22 pages, (2009).
- [20] Huillet, T.; Möhle, M. Duality and asymptotics for a class of nonneutral discrete Moran models. J. Appl. Probab. 46, (2009), no. 3, 866–893.
- [21] Huillet, T. Reversing the drift of the Ehrenfest urn model and three conditionings. J. Phys. A 42 (2009), no. 34, 345005, 8 pp.
- [22] Huillet, T.; Martínez, S. Duality and Intertwining for discrete Markov kernels: a relation and examples. arXiv hal-00401732, (2009).
- [23] Karlin, S.; McGregor, J. On a genetics model of Moran. Proc. Cambridge Philos. Soc. 58, 299–311, (1962).
- [24] Keilson, J. *Markov chain models - rarity and exponentiality*. Applied Mathematical Sciences, 28. Springer-Verlag, New York-Berlin, 1979.
- [25] Lambert, A. Population dynamics and random genealogies. Stochastic Models 24, suppl. 1, 45–163, (2008).
- [26] Lamperti, J.; Ney, P. Conditioned branching processes and their limiting diffusions. Teor. Veroyatnost. i Primenen. 13, (1968), 126–137.
- [27] Liggett, T. M. *Interacting particle systems*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 276. Springer-Verlag, New York, 1985.
- [28] Maruyama, T. *Stochastic problems in population genetics*. Lecture Notes in Biomathematics, 17. Springer-Verlag, Berlin-New York, 1977.
- [29] Möhle, M. The concept of duality and applications to Markov processes arising in neutral population genetics models. Bernoulli 5, (1999), no. 5, 761–777.
- [30] Siegmund, D. The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes. Ann. Probability 4, (1976), no. 6, 914–924.
- [31] Sudbury, A.; Lloyd, P. Quantum operators in classical probability theory. II. The concept of duality in interacting particle systems. Ann. Probab. 23, (1995), no. 4, 1816–1830.
- [32] Spitzer, F. Interaction of Markov processes. Advances in Math. 5, (1970), 246–290.

## ONE-DIMENSIONAL DIFFUSIONS: SCALING LIMITS

### 14. DIFFUSION PROCESSES ON THE UNIT INTERVAL AND DOOB TRANSFORMS

We start with generalities on one-dimensional diffusions with the WF model and its relatives in mind. For more technical details, we refer to [6], [7], [13] and [19]. We also introduce Doob transforms as particular instances of the modification of the original diffusion process through a multiplicative functional.

**14.1. One-dimensional diffusions on the interval**  $[0, 1]$ . Let  $(w_t; t \geq 0)$  be a standard one-dimensional Brownian (Wiener) motion. We consider a 1–dimensional Itô diffusion driven by  $(w_t; t \geq 0)$  on the interval say  $[0, 1]$ , see [11]. We assume it has locally Lipschitz continuous drift  $f(x)$  and local standard deviation (volatility)  $g(x)$ , namely we consider the stochastic differential equation (SDE):

$$(52) \quad dx_t = f(x_t) dt + g(x_t) dw_t, \quad x_0 = x \in I := (0, 1).$$

The condition on  $f(x)$  and  $g(x)$  guarantees in particular that there is no point  $x_*$  in  $I$  for which  $|f(x)|$  or  $|g(x)|$  would blow up and diverge as  $|x - x_*| \rightarrow 0$ .

The Kolmogorov backward infinitesimal generator of (52) is  $G = f(x) \partial_x + \frac{1}{2} g^2(x) \partial_x^2$ . As a result, for all suitable  $\psi$  in the domain of the operator  $S_t := e^{tG}$ ,  $u := u(x, t) = \mathbf{E}\psi(x_{t \wedge \tau_x})$  satisfies the Kolmogorov backward equation (KBE)

$$\partial_t u = G(u); \quad u(x, 0) = \psi(x).$$

In the definition of the mathematical expectation  $u$ , we have  $t \wedge \tau_x := \inf(t, \tau_x)$  where  $\tau_x$  indicates a random time at which the process should possibly be stopped (absorbed), given the process was started in  $x$ . The description of this (adapted) absorption time is governed by the type of boundaries which  $\partial I := \{0, 1\}$  are to  $(x_t; t \geq 0)$ . A classification of the boundaries exists, due to Feller (see [13] pp. 226): they can be either accessible (namely exit or regular), or inaccessible (namely entrance or natural).

**14.2. Natural coordinate, scale, speed measure, time change.** For such Markovian diffusions, it is interesting to consider the  $G$ –harmonic coordinate  $\varphi \in C^2$  belonging to the kernel of  $G$ , i.e. satisfying  $G(\varphi) = 0$ . For  $\varphi$  and its derivative  $\varphi' := d\varphi/dy$ , with  $(x_0, y_0) \in (0, 1)$ , one finds

$$\begin{aligned} \varphi'(y) &= \varphi'(y_0) e^{-2 \int_{y_0}^y \frac{f(z)}{g^2(z)} dz} \\ \varphi(x) &= \varphi(x_0) + \varphi'(y_0) \int_{x_0}^x e^{-2 \int_{y_0}^y \frac{f(z)}{g^2(z)} dz} dy. \end{aligned}$$

One should choose a version of  $\varphi$  satisfying  $\varphi'(y) > 0$ ,  $y \in I$ . The function  $\varphi$  kills the drift  $f$  of  $(x_t; t \geq 0)$  in the sense that, considering the change of variable  $y_t = \varphi(x_t)$ ,

$$dy_t = (\varphi' g)(\varphi^{-1}(y_t)) dw_t, \quad y_0 = \varphi(x).$$

The drift-less diffusion  $(y_t; t \geq 0)$  is often termed the diffusion in natural coordinates with state-space  $[\varphi(0), \varphi(1)] =: \varphi(I)$ . Its volatility is  $\tilde{g}(y) := (\varphi' g)(\varphi^{-1}(y))$ . The function  $\varphi$  is often called the scale function.

Whenever  $\varphi(0) > -\infty$  and  $\varphi(1) < +\infty$ , one can choose the integration constants defining  $\varphi(x)$  so that

$$\varphi(x) = \frac{\int_0^x e^{-2 \int_0^y \frac{f(z)}{g^2(z)} dz} dy}{\int_0^1 e^{-2 \int_0^y \frac{f(z)}{g^2(z)} dz} dy},$$

with  $\varphi(0) = 0$  and  $\varphi(1) = 1$ . In this case, the state-space of  $(y_t; t \geq 0)$  is again  $[0, 1]$ , the same as for  $(x_t; t \geq 0)$ .

Finally, considering the random time change  $t \rightarrow \theta_t$  with inverse:  $\theta \rightarrow t_\theta$  defined by  $\theta_{t_\theta} = \theta$  and

$$\theta = \int_0^{t_\theta} \tilde{g}^2(y_s) ds,$$

the novel diffusion  $(w_\theta := y_{t_\theta}; \theta \geq 0)$  is easily checked to be identical in law to a standard Brownian motion on  $\varphi(I)$ . The random time  $t_\theta$  can be expressed as

$$t_\theta = \int_0^\theta m(\varphi^{-1}(w_\tau)) (\varphi^{-1})'(w_\tau) d\tau$$

where  $m(x) := 1/(g^2 \varphi')(x)$  is the (positive) speed density at  $x = \varphi^{-1}(y)$ . Both the scale function  $\varphi$  and the speed measure  $d\mu = m(x) \cdot dx$  are therefore essential ingredients to reduce the original stochastic process  $(x_t; t \geq 0)$  to the standard Brownian motion  $(w_\theta; \theta \geq 0)$ . The Kolmogorov backward infinitesimal generator  $G$  may then be written in Feller form

$$G(\cdot) = \frac{1}{2} \frac{d}{d\mu} \left( \frac{d}{d\varphi} \cdot \right).$$

**Examples** (from population genetics):

- Assume  $f(x) = 0$  and  $g^2(x) = x(1-x)$ . This is the neutral WF model discussed at length later. This diffusion is already in natural scale and  $\varphi(x) = x$ ,  $m(x) = [x(1-x)]^{-1}$ . The speed measure is not integrable.

- With  $\pi_1, \pi_2 > 0$ , assume  $f(x) = \pi_1 - (\pi_1 + \pi_2)x$  and  $g^2(x) = x(1-x)$ . This is the WF model with mutation. The parameters  $\pi_1, \pi_2$  can be interpreted as mutation rates. The drift vanishes when  $x = \pi_1/\pi$  (where  $\pi := \pi_1 + \pi_2$  is the total mutation pressure) which is an attracting point for the dynamics. Here:

$\varphi'(y) = \varphi'(y_0) y^{-2\pi_1} (1-y)^{-2\pi_2}$ ,  $\varphi(x) = \varphi(x_0) + \varphi'(y_0) \int_{x_0}^x y^{-2\pi_1} (1-y)^{-2\pi_2} dy$ , with  $\varphi(0) = -\infty$  and  $\varphi(1) = +\infty$  if  $\pi_1, \pi_2 > 1/2$ . The speed measure density is  $m(x) \propto x^{2\pi_1-1} (1-x)^{2\pi_2-1}$  and so is always integrable. After normalization to 1,  $m(x)$  is the beta( $2\pi_1, 2\pi_2$ ) density.

- With  $\sigma \in \mathbf{R}$ , assume a model with quadratic logistic drift  $f(x) = \sigma x(1-x)$  and local variance  $g^2(x) = x(1-x)$ . This is the WF model with selection or selection. For this diffusion (see [16]),  $\varphi(x) = \frac{1-e^{-2\sigma x}}{1-e^{-2\sigma}}$  and  $m(x) \propto [x(1-x)]^{-1} e^{2\sigma x}$  is not integrable. Here,  $\sigma$  is a selection or fitness parameter.

**Time change and subordination.** We start from the diffusion (1) with infinitesimal generator  $G = f\partial_x + \frac{1}{2}g^2\partial_x^2$  and consider the time change problem without

passing first in natural coordinate. Let the random time change

$$t \rightarrow \theta_t = \int_0^t g^2(x_s) ds.$$

Its inverse:  $\theta \rightarrow t_\theta$  defined by  $\theta_{t_\theta} = \theta$  is given by  $\theta = \int_0^{t_\theta} g^2(x_s) ds$ .

In this new stochastic time clock, the subordinated diffusion ( $y_\theta := x_{t_\theta}; \theta \geq 0$ ) obeys the Langevin SDE with potential  $U(y) := -2 \int_0^y \frac{f(z)}{g^2(z)} dz$

$$dy_\theta = \frac{f}{g^2}(y_\theta) d\theta + dw_\theta,$$

with backward infinitesimal generator  $\tilde{G} = g^{-2}G = \frac{f}{g^2}\partial_x + \frac{1}{2}\partial_x^2$  (See [12], pp. 164-169).

We have  $\dot{\theta}_t = g^2(x_t)$  meaning that at each point  $x_t$  of the former motion, the motion of the path is accelerated or decelerated, depending on the rate  $g^2(x_t) \leq 1$ .

Note that conversely  $t_\theta = 1/g^2(y_\theta)$ . Under the time substitutions, the road maps of the paths of both  $(x_t; t \geq 0)$  and  $(y_\theta; \theta \geq 0)$  remain exactly the same. If a path of the former process is accelerated or decelerated by its squared volatility  $g^2$  (its local variance) at each locality, then this process boils down to the latter one. Stated differently, if we measure time by the amount of squared volatility accumulated within each of its path, the process  $(x_t; t \geq 0)$  becomes  $(y_\theta; \theta \geq 0)$ , both with state-space  $I$ .

**14.3. The transition probability density.** Assume that  $f(x)$  and  $g(x)$  are now differentiable in  $I$ . Let then  $p(x; t, y)$  stand for the transition probability density function of  $x_t$  at  $y$  given  $x_0 = x$ . Then  $p := p(x; t, y)$  is the smallest solution to the Kolmogorov forward (Fokker-Planck) equation (KFE):

$$(53) \quad \partial_t p = G^*(p), \quad p(x; 0, y) = \delta_y(x)$$

where  $G^*(\cdot) = -\partial_y(f(y)\cdot) + \frac{1}{2}\partial_y^2(g^2(y)\cdot)$  is the adjoint of  $G$  ( $G^*$  acts on the terminal value  $y$  whereas  $G$  acts on the initial value  $x$ ). The way one can view this partial differential equation (PDE) depends on the type of boundaries that  $\{0, 1\}$  are.

Suppose for example that the boundaries  $\circ := 0$  or  $1$  are both exit (or absorbing) boundaries. From the Feller classification of boundaries, this will be the case if  $\forall y_0 \in (0, 1)$ :

$$(54) \quad (i) \ m(y) \notin L_1(y_0, \circ) \text{ and } (ii) \ \varphi'(y) \int_{y_0}^y m(z) dz \in L_1(y_0, \circ),$$

where a function  $f(y) \in L_1(y_0, \circ)$  if  $\int_{y_0}^\circ |f(y)| dy < +\infty$ .

In this case, a sample path of  $(x_t; t \geq 0)$  can reach  $\circ$  from the inside of  $I$  in finite time but cannot reenter. The sample paths are absorbed at  $\circ$ . There is an absorption at  $\circ$  at time  $\tau_{x,\circ} = \inf(t > 0 : x_t = \circ | x_0 = x)$  and  $\mathbf{P}(\tau_{x,\circ} < \infty) = 1$ . Whenever both boundaries  $\{0, 1\}$  are absorbing, the diffusion  $x_t$  should be stopped at  $\tau_x := \tau_{x,0} \wedge \tau_{x,1}$ . Would none of the boundaries  $\{0, 1\}$  be absorbing, then  $\tau_x = +\infty$ . This occurs when the boundaries are inaccessible.

Examples of diffusion with exit boundaries are the WF model and the WF model with selection. In the WF model including mutations, the boundaries are entrance boundaries and so are not absorbing.

When the boundaries are absorbing, then  $p(x; t, y)$  is a sub-probability. Letting  $\rho_t(x) := \int_0^1 p(x; t, y) dy$ , we clearly have  $\rho_t(x) = \mathbf{P}(\tau_x > t)$ . Such models are non-conservative.

For one-dimensional diffusions, the transition density  $p(x; t, y)$  is reversible with respect to the speed density ([13], Chapter 15, Section 13) and so detailed balance holds:

$$(55) \quad m(x) p(x; t, y) = m(y) p(y; t, x), \quad 0 < x, y < 1.$$

The speed density  $m(y)$  satisfies  $G^*(m) = 0$ . It may be written as a Gibbs measure with density:  $m(y) \propto \frac{1}{g^2(y)} e^{-U(y)}$  where the potential function  $U(y)$  reads:

$$(56) \quad U(y) = -2 \int_0^y \frac{f(z)}{g^2(z)} dz, \quad 0 < y < 1$$

and with the measure  $\frac{dy}{g^2(y)}$  standing for the reference measure.

Furthermore, if  $p(s, x; t, y)$  is the transition probability density from  $(s, x)$  to  $(t, y)$ ,  $s < t$ , then  $-\partial_s p = G(p)$ , with terminal condition  $p(t, x; t, y) = \delta_y(x)$  and so  $p(s, x; t, y)$  also satisfies the KBE when looking at it backward in time. The Feller evolution semigroup being time-homogeneous, one may as well observe that with  $p := p(x; t, y)$ , operating the time substitution  $t - s \rightarrow t$ ,  $p$  itself solves the KBE

$$(57) \quad \partial_t p = G(p), \quad p(x; 0, y) = \delta_y(x).$$

In particular, integrating over  $y$ ,  $\partial_t \rho_t(x) = G(\rho_t(x))$ , with  $\rho_0(x) = \mathbf{1}(x \in (0, 1))$ .

$p(x; t, y)$  being a sub-probability, we may define the normalized conditional probability density  $q(x; t, y) := p(x; t, y) / \rho_t(x)$ , now with total mass 1. We get

$$\partial_t q = -\partial_t \rho_t(x) / \rho_t(x) \cdot q + G^*(q), \quad q(x; 0, y) = \delta_y(x).$$

The term  $b_t(x) := -\partial_t \rho_t(x) / \rho_t(x) > 0$  is the time-dependent birth rate at which mass should be created to compensate the loss of mass of the original process due to absorption of  $(x_t; t \geq 0)$  at the boundaries. In this creation of mass process, a diffusing particle started in  $x$  dies at rate  $b_t(x)$  at point  $(t, y)$  where it is duplicated in two new independent particles both started at  $y$  (resulting in a global birth) evolving in the same diffusive way<sup>2</sup>. The birth rate function  $b_t(x)$  depends here on  $x$  and  $t$ , not on  $y$ .

When the boundaries of  $x_t$  are absorbing, the spectra of both  $-G$  and  $-G^*$  are discrete (see [13] pp. 330): There exist positive eigenvalues  $(\lambda_k)_{k \geq 1}$  ordered in ascending sizes and eigenvectors  $(v_k, u_k)_{k \geq 1}$  of both  $-G^*$  and  $-G$  satisfying  $-G^*(v_k) =$

---

<sup>2</sup>Consider a diffusion process with forward infinitesimal generator  $G^*$  governing the evolution of  $p(x; t, y)$ . Suppose that a sample path of this process has some probability that it will be killed or create a new copy of itself, and that the killing and birth rates  $d$  and  $b$  depend on the current location  $y$  of the path. Then the process with the birth and death opportunities of a path has the infinitesimal generator  $\lambda(y) \cdot +G^*(\cdot)$ , where  $\lambda(y) = b(y) - d(y)$ . The rate can also depend on  $t$  and  $x$ .

$\lambda_k v_k$  and  $-G(y_k) = \lambda_k u_k$  such that, with  $\langle u_k, v_k \rangle := \int_0^1 u_k(x) v_k(x) dx$  and  $b_k := \langle u_k, v_k \rangle^{-1}$ , the spectral representation

$$(58) \quad p(x; t, y) = \sum_{k \geq 1} b_k e^{-\lambda_k t} u_k(x) v_k(y)$$

holds.

Let  $\lambda_1 > \lambda_0 = 0$  be the smallest non-null eigenvalue of the infinitesimal generator  $-G^*$  (and of  $-G$ ). Clearly,  $-\frac{1}{t} \log \rho_t(x) \xrightarrow{t \rightarrow \infty} \lambda_1$  and by L' Hospital rule therefore  $b_t(x) \xrightarrow{t \rightarrow \infty} \lambda_1$ . Putting  $\partial_t q = 0$  in the latter evolution equation, independently of the initial condition  $x$

$$(59) \quad q(x; t, y) \xrightarrow{t \rightarrow \infty} q_\infty(y) = v_1(y),$$

where  $v_1$  is the eigenvector of  $-G^*$  associated to  $\lambda_1$ , satisfying  $-G^* v_1 = \lambda_1 v_1$ . The limiting probability  $v_1/\text{norm}$  (after a proper normalization) is called the quasi-stationary Yaglom limit law of  $(x_t; t \geq 0)$  conditioned on being currently alive at all time  $t$  (see [23]).

**14.4. Additive functionals along sample paths.** Let  $(x_t; t \geq 0)$  be the diffusion model defined by (52) on the interval  $I$  where both endpoints are assumed absorbing (exit). This process is thus transient and non-conservative. We wish to evaluate the nonnegative additive quantities

$$\alpha(x) = \mathbf{E} \left( \int_0^{\tau_x} c(x_s) ds + d(x_{\tau_x}) \right),$$

where the functions  $c$  and  $d$  are both assumed nonnegative on  $I$  and  $\partial I = \{0, 1\}$ . The functional  $\alpha(x) \geq 0$  solves the Dirichlet problem:

$$\begin{aligned} -G(\alpha) &= c \text{ if } x \in I \\ \alpha &= d \text{ if } x \in \partial I, \end{aligned}$$

and  $\alpha$  is a super-harmonic function for  $G$ , satisfying  $-G(\alpha) \geq 0$ .

**Some examples:**

1. Assume  $c = 1$  and  $d = 0$  : here,  $\alpha = \mathbf{E}(\tau_x)$  is the mean time of absorption (average time spent in  $(0, 1)$  before absorption).
2. Whenever both  $\{0, 1\}$  are exit boundaries, it is of interest to evaluate the probability that  $x_t$  first hits  $[0, 1]$  (say) at 1, given  $x_0 = x$ . This can be obtained by choosing  $c = 0$  and  $d(\circ) = \mathbf{1}(\circ = 1)$ .
3. Let  $y \in I$  and put  $c = \frac{1}{2\varepsilon} \mathbf{1}(x \in (y - \varepsilon, y + \varepsilon))$  and  $d = 0$ . As  $\varepsilon \rightarrow 0$ ,  $c$  converges weakly to  $\delta_y(x)$  and,  $\alpha =: \mathbf{g}(x, y) = \mathbf{E} \left( \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_0^{\tau_x} \mathbf{1}_{(y-\varepsilon, y+\varepsilon)}(x_s) ds \right) = \int_0^\infty p(x; s, y) ds$  is the Green function, solution to:

$$\begin{aligned} -G(\mathbf{g}) &= \delta_y(x) \text{ if } x \in I \\ \mathbf{g} &= 0 \text{ if } x \in \partial I. \end{aligned}$$

$\mathbf{g}$  is therefore the mathematical expectation of the local time at  $y$ , starting from  $x$  (the sojourn time density at  $y$ ). The solution is known to be (see [13], pp. 198 or

[5], pp. 280)

$$(60) \quad \mathbf{g}(x, y) = 2m(y) (\varphi(1) - \varphi(y)) \frac{\varphi(x) - \varphi(0)}{\varphi(1) - \varphi(0)} \text{ if } x \leq y$$

$$\mathbf{g}(x, y) = 2m(y) (\varphi(y) - \varphi(0)) \frac{\varphi(1) - \varphi(x)}{\varphi(1) - \varphi(0)} \text{ if } x \geq y.$$

The Green function is of particular interest to solve the general problem of evaluating additive functionals  $\alpha(x)$ . Indeed, as is well-known, see [13] for example, the integral operator with respect to the Green kernel inverts the second order operator  $-G$  leading to

$$\alpha(x) = \int_I \mathbf{g}(x, y) c(y) dy \text{ if } x \in I$$

$$\alpha = d \text{ if } x \in \partial I.$$

Under this form,  $\alpha(x)$  appears as a potential function and any potential function is super-harmonic. Note that for all harmonic function  $h \geq 0$  satisfying  $-G(h) = 0$ ,

$$\alpha_h(x) := \int_I \mathbf{g}(x, y) c(y) dy + h(x)$$

is again super-harmonic because  $-G(\alpha_h) = c \geq 0$ .

**14.5. Transformation of sample paths (Doob-transform) producing killing and/or branching.** In the preceding Subsections, we have dealt with a given process and recalled the various ingredients for the expectations of various quantities of interest, summing over the history of paths. In this setup, there is no distinction among paths with different destinations nor did we allow for annihilation or creation of paths inside the domain before the process reached one of the boundaries. The Doob transform of paths allows to do so.

Consider a one-dimensional diffusion  $(x_t; t \geq 0)$  as in (52). Let  $p(x; t, y)$  be its transition probability. Let  $\alpha(x) \geq 0$  as  $x \in [0, 1]$ .

Define a new transformed stochastic process  $(\bar{x}_t; t \geq 0)$  by its transition probability

$$(61) \quad \bar{p}(x; t, y) = \frac{\alpha(y)}{\alpha(x)} p(x; t, y).$$

In this construction of  $(\bar{x}_t; t \geq 0)$  through a change of measure, sample paths  $x \rightarrow y$  of  $(x_t; t \geq 0)$  with a large value of the ratio  $\alpha(y)/\alpha(x)$  are favored. This is a selection of paths procedure due to Doob (see [6]).

The KFE for  $\bar{p}$  clearly is  $\partial_t \bar{p} = \bar{G}^*(\bar{p})$ , with  $p(x; 0, y) = \delta_y(x)$  and  $\bar{G}^*(\bar{p}) = \alpha(y) G^*(\bar{p}/\alpha(y))$ . The adjoint Kolmogorov backward operator of the transformed process is therefore by duality

$$(62) \quad \bar{G}(\cdot) = \frac{1}{\alpha(x)} G(\alpha(x) \cdot).$$

Developing, with  $\alpha'(x) := d\alpha(x)/dx$  and  $\tilde{G}(\cdot) := \frac{\alpha'}{\alpha} g^2 \partial_x(\cdot) + G(\cdot)$ , we get

$$(63) \quad \bar{G}(\cdot) = \frac{1}{\alpha} G(\alpha) \cdot + \tilde{G}(\cdot) =: \lambda(x) \cdot + \tilde{G}(\cdot)$$

and the new KB operator can be obtained from the latter by adding a drift term  $\frac{\alpha'}{\alpha}g^2\partial_x$  to the one in  $G$  of the original process to form a new process  $(\tilde{x}_t; t \geq 0)$  with the KB operator  $\tilde{G}$  and by killing or branching its sample paths at rate  $\lambda(x) := G(\alpha)/\alpha$ . In others words, with  $\tilde{f}(x) := f(x) + \frac{\alpha'}{\alpha}g^2(x)$ , the novel time-homogeneous SDE to consider is

$$(64) \quad d\tilde{x}_t = \tilde{f}(\tilde{x}_t) dt + g(\tilde{x}_t) dw_t, \tilde{x}_0 = x \in (0, 1),$$

possibly killed or branching at rate  $\lambda(x)$  as soon as  $\lambda \neq 0$ . Whenever  $(\tilde{x}_t; t \geq 0)$  is killed, it enters conventionally into some coffin state  $\{\partial\}$  added to the state-space.

Let us look at special cases:

(i) Suppose  $\alpha \geq 0$  is such that  $-G(\alpha) \geq 0$  (By  $\alpha \geq 0$ , we mean  $\alpha > 0$  in  $I$ , possibly with  $\alpha(0)$  or  $\alpha(1)$  equal 0). Then  $\alpha$  is called a super-harmonic (or excessive) function for the process with infinitesimal generator  $G$ .

In this case, the rate  $\lambda(x) =: -d(x)$  satisfies  $\lambda(x) \leq 0$  and only killing occurs at rate  $d(x)$ . Let  $\tilde{\tau}_x$  be the new absorption time at the boundaries of  $(\tilde{x}_t; t \geq 0)$  started at  $x$  (with  $\tilde{\tau}_x = \infty$  would the boundaries be inaccessible to the new process  $\tilde{x}_t$ ). Let  $\tilde{\tau}_{x,\partial}$  be the killing time of  $(\tilde{x}_t; t \geq 0)$  started at  $x$  (the hitting time of  $\partial$ ), with  $\tilde{\tau}_{x,\partial} = \infty$  if  $G(\alpha) \equiv 0$ . Then  $\bar{\tau}_x := \tilde{\tau}_x \wedge \tilde{\tau}_{x,\partial}$  is the novel stopping time for  $(\tilde{x}_t; t \geq 0)$ . The SDE for  $(\tilde{x}_t; t \geq 0)$ , together with its global stopping time  $\bar{\tau}_x$  characterize the new process  $(\bar{x}_t; t \geq 0)$  with full generator  $\bar{G}$  to consider.

(ii) Suppose  $\alpha \geq 0$  is such that  $-G(\alpha) \leq 0$ . Then  $\alpha$  is called a sub-harmonic function for the process with generator  $G$ .

In this case, the rate  $\lambda(x) =: b(x)$  satisfies  $\lambda(x) \geq 0$  and only branching occurs at rate  $b(x)$ . The transformed process (with infinitesimal backward generator  $\bar{G}$ ) accounts for a branching diffusion where a diffusing mother particle (with generator  $\tilde{G}$  and started at  $x$ ) lives a random exponential time with constant rate 1. When the mother particle dies, it gives birth to a spatially dependent random number  $M(x)$  of particles, with mean  $\mu(x) = 1 + \lambda(x)$  (where  $M(x) \stackrel{d}{=} 1 + \Delta(\lambda(x))$ ) and  $\Delta(\lambda(x))$  is a geometrically distributed random variable on  $\{0, 1, 2, \dots\}$  with mean  $\lambda(x)$ . Then  $M(x)$  independent daughter particles are started afresh where their mother particle died, with the event  $M(x) = 0$  impossible; they move along a diffusion governed by  $\tilde{G}$  and reproduce, independently and so on for the subsequent particles.

(iii) If  $\lambda(x) =: b(x)$  is bounded above,  $\lambda$  may be put under the alternative form

$$\lambda(x) = \lambda_* (\mu(x) - 1),$$

where  $\lambda_* = \sup_{x \in [0,1]} \lambda(x)$  and  $1 \leq \mu(x) \leq 2$ . In this case, we can assume that  $M(x)$  can only take the values 1 or 2 with probability  $p_1(x)$  and  $p_2(x)$  respectively, with  $p_1(x) + p_2(x) = 1$ . Then,  $\mu(x) = \mathbf{E}(M(x)) = p_1(x) + 2p_2(x) = 1 + p_2(x)$  and

$$\lambda(x) = \lambda_* p_2(x).$$

Note that  $\mu(x) - 1 = p_2(x) > p_1(x) = 2 - \mu(x)$ . We get a binary branching process at rate  $\lambda_*$  with the event to produce two particles being more likely than the one

to produce a single one, whatever is  $x$ .

(iv) Whenever  $\alpha$  is such that  $-G(\alpha)$  has no constant sign, then both killing and branching can simultaneously occur at the death of the mother particle.  $\lambda(x)$  may be put under the form  $\lambda(x) = b(x) - d(x)$  where  $b(x)$  and  $d(x)$  are the birth (branching) and death (killing) components of  $\lambda(x)$ .

(v) Suppose  $\lambda(x)$  is bounded below and let  $\lambda_* = -\inf_{x \in [0,1]} \lambda(x) > 0$ . Then one may view  $\lambda$  as

$$\lambda(x) = \lambda_*(\mu(x) - 1),$$

where  $\mu(x) \geq 0$ . In this case, branching occurs at rate  $\lambda_*$ . When the mother particle dies, it gives birth to a spatially dependent random number  $M(x)$  of particles (where  $M(x) \stackrel{d}{=} \Delta(\mu(x))$  and  $\Delta(\mu(x))$  is a geometrically distributed random variable on  $\{0, 1, 2, \dots\}$  with mean  $\mu(x) = 1 + \lambda(x)/\lambda_*$ ). With  $p_m(x) = \mathbf{P}(M(x) = m) = p_0(x) q(x)^m$ ,  $m \geq 0$ ,  $p_0(x) = \frac{1}{1+\mu(x)}$ ,  $q_0(x) = 1 - p_0(x)$

$$\lambda(x) = \lambda_* \left( \sum_{m \geq 1} m p_m(x) - 1 \right) = \lambda_* \left( \sum_{m \geq 2} (m-1) p_m(x) - p_0(x) \right).$$

Thus, the decomposition  $\lambda(x) = b(x) - d(x)$  holds, where  $b$  and  $d$  can be read from

$$\lambda(x) = \lambda_* \left( \frac{\mu(x)^2}{1 + \mu(x)} - \frac{1}{1 + \mu(x)} \right).$$

(vi) In some other examples, the killing/branching rate  $\lambda = G(\alpha)/\alpha$  is bounded above and below. Then  $\lambda$  may be put under the form

$$\lambda(x) = \lambda_*(\mu(x) - 1),$$

where  $\lambda_* = \sup_{x \in [0,1]} |\lambda(x)|$  and  $0 \leq \mu(x) \leq 2$ . In this case, we can assume that  $M(x)$  can only take the values 0 or 2 with probability  $p_0(x)$  and  $p_2(x)$  respectively, with  $p_0(x) + p_2(x) = 1$ . Then,  $\mu(x) = \mathbf{E}(M(x)) = 2p_2(x)$  and

$$\lambda(x) = \lambda_*(2p_2(x) - 1) = \lambda_*(p_2(x) - p_0(x)),$$

giving a simple decomposition of  $\lambda$  in the form  $\lambda(x) = b(x) - d(x)$  with the mother particle living a random exponential time now with constant rate  $\lambda_*$  before giving birth to none or two descending particles (a binary branching process). Note that  $p_0(x) \geq p_2(x)$  (respectively  $p_0(x) \leq p_2(x)$ ) when  $\mu(x) \leq 1$  ( $\mu(x) \geq 1$ ).

**Examples of  $\alpha$ .** When  $(x_t; t \geq 0)$  is non-conservative, consider

$$\alpha(x) = \mathbf{E} \left( \int_0^{\tau_x} c(x_s) ds + d(x_{\tau_x}) \right),$$

where the functions  $c$  and  $d$  are both assumed nonnegative on  $I$  and  $\partial I = \{0, 1\}$ . Then  $\alpha \geq 0$  solves the Dirichlet equation  $-G\alpha(x) = c(x) \geq 0$  on  $I$  ( $= d(x)$  on  $\partial I$ ) and so  $\alpha$  is super-harmonic or excessive. We refer to [10] for examples of Doob transforms based on such super-harmonic functions allowing to understand various

conditionings of interest when the starting point process  $(x_t; t \geq 0)$  is a neutral WF diffusion or a WF diffusion with selection.

Whenever  $\alpha$  is super-harmonic for  $G$ , then  $\beta = 1/\alpha \geq 0$  is sub-harmonic for  $\tilde{G} = G + \frac{\alpha'}{\alpha} g^2 \partial_x$ . This results from the obvious identity

$$\beta^{-1} \tilde{G}(\beta) = -\alpha^{-1} G(\alpha),$$

showing that  $-G\alpha \geq 0$  entails  $-\tilde{G}(\beta) \leq 0$ .

Whenever  $(x_t; t \geq 0)$  is conservative and ergodic

$$\frac{1}{t} \mathbf{E}_x \int_0^t c(x_s) ds \xrightarrow{t \rightarrow \infty} \mu(c) := \int_0^1 c(y) d\mu(y)$$

where  $d\mu = m(y) dy$  is the invariant probability measure of  $(x_t; t \geq 0)$ . Define

$$\lim_{t \rightarrow \infty} \mathbf{E}_x \int_0^t c(x_s) ds - t\mu(c) = \alpha(x).$$

Thus

$$\alpha(x) := \int_0^\infty (\mathbf{E}_x(c(x_s)) - \mu(c)) ds$$

solves the Poisson equation

$$-G\alpha(x) = \tilde{c}(x) := c(x) - \mu(c).$$

We conclude that  $\alpha$  is  $G$ -super-harmonic if ever  $c(x) \geq \mu(c)$ ,  $\forall x$ .  $\diamond$

**Background (multiplicative functional and path integral).** The Doob transforms used here are particular instances of more general transformations based on multiplicative functionals. Let  $x_t$  be the diffusion process (52) governed by  $G = f\partial_x + \frac{1}{2}g^2\partial_x^2$  with  $x_0 = x$ .

Define the multiplicative functional  $M_t$  as the solution of the differential equation

$$dM_t = M_t \cdot (a(x_t) dt + b(x_t) dw_t), \quad M_0 = 1,$$

where  $a$  and  $b$  are arbitrary twice differentiable functions. Integrating, we get

$$M_t = e^{\int_0^t (a - \frac{1}{2}b^2)(x_s) ds + \int_0^t b(x_s) dw_s}.$$

Let  $B$  be a Borel subset of  $I$ . Define a new process whose density  $\bar{p}$  is obtained after a modification of the original one while using the multiplicative modulation factor  $M_t$  as

$$\int_B \bar{p}(x; t, y) dy := \mathbf{E}_x [M_t \mathbf{1}(x_t \in B)] = \int_B \mathbf{E}_x [M_t | x_t = y] p(x; t, y) dy.$$

Integrating  $M_t$  over paths with fixed two endpoints  $x$  and  $y$ ,  $\mathbf{E}_x [M_t | x_t = y]$  can be interpreted as the Radon-Nykodym derivative of  $\bar{p}$  with respect to  $p$ , the density of  $x_t$ . By duality, let

$$v(x, t) = \mathbf{E}_x [M_t \psi(x_t)], \quad v(x, 0) = \psi(x).$$

Applying Itô calculus, we get

$$\partial_t v = \bar{G}(v) = (G + gb\partial_x + a)(v) =: (\tilde{G} + a)(v),$$

where the modified backward infinitesimal generator  $\bar{G}$  is obtained by adding a drift term  $gb\partial_x$  to  $G$  to produce  $\tilde{G}$  and a multiplicative part  $a$ . The adjoint KFE giving the evolution of  $\bar{p}$  is thus

$$\partial_t \bar{p} = \bar{G}^* (\bar{p}) = \left( \tilde{G}^* + a \right) (\bar{p}), \quad \bar{p}(x; 0, y) = \delta_y(x).$$

- (Cameron-Martin-Girsanov) For instance, when  $a = 0$  and  $b = -f/g$ , the generator of the transformed diffusion is  $\bar{G} = \frac{1}{2}g^2\partial_x^2$  killing the drift term of the original process governed by  $G$ . In this case,

$$M_t = e^{-\frac{1}{2} \int_0^t \left[ \left( \frac{f}{g} \right)^2 (x_s) ds + 2 \frac{f}{g} (x_s) dw_s \right]}.$$

Clearly in this case  $M_t$  is a martingale with  $\mathbf{E}_x(M_t) = 1$ , assuming  $b$  to be bounded. This construction kills the drift of the original process while using a change of measure.

- (Feynman-Kac) When  $b = 0$ , the generator of the transformed diffusion is  $\bar{G} = G + a$  adding a multiplicative component  $a$  to the one  $G$  governing the original process. In this case

$$M_t = e^{\int_0^t a(x_s) ds}$$

is the exponential of the integrated rate. If  $v(x, t) = \mathbf{E}_x[M_t \psi(x_t)]$ ,  $v(x, 0) = \psi(x)$ , then  $v$  solves

$$\partial_t v = \bar{G}(v) = (G + a)(v), \quad v(x, 0) = \psi(x).$$

In particular, if  $v(x, t) = \mathbf{E}_x[M_t]$ ,  $v(x, 0) = \mathbf{1}(x \in (0, 1))$ , then  $v$  solves

$$\partial_t v = \bar{G}(v) = (G + a)(v), \quad v(x, 0) = \mathbf{1}(x \in (0, 1)).$$

- (Doob) Suppose now

$$dM_t = M_t (\alpha^{-1}(x_t) d\alpha(x_t)), \quad M_0 = 1.$$

This  $M_t$  is a particular instance of the general  $M_t$  introduced above. Indeed, applying Itô calculus,

$$\alpha^{-1}(x_t) d\alpha(x_t) = \alpha^{-1}\alpha' [f dt + g dw] + \frac{1}{2} \alpha^{-1}\alpha'' g^2 dt,$$

leading to

$$\begin{aligned} a &= \alpha^{-1} \left( f\alpha' + \frac{g^2}{2} \alpha'' \right) = G(\alpha) / \alpha =: \lambda(x) \\ b &= \alpha^{-1} \alpha' g. \end{aligned}$$

Thus  $\bar{G} = G + gb\partial_x + a = G + \alpha^{-1}\alpha'g^2\partial_x + \lambda(x)$  as already observed earlier.

Now, from the differential generation of  $M_t$ ,

$$M_t = \frac{\alpha(x_t)}{\alpha(x)}, \quad M_0 = 1$$

only depends on the terminal and initial values of  $(x_s; 0 \leq s \leq t)$  and not on its intermediate values (such a particular Doob transformation is thus a gauge). Thus here

$\mathbf{E}_x [M_t \mid x_t = y] = \frac{\alpha(y)}{\alpha(x)}$  consistently with the definition  $\bar{p}(x; t, y) = \frac{\alpha(y)}{\alpha(x)} p(x; t, y)$ .

### A super-harmonic example.

Although this work chiefly focuses on Doob-transforms where branching is present in  $\lambda$ , let us give a significant example where the Doob transform just produces killing like in (i). Suppose  $(x_t; t \geq 0)$  is a non-conservative diffusion. Let  $\lambda_1$  be the smallest non-null eigenvalue of the infinitesimal generator  $G$  of  $(x_t; t \geq 0)$ . Let  $\alpha = u_1$  be the corresponding eigenvector, that is satisfying  $-Gu_1 = \lambda_1 u_1 \geq 0$  with boundary conditions  $u_1(0) = u_1(1) = 0$ . Then  $c = \lambda_1 u_1$ . The new KB operator associated to the transformed process  $(\bar{x}_t; t \geq 0)$  is

$$(65) \quad \bar{G}(\cdot) = \frac{1}{\alpha} G(\alpha) \cdot + \tilde{G}(\cdot) = -\lambda_1 \cdot + \tilde{G}(\cdot),$$

obtained while killing the sample paths of the process  $(\tilde{x}_t; t \geq 0)$  governed by  $\tilde{G}$  at constant death rate  $d = \lambda_1$ . The transition probability of the transformed stochastic process  $(\bar{x}_t; t \geq 0)$  is

$$\bar{p}(x; t, y) = \frac{u_1(y)}{u_1(x)} p(x; t, y).$$

Define  $\tilde{p}(x; t, y) = e^{\lambda_1 t} \bar{p}(x; t, y)$ . It is the transition probability of the process  $(\tilde{x}_t; t \geq 0)$  governed by  $\tilde{G}$ ; it corresponds to the original process  $(x_t; t \geq 0)$  conditioned on never hitting the boundaries  $\{0, 1\}$  (the so-called  $Q$ -process of  $(x_t; t \geq 0)$ , see [18]). It is simply obtained from  $(x_t; t \geq 0)$  by adding the additional drift term  $\frac{u_1'}{u_1} g^2$  to  $f$ , where  $u_1$  is the eigenvector of  $G$  associated to its smallest non-null eigenvalue. The determination of  $\alpha = u_1$  is a Sturm-Liouville problem. When  $t$  is large, to the dominant order

$$p(x; t, y) \sim e^{-\lambda_1 t} \frac{u_1(x) v_1(y)}{\langle u_1, v_1 \rangle},$$

where  $v_1$  is the Yaglom limit law of  $(x_t; t \geq 0)$ . Therefore

$$(66) \quad \tilde{p}(x; t, y) \sim e^{\lambda_1 t} \frac{u_1(y)}{u_1(x)} e^{-\lambda_1 t} \frac{u_1(x) v_1(y)}{\langle u_1, v_1 \rangle} = \frac{u_1(y) v_1(y)}{\langle u_1, v_1 \rangle}.$$

Thus the limit law of the  $Q$ -process  $(\tilde{x}_t; t \geq 0)$  is the normalized Hadamard product of the eigenvectors  $u_1$  and  $v_1$  associated respectively to  $G$  and  $G^*$ . On the other hand, the limit law of  $(\bar{x}_t; t \geq 0)$  is directly given by

$$(67) \quad \tilde{p}(x; t, y) \xrightarrow{t \rightarrow \infty} \tilde{p}(y) = \frac{1}{Z g^2(y)} e^{2 \int_0^y \frac{f(z) + \left(\frac{u_1'}{u_1} g^2\right)(z)}{g^2(z)} dz} = \frac{u_1^2(y)}{Z g^2(y)} e^{2 \int_0^y \frac{f(z)}{g^2(z)} dz},$$

where  $Z$  is the appropriate normalizing constant. Comparing (66) and (67)

$$v_1(y) = \frac{u_1(y)}{g^2(y)} e^{2 \int_0^y \frac{f(z)}{g^2(z)} dz} = u_1(y) m(y).$$

The eigenvector  $v_1$  associated to  $G^*$  is therefore equal to the eigenvector  $u_1$  associated to  $G$  times the speed density of  $(x_t; t \geq 0)$ .

When dealing for example with the neutral WF diffusion (see Section 4 for additional details), it is known that  $\lambda_1 = 1$  with  $u_1 = x(1-x)$  and  $v_1 \equiv 1$ . The

$Q$ -process  $(\tilde{x}_t; t \geq 0)$  in this case obeys

$$(68) \quad d\tilde{x}_t = (1 - 2\tilde{x}_t) dt + \sqrt{\tilde{x}_t(1 - \tilde{x}_t)} dw_t,$$

with an additional stabilizing drift toward  $1/2$ :  $\tilde{f}(x) = \frac{u_1'}{u_1} g^2(x) = 1 - 2x$ .

The limit law of the  $Q$ -process  $(\tilde{x}_t; t \geq 0)$  in this case is  $6y(1 - y)$ . The latter conditioning is more stringent than the Yaglom conditioning and so the limiting law has more mass away from the boundaries (compare with the uniform quasi-stationary Yaglom limit (59) with  $v_1 \equiv 1$ ).

## 15. THE WRIGHT-FISHER EXAMPLE

In this Section, we briefly and informally recall that the celebrated WF diffusion process with or without a drift may be viewed as a scaling limit of a simple two alleles discrete space-time branching process preserving the total number  $N$  of individuals in the subsequent generations (see [13], [7], for example).

**15.1. The neutral Wright-Fisher model.** Consider a discrete-time Galton Watson branching process preserving the total number of individuals in each generation. We start with  $N$  individuals. The initial reproduction law is defined as follows: Let  $|\mathbf{k}_N| := \sum_{m=1}^N k_m = N$  and  $\mathbf{k}_N := (k_1, \dots, k_N)$  be integers. Assume the first-generation random offspring numbers  $\nu_N := (\nu_N(1), \dots, \nu_N(N))$  admit the following joint exchangeable polynomial distribution on the discrete simplex  $|\mathbf{k}_N| = N$ :

$$\mathbf{P}(\nu_N = \mathbf{k}_N) = \frac{N! \cdot N^{-N}}{\prod_{n=1}^N k_n!}.$$

This distribution can be obtained by conditioning  $N$  independent Poisson distributed random variables on summing to  $N$ . Assume subsequent iterations of this reproduction law are independent so that the population is with constant size for all generations.

Let  $N_r(n)$  be the offspring number of the  $n$  first individuals at the discrete generation  $r \in \mathbf{N}_0$  corresponding to (say) allele  $A_1$  (the remaining number  $N - N_r(n)$  counts the number of alleles  $A_2$  at generation  $r$ ). This sibship process is a discrete-time Markov chain with binomial transition probability given by:

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

Assume next that  $n = [Nx]$  where  $x \in (0, 1)$ . Then, as well-known, the dynamics of the continuous space-time re-scaled process  $x_t := N_{[Nt]}(n)/N$ ,  $t \in \mathbf{R}_+$  can be approximated for large  $N$ , to the leading term in  $N^{-1}$ , by a Wright-Fisher-Itô diffusion on  $[0, 1]$  (the purely random genetic drift case):

$$(69) \quad dx_t = \sqrt{x_t(1 - x_t)} dw_t, \quad x_0 = x.$$

Here  $(w_t; t \geq 0)$  is a standard Wiener process. For this scaling limit process, a unit laps of time  $t = 1$  corresponds to a laps of time  $N$  for the original discrete-time process; thus time is measured in units of  $N$ . If the initial condition is  $x = N^{-1}$ ,  $x_t$  is the diffusion approximation of the offspring frequency of a singleton at generation  $[Nt]$ .

Equation (69) is a 1-dimensional diffusion as in (52) on  $[0, 1]$ , with zero drift  $f(x) = 0$  and volatility  $g(x) = \sqrt{x(1-x)}$ . This diffusion is already in natural coordinate and so  $\varphi(x) = x$ . The scale function is  $x$  and the speed measure  $[x(1-x)]^{-1} dx$ . One can check that both boundaries are exit in this case: The stopping time is  $\tau_x = \tau_{x,0} \wedge \tau_{x,1}$  where  $\tau_{x,0}$  is the extinction time and  $\tau_{x,1}$  the fixation time. The corresponding infinitesimal generators are  $G(\cdot) = \frac{1}{2}x(1-x)\partial_x^2(\cdot)$  and  $G^*(\cdot) = \frac{1}{2}\partial_y^2(y(1-y)\cdot)$ .

**15.2. Non-neutral cases.** Two alleles WF models (with non-null drifts) are classically obtained by considering the binomial transition probabilities  $\text{bin}(N, p_N)$ :

$$\mathbf{P}(N_{r+1}(n) = k' \mid N_r(n) = k) = \binom{N}{k'} \left( p_N \left( \frac{k}{N} \right) \right)^{k'} \left( 1 - p_N \left( \frac{k}{N} \right) \right)^{N-k'}$$

where

$$p_N(x) : x \in (0, 1) \rightarrow (0, 1)$$

is now some state-dependent probability (which is different from the identity  $x$ ) reflecting some deterministic evolutionary drift from the allele  $A_1$  to the allele  $A_2$ . For each  $r$ , we have

$$\begin{aligned} \mathbf{E}(N_{r+1}(n) \mid N_r(n) = k) &= N p_N \left( \frac{k}{N} \right) \\ \sigma^2(N_{r+1}(n) \mid N_r(n) = k) &= N p_N \left( \frac{k}{N} \right) \left( 1 - p_N \left( \frac{k}{N} \right) \right) \end{aligned}$$

which is amenable to a diffusion approximation in terms of  $x_t := N_{[Nt]}(n)/N$ ,  $t \in \mathbf{R}_+$  under suitable conditions.

For instance, taking

$$p_N(x) = (1 - \pi_{2,N})x + \pi_{1,N}(1 - x)$$

where  $(\pi_{1,N}, \pi_{2,N})$  are small ( $N$ -dependent) mutation probabilities from  $A_2$  to  $A_1$  (respectively  $A_1$  to  $A_2$ ). Assuming  $(N \cdot \pi_{1,N}, N \cdot \pi_{2,N}) \xrightarrow{N \rightarrow \infty} (\pi_1, \pi_2)$ , leads after

scaling to the drift of WF model with positive mutations rates  $(\pi_1, \pi_2)$ .

Taking

$$p_N(x) = \frac{(1 + s_{1,N})x}{1 + s_{1,N}x + s_{2,N}(1-x)}$$

where  $s_{i,N} > 0$  are small  $N$ -dependent selection parameter satisfying  $N \cdot s_{i,N} \xrightarrow{N \rightarrow \infty} \sigma_i > 0$ ,  $i = 1, 2$ , leads, after scaling, to the WF model with selective drift  $f(x) = \sigma x(1-x)$ , where  $\sigma := \sigma_1 - \sigma_2$ . Typically, the drift  $f(x)$  is a large  $N$  approximation of the bias:  $N(p_N(x) - x)$ . The WF diffusion with selection is thus:

$$(70) \quad dx_t = \sigma x_t(1-x_t) dt + \sqrt{x_t(1-x_t)} dw_t$$

where time is measured in units of  $N$ . Letting  $\theta_t = Nt$  define a new time-scale with inverse  $t_\theta = \theta/N$ , the time-changed process  $y_\theta = x_{\theta/N}$  now obeys the SDE

$$dy_\theta = s y_\theta(1-y_\theta) d\theta + \sqrt{\frac{1}{N} y_\theta(1-y_\theta)} dw_\theta,$$

with a small diffusion term. Here  $s = s_1 - s_2$  and time  $\theta$  is the usual time-clock.

The WF diffusion with selection (70) tends to drift to  $\circ = 1$  (respectively  $\circ = 0$ ) if allele  $A_1$  is selectively advantageous over  $A_2$  :  $\sigma_1 > \sigma_2$  (respectively  $\sigma_1 < \sigma_2$ ) in the following sense: if  $\sigma > 0$  (respectively  $< 0$ ), the fixation probability at  $\circ = 1$ , which is [16]

$$\mathbf{P}(\tau_{x,1} < \tau_{x,0}) = \frac{1 - e^{-2\sigma x}}{1 - e^{-2\sigma}},$$

increases (decreases) with  $\sigma$  taking larger (smaller) values.

The usual way to look at the WF diffusion with mutation and selection is to compose the two above mechanisms  $p_N(x)$  corresponding to mutation and selection respectively. In the scaling limit, one obtains the standard WF diffusion model including mutations and selection as:

$$(71) \quad dx_t = [(\pi_1 - (\pi_1 + \pi_2)x_t) + \sigma x_t(1 - x_t)] dt + \sqrt{x_t(1 - x_t)} dw_t.$$

## 16. THE NEUTRAL WF MODEL

In this Section, we particularize the general ideas developed in the introductory Section 2 to the neutral WF diffusion (69) and draw some straightforward conclusions most of which are known which illustrate the use of Doob transforms.

**16.1. Explicit solutions of the neutral KFE.** As shown by Kimura in ([15]), the Kolmogorov forward (and backward) equation is exactly solvable in this case, using spectral theory. The solutions involve a series expansion in terms of eigenfunctions of the KB infinitesimal generator with discrete eigenvalues spectrum.

Let  $\lambda_k = k(k+1)/2$ ,  $k \geq 0$ . There exist  $u_k = u_k(x)$  and  $v_k = v_k(y)$  solving the eigenvalue problem:  $-G(u_k) = \lambda_k u_k$  and  $-G^*(v_k) = \lambda_k v_k$ . With  $\langle v_k, u_k \rangle = \int_0^1 u_k(x) v_k(x) dx$ , the transition probability density  $p(x; t, y)$  of the neutral WF models admits the spectral expansion

$$p(x; t, y) = \sum_{k \geq 1} b_k e^{-\lambda_k t} u_k(x) v_k(y) \text{ where } b_k = \frac{1}{\langle v_k, u_k \rangle}$$

where  $u_k(x)$  are the Gegenbauer polynomials rescaled on  $[0, 1]$  and normalized to have value 1 at  $x = 0$ . In particular,  $u_0(x) = x$ ,  $u_1(x) = x - x^2$ ,  $u_2(x) = x - 3x^2 + 2x^3$ ,  $u_3(x) = x - 6x^2 + 10x^3 - 5x^4$ ,  $u_4(x) = x - 10x^2 + 30x^3 - 35x^4 + 14x^5, \dots$

Next,  $v_k(y) = m(y) u_k(y)$  where  $m(y) = 1/(y(1-y))$  is the speed density of the neutral WF diffusion. For instance,  $v_0(y) = \frac{1}{1-y}$ ,  $v_1(y) = 1$ ,  $v_2(y) = 1 - 2y$ ,  $v_3(y) = 1 - 5y + 5y^2$ ,  $v_4(y) = 1 - 9y + 21y^2 - 14y^3, \dots$

Although  $\lambda_0 = 0$  really constitutes an eigenvalue, only  $v_0(y)$  is not a polynomial and the spectral expansion of  $p$  should start at  $k = 1$ , expressing that  $p$  is a sub-probability. When  $k \geq 1$ , from their definition, the  $u_k(x)$  polynomials satisfy  $u_k(0) = u_k(1) = 0$  in such a way that  $v_k(y) = m(y) \cdot u_k(y)$ ,  $k \geq 1$  is a polynomial with degree  $k - 1$ .

The series expansion for  $p(x; t, y)$  solves the KFE of the WF model.

We have  $\mathbf{P}(\tau_x > t) = \int_0^1 \mathbf{P}(x_t \in dy)$  and so

$$\rho_t(x) := \mathbf{P}(\tau_x > t) = \sum_{k \geq 1} \frac{\int_0^1 v_k(y) dy}{\langle v_k, u_k \rangle} e^{-\lambda_k t} u_k(x)$$

is the exact tail distribution of the absorption time.

Since  $v_1(y) = 1$ , to the leading order in  $t$ , for large time

$$\mathbf{P}(x_t \in dy) = 6e^{-t} \cdot x(1-x) dy + \mathcal{O}(e^{-3t})$$

which is independent of  $y$ . Integrating over  $y$ ,  $\rho_t(x) := \mathbf{P}(\tau_x > t) \sim 6e^{-t} \cdot x(1-x)$  so that the conditional probability

$$(72) \quad \mathbf{P}(x_t \in dy \mid \tau_x > t) \underset{t \rightarrow \infty}{\sim} dy$$

is asymptotically uniform in the Yaglom limit. As time passes by, given absorption did not occur in the past,  $x_t \xrightarrow{d} x_\infty$  (as  $t \rightarrow \infty$ ) which is a uniformly distributed random variable on  $[0, 1]$ .

**16.2. Additive functionals for the neutral WF and Doob transforms.** Let  $(x_t; t \geq 0)$  be the neutral WF diffusion model defined by (69) on the interval  $I = [0, 1]$  where both endpoints are absorbing (exit). Consider the additive quantities

$$\alpha(x) = \mathbf{E} \left( \int_0^{\tau_x} c(x_s) ds + d(x_{\tau_x}) \right),$$

where functions  $c$  and  $d$  are both nonnegative. With  $G = \frac{1}{2}x(1-x)\partial_x^2$ ,  $\alpha(x)$  solves:

$$\begin{aligned} -G(\alpha) &= c \text{ if } x \in I \\ \alpha &= d \text{ if } x \in \partial I. \end{aligned}$$

Therefore  $\alpha$  is a super-harmonic function for  $G$ .

Take  $c = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \mathbf{1}_{(y-\varepsilon, y+\varepsilon)}(x) =: \delta_y(x)$  and  $d = 0$ , when  $y \in I$ : in this case,  $\alpha := \mathfrak{g}(x, y)$  is the Green function (the mean local time at  $y$  given the process started at  $x$ ). The solution takes the simple form

$$\begin{aligned} \mathfrak{g}(x, y) &= 2\frac{x}{y} \text{ if } x < y \\ \mathfrak{g}(x, y) &= 2\frac{1-x}{1-y} \text{ if } x > y. \end{aligned}$$

The Green function solves the above general problem of evaluating additive functionals  $\alpha(x)$ :

$$\begin{aligned} \alpha(x) &= \int_I \mathfrak{g}(x, y) c(y) dy \text{ if } x \in I \\ \alpha &= d \text{ if } x \in \partial I. \end{aligned}$$

There are many interesting choices of  $c$  therefore leading to  $\alpha$ , allowing to compute for example the mean time till absorption for the neutral WF diffusion, the probability to hit state 1 before 0... For each choice of  $\alpha$ , it is interesting to study the transformed process  $(\bar{x}_t; t \geq 0)$  whose transition probability is given by

$$\bar{p}(x; t, y) = \frac{\alpha(y)}{\alpha(x)} p(x; t, y),$$

in terms of the original process transition probability  $p(x; t, y)$ . This allows for example to understand the neutral WF process conditioned on exit at some boundary and to evaluate for this new process interesting average additive functionals such as the mean time needed to hit the exit boundary...For detailed similar examples arising in the context of WF diffusions and related ones, see [10].

## 17. THE WF MODEL WITH SELECTION

Now we briefly focus on the diffusion process (70). Let  $(v_k(y))_{k \geq 1}$  be the Gegenbauer eigen-polynomials of the KF operator corresponding to the neutral WF diffusion (69), so with eigenvalues  $\lambda_k = k(k+1)/2$ ,  $k \geq 1$ . Define the oblate spheroidal wave functions on  $[0, 1]$  as

$$(73) \quad w_k^\sigma(y) = \sum_{l \geq 1} f_k^l v_l(y),$$

where  $f_k^l$  obey the three-term recurrence defined in [20]. In the latter equality, the  $l$  summation is over odd (even) values if  $k$  is even (odd).

Define  $v_k^\sigma(y) = e^{\sigma y} w_k^\sigma(y)$  and  $u_k^\sigma(x) = \frac{1}{m(x)} v_k^\sigma(x)$  where  $m(x) = e^{2\sigma x} / (x(1-x))$  is the speed measure density of the WF model with selection (70).

The system  $(u_k^\sigma(x), v_k^\sigma(x))_{k \geq 1}$  constitute a system of eigen-functions for the WF with selection generators  $-G$  and  $-G^*$  with eigenvalues  $\lambda_k^\sigma$  implicitly defined in [20], thus with  $-G(u_k^\sigma) = \lambda_k^\sigma u_k^\sigma$  and  $-G^*(v_k^\sigma) = \lambda_k^\sigma v_k^\sigma$ . The eigen-function expansion of the transition probability density of the WF model with selection is thus, [14]:

$$(74) \quad p(x; t, y) = \sum_{k \geq 1} b_k^\sigma e^{-\lambda_k^\sigma t} u_k^\sigma(x) v_k^\sigma(y)$$

where  $b_k^\sigma = \langle v_k^\sigma, u_k^\sigma \rangle^{-1}$ . The WF model with selection can be viewed as a perturbation problem of the neutral WF model (see [21]). There exist perturbation developments of  $\lambda_k^\sigma$  around  $\lambda_k$  with respect to  $\sigma^2$ , [14]. They are valid and useful for small  $\sigma$ .

The WF diffusion process  $x_t$  with selection (70) is non-conservative, with finite hitting time  $\tau_x$  of one of the boundaries. Following the general arguments developed in Section 2, the Yaglom limit of  $x_t$  conditioned on  $\tau_x > t$  is the normalized version of

$$(75) \quad v_1^\sigma(y) = e^{\sigma y} w_1^\sigma(y).$$

The limit law of  $x_t$  conditioned on never hitting the boundaries in the remote future is the normalized version of

$$(76) \quad u_1^\sigma(y) v_1^\sigma(y) = \frac{1}{m(y)} v_1^\sigma(x)^2 = y(1-y) w_1^\sigma(y)^2.$$

Because the latter conditioning is more stringent than the former, the probability mass of (76) is more concentrated inside the interval than (75). Compare with the statements at the end of Section 2 concerning the neutral WF diffusion.

## 18. DOOB TRANSFORM OF THE NEUTRAL WF MODEL: SUB-CRITICAL BD

In this Section, we define the branching WF diffusion model with selection while applying a Doob transform to the neutral WF model, based on the sub-harmonic additive functional  $\alpha(x) = e^{\sigma x}$ , say with  $\sigma > 0$ . We then study in detail the obtained branching process.

The starting point is thus the neutral WF diffusion:  $dx_t = \sqrt{x_t(1-x_t)}dw_t$ ,  $x_0 = x \in (0, 1)$ .

For this model,  $G = \frac{1}{2}x(1-x)\partial_x^2$  and both boundaries are exit. With  $\lambda_k = k(k+1)/2$ ,  $k \geq 0$ , its transition density  $p(x; t, y)$  admits the spectral representation

$$(77) \quad p(x; t, y) = \sum_{k \geq 1} b_k e^{-\lambda_k t} u_k(x) v_k(y),$$

in terms of the Gegenbauer eigen-polynomials (see Subsection 4.1). We shall consider the following transformation of paths on the neutral WF model: Let  $\alpha(x) = e^{\sigma x}$ ,  $\sigma > 0$  and consider  $\bar{G}(\cdot) = \alpha^{-1}G(\alpha \cdot) = \tilde{G}(\cdot) + b(x) \cdot$ . We now have  $G(\alpha) = \frac{1}{2}\sigma^2 x(1-x)e^{\sigma x}$  and so  $b(x) = G(\alpha)/\alpha = \frac{\sigma^2}{2}x(1-x) \geq 0$ .

Note that  $-G(\alpha) \leq 0$  indicating that  $\alpha$  is sub-harmonic for  $G$ .

In this case study, one selects sample paths of  $(x_t; t \geq 0)$  with large  $\alpha(y)$  and we claim that this is an alternative interesting way to introduce selection in the neutral WF diffusion process.

The dynamics of  $(\tilde{x}_t; t \geq 0)$  governed by  $\tilde{G}$  is easily seen to be the standard WF with selection dynamics (70)

$$d\tilde{x}_t = \sigma \tilde{x}_t(1-\tilde{x}_t)dt + \sqrt{\tilde{x}_t(1-\tilde{x}_t)}dw_t,$$

subject to additional quadratic branching at rate  $b(x) = \frac{1}{2}\sigma^2 x(1-x)$  inside  $I$ . We indeed have

$$\bar{G}(\cdot) = e^{-\sigma x} G(e^{\sigma x} \cdot) = b(x) \cdot + \tilde{G}(\cdot),$$

where

$$\tilde{G} =: \tilde{f}\partial_x + \frac{1}{2}\tilde{g}^2\partial_x^2 = \sigma x(1-x)\partial_x + \frac{1}{2}x(1-x)\partial_x^2$$

is the KBE operator of the dynamics  $(\tilde{x}_t; t \geq 0)$ . Recall that  $\tilde{x}_t$  is transient and so hits one of the boundaries  $\{0, 1\}$  in finite time  $\tilde{\tau}_x$ .

To summarize, in our branching diffusion way to look at selection, we move from the neutral WF diffusion  $(x_t; t \geq 0)$  to the standard WF diffusion with selection  $(\tilde{x}_t; t \geq 0)$  but subject to additional branching at rate  $b(x)$ .

**Remark.** With  $\beta(x) := \alpha(x)^{-1} = e^{-\sigma x}$ , we clearly have

$$\bar{G}(\beta(x)) = 0$$

and  $\beta$  is an harmonic function for  $\bar{G}$  and as a result, Doob-transforming  $\bar{G}$  by  $\beta$ , we get

$$\beta^{-1}\bar{G}(\beta \cdot) = (\alpha\beta)^{-1}\bar{G}(\alpha\beta \cdot) = G(\cdot)$$

which is the infinitesimal generator of the original neutral WF martingale.  $\diamond$

The birth (creating) rate  $b \geq 0$  in  $\bar{G}$  is bounded from above on  $(0, 1)$ . It may be put into the canonical form  $b(x) = b_*(\mu(x) - 1)$  where  $b_* = \max_{x \in [0,1]} (b(x)) = \frac{\sigma^2}{8} > 0$  and

$$(78) \quad \mu(x) = 1 + 4x(1 - x),$$

whose range is the interval  $[1, 2]$  as  $x \in [0, 1]$ .

The density of the transformed process is  $\bar{p}(x; t, y) = \frac{\alpha(y)}{\alpha(x)} p(x; t, y)$ . It is exactly known because so is  $p$  is from (77).

The transformed process (with infinitesimal backward generator  $\bar{G}$ ) accounts for a branching diffusion (BD) where a diffusing mother particle (with generator  $\tilde{G}$  and started at  $x$ ) lives a random exponential time with constant rate  $b_*$ . When the mother particle dies, it gives birth to a spatially dependent random number  $M(x)$  of particles (with mean  $\mu(x)$ ).  $M(x)$  independent daughter particles are started where their mother particle died; they move along a WF diffusion with selection and reproduce, independently and so on.

Because  $\mu(x)$  is bounded above by 2 and larger than 1 (indicating a super-critical branching process), we actually get a BD with binary scission whose random offspring number satisfies ('w.p.' meaning 'with probability')

$$\begin{aligned} M(x) = 0 \text{ w.p. } p_0(x) &= 0 \\ M(x) = 1 \text{ w.p. } p_1(x) &= 2 - \mu(x) \\ M(x) = 2 \text{ w.p. } p_2(x) &= \mu(x) - 1, \end{aligned}$$

with  $p_2(x) \geq p_1(x)$  (the event that 2 particles are generated in a splitting event is more probable than a single one).

For such a transformed process, the trade-off is as follows: there is a competition between the boundaries  $\{0, 1\}$  which are absorbing for the particle system and the number of particles  $N_t(x)$  in the system at each time  $t$ , which may grow due to binary branching events (or remain steady when  $M(x) = 1$ ).

The density  $\bar{p}$  of the transformed process has the following interpretation

$$(79) \quad \bar{p}(x; t, y) = \mathbf{E} \left[ \sum_{n=1}^{N_t(x)} p^{(n)}(x; t, y) \right],$$

where  $p^{(n)}(x; t, y)$  is the density at  $(t, y)$  of the  $n$ th alive particle descending from the ancestral one (Eve), started at  $x$ . In the latter formula, the sum vanishes if  $N_t(x) = 0$ . A particle is alive at time  $t$  if it came to birth before  $t$  and has not been yet absorbed by the boundaries.

Let  $\bar{\rho}_t(x) = \int_{(0,1)} \bar{p}(x; t, y) dy$ . Then  $\bar{\rho}_t(x)$  is the expected number of particle alive at time  $t$ . We have

$$\partial_t \bar{\rho}_t(x) = \bar{G}(\bar{\rho}_t(x)), \quad \bar{\rho}_0(x) = \mathbf{1}(x \in (0, 1)).$$

**Remark.** From the Feynman-Kac formula,  $\bar{p}$  in (79) is also

$$\bar{p}(x; t, y) = \mathbf{E}_x \left( e^{\int_0^{t \wedge \bar{\tau}_x} b(\tilde{x}_s) ds} \mid \tilde{x}_t = y \right) p(x; t, y)$$

and

$$\bar{\rho}_t(x) = \mathbf{E}_x \left( e^{\int_0^{t \wedge \bar{\tau}_x} b(\bar{x}_s) ds} \right). \diamond$$

But then  $\bar{q}(x; t, y) := \bar{p}(x; t, y) / \bar{\rho}_t(x)$  obeys the forward PDE

$$\partial_t \bar{q}(x; t, y) = \left( -\frac{\partial_t \bar{\rho}_t(x)}{\bar{\rho}_t(x)} + b(y) \right) \bar{q}(x; t, y) + \tilde{G}^*(\bar{q}(x; t, y))$$

as a result of  $\partial_t \bar{p}(x; t, y) = \bar{G}^*(\bar{p}(x; t, y))$ . We have

$$(80) \quad \bar{q}(x; t, y) = \frac{\mathbf{E} \left[ \sum_{n=1}^{N_t(x)} p^{(n)}(x; t, y) \right]}{\mathbf{E} [N_t(x)]}$$

showing that  $\bar{q}(x; t, y)$  is the average presence density at  $(t, y)$  of the system of particles all descending from Eve started at  $x$ .

Clearly  $-\frac{\log \bar{\rho}_t(x)}{t} \xrightarrow{t \rightarrow \infty} \lambda_1 = 1$  (and therefore also  $-\frac{\partial_t \bar{\rho}_t(x)}{\bar{\rho}_t(x)}$  by L' Hospital rule), because

$$\bar{\rho}_t(x) = \frac{1}{\alpha(x)} \sum_{k \geq 1} b_k e^{-\lambda_k t} u_k(x) \int_0^1 \alpha(y) v_k(y) dy.$$

The expected number of particles in the system decays globally and exponentially at rate  $\lambda_1$ .

The BD transformed process therefore admits an integrable Yaglom limit  $\bar{q}_\infty$ , solution to  $-\tilde{G}^*(\bar{q}_\infty) = (\lambda_1 + b(y)) \bar{q}_\infty$  or  $-\bar{G}^*(\bar{q}_\infty) = \lambda_1 \bar{q}_\infty$ . With  $v_1(y) = 1$ , the first eigenvector of  $-\bar{G}^*$  associated to the smallest positive eigenvalue  $\lambda_1 = 1$ ,  $\bar{q}_\infty$  is of the product form

$$(81) \quad \bar{q}_\infty(y) = C_* e^{\sigma y} v_1(y) = \frac{\sigma e^{\sigma y}}{e^\sigma - 1}.$$

The arbitrary multiplicative constant  $C_*$  was chosen in such a way that  $\bar{q}_\infty(y)$  is a probability.

By analogy with the Yaglom construction, this limiting probability  $\bar{q}_\infty$  can be called the quasi-stationary Yaglom average density at  $(t, y)$  for the BD particle system (it is also the ground state for  $\bar{G}^*$ ).

There is also a natural eigenvector  $\bar{\phi}_\infty$  of the backward operator  $-\bar{G}$ , satisfying  $-\bar{G}(\bar{\phi}_\infty) = \lambda_1 \bar{\phi}_\infty$  (the ground state for  $\bar{G}$ ). It is explicitly here

$$(82) \quad \bar{\phi}_\infty(x) = \frac{C}{\alpha(x)} u_1(x) = \frac{6(e^\sigma - 1)}{\sigma} e^{-\sigma x} x(1-x).$$

The arbitrary multiplicative constant  $C = 6/C_*$  was chosen in such a way that  $\int_0^1 \bar{q}_\infty(y) \bar{\phi}_\infty(y) dy = 1$ . Note that the spectral structures of both  $\bar{G}^*$  and  $\bar{G}$  are easily obtainable from the ones of  $G^*$  and  $G$  thanks to the Doob transform structure.

In the terminology of [22], both operators  $\bar{G}(\cdot) + \lambda_1 \cdot$  and its adjoint are critical<sup>3</sup>. In this context, the constant  $\lambda_1$  is called the generalized principal eigenvalue. The

---

<sup>3</sup> $\bar{G}(\cdot) + \lambda_1 \cdot$  ( $\bar{G}^*(\cdot) + \lambda_1 \cdot$ ) is said to be critical if there exists some function  $\bar{\phi}_\infty \in C^2$  (respectively  $\bar{q}_\infty \in C^2$ ), strictly positive in  $(0, 1)$ , such that:  $\bar{G}(\bar{\phi}_\infty) + \lambda_1 \bar{\phi}_\infty = 0$  (respectively  $\bar{G}^*(\bar{q}_\infty) + \lambda_1 \bar{q}_\infty = 0$ ) and the operators do not possess a minimal positive Green function.

eigen-functions  $(\bar{\phi}_\infty, \bar{q}_\infty)$  are their associated ground states. We note that we have the  $L^1$ -product property (See [22], Subsection 4.9).

$$\int_0^1 \bar{\phi}_\infty(x) \bar{q}_\infty(x) dx = 6 \int_0^1 u_1(x) v_1(x) dx = 1 < \infty.$$

**Remark.** Using the Feynman-Kac representation of  $\bar{\rho}_t(x)$ , we get

$$\begin{aligned} -\frac{1}{t} \log \mathbf{E}_x \left( e^{\int_0^{t \wedge \bar{\tau}_x} b(\tilde{x}_s) ds} \right) &\xrightarrow{t \rightarrow \infty} \lambda_1 = 1 \text{ and} \\ e^{\lambda_1 t} \mathbf{E}_x \left( e^{\int_0^{t \wedge \bar{\tau}_x} b(\tilde{x}_s) ds} \right) &\xrightarrow{t \rightarrow \infty} \bar{\phi}_\infty(x). \diamond \end{aligned}$$

With  $p_m(x) = \mathbf{P}(M(x) = m)$ , let

$$l(x) = \sum_{m \geq 1} p_m(x) m \log m = 2 \log 2 p_2(x).$$

We have the  $x \log x$  condition:

$$(83) \quad \int_0^1 l(x) \bar{\phi}_\infty(x) \bar{q}_\infty(x) dx = 48 \log 2 \int_0^1 x(1-x) u_1(x) v_1(x) dx < \infty.$$

We conclude (following [1] and [2]) that, as a result of the condition (83) being trivially satisfied, global extinction holds in the following sense:

- (i)  $\mathbf{P}(N_t(x) = 0) \xrightarrow{t \rightarrow \infty} 1$ , uniformly in  $x$ .
- (ii) there exists a constant  $\gamma > 0$ :  $e^{\lambda_1 t} [1 - \mathbf{P}(N_t(x) = 0)] \xrightarrow{t \rightarrow \infty} \gamma \bar{\phi}_\infty(x)$ , uniformly in  $x$ .
- (iii) For all bounded measurable function  $\psi$  on  $I$ :

$$\mathbf{E} \left[ \sum_{n=1}^{N_t(x)} \psi(\tilde{x}_t^{(n)}) \mid N_t(x) > 0 \right] \xrightarrow{t \rightarrow \infty} \gamma^{-1} \int_{(0,1)} \psi(y) \bar{q}_\infty(y) dy.$$

From (i), it is clear that the process gets ultimately extinct with probability 1. In the trade-off between pure branching and absorption at the boundaries, all particles get eventually absorbed and the global BD process turns out to be sub-critical (even though  $\mu(x) = \mathbf{E}M(x) > 1$  for all  $x \in (0, 1)$ ): Probability mass escapes out of  $I$  although the BD survives with positive probability.

In the statement (ii), the quantity  $1 - \mathbf{P}(N_t(x) = 0) = \mathbf{P}(N_t(x) > 0)$  is also  $\mathbf{P}(T(x) > t)$  where  $T(x)$  is the global extinction time of the particle system descending from an Eve particle started at  $x$ . The number  $-\lambda_1$  is the usual Malthus exponential decay rate parameter. From (ii),  $\bar{\phi}_\infty(x)$  has a natural interpretation in terms of the propensity of the particle system to survive to its extinction fate: the so-called reproductive value in demography.

(iii) with  $\psi = 1$  reads  $\mathbf{E}[N_t(x) | N_t(x) > 0] \xrightarrow{t \rightarrow \infty} \gamma^{-1}$  giving an interpretation of the constant  $\gamma$  (which may be hard to evaluate in practise).

The ground states of  $\bar{G} + \lambda_1$  and its adjoint are thus  $(\bar{\phi}_\infty, \bar{q}_\infty)$  and explicit here. It is useful to consider the process whose infinitesimal generator is given by the Doob-transform

$$\bar{\phi}_\infty^{-1} (\bar{G} + \lambda_1) (\bar{\phi}_\infty \cdot) = \bar{\phi}_\infty^{-1} (\tilde{G} + b + \lambda_1) (\bar{\phi}_\infty \cdot),$$

because product-criticality is preserved under this transformation. The ground states associated to this new operator and its dual are  $(1, \bar{\phi}_\infty \bar{q}_\infty)$ . Developing, we obtain a process whose infinitesimal generator is

$$\tilde{G} + \frac{\bar{\phi}'_\infty}{\bar{\phi}_\infty} g^2 \partial_x = G + \frac{u'_1}{u_1} g^2 \partial_x,$$

with no multiplicative part. In our case study, we get  $\frac{1}{2}x(1-x)\partial_x^2 + (1-2x)\partial_x$  adding a stabilizing drift towards 1/2 to the original neutral WF model. The associated diffusion process is positive recurrent and so its invariant measure  $\bar{\phi}_\infty \bar{q}_\infty = 6u_1v_1 = 6y(1-y)$  is integrable with mass 1. It is the beta(2,2) limit law of the  $Q$ -process (see (68) and the comments at the end of Section 2 relative to the neutral WF diffusion).

### Remarks.

(i) At time  $t$ , let  $(\tilde{x}_t^{(n)})_{n=1}^{N_t(x)}$  denote the positions of the BD particle system. Let  $u(x, t; z) = \mathbf{E} \left[ \prod_{n=1}^{N_t(x)} z^{\psi(\tilde{x}_t^{(n)})} \right]$  stand for the functional generating function ( $|z| \leq 1$ ) of the measure-valued branching particle system.  $u(x, t; z)$  obeys the nonlinear (quadratic) Kolmogorov-Petrovsky-Piscounoff PDE, [17]:

$$\partial_t u(x, t; z) = b_* \theta(x, u(x, t; z)) + \tilde{G}(u(x, t; z)); \quad u(x, 0; z) = z^{\psi(x)},$$

where  $\theta(x, z) = \mathbf{E}[z^{M(x)}] - z = (p_2(x)z^2 + p_1(x)z) - z$  or

$$\theta(x, z) = 4x(1-x)z(z-1)$$

is the shifted probability generating function of the branching law of  $M(x)$ . Thus, the nonlinear part reads  $b_* \theta(x, u(x, t; z)) = b(x)u(x, t; z)(u(x, t; z) - 1)$  which is quadratic in  $u$ .

In particular, if  $u(x, t) := \partial_z u(x, t; z)|_{z=1} = \mathbf{E} \left[ \sum_{n=1}^{N_t(x)} \psi(\tilde{x}_t^{(n)}) \right]$ ,  $u(x, t)$  obeys the linear backward PDE

$$\partial_t u(x, t) = b(x)u(x, t) + \tilde{G}(u(x, t)); \quad u(x, 0) = \psi(x)$$

involving  $\bar{G}(\cdot) = \tilde{G}(\cdot) + b(x)\cdot$ . We have the Feynman-Kac interpretation

$$u(x, t) = \mathbf{E}_x \left( e^{\int_0^{t \wedge \bar{\tau}_x} b(\tilde{x}_s) ds} \psi(\tilde{x}_t) \right).$$

The latter evolution equation is the backward version of the forward PDE giving the evolution of  $\bar{p}(x; t, y)$  as  $\partial_t \bar{p}(x; t, y) = \bar{G}^*(\bar{p}(x; t, y))$ ,  $\bar{p}(x; 0, y) = \delta_x(y)$ .

(ii) Let us look at the branching diffusion process governed by  $\bar{G}$  would time be measured using the time substitution

$$\theta_t = \int_0^t g^2(\tilde{x}_s) ds = \int_0^t \tilde{x}_s(1 - \tilde{x}_s) ds$$

for each of the particles that came to birth before  $t$ .

Then  $\bar{G} \rightarrow \bar{\mathcal{G}} := \frac{1}{x(1-x)}\bar{G} = \sigma\partial_x + \frac{1}{2}\partial_x^2 + \frac{1}{2}\sigma^2 \cdot$ . In particular, each motion  $y_\theta = \tilde{x}_{t_\theta}$  is a Brownian motion with constant drift (a Gaussian process). This new  $\bar{\mathcal{G}}$  is the one of absorbing Brownian motion with drift  $\sigma$  on  $[0, 1]$ , including branching at constant rate  $\frac{1}{2}\sigma^2$ . The Sturm-Liouville problem for  $\bar{\mathcal{G}}$  admits the eigenvalues  $\lambda_k = \frac{k^2 + \sigma^2}{2}$ ,  $k \geq 1$  with eigen-states  $u_k(x) \propto e^{-\sigma x} \sin(k\pi x)$  and  $v_k(y) \propto e^{\sigma y} \sin(k\pi y)$ . The spectral gap is  $\lambda_1 = \frac{1 + \sigma^2}{2} > 0$  and the time-changed branching diffusion also becomes eventually extinct, sub-critically: The time substitution changes the spectral structure of the model but not its qualitative features.  $\diamond$

## 19. DOOB TRANSFORM OF THE WF MODEL WITH MUTATIONS: CRITICAL BD

In this Section, we start from the WF model with mutations. Using the same Doob transform based on the additive functional  $\alpha(x) = e^{\sigma x}$  to introduce selection, we end up with a WF diffusion process with killing and branching describing the effect of selection on the WF model in the presence of mutations. We show that in this setup, the resulting branching diffusion process is no longer sub-critical; rather, it turns out to be critical.

Suppose the starting point model is now the WF diffusion with mutations:

$$dx_t = (\pi_1 - \pi x_t) dt + \sqrt{x_t(1-x_t)} dw_t, x_0 = x \in (0, 1),$$

with  $\pi := \pi_1 + \pi_2$ . For this model,  $G = (\pi_1 - \pi x)\partial_x + \frac{1}{2}x(1-x)\partial_x^2$  and both boundaries are chosen as being entrance (reflecting)<sup>4</sup>. The WF diffusion process with mutations is now ergodic. With

$$\lambda_k = \frac{k(k-1+\pi)}{2}, k \geq 0,$$

its transition density  $p(x; t, y)$  now admits the discrete spectral representation

$$(84) \quad p(x; t, y) = \sum_{k \geq 0} b_k e^{-\lambda_k t} u_k(x) v_k(y).$$

Here,  $u_k(x)$  are the Jacobi polynomials rescaled on  $[0, 1]$  and normalized to have value 1 at  $x = 0$ . In particular,  $u_0(x) = 1$ ,  $u_1(x) = 1 - \frac{\pi}{\pi_2}x$ ,  $u_2(x) = 1 - \frac{2(1+\pi)}{\pi_2}x + \frac{(1+\pi)(2+\pi)}{\pi_2(1+\pi_2)}x^2, \dots$  Next,  $v_k(y) = m(y) u_k(y)$  where

$$m(y) = \frac{\Gamma(2\pi)}{\Gamma(2\pi_1)\Gamma(2\pi_2)} y^{2\pi_1-1} (1-y)^{2\pi_2-1}$$

<sup>4</sup>When both the mutation rates  $u_1$  and  $u_2$  are greater than  $1/2$ , the boundaries are entrance. When either  $u_1$  or  $u_2$  is smaller than  $1/2$  the corresponding boundary is regular and one needs to specify whether it is reflecting or absorbing or a mixture of the two. We force here the regular boundaries to be entrance.

is the speed density of the ergodic WF diffusion with mutations (its normalized invariant measure). Note that the  $k = 0$  term in (84) is precisely  $m(y)$  as required. Because the transition probability density of the WF diffusion with mutations has also a discrete spectral representation, this model is amenable to a similar analysis than the neutral WF diffusion.

Proceeding as for the neutral case, we shall consider the following transformation of paths for the WF model with mutations: Let  $\alpha(x) = e^{\sigma x}$  and consider a transformed process with infinitesimal generator  $\bar{G}(\cdot) = \alpha^{-1}G(\alpha\cdot)$ . The multiplicative part of  $\bar{G}$  is now

$$\lambda(x) = G(\alpha)/\alpha = \sigma(\pi_1 - \pi x) + \frac{\sigma^2}{2}x(1-x).$$

Note that now  $\alpha$  is neither sub-harmonic nor super-harmonic for the infinitesimal generator  $G$  including mutations because the sign of  $\lambda(x)$  varies as  $x$  varies.

In this case study, one selects sample paths of the WF diffusion model with mutations  $(x_t; t \geq 0)$  with large terminal values of  $\alpha(y)$ . The dynamics of  $(\tilde{x}_t; t \geq 0)$  is easily seen to be the WF with mutation and selection dynamics of the type (71)

$$d\tilde{x}_t = [(\pi_1 - \pi\tilde{x}_t) + \sigma\tilde{x}_t(1 - \tilde{x}_t)]dt + \sqrt{\tilde{x}_t(1 - \tilde{x}_t)}dw_t,$$

subject to additional quadratic killing and branching at rate  $\lambda(x)$  inside  $I$ . We indeed have

$$\bar{G}(\cdot) = e^{-\sigma x}G(e^{\sigma x}\cdot) = \lambda(x)\cdot + \tilde{G}(\cdot),$$

where  $\tilde{G} = [(\pi_1 - \pi x) + \sigma x(1-x)]\partial_x + \frac{1}{2}x(1-x)\partial_x^2$  is the KBE operator of the dynamics  $(\tilde{x}_t; t \geq 0)$ .

To summarize, in our branching diffusion way to look at the action of selection, we move from the WF diffusion with mutations  $(x_t; t \geq 0)$  to the standard WF diffusion with mutation and selection  $(\tilde{x}_t; t \geq 0)$ , but subject to additional killing/branching at rate  $\lambda(x)$ .

**Remark.** With  $\beta = \alpha^{-1} = e^{-\sigma x}$ , again  $\bar{G}(\beta) = 0$  and  $\beta^{-1}\bar{G}(\beta\cdot) = G(\cdot)$  is the infinitesimal generator of the original WF model, now with mutations.  $\diamond$

The birth (creating) and death (annihilating) rate  $\lambda$  in  $\bar{G}$  is bounded from above and below on  $(0, 1)$ . It may now be put into the canonical form  $\lambda(x) = \lambda_*(\mu(x) - 1)$  where  $\lambda_* = \max_{x \in [0,1]} (|\lambda(x)|)$  and

$$(85) \quad \mu(x) = 1 + \frac{\lambda(x)}{\lambda_*}$$

whose range belongs to the interval  $[0, 2]$  as  $x \in [0, 1]$ .

Note that when  $\pi > \sigma/2$ ,  $\lambda_* = \sigma(\pi_1 \vee \pi_2)$  whereas when  $\pi < \sigma/2$ ,  $\lambda_* = \lambda(x_*) \vee \sigma\pi_2$  where  $x_* = 1/2 - \pi/\sigma > 0$ .

The density of the transformed process is  $\bar{p}(x; t, y) = \frac{\alpha(y)}{\alpha(x)}p(x; t, y)$ . It is exactly known because  $p$  is known from (84).

The transformed process (with infinitesimal backward generator  $\overline{G}$ ) accounts for a branching diffusion (BD) where a diffusing mother particle (with generator  $\tilde{G}$  and started at  $x$ ) lives a random exponential time with constant rate  $\lambda_*$ . When the mother particle dies, it gives birth to a spatially dependent random number  $M(x)$  of particles (with mean  $\mu(x)$ ). If  $M(x) \neq 0$ ,  $M(x)$  independent daughter particles are started where their mother particle died; they move along a WF diffusion with mutation and selection (with generator  $\tilde{G}$ ) and reproduce independently, and so on. Because  $\mu(x)$  is bounded above by 2 and larger than 0, we actually get a BD with binary scission whose random offspring number satisfies

$$M(x) = 0 \text{ w.p. } p_0(x) = 1 - \mu(x)/2$$

$$M(x) = 1 \text{ w.p. } p_1(x) = 0$$

$$M(x) = 2 \text{ w.p. } p_2(x) = \mu(x)/2.$$

Note that

$$\lambda(x) = \lambda_* (p_2(x) - p_0(x)) =: b(x) - d(x)$$

identifying the birth and death components of the full multiplicative rate  $\lambda(x)$ .

For such a transformed process, the trade-off is of a different nature: there is a competition between the boundaries  $\{0, 1\}$  which are now reflecting for the system of particles and the number of particles  $N_t(x)$  in the system at each time  $t$ , which may grow or diminish due either to branching or killing events. In the presence of mutations, the particles are no longer killed once they hit the boundaries, suggesting that there should be a greater amount of them alive in the system. However, in this new model, there is an opportunity to kill the particles inside the definition domain, when they branch. The question now being: does the new trade-off result in global extinction or global explosion of the particle system? We will now show that critical global extinction occurs.

The density  $\bar{p}$  of the transformed process again has the interpretation (79), where  $p^{(n)}(x; t, y)$  is the density at  $(t, y)$  of the  $n$ th alive particle descending from the ancestral one (Eve), started at  $x$ . In the latter formula, the sum vanishes if  $N_t(x) = 0$ . A particle is alive at time  $t$  if it came to birth before  $t$  and has not yet been killed by a killing event.

Let  $\bar{p}_t(x) = \int_{(0,1)} \bar{p}(x; t, y) dy$ . Then  $\bar{p}_t(x)$  is the expected number of particle alive at time  $t$ . We have

$$\partial_t \bar{p}_t(x) = \overline{G}(\bar{p}_t(x)), \quad \bar{p}_0(x) = \mathbf{1}(x \in (0, 1)).$$

But then  $\bar{q}(x; t, y) := \bar{p}(x; t, y) / \bar{p}_t(x)$  obeys the forward PDE

$$\partial_t \bar{q}(x; t, y) = \left( -\frac{\partial_t \bar{p}_t(x)}{\bar{p}_t(x)} + b(y) \right) \bar{q}(x; t, y) + \tilde{G}^*(\bar{q}(x; t, y))$$

as a result of  $\partial_t \bar{p}(x; t, y) = \overline{G}^*(\bar{p}(x; t, y))$ . We again have (80), with  $\bar{q}(x; t, y)$  the average presence density at  $(t, y)$  of the system of particles all descending from Eve started at  $x$ .

Clearly  $-\frac{\log \bar{\rho}_t(x)}{t} \xrightarrow{t \rightarrow \infty} \lambda_0 = 0$  (and therefore also  $-\frac{\partial_t \bar{\rho}_t(x)}{\bar{\rho}_t(x)}$ ), because

$$\bar{\rho}_t(x) = \frac{1}{\alpha(x)} \sum_{k \geq 0} b_k e^{-\lambda_k t} u_k(x) \int_0^1 \alpha(y) v_k(y) dy.$$

The expected number of particles in the system decays globally at rate  $\lambda_1$  towards the non-zero limiting value

$$\bar{\rho}_\infty(x) := \alpha(x)^{-1} b_0 u_0(x) \int_0^1 \alpha(y) v_0(y) dy = e^{-\sigma x} \int_0^1 e^{\sigma y} m(y) dy.$$

The BD transformed process therefore admits an integrable Yaglom limit  $\bar{q}_\infty$ , solution to  $-\tilde{G}^*(\bar{q}_\infty) = \lambda(y) \bar{q}_\infty$  or  $-\tilde{G}^*(\bar{q}_\infty) = 0$ . With  $v_0(y) = m(y)$ , the first eigenvector of  $-\tilde{G}^*$  associated to the smallest positive eigenvalue  $\lambda_0 = 0$  (the equilibrium density of the WF diffusion with mutations),  $\bar{q}_\infty$  is of the product form

$$(86) \quad \bar{q}_\infty(y) = \frac{e^{\sigma y} m(y)}{\int_0^1 e^{\sigma y} m(y) dy}.$$

This explicit limiting probability  $\bar{q}_\infty$  is the Yaglom limiting average presence density at  $(t, y)$  for the BD system of particles (it is also the ground state for  $\tilde{G}^*$ ).

There is also a natural eigenvector  $\bar{\phi}_\infty$  of the backward operator  $-\bar{G}$ , satisfying  $-\bar{G}(\bar{\phi}_\infty) = 0$  (the ground state for  $\bar{G}$ ). It is explicitly here

$$(87) \quad \bar{\phi}_\infty(x) = \frac{1}{\alpha(x)} u_0(x) \int_0^1 e^{\sigma y} m(y) dy = e^{-\sigma x} \int_0^1 e^{\sigma y} m(y) dy.$$

Both operators  $\bar{G}(\cdot)$  and its adjoint are again critical. The constant  $\lambda_0 = 0$  is the new generalized principal eigenvalue; The eigen-functions  $(\bar{\phi}_\infty, \bar{q}_\infty)$  are the new associated ground states.

We note that we have the  $L^1$ -product property

$$\int_0^1 u_0(x) v_0(x) dx = \int_0^1 \bar{\phi}_\infty(x) \bar{q}_\infty(x) dx = 1 < \infty.$$

Clearly the ground states of  $-\tilde{G}^*$  and  $-\bar{G}$  are defined up to arbitrary multiplicative constants. Note that we chose these constants in such a way that  $\int_0^1 \bar{q}_\infty(y) dy = 1$  and  $\int_0^1 \bar{\phi}_\infty(x) \bar{q}_\infty(x) dx = 1$ .

With  $p_m(x) = \mathbf{P}(M(x) = m)$ , let

$$\kappa(x) = \sum_{m \geq 2} m(m-1) p_m(x) = 2p_2(x).$$

Because  $p_2(x)$  is a degree two polynomial in  $x$ , we have the condition:

$$(88) \quad \int_0^1 \kappa(x) \bar{\phi}_\infty(x) \bar{q}_\infty(x) dx = 2 \int_0^1 p_2(x) u_0(x) v_0(x) dx < \infty.$$

We conclude (following [1] and [2]) that, as a result of the condition (88) being trivially satisfied, global extinction holds critically, in the following sense:

(i)  $\mathbf{P}(N_t(x) = 0) \xrightarrow{t \rightarrow \infty} 1$ , uniformly in  $x$ .

(ii) Let  $\mu_t = \sum_{n=1}^{N_t(\cdot)} \delta_{x_t^{(n)}}$ , with  $\mu_t(\psi) = \sum_{n=1}^{N_t(\cdot)} \psi(x_t^{(n)})$ .

There exists a finite positive constant :

$$\mu = \frac{1}{2t} \int_0^1 \mathbf{E}_x \left[ \mu_t(\phi)^2 - \mu_t(\phi^2) \right] \bar{q}_\infty(x) dx = \frac{1}{2t} \mathbf{E}_{\bar{q}_\infty} \left[ \mu_t(\phi)^2 - \mu_t(\phi^2) \right]$$

such that  $t[1 - \mathbf{P}(N_t(x) = 0)] \xrightarrow[t \rightarrow \infty]{} \mu^{-1} \bar{\phi}_\infty(x)$ , uniformly in  $x$ .

(iii) For all bounded measurable function  $\psi$  on  $I$  :

$$\frac{1}{t} \mathbf{E} \left[ \sum_{n=1}^{N_t(x)} \psi(\tilde{x}_t^{(n)}) \mid N_t(x) > 0 \right] \xrightarrow[t \rightarrow \infty]{} \mu \int_{(0,1)} \psi(y) \bar{q}_\infty(y) dy.$$

From (i), it is clear that the process gets ultimately extinct with probability 1. In the trade-off between killing-branching and reflection at the boundaries, all particles get eventually absorbed but the global BD process turns out to be critical. Thus, the killing part of  $\lambda(x)$  is strong enough to avoid the explosion of the number of particles inside the unit interval, resulting in an overall critical process where global extinction still holds.

In the statement (ii),  $1 - \mathbf{P}(N_t(x) = 0) = \mathbf{P}(T(x) > t)$  where  $T(x)$  is the global extinction time of the particle system. The Pareto tails of  $T(x)$  decay like  $t^{-1}$ , thus algebraically slowly: the time till extinction in this critical model is much longer than in the previous neutral sub-critical case (with exponential tails). From (ii),  $\bar{\phi}_\infty(x)$  has again a natural interpretation in terms of the propensity of the particle system to survive to its extinction fate.

(iii) with  $\psi = 1$  reads  $\frac{1}{t} \mathbf{E}[N_t(x) \mid N_t(x) > 0] \xrightarrow[t \rightarrow \infty]{} \mu$  giving an interpretation of the constant  $\mu$ . The constant  $\mu$  is also ([2], page 287)

$$\mu = \frac{1}{2} \lambda_* \int_0^1 \kappa(x) \bar{\phi}_\infty(x)^2 \bar{q}_\infty(x) dx = \lambda_* \int_0^1 p_2(x) \bar{\phi}_\infty(x)^2 \bar{q}_\infty(x) dx < \infty$$

and so is explicitly available in our case.

The ground states of  $\bar{G} + \lambda_0$  and its adjoint are thus  $(\bar{\phi}_\infty, \bar{q}_\infty)$  and explicit here. It is also useful to consider the process whose infinitesimal generator is given by the Doob-transform

$$\bar{\phi}_\infty^{-1} \bar{G} (\bar{\phi}_\infty \cdot) = \bar{\phi}_\infty^{-1} (\tilde{G} + \lambda) (\bar{\phi}_\infty \cdot),$$

because product-criticality is preserved under this transformation. The ground states associated to this new operator and its dual are  $(1, \bar{\phi}_\infty \bar{q}_\infty)$ . Developing, we obtain a process whose infinitesimal generator is

$$\tilde{G} + \frac{\bar{\phi}_\infty'}{\bar{\phi}_\infty} g^2 \partial_x = G + \frac{u_0'}{u_0} g^2 \partial_x = G,$$

with no multiplicative part. The associated diffusion process is the starting point WF diffusion with mutations, which is positive recurrent and so its invariant measure  $\bar{\phi}_\infty \bar{q}_\infty = u_0 v_0 = m(y)$  is integrable.

**Remark.** The functional generating function  $u(x, t; z) = \mathbf{E} \left[ \prod_{n=1}^{N_t(x)} z^{\psi(\tilde{x}_t^{(n)})} \right]$  of the measure-valued branching particle system obeys now the nonlinear (quadratic) PDE:

$$\partial_t u(x, t; z) = \lambda_* \theta(x, u(x, t; z)) + \tilde{G}(u(x, t; z)); \quad u(x, 0; z) = z^{\psi(x)},$$

where  $\theta(x, z) = \mathbf{E} [z^{M(x)}] - z = (p_2(x)z^2 + p_0(x)) - z$  or

$$\theta(x, z) = (z - 1)(p_2(x)z - p_0(x))$$

is the shifted probability generating function of the branching law of  $M(x)$ .

If  $u(x, t) := \partial_z u(x, t; z)|_{z=1} = \mathbf{E} \left[ \sum_{n=1}^{N_t(x)} \psi(\tilde{x}_t^{(n)}) \right]$ , recalling  $\lambda(x) = \lambda_* (p_2(x) - p_0(x))$ ,  $u(x, t)$  obeys the linear backward PDE

$$\partial_t u(x, t) = \lambda(x) u(x, t) + \tilde{G}(u(x, t)); \quad u(x, 0) = \psi(x)$$

involving  $\bar{G}(\cdot) = \tilde{G}(\cdot) + \lambda(x)\cdot$ . It holds that

$$u(x, t) = \mathbf{E}_x \left( e^{\int_0^t \lambda(\tilde{x}_s) ds} \psi(\tilde{x}_t) \right). \diamond$$

## 20. KINGMAN COALESCENT AND DUALITY

Let us now briefly mention some basic facts if one looks at this process backward in time. The latter discrete space-time process can be extended while assuming that  $t \in \mathbb{Z}$ . Take then a sub-sample of size  $n$  from  $[N] := \{1, \dots, N\}$  at generation 0. Identify two individuals from  $[n]$  at each step if they share a common ancestor one generation backward in time. This defines an equivalence relation between two individuals from  $[n]$ . It is of interest to study the induced ancestral backward count process. Let then  $\hat{x}_t^{(N)} = \hat{x}_t^{(N)}(n)$  count the number of ancestors at generation  $t \in \mathbb{N}$ , backward in time, starting from  $\hat{x}_0^{(N)} = n \leq N$ . This backward counting process is again a discrete-time Markov chain (with state-space  $\{1, \dots, N\}$ ) whose lower-triangular transition matrix  $\hat{P}_{i,j}^{(N)}$  can easily be written down under our assumptions on  $\nu$ . The process  $\hat{x}_t^{(N)}$  thus shrinks by random amounts till it hits 1 which is an absorbing state. Of particular interest in  $\hat{P}_{i,j}^{(N)}$  is the coalescence probability  $c_N := \hat{P}_{2,1}^{(N)} = 1/N_e$ . It is the probability that two individuals chosen at random from some generation have a common parent. The probability  $d_N := \hat{P}_{3,1}^{(N)}$  that 3 individuals chosen at random from some generation share a common parent is also relevant. For scaling limits  $N \rightarrow \infty$ , whether  $c_N \rightarrow 0$  or not and whether triple mergers are asymptotically negligible compared to double ones ( $\frac{d_N}{c_N} \rightarrow 0$ ) or not ( $\frac{d_N}{c_N} \not\rightarrow 0$ ) is important, (Sagitov). Under our assumptions on  $\nu$ , both  $c_N \rightarrow 0$  and  $\frac{d_N}{c_N} \rightarrow 0$ , leading to the well-known conclusion that as  $N \rightarrow \infty$

$$\hat{x}_{\lfloor t/c_N \rfloor}^{(N)}(n) \xrightarrow{\mathcal{D}} \hat{x}_t, \quad \hat{x}_0 = n, \quad t \in \mathbb{R}_+,$$

where  $\hat{x}_t$  is the continuous-time Kingman coalescent. This process is a Markov one with semi-infinite lower-triangular rate matrix:  $\hat{Q}_{i,i-1} = -\frac{1}{2}i(i-1)$ ,  $\hat{Q}_{i,i} = \frac{1}{2}i(i-1)$  and  $\hat{Q}_{i,j} = 0$  if  $j \neq \{i-1, i\}$ . The effective population size  $N_e := 1/c_N$  fixes the time scale of the time-scaled process  $\hat{x}_t$ . For the Kingman coalescent tree, only binary collisions (mergers) can occur and never simultaneously. Of interest,

among other things on this coalescent, are the time to most recent common ancestor:  $\hat{\tau}_{n,1} := \inf(t \in \mathbb{R}_+ : \hat{x}_t = 1 \mid \hat{x}_0 = n)$ , the length of the coalescent tree, the number of collisions till  $\hat{\tau}_{n,1}$ ... (see Tavaré for the computation of the law of these variables and various asymptotics as  $n \rightarrow \infty$ ).

The scaled continuous-time Wright-Fisher and Kingman processes are well-known to be dual with respect to one another in the sense that:

$$(89) \quad \mathbf{E}_x(x_t^n) = \mathbf{E}_n(x^{\hat{x}_t}), \text{ for all } (n, t) \in \mathbb{N}_+ \times \mathbb{R}_+, x \in [0, 1].$$

For instance, from the knowledge of the  $n$ -th moment of  $x_t$  started at  $x$ , one can obtain the probability generating function (pgf)  $\mathbf{E}_n(x^{\hat{x}_t})$  of  $\hat{x}_t$  started at  $\hat{x}_0 = n$ .

#### REFERENCES

- [1] Asmussen, S.; Hering, H. Strong limit theorems for general supercritical branching processes with applications to branching diffusions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 36, no. 3, 195-212, 1976.
- [2] Asmussen, S.; Hering, H. Some modified branching diffusion models. *Math. Biosci.* 35, no. 3-4, 281-299, 1977.
- [3] Crow, J.; Kimura, M. *An introduction to population genetics theory*. Harper & Row, Publishers, New York-London 1970.
- [4] Crow, J.; Kimura, M. Some genetic problems in natural populations. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954-1955*, vol. IV, pp. 1-22. University of California Press, Berkeley and Los Angeles, 1956.
- [5] Durrett, R. *Probability models for DNA sequence evolution*. Second edition. Probability and its Applications (New York). Springer, New York, 2008.
- [6] Dynkin, E. B. *Markov processes*. Vols. I, II. Translated with the authorization and assistance of the author by J. Fabius, V. Greenberg, A. Maitra, G. Majone. Die Grundlehren der Mathematischen Wissenschaften, Bände 121, 122 Academic Press Inc., Publishers, New York; Springer-Verlag, Berlin-Göttingen-Heidelberg 1965 Vol. I: xii+365 pp.; Vol. II: viii+274 pp.
- [7] Ethier, S. N.; Kurtz, T. G. *Markov processes. Characterization and convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986.
- [8] Ewens, W. J. *Mathematical population genetics. I. Theoretical introduction*. Second edition. Interdisciplinary Applied Mathematics, 27. Springer-Verlag, New York, 2004.
- [9] Gillespie, J. H. *The Causes of Molecular Evolution*. New York and Oxford: Oxford University Press, 1991.
- [10] Huillet, T. On Wright-Fisher diffusion and its relatives. *J. Stat. Mech., Th. and Exp.* P11006, vol.11, 2007.
- [11] Itô, K. On stochastic differential equations. *Mem. Am. Math. Soc.* No. 4 (1951).
- [12] Itô, K.; McKean, H. P. *Diffusion processes and their sample paths*. Die Grundlehren der Mathematischen Wissenschaften, Band 125 Academic Press, Inc., Publishers, New York; Springer-Verlag, Berlin-New York 1965.
- [13] Karlin, S.; Taylor, H. M. *A second course in stochastic processes*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1981.
- [14] Kimura, M. Diffusion Models in Population Genetics. *J. Appl. Prob.* 1, 177-232, 1964.
- [15] Kimura, M. Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harbor Symposia on Quantitative Biology* 20, 33-53, 1955. Also, contribution 114 of the National Institute of Genetics, Mishima-shi, Japan.
- [16] Kimura, M. On the probability of fixation of mutant genes in a population. *Genetics*, 47, 713-19, 1962.
- [17] Kolmogorov, A., Petrovsky, I.; Piscounov, N. Étude de l'équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique. *Moscov Univ. Bull. Math.* 1, pp. 125, 1937.
- [18] Lambert, A. Population dynamics and random genealogies. *Stochastic Models* 24, suppl. 1, 45-163, (2008).

- [19] Mandl, P. *Analytical treatment of one-dimensional Markov processes*. Die Grundlehren der mathematischen Wissenschaften, Band 151 Academia Publishing House of the Czechoslovak Academy of Sciences, Prague; Springer-Verlag New York Inc., New York 1968.
- [20] Mano, S. Duality, ancestral and diffusion processes in models with selection. *Theoretical Population Biology* 75, 164-175, 2009.
- [21] Maruyama, T. *Stochastic problems in population genetics*. Lecture Notes in Biomathematics, 17. Springer-Verlag, Berlin-New York, 1977.
- [22] Pinsky, R. G. *Positive harmonic functions and diffusion*. Cambridge Studies in Advanced Mathematics, 45. Cambridge University Press, Cambridge, 1995.
- [23] Yaglom, A. M. Certain limit theorems of the theory of branching random processes. *Doklady Akad. Nauk SSSR (N.S.)* 56, 795–798, (1947).

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