Journal of Statistics: Advances in Theory and Applications
Volume 7, Number 2, 2012, Pages 85-154

DIFFUSION VERSUS JUMP PROCESSES ARISING AS SCALING LIMITS IN POPULATION GENETICS

THIERRY E. HUILLET

Laboratoire de Physique Théorique et Modélisation
CNRS-UMR 8089 et Université de Cergy-Pontoise
2 Avenue Adolphe Chauvin
F-95302, Cergy-Pontoise
France
e-mail: thierry.huillet@u-cergy.fr

Abstract

When the reproduction law of a discrete branching process preserving the total size N of a population is 'balanced', scaling limits of the forward and backward in time processes are known to be the Wright-Fisher diffusion and the Kingman coalescent.

When the reproduction law is 'unbalanced', depending on extreme reproduction events occurring either occasionally or systematically, then various forward and backward jump processes, either in continuous time or in discrete time arise as scaling limits in the large N limit. This is in sharp contrast with diffusion limits, whose sample paths are continuous. We study some aspects of these limiting jump processes both forward and backward, especially the discrete-time ones. In the forward in time approach, because the absorbing boundaries are not hit in finite time, the analysis of the models together with the conclusions, which can be drawn deviate significantly from the ones available in the diffusion context.

2010 Mathematics Subject Classification: 87.23.Cc, 02.50.Ey, 87.23.

Keywords and phrases: mutational and evolutionary processes (theory), population dynamics (theory), phylogeny (theory).

Received March 22, 2012

1. Introduction

The discrete Wright-Fisher (WF) model for bi-allelic haploid dynamics (and its diffusion limit) is at the heart of theoretical population genetics (see [28], [5], [26], [12], and [10] for instance). It describes the temporal evolution of the number of type 1 (allele 1) individuals in generation t among a population of fixed size N, subject to exchangeable reproduction laws. When the reproduction law of individuals is 'balanced' in some sense made precise, an appropriate space-time scaling gives rise in the large N limit to the WF diffusion model. When looking at the genealogy of this process backward in time, upon scaling time only, the Kingman coalescent pops in, see [23]. The Wright-Fisher diffusion and the Kingman coalescent in continuous time are dual processes in some sense. The WF diffusion on the unit interval is a transient martingale, which hits the boundaries $\{0, 1\}$ in finite time. It belongs to a class of well-studied one-dimensional absorbed diffusion process on an interval, see [25]. In this one-dimensional diffusion context (possibly with additional drifts killing the martingale property), it is of common use to study various positive additive functionals $\alpha(x)$ of the process started at x, the expected time to absorption being one of them. The Green function $\mathfrak{g}(x, y)$, which is the expected local time of the process at y given it started at x is another one, which is the most important, as any $\alpha(x)$ can be expressed in terms of an integral against $\mathfrak{g}$. It is also of interest to look at, say, the WF diffusion process conditioned on its non-absorption (whose limit law is the uniform Yaglom quasi-stationary distribution). Thanks to known spectral information on the WF diffusion, this program can be achieved, to a large extent. It makes use of the well-known fact that the Kolmogorov backward and forward elliptic generators of the WF diffusion have a purely discrete (atomic) spectrum.

Then some other questions pertaining to conditionings become relevant: What is the WF diffusion conditioned on extinction or fixation, for instance? or what is the WF diffusion conditioned on being killed when it quits some state y for the last time? It turns out that the tool needed to

formulate and understand these questions is the Doob transform, based on various (super)-harmonic additive functionals α. The Doob transforms allow to modify the sample paths $x \to y$ of the original process while favouring large values of the ratio $\alpha(y)/\alpha(x)$. The transformed process is obtained from the original one while adding a drift term to it and while possibly killing the drifted new diffusion at some additional killing rate. It gives rise to a large number of questions of interest in population genetics, such as the expected fixation time of a WF diffusion conditioned on fixation or the age of a mutant currently observed at some frequency y, or the Yaglom limits of the transformed process conditioned on its current survival. To answer such questions, it is relevant to consider the evaluation of additive functionals for the transformed process. We develop and illustrate some of these ideas in the (WF) diffusion context and we refer to [14] for additional examples and details.

When dealing with discrete (size N) Markov models with 'unbalanced' reproduction laws, the point of view turns out to be quite different. By 'unbalanced', we mean that one individual in the discrete model is allowed to give birth to a 'significant' number of individuals among the N possible ones of the next generation, the others adapting their descendance so as to fulfill the global conservation of the total number N. It turns out that there are two possible 'unbalanced' models with such extreme reproduction events: One is occasional extreme events and the other is systematic extreme events, when the very productive individual produces a random fraction U of the whole population at each step. When the reproduction law is 'unbalanced', depending on extreme reproduction events occurring occasionally or systematically, then various forward and backward jump processes, either in continuous time or in discrete time arise as scaling limits in the large N limit. We give some details. When running time backward, it was shown in [17] (developing some ideas first discussed in [7]), that these limiting processes were continuous and discrete-time Λ-coalescents [32], respectively. We give some additional information on these processes, especially in the discrete-time case. All scaled processes depend on the measure Λ.

We will also consider the forward in time scaled jump processes on the unit interval. Due to extreme events, the scaled processes are no longer of diffusion kind with continuous sample paths, rather they are jump processes on the interval. We shall mainly focus on the discrete-time version (arising then when systematic extreme events occur). In this latter case, we show that such processes are again transient, but that in sharp contrast with the WF diffusion, the time to absorption occurs in infinite time, with probability 1. There is indeed a positive probability that this motion only makes move to the right or to the left, so that there is a positive probability never to visit a neighbourhood of y starting from any x inside the interval: Such processes are thus transient for any choice of the measure Λ. They eventually end up their life at either boundaries. The scaled forward process in discrete time depends on this measure Λ in the following way: The very productive individual produces a random fraction U of the whole population and the law of U is $u^{-2}\Lambda(du)$.

In the last section, we study in some detail the particular scaled discrete-time forward process when U is assumed uniform. We call it the special case and because of its 'simplicity', the analysis can be carried out. We give its Kolmogorov backward Fredholm generator L and its adjoint L^*. Because, we deal here with a jump process, the generators are no longer local second-order differential ones (as in the diffusion case), rather they are integral Fredholm operators of a special singular kind. We investigate some of their spectral properties. The operator L is not compact and it has now a point spectrum, which is a whole closed sub-disk of the unit disk. However, L leaves invariant some polynomials associated to a discrete real subset of the point spectrum, akin to the eigenvalues of the transition matrix of its associated discrete-time limiting coalescent. As for its WF diffusion process counterpart, the Green function is a key quantity to evaluate additive functionals of the new process under study. We compute it and give some examples of applications. Then, following the path borrowed in the WF diffusion context, we investigate the questions of conditionings via Doob transform in the context of the special process. Finally, we address the questions of integrating drifts either due to mutation or selection, deviating thereby from neutrality.

2. The Wright-Fisher and Related Models

In this section, we briefly review some basic facts concerning the Wright-Fisher (WF) diffusion as a scaling limit of 'balanced' reproduction laws as the size of the population goes to infinity.

2.1. The neutral Wright-Fisher model

We first consider a discrete-time Galton-Watson branching process preserving the total number of individuals at each generation. It can be defined as follows. Start with a population of N individuals at generation 0. Each individual can then give birth to a random number ξ_n of individuals, $n = 1, \ldots, N$, where the ξs are mutually independent and identically distributed (iid). Because a parent dies in the process of giving birth, we can interpret the event $\xi_n = 0$ as the death of individual n. In order to fulfill the requirement that the population size remains constant over time, we can assume a conditional Cannings reproduction law (see [3], [4]), that is: The first-generation random offspring numbers is $\boldsymbol{\nu} := (\nu_1, \ldots, \nu_N)$, whose law is obtained as $\boldsymbol{\nu} = (\xi_1, \ldots, \xi_N \mid \sum \xi_n = N)$, while conditioning N iid random variables on summing to N (we will come back later to the notion of a Cannings model for $\boldsymbol{\nu}$ while considering a very different class). Subsequent iterations of this reproduction law are applied independently. Would the ξs be Poisson-distributed, for example, regardless of their common means, $\boldsymbol{\nu}$ would have the joint exchangeable polynomial distribution on the simplex $|\mathbf{i}| := N$, where $\mathbf{i} := (k_1, \ldots, k_N)$

$$\mathbf{P}(\boldsymbol{\nu} = \mathbf{i}) = \frac{N! \cdot N^{-N}}{\prod_{n=1}^{N} i_n!}. \tag{1}$$

Let $x_t^{(N)}(n)$ denote the offspring number of the n first individuals at discrete generation $t \in \mathbb{N}_0$ corresponding to (say) allele A_1 or type 1 individuals. $N - x_t^{(N)}(n)$ is therefore the offspring number at t of the

$N-n$ type 2 original individuals[1]. $x_t^{(N)}(n)$ is a discrete-time homogeneous Markov chain on $\{0, \ldots, N\}$ with transition probability $\mathbf{P}(\nu_1 + \ldots + \nu_i = j)$, so when the ξs are Poisson, with

$$\mathbf{P}\left(x_{t+1}^{(N)}(n) = j \,|\, x_t^{(N)} = i\right) = \binom{N}{j}\left(\frac{i}{N}\right)^j\left(1-\frac{i}{N}\right)^{N-j}.$$

The discrete Wright-Fisher process $x_t^{(N)}$ clearly is a martingale with absorbing states $\{0, N\}$, which are being hit in finite time with probability 1.

With $B(N, p) \overset{d}{\sim} \text{bin}(N, p)$ a binomial random variable (r.v.), the dynamics of $x_t^{(N)}$ is

$$x_{t+1}^{(N)} = B\left(N, \frac{x_t^{(N)}}{N}\right), \quad x_0^{(N)} = n.$$

Let $n = \lfloor Nx \rfloor$ for some $x \in (0, 1)$. The dynamics of the continuous space-time re-scaled process $x_{\lfloor Nt \rfloor}^{(N)}(\lfloor Nx \rfloor)/N$, $t \in \mathbb{R}_+$ can be approximated for large N, to the leading term in N^{-1}, by a Wright-Fisher-Itô diffusion on $[0, 1]$ driven by standard Brownian motion w_t (the random genetic drift)

$$dx_t = \sqrt{x_t(1-x_t)}\, dw_t, \quad x_0 = x, \tag{2}$$

where x_t is the diffusion martingale approximation of the offspring frequency at generation $\lfloor Nt \rfloor$ (the integral part of Nt) when the initial

[1] This model and its forthcoming Wright-Fisher scaling limit typically accounts for the intrinsic temporal fluctuations of the one allele count or frequency in a simple bi-allelic population. Recently, see [1], an interesting political interpretation of this model was given in terms of the bi-partition of some population with respect to two political beliefs.

frequency is x. In the scaling limit process, time t is thus measured in units of N. The global stopping time of $x_t(x)$ is $\tau_x = \tau_{x,0} \wedge \tau_{x,1}$, where $\tau_{x,0}$ is the extinction time and $\tau_{x,1}$ is the fixation time of x_t when the process starts in x. The boundaries of x_t are absorbing and they are hit in finite time with probability 1.

This well-known result (see [10], for example), which is valid when the reproduction law $\boldsymbol{\nu}$ is built from ξs, which are iid Poisson distributed, extends to a much larger class of $\boldsymbol{\nu}$ also obtained from conditioning N iid discrete random variables ξ on summing to N. The only difference is that the time scaling should be $\lfloor N_e t \rfloor$ with $N_e = \rho N$ instead of simply N, for some $\rho > 0$ with possibly $\rho \neq 1$ (see [16] and [15], Theorem 3.2). Note that these models for $\boldsymbol{\nu}$ are balanced in that there is no individual, whose offspring number is statistically different from the one of the others.

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 $[N] := \{1, \ldots, N\}$ 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 $\{1, \ldots, N\}$), whose lower-triangular transition matrix $\hat{P}_{i,j}^{(N)}$ can easily be written down under our assumptions on $\boldsymbol{\nu}$. 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

$c_N := \widehat{P}_{2,1}^{(N)} = 1/N_e$. It is the probability that two individuals chosen at random from some generation have a common parent. The probability $d_N := \widehat{P}_{3,1}^{(N)}$ that 3 individuals chosen at random from some generation share a common parent is also relevant. For scaling limits $N \to \infty$, whether $c_N \to 0$ or not and whether triple mergers are asymptotically negligible compared to double ones ($\frac{d_N}{c_N} \to 0$) or not ($\frac{d_N}{c_N} \not\to 0$) is important, [33]. Under our assumptions on ν, both $c_N \to 0$ and $\frac{d_N}{c_N} \to 0$, leading to the well-known conclusion that as $N \to \infty$

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

where $\hat{x}_t$ is the continuous-time Kingman coalescent [23]. This process is a Markov one with semi-infinite lower-triangular rate matrix: $\widehat{Q}_{i,i-1} = -\frac{1}{2} i(i-1)$, $\widehat{Q}_{i,i} = \frac{1}{2} i(i-1)$, and $\widehat{Q}_{i,j} = 0$ 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: $\hat{\tau}_{n,1} := \inf(t \in \mathbb{R}_+ : \hat{x}_t = 1 \,|\, \hat{x}_0 = n)$, the length of the coalescent tree, the number of collisions till $\hat{\tau}_{n,1 \cdots}$ (see [34] for the computation of the law of these variables and various asymptotics as $n \to \infty$).

The scaled continuous-time Wright-Fisher and Kingman processes are well-known to be dual with respect to one another in the sense that (see [30] and [14], for example)

$$\mathbf{E}_x(x_t^n) = \mathbf{E}_n\left(x^{\hat{x}_t}\right), \text{ for all } (n, t) \in \mathbb{N}_+ \times \mathbb{R}_+, \quad x \in [0, 1]. \tag{3}$$

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

2.2. Non-neutral cases

The neutral case accounts for the so-called random genetic 'drift'. The presence of additional evolutionary 'forces' results typically in adding to the SDE (4) a true (non-random) drift. The two alleles Wright-Fisher models (with non-null drifts) have binomial transition probabilities $\mathrm{bin}(N,\ p_N)$

$$\mathbf{P}\left(x_{t+1}^{(N)}(n) = k' \mid x_t^{(N)}(n) = k\right) = \binom{N}{k'}\left(p_N\left(\frac{k}{N}\right)\right)^{k'}\left(1 - p_N\left(\frac{k}{N}\right)\right)^{N-k'},$$

where

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

is some state-dependent probability different from the identity x: This continuous mapping accounts for a deterministic evolutionary drift from allele A_1 to allele A_2 due to external evolutionary forces. For each t, we thus have

$$\mathbf{E}\left(x_{t+1}^{(N)}(n) \mid x_t^{(N)}(n) = k\right) = Np_N\left(\frac{k}{N}\right),$$

$$\sigma^2\left(x_{t+1}^{(N)}(n) \mid x_t^{(N)}(n) = k\right) = Np_N\left(\frac{k}{N}\right)\left(1 - p_N\left(\frac{k}{N}\right)\right),$$

and the martingale property is lost. $x_t^{(N)}(n)$ is also amenable to a diffusion approximation x_t as the scaling limit of $x_{\lfloor Nt \rfloor}^{(N)}(n) / N,\ t \in \mathbb{R}_+$ under suitable conditions on $p_N(x)$.

(i) Take for instance $p_N(x) = (1 - \pi_{2,N})x + \pi_{1,N}(1 - x)$, 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 $(N \cdot \pi_{1,N},\, N \cdot \pi_{2,N}) \underset{N\to\infty}{\to} (u_1,\, u_2)$, this leads after scaling to a Wright-Fisher diffusion model with an additional drift: $f(x) = u_1 - (u_1 + u_2)x$, involving positive mutations rates $(u_1,\, u_2)$. Thus, the Wright-Fisher diffusion with mutations is

$$dx_t = (u_1 - (u_1 + u_2)x_t)dt + \sqrt{x_t(1 - x_t)}dw_t, \quad x_0 = x.$$

(ii) 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 $N \cdot s_{i,N} \underset{N\to\infty}{\to} \sigma_i > 0$, $i = 1, 2$, leads, after scaling, to the WF model with selective logistic drift $f(x) = \sigma x(1 - x)$. Here $\sigma := \sigma_1 - \sigma_2$ is the selective advantage of allele A_1 over allele A_2. The drift term $f(x)$ is a large N approximation of the bias to neutrality: $N(p_N(x) - x)$. The Wright-Fisher diffusion with selection is

$$dx_t = \sigma x_t(1 - x_t)dt + \sqrt{x_t(1 - x_t)}dw_t, \tag{4}$$

with time t measured in units of N.

Like the neutral Wright-Fisher diffusion, (4) has two absorbing barriers. It tends to drift to the boundary $\{1\}$ (respectively $\{0\}$), if allele A_1 is selectively advantageous over A_2 : $\sigma_1 > \sigma_2$ (respectively, $\sigma_1 < \sigma_2$) : if $\sigma > 0$ (respectively $\sigma < 0$), the fixation probability at $\{1\}$, which is known to be [21]

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

increases (decreases) with σ taking larger (smaller) values.

3. Diffusions on $[0, 1]$

From now on, we discuss some general facts about diffusion processes on the unit interval, the Wright-Fisher diffusion process (either neutral or with various drifts) being one of them that one should keep in the background.

3.1. Kolmogorov backward equation

Let w_t denote the standard Brownian motion. Consider the Itô diffusion process on $[0, 1]$

$$dx_t = f(x_t)dt + g(x_t)dw_t, \quad x_0 = x \in (0, 1), \tag{5}$$

where we assume $g(0) = g(1) = 0$ (see [25]). The Kolmogorov-backward (KB) infinitesimal generator of (5) is

$$G = f(x)\partial_x + \frac{1}{2}g^2(x)\partial_x^2.$$

The quantity $u := u(x, t) = \mathbf{E}\psi(x_{t\wedge\tau_x})$ satisfies Kolmogorov-backward equation (KBE)

$$\partial_t u = G(u); \quad u(x, 0) = \psi(x). \tag{6}$$

In the definition of u, $t \wedge \tau_x := \inf(t, \tau_x)$, where $\tau_x = \tau_{x,0} \wedge \tau_{x,1}$ is the random time at which, the process should possibly be stopped, given the process was started at x. τ_x is thus the adapted absorption time, governed by the type of boundaries, which $\{0, 1\}$ are to x_t. The KBE equation may not have unique solutions, unless one specifies the conditions at the boundaries $\{0, 1\}$. For 1-dimensional diffusions as in (5) on $[0, 1]$, the boundaries $\{0, 1\}$ are of two types: Either accessible or inaccessible. Accessible boundaries are either regular or exit boundaries, whereas inaccessible boundaries are either entrance or natural boundaries. Integrability criteria based on both the scale function and the speed measure are essential in the classification of boundaries due to Feller [11]. We now define these quantities.

3.2. Scale function and speed measure

When dealing with such diffusion processes, one introduces the G-harmonic coordinate $\varphi \in C^2$, i.e., satisfying $G(\varphi) = 0$. It is

$$\varphi'(y) = e^{-2\int^y \frac{f(z)}{g^2(z)}dz} > 0 \text{ and } \varphi(x) = \int^x e^{-2\int_{y_0}^y \frac{f(z)}{g^2(z)}dz} dy. \tag{7}$$

The function φ kills the drift f of x_t in that: $y_t := \varphi(x_t)$ is a martingale obeying the new drift-less stochastic differential equation (SDE)

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

Also of interest is the speed density: $m(y) = 1/\left(g^2\varphi'\right)(y)$. The speed density m is in the kernel of the adjoint KB generator

$$G^*(\cdot) = -\partial_y(f(y)\cdot) + \frac{1}{2}\partial_y^2(g^2(y)\cdot\ ), \tag{8}$$

so $G^*(m) = 0$.

Defining now the random time change: $t \to \theta_t$, with inverse: $\theta \to t_\theta$ defined by $\theta_{t_\theta} = \theta$ and

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

the time-changed process $(w_\theta := y_{t_\theta};\ \theta \geq 0)$ is easily seen to coincide with the standard Brownian motion. Both the scale function $\varphi(x)$ and the speed measure $d\mu := m(y)\cdot dy$ are thus essential ingredients to reduce the original stochastic process x_t to standard Brownian motion w_θ. The KB infinitesimal generator G can be written in Feller form

$$G(\cdot) = \frac{1}{2}\frac{d}{d\mu}\left(\frac{d}{d\varphi}\cdot\right). \tag{9}$$

Examples (From population genetics)**.** See [5], [26], [12], and [10].

• $f(x) = 0$ and $g^2(x) = x(1 - x)$. This is the neutral WF model already encountered. Here, $\varphi(x) = x$ and $m(y) = [y(1 - y)]^{-1}$. The speed measure is not integrable.

• With $u_1, u_2 > 0$, $f(x) = u_1 - (u_1 + u_2)x$ and $g^2(x) = x(1 - x)$. This is the WF diffusion with mutation rates u_1, u_2. The drift vanishes when $x = u_1 / (u_1 + u_2)$, which is an attracting point for the dynamics.

Here, $\varphi'(y) = y^{-2u_1}(1 - y)^{-2u_2}$, $\varphi(x) = \int^x y^{-2u_1}(1 - y)^{-2u_2}\,dy$, with $\varphi(0) = -\infty$ and $\varphi(1) = +\infty$, if $u_1, u_2 > 1/2$. The speed density $m(y) \propto y^{2u_1 - 1}(1 - y)^{2u_2 - 1}$ is always integrable.

• With $\sigma \in \mathbb{R}$, consider a diffusion process 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. Here, $\varphi(x) \propto e^{-2\sigma x}$ and $m(y) \propto [y(1 - y)]^{-1} e^{2\sigma y}$ is not integrable. σ is the selection or fitness differential parameter.

• The WF model with $f(x) = \sigma x(1 - x) + u_1 - (u_1 + u_2)x$ and $g^2(x) = x(1 - x)$ is WF model with mutations and selection parameters $(u_1, u_2; \sigma)$. Here, $\varphi(x) = \int^x e^{-2\sigma y} y^{-2u_1}(1 - y)^{-2u_2}\,dy$. The speed density $m(y) \propto y^{2u_1 - 1}(1 - y)^{2u_2 - 1} e^{2\sigma y}$ is not integrable. □

3.3. Transition sub-probability density and Yaglom limits

Assume $\tau_x = \tau_{x,0} \wedge \tau_{x,1} < \infty$ with probability one (the boundaries are absorbing). Let $f(x)$ and $g(x)$ be differentiable in $(0, 1)$. Let $p(x; t, y)$ stand for the transition probability density of $x_{t \wedge \tau_x}$ at y, given $x_0 = x$. Then $p := p(x; t, y)$ is the smallest solution to the Kolmogorov-forward equation (KFE)

$$\partial_t p = G^*(p), \quad p(x;\, 0,\, y) = \delta_y(x), \tag{10}$$

with $G^*(\cdot)$, the adjoint of G, defined in (8).

The density $p(x;\, t,\, y)$ is reversible with respect to the speed density m in the sense that

$$m(x)p(x;\, t,\, y) = m(y)p(y;\, t,\, x), \quad 0 < x,\, y < 1, \tag{11}$$

with $m(y)$ satisfying $G^*(m) = 0$. The speed measure is a Gibbs measure with density: $m(y) \propto \frac{1}{g^2(y)} e^{-U(y)}$ associated to the potential function $U(y)$ such that

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

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

Under our assumption on τ_x, $p(x;\, t,\, y)$ is a sub-probability density, losing mass at the boundaries. Let $\rho_t(x) := \int_{(0,1)} p(x;\, t,\, y)dy$; then $\rho_t(x) = \mathbf{P}(\tau_x > t)$ is the tail distribution of the stopping time τ_x. This quantity obeys

$$\partial_t \rho_t(x) = G(\rho_t(x)), \text{ with } \rho_0(x) = \mathbf{1}_{(0,1)}(x).$$

Whenever $p(x;\, t,\, y)$ is a sub-probability, while normalizing, define $q(x;\, t,\, y) := p(x;\, t,\, y) / \rho_t(x)$, now with total mass 1 for each t. It holds that

$$\partial_t q = -\partial_t \rho_t(x) / \rho_t(x) \cdot q + G^*(q), \quad q(x;\, 0,\, y) = \delta_y(x), \tag{12}$$

where $-\partial_t \rho_t(x) / \rho_t(x) > 0$ 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 at the boundaries.

When the boundaries of x_t are absorbing (and under our assumption that $g(0) = g(1) = 0$), the spectra of $-G$ and $-G^*$ are discrete or atomic: There exist non-negative 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 $-G^*(v_k) = \lambda_k v_k$ and $-G(y_k) = \lambda_k u_k$. Note that $\lambda_1 = 0$. 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 expansion of $p(x; t, y)$ is

$$p(x; t, y) = \sum_{k\geq 2} b_k e^{-\lambda_k t} u_k(x) v_k(y). \tag{13}$$

Let $\lambda_2 > \lambda_1 = 0$ be the smallest non-null eigenvalue of the infinitesimal generator $-G^*$ (and of $-G$). With $b_2' := b_2 \int_{(0,1)} v_2(y)dy$, using (13), we have

$$e^{\lambda_2 t} \rho_t(x) \underset{t\to\infty}{\to} b_2' u_2(x),$$

and therefore τ_x is tail-equivalent to an exponential distribution with rate λ_2. The right-hand-side term in the latter limit has a natural interpretation in terms of the propensity of x_t to survive to its ultimate fate of being absorbed (the so-called reproductive value in demography). Note that $\rho_t(x)$ admits the expansion

$$\rho_t(x) = \sum_{k\geq 2} b_k e^{-\lambda_k t} u_k(x) \int_{(0,1)} v_k(y)dy,$$

and that so does therefore the mean of τ_x

$$\mathbf{E}(\tau_x) = \int_0^\infty \rho_t(x)dt = \sum_{k\geq 2} b_k \lambda_k^{-1} u_k(x) \int_{(0,1)} v_k(y)dy. \tag{14}$$

Clearly, $-\frac{1}{t}\log \rho_t(x) \underset{t\to\infty}{\to} \lambda_2$ and by L'Hospital rule therefore $-\partial_t \rho_t(x) / \rho_t(x) \underset{t\to\infty}{\to} \lambda_2$. Putting $\partial_t q = 0$ in the evolution equation (12) of q, independently of the initial condition x

$$q(x;\, t,\, y) \underset{t\to\infty}{\to} q_\infty(y) \propto v_2(y), \tag{15}$$

with v_2 the eigenvector of $-G^*$ associated to λ_2, satisfying $-G^*(v_2) = \lambda_2 v_2$.

The limiting probability density $q_\infty(y) = v_2(y) / \int_{(0,1)} v_2(y)dy$ is called the *Yaglom limit law* of x_t conditioned on being currently alive at time t. Note that, due to the orthogonality relations between the v_ks and the u_ks

$$\mathbf{P}_{q_\infty}(\tau > t) = \int_{(0,1)} q_\infty(x) \int_{(0,1)} p(x;\, t,\, y)dy = e^{-\lambda_2 t}. \tag{16}$$

Would the process x_t be started with the Yaglom limit law q_∞ (the quasi-stationary distribution), its absorption time τ would be exactly exponentially distributed with rate λ_2.

Example. L^2 theory and the neutral Wright-Fisher diffusion.

The degree-k Gegenbauer polynomials constitute a system of eigenfunctions for the KB operator $G = \frac{1}{2} x(1-x)\partial_x^2$ with the eigenvalues $\lambda_k = k(k-1)/2,\ k \geq 1,$ thus with $-G(u_k(x)) = \lambda_k u_k(x)$. In particular, $u_1(x) = x,\ u_2(x) = x - x^2,\ u_3(x) = x - 3x^2 + 2x^3,\ u_4(x) = x - 6x^2 + 10x^3 - 5x^4, \ldots$.

The eigenfunctions for the same eigenvalues of the KF operator $G^*(\cdot) = \frac{1}{2}\partial_x^2[y(1-y)\cdot]$ are given by $v_k(y) = m(y) \cdot u_k(y),\ k \geq 1,$ where the Radon measure of weights $m(y)dy$ is the speed measure: $m(y)dy = \dfrac{dy}{y(1-y)}$. For instance, $v_1(y) = \dfrac{1}{1-y}$, $v_2(y) = 1,\ v_3(y) = 1 - 2y,$ $v_4(y) = 1 - 5y + 5y^2, \ldots$.

Although $\lambda_1 = 0$ really constitutes an eigenvalue, only $v_1(y)$ is not a polynomial. When $k \geq 2$, from their definition, the u_ks and the v_ks polynomials satisfy $v_k(y) = m(y) \cdot u_k(y)$.

We note that, $\langle v_j, u_k \rangle = \langle u_j, u_k \rangle_m = 0$ if $j \neq k$ and so the system $u_k(x)$; $k \geq 2$ is a complete orthogonal set of eigenvectors. Therefore, for any square-integrable function $\psi(x) \in L_2([0, 1], m(y)dy)$ admitting the decomposition $\psi(x) = \sum_{k\geq 2} c_k u_k(x)$ in the basis $u_k(x)$, $k \geq 2$

$$\mathbf{E}_x \psi(x_t) = \sum_{k\geq 2} c_k e^{-\lambda_k t} u_k(x), \text{ where } c_k = \frac{\langle \psi, u_k \rangle_m}{\langle v_k, u_k \rangle} = \frac{\int_{(0,1)} \psi(y) u_k(y) m(y) dy}{\int_{(0,1)} v_k(y) u_k(y) dy}.$$

This series expansion solves KBE: $\partial_t u = G(u)$; $u(x, 0) = \psi(x)$, where $u = u(x, t) := \mathbf{E}^x \psi(x_t)$. [2]

Similarly, we have the series expansion of the transition probability density

$$p(x; t, y) = \sum_{k\geq 2} b_k e^{-\lambda_k t} u_k(x) v_k(y), \text{ where } b_k = \frac{1}{\int_{(0,1)} v_k(y) u_k(y) dy},$$

solving the KFE of the WF model. This transition density is clearly reversible with respect to the speed density since for $0 < x, y < 1$

---

[2] Whenever $\psi(x) = x^n$ is a degree-n polynomial, it can be uniquely decomposed on the n first eigenpolynomials $u_k(x)$ and therefore $\mathbf{E}_x(x_t^n) = \sum_{k=2}^{n} c_k e^{-\lambda_k t} u_k(x)$ is exactly known. Using the duality relationship (3) with Kingman coalescent $\hat{x}_t$, its pgf (and so its law): $\mathbf{E}_n(x^{\hat{x}_t})$ follows in principle. Considering $[x]\mathbf{E}_n(x^{\hat{x}_t}) = \mathbf{P}_n(\hat{x}_t = 1)$, one obtains the probability that the time to most recent common ancestor of $\hat{x}_t$ with $\hat{x}_0 = n$ occurred before time t.

$$m(x)p(x;\, t,\, y) = m(y)p(y;\, t,\, x) = \sum_{k\geq 2} b_k e^{-\lambda_k t} v_k(x) v_k(y).$$

The functions $v_k(y)$, $k \geq 2$ are not probability densities because $v_k(y)$ is not even necessarily positive over $[0, 1]$. The decomposition of p is not a mixture. We have $\langle v_k,\, u_k \rangle = \|u_k\|^2_{2,m}$ the 2-norm for the weight function m. We notice that $\langle v_1,\, u_1 \rangle = \int_{(0,1)} \frac{y}{1-y}\, dy = \infty$ so that $c_1 = b_1 = 0$; although $\lambda_1 = 0$ is indeed an eigenvalue, the above sums should be started at $k = 2$ (expressing the lack of an invariant measure for the WF model as a result of explosion and mass loss of the density p at the boundaries).

For the neutral Wright-Fisher diffusion, $\lambda_2 = 1$ with $v_2 \equiv 1$. The Yaglom limit $q_\infty(y)$ in this case is thus the uniform measure. □

3.4. Additive and multiplicative functionals along sample paths

Let x_t as from (5) on $[0, 1]$, with both endpoints $\{0, 1\}$ absorbing (exit). This process is transient. We wish to evaluate non-negative additive functionals of the type

$$\alpha(x) = \mathbf{E}\left(\int_0^{\tau_x} c(x_s)ds + d(x_{\tau_x}) \right), \tag{17}$$

where the functions c and d are both assumed non-negative. Thus, $\alpha(x) > 0$ in $(0, 1)$ solves the Dirichlet problem

$$-G(\alpha) = c, \text{ if } x \in (0, 1),$$

$$\alpha = d, \text{ if } x \in \{0, 1\}.$$

Examples. (1) Let $y \in (0, 1)$. Let $\alpha(x) = \mathbf{E}(\int_0^{\tau_x} \delta_y(x_s)ds)$ be the mean value of the local time $l_x(y) := \int_0^{\tau_x} \delta_y(x_s)ds$ of x_t at y, starting from x. Then $\alpha =: \mathfrak{g}(x, y) = \int_0^\infty p(x;\, s,\, y)ds$ is the Green function, solution to

$$-G(\mathfrak{g}) = \delta_y(x), \text{ if } x \in (0, 1),$$

$$\mathfrak{g} = 0, \text{ if } x \in \{0, 1\}.$$

Following ([19], p. 198), with $\alpha_0(x) := \mathbf{P}(\tau_{x,0} < \tau_{x,1})$ and $\alpha_1(x) = 1 - \alpha_0(x)$

$$\mathfrak{g}(x, y) = 2\alpha_0(x)m(y)(\varphi(y) - \varphi(0)), \text{ if } 0 \le y \le x,$$

$$\mathfrak{g}(x, y) = 2\alpha_0(x)m(y)(\varphi(1) - \varphi(y)), \text{ if } x < y \le 1. \tag{18}$$

Note that indeed $\mathfrak{g}$ vanishes at the boundaries: $\mathfrak{g}(0, y) = \mathfrak{g}(1, y) = 0$.

When dealing, for example, with the neutral WF diffusion

$$\mathfrak{g}(x, y) = 2\frac{1-x}{1-y}1_{0\le y\le x} + 2\frac{x}{y}1_{x<y\le 1}.$$

The Green function is useful to evaluate additive functionals $\alpha(x)$ such as the ones appearing in (17): The integral operator with respect to the Green kernel inverts the second order operator $-G$, leading to

$$\alpha(x) = \int_{(0,1)} \mathfrak{g}(x, y)c(y)dy + d(0) + (d(1) - d(0))x. \tag{19}$$

Note that indeed, $\alpha = d$ if $x \in \{0, 1\}$.

(2) If both $\{0, 1\}$ are exit boundaries, we wish to evaluate the probability that x_t first hits $[0, 1]$ (say) at 1, given $x_0 = x$. Choose then $c = 0$ and $d(x) = \mathbf{1}(x = 1)$. Then

$$\alpha =: \alpha_1(x) = \mathbf{P}(\tau_{x,1} < \tau_{x,0}).$$

$\alpha_1(x)$ is a G-harmonic solution to $G(\alpha_1) = 0$, with boundary conditions $\alpha_1(0) = 0$ and $\alpha_1(1) = 1$. From (19) and (7), we get

$$\alpha_1(x) = \frac{\varphi(x) - \varphi(0)}{\varphi(1) - \varphi(0)} = \int_0^x dy e^{-2\int_0^y \frac{f(z)}{g^2(z)}dz} \Big/ \int_{(0,1)} dy e^{-2\int_0^y \frac{f(z)}{g^2(z)}dz}.$$

Conversely, if $\alpha_0(x)$ is a G-harmonic function with boundary conditions $\alpha_0(0) = 1$ and $\alpha_0(1) = 0$, then

$$\alpha_0(x) = \mathbf{P}(\tau_{x,0} < \tau_{x,1}) = 1 - \alpha_1(x).$$

(3) Assume $c = 1$ and $d = 0$: here, $\alpha(x) = \mathbf{E}(\tau_x)$ is the mean time of absorption (average time spent in $[0, 1]$ before absorption), solution to

$$-G(\alpha) = 1, \text{ if } x \in (0, 1),$$

$$\alpha = 0, \text{ if } x \in \{0, 1\}.$$

From (19), $\alpha(x) = \int_{(0,1)} \mathfrak{g}(x, y)dy$, which is an alternative and much simpler expression of $\alpha(x) = \mathbf{E}(\tau_x)$ than the one displayed in (14) and which does not requires the knowledge of the full spectra of both $-G$ and $-G^*$.

As an illustrative example, if x_t is the WF diffusion, $\alpha(x)$ is easily seen to take the well-known entropy-like form

$$\alpha(x) = -2(x \log x + (1 - x)\log(1 - x)).$$

(4) Also of interest are the additive functionals

$$\alpha_\lambda(x) = \mathbf{E}\left(\int_0^{\tau_x} e^{-\lambda s} c(x_s)ds + d(x_{\tau_x})\right),$$

where c and d are again non-negative. $\alpha_\lambda(x) \geq 0$ solves the Dynkin problem

$$(\lambda I - G)(\alpha_\lambda) = c, \text{ if } x \in (0, 1),$$

$$\alpha_\lambda = d, \text{ if } x \in \{0, 1\},$$

involving the action of the resolvent operator $(\lambda I - G)^{-1}$ on c.

Whenever $c(x) = \delta_y(x)$, $d = 0$, then

$$\alpha_\lambda =: \mathfrak{g}_\lambda(x,\ y) = \mathbf{E}\left(\int_0^{\tau_x} e^{-\lambda s}\delta_y(x_s)ds\right) = \int_0^\infty e^{-\lambda s}p(x;\ s,\ y)ds,$$

is the λ-potential function, solution to

$$(\lambda I - G)(\mathfrak{g}_\lambda) = \delta_y(x), \text{ if } x \in (0,\ 1),$$

$$\mathfrak{g}_\lambda = 0, \text{ if } x \in \{0,\ 1\}.$$

The function $\mathfrak{g}_\lambda$ is the mathematical expectation of the exponentially damped local time at y, starting from x (the temporal Laplace transform of the transition probability density from x to y at t), with $\mathfrak{g}_0 = \mathfrak{g}$. Then

$$\alpha_\lambda(x) = \int_{(0,\ 1)} \mathfrak{g}_\lambda(x,\ y)c(y)dy + d(0) + (d(1) - d(0))x.$$

The λ-potential function is useful in the computation of the law of the first-passage time $\tau_{x,\,y}$ to y starting from x. Consider indeed the convolution formula

$$p(x;\ t,\ y) = \int_0^t \mathbf{P}(\tau_{x,\,y} \in ds)p(y;\ t - s,\ y).$$

Taking the temporal Laplace transform of both sides, we get the Laplace-Stieltjes transform of the law of $\tau_{x,\,y}$ as

$$\mathbf{E}(e^{-\lambda\tau_{x,\,y}}) = \frac{\mathfrak{g}_\lambda(x,\ y)}{\mathfrak{g}_\lambda(y,\ y)}. \tag{20}$$

Putting $\lambda = 0$, we have $\mathbf{P}(\tau_{x,\,y} < \infty) = \dfrac{\mathfrak{g}(x,\ y)}{\mathfrak{g}(y,\ y)} \in (0,\ 1)$ as a result of both terms in the ratio being finite and $x,\ y$ belonging to the same transience class of the process (under our assumptions that the boundaries are absorbing).

(5) (Multiplicative functionals). Multiplicative functionals are useful to evaluate the higher order moments of the additive functionals $\int_0^{\tau_x} c(x_s)ds$. Let indeed

$$\beta_\lambda(x) = \mathbf{E}\left(e^{\lambda\int_0^{\tau_x} c(x_s)ds}\right),$$

be now a multiplicative Kac functional. It is known (see [19]) that $\beta_\lambda(x)$ now solves

$$-G(\beta_\lambda(x)) = \lambda c(x)\beta_\lambda(x), \quad \beta_\lambda(0) = \beta_\lambda(1) = 1. \tag{21}$$

It holds that

$$\beta_\lambda(x) = 1 + \sum_{k\geq 1} \frac{\lambda^k}{k!}\alpha_k(x),$$

where $\alpha_k(x) = \mathbf{E}\left((\int_0^{\tau_x} c(x_s)ds)^k\right)$ are the k-moments of $\int_0^{\tau_x} c(x_s)ds$. Taking successive derivatives of (21) with respect to λ and putting $\lambda = 0$, one gets the recurrence for the moments

$$-G(\alpha_k(x)) = kc(x)\alpha_{k-1}(x), \quad k \geq 1.$$

Using the Green function $\mathfrak{g}$, with $x := x_0$

$$\alpha_k(x) = k!\int_{[0,1]^k}\prod_{l=1}^{k}\mathfrak{g}(x_{l-1}, x_l)c(x_l)\cdot dx_1\ldots dx_l,$$

giving an explicit expression of the moments.

If $c(x) = \delta_y(x)$, $\beta_{-\lambda}(x)$ is the Laplace-Stieltjes transform of the local time $l_x(y)$ of x_t at y, starting from x. In this case, we get

$$\alpha_k(x) = k!\int_{[0,1]^k}\prod_{l=1}^{k}\mathfrak{g}(x_{l-1}, x_l)\delta_y(x_l)\cdot dx_1\ldots dx_l$$

$$= k!\,\mathfrak{g}(x, y)\mathfrak{g}(x, y)^{k-1}.$$

Thus,

$$\mathbf{E}\left(e^{-\lambda l_x(y)}\right) = 1 - \frac{\lambda\mathfrak{g}(x, y)}{1+\lambda\mathfrak{g}(y, y)},$$

and, upon inverting this Laplace transform, recalling $\mathbf{P}(\tau_{x,y} < \infty) = \frac{\mathfrak{g}(x, y)}{\mathfrak{g}(y, y)}$,

$$\mathbf{P}(l_x(y) \in dt) = \mathbf{P}(\tau_{x,y} = \infty)\delta_0 + \mathbf{P}(\tau_{x,y} < \infty)\frac{1}{\mathfrak{g}(y, y)} e^{-t/\mathfrak{g}(y, y)}dt.$$

Given $\tau_{x,y} < \infty$, the local time $l_x(y)$ is exponentially distributed with mean $\mathfrak{g}(y, y)$ (see [29]). Without conditioning, the mean value of $l_x(y)$ is $\mathfrak{g}(x, y)$, with of course $\mathfrak{g}(x, y) < \mathfrak{g}(y, y)$. □

3.5. Doob transformation of paths

Consider x_t as in (5) with absorbing barriers. Let $p := p(x; t, y)$ be its transition probability density and let τ_x be its absorbing time at the boundaries.

Let $\alpha(x) := \mathbf{E}(\int_0^{\tau_x} c(x_s)ds + d(x_{\tau_x}))$ be a non-negative additive functional solving

$$-G(\alpha) = c, \text{ if } x \in (0, 1),$$

$$\alpha = d, \text{ if } x \in \{0, 1\}.$$

Recall the functions c and d are both chosen non-negative so that so is α is positive inside the unit interval (α is super-harmonic), possibly vanishing at the boundaries. Define a new transformed stochastic process, say $\overline{x}_t$, by its transition probability density

$$\overline{p}(x; t, y) = \frac{\alpha(y)}{\alpha(x)} p(x; t, y). \tag{22}$$

In this construction of $\overline{x}_t$ (relevant to a change of measure), sample paths $x \to y$ of x_t with large $\alpha(y)/\alpha(x)$ are favoured. This is a selection of paths procedure due to Doob (see [6]). The KFE for $\overline{p}$ is: $\partial_t \overline{p} = \overline{G}^*(\overline{p})$, with $p(x; 0, y) = \delta_y(x)$ and $\overline{G}^*(\overline{p}) := \alpha(y)G^*(\overline{p}/\alpha(y))$. The adjoint KBE of the transformed process is

$$\overline{G}(\cdot) = \frac{1}{\alpha(x)} G(\alpha(x)\,\cdot).$$

With $\alpha'(x) := d\alpha(x)/dx$ and $\widetilde{G}(\cdot) := \frac{\alpha'}{\alpha} g^2 \partial_x(\cdot) + G(\cdot)$,

$$\overline{G}(\cdot) = \frac{1}{\alpha} G(\alpha)\,\cdot + \widetilde{G}(\cdot) = -\frac{c}{\alpha}\,\cdot + \widetilde{G}(\cdot). \tag{23}$$

With $\widetilde{f}(x) := f(x) + \frac{\alpha'}{\alpha} g^2(x)$ the modified drift, the novel time-homogeneous SDE to consider is therefore

$$d\widetilde{x}_t = \widetilde{f}(\widetilde{x}_t)dt + g(\widetilde{x}_t)dw_t, \quad \widetilde{x}_0 = x \in (0, 1), \tag{24}$$

possibly killed at rate $\delta(x) := \frac{c}{\alpha}(x)$ as soon as $c \neq 0$.

Whenever $\widetilde{x}_t$ is killed, it enters into a coffin state $\{\partial\}$.

Let $\widetilde{\tau}_x$ be the new absorbing time at the boundaries of $\widetilde{x}_t$, started at x, with $\widetilde{\tau}_x = \infty$, if the boundaries are inaccessible to the new process $\widetilde{x}_t$.

Let $\widetilde{\tau}_{x,\partial}$ be the killing time of $(\widetilde{x}_t; t \geq 0)$ started at x (the hitting time of ∂), with $\widetilde{\tau}_{x,\partial} = \infty$ if $c = 0$.

Then $\overline{\tau}_x := \widetilde{\tau}_x \wedge \widetilde{\tau}_{x,\partial}$ is the novel stopping time of $\widetilde{x}_t$ to consider. The SDE for $\widetilde{x}_t$, together with its global stopping time $\overline{\tau}_x$ characterize the transformed process $\overline{x}_t$, with generator $\overline{G}$.

For the new process $\overline{x}_t$, one may also wish to evaluate

$$\widetilde{\alpha}(x) := \widetilde{\mathbf{E}}\left(\int_0^{\overline{\tau}_x} \widetilde{c}(\widetilde{x}_s)ds + \widetilde{d}(\widetilde{x}_{\overline{\tau}(x)})\right).$$

This is an additive functional, where the functions $\widetilde{c}$ and $\widetilde{d}$ are themselves both non-negative. $\widetilde{\alpha}$ is positive in $(0, 1)$. It solves the Dirichlet problem

$$-\overline{G}(\widetilde{\alpha}) = \widetilde{c}, \text{ if } x \in (0, 1),$$

$$\widetilde{\alpha} = \widetilde{d}, \text{ if } x \in \{0, 1\}.$$

With $\mathfrak{g}(x, y)$, the Green function of x_t, we get

$$\widetilde{\alpha}(x) = \frac{1}{\alpha(x)} \int_{(0,1)} \mathfrak{g}(x, y)\alpha(y)\widetilde{c}(y)dy + \widetilde{d}(0) + \left(\widetilde{d}(1) - \widetilde{d}(0)\right)x. \tag{25}$$

A particular quantity of interest is the distribution of $\overline{\tau}_x$ itself. From (22), we have

$$\mathbf{P}(\overline{\tau}_x > t) = \frac{1}{\alpha(x)} \int_{(0,1)} \alpha(y)p(x; t, y)dy,$$

and also from (13),

$$\mathbf{P}(\overline{\tau}_x > t) = \sum_{k\geq 2} b_k e^{-\lambda_k t} \frac{u_k(x)}{\alpha(x)} \int_{(0,1)} \alpha(y)v_k(y)dy.$$

In particular,

$$\mathbf{E}(\overline{\tau}_x) = \int_0^\infty \mathbf{P}(\overline{\tau}_x > t)dt = \frac{1}{\alpha(x)} \sum_{k\geq 2} b_k \frac{u_k(x)}{\lambda_k} \int_{(0,1)} \alpha(y)v_k(y)dy.$$

However, from (25) with $\widetilde{c} = 1$ and $\widetilde{d} = 0$, this complicate expression is also more compactly

$$\mathbf{E}(\overline{\tau}_x) = \frac{1}{\alpha(x)} \int_{(0,1)} \mathfrak{g}(x, y)\alpha(y)dy,$$

which is easy to evaluate from the knowledge of $\mathfrak{g}$.

3.5.1. Normalizing and conditioning: Yaglom limits of the transformed process

Consider the process $\overline{G}$ losing mass due either to absorption at the boundaries and/or to killing. Let $\overline{\rho}_t(x) := \int_{(0,1)} \overline{p}(x; t, y)dy = \widetilde{\mathbf{P}}(\overline{\tau}_x > t)$ be the tail distribution of the full stopping time $\overline{\tau}_x$. Then,

$$\partial_t \overline{\rho}_t(x) = \overline{G}(\overline{\rho}_t(x)) = -\delta(x)\overline{\rho}_t(x) + \widetilde{G}(\overline{\rho}_t(x)), \tag{26}$$

with $\overline{\rho}_0(x) = 1_{(0,1)}(x)$.

Introduce the conditional probability density: $\overline{q}(x;\, t,\, y) := \overline{p}(x;\, t,\, y) / \overline{\rho}_t(x)$, now with total mass 1. With $\overline{q}(x;\, 0,\, y) = \delta_y(x)$, $\overline{q}$ obeys

$$\begin{aligned}\partial_t \overline{q} &= -\partial_t \overline{\rho}_t(x) / \overline{\rho}_t(x) \cdot \overline{q} + \overline{G}^*(\overline{q}) \\ &= (-\partial_t \overline{\rho}_t(x) / \overline{\rho}_t(x) - \delta(y)) \cdot \overline{q} + \widetilde{G}^*(\overline{q}).\end{aligned}$$

Here, $-\partial_t \overline{\rho}_t(x) / \overline{\rho}_t(x) > 0$ is again the rate at which mass should be created to compensate the loss of mass of the process $\widetilde{x}_t$ due to its possible absorption at the boundaries and/or to its killing. With $b_2'' := b_2 \int_{(0,1)} \alpha(y) v_2(y) dy$, we now clearly have

$$e^{\lambda_2 t} \overline{\rho}_t(x) \underset{t\to\infty}{\to} b_2'' \frac{u_2(x)}{\alpha(x)}.$$

Again therefore; $-\frac{1}{t} \log \overline{\rho}_t(x) \underset{t\to\infty}{\to} \lambda_2$ and by L'Hospital rule, $-\partial_t \overline{\rho}_t(x) / \overline{\rho}_t(x) \to \lambda_2$ (λ_2 being again the smallest positive eigenvalue of $-G$). Putting $\partial_t \overline{q} = 0$ in the evolution equation of $\overline{q}$, independently of the initial condition x

$$\overline{q}(x;\, t,\, y) \underset{t\to\infty}{\to} \overline{q}_\infty(y), \tag{27}$$

where $\overline{q}_\infty(y)$ is the solution to

$$-\widetilde{G}^*(\overline{q}_\infty) = (\lambda_2 - \delta(y)) \cdot \overline{q}_\infty, \text{ or } -\overline{G}^*(\overline{q}_\infty) = \lambda_2 \cdot \overline{q}_\infty.$$

With v_2 the eigenvector of $-G^*$ associated to λ_2, $\overline{q}_\infty(y)$ is of the product form

$$\overline{q}_\infty(y) = \alpha(y) v_2(y) / \int_{(0,1)} \alpha(y) v_2(y) dy, \tag{28}$$

because $\overline{G}^*(\cdot) = \alpha(y)G^*(\cdot / \alpha(y))$ and v_2 is the stated eigenvector of $-G^*$.

The limiting probability density $\overline{q}_\infty = \alpha v_2 / \int_{(0,1)} \alpha(y)v_2(y)dy$ is thus the Yaglom limit law of $\overline{x}_t$, now conditioned on the event $\overline{\tau}_x > t$.

3.5.2. Illustrative transformations of interest

(i) The case $c = 0$: In this case, $\widetilde{\tau}_{x,\partial} = \infty$ and so $\overline{\tau}_x := \widetilde{\tau}_x$, the absorption time for the process $\widetilde{x}_t$ governed by the new SDE. Here $\overline{G} = \widetilde{G}$. Assuming α solves $-G(\alpha) = 0$ if $x \in (0, 1)$ with boundary conditions $\alpha(0) = 0$ and $\alpha(1) = 1$ (respectively, $\alpha(0) = 1$ and $\alpha(1) = 0$), new process $\widetilde{x}_t$ is x_t conditioned on exit at $x = 1$ (respectively, at $x = 0$). In the first case, boundary 1 is exit, whereas 0 is entrance; $\alpha = \alpha_1$ reads

$$\alpha_1(x) = \frac{\varphi(x) - \varphi(0)}{\varphi(1) - \varphi(0)},$$

with new drift

$$\widetilde{f}(x) = f(x) + \frac{g^2(x)\alpha_1'(x)}{\alpha_1(x)}.$$

In the second case, $\alpha(x) = \alpha_0(x)$ and boundary 0 is exit, whereas 1 is entrance. Thus $\widetilde{\tau}_x$ is just the exit time at $x = 1$ (respectively, at $x = 0$). Let $\widetilde{\alpha}(x) := \widetilde{\mathbf{E}}(\widetilde{\tau}_x)$. Then, $\widetilde{\alpha}(x)$ solves $-\widetilde{G}(\widetilde{\alpha}) = 1$, whose explicit solution is ($\alpha(x) = \alpha_1(x)$ or $\alpha_0(x)$)

$$\widetilde{\alpha}(x) = \frac{1}{\alpha(x)} \int_{(0,1)} \mathfrak{g}(x, y)\alpha(y)dy,$$

in terms of $\mathfrak{g}(x, y)$, the Green function of x_t.

Examples. (i) Starting from the WF diffusion on $[0, 1]$, these constructions are important to understand the WF diffusion $\widetilde{x}_t$ conditioned on either extinction or fixation, adding an appropriate linear

drift and avoiding killing. See [14], for the determination of the beta Yaglom limits of the conditioned processes in these cases and the corresponding expected fixation and extinction times, related to the Kimura-Ohta formulae for the age of a mutant, using reversibility [13].

(ii) Consider the WF model on $[0, 1]$ with selection for which, with $\sigma \in \mathbb{R}$, $f(x) = \sigma x(1-x)$ and $g^2(x) = x(1-x)$. Assume α solves $-G(\alpha) = 0$ if $x \in (0, 1)$ with $\alpha(0) = 0$ and $\alpha(1) = 1$; then $\alpha_1(x) = (1 - e^{-2\sigma x}) / (1 - e^{-2\sigma})$. The diffusion corresponding to (24) has new drift: $\widetilde{f}(x) = \sigma x(1-x) \coth(\sigma x)$, independently of the sign of σ. This is the WF diffusion with selection conditioned on exit at $\{1\}$.

(iii) Assume α now solves $-G(\alpha) = 1$ if $x \in (0, 1)$ with boundary conditions $\alpha(0) = \alpha(1) = 0$. Proceeding in this way, one selects sample paths of x_t with a large mean absorption time $\alpha(x) = \mathbf{E}(\tau_x)$. Sample paths with large sojourn time in $(0, 1)$ are favoured. We have

$$\alpha(x) = \int_{(0,1)} \mathfrak{g}(x, y) dy,$$

where $\mathfrak{g}(x, y)$ is the Green function (18). The boundaries of $\widetilde{x}_t$ are both entrance, so $\widetilde{\tau}_x = \infty$ and $\widetilde{x}_t$ is not absorbed at the boundaries. The stopping time $\overline{\tau}_x$ of $\widetilde{x}_t$ is just its killing time $\widetilde{\tau}_{x,\partial}$. Let $\widetilde{\alpha}(x) := \widetilde{\mathbf{E}}(\widetilde{\tau}_{x,\partial})$. Then, $\widetilde{\alpha}(x)$ solves $-\overline{G}(\widetilde{\alpha}) = 1$, $\widetilde{\alpha}(0) = \widetilde{\alpha}(1) = 0$, with explicit solution

$$\widetilde{\alpha}(x) = \frac{1}{\alpha(x)} \int_{(0,1)} \mathfrak{g}(x, y)\alpha(y) dy.$$

(iv) Assume α now solves $-G(\alpha) = \delta_y(x)$ if $x \in (0, 1)$ with boundary conditions $\alpha(0) = \alpha(1) = 0$. Using this α, one selects sample paths of x_t with a large sojourn time density at y, recalling $\alpha(x) =: \mathfrak{g}(x, y) = \mathbf{E}(\int_0^{\tau_x} \delta_y(x_s) ds)$. The drift of $\widetilde{x}_t$ is

$$\widetilde{f}(x) = f(x) + g^2(x)\frac{\alpha_0'(x)}{\alpha_0(x)}, \text{ if } y \le x$$

$$= f(x) + g^2(x)\frac{\alpha_1'(x)}{\alpha_1(x)}, \text{ if } x < y,$$

$\widetilde{x}_t$ is thus x_t conditioned on exit at $\{1\}$ if $x < y$ and x_t conditioned on exit at $\{0\}$ if $x > y$.

The stopping time $\overline{\tau}_y(x)$ of $\widetilde{x}_t$ occurs at rate $\delta_y(x)/\mathfrak{g}(x, y)$. It is a killing time when the process met y at least once and is at y for the last time before entering ∂. Let $\widetilde{\alpha}_y(x) := \widetilde{\mathbf{E}}(\overline{\tau}_y(x))$. Then, $\widetilde{\alpha}_y(x)$ solves $-\overline{G}(\widetilde{\alpha}) = 1$, with explicit solution

$$\widetilde{\alpha}_y(x) = \frac{1}{\mathfrak{g}(x, y)}\int_{(0,1)} \mathfrak{g}(z, y)\mathfrak{g}(x, z)dz.$$

When $x = 1/N$, $\widetilde{\alpha}_y(1/N)$ gives the age of a mutant currently observed to the present frequency y.

As an illustrative example, if x_t is the WF diffusion, $\widetilde{\alpha}_y(x)$ can easily be found to be

$$\widetilde{\alpha}_y(x) = -2\left(1 + \frac{1-x}{x}\log(1-x) + \frac{y}{1-y}\log y\right),$$

which, if $x = 1/N$, gives back the celebrated Kimura and Ohta formula: $\widetilde{\alpha}_y(1/N) = -2\left(\frac{y}{1-y}\log y + O(\frac{1}{N})\right)$, which can be obtained differently, see [13].

(v) Let λ_2 be the smallest non-null eigenvalue of G. Let $\alpha = u_2$ correspond to the second eigenvector: $-G(u_2) = \lambda_2 u_2$, with boundary conditions $u_2(0) = u_2(1) = 0$. Then $c(c) = \lambda u_2(x)$. The KB generator associated to $\overline{x}_t$ is

$$\overline{G}(\cdot) = \frac{1}{\alpha}G(\alpha)\cdot + \widetilde{G}(\cdot) = -\lambda_2 \cdot + \widetilde{G}(\cdot),$$

obtained while killing the sample paths of the process $\widetilde{x}_t$ governed by $\widetilde{G}$ at a constant death rate $\delta(x) = \lambda_2$. The transition probability density of $\overline{x}_t$ is

$$\overline{p}(x;\, t,\, y) = \frac{u_2(y)}{u_2(x)}\, p(x;\, t,\, y).$$

Define $\widetilde{p}(x;\, t,\, y) = e^{\lambda_2 t}\overline{p}(x;\, t,\, y)$. This is the transition probability density of $\widetilde{x}_t$, governed by $\widetilde{G}$, corresponding to the original process x_t conditioned on never hitting the boundaries $\{0, 1\}$ (the so-called Q-process of x_t).

The process $\widetilde{x}_t$ is obtained from x_t while adding the additional drift term $\frac{u_2'}{u_2}\, g^2$ to the original drift f. The determination of $\alpha = u_2$ is a Sturm-Liouville problem. When t is large, to the dominant order

$$p(x;\, t,\, y) \sim e^{-\lambda_2 t}\, \frac{u_2(x)v_2(y)}{\int_{(0,1)} u_2(y)v_2(y)dy},$$

where v_2 is the Yaglom limit law of $(x_t;\, t \geq 0)$. Therefore,

$$\widetilde{p}(x;\, t,\, y) \sim e^{\lambda_2 t}\, \frac{u_2(y)}{u_2(x)}\, e^{-\lambda_2 t}\, \frac{u_2(x)v_2(y)}{\int_{(0,1)} u_2(y)v_2(y)dy} = \frac{u_2(y)v_2(y)}{\int_{(0,1)} u_2(y)v_2(y)dy}. \tag{29}$$

The limit law of the Q-process $\widetilde{x}_t$ is thus the normalized Hadamard product of the eigenvectors u_2 and v_2, associated, respectively, to G and G^*.

Example. When dealing, for example, with the neutral Wright-Fisher diffusion, it is known that $\lambda_2 = 1$ with $u_2 = x(1 - x)$ and $v_2 \equiv 1$. The limit law of the Q-process $\widetilde{x}_t$ in this case is $6y(1 - y)$, which is a beta(2, 2) density. □

4. Extreme Reproduction Events

So far, the forward scaling limits of the discrete space-time Markov chains obtained as a conservative Galton-Watson branching process with balanced reproduction laws, were of the diffusion type. We now discuss an 'unbalanced' case, where one individual is allowed to be very productive compared to the other ones.

Assume again a random dynamical population model with non-overlapping generations $t \in \mathbb{Z}$ and a constant population size N. That is, starting with N individuals at generation $t = 0$, we assume that each individual can die or give birth to a random number of descendants while preserving the total number of individuals at the next generation. Suppose the random reproduction law at generation 0 is $\boldsymbol{\nu} := (\nu_1, \ldots, \nu_N)$, therefore obeying the conservation law

$$\sum_{n=1}^{N} \nu_n = N.$$

Here, ν_n is the random number of offspring of the individual number n. Iterate the reproduction law at each times. With $[N] := \{1, \ldots, N\}$, tracing the number of offspring of a subset of individuals from $[N]$ leads to a conservative branching Galton-Watson process in $(0 \cup [N])^{\mathbb{Z}}$ first introduced in [18]. If one adds the following assumptions:

(1) Exchangeability of $\boldsymbol{\nu}$.

(2) Homogeneity: The reproduction laws are i.i.d. for each generation $t \in \mathbb{Z}$.

(3) Neutrality (no mutation, no selection,...), then we get a Cannings process with reproduction law $\boldsymbol{\nu}$ on the N-simplex ([3], [4]).

4.1. The discrete extended Moran model

The extended Moran model is a special class of Cannings model defined as follows ([17]):

Consider a population of N individuals. Let $M_N > 1$ be an r.v. taking values in $\{2, \ldots, N\}$ and define the offspring vector $\boldsymbol{\mu} := (\mu_1, \ldots, \mu_N)$ via $\mu_1 := M_N$, $\mu_n = 0$ for $n \in \{2, \ldots, M_N\}$ and $\mu_n = 1$ for $n \in \{M_N + 1, \ldots, N\}$. μ_n is the number of descendants at generation 0 of the n-th individual. Consider the exchangeable Cannings reproduction model $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$ obtained as a random permutation of $\boldsymbol{\mu}$. For such a model for $\boldsymbol{\nu}$, one individual taken at random from $[N]$ is allowed to produce a (possibly) large number M_N of descendants, the other ones fitting their descendance, either 0 or 1, to guarantee the total number conservation.

This allows to define two Markov chains.

• **Forward in time.** Take a sub-sample of $n \le N$ individuals and let $x_t^{(N)}(n)$ denote the number of descendants of these n out of N individuals, t generations forward in time. Then $x_t^{(N)}$ with $x_0^{(N)} = n$, is a discrete-time Markov chain (with state-space $\{0, \ldots, N\}$ and absorbing barriers $\{0, N\}$), whose transition probabilities $P_{i,j}^{(N)} := \mathbf{P}\left(x_{t+1}^{(N)} = j \,|\, x_t^{(N)} = i\right)$ are given by (see [[17], page 2, (1)])

$$
\begin{aligned}
P_{i,j}^{(N)} &= \frac{1}{\binom{N}{i}} \mathbf{E}\left[\binom{N - M_N}{j}\binom{M_N - 1}{i - j}\right], \text{ if } j < i,\\
P_{i,j}^{(N)} &= \frac{1}{\binom{N}{i}} \mathbf{E}\left[\binom{N - M_N}{i} + \binom{N - M_N}{N - i}\right], \text{ if } j = i, \qquad (30)\\
P_{i,j}^{(N)} &= \frac{1}{\binom{N}{i}} \mathbf{E}\left[\binom{N - M_N}{N - j}\binom{M_N - 1}{j - i}\right], \text{ if } j > i.
\end{aligned}
$$

For $M_N \equiv 2$, this model reduces to the standard Moran model [28] with forward transition probabilities $P_{i,i-1}^{(N)} = i(N-i)/(N(N-1))$, $i \in \{1, \ldots, N\}$, $P_{i,i+1}^{(N)} = i(N - i)/(N(N - 1))$, $i \in \{0, \ldots, N - 1\}$, $P_{i,i}^{(N)} = 1 - 2i(N - i)/$

$(N(N-1))$, $i \in \{0, \ldots, N\}$, and $P_{i,j}^{(N)} = 0$, otherwise. In the Moran model, two individuals are chosen at random; one is bound to generate 2 offspring, while the other one dies out. The rest of the population generates one offspring.

Allowing $M_N > 2$ random and possibly of order N, the extended Moran model provides the opportunity that one individual is very productive compared to the other ones, who either survive or die in the next generation. Note that under our assumption $M_N \geq 2$, M_N is the largest of the νs : $M_N = \max(\nu_1, \ldots, \nu_N)$.

• **Backward in time.** Take a sub-sample of size n from $[N]$ 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 genes from $[n]$. Define the induced ancestral backward process as:

$\mathcal{A}_t(n) \in \mathcal{E}_n =$ {equivalence relations on $[n] \subset [N]$}, $t \in \mathbb{N}$, backward in time. The ancestral process is a discrete-time-t Markov chain with transition probability

$$\mathbf{P}\left(\mathcal{A}_{t+1}(n) = \alpha \mid \mathcal{A}_t(n) = \beta\right) = \widehat{P}_{\beta,\alpha}^{(N)}; \quad \text{with } (\alpha, \beta) \in \mathcal{E}_n, \quad \alpha \subseteq \beta,$$

where, with $(n)_j := n(n-1)\ldots(n-j+1)$, and

$j = |\alpha| =$ the number of equivalence classes of α;

$i = |\beta| =$ the number of equivalence classes of β;

$\mathbf{i}_j := (i_1, \ldots, i_j)$ the clusters (blocks) sizes of β;

the transition probability reads

$$\widehat{P}_{\beta,\alpha}^{(N)} =: \widehat{P}_{i,j}^{(N)}(\mathbf{i}_j) = \frac{(N)_j}{(N)_i}\,\mathbf{E}\left(\prod_{l=1}^{j}(\nu_l)_{i_l}\right).$$

Let $\widehat{x}_t^{(N)} = \widehat{x}_t^{(N)}(n)$ count the number of ancestors at generation $t \in \mathbb{N}$, backward in time, starting from $\widehat{x}_0^{(N)} = n \le N$. Then, $\widehat{x}_t^{(N)}$ also counts the number of blocks of $\mathcal{A}_t(n)$, $\widehat{x}_t^{(N)} = |\mathcal{A}_t(n)|$. This backward counting process is a discrete-time Markov chain with state-space $\{0, \dots, N\}$ and transition probability

$$\mathbf{P}\left(\widehat{x}_{t+1}^{(N)} = j \,|\, \widehat{x}_t^{(N)} = i\right) =: \widehat{P}_{i,j}^{(N)} = \frac{i!}{j!} \sum_{\substack{i_1,\dots,i_j \in \mathbb{N}_+ \\ i_1+\dots+i_j = i}} \frac{\widehat{P}_{i,j}^{(N)}(\mathbf{i}_j)}{i_1! \dots i_j!}.$$

When the reproduction law ν is the one of an extended Moran model, for $i, j \in \{1, \dots, N\}$, (see [[17], page 2])

$$\widehat{P}_{i,j}^{(N)} = \frac{\mathbf{E}\left[\binom{N - M_N}{j-1}\binom{M_N}{i-j+1}\right]}{\binom{N}{i}}, \text{ if } j < i,$$

$$\widehat{P}_{i,j}^{(N)} = \frac{\mathbf{E}\left[\binom{N - M_N}{i} + M_N\binom{N - M_N}{i-1}\right]}{\binom{N}{i}}, \text{ if } j = i, \tag{31}$$

$$\widehat{P}_{i,j}^{(N)} = 0, \text{ if } j > i.$$

Both matrices $P^{(N)}$ and $\widehat{P}^{(N)}$ can be shown to be similar and so they share the same eigenvalues: thus $\widehat{P}_{i,i}^{(N)}$ in (31) are also the eigenvalues of $P^{(N)}$.

Note that $\widehat{P}_{i,1}^{(N)} = \mathbf{E}[(M_N)_i]/(N)_i$, $i \in \{2, \dots, N\}$ is the probability that i individuals chosen at random from some generation share a common parent. In particular, the coalescence probability is $c_N := \widehat{P}_{2,1}^{(N)} =$

$\mathbf{E}[(M_N)_2]/(N(N-1))$, in agreement with [7]. c_N is the probability that two individuals chosen at random from some generation have a common parent. The effective population size $N_e := 1/c_N$ is a relevant quantity. Introduce also the probability $d_N := \widehat{P}_{3,1}^{(N)}$ that 3 individuals chosen at random from some generation share a common parent. For scaling limits, whether $c_N \to 0$ or not and whether triple mergers are asymptotically negligible compared to double ones $(\frac{d_N}{c_N} \to 0)$ or not $(\frac{d_N}{c_N} \not\to 0)$ is important, [33].

4.2. Scaling. Depending on whether

(i) (Occasional extreme events): $M_N / N \xrightarrow{d} 0$ (convergence in distribution as $N \to \infty$) or

(ii) (Systematic extreme events): $M_N / N \xrightarrow{d} U$ (as $N \to \infty$), where U is a non-degenerate $[0, 1]$-valued r.v. with $\mathbf{E}(U) > 0$, different scaling processes both forward and backward in time can arise in the large N population limit.

4.2.1. Occasional extreme events. Let us first discuss the case (i).

- **Backward in time.** In this first case (i), if in addition, the limits

$$\phi(k) := \lim_{N\to\infty} \frac{\mathbf{E}[(M_N)_k]}{N^{k-2}\mathbf{E}[(M_N)_2]}, \tag{32}$$

exist for all $k \in \{2, 3, \ldots\}$, then the extended Moran model is in the domain of attraction of a continuous-time Λ-coalescent $\widehat{x}_t$, with Λ a probability measure on $[0, 1]$ uniquely determined by its moments: $\int_0^1 u^{k-2}\Lambda(du) = \phi(k)$ (see [17], page 3). Λ-coalescents allow for multiple,

but no simultaneous collisions (see [32] for a precise definition). All continuous-time Λ-coalescents can be produced as such a limiting extended Moran process. They are obtained from the discrete-time $\hat{x}_t^{(N)}$ of Subsection 4.1 as

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

where (see [17], page 5)

$$1/c_N = \sum_{j=1}^{N-1} \binom{N}{j-1} \int_0^1 u^{N-j-1}(1-u)^{j-1} \Lambda(du).$$

The limiting process Λ-coalescent $\hat{x}_t$ is integral-valued at all (continuous) times and non-increasing (it is a pure death process). It has 1 as an absorbing state. It has transition rate matrix $\widehat{Q}^{\infty}$, which is a lower tri-diagonal semi-infinite matrix with non-null entries

$$\widehat{Q}^{\infty}_{i,j} = \binom{i}{j-1} \int_0^1 u^{i-j-1}(1-u)^{j-1} \Lambda(du), \text{ if } 2 \le j < i, \tag{33}$$

$$\widehat{Q}^{\infty}_{i,i} = -\sum_{j=1}^{i-1} \widehat{Q}^{\infty}_{i,j} =: -Q_i, \tag{34}$$

where $Q_i := 1/c_i$ is the total death rate of $\hat{x}_t$ starting from state i. When $\Lambda(\{0\}) = 0$ (excluding the Kingman coalescent), its dynamics when started at n is given by $\hat{x}_0 = n$ and

$$\begin{aligned} \hat{x}_t - \hat{x}_0 &= -\int_{(0,t]\times(0,1]} \left(B(\hat{x}_{s_-}, u) - 1_{B(\hat{x}_{s_-}, u)>0} \right) \mathcal{N}(ds \times du) \\ &= -\int_{(0,t]\times(0,1]} (B(\hat{x}_{s_-}, u) - 1)_+ \mathcal{N}(ds \times du). \end{aligned} \tag{35}$$

Here, $x_+ = \max(x, 0)$, $\mathcal{N}$ is a random Poisson measure on $[0, \infty) \times (0, 1]$ with intensity $ds \times \frac{1}{u^2} \Lambda(du)$ and $B(\hat{x}_{s\_}, u) \overset{d}{\sim}$ bin $(\hat{x}_{s\_}, u)$ is a binomial r.v. with parameters $(\hat{x}_{s\_}, u)$. As a result, with

$$r(y) := \int_{(0,1]} (uy - 1 + (1-u)^y) \frac{1}{u^2} \Lambda(du), \quad y > 0, \tag{36}$$

upon taking the expectation in (35), it holds that

$$\mathbf{E}(d\hat{x}_t \mid \hat{x}_{t\_}) = -r(\hat{x}_{t\_})\, dt.$$

From this, the quantity $r(y)$ is the rate at which size y blocks are being lost as time passes by. Note that r is also

$$r(i) = iQ_i - \sum_{j=1}^{i-1} j\hat{Q}^{\infty}_{i,j} = \sum_{j=1}^{i-1} (i-j) \binom{i}{j-1} \int_0^1 u^{i-j-1} (1-u)^{j-1} \Lambda(du). \tag{37}$$

Consequently, the reciprocal function $1/r(y)$ of the rate r interprets as the expected time spent by $\hat{x}_t$ in a state with y lineages and therefore $\sum_{y=2}^{\hat{x}_0=n} r(y)^{-1}$ will give the expected time to the most recent common ancestor

$$\hat{\tau}_{n,1} := \inf(t \in \mathbb{R}_+ : \hat{x}_t = 1 \mid \hat{x}_0 = n).$$

In some cases, the latter sum can be estimated by $\int_1^{\hat{x}_0=n} r(y)^{-1}\, dy$; it will give the (large-n) order of magnitude of the expected time to the most recent common ancestor. Similarly, one expects that $\int_1^{\hat{x}_0=n} yr(y)^{-1} dy$ will give the order of magnitude of the expected length of the coalescent, which is the additive functional $L_n = \mathbf{E}\int_0^{\hat{\tau}_{n,1}} \hat{x}_s\, ds$ and more generally

that $\int_1^{\hat{x}_0=n} c(y)\, r(y)^{-1}\, dy$ will give the order of magnitude of the additive functional $\mathbf{E}\int_0^{\hat{\tau}_{n,1}} c(\hat{x}_s)\, ds$. In other words, with r given either by (36) or by (37), the Green function of the continuous-time Λ-coalescent $\hat{x}_t$ is

$$\mathfrak{g}(x,\, y) = r(y)^{-1} \cdot 1_{y\in\{2,\ldots,x=n\}}, \quad x,\, y \in \{2,\, 3,\, \ldots\}.$$

There are lots of detailed studies in the literature on the length of the Λ-coalescent, the length of its external branch, the number of collisions till time to most recent common ancestor, …. Famous examples include Λ-coalescents for which:

• (Lebesgue) $\Lambda(du) = 1_{[0,1]}(u)du$: This is the Bolthausen-Sznitman coalescent. In this case, $u^{-2}\Lambda(du)$ is not integrable.

• $\Lambda(du) = B(2-\alpha,\, \alpha)u^{1-\alpha}(1-u)^{\alpha-1}1_{[0,1]}(u)du$, $(\alpha \in (0,1)\cup(1,2))$, with $B(a,\, b)$ the beta function; this is the beta(α) coalescent. In this case, $u^{-2}\Lambda(du)$ is not integrable either.

• $\Lambda(du) = B(a,\, b)u^{a-1}(1-u)^{b-1}1_{[0,1]}(u)du$: We get the beta$(a,\, b)$ coalescent. In this case, $u^{-2}\Lambda(du)$ is integrable only if $a > 2$. □

• **Forward in time.** Let us now briefly look at the scaling limits forward in time. In case (i), the coalescence probability c_N tends to 0 and the space-time scaled forward process x_t, as a scaling limit of $x^{(N)}_{\lfloor t/c_N \rfloor}(\lfloor Nx \rfloor)/N$ with $x_0 = x$ and $t \in \mathbb{R}_+$, is a well-defined (two-types neutral Λ-Fleming-Viot) continuous-time Markov process with state-space $[0,\, 1]$ (see [2], [9]). More precisely, x_t has backward infinitesimal generator

$$\psi \in C^2([0,1]) \to G(\psi)(x) = \frac{\Lambda(\{0\})}{2} x(1-x)\partial_x^2 \psi(x)$$

$$+ \int_{[0,1]\setminus\{0\}} [x\psi(x+(1-x)u) + (1-x)\psi(x(1-u)) - \psi(x)] \frac{1}{u^2} \Lambda(du),$$

which is the one of a pure jump process if Λ has no atom at $\{0\}$, so with

$$u(x,t) = \mathbf{E}_x \psi(x_t) \text{ obeying } \partial_t u = G(u); \quad u(x,0) = \psi(x).$$

Equivalently, the sample-paths of x_t obey the stochastic evolution

$$x_t - x_0 = \int_0^t \sqrt{\Lambda(\{0\}) x_s (1-x_s)}\, dw_s$$

$$+ \int_{(0,t]\times(0,1]\times[0,1]} \left(1_{v \le x_{s_-}} u(1-x_{s_-}) - 1_{v > x_{s_-}} u x_{s_-}\right) \mathcal{N}(ds \times du \times dv), \quad (38)$$

where $\mathcal{N}$ is a random Poisson measure on $[0,\infty)\times(0,1]\times[0,1]$ with intensity $ds \times \frac{1}{u^2}\Lambda(du) \times dv$, independent of the standard Brownian motion w_t. If $\Lambda(0) \neq 0$ term accounting for the Wright-Fisher diffusion has to be included. In (38), the clock-time t is measured in units of $N_e = c_N^{-1}$.

Example. The Eldon and Wakeley model, [7]. Let $\gamma > 0$. Take for M_N, the following mixture model:

$$M_N = 2 \text{ with probability } 1 - N^{-\gamma} \text{ (Moran model)},$$

$$M_N = 2 + \lfloor (N-2)V \rfloor \text{ with probability } N^{-\gamma},$$

where V is an r.v. on $[0,1]$ with distribution $\alpha(du)$. The law of M_N itself is

$$\pi(du) = (1-N^{-\gamma})\delta_2 + N^{-\gamma} \cdot \delta_2 * \mathcal{L}(\lfloor (N-2)V \rfloor),$$

leading to occasional extreme reproduction events, with probability $N^{-\gamma}$, with clearly $M_N / N \xrightarrow{d} 0$. In [7], the law of V is a Dirac mass at some $\psi \in (0, 1)$. For all $\gamma > 0$, one can check that $c_N \underset{N\to\infty}{\to} 0$. The larger the values of γ, the smaller the contribution of extreme events. By computing the limiting behaviour $(N \to \infty)$ of $\frac{d_N}{c_N}$ in each case, one can conclude

• If $\gamma > 2$: We are in the attraction basin of the Kingman coalescent ($\frac{d_N}{c_N} \to 0$), [23]. So, no jump process in the scaling limit of the forward process, only the Wright-Fisher diffusion.

• If $\gamma \le 2$: We are not in the attraction basin of the Kingman coalescent ($\frac{d_N}{c_N} \not\to 0$), rather of the full Λ-coalescent. So, here a jump process in the scaling limit of the forward process, to which a Wright-Fisher diffusion term should be superposed if π has mass at 0. □

4.2.2. Systematic extreme events. Let us now investigate the case (ii).

• **Backward in time.** In the second case (ii), $\widehat{P}^{(N)}_{i,1} \to \mathbf{E}\left(U^i\right) > 0$ and the extended Moran model is in the domain of attraction of a discrete-time Λ-coalescent with $\Lambda(du) = u^2\pi(du)$ and $\pi(du)$ the probability distribution of U, [17]. Here extreme events are not occasional, but systematic because M_N is a random fraction of N. In particular, the coalescence probability c_N tends to a limit $c = \mathbf{E}\left(U^2\right) > 0$. It holds that

$$\left(\hat{x}^{(N)}_t,\, t \in \mathbb{N}\right) \xrightarrow{\mathcal{D}} (\,\hat{x}_t,\, t \in \mathbb{N}\,),$$

which is a discrete-time t limiting Λ-coalescent, whose transition matrix $\widehat{P}^{\infty}$ is a lower tri-diagonal semi-infinite stochastic matrix with non-null entries (see (9) and Theorem 2.1 of [27])

$$\widehat{P}^{\infty}_{i,j} = \binom{i}{j-1} \int_0^1 u^{i-j+1} (1-u)^{j-1} \, \pi(du), \text{ if } 1 \le j < i, \tag{39}$$

$$\widehat{P}^{\infty}_{i,i} = \int_0^1 (1-u)^{i-1} (1-u+iu) \, \pi(du), \text{ if } j = i. \tag{40}$$

The corresponding discrete-time dynamics of $\hat{x}_t$ is easily seen to be

$$\hat{x}_{t+1} = \hat{x}_t - (B(\hat{x}_t, U_{t+1}) - 1) 1_{B(\hat{x}_t, U_{t+1}) > 1}, \quad \hat{x}_0 = n,$$

where $B(\hat{x}_t, u) \overset{d}{\sim} \text{bin}(\hat{x}_t, u)$ is a binomial r.v. with parameters $(\hat{x}_t, u)$.

From the expressions (39) and (40), it hold that

$$\mathbf{E}(\hat{x}_{t+1} \mid \hat{x}_t) - \hat{x}_t = -r(\hat{x}_t),$$

where, with $\overline{U} := 1 - U$.

$$r(y) = y\mathbf{E}(U) - 1 + \mathbf{E}\left(\overline{U}^{\,y}\right), \quad y \ge 0,$$

involving the Laplace-Stieltjes transform $\mathbf{E}\left(\overline{U}^{\,x}\right)$ of $-\log(\overline{U})$. The function $r(y)$ is convex with $r'(0) = \mathbf{E}(U) + \mathbf{E}\left(\log \overline{U}\right) < 0$, $r(y) \ge 0$, if $y \ge 1$ and $r(y) \sim y\mathbf{E}(U)$ as $y \to \infty$. It is the discrete-time analogue to the rate r function defined in the continuous-time setting.

Of interest also on this discrete-time coalescent, are the time to most recent common ancestor: $\hat{\tau}_{n,1} := \inf(t \in \mathbb{N} : \hat{x}_t = 1 \mid \hat{x}_0 = n)$, the length of the coalescent tree $L_n = \sum_{t=0}^{\hat{\tau}_{n,1}} \hat{x}_t$, the number of internal nodes, the number of collisions till $\hat{\tau}_{n,1}$ ….

Example. It is not so clear which model π for the law of U is meaningful in population genetics. However, a special 'canonical' case of interest is when π is uniform on $[0, 1]$. One can then check the simple expression

$$\widehat{P}^{\infty}_{i,j} = \frac{1}{i+1}, \text{ if } 1 \le j < i \text{ and } \widehat{P}^{\infty}_{i,i} = \frac{2}{i+1}. \tag{41}$$

□

• **Forward in time.** Whatever π really is, the space-scaled forward process $x_t^{(N)}(\lfloor Nx \rfloor)/N$ with $x_0^{(N)} = x$ has a well-defined scaling limit, which is a discrete-time-t Markov process x_t with state-space $[0, 1]$, defined as follows. Let $(U_t, V_t)_{t\ge1}$ be two mutually independent random sequences with respective common laws: $U_1 \overset{d}{\sim} \pi(du) = \frac{1}{u^2}\Lambda(du)$ and $V_1 \overset{d}{\sim}$ uniform on $[0, 1]$. If π has no atom at $\{0\}$, then x_t is the Markov chain (with state-space $[0, 1]$) driven by $(U_t, V_t)_{t\ge1}$

$$x_{t+1} = x_t + U_{t+1}(1 - x_t)1(V_{t+1} \le x_t) - U_{t+1}x_t 1(V_{t+1} > x_t); \quad x_0 = x. \tag{42}$$

So, depending on the current state of the process and independently, V_{t+1} allows to decide whether x_t moves up or down and then U_{t+1} governs the amplitude of the jump.

Example. If $M_N = 2 + M'_N$, where M'_N is binomially distributed with parameters $N - 2$ and $p \in (0, 1)$, then $M_N / N \overset{d}{\to} U \sim \delta_p$ and $c_N \to p^2$ (relevant for case (ii)). Note that $p = p_N$ may depend on N with $p_N \to 0$ (relevant for case (i)). In the latter case, this model can have a wide variety of effective population sizes $N_e = 1/p_N^2$. For instance, if $p_N = N^{-\alpha}$, $\alpha > 0$, then $N_e = N^{2\alpha}$ is sub-linear for $\alpha < 1/2$ and super-linear for $\alpha > 1/2$. If $p_N = \lambda^N$, $\lambda < 1$, then $N_e = \lambda^{-2N}$ grows exponentially. □

When $\Lambda(du)$ is not reduced to δ_0, both limiting space-scaled forward models (35) and (42) account for jump processes on the unit interval, either with time continuous or discrete, contrasting with the standard

Wright-Fisher diffusion, whose sample-paths are continuous. When dealing with populations with a very productive individual (either occasional or systematic extreme events), the Latin principle of natural philosophy 'Natura non facit saltus' breaks down. When dealing with extreme reproduction events, it is not even possible to scale time.

5. Forward Process Associated to the Discrete Λ-Coalescent

We now study some properties of the discrete-time forward process defined in (42). Let $(U_t, V_t)_{t\geq 1}$ be two mutually independent random sequences with respective common laws: $U_1 \overset{d}{\sim} \pi(du) = \frac{1}{u^2}\Lambda(du)$ and $V_1 \overset{d}{\sim}$ uniform on $[0, 1]$. We shall assume that π is absolutely continuous with density f so that, in particular, it has no atom at $\{0\}$. We also assume that $\{u : f(u) > 0\} \supset (0, 1)$, ($U$ has full support). Consider then the Markov chain (with state-space $[0, 1]$) driven by $(U_t, V_t)_{t\geq 1}$

$$x_{t+1} = x_t + U_{t+1}(1 - x_t)1(V_{t+1} \leq x_t) - U_{t+1}x_t 1(V_{t+1} > x_t); \quad x_0 = x.$$

From this model, if at some (discrete) time t, the process x_t has got close to say 1, there is a big chance (x_t) that in the next step, it will even get closer to 1 by a small move, but there is always some small probability $(1 - x_t)$ that the process can move back abruptly in the bulk of the state-space (by a big move of amplitude $-U_{t+1}x_t$) in which case the whole process starts afresh. By symmetry, a similar argument can be applied when the particle happens to be very close to 0. A rare jump will drive it back at some point inside the state-space, closer to 1 then. The process x_t can be either recurrent or transient on $[0, 1]$ and we would like to fix what it is.

A question maybe important for that purpose is how large is the jump that brings the particle back inside at resetting time?

One could think that would the probability mass of U be concentrated 'close to 1' (U large), the amplitude $-U_{t+1}x_t$ of the rare resetting jump from x_t (already close to 1) is relatively large, so that the particle would sample again the whole interval in this case, leading maybe to a recurrent process x_t. On the other hand, if U is large, x_t can first move fast to the boundary 1 making it harder to escape in the future, and whenever it escapes, then x_t will be trapped near 0 with difficulties to escape 0 then.

Would U be too small (with probability mass concentrated near 0) at resetting time, the particle would still remain too close to 1 after the resetting, thereby ruining the global chance of a real mixing or here of positive-recurrence of x_t.

In fact, we will see latter that whatever the size of the jump at the resetting time (whatever the law of U), the process x_t is always transient: There is a positive probability not to ever visit a neighbourhood of a point $y \neq x$, when $x_0 = x$.

Before that, let us first investigate some immediate properties of x_t.

First, we have $\mathbf{E}(x_{t+1} \mid x_t = x) = x + \mathbf{E}(U_{t+1})(1-x)x - \mathbf{E}(U_{t+1})x(1-x) = x$ and so x_t is a martingale. As for the variance: $\sigma^2(x_{t+1} \mid x_t = x) = \sigma^2(U_{t+1})(1-x)x$. Clearly also, would x_t be transient (and it is), it will eventually hit the boundaries either $\{0\}$ or $\{1\}$, but not in finite time, so that $\tau_x = \tau_{x,0} \wedge \tau_{x,1}$ is ∞ with probability 1. Both boundaries are clearly absorbing. From the martingale property, x_t will eventually hit first the boundary $\{0\}$ (respectively, $\{1\}$) with probability $1-x$ (respectively, x).

Let $\psi \in C_0([0, 1])$. With $t \geq 1$, we have

$$u(x, t) = \mathbf{E}_x \psi(x_t) = (L^t \psi)(x), \quad u(x, 0) = \psi(x),$$

where the backward generator L is given by

$$(L\psi)(x) = \mathbf{E}_x \psi(x_1)$$

$$= x\int_0^1 \psi(x + (1-x)u) f(u)du + (1-x)\int_0^1 \psi(x - xu) f(u)du. \qquad (43)$$

From the expression (39), it is clear that if $\psi(x) = x^k$ is a monomial, $(L\psi)(x)$ remains a degree k polynomial because $[x^{k+1}](L\psi)(x) = 0$. More precisely, we easily get

$$(L\psi)(x) = \sum_{l=1}^{k-1} x^l \binom{k}{l-1} \mathbf{E}\left[U^{k-l+1}(1-U)^{l-1}\right] + [x^k](L\psi)(x).$$

Here, from (39), $[x^l](L\psi)(x) = \widehat{P}^{\infty}_{k,l}$ and

$$[x^k](L\psi)(x) = \mathbf{E}\left[(1-U)^{k-1}(1-U+kU)\right],$$

which, from (40), coincide with the eigenvalues $\widehat{P}^{\infty}_{k,k}$ of $\widehat{P}^{\infty}$. From these facts, it is clear that there exist degree-k eigenpolynomials $u_k(x)$ to the eigenvalue equations: $(\lambda_k I - L)u_k = 0$ with $\lambda_k = \widehat{P}^{\infty}_{k,k}$. Except in the special case $\widehat{P}^{\infty}_{k,k} = 2/(k+1)$ corresponding to U uniform, we were not able to compute them in detail.

Clearly, with a and b real numbers, if $\psi(x) = a + bx$, $(L\psi)(x) = \psi(x)$, showing that the affine functions $a + bx$ are the harmonic functions of L, as required for a martingale.

The operator L as in (43) also takes the form

$$(L\psi)(x) = \frac{1-x}{x}\int_0^x f\left(\frac{x-y}{x}\right)\psi(y)dy + \frac{x}{1-x}\int_x^1 f\left(\frac{y-x}{1-x}\right)\psi(y)dy, \qquad (44)$$

observing that with probability $1 - x$ (the event $V_1 > x$), $x_1 = x(1 - U_1)$ has range $[0, x)$ and law given by the image measure of $U_1 = \dfrac{x - x_1}{x}$, while with probability x (the event $V_1 \le x$), $x_1 = x + U_1(1 - x)$ has range

$(x, 1]$ and law given by the image measure of $U_1 = \frac{x_1 - x}{1 - x}$. Under the latter form, the operator L turns out to be an integral Fredholm operator [24] with kernel

$$K(x, y) = \frac{1-x}{x} f\left(\frac{x-y}{x}\right) 1(0 \le y \le x) + \frac{x}{1-x} f\left(\frac{y-x}{1-x}\right) 1(x < y \le 1), \quad (45)$$

that is; $(L\psi)(x) = \int_0^1 K(x, y)\psi(y)dy$. This operator acts on the Banach space of continuous (and so bounded) functions $C_0([0, 1])$ and it is bounded with norm $\|L\|_\infty = \sup_{\|\psi\|_\infty = 1} \|L\psi\|_\infty = 1$. Because L is associated to a stochastic kernel, $(L1)(x) = 1$ and so the spectral radius of L is also one.

The forward generator L^*, which is the adjoint of L, acts on the space of positive Radon measures and it is easily seen to be given by

$$(L^*\mu)(y) = \int_0^y \frac{z}{1-z} f\left(\frac{y-z}{1-z}\right)\mu(dz) + \int_y^1 \frac{1-z}{z} f\left(\frac{z-y}{z}\right)\mu(dz). \quad (46)$$

The operator L is not self-adjoint, nor is it normal.

As a result of the expression of L^*, the density $p(x; t, y)$ of x_t at y (started at $x_0 = x$) obeys the Fokker-Planck equation $(t \ge 1)$

$$p(x; t+1, y) = \int_0^y \frac{z}{1-z} f\left(\frac{y-z}{1-z}\right) p(x; t, z)dz + \int_y^1 \frac{1-z}{z} f\left(\frac{z-y}{z}\right) p(x; t, z)dz$$

$$= (L^* p(x; t, \cdot))(y).$$

When $t = 0$, with $p(x; 0, \cdot) = \delta_x(\cdot)$

$$p(x; 1, y) = \int_0^y \frac{z}{1-z} f\left(\frac{y-z}{1-z}\right)\delta_x(z)dz + \int_y^1 \frac{1-z}{z} f\left(\frac{z-y}{z}\right)\delta_x(z)dz$$

$$= \frac{1-x}{x} f\left(\frac{x-y}{x}\right) 1(y \le x) + \frac{x}{1-x} f\left(\frac{y-x}{1-x}\right) 1(y > x),$$

is the density of x_1 given $x_0 = x$, which is the kernel $K(x, y)$. Because L^* maps a Dirac measure into a density, there exists a (speed) measure μ with density m satisfying $(L^*\mu)(y) = \mu$ with m obeying

$$m(y) = \int_0^y \frac{z}{1-z} f\left(\frac{y-z}{1-z}\right) m(z)dz + \int_y^1 \frac{1-z}{z} f\left(\frac{z-y}{z}\right) m(z)dz. \qquad (47)$$

Note that the process x_t is not reversible with respect to the speed measure $m(y)dy$

$$m(x)p(x; 1, y) \neq m(y)p(y; 1, x).$$

The expression of $K(x, y)$ is useful for the following statement, which shows that in fact x_t is always transient.

Let $l(x)$ be the probability that the particle always moves to the left (towards 0) starting from x : $l(x) = \mathbf{P}(\ldots < x_2 < x_1 < x_0 = x)$. Then $l(x) \leq 1$ obeys the functional equation

$$l(x) = \frac{1-x}{x} \int_0^x f\left(\frac{x-y}{x}\right) l(y)dy,$$

where in the right-hand-side, $l(y)$ is the same probability when the process is started from $y < x$ after the first jump to the left of x. Clearly, $l(x) \leq 1 - x$. We look for the largest non-null solution to this functional equation. The question is: Is $l(x) > 0$?

Assume $\mathbf{E}(-\log(1 - U_1)) < \infty$. Let $T_x = \inf(t \geq 1 : x_{t+1} > x_t \mid x_0 = x)$ be the first time of a jump to the right given the process started at x. With $\kappa \geq 0$, we have

$$\mathbf{P}(T_x > \kappa + 1) = \mathbf{P}(V_1 > x, V_2 > x\overline{U}_1, \ldots, V_{\kappa+1} > x\overline{U}_1 \cdots \overline{U}_\kappa)$$

$$= (1-x)\mathbf{E}\left(\prod_{k=1}^{\kappa} (1 - x\overline{U}_1 \cdots \overline{U}_k)\right).$$

Developing the product, with $\lambda_l := \mathbf{E}\left(\overline{U}_1^l\right)/\mathbf{E}\left(\overline{U}_1^{l-1}\right)$, the ratio of consecutive moments of $\overline{U}_1$

$$\mathbf{P}(T_x > \kappa + 1) = (1-x)\sum_{k=1}^{\kappa}(-1)^k x^k \sum_{1\le n_1<\ldots<n_k\le\kappa}\prod_{l=1}^{k}\lambda_l^{n_l}.$$

Clearly, $l(x) = \lim_{\kappa\to\infty}\mathbf{P}(T_x > \kappa+1) = (1-x)\mathbf{E}(\prod_{k=1}^{\infty}(1-x\overline{U}_1\cdots\overline{U}_k))$ and the question is if a nonzero limit exists. Let $W_k := \overline{U}_1\cdots\overline{U}_k$. Almost surely (a.s.), we have $W_k \to 0$ and by the Strong law of large numbers: $W_k^{1/k} \to e^{-\mathbf{E}(-\log\overline{U}_1)} \in (0,1)$ (a.s.). Thus, the series W_k is a.s. convergent, with

$$\sum_{k=1}^{\kappa} W_k \underset{\kappa\to\infty}{\to} S,$$

where $S = \sum_{k=1}^{\infty}\overline{U}_1\cdots\overline{U}_k$ is a non-degenerate limiting r.v. obeying: $S \overset{d}{=} \overline{U}_1(1+S')$ with $S' \overset{d}{=} S$ a copy of S, independent of U_1. Note that the moments $m_n := \mathbf{E}(S^n)$ of S can recursively be computed with $m_0 = 1$, and

$$m_n = \frac{\mathbf{E}\left(\overline{U}_1^n\right)}{1-\mathbf{E}\left(\overline{U}_1^n\right)}\sum_{l=0}^{n-1}\binom{n}{l}m_l, \quad n \ge 1.$$

For all $x \in (0, \overline{U}_1)$, we have $e^{-\lambda x} < 1-x < e^{-x}$, where $\lambda > 1$ is defined a.s. by $1 - e^{-\lambda\overline{U}_1} = \overline{U}_1$. Therefore, for all $x \in (0,1)$, the infinite product giving $l(x)$ can be bounded above and below with $l(x) < l_+(x) := (1-x)\mathbf{E}(e^{-xS})$ and, with $S' \overset{d}{=} S$ a copy of S, independent of U_1

$$l(x) > l_-(x) := (1-x)\mathbf{E}\left(e^{-\lambda x\overline{U}_1(1+\sum_{k\ge2}\prod_{l=2}^{k}\overline{U}_l)}\right)$$

$$= (1-x)\mathbf{E}\left(e^{-x(-\log(U_1)\cdot(1+S'))}\right).$$

We have $-\log(U_1) > 1 - U_1$ a.s. and therefore $-\log(U_1)\cdot(1+S') > \overline{U}_1$ $(1+S') \stackrel{d}{=} S$. By the monotonicity of the expectation, for all $x \in (0,1)$

$$l_-(x) := (1-x)\mathbf{E}\left(e^{-x(-\log(U_1)\cdot(1+S'))}\right) < l_+(x) = (1-x)\mathbf{E}(e^{-xS}).$$

Besides, both $\mathbf{E}(e^{-xS})$ and $\mathbf{E}\left(e^{-x(-\log(U_1)\cdot(1+S'))}\right)$ belong to $(0,1)$ as $x \in (0,1)$ because they are the Laplace-Stieltjes transforms of positive random variables evaluated at x. This shows that $0 < l_-(x) < l(x) < l_+(x) < 1$ for all $U_1 \in (0,1)$ satisfying $\mathbf{E}(-\log \overline{U}_1) < \infty$.

So, under the latter assumption on U_1, there is a non trivial positive probability solution $l(x)$, with, as required, $l(x) \to 1$ as $x \to 0$.

Whenever $\mathbf{E}(-\log \overline{U}_1) = \infty$ (which entails $\mathbf{E}(1/\overline{U}_1) = \infty$ and so U_1 very close to 1), $S = 0$ and $l(x) = 1 - x$ corresponding to the probability of the first jump being to the left, where the process is instantaneously brought very close to the 0 boundary, where it remains stuck.

Similarly, the probability $r(x) = \mathbf{P}(x = x_0 < x_1 < x_2 < \dots)$ that the particle always moves to the right starting from x obeys

$$r(x) = \frac{x}{1-x}\int_x^1 f\left(\frac{y-x}{1-x}\right) r(y)\,dy.$$

Proceeding as for $l(x)$, the formal solution is

$$r(x) = x\mathbf{E}\left(\prod_{k=1}^{\infty}(1-(1-x)\overline{U}_1\cdots\overline{U}_k)\right),$$

so $r(x) = l(1-x) > 0$. With $r_-(x) = x\mathbf{E}\left(e^{-(1-x)(-\log(U_1)\cdot(1+S'))}\right)$ and $r_+(x) = x\mathbf{E}(e^{-(1-x)S})$, we therefore have $0 < r_-(x) < r(x) < r_+(x) < 1$ for all $U_1 \in (0,1)$ satisfying $\mathbf{E}(-\log \overline{U}_1) < \infty$ and $r(x) \to 1$ as $x \to 1$.

We conclude from this that when $\mathbf{E}(-\log \overline{U}_1) \leq \infty$, x_t is transient because for all $y \neq x$, there is a positive probability that x_t started at $x_0 = x$ never visits a neighbourhood of y. This probability turns out to be larger than $l(x) > 0$ (respectively, $r(x) > 0$), if y is to the right (to the left) of x.

Remarks. (i) When the law of U is atomic, then things in a way turn out to be simpler; assume, for instance, that the law of U is concentrated on some $\psi \in (0, 1)$. Then, the functional equation giving $l(x)$ can easily be put under the form

$$l(x) = (1 - x)\int_0^1 l(x(1 - z))\delta_\psi(dz) = (1 - x)l(x(1 - \psi)),$$

whose solution (with $\overline{\psi} := 1 - \psi$), is formally the infinite product

$$l(x) = (1 - x)\prod_{k\geq 1}\left(1 - x\overline{\psi}^k\right).$$

For all $x \in (0, \overline{\psi})$, we have $e^{-\lambda x} < 1 - x < e^{-x}$, where $\lambda > 1$ is defined by $1 - e^{-\lambda\overline{\psi}} = \overline{\psi}$. Therefore, for all $x \in (0, 1)$, the infinite product is convergent with

$$(1 - x)e^{-x(-\log\psi/\psi)} = (1 - x)e^{-\lambda x\overline{\psi}/\psi} < l(x) < (1 - x)e^{-x\overline{\psi}/\psi},$$

showing that $1 > l(x) > 0$ for all $\psi \in (0, 1)$; clearly $l(x) \to 0^+$ for all $x \in (0, 1)$, as $\psi \to 0^+$. Clearly also, $xe^{-(1-x)(-\log\psi/\psi)} < r(x) < xe^{-(1-x)\overline{\psi}/\psi}$ are bounds for the probability $r(x)$ of all moves to the right.

When U is atomic at ψ, the process x_t is thus also transient.

(ii) When U is uniformly distributed on $[0, 1]$, the functional equation giving $l(x)$ simply is: $l(x) = \frac{1 - x}{x}\int_0^x l(y)dy$. It has an exact solution obtained while integrating: $l'(x) = \left(-\frac{1}{x(1 - x)} + \frac{1 - x}{x}\right)l(x)$ with boundary conditions $l(0) = 1$. One finds $l(x) = (1 - x)e^{-x}$.

6. The Special Transient Case (U Uniform)

We limit ourselves in the following study to the discrete-time canonical case, where U is uniformly distributed. The limiting discrete Λ-coalescent is thus characterized by (41), whereas, the limiting forward process obeys the discrete-time dynamics (42), so with $U \overset{d}{\sim} \pi(du) = du$.

6.1. The model and its main properties

If we limit ourselves to the case $\pi(du) = du$ (U_1 is also uniform on $[0, 1]$ with $f \equiv 1$), then

$$(L\psi)(x) = \mathbf{E}_x \psi(x_1) = \frac{1-x}{x} \int_0^x \psi(y)dy + \frac{x}{1-x} \int_x^1 \psi(y)dy, \tag{48}$$

which is $(L\psi)(x) = \int_0^1 K(x, y)\psi(y)dy$, where

$$K(x, y) = \frac{1-x}{x} 1(0 \le y \le x) + \frac{x}{1-x} 1(x < y \le 1). \tag{49}$$

Clearly, the kernel K is not totally positive (see [31]), for instance, because

$$\det \begin{bmatrix} K(x_1, y_1) & K(x_1, y_2) \\ K(x_2, y_1) & K(x_2, y_2) \end{bmatrix},$$

with $x_1 < x_2$ and $y_1 < y_2$ is not always positive: K does not either possess the nice spectral properties of totally positive kernels.

We note that $K(x, y)$ is singular in the sense that it is neither bounded nor continuous, on $[0, 1]^2$, nor does it fulfill $\int_{[0,1]^2} K(x, y)^2 dx\, dy < \infty$. Due to the divergence near the boundaries of K, the operator L can easily be checked not to be compact and not even quasi-compact. So K is a singular kernel, which is not reminiscent of the classical (say Hilbert-Schmidt) Fredholm theory for integral operators, [20], [24]. In the special case, with $t \ge 1$, we also have

$$p(x;\, t,\, y) = \left((L^*)^t\, p(x;\, 0,\, \cdot)\right)(y), \qquad p(x;\, 0,\, z) = \delta_x(z),$$

with $p(x;\, 1,\, y) = (L^* p(x;\, 0,\, \cdot))(y)$ given by

$$(L^* p(x;\, 0,\, \cdot))(y) = \int_0^y \frac{z}{1-z}\, p(x;\, 0,\, z)dz + \int_y^1 \frac{1-z}{z}\, p(x;\, 0,\, z)dz$$

$$= \frac{1-x}{x} 1(y \le x) + \frac{x}{1-x} 1(y > x). \tag{50}$$

If the particle is originally at $x < 1/2\,(x > 1/2)$, the probability density of a further move to the left (to the right) is $(1-x)/x$ (respectively, $x/(1-x)$) with $(1-x)/x > x/(1-x)$ (respectively, $x/(1-x) > (1-x)/x$); the process x_t is stochastically monotone. If initially x_0 is uniform, then at step 1, x_1 has density $-\log(y(1-y)) - 1$, diverging symmetrically at both ends, logarithmically.

The probability $l(x)$ that the particle always moves to the left starting from x obeys the functional equation

$$l(x) = \frac{1-x}{x} \int_0^x l(y)dy,$$

whose solution is $l(x) = (1-x)e^{-x}$. Similarly, the probability $r(x)$ that the particle always moves to the right starting from x is by symmetry $r(x) = xe^{-(1-x)}$.

6.2. Resolvent and spectral aspects of the special model

Let

$$(L\psi)(x) = \mathbf{E}_x \psi(x_1) = \frac{1-x}{x} \int_0^x \psi(y)dy + \frac{x}{1-x} \int_x^1 \psi(y)dy.$$

Let $\lambda \in \mathbb{C}$. Let c be a bounded function on $[0, 1]$ satisfying $c(0) = c(1) = 0$. We look for continuous solutions α to the Fredholm problem: $(\lambda I - L)\alpha = c$ or, with $z = \lambda^{-1}$, to

$$(I - zL)\,\alpha = zc. \tag{51}$$

When $|z| < 1$, α takes the converging Liouville-Neumann power-series form

$$\alpha(x) = \sum_{n\geq 0} z^{n+1} L^n(c)(x),$$

involving iterates of L.

Let $A(x) = \int_0^x \alpha(y)dy$ so that $\alpha = A'.(I - zL)\alpha = zc$ is also the linear differential system

$$A'(x) - zA(x)\left(\frac{1}{x} - \frac{1}{1-x}\right) = z\left(c(x) + \frac{x}{1-x}A(1)\right) =: zf(x).$$

Let $A_0(x)$ be the solution of the homogeneous system: $A_0'(x) - zA_0(x)\left(\frac{1}{x} - \frac{1}{1-x}\right) = 0$. With C some constant, we get

$$A_0(x) = C(x(1-x))^z.$$

Applying the method of variation of the constant, let $A(x) = K(x)A_0(x)$. Then, $K'(x) = \frac{zf(x)}{A_0(x)}$, leading to

$$K(x) = \frac{z}{C}\int_{1/2}^x (y(1-y))^{-z} f(y)dy + K(1/2).$$

Therefore, with $f(x) = c(x) + \frac{x}{1-x}A(1)$,

$$A(x) = z(x(1-x))^z \int_{1/2}^x (y(1-y))^{-z} f(y)dy + 4^z(x(1-x))^z A(1/2), \tag{52}$$

$$\alpha(x) = zA(x)\left(\frac{1}{x} - \frac{1}{1-x}\right) + zf(x). \tag{53}$$

• Assume first $|z| < 1$.

Looking at the behaviour near 0 and 1 of $A(x)$ and then of $\alpha(x)$, we find

$$A(x) \underset{x\downarrow 0}{\sim} -\frac{z}{2-z} A(1) 2^{-(2-z)} x^z + 4^z x^z A(1/2),$$

$$\alpha(x) \underset{x\downarrow 0}{\sim} -\frac{z^2}{2-z} A(1) 2^{-(2-z)} x^{z-1} + z 4^z x^{z-1} A(1/2);$$

and

$$A(x) \underset{x\uparrow 1}{\sim} A(1)\left(1 - 2^z (1-x)^z\right) + 4^z (1-x)^z A(1/2),$$

$$\alpha(x) \underset{x\uparrow 1}{\sim} zA(1) 2^z (1-x)^{z-1} - z 4^z (1-x)^{z-1} A(1/2).$$

The solution α is continuous at 0 and 1 only if $A(1/2) = A(1) = 0$. Therefore, when $|z| < 1$, the solution α is unique and takes the simple form

$$\alpha(x) = zc(x) + z^2 (1-2x)\left(x(1-x)\right)^{z-1} \int_{1/2}^{x} \left(y(1-y)\right)^{-z} c(y)\, dy, \qquad (54)$$

which may be viewed an alternative representation to the Liouville-Neumann power series. Recalling $\lambda = z^{-1}$, the domain $\left|\lambda^{-1}\right| < 1$ is the complementary of the unit disk of $\mathbb{C}$ centered at 0. Such λs are regular points of L for which $(\lambda I - L)^{-1}$ exists, is bounded and is defined on the whole space $C_0([0, 1])$.

• Assume now $\mathrm{Re}(z) \geq 1$ and $c \equiv 0$.

When $\mathrm{Re}(z) = 1$ and $c \equiv 0$, we already know that $\alpha(x) = a + bx$ are the harmonic function solutions.

Then, from (52)-(53) with $\mathrm{Re}(z) > 1$ and $c \equiv 0$,

$$A(x) = zA(1)(x(1-x))^z \int_{1/2}^{x} y^{1-z}(1-y)^{-(z+1)}dy + 4^z(x(1-x))^z A(1/2),$$

$$\alpha(x) = zA(x)\left(\frac{1}{x} - \frac{1}{1-x}\right) + z\frac{x}{1-x}A(1),$$

and one expects, by symmetry, that $\alpha(x) = \pm\alpha(1-x)$.

The behaviours near 0 and 1 of $\alpha(x)$ are found to be

$$\alpha(x) \underset{x\tilde{\downarrow}0}{} x^{z-1}\left(-\frac{z^2}{2-z}A(1)2^{-(2-z)} + z4^z A(1/2)\right); \quad 2 > \mathrm{Re}(z) > 1,$$

$$\alpha(x) \underset{x\tilde{\downarrow}0}{} xA(1)\left(\frac{z^2}{2-z} + z\right) + x^{z-1}\left(-\frac{z}{2-z}A(1)2^z + z4^z A(1/2)\right); \quad \mathrm{Re}(z) > 2,$$

$$\alpha(x) \underset{x\tilde{\uparrow}1}{} (1-x)^{z-1}(zA(1)2^z - z4^z A(1/2)); \quad \mathrm{Re}(z) > 1.$$

When $2 > \mathrm{Re}(z) > 1$, assuming $\alpha(x) = -\alpha(1-x)$ forces the coefficients of $x^{z-1}((1-x)^{z-1})$ of the behaviours of $\alpha(x)$ near 0 and 1 to be opposite, which is possible only if $A(1) = 0$. In this case,

$$\alpha(x) \propto (1-2x)(x(1-x))^{z-1}, \tag{55}$$

which are indeed anti-symmetric continuous solutions. These solutions are eigenstates of L associated to the eigenvalues $\lambda = z^{-1}$. Maybe there are other symmetric solutions.

When $\mathrm{Re}(z) > 2$, necessarily the leading coefficients of x from the behaviour of $\alpha(x)$ near 0 is 0, forcing again $A(1) = 0$. In this case, $\alpha(x) \propto (1-2x)(x(1-x))^{z-1}$ are again anti-symmetric continuous solutions. A similar conclusion is obtained when $\mathrm{Re}(z) = 2$.

We conclude that when $\mathrm{Re}(z) \geq 1$, there are continuous solutions α (eigenstates) to $(I - zL)\alpha = 0$, defined up to a multiplicative constant. Recalling $\lambda = z^{-1}$, we get that the closed disk of $\mathbb{C}$ centered at $(1/2, 0)$

with radius $1/2$ (which is: $\mathrm{Re}\left(\lambda^{-1}\right) \geq 1$) constitutes the point spectrum of L. When λ belongs to the latter disk with radius $1/2$, $(\lambda I - L)^{-1}$ does not exist. Because there is a continuum of eigenvalues in the latter disk, the corresponding neutral Fleming-Viot model has no spectral gap. The points λ belonging to the complementary of the latter disk to the unit disk centered at 0 constitute the continuous spectrum, where $(\lambda I - L)^{-1}$ exists, but is not defined on the whole space $C_0([0, 1])$: The operator $\lambda I - L$ is not surjective.

- Assume finally $z = 1$ and c not identically 0.

Then, (52)-(53), with $f(x) = c(x) + \dfrac{x}{1-x} A(1)$,

$$A(x) = x(1-x)\int_{1/2}^{x} (y(1-y))^{-1} f(y)dy + 4x(1-x)A(1/2),$$

$$\alpha(x) = A(x)\left(\frac{1}{x} - \frac{1}{1-x}\right) + f(x)$$

$$= (1-2x)\int_{1/2}^{x} (y(1-y))^{-1} c(y)dy + A(1)(4x-1)$$

$$+ 4A(1/2)(1-2x) + c(x), \tag{56}$$

where the constants $A(1/2)$ and $A(1)$ should be determined from the imposed values $\alpha(0)$ and $\alpha(1)$ of α at the boundaries. $\alpha(x)$ in (56) solves

$$-(L-I)\alpha = c, \text{ if } x \in (0, 1); \quad \alpha = d, \text{ if } x \in \{0, 1\}, \tag{57}$$

and so α can be interpreted as the additive functional

$$\alpha(x) = \mathbf{E}_x\left[\sum_{t \geq 0} c(x_t) + d(x_\infty)\right].$$

Examples. (i) Let $\varepsilon > 0$, small and $I_\varepsilon = (\varepsilon, 1 - \varepsilon)$. Let $c(y) = 1(y \in I_\varepsilon)$ and suppose the initial condition x belongs to the interval I_ε. Then $\alpha(x)$ represents the expected time till x_t first exits out of the interval I_ε, starting from x within the interval. Using (56), we find

$$\alpha(x) = (1 - 2x)\log\frac{x}{1-x} + A(1)(4x - 1) + 4A(1/2)(1 - 2x) + 1.$$

Putting $\alpha(\varepsilon) = \alpha(1 - \varepsilon) = 0$ fixes the constants and we finally find

$$\alpha(x) = (1 - 2x)\log\frac{x}{1-x} - (1 - 2\varepsilon)\log\frac{\varepsilon}{1-\varepsilon},$$

which is of order $-\log\varepsilon$, with a symmetric initial condition dependent correcting term $(1 - 2x)\log\frac{x}{1-x}$, which is maximal when $x = 1/2$.

(ii) (Green function). Let $y \in (0, 1)$ and $I_\delta(y) = [y - \delta, y + \delta]$ be an interval of width 2δ, where δ was chosen small enough in such a way that $I_\delta(y) \subset (0, 1)$. Let $x \in (0, 1)$ but not to $I_\delta(y)$. Let $c_y(z) = 1(z \in I_\delta(y))$ and $d = 0$. Then $\alpha(x) =: \alpha_{I_\delta(y)}(x)$ represents the expected sojourn time spent by x_t in the interval $I_\delta(y)$, starting from x outside this interval. We look for solutions of α with boundary conditions $\alpha(0) = \alpha(1) = 0$ translating that $\{0, 1\}$ are absorbing states, so that if $x_0 \in \{0, 1\}$, x_t will never visit $I_\delta(y)$. We first look for an expression of $\alpha(x) =: \mathfrak{g}(x, y)$ when $c(z) = \delta_y(z)$, satisfying the same boundary conditions. Using (56) with

$$H(x) := \int_{1/2}^{x} (z(1 - z))^{-1}\delta_y(z)dz,$$

we find $\alpha(x) = (1 - 2x)H(x) - H(0) + (H(0) + H(1))x$. After some easy computations pertaining to $y < x$ or $y > x$, and recalling $m(y) = (y(1 - y))^{-1}$, we find the Green function $\alpha(x) =: \mathfrak{g}(x, y)$ as

$$\mathfrak{g}(x,\ y) = m(y)(1 - x),\ \text{if}\ \ y < x,$$

$$\mathfrak{g}(x,\ y) = m(y)x,\ \text{if}\ \ y > x.$$

Therefore, the expected time spent by x_t in $I_\delta(y)$, starting from $x \notin I_\delta(y)$ is

$$\alpha_{I_\delta(y)}(x) = \int_{I_\delta(y)} \mathfrak{g}(x,\ z)dz.$$

More generally, the solution to (57) is

$$\alpha(x) = \int_0^1 \mathfrak{g}(x,\ y)c(y)dy + d(0) + (d(1) - d(0))x,$$

where $\mathfrak{g}(x,\ y)$ is the Green kernel just defined satisfying $\mathfrak{g}(0,\ y) = \mathfrak{g}(1,\ y) = 0$.

□

6.3. Eigenpolynomials

Recalling

$$(L\psi)(x) = \mathbf{E}_x\psi(x_1) = \frac{1-x}{x}\int_0^x \psi(y)dy + \frac{x}{1-x}\int_x^1 \psi(y)dy,$$

let $\psi(x) = x^k$ be a monomial of degree $k \geq 1$. We have

$$(L\psi)(x) = \frac{1}{k+1}(x + \ldots + x^{k-1} + 2x^k),$$

and the action of L on x^k does not change the degree of the polynomial image. Thus, there are polynomials $u_k(x)$ of degree k such that, with $\lambda_k := 2/(k+1),\ k \geq 1$

$$(\lambda_k I - L)u_k = 0. \tag{58}$$

These values of λ are particular (real and rational) values of the point spectrum of L [note that $\lambda_k = \widehat{P}^{\infty}_{k,k}$ coincide with the diagonal terms of $\widehat{P}^{\infty}$].

When k is odd, one can check that $u_1(x) = x$ and

$$u_k(x) = (1-2x)(x(1-x))^{(k-1)/2}, \quad k \geq 3, \tag{59}$$

with u_k anti-symmetric: $u_k(x) = -u_k(1-x)$. Note that u_k is a special incarnation of (55) when $z = \lambda^{-1} = (k+1)/2$.

When k is even, (55) with $z = \lambda^{-1} = (k+1)/2$ is not a polynomial solution of (58). There exist other solutions u_ks, which are symmetric $(u_k(x) = u_k(1-x))$ polynomials, namely,

$$u_{2p}(x) = x(1-x)\sum_{q=1}^{p-1}(a_{q,p} + b_{q,p}(x(1-x))^q), \quad p \geq 1. \tag{60}$$

Here, $(a_{q,p}, b_{q,p})_{q=1,\ldots,p}$ constitute some sequences of real numbers, which can be computed recursively by iterated Euclidean division of u_{2p} by $x(1-x)$.

These polynomials all satisfy $u_k(0) = u_k(1) = 0$ for $k \geq 1$. For instance

$$u_1(x) = x,\ u_2(x) = x(1-x),\ u_3(x) = (1-2x)x(1-x),$$

$$u_4(x) = x(1-x)(-\frac{1}{8} + x(1-x)),\ u_5(x) = (1-2x)(x(1-x))^2,$$

$$u_6(x) = x(1-x)\left[x(1-x)(-\frac{1}{8} + x(1-x)) - \frac{1}{8\times 16}\right],$$

$u_7(x) = (1-2x)(x(1-x))^3$ are the seven first eigenpolynomials.

For all $\psi \in C_0([0,1])$ vanishing at $\{0,1\}$, with $\psi(x) = \sum_{l\geq 1} c_l u_l(x)$ the development of ψ along the complete set of polynomials u_l

$$(L^t\psi)(x) = \mathbf{E}_x\psi(x_t) = \sum_{l\geq 1}\left(\frac{2}{l+1}\right)^t c_l \cdot u_l(x).$$

We also note that the functions $v_k(y) = (y(1-y))^{-(k+1)/2}$ are eigenstates of the operator L^* associated to the eigenvalues $\lambda_k = 2/(k+1)$, $k \geq 1$: $(L^* v_k)(y) = \lambda_k v_k(y)$. In particular, $v_1(y) = (y(1-y))^{-1} = m(y)$, the speed measure density.

Examples. (i) From this, we easily get the dynamics of heterozygosity $\mathbf{E}_x(2x_t(1-x_t)) = 2\left(\frac{2}{3}\right)^t x(1-x)$, which tends to 0 exponentially fast as $t \to \infty$.

(ii) Observing $(2x_t(1-x_t))^2 = 4(u_4(x_t) + \frac{1}{8}u_2(x_t))$, the variance of heterozygosity is found to be

$$\boldsymbol{\sigma}_x^2(2x_t(1-x_t)) = 4\mathbf{E}_x\left[u_4(x_t) + \frac{1}{8}u_2(x_t)\right] - 4\mathbf{E}_x[u_2(x_t)]^2$$

$$= 4x(1-x)\left[\frac{1}{8}\left(\frac{2}{3}\right)^t + \left(x(1-x) - \frac{1}{8}\right)\left(\frac{2}{5}\right)^t - x(1-x)\left(\frac{2}{3}\right)^{2t}\right].$$

It starts at $t = 1$ from

$$\boldsymbol{\sigma}_x^2(2x_1(1-x_1)) = 4x(1-x)\left[\frac{1}{30} - \frac{2}{45}x(1-x)\right] > 0,$$

where it is not maximal $(\boldsymbol{\sigma}_x^2(2x_2(1-x_2)) > \boldsymbol{\sigma}_x^2(2x_1(1-x_1)))$ and then decays exponentially at rate $2/3$ when $t \to \infty$. The fluctuations start growing and then there is an intermediate time $t_* > 1$ at which they reach a maximum, before decaying to 0. Similar conclusions can be found in the context of Wright-Fisher diffusions, [26].

(iii) In particular also, if $\psi(x) = x^n$ and $x^n = \sum_{k=1}^n c_{k,n}u_k(x)$, then

$$(L^t\psi)(x) = \mathbf{E}_x(x_t^n) = \sum_{k=1}^n \left(\frac{2}{k+1}\right)^t c_{k,n} \cdot u_k(x).$$

Defining by duality the integral-valued process $\hat{x}_t$ by

$$\mathbf{E}_x\,(\,x_t^n\,) = \mathbf{E}_n\left(x^{\hat{x}_t}\right),\ \text{for all}\ (n,\,t)\in\mathbb{N}_+,\quad x\in[0,\,1], \tag{61}$$

we obtain the pgf $\mathbf{E}_n\left(x^{\hat{x}_t}\right)$ of $\hat{x}_t$ started at $\hat{x}_0 = n$. Proceeding as in [30] p. 65, [dealing with the well-known duality between Kingman coalescent and the Wright-Fisher diffusion], using martingale arguments $\hat{x}_t$ is seen to be precisely the limiting discrete-time Λ-coalescent with $\pi = u^{-2}\Lambda$ uniform on $[0,\,1]\,(\Lambda(du) = u^2 du)$. (61) is clearly already true for $t = 1$, using (41) and the above expression of $(L\psi)(x)$ when $\psi = x^n$ [3].

In particular,

$$[x]\,\mathbf{E}_n\left(x^{\hat{x}_t}\right) = [x]\,\mathbf{E}_x\,(\,x_t^n\,),$$

is the probability that $\hat{x}_t = 1$ (starting from $\hat{x}_0 = n$) or else that the time to most recent common ancestor $\hat{\tau}_{n,1}$ of $\hat{x}_t$ is $\le t$. More generally,

$$\mathbf{P}_n\,(\,\hat{x}_t = i\,) = [\,x^i\,]\,\mathbf{E}_x\,(\,x_t^n\,) = \sum_{k=1}^{n}\left(\frac{2}{k+1}\right)^t c_{k,n}\cdot[\,x^i\,]\,u_k\,(x).$$

Noting that $[x]u_k(x) = 0$ if k is odd $\ge 5\,([x]u_1(x) = [x]u_3(x) = 1)$, only the even terms essentially contribute to $\mathbf{P}\,(\,\hat{\tau}_{n,1} \le t\,) = \mathbf{P}_n\,(\,\hat{x}_t = 1\,) = [x]$ $\mathbf{E}_x\,(\,x_t^n\,)$ and the tail of $\hat{\tau}_{n,1}$ decays like $\left(\frac{2}{3}\right)^t$. □

6.4. Two conditionings

(i) Proceeding as for the Wright-Fisher diffusion, it is clear that the new process with the modified kernel

$$p(x;\,1,\,y) \to \overline{p}_1(x;\,1,\,y) := \frac{y}{x}\,p(x;\,1,\,y),$$

---

[3] This duality relationship is not limited to the special case. It carries over to the cases, where the density of U is not uniform.

corresponds to (42) conditioned on exit eventually at 1 (ultimate fixation of allele A_1). Note that $\overline{p}_1$ is of the Doob-transform type $\frac{\alpha(y)}{\alpha(x)}p(x;1,y)$, where $\alpha(x)=x$ is a harmonic function of L giving the probability that the original process x_t hits first the boundary $\{1\}$ before $\{0\}$ given $x_0=x$.

Let us call the new process with kernel $\overline{p}_1$ say $\widetilde{x}_t$. Its corresponding backward generator is

$$(\overline{L}\psi)(x)=\mathbf{E}_x\psi(\widetilde{x}_1)=\frac{1-x}{x^2}\int_0^x y\psi(y)dy+\frac{1}{1-x}\int_x^1 y\psi(y)dy. \qquad (62)$$

Note that $(\overline{L}1)(x)=1$, so $\overline{L}$ is a true stochastic kernel (no mass loss nor creation). In particular, the mean is

$$\mathbf{E}_x(\widetilde{x}_1)=\frac{1-x}{x^2}\int_0^x y^2dy+\frac{1}{1-x}\int_x^1 y^2dy=\frac{1}{3}(2x+1).$$

So $\widetilde{x}_t$ presents an additional drift, which is $\mathbf{E}_x(\widetilde{x}_1)-x=\frac{1}{3}(1-x)$.

Similarly, the process with the modified kernel

$$p(x;1,y)\to\overline{p}_0(x;1,y):=\frac{1-y}{1-x}p(x;1,y),$$

corresponds to (42) conditioned on exit eventually at 0 (ultimate extinction of allele A_1). This process presents an additional drift, which is $-\frac{1}{3}x$, pushing x_t towards 0.

(ii) Recall the eigenvector u_2 of L associated to the eigenvalue $\lambda_2=2/3$ is $u_2=x(1-x)$. Consider a new process with the modified kernel

$$p(x;1,y)\to\overline{p}(x;1,y):=\lambda_2^{-1}\frac{y(1-y)}{x(1-x)}p(x;1,y).$$

Let us call again the new process $\widetilde{x}_t$. Its corresponding backward generator is

$$(\overline{L}\psi)(x) = \mathbf{E}_x\psi(\widetilde{x}_1) = \frac{\lambda_2^{-1}}{x^2}\int_0^x y(1-y)\psi(y)dy + \frac{\lambda_2^{-1}}{(1-x)^2}\int_x^1 y(1-y)\psi(y)dy.$$

One can check that $(\overline{L}1)(x) = 1$, so $\overline{L}$ is again a true stochastic kernel (no mass loss nor creation). It corresponds to (42) conditioned on never hitting neither $\{0\}$ or $\{1\}$, even at $t = \infty$ (the so-called Q-process of x_t). In particular, the mean is $(\psi(y) = y)$

$$\left(\overline{L}\psi\right)(x) = \mathbf{E}_x(\widetilde{x}_1) = \frac{3}{4}\left(x + \frac{1}{6}\right).$$

So $\widetilde{x}_t$ presents an additional stabilizing drift towards $1/2$, which is $\mathbf{E}_x(\widetilde{x}_1) - x = \frac{1}{4}(\frac{1}{2} - x)$.

The limit law m of $\widetilde{x}_t$ obeys $\left(\overline{L}^* m\right)(y) = m(y)$, where

$$\left(\overline{L}^* m\right)(y) = \lambda_2^{-1} y(1-y)\left[\int_0^y \frac{m(x)dx}{(1-x)^2} + \int_y^1 \frac{m(x)dx}{x^2}\right].$$

One gets $m(y) \propto (y(1-y))^{-1/2}$, which is an integrable beta$(\frac{1}{2}, \frac{1}{2})$ density (the Arcsine law), suggesting that $\widetilde{x}_t$ is positive-recurrent.

6.5. Doob transforms

The example (i) is a particular Doob transform making use of the harmonic function $\alpha(x) = x$ of L. Let $\alpha \geq 0$ solve (57)

$$-(L - I)\alpha = c,$$

for some bounded c. If $c > 0\,(c < 0)$ on $(0, 1)$, α is called *super-harmonic* (*sub-harmonic*). It is harmonic if $c = 0$. With L the backward generator of x_t, define the generator of some new process $\overline{x}_t$ as

$$(\overline{L}\psi)(x) = \frac{1}{\alpha(x)} L(\alpha\psi)(x).$$

Note that the time iterates are obtained as $\left(\overline{L}^t\psi\right)(x) = \frac{1}{\alpha(x)} L^t(\alpha\psi)(x)$.

We have $(\overline{L}1)(x) - 1 = \frac{1}{\alpha(x)} L(\alpha)(x) - 1 = -c/\alpha =: \lambda(x)$, with $|\lambda(x)| < 1$.

Therefore,

$$(\overline{L}\psi)(x) = \left(\widetilde{L}\psi\right)(x) + \lambda(x)\cdot\psi,$$

where

$$\left(\widetilde{L}\psi\right)(x) = (I - (\overline{L}1)(x))\psi(x) + (\overline{L}\psi)(x)$$

$$= \psi(x) + \frac{1-x}{x\alpha(x)}\int_0^x \alpha(y)(\psi(y) - \psi(x))dy + \frac{x}{(1-x)\alpha(x)}\int_x^1 \alpha(y)(\psi(y) - \psi(x))dy,$$

is the backward generator of some new stochastic process $\widetilde{x}_t$, noting that $\left(\widetilde{L}1\right)(x) = 1$.

Depending on whether $\lambda(x) > 0\,(\lambda < 0)$ on $(0, 1)$, which is obtained when α is sub-harmonic (super-harmonic), the multiplicative term $\psi \to \lambda(x)\cdot\psi$ accounts either for (binary) branching or for killing of $\widetilde{x}_t$ inside $(0, 1)$, with probability $\pm\lambda(x)$, respectively. Therefore, $\overline{x}_t$ may be viewed as $\widetilde{x}_t$ with additional branching or killing inside the state-space. Note that $\overline{L} = \widetilde{L}$ when $c = 0$ (in the harmonic case).

Whenever c has no specific sign on $(0, 1)$, $\overline{x}_t$ may be viewed as $\widetilde{x}_t$ with both binary branching and killing. Because $|\lambda(x)| < 1$, one can uniquely write: $\lambda(x) = p_2(x) - p_0(x)$, where $p_2(x)$ interprets as the probability that the particle splits at x, $p_0(x)$ that it is killed at x, $p_0(x) + p_2(x) = 1$.

The kernel $\overline{K}(x, y)$ associated to the process with backward generator $\overline{L}$ is obtained from the substitution (K as in (49))

$$K(x, y) = p(x; 1, y) \to \overline{K}(x, y) = \overline{p}(x; 1, y) := \frac{\alpha(y)}{\alpha(x)} p(x; 1, y).$$

It allows to define the adjoint $\overline{L}^*$ of $\overline{L}$ as: $\left(\overline{L}^* m\right)(y) = \int_0^1 \overline{K}(x, y) m(x) dx$.

In this selection of paths construction of $\overline{x}_t$ through a change of measure, sample paths $x \to y$ of x_t with large $\alpha(y) / \alpha(x)$ are favoured.

6.6. Deviation from neutrality (drifts)

Consider the discrete-time Markovian dynamics with $x_0 = x$ and driven by $(U_t, V_t)_{t \geq 1}$

$$x_{t+1} = p(x_t) + U_{t+1}(1 - p(x_t)) 1(V_{t+1} \leq x_t) - U_{t+1} p(x_t) 1(V_{t+1} > x_t). \quad (63)$$

Here $x \to p(x)$ is an invertible non-decreasing mapping from $[0, 1]$ to an interval $I \subseteq [0, 1]$. Once the choice (governed by V) to move to the left or to the right is made (based on the current frequency x_t) the amplitude of the jump (governed by U) applies not to x_t, but to a (small) deformation $p(x_t)$ translating that some external shift (such as mutation or selection) occurred in the mean time. From these drift effects, x_t is no longer a martingale

$$\mathbf{E}(x_{t+1} \mid x_t = x) = p(x) + \frac{1}{2}(1 - p(x))x - \frac{1}{2} p(x)(1 - x) = \frac{1}{2}(x + p(x)).$$

As for the variance: $\sigma^2(x_{t+1} \mid x_t = x) = \sigma^2(U_{t+1}) \left[(1 - x)x + (p(x) - x)^2\right]$.

The backward generator associated to (63) is

$$(L\psi)(x) = \frac{x}{1 - p(x)} \int_{p(x)}^1 f\left(\frac{y - p(x)}{1 - p(x)}\right) \psi(y) dy$$

$$+\frac{1-x}{p(x)}\int_0^{p(x)} f\left(\frac{p(x)-y}{p(x)}\right)\psi(y)dy.$$

The corresponding forward generator L^*, which is the adjoint of L, acts on the space of positive Radon measures and it is easily seen to be given by

$$(L^*\mu)(y)=\int_0^{p^{-1}(y)}\frac{z}{1-p(z)}f\left(\frac{y-p(z)}{1-p(z)}\right)\mu(dz)$$

$$+\int_{p^{-1}(y)}^1\frac{1-z}{p(z)}f\left(\frac{p(z)-y}{p(z)}\right)\mu(dz). \tag{64}$$

When $f\equiv 1$ (U uniform) as in the special case,

$$(L^*\mu)(y)=\int_0^{p^{-1}(y)}\frac{z}{1-p(z)}\mu(dz)+\int_{p^{-1}(y)}^1\frac{1-z}{p(z)}\mu(dz).$$

In this latter case, clearly, there exists a (speed) measure μ with density m satisfying $(L^*\mu)(y)=\mu$ with m obeying the functional equation

$$m'(y)=p^{-1}(y)'\left(\frac{p^{-1}(y)}{1-y}-\frac{1-p^{-1}(y)}{y}\right)m(p^{-1}(y)).$$

Let us look at the classical examples.

Small mutations. Take $p(x)=\pi_1(1-x)+(1-\pi_2)x$, where (π_1,π_2) are very small (N-dependent) mutation probabilities from A_2 to A_1 (respectively, A_1 to A_2). Let $\pi:=\pi_1+\pi_2\ll 1$. Then

$$p^{-1}(y)=\frac{y-\pi_1}{1-\pi},\quad p^{-1}(y)'=\frac{1}{1-\pi}.$$

Approximating $m(p^{-1}(y))$ by $m(y)$ to leading order leads to the speed measure density, defined up to a multiplicative constant: $m(y)\propto y^{-\alpha_1}$

$(1-y)^{-\alpha_2}$, where $\alpha_1 := \frac{1-\pi_2}{(1-\pi)^2} \sim 1 + 2\pi_1 + \pi_2$ and $\alpha_2 := \frac{1-\pi_1}{(1-\pi)^2} \sim 1 + 2\pi_2 + \pi_1$. Both exponents of m are smaller than -1 and so m is not an integrable beta density (the jump process x_t with mutations is not ergodic). The α_is are increasing functions of the π_is and so increasing mutation probabilities puts more probability mass of x_t to the endpoints, which looks counter-intuitive. However, one can perhaps understand this as follows: Whenever x_t has got close to say 1, mutations tend to push x_t inside the interval and so to attenuate the amplitude $-U_{t+1}x_t$ of the rare resetting jump from x_t, whose large size, at the end, would have been chiefly responsible of positive recurrence.

Small selection. Take $p(x) = (1+s_1)x / (1 + s_1x + s_2(1-x))$, where (s_1, s_2) are small (N-dependent) fitness parameters of A_1 (respectively, A_2). Let $s = s_1 - s_2 > 0$ corresponding to a small fitness advantage of A_1 over A_2. We have

$$p(x) \sim x + sx(1-x),$$

and

$$p^{-1}(y) \sim y - 2sy^2, \quad p^{-1}(y)' \sim 1 - 4sy.$$

To the dominant order in s, this leads to the speed measure density, defined up to a multiplicative constant

$$m(y) \propto \frac{1}{y(1-y)}(1-y)^{-6s}e^{10sy}.$$

It is biased to the right (allele A_1 is eventually favoured) and not integrable.

Acknowledgement

During the course of the elaboration of this draft, the author benefited in various places from fruitful discussions with Anthony Quas (Victoria U.), Alexandre Jollivet (Cergy U.), Servet Martinez (U. de Chile), Pierre Collet (X, Palaiseau), and Martin Möhle (Tübingen U.). Many thanks to all of them. The author also acknowledges partial support from the ANR Modélisation Aléatoire en Ecologie, Génétique et Evolution (ANR-Manége- 09-BLAN-0215 project).

References

[1] M. Benhenda, A model of deliberation based on Rawls's political liberalism, Social Choice and Welfare, Springer, 36 (2011), 121-178.

[2] M. Birkner and J. Blath, Computing likelihoods for coalescents with multiple collisions in the infinitely many sites model, J. Math. Biol. 57(3) (2008), 435-465.

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

[4] C. Cannings, The latent roots of certain Markov chains arising in genetics: A new approach, II, Further haploid models, Advances in Appl. Probability 7 (1975), 264-282.

[5] J. F. Crow and M. Kimura, An Introduction to Population Genetics Theory, Harper & Row, Publishers, New York-London, 1970.

[6] E. B. Dynkin, Markov Processes, Vols. I, II, Translated with the authorization and assistance of the author by J. Fabius, V. Greenberg, A. Maitra and G. Majone, Die Grundlehren der Mathematischen Wissenschaften, Bände 121, 122, Academic Press Inc., Publishers, New York; Springer-Verlag, Berlin-Göttingen-Heidelberg, Vol. I: xii+365 pp.; Vol. II: viii+274 pp., 1965.

[7] B. J. Eldon and J. Wakeley, Coalescent processes when the distribution of offspring number among individuals is highly skewed, Genetics 172 (2006), 2621-2633.

[8] S. N. Ethier and T. G. Kurtz, Markov Processes, Characterization and Convergence, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986.

[9] S. N. Ethier and T. G. Kurtz, Fleming-Viot processes in population genetics, SIAM J. Control Optim. 31(2) (1993), 345-386.

[10] W. J. Ewens, Mathematical Population Genetics, I, Theoretical Introduction, Second Edition, Interdisciplinary Applied Mathematics, 27, Springer-Verlag, New York, 2004.

[11] W. Feller, The parabolic differential equations and the associated semi-groups of transformations, Ann. of Math. 55(2) (1952), 468-519.

[12] J. H. Gillespie, The Causes of Molecular Evolution, Oxford University Press, New York and Oxford, 1991.

[13] R. C. Griffiths, The frequency spectrum of a mutation, and its age, in a general diffusion model, Theoretical Population Biology 64(2) (2003), 241-251.

[14] T. Huillet, On Wright-Fisher diffusion and its relatives, J. Stat. Mech., Th. and Exp. 11 (2007), 11006.

[15] T. Huillet and M. Möhle, Erratum on population genetics models with skewed fertilities: A forward and backward analysis [hal-00646215/fr], (2011).

[16] T. Huillet and M. Möhle, Population genetics models with skewed fertilities: A forward and backward analysis, Stochastic Models 27 (2011), 521-554.

[17] T. Huillet and M. Möhle, On the extended Moran model and its relation to coalescents with multiple collisions, Theor. Popul. Biol., Online first papers, to appear in (2012).

[18] S. Karlin and J. McGregor, Direct product branching processes and related Markov chains, Proc. Nat. Acad. Sci. USA 51 (1964), 598-602.

[19] S. Karlin and H. M. Taylor, A Second Course in Stochastic Processes, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1981.

[20] B. V. Khvedelidze, Fredholm Theorems for Integral Equations, in Hazewinkel, Michiel, Encyclopedia of Mathematics, Springer, 2001.

[21] M. Kimura, On the probability of fixation of mutant genes in a population, Genetics 47 (1962), 713-719.

[22] M. Kimura and T. Ohta, The age of a neutral mutant persisting in a finite population, Genetics 75 (1973), 199-212.

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

[24] A. Kolmogorov, S. Fomine and V. M. Tihomirov, Eléments de la théorie des fonctions et de l'analyse fonctionnelle, (French) Avec un complément sur les algèbres de Banach, par V. M. Tikhomirov, Éditions Mir, Moscow, 536 pp, 1974.

[25] P. Mandl, 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.

[26] T. Maruyama, Stochastic Problems in Population Genetics, Lecture Notes in Biomathematics, 17, Springer-Verlag, Berlin-New York, 1977.

[27] M. Möhle and S. Sagitov, A classification of coalescent processes for haploid exchangeable population models, Ann. Probab. 29(4) (2001), 1547-1562.

[28] P. A. P. Moran, Random processes in genetics, Proc. Cambridge Philos. Soc. 54 (1958), 60-71.

[29] T. Nagylaki, The moments of stochastic integrals and the distribution of sojourn times, Proc. Nat. Acad. Sci. USA 71 (1974), 746-749.

[30] E. Pardoux, Probabilistic models of population genetics, 2009.

www.latp.univ-mrs.fr/~pardoux/enseignement/cours_genpop.pdf

[31] A. Pinkus, Spectral Properties of Totally Positive Kernels and Matrices, In: Total Positivity and its Applications, M. Gasca and A. Micchelli (Eds), 1-35, 1995.

[32] J. Pitman, Coalescents with multiple collisions, Ann. Probab. 27(4) (1999), 1870-1902.

[33] S. Sagitov, The general coalescent with asynchronous mergers of ancestral lines, J. Appl. Probab. 36(4) (1999), 1116-1125.

[34] S. Tavaré, Ancestral Inference in Population Genetics, Lectures on Probability Theory and Statistics, Saint-Flour 2001, Lecture Notes in Math., 1837, (1-188) Springer, 2004.

■