ON WRIGHT-FISHER DIFFUSION AND ITS RELATIVES

Thierry Edmond Arnold Huillet

ON WRIGHT-FISHER DIFFUSION AND ITS RELATIVES

THIERRY HUILLET

Abstract. We present a series of elementary stochastic models arising from population genetics. Various mathematical aspects are investigated at the intuitive level. Special emphasis is put on the diffusion method. Focus is on the Wright-Fisher diffusion model and its variations, describing the forward evolution of one colony undergoing random sampling, possibly under the additional forces arising from mutation, selection with or without dominance. Some aspects of their dual coalescents obtained while running the diffusion process backward in time are also investigated.

Keywords: Mutational and evolutionary processes (theory), Population dynamics (Theory), Phylogeny (Theory).

1. Introduction

The goal of this manuscript is to present a series of elementary stochastic models from population dynamics which found their way over the last sixty years, chiefly in mathematical population genetics. Related biological material and various mathematical techniques are discussed at the simple intuitive level. Special emphasis is put on the diffusion method with a tentative emphasis on the underlying unity of various problems, based on Kolmogorov backward and forward equations. Most of the manuscript's content focuses on the specific Wright-Fisher (WF) diffusion model and its variations, describing the evolution of one colony undergoing random mating, possibly under the additional actions of mutation, selection with or without dominance. Some aspects of the coalescents obtained while running the diffusion process backward in time are also investigated, using duality techniques. A non-exhaustive list of references to the vast existing literature will be given, when necessary, in the body of the text. We now describe the content of this work in some more details.

Section 2 is devoted to generalities on one-dimensional diffusions. It is designed to fix the background and notations. Special emphasis is put on the Kolmogorov backward and forward equations, while stressing the crucial role played by the boundaries in such one-dimensional diffusion problems. Some questions such as the meaning of speed and scale functions, existence of an invariant measure, validity of detailed balance, are addressed in the light of Feller classification of boundaries. The important problem of evaluating additive functionals along sample paths is then briefly discussed, emphasizing the prominent role played by the Green function of the model; several simple illustrative examples are supplied. As a by-product, the transformation (selection) of sample paths techniques, deriving from specific

additive functionals, are next briefly introduced in the general diffusion context. Some transformations of interest are then investigated, together with the problem of evaluating additive functionals of the transformed diffusion process itself.

Roughly speaking, the transformation of paths procedure allows to select sample paths of the original process with, say, a fixed destination and/or, more generally, to kill certain sample paths that do not fit the integral criterion encoded by the additive functional. One should therefore see it as a selection of paths procedure leading to new processes described by an appropriate modification of the infinitesimal generator of the original process. It turns out therefore that the same diffusion methods used in the previous discussions apply to the transformed processes, obtained after a change of measure. When particularized to the WF model, these new processes favoring large values of the additive functional will reveal some biological phenomena and problems of interest, examples of which are discussed below in some detail.

Section 3 is concerned with the specific Wright-Fisher diffusion model and its relatives, allowing various drifts of biological interest to force the neutral WF model in specific directions. These continuous space-time models can be obtained as scaling limits of a biased discrete Galton-Watson model with a conservative number of offsprings over generations. The purpose of this Section is to illustrate the general techniques introduced in Section 2, allowing to address some important questions raised in population genetics, such as for example: times to extinction and/or fixation of an allele, dynamics of heterozygosity, time spent in some frequency range, fixation probability when extinction is most likely, conditional limiting frequency distributions given neither fixation nor extinction occurred in the past... The selection of paths procedure based on specific additive functionals is next illustrated in the following problems, starting with the simplest: neutral WF diffusion conditioned on exit at some boundary, neutral WF sample paths favoring a large exit time, selection of WF sample paths with large heterozygosity, selection of WF sample paths with large sojourn time density at some point of the state-space, conditioned Wright-Fisher diffusion with irreversible (one-way) mutation. In solving some of these problems, some use is made of the explicit spectral decomposition of the neutral WF Kolmogorov infinitesimal generators. In each cases, the characteristics of the transformed process are discussed and analyzed.

Last Section is devoted to the various coalescents obtained while running the Wright-Fisher diffusions backward in time. It makes use of the duality techniques. The advantage is that the structure of the dual process is often of great simplicity as compared to the one of WF diffusions themselves. We illustrate these classical ideas firstly on Kingman coalescent (the dual pure death process of the pure random genetic drift model), then on the ancestral selection graph (the dual birth and death process) when both mutation and selection are present in WF model, and finally on the ancestral graph of WF with selective dominance, in some range of the dominance parameter. Some of their mathematical features are briefly analyzed.

2. Preliminaries on diffusions

Before particularizing our study to the Wright-Fisher model and its relatives, we start with generalities on one-dimensional diffusions. For more technical details, we refer to [3], [4], [9] and [11]. This Section is designed to fix the background and notations for the rest of the paper.

2.1. **Generalities on one-dimensional diffusions on an interval.** Let $(w_t; t \geq 0)$ be a standard one-dimensional Brownian (Wiener) motion. Consider a $1-$dimensional Itô diffusion driven by $(w_t; t \geq 0)$ on an interval $I = [a, b]$ with $-\infty \leq a < b \leq \infty$. Assume it has locally Lipschitz continuous drift $f(x)$ and local standard deviation (volatility) $g(x)$, namely consider the stochastic differential equation (SDE):

$$dx_t = f(x_t)\, dt + g(x_t)\, dw_t,\ x_0 = x \in (a, b). \tag{1}$$

The condition on $f(x)$ and $g(x)$ guarantees in particular that there is no point x_* in the interior $\overset{\circ}{I} := (a, b)$ of I for which $|f(x)|$ or $|g(x)|$ would blow up and diverge as $|x - x_*| \to 0$.

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

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

In the definition of the mathematical expectation u, we have $t \wedge \tau(x) := \inf(t, \tau(x))$ where $\tau(x)$ indicates a random time at which the process should eventually be stopped, given the process was started at x. The description of this (adapted) explosion time is governed by the type of boundaries which $\{a, b\}$ are to $(x_t; t \geq 0)$. We shall return to this point later.

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

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

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

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

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

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

$$\varphi(x) = a + (b-a)\frac{\int_a^x e^{-2\int_a^y \frac{f(z)}{g^2(z)}dz}dy}{\int_a^b e^{-2\int_a^y \frac{f(z)}{g^2(z)}dz}dy},$$

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

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

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

the novel diffusion $(w_\theta := y_{t_\theta}; \theta \geq 0)$ is easily checked to be identical in law to a standard Brownian motion. Let now $\delta_y(\cdot) = \text{weak}-\lim_{\varepsilon\downarrow 0} \frac{1}{2\varepsilon}\mathbf{1}(\cdot \in (y-\varepsilon, y+\varepsilon))$ stand for the Dirac delta mass at y. The random time θ_t can be expressed as

$$\theta_t = \int_a^b dx \cdot m(x) \int_0^t \delta_{\varphi(x)}(w_s)\, ds = \int_0^t m\left(\varphi^{-1}(w_s)\right) ds$$

where $m(x) := 1/\left(g^2\varphi'\right)(x)$ is the (positive) speed density at $x = \varphi^{-1}(y)$ and $L_t(y) := \lim_{\varepsilon\downarrow 0} \frac{1}{2\varepsilon}\int_0^t \mathbf{1}(w_s \in (y-\varepsilon, y+\varepsilon))\, ds$ the local time at y of Brownian motion before time t. Both scale function φ and speed measure $m(x)\cdot dx$ therefore are essential ingredients to reduce the original stochastic process $(x_t; t \geq 0)$ to standard Brownian motion $(w_t; t \geq 0)$. Indeed, it follows from the above arguments that if $\theta_t = \int_0^t m(x_s)\, ds$, then $(\varphi(x_{\theta_t}); t \geq 0)$ is a Brownian motion.

Examples (from population genetics, with $I = [0, 1]$):

• Assume $f(x) = 0$ and $g^2(x) = x(1-x)$. This is the neutral Wright-Fisher (WF) model discussed at length later. This diffusion is already in natural scale and $\varphi(x) = x$, $m(x) = [x(1-x)]^{-1}$. The speed measure is not integrable. Due to symmetries of this particular diffusion, we observe that defining $\overline{x}_t := 1 - x_t$, $(\overline{x}_t; t \geq 0)$ with $\overline{x}_0 = \overline{x} := 1 - x$ obeys the same SDE as $(x_t; t \geq 0)$.

• With $u_1, u_2 > 0$, assume $f(x) = u_1 - (u_1 + u_2)x$ and $g^2(x) = x(1-x)$. This is the Wright-Fisher model with mutation. Parameters u_1, u_2 interpret as mutation rates. The drift vanishes when $x = u_1/(u_1 + u_2)$ which is an attracting point for the dynamics. Here:

$\varphi'(y) = \varphi'(y_0)\, y^{-2u_1}(1-y)^{-2u_2}$, $\varphi(x) = \varphi(x_0) + \varphi'(y_0)\int_{x_0}^x y^{-2u_1}(1-y)^{-2u_2}\, dy$, with $\varphi(0) = -\infty$ and $\varphi(1) = +\infty$ if $u_1, u_2 > 1/2$. The speed measure density is $m(x) \propto x^{2u_1-1}(1-x)^{2u_2-1}$ and so is always integrable.

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

• The WF model for which $f(x) = \sigma x(1-x) + u_1 - (u_1 + u_2)x$ and $g^2(x) = x(1-x)$ is called WF model with mutations and selection (σ, u_1, u_2).

We have: $\varphi(x) = \varphi(x_0) + \varphi'(y_0)\int_{x_0}^{x} e^{-2\sigma y} y^{-2u_1}(1-y)^{-2u_2}\,dy$ and speed density $m(x) \propto x^{2u_1-1}(1-x)^{2u_2-1} e^{2\sigma x}$ is integrable.

Considering as for the WF model, $\overline{x}_t := 1 - x_t$, one can check that the diffusion which governs $(\overline{x}_t; t \geq 0)$ is the in the same class as the one governing $(x_t; t \geq 0)$ after the following substitutions: $(\sigma, u_1, u_2) \to (-\sigma, u_2, u_1)$.

• (diploidy) The WF model for which $f(x) = \sigma x(1-x)(h - x(2h-1))$ and $g^2(x) = x(1-x)$ is called WF model with selection and dominance $h \in \mathbf{R}\backslash\left\{\frac{1}{2}\right\}$. When $h > 1$ (overdominance), the drift vanishes at $x_* = h/(2h-1)$ which lies inside $(1/2, 1)$. When $\sigma > 0$ ($\sigma < 0$), x_* is a stable (unstable) equilibrium point for the underlying deterministic dynamics. We have $\varphi(x) = \frac{\int_0^x e^{\widetilde{\sigma}(y-x_*)^2} dy}{\int_0^1 e^{\widetilde{\sigma}(y-x_*)^2} dy}$ and $m(x) \propto [x(1-x)]^{-1} e^{-\widetilde{\sigma}(x-x_*)^2}$ where $\widetilde{\sigma} := \sigma(2h-1)$. When passing from $(x_t; t \geq 0)$ to $(\overline{x}_t; t \geq 0)$ for a WF model with selection and dominance, one remains in the same class of models after the substitution $(\sigma, h) \to (-\sigma, 1-h)$ in parameter space (from which $x_* \to 1 - x_*$). Finally, considering the random time change: $\theta \to t_\theta$ defined by

$$\theta = \int_0^{t_\theta} x_s(1-x_s)\,ds,$$

the novel process $(y_\theta := x_{t_\theta} - \sigma x_*; \theta \geq 0)$ is governed by the familiar Ornstein-Uhlenbeck linear SDE: $dy_\theta = -\widetilde{\sigma} y_\theta d\theta + dw_\theta$, $y_0 = x_0 - \sigma x_*$.

Probability density: Assume $f(x)$ and $g(x)$ are now differentiable in $\overset{\circ}{I}$. Let then $p(x; t, y)$ stand for the transition probability density function 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 (Fokker-Planck) equation (KFE):

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

where $G^*(\cdot) = -\partial_y(f(y)\cdot) + \frac{1}{2}\partial_y^2\left(g^2(y)\cdot\right)$ is the adjoint of G (G^* acts on the terminal value y whereas G acts on the initial value x). In general, $p(x; t, y)$ is a sub-probability because, letting $\overline{\pi}_t(x) := \int_0^1 p(x; t, y)\,dy$, we have $\overline{\pi}_t(x) = \mathbf{P}(\tau(x) > t)$ and this tail distribution is different from 1 unless stopping time $\tau(x) = \infty$ with probability 1.

For one-dimensional diffusions, the transition density $p(x; t, y)$ is reversible with respect to the speed density ([9], Chapter 15, Section 13) and so detailed balance holds:

$$m(x)\,p(x; t, y) = m(y)\,p(y; t, x),\ a < x, y < b.$$

The speed density $m(y)$ satisfies $G^*(m) = 0$. It may be written as: $m(y) \propto e^{-U(y)}$ where potential function $U(y)$ reads:

$$U(y) := 2\int_a^y \frac{(gg')(z) - f(z)}{g^2(z)} dz,\ a < y < b.$$

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

substitution $t-s\to t$, p itself solves KBE

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

In particular, integrating over y, $\partial_t \overline{\pi}_t(x) = G(\overline{\pi}_t(x))$, with $\overline{\pi}_0(x)=\mathbf{1}(x\in(0,1))$. Defining the normalized conditional probability density $p^c(x;t,y):=p(x;t,y)/\overline{\pi}_t(x)$, now with total mass 1, we get

$$\partial_t p^c = -\partial_t \overline{\pi}_t(x)/\overline{\pi}_t(x)\cdot p^c + G^*(p^c)\,,\ p^c(x;0,y)=\delta_y(x)\,.$$

The term $\rho_t(x):=-\partial_t\overline{\pi}_t(x)/\overline{\pi}_t(x)>0$ is the time-dependent rate at which mass should be created to compensate the loss of mass of the original process due, say, to absorption of $(x_t;t\geq 0)$ at the boundaries. In the creation of mass process, a diffusing particle dies at rate $\rho_t(y)$ at point (t,y) where it is duplicated in two independent particles both started at y, evolving in the same diffusive way.

Let us draw again attention on KFE for p. One has $\partial_t p=-\partial_y J$ where $J=f(y)\,p(x;t,y)-\frac{1}{2}\partial_y\left(g^2(y)\,p(x;t,y)\right)$ is the probability current at (t,y). Assuming a stationary solution $p_{st}(y)$ to exist, independently of the initial condition, it must solve $f(y)\,p_{st}(y)-\frac{1}{2}\partial_y\left(g^2(y)\,p_{st}(y)\right)=J_{st}$ where J_{st} is the probability current at some boundary $\{a,b\}$. Integrating, one gets

$$p_{st}(y)=\frac{1}{g^2(y)}e^{2\int_{y_0}^{y}\frac{f(z)}{g^2(z)}dz}\left[C-2J_{st}\int_{y_0}^{y}dze^{2\int_{y_0}^{z}\frac{f(x)}{g^2(x)}dx}\right]$$

which reduces to $p_{st}(y)=\frac{C}{g^2(y)}e^{2\int_{y_0}^{y}\frac{f(z)}{g^2(z)}dz}\propto m(y)$ if $J_{st}=0$ at both boundaries. In this case, one can check that the probability current vanishes at all points of $[a,b]$. The constant C is the normalization constant which eventually renders $p_{st}(y)$ of total mass 1.

To take an example, the WF model with mutation rates $u_1,u_2>0$ has invariant measure on $[0,1]$: $\frac{\Gamma(2(u_1+,u_2))}{\Gamma(2u_1)\Gamma(2,u_2)}y^{2u_1-1}(1-y)^{2u_2-1}$ which is an integrable probability density known as the beta$(2u_1,2u_2)$ density.

2.2. **Feller classification of boundaries.** The KBE equation may not have unique solutions, unless one specifies the conditions at the boundaries $\{a,b\}$.

For 1−dimensional diffusions as in (1) on $[a,b]$, the boundaries $\partial I:=\{a,b\}$ 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 of the scale function and the speed measure turn out to be essential in the classification of boundaries due to Feller [6].

In the sequel, the symbol $\circ$ will designate either a or b. We shall say that a function $f(y)\in L_1(y_0,\circ)$ if $-\infty<\int_{y_0}^{\circ}f(y)\,dy<+\infty$.

$(A1)$ The boundary $\circ$ is a regular boundary if $\forall y_0\in(a,b)$:

$$(i)\ \ \varphi'(y)\in L_1(y_0,\circ)\ \text{ and }\ (ii)\ \ \frac{1}{(g^2\varphi')(y)}\in L_1(y_0,\circ)$$

In this case, a sample path of $(x_t;t\geq 0)$ can reach $\circ$ from the interior $\overset{\circ}{I}$ of I and reenter inside I, in finite time. The WF model with mutation has both regular

boundaries whenever $u_1, u_2 < 1/2$.

Remarks:

(i) If $\circ$ is not a regular boundary, it is unbridgeable and a sample path of $(x_t; t \geq 0)$ will never quit nor reenter I at $\circ$. For such an unbridgeable boundary at least, for all $t > 0$: $f(y)\, p(x;t,\circ) - \frac{1}{2}\partial_y \left(g^2(y)\, p(x;t,\circ)\right) = 0$ and the probability current vanishes at $(t,\circ)$.

(ii) For diffusion processes with regular boundaries, one may think in some cases that allowing the particle to quit the definition domain I and reentering later on, lacks physical meaning. In this case, if $\circ$ is found to be a regular boundary, one may force it *a posteriori* to be a reflecting or absorbing barrier or a mixture of them. In this case, one needs to impose boundary conditions on KBE at $\circ$; we shall return to this point later. $\diamondsuit$

$(A2)$ The boundary $\circ$ is an exit boundary if $\forall y_0 \in (a,b)$:

$$(i)\ \frac{1}{(g^2\varphi')(y)} \notin L_1(y_0,\circ) \text{ and } (ii)\ \varphi'(y) \int_{y_0}^{y} \frac{1}{(g^2\varphi')(z)} dz \in L_1(y_0,\circ)$$

In this case, a sample path of $(x_t; t \geq 0)$ can reach $\circ$ from the inside of I in finite time but cannot reenter. The sample paths are absorbed at $\circ$. There is an explosion at $\circ$ at time $\tau_\circ(x) = \inf(t > 0 : x_t = \circ \mid x_0 = x)$ and $\mathbf{P}(\tau_\circ(x) < \infty) = 1$. Whenever both boundaries $\{a,b\}$ are absorbing, the diffusion x_t should be stopped at $\tau(x) = \tau_a(x) \wedge \tau_b(x)$. When at least one of the boundaries is an exit boundary, the diffusion is transient and the process stops with probability 1 when hitting one of these exit boundaries. Whenever none of the boundaries $\{a,b\}$ is absorbing, $\tau(x) = +\infty$. Examples of diffusion with exit boundaries is WF model and WF model with selection.

$(I1)$ The boundary $\circ$ is an entrance boundary if $\forall y_0 \in (a,b)$:

$$(i)\ \varphi'(y) \notin L_1(y_0,\circ),\ (ii)\ \frac{1}{(g^2\varphi')(y)} \in L_1(y_0,\circ)$$

$$(iii)\ \frac{1}{(g^2\varphi')(y)} \int_{y_0}^{y} \varphi'(z)\, dz \in L_1(y_0,\circ).$$

An entrance boundary clearly is not a regular boundary.

In case $\circ$ is entrance, a sample path of $(x_t; t \geq 0)$ can enter from $\circ$ to the interior of $[a,b]$ but cannot return to $\circ$ from the interior of $[a,b]$. The WF model with mutation has both entrance boundaries whenever $u_1, u_2 > 1/2$.

When both boundaries are entrance boundaries, the diffusion $(x_t; t \geq 0)$ is positive recurrent inside $[a,b]$; note that condition (ii) guarantees the integrability of the (unique) invariant measure. In natural coordinate, $(y_t = \varphi(x_t); t \geq 0)$ is a diffusion in $\mathbf{R}$, since $\varphi(a) = -\infty$ and $\varphi(b) = +\infty$.

$(I2)$ The boundary $\circ$ is natural in all other cases. When $\circ$ is natural, sample paths cannot enter nor quit $[a,b]$ and sample paths are trapped inside $[a,b]$ with

$\{a,b\}$ inaccessible; the 'simplest' case is when $(x_t; t \geq 0)$ is itself a Brownian motion.

Restriction of the diffusion to a sub-interval: Let $I_* \subset I$ with $I_* := [x_*, x^*]$ and $-\infty \leq a \leq x_* < x^* \leq b \leq \infty$. It is sometimes of interest to consider the restriction of the diffusion (1) to the interval I_*. Then, one must specify what happens to the diffusing particle whenever it hits x_* or x^*, since one of the new boundaries at least is different from $\{a,b\}$. In the sequel, the symbol $*$ will designate either x_* or x^* whenever it differs from the pair $\{a,b\}$. It is often of interest to impose that the particle is either reflected (with probability π) or absorbed (with probability $1-\pi$) at $*$. In this case, the KBE equation with infinitesimal generator G must be considered on I_* together with the additional boundary condition(s): $\pi\partial_x u(*,t) + (1-\pi)u(*,t) = 0$, for all $t \geq 0$.

For instance, if both x_* and x^* differ from $\{a,b\}$, assuming x_* is reflecting and x^* absorbing, then $\partial_x u(x_*,t) = 0$ and $u(x^*,t) = 0$, $t \geq 0$. Note that reflecting or absorbing barriers are not regular barriers. Therefore the probability current vanishes at these points. The canonical coordinate to consider is the restriction of φ to I_*; finally, note that necessarily $\varphi(x_*) > -\infty$ and $\varphi(x^*) < +\infty$, whenever x_* and x^* both differ from $\{a,b\}$.

Remark: When $I = I_*$, it still makes sense to impose additional boundary conditions only if one of the boundaries $\circ$ is a regular boundary. By imposing a reflecting/absorbing condition at $\circ$: $\pi\partial_x u(\circ,t) + (1-\pi)u(\circ,t) = 0$, for all $t \geq 0$, we force the diffusion inside I although its natural tendency is to quit I to reenter later on. $\diamondsuit$

2.3. **Evaluation of additive functionals along sample paths.** Let $(x_t; t \geq 0)$ be the diffusion model defined by (1) on the interval I where both endpoints are assumed absorbing (exit). We wish to evaluate the non-negative additive quantities

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

where functions c and d are both assumed non-negative. As is well-known, functional $\alpha(x) \geq 0$ solves:

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

Important examples are:

1. Assume $c = 1$ and $d = 0$: here, $\alpha = \mathbf{E}^x\tau(x)$ is the mean time of explosion (average time spent in I before explosion), solution to:

$$\begin{aligned} -G(\alpha) &= 1 \text{ if } x \in \overset{\circ}{I} \\ \alpha &= 0 \text{ if } x \in \partial I. \end{aligned}$$

More generally, if $\alpha_n := \mathbf{E}^x\left[\tau(x)^n\right]$ is the $n-$th moment of $\tau(x)$, by Nagylaki formula

$$-G(\alpha_n) = n\alpha_{n-1},\ \alpha_0 = 1,$$

allowing to compute recursively α_n once α_{n-1} is known.

2. Assume $c(x) = \mathbf{1}(x \in (x_*, x^*))$ where $a < x_* < x^* < b$ and $d = 0$: here, $\alpha = \mathbf{E}^x\left(\int_0^{\tau(x)} \mathbf{1}(x_s \in (x_*, x^*))\, ds\right)$ is the mean time spent in (x_*, x^*) before explosion, solution to:

$$\begin{aligned} -G(\alpha) &= \mathbf{1}(x \in (x_*, x^*)) \text{ if } x \in \overset{\circ}{I} \\ \alpha &= 0 \text{ if } x \in \partial I. \end{aligned}$$

If $x \in (a, x_*]$ or $x \in [x^*, b)$ and if $x_* - a$ or $b - x^*$ is small, one expects $\alpha(x)$ to be short because the diffusion starts at a point close to a boundary where it is likely to explode and should therefore find little time to visit (x_*, x^*).

3. Whenever both $\{a, b\}$ are exit boundaries, it is of interest to evaluate the probability that x_t first hits $[a, b]$ (say) at b, given $x_0 = x$. This can be obtained by choosing $c = 0$ and $d(\circ) = \mathbf{1}(\circ = b)$.

Let then $\alpha =: \alpha_b(x) = \mathbf{P}(x_t \text{ first hits } [a,b] \text{ at } b \mid x_0 = x)$. $\alpha_b(x)$ is a $G-$harmonic function solution to $G(\alpha_b) = 0$, with boundary conditions $\alpha_b(a) = 0$ and $\alpha_b(b) = 1$. Solving this problem, we get: $\alpha_b(x) = \int_a^x dy e^{-2\int_a^y \frac{f(z)}{g^2(z)}dz} / \int_a^b dy e^{-2\int_a^y \frac{f(z)}{g^2(z)}dz}$.

On the contrary, choosing $\alpha_a(x)$ to be a $G-$harmonic function with boundary conditions $\alpha_a(a) = 1$ and $\alpha_a(b) = 0$, $\alpha_a(x) = \mathbf{P}(x_t \text{ first hits } [a,b] \text{ at } a \mid x_0 = x) = 1 - \alpha_b(x)$.

4. Let $y \in \overset{\circ}{I}$ and put $c = \frac{1}{2\varepsilon}\mathbf{1}(x \in (y-\varepsilon, y+\varepsilon))$ and $d = 0$. As $\varepsilon \downarrow 0$, c converges weakly to $\delta_y(x)$ and, $\alpha =: \mathfrak{g}(x,y) = \mathbf{E}^x\left(\lim \frac{1}{2\varepsilon}\int_0^{\tau(x)} \mathbf{1}(x_s \in (y-\varepsilon, y+\varepsilon))\, ds\right) = \int_0^\infty p(x; s, y)\, ds$ is the Green function, solution to:

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

$\mathfrak{g}$ therefore is the mathematical expectation of the local time at y, starting from x (the sojourn time density at y). The solution is easily seen to be

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

The Green function is of particular interest to solve the general problem of evaluating additive functionals $\alpha(x)$. Indeed, one easily finds

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

Consider now the restriction of $(x_t; t \geq 0)$ to the interval I_*. Assume $a < x_* < x^* < b$ and that x_* is reflecting and x^* absorbing. With $x \in \overset{\circ}{I}_*$, we now wish to evaluate the quantities

$$\alpha(x) = \mathbf{E}^x \left(\int_0^{\tau(x)} c(x_s)\, ds \right),$$

where $\tau(x)$ now is the exit time of I_*, necessarily at x^*, starting from x. $\alpha(x)$ now solves:

$$-G(\alpha) = c \text{ if } x \in \overset{\circ}{I}_*$$

whose solution in the bulk is

$$\alpha(x) = \int_{\overset{\circ}{I}_*} \mathfrak{g}(x,y)\, c(y)\, dy \text{ if } x \in \overset{\circ}{I}_*$$

The Green function in this particular case also solves $-G(\mathfrak{g}) = \delta_y(x)$ if $x \in \overset{\circ}{I}_*$, but now with the additional boundary conditions $\partial_x \mathfrak{g}(x_*, y) = 0$ and $\mathfrak{g}(x^*, y) = 0$, reflecting the nature of the novel boundaries $\{x_*, x^*\}$. Solving in $\overset{\circ}{I}_*$ this Cauchy problem, $\mathfrak{g}$ takes the explicit form

$$\begin{aligned} \mathfrak{g}(x,y) &= \frac{2(\varphi(x^*) - \varphi(y))}{(g^2\varphi')(y)} \text{ if } x < y \\ \mathfrak{g}(x,y) &= \frac{2(\varphi(x^*) - \varphi(x))}{(g^2\varphi')(y)} \text{ if } x > y \end{aligned}$$

in terms of the $G-$harmonic function φ. An explicit expression of the Green function also exists for all combinations of $\{x_*, x^*\}$ either absorbing or reflecting.

2.4. **Transformation of sample paths.** Consider a one-dimensional diffusion $(x_t; t \geq 0)$ as in (1). Let $p := p(x; t, y)$ be its transition probability and let $\tau(x)$ be its explosion time.

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

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

Recall functions c and d are both chosen non-negative so that so is α.

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

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

In this construction of $(\breve{x}_t; t \geq 0)$ through a change of measure, sample paths of $(x_t; t \geq 0)$ for which $\alpha(y)$ is large are favored. This is a selection of paths procedure due to Doob (see [3]).

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

therefore is

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

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

$$\breve{G}(\cdot) = \frac{1}{\alpha} G(\alpha)\cdot + \widetilde{G}(\cdot) = -\frac{c}{\alpha}\cdot + \widetilde{G}(\cdot)$$

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

$$d\widetilde{x}_t = \widetilde{f}(\widetilde{x}_t)\, dt + g(\widetilde{x}_t)\, dw_t,\ \widetilde{x}_0 = x \in (a,b), \tag{2}$$

eventually killed at rate $\frac{c}{\alpha}$ as soon as $c \neq 0$. Whenever $(\widetilde{x}_t; t \geq 0)$ is killed, it enters conventionally the coffin state $\{\partial\}$. Let $\widetilde{\tau}(x)$ be the new explosion time at the boundaries of $(\widetilde{x}_t; t \geq 0)$ started at x, with $\widetilde{\tau}(x) = \infty$ if the boundaries are now inaccessible to the new process. Let $\widetilde{\tau}_\partial(x)$ be the killing time of $(\widetilde{x}_t; t \geq 0)$ started at x (hitting time of ∂), with $\widetilde{\tau}_\partial(x) = \infty$ if $c = 0$. Then $\breve{\tau}(x) := \widetilde{\tau}(x) \wedge \widetilde{\tau}_\partial(x)$ is the novel stopping time of killed $(\widetilde{x}_t; t \geq 0)$. The SDE for $(\widetilde{x}_t; t \geq 0)$, together with its stopping time $\breve{\tau}(x)$ characterize the new process $(\breve{x}_t; t \geq 0)$ to consider.

Normalizing: Integrating over y, with $\breve{\pi}_t(x) = \int \breve{p}(x;t,y)\, dy := \widetilde{\mathbf{P}}^x(\breve{\tau}(x) > t)$, we have $\partial_t \breve{\pi}_t(x) = \breve{G}(\breve{\pi}_t(x))$, with $\breve{\pi}_0(x) = \mathbf{1}(x \in (0,1))$. This gives the tail distribution of stopping time $\breve{\tau}(x)$.

Defining the conditional probability density $\breve{p}^c(x;t,y) = \breve{p}(x;t,y)/\breve{\pi}_t(x)$, now with total mass 1, we get

$$\partial_t \breve{p}^c = -\partial_t \breve{\pi}_t(x)/\breve{\pi}_t(x) \cdot \breve{p}^c + \breve{G}^*(\breve{p}^c),\ \breve{p}^c(x;0,y) = \delta_y(x).$$

The term $-\partial_t \breve{\pi}_t(x)/\breve{\pi}_t(x) > 0$ is the rate at which mass should be created to compensate the loss of mass of the process $(\widetilde{x}_t; t \geq 0)$ due to eventual absorption at the boundaries and/or killing.

Additive functionals of transformed process: For the new process $(\widetilde{x}_t; t \geq 0)$, it is also of interest to evaluate additive functionals along their own sample paths. Let then $\widetilde{\alpha}(x) := \widetilde{\mathbf{E}}^x \left(\int_0^{\breve{\tau}(x)} \widetilde{c}(\widetilde{x}_s)\, ds + \widetilde{d}\left(\widetilde{x}_{\breve{\tau}(x)}\right) \right)$ be such an additive functional where functions $\widetilde{c}$ and $\widetilde{d}$ are themselves both non-negative. It solves

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

Then, recalling the expression of $\mathfrak{g}(x,y)$, the Green function of $(x_t; t \geq 0)$:

$$\begin{aligned}\mathfrak{g}(x,y) &= 2\frac{(\varphi(x)-\varphi(a))(\varphi(b)-\varphi(y))}{(g^2\varphi')(y)(\varphi(b)-\varphi(a))} \text{ if } x<y \\ \mathfrak{g}(x,y) &= 2\frac{(\varphi(b)-\varphi(x))(\varphi(y)-\varphi(a))}{(g^2\varphi')(y)(\varphi(b)-\varphi(a))} \text{ if } x>y,\end{aligned}$$

we find explicitly

$$\widetilde{\alpha}(x) = \frac{1}{\alpha(x)}\int_{\overset{\circ}{I}} \mathfrak{g}(x,y)\,\alpha(y)\,\widetilde{c}(y)\,dy.$$

Remark: It results of the following simple formula:

$$G(\alpha\widetilde{\alpha}) = \alpha G(\widetilde{\alpha}) + \widetilde{\alpha}G(\alpha) + g^2\alpha'\widetilde{\alpha}',$$

where $\breve{G}(\widetilde{\alpha}) = -\widetilde{c}$ and $G(\alpha) = -c$, that: $\alpha G(\widetilde{\alpha}) = c\widetilde{\alpha} - \alpha\widetilde{c} - g^2\alpha'\widetilde{\alpha}'$. Consequently, with $\widetilde{G}(\cdot) := \frac{\alpha'}{\alpha}g^2\partial_x(\cdot) + G(\cdot)$:

$$-\widetilde{G}(\widetilde{\alpha}) = -c\alpha^{-1}\widetilde{\alpha} + \widetilde{c}.$$

As a result:

$$\widetilde{\alpha}(x) = \widetilde{\mathbf{E}}^x\left(\int_0^{\widetilde{\tau}(x)} dt\cdot\widetilde{c}(\widetilde{x}_t)\,e^{-\int_0^t\left(c\alpha^{-1}\right)(\widetilde{x}_s)ds} + \widetilde{d}\left(\widetilde{x}_{\breve{\tau}(x)}\right)\right),$$

where $(\widetilde{x}_t; t \geq 0)$ is the process (2) with infinitesimal generator $\widetilde{G}$ (including the additional drift) with explosion time $\widetilde{\tau}(x)$. $\diamondsuit$

Specific transformations of interest:

(i) The case $c = 0$ deserves a special treatment. Indeed, in this case, $\widetilde{\tau}_\partial(x) = \infty$ and so $\breve{\tau}(x) := \widetilde{\tau}(x)$, the explosion time for the process $(\widetilde{x}_t; t \geq 0)$ governed by the new SDE. Here $\breve{G} = \widetilde{G}$. Assuming α solves $-G(\alpha) = 0$ if $x \in \overset{\circ}{I}$ with boundary conditions $\alpha(a) = 0$ and $\alpha(b) = 1$ (respectively $\alpha(a) = 1$ and $\alpha(b) = 0$), the new process $(\widetilde{x}_t; t \geq 0)$ is just $(x_t; t \geq 0)$ conditioned on exiting at $x = b$ (respectively at $x = a$). In the first case, boundary b is exit whereas a is entrance; α reads

$$\alpha(x) = \frac{\int_a^x e^{-2\int_a^y \frac{f(z)}{g^2(z)}dz}dy}{\int_a^b e^{-2\int_a^y \frac{f(z)}{g^2(z)}dz}dy}$$

with

$$\widetilde{f}(x) = f(x) + \frac{g^2(x)\,e^{-2\int_a^x \frac{f(z)}{g^2(z)}dz}}{\int_a^x e^{-2\int_a^y \frac{f(z)}{g^2(z)}dz}dy}$$

giving the new drift. In the second case, $\alpha(x) = \frac{\int_x^b e^{-2\int_a^y \frac{f(z)}{g^2(z)}dz}dy}{\int_a^b e^{-2\int_a^y \frac{f(z)}{g^2(z)}dz}dy}$ and boundary a is exit whereas b is entrance. Thus $\widetilde{\tau}(x)$ is just the exit time at $x = b$ (respectively at $x = a$). Let $\widetilde{\alpha}(x) := \widetilde{\mathbf{E}}^x(\widetilde{\tau}(x))$. Then, $\widetilde{\alpha}(x)$ solves $-\widetilde{G}(\widetilde{\alpha}) = 1$, whose explicit solution is:

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

in terms of $\mathfrak{g}(x,y)$, the Green function of $(x_t; t \geq 0)$.

Example: Consider the WF model on $[0,1]$ with selection for which, with $\sigma \in \mathbf{R}$, $f(x) = \sigma x(1-x)$ and $g^2(x) = x(1-x)$. Assume α solves $-G(\alpha) = 0$ if $x \in (0,1)$ with $\alpha(0) = 0$ and $\alpha(1) = 1$; one gets, $\alpha(x) = \left(1 - e^{-2\sigma x}\right) / \left(1 - e^{-2\sigma}\right)$. The diffusion corresponding to (2) has the new drift: $\widetilde{f}(x) = \sigma x(1-x)\coth(2\sigma x)$, independent of the sign of σ. It models WF diffusion with selection conditioned on exit at $\circ = 1$. If σ is small with (say) $\sigma = c/n$, where $c \in \mathbf{R}$, then: $\widetilde{f}(x) \sim \frac{1}{2}(1-x)\left(1 + c^2x^2/n^2\right)$ only depends on c^2 and not on c.

(ii) Assume α now solves $-G(\alpha) = 1$ if $x \in \overset{\circ}{I}$ with boundary conditions $\alpha(a) = \alpha(b) = 0$. In this case study, one selects sample paths of $(x_t; t \geq 0)$ with a large mean explosion time $\alpha(x) = \mathbf{E}^x \tau(x)$. Sample paths with large sojourn time in $\overset{\circ}{I}$ are favored. We have

$$\alpha(x) = \int_{\overset{\circ}{I}} \mathfrak{g}(x,y)\, dy$$

where $\mathfrak{g}(x,y)$ is the above Green function. The boundaries of $(\widetilde{x}_t; t \geq 0)$ are now both entrance boundaries and so $\widetilde{\tau}(x) = \infty$. $(\widetilde{x}_t; t \geq 0)$ is not absorbed at the boundaries. The stopping time $\breve{\tau}(x)$ of $(\widetilde{x}_t; t \geq 0)$ is just its killing time $\widetilde{\tau}_\partial(x)$. Let $\widetilde{\alpha}(x) := \widetilde{\mathbf{E}}^x\left(\widetilde{\tau}_\partial(x)\right)$. Then, $\widetilde{\alpha}(x)$ solves $-\breve{G}(\widetilde{\alpha}) = 1$, $\widetilde{\alpha}(a) = \widetilde{\alpha}(b) = 0$, with explicit solution:

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

(iii) Assume α now solves $-G(\alpha) = \delta_y(x)$ if $x \in \overset{\circ}{I}$ with boundary conditions $\alpha(a) = \alpha(b) = 0$. In this case study, one selects sample paths of $(x_t; t \geq 0)$ with a large sojourn time density at y since $\alpha(x) =: \mathfrak{g}(x,y) = \mathbf{E}^x\left(\int_0^{\tau(x)} \delta_y(x_s)\, ds\right)$. The stopping time $\widetilde{\tau}_y(x)$ of $(\widetilde{x}_t; t \geq 0)$ is just its killing time when the process is at y for the last time. Let $\widetilde{\alpha}_y(x) := \widetilde{\mathbf{E}}^x\left(\widetilde{\tau}_y(x)\right)$. Then, $\widetilde{\alpha}_y(x)$ solves $-\breve{G}(\widetilde{\alpha}) = 1$, with explicit solution:

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

The Green function at $x_0 \in (0,1)$ of the transformed process $(\widetilde{x}_t; t \geq 0)$ is $\widetilde{\mathfrak{g}}_y(x, x_0)$ solution to: $-\breve{G}(\widetilde{\mathfrak{g}}_y) = \delta_{x_0}(x)$. It takes the simple form:

$$\widetilde{\mathfrak{g}}_y(x, x_0) = \frac{1}{\mathfrak{g}(x,y)} \int_{\overset{\circ}{I}} \mathfrak{g}(x,z)\, \mathfrak{g}(z,y)\, \delta_{x_0}(z)\, dz = \frac{\mathfrak{g}(x_0,y)}{\mathfrak{g}(x,y)} \mathfrak{g}(x,x_0).$$

3. The Wright Fisher example

In this Section, we shall largely particularize the general diffusion model to the pure neutral genetic model, originally due to Wright and Fisher (WF). Some of its relatives including various drifts are also examined. To some extent, it is possible to find a transformation from the general diffusion model (1) to the specific WF type

model with quadratic volatility $g^2(x) = x(1-x)$. (see [8], Appendix 1). When particularized to WF models, the general techniques introduced in Section 2 will lift the veil of obscurity on some problems of interest in population genetics at large. This is the main purpose of this Section.

We refer to [13] and to its extensive and exhaustive list of references for historical issues, ownership of evoked results (after Wright, Fisher, Crow, Kimura, Nagylaki, Maruyama, Ohta, Watterson, Ewens, Kingman, Griffiths, Tavaré...) and for the role played by french geneticist Gustave Malécot in the development of modern mathematical population genetics. See also the general monographs [2], [12], [5] and [7]. For recent related works with statistical physics' motivations, see [1] and [15]. In the first reference, focus chiefly is on neutral populations with fixed population sizes with a mapping with ecological and linguistic models, whereas the second work includes mutations and discusses how the transient time to Most Recent Common Ancestor (MRCA) is related to genetic diversity. Both works therefore consider the problem of running the descendant process backward in time to trace back the ancestral lineages; we shall also address this point of view in Section 4.

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

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

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

Let $N_r(m)$ be the offspring number of the m first individuals at discrete generation $r \in \mathbf{N}_0$. This sibship process is a discrete-time Markov chain with binomial transition probability given by:

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

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

$$dx_t = \sqrt{x_t(1-x_t)}dw_t,\ x_0 = x. \tag{3}$$

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

Equation (3) is a 1−dimensional diffusion as in (1) on $I = [a = 0, b = 1]$, with zero drift $f(x) = 0$ and volatility $g(x) = \sqrt{x(1-x)}$. This diffusion is already in natural coordinate and so $\varphi(x) = x$. The scale function is x and the speed measure $[x(1-x)]^{-1} dx$. One can check that both boundaries are exit in this case: Stopping time is $\tau(x) = \tau_0(x) \wedge \tau_1(x)$ where $\tau_0(x)$ is the extinction time and $\tau_1(x)$ the fixation time. The corresponding infinitesimal generators are $G(\cdot) = \frac{1}{2} x(1-x) \partial_x^2(\cdot)$ and $G^*(\cdot) = \frac{1}{2} \partial_y^2 (y(1-y)\cdot)$.

Remark: Non-neutral Wright-Fisher models (with non-null drifts) can be obtained by considering the binomial transition probabilities

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

where

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

now is some state-dependent probability which is different from identity x. For instance, taking $p_n(x) = (1-\pi_{1,n})x + \pi_{2,n}(1-x)$ where $(\pi_{1,n}, \pi_{2,n})$ are small (n-dependent) mutation probabilities, assuming $(n \cdot \pi_{1,n}, n \cdot \pi_{2,n}) \to_{n\uparrow\infty} (u_1, u_2)$, leads after scaling to the drift of WF model with mutations rates (u_1, u_2). Taking $p_n(x) = (1+s_n)x/(1+s_n x)$ where $s_n > 0$ is a small $n-$dependent selection parameter satisfying $n \cdot s_n \to_{n\uparrow\infty} \sigma > 0$, leads, after scaling, to the WF model with selective drift $\sigma x(1-x)$. Essentially, the drift $f(x)$ is a large n approximation of the bias: $n(p_n(x) - x)$.

3.2. **Explicit solutions to KBE and KFE.** As shown by Kimura in 1955, it turns out that both Kolmogorov equations are exactly solvable in this case, using spectral theory. Indeed, solutions involve a series expansion in terms of eigenfunctions of KB and KF infinitesimal generators with discrete eigenvalues spectrum. In principle, such a spectral expansion of the solutions is possible as soon as the diffusion process under study has no natural boundaries. We now consider the specific WF model.

With $z \in (-1,1)$, let $(P_k(z); k \geq 1)$ be the degree-k Gegenbauer polynomials solving $(1-z^2) P_k''(z) + k(k-1) P_k(z) = 0$ with $P_k'(\pm 1) = \mp 1/2$, $k \geq 2$; we let $P_1(z) := (1-z)/2$. When $k \geq 2$, we have $P_k(\pm 1) = 0$ and so $P_k(z) = (1-z^2) Q_k(z)$ where $Q_k(z)$ is a polynomial with degree $k-2$. With $x \in (0,1)$, let $(u_k(x); k \geq 1)$ be defined by: $u_k(x) = P_k(1-2x)$. These polynomials clearly constitute a system of eigenfunctions for the KB operator $G = \frac{1}{2} x(1-x) \partial_x^2$ with eigenvalues $\lambda_k = -k(k-1)/2$, $k \geq 1$, thus with $G(u_k(x)) = \lambda_k u_k(x)$. In particular, $u_1(x) = x$, $u_2(x) = x - x^2$, $u_3(x) = x - 3x^2 + 2x^3$, $u_4(x) = x - 6x^2 + 10x^3 - 5x^4$,...With $k \geq 2$, we have $u_k(0) = u_k(1) = 0$ and $u_k'(0) = 1$ and $u_k'(1) = -1$.

The eigenfunctions of 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: $m(y)\, dy = \frac{dy}{y(1-y)}$, for the same eigenvalues. For instance, $v_1(y) = \frac{1}{1-y}$, $v_2(y) = 1$, , $v_3(y) = 1 - 2y$, $v_4(y) = 1 - 5y + 5y^2$,...

Although $\lambda_1 = 0$ really constitutes an eigenvalue, only $v_1(y)$ is not a polynomial. When $k \geq 2$, from their definition, the $u_k(x)$ polynomials satisfy $u_k(0) = u_k(1) = 0$ in such a way that $v_k(y) = m(y) \cdot u_k(y)$, $k \geq 2$ is a polynomial with degree $k-2$.

We note that, $\langle v_j, u_k \rangle = \langle u_j, u_k \rangle_m = 0$ if $j \neq k$ and 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 a decomposition in the basis $u_k(x)$, $k \geq 2$

$$\mathbf{E}^x \psi\left(x_{t\wedge\tau(x)}\right) = \sum_{k\geq 2} c_k e^{\lambda_k t} u_k(x) \text{ where } c_k = \frac{\langle \psi, u_k \rangle_m}{\langle v_k, u_k \rangle} = \frac{\int_0^1 \psi(y)\, u_k(y)\, m(y)\, dy}{\int_0^1 v_k(y)\, u_k(y)\, dy},$$

and $\psi(x) = \sum_{k\geq 2} c_k u_k(x)$. This series expansion solves KBE: $\partial_t u = G(u)$; $u(x,0) = \psi(x)$ where $u = u(x,t) := \mathbf{E}^x \psi\left(x_{t\wedge\tau(x)}\right)$.

Moreover, with $\mathbf{P}^x\left(x_{t\wedge\tau(x)} \in dy\right) := p(x;t,y)\,dy$, we have the decomposition of this measure on the series of measures $v_k(y)\,dy$:

$$\mathbf{P}^x\left(x_{t\wedge\tau(x)} \in dy\right) = \sum_{k\geq 2} b_k e^{\lambda_k t} u_k(x)\, v_k(y)\, dy \text{ where } b_k = \frac{1}{\int_0^1 v_k(y)\, u_k(y)\, dy}.$$

This series expansion solves KFE of the WF model. The transition density $p(x;t,y) = \mathbf{P}^x\left(x_{t\wedge\tau(x)} \in dy\right)/dy$ is reversible with respect to the speed density since for $0 < x, y < 1$

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

The measures $v_k(y)\,dy$, $k \geq 2$ are not probability measures because $v_k(y)$ is not necessarily positive over $[0,1]$. This decomposition 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 at the boundaries).

We have $\mathbf{P}^x(\tau(x) > t) = \int_0^1 \mathbf{P}^x\left(x_{t\wedge\tau(x)} \in dy\right)$ and so

$$\overline{\pi}_t(x) := \mathbf{P}^x(\tau(x) > t) = \sum_{k\geq 2} \frac{\int_0^1 v_k(y)\, dy}{\int_0^1 v_k(y)\, u_k(y)\, dy} e^{\lambda_k t} u_k(x)$$

is the exact tail distribution of the explosion time.

Since $v_2(y) = 1$, to the leading order in t, for large time

$$\mathbf{P}^x\left(x_{t\wedge\tau(x)} \in dy\right) \sim 6e^{-t} \cdot x(1-x)\, dy + \mathcal{O}\left(e^{-3t}\right)$$

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

$$\mathbf{P}^x(x_t \in dy \mid \tau(x) > t) \underset{t\uparrow\infty}{\sim} dy$$

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

We finally observe that $\mathbf{E}^x u_l\left(x_{t\wedge\tau(x)}\right) = e^{\lambda_l t}u_l(x)$, $l \geq 1$ so that the law has all its moments given by:

$$m_{k,t}(x) := \mathbf{E}^x\left(x^k_{t\wedge\tau(x)}\right) = x + \sum_{l=2}^{k} c_{k,l}e^{\lambda_l t}u_l(x).$$

Here $c_{k,l} = \frac{\int_0^1 (y^k - y)u_l(y)m(y)dy}{\int_0^1 v_k(y)u_l(y)dy}$, $l = 2,..,k$, are the rational coefficients of the projection of $x^k - x$ on the eigenfunctions $u_l(x)$: $x^k = x + \sum_{l=2}^k c_{k,l}u_l(x)$.
For instance, $m_{1,t}(x) = u_1(x) = x$, $m_{2,t}(x) = u_1(x) - e^{-t}u_2(x) = x\left(1-(1-x)e^{-t}\right)$, $m_{3,t}(x) = u_1(x) - \frac{3}{2}e^{-t}u_2(x) + \frac{1}{2}e^{-3t}u_3(x)$...Note that, for all $k \geq 1$, $m_{k,0}(x) = x^k$ and $m_{k,t}(x) \to_{t\uparrow\infty} \mathbf{E}^x\left(x^k_{\tau(x)}\right) = \mathbf{P}(\tau_1(x) < \tau_0(x)) = x$, the fixation probability. From this, we get the dynamics of heterozygosity $\mathbf{E}^x\left(2x_{t\wedge\tau(x)}\left(1-x_{t\wedge\tau(x)}\right)\right) = 2x(1-x)e^{-t}$ which tends to 0 exponentially fast as $t \to \infty$. Its variance can be found to be

$$\boldsymbol{\sigma}^2_x\left(2x_{t\wedge\tau(x)}\left(1-x_{t\wedge\tau(x)}\right)\right) = \frac{4}{5}x(1-x)\left[e^{-t} - (1-5x(1-x))e^{-6t} - 5x(1-x)e^{-2t}\right].$$

It vanishes linearly when $t \to 0$ and exponentially when $t \to \infty$; it is maximal at some intermediate time $t = t_*(x)$.

3.3. **Additive functionals for WF.** Let $(x_t; t \geq 0)$ be the WF diffusion model defined by (1) on the interval $I = [0,1]$ where both endpoints are absorbing (exit). We wish to evaluate the additive quantities

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

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

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

1. $c = \lim_{\varepsilon\downarrow 0}\frac{1}{2\varepsilon}\mathbf{1}(x \in (y-\varepsilon, y+\varepsilon)) =: \delta_y(x)$ and $d = 0$, when $y \in \overset{\circ}{I}$: in this case, $\alpha := \mathfrak{g}(x,y)$ is the Green function. The solution takes the simple form

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

The Green function is of particular interest to solve the above general problem of evaluating additive functionals $\alpha(x)$. Indeed, one easily finds

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

2. $c=1$ and $d=0$: here, $\alpha=\mathbf{E}^x\tau(x)$ is the mean time of explosion (average time spent in I before explosion). The solution is (Crow and Kimura formula)

$$\alpha(x)=2x\int_x^1\frac{dy}{y}+2(1-x)\int_0^x\frac{dy}{1-y}=-2(x\log x+(1-x)\log(1-x)).$$

3. Let $c=0$ and $d(\circ)=\mathbf{1}(\circ=1)$. Let $\alpha(x)=\mathbf{P}(x_t \text{ first hits } [0,1] \text{ at } 1\mid x_0=x)$. Then $\alpha(x)$ is a $G-$harmonic function solution to $G(\alpha)=0$, with boundary conditions $\alpha(0)=0$ and $\alpha(1)=1$. The solution for WF model is: $\alpha(x)=x$. Stated differently, $x=\mathbf{P}^x(\tau_1(x)<\tau_0(x))$ is the probability that the exit time at $\circ=1$ is less than the one at $\circ=0$, starting from x.

On the contrary, choosing $\alpha(x)$ to be a $G-$harmonic function with boundary conditions $\alpha(0)=1$ and $\alpha(1)=0$, $\alpha(x)=\mathbf{P}(x_t \text{ first hits } [0,1] \text{ at } 0\mid x_0=x)=1-x$. Thus, $1-x=\mathbf{P}^x(\tau_0(x)<\tau_1(x))$.

4. Let $c(x)=2x(1-x)$ measure the heterozygosity of the WF process and assume $d(0)=d(1)=1$. The average heterozygosity over sample paths is

$$\alpha(x)=4x\int_x^1(1-y)\,dy+4(1-x)\int_0^x y dy=2x(1-x).$$

5. Assume $c(x)=\mathbf{1}\left(x\in\left(n^{-1},1-n^{-1}\right)\right)$ and $d=0$:

Then, $\alpha=\mathbf{E}^x\left(\int_0^{\tau(x)}\mathbf{1}\left(x_s\in\left(n^{-1},1-n^{-1}\right)\right)ds\right)$ is the mean time spent in interval $\left(n^{-1},1-n^{-1}\right)$ before explosion, solution to:

$$\begin{aligned}-G(\alpha) &= \mathbf{1}\left(x\in\left(n^{-1},1-n^{-1}\right)\right) \text{ if } x\in\overset{\circ}{I}\\ \alpha &= 0 \text{ if } x\in\partial I.\end{aligned}$$

If $x=n^{-1}$, one finds $\alpha\left(n^{-1}\right)=2n^{-1}\int_{n^{-1}}^{1-n^{-1}}dy/y\sim 2\log n/n$ which is small when n is large. The diffusion represents the offspring frequency of a singleton; it starts at n^{-1} which is close to boundary 0 where x_t gets extinct. It therefore has little opportunity to visit $\left(n^{-1},1-n^{-1}\right)$.

3.4. **Transformation of WF sample paths (**[12]**).** With $p(x;t,y)$ the transition probability density of WF model, define a new $\alpha-$transformed stochastic process $(\breve{x}_t;t\geq 0)$ by its transition probability

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

(i) *Conditioned WF on exit at some boundary:* Assume first α solves $-G(\alpha)=0$ with boundary conditions $\alpha(0)=0$ and $\alpha(1)=1$; hence, α reads $\alpha(x)=x$. In this case, $\widetilde{\tau}_\partial(x)=\infty$ (no killing) and so $\breve{\tau}(x):=\widetilde{\tau}(x)$ is the explosion time for a process $(\widetilde{x}_t;t\geq 0)$ governed by a new SDE with a drift term. The new process $(\widetilde{x}_t;t\geq 0)$ is just $(x_t;t\geq 0)$ conditioned on exiting at $\circ=1$. Boundary 1 is exit whereas 0 is entrance. Thus the model for $(\widetilde{x}_t;t\geq 0)$ becomes $d\widetilde{x}_t=(1-\widetilde{x}_t)\,dt+\sqrt{\widetilde{x}_t(1-\widetilde{x}_t)}dw_t$, $\widetilde{x}_0=x\in(0,1)$ now with linear drift $\widetilde{f}(x)=1-x$ and $g(x)=\sqrt{x(1-x)}$. Its transition probability is

$$\breve{p}_1(x;t,y)=\frac{y}{x}p(x;t,y),$$

where subscript 1 indicates that this is the conditional transition probability of sample paths whose exit is necessarily at boundary 1.

Assuming now α solves $-G(\alpha)=0$ if $x\in\overset{\circ}{I}$ with boundary conditions $\alpha(0)=1$ and $\alpha(1)=0$, the new process $(\widetilde{x}_t;t\geq 0)$ is just $(x_t;t\geq 0)$ conditioned on exiting at $x=0$. Boundary 0 is exit whereas 1 is entrance; in this case, α is $\alpha(x)=1-x$. Thus the model for $(\widetilde{x}_t;t\geq 0)$ becomes $d\widetilde{x}_t=-\widetilde{x}_t dt+\sqrt{\widetilde{x}_t(1-\widetilde{x}_t)}dw_t$, $\widetilde{x}_0=x\in(0,1)$ with $\widetilde{f}(x)=-x$ and $g(x)=\sqrt{x(1-x)}$. Its transition probability is

$$\breve{p}_0(x;t,y)=\frac{1-y}{1-x}p(x;t,y),$$

where subscript 0 indicates that this is the conditional transition probability of WF sample paths whose exit now is at $\circ=0$. Recalling that, starting from x, $(x_t;t\geq 0)$ gets absorbed at $\circ=1$ (respectively 0) with probability x (respectively $1-x$), we recover that

$$p(x;t,y)=x\cdot\breve{p}_1(x;t,y)+(1-x)\cdot\breve{p}_0(x;t,y).$$

Using the solution to KFE for p, we obtain an expression for both $\breve{p}_1(x;t,y)$ and $\breve{p}_0(x;t,y)$, simply by pre-multiplying it by the corresponding right factor. Integrating the results over y, we get the conditional tail distributions of the exit times at $\circ=1$ or 0, given the exit is at $\circ=1$ or 0.

Exploiting the large time behavior of $p(x;t,y)$, to the first order in t, we get

$$\begin{aligned}\breve{p}_1(x;t,y) &\sim 6e^{-t}\cdot(1-x)\,y\\ \breve{p}_0(x;t,y) &= 6e^{-t}\cdot x(1-y).\end{aligned}$$

Integrating over y, $\overline{\pi}_{t,1}(x):=\widetilde{\mathbf{P}}_1^x(\widetilde{\tau}(x)>t)\sim 3e^{-t}\cdot(1-x)$ and $\overline{\pi}_{t,0}(x):=\widetilde{\mathbf{P}}_0^x(\widetilde{\tau}(x)>t)\sim 3e^{-t}\cdot x$ are the large time behaviors of the absorption times at 1 and 0 respectively. Using this, we get the large time behaviors of the conditional probabilities

$$\begin{aligned}\widetilde{\mathbf{P}}_1^x(\widetilde{x}_t\in dy\mid\widetilde{\tau}(x)>t) &\sim 2ydy\\ \widetilde{\mathbf{P}}_0^x(\widetilde{x}_t\in dy\mid\widetilde{\tau}(x)>t) &\sim 2(1-y)\,dy,\end{aligned}$$

where we recognize the densities of specific beta-distributed random variables. Specifically, we conclude that, as time passes by, given explosion occurs at $\circ=1$ and given it has not occurred in the past, $\widetilde{x}_t\overset{d}{\to}$ beta$(2,1)$ distribution on $[0,1]$. Similarly, given explosion occurs at $\circ=0$ and given it has not occurred previously, $\widetilde{x}_t\overset{d}{\to}$ beta$(1,2)$ distribution on $[0,1]$.

In the previously displayed formula, $\widetilde{\tau}(x)$ is just the exit time at $\circ=1$ (respectively at $\circ=0$) of the conditional transformed WF diffusions. Let $\widetilde{\alpha}(x):=\widetilde{\mathbf{E}}^x(\widetilde{\tau}(x))$. Then, $\widetilde{\alpha}(x)$ solves $-\breve{G}(\widetilde{\alpha})=1$, whose explicit solution is:

$$\widetilde{\alpha}(x)=\frac{1}{\alpha(x)}\int_0^1\mathfrak{g}(x,y)\,\alpha(y)\,dy$$

in terms of $\mathfrak{g}(x,y)$, the Green function of $(x_t; t \geq 0)$. For the WF model conditioned on exit at $\circ = 1$ (respectively 0), we find respectively Kimura and Ohta's formulae

$$\begin{aligned}\widetilde{\alpha}_1(x) &= -\frac{2}{x}(1-x)\log(1-x)\\ \widetilde{\alpha}_0(x) &= -\frac{2}{1-x}x\log x.\end{aligned}$$

This result could have been guessed by observing that $x\widetilde{\alpha}_1(x) + (1-x)\widetilde{\alpha}_0(x)$ is the expected explosion time of the original WF model. When $x \downarrow 0$, (respectively $x \uparrow 1$), it takes an average time 2 to reach 1 (respectively 0) for WF conditioned on exit at $\circ = 1$ (respectively 0).

(*ii*) *WF sample paths favoring large exit time:* Assume α now solves $-G(\alpha) = 1$ if $x \in \overset{\circ}{I}$ with boundary conditions $\alpha(0) = \alpha(1) = 0$ and consider the associated $\alpha-$transformed process. In this case study, one selects sample paths of $(x_t; t \geq 0)$ with a large mean explosion time $\alpha(x) = \mathbf{E}^x\tau(x)$. Sample paths with large sojourn time within $\overset{\circ}{I}$ are favored. For the neutral WF model, we have

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

With $\widetilde{f}(x) = \frac{x(1-x)\log(x/(1-x))}{x\log x+(1-x)\log(1-x)}$, the dynamics of $(\widetilde{x}_t; t \geq 0)$ is

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

with killing rate $1/\alpha(x)$. The new drift is symmetric around 1/2 ($\widetilde{f}(1-x) = -\widetilde{f}(x)$); it tends to concentrate the probability mass of $\widetilde{x}_t$ at this point. The boundaries of $(\widetilde{x}_t; t \geq 0)$ are now both entrance boundaries and so $\widetilde{\tau}(x) = \infty$. $(\widetilde{x}_t; t \geq 0)$ is not absorbed at the boundaries. The stopping time $\breve{\tau}(x)$ of $(\widetilde{x}_t; t \geq 0)$ is just its killing time $\widetilde{\tau}_\partial(x)$. Let $\widetilde{\alpha}(x) := \widetilde{\mathbf{E}}^x(\widetilde{\tau}_\partial(x))$ be its expected value. Then, $\widetilde{\alpha}(x)$ solves $-\breve{G}(\widetilde{\alpha}) = 1$ whose explicit solution is:

$$\widetilde{\alpha}(x) = 2\frac{x\int_x^1 \frac{y\log y+(1-y)\log(1-y)}{y}dy + (1-x)\int_0^x \frac{y\log y+(1-y)\log(1-y)}{1-y}dy}{x\log x + (1-x)\log(1-x)}.$$

We get the large time behavior of the conditional probabilities

$$\widetilde{\mathbf{P}}^x(\widetilde{x}_t \in dy \mid \widetilde{\tau}_\partial(x) > t) \sim -2(y\log y + (1-y)\log(1-y))\,.$$

As time passes by, killing occurs, and given killing has not occurred in the past, $\widetilde{x}_t \overset{d}{\to} x_\infty$ a random variable with logarithmic density $-2(y\log y + (1-y)\log(1-y))$ on $[0,1]$.

(*iii*) *Selection of WF sample paths with large heterozygosity.* Assume α now solves $-G(\alpha) = 2x(1-x)$ if $x \in \overset{\circ}{I}$ with boundary conditions $\alpha(0) = \alpha(1) = 0$. Then, $\alpha = 2x(1-x)$. In this case study, one selects sample paths of $(x_t; t \geq 0)$ with large heterozygosity. The dynamics of $(\widetilde{x}_t; t \geq 0)$ is

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

subject to a constant killing rate 1. The boundaries of $(\widetilde{x}_t; t \geq 0)$ are now both entrance boundaries and so $\widetilde{\tau}(x) = \infty$. $(\widetilde{x}_t; t \geq 0)$ is not absorbed at the boundaries.

The stopping time $\breve{\tau}(x)$ of $(\widetilde{x}_t; t \geq 0)$ is just its killing time $\widetilde{\tau}_\partial(x)$ which is mean 1 exponentially distributed: $\widetilde{\mathbf{P}}^x(\widetilde{\tau}_\partial(x) > t) =: \breve{\pi}_t(x) = e^{-t}$, independently of the starting point x. Let for instance $\widetilde{\alpha}(x) := \widetilde{\mathbf{E}}^x(\widetilde{\tau}_\partial(x))$ be its expected value. Then, $\widetilde{\alpha}(x)$ solves $-\breve{G}(\widetilde{\alpha}) = 1$, whose explicit solution is, as required:

$$\widetilde{\alpha}(x) = 2\frac{x\int_x^1 \frac{y(1-y)}{y}dy + (1-x)\int_0^x \frac{y(1-y)}{1-y}dy}{x(1-x)} = 1.$$

As time passes, killing of $\widetilde{x}_t$ occurs, and given killing has not yet occurred, $\widetilde{x}_t \xrightarrow{d} x_\infty$ a random variable with density $6y(1-y)$ on $[0,1]$ which is a beta$(2,2)$ density. In this selection of paths procedure, the conditional density of $(\widetilde{x}_t; t \geq 0)$ given $\widetilde{\tau}_\partial(x) > t$ is $\breve{p}^c(x;t,y) := \breve{p}(x;t,y)/\breve{\pi}_t(x)$ where $\breve{p}(x;t,y) = \frac{y(1-y)}{x(1-x)}p(x;t,y)$ and $\breve{\pi}_t(x) = e^{-t}$. Using the reversibility property, $\breve{p}^c(x;t,y) = e^t p(y;t,x)$ takes the simple explicit form

$$\breve{p}^c(x;t,y) = \sum_{k\geq 2} b_k e^{(\lambda_k+1)t} v_k(x)\, u_k(y).$$

(*iv*) *Selection of WF sample paths with large sojourn time density at* y. Assume now α solves $-G(\alpha) = \delta_y(x)$ if $x \in \overset{\circ}{I}$ and so $\alpha(x) =: \mathfrak{g}(x,y)$. The stopping time $\widetilde{\tau}_y(x)$ of $(\widetilde{x}_t; t \geq 0)$ is just its killing time when the process is at y for the last time. Let $\widetilde{\alpha}_y(x) := \widetilde{\mathbf{E}}^x(\widetilde{\tau}_y(x))$ be the expected such time. Then, $\widetilde{\alpha}_y(x)$ solves $-\breve{G}(\widetilde{\alpha}_y) = 1$, whose explicit solution is:

$$\begin{aligned}
\widetilde{\alpha}_y(x) &= \frac{1}{\mathfrak{g}(x,y)}\int_0^1 \mathfrak{g}(x,z)\,\mathfrak{g}(z,y)\,dz \\
&= -2\left(1 + \frac{y}{1-y}\log y + \frac{1-x}{x}\log(1-x)\right) \text{ if } x < y \\
&= -2\left(1 + \frac{1-y}{y}\log(1-y) + \frac{x}{1-x}\log x\right) \text{ if } x > y.
\end{aligned}$$

The Green function at $x_0 \in (0,1)$ of the transformed WF process $(\widetilde{x}_t; t \geq 0)$ is $\widetilde{\mathfrak{g}}_y(x, x_0)$ solution to: $-\breve{G}(\widetilde{\mathfrak{g}}_y) = \delta_{x_0}(x)$. It takes the simple form:

$$\widetilde{\mathfrak{g}}_y(x,x_0) = \frac{1}{\mathfrak{g}(x,y)}\int_0^1 \mathfrak{g}(x,z)\,\mathfrak{g}(z,y)\,\delta_{x_0}(z)\,dz = \frac{\mathfrak{g}(x_0,y)}{\mathfrak{g}(x,y)}\mathfrak{g}(x,x_0).$$

Explicitly,

$$\begin{aligned}
\widetilde{\mathfrak{g}}_y(x,x_0) &= 2\frac{x\mathfrak{g}(x_0,y)}{x_0\mathfrak{g}(x,y)} \text{ if } x < x_0 \\
&= 2\frac{(1-x)\,\mathfrak{g}(x_0,y)}{(1-x_0)\,\mathfrak{g}(x,y)} \text{ if } x > x_0.
\end{aligned}$$

Depending on the position of y with respect to x and x_0, there are six different expressions of $\widetilde{\mathfrak{g}}_y(x,x_0)$. For instance, if $x=n^{-1}$:

$$\begin{aligned}\widetilde{\mathfrak{g}}_y\left(n^{-1},x_0\right) &= 2 \text{ if } n^{-1}<x_0<y \\ &= 2\frac{y(1-x_0)}{x_0(1-y)} \text{ if } n^{-1}<y<x_0\end{aligned}$$

is the Green function of singleton sample-paths at x_0, started at $x=n^{-1}$, for transformed WF model favoring large sojourn time density at y, $x_0\wedge y>n^{-1}$. It is insensitive to y, on the whole x_0-range $n^{-1}<x_0<y$.

(v) *Transformation of WF sample paths with selection:* Consider the Wright-Fisher diffusion with selection: $dx_t=\sigma x_t(1-x_t)\,dt+\sqrt{x_t(1-x_t)}dw_t$, $x_0=x\in(0,1)$. For this model, both boundaries are exit. Suppose α solves $-G(\alpha)=\frac{1}{2}\sigma^2x(1-x)e^{-\sigma x}$, with solution $\alpha(x)=e^{-\sigma x}$. In this case study, one selects sample paths of $(x_t;t\geq 0)$ with large heterozygosity weighted by the selection factor $\sigma^2e^{-\sigma x}/4$. The dynamics of $(\widetilde{x}_t;t\geq 0)$ is the drift-less WF dynamics $d\widetilde{x}_t=\sqrt{\widetilde{x}_t(1-\widetilde{x}_t)}dw_t$, subject to the quadratic killing at rate $\frac{1}{2}\sigma^2x(1-x)$ inside I. The boundaries of $(\widetilde{x}_t;t\geq 0)$ are still exit and the stopping time $\breve{\tau}(x)$ of $(\widetilde{x}_t;t\geq 0)$ is $\breve{\tau}(x)=\widetilde{\tau}(x)\wedge\widetilde{\tau}_\partial(x)$ where $\widetilde{\tau}(x)$ is its (known) explosion time at the boundaries and $\widetilde{\tau}_\partial(x)$ its killing time.

3.5. **Wright-Fisher diffusion with irreversible (one-way) mutation.** We now consider another interesting occurrence of conditioning**.** Let $u>0$ be some mutation rate. Consider the Wright-Fisher diffusion with irreversible mutation: $dx_t=-ux_tdt+\sqrt{x_t(1-x_t)}dw_t$, $x_0=x\in(0,1)$. For this model, $\circ=0$ is an exit absorbing boundary. The drift term $-ux$ attracts the particle at 0. When $u<1/2$, boundary $\circ=1$ is regular whereas it is an exit boundary when $u>1/2$. When $u<1/2$ the drift term is not strong enough to fix definitively the sample paths of x_t when it first hits $\circ=1$.

Assume next that $u<1/2$. We want to compute the expected extinction time of $(x_t;t\geq 0)$ at $\circ=0$ given fixation at $\circ=1$ did not occur in the past. To do this, we first force its boundary $\circ=1$ to be itself absorbing. Next, for this model with both boundaries absorbing, we compute the probability $\alpha(x)$ that exit is at $\circ=0$ rather than at $\circ=1$, starting from x. $\alpha(x)$ satisfies $G(\alpha)=0$, $\alpha(0)=1$, $\alpha(1)=0$, where $G=-ux\partial_x+\frac{1}{2}x(1-x)\partial_x^2$. We find: $\alpha(x)=(1-x)^v$ where $v:=1-2u>0$. We consider the new conditioned diffusion with infinitesimal generator $\breve{G}(\cdot)=\alpha^{-1}G(\alpha\cdot)$. The associated diffusion is: $d\widetilde{x}_t=-\widetilde{u}\cdot\widetilde{x}_tdt+\sqrt{\widetilde{x}_t(1-\widetilde{x}_t)}dw_t$, $\widetilde{x}_0=x\in(0,1)$. The new drift term is linear $-\widetilde{u}\cdot x$ where $\widetilde{u}:=1-u>1/2$. This diffusion is WF model with irreversible mutation conditioned on exit at $\circ=0$. For $(\widetilde{x}_t;t\geq 0)$, $\circ=0$ still is an exit absorbing boundary but $\circ=1$ now is an entrance boundary. Let $\widetilde{\alpha}(x)$ be the expected extinction time at $\circ=0$ given fixation at $\circ=1$ did not occur in the past. $\widetilde{\alpha}(x)$ solves $\breve{G}(\widetilde{\alpha})=-1$ and so

$$\widetilde{\alpha}(x)=\frac{1}{\alpha(x)}\int_0^1\mathfrak{g}(x,y)\,\alpha(y)\,dy$$

where $\mathfrak{g}(x,y)$ is the Green function of WF with irreversible mutation, admitting both absorbing boundaries. The natural coordinate being $\varphi(x) = 1-(1-x)^v$, we have

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

The final searched result is (Maruyama formula):

$$\widetilde{\alpha}(x) = \frac{2}{v}\int_0^x \frac{1-(1-y)^v}{y}dy + \frac{2}{v}\frac{1-(1-x)^v}{(1-x)^v}\int_x^1 \frac{(1-y)^v}{y}dy.$$

4. Backward in time: the coalescents

Based on the forward discrete time Galton-Watson process whose offspring is n and constant at all generations, a coalescent process can be defined (together with its scaling limit). It is a backward discrete-time Markov process whose state-space is the set of equivalence classes (partitions) of set $[n] := \{1,..,n\}$, identifying two labels from $[n]$ at each step if they share a common ancestor one generation ago. If one views the descendant process backward in time, individuals are seen to choose their parents independently and at random from the individuals in the previous generation, successive choices being independent from generation to generation. A scaling limit of the ancestral process can be obtained. Let us first illustrate this point, starting with Kingman coalescent (see [10]) associated to the pure genetic drift WF model.

Transition probabilities: Let us start with the discrete case already discussed. Let $\nu_n(m)$, $m=1,..,n$ be the offspring number of individual number m, one generation ahead with exchangeable law, assuming the number n of individuals is conserved over time. Let $b \geq a \geq 1$ both belonging to $[n]$. Moving backwards in time, the probability that b randomly chosen individuals out of n have a distinct parents, c merging classes and cluster sizes $b_1 \geq .. \geq b_c \geq 2$, $b_{c+1} = .. = b_a = 1$ is

$$P_{b;a,\mathbf{b}_a} = \frac{\{n\}_a}{\{n\}_b}\mathbf{E}\left(\prod_{m=1}^a \{\nu_n(m)\}_{b_m}\right)$$

where $\mathbf{b}_a := b_1,...,b_c,\ b_{c+1},...,b_a$, all larger than 1, satisfy $\sum_{m=1}^a b_m = b$. Here $\{n\}_a := n(n-1)..(n-a+1)$ is the falling factorial. When the exchangeable reproduction law is Cannings', we simply have

$$\mathbf{E}\left(\prod_{m=1}^a \{\nu_n(m)\}_{b_m}\right) = \frac{\{n\}_b}{n^b},$$

independently of $\mathbf{b}_a$ (for given a, all arrival states $(a,\mathbf{b}_a)$ are equally likely). Thus,

$$P_{b;a,\mathbf{b}_a} = \frac{\{n\}_a}{n^b},$$

is the one-step transition probability from b to $(a,\mathbf{b}_a)$.

Let $\nabla_{b;a,\mathbf{b}_a} := \{\text{partitions of } [b] \text{ into } a \text{ clusters with sizes } \mathbf{b}_a \geq 1\}$ with: $^{\#}\nabla_{b;a,\mathbf{b}_a} = \frac{b!}{a!}\frac{1}{\prod_{m=1}^{a} b_m!}$. Let also $\nabla_{b;a} := \{\text{partitions of } [b] \text{ into } a \text{ clusters}\}$. Then, with $S_{b,a}$ the second-kind Stirling numbers,

$$^{\#}\nabla_{b;a} = \sum_{\mathbf{b}_a} {}^{\#}\nabla_{b;a,\mathbf{b}_a} = S_{b,a},$$

where the summation runs over sequences $\mathbf{b}_a \geq 1$ satisfying $\sum_{m=1}^{a} b_m = b$. Let now

$$P_{b;a} := \frac{b!}{a!}\sum_{\mathbf{b}_a}\frac{1}{\prod_{m=1}^{a} b_m!}P_{b;a,\mathbf{b}_a}$$

be the one-step transition probability from b to a. For the Cannings model, we obtain:

$$P_{b;a} = \frac{\{n\}_a}{n^b}S_{b,a}.$$

This expression allows to compute some quantities of interest.

• First, the coalescence probability for pairs is: $c_n := P_{2;1} = n^{-1}$.

• Another transition of interest is the one from b to $b-1$ (binary collisions): we have

$$P_{b;b-1} = \frac{\{n\}_{b-1}}{n^b}S_{b,b-1}, \text{ with } S_{b,b-1} = \binom{b}{2}.$$

Note that, when n is large

$$P_{b;b-1} \sim_{n\uparrow\infty} \frac{1}{n}\binom{b}{2}$$

and $P_{b;a} = o\left(n^{-1}\right)$ if $a \neq b-1$. As a result, $P_{b;b} \sim_{n\uparrow\infty} 1 - P_{b;b-1}$.

• Finally,

$$\frac{P_{3;1}}{P_{2;1}} = \frac{1}{n} \underset{n\uparrow\infty}{\to} 0,$$

and ancestral triple mergers of ancestral lines are asymptotically negligible in comparison with binary mergers.

The ancestral Markov chain: Let P_n be a lower-triangular $n \times n$ stochastic matrix with entries $(P_n)_{b,a} = P_{b;a}$, $n \geq b \geq a \geq 1$. Let $\mathcal{A}_r \in [n]$ be the number of ancestors of $[n]$, r generations ahead, satisfying $\mathcal{A}_0 = n$. Let $\mathbf{P}_r(\mathbf{n}) := (\mathbf{P}(\mathcal{A}_r = 1), ..., \mathbf{P}(\mathcal{A}_r = n))$ be its row vector of probabilities. We have

$$\mathbf{P}_{r+1}(\mathbf{n}) = \mathbf{P}_r(\mathbf{n})P_n, \text{ with } \mathbf{P}_0(\mathbf{n}) = (0, ..., 0, 1) \text{ and } \mathbf{P}_\infty(\mathbf{n}) = (1, ..., 0, 0).$$

The pure death Markov chain $\mathcal{A}_r$ evolves backwards in time r, starting from probability state $(0, ..., 0, 1)$ till it reaches state $(1, ..., 0, 0)$ where $\mathcal{A}_r$ is reduced to a single common ancestor for ever.

Kingman limiting coalescent Markov chain (large sample): Recall that $P_n =_{n\uparrow\infty} I + c_n Q_n + o(c_n)$ where the $n \times n$ matrix Q_n is defined by its entries $(Q_n)_{b,a} = -\binom{b}{2}$ if $a = b$, $(Q_n)_{b,a} = \binom{b}{2}$ if $a = b-1$ and $(Q_n)_{b,a} = 0$ otherwise. Consequently, as

$n \uparrow \infty$, with $s \in \mathbf{R}_+$, the re-scaled discrete-time version of the ancestral process $A_s := \mathcal{A}_{\lfloor s/c_n \rfloor}$ converges weakly to some continuous-time Markov process $(A_s; s \geq 0)$ known as the standard Kingman tree coalescent (see [16] for a review).

Consider a $k \times k$ sub-matrix Q_k of Q_∞, choosing a sub-sample of k leaves out of Kingman tree with infinitely many of them. With $A_0 = k$, let $(A_s; s \geq 0)$ with values in $\mathbf{N}$ be the continuous-time pure death Markov process with sub-diagonal transition matrix Q_k. The process $(A_s; s \geq 0)$ models the coalescence ancestral process of a sub-sample of k leaves, running the descendant process backward in time. Given the ancestral process is in state b, only pairs of particles will merge at rate $\binom{b}{2}$ and one at a time. State 1 is absorbing.

With $\mathbf{P}_s(\mathbf{k}) := \left(\mathbf{P}^k(A_s = 1), ..., \mathbf{P}^k(A_s = k)\right)$, the Markov chain is described by its Chapman-Kolmogorov equation

$$\frac{d}{ds}\mathbf{P}_s(\mathbf{k}) = \mathbf{P}_s(\mathbf{k}) Q_k, \text{ with } \mathbf{P}_0(\mathbf{k}) = (0, ..., 0, 1) \text{ and } \mathbf{P}_\infty(\mathbf{k}) = (1, 0, ..., 0).$$

Initially ($s = 0$), $A_s = k$ with probability 1 whereas, at the end of the coalescence process ($s \uparrow \infty$), $A_s = 1$ with probability 1. The continuous-time Markov chain $(A_s; s \geq 0)$ is Kingman coalescent tree. It has state-space $\{1, .., k\}$ and backward generator $G_A\psi(b) = \binom{b}{2}(\psi(b-1) - \psi(b))$ for all suitable $\psi : [k] \to \mathbf{C}$.

Let $x \in [0, 1]$ and assume $\psi(b) = x^b$. Let $\phi_k(s, x) = \mathbf{E}^k\left(x^{A_s}\right) = \sum_{a=1}^k x^a \mathbf{P}^k(A_s = a)$ be the generating function of the ancestral process $(A_s; s \geq 0)$, starting from k descendants. From the above Chapman-Kolmogorov formula, its dynamics may also be written compactly as:

$$\partial_s \phi_k(s, x) = \frac{1}{2} x(1 - x) \partial_x^2 \phi_k(s, x),$$

with initial condition $\phi_k(0, x) = x^k$.

This shows that, if $(x_s; s \geq 0)$ and $(\overline{x}_s; s \geq 0)$ are the Wright-Fisher diffusions defined by equation (3), the duality relations

$$\mathbf{E}^x\left(x^k_{s\wedge\tau(x)}\right) = \mathbf{E}^k\left(x^{A_s}\right) \text{ and } \mathbf{E}^{\overline{x}}\left(\overline{x}^k_{s\wedge\tau(\overline{x})}\right) = \mathbf{E}^k\left(\overline{x}^{A_s}\right)$$

hold, relating moments of the Wright-Fisher descendant processes to the generating function of the ancestral process.

The process $(x_s; s \geq 0)$ started at x describes the evolution of the type of individuals constituting a large partitioned population: those (first type) descending from the first $\lfloor nx \rfloor$ initiators and those (second type) descending from the remaining part. If we sample k individuals from the population at time s, then $\mathbf{E}^x\left(x^k_{s\wedge\tau(x)}\right)$ is the probability to obtain only type 1 individuals in the k-sample. But this probability can be computed in a different way: if we know that the k sampled individuals are descendants of A_s different ancestors at time $s = 0$ and if the initial fraction of type 1 individuals is x, then the probability that all A_s ancestors are of type 1 can be obtained by averaging over the random genealogy to get $\mathbf{E}^k\left(x^{A_s}\right)$. The advantage is that the structure of the dual process $(A_s; s \geq 0)$ is of great simplicity compared to the one of the original WF diffusion itself. Using symmetry, the same holds true by replacing x by $\overline{x}$ and $(x_s; s \geq 0)$ by $(\overline{x}_s; s \geq 0)$.

The Wright-Fisher model first hits the state 1 (respectively 0) in finite time with probability x (respectively $1-x$). Thus, as $s \uparrow \infty$, $\mathbf{E}^x\left(x^k_{s\wedge\tau(x)}\right) \to 1^k x + 0^k(1-x) = x$, showing as required that $A_s \to 1$ with probability 1.

Recalling $\mathbf{E}^x\left(x^k_{s\wedge\tau(x)}\right) = x + \sum_{l=2}^k c_{k,l} e^{\lambda_l s} u_l(x)$, with $i \in \{2,k\}$: $\mathbf{P}^k(A_s = i) = \left[x^i\right] \sum_{l=2}^k c_{k,l} e^{\lambda_l s} u_l(x)$ where $\left[x^i\right] f(x)$ is the coefficient of x^i for the power-series expansion of $f(x)$. Let $\tau(k) = \inf(s>0 : A_s = 1 \mid A_0 = k)$ be the time to MRCA of the ancestral process with k initial descendants. We have: $\mathbf{P}^k(\tau(k) > s) = \sum_{i=2}^k \mathbf{P}^k(A_s = i)$ which expresses as a linear combination of exponentials $e^{\lambda_i s}$, $i \in \{2,k\}$ with rational coefficients. For instance ($k=3$), $\mathbf{P}^3(\tau(3) > s) = \frac{3}{2}e^{-s} - \frac{1}{2}e^{-3s}$. This is consistent with the fact that, from the construction of $(A_s; s \geq 0)$, $\tau(k) \overset{d}{=} \sum_{l=2}^k E_l$, where $(E_l; l = 2,..,k)$ are independent random variables each exponentially distributed with parameter $\binom{l}{2}$. Kingman's tree length with k leaves is $L(k) \overset{d}{=} \sum_{l=2}^k l E_l$.

The ancestral selection graph. We now come to the question of looking at the genealogy of a Wright-Fisher model in the presence of mutations and selection.

Let (σ, u_1, u_2) be non-negative parameters. Consider now the birth and death process $(A_s; s \geq 0)$ on $\mathbf{N}$ satisfying $A_0 = k$ and with backward generator:

$$G_A\psi(b) = \left[\binom{b}{2} + bu_2\right](\psi(b-1) - \psi(b)) + b\sigma(\psi(b+1) - \psi(b)) - bu_1\psi(b)$$

for all suitable bounded $\psi : \mathbf{N} \to \mathbf{C}$. For such a process in state $b \in \mathbf{N}$, particles merge at rate $\binom{b}{2} + bu_2$, split (branch) at rate $b\sigma$ and overall killing occurs at rate bu_1.

With $\overline{x} \in (0,1)$, let $\phi_k(s,\overline{x}) = \mathbf{E}^k\left(\overline{x}^{A_s}\right) = \sum_{a=1}^k \overline{x}^a \mathbf{P}^k(A_s = a)$ be the generating function of the process $(A_s; s \geq 0)$. From the Chapman-Kolmogorov formula corresponding to G_A with tri-diagonal transition matrix, the dynamics of $\phi_k(s,\overline{x})$ reads:

$$\partial_s \phi_k(s,\overline{x}) = \left[-\sigma\overline{x}(1-\overline{x}) + u_2 - (u_1+u_2)\overline{x}\right]\partial_{\overline{x}}\phi_k(s,\overline{x}) + \frac{1}{2}\overline{x}(1-\overline{x})\partial^2_{\overline{x}}\phi_k(s,\overline{x}),$$

with initial condition $\phi_k(0,\overline{x}) = \overline{x}^k$.

Consider WF model $(x_s; s \geq 0)$ with mutations and selection for which $(\sigma, u_1, u_2) > 0$ and $f(x) = \sigma x(1-x) + u_1 - (u_1+u_2)x$ and $g^2(x) = x(1-x)$. Recall $(\overline{x}_s; s \geq 0)$ is also a WF model with mutations and selection for which $f(\overline{x}) = -\sigma\overline{x}(1-\overline{x}) + u_2 - (u_1+u_2)\overline{x}$ and $g^2(\overline{x}) = \overline{x}(1-\overline{x})$. From the dynamics of $\phi_k(s,\overline{x})$, we have:

$$\mathbf{E}^{\overline{x}}\left(\overline{x}^k_{s\wedge\tau(\overline{x})}\right) = \mathbf{E}^k\left(\overline{x}^{A_s}\right),$$

relating moments of the Wright-Fisher process $(\overline{x}_s; s \geq 0)$ to the generating function of $(A_s; s \geq 0)$. Process $(A_s; s \geq 0)$ represents the ancestral lines of the ancestral selection graph of a WF model with mutation selection parameters $(-\sigma, u_2, u_1)$ (see [14]). This graph is a generalization of Kingman's binary tree in the neutral case. Process $(A_s; s \geq 0)$ could be stopped at time $\tau(k) = \inf(s>0 : A_s = 1 \mid A_0 = k)$ which is the (almost surely finite) time to Ultimate Ancestor of the ancestral process with k initial descendants.

Proceeding further in time, as $s \to \infty$, $\overline{x}_{s\wedge\tau(\overline{x})} \overset{d}{\to} \overline{x}_\infty$ with density: $p_{st}(y) = Cy^{2u_2-1}(1-y)^{2u_1-1}e^{-2\sigma y}$, which is the normalizable invariant measure of $(\overline{x}_s; s \geq 0)$, independent of the initial condition $\overline{x}_0 = \overline{x}$. As a result, $A_s \overset{d}{\to} A_\infty$, with generating function $\mathbf{E}^k\left(\overline{x}^{A_\infty}\right) = \int_0^1 y^k p_{st}(y)\,dy$. From this, we get that, given $A_0 = k$, A_∞ is degenerate. Indeed, $A_\infty = 0$ (the absorbing state) with probability $\int_0^1 y^k p_{st}(y)\,dy$ and $A_\infty = \partial$ (the 'coffin state' reached when the chain is killed in finite time) with complementary probability. Due to selection, splitting (binary branching) occurs at rate $b\sigma$ and so the ancestral process is now allowed to increase (by one unit). There is an overall balance of branching events with merging and killing ones (which tends to shrink $(A_s; s \geq 0)$ by one unit), producing a degenerate equilibrium state with masses concentrated at 0 and ∂. The explosion time of $(A_s; s \geq 0)$ is $T_A = T_0 \wedge T_\partial$ where T_0, T_∂ are its times till absorption and killing respectively.

Particular cases:

• Assuming $u_1 = u_2 = 0$ (no mutation, only selection), as $s \to \infty$, $\overline{x}_{s\wedge\tau(\overline{x})} \overset{d}{\to} \overline{x}_\infty$ where the law of $\overline{x}_\infty$, although now degenerate, depends on the initial condition $\overline{x}_0 = \overline{x}$: indeed, $\overline{x}_\infty = 0$ with probability $\frac{e^{2\sigma}-e^{2\sigma\overline{x}}}{e^{2\sigma}-1}$, $\overline{x}_\infty = 1$ with probability $\frac{e^{2\sigma\overline{x}}-1}{e^{2\sigma}-1}$. In this case, $A_s \overset{d}{\to} A_\infty$ with generating function $\mathbf{E}^k\left(\overline{x}^{A_\infty}\right) = \frac{e^{2\sigma\overline{x}}-1}{e^{2\sigma}-1}$ (the generating function of a non-degenerate Poisson distribution with parameter 2σ, restricted to the non-null integers: $\mathbf{P}^k(A_\infty = l) = \frac{(2\sigma)^l}{(e^{2\sigma}-1)l!}$, $l \in \{1, 2, ...\}$). $(A_s; s \geq 0)$ has this distribution for invariant measure. In the absence of mutations, state 0 cannot be attained and the law of A_∞ has support $\{1, 2, ...\}$, independently of $A_0 = k$.

• Assuming $\sigma = 0$ (no selection, only mutation), we get a binary tree again, corresponding to Kingman coalescent with mutations. The support of A_∞ is again $\{0, \partial\}$ with

$$\mathbf{P}^k(A_\infty = 0) = \frac{\Gamma(2(u_1+u_2))}{\Gamma(2u_1)}\frac{\Gamma(2u_1+k)}{\Gamma(2(u_1+u_2)+k)},$$

decreasing with k.

Process $(A_s; s \geq 0)$ could be prolonged after T_∂ and stopped when it hits 1 for the first time (time to MRCA with mutation for the $k-$subsample). The Kingman coalescent with mutations till MRCA can be seen as follows: let $(E_l; l = 2, .., k)$ be independent random variables each exponentially distributed with parameter $\binom{l}{2} + lu_2$. The random variables $(E_l; l = 2, .., k)$ are the times it takes for Kingman tree with mutation rate u_2 and k leaves to pass from state l to $l-1$ as a result of a coalescence. Kingman's tree length with k leaves till MRCA is $L(k) \overset{d}{=} \sum_{l=2}^k lE_l$. Mutations occur at rate u_1 on the edges of this tree according to a Poisson point process given its edge lengths. Given $L(k)$, the number M of mutations at rate u_1 is thus Poisson distributed (with mean $L(k)$) with shifted harmonic average: $\mathbf{E}(M) = u_1 \sum_{l=2}^k \frac{1}{u_2+(l-1)/2}$.

Assume again $(\sigma, u_1, u_2) > 0$. Consider now the logistic birth and death process $(B_s; s \geq 0)$ on $\mathbf{N}$ satisfying $B_0 = k$ and with backward conservative generator:

$$G_B\psi(b) = \left[\binom{b}{2} + bu_2\right](\psi(b-1) - \psi(b)) + b\sigma(\psi(b+1) - \psi(b)).$$

This generator is the one of $(A_s; s \geq 0)$ except that no killing occurs; the overall rate of change of $(B_s; s \geq 0)$ is $\rho = \binom{b}{2} + b(u_2 + \sigma)$ and the chain's increment is $+1$ with probability $(b\sigma)/\rho$ and -1 with probability $\left[\binom{b}{2} + bu_2\right]/\rho$. We have:

$$\mathbf{E}^k\left(\overline{x}^{A_s}\right) = \mathbf{E}^k\left(\overline{x}^{B_s} e^{-u_1 \int_0^s B_\tau d\tau}\right),$$

killing $(B_s; s \geq 0)$ at rate $u_1 B_s$ to recover $(A_s; s \geq 0)$. It can be checked that the birth and death process $(B_s; s \geq 0)$ (and a fortiori $(A_s; s \geq 0)$) does not blow up at ∞ in finite time; it hits $b = 0$ in finite time, almost surely (where it gets absorbed). Letting $T_B = \inf(s > 0 : B_s = 0)$, we obtain the Laplace-Stieltjes transform of $\int_0^{T_B} B_\tau d\tau$ under the form of the $k-$th moment of $\overline{x}_\infty$:

$$\mathbf{E}^k\left(e^{-u_1 \int_0^{T_B} B_\tau d\tau}\right) = C \int_0^1 y^{2u_2+k-1} (1-y)^{2u_1-1} e^{-2\sigma y} dy.$$

The random variable $\int_0^{T_B} B_\tau d\tau$ is the area under the profile of $(B_s; s \geq 0)$ on the time interval $[0, T_B]$, given $B_0 = k$. Note that $\mathbf{E}^k\left(e^{-u_1 \int_0^{T_B} B_\tau d\tau}\right) = \mathbf{P}^k(A_\infty = 0) = \mathbf{P}^k(T_0 < T_\partial)$, the probability that $(A_s; s \geq 0)$ started at k gets absorbed before it gets killed.

Ancestral graph for WF with dominance. Let $(\sigma > 0, h \in (1/2, 1))$ be parameters. Let $\overline{h} := 1 - h$. Consider now the integral-valued process $(C_s; s \geq 0)$ satisfying $C_0 = k$ and with conservative backward generator $G_C\psi(b)$ given by:

$$\binom{b}{2}(\psi(b-1) - \psi(b)) + b\sigma\overline{h}(\psi(b+1) - \psi(b)) + b\sigma\left(1 - 2\overline{h}\right)(\psi(b+2) - \psi(b+1)),$$

for all suitable bounded $\psi : \mathbf{N} \to \mathbf{C}$. For such a process in state $b \in \mathbf{N}$, particles merge at rate $\binom{b}{2}$, split (branch) at rate $b\sigma\overline{h}$ and overall intput/output occurs at rate $b\sigma\left(1 - 2\overline{h}\right)$. Let $\mathbf{P}_s := (\mathbf{P}(C_s = 1), ..., \mathbf{P}(C_s = b), ..)$ be its row vector of probabilities. This Markov chain can also be described by its Chapman-Kolmogorov equation:

$$\frac{d}{ds}\mathbf{P}_s = \mathbf{P}_s Q + \mathbf{P}_s(Q_+ - Q_-), \ \mathbf{P}_0 = \left(0, .., 0, \underset{\to k \leftarrow}{1}, 0, ..\right).$$

Here Q is a tri-diagonal stochastic (selection) matrix defined by its entries $Q_{b,a} = -\left[\binom{b}{2} + b\sigma\overline{h}\right]$ if $a = b$, $Q_{b,a} = \binom{b}{2}$ if $a = b-1$, $Q_{b,a} = b\sigma\overline{h}$ if $a = b+1$ and $Q_{b,a} = 0$ otherwise. Matrix Q_+ is an input (creation) matrix satisfying $(Q_+)_{b,a} = 0$ except for $a = b+2$ for which $(Q_+)_{b,b+2} = b\sigma\left(1 - 2\overline{h}\right)$. Matrix Q_- is an output (killing) matrix satisfying $(Q_-)_{b,a} = 0$ except for $a = b+1$ for which $(Q_-)_{b,b+1} = b\sigma\left(1 - 2\overline{h}\right)$. Note that, for each b : $\sum_a (Q + Q_+ - Q_-)_{b,a} = 0$ so that $\mathbf{P}_s$ is indeed a conserved probability vector satisfying $\sum_b \mathbf{P}(C_s = b) = 1$ for each s (as a result of $G_C(1) = 1$). The probability mass adjunction at rate Q_+ is balanced by mass loss at rate Q_-. The birth and death process with selection governed by Q undergoes an additional transition at rate $2b\sigma\left(1 - 2\overline{h}\right)$. When creating mass (Q_+) with probability $1/2$, it passes from state b to $b+2$ where it is duplicated in starting afresh independent copies. Conversely, when loosing mass (Q_-) with probability $1/2$, it passes from state b to $b+1$ where it is killed.

With $\overline{x} \in (0,1)$, let $\phi_k(s,\overline{x}) = \mathbf{E}^k\left(\overline{x}^{C_s}\right) = \sum_{a=1}^k \overline{x}^a \mathbf{P}^k(C_s = a)$ be the generating function of the process $(C_s; s \geq 0)$. From the Chapman-Kolmogorov formula corresponding to G_C, the dynamics of $\phi_k(s,\overline{x})$ reads:

$$\partial_s \phi_k(s,\overline{x}) = \left[-\sigma\overline{x}(1-\overline{x})\left(\overline{h} - \overline{x}\left(2\overline{h}-1\right)\right)\right]\partial_{\overline{x}}\phi_k(s,\overline{x}) + \frac{1}{2}\overline{x}(1-\overline{x})\,\partial^2_{\overline{x}}\phi_k(s,\overline{x}),$$

with initial condition $\phi_k(0,\overline{x}) = \overline{x}^k$.

Consider WF model $(x_s; s \geq 0)$ with under-dominant selection for which we have $(\sigma > 0, h \in (1/2,1))$, $f(x) = \sigma x(1-x)(h - x(2h-1))$ and $g^2(x) = x(1-x)$. Recall $(\overline{x}_s; s \geq 0)$ is also a WF model with dominance the new drifts are $f(\overline{x}) = -\sigma\overline{x}(1-\overline{x})\left(\overline{h} - \overline{x}\left(2\overline{h}-1\right)\right)$ and $g^2(\overline{x}) = \overline{x}(1-\overline{x})$. From the dynamics of $\phi_k(s,\overline{x})$, we get:

$$\mathbf{E}^{\overline{x}}\left(\overline{x}^k_{s\wedge\tau(\overline{x})}\right) = \mathbf{E}^k\left(\overline{x}^{C_s}\right),$$

relating moments of the Wright-Fisher process $(\overline{x}_s; s \geq 0)$ to the generating function of $(C_s; s \geq 0)$. Process $(C_s; s \geq 0)$ represents the ancestral lines of the ancestral selection graph of a WF model with dominance parameters $\left(-\sigma < 0, 0 < \overline{h} < \frac{1}{2}\right)$.

For this parameter range, as $s \to \infty$, $\overline{x}_{s\wedge\tau(\overline{x})} \overset{d}{\to} \overline{x}_\infty$ where the law of $\overline{x}_\infty$, is given by $\overline{x}_\infty = 0$ with probability $\frac{\int_0^{\overline{x}} e^{\sigma(1-2\overline{h})(y-\overline{x}_*)^2}dy}{\int_0^1 e^{\sigma(1-2\overline{h})(y-\overline{x}_*)^2}dy}$, $\overline{x}_\infty = 1$ with complementary probability. In this case, $C_s \overset{d}{\to} C_\infty \in \{1,2,..\}$ with absolutely monotone generating function $\mathbf{E}^k\left(\overline{x}^{C_\infty}\right) = \frac{\int_0^{\overline{x}} e^{\sigma(1-2\overline{h})(y-\overline{x}_*)^2}dy}{\int_0^1 e^{\sigma(1-2\overline{h})(y-\overline{x}_*)^2}dy}$, $\overline{x} \in [0,1]$. The law of C_∞ is the invariant measure for $(C_s; s \geq 0)$.

5. Concluding Remarks

We have presented a series of elementary stochastic models arising from population genetics, with emphasis on the diffusion method. After some generalities on one-dimensional Markovian diffusions, focus was put on the specific Wright-Fisher neutral diffusion and some of its variations including various drifts of interest in genetics, specifically: mutation, selection, with and without dominance. Using similar diffusion techniques, we have discussed a general selection of paths procedure leading, when applied to WF models, to puzzling questions of biological interest. Some statistical characteristics of the underlying transformed processes were exhibited. WF diffusions describes the forward evolution of some descendant process which is the scaling limit of some discrete-time branching process with constant population size. Some aspects of their dual coalescents obtained while running the diffusion process backward in time have also been investigated. This led us to birth and death Markov processes supplying statistical insight into the ancestral lineages of the WF models presenting the drifts under study.

References

[1] Blythe, R. A.; McKane, A. J. Stochastic models of evolution in genetics, ecology and linguistics. J. Stat. Mech. No 07, P07018, 2007.

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

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

[4] Ethier, S. N.; Kurtz, T. G. *Markov processes. Characterization and convergence.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986.

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

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

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

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

[9] Karlin, S.; Taylor, H. M. *A second course in stochastic processes.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1981.

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

[11] Mandl, P. *Analytical treatment of one-dimensional Markov processes.* Die Grundlehren der mathematischen Wissenschaften, Band 151 Academia Publishing House of the Czechoslovak Academy of Sciences, Prague; Springer-Verlag New York Inc., New York 1968.

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

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

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

[15] Simon, D.; Derrida, B. Evolution of the most recent common ancestor of a population with no selection. J. Stat. Mech. No 05, P05002, 2006

[16] Tavaré, S. Ancestral inference in population genetics. Lectures on probability theory and statistics, Saint-Flour 2001, Lecture Notes in Math., 1837, (1-188) Springer, 2004.

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