On the Karlin-Kimura approaches to the Wright-Fisher diffusion with fluctuating selection

Thierry Huillet

ON THE KARLIN-KIMURA APPROACHES TO THE WRIGHT-FISHER DIFFUSION WITH FLUCTUATING SELECTION

THIERRY HUILLET

Abstract. The goal of this manuscript is a comparative study of two Wright-Fisher-like diffusion processes on the interval, one due to Karlin and the other one due to Kimura. Each model accounts for the evolution of one two-locus colony undergoing random mating, under the additional action of selection in random environment. In other words, we study the effect of disorder on the usual Wright-Fisher model with fixed (nonrandom) selection. There is a drastic qualitative difference between the two models and between the random and nonrandom selection hypotheses.

We first present a series of elementary stochastic models and tools that are needed to undergo this study in the context of diffusion processes theory, including: Kolmogorov backward and forward equations, scale and speed functions, classification of boundaries, Doob-transformation of sample paths using additive functionals. In this spirit, we briefly revisit the neutral Wright-Fisher diffusion and the Wright-Fisher diffusion with nonrandom selection.

With these tools at hand, we first deal with the Karlin approach to the Wright-Fisher diffusion model with randomized selection differentials. The specificity of this model is that in the large population case, the boundaries of the state-space are natural, hence inaccessible and so quasi-absorbing only. We supply some limiting properties pertaining to hitting times of points close to the boundaries.

Next, we study the Kimura approach to the Wright-Fisher model with randomized selection, which may be viewed as a modification of the Karlin model, using an appropriate Doob transform which we describe. This model also has natural boundaries but they turn out to be much more attracting and sticky than in its Karlin's version. This leads to a faster approach to the quasi-absorbing states, a larger time needed to move from the vicinity of one boundary to the other and to a local critical behavior of the branching diffusion obtained after the relevant Doob transformation.

Running title: Karlin-Kimura Models with Fluctuating Selection.

Keywords: Wright-Fisher diffusion, fluctuating selection, Karlin, Kimura, quasi-fixation, hitting times, natural boundaries, Doob transform, branching diffusion, local criticality.

1. Introduction and outline of the results

The goal of this manuscript is the study of two diffusion processes on the unit interval, due to *Karlin* for the first and to *Kimura* for the second. *Each model describes the evolution of one two-locus colony undergoing random mating, under the additional action of selection in random environment. The effect of disorder*

on the usual Wright-Fisher model with fixed (nonrandom) selection is huge and the conclusions to be drawn in each case radically different.

Besides the pioneer works [18] and [21] on which our own work chiefly relies, we wish to cite the related works [19], [28], [24], [11] and [34], also dealing with temporally varying selection coefficients.

Before proceeding with this program, we first revisit a series of elementary stochastic diffusion models and needed tools which found their way over the last sixty years, chiefly in mathematical population genetics [In this context, we refer to [30] and to its extensive and non-exhaustive list of references for historical issues in the development of modern mathematical population genetics (after Wright, Fisher, Crow, Kimura, Nagylaki, Maruyama, Ohta, Watterson, Ewens, Kingman, Griffiths, Tavaré...). See also the general monographs [3], [26], [9], [35] and [12]].

Section 2 is therefore devoted to generalities on one-dimensional diffusions on the unit interval $[0,1]$, with the usual Wright-Fisher (WF) diffusion in mind. It is designed to fix the background and notations, following and completing a similar treatment in [15]. *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 the speed and scale functions, existence of an invariant measure, validity of detailed balance, are addressed in the light of the Feller classification of boundaries. The important problem of evaluating additive functionals along sample paths is then briefly recalled, 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. We will make use of one of them in Section 5.

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. They are very useful in the context of conditioning the original process in various ways. *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, including a multiplicative killing part, in general.* 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. For example, *we show how to obtain the Yaglom limit of the transformed process conditioned on not being yet killed nor absorbed, exactly in the same way as the Yaglom limit of the original process simply conditioned on not being yet absorbed can be obtained.*

In Section 3, we briefly and informally recall how *the specific continuous space-time Wright-Fisher diffusion models, either neutral on nonneutral, can be obtained as scaling limits of a biased discrete Poisson-Galton-Watson model with a conservative number of offsprings over the generations.* This allows to introduce various drifts of biological interest to force the neutral WF model in specific directions.

Because our study is chiefly about selection, we focus on the bias to neutrality arising from (deterministic) selection or fitness.

In the forthcoming three Sections, we study the announced problem of WF diffusion in random uncorrelated environment. For interested readers not willing to go through all technical details, we now summarize the obtained results. Whenever it is the case, pointers to existing known results borrowed from the literature will be introduced, in the course of the exposition of the results.

In Section 4, *we deal with the Wright-Fisher model with randomized selection differentials, as first studied by Karlin,* [18]*. In Karlin's approach, the constant selection intensities at each generation appearing in the usual WF model with selection are regarded as a random iid selection sequence. By doing so, we study the effect of disorder on the WF diffusion model with nonrandom selection.*

In the diffusion approximation of the Karlin model, the local variance term turns out to be the sum of two contributions, respectively one due to binomial sampling (the genetic drift) and the other to the within generation selection variance. In the latter contribution, the amplitude parameter ρ quantifying the selection intensities fluctuations appears. The obtained drift is cubic. It is also the sum of two contributions. In the first one, a parameter γ representing selective advantage of allele 1 over allele 2 comes in. It is not simply the difference between the mean selection intensities as it also involves their second-order moment properties (this is a characteristic ingredient of random environments). The second contribution to the drift term is also affected by the amplitude of the selection intensities fluctuations. It accounts for a stabilizing drift toward $1/2$. Both boundaries $\{0,1\}$ of this new process being exit, this process is transient, just like the usual WF model with nonrandom selection coefficients. Using techniques developed in Section 2, we compute the expected time to fixation at the boundaries with the help of the Green function of the model. In the symmetric or neutral case with equality in distribution of the selection differentials (no selective advantage of any allele and $\gamma = 0$), the expected time to fixation is seen to decrease with the amplitude ρ of their fluctuations (a result first discussed in [17] *and* [14]*). In this neutral case, the drift term of the diffusion model is restricted to the stabilizing drift toward $1/2$ proportional to ρ, tending to increase the fixation time, competing with an increase of the selection intensities fluctuations tending to make it shorter.*

In the large population setup, we can drop the binomial sampling contribution to the variance term in the former diffusion model with randomized selection. For this new model, the boundaries become inaccessible and natural. The transition probability kernel can easily be computed showing that if $\gamma > 0$ (< 0), the mass of the law of the process accumulates near 1 (respectively 0) as time passes by. In the symmetric or neutral case ($\gamma = 0$), this law forms two symmetric peaks about both 0 and 1, but without reaching the boundaries in finite time, expressing that quasi-fixation (or quasi-loss) occurs, a phenomenon first observed and described by Kimura in [21].

Still with $\gamma = 0$, we estimate the probability starting from $x \in I_\varepsilon = [\varepsilon, 1-\varepsilon]$ to hit first $1-\varepsilon$ before ε ($\varepsilon > 0$ small). We conclude that, independently of ρ, when $x < \frac{1}{2}$

($x > \frac{1}{2}$), this probability is slightly smaller (larger) than $1/2$ *with a correcting term of order* $\frac{1}{-\log\varepsilon} \cdot \log\left(\frac{x}{1-x}\right)$.

Then we proceed with estimating the expected first exit time of I_ε. *We show that it diverges like* $\frac{1}{\rho}\left[\log(\varepsilon)\right]^2$ *as* $\varepsilon \to 0$, *somehow quantifying how inaccessible the natural boundaries of this model are. As intuitively required, this quantity is a decreasing function of* ρ.

Assume the process does not leave I_ε. *As a measure of the process spending a substantial amount of time in the neighborhood of the boundaries of* I_ε *where homozygosity is largest, we compute a conditional empirical average of the heterozygosity over the sample paths. We show that it tends to* 0 *like* $\frac{1}{-\log(\varepsilon)}$ *as* $\varepsilon \to 0$, *independently of* ρ.

We finally show that the time to move from ε *to* $1-\varepsilon$ *is large in the following sense. Using the Green potential function of the Kimura model, we compute the law of* $\tau_{\varepsilon,1-\varepsilon}$ *which is the first time that the process hits* $1-\varepsilon$ *starting from* ε. *We get that, as* $\varepsilon \to 0$, $\frac{\rho}{2^3[\log(1/\varepsilon)]^2}\tau_{\varepsilon,1-\varepsilon}$ *converges in distribution to a stable law of index* $1/2$. *The order of magnitude of* $\tau_{\varepsilon,1-\varepsilon}$ *is thus* $2^3\left[\log(1/\varepsilon)\right]^2\rho^{-1} \underset{\varepsilon\to 0}{\to} \infty$. *This quantity is decreasing with* ρ.

In Section 5, *we introduce the Wright-Fisher-Kimura model with fluctuating selection intensities which turns out to be a significant modification of the Karlin model. In the Karlin diffusion models, the stochastic differential equations were understood in the sense of Itô. Looking at their equivalent Stratonovitch form, the drift is modified accordingly; it becomes quadratic and the* $\rho-$*dependent part of the Karlin drift vanishes. Kimura considered a selection model in random environment where the latter Stratonovitch diffusion was now understood in the classical sense of Itô; see* [21]. *The drift term of the Kimura model no longer depends on the amplitude of the selection intensities fluctuations, killing the stabilizing effect towards* $1/2$ *of the Karlin model studied in Section* 4. *As a result, one expects the boundaries of the Kimura model to be much more attracting and sticky than in its previous Karlin version.*

Finally, in the last Section 6, *we study the Kimura model in itself proposing in the process a systematic comparative study of the Karlin and Kimura models.*

We start with the easier symmetric case when $\gamma = 0$ *(the Kimura martingale). We show that it has again two natural inaccessible boundaries; the process is still null-recurrent. For this driftless Kimura model, the transition probability density can easily be obtained in closed-form. It converges more rapidly than for its Karlin version to the quasi-absorption states* $\{0,1\}$.

We show that the probability starting from $x \in I_\varepsilon = [\varepsilon, 1-\varepsilon]$ *to hit first* $1-\varepsilon$ *before* ε *is no longer of order* $1/2$. *Rather, still independently of* ρ, *it is given by* $\frac{x-\varepsilon}{1-2\varepsilon}$ *showing that the initial condition has a greater influence on this event than in the Karlin model.*

Next, we estimate the expected first exit time of I_ε. *We show that it diverges like* $-\frac{2}{\rho}\log(\varepsilon)$, *which is smaller than* $\frac{1}{\rho}[\log(\varepsilon)]^2$ *obtained previously in the Karlin context. This quantity is again a decreasing function of* ρ. *The Kimura model hits therefore the boundaries of* I_ε *in a much shorter time on average than its Karlin's counterpart.*

The conditional empirical average measure of heterozygosity for the Kimura martingale also tends to 0 *when* $\varepsilon \to 0$, *but like* $-2\varepsilon\log\varepsilon$, *which is much faster than in the Karlin model. In this sense, the Kimura martingale spends much more time close to the boundaries than the corresponding Karlin process, because in the latter case there is no more a stabilizing drift toward* $1/2$.

Using the Green potential function technique, we compute next the law of the time needed to move from ε *to* $1-\varepsilon$ *for the Kimura martingale. We show that a limiting damped* $\frac{1}{2}-$*stable law is involved. We find in particular that the probability that a move from* ε *to* $1-\varepsilon$ *occurs in finite time is of order* ε : *Whenever the Kimura martingale approaches a boundary closely, its probability not to move back to the opposite boundary approaches* 1 *accordingly. This opportunity did not exist when dealing with the Karlin model. It shows how sticky the boundaries became, with a behavior closer to the one expected would the quasi-absorbing boundaries of the Kimura model be simply absorbing. Moreover, given the switch from* ε *to* $1-\varepsilon$ *occurs (an event with small probability), the expected waiting time is shown to be of order* $-4\log(\varepsilon)/\rho \underset{\varepsilon\to 0}{\to} \infty$.

Finally we study the non-symmetric Kimura model with a drift ($\gamma \neq 0$*). We first check that the boundaries are always natural. We show next that the Kimura process can be obtained from the Karlin model with fluctuating selection after using a selection of paths procedure favoring paths whose square-root of the reciprocal heterozygosity is large. This Doob transform therefore favors Karlin's model sample paths staying close to the boundaries, resulting in a process whose boundaries are much more sticky than in its non-modified Karlin version. This construction connects the two models in a clear way. The Doob transform produces an additive multiplicative component in the Kimura generator* $\overline{G}$ *of the transformed process, which is a birth and death rate: the new process can therefore be interpreted in terms of a branching diffusion process of Kimura particles which we describe precisely.*

We show that this branching diffusion process does not fit into the general framework of positively regular branching models developed in [1] *and* [2], *leading to strong law of large numbers for global population growth or extinction. Rather, it fits to the general framework developed in* [8] *for local population growth or extinction. More precisely, we conclude that* $|\gamma| = \rho/2$ *is a case of local criticality. Would* $|\gamma| < \rho/2$, *then local supercriticality (growth) holds: the expected number of Kimura particles grows exponentially locally within each Borel subset with closure inside* $(0,1)$ *at rate* $-\lambda_c = \frac{\rho}{8}\left(1-4\left(\frac{\gamma}{\rho}\right)^2\right) > 0$. *Would* $|\gamma| > \rho/2$, *then local subcriticality (extinction) holds at the decay rate* λ_c: *the drift towards the boundaries (either* $\{0,1\}$*) is so strong that it pushes all the Kimura particles very close to either boundaries where they remain stuck, all ending up outside any such Borel subset in finite time. This approach relies on the criticality of the operator* $\overline{G}(\cdot) + \lambda_c\cdot$ *(and its adjoint), in*

the sense of [31]. *The ground states of the critical branching Kimura operators are exactly identified.*

2. Preliminaries on diffusions on the unit interval

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 [6], [7], [20] and [25]. This Section is designed to fix the background and notations for the rest of the paper.

2.1. **Generalities on one-dimensional diffusions on the interval** $[0,1]$**.** 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 the interval say $I = [0,1]$, see [16]. Assume it has locally Lipschitz continuous drift $f(x)$ and local standard deviation (volatility) $g(x)$, namely consider the stochastic differential equation (SDE):

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

The condition on $f(x)$ and $g(x)$ guarantees in particular that there is no point x_* in the interior $\overset{\circ}{I} := (0,1)$ 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 $S_t := e^{tG}$, $u := u(x,t) = \mathbf{E}\psi(x_{t\wedge\tau_x})$ satisfies the Kolmogorov backward equation (KBE)

$$\partial_t u = G(u)\,;\ u(x,0) = \psi(x)\,. \tag{2}$$

In the definition of the mathematical expectation u, we have $t \wedge \tau_x := \inf(t, \tau_x)$ where τ_x indicates a random time at which the process should possibly be stopped, given the process was started at x. The description of this (adapted) absorption time is governed by the type of boundaries which $\{0,1\}$ are to $(x_t; t \geq 0)$. We shall return to this point later.

2.2. **Natural coordinate, scale and speed measure.** 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 $(x_0, y_0) \in (0,1)$, one finds

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

One should choose a version of 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 $[\varphi(0), \varphi(1)]$. Its volatility is $\widetilde{g}(y) := (\varphi' g)\left(\varphi^{-1}(y)\right)$. The function φ is often called the scale function.

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

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

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

Finally, considering the random time change $t \to \theta_t$ 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 $\delta_y(\cdot) = \text{weak}-\lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon}\mathbf{1}_{(y-\varepsilon, y+\varepsilon)}(\cdot)$ stand for the Dirac delta mass at y. The random time θ_t can be expressed as

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

where $m(x) := 1/\left(g^2\varphi'\right)(x)$ is the (positive) speed density at $x = \varphi^{-1}(y)$ and $L_t(y) := \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon}\int_0^t \mathbf{1}_{(y-\varepsilon, y+\varepsilon)}(w_s)\, ds$ the local time at y of Brownian motion before time t. Both the scale function φ and the speed measure $d\mu = m(x) \cdot dx$ are therefore essential ingredients to reduce the original stochastic process $(x_t; t \geq 0)$ to the standard Brownian motion $(w_t; t \geq 0)$. 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. The Kolmogorov backward infinitesimal generator G may be written in Feller form

$$G(\cdot) = \frac{1}{2}\frac{d}{d\mu}\left(\frac{d}{d\varphi}\cdot\right). \tag{3}$$

Examples (from population genetics).

• 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.

• 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. The parameters u_1, u_2 can be interpreted 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 $g^2(x) = x(1-x)$. This is the WF model with selection. For this diffusion (see, [23]), $\varphi(x) = \frac{1-e^{-2\sigma x}}{1-e^{-2\sigma}}$ 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 the WF model with mutations and selection parameters $(u_1, u_2; \sigma)$.

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 the speed density $m(x)\propto x^{2u_1-1}(1-x)^{2u_2-1}e^{2\sigma x}$ is integrable. ♦

2.3. **The transition probability density.** Assume that $f(x)$ 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):

$$\partial_t p=G^*(p)\,,\ p(x;0,y)=\delta_y(x) \tag{4}$$

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). It can be that $p(x;t,y)$ is a sub-probability in which case, letting $\rho_t(x):=\int_0^1 p(x;t,y)\,dy$, we have $\rho_t(x)=\mathbf{P}(\tau_x>t)$ and this tail distribution is different from 1 unless the 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 ([20], Chapter 15, Section 13) and so detailed balance holds:

$$m(x)\,p(x;t,y)=m(y)\,p(y;t,x)\,,\ 0<x,y<1. \tag{5}$$

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

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

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

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

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

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

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

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

The term $b_t(x):=-\partial_t\rho_t(x)/\rho_t(x)>0$ is the time-dependent birth rate at which mass should be created to compensate the loss of mass of the original process due, say, to absorption of $(x_t;t\geq 0)$ at the boundaries. In this creation of mass process, a diffusing particle started in x dies at rate $b_t(x)$ at point (t,y) where it is duplicated in two new independent particles both started at y (resulting in a global birth) evolving in the same diffusive way [1]. The birth rate function $b_t(x)$ depends here on x and t, not on y.

[1]Consider a diffusion process with forward infinitesimal generator G^* governing the evolution of $p(x;t,y)$. Suppose that a sample path of this process has some probability that it will be killed or create a new copy of itself, and that the killing and birth rates d and b depend on the current location y of the path. Then the process with the birth and death opportunities of a path has the

When the boundaries of x_t are absorbing, the spectrum of both $-G$ and $-G^*$ is discrete, meaning that there exist positive eigenvalues $(\lambda_k)_{k\geq 1}$ ordered in ascending sizes and eigenvectors $(v_k, u_k)_{k\geq 1}$ of both $-G^*$ and $-G$ satisfying $-G^*(v_k) = \lambda_k v_k$ and $-G(y_k) = \lambda_k u_k$. Let $\lambda_1 > \lambda_0 = 0$ be the smallest non-null eigenvalue of the infinitesimal generator $-G^*$ (and of $-G$). Clearly, $-\frac{1}{t}\log \rho_t(x) \underset{t\to\infty}{\to} \lambda_1$ and by L' Hospital rule therefore $b_t(x) \underset{t\to\infty}{\to} \lambda_1$. Putting $\partial_t q = 0$ in the latter evolution equation, independently of the initial condition x

$$q(x;t,y) \underset{t\to\infty}{\to} q_\infty(y) = v_1(y), \tag{6}$$

where v_1 is the eigenvector of $-G^*$ associated to λ_1, satisfying $-G^*(v_1) = \lambda_1 v_1$. The limiting probability v_1/norm (after a proper normalization) is called the Yaglom limit law of $(x_t; t \geq 0)$ conditioned on being currently alive at all time t.

Example. When dealing with the neutral Wright-Fisher diffusion, it is known that $\lambda_1 = 1$ with $v_1 \equiv 1$. The Yaglom limit in this case is the uniform measure. ♦

2.4. **Feller classification of boundaries.** The KBE equation may not have unique solutions, unless one specifies the conditions at the boundaries $\{0,1\}$.

For $1-$dimensional diffusions as in (1) on $[0,1]$, the boundaries $\partial I := \{0,1\}$ are of two types: either accessible or inaccessible. Accessible boundaries are either regular or exit boundaries, whereas inaccessible boundaries are either entrance or natural boundaries. Integrability of the scale function and the speed measure turn out to be essential in the classification of boundaries due to Feller [10].

In the sequel, the symbol $\circ$ will designate either 0 or 1. 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 $\forall y_0 \in (0,1)$:

$$(i)\ \ \varphi'(y) \in L_1(y_0,\circ) \ \text{ and } \ (ii)\ \ m(y) \in L_1(y_0,\circ)$$

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 the KBE at $\circ$; we shall

infinitesimal generator $\lambda(y)\cdot + G^*(\cdot)$, where $\lambda(y) = b(y) - d(y)$. The rate can also depend on t and x.

return to this point later. $\diamond$

$(A2)$ The boundary $\circ$ is an exit boundary if $\forall y_0 \in (0,1)$:

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

In this case, a sample path of $(x_t; t \geq 0)$ can reach $\circ$ from the inside of I in finite time but cannot reenter. The sample paths are absorbed at $\circ$. There is an absorption at $\circ$ at time $\tau_\circ(x) = \inf(t > 0 : x_t = \circ \mid x_0 = x)$ and $\mathbf{P}(\tau_\circ(x) < \infty) = 1$. Whenever both boundaries $\{0,1\}$ are absorbing, the diffusion x_t should be stopped at $\tau_x = \tau_{x,0} \wedge \tau_{x,1}$. 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 $\{0,1\}$ is absorbing, $\tau_x = +\infty$. Examples of diffusion with exit boundaries are the driftless neutral WF model and the WF model with selection.

$(I1)$ The boundary $\circ$ is an entrance boundary if $\forall y_0 \in (0,1)$:

$$(i)\ \ \varphi'(y) \notin L_1(y_0, \circ), \ \ (ii)\ \ m(y) \in L_1(y_0, \circ)$$
$$(iii)\ \ m(y) \int_{y_0}^{y} \varphi'(z)\, dz \in L_1(y_0, \circ).$$

An entrance boundary is clearly 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 $[0,1]$ but cannot return to $\circ$ from the interior of $[0,1]$. 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 $[0,1]$; 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 $\varphi(0) = -\infty$ and $\varphi(1) = +\infty$.

$(I2)$ The boundary $\circ$ is natural in all other cases. When $\circ$ is natural, sample paths cannot enter nor quit $[0,1]$ and sample paths are trapped inside $[0,1]$ with $\{0,1\}$ inaccessible; the 'simplest' case is when $(x_t; t \geq 0)$ is itself a Brownian motion. In Sections 4 and 5 dedicated to WF diffusions with selection in random environment, we will encounter some other examples of diffusions with natural boundaries.

2.5. **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). This process is thus transient. We wish to evaluate the non-negative additive quantities

$$\alpha(x) = \mathbf{E}\left(\int_0^{\tau_x} c(x_s)\, ds + d(x_{\tau_x})\right), \tag{7}$$

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

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

Examples.

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

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

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

Let then $\alpha =: \alpha_1(x) = \mathbf{P}(x_t \text{ first hits } [0,1] \text{ at } 1 \mid x_0 = x)$. $\alpha_1(x)$ is a $G-$harmonic function solution to $G(\alpha_1) = 0$, with boundary conditions $\alpha_1(0) = 0$ and $\alpha_1(1) = 1$. Solving this problem, we get: $\alpha_1(x) = \int_0^x dy e^{-2\int_0^y \frac{f(z)}{g^2(z)}dz} / \int_0^1 dy e^{-2\int_0^y \frac{f(z)}{g^2(z)}dz}$.

On the contrary, choosing $\alpha_0(x)$ to be a $G-$harmonic function with boundary conditions $\alpha_0(0) = 1$ and $\alpha_0(1) = 0$, $\alpha_0(x) = \mathbf{P}(x_t \text{ first hits } [0,1] \text{ at } 0 \mid x_0 = x) = 1 - \alpha_1(x)$.

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

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

$\mathfrak{g}$ is therefore 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

$$\mathfrak{g}(x,y) = 2\frac{(\varphi(x) - \varphi(0))(\varphi(1) - \varphi(y))}{(g^2\varphi')(y)(\varphi(1) - \varphi(0))} \text{ if } x < y$$

$$\mathfrak{g}(x,y) = 2\frac{(\varphi(1) - \varphi(x))(\varphi(y) - \varphi(0))}{(g^2\varphi')(y)(\varphi(1) - \varphi(0))} \text{ if } x > y \tag{8}$$

$$\mathfrak{g}(y,y) = 2\frac{(\varphi(1) - \varphi(y))(\varphi(y) - \varphi(0))}{(g^2\varphi')(y)(\varphi(1) - \varphi(0))} \text{ if } x = y$$

The Green function is of particular interest to solve the general problem of evaluating additive functionals $\alpha(x)$. Indeed, as is well-known, see ([20]) for example, the integral operator with respect to the Green kernel inverts the second order operator $-G$ leading to

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

4. Also of interest are the additive functionals of the type

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

where the functions c and d are again both assumed to be non-negative. The functional $\alpha_\lambda(x) \geq 0$ solves the Dynkin problem ([20]):

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

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

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

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

is the $\lambda-$potential function, solution to:

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

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

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

The $\lambda-$potential function is also useful in the computation of the distribution of the first-passage time $\tau_{x,y}$ to y starting from x. From the convolution formula

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

and taking the Laplace transform of both sides with respect to time, we obtain the Laplace-Stieltjes transform (LST) of the law of $\tau_{x,y}$ as

$$\mathbf{E}\left(e^{-\lambda\tau_{x,y}}\right) = \frac{\mathfrak{g}_\lambda(x,y)}{\mathfrak{g}_\lambda(y,y)}. \tag{9}$$

We have $\mathbf{P}(\tau_{x,y} < \infty) = \frac{\mathfrak{g}_0(x,y)}{\mathfrak{g}_0(y,y)} \in (0,1)$ as a result of both terms in the ratio being finite and x, y belonging to the same transience class of the process (under our assumptions that the boundaries are absorbing). ♦

2.6. **Transformation of sample paths (Doob transform).** Consider a one-dimensional diffusion $(x_t; t \geq 0)$ as in (1) with absorbing barriers. Let $p := p(x;t,y)$ be its transition probability and let τ_x be its absorption time at the boundaries.

Let $\alpha(x) := \mathbf{E}\left(\int_0^{\tau_x} c(x_s)\, ds + d(x_{\tau_x})\right)$ 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 the functions c and d are both chosen non-negative so that so is α.

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

$$\overline{p}(x;t,y) = \frac{\alpha(y)}{\alpha(x)} p(x;t,y). \tag{10}$$

In this construction of $(\overline{x}_t; t \geq 0)$ 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 [6]).

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

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

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

$$\overline{G}(\cdot) = \frac{1}{\alpha} G(\alpha)\cdot + \widetilde{G}(\cdot) = -\frac{c}{\alpha}\cdot + \widetilde{G}(\cdot) \tag{11}$$

and the new KB operator can be obtained from the latter by adding a drift term $\frac{\alpha'}{\alpha} g^2 \partial_x$ to the one in G of the original process to form a new process $(\widetilde{x}_t; t \geq 0)$ with the KB operator $\widetilde{G}$ and by killing its sample paths at death rate $d(x) := \frac{c}{\alpha}(x)$ (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

$$d\widetilde{x}_t = \widetilde{f}(\widetilde{x}_t)\, dt + g(\widetilde{x}_t)\, dw_t,\ \widetilde{x}_0 = x \in (0,1), \tag{12}$$

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

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 $\widetilde{\alpha}(x) := \widetilde{\mathbf{E}}^x \left(\int_0^{\overline{\tau}(x)} \widetilde{c}(\widetilde{x}_s)\, ds + \widetilde{d}\left(\widetilde{x}_{\overline{\tau}(x)}\right) \right)$ be such an additive functional where functions $\widetilde{c}$ and $\widetilde{d}$ are themselves both non-negative. It solves

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

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

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

Normalizing and conditioning.

Consider again the process with infinitesimal generator $\overline{G}$ losing mass due to absorption and/or killing at the boundaries. Integrating over y, with $\overline{\rho}_t(x) := \int_{\mathring{I}} \overline{p}\,(x;t,y)\,dy = \widetilde{\mathbf{P}}(\overline{\tau}_x > t)$, we have

$$\partial_t \overline{\rho}_t\,(x) = \overline{G}(\overline{\rho}_t\,(x)) = -d\,(x)\,\overline{\rho}_t\,(x) + \widetilde{G}(\overline{\rho}_t\,(x)), \tag{13}$$

with $\overline{\rho}_0\,(x) = \mathbf{1}_{(0,1)}\,(x)$. This gives the tail distribution of the full stopping time $\overline{\tau}_x$.

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

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

The term $\overline{b}_t\,(x) = -\partial_t \overline{\rho}_t\,(x)\,/\overline{\rho}_t\,(x) > 0$ 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 its possible absorption at the boundaries and/or killing. Again, we have $\overline{b}_t\,(x) \to \lambda_1$ where λ_1 is the smallest positive eigenvalue of $-G$ and therefore, putting $\partial_t \overline{q} = 0$ in the latter evolution equation, we get that, independently of the initial condition x

$$\overline{q}\,(x;t,y) \underset{t\to\infty}{\to} \overline{q}_\infty\,(y)\,, \tag{14}$$

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

$$\begin{aligned} -\widetilde{G}^*(\overline{q}_\infty) &= (\lambda_1 - d\,(y)) \cdot \overline{q}_\infty, \text{ or} \\ -\overline{G}^*\,(\overline{q}_\infty) &= \lambda_1 \cdot \overline{q}_\infty. \end{aligned}$$

With v_1 the eigenvector of $-G^*$ associated to λ_1, $\overline{q}_\infty\,(y)$ is of the product form

$$\overline{q}_\infty\,(y) = \alpha\,(y)\,v_1\,(y)\,/\int_0^1 \alpha\,(y)\,v_1\,(y)\,dy. \tag{15}$$

This results directly from the fact that $\overline{G}^*(\cdot) = \alpha\,(y)\,G^*(\cdot/\alpha\,(y))$ and that v_1 is the stated eigenvector of $-G^*$. A different way to see this is as follows. We have

$$\overline{\rho}_t\,(x) = \frac{1}{\alpha\,(x)} \int_0^1 \alpha\,(y)\,p\,(x;t,y)\,dy$$

and the conditional density of $\overline{x}_t$ given $\overline{\tau}_x > t$ is therefore:

$$\overline{q}\,(x;t,y) = \frac{\alpha\,(y)\,p\,(x;t,y)}{\int_0^1 \alpha\,(y)\,p\,(x;t,y)\,dy}.$$

The rest follows from observing that, to the leading order in t, for large time

$$p\,(x;t,y) \sim b_1 e^{-\lambda_1 t} \cdot u_1\,(x)\,v_1\,(y)\,,$$

where u_1 (respectively v_1) is the eigenvector of $-G$ (respectively $-G^*$) associated to λ_1 and $b_1 = \left(\int_0^1 dy u_1\,(y)\,v_1\,(y)\right)^{-1}$. From this, it is clear that $-\frac{1}{t}\log \overline{\rho}_t\,(x) \underset{t\to\infty}{\to} \lambda_1$ and

$$\overline{q}\,(x;t,y) \sim \frac{e^{-\lambda_1 t} \cdot \alpha\,(y)\,v_1\,(y)}{e^{-\lambda_1 t} \cdot \int_0^1 \alpha\,(y)\,v_1\,(y)\,dy} = \overline{q}_\infty\,(y)\,.$$

The limiting probability $\overline{q}_\infty = \alpha v_1/\text{norm}$ can therefore be interpreted as the Yaglom limit law of $(\overline{x}_t; t \geq 0)$ conditioned on the event $\overline{\tau}_x > t$. See [32].

Specific transformations of interest.

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

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

with

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

giving the new drift. In the second case, $\alpha(x) = \frac{\int_x^1 e^{-2\int_0^y \frac{f(z)}{g^2(z)}dz} dy}{\int_0^1 e^{-2\int_0^y \frac{f(z)}{g^2(z)}dz} dy}$ and boundary 0 is exit whereas 1 is entrance. Thus $\widetilde{\tau}_x$ is just the exit time at $x = 1$ (respectively at $x = 0$). Let $\widetilde{\alpha}(x) := \widetilde{\mathbf{E}}(\widetilde{\tau}_x)$. Then, $\widetilde{\alpha}(x)$ solves $-\widetilde{G}(\widetilde{\alpha}) = 1$, whose explicit solution is:

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

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

Examples.

(i) 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 (12) has the new drift: $\widetilde{f}(x) = \sigma x(1 - x) \coth(\sigma x)$, independently of the sign of σ. It models the WF diffusion with selection conditioned on exit at $\circ = 1$.

(ii) Assume α now solves $-G(\alpha) = 1$ if $x \in \overset{\circ}{I}$ with boundary conditions $\alpha(0) = \alpha(1) = 0$. In this case study, one selects sample paths of $(x_t; t \geq 0)$ with a large mean absorption time $\alpha(x) = \mathbf{E}(\tau_x)$. 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 Green function (8). 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 $\overline{\tau}_x$ of $(\widetilde{x}_t; t \geq 0)$ is just its killing time $\widetilde{\tau}_{x,\partial}$. Let $\widetilde{\alpha}(x) := \widetilde{\mathbf{E}}(\widetilde{\tau}_{x,\partial})$. Then, $\widetilde{\alpha}(x)$ solves $-\overline{G}(\widetilde{\alpha}) = 1$, $\widetilde{\alpha}(0) = \widetilde{\alpha}(1) = 0$, with explicit solution:

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

(*iii*) Assume α now solves $-G(\alpha) = \delta_y(x)$ if $x \in \overset{\circ}{I}$ with boundary conditions $\alpha(0) = \alpha(1) = 0$. In this case study, one selects sample paths of $(x_t; t \geq 0)$ with a large sojourn time density at y recalling $\alpha(x) =: \mathfrak{g}(x,y) = \mathbf{E}\left(\int_0^{\tau_x} \delta_y(x_s)\, ds\right)$. The stopping time $\widetilde{\tau}_y(x)$ of $(\widetilde{x}_t; t \geq 0)$ occurs at rate $\delta_y(x)/\mathfrak{g}(x,y)$. It is a killing time when the process is at y for the last time after a geometrically distributed number of passages there with rate $1/\mathfrak{g}(x,y)$ (or with success probability $1/(1+\mathfrak{g}(x,y))$). Let $\widetilde{\alpha}_y(x) := \widetilde{\mathbf{E}}(\widetilde{\tau}_y(x))$. Then, $\widetilde{\alpha}_y(x)$ solves $-\overline{G}(\widetilde{\alpha}) = 1$, with explicit solution:

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

When $x = 1/N$, $\widetilde{\alpha}_y(1/N)$ may be viewed as the age of a mutant currently observed to the present frequency y, see [13].

(*iv*) Let λ_1 be the smallest non-null eigenvalue of the infinitesimal generator G. Let $\alpha = u_1$ be the corresponding eigenvector, that is satisfying $-G(u_1) = \lambda_1 u_1$ with boundary conditions $u_1(0) = u_1(1) = 0$. Then $c = \lambda u_1$. The new KB operator associated to the transformed process $(\overline{x}_t; t \geq 0)$ is

$$\overline{G}(\cdot) = \frac{1}{\alpha} G(\alpha) \cdot + \widetilde{G}(\cdot) = -\lambda_1 \cdot + \widetilde{G}(\cdot),$$

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

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

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

$$p(x;t,y) \sim e^{-\lambda_1 t} \frac{u_1(x)\, v_1(y)}{\int_0^1 u_1(y)\, v_1(y)\, dy},$$

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

$$\widetilde{p}(x;t,y) \sim e^{\lambda_1 t} \frac{u_1(y)}{u_1(x)} e^{-\lambda_1 t} \frac{u_1(x)\, v_1(y)}{\int_0^1 u_1(y)\, v_1(y)\, dy} = \frac{u_1(y)\, v_1(y)}{\int_0^1 u_1(y)\, v_1(y)\, dy}. \tag{16}$$

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

$$\widetilde{p}(x;t,y) \underset{t\to\infty}{\to} \widetilde{p}(y) = \frac{1}{Zg^2(y)} e^{2\int_0^y \frac{f(z) + \left(\frac{u_1'}{u_1} g^2\right)(z)}{g^2(z)} dz} = \frac{u_1^2(y)}{Zg^2(y)} e^{2\int_0^y \frac{f(z)}{g^2(z)} dz}, \tag{17}$$

where Z is the appropriate normalizing constant. Comparing (16) and (17)

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

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

When dealing for example with the neutral Wright-Fisher diffusion, it is known that $\lambda_1 = 1$ with $u_1 = x(1-x)$ and $v_1 \equiv 1$. The limit law of the $Q-$process $(\widetilde{x}_t; t \geq 0)$ in this case is $6y(1-y)$. For additional similar examples in the context of WF diffusions and related ones, see [15]. ♦

3. The Wright-Fisher example

In this Section, we briefly and informally recall how the specific continuous space-time Wright-Fisher diffusion like models, either neutral on nonneutral, can be obtained as scaling limits of a biased discrete Poisson-Galton-Watson model with a conservative number of offsprings over the generations. This allows to introduce various drifts of biological interest to force the neutral WF model in specific directions. Because our study is chiefly about selection, we focus on the bias to neutrality arising from (deterministic) selection or fitness.

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 $|\mathbf{k}_N| = N$:

$$\mathbf{P}(\boldsymbol{\nu}_N = \mathbf{k}_N) = \frac{N! \cdot N^{-N}}{\prod_{n=1}^N k_n!}. \tag{18}$$

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 remains with constant size at all generations.

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

$$\mathbf{P}(N_{r+1}(n) = k' \mid N_r(n) = k) = \mathbf{P}(\nu_1 + .. + \nu_k = k') = \binom{N}{k'} \left(\frac{k}{N}\right)^{k'} \left(1 - \frac{k}{N}\right)^{N-k'}.$$

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

$$dx_t = \sqrt{x_t(1-x_t)} dw_t,\ x_0 = x. \tag{19}$$

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

is the diffusion approximation of the offspring frequency of a singleton at generation $[Nt]$.

Equation (19) is a $1-$dimensional diffusion as in (1) on $I=[0,1]$, with zero drift $f(x)=0$ and volatility $g(x)=\sqrt{x(1-x)}$. This diffusion is already in natural coordinate and so $\varphi(x)=x$. The scale function is x and the speed measure $[x(1-x)]^{-1}dx$. One can check that both boundaries are exit in this case: The stopping time is $\tau_x=\tau_{x,0}\wedge\tau_{x,1}$ where $\tau_{x,0}$ is the extinction time and $\tau_{x,1}$ the fixation time. The corresponding infinitesimal generators are $G(\cdot)=\frac{1}{2}x(1-x)\partial_x^2(\cdot)$ and $G^*(\cdot)=\frac{1}{2}\partial_y^2(y(1-y)\cdot)$.

3.2. **Non-neutral cases.** Two alleles Wright-Fisher models (with non-null drifts) can be obtained by considering the binomial transition probabilities $\text{bin}(N,p_N)$:

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

where

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

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

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

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

For instance, taking $p_N(x)=(1-\pi_{2,N})x+\pi_{1,N}(1-x)$ where $(\pi_{1,N},\pi_{2,N})$ are small (N-dependent) mutation probabilities from A_1 to A_2 (respectively A_2 to A_1). Assuming $(N\cdot\pi_{1,N},N\cdot\pi_{2,N})\underset{N\to\infty}{\to}(u_1,u_2)$, leads after scaling to the drift of WF model with positive mutations rates (u_1,u_2).

Taking

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

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

$$dx_t=\sigma x_t(1-x_t)\,dt+\sqrt{x_t(1-x_t)}dw_t \tag{20}$$

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

$$dy_\theta=sy_\theta(1-y_\theta)\,d\theta+\sqrt{\frac{1}{N}y_\theta(1-y_\theta)}dw_\theta, \tag{21}$$

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

The WF diffusion with selection (20) has two absorbing barriers. It tends to drift to $\circ = 1$ (respectively $\circ = 0$) if allele A_1 is selectively advantageous over A_2 : $\sigma_1 > \sigma_2$ (respectively $\sigma_1 < \sigma_2$) in the following sense: if $\sigma > 0$ (respectively < 0), the fixation probability at $\circ = 1$, which is [23]

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

increases (decreases) with σ taking larger (smaller) values. Putting $x = 1/N$, the fixation probability at 1of an allele A_1 mutant is of order: $2\sigma/N$, see [23].

4. The Wright-Fisher-Karlin model with randomized fitness

We wish now to study the WF model with selection in random environment, using the reminders of the latter two Sections. In other words, the main purpose of this Section is to study the effect of disorder on the WF model with selection just alluded to in (20). Such models were first introduced in [18] and [21]. We start with the construction of [18] before switching to the related one of [21] in the next Section.

4.1. **Karlin model: The small population case.** The introduction of disorder is the simplest possible: One replaces the constant selection intensities $(s_{1,N}, s_{2,N})$ at each generation r by the random iid sequence $\left(s_{1,N}^{(r)}, s_{2,N}^{(r)}\right)_{r\geq 1}$. We assume next that the following conditions (C) hold

$$N \cdot \mathbf{E}\left(s_{i,N}\right) \underset{N\to\infty}{\to} \sigma_i > 0, i = 1, 2$$

$$N \cdot \mathbf{E}\left(s_{i,N}^2\right) \underset{N\to\infty}{\to} \mu_i > 0, i = 1, 2$$

$$N \cdot \mathbf{E}\left(s_{1,N} s_{2,N}\right) \underset{N\to\infty}{\to} \mu_{1,2}.$$

and that all moment terms higher than second degree are of smaller order than $1/N$.

A straightforward computation shows (see [18], [5] Subsection 7.2) that the diffusion approximation in terms of $x_t := N_{[Nt]}(n)/N$, $t \in \mathbf{R}_+$ is of the type (1), now with volatility

$$g(x) = \sqrt{x(1-x) + \rho x^2 (1-x)^2} \tag{22}$$

and drift

$$f(x) = x(1-x)[\eta - \rho x]. \tag{23}$$

Again, time is measured in units of N. Under the conditions (C), the parameters appearing in (22), (23) are

$$\begin{aligned} \eta &= \sigma_1 - \sigma_2 + \mu_2 - \mu_{1,2} = \lim_{N\to\infty} N\mathbf{E}\left(\left(1 - s_{2,N}\right)\left(s_{1,N} - s_{2,N}\right)\right) \\ \rho &= \mu_1 + \mu_2 - 2\mu_{1,2} = \lim_{N\to\infty} N\mathbf{E}\left(\left(s_{1,N} - s_{2,N}\right)^2\right) > 0. \end{aligned}$$

The drift may also be written as:

$$f(x) = x(1-x)\left[\gamma + \rho\left(\frac{1}{2} - x\right)\right] \tag{24}$$

where $\gamma = \gamma_1 - \gamma_2$, with $\gamma_i = \sigma_i - \mu_i/2$, $i = 1, 2$. It is the sum of two contributions, one involving γ, the other one ρ. The latter one introduces a stabilizing drift towards $1/2$.

The local variance term $g^2(x)$ is now the sum of two contributions respectively, one due to binomial sampling and the other the within generation selection variance. When ρ is not large compared to 1 (the small population size case) both terms contribute equally likely and we shall first deal with this case.

The selective advantage of allele A_1 over allele A_2 is now given by $\gamma_1 > \gamma_2$. Because $\gamma_i = \sigma_i - \mu_i/2$, each involves the second-order moment of the $s_{i,N}$, not only the means σ_i.

We note that if $\eta \leq 0$ ($\gamma \leq -\rho/2$), the drift term $f(x)$ is < 0 when $x \in (0,1)$, whereas when $\eta \geq \rho$ ($\gamma \geq \rho/2$), the drift term $f(x)$ is > 0. In the former case, 0 is a stable equilibrium of f whereas in the latter case, 1 is stable. When $\rho > \eta > 0$, ($|\gamma| < \rho/2$), the drift term vanishes inside the interval $(0,1)$ at some point $\widehat{x} = \eta/\rho = \gamma/\rho + 1/2$ corresponding to a stable equilibrium point of f. In particular $\widehat{x} = 1/2$ when $\gamma = 0$, $\widehat{x} \in (1/2, 1]$ (respectively $\widehat{x} \in [0, 1/2)$) when $\gamma \in (0, \rho/2]$ ($\gamma \in [-\rho/2, 0)$).

The key parameters (γ, ρ) appearing in the new diffusion model (1), (22), (24), arise in the following setting. Consider the discrete-time updating of the allele A_1 frequency under fluctuating selection:

$$x_{r+1} = p_N(x_r) = \frac{\left(1 + s_{1,N}^{(r+1)}\right) x_r}{1 + s_{1,N}^{(r+1)} x_r + (1 - x_r)\, s_{2,N}^{(r+1)}}.$$

Here $\left(s_{1,N}^{(r)}, s_{2,N}^{(r)}\right)_{r \geq 1}$ is the random iid sequence introduced earlier. Assume x_0 is fixed (non-random). Let $y_r = x_r/(1 - x_r)$. Then

$$y_{r+1} = y_r \frac{1 + s_{1,N}^{(r+1)}}{1 + s_{2,N}^{(r+1)}}$$

and with $z_r = \log y_r$,

$$z_r = z_0 + \sum_{t=1}^{r} \log\left(\frac{1 + s_{1,N}^{(t)}}{1 + s_{2,N}^{(t)}}\right).$$

Under the assumptions (C) on $(s_{1,N}, s_{2,N})$, we have

$$\begin{aligned} \mathbf{E}(z_r) &= z_0 + \sum_{t=1}^{r} \mathbf{E} \log\left(\frac{1 + s_{1,N}^{(t)}}{1 + s_{2,N}^{(t)}}\right) = z_0 + r\mathbf{E} \log \frac{1 + s_{1,N}}{1 + s_{2,N}} \\ &= z_0 + \frac{r}{N}\left(\gamma + o\left(\frac{1}{N}\right)\right). \end{aligned}$$

Therefore,

$$\frac{1}{t}\mathbf{E}\left(z_{[Nt]}\right) \underset{t\to\infty}{\to} \gamma, \tag{25}$$

and γ interprets as an asymptotic Malthus mean growth parameter for y_r (and therefore also for x_r when x_r is rare). Similarly,

$$\sigma^2\left(z_r\right) = r\sigma^2\left(\log\frac{1+s_{1,N}}{1+s_{2,N}}\right) = \frac{r}{N}\left(\rho + o\left(\frac{1}{N}\right)\right).$$

Therefore,

$$\frac{1}{t}\sigma^2\left(z_{[Nt]}\right) \underset{t\to\infty}{\to} \rho. \tag{26}$$

Additive functionals. We now study some aspects of the diffusion model in random environment defined by (22), (23). We have

$$-\frac{2f\left(x\right)}{g^2\left(x\right)} = -2\frac{\eta-\rho x}{1+\rho x\left(1-x\right)} = \frac{\kappa_1}{x-r_1} + \frac{\kappa_2}{r_2-x}$$

where the roots $r_i = \frac{1\mp\sqrt{1+4/\rho}}{2}$, $i=1,2$, satisfy $r_1<0<1<r_2$ and $r_1+r_2=1$. We shall let $r=\sqrt{1+4/\rho}>1$ so that $r_i=\frac{1\mp r}{2}$. Thus

$$\kappa_1 = -1-\frac{2\gamma}{\rho r},\ \kappa_2 = 1-\frac{2\gamma}{\rho r}.$$

It is possible to define the normalized scale function

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

satisfying $\varphi\left(0\right)=0$ and $\varphi\left(1\right)=1$. With $Z=\int_0^1\left(y-r_1\right)^{-1-\frac{2\gamma}{\rho r}}\left(1-y-r_1\right)^{-1+\frac{2\gamma}{\rho r}}dy<\infty$, we get

$$\varphi\left(x\right) = \frac{1}{Z}\int_0^x\left(y-r_1\right)^{-1-\frac{2\gamma}{\rho r}}\left(1-y-r_1\right)^{-1+\frac{2\gamma}{\rho r}}dy \tag{27}$$

with $\varphi\left(1/2\right)\neq 1/2$. The speed measure density reads

$$m\left(x\right) = \frac{1}{\left(g^2\varphi'\right)\left(x\right)} = \frac{1}{Z}\frac{\left(x-r_1\right)^{\frac{2\gamma}{\rho r}}\left(1-x-r_1\right)^{-\frac{2\gamma}{\rho r}}}{\rho x\left(1-x\right)}. \tag{28}$$

Clearly with $\circ=$ either 0 or 1

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

showing that both boundaries are exit. As a result, $\varphi\left(x\right)=\mathbf{P}\left(x_{\tau_x}=1\right)=\mathbf{P}\left(\tau_{x,1}<\tau_{x,0}\right)$ and the process with small population size is transient.

With φ and $g^2\varphi'$ given by (27) and (28), the Green function can be explicitly computed from (8). The expected time to fixation is therefore

$$\mathbf{E}\left(\tau_x\right) = 2\varphi\left(x\right)\int_x^1\frac{\left(1-\varphi\left(y\right)\right)}{\left(g^2\varphi'\right)\left(y\right)}dy + 2\left(1-\varphi\left(x\right)\right)\int_0^x\frac{\varphi\left(y\right)}{\left(g^2\varphi'\right)\left(y\right)}dy. \tag{29}$$

Remark. Consider the random time change $t \to \theta_t$ with inverse: $\theta \to t_\theta$ defined by $\theta_{t_\theta} = \theta$ and

$$\theta = \int_0^{t_\theta} g^2(x_s)\, ds = \int_0^{t_\theta} \left[x_s(1-x_s) + \rho x_s^2 (1-x_s)^2 \right] ds.$$

The novel diffusion $(u_\theta := x_{t_\theta}; \theta \geq 0)$ is easily checked to satisfy the Langevin SDE on $[0,1]$

$$du_\theta = \frac{\eta - \rho u_\theta}{1 + \rho u_\theta (1-u_\theta)} d\theta + dw_\theta = -\frac{1}{2} \partial_u U(u_\theta)\, d\theta + dw_\theta,$$

with additive noise. The invariant measure is therefore $e^{-U(u)}$ where

$$U(u) = \log \left[(u-r_1)^{-1-\frac{2\gamma}{\rho r}} (1-u-r_1)^{-1+\frac{2\gamma}{\rho r}} \right]$$

if $u \in [0,1] \subset (r_1, r_2)$. $\diamond$

The symmetric case. suppose $s_{1,N} \stackrel{d}{=} s_{2,N}$ (equality in distribution). Then, $\sigma_1 = \sigma_2$, $\mu_1 = \mu_2$ and

$$\eta = \mu_2 - \mu_{1,2} \text{ and } \rho = 2\left(\mu_2 - \mu_{1,2}\right).$$

Thus $\gamma = 0$ and this particular model consists in the neutral WF model in random environment with drift and local variance

$$f(x) = \rho x(1-x)\left(\frac{1}{2} - x\right);\ g^2(x) = x(1-x) + \rho x^2 (1-x)^2$$

involving only ρ. In this case, $m(x) = \frac{r}{2\rho \log\left[\frac{1-r_1}{-r_1}\right]} \frac{1}{x(1-x)}$ and

$$\mathbf{P}(\tau_{x,1} < \tau_{x,0}) = \varphi(x) = \frac{\log\left[\frac{x-r_1}{-r_1}\frac{1-r_1}{1-x-r_1}\right]}{2\log\left[\frac{1-r_1}{-r_1}\right]} = \frac{1}{2}\left(1 + \frac{\log\left[\frac{x-r_1}{1-x-r_1}\right]}{\log\left[\frac{1-r_1}{-r_1}\right]}\right) \tag{30}$$

with $\varphi(1/2) = 1/2$ (See (8) in [17]). As $\rho \to \infty$, $r_1 \to 0^-$ and $r_2 \to 1^+$ with $\varphi(x) \to 1/2$.

The expected time to fixation is

$$\mathbf{E}(\tau_x) = 2\varphi(x) \int_x^1 \frac{(1-\varphi(y))}{(g^2\varphi')(y)} dy + 2(1-\varphi(x)) \int_0^x \frac{\varphi(y)}{(g^2\varphi')(y)} dy$$

which, thanks to the symmetries of φ, takes the simple form (See [18], [5] and [14])

$$\mathbf{E}(\tau_x) = 2\int_0^x \frac{\log((1-y)/y)}{1+\rho y(1-y)} dy \tag{31}$$

Whatever the value of x, this quantity decreases with ρ, consistently with the fact that fluctuations in differential selection intensities tend to decrease the expected fixation time (despite the presence of a competing drift toward 1/2). Note that if $\rho \to 0$, $\mathbf{E}(\tau_x)$ gets close to $-2(x\log x + (1-x)\log(1-x))$ which is the expected fixation time of the usual neutral WF model (19). This can be understood as follows; recall $\rho = \lim N\mathbf{E}\left((s_{1,N} - s_{2,N})^2\right)$. If $\rho \to 0$, then $(s_{1,N} - s_{2,N}) \stackrel{d}{\to} \delta_0$ and therefore $s_{1,N} = s_{2,N}$ almost surely. Thus, γ itself tends to 0 simultaneously and

the model with random selection boils down to the neutral WF model (19) in the limit.

4.2. **The large population case $\rho \gg 1$.** In this case, we are led to the study of the diffusion model (1) with volatility and drift

$$g(x) = \sqrt{\rho} x(1-x) \ ; \ f(x) = x(1-x)\left[\gamma + \rho\left(\frac{1}{2} - x\right)\right]. \tag{32}$$

Here we simply dropped the binomial sampling contribution to the variance term $g^2(x)$ in (22) because it is small under the large population case assumption. Consider the change of variable $y_t = \int_0^{x_t} \frac{dx}{x(1-x)} = \log\left(\frac{x_t}{1-x_t}\right)$. Applying Itô calculus

$$dy_t = \gamma dt + \sqrt{\rho} dw_t, \tag{33}$$

with a Gaussian transition kernel. Using this fact, we easily conclude that the solution to the KF equation for x_t defined by (32) is explicitly given by

$$p(x;t,y) = \frac{1}{\sqrt{2\pi\rho t}} \frac{1}{y(1-y)} e^{-\frac{1}{2\rho t}\left(\log\left(\frac{y(1-x)}{(1-y)x}\right) - \gamma t\right)^2}. \tag{34}$$

If $\gamma > 0$ (< 0), the mass of the law of x_t accumulates near $y = 1$ ($y = 0$). When $\gamma = 0$, the law of x_t forms two symmetric peaks about both $y = 1$ and $y = 0$ as time increases, but without reaching the boundaries since $p(x;t,y)$ vanishes at both $y = 1$ and $y = 0$.

It turns out that this can be understood by noting that both boundaries are natural, therefore inaccessible. Using this solution, it can be checked that, whatever $\varepsilon > 0$

$$\begin{aligned}
\mathbf{P}(x_t \in (1-\varepsilon, 1) \mid x_0 = x) &\underset{t\to\infty}{\to} 1 \text{ if } \gamma > 0 \\
\mathbf{P}(x_t \in (0, \varepsilon) \mid x_0 = x) &\underset{t\to\infty}{\to} 1 \text{ if } \gamma < 0 \\
\mathbf{P}(x_t \in (1-\varepsilon, 1) \mid x_0 = x) &\underset{t\to\infty}{\to} 1/2 \text{ if } \gamma = 0 \\
\mathbf{P}(x_t \in (0, \varepsilon) \mid x_0 = x) &\underset{t\to\infty}{\to} 1/2 \text{ if } \gamma = 0
\end{aligned}$$

expressing that, at the boundaries, quasi-fixation (or quasi-extinction) occurs. Note that the limits do not depend on the initial condition x. We conclude that under the conditions of randomly varying selection, quasi-fixation of allele A_1 possessing the selective advantage $\gamma_1 > \gamma_2$ over A_2 ($\gamma > 0$) will occur with probability 1, regardless what its initial frequency is and no matter on how large the fluctuations in selection intensities really are. In this case, the density function $p(x;t,y)$ increasingly concentrates near the boundary $\circ = 1$ which is stochastically locally stable in the sense of [19]. The boundaries, either $\{0,1\}$ or both are eventually attained although not in finite time. When there is no selective advantage ($\gamma = 0$), quasi-absorption at both endpoints of I occurs equally likely, whatever the initial condition.

Let us now check that the boundaries are natural (and therefore that $-G$ and $-G^*$ corresponding to the Karlin diffusion no longer have a discrete spectrum). We have

$$-\frac{2f(x)}{g^2(x)} = -2\frac{\eta - \rho x}{\rho x(1-x)} = \frac{\kappa_1}{x} + \frac{\kappa_2}{1-x}$$

where

$$\kappa_1 = -1 - \frac{2\gamma}{\rho}, \ \kappa_2 = 1 - \frac{2\gamma}{\rho}.$$

Thus

$$\varphi'(y) = e^{-2\int^y \frac{f(z)}{g^2(z)}dz} = y^{-\left(1+\frac{2\gamma}{\rho}\right)}(1-y)^{-\left(1-\frac{2\gamma}{\rho}\right)} \tag{35}$$

and there is no way to define a normalized scale function satisfying $\varphi(0) = 0$ and $\varphi(1) = 1$. The speed measure density reads

$$m(x) = \frac{1}{(g^2\varphi')(x)} = \frac{1}{\rho}x^{\frac{2\gamma}{\rho}-1}(1-x)^{-\frac{2\gamma}{\rho}-1}. \tag{36}$$

When $\gamma > 0$, m is not integrable at $\circ = 1$ and $\varphi'(y)\int^y m(x)\,dx \underset{y\to 1}{\sim} (1-y)^{-1}$ is not integrable near $y = 1$. Thus $\{1\}$ is a natural boundary. However, m is integrable at $\circ = 0$, but φ' is not integrable and $m\varphi$ is not integrable there showing that $\{0\}$ is also a natural boundary. Similar conclusions may be drawn when $\gamma < 0$: both boundaries are natural. When $\gamma = 0$, $m(x)\,dx$ is the speed measure of the neutral WF diffusion which is not integrable near either $x = 0$ or $x = 1$. Moreover, $\varphi'(y)\int^y m(x)\,dx \underset{y\to 0}{\sim} y^{-1}\log y$ ($\underset{y\to 1}{\sim} -(1-y)^{-1}\log(1-y)$) is not integrable neither at $y = 0$ (nor at $y = 1$). Thus both boundaries are again natural and therefore inaccessible and $\tau_x = \tau_{x,0} \wedge \tau_{x,1} = \infty$ with probability 1.

The symmetric (neutral) case. Assume now $\gamma = 0$. Then the realizations of x_t oscillate back and forth between the boundaries, infinitely often, so that there is a substantial amount of time spent in their neighborhood. The process is null-recurrent. The diffusion model (32) may be recast as:

$$dx_t = \rho x_t(1-x_t)\left(\frac{1}{2} - x_t\right)dt + \sqrt{\rho}x_t(1-x_t)\,dw_t \tag{37}$$

with a stabilizing drift toward 1/2.

Let $\varepsilon > 0$ be a small parameter. Let $x \in I_\varepsilon$. Because the extreme boundaries $\{0, 1\}$ are inaccessible, we shall work (with the tools developed in Section 2) on I_ε rather than on I.

Let $\tau_{x,I_\varepsilon} = \tau_{x,\varepsilon} \wedge \tau_{x,1-\varepsilon}$ be the first exit time of I_ε. We wish to estimate the probability $\mathbf{P}(\tau_{x,1-\varepsilon} < \tau_{x,\varepsilon})$ as $\varepsilon \to 0$, together with $\mathbf{E}(\tau_{x,I_\varepsilon})$.

The scale function of (37) is $\varphi(x) = \log\left(\frac{x}{1-x}\right)$. Defining a normalized scale function $\varphi_\varepsilon(x) = \frac{\varphi(x)-\varphi(\varepsilon)}{\varphi(1-\varepsilon)-\varphi(\varepsilon)}$ satisfying $\varphi_\varepsilon(\varepsilon) = 0$ and $\varphi_\varepsilon(1-\varepsilon) = 1$, we easily get

$$\mathbf{P}(\tau_{x,1-\varepsilon} < \tau_{x,\varepsilon}) = \varphi_\varepsilon(x) = \frac{1}{2}\left(1 - \frac{\log\left(\frac{x}{1-x}\right)}{\log\left(\frac{\varepsilon}{1-\varepsilon}\right)}\right). \tag{38}$$

We conclude that, independently of ρ:

If $x < \frac{1}{2}$, $\mathbf{P}(\tau_{x,1-\varepsilon} < \tau_{x,\varepsilon}) \underset{\varepsilon\to 0}{\sim} \frac{1}{2}\left(1 - \frac{\log\left(\frac{1-x}{x}\right)}{-\log\varepsilon}\right)$ which is slightly less than 1/2 with a correcting term of order $-1/\log\varepsilon$. If $\varepsilon = 1/(2N)$ and $x = 1/N$, the quasi-fixation probability at $1-\varepsilon$ of a mutant is of order:

$$\frac{1}{2}\left(1 - \frac{\log\left(\frac{1}{N}\right)}{\log\left(\frac{2}{N}\right)}\right) \sim \frac{1}{\log N}. \tag{39}$$

If $x > \frac{1}{2}$, $\mathbf{P}(\tau_{x,1-\varepsilon} < \tau_{x,\varepsilon}) \underset{\varepsilon\to 0}{\sim} \frac{1}{2}\left(1 + \frac{\log\left(\frac{x}{1-x}\right)}{-\log\varepsilon}\right)$ which is slightly greater than $1/2$.

Using the Green function

$$\begin{aligned}
\mathfrak{g}_{(\varepsilon)}(x,y) &= 2\frac{(\varphi_\varepsilon(x)-\varphi_\varepsilon(\varepsilon))(\varphi_\varepsilon(1-\varepsilon)-\varphi_\varepsilon(y))}{(g^2\varphi'_\varepsilon)(y)} \text{ if } \varepsilon < x < y < 1-\varepsilon \\
\mathfrak{g}_{(\varepsilon)}(x,y) &= 2\frac{(\varphi_\varepsilon(1-\varepsilon)-\varphi_\varepsilon(x))(\varphi_\varepsilon(y)-\varphi_\varepsilon(\varepsilon))}{(g^2\varphi'_\varepsilon)(y)} \text{ if } 1-\varepsilon > x > y > \varepsilon
\end{aligned}$$

the expected exit time of I_ε is obtained as

$$\begin{aligned}
\mathbf{E}(\tau_{x,I_\varepsilon}) &= \frac{2}{\rho}\int_\varepsilon^x \frac{\log((1-y)/y)}{y(1-y)}dy \\
&= \frac{1}{\rho}\left(\log(((1-\varepsilon)/\varepsilon))^2 - (\log((1-x)/x))^2\right) \\
&\underset{\varepsilon\to 0}{\sim} \frac{1}{\rho}[\log(\varepsilon)]^2.
\end{aligned}$$

Thus this expected time diverges like $[\log(\varepsilon)]^2$, somehow quantifying how inaccessible the natural boundaries are. Note that $\mathbf{E}(\tau_{x,I_\varepsilon})$ is a decreasing function of ρ.

• Empirical average measure of heterozygosity. As noted above, the realizations of x_t oscillate back and forth between the boundaries, infinitely often. We therefore expect that the empirical average of heterozygosity should be close to 0, as result of x_t spending a substantial amount of time in the neighborhood of the boundaries. Let us quantify this intuition.

Let a and b be two functions which are integrable with respect to the speed measure $m(x)\,dx$, with, from (36) with $\gamma = 0$

$$m(x) = \frac{1}{\rho x(1-x)}.$$

Then, by the ergodic Chacon-Ornstein ratio theorem for null-recurrent processes

$$\frac{\int_0^t a(x_s)\,ds}{\int_0^t b(x_s)\,ds} = \frac{t^{-1}\int_0^t a(x_s)\,ds}{t^{-1}\int_0^t b(x_s)\,ds} \underset{t\to\infty}{\to} \frac{\int_0^1 a(x)\,m(x)\,dx}{\int_0^1 b(x)\,m(x)\,dx}.$$

Let $\varepsilon > 0$ be small. Let us choose $a(x) = 2x(1-x)1_{x\in(\varepsilon,1-\varepsilon)}$ and $b(x) = 1_{x\in(\varepsilon,1-\varepsilon)}$, in such a way that the above ratio represents the conditional empirical average of heterozygosity of the process x_t given it keeps remaining within the interval $(\varepsilon, 1-\varepsilon)$. Would the process spend most of the time close to 0 and 1 where the heterozygosity vanishes, one would expect the empirical average of heterozygosity to tend to 0 as $\varepsilon \to 0$. By the ergodic Chacon-Ornstein ratio theorem

$$\frac{t^{-1}\int_0^t 2x_s(1-x_s)1_{x_s\in(\varepsilon,1-\varepsilon)}ds}{t^{-1}\int_0^t 1_{x_s\in(\varepsilon,1-\varepsilon)}ds} \underset{t\to\infty}{\to} \frac{2\int_\varepsilon^{1-\varepsilon}dx}{\int_\varepsilon^{1-\varepsilon}\frac{1}{x(1-x)}dx} \underset{\varepsilon\to 0}{\sim} \frac{1}{-\log(\varepsilon)} \tag{40}$$

which indeed tends to 0 when $\varepsilon \to 0$, independently of ρ.

• If the particle spends a substantial amount of time near the boundaries, this is also because the time to move from ε to $1-\varepsilon$ is large. Let us quantify this point.

Consider (34) with $x<y$ and $\gamma=0$. Let

$$\mathfrak{g}_\lambda(x,y)=\int_0^\infty e^{-\lambda t}p(x;t,y)\,dt$$

be the Green potential function of the neutral Kimura model.

With $\delta_2:=\frac{1}{\rho}\left[\log\left(\frac{y(1-x)}{x(1-y)}\right)\right]^2$, we have

$$\begin{aligned}\mathfrak{g}_\lambda(x,y)&=\frac{1}{\sqrt{2\pi\rho}}\frac{1}{y(1-y)}\int_0^\infty t^{-\frac{1}{2}}e^{-\frac{1}{2}\left(2\lambda t+\frac{\delta_2}{t}\right)}dt\\&=\frac{2}{\sqrt{2\pi\rho}}\frac{1}{y(1-y)}K_{\frac{1}{2}}\left(\sqrt{2\lambda\delta_2}\right)\left(\frac{\delta_2}{2\lambda}\right)^{1/4}\end{aligned}$$

where $K_{\frac{1}{2}}(x)=\sqrt{\frac{\pi}{2x}}e^{-x}$ is a particular modified Bessel function of the third kind

$$K_\alpha(x)=\frac{1}{2}\int_0^\infty t^{\alpha-1}e^{-x(t+1/t)}dt,$$

with index $\alpha=1/2$. Hence

$$\mathfrak{g}_\lambda(x,y)=\frac{1}{y(1-y)}\frac{1}{(2\rho\lambda)^{1/2}}e^{-\sqrt{2\lambda\delta_2}}.$$

Assume $x=y$ and let us compute $\mathfrak{g}_\lambda(y,y)=\int_0^\infty e^{-\lambda t}p(y;t,y)\,dt$. We get

$$\begin{aligned}\mathfrak{g}_\lambda(y,y)&=\frac{1}{\sqrt{2\pi\rho}}\frac{1}{y(1-y)}\int_0^\infty t^{-\frac{1}{2}}e^{-\lambda t}dt\\&=\frac{1}{\sqrt{2\rho\lambda}}\frac{1}{y(1-y)}.\end{aligned}$$

Note that $\mathfrak{g}_0(x,y)=\mathfrak{g}_0(y,y)=\infty$ (the process under study is recurrent): the expected local time at y starting from x is ∞. This suggests that for all $x<y$ ($x>y$), the expected value of $\tau_{x,y}$ should be $\int_0^y\mathfrak{g}_0(x,z)\,dz=\infty$ (respectively $\int_y^1\mathfrak{g}_0(x,z)\,dz=\infty$).

Indeed, with $\tau_{x,y}$ the first time x_s hits y starting from x

$$\mathbf{E}\left(e^{-\lambda\tau_{x,y}}\right)=\frac{\mathfrak{g}_\lambda(x,y)}{\mathfrak{g}_\lambda(y,y)}=e^{-\sqrt{2\delta_2\lambda}}\tag{41}$$

showing that $\tau_{x,y}\overset{d}{=}bS_{1/2}$ where $S_{1/2}$ is a standard random variable with stable law of index $1/2$ and b is the scale parameter $b=2\delta_2=\frac{2}{\rho}\left[\log\left(\frac{y(1-x)}{x(1-y)}\right)\right]^2$. The probability of the event $\tau_{x,y}<\infty$ is 1. When $x=\varepsilon$ and $y=1-\varepsilon$, the scale parameter is

$$b=\frac{2^3}{\rho}\left[\log\left(\frac{1-\varepsilon}{\varepsilon}\right)\right]^2\underset{\varepsilon\to0}{\sim}\frac{2^3}{\rho}\left[\log(1/\varepsilon)\right]^2\to\infty.$$

It takes a long time to move from ε to $1-\varepsilon$ and back, but this move occurs with probability 1. We have

$$\frac{\rho}{2^3\left[\log(1/\varepsilon)\right]^2}\tau_{\varepsilon,1-\varepsilon}\underset{\varepsilon\to0}{\overset{d}{\to}}S_{1/2}.$$

Proceeding similarly

$$\frac{\rho}{2^5\varepsilon^2}\tau_{\frac{1}{2}\pm\varepsilon,\frac{1}{2}}\underset{\varepsilon\to0}{\overset{d}{\to}}S_{1/2}.$$

quantifying how small the first return time to $x = 1/2$ is.

5. Karlin versus Kimura models

Consider the Karlin diffusion model (1), in the sense of Itô, therefore with volatility and drift (32)

$$g(x) = \sqrt{\rho} x(1-x) \ ; \ f(x) = x(1-x)\left[\gamma + \rho\left(\frac{1}{2} - x\right)\right].$$

It has the equivalent Stratonovitch form

$$dx_t = \left[f(x_t) - \frac{1}{2} g g'(x_t)\right] dt + g(x_t) \circ dw_t, x_0 = x \tag{42}$$

where $\int_0^t g(x_s) \circ dw_s$ is now to be understood as a Stratonovitch integral [33]. The equivalent Stratonovitch form of (32) therefore is (see [36])

$$dx_t = \gamma x_t (1-x_t)\, dt + \sqrt{\rho} x_t (1-x_t) \circ dw_t, x_0 = x. \tag{43}$$

Kimura considered the related model ([21], [22], [4])

$$dx_t = \gamma x_t (1-x_t)\, dt + \sqrt{\rho} x_t (1-x_t)\, dw_t, x_0 = x. \tag{44}$$

where the latter diffusion is now understood in the classical sense of Itô. This model was also viewed by Kimura as a WF model under fluctuating selection. It can be interpreted as follows (See also [20], page 361-362). Consider the continuous-time deterministic evolution equation for the gene frequency of allele A_1, driven by fitness σ:

$$dx_t = \sigma x_t (1-x_t)\, dt.$$

Assume that the selection differential σdt is random and that it can be modelled by some random Gaussian differential $d\widetilde{w}_t$ satisfying $\mathbf{E}(d\widetilde{w}_t) = \gamma dt$ and $\sigma^2(d\widetilde{w}_t) = \rho dt$. Then we obtain (44).

Clearly however, the Kimura model (44) is of a very different nature than its Karlin counterpart defined in (32) and we wish to analyze this discrepancy.

Remark. Historically, the first model to describe WF diffusion in uncorrelated random environment is Kimura's model in its Itô's version (44). Looking at (44) in the sense of Stratonovitch as in (43) and taking its equivalent Itô version, brings one back to the Karlin model. The discrepancy between SDEs interpreted in the sense of Itô and Stratonovitch is not new. Similar observations were made in [27] in the context of random population growth models. One could ask which model to use in practise? Although some authors like [29] argue that, for practical purposes, Stratonovitch models like in (43) (and therefore Karlin model) are best-suited for physical-world models, this non-uniqueness problem is not completely and satisfactorily solved, to the best of the author's knowledge. Still, in our context, one argument pleading in favor of Karlin's model is the fact that it arises as a proper scaling limit of the discrete space-time construction briefly sketched at the beginning of Subsection 4.1. $\diamond$

Curiously, the Kimura model is more tricky than the Karlin one. For instance, when dealing with (44), we could first think of applying the change of variable

$$y_t = \int_0^{x_t} \frac{1}{z\,(1-z)} dz = \log\left(\frac{x_t}{1-x_t}\right),$$

following the path already used in the context of Karlin's model. The new process y_t then obeys the Langevin equation

$$dy_t = \left(\gamma + \frac{\rho}{2}\tanh\left(\frac{y_t}{2}\right)\right)dt + \sqrt{\rho}dw_t,\ y_0 = \log\left(\frac{x_0}{1-x_0}\right) \tag{45}$$

with state-independent volatility but with a non-linear drift that will make it difficult to handle.

6. The Kimura model

In this last Section, we deal with the Kimura model itself, applying to it similar techniques than the ones used for the Karlin model and drawing the appropriate conclusions.

6.1. **The symmetric case (Kimura martingale).** We start with the easier case when $\gamma = 0$. The obtained diffusion (44) reduces to the Kimura martingale $dx_t = \sqrt{\rho}x_t\,(1-x_t)\,dw_t$. It is also of a very different nature than (37). As we will show next, it has again two natural inaccessible boundaries; the process is still null-recurrent. For the driftless model (44), using an elegant change of variable bringing the problem into the heat equation, the solution to the associated KFE was obtained by ([21]) [2] as

$$\widetilde{p}\,(x;t,y) = \frac{1}{\sqrt{2\pi\rho t}}\frac{(x\,(1-x))^{1/2}}{(y\,(1-y))^{3/2}}e^{-\left(\frac{\rho t}{8}+\frac{1}{2\rho t}\left[\log\left(\frac{y(1-x)}{x(1-y)}\right)\right]^2\right)}. \tag{46}$$

The density (46) converges more rapidly than its Karlin version (34) to the quasi-absorption states $\{0,1\}$. Its mean and variance satisfy (see [21]):

$$\int_0^1 y\widetilde{p}\,(x;t,y)\,dy = x$$

$$\int_0^1 (y-x)^2\,\widetilde{p}\,(x;t,y)\,dy \underset{t\to\infty}{\to} x\,(1-x)\,,$$

in accordance with the fact

$$\widetilde{p}\,(x;t,y)\,dy \underset{t\to\infty}{\to} (1-x)\,\delta_0 + x\delta_1.$$

Note that this is not absorption because the absorption time is ∞ with probability one, rather it is quasi-absorption. Would the initial condition be chosen random with mean $1/2$, the limiting law would be symmetric with mass $1/2$ at both endpoints (like in the neutral Karlin model with $\gamma = 0$). Based on this solution (46), Kimura [21] and Tuckwell [36] also argued that

$$\mathbf{P}\,(x_t \in (0,\varepsilon) \mid x_0 = x) \underset{t\to\infty}{\to} 1-x$$
$$\mathbf{P}\,(x_t \in (1-\varepsilon,1) \mid x_0 = x) \underset{t\to\infty}{\to} x$$

[2] With $\widetilde{p}$ obeying the corresponding KFE equation $\partial_t\widetilde{p} = \rho/2\partial_y^2\left(y^2\left(1-y^2\right)\widetilde{p}\right)$, putting $\phi = 1/2e^{(\rho t)/8}y^{3/2}\left(1-y\right)^{3/2}\phi$ and $z = \log\left(y/\left(1-y\right)\right)$, ϕ obeys the heat equation $\partial_t\phi = \rho/2\partial_z^2\left(\phi\right)$.

where the limiting quantities now depend on the initial condition. The approach of the limiting measure occurs with the exponential rate $\frac{\rho t}{8}$ in the sense for instance that

$$\mathbf{P}\left(x_t \in (0,\varepsilon) \mid x_0 = x\right) = 1 - x + O\left(\frac{e^{-\frac{\rho t}{8}}}{\sqrt{2\pi\rho t}}\right).$$

The scale function associated to the Kimura diffusion is the identity $\varphi(x) = x$. Its speed measure density is $m(x) = \frac{1}{\rho x^2(1-x)^2}$ which is not integrable at the boundaries. Because $\varphi'(y)\int^y m(x)\,dx$ is not integrable neither at $y = 0$ nor at $y = 1$, both boundaries are natural.

Let $x \in I_\varepsilon$. Let $\tau_{x,I_\varepsilon} = \tau_{x,\varepsilon} \wedge \tau_{x,1-\varepsilon}$ be the first exit time of I_ε. We wish now to estimate the probability $\mathbf{P}\left(\tau_{x,1-\varepsilon} < \tau_{x,\varepsilon}\right)$ as $\varepsilon \to 0$, together with $\mathbf{E}\left(\tau_{x,I_\varepsilon}\right)$, for the Kimura martingale model.

Defining a normalized scale function $\varphi_\varepsilon(x) = \frac{\varphi(x)-\varphi(\varepsilon)}{\varphi(1-\varepsilon)-\varphi(\varepsilon)}$ (with $\varphi(x) = x$), satisfying $\varphi_\varepsilon(\varepsilon) = 0$ and $\varphi_\varepsilon(1-\varepsilon) = 1$, we easily get

$$\mathbf{P}\left(\tau_{x,1-\varepsilon} < \tau_{x,\varepsilon}\right) = \varphi_\varepsilon(x) = \frac{x-\varepsilon}{1-2\varepsilon}, \tag{47}$$

independently of ρ. The result is thus very different from the one displayed in (38) which was close to $1/2$. The origin of this difference is to be searched in the presence of an attracting drift to $1/2$ in the Karlin model (37), which is not present in the driftless Kimura martingale. If $\varepsilon = 1/(2N)$ and $x = 1/N$, the quasi-fixation probability at $1-\varepsilon$ of a mutant is now of order $1/(2N)$, much smaller than in the Karlin case (39).

• Let $\alpha(x) = \mathbf{E}\left(\tau_{x,I_\varepsilon}\right)$ be the expected exit time of I_ε. It solves $-G\alpha(x) = 1$ where $G = \frac{\rho}{2}x^2(1-x)^2\partial_x^2$ and $\alpha(\varepsilon) = \alpha(1-\varepsilon) = 0$. Thus

$$\partial_x^2\alpha(x) = -\frac{2}{\rho x^2(1-x)^2} = -\frac{2}{\rho}\left[\frac{1}{x^2} + \frac{1}{(1-x)^2} + \frac{2}{x} + \frac{2}{1-x}\right].$$

Integrating twice and plugging in the boundary conditions, we find

$$\alpha(x) = \mathbf{E}\left(\tau_{x,I_\varepsilon}\right) = \frac{2}{\rho}\left(h(\varepsilon) - h(x)\right) \tag{48}$$

where

$$h(x) = 2x\log x + 2(1-x)\log(1-x) - \log\left(x(1-x)\right). \tag{49}$$

Thus this expected time diverges like $-\frac{2}{\rho}\log(\varepsilon)$, which is smaller than $\frac{1}{\rho}\left[\log(\varepsilon)\right]^2$ obtained previously for its Karlin counterpart. The Kimura model hits therefore the boundaries of I_ε in a shorter time. Note that $\mathbf{E}\left(\tau_{x,I_\varepsilon}\right)$ is again a decreasing function of ρ.

• The empirical average measure of heterozygosity for the Kimura martingale x_t as in (44) with $\gamma = 0$. Again the realizations of x_t also oscillate back and forth between the boundaries. We therefore also expect that the empirical average of heterozygosity should also be close to 0, as result of x_t spending a substantial

amount of time in the neighborhood of the boundaries. For this process indeed, the speed measure is

$$m(x) = \frac{1}{\rho x^2 (1-x)^2}.$$

By the ergodic Chacon-Ornstein ratio theorem

$$\frac{t^{-1}\int_0^t 2x_s(1-x_s)1_{x_s\in(\varepsilon,1-\varepsilon)}ds}{t^{-1}\int_0^t 1_{x_s\in(\varepsilon,1-\varepsilon)}ds} \underset{t\to\infty}{\to} \frac{2\int_\varepsilon^{1-\varepsilon}\frac{1}{x(1-x)}dx}{\int_\varepsilon^{1-\varepsilon}\frac{1}{x^2(1-x)^2}dx} \underset{\varepsilon\to 0}{\sim} -2\varepsilon\log\varepsilon \tag{50}$$

which also tends to 0 when $\varepsilon \to 0$, but much faster than in (40). Therefore, the Kimura martingale spends much more time close to the boundaries than the Karlin process defined in (36) with $\gamma = 0$, presenting a stabilizing drift.

• The time to move from ε to $1-\varepsilon$. Consider now (46) with $x < y$. Let

$$\mathfrak{g}_\lambda(x,y) = \int_0^\infty e^{-\lambda t}\widetilde{p}(x;t,y)\,dt$$

be the Green function of the driftless Kimura model.

With $\delta_1 := \frac{\rho}{4}$ and $\delta_2 := \frac{1}{\rho}\left[\log\left(\frac{y(1-x)}{x(1-y)}\right)\right]^2$, we have

$$\mathfrak{g}_\lambda(x,y) = \frac{1}{\sqrt{2\pi\rho}}\frac{(x(1-x))^{1/2}}{(y(1-y))^{3/2}}\int_0^\infty t^{-\frac{1}{2}}e^{-\frac{1}{2}\left((2\lambda+\delta_1)t+\frac{\delta_2}{t}\right)}dt$$

$$= \frac{2}{\sqrt{2\pi\rho}}\frac{(x(1-x))^{1/2}}{(y(1-y))^{3/2}}K_{\frac{1}{2}}\left(\sqrt{(2\lambda+\delta_1)\delta_2}\right)\left(\frac{\delta_2}{2\lambda+\delta_1}\right)^{1/4}$$

where $K_{\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2x}}e^{-x}$. Hence

$$\mathfrak{g}_\lambda(x,y) = \frac{1}{\sqrt{\rho}}\frac{(x(1-x))^{1/2}}{(y(1-y))^{3/2}}\frac{1}{(2\lambda+\delta_1)^{1/2}}e^{-\sqrt{(2\lambda+\delta_1)\delta_2}}.$$

Assume $x = y$ and let us compute $\mathfrak{g}_\lambda(y,y) = \int_0^\infty e^{-\lambda t}\widetilde{p}(y;t,y)\,dt$. We get

$$\begin{aligned}\mathfrak{g}_\lambda(y,y) &= \frac{1}{\sqrt{2\pi\rho}}\frac{1}{y(1-y)}\int_0^\infty t^{-\frac{1}{2}}e^{-\frac{1}{2}(2\lambda+\delta_1)t}dt \\ &= \frac{1}{\sqrt{2\pi\rho}}\frac{1}{y(1-y)}\frac{\sqrt{\pi}}{\sqrt{\frac{1}{2}(2\lambda+\delta_1)}} = \frac{1}{\sqrt{\rho}}\frac{1}{y(1-y)}\frac{1}{(2\lambda+\delta_1)^{1/2}}.\end{aligned}$$

Thus

$$\mathbf{E}\left(e^{-\lambda\tau_{x,y}}\right) = \frac{\mathfrak{g}_\lambda(x,y)}{\mathfrak{g}_\lambda(y,y)} = \left(\frac{x(1-x)}{y(1-y)}\right)^{1/2}e^{-\sqrt{(2\lambda+\delta_1)\delta_2}}$$

$$= \left(\frac{x}{y}\right)^{\frac{1}{2}\left(1+\sqrt{1+\frac{8\lambda}{\rho}}\right)}\left(\frac{1-x}{1-y}\right)^{\frac{1}{2}\left(1-\sqrt{1+\frac{8\lambda}{\rho}}\right)} \tag{51}$$

$$= \frac{x}{y}\cdot\left[\left(\frac{x(1-y)}{y(1-x)}\right)^{1/2}\right]^{\sqrt{1+\frac{8\lambda}{\rho}}-1}$$

recalling $\delta_1\delta_2 = \frac{1}{4}\left[\log\left(\frac{y(1-x)}{x(1-y)}\right)\right]^2$. Compare with (41). Thus $\tau_{x,y} < \infty$ with probability $\frac{x}{y} \in (0,1)$ and given $\tau_{x,y} < \infty$, the LST of the law of $\tau_{x,y}$ is the one of an exponentially damped $\frac{1}{2}-$stable law with scale parameter $b = \frac{2}{\rho}\left[\log\left(\frac{x(1-y)}{y(1-x)}\right)\right]^2$ and exponential damping parameter $a = \frac{\rho}{8}$ and finite mean $\frac{1}{2}\sqrt{\frac{b}{a}}$ [3]. If $x = \varepsilon < 1/2$ and $y = 1-\varepsilon$, the probability that a move from ε to $1-\varepsilon$ occurs in finite time is $\pi = \frac{\varepsilon}{1-\varepsilon} \underset{\varepsilon\to 0}{\sim} \varepsilon$. We have

$$\mathbf{E}\left(e^{-\lambda\tau_{\varepsilon,1-\varepsilon}}\right) = \pi \cdot e^{-(-\log\pi)\left(\sqrt{1+\frac{8\lambda}{\rho}}-1\right)}.$$

Whenever the Kimura martingale approaches a boundary closely, its probability not to move back to the opposite boundary increases and approaches 1 accordingly. This opportunity does not exist when dealing with its Karlin counterpart (37). Given the switch from ε to $1-\varepsilon$ occurs, the expected waiting time is seen to be of order

$$\frac{1}{2}\sqrt{\frac{b}{a}} \sim -\frac{4}{\rho}\log(\varepsilon) \underset{\varepsilon\to 0}{\to} \infty,$$

decreasing with ρ.

6.2. **The non-symmetric Kimura model with a drift.** Consider the full Kimura model (44) with $\gamma \neq 0$.

Natural boundaries. Let us first check that the boundaries are natural. We have

$$-\frac{2f(x)}{g^2(x)} = -\frac{2\gamma}{\rho}\frac{1}{x(1-x)} = -\frac{2\gamma}{\rho}\left(\frac{1}{x}+\frac{1}{1-x}\right).$$

Thus

$$\varphi'(y) = e^{-2\int^y \frac{f(z)}{g^2(z)}dz} = y^{-\frac{2\gamma}{\rho}}(1-y)^{\frac{2\gamma}{\rho}}. \tag{52}$$

The speed measure density reads

$$m(x) = \frac{1}{(g^2\varphi')(x)} = \frac{1}{\rho}x^{\frac{2\gamma}{\rho}-2}(1-x)^{-\frac{2\gamma}{\rho}-2}. \tag{53}$$

We have:

$$\varphi'(y)\int^y m(x)\,dx \underset{y\to 0}{\sim} y^{-1} \quad \notin \quad L_1(0) \text{ if } \gamma \neq \rho/2$$

$$\varphi'(y)\int^y m(x)\,dx \underset{y\to 1}{\sim} (1-y)^{-1} \quad \notin \quad L_1(1) \text{ if } \gamma \neq -\rho/2$$

and

$$m \in L_1(0) \Leftrightarrow \gamma > \rho/2$$
$$m \in L_1(1) \Leftrightarrow \gamma < -\rho/2$$

[3] Let $\phi(\lambda) = \exp\left[-\sqrt{b\lambda}\right]$ be the LST of a stable law with scale parameter b and index $1/2$. Then, with $a > 0$ $\phi(\lambda+a)/\phi(a) = \exp\left[-\sqrt{ab}\left(\sqrt{1+\lambda/a}-1\right)\right]$ is the LST of a damped $\frac{1}{2}-$stable law, with scale parameter b and exponential damping parameter a. Such distributions possess a finite mean, namely $\frac{1}{2}\sqrt{\frac{b}{a}}$.

and

$$\begin{aligned}\varphi' &\in L_1(0) \Leftrightarrow \gamma < \rho/2\\ \varphi' &\in L_1(1) \Leftrightarrow \gamma > -\rho/2.\end{aligned}$$

When $\gamma < \rho/2$, $m \notin L_1(0)$ and $\varphi'(y)\int^y m(x)\,dx \notin L_1(0)$. Thus $\{0\}$ is a natural boundary. When $\gamma > \rho/2$, $m \in L_1(0)$, $\varphi' \notin L_1(0)$ and $m\varphi \notin L_1(0)$. Thus $\{0\}$ is again a natural boundary.

When $\gamma = \rho/2$, $m \underset{y\to 0}{\sim} y^{-1} \notin L_1(0)$ and $\varphi'(y)\int^y m(x)\,dx \underset{y\to 0}{\sim} y^{-1}\log y \notin L_1(0)$. Thus $\{0\}$ is a natural boundary.

When $\gamma > -\rho/2$, $m \notin L_1(1)$ and $\varphi'(y)\int^y m(x)\,dx \notin L_1(1)$. Thus $\{1\}$ is a natural boundary. When $\gamma < -\rho/2$, $m \in L_1(1)$, $\varphi' \notin L_1(1)$ and $m\varphi \notin L_1(1)$. Thus $\{1\}$ is again a natural boundary.

When $\gamma = -\rho/2$, we have $m \underset{y\to 1}{\sim} (1-y)^{-1} \notin L_1(0)$ and $\varphi'(y)\int^y m(x)\,dx \underset{y\to 0}{\sim} -(1-y)^{-1}\log(1-y) \notin L_1(0)$. Thus $\{1\}$ is a natural boundary.

We conclude that $\{0,1\}$ are always natural boundaries for the Kimura model with a drift although for different reasons (Feller criteria) when γ passes through the values $-\rho/2$ and $\rho/2$.

When $\gamma \neq 0$, to the best of the author knowledge, there is, up to now, no known solution of the transition probability density of (1) associated to (44), [see [22], page 37]. However, for the Kimura model with a drift, Tuckwell [36] argues that

$$\mathbf{P}(x_t \in (0,\varepsilon) \mid x_0 = x) \underset{t\to\infty}{\to}$$

$$\left(1 \text{ if } \gamma < -\rho/2;\ \tfrac{1-x}{2} \text{ if } \gamma = -\rho/2;\ 1-x \text{ if } \rho/2 > \gamma > -\rho/2 \text{ and } 0 \text{ if } \gamma > \rho/2\right)$$

and

$$\mathbf{P}(x_t \in (1-\varepsilon,1) \mid x_0 = x) \underset{t\to\infty}{\to}$$

$$\left(1 \text{ if } \gamma > \rho/2;\ \tfrac{x}{2} \text{ if } \gamma = \rho/2;\ x \text{ if } \rho/2 > \gamma > -\rho/2 \text{ and } 0 \text{ if } \gamma < -\rho/2\right),$$

suggesting that the limiting law of the Kimura process with drift is

$$\begin{aligned}\delta_1 \text{ if } \gamma &> \rho/2\\ (1-x)\,\delta_0 + x\delta_1 \text{ if } |\gamma| &< \rho/2\\ \frac{1-x}{2}\delta_0 + \frac{1+x}{2}\delta_1 \text{ if } \gamma &= \rho/2\\ \left(1-\frac{x}{2}\right)\delta_0 + \frac{x}{2}\delta_1 \text{ if } \gamma &= -\rho/2\\ \delta_0 \text{ if } \gamma &< -\rho/2.\end{aligned}$$

For the nonneutral Kimura model therefore, there is a non-null probability that an allele gets quasi-fixed (quasi-extinct) even if its selective differential γ is negative (positive), depending on the initial allele frequency. This differential simply needs to be larger (smaller) than $-\rho/2$ (respectively $\rho/2$). Both preliminary arguments suggest that something special should occur when $|\gamma| = \rho/2$.

From Karlin to Kimura model using an appropriate Doob transform.

We shall use here the ideas on Doob-transforms developed in Subsection 2.6. Consider the Karlin model (1) with (32):

$$g(x) = \sqrt{\rho}x(1-x) \ ; \ f(x) = x(1-x)\left[\gamma + \rho\left(\frac{1}{2} - x\right)\right].$$

Let $\alpha(x) = g(x)^{-1/2} = \rho^{-1/4}(x(1-x))^{-1/2}$. With $G = f\partial_x + \frac{1}{2}g^2\partial_x^2$, we have

$$G\alpha = \frac{1}{2}f\frac{g'}{g} - \frac{3}{8}g'^2 + \frac{1}{4}gg''.$$

We wish to study a transformed version of the Karlin diffusion model (32) using the additive functional $\alpha(x)$. By doing so, we move from G to

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

This choice of α is because, using the preliminaries of Section 2, the drift of the transformed process becomes

$$\widetilde{f}(x) = f(x) + \frac{\alpha'(x)}{\alpha(x)}g^2(x) = f(x) - \frac{1}{2}gg'(x),$$

as required to switch from the Karlin model (32) to the Kimura one. Indeed, the drift of the transformed process is

$$\widetilde{f}(x) = x(1-x)\left[\gamma + \rho\left(\frac{1}{2} - x\right)\right] - \frac{1}{2}\rho x(1-x)(1-2x) = \gamma x(1-x),$$

which is the one appearing in (44) for the Kimura model. In this transformation of paths process, we additionally get the affine creating-annihilating paths rate function

$$\lambda(x) = \frac{G\alpha}{\alpha}(x) = -\frac{1}{2}\left(\gamma - \frac{\rho}{4}\right) + \gamma x. \tag{54}$$

The birth and death rate λ is bounded above. It may be put into the canonical form $\lambda(x) = \lambda_*(\mu(x) - 1)$ where $\lambda_* = \frac{\rho}{8} + \frac{|\gamma|}{2} > 0$ and

$$\mu(x) = 2 - \frac{2|\gamma|}{|\gamma| + \frac{\rho}{4}}\left(1 - x^{\mathbf{1}(\gamma \geq 0)}(1-x)^{\mathbf{1}(\gamma<0)}\right)$$

which is > 0 for all $x \in (0,1)$. Therefore, the transformed process accounts for a branching diffusion (BD) where a diffusing Kimura mother particle (started in x) lives a random exponential time with constant rate λ_*. When the mother particle dies, it gives birth to a spatially dependent random number $M(x)$ of particles (with mean $\mu(x)$). If $M(x) \neq 0$, $M(x)$ independent daughter particles are started afresh where their mother particle died; they move along a Kimura diffusion and reproduce, independently and so on for the subsequent generation particles. If $M(x) = 0$, the process stops in the first generation.

We note that $\mu(x) \geq 1$ for all $x \in (0,1)$ if and only if $|\gamma| \leq \rho/4$ and that $\mu(x)$ is largest equal to 2 when $\gamma = 0$. So we actually get a BD with binary scission whose random offspring number satisfies

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

with $p_2(x) \geq p_0(x)$ for all x if and only if $|\gamma| \leq \rho/4$.

While modifying the Karlin model x_t using $\alpha(x) = g(x)^{-1/2}$, the law $p(x;t,y)$ of x_t is transformed into

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

and so is explicitly known because so is p from (34). The transformed process is therefore a BD where individual particles diffuse according to the Kimura model and branch at rate $\lambda(y) = \lambda_*(p_2(y) - p_0(y))$. However, this BD model does not fit into the framework of [1] and [2] for positively regular branching diffusions, leading to global population growth.

Suppose it does and let us see what should be expected. Integrating over y, $\overline{\rho}_t(x) := \int_{\overset{\circ}{I}} \overline{p}(x;t,y)\, dy$ would be the global expected number $\mathbf{E}(N_t(x))$ of Kimura particles alive at time t in $\overset{\circ}{I}$, and we would have

$$\partial_t \overline{\rho}_t(x) = \overline{G}(\overline{\rho}_t(x)) = \lambda(x)\,\overline{\rho}_t(x) + \widetilde{G}(\overline{\rho}_t(x)),\ \overline{\rho}_0(x) = \mathbf{1}_{(0,1)}(x).$$

Using (34) and performing the change of variable $z = \log\left(\frac{y}{1-y}\right)$, we are led to evaluate integrals with respect to the Gaussian kernel. We easily find [4]

$$\overline{\rho}_t(x) = xe^{t(\gamma/2+\rho/8)} + (1-x)\, e^{-t(\gamma/2-\rho/8)},$$

with

$$-\frac{1}{t}\log \overline{\rho}_t(x) \underset{t\to\infty}{\to} \lambda_1 := -\left(\frac{|\gamma|}{2} + \frac{\rho}{8}\right) = -\lambda_* < 0.$$

This suggests that $-\lambda_1$ should be the global Malthus exponential rate of growth of the global expected number of particles within the whole system. Defining then the conditional probability presence density $\overline{q}(x;t,y) := \overline{p}(x;t,y)/\overline{\rho}_t(x)$, now with total mass 1, with $\overline{q}(x;0,y) = \delta_y(x)$, we would get

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

The term $\overline{d}_t(x) = -\partial_t \overline{\rho}_t(x)/\overline{\rho}_t(x) < 0$ is the rate at which mass should be removed to compensate the creation of mass of the BD process $\left(\left(\widetilde{x}_t^{(n)}\right)_{n=1}^{N_t(x)};\ t \geq 0\right)$ arising from splitting. In other words,

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

where $p^{(n)}(x;t,y)$ would be the density at (t,y) of the nth alive particle in the system, descending from the ancestral one, started at x. Thus $\overline{q}(x;t,y)$ would be the average presence density at (t,y) of the branching system of Kimura particles.

We have $\overline{d}_t(x) \to \lambda_1$ where λ_1 should then be the largest negative eigenvalue of $-G$ and therefore, putting $\partial_t \overline{q} = 0$ in the latter evolution equation, we would get that, independently of the initial condition x

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

[4] When $\gamma = 0$, $\overline{\rho}_t(x) = e^{(\rho t)/8}$, as required from the neutral theory.

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

$$\begin{aligned} -\widetilde{G}^{*}(\overline{q}_{\infty}) &= (\lambda_1 + \lambda(y)) \cdot \overline{q}_{\infty}, \text{ or} \\ -\overline{G}^{*}(\overline{q}_{\infty}) &= \lambda_1 \cdot \overline{q}_{\infty}. \end{aligned}$$

This solution could exist because the Kimura diffusion (with generator $\widetilde{G}$ with smooth drift and local variance) being recurrent on a bounded domain and the birth and death rate λ being bounded above, the operator $\overline{G}(\cdot) + \lambda_1 \cdot$ (and its adjoint) could be critical [5] (in the sense that it is critical for some λ_c which could be λ_1, see Th. 3.2 of [31], Chapter 4). λ_1 would then be called the generalized principal eigenvalue for $\overline{G}$ (or $\overline{G}^{*}$).

From the construction of $\overline{G}$, one would expect (proceeding like in (13-15)) that $\overline{q}_{\infty}(y)$ should be of the product form

$$\overline{q}_{\infty}(y) = \alpha(y) v_1(y) / \int_0^1 \alpha(y) v_1(y) dy, \tag{55}$$

would v_1 be the eigenfunction of $-G^{*}$ associated to $\lambda_1 < 0$. This results directly from the fact that $\overline{G}^{*}(\cdot) = \alpha(y) G^{*}(\cdot/\alpha(y))$. As a result, the infinitesimal generator $-G^{*}$ of the nonneutral Karlin diffusion with natural boundaries should itself possess a unique (up to multiplicative constant) eigenvector associated to $\lambda_1 = -\lambda_{*}$. Similarly, $-G$ also should have an eigenvector associated to $\lambda_1 = -\lambda_{*}$ which is $u_1(x) = v_1(x)/m(x)$, where $m(x)$ is the speed measure density of the Karlin diffusion, (36). Furthermore, we would hope the product criticality property $\int_{(0,1)} u_1(x) \cdot v_1(x) dx < \infty$. As a result, the Karlin operator $G^{*}(\cdot) + \lambda_1 \cdot$ and its adjoint would themselves be critical and product critical.

Note that there also should exist $\overline{\phi}_{\infty}(x)$ such that $-\overline{G}\left(\overline{\phi}_{\infty}\right) = \lambda_1 \overline{\phi}_{\infty}$ and that $\overline{\phi}_{\infty}(x) = u_1(x)/\alpha(x)$. All this would make sense if the product criticality property

$$\int_{(0,1)} \overline{\phi}_{\infty}(x) \cdot \overline{q}_{\infty}(x) dx = \int_{(0,1)} u_1(x) \cdot v_1(x) dx < \infty$$

would hold, in which case we could choose

$$\int_{(0,1)} \overline{\phi}_{\infty}(x) \cdot \overline{q}_{\infty}(x) dx = 1 \text{ and } \int_{(0,1)} \overline{q}_{\infty}(x) dx = 1,$$

(See [31], Subsection 4.9).

Would all this be the case, $e^{\lambda_1 t} \sum_{n=1}^{N_t(x)} \overline{\phi}_{\infty}\left(\widetilde{x}_t^{(n)}\right)$ would be a martingale converging a.s. to a nondegenerate random variable $W(x)$ satisfying $\mathbf{E}(W(x)) = \overline{\phi}_{\infty}(x)$, (see [1]). For any a.e. continuous bounded measurable function ψ on I

$$e^{\lambda_1 t} \sum_{n=1}^{N_t(x)} \psi\left(\widetilde{x}_t^{(n)}\right) \underset{t \to \infty}{\overset{a.s.}{\to}} W(x) \frac{\int_{(0,1)} \psi(x) \cdot \overline{q}_{\infty}(x) dx}{\int_{(0,1)} \overline{q}_{\infty}(x) dx}.$$

In particular,

$$e^{\lambda_1 t} N_t(x) \underset{t \to \infty}{\overset{a.s.}{\to}} W(x),$$

[5] $\overline{G}(\cdot) + \lambda_c \cdot (\overline{G}^{*}(\cdot) + \lambda_c \cdot)$ is said to be critical if there exists some function $\overline{\phi}_{\infty} \in C^2$ (respectively $\overline{q}_{\infty} \in C^2$), strictly positive in $(0,1)$, such that: $\overline{G}\left(\overline{\phi}_{\infty}\right) + \lambda_c \overline{\phi}_{\infty} = 0$ (respectively $\overline{G}^{*}(\overline{q}_{\infty}) + \lambda_c \overline{q}_{\infty} = 0$).

making more precise how fast the global expected number of particles would grow within the whole system.

But all this global picture does **not** hold simply because one can check that there is no positive $(u_1(x); v_1(y))$ satisfying $-G(u_1) = \lambda_1 u_1$ and $-G^*(v_1) = \lambda_1 v_1$ for $\lambda_1 = -\left(\frac{|\gamma|}{2} + \frac{\rho}{8}\right)$. In fact, these eigenvectors exist but for some alternative eigenvalue $\lambda_c > \lambda_1$ that we will give below (actually, because there can be only one critical value, see [31], Th. 3.2 on p. 146, this rules out de facto λ_1). So criticality of $\overline{G}(\cdot) + \lambda_1 \cdot$ is not valid and the above global approach fails. However the criticality of $\overline{G}(\cdot) + \lambda_c \cdot$ indeed holds and we will rather need to focus on a local approach. Let us first introduce

$$\lambda_c = -\frac{\rho}{8}\left(1 - 4\left(\frac{\gamma}{\rho}\right)^2\right) > \lambda_1 = -\frac{\rho}{8}\left(1 + 4\frac{|\gamma|}{\rho}\right), \tag{56}$$

with $\lambda_c < 0$ if and only if $|\gamma| < \rho/2$.

It turns out that $\overline{G}(\cdot) + \lambda_c \cdot$ and $\overline{G}^*(\cdot) + \lambda_c \cdot$ are indeed critical with respective ground states $\overline{\phi}_\infty(x) > 0$ and $\overline{q}_\infty(y) > 0$ satisfying $\int_{(0,1)} \overline{\phi}_\infty(x)\, \overline{q}_\infty(x)\, dx = \infty$. So λ_c is the effective generalized principal eigenvalue. Let us check this point first and compute $\overline{\phi}_\infty(x)$ and $\overline{q}_\infty(y)$.

We have $-G(u_1) = \lambda_c u_1$, $-G^*(v_1) = \lambda_c v_1$ with $u_1(x) = v_1(x)/m(x)$ and $m(x)$ given again by (36) and a straightforward computation shows that

$$\begin{aligned} u_1(x) &= x^{-\frac{\gamma}{\rho}-\frac{1}{2}}(1-x)^{\frac{\gamma}{\rho}-\frac{1}{2}} \\ v_1(y) &= m(y)\, u_1(y) = y^{\frac{2\gamma}{\rho}-1}(1-y)^{-\frac{2\gamma}{\rho}-1} u_1(y) \\ &= y^{\frac{\gamma}{\rho}-\frac{3}{2}}(1-y)^{-\frac{\gamma}{\rho}-\frac{3}{2}}. \end{aligned}$$

Therefore, following the arguments leading to (55), up to a multiplicative constant, we get the positive ground states solutions on $\overset{\circ}{I} = (0,1)$ as

$$\overline{\phi}_\infty(x) = \frac{1}{\alpha(x)} u_1(x) = x^{-\frac{\gamma}{\rho}}(1-x)^{\frac{\gamma}{\rho}} \tag{57}$$

$$\overline{q}_\infty(y) = \alpha(y)\, v_1(y) = y^{\frac{\gamma}{\rho}-2}(1-y)^{-\frac{\gamma}{\rho}-2} \tag{58}$$

with

$$\int_{(0,1)} u_1(x) \cdot v_1(x)\, dx = \int_{(0,1)} \overline{\phi}_\infty(x)\, \overline{q}_\infty(x)\, dx = \int_{(0,1)} x^{-2}(1-x)^{-2}\, dx = \infty.$$

So again, the product criticality property does not hold but this is not so disturbing because the growth property under concern is only local in the following sense.

Let B be any Borel subset of $\overset{\circ}{I}$ with closure $\overline{B} \subset \overset{\circ}{I}$ [A suitable choice of B could typically be the interior of I_ε]. Let $N_t(x, B) = \sum_{n=1}^{N_t(x)} \mathbf{1}_B\left(\widetilde{x}_t^{(n)}\right)$ count the local number of Kimura particles within B at time t given Eve started at x. Let $\overline{\phi}_\infty^B(x)$ and $\overline{q}_\infty^B(y)$ denote the above functions where the multiplicative constants were adjusted in such a way that $\int_B \overline{\phi}_\infty^B(x) \cdot \overline{q}_\infty^B(x)\, dx = \int_B \overline{q}_\infty^B(x)\, dx = 1$. Then, we have the local version of the Asmussen-Hering result:

Local supercriticality (growth). Suppose $\lambda_c < 0$ or $|\gamma| < \rho/2$.

In our case study, for each B, $e^{\lambda_c t}\sum_{n=1}^{N_t(x)} \overline{\phi}_\infty^B\left(\widetilde{x}_t^{(n)}\right)\mathbf{1}_B\left(\widetilde{x}_t^{(n)}\right)$ is a martingale converging a.s. to a nondegenerate random variable $W_B(x)$ satisfying $\mathbf{E}(W_B(x)) = \overline{\phi}_\infty^B(x)$ (see [8], p. 84).

For any a.e. continuous bounded measurable function ψ on I,

$$e^{\lambda_c t}\sum_{n=1}^{N_t(x)} \psi\left(\widetilde{x}_t^{(n)}\right)\mathbf{1}_B\left(\widetilde{x}_t^{(n)}\right) \underset{t\to\infty}{\overset{a.s.}{\to}} W_B(x)\,\frac{\int_B \psi(x)\cdot \overline{q}_\infty^B(x)\,dx}{\int_B \overline{q}_\infty^B(x)\,dx}. \tag{59}$$

In particular ($\psi \equiv 1$),

$$e^{\lambda_c t} N_t(x,B) \underset{t\to\infty}{\overset{a.s.}{\to}} W_B(x), \tag{60}$$

making more precise how fast the expected number of particles grows locally within each B of the open interval.

Note that $-\lambda_c > 0$ is the local Malthus growth parameter of $N_t(x,B)$. As conventional wisdom suggests, it is smaller than the global one $-\lambda_1$.

Local subcriticality (extinction). Suppose $\lambda_c > 0$ or $|\gamma| > \rho/2$.

For each B :

(i)

$$\mathbf{P}(N_t(x,B) = 0) \underset{t\to\infty}{\to} 1, \tag{61}$$

uniformly in x.

(ii) Suppose $x \in B$. There exists a constant $\gamma_B > 0$ such that:

$$e^{\lambda_c t}\left[1 - \mathbf{P}(N_t(x,B) = 0)\right] \underset{t\to\infty}{\to} \gamma_B \overline{\phi}_\infty^B(x), \tag{62}$$

uniformly in x.

(iii) For all bounded measurable function ψ on I :

$$\mathbf{E}\left[\sum_{n=1}^{N_t(x)} \psi\left(\widetilde{x}_t^{(n)}\right)\mathbf{1}_B\left(\widetilde{x}_t^{(n)}\right) \mid N_t(x,B) > 0\right] \underset{t\to\infty}{\to} \gamma_B^{-1}\int_B \psi(y)\,\overline{q}_\infty^B(y)\,dy. \tag{63}$$

From (i), it is clear that, under the assumption $|\gamma| > \rho/2$, the process gets ultimately extinct with probability 1, locally for each B. In this subcritical regime, the drift is so strong (and the affinity of Kimura particles for the boundaries so large) that it pushes all the particles very close to either boundaries, all ending up eventually outside B.

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

(iii) with $\psi \equiv 1$ reads $\mathbf{E}\left[N_t\left(x,B\right) \mid N_t\left(x,B\right)>0\right] \underset{t\to\infty}{\to} \gamma_B^{-1}$ giving an interpretation of the constant γ_B (which may be hard to evaluate in practise).

Thus $\lambda_c = 0$ or $|\gamma| = \rho/2$ is a case of local criticality where the process also gets ultimately locally extinct with probability 1 but at a smaller-$1/t$ speed than in the subcritical regime.

Remark. The ground states of $\overline{G}+\lambda_c$ and its adjoint are thus $\left(\overline{\phi}_\infty, \overline{q}_\infty\right)$ and explicit here; see (57) and (58). We note that $\overline{\phi}_\infty$ is unbounded and that $\overline{q}_\infty$ is not integrable on $[0,1]$. Nor is the product of the two integrable. It is useful to consider the process whose infinitesimal generator is given by the Doob-transform

$$\overline{\phi}_\infty^{-1}\left(\overline{G}+\lambda_c\right)\left(\overline{\phi}_\infty \cdot\right) = \overline{\phi}_\infty^{-1}\left(\widetilde{G}+\lambda+\lambda_c\right)\left(\overline{\phi}_\infty \cdot\right),$$

because product-criticality is preserved under this transformation. The ground states associated to this new operator and its adjoint are $\left(1, \overline{\phi}_\infty \overline{q}_\infty\right)$. Developing the Doob transform, we obtain a process whose infinitesimal elliptic generator is $\widetilde{G}+\frac{\overline{\phi}'_\infty}{\overline{\phi}_\infty} g^2 \partial_x$,with no multiplicative component.

Recalling $g^2 = \frac{\rho}{2}x^2\left(1-x\right)^2$, we have $\widetilde{G} = \gamma x\left(1-x\right)\partial_x + \frac{\rho}{2}x^2\left(1-x\right)^2\partial_x^2$ and $\frac{\overline{\phi}'_\infty}{\overline{\phi}_\infty} = -\frac{\gamma}{\rho}\frac{1}{x(1-x)}$. Thus

$$\widetilde{G}+\frac{\overline{\phi}'_\infty}{\overline{\phi}_\infty} g^2 \partial_x = \frac{\rho}{2}x^2\left(1-x\right)^2\partial_x^2$$

is the driftless infinitesimal generator of the Kimura martingale. The associated diffusion process is thus null recurrent with a constant harmonic function 1 and invariant measure $\overline{\phi}_\infty \overline{q}_\infty = u_1 v_1 \propto x^{-2}\left(1-x\right)^{-2}$ which is not integrable. $\diamond$

Acknowledgments: The author thanks his Referees for the careful reading of the manuscript, leading to some significant improvements in the presentation. The author also acknowledges support from the ANR Modélisation Aléatoire en Ecologie, Génétique et Evolution (ANR-Manège- 09-BLAN-0215).

References

[1] Asmussen, S.; Hering, H. Strong limit theorems for general supercritical branching processes with applications to branching diffusions. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 36, no. 3, 195-212, 1976.

[2] Asmussen, S.; Hering, H. Some modified branching diffusion models. Math. Biosci. 35, no. 3-4, 281-299, 1977.

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

[4] Crow, J.; Kimura, M. Some genetic problems in natural populations. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. IV, pp. 1–22. University of California Press, Berkeley and Los Angeles, 1956.

[5] Durrett, R. Probability models for DNA sequence evolution. Second edition. Probability and its Applications (New York). Springer, New York, 2008.

[6] Dynkin, E. B. *Markov processes.* Vols. I, II. Translated with the authorization and assistance of the author by J. Fabius, V. Greenberg, A. Maitra, G. Majone. Die Grundlehren der Mathematischen 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.

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

[8] Engländer, J.; Kyprianou, A. E. Local extinction versus local exponential growth for spatial branching processes. Ann. Probab. 32 (2004), no. 1A, 78-99.

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

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

[11] Gillespie, J. H. The effects of stochastic environments on allele frequencies in natural populations. Theoretical Population Biology, Volume 3, Issue 3, 241-248, 1972.

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

[13] 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.

[14] Huerta-Sanchez, E.; Durrett, R.; Bustamante C.D. Population genetics of polymorphism and divergence under fluctuating selection. Genetics, 178(1): 325337, 2008.

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

[16] Itô, K. On stochastic differential equations. Mem. Am. Math. Soc. No. 4 (1951).

[17] Jensen, L. Random selective advantages of genes and their probabilities of fixation. Genet. Res. 21, pp. 215-219, 1973.

[18] Karlin, S.; Levikson B. Temporal fluctuations in selection intensities: case of small population size. Theoretical Population Biology, 6: 383-412, 1974.

[19] Karlin, S.; Liberman, U. Random temporal variation in selection intensities: one-locus two-allele model. J. Math. Biol. 2, no. 1, 1–17, 1975.

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

[21] Kimura, M. Process leading to quasi-fixation of genes in natural populations due to random fluctuations of selection intensities. Genetics, 39, 280-295, 1954.

[22] Kimura, M. Stochastic processes and distribution of gene frequencies under natural selection. Cold Spring Harbor Symposia on Quantitative Biology 20, 33-53, 1955. Also, contribution 114 of the National Institute of Genetics, Mishima-shi, Japan.

[23] Kimura, M. On the probability of fixation of mutant genes in a population. Genetics, 47, 713-19, 1962.

[24] Kussell, E.; Leibler; S. Phenotypic diversity, population growth, and information in fluctuating environments. Science, 309(5743), 2005.

[25] 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.

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

[27] May, R. M. Stability in model ecosystems, Proc. Ecol. Soc. Aust. 6, 18-56, 1971.

[28] Mustonen V., Lässig M. Molecular evolution under fitness fluctuations. Phys. Rev. Letters, 100, 108101, 2008.

[29] Mortensen, R. E. Mathematical problems of modeling stochastic nonlinear dynamic systems, J. Stat. Phys. 1, 271-296, 1969.

[30] 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.

[31] Pinsky, R. G. *Positive harmonic functions and diffusion.* Cambridge Studies in Advanced Mathematics, 45. Cambridge University Press, Cambridge, 1995.

[32] Steinsaltz, D.; Evans, S. N. Quasistationary distributions for one-dimensional diffusions with killing. Trans. Amer. Math. Soc. 359 (2007), no. 3, 12851324 (electronic).

[33] Stratonovich, R.L. A new representation for stochastic integrals and equations. SIAM J. Control 4 (1966), pp. 362371.
[34] Takahata, N.; Ishii, K.; Matsuda, H. Effect of temporal fluctuation of selection coefficient on gene frequency in a population. Proc. Nat. Acad. Sci., USA, vol 72 No 11, pp. 4541-4545, 1975.
[35] 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.
[36] Tuckwell, H. C. The effects of random selection on gene frequency. Math. Biosciences, 30, 113-128, 1976.

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