A MODEL FOR MENDELIAN POPULATIONS DEMOGENETICS [*]

Camille Coron[1]

Abstract. This proceeding introduces and studies a diffusion process that models the demogenetic behavior of a diploid population with sexual Mendelian reproduction. This process is defined as the slow limit of a slow-fast dynamics derived from the rescaling of a birth-and-death process with interactions.

Résumé. Cet acte de conférence introduit et étudie un processus de diffusion modélisant le comportement démo-génétique d'une population diploïde à reproduction sexuée. Ce processus est obtenu comme la limite lente d'une dynamique stochastique lente-rapide obtenue par changement d'échelle d'un processus de naissance et mort avec interactions.

1. Context

We are interested in modeling and studying the demogenetic dynamics of a diploid Mendelian population with competition. Our motivation is more precisely to understand first, the interactions between population size and population genetics dynamics, and second, the impact of sexual reproduction on population demogenetics. This proceeding is based on the article Coron (2013) in which proofs of the presented results are detailed, though we consider here a slightly more general model, taking into account breeding preferences.

Let us consider a population of diploid individuals that are characterized by their genotype at one locus for which there exist 2 alleles, A and a (also denoted respectively 1 and 2). We study the genetic dynamics of the population, i.e. the dynamics of the respective numbers of individuals with genotype AA, Aa, and aa. First, let us recall some models of genetic or demogenetic dynamics of diploid and/or haploid populations.

The following well-known neutral Wright-Fisher diffusion models the dynamics of a bi-allelic locus in a haploid population with constant size (Wright (1931)). In this model, each individual has genotype A or a, X_t is the proportion of allele A at time t, N_e is the (constant) effective population size, s is the selective advantage of allele A over allele a and $(B_t)_{t \geq 0}$ is a standard Brownian motion.

$$dX_t = \sqrt{\frac{X_t(1 - X_t)}{N_e}} dB_t + sX_t(1 - X_t)dt. \tag{1}$$

This diffusion is obtained by the rescaling of a discrete time population model with constant size and non overlapping generations. For diploid populations, an analogous version of this model is presented notably in Crow and Kimura (1970), p. 408. More precisely, the selective advantage of genotype AA (resp. Aa) over

[1] Laboratoire de Mathématiques d'Orsay
Université Paris Sud
91405 Orsay Cedex
camille.coron@math.u-psud.fr

genotype aa is denoted by s (resp. hs, h is called the dominance coefficient) and the frequence $(X_t)_{t\geq 0}$ of allele A then satisfies

$$dX_t = \sqrt{\frac{X_t(1-X_t)}{2N_e}}dB_t + s(h + (1-2h)X_t)X_t(1-X_t)dt. \tag{2}$$

Wright-Fisher diffusions are therefore equivalent for haploid populations and for diploid populations with additive selection (i.e. when $h = 1/2$), when changing the effective haploid population size N_e to $2N_e$ and the haploid selection coefficient s to $s/2$. In other words, studying the genetic dynamics of a population of diploid individuals with constant effective size and additive selection is equivalent to studying the set of alleles carried by these individuals.

Now, Cattiaux and Méléard (2010) studied the following model of haploid population with one bi-allelic locus, stochastically varying size, clonal reproduction, and interaction. In this model, each individual has genotype A or a, and the process $(N_t^h, X_t^h)_{t\geq 0}$ models simultaneously the dynamics of population size and of the proportion of allele A in the population:

$$dN_t^h = \sqrt{2\gamma N_t^h}d\tilde{B}_t^1 + N_t^h\big[\beta - \delta + sX_t^h - N_t^h\big(c_{11}(X_t^h)^2 + c_{22}(1-X_t^h)^2 + (c_{12}+c_{21})X_t^h(1-X_t^h)\big)\big]dt$$

$$dX_t^h = \sqrt{\frac{2\gamma X_t^h(1-X_t^h)}{N_t^h}}d\tilde{B}_t^2 + X_t^h(1-X_t^h)\big[s - N_t^h\big(c_{11}X_t^h - c_{22}(1-X_t^h) + c_{12}(1-X_t^h) - c_{21}X_t^h\big)\big]dt \tag{3}$$

Here $(\tilde{B}_t^1, \tilde{B}_t^2)_{t\geq 0}$ is a standard bidimensional Brownian motion, $\beta - \delta$ (resp. $\beta - \delta + s$) is the growth rate of individuals with genotype a (resp. A), and c_{ij} is the competition rate of an individual with genotype j over an individual with genotype i (we recall that alleles A and a can also be respectively denoted by 1 and 2).

In this work we introduce a diffusion model to study the demogenetic dynamics of a diploid population with sexual reproduction. This diffusion (given in Equation (4)) is obtained as the slow limiting dynamics of a rescaled birth-and-death process with competition and Mendelian reproduction. More precisely, the population is initially modeled by a 3-type birth-and-death process denoted by $\nu^K = (\nu_t^K, t \geq 0)$. Following an infinite population size approximation (as in Champagnat (2006) notably) we assume that the initial number of individuals is of order K where K is a scale parameter that will go to infinity and we consider the sequence of stochastic processes $Z^K = \nu^K/K$. At each time t and for all K, we define the deviation Y_t^K of the population Z_t^K from a so-called Hardy-Weinberg structure. Under a weak-selection regime, the sequence of stochastic processes $(Z^K)_{K\geq 1}$ converges toward a slow-fast dynamics: the sequence of random variables $(Y_t^K)_{K\in\{1,2,\dots\}}$ goes to 0 as K goes to infinity, while the sequence of processes $(N_t^K, X_t^K)_{t\geq 0}$ giving respectively the population size and the proportion of allele A converges in law toward the 2-dimensional diffusion process $(N_t, X_t)_{t\geq 0}$ we are interested in. This convergence toward Hardy-Weinberg structure was obtained for deterministic models in Norton (1928) and for stochastic models with large constant population size in Nagylaki (1992) (Chap. 4.10) and therefore remains true with stochastic population size dynamics. However, due to this stochasticity, the limiting diffusion $(N_t, X_t)_{t\geq 0}$ is not equivalent to the corresponding haploid one obtained from Cattiaux and Méléard (2010), contrarily to what was shown previously for populations with constant population size (Equations (1) and (2)).

In Section 3, absorption behaviors of diploid and haploid populations are compared. Due to competition, these populations get extinct almost surely in finite time and we therefore study the quasi-stationary behavior of the diffusion process $(N_t, X_t)_{t\geq 0}$ conditioned on the non extinction of the population. In particular, we present numerical applications that show the long-time coexistence of the two alleles for three biologically relevant cases: a pure competition neutral case, a case in which each genotype has its own ecological niche, and an overdominance case. Notably, we show that a long-term coexistence of alleles is possible even in some full competition cases, which is not true for haploid clonal reproduction (Cattiaux and Méléard, 2010).

2. MODEL AND SLOW-FAST DYNAMICS

2.1. **Model**

The population at any time t is represented by a 3-dimensional birth-and-death process giving the respective numbers of individuals with genotype AA, Aa and aa. As in Champagnat and Méléard (2007); Fournier and Méléard (2004) or Collet et al. (2013), we consider an infinite population size approximation. To this end we index the population by a scaling parameter $K \in \{1, 2, ...\}$ that will go to infinity. The initial numbers of individuals of each type will be of order K and we then consider the rescaled stochastic processes

$$\left(Z_t^K\right)_{t \geq 0} = \left(Z_t^{1,K}, Z_t^{2,K}, Z_t^{3,K}\right)_{t \geq 0}$$

giving for each time t the respective numbers of individuals with genotypes AA, Aa, and aa divided by K. As in Collet et al. (2013) or Coron (2014), the birth rates of $(Z_t^K)_{t \geq 0}$ model Mendelian reproduction. More precisely, if we set $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1)$, then for all $i \in \{1, 2, 3\}$, the rates $\lambda_i^K(Z)$ at which the stochastic process Z^K jumps from $z = (z_1, z_2, z_3) \in \left(\frac{\mathbb{Z}_+}{K}\right)^3$ to $z + e_i/K$, as long as $z_1 + z_2 + z_3 = n \neq 0$, are given by:

$$\lambda_1^K(z) = \frac{K b_1^K}{n}\left[r_{11}^K z_1^2 + r_{12}^K z_1 z_2 + r_{22}^K \left(\frac{z_2}{2}\right)^2\right],$$

$$\lambda_2^K(z) = \frac{2K b_2^K}{n}\left[r_{12}^K \frac{z_1 z_2}{2} + r_{32}^K \frac{z_3 z_2}{2} + r_{22}^K \left(\frac{z_2}{2}\right)^2 + r_{13}^K z_1 z_3\right],$$

$$\lambda_3^K(z) = \frac{K b_3^K}{n}\left[r_{33}^K z_3^2 + r_{32}^K z_3 z_2 + r_{22}^K \left(\frac{z_2}{2}\right)^2\right].$$

These birth rates are naturally set to 0 if $n = 0$, and the demographic parameters $r_{ij}^K = r_{ji}^K \in \mathbb{R}_+$ (resp. b_i^K) model fertility and breeding preferences (resp. viability at birth) of individuals. Now individuals can die either naturally or due to competition with other individuals. More precisely, for all $i \in \{1, 2, 3\}$, the rates $\mu_i^K(z)$ at which the stochastic process Z^K jumps from $z = (z_1, z_2, z_3) \in (\mathbb{Z}_+)^3/K$ to $z - e_i/K$ for $i \in \{1, 2, 3\}$ are given by:

$$\mu_1^K(z) = K z_1(d_1^K + K(c_{11}^K z_1 + c_{12}^K z_2 + c_{13}^K z_3)),$$

$$\mu_2^K(z) = K z_2(d_2^K + K(c_{21}^K z_1 + c_{22}^K z_2 + c_{23}^K z_3)),$$

$$\mu_3^K(z) = K z_3(d_3^K + K(c_{31}^K z_1 + c_{32}^K z_2 + c_{33}^K z_3)).$$

The demographic parameter $d_i^K \in \mathbb{R}_+$ (resp. c_{ij}^K) is called the intrinsic death rate of individuals of type i (resp. the competition rate of individuals of type j over type i). From now on, we say that the stochastic process Z^K is neutral if its demographic parameters do not depend on individuals genotypes, i.e.

$$b_i^K = b^K, \quad r_{ij}^K = 1, \quad d_i^K = d^K \quad \text{and} \quad c_{ij}^K = c^K \quad \forall i, j \in \{1, 2, 3\}.$$

The pure jump process $(Z_t^K)_{t \geq 0}$ is well defined for all $t \in \mathbb{R}_+$ and is a $\frac{(\mathbb{Z}_+)^3}{K}$-valued pure jump Markov process absorbed at $(0, 0, 0)$, with trajectories in the space $\mathbb{D}(\mathbb{R}_+, (\mathbb{Z}_+)^3/K)$ of left-limited and right-continuous functions from $\mathbb{R}_+$ to $(\mathbb{Z}_+)^3/K$, endowed with the Skorohod topology. The infinitesimal generator L^K of Z^K satisfies for every bounded measurable function f from $(\mathbb{Z}_+)^3/K$ to $\mathbb{R}$ and for every $z \in \frac{(\mathbb{Z}_+)^3}{K}$:

$$L^K f(z) = \sum_{i \in \{1,2,3\}} \left[\lambda_i^K(z)\left(f\left(z + \frac{e_i}{K}\right) - f(z)\right) + \mu_i^K(z)\left(f\left(z - \frac{e_i}{K}\right) - f(z)\right)\right].$$

To end with model description, let us introduce three quantities of interest. First, the rescaled population size at time t and the proportion of allele A at time t are respectively denoted by

$$N_t^K = Z_t^{1,K} + Z_t^{2,K} + Z_t^{3,K} \in \frac{\mathbb{Z}_+}{K} \quad \text{and } X_t^K = \frac{2Z_t^{1,K} + Z_t^{2,K}}{2(Z_t^{1,K} + Z_t^{2,K} + Z_t^{3,K})} \in [0,1].$$

Next, let us define the stochastic processes $(Y_t^K, t \geq 0)$ such that for every $t \geq 0$, as long as $N_t^K > 0$,

$$Y_t^K = \frac{4Z_t^{1,K} Z_t^{3,K} - (Z_t^{2,K})^2}{4N_t^K}.$$

If $N_t^K = 0$, we naturally set $Y_t^K = 0$ as $|Y_t^K| \leq N_t^K$ for all $t \geq 0$. This stochastic process will play a main role afterwards; note first that

$$Y_t^K = Z_t^{1,K} - \frac{(2Z_t^{1,K} + Z_t^{2,K})^2}{4N_t^K} = \left(p_t^{AA,K} - (X_t^K)^2 \right) N_t^K$$

if $p_t^{AA,K}$ is the proportion of genotype AA in the population at time t. Similarly,

$$Y_t^K = \left(p_t^{aa,K} - (1 - X_t^K)^2 \right) N_t^K = \left(2X_t^K(1 - X_t^K) - p_t^{Aa,K} \right) N_t^K.$$

Then if $Y_t^K = 0$, the proportion of each genotype in the population is equal to the proportion of pairs of alleles forming this genotype. If $Y_t^K = 0$ the population Z_t^K is called to satisfy Hardy-Weinberg structure. We prove easily that the triplet (N_t^K, X_t^K, Y_t^K) characterizes the population at time t and from now on, we will mostly use this representation.

2.2. Slow-fast stochastic dynamics and limiting diffusion

In this section, we investigate a diffusive scaling under which both population size and proportion of allele a evolve stochastically with time (in particular the population can get extinct and one of the two alleles can eventually get fixed), while the population still converges rapidly toward Hardy-Weinberg structure.
We assume that individual birth and natural death rates are of order K, while Z_0^K converges in law toward a random vector Z_0 as $K \to \infty$. More precisely, we set for $\gamma > 0$ and for a $(\mathbb{R}_+)^3$-valued random variable Z_0:

$$b_i^K = \gamma K + \beta_i \in [0, \infty[$$
$$r_{ij}^K = 1 + \frac{\rho_{ij}}{K} \in [0, \infty[$$
$$d_i^K = \gamma K + \delta_i \in [0, \infty[$$
$$c_{ij}^K = \frac{\alpha_{ij}}{K} \in \mathbb{R}$$
$$Z_0^K \underset{K \to \infty}{\to} Z_0 \quad \text{in law.}$$

Therefore birth and natural death events are now happening faster and compensate each other, which will introduce some stochasticity in the limiting process. A biological interpretation for the scaling of the interaction coefficients c_{ij}^K is that a constant quantity of resources is shared by small individuals whose biomass is equal to $1/K$ (as presented in Champagnat et al. (2006)); these coefficients will only step in limiting drift terms. Under this scaling of demographic parameters, Y^K will be a "fast" variable that converges directly toward the long time equilibrium of $\mathcal{Y}$ (equal to 0, as studied in Ethier and Nagylaki (1980, 1988); Katzenberger (1991)), while X^K and N^K will be "slow" variables, converging toward a non-deterministic process.
Population size stochasticity induces dificulties linked to both population extinction and population size control. From now on, we assume the following initial 3-rd-order moments condition:

$$\text{there exists } C < \infty \quad \text{such that} \quad \sup_K \mathbb{E}((N_0^K)^3)) \le C. \tag{H1}$$

Lemma 1 of Champagnat (2006) and the proof of Theorem 5.3 of Fournier and Méléard (2004)), ensure the propagation of population size 3-rd-order moments control, which is needed to prove convergence toward a slow-fast dynamics. By domination, this result is generalized in Coron (2013) in a case where cooperation is also possible but competition remains stronger. Then, the following proposition gives that $(Y_t^K, t \ge 0)$ is a fast variable that converges toward the deterministic value 0 as K goes to infinity.

Proposition 2.1. *Under* (H1)*, for all* $s, t > 0$, $\displaystyle\sup_{t \le u \le t+s} \mathbb{E}((Y_u^K)^2) \to 0$ *as* K *goes to infinity.*

The proof of this proposition relies on Kolmogorov-forward equation, and we now study the behavior of the sequence of remaining stochastic processes $((N_t^K, X_t^K)_{t \ge 0})_{K \ge 1}$.

Theorem 2.2. *For any* $\epsilon > 0$ *and* $T > 0$*, let* $T_\epsilon^K = \inf\{t \in [0, T] : N_t^K \le \epsilon\}$*. If the sequence of random variables* $(N_0^K, X_0^K) \in [\epsilon, +\infty[\times[0, 1]$ *converges in law toward a random variable* $(N_0, X_0) \in]\epsilon, +\infty[\times[0, 1]$ *as* K *goes to infinity, then the sequence of stopped stochastic processes* $\{(N^K, X^K)_{.\wedge T_\epsilon^K}\}_{K \ge 1}$ *converges in law in* $\mathbb{D}([0, T], [\epsilon, \infty[\times[0, 1])$ *as* K *goes to infinity, toward the stopped diffusion process* $(N, X)_{.\wedge T_\epsilon}$ $(T_\epsilon = \inf\{t \in [0, T] : N_t = \epsilon\})$*, starting from* (N_0, X_0) *and satisfying the following diffusion equation:*

$$
\begin{aligned}
dN_t = {}& \sqrt{2\gamma N_t}dB_t^1 \\
& + N_t \Big[X_t^2(\beta_1 - \delta_1) + 2X_t(1 - X_t)(\beta_2 - \delta_2) + (1 - X_t)^2(\beta_3 - \delta_3) \\
& \qquad + \gamma \left[\rho_{11}X_t^4 + 4\rho_{12}X_t^3(1 - X_t) + 4\rho_{22}X_t^2(1 - X_t)^2 + 2\rho_{13}X_t^2(1 - X_t)^2 + 4\rho_{23}X_t(1 - X_t)^3 + \rho_{33}(1 - X_t)^4 \right] \\
& \qquad - N_t \big[X_t^2 \left(\alpha_{11}X_t^2 + \alpha_{21}2X_t(1 - X_t) + \alpha_{31}(1 - X_t)^2 \right) \\
& \qquad\qquad + 2X_t(1 - X_t) \left(\alpha_{12}X_t^2 + \alpha_{22}2X_t(1 - X_t) + \alpha_{32}(1 - X_t)^2 \right) \\
& \qquad\qquad + (1 - X_t)^2 \left(\alpha_{13}X_t^2 + \alpha_{23}2X_t(1 - X_t) + \alpha_{33}(1 - X_t)^2 \right) \big] \Big] dt \\[1em]
dX_t = {}& \sqrt{\frac{\gamma X_t(1 - X_t)}{N_t}}dB_t^2 \\
& + (1 - X_t)X_t \Big[(\beta_1 - \delta_1)X_t + (\beta_2 - \delta_2)(1 - 2X_t) + (\beta_3 - \delta_3)(1 - X_t) \\
& \qquad + \gamma \big[\rho_{11}X_t^3 + \rho_{12}X_t^2(3 - 4X_t) + 2\rho_{22}X_t(1 - X_t)(1 - 2X_t) + \rho_{23}(1 - X_t)^2(1 - 4X_t) \\
& \qquad\qquad + \rho_{13}X_t(1 - X_t)(1 - 2X_t) - \rho_{33}(1 - X_t)^3 \big] \\
& \qquad - N_t \big[X_t^2(\alpha_{11}X_t + \alpha_{12}(1 - 2X_t) - \alpha_{13}(1 - X_t)) \\
& \qquad\qquad + 2X_t(1 - X_t)(\alpha_{21}X_t + \alpha_{22}(1 - 2X_t) - \alpha_{23}(1 - X_t)) \\
& \qquad\qquad + (1 - X_t)^2(\alpha_{31}X_t + \alpha_{32}(1 - 2X_t) - \alpha_{33}(1 - X)) \big] \Big] dt,
\end{aligned}
\tag{4}
$$

where $(B_t^1, B_t^2)_{t \ge 0}$ *is a 2-dimensional standard Brownian motion.*

The diffusion equation (4) can be simplified in the neutral case:

Remark 2.3. *If* $\beta_i = \beta$*,* $\rho_{ij} = 0$*,* $\delta_i = \delta$ *and* $\alpha_{ij} = \alpha$ *for all* i, j*, the diffusion process* $(N_t, X_t)_{t \ge 0}$ *satisfies:*

$$
\begin{aligned}
dN_t &= \sqrt{2\gamma N_t}dB_t^1 + N_t(\beta - \delta - \alpha N_t)dt \\
dX_t &= \sqrt{\frac{\gamma X_t(1 - X_t)}{N_t}}dB_t^2.
\end{aligned}
\tag{5}
$$

This diffusion can be seen as a Wright-Fisher diffusion (Equation (2)) associated to a population size evolving stochastically with time. This diffusion will later be compared (Remark 2.5 and next section) to the model of haploid population studied in Cattiaux and Méléard (2010) and recalled in Equation (3).

Elements of proof of Theorem 2.2. Stopping times T_ϵ^K are meant to deal with the explosion of the diffusion coefficient $\sqrt{\frac{\gamma X_t(1-X_t)}{N_t}}$ when N_t goes to 0. To avoid such complications, the convergence result is proved initially for the sequence of processes $((2N_t^K(1-X_t^K), 2N_t^K X_t^K)_{t\geq 0})_{K\geq 1}$ giving the respective quantities of allele A and a, for which limiting diffusion coefficients are bounded near $\{N_t = 0\}$, and is next derived for the sequence $((N_t^K, X_t^K)_{t\geq 0})_{K\geq 1}$. The proof of the convergence for the sequence $((2N_t^K(1-X_t^K), 2N_t^K X_t^K)_{t\geq 0})_{K\geq 1}$, using the Rebolledo-Aldous criterion (Theorem 2.3.2 Joffe and Métivier (1986)), consists in proving first the tightness of this sequence of processes and second its convergence (using Proposition 2.1) toward the unique continuous solution of the appropriate martingale problem. $\qquad\square$

Remark 2.4. As noted in Depperschmidt et al. (2012) (Remark 2.1), diploid additive and haploid selection are equivalent for populations with constant size. In our model, in an additive case for which $\beta_2 - \delta_2 = (\beta_1 - \delta_1) - s$, $\beta_3 - \delta_3 = (\beta_1 - \delta_1) - 2s$, $\rho_{ij} = 0$ and $\alpha_{ij} = \alpha$ for all i, j, the limiting diffusion (N, X) given in Equation (4) satisfies:

$$dN_t = \sqrt{2\gamma N_t}dB_t^1 + N_t(\beta_1 - \delta_1 - \alpha N_t)dt + 2sN_t X_t dt$$

$$dX_t = \sqrt{\frac{\gamma X_t(1-X_t)}{N_t}}dB_t^2 + sX_t(1-X_t)dt.$$

In this case of diploid additive selection, the drift coefficient of the proportion X_t has indeed the same form than the drift of a haploid Wright-Fisher diffusion with selection (Equation (1)). However, the following remark shows that due to interactions between demography and genetics, our model of diploid population with additive selection is different than the model of haploid population with stochastic population size of Equation (3).

Remark 2.5. Consider a neutral version of the haploid model given in Equation (3):

$$dN_t^h = \sqrt{2\gamma N_t^h}d\tilde{B}_t^1 + (\beta - \delta - \alpha N_t^h)N_t^h dt$$

$$dX_t^h = \sqrt{\frac{2\gamma X_t^h(1-X_t^h)}{N_t^h}}d\tilde{B}_t^2 \tag{6}$$

The only difference between diploid and haploid neutral models is that the variation of proportion of allele A is divided by $\sqrt{2}$ in the diploid population (Equations (5) and (6)). However, the respective numbers of alleles A and a are directed by correlated Brownian motions in a diploid population which is not the case in a haploid population. Indeed, setting $N_t^{A,h} = N_t^h(1-X_t^h)$, $N_t^{a,h} = N_t^h X_t^h$, $N_t^A = N_t(1-X_t)$ and $N_t^a = N_t X_t$, we obtain

$$dN_t^{A,h} = \sqrt{2\gamma N_t^{A,h}}dB_t^{1,h} + (\beta - \delta - \alpha(N^{A,h} + N^{a,h}))N^{A,h}dt$$

$$dN_t^{a,h} = \sqrt{2\gamma N_t^{a,h}}dB_t^{2,h} + (\beta - \delta - \alpha(N_t^{A,h} + N_t^{a,h}))N_t^{a,h}dt,$$

where $(B_t^{1,h}, B_t^{2,h})_{t\geq 0}$ is a bidimensional standard Brownian motion, while

$$dN_t^A = \sqrt{\frac{4\gamma}{N_t^A + N_t^a}}N_t^A dB_t^1 + \sqrt{2\gamma\frac{N_t^A N_t^a}{N_t^A + N_t^a}}dB_t^2 + N_t^A\left(\beta - \delta - \alpha\frac{N_t^A + N_t^a}{2}\right)dt$$

$$dN_t^a = \sqrt{\frac{4\gamma}{N_t^A + N_t^a}}N_t^a dB_t^1 - \sqrt{2\gamma\frac{N_t^A N_t^a}{N_t^A + N_t^a}}dB_t^2 + N_t^a\left(\beta - \delta - \alpha\frac{N_t^A + N_t^a}{2}\right)dt,$$

In particular, our model of neutral diploid population with stochastic size does not have the same law as a neutral haploid population with sampling parameter γ replaced by $\gamma/2$.

3. Behavior of the model

In this section we study the long-time behavior of the diffusion process $(N_t, X_t)_{t \geq 0}$ introduced in Theorem 2.2. For any process U, we denote by $\mathbb{P}_x^U$ the law of U starting from a point x, and $\mathbb{E}_x^U$ the associated expectation.

3.1. Extinction and fixation times

Let $T_0 = \inf\{t \geq 0 : N_t = 0\}$ and $T_{\{0,1\}}^X = \inf\{t \geq 0 : X_t \in \{0,1\}\}$ be respectively the extinction time of the population and the fixation time of one of the two alleles A and a. First of all, from Lemma 4.1 of Etheridge (2004) and domination arguments, the population gets extinct almost surely in finite time, i.e. $\mathbb{P}_{(n,x)}^{(N,X)}(T_0 < +\infty) = 1$. The long-time behavior of the diffusion process $(N_t, X_t)_{t \geq 0}$ is therefore trivial, and we study allele fixation before extinction (next proposition) and conditioned on non-extinction (Section 3.3).

Proposition 3.1. *For all $n > 0$, $\mathbb{P}_{(n,x)}^{(N,X)}(T_{\{0,1\}}^X < T_0) = 1$.*

Remark 3.2. The proof of the previous proposition in the neutral case (Equation (5)) relies on the following time change representation: if we define the time change τ such that $\int_0^{\tau(t)} \frac{1}{N_s} ds = t$ for all $t \geq 0$ and set $\hat{X}_t = X_{\tau(t)}$ then $(\hat{X}_t)_{t \geq 0}$ is a diploid Wright-Fisher diffusion with effective size $1/2\gamma$. Similarly, $(\hat{X}_t^h)_{t \geq 0} = (X_{\tau(t)}^h)_{t \geq 0}$ is a haploid Wright-Fisher diffusion with effective size $1/2\gamma$. In particular this representation implies from Equations (1) and (2) that alleles fixation time in a neutral haploid population with demographic parameters γ, β, δ and α is stochastically dominated by alleles fixation time in a neutral diploid population with same demographic parameters.

3.2. New change of variables

We now study the long-time behavior of the diffusion process $(N_t, X_t)_{t \geq 0}$ conditioned on non-extinction, i.e. conditioned on not reaching the absorbing state $(0, 0)$. In particular, we are interested in studying the possibility of a long-time coexistence of the two alleles A and a in the population conditioned on non-extinction. Cattiaux et al. (2009) and Cattiaux and Méléard (2010) studied the quasi-stationary behavior of multi-dimensional Kolmogorov diffusions (i.e. diffusion processes with unit diffusion coefficient and gradient drift). We therefore change variables to obtain a 2-dimensional Kolmogorov diffusion (under conditions on the interactions parameters that will be discussed later). Let us define $(S_t, t \geq 0) = ((S_t^1, S_t^2), t \geq 0)$, with

$$S_t^1 = \sqrt{\frac{2N_t}{\gamma}} \cos\left(\frac{\arccos(2X_t - 1)}{\sqrt{2}}\right)$$

$$S_t^2 = \sqrt{\frac{2N_t}{\gamma}} \sin\left(\frac{\arccos(2X_t - 1)}{\sqrt{2}}\right).$$

To begin with, simple calculations give the following proposition, illustrated in Figure 1.

Proposition 3.3. *For all $t \geq 0$, one has $S_t^2 \geq 0$ and $S_t^2 \geq u S_t^1$, with $u = \tan\left(\frac{\pi}{\sqrt{2}}\right) < 0$.*

Remark 3.4. Let us define for all $(s_1, s_2) \in \mathbb{R}^2$, the sets $\mathbf{A} = \{s_2 = 0, s_1 > 0\}$, $\mathbf{a} = \{s_2 = us_1, s_2 > 0\}$ and $\mathbf{0} = \{s_1 = s_2 = 0\}$. The events $\{S_t \in \mathbf{A}\}$, $\{S_t \in \mathbf{a}\}$, and $\{S_t \in \mathbf{0}\}$ are respectively equal to the events $\{X_t = 1\}$ (fixation of allele A), $\{X_t = 0\}$ (fixation of allele a) and $\{N_t = 0\}$ (extinction of the population).
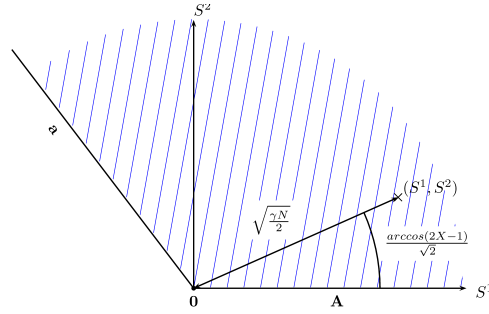
FIGURE 1. Set $\mathcal{D}$ of the possible values of S_t, for $t \geq 0$.

We denote by $\mathcal{D} = \mathbf{R} \times \mathbf{R}_+ \cap \{(s_1, s_2) : s_2 \geq u s_1\}$ the set of values taken by S_t for $t \geq 0$, $\partial \mathcal{D} = \mathbf{A} \cup \mathbf{a} \cup \mathbf{0}$ its boundary in $\mathbb{R}^2$, and $T_{\mathbf{D}}$ the hitting time of $\mathbf{D}$ by $(S_t)_{t \geq 0}$ for any $\mathbf{D} \subset \mathcal{D}$. $\mathbf{0}$, $\mathbf{A} \cup \mathbf{0}$ and $\mathbf{a} \cup \mathbf{0}$ are absorbing sets and are reached almost surely in finite time by $(S_t)_{t \geq 0}$.

We define the following symmetry hypothesis for competition parameters:

$$\alpha_{ij} = \alpha_{ji}, \quad \forall i, j \in \{1, 2, 3\}. \tag{H2}$$

From Itô's formula, under (H2) the diffusion process $(S_t, t \geq 0)$ satisfies the following diffusion equation:

$$dS_t = dW_t - \nabla Q(S_t)dt,$$

where the potential Q satisfies in the neutral case and for all $s = (s_1, s_2) \in \mathcal{D}$,

$$Q(s) = \begin{cases} \frac{\ln((s_1)^2 + (s_2)^2)}{2} + \frac{1}{2}\ln\left(\sin\left(\sqrt{2}\arctan\left(\frac{s_2}{s_1}\right)\right)\right) \\ \qquad - (\beta - \delta - \frac{\alpha\gamma}{4}((s_1)^2 + (s_2)^2))\frac{(s_1)^2 + (s_2)^2}{4} & \text{if } s_1 \geq 0 \\[2ex] \frac{\ln((s_1)^2 + (s_2)^2)}{2} + \frac{1}{2}\ln\left(\sin\left(\sqrt{2}\left(\arctan\left(\frac{s_2}{s_1}\right) + \pi\right)\right)\right) \\ \qquad - (\beta - \delta - \frac{\alpha\gamma}{4}((s_1)^2 + (s_2)^2))\frac{(s_1)^2 + (s_2)^2}{4} & \text{if } s_1 \leq 0. \end{cases}$$

The corresponding potential Q in the non-neutral case is given in Coron (2013).

3.3. Quasi-stationary behavior

From Cattiaux and Méléard (2010) and since some additional properties of the potential Q are proved in Proposition 8 of Coron (2013), the law of S_t conditioned on $S_t \notin \mathbf{0}$ converges.

Theorem 3.5. *There exists a unique probability measure ν_0 on $\mathcal{D} \setminus \mathbf{0}$ such that for all $s \in \mathcal{D} \setminus \partial \mathcal{D}$ and for all $E \subset \mathcal{D} \setminus \mathbf{0}$,*

$$\lim_{t \to \infty} \mathbb{P}_s^S(S_t \in E | T_{\mathbf{0}} > t) = \nu_0(E).$$

The quasi-stationary behavior of the diffusion process $((N_t, X_t), t \geq 0)$ conditioned on non extinction is then obtained easily since the variables (N_t, X_t) are obtained from (S_t^1, S_t^2) by a change of variables (Figure 1). Let us remind that we are interested in studying the possibility of a long-time coexistence of the two alleles A and a in the population conditioned on non-extinction. This means that we want to approximate the quasi-stationary distribution ν_X such that

$$\nu_X(.) := \lim_{t \to \infty} \mathbb{P}_{(n,x)}^{N,X}(X_t \in . | N_t > 0) \tag{7}$$

and we are see whether $\nu_X(]0,1[) = 0$ or not. Indeed, if $\nu_X(]0,1[) \neq 0$ then we can observe a long-time coexistence of the two alleles in the population conditioned on non-extinction whereas if $\nu_X(]0,1[) = 0$, no such coexistence is possible.

4. Numerical results

Numerical simulations of ν_X are obtained following the Fleming-Viot algorithm introduced in Burdzy et al. (1996) and studied in Villemonais (2011, 2013). This approach consists in approximating the conditioned distribution $\mathbb{P}_{(n,x)}^{N,X}((N_t, X_t) \in .|T_0 > t)$ by the empirical distribution of an interacting particle system. Here we present three biologically relevant examples. For each case, we plot the empirical distribution of proportion of allele a, at a large time T. First, we consider a neutral competitive case (Equation (5)). Here, our simulation of the quasi-stationary distribution ν_X of the proportion X is a sum of two Dirac functions in 0 and 1 (Figure 2), i.e. alleles A and a do not coexist in a long time limit. Second (Figure 3), we show a case in which individuals only
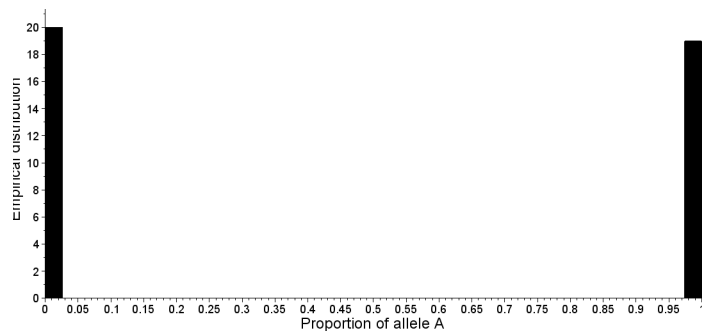

Figure 2. Approximation of the quasi-stationary distribution ν_X of the proportion X of allele a (Equation (7)), in a neutral competitive case. In this figure, $\beta_i = 1$, $\rho_{ij} = 0$, $\delta_i = 0$, and $\alpha_{ij} = 0.1$ for all i, j, and $T = 40$.

compete with individuals with same genotype; this can happen if different genotypes feed differently and/or have different predators. In this case, we can observe either a coexistence of the two alleles A and a or an elimination of one of the alleles, since the distribution ν_X charges both $\{0\} \cup \{1\}$ and $]0,1[$. Third (Figure 4), we show an overdominance case: every individual competes equally with every other ones but heterozygotes (individuals with genotype Aa) are favored compared to homozygotes (individuals with genotype AA or aa), as their reproduction rate is higher. In this case, our simulation of the quasi-stationary distribution ν_X charges only points of $]0,1[$, i.e. alleles A and a coexist with probability 1 or close to 1. This behavior is specific to the Mendelian reproduction: in Cattiaux and Méléard (2010), the authors proved that no coexistence of alleles is possible in a haploid population with clonal reproduction, if every individual is in competition with every other one.

References

Burdzy, K., R. Holyst, D. Ingerman, and P. March (1996). Configurational transition in a fleming-viot-type model and probabilistic interpretation of laplacian eigenfunctions. *J. Phys. A 29*(11), 2633–2642.

Cattiaux, P., P. Collet, A. Lambert, S. Martínez, S. Méléard, and J. San Martín (2009). Quasi-stationary distributions and diffusion models in population dynamics. *Ann. Probab. 37*(5), 1926–1969.
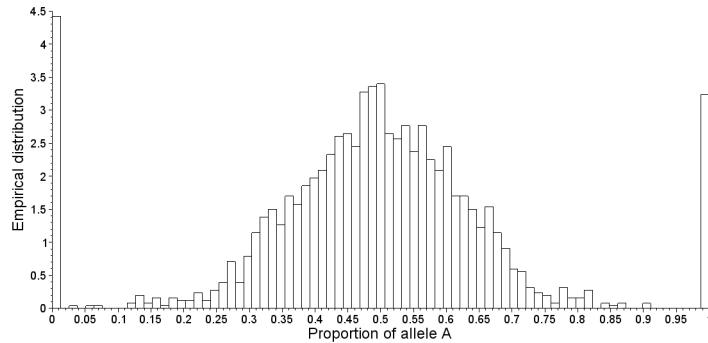
FIGURE 3. Approximation of the quasi-stationary distribution ν_X of the proportion X of allele a (Equation (7)), in a case where individuals with different genotypes do not compete or cooperate with each other. In this figure, $\beta_i = 1$, $\rho_{ij} = 0$, $\delta_i = 0$, $\alpha_{ii} = 0.1$ for all i, $\alpha_{ij} = 0$ for all $i \neq j$, and $T = 2500$.
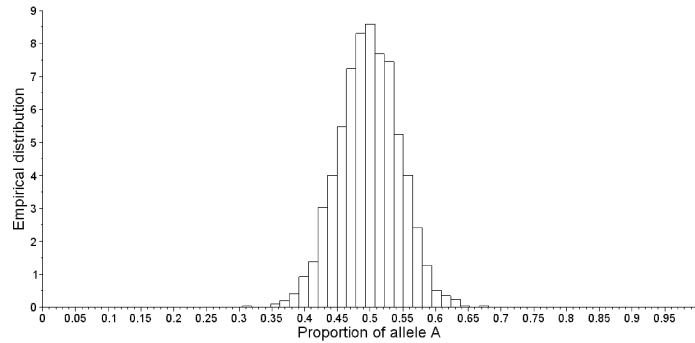

FIGURE 4. Approximation of the quasi-stationary distribution ν_X of the proportion X of allele a (Equation (7)), in an overdominance case. In this figure, $\beta_i = 1$ for all $i \neq 2$, $\beta_2 = 5$, $\delta_i = 0$ for all i, $\alpha_{ij} = 0.1$ for all (i, j), and $T = 500$.

Cattiaux, P. and S. Méléard (2010). Competitive or weak cooperative stochastic Lotka-Volterra systems conditioned on non-extinction. *J. Math. Biol. 60*(6), 797–829.

Champagnat, N. (2006). A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stochastic Process. Appl. 116*(8), 1127–1160.

Champagnat, N., R. Ferrière, and S. Méléard (2006). Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models. *Theor. Popul. Biol. 69*, 297–321.

Champagnat, N. and S. Méléard (2007). Invasion and adaptive evolution for individual-based spatially structured populations. *Journal of Mathematical Biology 55*, 147–188.

Collet, P., S. Méléard, and J. A. Metz (2013). A rigorous model study of the adaptive dynamics of Mendelian diploids. *J. Math. Biol. 67*(3), 569–607.

Coron, C. (2013). Slow-fast stochastic diffusion dynamics and quasi-stationary distributions for diploid populations. *Prepublication, ArXiv 1309.3405*.

Coron, C. (2014). Stochastic modeling of density-dependent diploid populations and extinction vortex. *Adv. in Appl. Probab. 46*, 446–477.

Crow, J. F. and M. Kimura (1970). *An introduction to population genetics theory*. New York: Harper & Row Publishers.

Depperschmidt, A., A. Greven, and P. Pfaffelhuber (2012). Tree-valued Fleming-Viot dynamics with mutation and selection. *Ann. Appl. Probab. 22*(6), 2560–2615.

Etheridge, A. M. (2004). Survival and extinction in a locally regulated population. *Ann. Appl. Probab. 14*(1), 188–214.

Ethier, S. N. and T. Nagylaki (1980). Diffusion approximations of markov chains with two time scales and applications to population genetics. *Adv. in App. Probab. 12*(1), 14–49.

Ethier, S. N. and T. Nagylaki (1988). Diffusion approximations of markov chains with two time scales and applications to population genetics, ii. *Advances in Applied Probability 20*(3), 525–545.

Fournier, N. and S. Méléard (2004). A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab. 14*(4), 1880–1919.

Joffe, A. and M. Métivier (1986). Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Adv. in Appl. Probab. 18*(1), 20–65.

Katzenberger, G. S. (1991, 10). Solutions of a stochastic differential equation forced onto a manifold by a large drift. *Ann. Probab. 19*(4), 1587–1628.

Nagylaki, T. (1992). *Introduction to theoretical population genetics*. Berlin: Springer-Verlag.

Norton, H. (1928). Natural selection and mendelian variation. *Proc. Lond. Math. Soc. 2*(1), 1–45.

Villemonais, D. (2011). Interacting particle systems and Yaglom limit approximation of diffusions with unbounded drift. *Electron. J. Probab. 16*, no. 61, 1663–1692.

Villemonais, D. (2013, 7). General approximation method for the distribution of markov processes conditioned not to be killed. *ESAIM: Probability and Statistics eFirst*.

Wright, S. (1931). Evolution in mendelian populations. *Genetics 16*(2), 97–159.