

Diffusion approximation of a multilocus model with assortative mating

Alison Etheridge, Sophie Lemaire

▶ **To cite this version:**

Alison Etheridge, Sophie Lemaire. Diffusion approximation of a multilocus model with assortative mating. 2011. hal-00560285v2

HAL Id: hal-00560285
https://hal.science/hal-00560285v2

Preprint submitted on 31 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Diffusion approximation of a multilocus model with assortative mating

A. M. Etheridge* S. Lemaire†

Abstract

To understand the effect of assortative mating on the genetic evolution of a population, we consider a finite population in which each individual has a type, determined by a sequence of n diallelic loci. We assume that the population evolves according to a Moran model with weak assortative mating, strong recombination and low mutation rates. With an appropriate rescaling of time, we obtain that the evolution of the genotypic frequencies in a large population can be approximated by the evolution of the product of the allelic frequencies at each locus, and the vector of the allelic frequencies is approximately governed by a diffusion. We present some features of the limiting diffusions (in particular their boundary behaviour and conditions under which the allelic frequencies at different loci evolve independently). If mutation rates are strictly positive then the limiting diffusion is reversible and, under some assumptions, the critical points of the stationary density can be characterised.

AMS 2000 Subject Classification: 60J20, 92D25, 60J70

Key words: Moran model, population genetics, multilocus models, assortative mating, diffusion approximation.

*Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK; email: etheridg@stats.ox.ac.uk. AME supported in part by EPSRC grant no. EP/G052026/1

†Univ. Paris-Sud, Laboratoire de Mathématiques, UMR 8628, Orsay F-91405; CNRS, Orsay, F-91405; email: sophie.lemaire@math.u-psud.fr.

1 Introduction

The aim of this paper is to construct and analyse a diffusion approximation for a diallelic multilocus reproduction model with assortative mating, recombination and mutation. Our starting point is a variant of the Moran model. We suppose that the population is monoecious[1], haploid[2] and of constant size N. This will be an overlapping generation model, but, in contrast to the usual Moran framework, we suppose that reproduction takes place at discrete times 1, 2, ... In each time step, a mating event occurs between two individuals I_1 and I_2; I_1 is replaced by an offspring, so that the size of the population is kept constant. The genotype of the offspring is obtained from those of I_1 and I_2 through a process of recombination followed by mutation which we make precise in §2. In the classical Moran model, the two individuals I_1 and I_2 are chosen at random from the population. Here, to study the effects of assortative mating, we assume that the first individual, I_1, is still chosen at random, but the second individual, I_2, is sampled with a probability that depends on its genotype and on the genotype of the first selected individual. The genotype of an individual is composed of a finite number, n, of loci with two alleles per locus denoted by 0 and 1. To characterise the assortative mating, we introduce a real parameter $s_{\boldsymbol{i},\boldsymbol{j}}$ for every pair of genotypes $(\boldsymbol{i},\boldsymbol{j})$. If I_1 has genotype $\boldsymbol{i}$, then, in the draw of I_2, an individual with genotype $\boldsymbol{j}$ has a probability proportional to $1+\frac{1}{N}s_{\boldsymbol{i},\boldsymbol{j}}$ of being selected.

Diffusion approximations for different selection-mutation models have been studied extensively in the one-locus case (see, for example, Ethier & Kurtz 1986, Chapter 10). The coefficients $1+\frac{1}{N}s_{\boldsymbol{i},\boldsymbol{j}}$ of our model play the same rôle as the (viability) selection coefficients in a Wright-Fisher model for a diploid population. Since they depend on the types of both parents they result in nonlinear (frequency dependent) selection (see §4.4). Ethier & Nagylaki (1989) study two-locus Wright-Fisher models for a panmictic[3], monoecious, diploid population of constant size under various assumptions on selection and recombination. Depending on the strength of the linkage between the two loci, they obtain different types of diffusion approximation: limiting diffusions for gametic[4] frequencies if the recombination fraction multiplied by the population size tends to a constant as the size tends to $+\infty$ (so-called *tight linkage*)

[1] Every individual has both male and female sexual organs.

[2] Each cell has one copy of each chromosome.

[3] Every individual is equally likely to mate with every other.

[4] Gametes are produced during reproduction. A gamete contains a single copy of each chromosome, composed of segments of the two chromosomes in the corresponding parent. Two gametes, one from each parent, fuse to produce an offspring.

and limiting diffusions for allelic frequencies if the recombination fraction multiplied by the population size tends to $+\infty$ (so-called *loose linkage*). To our knowledge, this work has not been extended to more general multilocus models with recombination and assortative mating. Nevertheless, there is a large body of work on multilocus genetic systems. Most theoretical investigations assume that the size of the population is infinite, so that the random genetic drift can be ignored; the evolution of genotypic frequencies is then described by recursive equations or by differential equations (see Christiansen 2000 and references within). A comparison between infinite and finite population models with random mating is presented in Baake & Herms (2008). A review of several simulation studies can be found in the introduction of Devaux & Lande (2008). Among these, the 'species formation model', introduced by Higgs & Derrida (1992) inspired our work. In their model, mating is only possible between individuals with sufficiently similar genotypes, so that from the point of view of reproduction the population is split into isolated subgroups. Their simulations display a succession of divisions and extinctions of subgroups. In this paper we generalise their assortative mating criterion to one defined through the family of parameters $s_{\boldsymbol{i},\boldsymbol{j}}$ and provide a general theoretical treatment.

To give an overview of our results, we first consider a particular pattern of assortative mating. Let us assume that the frequency of matings between two individuals of types $\boldsymbol{i}$ and $\boldsymbol{j}$ depends only on the number of loci at which their allelic types differ (and not on the positions of those loci along the genome). We then have a model with $n+1$ assortment parameters, denoted by $s_0, \ldots, s_n$, obtained by setting $s_{\boldsymbol{i},\boldsymbol{j}} = s_k$ if the genotypes $\boldsymbol{i}$, $\boldsymbol{j}$ are different at exactly k loci (regardless of their positions). This mating criterion will be called the *Hamming criterion* in what follows. A decreasing sequence $s_0 \geq s_1 \geq \ldots \geq s_n$ will describe a positive assortative mating (individuals mate preferentially with individuals that are similar). An increasing sequence $s_0 \leq s_1 \leq \ldots \leq s_n$ will describe a negative assortative mating (individuals mate preferentially with individuals that are dissimilar).

We establish a weak convergence of the Markov chain describing the genetic evolution of the population as its size tends to $+\infty$, under a hypothesis on the recombination distribution that corresponds to loose linkage (during each reproduction event, recombination between any pair of loci occurs with a positive probability) and under the assumption that mutations occur independently at each locus with the same rates (at each locus, the rate of mutation of a type 0 allele to a type 1 allele is $\frac{\mu_0}{N}$ and the rate of mutation of a type 1 to a type 0 is $\frac{\mu_1}{N}$). In particular, while mutation and assortment parameters are rescaled with population size, recombination is not. As a result, we see a *separation of timescales*. Due to recombination,

the genotypic frequencies rapidly converge to a product distribution which is characterised by its marginals, that is by the 0-allelic frequencies at each locus. We show that, at a slower rate, the allelic frequencies converge to a multidimensional diffusion, whose components are coupled only through an infinitesimal drift term (in the mathematical sense) arising from the assortative mating.

Let us describe some features of the limiting diffusion. If $s_1 - s_0 = s_2 - s_1 = \ldots = s_n - s_{n-1}$ then the frequencies of the 0-allele at each locus evolve according to independent Wright-Fisher diffusions with mutation rates μ_0 and μ_1 and symmetric balancing selection with strength $\frac{1}{2}(s_1 - s_0)$; that is they solve the following stochastic differential equation:

$$dx_t \;=\; \sqrt{x_t(1-x_t)}dW_t \;+\; \Big(\mu_1(1 \,-\, x_t) \,-\, \mu_0 x_t \,+\, (s_1 \,-\, s_0)(1/2 \,-\, x_t)x_t(1 \,-\, x_t)\Big)dt,$$

where $(W_t)_{t\geq 0}$ is a standard Brownian motion. In all other cases, the allelic frequencies at different loci no longer evolve independently. Instead the vector of 0-allelic frequencies $(x_t(1), \ldots, x_t(n))$ is governed by the stochastic differential equation:

$$\begin{aligned} dx_t(i) = \sqrt{x_t(i)(1-x_t(i))}dW_t(i) \\ + \Big(\mu_1(1-x_t(i)) - \mu_0 x_t(i) + (1/2 - x_t(i))x_t(i)(1-x_t(i))P_{i,s}(x_t)\Big)dt, \quad (1.1) \end{aligned}$$

where $(W_t(1))_{t\geq 0}, \ldots, (W_t(n))_{t\geq 0}$ denote n independent standard Brownian motions and $P(\hat{x}^{(i)})$ is a symmetric polynomial function of the $n-1$ variables $x(j)(1-x(j))$, $j \in \{1, \ldots, n\} \setminus \{i\}$ whose coefficients depend only on the parameters $s_1 - s_0, \ldots, s_n - s_{n-1}$. More precisely, $P(\hat{x}^{(i)})$ is an increasing function of each parameter $s_1 - s_0, \ldots, s_n - s_{n-1}$ (see Theorem 4.1 for an explicit formula of $P(\hat{x}^{(i)})$). When the mutation rates μ_0 and μ_1 are strictly positive, the limiting diffusion has a reversible stationary measure, the density of which is explicit. When the two mutation rates are equal to $\mu > 0$, we describe the properties of the critical points of the density of the stationary measure. In particular, we find sufficient conditions on μ and $s_1 - s_0, \ldots, s_n - s_{n-1}$ for the state where the frequencies of the two alleles are equal to $1/2$ at each locus to be a global maximum and for the stationary measure to have 2^n modes. These sufficient conditions generalise the independent case. For example, when $\mu > 1/2$ they imply the following results:

1. if $s_\ell - s_{\ell-1} \geq -(8\mu - 4)$ for every $\ell \in \{1, \ldots, n\}$, then $(1/2, \ldots, 1/2)$ is the only mode of the stationary measure;

2. if $s_n - s_{n-1} \leq \ldots \leq s_1 - s_0 < -(8\mu - 4)$ then the stationary measure has 2^n modes.

These results can be extended to other patterns of assortative mating. In fact, we need only make the following assumption on the parameters $s_{\boldsymbol{i},\boldsymbol{j}}$: the value of the assortment parameter $s_{\boldsymbol{i},\boldsymbol{j}}$ between two genotypes $\boldsymbol{i}$ and $\boldsymbol{j}$ is assumed to be the same as the value of $s_{\boldsymbol{j},\boldsymbol{i}}$ and to depend only on the loci at which $\boldsymbol{i}$ and $\boldsymbol{j}$ differ. In particular this implies that the value of $s_{\boldsymbol{i},\boldsymbol{i}}$ is the same for every genotype $\boldsymbol{i}$. This generalises the Hamming criterion and allows us to consider more realistic situations in which the influence on mating choice differs between loci (see §2.3). It transpires that, under these assumptions, the limiting diffusion does not depend on the whole family of assortment parameters, but only on one coefficient per subgroup of loci L. We denote this coefficient $m_L(s)$. It is the mean of the assortment parameters for pairs of genotypes that carry different alleles on each locus in L and identical alleles on all other loci. The stochastic differential equation followed by the limiting diffusion can still be described by equation (1.1) if the symmetric polynomial term $P(\hat{x}^{(i)})$ in the drift of the i-th coordinate is replaced by a non-symmetric polynomial term $P_i(\hat{x}^{(i)})$ in the coefficients of which the quantities $m_{L\cup\{i\}}(s) - m_L(s)$ for $L \subset \{1, \ldots, n\}$ replace $s_1 - s_0, \ldots, s_n - s_{n-1}$.

The rest of the paper is organized as follows. In §2, we present our multilocus Moran model. In §3, we describe the diffusion approximation for the one-locus model and compare it with a diffusion approximation for a population undergoing mutation and 'balancing selection'. We recall some well-known properties of this diffusion, in particular the boundary behaviour and the form of the stationary measure, for later comparison with the multilocus case. In §4, we state our main result concerning convergence to a diffusion approximation in the multilocus case (Theorem 4.1) and give two equivalent expressions for the limiting diffusion. We then compare with the two-locus diffusion approximation obtained in Ethier & Nagylaki (1989). The proof of Theorem 4.1 is postponed until §7. In §5, we derive some general properties of the limiting diffusion. §6 is devoted to the study of the density of the stationary measure. An appendix collects some technical results used in the description of the limiting diffusion.

2 The discrete model

This section is devoted to a detailed presentation of the individual based model. The assumptions on assortative mating, recombination and mutation that we will require to establish a diffusion approximation for the allelic frequencies are discussed at the end of the section.

2.1 Description of the model

We consider a monoecious and haploid population of size N where the type of each individual is described by a sequence of n diallelic loci. For the sake of brevity, let the set of loci be identified with the set of integers $[\![1\,;n]\!] := \{1,\dots,n\}$ and let the two alleles at each locus be labelled 0 and 1. The type of an individual is then identified by an n-tuple $\boldsymbol{k} := (k_1,...,k_n) \in \{0,1\}^n$. Let $\mathcal{A} = \{0,1\}^n$ be the set of possible types. The proportion of individuals of type $\boldsymbol{k}$ at time $t \in \mathbb{N}$ will be denoted by $Z_t^{(N)}(\boldsymbol{k})$ so that the composition of the population is described by the set $Z_t^{(N)} = \{Z_t^{(N)}(\boldsymbol{k}),\ \boldsymbol{k} \in \mathcal{A}\}$. At each unit of time the population evolves under the effect of assortative mating, recombination and mutation as follows.

Assortative mating: at each time t, two individuals are sampled from the population in such a way that:

1. the first individual has probability $Z_t^{(N)}(\boldsymbol{i})$ of being of type $\boldsymbol{i}$;
2. given that the first individual chosen is of type $\boldsymbol{i}$, the probability that the second individual is of type $\boldsymbol{j}$ is
$$\frac{\left(1+s_{\boldsymbol{i},\boldsymbol{j}}^{(N)}\right)Z_t^{(N)}(\boldsymbol{j})}{\sum_{\boldsymbol{k}\in\mathcal{A}}\left(1+s_{\boldsymbol{i},\boldsymbol{k}}^{(N)}\right)Z_t^{(N)}(\boldsymbol{k})},$$
where the *assortment parameters* $\{s_{\boldsymbol{i},\boldsymbol{j}}^{(N)}, \boldsymbol{i},\boldsymbol{j} \in \mathcal{A}\}$ are fixed nonnegative real numbers[5].

The population at time $t+1$ is obtained by replacing the first chosen individual with an offspring whose type is the result of the following process of recombination followed by mutation.

Recombination: for each subset L of $[\![1\,;n]\!]$, let r_L denote the probability that the offspring inherits the genes of the first chosen parent at loci $\ell \in L$ and the genes of the second parent at loci $\ell \notin L$. The family of parameters $\{r_L,\ L \subset [\![1\,;n]\!]\}$ defines a probability distribution, called the *recombination distribution*, on the power set $\mathcal{P}([\![1\,;n]\!])$ (it was first introduced in this manner by Geiringer (1944) to describe the recombination-segregation of gametes in a diploid population). It is natural to assume that the two parents contribute symmetrically to the offspring genotype, that is:

Assumption H1: *for each subset L of $[\![1\,;n]\!]$, $r_L = r_{\bar{L}}$ where $\bar{L}$ denotes the complementary set of loci, $[\![1\,;n]\!] \setminus L$.*

[5] We are allowing a small chance of self-fertilisation.

With this notation, the probability that, before mutation, the offspring of a pair of individuals of types $(\boldsymbol{i}, \boldsymbol{j})$ is of type $\boldsymbol{k}$ is

$$q((\boldsymbol{i}, \boldsymbol{j}); \boldsymbol{k}) = \sum_{L \subset [\![1;n]\!]} r_L \, \mathbb{I}_{\{\boldsymbol{k} = (\boldsymbol{i}_{|L}, \boldsymbol{j}_{|\bar{L}})\}} \, .$$

Let us express some classical examples of recombination distributions in this notation:

Examples 2.1.

1. $r_\emptyset = r_{[\![1;n]\!]} = \frac{1}{2}$ (no recombination, also called absolute linkage)
2. $r_I = 2^{-n}$ for each $I \in \mathcal{P}([\![1\,;n]\!])$ (free recombination)
3. $r_{[\![1;x]\!]} = r_{[\![x+1;n]\!]} = \frac{r}{2(n-1)}$ for $1 \leq x \leq n-1$ and $r_\emptyset = r_{[\![1;n]\!]} = 1/2(1-r)$ where r denotes an element of $]0, 1]$ (at most one exchange between the sequence of loci which occurs with equal probability at each position).

Finally we superpose mutation.

Mutation: we assume that mutations occur independently and at the same rate at each locus: $\mu_1^{(N)}$ will denote the probability that an allele 1 at a given locus of the offspring changes into allele 0 and $\mu_0^{(N)}$ the probability of the reverse mutation. The probability that the mutation process changes a type $\boldsymbol{k}$ into a type $\boldsymbol{\ell}$ is:

$$\mu^{(N)}(\boldsymbol{k}, \boldsymbol{\ell}) := \prod_{i=1}^{n} (\mu_{k_i}^{(N)})^{|\ell_i - k_i|} (1 - \mu_{k_i}^{(N)})^{1-|\ell_i - k_i|}.$$

2.2 Expression for the transition probabilities

It is now elementary to write down an expression for the transition probabilities of our model. In the notation above, if $z = \{z(\boldsymbol{k}), \boldsymbol{k} \in \mathcal{A}\}$ describes the proportion of individuals of each type in the population at a given time, then the probability that, in the next time step, the number of individuals of type $\boldsymbol{j}$ increases by one and the number of individuals of type $\boldsymbol{i} \neq \boldsymbol{j}$ decreases by one is

$$f_N(z, \boldsymbol{i}, \boldsymbol{j}) := \sum_{\boldsymbol{k}, \boldsymbol{\ell} \in \mathcal{A}} z(\boldsymbol{i}) z(\boldsymbol{k}) w^{(N)}(z, \boldsymbol{i}, \boldsymbol{k}) q((\boldsymbol{i}, \boldsymbol{k}); \boldsymbol{\ell}) \mu^{(N)}(\boldsymbol{\ell}, \boldsymbol{j})$$

where

$$w^{(N)}(z, \boldsymbol{i}, \boldsymbol{k}) = \frac{1 + s_{\boldsymbol{i}, \boldsymbol{k}}^{(N)}}{\sum_{\boldsymbol{\ell} \in \mathcal{A}} (1 + s_{\boldsymbol{i}, \boldsymbol{\ell}}^{(N)}) z(\boldsymbol{\ell})}.$$

2.3 Assumptions on assortative mating, recombination and mutation

In order to obtain a diffusion approximation for a large population, we assume that mutation and assortment parameters are both $O(N^{-1})$, so we set

Assumption H2: $\mu_\epsilon^{(N)} = \frac{\mu_\epsilon}{N}$ *for* $\epsilon \in \{0,1\}$ *and* $s_{\boldsymbol{i},\boldsymbol{j}}^{(N)} = \frac{s_{\boldsymbol{i},\boldsymbol{j}}}{N}$ *for* $\boldsymbol{i},\boldsymbol{j} \in \mathcal{A}$.

Just as in the two-locus case studied by Ethier & Nagylaki (1989), we can expect diffusion approximations to exist under two quite different assumptions on recombination, corresponding to tight and loose linkage. Here we focus on loose linkage. More precisely, we assume that the recombination distribution does not depend on the size of the population and that recombination can occur between any pair of loci:

Assumption H3: *For every* $I \in \mathcal{P}(\llbracket 1\,;n \rrbracket)$, r_I *does not depend on* N *and for any distinct integers* $h,k \in \llbracket 1\,;n \rrbracket$, *there exists a subset* $I \in \mathcal{P}(\llbracket 1\,;n \rrbracket)$ *such that* $h \in I$, $k \notin I$ *and* $r_I > 0$.

This assumption is satisfied for the last two examples of recombination distribution presented in Example 2.1, but not in the absolute linkage case. In infinite population size multilocus models with random mating, and without selection, this condition is known to ensure that in time the genotype frequencies will converge to linkage equilibrium, where they are products of their respective marginal allelic frequencies (see Geiringer 1944 and Nagylaki 1993 for a study of the evolution of multilocus linkage disequilibria under weak selection).

In order that the generator of the limiting diffusion has a tractable form, we shall make two further assumptions on the family of assortment coefficients $s = \{s_{\boldsymbol{i},\boldsymbol{j}}, (\boldsymbol{i},\boldsymbol{j}) \in \mathcal{A}^2\}$:

Assumption H4: *for every* $(\boldsymbol{i},\boldsymbol{j}) \in \mathcal{A}^2$,

1. $s_{\boldsymbol{i},\boldsymbol{j}} = s_{\boldsymbol{j},\boldsymbol{i}}$
2. *the value of* $s_{\boldsymbol{i},\boldsymbol{j}}$ *depends only on the set of loci* k *at which* $i_k = 0$ *and* $j_k = 1$ *and on the set of loci* ℓ *at which* $i_\ell = 1$ *and* $j_\ell = 0$.

These conditions mean that the probability of mating between two individuals at a fixed time depends only on the difference between their types. In particular, two individuals of the same type will have a probability of mating that does not depend on their common type: $s_{\boldsymbol{i},\boldsymbol{i}} = s_{\boldsymbol{j},\boldsymbol{j}}$ for every $\boldsymbol{i},\boldsymbol{j} \in \mathcal{A}$. In the one-locus case, this assumption means that the model distinguishes only two classes of pairs of individuals since $s_{0,1} = s_{1,0}$ and $s_{0,0} = s_{1,1}$.

In the two-locus case, this assumption leads to a model with five assortment parameters:

- one parameter, $s_{00,00} = s_{11,11} = s_{10,10} = s_{01,01}$, for pairs of individuals having the same genotype,
- one parameter $s_{00,10} = s_{10,00} = s_{11,01} = s_{01,11}$ for pairs of individuals whose genotypes only differ on the first locus,
- one parameter, $s_{00,01} = s_{01,00} = s_{11,10} = s_{10,11}$, for pairs of individuals whose genotypes only differ on the second locus,
- two parameters $s_{01,10} = s_{10,01}$ and $s_{00,11} = s_{11,00}$ for pairs of individuals whose genotypes differ on the two loci.

To describe positive or negative assortative mating we have to choose how to quantify similarities between two types. Let us present two criteria that provide assortment parameters for which assumption H4 is fulfilled:

1. **Hamming Criterion.** One simple measure to quantify similarities between two types is the number of loci with distinct alleles: $s_{\boldsymbol{i},\boldsymbol{j}}$ will be defined as nonnegative reals depending only on the Hamming distance between $\boldsymbol{i}$ and $\boldsymbol{j}$ denoted by $d_h(\boldsymbol{i},\boldsymbol{j}) := \sum_{\ell=1}^{n} |i_\ell - j_\ell|$. A positive assortative mating will be described by a sequence of $n+1$ nonnegative reals $s_0 \geq s_1 \geq \ldots \geq s_n$ by setting $s_{\boldsymbol{i},\boldsymbol{j}} = s_{d_h(\boldsymbol{i},\boldsymbol{j})}$ for every $\boldsymbol{i},\boldsymbol{j} \in \mathcal{A}$. This criterion will be called *Hamming criterion*.
2. **Additive Criterion.** If we assume that the assortment is based on a phenotypic trait which is determined by the n genes whose effects are similar and additive, then a convenient measure of the difference between individuals of type $\boldsymbol{i}$ and $\boldsymbol{j}$ is $d_a(\boldsymbol{i},\boldsymbol{j}) := |\sum_{\ell=1}^{n}(i_\ell - j_\ell)|$. A positive assortative mating will be described by a sequence of $n+1$ nonnegative reals $s_0 \geq s_1 \geq \ldots \geq s_n$ by setting $s_{\boldsymbol{i},\boldsymbol{j}} = s_{d_a(\boldsymbol{i},\boldsymbol{j})}$ for every $\boldsymbol{i},\boldsymbol{j} \in \mathcal{A}$. This criterion will be called *additive criterion*.

The assortative mating in the species formation model of Higgs & Derrida (1992) is a special case of the Hamming criterion. The additive criterion is widely used in models in which assortative mating is determined by an additive genetic trait. For example, Devaux & Lande (2008) use it to investigate speciation in flowering plants due to assortative mating determined by flowering time. Flowers can only be pollinated by other flowers that are open

at the same time. Modelling flowering time as an additive trait, they observe an effect that is qualitatively similar to that observed in the simulations of Higgs & Derrida (1992) for the Hamming criterion, namely continuous creation of reproductively isolated subgroups.

With the Hamming and additive criteria, every locus is assumed to have an identical positive or negative influence on the assortment. As we have defined a general family of assortment parameters, it is possible to consider more complex situations. For instance, we can take into account that some loci have a greater influence on the mating choice than others by dividing the set of loci into two disjoint subgroups G_1 and G_2; we introduce two sets of assortment parameters $s^{(1)}$ and $s^{(2)}$ that satisfy assumption H4 for the subgroups of loci G_1 and G_2 respectively. If we assume that the effects of the two subgroups are additive we set $s_{\boldsymbol{i},\boldsymbol{j}} = s^{(1)}_{\boldsymbol{i}_{|G_1},\boldsymbol{j}_{|G_1}} + s^{(2)}_{\boldsymbol{i}_{|G_2},\boldsymbol{j}_{|G_2}}$ for every $\boldsymbol{i},\boldsymbol{j} \in \mathcal{A}$. This defines a set of assortment parameters that satisfies assumption H4. More generally, any set of assortment parameters defined as a function of $s^{(1)}$ and $s^{(2)}$ satisfies assumption H4.

3 The one-locus diffusion approximation

Before studying the multilocus case, for later comparison, in this section we record some properties of the one-locus model.

3.1 The generator of the one-locus diffusion

In the case of one locus ($n = 1$), under assumption H2, the frequency of 0-alleles satisfies:

$$\begin{aligned}
\mathbb{E}_z[Z_1^{(N)}(0) - z] &= \frac{1}{N^2}\Big((1-z)\mu_1 - z\mu_0 \\
&\qquad + \frac{1}{2}z(1-z)((s_{1,0} - s_{1,1})(1-z) - (s_{0,1} - s_{0,0})z)\Big) + O(1/N^3) \\
\mathbb{E}_z[(Z_1^{(N)}(0) - z)^2] &= \frac{1}{N^2}z(1-z) + O(1/N^3) \\
\mathbb{E}_z[(Z_1^{(N)}(0) - z)^4] &= O(1/N^4)
\end{aligned}$$

uniformly in z.

Therefore the distribution of the frequency of 0-alleles at time $[N^2 t]$ is approximately governed,

when N is large, by a diffusion with generator:

$$\mathcal{G}_{1,s} = \frac{1}{2}x(1-x)\frac{d^2}{dx^2} + \Big((1-x)\mu_1 - x\mu_0 + 1/2x(1-x)((s_{1,0}-s_{1,1})(1-x) - (s_{0,1}-s_{0,0})x)\Big)\frac{d}{dx}. \quad (3.1)$$

More precisely, if $Z_0^{(N)}$ converges in distribution in $[0,1]$ as N tends to $+\infty$, then $(Z^{(N)}_{[N^2t]})_{t\geq 0}$ converges in distribution in the Skorohod space of càdlàg functions $D_{[0,1]}([0,+\infty))$ to a diffusion with generator $\mathcal{G}_{1,s}$ (see, for example, Ethier & Kurtz 1986, Chapter 10).

If we assume that s satisfies assumption H4, that is $s_{0,0} = s_{1,1}$ and $s_{0,1} = s_{1,0}$, and if we denote their common values by s_0 and s_1 respectively, then the drift has a simpler form and we obtain

$$\mathcal{G}_{1,s} = \frac{1}{2}x(1-x)\frac{d^2}{dx^2} + \Big((1-x)\mu_1 - x\mu_0 + (s_1-s_0)(1/2-x)x(1-x)\Big)\frac{d}{dx}.$$

Remark 3.1. This diffusion can also be obtained as an approximation of a diploid model with random mating, mutation and weak selection in favour of homozygosity[6] (when $s_0 - s_1 > 0$) or in favour of heterozygosity (when $s_0 - s_1 < 0$) (see, for example, Ethier & Kurtz 1986, Chapter 10).

3.2 Properties of the one-locus diffusion

Stationary measure. If μ_0 and μ_1 are strictly positive, this diffusion has a reversible stationary measure. Its density with respect to Lebesgue measure on $[0,1]$ is given by Wright's formula:

$$g_{\mu,s}(x) = C_{\mu,s}x^{2\mu_1-1}(1-x)^{2\mu_0-1}\exp\Big(-1/2((s_{1,0}-s_{1,1})(1-x)^2 + (s_{0,1}-s_{0,0})x^2)\Big)$$

where the constant $C_{\mu,s}$ is chosen so that $\int_0^1 g_{\mu,s}(x)dx = 1$. This is plotted, for various parameter values, in Fig. 1 under the assumptions $\mu_1 = \mu_0 = \mu$, $s_{0,0} = s_{1,1} = s_0$ and $s_{0,1} = s_{1,0} = s_1$.

Boundary behaviour. According to the Feller boundary classification for one-dimensional diffusions (see e.g. Ethier & Kurtz 1986):

[6] A diploid individual is homozygous at a gene locus when its cells contain two identical alleles at the locus.

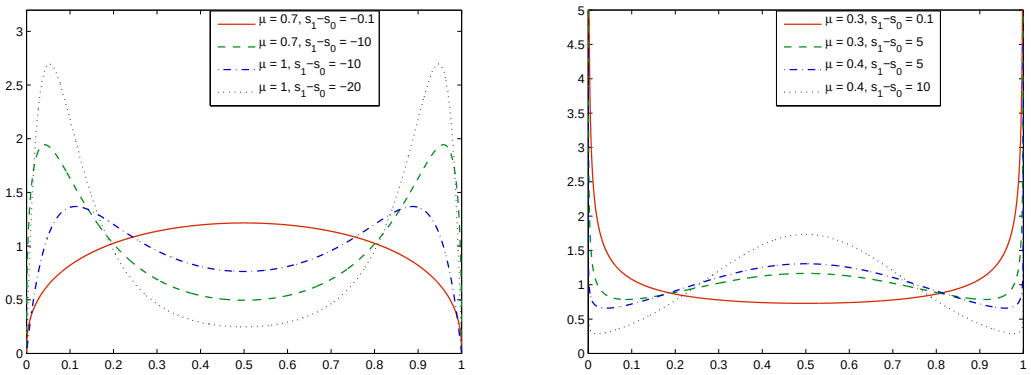


Figure 1: Representation of the invariant density $g_{\mu,s}$ for the one-locus diffusion when the two mutations rates $\mu_1 = \mu_0 = \mu$, $s_{0,0} = s_{1,1} = s_0$ and $s_{0,1} = s_{1,0} = s_1$. In the figure on the left, $\mu > 1/2$ and matings between individuals of the same allelic type are favoured. The density is bimodal if and only if $s_0 - s_1 > 8\mu - 4$. In the figure on the right, $0 < \mu < 1/2$ and matings between individuals of different allelic types are favoured. The density tends to $+\infty$ at the boundaries and has a global minimum at $1/2$ if and only if $s_1 - s_0 \leq 4 - 8\mu$.

(i) if $\mu_1 = 0$ then 0 is an absorbing state and the diffusion exits from $]0,1[$ in a finite time almost surely;

(ii) if $\mu_1 \geq 1/2$ then 0 is an entrance boundary (started from a point in $]0,1[$ the diffusion will not reach 0 in finite time, but the process started from 0 is well-defined);

(iii) if $0 < \mu_1 < 1/2$ then 0 is a regular boundary (starting from a point $z_0 \in]0,1[$ the diffusion has a positive probability of reaching 0 before any point $b \in]z_0, 1]$ in a finite time and the diffusion started from 0 is well-defined);

with the obvious symmetric definitions at 1.

4 Convergence to a diffusion in the multilocus case

In the case of several loci, under assumptions H2 and H3, a Taylor expansion shows that the drift $\mathbb{E}[Z_{t+1}^{(N)}(\boldsymbol{i}) - Z_t^{(N)}(\boldsymbol{i}) \mid Z_t^{(N)} = z]$ is of order $\frac{1}{N^2}$ only inside the set of product distributions on $\{0,1\}^n$. This set is often called the *Wright manifold* or the *linkage-equilibrium manifold* and denoted by $\mathcal{W}_n$ (a population is said to be in linkage equilibrium if the geno-

type distribution z is in $\mathcal{W}_n$, that is if $z(\boldsymbol{i}) = z_1(i_1)\cdots z_n(i_n)\ \forall \boldsymbol{i} = (i_1,\ldots,i_n) \in \mathcal{A}$ where $z_j(x) = \sum_{\boldsymbol{k}\in\{0,1\}^{n-1}} z(k_1,\ldots,k_{j-1},x,k_{j+1},\ldots,k_n)$ denotes the frequency of individuals having the allele x at the j-th locus).
Outside this manifold, the drift pushes the process towards the Wright manifold at an exponential speed. Therefore to extend the diffusion approximation to the n-locus case, we introduce a change of variables composed of the n 0-allelic frequencies and of $2^n - n - 1$ processes that measure the deviation from the linkage equilibrium.
For a nonempty subset L of $\{1,\ldots,n\}$,

- let $X_t^{(N)}(L) = \sum_{\boldsymbol{j}\in\mathcal{A},\ \boldsymbol{j}_{|L}\equiv 0} Z_t^{(N)}(\boldsymbol{j})$ denote the proportion of individuals having the allele 0 on all loci in L at time t;
- let $Y_t^{(N)}(L) = \prod_{i\in L} X_t^{(N)}(\{i\}) - X_t^{(N)}(L)$ for $|L| \geq 2$ describe the linkage disequilibrium between the loci in L at time t. (This is just one of many ways to measure the linkage disequilibrium, see for example Bürger (2000), Chapter V.4.2, for other measures.)

The vector of 0-allelic frequencies at time t is $X_t^{(N)} := \left(X_t^{(N)}(\{1\}),...,X_t^{(N)}(\{n\})\right)$.
The process $Y^{(N)}$ defined by $Y_t^{(N)} := \left\{Y_t^{(N)}(L),\ L \subset [\![1\,;n]\!] \text{ such that } |L| \geq 2\right\}$ for $t \geq 0$ vanishes on the Wright manifold.
We shall show that if t_N tends to $+\infty$ faster than N then $Y_{[t_N]}^{(N)}$ converges to 0 while if time is sped up by N^2 then $X^{(N)}$ converges to a diffusion as N tends to $+\infty$.
Before giving a precise statement of the convergence result for the two processes $X^{(N)}$ and $Y^{(N)}$ (Theorem 4.1), let us introduce some notation in which to express the parameters of the limiting diffusion.

4.1 Mean assortment parameters

For a subset L of loci, consider the set of pairs of genotypes that differ at each locus $\ell \in L$ and are equal at each locus $\ell \notin L$:

$$F_L = \left\{(\boldsymbol{i},\boldsymbol{j}) \in \mathcal{A}^2 :\ i_u = 1 - j_u\ \forall u \in L \text{ and } i_u = j_u\ \forall u \in \bar{L}\right\}.$$

Let $m_L(s)$ denote the mean value of the assortment parameters for all pairs in this set F_L:

$$m_L(s) = 2^{-n} \sum_{(\boldsymbol{i},\boldsymbol{j})\in F_L} s_{\boldsymbol{i},\boldsymbol{j}}.$$

Examples 4.1.

1. In the two-locus case,

$$\begin{aligned} m_{\emptyset}(s) &= \frac{1}{4}(s_{00,00} + s_{01,01} + s_{10,10} + s_{11,11}), \\ m_{\{1\}}(s) &= \frac{1}{4}(s_{00,10} + s_{10,00} + s_{01,11} + s_{11,01}), \end{aligned}$$

$$m_{\{2\}}(s) = \frac{1}{4}(s_{00,01} + s_{01,00} + s_{11,10} + s_{10,11}).$$

   In each of these expressions the four coefficients are equal by assumption H4.

$$m_{\{1,2\}}(s) = \frac{1}{4}(s_{00,11} + s_{11,00} + s_{01,10} + s_{10,01}).$$

   In this expression the first (resp. last) two coefficients are equal by H4.

2. With the Hamming criterion, $m_L(s) = s_{|L|}$ for every $L \subset [\![1\,;n]\!]$, where $|L|$ denotes the number of loci in L.

3. With the additive criterion, $m_{\emptyset}(s) = s_0$, $m_{\{\ell\}}(s) = s_1$ $\forall \ell \in [\![1\,;n]\!]$ and more generally $m_L(s) = 2^{-|L|} \sum_{k=0}^{|L|} \binom{|L|}{k} s_{|2k-|L||}$ for every $L \subset [\![1\,;n]\!]$.

4.2 Convergence to a diffusion

The following theorem provides convergence results for the two processes $X^{(N)}$ and $Y^{(N)}$ as the population size N tends to $+\infty$. The proof, based on Theorem 3.3 of Ethier & Nagylaki (1980), is postponed to §7.

Theorem 4.1. *Assume that hypotheses H1, H2, H3 and H4 hold.*

(a) For $i \in [\![1\,;n]\!]$ let $P_{i,s}(x)$ denote a polynomial function in the $n-1$ variables $x_k(1-x_k)$ for $k \in [\![1\,;n]\!] \setminus \{i\}$. Then, the operator

$$\mathcal{G}_{n,s} = \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial_{x_i}\partial_{x_i}} + \sum_{i=1}^{n}\Big((1-x_i)\mu_1 - x_i\mu_0 + (1/2-x_i)x_i(1-x_i)P_{i,s}(x)\Big)\frac{\partial}{\partial_{x_i}} \quad (4.1)$$

with domain $\mathcal{D}(\mathcal{G}_{n,s}) = C^2([0,1]^n)$ is closable in $C([0,1]^n)$ and its closure is the generator of a strongly continuous semigroup of contractions.

(b) If $X_0^{(N)}$ converges in distribution in $[0,1]^n$, then $(X^{(N)}_{[N^2t]})_t$ converges in distribution in the Skorohod space of càdlàg functions $D_{[0,1]^n}([0,\infty))$ to a diffusion process X with generator $\mathcal{G}_{n,s}$ where the polynomial function $P_{i,s}(x)$ has the following expression:

$$P_{i,s}(x) = \sum_{A\in\mathcal{P}(\llbracket 1;n\rrbracket\setminus\{i\})} \big(m_{A\cup\{i\}}(s) - m_A(s)\big) \prod_{k\in A}(2x_k(1-x_k)) \prod_{\ell\in\llbracket 1;n\rrbracket\setminus\{A\cup\{i\}\}} (1-2x_\ell(1-x_\ell)). \quad (4.2)$$

(c) For every positive sequence $(t_N)_N$ that converges to $+\infty$, $Y^{(N)}_{[Nt_N]}$ converges in distribution to 0.

Remark 4.1.

1. The recombination distribution $(r_I)_{I\subset\llbracket 1;n\rrbracket}$ does not appear in the expression for the limiting diffusion. Nevertheless, the proof of Theorem 4.1 will show that it has an influence on the speed of convergence of the linkage disequilibrium to 0.

2. The limiting diffusion depends on the assortment parameters only via the quantities $m_A(s)$ for every $A \subset \llbracket 1\,;n\rrbracket$. A set of assortment parameters for which

$$m_{A\cup\{i\}}(s) - m_A(s) < 0 \text{ for every } i \in \llbracket 1\,;n\rrbracket \text{ and } A \subset \llbracket 1\,;n\rrbracket \setminus \{i\}$$

favours homozygous mating with respect to the genotype at the i-th locus. It is therefore no surprise that by increasing the value of $m_{A\cup\{i\}}(s) - m_A(s)$ for a fixed subset A, we increase the value of the i-th coordinate of the drift at a point x for which $x_i < 1/2$ and decrease it at a point x for which $x_i > 1/2$.

4.3 Another expression for the polynomial term $P_{i,s}(x)$ of the drift

An expansion of the polynomial function $P_{i,s}(x)$ in terms of the variables $x_k(1-x_k)$, $k \in \llbracket 1\,;n\rrbracket \setminus \{i\}$ yields the following expression:

$$P_{i,s}(x) = \sum_{L\in\mathcal{P}(\llbracket 1;n\rrbracket\setminus\{i\})} \alpha_{i,L}(s) \prod_{\ell\in L} x_\ell(1-x_\ell) \quad (4.3)$$

with

$$\alpha_{i,L}(s) = 2^{|L|} \sum_{A\subset L} (-1)^{|L|-|A|}(m_{A\cup\{i\}}(s) - m_A(s)). \quad (4.4)$$

The details of the proof are provided in §7.2.

The coefficients $\alpha_{i,L}(s)$ can be compactly expressed using difference operators. Let us introduce some notation: for a function f defined on the subsets of a finite set E and for an element i of E, we denote by $\delta_i[f]$ the function on $\mathcal{P}(E)$ defined by

$$\delta_i[f](A) = f(A \cup \{i\}) - f(A), \ \forall A \in \mathcal{P}(E).$$

Since $\delta_i \circ \delta_j = \delta_j \circ \delta_i$ for every $i, j \in E$, we can, more generally, introduce a difference operator δ_B for each subset $B \in \mathcal{P}(E)$ by setting $\delta_\emptyset = Id$, and $\delta_B = \delta_{b_1} \circ \cdots \circ \delta_{b_r}$ if $B = \{b_1, \ldots, b_r\}$. A proof by induction on $|B|$ provides the following formula for δ_B:

$$\delta_B[f](A) = \sum_{J \subset B} (-1)^{|B|-|J|} f(A \cup J) \quad \forall A \subset E. \tag{4.5}$$

Let $m(s)$ denote the function $A \mapsto m_A(s)$ defined on the subsets of $[\![1\,;n]\!]$. In this notation, for every $A \subset [\![1\,;n]\!] \setminus \{i\}$,

$$m_{A\cup\{i\}}(s) - m_A(s) = \delta_i[m(s)](A) \text{ and } \alpha_{i,A}(s) = 2^{|A|}\delta_{A\cup\{i\}}[m(s)](\emptyset). \tag{4.6}$$

If, for each subset A of loci, the coefficient $m_A(s)$ depends only on the number of loci in A, then it follows from expression (4.3) that $P_{i,s}(x)$ is a symmetric polynomial function, the coefficients of which do not depend on i. This is the case for instance with the Hamming and additive criteria (see Example 4.1 for the corresponding expressions for $m_A(s)$). Let us give the expanded form of $P_{i,s}(x)$ for the Hamming criterion:

$$P_{i,s}(x) = \sum_{\ell=0}^{n-1} \tilde{\alpha}_\ell \sum_{L \subset [\![1;n]\!]\setminus\{i\},\ |L|=\ell} \prod_{\ell \in L} x_\ell(1-x_\ell). \tag{4.7}$$

where $\tilde{\alpha}_k(s) = 2^k \sum_{\ell=0}^{k} (-1)^\ell \binom{k}{\ell} (s_{k-\ell+1} - s_{k-\ell})$.
As in the general case, the coefficient $\tilde{\alpha}_k(s)$ has a compact expression in terms of difference operators. Let $\delta^{(1)}$ denote the forward difference operators: $\delta^{(1)}[s](i) = s_{i+1} - s_i$ for every $i \in [\![0\,;n-1]\!]$. The forward difference operators of higher orders are defined iteratively: $\delta^{(k+1)}[s] = \delta^{(k)} \circ \delta^{(1)}[s]$ for $k \in \mathbb{N}^*$. With this notation, $\tilde{\alpha}_k(s) = 2^k \delta^{(k+1)}[s](0)$ for $k \in [\![0\,;n-1]\!]$.

4.4 Comparison with the two-locus Wright-Fisher diffusion

Ethier & Nagylaki (1989) established convergence results for a general multiallelic two-locus Wright-Fisher model of a panmictic, monoecious, diploid population of N individuals (identified with $2N$ haploids) undergoing mutation and selection. In their model, a gamete is described by a pair $\boldsymbol{i} = (i_1, i_2) \in [\![1\,;r_1]\!] \times [\![1\,;r_2]\!]$ where r_1 is the number of alleles in the first locus and r_2 is the number of alleles in the second locus. The parameters of their model are:

1. the viability of a pair of gametes $(\boldsymbol{i},\boldsymbol{j})$ denoted by $w_{N,\boldsymbol{i},\boldsymbol{j}} = 1-\sigma_{N,\boldsymbol{i},\boldsymbol{j}}$ with the assumption $\sigma_{N,\boldsymbol{i},\boldsymbol{j}} = \sigma_{N,\boldsymbol{j},\boldsymbol{i}}$ and $\sigma_{N,\boldsymbol{i},\boldsymbol{i}} = 0$ for every $\boldsymbol{i},\boldsymbol{j} \in [\![1\,;r_1]\!] \times [\![1\,;r_2]\!]$ (after viability selection the proportion of a pair of gametes $(\boldsymbol{i},\boldsymbol{j})$ is assumed to be $P^*_{\boldsymbol{i},\boldsymbol{j}} = \dfrac{w_{N,\boldsymbol{i},\boldsymbol{j}} P_{\boldsymbol{i}} P_{\boldsymbol{j}}}{\sum_{\boldsymbol{k},\boldsymbol{\ell}} w_{N,\boldsymbol{k},\boldsymbol{\ell}} P_{\boldsymbol{k}} P_{\boldsymbol{\ell}}}$ if $P_{\boldsymbol{k}}$ denotes the frequency of gametes $\boldsymbol{k}$ in the population $\forall \boldsymbol{k} \in [\![1\,;r_1]\!] \times [\![1\,;r_2]\!]$);
2. the recombination fraction c_N;
3. the probability $(2N)^{-1}\nu^{(i)}_{j,k}$ that the j-th allele in the i-th locus mutates to the k-th allele.

The population at the generation $t+1$ is obtained by choosing $2N$ gametes uniformly at random with replacement from the pool of gametes of the generation t after the steps of viability selection, recombination and mutation.

They studied the diffusion approximation under several assumptions on selection and recombination coefficients. In the case of weak selection ($2N\sigma_{N,\boldsymbol{i},\boldsymbol{j}}$ converges to a real number denoted by $\sigma_{\boldsymbol{i},\boldsymbol{j}}$ for every $\boldsymbol{i},\boldsymbol{j}$) and loose linkage ($c_N$ converges to a finite limit and Nc_N tends to $+\infty$) they obtained a limiting diffusion for the allelic frequencies $(p_1,\dots,p_{r_1-1},q_1,\dots,q_{r_2-1})$ of the alleles $1,\dots,r_1-1$ in the first locus and the alleles $1,\dots,r_2-1$ in the second locus. In the case of two alleles at each locus ($r_1=r_2=2$), the generator of the limiting diffusion is

$$\mathcal{L} = \frac{1}{2}p_1(1-p_1)\partial^2_{p_1,p_1} + \frac{1}{2}q_1(1-q_1)\partial^2_{q_1,q_1} + b_1(p_1,q_1)\partial_{p_1} + b_2(p_1,q_1)\partial_{q_1}$$

with

$$\begin{aligned}
b_1(p_1,q_1) =& \nu^{(1)}_{2,1}(1-p_1) - \nu^{(1)}_{1,2}p_1 \\
&- p_1(1-p_1)(1-2p_1)\Big((\sigma_{12,21}+\sigma_{11,22})q_1(1-q_1) + \sigma_{11,21}q_1^2 + \sigma_{12,22}(1-q_1)^2\Big) \\
&- 2p_1(1-p_1)q_1(1-q_1)\Big(\sigma_{11,12}p_1 - \sigma_{21,22}(1-p_1)\Big). \\
b_2(p_1,q_1) =& \nu^{(2)}_{2,1}(1-q_1) - \nu^{(2)}_{1,2}q_1 \\
&- q_1(1-q_1)(1-2q_1)\Big((\sigma_{12,21}+\sigma_{11,22})p_1(1-p_1) + \sigma_{11,12}p_1^2 + \sigma_{21,22}(1-p_1)^2\Big) \\
&- 2q_1(1-q_1)p_1(1-p_1)\Big(\sigma_{11,21}q_1 - \sigma_{12,22}(1-q_1)\Big).
\end{aligned}$$

Accordingly, the generator $\mathcal{L}$ coincides with $\mathcal{G}_{2,s}$ if we assume

(a) that the mutation rates $\nu^{(i)}_{j,k}$ do not depend on the locus i and set $\nu^{(i)}_{1,2} = \mu_0$ and $\nu^{(i)}_{2,1} = \mu_1$,

(b) that the coefficients of selection satisfy $\sigma_{11,21} = \sigma_{12,22}$ and $\sigma_{11,12} = \sigma_{21,22}$ (second condition of assumption H4)

and set $\sigma_{\boldsymbol{i},\boldsymbol{j}} = -\frac{1}{2}s_{\boldsymbol{i}-\mathbf{1},\boldsymbol{j}-\mathbf{1}}$, for every $\boldsymbol{i},\boldsymbol{j} \in \{1,2\}^2$ (with the notation $\mathbf{1} = (1,\dots,1)$).
This comparison suggests that the effect of assortative mating on the genotype evolution of a large population in our model is similar to the effect of weak viability selection in a diploid Wright-Fisher model with mutation.

5 Description of the limiting diffusion

This section collects some properties that can be deduced from the form of the generator, $\mathcal{G}_{n,s}$, of the limiting diffusion.

5.1 The set of generators arising from the model

Lemma 5.1. *Any generator on $C^2([0,1]^n)$ of the form*

$$\frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial_{x_i}\partial_{x_i}} + \sum_{i=1}^{n}\Big((1-x_i)\mu_1 - x_i\mu_0 + (1/2-x_i)x_i(1-x_i)\sum_{L\in\mathcal{P}(\llbracket 1;n\rrbracket\setminus\{i\})}\alpha_{L\cup\{i\}}\prod_{k\in L}x_k(1-x_k)\Big)\frac{\partial}{\partial_{x_i}},$$

where $\{\alpha_A,\ A\subset\llbracket 1\,;n\rrbracket,\ A\neq\emptyset\}$ is a family of real numbers, can be interpreted as the generator of the diffusion approximation of an n-locus Moran model as defined in §2.

Proof. We may, for instance, take the following set of assortment parameters $\{s_{\boldsymbol{i},\boldsymbol{j}},\ \boldsymbol{i},\boldsymbol{j}\in\mathcal{A}\}$:

- $s_{\boldsymbol{i},\boldsymbol{i}} = 0$ for every $\boldsymbol{i}\in\mathcal{A}$.
- $s_{\boldsymbol{i},\boldsymbol{j}} = \sum_{B\subset L,\ |B|\geq 1} 2^{-|B|+1}\alpha_B$ for every $(\boldsymbol{i},\boldsymbol{j})\in F_L$ and for every nonempty subset L of $\llbracket 1\,;n\rrbracket$.

Let us check that this family satisfies $2^{|L|-1}\delta_L[m(s)](\emptyset) = \alpha_L$ for every nonempty subset L of $\llbracket 1\,;n\rrbracket$. First, $m_L(s) = \sum_{B\subset L,\ |B|\geq 1} 2^{-|B|+1}\alpha_B$. For every $i\in\llbracket 1\,;n\rrbracket$ and $L\subset\llbracket 1\,;n\rrbracket\setminus\{i\}$

$$\delta_{L\cup\{i\}}[m(s)](\emptyset) = \sum_{A\subset L}(-1)^{|L|-|A|}(m_{A\cup\{i\}}(s) - m_A(s)) = \sum_{A\subset L}(-1)^{|L|-|A|}\sum_{B\subset A}2^{-|B|}\alpha_{B\cup\{i\}}.$$

We invert the double sum and use the formula $\displaystyle\sum_{A\subset L,\text{ s. t. }B\subset A}(-1)^{|L|-|A|} = \mathbb{1}_{\{L=B\}}$ to obtain:

$$\delta_{L\cup\{i\}}[m(s)](\emptyset) = \sum_{B\subset L}2^{-|B|}\alpha_{B\cup\{i\}}\,\mathbb{1}_{\{L=B\}} = 2^{-|L|}\alpha_{L\cup\{i\}}.$$

□

In particular, the n-locus Moran model with assortative mating based on the Hamming criterion allows us to obtain, through diffusion approximation, any generator on $C^2([0,1]^n)$ of the form:

$$\frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\frac{\partial^2}{\partial_{x_i}\partial_{x_i}} + \sum_{i=1}^{n}\Big((1-x_i)\mu_1 - x_i\mu_0 + (1/2-x_i)x_i(1-x_i)\sum_{\ell=0}^{n-1}\alpha_\ell \sum_{\substack{L\subset[\![1;n]\!]\setminus\{i\}\\ \text{s.t.}|L|=\ell}}\prod_{k\in L} x_k(1-x_k)\Big)\frac{\partial}{\partial_{x_i}}.$$

To see this, given any sequence $\alpha_0,\dots,\alpha_{n-1}$ of n reals, we have to find $n+1$ real numbers $s_0,\dots,s_n$ such that $\alpha_\ell = 2^\ell\delta^{(\ell+1)}[s](0)$. These are given by the inversion formula (A.3) in the Appendix, from which we see that we may set $s_0=0$ and $s_k=\sum_{\ell=1}^{k}2^{1-\ell}\binom{k}{\ell}\alpha_{\ell-1}$ for $k\in[\![1\,;n]\!]$.

5.2 The generator for two groups of loci

Let us consider a partition of the set of loci into two subgroups, $G_1 = [\![1\,;k]\!]$ and $G_2=[\![k+1\,;n]\!]$, say. We introduce two sets of assortment parameters $s^{(1)}$ and $s^{(2)}$ depending on subgroups of loci from G_1 and from G_2 respectively and satisfying assumption H4. If we assume that the assortment parameter between two individuals of type $\boldsymbol{i}$ and $\boldsymbol{j}$ is $s_{\boldsymbol{i},\boldsymbol{j}} = s^{(1)}_{\boldsymbol{i}_{|G_1},\boldsymbol{j}_{|G_1}} + s^{(2)}_{\boldsymbol{i}_{|G_2},\boldsymbol{j}_{|G_2}}$ for every $\boldsymbol{i},\boldsymbol{j}\in\mathcal{A}$, then $m_L(s) = m_{L\cap G_1}(s^{(1)}) + m_{L\cap G_2}(s^{(2)})$ for every subset L of $[\![1\,;n]\!]$. This implies that the first k coordinates of diffusion limit evolve independently of the last $n-k$ coordinates and that the generator of the diffusion limit is:

$$\mathcal{G}_{n,s} = \mathcal{G}_{k,s_1}\otimes\mathcal{G}_{n-k,s_2}.$$

Therefore, with these choices we can reduce our study to subgroups of loci having the same influence on assortment.

5.3 Conditions for independent coordinates

For some patterns of assortment, the allelic frequencies at each locus in a large population evolve approximately as independent diffusions:

Proposition 5.1. *Assume that the assortment parameters $s=\{s_{\boldsymbol{i},\boldsymbol{j}},\ \boldsymbol{i},\boldsymbol{j}\in\mathcal{A}\}$ satisfy the assumption H4.*

1. *The n coordinates of the diffusion associated with the generator $\mathcal{G}_{n,s}$ are independent if and only if the following condition holds:*

(H5) for every $i \in [\![1\,;n]\!]$, $m_{L\cup\{i\}}(s) - m_L(s)$ does not depend on the choice of the subset L of $[\![1\,;n]\!] \setminus \{i\}$.

2. *If condition (H5) holds, the i-th coordinate behaves as the one-locus diffusion with assortment coefficients $s_0 = s_{\mathbf{1},\mathbf{1}}$ and $s_1 = s_{u_i,\mathbf{1}}$ where $u_i = (0_{\{i\}}, \mathbf{1}_{[\![1;n]\!]\setminus\{i\}})$ denotes the genotype which differs from the genotype $\mathbf{1}$ only on the locus i; its generator is*

$$\frac{1}{2}x(1-x)\frac{d^2}{dx^2} + \Big((1-x)\mu_1 - x\mu_0 + (s_{u_i,\mathbf{1}} - s_{\mathbf{1},\mathbf{1}})(1/2-x)x(1-x)\Big)\frac{d}{dx}.$$

3. *In particular,*

   (a) *with the Hamming criterion, $\mathcal{G}_{n,s}$ is the generator of n independent one-dimensional diffusions if and only if the value of $s_{\ell+1} - s_\ell$ does not depend on ℓ;*

   (b) *with the additive criterion, $\mathcal{G}_{n,s}$ is the generator of n independent one-dimensional diffusions if and only if there exists a constant c such that $s_{\ell+1} - s_\ell = c(2\ell+1)$ for every $\ell \in [\![0\,;n-1]\!]$.*

Proof. First note that $\mathcal{G}_{n,s}$ is the generator of n independent diffusions if and only if the polynomial term $P_{i,s}(x)$ is a constant function for every $i \in [\![1\,;n]\!]$.

1. According to the formula (4.2), the polynomial term $P_{i,s}(x)$ is a constant function for every $i \in [\![1\,;n]\!]$ whenever condition H5 holds. Conversely, assume that the polynomial term $P_{i,s}(x)$ is a constant function for every $i \in [\![1\,;n]\!]$. By formulae (4.3) and (4.6), $\delta_L[m(s)](\emptyset) = 0$ for every subset L of $[\![1\,;n]\!]$ having at least two elements. We derive, from the inversion formula (A.2) stated in the Appendix, that for every subset $A \in \mathcal{P}([\![1\,;n]\!])$,

$$\delta_i[m(s)](A) = \sum_{B\subset A} \delta_{B\cup\{i\}}[m(s)](\emptyset) = \delta_i[m(s)](\emptyset).$$

   Therefore, condition H5 is satisfied.

2. With the Hamming criterion, $m_A(s) = s_{|A|}$ and condition H5 is equivalent to

$$s_{k+1} - s_k = s_1 - s_0 \text{ for every } k \in [\![1\,;n-1]\!].$$

3. With the additive criterion, for a subset L with ℓ elements $m_L(s) = 2^{-\ell}\sum_{j=0}^{\ell}\binom{\ell}{j}s_{|2j-\ell|}$.

After some computation, we obtain for $i \notin L$,

$$m_{L\cup\{i\}}(s) - m_L(s) = 2^{-\ell} \sum_{j=0}^{\ell} \binom{\ell}{j} (s_{|2j-\ell+1|} - s_{|2j-\ell|})$$
$$= \begin{cases} 2^{-\ell} \sum_{j=1}^{\frac{\ell+1}{2}} \binom{\ell}{\frac{\ell+1}{2}-j} \delta^{(2)}[s](2j-2) & \text{if } \ell \text{ is odd,} \\ 2^{-\ell} \Big(\sum_{j=1}^{\frac{\ell}{2}} \binom{\ell}{\frac{\ell}{2}-j} \delta^{(2)}[s](2j-1) + \binom{\ell}{\frac{\ell}{2}} \delta^{(1)}[s](0) \Big) & \text{if } \ell \text{ is even.} \end{cases} \tag{5.1}$$

It follows from (5.1) that for every $c \in \mathbb{R}$, the system defined by

$$m_{L\cup\{i\}}(s) - m_L(s) = c \text{ for every } i \in [\![1\,;n]\!] \text{ and } L \subset [\![1\,;n]\!] \setminus \{i\}$$

has a unique solution which is $\delta^{(1)}[s](k) = c(2k+1)$ for every $k \in [\![0\,;n-1]\!]$.

□

5.4 Behaviour at the boundaries

In this section the trajectories of the coordinates of the limiting diffusion are compared with those of one-dimensional diffusions in order to investigate whether an allele can be (instantaneously) fixed at one of the loci.

Consider the stochastic differential equations associated with the generator $\mathcal{G}_{n,s}$:

$$dx_t(i) = \sqrt{x_t(i)(1-x_t(i))}dW_t(i) + b_i(x_t)dt \quad \forall i \in [\![1\,;n]\!], \tag{5.2}$$

where $(W_t(1))_{t\geq 0}, \ldots, (W_t(n))_{t\geq 0}$ denote n independent standard Brownian motions, and

$$b_i(x) = \mu_1(1-x(i)) - \mu_0 x(i) + (1/2 - x(i))x(i)(1-x(i))P_{i,s}(x) \text{ for } i \in [\![1\,;n]\!].$$

Theorem 1 of Yamada & Watanabe (1971) ensures pathwise uniqueness for the stochastic differential equation (5.2), since the drift is Lipschitz and the diffusion matrix is a diagonal matrix of the form

$$\sigma(x) = \mathrm{diag}(\sigma_1(x(1)), \ldots, \sigma_n(x(n))),$$

where the functions σ_i are 1/2-Hölder continuous functions.

The following proposition shows that, just as for the one-locus case, the boundary behaviour of the solution to (5.2) depends only on the values of the mutation rates μ_0 and μ_1.

Proposition 5.2. *Let $(x_t)_{t\geq 0}$ denote a solution of the stochastic differential equation (5.2) starting from a point $x_0 \in]0,1[^n$.*

(i) If $\mu_1 = \mu_0 = 0$ then the diffusion process $(x_t)_t$ exits from $]0,1[^n$ in a finite time almost surely.

(ii) If $\mu_1 = 0$ and $\mu_0 > 0$ then each coordinate of $(x_t)_t$ reaches the point 0 in a finite time almost surely.

(iii) If $0 < \mu_1 < 1/2$ then 0 is attainable for each coordinate of the diffusion process:

$$\mathbb{P}[\exists t > 0,\ x_t(i) = 0] > 0 \quad \forall i \in [\![1\,;n]\!].$$

(iv) If $\mu_1 \geq 1/2$ then 0 is inaccessible for each coordinate of the diffusion process:

$$\mathbb{P}[\exists t > 0,\ x_t(i) = 0] = 0 \text{ and } \mathbb{P}[\lim_{t\to+\infty} x_t(i) = 0] = 0 \text{ for every } i \in [\![1\,;n]\!].$$

Similar statements to (ii), (iii) and (iv) hold for the point 1 on exchanging the rôles of μ_1 and μ_0.

Proof. Let $i \in [\![1\,;n]\!]$. On $[0,1]^n$ the polynomial function $P_{i,s}$ is bounded above by

$$M_i^+ = \sum_{A\subset[\![1;n]\!]\setminus\{i\}} 2^{-|A|} \max\big\{m_{A\cup\{i\}}(s) - m_A(s), 0\big\}$$

and is bounded below by

$$M_i^- = -\sum_{A\subset[\![1;n]\!]\setminus\{i\}} 2^{-|A|} \max\big\{-(m_{A\cup\{i\}}(s) - m_A(s)), 0\big\}.$$

Let b_i^+ and b_i^- denote the functions defined on $[0,1]$ by

$$\begin{aligned} b_i^+(u) &= \mu_1(1-u) - \mu_0 u + (1/2-u)u(1-u)(M_i^+ \mathbb{1}_{\{u<1/2\}} + M_i^- \mathbb{1}_{\{u>1/2\}}),\\ b_i^-(u) &= \mu_1(1-u) - \mu_0 u + (1/2-u)u(1-u)(M_i^+ \mathbb{1}_{\{u>1/2\}} + M_i^- \mathbb{1}_{\{u<1/2\}}), \end{aligned}$$

for every $u \in [0,1]$. For every $i \in [\![1\,;n]\!]$, pathwise uniqueness holds for the following two stochastic differential equations:

$$du_t = \sqrt{u_t(1-u_t)}dW_t(i) + b_i^+(u_t)dt \tag{5.3}$$

and

$$du_t = \sqrt{u_t(1-u_t)}dW_t(i) + b_i^-(u_t)dt. \tag{5.4}$$

Let $\xi_t^+(i)$ and $\xi_t^-(i)$ be the solution starting from $x_0(i)$ of the stochastic differential equations (5.3) and (5.4) respectively. As the i-th coordinate of the drift is bounded above by b_i^+ and is bounded below by b_i^-, the comparison theorem of Ikeda & Watanabe (1977) ensures that the following inequalities hold with probability one:

$$\xi_t^-(i) \leq x_t(i) \leq \xi_t^+(i) \quad \forall t \geq 0,\ \forall i \in [\![1\,;n]\!]. \tag{5.5}$$

The nature of the points 0 and 1 as described by the Feller classification is the same for $(\xi_t^-(i))_t$ and $(\xi_t^+(i))_t$ and depends only on μ_1 and μ_0. To describe their behaviours near 0, let $\tau_z^{\pm,i}(a,b)$ denote the first time the process $(\xi_t^\pm(i))_t$, starting from z, exits (a,b) for $0 \leq a < z < b \leq 1$.

1. If $\mu_1 = \mu_0 = 0$ then 0 and 1 are absorbing points; $(\xi_t^\pm(i))_t$ reaches 0 or 1 in a finite time with probability one and

$$\mathbb{P}\left[\lim_{t\to\tau_z^{\pm,i}(0,1)} \xi_t^\pm(i) = 0\right] = \frac{\int_z^1 \exp\Big(-\int_{1/2}^x \frac{2b_i^\pm(u)}{u(1-u)}du\Big)dx}{\int_0^1 \exp\Big(-\int_{1/2}^x \frac{2b_i^\pm(u)}{u(1-u)}du\Big)dx}.$$

2. If $\mu_1 = 0$ and $\mu_0 > 0$ then 0 is the only absorbing point and $(\xi_t^\pm(i))_t$ reaches 0 in a finite time with probability one.

3. If $0 < \mu_1 < 1/2$ and $\mu_0 > 0$ then 0 is attainable: for every $0 < z < b < 1$,

$$\mathbb{P}\left[\tau_z^{\pm,i}(0,b) < +\infty \text{ and } \lim_{t\to\tau_z^{\pm,i}(0,b)} \xi_t^\pm(i) = 0\right] > 0.$$

4. If $\mu_1 \geq 1/2$ and $\mu_0 > 0$ then 0 is inaccessible: for every $0 < z < 1$,

$$\mathbb{P}_z[\exists t > 0,\ \xi_t^\pm(i) = 0] = 0 \text{ and } \mathbb{P}_z\left[\lim_{t\to+\infty} \xi_t^\pm(i) = 0\right] = 0.$$

Similar properties hold for the behaviour near the point 1.
These properties are sufficient to prove the boundary behaviour claimed for $(x_t)_t$. Since $x_t(i) \leq \xi_t^+(i)$ for every $t \geq 0$, if $(\xi_t^+(i))_t$ reaches 0 in a finite time then so must $(x_t(i))_t$. Similarly, if 0 is attainable for $(\xi_t^+(i))_t$ then 0 is also attainable for $(x_t(i))_t$. In the same way, since $x_t(i) \geq \xi_t^-(i)$ for every $t \geq 0$, if 0 is inaccessible for $(\xi_t^-(i))_t$ then 0 is also inaccessible for $(x_t(i))_t$.

It remains to prove that $(x_t)_t$ exits from $]0,1[$ in a finite time with probability one if $\mu_1 = \mu_0 = 0$. Let $\epsilon > 0$ be small enough that $x_0 \in [\epsilon, 1-\epsilon]^n$. The diffusion x_t exits from the compact $[\epsilon, 1-\epsilon]^n$ in a finite time with probability one. Let x_ϵ be a point on the

boundary of $[\epsilon, 1-\epsilon]^n$. There exists $i \in [\![1\,;n]\!]$ such that $x_\epsilon(i) \in \{\epsilon, 1-\epsilon\}$. For $z \in]0,1[$, set $\phi_i^{\pm}(z) := \mathbb{P}_z[\lim_{t\to\tau^{\pm,i}(0,1)} \xi_t^{\pm}(i) = 0]$. By the comparison theorem applied to the solutions of the stochastic differential equations (5.2), (5.3) and (5.4) starting from x_ϵ, the probability that the solution of (5.2) starting at x_ϵ reaches the boundary of $[0,1]^n$ in a finite time is greater than $\phi_i^+(\epsilon)$ if $x_i^\epsilon = \epsilon$ and is greater than $1-\phi_i^-(1-\epsilon)$ if $x_i^\epsilon = 1-\epsilon$. By the strong Markov property, the probability that (x_t) reaches the boundary in a finite time is greater than $\min\{\min(\phi_i^+(\epsilon), 1-\phi_i^-(1-\epsilon)), i \in [\![1\,;n]\!]\}$ for every $\epsilon > 0$. Therefore $(x_t)_t$ reaches the boundary in a finite time with probability one. □

6 The stationary measure of the limiting diffusion

6.1 Existence of a stationary distribution and an expression for its density

As in the one-locus case, when the mutation rates are strictly positive, the diffusion has a reversible stationary distribution:

Proposition 6.1. *Assume that the hypothesis H4 holds and that the mutation rates μ_0 and μ_1 are strictly positive. Set $\tilde{s}_{\boldsymbol{i},\boldsymbol{j}} = s_{\boldsymbol{i},\boldsymbol{j}} - s_{\mathbf{1},\mathbf{1}}$ for every pair of types $\boldsymbol{i},\boldsymbol{j} \in \mathcal{A}$. The diffusion with generator $\mathcal{G}_{n,s}$ has a unique reversible stationary distribution which has the following density with respect to the Lebesgue measure on $[0,1]^n$:*

$$g_{n,\mu,s}(x) = C_{n,\mu,s} \prod_{i=1}^{n} x_i^{2\mu_1-1}(1-x_i)^{2\mu_0-1} \exp(H_{n,s}(x))$$

where

- $H_{n,s}(x) = \dfrac{1}{2} \displaystyle\sum_{L\subset[\![1;n]\!],\ |L|\geq 1} m_L(\tilde{s}) \prod_{\ell\in L}(2x_\ell(1-x_\ell)) \prod_{k\in[\![1;n]\!]\setminus L}(1-2x_k(1-x_k))$;
- *$C_{n,\mu,s}$ is chosen so that* $\displaystyle\int_{[0,1]^n} g_{n,\mu,s}(x_1,\ldots,x_n)dx_1\cdots dx_n = 1$.

Remark 6.1. An expansion of the polynomial function $H_{n,s}$ yields:

$$H_{n,s}(x) = \sum_{L\subset[\![1;n]\!],\ |L|\geq 1} 2^{|L|-1}\delta_L[m(s)](\emptyset) \prod_{\ell\in L} x_\ell(1-x_\ell).$$

Proof of Proposition 6.1. Let $\mathcal{G}_{n,0}$ denote the generator of the limiting diffusion in the random mating case ($s_{\boldsymbol{i},\boldsymbol{j}} = 0$ for every $\boldsymbol{i},\boldsymbol{j} \in \mathcal{A}$). The diffusion associated with this generator is ergodic

and has a reversible stationary distribution $m_{\mu,0}$ which is the product of Beta distributions: $m_{\mu,0} := (\mathrm{Beta}(2\mu_0, 2\mu_1))^{\otimes n}$. In the general case, the generator $\mathcal{G}_{n,s}$ can be decomposed as

$$\mathcal{G}_{n,s} = \mathcal{G}_{n,0} + \frac{1}{2}\sum_{i=1}^{n} x_i(1-x_i)\partial_i h(x)\partial_i$$

where

$$h(x) = \sum_{L\subset[\![1;n]\!],\ |L|\geq 1} 2^{|L|-1}\delta_L[m(s)](\emptyset)\prod_{\ell\in L} x_\ell(1-x_\ell).$$

Therefore, as explained in Ethier & Nagylaki (1989), we can apply a result of Fukushima & Stroock (1986) to deduce that the diffusion associated with $\mathcal{G}_{n,s}$ has a unique reversible stationary distribution $m_{\mu,s}$ given by

$$m_{\mu,s}(dx) = C\exp(h(x))m_{\mu,0}(dx),$$

where C is chosen so that $m_{\mu,s}$ is a probability distribution. □

6.2 Description of the density of the stationary measure

We analyse the density of the stationary measure under two supplementary assumptions:

Assumption H6: *The two mutation rates μ_0 and μ_1 are assumed to be equal to a strictly positive real number μ.*

Assumption H7: *For every $L \in \mathcal{P}([\![1\,;n]\!])$, $m_L(s)$ depends only on $|L|$. We write $m(\ell)$ for the common value of $m_L(s)$ for $L \in \mathcal{P}([\![1\,;n]\!])$ such that $|L| = \ell$.*

Assumption H7 holds if the assortment parameters satisfy the Hamming criterion or the additive criterion.

Under the hypotheses H1, H2, H3, H4, H6 and H7, the density of the invariant measure can be written as $g_{n,\mu,s}(x) = C\exp(h_{n,\mu,s}(x))$ with

$$h_{n,\mu,s}(x) = (2\mu-1)\sum_{i=1}^{n}\ln(\rho(x_i)) + \sum_{\ell=0}^{n-1}\alpha_\ell \sum_{L\subset[\![1;n]\!],\ |L|=\ell+1}\prod_{k\in L}\rho(x_k),$$

where $\rho(x_i) = x_i(1-x_i)$ and $\alpha_\ell = 2^\ell\delta^{(\ell+1)}[m](0)$.

The study of the invariant measure in the one-locus case already provides a precise image of the graph of $g_{n,\mu,s}$ when the n coordinates of the diffusion are independent, that is when the assortment coefficients are chosen so that

$$\text{for every } \ell\in\{0,\dots,n-1\},\ m(\ell+1)-m(\ell) = m(1)-m(0).$$

There are then at least four different types of graph depending on the respective contributions to allelic diversity of mutations ($\mu > 1/2$ or $0 < \mu < 1/2$) and assortment parameters ($m(1)-m(0)$ smaller than $|8\mu - 4|$ or not) as shown in Fig. 1.

Proposition 6.2 gives conditions on the assortment parameters under which $(1/2, \dots, 1/2)$ is the only critical point of the density, as in the random mating case. Proposition 6.3 deals with situations far from the random mating case (the proofs are postponed to §6.4).

Proposition 6.2. *We assume that the hypotheses H1, H2, H3, H4, H6 and H7 hold. Set* $V_n = 2\mu - 1 + 2^{-(n+1)} \sum_{k=0}^{n-1} \binom{n-1}{k} \delta^{(1)}[m](k)$.

1. *If* $V_n > 0$, *then* $(1/2, \dots, 1/2)$ *is a local maximum of* $g_{n,\mu,s}$.
2. *If* $V_n < 0$, *then* $(1/2, \dots, 1/2)$ *is a local minimum of* $g_{n,\mu,s}$.
3. *If* $\mu > 1/2$ *and if* $\delta^{(1)}[m](\ell) \geq -(8\mu - 4)\ \forall \ell \in [\![0\,; n-1]\!]$, *then* $(1/2, \dots, 1/2)$ *is a global maximum and is the only critical point of* $g_{n,\mu,s}$.
4. *If* $0 < \mu < 1/2$ *and if* $\delta^{(1)}[m](\ell) \leq -(8\mu - 4)\ \forall \ell \in [\![0\,; n-1]\!]$, *then* $(1/2, \dots, 1/2)$ *is a global minimum and is the only critical point of* $g_{n,\mu,s}$.

Example 6.1. Let us consider the additive criterion with the assortment sequence $s_\ell = b\ell$ for $\ell \in [\![0\,; n]\!]$. Then $\delta^{(1)}[m](\ell) = 2^{-\ell}\binom{\ell}{\ell/2} b\, \mathbb{1}_{\{\ell \text{ is even}\}}$. As $2^{-\ell}\binom{\ell}{\ell/2}$ is a strictly decreasing sequence smaller than 1, $b < 0$ implies $V_n > 2\mu - 1 + \frac{1}{8}b$. Thus, it follows from Proposition 6.2 that if $\mu > 1/2$ and $b \geq -8(2\mu - 1)$, the point $(1/2, \dots, 1/2)$ is a local maximum of $g_{n,s,\mu}$. Let us note that if we consider the same sequence $s_\ell = b\ell$ but with the Hamming criterion, then for $\mu > 1/2$ and $b < -4(2\mu - 1)$, $(1/2, \dots, 1/2)$ is a local minimum of $g_{n,s,\mu}$.

Remark 6.2. The statement of Proposition 6.2 can be easily extended to a family of assortment parameters for which H7 does not hold: V_n must be replaced by

$$V_{n,i} = 2\mu - 1 + 2^{-(n+1)} \sum_{B \subset [\![1;n]\!] \setminus \{i\}} \delta_i[m(s)](B)$$

for every $i \in [\![1\,; n]\!]$ and the conditions on $\delta[m](\ell)$ in assertions 3 and 4 are replaced by a condition on $\delta_i[m(s)](A)$ for every $i \in [\![1\,; n]\!]$ and $A \in \mathcal{P}([\![1\,; n]\!] \setminus \{i\})$.

The following proposition describes the properties of the critical points of the density in two cases, (1) $\mu > 1/2$ and a condition on the assortment parameters which strongly favours mating between individuals carrying similar types:

$$\delta^{(1)}[m](n-1) \leq \delta^{(1)}[m](n-2) \leq \ldots \leq \delta^{(1)}[m](0) \leq 0 \text{ and } \delta^{(1)}[m](n-2) < 0,$$

and (2) $0 < \mu < 1/2$ and a condition on the assortment parameters which strongly favours mating between individuals with dissimilar types:

$$\delta^{(1)}[m](n-1) \geq \delta^{(1)}[m](n-2) \geq \ldots \geq \delta^{(1)}[m](0) \geq 0 \text{ and } \delta^{(1)}[m](n-2) > 0.$$

To simplify the statement, the description is limited to the hypercube $[0,1/2]^n$. The description on the whole space $[0,1]^n$ can be deduced from this since $g_{n,\mu,s}(x)$ is invariant if we replace any coordinate x_i with $1-x_i$.

Proposition 6.3. *Assume that conditions H1, H2, H3, H4, H6 and H7 hold. Set*

$$V_n = 2\mu - 1 + 2^{-(n+1)} \sum_{k=0}^{n-1} \binom{n-1}{k} \delta^{(1)}[m](k).$$

1. *Case $\mu > 1/2$. Assume furthermore that:*
   $\delta^{(1)}[m](n-1) \leq \delta^{(1)}[m](n-2) \leq \ldots \leq \delta^{(1)}[m](0) \leq 0$ *and* $\delta^{(1)}[m](n-2) < 0$.

   (a) *If $V_n > 0$ then $(1/2,\ldots,1/2)$ is a global maximum and is the only critical point of the density $g_{n,\mu,s}$.*

   (b) *If $V_n < 0$ then*

   i. *$g_{n,\mu,s}$ has a local minimum at $(1/2,\ldots,1/2)$.*

   ii. *In $[0,1/2]^n$, $g_{n,\mu,s}$ takes its maximum value at a unique point of the form $(\xi_0,\ldots,\xi_0)$.*

   iii. *The other critical points of $g_{n,\mu,s}$ in $[0,1/2]^n$ are saddle points: for every $\ell \in [\![1\,;n-1]\!]$, $g_{n,\mu,s}$ has $\binom{n}{\ell}$ saddle points of index $n-\ell$ in $[0,1/2]^n$. The saddle points of index $n-\ell$ have ℓ coordinates equal to $1/2$ and the other coordinates have the same value denoted by ξ_ℓ.*

   iv. *The relative positions of the coordinates of the critical points in $[0,1/2]^n$ satisfy $0 < \xi_{n-1} < \cdots < \xi_0 < 1/2$.*

   v. *The value of $g_{n,\mu,s}$ is the same at any saddle point of index $n-\ell$ and decreases as ℓ increases.*

2. *Case $0 < \mu < 1/2$. Assume furthermore that:*

   $$\delta^{(1)}[m](n-1) \geq \delta^{(1)}[m](n-2) \geq \ldots \geq \delta^{(1)}[m](0) \geq 0 \text{ and } \delta^{(1)}[m](n-2) > 0.$$

   (a) *If $V_n < 0$ then $(1/2,\ldots,1/2)$ is a global minimum and is the only critical point of the density $g_{n,\mu,s}$.*

(b) If $V_n > 0$ then

i. $g_{n,\mu,s}$ has a local maximum at $(1/2, \dots, 1/2)$.

ii. In $[0, 1/2]^n$, $g_{n,\mu,s}$ takes its minimum value at a unique point of the form $(\xi_0, \dots, \xi_0)$.

iii. The other critical points of $g_{n,\mu,s}$ in $[0, 1/2]^n$ are saddle points: for every $\ell \in [\![1\,; n-1]\!]$, $g_{n,\mu,s}$ has $\binom{n}{\ell}$ saddle points of index ℓ in $[0, 1/2]^n$. The saddle points of index ℓ have ℓ coordinates equal to $1/2$ and the other coordinates have the same value denoted by ξ_ℓ.

iv. The relative positions of the coordinates of the critical points in $[0, 1/2]^n$ satisfy $0 < \xi_{n-1} < \cdots < \xi_0 < 1/2$.

v. The value of $g_{n,\mu,s}$ is the same at any saddle point of index $n - \ell$ and increases as ℓ increases.

Remark 6.3.

1. $\xi_0 = 1/2 - 1/2\sqrt{1 - 4\lambda_0}$ where λ_0 is the unique solution in $]0, 1/4[$ of the equation:

$$2\mu - 1 + x \sum_{i=0}^{n-1} \delta^{(1)}[m](i) \binom{n-1}{i} (2x)^i (1-2x)^{n-1-i} = 0 \qquad (\mathcal{E}_0)$$

More generally, for every $\ell \in [\![0\,; n-1]\!]$, $\xi_\ell = 1/2 - 1/2\sqrt{1 - 4\lambda_\ell}$ where λ_ℓ is the unique solution in $]0, 1/4[$ of the equation:

$$2\mu - 1 + x \sum_{i=0}^{n-1} B_{n-1,\ell,i}(2x) \delta^{(1)}[m](i) = 0 \qquad (\mathcal{E}_\ell)$$

and $B_{n,\ell,i}(x) = 2^{-\ell} \sum_{j=\max(0,i-n+\ell)}^{\min(i,\ell)} \binom{\ell}{j} \binom{n-\ell}{i-j} x^{i-j} (1-x)^{n-\ell-(i-j)}$.
Let us note that $(B_{n,\ell,i}(x))_{i=0,\dots,n}$ are positive on $]0, 1[$ and their sum is equal to 1.

2. The assumption that $\delta^{(1)}[m](i)$ is a decreasing function of i cannot be removed since one can find examples of assortment parameters satisfying $\delta^{(1)}[m](i) < 0$ for every $i \in [\![0\,; n-1]\!]$ and such that:

   (a) $\mu > 1/2$, $V_n > 0$, but $(1/2, \dots, 1/2)$ is not the only local maximum,

   (b) $V_n < 0$ and $g_{n,\mu,s}$ has more than 2^n local maxima.

3. If x_i is the proportion of the population with allele 0 at the i-th locus, $2x_i(1-x_i)$ is the probability that two individuals sampled at random from the population carry different alleles at the ith locus. The density function of the reversible measure takes its maximum value at a point x such that for each $i \in [\![1\,;n]\!]$, $x_i(1-x_i) = \lambda_0$.

Example 6.2. Let us consider a quadratic sequence of parameters $s_\ell = s_0 - (b\ell + c\ell^2)$ $\forall \ell \in [\![0\,;n]\!]$ and let us define the assortment with this sequence by means of the Hamming criterion. If $c > 0$, $b+c \geq 0$ and $\mu > 1/2$ then $g_{n,\mu,s}$ has 3^n critical points if and only if $b+nc > 8\mu-4$. In this case, $\lambda_0 = n^{-1/2}\sqrt{\frac{2\mu-1}{4c}} + O(n^{-1})$. If $h_{n,k}$ denotes the value of the function $h_{n,\mu,s}$ at a critical point of index $n-k$ then $h_{n,0}-h_{n,n} \underset{n\to+\infty}{\sim} \frac{c}{8}n^2$ and $h_{n,0}-h_{n,1} \underset{n\to+\infty}{\sim} n^{1/2}1/2\sqrt{c(2\mu-1)}$ (see Appendix A.2 for more details).

6.3 Graphs of the density and simulations of trajectories in the two and three locus cases

Figures 2 to 4 show graphs of the density of the reversible stationary measure in the two-locus case for $\mu = 0.6$ and for several values of $s_1 - s_0$ and $s_2 - s_1$, the assortative mating being defined by the Hamming distance. Figures 2 and 3 illustrate the two situations considered in Proposition 6.3 when $\mu > 1/2$. When $s_1 - s_0 = 0$, the density may have a continuum of critical points as in Fig. 4; this corresponds to a case in which the assumption $\delta^{(1)}[m](n-2) < 0$ of Proposition 6.3 is not satisfied.

To illustrate the evolution of the 0-allelic frequency when $\mu > 1/2$ and the assortative mating strongly favours pairing between similar types, simulations were run in a population of size $N = 10^3$ with the two-locus model (Fig. 5) and with the three-locus model (Fig. 6). For these simulations, every individual initially carries the allele 0 at every locus, recombination occurs independently at each locus and the assortative mating is defined by the Hamming criterion. The trajectory is plotted at intervals of size N between the iterations N^2 and $33N^2$. To help to visualize the evolution, the colour of the plot changes every $\frac{1}{2}N^2$ iterations. The form of the density of the stationary measure here is highly reminiscent of that of the fitness landscapes studied in the adaptive evolution literature in modelling additive traits under frequency dependent intraspecific competition, see e.g. Schneider (2007) and references therein. In the deterministic setting the existence of multiple 'long term equilibria' renders the behaviour of the system very sensitive to assumptions about the initial conditions. In our setting, the presence of genetic drift is sufficient for the population to (eventually) explore the neighbourhoods of all the maxima, irrespective of its starting point. The time spent by the population in the neighbourhood of a maximum depends on the assortment parameters (Fig. 6a and 6b).

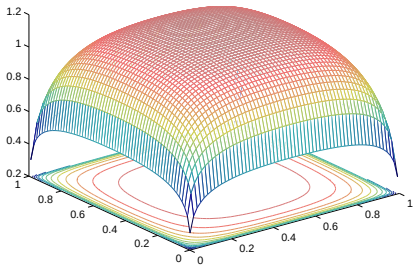


Figure 2: Graph of $g_{2,\mu,s}$ when $\mu = 0.6$, $s_1 - s_0 = -0.4$ and $s_2 - s_1 = -0.6$ so that the point $(1/2, 1/2)$ is the only critical point of the density $g_{2,s,\mu}$.

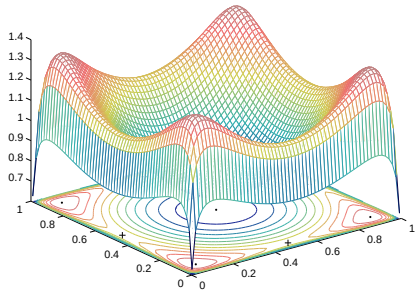


Figure 3: Graph of $g_{2,\mu,s}$ when $\mu = 0.6$, $s_1 - s_0 = -2$ and $s_2 - s_1 = -6$ so that $\lambda_0 \simeq 0.0766$. A black dot marks the position of each extremum and a cross is plotted at each saddle point.

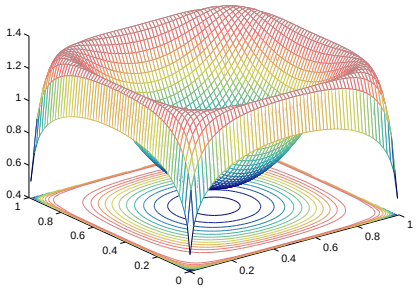


Figure 4: Graph of $g_{2,\mu,s}$ when $\mu = 0.6$, $s_1 - s_0 = 0$ and $s_2 - s_1 = -12$; there is a continuum of critical points.

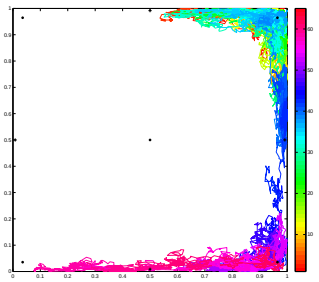
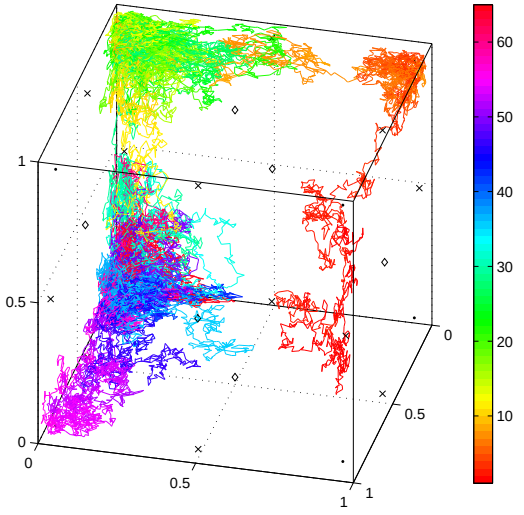


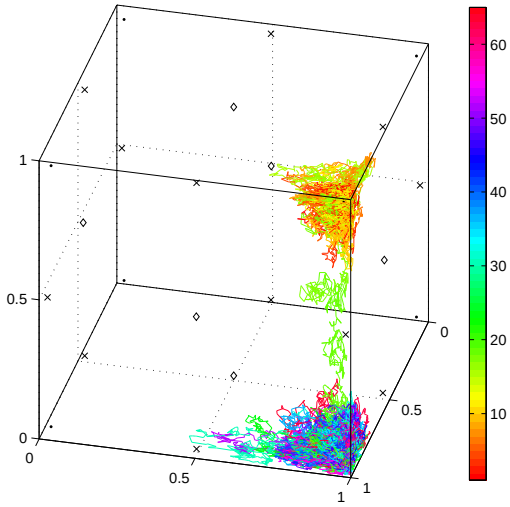
Figure 5: Simulation of the evolution of the 0-allelic frequency in the two-locus model. The population size is $N = 10^3$, $\mu = 1$, $s_1 - s_0 = -15$, $s_2 - s_1 = -210$. A black dot marks the position of each extremum and a cross is plotted at each saddle point. In this example, $\lambda_0 \simeq 0.034$ and $\lambda_1 \simeq 0.008$.



(a)

Assortment parameters: $s_1 - s_0 = -20$, $s_2 - s_1 = -40$ and $s_3 - s_2 = -60$.

Characteristics of the stationary density: $\lambda_0 \simeq 0.043$, $\lambda_1 \simeq 0.031$ and $\lambda_2 \simeq 0.025$. $h_0 - h_1 = 7.9$, $h_0 - h_2 \simeq 24.3$ and $h_0 - h_3 \simeq 49.8$.



(b)

Assortment parameters: $s_1 - s_0 = -30$, $s_2 - s_1 = -60$ and $s_3 - s_2 = -90$.

Characteristics of the stationary density: $\lambda_0 \simeq 0.030$, $\lambda_1 \simeq 0.021$ and $\lambda_2 \simeq 0.017$. $h_0 - h_1 = 12.6$, $h_0 - h_2 \simeq 38.6$ and $h_0 - h_3 \simeq 78.7$.

Figure 6: Simulations of the evolution of the 0-allelic frequency with the three-locus model for two different sets of assortment parameters. The assortative mating favours more strongly pairing between similar types in Fig. 6b. The size of the population is $N = 10^3$ and the mutation rate is $\mu = 1$. A black dot marks the position of each global maximum of the stationary density, a cross the position of each saddle point of index 2 and a diamond the position of each saddle point of index 1. Some numerical characteristics of the stationary density are presented to the right of each figure: for $i \in \{1, 2, 3\}$, the value of $\lambda_i = \xi_i(1 - \xi_i)$ provides the position of the critical points of index $3 - i$ (see Proposition 6.3) and h_i is the value of the log-density $h_{n,\mu,s}$ at a critical point of index $3 - i$.

6.4 Proofs of Propositions 6.2 and 6.3

Proof of Proposition 6.2 Let us introduce some notation in order to shorten the expressions. We set $\nu = 2\mu - 1$, $\rho(u) = u(1-u)$ for $u \in [0,1]$, $\boldsymbol{\rho}(x) = (\rho(x_1), \dots, \rho(x_n))$,

$$\bar{h}(x) = \nu \sum_{i=1}^{n} \log(x_i) + \frac{1}{2} \sum_{\ell=1}^{n} (m(\ell) - m(0)) \sum_{L \subset [\![1;n]\!],\ |L|=\ell} \prod_{j \in L} (2x_j) \prod_{k \in [\![1;n]\!] \setminus L} (1 - 2x_k)$$

and $h(x) = \bar{h}(\boldsymbol{\rho}(x))$ for $x = (x_1, \dots, x_n) \in]0,1[^n$. With this notation, $g_{n,\mu,s}(x) = C_{n,\mu,s} \exp(h(x))$.

1. For every $x \in]0,1[^n$ and $i \in [\![1\,;n]\!]$, $\partial_i h(x) = (1-2x_i)\partial_i \bar{h}(\boldsymbol{\rho}(x))$ where

$$\partial_i \bar{h}(x) = \frac{\nu}{x_i} + \sum_{\ell=0}^{n-1} \delta^{(1)}[m](\ell) \sum_{L \subset [\![1;n]\!] \setminus \{i\}, |L|=\ell} \prod_{j \in L} (2x_j) \prod_{k \in [\![1;n]\!] \setminus (L \cup \{i\})} (1 - 2x_k).$$

   First, the point $u_n = (1/2, \dots, 1/2)$ is a critical point of $h_{n,\mu,s}$ and the Hessian matrix at this point is the diagonal matrix $-2\Delta I_n$ where

$$\Delta = 4\nu + 2^{-(n-1)} \sum_{i=0}^{n-1} \binom{n-1}{i} \delta^{(1)}[m](i) = 4V_n.$$

   This proves the first two assertions of the proposition.

2. The last two assertions follow from the fact that $\partial_i \bar{h}(x)$ and Δ are increasing functions of $\delta^{(1)}[m](\ell)$ for every ℓ. Let us prove assertion 3 to illustrate the method. First, if $\delta^{(1)}[m(s)](\ell) = -(8\mu - 4)$ for every $\ell \in [\![0\,;n-1]\!]$ then the n coordinates of the diffusion are independent. In this case, $\Delta = 0$ and the stationary density has only one critical point at $(1/2, \dots, 1/2)$ which is a maximum. If $\{s_{\boldsymbol{i},\boldsymbol{j}}, (\boldsymbol{i},\boldsymbol{j}) \in \mathcal{A}^2\}$ is a family of assortment parameters such that $\partial_i \bar{h}(x)$ is nonnegative for every $x \in]0,1/4]^n$ and the density $g_{n,s,\mu}$ has a unique critical point at $(1/2, \dots, 1/2)$ which is a maximum, then the same is true for any family of assortment parameters $\{\hat{s}_{\boldsymbol{i},\boldsymbol{j}}, (\boldsymbol{i},\boldsymbol{j}) \in \mathcal{A}^2\}$ such that $\delta^{(1)}[m(\hat{s})](\ell) \geq \delta^{(1)}[m(s)](\ell)$ for every $\ell \in [\![0\,;n-1]\!]$.

Proof of Proposition 6.3 We retain the notation introduced in the proof of Proposition 6.2. For $k \in [\![1\,;n]\!]$, we set $\alpha_k = 2^k \delta^{(k+1)}[m](0)$ and denote by $e_{n,k}$ the elementary symmetric polynomial function in n variables of degree k:

$$e_{n,0}(x) = 1 \text{ and } e_{n,k}(x) = \sum_{\substack{L \subset [\![1;n]\!], \\ |L|=k}} \prod_{\ell \in L} x_\ell \quad \text{for } k \in [\![1\,;n]\!].$$

For instance, $e_{n,1}(x) = x_1 + \ldots + x_n$, $e_{n,2}(x) = \sum_{1\le i<j\le n} x_i x_j$.
With this notation

$$\bar{h}(x) = \nu \sum_{i=1}^{n} \ln(x_i) + \sum_{\ell=0}^{n-1} \alpha_\ell e_{n,\ell+1}(x).$$

In the proof we shall use (several times) the following identity for elementary symmetric polynomial functions:

Lemma 6.1. *Let n be an integer greater than 1 and let $k \in [\![0\,;n-2]\!]$. For every $x \in \mathbb{R}^n$, set $\hat{x}^{(i)} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ for $i \in [\![1\,;n]\!]$ and*

$$\hat{x}^{(i,j)} = \hat{x}^{(j,i)} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) \text{ for } i,j \in [\![1\,;n]\!] \text{ such that } i<j.$$

Then,

$$x_i e_{n-1,k}(\hat{x}^{(i)}) - x_j e_{n-1,k}(\hat{x}^{(j)}) = (x_i - x_j) e_{n-2,k}(\hat{x}^{(i,j)}). \tag{6.1}$$

We shall also use the following alternative expression for symmetric polynomial functions that are similar to the polynomial term in h:

Lemma 6.2. *Let $n \in \mathbb{N}^*$ and let $a_0, \ldots, a_n$ be real numbers. Then for every $x \in \mathbb{R}^n$,*

$$\sum_{k=0}^{n} 2^k \delta^{(k)}[a](0) e_{n,k}(x) = \sum_{i=0}^{n} a_i \sum_{I\subset[\![1;n]\!],\ |I|=i} \prod_{i\in I} 2x_i \prod_{j\notin I}(1-2x_j).$$

In particular, for every $y \in \mathbb{R}$ and $\ell \in [\![0\,;n]\!]$,

$$\sum_{k=0}^{n} 2^k \delta^{(k)}[a](0) e_{n,k}((1/4)^{\otimes \ell}, y^{\otimes(n-\ell)}) = \sum_{i=0}^{n} a_i B_{n,\ell,i}(2y)$$

where $B_{n,\ell,i}(y) = 2^{-\ell} \displaystyle\sum_{j=\max(0,i-n+\ell)}^{\min(i,\ell)} \binom{\ell}{j}\binom{n-\ell}{i-j} y^{i-j}(1-y)^{n-\ell-(i-j)}$.

Proof. See Corollary A.2. □

1. Let us assume that $x = (x_1, \ldots, x_n)$ is a critical point of $g_{n,\mu,s}$ different from u_n. Let ℓ denote the number of coordinates equal to $1/2$ ($\ell \in [\![0\,;n-1]\!]$). Every coordinate x_i different from $1/2$ has to satisfy: $\partial_i h_{n,\mu,s}(\boldsymbol{\rho}(x)) = 0$, that is

$$\nu + \rho(x_i) \sum_{k=0}^{n-1} \alpha_k e_{n-1,k}(\widehat{\boldsymbol{\rho}(x)}^{(i)}) = 0.$$

In particular, it follows from Lemma 6.1 that if x_i and x_j are two coordinates of the critical point x not equal to $1/2$ then

$$\rho(x_i) = \rho(x_j) \text{ or } \sum_{k=0}^{n-2} \alpha_k e_{n-2,k}(\widehat{\boldsymbol{\rho}(x)}^{(i,j)}) = 0.$$

By Lemma 6.2,

$$\sum_{k=0}^{n-2} \alpha_k e_{n-2,k}(x) = \sum_{\ell=0}^{n-2} \delta^{(1)}[a](\ell) Q_\ell(x),$$

where Q_ℓ denotes a polynomial function which is positive on $x \in]0, 1/4[^{n-2}$ for every $\ell \in [\![0\,; n-2]\!]$. Thus this sum cannot vanish in $]0, 1/4[^{n-2}$ under the assumption that all coefficients $\delta^{(1)}[m](i)$ have the same sign and that for at least one $i \leq n-2$, $\delta^{(1)}[m](i)$ is non-zero. Therefore, such a critical point exists only if there exists a solution in the interval $]0, 1/4[$ of

$$\nu + y \sum_{k=0}^{n-1} \alpha_k e_{n-1,k}\left((\frac{1}{4})^{\otimes \ell}, y^{\otimes(n-\ell-1)}\right) = 0. \tag{$\mathcal{E}'_\ell$}$$

In order to study the solutions of ($\mathcal{E}'_\ell$), let $\phi_\ell(y)$ denote the left-hand side of ($\mathcal{E}'_\ell$):

$$\phi_\ell(y) = \nu + y \sum_{k=0}^{n-1} \alpha_k e_{n-1,k}((\frac{1}{4})^{\otimes \ell}, y^{\otimes(n-\ell-1)}) \tag{6.2}$$

By Lemma 6.2,

$$\phi_\ell(y) = \nu + y \sum_{i=0}^{n-1} B_{n-1,\ell,i}(2y) \delta^{(1)}[m](i). \tag{6.3}$$

Therefore, ($\mathcal{E}'_\ell$) coincides with ($\mathcal{E}_\ell$) of Remark 6.3. The derivative of ϕ_ℓ is equal to:

$$\phi'_\ell(y) = \sum_{i=0}^{n-1} B_{n-1,\ell,i}(2y) \delta^{(1)}[m](i) + 2y(n-1-\ell) \sum_{i=0}^{n-2} B_{n-2,\ell,i}(2y)(\delta^{(1)}[m](i+1) - \delta^{(1)}[m](i)).$$

If $\delta^{(1)}[m](n-1) \leq \cdots \leq \delta^{(1)}[m](0) \leq 0$ (respectively $\delta^{(1)}[m](n-1) \geq \cdots \geq \delta^{(1)}[m](0) \geq 0$), ϕ_ℓ is a decreasing function on the interval $[0, 1/2]$ (resp. an increasing function on the interval $[0, 1/2]$). The value of ϕ_ℓ at 0 is ν and the value at $1/4$ is V_n. Therefore, under the assumptions of 1 or 2 of the proposition, for every $\ell \in \{0, \ldots, n-1\}$ ($\mathcal{E}'_\ell$) has no solution in $]0, 1/4[$ if V_n and ν have the same sign and has exactly one solution in $]0, 1/4[$ denoted by λ_ℓ if V_n and ν have opposite signs. This proves assertions 1.(a) and 2.(a).

For every pair of disjoint subsets I and J of $[\![1\,;n]\!]$, let us introduce the following point: $u_{I,J} = (x_1, \dots, x_n)$ with

$$x_i = \begin{cases} 1/2 \text{ if } i \in I, \\ 1/2 + 1/2\sqrt{1-4\lambda_{|I|}} \text{ if } i \in J, \\ 1/2 - 1/2\sqrt{1-4\lambda_{|I|}} \text{ if } i \in [\![1\,;n]\!] \setminus (I \cup J). \end{cases}$$

We have shown that if V_n and ν have opposite signs, then every point $u_{I,J}$ is a critical point and any critical point is one of these points $u_{I,J}$.

So that we may use our conclusions above, from now on, we assume that the hypotheses stated in point 1 of the proposition are satisfied. However, the computations that follow do not depend on these hypotheses, and so our proof is easily modified to the setting of point 2.

2. Let us study the Hessian matrix of $h_{n,\mu,s}$ at a critical point $u_{I,J}$ such that $|I| \leq n-1$. For that, set $\ell = |I|$, $\ell^+ = |J|$ and $\ell^- = n - \ell - \ell^+$ and let us introduce the following notations:

$$a_\ell = \partial_1 \bar{h}((\frac{1}{4})^{\otimes \ell}, (\lambda_\ell)^{\otimes(n-\ell)}), \qquad b_\ell = -(1-4\lambda_\ell)\frac{\nu}{\lambda_\ell^2},$$
$$c_\ell = (1-4\lambda_\ell)\partial^2_{n,n-1}\bar{h}((\frac{1}{4})^{\otimes \ell}, (\lambda_\ell)^{\otimes(n-\ell)}).$$

The Hessian matrix of $h_{n,\mu,s}$ at $u_{I,J}$ is permutation-similar to the following block matrix:

$$\mathcal{H}_{I,J} = \begin{pmatrix} \boldsymbol{A}_\ell & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B}_{\ell,\ell^+} & \boldsymbol{C}_\ell \\ \mathbf{0} & \boldsymbol{C}_\ell & \boldsymbol{B}_{\ell,\ell^-} \end{pmatrix}$$

where

- $\boldsymbol{A}_\ell$ denotes the scalar matrix $-2a_\ell I_\ell$ with $a_\ell = \partial_1 \bar{h}((\frac{1}{4})^{\otimes \ell}, (\lambda_\ell)^{\otimes(n-\ell)})$,
- $\boldsymbol{B}_{\ell,k}$ denotes the following k-by-k matrix : $\boldsymbol{B}_{\ell,k} = \begin{pmatrix} b_\ell & c_\ell & \cdots & c_\ell \\ c_\ell & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & c_\ell \\ c_\ell & \cdots & c_\ell & b_\ell \end{pmatrix}$,
- $\boldsymbol{C}_\ell$ denotes the ℓ^+-by-ℓ^- matrix all the elements of which are equal to $-c_\ell$.

By assumption on μ, $b_\ell < 0$. To complete the proof of assertions (i) and (ii) of 1-(b), we shall prove that $a_\ell < 0$ and that $b_\ell < c_\ell < 0$. That will imply that the submatrix $\begin{pmatrix} \boldsymbol{B}_{\ell,\ell^+} & \boldsymbol{C}_\ell \\ \boldsymbol{C}_\ell & \boldsymbol{B}_{\ell,\ell^-} \end{pmatrix}$ is negative definite (for more details, see Lemma A.2) hence that the Hessian matrix of $h_{n,\mu,s}$ at a point $u_{I,J}$ has $|I|$ positive eigenvalues and $n - |I|$ negative eigenvalues.

First, let us study the sign of $a_\ell = 4\nu + \sum_{i=0}^{n-1} \alpha_i e_{n-1,i}((\frac{1}{4})^{\otimes(\ell-1)}, \lambda_\ell^{\otimes(n-\ell)})$. As $\phi_\ell(\lambda_\ell) = 0$, an application of Lemma 6.1 yields:

$$a_\ell = (1 - 4\lambda_\ell) \sum_{i=0}^{n-2} \alpha_i e_{n-2,i}((\frac{1}{4})^{\otimes(\ell-1)}, \lambda_\ell^{\otimes(n-1-\ell)}).$$

The right-hand side can be rewritten using Lemma 6.2:

$$a_\ell = (1 - 4\lambda_\ell) \sum_{i=0}^{n-2} \delta^{(1)}[m](i) B_{n-2,\ell-1,i}(2\lambda_\ell).$$

The conditions on $\delta^{(1)}[m](i)$ imply that a_ℓ is negative.

Let us now study the coefficients $\tilde{b}_\ell = (1 - 4\lambda_\ell)^{-1} b_\ell$ and $\tilde{c}_\ell = (1 - 4\lambda_\ell)^{-1} c_\ell$. As in the study of a_ℓ we use that $\phi_\ell(\lambda_\ell) = 0$ and Lemma 6.2 to write $\tilde{b}_\ell$ and $\tilde{c}_\ell$ in terms of the coefficients $\delta^{(1)}[m](s)(i)$:

$$\tilde{b}_\ell = \frac{1}{\lambda_\ell} \sum_{i=0}^{n-1} \delta^{(1)}[m](i) B_{n-1,\ell,i}(2\lambda_\ell),$$
$$\tilde{c}_\ell = 2 \sum_{i=0}^{n-2} (\delta^{(1)}[m](i+1) - \delta^{(1)}[m](i)) B_{n-2,\ell,i}(2\lambda_\ell).$$

As $\delta^{(1)}[m(s)](i)$ is assumed to be a decreasing sequence, $\tilde{c}_\ell < 0$. After some computations, we obtain:

$$\lambda_\ell(\tilde{c}_\ell - \tilde{b}_\ell) = -\sum_{i=0}^{n-2} \delta^{(1)}[m](i) B_{n-2,\ell,i}(2\lambda_\ell).$$

The conditions on $\delta^{(1)}[m](i)$ imply that $\tilde{c}_\ell > \tilde{b}_\ell$.

3. Let us prove that $0 < \lambda_{n-1} < \cdots < \lambda_0 < 1/4$, which gives the relative positions of the coordinates of the critical points.

   Let $\ell \in [\![0\,; n-2]\!]$. If we return to the expression (6.2) of ϕ_ℓ, use Lemma 6.1 and then

Lemma 6.2, we obtain:

$$\phi_{\ell+1}(y) - \phi_\ell(y) = y(1/4-y)\sum_{i=0}^{n-2} \alpha_{i+1} e_{n-2,i}((1/4)^{\otimes \ell}, y^{\otimes(n-2-\ell)})$$
$$= 2y(1/4-y)\sum_{i=0}^{n-2} \delta^{(2)}[m](i) B_{n-2,\ell,i}(2y).$$

By assumption, $\delta^{(2)}[m](i) \leq 0$ for every $i \in [\![0\,;n-2]\!]$ hence $\phi_{\ell+1}(y) \leq \phi_\ell(y)$ for every $y \in [0,1/4]$. As the functions ϕ_ℓ are decreasing on $[0,1/4]$, we deduce that $\lambda_{\ell+1} \leq \lambda_\ell$ for every $\ell \in [\![0\,;n-2]\!]$. As the two critical points $u_{[\![1;\ell]\!],\emptyset}$ and $u_{[\![1;\ell+1]\!],\emptyset}$ have not the same properties, they cannot coincide and thus $\lambda_{\ell+1} < \lambda_\ell$ for every $\ell \in [\![0\,;n-2]\!]$.

4. *Proof of assertion* 1.(*b*).*v:* let h_ℓ denote the value of $h_{n,\mu,s}$ at a saddle point of index $n-\ell$: $p_\ell = ((1/2)^{\otimes \ell}, (\xi_\ell)^{\otimes(n-\ell)})$. To prove that $h_\ell > h_{\ell+1}$ for every $\ell \in [\![0\,;n-2]\!]$, we shall use the properties of the gradient dynamical system $\frac{dx(t)}{dt} = -\nabla \tilde{h}(x)$ with $\tilde{h} = -h_{n,\mu,s}$. Fix a positive value M large enough so that $U_M = \tilde{h}^{-1}([-M,M])$ contains all critical points of h (such an M exists since $\tilde{h}(x)$ tends to infinity as x tends to the boundary of $[0,1]^n$). The function $\tilde{h}$ decreases along trajectories and a trajectory of a point $x \in M$ converges to a critical point of $\tilde{h}$ as t tends towards $+\infty$, since $\tilde{h}$ has only isolated critical points. For $k \in \{0,\dots,n-1\}$, let $U_M^{(k)}$ denote the subset:

$$U_M^{(k)} = \{x \in U_M,\ x_1 = \cdots = x_k = 1/2 \text{ and } x_i < 1/2\ \forall i > k\}.$$

Every subset $U_M^{(k)}$ contains exactly one critical point, the saddle point p_k. As $\partial_i \tilde{h}(x) = 0$ at points x such that $x_i = 1/2$, the subset $U_M^{(k)}$ is positively invariant by the gradient flow. Therefore, to prove that $h_k > h_{k+1}$, it is enough to show that there exists $0 < y_0 < 1/2$ such that for $y \in]y_0, 1/2[$, $\tilde{h}((1/2)^{\otimes k}, y, \xi_{k+1}^{\otimes n-k-1}) < \tilde{h}(p_{k+1})$.
As $\tilde{h}((1/2)^{\otimes k}, y, \xi_{k+1}^{\otimes n-k-1}) = -\bar{h}_{n,\mu,s}((1/4)^{\otimes k}, y(1-y), \lambda_{k+1}^{\otimes n-k-1})$, it is enough to show that $\partial_{k+1}\bar{h}_{n,\mu,s}((1/4)^{\otimes(k+1)}, \lambda_{k+1}^{\otimes(n-k-1)}) < 0$. Using that λ_{k+1} is solution of the equation $(\mathcal{E}_{k+1})$, we obtain

$$\partial_{k+1}\bar{h}_{n,\mu,s}((1/4)^{\otimes(k+1)}, \lambda_{k+1}^{\otimes(n-k-1)}) = (1-4\lambda_{k+1})\sum_{i=0}^{n-2} \delta^{(1)}[m](i) B_{n-2,k,i}(2y) < 0.$$

7 Proof of convergence to the diffusion

In this section, we prove convergence to the diffusion approximation in the n-locus case (Theorem 4.1). We also establish the two simple expressions for the drift presented in §4.

First, the properties of the generator $\mathcal{G}_{n,s}$ stated in assertion (a) of Theorem 4.1 can be obtained by applying the following theorem established by Cerrai and Clément:

Theorem 7.1 (Cerrai & Clément 2004)**.** *Let $\mathcal{S}^+(\mathbb{R}^n)$ be the space of symmetric, non-negative definite, $n \times n$ matrices. Let $A : [0,1]^n \to \mathcal{S}^+(\mathbb{R}^n)$ and $b : [0,1]^n \to \mathbb{R}^n$ be mappings of class C^2. For $i \in \{1,\dots,n\}$ and $\epsilon \in \{0,1\}$, let ν^i_ϵ denote the unit inward normal vector of the hypercube $C^i_\epsilon = \{x \in [0,1]^n,\ x_i = \epsilon\}$. Let us assume the following two conditions:*

- *for every $i \in \{1,\dots,n\}$, $\epsilon \in \{0,1\}$ and $x \in C^i_\epsilon$, $A(x)\nu^i_\epsilon(x) = \mathbf{0}$ and $\langle b(x), \nu^i_\epsilon(x)\rangle \geq 0$;*
- *for every $i,j \in \{1,\dots,n\}$, $A_{i,j}(x)$ depends only on x_i and x_j.*

Then the operator

$$L = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{i,j}(x)\frac{\partial^2}{\partial_{x_i}\partial_{x_j}} + \sum_{i=1}^{n} b_i(x)\frac{\partial^2}{\partial_{x_i}}$$

is closable in $C([0,1]^n)$ and its closure is the generator of a strongly continuous semigroup of contractions.

To prove the convergence result, we use the following theorem, due to Ethier and Nagylaki, on diffusion approximations for Markov chains with two time scales.

Theorem 7.2 (Ethier & Nagylaki 1980, Theorem 3.3)**.** *For $N \in \mathbb{N}^*$, let $\{Z^N_k,\ k \in \mathbb{N}\}$ be a homogeneous Markov chain in a metric space E_N with Feller transition function. Let F_1 and F_2 be compact convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, having non-empty interiors. Assume further that $0 \in \overset{\circ}{F}_2$. Let $\Phi_N : E_N \to F_1$ and $\Psi_N : E_N \to F_2$ be continuous functions. Define $X^N_k = \Phi_N(Z^N_k)$ and $Y^N_k = \Psi_N(Z^N_k)$ for each $k \in \mathbb{N}$. Let $(\epsilon_N)_N$ and $(\delta_N)_N$ be two positive sequences such that $\delta_N \to 0$ and $\epsilon_N/\delta_N \to 0$.*
Assume that there exist continuous functions $a : F_1 \times \mathbb{R}^m \to \mathbb{R}^n \otimes \mathbb{R}^n$, $b : F_1 \times \mathbb{R}^m \to \mathbb{R}^n$ and $c : F_1 \times \mathbb{R}^m \to \mathbb{R}^m$ such that for $i,j \in [\![1\,;n]\!]$ and $\ell \in [\![1\,;m]\!]$ the following properties (a)-(e) hold as $N \to +\infty$ uniformly in $z \in E_N$ where $x = \Phi_N(z)$ and $y = \Psi_N(z)$:

(a) $\epsilon_N^{-1}\,\mathbb{E}_z[X^N_1(i) - x(i)] = b_i(x,y) + o(1)$,

(b) $\epsilon_N^{-1}\,\mathbb{E}_z\left[(X^N_1(i) - x(i))(X^N_1(j) - x(j))\right] = a_{i,j}(x,y) + o(1)$,

(c) $\epsilon_N^{-1}\,\mathbb{E}_z[(X^N_1(i) - x(i))^4] = o(1)$,

(d) $\delta_N^{-1}\,\mathbb{E}_z[Y^N_1(\ell) - y(\ell)] = c_\ell(x,y) + o(1)$,

(e) $\delta_N^{-1}\,\mathbb{E}_z[(Y_1^N(\ell)-y(\ell))^2]=o(1)$.

Assume further that

(f) *c is of class C^2, $c(x,0)=0$ for all $x\in\mathbb{R}^m$ and the solution of the differential equation*

$$\frac{d}{dt}u(t,x,y)=c(x,u(t,x,y)),\quad u(0,x,y)=y.$$

exists for all $(t,x,y)\in[0,+\infty[\times F_1\times F_2$ and satisfies

$$\lim_{t\to+\infty}\sup_{(x,y)\in F_1\times F_2}|u(t,x,y)|=0.$$

(g) *The closure of the following operator*

$$\mathcal{L}=\frac{1}{2}\sum_{i,j=1}^{n}a_{i,j}(x,0)\frac{\partial^2}{\partial_{x_i}\partial_{x_j}}+\sum_{i=1}^{n}b_i(x,0)\frac{\partial}{\partial_{x_i}},\quad \mathcal{D}(\mathcal{L})=C^2(F_1),$$

generates a strongly continuous semigroup on $C(F_1)$ corresponding to a diffusion process X in F_1.

Then the following conclusions in which the symbol $\Rightarrow$ denotes convergence in distribution, hold:

(i) *If $X_0^N\Rightarrow X(0)$ then $\{X^N_{[t/\epsilon_N]},t\geq 0\}\Rightarrow X(\cdot)$ in $D_{F_1}([0,+\infty[)$ (where $D_{F_1}([0,+\infty[)$ is the space of càdlàg paths $\omega:[0,\infty)\to F_1$ with the Skorohod topology),*

(ii) *For every positive sequence $(t_N)_N$ that converges to $+\infty$, $Y^N_{[t_N/\delta_N]}\Rightarrow 0$.*

Remark 7.1. We have only stated the part of Ethier and Nagylaki's theorem that we need. The full statement also gives a convergence result when the sequence $(\delta_N)_N$ converges to a positive real number.

To apply this theorem, we consider the two sequences $\epsilon_N=N^{-2}$ and $\delta_N=N^{-1}$, we set $E_N=\{z\in(N^{-1}\,\mathbb{N})^{\mathcal{A}},\ \sum_{\boldsymbol{i}\in\mathcal{A}}z(\boldsymbol{i})=1\}$, and we define by (Φ_N,Ψ_N) a change of coordinates such that $\Psi_N^{-1}(\{0\})$ is the linkage equilibrium manifold:

$$\begin{array}{rccccccc}\Phi_N: & E_N & \to & [0,1]^n & \text{and} & \Psi_N: & E_N & \to & [-1,1]^{2^n-n-1}\\ & z & \mapsto & (u_1,\dots,u_n) & & & z & \mapsto & (u_I,\ I\subset[\![1\,;n]\!] \text{ s. t. } |I|\geq 2)\end{array}$$

where $u_i=\sum_{\boldsymbol{\ell},\ell_i=0}z(\boldsymbol{\ell})$ for $i\in[\![1\,;n]\!]$ and $u_I=\prod_{i\in I}u_i-\sum_{\boldsymbol{\ell},\boldsymbol{\ell}_{|I}\equiv 0}z(\boldsymbol{\ell})$ for each $I\subset[\![1\,;n]\!]$ having at least two elements.

First (in §7.1), we shall check that $X_1^{(N)} = \Phi_N(Z_1^{(N)})$ and $Y_1^{(N)} = \Psi_N(Z_1^{(N)})$ satisfy the conditions (a)-(f) of Ethier and Nagylaki's theorem with the following expressions for the functions $a_{i,j}(x,0)$ and $b_i(x,0)$:

$$a_{i,j}(x,0) = x(i)(1-x(i))\,\mathbb{1}_{\{i=j\}}, \tag{7.1}$$

$$b_i(x,0) = (1-x(i))\mu_1 - x(i)\mu_0 + (1/2 - x(i))x(i)(1-x(i))P_{i,s}(x), \tag{7.2}$$

where

$$P_{i,s}(x) = \sum_{J\subset[\![1;n]\!]\setminus\{i\}} \sum_{H\subset[\![1;n]\!]\setminus\{i\}} (s_{J\cup\{i\},H} - s_{J,H}) \prod_{j\in J} x(j) \prod_{h\in H} x(h) \prod_{\substack{j\in[\![1;n]\!],\\ j\notin J\cup\{i\}}} (1-x(j)) \prod_{\substack{h\in[\![1;n]\!],\\ h\notin H\cup\{i\}}} (1-x(h)),$$

and, for two subsets I and J of $[\![1\,;n]\!]$, $s_{I,J}$ denotes the assortment parameter $s_{\boldsymbol{i},\boldsymbol{j}}$ for the types $\boldsymbol{i} = (\mathbf{0}_I, \mathbf{1}_{\bar{I}})$ and $\boldsymbol{j} = (\mathbf{0}_J, \mathbf{1}_{\bar{J}})$.

In §7.2 we shall show that $P_{i,s}$ has the following two equivalent expressions:

$$\begin{aligned} P_{i,s}(x) &= \sum_{A\subset[\![1;n]\!]\setminus\{i\}} 2^{|A|}\delta_{A\cup\{i\}}[m(s)](\emptyset) \prod_{\ell\in A} x(\ell)(1-x(\ell)) \\ &= \sum_{A\subset[\![1;n]\!]\setminus\{i\}} \delta_i[m(s)](A) \prod_{k\in A} 2x(k)(1-x(k)) \prod_{\ell\notin A\cup\{i\}} \Big(1 - 2x(\ell)(1-x(\ell))\Big). \end{aligned}$$

7.1 Verification of the conditions (a)-(f) of Ethier and Nagylaki's theorem

As the proportion of individuals of a given type $\boldsymbol{i}$ can only change by $\pm 1/N$ in one step:

- If $r \in \mathbb{N}^*$ and $\boldsymbol{i} \in \mathcal{A}$, then

$$\mathbb{E}_z\left[(Z_1^{(N)}(\boldsymbol{i}) - z(\boldsymbol{i}))^r\right] = N^{-r} \sum_{\boldsymbol{j}\in\mathcal{A}\setminus\{\boldsymbol{i}\}} \Big(f_N(z,\boldsymbol{j},\boldsymbol{i}) + (-1)^r f_N(z,\boldsymbol{i},\boldsymbol{j})\Big) \tag{7.3}$$

- if $r, u \in \mathbb{N}^*$, $\boldsymbol{i}, \boldsymbol{j} \in \mathcal{A}$ so that $\boldsymbol{i} \neq \boldsymbol{j}$, then

$$\begin{aligned} &\mathbb{E}_z\left[(Z_1^{(N)}(\boldsymbol{i}) - z(\boldsymbol{i}))^r (Z_1^{(N)}(\boldsymbol{j}) - z(\boldsymbol{j}))^u\right] \\ &\qquad = N^{-(r+u)}\Big((-1)^r f_N(z,\boldsymbol{i},\boldsymbol{j}) + (-1)^u f_N(z,\boldsymbol{j},\boldsymbol{i})\Big) \end{aligned} \tag{7.4}$$

- if $r \geq 3$ and $\boldsymbol{i}^{(1)},\ldots,\boldsymbol{i}^{(r)} \in \mathcal{A}$ so that at least three of them are distinct, then

$$\mathbb{E}_z\Big[\prod_{u=1}^{r}\big(Z_1^{(N)}(\boldsymbol{i}^{(u)}) - z(\boldsymbol{i}^{(u)})\big)\Big] = 0. \tag{7.5}$$

Condition (a). To show that condition (a) of Theorem 7.2 holds, we first examine the drift of $Z^{(N)}$. A Taylor expansion of the transition probabilities of the Markov chain $(Z_t^{(N)})_{t\in\mathbb{N}}$ using assumption H2 yields the following formula:

Lemma 7.1. *For every $\boldsymbol{i} \in \mathcal{A}$,*

$$N^2\,\mathbb{E}_z[Z_1^{(N)}(\boldsymbol{i}) - z(\boldsymbol{i})] = NB_{\boldsymbol{i}}^{(0)}(z) + B_{\boldsymbol{i}}^{(1)}(z) + O(N^{-1}),\ \textit{uniformly on } z \in E_N,$$

where

$$\begin{aligned}
B_{\boldsymbol{i}}^{(0)}(z) &= \sum_{\boldsymbol{k}\in\mathcal{A}}\sum_{\boldsymbol{j}\in\mathcal{A}} z(\boldsymbol{j})z(\boldsymbol{k})q((\boldsymbol{j},\boldsymbol{k});\boldsymbol{i}) - z(\boldsymbol{i})\\
B_{\boldsymbol{i}}^{(1)}(z) &= \sum_{\boldsymbol{k}\in\mathcal{A}}\sum_{\boldsymbol{j}\in\mathcal{A}} z(\boldsymbol{j})z(\boldsymbol{k})\Big(\sum_{u=1}^{n} q((\boldsymbol{j},\boldsymbol{k});(1-i_u,\boldsymbol{i}_{[\![1;n]\!]\setminus\{u\}}))\mu_{1-i_u} - q((\boldsymbol{j},\boldsymbol{k});\boldsymbol{i})\sum_{u=1}^{n}\mu_{i_u}\Big)\\
&+ \sum_{\boldsymbol{k}\in\mathcal{A}}\sum_{\boldsymbol{j}\in\mathcal{A}} s_{\boldsymbol{j},\boldsymbol{k}}z(\boldsymbol{j})z(\boldsymbol{k})q((\boldsymbol{j},\boldsymbol{k});\boldsymbol{i}) - z(\boldsymbol{i})\sum_{\boldsymbol{k}\in\mathcal{A}} s_{\boldsymbol{i},\boldsymbol{k}}z(\boldsymbol{k})\\
&- \sum_{\boldsymbol{k}\in\mathcal{A}}\sum_{\boldsymbol{j}\in\mathcal{A}}\sum_{\boldsymbol{h}\in\mathcal{A}} s_{\boldsymbol{j},\boldsymbol{h}}z(\boldsymbol{j})z(\boldsymbol{h})z(\boldsymbol{k})q((\boldsymbol{j},\boldsymbol{k});\boldsymbol{i}) + z(\boldsymbol{i})\sum_{\boldsymbol{h}\in\mathcal{A}}\sum_{\boldsymbol{k}\in\mathcal{A}} s_{\boldsymbol{i},\boldsymbol{h}}z(\boldsymbol{k})z(\boldsymbol{h})
\end{aligned}$$

Proof. By assumption H2, for two different types $\boldsymbol{i},\boldsymbol{j} \in \mathcal{A}$

$$f_N(z,\boldsymbol{i},\boldsymbol{j}) := \sum_{\boldsymbol{k},\boldsymbol{\ell}\in\mathcal{A}} z(\boldsymbol{i})z(\boldsymbol{k})w^{(N)}(z,\boldsymbol{i},\boldsymbol{k})q((\boldsymbol{i},\boldsymbol{k});\boldsymbol{\ell})\mu^{(N)}(\boldsymbol{\ell},\boldsymbol{j}).$$

where $w^{(N)}(z,\boldsymbol{i},\boldsymbol{k}) = 1 + \frac{1}{N}\big(s_{\boldsymbol{i},\boldsymbol{k}} - \sum_{\boldsymbol{h}\in\mathcal{A}} s_{\boldsymbol{i},\boldsymbol{h}}z(\boldsymbol{h})\big) + O(N^{-2})$ and

$$\mu^{(N)}(\boldsymbol{\ell},\boldsymbol{j}) = \begin{cases} 1 - \frac{1}{N}\sum_{u=1}^{n}\mu_{j_u} + O(N^{-2}) & \text{if } d_h(\boldsymbol{\ell},\boldsymbol{j}) = 0\\ \frac{1}{N}\mu_{1-j_i} + O(N^{-2}) & \text{if } d_h(\boldsymbol{\ell},\boldsymbol{j}) = 1 \text{ and } \ell_i = 1-j_i\\ O(N^{-2}) & \text{if } d_h(\boldsymbol{\ell},\boldsymbol{j}) \geq 2 \end{cases}$$

To prove Lemma 7.1, it suffices to use these expansions in

$$\mathbb{E}_z\left[Z_1^{(N)}(\boldsymbol{i}) - z(\boldsymbol{i})\right] = N^{-1}\sum_{\boldsymbol{j}\neq\boldsymbol{i}}\Big(f_N(z,\boldsymbol{j},\boldsymbol{i}) - f_N(z,\boldsymbol{i},\boldsymbol{j})\Big)$$

and to simplify. □

Let $u \in [\![1\,;n]\!]$. To establish an expression for the drift of $X^{(N)}(u)$, we must compute $\sum_{\boldsymbol{i}\in\mathcal{A},i_u=0} B_{\boldsymbol{i}}^{(0)}(z)$ and $\sum_{\boldsymbol{i}\in\mathcal{A},i_u=0} B_{\boldsymbol{i}}^{(1)}(z)$. Direct computations yield:

Lemma 7.2. *For every $u \in [\![1\,;n]\!]$ and $z \in E_N$,*

$$\sum_{\boldsymbol{i}\in\mathcal{A},\ i_u=0} B^{(0)}_{\boldsymbol{i}}(z) = 0, \tag{7.6}$$

$$\sum_{\boldsymbol{i}\in\mathcal{A},\ i_u=0} B^{(1)}_{\boldsymbol{i}}(z) = (1-x(u))\mu_1 - x(u)\mu_0 + \frac{1}{2}G_u(z), \tag{7.7}$$

where

$$x(u) = \sum_{\boldsymbol{i}\in\mathcal{A},\ i_u=0} z(\boldsymbol{i}) \quad \textit{and} \quad G_u(z) = \sum_{\boldsymbol{j}\in\mathcal{A}}\sum_{\boldsymbol{h}\in\mathcal{A}} z(\boldsymbol{j})z(\boldsymbol{h})s_{\boldsymbol{j},\boldsymbol{h}}(\mathbb{1}_{\{j_u=0\}} - x(u)).$$

Proof. For $\epsilon \in \{0,1\}$ and $\boldsymbol{i} \in \mathcal{A}$, let $\sigma^{(\epsilon)}_u(\boldsymbol{i})$ denote the type $\boldsymbol{i}$ modified by setting the allele ϵ at the locus u. We shall use the following formula several times:

$$\sum_{\boldsymbol{i}\in\mathcal{A},\ i_u=0} q((\boldsymbol{j},\boldsymbol{k});\sigma^{(\epsilon)}_u(\boldsymbol{i})) = \mathbb{1}_{\{j_u=\epsilon\}} + \bar{r}(u)(\mathbb{1}_{\{k_u=\epsilon\}} - \mathbb{1}_{\{j_u=\epsilon\}}) \tag{7.8}$$

with $\bar{r}(u) = \sum_{I\subset[\![1;n]\!]\setminus\{u\}} r_I = \frac{1}{2}$ by assumption H1.
First, formula (7.8) with $\epsilon = 0$ provides

$$\sum_{\boldsymbol{i}\in\mathcal{A},\ i_u=0} B^{(0)}_{\boldsymbol{i}}(z) = \sum_{\boldsymbol{j}\in\mathcal{A},\ j_u=0} z(\boldsymbol{j}) + \bar{r}(u)\sum_{\boldsymbol{j}\in\mathcal{A}}\sum_{\boldsymbol{k}\in\mathcal{A}}(\mathbb{1}_{\{k_u=\epsilon\}} - \mathbb{1}_{\{j_u=\epsilon\}}) - \sum_{\boldsymbol{i}\in\mathcal{A},\ i_u=0} z(\boldsymbol{i}) = 0.$$

Let $B^{(1,j)}_{\boldsymbol{i}}(z)$ denote the j-th line of the expression of $B^{(1)}_{\boldsymbol{i}}(z)$ for $j \in \{1,2,3\}$.
As $\sum_{i\in\mathcal{A},\ i_u=0,i_x=a} q((j,k);\sigma^{\epsilon}_x(\boldsymbol{i}))$ does not depend on the value of a if $u \neq x$:

$$\sum_{\boldsymbol{i}\in\mathcal{A},\ i_u=0} B^{(1,1)}_{\boldsymbol{i}}(z) = \sum_{\boldsymbol{k}\in\mathcal{A}}\sum_{\boldsymbol{j}\in\mathcal{A}} z(\boldsymbol{j})z(\boldsymbol{k}) \sum_{\boldsymbol{i}\in\mathcal{A},\ i_u=0} \Big(q((\boldsymbol{j},\boldsymbol{k});\sigma^{(1)}_u(\boldsymbol{i}))\mu_1 - q((\boldsymbol{j},\boldsymbol{k});\sigma^{(0)}_u(\boldsymbol{i}))\mu_0\Big).$$

Applying (7.8) again, we obtain:

$$\sum_{\boldsymbol{i}\in\mathcal{A},\ i_u=0} B^{(1,1)}_{\boldsymbol{i}}(z) = (1-x(u))\mu_1 - x(u)\mu_0.$$

Due to the symmetry of the parameters: $s_{\boldsymbol{i},\boldsymbol{j}} = s_{\boldsymbol{j},\boldsymbol{i}}$ for $\boldsymbol{i},\boldsymbol{j} \in \mathcal{A}$, we have:

$$\sum_{\boldsymbol{i}\in\mathcal{A},\ i_u=0} B^{(1,2)}_{\boldsymbol{i}}(z) = 0.$$

Finally, computations using (7.8) yet again yield:

$$\sum_{\boldsymbol{i}\in\mathcal{A},\ i_u=0} B^{(1,3)}_{\boldsymbol{i}}(z) = \frac{1}{2}G_u(z).$$

□

To obtain condition (a), it remains to express $G_u(z)$ in the new coordinates. The following lemma describes the inverse of the change of coordinates (Φ_N, Ψ_N):

Lemma 7.3. *For $z \in E_N$ and $L \subset [\![1\,;n]\!]$, set $x(L) = \sum_{\boldsymbol{i},\ i_{|L}\equiv 0} z(\boldsymbol{i})$ with the convention $x(\emptyset) = 1$ and $y(L) = \prod_{\ell\in L} x(\ell) - x(L)$ if $|L| \geq 2$. Then for every $J \subset [\![1\,;n]\!]$,*

$$z(\mathbf{0}_J, \mathbf{1}_{\bar{J}}) = \prod_{i\in J} x(i) \prod_{i\in\bar{J}} (1 - x(i)) - \sum_{I\subset[\![1;n]\!] \text{ s. t. } J\subset I,\ |I|\geq 2} (-1)^{|I|-|J|} y(I). \tag{7.9}$$

Proof. First, by induction on $n - |J|$, we show that

$$z(\mathbf{0}_J, \mathbf{1}_{\bar{J}}) = \sum_{I\subset[\![1;n]\!] \text{ s. t. } J\subset I} (-1)^{|I|-|J|} x(I). \tag{7.10}$$

Since $z(\mathbf{0}) = x([\![1\,;n]\!])$, the equality (7.10) holds for $J = [\![1\,;n]\!]$.
Let $m \in [\![1\,;n]\!]$. Assume that the formula (7.10) holds for every subset J of $[\![1\,;n]\!]$ such that $|J| \geq m$. Let K be a subset of $[\![1\,;n]\!]$ with $m - 1$ elements.

$$z(\mathbf{0}_K, \mathbf{1}_{\bar{K}}) = x(K) - \sum_{L\subset[\![1;n]\!] \text{ s. t. } K\subsetneq L} z(\mathbf{0}_L, \mathbf{1}_{\bar{L}})$$

We apply the formula (7.10) to every term in the sum and we invert the double sum we have obtained:

$$z(\mathbf{0}_K, \mathbf{1}_{\bar{K}}) = x(K) - \sum_{H\subset[\![1;n]\!] \text{ s. t. } K\subsetneq H} x(H)\Big(\sum_{L\subset[\![1;n]\!] \text{ s. t. } K\subsetneq L\subset H} (-1)^{|H|-|L|}\Big).$$

The sum between parentheses is equal to

$$\sum_{v=1}^{|H|-|K|} (-1)^{|H|-|K|-v} \binom{|H| - |K|}{v} = -(-1)^{|H|-|K|}.$$

Thus the formula (7.10) is also satisfied for the subset K which completes the induction.
To complete the proof, we replace $x(I)$ in (7.10) with $\prod_{i\in I} x(i) - y(I)$ for every subset I having at least two elements and use the following equality:

$$\sum_{I\subset[\![1;n]\!],\ J\subset I} (-1)^{|I|-|J|} \prod_{i\in I} x(i) = \prod_{j\in J} x(j)\Big(\sum_{L\subset[\![1;n]\!]\setminus J} (-1)^{|L|} \prod_{\ell\in L} x(\ell)\Big) = \prod_{j\in J} x(j) \prod_{i\in[\![1;n]\!]\setminus J} (1 - x(i)).$$

□

To shorten the notation, set

- $\Lambda_u = [\![1\,;n]\!] \setminus \{u\}$ for $u \in [\![1\,;n]\!]$
- $\Pi_J(v) = \prod_{j\in J} v(j)$ for $v \in [0,1]^n$ and $J \in \mathcal{P}([\![1\,;n]\!])$ with the usual convention $\Pi_\emptyset = 1$,
- $s_{I,J} = s_{\boldsymbol{i},\boldsymbol{j}}$ for $\boldsymbol{i} = (\mathbf{0}_I, \mathbf{1}_{\bar{I}})$ and $\boldsymbol{j} = (\mathbf{0}_J, \mathbf{1}_{\bar{J}})$.

With this notation, for every $J \subset \Lambda_u$,

- $z(\mathbf{0}_J, \mathbf{1}_{\bar{J}}) = (1-x(u))\Pi_J(x)\Pi_{\Lambda_u\setminus J}(1-x) - R_J(y)$,
- $z(\mathbf{0}_{J\cup\{u\}}, \mathbf{1}_{\overline{J\cup\{u\}}}) = x(u)\Pi_J(x)\Pi_{\Lambda_u\setminus J}(1-x) - R_{J\cup\{u\}}(y)$,

where $R_J(y)$ and $R_{J\cup\{u\}}(y)$ denote polynomial functions that vanish at $y \equiv 0$. Therefore,

$$G_u(z) = x(u)(1-x(u)) \sum_{J\subset\Lambda_u}\sum_{H\subset\Lambda_u} \Pi_J(x)\Pi_H(x)\Pi_{\Lambda_u\setminus J}(1-x)\Pi_{\Lambda_u\setminus H}(1-x)\times \Big(x(u)(s_{J\cup\{u\},H\cup\{u\}} - s_{J,H\cup\{u\}}) + (1-x(u))(s_{J\cup\{u\},H} - s_{J,H})\Big) + R_u(x,y),$$

where $R_u(x,y)$ is a polynomial function in the variables $x(1),\dots,x(n)$ and $y(I)$ for $I \subset [\![1\,;n]\!]$ such that $|I| \geq 2$, that vanishes in the equilibrium manifold: $R_u(x,0) = 0$.
The expression for $G_u(z)$ can be simplified by using the two assumptions H4 on the assortment parameters, that is $s_{J,H} = s_{H,J}$ for every $J, H \subset [\![1\,;n]\!]$ and $s_{J\cup\{u\},H\cup\{u\}} = s_{J,H}$ for every $u \in [\![1\,;n]\!]$ and $J, H \subset \Lambda_u$:

$$G_u(z) = (1-2x(u))x(u)(1-x(u))\times \sum_{J\subset\Lambda_u}\sum_{H\subset\Lambda_u} \Pi_J(x)\Pi_H(x)\Pi_{\Lambda_u\setminus J}(1-x)\Pi_{\Lambda_u\setminus H}(1-x)(s_{J\cup\{u\},H} - s_{J,H}) + R_u(x,y).$$

In summary, we have established the following expansion of the drift of $X^{(N)}$:

Lemma 7.4. *Assume that hypotheses H1, H2, H3 and H4 hold. For every $i \in [\![1\,;n]\!]$,*

$$N^2\,\mathbb{E}_z[X^{(N)}(i) - x(i)] = (1-x(i))\mu_1 - x(i)\mu_0 + (\frac{1}{2} - x(i))x(i)(1-x(i))P_{i,s}(x) + R_i(x,y) + O(N^{-1}) \quad (7.11)$$

uniformly on $z \in E_N$ where

$$P_{i,s}(x) = \sum_{J\subset\Lambda_u}\sum_{H\subset\Lambda_u} \Pi_J(x)\Pi_H(x)\Pi_{\Lambda_u\setminus J}(1-x)\Pi_{\Lambda_u\setminus H}(1-x)(s_{J\cup\{u\},H} - s_{J,H})$$

and $R_i(x,y)$ is a polynomial function in the variables $x(1),\dots,x(n)$ and $y(I)$ for $I \in \mathcal{P}([\![1\,;n]\!])$ with at least two elements such that $R_i(x,\mathbf{0}) = 0$.

Condition (b). Computations similar to those used to obtain (7.6) lead to the following expansion of the second moments of $X_1^{(N)} - x$, showing that condition (b) holds:

Lemma 7.5. $N^2\,\mathbb{E}_z\left[(X_1^{(N)}(i) - x(i))(X_1^{(N)}(j) - x(j))\right] = a_{i,j}(x,y) + O(N^{-1})$, *with*

$$\begin{cases} a_{i,i}(x,y) = x(i)(1-x(i)) + O(N^{-1}) \\ a_{i,j}(x,y) = -2\Big(\sum\limits_{I \subset [\![1;n]\!] \setminus \{i,j\}} r_I\Big) y(\{i,j\}) + O(N^{-1}) \ \text{ if } i \neq j \end{cases}$$

uniformly on $z \in E_N$.

Proof. Let $i, j \in [\![1\,;n]\!]$ and $z \in E_N$. By definition of $X^{(N)}$,

$$N^2\,\mathbb{E}_z\left[(X_1^{(N)}(i) - x(i))(X_1^{(N)}(j) - x(j))\right] = N^2 \sum_{\boldsymbol{k}\in\mathcal{A},\ k_i=0} \ \sum_{\boldsymbol{\ell}\in\mathcal{A},\ \ell_j=0} \mathbb{E}_z\left[(Z_1^{(N)}(\boldsymbol{k}) - z(\boldsymbol{k}))(Z_1^{(N)}(\boldsymbol{\ell}) - z(\boldsymbol{\ell}))\right]$$

Using formulae (7.3) and (7.4) and assumption H2, we obtain

$$\begin{aligned} N^2\,\mathbb{E}_z\left[(X_1^{(N)}(i) - x(i))(X_1^{(N)}(j) - x(j))\right] &= \sum_{\boldsymbol{k}\in\mathcal{A}} \sum_{\boldsymbol{\ell}\in\mathcal{A}} (f_N(z,\boldsymbol{\ell},\boldsymbol{k}) + f_N(z,\boldsymbol{k},\boldsymbol{\ell}))(\mathbb{1}_{\{k_i=0,k_j=0\}} - \mathbb{1}_{\{k_i=0,\ell_j=0\}}) \\ &= T_{i,j}^{(1)} + T_{i,j}^{(2)} - T_{i,j}^{(3)} - T_{j,i}^{(3)} + O(N^{-1}), \end{aligned}$$

where

$$\begin{aligned} T_{i,j}^{(1)} &= \sum_{\boldsymbol{t}\in\mathcal{A}} z(\boldsymbol{t}) \sum_{\boldsymbol{\ell}\in\mathcal{A}} z(\boldsymbol{\ell}) \sum_{\boldsymbol{k}\in\mathcal{A},\ k_i=k_j=0} q((\boldsymbol{\ell},\boldsymbol{t});\boldsymbol{k}), \\ T_{i,j}^{(2)} &= \sum_{\boldsymbol{t}\in\mathcal{A}} z(\boldsymbol{t}) \sum_{\boldsymbol{k}\in\mathcal{A},\ k_i=k_j=0} z(\boldsymbol{k}) \sum_{\boldsymbol{\ell}\in\mathcal{A}} q((\boldsymbol{k},\boldsymbol{t});\boldsymbol{\ell}), \\ T_{i,j}^{(3)} &= \sum_{\boldsymbol{t}\in\mathcal{A}} z(\boldsymbol{t}) \sum_{\boldsymbol{\ell}\in\mathcal{A},\ \ell_j=0} z(\boldsymbol{\ell}) \sum_{\boldsymbol{k}\in\mathcal{A},\ k_i=0} q((\boldsymbol{\ell},\boldsymbol{t});\boldsymbol{k}). \end{aligned}$$

With the convention $x(\{i,j\}) = x(i)$ if $i = j$, we have $T_{i,j}^{(2)} = x(\{i,j\})$ and it follows from assumption H1 ($r_I = r_{\bar{I}}$ for every $I \subset [\![1\,;n]\!]$) that

$$\begin{aligned} T_{i,j}^{(1)} &= x(i)x(j) + \sum_{I\subset[\![1;n]\!]} r_I\big(\mathbb{1}_{\{i\in I, j\in I\}} + \mathbb{1}_{\{i\notin I, j\notin I\}}\big)\big(x(\{i,j\}) - x(i)x(j)\big) \\ &= x(i)x(j) + 2\Big(\sum_{I\subset[\![1;n]\!]\setminus\{i,j\}} r_I\Big)\Big(x(\{i,j\}) - x(i)x(j)\Big), \\ T_{i,j}^{(3)} &= x(\{i,j\}) + \Big(\sum_{I\subset[\![1;n]\!]\setminus\{i\}} r_I\Big)\big(x(i)x(j) - x(\{i,j\})\big) = \frac{1}{2}\big(x(i)x(j) + x(\{i,j\})\big). \end{aligned}$$

Therefore, for every $i, j \in [\![1\,;n]\!]$,

$$N^2\,\mathbb{E}_z\left[(X_1^{(N)}(i)-x(i))(X_1^{(N)}(j)-x(j))\right] \\ = 2\Big(\sum_{I\subset[\![1;n]\!]\setminus\{i,j\}} r_I\Big)\big(x(\{i,j\})-x(i)x(j)\big)+O(N^{-1}).$$

If $i=j$ then $x(\{i,j\})-x(i)x(j)=x(i)(1-x(i))$ and $\displaystyle\sum_{I\subset[\![1;n]\!]\setminus\{i,j\}} r_I=\frac{1}{2}$. □

Condition (d). Let I be a subset of $[\![1\,;n]\!]$ with at least two elements. To compute the drift of $Y^{(N)}(I)$, we use the following lemma and formulae (7.3), (7.4) and (7.5) describing the moments of $Z_1^{(N)}-z$.

Lemma 7.6. *Let J be a finite set. Consider two families of reals $\{a_j, j\in J\}$ and $\{b_j, j\in J\}$. The following identity holds:*

$$\prod_{j\in J} a_j - \prod_{j\in J} b_j = \sum_{K\subset J,\ K\neq\emptyset}\ \prod_{k\in K}(a_k-b_k)\prod_{\ell\in J\setminus K} b_\ell. \tag{7.12}$$

Computations yield:

$$N\,\mathbb{E}_z[Y_1^{(N)}(I)-y(I)] = \sum_{i\in I}\Big(\prod_{\ell\in I\setminus\{i\}} x(\ell)\sum_{\boldsymbol{j}\in\mathcal{A},\ j_i=0} B_{\boldsymbol{j}}^{(0)}(z)\Big) \\ -\sum_{\boldsymbol{j}\in\mathcal{A},\ \boldsymbol{j}_{|I}\equiv 0} B_{\boldsymbol{j}}^{(0)}(z)+O(N^{-1}). \tag{7.13}$$

uniformly on $z\in E_N$. As we have shown that $\sum_{\boldsymbol{j}\in\mathcal{A},\ j_i=0} B_{\boldsymbol{j}}^{(0)}(z)=0$ for every $i\in[\![1\,;n]\!]$ (equation (7.6)),

$$N\,\mathbb{E}_z[Y_1^{(N)}(I)-y(I)] = -\sum_{\boldsymbol{j}\in\mathcal{A},\ \boldsymbol{j}_{|I}\equiv 0} B_{\boldsymbol{j}}^{(0)}(z)+O(N^{-1}) \tag{7.14}$$

uniformly on $z\in E_N$.

Direct computations provide the following expression of the sum on the right-hand side of (7.14) using the variables $x(L)=\sum_{\boldsymbol{j}\in\mathcal{A},\ j_{|L}\equiv 0} x(j)$ for $L\in\mathcal{P}([\![1\,;n]\!])$:

$$\sum_{\boldsymbol{j}\in\mathcal{A},\ \boldsymbol{j}_{|I}\equiv 0} B_{\boldsymbol{j}}^{(0)}(z) = \sum_{L\subset[\![1;n]\!]\text{ s. t. }I\cap L\neq\emptyset,\ I\cap\bar{L}\neq\emptyset} r_L\Big(x(I\cap L)x(I\cap\bar{L})-x(I)\Big) \tag{7.15}$$

To obtain an expression for $\mathbb{E}_z[Y_1^{(N)}(I)-y(I)]$ in the new coordinates, it remains to replace each term $x(L)$ for $|L|\geq 2$ with $\prod_{\ell\in L}x(\ell)-y(L)$ in (7.15). This leads to the following lemma and shows that condition (d) holds.

Lemma 7.7. *For a subset I of $[\![1\,;n]\!]$ having at least two elements,*

$$N\,\mathbb{E}_z[Y_1^{(N)}(I) - y(I)] = c_{n,I}(x,y) + O(N^{-1}) \tag{7.16}$$

where

$$c_{n,I}(x,y) = -\Big(\sum_{\substack{L\subset[\![1;n]\!],\\ L\cap I\neq\emptyset,\ \bar{L}\cap I\neq\emptyset}} r_L\Big)y(I) - \mathbb{1}_{\{|I|\geq 4\}} \sum_{\substack{L\subset[\![1;n]\!],\\ |I\cap L|\geq 2, |I\cap\bar{L}|\geq 2}} r_L y(I\cap L)y(I\cap\bar{L})$$
$$+ \mathbb{1}_{\{|I|\geq 3\}} \sum_{\substack{L\subset[\![1;n]\!],\\ |I\cap L|\geq 2, |I\cap\bar{L}|\geq 1}} (r_L + r_{\bar{L}})y(L\cap I) \prod_{\ell\in I\cap\bar{L}} x(\ell).$$

Condition (f). The following lemma shows that the condition (f) holds under the assumption H3:

Lemma 7.8. *For two distinct loci k,ℓ, let $r_{k,\ell}$ denote the probability that the offspring does not inherit the genes at the loci k and ℓ from the same parent,*

$$r_{k,\ell} = \sum_{I\subset[\![1;n]\!],\ k\in I \text{ and } \ell\notin I} (r_I + r_{\bar{I}}),$$

and set $r(n) = \min(r_{k,h}\ k,h\in[\![1\,;n]\!] \text{ and } h\neq k)$.
If $r(n) > 0$ then the following system of differential equations

$$(S_{n,I})\begin{cases} \frac{dv_{n,I}}{dt}(t,x,y) = c_{n,I}(x, v_{n,I}(t,x,y)) \\ v_{n,I}(0,x,y) = y(I) \end{cases} \quad \forall I\subset[\![1\,;n]\!] \text{ s. t. } |I|\geq 2$$

has a unique solution $v_n = \{v_{n,I},\ I\subset[\![1\,;n]\!] \text{ and } |I|\geq 2\}$ which is of the form:

$$v_{n,I}(t,x,y) = \exp(-r(n)t) f_{n,I}(t,x,y),$$

where $f_{n,I}$ is a continuous and bounded function on $\mathbb{R}\times[0,1]^n\times[-1,1]^{2^n-n-1}$ so that the value of $f_{n,I}(t,x,y)$ depends on x and y only via the coordinates $x(i)$ for $i\in I$ and $y(J)$ for $J\subset I$ such that $|J|\geq 2$.

Remark 7.2. For every subset $I\subset[\![1\,;n]\!]$ with two elements say k and ℓ,

$$\frac{dv_{n,I}}{dt}(t,x,y) = -r_{k,\ell}\ v_{n,I}(t,x,y).$$

Therefore if $r(n) = 0$ then there exists a subset I of $[\![1\,;n]\!]$ with two elements such that $v_{n,I}(t,x,y) = y(I)$. Thus the assumption $r(n) > 0$ is a necessary condition for the solution of $(S_{n,I})$ to converge to 0 as t tends to $+\infty$ for any initial values.

Proof. Let $n \geq 2$ and let $I \subset [\![1\,;n]\!]$ be such that $|I| \geq 2$. As $c_{n,I}(x,y)$ depends only on the coordinates $x(\ell)$ for $\ell \in I$ and $y(L)$ for $L \subset I$ such that $|L| \geq 2$, we shall prove by induction on the number of elements of I that for any $J \subset I$, $(S_{n,J})$ has a unique solution of the form $v_{n,J}(t,x,y) = \exp(-r(n)t) f_{n,J}(t,x,y)$, where $f_{n,J}$ is a continuous and bounded function on $\mathbb{R} \times [0,1]^n \times [-1,1]^{2^n-n-1}$ such that the value of $f_{n,J}(t,x,y)$ depends on x and y only through the values of the coordinates $x(j)$ for $j \in J$ and $y(L)$ for $L \subset J$ such that $|L| \geq 2$.

- If I has two elements say k and ℓ, then $(S_{n,I})$ is the following differential equation:

$$\begin{cases} \frac{dv_{n,I}}{dt}(t,x,y) = -r_{k,\ell}\, v_{n,I}(t,x,y) \\ v_{n,I}(0,x,y) = y(I) \end{cases}$$

  It has a unique solution $v_{n,I}(t,x,y) = y(I)e^{-r(2)t} f_{n,I}(t,x,y)$ where

$$f_{n,I}(t,x,y) = e^{-(r_{k,\ell}-r(2))t} y(I).$$

  By assumption $r(k,\ell) \geq r(2) > 0$, hence the result holds.

- Let $2 \leq m < n$. Assume that the inductive hypothesis holds for any subsets J with m elements. Let I be a subset of $[\![1\,;n]\!]$ with $m+1$ elements. Then

$$\frac{dv_{n,I}}{dt}(t,x,y) = -\bar{r}_I v_{n,I}(t,x,y) + e^{-tr(n)} g(t,x,y)$$

  where $\bar{r}_I = \sum_{\substack{L \subset [\![1;n]\!] \text{ s. t.} \\ L \cap I \neq \emptyset,\ \bar{L} \cap I \neq \emptyset}} r_L$ and

$$\begin{aligned} g(t,x,y) = &- \mathbb{1}_{\{|I| \geq 4\}} \sum_{\substack{L \subset [\![1;n]\!] \text{ s. t.} \\ |I \cap L| \geq 2, |I \cap \bar{L}| \geq 2}} r_L e^{-tr(n)} f_{n,I\cap L}(t,x,y) f_{n,I\cap \bar{L}}(t,x,y) \\ &+ \mathbb{1}_{\{|I| \geq 3\}} \sum_{\substack{L \subset [\![1;n]\!] \text{ s. t.} \\ |I \cap L| \geq 2, |I \cap \bar{L}| \geq 1}} (r_L + r_{\bar{L}}) f_{n,L\cap I}(t,x,y) \prod_{\ell \in I \cap \bar{L}} x(\ell). \end{aligned}$$

  As $\bar{r}_I$ is the probability that the offspring does not inherit all the genes at loci $i \in I$ from the same parent, $\bar{r}_I \geq r(n)$. Therefore the differential equation $(S_{n,I})$ has a unique solution:

$$v_{n,I}(t,x,y) = y(I)e^{-\bar{r}_I t} + e^{-\bar{r}_I t} \int_0^t g(s,x,y) e^{(\bar{r}_I - r(n))s} ds.$$

  By our assumptions on the functions $f_{n,J}$ for $J \subsetneq I$, g is a bounded continuous function on $\mathbb{R}_+ \times [0,1]^n \times [-1,1]^{2^n-n-1}$ such that the value of $g(t,x,y)$ depends on x and y only through the coordinates $x(i)$ for $i \in I$ and $y(L)$ for $L \subset I$ such that $|L| \geq 2$. Therefore, the function $f_{n,I}(t,x,y) = e^{r(n)t} v_{n,I}(t,x,y)$ has the asserted properties.

Conditions (c) and (e). Condition (c) is easy to verify using formulae (7.3), (7.4), (7.5) describing the moments of $Z_1^{(N)} - z$. This leads to:

$$N^2\,\mathbb{E}_z[(X_1^{(N)}(i) - x(i))^4] = O(N^{-2}) \quad \forall i \in [\![1\,;n]\!],\ \text{uniformly on } z \in E_N.$$

Similarly, using Lemma 7.6, we obtain

$$N\,\mathbb{E}_z[(Y_1^{(N)}(I) - y(I))^2] = O(N^{-1}) \quad \forall I \subset [\![1\,;n]\!],\ s.t.\ |I| \geq 2,\ \text{uniformly on } z \in E_N.$$

□

7.2 Expressions for the drift

We have shown that the i-th coordinate of the drift of the limiting diffusion is

$$(1 - x(i))\mu_1 - x(i)\mu_0 + (1/2 - x(i))x(i)(1 - x(i))P_{i,s}(x)$$

where

$$P_{i,s}(x) = \sum_{J \subset [\![1;n]\!]\setminus\{i\}} \sum_{H \subset [\![1;n]\!]\setminus\{i\}} (s_{J\cup\{i\},H} - s_{J,H}) \times \prod_{j\in J} x(j) \prod_{h\in H} x(h) \prod_{\substack{j\in[\![1;n]\!],\\ j\notin J\cup\{i\}}} (1 - x(j)) \prod_{\substack{h\in[\![1;n]\!],\\ h\notin H\cup\{i\}}} (1 - x(h)),$$

and, for two subsets I and J of $[\![1\,;n]\!]$, $s_{I,J}$ denotes the assortment parameter $s_{\boldsymbol{i},\boldsymbol{j}}$ for the types $\boldsymbol{i} = (\mathbf{0}_I, \mathbf{1}_{\bar{I}})$ and $\boldsymbol{j} = (\mathbf{0}_J, \mathbf{1}_{\bar{J}})$. The following lemma states that $P_{i,s}(x)$ is actually a polynomial function in the variables $x(i)(1 - x(i))$ for $i \in [\![1\,;n]\!] \setminus \{u\}$:

Lemma 7.9. *Let Λ be a finite subset of $\mathbb{N}$. Consider a family of reals $\beta = \{\beta_{I,J},\ I, J \subset \Lambda\}$ such that $\beta_{I,J} = \beta_{I\setminus J, J\setminus I}$ for every $I, J \subset \Lambda$. Then,*

$$\sum_{J\subset\Lambda}\sum_{H\subset\Lambda}\left(\beta_{J,H}\prod_{j\in J} x(j)\prod_{h\in H} x(h)\prod_{j\in\Lambda\setminus J}(1 - x(j))\prod_{h\in\Lambda\setminus H}(1 - x(h))\right) = \sum_{L\subset\Lambda} C_L(\beta)\prod_{\ell\in L} x(\ell)(1 - x(\ell)) \quad (7.17)$$

where

$$C_L(\beta) = \sum_{T\subset L}(-2)^{|T|-|L|}\sum_{A\subset T}\beta_{A,T\setminus A}.$$

Proof. Let $P_\Lambda(\beta)$ denote the polynomial function on the right-hand side. The proof is by induction on $|\Lambda|$. First, $P_\emptyset(\beta)(x) = \beta_{\emptyset,\emptyset} = C_\emptyset(\beta)$.
Let $n \in \mathbb{N}$. Assume that the equality (7.17) holds for every subset Λ of $\mathbb{N}$ with at most n elements and every family of reals β satisfying the assumptions of the lemma.
Let Λ be a subset of $\mathbb{N}$ with $n+1$ elements, let j be an element of Λ and let $\eta = \{\eta_{I,J},\ I, J \subset \Lambda\}$ be a family of reals such that $\eta_{I,J} = \eta_{I\setminus J, J\setminus I}$ for every $I, J \subset \Lambda$. We split $P_\Lambda(\eta)$ into a sum over the subsets of Λ containing j and a sum over the subsets of $\Lambda \setminus \{j\}$ to obtain the following expression:

$$P_\Lambda(\eta)(x) = \sum_{K\subset\Lambda\setminus\{j\}} \sum_{L\subset\Lambda\setminus\{j\}} \prod_{k\in K} x(k) \prod_{\ell\in L} x(\ell) \prod_{k\in\Lambda\setminus K} (1-x(k)) \prod_{h\in\Lambda\setminus L} (1-x(h)) \times \Big(x(j)^2 \eta_{K\cup\{j\},L\cup\{j\}} + (1-x(j))^2 \eta_{K,L} + x(j)(1-x(j))(\eta_{K\cup\{j\},L} + \eta_{K,L\cup\{j\}}) \Big).$$

This expression can be simplified by using that $\eta_{K\cup\{j\},L\cup\{j\}} = \eta_{K,L}$:

$$P_\Lambda(\eta)(x) = P_{\Lambda\setminus\{j\}}(\eta^{(0)})(x) + x(j)(1-x(j)) \big(P_{\Lambda\setminus\{j\}}(\eta^{(1)})(x) + P_{\Lambda\setminus\{j\}}(\eta^{(2)})(x) - 2P_{\Lambda\setminus\{j\}}(\eta^{(0)})(x) \big),$$

where $\eta^{(0)}$, $\eta^{(1)}$ and $\eta^{(2)}$ are the following three families of reals indexed by the pairs of subsets of $\Lambda \setminus \{j\}$:

$$\eta^{(0)}_{A,B} = \eta_{A,B}, \quad \eta^{(1)}_{A,B} = \eta_{A\cup\{j\},B} \text{ and } \eta^{(2)}_{A,B} = \eta_{A,B\cup\{j\}} \text{ for every } A, B \subset \Lambda \setminus \{j\}.$$

The inductive hypothesis applies to $\Lambda \setminus \{j\}$ and the three families of reals $\eta^{(0)}$, $\eta^{(1)}$ and $\eta^{(2)}$:

$$P_\Lambda(\eta)(x) = \sum_{L\subset\Lambda\setminus\{j\}} C_L(\eta) \prod_{\ell\in L} x(\ell)(1-x(\ell)) + \sum_{L\subset\Lambda,\ j\in L} \tilde{C}_L \prod_{\ell\in L} x(\ell)(1-x(\ell)),$$

where

$$\tilde{C}_L = \sum_{T\subset L\setminus\{j\}} (-2)^{|L|-1-|T|} \sum_{A\subset T} (\eta_{A\cup\{j\},T\setminus A} + \eta_{A,(T\cup\{j\})\setminus A} - 2\eta_{A,T\setminus A}).$$

The double sum of the terms $\eta_{A\cup\{j\},T\setminus A} + \eta_{A,(T\cup\{j\})\setminus A}$ is equal to:

$$\sum_{T\subset L,\ j\in T} (-2)^{|L|-|T|} \sum_{A\subset T} \eta_{A,T\setminus A}.$$

Therefore, $\tilde{C}_L = C_L(\eta)$ and $P_\Lambda(\eta)(x) = \sum_{L\subset\Lambda} C_L(\eta) \prod_{\ell\in L} x(\ell)(1-x(\ell))$ which completes the proof by induction. □

By Lemma 7.9, the expanded form of $P_{i,s}$ as a polynomial function of the $n-1$ variables $x(j)(1-x(j))$, $j \neq i$ is:

$$P_{i,s}(x) = \sum_{L \subset [\![1;n]\!] \setminus \{i\}} \alpha_{i,L}(s) \prod_{\ell \in L} x(\ell)(1-x(\ell)) \tag{7.18}$$

where

$$\alpha_{i,L}(s) = \sum_{T \subset L} (-2)^{|L|-|T|} \sum_{A \subset T} (s_{A \cup \{i\}, T \setminus A} - s_{A, T \setminus A}).$$

The coefficient $\alpha_{i,L}(s)$ can be rewritten in terms of the mean values of the assortment parameters $m_T(s)$ for $T \subset L$:

$$\alpha_{i,L}(s) = 2^{|L|} \sum_{T \subset L} (-1)^{|L|-|T|} (m_{T \cup \{i\}}(s) - m_T(s)) = 2^{|L|} \sum_{T \subset L} (-1)^{|L|-|T|} \delta_i[m(s)](T).$$

Indeed, it follows from the assumption H4 that for every $i \in [\![1\,;n]\!]$ and $T \subset [\![1\,;n]\!] \setminus \{i\}$,

$$m_T(s) = 2^{-|T|} \sum_{A \subset T} s_{A, T \setminus A} \text{ and } m_{T \cup \{i\}}(s) = 2^{-|T|} \sum_{A \subset T} s_{A \cup \{i\}, T \setminus A}.$$

Using formula (4.5), we obtain $\alpha_{i,L}(s) = 2^{|L|} \delta_{L \cup \{i\}}[m(s)](\emptyset)$.

The following factorised form of the polynomial function $P_{i,s}$ can be derived from a general identity stated in Lemma A.1:

$$P_{i,s}(x) = \sum_{A \subset [\![1;n]\!] \setminus \{i\}} \delta_i[m(s)](A) \prod_{k \in A} 2x(k)(1-x(k)) \prod_{\ell \notin A \cup \{i\}} \big(1 - 2x(\ell)(1-x(\ell))\big).$$

A Appendix

A.1 Combinatorial formulae for difference operators

This section collects some combinatorial formulae used to study the limiting diffusion. Let E be a finite set and t be a real. For a function f defined on $\mathcal{P}(E)$, we set

$$S_t(f)(A) = \sum_{B \subset A} t^{|A|-|B|} f(B) \text{ for every } A \in \mathcal{P}(E)$$

(with the usual convention $a^0 = 1$ for every $a \in \mathbb{R}$). Most of the combinatorial formulae used in the paper can be deduced from this general identity:

Lemma A.1. *Let U be a subset of E and let $\{x_u, u \in U\}$ be a family of reals.*

$$\sum_{A \subset U} S_t(f)(A) \prod_{i \in A} x_i = \sum_{B \subset U} f(B) \prod_{i \in B} x_i \prod_{j \in U \setminus B} (1 + t x_j). \tag{A.1}$$

Proof. One way to derive this equality is to interchange the sum on the right-hand side of the equation with the sum that appears in the definition of $S_t(f)(A)$, to use the new summation index $C = A \setminus B$ and to recognize the following expansion of the product of the terms $1 + tx_i$:

$$\prod_{i \in U \setminus B} (1 + tx_i) = \sum_{C \subset U \setminus B} t^{|C|} \prod_{i \in C} x_i.$$

□

As $S_{-1}(f)(A)$ is nothing other than $\delta_A[f](\emptyset)$ by (4.5), if we apply Lemma A.1 with $t = -1$, $f(A) = \delta_i[m(s)](A)$ and the family of reals $\{2x(j)(1 - x(j)),\ j \in [\![1\,;n]\!] \setminus \{i\}\}$, we obtain the following equality

$$\sum_{A \subset [\![1;n]\!] \setminus \{i\}} 2^{|A|} \delta_{A \cup \{i\}}[m(s)](\emptyset) \prod_{\ell \in A} x(\ell)(1 - x(\ell))$$
$$= \sum_{A \subset [\![1;n]\!] \setminus \{i\}} \delta_i[m(s)](A) \prod_{k \in A} 2x(k)(1 - x(k)) \prod_{\ell \notin A \cup \{i\}} \Big(1 - 2x(\ell)(1 - x(\ell))\Big).$$

This shows the equality between the expanded form (4.3) and factorised form (4.1) of the polynomial term $P_{i,s}(x)$ appearing in the drift of the limiting diffusion.

By taking $x_i = -1/t$ for every $i \in U$ in Lemma A.1, we can deduce the inverse of the operator S_t. This gives a useful formula for inverting a relation between two sequences indexed by the subsets of a finite set.

Corollary A.1. *The inverse of the operator S_t is S_{-t}, that is*

$$f(A) = \sum_{B \subset A} (-t)^{|A| - |B|} S_t(f)(B) \text{ for every } A \subset E.$$

From Corollary A.1 we can deduce the following identity for the finite difference operator:

$$f(A) = \sum_{B \subset A} \delta_B[f](\emptyset) \text{ for every } A \in \mathcal{P}(E). \tag{A.2}$$

By considering the operator S_t for a function f which is constant on subsets having the same number of elements, we can rewrite the previous relations to obtain useful formulae relating two sequences indexed by the integers $0, 1, \ldots, n$.

Corollary A.2. *Let t be a real number. Let $n \in \mathbb{N}^*$. For a function f defined on $[\![0\,;n]\!]$, let $s_t(f)$ be the function defined by:*

$$s_t(f)(k) = \sum_{\ell=0}^{k} \binom{k}{\ell} t^{k-\ell} f(\ell) \text{ for every } k \in [\![1\,;n]\!].$$

Then,

1. *For every* $x \in \mathbb{R}^n$

$$\sum_{j=0}^{n} s_t(f)(j) e_{n,j}(x) = \sum_{\ell=0}^{n} f(\ell) \sum_{L \subset [\![1;n]\!] \text{ s. t. } |L|=\ell} \prod_{i \in L} x_i \prod_{j \in [\![1;n]\!] \setminus L} (1 + t x_j)$$

where $e_{n,j}$ *denotes the elementary polynomial of degree* j *in* n *variables:*

$$e_{n,j}(x) = \sum_{J \subset [\![1;n]\!] \text{ s. t. } |J|=j} \prod_{i \in J} x_i.$$

2. *The operator* s_{-t} *is the inverse of the operator* s_t*:*

$$f(k) = \sum_{\ell=0}^{k} \binom{k}{\ell} (-t)^{k-\ell} s_t(f)(\ell) \text{ for every } k \in [\![1\,;n]\!].$$

This corollary provides identities for the forward finite difference operators of any orders since $s_{-1}(f)(k) = \delta^{(k)}[f](0)$ for every $k \in [\![0\,;n]\!]$. In particular, this leads to the following formula used in the proof of Proposition 5.1:

$$\sum_{\ell=0}^{k} \binom{k}{\ell} \delta^{(\ell)}[f](0) = f(k) \text{ for every } k \in [\![1\,;n]\!] \tag{A.3}$$

and Lemma 6.2 used in the proof of Proposition 6.3.

A.2 Example 6.2

Under the hypotheses of the assertion 1-(b) of Proposition 6.3, the logarithm of the stationary density $h_{n,s,\mu}$ takes its maximum value in $[0,1/2]^n$ at a unique point $(\xi_0, \ldots, \xi_0)$ such that $\lambda_0 = \xi_0(1-\xi_0)$ is the unique solution in $]0,1/4[$ of the equation $\mathcal{E}_0'$:

$$2\mu - 1 + \sum_{k=0}^{n-1} 2^k \delta^{(k+1)}[m](0) \binom{n-1}{k} y^{k+1} = 0.$$

In $[0,1/2]^n$ the saddle points of index $n-1$ has one coordinate equal to $1/2$ and $(n-1)$ coordinates equal to ξ_1 where $\lambda_1 = \xi_1(1-\xi_1)$ is the unique solution in $]0,1/4[$ of the equation $\mathcal{E}_1'$:

$$2\mu - 1 + \sum_{k=0}^{n-2} \binom{n-2}{k} \left(2^{k-1} \delta^{(k+2)}[m](0) + 2^k \delta^{(k+1)}[m](0)\right) y^{k+1} = 0.$$

If we denote by $h_{n,i}$ the value of $h_{n,s,\mu}$ at a critical point of index $n-i$ then

$$h_{n,0}-h_{n,n}=(2\mu+1)n\ln(4\lambda_0)+\sum_{k=0}^{n-1}2^k\delta^{(k+1)}[m](0)\binom{n}{k+1}(\lambda_0^{k+1}-(1/4)^{k+1}),$$

$$h_{n,0}-h_{n,1}=(2\mu+1)\Big(n\ln(\frac{\lambda_0}{\lambda_1})+\ln(4\lambda_1)\Big)$$
$$+\sum_{k=0}^{n-1}2^k\delta^{k+1}[m](0)\left(\binom{n-1}{k+1}(\lambda_0^{k+1}-\lambda_1^{k+1})\,\mathbb{1}_{\{k\leq n-2\}}+\binom{n-1}{k}(\lambda_0^{k+1}-\frac{1}{4}\lambda_1^k)\right).$$

If we define the assortment by means of the Hamming criterion with the quadratic sequence of parameters: $s_k=s_0-(bk+ck^2)\ \forall k\in[\![0\,;n]\!]$ with $c>0$ and $b+c>0$, then

$$\delta^{(1)}[m](k)=-(b+c+2kc)\ \forall k\in[\![0\,;n-1]\!],\ \delta^{(2)}[m](0)=-2c \text{ and } \delta^{(r)}[m](0)=0\ \forall r\geq 3.$$

In this case, λ_0 and λ_1 are solutions of quadratic functions: $2\mu-1-(b+c)\lambda_0-4c(n-1)\lambda_0^2=0$ and $2\mu-1-(b+2c)\lambda_1-4c(n-2)\lambda_1^2=0$. After some computations, we obtain: $h_{n,0}-h_{n,n}\underset{n\to+\infty}{\sim}\frac{c}{8}n^2$ and $h_{n,0}-h_{n,1}\underset{n\to+\infty}{\sim}n^{1/2}1/2\sqrt{c(2\mu-1)}$.

A.3 Property of a symmetric matrix

The following lemma is used to determine the nature of the critical points of the density of the invariant measure (Proposition 6.3).

Lemma A.2. *For a real a and two integers k and n so that $n\geq 1$ and $0\leq k\leq n$, let $M_{n,k}(a)$ denote the following symmetric matrix:*

$$M_{n,k}=\begin{pmatrix}\boldsymbol{A}_k & \boldsymbol{B}_{k,n-k}\\ \boldsymbol{B}_{n-k,k} & \boldsymbol{A}_{n-k}\end{pmatrix}$$

where

- *$\boldsymbol{A}_k$ denotes the following k-by-k matrix: $\boldsymbol{A}_k=\begin{pmatrix}1 & a & \cdots & a\\ a & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & a\\ a & \cdots & a & 1\end{pmatrix}$*
- *$\boldsymbol{B}_{k_1,k_2}$ denotes the k_1-by-k_2 matrix all the elements of which are equal to $-a$.*

If $0\leq a<1$ then $M_{n,k}(a)$ is positive definite.

Proof. Let $Q_{n,k,a}$ denote the quadratic form with matrix $M_{n,k}(a)$ in the canonical basis. For every $x \in \mathbb{R}^n$, $Q_{n,k,a}(x) = \sum_{i=1}^n x_i^2 + 2a \sum_{1 \le i < j \le n} \epsilon_i \epsilon_j x_i x_j$, where $\epsilon_1 = \ldots = \epsilon_k = 1$ and $\epsilon_{k+1} = \ldots = \epsilon_n = -1$. This lemma can be established by induction on n by using the following decomposition of $Q_{n,k,a}(x)$:

$$Q_{n,k,a}(x) = (x_n + a\epsilon_n \sum_{i=1}^{n-1} \epsilon_i x_i)^2 + (1-a^2)\Big(\sum_{i=1}^{n-1} x_i^2 + 2b \sum_{1 \le i < j \le n-1} \epsilon_i \epsilon_j x_i x_j \Big).$$

where $b = \frac{a}{1+a} \in [0, 1[$. □

References

Baake, E. and Herms, I. (2008). Single-crossover dynamics: finite versus infinite populations. *Bull. Math. Biol.*, 70(2):603–624.

Bürger, R. (2000). *The mathematical theory of selection, recombination, and mutation*. Wiley Series in Mathematical and Computational Biology. John Wiley & Sons Ltd., Chichester.

Cerrai, S. and Clément, P. (2004). Well-posedness of the martingale problem for some degenerate diffusion processes occuring in dynamics of populations. *Bull. Sci. Math.*, 128:355–389.

Christiansen, F. B. (2000). *Population Genetics of Multiple Loci*. Wiley Series in Mathematical & Computational Biology. John Wiley & Sons Inc., New York.

Devaux, C. and Lande, R. (2008). Incipient allochronic speciation due to non-selective assortative mating by flowering time, mutation and genetic drift. *Proc. R. Soc. B*, 275:2723–2732.

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

Ethier, S. N. and Nagylaki, T. (1980). Diffusion approximations of Markov chains with two time scales and applications to population genetics. *Adv. Appl. Prob.*, 12:14–49.

Ethier, S. N. and Nagylaki, T. (1989). Diffusion approximations of the two-locus Wright-Fisher model. *J. Math Biol.*, 27:17–28.

Fukushima, M. and Stroock, D. (1986). Reversibility of solutions to martingale problems. In *Probability, statistical mechanics, and number theory*, volume 9 of *Adv. Math. Suppl. Stud.*, pages 107–123. Academic Press, Orlando, FL.

Geiringer, H. (1944). On the probability theory of linkage in Mendelian heredity. *Ann. Math. Statistics.*, 15:25–57.

Higgs, P. and Derrida, B. (1992). Genetic distance and species formation in evolving populations. *J. Mol. Evol.*, 35:454–465.

Ikeda, N. and Watanabe, S. (1977). A comparison theorem for solutions of stochastic differential equations and its applications. *Osaka J. Math.*, 14(3):619–633.

Nagylaki, T. (1993). The evolution of multilocus systems under the weak selection. *Genetics*, 134:627–647.

Schneider, K. (2007). Long-term evolution of polygenic traits under frequency-dependent intraspecific competition. *Theor. Pop. Biol.*, 71:342–366.

Yamada, T. and Watanabe, S. (1971). On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11:155–167.