THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'École Normale Supérieure
Dans le cadre d'une cotutelle avec Université de Vienne

Guide d'un probabiliste vers la limite polygénique

A probabilist's roadmap to the polygenic limit

Soutenue par
Philibert COURAU
Le 18 Décembre 2025

Ecole doctorale n° 386
École doctorale de Sciences Mathématiques de Paris Centre

Spécialité
Mathématiques

Composition du jury :

Paul, JENKINS PU, University of Warwick	*Rapporteur*
Charline, SMADI DR INRAE, Université Grenoble Alpes	*Rapporteur*
Amandine, VÉBER DR CNRS, Université Paris Cité	*Examinatrice*
Joachim, HERMISSON PU, Universität Wien	*Examinateur*
Thierry, MORA DR CNRS, École Normale Supérieure	*Examinateur*
Luis-Miguel, CHEVIN DR CNRS, CEFE	*Examinateur*
Emmanuel, SCHERTZER PU, Universität Wien	*Co-directeur*
Amaury, LAMBERT PU, École Normale Supérieure	*Co-directeur*

IBENS, Collège-de-France & University of Vienna

ED386

Équipe SMILE

A probabilist's roadmap to the polygenic limit

par
Philibert COURAU

Thèse de doctorat de Mathématiques

dirigée par
Amaury LAMBERT & Emmanuel SCHERTZER

Présentée et soutenue publiquement le Jeudi 18 Décembre 2025

Devant un jury composé de :

Paul JENKINS	PU, University of Warwick	Rapporteur
Charline SMADI	DR INRAE, Université Grenoble Alpes	Rapporteuse
Luis-Miguel CHEVIN	DR CNRS, CEFE	Examinateur
Joachim HERMISSON	PU Universität Wien	Examinateur
Thierry MORA	DR CNRS, ENS	Examinateur
Amandine VÉBER	DR CNRS, Université Paris Cité	Examinatrice
Amaury LAMBERT	PU École Normale Supérieure (ENS)	Directeur
Emmanuel SCHERTZER	PU Universität Wien	Directeur

À Blaise, merci pour Joyce.

Acknowledgments/Remerciements

I am imensely grateful to the two referees who agreed to review this document.

Merci à mes deux encadrants Amaury Lambert et Emmanuel Schertzer pour leur patience sans failles, leur soutien à toute épreuve et leurs efforts pour stabiliser mes excursions hasardeuses.

I am immensely grateful to all the administrative staff of ENS, ED386 and the University of Vienna whose patience and help were priceless. I would like to specifically name Mme Fouzia Bouzid, Mr. David Zuser and Mr. Andreas Cap.

Merci à mon financement CDSN pour la liberté académique qu'il m'a accordée, et notamment de m'avoir autorisé à parler de ”biais mutationnel” et ”diversité génétique”.

Merci à Louis-Pierre Chaintron, Reinhard Bürger, Nicolas Fournier et Himani Sachdeva dont les excellents conseils ont grandement servi mes projets. Merci à François Ged de m'avoir aidé à rédiger et Félix Foutel-Rodier pour des discussions importantes.

Merci à l'équipe SMILE officielle et officieuse, au sixième et neuvième étage de l'Université de Vienne. Merci particulièrement à mes co-thésard·es Élisa Couvert, Mathilde André, Yannic Wenzel, Augustin Chen, Abdelmajid Omarjee, Thomas Forest et Frith Edbrooke.

Merci à tous les professeurs qui ont marqué mon parcours. Merci à mes professeurs MM. Faury, Alarcon, Rouillon, Boucekkine, Campomar, Dejean de la Bâtie, Lascroux. Je suis reconaissant envers Pierre-Henri Gouyon de m'avoir introduit à la génétique et à Mme Gazeau de ses cours de biologie.

Merci à ma famille, et toustes les ami·es qui m'ont soutenu de Vienne à Lausanne, Grenoble, Paris, Toulouse, Moulis, Avignon, Marseille et ailleurs.

Merci à Wikipédia, la plus belle réussite scientifique et sociale des dernières décennies.

Contents

Chapter 1

Introduction

The polygenic limit describes the limit of a model for a biological population of size N in which each organism has a number L of genes, as N and L go to infinity. This double scaling ($N \gg 1, L \gg 1$) is at the heart of the rich field of quantitative genetics [1, 2], but has so far eluded much of the mathematical community (with some notable exceptions to be mentioned). One possible reason for this is a rather murky vision of what scalings are needed, in what order, to get from an individual-based model to the polygenic limit. This work aims at

- offering context to the mathematically-minded audience on the motivation and importance of the polygenic limit

- providing a roadmap to the polygenic limit from an individual-based model, clarifying the type of approximations used (see Figure 1.3)

- clarifying the biologically reasonable scaling relationships of the different observables

- deriving formal and less formal proofs of some of the approximations used.

- characterizing the equilibrium under stabilizing selection in terms of fixed point equations

- promote a vision of the polygenic limit based on the "typical locus", as will be defined in Chapter 4.

My focus will mostly be on a population at (or close to) equilibrium under stabilizing selection, which as I will explain can serve as a benchmark of a "typical" wild population at stationarity.

This introduction is structured as follows. In Section 1.1, we will start by presenting phenotypic models of evolution, used to describe the evolution of continuous traits. In Section 1.2, we will discuss classical population genetics models, which describe the evolution of discrete "heredity particles", that is, genes. In Section 1.3, we will discuss how polygenic models are obtained by letting the number of genes of each organism go to infinity, and how this limit can reconcile population genetics models with phenotypic models. In Section 1.4, we will summarize the duality of the "trait-based" and "gene-based" approaches with a roadmap to the polygenic limit.

1.1 Evolution from the trait's eye-view

> As man can produce a great result with his domestic animals and plants by adding up in any given direction individual differences, so could natural selection, but far more easily from having incomparably longer time for action. [...] As man can produce, and certainly has produced, a great result by his methodical and unconscious means of selection, what may not natural selection effect ? [...] Man selects only for his own good, Nature only for that of the being which she tends.
>
> Charles Darwin, *The Origin of Species*,
> 150th anniversary edition.

These words from Chapter IV of *The Origin of Species* establish natural selection as the wild counterpart to human breeding. The extraordinary complexity of the living world is paralleled with the incredible achievements of breeders described the first chapter of *The Origin of Species*. Given the capabilities of the latter on the scale of decades, what incredible power must have natural selection, acting on the scale of millions of years ?

In some cases, breeding is about finding some strange new heritable form and selecting it until the whole population presents this new trait. The most famous example is the Ancon sheep, which due to a mutation had very short legs. This trait was selected because such sheep couldn't escape their owners [3]. Similarly, in the well-known case of the peppered moth (*Biston betularia*), natural selection has selected a black (melanic) form of the moth over the typical white form in regions where the trees had gone dark due to soot pollution [4]. Such a situation where several clearly distinguishable types exist in the population is referred to as a polymorphism. In such a case, breeding is about picking the best type, with the goal of obtaining a population in which all organisms have the desired type.

But most of the work of breeders is concerned with quantitative traits, that is, measurable characteristics of an organism which take continuous values. In this case, there are no clearly-defined types, rather a continuum of trait values. Darwin cites the short-beaked pigeon (selected to have a short beak), the pouter (selected to have a large crop), the fantail (selected to have more tail feathers). In the case of the fantail, one could argue that the number of tail feathers is not a quantitative trait, and should rather be seen as a polymorphism (the number of tail feathers is discrete and not continuous), we will return to this when presenting the infinitesimal model.

The fundamental law of breeding can be summarized as:

If a quantitative trait is variable within a population, and if part of that variation is heritable, then selective breeding can increase or decrease the mean value of the trait within the population well beyond its current range.

Here, "the variation is heritable" means the offspring of organisms with high trait values have, on average, a higher trait value than the offspring of organisms with low trait values.

The fundamental law of breeding was verified in countless experimental evolution experiments. The longest running experiment to increase a quantitative trait is the experiment from the University of Illinois to alter seed properties of maize, which has run since 1896. Selecting for higher oil concentration led to a sustained increase until 2008 when this experiment was discontinued, at which point seeds had an average oil concentration of 20%

(compared to 5% at the beginning of the experiment)[1]. Another long-term experiment was the domestication of the silver fox in the Institute of Cytology and Genetics of Russia, which since 1952 selected silver foxes to be human-friendly. By the thirtieth generation, over 70% of the population was “eager to establish human contact” [5]. Even more abstract traits such as maternal effects (that is, the capacity for a mother to influence the traits of her offspring) [6], plasticity (the ability of an organism to respond to the environment) [7] or bet-hedging (the ability of an organism to switch between different phenotypes) [8], all respond to selection.

Confirming Darwin’s parallel between breeding and evolution required obtaining empirical proofs that natural selection acts on quantitative traits. This was the goal of the biometrician school, championed by Weldon in [9]. Proving that natural selection acts to increase or decrease a given trait in a natural population requires the following steps:

1. Prove that there is covariance between the trait and fitness.
2. Prove that the trait is heritable.

The first point was famously illustrated with experimental measurement of natural selection by Bumpus [10]. During the particularly severe winter of 1898, he measured various traits of sparrows immobilized by the cold, some of which survived while the rest perished. He concluded that natural selection favored smaller, lighter birds, with longer feathers and larger brains for survival. The second point requires a model for the inheritance of traits: for quantitative traits, the infinitesimal model.

1.1.1 The infinitesimal model (diploids)

Clarifying the long-term consequences of natural selection required a better understanding of the heredity of complex traits. A crucial step was to find a suitable model to predict the distribution of trait values among the offspring of two organisms in a sexual species. This was achieved with the infinitesimal model [11], which was gradually developed from Galton’s revolutionary study on human height [12], the biometrician take on the law of ancestral heredity [13], and later on Bulmer’s work on the joint distribution of phenotypes among relatives [14].

The goal is to model a population of organisms, each of which is described by a phenotypic trait value in $\mathbb{R}$. The infinitesimal model is defined with two parameters

- The segregation variance $V_S \in (0, +\infty)$.
- The environmental variance $V_E \geq 0$.

The segregation variance quantifies the genetic diversity within the population, while the environmental variance quantifies the non-heritable sources of phenotypic diversity. We will focus on $V_E = 0$ for simplicity.

Describing the infinitesimal model requires the definition of the **pedigree** and **inbreeding coefficient**. The pedigree encodes the genealogical relationship between the organisms of the population.

Definition 1.1.1. *A* ***pedigree*** *is a finite directed acyclic graph* $(\mathscr{P}, E)$*, such that every node has 0, 1 or 2 parents (See Figure 1.1).*

The nodes of $\mathscr{P}$ *correspond to the past and present organisms of the population. If there is a path of arrows going from node* A *to node* B*, we say* A *is a descendent of* B*. Following [11], we will consider that there is a foundational generation of unrelated individuals, which consists of all nodes with no parents. An organism has only one parent if it results from selfing.*

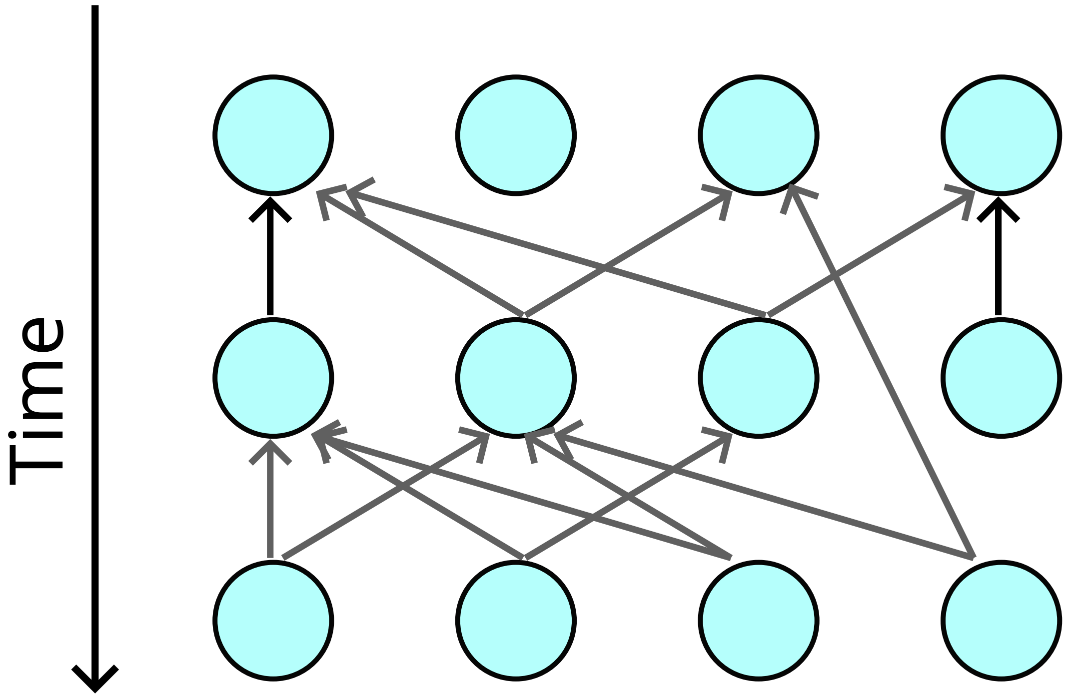

Figure 1.1: Example of a pedigree. Each node has two parents (grey arrows), or one parent if it is the result of selfing (black arrow), or zero parent if it is in the founding generation.

The inbreeding coefficient quantifies the percentage of genetic material shared between two organisms. It is typically computed under the assumption that offspring inherit a proportion 1/2 of its genetic material from each parent.

Definition 1.1.2 (See (1) and (2) in [11])**.** *Conditional on the pedigree* $\mathscr{P}$*, the inbreeding coefficient is a symmetric function* $\mathcal{F} : \mathscr{P}^2 \to [0,1]$ *which we will define recursively as follows*

- *If* $A, B \in \mathscr{P}$ *have no parents then* $\mathcal{F}(A,B) = 0$ *(even if* $A = B$*).*
- *If* A *has two parents* A_1, A_2 *(with* $A_1 = A_2$ *in the case of selfing) then for any* $B \neq A$ *such that* B *is not a descendent of* A *we define*

$$\mathcal{F}(A,B) = \mathcal{F}(B,A) = \frac{1}{2}\left(\mathcal{F}^*(A_1,B) + \mathcal{F}^*(A_2,B)\right).$$

where for $i \in [2]$

$$\mathcal{F}^*(A_i,B) := \begin{cases} \mathcal{F}(A_i,B) & \text{if } A_i \neq B \\ \frac{1+\mathcal{F}(A_i,A_i)}{2} & \text{otherwise} \end{cases}$$

Also

$$\mathcal{F}(A,A) = \mathcal{F}^*(A_1,A_2)$$

In this definition, the coefficient $\mathcal{F}^*(A_i,B)$ appears because if $B = A_i$, that is, B is a parent of A, then A automatically shares half its genetic material with B (in diploids).

The infinitesimal model can be defined in a condensed way as follows.

Definition 1.1.3 (The infinitesimal model)**.**

1. *Organism number* k *of the population has a measurable trait* $z_k \in \mathbb{R}$*.*

---

[1] see the team's website http://mooselab.cropsci.illinois.edu/longterm.html(consulted on the 08/08/2025)

2. *Organism number k of the population has a hidden* ***genetic trait value*** $z_k^G \in \mathbb{R}$*. This value corresponds to the thought experiment: if we were to breed many genetic clones of this organism, letting the clones live in the same conditions as their progenitor, what would the average of the trait values of the clones be?*

3. *The value of* $e_k := z_k - z_k^G$ *is called the* ***environmental component*** *of* k*.*

4. *Suppose we know the pedigree* $\mathscr{P}$ *of two organisms labelled* k_1 *and* k_2*, as well as their genetic trait values. If they have an offspring, then the environmental component of this offspring is sampled randomly from a normal distribution* $\mathcal{N}(0, V_E)$*. The genetic trait value of the offspring is independently sampled from a normal distribution*

$$\mathcal{N}\left(\frac{z_{k_1}^G + z_{k_2}^G}{2}, V_S\left(1 - \frac{\mathcal{F}(k_1,k_1) + \mathcal{F}(k_2,k_2)}{2}\right)\right).$$

Remark 1. *This model is named "infinitesimal" in that there are no atoms of inheritance: no matter how small the inbreeding coefficient* $\mathcal{F}(k_1, k_2)$*, it is always assumed that the organisms labelled* k_1 *and* k_2 *share exactly a fraction* $\mathcal{F}(k_1, k_2)$ *of their genetic material.*

For the purpose of this PhD, it will not be needed to get into additional subtleties due to sex (in reality, the trait of the offspring typically depends on its sex [15]). We will also ignore environmental effects, setting $V_E = 0$. Furthermore, rather than working with the inbreeding coefficient $\mathcal{F}$, we will work with the segregation variance $c(k_1, k_2) := V_S(1 - \mathcal{F}(k_1, k_2))$, which is more suited to account for mutations (see Section 1.1.2).

The parameter V_S is the keystone of the infinitesimal model. Let us ignore for now inbreeding (setting $\mathcal{F} = 0$) to discuss its rôle. If $V_S = 0$, the genetic value of the offspring is exactly the mean of that of its parents. After a while, we expect all organisms to have the same genetic trait value and the only source of difference between organisms to be the random uncorrelated environmental components. Somewhat paradoxically, perfect inheritance leads to a population with no apparent inheritance.

If instead $V_S > 0$, inheritance is imperfect. Because the normal distribution is not bounded, there is a non-zero (albeit small) probability that two organisms with small trait values have an offspring with a large trait value, or vice-versa. Now, consider a breeder trying to increase the trait value of the population. If $V_S = 0$, and if the parents have genetic trait values $z_{k_1}^G$ and $z_{k_2}^G$, then the genetic trait value of the offspring is $\frac{z_1^G + z_2^G}{2} \leq \max\{z_1^G, z_2^G\}$. In particular, the breeder cannot increase the genetic value of the trait beyond the maximum of the starting population. Conversely, if $V_S > 0$, then if the pair with genetic trait values (z_1^G, z_2^G) has many offspring, it is likely at least one of them has trait value larger than $\max\{z_1^G, z_2^G\}$. In fact, the larger V_S, the larger this probability. So the breeder can increase the maximum genetic trait value of the population. In this sense, V_S can be seen to encode the **ability of the population to respond to directional selection**.

Now, consider the effect of inbreeding through the coefficient $\mathcal{F}$. Notice how, in the recursion formula from Definition 1.1.2, we always have $\mathcal{F}(A, B) \geq \min\{\mathcal{F}(A_1, B), \mathcal{F}(A_2, B)\}$ for two nodes A, B such that A_1, A_2 are the parents of A. In particular, the minimal inbreeding coefficient within the population $\min\{\mathcal{F}(A, B)\}$, where the minimum is over all pairs (A, B) of organisms alive at time t, can only increase with t. The larger $\mathcal{F}(A, A) + \mathcal{F}(B, B)$, the lower the variance of the trait value from the offspring of A and B, and the lower the probability that A and B have an offspring with larger trait value than either of its parents. In this sense, inbreeding hampers the response to selection.

In fact, it can be seen that uniform inbreeding is equivalent to a decrease in V_S as follows. For a given pedigree $\mathscr{P}$, define an alternative parameterization $(\hat{\mathcal{F}}, \hat{V}_S)$ as follows

- If A, B have no parents, then $\hat{\mathcal{F}}(A, B) = \hat{F}_0$ for some fixed value $\hat{F}_0 \in (0, 1)$. Define $\hat{\mathcal{F}}(A, B)$ for all other pairs (A, B) by the same recursion as in Definition 1.1.2.

- $V_S = \hat{V}_S(1 - \hat{F}_0)$.

Then it can be checked that for any A, B, we have

$$V_S\left(1 - \frac{\mathcal{F}(A,A) + \mathcal{F}(B,B)}{2}\right) = \hat{V}_S\left(1 - \frac{\hat{\mathcal{F}}(A,A) + \hat{\mathcal{F}}(B,B)}{2}\right).$$

In other words, it is strictly equivalent to use the pair $(\mathcal{F}, V_S)$ or the pair $(\hat{\mathcal{F}}, \hat{V}_S)$ in the infinitesimal model. Uniform inbreeding is equivalent to a reduction of the segregation variance.

To use the infinitesimal model to describe the evolution of a real population, we must specify which traits satisfy the infinitesimal model and how they relate to fitness. If we wish to predict the long-term behavior of the population, we furthermore need to account for the replenishment of the segregation variance V_S by mutations, countering the effect of inbreeding.

Which traits satisfy the infinitesimal model ?

The reader may be skeptical that "traits" can be treated as a general object of study given the vague definition we used so far, for which any measurable quantity qualifies. For instance, as noted in [11], if a given trait satisfies the infinitesimal model, then the square of that trait cannot satisfy the infinitesimal model. An ecologist studying a population needs a model to classify traits, and in particular to determine which traits will satisfy the infinitesimal model.

In the nineties it became accepted that suitably transformed **morphological traits** (such as the relative lengths of different body parts) generally obey the infinitesimal model [16] (or to frame it in quantitative genetics terms, that the genetic variance for such traits is mostly additive - see Section 1.3.4), whereas **life history traits** such as number of offspring and longevity do not (in quantitative genetics terminology, their genetic variance contains a substantial amount of dominance or epistasis variance). See also Chapter 7 of [17].

The model that emerged is described in Figure 1.2. The fitness of an organism is determined by life-history traits, which are themselves determined by morphological traits, which when suitably scaled obey the infinitesimal model. Even the inheritance of traits such as "log-bristle number in an abdominal segment of *Drosophila melanogaster*", which are discrete, is well approximated by the infinitesimal model [18] (that is, its genetic variance is mostly additive). However, if sustained selection is applied to increase or decrease a morphological trait (as for domestic species), then eventually the infinitesimal model will break down (in quantitative genetics terminology, substantial dominance and epistasis variance will appear) [16].

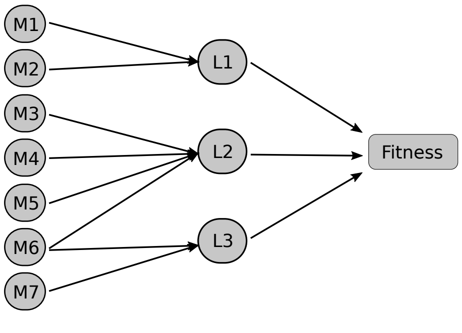

Figure 1.2: Fitness is determined by life history trait L_1, L_2, L_3. Each of these traits is determined by morphological traits $M_1, \dots, M_7$. Each morphological trait obeys the infinitesimal model, but the life history traits do not. See [16].

More recently, the advent of mass transcriptomics has led to large datasets recording the level of gene expression in a given organism, in a specific tissue at a specific developmental time. The interpretation of these datasets typically treats the log-level of gene expression as a quantitative trait satisfying the infinitesimal model [19].

It should be mentioned that an important class of traits, particularly in humans, does **not** satisfy the infinitesimal model: they are the so-called behavioral traits. In humans, these include IQ or educational attainment[2].

To model the evolution of the population, we must further specify how the traits relate to fitness, in order to incorporate the effect of natural selection.

On the importance of stabilizing selection.

Selective breeding can lead to rapid evolution on the scale of dozens of generations. Yet modern wild species closely resemble their ancestors living thousands of years ago. Already at the time of Darwin palaeontologists had observed a puzzling phenomenon: measurements from fossils indicated that the traits of a species tend to remain remarkably constant over time, except during abrupt transitions. Darwin believed evolution to proceed very slowly, even on geological timescales, and blamed these abrupt transitions on the imperfection of the geological record, as well as immigration: if an invasive species replaces its local relative, the local fossil record will show an abrupt transition. He discusses this in Chapter IX of *The Origin of Species*. The history of the debate on rates of evolution is admirably summed up in the introduction of [20], from Lamarck's *peu à peu* to the theory of punctuated equilibria as formulated by Eldredge and Gould.

The modern opinion on this debate as summarized in [20] is

1. Traits evolve very rapidly over short timescales. Field studies typically show a change in the mean trait value over a generation of order 0.1σ, where σ^2 is the population variance of the trait.
2. Traits remain approximately constant over long timescales, during very long periods of evolutionary stasis. When comparing trait values z_1 and z_2 of fossils separated by a time interval of Δt generations, the typical value of $|z_1 - z_2|/\Delta t$ is $10^{-4}\sigma$.

The simplest way to reconcile these two observations is the concept of stabilizing selection (though evolutionary stasis can be influenced by other causes like mutation rates [21]). Stabilizing selection corresponds to selection favouring an intermediate trait value. It maintains the mean value of the trait within the population close to some optimum. So if the abiotic and biotic environment of the population remain the same over geological time, the mean trait value of the population will remain close to the optimum, while a sudden change can lead to a rapid evolution until a new optimum is reached (there are now examples of such transitions which have been perfectly conserved in fossils, see Chapter 9 of [20]).

Another line of evidence in favour of pervasive stabilizing selection is from empirical studies. In [22], Weldon famously found a significant effect of natural selection in reducing the genetic variance in morphological traits for crabs. Generally speaking, one expects organisms with extreme trait values to have lower fitness. This correlative evidence could be biased if genes responsible for extreme trait values also tend to decrease fitness for other causes (for instance, if a gene shuts down a development pathway, it may lead to short height and also severe disabilities) [23]. Under this correlative approach, the joint distribution of fitness F and trait value z are measured in a population, and fitness is fitted as a quadratic polynomial of trait value

$$F = a + bz + cz^2 + \varepsilon$$

---

[2] An excellent review of the literature on this subject can be found on Sasha Gusev's substack https://theinfinitesimal.substack.com (consulted on the 08/08/2025)

where a, b, c are fitted constants and ε is the residual [24]. It should be noted that field studies often report apparent directional selection on a morphological traits (that is, positive covariance between trait value and fitness) instead of stabilizing selection, without observing a response to selection [25]. This phenomenon could be due to the time fluctuations of selection: one year it is better to be tall, the next it is better to be short, and therefore on average it is better to be of intermediate height [26]. In fact, in Figure 6. of [25], a review of field studies offers a rather puzzling observation: stabilizing selection does not seem to be more frequently detected than its converse, diversifying selection (which would favor extreme trait values), except for height. There are a number of reasons why this should be taken with caution, including biased trait reporting and inadequate statistical methods (see [25]). More recently, [15] has found stabilizing selection on a large number of human traits using the same statistical approach on very large datasets. The strongest evidence for stabilizing selection is from GWAS data and will be discussed in Section 1.3.

There is also evidence that gene expression (which can also be modelled with the infinitesimal model) is under stabilizing selection [27, 28]. Specifically, [27] compared gene expression patterns in different strains/species of Drosophila and found greater divergence in gene regulation than in gene expression. That is, for a given gene A, the species had diverged in the genetic mechanisms to control the expression of A, but kept similar levels of expression of A.

Mutations can generate genetic variability

So far, we have not discussed the cause of genetic variability. If we try to describe the evolution of a population using the infinitesimal model as in Definition 1.1.3, it can easily be seen that under panmixia, after a while the population will be completely inbred ($\mathcal{F}(k_1, k_2) = 1$ for all pairs (k_1, k_2) of organisms), and then the trait will cease to evolve.

And yet, fully inbred populations respond to selection after a sufficiently large number of generations. For instance, in Drosophila populations, [29] found the genetic variance in abdominal or sternopleural bristle numbers could be restored to its typical value in 400 generations. This points to the process of mutation, which generates genetic variability in the trait. Mutations generate genetic differences between organisms. They can be modelled as a force countering inbreeding, decreasing the value of the inbreeding coefficient $\mathcal{F}$ and replenishing the segregation variance. In the next section, we will discuss how this process can be incorporated into phenotypic models.

1.1.2 Phenotypic models for the evolution of a trait

Using the infinitesimal model, it is possible to model the evolution of a population from the trait's eye-view. In this section, we will introduce an individual-based model, give heuristics for the simplifications which are expected to hold when the population is large, and describe the large population limit following the work of Lande [30]. I warn the reader that the notation introduced in this section will not be used in the rest of this dissertation.

A population is described as a probability distribution p over the trait space (here taken to be $\mathbb{R}$ for simplicity). The trait of an organism uniformly sampled within the population is a random variable with law p. The typical biological forces shaping p in an isolated population are genetic drift (D), reproduction and selection (S), mutation (M).

Let $\mathcal{P}(\mathbb{R})$ (resp. $\mathcal{M}(\mathbb{R})$) be the set of probability distributions (resp. measures) on $\mathbb{R}$. For $p \in \mathcal{M}(\mathbb{R})$ and a measurable function $f : \mathbb{R} \to \mathbb{R}$, we write

$$p[f] := \int f(x) p(\mathrm{d}x)$$

The population at generation n is described by a pair (p_n, C_n) where

- $p_n \in \mathcal{P}(\mathbb{R})$ is the distribution of the trait within the population.

- C_n is the distribution of segregation variances. We will take it to be a probability kernel from $\mathbb{R}^2$ to $\mathbb{R}$, that is, for any $z^1, z^2 \in \mathbb{R}$, $C_n((z^1, z^2), \cdot)$ is the distribution of the inbreeding coefficient between pairs of organisms such that the first one has trait value z^1 and the second one has trait value z^2. If no organism has trait value z^i for $i \in \{1, 2\}$, then we set $C_n((z^1, z^2), \cdot) = 0$.

An individual-based model

The model will be specified with the following parameters

(D) the population size N

(S) the fitness function $F : \mathbb{R} \to (0, +\infty)$

(M) the mutation probability $\mu \in (0, 1)$, mean mutation effect $\check{z}$ and mutation variance $V_m > 0$.

Following [31], a natural and simple mathematical model to represent a finite population under reproduction, selection, mutation, and genetic drift is as follows. At time t, the population contains N organisms with (genetic) trait values $(z_n^i)_{i\in[N]}$ and segregation variances $(c_n(i,j))_{i,j\in[N]}$. The population is therefore

$$p_n^{(N)} = \frac{1}{N} \sum_{k\in[N]} \delta_{z_n^k}$$

where δ is the Dirac distribution. Every time step

1. *Selection+Genetic drift.* For each $k_0 \in [N]$, two parents k_1, k_2 are independently picked at random, such that the probability that organism number k is picked is proportional to $F(z_n^k)$.

2. *Reproduction.* Following the infinitesimal model, the trait $z_{n+1}^{k_0}$ of the offspring of $z_n^{k_1}, z_n^{k_2}$ is taken to have distribution $\mathcal{N}\left((1-\mu)\frac{z_n^{k_1}+z_n^{k_2}}{2} + \mu\check{z}, c_n(k_1, k_2)\right)$.

3. *Segregation variance.* The segregation variance between z_{n+1}^i and z_{n+1}^j is defined as follows. Write $i[1], i[2]$ for the two parents of organism i, and similarly define $j[1], j[2]$. Then we have for $i \neq j$

$$\begin{aligned} c_{n+1}(i,j) :=& 2\mu V_m \\ &+ \frac{(1-\mu)^2}{4}\left(c_n^*(i[1], j[1]) + c_n^*(i[2], j[1]) + c_n^*(i[1], j[2]) + c_n^*(i[2], j[2])\right) \end{aligned} \tag{1.1}$$

$$c_{n+1}(i,i) := 2\mu V_m + (1-\mu)^2 c_n^*(i[1], i[2])$$

where $c_n^*(i,j) := \frac{1}{2}(2 - \mathbb{1}_{[i=j]})c_n(i,j)$.

Remark 2. *This simplified model assumes a panmictic hermaphroditic isolated unstructured diploid*[3] *population with discrete generations, in a constant environment. Furthermore, the effect of mutations on the offspring k_0 is assumed to be independent of the trait values of its parents.*

---

[3] For haploids, simply replace c_n^* with c_n.

Remark 3. *The formula for the segregation variance accounting for the effect of mutations is adapted from Section 2.2.6 of [11] for diploids. This formula should be interpreted as follows: the segregation variance $c_{n+1}(i,j)$ quantifies the genetic divergence between organisms i and j. Mutations create divergence between the two by replacing a fraction μ of the genome with mutations with variance V_m. The rest of the divergence is made up of the divergence between the inherited parts of the genetic material. Because each organism inherits a proportion $(1-\mu)/2$ of its genome from each parent, the divergence between the inherited parts of the genetic material is $(1-\mu)^2$ times the mean divergence between the parents of i and those of j.*

Large population limit and concentration of the segregation variance.

Taking the limit $N \to +\infty$ in the individual-based model has to my knowledge not yet been rigorously done. It would be desirable to obtain a continuous limit process which preserves some randomness, because such systems are often much more mathematically tractable than the discrete individual-based model, and more realistic in that they describe the fluctuations of the system. What has been done is to describe the segregation of the pedigrees in similar biparental models with weak selection and weak population structure. We assume that the mutation probability μ is of order at most $1/N$.

By segregation of the pedigrees, we mean how two present-day organisms have many distinct ancestors at generation n such that $1 \ll n \ll \ln(N)/\ln(2)$, and their respective inbreeding coefficients are determined by average properties of the population this time ago. At equilibrium, this lets us argue that we may effectively consider the empirical distribution of the segregation variances

$$\frac{1}{N(N-1)} \sum_{i,j\in[N]} \delta_{c_n(i,j)}$$

as very concentrated around some value V_S^*, which remains constant through time.

As an illustrative example, consider a neutral discrete biparental Wright-Fisher model: each organism chooses two parents independently in the previous generation. Suppose that at generation 1, the segregation variances $(c_1(i,j))_{i,j}$ are i.i.d with distribution d_1 on $[0, V_S]$ and expectation V_S^*. Consider two organisms A, B sampled uniformly at random at generation n with $1 \ll n \ll \ln(N)/\ln(2)$. We argue that $c_n(A,B) \simeq c_n(A,A) \simeq V_S^*$. In particular, this means the distribution of segregation variances becomes concentrated around its expectation. The probability that A and B have 4 distinct parents is

$$f_1 = \left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\left(1-\frac{3}{N}\right).$$

The probability that they have 8 distinct grandparents is the previous probability, times the probability that the 4 parents have distinct parents

$$f_2 = \left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\left(1-\frac{3}{N}\right)\times\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\dots\left(1-\frac{7}{N}\right).$$

More generally, the probability that A and B have 2^{n+1} distinct ancestors n generations in the past is

$$f_n = \prod_{t=1}^{n}\left(\prod_{k=1}^{2^{t+1}-1}\left(1-\frac{k}{N}\right)\right).$$

A standard Taylor expansion yields that for $N \gg 1$, $n \gg 1$

$$f_n = e^{C\frac{2^n}{N}+o\left(\frac{2^n}{N}\right)}$$

for some constant $C > 0$. In particular, whenever $1 \ll n \ll \ln(N)/\ln(2)$, with high probability A and B have $2^n \gg 1$ distinct ancestors n generations in the past, and inherited the same proportion $1/2^n$ of genetic material from each of them, and A and B share no ancestry in the past n generations. On this event, if we neglect mutations, the segregation variance $c_n(A, A)$ is equal to the average segregation variance between one uniformly-picked paternal ancestor and one uniformly-picked maternal ancestor of A, and the segregation variance $c_n(A, B)$ is equal to the average segregation variance between one uniformly-picked ancestor of the 2^n ancestors of A and one uniformly-picked ancestor of the 2^n ancestors of B. Since we considered that at time 1, the $(c_1(i, j))_{i,j}$ are i.i.d with mean V_S^*, the law of large numbers implies that the segregation variances $c_n(i, i)$ and $c_n(i, j)$ will both be very close to the expectation of a random variable with distribution d_1. We thus obtain that the distribution of the segregation variances at generation n $(c_n(i, j))_{i,j}$ is very concentrated around V_S^*. This justifies assuming that a single value V_S^* can effectively describe the segregation variance between two randomly sampled organisms at any given time, which remains constant over timescales shorter than $\ln(N)/\ln(2)$. A rigorous justification that this property propagates through time would require showing the same result for a more general structure of $(c_0(i, j))_{i,j}$ than i.i.d.

The behavior of the biparental pedigree of a neutral population, at generation $n \sim \ln(N)/\ln(2)$, was famously studied in [32, 33]. In these articles, it is found that all present-day organisms have the same ancestors at generation $n \sim \ln(N)/\ln(2)$, with similar multiplicity. In particular, the contribution of an ancestor n generations in the past to the present-day population is entirely summed up by a single value called the reproductive value (see [34]). It was later shown in [35] that this result is robust to moderate population structure. Finally, [34] showed that this is also robust to directional selection acting on the trait. On this timescale, V_S^* is still expected to remain constant. This is because, on one hand, mutations are still rare. On the other hand, we argue that inbreeding still has no effect. Indeed, though two randomly-sampled organisms at generation n share close to 100% of their ancestry at generation 0, each ancestor only contributes a little bit. To take an extreme example, consider two organisms k_1, k_2 from generation n which have the same 2^n unrelated ancestors at generation 0, each ancestor contributing only once to k_1 and k_2. Then for any ancestor, k_1 and k_2 inherited a fraction 2^{-n} of genetic material from that ancestor. In particular, a fraction 2^{-2n} of genetic material of that ancestor is shared by both k_1 and k_2. So the inbreeding coefficient of $\mathcal{F}(k_1, k_2)$ is the number of ancestors (2^n) times the fraction of genetic material shared between k_1, k_2 and that ancestor (2^{-2n}) and is therefore of order $2^{-n} \ll 1$.

On a larger timescale of n, the effect of inbreeding and mutations on V_S^* kick in, which can also be investigated in the neutral model, assuming all of the segregation variances at generation 0 are equal to V_S^*. If we sample two distinct organisms A, B at generation 1 at random,

- if they share 0 parents, then $c_1(A, B) = V_S^*$
- they share 1 parent and are not the result of selfing with probability

  $$\frac{4}{N}\left(1 - \frac{1}{N}\right)^2 \simeq \frac{4}{N}$$

  in which case (1.1) yields

  $$c_1(A, B) = 2\mu V_m + \frac{(1-\mu)^2}{4} V_S^* \left(3 + \frac{1}{2}\right).$$

- all other events have probability of order $1/N^2$.

We therefore find that that on a leading order of $1/N$ and μ,

$$\mathbb{E}\left[c_1(A,B)\right] \simeq 2\mu V_m + (1-2\mu)\left(1 - \frac{4}{N} + \frac{4}{N} \times \frac{7}{8}\right) V_S^*$$

which yields

$$\mathbb{E}\left[c_1(A,B)\right] \simeq V_S^* + 2\mu(V_m - V_S^*) - \frac{1}{2N} V_S^*.$$

Thus, assuming μ has order $1/(2N)$, we find that V_S^* at equilibrium satisfies

$$V_S^* = \frac{4\mu N}{1+4\mu N} V_m.$$

This sort of reasoning was used in [36] (equation (10)). It could be adapted to account for selection, but the traditional approaches to finding the equilibrium value of the genetic variance within a population with selection is to use polygenic models (see Section 1.3.9).

Assuming the segregation variances are very concentrated around V_S^* lets us model the evolution of the large population p_t under reproduction with the infinitesimal model as $p_{t+1} = R(p_t)$ with the following operator

$$R : \begin{cases} \mathcal{P}(\mathbb{R}) & \longrightarrow \mathcal{M}(\mathbb{R}) \\ p & \longmapsto \int_{\mathbb{R}\times\mathbb{R}} \varphi\left(\cdot - \frac{z_1+z_2}{2}, V_S^*\right) p(\mathrm{d}z_1) p(\mathrm{d}z_2) Leb \end{cases}$$

where Leb is the Lebesgue measure and

$$\varphi(z, V_S^*) := \frac{1}{\sqrt{2\pi V_S^*}} e^{-\frac{z^2}{2V_S^*}}.$$

One remarkable property of R is that normal distributions are stable under R. Formally,

$$\forall (\bar{z}, V) \in \mathbb{R}\times\mathbb{R}_+, \qquad R(\mathcal{N}(\bar{z}, V)) = \mathcal{N}\left(\bar{z}, \frac{V}{2} + V_S^*\right).$$

For this reason, if the system is dominated by recombination (meaning we can write $p_{t+1} \simeq R(p_t)$) we expect the trait distribution to be close to a normal distribution.

In the next section, we show how this simplification lets us efficiently describe a population under stabilizing selection.

Large population under stabilizing selection

In 1976, Lande [30] suggested the first autonomous model of a trait evolving under stabilizing selection and genetic drift of a large population ($N \gg 1$). The effect of selection is modelled using the quadratic logfitness function

$$W(z) = -\frac{(z-\eta)^2}{2\omega^2} \qquad\qquad F = e^W \tag{1.2}$$

with ω^{-2} the strength of selection, and η the selection optimum. This type of quadratic selection with a single optimum is known as Fisher's Geometric Model [37].

Lande suggests the description of the population at time t with normal distribution $\mathcal{N}(\bar{z}_t, \sigma^2)$ for some fixed parameter $\sigma^2 \ll \omega^2$ and a mean $\bar{z}_t$, which evolves following the Stochastic Differential Equation (SDE)

$$\mathrm{d}\bar{z}_t = \frac{\sigma^2}{\omega^2}(\eta - \bar{z}_t)\mathrm{d}t + \sqrt{\frac{\sigma^2}{N}}\mathrm{d}B_t \tag{1.3}$$

This is an Ornstein-Uhlenbeck process.

For this model to be theoretically justified, it should be obtained as a scaling limit from an individual-based model. Specifically, rigorously obtaining this model would require showing the following decomposition

- On a very short timescale, the pedigrees segregate so that the typical segregation variance between two organisms is V_S^*.
- On a short timescale, reproduction mixes the phenotypic distribution p_t, maintaining it close to a normal distribution with constant variance $\sigma^2 = 2V_S^*$ and mean $\bar{z}_t$.
- On a larger timescale, stabilizing selection and genetic drift jointly act on $\bar{z}_t$ as in (1.3).

In [38], a very precise description of the evolution of a population under stabilizing selection is achieved, under the assumption that the population trait distribution p_t remains Gaussian with constant variance, but allowing the population size to fluctuate. Specifically, in this article the average number of offspring of a given organism is a function of the trait z_G and the population size N_t to prevent endless growth of the population. A coupled set of SDEs for the trait mean $\bar{z}_t$ and the population size N_t are obtained, describing the establishment of migrants in a new population and the stationary population, for which the joint distribution of $(\bar{z}_t, N_t)$ is obtained.

Let us now mention some of the possible applications of this model.

1.1.3 Applications of phenotypic models

The infinitesimal model (suitably extended) is at the heart of quantitative genetics, which has had immense success for breeding, laboratory studies and field studies in the wild [1, 2], and can be considered as one of the pillars of modern evolutionary biology. I will only briefly mention examples of applications of the stabilizing selection model.

Application to model the divergence of populations

Lande's Ornstein-Uhlenbeck model (1.3) has been used to compare closely-related species' morphological traits, under the assumption that since the time at which two species split, their mean morphological traits values have evolved as independent Ornstein-Uhlenbeck processes [39, 40] (though see [41]). Since the time of the split is known from molecular clocks, this lets researchers estimate with a Bayesian approach the timescales at which traits fluctuate, and the strength of stabilizing selection.

Extension to multiple traits.

In a landmark article [24], Lande and Arnold extended the infinitesimal model to a system with multiple traits as described in Figure 1.2, and proposed a method to infer stabilizing selection from measurements in wild populations. Now the traits of an organism are given by a vector $z \in \mathbb{R}^d$, and the covariances between the different components of z are given by an element of $\mathscr{M}_d(\mathbb{R})$, the set of positive-definite matrices. The idea, from Lande's 1979 article [42], can be formulated as follows

1. The population traits at time t are distributed following a normal distribution $\mathcal{N}(\bar{z}_t, V_P)$ where $V_P \in \mathscr{M}_d(\mathbb{R})$ is the phenotypic covariance matrix.
2. The population genetic trait values are distributed following a normal distribution $\mathcal{N}(\bar{z}_t, V_G)$ where $V_G \in \mathscr{M}_d(\mathbb{R})$ is the genetic covariance matrix.
3. The effect of stabilizing selection is obtained as in (1.2) through the fitness function

$$F(z) = \exp\left[-\frac{1}{2}(z-\eta)^\top \omega^{-2}(z-\eta)\right]$$

where $\eta \in \mathbb{R}^d$ is the selection optimum and $\omega^{-2} \in \mathscr{M}_d(\mathbb{R})$ is a positive-definite matrix specifying the strength of stabilizing selection. The model also allows for other forms of selection, most notably directional selection (replacing $\omega^{-2}(z-\eta)$ with some vector $\beta \in \mathbb{R}^d$) and disruptive selection (replacing ω^{-2} with $-\omega^{-2}$).

From a modelling perspective, this model can be obtained by updating the infinitesimal model from Definition 1.1.3, replacing the environmental and segregation variances (respectively V_E and V_S) with covariance matrices in $\mathscr{M}_d(\mathbb{R})$, and assuming that the population is well-mixed.

This model has had enormous success, and Lande and Arnold's paper [24] became one of the most cited articles of *Evolution* [43], the basis of field studies on eco-evolutionary dynamics of quantitative traits [17]. In a sense, it completes the picture from Figure 1.2, allowing field experimentalists to disentangle the way fitness is determined by morphological traits. It also saw use on gene expression data from transcriptomics [19].

Beyond normally-distributed traits

We have only mentioned situations where it is assumed that the trait distribution p_t remains normal within the population. The infinitesimal model assumes that the trait value of the offspring of two organisms is normally distributed around the mean of its parents, and can still accurately describe situations such as disruptive selection where the population p_t is no longer a normal distribution [44]. For instance, [45] emphasizes that the most computationally efficient way to keep track of the evolution of the population is the Fourier transform, which transforms convolutions due to the R operator to products. There has been remarkable deterministic work from the PDE community to prove the robustness of the normal distribution of the trait within a population to weak selection [46, 47], competition [48], or immigration in a spatially structured population with heterogeneous environments [49]. These works in deterministic settings all assume that the pedigrees segregate as discussed in Section 1.1.2, meaning the effect of reproduction can be modelled with the operator R.

Applying the infinitesimal model directly to phenotypic data in selection experiments or long-term measurements in wild population is possible using linear mixed models. This approach, known as the "animal model", can account for many complications due to phenotypic plasticity, shared environment, maternal effects, environmental fluctuations (see for instance Chapter 19 of [2]).

1.2 Evolution from the gene's eye-view

As mentioned at the beginning of this Introduction, some traits present a mode of inheritance radically different from the infinitesimal model. Under the infinitesimal model, extreme trait values occur rarely. When such a rare event occurs, which leads to an organism with an extreme trait value, the offspring of that organism tend to have less extreme trait values (because its other parent does not have an extreme trait value), and after a few generations the system loses all memory of this special event[4].

In biological systems however, some extreme events can have long-lasting consequences. In the case of the Ancon sheep, one should imagine a sheep born with a trait very different from that of its parents or any other sheep within the population: an extreme form of dwarfism. And among the offspring of that sheep, some would develop the same kind of dwarfism as their parent, while others would be indistinguishable from the rest of the population. In such a setting, the most appropriate description of the population seems to be a binning into dwarf and non-dwarf discrete categories, rather than a continuous distribution.

---

[4] This is reminiscent, albeit distinct, of Galton's regression to the mean [12]. Galton dubbed regression to the mean a phenomenon due to the environmental component (see Definition 1.1.3), summarized as follows. Consider the tallest midparent in the population, that is, the breeding pair with the greatest mean height. The elements of this pair tend to have high genetic value z_k^G **and** high environmental component e_k. The offspring will most likely have a genetic trait value close to that of the midparent and a non-extreme environmental component. Therefore, it will be shorter than its midparent.

Historically, such discrete events, incompatible with the infinitesimal model, were the main arguments used against Darwin's theory of natural selection (though the infinitesimal model had not yet been formalized as such). On one hand, there was still a lot of confusion on the inheritance patterns of quantitative traits, which led to incorrect claims that natural selection could not effectively act on a quantitative trait. On the other hand, Huxley, Galton, de Vries and later Bateson emphasized the importance of discrete variants (which Galton called "sports" as opposed to "variation proper"). See Chapters 1-2 of [50] for details.

At the turn of the twentieth century, the rediscovery of Mendel's works by Hugo de Vries, Carl Correns, and Erich von Tschermak led Hugo de Vries to formulate the concept of pangene [51], which would eventually become genes. Genes are defined as unalterable microscopic particles of inheritance, which are transmitted following Mendel's laws and determine the genetic trait values of organisms. The work of Morgan in the 1910s clarified the chromosomic structure of the genome and the concept of locus. This ultimately led to the development of population genetics by Wright, Fisher, and Haldane, which would lead to the modern synthesis of evolution (see Chapter 5 of [50]). This work reached further maturity with the use of the mathematical theory of diffusions, pioneered by Feller [52] and Kimura [53], building up on the work of Wright and Fisher.

In this section, we will present Mendelian inheritance in Section 1.2.1 and suggest a simple model for the evolution of the population in Section 1.2.2, with the corresponding large-population limit. In Section 1.2.3 we will discuss how the model can be adapted to account for diploidy.

1.2.1 The particles of heredity (haploids)

The microscopic model representing the particles of heredity in haploids can be described as follows

- Each organism has a **genome**, which has L positions called **loci** labelled 1 to L.
- At each position, the organism has a gene. This gene has a finite number of possible types, called **alleles**.

We will only be concerned with **biallelic loci**, that is, there are two alleles at each locus, labelled $+1$ and -1. As will be discussed in Section 1.3.10, this is a standard assumption in modern efforts to model polygenic systems.

The genome of the organism can therefore be represented as an element $g \equiv (g_\ell)_{\ell\in[L]} \in \{-1,+1\}^L =: \square_{[L]}$. We call a given element of $\square_{[L]}$ a **genotype**, denoted γ, whereas a $\square_{[L]}$-valued random variable is a **genome**, denoted g. Similarly, a $\{-1,+1\}-$valued random variable is a **gene** whereas -1 and $+1$ are **alleles**.

The basic model of heredity for haploids (counterpart to the infinitesimal model from Definition 1.1.3) can be described as follows

Definition 1.2.1 (Mendelian inheritance in haploids). *If two genomes $g^1, g^2 \in \square_{[L]}$ have an offspring with genome g^*, then the offspring samples randomly a certain subset $\mathcal{I} \subseteq [L]$ with some probability distribution ν. Then for every locus $\ell \in [L]$, if $\ell \in \mathcal{I}$ then $g^*_\ell = g^1_\ell$, otherwise $g^*_\ell = g^2_\ell$.*

This model is parameterized by the recombination measure ν, which specifies the pattern of inheritance (examples are given below). It is typically assumed that ν is non-degenerate, meaning for any $\ell_1, \ell_2 \in [L]$ with $\ell_1 \neq \ell_2$, we can find a subset $A \subset [L]$ with $\ell_1 \in A, \ell_2 \notin A$ and $\nu(A) > 0$.

Note the contrast with the infinitesimal model from Definition 1.1.3: here, g^*_ℓ is either an exact copy of g^1_ℓ or an exact copy of g^2_ℓ, whereas under the infinitesimal model, the genetic trait value of the offspring z^G_* is distributed around the midparent $\frac{z^G_1+z^G_2}{2}$.

The recombination measure

I mention a few of the classical recombination models to provide intuition on what this object means.

- The model known to population geneticists as **free recombination** corresponds to the case where $\nu(\mathcal{I}) = \frac{1}{2^L}$ for any $\mathcal{I} \subseteq [L]$. This recombination measure is typically suited if there is one locus per chromosome.

- **Single crossover.** Let μ be a probability measure on $[0,1]$ with positive continuous density. Then ν is the law of the random set

  $$\mathcal{J} = \left\{ i \in [L], \frac{i}{L} \leq X \right\}, \quad \text{where } \mathcal{L}(X) = \mu.$$

  This recombination measure seems suited to describe one chromosome of some species such as butterfly [54].

- **Multiple crossovers.** Consider a Point Process with an intensity measure with a strictly positive continuous density on $[0,1]$, seen as a random set of points $\lambda_1 < \cdots < \lambda_N$. We add the boundary points $\lambda_0 := 0$ and $\lambda_{N+1} := 1$. Then ν is the law of the random set

  $$\mathcal{J} = \left\{ i \in [L] : \exists k \leq \frac{N+1}{2} \ \text{ s.t. } \ \frac{i}{L} \in [\lambda_{2k}, \lambda_{2k+1}) \right\}$$

  This is the most general recombination measure one can wish for.

Biologically speaking, a crucial feature of ν is that except under free recombination, loci which are very close do not recombine very often. Specifically, in the examples given (except free recombination) one may check that if ℓ and ℓ' are very close, then picking $\mathcal{I}$ with law ν, we have

$$\mathbb{P}[\{\ell, \ell'\} \subset \mathcal{I}] + \mathbb{P}[\{\ell, \ell'\} \subset \mathcal{I}^c] \sim \frac{|\ell - \ell'|}{L-1}.$$

where $\mathcal{I}^c := [L] \setminus \mathcal{I}$ is the complement of $\mathcal{I}$. This means, in the notation of Definition 1.2.1, that if the new genome g^* inherits its content at locus ℓ from g^1, then it is very likely to also inherit its content at locus ℓ' from g^1.

We now illustrate how Mendelian inheritance can be used to model the evolution of a population.

1.2.2 Models from genetics

The population is described as a probability on the set of genotypes, that is, an element of $\mathbb{X}^{[L]} := \mathcal{P}(\square_{[L]})$. The typical biological forces shaping an isolated population are recombination (R), selection (S), mutation (M), and genetic drift (D).

For a population $\mathbf{x} \equiv (x(\gamma))_{\gamma \in \square_{[L]}} \in \mathbb{X}^{[L]}$ and a function f on $\square_{[L]}$, we write

$$\mathbf{x}[f(g)] = \sum_{\gamma \in \square_{[L]}} f(\gamma) x(\gamma).$$

Mathematically speaking, $\mathbf{x}[f(g)]$ is the expectation of $f(g)$, where g is a $\square_{[L]}-$valued random variable with law $\mathbf{x}$. Under this notation, the frequency of the $+1$ allele at locus ℓ is $\mathbf{x}[\mathbb{1}_{[g_\ell=+1]}]$.

In this section, we formally introduce an individual-based model and describe its large-population limit, obtaining a diffusion approximation (1.5) which jointly models mutation, recombination, selection and genetic drift. We discuss a collision-based interpretation of this SDE and describe the large-recombination limit.

An individual-based model

Here we introduce the individual-based model which will be used in all of our simulations (parallel to the one presented in Section 1.1.2). The model will be specified with the following parameters

(R) the recombination measure $\nu^{(N)} \in \mathcal{P}(\mathbb{R})$ and recombination probability $\rho^{(N)} \in (0,1)$

(S) the log-fitness function $W^{(N)} : \square_{[L]} \to \mathbb{R}$

(M) the mutation probabilities at each locus $(\mu_\ell^{(N)+}, \mu_\ell^{(N)-}) \in [0,1]^2$.

(D) the population size N.

At time t, the population contains N genomes with genotypes $g^1, \ldots, g^N$. The population is therefore

$$\mathbf{X}_t^{(N)} = \frac{1}{N} \sum_{k \in [N]} \delta_{g^k}.$$

Every time step

1. *Selection+Genetic drift.* For each $k_0 \in [N]$, two parents k_1, k_2 are picked at random independently, such that the probability that organism number k is picked is proportional to $e^{W^{(N)}(g^k)}$

2. *Recombination.* With probability $\rho^{(N)}$, the genotype of the offspring g^* of (k_1, k_2) is obtained following Mendelian inheritance from Definition 1.2.1. Otherwise, $g^* = g^{k_1}$.

3. *Mutation.* For each locus $\ell \in [L]$ independently, with probability $\mu_\ell^{(N)+} + \mu_\ell^{(N)-}$, the locus mutates. This means re-sampling g_ℓ^* with law

$$\mathcal{L}_\ell := \frac{\mu_\ell^{(N)-}}{|\mu_\ell^{(N)}|} \delta_{-1} + \frac{\mu_\ell^{(N)+}}{|\mu_\ell^{(N)}|} \delta_{+1}$$

   where $|\mu_\ell^{(N)}| := \mu_\ell^{(N)+} + \mu_\ell^{(N)-}$. We thus obtain a new offspring genome g^{*k_0}.

4. The new population is $\frac{1}{N} \sum_{k \in [N]} \delta_{g^{*k}}$.

Remark 4. *Here as in Section 1.1.2, we assume a panmictic hermaphroditic isolated unstructured population with discrete generations, in a constant environment. The model does not account for non-Mendelian inheritance, such as epigenetics or mitochondrial DNA [55].*

Remark 5. *I am not aware of any work on the polygenic limit which accounts for the fact that the mutation probabilities* $(\mu_\ell^{(N)})_{\ell \in [L]}$ *are not constant across loci. Though, as will be argued in Chapter 4, this is easy to add to the model, and biologically realistic [56].*

Scaling to the LD-Wright-Fisher diffusion

Let $N \to +\infty$ with

$$\rho^{(N)} = \frac{\rho}{N} \qquad W^{(N)} = \frac{W}{N} \qquad \forall \ell \in [L], \quad \mu_\ell^{(N)\pm} = \frac{\theta_\ell^\pm}{N} \tag{1.4}$$

for some constant $\rho, \theta_\ell \equiv (\theta_\ell^+, \theta_\ell^-)$ and a function $W : \square_{[L]} \to \mathbb{R}$. Then fundamental diffusion theory tells us $(\mathbf{X}_{\lfloor tN \rfloor}^{(N)})_{t \geq 0}$ converges to the following process

$$\mathrm{d}\mathbf{X}_t = (\rho R(\mathbf{X}_t) + \Theta(\mathbf{X}_t) + S(\mathbf{X}_t))\mathrm{d}t + \Sigma(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t \tag{1.5}$$

with the operators that we now describe. The convergence is here implied in the Skorokhod J1 topology [57]. It should be remembered that $\mathbb{X}^{[L]}$ is isomorphic to the simplex of $\mathbb{R}^{2^L}$, so we may think of the operators R, Θ, S as being $\mathbb{R}^{2^L}-$valued.

Recombination. For a subset $\mathcal{I} \subseteq [L]$ and $\mathbf{x} \in \mathbb{X}_L$, define $\mathbf{x}^{\mathcal{I}}$ the marginal of $\mathbf{x}$ on the hypercube $\square_{\mathcal{I}} := \{-1,+1\}^{\mathcal{I}}$. Let $\mathbf{x}^{\mathcal{I}} \otimes \mathbf{x}^{\mathcal{I}^c}$ be the product measure on $\square_{[L]}$ of $\mathbf{x}^{\mathcal{I}}$ and $\mathbf{x}^{\mathcal{I}^c}$. Then we define the recombinator as

$$R : \begin{cases} \mathbb{X}^{[L]} & \longrightarrow \mathbb{R}^{\square_{[L]}} \\ \mathbf{x} & \longmapsto \sum\limits_{\emptyset \subsetneq \mathcal{I} \subsetneq [L]} \nu(\mathcal{I})(\mathbf{x}^{\mathcal{I}} \otimes \mathbf{x}^{\mathcal{I}^c} - \mathbf{x}) \end{cases}$$

Note that up to replacing ν with $\tilde{\nu} : \mathcal{I} \mapsto \frac{\nu(\mathcal{I})+\nu(\mathcal{I}^c)}{2}$, we can and will assume that for any $\mathcal{I} \subseteq [L]$, $\nu(\mathcal{I}) = \nu(\mathcal{I}^c)$. The recombinator has been extensively studied in the deterministic setting [58, 59, 60]. It is a mixing operator for which entropy production bounds are known [61].

Mutation. The mutator is defined as

$$\Theta : \begin{cases} \mathbb{X}^{[L]} & \longrightarrow \mathbb{R}^{\square_{[L]}} \\ \mathbf{x} & \longmapsto \sum\limits_{\ell \in [L]} |\theta_\ell| \left(\mathbf{x}^{[L]\setminus\{\ell\}} \otimes \mathcal{L}_\ell - \mathbf{x}\right) \end{cases}$$

where

$$|\theta_\ell| = \theta_\ell^+ + \theta_\ell^- \qquad\qquad \mathcal{L}_\ell := \frac{\theta_\ell^-}{|\theta_\ell|}\delta_{-1} + \frac{\theta_\ell^+}{|\theta_\ell|}\delta_{+1}.$$

Selection. The operator $S : \square_{[L]} \to \mathbb{R}^{\square_{[L]}}$ is the selector defined with

$$S(\mathbf{x})(\gamma) := x(\gamma)(W(\gamma) - \mathbf{x}[W(g)]) = \mathbf{Cov}_{\mathbf{x}}\left[W(g), \mathbb{1}_{[g=\gamma]}\right]$$

where $\mathbf{Cov}_{\mathbf{x}}[\cdot,\cdot]$ are the expectation and the covariance function for a random genotype g with law $\mathbf{x}$.

Genetic Drift. The stochastic term is the traditional multiallele Wright-Fisher diffusion term [62], which corresponds to the Fleming-Viot noise term. We consider a Gaussian process $\mathbf{B} \equiv (B_t(\gamma^1,\gamma^2))_{t\in[0,T];\gamma^1,\gamma^2\in\square_{[L]}}$ indexed by $\square_{[L]} \times \square_{[L]}$ such that

$$\forall \gamma^1,\gamma^2,\gamma^3,\gamma^4 \in \square_{[L]}, \qquad \mathrm{d}\left\langle B(\gamma^1,\gamma^2), B(\gamma^3,\gamma^4)\right\rangle_t = (\mathbb{1}_{\begin{bmatrix}\gamma^1=\gamma^3\\ \gamma^2=\gamma^4\end{bmatrix}} - \mathbb{1}_{\begin{bmatrix}\gamma^1=\gamma^4\\ \gamma^2=\gamma^3\end{bmatrix}})\mathrm{d}t$$

To put it differently, $B(\gamma^1,\gamma^2) = -B(\gamma^2,\gamma^1)$, and $B(\gamma^1,\gamma^2), B(\gamma^3,\gamma^4)$ are independent Brownian motions if $(\gamma^1,\gamma^2) \notin \{(\gamma^3,\gamma^4),(\gamma^4,\gamma^3)\}$.

Finally, let $\mathcal{M}(\square_{[L]} \times \square_{[L]}, \mathbb{R}^{\square_{[L]}})$ denote the space of linear functions from $\square_{[L]} \times \square_{[L]}$ to $\mathbb{R}^{\square_{[L]}}$. Then

$$\Sigma : \mathbb{X}^{[L]} \to \mathcal{M}\left(\square_{[L]} \times \square_{[L]}, \mathbb{R}^{\square_{[L]}}\right)$$

is defined such that

$$\forall \gamma \in \square_{[L]}, \qquad (\Sigma(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t)(\gamma) := \sum_{\hat{\gamma}\neq\gamma} \sqrt{X_t(\gamma)X_t(\hat{\gamma})}\mathrm{d}B_t(\gamma,\hat{\gamma}).$$

Comments on equation (1.5). Equation (1.5) has been remarkably under-studied in full generality by mathematicians. It is admittedly unrealistic for biological purposes, as it assumes the 2^L possible genotypes to simultaneously exist at any given time point.

Nevertheless it presents the mathematical convenience of diffusion theory and simultaneously accounts for all four forces of biological interest: recombination, selection, mutation and genetic drift. Existence and uniqueness of solutions to (1.5) was obtained from the martingale problem in [63]. When $L = 1$, it reduces to the standard Wright-Fisher diffusion (see Chapter 4). The case $L = 2$ was studied by Ohta and Kimura in [64, 65], and their work is a testament to how involved computations become even with just two loci. For a general L, $\mathbb{X}^{[L]}$ is of dimension $2^L - 1$ and computations become even more complex. When $S = 0$, [66] derived a dual process. When $\Sigma = 0$, there is a rich literature reviewed in [67] and Chapter II of [68].

One important feature of $(\mathbf{X}_t)_{t\geq 0}$ is the presence of **linkage disequilibrium (LD)**

$$\mathbf{Cov}_{\mathbf{X}_t}[g_{\ell_1}, g_{\ell_2}] \neq 0$$

For this reason, we will call (1.5) the **LD-Wright-Fisher diffusion**. It will be the focus of Chapter 2.

The collision point of view

Equation (1.5) is interesting from the perspective of kinetics theory, in that the system can be seen as a particle system evolving under three types of two-particle collisions and a Markovian generator. Specifically, it can be obtained from a finite Moran-style model in continuous time, in which the population at time t is composed of $g_t^1, \dots, g_t^N \in \square_{[L]}$ and has the following transitions

- *(Mixing collisions).* For any $i, j \in [N]$ with $i \neq j$, and any subset $\mathcal{I} \subseteq [L]$, we have the transition
$$(g_t^i, g_t^j) \longrightarrow (g_t^{i|\mathcal{I}} \otimes g_t^{j|\mathcal{I}^c}, g_t^{j|\mathcal{I}} \otimes g_t^{i|\mathcal{I}^c})$$
with rate $\rho\nu(\mathcal{I})/N$, where $g^{|\mathcal{I}} := (g_\ell)_{\ell\in\mathcal{I}}$ is the restriction of g to $\mathcal{I}$ and for $\mathcal{I} \subseteq [L]$, $g^a \in \square_{\mathcal{I}}, g^b \in \square_{\mathcal{I}^c}$, we let $g^a \otimes g^b$ be the element of $\square_{[L]}$ such that
$$\forall \ell \in [L], \qquad (g^a \otimes g^b)_\ell = \begin{cases} g_\ell^a & \text{if } \ell \in \mathcal{I} \\ g_\ell^b & \text{otherwise.} \end{cases}$$
- *(Selective collisions).* For any $i, j \in [N]$ with $W(g_t^i) > W(g_t^j)$, we have
$$(g_t^i, g_t^j) \longrightarrow (g_t^i, g_t^i)$$
with rate $(W(g_t^i) - W(g_t^j))/N$.
- *(Mutations).* For any $\ell \in [L], i \in [N]$,
$$g_{t,\ell}^i \longrightarrow \breve{g}_\ell$$
with rate $|\theta_\ell|$, where $\breve{g}_\ell$ is randomly sampled with law $\mathcal{L}_\ell$.
- *(Cloning collisions).* For any $i, j \in [N]$ with $i \neq j$ we have
$$(g_t^i, g_t^j) \longrightarrow (g_t^i, g_t^i)$$
with rate 1/2.

Define $\mathbf{X}_t \equiv (X_t(\gamma))_{\gamma\in\square_{[L]}}$ such that $X_t(\gamma)$ is the frequency of the γ genotype at time t

$$X_t(\gamma) := \frac{1}{N} \sum_{i\in[N]} \mathbb{1}_{[\gamma=g_t^i]}.$$

Scaling $N \to +\infty$, first and second-moment computations show that $(\mathbf{X}_t)_{t\geq 0}$ converges to a solution of (1.5). This shows in particular that the selector S and the recombinator R belong to a class of operators known as **collision operators** [69]. More insight on this point of view will be obtained on this in Chapter 3.

Scaling to the LE-Wright-Fisher diffusion

If we write $P_t^\ell := \mathbf{X}_t[\mathbb{1}_{[g_\ell=+1]}]$ for the frequency of the $+1$ allele at locus $\ell \in [L]$, it will be seen in Corollary 2.3.5 (Chapter 2) that (1.5) implies

$$\mathrm{d}P_t^\ell = s_\ell(\mathbf{X}_t)P_t^\ell(1-P_t^\ell)\mathrm{d}t + (\theta_\ell^+(1-P_t^\ell) - \theta_\ell^- P_t^\ell)\mathrm{d}t + \sqrt{P_t^\ell(1-P_t^\ell)}\mathrm{d}B_t^\ell$$

where B^ℓ is a Brownian motion and

$$s_\ell(\mathbf{x}) := \frac{\mathbf{Cov_x}\left[W(g), g_\ell\right]}{\mathbf{Var_x}[g_\ell/2]}.$$

If we let the recombination rate ρ go to infinity, assuming the recombination measure ν is non-degenerate, two crucial events occur (see Chapter 2 for details)

- $\mathbf{X}_t$ is forced on a stable manifold in which it is a product measure, entirely determined by $(P_t^\ell)_{\ell\in[L]}$.

  $$\mathbf{X}_t = \bigotimes_{\ell\in[L]} ((1-P_t^\ell)\delta_{\{-1\}} + P_t^\ell\delta_{\{+1\}}) \tag{1.6}$$

  The manifold determined by (1.6) is the manifold of **Linkage Equilibrium**[5] **(LE)**, sometimes called the Wright manifold [71], which will be denoted $\Gamma^{[L]}$.

- The Brownian motions $(B^\ell)_{\ell\in[L]}$ become independent.

As a consequence, the dynamics of $(P_t^\ell)_{\ell\in[L]}$ can be treated as an autonomous set of L SDE, which we call the **LE-Wright-Fisher diffusion**. Specifically, we write

$$\mathrm{d}P_t^\ell = \bar{s}_\ell(\mathbf{P}_t)P_t^\ell(1-P_t^\ell)\mathrm{d}t + (\theta_\ell^+(1-P_t^\ell) - \theta_\ell^- P_t^\ell)\mathrm{d}t + \sqrt{P_t^\ell(1-P_t^\ell)}\mathrm{d}B_t^\ell \tag{1.7}$$

where $\mathbf{P}_t \equiv (P_t^\ell)_{\ell\in[L]}$, B^ℓ is a Brownian motion and

$$\bar{s}_\ell(\mathbf{p}) := \frac{\mathbf{Cov_p}\left[W(g), g_\ell\right]}{\mathbf{Var_p}[g_\ell/2]}. \tag{1.8}$$

where here, $\mathbf{Var_p}$ and $\mathbf{Cov_p}$ are the variance and covariance of the random variable g such that $(g_\ell)_{\ell\in[L]}$ are independently sampled such that g_ℓ has law $((1-p_\ell)\delta_{-1} + p_\ell\delta_{+1})$.

This system is convenient to work with, and in particular it has a known stationary distribution [72] which was obtained by Kimura in [53], confirming an earlier derivation by Wright in [73]. When $L = 1$, the (one-dimensional) Wright-Fisher diffusion given by (1.7) is very well known [74, 75, 76, 77].

The convergence to LE and stability of equilibria, when L is fixed and $\Sigma = \Theta = 0$ (no mutation, no genetic drift) has received much attention (reviewed in Chapter II and Chapter V.4 of [68]). It has recently been proved [78] that mixing by free recombination (see Section 1.2.1) leads to the "cutoff phenomenon", well-known to probabilists working on mixing systems. This means that considering $(\mathbf{X}_t)_{t\geq 0}$ evolving under recombination R alone, the distance between $\mathbf{X}_t$ and $\Gamma^{[L]}$ converges to a step function of t as $L \to +\infty$, the cutoff time being $t_c = \ln_2(L)$.

---

[5] Not to be confused with Quasi-Linkage Equilibrium (QLE), which is a manifold close to the LE manifold, defined by the first-order perturbations of the LE manifold due to selection (chapter II.6 of [68]). Statistical physicists use a similar but slightly different definition of QLE [70].

1.2.3 Diploidy

Diploidy refers to the situation in which an organism has two haploid genomes. In this section we discuss how the model from the previous section can be modified to account for diploidy. This will require that we adapt the notation. Specifically, compared with haploidy, the genome of an organism is described by an element $G = (G_{\ell,(1)}, G_{\ell,(2)})_{\ell\in[L]}$ of $(\{0,1\}^2)^L$ (here, the alleles are labelled 0 and 1 instead of -1 and $+1$ to stay closer to the terminology of the literature).

Under diploidy the law of inheritance is modelled as

Definition 1.2.2 (Mendelian inheritance in diploids)**.** *If two genomes $G^1, G^2 \in (\{0,1\}^2)^L$ have an offspring with genome G^*, then we sample independently two subset $\mathcal{I}_1, \mathcal{I}_2 \subseteq [L]$ with the same probability distribution ν. Then for every locus $\ell \in [L]$, and $i \in \{0,1\}$*

$$G^*_{\ell,(i)} = \begin{cases} G^i_{\ell,(1)} & \textit{if } \ell \in \mathcal{I}_i \\ G^i_{\ell,(2)} & \textit{otherwise} \end{cases}$$

It is then straightforward to adapt the individual-based model from Section 1.2.2, with the same parameters $N \in \mathbb{N}, \rho^{(N)} \in [0,1], (\mu_\ell^{(N)\pm})_{\ell\in[L]} \in [0,1]^{2L}$, a probability ν on the subsets of $[L]$ and a fitness function $W^{(N)}$.

In this setting, the population at generation n is a probability distribution $\mathbf{X}_n^{(N)}$ on $(\{0,1\}^2)^L$. Letting $2N$ go to $+\infty$ with strong recombination $\rho^{(N)} \gg 1/N$ and the following analog to (1.4)

$$W^{(N)} = \frac{W}{2N} \qquad\qquad \forall \ell \in [L], \qquad \mu_\ell^{(N)\pm} = \frac{\theta_\ell^\pm}{2N}$$

the population mixes analogously to what was seen in Section 1.2.2 and reaches LE, meaning $(\mathbf{X}^{(N)}_{\lfloor t2N\rfloor})_{t\in[0,T]}$ converges to a diffusion $(\mathbf{X}_t)_{t\in[0,T]}$ such that the $(G_\ell)_{\ell\in[L]}$ are independent under $\mathbf{X}_t$. Furthermore, if G is a random variable with law $\mathbf{X}_t$, the two haploid genomes $(G_{\ell,(1)})_{\ell\in[L]}$ and $(G_{\ell,(2)})_{\ell\in[L]}$ are independent. This follows from the fact that a genome G inherits its first chromosome $G_{\cdot,(1)}$ from its first parent and its second chromosome $G_{\cdot,(2)}$ from its second parent, which under panmixia are sampled independently. By analogy with the Wright manifold, we call the manifold of $\mathcal{P}((\{0,1\}^2)^L)$ on which this last condition is true is the **strong Hardy-Weinberg manifold (HW)**. Under HWLE, it is the same to randomly sample a genome G at time t in the population and to independently sample $(G_{\ell,(i)})_{\ell\in[L],i\in[2]}$ such that $G_{\ell,(i)}$ has law $Bernoulli(P_t^\ell)$, with P_t^ℓ the frequency of the $+1$ allele at time t. It follows that the population is entirely described by the frequencies of the $+1$ allele at each locus $(P_t^\ell)_{\ell\in[L]}$. For $\mathbf{p} \in [0,1]^L$, we write $\hat{\mathbf{E}}_\mathbf{p}$ for the corresponding population. In particular, for a measurable function $f : (\{0,1\}^2)^L \to \mathbb{R}$, $\hat{\mathbf{E}}_\mathbf{p}[f(G)]$ is the expectation of $f(G)$ where $G \equiv (G_{\ell,(1)}, G_{\ell,(2)})_{\ell\in[L]}$ has law $\hat{\mathbf{E}}_\mathbf{p}$.

The analog to the LE-Wright-Fisher diffusion in the diploid case is the **HWLE-Wright-Fisher diffusion** for the frequencies of the $+1$ allele at each locus $\mathbf{P}_t \equiv (P_t^\ell)_{\ell\in[L]}$ is

$$\mathrm{d}P_t^\ell = \hat{s}_\ell(\mathbf{P}_t)P_t^\ell(1-P_t^\ell)\mathrm{d}t + (\theta_\ell^+(1-P_t^\ell) - \theta_\ell^- P_t^\ell)\mathrm{d}t + \sqrt{P_t^\ell(1-P_t^\ell)}\mathrm{d}B_t^\ell \tag{1.9}$$

where $(B^\ell)_{\ell\in[L]}$ are independent Brownian motions and for any $\mathbf{p} \equiv (p^\ell)_{\ell\in[L]} \in [0,1]^L$

$$\hat{s}_\ell(\mathbf{p}) := \frac{\hat{\mathbf{Cov}}_\mathbf{p}\left[W(G), |G_\ell|\right]}{\hat{\mathbf{Var}}_\mathbf{x}[|G_\ell|]} \tag{1.10}$$

where for $\mathbf{p} \in [0,1]^L$, $\hat{\mathbf{Cov}}_\mathbf{p}$ and $\hat{\mathbf{Var}}_\mathbf{p}$ are the covariance and variance associated to $\hat{\mathbf{E}}_\mathbf{p}$.

1.3 Polygenic trait models

Here, I will motivate the introduction of polygenic model, formally define them, and briefly go over the historical development of these models as well as their current uses. I will explain how the main polygenic trait models used today were developed. I warn the reader not to expect mathematical rigor, and that some results will be simplified for ease of reading, at the cost of reducing their extent.

1.3.1 Historical motivation for polygenic trait models

Historically, the development of polygenic trait models can be traced back to the debate between the biometricians and the Mendelians at the turn of the twentieth century. The biometricians, as emphasized in Section 1.1, focused on evolution through natural selection acting on traits which satisfied the infinitesimal model. The Mendelians focused on simple genetic systems in which one or two Mendelian loci determine a polymorphism. For a variety of reasons including insufficiently specified models, mathematical mistakes, difficult communication and personal enmity, the two communities could not reconcile their respective visions into a unifying framework (see Chapter 3 of [50]). The missing ingredient was a model under which the genetic value of the trait z_G from the infinitesimal model (definition 1.1.3) is seen as a function $Z^{(L)}(g) \equiv Z(g)$ of $L \gg 1$ genes $(g_\ell)_{\ell \in [L]}$, which are inherited following Mendel's laws (definition 1.2.1). Such a model was pointed to by Yule in 1906 ([50], Chapter 3), but was only established as a unifying framework following Fisher's landmark article in 1918 [79].

Before discussing this, let us clarify what we mean by a polygenic model.

1.3.2 Definition

We will define the **polygenic limit** as the limit of a gene-centric system (either the LD-Wright-Fisher diffusion (1.5) or the LE-Wright-Fisher diffusion (1.7)) when the number of loci is large $L \gg 1$. If this limit is obtained jointly with the limit distribution of $Z(g)$, where $Z : \square_{[L]} \to \mathbb{R}$ is a trait function, then we speak of the **polygenic trait limit**.

I will call a model polygenic whenever this model explicitly describes

1. the stochastic evolution of allele frequencies at any given locus[6].
2. the dynamics of the macroscopic observables emerging from all loci.

I call a model a **polygenic trait model** if it models a third process on top of the two ones enumerated above

3. the dynamics of the distribution of $Z(g)$ under $\mathbf{X}_t$ for some trait function Z.

In the next section, we discuss the usual additivity assumption on the trait function Z.

1.3.3 The additive model

The simplest model for $Z(g)$ is the **additive model**, devised by Fisher [79] and Wright [84]. Under the additive model, the genetic value of a trait is obtained as the sum of the contributions of a large number of underlying genes, which are inherited following Mendel's

---

[6] This excludes a large class of models which assume a deterministic evolution at a given locus, due to selection being much stronger than genetic drift (the classical example being the selective sweep [80]). This is known in the literature as the oligogenic regime [81]. Our definition also excludes Kimura's infinite-alleles model [82] which considers the evolution of the population at one locus to be deterministic. This model requires the per-locus mutation rates $(\theta_\ell^\pm)_{\ell \in [L]}$ to be much greater than the per-locus recombination rate ρ/L, which is now deemed biologically unrealistic [83].

laws (Definition 1.2.1). As will be seen in Section 1.3.5, this model lets us recover the infinitesimal model from Definition 1.1.3 when the number of loci L is large.

To clarify, we take the Mendelian model of inheritance, and we define the following function in the haploid case

$$Z : \begin{cases} \square_{[L]} & \longrightarrow \mathbb{R} \\ \gamma & \longmapsto z_0 + \sum_{\ell \in [L]} \alpha_\ell \gamma_\ell \end{cases} \tag{1.11}$$

where $\alpha_\ell \in \mathbb{R}$ is the additive effect at locus ℓ. Functions of the form (1.11) are called **(haploid) additive traits**.

Similarly, an additive diploid trait function is defined as

$$Z : \begin{cases} (\{0,1\}^2)^L & \longrightarrow \mathbb{R} \\ G & \longmapsto z_0 + \sum_{\ell \in [L]} \alpha_\ell |G_\ell| \end{cases} \tag{1.12}$$

Common extensions of the additive model will be considered in Chapter 4: these allow us to incorporate pleiotropy, epistasis and dominance. We quickly go through them.

Pleiotropy

Pleiotropy refers to models in which a given locus influences more than one trait. To this end, we take the parameter α_ℓ to be an element of $\mathbb{R}^d$ for some parameter d, and Z defined as in (1.11) is a function from $\square_{[L]}$ to $\mathbb{R}^d$. Proponents of the omnigenic model [85] argue that pleiotropy is pervasive, meaning loci which affect one additive trait tend to affect all other additive traits.

Epistasis

The haploid trait function Z is said to be **functionally epistatic** if we can find $A \subsetneq [L]$, $g^1, g^2 \in \square_A$ and $g^3, g^4 \in \square_{[L]\smallsetminus A}$ such that

$$Z(g^1 \otimes g^3) + Z(g^2 \otimes g^4) \neq Z(g^1 \otimes g^4) + Z(g^2 \otimes g^3).$$

In particular, additive traits (1.11) are not functionally epistatic: for g^1, g^2, g^3, g^4 as above, we have

$$\begin{aligned} Z(g^1 \otimes g^3) + Z(g^2 \otimes g^4) =& \sum_{\ell \in A} g^1_\ell + \sum_{\ell \in [L]\smallsetminus A} g^3_\ell + \sum_{\ell \in A} g^2_\ell + \sum_{\ell \in [L]\smallsetminus A} g^4_\ell \\ =& Z(g^1 \otimes g^4) + Z(g^2 \otimes g^3) \end{aligned}$$

Functional epistasis in the diploid case is defined similarly.

Dominance

Recall from Section 1.2.3 that for **diploids**, the gene content $|G_\ell|$ at locus ℓ takes values in $\{0,1,2\}$ instead of $\{-1,+1\}$. The definition of an additive trait and epistasis in diploids is the analog of (1.11). A non-additive trait Z is said to exhibit **functional dominance** at locus ℓ if Z is not a linear function of $|G_\ell|$. Formally, we can find $G \in (\{0,1\}^2)^{[L]\smallsetminus\{\ell\}}$ such that

$$Z(G \otimes (1,1)) + Z(G \otimes (0,0)) \neq Z(G \otimes (1,0)) + Z(G \otimes (0,1)).$$

It can be checked again that additive traits (1.12) do not display functional dominance because

$$Z(G \otimes (1,1)) - Z(G \otimes (1,0)) = \alpha_\ell = Z(G \otimes (0,1)) - Z(G \otimes (0,0)).$$

Diploidy also brings the possibility of *cis* effects of gene expression[28], which would correspond to a trait function Z satisfying for some $G \in (\{0,1\}^2)^{[L] \setminus \{\ell\}}$

$$Z(G \otimes (0,1)) \neq Z(G \otimes (1,0))$$

Cis effects will be ignored throughout this work.

In the next section, we discuss how the additive model can be used in practice for real data. This crucially relies on the fact that any trait function Z in a given HWLE population can be locally decomposed into additive, dominance and epistasis components by linear regression, which is the method used in empirical studies to describe the genetic architecture of a trait in a given population.

1.3.4 The additive projection of a trait

One major goal of field work on quantitative traits is to determine the underlying **genetic architecture**, that is, to infer statistical properties of $(Z, \mathbf{X})$. The genetic architecture of a polygenic trait is important to answer the following questions

- How much of the genetic variance $\mathbf{Var}_{\mathbf{X}}[Z(g)]$ can be explained by an additive model ?
- Is the genetic variance mostly determined by a few loci with large effects, or many loci with small effects ?
- How is the variability within the population shaped by the forces of natural selection, random genetic drift, mutation and recombination ?

Fisher [79] suggested that a given pair $(Z, \mathbf{p})$, where $Z : \{0,1,2\}^L \to \mathbb{R}$ is a diploid trait function and $\mathbf{p} \in [0,1]^L$ is a HWLE population, can be decomposed into additive, dominance and epistasis components. I will summarize how this is done in diploid (the haploid case is analogous), following the method used in [86] for the decomposition. Recall that in a diploid model, $\mathbf{E_p}$ is a probability distribution on $(\{0,1\}^2)^L$ such that if G has law $\mathbf{E_p}$, then for any $\ell \in [L], i \in [2]$, $G_{\ell,(i)}$ is a *Bernoulli*(p^ℓ) variable, and the notation $|G_\ell| := G_{\ell,(1)} + G_{\ell,(2)}$.

Define the additive, dominance and epistasis projections of Z under $\mathbf{x}$ as follows

$$\begin{aligned}
&Z_{Add} : \left\{ \begin{array}{ll} (\{0,1\}^2)^L & \longrightarrow \mathbb{R} \\ \gamma & \longmapsto \mathbf{E_p}[Z(G)] + \sum_\ell \frac{\mathbf{Cov_p}[Z(G), |G_\ell|]}{\mathbf{Var_p}[|G_\ell|]} (|\gamma_\ell| - 2p_\ell) \end{array} \right. \\
&Z_{Dom} : \left\{ \begin{array}{ll} (\{0,1\}^2)^L & \longrightarrow \mathbb{R} \\ \gamma & \longmapsto \sum_\ell \frac{\mathbf{Cov_p}[Z(G), (G_{\ell,(1)} - p_\ell)(G_{\ell,(2)} - p_\ell)]}{\mathbf{Var_p}[G_{\ell,(1)}]^2} (\gamma_{\ell,(1)} - p_\ell)(\gamma_{\ell,(2)} - p_\ell) \end{array} \right. \\
&Z_{Epi} := Z - Z_A - Z_D
\end{aligned}$$

In particular, Z_{Add} is simply the linear regression of Z on the uncorrelated variables $(|G_\ell|)_{\ell \in [L]}$ under $\mathbf{E_p}$. It can be checked that the covariances between Z_{Add}, Z_{Dom} and Z_{Epi} are zero, which yields

$$\mathbf{Var}_{\mathbf{E_p}}[Z(G)] = \mathbf{Var}_{\mathbf{E_p}}[Z_{Add}(G)] + \mathbf{Var}_{\mathbf{E_p}}[Z_{Dom}(G)] + \mathbf{Var}_{\mathbf{E_p}}[Z_{Epi}(G)]$$

The term on the left-hand side is called the genetic variance of the trait Z in the population $\mathbf{p}$, the terms on the right-hand side are respectively the additive, dominance and epistasis variances. The pair $(Z, \mathbf{p})$ is said to be **statistically epistatic** if it has nonzero epistasis variance. Functional epistasis of Z is necessary but non-sufficient for statistical epistasis of $(Z, \mathbf{p})$. For instance, if $\mathbf{p} = (0, 0, \ldots, 0)$, then the pair $(Z, \mathbf{p})$ is never functionally epistatic.

This decomposition in additive, dominance and epistasis components can be extended to more than two alleles per locus, see Chapter II.3 of [68]. It is also known from [87] that the decomposition into additive and non-additive variance can be achieved outside HWLE, but contrary to what is sometimes claimed[7], it is not possible to decompose genetic variance into additive, dominance and epistasis variance in the general case [89]. Assuming the population is in HWLE, fitting the animal model to empirical measurements (Chapter 19 of [2]) yields estimates of the additive, dominance and epistasis variance of a given trait, which quantifies whether the trait is well approximated from an additive model.

We may now discuss the profound link between the additive model and the infinitesimal model from Definition 1.1.3.

1.3.5 Convergence to the infinitesimal model

The crucial result, proved in [11], is that as $L \to +\infty$, the inheritance pattern of $Z(G)$ converges to the infinitesimal model under reasonable assumptions on the distribution of α_ℓ (which depends on L). Informally, considering a diploid additive trait function Z, assuming

- the $(\alpha_\ell)_{\ell \in [L]}$ are i.i.d variables uniformly bounded by a constant $C/\sqrt{L}$.
- we start from a "well-mixed" population of independent genomes $G^{0,1}, \dots, G^{0,N} \in (\{0,1\}^2)^L$ independently sampled under a HWLE distribution $\mathbf{X}_0$.
- we produce new genomes $(G^{m,i})_{i \in [N], m \in [n]}$ such that $G^{m,i}$ is produced by Mendelian inheritance from two genomes G^{m-1,j_1}, G^{m-1,j_2} under free recombination (see Section 1.2.1). We keep track of the **pedigree** (see Definition 1.1.1), that is, an oriented graph $\mathscr{P}$ on $\mathbb{N}^2$ such that $(i, m+1)$ is connected to (j, m) iff $G^{m,j}$ is a parent of $G^{m,i}$.
- we produce a genome $G^{n+1,1}$ by Mendelian inheritance from $G^{n,1}$ and $G^{n,2}$.
- two technical conditions to account for inbreeding and unlikely recombination events

then, conditional on $Z(G^{n,1}) = z^1$ and $Z(G^{n,2}) = z^2$ and $\mathscr{P}$, letting $L \to +\infty$ with $n \ll \sqrt{L}$, the distribution of $Z(G^{n+1,1})$ converges to

$$\mathcal{N}\left(\frac{z^1 + z^2}{2}, V_S\left(1 - \mathcal{F}((n,1),(n,2))\right)\right)$$

where $\mathcal{F}((n,1),(n,2))$ is the inbreeding coefficients between $G^{n,1}$ and $G^{n,2}$, defined from $\mathscr{P}$ as in Definition 1.1.2 and the segregation variance is

$$V_S = \lim_{L \to +\infty} \sum_{\ell \in [L]} (\alpha_\ell)^2 P_0^\ell (1 - P_0^\ell).$$

This result can easily be extended to haploids and to accomodate mutations (in which case we recover (1.1)) and pleiotropy (in which case the segregation variance V_S is replaced by a covariance matrix), and has been extended to models where $(Z, \mathbf{x})$ is statistically epistatic and shows dominance [90], though in these cases the limit phenotypic model is much more involved[8].

---

[7] In 2019, [88] claimed to have achieved this, but I wrote a computer program available on https://github.com/PhCourau/Orthogonal_decomp_additive_variance which proves using their method that there is non-zero covariance between the dominance and additive components.

[8] Specifically, conditional on the trait values of the parents z^1, z^2 and the pedigree $\mathscr{P}$, the trait value of the offspring is no longer independent from the other trait values of the population, and its variance depends on more microscopic parameters than V_S.

1.3.6 Wright's formula

Taking the limit $L \gg 1$ in the LE-Wright-Fisher diffusion (1.7) requires specifying a log-fitness function W on $\square_{[L]}$. In the context of the polygenic trait limit, the most natural method is to specify W as a function of Z. Specifically we set

$$W(g) = U(Z(g)) \tag{1.13}$$

for some function $U : \mathbb{R} \to \mathbb{R}$ (or $U : \mathbb{R}^d \to \mathbb{R}$ if we have a pleiotropic model). The goal is to let L go to infinity, letting α (and therefore Z) as well as U depend on L. Wright [73] made an ingenious observation[9] which would prove crucial to develop polygenic models. Define the mean fitness function

$$\bar{W} : \begin{cases} [0,1]^L & \longrightarrow \mathbb{R} \\ \mathbf{p} & \longrightarrow \bar{\mathbf{E}}_{\mathbf{p}}[W(g)] \end{cases}$$

where $\bar{\mathbf{E}}_{\mathbf{p}}$ is the element of $\Gamma^{[L]}$ with marginals $(p_\ell)_{\ell \in [L]}$

$$\bar{\mathbf{E}}_{\mathbf{p}} := \bigotimes_{\ell \in [L]} ((1 - p_\ell)\delta_{\{-1\}} + p_\ell \delta_{\{+1\}}).$$

The notation $\bar{\mathbf{E}}_{\mathbf{p}}$ is the haploid analog of $\hat{\mathbf{E}}_{\mathbf{p}}$ from Section 1.2.3.

Then the following holds true

$$\partial_\ell \bar{W}(\mathbf{p}) = \bar{\mathbf{E}}_{\mathbf{p}}[W(g) | g_\ell = +1] - \bar{\mathbf{E}}_{\mathbf{p}}[W(g) | g_\ell = -1].$$

In particular, simple computations show that the selection coefficient at locus ℓ given by (1.8) is

$$\bar{s}_\ell(\mathbf{p}) = \partial_\ell \bar{W}(\mathbf{p}). \tag{1.14}$$

The analog of [1.14] is true for diploids with the HWLE-Wright-Fisher diffusion (1.9).

Application of the formula to the polygenic limit

Consider a haploid additive trait Z as in (1.11). If we consider a population $\mathbf{p} \in [0,1]^{[L]}$ in LE (1.6), then the central limit theorem tells us that under reasonable assumptions on $(\alpha_\ell, p_\ell)_{\ell \in [L]}$ such as uniform boundedness (see Chapter V of [91] for details) then $Z(g)$ under $\bar{\mathbf{E}}_{\mathbf{p}}$ can be approximated as $L \to +\infty$ with a normal distribution $\mathcal{N}(\bar{z}(\mathbf{p}), V(\mathbf{p}))$ for some parameters $(\bar{z}(\mathbf{p}), V(\mathbf{p})) \in \mathbb{R} \times \mathbb{R}_+$.

Suppose log-fitness is determined by Z as in (1.13)

$$W(g) = U(Z(g)).$$

Then the normal approximation means we can compute $\bar{W}$ as

$$\bar{W}(\mathbf{p}) \simeq \bar{U}(\bar{z}(\mathbf{p}), V(\mathbf{p}))$$

where $\bar{U}(\bar{z}, V)$ is the expectation of $U(\mathcal{Z})$ for some variable $\mathcal{Z}$ with law $\mathcal{N}(\bar{z}, V)$ and

$$\bar{z}(\mathbf{p}) := \sum_{\ell \in [L]} \alpha_\ell p_\ell \qquad V(\mathbf{p}) := \sum_{\ell \in [L]} \alpha_\ell^2 p_\ell (1 - p_\ell).$$

Then applying Wright's formula to $W = U \circ Z$ we get

$$\bar{s}_\ell(\mathbf{p}) = \partial_{\bar{z}} \bar{U}(\bar{z}, V) \times \partial_\ell \bar{z} + \partial_V \bar{U}(\bar{z}, V) \times \partial_\ell V \tag{1.15}$$

where we dropped $\mathbf{p}$ to alleviate notation.

The next two sections are examples of applications of Wright's formula, specifically to directional and stabilizing selection.

---

[9] In his work, he considers discrete time, and the fitness function e^W instead of the log-fitness function W. I here simplify in an attempt to concisely derive the polygenic limit.

1.3.7 Limits to artificial selection

The simplest model on which to apply the previous method is directional selection on diploids, in which $U(z) = \beta z$ for some $\beta \in \mathbb{R}$. In this setting, we find

$$\hat{s}_\ell(\mathbf{p}) = \beta \alpha_\ell p_\ell (1 - p_\ell).$$

This result was used in one of the earliest derivations of a polygenic model: Robertson's work on the limits to artificial selection [92], which is of crucial importance to breeders [93]. The fundamental question is: considering a population of size N, applying sustained selection to increase the value of a quantitative trait with initial mean $\bar{z}_0$, what is the asymptotic behavior of $\bar{z}_t$, the mean trait value at time t ? The goal is to predict this long-term behavior as a function of the short-term behavior of the population, that is, the mean trait value in generation 0 and 1, corresponding to $\bar{z}_0$ and $\bar{z}_{1/(2N)}$.

Robertson's derivation assumes no mutations ($\theta = 0$) in the LE-Wright-Fisher diffusion (1.7), and constant additive effects ($\alpha_\ell = 1$). We also assume the $(P_0^\ell)_{\ell \in [L]}$ are i.i.d. We then get

$$\mathrm{d}P_t^\ell = 2N\beta P_t^\ell (1 - P_t^\ell)\mathrm{d}t + \sqrt{P_t^\ell (1 - P_t^\ell)}\mathrm{d}B_t^\ell. \tag{1.16}$$

In particular, the $(P_t^\ell)_{\ell \in [L]}$ evolve independently. From (1.12), the trait mean is

$$\bar{z}_t = \sum_{\ell \in [L]} 2P_t^\ell.$$

The asymptotic value of $\bar{z}_t$ is therefore

$$\bar{z}_\infty = \sum_{\ell \in [L]} 2P_\infty^\ell$$

where P_∞^ℓ is 1 if the trait-increasing locus at locus ℓ has reached fixation, and 0 otherwise. From diffusion theory, it can be found from [53]

$$\mathbb{P}[P_\infty^\ell = 1] = \frac{1 - e^{-4N\beta P_0^\ell} - 1}{1 - e^{-4N\beta}}$$

Using a the law of large numbers (Chapter IX of [91]), we write the approximation

$$\begin{aligned}\bar{z}_\infty \simeq & 2L\mathbb{E}\left[P_0^1\right] \\ = & 2L\mathbb{P}[P_0^1 = 1] \\ \simeq & 2L\mathbb{E}\left[\frac{1 - e^{-4N\beta P_0^1}}{1 - e^{-4N\beta}}\right]\end{aligned}$$

If we assume $4N\beta \ll 1$, a Taylor expansion yields

$$\bar{z}_\infty \simeq 2L\mathbb{E}\left[P_0^1 + 2N\beta P_0^1 (1 - P_0^1)\right]$$

On the other hand we have

$$\begin{aligned}\bar{z}_{1/(2N)} - \bar{z}_0 \simeq & \frac{1}{2N}\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \bar{z}_t \\ \simeq & \frac{1}{2N}\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} 2L\mathbb{E}\left[P_t^\ell\right] \\ = & 2L\beta\mathbb{E}\left[P_0^1 (1 - P_0^1)\right]\end{aligned}$$

using the law of large numbers.

We thus obtain from (1.16) the key result of Robertson

$$\bar{z}_\infty \simeq \bar{z}_0 + 2N(\bar{z}_{1/(2N)} - \bar{z}_0) \tag{1.17}$$

for diploids (for haploids, replace $2N$ with N).

This remarkable result is straightforward to obtain, because in this model the selection coefficient at locus ℓ given by $\hat{s}_\ell(\mathbf{P}_t)$ in (1.10) is only a function of P_t^ℓ, and therefore the LE-Wright-Fisher diffusion (1.7) can be seen as a set of L independent diffusions. The same kind of tools can be used to model positive [80] and purifying [94] selection, in which the effect of selection on a locus is independent of macroscopic observables.

Studying the evolution of a polygenic trait under directional selection is still an active area of research [95].

1.3.8 Stabilizing selection and the Latter-Bulmer model

For stabilizing selection let us define W with (1.13), where Z is an additive trait (1.11) and U is defined as follows

$$U(z) := -\frac{1}{(\omega_e)^2}(z-\eta)^2 \tag{1.18}$$

where η is the optimal trait value and ω_e^{-2} is the rescaled selection strength. Because of the scaling (1.4), this is equivalent to (1.2) if we set $\omega_e := \omega/\sqrt{N}$ (for haploids, for diploids N should be replaced with $2N$).

Using the reasoning just described, as well as scaling approximations which will be made clearer in Chapter 4, we find

$$\bar{s}_\ell(\mathbf{p}) \simeq -\frac{\alpha_\ell}{(\omega_e)^2}(\eta - \bar{z}(\mathbf{p})) - \frac{\alpha_\ell^2}{(\omega_e)^2}\left(\frac{1}{2} - x_\ell\right). \tag{1.19}$$

where

$$\bar{z}(\mathbf{p}) = \sum_{\ell\in[L]} 2\alpha_\ell p_\ell.$$

This was obtained by Wright in [96] (equation (19)). The first term drives the population to the optimum η, whereas the second term drives allele frequencies to the boundary $\{0,1\}$. This second term is often called “Robertson’s underdominant term”, referring to Robertson’s derivation in [23].

Carrying this into the LE-Wright-Fisher diffusion (1.7), it is possible, as will be seen in Chapter 4, to describe the equilibrium distribution of the population. This was originally done by Latter [97] and Bulmer in [98]. In particular, we will explain in Chapter 4 how to obtain a system of equations (4.28-4.30) such that

- Selection acts on a locus in a way that depends on the mean-field behavior of the other loci.
- The trait mean $\bar{z}_t = \mathbf{X}_t[Z(g)]$ evolves as an Ornstein-Uhlenbeck process, consistent with (1.3).
- The distribution of the trait $Z(g)$ under $\mathbf{X}_t$ is close to a normal distribution.

We thus see how powerful Wright’s formula (1.15) is, letting us describe the limit system of the LE-Wright-Fisher diffusion (1.7). One major limitation is that it assumes LD is negligible, which is not the case in natural populations. This is why other methods have been developped, to account for some small amount of loose LD among pairs of loci, which we now turn to.

1.3.9 Cumulant-based approaches to LD

One problem of the previous approach is that its starting point is the LE-Wright-Fisher diffusion (1.7), which assumes LE, whereas in a biologically realistic setting LE cannot be attained. There has however been remarkable effort to account for LD with cumulant-based approaches (see Chapter V of [68]).

At the heart of cumulant-based approaches is the cumulant distribution function. Specifically, for $\mathbf{x} \in \mathbb{X}^{[L]}$, define the **moment-generating function** of $\mathbf{x}$ as

$$\varphi_{\mathbf{x}} : \begin{cases} \mathbb{R}^{[L]} & \longrightarrow \mathbb{R} \\ \lambda \equiv (\lambda_\ell)_{\ell\in[L]} & \longmapsto \mathbf{x}\left[e^{\sum_{\ell\in[L]}\lambda_\ell g_\ell}\right] \end{cases}$$

The **cumulant-generating function** of $\mathbf{x}$ is then

$$\psi_{\mathbf{x}} := \ln \varphi_{\mathbf{x}}.$$

The cumulant-generating function $\psi_{\mathbf{x}}$ entirely characterizes $\mathbf{x}$. For a subset $\mathcal{I} \subseteq [L]$, the $\mathcal{I}-$cumulant of $\mathbf{x}$ is defined as

$$\chi_{\mathcal{I}}(\mathbf{x}) := \partial_{\mathcal{I}}\psi_{\mathbf{x}}(0)$$

where $\partial_{\mathcal{I}}$ is the partial derivative with respect to all the elements of $\mathcal{I}$. It can be seen when $\#\mathcal{I} > 1$

$$\chi_{\mathcal{I}}(\mathbf{x}) = \mathbf{x}\left[\prod_{\ell\in\mathcal{I}}(g_\ell - \mathbf{x}[g_\ell])\right]$$

and $\chi_{\{\ell\}}(\mathbf{x}) = \mathbf{x}[g_\ell]$ for $\ell \in [L]$.

The goal of cumulant-based approaches can be summarized as follows

- For a system $\mathbf{X}_t$ evolving under recombination, selection, mutation, sometimes genetic drift (see chapter VII.2 of [68]) and possibly other forces, find recursion equations for the cumulants $\mathrm{d}\chi_{\mathcal{I}}(\mathbf{X}_t)$.
- Simplify this system, typically by neglecting all cumulants $\chi_{\mathcal{I}}$ such that $\#\mathcal{I} > c_{max}$ for some fixed c_{max} (typically $c_{max} = 2$).
- Solve for $\mathrm{d}\chi_{\mathcal{I}}(\mathbf{X}_t)$ with $\#\mathcal{I} \leq c_{max}$.
- Control departures of the distribution of $Z(g)$ under $\mathbf{X}_t$ from a normal distribution.

This method, as presented in [99], was developed in a very general setting, which could allow for infinitely-many alleles at each locus (in which case, instead of the $(\chi_{\mathcal{I}})_{\mathcal{I}\subseteq[L]}$, we must keep track of the $(\chi_{\mathbf{i}})_{\mathbf{i}\in\mathbb{N}^{[L]}}$ where the $\mathbf{i}$ are multi-index, see Chapter V of [68]), and epistasis (see [86]). In our system of biallelic loci, it can be seen that for an LE population $\mathbf{x} \in \Gamma^{[L]}$ all cumulants are zero except $(\chi_\ell(\mathbf{x}))_{\ell\in[L]} = (\mathbf{x}[g_\ell])_{\ell\in[L]}$. In particular, cumulant-based approaches recover the LE-Wright-Fisher diffusion (1.7) when $c_{max} = 1$, but allow for some loose LD when $c_{max} > 1$. In a sense, when $\#\mathcal{I} > 1$, $\chi_{\mathcal{I}}(\mathbf{x})$ quantifies the amount of LD of the marginal of $\mathbf{x}$ on $\mathcal{I}$. This method seems appropriate as long as the population is "well-mixed", meaning the dynamics at locus ℓ are well approximated by its intrinsic parameters $(\alpha_\ell, \theta_\ell)$ rather than the correlation under $\mathbf{X}_t$ between g_ℓ and fit allele combinations at other loci (the so-called hitch-hiking effect [100]).

The most recent theoretical developments using these methods are to be found in [101, 70]. All cumulants of order greater than c_{max} are neglected, meaning the population can be represented as

$$X_t(\gamma) := \frac{1}{\mathscr{Z}(\phi(t))}\exp\left[\sum_{\substack{\mathcal{I}\subseteq[L]\\ \#\mathcal{I}\leq c_{max}}} \phi_{\mathcal{I}}(t)\prod_{\ell\in\mathcal{I}}\gamma_\ell\right]$$

where $\mathscr{Z}$ is the partition function and the $(\phi_{\mathcal{I}}(t))_{\mathcal{I}\subseteq[L],\#\mathcal{I}\leq c_{max}}$ characterize $\mathbf{X}_t$. Such a representation of a probability distribution on $\Box_{[L]}$ with negligible higher-order cumulants is called a Potts model. It is commonly used in statistical physics to model complex systems [102].

These models are typically deemed appropriate when the log-fitness function is of the form

$$W(\gamma) = \sum_{\ell\in[L]} f_\ell \gamma_\ell + \sum_{\ell_1,\ell_2\in[L]} f_{\ell_1\ell_2}\gamma_{\ell_1}\gamma_{\ell_2}$$

where the $(f_\ell)_{\ell\in[L]}$ and $(f_{\ell_1\ell_2})_{\ell_1,\ell_2\in[L]}$ are described in terms of their distributions. If the $f_{\ell_1\ell_2}$ are independent Gaussian variables, this is akin to the Sherrington-Kirkpatrick form from spin-glass theory [103].

Statistical physicists have characterized the phase transition to regimes in which the general premise that high-order cumulants can be neglected is no longer satisfied. They have in particular used spin-glass theory to characterize a phase transition to “clonal condensation” [104], in which some very fit genotypes dominate the population, and described a transition to “non-random coexistence” [70], in which a diffuse cloud of fit genotypes dominates the population.

The papers mentioned so far focus on macroscopic observables, typically the joint distribution of $(P_t^\ell)_{\ell\in[L]}$ and average covariance. But how does LD manifest itself if we only look at one locus ? Very recently, [105] (not yet peer-reviewed) applied this cumulant-based method to an additive trait under stabilizing selection, to determine a parameter $\hat{\alpha}_\ell$, which should be thought of as an effective additive effect at locus ℓ in the presence of low levels of LD (when the population is at the optimum). Their approach follows that of Bulmer in [106], which assumes a version of the infinitesimal model with LD[10].

From a mathematical point of view, what is rather surprising is that these methods account for LD between pairs of loci which are very far away, but neglect LD between triplets of loci which are very close, which seems at odds with the typical assumptions that recombination mixes more effectively over long distances than short distances. That is, these methods will account for the cumulants $\chi_{\{\ell_1,\ell_2\}}(\mathbf{x})$ even when ℓ_1,ℓ_2 are very far, but neglect cumulants of the form $\chi_{\{\ell'_1,\ell'_2,\ell'_3\}}$ when ℓ'_1,ℓ'_2,ℓ'_3 are very close. This is loosely justified with the idea that though the recombinator R is a more efficient mixing force on $\{\ell_1,\ell_2\}$ than on $\{\ell'_1,\ell'_2,\ell'_3\}$, the selector S has a stronger action on the former set than on the latter. In [107], the theoretical validity of the cumulant-based approach when $c_{max}=2$ is discussed when genetic drift is neglected: in particular it is shown that this approach is self-consistent provided the typical value of $\chi_{\{\ell_1,\ell_2\}}$ is much smaller than $1/L$. More details on this approximation are given in Appendix A.

Application to detecting epistasis

Direct Coupling Analysis was applied to the Potts model for pathogens with high recombination rates such as HIV, to learn from massive genomics data the fitness coefficients $f_\ell, f_{\ell_1\ell_2}$ (reviewed in [70]). The real data can be seen as a distribution $\mathbf{x}$ on the set possible virus genotypes $\Box_{[L]}$. The goal is then to infer $(f_\ell, f_{\ell_1,\ell_2})_{\ell,\ell_1,\ell_2}$ from the cumulants $(\chi_{\{\ell\}}, \chi_{\{\ell_1,\ell_2\}})_{\ell,\ell_1,\ell_2}$ in a computationally efficient way, which is precisely what Direct Coupling Analysis is about.

Application to the maintenance of genetic variation

One fundamental goal of polygenic models is to predict the value of the additive genetic variance V_A (see Section 1.3.4), which is routinely measured in wild populations (chapter 2

---

[10] This infinitesimal model with LD has not yet been obtained rigorously as a scaling limit of a finite-L model. It is equivalent to neglecting all cumulants of order greater than 2, and assuming that the population trait distribution remains Gaussian.

of [17]) and underlies the trait-based model from Lande & Arnold (in their nomenclature, the G matrix [108]). Predictions have been obtained using the cumulant-based approach , see Chapter 28 of [2] and Chapter VI.7 of [68]. The goal is to express how genetic variance can be maintained by the interaction of natural selection, mutation, LD (when possible) and genetic drift.

This is still an active area of research: for instance, [26] (not yet peer-reviewed) recently proposed a new prediction for the additive genetic variance V_A accounting for environmental fluctuations.

1.3.10 Contributions and challenges from Genome-Wide Association Studies

The advent of high-throughput sequencing and the massive datasets that were subsequently obtained led to the development of Genome-Wide Association Studies (GWAS). These studies aim at explaining phenotypic data using genomic data, or, to frame it in our notation, to infer the coefficients $(\alpha_\ell)_{\ell\in[L]}$ from a joint distribution of $(g, Z(g))$ sampled within a population, accounting for confounders such as population structure [109]. The typical output from such studies is a joint distribution $(\hat{P}^\ell, \hat{\alpha}_\ell)_{\ell\in[L]}$ where $\hat{P}^\ell$ is the inferred frequency of a variant allele at locus ℓ and $\hat{\alpha}_\ell$ is the inferred additive effect of the variant allele. This can be used to construct a **polygenic score**

$$\hat{Z} : g \longmapsto \sum_{\ell\in[L]} \hat{\alpha}_\ell g_\ell.$$

In the population $\mathbf{X}_t$, the quality of the polygenic score can be measured as the proportion of genetic variance explained by the polygenic score

$$1 - \frac{\mathbf{Var}_{\mathbf{X}_t}[\hat{Z}(g) - Z(g)]}{\mathbf{Var}_{\mathbf{X}_t}[Z(g)]}.$$

For the reference trait of human height, GWAS has now achieved a saturated map, meaning[11] the proportion of genetic variance explained is close to 1 [110].

GWAS confirm polygenic additive traits under stabilizing selection

The first contribution of GWAS was to confirm the highly polygenic structure of typical traits. For instance, the saturated map for human height in [110] requires 12,111 loci. More generally, [111] found on 30 human traits that a chromosome's contribution to the additive genetic variance V_A is roughly proportional to its length (their Figure 5). This suggests that loci which contribute to the trait are spread on every chromosome [85].

GWAS also vindicated the additive model from Section 1.3.3. Though population structure can lead to substantial bias in polygenic scores, in particular for behavioral traits [112], GWAS results for traits such as height are very robust, meaning the estimates $(\hat{\alpha}_\ell)$ are consistent across diverse genetic backgrounds [113]. To put it differently, the estimate of $\hat{\alpha}_\ell$ for a population of European ancestry is consistent with that of a population of a different origin. Dominance effects are found to be rare (though their detection requires more power): [114] found only 175 loci exhibiting dominance patterns when considering 1,000 human traits.

The distribution of $(\hat{P}^\ell, \hat{\alpha}_\ell)_{\ell\in[L]}$ is particularly amenable to evolutionary interpretations [115]. In particular, it is generally observed that variants of large effects (large value of $\hat{\alpha}_\ell$)

---

[11] To be more specific, in real studies such as [110], this is further complicated by the fact that the measurement of Z includes environmental noise (see Definition 1.1.3). In this setting, the environmental variance V_E is obtained from heritability studies, which rely on the infinitesimal model. Saturation is achieved when $\mathbb{V}\mathbf{ar}_t[\hat{Z}(g) - z]$ is close to the environmental variance V_E, where $\mathbb{P}_t$ is the law of a pair (g, z) with g a genome and z the phenotypic trait value of a randomly sampled organism at time t.

tend to segregate at very low frequencies ($\hat{P}^\ell$ is close to $\{0, 1\}$), which is interpreted as the result of stabilizing selection (that is, Robertson's underdominant term in (1.19)). Some of the landmark studies on the subject are [116, 117, 118, 119] and more recently [120] (not yet peer-reviewed). In particular, [120] found evidence for stabilizing selection on medical traits which one would have intuitively assumed to be under directional selection, including liability to type II diabetes, arthritis or schizophrenia (see their Figure 3).

Why biallelic loci ?

The models mentioned for the evolutionary interpretation of GWAS usually consider that each locus only has two alleles segregating at any given time. This assumption is typically satisfied for GWAS, because alleles are defined with Single Nucleotide Polymorphisms (SNP): that is, if two humans have DNA sequence AATA and TATA at the same position of their genomes, it will be considered that there are two alleles (A and T) segregating within the population at this position. Situations in which a third human has DNA sequence, say, CATA, segregating at this position will be extremely rare, and a biallelic model thus seems appropriate.

The reader should be warned of the limitations of this two-allele model when applied to GWAS data: because the population is typically not at HWLE, a GWAS analysis usually segments the genome into genomic fragments which are sufficiently small to only have a small amount of SNPs each, but sufficiently large that one may ignore LD between genomic segments (see for instance [120]). One reference SNP is picked in each genomic segment, and then the statistical processing considers that there are only two alleles at this genomic segment, given by the two versions of the SNP. For this reason, and also because many DNA variants are not SNPs (for instance insertions, deletions), the reference SNP is a binary summary of a complex genomic region, and "tags" all other causal variants within this region (see also SI section 6 of [116]).

The infinite-site model

Underlying the models mentioned is the infinite-site model, in which two mutations never occur on the same position. In terms of the individual-based model from Section 1.2.2, each organism has a chromosome seen as a continuous segment $[0, 1]$, on which mutations occur as a Poissonian process (see for instance [121]), and the additive effect of each mutation is independently sampled with a certain common law. This model is also obtained as the limit of our genetic model from Section 1.2.2 when $L \to +\infty$, and the mutation rate is small $|\theta_\ell| \to 0$ (for instance in [122]).

I am not aware of a rigorous mathematical representation of the infinite-site model from the gene's eye-view, modelling genetic drift and selection, but I believe this can be achieved using the excursion theory of Wright-Fisher diffusions developped in [76, 77].

1.3.11 Out-of-equilibrium dynamics under stabilizing selection

One recent class of work aims at describing the evolution of the population after a sudden change of optimum. That is, start from a stationary population evolving under stabilizing selection (1.18), and suddenly change the value of the selection optimum η (say, increase its value). If the change is very large, then the first term of the selection coefficient (1.19) becomes very large and many alleles which were segregating at intermediate frequencies will suddenly come under strong selection, and will go very quickly to fixation: such a phenomenon is called genetic sweep [80] and the corresponding regime is called the oligogenic regime [81]. In the LE-Wright-Fisher diffusion (1.7), for any locus ℓ such that α_ℓ is large, P_t^ℓ will quickly go to one. On the other hand, if the change in optimum is small, then the response to selection will be through small shifts of allele frequencies. That is, P_t^ℓ will only change a little to accommodate the new optimum. This regime is called the polygenic regime [81], and this matches the terminology I introduced in Section 1.3.2 in that the dynamics at a given locus will be stochastic.

This subject has been much studied [123, 124, 81, 125, 126] (former work neglecting genetic drift was done in [127, 128, 129, 130, 131], and a simulation study was done in [132]). In particular, [124] has described the two phases of adaptation: one rapid initial phase which is roughly equivalent to directional selection on each locus, and one slower phase when the population is close to the optimum. Considering a larger parameter space, [81, 125] has characterized the transition from the oligogenic to the polygenic regimes of adaptation in terms of background mutation rates and genetic architecture. To be more precise, [81] use the infinite-site model which we described in the Section 1.3.10: in this model, the number of segregating loci is a function of the mutation rate. The oligogenic regime is then characterized by a sufficiently small number of segregating loci, that adaptation results in selective sweeps at each segregating locus. This is as opposed to the polygenic regime where the number of segregating loci is sufficiently large, that adaptation of the trait only translates in small shifts of allele frequencies. Finally, [126] has uses path integrals (tools from statistical physics) to more efficiently compute the transition semigroup of P_t^ℓ at a given locus ℓ.

Possible applications are finding footprints of past selection events on genomic data, though population stratification make this challenging [133, 134]. Traces of selective sweeps are rare in humans, which suggests that adaptation in humans in the past thousands of years has mostly occurred through the polygenic regime [135].

1.3.12 Bridging the gap towards ecology: accounting for spatial structure and demography.

Integrating ecological and evolutionary dynamics is essential to make sense of field studies, because of the fact that polygenic adaptation can occur on the scale of a few generations, that is, the same timescale as ecological factors such as environmental fluctuations. While much work has been done to reconcile trait-based quantitative genetics models with ecological factors [17], there has also been efforts to reconcile polygenic models with ecology.

In [136], a model is studied under which a population evolves under directional selection, mutation, genetic drift and migration. In the individual model from section 1.2.2, migration should be seen as a fixed probability m each generation for an offspring to turn into the $(-1, \dots, -1)$ genome (with minimal fitness). [136] shows that, in the equation for P_t^ℓ, this migration translates as an effective input m_e in trait-decreasing alleles. In this sense, the effect of m_e is the same as that of the negative mutation rate θ^-. The system is then equivalent to a system with no immigration, with mutation rate at locus ℓ given by $(\theta_\ell^+, \theta_\ell^- + m_e)$. The value of m_e is derived accounting for LD but neglecting genetic drift (that is, m_e is derived using the discrete-time equivalent of the LD-Wright-Fisher diffusion (1.5) with free recombination (see Section 1.2.1) under the assumption of large ρ, large m and large W, thus neglecting the Brownian term).

In [137], semi-deterministic approximations are developed to model a spatially structured population with fluctuating population size. Specifically, the population size N_t evolves as a stochastic logistic process, with growth rate proportional to mean fitness $\mathbf{X}_t[W(g)]$, and the genetic drift in (1.7) is inversely proportional to $\sqrt{N_t}$. In [137], an approximation is obtained for the joint stationary distribution of $(\mathbf{X}_t, N_t)$ by neglecting the fluctuations of N_t.

In [138, 139], the LE-Wright-Fisher diffusion (1.7) is studied in the polygenic limit ($L \gg 1$) under stabilizing selection, in a continuous spatial setting. Specifically

- $\mathbf{X}_t$ is a function of space, that is, a continuous function from $\mathbb{R}$ to $\mathbb{X}^{[L]}$.
- in (1.7) a Laplacian $\Delta \mathbf{X}_t$ accounts for migration

- Fitness is local: specifically, the optimum η in (1.19) is a linear function of space.

This model also accounts for varying population size (that is, the Brownian term in (1.7) depends on a locally fluctuating parameter determined by the local mean fitness $\mathbf{X}_t[W(g)]$). The appearance of range limits is described, that is, even though space is infinite, the population is only maintained within a finite range. In [140] (not yet peer-reviewed), this theoretical framework is applied to predict how climate change can impact a species' range and clarify what parameters will influence its resilience.

1.4 A probabilist's roadmap to the polygenic limit.

The roadmap to the polygenic limit is summarized in Figure 1.3. We start from the individual-based model from Section 1.2.2. There are three limit objects of interest: we want the distribution of the trait $Z(g)$ to be approximately normal, we want the mean trait value $\bar{z}_t$ to follow an Ornstein-Uhlenbeck process and we want a description of the behavior of the typical locus. Ideally, one would want to obtain the limit directly from the individual-based model, simultaneously letting N and L go to infinity, with an appropriate scaling of the log-fitness function W, the mutation rate μ_ℓ, and the recombination measure ν. This is at present too difficult. An alternative approach to the first two limit objects can be obtained from the trait's eye-view, by scaling $L \to +\infty$ to obtain the infinitesimal model (Section 1.3.5), and then scaling $N \to +\infty$ (Section 1.1.2, though this has not yet been proved rigorously). But such an approach precludes the modelling of a typical locus, because we lose all information on the underlying gene frequencies, which end up summarized on the macroscopic parameters such as mutational variance V_m, and the mean segregation variance V_S^*.

The goal of this work is to map the way to the three limit objects from the gene's eye-view, by letting $N \to +\infty$ to obtain the Wright-Fisher diffusion and then $L \to +\infty$. The gene's eye-view is the scholarly term to refer to the philosophical stance of Fisher and his followers, who believed that an appropriate description of a biological population could be achieved by only focusing on allele frequencies $(P_t^\ell)_{\ell\in[L]}$ and neglecting epistasis and population structure, or to put it another way, that the gene is the "true" unit of selection [141].

Two distinct approaches will be presented. In Chapter 2, we will start from the LD-Wright-Fisher diffusion (1.5) and let L go to infinity, while simultaneously scaling the recombination rate ρ, the log-fitness function W and the recombination measure ν. We will rigorously prove that a polygenic limit can be obtained this way, in which the selection coefficient at a locus s_ℓ given by (1.19) is replaced by a mean-field coefficient s^*. The reason for starting from the LD-Wright-Fisher diffusion is that we want to obtain the polygenic limit in as little steps as possible. In particular, we want to know how ρ must scale with L, that is, how strong recombination needs to be with respect to selection so that we may neglect LD and derive the polygenic limit. In Theorem 2.1.1, we will see that for most recombination measures, it is sufficient for ρ to be much greater than $L^2\ln(L)^2$ when the log-fitness function W is of order L (which, as will be discussed in Chapter 4, corresponds to weak selection). This criterion, though quite laughable from a biologist's perspective, presents several interesting features. In Appendix A.2, we show ρ must be greater than 2^L in order for the whole population $\mathbf{X}_t$ to be close to LE, that is, for the LD-Wright-Fisher diffusion (1.5) to be close to the LE-Wright-Fisher diffusion (1.7). The criterion of Theorem 2.1.1 shows that this is not needed to reach the polygenic limit. To put it another way, we know that $\mathbf{X}_t$ will not be globally at LE when ρ is a power of L, and yet we still derive the polygenic limit. Furthermore, our derivation is rigorous: to the best of my knowledge, this is the first proof in a setting with recombination, selection, and genetic drift (a noteworthy derivation in the case of recombination and general selection is given in [121]).

Our approach justifies the choice of the individual-based model in Section 1.2.2, in which we allowed organisms to reproduce clonally a fraction ρ/N of the time. This clonal reproduction is inserted in our model in order to guarantee that we can scale $N \to +\infty$ and obtain a well-defined object (the LD-Wright-Fisher diffusion (1.5)) which still retains LD. This object being a continuous diffusion, it is much more convenient to work with, which lets us derive our result.

Chapter 3 is a short addition to Chapter 2 in which we obtain an ancestral process when there is no genetic drift ($\Sigma = 0$), which we argue can be used to build intuition on the first-order perturbations to LE due to selection.

In Chapter 4, we will completely ignore the difficulties of LD, starting instead from the LE-Wright-Fisher diffusion (1.7). This chapter corresponds to an article soon to be submitted, which is aimed at a biological audience: in particular, we drop mathematical rigor and most results there presented are not properly derived. For a mathematician, this chapter should be seen as a research project: we present and classify all the arguments needed to get from the LE-Wright-Fisher diffusion to the polygenic limit, as illustrated in Figure 1.3, including a mean-field approximation, a separation of timescales between the evolution of the trait mean $(\bar{z}_t)_{t\geq 0}$ and the evolution of a given locus $(P_t^\ell)_{t\geq 0}$, and a central limit theorem. We also discuss how this behavior changes when instead of an additive trait Z, we take a general trait showing dominance or epistasis.

The notation used will be broadly consistent with that introduced in the past sections, the only notable difference being that we will focus in Chapters 2-3 exclusively on haploids, whereas Chapter 4 will be concerned with diploids. Furthermore, in Chapter 2 the log-fitness of an organism with trait z is expressed as $e^{\frac{L}{N}W(z)}$ whereas in (1.5) it is $e^{\frac{W(z)}{N}}$. The notation will be reintroduced in each chapter to make reading more convenient.

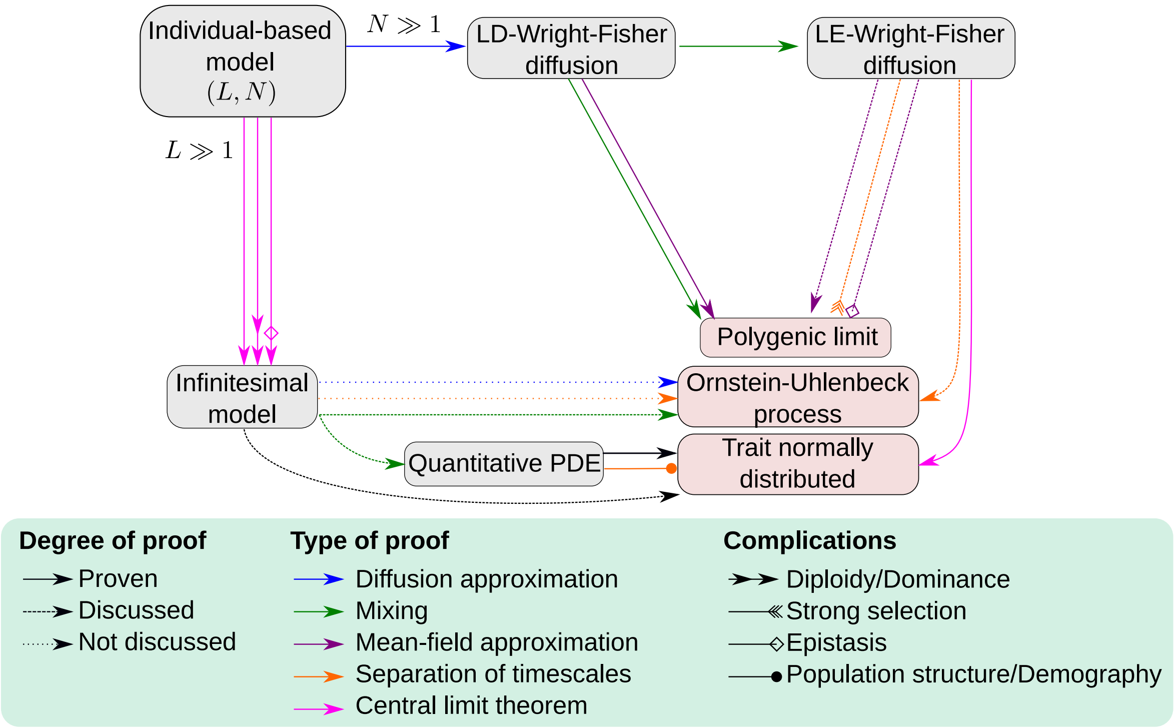

Figure 1.3: Roadmap to the polygenic limit. "Mixing" designates the situation where a strong force maintains the population on a submanifold of low dimensionality, such as strong recombination maintaining the population in LE or the shuffling of biparental pedigrees. "Mean-field approximation" designates approximations where we consider that the mean behavior of a large system of interacting particles is close to deterministic. "Separation of timescales" designates approximations where the fluctuations of a system are ignored when computing the evolution of a much slower system. Chapter 2 will be concerned in getting from the LD-Wright-Fisher diffusion to the polygenic limit, Chapter 4 will classify the required steps to get from the LE-Wright-Fisher diffusion to the polygenic limit and Ornstein-Uhlenbeck process for the trait mean $\bar{z}_t$ as in (1.3), including when facing specific complications due to dominance, epistasis, or very strong selection. The scaling from finite (L, N) to $L \gg 1$ has been described in [11, 90]. The scaling from the individual-based model to the LD-Wright-Fisher diffusion has been discussed for instance in [31]. The scaling from the LD-Wright-Fisher diffusion to the LE-Wright-Fisher diffusion can be proved using the same kind of arguments as in Chapter 2. The scaling from the infinitesimal model to PDEs for the trait distribution using the infinitesimal operator from Section 1.1.2 with segregation of the pedigrees is still lacking (though see [32, 33, 35, 34]). The scaling from PDEs on the trait distribution to a population with a normally distributed trait was discussed in Section 1.1.3. Quantifying the deviation from normality of a finite population is explored for instance in [45].

Chapter 2

The gene's eye-view of quantitative genetics

Contents

2.1 Introduction

Here, we consider the LD-Wright-Fisher diffusion (1.5), and we let the number of loci go to infinity under the assumption of strong recombination. We characterize the limit behavior of a given locus with a McKean-Vlasov SDE and the corresponding Fokker-Planck IPDE. In words, the selection on a typical locus depends on the mean behaviour of the other loci which can be approximated with the law of the focal locus. Results include the independence of two loci and explicit stationary distribution for allelic frequencies at a given locus (under some assumptions on the fitness function).

2.1.1 Definition of the model.

To motivate the definition of our model let us first consider a single locus/$\{-1,+1\}$ alleles. When the population N is large, under the joint action of mutation, selection and genetic drift, the evolution of the frequency of $+1$-allele is well approximated by the Wright-Fisher diffusion

$$\mathrm{d}X_t = s(X_t)X_t(1-X_t)\mathrm{d}t + \overline{\Theta}(X_t)\mathrm{d}t + \sqrt{X_t(1-X_t)}\mathrm{d}B_t \tag{2.1}$$

where s is a frequency dependent selection term, B is a Browian motion and

$$\overline{\Theta}(x) \ := \ \theta^{+}(1-x) - \theta^{-}x$$

with θ^{+}, θ^{-} the rates of mutations from the -1 allele to the $+1$ allele and back.

The aim of the present article is to consider a L-loci/$\{-1,+1\}$ allele model. In the following, we will consider a diffusion analog to the classical Wright-Fisher diffusion (2.1), but in order to provide more intuition on our continuum model, we first consider a discrete population in discrete time comprised of N individuals, each with L genes encoded by an element of the hypercube $\square_{[L]} := \{-1,+1\}^L$.

The evolution of the population results from the combined effect of Selection (S), Recombination (R), Mutation (M) and random sampling. Those fundamental evolutionary forces are encoded by the following parameters.

(S) Let $W \ : \ \square_{[L]} \to \mathbb{R}$.

(R) Let $\rho > 0$ and ν be a probability measure on $\{\mathcal{I} \subset [L] : \mathcal{I} \neq \emptyset, [L]\}$.

(M) Let $\theta^{+}, \theta^{-} \geq 0$.

Reproduction then occurs according to Wright-Fisher sampling. Start with N genomes at time $k = 0$, meaning a vector $(g^1, \ldots, g^N) \in (\square_{[L]})^N$. At every generation $k > 1$, each of the N genomes comprising the new generation independently picks two parent genomes γ^1, γ^2 with probability proportional to their fitnesses

$$\exp\left(\frac{L}{N}W(\gamma^1)\right), \ \exp\left(\frac{L}{N}W(\gamma^2)\right) \tag{2.2}$$

The function W is often referred to as the log-fitness function. As we will see, the $\frac{L}{N}$ factor is important so that the strength of selection felt by one locus is of order 1.

With probability $1 - \rho/N$, the new genome g^0 inherits the whole genome of one of its two parents g^1 and g^2 chosen uniformly at random. With probability ρ/N, a recombination occurs and g^0 inherits a mixture of the genetic material of the two parents according to ν. More precisely, a subset $\mathcal{I} \subset [L]$ is sampled according to ν and g^0 inherits the genetic material of g^1 at loci $\mathcal{I}$, and the material of g^2 at loci $\mathcal{I}^c$.

Finally, we assume that each generation, each locus on each allele can mutate with probability θ^{+}/N (resp., θ_2/N) from -1 to $+1$ (resp. $+1$ to -1) after reproduction.

Let $k \in \mathbb{N}$, and define $\mathbf{X}^{(N)}_k = (X^{(N)}_k(\gamma))_{\gamma \in \square_{[L]}}$ where $X^{(N)}_k(\gamma)$ is the frequency of genotype γ at generation k. The process $(\mathbf{X}^{(N)}_k)_{k \in \mathbb{N}}$ is valued in the space of probability distributions on the hypercube $\square_{[L]}$ that we denote by $\mathbb{X}^{[L]}$. Let us now consider the large population limit $(N \to +\infty)$ of the rescaled process $(\mathbf{X}^{(N)}_{\lfloor tN \rfloor})_{t \geq 0}$. A straightforward generator computation indicates that the process converges to a diffusive limit $(\mathbf{X}_t)_{t \geq 0}$ valued in $\mathbb{X}^{[L]}$ which is solution to a Stochastic Differential Equation (SDE) on $\mathbb{X}^{[L]}$ of the form

$$\mathrm{d}\mathbf{X}_t = (\rho R(\mathbf{X}_t) + \Theta(\mathbf{X}_t) + LS(\mathbf{X}_t))\mathrm{d}t + \Sigma(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t \tag{2.3}$$

that we now explain. This equation will be the focus of this article.

Recombination. For a subset $\mathcal{I} \subseteq [L]$ and $\mathbf{x} \in \mathbb{X}_L$, define $\mathbf{x}^{\mathcal{I}}$ the marginal of $\mathbf{x}$ on the hypercube $\square_{\mathcal{I}} := \{-1,+1\}^{\mathcal{I}}$. Let $\mathbf{x}^{\mathcal{I}} \otimes \mathbf{x}^{\mathcal{I}^c}$ be the product measure on $\square_{[L]}$ of $\mathbf{x}^{\mathcal{I}}$ and $\mathbf{x}^{\mathcal{I}^c}$. The recombinator operator has been extensively studied in the deterministic setting, see e.g., [142, 58, 60, 61, 78] and is defined as

$$R : \begin{cases} \mathbb{X}^{[L]} & \to \mathbb{R}^{\square_{[L]}} \\ \mathbf{x} & \mapsto \sum\limits_{\emptyset \subsetneq \mathcal{I} \subsetneq [L]} \nu(\mathcal{I})(\mathbf{x}^{\mathcal{I}} \otimes \mathbf{x}^{\mathcal{I}^c} - \mathbf{x}) \end{cases} \tag{2.4}$$

Note that up to replacing ν with $\tilde{\nu} : \mathcal{I} \mapsto \frac{\nu(\mathcal{I})+\nu(\mathcal{I}^c)}{2}$, we can and will assume that for any $\mathcal{I} \subseteq [L]$, $\nu(\mathcal{I}) = \nu(\mathcal{I}^c)$.

Mutation. The mutator is defined as

$$\Theta : \begin{cases} \mathbb{X}^{[L]} & \longrightarrow \mathbb{R}^{\square_{[L]}} \\ \mathbf{x} & \longmapsto |\theta| \sum\limits_{\ell \in [L]} \left(\mathbf{x}^{[L] \setminus \{\ell\}} \otimes \mathcal{L}_\theta - \mathbf{x} \right) \end{cases}$$

where $|\theta|$ is the total mutation rate per locus and $\mathcal{L}_\theta$ is the mutational law defined with

$$|\theta| := \theta^+ + \theta^- \qquad\qquad \mathcal{L}_\theta := \frac{\theta^-}{|\theta|}\delta_{-1} + \frac{\theta^+}{|\theta|}\delta_{+1}.$$

Selection. The operator $S : \square_{[L]} \to \mathbb{R}^{\square_{[L]}}$ is the selector defined with

$$S(\mathbf{x})(\gamma) := x(\gamma)(W(\gamma) - \mathbf{x}[W(g)]) = \mathbf{Cov}_{\mathbf{x}} \left[W(g), \mathbb{1}_{[g=\gamma]} \right]$$

where $\mathbf{x}[\cdot]$ and $\mathbf{Cov}_{\mathbf{x}}[\cdot,\cdot]$ are the expectation and the covariance function for a random genotype g with law $\mathbf{x}$. This can be thought of as an application of the Price equation [143] to the trait $F^\gamma(g) := \mathbb{1}_{[g=\gamma]}$. See for instance [144]. The factor L in front of S in (2.3) corresponds to the strength of selection, which stems from (2.2) and will be discussed in the next subsection.

Genetic Drift. The stochastic term is the traditional multiallele Wright-Fisher diffusion term [62] and with xthe following covariance structure.

$$\left\langle (\Sigma(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t)(\gamma), (\Sigma(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t)(\gamma') \right\rangle = \delta_{\gamma,\gamma'} X_t(\gamma) - X_t(\gamma)X_t(\gamma').$$

More precisely, we will consider $\mathbf{B} \equiv (B_t(\gamma^1,\gamma^2))_{t\in[0,T];\gamma^1,\gamma^2\in\square_{[L]}}$ a Gaussian process indexed by $\square_{[L]} \times \square_{[L]}$ such that $B(\gamma^1,\gamma^2) = -B(\gamma^2,\gamma^1)$, and $B(\gamma^1,\gamma^2), B(\gamma^3,\gamma^4)$ are independent Brownian motions if $(\gamma^1,\gamma^2) \notin \{(\gamma^3,\gamma^4),(\gamma^4,\gamma^3)\}$.

Finally, let $\mathcal{M}(\square_{[L]} \times \square_{[L]}, \mathbb{R}^{\square_{[L]}})$ denote the space of linear functions from $\square_{[L]} \times \square_{[L]}$ to $\mathbb{R}^{\square_{[L]}}$. Then

$$\Sigma : \mathbb{X}^{[L]} \to \mathcal{M}\left(\square_{[L]} \times \square_{[L]}, \mathbb{R}^{\square_{[L]}}\right)$$

is defined such that

$$\forall \gamma \in \square_{[L]}, \quad (\Sigma(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t)(\gamma) := \sum_{\hat{\gamma}\neq\gamma} \sqrt{X_t(\gamma)X_t(\hat{\gamma})}\mathrm{d}B_t(\gamma,\hat{\gamma}) \tag{2.5}$$

Remark 6. *Existence and uniqueness of solutions to (2.3) was obtained from the martingale problem in [63].*

2.1.2 Propagation of chaos.

For any $\ell \in [L]$, define

$$p^\ell(\mathbf{X}_t) := \sum_{\gamma \in \square_{[L]}} \mathbb{1}_{[\gamma_\ell = +1]} X_t(\gamma)$$

as the frequency of $+1$ allele at locus ℓ at time t. We are interested in describing the joint evolution of the $p^\ell(X_t)$'s together with the allelic averaged process $\mu_{\mathbf{X}_t}$, where for $\mathbf{x} \in \mathbb{X}^{[L]}$

$$\mu_{\mathbf{x}} := \frac{1}{L} \sum_{\ell \in [L]} \delta_{p^\ell(\mathbf{x})}$$

When focusing on just a few loci, the high dimensionality of the SDE (2.1) renders the problem practically intractable. However, as $L \to +\infty$ and recombination becomes sufficiently strong to neglect correlations between loci, a mean field approximation applies where any focal locus only experiences the averaged effect of the rest of the genome, a phenomenon commonly known as the propagation of chaos. See Fig 2.1.

Let us now describe our results in more details. We will consider a sequence of models indexed by L where the dependence in L is in

$$(\rho, \nu, W, \mathbf{X}_0) \equiv (\rho^L, \nu^L, W^L, \mathbf{X}_0^L),$$

whereas other parameters remain constant. Our theorem is concerned with the limit of (2.3) as $L \to +\infty$. We make several assumptions on the order of the parameters.

The most restrictive assumption is that the log-fitness W is a quadratic polynomial with bounded coefficients. More precisely,

$$\forall \gamma \in \square_{[L]}, \quad W(\gamma) := U(Z(\gamma)) \qquad ; \qquad Z(\gamma) := \frac{1}{L} \sum_{\ell \in [L]} \gamma_\ell \tag{2.6}$$

where $Z(\gamma)$ is the additive trait value associated with genotype γ and U is a polynomial of order 1 or 2. In particular, if U is of order 2 it can be written up to an additive constant

$$U(z) = -\kappa(z - z^*)^2 \tag{2.7}$$

for some $\kappa, z^* \in \mathbb{R}$. This assumption on the functional form for the log-fitness is classic in quantitative genetics. Take $z^* \in [-1, 1]$. The case $\kappa > 0$ (resp., $\kappa < 0$) corresponds to a scenario of stabilizing (resp., disruptive selection) on a quantitative additive trait [30, 116]. The assumption that the fitness within the population is mostly determined by stabilizing selection on a (multi-dimensional) highly polygenic trait is known as Fisher's geometric fitness model and has had countless applications [37].

Let us briefly discuss the scaling. With the functional form of the log-fitness, selection acts on the additive trait Z. Under sufficiently strong recombination, it is reasonable to assume that selection impacts all loci similarly. Due to the additive nature of the trait, this influence is evenly distributed across all loci. Therefore, if we start with a selection of order L at the trait level as in (2.3), we expect selection to have an effect of order 1 at the locus level. In particular, the rescaling in (2.2) ensures that the dynamics at the gene level remain well-defined and non-degenerate.

We will need to assume that recombination is strong as compared to selection. In order to quantify this relation, we first define some summary statistics related to the recombination measure ν. For $A \subseteq [L]$, let ν^A be the marginal of ν on A. Define the recombination rate between two distinct loci $\ell_1, \ell_2 \in [L]$

$$r_{\{\ell_1,\ell_2\}} := \nu^{\{\ell_1,\ell_2\}}(\{\ell_1\}) + \nu^{\{\ell_1,\ell_2\}}(\{\ell_2\})$$

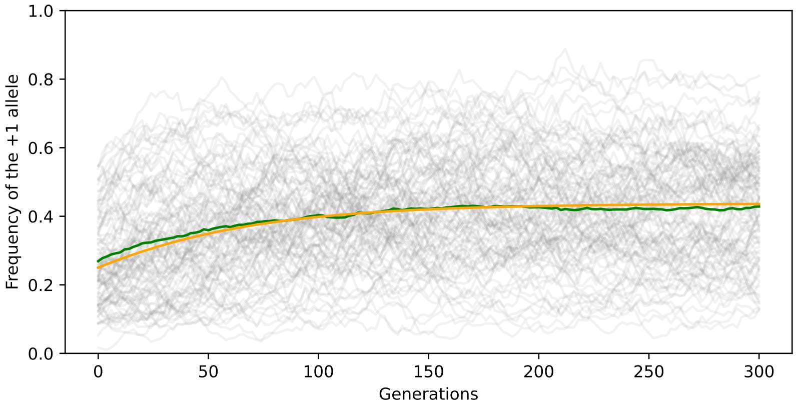

Figure 2.1: We consider the discete population of $N = 1000$ individuals with $L = 100$ genes each for $T = 1000$ generations, simulated as detailed in Section 2.1.1 with single uniform crossing-over (see below). The mutation rates are $\theta = (1.1, 3.3)$, the strength of stabilizing in (2.7) is $\kappa = 15$ and $z^* = 0$. At time $t = 0$, the population is distributed according to the neutral discrete Wright-Fisher equilibrium with mutation rate θ. The grey lines show the trajectories of the frequency of the +1 allele at each individual locus. The green line is the average of the grey lines. The orange line corresponds to the mean-field approximation (2.19), computed with an Euler approximation scheme. The code is available on `https://github.com/PhCourau/Gene-s_eye_view_of_quantitative_genetics`.

We assume that ν is non-degenerate, that is, $\inf_{\ell_1 \neq \ell_2} r_{\{\ell_1,\ell_2\}} > 0$. For a given locus $\ell_0 \in [L]$, define $r^*_{\ell_0}$ the harmonic recombination at ℓ_0 as

$$\frac{1}{r^*_{\ell_0}} := \frac{1}{L-1} \sum_{\ell_1 \in [L] \setminus \{\ell_0\}} \frac{1}{r_{\{\ell_0,\ell_1\}}} \tag{2.8}$$

and the average harmonic recombination rate along the genome as

$$\frac{1}{r^{**}} := \frac{1}{L} \sum_{\ell_0 \in [L]} \frac{1}{r^*_{\ell_0}} \tag{2.9}$$

The next theorem states that under strong enough recombination (see conditions (2.12-2.14)), a mean-field approximation applies and the $p^\ell(X_t)$'s converge to independent McKean-Vlasov diffusions [69]

$$\begin{aligned} \mathrm{d}f_t \; &= \; \bar{s}(\mathscr{L}(f_t))\; f_t(1-f_t)\mathrm{d}t + \overline{\Theta}(f_t)\mathrm{d}t + \sqrt{f_t(1-f_t)}\;\mathrm{d}B_t \\ \text{where} \;\; \bar{s}(\zeta) \; &= 2U'(<\zeta, 2\mathrm{Id}-1>) \qquad ; \qquad \overline{\Theta}(x) := \theta^+(1-x) - \theta^- x \end{aligned} \tag{2.10}$$

and $\mathscr{L}(f_t)$ denotes the law of the process f_t. This shows that in polygenic adapation, selection at the genome level "percolates" at the locus level where it dictates a non-linear Wright-Fisher dynamics. Further, the averaged process $\mu_{\mathbf{X}_t}$ converges to $\mathscr{L}(f_t)$ which in turn can be calculated as the weak solution of the non-linear IPDE (integro partial differential equation) corresponding to the Fokker-Planck equation associated to (2.10)

$$\partial_t u_t(x) = -\partial_x \left[\left(\bar{s}\left(u_t(\cdot)\right) x(1-x) + \overline{\Theta}(x) \right) u_t(x) \right] + \frac{1}{2} \partial_{xx} \left(x(1-x) u_t(x) \right). \tag{2.11}$$

We now state our main result on the convergence to (2.11). We let $\mathcal{P}([0,1])$ be the set of probability measures on $[0,1]$ equipped with the weak topology.

Theorem 2.1.1. *Assume that* $\mu_{\mathbf{X}_0}$ *converges in law to a deterministic measure* m_0*, and that*

$$\rho r^{**} \gg L^2 \ln(\rho) \tag{2.12}$$

Then

1. *For every* $T > 0$

   $$(\mu_{\mathbf{X}_t})_{t\in[0,T]} \quad \Longrightarrow \quad (\mathscr{L}(f_t))_{t\in[0,T]}$$

   where $\Longrightarrow$ *denotes weak convergence in* $\mathbb{D}([0,T],\mathcal{P}([0,1]))$ *for the Skorokhod J1 topology (see [57], Chapter 3), and* f_t *is the unique solution of the McKean-Vlasov equation (2.10) with initial distribution* m_0*. In particular,* $(\mathscr{L}(f_t))_{t\in[0,T]}$ *is the unique weak solution to (2.11) (see Section 2.2 for details).*

2. *Let* $n \in \mathbb{N}$*. Assume there exists a sequence integers* $\ell_1^L < \cdots < \ell_n^L$ *in* $[L]$*, such that* $(p^{\ell_i^L}(\mathbf{X}_0))_{i\in[n]}$ *has law converging to* $\mathfrak{P}_0 \in \mathcal{P}([0,1]^n)$ *and that*

   $$\min_{i\in[n]} \rho r^*_{\ell_i^L} \gg L^2 \ln(\rho) \tag{2.13}$$

   $$\min_{\substack{i,j\in[n]\\ i\neq j}} \rho r_{\{\ell_i^L,\ell_j^L\}} \gg L \tag{2.14}$$

   Then for every $T > 0$

   - $(p^{\ell_i^L}(\mathbf{X}_t))_{t\in[0,T];i\in[n]}$ *converges in distribution to n diffusions* $(\overline{p}_t^i)_{t\in[0,T];i\in[n]}$ *solutions to*

     $$\mathrm{d}\overline{p}_t^i = \overline{s}(\mathscr{L}(f_t))\overline{p}_t^i(1-\overline{p}_t^i)\mathrm{d}t + \overline{\Theta}(\overline{p}_t^i)\mathrm{d}t + \sqrt{\overline{p}_t^i(1-\overline{p}_t^i)}\,\mathrm{d}B_t^i \tag{2.15}$$

     with $(B^i)_{i\in[n]}$ *independent Brownian motions and initial conditions* $\mathscr{L}((\overline{p}_0^i)_{i\in[n]}) = \mathfrak{P}_0$*.*
   - *If* $\mathfrak{P}_0 = m_0^{\otimes n}$ *then* $(p^{\ell_i^L}(\mathbf{X}_t))_{t\in[0,T];i\in[n]}$ *converges in distribution to n independent McKean-Vlasov diffusions (2.10).*

Let us briefly discuss the previous assumptions. First, the theorem relies on the parameter r^{**}, the importance of which was already noted by Bulmer [106]. Secondly, the strong recombination conditions (2.12–2.14) are satisfied provided that recombination grows significantly with L.

The intuition is as follows: recombination needs to be strong enough to sufficiently break correlations between loci, allowing a mean-field approximation to be valid. The technical challenge arises because selection tends to induce correlations along the genome. For instance, if the optimum z^* is at 0 and one knows that $\gamma_\ell = +1$ at a given locus, selection will tend to favor -1 at other loci to compensate, keeping the trait near the optimum. In this way, selection introduces negative correlations across the genome.

Since the strength of selection scales with L, it becomes necessary to assume a recombination rate that also scales sufficiently-specifically, one that is large in L-to counteract these correlations.

To provide more intuition on the strong recombination conditions (2.12-2.14) of the previous theorem, we apply it to the classical recombination models.

- The model known to population geneticists as **free recombination** corresponds to the case where $\nu(\mathcal{I}) = \frac{1}{2^L}$ for any $\mathcal{I} \subseteq [L]$. In this case we can check that

  $$\forall \ell_0 \in [L], \quad r^*_{\ell_0} = r^{**} = \frac{1}{2}$$

  so that conditions the strong recombination conditions (2.12-2.14) are satisfied provided that $\rho \gg L^2 \ln(L)$.

- **Single crossing-over.** Let μ be a probability measure on $[0,1]$ with stricty positive continuous density. Then ν is the law of the random set
$$\mathcal{J} = \left\{ i \in [L], \frac{i}{L+1} \leq X \right\}, \quad \text{where } \mathscr{L}(X) = \mu.$$
The strong recombination conditions (2.12-2.14) are satisfied provided $\rho \gg L^2 \ln(L)^2$.

- **Multiple crossing-overs.** Consider a Poisson Point Process with an intensity measure with a strictly positive continuous density, seen as a random set of points $\lambda_1 < \cdots < \lambda_N$. We add the boundary points $\lambda_0 := 0$ and $\lambda_{N+1} := 1$. Then ν is the law of the random set
$$\mathcal{J} = \left\{ i \in [L] : \exists k \leq \frac{N+1}{2} \quad \text{s.t.} \quad \frac{i}{L+1} \in [\lambda_{2k}, \lambda_{2k+1}) \right\}$$
The strong recombination conditions (2.12-2.14) are satisfied provided $\rho \gg L^2 \ln(L)^2$.

2.1.3 Invariant distribution(s).

Assume that $\theta^+, \theta^- > 0$. For any $y \in \mathbb{R}$, define
$$\Pi_y(x) = C_y x^{2\theta^+ - 1}(1-x)^{2\theta^- - 1} e^{2xy} \mathrm{d}x \tag{2.16}$$
with C_y a normalization constant so that Π_y is a probability on $[0,1]$. It is well known that the classical Wright-Fisher SDE (2.1) with a constant selection term s has a unique invariant distribution Π_s. From there, the Lipschitzness of $\overline{s}$ trivially implies

Theorem 2.1.2. *Define*
$$\chi : y \longmapsto \overline{s}(\Pi_y) \tag{2.17}$$
where $\bar{s}$ is the function defined in the McKean-Vlasov equation (2.10). Then χ admits at least one fixed point. Futher,

- *The set of invariant distribution for (2.10) coincides with*
$$\{\Pi_{y^*} \ : \ \chi(y^*) = y^*\}$$

- *Assume that the initial condition of (2.10) is given by Π_{y^*} with $\chi(y^*) = y^*$. Then $(f_t)_{t \geq 0}$ is distributed as a classical Wright-Fisher diffusion (2.1), with a constant selection term $s \equiv \overline{s}(\Pi_{y^*})$ and initial distribution Π_{y^*}.*

We emphasize that there may exist several solutions to the fixed point problem of Theorem 2.1.2. For U as in (2.7) with $z^* = 0$ and $\theta^+ = \theta^-$, we show in Corollary 2.2.4 the existence of a critical $\kappa_c < 0$ at which the system undergoes a pitchfork-like bifurcation. See Fig 2.2.

2.1.4 Trait Variance.

A biologically important quantity is the variance of the trait $\mathbf{Var}_{\mathbf{X}_t}[Z(g)]$. Our mean-field approximation in Theorem 2.1.1(1) entails that the variance goes to 0. The next result provides a second order approximation.

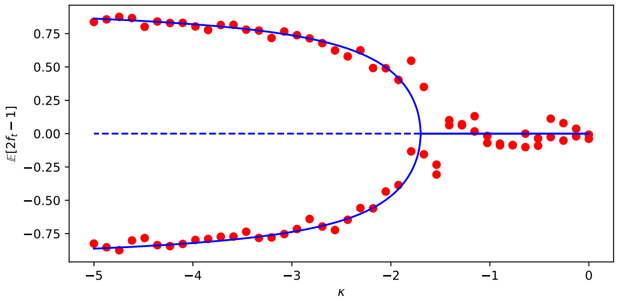

Figure 2.2: Supercritical pitchfork bifurcation in disruptive selection (2.7). For each value of κ on the x-axis, we simulated two discrete population (as detailed in Section 2.1.1) with $N = 200, L = 100$ after $T = 20N$ generations, with initial conditions "all +1" or "all −1". The red dots correspond to $< \mu_{\mathbf{X}_T}, 2\mathrm{Id} - 1 >$ at the end of the simulation. The mutation rates are fixed $\theta^+ = \theta^- = 0.6$, and the selection optimum is $z^* = 0$. The blue lines correspond to the possible values of $\mathbb{E}\left[2f_t - 1\right]$ for stationary solutions to the limit equation (2.10). Corollary 2.2.4 predicts a pitchfork-like bifurcation at $\kappa_c = -1.7$.

Theorem 2.1.3. *Assume that the assumptions of Theorem 2.1.1 part 1. hold. Set*

$$\varepsilon_L := \frac{1}{\sqrt{\rho r^{**}}} \tag{2.18}$$

Define the genetic variance $\sigma_t^2 := 4\mathbb{E}\left[f_t(1 - f_t)\right]$*. Then*

$$\mathbb{E}\left[\sup_{t \in [\varepsilon_L, T]} \left| L\mathbf{Var}_{\mathbf{X}_t}[Z(g)] - \sigma_t^2 \right| \right] \longrightarrow 0$$

Equation (2.11) has a biologically important corollary. Assume for simplicity there is no mutation $\theta^+ = \theta^- = 0$. The mean of the trait distribution satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[2f_t - 1\right] = U'(\mathbb{E}\left[2f_t - 1\right]) \times \sigma_t^2 \tag{2.19}$$

This is known as Lande's equation [124]. The term $U'(\mathbb{E}\left[2f_t - 1\right])$ is called the selection gradient.

2.1.5 Significance and extensions

Population genetics and quantitative genetics have mostly been developed in parallel throughout the twentieth century. Interpreting modern GWAS results requires understanding how the two interact, that is, how selection on a polygenic trait translates on the dynamics of allele frequencies [83]. The historical approach to this problem, which we call the trait's eye-view, can be traced back to the works of Latter [97] and Bulmer [98] and has been applied to GWAS in [116]. It consists in making assumptions on the distribution of the trait (typically a normality assumption), and conditional on the trait to model the evolution of the genes. Here we take the gene's eye-view [141], meaning we start from a finite model of coupled equations representing the evolution of gene frequencies, and then let the number of loci go to infinity. In this setting, the marginal fitness of an allele depends not just on intrinsic properties of this allele but on the mean behavior of the other alleles within the population. Theorem 2.1.1 describes this limit with equation (2.10), which is a Wright-Fisher equation typical of population genetics. The behavior of the trait then emerges from the distribution of allele frequencies. This is seen in Theorem 2.1.3 which describes the evolution of the trait once the behavior of the genes is known. From there we obtain equation (2.19) which is a typical equation of quantitative genetics. To the best of our knowledge, this gene-centric approach is new. We do note similarities with the ideas of the Dynamic Maximum Entropy (DME) approximation method [145]. This approximation from statistical physics consists in a joint modelling of deterministic macroscopic observables, corresponding in our setting to $\overline{s}(\mathscr{L}(f_t))$, and the evolution of a typical locus conditional on these observables.

Genetic structure

Our approach raises the prospects of future exciting developments. The gene's eye-view in equation (2.10) can be extended without any difficulty to incorporate diploidy and dominance. It would also be important in terms of application to incorporate unequal allelic effects, replacing the set of genotypes $\{-1,+1\}^L$ with $\prod_{\ell\in[L]}\{-\alpha_\ell,+\alpha_\ell\}$ where the α_ℓ have some distribution. This is not hard to do in our current setting, provided the allelic effects are bounded. However the ideal scaling would allow the distribution of allelic effects to have an exponential or even heavy (polynomial) tail. For instance, [146] found rough estimates of tail exponents in additive effects between 1 and 2.5, and [116, 124] assume exponential tails. Similarly, we could account for mutation rate variation, letting θ^+, θ^- vary between the different loci. We will explore some of these extensions in Chapter 4.

The strength of selection

We chose in (2.2) to let the logfitness of an organism be of order L. This let us accommodate both stabilizing, directional and disruptive selection. In the limit equation (2.19), we see that for any initial condition, the mean trait value will evolve on the same timescale as f_t in equation (2.10). In biological terms, the evolution of the trait is on the same timescale as genetic drift. This is in stark contrast with articles from the literature on stabilizing selection such as [124] which typically find that the evolution of the trait is much faster than the evolution of the underlying genes. There are differences with the underlying model, including in the way mutations are specified (they use the infinite-allele infinite-loci model and assume no mutational bias), but we believe that the results of [124] can be recovered from our model assuming the strength of selection to be of order L^2. Indeed, in Lemma 2.3.6, we compute the contribution of selection to the dynamics of locus ℓ under linkage equilibrium. Under stabilizing selection ($\kappa > 0$ in (2.7)), our result reads

$$4\kappa\left(\eta - \frac{1}{L}\sum_{\ell'\in[L]}(2X_t^{\ell'}-1)\right) + \mathcal{O}\left(\frac{1}{L}\right)$$

where we recall that $\mu_{\mathbf{X}_t}$ is the empirical distribution of $(p^\ell(\mathbf{X}_t))_{\ell\in[L]}$. But the same computations, replacing L with L^2 in the base equation (2.21) yield the following selection coefficient at locus ℓ

$$4\kappa\left(L\eta - \sum_{\ell'\in[L]}(2X_t^{\ell'}-1)\right) + 2\kappa\left(X_t^\ell - \frac{1}{2}\right) \tag{2.20}$$

This corresponds to the selection coefficient as computed from Wright's formula in [96]. The second term is known to biologist as Robertson's underdominant term [23], and has been crucial for the evolutionary interpretation of GWAS results [116].

Extending our results to the case where selection is of order L^2 requires handling the possibly degenerate first term of (2.20). Furthermore, our control on linkage disequilibrium from Section 2.3.4 will be affected, meaning the requirements on the strength of recombination ρ for propagation of chaos will be stronger than in Theorem 2.1.1.

Returning to the case when selection is of order L, if we assume $(\theta^+ - \theta^-)/(\theta^+ + \theta^-) \neq z^*$ in (2.7) and X has distribution given by the stationary solution in Corollary 2.2.4, then

$$\mathbb{E}\left[2X - 1 - z^*\right] = \mathcal{O}(1)$$

whereas Theorem 2.1.3 states

$$\sigma^2 = \mathcal{O}(1/L).$$

This means in particular that at equilibrium, the population is very far from the selection optimum $|\mathbb{E}\left[Z(g)\right] - z^*| \gg \sigma$. This corresponds to the drift-barrier hypothesis from [147]. The idea is that when selection is very weak, it cannot overpower the forces of mutation and genetic drift, and therefore the population at equilibrium remains very far from the optimum. This hypothesis was developed specifically for the evolution of the mutation rate, meaning the trait $Z(g)$ corresponds to the probability that g mutates, making the models required to study this much more involved than ours (see for instance [148]).

The suppression of linkage by recombination

Obtaining equation (2.10) required strong recombination, effectively enforcing independent between most loci. This prevents the formation of "linkage blocks". We do not believe our result to be optimal. For instance, if all loci are evenly spaced, except two loci ℓ_1, ℓ_2 so that $r_{\{\ell_1,\ell_2\}} = 0$, then $r^{**} = 0$, even though we expect that these two loci should not matter in the grand scheme of things. Furthermore, equation (2.10) seems to fit the dynamics of simulations even when ρ is of order L (see Fig 2.1).

Finding the optimal scaling for ρ will be a daunting task. The suppression of linkage blocks by recombination was masterfully described in [121], in a very broad deterministic setting with recombination, bounded selection (whereas ours is of order L) and point mutations (no genetic drift). This approach assumes very rare mutations: a newborn's mutations are given by a point process on $[0, 1]$ (representing the positions of the mutations along the chromosome). The main result is that linkage will be negligible if, loosely speaking, recombination separates two new mutations before a third one occurs. Statistical physicists have studied polygenic adaptation using Random Energy Models. These models typically assume the fitness of a genotype γ to be of the form

$$W(\gamma) = \sum_{\ell_1,\ell_2} f_{\ell_1\ell_2}\gamma_{\ell_1}\gamma_{\ell_2}$$

for i.i.d random coefficients $(f_{\ell_1\ell_2})_{\ell_1,\ell_2\in[L]}$. At least two phase transitions were identified for low recombination, which they call the transition from quasi-linkage equilibrium to clonal condensation or to non-random coexistence [70]. These transitions see the appearance of very fit combination of genes which disproportionally contribute to the population without recombination breaking them up fast enough. The distinction between the two is that clonal condensation sees the appearance of true clones whereas in non-random coexistence a cloud of fit genotypes dominates the population. The phase transition occurs when selection is of the same order as recombination, corresponding in our system to $\rho \sim L$.

2.1.6 Outline of the paper

Our paper is organized as follows. In Section 2.2, we formally introduce McKean-Vlasov diffusions which generalize (2.10) and prove well-posedness of the associated martingale problem. In Section 2.3, we prove Theorems 2.1.1 and 2.1.3.

On notation

Summary of notations

$\square_A$	For $A \subseteq [L]$, the set $\{-1,+1\}^A$
$\gamma^{\|A}$	For $A \subseteq [L]$ and $\gamma \in \square_{[L]}$, the restriction of γ to A
$\mathbb{X}^A$	The set of probability measures on $\square_A$, denoted $\mathbf{x} = (x(\gamma))_{\gamma \in \square_A}$
$\mathbf{x}^A = (x^A(\gamma))_{\gamma \in \square_A}$	For $\mathbf{x} \in \mathbb{R}^{\square_{[L]}}$, the marginal of $\mathbf{x}$ on $\square_A$
$p^\ell(\mathbf{x})$	The frequency of the $+1$ allele at locus ℓ: it is the same as $x^{\{\ell\}}(+1)$
$\pi(\mathbf{x})$	The linkage equilibrium projection of $\mathbf{x}$ (see Section 2.3.1)
$\mathbf{x}[\varphi(g)]$	For $\mathbf{x} \in \mathbb{X}^{[L]}$, φ a function on $\square_{[L]}$, this is the expectation of $\varphi(g)$ where g has law $\mathbf{x}$.
$\mathbf{Var}_{\mathbf{x}}, \mathbf{Cov}_{\mathbf{x}}$	The variance and covariances associated with $\mathbf{x}$ (applied to functions of g)
$\mu_{\mathbf{x}}$	The allelic law: $\mu_{\mathbf{x}} := \frac{1}{L} \sum\limits_{\ell \in [L]} \delta_{p^\ell(\mathbf{x})}$
$\mathscr{L}(X)$	The law of a variable X.
$\mathcal{O}$	We write $\varphi_1 = \mathcal{O}(\varphi_2)$ iff there is a constant $C > 0$ such that $\|\varphi_1\| \leq C\|\varphi_2\|$.
o	We write $\varphi_1 = o(\varphi_2)$ iff there is a function h such that $\|\varphi_1\| \leq h\|\varphi_2\|$ and $h \to 0$.
C	A positive constant **whose value may change from one line to the next**.

Marginals and restrictions

For a genotype $\gamma \in \square_{[L]}$ and a subset $A \subseteq [L]$, $\gamma^{|A} := (\gamma_\ell)_{\ell \in A} \in \square_A$ is the restriction of γ to A.

A population $\mathbf{x} \in \mathbb{X}^{[L]}$ can be seen as a vector of $\mathbb{R}^{\square_{[L]}}$, written $(x(\gamma))_{\gamma \in \square_{[L]}}$. It can also be seen as a probability on $\square_{[L]}$. We write $\mathbf{x}[F(g)]$ for the expectation of $F(g)$, where g is a random variable on $\square_{[L]}$ with law $\mathbf{x}$ and F is a function on $\square_{[L]}$. We similarly define $\mathbf{Cov}_{\mathbf{x}}$ and $\mathbf{Var}_{\mathbf{x}}$ to be the covariance and variance associated with $\mathbf{x}$, evaluated on functionals of g.

For $\mathcal{I} \subseteq [L]$, $\mathbf{x} \in \mathbb{R}^{\square_{[L]}}$, we let $\mathbf{x}^{\mathcal{I}} \equiv (x^{\mathcal{I}}(\gamma))_{\gamma \in \square_{\mathcal{I}}}$ be the marginal of $\mathbf{x}$ on $\mathcal{I}$. For $\mathbf{y} \in \mathbb{R}^{\square_{[L]}}$ we may then define $\mathbf{x}^{\mathcal{I}} \otimes \mathbf{y}^{\mathcal{I}^c}$ as the product vector, that is, the vector such that its $\gamma-$th coordinate is

$$\sum_{\substack{\gamma^1,\gamma^2 \in \square_{[L]} \\ \gamma^{1|\mathcal{I}} = \gamma^{|\mathcal{I}} \\ \gamma^{1|\mathcal{I}} = \gamma^{|\mathcal{I}^c}}} x(\gamma^1) y(\gamma^2)$$

The parameters

L	The number of loci and the strength of selection
	Recombinator R parameters
ρ	The strength of recombination (it depends on L, see Theorem 2.1.1)
ν	The measure associated to recombination (it depends on L)
	Mutator Θ parameters
θ^+, θ^-	The mutation rate from -1 to $+1$ and back.
	Selector S parameters
U	Under quadratic selection (2.6), $W(\gamma) = U(Z(\gamma))$ where U is a polynomial of order 1 or 2

2.2 Large genome limit

In this section, we describe McKean-Vlasov diffusions and the associated non-linear IPDEs, making sense of equation (2.10) which describes the limit behavior of our system as $L \to +\infty$. We fix two bounded measurable functions $a : \mathbb{R} \to \mathbb{R}_+$ and $b : \mathbb{R} \times \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ and a compactly supported probability measure $m_0 \in \mathcal{P}(\mathbb{R})$.

2.2.1 McKean-Vlasov diffusions

Our aim is to give a sense to the following McKean-Vlasov SDE

$$\begin{aligned} \mathrm{d}f_t = \quad & b(f_t, \xi_t)\mathrm{d}t + \sqrt{a(f_t)}\mathrm{d}B_t \qquad ; \qquad \mathscr{L}(f_0) = m_0 \\ & \text{with } \xi_t = \mathscr{L}(f_t), \text{ the law of } f_t \end{aligned} \tag{2.21}$$

Before going into (2.21), we first make a small detour and consider an alternative problem. Let us now consider the following SDE

$$\mathrm{d}\hat{f}_t = b(\hat{f}_t, \zeta_t)\mathrm{d}t + \sqrt{a(\hat{f}_t)}\mathrm{d}B_t \qquad ; \qquad \mathscr{L}(\hat{f}_0) = m_0 \tag{2.22}$$

with $\zeta \in \mathbb{D}([0,T], \mathcal{P}(\mathbb{R}))$. Recall that (2.22) admits a weak solution on $[0,T]$ iff there exists a filtration $(\mathscr{F}_t; t \geq 0)$ and an adapted pair $(\hat{f}, B)$, such that B is a Brownian motion and (2.22) is satisfied. Theorem 6.1.6 of [149] implies the existence of weak solutions to (2.22) for any initial condition m_0 as long as a and $b(\cdot, m)$ are continuous for any $m \in \mathcal{P}(\mathbb{R})$.

Given $\zeta \in \mathcal{P}(\mathbb{R})$, let us now define the following differential operator

$$\forall \varphi \in \mathcal{C}_c^2(\mathbb{R}), \quad G_\zeta \varphi(x) := b(x, \zeta)\varphi'(x) + \frac{1}{2}a(x)\varphi''(x).$$

where $\mathcal{C}_c^2(\mathbb{R})$ is the space of $\mathbb{R}-$valued, compactly supported, continuously twice differentiable functions on $\mathbb{R}$.

The existence and uniqueness in law of a solution to (2.22) can be investigated through the associated martingale problem. For $\zeta \in \mathbb{D}([0,T], \mathcal{P}(\mathbb{R}))$, we say $\hat{f}$ solves the martingale problem for $(G_{\zeta_t})_{t \in [0,T]}$ iff

$(*)$ $\forall \varphi \in \mathcal{C}_c^2(\mathbb{R}),$ $\varphi(\hat{f}_t) - \varphi(\hat{f}_0) - \int_0^t \mathrm{d}u \, G_{\zeta_u}\varphi(\hat{f}_u)$ is a martingale in the filtration of $\hat{f}$ and $\mathscr{L}(f_0) = m_0$

Existence and uniqueness in law of a weak solution to (2.22) is equivalent to existence and uniqueness of a solution to $(*)$ (see Theorem 4.5.2 in [149]).

Analogously, one can define a weak solution to the McKean-Vlasov SDE (2.21). Existence and uniqueness of weak solutions to the SDE (2.21) again boil down to the existence and uniqueness of solutions to the mean-field Martingale problem

$(**)$ $\forall \varphi \in \mathcal{C}_c^2(\mathbb{R}),$ $\varphi(f_t) - \varphi(f_0) - \int_0^t \mathrm{d}u \, G_{\xi_u}\varphi(f_u)$ is a martingale in the filtration of f with $\mathscr{L}(f_0) = m_0$ and $\mathscr{L}(f_t) = \xi_t$

2.2.2 Weak solutions to non-linear IPDEs

We say a measure-valued process $\xi \in \mathbb{D}([0,T],\mathcal{P}(\mathbb{R}))$ is a *weak* solution to the non-linear IPDE

$$\partial_t u_t(x) = -\partial_x(b(x,u_t(\cdot))u_t(x)) + \frac{1}{2}\partial_{xx}(a(x)u_t(x)) \qquad ; \qquad u_0 = m_0 \tag{2.23}$$

if $t \mapsto \xi_t$ is continuous, $\xi_0 = m_0$ and we have

$$\forall \varphi \in \mathcal{C}^2_c(\mathbb{R}), \qquad \frac{\mathrm{d}}{\mathrm{d}t} < \xi_t, \varphi > \;=\; < \xi_t, b(\cdot,\xi_t)\varphi' > + \frac{1}{2} < \xi_t, a\varphi'' >$$

Consider $(f_t)_{t\in[0,T]}$ a weak solution to the McKean-Vlasov equation (2.21). Then the process $(\mathscr{L}(f_t))_{t\in[0,T]}$ is a weak solution to the non-linear Fokker-Planck IPDE (2.23).

Let us now discuss the reciprocal statement: whether a weak solution to (2.23) is necessarily associated with a weak solution to (2.21). We will need

Theorem 2.2.1 (Superposition principle (see Theorem 2.5 of [150]))**.** *Consider $\tilde{a}, \tilde{b}$ measurable bounded functions on $\mathbb{R} \times [0,T]$ and define the generator*

$$G_t\varphi(x) := \tilde{b}(x,t)\varphi'(x) + \frac{1}{2}\tilde{a}(x,t)\varphi''(x)$$

Let $\xi \in \mathbb{D}([0,T],\mathcal{P}(\mathbb{R}))$ be a weak solution to

$$\partial_t u_t(x) = -\partial_x(\tilde{b}(x,t)u_t(x)) + \frac{1}{2}\partial_{xx}(\tilde{a}(x,t)u_t(x)) \qquad ; \qquad u_0 = m_0$$

Then there exists a solution to the martingale problem

$$\forall \varphi \in \mathcal{C}^2_c(\mathbb{R}), \quad \varphi(f_t) - \varphi(f_0) - \textstyle\int_0^t \mathrm{d}u\, G_u\varphi(f_u) \text{ is a martingale in the filtration of } f.$$

such that $\mathscr{L}(f_t) = \xi_t$.

Remark 7. *Our notion of weak solution corresponds to the notion of narrowly-continuous weak solution in [150].*

Let ξ_t be a weak solution to (2.23). Then in particular ξ is a weak solution to the linear PDE

$$\partial_t u_t(x) = -\partial_x(b(x,\xi_t)u_t(x)) + \frac{1}{2}\partial_{xx}(a(x)u_t(x)) \qquad ; \qquad u_0 = m_0$$

We may therefore apply the superposition principle to ξ to find a solution $(f_t)_{t\in[0,T]}$ to the martingale problem

$$\forall \varphi \in \mathcal{C}^2_c(\mathbb{R}), \quad \varphi(f_t) - \varphi(f_0) - \textstyle\int_0^t \mathrm{d}u\, G_{\xi_u}\varphi(f_u) \text{ is a martingale in the filtration of } f$$

such that f_t has law ξ_t. It follows that f_t is a solution to $(**)$.

This shows that if we prove uniqueness of the solutions to $(**)$, we will automatically get uniqueness of solutions to (2.23).

2.2.3 Uniqueness of solutions to Wright-Fisher-type McKean-Vlasov SDE

There is a large literature on McKean-Vlasov diffusions, but Wright-Fisher-type SDEs such as ours require special handling because the Brownian coefficient is degenerate and not Lipschitz. This motivates the next theorem.

We metrize the weak topology on $\mathcal{P}(\mathbb{R})$ with the total variation distance (see [69], proposition 4)

$$\forall \xi, \zeta \in \mathcal{P}(\mathbb{R}), \qquad \mathcal{D}_1(\xi,\zeta) := \sup_{||\varphi||_\infty \le 1} | < \xi - \zeta, \varphi > |$$

with $\langle \xi, \varphi \rangle = \int_0^1 \varphi(x)\xi(\mathrm{d}x)$, where the supremum is over all measurable functions φ bounded by 1.

Proposition 2.2.2. *Consider three bounded measurable function* $a, c, \tilde{\Theta} : \mathbb{R} \to \mathbb{R}_+$ *and a Lipschitz function* $\tilde{s} : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$, *and define* $b(x, \zeta_t) := \sqrt{a(x)}c(x)\tilde{s}(\zeta_t) + \tilde{\Theta}(x)$. *Assume that for any fixed* $\zeta \in \mathbb{D}([0,T], \mathcal{P}(\mathbb{R}))$, *the linear martingale problem (*) admits a unique weak solution. Then the McKean-Vlasov SDE (2.21) admits at most one weak solution and the non-linear PDE (2.23) admits at most one weak solution.*

Proof. We prove uniqueness of solutions with a Girsanov transform, followed by Grönwall's Lemma. This sort of proof was already used in [151].

Let us consider two solutions to the Martingale problem $(**)$ with initial distribution m_0 and denote by ξ and ζ their respective laws. We can construct two weak solutions (f^ξ, B^ξ) and (f^ζ, B^ζ) to the McKean-Vlasov SDE (2.21). Consider a test function φ. We wish to bound the distance

$$| < \xi_t - \zeta_t, \varphi > | = \left| \mathbb{E}\left[\varphi(f_t^\xi)\right] - \mathbb{E}\left[\varphi(f_t^\zeta)\right] \right|$$

and show that it must be 0.

Step 1: Girsanov transform. Consider the classical Cameron-Martin-Girsanov change of measure

$$M_t := \int_0^t \frac{b(f_u^\zeta, \xi_u) - b(f_u^\zeta, \zeta_u)}{\sqrt{a(f_u^\zeta)}} \mathrm{d}B_u^\zeta$$

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \quad := \quad \exp[M_T - \frac{1}{2}\langle M \rangle_T^{QV}]$$

where $\langle \cdot \rangle^{QV}$ is the quadratic variation. Since for any $x \in \mathbb{R}$,

$$\frac{b(x, \xi_u) - b(x, \zeta_u)}{\sqrt{a(x)}} = c(x)\left(\tilde{s}(\xi_u) - \tilde{s}(\zeta_u)\right)$$

is bounded, $\mathbb{Q}$ is well-defined since M_t satisfies Novikov's condition.

Theorem 6.4.2 of [149] implies that f^ζ under $\mathbb{Q}$ satisfies the martingale problem

$$\forall \varphi \in \mathcal{C}_c^2(\mathbb{R}), \quad \varphi(f_t) - \varphi(f_0) - \textstyle\int_0^t \mathrm{d}u\, G_{\xi_u}\varphi(f_u) \text{ is a martingale with } \mathscr{L}(f_0) = m_0$$

Since we assumed that this linear martingale problem has a unique solution, it follows that f^ζ under $\mathbb{Q}$ has the same law as f^ξ under $\mathbb{P}$. We thus obtain

$$\mathbb{E}\left[\varphi(f_t^\xi)\right] = \mathbb{Q}\left[\varphi(f_t^\zeta)\right] = \mathbb{E}\left[\varphi(f_t^\zeta)e^{M_t - \frac{1}{2}\langle M \rangle_t^{QV}}\right].$$

Step 2: Grönwall's Lemma. We now alleviate notations by writing $f^\zeta \equiv f$. From the previous discussion,

$$| < \xi_t - \zeta_t, \varphi > | = \left| \mathbb{E}\left[\varphi(f_t)e^{M_t - \frac{1}{2}\langle M \rangle_t^{QV}}\right] - \mathbb{E}\left[\varphi(f_t)\right] \right|$$

Write $\mathcal{E}_t := e^{M_t - \frac{1}{2}\langle M\rangle_t^{QV}}$. We have

$$\begin{aligned}
| < \xi_t - \zeta_t, \varphi > | =& |\mathbb{E}\left[\varphi(f_t)\left(\mathcal{E}_t - 1\right)\right]| \\
\leq& ||\varphi||_\infty \, \mathbb{E}\left[\left|e^{M_t - \frac{1}{2}\langle M\rangle_t^{QV}} - 1\right|\right] \\
=& ||\varphi||_\infty \, \mathbb{E}\left[\left|\int_0^t \mathrm{d}M_u \, \mathcal{E}_u\right|\right] \\
=& ||\varphi||_\infty \, \mathbb{E}\left[\left|\int_0^t \mathrm{d}B_u \, c(f_u)(\tilde{s}(\xi_u) - \tilde{s}(\zeta_u))\mathcal{E}_u\right|\right] \\
\leq& ||\varphi||_\infty \, \mathbb{E}\left[\left(\int_0^t \mathrm{d}B_u \, c(f_u)(\tilde{s}(\xi_u) - \tilde{s}(\zeta_u))\mathcal{E}_u\right)^2\right]^{\frac{1}{2}} \\
\leq& ||\varphi||_\infty \left(\mathbb{E}\left[\int_0^t \mathrm{d}u \, c(f_u)^2 \left(\tilde{s}(\xi_u) - \tilde{s}(\zeta_u)\right)^2 \mathcal{E}_u^2\right]\right)^{\frac{1}{2}} \\
\leq& C_{\tilde{s}} ||\varphi||_\infty \, ||c||_\infty \left(\int_0^t \mathrm{d}u \, \mathcal{D}_1(\xi_u, \zeta_u)^2 \mathbb{E}\left[\mathcal{E}_u^2\right]\right)^{\frac{1}{2}}
\end{aligned}$$

where $C_{\tilde{s}}$ is the Lipschitz constant of $\tilde{s}$. We used in the third line that $\mathcal{E}_t$ is the exponential martingale associated with M_t, and in the fifth line we used the Cauchy-Schwarz inequality. Since $\mathbb{E}\left[\mathcal{E}_u^2\right]$ has uniform bounds on $[0, T]$, we thus obtain

$$< \xi_t - \zeta_t, \varphi >^2 \leq ||\varphi||_\infty^2 C^2 \int_0^t \mathrm{d}u \, \mathcal{D}_1(\xi_u, \zeta_u)^2$$

for some constant $C > 0$. Taking the supremum over $||\varphi||_\infty \leq 1$ yields

$$\mathcal{D}_1(\xi_t, \zeta_t)^2 \leq C \int_0^t \mathrm{d}u \, \mathcal{D}_1(\xi_u, \zeta_u)^2$$

The result follows from Grönwall's lemma. □

We obtain as a Corollary the well-posedness of (2.10).

Corollary 2.2.3. *For $m_0 \in \mathcal{P}([0, 1])$, there exists a unique weak solution to the McKean Vlasov problem (2.10)*

$$\mathrm{d}f_t \; = \; \overline{s}(\mathscr{L}(f_t)) \; f_t(1 - f_t)\mathrm{d}t + \overline{\Theta}(f_t)\mathrm{d}t + \sqrt{f_t(1 - f_t)} \, \mathrm{d}B_t \qquad ; \qquad \mathscr{L}(f_0) = m_0$$

In particular $\mathscr{L}((f_t)_{t\in[0,T]})$ is a weak solution to the IPDE (2.11) on $[0, 1]$

$$\partial_t u_t(x) = -\partial_x\left[\left(\overline{s}(u_t(\cdot))x(1 - x) + \overline{\Theta}(x)\right) u_t(x)\right] + \frac{1}{2}\partial_{xx}\left(x(1 - x)u_t(x)\right)$$

with initial condition m_0.

Proof. Existence will be obtained from a convergence argument in the next section (see Theorem 2.1.1).

Let $a(x) = c(x)^2 = x(1 - x)\mathbf{1}_{[0,1]}(x)$ and $\tilde{\Theta} = \overline{\Theta}, \tilde{s} = \overline{s}$ in Proposition 2.2.2. Note that for $\xi \in \mathcal{P}([0, 1])$, $\xi \mapsto < \xi, 2\mathrm{Id} - 1 >$ is Lipschitz. Finally, note that for any fixed $\zeta \in \mathbb{D}([0, T], \mathcal{P}([0, 1]))$, the equation

$$\mathrm{d}f_t \; = \; \overline{s}(\zeta_t) \; f_t(1 - f_t)\mathrm{d}t + \overline{\Theta}(f_t)\mathrm{d}t + \sqrt{f_t(1 - f_t)} \, \mathrm{d}B_t \qquad ; \qquad \mathscr{L}(f_0) = m_0$$

is a Wright-Fisher diffusion with a unique weak solution.

We may therefore apply Proposition 2.2.2. The IPDE (2.11) is the Fokker-Planck equation associated with (2.10). The uniqueness of its solution is given by the discussion in Section 2.2.2. □

2.2.4 Stationary distribution for quadratic selection

We focus on quadratic selection.

Corollary 2.2.4. *Assume* $\theta^+, \theta^- > 0$ *and consider the case of symmetric quadratic selection*

$$\overline{s}(\xi) = -2\kappa\,(<\xi, 2Id - 1> - z^*)$$

for some parameter $\kappa \in \mathbb{R}, z^*$*. Then*

- *(stabilizing selection) Suppose* $\kappa \geq 0$*. Then* χ *has a unique fixed point at 0.*
- *(disruptive selection) Assume* $\kappa < 0$ *and no mutational bias:* $\theta^+ = \theta^-$ *and* $z^* = 0$*. Let* $\kappa_c := -\frac{4\theta^+ + 1}{2}$*. Then we can find* $\delta > 0$ *such that if* $\kappa \in (\kappa_c - \delta, \kappa_c]$*,* χ *has at least three fixed points.*

Proof. Recall the definition of Π_y in (2.16) and the definition of χ in (2.17). Let F be the cumulant generating function of Π_0

$$F(y) := \ln(<\Pi_0, \exp(y \text{ Id})>)$$

A quick computation shows that $\chi(y) = -2\kappa(2F'(2y) - 1 - z^*)$. In particular

$$\chi'(y) = -8\kappa F''(2y) \tag{2.24}$$

It is also easy to see that $F''(2y)$ is the variance of Π_y, and in particular is positive. In the $\kappa \geq 0$ case, we get that χ is non-increasing. In particular it will have a single fixed point.

We turn to the case $\kappa < 0, z^*$. First, notice that by symmetry we will always have $\overline{s}(\Pi_0) = 0$. In particular, for any κ, 0 is a fixed point of χ. Further,

$$\chi(y) = y\chi'(0) + \frac{y^2}{2}\chi''(0) + \frac{y^3}{6}\chi'''(0) + o(y^3) \tag{2.25}$$

The variance of Π_0 is $F''(0) = \frac{1}{4(4\theta^+ + 1)}$. From (2.24) we find

$$\chi'(0) = \frac{\kappa}{\kappa_c}$$

We similarly compute

$$\chi''(0) = -16\kappa\, F'''(0) \qquad ; \qquad \chi'''(0) = -32\kappa F''''(0)$$

Since $F'''(0)$ is the skew of Π_0, it is 0 by symmetry. Recall $F''''(0)$ is the fourth cumulant of Π_0. This can be seen to be negative for symmetric Beta distributions (using for instance (25.15d), p.217 of [152]). We then rewrite (2.25) as

$$\chi(y) - y = y\frac{\kappa - \kappa_c}{\kappa_c} - \frac{y^3}{6}\kappa 32 F''''(0) + o(y^3)$$

When $\kappa \leq \kappa_c$, the two terms on the right-hand side are of opposite signs. The result follows. □

2.3 Convergence to the McKean-Vlasov SDE under strong recombination

In this section we prove Theorem 2.1.1 (in Section 2.3.6), and Theorem 2.1.3 (in Section 2.3.7). We start with an outline of the main step of the proofs.

2.3.1 Heuristics and outline of the proof

Recall that we consider a multidimensional SDE of the form

$$\mathrm{d}\mathbf{X}_t \; = \; (\rho R(\mathbf{X}_t) + \Theta(\mathbf{X}_t) + LS(\mathbf{X}_t))\mathrm{d}t + \Sigma(\mathbf{x})\mathrm{d}\mathbf{B}_t$$

Step 1. For every $\mathbf{x} \in \mathbb{X}^{[L]}$, define $\pi(\mathbf{x})$ as

$$\pi(\mathbf{x}) := \bigotimes_{\ell \in [L]} \mathbf{x}^{\{\ell\}}$$

i.e., $\pi(\mathbf{x})$ is the product measure whose one dimensional marginals coincides with the ones of $\mathbf{x}$. We extend the definition of π to $\bigcup_{A \subseteq [L]} \mathbb{X}^A$, such that

$$\forall \mathbf{x} \in \mathbb{X}^A, \quad \pi(\mathbf{x}) := \bigotimes_{\ell \in A} \mathbf{x}^{\{\ell\}}. \tag{2.26}$$

π gives the attractors of the recombinator R as the following Lemma shows.

Lemma 2.3.1. *For $\mathrm{x}_0 \in \mathbb{X}^{[L]}$, define x_t to be the unique solution to $\frac{\mathrm{d}\mathrm{x}_t}{\mathrm{d}t} = R(\mathrm{x}_t)$ with initial condition x_0. Then, provided the recombination measure ν is non-degenerate, x_t converges to $\pi(\mathrm{x}_0)$ as $t \to +\infty$.*

Proof. We refer the reader to [61] and the references therein. ☐

Under the strong recombination assumption, the driving force is recombination. From the previous result, we may expect that

$$\mathbf{X}_t \; \approx \; \pi(\mathbf{X}_t)$$

so that the SDE should asymptotically diffuse on the stable manifold for the recombinator

$$\Gamma^{[L]} := \{\mathbf{x} \in \mathbb{X}^{[L]} \mid \quad \mathbf{x} = \pi(\mathbf{x})\}.$$

Biologically speaking, we expect the system to be at linkage equilibrium (LE) due to the overwhelming effect of recombination.

Step 2. For $\ell \in [L]$, define

$$S^\ell(\mathbf{x}) := S^{\{\ell\}}(\mathbf{x})(+1)$$

where we recall that $S^{\{\ell\}}$ is the marginal of S on $\{\ell\}$. We show in Corollary 2.3.5 that

$$\forall \ell \in [L], \quad \mathrm{d}p^\ell(\mathbf{X}_t) = \left(\overline{\Theta}(p^\ell(\mathbf{X}_t)) + LS^\ell(\mathbf{X}_t)\right)\mathrm{d}t + \sqrt{p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))}\mathrm{d}\hat{B}^\ell_t$$

where $\hat{B}^l$ is a Brownian motion. It is also easy to see that if $\mathbf{x}$ belongs to the LE manifold (see Lemma 2.3.6)

$$LS^\ell(\mathbf{x}) \approx \bar{s}(\mu_{\mathbf{x}})p^\ell(\mathbf{x})(1-p^\ell(\mathbf{x})) \tag{2.27}$$

It should follow that for any $\ell \in [L]$,

$$\mathrm{d}p^\ell(\mathbf{X}_t) \approx \left(\overline{\Theta}(p^\ell(\mathbf{X}_t)) + \bar{s}(\mu_{\mathbf{X}_t})p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))\right)\mathrm{d}t + \sqrt{p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))}\mathrm{d}\hat{B}^\ell_t$$

In order to derive a mean field approximation, it remains to prove that loci become decorrelated at the limit. Define the linkage disequilibrium between $\ell_1 \neq \ell_2 \in [L]$ as

$$D^{\ell_1,\ell_2}(\mathbf{x}) := \mathbf{Cov}_{\mathbf{x}}\left[\mathbb{1}_{[g_{\ell_1}=+1]}, \mathbb{1}_{[g_{\ell_2}=+1]}\right] \tag{2.28}$$

where we recall that $\mathbf{Cov}_{\mathbf{x}}$ is the covariance of functionals of a random variable g with law $\mathbf{x}$. We show (again in Corollary 2.3.5) that

$$\mathrm{d}\left\langle p^{\ell_1}(\mathbf{X}), p^{\ell_2}(\mathbf{X})\right\rangle_t^{QV} \ = \ D^{\ell_1,\ell_2}(\mathbf{X}_t)\mathrm{d}t = 0 \tag{2.29}$$

where the last equality holds provided that $\mathbf{X}_t$ is on the LE manifold $\Gamma^{[L]}$. Putting everything together, if recombination is strong enough, we should expect a propagation of chaos principle to hold.

Technical ingredients. The previous heuristics rely on the underlying assumption that $\mathbf{X}_t$ is on the LE manifold. Making this rigorous will raise one major difficulty. We need to derive the conditions on the recombinator so that $\mathbf{X}_t$ remain close enough to the boundary so that the previous estimates remain valid.

The first step is the following proposition which allows to justify (2.27) and (2.29) by controlling

$$Y_t^A := ||\mathbf{X}_t^A - \pi(\mathbf{X}_t^A)||_2 \tag{2.30}$$

on every subset $A \subset [L]$ of size 2 or 3.

Proposition 2.3.2. *We have*

$$\forall \ell_1 \neq \ell_2 \in [L], \qquad |D^{\ell_1,\ell_2}(\mathbf{x})| \leq \ ||\mathbf{x}^{\{\ell_1,\ell_2\}} - \pi(\mathbf{x}^{\{\ell_1,\ell_2\}})||_2 \tag{2.31}$$

$$\forall \ell_0 \in [L], \qquad \left|S^{\ell_0}(\mathbf{x}) - S^{\ell_0}(\pi(\mathbf{x}))\right| \leq \ C \sum_{\substack{A\subseteq[L]\setminus\{\ell_0\} \\ 1\leq \#A \leq 2}} \frac{1}{L^{\#A}} ||\mathbf{x}^{\{\ell_0\}\cup A} - \pi(\mathbf{x}^{\{\ell_0\}\cup A})||_2 \tag{2.32}$$

for some constant $C > 0$ independent of L.

To get control on (2.30), we will use the linearized recombinator ∇R. The eigenvalues of ∇R were computed in [153] p. 107. We give a full spectral characterization of ∇R in Section 2.3.3. In particular, we obtain that the system is always attracted towards the LE manifold at a rate at least

$$r_A := \min_{\substack{\ell_1,\ell_2 \in A \\ \ell_1 \neq \ell_2}} r_{\{\ell_1,\ell_2\}} \tag{2.33}$$

In Section 2.3.4, we use this estimate through a combination of Itô's and Grönwall's lemmas to get quantative bounds on (2.30). In Section 2.3.6 we conclude the proof of Theorem 2.1.1 with standard martingale arguments to prove convergence of $(\mu_{\mathbf{X}_t})_{t\in[0,T]}$ and $(p^\ell(\mathbf{X}_t))_{t\in[0,T]}$. Finally, we obtain Theorem 2.1.3 in Section 2.3.7.

Remark 8. *Katzenberger [154] considered a generic SDE with a strong drift attracting the dynamics on an invariant manifold and derived a slow-fast principle for the stochastic evolution on the manifold. This sort of proof was already used in a population genetics context, in the case $\rho \to +\infty, L = 2$ in [155]. Our system presents two additional complexities. The first difficulty is that the dimension of the problem explodes exponentially with L just as the strength of recombination becomes large. The second difficulty is that we not only need $\mathbf{x} \approx \pi(\mathbf{x})$, we actually need the difference to be small, of order $1/L$, because the strength of selection is of order L. We therefore require quantitative bounds on the linkage disequilibrium on any small set $A \subseteq [L]$.*

2.3.2 Evolution of the marginals

In this section we will derive the SDE for $p^\ell(\mathbf{X}_t)$. We will in fact study $\mathbf{X}_t^A$ for any $A \subseteq [L]$, of which $\{\ell\}$ is a special case. The reason why we need to study $\mathbf{X}_t^A$ for a general A is because we will need to control the divergence from LE of $\mathbf{X}^A$ for small sets A of size at most 3.

Recall the definition of R, Θ, Σ as defined in in the introductory Section 2.1.1. For any subset A, we define the same quantity $\hat{R}^A, \hat{\Theta}^A, \hat{\Sigma}^A$ but on the hypercube $\square_A$. For instance,the operators $\hat{R}^A, \hat{\Theta}^A : \mathbb{X}^A \to \mathbb{R}^{\square_A}$ read

$$\forall \mathbf{x} \in \mathbb{X}^A, \qquad \hat{\Theta}^A(\mathbf{x})(\gamma) := |\theta| \sum_{\ell \in A} \left(\mathbf{x}^{[L] \setminus \{\ell\}} \otimes \mathcal{L}_\theta - \mathbf{x} \right)$$

$$\forall \mathbf{x} \in \mathbb{X}^A \qquad \hat{R}^A(\mathbf{x}) := \sum_{\emptyset \subsetneq \mathcal{I} \subsetneq A} \nu^A(\mathcal{I}) \left(\mathbf{x}^{\mathcal{I}} \otimes \mathbf{x}^{A \setminus \mathcal{I}} - \mathbf{x} \right)$$

where we recall that ν^A is the marginal of ν on A, and we take an empty sum to be equal to zero. Similarly, $\hat{\Sigma}^A$ is a function

$$\hat{\Sigma}^A : \ \mathbb{X}^{\square_A} \to \ \mathcal{M}\left(\square_A \times \square_A, \mathbb{R}^{\square_A}\right)$$

Define $S^A : \mathbb{X}^L \to \mathbb{R}^{\square_A}$ as

$$\forall x \in \mathbb{X}^L, \qquad S^A(\mathbf{x}) \ = \ (S(\mathbf{x}))^A$$

so that $S^A(\mathbf{x})$ is the generalized marginal of $S(\mathbf{x})$ on A. A direct computation shows that

$$\forall \mathbf{x} \in \mathbb{X}^{[L]}, \qquad S^A(\mathbf{x})(\gamma) = x^A(\gamma) \ \left(\mathbf{x}\big[W(g) \mid g^{|A} = \gamma\big] - \mathbf{x}\big[W(g)\big] \right) \tag{2.34}$$

$$= \mathbf{Cov}_{\mathbf{x}}[W(g), \mathbb{1}_{[\ g^{|A} = \gamma]}] \tag{2.35}$$

Remark 9. *Note that $S^A(\mathbf{x})$ is the marginal on A of $S(\mathbf{x})$. However, it is not so for $\hat{R}^A$. $\hat{R}^A$ is defined on $\mathbb{X}^A$ (not on $\mathbb{X}^{[L]}$). To stress out the distinction, we write $\hat{R}^A$ and not R^A. The same goes for $\hat{\Theta}^A, \hat{\Sigma}^A$.*

Proposition 2.3.3. *For $A \subseteq [L]$, there exists $\hat{\mathbf{B}}^A = (\hat{B}^A(\gamma^1\gamma^2))_{\gamma^1 \neq \gamma^2 \in \square_A}$ a Gaussian process such that*

$$\forall \gamma \in \square_A, \qquad \mathrm{d}\mathbf{X}_t^A = \left(\rho \hat{R}^A(\mathbf{X}_t^A) + \hat{\Theta}^A(\mathbf{X}_t^A) + L S^A(\mathbf{X}_t) \right) \mathrm{d}t + \hat{\Sigma}^A(\mathbf{X}_t^A)\mathrm{d}\hat{\mathbf{B}}_t^A \tag{2.36}$$

Furthermore, $\hat{B}^A(\gamma^1, \gamma^2) = -\hat{B}^A(\gamma^2, \gamma^1)$ and $\hat{B}^A(\gamma^1, \gamma^2), \hat{B}^A(\gamma^3, \gamma^4)$ are independent Brownian motions whenever $(\gamma^1, \gamma^2) \notin \{(\gamma^3, \gamma^4), (\gamma^4, \gamma^3)\}$.

Proof. We have

$$\begin{aligned} \mathrm{d}X_t^A(\gamma) = \sum_{\hat{\gamma} \in \square_{[L]}} & \mathbb{1}_{[\hat{\gamma}^{|A} = \gamma]} \mathrm{d}X_t(\hat{\gamma}) \\ & = \left(R(\mathbf{X}_t) + \Theta(\mathbf{X}_t)\right)^A(\gamma)\mathrm{d}t + L S^A(\mathbf{X}_t)(\gamma)\mathrm{d}t \\ & \qquad + \sum_{\substack{\hat{\gamma}_1, \hat{\gamma}_2 \in \square_{[L]} \\ \hat{\gamma}_1 \neq \hat{\gamma}_2}} \mathbb{1}_{[\hat{\gamma}_1^A = \gamma]} \sqrt{X_t(\hat{\gamma}_1) X_t(\hat{\gamma}_1)} \mathrm{d}B_t(\hat{\gamma}_1, \hat{\gamma}_2) \end{aligned} \tag{2.37}$$

We first calculate the marginal effect of recombination. We use the Proposition 6 of [59], where the following consistency relation is shown

$$(R(\mathbf{x}))^A = \hat{R}^A(\mathbf{x}^A)$$

Secondly, recall that Θ is the generator corresponding to mutation: each locus mutates independently of the others, from -1 to $+1$ (resp. -1 to $+1$) at rate θ^+ (resp. θ^-). We can therefore expect a consistency property, by which taking the marginal effect of Θ on the loci in A, each locus in A mutates independently of the rest with rates θ^+, θ^-, which translates into

$$(\Theta(\mathbf{x}))^A = \hat{\Theta}^A(\mathbf{x}^A)$$

Formally, this can be proved as follows.

$$(\Theta(\mathbf{x}))^A(\gamma) = |\theta| \sum_{\ell \in [L]} \left(\left(\mathbf{x}^{[L] \setminus \{\ell\}} \otimes \mathcal{L}_\theta \right)^A - \mathbf{x}^A \right)$$

For any $\ell \notin A$, we have

$$(\mathbf{x}^{[L] \setminus \{\ell\}} \otimes \mathcal{L}_\theta)^A = \mathbf{x}^A$$

This means the sum on $\ell \in [L]$ can be restricted to A, which yields $\hat{\Theta}^A(\mathbf{x}^A)$.

Finally, we turn to the Brownian term. This term corresponds to the equation for a neutral Wright-Fisher diffusion with 2^L alleles. It is well-known that the multi-allele Wright-Fisher diffusion admits a consistency property, by which if we group alleles together into $2^{\# A}$ families, the frequencies of these families behave like a $2^{\# A}$-allele Wright-Fisher diffusion. For the unconvinced reader we give a sketch of the proof.

For $\gamma^1, \gamma^2 \in \square_A$, define

$$\mathrm{d}\hat{B}_t^A(\gamma^1, \gamma^2) := \sum_{\substack{\hat{\gamma}_1, \hat{\gamma}_2 \in \square_{[L]} \\ \hat{\gamma}_1^A = \gamma^1 \\ \hat{\gamma}_2^A = \gamma^2}} \sqrt{\frac{X_t(\hat{\gamma}_1) X_t(\hat{\gamma}_2)}{X_t^A(\gamma^1) X_t^A(\gamma^2)}} \mathrm{d}B_t(\hat{\gamma^1}, \hat{\gamma}_2)$$

Check that $\hat{B}^A$ is formally well-defined because

$$\mathbb{1}_{[\hat{\gamma}_1^A = \gamma^1, \hat{\gamma}_2^A = \gamma^2]} \frac{X_t(\hat{\gamma}_1) X_t(\hat{\gamma}_2)}{X_t^A(\gamma^1) X_t^A(\gamma^2)} \leq 1.$$

Recall that for any $\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3, \hat{\gamma}_4 \in \square_{[L]}$

$$\mathrm{d} \langle B(\hat{\gamma}_1, \hat{\gamma}_2), B(\hat{\gamma}_3, \hat{\gamma}_4) \rangle_t^{QV} = (\mathbb{1}_{[\hat{\gamma}_1 = \hat{\gamma}_3, \hat{\gamma}_2 = \hat{\gamma}_4]} - \mathbb{1}_{[\hat{\gamma}_1 = \hat{\gamma}_4, \hat{\gamma}_2 = \hat{\gamma}_3]}) \mathrm{d}t$$

From there, it is straightforward to obtain for $\gamma^1, \gamma^2, \gamma^3, \gamma^4 \in \square_A$

$$\mathrm{d} \langle \hat{B}^A(\gamma^1, \gamma^2), \hat{B}^A(\gamma^3, \gamma^4) \rangle = \left(\mathbb{1}_{[\gamma^1 = \gamma^3, \gamma^2 = \gamma^4]} - \mathbb{1}_{[\gamma^1 = \gamma^4, \gamma^2 = \gamma^3]} \right) \mathrm{d}t.$$

To conclude, let us show $\hat{\Sigma}^A(\mathbf{X}_t^A) \mathrm{d}\mathbf{B}_t^A$ is equal to the Brownian term from (2.37). We write for fixed $\gamma^1 \in \square_A$

$$\sum_{\gamma^2 \in \square_{[L]} \setminus \{\gamma^1\}} \sqrt{X_t^A(\gamma^1) X_t^A(\gamma^2)} \mathrm{d}\hat{B}_t^A(\gamma^1, \gamma^2) = \sum_{\substack{\hat{\gamma}_1, \hat{\gamma}_2 \in \square_{[L]} \\ \hat{\gamma}_1^A = \gamma^1 \neq \hat{\gamma}_2^A}} \sqrt{X_t(\hat{\gamma}_1) X_t(\hat{\gamma}_2)} \mathrm{d}B_t(\hat{\gamma}_1, \hat{\gamma}_2)$$

We can extend the sum to the cases where $[\hat{\gamma}_2^A = \gamma^1, \hat{\gamma}_2 \neq \hat{\gamma}_1]$ because the terms $(\hat{\gamma}_1, \hat{\gamma}_2)$ and $(\hat{\gamma}_2, \hat{\gamma}_1)$ cancel out. This yields the Brownian term from (2.37). □

Remark 10. *We may notice that if* $W = 0$*, then* $S^A = \mathbf{0}$ *and the equation for* $\mathbf{X}_t^A$ *is autonomous.*

We can apply the previous Proposition to $A = \{\ell\}$ and obtain an important Corollary. We need a Lemma

Lemma 2.3.4 (Fleming-Viot property [156]). *For functions F_1 and F_2 on $\square_{[L]}$ we have*

$$\mathrm{d}\left\langle \mathbf{X}[F_1(g)], \mathbf{X}[F_2(g)]\right\rangle_t = \mathbf{Cov}_{\mathbf{X}_t}[F_1(g), F_2(g)]\mathrm{d}t$$

Proof. We have

$$\begin{aligned}\mathrm{d}\left\langle \mathbf{X}[F_1(g)], \mathbf{X}[F_2(g)]\right\rangle_t &= \sum_{\gamma^1,\gamma^2\in\square_{[L]}} F_1(\gamma^1)F_2(\gamma^2)\left(X_t(\gamma^1)\mathbb{1}_{[\gamma^1=\gamma^2]} - X_t(\gamma^1)X_t(\gamma^2)\right)\mathrm{d}t \\ &= \sum_{\gamma^1\in\square_{[L]}} F_1(\gamma^1)F_2(\gamma^1)X_t(\gamma^1)\mathrm{d}t - \sum_{\gamma^1,\gamma^2\in\square_{[L]}} F_1(\gamma^1)F_2(\gamma^2)X_t(\gamma^1)X_t(\gamma^2)\mathrm{d}t\end{aligned}$$

where in the first equality we used that

$$\mathrm{d}\left\langle X(\gamma^1), X(\gamma^2)\right\rangle_t = \left(\mathbb{1}_{[\gamma^1=\gamma^2]}X_t(\gamma^1) - X_t(\gamma^1)X_t(\gamma^2)\right)\mathrm{d}t.$$

The last line is $\mathbf{Cov}_{\mathbf{X}_t}[F_1(g), F_2(g)]$. □

Corollary 2.3.5.

$$\forall \ell\in[L],\quad \mathrm{d}p^\ell(\mathbf{X}_t) = \left(\overline{\Theta}(p^\ell(\mathbf{X}_t)) + LS^\ell(\mathbf{X}_t)\right)\mathrm{d}t + \sqrt{p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))}\mathrm{d}\hat{B}^\ell_t \tag{2.38}$$

with $\overline{\Theta}$ from Theorem 2.1.1 and $(\hat{B}^\ell)_{\ell\in[L]}$ a L-dimensional Brownian motion with

$$\forall i\neq j\in[L],\ \mathrm{d}\left\langle p^{\ell_i}(\mathbf{X}), p^{\ell_j}(\mathbf{X})\right\rangle_t^{QV} = D^{\ell_1,\ell_2}(\mathbf{X}_t)\mathrm{d}t \tag{2.39}$$

where we recall D^{ℓ_1,ℓ_2} from (2.28).

Proof of Corollary 2.3.5. We apply Proposition 2.3.3 to $A=\{\ell\}$ and consider the +1 coordinate, recalling that $p^\ell(\mathbf{x}) = \mathbf{x}^{\{\ell\}}(+1))$. We also use the fact that $\hat{R}^{\{\ell\}} = \mathbf{0}$, i.e., recombination does not alter allele frequencies. To get (2.39), apply Lemma 2.3.4 with $F_i(g) = \mathbb{1}_{[g_{\ell_i}=+1]}$ for $i\in\{1,2\}$. □

Comparing equation (2.38) and the desired limit equation (2.10), we see that we need two things. On the one hand, for any $\ell_1\neq\ell_2$ we must obtain that $\left\langle p^{\ell_1}(\mathbf{X}), p^{\ell_2}(\mathbf{X})\right\rangle_t^{QV}\to 0$. This will be achieved by controlling $D^{\ell_1,\ell_2}(\mathbf{X}_t)$. On the other hand, we need that

$$|LS^\ell(\mathbf{X}_t) - p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))\overline{s}(\mu_{\mathbf{X}_t})| \longrightarrow 0$$

This is much more difficult to obtain. We will get it by showing $S^\ell(\mathbf{X}_t)\simeq S^\ell(\pi(\mathbf{X}_t))$ and using the following Lemma

Lemma 2.3.6. *We have*

$$\forall\ell\in[L],\qquad LS^\ell(\pi(\mathbf{x})) = p^\ell(\mathbf{x})(1-p^\ell(\mathbf{x}))\overline{s}(\mu_{\mathbf{x}}) + \mathcal{O}\left(\frac{1}{L}\right)$$

where $\mathcal{O}$ is uniform in $\mathbf{x}$.

Proof. Recall from the definition of π that $p^\ell(\pi(\mathbf{x})) = p^\ell(\mathbf{x})$. Applying this to equation (2.34) for $\ell\in[L]$ we get

$$\begin{aligned}LS^\ell(\pi(\mathbf{x})) =& Lp^\ell(\mathbf{x})\left(\pi(\mathbf{x})[U(Z(g))|g_\ell=+1] - \pi(\mathbf{x})[U(Z(g))]\right) \\ =& Lp^\ell(\mathbf{x})\Bigg(\pi(\mathbf{x})\left[U\left(\frac{1}{L}+\frac{1}{L}\sum_{\hat{\ell}\in[L]\setminus\{\ell\}} g_{\hat{\ell}}\right)\Bigg|g_\ell=+1\right] \\ &\qquad - \pi(\mathbf{x})\left[U\left(\frac{g_\ell}{L}+\frac{1}{L}\sum_{\hat{\ell}\in[L]\setminus\{\ell\}} g_{\hat{\ell}}\right)\right]\Bigg) \\ =& p^\ell(\mathbf{x})\times 2(1-p^\ell(\mathbf{x}))\times\pi(\mathbf{x})\left[U'\left(\frac{1}{L}\sum_{\hat{\ell}\in[L]\setminus\{\ell\}} g_{\hat{\ell}}\right)\right] + \mathcal{O}\left(\frac{1}{L}\right)\end{aligned}$$

where in the third equality we used that $g_{\hat{\ell}}$ and g_ℓ are independent under $\pi(\mathbf{x})$ and $\pi(\mathbf{x})[g_\ell] = 2p^\ell(\mathbf{x}) - 1$. To conclude, write

$$\pi(\mathbf{x}) \left[\frac{1}{L} \sum_{\hat{\ell} \in [L] \setminus \{\ell\}} g_{\hat{\ell}} \right] = < \mu_{\mathbf{x}}, 2\mathrm{Id} - 1 > + \mathcal{O}\left(\frac{1}{L}\right)$$

Since we assumed U to be a quadratic polynomial, then U' is of degree one and we get the result. □

We conclude this section by proving Proposition 2.3.2 which states that we may control both $D^{\ell_1,\ell_2}(\mathbf{X}_t)$ and $S^\ell(\mathbf{X}_t)$ with $||\mathbf{X}_t^A - \pi(\mathbf{X}_t^A)||_2$ for small sets $A \subseteq [L]$.

Proof of Proposition 2.3.2. The first inequality is readily obtained with

$$\begin{aligned}
\forall \ell_1 \neq \ell_2, \qquad \left| D^{\ell_1,\ell_2}(\mathbf{x}) \right| &= \left| \mathbf{Cov}_{\mathbf{x}} \left[\mathbb{1}_{[g_{\ell_1}=+1]}, \mathbb{1}_{[g_{\ell_2}=+1]} \right] \right| \\
&= \left| \mathbf{x} \left[\mathbb{1}_{[g_{\ell_1}=+1, g_{\ell_2}=+1]} \right] - \pi(\mathbf{x}) \left[\mathbb{1}_{[g_{\ell_1}=+1, g_{\ell_2}=+1]} \right] \right| \\
&\leq \left| \left| \mathbf{x}^{\{\ell_1,\ell_2\}} - \pi^{\{\ell_1,\ell_2\}}(\mathbf{x}) \right| \right|_2
\end{aligned}$$

For the second inequality, it is enough to prove the result when $W(g) = Z(g)$ or $W(g) = Z(g)^2$, and the general case will follow by linearity. We show this for $W(g) = Z(g)^2$. In this case we write from Proposition 2.3.3

$$\begin{aligned}
S^{\ell_0}(\mathbf{x}) =& \mathbf{Cov}_{\mathbf{x}} \left[Z(g)^2, \mathbb{1}_{[g_{\ell_0}=+1]} \right] \\
=& \frac{1}{L^2} \sum_{\ell_1,\ell_2 \in [L]} \mathbf{Cov}_{\mathbf{x}} \left[g_{\ell_1} g_{\ell_2}, \mathbb{1}_{[g_{\ell_0}=+1]} \right] \\
=& \frac{1}{L^2} \sum_{\ell_1,\ell_2 \in [L]} \mathbf{x} \left[g_{\ell_1} g_{\ell_2} \mathbb{1}_{[g_{\ell_0}=+1]} \right] - \mathbf{x} \left[g_{\ell_1} g_{\ell_2} \right] p^{\ell_0}(\mathbf{x})
\end{aligned}$$

The analog holds for $S^{\ell_0}(\pi(\mathbf{x}))$. It follows

$$\begin{aligned}
\left| S^{\ell_0}(\mathbf{x}) - S^{\ell_0}(\pi(\mathbf{x})) \right| \leq \frac{1}{L^2} \sum_{\ell_1,\ell_2 \in [L]} \Bigg(& \left| \mathbf{x} \left[g_{\ell_1} g_{\ell_2} \mathbb{1}_{[g_{\ell_0}=+1]} \right] - \pi(\mathbf{x}) \left[g_{\ell_1} g_{\ell_2} \mathbb{1}_{[g_{\ell_0}=+1]} \right] \right| \\
& + p^{\ell_0}(\mathbf{x}) \times \left| \mathbf{x}[g_{\ell_1} g_{\ell_2}] - \pi(\mathbf{x})[g_{\ell_1} g_{\ell_2}] \right| \Bigg)
\end{aligned}$$

We may remove the summand corresponding to $\ell_1 = \ell_2 = \ell_0$, because $\mathbf{x}$ and $\pi(\mathbf{x})$ have the same marginals on ℓ_0. We conclude by rewriting this

$$\left| S^{\ell_0}(\mathbf{x}) - S^{\ell_0}(\pi(\mathbf{x})) \right| \leq \frac{2}{L^2} \sum_{\substack{A \subset [L] \setminus \{\ell_0\} \\ 1 \leq \# A \leq 2}} \left| \left| \mathbf{x}^{\{\ell_0\} \cup A} - \pi(\mathbf{x}^{\{\ell_0\} \cup A}) \right| \right|_1$$

Similar calculations when $W(g) = Z(g)$ yield

$$\left| S^{\ell_0}(\mathbf{x}) - S^{\ell_0}(\pi(\mathbf{x})) \right| \leq \frac{2}{L} \sum_{\ell_1 \in [L] \setminus \{\ell_0\}} \left| \left| \mathbf{x}^{\{\ell_0,\ell_1\}} - \pi(\mathbf{x}^{\{\ell_0,\ell_1\}}) \right| \right|_1$$

We conclude using the equivalence of the L^1 and L^2 norms in $\mathbb{R}^{\square_A}$. □

2.3.3 Eigenvalues of the linearized recombinator

The goal of this section is to obtain some properties of the jacobian of the recombinator $\nabla \hat{R}^A$, which will allow us to find a lower bound for the contribution of recombination to the dynamics of $\mathbf{X}_t^A$. Because $\hat{R}^A$ is the analog of R, we will simplify our proofs without loss of generality by assuming $A = [L]$. The following Lemma motivates the study of the jacobian of the recombinator by relating it to the recombinator itself.

Lemma 2.3.7.

$$\forall \mathbf{x} \in \mathbb{X}^{[L]}, \qquad \nabla R(\mathbf{x})(\mathbf{x} - \pi(\mathbf{x})) - R(\mathbf{x}) = \sum_{\emptyset \subsetneq \mathcal{I} \subsetneq [L]} \nu(\mathcal{I})(\mathbf{x} - \pi(\mathbf{x}))^{\mathcal{I}} \otimes (\mathbf{x} - \pi(\mathbf{x}))^{\mathcal{I}^c}$$

In particular, if $\#A = 3$ then

$$\forall \mathbf{x} \in \mathbb{X}^{A}, \qquad \hat{R}^A(\mathbf{x}) = \nabla \hat{R}^A(\mathbf{x})(\mathbf{x} - \pi(\mathbf{x})) \tag{2.40}$$

Proof. A simple computation shows that for any $\mathbf{h} \in \mathbb{R}^{\square_{[L]}}$

$$\nabla R(\mathbf{x})\mathbf{h} = \sum_{\emptyset \subsetneq \mathcal{I} \subsetneq [L]} \nu(\mathcal{I})(\mathbf{x}^{\mathcal{I}} \otimes \mathbf{h}^{\mathcal{I}^c} + \mathbf{h}^{\mathcal{I}} \otimes \mathbf{x}^{\mathcal{I}^c} - \mathbf{h}) \tag{2.41}$$

And thus

$$\nabla R(\mathbf{x})\mathbf{h} - R(\mathbf{x}) = \sum_{\emptyset \subsetneq \mathcal{I} \subsetneq [L]} \nu(\mathcal{I}) \left(\mathbf{x}^{\mathcal{I}} \otimes \mathbf{h}^{\mathcal{I}^c} + (\mathbf{h} - \mathbf{x})^{\mathcal{I}} \otimes \mathbf{x}^{\mathcal{I}^c} - (\mathbf{h} - \mathbf{x})\right)$$

Applying this to $\mathbf{h} = \mathbf{x} - \pi(\mathbf{x})$ yields

$$\begin{aligned}
\nabla R(\mathbf{x})(\mathbf{x} - \pi(\mathbf{x})) - R(\mathbf{x}) &= \sum_{\emptyset \subsetneq \mathcal{I} \subsetneq [L]} \nu(\mathcal{I})(\mathbf{x}^{\mathcal{I}} \otimes (\mathbf{x} - \pi(\mathbf{x}))^{\mathcal{I}^c} - \pi(\mathbf{x})^{\mathcal{I}} \otimes \mathbf{x}^{\mathcal{I}^c} + \pi(\mathbf{x})) \\
&= \sum_{\emptyset \subsetneq \mathcal{I} \subsetneq [L]} \nu(\mathcal{I})(\mathbf{x}^{\mathcal{I}} \otimes (\mathbf{x} - \pi(\mathbf{x}))^{\mathcal{I}^c} - \pi(\mathbf{x})^{\mathcal{I}} \otimes (\mathbf{x} - \pi(\mathbf{x}))^{\mathcal{I}^c}) \\
&= \sum_{\emptyset \subsetneq \mathcal{I} \subsetneq [L]} \nu(\mathcal{I})(\mathbf{x} - \pi(\mathbf{x}))^{\mathcal{I}} \otimes (\mathbf{x} - \pi(\mathbf{x}))^{\mathcal{I}^c}
\end{aligned}$$

where in the second equality we used $\pi(\mathbf{x}) = \pi(\mathbf{x})^{\mathcal{I}} \otimes \pi(\mathbf{x})^{\mathcal{I}^c}$.

Because $\hat{R}^A$ is the analog of R on $\mathbb{X}^A$, we obtain

$$\forall \mathbf{x} \in \mathbb{X}^{A}, \qquad \nabla \hat{R}^A(\mathbf{x})(\mathbf{x} - \pi(\mathbf{x})) - \hat{R}^A(\mathbf{x}) = \sum_{\emptyset \subsetneq \mathcal{I} \subsetneq A} \nu^A(\mathcal{I})(\mathbf{x} - \pi(\mathbf{x}))^{\mathcal{I}} \otimes (\mathbf{x} - \pi(\mathbf{x}))^{A \setminus \mathcal{I}} \tag{2.42}$$

Recall from (2.26) that for any $\ell \in A$,

$$\mathbf{x}^{\{\ell\}} = \pi(\mathbf{x})^{\{\ell\}}$$

It follows that whenever $\#\mathcal{I} \in \{1, 2\}$, we necessarily have

$$(\mathbf{x} - \pi(\mathbf{x}))^{\mathcal{I}} \otimes (\mathbf{x} - \pi(\mathbf{x}))^{A \setminus \mathcal{I}} = 0$$

This yields (2.40). □

Let us define the following quantities

$$\forall \mathcal{I} \subseteq [L], \mathcal{I} \neq \emptyset, \qquad \beta_{\mathcal{I}} := \sum_{\emptyset \subsetneq \mathcal{K} \subsetneq \mathcal{I}} \nu^{\mathcal{I}}(\mathcal{K}) = 1 - \nu^{\mathcal{I}}(\mathcal{I}) - \nu^{\mathcal{I}}(\emptyset) = 1 - 2\nu^{\mathcal{I}}(\mathcal{I}) \tag{2.43}$$

We call $\beta_{\mathcal{I}}$ the probability that there is a recombination within $\mathcal{I}$. Also define $\beta_\emptyset := -\beta_{[L]}$. We note the following order property

$$\forall \mathcal{J} \subseteq \mathcal{I} \subseteq [L], \qquad \beta_{\mathcal{I}} \geq \beta_{\mathcal{J}}. \tag{2.44}$$

For $\mathcal{I} \subseteq [L]$, we define the **$\mathcal{I}$-linkage vector** with

$$\mathbf{w}_{\mathcal{I}} := \left(2^{-\frac{L}{2}} \prod_{\ell \in \mathcal{I}} \gamma_\ell \right)_{\gamma \in \Box_{[L]}} \in \mathbb{R}^{\Box_{[L]}} \tag{2.45}$$

We start with computational properties of the linkage vectors.

Lemma 2.3.8. *The* $(\mathbf{w}_{\mathcal{I}})_{\mathcal{I} \subseteq [L]}$ *form an orthonormal basis of* $\mathbb{R}^{\Box_{[L]}}$ *for the usual scalar product, which will be denoted* $\langle \cdot, \cdot \rangle$. *Furthermore, we have*

$$\forall \mathcal{J} \subseteq [L], \qquad \mathbf{w}_{\mathcal{I}} = 2^{-\frac{L}{2}} \mathbf{w}^{\mathcal{J}}_{\mathcal{I} \cap \mathcal{J}} \otimes \mathbf{w}^{\mathcal{J}^c}_{\mathcal{I} \cap \mathcal{J}^c} \tag{2.46}$$

$$\forall \mathcal{J}, \mathcal{I} \subset [L], \forall \mathbf{x} \in \mathbb{R}^{\Box_{[L]}}, \qquad \langle \mathbf{w}^{\mathcal{J}}_{\mathcal{I}}, \mathbf{x}^{\mathcal{J}} \rangle = 2^{L - \#\mathcal{J}} \langle \mathbf{w}_{\mathcal{I}}, \mathbf{x} \rangle \, \mathbb{1}_{[\mathcal{I} \subseteq \mathcal{J}]} \tag{2.47}$$

Proof. We compute

$$\begin{aligned} \langle \mathbf{w}_{\mathcal{I}}, \mathbf{w}_{\mathcal{J}} \rangle &= 2^{-L} \sum_{\gamma \in \Box_{[L]}} \prod_{\ell \in \mathcal{I}} \gamma_\ell \prod_{\ell \in \mathcal{J}} \gamma_\ell \\ &= 2^{-L} \sum_{\gamma \in \Box_{[L]}} \prod_{\ell \in \mathcal{I} \oplus \mathcal{J}} \gamma_\ell \end{aligned}$$

where we write $\mathcal{I} \oplus \mathcal{J} := (\mathcal{I} \cup \mathcal{J}) \setminus (\mathcal{I} \cap \mathcal{J})$. If $\mathcal{I} = \mathcal{J}$, this yields 1. Otherwise, we can find $\ell_0 \in \mathcal{I} \oplus \mathcal{J}$. Then we can cancel out the term γ of the sum with the term $\gamma^{-\ell_0}$, which we define to be γ with the ℓ_0-th coordinate flipped. This yields that $(\mathbf{w}_{\mathcal{I}})_{\mathcal{I} \subseteq [L]}$ is an orthonormal basis.

Let us observe that whenever $\gamma \in \Box_{\mathcal{J}}, \mathcal{I} \subseteq \mathcal{J}$ we have

$$\mathbf{w}^{\mathcal{J}}_{\mathcal{I}}(\gamma) = \sum_{\substack{\hat{\gamma} \in \Box_{[L]} \\ \hat{\gamma}^{|\mathcal{J}} = \gamma}} 2^{-\frac{L}{2}} \prod_{\ell \in \mathcal{I}} \gamma_\ell = 2^{-\frac{L}{2}} \times 2^{L - \#\mathcal{J}} \prod_{\ell \in \mathcal{I}} \gamma_\ell$$

We then get (2.46) with

$$\mathbf{w}^{\mathcal{J}}_{\mathcal{I} \cap \mathcal{J}} \otimes \mathbf{w}^{\mathcal{J}^c}_{\mathcal{I} \cap \mathcal{J}^c}(\gamma) \;=\; 2^{\frac{L}{2} - \#\mathcal{J}} \prod_{\ell \in \mathcal{I} \cap \mathcal{J}} \gamma_\ell \;\times\; 2^{\frac{L}{2} - \#\mathcal{J}^c} \prod_{\ell \in \mathcal{I} \cap \mathcal{J}^c} \gamma_\ell \;=\; 2^{\frac{L}{2}} \mathbf{w}_{\mathcal{I}}$$

To get (2.47) when $\mathcal{I} \subseteq \mathcal{J}$ we write

$$\left\langle \mathbf{w}^{\mathcal{J}}_{\mathcal{I}} \;,\; \mathbf{x}^{\mathcal{J}} \right\rangle = 2^{\frac{L}{2} - \#\mathcal{J}} \sum_{\gamma \in \Box_{[L]}} \left(\prod_{\ell \in \mathcal{I}} \gamma_\ell \right) x(\gamma) = 2^{L - \#\mathcal{J}} \left\langle \mathbf{w}_{\mathcal{I}} \;,\; \mathbf{x} \right\rangle$$

When $\mathcal{I} \nsubseteq \mathcal{J}$, then we can find $\ell_0 \in \mathcal{I} \setminus \mathcal{J}$. We then write

$$\mathbf{w}^{\mathcal{J}}_{\mathcal{I}}(\gamma) = \sum_{\substack{\hat{\gamma} \in \Box_{[L]} \\ \hat{\gamma}^{|\mathcal{J}} = \gamma}} 2^{-\frac{L}{2}} \prod_{\ell \in \mathcal{I}} \gamma_\ell = 0$$

where we cancelled out the $\hat{\gamma}$ term of the sum with the corresponding term with the ℓ_0-th coordinate flipped. $\Box$

We can now compute ∇R explicitly in the next Theorem.

Theorem 2.3.9. *We have for any $\mathcal{I}, \mathcal{J} \subseteq [L]$.*

$$\langle \mathbf{w}_{\mathcal{I}}, \nabla R(\mathbf{x}) \mathbf{w}_{\mathcal{J}} \rangle = -\mathbb{1}_{[\mathcal{I}=\mathcal{J}]} \beta_{\mathcal{I}} + \mathbb{1}_{[\mathcal{J} \subsetneq \mathcal{I}]} \sum_{\emptyset \subsetneq \mathcal{K} \subsetneq [L]} \nu(\mathcal{K}) 2^{1+\frac{L}{2}} \langle \mathbf{w}_{\mathcal{I} \cap \mathcal{K}} \, , \, \mathbf{x} \rangle \, \mathbb{1}_{[\mathcal{I} \cap \mathcal{K}^c = \mathcal{J}]}$$

In particular, we have

$$\forall \mathcal{J} \nsubseteq \mathcal{I}, \qquad \langle \mathbf{w}_{\mathcal{I}}, \nabla R(\mathbf{x}) \mathbf{w}_{\mathcal{J}} \rangle = 0 \tag{2.48}$$

Remark 11. *A well-known method to handle the recombinator is Haldane linearization [60]. This method relies on considering linkage, not on subsets $\mathcal{I} \subseteq A$, but rather on partitions of A. This approach may prove necessary if we want to control linkage on subsets A with arbitrary size, but will not be needed here.*

Proof. According to (2.41) (replacing $[L]$ by the set A),

$$\begin{aligned}
\langle \mathbf{w}_{\mathcal{I}}, \nabla R(\mathbf{x}) \mathbf{w}_{\mathcal{J}} \rangle &= \sum_{\emptyset \subsetneq \mathcal{K} \subsetneq [L]} \nu(\mathcal{K}) \left\langle \mathbf{w}_{\mathcal{I}} \, , \, \mathbf{x}^{\mathcal{K}} \otimes \mathbf{w}_{\mathcal{J}}^{\mathcal{K}^c} + \mathbf{w}_{\mathcal{J}}^{\mathcal{K}} \otimes \mathbf{x}^{\mathcal{K}^c} - \mathbf{w}_{\mathcal{J}} \right\rangle \\
&= \sum_{\emptyset \subsetneq \mathcal{K} \subsetneq [L]} \nu(\mathcal{K}) \left(\left\langle \mathbf{w}_{\mathcal{I}} \, , \, \mathbf{x}^{\mathcal{K}} \otimes \mathbf{w}_{\mathcal{J}}^{\mathcal{K}^c} \right\rangle + \left\langle \mathbf{w}_{\mathcal{I}} \, , \, \mathbf{w}_{\mathcal{J}}^{\mathcal{K}} \otimes \mathbf{x}^{\mathcal{K}^c} \right\rangle - \mathbb{1}_{[\mathcal{I}=\mathcal{J}]} \right)
\end{aligned}$$

Using $\nu(\mathcal{K}) = \nu(\mathcal{K}^c)$, we can rewrite this

$$\langle \mathbf{w}_{\mathcal{I}}, \nabla R(\mathbf{x}) \mathbf{w}_{\mathcal{J}} \rangle = \sum_{\emptyset \subsetneq \mathcal{K} \subsetneq [L]} \nu(\mathcal{K}) \left(2 \left\langle \mathbf{w}_{\mathcal{I}} \, , \, \mathbf{x}^{\mathcal{K}} \otimes \mathbf{w}_{\mathcal{J}}^{\mathcal{K}^c} \right\rangle - \mathbb{1}_{[\mathcal{I}=\mathcal{J}]} \right) \tag{2.49}$$

Using (2.46) we get

$$\begin{aligned}
\left\langle \mathbf{w}_{\mathcal{I}} \, , \, \mathbf{x}^{\mathcal{K}} \otimes \mathbf{w}_{\mathcal{J}}^{\mathcal{K}^c} \right\rangle &= 2^{-\frac{L}{2}} \left\langle \mathbf{w}_{\mathcal{I} \cap \mathcal{K}}^{\mathcal{K}} \, , \, \mathbf{x}^{\mathcal{K}} \right\rangle \left\langle \mathbf{w}_{\mathcal{I} \cap \mathcal{K}^c}^{\mathcal{K}^c}, \mathbf{w}_{\mathcal{J}}^{\mathcal{K}^c} \right\rangle \\
&= 2^{\frac{L}{2}} \langle \mathbf{w}_{\mathcal{I} \cap \mathcal{K}} \, , \, \mathbf{x} \rangle \, \mathbb{1}_{[\mathcal{I} \cap \mathcal{K}^c = \mathcal{J}]}
\end{aligned}$$

where in the second equality we used (2.47) twice. We thus obtain from (2.49)

$$\begin{aligned}
\langle \mathbf{w}_{\mathcal{I}}, \nabla R(\mathbf{x}) \mathbf{w}_{\mathcal{J}} \rangle &= \sum_{\emptyset \subsetneq \mathcal{K} \subsetneq [L]} \nu(\mathcal{K}) \left(2^{1+\frac{L}{2}} \langle \mathbf{w}_{\mathcal{I} \cap \mathcal{K}} \, , \, \mathbf{x} \rangle \, \mathbb{1}_{[\mathcal{I} \cap \mathcal{K}^c = \mathcal{J}]} - \mathbb{1}_{[\mathcal{I}=\mathcal{J}]} \right) \\
&= \sum_{\emptyset \subsetneq \mathcal{K} \subsetneq [L]} \nu(\mathcal{K}) 2^{1+\frac{L}{2}} \langle \mathbf{w}_{\mathcal{I} \cap \mathcal{K}} \, , \, \mathbf{x} \rangle \, \mathbb{1}_{[\mathcal{I} \cap \mathcal{K}^c = \mathcal{J}]} \quad - \quad \mathbb{1}_{[\mathcal{I}=\mathcal{J}]} \beta_{[L]}
\end{aligned} \tag{2.50}$$

where in the last equality we used

$$\sum_{\emptyset \subsetneq \mathcal{K} \subsetneq [L]} \nu(\mathcal{K}) = \beta_{[L]}$$

Notice that the sum in (2.50) can only be nonzero if $\mathcal{J} \subseteq \mathcal{I}$. When $\mathcal{J} \neq \mathcal{I}$, we get the result. When $\mathcal{I} = \mathcal{J}$ we write

$$\begin{aligned}
\langle \mathbf{w}_{\mathcal{I}}, \nabla R(\mathbf{x}) \mathbf{w}_{\mathcal{I}} \rangle &= \sum_{\emptyset \subsetneq \mathcal{K} \subsetneq [L]} \nu(\mathcal{K}) 2^{1+\frac{L}{2}} \langle \mathbf{w}_{\emptyset} \, , \, \mathbf{x} \rangle \, \mathbb{1}_{[\mathcal{I} \cap \mathcal{K}^c = \mathcal{I}]} \quad - \quad \beta_{[L]} \\
&= \sum_{\emptyset \subsetneq \mathcal{K} \subsetneq [L]} 2 \nu(\mathcal{K}) \mathbb{1}_{[\mathcal{I} \subseteq \mathcal{K}^c]} \quad - \quad \beta_{[L]}
\end{aligned}$$

where we used $\langle \mathbf{w}_{\emptyset}, \mathbf{x} \rangle = 2^{-\frac{L}{2}}$. If $\mathcal{I} = \emptyset$, the sum is equal to $2\beta_{[L]}$ and we get

$$\langle \mathbf{w}_{\emptyset}, \nabla R(\mathbf{x}) \mathbf{w}_{\emptyset} \rangle = \beta_{[L]} = -\beta_{\emptyset}$$

If $\mathcal{I} \neq \emptyset$, we can extend the sum to $\mathcal{K} \in \{\emptyset, [L]\}$ and write

$$\begin{aligned}\sum_{\emptyset \subsetneq \mathcal{K} \subsetneq [L]} 2\nu(\mathcal{K}) \mathbb{1}_{[\mathcal{I} \subseteq \mathcal{K}^c]} = & \sum_{\emptyset \subseteq \mathcal{K} \subseteq [L]} 2\nu(\mathcal{K}) \mathbb{1}_{[\mathcal{I} \subseteq \mathcal{K}^c]} \quad - \; 2\nu(\emptyset) \\ =& 2\nu^{\mathcal{I}}(\mathcal{I}) - (1 - \beta_{[L]})\end{aligned}$$

It follows

$$\begin{aligned}\langle \mathbf{w}_{\mathcal{I}}, \nabla R(\mathbf{x}) \mathbf{w}_{\mathcal{I}} \rangle =& 2\nu^{\mathcal{I}}(\mathcal{I}) - (1 - \beta_{[L]}) - \beta_{[L]} \\ =& 2\nu^{\mathcal{I}}(\mathcal{I}) - 1 \\ =& - \beta_{\mathcal{I}}\end{aligned}$$

□

2.3.4 Controlling linkage disequilibrium over a small subset

For any $A \subseteq [L]$ with $\#A \geq 2$, recall from (2.33)

$$r_A := \min_{\substack{\ell_1, \ell_2 \in A \\ \ell_1 \neq \ell_2}} r_{\{\ell_1, \ell_2\}}$$

where $r_{\{\ell_1,\ell_2\}}$ is the probability of a recombination event between ℓ_1 and ℓ_2, and in particular from (2.43) $r_{\{\ell_1,\ell_2\}} = \beta_{\{\ell_1,\ell_2\}}$. From (2.44) we have for any $\mathcal{J} \subseteq A$

$$r_A \leq \beta_{\mathcal{J}} \leq 1. \tag{2.51}$$

The goal of this subsection is to prove

Proposition 2.3.10. *Let $T > 0$. Assume that $\eta := \frac{1}{\sqrt{\rho r_A}} \leq 1$ and $\rho \geq e$. There exists a constant C independent of (A, L) such that for every $\varepsilon \in [0, T]$,*

$$\forall A \subseteq [L], \#A \leq 3, \qquad \mathbb{E}\left[\sup_{t \in [\varepsilon, T]} ||\mathbf{X}_t^A - \pi(\mathbf{X}_t^A)||_2\right] \leq C \left(\frac{1}{\rho r_A \varepsilon} + \frac{L}{\rho r_A} + \sqrt{\frac{\ln(\rho)}{\rho r_A}}\right).$$

We will prove the result for $\#A = 3$. We start with two Lemmas.

Lemma 2.3.11. *Consider $A \subset [L]$ with $\#A \leq 3$. Then*

$$\forall \mathbf{x} \in \mathbb{X}^A, \qquad (Id - \nabla\pi(\mathbf{x}))\, \hat{R}^A(\mathbf{x}) = \nabla \hat{R}^A(\mathbf{x})(\mathbf{x} - \pi(\mathbf{x})) \tag{2.52}$$

where we abusively write $\nabla\pi$ for the $\Box_A \times \Box_A$ jacobian of π, seen as a $\mathbb{R}^{\Box_A}-$valued function on $\mathbb{X}^A$.

Proof. Because of (2.40) in Lemma 2.3.7, we only need to show

$$\forall \mathbf{x} \in \mathbb{X}^A, \qquad \nabla\pi(\mathbf{x}) \hat{R}^A(\mathbf{x}) = 0 \tag{2.53}$$

Fix $\mathbf{x}_0 \in \mathbb{X}^A$. We define $(\mathbf{x}_t)_{t \geq 0}$ as the solution to

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{x}_t = \hat{R}^A(\mathbf{x}_t)$$

with initial condition $\mathbf{x}_0$. By Lemma 2.3.1, we have

$$\forall t \geq 0, \qquad \pi(\mathbf{x}_t) = \pi(\mathbf{x}_0).$$

Taking the time derivative of $\pi(\mathbf{x}_t)$, we get

$$\mathbf{0} = \frac{\mathrm{d}}{\mathrm{d}t} \pi(\mathbf{x}_t) = \nabla\pi(\mathbf{x}_t) \hat{R}^A(\mathbf{x}_t)$$

Evaluating the right-hand side at $t = 0$ gives (2.53). □

Lemma 2.3.12. *Consider a continuous martingale* $(M_t)_{t\geq 0}$ *with quadratic variation uniformly bounded by* $C_0 > 0$. *Then we can find a universal constant* $C(C_0)$ *such that for any* $\lambda \geq 1, T > 0$

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t e^{-\lambda(t-u)}\mathrm{d}M_u\right|\right] \leq C\sqrt{\frac{1+\ln_+(2\lambda T)}{\lambda}}$$

with $\ln_+(x) = (\ln(x)) \vee 0$.

Proof. Let us write

$$Z_t := \int_0^t e^{\lambda u}\mathrm{d}M_u \qquad ; \qquad Y_t := \int_0^t e^{-\lambda(t-u)}\mathrm{d}M_u$$

such that

$$Y_t = e^{-\lambda t} Z_t \tag{2.54}$$

The martingale $(Z_t)_{t\in[0,T]}$ has quadratic variation

$$\langle Z\rangle_t^{QV} = \int_0^t e^{2\lambda u}\mathrm{d}\,\langle M\rangle_u^{QV} \leq C_0\frac{e^{2\lambda t}}{2\lambda}$$

Let us define the iterated logarithm

$$\ln_{(2)}(x) := \ln_+(\ln_+(x)).$$

Since $\lambda \geq 1$

$$\ln_{(2)}\left(\frac{e^{2\lambda t}}{2\lambda}\right) \leq \ln_+(2\lambda T).$$

Therefore

$$F(t) := \langle Z\rangle_t^{QV}\left(1+\ln_{(2)}(\langle Z\rangle_t^{QV})\right) \leq C_0\frac{e^{2\lambda t}}{2\lambda}\left(1+\ln_{(2)}(C_0)+\ln_+(2\lambda T)\right)$$

It follows from (2.54)

$$\begin{aligned}\sup_{t\in[0,T]}|Y_t| &= \sup_{t\in[0,T]} e^{-\lambda t}|Z_t| \\ &= \sup_{t\in[0,T]} e^{-\lambda t}\sqrt{F(t)}\times\frac{|Z_t|}{\sqrt{F(t)}} \\ &\leq \sqrt{C_0}\sqrt{\frac{1+\ln_+(C_0)+\ln_+(2\lambda T)}{2\lambda}}\sup_{t\in[0,T]}\frac{|Z_t|}{\sqrt{F(t)}}\end{aligned}$$

We show in Appendix B the existence of a universal constant $C > 0$ (only depending on C_0) such that

$$\mathbb{E}\left[\sup_{t\in[0,T]}\frac{Z_t}{\sqrt{F(t)}}\right] \leq C.$$

This completes the proof of the lemma. □

Proof of Proposition 2.3.10. For $\mathcal{I} \subseteq A$, define by analogy with (2.45)

$$\hat{\mathbf{w}}_{\mathcal{I}}^{A} = \left(2^{-\#A/2}\prod_{\ell\in A}\gamma_\ell\right)_{\gamma\in\square_A}$$

One may check that $\hat{\mathbf{w}}^A_{\mathcal{I}} = 2^{(L-\#A)/2}\mathbf{w}^A_{\mathcal{I}}$. By the analog of Lemma 2.3.8, $(\hat{\mathbf{w}}^A_{\mathcal{I}})_{\mathcal{I}\subseteq A}$ is an orthonormal basis of $\mathbb{R}^{\square_A}$ and we have for any $\mathbf{y} \in \mathbb{R}^{\square_A}$

$$||\mathbf{y}||_2^2 = \sum_{\mathcal{I}\subseteq A} \langle \hat{\mathbf{w}}^A_{\mathcal{I}}, \mathbf{y}\rangle^2 .$$

Therefore, to get the result we only need to control, for all $\mathcal{I} \subseteq A$,

$$Y^{\mathcal{I}}_t := \langle \hat{\mathbf{w}}^A_{\mathcal{I}}, \mathbf{X}^A_t - \pi(\mathbf{X}^A_t)\rangle .$$

Note that because $\hat{\mathbf{w}}^A_{\mathcal{I}}$ and $\mathbf{X}^A_t - \pi(\mathbf{X}^A_t)$ have coefficients bounded by 1, and $\mathbb{R}^A$ has dimension at most 2^3, then $\langle \hat{\mathbf{w}}^A_{\mathcal{I}}, \mathbf{X}^A_t\rangle$ and $Y^{\mathcal{I}}_t$ are uniformly bounded by a constant C independent of (L, A).

Recall from the definition of π in (2.26) that for any $\mathbf{x} \in \mathbb{X}^A$

$$Y^{\emptyset}_t = \langle \hat{\mathbf{w}}^A_{\emptyset}, \mathbf{x}_t - \pi(\mathbf{x}_t)\rangle = 0 \qquad ; \qquad \forall \ell \in A, \qquad Y^{\{\ell\}}_t = \langle \hat{\mathbf{w}}^A_{\{\ell\}}, \mathbf{x}_t - \pi(\mathbf{x}_t)\rangle = 0 \tag{2.55}$$

It remains to consider $\#\mathcal{I} \in \{2, 3\}$. We apply Itô's formula to write

$$\mathrm{d}\left(\mathbf{X}^A_u - \pi(\mathbf{X}^A_u)\right) = \left(\mathrm{Id} - \nabla\pi(\mathbf{X}^A_u)\right) \mathrm{d}\mathbf{X}^A_u - \sum_{\gamma^1,\gamma^2\in\square_A} \partial_{\gamma^1,\gamma^2}\pi(\mathbf{X}^A_u)\, \mathrm{d}\left\langle X^A(\gamma^1), X^A(\gamma^2)\right\rangle^{QV}_u$$

Using Proposition 2.3.3 this can be rewritten

$$\mathrm{d}\left(\mathbf{X}^A_u - \pi(\mathbf{X}^A_u)\right) = \rho\left(\mathrm{Id} - \nabla\pi(\mathbf{X}^A_u)\right) \hat{R}^A(\mathbf{X}^A_u)\mathrm{d}t + \mathbf{F}_u\mathrm{d}u + \mathrm{d}\mathbf{M}_u$$

where

$$\begin{aligned}\mathbf{F}_t :=& \left(\mathrm{Id} - \nabla\pi(\mathbf{X}^A_t)\right)\left(\hat{\Theta}^A(\mathbf{X}^A_t) + LS^A(\mathbf{X}_t)\right) \\ &- \sum_{\gamma^1,\gamma^2\in\square_A} \partial_{\gamma^1,\gamma^2}\pi(\mathbf{X}^A_t)\frac{\mathrm{d}}{\mathrm{d}t}\left\langle X^A(\gamma^1), X^A(\gamma^2)\right\rangle^{QV}_t \\ \mathrm{d}\mathbf{M}_t :=& \left(\mathrm{Id} - \nabla\pi(\mathbf{X}^A_t)\right)\hat{\Sigma}^A(\mathbf{X}^A_t)\mathrm{d}\hat{\mathbf{B}}^A_t.\end{aligned}$$

In particular, $\mathbf{F}_t$ has coefficients uniformly bounded by CL and $\mathbf{M}_t$ is a continuous martinale with quadratic variation uniformly smaller than C for some constant $C > 0$ independent of (A, L). Finally, Lemma 2.3.11 yields for $\mathcal{I} \subseteq A$

$$\mathrm{d}Y^{\mathcal{I}}_u = \rho\left\langle \hat{\mathbf{w}}^A_{\mathcal{I}}, \nabla\hat{R}^A(\mathbf{X}^A_u)\left(\mathbf{X}^A_u - \pi(\mathbf{X}^A_u)\right)\right\rangle \mathrm{d}u + \left\langle \hat{\mathbf{w}}^A_{\mathcal{I}}, \mathbf{F}_u\right\rangle \mathrm{d}u + \left\langle \hat{\mathbf{w}}^A_{\mathcal{I}}, \mathrm{d}\mathbf{M}_u\right\rangle \tag{2.56}$$

Case $\#\mathcal{I} = 2$:
Theorem 2.3.9 and (2.55) imply

$$\left\langle \hat{\mathbf{w}}^A_{\mathcal{I}}, \nabla\hat{R}^A(\mathbf{X}^A_u)\left(\mathbf{X}^A_u - \pi(\mathbf{X}^A_u)\right)\right\rangle = -\beta_{\mathcal{I}}\left\langle \hat{\mathbf{w}}^A_{\mathcal{I}},\ \mathbf{X}^A_u - \pi(\mathbf{X}^A_u)\right\rangle = -\beta_{\mathcal{I}}Y^{\mathcal{I}}_u$$

Therefore (2.56) becomes

$$\mathrm{d}Y^{\mathcal{I}}_u = -\rho\beta_{\mathcal{I}}Y^{\mathcal{I}}_u\mathrm{d}u + \left\langle \hat{\mathbf{w}}^A_{\mathcal{I}}, \mathbf{F}_u\right\rangle \mathrm{d}u + \left\langle \hat{\mathbf{w}}^A_{\mathcal{I}}, \mathrm{d}\mathbf{M}_u\right\rangle$$

It can be checked that this implies

$$Y^{\mathcal{I}}_t = Y^{\mathcal{I}}_0 e^{-\rho\beta_{\mathcal{I}}t} + \int_0^t e^{-\rho\beta_{\mathcal{I}}(t-u)}\left\langle \hat{\mathbf{w}}^A_{\mathcal{I}}, \mathbf{F}_u\right\rangle \mathrm{d}u + \int_0^t e^{-\rho\beta_{\mathcal{I}}(t-u)}\left\langle \hat{\mathbf{w}}^A_{\mathcal{I}}, \mathrm{d}\mathbf{M}_u\right\rangle \tag{2.57}$$

The fact that $Y_0^{\mathcal{I}}$ is uniformly bounded means the first term is smaller than $Ce^{-\rho\beta_{\mathcal{I}}\varepsilon}$ for any $t \in [\varepsilon, T]$. We use $e^{-x} \leq \frac{1}{x}$ for $x \geq 0$ to write

$$Y_0^{\mathcal{I}} e^{-\rho\beta_{\mathcal{I}} t} \leq \frac{C}{\rho\beta_{\mathcal{I}}\varepsilon} \leq \frac{C}{\rho r_A \varepsilon}$$

from (2.51). Because $\mathbf{F}_t$ is smaller than CL, the first integral is smaller than $C\frac{L}{\rho\beta_{\mathcal{I}}}$. For the second integral, we use Lemma 2.3.12 applied with $\lambda = \rho\beta_{\mathcal{I}}$. Because $\rho\beta_{\mathcal{I}} \geq \rho r_A \geq 1$ by assumption, this Lemma yields a a bound on the second integral of

$$C\sqrt{\frac{1 + \ln_+(\rho\beta_{\mathcal{I}} 2T)}{\rho\beta_{\mathcal{I}}}} \leq C\sqrt{\frac{\ln(\rho)}{\rho r_A}}.$$

using (2.51) and $\rho \geq e$, for some constant C independent of (ε, A, L). We thus obtain from (2.57)

$$\mathbb{E}\left[\sup_{t\in[\varepsilon,T]} |Y_t^{\mathcal{I}}|\right] \leq C\left(\frac{1}{\rho r_A \varepsilon} + \frac{L}{\rho r_A} + \sqrt{\frac{\ln(\rho)}{\rho r_A}}\right)$$

Case $\#\mathcal{I} = 3$:
When $\mathcal{I} = A$, Theorem 2.3.9 and (2.55) imply

$$\begin{aligned}\left\langle \hat{\mathbf{w}}_A^A, \nabla \hat{R}^A(\mathbf{X}_u^A)\left(\mathbf{X}_u^A - \pi(\mathbf{X}_u^A)\right)\right\rangle &= -\beta_A \left\langle \hat{\mathbf{w}}_A^A,\ \mathbf{X}_u^A - \pi(\mathbf{X}_u^A)\right\rangle \\ &\quad + \sum_{\substack{\mathcal{J}\subsetneq A \\ \#\mathcal{J}=2}} \nu^A(\mathcal{J}) 2^{1+\frac{\#A}{2}} \left\langle \hat{\mathbf{w}}_{A\smallsetminus\mathcal{J}}^A,\ \mathbf{X}_u^A\right\rangle \left\langle \hat{\mathbf{w}}_{\mathcal{J}}^A,\ \mathbf{X}_u^A - \pi(\mathbf{X}_u^A)\right\rangle\end{aligned}$$

Therefore, (2.56) becomes

$$\begin{aligned}\mathrm{d}Y_t^A = -\rho\beta_A Y_t^A \mathrm{d}t + \rho \sum_{\substack{\mathcal{J}\subsetneq A \\ \#\mathcal{J}=2}} \nu^A(\mathcal{J}) 2^{1+\frac{\#A}{2}} \left\langle \hat{\mathbf{w}}_{A\smallsetminus\mathcal{J}}^A,\ \mathbf{X}_t^A\right\rangle Y_t^{\mathcal{J}} \mathrm{d}t \\ + \left\langle \hat{\mathbf{w}}_A^A, \mathbf{F}_t\right\rangle \mathrm{d}t + \left\langle \hat{\mathbf{w}}_A^A, \mathrm{d}\mathbf{M}_t\right\rangle\end{aligned}$$

which can be solved as

$$\begin{aligned}Y_t^A = Y_0^A e^{-\rho\beta_A t} + \rho \sum_{\substack{\mathcal{J}\subsetneq A \\ \#\mathcal{J}=2}} \nu^A(\mathcal{J}) 2^{1+\frac{\#A}{2}} \int_0^t e^{-\rho\beta_A(t-u)} \left\langle \hat{\mathbf{w}}_{A\smallsetminus\mathcal{J}}^A,\ \mathbf{X}_u^A\right\rangle Y_u^{\mathcal{J}} \mathrm{d}u \\ + \int_0^t e^{-\rho\beta_A(t-u)} \left\langle \hat{\mathbf{w}}_A^A, \mathbf{F}_u\right\rangle \mathrm{d}u + \int_0^t e^{-\rho\beta_A(t-u)} \left\langle \hat{\mathbf{w}}_A^A, \mathrm{d}\mathbf{M}_u\right\rangle.\end{aligned}$$

The first, third and fourth terms are handled as in the case $\#\mathcal{I} = 2$. For the second term, let us define

$$\begin{aligned}b_1(t) &:= \rho\beta_A \int_0^t e^{-\rho\beta_A(t-u)} \times \left|Y_0^{\mathcal{J}}\right| e^{-\rho\beta_{\mathcal{J}} u} \mathrm{d}u \\ b_2(t) &:= \rho\beta_A \int_0^t e^{-\rho\beta_A(t-u)} \left(\int_0^u e^{-\rho\beta_{\mathcal{J}}(u-v)} \left|\left\langle \hat{\mathbf{w}}_{\mathcal{J}}^A, \mathbf{F}_v\right\rangle\right| \mathrm{d}v\right) \mathrm{d}u \\ b_3(t) &:= \rho\beta_A \int_0^t e^{-\rho\beta_A(t-u)} \left|\int_0^u e^{-\rho\beta_{\mathcal{J}}(u-v)} \left\langle \hat{\mathbf{w}}_{\mathcal{J}}^A, \mathrm{d}\mathbf{M}_v\right\rangle\right| \mathrm{d}u\end{aligned}$$

Notice from (2.43) that $\nu^A(\mathcal{J}) \leq \beta_{\mathcal{J}}$. This, along with the boundedness of $\langle \hat{\mathbf{w}}^A_{A\setminus\mathcal{J}},\, \mathbf{X}^A_u \rangle$ and (2.57) lets us write for $\mathcal{J} \subseteq A$ with $\#\mathcal{J} = 2$

$$\rho\nu^A(\mathcal{J}) \int_0^t e^{-\rho\beta_A(t-u)} \left| \langle \hat{\mathbf{w}}^A_{A\setminus\mathcal{J}},\, \mathbf{X}^A_u \rangle Y^{\mathcal{J}}_u \right| \mathrm{d}u \leq C(b_1(t) + b_2(t) + b_3(t)) \tag{2.58}$$

where C is a constant independent of (ε, A, L). We now control b_1 by writing

$$\begin{aligned} b_1(t) = \rho\beta_{\mathcal{J}} \int_0^t e^{-\rho\beta_A(t-u)} e^{-\rho\beta_{\mathcal{J}}u} \mathrm{d}u \left|Y^{\mathcal{J}}_0\right| + \rho(\beta_A - \beta_{\mathcal{J}}) \int_0^t e^{-\rho\beta_A(t-u)} e^{-\rho\beta_{\mathcal{J}}u} \mathrm{d}u \left|Y^{\mathcal{J}}_0\right| \\ = \rho\beta_{\mathcal{J}} e^{-\rho\beta_{\mathcal{J}}t} \int_0^t e^{-\rho(\beta_A-\beta_{\mathcal{J}})(t-u)} \mathrm{d}u \left|Y^{\mathcal{J}}_0\right| \\ + e^{-\rho\beta_{\mathcal{J}}t} \int_0^t \rho(\beta_A - \beta_{\mathcal{J}}) e^{-\rho(\beta_A-\beta_{\mathcal{J}})(t-u)} \mathrm{d}u \left|Y^{\mathcal{J}}_0\right| \end{aligned}$$

Because of the order property (2.44), $\beta_A \geq \beta_{\mathcal{J}}$. So the first integral is smaller than t, the second one is smaller than one. We thus obtain

$$b_1(t) \leq C(1 + \rho\beta_{\mathcal{J}}t)e^{-\rho\beta_{\mathcal{J}}t}$$

We can use the inequality

$$\forall x, h \geq 0, \qquad (x+h)e^{-(x+h)} \leq \frac{C}{x}$$

with $x = \rho\beta_{\mathcal{J}}\varepsilon$ and $h = \rho\beta_{\mathcal{J}}(t - \varepsilon)$ to conclude

$$\sup_{t\in[\varepsilon,T]} b_1(t) \leq \frac{C}{\rho\beta_{\mathcal{J}}\varepsilon} \leq \frac{C}{\rho r_A \varepsilon} \tag{2.59}$$

for some universal constant C, using $r_A \leq \beta_{\mathcal{J}}$.

We now control b_2. Because $\mathbf{F}_t$ is of order L we can write

$$b_2(t) \leq CL\rho\beta_A \int_0^t e^{-\rho\beta_A(t-u)} \left(\int_0^u e^{-\rho\beta_{\mathcal{J}}(u-v)} \mathrm{d}v \right) \mathrm{d}u \quad \leq \frac{CL}{\rho\beta_{\mathcal{J}}} \leq \frac{CL}{\rho r_A} \tag{2.60}$$

for C independent of (ε, A, L). Finally, we turn to b_3. Because the martingale $\mathbf{M}_u$ has uniformly bouded quadratic variation, we can use Lemma 2.3.12 to write

$$\begin{aligned} \mathbb{E}\left[\sup_{t\in[0,T]} b_3(t)\right] \leq & \rho\beta_A \int_0^t e^{-\rho\beta_A(t-u)} \mathrm{d}u \mathbb{E}\left[\sup_{t\in[0,T]} \left| \int_0^u e^{-\rho\beta_{\mathcal{J}}(u-v)} \langle \hat{\mathbf{w}}^A_{\mathcal{J}}, \mathrm{d}\mathbf{M}_v \rangle \right|\right] \\ \leq & \sqrt{\frac{1 + \ln_+(\rho\beta_{\mathcal{J}}2T)}{\rho\beta_{\mathcal{J}}}} \\ \leq & C\sqrt{\frac{\ln(\rho)}{\rho r_A}} \end{aligned}$$

using (2.51) and $\rho \geq e$. We obtain the result by combining this with (2.59) and (2.60) in (2.58). □

2.3.5 Summing controls of linkage equilibrium across loci

We now prove the following estimates

Proposition 2.3.13. *Recall from (2.18)*

$$\varepsilon_L := \frac{1}{\sqrt{\rho r^{**}}}$$

Let us assume (2.12), that is,

$$\rho r^{**} \gg L^2 \ln(\rho)$$

We have

$$\lim_{L\to+\infty} \sum_{\ell\in[L]} \mathbb{E}\left[\int_0^T |S^\ell(\mathbf{X}_t) - S^\ell(\pi(\mathbf{X}_t))|\mathrm{d}t\right] =0 \tag{2.61}$$

$$\lim_{L\to+\infty} \frac{1}{L} \sum_{\substack{\ell_1,\ell_2\in[L]\\ \ell_1\neq\ell_2}} \mathbb{E}\left[\sup_{t\in[\varepsilon_L,T]} \left|D^{\ell_1,\ell_2}(\mathbf{X}_t)\right| \mathrm{d}t\right] =0 \tag{2.62}$$

Consider a sequence integers $\ell^L \equiv \ell \in [L]$ satisfying Assumptions (2.13). Then

$$\lim_{L\to+\infty} L\mathbb{E}\left[\int_0^T |S^\ell(\mathbf{X}_t) - S^\ell(\pi(\mathbf{X}_t))|\mathrm{d}t\right] =0 \tag{2.63}$$

Finally, consider $\ell_1^L \equiv \ell_1, \ell_2^L \equiv \ell_2$ such that

$$\rho r_{\{\ell_1,\ell_2\}} \gg L \tag{2.64}$$

Then

$$\lim_{L\to\infty} \mathbb{E}\left[\int_0^T |D^{\ell_1,\ell_2}(\mathbf{X}_t)|\right] = 0 \tag{2.65}$$

We use the following computational Lemma.

Lemma 2.3.14. *We can find a constant C independent of (L,ε) such that for any $\ell_0 \in [L]$*

$$\sum_{\substack{A\subseteq[L]\setminus\{\ell_0\}\\ 1\le\#A\le2}} \frac{1}{L^{\#A}} \left(\frac{1}{\rho r_{\{\ell_0\}\cup A}\varepsilon} + \frac{L}{\rho r_{\{\ell_0\}\cup A}} + \sqrt{\frac{\ln(\rho)}{\rho r_{\{\ell_0\}\cup A}}}\right)$$
$$\le C\left(\left(\frac{1}{\rho\varepsilon} + \frac{L}{\rho}\right)\left(\frac{1}{r^*_{\ell_0}} + \frac{1}{r^{**}}\right) + \sqrt{\frac{\ln(\rho)}{\rho r^*_{\ell_0}}} + \sqrt{\frac{\ln(\rho)}{\rho r^{**}}}\right) \tag{2.66}$$

Furthermore,

$$\sum_{\substack{A\subseteq[L]\\ 2\le\#A\le3}} \frac{1}{L^{\#A}} \left(\frac{1}{\rho r_A\varepsilon} + \frac{L}{\rho r_A} + \sqrt{\frac{\ln(\rho)}{\rho r_A}}\right) \le C\left(\frac{1}{\rho r^{**}\varepsilon} + \frac{L}{\rho r^{**}} + \sqrt{\frac{\ln(\rho)}{\rho r^{**}}}\right) \tag{2.67}$$

Proof. In the following, we write a_{ℓ_0} for the left-hand side of (2.66). Then

$$a_{\ell_0} = \left(\frac{1}{\rho\varepsilon} + \frac{L}{\rho}\right) \sum_{\substack{A\subseteq[L]\setminus\{\ell_0\}\\ 1\le\#A\le2}} \frac{1}{L^{\#A} r_{\{\ell_0\}\cup A}} + \sqrt{\frac{\ln(\rho)}{\rho}} \sum_{\substack{A\subseteq[L]\setminus\{\ell_0\}\\ 1\le\#A\le2}} \frac{1}{L^{\#A}\sqrt{r_{\{\ell_0\}\cup A}}}$$

From Jensen's inequality,

$$\sum_{\substack{A\subseteq[L]\setminus\{\ell_0\}\\1\leq\#A\leq 2}}\frac{1}{KL^{\#A}}\times\frac{1}{\sqrt{r_{\{\ell_0\}\cup A}}}\leq\left(\sum_{\substack{A\subseteq[L]\setminus\{\ell_0\}\\1\leq\#A\leq 2}}\frac{1}{KL^{\#A}}\times\frac{1}{r_{\{\ell_0\}\cup A}}\right)^{\frac{1}{2}}$$

where K is a normalization constant

$$K:=\sum_{\substack{A\subseteq[L]\setminus\{\ell_0\}\\1\leq\#A\leq 2}}\frac{1}{L^{\#A}}$$

Because K is of order 1, we can absorb it into a universal constant C and we get

$$a_{\ell_0}\leq C\left(\left(\frac{1}{\rho\varepsilon}+\frac{L}{\rho}\right)H^{\ell_0}+\sqrt{\ln(\rho)\frac{H^{\ell_0}}{\rho}}\right) \tag{2.68}$$

where

$$H^{\ell_0}:=\sum_{\substack{A\subseteq[L]\setminus\{\ell_0\}\\1\leq\#A\leq 2}}\frac{1}{L^{\#A}}\times\frac{1}{r_{\{\ell_0\}\cup A}}$$

We write

$$\begin{aligned}H^{\ell_0}=&\frac{1}{L}\sum_{\ell_1\in[L]\setminus\{\ell_0\}}\frac{1}{r_{\{\ell_0,\ell_1\}}}+\frac{1}{L^2}\sum_{\substack{\ell_1,\ell_2\in[L]\setminus\{\ell_0\}\\\ell_1\neq\ell_2}}\frac{1}{r_{\{\ell_0,\ell_1,\ell_2\}}}\\\leq&\frac{1}{r^*_{\ell_0}}+\frac{1}{L^2}\sum_{\substack{\ell_1,\ell_2\in[L]\setminus\{\ell_0\}\\\ell_1\neq\ell_2}}\frac{1}{r_{\{\ell_0,\ell_1,\ell_2\}}}\end{aligned}$$

from the definition of $r^*_{\ell_0}$ in (2.8) .

Since $r_A=\min_{\ell_1\neq\ell_2\in A}r_{\{\ell_1,\ell_2\}}$

$$\frac{1}{r_{\{\ell_0,\ell_1,\ell_2\}}}\leq\frac{1}{r_{\{\ell_0,\ell_1\}}}+\frac{1}{r_{\{\ell_0,\ell_2\}}}+\frac{1}{r_{\{\ell_1,\ell_2\}}}$$

It follows

$$\frac{1}{L^2}\sum_{\substack{\ell_1,\ell_2\in[L]\setminus\{\ell_0\}\\\ell_1\neq\ell_2}}\frac{1}{r_{\{\ell_0,\ell_1,\ell_2\}}}\leq\frac{1}{L^2}\sum_{\substack{\ell_1,\ell_2\in[L]\setminus\{\ell_0\}\\\ell_1\neq\ell_2}}\frac{1}{r_{\{\ell_1,\ell_2\}}}+\frac{1}{L^2}\sum_{\substack{\ell_1,\ell_2\in[L]\setminus\{\ell_0\}\\\ell_1\neq\ell_2}}\frac{1}{r_{\{\ell_0,\ell_1\}}}+\frac{1}{r_{\{\ell_0,\ell_2\}}}$$

The first sum is smaller than $\frac{1}{r^{**}}$, the second one is smaller than $\frac{2}{r^*_{\ell_0}}$. Putting it all together we find a constant C independent of (ε,L,ℓ_0) such that

$$H^{\ell_0}\leq C\left(\frac{1}{r^*_{\ell_0}}+\frac{1}{r^{**}}\right)$$

We thus obtain that for any $\ell_0\in[L]$

$$a_{\ell_0}\leq C\left(\left(\frac{1}{\rho\varepsilon}+\frac{L}{\rho}\right)\left(\frac{1}{r^*_{\ell_0}}+\frac{1}{r^{**}}\right)+\sqrt{\frac{\ln(\rho)}{\rho}}\left(\frac{1}{r^*_{\ell_0}}+\frac{1}{r^{**}}\right)^{\frac{1}{2}}\right)$$

for C independent of (ε, L, ℓ_0). To get (2.66), use that $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for any $x, y \geq 0$. We turn to (2.67). Write a^* for the left-hand side of (2.67). We have

$$\begin{aligned}
a^* =& \frac{1}{L} \sum_{\ell\in[L]} a_\ell \\
\leq& C\left(\left(\frac{1}{\rho\varepsilon}+\frac{L}{\rho}\right)\left(\frac{1}{L}\sum_{\ell\in[L]}\frac{1}{r^*_\ell}+\frac{1}{r^{**}}\right)+\sqrt{\frac{\ln(\rho)}{\rho r^{**}}}+\sqrt{\frac{\ln(\rho)}{\rho}}\frac{1}{L}\sum_{\ell\in[L]}\frac{1}{\sqrt{r^*_\ell}}\right) \\
\leq& C\left(\left(\frac{1}{\rho\varepsilon}+\frac{L}{\rho}\right)\frac{2}{r^{**}}+\sqrt{\frac{\ln(\rho)}{\rho r^{**}}}+\sqrt{\frac{\ln(\rho)}{\rho}}\frac{1}{L}\sum_{\ell\in[L]}\frac{1}{\sqrt{r^*_\ell}}\right)
\end{aligned}$$

We conclude with Jensen's inequality

$$\frac{1}{L}\sum_{\ell\in[L]}\frac{1}{\sqrt{r^*_\ell}} \leq \left(\frac{1}{L}\sum_{\ell\in[L]}\frac{1}{r^*_\ell}\right)^{\frac{1}{2}} = \frac{1}{\sqrt{r^{**}}}$$

□

Proof of Proposition 2.3.13. First, let us notice that (2.12) implies

$$\rho \gg \frac{L^2}{r^{**}} = \frac{L^2}{L(L-1)}\sum_{\substack{\ell_1,\ell_2\in[L]\\ \ell_1\neq\ell_2}}\frac{1}{r_{\{\ell_1,\ell_2\}}} \geq \max_{\substack{\ell_1,\ell_2\in[L]\\ \ell_1\neq\ell_2}}\frac{1}{r_{\{\ell_1,\ell_2\}}} = \frac{1}{r_{[L]}}$$

In particular, for any $A \subseteq [L]$ we have $\rho r_A \geq \rho r_{[L]} \gg 1$ so that we may apply Proposition 2.3.10 and Lemma 2.3.14. Second, we note that by (2.12), ε_L satisfies

$$\frac{1}{\rho r^{**}} \ll \varepsilon_L \ll \frac{1}{L} \tag{2.69}$$

Let us start with (2.61). We have

$$\begin{aligned}
\sum_{\ell\in[L]}\mathbb{E}\left[\int_0^T |S^\ell(\mathbf{X}_t)-S^\ell(\pi(\mathbf{X}_t))|\mathrm{d}t\right] =& \mathcal{O}(L\varepsilon_L)+\mathbb{E}\left[\sum_{\ell\in[L]}\int_{\varepsilon_L}^T |S^\ell(\mathbf{X}_t)-S^\ell(\pi(\mathbf{X}_t))|\mathrm{d}t\right] \\
\leq& o(1)+T\sum_{\ell\in[L]}\mathbb{E}\left[\sup_{t\in[\varepsilon_L,T]}\left|S^\ell(\mathbf{X}_t)-S^\ell(\pi(\mathbf{X}_t))\right|\right]
\end{aligned}$$

from (2.69). Using (2.32) from Proposition 2.3.2 we get

$$\sum_{\ell\in[L]}\mathbb{E}\left[\sup_{t\in[\varepsilon_L,T]}\left|S^\ell(\mathbf{X}_t)-S^\ell(\pi(\mathbf{X}_t))\right|\right] \leq CL\sum_{\substack{A\subseteq[L]\\ 2\leq\#A\leq 3}}\frac{1}{L^{\#A}}\mathbb{E}\left[\sup_{t\in[\varepsilon_L,T]}||\mathbf{X}^A_t-\pi(\mathbf{X}^A_t)||_2\right]$$

Then, Proposition 2.3.10 and (2.67) from Lemma 2.3.14 yield

$$\sum_{\ell\in[L]}\mathbb{E}\left[\sup_{t\in[\varepsilon_L,T]}\left|S^\ell(\mathbf{X}_t)-S^\ell(\pi(\mathbf{X}_t))\right|\right] \leq CL\left(\frac{1}{\rho r^{**}\varepsilon_L}+\frac{L}{\rho r^{**}}+\sqrt{\frac{\ln(\rho)}{\rho r^{**}}}\right)$$

By (2.12) and (2.69), we obtain that the right-hand side is $o(1)$. We obtain (2.62) the same way, using (2.31) from Proposition 2.3.2.

We now consider a sequence $\ell^L \equiv \ell$ satisfying (2.13). We define

$$\varepsilon_{\ell,L} \equiv \varepsilon_\ell := \frac{1}{\sqrt{\rho r_\ell^*}} \vee \varepsilon_L$$

Just as in (2.69) we have by (2.13)

$$\frac{1}{\rho r_\ell^*} \ll \varepsilon_\ell \ll \frac{1}{L}$$

It follows as before

$$\begin{aligned} L\mathbb{E}\left[\int_0^T |S^\ell(\mathbf{X}_t) - S^\ell(\pi(\mathbf{X}_t))|\mathrm{d}t\right] \leq & \mathcal{O}(L\varepsilon_\ell) + TL\mathbb{E}\left[\sup_{t\in[\varepsilon_\ell,T]}\left|S^\ell(\mathbf{X}_t) - S^\ell(\pi(\mathbf{X}_t))\right|\right] \\ \leq & o(1) + TL\mathbb{E}\left[\sup_{t\in[\varepsilon_\ell,T]}\left|S^\ell(\mathbf{X}_t) - S^\ell(\pi(\mathbf{X}_t))\right|\right] \end{aligned}$$

Using Proposition 2.3.10 and (2.66) from Lemma 2.3.14 we get

$$\begin{aligned} & TL\mathbb{E}\left[\sup_{t\in[\varepsilon_\ell,T]}\left|S^\ell(\mathbf{X}_t) - S^\ell(\pi(\mathbf{X}_t))\right|\right] \\ & \qquad \leq CL\left(\left(\frac{1}{\rho\varepsilon_L} + \frac{L}{\rho}\right)\left(\frac{1}{r_\ell^*} + \frac{1}{r^{**}}\right) + \sqrt{\frac{\ln(\rho)}{\rho r_\ell^*}} + \sqrt{\frac{\ln(\rho)}{\rho r^{**}}}\right) \\ & \qquad \leq C\left(\frac{L}{\rho r_\ell^*\varepsilon_L} + \frac{L}{\rho r_\ell^*} + \frac{L}{\rho r^{**}\varepsilon_L} + \frac{L}{\rho r^{**}} + \sqrt{\frac{\ln(\rho)L^2}{\rho r_\ell^*}} + \sqrt{\frac{\ln(\rho)L^2}{\rho r^{**}}}\right) \end{aligned}$$

for some constant C independent of (L,ℓ). The right-hand side is small from (2.12),(2.13) and the definition of ε_ℓ.

Finally, to prove (2.65) we define $\varepsilon_{\ell_1,\ell_2,L} \equiv \varepsilon_{\ell_1,\ell_2} := \frac{1}{\sqrt{\rho r_{\{\ell_1,\ell_2\}}}} \vee \frac{1}{\sqrt{\rho r^{**}}}$. We have

$$\int_0^t \left|D^{\ell_1,\ell_2}(\mathbf{X}_u)\right| \mathrm{d}u \leq \varepsilon_{\ell_1,\ell_2} + t \sup_{u\in[\varepsilon_{\ell_1,\ell_2},T]}\left|D^{\ell_1,\ell_2}(\mathbf{X}_t)\right|$$

The first term on the right-hand side is $o(1)$ from Assumption (2.64). For the second term, (2.31) from Proposition 2.3.2 and Proposition 2.3.10 tell us

$$\mathbb{E}\left[\sup_{u\in[\varepsilon_{\ell_1,\ell_2},T]}|D^{\ell_1,\ell_2}(\mathbf{X}_t)|\right] \leq C\left(\frac{1}{\rho r_{\{\ell_1,\ell_2\}}\varepsilon_{\ell_1,\ell_2}} + \frac{L}{\rho r_{\{\ell_1,\ell_2\}}} + \sqrt{\frac{\ln(\rho)}{\rho r_{\{\ell_1,\ell_2\}}}}\right)$$

The result follows from Assumption (2.64) and the definition of $\varepsilon_{\ell_1,\ell_2}$. □

2.3.6 Proof of Theorem 2.1.1

In this section, we make the following assumptions of Theorem 2.1.1:

- $\mu_{\mathbf{X}_0}$ converges in law to a deterministic measure m_0.
- We have $\rho r^{**} \gg L^2\ln(\rho)$ (2.12)

The proof is broken down into several lemmas. Let us define for $\mathbf{y} \in \mathbb{X}^{[L]}$ and $\varphi \in \mathcal{C}^2([0,1])$

$$G^\ell_{\mathbf{y}}\varphi(x) := x(1-x)LS^\ell(\mathbf{y})\varphi'(x) + \overline{\Theta}(x)\varphi'(x) + \frac{1}{2}x(1-x)\varphi''(x) \tag{2.70}$$

$G^\ell_{\mathbf{X}_t}$ can be thought of as the generator for the $\ell-$th locus. And the limit generator is

$$\forall x \in [0,1], \xi \in \mathcal{P}([0,1]), \qquad \overline{G}_\xi\varphi(x) := \left(\overline{s}(\xi)x(1-x) + \overline{\Theta}(x)\right)\varphi'(x) + \frac{1}{2}x(1-x)\varphi''(x) \tag{2.71}$$

Lemma 2.3.15. *Consider a sequence $\ell^L \equiv \ell \in [L]$ satisfying (2.13), that is*

$$\rho r^*_\ell \gg L^2\ln(\rho)$$

We have for $\varphi \in \mathcal{C}^2([0,1])$

$$\mathbb{E}\left[\int_0^T \mathrm{d}u \, \left|G^\ell_{\mathbf{X}_u}\varphi(p^\ell(\mathbf{X}_u)) - \overline{G}_{\mu_{\mathbf{X}_u}}\varphi(p^\ell(\mathbf{X}_u))\right|\right] \longrightarrow 0 \tag{2.72}$$

Furthermore, for any $\varphi \in \mathcal{C}^2([0,1])$ we have

$$\mathbb{E}\left[\frac{1}{L}\sum_{\ell\in[L]}\int_0^T \mathrm{d}u \, \left|G^\ell_{\mathbf{X}_u}\varphi(p^\ell(\mathbf{X}_u)) - \overline{G}_{\mu_{\mathbf{X}_u}}\varphi(p^\ell(\mathbf{X}_u))\right|\right] \longrightarrow 0 \tag{2.73}$$

Proof. We first consider the case where $\ell^L \equiv \ell$ is such that $\rho r^*_\ell \gg L^2\ln(\rho)$. For any $t \in [0,T]$ we have from (2.70)

$$\begin{aligned}&\left|G^\ell_{\mathbf{X}_t}\varphi(p^\ell(\mathbf{X}_t)) - \overline{G}_{\mu_{\mathbf{X}_t}}\varphi(p^\ell(\mathbf{X}_t))\right| \\ &\qquad = \left|LS^\ell(\mathbf{X}_t) - \overline{s}(\mu_{\mathbf{X}_t})p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))\right| \times |\varphi'(p^\ell(\mathbf{X}_t))|\end{aligned}$$

Observe that

$$\begin{aligned}\left|LS^\ell(\mathbf{X}_t) - \overline{s}(\mu_{\mathbf{X}_t})p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))\right| &\le L\left|S^\ell(\mathbf{X}_t) - S^\ell(\pi(\mathbf{X}_t))\right| \\ &\quad + \left|LS^\ell(\pi(\mathbf{X}_t)) - \overline{s}(\mu_{\mathbf{X}_t})p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))\right|\end{aligned}$$

The expectation of the integral of the first term is $o(1)$ from (2.63) in Proposition 2.3.13. The expectation of the second term is $\mathcal{O}\left(\frac{1}{L}\right)$ from Lemma 2.3.6. To prove (2.73), write

$$\begin{aligned}&\left|\left(G^\ell_{\mathbf{X}_t} - \overline{G}_{\mu_{\mathbf{X}_t}}\right)\varphi(p^\ell(\mathbf{X}_t))\right| \\ &\quad = \frac{1}{L}\sum_{\ell\in[L]}\left|LS^\ell(\mathbf{X}_t) - \overline{s}(\mu_{\mathbf{X}_t})p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))\right| \times |\varphi'(p^\ell(\mathbf{X}_t))| \\ &\quad \le \frac{||\varphi'||_\infty}{L}\sum_{\ell\in[L]} L\left|S^\ell(\mathbf{X}_t) - S^\ell(\pi(\mathbf{X}_t))\right| \\ &\qquad + \frac{||\varphi'||_\infty}{L}\sum_{\ell\in[L]}\left|LS^\ell(\pi(\mathbf{X}_t)) - \overline{s}(\mu_{\mathbf{X}_t})p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))\right|\end{aligned}$$

The expectation of the integral of the first term is $o(1)$ from (2.61) in Proposition 2.3.13, and the second term is small from Lemma 2.3.6. ☐

To get the convergence of $(\mu_{\mathbf{X}_t})_{t\in[0,T]}$ we follow a classical proof, with first a tightness Lemma, then proof that we get the correct martingale problem in the limit.

Lemma 2.3.16. *Consider a sequence $\ell^L \equiv \ell \in [L]$ such that (2.13) is satisfied, that is,*

$$\rho r^*_\ell \gg L^2 \ln(\rho)$$

The law of $(p^\ell(\mathbf{X}_t))_{t\in[0,T]} \in \mathbb{D}([0,T],[0,1])$ is tight for the Skorokhod J1 topology. Furthermore, the allelic law process $(\mu_{\mathbf{X}_t})_{t\in[0,T]} \in \mathbb{D}([0,T],\mathcal{P}([0,1]))$ is tight for the Skorokhod J1 topology.

Proof. To prove tightness of $p^\ell(X_t)_{t\in[0,T]}$ in $\mathbb{D}([0,T],[0,1])$ for the Skorokhod J1 topology, we use the classical Rebolledo criterion (see Theorem C.4 in [157]). We first separate martingale and finite variation part from the SDE (2.38).

$$p^\ell(\mathbf{X}_t) - p^\ell(\mathbf{X}_0) = \int_0^t G^\ell_{\mathbf{X}_u}\mathrm{Id}(p^\ell(\mathbf{X}_u)))\mathrm{d}u + \int_0^t \sqrt{p^\ell(\mathbf{X}_u)(1-p^\ell(\mathbf{X}_u))}\mathrm{d}\hat{B}^\ell_u =: V^\ell_t + M^\ell_t$$

The Rebolledo criterion has three conditions:

1. $p^\ell(\mathbf{X}_0)$ is tight.
2. V^ℓ_t and M^ℓ_t are tight for all $t \in [0,T]$.
3. For $A = V^\ell$ or $A = \langle M^\ell \rangle^{QV}$ we have

$$\forall \varepsilon > 0, \qquad \lim_{\delta\downarrow 0} \limsup_{L\to+\infty} \sup_{(t_1,t_2)\in\mathcal{S}_\delta} \mathbb{P}\left[|A_{t_2} - A_{t_1}| > \varepsilon\right] = 0$$

where $\mathcal{S}_\delta$ is the set of pairs of stopping times (t_1,t_2) for the natural filtration of the process $(\mathscr{F}_t)_{t\in[0,T]}$, such that $|t_1 - t_2| < \delta$ a.s.

Since $p^\ell(\mathbf{X}_0)$ lives in $[0,1]$, the first condition is trivially satisfied. We turn to the second condition. Since $\langle M^\ell \rangle^{QV}_t \leq t$ from Corollary 2.3.5, M^ℓ_t is tight. Furthermore, since $V^\ell_t = p^\ell(\mathbf{X}_t) - p^\ell(\mathbf{X}_0) - M^\ell_t$, it is necessarily tight. The third condition for $A = \langle M^\ell \rangle^{QV}$ is trivially satisfied for $\langle M^\ell \rangle^{QV}$ since for $t \in [0,T]$ we have $\frac{\mathrm{d}}{\mathrm{d}t}\langle M^\ell \rangle_t \leq 1$. For $A = V^\ell$, we write

$$\begin{aligned}\int_{t_1}^{t_2} &|G^\ell_{\mathbf{X}_u}\mathrm{Id}(p^\ell(\mathbf{X}_u)))|\mathrm{d}u \\ &\leq \int_{t_1}^{t_2} |\overline{G}_{\mu_{\mathbf{X}_u}}\mathrm{Id}(p^\ell(\mathbf{X}_u)))|\mathrm{d}u + \int_0^T \left|G^\ell_{\mathbf{X}_u}\mathrm{Id}(p^\ell(\mathbf{X}_u))) - \overline{G}_{\mu_{\mathbf{X}_u}}\mathrm{Id}(p^\ell(\mathbf{X}_u)))\right|\mathrm{d}u\end{aligned}$$

The first term can be bounded by $C|t_1 - t_2|$ for some deterministic constant C, the second term goes to 0 in probability from (2.72) in Lemma 2.3.15 uniformly in t_1, t_2. This yields the tightness of $(p^\ell(\mathbf{X}_t))_{t\in[0,T]}$.

To get the tightness of $(\mu_{\mathbf{X}_t})_{t\in[0,T]} \in \mathbb{D}([0,T],\mathcal{P}([0,1]))$, following Theorem 2.1 of [158], we only need to show that for any $\varphi \in \mathcal{C}^2([0,1])$, the process $(<\mu_{\mathbf{X}_t},\varphi>)_{t\in[0,T]}$ is tight. We may again apply the Rebolledo criterion, writing

$$<\mu_{\mathbf{X}_t},\varphi> - <\mu_{\mathbf{X}_0},\varphi> \; = \frac{1}{L}\sum_{\ell\in[L]} \varphi(p^\ell(\mathbf{X}_t)) - \varphi(p^\ell(\mathbf{X}_0)) = \frac{1}{L}\sum_{\ell\in[L]} V^{\varphi,\ell}_t + M^{\varphi,\ell}_t$$

with

$$V^{\varphi,\ell}_t := \int_0^t G^\ell_{\mathbf{X}_u}\varphi(p^\ell(\mathbf{X}_u))\mathrm{d}u \qquad M^{\varphi,\ell}_t := \int_0^t \varphi'(p^\ell(\mathbf{X}_u))\sqrt{p^\ell(\mathbf{X}_u)(1-p^\ell(\mathbf{X}_u))}\mathrm{d}\hat{B}^\ell_u$$

The condition 1 of the Rebolledo criterion is easily verified, since $<\mu_{\mathbf{X}_0},\varphi>$ is tight because bounded by $||\varphi||_\infty$. The condition 2 is obtained as above. Indeed, on the one hand, the quadratic variation of $M_t^{\varphi,\ell}$ is uniformly bounded by $T||\varphi'||_\infty$ so $M_t^{\varphi,\ell}$ is tight for any fixed $t \in [0,T]$. On the other hand, since $p^\ell(\mathbf{X}_t)$ is in $[0,1]$, it is also tight, and therefore $V_t^{\varphi,\ell}$ is tight.

Controlling $L^{-1} \sum_{\ell\in[L]} V_t^{\varphi,\ell}$ works just as above, using (2.73) instead of (2.72). We only need to control the martingale $L^{-1} \sum_{\ell\in[L]} M_t^{\varphi,\ell}$. We have for $t_1, t_2 \in [0,T]$

$$\left\langle \frac{1}{L} \sum_{\ell\in[L]} M^{\varphi,\ell} \right\rangle_{t_2}^{QV} - \left\langle \frac{1}{L} \sum_{\ell\in[L]} M^{\varphi,\ell} \right\rangle_{t_1}^{QV} \\ = \frac{1}{L^2} \sum_{\ell_1,\ell_2\in[L]} \langle M^{\varphi,\ell_1}, M^{\varphi,\ell_2}\rangle_{t_2}^{QV} - \langle M^{\varphi,\ell_1}, M^{\varphi,\ell_2}\rangle_{t_1}^{QV}$$

Using the Kunita-Watanabe inequality (see Corollary 1.16, chapter IV in [159]), we have

$$\left| \langle M^{\varphi,\ell_1}, M^{\varphi,\ell_2}\rangle_{t_2}^{QV} - \langle M^{\varphi,\ell_1}, M^{\varphi,\ell_2}\rangle_{t_1}^{QV} \right| \\ \leq \sqrt{\left| \langle M^{\varphi,\ell_1}\rangle_{t_2}^{QV} - \langle M^{\varphi,\ell_1}\rangle_{t_1}^{QV} \right| \times \left| \langle M^{\varphi,\ell_2}\rangle_{t_2}^{QV} - \langle M^{\varphi,\ell_2}\rangle_{t_1}^{QV} \right|}$$

This is uniformly bounded by $||\varphi'||_\infty^2 \, |t_2 - t_1|$, which yields the result. □

We can now show that $(\mu_{\mathbf{X}_t})_{t\in[0,T]}$ converges to a solution of (2.10).

Proof of Theorem 2.1.1 part 1. We will prove for any function $\varphi \in \mathcal{C}^2([0,1])$

$$< \mu_{\mathbf{X}_t}, \varphi > - < \mu_{\mathbf{X}_0}, \varphi > - \int_0^t < \mu_{\mathbf{X}_u}, \overline{G}_{\mu_{\mathbf{X}_u}} \varphi > \mathrm{d}u \longrightarrow 0$$

in probability. Notice that

$$\begin{aligned} M_t^\varphi :=& \frac{1}{L} \sum_{\ell\in[L]} \varphi(p^\ell(\mathbf{X}_t)) - \varphi(p^\ell(\mathbf{X}_0)) - \int_0^t \mathrm{d}u G_{\mathbf{X}_u}^\ell \varphi(p^\ell(\mathbf{X}_u)) \\ =& < \mu_{\mathbf{X}_t}, \varphi > - < \mu_{\mathbf{X}_0}, \varphi > - \int_0^t \mathrm{d}u \frac{1}{L} \sum_{\ell\in[L]} G_{\mathbf{X}_u}^\ell \varphi(p^\ell(\mathbf{X}_u)) \end{aligned}$$

is a martingale, which by Itô's formula and Corollary 2.3.5 satisfies

$$\mathrm{d}M_t^\varphi := \frac{1}{L} \sum_{\ell\in[L]} \sqrt{p^\ell(\mathbf{X}_t)(1 - p^\ell(\mathbf{X}_t))} \, \varphi'(p^\ell(\mathbf{X}_t)) \mathrm{d}\hat{B}_t^\ell.$$

We split the equation between a martingale and non-martingale part

$$\left| < \mu_{\mathbf{X}_t}, \varphi > - < \mu_{\mathbf{X}_0}, \varphi > - \int_0^t \mathrm{d}u < \mu_{\mathbf{X}_u}, \overline{G}_{\mu_{\mathbf{X}_u}} \varphi > \right| \\ = |M_t^\varphi| + \int_0^t \mathrm{d}u \left| < \mu_{\mathbf{X}_u}, \overline{G}_{\mu_{\mathbf{X}_u}} \varphi > - \frac{1}{L} \sum_{\ell\in[L]} G_{\mathbf{X}_u}^\ell \varphi(p^\ell(\mathbf{X}_u)) \right|$$

Let us show M_t^φ goes to 0. We have

$$\begin{aligned}
\mathrm{d}\left\langle M^\varphi \right\rangle_t^{QV} &= \frac{1}{L^2} \sum_{\ell \in [L]} \varphi'(p^\ell(\mathbf{X}_t))^2 p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))\mathrm{d}t \\
&\quad + \frac{1}{L^2} \sum_{\substack{\ell_1,\ell_2 \in [L] \\ \ell_1 \neq \ell_2}} \varphi'(p^{\ell_1}(\mathbf{X}_t))\varphi'(p^{\ell_2}(\mathbf{X}_t))\, \mathrm{d}\left\langle p^{\ell_1}(\mathbf{X}), p^{\ell_2}(\mathbf{X}) \right\rangle_t^{QV} \\
&\qquad\qquad \leq \frac{1}{L}||\varphi'||_\infty^2 + \frac{1}{L^2} \sum_{\substack{\ell_1,\ell_2 \in [L] \\ \ell_1 \neq \ell_2}} ||\varphi'||_\infty^2 \left| D^{\ell_1,\ell_2}(\mathbf{X}_t) \right| \mathrm{d}t
\end{aligned}$$

where we used (2.39) from Corollary 2.3.5. It follows

$$\left\langle M^\varphi \right\rangle_T^{QV} \leq \frac{1}{L}||\varphi'||_\infty^2 T + \frac{1}{L^2} \sum_{\substack{\ell_1,\ell_2 \in [L] \\ \ell_1 \neq \ell_2}} ||\varphi'||_\infty^2 \int_0^T \left| D^{\ell_1,\ell_2}(\mathbf{X}_t) \right| \mathrm{d}t$$

The first term goes to 0. The expectation of the second term is $o_{\mathbb{P}}(1)$ from (2.62) in Proposition 2.3.13. Since the quadratic variation goes to 0, it follows that M^φ goes to zero in the Skorokhod toplogy from the Dambis, Dubins-Schwartz theorem (Theorem 1.6, chapter V of [159]).

We now need to control

$$\begin{aligned}
&\int_0^t \mathrm{d}u \left| < \mu_{\mathbf{X}_u}, \overline{G}_{\mu_{\mathbf{X}_u}} \varphi > - \frac{1}{L} \sum_{\ell \in [L]} G_{\mathbf{X}_u}^\ell \varphi(p^\ell(\mathbf{X}_u)) \right| \\
&\qquad\qquad \leq \frac{1}{L} \sum_{\ell \in [L]} \int_0^T \mathrm{d}u \left| \overline{G}_{\mu_{\mathbf{X}_u}} \varphi(p^\ell(\mathbf{X}_u)) - G_{\mathbf{X}_u}^\ell \varphi(p^\ell(\mathbf{X}_u)) \right|
\end{aligned}$$

The result follows from (2.73) from Lemma 2.3.15.

We thus obtain that any subsequential limit of $(\mu_{\mathbf{X}_t})_{t \in [0,T]}$ must satisfy equation (2.10), with initial law m_0. In particular, this yields existence of a solution to (2.10), which is unique from Proposition 2.2.3. This completes the proof of the first part Theorem 2.1.1 □

We conclude with the second part of the Theorem.

Proof of Theorem 2.1.1 part 2. First, notice that for every $i \in [n]$, Lemma 2.3.16 implies the tightness of $(p^{\ell_i^L}(\mathbf{X}))_{L \geq 1}$. Since a finite union of tight families is tight, we also get tightness for $(p^{\ell_i^L}(\mathbf{X}))_{L \geq n, i \in [n]}$. We will write $\ell_i^L \equiv \ell_i$ to alleviate notations.

For $\varphi \in \mathcal{C}^2([0,1]^n)$, $\mathbf{x} \in \mathbb{X}^{[L]}$ and $\mathbf{y} \in [0,1]^n$ we define

$$\mathcal{G}_{\mathbf{x}} \varphi(\mathbf{y}) := \sum_{i \in [n]} G_{\mathbf{x}}^{\ell_i} \diamond_i \varphi(\mathbf{y}) + \frac{1}{2} \sum_{\substack{i,j \in [n] \\ i \neq j}} \partial_{i,j} \varphi(\mathbf{y}) D^{\ell_i,\ell_j}(\mathbf{x})$$

where $G_{\mathbf{x}}^\ell \diamond_i \varphi(\mathbf{y})$ means $G_{\mathbf{x}}^\ell$ applied to the function $z \mapsto \varphi(y^1, \ldots, y^{i-1}, z, y^{i+1}, \ldots, y^n)$, evaluated in $z = y^i$. Define $\mathbf{Y}_t := (p^{\ell_i}(\mathbf{X}_t))_{i \in [n]}$. We know from Itô's formula that

$$M_t^\varphi := \varphi(\mathbf{Y}_t) - \varphi(\mathbf{Y}_0) - \int_0^t \mathrm{d}u\, \mathcal{G}_{\mathbf{X}_u} \varphi(\mathbf{Y}_u)$$

is a martingale for any $\varphi \in \mathcal{C}^2([0,1]^n)$. Furthermore, using $\langle p(\mathbf{X})\rangle_t \leq t$, we can find uniform bounds for the quadratic variation of M_t^{φ}. It remains to show

$$\int_0^t \mathrm{d}u\, \mathcal{G}_{\mathbf{X}_u}\varphi(\mathbf{Y}_u) - \int_0^t \mathrm{d}u \sum_{i\in[n]} \overline{G}_{\xi_u} \diamond_i \varphi(\mathbf{Y}_u) \longrightarrow 0$$

where ξ is the limit of $\mu_{\mathbf{X}}$. We write

$$\begin{aligned}
&\left|\int_0^t \mathrm{d}u\, \mathcal{G}_{\mathbf{X}_u}\varphi(\mathbf{Y}_u) - \int_0^t \mathrm{d}u \sum_{i\in[n]} \overline{G}_{\xi_u} \diamond_i \varphi(\mathbf{Y}_u)\right| \\
&\qquad\qquad \leq \sum_{i\in[n]} \int_0^t \mathrm{d}u \left| G^{\ell_i}_{\mathbf{X}_u} \diamond_i \varphi(\mathbf{Y}_u) - \overline{G}_{\mu_{\mathbf{X}_u}} \diamond_i \varphi(\mathbf{Y}_u)\right| \\
&+ \sum_{i\in[n]} \int_0^t \mathrm{d}u \left|\overline{G}_{\mu_{\mathbf{X}_u}} \diamond_i \varphi(\mathbf{Y}_u) - \overline{G}_{\xi_u} \diamond_i \varphi(\mathbf{Y}_u)\right| + \frac{1}{2} \sum_{\substack{i,j\in[n]\\ i\neq j}} \int_0^t \mathrm{d}u \, \left|\partial_{i,j}\varphi(\mathbf{Y}_u) D^{\ell_i,\ell_j}(\mathbf{X}_u)\right|
\end{aligned}$$

The first term on the right-hand side goes to zero from (2.72) in Lemma 2.3.15. The second term goes to zero from the first part of the theorem. For the third, we use (2.65) in Proposition 2.3.13. □

2.3.7 Convergence of the genetic variance

Here, we show Theorem 2.1.3, which states that the population trait variance converges as $L \to +\infty$ to the genetic variance.

Theorem 2.1.3. *Set* $\varepsilon_L := \frac{1}{\sqrt{\rho r^{**}}}$ *and define the genetic variance* $\sigma_t^2 := 4\mathbb{E}\left[f_t(1-f_t)\right]$ *where* $(f_t)_{t\in[0,T]}$ *is solution to (2.10). Under the assumptions of Theorem 2.1.1 we have*

$$\mathbb{E}\left[\sup_{t\in[\varepsilon_L,T]} \left|L\mathbf{Var}_{\mathbf{X}_t}[Z(g)] - \sigma_t^2\right|\right] \longrightarrow 0$$

Proof. First, note that

$$\mathbf{Var}_{\mathbf{X}_t}[Z(g)] = \frac{1}{L^2} \sum_{\ell_1,\ell_2\in[L]} \mathbf{Cov}_{\mathbf{X}_t}[g_{\ell_1}, g_{\ell_2}]$$

This implies

$$\begin{aligned}
&\left|L\mathbf{Var}_{\mathbf{X}_t}[Z(g)] - 4\mathbb{E}\left[f_t(1-f_t)\right]\right| \\
&\quad \leq \left|\frac{1}{L}\sum_{\ell\in[L]} \mathbf{Var}_{\mathbf{X}_t}[g_\ell] \;-\; 4\mathbb{E}\left[f_t(1-f_t)\right]\right| \;+\; \frac{1}{L} \sum_{\substack{\ell_1,\ell_2\in[L]\\ \ell_1\neq\ell_2}} \left|\mathbf{Cov}_{\mathbf{X}_t}[g_{\ell_1}, g_{\ell_2}]\right| \\
&\quad \leq \left|\frac{1}{L}\sum_{\ell\in[L]} 4p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t)) \;-\; 4\mathbb{E}\left[f_t(1-f_t)\right]\right| + \frac{1}{L} \sum_{\substack{\ell_1,\ell_2\in[L]\\ \ell_1\neq\ell_2}} \left|D^{\ell_1,\ell_2}(\mathbf{X}_t)\right| \\
&\qquad\qquad \leq\; 4\left|< \mu_{\mathbf{X}_t}, \mathrm{Id}(1-\mathrm{Id}) > - \mathbb{E}\left[f_t(1-f_t)\right]\right| \;+ \frac{1}{L} \sum_{\substack{\ell_1,\ell_2\in[L]\\ \ell_1\neq\ell_2}} \left|D^{\ell_1,\ell_2}(\mathbf{X}_t)\right|
\end{aligned}$$

where in the second inequality we used that under $\mathbf{X}_t$, g_ℓ has law $2\mathrm{Ber}(p^\ell(\mathbf{X}_t)) - 1$. Taking the expectation on both sides of the inequality, the convergence of the first term on the right-hand side follows from Theorem 2.1.1. For the second term we use (2.62) in Proposition 2.3.13. □

Chapter 3

An ancestral process in an infinite population

Contents

Ancestral processes are useful tools to build intuition about the behavior of a population by describing the ancestry of a randomly sampled organism at present. Unfortunately, no such process exists for the LD-Wright-Fisher diffusion (2.3). When $S = 0$, the dual process is the **discrete ancestral recombination graph** [160]. When $R = 0$, the dual process is the **ancestral selection graph** [161]. The dual of the LE-Wright-Fisher diffusion (1.7) with quadratic fitness was obtained in [162]. In this chapter we will present an ancestral process when $\Sigma = 0$ (no genetic drift). This process helps build intuition about the evolution of a population and the interaction between selection and recombination, but the greatest interest of this ancestral process is negative: in Section 3.2.2, we will see that it is simply not possible to find an ancestral process to the general LD-Wright-Fisher diffusion with genetic drift. The ancestral process we will study works as follows (see Figure 3.1): to sample a genome at time T, we do

1. Go back to time 0 under the action of recombination and mutation

2. Sample ancestral genomic components with law $\mathbf{X}_0$ and mutations with law $(\mathcal{L}_\ell)_{\ell\in[L]}$.
3. Going forward in time, decide selection events based on the deterministic probabilities that a block is outcompeted by an organism randomly sampled within the population.

This method can help clarifying the use by articles such as [163] of heuristics to determine the "effective chromosome block length", that is, the typical length of a chromosome block on which the effect of selection dominates that of recombination.

3.1 Definitions and notation

We will use the notation $\Box_{[L]} := \{-1,+1\}^L, \mathbb{X}^{[L]} := \mathcal{P}(\Box_{[L]})$ from Section 1.2.2 (though the model can trivially be extended to a finite number of alleles per locus as long as mutations are specified by the House-of-Cards model [164]). We recall that for $\mathbf{X} \in \mathbb{X}^{[L]}$, $\mathbf{X}^A$ is the marginal of $\mathbf{X}$ on $\Box_A := \{-1,1\}^A$. We recall the recombinator R, the selector S and the mutator Θ

$$R :\mathbf{x} \longmapsto \sum_{\emptyset\subsetneq\mathcal{I}\subsetneq[L]} \nu(\mathcal{I})(\mathbf{x}^{\mathcal{I}} \otimes \mathbf{x}^{\mathcal{I}^c} - \mathbf{x}) \tag{3.1}$$

$$S :\mathbf{x} \longmapsto \left(\mathbf{Cov}_{\mathbf{x}}\left[W(g), \mathbb{1}_{[g=\gamma]}\right]\right)_{\gamma\in\Box_{[L]}} \tag{3.2}$$

$$\Theta :\mathbf{x} \longmapsto \sum_{\ell\in[L]} |\theta_\ell| \left(\mathbf{x}^{[L]\setminus\{\ell\}} \otimes \mathcal{L}_\ell - \mathbf{x}\right) \tag{3.3}$$

where $W : \Box_{[L]} \to \mathbb{R}$ is the log-fitness function (on which no assumption is made), ν is the recombination measure, $\theta_\ell^+, \theta_\ell^-$ are mutation rates, and we define the total mutation rate and the mutational distribution at locus ℓ as

$$|\theta_\ell| := \theta_\ell^+ + \theta_\ell^- \qquad\qquad \mathcal{L}_\ell := \frac{\theta_\ell^-}{|\theta_\ell|}\delta_{-1} + \frac{\theta_\ell^+}{|\theta_\ell|}\delta_{+1}$$

We consider that ν is non-degenerate, that is, that for any pair $\ell_1, \ell_2 \in [L]$ with $\ell_1 \neq \ell_2$, $\nu^{\{\ell_1,\ell_2\}}(\{\ell_1\}) > 0$ where ν^A is the marginal of ν on A. We define $\hat{\nu}^A$ as the marginal of ν on A, excluding $(\emptyset, A)$, that is

$$\forall \mathcal{I} \subseteq A, \qquad \hat{\nu}^A(\mathcal{I}) = \nu^A(\mathcal{I})\mathbb{1}_{[\mathcal{I}\notin\{\emptyset,A\}]}. \tag{3.4}$$

We recall the following definition for any subset $A \subseteq [L]$

$$\beta_A := \sum_{\emptyset\subsetneq\mathcal{I}\subsetneq A} \nu^A(\mathcal{I}). \tag{3.5}$$

Throughout, we fix $(\mathbf{X}_t)_{t\in[0,T]}$ as the solution to the recombination-selection-mutation equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{X}_t = R(\mathbf{X}_t) + S(\mathbf{X}_t) + \Theta(\mathbf{X}_t) \tag{3.6}$$

with initial condition $\mathbf{X}_0 \in \mathbb{X}^{[L]}$.

In particular, with respect to the LD-Wright-Fisher diffusion (1.5), we take the recombination rate ρ to be equal to 1. In this setting, the strong recombination limit corresponds to the limit $\max\limits_{\ell\in[L]} |\theta_\ell| + \max\limits_{\Box_{[L]}} |W| \ll 1$.

3.1.1 Forward and backwards processes.

We will be handling processes which are either defined from an initial condition at time 0 (forward in time) or from an initial condition at time T (backwards in time). To make things more convenient, we will always denote forward time with the letter t or τ and backwards time with the letter u or v.

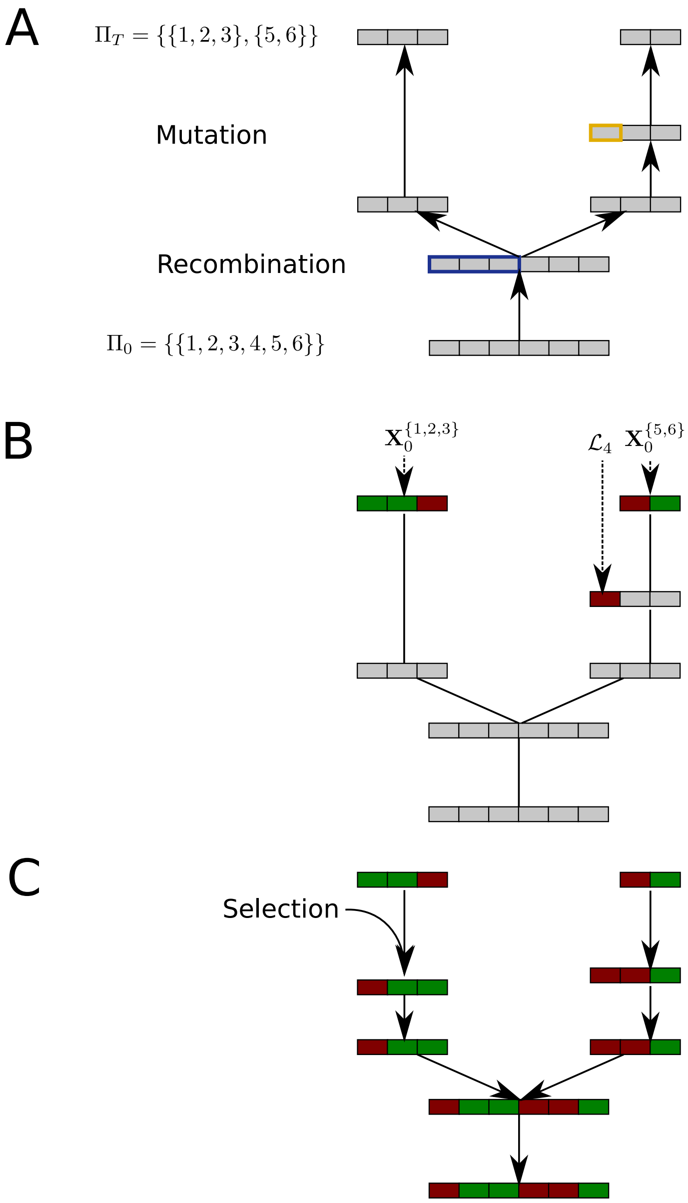

Figure 3.1: The ancestral process. First, we go backwards in time under mutation and selection. Then we pick the ancestors and the mutants. Then we go forward in time, with possible selection events replacing a block with a different one.

3.1.2 Sub-partitions and recombination-mutation process

A **sub-partition** of $[L]$ is a set of disjoint subsets of $[L]$. We let $\mathscr{S}$ be the set of sub-partitions of $[L]$. We let $\Pi \setminus_r \mathcal{I}$ be the subpartition obtained by removing $\mathcal{I}$ from the block of Π. Formally

$$\Pi \setminus_r \mathcal{I} := \{A \cap ([L] \setminus \mathcal{I}) \mid A \in \Pi, A \neq \mathcal{I}\}.$$

We say a subpartition Π_1 is **finer** than a subpartition Π_2 if for any block $\mathcal{I} \in \Pi_1$, we have a block $\mathcal{J} \in \Pi_2$ with $\mathcal{I} \subseteq \mathcal{J}$. We then write $\Pi_1 \preceq \Pi_2$. A right-continuous Markov process $(\Pi_u)_{u\in[0,T]}$ such that $\Pi_{u_1} \preceq \Pi_{u_2}$ whenever $u_1 \leq u_2$ is called a **subpartitioning process**.

We can now define the recombination-mutation process, which takes us backwards from time T to time 0.

Definition 3.1.1 (Recombination-mutation process). *For any subset $A \subseteq [L]$, we define the **recombination-mutation process** started at $\Xi \in \mathscr{S}$ as the subpartitioning process $(\Pi_u)_{u\in[0,T]}$ such that $\Pi_0 = \Xi$, and with transition rates*

1. *$\Pi_u \to \Pi_u \setminus_r \{\ell\}$ with rate $|\theta_\ell|$, for each $\ell \in \Pi_u$.*
2. *$\Pi_u \to (\Pi_u \setminus \mathcal{I}) \cup \{\mathcal{I} \cap \mathcal{J}, \mathcal{I} \cap \mathcal{J}^c\}$ with rate $\nu^{\mathcal{I}}(\mathcal{J})$, for each $\mathcal{I} \in \Pi_u$ and subset $\mathcal{J}$ with $\emptyset \subsetneq \mathcal{J} \subsetneq \mathcal{I}$.*

We fix a probability space $(\Omega, \mathscr{F}, \mathbb{P}_{[0,T],\Xi})$ on which a process $(\Pi_u)_{u\in[0,T]}$ is defined such that for any $\Xi \in \mathscr{S}$, $(\Pi_u)_{u\in[0,T]}$ under $\mathbb{P}_{[0,T],\Xi}$ is the recombination-mutation process started at Ξ.

We let $(\mathscr{F}_u^-)_{u\in[0,T]}$ be the filtration of $(\Pi_u)_{u\in[0,T]}$, called the **backwards filtration**.

3.1.3 Warm-up: duality for the recombination-mutation process

The following Theorem is taken from [160] when there is no coalescence ($\Sigma = 0$).

Theorem 3.1.2. *The recombination-mutation process $\mathbf{X}_t$ solution to*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{X}_t = R(\mathbf{X}_t) + \Theta(\mathbf{X}_t)$$

is dual to the process Π_u. In particular, sample $\mathcal{G} \in \square_{[L]}$ such that conditional on $\mathscr{F}_T^-$, we have

- *$\forall A \in \Pi_T$, $\mathcal{G}^{|A}$ has law $\mathbf{X}_0^A$*
- *for any $\ell \in [L]$ not in Π_T, $\mathcal{G}_\ell$ has law $\mathcal{L}_\ell$.*

Then $\mathcal{G}$ has law $\mathbf{X}_T$.

3.1.4 Genomic components process

We let $\tilde{\Pi}_t := \Pi_{T-t}$ be the forward-time process associated with Π. We recall that for an element $\gamma \in \square_{[L]}$ and a subset $\mathcal{I} \subseteq [L]$, $\gamma^{|\mathcal{I}} = (\gamma_\ell)_{\ell\in\mathcal{I}}$ is the restriction of γ to $\mathcal{I}$.

Definition 3.1.3 (Genomic components process). *The genomic components process is the $\square_{[L]}$−valued process $(\mathcal{G}_t)_{t\in[0,T]} \equiv ((\mathcal{G}_{t,\ell})_{\ell\in[L]})_{t\in[0,T]}$ generated conditional on $(\Pi_u)_{u\in[0,T]}$ as follows*

- *For any $A \in \tilde{\Pi}_0$, $(\mathcal{G}_0^{|A})$ is independently sampled with law $\mathbf{X}_0^A$.*
- *For ℓ not in $\tilde{\Pi}_0$, $\mathcal{G}_{0,\ell}$ is independently sampled with law $\mathcal{L}_\ell$.*

- *At time $t \in [0,T]$, if $A \in \tilde{\Pi}_t$, then for any $\gamma \in \Box_A$, we have the transition $\mathcal{G}_t^{|A} \to \gamma$ with rate*
$$\left(\mathbf{X}_t\left[W(g) \mid g^{|A} = \gamma\right] - \mathbf{X}_t\left[W(g) \mid g^{|A} = \mathcal{G}_t^{|A}\right]\right)_+ \mathbf{X}_t[\mathbb{1}_{[g^{|A}=\gamma]}]$$
*where $\mathbf{X}_t[\cdot|\cdot]$ is the conditional expectation and $a_+ := \max\{a,0\}$. Such a transition is called a **selective collision on the block** A of $\tilde{\Pi}_t$.*

The smallest filtration $(\mathscr{F}_t^+)_{t\in[0,T]}$ such that $\mathscr{F}_T^- \subset \mathscr{F}_0^+$ and $(\mathcal{G}_t)_{t\in[0,T]}$ is $\mathscr{F}^+-$adapted is called the forward filtration.

3.1.5 No-singletons genomic components process

Here we build another ancestral process, which only accounts for collisions on blocks of size greater than 2. Specifically, conditional on $(\Pi_u)_{u\in[0,T]}$, we define for $\ell \in [L]$ the last time at which ℓ is in a singleton as

$$\tau_\ell = \sup\{t \in [0,T], \quad \{\ell\} \in \tilde{\Pi}_t\} \tag{3.7}$$

with $\sup \emptyset = 0$. We then define $(\mathcal{H}_t)_{t\in[0,T]}$ as follows

- Initial conditions:
  - If $A \in \tilde{\Pi}_0$, then sample $\mathcal{H}_0^{|A}$ with distribution $\mathbf{X}_0^A$.
  - For ℓ such that $\tau_\ell > 0$, sample $\mathcal{H}_{0,\ell}$ with law $\mathbf{X}_{\tau_\ell}^{\{\ell\}}$
  - If ℓ is such that $\mathcal{H}_{0,\ell}$ was not yet sampled, sample it with law $\mathcal{L}_\ell$.
- For $t \in [0,T]$, if $A \in \tilde{\Pi}_t$ is such that $\#A \geq 2$, then $\mathcal{H}_t^{|A}$ jumps to $\gamma \in \Box_A$ with rate
$$\left(\mathbf{X}_t\left[W(g) \mid g^{|A} = \gamma\right] - \mathbf{X}_t\left[W(g) \mid g^{|A} = \mathcal{H}_t^{|A}\right]\right)_+ \mathbf{X}_t[\mathbb{1}_{[g^{|A}=\gamma]}]$$

3.2 Main result

We state the main result fixing an initial condition Ξ for $(\Pi_u)_{u\in[0,T]}$.

Theorem 3.2.1. *The law of $\mathcal{G}_T$ under $\mathbb{P}_{[0,T],\Xi}$ is $\bigotimes_{\mathcal{I}\in\Xi}\mathbf{X}_T^{\mathcal{I}}$. Furthermore, consider υ a $[0,T]-$valued $\mathscr{F}^-{-}$stopping time. Then conditional on $\mathscr{F}_\upsilon^-$, the $(\mathcal{G}_{T-\upsilon}^{|\mathcal{I}})_{\mathcal{I}\in\Pi_\upsilon}$ are independent $\mathscr{F}_{T-\upsilon}^+$-measurable variables such that $\mathcal{G}_{T-\upsilon}^{|\mathcal{I}}$ has law $\mathbf{X}_{T-\upsilon}^{\mathcal{I}}$.*
The same holds replacing $(\mathcal{G}_t)_{t\in[0,T]}$ with $(\mathcal{H}_t)_{t\in[0,T]}$.

The proof is in Section 3.3.

One conceptually interesting consequence of this Theorem is that LD between loci is created by selection collisions on $(\mathcal{H}_t)_{t\in[0,T]}$. In particular, if recombination is strong, such collisions will become rarer, which implies that the population will get closer to LE. Conditional on a rare collision having occurred, we may ask whether this collision took place on a large or a small block.

3.2.1 Application: the fundamental difference between free recombination and single-point crossing-over

Here we will briefly illustrate one application of Theorem 3.2.1. We will consider two possible recombination mechanisms, namely free recombination ν_{fr} and single-point crossing-over ν_{sp} defined with

$$\nu_{fr} := \frac{1}{2^L}\sum_{\mathcal{I}\subseteq[L]}\delta_{\mathcal{I}} \qquad\qquad \nu_{sp} := \frac{1}{L-1}\sum_{\ell\in[L-1]}\delta_{[\ell]}$$

We pick a fitness function corresponding to stabilizing/disruptive selection

$$W(g) = \kappa Z(g)^2 \qquad\qquad Z(g) = \sum_{\ell\in[L]} g_\ell$$

for some parameter $\kappa \in \mathbb{R}$. Under stabilizing selection ($\kappa < 0$), the appearance of LD is known as the Bulmer effect [98], and under disruptive selection ($\kappa > 0$) the appearance of LD has been compared to the paramagnetic/ferromagnetic phase transition of the Ising model [107]. We will heuristically show the following dichotomy

- Under free recombination, if we are close to LE, most selection collisions occur on large blocks.
- Under single-point crossing-over, if we are close to LE, most selection collisions occur on small blocks.

Method

Fix a locus $\ell_0 \in [L]$, assuming $L \gg 1$. For $u \in [0,T]$, write $A_{\ell_0}(u)$ for the block of Π_u containing ℓ_0 with $A_{\ell_0}(u) = \emptyset$ if ℓ_0 is not in Π_u. Then we know from Theorem 3.2.1 that $\mathcal{G}_{T-u}$ has law $\mathbf{X}_{T-u}^{A_{\ell_0}(u)}$. In particular, the rate at which a selection collision occurs at $\mathcal{G}_{T-u}^{A_{\ell_0}(u)}$ is $V_{A_{\ell_0}(u),T-u}$ where for $A \subseteq [L]$ we define the rate of collisions

$$V_{A,t} := \mathbf{X}_t^{A\otimes 2}\left[\left|\mathbf{X}_t\left[W(g) \mid g^{|A} = g^{(1)}\right] - \mathbf{X}_t\left[W(g) \mid g^{|A} = g^{(2)}\right]\right|\right]$$

where under $\mathbf{X}_t^{A\otimes 2}$, $g^{(1)}, g^{(2)}$ are two i.i.d random variables with law $\mathbf{X}_t^A$.

Understanding the appearance of linked selection requires tackling two questions

- What is the typical size of $A_{\ell_0}(u)$?
- What is the typical value of $V_{A,t}$ for a given block A ?

In other words, we want to approximate the distribution of selective collisions on $[0,T]$ with

$$D_{\ell_0}(\mathrm{d}u) := V_{A_{\ell_0}(u),T-u}\mathrm{d}u.$$

Size of the $A_{\ell_0}(u)$

The typical size of $A_{\ell_0}(u)$ can be approximated deterministically. We claim

Claim 3.2.2. *Let us assume $\Theta = 0$ (negligible mutations), and $L \gg 1$. Under free recombination if $0 \leq u_L$, we have*

$$\mathbb{E}\left[\#A_{\ell_0}(u_L)\right] = 1 + (L-1)e^{-u_L/2}.$$

Under single-point crossing-over if $1 \ll u_L \leq L$ we have

$$\#A_{\ell_0}(u_L) \asymp \frac{L}{1+u_L}.$$

Here, the notation $X \asymp Y$ means X/Y is tight and does not converge to zero.

Proof under free recombination. The process $(A_{\ell_0}(u))_{u\in[0,T]}$ is a Markov process such that the transition $\mathcal{I} \to \mathcal{J}$ occurs with rate $\nu^{\mathcal{I}}(\mathcal{J})$ for $\ell_0 \in \mathcal{J} \subsetneq \mathcal{I} \subseteq [L]$.

For free recombination, one easily finds $\nu^{\mathcal{I}}(\mathcal{J}) = 2^{1-\#\mathcal{I}}$. In particular, the process $(\#A_{\ell_0}(u))_{u\in[0,T]}$ is a pure-death process with initial condition L and transition rates

$$i \to j \text{ with rate } \binom{i}{j} 2^{1-i} \text{ for } i > j \geq 1.$$

This can be represented as a process $(N_u)_{u \geq 0}$ such that at rate 1, N_u jumps to

$$1 + \sum_{i \in [N_u - 1]} B_i$$

for some independent *Bernoulli*$(1/2)$ variables $(B_i)_{i \in [N_u]}$. At time u_L, the number K_{u_L} of jumps of N_{u_L} is *Poisson*(u_L) distributed. Furthermore, we have

$$N_{u_L} = \sum_{i \in [L]} \prod_{n \in [K_{u_L}]} B_{i,n}$$

for i.i.d Bernoulli variables $(B_{i,n})_{i \in [L], n \in [K_{u_L}]}$. In particular,

$$\mathbb{E}\left[N_{u_L}\right] = 1 + (L-1)\mathbb{E}\left[2^{-K_{u_L}}\right]$$

Because K_{u_L} is a *Poisson*(u_L) variable, we have $\mathbb{E}\left[2^{-K_{u_L}}\right] = e^{-u_L/2}$. This yields the result. □

Proof under single-point crossing-over. Under single-point crossing-over, let us consider the left-hand size of $A_{\ell_0}(u)$ defined as

$$\#_l A_{\ell_0}(u) := \#([\ell_0 - 1] \cap A_{\ell_0}(u)).$$

For $n < \ell_0$, the event $[\#_l A_{\ell_0}(u_L) > n]$ corresponds to the fact that Π no recombination events on the block $\{\ell_0 - n - 1, \dots, \ell_0\}$ during the interval $[0, u_L]$. This is the probability that a Poisson variable with parameter $(n-1)u_L$ is equal to zero, and therefore

$$\mathbb{P}[\#_l A_{\ell_0}(u_L) > n] = e^{-(n-1)u_L}.$$

In particular, because $L \gg 1$, we see that $\#_l A_{\ell_0}$ is of order L/u_L. A symmetric reasoning yields the result for the right-hand size of $A_{\ell_0}(u)$. □

Rate of collisions

Let us now compute $V_{A,t}$ for A such that $\#A \gg 1$. We assume

- that $\mathbf{X}_t[g_\ell] = 0$ for all $\ell \in [L]$
- that $\mathbf{X}_t \simeq \bigotimes_{\ell \in [L]} \mathbf{X}_t^{\{\ell\}}$ (the population is close to LE)
- that $\mathbf{X}_t \simeq \mathbf{X}_0$ (the population does not evolve much on the timescale considered).

The first assumption is made for simplicity. If we consider a large subset $A \subseteq [L]$, we can approximate $Z(g)$ under $\mathbf{X}_t$ with $z^A + z^{A^c}$, where z^A and z^{A^c} are independent normally distributed variables with mean zero and respective variances $\sigma_A^2 = \frac{\#A}{L}\sigma_L^2$ and $\sigma_{A^c}^2 = \frac{(L - \#A)}{L}\sigma_L^2$, where σ_L^2 is the variance of g_ℓ for a typical locus ℓ.

We then compute

$$V_{A,t} \simeq |\kappa| \mathbb{E}\left[\left|\mathbb{E}\left[(z_1^A + z^{A^c})^2 \mid z^{A^c}\right] - \mathbb{E}\left[(z_2^A + z^{A^c})^2 \mid z^{A^c}\right]\right|\right]$$

where z_1^A, z_2^A are independent with law $\mathcal{N}(0, \sigma_A^2)$. We find

$$V_{A,t} \simeq |\kappa| \mathbb{E}\left[\left|(z_1^A - z_2^A)(z_1^A + z_2^A)\right|\right]$$

which yields

$$V_{A,t} \asymp |\kappa| \sigma_A^2 \simeq |\kappa| \sigma_L^2 \frac{\#A}{L}.$$

where $\asymp$ is heuristically used from now on to describe the expected order of magnitude of $V_{A,t}$.

Conclusion: distribution of collisions

We find that $D_{\ell_0}(\mathrm{d}u)$ is of order

$$|\kappa|\sigma_L^2 \frac{\# A_{\ell_0}(u)}{L}\,\mathrm{d}u$$

We may now state the following crucial dichotomy

- Under **free recombination**, we have for $u \in [0, \log_2(L)]$

  $$D_{\ell_0}(\mathrm{d}u) \asymp |\kappa|\sigma_L^2 e^{-u/2}\mathrm{d}u$$

  The total rate of collisions is

  $$\int_0^{\log_2(L)} D_{\ell_0}(\mathrm{d}u) \asymp |\kappa|\sigma_L^2$$

  In particular, selection collisions may appear as soon as $|\kappa|\sigma_L^2$ is non-negligible, and the time at which these selection events occur is distributed as an *Exponential*(ln(2)) variable. At such a time, the typical size of A_{ℓ_0} is $Le^{-u/2} \asymp L$.

- Under **single-point crossing-over**, we have for $u \in [0, L]$

  $$D_{\ell_0}(\mathrm{d}u) \asymp |\kappa|\sigma_L^2 \frac{\mathrm{d}u}{1+u}$$

  The total rate of collisions is

  $$\int_0^{L} D_{\ell_0}(\mathrm{d}u) \asymp |\kappa|\sigma_L^2 \ln(L)$$

  In particular, selection collisions may appear as soon as $|\kappa|\sigma_L^2 \ln(L)$ is non-negligible. The time at which these selection collisions occur is of order L. At such a time, the typical size of A_{ℓ_0} is of order 1.

3.2.2 Why can't we add genetic drift ?

The algorithm used to generate $\mathcal{G}_T$ seems at first glance like it could easily incorporate genetic drift: one would simply need to allow coalescence in the subpartition process $(\Pi_u)_{u\in[0,T]}$. That is, one would allow two blocks of Π_u to coagulate with rate 1. This is formally possible, and the resulting object may present interesting features. Nevertheless, Theorem 3.2.1 cannot hold in this setting because the process $(\mathbf{X}_t)_{t\in[0,T]}$ used to define $(\mathcal{G}_t)_{t\in[0,T]}$ should not be independent of the coalescence process. This is because the coalescent process comes down from infinity, which means a coalescence event between two organisms some time u in the past which affects the present potentially impacts a significant fraction of the population [165]. To put it another way, we should either generate $(\mathbf{X}_t)_{t\in[0,T]}$ conditional on $(\Pi_u)_{u\in[0,T]}$ or the other way around, before we can generate $(\mathcal{G}_t)_{t\in[0,T]}$.

3.3 Proof of Theorem 3.2.1

We will obtain the result for $(\mathcal{G}_t)_{t\in[0,T]}$ assuming no mutations ($\Theta = 0$), and discuss how to adapt the proof for mutations in Section 3.3.6. In Section 3.3.7, we obtain as a corollary the result for $(\mathcal{H}_t)_{t\in[0,T]}$.

The proof is structured as follows: we construct a forward stochastic process $(\tilde{\mathcal{G}}_t)_{t\in[0,T]}$ such that $\tilde{\mathcal{G}}_t$ has law $\mathbf{X}_t^A$ for some subset $A \subseteq [L]$, then we show that we can construct a process $(\tilde{\mathcal{G}}_t)_{t\in[0,T]}$ from Poisson point processes. Standard properties from Poisson point processes let us reverse time, that is, we may generate τ the "last recombination event on $[0, T]$", and then generate $(\tilde{\mathcal{G}}_t)_{t\in[\tau,T]}$ conditional on this last recombination event. Iterating this construction, we obtain $(\mathcal{G}_t)_{t\in[0,T]}$.

3.3.1 The marginal forward genome process

We fix a subset $A \subseteq [L]$.

For $\mathcal{I} \subseteq A$, $\gamma^{(1)} \in \square_{\mathcal{I}}$, $\gamma^{(2)} \in \square_{A \setminus \mathcal{I}}$, we let $\gamma^{(1)} \otimes \gamma^{(2)}$ be the element of $\square_A$ such that its restriction to $\mathcal{I}$ is $\gamma^{(1)}$ and its restriction to $A \setminus \mathcal{I}$ is $\gamma^{(2)}$.

For $A \subseteq [L]$, we define the A-**marginal forward genome process** to be the right-continuous $\square_A-$valued Markov process $(\tilde{\mathcal{G}}_t)_{t\in[0,T]}$ with initial condition $\tilde{\mathcal{G}}_0$ sampled with law $\mathbf{X}_0^A$ and transition rates

- $\tilde{\mathcal{G}}_t \to \tilde{\mathcal{G}}_t^{|\mathcal{I}} \otimes \gamma$ with rate $\hat{\nu}^A(\mathcal{I}) X_t^{A\setminus\mathcal{I}}(\gamma)$ for any $\mathcal{I}$ with $\emptyset \subsetneq \mathcal{I} \subsetneq A$ and $\gamma \in \square_{A\setminus\mathcal{I}}$.
- $\tilde{\mathcal{G}}_t \to \gamma$ with rate $(\mathbf{X}_t[W(g)|g^{|A} = \gamma] - \mathbf{X}_t[W(g)|g^{|A} = \tilde{\mathcal{G}}_t])_+ X_t^A(\gamma)$ for $\gamma \in \square_A$.

Proposition 3.3.1. *If $\tilde{\mathcal{G}}_t$ is the $A-$marginal forward genome process, then the law of $\tilde{\mathcal{G}}_t$ is $\mathbf{X}_t^A$.*

Proof. We prove the result for $A = [L]$ to alleviate notation, the general result follows from the same computations using the marginal operators $\hat{R}^A$ and S^A from Proposition 2.3.3.

We simply compute the generator $\tilde{Q}_t$ of $\tilde{\mathcal{G}}_t$ at time t, seen as a transition-rate matrix $(\tilde{Q}_t(\gamma^{(1)}, \gamma^{(2)}))_{\gamma_1,\gamma^{(2)}\in\square_{[L]}}$ such that $\tilde{Q}_t(\gamma^{(1)}, \gamma^{(2)})$ is the rate of transition from $\gamma^{(1)}$ to $\gamma^{(2)}$. We write the decomposition

$$\tilde{Q}_t = \tilde{R}_t + \tilde{S}_t$$

where $\tilde{R}_t, \tilde{S}_t$ correspond to the two types of jumps of $\tilde{\mathcal{G}}_t$. We must show

$$(\tilde{R}_t)^\top \mathbf{X}_t = R(\mathbf{X}_t) \qquad\qquad (\tilde{S}_t)^\top \mathbf{X}_t = S(\mathbf{X}_t)$$

where $^\top$ denotes matrix transposition and $\mathbf{X}_t$ is seen as a column vector.

It is easy to find that if $\gamma^{(1)}, \gamma^{(2)} \in \square_{[L]}$ with $\gamma^{(1)} \neq \gamma^{(2)}$, then the recombination jump from $\gamma^{(1)}$ to $\gamma^{(2)}$ occurs with rate

$$\tilde{R}_t(\gamma^{(1)}, \gamma^{(2)}) = \sum_{\emptyset\subsetneq\mathcal{I}\subsetneq A} \nu(\mathcal{I}) \mathbb{1}_{[\gamma^{(2)|\mathcal{I}} = \gamma^{(1)|\mathcal{I}}]} X_t^{\mathcal{I}^c}(\gamma^{(2)|\mathcal{I}^c})$$

and the diagonal coefficient is

$$\tilde{R}_t(\gamma^{(1)}, \gamma^{(1)}) = -\sum_{\gamma^{(2)}\in\square_{[L]}\setminus\{\gamma^{(1)}\}} \tilde{R}_t(\gamma^{(1)}, \gamma^{(2)})$$

From there we obtain for any $\gamma \in \square_{[L]}$

$$\begin{aligned}
\left((\tilde{R}_t)^\top \mathbf{X}_t\right)(\gamma) &= \sum_{\gamma'\in\square_{[L]}\setminus\{\gamma\}} \tilde{R}_t(\gamma', \gamma) X_t(\gamma') - \tilde{R}_t(\gamma, \gamma') X_t(\gamma) \\
&= \sum_{\emptyset\subsetneq\mathcal{I}\subsetneq A} \nu(\mathcal{I}) \sum_{\gamma'\in\square_{[L]}\setminus\{\gamma\}} \Big(\mathbb{1}_{[\gamma^{|\mathcal{I}} = \gamma'^{|\mathcal{I}}]} X_t^{\mathcal{I}^c}(\gamma^{|\mathcal{I}^c}) X_t(\gamma') \\
&\qquad\qquad - \mathbb{1}_{[\gamma'^{|\mathcal{I}} = \gamma^{|\mathcal{I}}]} X_t^{\mathcal{I}^c}(\gamma'^{|\mathcal{I}^c}) X_t(\gamma)\Big) \\
&= \sum_{\emptyset\subsetneq\mathcal{I}\subsetneq A} \nu(\mathcal{I}) \left(X_t^{\mathcal{I}}(\gamma^{|\mathcal{I}}) X_t^{\mathcal{I}^c}(\gamma^{|\mathcal{I}^c}) - X_t(\gamma)\right)
\end{aligned}$$

which corresponds to the recombinator R.

Similarly, if $\gamma^{(1)}, \gamma^{(2)} \in \square_{[L]}$ with $\gamma^{(1)} \neq \gamma^{(2)}$, then the selection jump from $\gamma^{(1)}$ to $\gamma^{(2)}$ occurs with rate

$$\tilde{S}_t(\gamma^{(1)}, \gamma^{(2)}) = \left(W(\gamma^{(2)}) - W(\gamma^{(1)})\right)_+ X_t(\gamma^{(2)})$$

and the diagonal coefficient is

$$\tilde{S}_t(\gamma^{(1)}, \gamma^{(1)}) = - \sum_{\gamma^{(2)} \in \square_{[L]} \setminus \{\gamma^{(1)}\}} \tilde{S}_t(\gamma^{(1)}, \gamma^{(2)}).$$

It follows for any $\gamma \in \square_{[L]}$

$$\begin{aligned}\left((\tilde{S}_t)^\top \mathbf{X}_t\right)(\gamma) &= \sum_{\gamma' \in \square_{[L]} \setminus \{\gamma\}} \tilde{S}_t(\gamma', \gamma) X_t(\gamma') - \tilde{S}_t(\gamma, \gamma') X_t(\gamma) \\ &= \sum_{\gamma' \in \square_{[L]} \setminus \{\gamma\}} \Big(\big(W(\gamma) - W(\gamma')\big)_+ X_t(\gamma) X_t(\gamma') \\ &\qquad\qquad - \big(W(\gamma') - W(\gamma)\big)_+ X_t(\gamma') X_t(\gamma) \Big)\end{aligned}$$

which can be rewritten

$$\begin{aligned}\left((\tilde{S}_t)^\top \mathbf{X}_t\right)(\gamma) &= \sum_{\gamma' \in \square_{[L]} \setminus \{\gamma\}} \left(W(\gamma) - W(\gamma')\right) X_t(\gamma) X_t(\gamma') \\ &= W(\gamma) X_t(\gamma) - \left(\sum_{\gamma' \in \square_{[L]}} W(\gamma') X_t(\gamma') \right) X_t(\gamma) \\ &= \mathbf{Cov}_{\mathbf{X}_t}[W(g), \mathbb{1}_{[g=\gamma]}]\end{aligned}$$

which yields the result. ☐

3.3.2 Construction from Poisson point processes

Here we construct $\tilde{\mathcal{G}}_T$ from two independent Poisson point processes (PPP) representing recombination and selection. Specifically, we let η^R, η^S be PPP defined as follows

- η^R is a PPP on $[0, T] \times \mathscr{P}(A)$ with intensity $\mathrm{d}t \otimes \hat{\nu}^A$, where $\mathscr{P}(A)$ is the set of subsets of A.
- η^S is a PPP on $[0, T] \times \square_A \times \mathbb{R}_+$ with intensity $\sum_{\gamma \in \square_A} \mathrm{d}t \otimes \delta_\gamma \otimes \mathrm{d}x$.

We construct a process $(\hat{\mathcal{G}}_t)_{t \in [0,T]}$ as follows

- Sample $\hat{\mathcal{G}}_0$ with law $\mathbf{X}_0$.
- If η^R has a point $(t, \mathcal{I})$ then sample g'_t with law $\mathbf{X}_t^{A \setminus \mathcal{I}}$, and $\hat{\mathcal{G}}_t^{|A \setminus \mathcal{I}}$ jumps to g'_t
- If η^S has a point (t, γ, x), and if

  $$x < (\mathbf{X}_t[W(g) | g^{|A} = \gamma] - \mathbf{X}_t[W(g) | g^{|A} = \hat{\mathcal{G}}_t]) X_t^A(\gamma)$$

  then $\hat{\mathcal{G}}_t$ jumps to γ.

Proposition 3.3.2. *The process $(\hat{\mathcal{G}}_t)_{t \in [0,T]}$ is equal in law to the process $(\tilde{\mathcal{G}}_t)_{t \in [0,T]}$.*

Proof. This is a standard result from Poisson point processes, see for instance Theorem 7.6 of [166]. ☐

The set on which η^S and η^R have points at the same time-coordinate has probability zero, and will therefore be ignored.

3.3.3 Conditioning on the last recombination event

For $0 \leq a \leq b \leq T$, define $\eta_{(a,b]}$ as the restriction of η to $(a, b] \times E$ with $\eta_{\emptyset} = 0$.

The goal of this section is to show the following proposition

Proposition 3.3.3. *Let $(\tau, \mathcal{I})$ be the last point of η^R on $[0, T]$, with $\tau = 0$ and $\mathcal{I} = [L]$ if η^R is empty.*

Then conditional on $(\tau, \mathcal{I})$, $\hat{\mathcal{G}}_\tau$ has law $\mathbf{X}_\tau^{\mathcal{I}} \otimes \mathbf{X}_\tau^{A \setminus \mathcal{I}}$ and is independent of $\eta^S_{(\tau,T]}$.

This will be obtained as a corollary of the following proposition. For $A \subseteq [L]$, recall β_A the probability that there is a recombination event within A from (3.5).

Proposition 3.3.4. *Let $(\tau, \mathcal{I})$ be the last point of η^R on $[0, T]$. Consider $(\tilde{\tau}, \tilde{\mathcal{I}})$ sampled independently of $(\hat{\mathcal{G}}_t)_{t \in [0,T]}$ such that*

- *$\tilde{\tau} = (0 \vee (T - \mathcal{E}))$ where $\mathcal{E}$ has law Exponential(β_A)*
- *If $\tilde{\tau} = T$, $\tilde{\mathcal{I}} = [L]$. Otherwise $\tilde{\mathcal{I}}$ is independently sampled with law $\frac{1}{\beta_A} \hat{\nu}^A(\cdot)$.*

For $t \in [0, T]$, let $\hat{\mathcal{G}}_{t-}$ be the left limit of $\hat{\mathcal{G}}$ at t with $\hat{\mathcal{G}}_{0-} = \hat{\mathcal{G}}_0$.

Then $\mathscr{L}(\tau, \mathcal{I}, \eta^S, \hat{\mathcal{G}}_{\tau-}, g'_\tau) = \mathscr{L}(\tilde{\tau}, \tilde{\mathcal{I}}, \eta^S, \hat{\mathcal{G}}_{\tilde{\tau}-}, \tilde{\mathcal{G}}'_{\tilde{\tau}})$ where, conditional on $(\tilde{\tau}, \tilde{\mathcal{I}})$, $\tilde{\mathcal{G}}_{\tilde{\tau}'}$ has law $\mathbf{X}_{\tilde{\tau}}^{\tilde{\mathcal{I}}}$.

We need the following Theorem (see Theorem 7.4 of [166]).

Theorem 3.3.5 (Memoryless property of PPP)**.** *Consider some measurable space E, $T > 0$ and a non-zero finite measure Λ on E. Consider η a PPP on $\mathbb{R}_+ \times E$ with intensity $\mathrm{d}t \otimes \Lambda$. Generate a process $\hat{\eta}$ as follows*

- *Generate an exponential variable υ with parameter $\Lambda(E)$.*
- *Generate a random variable Y with law $\Lambda(\cdot)/\Lambda(E)$*
- *Generate an independent PPP η' on $\mathbb{R}_+ \times E$ with intensity $\mathrm{d}t \otimes \Lambda$*
- *Set $\hat{\eta} := \vartheta_\upsilon \eta' + \delta_{(\upsilon, Y)}$, where ϑ_t is the linear time-shift operator defined as $\vartheta_t \delta_{(t_0, y_0)} = \delta_{(t_0 + t, y_0)}$.*

Then $\mathscr{L}(\eta) = \mathscr{L}(\hat{\eta})$ almost surely.

With this theorem, we may prove the following Lemma

Lemma 3.3.6. *Consider some measurable space E, $T > 0$ and a non-zero finite measure Λ on E. Consider a PPP η on $[0, T] \times E$ with intensity $\mathrm{d}t \otimes \Lambda$. Consider υ the first time such that η has a point (υ, Y), with $(\upsilon, Y) = (+\infty, \tilde{Y})$ if η is empty, where $\tilde{Y}$ is independently sampled with law $\Lambda(\cdot)/\Lambda(E)$.*

Now, consider an independent variable $\tilde{\upsilon}$ such that with probability $\mathrm{e}^{-\Lambda(E)}$, $\tilde{\upsilon} = +\infty$, otherwise $\tilde{\upsilon}$ has law Exponential$(\Lambda(E))$ conditioned on being smaller than T.

Then $\mathscr{L}(\eta_{(\upsilon,T]}, \upsilon, Y) = \mathscr{L}(\eta_{(\tilde{\upsilon},T]}, \tilde{\upsilon}, \tilde{Y})$.

Proof. We consider a PPP $\hat{\eta}$ on $\mathbb{R}_+ \times E$, and in particular almost surely

$$\mathscr{L}(\hat{\eta}_{(0,T]}) = \mathscr{L}(\eta).$$

Then Theorem 3.3.5 tells us that $\hat{\eta}$ is equal in distribution to $\vartheta_{\hat{\upsilon}} \eta' + \delta_{(\hat{\upsilon}, \tilde{Y})}$ where η' is independent of $(\hat{\upsilon}, \tilde{Y})$ and $\hat{\upsilon}$ has law *Exponential*$(\Lambda(E))$.

The restriction of $\hat{\eta}$ to $(\upsilon, T]$ is equal in distribution to η_υ, and is also equal in distribution to the restriction of η' to $(0, T - \tilde{\upsilon}]$. The fact that the restriction of η' to $(0, T - \tilde{\upsilon}]$ is equal in distribution to $\eta_{(\tilde{\upsilon},T]}$ follows from the invariance of time-homogeneous PPP by time translations (see Theorem 7.4 of [166]). □

Proof of Proposition 3.3.4. The event $[\tau = 0]$ and $[\tilde{\tau} = 0]$ have the same probability, and in both cases $\hat{\mathcal{G}}_0$ has distribution $\mathbf{X}_0^A$. We therefore work on the events $[\tau > 0]$ and $[\tilde{\tau} > 0]$.

We know that $\hat{\mathcal{G}}_{\tau-}$ is a function of $\eta^R_{(0,\tau)}, \eta^S_{(0,\tau)}$ and $(g'_{t_i})_{i:t_i<\tau}$, where t_i are the times at which $\eta^R_{(0,\tau)}$ has points. In particular, conditional on τ, $\hat{\mathcal{G}}_{\tau-}$ is independent of $\eta^R_{[\tau,T]}$. From Lemma 3.3.6 applied to $\upsilon := T - \tau$, we have

$$\mathscr{L}(\tau, \mathcal{I}, \eta^R_{(0,\tau)}, \eta^S, (g'_{t_i})_{t_i<\tau}, g'_\tau) = \mathscr{L}(\tilde{\tau}, \tilde{\mathcal{I}}, \eta^R_{(0,\tilde{\tau})}, \eta^S, (g'_{t_i})_{i:t_i<\tilde{\tau}}, \tilde{\mathcal{G}}'_{\tilde{\tau}}).$$

This yields the result. □

Proof of Proposition 3.3.3. From Proposition 3.3.4 and Propostion 3.3.2, we know that $\hat{\mathcal{G}}_{\tau-}$ conditional on $(\tau, \mathcal{I}, g'_\tau)$ has law $\mathbf{X}_\tau$. It follows from the definition of $\hat{\mathcal{G}}$ that $\hat{\mathcal{G}}_\tau^{|\mathcal{I}}$ conditioned on τ has law $\mathbf{X}_\tau^{\mathcal{I}}$, whereas $\hat{\mathcal{G}}_\tau^{|A\setminus\mathcal{I}} = g'_\tau$ conditioned on $(\tau, \mathcal{I})$ has law $\mathbf{X}_\tau^{A\setminus\mathcal{I}}$. The independence of $\hat{\mathcal{G}}_\tau$ and $\eta^S_{(\tau,T]}$ conditional on $(\tau, \mathcal{I})$ is obtained from the independence of $\eta^S_{(\tau,T]}$ and $\eta^S_{(0,\tau]}$ conditional on τ, and the fact that $\hat{\mathcal{G}}_\tau$ is only a function of $(\eta^R_{(0,\tau)}, \eta^S_{(0,\tau)}, (g'_{t_i})_{i:t_i\leq\tau}, \mathcal{I})$. □

3.3.4 Consistency property

We now turn to $(\mathcal{G}_t)_{t\in[0,T]}$ to obtain a consistency result. For $0 \leq a < b \leq T$ and $\Xi \in \mathscr{S}$, we write $\mathbb{P}_{[a,b],\Xi}$ for the law of $((\Pi_u)_{u\in[T-b,T-a]}, (\mathcal{G}_t)_{t\in[a,b]})$ defined analogously to the general process, but on the interval $[a, b]$ with initial condition $\Pi_{T-b} = \Xi$.

Proposition 3.3.7. *Consider* $((\Pi_u)_{u\in[\upsilon,T]}, (\mathcal{G}_t)_{t\in[0,T-\upsilon]})$ *with law* $\mathbb{P}_{[0,T],\Xi}$ *for some* $\Xi \in \mathscr{S}$.

For any $\mathscr{F}^-$*−stopping time* υ, $((\Pi_u)_{u\in[\upsilon,T]}, (\mathcal{G}_t)_{t\in[0,T-\upsilon]})$ *conditional on* $\mathscr{F}_\upsilon^-$ *has law* $\mathbb{P}_{[0,T-\upsilon],\Pi_\upsilon}$.

Proof. The fact that $(\Pi_u)_{u\in[\upsilon,T]}$ under $\mathbb{P}_{[0,T],\Xi}$, conditional on $\mathscr{F}_\upsilon^-$, has the same law as $(\Pi_u)_{u\in[0,T-\upsilon]}$ under $\mathbb{P}_{[0,T-\upsilon],\Pi_\upsilon}$ is a standard application of the strong Markov property. Furthermore, by construction $\mathcal{G}_{T-\upsilon}$ conditional on $(\Pi_u)_{u\in[\upsilon,T]}$ is independent of $\mathscr{F}_\upsilon^-$. The result follows. □

3.3.5 Iterative backwards construction and proof of the theorem

Proof of Theorem 3.2.1. We will prove that for $\Xi \in \mathscr{S}$, $\mathcal{G}_T$ under $\mathbb{P}_{[0,T],\Xi}$ has the same law as $(\tilde{\mathcal{G}}_T^{\mathcal{I}})_{\mathcal{I}\in\Xi}$ where $(\tilde{\mathcal{G}}_t^{\mathcal{I}})_{t\in[0,T],\mathcal{I}\in\Xi}$ are independent processes such that $(\tilde{\mathcal{G}}_t^{\mathcal{I}})_{t\in[0,T]}$ is an $\mathcal{I}$−marginal forward genome process. The theorem follows from Proposition 3.3.1 and Proposition 3.3.7.

We prove the result by recursion on Ξ. Suppose Ξ only has singletons

$$\Xi = \{\ell_i, i \in [n]\}.$$

for some $(\ell_i)_{i\in[n]}$. Then η^R has intensity 0 and in particular $\mathcal{G}_T$ has the same law as independent processes $(\tilde{\mathcal{G}}_t^{\mathcal{I}})_{t\in[0,T],\mathcal{I}\in\Pi_T}$.

Suppose the result holds when Ξ contains sets of size at most k for $k \geq 1$. Consider $\Xi = \{A\}$ where $\#A = k + 1$. Define υ as the time of the first jump of $(\Pi_u)_{u\in[0,T]}$, with $\upsilon = +\infty$ if $\Pi_T = \Pi_0$. On the event $[\upsilon = +\infty]$, $(\mathcal{G}_t)_{t\in[0,T]}$ is equal in distribution to the A-marginal forward process $(\tilde{\mathcal{G}}_t^A)_{t\in[0,T]}$, conditioned not to have recombination events.

On the event $[\upsilon < T]$, we know from Proposition 3.3.7 that $\mathcal{G}_{T-\upsilon}$ has law $\mathbb{P}_{[0,T-\upsilon],\Pi_\upsilon}$. From the recursion hypothesis, writing $\Pi_\upsilon = \{\mathcal{I}, A \setminus \mathcal{I}\}$, this implies that $\mathcal{G}_{T-\upsilon}$ has law $\mathbf{X}_{T-\upsilon}^{\mathcal{I}} \otimes \mathbf{X}_{T-\upsilon}^{A\setminus\mathcal{I}}$. But this means $\mathcal{G}_T$ is equal in distribution to a process obtained by sampling $(\upsilon, \mathcal{I})$, sampling $\mathcal{G}_{T-\upsilon}$ with law $\mathbf{X}_{T-\upsilon}^{\mathcal{I}} \otimes \mathbf{X}_{T-\upsilon}^{A\setminus\mathcal{I}}$, sampling $\eta^S_{(T-\upsilon,T]}$, and then generating $(\mathcal{G}_t)_{t\in[T-\upsilon,T]}$. It follows that $\mathcal{G}_T$ is equal in distribution to $\hat{\mathcal{G}}_T$ from Proposition 3.3.4. This proves the result when $\Xi = \{A\}$ with $\#A = k+1$. This result is straightforwardly extended to any $\Xi \in \mathscr{S}$ such that the largest block of Ξ has length $k+1$ by noticing that the restrictions $(\Pi_u^{|\mathcal{I}})_{u\in[0,T],\mathcal{I}\in\Xi}$ are independent, where for $\Pi \in \mathscr{S}$ and $\mathcal{I} \subseteq [L]$ we define the marginal of Π on $\mathcal{I}$ as

$$\Pi^{|\mathcal{I}} := \{A \cap \mathcal{I}, A \in \Pi\}.$$

□

3.3.6 Adding mutations

Mutations are handled with mutation collisions. In the definition of the marginal forward genomic process $(\tilde{\mathcal{G}}_t)_{t\in[0,T]}$ in Section 3.3.1, a third kind of jump occurs: with rate $|\theta_\ell|$, $\tilde{\mathcal{G}}_{t,\ell}$ jumps to $\breve{g}_{t,\ell}$ independently sampled with law $\mathcal{L}_\ell$. This process can be represented using a PPP η^Θ on $[0,T] \times [L] \times \{-1,+1\}$ with intensity $\sum_{\ell\in A}|\theta_\ell|\mathrm{d}t \times \delta_\ell \times \mathcal{L}_\ell$. Recall (η^R, η^Θ) can be represented as a couple $(\eta^{R+\Theta}, Y)$, where $\eta^{R+\Theta}$ is a PPP on $[0,T]$ with intensity $\beta_A + \sum_{\ell\in A}|\theta_\ell|$ and for any point t of $\eta^{R+\Theta}$, Y_t is an independently sampled random variable determining whether the jump at t is a mutational collision or recombination collision (see Theorem 5.6 of [166]). Therefore, we can use the same kind of arguments as in Section 3.3.3, to obtain that if υ is the last jump of $\eta^R + \eta^\Theta$, then $\tilde{\mathcal{G}}_{\upsilon-}$ has law $\mathbf{X}_{\upsilon-}$ and is independent of what the event at υ is.

We may then conclude the proof of Theorem 3.2.1 as above with a recursion on the size of the largest block of Ξ.

3.3.7 Proof for $\mathcal{H}_t$

Define the set of singletons

$$\mathscr{T} := \{\ell \in [L], \quad \tau_\ell > 0\}$$

Let ℓ_0 be a latest singleton, that is, such that $\tau_{\ell_0} \geq \tau_\ell$ for $\ell \neq \ell_0$. Conditional on $\tau_{\ell_0} = 0$, $(\mathcal{H}_t)_{t\in[0,T]}$ and $(\mathcal{G}_t)_{t\in[0,T]}$ clearly have the same distribution.

Notice $T - \tau_{\ell_0}$ is a $\mathscr{F}^-$–stopping time, for which ℓ_0 is measurable. It follows from the first part of the Theorem that $\mathcal{G}_{\tau_{\ell_0},\ell_0}$ has law $\mathbf{X}_{\tau_{\ell_0}}^{\{\ell_0\}}$, and that $\mathcal{G}_{\tau_{\ell_0}}^{[L]\smallsetminus\{\ell_0\}}$ is independent of $\mathcal{G}_{\tau_{\ell_0},\ell_0}$, conditional on $\mathscr{F}_{T-\tau_{\ell_0}}^-$. We may then iterate the argument to show that conditional on $(\tau_\ell)_{\ell\in[L]}$, the $(\mathcal{G}_{\tau_\ell,\ell})_{\ell\in\mathscr{T}}$ are independent such that $\mathscr{L}(\mathcal{G}_{\tau_\ell,\ell}) = \mathbf{X}_{\tau_\ell}^{\{\ell\}}$, and are independent of $\mathcal{G}_0^{[L]\smallsetminus\mathscr{T}}$.

In particular, we have from the definition of $\mathcal{H}$:

$$\mathscr{L}\left((\mathcal{G}_{\tau_\ell,\ell})_{\ell\in\mathscr{T}}, \mathcal{G}_0^{[L]\smallsetminus\mathscr{T}}\right) = \mathscr{L}\left((\mathcal{H}_{\tau_\ell,\ell})_{\ell\in\mathscr{T}}, \mathcal{H}_0^{[L]\smallsetminus\mathscr{T}}\right) \tag{3.8}$$

As in Section 3.3.2, we can construct $\mathcal{G}_T$ from $(\mathscr{F}^-, \mathcal{G}_0)$ using an independent PPP η^S on $[0,T] \times \bigcup_{A\subseteq[L]} \Box_A \times \mathbb{R}_+$ with constant intensity. Specifically, if $A \in \tilde{\Pi}_t$ and if η^S has a point on (t, γ, x) for $\gamma \in \Box_A$, then $\mathcal{G}_t^{|A}$ jumps to γ as long as

$$x < \left(\mathbf{X}_t\left[W(g) \mid g^{|A} = \gamma\right] - \mathbf{X}_t\left[W(g) \mid g^{|A} = \mathcal{H}_t^{|A}\right]\right)_+ \mathbf{X}_t[\mathbb{1}_{[g^{|A}=\gamma]}].$$

and similarly $\mathcal{H}_t^{|A}$ jumps to γ with the same rate, provided $\#A > 1$.

We get the result by combining (3.8) with the fact that the restriction of η^S to $[0,T] \times \bigcup_{A\subset[L];\#A>1} \Box_A \times \mathbb{R}$ is independent from $(\mathscr{F}^-, (\mathcal{G}_{\tau_\ell,\ell})_{\ell\in\mathscr{T}}, \mathcal{G}_0^{[L]\smallsetminus\mathscr{T}})$.

Chapter 4

Focus on stabilizing selection.

Contents

4.1 Introduction

Historically, the infinitesimal model [11] has been considered to bridge population genetics, which is the field that studies the evolution of the genetic composition of populations, and quantitative genetics, which is the field that studies the evolution of quantitative traits. In the infinitesimal model, (suitably transformed) quantitative traits are determined by the sum of a genetic trait value and an independent environmental effect. The genetic trait value is inherited as follows: if two unrelated organisms with respective genetic trait values z^A and z^B reproduce, their offspring has normally distributed genetic trait value with mean $(z^A + z^B)/2$ and variance V_S called the segregation variance. It was long conjectured [79] and proved in [11] that this model is compatible with Mendelian genetics, provided the quantitative trait is determined by a large number of loci acting additively on the trait, with no major-effect locus.

This allows for an autonomous description of the evolution of the trait, which we call the **trait's eye view**. This was historically done by Wright [96] and Lande [30], and a wealth of mathematical models extending this approach has been developed since [38] with recent efforts from the partial differential equation community to determine their analytical properties and robustness (see, for instance, [47, 46]).

The segregation variance V_S is the key component of the infinitesimal model. Modeling its evolution requires understanding how mutations generate genetic variability, which itself is eroded by genetic drift and selection. Therefore, it should require modeling jointly the genetic composition of the population with the evolution of the trait [122]. However, in trait-based models, such as Lande's famous study of stabilizing selection [30], the segregation variance is a fixed parameter of the model.

Recently, some authors have used tools such as Wright's formula [73] to study the dynamics of the frequency of a single allele *conditional on the evolution of the trait.* Assuming linkage equilibrium, this formula gives the effective selection coefficient at a biallelic locus as a function of the gradient of mean logfitness in the population. This approach was used to interpret Genome-Wide Association Studies (GWAS) results in [116], and to model adaptation after a change of environment in [123, 124].

In principle, since the genes determine the trait and only the genes are transmitted, the genetic composition of the population evolves autonomously, without the need to specify any auxiliary variable, and from it the trait distribution should be deduced. In the present article, we seek to fullfil this program and start directly from the allele frequencies at all loci underlying a polygenic trait. We assume stabilizing selection which, as emphasized in [115], is a realistic assumption for wild populations.

The difficulty of this approach is that genes are coupled in a nontrivial manner by recombination and selection. Now, if we assume that genes are in perfect linkage equilibrium (see Appendix 4.C.E.1 for a mathematical discussion about the domain of validity of this assumption), then selection remains as the only coupling force–the selective advantage of an allele at a given locus depends on the entire trait distribution, which itself depends on all other loci.

We show using varied mathematical tools (diffusion approximation, mean field theory, fixed-point equations, slow-fast approximations) that this dependence vanishes with the number of loci.

We call the **gene's eye-view** this approach, which studies the distribution of a polygenic trait determined by a large number of unlinked loci, through the dynamics of their allele frequencies, to which diffusion approximations and mean-field approximations are applied. This approach arose recently, notably in the seminal paper by Charlesworth [122], and to describe spatial populations in [167, 168]. It lets us characterize properties of the trait distribution that emerge directly from the sole description of genes, such as the genetic variance, the distance to the selection optimum, and the magnitude of the fluctuations of the population mean trait value at statistical equilibrium–though our model can in principle be extended to describe the dynamics out-of-equilibrium.

Our computations distinguish three regimes which we call weak, moderate and strong selection, based on the value of the **selection-drift ratio**

$$\omega_e^{-2} := 2N(L\bar{\alpha})^2\omega^{-2}$$

compared to powers of L, where N is the (haploid) effective population size, ω^{-2} is the strength of stabilizing selection, $\bar{\alpha}$ is the average effect size of a locus and L is the number of loci.

4.2 Description of the population model

4.2.1 The individual-based model

We consider a large, panmictic population composed of N diploid organisms, displaying an additive quantitative trait z controlled by L biallelic loci. We assume that $N \gg 1$ and $L \gg 1$.

A genotype $g = (g_\ell)_{\ell\in[L]}$ determines the value $Z(g)$ of the trait as

$$Z(g) = z_0 + \sum_{\ell\in[L]} \alpha_\ell g_\ell,$$

where $g_\ell \in \{0, 1, 2\}$ is the gene content and $\alpha_\ell \geq 0$ is the genetic additive effect of the trait-increasing allele at locus ℓ. We define the **mean additive effect** $\bar{\alpha} := (1/L)\sum_\ell \alpha_\ell$. The model can accomodate extensions to polyploidy, pleiotropy, dominance or some forms of epistasis (see Appendix 4.I). We assume that $z_0 = 0$ without loss of generality.

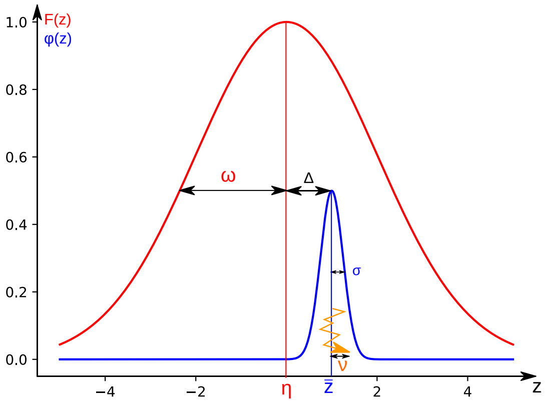

Figure 4.1: Phenotypic observables. The wide curve is the Gaussian fitness function $F(z) = e^{W(z)}$, which has typical width ω and is centered on η. The narrow curve is the distribution of the trait in the population, which has mean $\bar{z}$ and variance σ^2. The deviation of $\bar{z}$ from the optimum is $\Delta = \bar{z} - \eta$. The broken line represents the fluctuations of $\bar{z}$ through time, which have magnitude ν.

We define the probability of mutation per generation from the trait-decreasing allele to the trait-increasing allele at locus ℓ, and back, as $\mu_\ell = (\mu_\ell^+, \mu_\ell^-)$. We assume that the $(\alpha_\ell, \mu_\ell)_{\ell \in [L]}$ are exchangeable, such that (α_ℓ, μ_ℓ) has a prescribed distribution which is not heavy-tailed, so that in particular there are no large-effect mutations.

The logfitness of an organism with trait value z is given by

$$W(z) = -\frac{1}{2\omega^2}(z-\eta)^2 \tag{4.1}$$

with ω^{-2} the **selection strength** and η the **selection optimum**.

Reproduction occurs via Wright-Fisher sampling: at each generation, organisms independently choose two parents with probability proportional to their fitness and the two parental genomes recombine with one or several crossovers.

Importantly, we assume that recombination is strong enough to allow us to neglect linkage disequilibrium (this assumption is discussed in Appendix 4.F.2).

4.2.2 Three trait values

A model studying the dynamics of a polygenic trait under stabilizing selection from the gene's eye-view distinguishes three important theoretical trait values.

The first is the **selection optimum** η.

The second is the **mutational optimum** z_M defined as the mean trait value that the population would reach at equilibrium if selection were completely absent

$$z_M := \sum_{\ell=1}^{L} \alpha_\ell \frac{2\mu_\ell^+}{\mu_\ell^+ + \mu_\ell^-}, \tag{4.2}$$

because $2\mu_\ell^+/(\mu_\ell^+ + \mu_\ell^-)$ is the mean value of g_ℓ at mutational equilibrium. This value has no *a priori* reason to coincide with η, and we call $z_M - \eta$ the **trait mutational bias**.

The third is the **heterozygous trait value** z_H, which is the trait value of an organism heterozygous at every locus

$$z_H := \sum_{\ell=1}^{L} \alpha_\ell = L\bar{\alpha}. \tag{4.3}$$

We assume that $0 < \eta < 2z_H$, so that many different genotypes can realize the selection optimum and the population close to the optimum is not depleted in genetic variability. We may also assume that the trait is measured in units such that $z_H = 1$ so that in particular $\bar{\alpha} = 1/L$.

Other key ecological observables are represented in Fig. 4.1 and are explicitly defined in Section 4.3.

Assumptions The details of our explicit quantitative assumptions are given in Appendix 4.B. Some assumptions are made for mathematical convenience, others are necessary to reach the polygenic limit. Here, we informally explain the underlying philosophy.

The mathematically convenient assumptions are as follows. We assume that (A1) all loci have mutation rates of order $|\bar{\theta}|$ and additive effect $\alpha_\ell \sim 1/L$, so that no one locus disproportionately contributes to the trait or the genetic variance. We assume that (A2) mutations are weaker than genetic drift ($|\bar{\theta}| \lesssim 1$) and that (A3, A6) the mutational bias is not too large or too small. We assume that (A4) the selection optimum η is accessible, meaning many different genotypes can realize the selection optimum (in particular, the population close to the optimum will not be depleted in genetic variability). We assume that (A5) selection is strong enough to have a detectable footprint on the genes, but not so strong as to completely deplete genetic variability in the trait.

We argue in Appendix 4.F that the equations we obtain further require a sufficiently large population size (N1) and a sufficient mutational input (N2, N3) to provide an accurate description of the system.

4.2.3 Diffusion approximation

We let P_t^ℓ denote the frequency of the trait-increasing allele at locus ℓ, where time is now measured in units of $2N$ generations; that is, P_t^ℓ represents the frequency at generation $\lfloor 2Nt \rfloor$. We define

$$\theta_\ell := 2N\mu_\ell, \quad \omega_e := \frac{\omega}{\sqrt{2N}},$$

so that θ_ℓ is the effective mutation rate, and $\omega_e^{-2} = 2N\omega^{-2}$ is the **selection-drift ratio**.

As explained in Appendix 4.C), neglecting linkage disequilibrium allows approximating allele frequency dynamics of the individual loci by Wright-Fisher diffusions which are only coupled by selection

$$\begin{aligned} \mathrm{d}P_t^\ell &= s_\ell(\mathbf{P_t})P_t^\ell(1-P_t^\ell)\,\mathrm{d}t \\ &\quad + \left(\theta_\ell^+(1-P_t^\ell) - \theta_\ell^- P_t^\ell\right)\mathrm{d}t + \sqrt{P_t^\ell(1-P_t^\ell)}\,\mathrm{d}B_t^\ell, \end{aligned} \tag{4.4}$$

where $(B^\ell)_\ell$ are independent Brownian motions. Classically, the selection coefficient s_ℓ is the regression coefficient of the fitness W on genotype g_ℓ

$$s_\ell(\mathbf{P}_t) := 2N\frac{\mathbf{Cov}_{\mathbf{P_t}}\left[W(Z(g)), g_\ell\right]}{\mathbf{Var}_{\mathbf{P_t}}[g_\ell]},$$

where for $\mathbf{p} = (p_\ell)_{\ell\in[L]} \in [0,1]^L$, $\mathbf{Cov_p}$ and $\mathbf{Var_p}$ are the covariance and variance of a vector of independent random variables $g = (g_\ell)_{\ell\in[L]}$ such that g_ℓ is *Binomial*$(2, p_\ell)$ (Hardy-Weinberg Linkage Equilibrium). We can show that s_ℓ (see Appendix 4.C.2) satisfies

$$s_\ell(\mathbf{P}_t) = -\frac{\alpha_\ell}{\omega_e^2}(\bar{z}_t - \eta) + \frac{\alpha_\ell^2}{\omega_e^2}\left(P_t^\ell - \frac{1}{2}\right), \tag{4.5}$$

where $\bar{z}_t$ is the mean trait value at time t

$$\bar{z}_t = \sum_{\ell=1}^{L} \alpha_\ell \mathbf{E}_{\mathbf{P_t}}(g_\ell) = \sum_{\ell=1}^{L} 2\alpha_\ell P_t^\ell.$$

As a result, allele frequencies (P_t^ℓ) are now coupled only by their weighted mean $\bar{z}_t$. This coupling can be further simplified when L is large thanks to a mean-field approximation.

4.3 Macroscopic observables from gene's eye view

4.3.1 Locus dynamics and mean field approximation

Mean-field theory is a powerful approach for analyzing complex systems composed of a large number L of interacting components. When the constituents interact through a weighted mean of their individual characteristics, as $L \to \infty$ this weighted mean behaves deterministically as by the law of large numbers, and the constituents effectively behave independently [69].

In the case where the genetic effects and mutation rates $(\alpha_\ell, \theta_\ell) \equiv (\alpha, \theta)$ are equal across loci, by replacing the mean with its expectation

$$\bar{z}_t = \sum_{\ell=1}^{L} 2\alpha P_t^\ell = 2L\alpha \times \left(\frac{1}{L}\sum_{\ell=1}^{L} P_t^\ell\right) \approx 2L\alpha \mathbb{E}\left[P_t\right], \tag{4.6}$$

where P_t has the law of P_t^ℓ (which here does not depend on ℓ). When the $(\alpha_\ell, \theta_\ell)$ are now distributed, we replace the *weighted* mean with an 'augmented' expectation which includes averaging over the distribution of $(\alpha_\ell, \theta_\ell)$

$$\bar{z}_t \approx 2L\mathbb{E}\left[\alpha P_t\right]. \tag{4.7}$$

We can now define $\Delta_t := \bar{z}_t - \eta$ as the deviation of the trait mean from the optimum and Δ^* as the **mean deviation at statistical equilibrium**

$$\Delta^* := 2L\mathbb{E}^*[\alpha P_t] - \eta \tag{4.8}$$

where $\mathbb{E}^*$ is the expectation at equilibrium, which is also the time average

$$\mathbb{E}^*[\alpha P_t] = \lim_{T \to +\infty} \frac{1}{T}\sum_{t=0}^{T} \frac{1}{L}\sum_{\ell=1}^{L} \alpha_\ell P_t^\ell.$$

Then, the allele frequency P_t^ℓ at each locus ℓ independently obeys the following stochastic differential equation which is autonomous in the sense that each term only depends on P_t^ℓ itself

$$\begin{aligned} \mathrm{d}P_t^\ell = {} & \xi_{\Delta^*,\alpha_\ell}(P_t^\ell) P_t^\ell (1 - P_t^\ell)\,\mathrm{d}t \\ & + \left(\theta_\ell^+ (1 - P_t^\ell) - \theta_\ell^- P_t^\ell\right)\mathrm{d}t + \sqrt{P_t^\ell (1 - P_t^\ell)}\,\mathrm{d}B_t^\ell, \end{aligned} \tag{4.9}$$

where

$$\xi_{\delta,a}(x) := -\frac{a}{\omega_e^2}\delta + \frac{a^2}{\omega_e^2}\left(x - \frac{1}{2}\right). \tag{4.10}$$

A formal justification for this mean-field approximation is spelled out in Section 4.3.4, where we also describe the dynamics of the trait.

Let us briefly discuss the expression of $\xi_{\Delta^*,\alpha_\ell}$ given in (4.10). The second term is known as Robertson's underdominant effect [23]. The term $-\alpha_\ell\Delta^*/\omega_e^2$ can be interpreted as a *bias-correcting selection coefficient.* It reflects the selective advantage of trait-increasing alleles when $\Delta^* < 0$ (the trait is on average below the optimum) and their disadvantage when $\Delta^* > 0$ (the trait is on average above the optimum). We define s^* as its mean value

$$s^* = -\frac{\Delta^*}{L\omega_e^2}. \tag{4.11}$$

In the absence of selection, mutations naturally drive the trait to the mutational optimum z_M defined in (4.2). In view of (4.10) and (4.11), this occurs when $\omega_e^{-2} \ll L$. Then $\Delta^* \approx z_M - \eta = O(1)$, and selection is negligible at the genetic and trait levels (see Appendix 4.E.2). As soon as the order of magnitude of ω_e^{-2} is equal to L or higher, what shapes the value of Δ^* is the balance between mutation (pushing $\bar{z}_t$ to z_M) and selection (pushing $\bar{z}_t$ to η).

4.3.2 Bridging microscopic and macroscopic scales

The consistency between microscopic (locus-level) and macroscopic (trait-level) dynamics is central to the mean-field approach and allows for an explicit determination of Δ^* as the solution to a fixed-point problem.

In (4.9), take $(\Delta^*, \alpha_\ell, \theta_\ell) = (\delta, \alpha, \theta)$. The stationary density of this diffusion [53] is then given for $p \in (0, 1)$ by

$$\Pi_{\delta,\alpha,\theta}(p) := C_{\delta,\alpha,\theta}\, p^{2\theta^+-1}(1-p)^{2\theta^--1}e^{2\int_0^p \xi_{\delta,\alpha}(u)\,\mathrm{d}u}, \tag{4.12}$$

where $C_{\delta,\alpha,\theta}$ is a normalizing constant. Using the mean-field approximation to average trait contributions across loci, from (4.8) we deduce that Δ^* satisfies the following fixed-point equation:

$$\Delta^* = 2L\mathbb{E}\left[\alpha \int_0^1 x\,\Pi_{\Delta^*,\alpha,\theta}(x)dx\right] - \eta, \tag{4.13}$$

where the expectation is now only taken over the distribution of $(\alpha_\ell, \theta_\ell)$. The latter relation embodies a *self-consistency condition* from the gene's eye-view of quantitative genetics that allows tying the dynamics of allelic frequencies with macroscopic observables. We deduce from this fixed-point equation in Appendix 4.E.3 that (except when $\omega_e^{-2} \ll L$, where $\Delta^* \approx z_M - \eta$) the mean trait deviation always scales as

$$\Delta^* \sim \omega_e^2 L = \omega^2 \frac{L}{2N}. \tag{4.14}$$

In this equation, $L/(2N)$ can be interpreted as the genetic drift accumulated along the genome, while ω^2 is the inverse of the strength of directional selection felt by the trait away from the optimum.

Similarly, we can deduce the trait variance σ^2 as a result of the independent, genetic additive contributions

$$\sigma^2 = 2L\mathbb{E}\left[\alpha^2 \int_0^1 x(1-x)\Pi_{\Delta^*,\alpha,\theta}(x)dx\right] \sim \frac{|\bar{\theta}|}{L}. \tag{4.15}$$

where $|\bar{\theta}|$ is the mean mutation rate

$$|\bar{\theta}| := \mathbb{E}\left[\theta^+ + \theta^-\right].$$

We plot in Fig. 4.3 the result of simulations using the individual-based model against our theoretical predictions for different selection strengths. As expected, stronger selection leads to a smaller distance to the optimum Δ^*. We also see that when selection is sufficiently strong, it depletes the genetic variance σ^2. More precisely, we see that when selection is weak, our theoretical predictions for Δ_t and σ_t match simulation results even when $L = 100$ and $N = 50$. On the other hand, when selection is strong, we see a mismatch in the genetic variance σ^2 due to the build-up of negative Linkage Disequilibrium (LD): this is the so-called Bulmer effect [106]. The mismatch is decreased when the population size is increased ($N = 500$). Furthermore, the decrease in σ^2 due to the Bulmer effect decreases the effectiveness of the bias-correcting coefficient of selection, increasing the equilibrium distance to the optimum Δ^* with respect to neutral expectations. Following the Quasi-Linkage Equilibrium approach [101], we argue in Appendix 4.F.2 that the Bulmer effect can be neglected if $N \gg |\bar{\theta}| \ln(L)/(L\omega_e^2)$.

4.3.3 Selection regimes

As noted above, when $\omega_e^{-2} \ll L$, selection has no substantial effect on the system (Appendix 4.E.2). In contrast, when $\omega_e^{-2} \gg L^2$, Robertson's underdominant effect becomes degenerate ($\frac{\alpha^2}{\omega_e^2} \gg 1$), resulting in a depletion of genetic variance with a high concentration of allelic frequencies at 0 and 1. In this "ultra-strong" selection regime, the intensity of selection is the main determinant of genetic variance. Inbetween these two extremes, our analysis identifies three distinct regimes, as illustrated in Fig. 4.2 and summarized in Table 4.1. Recall from (4.15) that the trait population variance σ^2 is always of order $|\bar{\theta}|/L$.

1. *Weak selection regime* ($\omega_e^{-2} \sim L$): In this regime, the population is concentrated away from the fitness optimum:
$$\sigma \ll \Delta^* \sim 1.$$
2. *Strong selection regime* ($\omega_e^{-2} \sim L^2$): Here, the population is tightly concentrated around the fitness optimum:
$$\Delta^* \ll \sigma \ll 1.$$
   It is only in this regime that Robertson's underdominant effect becomes significant, so that genes also experience disrupting selection in addition to genic selection against mutational bias.
3. *Moderate selection regime* ($L \ll \omega_e^{-2} \ll L^2$): This regime interpolates between the weak and the strong regime, with Δ^* and σ of the same order when $\omega_e^{-2}|\bar{\theta}|^{\frac{1}{2}} \sim L^{3/2}$.

Persistence of s^* across scales. At the locus level, the mean selection coefficient of trait-increasing alleles $s^* = -\Delta^*/(L\omega_e^2)$ is proportional to the product of the trait deviation Δ^* from the optimum and of the strength of selection ω_e^{-2}. At the trait level, we see from (4.14) that the strength of selection has a direct effect on Δ^* which keeps the product $\Delta^*\omega_e^{-2}$ of order L. As a result, s^* remains of order 1 throughout the three regimes and as L increases. In other words, the mutational bias always has a substantial effect at the locus level. This effect is asymptotically ($L \to \infty$) independent of the strength of selection since s^* approaches some value independent of ω_e as long as $L \ll \omega_e^{-2} \ll L^2$ (Appendix 4.E.4).

In particular, in this regime, which we call the **moderate selection regime**, the same genetic architecture, and accordingly similar macroscopic observables, can arise for very different values of ω_e^{-2}. We illustrate this in Fig. 4.4, in which the theoretical prediction for the rescaled genetic variance $L\sigma^2$ is plotted in the limit $L \to +\infty$, as a function of the **selection power** defined with $\omega_e^{-2} = L^b$, equivalent to

$$b := \frac{\ln(\omega_e^{-2})}{\ln(L)} \tag{4.16}$$

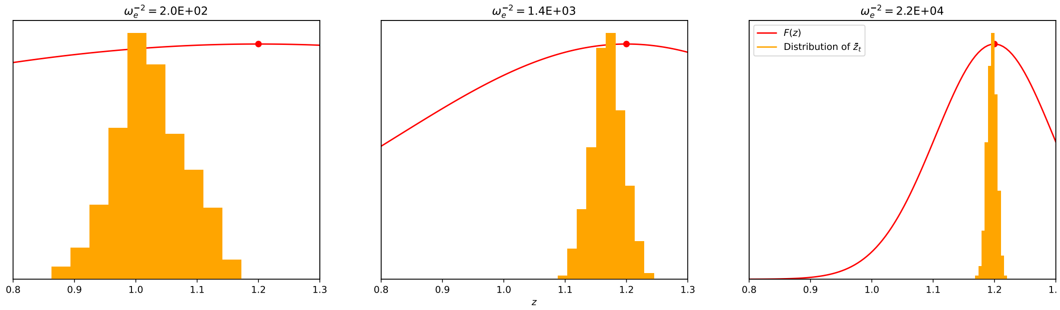

Figure 4.2: The histogram shows the distribution of the population mean trait $\bar{z}_t$ and not a snapshot of the trait distribution at a given time. For three different values of ω_e corresponding respectively to weak, moderate and strong selection, we show a histogram of $10,000$ values of $\bar{z}_t$ after a burn-in of $24,000$ generations. The parameters used are $N = 100, L = 100$, $\eta = 1.2, \theta = (0.1, 0.2)$, and the parameters $(\alpha_\ell)_{\ell\in[L]}$ were sampled with distribution $Exponential(L)$. We superimpose in red the fitness function $F(z)$, and the dot corresponds to η. In particular, Δ^* is the distance between the mean of the orange distribution and the red dot, and ν is the width of the orange distribution. As the strength of selection increases, both Δ^* and ν decrease. Only in strong selection (right panel) do we see Δ^* and ν being of the same order of magnitude.

In this figure, it can be seen that $L\sigma^2$ is expected to converge to a step function of b, with a single value corresponding to $1 < b < 2$ and discontinuities at $b = 1$ (weak selection regime) and $b = 2$ (strong selection regime). Similarly, in the moderate selection regime we expect $\Delta^* L^{b-1}$ to be independent of b (see Table 4.1). In particular, the mutational bias only has a macroscopically detectable effect $\Delta^* \gtrsim 1$ when $\sigma \lesssim \Delta^*$, which corresponds to $\omega_e^{-2}|\bar{\theta}|^{1/2} \lesssim L^{3/2}$ (see Table 4.1). This means that the strength of selection ω_e^{-2} cannot be inferred when $|\bar{\theta}|^{-1/2}L^{3/2} \ll \omega_e^{-2} \ll L^2$, neither from macroscopic data, because Δ^* cannot be measured, nor from genomic data, because the distribution of P_t is independent of ω_e^{-2}.

Non-monotonicity of the variance. Another counterintuitive prediction of our model is that σ does not always decrease with higher selection pressure. In the absence of selection, when mutational bias is strong, allele frequencies tend to accumulate close to 0 or to 1 depending on the sign of mutational bias. Increasing selection at the trait level (parameter b in Fig. 4.4, panels D, E, F) induces bias-correcting directional selection at the locus level that recenters the frequencies, thus increasing the population variance. We prove in Appendix 4.H that when the $(\alpha_\ell, \theta_\ell)$ are constant across loci, the criterion for some level of weak selection to increase σ is that going from the mutational optimum z_M to the selection optimum η brings the trait closer to the heterozygous trait value z_H. That is: $\eta, z_H < z_M$ or $\eta, z_H > z_M$. To put it another way, starting from a situation with very weak selection, where $\mathbb{E}^*[\bar{z}] \simeq z_M$, decreasing ω (increasing the strength of selection) brings $\mathbb{E}^*[\bar{z}]$ closer to η. If, in so doing, $\mathbb{E}^*[\bar{z}_t]$ also gets closer to z_H, then the action of selection will lead some loci which the mutational bias kept at one boundary closer to 1/2, and σ will increase for entropic reasons.

Heterogeneity across loci. In Fig. 4.5 we plot the joint distribution of (P_t^ℓ, α_ℓ), assuming that mutation rates are symmetric ($\theta^+ = \theta^- = 1/2$) and that the trait mutational bias $z_M - \eta$ is negative, thereby selecting trait-increasing alleles.

As expected, the distribution of allele frequencies is observed to be approximately uniform in the absence of selection and to be biased to the right in the presence of substantial selection. In the moderate regime, the bias-correcting selection term indeed shifts the distribution to the

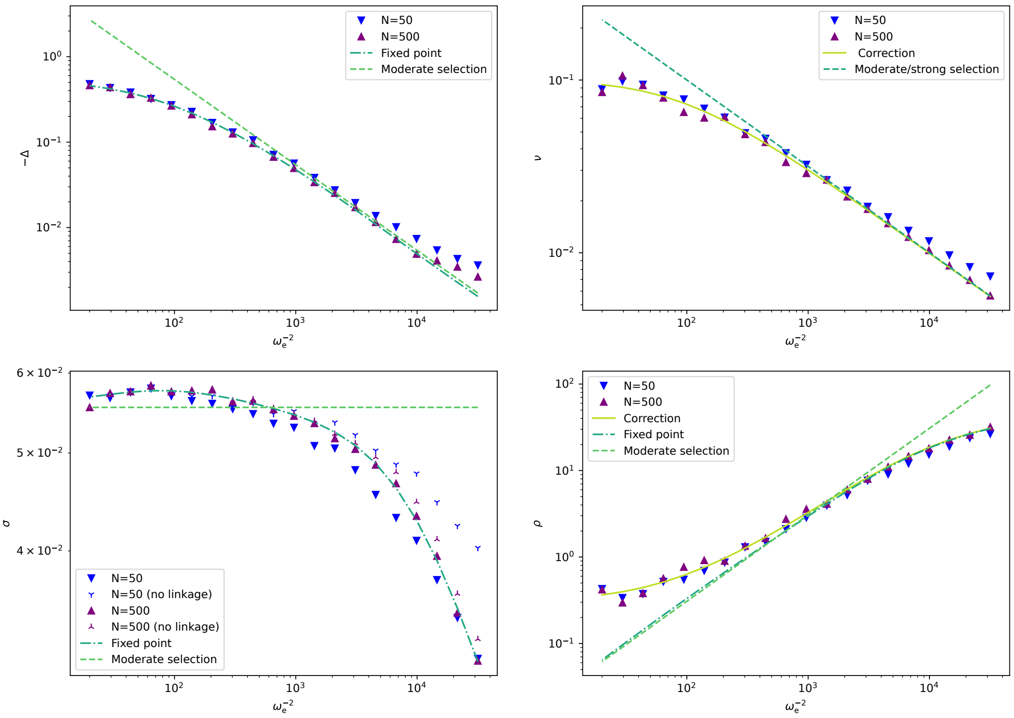

Figure 4.3: Comparison of theoretical predictions with numerical simulations for the macroscopic observables of Fig. 4.1 for varying selection regimes, from weak selection (left) to strong selection (right). The simulations were carried out with $L = 100$, the same mutation probabilities for all loci $\theta = (0.1, 0.2)$, and additive effects $(\alpha_\ell)_{\ell\in[L]}$ exponentially distributed with parameter L (in particular $\bar{\alpha} = 1/L$), so that $z_M = 2/3$. The selection optimum is $\eta = 1.2$. The same $(\alpha_\ell)_{\ell\in[L]}$ were used in all simulations, and the predictions were made conditional on the $(\alpha_\ell)_{\ell\in[L]}$. The simulations were run for $T = 500N$ generations, and each observable was measured as an average over the last $250N$ generations. For the magnitude of the fluctuations ν and the autocorrelation parameter ρ, the predictions for weak selection use the corrections derived in Appendix 4.F.5. The predictions of the fixed point equation are derived in Appendix 4.E.1. The predictions for moderate selection are derived in Appendix 4.E.4. For the genetic variance, the simulation results distinguish between the genetic variance in the trait within the population (filled triangles), and the genic variance (three-pointed stars), which is the variance in the trait if the population was in linkage equilibrium (neglecting correlations between pairs of loci).

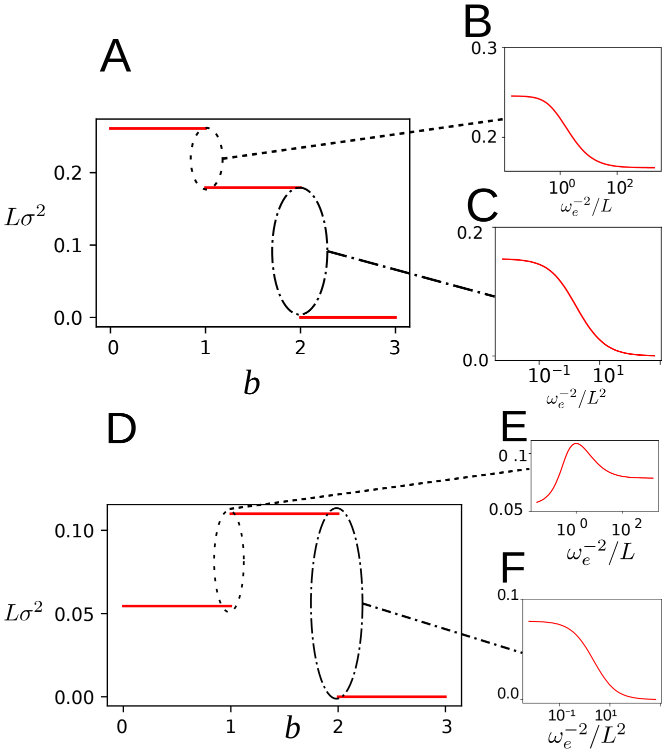

Figure 4.4: The predicted genetic variance at equilibrium in a limit system with $L \to +\infty, N \to +\infty$ with no mutational bias (**A**,**B**,**C** with $\theta_\ell = (0.1, 0.1)$) or strong mutational bias (**D**,**E**,**F** with $\theta_\ell = (0.01, 0.1)$), as a function of the selection power b (see (4.16)) and the selection-drift ratio ω_e^{-2}. We set $\eta = 1.5$ and let $(\alpha_\ell)_{\ell\in[L]}$ be exponentially distributed with parameter L (in particular $\bar{\alpha} = \frac{1}{L}$). **A&D** We expect the rescaled genetic variance $L\sigma^2$ in the limit to converge to a step function of the parameter b. When $b < 1$ the genetic variance is equivalent to that of a neutral model (no selection), when $b > 2$ we expect the genetic variance to be completely depleted, and when $b \in (1, 2)$, $L\sigma^2$ converge to the variance corresponding to moderate selection. The behavior at the critical points $b = 1$ and $b = 2$ correspond respectively to weak selection (**B** and **E**) and strong selection (**C** and **F**).

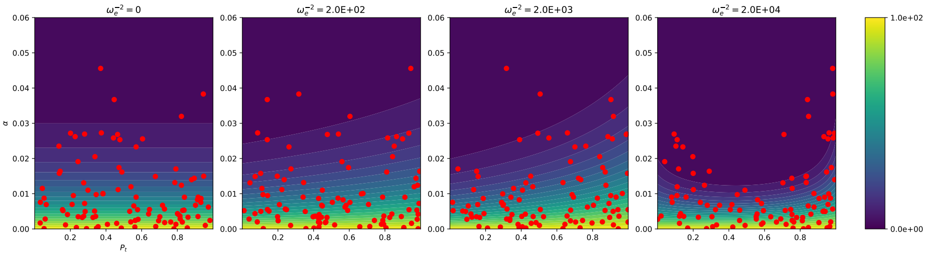

Figure 4.5: For no selection, weak, moderate and strong selection respectively, we plot the joint distribution of (P_t^ℓ, α_ℓ) after a single run. We chose symmetric mutation rates to make the transition between the selection regimes more apparent $\theta = (0.5, 0.5)$. The other parameters are $L = 100, N = 500, \eta = 1.2$ and α_ℓ is distributed as $\mathit{Exponential}(L)$. The colour plot correspond to the predicted densities in the corresponding regimes.

right, effectively counteracting the asymmetry of mutation rates. In the strong regime, the disruptive selection driven by Robertson's underdominant effect (4.10) results in a U-shape distribution, increasing the weight on extreme allelic frequencies.

4.3.4 Trait dynamics

In Appendix 4.D, we show that in the moderate and strong regimes ($\omega_e^{-2} \gg L$), the fluctuations of the mean deviation Δ_t away from its statistical average Δ^* evolves according to an Ornstein–Uhlenbeck process:

$$\mathrm{d}\Delta_t = \rho(\Delta^* - \Delta_t)\,\mathrm{d}t + \nu\sqrt{2\rho}\,dB_t, \tag{4.17}$$

In this equation, ν captures the magnitude of the stochastic fluctuations and ρ the autocorrelation parameter of $\bar{z}_t$ such that

$$\mathbb{C}\mathbf{ov}[\bar{z}_t, \bar{z}_{t+u}] = \nu^2 e^{-\rho u}. \tag{4.18}$$

We find in Appendix 4.E.3 that

$$\nu \sim \omega_e \quad \text{and} \quad \rho \sim \frac{1}{\tau} \tag{4.19}$$

where we define the characteristic timescale of the trait $\tau := \frac{L\omega_e^2}{|\theta|}$. Under weak selection ($\omega_e^{-2} \sim L$), we show in Appendix 4.F.5 that (4.17) is no longer justified mathematically (see Fig. 4.12). We derive a correction for ν^2 and ρ^2 which, though not mathematically rigorous, yield a good mumerical approximation when mutation rates $|\theta_\ell|$ are constant across loci. See the right panels of Fig. 4.3 for a comparison between theoretical predictions and individual-based simulations for ρ and ν. Increasing the strength of stabilizing selection decreases the magnitude of the fluctuations of $\bar{z}_t$ (ν^2), but accelerates them (increase in ρ). We plot in Fig. 4.12 the validity of equation 4.18 for the autocorrelation: it fits well even for weak selection, in which we expect this equation not to be valid.

Justifying the mean-field approximation. The previous observations provide further insight into justifying the mean-field approximation discussed in section 4.3.1. In Section 4.3.1, the locus dynamics was obtained by formally replacing Δ_t with its average Δ^*. How do we justify this approximation? From (4.5), we have

$$s_\ell(\mathbf{P}_t) = -\frac{\alpha}{\omega_e^2}\left(\Delta^* + \varepsilon_t\right) + \frac{\alpha^2}{\omega_e^2}\left(P_t^\ell - \frac{1}{2}\right), \tag{4.20}$$

Observable	**Symbol**	**Magnitude**	**Eq.**
Mean trait deviation from optimum	Δ^*	$L^{-(b-1)}$	(4.8)
Std deviation of trait distribution	σ	$\lvert\bar{\theta}\rvert^{1/2}L^{-1/2}$	(4.15)
Timescale of trait mean fluctuation	τ	$\lvert\bar{\theta}\rvert^{-1}L^{-(b-1)}$	(4.19)
Amplitude of trait mean fluctuation	ν	$L^{-b/2}$	(4.19)
Bias-correcting selection coeff.	s^*	1	(4.11)
Robertson's underdominant effect	$\frac{\bar{\alpha}^2}{\omega_e^2}$	$L^{-(2-b)}$	(4.10)

Table 4.1: Order of magnitude of macroscopic observables as a function of the number L of loci and $b \in [1,2]$, assuming $\omega_e^{-2} \sim L^b$, interpolating between the weak regime ($b = 1$), the moderate regime ($b \in (1,2)$) and the strong regime ($b = 2$). This table is derived in Appendix 4.E.3

where $\varepsilon_t := \Delta_t - \Delta^*$ is the fluctuation of the trait mean.

In the weak/moderate regimes ($L \lesssim \omega_e^{-2} \ll L^2$), the fluctuations are negligible compared to the mean deviation Δ^* ($\nu \ll \Delta^*$ in Table 4.1) and a classical mean-field approximation applies.

The strong regime ($\omega_e^{-2} \sim L^2$) is more subtle and can only be justified by a *dynamical* mean-field approach. As already discussed, fluctuations can not be ignored ($\nu \sim \Delta^*$, see Table 4.1) and at first sight, it seems that Δ_t is not approximately constant equal to Δ^* in (4.20). To address this issue, we adopt a *slow-fast approximation* [169]. Since the trait mean Δ_t fluctuates on a much faster timescale than the locus-level dynamics, this separation of timescales leads to a *slow-fast averaging effect*, allowing us to effectively decouple macroscopic (trait-level) and microscopic (locus-level) dynamics. As a result, in (4.5), we may still, despite non-negligible fluctuations, approximate Δ_t by Δ^*, now seen as a time average.

4.4 Discussion

4.4.1 Describing a polygenic system from the gene's eye-view

Our paper offers a comprehensive framework to describe a polygenic system from the gene's eye-view. This lets us describe the distribution of a gene at stationarity as well as macroscopic observables. These ideas were already present in the case of constant (α, θ) across loci, small mutation rates ($|\bar{\theta}| \ll 1$) and moderate/strong selection ($\omega_e^{-2} \gg L$) in [122] (see Appendix 4.E.4). We show in Appendix 4.G that the gene's eye-view and the traditional trait's eye-view from [96] lead to the same equations. In Appendix 4.F, we discuss the breakdown of our equations. Qualitative breakdown in which the population dynamics completely shift to genotype selection was described as clonal condensation in [104] and non-random coexistence [70], and the appearance of sweeps at individual loci was described [170, 81, 125]. The first two phenomena depend on linkage disequilibrium and are discussed in Appendix 4.F.2. The appearance of sweeps is what we expect to occur in the "ultra-strong" selection regime $\omega_e^{-2} \gg L^2$, when selection completely overwhelms genetic drift, or in the out-of-equilibrium dynamics if the system is originally very far from the optimum [81]. We also require the mutation rate to be large enough to maintain genetic diversity (we show in Appendix 4.F Fig. 4.12 that the model is robust to low mutation rates and significant mutational bias is high), and the $(\alpha_\ell)_{\ell\in[L]}$ to have well-defined moments (our plot in Appendix 4.F Fig. 4.11 suggests that the model is still accurate when the distribution of $(\alpha_\ell)_{\ell\in[L]}$ is heavy-tailed). This last assumption is of questionable biological relevance: [146] found rough estimates of tail exponents in additive effects between 1 and 2.5 (see also [171]).

The equation for the evolution of an allele depends on microscopic, intrinsic properties of the allele (the mutation probability μ, the genetic effect α) and also on an external variable, the bias-correcting coefficient s^*, which is determined by the value of a macroscopic observable (the mean trait value), which is itself a solution to a fixed point equation. This sort of approach from the gene's eye-view has seen increasing use in the past few years: for instance, [167] and [168] have used a similar approach to describe the stationary distribution of a population under directional selection and immigration. Another example is [26] (not yet peer-reviewed), in which the macroscopic observable is the genetic variance σ^2 under fluctuating stabilizing selection, which determines the allele frequencies, which in turn determine the genetic variance, thus yielding a fixed-point equation (their equation 16).

Taking a different approach based on the stationary solution to the multi-locus Wright-Fisher diffusion (4.4), [172] has recently found an explicit approximate expression for the solution Δ^* to the fixed-point equation (4.13) when the mutational bias is small ($|z_M - \eta| \ll 1$) and $\theta^+ = \theta^-$: this approximation is equivalent to computing the variance $\mathbb{V}\mathbf{ar}^*[P_t|\alpha, \theta]$ by approximating the distribution of P_t with the neutral distribution $\Pi_{0,0,\theta}$. Using this idea coupled with the central limit theorem leads to the first-order perturbation to $\Pi_{0,0,\theta}$ due to selection.

4.4.2 On the importance of trait mutational bias

Trait mutational bias in a quantitative trait describes the situation in which the additive effect of mutations on the trait has a nonzero mean. This should not be confused with mutational bias in fitness. Mutational bias in fitness refers to the fact that if the population is close to the optimum, new mutations will tend to be deleterious (see for instance [173]). Even if there was no trait mutational bias (say $z_H = \eta = z_M$), due to Robertson's underdominant term in the selection coefficient (4.10), we would still have that, on average, a mutation with large effect α is deleterious. The impact of trait mutational bias on a polygenic trait has been described in [174] numerically, and [122] has provided a theoretical treatment which includes approximate solutions assuming low mutation rates ($|\bar{\theta}| \ll 1$) and constant $(\theta_\ell, \alpha_\ell)$ across loci. [122] suggests two methods to approximate Δ^*, one (Eq. A3.b) which coincides with our own (see our Appendix 4.E.4) and one which requires approximating $\mathbb{E}\left[P_t(1-P_t)|(\alpha, \theta)\right]$ with its neutral expectation (Eq. S3.3b).

Apart from these two exceptions, trait mutational bias is typically neglected in theoretical studies of polygenic traits under stabilizing selection. For instance, [23, 97, 98, 175, 176, 177, 116] all study a quantitative trait under stabilizing selection, and always assume that at equilibrium the mutation rate does not push the population in any specific direction away from the optimum ([116] consider anisotropic pleiotropic mutations as an extension of their base model, but still assume that the mean effect of any new mutation is zero). The models for quantitative traits from statistical physics such as [101, 177, 70] also typically assume symmetric mutation rates. This seems to be motivated not just by mathematical tractability, but by the idea that when the mutation rate is very low, the directional effect of mutations can be neglected. It would be unreasonable to argue that mutational bias is typically small, as this would imply that mutation favors the same trait values as selection, which would be a strange coincidence, in particular if this were satisfied across distinct environments. In fact, experimental data on this can be obtained from the evolution of morphological traits in mutation accumulation (MA) experiments. Morphological quantitative traits are typically assumed to be under stabilizing selection [115]. In MA lines, such traits generally evolve in a deterministic direction. For instance, table 2 of [178] reviews the case of drosophila, in which the proportion of trait-increasing mutations for sternopleural/abdominal bristle number and wing length range between 0.4 and 0.07. More recently, [179] has reported significant mutational bias in locomotion traits in *Caenorhabditis elegans*. Mutational bias could also account for cases in which directional selection and stabilizing selection are simultaneously observed on phenotypic data [15]. Population genetic analyses also suggest significant mutational bias for codon usage [180, 181, 182, 183]. Finally, [118] has established an empirical test of mutational bias for GWAS in humans, and successfully detected a bias for traits such as height, BMI, and educational attainment.

In the presence of mutational bias, which we expect to be pervasive, weak selection cannot be distinguished from directional selection [174], because the population is too far from the optimum ($\frac{\Delta^*}{\sigma} \gg 1$). Such an equilibrium population would be in selection/mutation balance for a quantitative trait. While stable quantitative traits under directional selection are sometimes found in natural populations, mutational bias is however rarely mentioned as a putative explanation, presumably due to the relative strength of observed selection with respect to mutation (it is for instance never mentioned on reviews such as [25]). The possibility of mutations and genetic drift preventing a population from reaching the optimal trait value is discussed in the literature as the drift-barrier hypothesis. The typical quantitative trait for which this is deemed plausible is the mutation rate [184]. Our framework could be adapted to model this, following [148].

4.4.3 Example of a practical application: moderate selection and human height

One convenient aspect of our classification is that the qualitative behavior of the stationary system relies on parameters $\omega_e, |\bar{\theta}|, \tau$, which could be empirically estimated.

If we take the example of the human population, the effective population size is of order $N_e \sim 20,000$ [185]. For a trait like height, [117] estimate the mutational target to be of order 10^7 ([110] obtained a saturated GWAS map for height from $12,111$ loci, so a certain lower bound for L is 10,000). The average effect of a locus is taken to be of order $\bar{\alpha} = 10^{-1}cm$, in line with previous estimates (the effect sizes vary between $.03cm$ and $1cm$ according to [186]). EXPRESS IN NATURAL UNIT MEASURE? (METER) In [15], the UK biobank was used to estimate the parameter $\frac{\omega^{-2}}{2}$, which was found to be of order 1.5×10^{-2} in units of σ^{-2}, with σ of order $10cm$. Finally, we take the mean mutation rate $|\bar{\theta}|$ per locus per generation to be $10^{-8}N_e$, where N_e is the effective population size [187]. If we neglect environmental variation for simplicity, we may compute

$$\begin{aligned}\omega_e^{-2} =& N_e(L\bar{\alpha})^2\omega^{-2}\\ =& 20,000 \times (10^7 \times 10^{-1})^2 \times (2 \times 1.5 \times 10^{-2}/10^2)\\ =& 6 \times 10^{12}\end{aligned}$$

This is compatible with moderate selection. Of course, applying this crude computation to human data assumes equilibrium and ignores population structure, linkage, the fluctuations of selection, the environmental contribution to height, pleiotropy and the fact that the additive effects α have a heavy-tailed distribution.

4.4.4 Applicability to GWAS data

Our theoretical model yields a stationary distribution for the joint distribution of allele frequency and additive effect (P_t, α). This distribution is precisely what GWAS are measuring, and which the works of [118, 116, 117, 120] are interpreting from an evolutionary standpoint. Our contribution to their framework is the bias-correcting selection coefficient s^*. From (4.10), we expect the distribution of P_t conditional on α to be proportional to $\Pi[\xi_\alpha, \mu, N_e]$ where

$$\xi_\alpha(x) := A\alpha + B\alpha^2\left(\frac{1}{2} - x\right) \tag{4.21}$$

for some constants A and B which could be estimated empirically by maximal likelihood.

Empirically, [118] validates our core idea by detecting asymmetries in the distribution of (P_t, α). More precisely, their model is concerned with (Y_t, β) where Y_t is the frequency of the derived allele (as opposed to the ancestral allele) and β is its additive effect. Compared to our model, β can be negative, but $|\beta| = \alpha$. When $\beta < 0$, then $Y_t = 1 - P_t$, otherwise $Y_t = P_t$. They consistently reject the null hypothesis of unbiased mutations (same average value of β conditional on $Y_t = p$ or on $Y_t = 1 - p$) in 6 out of 9 GWAS of human traits. One would expect this to mean our model (4.21) would fare well in practice.

It turns out that model (4.21) has been tested in the case of quantitative disease GWAS in [120] (not yet peer-reviewed). In this preprint, the authors do not really justify model (4.21), but test it against the "traditional" model of [116] which corresponds to setting $A = 0$. Model (4.21) fares rather poorly, and for only 2 of 27 tested traits this model outperforms the $A = 0$ model in terms of restricted maximum likelihood (versus 21 traits that support the $A = 0$ model).

If it was confirmed that model (4.21) generally underperforms with respect to the model with $A = 0$, this could be interpreted in at least two different directions. First, we assume that genetic diversity is maintained by mutations. It can also be maintained by spatial structure [138]: if a very large population is subdivided into many small demes, then genetic diversity within each deme can be maintained by migration. In our model, if we assume a Levene migration model [188] we could add this effect by replacing the mutation probabilities (μ_ℓ^+, μ_ℓ^-) with $(\mu_\ell^+ + m\mathbb{E}^*[P_t^\ell], \mu_\ell^- + m\mathbb{E}^*[1 - P_t^\ell])$ for some migration parameter $m \gg \mu_\ell^{(\pm)}$. In such a setting, mutational bias should become negligible. Second, it could be seen as a signature of selection on the distribution (α, θ) itself. For instance, the mutation rate is highly variable across the genome [56], shaped by evolvable factors such as chromatin structure [189]. The distribution of α be shaped through second-order selection favoring robust gene regulatory networks. A quantitative trait under stabilizing selection can evolve towards robustness in a stable environment, because selection favors genes which reduces the genetic variance in the trait [190]. The process by which stabilizing selection shapes the distribution of α is called genetic canalization [191] (see also [192]).

A framework for a systematic evolutionary analysis GWAS analysis was recently proposed [117] (currently in revision). The model considers that each locus has a highly pleiotropic effect, which determines its selection coefficient in Robertson's underdominant term (the pleiotropic equivalent of the second term of (4.10), see Appendix 4.I.2). It is assumed that conditional on this selection coefficient, the additive effect of the alleles at this locus on a given focal trait has a symmetric distribution. This assumption posits that selection and mutation act identically on trait-increasing and trait-decreasing alleles, implying that $z_M = z_H = \eta$ and that the distributions of (α, P_t) and $(\alpha, 1 - P_t)$ are identical. We would argue for a generalization of this approach including a possible bias in mutation and in allelic effects, for example through the addition of a hyperparameter.

4.4.5 Extensions and applications

Many extensions are possible. In Appendix 4.I.1 we mention polyploidy, for which the results are essentially unchanged, replacing $2N_e$ with kN_e where k is the ploidy. In Appendix 4.I.2 we add pleiotropy: each allele influences d traits. Future work could study this behavior when $d \gg 1$, which should agree with the conclusions of [116, 118]. In Appendix 4.I.3 we add dominance: the selection coefficient (4.10) is then a polynomial in P_t^ℓ of degree 3. In Appendix 4.I.4 we discuss epistasis. We argue that under some conditions, the effect of a locus on the trait can be locally approximated as additive. It will also be straightforward to add fluctuating environments as in [26] (provided the fluctuations are of the same order as σ or smaller), genes influencing plasticity (G×E interactions, which are studied from the trait's eye-view in [193]), spatial structure as in [138], and more sophisticated selection (for instance if $W(z)$ is skewed).

One could also hope to derive first-order corrections for linkage disequilibrium, following [106, 105]. Another mechanism to consider would be insertion-deletions, which would let the number of loci L evolve [194, 195, 196, 197, 198]. On another level, there is growing evidence for the rôle of introgression and admixture [199] as an alternative to mutation for generating genetic diversity. Theoretical models such as [200] offer powerful descriptions of the long-term effect of an admixture event on a polygenic trait. These models allow for the representation of migration as an input in genetic variability under assumptions of strong recombination and weak selection [136]. Finally, one should account for the evolution of the effective population size N_e, the fluctuations of which can be influenced directly by the effect of selection [38, 138, 137].

Future work will also describe the dynamics of the system out-of-equilibrium, extending the work of [124, 126]. This will allow us to tackle questions such as the response of a population to a decrease in population size or a change in optimum, and the corresponding genetic load [201]. Overall, our framework lets us account for most biologically relevant features of polygenic systems (with the exception of LD) and provides efficient theoretical predictions as well as explicit criteria on the parameters for our predictions to hold.

Material and methods

Our programs were run on Python 3.10.12. All of our code is available at https://github.com/PhCourau /gene_s_eye_view_of_quantitative_genetics_applications. Our derivations rely on the following three kinds of mathematical tools.

4.A Model and notation

To make this appendix self-contained, we recall the model and notation from the chapter.

4.A.1 Miscellaneous notations

For $a, b \in \mathbb{R}_+$, the notation $a \lesssim b$ is taken to mean that we don't have $a \gg b$, and $a \sim b$ means we neither have $a \ll b$ nor $b \ll a$. We use the notation $[L] = \{1, \dots, L\}$.

4.A.2 Individual-based model

The individual-based model as implemented in our numerical simulation is a classical diploid $L-$loci biallelic Wright-Fisher model with a population of size N. Each organism can be described by its **genome** $g = (g_\ell)_{\ell\in[L]} \in \{0, 1, 2\}^L$, with g_ℓ representing the number of trait-increasing allele at locus ℓ. The (genetic) trait value of a genome is given by the **trait function**

$$Z : \begin{cases} \{0,1,2\}^L & \longrightarrow \mathbb{R} \\ g & \longmapsto \sum_{\ell\in[L]} \alpha_\ell g_\ell \end{cases}$$

where $\alpha_\ell \geq 0$ is the additive coefficient at locus ℓ. We take the $(\alpha_\ell)_{\ell\in[L]}$ to be an exchangeable vector of random variables such that

$$\sum_{\ell\in[L]} \alpha_\ell = 1. \tag{4.22}$$

In particular, unless their distribution is degenerate (see Fig. 4.11) the typical value of α_ℓ is of order $1/L$. The assumption of (4.22) is made to guarantee that $Z(g)$ is always between 0 and 2. It can be seen as a form of rescaling: we measure the trait in units such that the maximal possible trait value is 2 and the smallest one is 0.

The **fitness** of an organism is obtained with the **fitness function**

$$F(z) : \begin{cases} \mathbb{R} & \longrightarrow \mathbb{R} \\ z & \longmapsto \exp\left(-\frac{\omega^{-2}}{2}(z-\eta)^2\right) \end{cases}$$

We let $W := \ln F$ be the **log-fitness** function.

Reproduction occurs by sampling two parents with probability proportional to fitness, and for each parent picking a crossover position uniformly on $[L]$. Mutations at locus ℓ to (resp. from) the trait-increasing allele occur with probability μ_ℓ^+ (resp. μ_ℓ^-) per locus per generation per haploid genome. A formal definition of the model will be given in Section 4.C.3.

4.A.3 Representation from allele frequencies

For $t \geq 0$ we let P_t^ℓ be the frequency of the trait-increasing allele at generation $\lfloor 2Nt \rfloor$ at locus ℓ.

For $\mathbf{p} = (p_\ell)_{\ell\in[L]} \in [0,1]^L$ we let $\mathbf{E_p}[\varphi(g)]$ be the expectation of $\varphi(g)$ when $g = (g_\ell)_{\ell\in[L]}$ is a vector of independent variables such that g_ℓ has law $Binomial(2, p_\ell)$. We write $\mathbf{Var_p}, \mathbf{Cov_p}$ for the variance and covariance associated to $\mathbf{E_p}$. For all other random variables the notation $\mathbb{E}, \mathbb{V}\mathbf{ar}$ and $\mathbb{C}\mathbf{ov}$ are used.

Consider $t \geq 0$. The population at time t is in **Hardy-Weinberg Linkage Equilibrium (HWLE)** if conditional on $\mathbf{P}_t = (P_t^\ell)_{\ell\in[L]}$, the law of of a uniformly sampled genome is $\mathbf{E}_{\mathbf{P}_t}$.

4.A.4 Typical locus and genetic architecture

For $\ell \in [L]$, we define the summary vector at locus ℓ at time t as the vector

$$\vec{P}_t^\ell := (P_t^\ell, \alpha_\ell, \theta_\ell).$$

This vector contains all the information we need from locus ℓ at time t. Consider ℓ_U a uniform random variable on $[L]$. We call

$$\vec{P}_t = (P_t, \alpha, \theta) := \vec{P}_t^{\ell_U}$$

the **typical locus at time** t. The **genetic architecture** at time t is the law of $\vec{P}_t$.

4.A.5 Wright-Fisher diffusion

A process $(P_t)_{t\geq 0}$ is a standard Wright-Fisher diffusion with parameters $s > 0, \theta^+ > 0, \theta^- > 0$ if it satisfies the Stochastic Differential Equation (SDE)

$$\mathrm{d}P_t = sP_t(1-P_t)\mathrm{d}t + (\theta^+(1-P_t) - \theta^- P_t)\mathrm{d}t + \sqrt{P_t(1-P_t)}\mathrm{d}B_t \tag{4.23}$$

where B_t is a Brownian motion.

Similarly, a process $(P_t)_{t\geq 0}$ is a frequency-dependent Wright-Fisher diffusion with parameters ξ, θ^+, θ^- where ξ is a continuous function on $[0,1]$ and $\theta^+ > 0, \theta^- > 0$ if it satisfies the SDE

$$\mathrm{d}P_t = \xi(P_t)P_t(1-P_t)\mathrm{d}t + (\theta^+(1-P_t) - \theta^- P_t)\mathrm{d}t + \sqrt{P_t(1-P_t)}\mathrm{d}B_t \tag{4.24}$$

4.B Outline of the derivation

We outline the content of Sections 4.C-4.F.

4.B.1 The polygenic equation from the gene's eye view

In Sections 4.C-4.D, we will derive an equation that jointly describes the evolution of the typical locus P_t and the mean trait value within the population. This will be done in two successive steps.

The diffusion equation

Our first step relies on deriving coupled diffusion equations for the evolutions of the allelic frequencies at various loci. Define Δ_t as the deviation from the optimum at time t

$$\Delta_t := \mathbf{E}_{\mathbf{P}_t}[Z(g)] - \eta = \sum_{\ell \in [L]} 2\alpha_\ell P_t^\ell \; - \eta. \tag{4.25}$$

In Section 4.C, we derive the following diffusion equation for P_t^ℓ on the time scale N

$$\mathrm{d}P_t^\ell = \xi_{\Delta_t,\alpha_\ell}(P_t^\ell) P_t^\ell(1-P_t^\ell)\,\mathrm{d}t + \;\left(\theta_\ell^+(1-P_t^\ell) - \theta_\ell^- P_t^\ell\right)\mathrm{d}t + \sqrt{P_t^\ell(1-P_t^\ell)}\,\mathrm{d}B_t^\ell, \tag{4.26}$$

where $(B^\ell)_{\ell\in[L]}$ are independent Brownian motions, $\theta_\ell^\pm = 2N\mu_\ell^\pm$, and the selection coefficient ξ is such that for $a \in \mathbb{R}_+, \delta \in \mathbb{R}$:

$$\xi_{\delta,a}(p) := -\frac{a\delta}{\omega_e^2} + \frac{a^2}{\omega_e^2}\left(p - \frac{1}{2}\right). \tag{4.27}$$

where

$$\omega_e^2 := \frac{\omega^2}{2N}.$$

The derivation of of this result relies on the following hypotheses

H 1. *The population remains close to HWLE.*

H 2. *The law of $F(Z(g))$ when g has law $\mathbf{E}_{\mathbf{P}_t}$ is very concentrated around $\hat{F}_t := e^{\bar{W}_t}$ where $\bar{W}_t = \mathbf{E}_{\mathbf{P}_t}[Z(g)]$, so that we may write*

$$F(Z(g)) \simeq \hat{F}_t(1 + W(Z(g)) - \bar{W}_t).$$

(H1) implies the $(P_t^\ell)_{\ell\in[L]}$ are sufficient to fully describe the population at time t. (H2) should be interpreted as: the typical fitness difference between two randomly sampled organisms is small.

Mean-field approximation

From now on, we simplify the problem by considering that the system is at statistical equilibrium, and we denote by $\mathbb{P}^*$ the equilibrium distribution of the typical locus $\vec{P}_t$. The diffusion (4.26) is of dimension $L-1$, which makes it inconvenient to work with when $L \gg 1$. In Section 4.D, we derive a simplified system for the dynamics of the typical locus from (4.26) using mean-field approximations.

Mean-field approximations are an important tool in the study of interacting particles [69, 157]. In our setting, the mean-field approximation reads

H 3 (Mean-field approximations)**.** *For any test function f,*

$$\frac{1}{L}\sum_{\ell\in[L]} f(\vec{P}_t^\ell) \simeq \mathbb{E}^*[f(\vec{P}_t)] = \mathbb{E}^*[f(\vec{P}_0)].$$

where the last equality follows from the fact that the system is assumed to be at equilibrium.

The left hand side is a random quantity, whereas the right-hand side is deterministic and is obtained by looking at the expected value of an observable of a typical locus at time 0 (under the equilibrium distribution $\mathbb{P}^*$).

Intuitively, in our system, the loci are coupled by Δ_t in the selection coefficient (4.27), and Δ_t is determined by all loci. But this dependence is diffuse and evenly spread across all loci, as each locus only has a very small influence on Δ_t. So the dependence between different loci is weak, and we may expect (H3) to hold. This phenomenon, known as **propagation of chaos** [69], has been rigorously proven to occur in our system when $\omega_e^{-2} \sim L$ in Chapter 2.

In Section 4.D, we formally derive the limit equation for the distance of the trait mean to the optimum and a typical locus P_t at stationarity, and the fluctuations of the mean trait value $\varepsilon_t := \Delta_t - \Delta^*$.

$$\mathrm{d}P_t = \xi_{\Delta^*,\alpha}(P_t)P_t(1-P_t)\,\mathrm{d}t + \left(\theta^+(1-P_t) - \theta^- P_t\right)\mathrm{d}t + \sqrt{P_t(1-P_t)}\,\mathrm{d}B_t^P \tag{4.28}$$

$$\Delta^* := \mathbb{E}^*[\Delta_0] = \mathbb{E}^*[\Delta_t] = 2L\mathbb{E}^*[\alpha P_t] - \eta \tag{4.29}$$

$$\mathrm{d}\varepsilon_t = -\rho\varepsilon_t\,\mathrm{d}t + \omega_e\sqrt{2\rho}\,\mathrm{d}B_t^\Delta \tag{4.30}$$

where B^P, B^Δ are Brownian motions. In particular, $(\varepsilon_t)_{t\geq 0}$ is an Ornstein-Uhlenbeck process with **autocorrelation parameter** ρ defined as

$$\rho := \frac{1}{\tau} \times \frac{\mathbb{E}^*[2(L\alpha)^2 P_t(1-P_t)]}{|\bar{\theta}|} \tag{4.31}$$

where the **characteristic timescale of the trait** τ are

$$\tau := \frac{L\omega_e^2}{|\bar{\theta}|}. \tag{4.32}$$

where $|\theta| := \theta^+ + \theta^-$ and $\bar{\theta} = (\mathbb{E}[\theta^+], \mathbb{E}[\theta^-])$. This system is much more convenient than the SDE (4.26), and in particular it has a simple stationary distribution (see Section 4.B.1).

The derivation will be based on the mean-field hypothesis (H3) and either one of the following two hypotheses

H 4. *We have* $\omega_e^{-2} \ll L^2$.

H 4'. *We have*

a) $\omega_e^{-2} \sim L^2$

b) $\mathbb{E}^*\left[(L\alpha)^2 P_t(1-P_t)\right] \sim |\bar{\theta}|$

c) $L\omega_e^2 \ll |\bar{\theta}|$

d) the timescale on which $(P_t)_{t\geq 0}$ *evolves is of order* 1 *or greater*

(H4) corresponds to what we call the weak/moderate selection regime. (H4'a) corresponds to the strong selection regime.

(H4'b) is a classical result of population genetics. If we take $\hat{P}_t$ a neutral Wright-Fisher diffusion with mutation rates θ, then at stationarity $\hat{P}_t$ has distribution $Beta(2\theta^+, 2\theta^-)$, and in particular

$$\mathbb{E}\left[\hat{P}_t(1-\hat{P}_t)\right] = \frac{\theta^+\theta^-}{|\theta|\left(\frac{1}{2} + |\theta|\right)}$$

In particular, provided $\theta^+ \sim \theta^- \lesssim 1$, we have

$$\mathbb{E}\left[\hat{P}_t(1-\hat{P}_t)\right] \sim |\theta|$$

The same holds if $\hat{P}_t$ is a frequency-dependent Wright-Fisher diffusion (4.24) with selection coefficient of order 1. In light of this, we may expect (H4'b) to hold as soon as $\alpha \sim 1/L$ and the mutational bias is small. See Section 4.E for further details.

(H4'c) can be rewritten $\tau \ll 1$. (H4'd) is verified for any Wright-Fisher diffusion in which genetic drift plays a significant rôle.

Stationary distribution

Using the decoupling of P_t and Δ_t, we can rewrite (4.28) as

$$\mathrm{d}P_t^\ell = \xi_{\Delta^*,\alpha_\ell}(P_t^\ell)P_t^\ell(1-P_t^\ell)\,\mathrm{d}t + \left(\theta_\ell^+(1-P_t) - \theta_\ell^- P_t\right)\mathrm{d}t + \sqrt{P_t^\ell(1-P_t^\ell)}\,\mathrm{d}B_t^\ell \tag{4.33}$$

This means that conditional on $(\alpha_\ell, \theta_\ell)$, $(P_t^\ell)_{t\geq 0}$ behaves as a frequency-dependent Wright-Fisher diffusion (4.24). In particular, its equilibrium distribution [53] is $\Pi_{\Delta^*,\alpha_\ell,\theta_\ell}$ where for $\delta \in \mathbb{R}, a \in \mathbb{R}_+, \theta \in (0,+\infty)^2$ we define

$$\Pi_{\delta,a,\theta}(p) := C_{\delta,a,\theta}\, p^{2\theta^+-1}(1-p)^{2\theta^--1}e^{2\int_0^p \xi_{\delta,a}(u)\,\mathrm{d}u}. \tag{4.34}$$

with $C_{\delta,a,\theta}$ a normalization constant. Furthermore, the distribution of a typical locus $\vec{P}_t$ is given by

$$\mathbb{E}^*[f(\vec{P}_t)] = \mathbb{E}\left[\int f(p,\alpha,\theta)\Pi_{\Delta^*,\alpha,\theta}(p)\mathrm{d}p\right]$$

Since we know the distribution of a typical locus, we can compute

$$\Delta^* \simeq 2L\mathbb{E}^*[\alpha P_t] - \eta.$$

This can be rewritten

$$\Delta^* \simeq 2\, I(\Delta^*) - \eta \tag{4.35}$$

with

$$I(\delta) := \mathbb{E}\left[\int L\alpha p \Pi_{\delta,\alpha,\theta}(p)\mathrm{d}p\right]. \tag{4.36}$$

With words, $I(\delta)$ is the expectation of $L\alpha\hat{P}_t^\delta$, where $\hat{P}_t^\delta$ is a stationary frequency-dependent Wright-Fisher diffusion with parameters $(\xi_{\delta,\alpha}, \theta)$ ((4.24)).

In mathematical terms, (4.35) is what is known as a fixed point equation: we obtain Δ^* as a function of Δ^*. The fact that this equation has a unique solution can be seen from the fact that I is continuous, non-increasing and bounded by 1 on $\mathbb{R}$.

4.B.2 Simplifying assumptions on the parameters

In Section 4.E, we consider the stationary solution of the system given by (4.28-4.30), and we derive the order of magnitude of macroscopic observables based on the following assumptions.

A 1 (Uniform boundedness)**.** *There is a constant* $C \sim 1$ *such that for any* $\ell \in [L]$, $|\theta_\ell| \leq C|\bar{\theta}|$ *and* $\alpha_\ell \leq \frac{C}{L}$.

A 2 (Mutations smaller than genetic drift)**.** $|\bar{\theta}| \lesssim 1$.

A 3 (Mutational bias not too extreme)**.** *There is a constant* $C \sim 1$ *such that for any* ℓ, $\theta_\ell^-/C \leq \theta_\ell^+ \leq C\theta_\ell^-$.

A 4 (Accessibility of the selection optimum)**.** $\eta \in (0,2)$ *satisfies* $\eta(2-\eta) \sim 1$.

A 5 (Weak/moderate/strong selection)**.** $L \lesssim \omega_e^{-2} \lesssim L^2$.

A 6 (Distance between the selection and the mutation optimum)**.** *We have*

$$|2\, I(0) - \eta| \sim 1.$$

(A1) is a strong assumption that guarantees no single locus disproportionally contributes to the genetic variance. It makes computations much more tractable. Future work should relax this assumption (see also Fig. 4.11). (A2-3) ensure mutations rates are not too large or too asymmetric. (A4) ensures that there are many different genotypes that can satisfy $Z(g) = \eta$. (A5) rules out ultra-weak ($\omega_e^{-2} \ll L$) selection, under which the system is close to neutrality (see Section 4.E.2), and ultra-strong $\omega_e^{-2} \gg L^2$, under which natural selection strongly depletes the genetic variability. (A6) is quite technical, but when $\omega_e^{-2} \ll L^2$ (weak/moderate selection) it can be rewritten $|z_M - \eta| \sim 1$ where $z_M := 2L\mathbb{E}\left[\alpha\frac{\theta^+}{\theta^+ + \theta^-}\right]$ is the mutational optimum (the mean trait value if there was no selection). In such a setting, (A6) precludes coincidental situations in which the mutational optimum would be very close to the selection optimum. Under strong selection ($\omega_e^{-2} \sim L^2$), the condition is more complex, but it similarly rules out a very coincidental situation.

4.B.3 Necessary assumptions for consistency

In Section 4.F, we will show that the following assumptions are necessary for (A1-6) to be consistent with (H1-4').

N 1 (Sufficiently large population)**.** $2N \gg |\bar{\theta}|\frac{\ln(L)}{L\omega_e^2} + L\sqrt{|\bar{\theta}|}$.

N 2 (Minimal mutational input every generation)**.** $|\bar{\theta}|L \gg 1$.

Furthermore, the following assumption (stronger than (N2)) is needed for (4.30) to describe the fluctuations of the distance to the optimum $(\varepsilon_t)_{t\geq 0}$.

N 3 (Sufficient mutational input)**.** $\frac{|\bar{\theta}|}{L\omega_e^2} \gg 1$.

Condition (N2) guarantees that sufficient genetic variability within the population is maintained by mutations. (N1) ensures HWLE (H1) and that the fitness variance is small (H2). Finally, (N3) guarantees that selection is the force stabilizing the fluctuations of $(\varepsilon_t)_{t\geq 0}$ (see Section 4.F for details), it can be rewritten as $\tau \ll 1$ with τ the timescale of the trait from (4.32).

4.C The diffusion approximation

In this section, we aim at obtaining the SDE (4.26), assuming $N \gg 1$, $\theta_\ell^\pm = \mu_\ell^\pm 2N \sim 1$, HWLE (H1) and that fitness is very concentrated (H2). Equation (4.26) is well-known to be the limit of the individual-based model under the hypothesis of HWLE as $N \to +\infty$, when selection and mutation are weak (see [72]), and the selection coefficient (4.27) was obtained by Wright in [96]. We recall the derivation of these equations in an effort to be self-contained.

We will obtain the diffusion equation by computing the first and second moments. Specifically, we must show

$$\mathbb{E}\left[P^\ell_{t+\frac{1}{2N}} - P^\ell_t \middle| \mathbf{P}_t\right] \simeq \frac{1}{2N}\left(\xi_{\Delta_t,\alpha_\ell}(P^\ell_t)P^\ell_t(1-P^\ell_t) + \theta^+_\ell(1-P^\ell_t) - \theta^-_\ell P^\ell_t\right) \tag{4.37}$$

$$\mathbb{V}\mathbf{ar}\left[P^\ell_{t+\frac{1}{2N}} \middle| \mathbf{P}_t\right] \simeq \frac{1}{2N}P^\ell_t(1-P^\ell_t) \tag{4.38}$$

$$\left|\mathbb{C}\mathbf{ov}\left[P^{\ell_1}_{t+\frac{1}{2N}}, P^{\ell_2}_{t+\frac{1}{2N}} \middle| \mathbf{P}_t\right]\right| \ll \frac{1}{2N} \tag{4.39}$$

4.C.1 First moment computation

We write $P_t^{*\ell}$ for the frequency of the trait-increasing allele at generation $\lfloor 2Nt \rfloor$ after reproduction but before mutation. Assume the population at time $2Nt$ is in HWLE (H1). Because a genome g from generation $\lfloor 2Nt \rfloor$ has an average number of offspring proportional to $F(Z(g))$, and because $P_t^{*\ell}$ is half the expectation of g_ℓ post-reproduction, we find

$$\mathbb{E}\left[P_t^{*\ell} \,|\mathbf{P}_t\right] = \frac{\mathbf{E}_{\mathbf{P}_t}[\frac{g_\ell}{2}F(Z(g))]}{\mathbf{E}_{\mathbf{P}_t}[F(Z(g))]}.$$

It follows from $P_t^\ell = \frac{1}{2}\mathbf{E}_{\mathbf{P}_t}[g_\ell]$

$$\mathbb{E}\left[P_t^{*\ell} - P_t^\ell \,|\mathbf{P}_t\right] = \frac{\mathbf{Cov}_{\mathbf{P}_t}[F(Z(g)), g_\ell]}{2\mathbf{E}_{\mathbf{P}_t}[F(Z(g))]}.$$

As a sidenote, this can also be seen as an application of the Price equation [143] to the trait $\frac{g_\ell}{2}$. Writing $\hat{F}_t = e^{\bar{W}_t}$ we get from (H2)

$$\frac{\mathbf{Cov}_{\mathbf{P}_t}[F(Z(g)), g_\ell]}{2\mathbf{E}_{\mathbf{P}_t}[F(Z(g))]} \simeq \frac{\mathbf{Cov}_{\mathbf{P}_t}[\hat{F}_t(1 + W(Z(g)) - \bar{W}_t), g_\ell]}{2\hat{F}_t} = \frac{\mathbf{Cov}_{\mathbf{P}_t}[W(Z(g)), g_\ell]}{2}$$

Using $\mathbf{Var}_{\mathbf{P}_t}[g_\ell] = 2P_t^\ell(1 - P_t^\ell)$ we find

$$\mathbb{E}\left[P_t^{*\ell} - P_t^\ell \,|\mathbf{P}_t\right] \simeq \frac{s_\ell(\mathbf{P}_t)}{2N} P_t^\ell(1 - P_t^\ell)$$

where s_ℓ is the selection coefficient at locus ℓ

$$s_\ell : \begin{cases} [0,1]^L & \longrightarrow \mathbb{R} \\ \mathbf{p} & \longmapsto 2N\frac{\mathbf{Cov}_{\mathbf{p}}[W(Z(g)), g_\ell]}{\mathbf{Var}_{\mathbf{p}}[g_\ell]} \end{cases} \tag{4.40}$$

In particular, $\frac{s_\ell}{2N}$ can be seen as the linear regression of $W(Z(g))$ on g_ℓ (see Fig. 4.6).

Taking into account the effect of mutation we find

$$\mathbb{E}\left[P_{t+\frac{1}{2N}}^\ell - P_t^\ell \middle| \, \mathbf{P}_t\right] \simeq \frac{1}{2N}\left(s_\ell(\mathbf{P}_t)P_t^\ell(1 - P_t^\ell) + \theta_\ell^+(1 - P_t^\ell) - \theta_\ell^- P_t^\ell\right)$$

To get (4.37), we must show $s_\ell(\mathbf{P}_t) = \xi_{\Delta_t, \alpha_\ell}(P_t^\ell)$.

4.C.2 Selection coefficient

We find from the definition of s_ℓ and W

$$\begin{aligned}
\frac{s_\ell(\mathbf{p})}{2N} &= \frac{\mathbf{Cov}_{\mathbf{p}}\left[-\frac{1}{2\omega^2}\left(\sum_{\ell' \in [L]} \alpha_{\ell'} g_{\ell'} - \eta\right)^2, g_\ell\right]}{2p_\ell(1 - p_\ell)} \\
&= -\frac{1}{4\omega^2 p_\ell(1 - p_\ell)}\left(\mathbf{Cov}_{\mathbf{p}}\left[\sum_{\ell_1, \ell_2 \in [L]} \alpha_{\ell_1} g_{\ell_1} \alpha_{\ell_2} g_{\ell_2}, g_\ell\right] - 2\eta\alpha_\ell \mathbf{Var}_{\mathbf{p}}\left[g_\ell\right]\right) \\
&= -\frac{1}{4\omega^2 p_\ell(1 - p_\ell)}\left(\mathbf{Cov}_{\mathbf{p}}\left[\sum_{\substack{\ell' \in [L] \\ \ell' \neq \ell}} 2\alpha_{\ell'} g_{\ell'} \alpha_\ell g_\ell, g_\ell\right] + \mathbf{Cov}_{\mathbf{p}}\left[(\alpha_\ell g_\ell)^2, g_\ell\right] - 4\eta\alpha_\ell p_\ell(1 - p_\ell)\right)
\end{aligned}$$

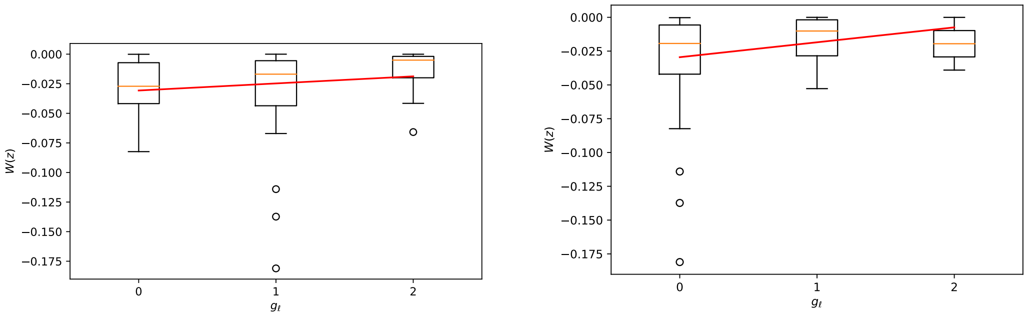

Figure 4.6: A population of $N = 100$ individuals with $L = 100$ biallelic loci, was evolved for $T = 100$ generations under stabilizing selection with parameters $\eta = 1.2, \omega_e^{-2} = 10^3$ and $\theta = (0.1, 0.2)$. The logfitness of the organisms within the population is plotted as a function of g_ℓ. The left and right figure correspond to two different loci ($\ell = 0$ and $\ell = 7$ respectively). The selection coefficient $s_\ell(\mathbf{P}_t)$ at generation t at locus ℓ is given by the linear regression coefficient of logfitness $W(Z(g))$ on g_ℓ (see Section 1A). This corresponds to the slope of the red line.

where in the last equality we used (H1) which guarantees the independence of $g_{\ell_1} g_{\ell_2}$ and g_ℓ when $\ell \notin \{\ell_1, \ell_2\}$. This independence further yields for $\ell' \neq \ell$

$$\begin{aligned}\mathbf{Cov}_{\mathbf{p}}\left[\alpha_{\ell'} g_{\ell'} \alpha_\ell g_\ell, g_\ell\right] =& \alpha_{\ell'} \mathbf{E}_{\mathbf{p}}\left[g_{\ell'}\right] \alpha_\ell \mathbf{Var}_{\mathbf{p}}[g_\ell] \\ =& \alpha_{\ell'} \times 2p_{\ell'} \times \alpha_\ell \times 2p_\ell(1-p_\ell)\end{aligned}$$

Similarly we can compute

$$\mathbf{Cov}_{\mathbf{p}}\left[\left(\alpha_\ell g_\ell\right)^2, g_\ell\right] = \alpha_\ell^2\left(2+4p_\ell\right)p_\ell(1-p_\ell)$$

From there it is elementary to obtain

$$\frac{s_\ell(\mathbf{p})}{2N} = \frac{\alpha_\ell}{\omega^2}\left(\eta - 2\sum_{\ell' \in [L]} \alpha_{\ell'}\, p_{\ell'}\right) + \frac{\alpha_\ell^2}{\omega^2}\left(p_\ell - \frac{1}{2}\right)$$

We thus get (4.27) using (4.25).

4.C.3 Second moment

Here we compute $\mathbb{V}\mathbf{ar}\left[P^\ell_{t+\frac{1}{2N}} \,\middle|\, \mathbf{P}_t\right]$ and $\mathbb{C}\mathbf{ov}\left[P^{\ell_1}_{t+\frac{1}{2N}}, P^{\ell_2}_{t+\frac{1}{2N}} \,\middle|\, \mathbf{P}_t\right]$ for $\ell_1 \neq \ell_2$. For this, it will be useful to give a more formal definition of the individual-based model.

Formal definition of the individual-based model.

The population is described with an array $(G^{n,i}_{\ell,(j)})_{\ell\in[L], j\in[2], i\in[N]}$ in which

- $n \in \mathbb{N}$ is a time-coordinate, it denotes the generation under consideration
- $i \in [N]$ is the label of the organism under consideration
- $j \in [2]$ is the label of the chromosome
- $\ell \in [L]$ is the label of the locus.

In particular, we have $G^{n,i}_{\ell,(j)} = 1$ (resp. 0) if the $i-$th organism at generation n has the trait-increasing (resp. decreasing) allele on its $j-$th chromosome, at locus ℓ. Then $g^{n,i} = (g^{n,i}_\ell)_{\ell\in[L]} := (G^{n,i}_{\ell,(1)} + G^{n,i}_{\ell,(2)})_{\ell\in[L]}$ is the genome of organism i at generation n.

The generation $n+1$ is generated from generation n as follows. For $i \in [N], j \in [2]$, the $j-$th chromosome of the $i-$th organism $(G^{n+1,i}_{\ell,(j)})_{\ell\in[L]}$ is independently generated in two steps

- *(Reproduction).* we sample a parental genome $I_{i,j}$ with probability proportional to $F(Z(g^{n,I_{i,j}}))$. A crossover positions $\ell^{(j)}_U$ is uniformly sampled on $[L]$, as well as a *Bernoulli*$(1/2)$ variable b_j. Then we set

$$\forall \ell \in [L], \qquad G^{n+1,i}_{\ell,(j)} = \begin{cases} G^{n,I_{i,j}}_{\ell,(b_j)} & \text{if } \ell < \ell^{(j)}_U \\ G^{n,I_{i,j}}_{\ell,(1-b_j)} & \text{otherwise} \end{cases}$$

- *(Mutation).* With probability $|\mu_\ell| := \mu^+_\ell + \mu^-_\ell$, $G^{n,i}_{\ell,(j)}$ mutates and is replaced by an independently sampled variable with law *Bernoulli*$(\mu^+_\ell/(\mu^+_\ell + \mu^-_\ell))$.

Second moment: diagonal coefficients.

From the formal model, we see that each $(G^{\lfloor 2Nt\rfloor+1,i}_{\ell,(j)})_{i\in[N],j\in[2]}$ is independently generated with the same procedure. We therefore find

$$\mathbb{V}\mathbf{ar}\left[P^\ell_{t+\frac{1}{2N}} \,\middle|\, \mathbf{P}_t\right] = \frac{1}{(2N)^2}\sum_{i\in[N],j\in[2]} \mathbb{V}\mathbf{ar}\left[G^{\lfloor 2Nt\rfloor+1,i}_{\ell,(j)} \,\middle|\, \mathbf{P}_t\right] = \frac{1}{2N}\mathbb{V}\mathbf{ar}\left[G^{\lfloor 2Nt\rfloor+1,1}_{\ell,(1)} \,\middle|\, \mathbf{P}_t\right].$$

Because $G^{\lfloor 2Nt\rfloor+1,1}_{\ell,(1)}$ is a Bernoulli variable we find

$$\mathbb{V}\mathbf{ar}\left[P^\ell_{t+\frac{1}{2N}} \,\middle|\, \mathbf{P}_t\right] = \frac{1}{2N}\mathbb{E}\left[G^{\lfloor 2Nt\rfloor+1,1}_{\ell,(1)} \,\middle|\, \mathbf{P}_t\right]\mathbb{E}\left[1 - G^{\lfloor 2Nt\rfloor+1,1}_{\ell,(1)} \,\middle|\, \mathbf{P}_t\right].$$

Since conditional on $P^\ell_{t+\frac{1}{2N}}$, $G^{\lfloor 2Nt\rfloor+1,1}_{\ell,(1)}$ has law *Bernoulli*$(P^\ell_{t+\frac{1}{2N}})$, we find

$$\mathbb{V}\mathbf{ar}\left[P^\ell_{t+\frac{1}{2N}} \,\middle|\, \mathbf{P}_t\right] = \frac{1}{2N}\mathbb{E}\left[P^\ell_{t+\frac{1}{2N}} \,\middle|\, \mathbf{P}_t\right]\mathbb{E}\left[1 - P^\ell_{t+\frac{1}{2N}} \,\middle|\, \mathbf{P}_t\right]. \tag{4.41}$$

Using the first-order approximation $\mathbb{E}\left[P^\ell_{t+\frac{1}{2N}} \,\middle|\, \mathbf{P}_t\right] \simeq P^\ell_t$ yields (4.38).

Second moment: cross coefficients

We can use again the fact that $G^{\lfloor 2Nt\rfloor+1,i_1}_{\ell_1,(j_1)}$ and $G^{\lfloor 2Nt\rfloor+1,i_2}_{\ell_1,(j_2)}$ are independently generated whenever $j_1 \neq j_2$ or $i_1 \neq i_2$ to find

$$\mathbb{C}\mathbf{ov}\left[P^{\ell_1}_{t+\frac{1}{2N}}, P^{\ell_2}_{t+\frac{1}{2N}} \,\middle|\, \mathbf{P}_t\right] = \frac{1}{2N}\mathbb{C}\mathbf{ov}\left[G^{\lfloor 2Nt\rfloor+1,1}_{\ell_1,(1)}, G^{\lfloor 2Nt\rfloor+1,1}_{\ell_2,(1)} \,\middle|\, \mathbf{P}_t\right].$$

If the population at time $t + \frac{1}{2N}$ is in HWLE (H1), then this last term is zero, which yields (4.39).

4.D The polygenic equation from the gene's eye-view

We now assume that the system is at statistical equilibrium, writing $\mathbb{P}^*$ for the corresponding probability. We use mean-field approximations to obtain the system (4.28-4.30) for $(P_t, \Delta^*, \varepsilon_t)$ where $\varepsilon_t := \Delta_t - \Delta^*$.

This section is structured as follows

- In Section 4.D.1, we apply the mean-field hypothesis (H3) to $(\Delta_t)_{t\geq 0}$.
- In Section 4.D.2, we discuss the decoupling of $(\Delta_t)_{t\geq 0}$ and $\vec{P}_t$ under (H4) and (H4'), which lets us obtain (4.28) by replacing Δ_t in (4.26) with its mean value Δ^*.
- In Section 4.D.3, we recover (4.30) for $(\varepsilon_t)_{t\geq 0}$.

4.D.1 Dynamics of the trait mean

Here we will derive the following equation for Δ_t

$$\begin{aligned}\mathrm{d}\Delta_t = \frac{1}{\tau}\Bigg(&- \Delta_t \times \frac{1}{|\bar{\theta}|} \times \mathbb{E}^*\left[2(L\alpha)^2 P_t(1-P_t)\right] \ + \ \frac{1}{|\bar{\theta}|} \times \mathbb{E}^*\left[2L^2\alpha^3\left(P_t - \frac{1}{2}\right)P_t(1-P_t)\right] \\ &+ \frac{L\omega_e^2}{|\bar{\theta}|} \times \mathbb{E}^*\left[2L\alpha\left(\theta^+(1-P_t) - \theta^- P_t\right)\right]\Bigg)\mathrm{d}t \ + \ \frac{1}{\sqrt{\tau}} \times \sqrt{\frac{\omega_e^2}{|\bar{\theta}|}\mathbb{E}^*\left[(2L\alpha)^2 P_t(1-P_t)\right]}\mathrm{d}B_t^{\Delta}\end{aligned} \tag{4.42}$$

The derivation is as follows. From (4.25), the dynamics of Δ_t are given by

$$\begin{aligned}\mathrm{d}\Delta_t &= \sum_{\ell=1,\dots,L} \alpha_\ell 2\mathrm{d}P_t^\ell \\ &= 2\sum_{\ell\in[L]} \alpha_\ell \Big(\xi_{\Delta_t,\alpha_\ell}(P_t^\ell)P_t^\ell(1-P_t^\ell)P_t^\ell(1-P_t^\ell)\mathrm{d}t + (\theta_\ell^+(1-P_t^\ell) - \theta_\ell^- P_t^\ell)\mathrm{d}t \\ &\qquad\qquad + \sqrt{P_t^\ell(1-P_t^\ell)}\mathrm{d}B_t^\ell\Big)\end{aligned}$$

We find from (4.27)

$$\begin{aligned}\mathrm{d}\Delta_t = 2\sum_{\ell\in[L]} -\Delta_t\frac{\alpha_\ell^2}{\omega_e^2}P_t^\ell(1-P_t^\ell)\mathrm{d}t \ &+ \ 2\sum_{\ell\in[L]}\frac{\alpha_\ell^3}{\omega_e^2}\left(P_t^\ell - \frac{1}{2}\right)P_t^\ell(1-P_t^\ell)\mathrm{d}t \\ + 2\sum_{\ell\in[L]}\alpha_\ell\left(\theta_\ell^+(1-P_t^\ell) - \theta_\ell^- P_t^\ell\right)\mathrm{d}t \ &+ \ 2\sum_{\ell\in[L]}\alpha_\ell\sqrt{P_t^\ell(1-P_t^\ell)}\mathrm{d}B_t^\ell\end{aligned}$$

We now express this as a function of τ from (4.32)

$$\begin{aligned}\mathrm{d}\Delta_t = \frac{1}{\tau}\Bigg(&\frac{2L\omega_e^2}{|\bar{\theta}|}\sum_{\ell\in[L]} -\Delta_t\frac{\alpha_\ell^2}{\omega_e^2}P_t^\ell(1-P_t^\ell) \ + \ \frac{2L\omega_e^2}{|\bar{\theta}|}\sum_{\ell\in[L]}\frac{\alpha_\ell^3}{\omega_e^2}\left(P_t^\ell - \frac{1}{2}\right)P_t^\ell(1-P_t^\ell) \\ &+ \frac{2L\omega_e^2}{|\bar{\theta}|}\sum_{\ell\in[L]}\alpha_\ell\left(\theta_\ell^+(1-P_t^\ell) - \theta_\ell^- P_t^\ell\right)\Bigg)\mathrm{d}t \ + \ \frac{1}{\sqrt{\tau}}\sum_{\ell\in[L]}2\alpha_\ell\sqrt{\frac{L\omega_e^2}{|\bar{\theta}|}P_t^\ell(1-P_t^\ell)}\mathrm{d}B_t^\ell\end{aligned}$$

which we rewrite

$$\begin{aligned}\mathrm{d}\Delta_t = \frac{1}{\tau}\Bigg(&-\frac{2}{|\bar{\theta}|}\Delta_t \times \frac{1}{L}\sum_{\ell\in[L]}(L\alpha_\ell)^2 P_t^\ell(1-P_t^\ell) \ + \ \frac{2}{|\bar{\theta}|} \times \frac{1}{L}\sum_{\ell\in[L]}L^2\alpha_\ell^3\left(P_t^\ell - \frac{1}{2}\right)P_t^\ell(1-P_t^\ell) \\ &+ \frac{L\omega_e^2}{|\bar{\theta}|} \times \frac{1}{L}\sum_{\ell\in[L]}2L\alpha_\ell\left(\theta_\ell^+(1-P_t^\ell) - \theta_\ell^- P_t^\ell\right)\Bigg)\mathrm{d}t \ + \ \frac{1}{\sqrt{\tau L}}\sum_{\ell\in[L]}2L\alpha_\ell\sqrt{\frac{\omega_e^2}{|\bar{\theta}|}P_t^\ell(1-P_t^\ell)}\mathrm{d}B_t^\ell.\end{aligned} \tag{4.43}$$

We now use (H3) on the first, second and third term

$$\frac{1}{L}\sum_{\ell\in[L]}(L\alpha_\ell)^2 P_t^\ell(1-P_t^\ell) \simeq \mathbb{E}^*\left[(L\alpha)^2 P_t(1-P_t)\right] \tag{4.44}$$

$$\frac{1}{L}\sum_{\ell\in[L]}L^2\alpha_\ell^3\left(P_t^\ell-\frac{1}{2}\right)P_t^\ell(1-P_t^\ell) \simeq \mathbb{E}^*\left[L^2\alpha^3\left(P_t-\frac{1}{2}\right)P_t(1-P_t)\right] \tag{4.45}$$

$$\frac{1}{L}\sum_{\ell\in[L]}2L\alpha_\ell\left(\theta_\ell^+(1-P_t^\ell)-\theta_\ell^-P_t^\ell\right) \simeq \mathbb{E}^*\left[2L\alpha\left(\theta^+(1-P_t)-\theta^-P_t\right)\right] \tag{4.46}$$

The last term of (4.43) is a Brownian term with quadratic variation

$$\frac{1}{\tau L}\sum_{\ell\in[L]}(2L\alpha_\ell)^2\times\frac{\omega_e^2}{|\bar{\theta}|}P_t^\ell(1-P_t^\ell) \simeq \frac{1}{\tau}\times\frac{\omega_e^2}{|\bar{\theta}|}\mathbb{E}^*\left[(2L\alpha)^2 P_t(1-P_t)\right] \tag{4.47}$$

where we used again (H3). We thus obtain (4.42) with B^Δ a Brownian motion such that for any $\ell\in[L]$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\langle\Delta,P^\ell\right\rangle_t = 2\alpha\frac{\mathrm{d}}{\mathrm{d}t}\left\langle P^\ell\right\rangle_t = 2\alpha\times P_t^\ell(1-P_t^\ell)\sim\frac{1}{L} \tag{4.48}$$

where $\langle\cdot,\cdot\rangle$ denotes quadratic variation.

4.D.2 Decoupling of the trait and the typical locus

We obtain the polygenic system (4.28-4.30) from the SDE for $(P_t^\ell)_{t\geq 0}$ in (4.26) and that for $(\Delta_t)_{t\geq 0}$ (4.42), by showing that we may replace $\xi_{\Delta_t,\alpha_\ell}$ with $\xi_{\Delta^*,\alpha_\ell}$ in the SDE (4.26) for P_t^ℓ.

Crucially, our derivation assumes either (H4) or (H4').

Derivation under (H4). At stationarity, equation (4.42) is an Ornstein-Uhlenbeck equation with fluctuations of order ω_e. It follows that

$$\xi_{\Delta_t,\alpha_\ell} = \xi_{\Delta^*,\alpha_\ell} + O(1/(L\omega_e))$$

In particular, because $\omega_e^{-2}\ll L^2$, we get that $\xi_{\Delta_t,\alpha_\ell}-\xi_{\Delta^*,\alpha_\ell}=o(1)$. ☐

Derivation under (H4'). We invoke the principle of **time-averaging** which we illustrate in Fig. 4.7.

This crucially relies on the fact that $(\Delta_t)_{t\geq 0}$ evolves on a timescale of $\tau\ll 1$. This can be seen as follows

- from (4.42) we know that $(\Delta_t)_{t\geq 0}$ is an Ornstein-Uhlenbeck process with autocorrelation ρ, and in particular it evolves on a characteristic timescale of $\frac{1}{\rho}$
- (H4'b) and (4.31) implies $\rho\sim\frac{1}{\tau}$
- (H4'c) implies $\tau\ll 1$.

Let us now recall the principle of time-averaging. Suppose we know (P_0,Δ_0), where $\Delta_0=\Delta^*+O(\omega_e)$. Let us consider what the first-order increments of P_t are, where t is chosen such that $\tau\ll t\ll 1$. The SDE for P_t can be written from (4.26) as

$$P_t = P_0+\int_0^t\left(\xi_{\Delta_u,\alpha}(P_u)P_u(1-P_u)\;+\;\left(\theta^+(1-P_u)-\theta^-P_u\right)\right)\mathrm{d}u+\int_0^t\sqrt{P_u(1-P_u)}\,\mathrm{d}B_u^P$$

for some Brownian motion B^P. We write from (4.27)

$$\xi_{\Delta_u,\alpha} = -\alpha\frac{\Delta_u-\Delta^*}{\omega_e^2}+\xi_{\Delta^*,\alpha}.$$

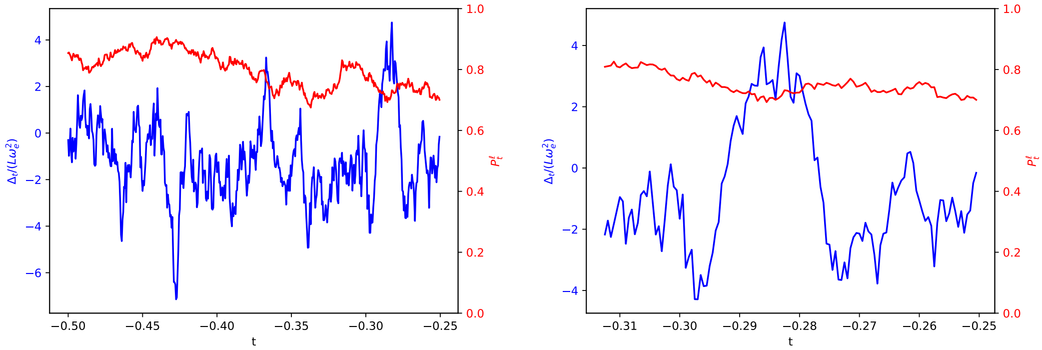

Figure 4.7: The slow/fast principle applied in strong selection. The population was evolved with $\omega_e^{-2} = L^2$ (strong selection) and $L = 1000, N = 1000, \eta = 1.2, \theta = (0.1, 0.2)$, and α_ℓ has law *Exponential*(L). In the left figure, we see that $(\Delta_t)_{t\geq 0}$ evolves with very short excursions away from its mean value before returning there. Meanwhile, P_t^ℓ only explores the segment $[0.6, 0.9]$, a small portion of the state space of P_t. On the right-hand side, we zoom in on a time window in which P_t^ℓ stays effectively constant, whereas Δ_t evolves very quickly. In particular, the fluctuations of Δ_t do not impact P_t. The time-averaging principle is then to consider a small time interval dt with $\tau \ll dt \ll 1$, to replace Δ_t in the equation of P_t^ℓ with its time average over $[t, t+dt]$, and consider that the law of P_t^ℓ does not evolve on $[t, t+dt]$.

Therefore, to replace $\xi_{\Delta_u,\alpha}$ by $\xi_{\Delta^*,\alpha}$ in the equation for P_t, we must show

$$\left| \int_0^t -\alpha \frac{\Delta_u - \Delta^*}{\omega_e^2} P_u(1 - P_u)\mathrm{d}u \right| \ll t$$

From (H4'd), on the interval $[0, t]$ the typical locus P_t scarcely evolves, and we may therefore suppose for $u \in [0, t]$

$$P_u \simeq P_0.$$

Therefore, we must show

$$\left| \int_0^t -\alpha \frac{\Delta_u - \Delta^*}{\omega_e^2} \mathrm{d}u \right| \ll t$$

Because the fluctuations of $(\Delta_u)_{u\leq t}$ are of order ω_e and since (H4'a) implies $\frac{\alpha}{\omega_e} \sim 1$, we only have to show that

$$\frac{1}{t}\int_0^t \Delta_u \mathrm{d}u \simeq \Delta^*$$

This ergodic theorem (see for instance [202]) crucially hinges on (H4'b), which tells us that the characteristic timescale for the evolution of $(\Delta_u)_{u\geq 0}$ is τ.

As we have done throughout this appendix, we defer to future work a rigorous proof of this time-averaging, but we do note that a result close to the one needed here has already been obtained by [169]. A rigorous proof in our system will face difficulties which are not tackled in [169], in particular the fact that B^Δ and B^P are not independent (see (4.48)).

□

4.D.3 Recovering the Ornstein-Uhlenbeck process

Here we recover the Ornstein-Uhlenbeck SDE (4.30) for $(\varepsilon_t)_{t\geq 0}$ from (4.42). If we define $\varepsilon_t := \Delta_t - \Delta^*$, we get from (4.42)

$$\begin{aligned}
\mathrm{d}\varepsilon_t = -\varepsilon_t \times \frac{1}{\tau|\bar{\theta}|} \times \mathbb{E}^*\left[2(L\alpha)^2 P_t(1-P_t)\right] \quad + \quad \frac{1}{\sqrt{\tau}} \times \sqrt{\frac{\omega_e^2}{|\bar{\theta}|}\mathbb{E}^*\left[(2L\alpha)^2 P_t(1-P_t)\right]}\mathrm{d}B_t^{\Delta} \\
+ \frac{1}{\tau}\left(\frac{1}{|\bar{\theta}|} \times \mathbb{E}^*\left[2L^2\alpha^3\left(P_t - \frac{1}{2}\right)P_t(1-P_t)\right] + \frac{L\omega_e^2}{|\bar{\theta}|} \times \mathbb{E}^*\left[2L\alpha\left(\theta^+(1-P_t) - \theta^- P_t\right)\right]\right. \\
\left. - \Delta^* \times \frac{1}{|\bar{\theta}|} \times \mathbb{E}^*\left[2(L\alpha)^2 P_t(1-P_t)\right]\right)\mathrm{d}t
\end{aligned}$$

It remains to show

$$\Delta^* = \frac{\mathbb{E}^*\left[2L^2\alpha^3\left(P_t - \frac{1}{2}\right)P_t(1-P_t)\right] + L\omega_e^2\mathbb{E}^*\left[2L\alpha\left(\theta^+(1-P_t) - \theta^- P_t\right)\right]}{\mathbb{E}^*\left[2(L\alpha)^2 P_t(1-P_t)\right]}. \tag{4.49}$$

where

$$\Delta^* = 2L\mathbb{E}^*[\alpha P_t] - \eta.$$

We start by noticing that because the system is at stationarity, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}^*[P_t \mid (\alpha,\theta)] = 0.$$

Applying (4.28) to $\mathbb{E}^*[P_t \mid (\alpha,\theta)]$, we find

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}^*[P_t \mid (\alpha,\theta)] = \mathbb{E}^*\left[\xi_{\Delta^*,\alpha}(P_t)P_t(1-P_t) + \theta^+(1-P_t) - \theta^- P_t \;\middle|\; (\alpha,\theta)\right] = 0.$$

From the definition of $\xi_{\Delta^*,\alpha}$ in (4.27) we find

$$\mathbb{E}^*\left[\left(-\frac{\alpha}{\omega_e^2}\Delta^* + \frac{\alpha^2}{\omega_e^2}\left(P_t - \frac{1}{2}\right)\right)P_t(1-P_t) + \left(\theta^+(1-P_t) - \theta^- P_t\right) \;\middle|\; (\alpha,\theta)\right] = 0.$$

We rewrite this, multiplying by $2L^2\alpha\omega_e^2$

$$\begin{aligned}
&\Delta^*\mathbb{E}^*\left[2(L\alpha)^2 P_t(1-P_t) \mid (\alpha,\theta)\right] \\
&\qquad = \mathbb{E}^*\left[2L^2\alpha^3\left(P_t - \frac{1}{2}\right)P_t(1-P_t) + L\omega_e^2 \times 2L\alpha\left(\theta^+(1-P_t) - \theta^- P_t\right) \;\middle|\; (\alpha,\theta)\right]
\end{aligned}$$

Taking the expectation with respect to (α,θ), this yields (4.49).

4.E Observables and scalings

Here, we consider that the polygenic limit holds, that is, that the system (4.28-4.29) holds for (P_t, Δ^*). We do not assume that (4.30) for $(\varepsilon_t)_{t\geq 0}$ holds unless stated otherwise.

In Section 4.E.1, we show how the macroscopic observables of the system can be computed from the stationary distribution of $\vec{P}_t$. In section 4.E.2 we briefly discuss the ultra-weak selection regime ($\omega_e^{-2} \ll L$). In section 4.E.3, we discuss how macroscopic observables scale under (A1-6). In section 4.E.4, we introduce the bias-correcting selection coefficient s^* and discuss its behavior in moderate selection ($L \ll \omega_e^{-2} \ll L^2$).

4.E.1 Observables at stationarity

With the fixed point equation (4.35), we showed how to compute $\Delta^* = \mathbb{E}[\Delta_t]$ and the distribution of the typical locus $\vec{P_t}$. In this subsection we detail how to compute the theoretical predictions for other observables. We assume (4.28-4.29) and HWLE (H1) hold.

We define the following observables at stationarity

$$\sigma_t^2 := \mathbf{Var}_{\mathbf{P}_t}[Z(g)] \qquad \bar{W}_t := \mathbf{E}_{\mathbf{P}_t}[W(Z(g))] \qquad V_t := \mathbf{Var}_{\mathbf{P}_t}[W(Z(g))] \tag{4.50}$$

$$\nu^2 := \mathbb{V}\mathbf{ar}[\Delta_t] \qquad \rho_u := -\ln\left(\frac{\mathbb{C}\mathbf{ov}\,[\Delta_t, \Delta_{t+u}]}{\mathbb{V}\mathbf{ar}[\Delta_t]}\right) \tag{4.51}$$

where $t, u \geq 0$. They are respectively the trait variance, the fitness load, the fitness variance, the fluctuations of the trait mean and its log-autocorrelation function.

We argue

$$\sigma_t^2 \simeq 2L\mathbb{E}^*[\alpha^2 P_t(1-P_t)] \tag{4.52}$$

$$\bar{W}_t = -\frac{1}{2\omega_e^2}\left(\sigma_t^2 + \Delta_t^2\right) \tag{4.53}$$

$$V_t \simeq \frac{\sigma_t^2(\sigma_t^2 + 2\Delta_t^2)}{2\omega^4} \tag{4.54}$$

Furthermore if (4.30) holds then

$$\nu = \omega_e \qquad \forall u \geq 0, \qquad \rho_u = u\rho \tag{4.55}$$

where ρ was defined in (4.31).

Remark 12. *Because $\mathbb{V}\mathbf{ar}^*[\sigma_t^2] \ll \mathbb{E}^*[\sigma_t^2]^2$, we will consider σ_t^2 as a constant in the rest of this work rather than a fluctuating quantity, writing $\sigma^2 = \sigma_t^2$.*

Let us start by computing a first-order approximation for the genetic variance σ_t^2. We have

$$\begin{aligned}
\sigma_t^2 =& \mathbf{Var}_{\mathbf{P}_t}\left[\sum_{\ell\in[L]} \alpha_\ell g_\ell\right] \\
\simeq& \sum_{\ell\in[L]} 2\alpha_\ell^2 P_t^\ell(1-P_t^\ell) \\
=& L \times \frac{1}{L}\sum_{\ell\in[L]} 2\alpha_\ell^2 P_t^\ell(1-P_t^\ell)
\end{aligned}$$

where we used the HWLE hypothesis (H1) to neglect cross-correlations. Using a mean-field approximation (H3) we get (4.52).

We turn to $\bar{W}_t$

$$\begin{aligned}
\bar{W}_t =& -\frac{1}{2\omega_e^2}\mathbf{E}_{\mathbf{P}_t}[(Z(g)-\eta)^2] \\
=& -\frac{1}{2\omega_e^2}\left(\mathbf{Var}_{\mathbf{P}_t}[Z(g)] + \mathbf{E}_{\mathbf{P}_t}[Z(g)-\eta]^2\right)
\end{aligned}$$

We thus get (4.53).

Third, we compute V_t

$$\begin{aligned}
V_t =& \frac{1}{4\omega^4}\mathbf{Var}_{\mathbf{P}_t}\left[(Z(g)-\eta)^2\right] \\
=& \frac{1}{4\omega^4}\mathbf{Var}_{\mathbf{P}_t}\left[(Z(g)-\bar{z}_t)^2 - 2(Z(g)-\bar{z}_t)\Delta_t\right] \\
=& \frac{1}{4\omega^4}\left(\mathbf{Var}_{\mathbf{P}_t}\left[(Z(g)-\bar{z}_t)^2\right] - 4\mathbf{Cov}_{\mathbf{P}_t}\left[(Z(g)-\bar{z}_t)^2, Z(g)-\bar{z}_t\right]\Delta_t + 4\sigma_t^2\Delta_t^2\right)
\end{aligned}$$

where in the second line we used $\bar{z}_t - \eta = \Delta_t$ and in the third line $\mathbf{Var}_{\mathbf{P}_t}[Z(g)] = \sigma_t^2$. We simplify the computations of V_t by approximating the law of $Z(g)$ under $\mathbf{E}_{\mathbf{P}_t}$ with a $\mathcal{N}(\bar{z}_t, \sigma_t^2)$ distribution (this is a consequence of the Central Limit Theorem under HWLE (H1), provided there is sufficient genetic variability). Under this approximation, the covariance between $(Z(g) - \bar{z}_t)^2$ and $Z(g) - \bar{z}_t$ is zero and

$$\mathbf{Var}_{\mathbf{P}_t}[(Z(g) - \bar{z}_t)^2] \simeq \mathbf{E}_{\mathbf{P}_t}[(Z(g) - \bar{z}_t)^4] - \sigma_t^4 \simeq 2\sigma_t^4$$

We thus obtain

$$V_t \simeq \frac{1}{4\omega^4}\left(2\sigma_t^4 + 4\sigma_t^2\Delta_t^2\right).$$

This yields the result.

The last two equalities are obtained from (4.30). When $\omega_e^{-2} \gg L$, (4.30) implies that ε_t is an Ornstein-Uhlenbeck. Furthermore, standard properties of Ornstein-Uhlenbeck processes, yield that Δ_t has autocorrelation structure

$$\forall t_1 < t_2, \qquad \mathbb{C}\mathbf{ov}[\varepsilon_{t_1}, \varepsilon_{t_2}] = \omega_e^2 e^{-\rho(t_2 - t_1)}$$

This yields (4.55).

4.E.2 Ultra-weak selection regime

Let us briefly discuss the case of ultra-weak-selection regime ($\omega_e^{-2} \ll L$).

Because of the definition of Δ_t and (4.22), we necessarily have $|\Delta_t| \leq 2$ as long as $\eta \in [0, 2]$. It follows from the definition of ξ (4.27) that when $\omega_e^{-2} \ll L$, $|\xi_{\Delta_t,\alpha}| \ll 1$, and therefore P_t^ℓ evolves as a neutral Wright-Fisher diffusion. In this regime, an individual locus is not affected by selection in a detectable way. In particular, the macroscopic observables Δ^*, σ can be computed from the neutral distribution $\Pi_{0,0,\theta}$.

4.E.3 Scaling of observables

We now assume (A2-5) hold. We claim

$$|\Delta^*| \sim L\omega_e^2 \qquad\qquad \sigma \sim \sqrt{\frac{|\bar{\theta}|}{L}} \tag{4.56}$$

If furthermore (4.30) holds for $(\varepsilon_t)_{t\geq 0}$ then

$$\rho \sim \frac{1}{\tau} \sim |\bar{\theta}|\omega_e^2 \quad \mathbb{E}^*[\bar{W}_t] \sim -\frac{1}{2N}\left(\frac{|\bar{\theta}|}{L\omega_e^2} + L^2\omega_e^2\right) \quad \mathbb{E}^*[V_t] \sim \frac{|\bar{\theta}|}{(2N)^2}\left(\frac{|\bar{\theta}|}{(L\omega_e^2)^2} + L\right). \tag{4.57}$$

Let us detail the computations.

Magnitude of Δ^**.* We must show $|\Delta^*| \sim L\omega_e^2$. Specifically, we will show that if Δ^* satisfies the fixed-point equation (4.35), then $x^* := \frac{\Delta^*}{L\omega_e^2}$ satisfies a non-degenerate equation and in particular has order 1. Without loss of generality, we assume $2I(0) - \eta > 0$, so that in particular $\Delta^* > 0$.

We rewrite (4.35) as

$$\Delta^* = 2\mathbb{E}^*\left[(L\alpha)\int p\Pi_{\Delta^*,\alpha,\theta}(p)\mathrm{d}p\right] - \eta.$$

From the definition of $\Pi_{\delta,a,\theta}$ in (4.34) we have

$$\Pi_{\Delta^*,\alpha,\theta}(p) = \frac{\Pi_{0,\alpha,\theta}(p)}{\int \Pi_{0,\alpha,\theta}(p')e^{-2\frac{\Delta^*}{\omega_e^2}\alpha p'}\mathrm{d}p'} e^{-2\frac{\Delta^*}{\omega_e^2}\alpha p}$$

In particular, the fixed point equation becomes

$$\Delta^* = 2\mathbb{E}^*\left[(L\alpha)\frac{\int p\Pi_{0,\alpha,\theta}(p)e^{-2\frac{\Delta^*}{\omega_e^2}\alpha p}\mathrm{d}p}{\int \Pi_{0,\alpha,\theta}(p)e^{-2\frac{\Delta^*}{\omega_e^2}\alpha p}\mathrm{d}p}\right] - \eta.$$

Rewriting this in terms of x^*, we get

$$x^* L\omega_e^2 = 2\mathbb{E}^*\left[(L\alpha)\frac{\int p\Pi_{0,\alpha,\theta}(p)e^{-2(L\alpha)x^*p}\mathrm{d}p}{\int \Pi_{0,\alpha,\theta}(p)e^{-2(L\alpha)x^*p}\mathrm{d}p}\right] - \eta. \tag{4.58}$$

Let us show that x^* necessarily satisfies $x^* \sim 1$, by showing that we neither have $x^* \ll 1$ nor $x^* \gg 1$.

If x^* satisfied $x^* \ll 1$, from (A5) we have $x^* L\omega_e^2 \ll 1$ and therefore (4.58) yields

$$|2I(0) - \eta| \ll 1$$

which is impossible from (A6). Let us now show that we cannot have $x^* \gg 1$. It can be checked that the mass of $\Pi_{0,\alpha,\theta}$ between 0 and $\varepsilon < 1/2$ is of order $\frac{\theta^-}{\theta^+ + \theta^-}\varepsilon^{2\theta^+}$. For $\alpha \sim 1/L$ and θ such that $\theta^+ \sim \theta^- \lesssim 1$, it follows that

$$\frac{\Pi_{0,\alpha,\theta}(p)}{\int \Pi_{0,\alpha,\theta}(p')e^{-2(L\alpha)xp'}\mathrm{d}p'}e^{-2(L\alpha)xp}$$

is close to a Dirac mass on 0 whenever $x \gg 1$. In particular,

$$\frac{\int p\Pi_{0,\alpha,\theta}(p)e^{-2(L\alpha)xp}\mathrm{d}p}{\int \Pi_{0,\alpha,\theta}(p)e^{-2(L\alpha)xp}\mathrm{d}p} \ll 1$$

If the solution to the fixed-point equation (4.58) satisfied $x^* \gg 1$, then we would have

$$x^* L\omega_e^2 \simeq -\eta$$

The left-hand side is non-negative whereas the right-hand side is negative of order 1 from (A4), which yields a contradiction. ☐

Magnitude of σ. Recall from (4.27) and (4.34)

$$\begin{aligned}\xi_{\delta,a}(p) :=& -\frac{a\delta}{\omega_e^2} + \frac{a^2}{\omega_e^2}\left(p - \frac{1}{2}\right)\\ \Pi_{\delta,a,\theta}(p) :=& C_{\delta,a,\theta}\, p^{2\theta^+-1}(1-p)^{2\theta^--1}e^{2\int_0^p \xi_{\delta,a}(u)\,\mathrm{d}u}.\end{aligned}$$

with $C_{\delta,a,\theta}$ a normalization constant. Then if $\alpha \sim \frac{1}{L}$ we find from $\Delta^* \sim L\omega_e^2$ and (5) that $\xi_{\Delta^*,\alpha} \sim 1$ and therefore

$$\int p(1-p)\Pi_{\Delta^*\alpha,\theta}(p)\mathrm{d}p \sim \frac{\int p^{2\theta^+}(1-p)^{2\theta^-}\mathrm{d}p}{\int p^{2\theta^+-1}(1-p)^{2\theta^--1}\mathrm{d}p} = \frac{Beta(2\theta^+ + 1, 2\theta^- + 1)}{Beta(2\theta^+, 2\theta^-)}$$

where *Beta* is the Beta function. In particular

$$\int p(1-p)\Pi_{\Delta^*\alpha,\theta}(p)\mathrm{d}p \sim \frac{\theta^+\theta^-}{|\theta|}$$

where $|\theta| = \theta^+ + \theta^-$. As a consequence, if we consider a typical locus $\vec{P}_t$ we have

$$\mathbb{E}^*[(L\alpha)^2 P_t(1-P_t)] \sim \mathbb{E}\left[(L\alpha)^2\frac{\theta^+\theta^-}{|\theta|}\right] \sim |\bar{\theta}|$$

where we used (A3). We thus find

$$\sigma^2 = L\mathbb{E}^*[\alpha^2 P_t(1-P_t)] \sim \frac{|\bar{\theta}|}{L}.$$

□

Derivation of (4.57). We obtain from (4.31) and $\sigma \sim \sqrt{|\bar{\theta}|/L}$ that $\rho \sim \frac{1}{\tau} = \frac{|\bar{\theta}|}{L\omega_e^2}$ from the definition of τ in (4.32). Finally, (4.53) implies

$$\mathbb{E}^*[\bar{W}_t] \sim -\frac{1}{\omega^2}\left(\sigma^2 + (\Delta^*)^2 + \omega_e^2\right)$$

using $\mathbb{E}^*[\Delta_t^2] = (\Delta^*)^2 + \omega_e^2$ from (4.55). Using $\omega^2 = 2N\omega_e^2$ and the previous estimates for σ^2, Δ^* we find

$$\begin{aligned}\mathbb{E}^*[\bar{W}_t] &\sim -\frac{1}{2N\omega_e^2}\left(\frac{|\bar{\theta}|}{L} + (L\omega_e^2)^2 + \omega_e^2\right)\\ &\sim -\frac{1}{2N}\left(\frac{|\bar{\theta}|}{L\omega_e^2} + L^2\omega_e^2 + 1\right).\end{aligned}$$

The result follows from the fact that under (A5), we have $L^2\omega_e^2 \gtrsim 1$, and therefore the last term can always be absorbed into the second term. Similarly (4.54) yields

$$\begin{aligned}\mathbb{E}^*[V_t] &\sim \frac{\sigma^2}{\omega^4}\left(\sigma^2 + (\Delta^*)^2 + \omega_e^2\right)\\ &\sim \frac{|\bar{\theta}|}{(2N)^2\omega_e^4 L}\left(\frac{|\bar{\theta}|}{L} + (L\omega_e^2)^2 + \omega_e^2\right)\\ &\sim \frac{|\bar{\theta}|}{(2N)^2}\left(\frac{|\bar{\theta}|}{(L\omega_e^2)^2} + L + \frac{1}{L\omega_e^2}\right)\end{aligned}$$

and the last term can be absorbed in L. □

4.E.4 The bias-correcting coefficient

Define the bias-correcting coefficient s^* as

$$s^* := -\frac{\Delta^*}{L\omega_e^2} \tag{4.59}$$

From (4.56), we know that s^* is always of order 1. In this subsection we claim that for $L \ll \omega_e^{-2} \ll L^2$ (moderate selection), s^* is solution to

$$2L\mathbb{E}\left[\alpha\frac{\theta^+}{|\theta|} \times \frac{{}_1F_1(2\theta^+ + 1\ ;\ 2|\theta| + 1\ ;\ 2s^*L\alpha)}{{}_1F_1(2\theta^+\ ;\ 2|\theta|\ ;\ 2s^*L\alpha)}\right] \simeq \eta$$

where ${}_1F_1$ is the confluent hypergeometric function. In particular, s^* is independent of ω_e^{-2}.

Furthermore, if we have $|\bar{\theta}| \ll 1$ and $\omega_e^{-2} \gg L$, then s^* is a solution to

$$2L\mathbb{E}\left[\alpha\frac{\theta^+}{\theta^+ e^{2s^*L\alpha} + \theta^-}e^{2s^*L\alpha}\right] \simeq \eta.$$

As noticed in [122] (Eq. A.2), assuming $(\alpha_\ell, \theta_\ell)$ is constant across loci, this equation can be solved for s^* explicitely to find

$$2\alpha L s^* \simeq \ln\left(\frac{\theta^-}{\theta^+}\right) + \ln\left(\frac{\eta}{2\alpha L - \eta}\right)$$

The derivation is as follows.

In moderate selection, we find from (4.56) that $|\Delta^*| \sim L\omega_e^2 \ll 1 \sim \eta$. It follows that (4.35) can be rewritten

$$2I(\Delta^*) \simeq \eta \tag{4.60}$$

On the other hand, we find from (4.27) that for a typical α,

$$\xi_{\Delta^*,\alpha}(p) \simeq -s^* L\alpha \sim 1$$

where we neglected $\frac{\alpha^2}{\omega_e^2} \sim \frac{1}{L^2\omega_e^2} \ll 1$. It follows

$$\Pi_{\Delta^*,\alpha,\theta}(p) \simeq \frac{p^{2\theta^+-1}(1-p)^{2\theta^--1}e^{2s^*L\alpha p}}{\int y^{2\theta^+-1}(1-y)^{2\theta^--1}e^{2s^*L\alpha y}\mathrm{d}y}$$

From the definition of I in (4.36) we get

$$\begin{aligned} I(\Delta^*) \simeq& \mathbb{E}\left[L\alpha \frac{\int p^{2\theta^+}(1-p)^{2\theta^--1}e^{2s^*L\alpha p}\mathrm{d}p}{\int p^{2\theta^+-1}(1-p)^{2\theta^--1}e^{2s^*L\alpha p}\mathrm{d}p}\right] \\ =& \mathbb{E}\left[\alpha\frac{\theta^+}{|\theta|} \times \frac{{}_1F_1(2\theta^+ + 1\ ;\ 2|\theta|+1\ ;\ 2s^*L\alpha)}{{}_1F_1(2\theta^+\ ;\ 2|\theta|\ ;\ 2s^*L\alpha)}\right] \end{aligned}$$

where we recall the notation $|\theta| := \theta^+ + \theta^-$. Carrying this into (4.60), we get the first result.

Now assume $|\bar{\theta}| \ll 1$ and $\omega_e^{-2} \gg L$. It is well-known that the distribution of P_t at equilibrium is concentrated on $\{0, 1\}$. This means conditional on (α, θ), the stationary distribution $\Pi_{\Delta^*,\alpha,\theta}$ of P_t is close to a Bernoulli law with parameter

$$\frac{\theta^+}{\theta^+ e^{2s^*L\alpha} + \theta^-} e^{2s^*L\alpha}$$

Plugging this into (4.29) we get

$$\Delta^* \simeq 2L\mathbb{E}\left[\alpha\frac{\theta^+}{\theta^+ e^{2s^*L\alpha} + \theta^-} e^{2s^*L\alpha}\right] - \eta$$

From (4.56), we have that $\Delta^* \ll 1$ when $\omega_e^{-2} \gg L$. This yields the second result.

4.F Breakdown of the polygenic limit

Here, we discuss (H1-4') in light of Section 4.E. We assume (A1-6) hold and the polygenic equation for $(P_t, \Delta^*)_{t\geq 0}$ in (4.28-4.30) hold. In particular, the orders of magnitude derived in Section 4.E let us characterize the parameter values for which (A1-6) are consistent with (H1-4'). We argue that if (N1-2) is not satisfied, then (4.56) is incompatible with (H1-4'), and in particular that the polygenic system in (4.28-4.29) breaks down. In Section 4.F.5, we argue that if (N3) is not satisfied, then (4.30) cannot provide an accurate description of the fluctuations of $(\varepsilon_t)_{t\geq 0}$.

4.F.1 Discussion of (H2)

We illustrate our method by discussing (H2), which assumes that the fitness $F(Z(g))$ is very concentrated around its mean value under $\mathbf{E}_{\mathbf{P}_t}$. In particular, this requires the fitness variance V_t to be very small. In light of (4.57), this requires

$$\left(\frac{|\bar{\theta}|}{2NL\omega_e^2}\right)^2 \ll 1 \qquad\qquad \frac{|\bar{\theta}|L}{(2N)^2} \ll 1$$

which we rewrite

$$2N \gg \frac{|\bar{\theta}|}{L\omega_e^2} \qquad\qquad 2N \gg \sqrt{|\bar{\theta}|L}.$$

Both these conditions are satisfied under (N1).

Let us be more precise. In Section 4.C, (H2) was used to write

$$\mathbf{Cov}_{\mathbf{P}_t}[F(Z(g)), g_\ell] \simeq \mathbf{Cov}_{\mathbf{P}_t}[\hat{F}_t(W(Z(g)) - \bar{W}_t), g_\ell] \qquad \mathbf{E}_{\mathbf{P}_t}[F(Z(g))] \simeq \hat{F}_t$$

By Taylor expansion, this is justified if

$$\sum_{k>1} \frac{1}{k!} \left|\mathbf{Cov}_{\mathbf{P}_t}[(W(Z(g)) - \bar{W}_t)^k, g_\ell]\right| \\ \ll \left|\mathbf{Cov}_{\mathbf{P}_t}[(W(Z(g)) - \bar{W}_t), g_\ell]\right| \sum_{k\geq 2} \frac{1}{k!} \mathbf{E}_{\mathbf{P}_t}[|W(Z(g)) - \bar{W}_t|^k] \ll 1$$

A proper justification of (H2) would require a control of $\mathbf{Cov}_{\mathbf{P}_t}[(W(Z(g)) - \bar{W}_t)^k, g_\ell]$ for $k \geq 1$ and $\mathbf{E}_{\mathbf{P}_t}[(W(Z(g)) - \bar{W}_t)^k]$ for $k \geq 3$. Such estimates can be obtained if we assume that $Z(g)$ is normally distributed under $\mathbf{E}_{\mathbf{P}_t}$ conditioned on g_ℓ. The trait's eye-view, which will be presented in Section 4.G, yields an alternative derivation of the diffusion equation (4.26) which relies precisely on this assumption.

4.F.2 Discussion of HWLE (H1)

HWLE (H1) is undoubtedly the most delicate assumption above. We will discuss it by referring to the rich literature on the subject.

Hypothesis (H1) was discussed in Chapter 2. There, the recombination rate ρ was defined as the number of recombination events for a given lineage over $2N$ generations. In our setting, in which recombination occurs every generation, this corresponds to $\rho = 2N$. The effect of recombination is to force the population close to the Wright manifold, on which the population is at LE. Theorem 2.1.1 of Chapter 2 adapted to our setting is that if

$$2N \gg \omega_e^{-4} \ln(L)^2 \tag{4.61}$$

then we can effectively assume LE when computing the dynamics of P_t^ℓ. That is, we may neglect the effect of linked selection on P_t^ℓ. (4.61) is admittedly biologically unrealistic, and it is likely not optimal, but it is derived entirely *a priori*. In our simulations (Fig. 4.1), $2N = 1,000$, $L = 100$ and $\omega_e^{-4} \in [L^2, L^4]$, and the criterion (4.61) reads $1,000 \gg [2 \times 10^5, 2 \times 10^9]$. In particular, (4.61) is assuredly not optimal. We can try and get a better idea of the conditions for the breakdown of (H1) using the Quasi-Linkage Equilibrium approach from statistical genetics [101], which is presented in the global Appendix A.1 of this PhD. This approach assumes *a priori* that the first-order effect of LD on gene dynamics and macroscopic observables can be described exclusively with two-loci correlations (neglecting cumulants of order 3 or greater).

Bulmer effect

When selection is sufficiently strong, we have $\Delta^* \ll \sigma$. Bulmer suggested in [106] that for an additive trait under stabilizing selection, LD can be neglected as long as $\ln(L)\frac{\sigma_t^2}{\omega^2} \ll 1$ where the $\ln(L)$ factor comes from the choice of single-point uniform crossover as a recombination mechanism (see (10) of [106], rederived in (A.4) in Appendix A.1). This criterion is derived from the by considering that LD only appears through selection picking advantageous combinations of genes, which when the population is close to the optimum results in negative LD (for stabilizing selection). The criterion can be rewritten

$$2N \gg \sigma_t^2 \ln(L) \omega_e^{-2}. \tag{4.62}$$

In light of (4.56), this can be rewritten

$$2N \gg |\bar{\theta}| \frac{\ln(L)}{L\omega_e^2}$$

which is satisfied under (N1). If this equation is not satisfied, then we expect to enter the Quasi-Linkage Equilibrium regime [101], in which LD appears between loci. This would mainly manifest itself in the decrease of the genetic variance in the trait σ^2 relative to the theoretical prediction due to negative LD [105].

In Fig. 4.1 bottom left, we see this effect appear around $\omega_e^{-2} \sim 200$ and $\omega_e^{-2} \sim 2000$, respectively for $2N = 100$ and $2N = 1000$. Translating this in terms of (4.62), we get that the critical value of ω_e^{-2} at which LD ceases to be negligible satisfies

$$|\bar{\theta}| \frac{\ln(L)}{2NL\omega_e^2} \sim 0.03.$$

The Hill-Robertson effect

When selection is sufficiently weak that $\Delta^* \gg \sigma$, the population remains far from the optimum and can be described using a linear selection model, in which the logfitness of an organism is approximated by linear selection

$$\tilde{W}(z) := -\frac{\Delta^*}{\omega^2}(z - \bar{z})$$

where $\bar{z} = 2L\mathbb{E}\left[\alpha P_t\right] = \Delta^* + \eta$.

In such a setting, the appearance of negative LD for small population sizes is known as Hill-Robertson effect [203]. The rationale is as follows: LD leads the genetic variance σ_t^2 to oscillates randomly and quickly with respect to the genic variance $\sigma_{G,t}^2$ defined with

$$\sigma_{G,t}^2 := \sum_{\ell \in [L]} 2\alpha_\ell^2 P_t^\ell (2 - P_t^\ell).$$

We illustrate this oscillation in Fig. 4.8 in an extreme scenario in which $L = 1,000$ and $N = 20$, for weak stabilising selection and directional selection of equivalent magnitude. When LD is positive, then there is a lot of genetic variance, in which case selection will act efficiently to decrease the genetic variance (and in particular LD). This explains why positive LD only manifests itself in very short excursions. On the other hand, when LD is negative, then the genetic variance σ_t^2 is reduced with respect to the genic variance $\sigma_{G,t}^2$, and selection is inefficient. In this setting, only recombination will, on a longer timescale, bring LD back to zero.

We illustrate the Hill-Robertson effect and the corresponding breakdown of the polygenic limit of (4.28-4.29) in Fig. 4.9. We see that this breakdown is quite limited, since even with as few as $N = 10$ organisms with $L = 100$ loci each, LD only decreases the genetic variance by about 5%.

Using the theory of Quasi-Linkage Equilibrium, we can try and obtain a quantitative criterion for the appearance of the Hill-Robertson effect. In (44) of Section VI.B of [101] (see

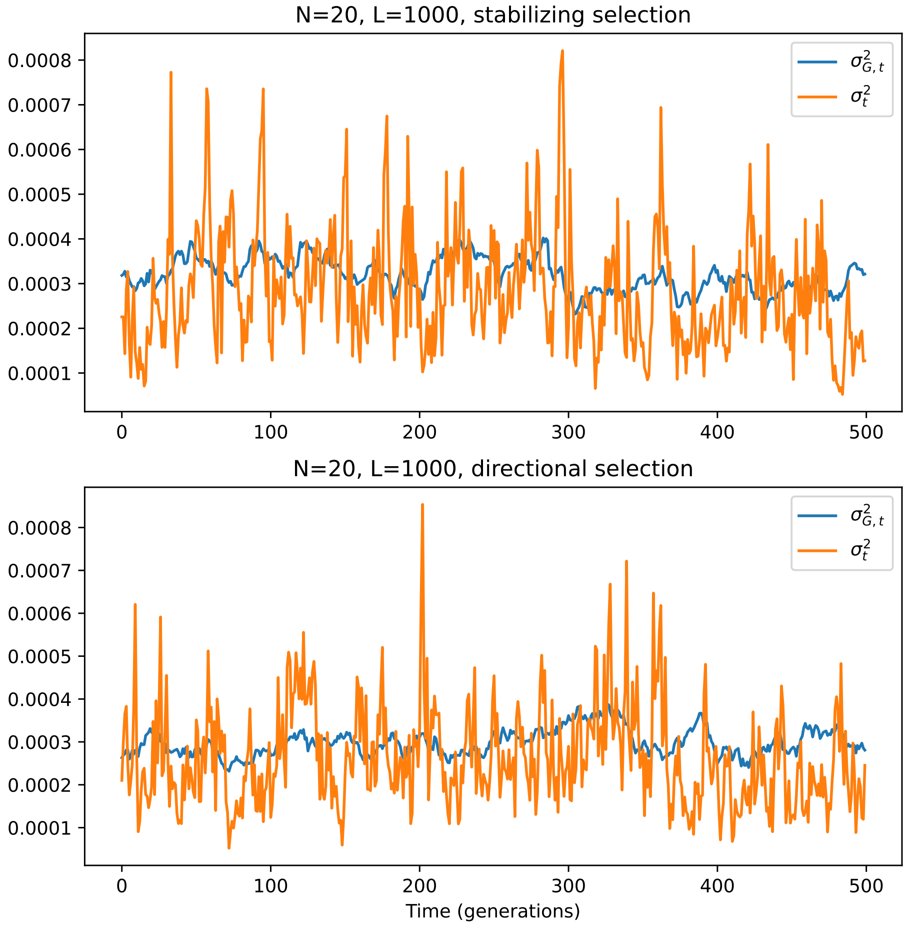

Figure 4.8: The Hill-Robertson effect. The population was evolved for $T = 500N$ generations either under stabilizing selection ($\eta = 1.2, \omega_e^{-2} = 2/L$) or on the corresponding directional selection regime. The mutation rate is $\theta = (0.1, 0.2)$ (weak selection) and the $(\alpha_\ell)_{\ell \in [L]}$ have law $Exponential(L)$. The **genetic trait variance** is $\sigma_t^2 := \mathbf{Var}_{\mathbf{X}_t}[Z(g)]$, whereas the **genic trait variance** $\sigma^2_{G,t}$ corresponds to the variance the population would have if it was in HWLE. When blue is on top of orange, the population displays negative LD, whereas when orange is on top of blue, the population displays positive LD.

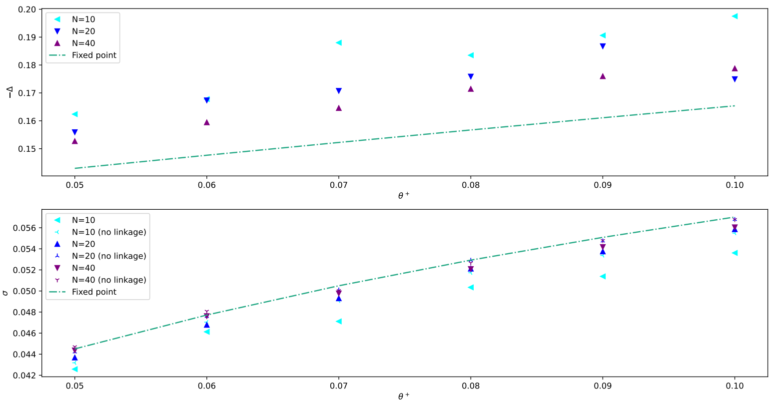

Figure 4.9: The Hill-Robertson effect: the population is evolves with $\omega_e^{-2} = 2/L$ (weak selection) with $L = 100, \theta^- = 2\theta^+$, $\eta = 1.2$, for a time of $T = 4,000N$ at various population sizes, half of which were used as burn-in. Below, the plot distinguishes between genetic variance (the variance in the trait) and the genic variance, corresponding to the label "no linkage". The genic variance is the variance of the trait if the allele frequencies were kept unchanged but linkage equilibrium was enforced. The difference between the genetic and the genic variance is a measure of LD.

(A.5) in Appendix A.1), it is claimed that the Hill-Robertson effect at a locus $\ell \in [L]$ can be neglected as soon as

$$\sum_{\ell' \neq \ell} P_t^{\ell'}(1 - P_t^{\ell'}) \left(\frac{1}{r_{\ell,\ell'}} \xi_{\Delta^*,\alpha_{\ell'}}(P_t^{\ell'}) \right)^2 \ll 1$$

where $r_{\ell_1 \ell_2}$ is the rate of recombination between ℓ_1 and ℓ_2, in our system $r_{\ell_1 \ell_2} = \frac{|\ell_1 - \ell_2|}{L} 2N$. This criterion was originally obtained in [204] assuming $|\bar{\theta}| \ll 1$.

Taking the expectation, and using from (4.59) that $\xi_{\Delta^*,\alpha_{\ell'}}$ is typically of order 1, we obtain the criterion

$$\sum_{\ell' \in [L] \setminus \{\ell\}} \mathbb{E}\left[P_t(1 - P_t)\right] \times \left(\frac{L}{|\ell - \ell'| 2N} \right)^2 \ll 1$$

which yields from (4.56)

$$|\bar{\theta}| \frac{L^2}{(2N)^2} \ll 1$$

which we rewrite

$$2N \gg L\sqrt{|\bar{\theta}|}.$$

This is satisfied under (N1).

4.F.3 Discussion of the mean-field hypothesis (H3)

We will now show that our mean field approximation breaks down when mutation rates are small. As is usual with mean-field approximations [157], let us consider that the error of the mean-field hypothesis (H3) has variance of order $\frac{1}{L}$, that is

$$\mathbb{V}\mathbf{ar}\left[\frac{1}{L} \sum_{\ell \in [L]} f(\vec{P}_t^{\ell}) \right] \sim \frac{\mathbb{V}\mathbf{ar}[f(\vec{P}_t)]}{L} \tag{4.63}$$

We consider the mean-field approximation (H3) to be valid as long as this error is much smaller than the mean of $f(\vec{P}_t)$

$$\sqrt{\frac{\mathbb{V}\mathbf{ar}[f(\vec{P}_t)]}{L}} \ll \mathbb{E}\left[f(\vec{P}_t)\right]$$

Let us now check the validity of the mean-field approximations of (4.44-4.47). From (4.56), we find

$$\begin{aligned}
\mathbb{E}^*\left[(L\alpha)^2 P_t(1-P_t)\right] &\sim |\bar{\theta}| \\
\mathbb{E}^*\left[L^2\alpha^3\left(\frac{1}{2}-P_t\right)P_t(1-P_t)\right] &\lesssim \frac{|\bar{\theta}|}{L} \\
\mathbb{E}^*\left[2L\alpha(\theta^+(1-P_t)-\theta^- P_t)\right] &\sim |\bar{\theta}|
\end{aligned}$$

In Section 4.F.3 we argue

$$\mathbb{V}\mathbf{ar}^*\left[(L\alpha)^2 P_t(1-P_t)\right] \sim |\bar{\theta}| \tag{4.64}$$

$$\mathbb{V}\mathbf{ar}^*\left[L^2\alpha^3\left(\frac{1}{2}-P_t\right)P_t(1-P_t)\right] \lesssim \frac{|\bar{\theta}|}{L^2} \tag{4.65}$$

$$\mathbb{V}\mathbf{ar}^*\left[2L\alpha(\theta^+(1-P_t)-\theta^- P_t)\right] \sim |\bar{\theta}|^2 \tag{4.66}$$

It follows that the mean-field approximations (4.44-4.47) are valid iff

$$\sqrt{\frac{|\bar{\theta}|}{L}} \ll |\bar{\theta}| \qquad \sqrt{\frac{|\bar{\theta}|}{L^2}} \ll \frac{|\bar{\theta}|}{L} \qquad \sqrt{\frac{|\bar{\theta}|^2}{L}} \ll |\bar{\theta}|$$

This is equivalent to (N2). In Fig. 4.10, we illustrate the breakdown of the polygenic limit if (N2) is not satisfied.

Computing the errors on the mean-field approximationss

We start with (4.64). Recall from (4.34) that conditional on (α,θ), P_t has distribution $\Pi_{\Delta^*,\alpha,\theta}$ which is equivalent to a $Beta(2\theta^+, 2\theta^-)$ distribution when $\Delta^* \sim L\omega_e^2$ and $\alpha \sim 1/L$. If $\hat{P}_t$ has law $Beta(2\theta^+, 2\theta^-)$ for some θ, standard properties of the Beta distribution yield

$$\begin{aligned}
\mathbb{E}\left[\hat{P}_t^2(1-\hat{P}_t)^2\right] &= \frac{2\theta^+(2\theta^++1)\theta^-(2\theta^-+1)}{(\theta^++\theta^-)(2\theta^++2\theta^-+1)(2\theta^++2\theta^-+2)} \sim |\theta| \\
\mathbb{E}\left[\hat{P}_t(1-\hat{P}_t)\right]^2 &= \left(\frac{2\theta^+\theta^-}{\theta^++\theta^-}\right)^2 \sim |\theta|^2 \\
\mathbb{V}\mathbf{ar}[\hat{P}_t(1-\hat{P}_t)] &= \mathbb{E}\left[\hat{P}_t^2(1-\hat{P}_t)^2\right] - \mathbb{E}\left[\hat{P}_t(1-\hat{P}_t)\right]^2 \sim |\theta|
\end{aligned}$$

if $\theta^+ \sim \theta^- \lesssim 1$. Similarly, conditional on (α,θ), we have

$$\mathbb{V}\mathbf{ar}^*[P_t(1-P_t) \mid (\alpha,\theta)] \sim |\theta|$$

Using the decomposition

$$\begin{aligned}
&\mathbb{V}\mathbf{ar}^*\left[(L\alpha)^2 P_t(1-P_t)\right] \\
&\qquad = \mathbb{E}^*\left[\mathbb{V}\mathbf{ar}^*\left[(L\alpha)^2 P_t(1-P_t)\,\middle|(\alpha,\theta)\right]\right] + \mathbb{V}\mathbf{ar}^*\left[\mathbb{E}^*\left[(L\alpha)^2 P_t(1-P_t)\,\middle|(\alpha,\theta)\right]\right]
\end{aligned}$$

we get under (A1-3)

$$\mathbb{V}\mathbf{ar}^*\left[(L\alpha)^2 P_t(1-P_t)\right] \sim |\bar{\theta}|.$$

as requested.

(4.65) is obtained similarly. Let us turn to (4.66). As above, if $\hat{P}_t$ has law $Beta(2\theta^+, 2\theta^-)$ for some θ with $\theta^+ \sim \theta^- \lesssim 1$ then

$$\mathbb{V}\mathbf{ar}[\hat{P}_t] = \frac{\theta^+\theta^-}{|\bar{\theta}|^2(2|\bar{\theta}|+1)} \sim 1.$$

Consider $\tilde{P}_t$ with law $\Pi_{\Delta^*,\alpha,\theta}$ with $\alpha \sim 1/L$. Let us show that we also have $\mathbb{V}\mathbf{ar}[\tilde{P}_t] \sim 1$. Because $\Pi_{\Delta^*,\alpha,\theta}$ is equivalent to a $Beta(2\theta^+, 2\theta^-)$ distribution, we have

$$\begin{aligned}\mathbb{V}\mathbf{ar}[\tilde{P}_t] =& \mathbb{E}\left[\left(\tilde{P}_t - \mathbb{E}\left[\tilde{P}_t\right]\right)^2\right]\\ \sim& \mathbb{E}\left[\left(\hat{P}_t - \mathbb{E}\left[\tilde{P}_t\right]\right)^2\right]\\ =& \mathbb{V}\mathbf{ar}[\hat{P}_t] + \left(\mathbb{E}\left[\hat{P}_t\right] - \mathbb{E}\left[\tilde{P}_t\right]\right)^2\\ \sim& 1\end{aligned}$$

From this, we may conclude that conditional on (α, θ) we have

$$\mathbb{V}\mathbf{ar}^*\left[P_t \mid (\alpha,\theta)\right] \sim 1$$

Writing the decomposition

$$\begin{aligned}&\mathbb{V}\mathbf{ar}^*\left[2\alpha(\theta^+(1-P_t) - \theta^- P_t)\right]\\ &= \mathbb{E}^*\left[\mathbb{V}\mathbf{ar}^*\left[2\alpha(\theta^+(1-P_t) - \theta^- P_t)\ \big|(\alpha,\theta)\right]\right] + \mathbb{V}\mathbf{ar}^*\left[2\mathbb{E}^*\left[\alpha(\theta^+(1-P_t) - \theta^- P_t)\ \big|(\alpha,\theta)\right]\right]\end{aligned}$$

we get under (A1-3)

$$\mathbb{V}\mathbf{ar}^*\left[2L\alpha(\theta^+(1-P_t) - \theta^- P_t)\right] \sim |\bar{\theta}|^2$$

as requested.

Breakdown of the mean-field hypothesis under small mutation rates

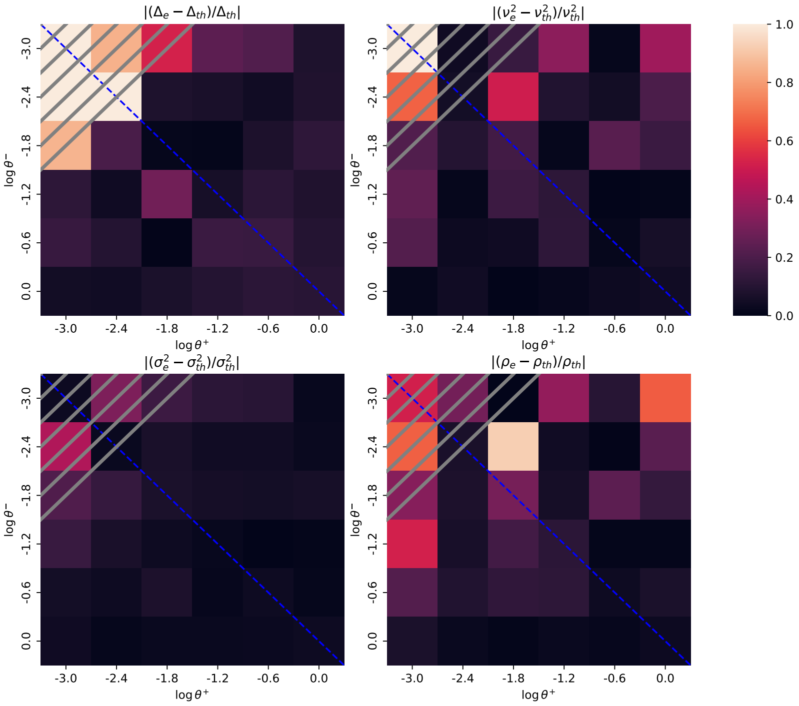

Figure 4.10: Breakdown of our approximation when hypotheses (N2) and (A3) fail. The population was evolved with $\omega_e^{-2} = 10^3$ (strong selection), $N = 500, L = 100, \eta = 1.2$ under different mutation rates for $T = 500N$ generations, including a burn-in of $250N$ generations. We took the same $(\alpha_\ell)_{\ell \in [L]}$ as in Fig. 4.1. We plot the empirical value of Δ from simulations (Δ_e) vs our theoretical prediction (Δ_{th}), and similarly for ν, σ, ρ. As in Fig. 4.1, the predictions were obtained conditional on the values of $(\alpha_\ell)_{\ell \in [L]}$. In this parameterization, assumption (N2) reads $\theta^+ + \theta^- \gg 10^{-2}$ (non-hashed zone) and (A3) reads $\theta^+ \sim \theta^-$ (blue dashed line).

Breakdown of the mean-field hypothesis when the allele effects have heavy tails.

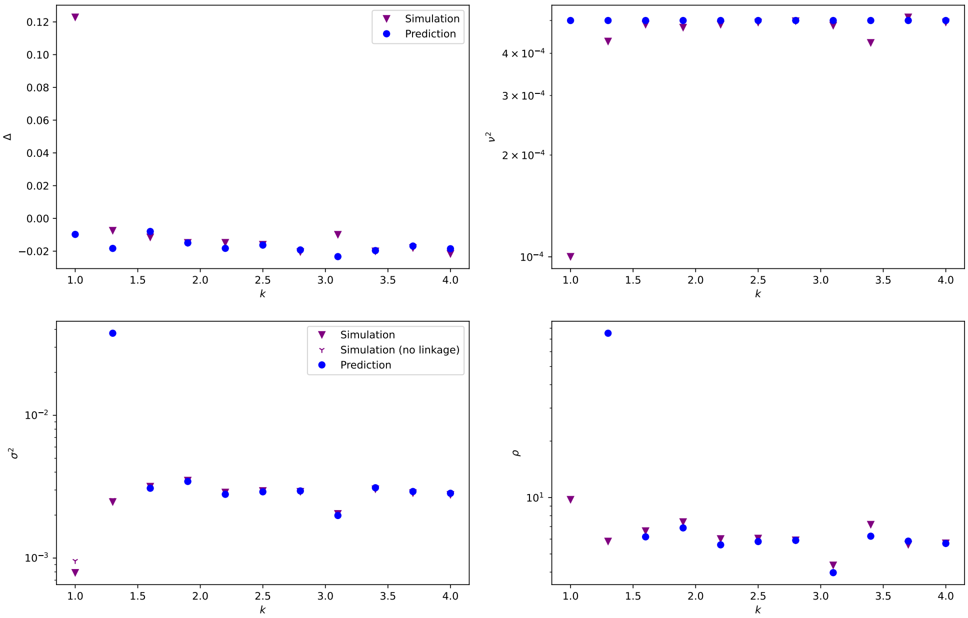

Figure 4.11: Breakdown of the approximation when $(\alpha_\ell)_{\ell\in[L]}$ has heavy tails. We simulated the system at stationarity for $N = 500, L = 100, \eta = 1.2, \omega_e^{-2} = 10^3, \theta = (0.1, 0.2)$. We sampled $\hat{\alpha}_\ell$ using the $Pareto(k)$ distribution for k between 1 and 4, and set the allelic effect as $\alpha_\ell := \hat{\alpha}_\ell/(\sum_\ell \hat{\alpha}_\ell) = 1$. In particular, the smaller k, the heavier the tail of α. The prediction was obtained using the same method as in Fig. 4.10, using the empirical distribution of $(\alpha_\ell)_{\ell\in[L]}$ as the distribution $\mathcal{L}$. When $k = 1$, a single locus can have a very large effect, $\alpha_\ell \sim 1$, and the mean-field approximation cannot be expected to hold. The prediction seems to hold quite well for $k > 1.5$. This is particularly surprising because the theoretical prediction for σ^2 and ρ are obtained using the expectation $\mathbb{E}^*[\alpha^2 P_t(1-P_t)]$ and equation (4.29) for Δ^* involves $\mathbb{E}^*[\alpha^3 P_t(1-P_t)(1-2P_t)]$, whereas $\mathbb{E}\left[\alpha^2\right]$ and $\mathbb{E}\left[\alpha^3\right]$ are ill-behaved when $k \leq 2$. We defer to future work a theoretical characterization of this breakdown.

4.F.4 Discussion of (H4-4')

In Section 4.D.2, (H4-4') were used to argue that $\xi_{\Delta_t,\alpha_\ell}$ can be replaced with $\xi_{\Delta^*,\alpha_\ell}$ in (4.26). Let us check that under (A5) and (N2), (H4) or (H4') are satisfied.

Under (A5), we have either $\omega_e^{-2} \ll L^2$ or $\omega_e^{-2} \sim L^2$ (strong selection). The former situation corresponds to (H4), while the latter corresponds to (H4'a). Furthermore, (H4'b) is satisfied from (4.56), and (H4'c) is satisfied under (H4'a) and (N2).

4.F.5 Breakdown of the equation for the dynamics of the trait mean

It can be seen from Fig. 4.1 that the theoretical predictions for (ν, ρ) from (4.57) fail for weak selection ($\omega_e^{-2} \sim L$), whereas those for σ^2, Δ^* still hold. This stems from the fact that the domain of validity of (4.30) is smaller than that of (4.28-4.30). Here we discuss the breakdown of (4.30) for the fluctuations of the trait mean $(\varepsilon_t)_{t\geq 0}$ when (N3) is not satisfied, and suggest a proxy equation when mutation rates are constant across loci.

Breakdown of (4.30) for weak selection

Here, we show why (N3) is necessary for the Ornstein-Uhlenbeck SDE (4.30) to hold.

Let us start by rewriting (4.42) as follows

$$\mathrm{d}\Delta_t = \frac{1}{\tau}\Bigg(-\Delta_t \times \frac{1}{|\bar\theta|} \times \mathbb{E}^*\left[2(L\alpha)^2 P_t(1-P_t)\right] \;+\; \frac{1}{|\bar\theta|} \times \mathbb{E}^*\left[2L^2\alpha^3\left(P_t - \frac{1}{2}\right)P_t(1-P_t)\right]$$
$$+\frac{L\omega_e^2}{|\bar\theta|}\times\mathbb{E}^*\left[2L\alpha\left(\theta^+(1-P_t) - \theta^- P_t\right)\right]\Bigg)\mathrm{d}t \;+\; \frac{1}{\sqrt{\tau}}\times\sqrt{\frac{\omega_e^2}{|\bar\theta|}\mathbb{E}^*\left[(2L\alpha)^2P_t(1-P_t)\right]}\mathrm{d}B_t^\Delta + \mathrm{d}E_t$$

where we added the term E_t, which is the **error term of the mean-field approximations** (4.44-4.47). Specifically, we define $E_t := E_t^1 + E_t^2 + E_t^3 + E_t^4$ where

$$\mathrm{d}E_t^1 := -\frac{1}{\tau} \times \Delta_t \times \frac{1}{|\bar\theta|} \times \left(\frac{1}{L}\sum_{\ell\in[L]} 2(L\alpha_\ell)^2 P_t^\ell(1-P_t^\ell) \;-\mathbb{E}^*\left[2(L\alpha)^2P_t(1-P_t)\right]\right)\mathrm{d}t$$

$$\mathrm{d}E_t^2 := \frac{1}{\tau|\bar\theta|}\times\left(\frac{1}{L}\sum_{\ell\in[L]} 2L^2\alpha_\ell^3\left(P_t^\ell - \frac{1}{2}\right)P_t^\ell(1-P_t^\ell) \;-\mathbb{E}^*\left[2L^2\alpha^3\left(P_t-\frac{1}{2}\right)P_t(1-P_t)\right]\right)\mathrm{d}t$$

$$\mathrm{d}E_t^3 := \frac{L\omega_e^2}{\tau|\bar\theta|}\times\left(\frac{1}{L}\sum_{\ell\in[L]} 2L\alpha_\ell\left(\theta_\ell^+(1-P_t^\ell) - \theta_\ell^- P_t^\ell\right) \;-\; \mathbb{E}^*\left[2L\alpha\left(\theta^+(1-P_t)-\theta^-P_t\right)\right]\right)\mathrm{d}t$$

$$\mathrm{d}E_t^4 := \sqrt{\frac{\omega_e^2}{\tau|\bar\theta|}}\left(\sqrt{\frac{1}{L}\sum_{\ell\in[L]}(2L\alpha_\ell)^2P_t^\ell(1-P_t^\ell)} - \sqrt{\mathbb{E}^*\left[(2L\alpha)^2P_t(1-P_t)\right]}\right)\mathrm{d}B_t^\Delta$$

Using $\varepsilon_t := \Delta_t - \Delta^*$ and Section 4.D.3, (4.42) can be rewritten

$$\mathrm{d}\varepsilon_t = -\rho\varepsilon_t + \omega_e\sqrt{2\rho}\,\mathrm{d}B_t^\Delta + \mathrm{d}E_t$$

We consider (4.30) for the description of $(\varepsilon_t)_{t\geq 0}$ to be valid as long as the error term E_t is negligible with respect to the other terms. Since (4.30) is an Ornstein-Uhlenbeck process with variance ω_e^2 and autocorrelation parameter $\rho \sim \frac{1}{\tau}$ (see (4.57)), this the contribution of the terms of (4.30) is of order ω over a timescale of order τ. Therefore, (4.30) is valid provided

$$\omega_e \gg \left|\tau\frac{\mathrm{d}}{\mathrm{d}t}E_t^1\right| + \left|\tau\frac{\mathrm{d}}{\mathrm{d}t}E_t^2\right| + \left|\tau\frac{\mathrm{d}}{\mathrm{d}t}E_t^3\right| + \sqrt{\tau\frac{\mathrm{d}}{\mathrm{d}t}\left\langle E^4\right\rangle_t}$$

which, in light of (4.63) we rewrite

$$\omega_e \gg \Delta_t \times \frac{1}{|\bar{\theta}|} \times \sqrt{\frac{\mathbb{V}\mathbf{ar}^*\left[2(L\alpha)^2 P_t(1-P_t)\right]}{L}} + \frac{1}{|\bar{\theta}|}\sqrt{\frac{\mathbb{V}\mathbf{ar}^*\left[2L^2\alpha^3\left(P_t - \frac{1}{2}\right)P_t(1-P_t)\right]}{L}}$$
$$+ \frac{L\omega_e^2}{|\bar{\theta}|} \times \sqrt{\frac{\mathbb{V}\mathbf{ar}^*\left[2L\alpha\left(\theta^+(1-P_t) - \theta^- P_t\right)\right]}{L}} + \sqrt{\frac{\omega_e^2}{|\bar{\theta}|}} \times \sqrt{\frac{\mathbb{V}\mathbf{ar}^*\left[(2L\alpha)^2 P_t(1-P_t)\right]}{L}}$$

In light of (4.64-4.66), this becomes

$$\omega_e \gg \frac{\Delta_t}{|\bar{\theta}|} \times \sqrt{\frac{|\bar{\theta}|}{L}} + \frac{1}{|\bar{\theta}|} \times \sqrt{\frac{|\bar{\theta}|}{L^3}} + \frac{L\omega_e^2}{|\bar{\theta}|} \times \sqrt{\frac{|\bar{\theta}|^2}{L}} + \sqrt{\frac{\omega_e^2}{|\bar{\theta}|}} \times \sqrt{\frac{|\bar{\theta}|}{L}}$$

which yields

$$\omega_e \gg \frac{\Delta_t}{\sqrt{|\bar{\theta}|L}} + \frac{1}{\sqrt{|\bar{\theta}|L^3}} + \omega_e^2\sqrt{L} + \frac{\omega_e}{\sqrt{L}}.$$

We can ignore the second term using (A5)+(N2) and the fourth term on the right-hand side using $L \gg 1$, and use that $\Delta_t \sim L\omega_e^2$ from (4.56) to get

$$\omega_e \gg \omega_e^2\sqrt{\frac{L}{|\bar{\theta}|}} + \omega_e^2\sqrt{L}.$$

Because $|\bar{\theta}| \lesssim 1$ (A2), the second term is smaller than the first one and we get

$$\omega_e \gg \omega_e^2\sqrt{\frac{L}{|\bar{\theta}|}}$$

which can be rewritten

$$1 \gg \frac{L\omega_e^2}{|\bar{\theta}|}.$$

This is precisely (N3).

A proxy equation for the fluctuations under weak selection

Assumption (N3) is not satisfied in many circumstances, for instance under weak selection ($\omega_e^{-2} \sim L$). From the previous section, this means the Ornstein-Uhlenbeck equation (4.30) cannot give a good description of the fluctuations of $(\Delta_t)_{t\geq 0}$ for weak selection. Here, we suggest a proxy equation for this regime. We will derive the proxy equation under the assumption

$$|\theta_\ell| = \theta_\ell^+ + \theta_\ell^- = |\bar{\theta}| \text{ is constant across loci.}$$

We will obtain the following proxy SDE for $(\varepsilon_t)_{t\geq 0}$

$$\mathrm{d}\varepsilon_t = -\tilde{\rho}\varepsilon_t\,\mathrm{d}t + \tilde{\nu}\sqrt{2\tilde{\rho}}\,\mathrm{d}B_t^\Delta \tag{4.67}$$

with

$$\tilde{\rho} := \frac{1}{\tau}\left(\frac{L\sigma^2}{|\bar{\theta}|} + L\omega_e^2\right) \tag{4.68}$$

$$\frac{1}{\tilde{\nu}^2} := \frac{1}{\omega_e^2} + \frac{|\bar{\theta}|}{\sigma^2} \tag{4.69}$$

By proxy equation, we mean that (4.67) is not the correct mathematical object to describe the limit, but can nevertheless yield a sufficient approximation for practical purposes (see Fig. 4.12). In particular, this approximation is used to compute the theoretical predictions under weak selection for ν^2 and ρ in Fig. 4.1. Notice how under moderate/strong selection ($L \ll \omega_e^{-2} \lesssim L^2$), we have from (4.56) $\omega_e^2 \ll \frac{\sigma^2}{|\bar{\theta}|}$ and therefore $\tilde{\rho} \simeq \rho, \tilde{\nu} \simeq \omega_e^2$, thus recovering (4.57). On the other hand, if $\omega_e^{-2} \ll L$ (ultra-weak selection), then we recover an Ornstein-Uhlenbeck process dominated by mutations.

Derivation of (4.67). Our proxy equation will account for the term E_t^3 from Section 4.F.5, but not for the term E_t^1. The missing term E_t^1 is precisely why this equation is not exact. In practice, this term slightly inflates the variance $\nu^2 = \mathbb{V}\mathbf{ar}[\Delta_t]$ in Fig. 4.1, and leads the log-autocorrelation function $(\rho_u)_{u\geq 0}$ to depart from linearity (Fig. 4.12).

Using that $|\theta_\ell|$ is constant across loci, we write

$$\begin{aligned}\sum_{\ell\in[L]} 2L\alpha_\ell(\theta_\ell^+(1-P_t^\ell) - \theta_\ell^- P_t^\ell) &= \sum_{\ell\in[L]} 2L\alpha_\ell\theta_\ell^+ \; - \; |\bar{\theta}|L \sum_{\ell\in[L]} 2\alpha_\ell P_t^\ell \\ &= \sum_{\ell\in[L]} 2L\alpha_\ell\theta_\ell^+ \; - \; |\bar{\theta}|L(\Delta_t + \eta)\end{aligned}$$

from the definition of Δ_t in (4.25). Using a mean-field approximation we get

$$\sum_{\ell\in[L]} 2L\alpha_\ell(\theta_\ell^+(1-P_t^\ell) - \theta_\ell^- P_t^\ell) \simeq L\mathbb{E}^*[2L\alpha\theta^+] - |\bar{\theta}|L(\Delta_t + \eta).$$

It can be checked that the error of this mean-field approximation is always negligible under (A1). It follows

$$\begin{aligned}\mathrm{d}E_t^3 =& \frac{L\omega_e^2}{\tau|\bar{\theta}|}\left(\mathbb{E}^*[2L\alpha\theta^+] - |\bar{\theta}|\Delta_t + |\bar{\theta}|\eta - \mathbb{E}^*\left[2L\alpha\left(\theta^+(1-P_t) - \theta^- P_t\right)\right]\right)\mathrm{d}t \\ =& \frac{L\omega_e^2}{\tau|\bar{\theta}|}\left(-|\bar{\theta}|\Delta_t + |\bar{\theta}|\eta - \mathbb{E}^*\left[2L\alpha|\theta|P_t\right]\right)\mathrm{d}t\end{aligned}$$

Therefore, accounting for E_t^3, (4.42) becomes

$$\begin{aligned}\mathrm{d}\Delta_t = \frac{1}{\tau}\Bigg(&-\Delta_t \times \frac{1}{|\bar{\theta}|} \times \mathbb{E}^*\left[2(L\alpha)^2 P_t(1-P_t)\right] \; + \; \frac{1}{|\bar{\theta}|} \times \mathbb{E}^*\left[2L^2\alpha^3\left(P_t - \frac{1}{2}\right)P_t(1-P_t)\right] \\ &+ \frac{L\omega_e^2}{|\bar{\theta}|} \times \mathbb{E}^*\left[2L\alpha\left(\theta^+(1-P_t) - \theta^- P_t\right)\right]\Bigg)\mathrm{d}t \; + \; \frac{1}{\sqrt{\tau}} \times \sqrt{\frac{\omega_e^2}{|\bar{\theta}|}\mathbb{E}^*\left[(2L\alpha)^2 P_t(1-P_t)\right]}\mathrm{d}B_t^\Delta \\ &+ \frac{L\omega_e^2}{\tau|\bar{\theta}|}\left(-|\bar{\theta}|\Delta_t + |\bar{\theta}|\eta - \mathbb{E}^*\left[2L\alpha|\theta|P_t\right]\right)\mathrm{d}t\end{aligned}$$

This can be rewritten

$$\begin{aligned}\mathrm{d}\Delta_t = \frac{1}{\tau}\Bigg(&-\left(\frac{\mathbb{E}^*\left[2(L\alpha)^2 P_t(1-P_t)\right]}{|\bar{\theta}|} + L\omega_e^2\right)\Delta_t \; + \; \frac{1}{|\bar{\theta}|}\times\mathbb{E}^*\left[2L^2\alpha^3\left(P_t - \frac{1}{2}\right)P_t(1-P_t)\right] \\ &+ L\omega_e^2\left(-\eta + \frac{\mathbb{E}^*\left[2L\alpha\theta^+\right]}{|\bar{\theta}|}\right)\Bigg)\mathrm{d}t \; + \; \frac{1}{\sqrt{\tau}} \times \sqrt{\frac{\omega_e^2}{|\bar{\theta}|}\mathbb{E}^*\left[(2L\alpha)^2 P_t(1-P_t)\right]}\mathrm{d}B_t^\Delta \qquad (4.70)\end{aligned}$$

This yields

$$\mathrm{d}\Delta_t = -\tilde{\rho}(\tilde{\Delta}^* - \Delta_t)\mathrm{d}t + \tilde{\nu}\sqrt{2\tilde{\rho}}\mathrm{d}B_t^\Delta$$

where

$$\begin{aligned}\tilde{\Delta}^* :=& \frac{1}{\tau\tilde{\rho}} \times \left(\frac{1}{|\bar{\theta}|} \times \mathbb{E}^* \left[2L^2\alpha^3 \left(P_t - \frac{1}{2} \right) P_t(1-P_t) \right] + L\omega_e^2 \left(-\eta + \frac{\mathbb{E}^*\left[2L\alpha\theta^+\right]}{|\bar{\theta}|} \right) \right) \\ \tilde{\rho} :=& \frac{1}{\tau} \left(\frac{\mathbb{E}^*\left[2(L\alpha)^2 P_t(1-P_t)\right]}{|\bar{\theta}|} + L\omega_e^2 \right) \\ \tilde{\nu}^2 =& \frac{\frac{\omega_e^2}{|\bar{\theta}|}\mathbb{E}^*\left[(2L\alpha)^2 P_t(1-P_t)\right]}{2\left(\frac{\mathbb{E}^*\left[2(L\alpha)^2 P_t(1-P_t)\right]}{|\bar{\theta}|} + L\omega_e^2 \right)}\end{aligned}$$

To obtain (4.67-4.69), it remains to show that $\tilde{\Delta}^* = \Delta^*$.

From (4.49) we know

$$\Delta^* = \frac{1}{\tau|\bar{\theta}|\rho} \left(\mathbb{E}^* \left[2L^2\alpha^3 \left(P_t - \frac{1}{2} \right) P_t(1-P_t) \right] + L\omega_e^2 \mathbb{E}^* \left[2L\alpha \left(\theta^+(1-P_t) - \theta^- P_t \right) \right] \right).$$

It follows

$$\tilde{\Delta}^* = \frac{1}{\tau\tilde{\rho}} \left(\tau\rho\Delta^* - \frac{L\omega_e^2}{|\bar{\theta}|}\mathbb{E}^* \left[2L\alpha \left(\theta^+(1-P_t) - \theta^- P_t \right) \right] + L\omega_e^2 \left(-\eta + \frac{\mathbb{E}^*\left[2L\alpha\theta^+\right]}{|\bar{\theta}|} \right) \right)$$

which we rewrite using the definition of $\tau = \frac{L\omega_e^2}{|\bar{\theta}|}$

$$\tilde{\Delta}^* = \frac{1}{\tilde{\rho}} \left(\rho\Delta^* - \mathbb{E}^* \left[2L\alpha \left(\theta^+(1-P_t) - \theta^- P_t \right) \right] - |\bar{\theta}|\eta + \mathbb{E}^*\left[2L\alpha\theta^+\right] \right).$$

This can be rewritten

$$\tilde{\Delta}^* = \frac{1}{\tilde{\rho}} \left(\rho\Delta^* + |\bar{\theta}|\mathbb{E}^*\left[2L\alpha P_t\right] - |\bar{\theta}|\eta \right).$$

Using $\Delta^* = \mathbb{E}^*[2L\alpha P_t] - \eta$, we get

$$\tilde{\Delta}^* = \frac{1}{\tilde{\rho}} \left(\rho\Delta^* + |\bar{\theta}|\Delta^* \right).$$

Because $\tilde{\rho} = \rho + |\bar{\theta}|$, we obtain as claimed

$$\tilde{\Delta}^* = \Delta^*.$$

□

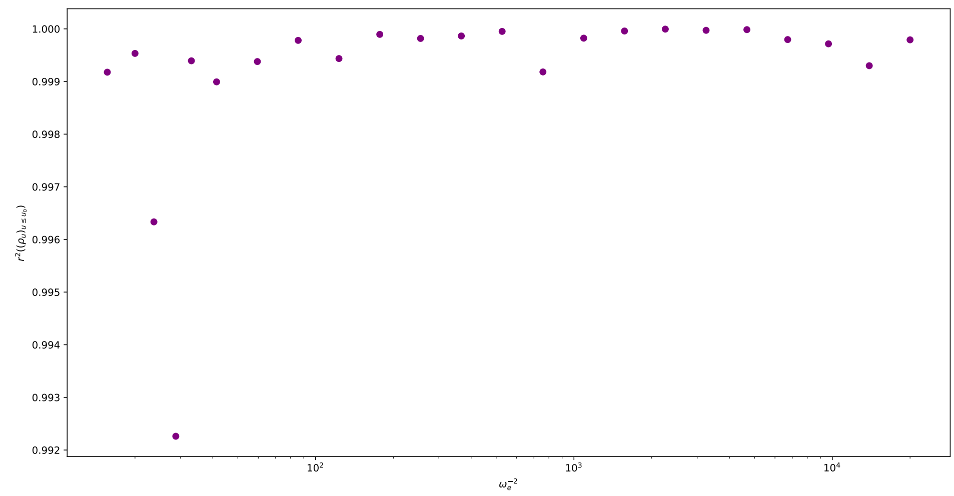

Figure 4.12: Here we plot the square of Pearson's correlation coefficient r^2 for the log-autocorrelation function $(\rho_u)_{u\leq u_0}$ defined in (4.51) as a function of the selection strength ω_e^{-2}. For each value of ω_e^{-2}, ρ_u is computed for times $u \leq u_0$ with $u_0 = \ln(2)/\rho_{1/(2N)}$, and r^2 was computed on $(\rho_u)_{u\leq u_0}$. The time u_0 was chosen such that we may expect $e^{-\rho_{u_0}} \simeq 1/2$ for an Ornstein-Uhlenbeck process.
For an Ornstein-Uhlenbeck process, we expect $r^2 = 1$. As expected, the deviation from the expectation of an Ornstein-Uhlenbeck process appears for weak selection.
The simulations were carried out with $L = 100, N = 500$, the same mutation probabilities for all loci $\theta = (0.1, 0.2)$, and additive effects $(\alpha_\ell)_{\ell\in[L]}$ exponentially distributed with parameter L (in particular $\bar{\alpha} = 1/L$). The selection optimum is $\eta = 1.2$. The same $(\alpha_\ell)_{\ell\in[L]}$ were used in all simulations. The simulations were run for $T = 1000N$ generations, and each observable was measured as an average over the last $500N$ generations.

4.G Equivalence with the trait's eye-view

Here we recover $\xi_{\Delta_t,\alpha_\ell}$ the selection coefficient of the diffusion equation (4.26) from the trait's eye-view, as originally proposed by Wright in [96]. This method has seen much use since (see for instance the supplementary data of [116, 118]). We will show that

$$s^0_\ell(\mathbf{P}_t) := 2N \frac{\mathbf{Cov}_{\mathbf{P}_t}[F(Z(g)), g_\ell]}{\mathbf{Var}_{\mathbf{p}}[g_\ell]}$$

satisfies

$$s^0_\ell(\mathbf{P}_t) \simeq \xi_{\Delta_t,\alpha_\ell}(P^\ell_t)$$

under the following hypotheses

(T1) We may write $Z(g) \simeq \mathcal{Z}_{\ell,t} + \alpha_\ell g_\ell - 2\alpha_\ell P^\ell_t$ where under $\mathbf{E}_{\mathbf{P}_t}$, the law of $\mathcal{Z}_{\ell,t}$ is well approximated by a normal distribution $\mathcal{N}(\bar{z}_t, \sigma^2_{\ell,t})$ for some parameters $\bar{z}_t, \sigma^2_{\ell,t}$.

(T2) We have $\sigma_{\ell,t} \ll \omega \qquad \frac{|\Delta_t|}{L} \ll \omega^2$

Discussion of the hypotheses. (T1) is interpreted as a consequence of the central limit theorem under HWLE (H1), meaning that if $(\alpha_{\ell'} g_{\ell'})_{\ell' \in [L]}$ are independent under $\mathbf{E}_{\mathbf{P}_t}$, then

$$\mathcal{Z}_{\ell,t} := 2\alpha_\ell P^\ell_t + \sum_{\ell' \in [L] \setminus \{\ell\}} \alpha_{\ell'} g_{\ell'}$$

suitably rescaled converges to a normally distributed variable. This use of the central limit theorem should be interpreted with caution, because **the error on the central limit theorem is of order** $\frac{\sigma_{\ell,t}}{\sqrt{L}}$ (see the Berry-Esseem inequality, for instance Theorem 3, Chapter V of [91], see also the discussion in Appendix C of [11]), which is precisely the order of $\alpha_\ell g_\ell$. Furthermore, the Central Limit Theorem cannot be expected to hold if the genetic variance is too small, that is, $\sigma^2_{\ell,t} \lesssim 1/L^2$. This last point is compatible with the scaling obtained in (4.56) when (N2) is satisfied.

Similarly, (T2) is compatible with (4.56) when (N1) is satisfied. □

Alternative derivation of (4.40). Let us define

$$\forall i \in \{0,1,2\}, \qquad w_{\ell,i}(\mathbf{P}_t) := \mathbf{E}_{\mathbf{P}_t}\left[e^{W(\mathcal{Z}_{\ell,t} + \alpha_\ell g_\ell - 2\alpha_\ell P^\ell_t)} \,\middle|\, g_\ell = i \right]$$

Standard computations from population genetics yield

$$s^0_\ell(\mathbf{P}_t) = 2N \frac{(w_{\ell,2}(\mathbf{P}_t) - w_{\ell,1}(\mathbf{P}_t))P^\ell_t + (w_{\ell,1}(\mathbf{P}_t) - w_{\ell,0}(\mathbf{P}_t))(1-P^\ell_t)}{w_{\ell,2}(\mathbf{P}_t)(P^\ell_t)^2 + w_{\ell,1}(\mathbf{P}_t)2P^\ell_t(1-P^\ell_t) + w_{\ell,0}(\mathbf{P}_t)(1-P^\ell_t)^2} \tag{4.71}$$

Furthermore, the Gaussian approximation (T1) allows us to compute for $i \in \{0,1,2\}$

$$w_{\ell,i}(\mathbf{P}_t) \propto \exp\left[-\frac{1}{2(\sigma^2_{\ell,t} + \omega^2)} (\alpha_\ell i - 2\alpha_\ell P^\ell_t + \Delta_t)^2 \right]$$

where $\Delta_t := \bar{z}_t - \eta$.

We then use (T2) to write $\sigma^2_{\ell,t} + \omega^2 \simeq \omega^2$ and $\alpha_\ell |\Delta_t| \ll \omega^2$. These two approximations yield

$$\begin{aligned} w_{\ell,i}(\mathbf{P}_t) \propto\ & e^{-\frac{\Delta_t^2}{2(\sigma^2_{\ell,t}+\omega^2)}} \left(1 - \frac{1}{2\omega^2} \left((\alpha_\ell i - 2\alpha_\ell P^\ell_t)^2 + 2\Delta_t(\alpha_\ell i - 2\alpha_\ell P^\ell_t) \right) \right) \\ \propto\ & 1 - (i - 2P^\ell_t)\frac{\alpha_\ell}{\omega^2}\Delta_t - \frac{\alpha_\ell^2}{2\omega^2}(i - 2P^\ell_t)^2 \end{aligned}$$

Carrying this in (4.71) we find

$$s^0_\ell(\mathbf{P}_t) \simeq -\frac{\alpha_\ell}{\omega_e^2}\Delta_t - \frac{\alpha_\ell^2}{\omega_e^2}\left(\frac{1}{2} - P^\ell_t\right)$$

We thus recover (4.27). $\square$

Remark 13. *This alternative derivation of the selection coefficient is much more computationally cumbersome, which is why we will use our first approach when discussing extensions of the model in Section 4.I.*

4.H Derivation of the criterion for selection to increase genetic variance.

Fix the values of L, η of the model. Assume $\theta_\ell = \theta$ and $\alpha_\ell = 1/L$ for some fixed $\theta \in (0, +\infty)^2$. In this section, we will alleviate notation by writing $\mathbb{E}_s$ for the expectation of a typical locus P_t with distribution

$$C_{s,\theta} p^{2\theta^+ - 1}(1-p)^{2\theta^- - 1} e^{2sp}\mathrm{d}p$$

where $C_{s,\theta}$ is a normalization constant.

Define the mutational optimum

$$z_M := 2L\mathbb{E}_0\left[P_t\right]$$

Here we prove the following: assume

$$(1 - z_M)(\eta - z_M) > 0 \tag{4.72}$$

Then there exists an interval $I = (0, \varepsilon]$ such that for any $s \in I$

$$\mathbb{E}_s[P_t(1-P_t)] > \mathbb{E}_0[P_t(1-P_t)]$$

The derivation is as follows. By symmetry, it is enough to treat the case where $1 > z_M$, which is equivalent to

$$2\theta^+ < |\theta| \tag{4.73}$$

In this case, the criterion 4.72 translates to $\eta > z_M$. This implies $\Delta^* < 0$, and therefore from (4.59) $s^* > 0$. We therefore only need to show

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0}\mathbb{E}_s[P_t(1-P_t)] > 0$$

Simple computations show

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0}\mathbb{E}_s[P_t(1-P_t)] = 2\left(\mathbb{E}_0[P_t^2(1-P_t)] - \mathbb{E}_0[P_t(1-P_t)]\mathbb{E}_0[P_t]\right). \tag{4.74}$$

Under $\mathbb{E}_0$, P_t has distribution $Beta(2\theta^+, 2\theta^-)$. In particular this yields for any $a, b > 0$

$$\mathbb{E}_0[P_t^a(1-P_t)^b] = \frac{2\theta^+(2\theta^+ + 1)\dots(2\theta^+ + a)2\theta^-(2\theta^- + 1)\dots(2\theta^- + b)}{2|\theta|(2|\theta| + 1)\dots(2|\theta| + a + b)}$$

where $|\theta| = \theta^+ + \theta^-$. In particular

$$\mathbb{E}_0[P_t] = \frac{\theta^+}{|\theta|}$$

$$\mathbb{E}_0[P_t(1-P_t)] = \frac{2\theta^+ 2\theta^-}{2|\theta|(2|\theta| + 1)}$$

$$\mathbb{E}_0[P_t^2(1-P_t)] = \frac{2\theta^+(2\theta^+ + 1)2\theta^-}{2|\theta|(2|\theta| + 1)(2|\theta| + 2)}$$

It follows from (4.73)

$$\begin{aligned}\mathbb{E}_0[P_t^2(1-P_t)]-\mathbb{E}_0[P_t(1-P_t)]\mathbb{E}_0[P_t] =& 2\frac{\theta^+\theta^-\left(|\theta|(2\theta^++1)-\theta^+(2|\theta|+2)\right)}{|\theta|^2(2|\theta|+1)(2|\theta|+2)}\\ =& 2\frac{\theta^+\theta^-}{|\theta|^2(2|\theta|+1)(2|\theta|+2)}\left(|\theta|-2\theta^+\right)\\ >& 0\end{aligned}$$

where we used (4.73). The result follows, carrying this in (4.74).

4.I Some extensions

Here we show how our computations can be extended to handle special cases. In Section 4.I.1 we mention polyploidy, in Section 4.I.2 pleiotropy, in Section 4.I.3 dominance, and in Section 4.I.4 epistasis. We always start from an analog of the diffusion approximation (4.26), with selection coefficient given by (4.40). In each extension, the analog of assumptions (A1-6) and (N1-3) is assumed to hold.

4.I.1 Polyploidy

For a polyploid (or haploid) population in which each organism has k haploid genomes with $k \geq 1$, the following alterations need to be made to the diffusion equation (4.26)

- For $\mathbf{p} \in [0,1]^L$, g_ℓ under $\mathbf{E}_\mathbf{p}$ has law *Binomial*(k, p_ℓ).
- In (4.26), the definition of ω_e is $\omega_e^{-2} := (kN)\omega^{-2}$, and the definition of θ_ℓ is $\theta_\ell := kN\mu_\ell$.

In this case the computations of (4.27) yield

$$s_\ell(\mathbf{p}) = \frac{\alpha_\ell}{\omega^2}\left(\eta - k\sum_{\ell'\in[L]}\alpha_{\ell'}p_{\ell'}\right) + \frac{\alpha_\ell^2}{\omega^2}\left(p_\ell - \frac{1}{2}\right)$$

All of the previous results can be obtained from this, defining

$$\Delta_t = k\sum_{\ell'\in[L]}\alpha_{\ell'}P_t^{\ell'} - \eta$$

4.I.2 Pleiotropy

We give here an outlook as to how the theory can be adapted to account for pleiotropy, deferring to future work a rigorous treatment of edge cases.

We now assume that $\alpha_\ell = (\alpha_\ell^i)_{i\in[d]}$ takes values in $\mathbb{R}^d$ with d fixed. In particular, we allow $\alpha_\ell^i < 0$. This is because when $d = 1$, up to replacing $(\alpha_\ell, (\theta_\ell^+, \theta_\ell^-), P_t^\ell)$ with $(-\alpha_\ell, (\theta_\ell^-, \theta_\ell^+), 1-P_t^\ell)$, we may always assume $\alpha_\ell > 0$, but that is no longer true when α_ℓ has dimension d.

In this setting, $Z(g) = \sum_\ell \alpha_\ell g_\ell$ is a $d-$dimensional vector encoding d traits. The fitness function is given by

$$\forall z \in \mathbb{R}^d, \qquad W(z) = -\frac{1}{2}(z-\eta)^\top\omega^{-2}(z-\eta)$$

where $^\top$ denotes transposition, ω^2 is a positive definite matrix and $\eta \in \mathbb{R}^d$. The same computations as in Section 4.C.2 show that the selection coefficient at locus ℓ is $\xi_{\Delta_t,\alpha_\ell}(P_t^\ell)$ where Δ_t is as in (4.25) and for $a, \delta \in \mathbb{R}^d$ we define

$$\xi_{\delta,a} : \left\{\begin{array}{lll} [0,1] & \longrightarrow \mathbb{R} \\ p & \longmapsto -a^\top\omega_e^{-2}\delta \; + \; a^\top\omega_e^{-2}a\left(p-\frac{1}{2}\right)\end{array}\right. \tag{4.75}$$

with $\omega_e^{-2} := 2N\omega^{-2}$.

Using the same kind of reasoning as in Section 4.D.1, we argue that we may write

$$\xi_{\Delta_t,\alpha_\ell}(P_t^\ell) \simeq \xi_{\Delta^*,\alpha_\ell}(P_t^\ell) \tag{4.76}$$

where $\Delta^* := \mathbb{E}^*[\Delta_t]$.

It follows that P_t^ℓ is evolves according to a frequency-dependent Wright-Fisher diffusion (4.24) with selection coefficient $\xi_{\Delta^*,\alpha_\ell}$. In particular, P_t^ℓ has stationary distribution $\Pi_{\Delta^*,\alpha_\ell,\theta_\ell}$ where Π is defined as in (4.34) replacing the priginal definition of ξ by its multi-dimensional counterpart (4.75). We can then find the d-dimensional fixed point equation for Δ^* as

$$\Delta^* = 2L\mathbb{E}\left[\int \alpha p \Pi_{\Delta^*,\alpha,\theta}(p)\mathrm{d}p\right] - \eta \tag{4.77}$$

Global observables such as the genetic variance-covariance matrix $\sigma^2 = 2L\mathbb{E}^*[\alpha\alpha^\top P_t(1-P_t)]$ can be obtained from this as in Section 4.E.3.

Future work could tackle the analysis of this system when $d \gg 1$ as as was explored in [116]. We could also investiagte the effect of the variance-covariance matrix of the $(\alpha^i)_{i\in[d]}$, in particular when some traits are very correlated. We also note that proving existence and uniqueness of solutions to (4.77) will require more work than what we did in the single-trait case in Section 4.B.1.

4.I.3 Dominance

We introduce the model in Section 4.I.3, compute the selection coefficient in Section 4.I.3, the evolution of the trait under moderate/strong selection ($\omega_e^{-2} \gg L$) in Section 4.I.3, and obtain the fixed point equation for Δ^* and the stationary distribution of the typical locus in Section 4.I.3.

Model

We now account for dominance. Locus ℓ is now characterized by additive effect α_ℓ, dominance effect D_ℓ and mutation rate θ_ℓ. We assume $|D|$ is of order $\frac{1}{L}$ as α. The trait value and logfitness of a genotype $g = (g_\ell)_{\ell\in[L]}$ is then

$$Z(g) = \sum_{\ell\in[L]} \alpha_\ell g_\ell + D_\ell \mathbb{1}_{[g_\ell=1]} \qquad\qquad W(z) = -\frac{1}{2\omega^2}(z-\eta)^2$$

For a pair $(a, D) \in \mathbb{R}_+ \times \mathbb{R}$ we define the corresponding **average effect of gene substitution** as

$$\beta_{a,D}(p) := a + D(1-2p) \tag{4.78}$$

Selection coefficient

Here we show the selection coefficient $s_\ell(\mathbf{p})$ defined from (4.40) can be rewritten $s_\ell(\mathbf{P}_t) = \xi_{\Delta_t,\alpha_\ell,D_\ell}(P_t^\ell)$ where for $\delta \in \mathbb{R}, a \in \mathbb{R}_+, D \in \mathbb{R}$ we define

$$\xi_{\delta,a,D}(p) := -\frac{\beta_{a,D}(p)}{\omega_e^2}\delta + \omega_e^{-2}\left(\beta_{a,D}(p)^2\left(p-\frac{1}{2}\right) + 2aDp(1-p)\right) \tag{4.79}$$

$$\Delta_t := \mathbf{E}_{\mathbf{P}_t}[Z(g)] - \eta \;=\; 2\sum_{\ell\in[L]}\left(\alpha_\ell P_t^\ell + D_\ell P_t^\ell(1-P_t^\ell)\right) \;-\; \eta. \tag{4.80}$$

Derivation. For a general $\mathbf{p} \in [0,1]^L$ we write

$$\begin{aligned}\mathbf{Cov}_{\mathbf{p}}[(Z(g)-\eta)^2, g_\ell] =& \mathbf{Cov}_{\mathbf{p}}[\mathbf{E}_{\mathbf{p}}[(Z(g)-\eta)^2 \;|g_\ell], g_\ell] \\ =& \mathbf{Cov}_{\mathbf{p}}[\mathbf{E}_{\mathbf{p}}[Z(g)-\eta \;|g_\ell]^2, g_\ell] + \mathbf{Cov}_{\mathbf{p}}[\mathbf{Var}_{\mathbf{p}}[Z(g)-\eta \;|g_\ell], g_\ell]\end{aligned}$$

We can write

$$\mathbf{Var}_{\mathbf{p}}[Z(g)-\eta \;|g_\ell] = \sum_{\ell' \in [L] \setminus \{\ell\}} \mathbf{Var}_{\mathbf{p}}[\alpha_{\ell'} g_{\ell'} + D_{\ell'} \mathbb{1}_{[g_{\ell'}=1]}]$$

In particular

$$\mathbf{Cov}_{\mathbf{p}}[\mathbf{Var}_{\mathbf{p}}[Z(g)-\eta \;|g_\ell], g_\ell] = 0.$$

We are then left with

$$\mathbf{Cov}_{\mathbf{p}}[(Z(g)-\eta)^2, g_\ell] = \mathbf{Cov}_{\mathbf{p}}[\mathbf{E}_{\mathbf{p}}[Z(g)-\eta \;|g_\ell]^2, g_\ell].$$

Appyling this to $\mathbf{P}_t$, we get

$$\mathbf{E}_{\mathbf{P}_t}[Z(g)-\eta \;|g_\ell] = \Delta_t + \alpha_\ell(g_\ell - 2P_t^\ell) + D_\ell(\mathbb{1}_{[g_\ell=1]} - 2P_t^\ell(1-P_t^\ell))$$

using the definition of Δ_t in (4.80). It follows

$$\begin{aligned}\mathbf{Cov}_{\mathbf{P}_t}[(Z(g)-\eta)^2, g_\ell] = \mathbf{Cov}_{\mathbf{P}_t}\Big[&\alpha_\ell^2(g_\ell - 2P_t^\ell)^2 + D_\ell^2(\mathbb{1}_{[g_\ell=1]} - 2P_t^\ell(1-P_t^\ell))^2 \\ &+ 2\alpha_\ell(g_\ell - 2P_t^\ell)D_\ell(\mathbb{1}_{[g_\ell=1]} - 2P_t^\ell(1-P_t^\ell)) \\ &\quad + 2\Delta_t\left(\alpha_\ell(g_\ell - 2P_t^\ell) + D_\ell(\mathbb{1}_{[g_\ell=1]} - 2P_t^\ell(1-P_t^\ell))\right), g_\ell\Big].\end{aligned}$$

We compute

$$\begin{aligned}\mathbf{Cov}_{\mathbf{P}_t}[(g_\ell - 2P_t^\ell)^2, g_\ell] =& (1-2P_t^\ell)2P_t^\ell(1-P_t^\ell) \\ \mathbf{Cov}_{\mathbf{P}_t}[\mathbb{1}_{[g_\ell=1]}, g_\ell] =& (1-2P_t^\ell)2P_t^\ell(1-P_t^\ell) \\ \mathbf{Cov}_{\mathbf{P}_t}[\mathbb{1}_{[g_\ell=1]}g_\ell, g_\ell] =& \mathbf{Cov}_{\mathbf{P}_t}[\mathbb{1}_{[g_\ell=1]}, g_\ell]\end{aligned}$$

In particular

$$\begin{aligned}\mathbf{Cov}_{\mathbf{P}_t}[(g_\ell - 2P_t^\ell)(\mathbb{1}_{[g_\ell=1]} - 2P_t^\ell(1-P_t^\ell)), g_\ell] =& \mathbf{Cov}_{\mathbf{P}_t}[(1-2P_t^\ell)\mathbb{1}_{[g_\ell=1]} - 2P_t^\ell(1-P_t^\ell)g_\ell, g_\ell] \\ =& (1-2P_t^\ell)^2 2P_t^\ell(1-P_t^\ell) - \big(2P_t^\ell(1-P_t^\ell)\big)^2.\end{aligned}$$

It follows

$$\begin{aligned}\mathbf{Cov}_{\mathbf{P}_t}[(Z(g)-\eta)^2, g_\ell] = \;& \alpha_\ell^2(1-2P_t^\ell)2P_t^\ell(1-P_t^\ell) + D_\ell^2(1-2P_t^\ell)^3 2P_t^\ell(1-P_t^\ell) \\ &+ 2\alpha_\ell D_\ell\left((1-2P_t^\ell)^2 2P_t^\ell(1-P_t^\ell) - (2P_t^\ell(1-P_t^\ell))^2\right) \\ &\quad + 2\Delta_t\left(\alpha_\ell + D_\ell(1-2P_t^\ell)\right)2P_t^\ell(1-P_t^\ell).\end{aligned}$$

This lets us obtain

$$\begin{aligned}&\frac{\mathbf{Cov}_{\mathbf{P}_t}[(Z(g)-\eta)^2, g_\ell]}{2P_t^\ell(1-P_t^\ell)} \\ &\quad = \left(\alpha_\ell + D_\ell(1-2P_t^\ell)\right)^2(1-2P_t^\ell) - 4\alpha_\ell D_\ell P_t^\ell(1-P_t^\ell) + 2(\alpha_\ell + D_\ell(1-2P_t^\ell))\Delta_t\end{aligned}$$

We get from (4.78)

$$s_\ell(\mathbf{P}_t) = \omega_e^{-2}\Big(\beta_{\alpha_\ell,D_\ell}(P_t^\ell)^2\left(P_t^\ell - \frac{1}{2}\right) + 2\alpha_\ell D_\ell P_t^\ell(1-P_t^\ell) - \beta_{\alpha_\ell,D_\ell}(P_t^\ell)\Delta_t\Big)$$

The result follows. □

Genetic architecture

The same argument as in Section 4.D.2 lets us write

$$\xi_{\Delta_t,\alpha_\ell,D_\ell}(P_t^\ell) \simeq \xi_{\Delta^*,\alpha_\ell,D_\ell}(P_t^\ell)$$

where $\Delta^* = \mathbb{E}^*[\Delta_t]$. Then P_t^ℓ follows a frequency-dependent Wright-Fisher diffusion (4.24) with selection coefficient $\xi_{\Delta^*,\alpha_\ell,D_\ell}$. It follows that P_t^ℓ has stationary distribution $\Pi_{\Delta^*,a_\ell,D_\ell,\theta_\ell}$ where we define for $\delta \in \mathbb{R}, a \in \mathbb{R}_+, D \in \mathbb{R}, \theta \in (0,+\infty)^2$

$$\Pi_{\delta,a,D,\theta}(p) := C_{\delta,a,D,\theta}p^{2\theta^+-1}(1-p)^{2\theta^--1}e^{2\int_0^p \xi_{\delta,a,D}(u)\,\mathrm{d}u}$$

with $C_{\delta,a,D,\theta}$ a normalization constant.

On the other hand we can compute from (4.80)

$$\Delta^* \simeq 2L\mathbb{E}^*[\alpha P_t + DP_t(1-P_t)] - \eta$$

From there we obtain the fixed point equation

$$\Delta^* \simeq 2I(\Delta^*) - \eta \tag{4.81}$$

where

$$I(\delta) := \mathbb{E}^*\left[\int L(\alpha p + Dp(1-p))\Pi_{\delta,\alpha,D,\theta}(p)\mathrm{d}p\right].$$

The macroscopic observables can be computed from this as in Section 4.E.3, in particular the genetic variance

$$\sigma^2 = 2L\mathbb{E}^*\left[\beta_{\alpha,D}(P_t)^2P_t(1-P_t)\right] + 2L\mathbb{E}^*\left[2D^2\left(P_t(1-P_t)\right)^2\right].$$

The first term is the **additive genetic variance**, the second term is the **dominance variance**.

Evolution of the trait under moderate/strong selection

Here obtain the following stationary system analog to (4.28-4.17)

$$\begin{aligned}
\mathrm{d}P_t =&\xi_{\Delta^*,\alpha,D}(P_t)P_t(1-P_t)\,\mathrm{d}t + \left(\theta^+(1-P_t) - \theta^-P_t\right)\mathrm{d}t + \sqrt{P_t(1-P_t)}\,\mathrm{d}B_t^P\\
\Delta^* =&2L\mathbb{E}^*[\alpha P_t + DP_t(1-P_t)] - \eta\\
\mathrm{d}\varepsilon_t =&-\rho\varepsilon_t\,\mathrm{d}t + \omega_e\sqrt{2\rho}\,\mathrm{d}B_t^\Delta
\end{aligned}$$

where B^P, B^Δ are Brownian motions, $\varepsilon_t := \Delta_t - \Delta^*$ and

$$\rho := \frac{1}{\tau} \times \frac{\mathbb{E}^*\left[2(L\beta_{\alpha,D}(P_t))^2P_t(1-P_t)\right]}{|\bar{\theta}|}$$

Derivation. Applying Itô's formula yields

$$\begin{aligned}
\mathrm{d}\Delta_t =&2\sum_{\ell\in[L]}\left(\alpha_\ell + D_\ell(1-2P_t^\ell)\right)\mathrm{d}P_t^\ell \;-\; \sum_{\ell\in[L]} 2D_\ell P_t^\ell(1-P_t^\ell)\mathrm{d}t\\
=&2\sum_{\ell\in[L]}\beta_{\alpha_\ell,D_\ell}(P_t^\ell)\mathrm{d}P_t^\ell \;-\; \sum_{\ell\in[L]} 2D_\ell P_t^\ell(1-P_t^\ell)\mathrm{d}t
\end{aligned}$$

We define as before

$$\tau := \frac{L\omega_e^2}{|\bar{\theta}|}$$

We can then proceed just as in Section 4.D, using mean-field approximations and time-averaging to write

$$\begin{aligned}
\mathrm{d}\Delta_t = \frac{1}{\tau}\Bigg(& -\Delta_t \times \frac{1}{|\bar{\theta}|} \times \mathbb{E}^*\left[2(L\beta_{\alpha,D}(P_t))^2 P_t(1-P_t)\right] \\
& + \frac{L\omega_e^2}{|\bar{\theta}|} \times \mathbb{E}^*\left[2L\beta_{\alpha,D}(P_t)\left(\theta^+(1-P_t)-\theta^- P_t\right)\right] \\
& + \frac{1}{|\bar{\theta}|} \times \mathbb{E}^*\left[2L^2\beta_{\alpha,D}(P_t)^3\left(P_t-\frac{1}{2}\right)P_t(1-P_t) + 4L^2\alpha D\beta_{\alpha,D}(P_t)P_t^2(1-P_t)^2\right] \\
& - \frac{L\omega_e^2}{|\bar{\theta}|}\mathbb{E}^*[2LDP_t(1-P_t)]\Bigg)\mathrm{d}t \\
& + \frac{1}{\sqrt{\tau}} \times \sqrt{\frac{\omega_e^2}{|\bar{\theta}|}\mathbb{E}^*\left[(2L\beta_{\alpha,D}(P_t))^2 P_t(1-P_t)\right]}\mathrm{d}B_t^\Delta
\end{aligned} \tag{4.82}$$

The result follows. □

4.I.4 Epistasis

Adding epistasis to the model means the trait $Z(g)$ involves interactions between loci. Much work has been done on the subject, for instance [86]. There are many ways such interactions can go, depending on assumptions about gene regulatory networks (the typical model for such networks is the LK fitness landscape [205]). We briefly mention two extreme figures.

Epistasis between small clumps of loci

Under the assumption of small clumps of loci, we assume there exists a partition of $[L]$ into small subsets $(\mathcal{I}_k)_{k\in[n]}$ with size at most $i_{max} \sim 1$ such that the trait value $Z(g)$ is the sum from contributions from each subset

$$Z(g) = z_0 + \sum_{k=1}^{n} Z_k(g_{\mathcal{I}_k}).$$

where $g_{\mathcal{I}_k} := (g_\ell)_{\ell\in\mathcal{I}_k}$ and Z_k is a function on $\{0,1,2\}^{\mathcal{I}_k}$. We assume that there are many clumps of loci ($n \gg 1$), and that for each k, Z_k is independently sampled as a random function on $\{0,1,2\}^{\mathcal{I}_k}$ with a distribution that only depends on $\#\mathcal{I}_k$. In this subsection we derive the fixed-point equation for Δ^* in this system.

We can write the selection coefficient at locus $\ell \in \mathcal{I}_k$ as in (4.40)

$$s_\ell(\mathbf{p}) = -\frac{1}{2\omega_e^2} \times \frac{\mathbf{Cov_p}\left[(Z(g)-\eta)^2, g_\ell\right]}{2p_\ell(1-p_\ell)}$$

Using the independence of $g_{\mathcal{I}_k}$ and $(g_{\mathcal{I}_{k'}})_{k'\neq k}$, this can be rewritten

$$s_\ell(\mathbf{p}) = -\frac{\mathbf{Cov_p}\left[\left(Z_k(g_{\mathcal{I}_k}) - \mathbf{E_p}\left[Z_k(g_{\mathcal{I}_k})\right]\right)^2, g_\ell\right] + 2\mathbf{Cov_p}\left[Z_k(g_{\mathcal{I}_k}), g_\ell\right]\mathbf{E_p}\left[Z(g)\ -\eta\right]}{4\omega_e^2 p_\ell(1-p_\ell)}.$$

Then a mean-field approximation or a time-averaging principle lets us write as in Section 4.D.2

$$\mathbf{E_p}\left[Z(g)\ -\eta\right] \simeq \Delta^* \tag{4.83}$$

for some stationary value Δ^*. We are left with

$$s_\ell(\mathbf{p}) \simeq -\frac{\mathbf{Cov_p}\left[(Z_k(g_{\mathcal{I}_k}) - \mathbf{E_p}\left[Z_k(g_{\mathcal{I}_k})\right])^2, g_\ell\right] + 2\mathbf{Cov_p}\left[Z_k(g_{\mathcal{I}_k}), g_\ell\right]\Delta^*}{4\omega_e^2 p_\ell(1-p_\ell)}$$

Let us define

$$W_k(\mathbf{p}_{\mathcal{I}_k}, \Delta^*) := -\frac{1}{2\omega_e^2}\left((Z_k(g_{\mathcal{I}_k}) - \mathbf{E_p}\left[Z_k(g_{\mathcal{I}_k}) \ - \eta\right])^2 + 2Z_k(g_{\mathcal{I}_k})\Delta^*\right)$$

where $\mathbf{p}_{\mathcal{I}_k} = (p_\ell)_{\ell\in\mathcal{I}_k}$. Then we find

$$s_\ell(\mathbf{p}) \simeq \frac{\mathbf{Cov_p}[W_k(g), g_\ell]}{2p_\ell(1-p_\ell)}$$

Carrying this into our base SDE (4.26), we find that the system $(P_t^\ell)_{\ell\in\mathcal{I}_k}$ is an autonomous multiloci Wright-Fisher diffusion, similar to that studied in [72]. In particular, the stationary density of $(P_t^\ell)_{\ell\in\mathcal{I}_k}$ can be obtained as

$$\forall \mathbf{p} \in [0,1]^{\mathcal{I}_k}, \qquad \Pi_{\Delta^*, W_k, \theta_{\mathcal{I}_k}}(\mathbf{p}) = C_{\Delta^*, W_k, \theta_{\mathcal{I}_k}} \prod_{\ell\in\mathcal{I}_k} p_\ell^{2\theta_\ell^+ - 1}(1-p_\ell)^{2\theta_\ell^- - 1}\exp\left[\mathbf{E_p}[W_k(g_{\mathbf{I}_k}, \Delta^*)]\right]$$

where $C_{\Delta^*, W_k, \theta_{\mathcal{I}_k}}$ is a normalization constant. Finally, we may obtain a fixed point equation for Δ^* from (4.83)

$$\sum_k \int \mathbf{E_p}[Z_k(g)]\Pi_{\Delta^*, W_k, \theta_{\mathcal{I}_k}}(\mathrm{d}\mathbf{p}) \ - \eta = \Delta^*$$

where, for $\mathbf{p} \in [0,1]^{\mathcal{I}_k}$, $\mathbf{E_p}$ is the expectation of a $\{0,1,2\}^{\mathcal{I}_k}$-valued vector g of independent random variables such that g_ℓ has law $Binomial(2, p_\ell)$.

General diffuse epistasis

We now take Z to be some unspecified function on $\{0,1,2\}^L$. We start by decomposing the selection coefficient at locus ℓ into an additive/dominant component and an epistatic component. Then we illustrate with an example from spin-glass theory how to obtain and solve the fixed-point equation. Finally, we compute the magnitude of the stochastic fluctuations of the additive/dominant components.

Decomposition of the selection coefficient. Let us define the following coefficients

$$\hat{\alpha}_\ell(t) := \frac{1}{2}\left(\mathbf{E}_{\mathbf{P}_t}[Z(g)\ |g_\ell = 2] - \mathbf{E}_{\mathbf{P}_t}[Z(g)\ |g_\ell = 0]\right) \tag{4.84}$$

$$\hat{D}_\ell(t) := \frac{1}{2}\left(\mathbf{E}_{\mathbf{P}_t}[Z(g)\ |g_\ell = 2] + \mathbf{E}_{\mathbf{P}_t}[Z(g)\ |g_\ell = 0] - 2\mathbf{E}_{\mathbf{P}_t}[Z(g)\ |g_\ell = 1]\right) \tag{4.85}$$

$$\hat{\beta}_\ell(t) := \hat{\alpha}_\ell(t) + \hat{D}_\ell(t)(1 - 2P_t^\ell) \tag{4.86}$$

$$\epsilon_\ell(t) := \mathbf{Cov}_{\mathbf{P}_t}\left[\mathbf{Var}_{\mathbf{P}_t}\left[Z\ |g_\ell\right], g_\ell\right] \tag{4.87}$$

$$\Delta_t := \mathbf{E}_{\mathbf{P}_t}[Z(g)] \ - \eta \tag{4.88}$$

We derive the following decomposition

Claim 4.I.1. *We have*

$$s_\ell(\mathbf{P}_t)P_t^\ell(1-P_t^\ell) \simeq s_\ell^{Dom}(t)P_t^\ell(1-P_t^\ell) - \frac{\epsilon_\ell(t)}{4\omega_e^2} \tag{4.89}$$

where, by analogy with (4.79) *we define*

$$s_\ell^{Dom}(t) := \omega_e^{-2}\left(\hat{\beta}_\ell(t)^2\left(P_t^\ell - \frac{1}{2}\right) + 2\hat{\alpha}_\ell\hat{D}_\ell P_t^\ell(1-P_t^\ell) - \hat{\beta}_\ell(P_t^\ell)\Delta_t\right).$$

Remark 14. *If we neglect $\varepsilon_\ell(t)$, then we recover the same coefficient as for dominance (Section 4.I.3), with the difficulty that $\hat{\alpha}_\ell$ and $\hat{D}_\ell$ are functions of time. Looking at (4.84-4.87), we see neglecting $\varepsilon_\ell(t)$ is acceptable if g_ℓ has more impact on the mean trait through $\mathbf{E_p}[Z(g)|g_\ell]$ than on the variance of the trait through $\mathbf{Var_p}[Z(g)|g_\ell]$.*

Derivation of Claim 4.I.1. Let us compute the selection coefficient at locus ℓ. We have

$$\begin{aligned}
s_\ell(\mathbf{P}_t)P_t^\ell(1-P_t^\ell) =& 2N\frac{\mathbf{Cov}_{\mathbf{P}_t}[W(Z(g)),g_\ell]}{\mathbf{Var}_{\mathbf{P}_t}[g_\ell]}P_t^\ell(1-P_t^\ell)\\
=& 2N\frac{\mathbf{Cov}_{\mathbf{P}_t}[W(Z(g)),g_\ell]}{2}\\
=& -\frac{1}{4\omega_e^2}\mathbf{Cov}_{\mathbf{P}_t}\left[(Z(g)-\eta)^2,g_\ell\right]
\end{aligned}$$

We can write the decomposition as in Section 4.I.3

$$\mathbf{Cov_p}\left[(Z(g)-\eta)^2,g_\ell\right] = \mathbf{Cov_p}\left[(Z(g)-\mathbf{E_p}\left[Z(g)|g_\ell\right])^2,g_\ell\right] + \mathbf{Cov_p}\left[(\mathbf{E_p}\left[Z(g)|g_\ell\right]-\eta)^2,g_\ell\right] \tag{4.90}$$

The first term is $\epsilon_\ell(t)$. We thus obtain

$$\begin{aligned}
s_\ell(\mathbf{P}_t)P_t^\ell(1-P_t^\ell) =& -\frac{1}{4\omega_e^2}\left(\epsilon_\ell(t)+\mathbf{Cov_p}\left[(\mathbf{E_p}\left[Z(g)|g_\ell\right]-\eta)^2,g_\ell\right]\right)\\
=& -\frac{\epsilon_\ell(t)}{4\omega_e^2}+\mathbf{Cov_p}\left[W\left(\mathbf{E_p}\left[Z(g)|g_\ell\right]\right),g_\ell\right]
\end{aligned}$$

Note that with the notation (4.84-4.85), we have

$$\mathbf{E_p}\left[Z(g)|g_\ell\right] = \mathbf{E_p}\left[Z(g)|g_\ell=0\right] + g_\ell\hat{\alpha}_\ell(t) + \mathbb{1}_{[g_\ell=1]}\hat{D}_\ell(t) \tag{4.91}$$

From there, the same computations as in Section 4.I.3 yield

$$\mathbf{Cov_p}\left[W\left(\mathbf{E_p}\left[Z(g)|g_\ell\right]\right),g_\ell\right] = s_\ell^{Dom}(t)P_t^\ell(1-P_t^\ell).$$

□

Example from spin-glass theory. As an illustrative example, we take pairwise epistasis as in Sherrington-Kirkpatrick theory [103]

$$Z(g) = \sum_{\substack{\ell,\ell'\in[L]\\ \ell\neq\ell'}} f_{\ell\ell'}g_\ell g_{\ell'}$$

where $(f_{\ell\ell'})_{\ell,\ell'\in[L]}$ are i.i.d with uniform bounds of order $1/L^2$ and mean of order $1/L^2$. It can be checked that $\hat{D}_\ell = 0$, because for any $\ell' \neq \ell$

$$\mathbf{E}_{\mathbf{P}_t}[g_\ell g_{\ell'}\;|g_\ell=2] + \mathbf{E}_{\mathbf{P}_t}[g_\ell g_{\ell'}\;|g_\ell=0] - 2\mathbf{E}_{\mathbf{P}_t}[g_\ell g_{\ell'}\;|g_\ell=1] = 2\mathbf{E}_{\mathbf{P}_t}[g_{\ell'}] - 2\mathbf{E}_{\mathbf{P}_t}[g_{\ell'}] = 0.$$

We prove the following claim at the end of this section.

Claim 4.I.2. *In the equation for P_t^ℓ, we may neglect $\epsilon_\ell(t)$ if $\omega_e^{-2}|\bar{\theta}| \ll L^2$.*

Let us comment this result. The fact that we may neglect $\varepsilon_\ell(t)$ if $\omega_e^{-2}|\bar{\theta}| \ll L^2$ tells us that in weak and moderate selection ($\omega_e^{-2} \ll L^2$), or when mutation rates are small ($|\bar{\theta}| \ll 1$), modelling the trait as being determined by $\hat{\alpha}, \hat{D}$ is a good approximation.

Let us now compute the effective selection coefficient $\hat{\alpha}_\ell(t)$. We have from (4.84)

$$\begin{aligned}\hat{\alpha}_\ell(t) = & \sum_{\ell' \in [L] \smallsetminus \{\ell\}} f_{\ell\ell'} \mathbf{E}_{\mathbf{P}_t}[g_{\ell'}] \\ =& 2 \sum_{\ell' \in [L] \smallsetminus \{\ell\}} f_{\ell\ell'} P_t^{\ell'}.\end{aligned}$$

Using a mean-field approximation (H3) we find

$$\hat{\alpha}_\ell(t) \simeq 2L\mathbb{E}\left[f\right]\mathbb{E}\left[P_t\right] \tag{4.92}$$

where P_t is P_t^ℓ for a randomly sampled locus and f is independently sampled with the same law as $f_{1,2}$. In particular, we expect the effective additive coefficient $\hat{\alpha}_\ell(t)$ to be approximately the same at all loci.

We now consider the system at statistical equilibrium (as before denoted $\mathbb{E}^*$). Let us assume that conditional on $\hat{\alpha}_\ell(t), \theta_\ell$, P_t^ℓ has distribution $\Pi_{\Delta^*, \hat{\alpha}_\ell(t), \theta_\ell}$ with

$$\Delta^* := \mathbb{E}^*\left[\mathbf{E}_{\mathbf{P}_t}[Z(g)]\right] - \eta.$$

From (4.92), we expect $\hat{\alpha}_\ell(t)$ to be a constant $\hat{\alpha}^*$. We rewrite this

$$\Delta^* = 2L\mathbb{E}^*[\hat{\alpha}^* P_t] - \eta.$$

It follows from (4.92)

$$\Delta^* \simeq (2L)^2\mathbb{E}\left[f\right]\mathbb{E}\left[P_t\right]^2 - \eta.$$

We thus obtain the fixed-point system

$$\begin{aligned}\Delta^* =& (2L)^2\mathbb{E}\left[f\right]\mathbb{E}^*[P_t]^2 - \eta. \\ \hat{\alpha}^* =& 2L\mathbb{E}\left[f\right]\mathbb{E}^*[P_t] \\ \mathbb{E}^*[P_t] =& \mathbb{E}\left[\int_0^1 p\Pi_{\Delta^*, \hat{\alpha}^*, \theta}(p)\mathrm{d}p\right]\end{aligned}$$

Derivation of Claim 4.I.2. We write

$$\mathbf{Var}_{\mathbf{P}_t}[Z(g)|g_\ell] = g_\ell^2 \sum_{\ell' \neq \ell} \mathbf{Var}_{\mathbf{P}_t}[g_{\ell'}] f_{\ell\ell'}^2 \quad + \quad 2g_\ell \sum_{\substack{\ell'_1, \ell'_2 \in [L] \smallsetminus \{\ell\} \\ \ell'_1 \neq \ell'_2}} \mathbf{Cov}_{\mathbf{P}_t}[g_{\ell'_1}, g_{\ell'_1} g_{\ell'_2}] f_{\ell\ell'_1} f_{\ell'_1\ell'_2} \quad + \quad R$$

where R is independent of g_ℓ. Taking the variance of $g_{\ell'}$ to be of order $|\bar{\theta}|$ as in (4.56), this yields

$$\epsilon_\ell(t) = \mathbf{Cov}_{\mathbf{P}_t}\left[\mathbf{Var}_{\mathbf{P}_t}[Z(g)|g_\ell], g_\ell\right] \sim (L\mathbb{E}\left[f^2\right]|\bar{\theta}| + L^2\mathbb{E}\left[f\right]^2|\bar{\theta}|)P_t^\ell(1 - P_t^\ell) \sim |\bar{\theta}|/L^2 P_t^\ell(1 - P_t^\ell)$$

where we used a mean-field approximation and that f is typically of order $1/L^2$. This yields the result ☐

Fluctuations of the effective additive and dominance effect. Here, we assume we can neglect the term $\varepsilon_\ell(t)$ in the selection coefficient $s_\ell(\mathbf{P}_t)$ from (4.89). We make no other assumption on Z, and we compute the quadratic variation of $\hat{\alpha}_\ell(t)$ and $\hat{D}_\ell(t)$.

We need to introduce a notation. Recall that under $\mathbf{E}_\mathbf{p}$, g_ℓ has law *Binomial*$(2, p_\ell)$. We extend this probability space by writing $g_\ell = G_{\ell,(1)} + G_{\ell,(2)}$ where $G_{\ell,(1)}, G_{\ell,(2)}$ under $\mathbf{E}_\mathbf{p}$ are two independent *Bernoulli*(p_ℓ) variables.

We will show

Claim 4.I.3. *The quadratic variation of $\hat{\alpha}_\ell$ and $\hat{D}_\ell$ are given by*

$$\mathrm{d}\,\langle\hat{\alpha}_\ell\rangle_t = \sum_{\ell'\neq\ell} \mathbf{Var}_{\mathbf{P}_t}\left[\mathbf{E}_{\mathbf{P}_t}\left[Z(g) \mid g_\ell = 2,\ G_{\ell',(1)}\right] - \mathbf{E}_{\mathbf{P}_t}\left[Z(g) \mid g_\ell = 0,\ G_{\ell',(1)}\right]\right]\mathrm{d}t \tag{4.93}$$

$$\begin{aligned}\mathrm{d}\,\langle\hat{D}_\ell\rangle_t = \sum_{\ell'\neq\ell} \mathbf{Var}_{\mathbf{P}_t}\Big[2\mathbf{E}_{\mathbf{P}_t}&\left[Z(g) \mid g_\ell = 1,\ G_{\ell',(1)}\right] \\ &-\mathbf{E}_{\mathbf{P}_t}\left[Z(g) \mid g_\ell = 2,\ G_{\ell',(1)}\right] - \mathbf{E}_{\mathbf{P}_t}\left[Z(g) \mid g_\ell = 0,\ G_{\ell',(1)}\right]\Big]\mathrm{d}t\end{aligned} \tag{4.94}$$

This equation is interesting because, for a given ℓ, the right-hand side is computed from the distribution of $(Z(g), g_\ell)$, which can be obtained directly from data. If the right-hand side is small with respect to $\frac{1}{2N_e}$, where N_e is the effective size of the population ($\sim 10^4$ for humans, see Discussion in Section 4.4), then we claim the stationary distribution of P_t^ℓ conditional on $\hat{\alpha}_\ell(t)$ is well approximated by $\Pi_{\Delta^*,\hat{\alpha}_\ell(t),\theta_\ell}$.

The ℓ' term of the right-hand side of (4.93) quantifies how much knowing the value of $G_{\ell',(1)}$ gives us information about the way $Z(g)$ responds to g_ℓ. So (4.93) tells us we can neglect the fluctuations of $\hat{\alpha}_\ell$ and $\hat{D}_\ell$ provided the effect of each locus is broadly consistent against distinct genetic backgrounds. This is typically the same kind of condition which is required for the infinitesimal model to hold [11].

Derivation of Claim 4.I.3 We now derive Claim 4.I.3. We will proceed through two claims.

Claim 4.I.4. *For any function f on $\{0,1,2\}$, the quadratic variation of $\mathbf{E}_{\mathbf{P}_t}[f(g_\ell)]$ is given by*

$$\mathrm{d}\,\langle\mathbf{E}_{\mathbf{P}}[f(g_\ell)]\rangle_t = 4\mathbf{Var}_{\mathbf{P}_t}\left[\mathbf{E}_{\mathbf{P}_t}\left[f(g_\ell) \;\middle|\; G_{\ell,(1)}\right]\right]\mathrm{d}t$$

Remark 15. *In classical quantitative genetics terminology, the right-hand side is called the variance contributed to $f(g_\ell)$ by an allelic substitution at locus ℓ.*

Derivation. We write

$$\mathbf{E}_{\mathbf{P}_t}[f(g_\ell)] = (P_t^\ell)^2 f(2) + 2P_t^\ell(1-P_t^\ell)f(1) + (1-P_t^\ell)^2 f(0)$$

From there, we compute with Itô's formula that the martingale part of $\mathrm{d}\,\langle\mathbf{E}_{\mathbf{P}_t}[f(g_\ell)]\rangle_t$ is

$$\Big(2P_t^\ell f(2) + 2(1-P_t^\ell)f(1) - 2(1-P_t^\ell)f(0))\Big)\sqrt{P_t^\ell(1-P_t^\ell)}\mathrm{d}B_t^\ell$$

It follows

$$\mathrm{d}\,\langle\mathbf{E}_{\mathbf{P}}[f(g_\ell)]\rangle_t = \Big(2P_t^\ell f(2) + 2(1-P_t^\ell)f(1) - 2(1-P_t^\ell)f(0))\Big)^2 P_t^\ell(1-P_t^\ell)\mathrm{d}t$$

This can be rewritten

$$\begin{aligned}\mathrm{d}\,\langle\mathbf{E}_{\mathbf{P}}[f(g_\ell)]\rangle_t =& 4\left(\Big(P_t^\ell f(2) + (1-P_t^\ell)f(1)\Big) - \Big(P_t^\ell f(1) + (1-P_t^\ell)f(0)\Big)\right)^2 P_t^\ell(1-P_t^\ell)\mathrm{d}t \\ =& 4\left(\mathbf{E}_{\mathbf{P}_t}[f(g_\ell) \mid G_{\ell,(1)} = 1] - \mathbf{E}_{\mathbf{P}_t}[f(g_\ell) \mid G_{\ell,(1)} = 0]\right)^2 P_t^\ell(1-P_t^\ell)\mathrm{d}t\end{aligned}$$

On the other hand, the variance of $\mathbf{E}_{\mathbf{P}_t}[f(g_\ell)\ |G_{\ell,(1)}]$ is the variance of the random variable

$$\left(\mathbf{E}_{\mathbf{P}_t}[f(g_\ell) \mid G_{\ell,(1)} = 1] - \mathbf{E}_{\mathbf{P}_t}[f(g_\ell) \mid G_{\ell,(1)} = 0]\right) G_{\ell,(1)} + \mathbf{E}_{\mathbf{P}_t}[f(g_\ell) \mid G_{\ell,(1)} = 0]$$

which is equal to

$$\left(\mathbf{E}_{\mathbf{P}_t}[f(g_\ell)\ |G_{\ell,(1)} = 1] - \mathbf{E}_{\mathbf{P}_t}[f(g_\ell)\ |G_{\ell,(1)} = 0]\right)^2 P_t^\ell(1-P_t^\ell).$$

This yields the result. □

We extend the previous result to the situation where f is a function of $(g_\ell)_{\ell\in A}$ for some subset $A \subseteq [L]$.

Claim 4.I.5. *Consider a subset* $A \subseteq [L]$*. Write* $g_A = (g_\ell)_{\ell\in A} \in \{0,1,2\}^A$*. For any function* f *on* $\{0,1,2\}^A$*, the quadratic variation of* $\mathbf{E}_{\mathbf{P}_t}[f(g_A)]$ *is given by*

$$\mathrm{d}\left\langle \mathbf{E}_{\mathbf{P}}[f(g_A)]\right\rangle_t = \sum_{\ell\in A} 4\mathbf{Var}_{\mathbf{P}_t}\left[\mathbf{E}_{\mathbf{P}_t}\left[f(g) \mid G_{\ell,(1)}\right]\right]\mathrm{d}t$$

Proof. We write

$$\mathbf{E}_{\mathbf{P}_t}[f(g_A)] = \sum_{\hat{g}\in\{0,1,2\}^A} f(\hat{g}) \prod_{\ell\in A}\left((P_t^\ell)^2\mathbb{1}_{[\hat{g}_\ell=2]} + 2P_t^\ell(1-P_t^\ell)\mathbb{1}_{[\hat{g}_\ell=1]} + (1-P_t^\ell)^2\mathbb{1}_{[\hat{g}_\ell=0]}\right)$$

Because the $(B^\ell)_{\ell\in[L]}$ are independent, the martingale part of $\mathrm{d}\mathbf{E}_{\mathbf{P}_t}[f(g_A)]$ is given by

$$\begin{aligned}\sum_{\ell\in A}\sum_{\hat{g}\in\{0,1,2\}^A} f(\hat{g}) \prod_{\ell'\in A\setminus\{\ell\}} &\left((P_t^{\ell'})^2\mathbb{1}_{[\hat{g}_{\ell'}=2]} + 2P_t^{\ell'}(1-P_t^{\ell'})\mathbb{1}_{[\hat{g}_{\ell'}=1]} + (1-P_t^{\ell'})^2\mathbb{1}_{[\hat{g}_{\ell'}=0]}\right)\\ &\times\left(2P_t^\ell\mathbb{1}_{[\hat{g}_\ell=2]} + 2(1-P_t^\ell)\mathbb{1}_{[\hat{g}_\ell=1]} - 2(1-P_t^\ell)\mathbb{1}_{[\hat{g}_\ell=0]}\right)\sqrt{P_t^\ell(1-P_t^\ell)}\mathrm{d}B_t^\ell\end{aligned}$$

This can be rewritten

$$\sum_{\ell\in A}\mathbf{E}_{\mathbf{P}_t}\left[f(g_A)\left(2P_t^\ell\mathbb{1}_{[g_\ell=2]} + 2(1-P_t^\ell)\mathbb{1}_{[g_\ell=1]} - 2(1-P_t^\ell)\mathbb{1}_{[g_\ell=0]}\right)\right]\sqrt{P_t^\ell(1-P_t^\ell)}\mathrm{d}B_t^\ell$$

The same computations as in the previous derivation yield the result. ☐

Derivation of Claim 4.I.3. We write

$$\begin{aligned}\alpha_\ell(t) =& \frac{1}{2}\mathbf{E}_{\mathbf{P}_t}\left[Z(g_1,\dots,g_{\ell-1},2,g_{\ell+1},\dots,g_L) - Z(g_1,\dots,g_{\ell-1},0,g_{\ell+1},\dots,g_L)\right]\\ D_\ell(t) =& \frac{1}{2}\mathbf{E}_{\mathbf{P}_t}\Big[2Z(g_1,\dots,g_{\ell-1},1,g_{\ell+1},\dots,g_L) - Z(g_1,\dots,g_{\ell-1},2,g_{\ell+1},\dots,g_L)\\ &\qquad - Z(g_1,\dots,g_{\ell-1},0,g_{\ell+1},\dots,g_L)\Big]\end{aligned}$$

Applying claim 4.I.5 with $A = [L]\setminus\{\ell\}$ we get the result. ☐

4.J A small note on integral computations

Computing Δ^* in the fixed-point equation (4.35) and σ^2 with (4.52) requires computing integrals of the form

$$\mathcal{I}_{a,b,c,d} := \frac{1}{Beta(a,b)}\int_0^1 x^{a-1}(1-x)^{b-1}e^{cx+dx(1-x)}\mathrm{d}x$$

with $a, b > 0$ possibly small and $c, d \in \mathbb{R}$ and $Beta$ is the beta function (added here for mathematical convenience). For want of a better method, we compute these integrals as

$$\begin{aligned}\mathcal{I}_{a,b,c,d} &\simeq \sum_{k\in[k_{max}]}\frac{d^k}{k!}\int_0^1 x^{a+k-1}(1-x)^{b+k-1}e^{cx}\mathrm{d}x\\ &= \sum_{k\in[k_{max}]}\frac{d^k}{k!}\frac{(a)_k(b)_k}{(a+b)_{2k}}\,{}_1F_1(a+k;a+b+2k;c)\end{aligned}$$

where k_{max} is as large as computationally possible, ${}_1F_1$ is the confluent hypergeometric function, and $(a)_k := a(a+1)\dots(a+k-1)$ is the rising factorial, with $(a)_0 = 1$. We were satisfied with the result, but would enthusiastically accept any suggestion for a more efficient method of computation.

Chapter 5

Open problems

I have highlighted in Figure 1.3 a number of conjectures. Yet there are three open problems for which I cannot even conjecture what the answer should be.

5.1 A Fleming-Viot-Wright-Fisher-McKean-Vlasov system ?

In Figure 1.3, the paths to the polygenic limit involve two possible distinct diffusion processes based on the trait's and the gene's eye-views, respectively. The problem with the first approach is that it summarizes the genetic architecture with a single parameter, the segregation variance, losing much information in the process. The problem with the second approach is that we cannot reasonably expect the diffusion approximation of the Wright-Fisher diffusion to hold simultaneously across all loci. It would be much better to only use the diffusion approximation on one typical locus.

There are remarkable deterministic models which jointly keep track of the evolution of a gene in a quantitative genetics background [127, 206]. In particular, [206] considers a major-effect biallelic locus (with alleles A and a) on a quantitative genetics background, the effect size α of which is of the same order as the genetic variance in the trait σ^2. The model is based on two PDEs which model the evolution of n^a, n^A, where $n^a(t,z)$ is the number of organisms carrying the a allele with trait value z. The segregation variance of the infinitesimal model V_S is a fixed parameter of the model. They describe the evolution of the system in the regime where V_S is very small, and prove in particular a separation of timescale, the dynamics of the gene being much slower than the dynamics of the trait z. Their model even allows for spatial structure, environmental heterogeneity and demography.

The two missing ingredients with respect to our roadmap are the stochasticity (both of the quantitative trait and of the gene) and the feedback between the law of the frequency of A, and the segregation variance V_S. One would hope to find a model according to which

- The trait distribution evolves like a Fleming-Viot model [156]
- A typical locus evolves like a Wright-Fisher diffusion
- The evolution of the trait distribution depends on the law of the typical locus, McKean-Vlasov style, and the evolution of the typical locus depends on the trait distribution.

Is this possible ?

5.2 What happens when LD is non-negligible from the gene's eye-view ?

When does LD cease to be negligible ? If we are to believe the results from the literature in quantitative genetics and statistical genetics [106, 70], the phase transition is expected when selection is of the same order as the rate of recombination. Our simulations indicate an interaction between N, L and ω_e^{-2} (see Figure 4.1), which we discussed in Section 4.F.2.

As discussed in Chapter 4, when LD is neglected we expect the typical locus to evolve as

$$\mathrm{d}P_t = \left(\alpha L \tilde{s}^* + \frac{\alpha^2}{\omega^2}\left(P_t - \frac{1}{2}\right)\right) P_t(1-P_t)\mathrm{d}t + \left(\mu^{(+)}(1-P_t) - \mu^{(-)}P_t\right)\mathrm{d}t + \sqrt{\frac{1}{2N}P_t(1-P_t)}\mathrm{d}B_t$$

where μ is the mutation rates, α the additive effect, ω^{-2} the strength of selection, N the effective population size and s^* the bias-correcting selection coefficient. Here, with respect to (4.28), we rescaled the time by $2N$ and defined $\tilde{s}^* := s^*/(2N)$ with the bias-correcting coefficient from Section 4.E.4.

At the phase transition where LD ceases to be negligible, we discussed in Section 4.F.2 the tools to predict the general architecture [45, 101, 70] and macroscopic observables [106, 207]. A recent preprint [105] suggests that at the phase transition, the system can be well described by replacing the additive effect at locus ℓ with some effective coefficient $\hat{\alpha}_\ell$. This deserves further investigation: how should the Wright-Fisher diffusion for one typical locus be modified at the phase transition ? The pessimistic scenario would be to add on the right-hand side a $\mathrm{d}D_t$ term for some process $(D_t)_{t\geq 0}$ which cannot be easily described (one could imagine an infinite-dimensional Ornstein-Uhlenbeck process, as occurs in the fluctuations of mean-field approximations, see [157]). An optimistic scenario would be to imagine that the parameters α, ω, μ or N can be replaced with "effective coefficients" $\hat{\alpha}, \hat{\omega}, \hat{\mu}, \hat{N}$ which account for LD. These effective coefficients would possibly be random and fluctuating, but characterizing them would yield important insights, because of how they could be interpreted. I would like to argue that there are reasons to expect each of these coefficients to be potentially impacted by LD.

- $\hat{\alpha}$ could account for the effective **additive value of the genetic block** in which the locus is. If the genome is segmented into many small blocks within which recombination is very rare, for which only two haplotypes segregate at the same time, then all alleles within a given haplotype will have the same effective additive value. Following [106], we expect selection to create negative LD, meaning the extreme values of α would be buffered by negative LD. We would therefore expect $|\hat{\alpha}| \leq |\alpha|$, as in [105].

- $\hat{\omega}^{-2}$ would account for genome-wide loose LD, following Bulmer [106]. Our simulations suggest (Figure 4.1) that LD affects the distance to the optimum Δ_t and the genetic variance σ^2 differently: σ^2 seems to respond more. This would suggest that the bias-countering term $\frac{\alpha}{\bar{\alpha}}s^*$ remains unchanged by LD while the underdominant term $\frac{\alpha^2}{\omega^2}$ is decreased.

- $\hat{\mu}$ could be interpreted as an early hitch-hiking effect [100]. At first glance, it seems impossible for LD to impact the mutation rate μ, because LD acts through the selection term, which is proportional to $P_t(1-P_t)$. In particular, selection only acts on intermediate allele frequencies (P_t close to 1/2) whereas the effect of mutation is felt most at the borders $\{0, 1\}$. But when P_t is close to 1/2, we may expect that the corresponding allele exists on many different genetic backgrounds. On the other hand, when the allele is very rare $P_t \ll 1$, this is where we could expect the fluctuations due to LD to have the greatest importance. If a new mutation occurs on a highly advantageous background, it will have more chances of surviving the initial branching phase [208], but this advantage will be lost through recombination once P_t reaches intermediate frequencies. In fact, the stationary distribution obtained in equation (43) of [101] which accounts for LD, comprises an interaction term between mutation and selection.

- $\hat{N}$ would correspond to the phenomenon of **genetic draft** [209, 101]. This general term describes randomness in the trajectory of an allele due to selection on its fluctuating genetic background.

Can we hope to define such effective coefficients ? Could there be more than one phase transition, corresponding to different effective coefficients entering play ? We argued in Section 4.F.2 that the transition to genetic draft may occur when $2N \sim L\sqrt{|\bar{\theta}|}$ whereas the transition to the regime of effective additive effects $\hat{\alpha}$ from [105] is expected when $2N \sim \ln(L)(L\omega_e^2)^{-1}|\bar{\theta}|$. But our simulations were computationally limited, with L and θ barely spanning one order of magnitude: further work is needed to confirm this conjecture.

5.3 How degenerate can the joint distribution of effect sizes and mutation rates be ?

In Chapter 4, we assumed all allelic effects and mutation rates were of the same order. But as our simulations illustrated in Figure 4.56, the polygenic limit seems to hold even when the distribution of α has a heavy tail. We expect some mathematical difficulties to appear under this general assumption. For instance, recall the dynamics of Δ_t from Chapter 4.

$$\begin{aligned} \mathrm{d}\Delta_t = \frac{1}{\tau}\Bigg(&- \Delta_t \times \frac{1}{|\bar{\theta}|} \times \mathbb{E}^*\left[2(L\alpha)^2 P_t(1-P_t)\right] \quad + \quad \frac{1}{|\bar{\theta}|} \times \mathbb{E}^*\left[2L^2\alpha^3\left(P_t - \frac{1}{2}\right)P_t(1-P_t)\right] \\ &+ \frac{L\omega_e^2}{|\bar{\theta}|} \times \mathbb{E}^*\left[2L\alpha\left(\theta^+(1-P_t) - \theta^- P_t\right)\right]\Bigg)\mathrm{d}t \quad + \quad \frac{1}{\sqrt{\tau}} \times \sqrt{\frac{\omega_e^2}{|\bar{\theta}|}\mathbb{E}^*\left[(2L\alpha)^2 P_t(1-P_t)\right]\mathrm{d}B_t^\Delta} \end{aligned}$$

The third term of this equation involves the expectation of α^3, so it would require special handling. When $\mathbb{E}\left[\alpha^3\right] = \infty$, is it possible for $\mathbb{E}^*[\alpha^3 P_t(1-P_t)]$ to be well-defined, if the large values of α correspond to very small values of $P_t(1-P_t)$?

Another important regime to explore is the regime of strong mutational bias ($\theta^+ \ll \theta^-$). What is the genetic variance of the population $\mathbb{E}^*\left[\alpha^2 P_t(1-P_t)\right]$ in this regime if $\eta(2-\eta) \sim 1$?

5.4 Personal conclusion

From my personal experience, I find there is a form of hostility in the general population towards gene-based models of evolution, which I would impute to the insistence by science communicators on genetic directional selection, that is, situations where there is an allele that is ”good” on average and the alternative allele is ”bad” on average. Some people seem to view trait-based models as a refreshing alternative vision to gene-based models: I have heard some researchers claim that in the infinitesimal limit ($L \to \infty$), natural selection does not act on gene frequencies, which would imply that the ”true” action of natural selection, driving adaptation, cannot be captured from the gene's eye-view.

Here, we have shown how, unless we are in the ”ultra-weak” selection regime (see Chapter 4), we do expect the gene's eye-view to provide an accurate representation of the population. But this vision does not quite match the one of ”good” and ”bad” alleles. The gene perceives directional selection through the bias-correcting selection coefficient, and a eugenist trying to maximize fitness observing genomic data would therefore strive to fix the ”good” alleles (under positive selection). But this would lead to an overshoot of the selection optimum, and result in a decrease of fitness.

This gene's eye-view approach may be of particular interest in modelling the response of the genetic variance within a population to a change of environment and/or population size, the importance of which to the modern era need not be stated. Specifically, assuming

- most of the current environmental changes act on organisms' fitnesses through polygenic traits[1]
- the genetic architecture of these polygenic traits can be described with rough models of genetic architecture accomodating pleiotropy, and epistasis in these traits can be neglected
- the population structure can be neglected, at least locally [139]
- the diffusion approximation is acceptable (in particular, the noise due to genetic drift must be well-described by $\sqrt{P_t^\ell(1-P_t^\ell)}\mathrm{d}B_t^\ell$, which precludes situations where a single family has so many offspring, they make up a significant proportion of the whole population)

then the polygenic limit may provide an accurate model to determine the fate of the population. The debate of whether the genetic load of a population is best described as purifying selection or polygenic traits under stabilizing selection has not yet been resolved [201].

---

[1] This excludes, for instance, insecticide resistance which usually evolves through very few mutations[210]

List of Figures

List of Tables

Bibliography

[1] Michael Lynch and Bruce Walsh. *Genetics and analysis of quantitative traits.* Sinauer, Sunderland, Ma, 1998.

[2] Bruce Walsh and Michael Lynch. *Evolution and Selection of Quantitative Traits.* Oxford University Press, 07 2018.

[3] L. Gidney. Earliest archaeological evidence of the ancon mutation in sheep from leicester, uk. *International Journal of Osteoarchaeology*, 17(3):318–321, 2007.

[4] L. M. Cook, B. S. Grant, I. J. Saccheri, and J. Mallet. Selective bird predation on the peppered moth: the last experiment of michael majerus. *Biology Letters*, 8(4):609–612, 2012.

[5] Lyudmila N. Trut. Early canid domestication: The farm-fox experiment: Foxes bred for tamability in a 40-year experiment exhibit remarkable transformations that suggest an interplay between behavioral genetics and development. *American Scientist*, 87(2):160–169, 1999.

[6] Snigdhadip Dey, Stephen R. Proulx, and Henrique Teotónio. Adaptation to temporally fluctuating environments by the evolution of maternal effects. *PLOS Biology*, 14(2):1–29, 02 2016.

[7] Christelle Leung, Daphné Grulois, Leandro Quadrana, and Luis-Miguel Chevin. Phenotypic plasticity evolves at multiple biological levels in response to environmental predictability in a long-term experiment with a halotolerant microalga. *PLOS Biology*, 21(3):1–18, 03 2023.

[8] Hubertus JE Beaumont, Jenna Gallie, Christian Kost, Gayle C Ferguson, and Paul B Rainey. Experimental evolution of bet hedging. *Nature*, 462(7269):90–93, 2009.

[9] Walter Frank Raphael Weldon. Ii. on certain correlated variations in *Carcinus mænas. Proceedings of the Royal Society of London*, 54(326-330):318–329, 1894.

[10] Hermon C. Bumpus. The elimination of the unfit as illustrated by the introduced sparrow, *Passer domesticus. Biological Lectures from the Marine Biological Laboratory of Wood's Holl*, pages 209–226, 1899.

[11] N H Barton, A M Etheridge, and A Véber. The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology*, 118:50–73, 12 2017.

[12] Francis Galton. Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15:246–263, 1886.

[13] KARL PEARSON. The law of ancestral heredity. *Biometrika*, 2(2):211–228, 02 1903.

[14] M. G. Bulmer. The effect of selection on genetic variability. *The American Naturalist*, 105(943):201–211, 1971.

[15] Jaleal S. Sanjak, Julia Sidorenko, Matthew R. Robinson, Kevin R. Thornton, and Peter M. Visscher. Evidence of directional and stabilizing selection in contemporary humans. *Proceedings of the National Academy of Sciences*, 115(1):151–156, 2018.

[16] Juha Merilä and Ben C. Sheldon. Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. *Heredity*, 83(2):103–109, Aug 1999.

[17] Anne Charmantier, Dany Garant, and Loeske E. B. Kruuk. *Quantitative Genetics in the Wild*. Oxford University Press, 04 2014.

[18] Trudy F.C Mackay and Richard F Lyman. *Drosophila* bristles and the nature of quantitative genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1459):1513–1527, 2005.

[19] John R. Stinchcombe and John K. Kelly. Measuring natural selection on the transcriptome. *New Phytologist*, n/a(n/a), 2025.

[20] Philip D Gingerich. *Rates of evolution: a quantitative synthesis*. Cambridge University Press, 2019.

[21] Chase D Brownstein, Daniel J MacGuigan, Daemin Kim, Oliver Orr, Liandong Yang, Solomon R David, Brian Kreiser, and Thomas J Near. The genomic signatures of evolutionary stasis. *Evolution*, 78(5):821–834, 03 2024.

[22] Professor Weldon. Report of the committee, consisting of mr. galton (chairman), mr. f. darwin, professor macalister, professor meldola, professor poulton, and professor weldon, "for conducting statistical inquiries into the measurable characteristics of plants and animals." part i. "an attempt to measure the death-rate due to the selective destruction of carcinus moenas with respect to a particular dimension". *Proceedings of the Royal Society of London*, 57:360–379, 1894.

[23] Alan Robertson. The effect of selection against extreme deviants based on deviation or on homozygosis. *Journal of Genetics*, 54:236–248, 1956.

[24] Russell Lande and Stevan J. Arnold. The measurement of selection on correlated characters. *Evolution*, 37(6):1210–1226, 1983.

[25] Joel G. Kingsolver and Sarah E. Diamond. Phenotypic selection in natural populations: What limits directional selection? *The American Naturalist*, 177(3):346–357, 2011. PMID: 21460543.

[26] Jason Bertram and Zahra Shafiei. Strong amplification of quantitative genetic variation under a balance between mutation and fluctuating stabilizing selection. *bioRxiv*, 2025.

[27] Joseph D Coolon, C Joel McManus, Kraig R Stevenson, Brenton R Graveley, and Patricia J Wittkopp. Tempo and mode of regulatory evolution in drosophila. *Genome research*, 24(5):797–808, 2014.

[28] Sarah A. Signor and Sergey V. Nuzhdin. The evolution of gene expression in *cis* and *trans*. *Trends in Genetics*, 34(7):532–544, Jul 2018.

[29] CARMEN AMADOR, AURORA GARCÍA-DORADO, DIEGO BERSABÉ, and CARLOS LÓPEZ-FANJUL. Regeneration of the variance of metric traits by spontaneous mutation in a drosophila population. *Genetics Research*, 92(2):91–102, 2010.

[30] R Lande. Natural selection and random genetic drift in phenotypic evolution. *Evolution*, pages 314–334, 1976.

[31] S. N. Ethier and Thomas G. Kurtz. Fleming–viot processes in population genetics. *SIAM Journal on Control and Optimization*, 31(2):345–386, 1993.

[32] Joseph T. Chang. Recent common ancestors of all present-day individuals. *Advances in Applied Probability*, 31(4):1002–1026, 1999.

[33] BERNARD DERRIDA, SUSANNA C. MANRUBIA, and DAMIÁN H. ZANETTE. On the genealogy of a population of biparental individuals. *Journal of Theoretical Biology*, 203(3):303–315, 2000.

[34] Nicholas H Barton and Alison M Etheridge. The relation between reproductive value and genetic contribution. *Genetics*, 188(4):953–973, 08 2011.

[35] Douglas LT Rohde, Steve Olson, and Joseph T Chang. Modelling the recent common ancestry of all living humans. *Nature*, 431(7008):562–566, 2004.

[36] Michael Lynch and William G. Hill. Phenotypic evolution by neutral mutation. *Evolution*, 40(5):915–935, 1986.

[37] O. Tenaillon. The utility of fisher's geometric model in evolutionary genetics. *Annual Review of Ecology, Evolution, and Systematics*, 45(Volume 45, 2014):179–201, 2014.

[38] N.H. Barton and A.M. Etheridge. Establishment in a new habitat by polygenic adaptation. *Theoretical Population Biology*, 122:110–127, 2018. Paul Joyce.

[39] Marguerite A. Butler and Aaron A. King. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *The American Naturalist*, 164(6):683–695, 2004. PMID: 29641928.

[40] Dean C. Adams. Comparing evolutionary rates for different phenotypic traits on a phylogeny using likelihood. *Systematic Biology*, 62(2):181–192, 11 2012.

[41] Natalie Cooper, Gavin H. Thomas, Chris Venditti, Andrew Meade, and Rob P. Freckleton. A cautionary note on the use of ornstein uhlenbeck models in macroevolutionary studies. *Biological Journal of the Linnean Society*, 118(1):64–77, 03 2016.

[42] Russell Lande. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution*, 33(1Part2):402–416, 1979.

[43] Erik I Svensson. Phenotypic selection in natural populations: what have we learned in 40 years? *Evolution*, 77(7):1493–1504, 04 2023.

[44] Michael Turelli. Commentary: Fisher's infinitesimal model: A story for the ages. *Theoretical Population Biology*, 118:46–49, 2017.

[45] M Turelli and N H Barton. Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? *Genetics*, 138(3):913–941, 11 1994.

[46] Vincent Calvez, Jimmy Garnier, and Florian Patout. Asymptotic analysis of a quantitative genetics model with nonlinear integral operator. *Journal de l'École polytechnique-Mathématiques*, 6:537–579, 2019.

[47] Florian Patout. The cauchy problem for the infinitesimal model in the regime of small variance. *Analysis & PDE*, 16(6):1289–1350, 2023.

[48] Jessica Guerand, Matthieu Hillairet, and Sepideh Mirrahimi. A moment-based approach for the analysis of the infinitesimal model in the regime of small variance. *Kinetic and Related Models*, 18(3):389–425, 2025.

[49] Léonard Dekens. Evolutionary dynamics of complex traits in sexual populations in a heterogeneous environment: how normal? *Journal of Mathematical Biology*, 84(3):15, Feb 2022.

[50] W B Provine. *The origins of theoretical population genetics.* University of Chicago Press, 2001.

[51] Hugo de Vries. *Intracellular Pangenesis.* The open court publishing co, 1910.

[52] William Feller et al. Diffusion processes in genetics. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, 1951.

[53] Motoo Kimura. Diffusion models in population genetics. *Journal of Applied Probability*, 1(2):177–232, 1964.

[54] Tianzhu Xiong, Shreeharsha Tarikere, Neil Rosser, Xueyan Li, Masaya Yago, and James Mallet. A polygenic explanation for haldane's rule in butterflies. *Proceedings of the National Academy of Sciences*, 120(44):e2300959120, 2023.

[55] Susan Strome, Needhi Bhalla, Rohinton Kamakaka, Upasna Sharma, and William Sullivan. Clarifying mendelian vs non-mendelian inheritance. *Genetics*, 227(3):iyae078, 05 2024.

[56] Gustavo V Barroso and Julien Y Dutheil. The landscape of nucleotide diversity in *drosophila* melanogaster is shaped by mutation rate variation. *Peer Community Journal*, 3, 2023.

[57] Patrick Billingsley. *Convergence of probability measures.* John Wiley & Sons, 1968.

[58] Servet Martínez. A probabilistic analysis of a discrete-time evolution in recombination. *Advances in Applied Mathematics*, 91:115–136, 2017.

[59] Ellen Baake, Michael Baake, and Majid Salamat. The general recombination equation in continuous time and its solution. *Discrete and Continuous Dynamical Systems*, 36(1):63–95, 2016.

[60] Ellen Baake and Michael Baake. Haldane linearisation done right: Solving the nonlinear recombination equation the easy way. *Discrete and Continuous Dynamical Systems*, 36(12):6645–6656, 2016.

[61] P Caputo and A Sinclair. Entropy production in nonlinear recombination models. *Bernoulli*, 24(4B):3246 – 3282, 2018.

[62] Matthias Steinrücken, YX Rachel Wang, and Yun S Song. An explicit transition density expansion for a multi-allelic wright–fisher diffusion with general diploid selection. *Theoretical population biology*, 83:1–14, 2013.

[63] S. N. Ethier. A class of degenerate diffusion processes occurring in population genetics. *Communications on Pure and Applied Mathematics*, 29(5):483–493, 1976.

[64] Tomoko Ohta and Motoo Kimura. Linkage disequilibrium due to random genetic drift. *Genetical Research*, 13(1):47–55, 1969.

[65] Tomoko Ohta and Motoo Kimura. Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. *Genetics*, 63(1):229–238, 09 1969.

[66] Robert C. Griffiths, Paul A. Jenkins, and Sabin Lessard. A coalescent dual process for a wright-fisher diffusion with recombination and its application to haplotype partitioning. *Theoretical Population Biology*, 112:126–138, 2016.

[67] Reinhard Bürger. Multilocus population-genetic theory. *Theoretical Population Biology*, 133:40–48, 2020. Fifty years of Theoretical Population Biology.

[68] R. Bürger. *The Mathematical Theory of Selection, Recombination, and Mutation.* Wiley Series in Mathematical & Computational Biology. Wiley, 2000.

[69] LP Chaintron and A Diez. Propagation of chaos: A review of models, methods and applications. i. models and methods. *Kinetic and Related Models*, 15(6):895–1015, 2022.

[70] V Dichio, H-L Zeng, and E Aurell. Statistical genetics in and out of quasi-linkage equilibrium. *Reports on Progress in Physics*, 86(5):052601, apr 2023.

[71] S. Shahshahani. *A new mathematical framework for the study of linkage and selection.* Memoirs of the American Mathematical Society ; number 211. American Mathematical Society, Providence, 1st ed. edition, 1979. Bibliography: pages 34.

[72] Erik Aurell, Magnus Ekeberg, and Timo Koski. On a multilocus wright-fisher model with mutation and a svirezhev-shahshahani gradient-like selection dynamics, 2019.

[73] S Wright. The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences*, 23(6):307–320, 1937.

[74] Thierry Huillet. On wright–fisher diffusion and its relatives. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(11):P11006, nov 2007.

[75] Paul A. Jenkins and Dario Spanò. Exact simulation of the Wright–Fisher diffusion. *The Annals of Applied Probability*, 27(3):1478 – 1509, 2017.

[76] Paul A. Jenkins, Jere Koskela, Victor M. Rivero, Jaromir Sant, Dario Spanò, and Ivana Valentić. Excursion theory for the Wright–Fisher diffusion. *Electronic Journal of Probability*, 30(none):1 – 31, 2025.

[77] Paul A. Jenkins. The mutual arrangement of wright-fisher diffusion path measures and its impact on parameter estimation, 2024.

[78] Pietro Caputo, Cyril Labbé, and Hubert Lacoin. Cutoff phenomenon in nonlinear recombinations, 2024.

[79] R A Fisher. The correlation between relatives on the supposition of mendelian inheritance. *Transaction Royal Society of Edinburgh*, 52:399–433, 1918.

[80] Rasmus Nielsen, Scott Williamson, Yuseob Kim, Melissa J Hubisz, Andrew G Clark, and Carlos Bustamante. Genomic scans for selective sweeps using snp data. *Genome research*, 15(11):1566–1575, 2005.

[81] Ilse Höllinger, Benjamin Wölfl, and Joachim Hermisson. A theory of oligogenic adaptation of a quantitative trait. *Genetics*, 225(2):iyad139, 08 2023.

[82] Motoo Kimura. A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proceedings of the National Academy of Sciences of the United States of America*, 54(3):731–736, 1965.

[83] Guy Sella and Nicholas H. Barton. Thinking about the evolution of complex traits in the era of genome-wide association studies. *Annual Review of Genomics and Human Genetics*, 20:461–493, 8 2019.

[84] Sewall Wright. Systems of mating. i. the biometric relations between parent and offspring. *Genetics*, 6(2):111–123, 03 1921.

[85] Evan A Boyle, Yang I Li, and Jonathan K Pritchard. An expanded view of complex traits: from polygenic to omnigenic. *Cell*, 169(7):1177–1186, 2017.

[86] N.H. Barton and Michael Turelli. Effects of genetic drift on variance components under a general model of epistasis. *Evolution*, 58(10):2111–2132, 10 2004.

[87] A.M. Castilloux and S. Lessard. The fundamental theorem of natural selection in ewens' sense (case of many loci). *Theoretical Population Biology*, 48(3):306–315, 1995.

[88] José M. Álvarez-Castro and Rosa M. Crujeiras. Orthogonal decomposition of the genetic variance for epistatic traits under linkage disequilibrium—applications to the analysis of bateson-dobzhansky-müller incompatibilities and sign epistasis. *Frontiers in Genetics*, Volume 10 - 2019, 2019.

[89] B. S. Weir and C. C. Cockerham. Two-locus theory in quantitative genetics., 1977.

[90] Nicholas H Barton, Alison M Etheridge, and Amandine Véber. The infinitesimal model with dominance. *Genetics*, 225(2):iyad133, 07 2023.

[91] VV Petrov. Sums of independent random variables. *Yu. V. Prokhorov. V. StatuleviCius (Eds.)*, 1972.

[92] A. Robertson and Conrad Hal Waddington. A theory of limits in artificial selection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 153(951):234–249, 1960.

[93] Bruce Walsh. Population-and quantitative-genetic models of selection limits. *Plant breeding reviews*, 24(1):177–226, 2004.

[94] Ivana Cvijović, Benjamin H Good, and Michael M Desai. The effect of strong purifying selection on genetic diversity. *Genetics*, 209(4):1235–1278, 05 2018.

[95] Hannah Götsch and Reinhard Bürger. Polygenic dynamics underlying the response of quantitative traits to directional selection. *Theoretical Population Biology*, 158:21–59, 2024.

[96] Sewall Wright. The analysis of variance and the correlations between relatives with respect to deviations from an optimum. *Journal of Genetics*, 30(2):243–256, Mar 1935.

[97] BDH Latter. Natural selection for an intermediate optimum. *Australian Journal of Biological Sciences*, 13(1):30–35, 1960.

[98] M. G. Bulmer. The genetic variability of polygenic characters under optimizing selection, mutation and drift. *Genetical Research*, 19(1):17–25, 1972.

[99] N H Barton and M Turelli. Natural and sexual selection on many loci. *Genetics*, 127(1):229–255, 01 1991.

[100] John Maynard Smith and John Haigh. The hitch-hiking effect of a favourable gene. *Genetical Research*, 23(1):23–35, 1974.

[101] Richard A. Neher and Boris I. Shraiman. Statistical genetics and evolution of quantitative traits. *Rev. Mod. Phys.*, 83:1283–1300, Nov 2011.

[102] H. Chau Nguyen, Riccardo Zecchina, and Johannes Berg. Inverse statistical problems: from the inverse ising problem to data science. *Advances in Physics*, 66(3):197–261, 2017.

[103] Tommaso Castellani and Andrea Cavagna. Spin-glass theory for pedestrians. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(05):P05012, may 2005.

[104] Richard A Neher, Marija Vucelja, Mark Mezard, and Boris I Shraiman. Emergence of clones in sexual populations. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(01):P01008, jan 2013.

[105] Sherif Negm and Carl Veller. The effect of long-range linkage disequilibrium on allele-frequency dynamics under stabilizing selection. *bioRxiv*, 2024.

[106] M. G. Bulmer. Linkage disequilibrium and genetic variability. *Genetical Research*, 23(3):281–289, 1974.

[107] E. Mauri, S. Cocco, and R. Monasson. Gaussian closure scheme in the quasi-linkage equilibrium regime of evolving genome populations. *Europhysics Letters*, 132(5):56001, jan 2021.

[108] Stevan J. Arnold, Reinhard Bürger, Paul A. Hohenlohe, Beverley C. Ajie, and Adam G. Jones. Understanding the evolution and stability of the g-matrix. *Evolution*, 62(10):2451–2461, 10 2008.

[109] Jae Hoon Sul, Lana S. Martin, and Eleazar Eskin. Population structure in genetic studies: Confounding factors and mixed models. *PLOS Genetics*, 14(12):1–22, 12 2018.

[110] Loïc Yengo, Sailaja Vedantam, and Eirini et al. Marouli. A saturated map of common genetic variants associated with human height. *Nature*, 610(7933):704–712, Oct 2022.

[111] Huwenbo Shi, Gleb Kichaev, and Bogdan Pasaniuc. Contrasting the genetic architecture of 30 complex traits from summary association data. *The American Journal of Human Genetics*, 99(1):139–153, Jul 2016.

[112] Hakhamanesh Mostafavi, Arbel Harpak, Ipsita Agarwal, Dalton Conley, Jonathan K Pritchard, and Molly Przeworski. Variable prediction accuracy of polygenic scores within an ancestry group. *eLife*, 9:e48376, jan 2020.

[113] Sile Hu, Lino A. F. Ferreira, Sinan Shi, Garrett Hellenthal, Jonathan Marchini, Daniel J. Lawson, and Simon R. Myers. Fine-scale population structure and widespread conservation of genetic effect sizes between human groups across traits. *Nature Genetics*, 57(2):379–389, Feb 2025.

[114] Duncan S. Palmer, Wei Zhou, Liam Abbott, Emilie M. Wigdor, Nikolas Baya, Claire Churchhouse, Cotton Seed, Tim Poterba, Daniel King, Masahiro Kanai, Alex Bloemendal, and Benjamin M. Neale. Analysis of genetic dominance in the uk biobank. *Science*, 379(6639):1341–1348, 2023.

[115] Guy Sella and Nicholas H Barton. Thinking about the evolution of complex traits in the era of genome-wide association studies. *Annual review of genomics and human genetics*, 20(1):461–493, 2019.

[116] Y B. Simons, K Bullaughey, R R. Hudson, and G Sella. A population genetic interpretation of gwas findings for human quantitative traits. *PLoS Biology*, 16:e2002985, 4 2017.

[117] Yuval B. Simons, Hakhamanesh Mostafavi, Courtney J. Smith, Jonathan K. Pritchard, and Guy Sella. Simple scaling laws control the genetic architectures of human complex traits. *bioRxiv*, 2022.

[118] Lawrence H. Uricchio, Hugo C. Kitano, Alexander Gusev, and Noah A. Zaitlen. An evolutionary compass for detecting signals of polygenic selection and mutational bias. *Evolution Letters*, 3(1):69–79, 02 2019.

[119] Sivan Yair and Graham Coop. Population differentiation of polygenic score predictions under stabilizing selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1852):20200416, 2022.

[120] E. Koch, N. J. Connally, N. Baya, M. P. Reeve, M. Daly, B. Neale, E. S. Lander, A. Bloemendal, and S. Sunyaev. Genetic association data are broadly consistent with stabilizing selection shaping human common diseases and traits. *bioRxiv*, 2024.

[121] Steven N Evans, David Steinsaltz, and Kenneth Wachter. *A mutation-selection model with recombination for general genotypes*, volume 222. American Mathematical Society, 2013.

[122] Brian Charlesworth. Stabilizing selection, purifying selection, and mutational bias in finite populations. *Genetics*, 194(4):955–971, 08 2013.

[123] Laura Katharine Hayward and Guy Sella. Polygenic adaptation after a sudden change in environment. *eLife*, 11:e66697, sep 2022.

[124] Laura Katharine Hayward and Guy Sella. Polygenic adaptation after a sudden change in environment. *Elife*, 11:e66697, 2022.

[125] William R Milligan, Laura K Hayward, and Guy Sella. When should adaptation arise from a polygenic response versus few large effect changes? *bioRxiv*, page 2025.05.15.654234, 2025. https://www.biorxiv.org/content/biorxiv/early/2025/05/17/2025.05.15.654234.full.pdf.

[126] Nathan W Anderson, Lloyd Kirk, Joshua G Schraiber, and Aaron P Ragsdale. A path integral approach for allele frequency dynamics under polygenic selection. *Genetics*, 229(1):1–63, 11 2024.

[127] Luis-Miguel Chevin and Frédéric Hospital. Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics*, 180(3):1645–1660, 11 2008.

[128] Harold P de Vladar and Nick Barton. Stability and response of polygenic traits to stabilizing selection and mutation. *Genetics*, 197(2):749–767, 06 2014.

[129] Kavita Jain and Wolfgang Stephan. Rapid adaptation of a polygenic trait after a sudden environmental shift. *Genetics*, 206(1):389–406, 05 2017.

[130] Kavita Jain and Wolfgang Stephan. Modes of rapid polygenic adaptation. *Molecular Biology and Evolution*, 34(12):3169–3175, 09 2017.

[131] Kavita Jain and Wolfgang Stephan. Rapid adaptation of a polygenic trait after a sudden environmental shift. *Genetics*, 206(1):389–406, 05 2017.

[132] Markus G. Stetter, Kevin Thornton, and Jeffrey Ross-Ibarra. Genetic architecture and selective sweeps after polygenic adaptation to distant trait optima. *PLOS Genetics*, 14(11):1–24, 11 2018.

[133] Mashaal Sohail, Robert M Maier, Andrea Ganna, Alex Bloemendal, Alicia R Martin, Michael C Turchin, Charleston WK Chiang, Joel Hirschhorn, Mark J Daly, Nick Patterson, Benjamin Neale, Iain Mathieson, David Reich, and Shamil R Sunyaev. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife*, 8:e39702, mar 2019.

[134] Jeremy J Berg, Arbel Harpak, Nasa Sinnott-Armstrong, Anja Moltke Joergensen, Hakhamanesh Mostafavi, Yair Field, Evan August Boyle, Xinjun Zhang, Fernando Racimo, Jonathan K Pritchard, et al. Reduced signal for polygenic adaptation of height in uk biobank. *Elife*, 8:e39725, 2019.

[135] Jonathan K. Pritchard, Joseph K. Pickrell, and Graham Coop. The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, 20(4):R208–R215, Feb 2010.

[136] Himani Sachdeva. Reproductive isolation via polygenic local adaptation in sub-divided populations: Effect of linkage disequilibria and drift. *PLOS Genetics*, 18(9):1–29, 09 2022.

[137] Enikö Szép, Himani Sachdeva, and Nicholas H. Barton. Polygenic local adaptation in metapopulations: A stochastic eco-evolutionary model. *Evolution*, 75(5):1030–1045, 05 2021.

[138] Jitka Polechová and Nicholas H. Barton. Limits to adaptation along environmental gradients. *Proceedings of the National Academy of Sciences*, 112(20):6401–6406, 2015.

[139] Jitka Polechová. Is the sky the limit? on the expansion threshold of a species' range. *PLOS Biology*, 16(6):1–18, 06 2018.

[140] Jitka Polechová. Is the sky the limit? on the expansion threshold of a species' range. *PLOS Biology*, 16(6):1–18, 06 2018.

[141] J. Arvid Ågren. Sewall wright's criticism of the gene's-eye view of evolution. *Evolution*, 75(10):2326–2334, 10 2021.

[142] Yuval Rabani, Yuri Rabinovich, and Alistair Sinclair. A computational view of population genetics. In *Proceedings of the twenty-seventh annual ACM symposium on Theory of computing*, pages 83–92, 1995.

[143] George R. Price. Selection and covariance. *Nature*, 227(5257):520–521, Aug 1970.

[144] Sabin Lessard. Fisher's fundamental theorem of natural selection revisited. *Theoretical Population Biology*, 52(2):119–136, 1997.

[145] Katarína Bod'ová, Enikő Szép, and Nicholas H Barton. Dynamic maximum entropy provides accurate approximation of structured population dynamics. *PLoS Computational Biology*, 17(12):e1009661, 2021.

[146] Todd L. Parsons and Peter L. Ralph. Large effects and the infinitesimal model. *Theoretical Population Biology*, 156:117–129, 2024.

[147] Michael Lynch. The lower bound to the evolution of mutation rates. *Genome Biology and Evolution*, 3:1107–1118, 08 2011.

[148] Kevin J. Dawson. Evolutionarily stable mutation rates. *Journal of Theoretical Biology*, 194(1):143–157, 1998.

[149] D W Stroock and S R S Varadhan. *Multidimensional diffusion processes*, volume 233. Springer Science & Business Media, 1997.

[150] Dario Trevisan. Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients. *Electronic Journal of Probability*, 21(none):1 – 41, 2016.

[151] Zimo Hao, Xicheng Zhang, Rongchan Zhu, and Xiangchan Zhu. Singular kinetic equations and applications. *The Annals of Probability*, 52(2):576 – 657, 2024.

[152] Norman L Johnson, Samuel Kotz, and Narayanaswamy Balakrishnan. *Continuous univariate distributions, volume 2*, volume 2. John wiley & sons, 1995.

[153] Yu I Lyubich. Basic concepts and theorems of the evolutionary genetics of free populations. *Russian Mathematical Surveys*, 26(5):51, oct 1971.

[154] Gary Shon Katzenberger. *Solutions of a stochastic differential equation forced onto a manifold by a large drift.* The University of Wisconsin-Madison, 1990.

[155] SN Ethier. A limit theorem for two-locus diffusion models in population genetics. *Journal of Applied Probability*, 16(2):402–408, 1979.

[156] WENDELL H. FLEMING and MICHEL VIOT. Some measure-valued markov processes in population genetics theory. *Indiana University Mathematics Journal*, 28(5):817–843, 1979.

[157] LP Chaintron and A Diez. Propagation of chaos: A review of models, methods and applications. ii. applications. *Kinetic and Related Models*, 15(6):1017–1173, 2022.

[158] Sylvie Roelly-Coppoletta. A criterion of convergence of measure-valued processes: application to measure branching processes. *Stochastics: An International Journal of Probability and Stochastic Processes*, 17(1-2):43–65, 1986.

[159] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

[160] Robert C. Griffiths, Paul A. Jenkins, and Sabin Lessard. A coalescent dual process for a wright-fisher diffusion with recombination and its application to haplotype partitioning. *Theoretical Population Biology*, 112:126–138, 2016.

[161] Stephen M Krone and Claudia Neuhauser. Ancestral processes with selection. *Theoretical Population Biology*, 51(3):210–237, 1997.

[162] Martina Favero, Henrik Hult, and Timo Koski. A dual process for the coupled wright–fisher diffusion. *Journal of Mathematical Biology*, 82(1):6, Jan 2021.

[163] Richard A. Neher, Taylor A. Kessinger, and Boris I. Shraiman. Coalescence and genetic diversity in sexual populations under selection. *Proceedings of the National Academy of Sciences*, 110(39):15836–15841, 2013.

[164] Michael Turelli. Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theoretical Population Biology*, 25(2):138–193, 1984.

[165] R. C. Griffiths and D. Spanó. Diffusion processes and coalescent trees. In N. H. Bingham and C. M.Editors Goldie, editors, *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman*, London Mathematical Society Lecture Note Series, page 358–379. Cambridge University Press, 2010.

[166] Günter Last and Mathew Penrose. *Lectures on the Poisson process*, volume 7. Cambridge University Press, 2018.

[167] Arthur Zwaenepoel, Himani Sachdeva, and Christelle Fraïsse. The genetic architecture of polygenic local adaptation and its role in shaping barriers to gene flow. *Genetics*, 228(3):iyae140, 08 2024.

[168] Parvathy Surendranadh and Himani Sachdeva. Effect of assortative mating and sexual selection on polygenic barriers to gene flow. *Evolution*, 79(7):1185–1198, 05 2025.

[169] Wei Hong, Shihu Li, and Wei Liu. Strong convergence rates in averaging principle for slow-fast mckean-vlasov spdes. *Journal of Differential Equations*, 316:94–135, 2022.

[170] Ilse Höllinger, Pleuni S Pennings, and Joachim Hermisson. Polygenic adaptation: From sweeps to subtle frequency shifts. *PLoS genetics*, 15(3):e1008035, 2019.

[171] Jeffrey P. Spence, Nasa Sinnott-Armstrong, Themistocles L. Assimes, and Jonathan K. Pritchard. A flexible modeling and inference framework for estimating variant effect sizes from gwas summary statistics, 2022.

[172] Archana Devi and Kavita Jain. Polygenic adaptation dynamics in large, finite populations, 2023.

[173] Eloïse Vanhoenacker, Linnéa Sandell, and Denis Roze. Stabilizing selection, mutational bias, and the evolution of sex*. *Evolution*, 72(9):1740–1758, 09 2018.

[174] Xu-Sheng Zhang and William G Hill. The anomalous effects of biased mutation revisited: Meanoptimum deviation greatly, however, and that the apparent directional selection thereby caused can be substantial. *Genetics*, 179, 2008.

[175] Harold P. de Vladar and Nick H. Barton. The statistical mechanics of a polygenic character under stabilizing selection, mutation and drift. *Journal of The Royal Society Interface*, 8(58):720–739, 2011.

[176] Katarína Bod'ová, Enikő Szép, and Nicholas H Barton. Dynamic maximum entropy provides accurate approximation of structured population dynamics. *PLOS Computational Biology*, 17(12):e1009661, 2021.

[177] Armita Nourmohammad, Stephan Schiffels, and Michael Lässig. Evolution of molecular phenotypes under stabilizing selection. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(01):P01012, jan 2013.

[178] A. García-Dorado, C. López-Fanjul, and A. Caballero. Properties of spontaneous mutations affecting quantitative traits. *Genetical Research*, 74(3):341–350, 1999.

[179] François Mallard, Luke Noble, Charles F Baer, and Henrique Teotónio. Variation in mutational (co)variances. *G3 Genes—Genomes—Genetics*, 13(2):jkac335, 12 2022.

[180] Paul M. Sharp, Elizabeth Bailes, Russell J. Grocock, John F. Peden, and R. Elizabeth Sockett. Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Research*, 33(4):1141–1153, 01 2005.

[181] Ruth Hershberg and Dmitri A. Petrov. Selection on codon bias. *Annual Review of Genetics*, 42(Volume 42, 2008):287–299, 2008.

[182] Kai Zeng and Brian Charlesworth. Estimating selection intensity on synonymous codon usage in a nonequilibrium population. *Genetics*, 183(2):651–662, 10 2009.

[183] Kai Zeng. A simple multiallele model and its application to identifying preferred–unpreferred codons using polymorphism data. *Molecular Biology and Evolution*, 27(6):1327–1337, 01 2010.

[184] Way Sung, Matthew S Ackerman, Samuel F Miller, Thomas G Doak, and Michael Lynch. Drift-barrier hypothesis and mutation-rate evolution. *Proceedings of the National Academy of Sciences*, 109(45):18488–18492, 2012.

[185] Romain Fournier, Zoi Tsangalidou, David Reich, and Pier Francesco Palamara. Haplotype-based inference of recent effective population size in modern and ancient dna samples. *Nature Communications*, 14(1):7945, Dec 2023.

[186] Peter Visscher. Quantitative genetics isn't dead yet. *International Conference on Quantitative Genetics 6*, March 2021. https://www.youtube.com/watch?v=tDcTjatiWjg.

[187] Laure Ségurel, Minyoung J. Wyman, and Molly Przeworski. Determinants of mutation rate variation in the human germline. *Annual Review of Genomics and Human Genetics*, 15(Volume 15, 2014):47–70, 2014.

[188] Thomas Nagylaki. Evolution under the multilocus levene model without epistasis. *Theoretical Population Biology*, 76(3):197–213, 2009.

[189] Kateryna D. Makova and Ross C. Hardison. The effects of chromatin organization on variation in mutation rates in the genome. *Nature Reviews Genetics*, 16(4):213–223, Apr 2015.

[190] Günter P. Wagner, Ginger Booth, and Homayoun Bagheri-Chaichian. A population genetic theory of canalization. *Evolution*, 51(2):329–347, 04 1997.

[191] Thomas Flatt. The evolutionary genetics of canalization. *The Quarterly Review of Biology*, 80(3):287–316, 2005. PMID: 16250465.

[192] Kerry Geiler-Samerotte, Federica M.O. Sartori, and Mark L. Siegal. Decanalizing thinking on genetic canalization. *Seminars in Cell & Developmental Biology*, 88:54–66, 2019. Canalization, a central concept in biology.

[193] Sara Via and Russell Lande. Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, 39(3):505–522, 1985.

[194] Dmitri A. Petrov, Elena R. Lozovskaya, and Daniel L. Hartl. High intrinsic rate of dna loss in drosophila. *Nature*, 384(6607):346–349, Nov 1996.

[195] D A Petrov, Y C Chao, E C Stephenson, and D L Hartl. Pseudogene evolution in drosophila suggests a high rate of dna loss. *Molecular Biology and Evolution*, 15(11):1562–1567, 11 1998.

[196] Josep M Comeron and Martin Kreitman. The correlation between intron length and recombination in drosophila: Dynamic equilibrium between mutational and selective forces. *Genetics*, 156(3):1175–1190, 11 2000.

[197] Susan E Ptak and Dmitri A Petrov. How intron splicing affects the deletion and insertion profile in drosophila melanogaster. *Genetics*, 162(3):1233–1244, 11 2002.

[198] Evgeny V. Leushkin, Georgii A. Bazykin, and Alexey S. Kondrashov. Strong mutational bias toward deletions in the drosophila melanogaster genome is compensated by selection. *Genome Biology and Evolution*, 5(3):514–524, 02 2013.

[199] Nathaniel B. Edelman and James Mallet. Prevalence and adaptive impact of introgression. *Annual Review of Genetics*, 55(Volume 55, 2021):265–283, 2021.

[200] Himani Sachdeva and Nicholas H Barton. Introgression of a block of genome under infinitesimal selection. *Genetics*, 209(4):1279–1303, 2018.

[201] Brian Charlesworth. Why we are not dead one hundred times over. *Evolution*, 67(11):3354–3361, 2013.

[202] Sean Meyn, Richard L. Tweedie, and Peter W. Glynn. *Ergodicity*, page 313–335. Cambridge Mathematical Library. Cambridge University Press, 2009.

[203] W. G. Hill and Alan Robertson. The effect of linkage on limits to artificial selection. *Genetical Research*, 8(3):269–294, 1966.

[204] N H Barton. Linkage and the limits to natural selection. *Genetics*, 140(2):821–841, 06 1995.

[205] Stuart A. Kauffman and Edward D. Weinberger. The nk model of rugged fitness landscapes and its application to maturation of the immune response. *Journal of Theoretical Biology*, 141(2):211–245, 1989.

[206] L. Dekens, S.P. Otto, and V. Calvez. The best of both worlds: Combining population genetic and quantitative genetic models. *Theoretical Population Biology*, 148:49–75, 2022.

[207] Enrique Santiago. Linkage and the maintenance of variation for quantitative traits by mutation–selection balance: an infinitesimal model. *Genetics Research*, 71(2):161–170, 1998.

[208] Felix Hermann, Adrián Gonzalez Casanova, Renato Soares dos Santos, András Tóbiás, and Anton Wakolbinger. From clonal interference to poissonian interacting trajectories, 2025.

[209] Richard A. Neher. Genetic draft, selective interference, and population genetics of rapid adaptation. *Annual Review of Ecology, Evolution, and Systematics*, 44(Volume 44, 2013):195–215, 2013.

[210] Virginie Courtier-Orgogozo. The loci of insect phenotypic evolution. *Current Opinion in Insect Science*, 60:101134, 2023.

[211] Mark S Pinsker. Information and information stability of random variables and processes. *Holden-Day*, 1964.

[212] Alexei A Fedotov, Peter Harremoës, and Flemming Topsoe. Refinements of pinsker's inequality. *IEEE Transactions on Information Theory*, 49(6):1491–1498, 2003.

[213] Paul Lévy. *Le mouvement brownien.* Gauthier-Villars, 1954.

[214] Henry J Landau and Lawrence A Shepp. On the supremum of a gaussian process. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 369–378, 1970.

Appendix A

More on the LD-Wright-Fisher diffusion

In Section A.1 we explain how the Quasi-Linkage Equilibrium approach is obtained from the LD-Wright-Fisher diffusion, and in Section A.2 we explain why we cannot LE to hold globally unless ρ is of order 2^L. We use the notations from Chapter 2.

A.1 Recovering the Quasi-Linkage Equilibrium approach

Here we discuss what approximations are necessary to recover the Quasi-Linkage Equilibrium approach of [101] from the LD-Wright-Fisher diffusion (2.3) as used in Chapter 2, which yields first-order corrections for the behavior of $\mathbf{X}_t$ to account for LD. Specifically, we want to recover the closed system of equations (36-37) from [101] for $((p^\ell(\mathbf{X}_t))_{\ell\in[L]}, (D^{\ell_1,\ell_2}(\mathbf{X}_t))_{\ell_1\neq\ell_2})$. We warn the reader that a rigorous proof is at present unattainable. We will follow [101] in writing

$$LW(g) = \sum_{\ell\in[L]} f_\ell g_\ell + \sum_{1\leq\ell_1<\ell_2\leq L} f_{\ell_1\ell_2} g_{\ell_1} g_{\ell_2}$$

for some parameters $(f_\ell)_{\ell\in[L]}, (f_{\ell_1,\ell_2})_{\ell_1,\ell_2\in[L]}$ with $f_{\ell,\ell} = 0$ for any $\ell \in [L]$. The goal is to obtain a closed equation for $(\mathbf{P}_t, \mathbf{D}_t) \equiv ((P_t^\ell)_{\ell\in[L]}, (D_t^{\ell_1,\ell_2})_{\ell_1\neq\ell_2})$ where for $\ell, \ell_1, \ell_2 \in [L]$ with $\ell_1 \neq \ell_2$

$$P_t^\ell := p^\ell(\mathbf{X}_t) \qquad D_t^{\ell,\ell} := P_t^\ell(1-P_t^\ell) \qquad D_t^{\ell_1,\ell_2} := D^{\ell_1,\ell_2}(\mathbf{X}_t)$$

We claim

Claim A.1.1 (Equations (36-37) of [101])**.** *Assume all terms of the form*

$$\chi_{\ell_1,\ell_2,\ell_3} := \mathbf{X}_t[(g_{\ell_1} - \mathbf{X}_t[g_{\ell_1}])(g_{\ell_2} - \mathbf{X}_t[g_{\ell_2}])(g_{\ell_3} - \mathbf{X}_t[g_{\ell_3}])]$$

for ℓ_1, ℓ_2, ℓ_3 distinct can be neglected, and that $D_t^{\ell_1,\ell_2}$ is small. Then we may find the following first-order equation for the dynamics for $\ell_1 \neq \ell_2$

$$\mathrm{d}P_t^\ell \simeq \left(\overline{\Theta}(P_t^\ell) + \tilde{S}^\ell(\mathbf{P}_t, \mathbf{D}_t)\right)\mathrm{d}t + \sqrt{P_t^\ell(1-P_t^\ell)}\mathrm{d}\hat{B}_t^\ell \tag{A.1}$$

$$\begin{aligned}\mathrm{d}D_t^{\ell_1,\ell_2} \simeq &- h^{\ell_1,\ell_2}(P_t^{\ell_1}, P_t^{\ell_2}) D_t^{\ell_1,\ell_2}\mathrm{d}t \\ &+ f_{\ell_1\ell_2} 4P_t^{\ell_1}(1-P_t^{\ell_1})P_t^{\ell_2}(1-P_t^{\ell_2})\mathrm{d}t \\ &+ \sqrt{P_t^{\ell_1}(1-P_t^{\ell_1})P_t^{\ell_2}(1-P_t^{\ell_2})}\mathrm{d}\tilde{B}_t^{\ell_1,\ell_2}\end{aligned} \tag{A.2}$$

where $\hat{B}^{\ell}, \tilde{B}^{\ell_1,\ell_2}$ are Brownian motions with quadratic variations such that

$$\mathrm{d}\left\langle P^{\ell_1}, P^{\ell_2}\right\rangle_t = D_t^{\ell_1,\ell_2}\mathrm{d}t$$

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\left\langle P^{\ell}, D^{\ell_1,\ell_2}\right\rangle_t\right| \ll 1 \qquad \left|\frac{\mathrm{d}}{\mathrm{d}t}\left\langle D^{\ell_1,\ell_2}, D^{\ell_3,\ell_4}\right\rangle_t\right| \ll 1$$

for any $\{\ell_1,\ell_2\} \neq \{\ell_3,\ell_4\}$. In (A.1) we defined for $\mathbf{p} \in [0,1]^L, \mathbf{d} \in [-1,1]^{L^2}$

$$\tilde{S}^{\ell}(\mathbf{p},\mathbf{d}) := \sum_{\ell' \in [L]} 2f_{\ell'}d^{\ell,\ell'} + \sum_{1\leq \ell_1<\ell_2\leq L} 2f_{\ell_1,\ell_2}\left(d^{\ell,\ell_1}(2p^{\ell_2}-1) + d^{\ell,\ell_2}(2p^{\ell_1}-1)\right) - \sum_{\ell'\neq\ell} 2f_{\ell,\ell'}d^{\ell,\ell'}(2p^{\ell}-1)$$

$$h^{\ell_1,\ell_2}(p_1,p_2) := 1 + \rho r_{\{\ell_1,\ell_2\}} + |\theta_{\ell_1}| + |\theta_{\ell_2}| + f_{\ell_1}2(2p_1-1) + f_{\ell_2}2(2p_2-1) - f_{\ell_1\ell_2}(2p_1-1)(2p_2-1)$$

Remark 16. *To recover (36-37) of [101], one should additionally assume $\rho r_{\{\ell_1,\ell_2\}}$ dominates all other terms in h^{ℓ_1,ℓ_2}, and account for a small typo in equation (37) of [101] (χ_{ij} should be on the left of the bracket).*

Remark 17. *The assumption that the third cumulants are negligible cannot be proved as straightforwardly as (2.61) from Proposition 2.3.13, for two reasons*

- *It requires controlling LD on subsets of size four (replacing $\{\ell_0\}$ by $\{\ell_1,\ell_2\}$ in (2.32) from Proposition 2.3.2)*
- *If we assume that W is of order 1 (as in Chapter 2) and that all $f_{\ell_1\ell_2}$ have the same order, then we necessarily have that $f_{\ell_1\ell_2}$ is of order L^{-1}. Therefore the first-order effect of selection on $D_t^{\{\ell_1,\ell_2\}}$ is $LS^{\{\ell_1,\ell_2\}}(\pi(\mathbf{X}_t)) = O(1/L)$. The Quasi-Linkage Equilibrium approach assumes this to be larger than the error of neglecting LD. This is a much stronger requirement than what was given in Section 2.3.4, where we wanted LD to be negligible with respect to $LS^{\ell}(\pi(\mathbf{X}_t))$, which is of order 1.*

In Chapter 4., we use the two following observations from (A.2) in [101] in the setting of stabilizing selection, where the directional fitness coefficient felt by a locus satisfies $f_\ell \sim 1$ (Section 4.E.4) and the epistatic fitness coefficient is $f_{\ell_1\ell_2} = -\alpha_{\ell_1}\alpha_{\ell_2}/(2\omega_e^2) \sim 1/(L\omega_e)^2$. Furthermore, in this chapter we have uniform cross-over rates $r_{\{\ell_1,\ell_2\}} = |\ell_1-\ell_2|/L$ and[1] $\rho \simeq N$.

- Let us neglect the Brownian term, and assume $\rho r_{\{\ell_1,\ell_2\}}$ dominates all other terms in (A.2). Let us assume $P_t^{\ell_1}$ evolves more slowly than $D_t^{\ell_1,\ell_2}$, then

$$D_t^{\ell_1,\ell_2} \sim \frac{f_{\ell_1\ell_2}}{\rho r_{\{\ell_1,\ell_2\}}} P_t^{\ell_1}(1-P_t^{\ell_1})P_t^{\ell_2}(1-P_t^{\ell_2}). \tag{A.3}$$

  Considering stabilizing selection ($f_{\ell_1\ell_2} = -\alpha_{\ell_1}\alpha_{\ell_2}/(2\omega_e^2)$), we define the total amount of two-loci LD as

$$D_t^{tot} := \sum_{\ell_1\neq\ell_2} \alpha_{\ell_1}\alpha_{\ell_2}D_t^{\ell_1,\ell_2}$$

---

[1] Recall the recombination rate is related to the population size and the probability of outcrossing $\rho^{(N)}$ with $\rho = N\rho^{(N)}$, see (1.4). In Chapter 4, outcrossing occurs with probability $\rho^{(N)} \sim 1$.

so that the genetic variance given by

$$\mathbf{Var}_{\mathbf{X}_t}[Z(g)] = \sum_{\ell \in [L]} \alpha_\ell^2 P_t^\ell (1 - P_t^\ell) + D_t^{tot}$$

We find from (A.3)

$$|D_t^{tot}| \sim \sum_{\ell_1 \neq \ell_2} \frac{1}{\rho r_{\{\ell_1,\ell_2\}} \omega_e^2} (\alpha_{\ell_1} \alpha_{\ell_2})^2 P_t^{\ell_1} (1 - P_t^{\ell_1}) P_t^{\ell_2} (1 - P_t^{\ell_2}).$$

If we take $\rho r_{\{\ell_1,\ell_2\}} = (2N)|\ell_1 - \ell_2|/L$ we find using a mean-field approximation (see (H3) in Chapter 4)

$$|D_t^{tot}| \sim \frac{L^2 \ln(L)}{2N\omega_e^2} \mathbb{E}\left[\alpha^2 P_t (1 - P_t)\right]^2.$$

where $P_t = P_t^{\ell_U}$ for ℓ_U an independent uniform variable on $[L]$, representing the typical locus. For LD to be negligible, we require $\sum_{\ell_1 \neq \ell_2} \alpha_{\ell_1} \alpha_{\ell_2} D_t^{\ell_1,\ell_2}$ to be negligible with respect to the genetic variance $\sigma_t^2 \simeq 2L\mathbb{E}\left[\alpha^2 P_t (1 - P_t)\right]$ (see (4.52)). We find from $\alpha \sim 1/L$

$$\begin{aligned}\frac{|D_t^{tot}|}{\sigma_t^2} &\sim \frac{L^2 \ln(L)}{2N\omega_e^2} \frac{\mathbb{E}\left[(1/L)^2 P_t (1 - P_t)\right]^2}{L\mathbb{E}\left[(1/L)^2 P_t (1 - P_t)\right]} \\ &\sim \frac{\ln(L)}{2NL\omega_e^2} \mathbb{E}\left[P_t (1 - P_t)\right].\end{aligned}$$

This yields from $\sigma_t^2 \sim \mathbb{E}\left[P_t (1 - P_t)\right]/L$

$$\frac{|D_t^{tot}|}{\sigma_t^2} \sim \frac{\ln(L)}{2N\omega_e^2} \sigma_t^2. \tag{A.4}$$

- If we neglect the epistatic fitness coefficients $(f_{\ell_1,\ell_2})_{\ell_1,\ell_2}$, assume $\rho r_{\{\ell_1,\ell_2\}}$ dominates all other terms and assume P_t evolves more slowly than D_t, then

$$\mathbb{V}\mathbf{ar}[D_t^{\ell_1,\ell_2} | P_t^{\ell_1}, P_t^{\ell_2}] \sim \frac{P_t^{\ell_1} (1 - P_t^{\ell_1}) P_t^{\ell_2} (1 - P_t^{\ell_2})}{(\rho r_{\{\ell_1,\ell_2\}})^2}$$

Plugging this in (A.1) and neglecting the epistatic fitness terms, we see that the contribution of the $(f_{\ell'})_{\ell' \neq \ell}$ in $\tilde{S}^\ell(\mathbf{P}_t, \mathbf{D}_t)$ is of order

$$\sum_{\ell' \in [L]} f_{\ell'}^2 \mathbb{V}\mathbf{ar}[D_t^{\ell,\ell'} | P_t^\ell P_t^{\ell'}] \sim \sum_{\ell' \in [L]} f_{\ell'}^2 \frac{P_t^\ell (1 - P_t^\ell) P_t^{\ell'} (1 - P_t^{\ell'})}{(\rho r_{\{\ell,\ell'\}})^2}.$$

This corresponds to equation (44) in [101]. For the effect of hitch-hiking randomness on the dynamics of P_t^ℓ to be negligible with respect to genetic drift, we require this quantity to be negligible with respect to $P_t^\ell (1 - P_t^\ell)$. When f_ℓ is typically of order 1, and $\rho r_{\{\ell_1,\ell_2\}} = 2N|\ell_1 - \ell_2|/L$, we find

$$\frac{\sum_{\ell' \in [L]} f_{\ell'}^2 \mathbb{V}\mathbf{ar}[D_t^{\ell,\ell'} | P_t^\ell P_t^{\ell'}]}{P_t^\ell (1 - P_t^\ell)} \sim \sum_{\ell' \in [L] \setminus \{\ell\}} f_{\ell'}^2 \left(\frac{L}{2N|\ell - \ell'|}\right)^2 P_t^{\ell'} (1 - P_t^{\ell'}). \tag{A.5}$$

Derivation of Claim A.1.1. First, from Corollary 2.3.5 we have

$$\mathrm{d}P_t^{\ell} = \left(\overline{\Theta}(P_t^{\ell}) + LS^{\ell}(\mathbf{X}_t)\right)\mathrm{d}t + \sqrt{P_t^{\ell}(1-P_t^{\ell})}\,\mathrm{d}\hat{B}_t^{\ell}$$

We write

$$LS^{\ell}(\mathbf{X}_t) = \sum_{\ell' \in [L]} f_{\ell'} \mathbf{Cov}_{\mathbf{X}_t}[\mathbb{1}_{[g_\ell=+1]}, g_{\ell'}] + \sum_{1 \le \ell_1 < \ell_2 \le L} f_{\ell_1 \ell_2} \mathbf{Cov}_{\mathbf{X}_t}\left[\mathbb{1}_{[g_\ell=+1]}, g_{\ell_1} g_{\ell_2}\right] \qquad \text{(A.6)}$$

Using $g_\ell = 2\mathbb{1}_{[g_\ell=+1]} - 1$ we write

$$\mathbf{Cov}_{\mathbf{X}_t}[\mathbb{1}_{[g_\ell=+1]}, g_{\ell'}] = 2D_t^{\ell,\ell'}.$$

Furthermore

$$\begin{aligned}\mathbf{Cov}_{\mathbf{X}_t}\left[\mathbb{1}_{[g_\ell=+1]}, g_{\ell_1}g_{\ell_2}\right] = \mathbf{Cov}_{\mathbf{X}_t}&\left[\mathbb{1}_{[g_\ell=+1]}, g_{\ell_1}\right]\mathbf{X}_t[g_{\ell_2}] \\ &+ \mathbf{Cov}_{\mathbf{X}_t}\left[\mathbb{1}_{[g_\ell=+1]}, (g_{\ell_2} - \mathbf{X}_t[g_{\ell_2}])\right]\mathbf{X}_t[g_{\ell_1}] \\ &\qquad + \mathbf{Cov}_{\mathbf{X}_t}\left[\mathbb{1}_{[g_\ell=+1]}, (g_{\ell_1} - \mathbf{X}_t[g_{\ell_1}])(g_{\ell_2} - \mathbf{X}_t[g_{\ell_2}])\right]\end{aligned}$$

which can be rewritten with $\mathbf{X}_t[\mathbb{1}_{[g_\ell]}] = 2P_t^{\ell} - 1$

$$\begin{aligned}\mathbf{Cov}_{\mathbf{X}_t}\left[\mathbb{1}_{[g_\ell=+1]}, g_{\ell_1}g_{\ell_2}\right] &= 2D_t^{\ell,\ell_1}(2P_t^{\ell_2} - 1) + 2D_t^{\ell,\ell_2}(2P_t^{\ell_1} - 1) \\ &\quad + \frac{1}{2}\mathbf{X}_t\left[(g_\ell - \mathbf{X}_t[g_\ell])(g_{\ell_1} - \mathbf{X}_t[g_{\ell_1}])(g_{\ell_2} - \mathbf{X}_t[g_{\ell_2}])\right]\end{aligned}$$

The last term is a third-order cumulant, which is assumed to be negligible when ℓ, ℓ_1, ℓ_2 are distinct. If $\ell = \ell_1$ we find from $(g_\ell)^2 = 1$

$$\begin{aligned}\mathbf{X}_t\left[(g_\ell - \mathbf{X}_t[g_\ell])^2(g_{\ell_2} - \mathbf{X}_t[g_{\ell_2}])\right] =& \mathbf{X}_t\left[(1 - 2\mathbf{X}_t[g_\ell]g_\ell + \mathbf{X}_t[g_\ell]^2)(g_{\ell_2} - \mathbf{X}_t[g_{\ell_2}])\right] \\ =& -4D_t^{\ell,\ell_2}(2P_t^{\ell} - 1)\end{aligned}$$

We thus obtain from (A.6)

$$\begin{aligned}LS^{\ell}(\mathbf{X}_t) \simeq \sum_{\ell' \in [L]} 2f_{\ell'}D_t^{\ell,\ell'} + \sum_{1 \le \ell_1 < \ell_2 \le L} 2f_{\ell_1,\ell_2}&\left(D_t^{\ell,\ell_1}(2P_t^{\ell_2} - 1) + D_t^{\ell,\ell_2}(2P_t^{\ell_1} - 1)\right) \\ &- \sum_{\ell' \ne \ell} 2f_{\ell,\ell'}D_t^{\ell,\ell'}(2P_t^{\ell} - 1)\end{aligned}$$

This yields (A.1).

We now turn to (A.2). Tedious computations based on Proposition 2.3.3 yield the dynamics of the second-order cumulants $D_t^{\ell_1,\ell_2}$

$$\mathrm{d}D_t^{\ell_1,\ell_2} = -(1 + \rho r_{\{\ell_1,\ell_2\}} + 2|\theta|)D_t^{\ell_1,\ell_2}\mathrm{d}t + LS^{\ell_1,\ell_2}(\mathbf{X}_t)\mathrm{d}t + \mathrm{d}M_t^{\ell_1,\ell_2} \qquad \text{(A.7)}$$

where

$$\begin{aligned}S^{\ell_1,\ell_2}(\mathbf{X}_t) :=& \mathbf{Cov}_{\mathbf{X}_t}\left[W(g), (g_{\ell_1} - \mathbf{X}_t[g_{\ell_1}])(g_{\ell_2} - \mathbf{X}_t[g_{\ell_2}])\right] \\ \mathrm{d}M_t^{\ell_1,\ell_2} :=& \sum_{\substack{\gamma^{(1)},\gamma^{(2)} \in \square_{[L]} \\ \gamma^{(1)} \ne \gamma^{(2)}}} \left(K_t^{\ell_1,\ell_2}(\gamma^{(1)}) - K_t^{\ell_1,\ell_2}(\gamma^{(2)})\right)\sqrt{X_t(\gamma^{(1)})X_t(\gamma^{(2)})}\,\mathrm{d}B_t(\gamma^{(1)},\gamma^{(2)})\end{aligned}$$

where

$$K_t^{\ell_1,\ell_2}(\gamma) := \mathbb{1}_{[\gamma_{\ell_1}=\gamma_{\ell_2}=+1]} - P_t^{\ell_1}\mathbb{1}_{[\gamma_{\ell_2}=+1]} - P_t^{\ell_2}\mathbb{1}_{[\gamma_{\ell_1}=+1]}.$$

Equation (A.7) is equivalent to equation (10) of [101], with genetic drift added. The first term $-D_t^{\ell_1,\ell_2}\mathrm{d}t$ comes from the quadratic variation between $P_t^{\ell_1}$ and $P_t^{\ell_2}$.

Simple computations show that when the cumulants of order 3 or greater are neglected, we have

$$\begin{aligned}LS^{\ell_1,\ell_2}(\mathbf{X}_t) \simeq f_{\ell_1}\mathbf{Cov}_{\mathbf{X}_t}\left[g_{\ell_1},(g_{\ell_1}-\mathbf{X}_t[g_{\ell_1}])(g_{\ell_2}-\mathbf{X}_t[g_{\ell_2}])\right]\\ + f_{\ell_2}\mathbf{Cov}_{\mathbf{X}_t}\left[g_{\ell_2},(g_{\ell_1}-\mathbf{X}_t[g_{\ell_1}])(g_{\ell_2}-\mathbf{X}_t[g_{\ell_2}])\right]\\ + f_{\ell_1\ell_2}\mathbf{Cov}_{\mathbf{X}_t}\left[g_{\ell_1}g_{\ell_2},(g_{\ell_1}-\mathbf{X}_t[g_{\ell_1}])(g_{\ell_2}-\mathbf{X}_t[g_{\ell_2}])\right]\end{aligned}$$

We then compute

$$\begin{aligned}\mathbf{Cov}_{\mathbf{X}_t}\left[g_{\ell_1},(g_{\ell_1}-\mathbf{X}_t[g_{\ell_1}])(g_{\ell_2}-\mathbf{X}_t[g_{\ell_2}])\right] &= \mathbf{X}_t\left[(g_{\ell_1}-\mathbf{X}_t[g_{\ell_1}])^2(g_{\ell_2}-\mathbf{X}_t[g_{\ell_2}])\right]\\ &= -2\mathbf{X}_t[g_{\ell_1}]D_t^{\ell_1,\ell_2}\end{aligned}$$

$$\begin{aligned}&\mathbf{Cov}_{\mathbf{X}_t}\left[g_{\ell_1}g_{\ell_2},(g_{\ell_1}-\mathbf{X}_t[g_{\ell_1}])(g_{\ell_2}-\mathbf{X}_t[g_{\ell_2}])\right]\\ &\qquad= \left(1-\mathbf{X}_t[g_{\ell_1}]^2\right)\left(1-\mathbf{X}_t[g_{\ell_2}]^2\right)+\mathbf{X}_t[g_{\ell_1}]\mathbf{X}_t[g_{\ell_2}]D_t^{\ell_1,\ell_2}\end{aligned}$$

We thus find

$$\begin{aligned}LS^{\ell_1,\ell_2}(\mathbf{X}_t) \simeq \left(-f_{\ell_1}2\mathbf{X}_t[g_{\ell_1}]-f_{\ell_2}2\mathbf{X}_t[g_{\ell_2}]+f_{\ell_1\ell_2}\mathbf{X}_t[g_{\ell_1}]\mathbf{X}_t[g_{\ell_2}]\right)D_t^{\ell_1,\ell_2}\\ + f_{\ell_1\ell_2}(1-\mathbf{X}_t[g_{\ell_1}]^2)(1-\mathbf{X}_t[g_{\ell_2}]^2).\end{aligned}$$

We thus find from (A.7)

$$\begin{aligned}&\mathrm{d}D_t^{\ell_1,\ell_2} \simeq\\ &-\left(1+\rho r_{\{\ell_1,\ell_2\}}+|\theta_{\ell_1}|+|\theta_{\ell_2}|+f_{\ell_1}2\mathbf{X}_t[g_{\ell_1}]+f_{\ell_2}2\mathbf{X}_t[g_{\ell_2}]-f_{\ell_1\ell_2}\mathbf{X}_t[g_{\ell_1}]\mathbf{X}_t[g_{\ell_2}]\right)D_t^{\ell_1,\ell_2}\mathrm{d}t\\ &\quad+f_{\ell_1\ell_2}(1-\mathbf{X}_t[g_{\ell_1}]^2)(1-\mathbf{X}_t[g_{\ell_2}]^2)\mathrm{d}t\\ &\quad+\sqrt{\mathbf{Var}_{\mathbf{X}_t}\left[\mathbb{1}_{[g_{\ell_1}=+1]}\mathbb{1}_{[g_{\ell_2}=+1]}-P_t^{\ell_1}\mathbb{1}_{[g_{\ell_2}=+1]}-P_t^{\ell_2}\mathbb{1}_{[g_{\ell_1}=+1]}\right]}\mathrm{d}\tilde{B}_t^{\ell_1,\ell_2}\end{aligned}$$

Using $\mathbf{X}_t[g_\ell]=2P_t^\ell-1$, we see that it only remains to handle the Brownian term. We find

$$\mathrm{d}\left\langle M\right\rangle_t^{\ell_1,\ell_2}=\mathbf{Var}_{\mathbf{X}_t}\left[\mathbb{1}_{[g_{\ell_1}=+1]}\mathbb{1}_{[g_{\ell_2}=+1]}-P_t^{\ell_1}\mathbb{1}_{[g_{\ell_2}=+1]}-P_t^{\ell_2}\mathbb{1}_{[g_{\ell_1}=+1]}\right]$$

which justifies writing

$$\mathrm{d}\left\langle M\right\rangle_t^{\ell_1,\ell_2}=\sqrt{\mathbf{Var}_{\mathbf{X}_t}\left[\mathbb{1}_{[g_{\ell_1}=+1]}\mathbb{1}_{[g_{\ell_2}=+1]}-P_t^{\ell_1}\mathbb{1}_{[g_{\ell_2}=+1]}-P_t^{\ell_2}\mathbb{1}_{[g_{\ell_1}=+1]}\right]}\mathrm{d}\tilde{B}_t^{\ell_1,\ell_2}$$

for some Brownian motion $\tilde{B}^{\ell_1,\ell_2}$. We now compute

$$\begin{aligned}&\mathbf{Var}_{\mathbf{X}_t}\left[\mathbb{1}_{[g_{\ell_1}=+1]}\mathbb{1}_{[g_{\ell_2}=+1]}-P_t^{\ell_1}\mathbb{1}_{[g_{\ell_2}=+1]}-P_t^{\ell_2}\mathbb{1}_{[g_{\ell_1}=+1]}\right]\\ &\quad\simeq\mathbf{Var}_{\mathbf{X}_t}\left[\mathbb{1}_{[g_{\ell_1}=+1]}\mathbb{1}_{[g_{\ell_2}=+1]}\right]+\mathbf{Var}_{\mathbf{X}_t}[P_t^{\ell_1}\mathbb{1}_{[g_{\ell_2}=+1]}+P_t^{\ell_2}\mathbb{1}_{[g_{\ell_1}=+1]}]\\ &\qquad-2\mathbf{Cov}_{\mathbf{X}_t}\left[\mathbb{1}_{[g_{\ell_1}=+1]}\mathbb{1}_{[g_{\ell_2}=+1]},P_t^{\ell_1}\mathbb{1}_{[g_{\ell_2}=+1]}+P_t^{\ell_2}\mathbb{1}_{[g_{\ell_1}=+1]}\right]\end{aligned}$$

One may then compute

$$\begin{aligned}\mathbf{Var}_{\mathbf{X}_t}\left[\mathbb{1}_{[g_{\ell_1}=+1]}\mathbb{1}_{[g_{\ell_2}=+1]}\right]=&(P_t^{\ell_1}P_t^{\ell_2}+D_t^{\ell_1,\ell_2})(1-P_t^{\ell_1}P_t^{\ell_2}-D_t^{\ell_1,\ell_2})\\ \mathbf{Var}_{\mathbf{X}_t}[P_t^{\ell_1}\mathbb{1}_{[g_{\ell_2}=+1]}+P_t^{\ell_2}\mathbb{1}_{[g_{\ell_1}=+1]}]=&P_t^{\ell_1}P_t^{\ell_2}\left(P_t^{\ell_1}(1-P_t^{\ell_2})+P_t^{\ell_2}(1-P_t^{\ell_1})+2D_t^{\ell_1,\ell_2}\right)\\ \mathbf{Cov}_{\mathbf{X}_t}\left[\mathbb{1}_{[g_{\ell_1}=+1]}\mathbb{1}_{[g_{\ell_2}=+1]},P_t^{\ell_1}\mathbb{1}_{[g_{\ell_2}=+1]}\right]=&P_t^{\ell_1}\left(P_t^{\ell_1}P_t^{\ell_2}-D_t^{\ell_1,\ell_2}\right)(1-P_t^{\ell_2})\end{aligned}$$

We then find

$$\begin{aligned}\mathbf{Var}_{\mathbf{X}_t} &\left[\mathbb{1}_{[g_{\ell_1}=+1]}\mathbb{1}_{[g_{\ell_2}=+1]} - P_t^{\ell_1}\mathbb{1}_{[g_{\ell_2}=+1]} - P_t^{\ell_2}\mathbb{1}_{[g_{\ell_1}=+1]}\right] \\ &= P_t^{\ell_1}P_t^{\ell_2}\left(1-P_t^{\ell_1}\right)\left(1-P_t^{\ell_2}\right) + D_t^{\ell_1,\ell_2}\left(1+2(P_t^{\ell_1}+P_t^{\ell_2}) - 4P_t^{\ell_1}P_t^{\ell_2} - D_t^{\ell_1,\ell_2}\right)\end{aligned}$$

Neglecting the terms of order $D_t^{\ell_1,\ell_2}$, which are assumed to be much smaller than $P_t^{\ell_1}(1-P_t^{\ell_1})$ and $P_t^{\ell_2}(1-P_t^{\ell_2})$, we get the result.

To conclude, we compute the quadratic variation between $P_t^{\ell_1}$ and $D_t^{\ell_2,\ell_3}$ to be

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle P^{\ell_1}, D^{\ell_2,\ell_3}\rangle_t = \mathbf{Cov}_{\mathbf{X}_t}\left[\mathbb{1}_{[g_{\ell_1}=+1]}, \mathbb{1}_{[g_{\ell_2}=+1]}\mathbb{1}_{[g_{\ell_3}=+1]} - \mathbb{1}_{[g_{\ell_2}=+1]}P_t^{\ell_3} - \mathbb{1}_{[g_{\ell_3}=+1]}P_t^{\ell_2}\right]$$

If we neglect the cumulants of order 3 or greater, this can be found to be negligible. A similar computation yields the result for $\mathrm{d}\langle D^{\ell_1,\ell_2}, D^{\ell_3,\ell_4}\rangle_t$ when $\{\ell_1,\ell_2\} \neq \{\ell_3,\ell_4\}$. □

A.2 Itô's formula applied to the relative entropy

In Chapter 2, we obtained the polygenic limit from the LD-Wright-Fisher diffusion (2.3) under the assumption that the number of loci L and the recombination rate ρ scale as $\rho r^{**} \gg L^2\ln(L)$ for some parameter r^{**} which depends on the recombination measure ν.

Here we briefly justify why when $\rho \ll 2^L$ in (2.3), we cannot expect that $\mathbf{X}_t$ is globally at LE. This means that we managed to reach the polygenic limit despite the fact that recombination cannot entirely eliminate LD, only first-order LD.

The discussion relies on the **relative entropy**

$$H(\mathbf{X}_t) := \sum_{\gamma\in\square_{[L]}} X_t(\gamma)\ln\left(\frac{X_t(\gamma)}{\pi(\mathbf{X}_t)(\gamma)}\right) \tag{A.8}$$

where $\pi(\mathbf{X}_t)$ is the LE projection of $\mathbf{X}_t$, see (2.26).

The relative entropy has two remarkable features that make it particularly convenient to study dynamical systems on probability spaces. The first is that it dominates the distance to LE, with the famous Pinsker inequality [211, 212]

$$||\mathbf{X}_t - \pi(\mathbf{X}_t)||_1 \leq \sqrt{\frac{1}{2}H(\mathbf{X}_t)}$$

where $||\cdot||_1$ is the total variation norm. We have that $H(\mathbf{X}_t)=0$ iff $\mathbf{X}_t = \pi(\mathbf{X}_t)$.

The second important feature of relative entropy is that in mixing systems, it is often possible to define an **entropy production bound**. Specifically, consider $\mathbf{X}_t$ evolving under the deterministic equation $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{X}_t = R(\mathbf{X}_t)$ for some operator R. An entropy production bound is a constant $\delta > 0$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t}H(\mathbf{X}_t) \leq -\delta H(\mathbf{X}_t)$$

It is known that the recombinator R under a typical recombination measure ν such as uniform crossing-over (see Section 2.1.2) has **entropy production bound** of order $1/L$ (see [61]).

The entropy production bound lets us claim that the strength of recombination acting to bring the population to LE in the LD-Wright-Fisher diffusion (2.3) is of order ρ/L.

Here we will show that genetic drift has a typical contribution to $H(\mathbf{X}_t)$ of order 2^L. Specifically, we consider $(\mathbf{X}_t)_{t\geq 0}$ satisfying the pure-genetic-drift equation

$$\mathrm{d}\mathbf{X}_t = \Sigma(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t$$

with $\Sigma, \mathbf{B}$ as in (2.3). We define the following quantities

$$\begin{aligned}
H_t :=& H(\mathbf{X}_t) \\
V_t :=& \sum_{\gamma \in \square_{[L]}} X_t(\gamma) \left(\ln(X_t(\gamma)) - H_t\right)^2 \\
\mathscr{A}_t :=& \left\{\gamma \in \square_{[L]},\ X_t(\gamma) \in (0,1)\right\} \\
\mathscr{B}_t :=& \left\{\ell \in [L],\ p^{\ell}(\mathbf{X}_t) \in (0,1)\right\}
\end{aligned}$$

With words, H_t is the **relative entropy**, V_t is the **varentropy**, $\mathscr{A}_t$ is the set of **segregating genotypes** and $\mathscr{B}_t$ is the set of **segregating loci**. In particular, if all genotypes are present in the population at time t we have $\#\mathscr{A}_t = 2^L$ and $\#\mathscr{B}_t = L$. We will show

Theorem A.2.1. *We have*

$$\mathrm{d}H_t = \mathrm{d}M_t + \frac{1}{2}\left(\#\mathscr{A}_t - 2\#\mathscr{B}_t\right)\mathrm{d}t$$

where M_t is a continuous martingale with quadratic variation

$$\mathrm{d}\left\langle M\right\rangle_t = 2V_t \mathrm{d}t.$$

The proof is a straightforward application of the following generalized version of Itô's formula. Define

$$\psi : \left\{ \begin{array}{ll} \mathbb{R}_+ & \longrightarrow \mathbb{R} \\ x & \longmapsto x\ln(x) \end{array} \right.$$

with the convention $0 \cdot \infty = 0$.

We will proceed through two lemmas. The first is an extension of Itô's formula.

Lemma A.2.2. *Consider a neutral Wright-Fisher diffusion $(Y_t)_{t\geq 0}$*

$$\mathrm{d}Y_t = \sqrt{Y_t(1-Y_t)}\mathrm{d}B_t$$

with B a Brownian motion. Then for $t > 0$ the stochastic integral

$$\int_0^1 \left(1 + \ln(Y_u)\right)\mathrm{d}Y_u$$

with $\ln(0) = -\infty$ *is well defined and we have*

$$\psi(Y_t) - \psi(Y_0) = \int_0^1 \left(1 + \ln(Y_u)\right)\mathrm{d}Y_u + \frac{1}{2}\int_0^1 \mathbb{1}_{[Y_u \in (0,1)]}(1 - Y_u)\mathrm{d}u \tag{A.9}$$

Remark 18. *This result can be extended to the case*

$$\mathrm{d}Y_t = b_t \mathrm{d}t + \sqrt{Y_t(1-Y_t)}\mathrm{d}B_t$$

when $(b_t)_{t\geq 0}$ is a continuous process satisfying $b_t \geq -CY_t$ for some constant C, which would be applicable to $X_t(\gamma)$ for $\mathbf{X}_t$ the general LD-Wright-Fisher diffusion (1.5)*.*

Proof. For small $\varepsilon > 0$ we define

$$\psi_\varepsilon : \left\{ \begin{array}{ll} [0,1] & \longrightarrow \mathbb{R}_+ \\ x & \longmapsto \frac{\psi(\varepsilon + (1-\varepsilon)x)}{1-\varepsilon} \end{array} \right.$$

Then ψ_ε is $\mathcal{C}^2$ on $[0,1]$ and

$$\psi_\varepsilon'(x) = 1 + \ln(\varepsilon + (1-\varepsilon)x) \qquad ; \qquad \psi''(x) = \frac{1-\varepsilon}{\varepsilon + (1-\varepsilon)x}$$

Itô's formula therefore yields

$$\mathrm{d}\psi_\varepsilon(Y_t) = \psi_\varepsilon'(Y_t)\mathrm{d}Y_t + \frac{1}{2}\psi_\varepsilon''(Y_t)\mathrm{d}\left\langle Y\right\rangle_t$$

In particular

$$\psi_\varepsilon(Y_t) - \psi_\varepsilon(Y_0) = \int_0^t \psi_\varepsilon'(Y_u)\mathrm{d}Y_u + \frac{1}{2}\int_0^t \frac{1-\varepsilon}{\varepsilon + (1-\varepsilon)Y_u} Y_u(1-Y_u)\mathrm{d}u \tag{A.10}$$

As $\varepsilon \to 0$ we have from the continuity of ψ

$$\psi_\varepsilon(Y_t) - \psi_\varepsilon(Y_0) \longrightarrow \psi(Y_t) - \psi(Y_0) \tag{A.11}$$

Furthermore, since

$$\varepsilon \longmapsto \frac{1-\varepsilon}{\varepsilon + (1-\varepsilon)x} x(1-x) = \frac{1-\varepsilon}{\varepsilon(1-x) + x} x(1-x)$$

is decreasing in ε, we have by monotone convergence

$$\frac{1}{2}\int_0^t \frac{1-\varepsilon}{\varepsilon + (1-\varepsilon)Y_u} Y_u(1-Y_u)\mathrm{d}u \longrightarrow \frac{1}{2}\int_0^t \mathbb{1}_{[Y_u>0]}(1-Y_u)\mathrm{d}u \tag{A.12}$$

Since $\mathbb{1}_{[Y_u>0]}(1-Y_u) = \mathbb{1}_{[Y_u\in(0,1)]}(1-Y_u)$, we get the last integral of (A.9).

The remaining term of (A.10) is

$$\int_0^t \psi_\varepsilon'(Y_u)\mathrm{d}Y_u = \int_0^t \psi_\varepsilon'(Y_u)\sqrt{Y_u(1-Y_u)}\mathrm{d}B_u \tag{A.13}$$

Observe that the following limit integral is well-defined

$$\int_0^t \psi'(Y_u)\sqrt{Y_u(1-Y_u)}\mathrm{d}B_u = \int_0^t (1+\ln(Y_u))\sqrt{Y_u(1-Y_u)}\mathrm{d}B_u \tag{A.14}$$

since the integrand is bounded and continuous.

To check that the martingale part of (A.13) does converge to this, write

$$\begin{aligned}
\mathbb{E}&\left[\left(\int_0^t \psi_\varepsilon'(Y_u)\sqrt{Y_u(1-Y_u)}\mathrm{d}B_u - \int_0^t \psi'(Y_u)\sqrt{Y_u(1-Y_u)}\mathrm{d}B_u\right)^2\right] \\
&= \mathbb{E}\left[\int_0^t \left(\psi_\varepsilon'(Y_u) - \psi'(Y_u)\right)^2 Y_u(1-Y_u)\mathrm{d}u\right] \\
&= \mathbb{E}\left[\int_0^t \left(1+\ln(\varepsilon+(1-\varepsilon)Y_u) - 1 - \ln(Y_u)\right)^2 Y_u(1-Y_u)\mathrm{d}u\right] \\
&= \mathbb{E}\left[\int_0^t \left(\ln(\varepsilon+(1-\varepsilon)Y_u) - \ln(Y_u)\right)^2 Y_u(1-Y_u)\mathrm{d}u\right]
\end{aligned}$$

We then conclude using the following dichotomy for a small enough ε

$$\sup_{x\in[0,\sqrt{\varepsilon}]} \left(\ln(\varepsilon+(1-\varepsilon)x) - \ln(x)\right)^2 x(1-x) \leq \sup_{x\in[0,\sqrt{\varepsilon}]} 2(\ln(\varepsilon+(1-\varepsilon)\sqrt{\varepsilon})^2 + \ln(\sqrt{\varepsilon})^2)\sqrt{\varepsilon}$$

$$\sup_{x\in[\sqrt{\varepsilon},1]} \left(\ln(\varepsilon+(1-\varepsilon)x) - \ln(x)\right)^2 x(1-x) \leq \sup_{x\in[\sqrt{\varepsilon},1]} \ln\left(\frac{\varepsilon}{x}+1-\varepsilon\right)^2 \leq \ln\left(\frac{\varepsilon}{\sqrt{\varepsilon}}+1-\varepsilon\right)^2$$

In the first line we used $(a+b)^2 \leq 2a^2+2b^2$. In the third line we used that $x \mapsto \ln\left(\frac{\varepsilon}{x}+1-\varepsilon\right)^2$ is decreasing in x on $[\sqrt{\varepsilon},1]$. □

The second Lemma connects ψ to the entropy process $(H_t)_{t\geq 0}$.

Lemma A.2.3. *We have*

$$H_t = \sum_{\gamma\in\square_{[L]}} \psi(X_t(\gamma)) - \sum_{\ell\in[L]} \psi(p^\ell(\mathbf{X}_t)) + \psi(1-p^\ell(\mathbf{X}_t)) \tag{A.15}$$

Proof of Lemma A.2.3. We have from (A.8)

$$H_t = \sum_{\gamma\in\square_{[L]}} X_t(\gamma)\ln(X_t(\gamma)) - X_t(\gamma)\ln(\pi(\mathbf{X}_t)(\gamma))$$

which we rewrite

$$H_t = \sum_{\gamma\in\square_{[L]}} \psi(X_t(\gamma)) - X_t(\gamma)\ln(\pi(\mathbf{X}_t)(\gamma)).$$

Recall from the definition of π (2.26)

$$\pi(\mathbf{X}_t) = \bigotimes_{\ell\in[L]} (\delta_{-1}(1-p^\ell(\mathbf{X}_t)) + \delta_{+1}p^\ell(\mathbf{X}_t))$$

This yields

$$\ln(\pi(\mathbf{X}_t)(\gamma)) = \sum_{\ell\in[L]} \mathbb{1}_{[\gamma_\ell=-1]}\ln(1-p^\ell(\mathbf{X}_t)) + \mathbb{1}_{[\gamma_\ell=+1]}\ln(p^\ell(\mathbf{X}_t))$$

We thus find

$$\begin{aligned}\sum_{\gamma\in\square_{[L]}} X_t(\gamma)\ln(\pi(\mathbf{X}_t)(\gamma)) &= \sum_{\ell\in[L]} (1-p^\ell(\mathbf{X}_t))\ln(1-p^\ell(\mathbf{X}_t)) + p^\ell(\mathbf{X}_t)\ln(p^\ell(\mathbf{X}_t))\\ &= \sum_{\ell\in[L]} \psi(1-p^\ell(\mathbf{X}_t)) + \psi(p^\ell(\mathbf{X}_t))\end{aligned}$$

This yields the result ☐

Proof of Theorem A.2.1. Because of Corollary 2.3.5, we know that

$$\mathrm{d}p^\ell(\mathbf{X}_t) = \sqrt{p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))}\mathrm{d}B_t^\ell$$

and this implies

$$\mathrm{d}(1-p^\ell(\mathbf{X}_t)) = -\sqrt{p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))}\mathrm{d}B_t^\ell$$

so that we may apply Lemma A.2.2 to get

$$\begin{aligned}&\mathrm{d}\psi(p^\ell(\mathbf{X}_t)) + \mathrm{d}\psi(1-p^\ell(\mathbf{X}_t))\\ &\quad= \Big(\ln(p^\ell(\mathbf{X}_t)) - \ln(1-p^\ell(\mathbf{X}_t))\Big)\sqrt{p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))}\mathrm{d}B_t^\ell + \mathbb{1}_{[p^\ell(\mathbf{X}_t)\in(0,1)]}\mathrm{d}t\end{aligned} \tag{A.16}$$

Similarly, one can show that for any $\gamma\in\square_{[L]}$, $X_t(\gamma)$ is also a neutral Wright-Fisher diffusion

$$\mathrm{d}X_t(\gamma) = \sqrt{X_t(\gamma)(1-X_t(\gamma))}\mathrm{d}\tilde{B}_t^\gamma$$

where

$$\mathrm{d}\tilde{B}_t^\gamma := \sum_{\hat{\gamma}\in\square_{[L]}\setminus\{\gamma\}} \sqrt{\frac{X_t(\hat{\gamma})}{1-X_t(\gamma)}}\mathrm{d}B_t(\gamma,\hat{\gamma}).$$

It follows from Lemma A.2.2

$$\mathrm{d}\psi(X_t(\gamma)) = (1+\ln(X_t(\gamma))) \sum_{\gamma'\in\square_{[L]}\setminus\{\gamma\}} \sqrt{X_t(\gamma)X_t(\gamma')}\mathrm{d}B_t(\gamma,\gamma') + \frac{1}{2}\mathbb{1}_{[X_t(\gamma)\in(0,1)]}\mathrm{d}t$$

It follows

$$\begin{aligned}
&\sum_{\gamma\in\square_{[L]}} \mathrm{d}\psi(X_t(\gamma)) \\
&= \sum_{\substack{\gamma,\gamma'\in\square_{[L]}\\ \gamma<\gamma'}} \big(\ln(X_t(\gamma)) - \ln(X_t(\gamma'))\big)\sqrt{X_t(\gamma)X_t(\gamma')}\mathrm{d}B_t(\gamma,\gamma') + \frac{1}{2}\sum_{\gamma\in\square_{[L]}} \mathbb{1}_{[X_t(\gamma)\in(0,1)]}\mathrm{d}t \\
&\qquad = \sum_{\substack{\gamma,\gamma'\in\square_{[L]}\\ \gamma<\gamma'}} \big(\ln(X_t(\gamma)) - \ln(X_t(\gamma'))\big)\sqrt{X_t(\gamma)X_t(\gamma')}\mathrm{d}B_t(\gamma,\gamma') + \frac{1}{2}\#\mathscr{A}_t\mathrm{d}t
\end{aligned}$$

where in the first line we used $B(\gamma,\gamma') = -B(\gamma',\gamma)$. Combining this with (A.16) into (A.15) we get

$$\begin{aligned}
\mathrm{d}H_t = \sum_{\substack{\gamma,\gamma'\in\square_{[L]}\\ \gamma<\gamma'}} &\Bigg(\big(\ln(X_t(\gamma)) - \ln(X_t(\gamma'))\big)\sqrt{X_t(\gamma)X_t(\gamma')}\mathrm{d}B_t(\gamma,\gamma') \\
&+ \sum_{\ell\in[L]} \Big(\ln(p^\ell(\mathbf{X}_t)) - \ln(1-p^\ell(\mathbf{X}_t))\Big)\sqrt{p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))}\mathrm{d}B_t^\ell\Bigg) \\
&\qquad + \left(\frac{1}{2}\#\mathscr{A}_t - \sum_{\ell\in[L]} \mathbb{1}_{[p^\ell(\mathbf{X}_t)\in(0,1)]}\right)\mathrm{d}t
\end{aligned}$$

The final term is equal to $\frac{1}{2}(\#\mathscr{A}_t - \#\mathscr{B}_t)$ as claimed. It remains to check that the quadratic variation of the martingale part M_t is indeed V_t, where

$$\begin{aligned}
\mathrm{d}M_t := \sum_{\substack{\gamma,\gamma'\in\square_{[L]}\\ \gamma<\gamma'}} &\Bigg(\big(\ln(X_t(\gamma)) - \ln(X_t(\gamma'))\big)\sqrt{X_t(\gamma)X_t(\gamma')}\mathrm{d}B_t(\gamma,\gamma') \\
&+ \sum_{\ell\in[L]} \Big(\ln(p^\ell(\mathbf{X}_t)) - \ln(1-p^\ell(\mathbf{X}_t))\Big)\sqrt{p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))}\mathrm{d}B_t^\ell\Bigg)
\end{aligned}$$

Simple computations using the definition of B^ℓ in Proposition 2.3.3 will show

$$\begin{aligned}
&\sum_{\ell\in[L]} \Big(\ln(p^\ell(\mathbf{X}_t)) - \ln(1-p^\ell(\mathbf{X}_t))\Big)\sqrt{p^\ell(\mathbf{X}_t)(1-p^\ell(\mathbf{X}_t))}\mathrm{d}B_t^\ell \\
&\qquad = \sum_{\ell\in[L]} \sum_{\substack{\gamma,\gamma'\in\square_{[L]}\\ \gamma_\ell=+1,\gamma'_\ell=-1}} \Big(\ln(p^\ell(\mathbf{X}_t)) - \ln(1-p^\ell(\mathbf{X}_t))\Big)\sqrt{X_t(\gamma)X_t(\gamma')}\mathrm{d}B_t(\gamma,\gamma')
\end{aligned}$$

The martingale M_t can then be rewritten

$$\begin{aligned}&\mathrm{d}M_t\\&=\sum_{\substack{\gamma,\gamma'\in\square_{[L]}\\ \gamma<\gamma'}}\left(\left(\ln(X_t(\gamma))-\ln(X_t(\gamma'))-\sum_{\ell\in[L]}\frac{\gamma_\ell-\gamma'_\ell}{2}\times\Big(\ln(p^\ell(\mathbf{X}_t))-\ln(1-p^\ell(\mathbf{X}_t))\Big)\right)\right.\\&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\left.\times\sqrt{X_t(\gamma)X_t(\gamma')}\mathrm{d}B_t(\gamma,\gamma')\right)\end{aligned}$$

It can be checked

$$\sum_{\ell\in[L]}\frac{\gamma_\ell-\gamma'_\ell}{2}\times\Big(\ln(p^\ell(\mathbf{X}_t))-\ln(1-p^\ell(\mathbf{X}_t))\Big)=\ln(\pi(\mathbf{X}_t)(\gamma))-\ln(\pi(\mathbf{X}_t)(\gamma')).$$

We thus obtain

$$\mathrm{d}M_t=\sum_{\substack{\gamma,\gamma'\in\square_{[L]}\\ \gamma<\gamma'}}\left(\ln\left(\frac{X_t(\gamma)}{\pi(\mathbf{X}_t)(\gamma)}\right)-\ln\left(\frac{X_t(\gamma')}{\pi(\mathbf{X}_t)(\gamma')}\right)\right)\sqrt{X_t(\gamma)X_t(\gamma')}\mathrm{d}B_t(\gamma,\gamma').$$

It follows

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\langle M\right\rangle_t=\sum_{\substack{\gamma,\gamma'\in\square_{[L]}\\ \gamma<\gamma'}}\left(\ln\left(\frac{X_t(\gamma)}{\pi(\mathbf{X}_t)(\gamma)}\right)-\ln\left(\frac{X_t(\gamma')}{\pi(\mathbf{X}_t)(\gamma')}\right)\right)^2X_t(\gamma)X_t(\gamma')$$

This is the same as the expectation of $(f(g^{(1)})-f(g^{(2)}))^2$, where $g^{(1)},g^{(2)}$ are independent with law X_t and $f(\gamma):=\ln(X_t(\gamma)/\pi(\mathbf{X}_t)(\gamma))$. We find

$$\begin{aligned}\mathbb{E}\left[f(g^{(1)}-f(g^{(2)}))^2\right]=&2\mathbb{E}\left[f(g^{(1)})^2\right]-2\mathbb{E}\left[f(g^{(1)})\right]^2\\=&2\mathbb{V}\mathbf{ar}[f(g^{(1)})]\end{aligned}$$

which yields $2V_t$ as promised. □

Appendix B

The L1 law of the iterated logarithm

Theorem B.0.1. *There is a universal constant $C > 0$ such that for any a continuous martingale $(M_t)_{t\in[0,T)}$ with $0 \le T \le +\infty$ we have*

$$\mathbb{E}\left[\sup_{t\in[0,T)} \frac{M_t}{\sqrt{1+\left(\langle M\rangle_t^{QV}\ln_{(2)}\left(\langle M\rangle_t^{QV}\right)\right)}}\right] \le C$$

Proof. First, let us show that we can assume M to be a Brownian motion and $T = +\infty$. Indeed, the Dambis, Dubins-Schwartz theorem (Theorem 1.6, chapter V of [159]) states that if we define $\tau_t := \langle M\rangle_t^{QV}$ and $\sigma_u := \inf\{t \ge 0, \tau_t \ge u\}$, then $(M_{\sigma_u})_{0\le u\le\tau_T}$ is a Brownian motion on $[0,\tau_T]$, noted $(B_t)_{t\in[0,\tau_T]}$. Evidently

$$\mathbb{E}\left[\sup_{t\in[0,T)} \frac{M_t}{\sqrt{1+(\tau_t^{QV}\ln_{(2)}(\tau_t^{QV}))}}\right] = \mathbb{E}\left[\sup_{u\in[0,\tau_T)} \frac{B_u}{\sqrt{1+(u\ln_{(2)}(\sqrt{u}))}}\right]$$

Up to increasing the probability space, we may assume B is well defined on all of $\mathbb{R}_+$. Furthermore, as u goes to infinity $\ln_{(2)}(\sqrt{u})$ and $\ln_{(2)}(u)$ are equivalent, so we can replace one by the other.

We know from the global law of the iterated logarithm (see [213], p.13) that

$$\mathbb{P}\left[\limsup_{u\to+\infty}\left|\frac{B_u}{\sqrt{2u\ln_{(2)}(u)}}\right| = 1\right] = 1$$

It follows that $\left(\frac{B_u}{\sqrt{1+(2u\ln_{(2)}(u))}}\right)_{t\ge 0}$ is a continuous Gaussian process, with mean 0 and asymptotically bounded by 1 as $u \to +\infty$. In particular its supremum (resp. infimum) is almost surely finite. To see that the supremum is of finite expectation, we use [214]. In this paper, the authors show in (1.2) that for any sequence of (possibly correlated) jointly gaussian random variables $(X_n)_{n\in\mathbb{N}}$, such that $\mathbb{P}[\sup_n |X_n| < +\infty] = 1$, we necessarily have $\mathbb{E}[\sup_n |X_n|] < +\infty$ (even stronger, they show that we can find small enough $\varepsilon > 0$ such that $\mathbb{E}\left[e^{\varepsilon \sup_n |X_n|^2}\right] < +\infty$). We can apply this to $\left(\frac{B_u}{\sqrt{1+(2u\ln_{(2)}(u))}}\right)_{u\in\mathbb{Q}_+}$ and obtain the result. □

RÉSUMÉ

Les modèles polygéniques décrivent les grandes populations d’organismes, dont chacun porte un grand nombre de gènes. Dans les modèles de génétique quantitative, la valeur sélective d’un organisme est déterminée par un trait quantitatif, typiquement la somme de contributions de tous les gènes de l’organisme. Cette thèse clarifiera comment les modèles polygéniques pour la génétique quantitative peuvent être obtenus à partir d’un modèle individu-centré par des limites d’échelles appropriées. Dans le chapitre 2, nous obtiendrons rigoureusement la limite polygénique à partir d’une diffusion simplifiée qui modélise la dérive génétique, la mutation, la sélection et la recombinaison, en faisant tendre jointement vers l’infini le nombre de loci et le taux de recombinaison. Dans le chapitre 3, nous introduisons un objet probabiliste qui permet une représentation en temps renversé d’une population évoluant sous l’effet de la mutation, la sélection et la recombinaison. Cette représentation permet de penser les perturbations de premier ordre à l’équilibre de liaison, quand la dérive génétique est négligée. Dans le chapitre 4, nous utiliserons des simulations et des heuristiques pour discuter du domaine de validité de la limite polygénique sous sélection stabilisante à l’équilibre. Nous discuterons heuristiquement les difficultés qui peuvent émerger dues à l’épistasie, la pléiotropie, et le déséquilibre de liaison.

MOTS CLÉS

Génétique quantitative-Génétique des populations-Limite d’échelle-Propagation du chaos

ABSTRACT

Polygenic models describe large populations of organisms, such that every organism carries a large number of genes. In quantitative genetics models, the fitness of one organism is deter- mined by a quantitative trait, which is typically obtained by summing contributions from all the genes of this organism. This thesis will clarify how polygenic models for quantitative ge- netics can be obtained from a discrete individual-based model by suitable scalings. In Chapter 2, we rigorously derive the polygenic limit from a simplified diffusion equation accounting for genetic drift, mutation, selection and recombination, by letting the number of loci and the rate of recombination go to infinity. We characterize the stationary distribution(s) using a fixed point equation. In Chapter 3, we introduce a probabilistic object which provides a backwards representation of the deterministic dynamics of a population evolving under mutation, selection and recombination. This representation clarifies how selection may generate first-order LD when genetic drift is neglected. In Chapter 4, we use simulations and heuristics to discuss the full domain of validity of the polygenic limit under stabilizing selection at statistical equilibrium. We discuss heuristically the various difficulties that may arise due to epistasis, pleiotropy, and linkage disequilibrium.

KEYWORDS

Quantitative genetics-Population genetics-Scaling limits-Propagation of chaos