# Spatial population models

## Alison Etheridge

**Abstract.** These notes are intended to provide a very rapid introduction to some of the mathematical models of theoretical population genetics. We shall be particularly concerned with models that are designed to capture the interplay between spatial structure and various forces of evolution. However, we begin with a brief review of some classical non-spatial models (the Wright-Fisher and Moran models and their diffusion approximations) and the corresponding (backwards in time) Kingman coalescent which describes the genealogical relationships between individuals sampled at random from the population.

Our first spatial model is Kimura's stepping stone model of subdivided populations, which we use to capture information about the way in which the correlation between genetic types of individuals sampled from the population decays with spatial separation.

The stepping stone model is not a particularly natural model for populations evolving in a spatial continuum and we next turn our attention to the challenges of modelling evolution in a spatial continuum, describing, in particular, Felsenstein's 'pain in the torus'. The spatial Lambda-Fleming-Viot model is then presented as one resolution of the pain in the torus, and the rest of the notes are devoted to developing that model and illustrating how it captures some of the features seen in data, before using it to investigate the interplay between spatial structure and natural selection.

# Contents

# Chapter 1

# Introduction

When Mendelian genetics was rediscovered at the beginning of the 20th
Century, it was widely believed to be incompatible with Darwin's theory
of natural selection. The mathematical sciences, in the hands of pioneers
such as Fisher, Haldane and Wright, played a fundamental rôle in
the reconciliation of the two theories, and the new field of theoretical
population genetics was born.

The intervening century has seen a remarkable interplay between
population genetics and the mathematical sciences. The fundamental
importance to population genetics of Fisher's paper *The correlation
between relatives on the supposition of Mendelian inheritance* [44], is
matched by its underpinning rôle in statistics, introducing the term
variance (and proposing its formal analysis); the Fisher-KPP equation,
introduced by Fisher [46] and Kolmogorov, Petrovskii and Piscounov [56]
to model the spread of an advantageous gene through a population living
in Euclidean space, provides us with the canonical example of a semilinear
heat equation that exhibits travelling wave solutions; Kimura's innovative
use of diffusions to understand the interplay between natural selection
and the randomness due to reproduction in a finite population (Wright's
genetic drift) was one of the earliest applications of the emerging field
of stochastic differential equations; Feller's diffusion, which describes the
evolution of a rare mutant, appears in the celebrated Ray-Knight Theorem,
is the canonical continuous state branching process, and much more
besides; and Kingman's coalescent, which describes how genes sampled
from an idealised population are related to one another, not only sits at
the heart of statistical genetics, but its elegant mathematical structure has
been the seed for the development of a huge literature on exchangeable
coalescents.

In spite of this remarkable history, fundamental questions remain
unresolved. The genetic composition of a population can be changed

by natural selection, mutation, mating, and other genetic, ecological and evolutionary mechanisms. How do they interact with one another, and what was their relative importance in shaping the patterns that we see today? And for a mathematician, these questions continue to inspire the development of whole new classes of models, with rich mathematical structure, often revealing unexpected connections with other areas.

In these lectures, we shall focus on the models which arise when we try to capture the interplay between the forces of evolution (mutation, selection, random genetic drift etc.) acting on a population and the spatial structure of that population. Whereas the pioneers in the subject could only observe genetic variation indirectly, through phenotype, modern geneticists can observe genetic variation directly. Typically, differences in DNA sequences between individuals in a sample from the population are used to infer something about the genealogical relationships between those individuals. The mathematical challenge is to provide consistent forwards in time models for the way in which the frequencies of different genetic types evolve in the population, and backwards in time models for the ways in which genes in individuals sampled from the population are related to one another. In practice, there are very significant mathematical and computational challenges associated with comparing the predictions of the resultant backwards in time 'coalescent' models with data, that have stimulated a great deal of statistical and computational innovation. Meanwhile, forwards in time mathematical models of theoretical population genetics provide an invaluable tool for understanding how different forces of evolution will interact for different parameter regimes and timescales.

Mathematical population genetics is a catalogue of examples of the power of mathematical caricatures in explaining biological phenomena, and so the first step is to distil our understanding of evolutionary forces into simple, workable models. Even the simplest models often exhibit rich mathematical structure, worthy of investigation in its own right, but from time to time we shall try to convince ourselves that they are also capturing the essence of the biology that they are trying to model.

Our focus is going to be on understanding why results should be true, rather than the full details of what are often quite technical proofs. Thus we shall primarily be trying to learn how to perform the calculations that tell us whether there is an interesting result to be proved. The details of the analysis can be found in the original papers.

The plan of the rest of these notes is as follows. We shall begin in Section 2 by describing some of the simplest imaginable models of inheritance: the Wright-Fisher model and the Moran model. These two models form the basis of our understanding of what is known as genetic drift. This is a convenient setting in which to illustrate some desirable features of models in this area. In Section 3 we then extend to our

first spatial population model, the classical Kimura stepping stone model, which can be applied to subdivided populations. This allows us to perform some exact calculations. However, many populations are not subdivided, and so in Section 4 we then turn to a model of Malécot and Wright for populations distributed across spatial continua. They (separately) used their model to derive an expression for the way in which the correlation between the genetic types of two individuals sampled from the population decays with their spatial separation. The difficulty is that there are inconsistencies in their assumptions, consideration of which resulted in Felsenstein's famous paper 'A pain in the torus' [43]. We'll explore the implications of the pain in the torus, focussing particularly on the implications for models based on branching processes, before turning to one possible resolution, the spatial Lambda-Fleming-Viot model (SLFV), in Section 5. This model will be the main focus of the rest of the lectures. In particular, in Section 6 we shall use it to explore the interactions between genetic drift, natural selection, and spatial structure.

## Acknowledgement

# Terminology

We shall use a minimal amount of biological language. The term *locus* will be used to refer to a location on the genome. We shall be thinking of a stretch of DNA, longer than a single base, but short enough that is passed undivided from parent to offspring. We shall use the term *gene* essentially interchangeably with locus, but this is laziness; in general a locus could contain no genes or several genes.

A gene can accumulate mutations that alter the corresponding DNA sequence (just a string of letters from the alphabet A, C, G, T), which can therefore occur in different forms, that we shall call *alleles*.

An excellent introduction to the underlying biology is Barton et al. [2].

# Chapter 2

# Wright-Fisher and Moran models and the Kingman Coalescent

## 2.1   Pedigrees or Genealogies?

For populations such as our own, in which individuals have two parents, the ancestry of an individual is determined by tracing parents, grandparents, great grandparents, and so on. Since populations are finite, ultimately there must be individuals that appear multiple times in this 'pedigree' and so we end up with a rather complicated branching and coalescing structure.

To investigate this further, we consider an extremely simple model of reproduction. We're going to suppose that our population is hermaphrodite, so that we don't have to worry about distinguishing males and females. The basic conclusion would not change if we were to drop this assumption. 'Diploid' refers to the fact that each individual carries two copies of each chromosome.

**Definition 2.1.1 (Diploid Wright-Fisher model).** *Consider a large diploid (but for simplicity hermaphrodite) population of size $N$. Under the diploid Wright-Fisher model, the population evolves in discrete generations. In each generation, independently, each individual has two parents, chosen uniformly at random from the previous generation.*

This is illustrated in Figure 2.1. For this (rather small) population, after five generations, three individuals in the ancestral population are included in the pedigree of *everyone* in the current population.
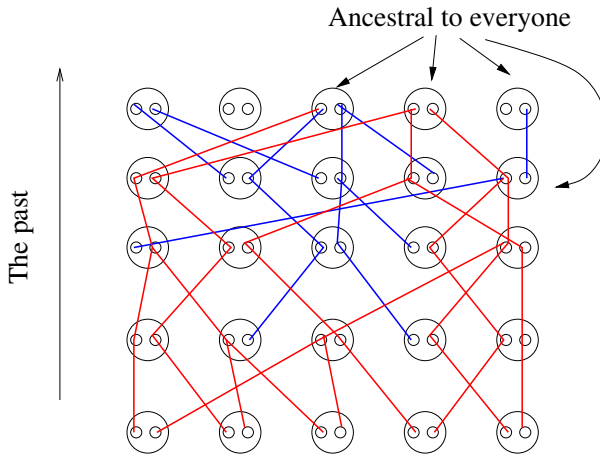
Figure 2.1: **The diploid Wright-Fisher model.** Time for the Wright-Fisher model runs down the picture. The large circles represent individuals, and the small circles represent the two copies of a particular gene carried by each individual; one copy from each parent. The lines join each gene to the corresponding parental gene; those coloured red pick out the genetic ancestry of the present-day population. For this small population, after just five generations, three individuals are in the pedigree of everyone in the population, but the rightmost of those individuals has not transmitted any genetic material to the current generation. Moreover, no ancestor has contributed genetic material to everyone in the present-day population.

**Lemma 2.1.2.** *Under the diploid Wright-Fisher model, for sufficiently large $N$, the probability that a randomly chosen individual from the population $t$ generations in the past is in the pedigree of a given individual in the current population converges to about $0.8$ as $t \to \infty$.*

**Heuristic explanation**

Let $P(t)$ be the probability that an individual alive $t$ generations in the past does *not* belong to the pedigree of our chosen individual. We write down an approximate recursion for $P(t)$.

Note that for an individual $t + 1$ generations in the past to not be in the pedigree of our chosen individual, none of that individual's direct descendants can be in the pedigree. And since, for large $N$, the random number of descendants left by a single individual in the next generation is $\texttt{Binom}(2N, 1/N) \approx \texttt{Poiss}(2)$, $P(t+1) \approx \exp(-2 + 2P(t))$. We are making two approximations here: first the Poisson approximation to the binomial distribution, and second that the probability of sitting in the pedigree of the selected individual is independent for each of the different descendants in the first generation. We look for a fixed point of this recursion.

To solve the equation $p = \exp(-2 + 2p)$, we first rearrange to obtain $(-2p)\exp(-2p) = -2\exp(-2)$. Now $z = W(z)\exp(W(z))$ defines the *Lambert W function* (or *product log* function). Although it is, in general, multivalued, for $z \in (-1/e, 0)$ there are just two branches and choosing the one with $W(z) \geq -1$ gives a unique solution. This yields $p = -\frac{1}{2}W(-2e^{-2})$ which is close to 0.2, and so the probability that the individual *is* in the pedigree is $\approx 1 - p \approx 0.8$. $\qquad\square$

However, being in the pedigree of an individual does not guarantee transmission of genetic material to that individual. Indeed, our calculation above suggests that we can expect 80% of individuals $t$ generations in the past to be in the pedigree of a particular individual, but that individual only has two copies of the relevant gene, so can only inherit genetic material from at most two of those ancestors.

Under Mendelian inheritance, the copy of a gene transmitted from parent to offspring is equally likely to be either of the parental copies. For the small population in Figure 2.1, in which the small circles represent genes, we see that of the three individuals that are pedigree ancestors of everyone, one has transmitted *no* genetic material to the current population, and none has contributed genetic material to *everyone* in the present population. What we are seeing is a difference between the timescales over which pedigree ancestry and genetic ancestry are determined. As we shall see shortly, for a population evolving according to the Wright-Fisher model, genetic ancestry is decided over timescales of the order of $N$ generations, but pedigree ancestry is determined much more quickly.

Evidently, under this simple model, individuals must start to occur multiple times in a given individual's pedigree after $\log_2 N$ generations. In fact much finer results are known. A careful justification and analysis, due to Chang [15], of the branching process approximation that lies behind our heuristic explanation of Lemma 2.1.2, shows that, with probability tending to one as $N \to \infty$, if we go back $\sim \log_2 N$ generations, then we can expect to see an individual in the population who is ancestral to *every* present-day individual; tracing back $\sim 1.77 \log_2 N$ generations, all individuals in the ancestral population are ancestral to either everyone or noone in the present day population.

Although in a large population, most individuals will have descendants alive into the indefinite future, a particular gene is highly unlikely to be transmitted.

## 2.2   The haploid Wright-Fisher model

The lesson learned so far is that the pedigree is not necessarily the right structure to consider if we are interested in genetic ancestry. An alternative is to model the genes themselves. Recall that in Figure 2.1, the small circles represent individual genes. There are $2N$ of them in any generation, and, in our simple diploid Wright-Fisher model, the subdivision into diploid individuals is unimportant; we can think of each gene as selecting a parental gene uniformly at random from the previous generation. (If we wish to recover the diploid structure, then we allow the $2N$ genes in the offspring population to fuse at random into $N$ pairs.) This leads us to study the *haploid* Wright-Fisher model. ('Haploid' means that each individual carries one copy of each gene.)

> **Assumption:** So that we don't need to keep track of factors of 2, we shall suppose that we are modelling $N$ genes.

We have reduced our model to the simplest imaginable model of inheritance:

**Definition 2.2.1 (The (haploid) Wright-Fisher model).** *Consider a large haploid population of size $N$. Under the (haploid) Wright-Fisher model, the population evolves in discrete generations. In each generation, each individual, independently, selects a parent uniformly at random from the previous generation. The offspring inherits the genetic type of the parent.*

**Remark 2.2.2.** *Of course, offspring don't choose their parents. There are several ways of achieving the same distribution.*

1. *It is equivalent to (randomly) label parents $1, \dots, N$ and assign families of sizes $\nu_1, \dots, \nu_N$, where $(\nu_1, \dots, \nu_N)$ has the multinomial distribution with $N$ trials and equal weights:*

$$\mathbb{P}[(\nu_1, \dots, \nu_N) = (k_1, \dots, k_N)] = \frac{1}{N^N} \binom{N}{k_1, \dots, k_N} \mathbf{1}_{\sum_{i=1}^{N} k_i = N}.$$

   *Indeed this is what lies behind the usual biological justfication for the model. For many organisms, during reproduction, each individual produces a very large (effectively infinite) number of gametes, which are combined to form a pool from which the next generation is sampled. If everyone is equally fit, each makes the same contribution to this pool, and so each has equal probability of having one of their gametes sampled. Later we shall use this type of reasoning to justify more complex models that include, for example, natural selection, in which parents of different types will be assigned different weights.*

2. *There is an interesting route to these family sizes that lies behind many of the branching process approximations employed in population genetics. Suppose that each individual in the parental population has a* `Poiss(1)` *number of offspring, conditioned on the total number of offspring across the population being $N$. Then, by Bayes' rule,*

$$
\begin{aligned}
\mathbb{P}[(\nu_1, \dots, \nu_N) = (k_1, \dots, k_N)] &= \frac{N!}{N^N e^{-N}} \prod_{i=1}^{N} \frac{1}{k_i!} e^{-1} \mathbf{1}_{\sum_{i=1}^{N} k_i = N} \\
&= \frac{1}{N^N} \frac{N!}{k_1! \cdots k_N!} \mathbf{1}_{\sum_{i=1}^{N} k_i = N}.
\end{aligned}
$$

   *In other words, we have recovered the multinomial distribution. If the population size is large, then, even after conditioning, the marginal probability that the first n individuals, say, have family sizes $k_1, \dots k_n$ is well-approximated by independent Poisson probabilities. And iterating over generations, the number of descendants of a given ancestor is approximately a Galton-Watson branching process with* `Poiss(1)` *offspring distribution, at least until the total number of descendants is non-negligible compared to $N$.*

So far we have ignored types, except to say that under this model offspring inherit the genetic type of their parent. Before investigating what happens to frequencies of different genetic types, we take a small diversion.

## 2.3 A brief diversion: one dimensional diffusions

A one-dimensional diffusion $\{X(t)\}_{t \geq 0}$ is a Markov process taking values in (a subset of) $\mathbb{R}$, that is continuous in both space and time. It is characterised in terms of two quantities that describe the mean and variance of the change in $X_t$ over an infinitesimally small time period. We shall call these the drift and diffusion coefficients, respectively. Writing $\Delta_h X(t) = X(t+h) - X(t)$ for the change in $X$ over the time interval $(t, t+h)$, the drift coefficient is

$$a(t,x) = \lim_{h \to 0} \frac{1}{h} \mathbb{E}[\Delta_h X(t) | X(t) = x], \tag{2.1}$$

and the diffusion coefficient is

$$b(t,x) = \lim_{h \to 0} \frac{1}{h} \mathbb{E}[(\Delta_h X(t))^2 | X(t) = x].$$

For this class of processes, for very small $h$, $\Delta_h X(t)$ is approximately normally distributed, and so we may write

$$X(t+h) \approx X(t) + h\, a(t, X(t)) + \sqrt{h\, b(t, X(t))} \xi(t)$$

where $\xi(t)$ is a standard normal random variable (and $\xi(t')$ is independent of $\xi(t)$ if $t \neq t'$). We often write this as a stochastic differential equation,

$$dX_t = a(t, X_t)dt + \sqrt{b(t, X_t)}dB_t, \tag{2.2}$$

where $\{B_t\}_{t \geq 0}$ is Brownian motion.

In many settings we arrive at a diffusion process as a limit of a sequence of discrete time and/or space Markov processes that move through a sequence of frequent small jumps. (The prototype is Brownian motion, which can be seen as a scaling limit of simple random walk.) Suppose that we have a sequence of discrete time processes $\{\widetilde{X}^{(h)}(nh)\}_{n \in \mathbb{N}}$, indexed by $h$, and write $\Delta \widetilde{X}^{(h)}(t) = \widetilde{X}^{(h)}(t+h) - \widetilde{X}^{(h)}(t)$. Then if

$$a(t,x) = \lim_{h \to 0} \frac{1}{h} \mathbb{E}[\Delta_h \widetilde{X}^{(h)}(t) | \widetilde{X}^{(h)}(t) = x], \tag{2.3}$$

and

$$b(t,x) = \lim_{h \to 0} \frac{1}{h} \mathbb{E}[(\Delta_h \widetilde{X}^{(h)}(t))^2 | \widetilde{X}^{(h)}(t) = x],$$

both exist, and

$$\lim_{h \to 0} \frac{1}{h} \mathbb{E}[(\Delta_h \widetilde{X}^{(h)}(t))^4 | \widetilde{X}^{(h)}(t) = x] = 0, \tag{2.4}$$

then as $h$ tends to zero, we can approximate $\widetilde{X}^{(h)}$ by the diffusion process with drift and diffusion coefficients $a$ and $b$ respectively. The fourth moment condition in (2.4) can be relaxed, but is often rather easy to check in our applications.

There are similar conditions to check for a sequence of continuous time Markov chains $\{X_t^h\}_{t \geq 0}$. In that setting, we have a sequence of transition rates:

$$\frac{d}{dt}\,\mathbb{P}[X_t^h \in A | X_0^h = x]\big|_{t=0} = Q_h(x, A), \qquad \text{for } x \in \mathtt{supp}(X^h), x \notin A,$$

and we require that

$$\int (y - x) Q_h(x, dy) = a^h(x) \to a(x),$$

$$\int (y - x)^2 Q_h(x, dy) = b^h(x) \to b(x),$$

and

$$\lim_{h \downarrow 0} \sup_{x \in \mathtt{supp}(X^h), |x| \leq R} Q_h(x, B(x, \varepsilon)^c) = 0, \ \forall \ R, \varepsilon > 0. \qquad (2.5)$$

The last condition guarantees that there are no jumps in the limit.

Often we combine the first two calculations and directly evaluate the infinitesimal generator, $\frac{d}{dt}\mathbb{E}[f(X_t^h)]|_{t=0}$. The limiting generator will be

$$\mathcal{L}f(x) := \frac{d}{dt}\mathbb{E}[f(X_t)|X_0 = x]\big|_{t=0} = a(x)f'(x) + \frac{1}{2}b(x)f''(x). \qquad (2.6)$$

There are lots of convenient ways to characterise the distribution of the diffusion. One which readily extends to more complex settings, and is particularly well suited to situations in which one wants to prove weak convergence results, is through a *martingale problem*. If the diffusion $X$ has generator $\mathcal{L}$ given by (2.6), then for $f$ nice enough that (2.6) is bounded,

$$f(X_t) - f(X_0) - \int_0^t \mathcal{L}f(X_s)ds \qquad (2.7)$$

is a martingale. Indeed, by the Markov property,

$$\mathbb{E}\left[ f(X_t) - f(X_0) - \int_0^t \mathcal{L}f(X_u)du \,\middle|\, \mathcal{F}_s \right]$$

$$= f(X_s) - f(X_0) - \int_0^s \mathcal{L}f(X_u)du + \mathbb{E}\left[ f(X_t) - f(X_s) - \int_s^t \mathcal{L}f(X_u)du \,\middle|\, \mathcal{F}_s \right],$$

and

$$\mathbb{E}\left[ f(X_t) - f(X_s) - \int_s^t \mathcal{L}f(X_u)du \,\middle|\, \mathcal{F}_s \right] = 0$$

by the Fundamental Theorem of Calculus.

Sometimes it is convenient to consider an equivalent formulation of the martingale problem. Applying the calculation above to $f(x) = x$,

$$M_t := X_t - \int_0^t a(X_s)ds \tag{2.8}$$

is a martingale.

Now take $f(x) = x^2$, so that

$$X_t^2 - X_0^2 - \int_0^t (2X_s a(X_s) + b(X_s))\, ds \tag{2.9}$$

is a martingale.

Combining (2.8) and (2.9),

$$M_t^2 - \int_0^t b(X_s)ds$$

$$= \left( X_t - \int_0^t a(X_s)ds \right)^2 - \left( X_t^2 - \int_0^t 2X_s a(X_s)ds \right) + \text{ martingale}$$

$$= -2X_t \int_0^t a(X_s)ds + 2 \int_0^t a(X_s) \int_s^t a(X_u)du\, ds$$

$$+ 2 \int_0^t X_s a(X_s)ds + \text{ martingale}$$

$$= -2 \int_0^t a(X_s) \left( X_t - X_s - \int_s^t a(X_u)du \right) ds + \text{ martingale.} \tag{2.10}$$

Now for $r < t$ observe that, by the tower property,

$$\mathbb{E}\left[ \int_r^t a(X_s) \left( X_t - X_s - \int_s^t a(X_u)du \right) ds \middle| \mathcal{F}_r \right]$$

$$= \mathbb{E}\left[ \int_r^t \mathbb{E}\left[ a(X_s) \left( X_t - X_s - \int_s^t a(X_u)du \right) \middle| \mathcal{F}_s \right] ds \middle| \mathcal{F}_r \right]$$

$$= \mathbb{E}\left[ \int_r^t a(X_s)\mathbb{E}\left[ \left( X_t - X_s - \int_s^t a(X_u)du \right) \middle| \mathcal{F}_s \right] ds \middle| \mathcal{F}_r \right]$$

$$= 0,$$

and note that

$$\mathbb{E}\left[ \int_0^r a(X_s) \left( X_t - X_s - \int_s^t a(X_u)du \right) ds \middle| \mathcal{F}_r \right]$$

$$= \int_0^r a(X_s) \left\{ \mathbb{E}\left[ X_t - X_r - \int_r^t a(X_u)du \middle| \mathcal{F}_r \right] + X_r - X_s - \int_s^r a(X_u)du \right\} ds$$

$$= \int_0^r a(X_s) \left( X_r - X_s - \int_s^r a(X_u)du \right) ds.$$

Thus, the integral in (2.10) is also a martingale, and we deduce that

$$M_t^2 - \int_0^t b(X_s)ds$$

is a martingale. In other words,

$$X_t - X_0 - \int_0^t a(X_s)ds \qquad (2.11)$$

is a martingale, with quadratic variation

$$\langle M \rangle_t = \int_0^t b(X_s)ds. \qquad (2.12)$$

That the two properties (2.11), (2.12) imply that (2.7) is a martingale is an application of Itô's formula and so we can use these two formulations of the martingale problem interchangeably. We shall use analogous representations of diffusions taking values in much more complicated state spaces later in these notes.

For more details of the material in this subsection, we refer to Karlin & Taylor [53], Durrett [29], or the lecture notes at `http://www.stats.ox.ac.uk/~etheridg/pdecdt.pdf`.

## 2.4 The Wright-Fisher diffusion

We now return to our Wright-Fisher population model. We should like to extend it to take account of different genetic types. Suppose that there are just two types (*alleles*), $P$ and $Q$, and write $p(t)$ for the proportion of $P$-alleles in generation $t$ and $\Delta p(t) := p(t+1) - p(t)$.

Since offspring inherit the type of their parent, given $p(t)$, $Np(t+1)$ is distributed as `Binom`$(N, p(t))$, and so

$$\mathbb{E}[\Delta p(t)] = 0, \quad \mathrm{var}\big(\Delta p(t)\big) = \frac{1}{N}p(t)\big(1 - p(t)\big). \qquad (2.13)$$

The zero expected change in allele frequency reflects neutrality; that the variance is $\mathcal{O}(1/N)$ tells us that we can only expect to see substantial changes in allele frequencies due to random fluctuations over timescales of the order of $N$ generations.

**Lemma 2.4.1.** *Measuring time in units of $N$ generations, as $N \to \infty$, the proportion of $P$-alleles in the neutral Wright-Fisher model above evolves approximately according to the Wright-Fisher diffusion:*

$$dp_t = \sqrt{p_t(1 - p_t)}dW_t,$$

*where $W_t$ is a standard Brownian motion.*

**Proof:** Since $\mathbb{E}[(\Delta p)^4] = \mathcal{O}(1/N^2)$, this follows from (2.13) and the results of the previous subsection. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ □

Eventually the Wright-Fisher diffusion will be absorbed in either $p = 0$ (with probability $1 - p(0)$)or $p = 1$ (with probability $p(0)$). This loss of diversity is due to Wright's *genetic drift* (the randomness due to reproduction in a finite population).

## 2.5   The Kingman coalescent

If we record only $p(t)$ in the Wright-Fisher model, then we lose almost all information about the way in which individuals in the population are related to one another, whereas our description in Definition 2.2.1 was in terms of an offspring 'choosing' parents and so, iterating, in terms of ancestry. From that description, it is straightforward to describe the genealogical trees relating individuals in a random sample from the population.

Consider first a sample of size two. The only information in the genealogical tree is the number of generations since their most recent common ancestor (MRCA). The chance that they 'chose' the same parent in the previous generation is just $1/N$. Since the dynamics are independent across generations, in any generation, given that they have not yet found a common ancestor, the chance that they do so across that generation is $1/N$. In other words, the number of generations that we must trace back to reach a common ancestor is $\mathtt{Geom}(1/N)$. This has mean $N$. Just as when we looked at allele frequencies forwards in time, we see that it is natural to measure time in units of $N$ generations. In these units, the time to the MRCA is $1/N\mathtt{Geom}(1/N) \approx \mathtt{Exp}(1)$.

For samples of size $k \geq 3$, say, there is the possibility of 'multiple mergers', by which we mean three or more lineages having a common parent. But the chance of such an event in a single generation is $\mathcal{O}(1/N^2)$ and so in fact we'd expect to wait $\mathcal{O}(N^2)$ generations before seeing one, and by that time all lineages will have merged through the pairwise mergers (at least for large $N$). Similarly, we don't expect to see *simultaneous* mergers of different pairs of lineages in the same generation, as such an event also has probability $\mathcal{O}(1/N^2)$.

Tracing back the ancestral lineages of a sample of size $k$, the first event will be the first of the pairwise mergers between lineages, so it will happen at the minimum of $\binom{k}{2}$ (almost) independent $\mathtt{Exp}(1)$ times; that is at an $\mathtt{Exp}\binom{k}{2}$ time. At that time, it is equally likely to be any of the $\binom{k}{2}$ possible pairs of lineages that merges. That leaves $k - 1$ lineages and we wait an additional $\mathtt{Exp}\binom{k-1}{2}$ time until a uniformly chosen pair of the lineages merges, and so on.

More formally, write $[k] = \{1, \ldots, k\}$.

**Definition 2.5.1 (Kingman coalescent [55]).** *A $k$-coalescent is a continuous time Markov chain on $\mathcal{E}_k$, the space of equivalence relations on $[k]$, with transition rates $q_{\xi,\eta}$ ($\xi,\eta \in \mathcal{E}_k$) given by*

$$q_{\xi,\eta} = \begin{cases} 1 \text{ if } \eta \text{ is obtained by coalescing two of the equivalence classes of } \xi, \\ 0 \text{ otherwise.} \end{cases}$$

*The* Kingman coalescent *on $\mathbb{N}$ is a process taking values in the space of equivalence relations on $\mathbb{N}$ with the property that, for each $k$, its restriction to $[k]$ is a $k$-coalescent. By convention, we take the initial condition to be the trivial partition into singletons.*

**Lemma 2.5.2.** *For a sample of fixed size $k$ from a population evolving according to the Wright-Fisher model of Definition 2.2.1, measuring time in units of $N$ generations, the genealogy of the sample can be approximated by a $k$-coalescent in which each block of the partition at time $t$ corresponds to an ancestral lineage and, assigning lables $1, \ldots, k$ to the individuals in the sample, the elements of a block are the labels of those individuals in the sample that are descended from that ancestor.*

**Remark 2.5.3 (Consistency).** *If we take a $(k+l)$-coalescent and restrict it to $[k]$, then we obtain a $k$-coalescent. This corresponds to sampling consistency - if we take a sample of size $k + l$ and restrict the genealogical tree to those individuals in the sample labelled $1, \ldots, k$, we arrive at a genealogy with the same distribution as if we had just taken a smaller sample in the first place.*

**Remark 2.5.4.** *The Kingman coalescent describes the genealogy of a* random *sample from the population. We do* not *get to assign types to the sample. If we condition on knowing the types in the sample, we no longer have a Kingman coalescent. (In particular, since we have not yet allowed mutation between types in our model, the ancestral lineage of an individual of type $P$ could never coalesce with that of an individual of type $Q$.)*

*If we are using the Kingman coalescent as a model, the distribution of types in the sample is determined by assigning types to the ancestral lineages at some time $t$ in the past, and then tracing their descent through the genealogy to the present.*

## Large sample sizes

In Lemma 2.5.2, we assume that the sample size $k$ is fixed and let $N \to \infty$. We justify the Kingman approximation by saying that multliple and/or simultaneous mergers will not be seen on the timescale of the coalescent. In the era of the 100,000 genomes project and other massive data sets, it is instructive to see what happens to the error that we are making as sample

sizes grow. This is all made rigorous in Melfi & Viswanath [63] (see also Chetwynd-Diggle [18]), but here we provide some heuristic arguments.

The probability of seeing two pairs of lineages within a sample of size $j$ merge into separate parents in a single generation is

$$\binom{j}{2}\binom{j-2}{2}\frac{1}{N}\frac{1}{N-1} \approx 6\binom{j}{4}\frac{1}{N^2}.$$

The probability of seeing a three-merger is $\binom{j}{3}\frac{1}{N^2}$. The probability of a single pairwise merger is $\approx \binom{j}{2}\frac{1}{N}$.

So the chance that we see a simultaneous merger (the more likely of our prohibited events) before the single pairwise merger is

$$\approx \frac{N}{\binom{j}{2}}\binom{j}{2}\binom{j-2}{2}\frac{1}{N^2} \approx \frac{j^2}{N},$$

up to a combinatorial factor which is independent of $j$. To see this, observe that the number of generations until we see a single pairwise merger or a simultaneous merger (or a multiple merger) are almost independent geometric random variables.

The chance that we see a simultaneous pairwise merger *somewhere* in the genealogical tree of a sample of size $k$ is found by summing $j = 2, \ldots, k$ to obtain

$$\sum_{j=1}^{k} \frac{j^2}{N} \approx \frac{k^3}{3N}.$$

In other words we expect the approximation to break down if the sample size is $\mathcal{O}(N^{1/3})$.

Melfi & Viswanath [63] show that we may see up to $c$ pairs of lineages merging as soon as the sample size exceeds $\mathcal{O}(N^{c/(2c+1)})$ and triple mergers once $k$ exceeds $\mathcal{O}(\sqrt{N})$.

In theory this could be problematic; in practice, although these anomalies may be present in the tree, our ability to detect them in data is restricted by the resolution at which we can reconstruct the tree, which in turn is controlled by the neutral mutation rate to which we now turn.

## 2.6   Adding mutation

In practice, reconstruction of genealogical trees from data is achieved by looking for differences between the DNA sequences of individuals in the sample. For our purposes we can think about point mutations which just change one base (letter) in the DNA sequence. We are going to assume a constant probability $\mu$ per individual per generation of a mutation at a given locus.

If we follow a single ancestral lineage, then we must wait a $\texttt{Geom}(\mu)$ number of generations until we see a mutation. If $\mu$ is such that $N\mu$ is $\mathcal{O}(1)$, then mutation and coalescence will affect a finite sample on the same timescale. Moreover, the chance of seeing a mutation and a coalescence in a single generation will be $\mathcal{O}(1/N^2)$, so (much as in our arguments above) we can treat mutation as falling on the branches of the coalescent tree as a Poisson process of rate $N\mu$. It is customary to set $\theta = 2N\mu$, so that the number of mutations that fall on a branch of length $L$ (in the Kingman coalescent timescale) is $\texttt{Poiss}(\theta L/2)$. The factor of 2 is a convention; it is chosen so that if two individuals had a common ancestor at time $T$ in the past, they will differ by a $\texttt{Poiss}(\theta T)$ number of mutations.

If $\theta$ is not $\mathcal{O}(1)$, we won't be able to 'see' a branch of length $\mathcal{O}(1)$ in our data. Patterns in data reflect evolution over timescales dictated by the neutral mutation rate.

## 2.7 The unreasonable effectiveness of the Kingman coalescent

Of course, no real population is going to evolve according to the Wright-Fisher model. Indeed, it would seem to be truly remarkable to find a laboratory population for which the Wright-Fisher model provided a reasonable approximation to the genetic variation. However, to investigate this, Buri [14] followed just over a hundred populations of *Drosophila melanogaster*, each propagated from 8 males and 8 females. In his experiment he maintained this population size in each generation through random sampling. He recorded the variance in allele frequency across populations for a gene that slightly alters eye colour (without affecting fitness).

Under the Wright-Fisher diffusion, $dp_t = \sqrt{p_t(1-p_t)}dW_t$,

$$\frac{d}{dt}\mathbb{E}[p_t] = 0, \quad \frac{d}{dt}\mathbb{E}[p_t^2] = \mathbb{E}[p_t(1-p_t)], \quad \frac{d}{dt}\mathbb{E}[p_t(1-p_t)] = -\mathbb{E}[p_t(1-p_t)].$$

In Buri's experiment, $p_0 = 1/2$ for all populations, and so the variance, $V_t$ satisfies

$$V_t = \mathbb{E}[p_0(1-p_0)]\,(1 - \exp(-t)) = \frac{1}{4}\,(1 - \exp(-t)). \tag{2.14}$$

This is, of course, in units of $N$ generations, so we should convert back to time units of single generations in order to compare to data. After $t$ generations, the predicted variance is

$$\frac{1}{4}\,(1 - \exp(-t/N)).$$
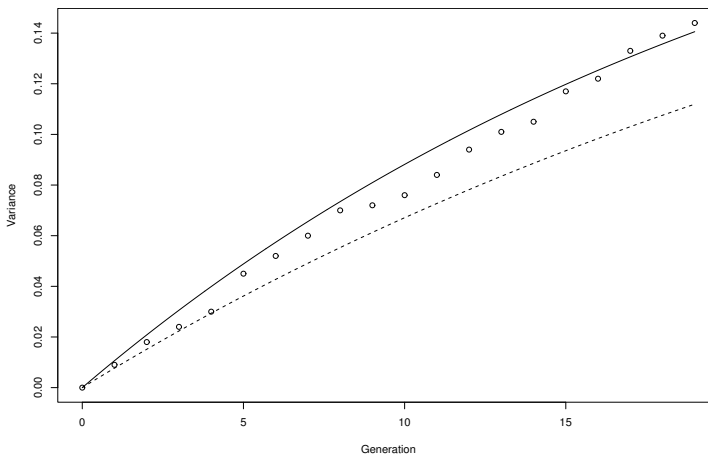
Figure 2.2:   The graph shows experimental results of Buri [14] based on just over 100 populations of *Drosophila melanogaster*, each propagated from 8 males and 8 females. At time zero, all populations were started from $p_0 = 1/2$. Variance in allele frequency across the populations is plotted against time (in generations). At time zero $V_0 = 0$. Eventually all individuals in a given population will have the same eye colour, with approximately half of the populations being of each colour, so we expect that $V_t \to 1/4$ for large $t$. Circles are data points, the dotted line is the theoretical prediction for $N = 16$, and the solid line is the theoretical prediction with $N_e = 11.5$.

In Figure 2.2, we see that if substitute $N = 16$, this prediction is a poor fit to Buri's data. However, if instead we substitute an *effective* population size, $N_e = 11.5$, the fit is quite good.

For Buri's data, one can identify $N_e = N/\sigma^2$, where $\sigma^2$ is the variance of the number of offspring of a single fly (forced to be $1 - 1/N$ in the Wright-Fisher model). What is more astonishing is that, at least if one samples from far enough apart, by substituting an effective population size in place of the census population size, the Kingman coalescent provides a surprisingly good model for genealogies of samples from natural populations. To give an idea of just how surprising this is, the effective population size of humans is around $10^4$ (Charlesworth [16]), as opposed to a census population size of around $7 \times 10^9$. Somehow, $N_e$ is capturing the effects of spatial structure, natural selection, population growth etc. in a single parameter.

We'll see the Kingman coalescent emerge from our models of spatially distributed populations when we sample over suitable scales, but we'd like

models that allow us to investigate smaller scale spatial patterns of genetic variation, and that will be a major focus of what follows.

## 2.8   The Moran model

The unreasonable effectiveness of the Kingman coalescent reflects the fact that it can be used to approximate the genealogies of populations with quite different local structures. Similarly, the Wright-Fisher diffusion, which we identified as a scaling limit of the Wright-Fisher model, approximates allele frequencies for populations with a wide variety of local structures/reproduction mechanisms (perhaps, as with the Kingman coalescent, up to a change of timescale, corresponding to replacing the census population size by an effective population size).

   If we are really interested in the behaviour of our population over the timescales of evolution, this suggests that we can choose a mathematically convenient caricature (perhaps at the expense of some biological realism) and expect the same asymptotic results. Before beginning our study of spatial structure, we introduce one such caricature. Whereas in the Wright-Fisher model, the population evolves in discrete non-overlapping generations, our next model allows for overlapping generations.

**Definition 2.8.1 (The neutral Moran model [65]).** *In the Moran model for a neutral haploid population of size $N$, reproduction events occur at the times of a Poisson process of rate $\binom{N}{2}$. At the time of such an event, a pair of genes is sampled uniformly at random from the population, one dies and the other produces one offspring. The offspring inherits the type of the parent.*

   In fact there is no accepted convention for the rate at which pairs of individuals are chosen. This choice has the advantage that we are already in the timescale of the Kingman coalescent. To see why, it is convenient to look at the graphical representation of the Moran model.

### The graphical representation

We suppose that individuals in our population at time zero are labelled by $1, \ldots, N$. Associated to each pair of labels $(i, j)$ is a rate one Poisson process, that we denote $\pi_{(i,j)}$. Since there are only finitely many labels, the points of the $\pi_{(i,j)}$ are all distinct.

   At a point of the Poisson process $\pi_{(i,j)}$, the individuals currently labelled $(i, j)$ are involved in a reproduction event in which one dies and the other splits in two (reproduces). The parent retains its label and the offspring adopts the label of the individual that died. In the graphical representation, each label (individual) is represented by a line.
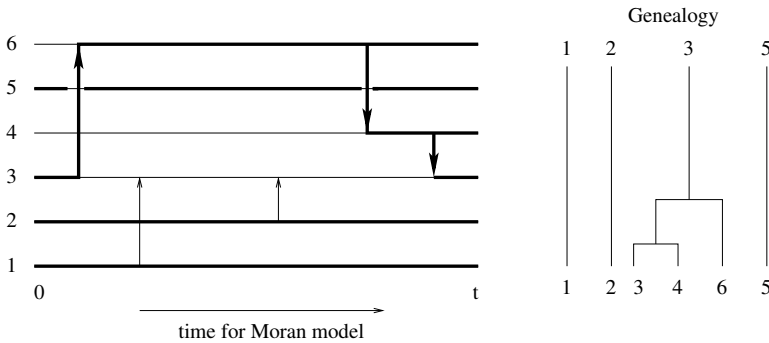
Figure 2.3: **Graphical representation of the Moran model.** Forwards in time, the individual at the tip of the arrow dies and is replaced by a copy of the individual at the tail. Genealogies have been picked out in bold and drawn separately on the right.

A reproduction event is represented by an arrow. The individual at the tip dies and is replaced by an offspring of the individual at the tail.

To trace back the ancestry of the population, we move backwards in time through the graphical representation; when we hit the tip of an arrow we follow that arrow; when we hit the tail of an arrow, our label corresponds to the individual that was the parent of the event, so we just keep going. This is illustrated in Figure 2.3.

The labels of individuals change as we move backwards in time, but, for example, suppose that we follow individuals with labels $(i(t), j(t))$. Because of lack of memory of the exponential distribution, the instantaneous rate at which an arrow falls between these two labels will always be one. And if we are following $k$ ancestral lineages, the instantaneous rate at which we see a new merger is $\binom{k}{2}$. Because the $\pi_{(i,j)}$ are independent, it is equally likely to be any pair of lineages that merge in this event. We have recovered the Kingman coalescent.

## The diffusion approximation

The calculation above tells us that we should be able to recover the Wright-Fisher diffusion by assigning types $P$, $Q$, to individuals in our population, writing $p(t)$ for the proportion of type $P$ at time $t$, and letting $N \to \infty$ with *no scaling of time*. We confirm this with a generator calculation.

Write $\mathcal{L}^N$ for the infinitesimal generator of the continuous time Markov chain $p(t)$ when the population is size $N$. Recall that events fall at rate $\binom{N}{2}$. If an event takes place at time $t$, there will only be a change in the allele frequencies if the pair of individuals involved in the event are

of different types (which happens with probability $2p(t-)\big(1 - p(t-)\big)$), in which case it is equally likely that $p$ increases or decreases by $1/N$. In other words,

$$
\begin{aligned}
\mathcal{L}^N f(p) &= \binom{N}{2} p(1-p) \left( f(p + \frac{1}{N}) - f(p) + f(p - \frac{1}{N}) - f(p) \right) \\
&= p(1-p) \binom{N}{2} \left( f(p) + \frac{1}{N} f'(p) + \frac{1}{2N^2} f''(p) - f(p) \right. \\
&\quad \left. + f(p) - \frac{1}{N} f'(p) + \frac{1}{2N^2} f''(p) - f(p) \right) + \mathcal{O}(\frac{1}{N}) \\
&= \frac{1}{2} p(1-p) f''(p) + \mathcal{O}(\frac{1}{N}).
\end{aligned}
$$

In the limit as $N \to \infty$ we recover the generator of the Wright-Fisher diffusion, as expected. Of course, to check that the limiting stochastic process really is the diffusion, we need to check that there are no jumps in the limit as in (2.5), but that is easy here as the chain only ever makes jumps of size $1/N$.

## Another small detour: the lookdown process

We have considered convergence of the forwards in time process of allele frequencies, and of the backwards in time coalescents, as separate limits. If we just look at allele frequencies, we lose sight of the genealogies completely. If we only have information about the genealogical trees, then that gives us information about how finite dimensional distributions of the process of allele frequencies are related at different time points. The converse is *not* true, and usually it is the forwards in time model of allele frequencies that we specify in order to test hypotheses about evolution. Ideally, we'd like a framework that investigates convergence of the allele frequencies and the genealogies simultaneously. This is provided by the Donnelly and Kurtz lookdown framework.

In this setting, typically, individuals are assigned (non-negative) integer or real-valued 'levels', with connections between the levels determining the genealogical trees. Lookdown constructions are motivated by the graphical representation of the Moran model, but with a very particular labelling of individuals. The term 'lookdown' comes about because during reproduction events offspring inserted at a given level typically 'look down' to individuals at lower levels to determine their parent.

When the reproductive dynamics are unaffected by the spatial locations or types of individuals in the population - as in the Wright-Fisher and Moran models that we have considered so far - the levels can be taken to be non-negative integer values, and that is the case that we shall briefly consider here, but these restrictions can be dropped if instead levels take

continuous values in $\mathbb{R}_+$ (generally the points of a Poisson process). This idea has been exploited extensively by Tom Kurtz and his coauthors. Etheridge & Kurtz [38] gives a set of 'lego blocks' for building very general classes of population processes in such a framework. Here we confine ourselves to the simplest example, discussed first in the original lookdown paper, Donnelly & Kurtz [27].

Consider a population of fixed size $N$. Individuals are assigned levels $1, \ldots, N$ by choosing uniformly at random from all possible assignments. In the same way as for the Moran model, we attach an independent Poisson process $\pi_{(i,j)}$, of rate 1, to each pair $(i,j)$ of levels. However, this time, at a point of $\pi_{(i,j)}$, the individual with the higher of the two levels $i$ and $j$ dies and is replaced by a copy of the individual *with the lower level*. In this sense it 'looks down'. Between replacement events, individuals (independently) accumulate mutations. Since the levels play such an important rôle in the dynamics, it is far from obvious that this gives rise to a reasonable population model. The key to seeing that it does, is to show that if $\{X_i(0)\}_{1 \le i \le N}$ is exchangeable (has the same distribution if we permute the labels), then so is $\{X_i(t)\}_{1 \le i \le N}$ for each $t > 0$. We then check that the probability measure-valued process

$$Z_N(t) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i(t)}$$

has the same distribution as the corresponding probability measure-valued process $\widehat{Z}_N$ obtained from a 'reasonable' population model. Moreover, because $Z_N$ is invariant under permutations of the labels, we can do this by comparing the generators of the two processes on symmetric functions.

Let us write $E$ for the space of possible genetic types in our population (corresponding to $\{P, Q\}$ above). For a vector $x = (x_1, \ldots x_N) \in E^N$, write $\Phi_{ij}(x)$ for the vector obtained from $x$ by replacing $x_j$ by $x_i$. Ignoring the possibility of mutations, the generator of the process described above is

$$A^N f(x) = \sum_{1 \le i < j \le N} \Big( f(\Phi_{ij}(x)) - f(x) \Big).$$

In this notation, the Moran model, has generator

$$\widehat{A}^N f(x) = \frac{1}{2} \sum_{1 \le i \ne j \le N} \Big( f(\Phi_{ij}(x)) - f(x) \Big).$$

For $x \in E^N$, let $z_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}$ be the corresponding empirical distribution. For $f \in \mathcal{B}(E^N)$ (bounded measurable functions on $E^N$), define

$$\alpha f(z_N) = \frac{1}{N!} \sum_{\sigma} f(x_{\sigma(1)}, \ldots, x_{\sigma(N)}),$$

where the sum is over all permutations of $\{1, \ldots, N\}$. In other words, we average out over the assignment of levels. The key observation is that

$$
\begin{aligned}
\alpha A^N f(z_N) =& \alpha \widehat{A}^N f(z_N) \\
=& \frac{1}{2} \sum_{1 \leq i \leq N} \left( \alpha f\big(z_N + \frac{1}{N}(\delta_{x_i} - \delta_{x_j})\big) - \alpha f(z_N) \right),
\end{aligned}
$$

for any choice of $x$ satisfying $z_N = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$. In other words, the generator of the Moran model and the lookdown process agree on symmetric functions.

From the point of view of the empirical distributions, provided we start from an exchangeable initial condition, it doesn't matter whether we consider the lookdown model or the Moran model to be driving the dynamics. The lookdown model is just the classical Moran model augmented with a very particular labelling of individuals in the population.

A nice property of the lookdown construction for the Moran model is that the model for a population of size $N$ is embedded in that for one of size $M$ for any $M > N$. And it is clear what happens to the bottom $n$ levels (corresponding, by exchangeability, to taking a random sample of size $n$ from the population) as $M \to \infty$, since their dynamics are unaffected by all levels greater than $n$.

# Chapter 3

# Adding space

## 3.1 Subdivided populations

Most of the models of spatially structured populations considered in theoretical population genetics take the same basic form: the population is subdivided into *demes* ('islands') sitting at the vertices of a graph and interaction between the corresponding subpopulations comes about through migration of individuals along the edges of the graph.

The most elementary example is Wright's island model (Wright [81]), in which the graph is the complete graph. More generally, the graph is chosen to caricature the actual spatial environment in which the population evolves so that, for example, we might take the graph to be $\mathbb{Z}^2$ if we were modelling a population that is distributed across $\mathbb{R}^2$.

We begin with a structured version of the Wright-Fisher model.

**Definition 3.1.1 (Structured Wright-Fisher model).** *Suppose that a population is subdivided into colonies situated at the vertices of a (possibly infinite) discrete graph. We suppose that the vertices of the graph are indexed by $\{i \in I\}$ and we write $i \sim j$ if $i$ and $j$ are neighbours in the graph. Let $N_i$ (assumed constant) denote the number of individuals living in deme (colony) $i$.*

*We say that the population evolves according to the* structured Wright-Fisher model *if it evolves in discrete generations; in each generation, each colony reproduces according to the neutral Wright-Fisher model, and then, after reproduction, a proportion $m_{ij}$ of the individuals in colony $i$ migrate to colony $j$.*

*In order to maintain constant population size in each colony we assume that*

$$N_i \sum_{j:j \sim i} m_{ij} = \sum_{j:j \sim i} N_j m_{ji}.$$

Consider the genealogy of a sample from such a population. First consider a single ancestral lineage currently in deme $i$. The chance that its immediate ancestor in the previous generation was in deme $j$ is $m_{ji}N_j/N_i$.

Now consider *two* lineages.

1. Suppose that they are both in deme $i$;

    (a) the chance that they coalesce in deme $j \neq i$ in the previous generation is
    $$\frac{\binom{m_{ji}N_j}{2}}{\binom{N_i}{2}} \frac{1}{N_j};$$

    (b) the chance that they coalesce in deme $i$ in the previous generation is
    $$\frac{\binom{N_i - \sum_{j:j\sim i} m_{ji}N_j}{2}}{\binom{N_i}{2}} \frac{1}{N_i}.$$

2. Suppose that they are in demes $i \neq j$;

    (a) the chance that they coalesce in $k \notin \{i,j\}$ in the previous generation is
    $$\frac{m_{ki}N_k}{N_i} \frac{m_{kj}N_k}{N_j} \frac{1}{N_k};$$

    (b) the chance that they coalesce in $j$ is
    $$\frac{m_{ji}N_j}{N_i} \frac{N_j - \sum_{l:l\sim j} m_{lj}N_l}{N_j} \frac{1}{N_j}.$$

In order to obtain a Kingman like limit, we suppose that $N_i$ is $\mathcal{O}(N)$ for some large $N$, that $N_i \asymp N_j$ for all $i, j$, and $m_{ij}$ is $\mathcal{O}(1/N)$.

Two lineages will simultaneously undergo a migration and a coalescence event with probability $\mathcal{O}(1/N^2)$, whereas a lineage at $i$ migrates with probability $\sum_{j:j\sim i} m_{ij}N_j/N_i$ which is $\mathcal{O}(1/N)$. In particular, we do not expect to see a simultaneous migration and merger. When in the same deme, lineages coalesce with probability $1/N_i$ which is $\mathcal{O}(1/N)$. Just as in our justification of the (unstructured) Kingman coalescent, we don't expect simultaneous or multiple mergers of lineages.

Measuring time in units of $N$ generations and letting $N \to \infty$, in the limit, the genealogy of a finite sample from the population will evolve according to a *structured coalescent*. Lineages follow random walks (with rates weighted by the ratios of population sizes in different demes) and can coalesce when within the same deme. Just as in our simple Kingman coalescent, to compare to data we need to revert to real (or at least effective) time units. This leads to:

**Definition 3.1.2 (Structured coalescent).** *We shall say that the genealogy of a finite sample from a population distributed across demes labelled by $i \in I$ as above evolves according to a* structured coalescent *if, writing $n_i(t)$ for the number of lineages in deme $i$ at time $t$ before the present,*

1. *for each $i \in I$, $n_i \mapsto n_i - 1$ at instantaneous rate $\frac{1}{N_e(i)}\binom{n_i}{2}$;*

2. *for each $i, j \in I$ with $i \neq j$,* $\begin{cases} n_i \mapsto n_i - 1 \\ n_j \mapsto n_j + 1 \end{cases}$ *at instantaneous rate* $n_i \frac{N_e(j)}{N_e(i)} m_{ji}$.

Notice that ancestral lineages are drawn into more populous demes.

**Remark 3.1.3.** *There are other modifications of the Kingman coalescent that account for other biological mechanisms. An important early contribution to this literature was Hudson [51].*

Just as for the simple Wright-Fisher model, we can also investigate what happens to allele frequencies forwards in time, as the intensity of the population in each deme tends to infinity.

Once again suppose that the population occurs in just two types $P$ and $Q$. Write $p_i(t)$ for the proportion of the population in deme $i$ at time $t$ that is of type $P$.

Consider the increment in $p_i$ over a single generation. Writing $\Delta p_i$ for the increment in the proportion of type $P$ in deme $i$ across a single generation, conditional on knowing the proportions $\{p_i\}_{i \in I}$ in the parental population,

$$
\begin{aligned}
\mathbb{E}[\Delta p_i] &= \frac{1}{N_i}\left( \left(1 - \sum_{j:j\sim i} m_{ij}\right)N_i p_i + \sum_{j:j\sim i} m_{ji} N_j p_j \right) - p_i \\
&= \sum_{j:j\sim i} \frac{N_j}{N_i} m_{ji}(p_j - p_i),
\end{aligned}
$$

where we have used that $N_i \sum_{j:j\sim i} m_{ij} = \sum_{j:j\sim i} N_j m_{ji}$. Under our assumptions above we see that $\mathbb{E}[\Delta p_i]$ is $\mathcal{O}(1/N)$. After a short calculation

$$
\mathbb{E}[(\Delta p_i)^2] = \frac{1}{N_i}\left( p_i(1 - p_i) + \mathcal{O}\left(\frac{1}{N}\right) \right),
$$

and for $i \neq j$,

$$
\mathrm{Cov}(\Delta p_i, \Delta p_j) = \mathcal{O}\left(\frac{1}{N^2}\right).
$$

Letting $N \to \infty$, under the assumption that $m_{ij}$ is $\mathcal{O}(1/N)$, we obtain a system of interacting Wright-Fisher diffusions, and so, as usual changing back into 'real' or at least 'effective' time units, the proportions follow the so-called stepping stone model, due to Kimura [54]:

**Definition 3.1.4 (Kimura's stepping stone model).** *We suppose that a population that is distributed across a collection of demes indexed by some set $I$ is also subdivided into two allelic types labelled $P$ and $Q$. The proportion of $P$-alleles in deme $i$ at time $t$ is denoted by $p_i(t)$. Under Kimura's stepping stone model:*

$$dp_i = \sum_{j:j\sim i} m_{ji} \frac{N_e(j)}{N_e(i)}(p_j - p_i)dt + \sqrt{\frac{1}{N_e(i)}p_i(1 - p_i)}dW_i, \qquad (3.1)$$

*where $\{W_i\}_{i\in I}$ are independent Brownian motions and we assume that*

$$N_e(i) \sum_{j:j\sim i} m_{ij} = \sum_{j:j\sim i} N_e(j)m_{ji}.$$

In other words we have a system of interacting Wright-Fisher diffusions.

**Remark 3.1.5.** *Kimura's stepping stone model is a cornerstone of the theoretical population genetics of spatially structured populations. Part of its power is that, just as for its non-spatial counterpart (the Wright-Fisher diffusion), it approximates a wide variety of local structures. In particular, it arises as a limit of Moran type models.*

**Example 3.1.6 (A structured Moran model).** *In the same way as above, suppose that the population lives at the vertices of a discrete graph with $N_i$ individuals living at the vertex labelled $i$, for each $i \in I$.*

*At rate $\frac{1}{N_i}$ each pair of individuals in deme $i$ is involved in a reproduction event in which one dies and the other splits in two (with equal probabilities).*

*In addition, there is exchange of individuals between demes. This can be achieved in multiple ways. For example, at rate $m_{ij}/2$ an individual in deme $i$ chooses a random individual from deme $j$ and exchanges places. As in the non-spatial setting, there are plenty of choices of time units that could have been made here. We have chosen 'real' time units.*

*In fact, if we are passing to a limit, supposing as above that $N_i$ is $\mathcal{O}(N)$ and that the rates $m_{ij}$ are chosen in such a way that $N_i \sum_{j:j\sim i} m_{ij} = \sum_{j:j\sim i} N_j m_{ji}$, we may assume that each individual in deme $i$ migrates to deme $j$ at rate $m_{ij}$. Although the population size is not preserved for $N < \infty$, the law of large numbers will ensure that portions of migrants 'balance out' for large $N$.*

*Adding types $P$ and $Q$, and assuming that individuals inherit the genetic type of their parent, it is an easy exercise to check that the generator of either of these two processes will converge to that of the Kimura stepping stone model as deme occupancy tends to infinity.*

It is common to suppose that the effective population size is the same in each deme in the Kimura stepping stone model, in which case we may

write

$$dp_i = \sum_{j:j\sim i} m_{ji}(p_j - p_i)dt + \sqrt{\frac{1}{N_e}p_i(1 - p_i)}dW_i, \qquad i \in I, \qquad (3.2)$$

and we assume $\sum_{j:j\sim i} m_{ij} = \sum_{j:j\sim i} m_{ji}$ for each $i$.

> From now on it is this form of the stepping stone model that we shall consider.

## 3.2 Duality

We have already given a justification for the claim that the genealogical tree relating individuals in a sample from the stepping stone model is given by a structured coalescent. In this section, we outline another connection between the two processes via what is known as *duality*.

The idea of duality is to express the distribution of one stochastic process, $p$, in terms of that of another process, $\underline{n}$ (which may lie on a completely different space). The key is to find a *duality function* $f(\underline{p}, \underline{n})$ for which, for each $t$,

$$\frac{d}{du}\mathbb{E}\left[f\left(\underline{p}(u), \underline{n}(t-u)\right)\right] = 0, \quad 0 \le u \le t, \qquad (3.3)$$

so that

$$\mathbb{E}\left[f\left(\underline{p}(t), \underline{n}(0)\right)\right] = \mathbb{E}\left[f\left(\underline{p}(0), \underline{n}(t)\right)\right]. \qquad (3.4)$$

If, as we vary $\underline{n}(0)$, $f(\cdot, \underline{n}(0))$ provides a wide enough class of functions, then this is enough to characterise the distribution of $\underline{p}$.

*Existence* of such a process $\underline{n}$ is often used to prove *uniqueness* (in distribution) of the original process $\underline{p}$. (If there are two such $\underline{p}$, then both must satisfy (3.4), so their distribution at time $t$ is characterised in terms of $\underline{p}(0)$ and $\underline{n}$ and so is unique.)

**Remark 3.2.1.** *In fact there are more general forms of duality in which (3.4) is replaced by*

$$\mathbb{E}\left[f(\underline{p}(t), \underline{n}(0)) \exp\left(\int_0^t \alpha(\underline{p}(s))ds\right)\right] = \mathbb{E}\left[f(\underline{p}(0), \underline{n}(t)) \exp\left(\int_0^t \beta(\underline{n}(s))ds\right)\right],$$

*for suitable $\alpha$ and $\beta$. However, some care is needed. It is often possible to find putative $f$, $\alpha$ and $\beta$ only to discover that the expectations are not defined. A good reference is Ethier & Kurtz [41], Chapter 4, Section 4.*

Finding a duality function $f(\cdot, \cdot)$ is usually a matter of trial and error; there is no systematic approach. An extensive survey is Jansen & Kurt

[52]. However, in our case, the process $\underline{p}$ in which we are interested is the vector of allele frequencies in different demes and a good starting point is often to look for a *moment dual*.

Moment duals provide expressions for the moments and mixed moments of allele frequencies,

$$\mathbb{E}\left[\prod_{i \in I} p_i^{n_i}\right], \tag{3.5}$$

where $\underline{n} = (n_i)_{i \in I}$ is a vector with non-negative integer entries, a finite number of which are non-zero.

If $p_i$ is the proportion of the population in deme $i$ that is of type $P$, the quantity in (3.5) answers the question 'What is the probability that a sample consisting of $n_i$ individuals picked at random from deme $i$ for each $i$ is all of type $P$?' In a model without mutation, if we knew the genealogical trees, we'd trace back to time zero and, since individuals inherit the types of their parents, the probability would be the same as the probability that all ancestors of the sample alive at time zero were of type $P$. So if $\underline{n}(t)$ traced out genealogies, we'd certainly have

$$\mathbb{E}[\underline{p}(t)^{\underline{n}(0)}] = \mathbb{E}[\underline{p}(0)^{\underline{n}(t)}],$$

where we have used the notation

$$\underline{p}^{\underline{n}} = \prod_{i \in I} p_i^{n_i}.$$

**Remark 3.2.2.** *We emphasize that the converse is* not *true. There are ways of constructing different genetic models in which the frequencies of the different types are the same, and so they have the same moment dual, but the genealogies are different (Taylor [78]). Nonetheless, moment duals can still tell us interesting things about our model.*

To see how duality works, we return to the stepping stone model. The quantity $n_i$ will represent a number of 'particles' in deme $i$. The function $f$ is defined by

$$f(\underline{p}, \underline{n}) = \underline{p}^{\underline{n}},$$

and we must find dynamics for the process $\underline{n}(t)$ that guarantee that equation (3.3) is satisfied. The first step is to calculate $d\underline{p}^{\underline{n}}$ with $\underline{n}$ held fixed. We use $\underline{e}_i$ to denote the vector $(\mathbf{1}_{k=i})_{k \in I}$, corresponding to a single individual in deme $i$.

$$d\left(\underline{p}^{\underline{n}}\right) = \sum_i n_i \underline{p}^{\underline{n}-\underline{e}_i} \left[\sum_j m_{ji}\left(p_j - p_i\right)\right] dt$$

$$+ \sum_i \frac{1}{2N_e} n_i\left(n_i - 1\right) \underline{p}^{\underline{n}-2\underline{e}_i} p_i\left(1 - p_i\right) dt + \sum_i \left(\ldots\right) dW_i.$$

Because we take the expectation in equation (3.3), we don't care about the exact form of the martingale term. Rearranging,

$$d\left(\underline{p}^{\underline{n}}\right) = \sum_i n_i \sum_j m_{ji} \left(\underline{p}^{\underline{n}+\underline{e}_j-\underline{e}_i} - \underline{p}^{\underline{n}}\right) dt$$

$$+ \sum_i \frac{1}{2N_e} n_i \left(n_i - 1\right) \left(\underline{p}^{\underline{n}-\underline{e}_i} - \underline{p}^{\underline{n}}\right) dt + \sum_i (\ldots) \, dW_i.$$

Now we identify the dynamics for $\underline{n}_t$ for which, if we are thinking of holding $\underline{p}$ fixed and varying $\underline{n}$, up to a mean zero martingale we obtain the same expression (bearing in mind that in our duality formula, one process runs 'backwards' in time, so the terms will then cancel). For the first term, particles should migrate according to the time reversal of the random walk that governed the forwards in time evolution of the individuals in our biological population. To deal with the second term we suppose that at rate $\frac{1}{N_e}$ each pair of particles in deme $i$ coalesces to form a single particle. In other words, we have recovered the structured coalescent.

**Lemma 3.2.3.** *Suppose that $(\underline{p}_t)_{t\geq 0}$ evolves according to the Kimura stepping stone model and that the process $(\underline{n}_t)_{t\geq 0}$, taking values in $\mathbb{Z}_+^I$ (that is vectors indexed by $I$ with non-negative integer components), evolves as follows:*

- $\begin{cases} n_i \mapsto n_i - 1 \\ n_j \mapsto n_j + 1 \end{cases}$ *at rate $n_i m_{ji}$*

- $n_i \mapsto n_i - 1$ *at rate $\frac{1}{2N_e} n_i \left(n_i - 1\right)$.*

*Then we have the duality relationship*

$$\mathbb{E}\left[\underline{p}_t^{\underline{n}_0}\right] = \mathbb{E}\left[\underline{p}_0^{\underline{n}_t}\right].$$

This duality can already be used to make some qualitative statements about the long term behaviour of the stepping stone model. We are primarily interested in populations living in either one or two dimensional Euclidean space (or subsets thereof). We use $\mathbb{Z}^1$ and $\mathbb{Z}^2$ to caricature these physical spaces and ask what happens to the process of allele frequencies under the stepping stone model in the case $m_{ij} = (m/2)\mathbf{1}_{\|i-j\|=1}$.

First we calculate $\mathbb{E}\left[p_i(t)p_j(t)\right]$ as $t \to \infty$. To do this, we start the dual process from $\underline{n}(0) = \underline{e}_i + \underline{e}_j$ and see what happens as $t \to \infty$. The distance between the corresponding two particles follows a continuous time simple random walk (run at twice the speed of a single particle). In one or two dimensions, this separation will reach zero in finite time. When this happens, there is some chance that the particles will coalesce before moving apart again. If they don't coalesce, after another (independent)

finite time, they will come back together and once again they will have some chance of coalescence. The number of times that they must come together before coalescing, and therefore the total time to coalescence, is finite. After coalescence, there will just be a single individual exploring $\mathbb{Z}$ or $\mathbb{Z}^2$. Thus, under reasonable initial conditions, $\mathbb{E}\left[p_i(t)p_j(t)\right] \to \overline{p}$ as $t \to \infty$ for some $\overline{p}$ that captures the initial prevalence of type $P$ in the population.

The same argument applies to *any* finite initial condition $\underline{n}(0)$. That is, provided that for a single random walker $X(t)$ (evolving in $\mathbb{Z}$ or $\mathbb{Z}^2$), $\mathbb{E}[p_{X(t)}(0)] \to \overline{p}$ as $t \to \infty$, we have

$$\mathbb{E}\left[\underline{p}(t)^{\underline{n}(0)}\right] \to \overline{p} \qquad \text{as } t \to \infty.$$

The only way that all the moments and mixed moments of the process can converge to the same value is if

$$\underline{p}(t) \to \left\{ \begin{array}{ll} \underline{1} & \text{with probability } \overline{p} \\ \underline{0} & \text{with probability } 1 - \overline{p} \end{array} \right. \qquad \text{as } t \to \infty,$$

where $\underline{1}$ is the vector all of whose entries are 1, and $\underline{0}$ is the vector consisting entirely of 0's. In other words, even though neither type has a selective advantage, eventually the population will consist entirely of $P$ or $Q$ alleles. (Things would be very different if natural populations lived in three dimensions or higher.)

This mirrors what we found in the non-spatial setting (below Lemma 2.4.1): the neutral Wright-Fisher diffusion will be absorbed at $p = 0$ with probability $1 - p(0)$ and $p = 1$ with probability $p(0)$.

## 3.3 Probability of identity in the stepping stone model

Our previous calculation was extremely crude and, in particular, ignored the fact that populations are continuously accumulating new mutations. In a population in which dispersal preferentially occurs between geographically close subpopulations, like our stepping stone model on $\mathbb{Z}^2$, the probability that two individuals will have the same allelic state will decrease as their separation increases, a phenomenon dubbed *isolation by distance* by Wright [81].

In this subsection we shall investigate the probability that two individuals sampled at a given separation are 'identical in state'.

---

**Assumption.** We shall suppose that each individual, in each generation, has probability $\mu$ of mutating to a new type, never before seen in the population.

---

(This is the so-called *infinitely many alleles* mutation model.) Under this model two individuals are identical in state if, following their ancestral lineages, neither has experienced a new mutation since their most recent common ancestor. Under other mutation models, in which the mutation does not always lead to a novel type, the quantity that we calculate here would be called the probability of *identity by descent*.

## One spatial dimension: unbounded domain

We follow Durrett [30], Chapter 5, and assume a structured Wright-Fisher model, in which there are $N$ individuals in each deme, evolving in discrete generations. (We consider demes with half the occupancy of those considered by Durrett, and that is reflected in the constants below.) As a first step, we work on $\mathbb{Z}$, with individuals following nearest neighbour random walk: $m_{ii} = 1 - m$, $m_{i\,i+1} = m_{i\,i-1} = m/2$. Write $\psi(i, j)$ for the probability that two individuals, one from colony $i$ and one from colony $j$ are identical in state. (When $i = j$ we suppose that two distinct individuals are sampled.) Evidently $\psi(i, j)$ will only depend on the difference $j - i$, and so without loss of generality we may consider just $\phi(i) = \psi(i, 0)$.

**Theorem 3.3.1 ([30, Theorem 5.1]).** *In the symmetric nearest neighbour Wright-Fisher model on $\mathbb{Z}$, the probability of identity for two lineages sampled at separation $i$ is*

$$\phi(i) \approx \frac{\lambda_2^{|i|}}{1 + 2N\mu + mN(1 - \lambda_2)}, \tag{3.6}$$

*where*

$$\lambda_2 = 1 + \frac{\mu}{m} - \sqrt{\frac{2\mu}{m} + \frac{\mu^2}{m^2}}.$$

**Proof.**

We suppose that $\mu$ and $m$ are small enough that we can ignore the probability that in a single generation two events affect the two individuals under consideration. This is why (3.6) is only an approximation. Now consider one generation. Under our assumptions,

$$\phi(i) = m\phi(i - 1) + m\phi(i + 1) + (1 - 2m - 2\mu)\phi(i), \quad \text{for } i \neq 0.$$

The factor of two in front of $m$ is because we are looking at the difference between the two lineages, which jumps with probability $2m$ (ignoring the double counting if both lineages jump). Rearranging,

$$\phi(i + 1) - \left(2 + \frac{2\mu}{m}\right)\phi(i) + \phi(i - 1) = 0, \quad \text{for } i \neq 0.$$

Restricting to $i \geq 1$, we have a second order difference equation whose general solution is $A\lambda_1^i + B\lambda_2^i$, where $\lambda_1 > \lambda_2$ are the roots of

$$\lambda^2 - \left(2 + \frac{2\mu}{m}\right)\lambda + 1 = 0; \tag{3.7}$$

that is

$$\lambda_i = \left(1 + \frac{\mu}{m}\right) \pm \sqrt{\frac{2\mu}{m} + \frac{\mu^2}{m^2}}.$$

Since $\lambda_1 > 1 > \lambda_2$, and we are looking for a probability, we must have $A = 0$. Using the symmetry $\phi(i) = \phi(-i)$, we have $\phi(i) = B\lambda_2^{|i|}$. It remains to find $B$, for which we need the analogue of our recursion for $\phi(0)$. Again ignoring the possibility of two events affecting our lineages in a single generation, and using now that $\phi(1) = \phi(-1)$, since the probability of coalescence of two lineages in the same colony in a single generation is $1/N + \mathcal{O}(m^2/N)$, we approximate by

$$\phi(0) = 2m\phi(1) + \left(1 - 2\mu - 2m - \frac{1}{N}\right)\phi(0) + \frac{1}{N}.$$

Rearranging, this gives

$$\left(2\mu + 2m + \frac{1}{N}\right)\phi(0) - 2m\phi(1) = \frac{1}{N},$$

and substituting in $\phi(i) = B\lambda_2^{|i|}$ we find

$$B\left[2\mu + 2m + \frac{1}{N} - 2m\lambda_2\right] = \frac{1}{N}.$$

That is

$$B = \frac{1/N}{2\mu + 2m(1 - \lambda_2) + 1/N}.$$

<div align="right">□</div>

**Remark 3.3.2.** *If the mutation probability is much smaller than the migration probability, $\mu \ll m$, then $\lambda_2 \approx 1 - \sqrt{2\mu/m}$, and $\mu \ll \sqrt{2\mu m} \approx m(1 - \lambda_2)$, so*

$$\phi(0) \approx \frac{1}{1 + 2N\sqrt{2\mu m}}, \quad \phi(i) \approx \phi(0)\left(1 - \sqrt{\frac{2\mu}{m}}\right)^{|i|},$$

*and we see an exponential decay of identity with separation.*

## One spatial dimension: bounded domain

Although some natural populations do live in essentially one dimensional habitats (e.g. populations living in the intertidal zone along coastlines, or adapted to live at a specific altitude), they do not have infinite range. Instead we often consider populations living on a torus (which is of course just a ring of colonies in one dimension). The choice of a torus is to avoid edge effects. Again following Durrett [30], we consider a ring of $L$ colonies with nearest neighbour random walk exactly as before, except that now $m_{L-1\,0} = m_{0\,L-1} = m/2$.

**Theorem 3.3.3 ([30, Theorem 5.2]).** *With a ring of $L$ colonies, the probability of identity for two lineages sampled from colonies that differ by $i$ (mod $L$) is*

$$\phi(i) \approx C\big(\lambda_1^{i-L/2} + \lambda_2^{i-L/2}\big),$$

*where $\lambda_1 > \lambda_2$ are*

$$1 + \frac{\mu}{m} \pm \sqrt{\frac{2\mu}{m} + \frac{\mu^2}{m^2}}$$

*and*

$$\frac{1}{C} = (1 + 2N\mu)\big(\lambda_1^{-L/2} + \lambda_2^{-L/2}\big) + 2N\mu\left[(1 - \lambda_1)\lambda_1^{-L/2} + (1 - \lambda_2)\lambda_2^{-L/2}\right].$$

**Proof.**

Defining $\phi(L) = \phi(0)$, for $0 < i < L$ we obtain exactly the same second order difference equation as before (with the same roots). The solution of interest to our recursion satisfies $\phi(i) = \phi(L - i)$, so using that $\lambda_1\lambda_2 = 1$, it is easy to see that the solution should take the form

$$\phi(i) = C\big(\lambda_1^{i-L/2} + \lambda_2^{i-L/2}\big).$$

To compute $C$ we use the same equation that determined $B$ in the infinite case:

$$\big(2\mu + 2m + \frac{1}{N}\big)\phi(0) - 2m\phi(1) = \frac{1}{N},$$

and substitute and rearrange to obtain the claimed result. $\qquad\square$

The time that it takes an ancestral lineage to travel a distance $L$ is about $L^2/m$ generations, so if $L^2/m \ll 1/\mu$ (the expected number of generations before we see a mutation), the lineages will have forgotten their starting point before we see a mutation and we can expect the genetic variation to be essentially the same as for a homogeneously mixing population.

For $L^2/m \gg 1/\mu$, lineages are likely to see a mutation before they 'notice' that they are living on a finite ring of colonies, and the behaviour should be much the same as on $\mathbb{Z}$. Both these statements are easily checked from Theorem 3.3.3, see Durrett [30] for the details.

## Two spatial dimensions: unbounded domain

The derivation above was very specific to nearest neighbour random walk on a one-dimensional structure. Two spatial dimensions requires a different approach. Here we follow Wright [81].

We work with a structured Wright-Fisher model on $\mathbb{Z}^2$, with $N$ genes in each deme. The population evolves in discrete generations. In each generation, first offspring are generated by Wright-Fisher sampling within each deme, then a proportion $g_1(x - y)$ of the offspring in deme $x$ migrate to deme $y$.

Under the infinitely many alleles mutation model, the probability of identity of two individuals sampled at separation $x$ is just the probability generating function (evaluated at $1 - 2\mu$) of the number of generations back to the MRCA of two individuals sampled at separation $x$ (now a two-dimensional vector). That is, writing $T$ for the time back to the MRCA, in an obvious notation,

$$\phi(x) = \mathbb{E}_x[(1 - 2\mu)^T].$$

Let $\psi_t(x)$ be the probability that two genes sampled at separation $x$ had their MRCA exactly $t$ generations in the past. For $t > 1$, we partition over the separation of the immediate ancestors of the two genes. The probability that they were migrants from the same deme is

$$G_1(x) = \int g_1(x, z) g_1(0, z) dz,$$

the convolution of two copies of $g_1$ (corresponding to modelling the separation of lineages), in which case they had a common ancestor with probability $1/N$. That is $\psi_1(x) = G_1(x)/N$. If, on the other hand, the two genes have distinct parents at separation $y$ in the previous generation, then the chance that their MRCA was $t$ generations in the past is $\psi_{t-1}(y)$. Combining these observations, for $t > 1$, we then have

$$\psi_t(x) = \sum_y G_1(x - y) \psi_{t-1}(y) - \frac{1}{N} G_1(x) \psi_{t-1}(0).$$

This can be rewritten as

$$\psi_t(x) = \frac{1}{N} \left( G_t(x) - \sum_{\tau=1}^{t-1} G_{t-\tau}(x) \psi_\tau(0) \right), \tag{3.8}$$

where $G_t$ is the $t$-fold convolution of $G_1$.

The generating function of $T$ (the time back to the MRCA), which of course depends on the sampling distance $x$, is

$$\phi(z, x) = \mathbb{E}_x[z^T],$$

and multiplying (3.8) by $z^t$ and summing over $t$ yields

$$\phi(z,x) = \frac{1}{N}\widetilde{G}(z,x)\big(1 - \phi(z,0)\big), \tag{3.9}$$

where $\widetilde{G}$ denotes the $Z$-transform (discrete Laplace transform) of $G$,

$$\widetilde{G}(z,x) = \sum_{t=1}^{\infty} G_t(x) z^t.$$

Setting $x = 0$ in (3.9) gives an expression for $\phi(z,0)$ and substituting this back into (3.9) yields

$$\phi(z,x) = \frac{\widetilde{G}(z,x)}{N + \widetilde{G}(z,0)}.$$

This takes a particularly simple form if $g_1$ is a discrete Gaussian kernel, which we then approximate by a strictly Gaussian dispersal kernel. On an infinite range this yields

$$\frac{1}{N} G_t(x) = \frac{1}{2\mathcal{N}t} \exp\big(-\frac{|x|^2}{4\sigma^2 t}\big),$$

where $\mathcal{N} = 2N\pi\sigma^2$ is Wright's *neighbourhood size*, which measures the number of 'potential parents' of an individual.

With this continuous approximation for $G_t$,

$$\frac{1}{N}\widetilde{G}(z,0) = \frac{1}{2\mathcal{N}} \sum_{t=1}^{\infty} \frac{z^t}{t} = \frac{1}{\mathcal{N}} \log\left(\frac{1}{\sqrt{1-z}}\right),$$

and

$$\frac{1}{N}\widetilde{G}(z,x) = \frac{1}{\mathcal{N}} \sum_{t=1}^{\infty} \frac{z^t}{2t} \exp\big(-\frac{|x|^2}{4\sigma^2 t}\big).$$

Provided that $|x|\sqrt{1-z}/\sigma$ is not too small, this latter quantity is approximately

$$\frac{1}{\mathcal{N}} K_0\left(\frac{|x|}{\sigma}\sqrt{1-z}\right),$$

where $K_0$ is the modified Bessel function of the second kind of degree zero. However, as $|x| \downarrow 0$, $\mathcal{N}\widetilde{G}(z,x)/N \to \log(1/\sqrt{1-z})$ whereas, since $K_0(y) \approx -\log y$ as $y \downarrow 0$, $K_0(|x|\sqrt{1-z}/\sigma)$ diverges, so this approximation breaks down at small scales.

We now have the ingredients for the generating function of the coalescence times:

$$\phi(z,0) = \mathbb{E}_0[z^T] = \frac{1}{1 - \frac{2\mathcal{N}}{\log(1-z)}}, \tag{3.10}$$

and, at least for sufficiently large $|x|$ and $z$ sufficiently close to 1,

$$\phi(z,x) = \mathbb{E}_x[z^T] = \frac{\tilde{G}(z,x)}{N + \tilde{G}(z,0)} \approx \frac{K_0\left(\frac{|x|}{\sigma}\sqrt{1-z}\right)}{\mathcal{N} - \log(\sqrt{1-z})}. \tag{3.11}$$

The expression (3.11) cannot apply for very small $|x|$ as it has the problem, inherited from $K_0$, of divergence at $x = 0$. The exact solution for these very small sampling distances will depend upon the details of the dispersal mechanism.

If we follow Barton et al. [3] and assume that there is a local scale $\kappa$ over which the probability of identity (i.e. the generating function) is approximately constant and equal to $\phi(z,0)$, then using equation (3.10) to rewrite equation (3.11) as

$$\phi(z,x) = \frac{1 - \phi(z,0)}{\mathcal{N}} K_0\left(\frac{|x|}{\sigma}\sqrt{1-z}\right),$$

equating $\phi(z,\kappa)$ to $\phi(z,0)$ and rearranging (using that $K_0(y) \approx -\log y$ as $y \downarrow 0$) we obtain

$$\phi(z,x) \approx \frac{K_0\left(\frac{|x|}{\sigma}\sqrt{1-z}\right)}{\mathcal{N} - \log\left(\frac{\kappa}{\sigma}\sqrt{1-z}\right)}. \tag{3.12}$$

It is more usual to set $z = e^{-2\mu}$ (approximately $1 - 2\mu$) then

$$\phi(e^{-2\mu}, x) = \mathbb{E}_x[e^{-2\mu T}] \approx \frac{K_0(x/\ell_\mu)}{\mathcal{N} + \log(\ell_\mu/\kappa)}, \quad \text{for } |x| > \kappa, \tag{3.13}$$

where $\ell_\mu = \sigma/\sqrt{2\mu}$ and

$$\phi(e^{-2\mu}, 0) = \frac{\log(\ell_\mu/\kappa)}{\mathcal{N} + \log(\ell_\mu/\kappa)}.$$

The expression (3.13) is called the *Wright-Malécot formula*. In particular, we see that the probability of identity, over intermediate scales, will decay approximately exponentially with sampling distance. If mutation rates are small, then this will provide a good approximation for a wide variety of local dispersal mechanisms - essentially we are approximating a random walk by a Brownian motion. In Figure 3.1, which is Figure 1 of Barton, Depaulis & Etheridge [3], we compare the predictions of the Wright-Malécot formula to the actual decay in identity for the Kimura stepping stone model for three different effective population sizes.

## Two spatial dimensions: bounded domain

Just as for one dimension, we can expect that over bounded domains the behaviour of the probability of identity will depend on the relative sizes
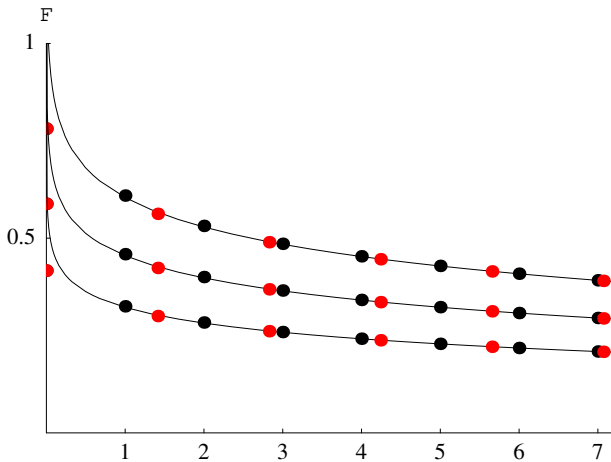
Figure 3.1: (Figure 1 of Barton, Depaulis & Etheridge [3]) Probability of identity in a stepping stone model on $\mathbb{Z}^2$ for three different values of $N_e$ (2, 5 and 10, top to bottom). The mutation rate is $\mu = 10^{-6}$ and migration is nearest neighbour with $m = 0.05$. The three sets of dots are from simulations of the stepping stone model. The curves are obtained from the Wright-Malécot formula with local scale $\kappa = 1/\sqrt{32}$.

of the time to the first mutation and the times for ancestral lineages to equilibrate over the domain. In particular, if $N$ is big and $\mu$ is very small, then identity may be insensitive to the sampling distance. Since results will be insensitive to local details, to investigate this we choose a convenient model for our purposes. Once again we (at least approximately) follow Durrett [30] to which we refer for more details.

We work on a torus of side $L$ in $\mathbb{Z}^2$. This time we take a structured Moran model (largely so that the separation of ancestral lineages is a continuous time random walk). We consider once again the time to the MRCA of a sample of size two. This can be separated into two parts: the time $T_0$ for the lineages to first meet in a deme (lattice point) and the additional time $t_0$ before they coalesce starting from separation zero. (Note this is not consistent with Durrett's notation; he uses $t_0$ to denote what in our notation corresponds to $T_0 + t_0$.)

Let's begin with $T_0$. Suppose that we sample two random walkers from random locations on the torus of side $L$ in $\mathbb{Z}^2$. The uniform distribution, which we shall denote by $\pi$, is stationary for the walkers on the torus, so since $\mathbb{P}_\pi[X_t = 0] = 1/L^2$, we know that the expected time during which two randomly chosen lineages are in the same colony up until time $L^2$ is

$$\int_0^{L^2} \mathbb{P}_\pi[X_t = 0]dt = 1.$$

If two lineages start at separation zero, then by a local central limit theorem, $\mathbb{P}_0[X_t = 0] \approx 1/(4\pi\sigma^2 t)$ (the '4' here because we are looking at the difference between two lineages). Thus the expected amount of time that two lineages chosen from colony zero are in the same colony up to time $L^2$ is

$$\int_0^{L^2} \mathbb{P}_0[X_t = 0]dt \approx \frac{\log(L^2)}{4\pi\sigma^2}.$$

We can now write

$$
\begin{aligned}
1 &= \int_0^{L^2} \mathbb{P}_\pi[X_t = 0]dt \\
&= \int_0^{L^2} \mathbb{P}_\pi[T_0 = s] \int_0^{L^2-s} \mathbb{P}_0[X_t = 0]dt\,ds \\
&\approx \mathbb{P}[T_0 \le L^2]\frac{\log(L^2)}{4\pi\sigma^2}.
\end{aligned}
$$

The approximation in the last line is easy to see up to a constant (just use that $\log(L^2 - s) \approx \log(L^2)$ for $s/L^2$ not too close to one), but in fact this more precise approximation can be justified. We then have that

$$\mathbb{P}_\pi[T_0 \le L^2] \approx \frac{2\pi\sigma^2}{\log L}.$$

From this we see that we should expect $T_0$ to be $\mathcal{O}(L^2 \log L)$. Cox & Durrett [21] show that the random walk on the torus equilibriates over a time of $o(L^2 \log L)$, so writing $\tau = T_0/(L^2 \log L)$,

$$\mathbb{P}[\tau > s + t | \tau > s] = \mathbb{P}[\tau > t] \qquad \text{as } L \to \infty;$$

i.e. $\tau$ has the lack of memory property, and so must be exponentially distributed. Combining these ideas, they show that

$$\mathbb{P}_\pi \left[ T_0 > \frac{L^2 \log L}{2\pi\sigma^2} t \right] \to e^{-t}.$$

Of course, having come together, the lineages must still actually coalesce. When two lineages are in the same colony, they coalesce at rate $1/N$ and migrate at rate $2m$. The probability that they coalesce before they jump is

$$\frac{\frac{1}{N}}{\frac{1}{N} + 2m},$$

and so the expected time to coalesce is

$$\frac{\frac{1}{N} + 2m}{\frac{1}{N}} \times \frac{1}{\frac{1}{N} + 2m} + \left( \frac{\frac{1}{N} + 2m}{\frac{1}{N}} - 1 \right) \mathbb{E}[R_0],$$

where $R_0$ is the return time to the same colony. By Kac's Lemma, $\mathbb{E}[R_0] = 1/(2m\pi_0)$, where $\pi_0$ is the stationary distribution of the difference random walk (and the $2m$ in the denominator reflects the rate of random walk jumps). Since $\pi_0$ is the uniform distribution, this gives $\mathbb{E}[R_0] = L^2/2m$ and so substituting $\mathbb{E}[t_0] = N(L^2 + 1)$.

We then have two possibilities: either the time $T_0$ to come together dominates, or, if $N/\log L \to \infty$, then

$$\mathbb{P}[(T_0 + t_0) > NL^2 t] - e^{-t} \to 0.$$

Typically we would not expect $N$ to grow with $L$ (as required for this second scenario) and so the time to coalesce will be dominated by $T_0$.

Zähle, Cox & Durrett [82] prove a much more detailed result. They show, in particular, that if one samples a finite number of individuals uniformly at random from the torus, then as $L \to \infty$, measuring time in units of order $L^2 \log L$, the genealogy of the sample converges to a Kingman coalescent. So far we have argued that for a sample of size two, firstly the time $T_0$ until the lineages are first in the same deme is such that $T_0/(L^2 \log L)$ converges to an exponentially distributed random variable as $L \to \infty$; secondly the additional time before the two lineages coalesce is asymptotically negligible in the timescale $L^2 \log L$. The extension to larger samples uses the fact (already reflected in the exponential distribution of $T_0/(L^2 \log L)$) that the

time $L^2 \log L$ is long enough for a random walk to reach its mixing time on the torus of side $L$ in $\mathbb{Z}^2$, and so at the time when a pair first come into a common deme, the positions of the lineages ancestral to the sample are no longer correlated with their starting point. This gives exchangeability: each pair of lineages is equally likely to coalesce. Moreover, when a first pair of lineages comes together, the others are still far apart and so we will not see 'multiple mergers' of lineages.

We shall see an analogue of this result in a slightly different context later on.

# Chapter 4

# Spatial continua: the pain in the torus

## 4.1 The Wright-Malécot model

Many biological populations are not subdivided into demes, but instead are distributed across a spatial continuum. For such populations, it is far from clear how best to approximate them through a stepping stone model: How should one choose the graph that supports the demes? What is $N_e$? What are the migration rates between demes? What we would really like is to be able to directly model a population that is distributed across a spatial continuum.

Wright and Malécot almost solved this problem in the 1940's. Their starting point was that the population should be distributed across (one or two-dimensional) Euclidean space as a Poisson random field, with intensity a constant times Lebesgue measure.

**Reminder:** In a Poisson random field of intensity $\lambda dx$, the number of points in a Borel set $A$ is Poisson with mean $\lambda|A|$ (where $|A|$ denotes the Lebesgue measure of $A$) and if $A \cap B = \emptyset$, then the numbers of points in $A$ and $B$ are independent.

In the Wright-Malécot model, the population evolves in discrete generations. In each generation, each individual, independently, dies and is replaced by a `Poiss`(1) number of offspring (compare to a Wright-Fisher model in which the number of offspring of an individual is `Binom`($N, 1/N$)). Offspring are distributed, independently, in a symmetric Gaussian distribution around the position of the parent.

In modern parlance, the population follows a branching random walk with `Poiss`(1) offspring distribution. Wright and Malécot assumed that the population would have the Poisson random field that they chose as

initial condition as a stationary distribution. Under that assumption, Malécot [60] wrote down recursions for the probability of identity of two individuals sampled from the population, in exactly the same way as we did for the stepping stone model, and given that we took a continuous approximation for the transition densities of the random walk, it is no surprise that he arrived at the same expression as an approximation for the probability of identity.

Perhaps not surprisingly, it was more than a quarter of a century before the inconsistencies of the Wright-Malécot assumptions were laid bare. In one and two dimensions, populations evolving according to the Wright-Malécot dynamics form clumps of arbitrary density and extent before locally dying out. This was first published by Felsenstein [43], although he credits Gillespie with bringing the clumping to his attention[1].

There is an easy way to understand this clumping in $d = 1$. Consider the descendants of a single individual. The total number of descendants is determined by a Galton-Watson branching process with a Poiss(1) offspring distribution; that is, each individual, independently, leaves a Poiss(1) distributed number of descendants in the next generation. The probability that this branching process is still alive after $t$ generations is $\propto 1/t$. The mean number of descendants is one, so conditional on being alive, the expected number of descendants must grow like $t$. These descendants will be spread out around the position of the original ancestor, but they will be no more spread out than as if they had followed *independent* Gaussian random walks, in which case they are spread over a region of diameter $\propto \sqrt{t}$, so we can expect the density to grow (at least) like $\sqrt{t}$.

In two dimensions, this argument doesn't quite work - we have to use a second moment argument that 'feels' the dependence between the spatial positions of indviduals in the branching process. We shall see shortly that in that setting we expect clumps with density that grows like $\log t$.

One's first guess might be that the clumping is a result of working on all of Euclidean space and that one could try to overcome it by working on a torus (a circle in one dimension and a square with opposite sides identified in two dimensions). To see what then happens, consider the simulation in Figure 4.1 (kindly supplied by Jerome Kelleher). At time zero, one thousand individuals are thrown down uniformly at random in the torus, resulting in the top left frame. Starting from this initial condition, the population evolves according to the Wright-Malécot model. For this realisation, the frames show the state of the population after 10, 100 and 1000 generations. Of course, since individuals reproduce independently of one another, the total population size in the torus is

---

[1]There is an error in Felsenstein's paper: he concluded that this clumping also took place in dimensions greater than two. In fact, in $d \geq 3$ there is a nontrivial stationary distribution. This was corrected by Sudbury [77].

$t = 0$   $N = 1000$   $t = 10$   $N = 1012$

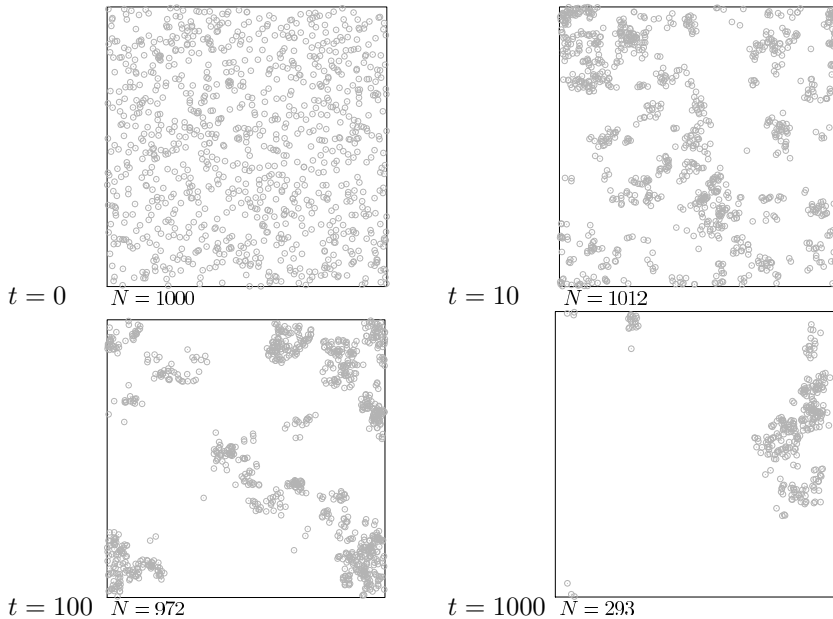$t = 100$   $N = 972$   $t = 1000$   $N = 293$

Figure 4.1: **The pain in the torus.**   Simulation, due to Jerome Kelleher, of a population evolving according to the Wright-Malécot model on a torus. In each generation, each individual (independently) produces `Poiss`(1) offspring, distributed (independently) in a Gaussian distribution around the location of the parent. Although the population size in this example remains fairly stable for timescales of hundreds of generations, the population develops 'clumps'. Top row, starting configuration of individuals and configuration after ten generations; bottom row, configuration after 100 and 1000 generations.

a critical Galton-Watson branching process, and so the population will eventually die out. However, this takes a long time and the population size is really quite stable over a fairly long time period. This gives hope that conditioning on the total population size being constant, say, we'd arrive at a sensible model; but as we see from our simulation, in which the population was very close to 1000 for the first 100 generations, this doesn't overcome the problem of clumping. Felsenstein dubbed the problem 'the pain in the torus'.

The lesson that we learn from Felsenstein is that in order to overcome clumping, we are going to need some sort of local population regulation. Before writing down one possible resolution of the pain in the torus, we try to understand the clumping a little better.

## 4.2 The Feller diffusion

If we ignore space, then a natural starting point for modelling a biological population is a Galton-Watson branching process in which each individual, independently, leaves behind a random number of offspring in the next generation. However, to specify such a model we must specify the whole offspring distribution - a countably infinite set of parameters. What we should like is an approximation that is valid for a wide class of offspring distributions, but which has only a small number of parameters. And as we have seen that evolution acts over very long timescales (on the order of population size in our Wright-Fisher models), we can expect a good model to be one that captures the large time behaviour.

In the seminal paper Feller [42], which initiated the cross-fertilisation of ideas between population genetics and diffusion theory, Feller observed that if one is observing a large population over sufficiently long timescales (a number of generations on the order of the population size), then, in suitable units, one can approximate the evolution of the population size by a one-dimensional diffusion. This mirrors our use of the Wright-Fisher diffusion to model frequencies of different genetic types and, indeed, Feller obtained the Wright-Fisher diffusion in his paper.

To identify Feller's diffusion, suppose that a population evolves according to a Galton-Watson branching process with offspring generating function $\Phi_N(s)$ (where the parameter $N$ will play the rôle of population size in our Wright-Fisher diffusion). That is,

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_i, \quad \xi_i \text{ are i.i.d}, \quad \mathbb{E}\left[s^{\xi_i}\right] = \Phi_N(s).$$

Suppose that $\Phi_N''(1) < \infty$ (so that the offspring distribution has finite variance).

If $\mathbb{E}[\xi_i] = \Phi_N'(1) = 1 + a_N$, say, then

$$\mathbb{E}[Z_n] = (1 + a_N)^n Z_0,$$

so that if we are going to measure time in units of $N$ generations, in order to obtain a nontrivial limit, we take $a_N = a/N$. We also suppose that $\Phi_N''(1)$ converges as $N \to \infty$ to a finite limit and that, for example $\Phi_N'''(1)$ is uniformly bounded.

Feller shows that as the unit $N$ tends to infinity, provided that the sequence $\{Z_0/N\}_{N \geq 1}$ converges, then so does the sequence of processes $\{Z_{\lfloor Nt \rfloor}/N\}_{N \geq 1}$, and he identifies the limit process, $\{Y(t)\}_{t \geq 0}$, often called the Feller diffusion in his honour, as the solution to the one-dimensional stochastic differential equation

$$dY_t = aY_t dt + \sqrt{\sigma Y_t} dB_t.$$

This is what we would now think of as standard Stroock-Varadhan theory as outlined in Section 2.3 (although Feller's work considerably predates that theory). In the notation that we used there, writing $Y_t^N = Z_{\lfloor Nt \rfloor}/N$,

$$\mathbb{E}[\Delta_{1/N} Y_t^N | Y_t^N = y] = \frac{1}{N} \, Ny\left(1 + \frac{a}{N}\right) - y = \frac{1}{N} ay;$$

and

$$\text{var}(\Delta_{1/N} Y_t^N | Y_t^N = y) = \frac{1}{N^2} \, Ny\sigma_N^2 = \frac{1}{N} y\sigma_N^2,$$

where $\sigma_N^2$ is the variance of the offspring distribution at the $N$th stage of the rescaling. Higher order moments of the increment are of order at most $1/N^2$ and so, since we are measuring time in units of $N$ generations, the result follows from Section 2.3.

**Remark 4.2.1.**     *1. It is easy to check that we obtain the same limit if we take a continuous time version of the branching process in which each individual has an independent exponentially distributed lifetime. In fact the same would be true for other lifetime distributions, but then we must work harder as the prelimiting model is no longer Markov.*

2. *We saw when we wrote down the Wright-Fisher model that the number of descendants of a single individual over a small number of generations could be approximated by a branching process. For a Wright-Fisher diffusion, we often use a Feller diffusion to approximate the frequency of a rare mutant as it becomes established in the population.*

Recall from Remark 2.2.2 (2) that the multinomial distribution that sits behind the Wright-Fisher model can be thought of as the result of conditioning $N$ independent Poisson random variables on their total mass. A non-rigorous calculation based on Itô's formula reveals a similar connection between Feller and Wright-Fisher diffusions. Let $X$, $Y$, be independent Feller diffusions and write $p = X/(X + Y)$. We consider $p$ conditional on $X + Y = N$.

First, by Itô's formula,

$$\begin{aligned}
dp &= \frac{dX}{X+Y} - \frac{Xd(X+Y)}{(X+Y)^2} - \frac{d\langle X, X+Y \rangle}{(X+Y)^2} + \frac{Xd\langle Y \rangle}{(X+Y)^3} \\
&= \frac{YdX - XdY}{(X+Y)^2} - \frac{Yd\langle X \rangle - Xd\langle Y \rangle}{(X+Y)^3} \\
&= \frac{Y\sqrt{\sigma^2 X}dW^X - X\sqrt{\sigma^2 Y}dW^Y}{(X+Y)^2},
\end{aligned}$$

where $W^X$ and $W^Y$ are independent Brownian motions (and the final term has cancelled). Rewriting the martingale in terms of a single Brownian motion $W$, we find

$$
\begin{aligned}
dp &= \sqrt{\frac{Y^2\sigma^2 X + X^2\sigma^2 Y}{(X+Y)^4}}dW \\
&= \sqrt{\sigma^2 \frac{X}{X+Y}\frac{Y}{X+Y}\frac{1}{X+Y}}dW \\
&= \sqrt{\frac{\sigma^2}{X+Y}p(1-p)}dW.
\end{aligned}
$$

So conditioning on $X + Y = N$ we expect

$$
dp = \sqrt{\frac{\sigma^2}{N}p(1-p)}dW.
$$

Of course, this is far from rigorous, but is a further indication of the deep connection between branching process models and models of population genetics.

## 4.3   The Dawson-Watanabe superprocess

The Wright-Malécot model that we saw in Section 4.1 is a natural extension of the Galton-Watson branching process to a spatial setting. Recall that in that model, each individual, independently, produces a `Poiss`(1) number of offspring (so the total population size is a branching process) and those offspring are scattered about the location of the parent according to a Gaussian distribution. Another way to think of this is that offspring are born at the location of the parent, but follow independent Brownian motions during their lifetimes.

It turns out that just as we can approximate a Galton-Watson branching process (for a large population and over large timescales) by a Feller diffusion, we can take the analogous scaling in this spatial context and approximate the behaviour of the population by a *measure-valued* diffusion. We generalise the Wright-Malécot model to allow for a more general offspring distribution. That is each individual leaves a random number of offspring (at the location where it dies) with generating function $\Phi_N$ such that $\Phi'_N(1) = 1 + a/N$, $\Phi''_N(1) \to \sigma^2$ and $\Phi'''_N(1) < C < \infty$ with $C$ independent of $N$. During their lifetime, individuals follow independent Brownian motions.

In the corresponding branching Brownian motion, we represent the individuals in the population by 'atoms' of mass, and the configuration

as a whole as a purely atomic measure:

$$\sum_{i=1}^{N_t} \delta_{\zeta_t^i},$$

where $\zeta_t^1, \ldots, \zeta_t^{N_t}$ are the spatial positions of the $N_t$ individuals in the population at time $t$. Under the Feller rescaling, each individual will have mass $1/N$, lifetime $1/N$ (i.e. time is measured in units of $N$ generations) and follow Brownian motion through its lifetime (or equivalently let offspring follow a random walk as in the Wright-Malécot model, but rescale space by $\sqrt{N}$). This leads to

$$X_t^N = \frac{1}{N} \sum_{i=1}^{N_{\lfloor Nt \rfloor}} \delta_{\zeta_{\lfloor Nt \rfloor}^i}.$$

Provided that $X_0^N \Rightarrow X$ (convergence in the sense of weak convergence of measures), then $X_t^N \Rightarrow X_t$ (weak convergence in the space of càdlàg paths in measure space) to a limiting process known as the *Dawson-Watanabe superprocess*.

**Definition 4.3.1 (Dawson-Watanabe superprocess).** *The Dawson-Watanabe superprocess is the continuous measure-valued process $(X_t)_{t \geq 0}$ taking values in the space $\mathcal{M}_F(\mathbb{R}^d)$ (finite measures on $\mathbb{R}^d$ endowed with the weak topology) for which, for all non-negative twice continuously differentiable functions $\phi$,*

$$\langle \phi, X_t \rangle - \langle \phi, X_0 \rangle - \int_0^t \langle D\Delta\phi, X_s \rangle ds - \int_0^t \langle a\phi, X_s \rangle ds \qquad (4.1)$$

*is a martingale with quadratic variation*

$$\int_0^t \langle \sigma^2 \phi^2, X_s \rangle ds.$$

*(The angle brackets here denote integration.)*

If the individuals in our population follow standard Brownian motions, then $D = 1/2$.

**Remark 4.3.2 (Infinite initial measures).** *For $p > d$, the Dawson-Watanabe superprocess can be extended to take values in $p$-tempered measures on $\mathbb{R}^d$, denoted $\mathcal{M}_p(\mathbb{R}^d)$. A measure $\mu$ is in $\mathcal{M}_p(\mathbb{R}^d)$ if*

$$\int \frac{1}{(1 + \|x\|^2)^{p/2}} \mu(dx) < \infty.$$

*In particular, we can construct the Dawson-Watanabe superprocess started from Lebesgue measure as initial condition, and it is often natural to do so.*

Taking $\phi$ to be identically one, and writing $Y_t = \langle 1, X_t \rangle$ for the resulting total population size process, we find that

$$M_t := Y_t - Y_0 - \int_0^t aY_s ds$$

is a martingale with quadratic variation

$$\langle Y \rangle_t = \int_0^t \sigma^2 Y_s ds.$$

Comparing to our calculations in Section 4.2 we see that $Y_t$ is a Feller diffusion.

**Remark 4.3.3.** *Like the Feller diffusion, the Dawson-Watanabe superprocess approximates a wide variety of local structures. The branching mechanism needs to be 'almost' critical and have finite variance (and a little bit more, e.g. finite $2 + \varepsilon$ moment to guarantee convergence), and the dispersion must look like Brownian motion over large spatial and temporal scales, but we can have discrete or overlapping generations, lifetimes don't have to be exponentially distributed, and so on. In spite of this permissible complexity, the limit is parametrised by just three parameters - the growth rate $a$, the variance $\sigma^2$ and the diffusion constant $D$. In the case of overlapping generations, $a$ and $\sigma^2$ will be compound parameters that comprise their discrete counterparts multiplied by the branching rate.*

There is a huge literature on superprocesses - they have a rich and beautiful mathematical structure. However, our immediate aim is simply to use the Dawson-Watanabe superprocess to better understand the clumping that we observed in the Wright-Malécot model, and so we refer to Dawson [25], Le Gall [59], Etheridge [32] and Perkins [70] for more thorough treatments.

Let's consider the Dawson-Watanabe superprocess with $a = 0$. For definiteness, set $D = 1$. First observe that, from (4.1), $\mathbb{E}[\langle \phi, X_t \rangle]$ solves the heat equation, so we can write $\mathbb{E}[\langle \phi, X_t \rangle] = \langle T_t \phi, X_0 \rangle$ where $\{T_t\}_{t \geq 0}$ denotes the heat semigroup.

In Definition 4.3.1, we took test functions $\phi$ to be constant in time, but it can be convenient to take functions which vary with time (in a deterministic way). Bearing in mind that the form of the martingale (4.1) is really telling us how $\mathbb{E}[\langle \phi, X_t \rangle]$ varies with time, we can guess that if we substitute $\langle \phi(t, x), X_t(dx) \rangle$, we should simply obtain an extra term $\langle \dot{\phi}(t, x), X_t(dx) \rangle$. This is really just because of the product rule of differentiation. Since we are assuming that $\phi$ changes deterministically with time, we do not expect a contribution from covariation between $\phi$ and $X$. The analogue for a one-dimensional diffusion would be that if $b_t$ is a deterministic differentiable function, then

$$d(b_t Y_t) = b_t dY_t + Y_t db_t.$$

For $\phi \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$, say, and $X$ the Dawson-Watanabe superprocess

$$\langle \phi(t, \cdot), X_t \rangle - \langle \phi(0, \cdot), X_0 \rangle - \int_0^t \langle (\dot{\phi}(s, X_s) + \Delta \phi(s, \cdot)), X_s \rangle \, ds \quad (4.2)$$

is a martingale with quadratic variation

$$\int_0^t \sigma^2 \langle \phi^2(s, \cdot), X_s \rangle ds.$$

We are going to choose a very special form of $\phi$. For a fixed time $T$, and $t < T$, we take $\phi(t, x) = T_{T-t}\psi(x)$ for some compactly supported $\psi$ (which is independent of time), where $T_t$ again denotes the heat semigroup, so that

$$\frac{\partial}{\partial t}\phi(t, x) = -\Delta \phi(t, x).$$

For definiteness, let's start our superprocess from Lebesgue measure at time zero. Then substituting in (4.2) and taking expectations,

$$\mathbb{E}[\langle \psi, X_T \rangle] = \langle T_T \psi, X_0 \rangle = \langle \psi, X_0 \rangle,$$

and the variance of $\langle \psi, X_T \rangle$ is

$$\mathbb{E}\left[ \int_0^T \langle \sigma^2 (T_{T-s}\psi)^2, X_s \rangle ds \right] = \int_0^T \sigma^2 \mathbb{E}\left[ \langle (T_{t-s}\psi)^2, X_s \rangle \right] ds$$

$$= \int_0^T \langle \sigma^2 T_s \left( (T_{T-s}\psi)^2 \right), X_0 \rangle ds$$

$$= \sigma^2 \int_0^T \int (T_{T-s}\psi)^2 (x) dx ds, \quad (4.3)$$

where we have used that $X_0$ is Lebesgue measure to integrate out $T_s$. This expression diverges as $T \to \infty$ in dimensions one and two. To see why, write $p(s, x, y)$ for the fundamental solution of the heat equation, then the last line of (4.3) becomes

$$\int_0^T \int \int \int p(s, x, y)\psi(y)p(s, x, z)\psi(z)dydzdxds.$$

Now observe that

$$p(s, x, y)p(s, x, z) = p(2s, y, z)p\left(\frac{1}{2}s, x, \frac{y+z}{2}\right),$$

so that (4.3) becomes

$$\int_0^T \int \int \int p(2s, y, z)p\left(\frac{1}{2}s, x, \frac{y+z}{2}\right)\psi(y)\psi(z)dydzdxds.$$

Now integrate out with respect to $x$ to leave

$$\int_0^T \int \int p(2s, y, z)\psi(y)\psi(z)dydzds,$$

which grows like $\sqrt{T}$ in $d = 1$ and like $\log T$ in $d = 2$.

Taking $\psi$ to be $\mathbf{1}_{B(0,1)}$, we see that although the expected population size in $B(0, 1)$ remains constant, the variance grows without bound, from which we conclude that with high probability there is no mass in the ball, but if there is some, there is a dense clump of population.

In fact, very detailed behaviour of the superprocess is known. In dimensions at least three, starting from Lebesgue measure there is a nontrivial stationary distribution. In dimensions one and two it suffers extinction, but on the way to extinction it develops clumps. In one dimension, there is a continuous density with respect to Lebesgue measure, but in two dimensions (and higher) the support is singular with respect to Lebesgue measure. In two dimensions, if $x$ is a typical point of the support then there exists $k$, independent of $x$ and $t$, such that

$$\lim_{r \downarrow 0} \mathbb{E}^{(x)} \left[ \frac{\langle \mathbf{1}_{B(x,r)}, X_t \rangle}{r^2 \log(\frac{1}{r})} \right] = k. \tag{4.4}$$

(The $(x)$ on the expectation is to emphasize that we are conditioning on $x$ being a point of the support of $X_t$.) This result is due to Dawson & Perkins [24]. A more detailed discussion of this and related results can be found, for example, in Chapter 6 of Etheridge [32].

## 4.4   Locally regulated populations

The lesson learned from the pain in the torus is that, in order to overcome the clumping of our population in one and two spatial dimensions, we need some *local* population regulation. In dimensions one and two, the population is dying out locally, but en route to extinction it is developing very big clumps. It seems natural to look for a model in which when the local population is sparse, an individual can have a large number of offspring, so that the population density grows; whereas if it is very dense, an individual can only have a small number of offspring, and the local population shrinks. For example, we could try to find a stochastic analogue of the heat equation with logistic growth,

$$\frac{\partial n}{\partial t} = \Delta n + \alpha n(M - n).$$

We saw for our Feller diffusion (and this feeds into our superprocess) that if the mean number of offspring of each individual is $1 + a/N$ at the $N$th

stage of the rescaling, then in the limiting diffusion, the drift is $aY_t dt$. This suggests that to obtain a superprocess with local population regulation, we should take $a$ to depend on the local population density.

In Bolker & Pacala [13], a model was introduced to the ecology literature that does just this. It is based on branching random walk, in which there will only ever be a single individual at any given point, so, necessarily, unlike the classical logistic growth equation, we cannot just measure crowdedness at a point, but instead we integrate against a test function centred on a potential parent to determine local crowdedness, and therefore their mean number of offspring.

To gain some insight, we work not with the Bolker-Pacala model itself, but instead with a superprocess analogue. Since the superprocess is singular with respect to Lebesgue measure, we once again measure local population density by integrating against a suitable test function. Our 'crowdedness' test function will be denoted by $h$ and we shall assume that it is radially symmetric (it might for example be $\mathbf{1}_{B(0,1)}$). If the maximum local population size that can be supported by the environment is $M$, then we define the growth rate at the point $x$ at time $s$ to be

$$a(s,x) = \alpha\big(M - \langle h(x,y), X_s(dy)\rangle\big).$$

One could replace this by any function of $M - \langle h, X_s \rangle$ but this is a natural starting point. If the process is to exist when started from suitable infinite initial measures (in particular Lebesgue measure), then one needs an integrability condition on the function $h$ to prevent an immediate catastrophe for the population. Since we have taken $h$ to be radially symmetric, we abuse notation by writing $h(r) = h(x,y)$ for $\|x - y\| = r$. The required condition is

$$\int_0^\infty h(r) r^{d-1} dr < \infty. \tag{4.5}$$

Of course such models considerably predate Bolker & Pacala, they can be obtained from the Dawson-Watanabe superprocess (and super-random walk) by the Dawson-Girsanov transform, but they are difficult to study because many of the powerful tools used in the study of superprocesses depend on the *branching property* (which essentially says that if we subdivide the population, then the descendants of each separate component will evolve independently), and this is destroyed by the interactions introduced by the nonlinear growth rate.

Here we simply present a heuristic argument which explains conditions under which we might hope for longterm survival in $d = 2$. We rescale according to a scaling under which the Dawson-Watanabe superprocess starting from Lebesgue measure is invariant.

Suppose then that $d = 2$. Define $X^\theta$ by

$$\langle \phi, X_t^\theta \rangle = \left\langle \frac{1}{\theta^2} \phi \left( \frac{x}{\theta} \right), X_{\theta^2 t}(dx) \right\rangle.$$

If we take $\phi = \mathbf{1}_{B(0,1)}$ say, then this corresponds to looking at the average density of $X$ over a large ball (of radius $\theta$) at times of order $\theta^2$.

We write $h^\theta(r) = \theta^2 h(\theta r)$. Then

$$\langle \phi, X_t^\theta \rangle - \langle \phi, X_0^\theta \rangle - \int_0^t \langle D\Delta\phi, X_s^\theta \rangle ds$$

$$- \int_0^t \left\langle \theta^2 \alpha \left( M - \langle h^\theta(\|x - y\|), X_s^\theta(dy) \rangle \right) \phi(x), X_s^\theta(dx) \right\rangle ds$$

is a martingale with quadratic variation

$$\int_0^t \langle \sigma^2 \phi^2, X_s^\theta \rangle ds.$$

Note then that if $r^2 h(r) \to \infty$ as $r \to \infty$, then $h^\theta$ grows without bound as $\theta \to \infty$, suggesting extinction.

In Etheridge [33], the extinction result is (almost) verified. Moreover, for a discrete space version of the model it is (essentially) shown that if $r^2 h(r)$ is bounded, then the process survives. Recalling (4.4), for a typical point in the support of the superprocess

$$\lim_{r \downarrow 0} \mathbb{E}^{(x)} \left[ \frac{\langle \mathbf{1}_{B(x,r)}, X_t \rangle}{r^2 \log(1/r)} \right] = k,$$

and so 'undoing' the rescaling,

$$\langle h^\theta(\|x - y\|), X_s(dy) \rangle \sim \log \theta.$$

If this were also true for the Bolker-Pacala model then we would expect extinction. This suggests that survival in two dimensions reflects successful eradication of clumping by the density dependent regulation term.

The proof of survival for the continuum version of the model has never been written down in detail, but, roughly, Etheridge [33] says that in order for the process to survive, individuals must spread out from their parents fast enough that they can colonise vacant regions before being killed by the crowding caused by their own close relatives. Although this gives us hope that we can find natural models, based on branching processes, which do not have clumps and which have a nontrivial stationary distribution, we have no tractable means of characterising the genealogical trees, and so no way to compare to genetic data.

There has been very considerable effort to find alternative models of locally regulated populations (we refer in particular to Birkner & Depperschmidt [11]), but analysis of such models has proved to be very difficult.

## Scaling the stepping stone model

In some sense, the stepping stone model is the ultimate model of population regulation - the population size is constrained to be constant in each deme - and so another possible approach to finding a locally regulated model in a spatial continuum involves scaling the stepping stone model according to the diffusive rescaling. In one dimension, this can be achieved. We start from the stepping stone model on $\mathbb{Z}$:

$$dp_i = \sum_{j \in \mathbb{Z}} m_{ji}(p_j - p_i)dt + \sqrt{\frac{1}{N_e}p_i(1 - p_i)}dW_i, \quad i \in \mathbb{Z},$$

where $\{W_i\}_{i \in \mathbb{Z}}$ is a collection of independent Brownian motions. Suppose that $m_{ji} = m_{ij}$ and that $\sum_{j \in \mathbb{Z}} m_{ji}|j - i|^2 = m$. We apply a diffusive scaling: $i \mapsto i/L$, $t \mapsto L^2 t$ and $N_e \mapsto LN_e$, then (provided the initial conditions converge) the sequence of scaled processes $p^L$ converges (weakly) as $L \to \infty$ to the solution to the stochastic p.d.e.

$$\frac{\partial p}{\partial t} = \frac{m}{2}\Delta p + +\sqrt{\frac{1}{N_e}p(1 - p)}\dot{W}, \tag{4.6}$$

where $W$ is space-time white noise. (See Walsh [80] for an introduction to stochastic p.d.e.s. The scaling $N_e \mapsto LN_e$ is replacing population size $N_e$ by population density.) This result was obtained by Shiga [73]. Convergence of the first and second moments was proved by Nagylaki [67, 68].

Rather than develop the theory of stochastic p.d.e., which is well beyond our scope, we use the dual process of ancestral lineages to understand why this scaling can be expected to lead to a nontrivial limit. Under the diffusive scaling, individually, the ancestral lineages converge to Brownian motions. In $d = 1$, these motions will meet, and to obtain the stochastic p.d.e. limit above, we arrange that they coalesce according to the local time that they spend together. To see what the right scaling should be, suppose that we scale time by $L^2$ and space by $L$. We need a lemma.

**Lemma 4.4.1.** *Consider continuous time simple random walk on $\mathbb{Z}$, started from the origin. The number of excursions away from zero that the walk makes before it reaches $\pm L$ is geometric with mean $L$.*

**Proof.**

Consider a continuous time simple random walk $X_t$ on $\mathbb{Z}$, with initial condition $X_0 \in [0, L]$ and stopped on exiting this interval. Note that $X_t$ is a martingale and so writing $\tau = \inf\{t > 0 : X_t \in \{0, L\}\}$ (which is a stopping time), by Doob's Optional Stopping Theorem $\mathbb{E}[X_\tau] = X_0$. In particular, taking $X_0 = 1$, the probability that started from 1, $X$ hits $L$ before it hits zero is $\mathbb{E}[X_\tau]/L = 1/L$. If every time the walk returns to

zero, we move it back to one, it will take of the order of $L$ attempts before we hit $L$.                                                                                            □

This lemma tells us that if two lineages meet in a deme, but move apart before coalescing, the chance that they move to a separation of order $L$ before coming back to the same deme is order $1/L$. In other words it takes order $L$ visits to the same deme before we see them move apart to a distance which will be of order 1 in our scaled units. If we arrange that the probability of coalescence at each visit is also order $1/L$, then coalescence and separation to distance $\mathcal{O}(1)$ have comparable probabilities and we can expect to see both in the limit. So in order to obtain the stochastic p.d.e. limit, on top of the diffusive scaling, we scale the population size in each deme by $L$. Once again, we can think of replacing population size in a deme by population density on $\mathbb{R}$.

In two dimensions, two Brownian motions will never meet and so we will not be able to find a scaling in which coalescence persists in the limit. Since coalescence reflects the *genetic drift*, that is the stochastic term in our stochastic p.d.e., we see that a diffusive scaling will lead to the deterministic heat equation. Indeed, the stochastic p.d.e. (4.6) has no solution in $d = 2$.

An alternative approach to finding a continuum analogue of the stepping stone model is to start with the backwards in time process of ancestral lineages, and try to write down a continuum analogue. Several authors have suggested taking systems of Brownian motions, with pairs of lineages coalescing at a rate that depends on their separation - with the resultant lineage typically being at the midpoint of the coalescing pair. The problem with this (other than the lack of a corresponding forwards in time model) is that this system of coalescing lineages lacks sampling consistency: suppose we take a sample of $n + m$ lineages and delete $m$ at random - the resultant tree is not the same as if we had just modelled $n$ lineages directly; in the full tree, whenever one of the $n$ lineages (or its ancestor) encounters one of the deleted lineages, it will jump - something that we would not see in the ancestral tree constructed directly from the $n$ lineages.

# Chapter 5

# Overcoming the pain in the torus: the spatial Lambda-Fleming-Viot model

## 5.1  Some features to incorporate

Although the Wright-Malécot model is inconsistent with the assumptions that Wright and Malécot were making, their calculations lead to a formula for the probability of identity that is completely consistent with what we observed for the stepping stone model. This suggests that it should be possible to write down a continuum model that is consistent with the Wright-Malécot formula.

Ideally, we'd like this model to display some of the other features that we observe in real data. For example, as we have already seen in Section 2.7, genetic diversity (for which effective population size can be seen as a proxy) is often orders of magnitude lower than predicted by census population size and a Wright-Fisher model. Moreover, the Wright-Malécot formula predicts that the probability of identity in state decays approximately exponentially with sampling distance for a two-dimensional population. However, what often happens in real data is that, although the decay over intermediate scales is indeed approximately exponential, the decay in correlations over larger scales is slower than predicted by the Wright-Malécot formula. This can, for example, be seen in human mitochondrial DNA data in Figure 4b of Simoni et al. [74].

One possible explanation for both of these observations is the large

scale fluctuations experienced by real populations, in which movement and reproductive success of many individuals can be correlated. For example, climate change has caused extreme extinction-recolonisation events that dominate the demographic history of humans and other species (Eller et al. [31]). As an example of the timescales involved, the last glacial maximum, in which much of Northern Europe was covered in ice, was of the order of 20,000 years ago. But the effective population size of humans is around $10^4$ and with an intergeneration time of, say, twenty years, that says that our genetic history was determined over much longer timescales, of the order of a hundred thousand years. The model that we are going to describe is not appropriate for modelling glacial maxima, but one might think of plants, living on a forest floor and being periodically eliminated by forest fires, after which recolonisation is typically extremely rapid - and indeed can be treated as instantaneous from the perspective of genetics.

Another factor that we may wish to incorporate into our models is that for a population distributed across a spatial continuum, neighbourhood size, which we recall measures the number of 'potential parents' of an individual, may be small. The reason that we only ever saw pairwise mergers of ancestral lineages in the Kingman coalescent was that the number of potential parents of each individual was very large. In a spatial continuum, if individuals are sampled from close to one another, pairwise mergers may no longer dominate.

## 5.2   An individual based model

We begin with an individual based model that was introduced in Berestycki, Etheridge & Hutzenthaler [8]. We suppose that the population evolves in $\mathbb{R}^d$ ('though this is far from necessary). We borrow ideas from the classical models of earlier sections. The most biologically relevant case is $d = 2$.

For initial condition, we follow Wright and Malécot and take a Poisson Point Process of constant intensity $\lambda dx$ on $\mathbb{R}^d$ ('though the initial condition is rapidly forgotten in our model). The key to the model is that reproduction will be based on *events* rather than on individuals. It will be this that overcomes the pain in the torus.

The model is parametrised by a real number $\lambda > 0$ and a measure which can be written as

$$\xi(dr, du) = \mu(dr)\nu_r(du) \tag{5.1}$$

on $(0, \infty) \times (0, 1]$. Here $\mu$ is a (possibly infinite) measure on $(0, \infty)$ which will determine, for each $r$, the rate at which 'events of radius $r$' fall on any given point. For each $r$, $\nu_r$ is a probability measure on $(0, 1]$ which will determine the 'impact' of an event on the local population. It tells us the

expected proportion of the population in the region covered by an event
that will die during the reproduction event.

To avoid trivialities we also assume that $\xi\left((0,\infty)\times(0,1]\right)>0$. The
dynamics are driven by a Poisson point process, $\Pi$, on $\mathbb{R}_+\times\mathbb{R}^d\times\mathbb{R}_+\times(0,1]$
with intensity $dt\otimes dx\otimes\xi(dr,du)$. In other words, the number of events in
$[s,t]\times A\times[r,R]\times B$ is Poisson with mean $(t-s)\mathrm{Vol}(A)\xi([r,R]\times B)$.

Each point $(t,x,r,u)$ of $\Pi$ corresponds to an event which affects
only individuals living within $B(x,r)$, the closed ball of radius $r$ and
centre $x\in\mathbb{R}^d$. Frequent 'small' events model ordinary reproduction,
whereas infrequent 'large' events mimic the effects of large-scale
extinction/recolonisations. More precisely, at a point $(t,x,r,u)$ of $\Pi$, if
$B(x,r)$ is empty do nothing. Otherwise:

1. choose a parent uniformly at random from those individuals present
   in the ball;

2. each individual in $B(x,r)$ (including the parent), independently, dies
   with probability $u$;

3. throw down new individuals (with the same type as the parent)
   according to an independent Poisson point process with intensity
   $u\lambda\mathbf{1}_{B(x,r)}(y)dy$.

This mechanism can be thought of as regulating the reproductive success of
individuals. If the ball $B(x,r)$ is crowded, then each individual living there
has only a small chance of reproducing. On the other hand if the ball is
only sparsely populated, an individual living there has a significant chance
of producing a Poisson number of offspring with mean $\lambda u\mathrm{Vol}(B(x,r))$.

Véber & Wakolbinger [79] have shown that this model exists for

$$\int_0^\infty\int_0^1 ur^d\nu_r(du)\mu(dr)<\infty.$$

(The original paper assumed a stronger condition.)

If we were to allow 'births' (creation of population) in events when
there is no potential parent present in the ball, then the population would
have a Poisson point process with intensity $\lambda dx$ as stationary distribution.
However, we do not. Nonetheless, because neighbourhoods overlap, an
empty region can subsequently become recolonised. The question is
whether this is enough to prevent the population from dying out. In
Berestycki, Etheridge & Hutzenthaler [8], it is shown that there is a critical
value of the parameter $\lambda$ below which extinction is certain, but above which
the population, started from a Poisson random measure with constant
intensity, survives (indeed there is an ergodic stationary distribution).
The proof is rather straightforward. An easy coupling shows that the
population is monotone in $\lambda$. For small $\lambda$ we can show that the population

dies out by comparing to a subcritical branching process. For large $\lambda$, we show survival by comparing to oriented percolation. This is a bit fiddly because of the discrete structure of percolation, and it turns out to be simpler to compare to edge percolation than the more common site percolation.

We should now like to identify the distribution of the genealogy of a sample from such a population. However, it turns out that this is complicated. Because regions can, and do, become empty, knowing that there is an individual in our sample at a point $x \in \mathbb{R}^d$ tells us something about the Poisson process of events that have occurred in the past. On the other hand, for sufficiently large $\lambda$, one expects that the distribution of the sample tells us very little about the Poisson Point Process of events. In that case, we expect the genealogy to be well approximated by a system of branching and coalescing random walks determined by the time reversal of $\Pi$. In this process, an ancestral lineage can only move when it is covered by the region affected by a reproduction event; it then jumps, with probability $u$, to the position of the parent of the event (which will be uniformly distributed on the affected region). Notice that if multiple lineages are in the affected region, then we could see a 'multiple merger' coalescence event, by which we mean a merger of three or more lineages.

Our approach will be to let $\lambda \to \infty$ so that this description of the genealogical trees becomes exact. We can also then describe the corresponding forwards in time model of allele frequencies. This will be the spatial Lambda-Fleming-Viot process. Etheridge & Kurtz [38] use a lookdown construction to prove the joint convergence of the forwards and backwards in time models (including for some more general models than those considered here), but that is beyond our scope.

## 5.3   Lambda-coalescents and the Lambda-Fleming-Viot process

To understand our description of the limiting spatial Lambda-Fleming-Viot process (SLFV), it is perhaps easier to first consider the non-spatial setting.

Recall two features that we require of our models for ancestral lineages: first we would like to allow for the possibility of small neighbourhood size, so that we may see more than two lineages merging in a single event; we call coalescents that allow for these events *multiple merger coalescents*. We also require *sampling consistency*: if we take the coalescent for a sample of size $n + k$ and delete $k$ leaves at random, then we recover the coalescent for a sample of size $n$.

In general, we might wish to allow for *simultaneous* mergers of lineages - especially resulting from an extinction-recolonisation event - but here,

for simplicity, we shall restrict ourselves to coalescents which don't allow simultaneous mergers. For us, this corresponds to only taking one parent in each reproduction event (and one founder in each extinction-recolonisation event). Subject to this restriction, in the non-spatial context, the most general form that a consistent coalescent can take is well-known - it is a so-called Lambda-coalescent. These were introduced independently by Donnelly & Kurtz [28], Pitman [71] and Sagitov [72].

**Definition 5.3.1 (Lambda coalescent).** *If there are currently* $n$ *ancestral lineages, each transition involving a particular subset of* $j$ *of them merging happens at rate*

$$\beta_{n,j} = \int_0^1 u^j (1-u)^{n-j} \frac{\Lambda(du)}{u^2},\qquad(5.2)$$

*where* $\Lambda$ *is a finite measure on* $[0,1]$.

If we take $\Lambda = \delta_0$, then we recover the Kingman coalescent.

The form (5.2) becomes considerably less mysterious when one writes down the corresponding forwards in time model for allele frequencies. We're going to suppose that there is no Kingman component, that is $\Lambda(\{0\}) = 0$. The forwards in time process, which appears from a lookdown construction in Donnelly & Kurtz [28], is dubbed the *generalised Fleming-Viot process* by Bertoin & Le Gall [10]. We shall call it the Lambda-Fleming-Viot process. When there is no Kingman component, it has a particularly simple form.

We write $K$ for the space of possible genetic types (assumed compact).

**Definition 5.3.2 (Lambda-Fleming-Viot process (with no Kingman component)).** *The Lambda-Fleming-Viot process takes values in* $\mathcal{M}_1(K)$, *the space of probability measures on* $K$. *We denote its state at time* $t$ *by* $\rho(t, \cdot)$. *Its dynamics are driven by a Poisson Point Process,* $\Pi$, *of events on* $[0, \infty) \times (0, 1]$, *with intensity* $dt \otimes u^{-2}\Lambda(du)$.

*If* $(t, u) \in \Pi$, *a reproduction event takes place at time* $t$ *in which a portion* $u$ *of the population is replaced by offspring of an individual chosen uniformly at random from the population immediately before the event. We call* $u$ *the* impact *of the event. Offspring inherit the type of the parent. That is for* $(t, u) \in \Pi$:

1. *sample* $k \sim \rho(t-, \cdot)$;

2. $\rho(t, \cdot) = (1-u)\rho(t-, \cdot) + u\delta_k$.

This process is illustrated in Figure 5.1. The duality with the Lambda-coalescent is now clear. For convenience, think of time as being extended to the whole real line, so that the Poisson Point Process that drives events is reversible. At a point $(t, u)$ of the reversed process, all the ancestral
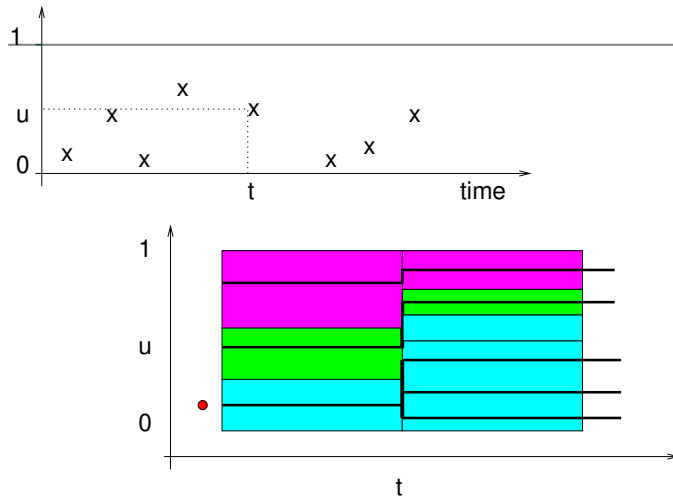
Figure 5.1: **The Lambda-Fleming-Viot process.** The upper graph depicts the Poisson point process of events; the lower picture illustrates the effect on the composition of the population of the event of impact $u$ at time $t$. The parent (depicted by the red dot) is a uniform pick from the population immediately before the event. After the event, the portion of the population corresponding to $[0, u]$ on the vertical axis are the offspring of the event. The proportions of different types in the remaining portion $[u, 1]$ mirror those before the event. The effect of the same event on five lineages ancestral to a random sample from the population is superposed as bold lines.

lineages in the portion of the population that are offspring of the event will coalesce into a single parent - the others will be unaffected by the event. If our lineages are a random sample, then each, independently, has probability $u$ of being among the offspring, so for a particular set of $j$ lineages from among $n$, the probability that exactly those lineages merge is $u^j(1-u)^{n-j}$, and integrating against the rate of events, we find that the rate at which exactly those $j$ lineages merge is

$$\int_0^1 u^j (1-u)^{n-j} \frac{\Lambda(du)}{u^2}.$$

One can obtain the Lambda-Fleming-Viot process of Definition 5.3.2 as a limit of a non-spatial version of the model described in Section 5.2. We start from a Poisson number of individuals with parameter $\lambda$. Events are driven by a Poisson Point Process $\Pi$ on $\mathbb{R}_+ \times [0,1]$ with intensity $dt \otimes \Lambda(du)/u^2$. When an event falls, we sample a parent uniformly at random from the population immediately before the event; each individual in the population, independently, is removed with probability $u$; and we replenish the population with an independent $\texttt{Poiss}(\lambda u)$ number of offspring, of the same type as the parent. Since the variance of a Poisson random variable is equal to its mean, if we measure the population in units of size $\lambda$, at time zero its size is $1 + \mathcal{O}(1/\sqrt{\lambda})$, the number removed during the reproduction event is $u + \mathcal{O}(1/\sqrt{\lambda})$, and the number of offspring is $u + \mathcal{O}(1/\sqrt{\lambda})$. We see that as $\lambda \to \infty$, the population size becomes fixed at 1 and at an event $(t, u) \in \Pi$ a proportion exactly $u$ of the population is replaced. In other words, the limiting dynamics are those of the Lambda-Fleming-Viot process of Definition 5.3.2. By choosing the replacement mechanism to be other than Poisson, we can obtain other limits, c.f. Section 4.2 of Etheridge & Kurtz [38].

## 5.4 The spatial Lambda-Fleming-Viot Process and its dual

The spatial Lambda-Fleming-Viot process (SLFV), introduced in Etheridge [34] and rigorously constructed in Barton, Etheridge and Véber [5], is an obvious extension of the Lambda-Fleming-Viot model of the last section. Once again the dynamics are driven by a Poisson Point Process $\Pi$, but now, in addition to specifying the time and impact of an event, a point of $\Pi$ encodes the spatial region that will be affected by the event.

**Definition 5.4.1 (Spatial Lambda-Fleming-Viot process (SLFV)).** *The spatial Lambda-Fleming-Viot process (SLFV) $\{\rho(t, x, \cdot); x \in \mathbb{R}^d, t \geq 0\}$, specifies a probability measure on the type space $K$ for each $x \in \mathbb{R}^d$ at each time $t \geq 0$. Its dynamics are driven by a Poisson Point Process*

*on* $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+ \times (0,1]$, *with intensity* $dt \otimes dx \otimes \xi(dr, du)$, *where* $\xi(dr, du) = \mu(dr)\nu_r(du)$ *with* $\mu$, $\nu$ *as in the individual based model of Section 5.2.*

*At a point* $(t, x, r, u) \in \Pi$, *a reproduction event takes place in the closed ball* $B(x,r)$ *with centre* $x$ *and radius* $r$. *During the event:*

1. *select a location* $z$ *for the parent uniformly from* $B_r(x)$; *that is* $z \sim U(B_r(x))$;

2. *select a type* $k$ *according to the distribution of types at* $z$ *immediately before the event; that is* $k \sim \rho(t-, z, \cdot)$;

3. *for all* $y \in B(x, r)$,

$$\rho(t, y, \cdot) = (1 - u)\rho(t-, y, \cdot) + u\delta_k.$$

*Sites outside* $B(x, r)$ *are not affected, so*

$$\rho(t, y, \cdot) = \rho(t-, y, \cdot), \qquad \forall y \notin B(x, r).$$

**Remark 5.4.2.** *Later on we shall be a bit more careful and specify the state space of the SLFV carefully, at least in the special case when the type space is* $P$, $Q$, *which we identify with* $K = \{0, 1\}$.

Note that all offspring are of the type of an individual that we sampled from $z$. In order for the process to exist, it is convenient to take the condition that we wrote down for our individual based model:

$$\int_0^\infty \int_0^1 ur^d \nu_r(du)\mu(dr) < \infty, \tag{5.3}$$

'though in fact this can be weakened somewhat (Etheridge & Kurtz [38]).

We can also write down the genealogical trees relating individuals in a sample from the population, and in view of the duality between the Lambda-Fleming-Viot process and the Lambda-coalescent, the form of the dual should come as no surprise.

Once again, we think of time for the Poisson Point Process $\Pi$ that dictates the dynamics of the model as extended to the whole real line, so that it is reversible. Tracing *backwards* in time, at each point $(t, x, r, u) \in \Pi$, each ancestral lineage that is currently in $B(x, r)$ will (independently) be marked with probability $u$. All marked lineages jump to the location of the parent of the event, $z \sim U(B(x, r))$. Crucially, this is true even if there is only a single lineage in the ball - it is this that maintains sampling consistency.

If we follow a single lineage, it evolves in a sequence of jumps with intensity

$$dt \otimes \left( \int_{[|x|/2, \infty)} \int_{(0,1]} \frac{L_r(x, 0)}{V_d(r)} u\nu_r(du)\mu(dr) \right) dx, \tag{5.4}$$

where $V_d(r)$ is the volume of a ball of radius $r$ in $\mathbb{R}^d$ and $L_r(x,0) = \mathrm{Vol}\big(B(x,r) \cap B(0,r)\big)$. To see why, suppose (without loss of generality) that the lineage is currently at the origin. It can only jump to $x$ if both 0 and $x$ are in the region affected by an event. For fixed $r$, the Lebesgue measure of the possible positions for the centres $y$ such that $\{0,x\} \in B(y,r)$ is $L_r(x,0)$ (which, in particular, is non-zero only if $r > |x|/2$). If 0 and $x$ are both covered by the same event, then the chance that the lineage is an offspring of the event is $u$, and the chance that $x$ is the location of the parent of the event is $dx/V_d(r)$. Integrating with respect to the intensity of $\Pi$ yields (5.4).

Two lineages currently at separation $y$ will coalesce if they are *both* affected by an event, which will happen at instantaneous rate

$$\int_{[|y|/2,\infty)} L_r(y,0) \int_{(0,1]} u^2 \nu_r(du)\mu(dr).$$

Evidently, if this is bounded, so is the rate of all other coalescence events, and since $L_r(x,0)$ is of order $r^d$ for large $r$, and $u < 1$, this is guaranteed under our condition (5.3).

The joint convergence of our individual based model to the SLFV and the embedded genealogical trees to the process of coalescing random walks described above is proved in Etheridge & Kurtz [38].

The moment duality between the stepping stone model and the structured coalescent will be mirrored in the SLFV setting. We defer a formal statement until we have specialised to $K = \{P,Q\}$. First we examine the behaviour of our coalescent when we sample uniformly at random from a population living on a large torus. In particular, we are interested in investigating the relationship between large scale extinction-recolonisation events and the effective population size, in an effort to substantiate our claim that such events could provide at least a partial explanation for lower than expected genetic diversity.

## 5.5   Large scale events and genetic diversity

At the end of Section 3.3 we described a result of Zähle, Cox, & Durrett [82] that says that if we sample a finite number of individuals uniformly at random from a population evolving according to the stepping stone model on the torus of side $L$ in $\mathbb{Z}^2$, then measuring time in units of $L^2 \log L$ generations, the genealogy of the sample converges to the Kingman coalescent. We emphasize that this timescale grows with the size of the range, and, in particular, if the population density is held constant, the timescale grows faster than the census population size, which will grow like $L^2$. Spatial structure on its own results in an effective population size which is greater than the census population size. To explain observed

effective population sizes, other effects must come into play. Our aim is to investigate the reduction in genetic diversity (or equivalently effective population size) resulting from large scale extinction recolonisation events in the SLFV. We mimic the setting of Zähle, Cox & Durrett [82] and work with a large torus in $\mathbb{R}^2$.

We write $\mathbb{T}(L)$ for the torus of side $L$ in $\mathbb{R}^2$. We shall consider two types of event. Small events will affect uniformly bounded regions. The rate at which we see small events of radius $r$ will be governed by a $\sigma$-finite measure $\mu_s(dr)$ on $[0, R_s]$. Large events will affect regions with a radius of $\mathcal{O}(L^\alpha)$ for some $0 \leq \alpha \leq 1$. The rate at which we see events of radius $L^\alpha r$ will be determined by a $\sigma$-finite measure $\mu_B(dr)$ on $[0, R_B]$.

More precisely, the dynamics of our population will be driven by *two* Poisson point processes:

1. small events are driven by $\Pi_L^s$, a Poisson point process on $\mathbb{R}_+ \times \mathbb{T}(L) \times [0, R_s] \times [0, 1]$ with intensity $dt \otimes dx \otimes \xi_s(dr, du)$ where $\xi_s(dr, du) = \mu_s(dr)\nu_r^s(du)$;

2. large events are driven by $\Pi_L^B$, a Poisson point process on $\mathbb{R}_+ \times L^{-\alpha}\mathbb{T}(L) \times [0, R_B] \times [0, 1]$ with intensity $\frac{1}{\rho_L}dt \otimes dx \otimes \xi_B(dr, du)$ where $\xi_B(dr, du) = \mu_B(dr)\nu_r^B(du)$.

The reproduction mechanism is as before except that at a point $(t, x, r, u)$ of $\Pi_L^B$, a reproduction event takes place in the ball centred at $L^\alpha x$ and of radius $L^\alpha r$. By defining $\Pi_L^B$ on the scaled torus, we have arranged that the parameter $\rho_L$ determines the relative frequency at which an ancestral lineage will be covered by small and large events and therefore their relative importance in shaping the genealogy of a sample. Thus the rate at which a lineage is hit by a small event is $\mathcal{O}(1)$. The rate at which it is hit by a large event is $\mathcal{O}(1/\rho_L)$, but when it is affected by such an event, it jumps a distance $\mathcal{O}(L^\alpha)$.

Since the sum of two Poisson processes is again a Poisson process, this is of course just the same model as before - we have just chosen to consider the events in two parts.

**Remark 5.5.1.** *Note that the effect of large scale events is very different from adding long range dispersal in the stepping stone model. Rather than a single offspring being born at a large displacement from its parent, here, as a result of a large scale event, offspring of a single parent replace a proportion of the population at* every *point within a large ball. We can approximate the stepping stone model with long range dispersal within this framework, but to do so we must scale in such a way that the impact of large events is small.*

The results that follow are from Amandine Véber's thesis and were published in Barton, Etheridge & Véber [5]. We're not going to give any

details of the proofs, but instead we'll try to explain why the results should be true. The results in the paper come hand in hand with exact constants that we don't make explicit here.

Without the large events, the model is very much like the stepping stone model and so it should come as no surprise that on timescales of $\mathcal{O}(L^2 \log L)$ the genealogy of a uniform sample from $\mathbb{T}(L)$ should converge to the Kingman coalescent.

Our first result says that for any $\alpha < 1$ the genealogy of a uniform sample will still be close to a Kingman coalescent, but for $\alpha > 0$ the timescale can depend on both big and small events. Roughly, if $\rho_L \gg L^2 \log L$, then we won't notice the big events; ancestral lineages will all coalesce due to small events before we see a big event. Since $L^{2\alpha}$ small events have a comparable effect to one big event, if $\rho_L \approx L^{2\alpha}$ then we see the effect of both. If $\rho_L \ll L^{2\alpha}$, then the coalescent is determined by the large events and the timescale is determined by the time that it takes for lineages jumping through big events to come close enough together that they can coalesce. The small events are negligible and the result is as though there were none and we had simply rescaled the torus and worked on $\mathbb{T}(L^{-\alpha})$.

To state a more precise result, let $\sigma_s^2$ (resp. $\sigma_B^2 L^{2\alpha}/\rho_L$) denote the variance in the displacement of a single ancestral lineage in one time unit due to small (resp. large) events. (Note that the rate at which a lineage is affected by big events scales with $1/\rho_L$ and the variance of each resulting jump scales with $L^{2\alpha}$.)

**Theorem 5.5.2 (Special case of Theorem 3.3 of Barton, Etheridge & Véber [5]).** *Define*

$$
\omega_L = \left\{
\begin{array}{ll}
\frac{(1-\alpha)\rho_L L^2 \log L}{2\pi \sigma_B^2 L^{2\alpha}} & \text{if } \frac{L^{2\alpha}}{\rho_L} \to \infty, \\[2ex]
\frac{(1-\alpha)L^2 \log L}{2\pi(\sigma_s^2 + b\sigma_B^2)} & \text{if } \frac{L^{2\alpha}}{\rho_L} \to b \in [0, \infty), \\[2ex]
\frac{L^2 \log L}{2\pi \sigma_s^2} & \text{if } \frac{L^2 \log L}{\rho_L} \to 0.
\end{array}
\right.
$$

*Then if we measure time in units of $\omega_L$ the genealogy of a uniform random sample from $\mathbb{T}(L)$ converges in law to the Kingman coalescent.*

In particular, we see that big events can change the effective population size, even though the genealogy is asymptotically determined by a Kingman coalescent.

For $\alpha = 1$, so that big events affect a significant proportion of the species range, the picture is mathematically much richer:

**Theorem 5.5.3 (Summary of Theorem 3.7 of Barton, Etheridge & Véber [5]).** *Suppose that $\alpha = 1$.*

1. *if $\rho_L/L^2 \to b$, rescaling the torus by $1/L$, on timescale $L^2$ the genealogy converges to a* spatial *Lambda-coalescent on $\mathbb{T}(1)$ in which, between mergers, lineages follow independent Brownian motions;*

2. *if $\rho_L/(L^2 \log L) \to \beta$, on timescale $L^2 \log L$ the genealogy converges to a (non-spatial) Lambda-coalescent;*

3. *if $\rho_L \gg L^2 \log L$ then on timescale $L^2 \log L$ the genealogy converges to a Kingman coalescent.*

In the first case, lineages don't equilibrate in between big events, so we still see the spatial structure in the limit. The Lambda-coalescent in the second case has a Kingman component. In the third case, lineages have all coalesced through small events before they are covered by a big event.

We earlier claimed that including large scale extinction-recolonisation events, could also explain the decay in the probability of identity that we see in real data, in which the approximately exponential decay over intermediate scales is replaced by a slower rate over large scales. Figure 5.2 is Figure 5 of Barton, Kelleher & Etheridge [7]. It is a simulation of the two-dimensional version of our model. (In fact it is a simulation of a slightly different version of the model than the one described above, but we expect the behaviour to be the same.) It shows the logarithm of the probability of identity of two individuals as a function of their separation $x$ under three different scenarios: just small events, just large events and a mixture of the two. With just one size of event, we see the approximately exponential decay of the Wright-Malécot formula, one of the characteristics one would hope for from a 'continuum stepping stone model'. When we have a mixture of small and large events we see the rate of decay of identity decrease at large spatial separation, suggesting that large scale extinction/recolonisation events really do provide one possible explanation of the pattern of large-scale correlations in observed allele frequencies.

## 5.6   Duality between the SLFV and a spatial Lambda coalescent

So far we have not been very precise about the state space for our process. It is often convenient, especially when proving convergence theorems, to think of the SLFV as a measure-valued process. We shall now carefully write down the space of measures and the moment duality with our spatial Lambda-coalescent in the most studied case, when the space of possible genetic types is $K = \{P, Q\}$ as in our earlier models. It is convenient to identify $\{P, Q\}$ with $\{0, 1\}$. We suppose that the population evolves in $\mathbb{R}^d$ (although the space of geographical locations could equally, for example, be taken to be some subset of $\mathbb{R}^d$, or a $d$-dimensional torus).

At each time $t$, the population is represented by a measure $M_t$ on $\mathbb{R}^d \times K$ whose first marginal is Lebesgue measure on $\mathbb{R}^d$. This corresponds to assuming that the population density is constant across $\mathbb{R}^d$ and for any
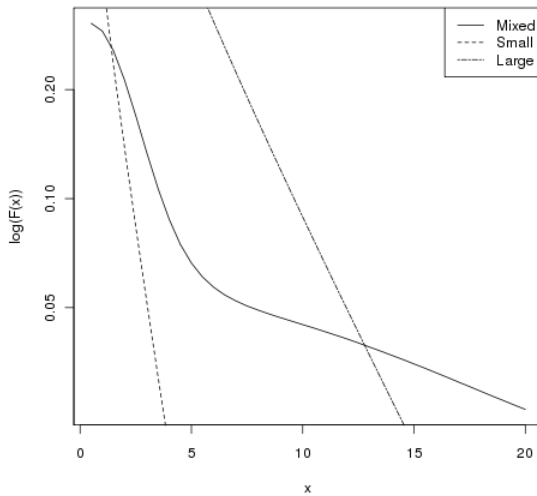
Figure 5.2: Figure 5 of Barton, Kelleher & Etheridge [7] showing the results of a simulation of the SLFV. The plots show the logarithm of the probability of identity of two individuals as a function of their spatial separation under three different scenarios: just small events, just large events and a mixture of the two. The impact of the large events in this last case is seen by the replacement of one initially approximately exponential rate of decay by a slower rate of decay at larger initial separations.

measurable subset $E$ of $\mathbb{R}^d$ and $\kappa \in \{0, 1\}$, $\mathrm{Vol}(E)^{-1}M_t(E \times \{\kappa\})$ gives the proportion of individuals of type $\kappa$ in $E$. The space

$$\mathcal{M}_\lambda := \Big\{ M \text{ measure on } \mathbb{R}^d \times \{0, 1\} : \ \forall f \in C_c(\mathbb{R}^d),$$
$$\int_{\mathbb{R}^d} f(x)M(dx, d\kappa) = \int_{\mathbb{R}^d} f(x)dx \Big\} \qquad (5.5)$$

of such measures is equipped with the topology of vague convergence, which makes it a compact set (c.f. Lemma 1.1 in Véber & Wakolbinger [79]). We have used $C_c(\mathbb{R}^d)$ to denote the space of all compactly supported continuous functions on $\mathbb{R}^d$. A standard disintegration theorem implies the existence of a measurable mapping $w_t : \mathbb{R}^d \to [0, 1]$ such that

$$M_t(dx, d\kappa) = \big(w_t(x)\delta_0(d\kappa) + (1 - w_t(x))\delta_1(d\kappa)\big)\, dx. \qquad (5.6)$$

We think of $w_t(x)$ as representing the local fraction of individuals of type 0 at site $x \in \mathbb{R}^d$ at time $t$, and we abuse terminology and call it the 'density' of $M$. Note that $w_t$ is only defined up to a Lebesgue null set, that is two mappings $w_t$ and $\tilde{w}_t$ will be equivalent if and only if

$$\mathrm{Vol}\big(\{x \in \mathbb{R}^d : w_t(x) \neq \tilde{w}_t(x)\}\big) = 0.$$

We often express everything in terms of a representative $w$ of the measure-valued evolution, but when we write down, for example, scaling limits, we are typically starting from a measure-valued process and proving weak convergence in the space of measure-valued processes.

The SLFV, thought of as a measure-valued process, can be characterised through its generator acting on test functions of the form

$$\Psi_{F,f}(M) := F(\langle f, w \rangle) = F\left( \int_{\mathbb{R}^d \times \{0,1\}} f(x)\mathbf{1}_{\{0\}}(\kappa)M(dx, d\kappa) \right), \quad (5.7)$$

where $w$ is any representative of the density of $M$, $f \in C_c(\mathbb{R}^d)$ and $F \in C^1(\mathbb{R})$ (the space of all continuously differentiable functions on $\mathbb{R}$). These test functions are dense in $C(\mathcal{M}_\lambda)$ (in the supremum norm topology).

For $w : \mathbb{R}^d \to [0, 1]$, $x \in \mathbb{R}^d$, $r > 0$ and $u \in [0, 1]$, define

$$\Theta^+_{x,r,u}(w) := \mathbf{1}_{B(x,r)^c}w + \mathbf{1}_{B(x,r)}((1 - u)w + u), \qquad \text{and}$$
$$\Theta^-_{x,r,u}(w) := \mathbf{1}_{B(x,r)^c}w + \mathbf{1}_{B(x,r)}(1 - u)w. \qquad (5.8)$$

These correspond to the value of $w$ immediately after a reproduction event of impact $u$ affecting $B(x, r)$, with '+' indicating that the parent was type 0 ($P$) and '−' indicating that it was type 1 ($Q$).

Writing as usual $V_d(r)$ for the volume of a ball of radius $r$, we can now write down the generator for the SLFV driven by a Poisson Point Process $\Pi$ of intensity $dt \otimes dx \otimes \mu(dr)\nu_r(du)$:

$$
\mathcal{L}\Psi_{F,f}(M) = \int_{\mathbb{R}^d} \int_0^\infty \int_0^1 \int_{B(x,r)} \frac{1}{V_d(r)} \Big[ w(y)F(\langle f, \Theta_{x,r,u}^+(w) \rangle)
$$
$$
+ (1 - w(y))F(\langle f, \Theta_{x,r,u}^-(w) \rangle) - F(\langle f, w \rangle) \Big] dy\, \nu_r(du)\, \mu(dr)\, dx.
$$

(5.9)

This is just the generator of a jump process.

## Duality

In our original description of the dual process of coalescing ancestral lineages, we argued heuristically that we could reverse the Poisson Point Process of events with respect to time. The simplest route to a rigorous statement of duality is in terms of generators, and yields 'weak duality' in the sense of Section 3.2. This will later extend to the case in which we include natural selection in our models.

First we define the process $(\Xi_t)_{t\geq 0}$ of 'ancestral lineages' driven by an *independent* copy, $\widetilde{\Pi}$, of $\Pi$ (the Poisson Point Process that drove reproduction events).

Let $\mathcal{M}_p(\mathbb{R}^d)$ denote the set of all finite point measures on $\mathbb{R}^d$, which we endow with the topology of weak convergence. The process $(\Xi_t)_{t\geq 0}$ will take its values in $\mathcal{M}_p(\mathbb{R}^d)$: each atom of $\Xi_t$ will represent the location of an ancestor, $t$ units of time in the past.

**Definition 5.6.1.** *Let $\Xi^0 \in \mathcal{M}_p(\mathbb{R}^d)$, and let us define the $\mathcal{M}_p(\mathbb{R}^d)$-valued process $(\Xi_t)_{t\geq 0}$ with initial value $\Xi^0$ as follows. We set $\Xi_0 = \Xi^0$ and, for convenience, at every time $t \geq 0$ we write*

$$
\Xi_t = \sum_{i=1}^{N_t} \delta_{\xi_t^i},
$$

*where $N_t = \Xi_t(\mathbb{R}^d)$ and some of the $\xi_t^i$ may be identical. Note that the ordering by $1, \ldots, N_t$ of the atoms is arbitrary and will play no rôle in the updating of $\Xi_t$.*
*For every $(t, x, r, u) \in \widetilde{\Pi}$:*

1. *each $\xi_{t-}^i \in B(x,r)$, independently, is given a mark with probability $u$ (or not with probability $1 - u$);*

2. *if at least one of the $\xi_{t-}^i$ is marked, to form $\Xi_t$ remove all the marked atoms from $\Xi_{t-}$ and add a new atom at a location drawn uniformly at random from within $B(x,r)$.*

*If no $\Xi_{t-}$ is marked, then nothing happens.*

The point measure $\Xi_t$ always has at least one atom.

The difficulty that we face in establishing a duality relation between our SLFV $(M_t)_{t\geq 0}$ and the process of coalescing lineages $(\Xi_t)_{t\geq 0}$ is that the density $w_t$ of the SLFV is only defined Lebesgue a.e. and so the obvious candidate for a duality function will not make sense. However, if, instead of taking deterministic points $x_1, \ldots, x_k$, we take random points, with a distribution which has a density $\psi$ with respect to Lebesgue measure on $(\mathbb{R}^d)^k$, then we can establish the corresponding duality.

**Proposition 5.6.2.** *Any $\mathcal{M}_\lambda$-valued Markov process $(M_t)_{t\geq 0}$ with generator $\mathcal{L}$ defined by (5.9) is dual to the process $(\Xi_t)_{t\geq 0}$, in the sense that for every $k \in \mathbb{N}$, $\psi \in C((\mathbb{R}^d)^k) \cap \mathbb{L}^1((\mathbb{R}^d)^k)$, $M_0 \in \mathcal{M}_\lambda$ and $t \geq 0$, we have for any choice of the representatives $w_0$ (resp., $w_t$) of the density of $M_0$ (resp., $M_t$):*

$$
\mathbb{E}_{M_0}\left[\int_{(\mathbb{R}^d)^k} \psi(x_1, \ldots, x_k)\left\{\prod_{j=1}^k w_t(x_j)\right\} dx_1 \cdots dx_k\right]
$$

$$
= \int_{(\mathbb{R}^d)^k} \psi(x_1, \ldots, x_k)\mathbb{E}_{\Xi[x_1,\ldots,x_k]}\left[\prod_{j=1}^{N_t} w_0(\xi_t^j)\right] dx_1 \cdots dx_k. \tag{5.10}
$$

**Remark 5.6.3.** *The right hand side of the duality relation is defined and independent of the choice of representative of the density $w_0$. This follows because if the law of $(\xi_0^1, \ldots, \xi_0^k)$ is absolutely continuous with respect to Lebesgue measure on $(\mathbb{R}^d)^k$, so too is the law of $\xi_t^1, \ldots, \xi_t^{N_t})$ on $(\mathbb{R}^d)^{N_t}$ (conditional on $N_t$). We do however need to include the 'extra' integration with respect to $\psi$ to ensure that everything is well defined.*

# Chapter 6

# Natural selection

## 6.1 Selection in the Wright-Fisher and Moran models

There are many ways of introducing selection into a model for allele frequencies, and indeed selection can take many forms. We begin with the simplest, in which there are just two alleles, $P$ and $Q$ say, and individuals carrying the $Q$-allele see a slight increase in their fecundity relative to those carrying the $P$-allele. For now, we assume a haploid population, with no spatial structure.

Recall that in the Wright-Fisher model, during reproduction, each individual produces a large number of juvenile offspring which go into an effectively infinite pool from which the next generation is sampled. To incorporate selection, suppose that individuals carrying the $Q$-allele produce $(1 + s)$-times as many juveniles as those carrying the $P$-allele. Thus, if the proportion of $P$-individuals at generation $t$ is $p$, the proportion of type $P$ juveniles is

$$p^* = \frac{p}{p + (1 + s)(1 - p)} = \frac{p}{1 + s(1 - p)},$$

for some $s > 0$, which we assume to be small. The number of type $P$ offspring will then be $\texttt{Binom}(N, p^*)$ and so, in particular, the expected proportion of type $P$ offspring is $p^*$.

Note that the family sizes will follow a multinomial distribution (c.f. Remark 2.2.2), but the weights will not be equal.

We may write

$$p^* - p = \frac{p}{1 + s(1 - p)} - p = -sp(1 - p) + \mathcal{O}(s^2),$$

and so to arrive at a diffusion limit for the allele freqencies, measuring
time in units of $N$ generations as we did in the neutral case, we must take
$s$ to be $\mathcal{O}(1/N)$. Indeed, setting $s = \overline{s}/N$ and writing $\Delta p$ for the change
in allele frequencies over a single generation, we find

$$
\begin{aligned}
\mathbb{E}[\Delta p] &= -\frac{1}{N}\overline{s}p(1-p) + \mathcal{O}\big(\frac{1}{N^2}\big), \\
\mathbb{E}[(\Delta p)^2] &= \frac{1}{N}p(1-p) + \mathcal{O}\big(\frac{1}{N^2}\big), \\
\mathbb{E}[(\Delta p)^4] &= \mathcal{O}\big(\frac{1}{N^2}\big),
\end{aligned}
$$

and as $N \to \infty$, using the results of Section 2.3, the process of allele
frequencies converges to the solution of the s.d.e.

$$
dp = -\overline{s}p(1-p)dt + \sqrt{p(1-p)}dW_t. \tag{6.1}
$$

## The ancestral selection graph (ASG)

It is straightforward to find a moment dual for this system. We mimic
what we did in Section 3.2 and apply Itô's formula to $p^n$ (with $p$ varying
and $n$ fixed),

$$
\begin{aligned}
d(p^n) &= np^{n-1}dp + \binom{n}{2}p^{n-2}d\langle p\rangle \\
&= \Big(np^{n-1}\big(-\overline{s}p(1-p) + \binom{n}{2}p^{n-2}p(1-p)\big)\Big)dt + \text{martingale} \\
&= \Big(n\overline{s}\big(p^{n+1} - p^n\big) + \binom{n}{2}\big(p^{n-1} - p^n\big)\Big)dt + \text{martingale}.
\end{aligned}
$$

We now think of $p$ as fixed and seek a process $(n_t)_{t\geq 0}$ satisfying the same
equation up to a mean zero martingale. We see that we should choose
$(n_t)_{t\geq 0}$ to be the birth-death process with rates

$$
n \mapsto n+1 \text{ at rate } n\overline{s}; \qquad n \mapsto n-1 \text{ at rate } \binom{n}{2}.
$$

In other words if we think of $n$ as counting particles, then each particle
branches at rate $\overline{s}$ and each *pair* of particles coalesces at rate one. It is
elementary to check that

$$
\mathbb{E}[p(t)^{n(0)}] = \mathbb{E}[p(0)^{n(t)}].
$$

Of course, since in a haploid population the number of ancestors in
a sample can only go down as we trace backwards in time, the dual
process does not have the same interpretation as the coalescent. Strictly,
all we have done is establish a representation for the finite-dimensional
distributions of $(p_t)_{t\geq 0}$.

## The Moran model with selection

We can arrive at the same limiting model for allele frequencies (and the same branching and coalescing dual) via the Moran model with selection. Recall that in the neutral Moran model, at rate $\binom{N}{2}$ a random pair is selected from the population, one dies and the other reproduces. To mimic the effect of reproduction in the Wright-Fisher model above, we bias the choice of parent. Thus if a pair is picked consisting of one type $P$ and one type $Q$ individual, with probability $(1+s)/2$ it is the type $Q$ that reproduces. It is sometimes convenient to think of there being two types of event: neutral events which occur at rate $(1-s)\binom{N}{2}$ and 'potential selective events' which happen at rate $s\binom{N}{2}$. At a potential selective event, if the pair of individuals chosen consists of one $P$ and one $Q$, then with probability *one* it is the $Q$ that reproduces.

Once again taking $s = \bar{s}/N$ and letting $N \to \infty$ we obtain (6.1) as a limiting model for allele frequencies. To see why, consider the generator of the continuous time Markov chain that governs the proportion $p$ of $P$-alleles.

$$
\begin{aligned}
\mathcal{L}^N f(p) &= \left(1 - \frac{\bar{s}}{N}\right)\binom{N}{2} p(1-p)\left(f\left(p + \frac{1}{N}\right) + f\left(p - \frac{1}{N}\right) - 2f(p)\right) \\
&\quad + \frac{\bar{s}}{N}\binom{N}{2} 2p(1-p)\left(f\left(p - \frac{1}{N}\right) - f(p)\right) \\
&= \frac{1}{2}p(1-p)f''(p) - \frac{\bar{s}}{N}\binom{N}{2}2p(1-p)\frac{1}{N}f'(p) + \mathcal{O}\left(\frac{1}{N}\right) \\
&\to \frac{1}{2}p(1-p)f''(p) - \bar{s}p(1-p)f'(p),
\end{aligned}
$$

which is the generator of the diffusion corresponding to (6.1).

The moment duality is easy to understand in the Moran context. If we take a sample of size $n(0)$ from the population at time $t$, then $\mathbb{E}[p(t)^{n(0)}]$ is the probability that they are all of type $P$. Neutral events correspond to coalescence of ancestral lineages, as before, but when one of our lineages is hit by a potential selective event - which happens in the limit at rate $\bar{s}$ per lineage - in order for that lineage, which is the *offspring* of the event, to be type $P$, *both* of the individuals sampled in the event must have been of type $P$. To confirm that this was the case, we must trace back the ancestry of both individuals - hence the birth in the moment dual. All individuals in our sample are of type $P$ if and only if all the 'potential ancestors' in the moment dual at time zero were of type $P$, and this has probability $\mathbb{E}[p(0)^{n(t)}]$.

We have only discussed a weak duality. The branching and coalescing structure defined above describes the number of lineages in the *ancestral selection graph* (ASG), introduced in the companion papers Krone & Neuhauser [58] and Neuhauser & Krone [69]. In fact the 'true ancestry'

of the sample can be recovered from the full description of the ASG, but that is beyond our scope here. Our main focus is on understanding the interplay between natural selection and spatial structure.

## Establishment probability

First we describe a classical argument which can be traced back at least to Fisher [45]. Recall our Wright-Fisher model with selection, in which the family sizes are determined by multinomial sampling with weights proportional to the relative fitness of the parents.

In a population of $N$ genes, suppose that a single favourable mutation arises. What is the probability that it establishes? By 'establishes' we mean reaches appreciable frequency. If $N$ is large, for a favoured allele this is tantamount to fixation (that is, everyone eventually carrying the allele). If favoured alleles are rare, the number of offspring of a given favoured individual is $\approx \texttt{Binom}(N, (1+s)/N) \approx \texttt{Poiss}(1+s)$. Moreover, still while the mutation is rare and population size $N$ is large, the number of offspring is approximately independent for each favoured allele, and so we can approximate the total number of favoured alleles by a Galton-Watson branching process with $\texttt{Poiss}(1+s)$ offspring distribution.

Using this branching process approximation, the probability of extinction of the favoured type satisfies

$$x = \exp(-(1+s)(1-x)),$$

and so the survival probability, $y = 1 - x$, satisfies

$$y = 1 - \exp(-(1+s)y) = (1+s)y - \frac{1}{2}(1+s)^2 y^2 + \mathcal{O}(y^3).$$

Rearranging:

$$\frac{1}{2}(1+s)^2 y^2 \approx sy \quad \Longrightarrow \quad y \approx 2s,$$

where we have used that since $s$ is small, so is $y$.

One of the questions that we are going to ask is 'does space matter?' So, for example, if we have a favourable mutation arising in a spatially structured population, how sensitive is its establishment probability to the spatial structure?

Maruyama [61], consider a population that is subdivided into demes of (large) constant size. He supposed that the contribution of each deme to the next generation is proportional to its size. Under these assumptions, he found that the probability of establishment of a newly arising favoured allele is independent of population subdivision. These are quite strong assumptions on the spatial structure and Barton [1] showed that Maryuama's conclusion breaks down if one adds extinction-recolonisation events to colonies. Our next aim is to investigate the

interaction of natural selection and spatial structure in the much richer
and more varied framework of the SLFV.

## 6.2   Introducing selection to the SLFV

There are *lots* of ways to introduce selection to the SLFV, and we can fairly
easily cook up examples where two different mechansms lead to the same
process of allele frequencies, but different genealogies, Miller [64]. Here
we begin by mimicking the approach that we took above in the classical
models; that is we weight the selection of the parent during a reproduction
event in order to give a selective advantage to one of two possible genetic
types. Later we'll talk briefly about other forms of selection that do not
simply favour one type over another.

Suppose that there are just two genetic types in the population, $P$ and
$Q$, and that $Q$ has a small selective advantage. Recall that in the SLFV,
reproduction events are driven by a Poisson Point Process $\Pi$ with intensity
$dt \otimes dx \otimes \xi(dr, du) = dt \otimes dx \otimes \mu(dr)\nu_r(du)$. If $(t, x, r, u) \in \Pi$, so that the
region $B(x, r)$ is affected by a reproduction event, then mimicking what
we did above for the Wright-Fisher model, if the proportion of $P$-alleles
in the region at time $t-$ is $\overline{w}$, then the probability that the parent (and
hence all the offspring of the event) is type $P$ is

$$\frac{\overline{w}}{1 + s(1 - \overline{w})} = \overline{w} - s\overline{w}(1 - \overline{w}) + \mathcal{O}(s^2) = (1 - s)\overline{w} + s\overline{w}^2 + \mathcal{O}(s^2).$$

Just as for the Moran model, we see that we can achieve the same
probability of type $P$ offspring in two stages. For each event, with
probability $1 - s$ we declare it to be neutral, in which case reproduction
is exactly as in the neutral SLFV. With probability $s$ it is selective - in
which case we want the probability of the offspring all being type $P$ to
be $\overline{w}^2$, which we achieve by sampling two 'potential' parents uniformly at
random from the affected region $B(x, r)$ and only if both are type $P$ are
the offspring type $P$.

Using the thinning property of Poisson processes, we see that we can
think of the SLFV with selection as being driven by two *independent*
Poisson Point Processes, $\Pi^n$ and $\Pi^s$ with intensities $(1-s)dt \otimes dx \otimes \xi(dr, du)$
and $sdt \otimes dx \otimes \xi(dr, du)$ respectively. As in the neutral case, to make contact
with the notation in the original papers, we identify the type space $\{P, Q\}$
with $\{0, 1\}$.

**Definition 6.2.1 (SLFV with fecundity selection (SLFVS)).** *Let $\mu$
be a $\sigma$-finite measure on $(0, \infty)$, and let $\nu = \{\nu_r, \ r > 0\}$ be a collection of
probability measures on $[0, 1]$ such that*

$$\int_0^\infty \int_0^1 r^d u \nu_r(du)\mu(dr) < \infty. \tag{6.2}$$

*Further, let* $\Pi^N$ *and* $\Pi^S$ *be two independent Poisson point processes on* $\mathbb{R} \times \mathbb{R}^d \times (0, \infty) \times [0, 1]$ *with respective intensity measures* $(1 - s)dt \otimes dx \otimes \mu(dr)\nu_r(du)$ *and* $sdt \otimes dx \otimes \mu(dr)\nu_r(du)$.

*The* spatial $\Lambda$-Fleming-Viot process with selection (SLFVS) *with initial condition* $M^0 \in \mathcal{M}_\lambda$ *is an* $\mathcal{M}_\lambda$-*valued process* $(M_t)_{t \geq 0}$ *such that* $M_0 = M^0$, *and whose dynamics are given as follows. If* $(t, x, r, u) \in \Pi^N$, *a neutral event occurs at time* $t$, *within the closed ball* $B(x, r)$:

1. *Sample a type* $\kappa$ *according to the type distribution within* $B(x, r)$ *just before the event. That is,* $\kappa = 0$ *with probability* $V_d(r)^{-1}M_{t-}(B(x, r) \times \{0\})$, *where* $V_d(r)$ *is the volume of a $d$-dimensional ball of radius* $r$; *otherwise,* $\kappa = 1$.

2. *Update the value of* $M_t$ *within* $B(x, r)$ *by setting*

$$M_t\Big|_{B(x,r) \times \{0,1\}} := (1 - u)M_{t-}\Big|_{B(x,r) \times \{0,1\}} + u\,dx\Big|_{B(x,r)} \otimes \delta_\kappa.$$

*In other words, if* $w_{t-}(y)$ *is a representative of the density of* $M_{t-}$, *then a representative of the density of* $M_t$ *can be taken to be* $w_t(y) = w_{t-}(y)$ *if* $y \notin B(x, r)$, *and*

$$w_t(y) = (1 - u)w_{t-}(y) + u\mathbf{1}_{\{\kappa=0\}} \qquad \text{if } y \in B(x, r).$$

*Similarly, if* $(t, x, r, u) \in \Pi^S$, *a selective event occurs at time* $t$, *within the closed ball* $B(x, r)$:

1. *Sample two types* $\kappa$ *and* $\kappa'$ *independently, according to the type distribution within* $B(x, r)$ *just before the event. We interpret them as the types of two 'potential' parents.*

2. *Update the value of* $M_t$ *(only) within* $B(x, r)$ *by setting*

$$M_t\Big|_{B(x,r) \times \{0,1\}} := (1 - u)M_{t-}\Big|_{B(x,r) \times \{0,1\}} + u\,dx\Big|_{B(x,r)} \otimes \delta_{\max\{\kappa,\kappa'\}}.$$

*That is, the offspring are of type* $0$ *if and only if both potential parents are of type* $0$. *This time, a representative of the density of* $M_t$ *can be taken to be* $w_t(y) = w_{t-}(y)$ *if* $y \notin B(x, r)$, *and*

$$w_t(y) = (1 - u)w_{t-}(y) + u\mathbf{1}_{\{\kappa=\kappa'=0\}} \qquad \text{if } y \in B(x, r).$$

**Remark 6.2.2.** *There are many obvious ways in which to generalise this model even if we are only interested in genic selection. For example, Etheridge, Véber & Yu [40], where this process is introduced, considers different driving Poisson Point Processes for neutral and selective events. Preprint versions of the paper are available on the ArXiv.*
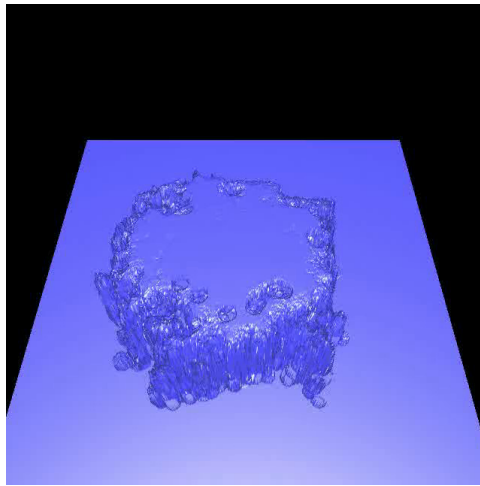
Figure 6.1: **The wave of advance of an advantageous allele under the SLFV with genic selection.** Simulation due to Jerome Kelleher. The population was seeded from a small region of type $Q$ individuals in a sea of type $P$. As time develops, as expected, the favoured type advances in an almost circular wave; behind the wavefront, the favoured type becomes fixed.

## Backwards in time

We can also write down a spatial version of the ASG. Evolution of the ancestry due to neutral events is exactly as before: lineages evolve in a series of jumps; they can coalesce when covered by the same event.

At selective events, each lineage covered by the event is marked with probability $u$. If at least one lineage is marked, then all marked lineages are replaced by two 'potential' ancestral lineages whose locations are chosen uniformly at random from $B(x, r)$.

Note that unlike the ASG that we wrote down in the non-spatial context, here we can see simultaneous branching and coalescence. Just as in the nonspatial context, it is useful to note that an individual is type $P$ if and only if all lineages in the corresponding ASG are type $P$ at any previous time. We can use this to investigate the establishment and spread of a favoured allele.

Figure 6.1 shows a snapshot of a population through which a favoured allele is spreading. The 'island' is the favoured type. As one might expect, the favoured type spreads in a roughly circular 'wave'; behind the wavefront the frequency of favoured individuals is eventually everywhere one.
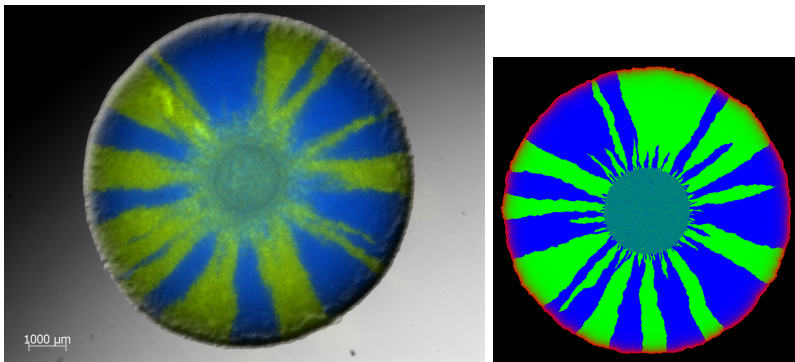
Figure 6.2: **Sectoring in an expanding population.** On the left hand side is a picture, courtesy of Kevin Foster of Oxford Zoology, of an expanding population of *Pseudomonas aeruginosa*. The population began with a drop of a mixture of cells labelled with two different fluorescent markers that do not effect the relative fitness of the cells. As the colony starts to expand, the population segregates into sectors of different colours. See Korolev et al. [57] for a full description of the experiments. On the right a simulation, by Jerome Kelleher, of the SLFV model, modified to capture range expansion.

## 6.3   Range expansion

Above we considered what happens when selecton is weak. At the other extreme one can ask what happens when selection is extremely strong. Thus, if an event falls on a region containing *any* type $Q$ individuals, then the parent is necessarily of type $Q$. Although not perhaps an interesting model of selection in biological populations, this can be thought of as a model of range expansion. We start from a cluster of individuals in our population. If an event falls on a region that contains any individuals, then a reproduction event takes place. This can be refined to include more than one type of individual in the expanding population.

In Figure 6.2 we consider two equally fit types in a population expanding into new territory. At time zero, we start from a small region in which both types are in a homogeneous mixture. As time progresses, the population develops 'sectors' of the two different types. This sectoring is ubiquitous in experiments on expanding colonies of bacteria. The population on the left in Figure 6.2 is a colony of *Pseudomonas aeruginosa*, that on the right is a simulation of our model. The similarity is striking. To obtain different numbers of sectors in our simulation, we can simply vary the ratio between the diameter of the original well-mixed population and that

of reproduction events.

## 6.4  Scaling limits of the SLFV with genic selection

The spread of a favoured allele is classically modelled through the (stochastic) Fisher-KPP equation:

$$dw = \Big( \frac{1}{2}\Delta w + sw(1-w) \Big) dt + \mathbf{1}_{d=1}\sqrt{\frac{1}{N_e}w(1-w)}W(dt, dx),$$

where $W$ is space-time white noise and $N_e$ is 'neighbourhood size'. One can think of this as a continuous space analogue of the stepping stone model with selection, c.f. Definition 3.1.4 in the neutral case. Once again, when we try to write down a continuum version of the equation in $d = 2$, we encounter a problem - the equation has no solution. However, our SLFV with selection makes sense in any spatial dimension and so one option is to use this model in place of the stochastic Fisher-KPP equation to capture the effects of genetic drift.

In order to justify this approach, it is natural to first ask whether over appropriate temporal and spatial scales the SLFV with selection will 'look like' a solution to the (stochastic) Fisher-KPP equation. The answer will be 'yes'. We won't prove this, but instead we'll demonstrate how to identify the appropriate scaling. A key tool is the spatial ASG. We'd like to find a scaling under which the motion of a single ancestral lineage converges to Brownian motion (to capture the Laplacian); in which lineages branch at rate one (to capture the 'logistic growth' term); and, in $d = 1$, for which coalescence is going to happen on the same timescale as branching (to retain the genetic drift corresponding to the stochastic term; recall from our discussion at the end of Section 4.4 that coalescence in the dual process reflects genetic drift in the forwards in time equation).

There are lots of interesting limits that one can take with large-scale extinction-recolonisation events, but our aim here is primarily to recover the classical model, so we shall suppose that the population evolves entirely through events that affect bounded regions in space. In fact it turns out that we don't really gain anything in the scaling limits that we care about by taking the radii of events to be random (with radius picked according to a measure of bounded support) and so we may as well take the radius of events to be fixed.

In the argument that follows, we shall once again appeal to Lemma 4.4.1, which told us that a simple random walk, started from zero, will make $\mathcal{O}(\sqrt{N})$ excursions away from zero before making one that reaches $\pm\sqrt{N}$. We also need the corresponding result in two dimensions. To see how it

will scale, we work by analogy with the radial part of a two-dimensional Brownian motion:

$$dR_t = \frac{1}{R_t}dt + dB_t,$$

where $B_t$ is a one-dimensional Brownian motion. This time we see that $\log R_t$ is a martingale and the same argument that we used to prove Lemma 4.4.1 suggests that it takes $\mathcal{O}(\log N)$ excursions away from zero for simple random walk to reach distance $\mathcal{O}(\sqrt{N})$ from the origin.

Although the constants will be different, these results should also reflect the number of excursions that a pair of ancestral lineages (under the SLFVS) will make away from regions where they have a chance to coalesce, before reaching a separation of order $\sqrt{N}$.

Armed with this, we return to understanding the scaling of our SLFVS. Let's set $u_n = u/n^\gamma$, $s_n = \bar{s}/n^\delta$, $w^{(n)}(t,x) = w(nt, n^\beta x)$, and try to identify the values of $\beta$, $\gamma$ and $\delta$ that lead to the classical (stochastic) Fisher-KPP equation in the limit.

We argue in terms of the behaviour of the ASG. A single ancestral lineage jumps at a rate proportional to $nu_n = n^{1-\gamma}$, and each jump is of size $\mathcal{O}(1/n^\beta)$. To recover a diffusive scaling (and hence Brownian motion of ancestral lineages in the limit), we take $2\beta = 1 - \gamma$.

Now we'd like to understand the branching. If we have a selective event, then the two resulting lineages are created at a separation of order $1/n^\beta$. With strictly positive probability they move apart without coalescing, but in order for us to 'see' this in the limit we need them to move apart to a distance of order 1 in scaled space (corresponding to a random walk making jumps of size order one escaping to distance order $n^\beta$). The argument above suggests that the number of excursions away from a region in which they can coalesce that they need before escaping to a distance of $\mathcal{O}(1)$ from one another is

1. $\mathcal{O}(n^\beta)$, in $d = 1$;

2. $\mathcal{O}(\log n)$, in $d = 2$;

3. $\mathcal{O}(1)$ in $d \geq 3$.

To control the chance that the lineages coalesce when they come back together, note that if two lineages are covered by the same event, given that one jumps, they coalesce with probability $\mathcal{O}(1/n^\gamma)$ (and after a small number of events in which only one jumps they will move apart). That is, it takes $\mathcal{O}(n^\gamma)$ returns together before they coalesce. So in $d \geq 2$ the probability that two lineages created in a selection event make a 'long' excursion away from one another before they coalesce is $\mathcal{O}(1)$.

In $d = 1$, we'd like to match the number of attempts required to reach separation $\mathcal{O}(1)$ with the number of attempts to coalesce, so that we'll see both branching and coalescence in the limit, and so we take $\beta = \gamma$.

Selection events cause a branch on a given lineage at rate $nu_n s_n$, so to make branching events occur at rate one in our limiting ASG, we take $nu_n s_n = \mathcal{O}(1)$; that is $1 - \gamma - \delta = 0$.

Putting this together, these arguments suggest that we should take $\beta = \gamma = 1/3$, $\delta = 2/3$. In Etheridge, Véber and Yu [40], it is shown that choosing $\xi(dr, du) = \delta_R \otimes \delta_u$ (so that all events have the same radius and impact), if we set $u_n = u/n^{1/3}$, $s_n = \bar{s}/n^{2/3}$, and set $w^{(n)}(t, x) = w(nt, n^{1/3}x)$, then as $n \to \infty$ we obtain convergence in an appropriate weak sense to the solution of the (stochastic in $d = 1$) Fisher-KPP equation. If we choose $1 - \gamma = 2\beta$, $1 - \gamma - \delta = 0$ and $0 < \beta < \delta$, then we obtain convergence to the deterministc Fisher-KPP equation in any dimension.

In fact we have to be a little careful. The proof requires some regularity, and it turns out to be convenient to instead show convergence of local averages. Let's state the result a little more carefully. Let $w_t$ be any representative of the density of the SLFVS. Define

$$\overline{w}_t^n(x) = \frac{n^{d/3}}{V_d(R)} \int_{B_n(x)} w_t^n(y) dy,$$

where $w_t^n(\cdot) = w_{nt}(n^{1/3}\cdot)$ and $B_n(x) = B(x, n^{-1/3}R)$. This scaling corresponds to scaling down the spatial coordinate by $n^{1/3}$ (so that distance one in the new units corresponds to distance $n^{1/3}$ in the original units); that is we are looking at spatial scale $n^{1/3}$ at time $n$. The random variable $\overline{w}_t^n(x)$ gives the local proportion of individuals of the unfavoured type 0 in a small neighbourhood (of radius $n^{-1/3}R$) of the point $x$ at time $t$ in these new units (corresponding to a ball of radius $R$ at time $nT$ in our original units).

We write $\overline{M}_t^n$ for the random measure (taking its values in $\mathcal{M}_\lambda$) with representative $\overline{w}_t^n$.

**Remark 6.4.1.** *Note that the process $\overline{M}_t^n$ is not Markov - the change in the value of $\overline{w}_t^n(y)$ due to an event affecting $B(x, R)$ will depend on the geometry of (and the genetic diversity within) the intersection $B(n^{1/3}y, R) \cap B(x, R)$.*

**Theorem 6.4.2 (Etheridge, Véber, Yu [40]).** *Suppose that $(\overline{M}_0^n)_{n \geq 1}$ converges in distribution to some $M_0 \in \mathcal{M}_\lambda$. Then, as $n \to \infty$, the process $(\overline{M}_t^n)_{t \geq 0}$ converges weakly in $D_{\mathcal{M}_\lambda}[0, \infty)$ towards a Markov process $(M_t^\infty)_{t \geq 0}$ with initial value $M_0^\infty = M_0$. The limiting process is characterised as follows. Let*

$$\Gamma_R = \frac{1}{V_d(R)} \int_{B(0,R)} \int_{B(x,R)} (z_1)^2 dz dx \qquad (6.3)$$

*(where $z_1$ denotes the first coordinate of $z$).*

*(i) When $d = 1$, $(M_t^\infty)_{t \geq 0}$ is the unique process for which, for every choice of the representative $w_s^\infty$ of the density of $M_s^\infty$ at every time $s$, and for every $f, g \in C_c^\infty(\mathbb{R})$,*

$$\mathcal{Z}^f := \left( \langle f, w_t^\infty \rangle - \langle f, w_0^\infty \rangle - \int_0^t \left\{ \frac{u\Gamma_R}{2} \langle \Delta f, w_s^\infty \rangle - 2Ru\bar{s} \langle f, w_s^\infty (1 - w_s^\infty) \rangle \right\} ds \right)_{t \geq 0}$$

*is a continuous zero-mean martingale with quadratic variation at time $t$ equal to*

$$4R^2 u^2 \int_0^t \langle f^2, w_s^\infty (1 - w_s^\infty) \rangle \, ds.$$

*Furthermore, the bracket process between $\mathcal{Z}^f$ and $\mathcal{Z}^g$ is given by*

$$\left[ \mathcal{Z}^f, \mathcal{Z}^g \right]_t = 4R^2 u^2 \int_0^t \langle fg, w_s^\infty (1 - w_s^\infty) \rangle \, ds.$$

*(ii) When $d \geq 2$, $(M_t^\infty)_{t \geq 0}$ is the unique (deterministic) process for which, for every choice of the representative $w_s^\infty$ of the density of $M_s^\infty$ at every time $s$, and for every $f \in C_c^\infty(\mathbb{R}^d)$ and $t \geq 0$,*

$$\langle f, w_t^\infty \rangle = \langle f, w_0^\infty \rangle + \int_0^t \left\{ \frac{u\Gamma_R}{2} \langle \Delta f, w_s^\infty \rangle - u\bar{s} V_d(R) \langle f, w_s^\infty (1 - w_s^\infty) \rangle \right\} ds.$$

Informally, in one space dimension, one can see the time-indexed family of densities of the limiting process $(M_t^\infty)_{t \geq 0}$ as a weak solution to the stochastic partial differential equation

$$\frac{\partial w}{\partial t} = \frac{u\Gamma_R}{2} \Delta w - 2Ru\bar{s}w(1 - w) + 2Ru\sqrt{w(1 - w)}\, \dot{W}$$

(independently of the representative chosen at every time $t$), where $W$ is a space-time white noise. In dimension $d \geq 2$, on the other hand, the noise term disappears in the limit and the time-indexed family of densities of $(M_t^\infty)_{t \geq 0}$ can be seen as a weak solution to the deterministic Fisher-KPP equation

$$\frac{\partial w}{\partial t} = \frac{u\Gamma_R}{2} \Delta w - u\bar{s} V_d(R)\, w(1 - w).$$

## Incorporating large-scale events

We can also obtain interesting limits when we allow for large-scale extinction recolonisation events. We choose these in such a way that when appropriately scaled, the motion of the ancestral lineages will sit in the domain of attraction of a symmetric stable process. For some $\alpha \in (1, 2)$, we set

$$\mu(dr) = \frac{\mathbf{1}_{\{r \geq 1\}}}{r^{d+\alpha+1}} \, dr, \tag{6.4}$$

and

$$\overline{w}_t^n(x) := \frac{1}{V_d(1)} M_{nt}\big(B(n^\beta x, 1) \times \{0\}\big) = \frac{1}{V_d(1)} \int_{B(n^\beta x, 1)} w_{nt}(y)\, dy, \quad (6.5)$$

with

$$\beta = \frac{1}{2\alpha - 1}, \quad \gamma = \frac{\alpha - 1}{2\alpha - 1} \quad \text{and} \quad \delta = \frac{\alpha}{2\alpha - 1}. \quad (6.6)$$

As before, we write $\overline{M}_t^n$ for the random measure (taking its values in $\mathcal{M}_\lambda$) with density $\bar{w}_t^n$.

We define

$$\Phi(|z - y|) := \int_{\frac{|z-y|}{2}}^{\infty} \frac{1}{r^{d+1+\alpha}} \frac{L_r(y, z)}{V_r}\, dr,$$

and for every $f \in C_c^\infty(\mathbb{R}^d)$ we set

$$\mathcal{D}^\alpha f(y) = u \int_{\mathbb{R}^d} \Phi(|z - y|)(f(z) - f(y)) dz. \quad (6.7)$$

It is checked in Lemma 6.1 of Etheridge, Véber & Yu [40] that this defines the infinitesimal generator of a symmetric stable process (that is, it is a constant multiple of the fractional Laplacian).

**Theorem 6.4.3 (Etheridge, Véber, Yu [40]). (Stable radii)** *Under the conditions above, suppose that $\overline{M}_0^n$ converges in distribution to some $M_0 \in \mathcal{M}_\lambda$. Then, as $n \to \infty$, the process $(\overline{M}_t^n)_{t \geq 0}$ converges weakly in $D_{\mathcal{M}_\lambda}[0, \infty)$ towards a Markov process $(M_t^\infty)_{t \geq 0}$ starting at $M_0$. Furthermore, if $\mathcal{D}^\alpha$ denotes the generator of the symmetric $\alpha$-stable process defined in (6.7), then*

*(i) When $d = 1$, $(M_t^\infty)_{t \geq 0}$ is the unique process for which, for every choice of the representative $w_s^\infty$ of the density of $M_s^\infty$ at every time $s$, and for every $f, g \in C_c^\infty(\mathbb{R})$,*

$$\mathcal{Z}^f := \left( \langle f, w_t^\infty \rangle - \langle f, w_0^\infty \rangle - \int_0^t \left\{ \langle \mathcal{D}^\alpha f, w_s^\infty \rangle - \frac{2u\bar{s}}{\alpha} \langle f, w_s^\infty(1 - w_s^\infty) \rangle \right\} ds \right)_{t \geq 0}$$

*is a continuous zero-mean martingale with quadratic variation at time $t$ equal to*

$$\frac{4u^2}{\alpha - 1} \int_0^t \langle f^2, w_s^\infty(1 - w_s^\infty) \rangle\, ds.$$

*Furthermore, the bracket process between $\mathcal{Z}^f$ and $\mathcal{Z}^g$ is given by*

$$[\mathcal{Z}^f, \mathcal{Z}^g]_t = \frac{4u^2}{\alpha - 1} \int_0^t \langle fg, w_s^\infty(1 - w_s^\infty) \rangle\, ds.$$

*(ii) When $d \geq 2$, $(M_t^\infty)_{t \geq 0}$ is the unique (deterministic) process for which, for every choice of the representative $w_s^\infty$ of the density of $M_s^\infty$ at every time $s$, and for every $f \in C_c^\infty(\mathbb{R}^d)$ and $t \geq 0$,*

$$\langle f, w_t^\infty \rangle = \langle f, w_0^\infty \rangle + \int_0^t \left\{ \langle \mathcal{D}^\alpha f, w_s^\infty \rangle - \frac{u\bar{s}V_d(1)}{\alpha} \langle f, w_s^\infty(1 - w_s^\infty) \rangle \right\} ds.$$

Our choice of values for $\beta$, $\gamma$ and $\delta$ in Theorem 6.4.3 is obtained by solving

$$1 - \gamma = \alpha\beta, \qquad 1 - \delta - \gamma = 0, \qquad \text{and } (\alpha - 1)\beta = \gamma$$

in order to obtain a limiting process $M^\infty$ which is stochastic in one dimension. If we replace the last condition by the inequality $0 < (\alpha-1)\beta < \gamma$, then (in any dimension) $(\overline{M}^n)_{n \geq 0}$ converges to a deterministic limit which is characterised as in the statement of Theorem 6.4.3(ii).

## Neighbourhood size revisited

In Theorem 6.4.2, $u_n \to 0$ as $n \to \infty$. This corresponds to high 'neighbourhood size'. We saw neighbourhood size before in the Wright-Malécot formula; roughly it is inversely proportional to the local coalescence rate. In a discrete time model, if we write $\eta(x)$ for the probability that two lineages at separation $x$ will coalesce in the previous generation, then in $d = 2$

$$\mathcal{N} = \frac{2\pi\sigma^2}{\int_{\mathbb{R}^2} \eta(x)dx},$$

where $\sigma^2$ is the variance of the jump of a lineage in a single generation. One can think of it as a measure of the number of 'potential parents' of an individual.

In continuous time, we define $\eta(x)$ to be the instantaneous rate of coalescence of two lineages at separation $x$. For the SLFV with fixed event size, this gives

$$\mathcal{N} = \frac{2\pi\sigma^2}{\int \eta(x)dx} = \frac{1}{u},$$

because

$$\eta(x) = \lambda u^2 \text{Vol}\big(B(0, R) \cap B(x, R)\big) = \lambda u^2 L_R(x, 0),$$

and one can calculate $\sigma^2 = \lambda u \pi R^4/2$ (see Barton et al. [4]).

## Establishment probability

Just as for the non-spatial case, one can ask about things like the fate of a newly arising favourable mutation. Consider $d = 1$. While the allele is rare, if

$$dw = \left(\frac{1}{2}\Delta w + \bar{s}w(1-w)\right)dt + \varepsilon\sqrt{w(1-w)}W(dt, dx),$$

we can approximate $1 - w \approx 1$, from which

$$dw \approx \left(\frac{1}{2}\Delta w + \bar{s}w\right)dt + \varepsilon\sqrt{w}W(dt, dx),$$

which is the density of a (supercritical) Dawson-Watanabe superprocess. This is of course heuristic, but Chetwynd-Diggle & Etheridge [20] show that in any dimension, if one traces the descent of a rare allele and scales appropriately, one can recover a superprocess sitting within the neutral SLFV and this can be extended to the case above.

The 'total mass', found by integrating against the constant function $f \equiv 1$, solves

$$dX \approx \bar{s}Xdt + \varepsilon\sqrt{X}dB_t,$$

a continuous state branching process, c.f. Remark 4.2.1(2). This suggests that establishment probability doesn't 'feel' the spatial structure when neighbourhood size is large. However, in a spatial continuum, neighbourhood size could be small (think of the number of potential parents) and then the situation is very different.

## Small neighbourhood size

In the previous results, we chose a parameter $n$ with which to scale time and expressed all the other scaling parameters in terms of $n$. We could equally think of the scaling as being relative to the neighbourhood size and taking an approximation when that neighbourhood size is very high. This would mirror using the population size as the scaling parameter in the Wright-Fisher model.

To understand how to interpret scalings when we hold neighbourhood size fixed, we return briefly to the biology. Recall that a geneticist uses differences between DNA sequences of individuals sampled from the population to infer something about the genealogical trees relating genes in the sample. So the neutral mutation rate $\mu$ sets the timescale over which we can observe the action of different forces of evolution.

Mutation rates are low, and scaling limits are insensitive to fine details of our models, so we scale relative to the neutral mutation rate. A natural question is then:

> When will we expect to see a signature of the existence of
> a favourable allele in data?

(That favourable allele may not have become 'fixed' in the population.)
To gain some insight into this, we consider the SLFV with selection and
*fixed* impact $u$. We shall use analogous arguments to those in the high
neighbourhood size scaling to understand under what circumstances we
can expect a nontrivial scaling limit (by which we mean one that differs
from the neutral case).

Fix $u \in (0,1)$. We shall once again only consider reproduction events
that affect bounded regions and, again, the possible scaling limits are all
captured if we take all reproduction events to affect regions of the same
radius. Set $n = 1/\mu$ and consider $w(nt, \sqrt{n}x)$. (Note that since $u$ is
fixed, $nu_n$ is order $n$ and so the scaling of space by $\sqrt{n}$ corresponds to the
diffusive scaling, and is forced upon us if we want a nontrivial limit for the
motion of ancestral lineages.)

When two branches arise in our ASG, they may coalesce again
(essentially) immediately, or they may move apart. In contrast to our
previous (large neighbourhood size) example, the number of excursions
that they make away from one another before coalescing does not change
under our scaling; it remains order one as we pass to the limit. We only
have a chance to 'see' that allele frequencies are different from those under
neutrality if we see mutations on branches of the ASG while they are
separated.

So in time units of order $n$, we need the rate at which we see branches
*that will separate to distance of order one (in scaled units)* appear in the
ASG to be order one. We can now use exactly the same heuristic as before.

At a branching event, lineages are born at separation order $1/\sqrt{n}$. The
chance that they separate to order one before coming back to a separation
at which they can coalesce is

1. $\mathcal{O}(1/\sqrt{n})$ in $d = 1$;

2. $\mathcal{O}(1/\log n)$ in $d = 2$;

3. $\mathcal{O}(1)$ in $d \geq 3$.

So in one dimension we need the overall rate at which branches appear
to be $\mathcal{O}(\sqrt{n})$ if ones that separate are to appear at rate $\mathcal{O}(1)$, whereas in
two dimensions we require rate $\mathcal{O}(\log n)$ and in $d \geq 3$ it suffices that they
appear at a rate of order one.

In other words, our ability to detect selection depends on dimension.
Branches appear at rate $uns_n$ and $u$ is fixed. One can prove that in

1. $d = 1$, selection is only visible if $s_n = \mathcal{O}(1/\sqrt{n})$, and if $u = 1$ the
   limiting ASG is then embedded in the Brownian net;

2. $d = 2$, selection is only visible if $s_n = \mathcal{O}(\log n/n)$, and the limiting ASG is a 'branching Brownian motion';

3. $d \geq 3$, selection is only visible if $s_n = \mathcal{O}(1/n)$, and the limiting ASG is branching Brownian motion.

If neighbourhood size is small, then the efficacy of selection depends on an individual's ability to escape the crowding from its own close family (c.f. the pain in the torus).

   The results above can be found in Daniel Straulino's thesis [76], or more rigorously in Etheridge, Freeman & Straulino [37] for $d = 1$, and Etheridge, Freeman, Penington & Straulino [36] for $d = 2$. There are technical challenges in dimensions $d = 1$ and $d = 2$ because $ns_n \to \infty$: the branching rate is infinite in the limiting ASG, but almost all branches are instantly annulled by coalescence. This is why we have put inverted commas around 'branching Brownian motion' in $d = 2$. The ASG is close to a branching Brownian motion, but in reality reflects very rapid branching and coalescence.

**Remark 6.4.4 (Large scale events and small neighbourhood size).**
*In Theorem 6.4.3 we scaled the impact parameter, corresponding to letting neighbourhood size tend to infinity, and we saw that the only trace of large scale events that survived in the limit was in the spatial motion of the ancestral lineages. In order to coalesce, lineages still had to 'meet'. As a result, in dimensions at least two, in which two independent symmetric stable processes of index $\alpha \in (1, 2)$ will not meet, the scaling limit of our SLFV is deterministic. If we do not scale the impact parameter, then we will also see the effect of the large scale events in the coalescence of ancestral lineages, which will now be able to coalesce when at arbitrarily large separations. In the neutral case, this is investigated in Berestycki, Etheridge & Véber [9] where it is shown that there are interesting stochastic limits in all dimensions.*

## 6.5   Hybrid zones

So far we have only considered the simplest possible form of selection, in which one allele confers a selective advantage to the carrier. There are many different forms of selection, often involving interactions between multiple genetic loci. Here we continue to consider just one locus, but now we consider the interaction between the two copies of a gene carried by a diploid individual.

   Our interest arises from the study of hybrid zones. A hybrid zone is a narrow geographic region where two genetically distinct populations are found close together and hybridise to produce offspring of mixed ancestry.

They are maintained by a balance between selection and dispersal. Hybrid zones are prevalent in nature - thousands have been catalogued. Some arise because of a change in the environment at the location where the two populations meet. But some are maintained because although the two populations can interbreed, the hybrids are less fit than either purebred population. In this second scenario the hybrid zone can move around over time.

Our aim in this section is to investigate the motion of the hybrid zone. This is probably of more mathematical than biological interest, but it will allow us to make contact with some other interesting models.

We begin with a mathematical model for selection against hybrids. Individuals now carry *two* copies of a gene that we shall suppose occurs in just two alleles: $a$ and $A$. We assume so-called Hardy-Weinberg proportions: if $\bar{w} =$ is the proportion of $a$-alleles across the population, then the proportions of individuals carrying the possible combinations of $a$ and $A$ alleles is

$$\begin{array}{c|c|c} aa & aA & AA \\ \hline \bar{w}^2 & 2\bar{w}(1-\bar{w}) & (1-\bar{w})^2 \end{array}.$$

In an infinite population of monoecious individuals (meaning the individual has both male and female parts as in many plants and invertebrates), at least provided male and female gametes are produced at equal frequency, then even if the parental population is not at Hardy-Weinberg equilibrium, the offspring will be. (This is derived in Hardy [50], a paper that mathematically debunked the view that dominant alleles would take over a population.)

Individuals of type $aa$ or $AA$ are called homozygotes, those of type $aA$ are heterozygotes. To model a hybrid zone that is maintained by selection against heterozygosity, we shall take relative fitnesses:

$$\begin{array}{c|c|c} aa & aA & AA \\ \hline 1 & 1-s & 1 \end{array}.$$

Suppose that before reproduction the proportion of type $a$ alleles in the population is $\bar{w}$. During reproduction, each heterozygote produces $(1-s)$ times as many germ cells (cells of the same genotype) as a homozygote. Germ cells split into an effectively infinite pool of gametes (containing just one copy of the gene). The proportion of type $a$ in this pool is

$$\begin{aligned} \bar{w}^* = \frac{\bar{w}^2 + \bar{w}(1-\bar{w})(1-s)}{1 - 2s\bar{w}(1-\bar{w})} \quad &= \quad (1-s)\bar{w} + s(3\bar{w}^2 - 2\bar{w}^3) + \mathcal{O}(s^2) \\ &= \quad \bar{w} + s\bar{w}(1-\bar{w})(2\bar{w}-1) + \mathcal{O}(s^2). \end{aligned}$$

In an infinite population, if $s = \frac{\alpha}{N}$ (where $N$ is large), measuring time in units of $N$ generations, as $N \to \infty$,

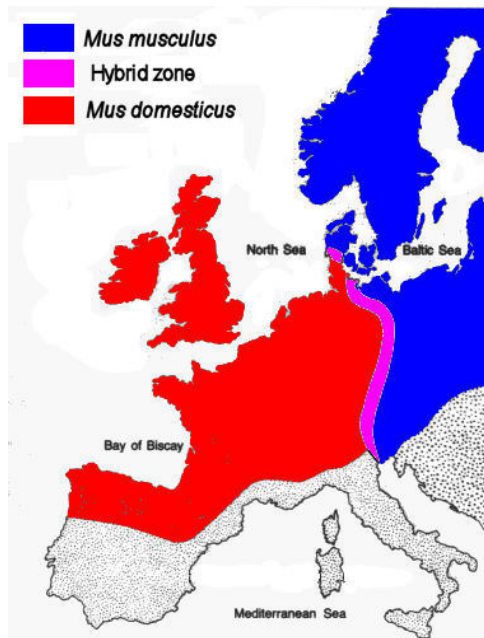$$\frac{d\bar{w}}{dt} = \alpha\bar{w}(1-\bar{w})(2\bar{w}-1).$$

Figure 6.3: The hybrid zone between two species of mice: *Mus musculus* (in the north) and *M. domesticus* (in the south). Picture © 2019 by Steven M Carr.

Adding dispersal yields

$$\frac{\partial w}{\partial t} = \frac{m}{2}\Delta w + \alpha w(1 - w)(2w - 1).$$

This is a special case of the *Allen-Cahn equation*. Figure 6.3 shows one of the most famous textbook examples of a hybrid zone. Hybrid zones are typically very narrow. Under the Allen-Cahn equation, the width of the zone is $\approx \sqrt{2m/\alpha}$. If we want to understand their motion over long time scales, then we may as well 'zoom out' so that they become one-dimensional. Applying a diffusive rescaling $t \mapsto \frac{t}{\varepsilon^2}$, $x \mapsto \frac{x}{\varepsilon}$, the Allen-Cahn equation becomes

$$\frac{\partial w}{\partial t} = \frac{m}{2}\Delta w + \frac{\alpha}{\varepsilon^2}w(1 - w)(2w - 1).$$

For convenience, set $m = 2$, $\alpha = 1$. It is well known in the p.d.e. literature that for sufficiently regular initial conditions, the solution to this equation converges, as $\varepsilon \to 0$, to the indicator function of a region whose boundary evolves according to *mean curvature flow*. We focus on two dimensions,

in which case mean curvature flow is just curvature flow which we now describe.

Let $S^1$ denote the unit circle in $\mathbb{R}^2$. Let $\boldsymbol{\Gamma}_t : S^1 \to \mathbb{R}^2$ be a family of smooth embeddings, indexed by $t \in [0, \mathcal{T})$. We write $\mathbf{n}_t(u)$ for the unit (inward) normal vector to $\boldsymbol{\Gamma}_t$ at $u$ and $\kappa = \kappa_t(u)$ for the curvature of $\boldsymbol{\Gamma}_t$ at $u$. Then $\boldsymbol{\Gamma}_t$ evolves according to curvature flow if

$$\frac{\partial \boldsymbol{\Gamma}_t(u)}{\partial t} = \kappa_t(u) \mathbf{n}_t(u).$$

It is defined up to a fixed time $T$. Curvature flow has a rich and beautiful mathematical theory. In particular, asymptotically as $t \uparrow T$, the curve $\boldsymbol{\Gamma}_t$ will tend towards a circle which then shrinks to a point, even if initially it does not enclose a convex region.

Let $d(x,t)$ be the signed distance from $x$ to $\boldsymbol{\Gamma}_t$. Choose $w_0$ such that $\boldsymbol{\Gamma}_0 = \{x \in \mathbb{R}^2 : w_0(x) = \frac{1}{2}\}$, $w_0 < \frac{1}{2}$ inside $\boldsymbol{\Gamma}$ and $> \frac{1}{2}$ outside. We have fixed parameters so that

$$\frac{\partial w}{\partial t} = \Delta w + \frac{1}{\varepsilon^2} w(1-w)(2w-1).$$

**Theorem 6.5.1 (Chen [17]).** *Fix $T^* \in (0,T)$. Let $k \in \mathbb{N}$. There exists $\varepsilon(k) > 0$, and $a(k), c(k) \in (0, \infty)$ such that for all $\varepsilon \in (0, \varepsilon(k))$ and $t$ satisfying $a\varepsilon^2 |\log \varepsilon| \le t \le T^*$,*

1. *for $x$ such that $d(x,t) \ge c\varepsilon |\log \varepsilon|$, we have $w(t,x) \ge 1 - \varepsilon^k$;*

2. *for $x$ such that $d(x,t) \le -c\varepsilon |\log \varepsilon|$, we have $w(t,x) \le \varepsilon^k$.*

If neighbourhood size were infinite, then this would suggest that hybrid zones maintained by selection against heterozygosity would evolve approximately according to curvature flow.

The Allen-Cahn equation originally arose in statistical physics, where it is more natural to consider solutions to the equation transformed so that solutions take values in $[-1, 1]$. Setting $v = 2w - 1$, we have

$$\frac{\partial v}{\partial t} = \Delta v + v - v^3.$$

Hairer, Ryser & Weber [49], consider this equation (still in two dimensions) with an additional noise term:

$$dv = (\Delta v + v - v^3)dt + \sigma d\mathcal{W},$$

where $\mathcal{W}$ is a mollified space-time white noise. They show that if the mollifier is removed, solutions converge weakly to zero; the structure that we see in the deterministic equation is destroyed. On the other hand, if

the intensity of $\mathcal{W}$ simultaneously converges to zero sufficiently quickly, then they recover the deterministic equation.

It is natural to ask whether hybrid zones will still evolve approximately according to curvature flow in the presence of random genetic drift? Additive mollified white noise is not a good model of genetic drift, so instead we add selection against heterozygosity to the SLFV.

Once again we start from the Wright-Fisher model. Recall that if the proportion of $a$-alleles in the parental population is $\bar{w}$, then under selection against heterozygosity, the proportion of $a$-alleles in the effectively infinite pool of gametes from which the generation is sampled is

$$
\begin{aligned}
\frac{\bar{w}^2 + \bar{w}(1-\bar{w})(1-s)}{1 - 2s\bar{w}(1-\bar{w})} &= (1-s)\bar{w} + s(3\bar{w}^2 - 2\bar{w}^3) + \mathcal{O}(s^2) \\
&= (1-s)\bar{w} + s(\bar{w}^3 + 3\bar{w}^2(1-\bar{w})) + \mathcal{O}(s^2).
\end{aligned}
$$

Mimicking what we did when we introduced genic selection, this suggests that for this form of selection too, we can think of there being two types of event in the SLFV:

1. with probability $(1-s)$ an event is neutral (as before);

2. with probability $s$, an event is selective and the probability that offspring are type $a$ is $\bar{w}^3 + 3\bar{w}^2(1-\bar{w})$, where $\bar{w}$ is the proportion of $a$-alleles in the affected region immediately before the event.

For the selective events, we can once again think of choosing 'potential parents' from the region. However, this time, we choose *three* potential parents, and the type of the offspring will be the type of the *majority* of the potential parents; that is offspring are of type $a$ if the three potential parents have types $\{a, a, a\}$ or $\{a, a, A\}$, otherwise offspring are type $A$.

This gives a means of investigating the interplay between genetic drift, spatial structure, and this form of selection. One might hope to be able to modify Chen's proof of convergence of the solution to the Allen-Cahn equation under the diffusive scaling, but that proof is too rigid to adapt to our setting. However, it turns out that there is a probabilistic approach to Chen's result, that is more flexible and can be adapted to include drift. The full details of this (and the corresponding stochastic result) can be found in Etheridge, Freeman & Penington [35]. Here we just present the main ideas.

The key to the probabilistic proof of the deterministic result is a representation of the solution to the Allen-Cahn equation in terms of a ternary branching Brownian motion, on which we superpose a 'majority voting' mechansim. This is similar to the Skorohod [75]/McKean [62] representation of the solution to the classical Fisher-KPP equation in terms of binary branching Brownian motion, except that here the voting
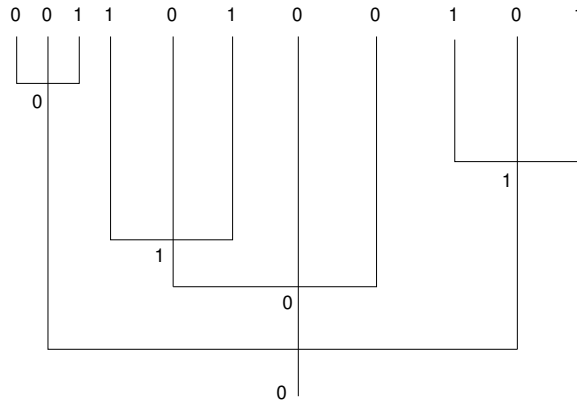
Figure 6.4: **Majority voting on a ternary tree.** Starting at the leaves, we move back through the tree. When we arrive at a branching event, the vote of the parent is taken to be the 'majority vote' of the three offspring. The resultant vote has been indicated at each branch point. In this example, the vote at the root is 0.

mechanism requires us to keep track of the entire history of the branching process. It is an adaptation of an idea of de Masi, Ferrari and Lebowitz [26]. In keeping with the original paper, we shall use $\{0, 1\}$ in place of $\{A, a\}$.

To see how it works, first we describe ternary branching Brownian motion started from a single individual at the point $x$. That individual has an exponential lifetime with parameter $1/\varepsilon^2$ (so mean $\varepsilon^2$), during which it follows a Brownian motion (with diffusion coefficient 2 in our scaling). At the end of its lifetime, it dies and is replaced, at the location where it died, by three offspring, which go on to behave, independently, in the same way as their parent.

We write $\boldsymbol{W}(t) =$ for the whole tree of paths up to time $t$. For a fixed function $p : \mathbb{R}^2 \to [0, 1]$, define a voting procedure on $\boldsymbol{W}(t)$ as follows:

1. each leaf, whose location we denote by $W_i(t)$, independently, votes 1 with probability $p(W_i(t))$ and otherwise votes 0;

2. at each branch point in $\boldsymbol{W}(t)$, the vote of the parent particle is the majority vote of the votes of its three children.

This defines an iterative voting procedure, which runs inwards from the leaves of $\boldsymbol{W}(t)$ to the root. Define $\mathbb{V}_p(\boldsymbol{W}(t))$ to be the vote associated to the root. This is illustrated in Figure 6.4.

**Proposition 6.5.2.** *If $\boldsymbol{W}(t)$ denotes historical branching Brownian*

*motion, with branching rate $\frac{1}{\varepsilon^2}$, and $p : \mathbb{R}^2 \to [0,1]$, then*

$$w(t,x) = \mathbb{P}_x^\varepsilon \left[ \mathbb{V}_p(\boldsymbol{W}(t)) = 1 \right]$$

*solves*

$$\frac{\partial w}{\partial t} = \Delta w + \frac{1}{\varepsilon^2} w(1-w)(2w-1), \quad w(0,x) = p(x).$$

**Idea of proof:**

The justification for this follows that of the corresponding justification of the Skorohod/McKean representation of the solution to the Fisher-KPP equation. Our aim is to evaluate

$$\lim_{\delta t \downarrow 0} \left( \frac{w(t+\delta t, x) - w(t,x)}{\delta t} \right).$$

Our approach is to partition over what happens to the ternary branching Brownian motion that determines $w(t+\delta t, x)$ in the first $\delta t$ of time: either the original individual dies, which happens with probability $\approx \delta t/\varepsilon^2$; or it does not.

Let $S$ be the first branch time of the branching Brownian motion. Let $V_1, V_2, V_3$, respectively, be the votes of the three offspring created at $S$, each corresponding to an independent branching Brownian motion running over the time interval $[S, t+\delta t]$. We have

$$\begin{aligned} w(t+\delta t, x) = & E_x \left[ \mathbb{P}_{W_S}^\varepsilon \left[ \text{at least 2 of the } (V_i) \text{ are } 1 \,|\, S \le \delta t \right] \right] \mathbb{P}\left[ S \le \delta t \right] \\ & + E_x \left[ \mathbb{P}_{W_{\delta t}}^\varepsilon \left[ \mathbb{V}(\boldsymbol{W}(t)) = 1 \right] \,|\, S > \delta t \right] (1 - \mathbb{P}\left[ S \le \delta t \right]), \end{aligned} \quad (6.8)$$

where $E_x$ denotes expectation with respect to a two-dimensional Brownian motion $W$ (with diffusion constant 2) started at the point $x$. The $V_1, V_2, V_3$ are conditionally independent (on the event $\{S \le \delta t\}$) random variables. Since we are only interested in events up to order $\delta t$, and $\mathbb{P}[S \le \delta t] = \mathcal{O}(\delta t)$, we only need evaluate each of their expectations to $\mathcal{O}(1)$, and so, in particular, we may suppose that no further branch has happened before time $\delta t$. Using the Markov property of the branching Brownian motion, and assuming some continuity (so that to first order we can approximate $V_i$ by assuming that the ancestor of the corresponding tree is still at $x$ at time $\delta t$), we find that we can use the approximation $\mathbb{E}[V_i] = w(t,x) + \mathcal{O}(\delta t)$.

On the event that there has been no branch by time $\delta t$, the ancestor of our branching Brownian motion has simply been following a Brownian motion and, by the Markov property, its vote is that determined by majority voting on a ternary branching Brownian motion started from $W_{\delta t}$, run over a time interval $0, t]$; that is, conditional on $W_{\delta t}$ it is $w(t, W_{\delta t})$.

Combining these observations,

$$\lim_{\delta t \downarrow 0} \frac{w(t+\delta t, x) - w(t,x)}{\delta t} = \frac{1}{\varepsilon^2}\Big( w(t,x)^3 + 3w(t,x)^2\big(1 - w(t,x)\big) - w(t,x)\Big)$$

$$+ \lim_{\delta t \downarrow 0}\left( \frac{E_x\left[w(t, W_{\delta t})\right] - w(t,x)}{\delta t}\right)$$

$$= \Delta w(t,x) + \frac{1}{\varepsilon^2}w(t,x)\big(1 - w(t,x)\big)\big(2w(t,x) - 1\big),$$

as required.                                                                          □

With this representation of solutions to the equation, Chen's result becomes:

1. for $x$ with $d(x,t) \geq c\varepsilon|\log\varepsilon|$, $\mathbb{P}^\varepsilon_x\left[\mathbb{V}_p(\boldsymbol{W}(t)) = 1\right] \geq 1 - \varepsilon^k$;

2. for $x$ with $d(x,t) \leq -c\varepsilon|\log\varepsilon|$, $\mathbb{P}^\varepsilon_x\left[\mathbb{V}_p(\boldsymbol{W}(t)) = 1\right] \leq \varepsilon^k$.

The proof proceeds in two steps. The first is to find a one-dimensional analogue in the case $p(x) = \mathbf{1}_{\{x \geq 0\}}$. This rests on symmetry, monotonicity for this $p$, and 'amplification of bias' through majority voting, by which we mean:

$$\text{if } p < \tfrac{1}{2}, \quad p^3 + 3p^2(1-p) < p;$$

$$\text{if } p > \tfrac{1}{2}, \quad p^3 + 3p^2(1-p) > p.$$

The second step is to couple a two-dimensional Brownian motion $W$ and a one-dimensional Brownian motion $B$ in such a way that $d(W_s, t-s)$ is well approximated by $B_s$ when $W_s$ is close to $\boldsymbol{\Gamma}_{t-s}$. This requires some regularity assumptions on the initial condition.

This approach parallels the approximation of the solution by a one-dimensional standing wave in Chen [17].

**Remark 6.5.3.** *The majority voting scheme works very well - the symmetry is a powerful tool. Not surprisingly, one can find voting schemes for other nonlinearities, although the representation may not be very useful. In his Oxford MMath dissertation in 2019, Zack O'Dowd showed that if $P(u)$ is any polynomial with $P(0) \geq 0$ and $P(1) \leq 0$ (or vice versa), then the solution to*

$$\frac{\partial u}{\partial t} = \Delta u + P(u)$$

*can be represented in terms of a historical n-ary branching Brownian motion and a voting scheme (possibly chosen at random from a collection of possible schemes at each branching event).*

In order to study our 'stochastic' hybrid zones, we consider the dual process of branching and coalescing lineages to the SLFV with selection against heterozygosity. As in the case of genic selection, we divide events into neutral events (which can cause jumps and coalescence) and selective

events. During a selective event, if any ancestral lineage is affected (that is marked), *three* 'potential' parents are born. We superpose majority voting on the resulting branching and coalescing structure in exactly the same way as we imposed it on the ternary branching Brownian motion, with the additional condition that all lineages that coalesce during a given event inherit the type determined by the (potential) parents of that event. If the impact of each event converges to zero sufficiently quickly, then with high probability we don't see any coalescence of ancestral lineages, and we are left with majority voting on a process that is very close to ternary branching Brownian motion.

## Relaxing the symmetry

If homozygotes *aa* and *AA* are not equally fit, then our Allen-Cahn equation must be modified. Assuming still that heterozygotes are less fit than either homozygote, we arrive at an equation of the form

$$
\begin{aligned}
\frac{\partial u}{\partial t} &= \frac{1}{2}\Delta u + su(1-u)(2au-1) \\
&= \frac{1}{2}\Delta u + sau(1-u)(2u-1) + s(a-1)u(1-u).
\end{aligned}
$$

The solution to this equation spreads across space with type *a* eventually taking over if $a > 1$. The timescale is faster than curvature flow. In fact the behaviour is quite sensitive to asymmetry. Mitch Gooding [48] shows that if

$$
\frac{\partial w}{\partial t} = \Delta w + \frac{1}{\varepsilon^2}w(1-w)(2w-(1+\nu\varepsilon)),
$$

then the limit is no longer curvature flow, but a mixture of curvature flow and flow at constant speed along the inward normal:

$$
\frac{\partial \mathbf{\Gamma}_t(u)}{\partial t} = \big(\nu + \kappa_t(u)\big)\mathbf{n}_t(u),
$$

(defined up to a fixed time $T$).

A smaller asymmetry takes us back to curvature flow; a larger asymmetry results in constant flow along the inward normal on a different timescale.

What we'd really like is to understand random fluctuations in the position of the hybrid zone caused by genetic drift and ultimately to find a description of the genealogies relating individuals sampled from the population.

# Chapter 7

# Further reading on the SLFV

We have only presented a fraction of the results available for the SLFV. A somewhat out of date survey is Barton, Etheridge & Véber [6]. Above all we should emphasize that the SLFV provides a flexible framework for modelling, into which we can readily incorporate many other mechanisms of biological relevance. For example, Etheridge & Véber [39] incorporates recombination and this structure is then exploited in Barton et al. [4] to develop inference methods for the model.

The relationship between the SLFV and the Dawson-Watanabe superprocess that we described in Section 6.4 mirrors results of Cox, Durrett & Perkins [22] for the voter model. More recently, Cox & Perkins [23] have shown how to recover a superprocess from the SLFV without scaling the impact. By considering more general distributions of event radii and impacts in a neutral SLFV, Chetwynd-Diggle & Etheridge [20] recover superprocesses with stable branching mechanisms as models for the evolution of rare alleles.

The forms of selection that we have considered all assume a constant environment. In Biswas, Etheridge & Klimek [12], we consider a population in which there are two genetic types, and which one is favoured depends on both space and time. In a regime in which selection fluctuates rapidly relative to the timescale of evolution, a scaling limit is obtained which is stochastic in all dimensions. Chetwynd-Diggle & Klimek [19] show that the dynamics of a rare allele in such a population, in appropriate scaling regimes, follows the superprocess in a random environment of Mytnik [66]. The proof of this result rests heavily on the lookdown construction of the SLFV in Etheridge & Kurtz [38], which in turn allows us to construct more general models in a similar vein (in which, for example, population density varies in space). That paper also relaxes

the condition (5.3) for existence of the SLFV.

Finally we mention that there are very different scaling regimes that one can consider. Forien & Penington [47] consider selection in favour of heterozygosity (the opposite of that considered in Section 6.5) in a regime in which selection is strong enough relative to genetic drift (governed by the impact of events) that the allele frequencies are maintained at an approximately constant intermediate frequency. They then prove a 'Central Limit Theorem' for the fluctuations about that limit and use it to investigate 'drift load', that is the loss of fitness in a population that is subject to selection as a result of genetic drift. They show that this behaves quite differently from the corresponding quantity in the absence of space. As well as being a technical tour de force, this paper is a powerful illustration of the importance of incorporating spatial structure into our models of populations.

# Bibliography

[1] Barton, N. H. (1993). The probability of fixation of a favoured allele in a subdivided population. *Genetical Research*, 62(02):149–157.

[2] Barton, N. H., Briggs, D. E. G., Eisen, J. A., Goldstein, D. B., and Patel, N. H. (2007). *Evolution*. Cold Spring Harbour Press.

[3] Barton, N. H., Depaulis, F., and Etheridge, A. M. (2002). Neutral evolution in spatially continuous populations. *Theor. Pop. Biol.*, 61:31–48.

[4] Barton, N. H., Etheridge, A. M., Kelleher, J., and Véber, A. (2013). Inference in two dimensions: allele frequencies versus lengths of shared sequence blocks. *Theor. Pop. Biol.*, 87:105–119.

[5] Barton, N. H., Etheridge, A. M., and Véber, A. (2010). A new model for evolution in a spatial continuum. *Electron. J. Probab.*, 15:162–216.

[6] Barton, N. H., Etheridge, A. M., and Véber, A. (2013). Modelling evolution in a spatial continuum. *J. Stat. Mech.*, PO1002.

[7] Barton, N. H., Kelleher, J., and Etheridge, A. M. (2010). A new model for extinction and recolonization in two dimensions: quantifying phylogeography. *Evolution*, 64(9):2701–2715.

[8] Berestycki, N., Etheridge, A. M., and Hutzenthaler, M. (2009). Survival, extinction and ergodicity in a spatially continuous population model. *Markov Processes and Related Fields*, 15:265–288.

[9] Berestycki, N., Etheridge, A. M., and Véber, A. (2013). Large-scale behaviour of the spatial $\Lambda$-Fleming-Viot process. *Ann. Inst. H. Poincaré*, 49(2):374–401.

[10] Bertoin, J. and Le Gall, J.-F. (2003). Stochastic flows associated to a coalescent process. *Prob. Theor. Rel. Fields*, 126:261–288.

[11] Birkner, M. and Depperschmidt, A. (2007). Survival and complete convergence for a spatial branching system with local regulation. *Ann. Appl. Probab.*, 17:1777–1807.

[12] Biswas, N., Etheridge, A., and Klimek, A. (2018). The spatial Lambda-Fleming-Viot process with fluctuating selection. *arXiv:1802.08188.*

[13] Bolker, B. M. and Pacala, S. W. (1997). Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor. Pop. Biol.*, 52(3):179–197.

[14] Buri, P. (1956). Gene frequency in small populations of mutant *Drosophila. Evolution*, 10:367–402.

[15] Chang, J. T. (1999). Recent common ancestors of all present day individuals. *Adv. Appl. Probab.*, 31:1002–1026.

[16] Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.*, 10:195–205.

[17] Chen, X. (1992). Generation and propagation of interfaces for reaction-diffusion equations. *J. Diff. Eq.*, 96:116–141.

[18] Chetwynd-Diggle, J. (2019). *Natural selection in spatially distributed populations.* Oxford DPhil Thesis.

[19] Chetwynd-Diggle, J. and Klimek, A. (2019). Rare mutations in the spatial Lambda-Fleming-Viot model in a fluctuating environment and SuperBrownian motion. *arXiv:1901.04374.*

[20] Chetwynd-Diggle, J. A. and Etheridge, A. M. (2018). SuperBrownian motion and the spatial Lambda-Fleming-Viot process. *Electron. J. Probab.*, 23:Paper no. 71, 36pp.

[21] Cox, J. T. and Durrett, R. (2002). The stepping stone model: new formulas expose old myths. *Ann. Appl. Probab.*, 12:1348–1377.

[22] Cox, J. T., Durrett, R., and Perkins, E. A. (2000). Rescaled voter models converge to super-Brownian motion. *Ann. Probab.*, 28:185–234.

[23] Cox, T. and Perkins, E. A. (2019). Rescaling the spatial Lambda-Fleming-Viot process and convergence to super-Brownian motion. *arXiv:1909.03277.*

[24] Dawson, D. and Perkins, E. (1991). Historical Processes. *Mem. A.M.S.*, 93(454).

[25] Dawson, D. A. (1993). Measure-valued Markov processes. In *École d'été de probabilités de Saint Flour*, volume 1541. Springer-Verlag.

[26] De Masi, A., Ferrari, P., and Lebowitz, J. (1986). Reaction-diffusion equations for interacting particle systems. *J Stat. Phys.*, 44(3-4):589–644.

[27] Donnelly, P. J. and Kurtz, T. G. (1996). A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Probab.*, 24:698–742.

[28] Donnelly, P. J. and Kurtz, T. G. (1999). Particle representations for measure-valued population models. *Ann. Probab.*, 27:166–205.

[29] Durrett, R. (1996). *Stochastic calculus. A practical introduction*. CRC Press.

[30] Durrett, R. (2008). *Probability models for DNA sequence evolution. 2nd Edition.* Springer-Verlag.

[31] Eller, E., Hawks, J., and Relethford, J. H. (2004). Local extinction and recolonization, species effective population size, and modern human origins. *Human Biology*, 76(5):689–709.

[32] Etheridge, A. M. (2000). *An introduction to superprocesses*, volume 20 of *University lecture notes.* Amer. Math. Soc.

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

[34] Etheridge, A. M. (2008). Drift, draft and structure: some mathematical models of evolution. *Banach Center Publ.*, 80:121–144.

[35] Etheridge, A. M., Freeman, N., and Penington, S. (2017). Branching Brownian motion, mean curvature flow and the motion of hybrid zones. *Electron. J. Probab.*, 22(103):40pp.

[36] Etheridge, A. M., Freeman, N., Penington, S., and Straulino, D. (2017). Branching Brownian motion and selection in the spatial $\Lambda$-Fleming-Viot process. *Ann. Appl. Probab.* Volume 27, number 5, 2605–2645.

[37] Etheridge, A. M., Freeman, N., and Straulino, D. (2017). The Brownian net and selection in the spatial $\Lambda$-Fleming-Viot process. *Electron. J. Probab.* Volume 22, paper no. 39, 36pp.

[38] Etheridge, A. M. and Kurtz, T. G. (2019). Genealogical constructions of population models. *Ann. Probab.*, 47(4):1827–1910.

[39] Etheridge, A. M. and Véber, A. (2012). The spatial Λ-Fleming-Viot process on a large torus: genealogies in the presence of recombination. *Ann. Appl. Probab.*, 22(6):2165–2209.

[40] Etheridge, A. M., Véber, A., and Yu, F. (2019+). Rescaling limits of the spatial Lambda-Fleming-Viot process with selection. *arXiv:1406.5884*.

[41] Ethier, S. N. and Kurtz, T. G. (1986). *Markov processes: characterization and convergence*. Wiley.

[42] Feller, W. (1951). Diffusion processes in genetics. *Proc. Second Berkeley Symp.*, pages 227–246.

[43] Felsenstein, J. (1975). A pain in the torus: some difficulties with the model of isolation by distance. *Amer. Nat.*, 109:359–368.

[44] Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Proc. Roy. Soc. Edinburgh*, 52:399–433.

[45] Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford University Press.

[46] Fisher, R. A. (1937). The wave of advance of advantageous genes. *Ann. Eugenics*, 7:355–369.

[47] Forien, R. and Penington, S. (2017). A central limit theorem for the spatial Λ-Fleming-Viot process with selection. *Electron. J. Probab.*, 20:68 pp.

[48] Gooding, M. (2018). *Long term behaviour of spatial population models with heterozygous or asymmetric homozygous selection*. Oxford DPhil Thesis.

[49] Hairer, M., Ryser, M. D., and Weber, H. (2012). On the triviality of the 2d stochastic Allen-Cahn equation. *Electron. J. Probab.*, 17:Paper no. 39, 14pp.

[50] Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science*, 28(706):49–50.

[51] Hudson, R. R. (1990). Gene genealogies and the coalescent process. In Futuyma, D. and Antonovics, J., editors, *Oxford Surveys in Evolutionary Biology*, pages 1–43. Oxford University Press.

[52] Jansen, S. and Kurt, N. (2014). On the notion(s) of duality for Markov processes. *Probability Surveys*, 11:59–120.

[53] Karlin, S. and Taylor, H. M. (1981). *A second course in stochastic processes*. Academic Press.

[54] Kimura, M. (1953). Stepping stone model of population. *Ann. Rep. Nat. Inst. Genetics Japan*, 3:62–63.

[55] Kingman, J. F. C. (1982). The coalescent. *Stoch. Proc. Appl.*, 13:235–248.

[56] Kolomogorov, A., Petrovsky, I., and Piscounov, N. (1937). Étude de l'equation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique. *Moscow Univ. Math. Bull.*, 1:1–25.

[57] Korolev, K. S., Xavier, J. B., Nelson, D. R., and Foster, K. R. (2011). A quantitative test of population genetics using spatiogenetic patterns in bacterial colonies. *American Naturalist*, 178(4):538–552.

[58] Krone, S. M. and Neuhauser, C. (1997). Ancestral processes with selection. *Theor. Pop. Biol.*, 51:210–237.

[59] Le Gall, J.-F. (1996). Superprocesses, Brownian snakes and partial differential equations. In *Lecture notes from the 11th Winter School on Stochastic Processes, Siegmundsburg. Prépublication no 337 du laboratoires de probabilités de l'université Paris VI.*

[60] Malécot, G. (1948). *Les Mathématiques de l'hérédité*. Masson et Cie, Paris.

[61] Maruyama, T. (1970). On the fixation probability of mutant genes in a subdivided population. *Genetical research*, 15(02):221–225.

[62] McKean, H. P. (1975). Application of Brownian motion to the equation of Kolmogorov-Petrovski-Piskunov. *Comm. Pure Appl. Math.*, 28:323–331.

[63] Melfi, A. and Viswanath, D. (2018). Single and simultaneous binary mergers in Wright-Fisher genealogies. *Theor. Pop. Biol.*, 121:60–71.

[64] Miller, L. (2015). *Evolution of highly fecund organisms*. Oxford DPhil Thesis.

[65] Moran, P. A. P. (1958). Random processes in genetics. *Proc. Camb. Phil. Soc.*, 54:60–71.

[66] Mytnik, L. (1996). Superprocesses in random environments. *Ann. Probab.*, 24:1953–1978.

[67] Nagylaki, T. (1978). A diffusion model for geographically structured populations. *J. Math. Biol.*, 6:375–382.

[68] Nagylaki, T. (1978). Random genetic drift in a cline. *Proc. Nat. Acad. Sci. USA*, 75:423–426.

[69] Neuhauser, C. and Krone, S. M. (1997). Genealogies of samples in models with selection. *Genetics*, 145:519–534.

[70] Perkins, E. A. (2002). Dawson-Watanabe Superprocesses and Measure-valued diffusions. In *École d'été de probabilités de Saint Flour XXIX-1999*. Springer-Verlag.

[71] Pitman, J. (1999). Coalescents with multiple collisions. *Ann. Probab.*, 27:1870–1902.

[72] Sagitov, S. (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.*, 26:1116–1125.

[73] Shiga, T. (1988). Stepping stone models in population genetics and population dynamics. In S. Alberverio et al, *Stochastic processes in physics and engineering*. Mathematics and its applications vol 42. Springer, Dordrecht.

[74] Simoni, L., Calafell, F., Pettener, D., Bertranpetit, J., and Barbujani, G. (2000). Geographic patterns of mtDNA diversity in Europe. *Am. J. Hum. Genet.*, 66:262–278.

[75] Skorohod, A. V. (1964). Branching diffusion processes. *Th. Prob. Appl.*, 9:492–497.

[76] Straulino, D. (2014). *Selection in a spatially structured population.* Oxford DPhil thesis.

[77] Sudbury, A. (1977). Clumping effects in models of isolation by distance. *J. Appl. Prob.*, 14(4):319–395.

[78] Taylor, J. E. (2009). The genealogical consequences of fecundity variance polymorphism. *Genetics*, 182(3):813–837.

[79] Véber, A. and Wakolbinger, A. (2015). The spatial Lambda-Fleming-Viot process: an event based construction and a lookdown representation. *Ann. Inst. H. Poincaré*, 51:570–598.

[80] Walsh, J. B. (1986). An introduction to stochastic partial differential equations. In *École d'été de probabilités de Saint Flour*, volume 1180. Springer-Verlag.

[81] Wright, S. (1943). Isolation by distance. *Genetics*, 28:114–138.

[82] Zähle, I., Cox, J. T., and Durrett, R. (2005). The stepping stone model ii: genealogies and the infinite sites model. *Ann. Appl. Probab.*, 15:671–699.

# Index

Alison Etheridge
Department of Statistics
University of Oxford
24-29 St Giles', Oxford OX1 3LB, United Kingdom
`http://www.stats.ox.ac.uk/~etheridg/`