The Moran model with random resampling rates

Driessen, Jannetje

Citation

Driessen, J. (2022). *The Moran model with random resampling rates*.

The Moran model with random resampling rates

Master thesis

J.E.P. DRIESSEN

Supervisor: Prof. Dr. F. den Hollander

Date of the exam: May 25, 2022

Mathematisch Instituut
Universiteit Leiden
The Netherlands
May 23, 2022

Abstract

In this master thesis, we study a stochastic model for genetic evolution. In particular, we add random resampling rates to the standard Moran model. Before the process starts, we let every individual in the population of size N choose at random a resampling rate from a finite set of size K of possible rates. We look at the K-vector of fractions of individuals with a given resampling rate and one of the two possible types. We show that, as $N \to \infty$, the scaled process converges in distribution in the Meyer-Zheng topology, which is a specific topology on the space of càdlàg paths. The limiting process lives on the K-dimensional simplex and its components are deterministic fractions of the total sum of components. The total sum performs a Wright-Fisher diffusion, with a diffusion constant that is the weighted average of the resampling rates. If the resampling rates scale with N and all converge to the same constant $r > 0$ as $N \to \infty$, then we obtain a similar result. In that case, the limiting process has diffusion constant r. If the resampling rates scale with N and converge to 0, then the random process converges in distribution in the Skorohod topology to a deterministic process.

Contents

1 Introduction

Humans, plants and animals have been around for a very long time. Over a period of 3.5 billion years, they all evolved from tiny multicellular organisms [1], [2]. The different species adapted to the environment through *natural selection* [3]. Small mutations in the genes of the offspring may result in traits that yield an advantage over others. When an organism has traits that make it have an advantage with respect to survival or reproduction, these traits are more likely to be passed on to future generations [4].

The study of genetic evolution was pioneered by Charles Darwin and Alfred Russel Wallace, who argued for their hypothesis of natural selection in careful detail [5]. However, how genetics worked precisely remained unclear until Friedrich Miescher first isolated DNA in 1869 [6]. In 1953, James Watson and Francis Crick, with the help of Rosalind Franklin and Maurice Wilkins, discovered the structure of DNA [7]. After this, genetic research went into high gear [8].

There is still a lot to learn. Further knowledge on evolution can be valuable for predicting the development of viruses [9], [10], creating advantageous modified crops [11], and making advances in the battle against genetic diseases [12]. That is why, in this master thesis, we study mathematical models that give insight into genetic evolution.

In Subsection 1.1, we present a few basic probabilistic models of genetic evolution and a few variations on these models. In Subsection 1.2, we explain what changes we make to these models and we state our main result.

1.1 Existing mathematical models

Reproduction and evolution are decidedly complex. There are so many factors at work that it would be impossible to take them all into account. However, this complexity can be encoded by using randomness. The probabilistic models that ensue from this idea are not perfect reflections of reality, but they do give us insight into the process of evolution.

In this section we discuss a few basic probabilistic models of genetic evolution. In particular, we look at the Wright-Fisher model and the Moran model. We refer to [13] for more background information.

The Wright-Fisher model and the Moran model consider a population of size N in which each of the individuals has one of two types of a trait: for instance, blue feathers or red feathers, green eyes or blue eyes, blonde hair or brown hair. As time progresses, the individuals in the population procreate and their children adopt the types of their parents. An interesting question is: How does the distribution of types evolve as the population size goes to infinity?

1.1.1 The Wright-Fisher model

The earliest versions of the Wright-Fisher model were created in the early 1940's [14]. To define this model, take a population of N individuals. Each individual has type $\heartsuit$ or type $\spadesuit$. After each unit of time, there is a completely new generation. Each of the new individuals chooses a parent uniformly at random from the previous population and adopts its type. Note that this is a model for haploid populations, which means that each individual has only one parent instead of two. When individuals have two parents, the population is called diploid. Under some constraints, the Wright-Fisher model can be used to represent a diploid population as well, when the population size is taken to be $2N$.

Let X_n be the number of individuals of type $\heartsuit$ at time n. For finite N,

$$\lim_{n\to\infty} \mathbb{P}(X_n = 0 \text{ or } X_n = N) = 1,$$

because 0 and N are traps. Namely, if there are no individuals with type ♡ in the population, no individual of the next generation can adopt type ♡. The reverse holds when no individual has type ♠.

Now that we have defined the random process $X = (X_n)_{n\in\mathbb{N}_0}$, we scale it in time and space to get

$$Y_t^N = \frac{1}{N} X_{\lceil Nt \rceil}, \quad t \geq 0.$$

Here, we speeded up time with a factor N, using $\lceil \cdot \rceil$ to get the upper integer part. We shrunk space by a factor N in order to be able to look at the fraction of individuals with type ♡ in the population, instead of counting the total number of individuals with type ♡. We speeded up time because significant changes in genetics only happen after a long time. The upper index N is there as a reminder that the process depends on N. It turns out that, in the limit as $N \to \infty$, the scaled process actually does something interesting. Namely, $Y^N = (Y_t^N)_{t\geq 0}$ converges to $Y = (Y_t)_{t\geq 0}$ in distribution on the space of càdlàg paths endowed with the Skorohod topology. The process $Y = (Y_t)_{t\geq 0}$ is the Wright-Fisher diffusion, defined by the stochastic differential equation

$$\mathrm{d}Y_t = \sqrt{Y_t(1-Y_t)}\,\mathrm{d}W_t.$$

Here, $W = (W_t)_{t\geq 0}$ is the standard Brownian motion. This statement is proven by showing that the infinitesimal generator L_N of $Y^N = (Y^N(t))_{t\geq 0}$ converges to the infinitesimal generator L of $Y = (Y_t)_{t\geq 0}$, which is given by

$$(Lf)(y) = \frac{1}{2} y(1-y) f''(y), \quad y \in [0,1],$$

for smooth test-functions f.

1.1.2 The Moran model

The Moran model is a continuous-time version of the Wright-Fisher model. It was first proposed by Moran in 1958 [15]. The process $X^N = (X_t^N)_{t\geq 0}$ is the same, except that each individual resamples and chooses a new parent at rate 1. Thus, the resampling is done sequentially instead of in parallel.

We again scale in time and space to get

$$Y_t^N = \frac{1}{N} X_{Nt}^N, \quad t \geq 0.$$

In the limit as $N \to \infty$, the scaled Moran model also converges to the Wright-Fisher diffusion. The only difference is that it runs twice as fast: the infinitesimal generator L_N of $Y^N = (Y_t^N)_{t\geq 0}$ converges to $2L$ instead of L.

1.1.3 Variations of the standard Moran model

The standard Moran model can be adapted to give a more accurate reflection of reality by adding *mutation*, *selection*, or *migration*.

Mutation is added to the model by letting each individual suffer a type mutation directly after resampling. With probability $u \in (0,1)$ it changes into type ♡ and with probability $v \in (0,1)$ it changes into type ♠. With probability $(1-u)(1-v)$, it does not mutate. Note that the mutations happen irrespective of the current type of the individual. When a mutation ♠ → ♠ or ♡ → ♡ occurs, we treat it as if the individual did not mutate.

In reality, when a gene mutates, the probability that it mutates into a type that was seen before is small. That is why Kimura introduced the infinite alleles model in 1968 [16]. In this model, any time an individual mutates, a new type is created.

With selection, one type resamples faster than the other. This means that the slower type has an advantage, as individuals with that type stay the same for a longer period of time. To implement this, we decree that each individual of type ♠ resamples at rate 1, while each individual of type ♡ resamples at rate $s \neq 1$.

Migration is added to the Moran model by creating different colonies. When an individual in a colony resamples, it chooses with probability $\nu \in (0,1]$ a different colony, and with probability $1-\nu$ its own colony. After that it chooses its parent uniformly at random from the chosen colony.

1.2 The Moran model with random resampling rates

In the models described above, all the parameters are deterministic. In particular, each individual has the same deterministic resampling rate. We wondered what would happen when the resampling rate of each individual is chosen at random. More specifically, what if individuals choose their resampling rate randomly, according to some law on a finite set of possible rates? In what follows, we study the *quenched* process in which the resampling rates are chosen first and do not change while the process runs.

It turns out, as we will see in Theorem 2.1, that the process converges to a process whose components are deterministic fractions of the total sum of the components. This sum performs a Wright-Fisher diffusion, with a diffusion constant that is determined by the relative significance of the different resampling rates. More formally, let K be the number of possible resampling rates, let r_k be the possible resampling rates for $k \in [K] = \{1, \dots, K\}$, and let μ_k be the probability that an individual has resampling rate r_k. Let $Y_k^N(t)$ denote the fraction of individuals at time Nt that have type ♡ and resampling rate r_k. Then the process $Y^N = (Y^N(t))_{t\geq 0}$ with $Y^N(t) = (Y_k^N(t))_{k\in[K]}$ converges in distribution on the space of càdlàg paths endowed with the Meyer-Zheng topology. The Meyer-Zheng topology is a specific topology that is weaker than the Skorohod topology, as we explain in Subsection 2.4.

The limiting process $Y = (Y(t))_{t\geq 0}$ lives on $[0,1]^K$ and is given by

$$Y_k(t) = \mu_k S(t)$$

for all $k \in [K]$, with $S(t) = \sum_{k\in[K]} Y_k(t)$. Note that each $Y_k(t)$ is a deterministic function of $S(t)$. It turns out that the process $S = (S(t))_{t\geq 0}$ is the scaled Wright-Fisher diffusion, defined by

$$\mathrm{d}S(t) = \sqrt{DS(t)(1-S(t))}\,\mathrm{d}W(t)$$

with $D = \sum_{k\in[K]} \mu_k r_k$. The state space of S is $[0,1]$.

The above statement is intuitive, because the only difference between the total fraction of individuals with type ♡ in this model and the standard Moran model is that some individuals may resample faster than others. The term $D = \sum_{k\in[K]} \mu_k r_k$ precisely accounts for this, as it matches the resampling rates with their relative weight in the population.

After proving the main theorem, we analyse what happens when the random resampling rates scale with the population size. We elaborate on this in Section 3. If the resampling rates scale with N such that $\lim_{N\to\infty} r_k = r$ for all $k \in [K]$ and for a certain $r > 0$, then we again obtain the Wright-Fisher diffusion, but with diffusion constant r instead of D. If the resampling rates scale such that $\lim_{N\to\infty} r_k = 0$ for all $k \in [K]$, then the process converges in distribution in the Skorohod topology to a deterministic process.

We do not consider the process in which there are countably infinite or uncountably infinite possible resampling rates. These and other possible future questions are discussed in Subsection 4.

2 Finitely many resampling rates

We introduce the random resampling rates in Subsection 2.1 and the random process in Subsection 2.2. After that, we scale the process and state the main result in Subsection 2.3. The proof of this result requires several steps of reasoning, which we split up into a main argument and several technical lemmas. The main argument is explained in Subsection 2.4. The lemmas are proven in Subsection 2.5 and 2.6.

2.1 The random resampling rates

Consider the Moran model, but with random resampling rates instead of fixed rates. We study the case in which there are finitely many resampling rates. Thus, each individual in the population has a random resampling rate that is drawn from a set $\Psi \subset \mathbb{R}_{>0}$ such that $|\Psi| = K \in \mathbb{N}$. We write $\Psi = \{r_1, \dots, r_K\}$.

In order to assign the resampling rates, we need a probability space and a random variable. So, define the sample space $\Omega = \Psi$ and add the σ-algebra $\mathcal{F} = \mathcal{P}(\Omega)$ to get the measurable space $(\Omega, \mathcal{F})$. Define the probability measure $\mathbb{P}$ by putting $\mathbb{P}(r_k) = \mu_k$ for $k \in [K]$, such that $\sum_{k\in[K]} \mu_k = 1$. We require that $\mu_k > 0$ for all $k \in [K]$. This gives us the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Now define the random variable $Z : \Omega \to \mathbb{R}$, that maps an element onto itself. Finally, define the law ν such that $\nu(B) = \mathbb{P}(Z \in B)$ for every $B \in \mathcal{B}(\mathbb{R})$.

To make the model as simple as possible, we assign the resampling rates to the individuals *independently* and before the process starts. To do this, define $R = (\tilde{r}_i)_{i\in\mathbb{N}}$ in which the $\tilde{r}_i$ are independent and identically distributed with law ν. Individual i of the population resamples at rate $\tilde{r}_i$.

2.2 The random process

In this subsection, we formally establish our random process and introduce the necessary notation.

Let $N \in \mathbb{N}$ be the size of the population. Individuals are labelled $\{1, \dots, N\}$ and carry either type $\heartsuit$ or type $\spadesuit$. The type of individual i at time t is denoted by $\eta_i(t)$. Define $X^{(N,R)}_{\rho,\eta}(t)$ as the number of individuals of type η with resampling rate ρ at time t. To monitor the process, we define

$$X(t) = (X^{(N,R)}_{r_k,\heartsuit}(t))_{k\in[K]}.$$

Note that we sort all the individuals with type $\heartsuit$ by their resampling rates. We drop the upper index (N, R) for now, but we remember that the population has size N and the assigned resampling rates are given by R. Further note that we may just look at type $\heartsuit$ instead of both $\heartsuit$ and $\spadesuit$, because

$$X_{r,\heartsuit}(t) + X_{r,\spadesuit}(t) = X_{r,\heartsuit}(0) + X_{r,\spadesuit}(0)$$

for all $r \in R$ and all $t \geq 0$, because the total number of individuals with a specific resampling rate does not change. Therefore, it is useful to define $\hat{n}_k \in \mathbb{N}_0$ for $k \in [K]$, where

$$\hat{n}_k = X_{r_k,\heartsuit}(0) + X_{r_k,\spadesuit}(0).$$

In other words, $\hat{n}_k$ is the total number of individuals with rate r_k. It is clear that $\sum_{k\in[K]} \hat{n}_k = N$.

The state space of $X = (X(t))_{t\geq 0}$ is given by

$$\hat{Q} = \{x = (x_k)_{k\in[K]} \,|\, x_k \in \mathbb{N}_0 \text{ and } 0 \leq x_k \leq \hat{n}_k \;\; \forall\, k \in [K]\}.$$

To analyse X, we define the operator $\hat{T}^{k,\Delta} : \hat{Q} \to \hat{Q}$ as follows: for $\Delta \in \{-1, +1\}$, $k \in [K]$, e_k the unit vector, and $x \in \hat{Q}$, put

$$\hat{T}^{k,\Delta} x = x + \Delta e_k.$$

This operator encodes the event that an individual with resampling rate r_k changes its type. Namely, when $\Delta = -1$, an individual with rate r_k and type ♡ chooses a parent with type ♠ and adopts its type. When $\Delta = +1$, an individual with rate r_k and type ♠ chooses a parent with type ♡ and adopts its type. The rate at which the process does this is given by $\hat{R}^{k,\Delta}$, with $\hat{R}^{k,\Delta} : \hat{Q} \to \mathbb{R}$ defined as

$$\hat{R}^{k,-1}(x) = r_k x_k \left(\sum_{l \in [K]} (\hat{n}_l - x_l) \right) \frac{1}{N},$$

$$\hat{R}^{k,+1}(x) = r_k (\hat{n}_k - x_k) \left(\sum_{l \in [K]} x_l \right) \frac{1}{N}.$$

It is easy to see that the change $\hat{T}^{k,-1}(x)$ happens at rate $\hat{R}^{k,-1}(x)$, because there are x_k individuals with type ♡ that can choose a new parent from the population at rate r_k. In order for such an individual to change to ♠, it needs to choose a parent with type ♠. There are $\sum_{l \in [K]} (\hat{n}_l - x_l)$ individuals of type ♠ in total, and the individual chooses one of them uniformly, which explains the factor $\frac{1}{N}$. The explanation for $\hat{R}^{k,+1}(x)$ goes analogously.

Note that the transition rates $\hat{R}^{k,\Delta}$ and the transition operators $\hat{T}^{k,\Delta}$ only depend on the current state of the process. Therefore, we conclude that $X = (X(t))_{t \geq 0}$ is a Markov chain.

Using what we found so far, we can now write down the infinitesimal generator $\hat{L}_N$ of X. Let $f : \hat{Q} \to \mathbb{R}$ be a bounded and continuous function, and let $x \in \hat{Q}$. Then

$$(\hat{L}_N f)(x) = \sum_{k \in [K]} \sum_{\Delta \in \{-1,+1\}} \hat{R}^{k,\Delta}(x) [f(\hat{T}^{k,\Delta} x) - f(x)].$$

This generator gives us both an intuitive idea and further knowledge of the random process X. The next step is to scale X, which we do in the next subsection. Later, the generator of the scaled process will be used in Subsection 2.5 in the proof of one of the lemmas.

2.3 The scaled process and its limit

Just like in Subsection 1.1, we scale the process in time and in space. We scale in space by a factor $\frac{1}{N}$, because if we take the limit $N \to \infty$, then the total number of individuals with type ♡ diverges, while the fraction of individuals with type ♡ might converge. We scale in time with a factor N to speed up the process. This is again intuitive, because genetic change is only significant after many generations.

Therefore, define

$$Y_k^N(t) = \frac{X_k^N(Nt)}{N}$$

as the fraction of individuals of type ♡ with resampling rate r_k at time t. Also define the variable $n_k^N = \frac{\hat{n}_k^N}{N}$, which is the fraction of individuals with resampling rate r_k. Note that $n_k^N \in [0, 1]$ and $\sum_{k \in [K]} n_k^N = 1$.

The random process $Y^N = (Y^N(t))_{t \geq 0}$ with $Y^N(t) = (Y_k^N(t))_{k \in [K]}$ lives on the state space

$$Q = \left\{ y = (y_k)_{k \in [K]} \;\middle|\; 0 \leq y_k \leq n_k^N \quad \forall\ k \in [K] \right\} \subset [0, 1]^K.$$

In fact, y_k is a multiple of $\frac{1}{N}$ for all $k \in [K]$.

Similarly as before, we get the transition operators $T^{k,\Delta} : Q \to Q$ with

$$T^{k,\Delta} y = y + \Delta \frac{e_k}{N},$$

and the transition rates

$$R^{k,-1}(y) = r_k y_k \left(\sum_{l \in [K]} (n_l^N - y_l) \right),$$
$$R^{k,+1}(y) = r_k (n_k^N - y_k) \left(\sum_{l \in [K]} y_l \right).$$

Finally, we substitute the operators and the transition rates into the infinitesimal generator for X, to find the infinitesimal generator for Y^N:

$$(L_N f)(y) = \sum_{k \in [K]} \sum_{\Delta \in \{-1,+1\}} R^{k,\Delta}(y) \left[f(T^{k,\Delta} y) - f(y) \right] N^2.$$

The term N^2 comes from the time scaling by N and the fact that $R^{k,\Delta} = \frac{1}{N} \hat{R}^{k,\Delta}$.

Now that we have established the necessary definitions and notations, we can state our main theorem. To do this, we note that $n_k^N = \frac{\hat{n}_k}{N}$ converges to μ_k in probability as $N \to \infty$ for all $k \in [K]$, by the law of large numbers. Furthermore, define the constant $D = \sum_{k \in [K]} \mu_k r_k$.

Theorem 2.1. *Given the assigned resampling rates $R = (\tilde{r}_i)_{i \in \mathbb{N}}$, where the $\tilde{r}_i$ are iid with law ν, the process $Y^N = (Y^N(t))_{t \geq 0}$ converges in distribution on the space of càdlàg trajectories endowed with the Meyer-Zheng topology to the process $Y = (Y(t))_{t \geq 0}$ that lives on $[0,1]^K$ and is defined by $Y(t) = (Y_k(t))_{k \in [K]}$ with $Y_k(t) = \mu_k S(t)$, $k \in [K]$. The process $S = (S(t))_{t \geq 0}$ lives on $[0,1]$ and solves the stochastic differential equation*

$$dS(t) = \sqrt{D S(t)(1 - S(t))}\, dW(t),$$

where $W = (W(t))_{t \geq 0}$ is the standard Brownian motion.

As mentioned before, this theorem is quite intuitive. The standard Moran model converges to the Wright-Fisher diffusion, and the only difference here is that individuals may resample faster or slower than others. Thus, we need to account for the relative influence of the resampling rates. This is precisely what is encoded in the factor $D = \sum_{k \in [K]} \mu_k r_k$.

We remark that, for all $k \in [K]$, $Y_k(t)$ is a deterministic function of $S(t)$. In fact, for $i, j \in [K]$, the ratio between $Y_i(t)$ and $Y_j(t)$ is $\frac{\mu_i}{\mu_j}$ for all $t \geq 0$. The total fraction $S(t)$ of type $\heartsuit$ performs a Wright-Fisher diffusion, and $Y_k(t)$ is precisely the corresponding μ_k fraction of $S(t)$.

2.4 Proof of the main theorem

Convergence of Markov processes is usually proven by showing that the infinitesimal generators converge, as explained in [17]. However, we will see that the infinitesimal generator of the process Y^N contains a factor N, and therefore diverges as $N \to \infty$. Instead, we will prove that the process converges in the Meyer-Zheng topology.

A natural topology on the space of càdlàg trajectories is induced by the supremum norm. However, this is inapplicable in the presence of jumps, which we do encounter in our process. Another topology on this space is the Skorohod topology, that gives flexibility in time and space. If two paths are mostly the same, but there is a small amount of time between their respective changes, then the Skorohod topology states that the paths are close. This topology is coarser than the topology induced by the supremum norm. Even weaker is the Meyer-Zheng topology, which was introduced by Meyer and Zheng in 1984 [18]. In addition to giving flexibility in time and space, this topology also gives flexibility in terms of local times. If for most of the time the process converges in space, the process converges in the Meyer-Zheng topology. For more background, we refer to [18], [19], [20].

2.4.1 The Meyer-Zheng topology

Let $v_N : [0,\infty) \to Q$ denote the path of Y^N, and let $v : [0,\infty) \to Q$ denote the path of Y. Define the corresponding pseudopaths ψ_{v_N} and ψ_v like in [18]. Let $f \in C_b(\Xi)$, where Ξ is the space of all pseudopaths. Then, according to [19],

$$\psi_{v_N}(f) = \int_{[0,\infty)} f(t, v_N(t)) \, \mathrm{e}^{-t} \, \mathrm{d}t.$$

Suppose that $f(t,x) = \mathbb{1}_{t\in[a,b]}\bar{f}(x)$, with $\bar{f} : Q \to \mathbb{R}$ bounded and continuous and $0 \leq a < b < \infty$. The set of functions defined in this way generates $C_b(\Xi)$. Results can thus be extended to all bounded and continuous functions by using the standard machinery of measure theory.

We can write

$$\psi_{v_N}(f) = \int_a^b \bar{f}(v_N(t)) \, \mathrm{e}^{-t} \, \mathrm{d}t.$$

To prove convergence in the Meyer-Zheng topology, we need to show that

$$\lim_{N\to\infty} \mathbb{E}[\psi_{v_N}(f)] = \mathbb{E}[\psi_v(f)].$$

By using [20] and doing some rearranging, we get that the latter is equivalent to

$$\lim_{N\to\infty} \mathbb{E}\left[\int_a^b [\bar{f}(Y^N(t)) - \bar{f}(Y(t))] \, \mathrm{e}^{-t} \, \mathrm{d}t\right] = 0.$$

Suppose that $\bar{f}$ is Lipschitz continuous with Lipschitz constant $C \in \mathbb{R}$. By the Stone-Weierstrass theorem, the space of polynomial functions defined on the set of closed intervals in $[0,1]^N$ is dense in the space of continuous functions defined on $[0,1]^N$. Furthermore, the space of Lipschitz functions contains the space of polynomial functions. This means that any result for Lipschitz functions can be generalised to continuous $\bar{f}$. We find that

$$\begin{aligned}
&\mathbb{E}\left[\int_a^b [\bar{f}(Y^N(t)) - \bar{f}(Y(t))] \, \mathrm{e}^{-t} \, \mathrm{d}t\right] \\
&\quad \leq C \, \mathbb{E}\left[\int_a^b |Y^N(t) - Y(t)| \, \mathrm{e}^{-t} \, \mathrm{d}t\right] = C \int_a^b \mathbb{E}[|Y^N(t) - Y(t)|] \, \mathrm{e}^{-t} \, \mathrm{d}t.
\end{aligned}$$

By Fubini's theorem we may indeed interchange the integrals, because both integrals are obviously finite: $t \mapsto \mathrm{e}^{-t}$ is integrable and Y^N and Y are contained in the set $[0,1]^N$. Note that we write $|\cdot|$ instead of $||\cdot||_1$, for ease of notation.

In summary, if we can prove that, for all $t \geq 0$,

$$\lim_{N\to\infty} \mathbb{E}[|Y^N(t) - Y(t)|] = 0,$$

then we have proven that $Y^N = (Y^N(t))_{t\geq 0}$ converges to $Y = (Y(t))_{t\geq 0}$ in the Meyer-Zheng topology.

2.4.2 Proof

Define the sum $S^N(t) = \sum_{k\in[K]} Y_k^N(t)$. Furthermore, define the random processes $\hat{Y}^N(t) = (\hat{Y}_k^N(t))_{k\in[K]}$ and $\tilde{Y}^N(t) = (\tilde{Y}_k^N(t))_{k\in[K]}$ by putting, for $k \in [K]$,

$$\hat{Y}_k^N(t) = n_k^N S^N(t),$$
$$\tilde{Y}_k^N(t) = \mu_k S^N(t).$$

Note that both $\hat{Y}^N$ and $\tilde{Y}^N$ take values in $[0,1]^K$. By the triangle inequality,

$$\mathbb{E}[|Y^N(t) - Y(t)|] \le \mathbb{E}[|Y^N(t) - \hat{Y}^N(t)|] + \mathbb{E}[|\hat{Y}^N(t) - \tilde{Y}^N(t)|] + \mathbb{E}[|\tilde{Y}^N(t) - Y(t)|].$$

Therefore, the proof is complete when we show that all three terms converge to zero.

The middle term is straightforward, because

$$|\hat{Y}^N(t) - \tilde{Y}^N(t)| \le \left| \sum_{k\in[K]} \left|n_k^N - \mu_k\right| S^N(t) \right|,$$

which converges to zero, because $|S^N(t)| \le 1$ and $n_k^N \to \mu_k$ almost surely as $N \to \infty$ by the law of large numbers. For the other two terms, we need the following lemmas.

Lemma 2.2. $\mathbb{E}[|Y^N(t) - \hat{Y}^N(t)|]$ *converges to zero as* $N \to \infty$ *uniformly in* t.

Lemma 2.3. *The sequence of processes* $S^N = (S^N(t))_{t\ge 0}$ *converges in distribution to the process* $S = (S(t))_{t\ge0}$ *in the Meyer-Zheng topology as* $N \to \infty$.

Note that, because $\sum_{k\in[K]} \mu_k = 1$,

$$\mathbb{E}[|\tilde{Y}^N(t) - Y(t)|] = \mathbb{E}\left[\left|\left(\mu_k S^N(t) - \mu_k S(t)\right)_{k\in[K]}\right|\right] = \mathbb{E}\left[\left|S^N(t) - S(t)\right|\right].$$

Thus, given Lemma 2.2 and 2.3, we conclude that $\mathbb{E}[|Y^N(t) - Y(t)|] \to 0$, so we indeed find that $Y^N = (Y^N(t))_{t\ge0}$ converges to $Y = (Y(t))_{t\ge0}$ in the Meyer-Zheng topology. □

We prove Lemma 2.2 and Lemma 2.3 in Subsection 2.5 and Subsection 2.6, respectively.

2.5 Proof of the first key lemma

The idea of the proof is to find a Lyapunov function that rapidly decays to zero. Background information on Lyapunov functions can be found in [21]. In order to be able to work with the infinitesimal generator, we first write down a Taylor expansion.

2.5.1 Lyapunov function

Let f be a continuous and bounded function. The Taylor expansion of $f(T^{k,\Delta}y) - f(y)$ around y is given by

$$\begin{aligned} f(T^{k,\Delta}y) - f(y) &= \frac{\Delta}{N}\frac{\partial}{\partial y_k} f(y) + \frac{1}{2N^2}\left(\frac{\partial}{\partial y_k}\right)^2 f(y) + \frac{\Delta}{6N^3}\left(\frac{\partial}{\partial y_k}\right)^3 f(y) + \dots \\ &= \sum_{j=1}^{\infty} \left(\frac{\Delta}{N}\right)^j \frac{1}{j!}\left(\frac{\partial}{\partial y_k}\right)^j f(y). \end{aligned}$$

Substituting this into $(L_N f)(y)$, we get

$$\begin{aligned}(L_N f)(y) &= \sum_{k\in[K]} \sum_{\Delta\in\{-1,+1\}} R^{k,\Delta}(y) \left[\sum_{j=1}^{\infty} \left(\frac{\Delta}{N}\frac{\partial}{\partial y_k}\right)^j \frac{f(y)}{j!}\right] N^2 \\ &= \sum_{j=1}^{\infty} \sum_{k\in[K]} N^{2-j} r_k \left(\frac{\partial}{\partial y_k}\right)^j \frac{f(y)}{j!} \left((-1)^j y_k \left(1 - \sum_{l\in[K]} y_l\right) + (n_k^N - y_k)\left(\sum_{l\in[K]} y_l\right)\right).\end{aligned}$$

Note that for $j \geq 3$ the term in the sum is $\mathcal{O}(\frac{1}{N})$, which converges to zero. For $j = 2$, we get

$$\sum_{k\in[K]} r_k \left(\left(\frac{\partial}{\partial y_k}\right)^2 \frac{f(y)}{2}\right) \left(y_k - 2y_k \left(\sum_{l\in[K]} y_l\right) + n_k^N \left(\sum_{l\in[K]} y_l\right)\right),$$

which has no factor N or $\frac{1}{N}$. However, for $j = 1$, we have

$$N \sum_{k\in[K]} r_k \left(\frac{\partial}{\partial y_k} f(y)\right) \left(-y_k + n_k^N \sum_{l\in[K]} y_l\right).$$

As $N \to \infty$, this term diverges and so we need to find out how to properly control the limiting process. Assume that f is a non-decreasing function. We see that if

$$y_k > n_k^N \sum_{l\in[K]} y_l,$$

then the term is very negative because of the factor N, while if

$$y_k < n_k^N \sum_{l\in[K]} y_l,$$

then the term is very positive. This shows that the system wants to move close the manifold where $y_k = n_k^N \sum_{l\in[K]} y_l$ for all $k \in [K]$.

To prove this more formally, we define a *Lyapunov function* and apply the infinitesimal generator to it. For ease of writing, we introduce the notation $\Delta_k = y_k - n_k^N \sum_l y_l$. Using this notation, we define the function $h : [0,1]^{[K]} \to \mathbb{R}$ given by

$$h(y) = \sum_{k\in[K]} |\Delta_k|.$$

Let $m \in [K]$ and note that

$$\frac{\partial \Delta_k}{\partial y_m} = \delta_{mk} - n_k^N,$$

where δ is the Kronecker delta.

The first-order partial derivative of h is given by

$$\frac{\partial h(y)}{\partial y_m} = \sum_{k\in[K]} \operatorname{sign}(\Delta_k)(\delta_{mk} - n_k^N) = \operatorname{sign}(\Delta_m) - \sum_{k\in[K]} \operatorname{sign}(\Delta_k) n_k^N.$$

We recognise that the derivative of the absolute value does not exist in 0, but a continuous-time process spends no time in a singleton. Thus, the above formula suffices as the partial derivative of $h(y)$.

The higher-order partial derivatives of h are all equal to 0. Indeed, the second-order derivative of h is equal to

$$\begin{aligned}\left(\frac{\partial}{\partial y_m}\right)^2 h(y) &= \frac{\partial}{\partial y_m}\left(\operatorname{sign}(\Delta_m) - \sum_{k\in[K]} \operatorname{sign}(\Delta_k) n_k^N\right)\\ &= (1-n_k^N)\frac{\partial}{\partial y_m}\operatorname{sign}(\Delta_m)\\ &= 0.\end{aligned}$$

Again, we recognise that this is not entirely true. In fact, the derivative of the sign is twice the Dirac distribution. However, a continuous-time process does not spend time in a singleton, so in practice, we can use 0 as the second- and higher-order derivative of $h(y)$.

Recall that

$$(L_N h)(y) = N \sum_{m\in[K]} r_m \left[\frac{\partial h(y)}{\partial y_m}\right]\left(-y_m + n_m^N\left(\sum_{l\in[K]} y_l\right)\right),$$

and recall that

$$\sum_{k\in[K]} n_k^N = 1.$$

Furthermore,

$$\sum_{k\in[K]} \Delta_k = \sum_{k\in[K]} y_k - \left(\sum_{k\in[K]} n_k^N\right)\left(\sum_{l\in[K]} y_l\right) = 0,$$

and, for each $k \in [K]$,

$$\operatorname{sign}(\Delta_k) \in \{-1, 0, 1\}.$$

Therefore, if there exists a $k \in [K]$ for which $\Delta_k \neq 0$, then there exists at least one $k \in [K]$ such that $\operatorname{sign}(\Delta_k) = -1$ and at least one $k \in [K]$ such that $\operatorname{sign}(\Delta_k) = +1$. So,

$$\sum_{k\in[K]} n_k^N \operatorname{sign}(\Delta_k) \in (-1, 1).$$

Consequently,

$$\begin{aligned}(L_N h)(y) &= -N \sum_{m\in[K]} r_m \Delta_m \left[\frac{\partial h(y)}{\partial y_m}\right]\\ &= -N \sum_{m\in[K]} r_m \Delta_m \left[\operatorname{sign}(\Delta_m) - \sum_{k\in[K]} n_k^N \operatorname{sign}(\Delta_k)\right]\\ &= -N \sum_{m\in[K]} r_m |\Delta_m| \underbrace{\left[1 - \operatorname{sign}(\Delta_m)\sum_{k\in[K]} n_k^N \operatorname{sign}(\Delta_k)\right]}_{>0}.\end{aligned}$$

Thus we see that $L_N h$ is very negative, unless $|\Delta_m|$ is of order $\mathcal{O}(\frac{1}{N})$ for all $m \in [K]$. In that case the term is still negative, but of order 1. Intuitively, this means that the process makes sure that $h(Y^N(t))$ quickly decays to zero, which means that the $Y_k^N(t) - n_k^N S^N(t)$ quickly decay to zero as well. Thus, we indeed see that the system wants to live close to the manifold where $y_k = n_k^N \sum_l y_l$ for all $k \in [K]$.

2.5.2 The process spends minimal time away from the manifold

For both the proof of Lemma 2.2 and the proof of Lemma 2.3, we need to know how quickly the process gets close to the manifold where $y_k = n_k^N \sum_l y_l$ for all $k \in [K]$.

By the law of large numbers, we can write $n_k^N = \mu_k + o(1)$. By using this and taking N large enough, we can find an $\epsilon = \epsilon((\mu_k)_{k\in[K]}, R) > 0$ such that, uniformly in $y = (y_k)_{k\in[K]}$,

$$\begin{aligned}
&r_m \left[1 - \operatorname{sign}(\Delta_m) \sum_{k\in[K]} n_k^N \operatorname{sign}(\Delta_k)\right] \\
&= r_m \left[1 - \operatorname{sign}(\Delta_m) \sum_{k\in[K]} (\mu_k + \mathrm{o}(1)) \operatorname{sign}(\Delta_k)\right] \\
&= r_m \left[1 - \operatorname{sign}(\Delta_m) \sum_{k\in[K]} \mu_k \operatorname{sign}(\Delta_k)\right] - r_m \left[\operatorname{sign}(\Delta_m) \sum_{k\in[K]} o(1) \operatorname{sign}(\Delta_k)\right] \\
&\geq \epsilon.
\end{aligned}$$

The inequalities are true, because r_m is fixed,

$$1 - \operatorname{sign}(\Delta_m) \sum_{k\in[K]} \mu_k \operatorname{sign}(\Delta_k) \in (0, 2)$$

for all $m \in [K]$ uniformly in $y = (y_k)_{k\in[K]}$, and we have chosen N large enough in order for the term with $o(1)$ to be negligible. Note that ϵ does not depend on N. Thus, we find

$$\begin{aligned}
(L_N h)(y) &= -N \sum_{m\in[K]} |\Delta_m| r_m \left[1 - \operatorname{sign}(\Delta_m) \sum_{k\in[K]} n_k^N \operatorname{sign}(\Delta_k)\right] \\
&\leq -N \sum_{m\in[K]} |\Delta_m| \epsilon \\
&= -N\epsilon h(y).
\end{aligned}$$

To see how the process behaves in expectation, define

$$g(t) = \mathbb{E}[h(Y^N(t))].$$

Then

$$g'(t) = \mathbb{E}[(L_N h)(Y^N(t))] \leq \mathbb{E}[-N\epsilon h(Y^N(t))] = -N\epsilon g(t).$$

Furthermore,

$$\int_{\cdot}^{t} \mathrm{d}s \, \frac{g'(s)}{g(s)} = \log g(t) + C_1,$$

for some constant $C_1 \in \mathbb{R}$. Hence,

$$\log g(t) + C_1 \leq -\int_{\cdot}^{t} \mathrm{d}s \; N\epsilon = -N\epsilon t + C_2,$$

for some constant $C_2 \in \mathbb{R}$. From this we get that

$$g(t) \leq \mathrm{e}^{C_2 - C_1} \mathrm{e}^{-N\epsilon t}.$$

By picking $t = 0$, we see that we need to take the constants C_1 and C_2 such that

$$\mathrm{e}^{C_2 - C_1} = g(0) = \mathbb{E}[h(Y_0^N)].$$

Thus,

$$g(t) \;\leq\; g(0)\;\mathrm{e}^{-N\epsilon t} \;\leq\; ||h||_\infty \mathrm{e}^{-N\epsilon t}.$$

Note that this means that, as $N \to \infty$, the probability to be away from the manifold at time $t > 0$ converges to zero.

We want to know how much time the process spends far away from the manifold. To that end, we define the neighbourhood

$$U_N = \left\{ y \in Q \,|\, h(y) < \frac{1}{N^2} \right\}.$$

Pick $\alpha > 0$. Use Markov's inequality to estimate

$$\begin{aligned}
\mathbb{E}\left[\int_\alpha^\infty \mathrm{d}t\, \mathbb{1}_{\{Y^N(t) \notin U_N\}}\right] &= \mathbb{E}\left[\int_\alpha^\infty \mathrm{d}t\, \mathbb{1}_{\{h(Y^N(t)) \geq \frac{1}{N^2}\}}\right] \\
&= \int_\alpha^\infty \mathrm{d}t\, \mathbb{P}\left(h(Y^N(t)) \geq \frac{1}{N^2}\right) \\
&\leq \int_\alpha^\infty \mathrm{d}t\, N^2\; \mathbb{E}(h(Y^N(t))) \\
&= N^2 \int_\alpha^\infty \mathrm{d}t\, g(t) \\
&\leq ||h||_\infty N^2 \int_\alpha^\infty \mathrm{d}t\;\mathrm{e}^{-N\epsilon t} \\
&= ||h||_\infty \frac{N}{\epsilon}\;\mathrm{e}^{-N\epsilon\alpha}.
\end{aligned}$$

The right-hand side converges to zero as $N \to \infty$. In other words, when the population gets larger, the process spends less time outside the neighbourhood U_N after time α. However, α is arbitrary, so we can take the limit $\alpha \downarrow 0$ after taking the limit $N \to \infty$, to see that the process spends no time outside the vanishing neighbourhood U_N.

Therefore, we see that Δ_k is $\mathcal{O}(\frac{1}{N^2})$ for all $k \in [K]$, so we have indeed proven that, for all $t > 0$,

$$\lim_{N\to\infty} \mathbb{E}[|Y^N(t) - \hat{Y}^N(t)|] = \lim_{N\to\infty} \mathbb{E}\left[\left|\left(Y_k^N(t) - n_k^N \sum_{l\in[K]} Y_l^N(t)\right)_{k\in[K]}\right|\right] = 0.$$

□

2.6 Proof of the second key lemma

The process S^N is not Markov, so it has no infinitesimal generator. Nonetheless, the set of functions that act on the sum of the components is a subset of the set of functions that act on the full process. Thereby, we can investigate how the infinitesimal generator of Y^N acts on S^N.

Recall that $y_k = n_k^N s + \Delta_k$ for all $k \in [K]$, by the definition of Δ_k, with $s = \sum_{l\in[K]} y_l$. Note that

$$\frac{\partial}{\partial y_k} f(s) = \frac{\partial}{\partial s} f(s)$$

for any $k \in [K]$ and any bounded and continuous f. Hence,

$$\begin{aligned}(L_N f)(s) &= \sum_{j=1}^{\infty} \sum_{k\in[K]} N^{2-j} r_k \left(\left(\frac{d}{ds}\right)^j \frac{f(s)}{j!} \right) \\ &\quad \times \left[(-1)^j (n_k^N s + \Delta_k) \left(1 - s - \left(\sum_{k\in[K]} \Delta_k \right) \right) + \left(s + \left(\sum_{k\in[K]} \Delta_k \right) \right) (n_k^N (1-s) - \Delta_k) \right] \\ &= \sum_{j=1}^{\infty} \sum_{k\in[K]} N^{2-j} r_k \left(\left(\frac{d}{ds}\right)^j \frac{f(s)}{j!} \right) \left[(-1)^j (n_k^N s + \Delta_k)(1-s) + s(n_k^N (1-s) - \Delta_k) \right].\end{aligned}$$

The last equality is true because

$$\sum_{k\in[K]} \Delta_k = \sum_{k\in[K]} y_k - \underbrace{\left(\sum_{k\in[K]} n_k^N \right)}_{=1} \left(\sum_{m\in[K]} y_m \right) = 0.$$

From the expression for $(L_N f)(s)$, we find that all third- and higher-order terms are $\mathcal{O}(\frac{1}{N})$ and thus converge to zero as $N \to \infty$.

The first-order term is

$$N f'(s) \left(\sum_{k\in[K]} r_k \Delta_k \right).$$

We know from Subsection 2.5 that Y^N spends most of its time in the neighbourhood U_N, on which Δ_k is $\mathcal{O}(\frac{1}{N^2})$ for all $k \in [K]$. Specifically, when N becomes larger, Y^N spends exponentially less time outside U_N. But on U_N, the first-order term disappears as $N \to \infty$.

The second-order term is

$$f''(s) \left[\left(\sum_{k\in[K]} r_k n_k^N \right) s(1-s) + \left(\sum_{k\in[K]} r_k \Delta_k \right) \left(\frac{1}{2} - s \right) \right].$$

So we can conclude that

$$\lim_{N\to\infty} (L_N f)(s) = D s(1-s) f''(s),$$

because the sum $\sum_{k\in[K]} r_k \Delta_k$ is of order $\mathcal{O}(\frac{1}{N^2})$ on U_N. Thus, $S^N = (S^N(t))_{t\geq 0}$ converges in distribution to $S = (S(t))_{t\geq 0}$ in the Meyer-Zheng topology. S performs the Wright-Fisher diffusion with diffusion constant $D = \sum_{k\in[K]} \mu_k r_k$. This concludes our proof of Lemma 2.3, and with it the proof of the main theorem. □

3 Scaled resampling rates

So far, we have only studied resampling rates that are constant in N. In this section, we look at random resampling rates that scale with N.

3.1 Strong resampling rates

Let $r \in \mathbb{R}$ and take $c = (c_k)_{k\in[K]}$ with $c_k \in \mathbb{R}_{>0}$. Define $r_k = r + \frac{c_k}{N}$ for all $k \in [K]$. Analogously to before, we can construct the random process $Y^N = (Y^N(t))_{t\geq 0}$ and we obtain a similar result.

Corollary 3.1. *When* $\lim_{N\to\infty} r_k = r$ *for all* $k \in [K]$*, the process* $Y^N = (Y^N(t))_{t\geq 0}$ *converges in distrubution to* $Y = (Y(t))_{t\geq 0}$ *in the space of càdlàg paths endowed with the Meyer-Zheng topology, where* $Y(t) = (Y_k(t))_{k\in[K]}$ *and* $Y_k(t) = \mu_k S(t)$*.* $S = (S(t))_{t\geq 0}$ *performs the Wright-Fisher diffusion with diffusion constant* r*.*

To prove this corollary, we use the same arguments as for Theorem 2.1. In the limit as $N \to \infty$, r_k converges to r for all $k \in [K]$, and thus

$$\sum_{k\in[K]} \mu_k r_k \longrightarrow r \sum_{k\in[K]} \mu_k = r.$$

□

3.2 Weak resampling rates

If instead we take $r_k = \frac{c_k}{N}$ and construct the process analogously to before, then we get the infinitesimal generator

$$(L_N f)(y) = \sum_{j=1}^{\infty} \sum_{k\in[K]} N^{1-j} c_k \left(\frac{\partial}{\partial y_k}\right)^j \frac{f(y)}{j!} \left((-1)^j y_k \left(1 - \sum_{l\in[K]} y_l\right) + (n_k^N - y_k)\left(\sum_{l\in[K]} y_l\right)\right).$$

To see why, recall the infinitesimal generator from Subsection 2.5 and substitute $r_k = \frac{c_k}{N}$.

However, the second- and higher order terms of $L_N f$ are $\mathcal{O}(\frac{1}{N})$, so

$$\lim_{N\to\infty} (L_N f)(y) = (Lf)(y) = \sum_{k\in[K]} c_k \left(\frac{\partial}{\partial y_k} f(y)\right) \left[-y_k + \mu_k \left(\sum_{k\in[K]} y_k\right)\right].$$

Note that the limiting process is deterministic, because L is a first-order partial differential equation operator.

Corollary 3.2. *When* $\lim_{N\to\infty} r_k = 0$ *for all* $k \in [K]$*, the process* $Y^N = (Y^N(t))_{t\geq 0}$ *converges to a deterministic process in distribution on the space of càdlàg paths endowed with the Skorohod topology.*

□

4 Conclusion

In this master thesis, we studied the Moran model with random resampling rates. The set-up was quenched, because the individuals first select their resampling rates at random, and then the process runs. We have proven that the scaled version of the process converges in distribution in the Meyer-Zheng topology as the population size N tends to infinity. The Meyer-Zheng topology is a topology on on the space of càdlàg trajectories which is coarser than the Skorohod topology.

The limiting process $Y = (Y(t))_{t\geq 0}$ lives on $[0,1]^K$, where K is the number of possible resampling rates. For $k \in [K]$, $\mu_k > 0$ is the probability that an individual has resampling rate r_k. The limiting process Y is defined by $Y(t) = (Y_k(t))_{k\in[K]}$ and $Y_k(t) = \mu_k S(t)$. Here, $S = (S(t))_{t\geq 0}$ performs the Wright-Fisher diffusion with diffusion constant $\sum_{k\in[K]} \mu_k r_k$. In other words, the scaled Moran model with random resampling rates converges to a deterministic functional of a Wright-Fisher diffusion with a diffusion constant that is the weighted average of the resampling rates.

When the resampling rates scale with N such that $\lim_{N\to\infty} r_k = r$, the same proof applies and we find that the limiting process is a deterministic functional of a Wright-Fisher diffusion with diffusion constant r. On the other hand, when the resampling rates scale such that $\lim_{N\to\infty} r_k = 0$, the scaled process converges in distribution in the Skorohod topology to a deterministic process.

A natural question to ask is: What happens when there are countably infinite or even uncountably infinite possible resampling rates? We would expect a similar result, but the proof needs work. In the former case, we could try to first take the limit $K \to \infty$ and after that take the limit $N \to \infty$. In the latter case, all processes need to be redefined, because they no longer live on the simplex $[0,1]^K$.

Other possible variations of the Moran model are when ♡ and ♠ have different resampling rate distributions and each individual chooses a new rate every time it resamples, or when each individual adopts the resampling rate of its parent. Furthermore, we could also add mutation to the model with random resampling rates, or add random mutation rates as well. To add selection, an individual could choose a pair of resampling rates and use one when it has type ♡ and use the other when it has type ♠. In conlusion, there are a lot of opportunities for further research on the Moran model with random resampling rates, or variations of it.

5 References

[1] Schopf, J.W. (2006). *Fossil evidence of Archaean life.* Philosophical Transactions of the Royal Society B. 361 (1470): 869–885. doi:10.1098/rstb.2006.1834

[2] Futuyma, D.J. (2005). *Evolution.* Sunderland, MA, Sinauer Associates. doi:10.1086/501261

[3] Orr, H.A. (2005). *The genetic theory of adaptation: a brief history.* Nature Reviews Genetics 6 (2): 119–127. doi:10.1038/nrg1523

[4] Hurst, L.D. (2009). *Fundamental concepts in genetics: genetics and the understanding of selection.* Nature Reviews Genetics 10 (2): 83–93. doi:10.1038/nrg2506

[5] Darwin, C.R., Wallace, A.R. (1858). *On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection.* Journal of the Proceedings of the Linnean Society of London. Zoology. 3 (9): 45–62. doi:10.1111/j.1096-3642.1858.tb02500.x

[6] Dahm, R. (2008). *Discovering DNA: Friedrich Miescher and the early years of nucleic acid research.* Human Genetics 122 (6): 565–81. doi:10.1007/s00439-007-0433-0

[7] Watson, J.D., Crick, F. H. C. (1953). *Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid.* Nature 171 (4356): 737–738. doi:10.1038/171737a0

[8] Sarkar, S. (1998). *Genetics and Reductionism.* Cambridge University Press p.140. doi:10.1017/cbo9781139173216

[9] Worobey, M., Holmes, E.C. (1999). *Evolutionary aspects of recombination in RNA viruses.* The Journal of General Virology 80 (10): 2535–43. doi:10.1099/0022-1317-80-10-2535

[10] Moss, R.B., Davey, R.T., Steigbigel, R.T., Fang, F. (2010). *Targeting pandemic influenza: a primer on influenza antivirals and drug resistance.* The Journal of Antimicrobial Chemotherapy 65 (6): 1086–93. doi:10.1093/jac/dkq100

[11] Klümper, W., Qaim, M. (2014). *A meta-analysis of the impacts of genetically modified crops.* PLOS ONE 9 (11): e111629. doi:10.1371/journal.pone.0111629

[12] Verma, I. M. (2013). *Gene Therapy That Works.* Science 341 (6148): 853–855. doi:10.1126/science.1242551

[13] Avena, L., Da Costa, C., Den Hollander, F. (2019). *Stochastic models of genetic evolution.* Lecture notes, Leiden University, The Netherlands, 2019.

[14] Wright, S. (1942). *Statistical Genetics and Evolution.* Bulletin of the American Mathematical Society (48): 223-246. doi:10.1090/S0002-9904-1942-07641-5

[15] Moran, P.A.P. (1958). *Random processes in genetics.* Mathematical Proceedings of the Cambridge Philosophical Society 54 (1): 60–71. doi:10.1017/S0305004100033193

[16] Kimura, M. (1968). *Evolutionary rate at the molecular level.* Nature. Nature Publishing Group 217 (5129): 624–6. doi:10.1038/217624a0

[17] Ethier, S., and Kurtz, T. (1986). *Markov Processes: Characterization and Convergence.* New York: Wiley. Chapter 4. doi:10.1002/9780470316658

[18] Meyer, P.A., and Zheng, W.A. (1984). *Tightness criteria for laws of semimartingales.* Annales de l'I.H.P. Probabilités et Statistiques 20(4): 353-372. zbl:0551.60046

[19] Kurtz, T.G. (1991). *Random time changes and convergence in distribution under the Meyer-Zheng conditions.* Annales de l'I.H.P. Probabilités et Statistiques 19(3): 1010-1034. doi:10.1214/aop/1176990333

[20] Greven, A., Den Hollander, F., Oomen, M. (2021). *Spatial populations with seed-bank; renormalisation on the hierarchical group.* To appear in the Memoirs of the American Mathematical Society.

[21] Meiss, J. D. (2008). *Differential dynamical systems.* Philadephia, PA: Society for Industrial and Applied Mathematics. Chapter 4.6. doi:10.1137/1.9780898718232

A Notations

N	number of individuals in the population
$W = (W_t)_{t\geq 0}$	Brownian motion
Ψ	sample set of resampling rates
K	cardinality of Ψ, number of distinct resampling rates
$[K]$	$\{1, \ldots, K\}$
ν	law that assigns the resampling rates
μ_k	probability that an individual has resampling rate r_k
$R = (\tilde{r}_i)_{i\in\mathbb{N}}$	assigned resampling rates
$\eta_i(t)$	type of individual i at time t
$X^{(N,R)}_{\rho,\eta}(t)$	number of individuals with type η and resampling rate ρ at time t
$X(t)$	$(X_{r_k,\heartsuit}(t))_{k\in[K]}$
$\hat{n}_k$	total number of individuals with resampling rate r_k
n^N_k	$= \frac{\hat{n}_k}{N}$, fraction of individuals with resampling rate r_k
$\hat{Q}$	state space of X
Δ	$\in \{-1, +1\}$
$\hat{T}^{k,\Delta}$	operator that encodes the possible transitions of X given its current state
$\hat{R}^{k,\Delta}$	transition rates of X given its current state
$\hat{L}_N$	infinitesimal generator of X
$Y^N(t)$	$\frac{1}{N}X(Nt)$
Q	state space of Y
$T^{k,\Delta}$	operator that encodes the possible transitions of Y given its current state
$R^{k,\Delta}$	transition rates of Y given its current state
L_N	infinitesimal generator of Y^N
D	$\sum_{k\in[K]} \mu_k r_k$
$S(t)$	Wright-Fisher diffusion with diffusion constant D
$Y(t)$	limiting process of Y^N for $N \to \infty$, with $Y_k(t) = \mu_k S(t)$
v_N	path of Y^N
v	path of Y
ψ_{v_N}	pseudopath of Y^N
ψ_v	pseudopath of Y
Ξ	space of pseudopaths
$S^N(t)$	$\sum_{k\in[K]} Y^N_k(t)$
$\hat{Y}^N_k(t)$	$n^N_k S^N(t)$
$\tilde{Y}^N_k(t)$	$\mu_k S^N(t)$
Δ_k	$y_k - n_k \sum_l y_l$
h	Lyapunov function $h : Q \to \mathbb{R}$ with $h(y) = \sum_{k\in[K]} \lvert\Delta_k\rvert$
δ_{mk}	Kronecker delta
$g(t)$	$\mathbb{E}[h(Y^N(t))]$
ϵ	lower bound for $r_m \left[1 - \operatorname{sign}(\Delta_m) \sum_{k\in[K]} n^N_k \operatorname{sign}(\Delta_k)\right]$
α	$\in \mathbb{R}_{>0}$ arbitrary