Population Dynamics in Random Environment, Random Walks on Symmetric Group, and Phylogeny Reconstruction

Arash Jamshidpey

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics [1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1]The Ph.D. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

This thesis concerns applications of some probabilistic tools to phylogeny reconstruction and population genetics. Modelling the evolution of species by continuous-time random walks on the signed permutation groups, we study the asymptotic medians of a set of random permutations sampled from simple random walks at time $0.25cn$, for $c > 0$. Running k independent random walks all starting at identity, we prove that the medians approximate the ancestor (identity permutation) up to time $0.25n$, while there exists a constant $c > 1$ after which the medians loose credibility as an estimator. We study the median of a set of random permutations on the symmetric group endowed with different metrics. In particular, for a special metric of dissimilarity, called breakpoint, where the space is not geodesic, we find a large group of medians of random permutations using the concept of partial geodesics (or geodesic patches).

Also, we study the Fleming-Viot process in random environment (FVRE) via martingale and duality methods. We develop the duality method to the case of time-dependent and quenched martingale problems. Using a family of dual processes we prove the convergence of the Moran processes in random environments to FVRE in Skorokhod topology. We also study the long-time behaviour of FVRE and prove the existence of equilibrium for the joint annealed-environment process and prove an ergodic theorem for the latter.

Acknowledgements

I have been very fortunate to have great supervisors to help me out. My deepest gratitude is to my PhD advisors, Prof. Don Dawson and Prof. David Sankoff, for their invaluable guidance and continuous encouragement. Their patience and support helped me overcome many crisis situations and finish this dissertation. I learned a lot of things from them. I could not have imagined having better advisors for my PhD. It was a great pleasure for me to be their student. So thanks Don. Thanks David. I would like to express my sincere appreciation to Prof. Vadim Kaimanovich, a member of my advisory committee, who taught me many things during my PhD. His door was always open to me, and I had this chance to discuss many parts of my research with him. I am grateful to him for his support, patience, and encouragement. I am also very much indebted to Prof. Miklós Csörgő, Prof. Xiaowen Zhou, and Prof. Mayer Alvo, for their excellent comments, suggestions, and help. I am proud, honored and extremely grateful to have been able to work with such talented researchers, but most of all, who are also kind human beings.

Huge thanks goes to my friends who helped me to stay strong during my studies, also to my fellow labmates in Sankoff Lab, with whom I had the pleasure to work with during the last years: Chunfang Zheng, Mona Meghdari, Krister Swenson, Ghada Badr, Zhe Yu, Yue Zhang, Alex Brandts, Eric Chen, Daniella Munoz, Caroline Larlee, and Baoyang Wang. I would like to express my especial thanks and appreciation to Poly Hannah da Silva for her suggestions and help in preparation of this dissertation.

At last but not least, to my parents, thanks for your unconditional love and support. During my life you were always there for me, day and night, and never gave up on me for even a moment. This dissertation is dedicated to you because this would not have been possible without your love and patience. I would like to acknowledge the rest of my family, my brothers and sister-in-law for their love and support. Thank you for all the moments you spent for me when I needed you.

Dedication

To my parents

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview of the thesis

This thesis reports the applications of some mathematical tools, from probability and discrete geometry in particular, to comparative genomics, phylogeny reconstruction, and population genetics. Today, the application of mathematics in both modelling biology structures and in analysing existing models and data is growing very fast. Among the biological fields in which mathematics has been applied successfully, one can mention molecular biology, genomics, cancer genomics, phylogeny reconstruction, population genetics, etc.[26, 19, 53, 25, 8, 62]. This thesis concerns some problems for both backward and forward stochastic models which arise, respectively, in genome-level phylogenetic reconstruction and long time behaviour in population dynamics with environmental changes. In the latter the role of natural selection is emphasized in a time-varying environment and the adaptation of the population with respect to this parameter is studied. Studying species from points of view of both its genetic variations and its large scale genomic evolution is a major research concern of contemporary biology. The role of mathematics in inferring phylogenetic trees has been highlighted in recent years, and many mathematical tools are developed to better understand the evolutionary history of species[48, 17]. On the other hand, the foundation of modern population genetics is based on mathematical tools, and one can see the trace of probability theory and the theory of differential equations in most parts of this field. In general, firstly, mathematics develops tools by which one can analyse the observed data and, secondly, serves as a measure of correctness and exactness for developing theories in biology. In the second case, combining mathematical and algorithmic tools, may not only finds the limitations of an informal theory, but can also lead the theory to become more precise.

From the mathematical point of view, in the case of genome rearrangement, we study certain relations between simple random walks on the Cayley graph of the symmetric group on n points and its asymptotic local geometry. In fact, because of some biological applications, we are interested in special subgroups of the symmetric groups. In this thesis, we concern the problem of the medians of random permutations sampled from simple random walks on the Cayley graph of symmetric groups with respect to certain generating sets. In fact the Cayley graphs of any finitely generated group are quasi-isometric. Therefore, it is not far from reality to have some asymptotically invariant geometric properties for finite symmetric groups with respect to different word metrics. In the level of asymptotic geometry, when one studies the behaviour of symmetric groups S_n as $n \to \infty$, with a convenient rescaling of the metrics, some behaviours of Cayley graphs with respect to different generating sets are similar. This can be measured by means of simple random walks. In fact simple random walks have valuable information about the local geometry of the group. In the biological context, the random walk could correspond to chromosome changes, for example reversal, under reproduction. For the purpose of this thesis, our objective is to measure the credibility of the geometric median of k random permutations sampled from k independent simple random walks at time cn (where $c > 0$, and all random walks start at a same permutation, say identity) as a means of estimating the starting state. That is we look for cases in which the median is not far from the starting state of the random walks. This is applied when the evolution of genomes is modelled by continuous-time random walks on the permutation groups when a permutation represents a genome. As discussed, most of the results in Chapters 2 and 3 remain true for different metrics in symmetric groups including the transposition metric. However for its biological importance, we formulate results for the reversal metric and the reversal random walk on a subgroup of the symmetric group on $2n$ elements, namely signed symmetric groups .

Another measure of dissimilarity of genomes is the breakpoint distance. Endowed with this pseudometric distance, the symmetric group is not a discrete geodesic space. Understanding the importance of discrete geodesics to determine the median points in the problems studied in Chapters 2 and 3, we try to generalize the notion of geodesics to partial geodesics (p-geodesics or geodesic patches). In fact the concept of non-trivial p-geodesics can serve as an interesting concept in many different problems corresponding to non-geodesic spaces. Chapter 4 studies the asymptotic breakpoint medians of finitely many random permutations.

Probabilistic models play a crucial role in population genetics. In particular, for a long time, different popular models in interacting particle systems have been used to model several population dynamics. In fact two important mechanisms of evolution in population dynamics, namely mutation and natural selection, are better to be understood as random time-varying parameters. The dynamics of a population is effected by environmental changes. In fact, the genetic variations exist in the genomes of species and these variations, in turn, are in interaction with the environments. Natural selection, as the most important mechanism of evolution, favors the fitter type in an organism. The fitness of different types determines the role of "natural selection" in a population and depends on an important environmental parameter.

It is a function of environmental changes and other evolutionary mechanisms, i.e. mutation and genetic drift. Subsequently, an important question is the effect of environmental changes on the structure of the population. "Adaptive processes have taken centre in molecular evolutionary biology. Time dependent fitness functions has opposing effects on adaptation. Rapid fluctuations enhance the stochasticity of the evolutionary process and impede long-term adaptation.[45]" In other words, living in rapidly varying environments, a population is not able to adapt to the environment. Because of simplicity, the existing probabilistic models in population genetics mainly concern problems in which the natural selection is not time-dependent. This decreases the validity of models and does not allow the study of the interactions between the environment and the population. In other words, they cannot explain the real effect of the environment on adaptation of a population system. In fact, it is both more realistic and also challenging to have a random environment varying in time.

Chapter 5 studies long time behaviours of some countable probabilistic population dynamics in random environment. For this purpose we make use of the martingale problem and the duality method and we develop a generalization of existing methods in the literature to the case of time-dependent Markov processes. Specially, the duality method has been studied for time-inhomogeneous Markov processes. In the case of their existence, dual processes are powerful tools to prove uniqueness of martingale problems and to understand the long-time behaviour of Markov processes. We apply these methods in order to define the Fleming-Viot process in random environment. In fact this process arises as a weak limit of the so called Moran processes in random environment which are natural generalizations of their counterpart in deterministic environment. Characterizing the solutions of Fleming-Viot process in random environment, we study its long-time behaviour through the long-time behaviour of its dual process.

1.2 Random walks on the symmetric group and the median problem

1.2.1 Some word metrics on (signed) symmetric groups

When there is no duplication, unichromosomal genomes are represented by permutations or signed permutations. A signed permutation π is a permutation on $\{\pm 1, ..., \pm n\}$ such that $\pi_{-i} = -\pi_i$, and is denoted by

$$\begin{pmatrix} -1 & +1 & ... & -n & +n \\ \pi_{-1} & \pi_{+1} & ... & \pi_{-n} & \pi_{+n} \end{pmatrix}, \tag{1.2.1}$$

or simply by

$$\pi_{-1}\pi_{+1}...\pi_{-n}\pi_{+n}. \tag{1.2.2}$$

Each number represents a gene or a marker in the genome while the signs indicate the orientation of genes. In the genome representation, we let $i_h := +i$ and $i_t := -i$, for $i = 1, ..., n$, indicating the head and the tail of the gene i, respectively. The set of all signed permutations of length n with the composition multiplication is a group called the signed symmetric group (or hyperoctahedral group) of order n denoted by $S_n^{\pm}$. Note that the composition of permutations is applied from right to left. In fact, there exists a group monomorphism from $S_n^{\pm}$ into S_{2n}, where S_m is the symmetric group of order $m \in \mathbb{N}$. Equivalently, $S_n^{\pm}$ can be defined as the wreath product of S_2 and S_n, that is

$$S_n^{\pm} = S_2 \wr S_n. \tag{1.2.3}$$

Genome rearrangement is the study of large scale mutations, rearrangements, over the set of genomes or (signed) permutations. An example of rearrangements is reversal (inversion). Let π be a (signed) permutation. A reversal is a permutation acting on π that reverses a segment $\pi_{-i}\pi_i, ..., \pi_{-j}\pi_j$ and keeps the other positions unchanged, that is, for $1 \leq i \leq j \leq n$, the reversal $\rho(i, j)$ is defined by the permutation

$$\begin{pmatrix} -1 & +1 \; ... & -(i-1) & +(i-1) & -i & +i \; ... & -j & +j & -(j+1) & +(j+1) \; ... & -n & +n \\ -1 & +1 \; ... & -(i-1) & +(i-1) & +j & -j \; ... & +i & -i & -(j+1) & +(j+1) \; ... & -n & +n \end{pmatrix}. \tag{1.2.4}$$

The set of all reversal permutations generates $S_n^{\pm}$. The reversal distance (metric) is the minimum number of reversal permutations needed to transform a permutation into another. In other words, the reversal distance is the graph metric on the Cayley graph of the symmetric group with respect to reversals as the generating set. Another popular metric on (signed) permutation groups is (signed) transposition distance which is the graph metric on Cayley graph of (signed) permutations with respect to transpositions. An example of a more complicated genome rearrangement is Double-Cut and Join (DCJ). Roughly speaking consider a linear genome as a linear segment. Double cut and join is a genomic operator that cuts the line in two different positions between genes and rejoins the segments in the freed positions in one of two possible ways, where one way gives the reversal while the other possible rejoining way gives two new chromosomes, one linear and the other one circular, i.e. one line segment and one circle. We can define the same operator not only on unichromosomal linear genomes, but also on multichromosomal genomes with linear or circular chromosomes. Starting at a linear chromosome, the iteration of DCJ operators always gives a union of a single linear chromosome along possibly with some other circular chromosomes. Adding the extra number 0 at one end of the linear chromosome and joining its two sides together, the linear chromosome can be represented as a circular chromosome including 0. Representing genes by numbers, cycles become the permutation cycles, and hence, whole genome can be represented as an element of the permutation group with an extra number 0. A restricted DCJ is a combination of two consecutive DCJ operators on a linear chromosome, that the first splits the linear chromosome to one linear and one circular chromosomes, and immediately after the creation of the circular chromosome, another DCJ between the linear and the circular chromosomes occurs which causes two chromosomes to be united once again as a new linear chromosome.

1.2.2 Geometric medians and approximating the true ancestors in special random evolutionary models

A geometric median of a finite subset A (with possible multiplicities) of a metric space (S, d) is a point of the space (not necessarily unique) that minimizes the total distance to points of A, i.e. $d_T(x, A) := \sum_{a \in A} d(x, a)$. The set of all medians of A is called the median set. The median problem has played an important role in parsimonious phylogenetics. However the question of credibility of medians in approximating the true ancestor remained unanswered for a long time. Some simulation studies suggest that, in certain random evolutionary models, medians fail to approximate the ancestor after genomes have been involved in the evolution for a long time. An example of such studies can be seen in work of Zheng and Sankoff [66] , in 2010, who observed that for three independently evolving genomes, all starting at a same ancestor, after time, approximately, $0.25n$, where n is the number of genes in the genomes, their heuristic median solution does not match the true ancestor, while before this time it does. Motivated by their work, it is natural to ask when medians are able to give information about the true ancestor. Chapters 2 and 3 are devoted to answer this question, when the genome space, $S_n^{\pm}$, is endowed with reversal distance and the evolution process is modelled by a continuous-time reversal random walk in which the jumps occur at Poisson times of rate 1. Upon a reversal jump two locations of the permutation are chosen uniformly at random (with replacement), and the segment including them (between them) will be reversed. In other words, we can think of a reversal random walk as a simple random walk on the Cayley graph of $S_n^{\pm}$ with respect to reversals as the generating set. Consider k independent reversal random walks $X^{i,n} = (X_t^{i,n})_{t \geq 0}$, $i = 1, ..., k$. We are interested in the asymptotic median set and the asymptotic median value of the permutations sampled from random walks $X^{i,n}$ at time cn, where $c > 0$ is a constant and $n \to \infty$. We are to see that the medians well approximate the true ancestors for $c < \frac{1}{4}$. We show that there exists a constant $0.75 \sim c^* \geq \frac{1}{4}$ for which, after time c^*n, the median cannot approximate the initial state anymore. In fact, we conjecture that a phase transition occurs at the time $\frac{n}{4}$ and, after this time, the median is no longer a good estimator of the ancestor.

Our studies show that, in a general continuous-time random walk $X = (X_t)_{t\geq 0}$ on a discrete metric space (S, d), the median remains a good estimator of the initial state, X_0, under the assumptions of linearity of the distance function

$$D(t) := d(X_0, X_t),$$

asymptotically almost surely (a.a.s.), and symmetry of the state space S. More explicitly, the median well approximates X_0 if

1) there exists a sequence of positive real numbers $(a_n)_{n\in\mathbb{N}}$, where $\lim_{n\to\infty} a_n = \infty$ and $a_n \ll n$ ($a_n/n \to 0$ as $n \to \infty$), such that for any $0 < c < c'$,

$$\frac{d(X_0, X_{cn}) + d(X_{cn}, X_{c'n}) - d(X_0, X_{c'n})}{a_n} \to 0 \quad \textit{in probability}, \tag{1.2.5}$$

and

2) for any $s, t \geq 0$,

$$d(X_s, Y_t) \stackrel{\mathrm{d}}{\equiv} d(X_0, X_{s+t}), \tag{1.2.6}$$

where $Y = (Y_r)_{r\geq 0}$, with $Y_0 = X_0$, is an independent version of X on S. That is, for any $s, t \geq 0$, $d(X_s, Y_t)$ and $d(X_0, X_{s+t})$ have the same distribution.

The latter condition holds when the state space is a Cayley graph of a group.

For simplicity, we formulate the results only for the reversal random walk on $S_n^{\pm}$, as all the transposition random walk on S_n, double cut and join (DCJ), and restricted DCJ random walks on $S_n^{\pm}$ have the same behaviour, and it can be shown that the results are true for them. It is shown by N. Berestycki and R. Durrett in "A phase transition in the random transposition random walk [2]" that the speed of escape of transposition random walk $(X_t)_{t\geq 0}$ on the symmetric group S_n endowed with the transposition metric $d_{tr} := d_{tr}^n$, starting at the identity permutation e_n, is close to the maximum value 1 up to time cn where $c \leq 1/2$. Therefore, the linearity of the distance can be concluded. A phase transition happens after this time, and the distance $d_{tr}(e_n, X_{cn})$ will be sublinear a.a.s.. They proved the similar result for the reversal random walk on $S_n^{\pm}$ endowed with the DCJ metric. We show that the same result is true for the reversal random walk on $S_n^{\pm}$ endowed with the reversal metric. Considering k independent reversal random walks $X^{i,n} = (X_t^{i,n})_{t\geq 0}$, $i = 1, ..., k$, all starting at identity, we prove that, after a convenient rescaling, the identity permutation (the true ancestor) is an asymptotic median of $V_{cn} := \{X_{cn}^{1,n}, ..., X_{cn}^{k,n}\}$ for $c \leq 1/4$ while this is not true when $c > 3/4$. In fact we conjecture that for any $c > 1/4$ the median value does not remain a good estimator of the true ancestor. We also find the approximate positions of all the medians of V_{cn} for $c \leq 1/4$. From the algorithmic point of view, this reduces the median search space significantly.

1.3 Break point median and random permutations

1.3.1 Non-geodesic structure of the symmetric group under the breakpoint metric

Breakpoints are among the first concepts used by biologists for comparing two genomes. In Chapter 4 we study the break point (bp) median value and median set of k independent random permutations, as random linear genomes. The method should carry over to other similar models such as signed permutations and cyclic permutations. For a permutation $\pi := \pi_1 ... \pi_n$ we define the set of adjacencies of π to be all the unordered pairs $\{\pi_i, \pi_{i+1}\} = \{\pi_{i+1}, \pi_i\}$. We denote by $\mathcal{A}_{x,y}$ the set of all common adjacencies of $x, y \in S_n$. The breakpoint (bp) distance between x and y is defined to be $d_{bp}^{(n)}(x, y) := n - 1 - |\mathcal{A}_{x,y}|$, where $|\mathcal{A}_{x,y}|$ is the cardinality of $\mathcal{A}_{x,y}$. The bp distance is a pseudometric because of non-reflexiveness. We say x is equivalent to y if and only if $d_{bp}^{(n)}(x, y) = 0$. The equivalence class containing π is represented by $[\pi]$ and contains exactly two permutations, $\pi_1, ..., \pi_n$ and $\pi_n, ..., \pi_1$. The bp distance, a metric on the set of all equivalence classes of S_n, denoted by $\hat{S}_n := S_n/\sim$, is defined by $d_{bp}^{(n)}([x], [y]) := d_{bp}^{(n)}(x, y)$. $\hat{S}_n$ is not a geodesic metric space. In the absence of geodesics we introduce the notion of partial geodesics (p-geodesics). In a metric space (S, d), a partial geodesic between $x, y \in S$ is a maximal subset of S containing x and y which is isometrically embedded in a subsegment (not necessarily contiguous) of the line segment $[0, 1, ..., d(x, y)]$. For any point z on a partial geodesic between x, y, we have $d(x, y) = d(x, z) + d(z, y)$. A non-trivial p-geodesic is one with at least three points. We denote the set of all permutations lying on p-geodesics connecting $x, y \in S_n$ by $\overline{[x, y]}$.

1.3.2 Breakpoint median value of k random permutations

For k random permutations, the bp median value takes its maximum possible value $(k-1)(n-1)$ after a convenient rescaling a.a.s.. By choosing two points at random, the number of common adjacencies is very small with high probability and the case is similar to when two permutations are at the maximum distance $n-1$. In the later, it is not hard to see that each median of k genomes with pairwisely maximum distances from each other, should take its adjacencies only from the union of adjacencies of those k permutations, and this is the reason for the fact that the median value takes its maximum possible value.

1.3.3 Breakpoint medians of k random permutations

To know the positions of bp medians, we need the concept of accessibility. Let $X := \{x_1, ..., x_k\}$ be a subset of $\hat{S}_n$. We say a permutation class $z \in \hat{S}_n$ is 1-accessible from X if there exists an $m \in \mathbb{N}$, a finite sequence $y_1, ..., y_m$ where $y_i \in X$ and $z_1, ..., z_m$, where $z_i \in \hat{S}_n$ such that $z_1 = y_1, z_m = z$ and $z_{i+1} \in \overline{[z_i, y_{i+1}]}$ for $i = 1, ..., m-1$. The set of all 1-accessible points from X is denoted by $Z_1(X) := Z(X)$. By induction we define the set of all r-accessible permutation classes to be $Z_{r+1}(X) := Z(Z_r(X))$. A permutation class z is said to be accessible from X if there exists $r \in \mathbb{N}$ such that $z \in Z_r(X)$. We denote the set of all accessible points by $\overline{Z}(X) = \cup_{r \in \mathbb{N} \cup \{0\}} Z_r(X)$. Roughly speaking, if X contains k random permutations, then we can show that $\overline{Z}(X)$ is contained in the bp median set of X a.a.s. with a convenient rescaling. One can ask if the converse is true. In other words, do we have $\overline{Z}(X) = M(X)$?

1.4 Population models in random environments

Chapter 5 studies the population models in random environment. The classic Moran process is a basic population model which models a population with N individuals with types in a metric space I. In this section we introduce a finite population system in random environment, namely the Moran process in random environment. For $N \in \mathbb{N}$, Let $M_N := \{1, ..., N\}$ be the set of individuals in a population. Let I be a metric space. In the sequel, we always assume that I is compact. Let $(\Omega, \mathbb{P}, \mathcal{F})$ be a probability measure space. In general, a bounded fitness process is a measurable stochastic process $e : \Omega \times \mathbb{R}_+ \times I \to [0, 1]$. For simplicity, we let $e_t(.) = e(t, .)$. Consider the families of independent Poisson point processes $\pi^{i,j}_{res}$ with rate $\gamma > 0$ and $\pi^{i,j}_{sel}$ with rate $\frac{\alpha}{N} > 0$ for every ordered pair of individuals (i, j), and π^{i}_{mut} with rate

$\beta > 0$ for every individual i. Let

$$\begin{aligned}\pi_{res} &= \{\pi_{res}^{i,j} : ordered\ i \neq j \in M_N\} \\ \pi_{sel} &= \{\pi_{sel}^{i,j} : ordered\ i \neq j \in M_N\} \\ \pi_{mut} &= \{\pi_{mut}^{i} : i \in M_N\}.\end{aligned} \tag{1.4.1}$$

A particle Moran process with N individuals, denoted by $Y = Y^{N,e} = (Y^{N,e}(t))_{t\geq 0}$ (we drop N and e when there is no risk of ambiguity), is a purely jump Markov process on the state space I^N with

$$Y^{N,e}(t) := (Y_1^{N,e}(t), ..., Y_N^{N,e}(t)) \tag{1.4.2}$$

for $t \geq 0$, where for any $i \in M_N$, $Y_i : \Omega \times \mathbb{R}_+ \to I$ is a purely jump Markov process. In fact the process Y_i can be interpreted as the configuration (type) process, namely each individual carries an allele or type (also called configuration) during the time $[0, \infty)$. The process Y evolves as a purely jump Markov process where the jumps occur at random Poisson times $\pi = \pi_{res} \cup \pi_{sel} \cup \pi_{mut}$. Let $Y_0 = (u_1, ..., u_N) \in I^N$ be the starting state. At every random time $\tau \in \pi_{res}^{i,j}$ the individuals i, j involve in a resampling event with probability $\frac{1}{2}$, and with probability $\frac{1}{2}$ no jump occurs. Upon a resampling event between i and j the individual j dies and is replaced by an offspring of the individual i. In other words, the type of j, i.e. $Y_j(\tau)$, is replaced by the type of i, $Y_i(\tau)$. Similarly, for any $i = 1, ..., N$, at a random time $\tau \in \pi_{mut}^{i}$, $Y_i(\tau)$ jumps according to a one step probability kernel on I, namely $q(x, dy)$. This is called the mutation kernel. Finally, at a random time $\tau \in \pi_{sel}^{i,j}$, for a pair of individuals (i, j), a selective event occurs with probability $e_\tau(Y_i(\tau))$ where the individual j dies and is replaced by an offspring of the individual i. At this time, there is no jump with probability $1 - e_\tau(Y_i(\tau))$.

Note that, always, there exist constants $\beta', \beta'' > 0$ and probability kernels $q'(dy)$ and $q''(x, dy)$, where

$$\beta q(x, dy) = \beta' q'(dy) + \beta'' q''(x, dy). \tag{1.4.3}$$

We call $q'(dy)$ the parent-independent component of the mutation.

Considering the frequency of alleles at each time, it is convenient to project $Y^{N,e}$ onto a purely atomic (with at most N atoms) measure-valued process on $\mathcal{P}^N(I)$, that is the space of probability measures on I with at most N atoms. More precisely, for any $t \geq 0$, let

$$\mu_N^e(t) = \frac{1}{N} \sum_{i=1}^{N} \delta_{Y_i^{N,e}(t)} \tag{1.4.4}$$

where, for $a \in I$, δ_a is the delta measure on a. For some results in this thesis, we assume that e is independent of the initial distribution of μ_N^e and also independent of Poisson times of jumps (for the dual process). Let E be a compact subset of $\mathcal{C}(I, [0,1])$ equipped with the sup-norm topology. We assume that the fitness process is a measurable stochastic process with sample paths in $D_E[0,\infty)$, the space of càdlàg functions endowed with the Skorokhod topology. Letting $N \to \infty$, the Fleming-Viot process in random environment arises as the weak limit of μ_N^e in $D_E[0,\infty)$. We characterize this process as a solution to a martingale problem in random environment (called quenched martingale problem). The main purpose of Chapter 5 is the study of the long-time behaviour of Fleming-Viot processes in random environment. In order to do that, we develop the duality method to the case of time-dependent martingale problems.

Chapter 2

Phase change for the accuracy of the median value in estimating divergence time [1]

In this chapter, we prove that for general models of random gene-order evolution of $k \geq 3$ genomes, as the number of genes n goes to ∞, the median value approximates k times the divergence time if the number of rearrangements is less than $cn/4$ for any $c \leq 1$. For some $c^* > 1$, if the number of rearrangements is greater than $c^*n/4$, this approximation does not hold.

2.1 Introduction

The iterative improvement of approximate solutions to the Steiner tree problem by optimizing one internal vertex at a time has a substantial history in the "small phylogeny" problem for parsimony-based phylogenetics, both at the sequence level [51] and the gene order level [4]. It has been generalized to iterative local subtree optimization methods, such as "tree-window-hill" [49] and "disc covering" [32, 60]. Here we focus on the "median problem" for gene order, where we estimate the location of a single point (the median) in a metric space, given the location of the three or

[1] This chapter is basically the paper published in [37].

more points connected to the median by an edge of the tree. Given $k \geq 3$ signed gene orders $G_1, \cdots, G_k$ on a single chromosome or several chromosomes, and a metric d such as breakpoints [50], inversions [52], inversions and translocations [6], or double-cut-and-join [65], we are to find the gene order M such that $\sum_{i=1}^{k} d(G_i, M)$ is minimized.

Although it plays a central role in gene order phylogeny, the median suffers from several liabilities. One is that it is hard to calculate in most metric spaces. Not only is it NP-hard [61], but exhaustive methods are costly for most instances, namely unless $G_1 \cdots, G_k$ are all relatively similar to each other, which we will refer to generically as the *similar genomes condition*. Another problem is that heuristics tend to produce inaccurate results unless a suitable similar genomes condition holds [66]. Still another, is the tendency in some metric spaces to degenerate solutions [29] unless the same conditions prevails.

In this chapter we add to this litany of difficulties by showing that as k genomes evolve over time, as modeled by any one of several biologically-motivated random walks, there is a phase change after $n/4$ steps, where n is the number of genes. With $u < n/4$ steps, the sum of the normalized distances $\sum_k d/n$ from each of the genomes to the starting point – the ancestor – converges to ku/n in probability, and this is the median value. When $u > c^* n/4$ steps, for a constant $c^* \geq 1$, the sum of the normalized distances to the median converges in probability to a value less than ku/n, and that the ancestor is no longer the median.

Our proof is inspired by a result of Berestycki and Durrett [2] in showing that the reversal distance between two signed permutations converges in probability to the actual number of steps, after rescaling, if and only if $u < n/2$. The technique is to construct a graph with genes as vertices and edges added between vertices according to how they are affected by reversals. Properties of the number of components of random Erdős-Rényi graphs can then be invoked to prove the result.

2.2 Definitions

We represent a unichromosomal genome by a signed permutation, where the sign indicates whether the gene is "read" from left to right (tail-to-head) or from right to left (head to tail) on the chromosome. Let $S_n^{\pm}$ be the signed symmetric group of order n, i.e. the space of all signed permutations of length n. A *reversal* operation applied to a signed permutation reverses the order, and changes the signs, of one or more adjacent terms in the permutation. A DCJ operation, which can apply not only to signed permutations but to more general genomes containing linear and circular chromosomes, *cuts* the genome in two places and rejoins pairs of the four "loose ends" in one of two possible new ways (one of which may be equivalent to a reversal). We define the reversal and DCJ distances, denoted by $d_r^{(n)}$ and $dcj^{(n)}$, to be the minimum number of reversal and DCJ operations, respectively, needed to transform one signed permutation (genome) to another. For simplicity, when it is not ambiguous, we drop the superscript "n".

For any permutation $\Pi \in S_n^{\pm}$, let $\Pi_0 := 0$ and $\Pi_{-(i+1)} = -(i+1)$. The *breakpoint graph* of signed permutations $\Pi, \Pi' \in S_n^{\pm}$, denoted by $BP(\Pi, \Pi')$, is a 2-regular graph with the set of vertices $\{+0, -1, +1, -2, +2, ..., -n, +n, -(n+1)\}$, black edges $\{\Pi_{+i}, \Pi_{-(i+1)}\}$, for $i = 0, ..., n$, and grey edges $\{\Pi'_{+i}, \Pi'_{-(i+1)}\}$, for $i = 0, ..., n$. In other words, $BP(\Pi, \Pi')$ contains vertices for the head and tail of each gene, *black edges* defined by the adjoining heads or tails of two adjacent genes in the genome Π and *grey edges* defined by two adjacent genes in the genome Π'. We use the usual representation of $BP(\Pi, \Pi')$ in which the vertices are located along a horizontal line in the plane, with order $\Pi_0, \Pi_{-1}, \Pi_{+1}, ..., \Pi_{-n}, \Pi_{+n}, \Pi_{-(n+1)}$ from left to right, and the black edges are represented by horizontal segments. Let $id^{(n)} = I_n$ be the identity permutation (when there is no risk of ambiguity, we drop n). Let $BP(\Pi) = BP(\Pi, id)$. It is well-known that

$$dcj(\Pi) = n + 1 - |cBP(\Pi)|, \tag{2.2.1}$$

where $cBP(\Pi)$ is the set of components of $BP(\Pi)$ with cardinality $|cBP(\Pi)|$[26]. It is clear, from the definition, that every element of $cBP(\Pi)$ is a cycle graph.

We need to define an orientation for grey (and black) edges of $BP(\Pi)$. We traverse a cycle $c \in cBP(\Pi)$ in a counter-clockwise manner if we start at the left-most vertex of $BP(\Pi)$ (in the usual representation), travel along its unique adjacent black edge and end at the same vertex through its unique adjacent grey edge. Then we say a black edge in c is positively oriented if we move along it from left to right in a counter-clockwise traversal. Otherwise we say it is negatively oriented. Similarly, for the grey edge $(+i, -(i+1))$ (i.e. $(i_t, (i+1)_h)$ with the genome representation) we say it is positively oriented if during a counter-clockwise traversal we move along it from $+i$ to $-(i+1)$ (i.e. from i_t to $(i+1)_h$, with the genome representation). Otherwise it is negatively oriented. We define the orientation function ξ on the edge u of $BP(\Pi)$ to be:

$$\xi(u) = \begin{cases} +1 & \text{if } u \text{ is positively oriented} \\ -1 & \text{if } u \text{ is negatively oriented.} \end{cases} \tag{2.2.2}$$

We say the black (grey) edges u, u' are parallel, denoted by $u \parallel u'$ if $\xi(u) = \xi(u')$. Otherwise we say they are crossing. This is just a reformulation of Hannenhalli and Pevzner's original concept of oriented cycles[30]. An oriented cycle in this definition is a cycle including at least one positively and one negatively oriented black edge. The mechanism by which a reversal affects a genome can easily be seen using the BP graph. Consider a reversal acting on two black edges u, u' in $BP(\Pi)$, i.e. it reverses the segment between u and u' in the breakpoint graph. If they are in two different cycles we have a merger of the two to construct a new cycle. But if u and u' are in a same cycle, that cycle either splits, if $u \nparallel u'$, or does not split if $u \parallel u'$.

2.3 Limit behavior of the median value

Let d^n be a metric on the space of all signed permutations length n. For a set $A \subset S_n^{\pm}$ of signed permutations, define the total distance to A by

$$d_T^n(.,A) : S_n^{\pm} \longrightarrow \mathbb{N}_0 = \mathbb{N} \cup \{0\}, \tag{2.3.1}$$

$$d_T^n(x,A) := \sum_{y \in A} d^n(x,y). \tag{2.3.2}$$

Then let

$$m^{d,n}(A) := min\{d_T^n(x,A) : x \in S_n^{\pm}\}. \tag{2.3.3}$$

$m^{d,n}(A)$ is called the median value of A under the metric d^n. A signed permutation which minimizes $d_T^n(.,A)$ is called a median of A.

Let $X_0^n = id$, the identity permutation, and let X_t^n be a continuous-time random walk on $S_n^{\pm}$, where at random Poisson times τ_κ, with rate 1, we choose two numbers $1 \leq i \leq j \leq n$ and let $\rho(i,j)$ operate on $X_{\tau_\kappa}^n$, that is

$$X_{\tau_\kappa^+}^n = X_{\tau_\kappa}^n \; o \; \rho(i,j), \tag{2.3.4}$$

where $\rho(i,j)$ is the reversal acting on i and j. We call X_t^n a reversal random walk (r.w.) on $S_n^{\pm}$. Let $X_t^{1,n}, \ldots, X_t^{k,n}$ be k independent reversal random walks, all starting at the identity element, id. Define

$$A_t^{(n)} := \{X_t^{1,n}, \ldots, X_t^{k,n}\} \tag{2.3.5}$$

and

$$\varepsilon_t^{d,n} := d_T^n(id, A_t^{(n)}) - m^{d,n}(A_t^{(n)}). \tag{2.3.6}$$

We investigate the time up to which the median value of $X_t^{1,n}, \ldots, X_t^{k,n}$, namely $m^{d,n}(A_t)$, remains a good estimator for the total divergence time, kt, as well as to the total distance of points in A_t to id, namely $d_T^n(id, A_t^{(n)})$. To answer this question we use the fact that the speed of escape of the r.w. up to some particular time, is the same from any point of the space and is close to 1, the maximum value. Berestycki and Durrett studied speed of transposition and reversal random walks with the related word distances, while in the latter they used "approximate reversal distance" instead of reversal itself, ignoring the effect of hurdles. This turns out to be the same as DCJ distance on single chromosomes. We have

$$d_r(\pi, I) = n + 1 - c(\pi) + h(\pi) + \tilde{f}(\pi), \tag{2.3.7}$$

while

$$dcj(\pi, I) = n + 1 - c(\pi), \tag{2.3.8}$$

where $h(\pi)$ and $\tilde{f}(\pi)$ are the number of hurdles and fortresses, respectively.

Although Berestycki and Durrett only proved their theorem for the random transposition r.w. on S_n, they suggested that the same method should carry over to reversal r.w. The following proposition is proved in [2] for *approximate* reversal distance (i.e., DCJ distance).

In this result, and in the ensuing discussion, $(a_n)_{n\in\mathbb{N}}$ is an arbitrary sequence such that $a_n \to \infty$ as $n \to 0$. When it is unambiguous, we drop n from $A_t^{(n)}$ and X_t^n.

Proposition 2.3.1. *[Berestycki-Durrett] Let c be fixed and let X_t be a reversal r.w. on $S_n^{\pm}$ starting at id. Then*

$$dcj(id, X_{cn/2}) = (1 - f(c))n + w(n), \tag{2.3.9}$$

where

$$f(c) := \frac{1}{c} \sum_{k=1}^{\infty} \frac{k^{k-2}}{k!} (ce^{-c})^k \tag{2.3.10}$$

and $\frac{w(n)}{a_n\sqrt{n}} \to 0$ in probability, as $n \to \infty$.

Remark 2.3.2. *The function $1 - f$ is linear for $c \leq 1$, $f(c) = 1 - c/2$, and sublinear for $c > 1$, $1 - f(c) < c/2$. This means that for $c \leq 1$*

$$dcj(id, X_{cn/2}) - \frac{cn}{2} = w(n), \tag{2.3.11}$$

and r.w. travels on an approximate geodesic (or parsimonious path) asymptotically almost surely. f is the function counting the number of tree components of an Erdős-Rényi random graph with n vertices for which the probability of having each edge is $\frac{c}{n}$, denoted by $G(c, n)$. See Theorem 12 in [5], Chapter V.

We extend the above theorem for the bonafide reversal distance. To do so, we need to estimate the number of hurdles of $X_{\frac{cn}{2}}$. Recall that an oriented cycle in a breakpoint graph is a cycle including an orientation edge, that is a grey edge with two black adjacency edges e, e', where a reversal involving e and e' splits the cycle [30]. As we discussed, this is equivalent to saying $e \nparallel e'$. It is not difficult to conclude the following result.

Lemma 2.3.3. *Let $C \in cBP(\pi)$, then C is oriented if and only if there exists exactly two equivalence classes of black edges, that is there exist at least two black edges with different signs.*

Proof: Suppose all the grey edges are unoriented. Start a travel on C at an arbitrary black edge e_1 counterclockwisely. Since the consecutive grey edge after e_1 is unoriented, we have $e_1 \parallel e_2$ where e_2 is the second black edge in the travel. Continuing this way, our travel covers all the black edges of c with $e_1 \parallel e_2 \parallel ... \parallel e_{|c|}$. Conversely, suppose there exists an oriented grey edge in c. Then, obviously, its black adjecancy edges, namely e and e', are not parallel, i.e. $e \nparallel e'$. ∎

Theorem 2.3.4. *Let $c > 0$ be fixed and let X_t be a reversal r.w. starting at id. Define $h_t := h(X_t)$ to be the number of hurdles in $BP(X_t)$. Then, as $n \to \infty$,*

$$\frac{h_{cn/2}}{a_n\sqrt{n}} \to 0 \quad \textit{in probability.} \tag{2.3.12}$$

Proof: Cycles of the BP that have never been involved in a fragmentation event must be oriented, since the two rejoined black edges resulting from an inversion-induced merger of cycles cannot be parallel.

Therefore we need only to count the number of edges that have been involved in a fragmentation event. To do so we apply the method of counting cycles in [2], Theorem 3. Hurdles occur only in those cycles with length more than one that have been involved in a fragmentation up to time $\frac{cn}{2}$. We call such cycles fragmented cycles. The number of fragmented cycles with length more than $\sqrt{n}$ is always less than $\sqrt{n}$. But to count all fragmented cycles in $X_{\frac{cn}{2}}$ with size less than $\sqrt{n}$, we need to find an upper bound for the rate of a fragmentation up to time $\frac{cn}{2}$. Since a fragmentation occurs when two black edges in one cycle are chosen, to fragment a cycle in BP for any chosen black edge e, we only can pick another black edge e' in the same cycle whose graph distance in the breakpoint graph is less than $2\sqrt{n}$. (The coefficient 2 arises from the fact that the cycles are alternating in BP.)

Thus the rate of fragmentation at an arbitrary time t is not more than $\frac{n}{n} \cdot \frac{2\sqrt{n}}{n} = \frac{2}{\sqrt{n}}$. Integrating up to time t, this gives us the expected number of fragmented cycles at time t which is $\frac{2}{\sqrt{n}} \cdot t$. For $t = \frac{cn}{2}$ this expectation is $c\sqrt{n}$. Now, dividing by $a_n\sqrt{n}$, the result follows from Chebyshev's inequality and the fact that hurdles only occur in fragmented cycles. ■

Theorem 2.3.5. *Let X_t be a reversal r.w. on $S_n^{\pm}$ starting at id and let $d_r := d_r^{(n)}$ denote the reversal distance on $S_n^{\pm}$. Then, with $c > 0$ fixed,*

$$d_r(id, X_{cn/2}) = (1 - f(c))n + w'(n), \tag{2.3.13}$$

where f is the same function as in the statement of Proposition 2.3.1 and $w'(n)$ is a function with $\frac{w'(n)}{a_n\sqrt{n}} \longrightarrow 0$ in probability, as $n \to \infty$.

Proof: Since $d_r(\Pi) = dcj(\Pi) + h(\Pi) + \tilde{f}(\Pi)$ by the proposition, we have $d_r(X_{cn/2}) = (1 - f(c))n + w(n) + h_{cn/2} + \tilde{f}(X_{cn/2})$. But

$$\frac{w'(n)}{a_n\sqrt{n}} := \frac{w(n) + h_{cn/2} + \tilde{f}(X_{cn/2})}{a_n\sqrt{n}} \longrightarrow 0 \tag{2.3.14}$$

in probability, as $n \to \infty$, by the convergence of $\frac{w(n)}{a_n\sqrt{n}}$ and $\frac{h_{cn/2}}{a_n\sqrt{n}}$ in Proposition 2.3.1 and Theorem 2.3.4 and $\tilde{f}(X_{cn/2}) \leqslant 1$. ∎

Theorem 2.3.6. *Let $X_t^{1,n}, \ldots, X_t^{k,n}$ be k independent reversal r.w. on $S_n^{\pm}$ starting at id. Suppose either*

a) $d^{(n)} := dcj^{(n)}$ dcj distance

or

b) $d^{(n)} := d_r^{(n)}$ reversal distance

Then for $c < \frac{1}{4}$ we have $\frac{\varepsilon_{cn}^{d,n}}{a_n\sqrt{n}} \longrightarrow 0$ in probability, as $n \to \infty$.

Proof: We prove the theorem only for d_r. The proof of the DCJ case is similar. For all $i, j \in \{1, \ldots, k\}$ and for a median x of $A_t^{(n)}$

$$d_r^{(n)}(X_t^{i,n}, X_t^{j,n}) \leq d_r^{(n)}(x, X_t^{i,n}) + d_r^{(n)}(x, X_t^{j,n}). \tag{2.3.15}$$

Therefore,

$$\sum_{i \neq j} d_r^{(n)}(X_t^{i,n}, X_t^{j,n}) \leq \sum_{i \neq j}(d_r^{(n)}(x, X_t^{i,n}) + d_r^{(n)}(x, X_t^{j,n})). \tag{2.3.16}$$

We now conclude

$$\sum_{i \neq j} d_r^n(X_t^{i,n}, X_t^{j,n}) \leq (k-1)m^{d,n}(A_t^{(n)}) \leq (k-1)d_T^n(id, A_t^{(n)}), \tag{2.3.17}$$

where d_T^n is the total distance with respect to the reversal metric. Let $c \leq \frac{1}{4}$. Then by Theorem 2.3.5 we have for all $i, j \quad i \neq j$

$$d_r^{(n)}(X_{cn}^{i,n}, X_{cn}^{j,n}) = 2cn - w(n) \tag{2.3.18}$$

and

$$d_r^{(n)}(id, X_{cn}^{i,n}) = cn - w(n), \tag{2.3.19}$$

as $n \to \infty$, where $\frac{w(n)}{a_n\sqrt{n}} \to 0$ in probability. Thus

$$\binom{k}{2}(2cn - w(n)) \leq (k-1)m^{d,n}(A_{cn}) \leq (k-1)k(cn - w(n)). \tag{2.3.20}$$

Then

$$|m^{d,n}(A_{cn}) - kcn| \leq k'w(n) \tag{2.3.21}$$

for a constant k'. Also $|d_T^n(id, A_{cn}) - kcn| \leq kw(n)$. Therefore, there exists a constant k^* such that

$$|m^{d,n}(A_{cn}) - d_T^n(id, A_{cn})| \leq k^*w(n). \tag{2.3.22}$$

This, as $n \to \infty$, implies

$$\frac{\varepsilon_{cn}}{a_n\sqrt{n}} = \frac{m^{d,n}(A_{cn}) - d_T^n(id, A_{cn})}{a_n\sqrt{n}} \to 0 \ \ in \ \ probability. \tag{2.3.23}$$

This proves the theorem. ■

Remark 2.3.7. *The statement of the theorem suggests that, ignoring the error of order $o(a_n\sqrt{n})$ for $a_n \longrightarrow \infty$, id remains as the median of leaves of k independent reversal random walks $X_t^{1,n}, \ldots, X_t^{k,n}$ up to time $\frac{n}{4}$ asymptotically almost surely.*

Corollary 2.3.8. *Let $c \leqslant \frac{1}{4}$ be fixed. Suppose d is either DCJ or reversal distance. Then, by the hypothesis of Theorem 2.3.6, as $n \to \infty$,*

$$\frac{kcn - m^{d,n}(A_{cn})}{a_n\sqrt{n}} \longrightarrow 0 \quad in\ probability,\ as\ n \longrightarrow \infty. \tag{2.3.24}$$

Proof: This follows directly from the fact that , as $n \to \infty$,

$$\frac{kcn - d_T^n(id, A_{cn}^{(n)})}{a_n\sqrt{n}} \longrightarrow 0 \tag{2.3.25}$$

in probability. ■

It is natural to ask whether the statement of Corollary 2.3.8 is also true for some time after $\frac{n}{4}$. In other words, is the median value kcn a fair estimator for the total time of divergence? We conjecture that the property is lost after time $\frac{n}{4}$, but for now can only prove a weaker upper bound.

Corollary 2.3.9. *Let $c > \frac{1}{2}$ be fixed. Suppose d is either DCJ or reversal distance. Then by the same hypothesis as in Theorem 2.3.6, there exists $\alpha_c > 0$ such that*

$$kcn - m^{d,n}(A_{cn}) \geq \alpha_c n + w(n), \tag{2.3.26}$$

where $\frac{w(n)}{a_n\sqrt{n}} \to 0$ in probability, as $n \to \infty$.

Proof:

$$kcn - m^{d,n}(A_{cn}) \geqslant kcn - d_T^n(id, A_{cn}) = kcn - k(1 - f(2c))n + w(n), \tag{2.3.27}$$

where $\frac{w(n)}{a_n\sqrt{n}} \longrightarrow 0$ in probability, as $n \to \infty$. Let $\alpha_c := k(c - (1 - f(2c)))$ which is strictly positive for $c > \frac{1}{2}$. ■

Remark 2.3.10. *This theorem shows that after time $\frac{n}{2}$, the error is of order n, and so the median value is not a good estimate of k times the divergence time.*

In fact, since $f(c)$ is decreasing for $c > 0$, and $f(c) = 1 - \frac{c}{2}$ for $c < 1$, it is easy to see that in the case $k = 3$, for $c > 0.75$, $\varepsilon_{cn}^{d,n}$ is of order $\beta_c n$ for some $\beta_c > 0$.

Theorem 2.3.11. *Let $k = 3$ and d be either dcj or d_r. Consider the same hypothesis as in Theorem 2.3.6. Let c^* be solution of*

$$f(\frac{x}{2}) = \frac{1}{3}. \tag{2.3.28}$$

Then, for all $c > c^*$, there exists $\beta_c > 0$ such that

$$\varepsilon_{\frac{cn}{4}}^{d,n} \geq \beta_c n + w(n), \tag{2.3.29}$$

where $\frac{w(n)}{a_n\sqrt{n}} \longrightarrow 0$ in probability, as $n \to \infty$.

Proof: We have

$$m^{d,n}(A_{\frac{cn}{4}}) \leqslant d(X^{1,n}_{\frac{cn}{4}}, X^{2,n}_{\frac{cn}{4}}) + d(X^{1,n}_{\frac{cn}{4}}, X^{3,n}_{\frac{cn}{4}}). \tag{2.3.30}$$

Because of symmetry and regularity of Cayley graphs (so reversibility of simple random walks on them), we have $d(X^{1,n}_{\frac{cn}{4}}, X^{i,n}_{\frac{cn}{4}}) \overset{\mathrm{d}}{\equiv} d(id, X^{1,n}_{\frac{cn}{2}})$, for $i = 2, 3$. Hence,

$$d(X^{1,n}_{\frac{cn}{4}}, X^{2,n}_{\frac{cn}{4}}) + d(X^{1,n}_{\frac{cn}{4}}, X^{3,n}_{\frac{cn}{4}}) = 2(1 - f(c))n + w'(n), \tag{2.3.31}$$

where $\frac{w'(n)}{a_n\sqrt{n}} \longrightarrow 0$ in probability, as $n \to \infty$. For $c > 0$, let

$$\beta_c = -2(1 - f(c)) + 3(1 - f(\frac{c}{2})). \tag{2.3.32}$$

If $\beta_c > 0$, then

$$\varepsilon^{d,n}_{\frac{cn}{4}} \geq d^n_T(id, A_{\frac{cn}{4}}) - (d(X^{1,n}_{\frac{cn}{4}}, X^{2,n}_{\frac{cn}{4}}) + d(X^{1,n}_{\frac{cn}{4}}, X^{3,n}_{\frac{cn}{4}})) = \beta_c n + w(n), \tag{2.3.33}$$

where $\frac{w(n)}{a_n\sqrt{n}} \to 0$ in probability, as $n \to \infty$. So it suffices to prove $\beta_c > 0$ for $x = c > c^*$. Since $f(x) > 0$ for all $x > 0$,

$$1 + 2f(x) - 3f(\frac{x}{2}) > 1 - 3f(\frac{x}{2}), \tag{2.3.34}$$

in which the right hand side is strictly increasing, Therefore, for all $c \geq c^*$,

$$1 + 2f(c) - 3f(\frac{c}{2}) > 1 - 3f(\frac{c^*}{2}) = 0. \tag{2.3.35}$$

This proves the statement. ■

Now, we would like to measure the volume of that part of the space $S^{\pm}_n$ for which median does well, compared with the whole space. The ratio of the two converges to 0 as n goes to ∞, showing that the median is only useful in a highly restricted region of the space. The following theorem is entailed by a theorem in [59]. Let $c_n = c_n(\Pi)$ be the number of cycles in the BP graph of a random $\Pi \in S^{\pm}_n$. Let d^n be a distance (metric) on $S^{\pm}_n$. Define

$$B^d_{cn} = B^{d,n}_{cn} := \{\Pi \in S^{\pm}_n, d^n(\Pi, id) \leq cn\} \tag{2.3.36}$$

to be the ball of radius cn in $S^{\pm}_n$.

Proposition 2.3.12. *Let $0 < c < 1$ be fixed. Then*

i)

$$\gamma_n = \frac{|B_{cn}^{dcj}|}{|S_n^{\pm}|} \longrightarrow 0 \quad as \quad n \longrightarrow \infty, \tag{2.3.37}$$

ii)

$$\gamma'_n = \frac{|B_{cn}^{d_r}|}{|S_n^{\pm}|} \longrightarrow 0 \quad as \quad n \longrightarrow \infty. \tag{2.3.38}$$

Proof:

i)

$$For\ all\ \Pi \in B_{cn}^{dcj} \quad , \quad |cBP(\Pi)| \geq (1-c)n. \tag{2.3.39}$$

Suppose γ_n does not converge to 0. Therefore there exists a subsequence $\{n_i\}_{i \in \mathbb{N}}$ such that $\gamma_{n_i} \geq \varepsilon$ for a constant $\varepsilon > 0$. This implies

$$E(c_{n_i}) \geq \varepsilon(1-c)n_i. \tag{2.3.40}$$

But by Theorem 2.2 in [59], we have

$$\frac{E(c_{n_i})}{n_i} \longrightarrow 0 \quad as \quad n_i \longrightarrow \infty. \tag{2.3.41}$$

That is in contradiction with the above inequality since , as $n \to \infty$,

$$\frac{\varepsilon(1-c)n_i}{n_i} \longrightarrow \varepsilon(1-c) > 0. \tag{2.3.42}$$

ii) For the second part it suffices to observe that for all $\Pi \in S_n^{\pm}$ we have

$$d_r(\Pi) \geqslant dcj(\Pi). \tag{2.3.43}$$

Therefore

$$B_{cn}^{d_r}(\Pi) \subset B_{cn}^{dcj}(\Pi) \tag{2.3.44}$$

and the result follows from part (i), since

$$\gamma'_n \leqslant \gamma_n \longrightarrow 0 \quad as \quad n \longrightarrow \infty. \tag{2.3.45}$$

■

2.4 Conclusion

We have shown that the median value for DCJ and for reversal distance for a reversal r.w. has good limiting properties if the number of steps remains below $cn/4$, for any $c \leq 1$, but for some value $c > 1$, more than this number of steps destroys these limiting properties. The critical value may indeed be $c = 1$, but for now we can only show that for $c > 3$ (and $c > 2$) the median value is no longer a good estimator of the total distance between the *id* and the current positions of the random walks (and k times the divergence time, respectively).

Note that a simulation strategy to estimate c is not available because of the hardness of calculating the median. As n increases even to moderate values, all exact methods require prohibitive computing time.

These results imply that the steinerization strategy for the small phylogeny problem may lead to poor estimates of the interior nodes of a phylogeny unless the taxon sampling is sufficient to assure that a "similar genomes condition" holds for every k-tuple of genomes used in the course of the iterative optimization search. This can be monitored prior to each step in the iterative optimization of the phylogeny through successive application of the median method.

Chapter 3

Asymptotic medians of random permutations sampled from reversal random walks [1]

3.1 Introduction

Medians can serve as good estimators of the ancestor (the initial state) for k independent reversal random walks on the space of signed permutations before time $\frac{n}{4}$, that is the initial state is a median of k random genomes sampled from k independent random walks at time cn where $c \leq 1/4$. In this chapter we relax the time scale of the individual random walks, investigate the positions of all possible medians other than the initial state, and reduce the state space necessary for median search algorithms.

The median plays an important role in the comparative genomics study of chromosomal rearrangements [61]. It is not only the archetypical phylogenetic instance – one unknown ancestor, $k \geq 3$ observed genomes – for the small phylogeny problem using unrooted trees, but it is also the innermost calculation for the iterative "steinerization" procedure for more larger instances of this problem, with several ancestral nodes [66].

[1] This chapter has been submitted to a scientific journal [38].

In a metric space, the median is a point whose sum of distances to k given points is minimized. In the simplest case, where a genome is a signed permutation on $\{1,\ldots,$ n$\}$, the biologically relevant metric is the reversal distance, where a reversal involves a contiguous subset of the elements in a permutation reversed in order and sign. The distance is the minimum number of reversals necessary to transform one genome to another.

The median problem is NP-hard for reversal distance [7], and the search space is very large. Reducing the search space to a much smaller subdomain would be a significant help in practice.

In a previous paper [37] we showed that the starting point of k independent reversal random walks on the signed permutation group remains approximately a median for the k current positions up to time $n/4$.

In the present chapter, we improve this result to allow the time scales to differ among the k random walks. We then show how to find the positions of all the other medians beside the starting point. In doing this we reduce the search space considerably for median search algorithms.

3.2 Main results

Where there is no duplication, unichromosomal genomes are represented by permutations or signed permutations. A signed permutation is a permutation π on $\{\pm 1, ..., \pm n\}$ such that $\pi_{-i} = -\pi_i$ (see Fertin *et al.* [26]). Each number represents a gene in the genome while its sign indicates its orientation or polarity (called strandedness in the biological literature). The set of all signed permutations of length n with the composition multiplication is a group called the signed symmetric group of order n denoted by $S_n^{\pm}$. Genome rearrangement is the study of large scale mutations, rearrangements, over the set of genomes or (signed) permutations. An example of a rearrangement is a reversal (called inversion in the biological literature). Let

$\pi = \pi_{-1}\pi_1...\pi_{-n}\pi_n$ be a (signed) permutation. A reversal is a permutation multiplying by π from right which reverses a segment $\pi_{-i}\pi_i, ..., \pi_{-j}\pi_j$ and keeps the other positions unchanged. Other words, for any $1 \leq i \leq j \leq n$, a reversal permutation reversing segment $\pi_{-i}\pi_i, ..., \pi_{-j}\pi_j$ is

$$-1 \ +1 \ ... \ -(i-1) \ \ +(i-1) \ +j \ -j \ ... \ +i \ -i \ -(j+1) \ +(j+1) \ ... \ -n \ +n \tag{3.2.1}$$

The reversal distance between two signed permutations π and π' is the minimum number of reversal permutations needed to transform π into π'. In fact, the reversal distance is a metric on $S_n^{\pm}$. Alternatively, one can define the reversal distance as the distance on the Cayley graph of $S_n^{\pm}$ with respect to reversals. More explicitly, the set of all reversal permutations generates $S_n^{\pm}$. We denote by G_n the Cayley graph of $S_n^{\pm}$ with respect to the reversal permutations as the generating set. The reversal distance is the graph metric on G_n. We denote by $d^{(n)}$ the reversal distance on $S_n^{\pm}$.

A reversal random walk on $S_n^{\pm}$ is a continuous-time simple random walk on G_n starting at identity element e_n where the jumps occur with rate 1, that is starting at e_n, at each position in $S_n^{\pm}$ the random walker chooses one of its $\binom{n}{2} + n$ neighbours with equal probability, and jumps to it in a Poisson time rate 1.

A median of a finite subset A (with possible multiplicities) of a metric space (S, ρ) is a point of S (not necessarily unique) that minimizes $\rho_T(., A) := \sum_{a \in A} \rho(., a)$. The function $\rho_T(x, A)$ is called the total distance of x to A. The set of all medians of A is called the median set. The total distance of a median of A to A is called the median value. In other words, the median value of A is the minimum value of $\rho_T(x, A)$ over all $x \in S$. The median problem has played an important role in phylogeny reconstruction to approximate the true ancestor, and more generally, to reconstruct the ancestral tree. The question that arises is: When does the median approximate the true ancestor? For $(S_n^{\pm}, d^{(n)})$ we denote the median set and median value of $A \subset S_n^{\pm}$ by $M_n(A)$ and $m_n(A)$ respectively.

Let $(\Omega, \mathcal{F}, P)$ be a Borel probability space. We denote by σ the Cartesian product $S_1^{\pm} \times S_2^{\pm} \times ....$ Let $(a_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of positive real numbers diverging to infinity, and let k be an integer. For a sequence of random sets of finite points on signed symmetric groups we define the concept of asymptotic almost surely median as follows.

Definition 1. *For any $n \in \mathbb{N}$ and $i \in \{1, ..., k\}$, let x_{in} be a random element of $S_n^{\pm}$, i.e. $x_{in} : \Omega \to S_n^{\pm}$ is a Borel measurable function. Let $A_n = \{x_{1n}, ..., x_{kn}\} \subset S_n^{\pm}$, and let $A = (A_n)_{n\in\mathbb{N}}$. We say $y := (y_n)_{n\in\mathbb{N}} \in \sigma$ is a median of A asymptotically almost surely (a.a.s.) if*

1. *For any $i \in \{1, ..., k\}$ there exists a function $f_i : \mathbb{N} \to \mathbb{R}_+$ such that*
$$\frac{d^{(n)}(y_n, x_{in}) - f_i(n)}{a_n\sqrt{n}} \to 0 \text{ in probability} \tag{3.2.2}$$
and $\lim_{n\to\infty} \frac{f_i(n)}{n}$ exists.

2. $$\frac{d_T^{(n)}(y_n, A_n) - m_n(A_n)}{a_n\sqrt{n}} \to 0 \text{ in probability} \tag{3.2.3}$$

In this case we say any function $\bar{f} : \mathbb{N} \to \mathbb{R}_+$ which satisfies
$$\frac{\bar{f}(n) - m_n(A_n)}{a_n\sqrt{n}} \to 0 \text{ in probability} \tag{3.2.4}$$
is an asymptotic median value of A. This includes the function $\bar{f}(n) = \sum_{i=1}^{k} f_i(n)$.

In [37], Theorem 3, we proved that the sequence of identity elements $e := (e_n)_{n\in\mathbb{N}}$ is a median of k random signed permutations sampled from k independent reversal random walks at time cn where $c \le 1/4$, a.a.s.. By a small modification in the proof of that theorem we can extend that result as follows.

Theorem 3.2.1. *Let $0 < c_1 \leq ... \leq c_k$ such that $c_{k-1} + c_k \leq \frac{1}{2}$. For any natural number n, let $X^{1,n}, ..., X^{k,n}$ be k independent reversal random walks on $S_n^{\pm}$, all starting at e_n. Let $x_{in} = X^{i,n}(c_i n)$, where $X^{i,n}(t)$ stands for the position of random walk at time t. Then $e = (e_n)_{n\in\mathbb{N}}$ is a median of $A := (A_n)_{n\in\mathbb{N}}$ a.a.s., where $A_n = \{x_{1n}, ..., x_{kn}\}$, and the function $\theta : \mathbb{N} \to \mathbb{R}_+$ defined by*

$$\theta(n) = (\sum_{i=1}^{k} c_i)n \tag{3.2.5}$$

is an asymptotic median value of A

Proof: For all $i, j \in \{1, \ldots, k\}$ and for a median solution $y = (y_n)_{n\in\mathbb{N}}$ of A

$$d^{(n)}(x_{in}, x_{jn}) \leq d^{(n)}(y_n, x_{in}) + d^{(n)}(y_n, x_{jn}) \tag{3.2.6}$$

Therefore,

$$\sum_{i\neq j} d^{(n)}(x_{in}, x_{jn}) \leq \sum_{i\neq j}(d^{(n)}(y_n, x_{in}) + d^{(n)}(y_n, x_{jn})) \tag{3.2.7}$$

Thus

$$\sum_{i\neq j} d^{(n)}(x_{in}, x_{jn}) \leq (k-1)m_n(A_n) \leq (k-1)d_T^{(n)}(e_n, A_n) \tag{3.2.8}$$

By Theorem 2 in [37]

$$\frac{d^{(n)}(e_n, x_{in}) - c_i n}{a_n\sqrt{n}} \to 0 \tag{3.2.9}$$

in probability, and

$$\frac{d^{(n)}(x_{in}, x_{jn}) - (c_i + c_j)n}{a_n\sqrt{n}} \to 0 \tag{3.2.10}$$

in probability, for any $i \neq j \in \{1, ..., k\}$. Therefore

$$\frac{m_n(A_n) - \theta(n)}{a_n\sqrt{n}} \to 0 \tag{3.2.11}$$

in probability, and hence e is a median of A a.a.s. since

$$\frac{d_T^{(n)}(e_n, A_n) - \theta(n)}{a_n\sqrt{n}} \to 0 \text{ in probability} \tag{3.2.12}$$

This proves the theorem. ■

By definition of an a.a.s. median, it is clear that if there exists $y = (y_n)_{n\in\mathbb{N}}$ such that for any $i \in \{1, ..., k\}$, as $n \to \infty$,

$$\frac{d^{(n)}(y_n, x_{in}) - c_i n}{a_n\sqrt{n}} \to 0 \text{ in probability,} \tag{3.2.13}$$

then y is a median of A a.a.s.. Is the converse true? The next result shows that the converse is true under the hypotheses of Theorem 3.2.1.

Theorem 3.2.2. *Let $k \geq 3$ and c_i, x_{in}, and A_n be as defined in the statement of Theorem 3.2.1 for $i \in \{1, ..., k\}$ and natural number n. Then $y := (y_n)_{n\in\mathbb{N}}$ is a median of A a.a.s. if and only if (3.2.13) holds.*

Proof: By the explanation given before the statement of the theorem, it suffices to prove the necessary part. Let $y = (y_n)_{n\in\mathbb{N}}$ be a median of A a.a.s.. Then, by definition, there exist functions $f_i : \mathbb{N} \to \mathbb{R}_+$ for $i = 1, ..., n$ such that, as $n \to \infty$,

$$\frac{d^{(n)}(y_n, x_{in}) - f_i(n)}{a_n\sqrt{n}} \to 0 \text{ in probability,} \tag{3.2.14}$$

and

$$\frac{f_i(n)}{n} \to c'_i \tag{3.2.15}$$

for real numbers $c'_i \geq 0$. Letting $a_n := \sqrt{n}$ and applying Theorem 2 in [37], for $0 \leq i, j \leq k$, as $n \to \infty$, we have

$$\frac{f_i(n) + f_j(n) - d^{(n)}(y_n, x_{in}) - d^{(n)}(y_n, x_{jn})}{n} \to 0 \text{ in probability} \tag{3.2.16}$$

and

$$\frac{d^{(n)}(x_{in}, x_{jn}) - (c_i + c_j)n}{n} \to 0 \text{ in probability.} \tag{3.2.17}$$

Therefore, as $n \to \infty$,

$$\delta_n \to (c'_i + c'_j) - (c_i + c_j) \text{ in probability,} \tag{3.2.18}$$

where

$$\delta_n = \frac{d^{(n)}(x_{in}, y_n) + d^{(n)}(y_n, x_{jn}) - d^{(n)}(x_{in}, x_{jn})}{n}, \tag{3.2.19}$$

which is positive for any $n \in \mathbb{N}$. Hence

$$c_i' + c_j' \geq c_i + c_j \tag{3.2.20}$$

Now, suppose there exists $i \in \{1, ..., k\}$ such that $c_i' < c_i$. Then for any $j \neq i$ $c_j < c_j'$, and hence $\sum_{j=1}^{k} c_j < \sum_{j=1}^{k} c_j'$, since $k \geq 3$. This contradicts with the fact that y is a median of A a.a.s., as we know $e = (e_n)_{n\in\mathbb{N}}$ is a median of A a.a.s., and therefore the function θ defined by $\theta(n) = \sum_{i=1}^{k} c_i n$ is an asymptotic median value of A by Theorem 3.2.1. Hence $c_i \leq c_i'$. Therefore, if there exists $1 \leq i \leq k$ such that $c_i < c_i'$, then y can not be a median of A a.a.s., since, as $n \to \infty$,

$$\frac{\sum_{i=1}^{k} c_i' n - \theta(n)}{a_n \sqrt{n}} \to \infty, \tag{3.2.21}$$

which is a contradiction again. Thus $c_i = c_i'$ for any $1 \leq i \leq k$. ■

To simplify the results, we define an equivalence relationship on σ. We say $x = (x_n)_{n\in\mathbb{N}}$ and $y = (y_n)_{n\in\mathbb{N}}$ in σ are equivalent and we denote it by $x \cong y$ if and only if, as $n \to \infty$,

$$\frac{d^{(n)}(x_n, y_n)}{a_n \sqrt{n}} \to 0. \tag{3.2.22}$$

We denote the qoutient space of all equivalent classes of σ by $\bar{\sigma}$, that is $\bar{\sigma} := \sigma / \cong$. It is clear that if $x \cong y$ for two random elements in σ and if x is a median of A a.a.s., then so is y. Motivated by this, we say that $\alpha \in \bar{\sigma}$ is a median of A a.a.s. if any $x \in \alpha$ is a median of A a.a.s.. Let

$$M(A) := \{\alpha \in \bar{\sigma} : \alpha \text{ is a median of } A \text{ a.a.s.}\}. \tag{3.2.23}$$

Let f be a positive real-valued function on $\mathbb{N}$, and let $o = (o_n)_{n\in\mathbb{N}}$ be a random element of σ. Define the asymptotic sphere of radius f centred at o by

$$\partial B(o,f) := \\ \{\alpha \in \bar{\sigma} : \frac{d^{(n)}(o_n, x_n) - f(n)}{a_n\sqrt{n}} \to 0 \text{ in probability for any } (x_n)_{n\in\mathbb{N}} \in \alpha,\ as\ n \to \infty\}. \tag{3.2.24}$$

Let $x_i^* := (x_{in})_{n\in\mathbb{N}}$, and define the function $c_i^* : \mathbb{N} \to \mathbb{R}_+$ by $c_i^*(n) = c_i n$. The following is an immediate consequence of Theorem 3.2.2, which reduces the median search space.

Corollary 3.2.3. *Let $k \geq 3$ and $A = (A_n)_{n\in\mathbb{N}}$ be as defined in the statement of Theorem 3.2.1. Then*

$$M(A) = \cap_{i=1}^k \partial B(x_i^*, c_i^*) \tag{3.2.25}$$

Proof: Trivial from Theorem 3.2.2. ∎

3.3 Conclusion

We have explored the set of all possible positions of the medians of k random genomes sampled from k independent reversal random walks starting at the identity, namely, $X^{1,n}(c_i n), ..., X^{k,n}(c_k n)$, where $X^{1,n}, ..., X^{k,n}$ are independent reversal random walks on $S_n^{\pm}$ and $0 < c_1 \leq ... \leq c_k$ such that $c_{k-1} + c_k \leq \frac{1}{2}$. Doing this makes a major difference in the volume of the median problem search space. This is normally the whole of $S_n^{\pm}$, but it is now reduced it to a much smaller search space (Corollary 3.2.3). More specifically, the proportion of the volume of the new search space to the volume of $S_n^{\pm}$, i.e. $2^n n!$, converges to 0. Also, in the case of existence of several a.a.s. medians for these random genomes, the other medians are located far from the identity. Settling for a single median, then, can mislead us in the search for the position of the true ancestor. Further investigation is needed to study the number of the medians for an arbitrary value of k.

Chapter 4

Sets of medians in the non-geodesic pseudometric space of unsigned genomes with breakpoints [1]

The breakpoint median in the set S_n of permutations on n terms is known to have some unusual behavior, especially if the input genomes are maximally different from each other. The mathematical study of the set of medians is complicated by the facts that breakpoint distance is not a metric but a pseudo-metric, and that it does not define a geodesic space. We introduce the notion of partial geodesic, or geodesic patch (p-geodesic) between two permutations, and show that if two permutations are medians, then every permutation on a p-geodesic between them is also a median. We also prove the conjecture that the input permutations themselves are medians.

4.1 Introduction

Among the common measures of gene order difference between two genomes, the edit distances, such as reversal distance or double-cut-and-join distance, contrast with the breakpoint distance in that the former are defined in a geodesic space while the latter is not. Another characteristic of breakpoint distance that it does not share with most

[1]This chapter is basically the paper published in [36]

other genomic distances is that it is a pseudometric rather than a metric.

A problem in computational comparative genomics that has been extensively studied under many definitions of genomic distance is the gene order median problem [61], the archetypical instance of the gene order small phylogeny problem. The median genome is meant, in the first instance, to embody the information in common among $k \geq 3$ given genomes, and second, to estimate the ancestral genome of these k genomes. We have shown that the second goal becomes unattainable as $n \to \infty$, where n is the length of the genomes, if there are more than $0.5n$ mutational steps changing the gene order [37]. Moreover, we have conjectured, and demonstrated in simulation studies, that where there is little or nothing in common among the k input genomes, the median tends to reflect only one (actually, any one) of them, with no incorporation of information from the other $k - 1$ [29].

In the present chapter, we investigate this conjecture mathematically in the context of a wider study of medians for the breakpoint distance between unsigned linear unichromosomal genomes, although the methods and results are equally valid for genomes with signed and/or circular chromosomes, as well as those with $\chi > 1$ chromosomes, where χ is a fixed parameter. Our approach involves first a rigorous treatment of the pseudometric character of the breakpoint distance. Then, given the non-geodesic nature of the space we are able to define a weaker concept of p-geodesic, that we use later, given two or more medians, to locate further medians. We also prove the conjecture that for k genomes containing no gene order information among them, the normalized (divided by n) median score tends to $k - 1$.

4.2 Results

4.2.1 From pseudometric to metric

We denote by S_n the set of all permutations of length n. Each permutation represents a unichromosomal linear genome where the numbers all represent different genes. For a permutation $\pi := \pi_1 \ldots \pi_n$ we define the set of adjacencies of π to be all the unordered pairs $\{\pi_i, \pi_{i+1}\} = \{\pi_{i+1}, \pi_i\}$ for $i = 1, ..., n-1$. For $I \subseteq S_n$ we denote by $\mathcal{A}_I := \mathcal{A}_I^{(n)}$ the set of all common adjacencies of the elements of I. Then $\mathcal{A}_{S_n} = \emptyset$, and we also write $\mathcal{A}_\emptyset$ for the set of all pairs $\{i, j\}, i \neq j$. For any $I, J \subseteq S_n$ $\mathcal{A}_{I \cup J} = \mathcal{A}_I \cap \mathcal{A}_J$. It will sometimes be convenient to write $\mathcal{A}_I$, the set of common adjacencies in $I = \{x_1, ..., x_k\}$, as $\mathcal{A}_{x_1,...,x_k}$. For example $\mathcal{A}_{x,y,z}$ represents the set of adjacencies common to permutations x, y and z.

For $x, y \in S_n$ we define the breakpoint distance (bp distance) between x and y by

$$d^{(n)}(x, y) := n - 1 - |\mathcal{A}_{x,y}|. \quad (4.2.1)$$

This distance is not a metric on S_n, but rather a pseudometric because of non-reflexiveness: cases where $d^{(n)}(x, y) = 0$ but $x \neq y$, namely $x = \pi_1 \ldots \pi_n$ and $y = \pi_n \ldots \pi_1$, for any $x \in S_n$. In these cases, the permutations x and y are said to be equivalent, denoted by $x \sim y$. The equivalence class containing π is represented by $[\pi]$ and contains exactly two permutations, $\pi_1, ..., \pi_n$ and $\pi_n, ..., \pi_1$. The number of classes is thus $n!/2$. For any π, we denote the other element of $[\pi]$ by $\bar{\pi}$. The bp distance, a metric on the set of all equivalence classes of S_n, denoted by $\hat{S}_n := S_n/\sim$, is defined by

$$d^{(n)}([x], [y]) := d^{(n)}(x, y). \quad (4.2.2)$$

Where there is no risk of ambiguity, we can simplify the notation by using x and y instead of $[x]$ and $[y]$, and/or drop the superscript n.

It is clear that the maximum possible bp distance between two permutation classes is $n-1$ when they have no common adjacencies. Bp distance is symmetric on S_n and hence on $\hat{S}_n$. By construction, it is reflexive on $\hat{S}_n$. To verify the triangle inequality, consider three permutations x, y, z. We have

$$\mathcal{A}_{x,z} \supseteq \mathcal{A}_{x,y,z} = \mathcal{A}_{x,y} \cap \mathcal{A}_{y,z} \tag{4.2.3}$$

Therefore

$$d(x,z) = n - 1 - |\mathcal{A}_{x,z}| \leq n - 1 - |\mathcal{A}_{x,y}| - |\mathcal{A}_{y,z}| + |\mathcal{A}_{x,y} \cup \mathcal{A}_{y,z}|. \tag{4.2.4}$$

But $|\mathcal{A}_{x,y} \cup \mathcal{A}_{y,z}| = |\mathcal{A}_y \cap (\mathcal{A}_x \cup \mathcal{A}_z)| \leq n-1$ and hence the triangle inequality holds.

We say a pseudometric (or a metric) $\tilde{\rho}$ is left-invariant on a group G if for any $x, y, z \in G$, $\tilde{\rho}(x,y) = \tilde{\rho}(zx, zy)$. The bp distance is a left-invariant pseudometric on S_n.

Definition 2. *Given a set* $\{x_1, \ldots, x_k\} \subseteq S$ *and a pseudometric* ρ *on* S*, a median for the set is* $\mu \in S$ *such that* $\sum_{i=1}^{k} \rho(\mu, x_i)$ *is minimal.*

4.2.2 Defining the p-geodesic

A discrete metric space (S, ρ) is a geodesic space if for any two points $x, y \in S$ there exists a finite subset of S containing x, y that is isometric with the discrete line segment $[0, 1, ..., \rho(x,y)]$ ($\mathbb{N} \cup \{0\}$ is endowed with the standard metric $dist(m,n) := |m-n|$). Any subset of S with this property, and there may be several, is called a geodesic between x and y. For example, all connected graphs are geodesic spaces. In a geodesic space the medians of two points x and y consist of all the points located on geodesics between x and y.

What can we say when the space is not a geodesic space? To answer this, we extend the concept of geodesic by introducing the concept of a p-geodesic. A p-geodesic between x and y is a maximal subset of S containing x, y which is isometric to a subsegment (not necessarily contiguous) of the line segment $[0, 1, ..., \rho(x,y)]$. For any two points x and y in an arbitrary metric space (S, ρ) there exists at least one p-geodesic between them because x, y is isometric to $\{0, \rho(x, y)\}$. In addition, any geodesic is a p-geodesic. A non-trivial p-geodesic is the one that contains at least three points. Any point z on a p-geodesic between x, y satisfies:

$$\rho(x, y) = \rho(x, z) + \rho(z, y). \tag{4.2.5}$$

Therefore all the medians of two points x and y must lie on a p-geodesic between them. We denote the set of all permutations lying on p-geodesics connecting $x, y \in S_n$ by $\overline{[x, y]}$.

$(\hat{S}_n, d)$ is not a geodesic space. For example there is no geodesic connecting the identity permutation id and $\pi := 1\ 2\ x_1\ x_2\ ...\ x_{n-4}\ n-1\ n$ when $x_1\ x_2\ ...\ x_{n-4}$ is a non-identical permutation on $\{3, ..., n-2\}$. The smallest change to id is to cut one of its adjacencies, say $\{i, i+1\}$, and rejoin the two segments in one of the three possible ways: 1 to n, 1 to $i+1$ or n to i. Now if we cut the adjacencies $\{1, 2\}$ or $\{n-1, n\}$ in id, the distance of the new permutation to both id and π increases. If, on the other hand, we cut one of the other adjacencies in id all the ways of rejoining, which increase the distance to id, either increase or leave unchanged the distance to π, since $\{1, n\}$, $\{1, i+1\}$ and $\{n, i\}$ are not adjacencies in $\mathcal{A}_\pi$. Therefore there is no geodesic connecting id to π.

Although $\hat{S}_n$ is not a geodesic space, there may still exist permutations with a geodesic between them. For example

$$\{id = 123456, 213456, 312456, 421356, 531246, \pi = 135246\} \tag{4.2.6}$$

is a geodesic between id and π. Note $d(id, \pi) = 5$, the maximum possible distance in $\hat{S}_6$.

4.2.3 The median value and medians of permutations with maximum pairwise distances

In this section we investigate the bp median problem in the case of k permutations with maximum pairwise distances. As we shall see later, this situation is very similar to the case of k uniformly random permutations. Let (S, ρ) be a pseudometric space. The total distance of a point $x \in S$ to a finite subset $\emptyset \neq B \subseteq S$ is defined to be

$$\rho(x, B) := \sum_{y \in B} \rho(x, y). \tag{4.2.7}$$

The median value of B, $m^{S,\rho}(B)$, is the infimum of the total distance when the infimum is over all the points $x \in S$, that is

$$m^{S,\rho}(B) := \inf_{x \in S} \rho(x, B). \tag{4.2.8}$$

We can extend this definition to sets with multiplicities. Let $\emptyset \neq B \subseteq S$. We define a multiplicity function n_B from B to $\mathbb{N}$ and write $n_B(x) = n_x$. We call $A = (B, n_B)$ a set with multiplicities. We define the total distance of a point $x \in S$ to A to be

$$\rho(x, A) := \sum_{y \in B} n_y \rho(x, y). \tag{4.2.9}$$

The definition of the median value in Equation (4.2.8) can be extended in an analogous way to the median value of a set with multiplicity A. When S is finite, then the total distance function takes its minimum on S and "inf" turns into "min" in the above formulation. The points of the space S that minimize the total distance to A are called the median points or medians of $A$and the set of all, these medians is called the median set of A, denoted by $M^{S,\rho}(A)$.

Let B and $A = (B, n_B)$ be a subset and a subset with multiplicities of S_n. We define $[B]$ to be the set of all permutation classes of S_n that have at least one of their permutations in B. That is

$$[B] = \{[x] \in \hat{S}_n \text{ such that } \exists \text{y} \in \text{B with x} \sim \text{y}\}. \tag{4.2.10}$$

Two nonempty subsets $B, B' \subseteq S_n$ are said to be equivalent, denoted by $B \sim B'$, if $[B] = [B']$. Also we define $[n_B]$ to be a function from $[B]$ to $\mathbb{N}$ with

$$[n_B]([x]) = n_{[x]} := \sum_{x \sim y \in B} n_y. \tag{4.2.11}$$

Then, for $A = (B, n_B)$, the definition of $[A]$ is straightforward:

$$[A] := ([B], [n_B]), \tag{4.2.12}$$

and we say two nonempty subsets of S_n with multiplicities, namely A and A' are equivalent, denoted by $A \sim A'$, if $[A] = [A']$. In fact $[A]$ is the equivalence class containing A. We call $[A]$ a subset of $\hat{S}_n$ with multiplicities. We use the notations "[ ]" and " $\sim$ " for all the above concepts without restriction.

With these definitions we can readily verify that in the context of bp distance, for $A \sim A'$ and $x \sim x'$, we have

$$d(x, A) = d(x', A') = d([x], [A]). \tag{4.2.13}$$

Recall that we use d as both a metric on $\hat{S}_n$ and a pseudometric on S_n. Therefore we can conclude that

$$m^{S_n,d}(A) = m^{S_n,d}(A') = m^{\hat{S}_n,d}([A]) \tag{4.2.14}$$

and similarly

$$[M^{S_n,d}(A)] = [M^{S_n,d}(A')] = M^{\hat{S}_n,d}([A]). \tag{4.2.15}$$

Henceforward, we will simplify by replacing the notation $m^{S_n,d}(A)$ and $M^{S_n,d}(A)$ by $m_n(A)$ and $M_n(A)$, respectively. Also for a subset $[A]$ of $\hat{S}_n$ with multiplicities, we will use the notation $m_n([A])$ and $M_n([A])$ instead of $m^{\hat{S}_n,d}([A])$ and $M^{\hat{S}_n,d}([A])$ respectively. Where there is no ambiguity, we will suppress the subscript n.

Proposition 4.2.1. *Suppose $X := \{x_1, ..., x_k\} \subset \hat{S}_n$ such that $d(x_i, x_j) = n - 1$ for any $i \neq j$. Then the bp median value of X is $(k-1)(n-1)$. Moreover, m^* is a median of X, $m^* \in M(X)$, if and only if $\mathcal{A}_{m^*} \subset \cup_{i=1}^{k} \mathcal{A}_{x_i}$.*

Proof: Let $\pi \in \hat{S}_n$ be an arbitrary permutation class. Since $\mathcal{A}_{\pi,x_i} \subset \mathcal{A}_{x_i}$ and $\mathcal{A}_{\pi,x_j} \subset \mathcal{A}_{x_j}$ for any $1 \leq i,j \leq k$, we have $\mathcal{A}_{\pi,x_i} \cap \mathcal{A}_{\pi,x_j} = \emptyset$. Also

$$\cup_{i=1}^{k} \mathcal{A}_{\pi,x_i} \subseteq \mathcal{A}_{\pi}. \tag{4.2.16}$$

Therefore

$$\sum_{i=1}^{k} |\mathcal{A}_{\pi,x_i}| \leq |\mathcal{A}_{\pi}| = n-1. \tag{4.2.17}$$

Hence

$$\sum_{i=1}^{k} d(\pi, x_i) \geq (k-1)(n-1). \tag{4.2.18}$$

The equality holds by letting $\pi = x_i$ for any $1 \leq i \leq k$. This proves the first part of the proposition. For the second part, we know that $m^* \in M(X)$ is equivalent with the fact that the total distance of m^* to X is $(k-1)(n-1)$, and this is equivalent to $\sum_{i=1}^{k} |\mathcal{A}_{m^*,x_i}| = n-1$ and $\cup_{i=1}^{k} \mathcal{A}_{m^*,x_i} = \mathcal{A}_{m^*}$. But $\cup_{i=1}^{k} \mathcal{A}_{m^*,x_i}$ can be written as $\mathcal{A}_{m^*} \cap (\cup_{i=1}^{k} \mathcal{A}_{x_i})$. This completes the proof of the equivalence relation in the proposition. ■

Lemma 4.2.2. *Let x, y, z be three permutation classes in $\hat{S}_n$ that are pairwise at a maximum distance $n-1$ from each other. Then for any $w \in \overline{[x,y]}$ we have $d(w,z) = n-1$.*

Proof: Having $w \in \overline{[x,y]}$, we have $\mathcal{A}_w \subset \mathcal{A}_x \cup \mathcal{A}_y$. Also, we know that $\mathcal{A}_z \cap (\mathcal{A}_x \cup \mathcal{A}_y) = \emptyset$. This concudes the result. ■

The above lemma simply indicates that for any two points x_i, x_j in the set X in the proposition above $\overline{[x_i, x_j]} \subset M(X)$, since the total distance of each point in $\overline{[x_i, x_j]}$ to X is $(k-1)(n-1)$.

Corollary 4.2.3. *Suppose* $X := \{x_1, ..., x_k\} \subset \hat{S}_n$ *such that* $d(x_i, x_j) = n - 1$ *for any* $i \neq j$*. Then* $\cup_{i,j}\overline{[x_i, x_j]} \subset M(X)$*.*

What can we say more about the median positions? The notion of "accessibility" will help us to keep track of some other medians of the set X that are not in $\cup_{i,j}\overline{[x_i, x_j]}$. Before defining this concept, we first need more information about the properties of $\overline{[x, y]}$ for $x, y \in \hat{S}_n$.

Lemma 4.2.4. *Let* $x, y \in \hat{S}_n$*. Then* $z \in \overline{[x, y]}$ *if and only if* $\mathcal{A}_{x,y} \subset \mathcal{A}_z \subset \mathcal{A}_x \cup \mathcal{A}_y$*.*

Proof: We know $z \in \overline{[x, y]}$ if and only if $d(x, z) + d(z, y) = d(x, y)$. On the other hand, we can write $\mathcal{A}_z$ as follows

$$\mathcal{A}_z = \mathcal{A}_{z,x,y} \cup (\mathcal{A}_{z,x} \setminus \mathcal{A}_y) \cup (\mathcal{A}_{z,y} \setminus \mathcal{A}_x) \cup (\mathcal{A}_z \setminus (\mathcal{A}_x \cup \mathcal{A}_y)), \tag{4.2.19}$$

where the pairwise intersection of the sets in the right hand side is empty. We can also write

$$d(x, z) = (n - 1) - |\mathcal{A}_{z,x,y}| - |\mathcal{A}_{z,x} \setminus \mathcal{A}_y| \tag{4.2.20}$$

and

$$d(z, y) = (n - 1) - |\mathcal{A}_{z,x,y}| - |\mathcal{A}_{z,y} \setminus \mathcal{A}_x|. \tag{4.2.21}$$

Furthermore

$$d(x, y) \leq (n - 1) - |\mathcal{A}_{z,x,y}| \tag{4.2.22}$$

and

$$(n - 1) - |\mathcal{A}_{z,x,y}| - |\mathcal{A}_{z,x} \setminus \mathcal{A}_y| - |\mathcal{A}_{z,y} \setminus \mathcal{A}_x| = |\mathcal{A}_z \setminus (\mathcal{A}_x \cup \mathcal{A}_y)|. \tag{4.2.23}$$

Now for "sufficiency", from $d(x, z) + d(z, y) = d(x, y)$, we have

$$(n - 1) - |\mathcal{A}_{z,x,y}| - |\mathcal{A}_{z,x} \setminus \mathcal{A}_y| + (n - 1) - |\mathcal{A}_{z,x,y}| - |\mathcal{A}_{z,y} \setminus \mathcal{A}_x| \tag{4.2.24}$$

$$= (n - 1) - |\mathcal{A}_{x,y}| \leq (n - 1) - |\mathcal{A}_{x,y,z}| \tag{4.2.25}$$

Therefore by Equation (4.2.23) we have

$$(n-1) - |\mathcal{A}_{z,x,y}| - |\mathcal{A}_{z,x} \setminus \mathcal{A}_y| - |\mathcal{A}_{z,y} \setminus \mathcal{A}_x| = |\mathcal{A}_z \setminus (\mathcal{A}_x \cup \mathcal{A}_y)| \leq 0 \qquad (4.2.26)$$

This results in $|\mathcal{A}_{x,y}| = |\mathcal{A}_{x,y,z}|$ and hence in $\mathcal{A}_{x,y} \subset \mathcal{A}_z$. Otherwise the inequality in (4.2.26) will be strict, which is impossible. On the other hand the inequality in (4.2.26) shows $\mathcal{A}_z \setminus (\mathcal{A}_x \cup \mathcal{A}_y) = \emptyset$ which results in $\mathcal{A}_z \subset \mathcal{A}_x \cup \mathcal{A}_y$.

For "necessity", we have

$$(n-1) - |\mathcal{A}_{z,x,y}| - |\mathcal{A}_{z,x} \setminus \mathcal{A}_y| - |\mathcal{A}_{z,y} \setminus \mathcal{A}_x| + (n-1) - |\mathcal{A}_{x,y}| = (n-1) - |\mathcal{A}_{x,y}|. \qquad (4.2.27)$$

This is true because of $\mathcal{A}_z \subset \mathcal{A}_x \cup \mathcal{A}_y$ and Equation (4.2.23). But since $\mathcal{A}_{x,y} \subset \mathcal{A}_z \subset \mathcal{A}_x \cup \mathcal{A}_y$, we have $|\mathcal{A}_{x,y}| = |\mathcal{A}_{x,y,z}|$ and we can replace $|\mathcal{A}_{x,y}|$ by $|\mathcal{A}_{x,y,z}|$ in the left hand side of the last equality. Hence, $d(x,z) + d(z,y) = d(x,y)$. This completes the "necessity" part of the proof. ■

Definition 3. *Let $X := \{x_1, ..., x_k\}$ be a subset of $\hat{S}_n$. We say a permutation class $z \in \hat{S}_n$ is 1-accessible from X if there exists an $m \in \mathbb{N}$, a finite sequence $y_1, ..., y_m$, where $y_i \in X$, and $z_1, ..., z_m$, where $z_i \in \hat{S}_n$ such that $z_1 = y_1$, $z_m = z$ and $z_{i+1} \in \overline{[z_i, y_{i+1}]}$ for $i = 1...m-1$. See Fig. 4.1.*

We denote the set of all 1-accessible points of X by $Z(X)$. We define $Z_0(X) := X$. Also, for $r \in \mathbb{N} \cup \{0\}$, by induction, we define $Z_{r+1}(X)$ to be $Z(Z_r(X))$ and we call it the set of all r-accessible permutation classes. That is to say, $Z_1(X) = Z(X)$, $Z_2(X) = Z(Z(X))$ and so on. It is clear that $Z_{r+1}(X)$ includes $Z_r(X)$ and also $\cup_{x,y \in Z_r(X)} \overline{[x,y]}$. A permutation class z is said to be accessible from X if there exists $r \in \mathbb{N}$ such that $z \in Z_r(X)$. We denote the set of all accessible points by $\bar{Z}(X) = \cup_{r \in \mathbb{N} \cup \{0\}} Z_r(X)$.

Note that $Z(\bar{Z}(X)) = \bar{Z}(X)$. This holds because for any 1-accessible permutation class z from $\bar{Z}(X)$, there must exist $m \in \mathbb{N}$, $r_0 \in \mathbb{N} \cup \{0\}$, $y_1, ..., y_m \in \bar{Z}_{r_0}(X)$ (the

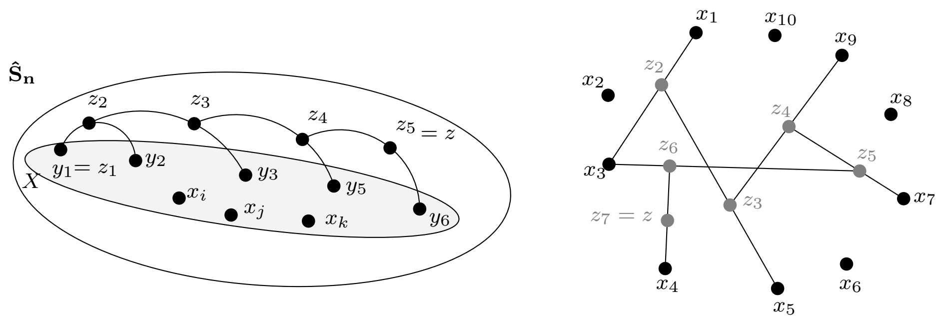

Figure 4.1: **Accessibility.** Illustration of how $\bar{Z}$ is constructed.

y_i's must be in $\bar{Z}(X)$, thus there must be such an r_0) and $z_1, ..., z_m$ where $z_i \in \hat{S}_n$ such that $z_1 = y_1$, $z_m = z$ and $z_{i+1} \in \overline{[z_i, y_{i+1}]}$. Therefore $z \in Z_{r_0+1}(X) \subset \bar{Z}(X)$. We can then conclude that $\bar{Z}(\bar{Z}(X)) = \bar{Z}(X)$.

Proposition 4.2.5. *Suppose $X := \{x_1, ..., x_k\} \subset \hat{S}_n$ such that $d(x_i, x_j) = n - 1$ for any $i \neq j$. Then for any permutation class $z \in \bar{Z}(X)$, the total distance $d(z, X)$ between z and X is $(k-1)(n-1)$, and hence $\bar{Z}(X) \subset M(X)$. Furthermore, if $m_1, m_2 \in M(X)$, then $\overline{[m_1, m_2]} \subset M(X)$.*

Proof: Suppose $m_1, m_2 \in M(X)$ and $m^* \in \overline{[m_1, m_2]}$. By Lemma 4.2.4 and Proposition 4.2.1 we have $\mathcal{A}_{m^*} \subset \mathcal{A}_{m_1} \cup \mathcal{A}_{m_2} \subset \cup_{i=1}^{k} \mathcal{A}_{x_i}$. Applying Proposition 4.2.1 again, we have $m^* \in M(X)$. Now it suffices to show that for any $r \in \mathbb{N} \cup \{0\}$, $Z_r(X) \subset M(X)$. We prove this by induction. For $r = 0$ this follows from Corollary 4.2.3. Suppose $Z_r(X) \subset M(X)$. By definition, we have $Z_{r+1}(X) = Z(Z_r(X))$. That is for $z \in Z_{r+1}(X)$ there exists an $m \in \mathbb{N}$,$y_1, ..., y_m \in Z_r(X)$ and $z_1, ..., z_m$, where $z_i \in \hat{S}_n$, such that $z_1 = y_1$, $z_m = z$ and $z_{i+1} \in \overline{[z_i, y_{i+1}]}$. $z_1 \in \overline{[y_1, y_2]}$, and by the fact we proved above $z_1 \in M(X)$ since $y_1, ..., y_m \in Z_r(X) \subset M(X)$. Continuing this, we conclude that $z_1, z_2, ..., z_m = z \in M(X)$. Hence $Z_{r+1}(X) \subset M(X)$. This completes the proof. ∎

Question: Is every median point of X accessible from X? Do we have $M(X) =$

$\bar{Z}(X)$?

4.2.4 The median value and medians of k random permutations

In this section we study the median value and median points of k independent random permutation classes uniformly chosen from $\hat{S}_n$. This is equivalent to studying the same problem for k random permutations sampled from S_n. All the results of this section carry over to permutations without any problem.

We make use of the fact that the bp distance of two independent random permutations tends to be close to its maximum value, $n-1$. Xu et al. [64] showed that if we fix a reference linear permutation id and pick a random permutation x uniformly, the expected number and variance of $|\mathcal{A}_{id,x}^{(n)}|$ both are very close to 2 for large enough n. Because of the symmetry of the group S_n and the fact that bp distance is a left-invariant pseudometric, the same results hold for two random permutations x and y. We first summarize the results we need from [64].

Suppose $\tilde{\nu}_n$ is the uniform measure on S_n. Let $\Pi : S_n \to \hat{S}_n$ be the natural surjective map sending each permutation onto its corresponding permutation class. Define

$$\nu_n := \Pi * \tilde{\nu}_n \tag{4.2.28}$$

to be the push-forward measure of $\tilde{\nu}_n$ induced by the map Π. It is clear that ν_n is the uniform measure on $\hat{S}_n$. The following proposition is a reformulation of Theorems 6 and 7 in [64].

Proposition 4.2.6. *[Xu-Alain-Sankoff] Let x and y be two independent random permutation classes (irpc) chosen uniformly from $\hat{S}_n$. Then*

$$\mathrm{E}[d(x,y)] = n - 3 - \frac{2}{n} - o(\frac{2}{n}), \tag{4.2.29}$$

$$\mathrm{Var}[d(x,y)] = 2 - \frac{2}{n} - o(\frac{2}{n}). \tag{4.2.30}$$

Define the error function for the distance of x, y by

$$\varepsilon_n(x,y) := (n-1) - d(x,y) = |\mathcal{A}_{x,y}|. \tag{4.2.31}$$

Corollary 4.2.7. *Suppose x and y are two irpc's sampled from the uniform measure ν_n and a_n is an arbitrary sequence of real numbers diverging to $+\infty$. Then $\frac{\varepsilon_n(x,y)}{a_n}$ converges to zero asymptotically ν_n^{*2}-almost surely (a.a.s.), that is , as $n \to \infty$,*

$$\frac{\varepsilon_n(x,y)}{a_n} \to 0 \text{ in probability.} \tag{4.2.32}$$

Proof: The proof is straightforward from [64] and Chebyshev's inequality. ■

Now we are ready to study the median value of k *irpc*'s. Let $[A]$ be a subset of $\hat{S}_n$ with multiplicities and with k elements. Define

$$e_n([A]) := (k-1)(n-1) - m_n([A]). \tag{4.2.33}$$

Theorem 4.2.8. *Let $X^{(n)} := \{x_1^{(n)}, x_2^{(n)}, ..., x_k^{(n)}\}$ be a set of k irpc in $\hat{S}_n$ sampled from the measure ν_n^{*k}. Then their breakpoint median value $m_n^* := m_n(X^{(n)})$ tends to be close to its maximum after a convenient rescaling with high probability, that is, for any arbitrary sequence $a_n \to \infty$ as $n \to \infty$, $\frac{e_n^*}{a_n} \to 0$ in ν_n^{*k}-probability where $e_n^* := e_n(X^{(n)})$.*

Proof: Let π be an arbitrary point of S_n. Let $\mathcal{A}_{\pi\setminus X} = \mathcal{A}_\pi \setminus \mathcal{A}_X$. We have

$$\sum_{i=1}^{k} |\mathcal{A}_{\pi,x_i}| \leq |\mathcal{A}_{\pi\setminus X}| + \sum_{i=1}^{k} |\mathcal{A}_{\pi,x_i}| \leq (n-1) + \binom{k}{2}\alpha_n, \tag{4.2.34}$$

where α_n is $\max_{i,j} \varepsilon_n(x_i, x_j)$. On the other hand $m_n(X^{(n)}) \leq (k-1)(n-1)$. The reason is the same as has already been discussed in the proof of Proposition 4.2.1.

Therefore, subtracting $(k-1)(n-1)$, we have

$$0 \leq e_n^* \leq \binom{k}{2}\alpha_n. \tag{4.2.35}$$

Dividing by a_n and letting n go to ∞, the result follows from the last corollary.

■

Theorem 4.2.9. *Let $X^{(n)} := \{x_1^{(n)}, x_2^{(n)}, ..., x_k^{(n)}\}$ be a set of k irpc's in $\hat{S}_n$ sampled from the measure ν_n^{*k}. Then for any permutation class $z^{(n)} \in \bar{Z}(X^{(n)})$, the total distance of $z^{(n)}$ to X is close to $(k-1)(n-1)$ with high probability after a convenient rescaling. More explicitly, for any arbitrary sequence of real numbers a_n diverging to ∞*

$$\frac{(k-1)(n-1) - d^{(n)}(z^{(n)}, X^{(n)})}{a_n} \to 0 \text{ in } \nu_n^{*k}-\text{probability}, \tag{4.2.36}$$

as $n \to \infty$. Therefore, as $n \to \infty$,

$$\frac{d^{(n)}(z^{(n)}, X^{(n)}) - m_n(X^{(n)})}{a_n} \to 0 \text{ in } \nu_n^{*k}-\text{probability}. \tag{4.2.37}$$

Furthermore, if $m_1^{(n)}, m_2^{(n)} \in M_n(X^{(n)})$ then for any $\tilde{m}^{(n)} \in \overline{[m_1^{(n)}, m_2^{(n)}]}$, as $n \to \infty$,

$$\frac{d^{(n)}(\tilde{m}^{(n)}, X^{(n)}) - m_n(X^{(n)})}{a_n} \to 0 \text{ in } \nu_n^{*k}-\text{probability}. \tag{4.2.38}$$

Proof: The structure of the proof is similar to the proof of Proposition 4.2.1. Suppose $o \in \hat{S}_n$ with $\mathcal{A}_o \subset_{i=1}^k \cup \mathcal{A}_{x_i}$. Let α_n be as defined in the proof of Theorem 4.2.8. Then by the same discussion we have

$$n-1 \leq \sum_{i=1}^{k} |\mathcal{A}_{o,x_i}| \leq n-1+\binom{k}{2}\alpha_n. \tag{4.2.39}$$

Therefore,

$$(k-1)(n-1) \geq d(o,X) \geq (k-1)(n-1) - \binom{k}{2}\alpha_n \tag{4.2.40}$$

and, as $n \to \infty$,

$$\frac{(k-1)(n-1) - d(o, X)}{a_n} \to 0 \;\; \textit{in probability.} \tag{4.2.41}$$

From Theorem 4.2.8, as $n \to \infty$, we have

$$\frac{(k-1)(n-1) - m_n(X)}{a_n} \to 0 \;\; \textit{in probability.} \tag{4.2.42}$$

Hence, as $n \to \infty$,

$$\frac{d(o, X^{(n)}) - m_n(X)}{a_n} \to 0 \;\; \textit{in probability.} \tag{4.2.43}$$

It suffices to show that $z := Z^{(n)} \in \bar{Z}(X)$ has the same property, that is $\mathcal{A}_z \in \cup_{i=1}^k \mathcal{A}_{x_i}$. But this is clear by induction. For the second part of the theorem, let $m^*_{1,n}, m^*_{2,n} \in M(X)$. Suppose $m^* \in \overline{[m^*_{1,n}, m^*_{2,n}]}$. By Theorem 4.2.8, as $n \to \infty$, $\frac{|\mathcal{A}_{m^*_{i,n} \setminus X}|}{a_n} \to 0$ in probability for $i = 1, 2$. On the other hand we have $\mathcal{A}_{m^* \setminus X} \subset \mathcal{A}_{m^*_{1,n} \setminus X} \cup \mathcal{A}_{m^*_{2,n} \setminus X}$. Therefore, as $n \to \infty$,

$$\frac{|\mathcal{A}_{m^* \setminus X}|}{a_n} \to 0 \;\; \textit{in probability.} \tag{4.2.44}$$

Therefore

$$(k-1)(n-1) \leq d(m^*, X) \leq (k-1)(n-1) + \binom{k}{2} \alpha_n \tag{4.2.45}$$

since, as $n \to \infty$,

$$\frac{|\mathcal{A}_{m^*, x_i} \cap \mathcal{A}_{m^*, x_j}|}{a_n} \to 0 \;\; \textit{in probability.} \tag{4.2.46}$$

The statement follows from the last inequality. ■

4.3 Conclusions

We have shown that the median value for a set of random permutations tends to be close to its extreme value with high probability. Also it has been shown that every permutation accessible from a set of random permutations can be considered as a median of that set asymptotically almost surely, and conjectured that the converse is true, namely that every median is accessible from the original set in this way.

Further work is needed to characterize the existence and size of non-trivial p-geodesics, in order to assess how extensive the set of medians is.

Chapter 5

Fleming-Viot process in random environment: quenched martingale approach

5.1 Introduction

In this chapter we study the Fleming-Viot (FV) processes in random environments. By random environment we mean that the fitness function is a stochastic process. We make use of the Stroock-Varadhan's characterization of Markov processes, namely, the martingale problem approach.

In the introduction we constructed the N-particle Moran process with resampling, mutation and selection, where $N \in \mathbb{N}$, and from that we derived the empirical measure-valued Moran process. When the number of individuals diverges to infinity, we can construct a countable particle system as a limit of the Moran process. In fact, this system is the $\mathbb{N}$-particle-valued system counterpart of the Fleming-Viot process. It was first introduced by Dawson and Hochberg [11] and later, considering an order for the particles, the countable particle system was improved by Donnelly and Kurtz in a series of papers [14, 15, 16]. The latter is called the look-down process. As we promised, in this chapter, we concentrate on a method called martingale characterization and duality method which has several advantages and applications in the literature. Our goal is to set up the martingale and duality method for measure-valued Moran and Fleming-Viot processes in random environments. We study the convergence and ergodic theorems for these processes. We organize the chapter as follows. In the rest of the first section, after introducing some general notations, we set up the time-inhomogeneous martingale problems and bring some criteria for existence and uniqueness of solutions. In subsection 1.3 we introduce the notion of operator-valued stochastic processes and generalize the time-inhomogeneous martingale problem to quenched martingale problems in order to characterize Markov processes in random environments as their solutions. Subsection 1.4 defines the joint annealed-environment process, where we consider the evolution of the annealed process together with its associated environment. Section 2 is devoted to martingale characterization of Moran and Fleming-Viot processes in random environments (r.e.). The statement of the main theorems will be given in this section as well. Section 3 develops the duality method in the case of general time-inhomogeneous and quenched martingale problems. Section 4 presents a function-valued dual for the Fleming-Viot process in random environment and studies its long-time behaviour. In section 5, we prove the convergence of infinitesimal generators of Moran process in random environments to that of the Fleming-Viot process in random environment. The proof of the well-posedness of the quenched Fleming-Viot martingale problem, along with the convergence of the Moran process in r.e. to Fleming-Viot process in r.e., will come in section 6. Section 7 is devoted to the proof of continuity of sample paths of the Fleming-Viot process in r.e. Finally, in section 8, we prove the ergodic theorem for the Fleming-Viot process in random environment.

5.1.1 General notations

Let (S, d_S) and $(S', d_{S'})$ be two metric spaces, and let $\mathcal{C}(S, S') = \mathcal{C}^0(S, S')$ and $\mathcal{C}^k(S, S')$ (for $k \geq 1$) be the space of all continuous, and k times continuously differentiable (Borel measurable) functions form S to S', respectively. In particular, when S' is the set of real numbers with the standard topology, we replace $\mathcal{C}(S, S')$ and $\mathcal{C}^k(S, S')$ by $\mathcal{C}(S) = \mathcal{C}^0(S)$ and $\mathcal{C}^k(S)$, respectively. Let $\mathcal{B}(S)$, $\bar{\mathcal{C}}(S) = \bar{\mathcal{C}}^0(S)$, and $\bar{\mathcal{C}}^k(S)$ (for $k \geq 1$) be the space of all bounded, bounded continuous, and bounded k times continuously differentiable Borel measurable real-valued functions on S, respectively, with norm $\|f\|_\infty = \|f\|_\infty^S := \sup_{x \in S} |f(x)|$. The topology induced by this norm is called the sup-norm topology. More explicitly, we define the sup-norm metric between two real-valued bounded Borel measurable (continuous, k times continuously differentiable) functions on S by $\|f - g\|_\infty$. We equip the space of all S-valued càdlàg functions, namely the space of all right continuous with left limits S-valued functions defined on $\mathbb{R}_+$, with Skorokhod topology, and denote it by $D_S[0, \infty)$. We denote by $\mathscr{B}(S)$ both the Borel σ-field and the space of all Borel measurable real-valued functions on S. Denote by $\mathcal{P}(S)$ the space of all (Borel) probability measures on S, equipped with the weak topology, and let "$\Rightarrow$" denote convergence in distribution. Also for $S_n \subset S$, for natural numbers n, say a sequence of S_n-valued random variables, namely $(Z_n)_{n \in \mathbb{N}}$, converges weakly to an S-valued random variable Z, if $\iota_n(Z_n) \Rightarrow Z$ as $n \to \infty$, where $\iota_n : S_n \to S$ is the natural embedding map. In general, for two topological spaces S_1 and S_2, by $S_1 \times S_2$ we mean the Cartesian product of two spaces equipped with the product topology, and by $\mathcal{P}(S_1 \times S_2)$ we mean the space of all Borel probability measures on $S_1 \times S_2$. Otherwise, we shall indicate it if we furnish the product space with another topology. Also, denote by $< m, f >$ or $< f, m >$ the integration $\int_S f dm$ for $m \in \mathcal{P}(S)$ and $f \in \mathcal{B}(S)$ (or more generally, when $m \in \mathcal{P}(S)$ is fixed, for all m-integrable functions.)

Throughout this chapter, S is a general Polish space, i.e. a separable completely metrizable topological space, with at least two elements (to avoid triviality), and we assume $(\Omega, \mathbb{P}, \mathcal{F})$ is a probability space, and all random variables and stochastic processes will be defined on this space. Also, we restrict random variables and stochastic processes to take values only on Polish spaces. We denote by $\mathbb{P}\zeta^{-1}$ the law of an S-valued random variable ζ (similarly, a measurable stochastic process $\zeta = (\zeta_t)_{t\geq 0}$). Let $m = \mathbb{P}\zeta^{-1}$. For an m-integrable real-valued function f on S, the expected value of $f(\zeta)$ is denoted by $\mathbb{E}[f(\zeta)]$, or to emphasise the law of ζ, by $\mathbb{E}_m[f(\zeta)]$. Also, by $\mathbb{E}^x[f(\zeta)]$ ($\mathbb{E}^{p_0}[f(\zeta)]$, respectively), we put emphasis on the initial state $x \in S$ (initial distribution $p_0 \in \mathcal{P}(S)$, respectively) of the process ζ.

5.1.2 Time-inhomogeneous martingale problem: existence and uniqueness

Let L be a Banach space. We can think of an operator G on L as a subset of $L \times L$. This definition allows G to be a multi-valued operator. A linear operator is one that is a linear subspace of $L \times L$. Observe that a linear operator G is single-valued, if the condition $(0, g) \in G$ implies $g = 0$. For a single-valued linear operator G, the domain of G, denoted by $\mathcal{D}(G)$, is the set of elements of L on which G is defined. Other words, $\mathcal{D}(G) = \{f \in L : (f, g) \in G\}$. Also, the range of an operator G is denoted by $\mathcal{R}(G) = \{g \in L : (f, g) \in G\}$. Let $\mathcal{D}$ be a linear subspace of L. A time-dependent single-valued linear operator G is a mapping from $\mathbb{R}_+ \times \mathcal{D}$ to L such that $G(t, .) : \mathcal{D} \to L$ is a single-valued linear operator. For simplicity, we set $G_t(.) = G(t, .)$ for any $t \geq 0$. In this chapter we only deal with time-dependent linear operators for which the domain of G_t is $\mathcal{D}$ for any $t \geq 0$. Therefore we define the domain of G to be $\mathcal{D}(G) = \mathcal{D}$.

As discussed above, we can identify the (time-dependent, respectively) generator of an S-valued (inhomogeneous, respectively) Markov process with the domain $\mathcal{D} \subset \mathcal{B}(S)$ as a subset of $\mathcal{B}(S) \times \mathcal{B}(S)$ (not necessarily linear). However for our purposes in this chapter we assume that all the operators are linear and single-valued, therefore their domains are linear subspaces of $\mathcal{B}(S)$. In the sequel we consider the generators of Markov processes as both single-valued linear operators and also linear subspaces of $\mathcal{B}(S) \times \mathcal{B}(S)$ for which containing $(0, g)$ implies $g = 0$. Similarly, for a time-dependent infinitesimal generator of a time-inhomogeneous Markov process, $G : \mathbb{R}_+ \times \mathcal{D} \to \mathcal{B}(S)$, for any $t \geq 0$, we assume that G_t is a single-valued linear operator and also a linear subspace of $\mathcal{B}(S) \times \mathcal{B}(S)$ for which containing $(0, g)$ implies $g = 0$.

The key point in the martingale method of Stroock-Varadhan is observing that, for a Markov process $\zeta(t)$ with the generator G, for $(f, g) \in G$

$$f(\zeta(t)) - \int_0^t g(\zeta(s))ds \tag{5.1.1}$$

is a martingale with respect to the filtration

$$\hat{\mathcal{F}}_t^\zeta := \mathcal{F}_t^\zeta \vee \sigma\Big(\int_0^s h(\zeta(r))dr \ s.t. \ h \in \mathcal{B}(S), r \leq t \Big), \tag{5.1.2}$$

where $\mathcal{F}_t^\zeta$ is the natural (canonical) filtration of ζ. In fact, for a progressive process ζ, we have $\hat{\mathcal{F}}_t^\zeta = \mathcal{F}_t^\zeta$. Looking conversely at the problem, the following question is imposed automatically. Can we restrict the set of functions in the domain of G such that ζ be the only "nice" stochastic process for which (5.1.1) is a martingale w.r.t. its canonical filtration? In fact the answer to this question is affirmative, and introduced by Stroock and Varadhan in [56, 57], this idea establishes a strong mechanism to deal with Markov processes, namely martingale method.

Here we speak of martingale problems for general Markov processes (possibly the time-inhomogeneous case). A martingale problem is identified by a triple $(G, \mathcal{D}, \mathbf{P}_0)$, where $\mathbf{P}_0 \in \mathcal{P}(S)$, $\mathcal{D} \subset \mathcal{B}(S)$, and $G : \mathbb{R}_+ \times \mathcal{D} \to \mathcal{B}(S)$ is a time-dependent linear operator with domain $\mathcal{D}$.

Definition 4. *An S-valued measurable stochastic process $\zeta = (\zeta_t)_{t\in\mathbb{R}}$, defined on $(\Omega, \mathbb{P}, \mathcal{F})$ is said to be a solution of the martingale problem $(G, \mathcal{D}, \boldsymbol{P}_0)$ if for ζ with sample paths in $D_S[0,\infty)$ and initial distribution $\boldsymbol{P}_0$, for any $f \in \mathcal{D}$*

$$f(\zeta(t)) - \int_0^t G_s f(\zeta(s)) ds \tag{5.1.3}$$

is a $\boldsymbol{P}$-martingale with respect to the canonical filtration, where $\boldsymbol{P} \in \mathcal{P}(D_S[0,\infty))$ is the law of ζ. We also say that $\boldsymbol{P}$ is a solution of $(\Omega, \mathbb{P}, \mathcal{F})$. The process ζ or its law $\boldsymbol{P}$ is said to be a general solution if the sample paths of ζ are not necessarily in $D_S[0,\infty)$, i.e. the support of $\boldsymbol{P}$ is not contained in $D_S[0,\infty)$. We say the martingale problem is well-posed if there is a unique solution (with paths in $D_S[0,\infty)$, general solutions not considered), that is there exists a unique $\boldsymbol{P} \in \mathcal{P}(D_S[0,\infty))$ that solves the martingale problem. It is said to be $\mathcal{C}([0,\infty), S)$-well posed, if it has a unique solution $\mathbf{P}$ in $\mathcal{P}(\mathcal{C}([0,\infty), S))$.

Remark 5.1.1. *If S is compact, $\mathcal{D}$ is dense in $\mathcal{C}(S)$, and $G : \mathbb{R}_+ \times \mathcal{D} \to \mathcal{C}(S)$, then well-posedness of martingale problems $\{(G, \mathcal{D}, \delta_x)\}_{x\in S}$ is equivalent to well-posedness of martingale problems $\{(G, \mathcal{D}, \boldsymbol{P}_0)\}_{\boldsymbol{P}_0 \in \mathcal{P}(S)}$ (See Ethier-Kurtz, question 49(a), chapter 4[23]). Also, for two time-dependent generators $G^1 : \mathbb{R}_+ \times \mathcal{D}_1 \to \mathcal{B}(S)$ and $G^2 : \mathbb{R}_+ \times \mathcal{D}_2 \to \mathcal{B}(S)$ with $G^1_t \subset G^2_t$ for any $t \geq 0$, any solution to the martingale problem $(G^2, \mathcal{D}_2, \boldsymbol{P}_0)$ is also a solution to the the martingale problem $(G^1, \mathcal{D}_1, \boldsymbol{P}_0)$ for an arbitrary initial distribution $\boldsymbol{P}_0 \in \mathcal{P}(S)$, but the converse does not necessarily holds.*

To have a better understanding of solutions of a martingale problem, we consider the forward equation for measure-valued functions defined on the positive real line. More explicitly, a solution to the forward equation with respect to the triple $(G, \mathcal{D}, \mathbf{P}_0)$, defined as above, is a measureable function $m : \mathbb{R}_+ \to \mathcal{P}(S)$ with $m(0) = \mathbf{P}_0$ satisfying the forward equation

$$< m_t, f > = < m_0, f > + \int_0^t < m_s, G_s f > ds \tag{5.1.4}$$

for any $f \in \mathcal{D}$ and $t \geq 0$, where $m_s := m(s)$ for any $s \geq 0$. It is clear that for any solution $\zeta = (\zeta_t)_{t \geq 0}$ to the martingale problem $(G, \mathcal{D}, \mathbf{P}_0)$, $(\mathbb{P}\zeta_t^{-1})_{t \geq 0}$ gives a solution to the corresponding forward equation, but not necessarily conversely (a number of sufficient conditions are given in [42]). Therefore, under certain conditions, the uniqueness of the second deduces the uniqueness of one-dimensional distributions for the solutions of the martingale problem which, in turn, implies the uniqueness of the martingale problem (see Proposition 5.1.8). If $G_t \subset \bar{\mathcal{C}}(S) \times \bar{\mathcal{C}}(S)$ for any time $t \geq 0$, and $\mathcal{D}$ is convergence-determining (Definition 6), then for any solution $(m_t)_{t \geq 0}$ of the forward equation, the mapping $t \mapsto m_t$ is continuous. So it is for $t \mapsto \mathbb{P}\zeta_t^{-1}$. Therefore, in this case, $t \mapsto \mathbb{E}_{\mathbf{P}}[\zeta_t]$ is continuous and differentiable. Regarding this, differentiating both sides in the forward equation under certain conditions to allow exchangeability of G_t and the integral for any $t \geq 0$, we derive $\mathbb{E}_{\mathbf{P}}[\zeta_t]$ as a solution of the Cauchy equation, that is

$$\frac{d}{dt}\mathbb{E}_{\mathbf{P}}[\zeta_t] = G_t \mathbb{E}_{\mathbf{P}}[\zeta_t] \tag{5.1.5}$$

for $t \geq 0$, and its uniqueness problem is well-studied and well-known when G is the infinitesimal generator of a contraction semigroup (Theorem 1.3 [21]).

To characterize a Markov processes with the martingale problem approach, we need to introduce a set of functions as the domain which is sufficiently large such that the corresponding martingale problem has a unique solution. In many cases we need this set to be an algebra. Uniqueness is essential in identification of Markov processes, since if a martingale problem has several solutions, it cannot identify a particular Markov process. The following concepts are useful in order to well establish the uniqueness of martingale problems. Also, we frequently will apply them in the rest of the chapter.

Definition 5. *We say a set of functions $U \subset \bar{C}(S)$ (more generally, $U \subset \mathcal{B}(S)$) separates points if for every $x, y \in S$ with $x \neq y$ there exists a function $f \in U$ for which $f(x) \neq f(y)$. In other words, U separates points if for every $x, y \in S$ we have $x = y$ if and only if $f(x) = f(y)$ for any $f \in U$. We also say U vanishes nowhere if for any $x \in S$ there exists a function $f \in U$ such that $f(x) \neq 0$.*

Definition 6. *A collection of functions $U \subset \bar{C}(S)$ (more generally, $U \subset \mathcal{B}(S)$) is said to be measure-determining on $\mathcal{M} \subset \mathcal{P}(S)$ if for $\boldsymbol{P}, \boldsymbol{P}' \in \mathcal{M}$, assuming*

$$\int_S f d\boldsymbol{P} = \int_S f d\boldsymbol{P}' \tag{5.1.6}$$

for all $f \in U$ implies $\boldsymbol{P} = \boldsymbol{P}'$. We say U is measure-determining, if it is measure-determining on $\mathcal{P}(S)$. Also, we say U is convergence-determining on $\mathcal{M} \subset \mathcal{P}(S)$ if for the sequence of probability measures $(\boldsymbol{P}_n)_{n\in\mathbb{N}}$ and the probability measure $\boldsymbol{P}$ in $\mathcal{M}$

$$\lim_{n\to\infty} \int_S f d\boldsymbol{P}_n = \int_S f d\boldsymbol{P} \quad for \quad all \quad f \in U \tag{5.1.7}$$

implies $\boldsymbol{P}_n \Rightarrow \boldsymbol{P}$. We say U is convergence-determining, if it is convergence-determining on $\mathcal{P}(S)$.

If $U \subset \bar{\mathcal{C}}(S)$ is convergence-determining then it is measure-determining , but the converse is not true in general. When S is a compact space, so is $\mathcal{P}(S)$, and the two concepts are equivalent by the following proposition.

Proposition 5.1.2. *Let S be a Polish space, and let $\boldsymbol{P}, \boldsymbol{P}_n \in \mathcal{P}(S)$ for $n \in \mathbb{N}$ such that $(\boldsymbol{P}_n)_{n\in\mathbb{N}}$ is tight. Let $U \subset \bar{\mathcal{C}}(S)$ be measure-determining. Then (5.1.7) implies $\boldsymbol{P}_n \Rightarrow \boldsymbol{P}$.*

Proof: Lemma 3.4.3 [23] ■

The following is an immediate consequence of the last proposition.

Corollary 5.1.3. *Let S be a compact space. Then $U \subset \mathcal{C}(S)$ is measure-determining if and only if it is convergence-determining.*

Proposition 5.1.4. *Let S be a Polish space and let $U \subset \bar{\mathcal{C}}(S)$ be an algebra of functions which separates points. Then U is measure-determining.*

Proof: Theorem 3.4.5 [23] ■

Corollary 5.1.5. *Let S be compact and let $U \subset \mathcal{C}(S)$ be an algebra of functions which separates points. Then U is convergence-determining.*

Returning to the uniqueness problem, in order to be able to transform some useful properties from the time-independent martingale problems to time-dependent ones, it is convenient to define the space-time process for the S-valued stochastic process $\zeta = (\zeta_t)_{t\geq 0}$ by $\zeta_t^* = (\zeta_t, t)$, which is an $S \times \mathbb{R}_+$-valued stochastic process. Consider the particular case when ζ is an S-valued time-inhomogeneous Markov process with time-dependent generator $G : \mathbb{R}_+ \times \mathcal{D}(G) \to \mathcal{B}(S)$, i.e. $(G_t : \mathcal{D}(G) \to \mathcal{B}(S))_{t\geq 0}$. Let $(T_{s,r})_{0\leq s\leq t}$ be the time-inhomogeneous semigroup of ζ, corresponding to G defined by

$$T_{s,r} f(x) = \int_S f(y) p(s, x; r, dy) \ for \ f \in \mathcal{B}(S), \tag{5.1.8}$$

where $p(s, x; r, dy)$ is the transition probability for ζ. From the definition of space-time process, ζ^* is also Markov with the transition probability

$$p^*(t, (x, s), d(y, r)) = \delta_{t+s}(r) p(s, x; r, dy) \tag{5.1.9}$$

where $\delta_u(.)$ is the delta measure on $\mathbb{R}_+$ for $u \geq 0$. Therefore, the semigroup of ζ^* is given by

$$\begin{aligned} T^* f^*(x,s) &= \int\limits_{S\times[0,\infty)} f^*(y,r)p^*(t,(x,s),d(y,r)) = \\ &\int\limits_S f^*(y,t+s)p(s,x;t+s,dy) \end{aligned} \tag{5.1.10}$$

for $f^* \in \mathcal{B}(S\times[0,\infty))$. Let G^* be the infinitesimal generator of ζ^* with domain $\mathcal{D}(G^*)$. Let $\mathcal{D}^* := \{fh \in \mathcal{B}(S\times[0,\infty)) : f \in \mathcal{D}(A), h \in \mathcal{C}_c^1([0,\infty),\mathbb{R})\}$ where $\mathcal{C}_c^1([0,\infty),\mathbb{R})$ is the set of all continuously differentiable real valued functions on $[0,\infty)$ with compact support. In particular for any $f^* = fh \in \mathcal{D}^*$ with $f \in \mathcal{D}(A)$ and $h \in \mathcal{C}_c^1([0,\infty),\mathbb{R})$, we have

$$T^* fh(x,s) = h(s+t)T_{s,s+t}f(x). \tag{5.1.11}$$

Therefore, $\mathcal{D}^* \subset \mathcal{D}(G^*)$, and the infinitesimal generator of ζ^* restricted to the domain $\mathcal{D}^*$ is given by

$$\begin{aligned} G^* fh(x,s) &= \lim_{t\to 0} \tfrac{h(s+t)T_{s,s+t}f(x) - h(s)f(x)}{t} = \\ &h(s)G_s f(x) + h'(s)f(x). \end{aligned} \tag{5.1.12}$$

Let $G' = G^*|_{\mathcal{D}^*}$. In other words, $G' \subset \mathcal{B}(S\times\mathbb{R}_+)\times\mathcal{B}(S\times\mathbb{R}_+)$ is defined by

$$G' = \{(fh, gh + fh') : (f,g) \in G, h \in \mathcal{C}_c^1([0,\infty),\mathbb{R})\}. \tag{5.1.13}$$

By Theorem 4.7.1 of (Ethier and Kurtz-1986 [23]) , ζ is a solution to the time-dependent martingale problem $(G, \mathcal{D}(G), \mathbf{P}_0)$ for $\mathbf{P}_0 \in \mathcal{P}(S)$ if and only if ζ^* is a solution to the martingale problem $(G^*, \mathcal{D}^*, \mathbf{P}_0^*)$, where $\mathbf{P}_0^* \in \mathcal{P}(S\times[0,\infty))$ is the image of $\mathbf{P}_0$ under the projection $x \mapsto (x,0)$, that is $\mathbf{P}_0^*(A,r) = \delta_0(r)\mathbf{P}_0(A)$ for $A \in \mathscr{B}(S)$. If, in addition, we assume $\mathcal{D}(G) \subset \bar{\mathcal{C}}(S)$ and that it also separates points and vanishes nowhere, then we can extend G' to a subset of G^* whose domain is an algebra which separates points. As G^* is linear and as $\mathcal{D}^*$ is closed under pointwise multiplication of functions, the algebra of functions generated by $\mathcal{D}^*$, denoted by $\mathcal{D}^{**}$, is a linear subspace of $\mathcal{D}(G^*)$. Also $\mathcal{D}^{**}$ separates points and vanishes nowhere. Hence $\mathcal{D}(G^*)$ is dense in $\bar{\mathcal{C}}(S\times\mathbb{R}_+)$ in the topology of uniform convergence on compact sets that co ncludes $G^* : \mathcal{D}(G^*) \to \bar{\mathcal{C}}(S \times \mathbb{R}_+)$. Let $G'' = G^*|_{\mathcal{D}^{**}}$. Consider the martingale problem $(G^*, \mathcal{D}^{**}, \mathbf{P}_0^*)$ (or, with an equivalent notation, $(G'', \mathcal{D}^{**}, \mathbf{P}_0^*)$). By linearity, any solution to $(G^*, \mathcal{D}^{**}, \mathbf{P}_0^*)$ is a solution to $(G^*, \mathcal{D}^*, \mathbf{P}_0^*)$ and vice versa. Hence ζ is a solution to the time-dependent martingale problem $(G, \mathcal{D}(G), \mathbf{P}_0)$ if and only if ζ^* is a solution to the martingale problem $(G^*, \mathcal{D}^{**}, \mathbf{P}_0^*)$.

The following lemma is useful to prove uniqueness in the case that we have a Markov solution of a time-dependent martingale problem.

Lemma 5.1.6. *Let S be a Polish space and $G = (G_t)_{t \geq 0}$ be a time-dependent linear operator on $\mathcal{B}(S)$ with the domain $\mathcal{D}(G) \subset \bar{\mathcal{C}}(S)$ that contains an algebra of functions that separates points. Suppose there exists an S-valued Markov process with generator G and initial distribution $\mathbf{P}_0 \in \mathcal{P}(S)$. Then the time-inhomogeneous martingale problem $(G, \mathcal{D}(G), \mathbf{P}_0)$ is well-posed.*

Remark 5.1.7. *The lemma remains true if S is a separable metric space.*

Proof: Without loss of generality, we assume $\mathcal{D}(G)$ is an algebra, separating points. Otherwise, we prove the theorem for the subalgebra of $\mathcal{D}(G)$ with this property, which has at least one Markov solution, and hence the uniqueness of the latter implies the uniqueness of the original martingale problem. Let a Markov process $\zeta = (\zeta_t)_{t \geq 0}$ be a solution to the $(G, \mathcal{D}(G), \mathbf{P}_0)$ martingale problem. Then by discussion before the above lemma, the Markov space-time process $\zeta^* = (\zeta_t^*)_{t \geq 0}$ defined by $\zeta_t^* = (\zeta_t, t)$ is a solution to the martingale problem $(G^*, \mathcal{D}^{**}, \mathbf{P}_0^*)$ (equivalently, $(G^*, \mathcal{D}^{**}, \mathbf{P}_0^*)$), where G^*, $\mathcal{D}^{**}$ and $\mathbf{P}_0^*$ are defined as before. Therefore, it is sufficient to prove that $(\zeta_t^*)_{t \geq 0}$ is the unique solution to this martingale problem. Since G'' is the infinitesimal generator of ζ^* restricted to the domain $\mathcal{D}^{**}$, it is dissipative and there is $\lambda > 0$ such that $\mathcal{R}(\lambda - G'') = \overline{\mathcal{D}(G'')} = \overline{\mathcal{D}^{**}}$. Thus, by theorem 4.4.1 in [23], we need only to show that $\mathcal{D}^{**}$ is measure-determining. But $\mathcal{D}^{**} \subset \bar{\mathcal{C}}(S \times \mathbb{R}_+)$ and the algebra $\mathcal{D}^{**}$ separates points. The latter follows from the fact that both

$\mathcal{D}(G)$ and $\mathcal{C}_c^1([0,\infty),\mathbb{R}_+)$ separate points. Hence by Proposition 5.1.4 $\mathcal{D}^{**}$ is measure-determining. This finishes the proof. ∎

In fact, we can see that the uniqueness of one-dimensional distributions of solutions of a martingale problem $(G,\mathcal{D},\mathbf{P}_0)$ guarantees the uniqueness of the finite-dimensional distributions which, in turn, implies uniqueness of the martingale problem. Another important fact about martingale formulation is that any unique solution of a martingale problem $(G,\mathcal{D},\mathbf{P}_0)$ is strongly Markovian. The following is a restatement of Theorem 4.4.2 and Corollary 4.4.3 (Ethier and Kurtz 1986 [23]) in the case of time-inhomogeneous martingale problems.

Proposition 5.1.8. *Let S be a separable metric space, and let $G:\mathbb{R}_+\times\mathcal{B}(S)\to\mathcal{B}(S)$ be a time-dependent linear operator. If the one-dimensional distributions of any two possible solutions $\zeta^{(1)}$ and $\zeta^{(2)}$ of the martingale problem $(G,\mathcal{D},\boldsymbol{P}_0)$ (with sample paths in $\mathcal{C}([0,\infty),S)$, respectively) coincide, i.e. if*

$$\mathbb{P}(\zeta^{(1)}(t)\in A)=\mathbb{P}(\zeta^{(2)}(t)\in A) \tag{5.1.14}$$

for any $t\geq 0$ and $A\in\mathscr{B}(S)$, then the martingale problem $(G,\mathcal{D},\boldsymbol{P}_0)$ has at most one solution (with sample paths in $\mathcal{C}([0,\infty),S)$, respectively). In the case of existence, the solution is a Markov process. In addition to the above assumptions, if G is a linear time-dependent operator satisfying $G:\mathbb{R}_+\times\bar{\mathcal{C}}(S)\to\mathcal{B}(S)$, then in the case of existence, the unique solution of $(G,\mathcal{D},\boldsymbol{P}_0)$, namely ζ, is strongly Markov, i.e. for any a.s. finite stopping time τ (with respect to the canonical filtration of ζ, namely $\{\mathcal{F}_t^\zeta\}_{t\in\mathbb{R}_+}$)

$$\mathbb{E}_{\mathbb{P}\zeta^{-1}}[f(\zeta(\tau+t)|\mathcal{F}_\tau^\zeta]=\mathbb{E}_{\mathbb{P}\zeta^{-1}}[f(\zeta(\tau+t)|\zeta(\tau)] \tag{5.1.15}$$

Proof: The same argument as the one used in the proof of Theorem 4.4.2 and Corollary 4.4.3 (Ethier and Kurtz 1986 [23]) (in homogeneous martingale problem case) proves the proposition. ∎

To have the uniqueness for a time-inhomogeneous martingale problem $(G, \mathcal{D}, \mathbf{P}_0)$, it is necessary and sufficient that the one-dimensional distributions of any two possible solutions of the martingale problem coincide. The uniqueness of the one-dimensional distributions concludes that every solution is a Markov process, and, hence, it implies the uniqueness of time-dependent semigroups of two solutions. For general theory of martingale problems one can see [58, 23, 8].

In the next subsection we develop the martingale problems with stochastic operator-valued processes.

5.1.3 Quenched martingale problem in random environment and stochastic operator process

As we restrict our attention to population dynamics in random time-varying environments (fitness functions), we are dealing with the Markov processes which are not only inhomogeneous in time but also their generators are random (depends on a random environment). Therefore, the idea of martingale problem should be extended to characterize time-inhomogeneous Markov processes in random environments with the mentioned characteristic. In the following paragraph, we describe what we mean by a stochastic process in a random environment.

Definition 7. *Let S' be a Polish space and let $\{\zeta^x\}_{x\in S'}$ be a family of S-valued measurable stochastic processes with laws $\{\boldsymbol{P}^x\}_{x\in S'}$ all defined on the Borel probability space $(\Omega, \mathbb{P}, \mathcal{F})$ and with sample paths in $D_S[0,\infty)$ a.s.. As the supports of all $\{\boldsymbol{P}^x\}_{x\in S'}$ are in $D_S[0,\infty)$, we regard these measures as the elements of $\mathcal{P}(D_S[0,\infty))$. Suppose the map $x \mapsto \boldsymbol{P}^x(B)$ is measurable for any Borel measurable subset B contained in $D_S[0,\infty)$. Let $X : \Omega \to S'$ be Borel measurable, i.e. X is an S'-valued random variable. The mapping $\zeta^X : \omega \mapsto \zeta^{X(\omega)}(\omega)$ is called an S-valued stochastic (annealed) process in random environment X with law $\boldsymbol{P}$ which is the average over* *all $\boldsymbol{P}^x$, i.e.*

$$\boldsymbol{P}(.) = \int_{\Omega} \boldsymbol{P}^x(.)\mathbb{P}^{env}(dx) \tag{5.1.16}$$

where $\mathbb{P}^{env} = X * \mathbb{P} = \mathbb{P}X^{-1}$ is the law of X. Recall that $X * \mathbb{P}$ is the push-forward measure of $\mathbb{P}$ under the random variable X. For any $x \in S'$, ζ^x is called the quenched process with the given environment x. As $\boldsymbol{P} \in \mathcal{P}(D_S[0,\infty))$, the annealed process has sample paths in $D_S[0,\infty)$.

The fact that the quenched processes are Markov does not guarantee that the annealed process is Markov. Also, ζ^X need not to be Markov when, for $\mathbb{P}^{env}$-a.e. $x \in S'$, ζ^x is Markov but X is not so.

We can consider a stochastic process in r.e. from the perspective of its generator which is a random time-varying generator. This hints us to think of the process by keeping the information of random variations for its generator. The following develops this concept.

Definition 8. *Consider a Banach space L with a linear subset $\mathcal{D} \subset L$ (i.e. $\mathcal{D}$ is closed under vector addition and scalar multiplication) and denote by $\mathscr{L}(\mathcal{D}, L)$ the set of all bounded linear operators with domain $\mathcal{D}$ on L, equipped with the operator norm. By a linear operator process on L with the domain $\mathcal{D} \subset L$ we mean an $\mathscr{L}(\mathcal{D}, L)$-valued stochastic process, i.e. a mapping $G : \Omega \times \mathbb{R}_+ \to \mathscr{L}(\mathcal{D}, L)$ such that $G(., t)$ is a measurable function for any $t \in \mathbb{R}_+$ and $G(\omega, t)$ is a bounded linear operator on L with domain $\mathcal{D}$ for any $t \in \mathbb{R}_+$ and $\omega \in \Omega$. A measurable linear operator process G is a linear operator process for which G is a measurable function. We, interchangeably, denote G either as above, or as a function from $\Omega \times \mathbb{R}_+ \times \mathcal{D}$ to L, i.e. for any $f \in \mathcal{D}$ we set $G(\omega, t, f) = G(\omega, t)f$. Also, by $(G_t)_{t\geq 0}$ we denote a general linear operator process.*

As we odeal withnly linear operators in this chapter, we call the process defined above an "operator process".

Definition 9. *For an S'-valued random variable $X : \Omega \to S'$, we say an operator process $G : \Omega \times \mathbb{R}_+ \to \mathscr{L}(\mathcal{D}, L)$ is consistent with X (or with its law $\mathbb{P}X^{-1} \in \mathcal{P}(S')$) if for any $t \in \mathbb{R}_+$ the function $G(.,t)$ is constant on every measurable pre-image $X^{-1}(x)$ for $x \in supp(\mathbb{P}X^{-1})$. In this case we set $G(X^{-1}(x), s) := G(\omega, s)$ for an arbitrary $\omega \in X^{-1}(x)$.*

Let S' be a Polish space and $\mathcal{D}$ be a linear subspace of $\mathcal{B}(S)$. Let a probability measure $\mathbb{P}^{env} \in \mathbb{P}(S')$ be the distribution of an S'-valued random variable $X : \Omega \to S'$ with law $\mathbb{P}^{env} = \mathbb{P}X^{-1}$. Consider the operator process $G : \Omega \times \mathbb{R}_+ \times \mathcal{D} \to \mathcal{B}(S)$ which is consistent with X. We identify a time-inhomogeneous martingale problem in random environment (r.e.), or a quenched martingale problem in r.e. X, with a quadruple $(G, \mathcal{D}, \mathbf{P}_0, \mathbb{P}^{env})$ where $\mathbf{P}_0 : S' \to \mathcal{P}(S)$ is measurable. From now on, when we speak of a quenched martingale problem $(G, \mathcal{D}, \mathbf{P}_0, \mathbb{P}^{env})$, we automatically assume that G is consistent with $\mathbb{P}^{env}$.

Definition 10. *(Quenched martingale problem in r.e.) Let X be an S'-valued random variable with law $\mathbb{P}^{env}$. An S-valued stochastic process in r.e. X, namely ζ^X, with the family of quenched laws $\{\boldsymbol{P}^x\}_{x\in S'}$ and initial distributions $\{\boldsymbol{P}_0(x)\}_{x\in S'}$ is said to be a solution of the quenched martingale problem $(G, \mathcal{D}, \boldsymbol{P}_0, \mathbb{P}^{env})$ if for any $f \in \mathcal{D}$*

$$f(\zeta_t^x) - \int_0^t G_x(s) f(\zeta_s^x) ds \tag{5.1.17}$$

is a $\boldsymbol{P}^x$-martingale with respect to the canonical filtration, for $\mathbb{P}^{env}$-almost all $x \in S'$, where

$$G_x(s) = G(X^{-1}(x), s) \tag{5.1.18}$$

In this case, we also say $\{\boldsymbol{P}^x\}_{x\in S'}$ is a solution to the quenched martingale problem. We say $\{\boldsymbol{P}^x\}_{x\in S'}$ is a general solution to the martingale problem, if there exists $U \in$

$\mathscr{B}(S')$ with $\mathbb{P}^{env}(U) > 0$ such that for any $x \in U$ the support of $\boldsymbol{P}^x$ is not in $D_S[0,\infty)$. We say the martingale problem is well-posed if there is a unique ζ^X solution with these properties (general solutions are not considered in the definition of uniqueness), i.e. there exists a unique ($\mathbb{P}^{env}$-a.s. uniquely determined) family $\{\boldsymbol{P}^x\}_{x\in S'}$ satisfying the above conditions.

Remark 5.1.9. *In fact, each quenched martingale problem in r.e. $(G, \mathcal{D}, \boldsymbol{P}_0, \mathbb{P}^{env})$, $\mathbb{P}^{env}$-a.s. uniquely determines an S'-indexed family of time-inhomogeneous martingale problems $\{(G_x, \mathcal{D}, \boldsymbol{P}_0(x))\}_{x\in S'}$ along with the environment probability measure $\mathbb{P}^{env}$, and vice versa. The problems of existence and uniqueness of the first are equivalent to the problems of existence and uniqueness of the second family $\mathbb{P}^{env}$-a.s..*

5.1.4 Markov process joint with its environment

Another process drawing our attention is a joint stochastic process and its environment.

Definition 11. *Suppose ζ^X be as defined in Definition 7, i.e. a stochastic process in random environment X, with an extra assumption $S' = D_{\mathbf{E}}[0,\infty)$ for a Polish space $\mathbf{E}$. The random variable X can be regarded as an $\mathbf{E}$-valued stochastic process with sample paths in S'. For $t \in \mathbb{R}_+$ and $\omega \in \Omega$, define the joint $S \times \mathbf{E}$-valued stochastic process by $\zeta_t(\omega) = (\zeta_t^{X(\omega)}(\omega), X_t(\omega))$ and call it the joint annealed-environment process. In fact for each $\omega \in \Omega$, $\zeta(\omega)$ gives a trajectory of environment and a trajectory of the process in that environment.*

Suppose we have the law of the joint process ζ. How can we write the law of ζ^X in term of the law of ζ? Let $\mathbf{P}^*$ be the law of ζ and $\mathbb{P}^{env}$ be the law of X. Since S and S' are Polish, by disintegration theorem, there exist a unique family of probability measures $\{\mathbf{P}^{*x}\}_{x\in S'} \subset \mathcal{P}(D_S[0,\infty) \times S')$ (unique w.r.t. $\mathbb{P}^{env}$, that is for any such family $\{\mathbf{P}^{**x}\}_{x\in S'}$, $\mathbb{P}^{env}(\{x \in S' : \mathbf{P}^{*x} = \mathbf{P}^{**x}\}) = 1$ satisfying

(i) The map $x \mapsto \mathbf{P}^{*x}(B)$ is Borel measurable for any Borel measurable set $B \subset D_S[0,\infty) \times S'$.

(ii) $\mathbf{P}^{*x}((D_S[0,\infty)\times S')\backslash Y_x) = 0$ for $\mathbb{P}^{env}$-almost all $x \in S'$ where $Y_x := \{(s,x)\}_{s\in D_S[0,\infty)}$.

(iii) For any Borel measurable subset B of $D_S[0,\infty) \times S'$ we have

$$\mathbf{P}^*(B) = \int_{S'} \mathbf{P}^{*x}(B)\mathbb{P}^{env}(dx). \tag{5.1.19}$$

Then, for $\mathbb{P}^{env}$-a.e. $x \in S'$, $\mathbf{P}^x$ will be the push-forward measure of $\mathbf{P}^{*x}$ under the measurable projection from $D_S[0,\infty) \times S'$ onto $D_S[0,\infty)$. We also can observe that the annealed measure $\mathbf{P}$ is the push-forward measure of $\mathbf{P}^*$ under the projection from $D_S[0,\infty) \times S'$ onto $D_S[0,\infty)$, and it can gives another way to construct quenched measures, as they are in fact conditional measures of $\mathbf{P}$ and can be derived by disintegration theorem for $\mathbf{P}$ and $\mathbb{P}^{env}$. The following diagram summarizes the relations of these measures.

$$\begin{array}{ccc} \mathbf{P}^* & \xrightarrow{push-forward} & \mathbf{P} \\ \downarrow & & \downarrow \\ \{\mathbf{P}^{*x}\}_{x\in S'} & \xrightarrow{push-forward} & \{\mathbf{P}^x\}_{x\in S'} \end{array} \tag{5.1.20}$$

5.2 Moran and Fleming-Viot processes in random environments: Martingale characterization

In this section, first, we identify the Moran process in r.e. as a quenched martingale problem and prove its wellposedness, and then we define the generator of the Fleming-Viot (FV) process in r.e. and the quenched martingale problem for it as well. Also, in this section, we state some main results of this chapter for Moran and FV processes in r.e.. This includes the wellposedness of the quenched martingale problem for the FV process in r.e., some properties of this process such as continuity of the sample paths almost surely, and the weak convergence of the quenched (and annealed) measure-valued Moran processes to the quenched (and annealed) FV process, when the environments of the first (fitness functions) converge to that of the second. Also, under the assumption of existence of a parent-independent component of the mutation process and certain assumptions for the environment process, we state an ergodic theorem for the annealed-environment process. The proofs of the main theorems will come in sections 5.6, 5.7, and 5.8.

Throughout this chapter we assume that I is a compact metric space, called type space. Any element of I is called a type or allele. We also assume that $E \subset \mathcal{C}(I,[0,1])$ is compact. Hence, I and E are Polish spaces as they are compact. A fitness function (or selection intensity function) is a Borel measurable function from I to $[0,1]$. In this chapter, we assume that fitness functions are in $C(I,[0,1])$.

Definition 12. *A fitness process is an E-valued measurable stochastic process defined on $(\Omega,\mathbb{P},\mathcal{F})$ with sample paths in $D_E[0,\infty)$. When the fitness process is Markov, we call it Markov fitness.*

Remark 5.2.1. *Restricting the fitness processes to have sample paths in $D_E[0,\infty)$ is an essential assumption to guarantee the generators of Moran and Fleming-Viot processes in random environments exist for a suitable set of functions.*

Let e be a fitness process. As E is a compact space and therefore separable, e can be regarded as a $D_E[0,\infty)$-valued random variable defined on $(\Omega,\mathbb{P},\mathcal{F})$, that is $e:\Omega\to D_E[0,\infty)$ be a measurable map. We denote by

$$\mathbb{P}^{env} = \mathbb{P}^{env,e} := e * \mathbb{P} = \mathbb{P}e^{-1} \in \mathcal{P}(D_E[0,\infty)) \tag{5.2.1}$$

the distribution of e. For simplicity of notation, we let $e_t = e(t)$ for a fitness process e. We frequently denote by $e = (e_t)_{t\geq 0}$ a fitness process and by $\tilde{e} \in D_E[0,\infty)$ a trajectory of e. Also, we denote by $\hat{e} \in E$ a fitness function. We emphasise that, in the sequel, a fitness process e is regarded as both an E-valued measurable stochastic process with sample paths in $D_E[0,\infty)$ and a $D_E[0,\infty)$-valued random variable with the law $\mathbb{P}^{env}$. We assume that the possible times of selection occur with rate α/N independently for every individual, and at a possible time t of selection for individual $i \leq N$, a selective event occurs with probability $e_t(a_i(t))$, where $a_i(t)$ is the type of individual i at time t. For the results in this chapter we assume that the fitness process is either a general E-valued stochastic process or a Markov process.

We continue this section with identifying the Moran process in r.e. (MRE) as a solution of a quenched martingale problem in r.e..

5.2.1 Moran process in random environments

In this chapter, by a Moran process we think of the measure-valued Moran process with resampling, mutation and selection with a compact type space I and the E-valued fitness process as constructed in Chapter 1 in detail. Recall that E is compact in this chapter. For $N \in \mathbb{N}$, let $\mathcal{P}^N(S)$ be the set of all purely atomic probability measures in $\mathcal{P}(S)$ with at most N atoms such that $Nm(.)$ is a counting measure. In other words, $\mathcal{P}^N(S)$ is the image of I^N under the map

$$(a_1, ..., a_N) \mapsto \frac{1}{N}\sum_{i=1}^{N} \delta_{a_i} \tag{5.2.2}$$

from I^N to $\mathcal{P}(S)$ where δ_a is the delta measure with support $\{a\} \subset S$. An element of $\mathcal{P}^N(S)$ is called an empirical measure on S (with at most N atoms). In this section we assume that the number of individuals $N \geq 1$ is fixed. With N individuals and type space I, let $(\mu_N^e(t))_{t\in\mathbb{R}_+}$ be a measure-valued ($\mathcal{P}^N(I)$-valued) Moran process with fitness process e whose law is given by $\mathbb{P}^{env} = \mathbb{P}^{env,e} \in \mathcal{P}(D_E[0,\infty))$ and let $\gamma, \alpha/N, \beta > 0$ be the selection, mutation, and resampling rates, respectively. We assume that the fitness process e evolves between jumps and is independent of the Poisson times of jumps (for resampling, mutation and selection), the initial distribution, and also of the mutation kernel, i.e. it is independent of the outcome of a mutation event that occurs on type a for every $a \in I$. Let $q(x, dy)$ be a stochastic kernel for the mutation process on the state space I, that is the type of an offspring of the allele a after a mutation event follows the transition function $q(x, dy)$. As $q(x, dy)$ can either depend on x or not, it is always possible to write the mutation kernel as

$$\beta q(x, dy) = \beta' q'(dy) + \beta'' q''(x, dy) \tag{5.2.3}$$

for $\beta', \beta'' \geq 0$ s.t. $\beta' + \beta'' = \beta > 0$. The first term in the right hand side of equation (5.2.3) is called parent-independent component of the mutation event. When there is no ambiguity in the notation and the fitness process is known, we drop the superscript e and denote MRE with the fitness function e by μ_N. Also we denote by $\mu_N^{\tilde{e}}$ the quenched Moran process with the deterministic fitness process $\tilde{e} \in D_E[0, \infty)$.

To study μ_N as a quenched martingale problem in r.e. e, we need to determine a convenient set of functions as the domain of its generator. We use the following domain for the generator of μ_N which has been used by several authors as a domain for the generator of the classic measure-valued Moran process. For an empirical measure $m \in \mathcal{P}^N(I)$, let $m^{(N)} \in \mathcal{P}(I^N)$ be the N times sampling measure without replacement from m, i.e. letting $da = d(a_1, ..., a_N)$

$$m^{(N)}(da) = m(da_1) \times \frac{Nm - \delta_{a_1}}{N-1}(da_2) \times ... \times \frac{Nm - \sum_{i=1}^{N-1} \delta_{a_i}}{1}(da_N) \tag{5.2.4}$$

Let $\tilde{\mathfrak{F}}_N$ be the algebra generated by all functions $\tilde{\Phi}_N^f : \mathcal{P}^N(I) \to \mathbb{R}$ for $f \in \mathcal{C}(I^N)$ with

$$\tilde{\Phi}_N^f(m) =< m^{(N)}, f >= \int_{I^N} f dm^{(N)}. \tag{5.2.5}$$

Note that $\bar{\mathcal{C}}(I) = \mathcal{C}(I)$, $\bar{\mathcal{C}}(I^N) = \mathcal{C}(I^N)$, $\bar{\mathcal{C}}(\mathcal{P}(I)) = \mathcal{C}(\mathcal{P}(I))$ and $\bar{\mathcal{C}}(\mathcal{P}^N(I)) =$

$\mathcal{C}(\mathcal{P}^N(I))$, since I and therefore $\mathcal{P}(I)$ and $\mathcal{P}^N(I)$ are compact. Also note that any function in $\tilde{\mathfrak{F}}_N$ is a restriction of a function in $\mathcal{C}(\mathcal{P}(I))$.

Proposition 5.2.2. *For any $N \in \mathbb{N}$, the algebra $\tilde{\mathfrak{F}}_N$ separates points, and hence is measure and convergence-determining on $\bar{\mathcal{C}}(\mathcal{P}^N(I))$. Also $\tilde{\mathfrak{F}}_N$ vanishes nowhere.*

Proof: Let $m_1, m_2 \in \mathcal{P}^N(I)$ such that $m_1 \neq m_2$. Let $\mathcal{R} := supp(m_1) \cap supp(m_2)$. There exists $a \in \mathcal{R}$ such that $m_1(a) \neq m_2(a)$. Since $|\mathcal{R}| \leq 2N$, there exists $r > 0$ such that the ball radius r centred at a (w.r.t. the metric of I), namely $B(a, r)$, excludes all the points of $\mathcal{R}$ except a. It is clear that there exists a function $f \in \mathcal{C}(I)$ with $f(a) = 1$ that vanishes outside of $B(a, r)$. Consider $\tilde{f} \in \mathcal{C}(I^N)$ which depends only on the first variable in I^N and defined by $\tilde{f}(x_1, x_2, ..., x_3) = f(x_1)$. Then

$$< m_1^{(N)}, \tilde{f} >=< m_1, f >= m_1(a) \neq m_2(a) =< m_2, f >=< m_2^{(N)}, \tilde{f} > . \quad (5.2.6)$$

Also, $\tilde{\mathfrak{F}}_N$ vanishes nowhere, since the constant function $1 \in \mathcal{C}(I^N)$ and for any $m \in \mathcal{P}^N(I)$, $< m, 1 >= 1 \neq 0$. ■

Remark 5.2.3. *Alternatively, the latter proposition can be proved by showing that $\tilde{\mathfrak{F}}_N$ strongly separates points. For the definition see [23], Section* 3.4.

It is straightforward to see that the generator of the MRE with fitness process e on $\tilde{\mathfrak{F}}_N$ is the operator process $\tilde{\mathcal{G}}^N : \Omega \times \mathbb{R}_+ \times \tilde{\mathfrak{F}}_N \to \tilde{\mathfrak{F}}_N$ consistent with the environment process e given by

$$\begin{aligned} &\tilde{\mathcal{G}}^N = \tilde{\mathcal{G}}^{N,e} := \\ &\tilde{\mathcal{G}}^{res,N,e} + \tilde{\mathcal{G}}^{mut,N,e} + \tilde{\mathcal{G}}^{sel,N,e} \end{aligned} \quad (5.2.7)$$

where $\tilde{\mathcal{G}}^{res,N,e}$ and $\tilde{\mathcal{G}}^{mut,N,e}$, i.e. the resampling and mutation generators, are linear operators from $\tilde{\mathfrak{F}}_N$ to $\tilde{\mathfrak{F}}_N$, and $\tilde{\mathcal{G}}^{sel,N,e}$, the selection generator, is an operator process consistent with e. We usually drop the superscript e, if there is no risk of ambiguity.

To be more explicit, let

$$\bar{I}_k = (\bigcup_{n=k}^{\infty} I^n) \cup I^{\mathbb{N}} \tag{5.2.8}$$

for $k \in \mathbb{N}$. For the resampling generator, we have

$$\tilde{\mathcal{G}}^{res,N}\tilde{\Phi}^f_N(m) = \frac{\gamma}{2}\sum_{i,j=1}^{N} < m^{(N)}, f \circ \sigma_{ij} - f > \tag{5.2.9}$$

where $\sigma_{ij} : \bar{I}_{i\vee j} \to \bar{I}_{i\vee j}$ is a map replacing the j-th component of $x \in \bar{I}_{i\vee j}$ with the i-th one ($i \vee j = max\{i,j\}$). In other words, defining another map $\sigma^y_j : \bar{I}_j \to \bar{I}_j$ for $j \in \mathbb{N}$ and $y \in I$ with $\sigma^y_j(x) := (x_1, ..., x_{j-1}, y, x_{j+1}, ..., x_n)$ if $j \leq n$ and $x = (x_1, ..., x_n)$, and with $\sigma^y_j(x) := (x_1, ..., x_{j-1}, y, x_{j+1}, ...)$ if $x = (x_1, x_2, ...)$, we have $\sigma_{ij}(x) = \sigma^{x_i}_j$ (the reason to define these functions to be so general is to use them later for Fleming-Viot processes). To show that the generator is as indicated above, we write

$$\begin{aligned}\tilde{\mathcal{G}}^{res,N}\tilde{\Phi}^f_N(m) = \lim_{s\to 0}\frac{1}{s}(\frac{\gamma}{2}s\sum_{0\leq i,j\leq N} < m^{(N)}, f \circ \sigma_{ij} > \\ + (1-\frac{\gamma}{2}s)\sum_{0\leq i,j\leq N} < m^{(N)}, f > - \sum_{0\leq i,j\leq N} < m^{(N)}, f >)\end{aligned} \tag{5.2.10}$$

that implies (5.2.9).

Similarly, for mutation, we have

$$\begin{aligned}\tilde{\mathcal{G}}^{mut,N}\tilde{\Phi}^f_N(m) &= \beta\sum_{i=1}^{N} < m^{(N)}, B_i f - f > \\ &= \beta'\sum_{i=1}^{N} < m^{(N)}, B'_i f - f > + \beta''\sum_{i=1}^{N} < m^{(N)}, B''_i f - f >,\end{aligned} \tag{5.2.11}$$

where

$$B_i, B'_i, B''_i : \mathcal{C}(I^{\mathbb{N}}) \cup (\bigcup_{n\geq i}\mathcal{C}(I^n)) \to \mathcal{C}(I^{\mathbb{N}}) \cup (\bigcup_{n\geq i}\mathcal{C}(I^n)) \tag{5.2.12}$$

are bounded linear operators defined by

$$B_i f(x) = \int_I f \circ \sigma^y_i(x) q(x_i, dy), \tag{5.2.13}$$

$$B'_i f(x) = B'_i f(x) := \int_I f \circ \sigma_i^y(x) q'(dy), \tag{5.2.14}$$

$$B''_i f(x) = B''_i f(x) := \int_I f \circ \sigma_i^y(x) q''(x_i, dy). \tag{5.2.15}$$

Note that B_i, B'_i, B''_i leaves $\mathcal{C}(I^{\mathbb{N}})$ and, for any $n \in \mathbb{N}$, $\mathcal{C}(I^n)$ invariant.

To have a generator process consistent with e we first need to specify the time-dependent generator of $\tilde{\mu}_N^{\tilde{e}}$ for any given (quenched) environment $\tilde{e}$. For any $t \geq 0$, let

$$\tilde{e}_i(t)(x) := \tilde{e}(t)(x_i) \tag{5.2.16}$$

and, for $x \in I^N$ and $s \geq 0$, define

$$\overrightarrow{e}_i^{t,t+s}(x) := \int_t^{t+s} \tilde{e}_i(r)(x) P^*_{t,t+s}(dr), \tag{5.2.17}$$

where $P^*_{t,t+s}$ is the distribution of the position of a Poisson point in the interval $[t, t+s]$, conditioned to have exactly one Poisson point in the interval. Then the generator process for the selection is

$$\begin{aligned} &\tilde{\mathcal{G}}_{\tilde{e}}^{sel,N}(t)\tilde{\Phi}_N^f(m) = \\ &\lim_{s\to 0} \tfrac{1}{s}\big(\tfrac{\alpha}{N} s \sum_{i,j=1}^N < m^{(N)}, \overrightarrow{e}_i^{t,t+s} f \circ \sigma_{ij} > + \tfrac{\alpha}{N} s \sum_{i,j=1}^N < m^{(N)}, (1 - \overrightarrow{e}_i^{t,t+s}) f > \\ &+ (1 - \tfrac{\alpha}{N} s) \sum_{i,j=1}^N < m^{(N)}, f > - \sum_{i,j=1}^N < m^{(N)}, f > \big) \\ &= \lim_{s\to 0} \tfrac{1}{s}\big(\tfrac{\alpha}{N} s \sum_{i,j=1}^N < m^{(N)}, \overrightarrow{e}_i^{t,t+s} f \circ \sigma_{ij} - \overrightarrow{e}_i^{t,t+s} f >\big). \end{aligned} \tag{5.2.18}$$

But $\tilde{e} \in D_E[0, \infty)$ is right-continuous (has a right limit), and hence so is $\tilde{e}_i$ for any $i \leq N$. Therefore for any $x \in I^N$

$$\inf_{t \leq r \leq t+s} \tilde{e}_i(r) \leq \overrightarrow{e}_i^{t,t+s}(x) \leq \sup_{t \leq r \leq t+s} \tilde{e}_i(r), \tag{5.2.19}$$

and hence

$$\overrightarrow{e}_i^{t,t+s}(x) \to \tilde{e}_i(t)(x) \tag{5.2.20}$$

as $s \to 0$, and furthermore $\overrightarrow{e}_i^{t,t+s} \to \tilde{e}_i(t)$ in the sup-norm topology. This concludes that the generator process for the selection, $\tilde{\mathcal{G}}_{\tilde{e}}^{sel,N} : \mathbb{R}_+ \times \tilde{\mathfrak{F}}_N \to \tilde{\mathfrak{F}}_N$, is given by

$$\tilde{\mathcal{G}}_{\tilde{e}}^{sel,N}(t)\tilde{\Phi}_N^f(m) = \frac{\alpha}{N}\sum_{i,j=1}^{N} < m^{(N)}, \tilde{e}_i(t)(f \circ \sigma_{i,j}^N - f) > . \tag{5.2.21}$$

For simplicity, similarly to the Definition 9, we denote

$$\tilde{\mathcal{G}}^{sel,N}(e^{-1}(\tilde{e}), t) := \tilde{\mathcal{G}}_{\tilde{e}}^{sel,N}(t). \tag{5.2.22}$$

Therefore, the selection generator process is the operator process $\tilde{\mathcal{G}}^{sel,N} = (\tilde{\mathcal{G}}_t^{sel,N})_{t\geq 0}$ or

$$\tilde{\mathcal{G}}^{sel,N} : \Omega \times \mathbb{R}_+ \times \tilde{\mathfrak{F}}_N \to \tilde{\mathfrak{F}}_N \tag{5.2.23}$$

that $\tilde{\mathcal{G}}^{sel,N}(\omega, t)$ is a linear operator from $\tilde{\mathfrak{F}}_N$ to $\tilde{\mathfrak{F}}_N$, defined by

$$\tilde{\mathcal{G}}^{sel,N}(\omega, t) = \tilde{\mathcal{G}}^{sel,N}(e^{-1}(\tilde{e}), t) \tag{5.2.24}$$

for any $\omega \in e^{-1}(\tilde{e})$ and any $\tilde{e}$ in the range of e. Note that the value of $\tilde{\mathcal{G}}^{sel,N}(\omega, t)$ for $\omega \in e^{-1}(\tilde{e})$ with $\tilde{e}$ out of range of e is not important and, actually, it can be any linear operator on $\tilde{\mathfrak{F}}_N$. Equivalently,

$$\tilde{\mathcal{G}}^{sel,N}(\omega, t)\tilde{\Phi}_N^f(m) = \frac{\alpha}{N}\sum_{i,j=1}^{N} < m^{(N)}, e_i(\omega, t)(f \circ \sigma_{i,j}^N - f) >, \tag{5.2.25}$$

where

$$e_i(\omega, t)(x) := e(\omega, t, x_i) = \tilde{\mathcal{G}}^{sel,N}(e^{-1}(\tilde{e}), t). \tag{5.2.26}$$

Note that, in order to ensure $\tilde{\mathcal{G}}_{\tilde{e}}^{sel,N}(t) : \tilde{\mathfrak{F}}_N \to \tilde{\mathfrak{F}}_N$, $\tilde{e}(t)$ must be in $\mathcal{C}(I, [0, 1])$. In fact we have assumed more, i.e. $\tilde{e} \in D_E[0, \infty)$ (recall $E \subset \mathcal{C}(I, [0, 1])$). The (quenched) linear generator of the Moran process with a deterministic fitness process $\tilde{e} \in D_E[0, \infty)$, namely $\tilde{\mathcal{G}}_{\tilde{e}} : \mathbb{R}_+ \times \tilde{\mathfrak{F}}_N \to \tilde{\mathfrak{F}}_N$, is given by

$$\tilde{\mathcal{G}}_{\tilde{e}}^N := \tilde{\mathcal{G}}^{res,N} + \tilde{\mathcal{G}}^{mut,N} + \tilde{\mathcal{G}}_{\tilde{e}}^{sel,N}. \tag{5.2.27}$$

Proposition 5.2.4. *Let $\tilde{\boldsymbol{P}}_0^N : D_E[0,\infty) \to \mathcal{P}(\mathcal{P}^N(I))$ be measurable and $\tilde{\mathcal{G}}^N$ be as defined above. The $(\tilde{\mathcal{G}}^N, \tilde{\mathfrak{F}}_N, \tilde{\boldsymbol{P}}_0^N, \mathbb{P}^{env})$-martingale problem is well-posed, and μ_N^e is identified as the solution of this martingale problem.*

Proof: The existence has been shown by construction. Note that the constructed quenched solutions $\mu_N^{\tilde{e}}$, for every $\tilde{e} \in D_E[0,\infty)$, is also Markov. It suffices to prove that, for any $\tilde{e} \in D_E[0,\infty)$, the time-inhomogeneous martingale problem $(\tilde{\mathcal{G}}_{\tilde{e}}^N, \tilde{\mathfrak{F}}_N, \tilde{\mathbf{P}}_0^N(\tilde{e}))$ is well-posed . But this follows from Lemma 5.1.6 and the fact that the algebra $\tilde{\mathfrak{F}}_N$ separates points. ■

Remark 5.2.5. *The latter proposition can also be proved using the duality method in the same way that we show the uniqueness of the quenched martingale problem for Fleming-Viot process in r.e.. See sections 5.3 and 5.4.*

5.2.2 Fleming-Viot process in random envirnoments

Following Fleming and Viot (1979)[27], one can define a similar process as the Moran process on countable number of individuals. In fact we can project the classic particle Moran process with N individuals and type space I on a measure-valued process, by looking at the frequency of types of individuals during the time. This gives rise to a measure-valued Moran process. Identifying this process as a solution of a well-posed martingale problem, we can see that the FV process arises as the weak limit (in $D_{\mathcal{P}(I)}[0,\infty)$) of $\mathcal{P}(I)$-valued Moran process with N individuals as $N \to \infty$. Similarly, for the Moran process in r.e., we can define the FV process in r.e. $(FVRE)$ as the weak limit of MRE process, denoted by μ_N^e, in Skorokhod topology. A more complicated problem is when the sequence of fitness processes e_N is converging weakly in $D_E[0,\infty)$ to a fitness process e and then $FVRE$ with the fitness process e, namely μ^e, arises as the weak limit of MRE processes $\mu_N^{e_N}$ in $D_{\mathcal{P}(I)}[0,\infty)$. The first step to prove this kind of theorems is to study the FVRE martingale problem. Here we set up the quenched martingale problem for FVRE.

Recall that by Tikhonov's theorem $I^{\mathbb{N}}$ is compact. Let $\mathcal{B}_n(I^{\mathbb{N}})$, $\bar{\mathcal{C}}_n(I^{\mathbb{N}}) = \mathcal{C}_n(I^{\mathbb{N}})$, and $\bar{\mathcal{C}}_n^k(I^{\mathbb{N}}) = \mathcal{C}_n^k(I^{\mathbb{N}})$ be the subsets of $\mathcal{B}(I^{\mathbb{N}})$, $\mathcal{C}(I^{\mathbb{N}})$, and $\mathcal{C}^k(I^{\mathbb{N}})$, respectively, depending on the first n variables of $I^{\mathbb{N}}$.

Definition 13. *For $f \in \mathcal{B}_n(I^{\mathbb{N}})$, a polynomial is a function*

$$\tilde{\Phi}^f = \tilde{\Phi} : \mathcal{P}(I) \to \mathbb{R} \tag{5.2.28}$$

defined by

$$\tilde{\Phi}^f(m) :=< m^{\otimes\mathbb{N}}, f > \quad for\ m \in \mathcal{P}(I), \tag{5.2.29}$$

where $m^{\otimes\mathbb{N}}$ is the $\mathbb{N}$-fold product measure of m. The smallest number n for which (5.2.29) holds is called the degree of $\tilde{\Phi}^f$.

Let

$$\tilde{\mathfrak{F}}^k = \{\tilde{\Phi}^f : f \in \bar{\mathcal{C}}_n^k(I^{\mathbb{N}})\ for\ some\ n \in \mathbb{N}\} \tag{5.2.30}$$

for $k \in \mathbb{N} \cup \{0, \infty\}$, and let $\tilde{\mathfrak{F}} = \tilde{\mathfrak{F}}^0$.

Proposition 5.2.6. *For $k \in \mathbb{N} \cup \{0, \infty\}$, $\tilde{\mathfrak{F}}^k$ is an algebra of functions that separates points and vanishes nowhere, therefore it is measure and convergence-determining.*

Proof: First, we show that $\tilde{\mathfrak{F}}^k$ is an algebra of function for every $k \in \mathbb{N} \cup \{0, \infty\}$. Let $\Phi^f, \Phi^g \in \tilde{\mathfrak{F}}^k$ with degree n_1 and n_2 respectively. Let $n := max\{n_1, n_2\}$. Then for any $m \in \mathcal{P}(I)$, we have

$$\Phi^f(m) + \Phi^g(m) =< m^{\otimes\mathbb{N}}, f > + < m^{\otimes\mathbb{N}}, g >=< m^{\otimes\mathbb{N}}, f + g >= \Phi^{f+g} \in \tilde{\mathfrak{F}}^k \tag{5.2.31}$$

as $f + g \in C_n(I^{\mathbb{N}})$. Also for $f \in \mathcal{C}_{n_1}(I^{\mathbb{N}})$ and $g \in \mathcal{C}_{n_2}(I^{\mathbb{N}})$

$$\Phi^f(m)\Phi^g(m) =< m^{\otimes\mathbb{N}}, f >< m^{\otimes\mathbb{N}}, g >=< m^{\otimes\mathbb{N}}, f. < m^{\otimes\mathbb{N}}, g >>=$$

$$< m^{\otimes \mathbb{N}}, f.(g \circ \tau_{n_2}) >= \Phi^{f.(g \circ \tau_{n_2})} \in \tilde{\mathfrak{F}}^k, \quad (5.2.32)$$

where $\tau_k : I^{\mathbb{N}} \to I^{\mathbb{N}}$ is the translation operator on $I^{\mathbb{N}}$ defined by $\tau_r(a_1, a_2, ...) = (a_{r+1}, a_{r+2}, ...)$ for $r \in \mathbb{N}$. This proves that $\tilde{\mathfrak{F}}^k$ is an algebra for any $k \in \mathbb{N}$. To prove the second part, first note that as $\mathcal{P}(I)$ is a compact space, then being convergence-determining and measure-determining are equivalent for $\tilde{\mathfrak{F}}^k \subset C(\mathcal{P}(I))$, by Corollary 5.1.3. Thus, by proposition 5.1.4, it suffices to show that $\tilde{\mathfrak{F}}^k \subset C(\mathcal{P}(I))$ separates points. But the latter follows from the fact that for any $m_1, m_2 \in \mathcal{P}(I)$, there exists $f \in \mathcal{C}_1^k(I^{\mathbb{N}})$, for $k \in \mathbb{N} \cup \{0, \infty\}$, such that

$$\Phi^f(m_1) =< m_1, f > \neq < m_2, f >= \Phi^f(m_2). \quad (5.2.33)$$

Also $1 \in \mathcal{C}_0^k(I^{\mathbb{N}})$, and for any $m \in \mathcal{P}(I)$, we have $1 =< m, 1 > \neq 0$. This proves the proposition. ■

Remark 5.2.7. *Clearly, $\tilde{\mathfrak{F}} \subset \mathcal{C}(\mathcal{P}(I))$, therefore it is dense in $\mathcal{C}(\mathcal{P}(I))$ in the topology of uniform convergence on compact sets.*

We are ready to define the generator of FVRE and state the quenched martingale problem in r.e. for it. For $n \in \mathbb{N}$ and $f \in \mathcal{C}_n(I^{\mathbb{N}})$, let $\tilde{\Phi}^f$ be a polynomial. The generator of the FVRE with a fitness process e is the operator process

$$\tilde{\mathcal{G}}^e : \Omega \times \mathbb{R}_+ \times \tilde{\mathfrak{F}} \to \tilde{\mathfrak{F}} \quad (5.2.34)$$

also denoted by $\tilde{\mathcal{G}} = (\tilde{\mathcal{G}}_t)_{t \geq 0}$, defined as

$$\tilde{\mathcal{G}}^e = \tilde{\mathcal{G}}^{res,e} + \tilde{\mathcal{G}}^{mut,e} + \tilde{\mathcal{G}}^{sel,e}, \quad (5.2.35)$$

where the first and the second terms on the right hand side are the linear operators corresponding to resampling and mutation (generators) from $\tilde{\mathfrak{F}}$ to $\tilde{\mathfrak{F}}$, and the third one is an operator process serving as the selection generator. Usually, we drop the superscript e, when there is no risk of confusion. For $\tilde{\Phi}^f \in \tilde{\mathfrak{F}}$, $m \in \mathcal{P}(I)$ and $x = (x_1, x_2, ...)$, the operator process is defined as follows.

The resampling generator is defined by

$$\tilde{\mathcal{G}}^{res}\tilde{\Phi}^f(m) = \frac{\gamma}{2}\sum_{i,j=1}^{n} < m^{\otimes\mathbb{N}}, f \circ \sigma_{i,j} - f > . \tag{5.2.36}$$

For mutation, put

$$\begin{aligned}\tilde{\mathcal{G}}^{mut}\tilde{\Phi}^f(m) &= \beta\sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, B_i f - f > \\ &= \beta'\sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, B_i^{'} f - f > +\beta''\sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, B_i^{''} f - f >\end{aligned} \tag{5.2.37}$$

Recall that

$$B_i f(x) = \int_I f \circ \sigma_i^y(x) q(x_i, dy) \tag{5.2.38}$$

$$B_i' f(x) = \int_I f \circ \sigma_i^y(x) q'(dy) \tag{5.2.39}$$

$$B_i'' f(x) = \int_I f \circ \sigma_i^y(x) q''(x_i, dy). \tag{5.2.40}$$

For the selection generator define the following operator process

$$\tilde{\mathcal{G}}^{sel} : \Omega \times \mathbb{R}_+ \times \tilde{\mathfrak{F}} \to \tilde{\mathfrak{F}} \tag{5.2.41}$$

consistent with e such that $\tilde{\mathcal{G}}^{sel}(\omega, t)$ is defined to be a linear operator from $\tilde{\mathfrak{F}}$ to $\tilde{\mathfrak{F}}$ as

$$\tilde{\mathcal{G}}^{sel}(\omega, t)\tilde{\Phi}^f(m) = \alpha\sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, e_i(\omega, t) f - e_{n+1}(\omega, t) f > . \tag{5.2.42}$$

Recall that $e_i(\omega, t)(x) = e(\omega, t, x_i)$ and $\tilde{e}_i(t)(x) = \tilde{e}(t)(x_i)$ for a given trajectory $\tilde{e} \in D_E[0, \infty)$. Then as denoted in Definition 9, we have

$$\tilde{\mathcal{G}}^{sel}_{\tilde{e}}(t)\tilde{\Phi}^f(m) = \tilde{\mathcal{G}}^{sel}(e^{-1}(\tilde{e}), t)\tilde{\Phi}^f(m) = \alpha\sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, \tilde{e}_i(t) f - \tilde{e}_{n+1}(t) f > . \tag{5.2.43}$$

Also, for a given trajectory $\tilde{e} \in D_E[0,\infty)$, let $\tilde{\mathcal{G}}_{\tilde{e}} : \mathbb{R}_+ \times \tilde{\mathfrak{F}} \to \tilde{\mathfrak{F}}$ be defined by

$$\tilde{\mathcal{G}}_{\tilde{e}} := \tilde{\mathcal{G}}^{res} + \tilde{\mathcal{G}}^{mut} + \tilde{\mathcal{G}}^{sel}_{\tilde{e}}. \tag{5.2.44}$$

The following theorems state the wellposedness of FVRE martingale problem and identify the limit of the measure-valued Moran processes in r.e. as the unique solution of a quenched martingale problem in r.e..

Theorem 5.2.8. *Let e be a fitness process and let $\tilde{\boldsymbol{P}}_0 : D_E[0,\infty) \to \mathcal{P}(\mathcal{P}(I))$ be measurable, and $\mathbb{P}^{env,e} \in \mathcal{P}(D_E[0.\infty))$. The $(\tilde{\mathcal{G}}^e, \tilde{\mathfrak{F}}, \tilde{\boldsymbol{P}}_0, \mathbb{P}^{env,e})$-martingale problem is well-posed. Furthermore, the unique solution is a strong Markov process.*

Definition 14. *Let e be an E-valued stochastic process with sample paths in $D_E[0,\infty)$ and law $\mathbb{P}^{env}$. The unique $\mathcal{P}(I)$-valued process which is the solution of the martingale problem $(\tilde{\mathcal{G}}^e, \tilde{\mathfrak{F}}, \tilde{\boldsymbol{P}}_0, \mathbb{P}^{env,e})$, denoted by μ^e, is called Fleming-Viot process in r.e. e (FVRE). When there is no risk of ambiguity, we drop e from the superscripts. For a given trajectory of e, namely $\tilde{e}$ picked by $\mathbb{P}^{env}$, $\mu^{\tilde{e}}$ represents the quenched FV process with the deterministic (fixed) environment $\tilde{e}$.*

Recall that we frequently denote by e, e_N stochastic fitness processes, and by $\tilde{e}$ a fixed time-dependent fitness function (an element of $D_E[0,\infty)$). Also note that measurable functions $\tilde{\mathbf{P}}_0 : D_E[0,\infty) \to \mathcal{P}(\mathcal{P}(I))$ and $\tilde{\mathbf{P}}_0^N : D_E[0,\infty) \to \mathcal{P}(\mathcal{P}^N(I))$, for $N \in \mathbb{N}$, are initial distributions of FVRE and MRE, respectively. Also, in the following, we assume

$$\begin{aligned} \mathbb{P}^{env} &= \mathbb{P}^{env,e} \in \mathcal{P}(D_E[0,\infty)) \\ \mathbb{P}^{env,N} &= \mathbb{P}^{env,e_N} \in \mathcal{P}(D_E[0,\infty)) \end{aligned} \tag{5.2.45}$$

are the laws of fitness processes e and e_N, for $N \in \mathbb{N}$, respectively. We usually use the environment e for FVRE and e_N for MRE with N individuals and assume e_N converges to e in Skororkhod topology. In particular, let $\mu_N^{e_N}$ be the unique solution to the quenched martingale $(\tilde{\mathcal{G}}^{N,e_N}, \tilde{\mathcal{F}}_N, \tilde{\mathbf{P}}_0^N, \mathbb{P}^{env,N}) = (\tilde{\mathcal{G}}^N, \tilde{\mathcal{F}}_N, \tilde{\mathbf{P}}_0^N, \mathbb{P}^{env,N})$ and μ^e be the unique solution to $(\tilde{\mathcal{G}}^e, \tilde{\mathcal{F}}, \tilde{\mathbf{P}}_0, \mathbb{P}^{env}) = (\tilde{\mathcal{G}}, \tilde{\mathcal{F}}, \tilde{\mathbf{P}}_0, \mathbb{P}^{env})$.

Theorem 5.2.9. *Let e be a fitness process and $\boldsymbol{P}$ be the law of μ^e with the family of quenched measures $\{\boldsymbol{P}^{\tilde{e}}\}_{\tilde{e}\in D_E[0,\infty)}$ for $\{\mu^{\tilde{e}}\}_{\tilde{e}\in D_E[0,\infty)}$. Then for $\mathbb{P}^{env}$-a.e. $\tilde{e}\in D_E[0,\infty)$, the process $\mu^{\tilde{e}}$ has continuous sample paths (in $\mathcal{C}([0,\infty),\mathcal{P}(I))$) $\boldsymbol{P}^{\tilde{e}}$-a.s., that is*

$$\mathbb{P}^{env}(\tilde{e}\in D_E[0,\infty):\boldsymbol{P}^{\tilde{e}}(t\mapsto\mu^{\tilde{e}}(t)\text{ is continuous})=1)=1. \tag{5.2.46}$$

Therefore,

$$\boldsymbol{P}(t\mapsto\mu^{e}(t)\text{ is continuous})=1. \tag{5.2.47}$$

Theorem 5.2.10. *Suppose $\tilde{\mathbf{P}}_0$ and $\tilde{\mathbf{P}}_0^N$ are continuous, for any $N\in\mathbb{N}$.*

(i) Let $\tilde{e}_N,\tilde{e}\in D_E[0,\infty)$, for $N\in\mathbb{N}$, such that $\tilde{e}_N\to\tilde{e}$ in $D_E[0,\infty)$. Then

a) If $\tilde{\boldsymbol{P}}_0^N(\tilde{e}_N)\to\tilde{\boldsymbol{P}}_0(\tilde{e})$ in $\mathcal{P}(\mathcal{P}(I))$, as $N\to\infty$, then $\mu_N^{\tilde{e}_N}\Rightarrow\mu^{\tilde{e}}$ in $D_{\mathcal{P}(I)}[0,\infty)$, as $N\to\infty$.

b) For $M\in\mathbb{N}$,

$$\mu_M^{\tilde{e}_N}\Rightarrow\mu_M^{\tilde{e}}\quad\text{as}\quad N\to\infty \tag{5.2.48}$$

in $D_{\mathcal{P}^M(I)}[0,\infty)$.

c) $\mu^{\tilde{e}_N}\Rightarrow\mu^{\tilde{e}}$ in $D_{\mathcal{P}(I)}[0,\infty)$, as $N\to\infty$.

(ii) Let e and $\{e_N\}_{N\in\mathbb{N}}$ be fitness processes (not necessarily Markov) such that $e_N\to e$ in $D_E[0,\infty)$ a.s., as $N\to\infty$. Then

a) If $\tilde{\boldsymbol{P}}_0^N(e_N)\to\tilde{\boldsymbol{P}}_0(e)$ in $\mathcal{P}(\mathcal{P}(I))$, as $N\to\infty$, a.s. then

$$\mu_N^{e_N}\Rightarrow\mu^{e} \tag{5.2.49}$$

in $D_{\mathcal{P}(I)}[0,\infty)$, as $N\to\infty$.

b) For $M\in\mathbb{N}$,

$$\mu_M^{e_N}\Rightarrow\mu_M^{e} \tag{5.2.50}$$

in $D_{\mathcal{P}^M(I)}[0,\infty)$, as $N\to\infty$.

c) $\mu^{e_N} \Rightarrow \mu^e$ *in* $D_{\mathcal{P}(I)}[0,\infty)$*, as* $N \to \infty$*.*

Remark 5.2.11. *We summarize the convergence theorem in the following diagram. If* $e_M \to e$*, as* $M \to \infty$*, a.s. in* $D_E[0,\infty)$*, then*

$$\begin{array}{ccc} \mu_N^{e_M} & \Rightarrow & \mu^{e_M} \\ \Downarrow & \searrow & \Downarrow \\ \mu_N^e & \Rightarrow & \mu^e \end{array} \tag{5.2.51}$$

as $M \to \infty$ *and* $N \to \infty$*, appropriately.*

Definition 15. *We say an* S*-valued Markov process* Z *is weakly ergodic if there exists* $m \in \mathcal{P}(S)$ *such that for every initial distribution of* Z

$$\lim_{t\to\infty} \int_\Omega f(Z_t(\omega))d\mathbb{P} =< m, f >, \qquad f \in \bar{\mathcal{C}}(S). \tag{5.2.52}$$

In other words, letting $\{T_t^Z\}$ *be the semigroup of* Z *on* $\bar{\mathcal{C}}(S)$*, there exists* $m \in \mathcal{P}(S)$ *such that*

$$\lim_{t\to\infty} T_t^Z f(x) =< m, f > \tag{5.2.53}$$

for any $x \in S$ *and* $f \in \bar{\mathcal{C}}(S)$*.*

Theorem 5.2.12. *Suppose there exists a parent-independent component in the mutation process, i.e.* $\beta' > 0$*, and let* e *either be a stationary fitness process (not necessarily Markov) or a weakly ergodic Markov fitness with semigroup* $\{T_t^{env}\}$ *such that* $T_t^{env} : \mathcal{C}(E) \to \mathcal{C}(E)$ *for any* $t \geq 0$ *. Then the following statement holds.*

(i) There exists a $\mathcal{P}(I)$*-valued random variable* $\mu^e(\infty)$ *such that*

$$\mu^e(t) \Rightarrow \mu^e(\infty) \tag{5.2.54}$$

as $t \to \infty$*, in* $\mathcal{P}(I)$*.*

(ii) By assumption on weak ergodicity of e*, there exists an* E*-valued random variable* $e(\infty)$ *such that the annealed-environment process converges weakly, that is*

$$(\mu^e(t), e(t)) \Rightarrow (\mu^e(\infty), e(\infty)) \tag{5.2.55}$$

as $t \to \infty$*, in* $\mathcal{P}(I) \times E$*, and the law of* $(\mu^e(\infty), e(\infty))$ *is the unique invariant distribution of* $(\mu^e(t), e(t))_{t \geq 0}$*.*

The strategy to prove these theorems for the annealed processes μ^e and $\mu_N^{e_N}$ is to prove them for quenched processes (processes with fixed environments), first, and then integrating over the elements of $D_E[0, \infty)$ we get the result for the annealed process. As each quenched (fixed) environment is a deterministic process and thus Markov, one can characterize the quenched processes as a quenched martingale problem in r.e. regardless of having Markovian property for the environments. This is one important advantage of this method. The technique that we apply is the combination of martingale problem and duality method. As the fitness process and hence the quenched generators depend on time, the dual process also must do so. Therefore, we need to understand the behaviour of the time-dependent dual process. The next section prepares some generalities about dual processes for time-inhomogeneous Markov processes.

5.3 Duality method for stochastic processes in random environments

The importance of uniqueness of solutions for martingale problems combined with relative hardness to prove it, has led many researchers to think of new methods and tools to deal with this issue. One of the most important methods which has been used by several authors is the method of duality. It has been used by many researchers including Holley and Liggett (1975)[31] who applied it in order to prove ergodic theorems for the voter model, by Shiga (1980-1981) [54, 55] who used it for the diffusions arising in population genetics, and by Dawson and Hochberg (1982)[11] who, for the first time, introduced function-valued duals for measure-valued processes. Many other dual processes have been introduced, later (see [8, 10]).

One of the goals of the duality method for martingale problems is to transform the problem of uniqueness for one martingale problem to the problem of existence for another one which is a dual of the first. Furthermore, in many examples, as the dual process is a relatively simpler process, it is much easier to understand its behaviour. This gives more information about the main process which is harder to deal with directly. One example is to prove ergodic theorems for the solution of a well-posed martingale problem when the dual of the main process reaches special absorbing states with probability one. Here we extend the method of duality for stochastic processes in random environments, and generalize the notion of time-dependent Feynman-Kac duals, namely we study general duals in which an exponential term appears. In applications, we usually cannot avoid appearance of Feynman-Kac term in our dual that makes the dual process harder to analyse. However, in the case of Fleming-Viot process when the fitness process is bounded, one can give duals in which there is no exponential term. In fact, when the fitness function (process) is unbounded, the existence of Feynman-Kac term is unavoidable.

In this section we assume that S' and S_1 are Polish metric spaces, S_2 is a separable metric space, and $\mathbb{P}^{env}$ is the law of the environment $X : \Omega \to S'$ which is an S'-valued random variable. Let $\mathcal{D}_i \subset \mathcal{B}(S_i)$ for $i = 1, 2$. We assume $\mathcal{G}^{i,t} : \mathbb{R}_+ \times \mathcal{D}_i \to \mathcal{B}(S_i)$ to be time-dependent linear operators, for $i = 1, 2$, and let $m_0^i \in \mathcal{P}(S_i)$. Also, we assume, for $i = 1, 2$ and for any real number $t \geq 0$, $G^{i,t} : \Omega \times \mathbb{R}_+ \times \mathcal{D}_i \to \mathcal{B}(S_i)$ are operator processes with domain $\mathcal{D}_i$, and $\mathbf{P}_0^i : S' \to \mathcal{P}(S_i)$ are measurable. Let

$f \in \mathcal{B}(S_1 \times S_2)$ be such that $f(.,v) \in \mathcal{D}_1$ and $f(u,.) \in \mathcal{D}_2$ for any $v \in S_2$ and $u \in S_1$. Let $g : \mathbb{R}_+ \times S_2 \to \mathbb{R}$ be a Borel measurable function. We start with the definition of duality for two families of time-inhomogeneous problems.

Two families of time-dependent martingale problems $\mathscr{G}^{*1} = \{(\mathcal{G}^{1,t}, \mathcal{D}_1, m_0^1)\}_{t\in\mathbb{R}_+}$ and $\mathscr{G}^{*2} = \{(\mathcal{G}^{2,t}, \mathcal{D}_2, m_0^2)\}_{t\in\mathbb{R}_+}$ are said to be dual with respect to (f, g), if for each family of solutions $\{\zeta^t\}_{t\in\mathbb{R}_+}$ to the martingale problem $\mathscr{G}^{*1}$, with respective laws $\{m^{1,t}\}_{t\in\mathbb{R}_+}$, and each family of solutions $\{\xi^t\}_{t\in\mathbb{R}_+}$ to the martingale problem $\mathscr{G}^{*2}$, with respective laws $\{m^{2,t}\}_{t\in\mathbb{R}_+}$, we have

$$\int\limits_{S_1} \mathbb{E}_{m^{2,t}}[|f(u,\xi^t(t))| \exp\{\int\limits_0^t g(s,\xi^t(s))ds\}] m_0^1(du)] < \infty \tag{5.3.1}$$

for any $t \in \mathbb{R}_+$, and

$$\int\limits_{S_2} \mathbb{E}_{m^{1,t}}[f(\zeta^t(t),v)] m_0^2(dv) = \int\limits_{S_1} \mathbb{E}_{m^{2,t}}[f(u,\xi^t(t)) \exp\{\int\limits_0^t g(s,\xi^t(s))ds\}] m_0^1(du). \tag{5.3.2}$$

We extend this idea to two families of quenched martingale problems in random environment. Let $M^i(x) \subset \mathcal{P}(S_i)$ be a collection of measures on S_i, for any $x \in S'$ and $i = 1, 2$. Set

$$M^i := \{\mathbf{P}_0^i \in \mathscr{B}(S', \mathcal{P}(S_i)) : \forall x \in S',\ \mathbf{P}_0^i(x) \in M^i(x)\}, \tag{5.3.3}$$

where $\mathscr{B}(S', \mathcal{P}(S_i))$ is the set of all Borel measurable functions from S' to $\mathcal{P}(S_i)$ for $i = 1, 2$.

Two families of quenched martingale problems in r.e. X, namely

$$\mathscr{G}^1 = \mathscr{G}^1(M^1) := \{(G^{1,t}, \mathcal{D}_1, \mathbf{P}_0, \mathbb{P}^{env})\}_{(t,\mathbf{P}_0)\in\mathbb{R}_+\times M^1}$$

and

$$\mathscr{G}^2 = \mathscr{G}^2(M^2) := \{(G^{2,t}, \mathcal{D}_2, \mathbf{Q}_0, \mathbb{P}^{env})\}_{(t,\mathbf{Q}_0)\in\mathbb{R}_+\times M^2},$$

are said to be (strongly) dual with respect to (f, g), if for each family of solutions $\{\zeta^{t,\mathbf{P}_0,X}\}_{(t,\mathbf{P}_0)\in\mathbb{R}_+\times M^1}$ to $\mathscr{G}^1$, where each solution $\zeta^{t,\mathbf{P}_0,X}$ has the family of quenched laws $\{\mathbf{P}^{t,\mathbf{P}_0,x}\}_{x\in S'}$, and for each family of solutions $\{\xi^{t,\mathbf{Q}_0,X}\}_{(t,\mathbf{Q}_0)\in\mathbb{R}_+\times M^2}$ to $\mathscr{G}^2$, where each solution $\xi^{t,\mathbf{Q}_0,X}$ has the family of quenched laws $\{\mathbf{Q}^{t,\mathbf{Q}_0,x}\}_{x\in S'}$, we have:

$$\mathbb{E}_{\mathbb{P}^{env}}[\int_{S_1} \mathbb{E}_{\mathbf{Q}^{t,\mathbf{Q}_0,X}}[|f(u,\xi^{t,\mathbf{Q}_0,X}(t))| \exp\{\int_0^t g(s,\xi^{t,\mathbf{Q}_0,X}(s))ds\}]\mathbf{P}_0(X)(du)] < \infty \tag{5.3.4}$$

for any $t \in \mathbb{R}_+$, $\mathbf{P}_0 \in M^1$, $\mathbf{Q}_0 \in M^2$ (Recall $\mathbf{P}^{t,\mathbf{P}_0,X} : \omega \mapsto \mathbf{P}^{t,\mathbf{P}_0,X(\omega)}$ and $\mathbf{Q}^{t,\mathbf{Q}_0,X} : \omega \mapsto \mathbf{Q}^{t,\mathbf{Q}_0,X(\omega)}$), and for $\mathbb{P}^{env}$-a.e. x

$$\int_{S_2} \mathbb{E}_{\mathbf{P}^{t,\mathbf{P}_0,x}}[f(\zeta^{t,\mathbf{P}_0,x}(t),v)]\mathbf{Q}_0(x)(dv) =$$

$$\int_{S_1} \mathbb{E}_{\mathbf{Q}^{t,\mathbf{Q}_0,x}}[f(u,\xi^{t,\mathbf{Q}_0,x}(t)) \exp\{\int_0^t g(s,\xi^{t,\mathbf{Q}_0,x}(s))ds\}]\mathbf{P}_0(x)(du) \tag{5.3.5}$$

for any $t \in \mathbb{R}_+$, $\mathbf{P}_0 \in M^1$ and $\mathbf{Q}_0 \in M^2$.

We say they are dual in average if for any $t \geq 0$, $\mathbf{P}_0 \in M^1$ and $\mathbf{Q}_0 \in M^2$ (5.3.4) holds, and

$$\mathbb{E}_{\mathbb{P}^{env}}[\int_{S_2} \mathbb{E}_{\mathbf{P}^{t,\mathbf{P}_0,X}}[f(\zeta^{t,\mathbf{P}_0,X}(t),v)]\mathbf{Q}_0(X)(dv)] =$$

$$\mathbb{E}_{\mathbb{P}^{env}}[\int_{S_1} \mathbb{E}_{\mathbf{Q}^{t,\mathbf{Q}_0,X}}[f(u,\xi^{t,\mathbf{Q}_0,X}(t)) \exp\{\int_0^t g(s,\xi^{t,\mathbf{Q}_0,X}(s))ds\}]\mathbf{P}_0(X)(du)] \tag{5.3.6}$$

for any $t \in \mathbb{R}_+$, $\mathbf{P}_0 \in M^1$ and $\mathbf{Q}_0 \in M^2$.

Remark 5.3.1. *For $i = 1,2$, $x \in S'$ and $t \in \mathbb{R}_+$, recall that $G_x^{i,t} : \mathbb{R}_+ \times \mathcal{D}_i \to \mathcal{B}(S_i)$ is defined by*

$$G_x^{i,t}(s,h) := G^{i,t}(X^{-1}(x), s, h). \tag{5.3.7}$$

For $i = 1,2$, $x \in S'$ and $m_0^i \in \mathcal{P}(S_i)$, let $\mathscr{G}_x^{*i}(m_0^i)$ be the family of martingale problems $\{(G_x^{i,t}, \mathcal{D}_i, m_0^i)\}_{t\in\mathbb{R}_+}$. In fact $\mathscr{G}^1(M^1)$ and $\mathscr{G}^2(M^2)$ are dual if and only if for $\mathbb{P}^{env}$-a.e. $x \in S'$, $\mathscr{G}_x^{*1}(m_0^1)$ and $\mathscr{G}_x^{*2}(m_0^2)$ are dual for any $m_0^1 \in M^1(x)$ and $m_0^2 \in M^2(x)$.

When there exist a time-dependent operator $\mathcal{G}^1 : \mathbb{R}_+ \times \mathcal{D}_1 \to \mathcal{B}(S_1)$ and an operator process $G^1 : \Omega \times \mathbb{R}_+ \times \mathcal{D}_1 \to \mathcal{B}(S_1)$ such that for any $t \geq 0$, $\mathcal{G}^{1,t} = \mathcal{G}^1$ and $G^{1,t} = G^1$, all the martingale problems in the families $\mathscr{G}^{*1}$ and $\mathscr{G}^1$ coincide with the ones in the families $\{(\mathcal{G}^1, \mathcal{D}_1, \mathbf{P}_0)\}_{\mathbf{P}_0\in M^1}$ and $\{(G^1, \mathcal{D}_1, \mathbf{P}_0, \mathbb{P}^{env})\}_{\mathbf{P}_0\in M^1}$, respectively. Because of the importance of these special cases, we give their definitions separately as follows.

Definition 16. *Suppose $\mathcal{G}^1 : \mathbb{R}_+ \times \mathcal{D}_1 \to \mathcal{B}(S_1)$ is a time dependent linear operator, and let $m_0^i \in \mathcal{P}(S_i)$ for $i = 1,2$. The martingale problem $\mathscr{G}^{*1} = (\mathcal{G}^1, \mathcal{D}_1, m_0^1)$ and the family of martingale problems $\mathscr{G}^{*2} = \{(\mathcal{G}^{2,t}, \mathcal{D}_2, m_0^2)\}_{t\in\mathbb{R}_+}$ are said to be dual with respect to (f,g), if for each solution ζ to the martingale problem $\mathscr{G}^{*1}$, with law m^1, and each family of solutions $\{\xi^t\}_{t\in\mathbb{R}_+}$ to the martingale problem $\mathscr{G}^{*2}$, with respective laws $\{m^{2,t}\}_{t\in\mathbb{R}_+}$, (5.3.1) holds for any $t \in \mathbb{R}_+$, and*

$$\int_{S_2} \mathbb{E}_{m^1}[f(\zeta(t),v)]m_0^2(dv) = \int_{S_1} \mathbb{E}_{m^{2,t}}[f(u,\xi^t(t))\exp\{\int_0^t g(s,\xi^t(s))ds\}]m_0^1(du) \tag{5.3.8}$$

for any $t \in \mathbb{R}_+$.

Remark 5.3.2. *If, in addition, we assume that $\mathcal{G}^1 : \mathcal{D}_1 \to \mathcal{B}(S_1)$ and there exists a linear operator $\mathcal{G}^2 : \mathcal{D}_2 \to \mathcal{B}(S_2)$ such that for any $t \geq 0$, $\mathcal{G}^{2,t} = \mathcal{G}^2$, then the duality* *in Definition 16 reduces to the classic time-homogeneous duality. In this case, it is still possible to find a family of time-dependent duals (not necessarily one dual).*

For the quenched martingale problem in random environment we have:

Definition 17. *Let f and g be as defined above and $G^1 : \Omega \times \mathbb{R}_+ \times \mathcal{D}_1 \to \mathcal{B}(S_1)$ be an operator process. We say a family of quenched martingale problems*

$$\mathscr{G}^1 = \mathscr{G}^1(M^1) = \{(G^1, \mathcal{D}_1, \boldsymbol{P}_0, \mathbb{P}^{env})\}_{\boldsymbol{P}_0 \in M^1}$$

and a family of quenched martingale problems

$$\mathscr{G}^2 = \mathscr{G}^2(M^2) = \{(G^{2,t}, \mathcal{D}_2, \boldsymbol{Q}_0, \mathbb{P}^{env})\}_{(t,\boldsymbol{Q}_0) \in \mathbb{R}_+ \times M^2}$$

are (strongly) dual with respect to (f, g) if for each family of solutions $\{\zeta^{\boldsymbol{P}_0,X}\}_{\boldsymbol{P}_0 \in M^1}$ to $\mathscr{G}^1$, with the respective families of quenched laws $\{\{\boldsymbol{P}^{\boldsymbol{P}_0,x}\}_{x \in S'}\}_{\boldsymbol{P}_0 \in M^1}$, and for each family of solutions $\{\xi^{t,\boldsymbol{Q}_0,X}\}_{(t,\boldsymbol{Q}_0) \in \mathbb{R}_+ \times M^2}$ to $\mathscr{G}^2$, with respective families of quenched laws $\{\{\boldsymbol{Q}^{t,\boldsymbol{Q}_0,x}\}_{x \in S'}\}_{(t,\boldsymbol{Q}_0) \in \mathbb{R}_+ \times M^2}$, we have:
For every $t \geq 0$, $\boldsymbol{P}_0 \in M^1$ and $\boldsymbol{Q}_0 \in M^2$, (5.3.4) holds and for $\mathbb{P}^{env}$-a.e. $x \in S'$

$$\int_{S_2} \mathbb{E}_{\boldsymbol{P}^{\boldsymbol{P}_0,x}}[f(\zeta^{\boldsymbol{P}_0,x}(t), v)]\, \boldsymbol{Q}_0(x) dv =$$

$$\int_{S_1} \mathbb{E}_{\boldsymbol{Q}^{t,\boldsymbol{Q}_0,x}}[f(u, \xi^{t,\boldsymbol{Q}_0,x}(t)) \exp\{\int_0^t g(s, \xi^{t,\boldsymbol{Q}_0,x}(s)) ds\}] \boldsymbol{P}_0(x) du \quad (5.3.9)$$

for any $t \in \mathbb{R}_+$, $\boldsymbol{P}_0 \in M^1$ and $\boldsymbol{Q}_0 \in M^2$.

They are said to be dual in average if (5.3.4) holds for any $t \geq 0$, $\boldsymbol{P}_0 \in M^1$, $\boldsymbol{Q}_0 \in M^2$ and

$$\mathbb{E}_{\mathbb{P}^{env}}[\int_{S_2} \mathbb{E}_{\boldsymbol{P}^{\boldsymbol{P}_0,X}}[f(\zeta^{\boldsymbol{P}_0,X}(t), v)]\, \boldsymbol{Q}_0(X) dv] =$$

$$\mathbb{E}_{\mathbb{P}^{env}}[\int_{S_1} \mathbb{E}_{\boldsymbol{Q}^{t,\boldsymbol{Q}_0,X}}[f(u,\xi^{t,\boldsymbol{Q}_0,X}(t))\exp\{\int_0^t g(s,\xi^{t,\boldsymbol{Q}_0,X}(s))ds\}]\boldsymbol{P}_0(X)du] \quad (5.3.10)$$

for any $t \geq 0$, $\boldsymbol{P}_0 \in M^1$ *and* $\boldsymbol{Q}_0 \in M^2$.

Remark 5.3.3. *For* $i = 1, 2$, $x \in S'$, *and* $t \in \mathbb{R}_+$, *as we already defined, let* $G^1_x : \mathbb{R}_+ \times \mathcal{D}_1 \to \mathcal{B}(S_1)$ *and* $G^{2,t}_x : \mathbb{R}_+ \times \mathcal{D}_2 \to \mathcal{B}(S_2)$ *with*

$$G^1_x(s,h) := G^1(X^{-1}(x), s, h) \quad (5.3.11)$$

and

$$G^{2,t}_x(s,h) := G^{2,t}(X^{-1}(x), s, h). \quad (5.3.12)$$

For $i = 1, 2$, $x \in S'$ *and* $m^i_0 \in \mathcal{P}(S_i)$, *let* $\mathscr{G}^{*1}_x(m^1_0) := (G^1_x, \mathcal{D}_1, m^1_0)$ *and* $\mathscr{G}^{*2}_x(m^2_0) := \{(G^1_x, \mathcal{D}_1, m^2_0)\}_{t\in\mathbb{R}_+}$. *We have that* $\mathscr{G}^1(M^1)$ *and* $\mathscr{G}^2(M^2)$ *are dual if and only if for* $\mathbb{P}^{env}$*-a.e.* $x \in S'$, $\mathscr{G}^{*1}_x(m^1_0)$ *and* $\mathscr{G}^{*2}_x(m^2_0)$ *are dual for any* $m^1_0 \in M^1(x)$ *and* $m^2_0 \in M^2(x)$.

When the family of functions $\{f(.,v) : v \in S_2\}$ is sufficiently nice, in other words measure-determining, the duality relation ensures the coincidence of the one-dimensional distributions of any two solutions of the martingale problem, which itself implies the uniqueness of finite dimensional distributions of those which is equivalent to well-posedness of the martingale problem. The following proposition transforms the problem of uniqueness for a martingale problem to the problem of existence of a dual process, or in other words, to the problem of existence of a dual martingale problem. This is a generalization of Lemma 5.5.1[8] and Proposition 4.4.7 [23].

Let $\mathcal{P}_c(S_1) = \{m \in \mathcal{P}(S_1) : m \textit{ has a compact support}\}$, and recall that, for $y \in S_2$, δ_y is the delta measure with the support on $\{y\}$.

Proposition 5.3.4. *Suppose that, for any $m_0 \in \mathcal{P}_c(S_1)$ and $y \in S_2$, the time-dependent martingale problem*

$$\mathscr{G}^{*1}(m_0) = (\mathcal{G}^1, \mathcal{D}_1, m_0)$$

and the families of time-dependent martingale problems

$$\mathscr{G}^{*2}(\delta_y) = \{(\mathcal{G}^{2,t}, \mathcal{D}_2, \delta_y)\}_{t\in\mathbb{R}_+}$$

are dual with respect to (f, g). Consider a collection of measures $\mathcal{M} \subset \mathcal{P}(S_1)$ containing $\mathbb{P}\zeta(s)^{-1}$ for every $s \geq 0$ and every solution ζ of $\mathscr{G}^{*1}(m_0)$ with $m_0 \in \mathcal{P}_c(S_1)$. Suppose that $\{f(.,y) : y \in S_2\}$ is measure-determining on $\mathcal{M}$. If for every $y \in S_2$ and $t \geq 0$ the martingale problem $(\mathcal{G}^{2,t}, \mathcal{D}_2, \delta_y)$ has a solution, then for any initial distribution $m_0 \in \mathcal{P}(S_1)$, the time-dependent martingale problem $(\mathcal{G}^1, \mathcal{D}_1, m_0)$ has at most one solution (a unique solution).

Proof: For $m_0 \in \mathcal{P}_c(S_1)$, let ζ and ζ' be two solutions to $\mathscr{G}^{*1}(m_0)$, and denote by $\xi^{t,y}$ an arbitrary solution to the martingale problem $(\mathcal{G}^{2,t}, \mathcal{D}_2, \delta_y)$ for $t \geq 0$ and $y \in S_2$. By the duality relation

$$\mathbb{E}[f(\zeta(t), y)] = \int_{S_1} \mathbb{E}[f(u, \xi^{t,y}(t)) \exp\{\int_0^t g(s, \xi^{t,y}(s))ds\}]m_0^1(du) = \mathbb{E}[f(\zeta'(t), y)] \quad (5.3.13)$$

that, as $\{f(.,y)\}_{y\in S_2}$ is measure-determining on $\mathcal{M}$, implies the uniqueness of one-dimensional distributions, i.e. $\mathbb{P}\zeta(t)^{-1}$ and $\mathbb{P}\zeta'(t)^{-1}$ coincide for any $t \in \mathbb{R}_+$. Hence, $\mathbb{P}\zeta^{-1} = \mathbb{P}\zeta'^{-1}$ that means uniqueness.

For general $m_0 \in \mathcal{P}(S_1)$, let ζ and ζ' be solutions to the martingale problem $\mathscr{G}^{*1}(m_0)$, and let $K \subset S_1$ be compact with $m_0(K) > 0$. Denote by ζ_K and ζ'_K, the processes ζ and ζ' conditioned on $\{\zeta(0) \in K\}$ and $\{\zeta'(0) \in K\}$, respectively. It is clear that ζ_K and ζ'_K are solutions to the martingale problem $\mathscr{G}^{*1}(m_0(.|K))$ with

$$m_0(.|K) = \frac{m_0(. \cap K)}{m_0(K)} \in \mathcal{P}_c(S_1). \tag{5.3.14}$$

Thus, as proved above, $\mathbb{P}\zeta_K(t)^{-1} = \mathbb{P}\zeta'_K(t)^{-1}$ which means

$$\mathbb{P}(\zeta(t) \in A|\zeta(0) \in K) = \mathbb{P}(\zeta'(t) \in A|\zeta'(0) \in K) \tag{5.3.15}$$

for any Borel measurable subset A of S_1. Since S_1 is a Polish space, from regularity of m_0, there exist a sequence of compact sets $(K_n)_{n\in\mathbb{N}}$ such that $m_0(K_n) \to 1$ as $n \to \infty$. Therefore,

$$\mathbb{P}(\zeta(t) \in A) = \lim_{n\to\infty} \mathbb{P}(\zeta(t) \in A|\zeta(0) \in K_n) = \lim_{n\to\infty} \mathbb{P}(\zeta'(t) \in A|\zeta'(0) \in K_n) = \mathbb{P}(\zeta'(t) \in A) \tag{5.3.16}$$

which implies uniqueness, by Proposition 5.1.8. ■

We easily generalize the last Proposition to the case of quenched martingale problems. For every $x \in S'$, let $M_c(x) = \mathcal{P}_c(S_1)$, and let $M_\delta(x) = \{\delta_y \in \mathcal{P}(S_2) : y \in S_2\}$. We define M_c and M_δ from $M_c(x)$ and $M_\delta(x)$ as already defined.

Proposition 5.3.5. *Suppose that the families of quenched martingale problems*

$$\mathscr{G}^1(M_c) = \{(G^1, \mathcal{D}_1, \boldsymbol{P}_0, \mathbb{P}^{env})\}_{\boldsymbol{P}_0 \in M_c}$$

and

$$\mathscr{G}^2(M_\delta) = \{(G^{2,t}, \mathcal{D}_2, \boldsymbol{P}_0, \mathbb{P}^{env})\}_{(t,\boldsymbol{P}_0)\in\mathbb{R}_+\times M_\delta}$$

are dual with respect to (f, g). Consider a collection of measures $\mathcal{M} \subset \mathcal{P}(S_1)$ such that for $\mathbb{P}^{env}$-a.e. $x \in S'$ contains $\mathbb{P}\zeta^x(s)^{-1}$ for all $s \geq 0$ and all solutions ζ^x of $(G^1_x, \mathcal{D}_1, \mathbb{P}\zeta^x(0)^{-1})$ for which $\mathbb{P}\zeta^x(0)^{-1} \in \mathcal{P}_c(S_1)$. Suppose that $\{f(., y) : y \in S_2\}$ is measure-determining on $\mathcal{M}$. If for $\mathbb{P}^{env}$-a.e. $x \in S'$, for every $y \in S_2$ and $t \geq 0$

the martingale problem $(G_x^{2,t}, \mathcal{D}_2, \delta_y)$ *has a solution, then for any initial distribution function* $\boldsymbol{P}_0 : S' \to \mathcal{P}(S_1)$ *the quenched martingale problem* $(G^1, \mathcal{D}_1, \boldsymbol{P}_0, \mathbb{P}^{env})$ *has at most one solution (a unique solution).*

Proof: First note that, as mentioned in Remark 5.3.3, $\mathscr{G}^1(M_c)$ and $\mathscr{G}^2(M_\delta)$ are dual with respect to (f, g) if and only if, for $\mathbb{P}^{env}$-a.e. $x \in S'$, $\mathscr{G}_x^{*1} = (G_x^1, \mathcal{D}_1, m_0)$ and $\mathscr{G}^{*2}(\delta_y) = \{(G_x^{2,t}, \mathcal{D}_2, \delta_y)\}_{t\geq 0}$ are dual with respect to (f, g) for every $m_0 \in \mathcal{P}_c(S_1)$ and $y \in S_2$. For any initial distribution function $\mathbf{P}_0 : S' \to \mathcal{P}(S_1)$, the quenched martingale problem $(G^1, \mathcal{D}_1, \mathbf{P}_0, \mathbb{P}^{env})$ has at most one solution if and only if $(G_x^1, \mathcal{D}_1, \mathbf{P}_0(x))$ has at most one solution for $\mathbb{P}^{env}$-a.e. $x \in S'$. But the latter follows from Proposition 5.3.4 and this finishes the proof. ■

Now we try to find conditions that guarantee the duality relation between two families of martingale problems. The following proposition is a natural extension of a theorem by D. Dawson and T. Kurtz [12] to the case of time-dependent duality relations.

Proposition 5.3.6. *Let* S_1 *and* S_2 *be two metric spaces, and let* $\mathcal{D}_1 \subset \mathcal{B}(S_1)$ *and* $\mathcal{D}_2 \subset \mathcal{B}(S_2)$. *Let* $\mathcal{G}^1 : \mathbb{R}_+ \times \mathcal{D}_1 \to \mathcal{B}(S_1)$ *and* $\mathcal{G}^{2,t} : \mathbb{R}_+ \times \mathcal{D}_2 \to \mathcal{B}(S_2)$, *for* $t \geq 0$, *be time-dependent linear operators. Consider functions* $g \in \mathcal{B}(\mathbb{R}_+ \times S_2)$ *and* $f \in \mathscr{B}(S_1 \times S_2)$ *such that, for any* $u \in S_1$ *and* $v \in S_2$, $f(., v) \in \mathcal{D}_1$ *and* $f(u, .) \in \mathcal{D}_2$, *and for any* $t \geq s \geq 0$

$$\mathcal{G}^1(s)f \in \mathscr{B}(S_1 \times S_2) \tag{5.3.17}$$

and

$$\mathcal{G}^{2,t}(s)f \in \mathscr{B}(S_1 \times S_2), \tag{5.3.18}$$

where for $(u, v) \in S_1 \times S_2$

$$\mathcal{G}^1(s)f(u, v) := \mathcal{G}^1(s)f(., v)(u) \tag{5.3.19}$$

and

$$\mathcal{G}^{2,t}(s)f(u,v) := \mathcal{G}^{2,t}(s)f(u,.)(v). \tag{5.3.20}$$

Let $m_0^i \in \mathcal{P}(S_i)$, for $i = 1,2$. Let ζ^1 and $\zeta^{2,t}$ be solutions to martingale problems $(\mathcal{G}^1, \mathcal{D}_1, m_0^1)$ and $(\mathcal{G}^{2,t}, \mathcal{D}_2, m_0^2)$, for any $t \geq 0$, respectively. Assume that for any $t \geq 0$, there exists an integrable random variable C_t such that

(i)

$$\sup_{s,r\leq t} |f(\zeta^1(s), \zeta^{2,t}(r))| \leq C_t \tag{5.3.21}$$

(ii)

$$\sup_{s,r\leq t} |\mathcal{G}^1(s)f(\zeta^1(s), \zeta^{2,t}(r))| \leq C_t \tag{5.3.22}$$

(iii)

$$\sup_{s,r\leq t} |\mathcal{G}^{2,t}(r)f(\zeta^1(s), \zeta^{2,t}(r))| \leq C_t. \tag{5.3.23}$$

If, for any $t \geq 0$, for a.e. $0 \leq s \leq t$

$$\mathcal{G}^1(s)f(.,v)(u) = \mathcal{G}^{2,t}(t-s)f(u,.)(v) + g(t-s,v) \tag{5.3.24}$$

for every $u \in S_1$ and every $v \in S_2$, then

$$\mathscr{G}^{*1}(m_0^1) = (\mathcal{G}^1, \mathcal{D}_1, m_0^1)$$

and

$$\mathscr{G}^{*2}(m_0^2) = \{(\mathcal{G}^{2,t}, \mathcal{D}_2, m_0^2)\}_{t\geq 0}$$

are dual with respect to (f,g).

Proof: We assume ζ^1 and $\{\zeta^{2,t}\}_{t\geq 0}$ are independent. For $t \geq 0$ and $s,r \leq t$ define

$$F(s,r) = \mathbb{E}[f(\zeta^1(s), \zeta^{2,t}(r)). \exp(\int_0^r g(u, \zeta^{2,t}(u))du)]. \tag{5.3.25}$$

Therefore, by martingale property

$$F(s,r)-F(0,r)=\int_0^s \mathbb{E}[\mathcal{G}^1(u)f(\zeta^1(u),\zeta^{2,t}(r))\exp\{\int_0^r g(\tilde{u},\zeta^{2,t}(\tilde{u}))d\tilde{u}\}]du \tag{5.3.26}$$

Let F_1 and F_2 be partial derivatives of F. Then

$$F_1(s,r)=\mathbb{E}[\mathcal{G}^1(s)f(\zeta^1(s),\zeta^{2,t}(r))\exp\{\int_0^r g(\tilde{u},\zeta^{2,t}(\tilde{u}))d\tilde{u}\}]. \tag{5.3.27}$$

We must also compute $F(s,r)-F(s,0)$. In order to do so, applying lemma 3.1.2 in [12], for $h\geq 0$ with $r+h\leq t$, we can write

$$\begin{aligned}
&F(s,r+h)-F(s,r)=\\
&\mathbb{E}[f(\zeta^1(s),\zeta^{2,t}(r+h))(\int_r^{r+h} g(u,\zeta^{2,t}(u))\exp\{\int_0^u g(\tilde{u},\zeta^{2,t}(\tilde{u}))d\tilde{u}\}du)]+\\
&\mathbb{E}[\int_r^{r+h} \mathcal{G}^{2,t}(u)(\zeta^1(s),\zeta^{2,t}(u))du.\exp\{\int_0^r g(\tilde{u},\zeta^{2,t}(\tilde{u}))d\tilde{u}\}]=\\
&\mathbb{E}[\int_r^{r+h} f(\zeta^1(s),\zeta^{2,t}(u)).g(u,\zeta^{2,t}(u)).\exp(\int_0^u g(\tilde{u},\zeta^{2,t}(\tilde{u}))d\tilde{u})du]+\\
&\mathbb{E}[\int_r^{r+h}\int_u^{r+h} \mathcal{G}^{2,t}(v)f(\zeta^1(s),\zeta^{2,t}(v))dv.g(u,\zeta^{2,t}(u).\exp(\int_0^u g(\tilde{u},\zeta^{2,t}(\tilde{u}))d\tilde{u})du]+\\
&\mathbb{E}[\int_r^{r+h} \mathcal{G}^{2,t}(u)f(\zeta^1(s),\zeta^{2,t}(u))\exp(\int_0^u g(\tilde{u},\zeta^{2,t}(\tilde{u}))d\tilde{u})du]+\\
&\mathbb{E}[\int_r^{r+h} \mathcal{G}^{2,t}(u)f(\zeta^1(s),\zeta^{2,t}(u)).\\
&\{\exp(\int_0^r g(\tilde{u},\zeta^{2,t}(\tilde{u}))d\tilde{u})-\exp(\int_0^u g(\tilde{u},\zeta^{2,t}(\tilde{u}))d\tilde{u})\}du].
\end{aligned} \tag{5.3.28}$$

Under the assumptions above integrals exist, and the second and the forth terms in the last equation are bounded by

$$\frac{1}{2}h^2\mathbb{E}[C_t]. \tag{5.3.29}$$

Writing $F(s,r)-F(s,0)$ as

$$\sum_{i=1}^{l}(F(s,r_i)-F(s,r_{i-1})) \tag{5.3.30}$$

for $l \in \mathbb{N}$ and an increasing sequence of real numbers $0 = r_0 < r_1 < ... < r_l = r$, and letting $l \to 0$ and $max_i(r_i - r_{i-1}) \to 0$, we get

$$F(s,r) - F(s,0) = \int_0^r \mathbb{E}[\{f(\zeta^1(s), \zeta^{2,t}(u))g(u, \zeta^{2,t}(u)) + \mathcal{G}^{2,t}(u)f(\zeta^1(s), \zeta^{2,t}(u))\}. \exp\{\int_0^u g(\tilde{u}, \zeta^{2,t}(\tilde{u}))d\tilde{u}\}]du. \tag{5.3.31}$$

Thus the partial derivative F_2 exist for a.e. $r \le t$ and

$$F_2(s,r) = \mathbb{E}[\{f(\zeta^1(s), \zeta^{2,t}(r))g(r, \zeta^{2,t}(r)) + \mathcal{G}^{2,t}(r)f(\zeta^1(s), \zeta^{2,t}(r))\}. \exp\{\int_0^r g(\tilde{u}, \zeta^{2,t}(\tilde{u}))d\tilde{u}\}]. \tag{5.3.32}$$

By Lemma 3.1.1 [12]

$$\begin{aligned} &F(t,0) - F(0,t) = \\ &\int_0^t F_1(s, t-s) - F_2(s, t-s))ds = \\ &\int_0^t \{\mathbb{E}[\mathcal{G}^1(s)f(\zeta^1(s), \zeta^{2,t}(t-s)) \exp\{\int_0^{t-s} g(\tilde{u}, \zeta^{2,t}(\tilde{u}))d\tilde{u}\}] - \\ &\mathbb{E}[\{f(\zeta^1(s), \zeta^{2,t}(t-s))g(t-s, \zeta^{2,t}(t-s)) + \mathcal{G}^{2,t}(t-s)f(\zeta^1(s), \zeta^{2,t}(t-s))\}. \\ &\exp\{\int_0^{t-s} g(\tilde{u}, \zeta^{2,t}(\tilde{u}))d\tilde{u}\}]\}dr. \end{aligned} \tag{5.3.33}$$

But this vanishes for a.e. $t \ge 0$ and a.e. $0 \le s \le t$ by (5.3.24). The statement follows, since $F(t,0)$ and $F(0,t)$ are continuous for $t \in \mathbb{R}_+$.

■

The following is an automatic extension of the last proposition to the case of quenched martingale problems.

Proposition 5.3.7. *Let S_1 and S_2 be two metric spaces, and let $\mathcal{D}_1 \subset \mathcal{B}(S_1)$ and $\mathcal{D}_2 \subset \mathcal{B}(S_2)$. Let $G^1 : \Omega \times \mathbb{R}_+ \times \mathcal{D}_1 \to \mathcal{B}(S_1)$ and $G^{2,t} : \Omega \times \mathbb{R}_+ \times \mathcal{D}_2 \to \mathcal{B}(S_2)$, for*

$t \geq 0$, be operator processes. Consider functions $g \in \mathcal{B}(\mathbb{R}_+ \times S_2)$ and $f \in \mathscr{B}(S_1 \times S_2)$ such that, for any $u \in S_1$ and $v \in S_2$, $f(.,v) \in \mathcal{D}_1$ and $f(u,.) \in \mathcal{D}_2$. Let $\mathbb{P}^{env} \in \mathcal{P}(S')$, and $m_0^i \in \mathcal{P}(S_i)$, for $i = 1,2$. Suppose $\{\zeta^{1,x}\}_{x \in S'}$ and $\{\zeta^{2,t,x}\}_{x \in S'}$ are solutions to quenched martingale problems $(G^1, \mathcal{D}_1, m_0^1, \mathbb{P}^{env})$ and $(G^{2,t}, \mathcal{D}_2, m_0^2, \mathbb{P}^{env})$, for any $t \geq 0$, respectively. Assume that for $\mathbb{P}^{env}$-a.e. $x \in S'$, for any $t \geq s \geq 0$

$$G_x^1(s)f \in \mathscr{B}(S_1 \times S_2) \tag{5.3.34}$$

and

$$G_x^{2,t}(s)f \in \mathscr{B}(S_1 \times S_2), \tag{5.3.35}$$

and $\mathbb{P}^{env}$-a.s. for any $t \geq 0$, there exists an integrable random variable C_t^x such that

(i)

$$\sup_{s,r \leq t} |f(\zeta^{1,x}(s), \zeta^{2,t,x}(r))| \leq C_t^x \tag{5.3.36}$$

(ii)

$$\sup_{s,r \leq t} |G_x^1(s) f(\zeta^{1,x}(s), \zeta^{2,t,x}(r))| \leq C_t^x \tag{5.3.37}$$

(iii)

$$\sup_{s,r \leq t} |G_x^{2,t}(r) f(\zeta^{1,x}(s), \zeta^{2,t,x}(r))| \leq C_t^x \tag{5.3.38}$$

If for $\mathbb{P}^{env}$-a.e. $x \in S'$, for any $t \geq 0$ and for a.e. $0 \leq s \leq t$

$$G_x^1(s) f(.,v)(u) = G_x^{2,t}(t-s) f(u,.)(v) + g(t-s, v) \tag{5.3.39}$$

for every $u \in S_1$ and every $v \in S_2$, then

$$\mathscr{G}^1(m_0^1) = (G^1, \mathcal{D}_1, m_0^1, \mathbb{P}^{env})$$

and

$$\mathscr{G}^2(m_0^2) = \{(G^{2,t}, \mathcal{D}_2, m_0^2, \mathbb{P}^{env})\}_{t \geq 0}$$

are dual with respect to (f, g).

Proof: The proof is an automatic application of Proposition 5.3.6 and Remark 5.3.3. ■

5.4 A function-valued dual for FVRE

The goal of this section is to construct a dual process in r.e. e which is a fitness process (not necessarily Markov). Recall that $\mathbb{P}^{env}$ is the law of e. For any $\tilde{e} \in D_E[0,\infty)$, we define the quenched dual family $\{\psi^{t,\tilde{e}}\}_{t\in\mathbb{R}_+}$ of Markov processes with the deterministic environment $\tilde{e} \in D_E[0,\infty)$, where $\psi^{t,\tilde{e}} = (\psi^{t,\tilde{e}}_s)_{s\in\mathbb{R}_+}$. The process $\psi^{t,\tilde{e}}$ is a Markov jump process with the state space $\mathcal{C}^* = \bigcup_{n\in\mathbb{N}} \mathcal{C}_n(I^{\mathbb{N}})$ without any jumps after time $t \in \mathbb{R}_+$, i.e. $\psi^t(t+s) = \psi^t(t)$ for any $s \geq 0$ (The process stays forever in its location at time t). Also, as before, we assume that e is independent of Poisson times of jumps and also independent of the initial distribution of the process.

In order to define the transition probabilities of $\psi^{t,\tilde{e}}$ at times of jumps, we need the following notations. For $\underline{a} = (a_1, a_2, ...) \in I^{\mathbb{N}}$, define the insertion function $\varrho_i^{ins} : I^{\mathbb{N}} \to I^{\mathbb{N}}$ to be

$$\varrho_i^{ins}(\underline{a}) = (a_{j-\mathbf{1}_{\{j>i\}}})_{j\in\mathbb{N}}, \tag{5.4.1}$$

where the value of $\mathbf{1}_{\{j>i\}}$ is 1 if $j > i$, and it is 0, otherwise. Also, the deletion function $\varrho_i^{del} : I^{\mathbb{N}} \to I^{\mathbb{N}}$ is defined by

$$\varrho_i^{del}(\underline{a}) = (a_{j+\mathbf{1}_{\{j>i\}}})_{j\in\mathbb{N}}. \tag{5.4.2}$$

For $n \in \mathbb{N}$, the process $\psi^{t,\tilde{e}}$ jumps from state $f \in \mathcal{C}_n(I^{\mathbb{N}}) \subset \mathcal{C}^*$ to

$$f \circ \sigma_{ij} \circ \varrho_j^{ins} \quad \textit{at rate } \frac{\gamma}{2} \textit{ for } i,j = 1,...,n \textit{ (resampling)} \tag{5.4.3}$$

to

$$B_i' f \quad \textit{at rate } \beta' \textit{ for } i = 1,..,n \textit{ (parent} - \textit{independent mutation)} \tag{5.4.4}$$

to

$$B_i'' f \quad at\ rate\ \beta''\ for\ i = 1, .., n\ (parent - dependent\ mutation) \tag{5.4.5}$$

(Recall $\beta = \beta' + \beta''$ and also recall the definition from (5.2.14) and (5.2.15)),
to

$$\tilde{e}_i(t-s)f + (1-\tilde{e}_i(t-s))(f \circ \varrho_i^{del}) \quad at\ rate\ \alpha\ for\ i = 1, ..., n\ (selection) \tag{5.4.6}$$

for a jump occurring at time $s \leq t$.

Having the fitness Markov e with law $\mathbb{P}^{env}$ which can be considered as a $D_E[0,\infty)$-valued random variable, we can think of a family of stochastic processes in random environment e, namely $\{\psi^{t,e}\}_{t \in \mathbb{R}_+}$. In fact, for $t \geq 0$, $\psi^{t,e}$ is a stochastic process in random environment e whose quenched processes, $\{\psi^{t,\tilde{e}}\}_{\tilde{e} \in D_E[0,\infty)}$, are defined as above.

We define the duality function $\tilde{H} : \mathcal{P}(I) \times \mathcal{C}^* \to \mathbb{R}$ by

$$\tilde{H}(m, f) = < m^{\otimes \mathbb{N}}, f > \tag{5.4.7}$$

for $f \in \mathcal{C}^*$ and $m \in \mathcal{P}(I)$.

For $f \in \mathcal{C}_n(I^{\mathbb{N}}) \subset \mathcal{C}^*$, in fact $\tilde{H}(m, f) = \tilde{\Phi}^f(m)$, where by definition $\Phi^f \in \tilde{\mathfrak{F}}$. Note that $\tilde{H}$ is a continuous function and hence measureable function but not bounded. The following is automatic.

Proposition 5.4.1. *The collection of functions $\{\tilde{H}(., f) : f \in \mathcal{C}^*\}$ is measure-determining on $\mathcal{P}(I)$.*

Proof: The set in the statement of proposition is in fact $\tilde{\mathfrak{F}}$ and, it was already proved that $\tilde{\mathfrak{F}}$ is measure-determining. ∎

Before proving the duality relation, we find the generator of $\psi^{t,\tilde{e}}$ for $t \geq 0$. For $\tilde{m} \in \mathcal{P}(I^{\mathbb{N}})$, define $\phi^{\tilde{m}} \in \mathcal{C}(\mathcal{C}^*)$ (remember $\mathcal{C}^*$ is equipped with sup-norm topology) as

$$\phi^{\tilde{m}}(f) = < \tilde{m}, f > \quad f \in \mathcal{C}^*. \tag{5.4.8}$$

From construction, for $t \geq 0$ and $\tilde{m} \in \mathcal{P}(I^{\mathbb{N}})$, the time-dependent generator of $\psi^{t,\tilde{e}}$ on function $\phi^{\tilde{m}}$, namely $\mathcal{G}^{*t}_{\tilde{e}}$, is computed as follows. For $0 \leq s \leq t$ and $f \in \mathcal{C}_n(I^{\mathbb{N}})$ for $n \in \mathbb{N}$

$$\begin{aligned} &\mathcal{G}^{*t}_{\tilde{e}}(s)\phi^{\tilde{m}}(f) = \tfrac{\gamma}{2} \sum_{0 \leq i,j \leq n} < \tilde{m}, f \circ \sigma_{ij} \circ \varrho^{ins}_j - f > \\ &+\beta' \sum_{i=1}^{n} < \tilde{m}, B'_i f - f > +\beta'' \sum_{i=1}^{n} < \tilde{m}, B''_i f - f > \\ &+\mathcal{G}^{*t,sel}_{\tilde{e}}(s)\phi^{\tilde{m}}(f) \end{aligned} \tag{5.4.9}$$

To continue, we must compute the last term that is the generator of the dual process $\psi^{t,\tilde{e}}$ corresponding to the selection jumps (5.4.6). Recall that the probability measure $P^*_{s,s+r}$ on $[s, s+r]$, for any $t - s \geq r \geq 0$, is the law of choosing one Poisson point in the interval $[s, s+r]$ conditioned on having only one Poisson point in that interval. For $x \in I^{\mathbb{N}}$, let

$$\overleftarrow{e}^{s,s+r}_i(x) = \int_s^{s+r} \tilde{e}_i(t-u)(x) P^*_{s,s+r}(du). \tag{5.4.10}$$

As before, since $\tilde{e} \in D_E[0, \infty)$, the left limit of $\tilde{e}$ exists for any $i \leq n$ and any time $s \in \mathbb{R}_+$ and

$$\inf_{s \leq u \leq s+r} \tilde{e}_i(t-u)(x) \leq \overleftarrow{e}^{t,t+s}_i(x) \leq \sup_{s \leq u \leq s+r} \tilde{e}_i(t-u)(x). \tag{5.4.11}$$

Therefore

$$\lim_{r \to 0} \int_s^{s+r} \tilde{e}(t-u)(.) P^*_{s,s+r}(du) = \tilde{e}(s^-)(.) \tag{5.4.12}$$

pointwisely, (and furthermore in sup-norm topology), where

$$\tilde{e}_i(s^-)(x) = \lim_{u \to s^-} \tilde{e}_i(u)(x). \tag{5.4.13}$$

Thus

$$\begin{aligned}
&\mathcal{G}^{*t,sel}_{\tilde{e}}(s)\phi^{\tilde{m}}(f) = \lim_{r\to 0}\tfrac{1}{r}\{\alpha r\sum_{i=1}^{n}\int_{s}^{s+r} < \tilde{m}, \tilde{e}_i(t-u)f+ \\
&(1-\tilde{e}_i(t-u))f\circ\varrho_i^{del} > P^*_{s,s+r}(du) + (1-\alpha r)\sum_{i=1}^{n} < \tilde{m}, f > - \sum_{i=1}^{n} < \tilde{m}, f >\} = \\
&\lim_{r\to 0}\alpha\sum_{i=1}^{n} < \tilde{m}, f\int_{s}^{s+r}\tilde{e}_i(t-u)P^*_{s,s+r}(du) + f\circ\varrho_i^{del}(1-\int_{s}^{s+r}\tilde{e}_i(t-u)P^*_{s,s+r}(du) >) \\
&-\alpha\sum_{i=1}^{n} < \tilde{m}, f >= \\
&\alpha\sum_{i=1}^{n} < \tilde{m}, f\tilde{e}_i(s^-) + f\circ\varrho_i^{del}(1-\tilde{e}_i(s^-)) - f >= \\
&\alpha\sum_{i=1}^{n} < \tilde{m}, f\tilde{e}_i(s^-) - f\circ\varrho_i^{del}\tilde{e}_i(s^-) > +\alpha\sum_{i=1}^{n}(< \tilde{m}, f\circ\varrho_i^{del} > - < \tilde{m}, f >).
\end{aligned} \tag{5.4.14}$$

If $\tilde{m} = m^{\otimes\mathbb{N}}$ for $m \in \mathcal{P}(I)$, then the right hand side of the equality will be

$$\alpha\sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, f\tilde{e}_i(s^-) - f\tilde{e}_{n+1}(s^-) > \tag{5.4.15}$$

Remark 5.4.2. *As we already mentioned, assuming that e has sample paths in $D_E[0,\infty)$ is essential in order to compute the operator process. Also note that under this assumption the operator processes $(\mathcal{G}^e_s)_{s\geq 0}$ and $(\mathcal{G}^{*t,e}_s)_{s\geq 0}$ take sample paths in $D_E[0,\infty)$, a.s..*

Applying above computations on $\tilde{H}(.,.)$, we have

$$\begin{aligned}
&\mathcal{G}^{*t}_{\tilde{e}}(s)\tilde{H}(m^{\otimes\mathbb{N}},.)(f) = \tfrac{\gamma}{2}\sum_{0\leq i,j\leq n} < m^{\otimes\mathbb{N}}, f\circ\sigma_{ij}\circ\varrho_j^{ins} - f > \\
&+\beta'\sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, B'_i f - f > +\beta''\sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, B''_i f - f > \\
&+\alpha\sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, f\tilde{e}_i(s^-) - f\tilde{e}_{n+1}(s^-) > .
\end{aligned} \tag{5.4.16}$$

Because of exchangeability, we can rewrite the first term of the last equation as

$$\frac{\gamma}{2}\sum_{0\leq i,j\leq n} < m^{\otimes\mathbb{N}}, f\circ\sigma_{ij}\circ\varrho_j^{ins} - f >= \frac{\gamma}{2}\sum_{0\leq i,j\leq n} < m^{\otimes\mathbb{N}}, f\circ\sigma_{ij} - f > . \tag{5.4.17}$$

On the other hand, for the function $\tilde{H}(.,f) = \Phi^f(.)$, where $f \in \mathcal{C}_n(I^{\mathbb{N}})$, we already saw that

$$\begin{aligned} &\tilde{\mathcal{G}}_{\tilde{e}}\tilde{H}(.,f)(m) = \\ &\frac{\gamma}{2}\sum_{0\le i,j\le n} < m^{\otimes\mathbb{N}}, f\circ\sigma_{ij} - f > + \\ &+\beta'\sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, B'_i f - f > +\beta''\sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, B''_i f - f > \\ &+\alpha\sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, f\tilde{e}_i(s) - f\tilde{e}_{n+1}(s) > . \end{aligned} \tag{5.4.18}$$

Since $\tilde{e}$ is in $D_E[0,\infty)$, it is right continuous with left limit. Also the number of discontinuity points of $\tilde{e}$ is at most countable. For any $t \geq 0$, this yields the equality

$$\tilde{\mathcal{G}}_{\tilde{e}}(s) = \mathcal{G}^{*t}_{\tilde{e}}(t-s) \tag{5.4.19}$$

for every $s \leq t$ except possibly at most countable points of discontinuity of $\tilde{e}$. Furthermore, constructing the corresponding operator process of $\psi^{t,\tilde{e}}$, namely $\tilde{\mathcal{G}}^{*t,e}$, which is consistent with e, for any $t \geq 0$, there exist at most countable times $s \leq t$ for which

$$\tilde{\mathcal{G}}^{e}(s) = \mathcal{G}^{*t,e}(t-s) \tag{5.4.20}$$

does not hold almost surely.

In fact we can deduce the duality relation between FV in environment $\tilde{e}$ and jump Markov processes $\{\psi^{t,\tilde{e}}\}_{t\geq 0}$. Before doing this, we need to know an easy property of the dual processes whose proof will be postponed, namely, for $t \geq 0$, starting at the state $\hat{\psi}_0 \in \mathcal{C}^*$, $\|\psi^{t,\tilde{e}}_s\|_\infty$ ($\|.\|_\infty$-supnorm on $\mathcal{C}^*$) remains bounded by $\|\hat{\psi}_0\|_\infty$ for any $s \geq 0$, a.s.. The following proposition states that for any $\tilde{e} \in D_E[0,\infty)$ and $t \geq 0$, conditioning on $\mu^{\tilde{e}}_0 = m_0 \in \mathcal{P}(I)$ and $\psi^{t,\tilde{e}}_0 = \hat{\psi}_0 \in \mathcal{C}^*$, the duality relation holds for $\mu^{\tilde{e}}$ and $\{\psi^{t,\tilde{e}}\}_{t\in\mathbb{R}_+}$.

Proposition 5.4.3. *For every $\tilde{e} \in D_E[0,\infty)$, $m_0 \in \mathcal{P}(I)$, and $\hat{\psi}_0 \in \mathcal{C}^*$, the time-dependent martingale problem*

$$(\tilde{\mathcal{G}}_{\tilde{e}}, \tilde{\mathfrak{F}}, \delta_{m_0}) \tag{5.4.21}$$

and one-parameter family of time-dependent martingale problems

$$\{(\mathcal{G}^{*t}_{\tilde{e}}, \{< m^{\otimes \mathbb{N}}, . >\}_{m \in \mathcal{P}(I)}, \delta_{\hat{\psi}_0})\}_{t \geq 0} \tag{5.4.22}$$

are dual with respect to $(\tilde{H}, 0)$*, that is for every* $t \geq 0$

$$\mathbb{E}^{m_0}[\tilde{H}(\mu^{\tilde{e}}_t, \hat{\psi}_0)] = \mathbb{E}^{\hat{\psi}_0}[\tilde{H}(m_0, \psi^{t,\tilde{e}}_t)] \tag{5.4.23}$$

Remark 5.4.4. *The statement of the theorem is stronger than Definition 17 as it guarantees the duality relation for every fixed (quenched) environment. Also, from the proof, it will be clear that, for any integrable* $\mathcal{C}^*$*-valued random variable* $\underline{\psi}$*, the duality relation holds between* $(\tilde{\mathcal{G}}_{\tilde{e}}, \tilde{\mathfrak{F}}, \delta_{m_0})$ *and*

$$\{(\mathcal{G}^{*t}_{\tilde{e}}, \{< m^{\otimes \mathbb{N}}, . >\}_{m \in \mathcal{P}(I)}, \mathbb{P}\underline{\psi}^{-1})\}_{t \geq 0}. \tag{5.4.24}$$

Proof: Boundedness of $\psi^{t,\tilde{e}}_s$ a.s. for any $t, s \geq 0$ yields that for any $t \geq 0$ and for any $s, r \leq t$, there exists a constant $C > 1$ such that

$$< \mu^{\tilde{e}}_s, \psi^{t,\tilde{e}}_r > \leq \|\hat{\psi}_0\|_\infty \leq C\|\hat{\psi}_0\|_\infty \quad a.s. \tag{5.4.25}$$

$$\|\tilde{\mathcal{G}}_{\tilde{e}}(s)\tilde{H}(\mu^{\tilde{e}}_s, \psi^{t,\tilde{e}}_r)\|_\infty \leq C < \mu^{\tilde{e}}_s, \psi^{t,\tilde{e}}_r > \leq C\|\hat{\psi}_0\|_\infty \quad a.s. \tag{5.4.26}$$

and

$$\|\tilde{\mathcal{G}}^{*t}_{\tilde{e}}(s)\tilde{H}(\mu^{\tilde{e}}_s, \psi^{t,\tilde{e}}_r)\|_\infty \leq C < \mu^{\tilde{e}}_s, \psi^{t,\tilde{e}}_r > \leq C\|\hat{\psi}_0\|_\infty \quad a.s. \tag{5.4.27}$$

since for any $f \in \mathcal{C}^*$, $\|f \circ \sigma_{ij} \circ \varrho^{ins}_j\|_\infty$, $\|B'_i f\|_\infty$, $\|B''_i f\|_\infty$, and $\|f\tilde{e}_i(s) - f\tilde{e}_{n+1}(s)\|_\infty$ are bonded by $\|f\|_\infty$ (cf. Proposition 5.4.8). Therefore the assumptions of Proposition 5.3.6 hold. Then the statement of theorem follows from (5.4.19) (for every t and a.e. $s \leq t$), and Proposition 5.3.6. ∎

Remark 5.4.5. *The same argument as the one in the proof of the last proposition shows that, in fact the following more general duality relation is true. For* $r \geq 0$*, let* $\tilde{\mathcal{G}}^{+r} : \Omega \times \mathbb{R}_+ \times \tilde{\mathfrak{F}} \to \tilde{\mathfrak{F}}$ *be an operator process defined by*

$$\tilde{\mathcal{G}}^{+r}(\omega, s) := \tilde{\mathcal{G}}(\omega, s + r) \text{ for } s \geq 0 \text{ and } \omega \in \Omega. \tag{5.4.28}$$

Then, the simple observation that for $t \geq 0$ and for a.e. $s \leq t$

$$\mathcal{G}^{*t}(s) = \tilde{\mathcal{G}}(t-s) = \tilde{\mathcal{G}}^{+r}(t-r-s) \qquad \mathbb{P}^{env} - a.s. \tag{5.4.29}$$

shows that for every $\tilde{e} \in D_E[0,\infty)$, $m_0 \in \mathcal{P}(I)$, $r \geq 0$, and $\hat{\psi}_0 \in \mathcal{C}^*$, the time-dependent martingale problem

$$(\tilde{\mathcal{G}}^{+r}_{\tilde{e}}, \tilde{\mathfrak{F}}, \delta_{m_0}) \tag{5.4.30}$$

and one-parameter family of time-dependent martingale problems

$$\{(\mathcal{G}^{*t}_{\tilde{e}}, \{< m^{\otimes \mathbb{N}}, . >\}_{m \in \mathcal{P}(I)}, \delta_{\hat{\psi}_0})\}_{t \geq r} \tag{5.4.31}$$

are dual with respect to $(\tilde{H}, 0)$, that is for every $t \geq r$

$$\mathbb{E}^{m_0}[\tilde{H}(\mu^{+r,\tilde{e}}_{t-r}, \hat{\psi}_0)] = \mathbb{E}^{\hat{\psi}_0}[\tilde{H}(m_0, \psi^{t,\tilde{e}}_{t-r})], \tag{5.4.32}$$

where $\mu^{+r,\tilde{e}}_s = \mu^{\tilde{e}}_{s+r}$ for $s \geq 0$.

The family of annealed stochastic processes $\{\psi^{t,e}\}_{t \in \mathbb{R}_+}$ is called the dual in r.e. e. Also for any $\tilde{e} \in D_E[0,\infty)$, the family of time-inhomogeneous Markov processes $\{\psi^{t,\tilde{e}}\}_{t \in \mathbb{R}_+}$ is called the dual in quenched environment (or with quenched fitness process) $\tilde{e}$.

Proposition 5.4.6. *For any measurable map*

$$\tilde{\mathbf{P}}_0 : D_E[0,\infty) \to \mathcal{P}(\mathcal{P}(E)),$$

the $(\tilde{\mathcal{G}}, \tilde{\mathfrak{F}}, \tilde{\boldsymbol{P}}_0, \mathbb{P}^{env})$-martingale problem has at most one solution.

Proof: Stronger than the statement of the theorem, we show that for every $\tilde{e} \in D_E[0,\infty)$, the time-dependent martingale problem $(\tilde{\mathcal{G}}_{\tilde{e}}, \tilde{\mathfrak{F}}, m_0)$ for any $m_0 \in \mathcal{P}(\mathcal{P}(I))$ is well-posed. Since $\mathcal{P}(I)$ is compact, this is equivalent to well-posedness of $(\tilde{\mathcal{G}}_{\tilde{e}}, \tilde{\mathfrak{F}}, \delta_\nu)$ for every $\nu \in \mathcal{P}(I)$. The latter is an immediate consequence of the duality relation (Proposition 5.4.3), Proposition 5.4.1, and Proposition 5.3.4. ■

In order to prove an ergodic theorem for FVRE we study the long-time behaviour of the dual family.

Proposition 5.4.7. *Suppose there exists a parent-independent mutation component, that is $\beta' > 0$. Then there exists an almost surely finite random time τ at which, for every $t \geq 0$ and $\tilde{e} \in D_E[0,\infty)$, $\psi_\tau^{t,\tilde{e}}$ does not depend on variables of $I^{\mathbb{N}}$, i.e. $\psi_\tau^{t,\tilde{e}}$ is a random constant function (a $\mathcal{C}_0(I^N)$-valued random variable), and τ is independent of t and $\tilde{e}$.*

Proof: First note that if there exists such a random time, then it is independent of the choice of $t \geq 0$ and $\tilde{e} \in D_E[0,\infty)$. This is true since the random time is only a function of Poisson jump processes which by assumption are independent of e and t. So it is enough to show the existence of τ for an arbitrary quenched process $\psi = (\psi_s)_{s\geq 0} := \psi_s^{t,\tilde{e}}$ for $\tilde{e} \in D_E[0,\infty)$ and $t \geq 0$. Note that constant functions , i.e. the elements of $\mathcal{C}_0(I^{\mathbb{N}})$, are absorbing states. For an arbitrary initial state $f \in \mathcal{C}_n(I^{\mathbb{N}})$, we prove that there exists a random almost surely finite time at which the process hits $\mathcal{C}_0(I^{\mathbb{N}})$. The degree of a function in $\mathcal{C}^*$ is the maximum number of variables (possibly 0) on which the function depends. Consider the natural surjective mapping from $\mathcal{C}^*$ onto $\mathbb{N}\cup\{0\}$ that corresponds to each function in $\mathcal{C}^*$, its degree in $\mathbb{N}\cup\{0\}$. This mapping induces a continuous time random walk on the state space $\mathbb{N}\cup\{0\}$, more precisely defined by $\varphi_t = n$ if the degree of ψ_t is n. for $t \geq 0$. Note that ψ hits a constant if and only if $\varphi = (\varphi_s)_{s\geq 0}$ hits 0. In fact, we can see that φ is a birth-death process with a quadratic rate of death and a linear rate of birth. In order to see this, we need to determine the degree of all the states to which ψ can jump from an arbitrary state $f \in \mathcal{C}_n(I^{\mathbb{N}})$. It is clear that, at any time $s \leq t$, f can jump only to states

$$\begin{aligned}
&f \circ \sigma_{ij} \circ \varrho_j^{ins} \in \mathcal{C}_{n-1} \textit{ with rate } \tfrac{\gamma}{2} \;\; \textit{for } i,j = 1,..,n \\
&B_i' f \in \mathcal{C}_{n-1} \quad \textit{with rate } \beta' \textit{ for } i = 1,..,n \\
&B_i'' f \in \mathcal{C}_n \quad \textit{with rate } \beta'' \textit{ for } i = 1,..,n \\
&\tilde{e}(t-s)f + (1-\tilde{e}(t-s))f \in \mathcal{C}_{n+1} \quad \textit{with rate } \alpha \textit{ for } i = 1,..,n \\
&.
\end{aligned} \tag{5.4.33}$$

Again, it is clear that time s and environment $\tilde{e}$ do not affect on the birth and death rates. Let g be a polynomial with degree n. There exists an $\tilde{n} \geq n$ such that $g \in \mathcal{C}_{\tilde{n}}(I^{\mathbb{N}})$. Therefore, at a time of jump, a birth occurs at state n with probability

$$\frac{n\alpha}{\binom{\tilde{n}}{2}\frac{\gamma}{2} + n\beta' + n\alpha}, \tag{5.4.34}$$

and a death occur with probability

$$\frac{\binom{\tilde{n}}{2}\frac{\gamma}{2} + n\beta'}{\binom{\tilde{n}}{2}\frac{\gamma}{2} + n\beta' + n\alpha}. \tag{5.4.35}$$

Thus, for any $n \geq 2$, φ starting in n will hit 1 in an a.s. finite random time. Suppose φ does not hit 0. Then the last line of argument implies that φ hits 1 infinitely many times without jumping afterwards to 0. But this is not possible due to the existence of the parent-independent mutation component which gives a positive probability, $\frac{\beta'}{\beta'+\alpha}$, of jumps from 1 to 0. ■

Similarly to [13], we show that the dual process is non-increasing a.s..

Proposition 5.4.8. *For any $\tilde{e} \in D_E[0,\infty)$ and $t \geq 0$, the dual process $\psi^{t,\tilde{e}}$, starting in $\hat{\psi}_0$, is non-increasing and bounded by $\|\hat{\psi}_0\|_\infty$ a.s..*

Proof: Let $n \in \mathbb{N}$ and $f \in \mathcal{C}_n(I^{\mathbb{N}}) \subset \mathcal{C}^*$ be arbitrary. For any $i, j \leq n$ and $a \in I$, $f \circ \sigma_{ij} \circ \varrho_i^{ins}$, $f \circ \sigma_i^a$, and $f \circ \varrho_i^{del}$ are defined by setting restrictions on the first n variables of f, that is they are restrictions of f on a subdomain and therefore

$$\begin{aligned} &\|f \circ \sigma_i^a\|_\infty \leq \|f\|_\infty \\ &\|f \circ \varrho_i^{del}\|_\infty \leq \|f\|_\infty \\ &\|f \circ \sigma_{ij} \circ \varrho_i^{ins}\|_\infty \leq \|f\|_\infty \\ &. \end{aligned} \tag{5.4.36}$$

Also

$$\begin{aligned} &\|B'f\|_\infty = \\ &\sup_{x \in I^{\mathbb{N}}} |\int_I f \circ \sigma_i^u(x) q'(du)| \leq \\ &\int_I \|f \circ \sigma_i^u\|_\infty q'(du) \leq \|f\|_\infty \end{aligned} \tag{5.4.37}$$

Similarly,

$$\|B''f\|_\infty \leq \|f\|_\infty. \tag{5.4.38}$$

For a selection jump at time $t \geq 0$, for $x \in I^{\mathbb{N}}$, if $f \circ \varrho_i^{del}(x) \leq f(x)$, then

$$\tilde{e}_i(t-s)(x)f(x) + (1-\tilde{e}_i(t-s)(x))f \circ \varrho_i^{del}(x) \leq f(x) \leq \|f\|_\infty, \tag{5.4.39}$$

and if, $f(x) \leq f \circ \varrho_i^{del}(x)$, then

$$\begin{aligned} &\tilde{e}_i(t-s)(x)f(x) + (1-\tilde{e}_i(t-s)(x))f \circ \varrho_i^{del}(x) \\ &\leq f \circ \varrho_i^{del}(x) \leq \|f \circ \varrho_i^{del}\|_\infty \leq \|f\|_\infty. \end{aligned} \tag{5.4.40}$$

Thus,

$$\|\tilde{e}_i(t-s)f + (1-\tilde{e}_i(t-s))f \circ \varrho_i^{del}\|_\infty \leq \|f\|_\infty. \tag{5.4.41}$$

Therefore, all jumps lead us to a function with a smaller sup-norm. In other words, $t \mapsto \|\psi_s^{t,\tilde{e}}\|$ is a non-increasing function, a.s.. In particular, for any $s \geq 0$, $\|\psi_s^{t,\tilde{e}}\|_\infty \leq \|\hat{\psi}_0\|_\infty$ a.s.. ■

To understand the long time behaviour of FVRE, we can study the long time behaviour of the dual process. We need the following lemma for this purpose.

Lemma 5.4.9. *Let Z^{ν_0} be an S-valued Markov process with initial distribution $\nu_0 \in \mathcal{P}(S)$ and with homogeneous transition probability function $p(t,x,dy)$ whose semigroup is denoted by $(T_t^Z : \bar{\mathcal{C}}(S) \to \bar{\mathcal{C}}(S))_{t\geq 0}$, with*

$$T_t^Z f(x) = \int_S f(y)p(t,x,dy) \ for \ f \in \bar{\mathcal{C}}(S). \tag{5.4.42}$$

Assume that Z^{ν_0} takes its sample paths in $D_S[0,\infty)$ a.s.. Suppose Z is weakly ergodic, i.e. there exists $\nu \in \mathcal{P}(S)$ such that for any $x \in S$ and $f \in \bar{\mathcal{C}}(S)$ we have $T_t^Z f(x) \to <\nu, f>$ as $t \to \infty$. Let m^{ν_0} be the law of Z^{ν_0} for $\nu_0 \in \mathcal{P}(S)$, and let $\tilde{Z} = Z^\nu$ and denote its law by $\tilde{m}$ (i.e. $\tilde{Z}$ is a stationary Markov process with law $\tilde{m} = m^\nu$). Denote by $m^{\nu_0}_{t_1,t_2,\dots,t_k}$ and $\tilde{m}_{t_1,t_2,\dots,t_k}$ the k-dimensional distributions of m^{ν_0} and $\tilde{m}$, respectively, for $k \in \mathbb{N}$ and real numbers $0 \leq t_1 < t_2 < \dots < t_k$. Then *(i) For any $\nu_0 \in \mathcal{P}(S)$, any $k \in \mathbb{N}$ and any sequence of real numbers $0 \le t_1 < t_2 < ... < t_k$*

$$m^{\nu_0}_{t_1+s,t_2+s,...,t_k+s} \Rightarrow \tilde{m}_{t_1,t_2,...,t_k} \tag{5.4.43}$$

as $s \to \infty$, for any $k \in \mathbb{N}$ and $t_1 < ... < t_k \in \mathbb{R}_+$.

(ii) In addition to above assumptions, let S be compact, and $\overline{\mathcal{D}(G^Z)}$, where G^Z is the generator of Z, contains an algebra that separates points and vanishes nowhere. Let $\hat{Z}^{\nu_0,t} := (\hat{Z}^{\nu_0,t}_s)_{s\ge 0}$, defined by $\hat{Z}^{\nu_0,t}_s := Z^{\nu_0}_{s+t}$(the initial distribution of $\hat{Z}^{\nu_0,t}$ is the law of $Z^{\nu_0}_t$). Then for any ν_0 the process $\hat{Z}^{\nu_0,t}$ weakly converges to $\tilde{Z}$ in $D_S[0,\infty)$ as $t \to \infty$.

Remark 5.4.10. *Recall that another equivalent definition of weak ergodicity for Z is that there exists a probability measure $\nu \in \mathcal{P}(S)$ such that for every initial distribution ν_0*

$$\lim_{t\to\infty} \mathbb{E}_{m^{\nu_0}}[f(Z^{\nu_0}_t)] =< \nu, f > \quad f \in \bar{\mathcal{C}}(S) \tag{5.4.44}$$

Proof:

(i) We must prove for arbitrary $\nu_0 \in \mathcal{P}(S)$, $k \in \mathbb{N}$, $0 \le t_1 < t_2 < ... < t_k$ and $f \in \bar{\mathcal{C}}(S^n)$

$$\int_{S^n} f dm^{\nu_0}_{t_1+s,t_2+s,...,t_k+s} \to \int_{S^n} f d\tilde{m}_{t_1,t_2,...,t_k}. \tag{5.4.45}$$

In order to prove the convergence, it suffices to prove it for a convergence-determining set of functions. In particular, we prove that convergence holds for

$$\{f \in \bar{\mathcal{C}}(S) : f(x_1,x_2,...,x_k) = f_1(x_1)f_2(x_2)...f_k(x_k), \quad k \in \mathbb{N},\ f_i \in \bar{\mathcal{C}}(S), i = 1,...,k\} \tag{5.4.46}$$

Set $s_i := t_{i+1} - t_i$ for $i = 1, ..., k-1$. For $f_1, ...f_k \in \tilde{\mathcal{C}}(S)$

$$\begin{aligned}
&\int_{S^n} f_1(x_1)...f_k(x_k)dm^{\nu_0}_{t_1+s,t_2+s,...,t_k+s}(dx_1,...,dx_k) = \\
&\int_S f_1(x_1)\int_S f_2(x_2)\int_S ... \int_S f_{k-1}(x_{k-1})\times \\
&\int_S f_k(x_k)p(s_{k-1},x_{k-1},dx_k)...p(s_1,x_1,dx_2)p(t_1+s,x_0,dx_1)\nu_0(dx_0) = \\
&\int_S T^Z_s T^Z_{t_1}(f_1(x)T^Z_{s_1}(...(f_{k-2}(x)T^Z_{s_{k-2}}(f_{k-1}(x)T^Z_{s_{k-1}}(f_k(x))))...))d\nu_0.
\end{aligned} \tag{5.4.47}$$

Under the assumptions, the function

$$g(x) := T^Z_{t_1}(f_1(x)T^Z_{s_1}(...(f_{k-2}(x)T^Z_{s_{k-2}}(f_{k-1}(x)T^Z_{s_{k-1}}(f_k(x))))...)) \tag{5.4.48}$$

is continuous, therefore $< \nu_0, T^Z_s g > \to < \nu, g >$ as $s \to \infty$, by weak ergodicity of Z.

(ii) It suffice to prove the tightness (cf. Theorem 3.7.8 [23]). But this follows Remark 4.5.2 in [23] and the fact that the generators of $\hat{Z}^{\nu_0,t}$, for any $t \geq 0$ is identical to G^Z.

■

Now, we are ready to state a main tool to study the long-time behaviour of the FVRE. We do this by the study of long time behaviour of the dual processes.

Theorem 5.4.11. *Suppose that there exists a parent-independent component in the mutation process, i.e.* $\beta' > 0$*, and let* e *either be a stationary fitness process (not necessarily Markov) or a weakly ergodic Markov fitness with semigroup* $\{T^{env}_t\}$ *such that* $T^{env}_t : \mathcal{C}(E) \to \mathcal{C}(E)$ *for any* $t \geq 0$ *. Then, conditioning on* $\psi^{t,e}_0 = \hat{\psi}_0 \in \mathcal{C}^*$*, the limit*

$$\lim_{t\to\infty} \mathbb{E}^{\hat{\psi}_0}[f(\psi^{t,e}_t)] \tag{5.4.49}$$

exists for any $f \in \bar{\mathcal{C}}(\mathcal{C}^*)$ *and is bounded by* $\|\hat{\psi}_0\|_\infty$*. In particular*

$$\lim_{t\to\infty} \mathbb{E}^{\hat{\psi}_0}[\psi^{t,e}_t] \tag{5.4.50}$$

exists (remember that $\psi_t^{t,e}$ hits a constant function, an absorbing state, in finite time a.s. and therefore (5.4.50) is meaningful).

Proof: Let $J_s^{t,e}$ be the set of all times of Poisson jumps for $\psi^{t,e}$ up to time s, including resampling, mutation , and selection times of jumps, that $J_s^{t,e}$ has all information of Poisson point processes, but not any information about e. Because of stationarity, the process $(J_s^{t,e})_{s\geq 0}$ is independent of t and also e (by assumption). Therefore, it is convenient to drop (t,e) from the superscript. Similarly, we define the following stochastic processes which are independent of e and t (for the same reason) and therefore we drop (t,e) from the superscript again. Let κ_s be the stochastic jump process counting the number of selective events of $\psi^{t,e}$ up to time $s \geq 0$, and let

$$T_s^{sel} := \{\varsigma_1^{sel} \leq \varsigma_2^{sel} \leq ... \leq \varsigma_{\kappa_s}^{sel}\} \tag{5.4.51}$$

be the times of selective events occurring for $\psi^{t,e}$ up to time s. As before, let τ be the stopping time at which $\psi^{t,e}$ hits a random constant function. Recall that (cf. Proposition 5.4.7) τ is independent of e and t, and it is a.s. finite. For any t and e, $\psi_\tau^{t,e}$ is a random constant time whose value is a function, F, of τ, J_τ, κ_τ, T_τ^{sel}, and $\{e_{t-\varsigma_i^{sel}}\}_{i=1}^{\kappa_\tau}$. Specially, fixing $\tau = s$, $J_\tau = \hat{J}_s$, $\kappa_\tau = k$, $T_\tau^{sel} = \{t_1, ..., t_k\}$, the function F is continuous with respect to $e_{t-t_1}, ..., e_{t-t_k}$, i.e it is continuous with respect to variables of E^k. Let $(e_s^*)_{s\geq 0}$ be the stationary process generated by the semigroup T^{env} and invariant initial distribution ν (In the case that e is stationary, let $e_s^* = e_s$, for $s \geq 0$, and continue the same proof). As e is weakly ergodic, by Lemma 5.4.9, for any continuous function $g \in \mathcal{C}(E^k)$

$$\begin{aligned}&\lim_{t\to\infty} \mathbb{E}[g(e_{t-t_k}, e_{t-t_{k-1}}, ..., e_{t-t_1})] = \\ &\mathbb{E}[g(e_0^*, e_{t_k-t_{k-1}}^*, ..., e_{t_k-t_1}^*)].\end{aligned} \tag{5.4.52}$$

Since e is independent of τ, J_τ, κ_τ, T_τ^{sel} (by assumption), the conditional process e

given values of J_τ, τ, κ_τ, T^{sel}_τ is still weakly ergodic, and hence

$$\begin{aligned}&\lim_{t\to\infty}\mathbb{E}^{\hat\psi_0}[\psi^{t,e}_\tau|\tau,J_\tau,\kappa,T^{sel}_\tau]=\\&\lim_{t\to\infty}\mathbb{E}^{\hat\psi_0}[F(e_{t-\varsigma_k},...,e_{t-\varsigma_1},\tau,J_\tau,\kappa,T^{sel}_\tau)|\tau,J_\tau,\kappa,T^{sel}_\tau]=\\&\mathbb{E}^{\hat\psi_0}[F(e^*_0,e^*_{\varsigma_k-\varsigma_{k-1}}...,e^*_{\varsigma_k-\varsigma_1},\tau,J_\tau,\kappa,T^{sel}_\tau)|\tau,J_\tau,\kappa,T^{sel}_\tau].\end{aligned} \tag{5.4.53}$$

Similarly, for any $f\in\bar{\mathcal{C}}(\mathcal{C}^*)$,

$$\begin{aligned}&\lim_{t\to\infty}\mathbb{E}^{\hat\psi_0}[f(\psi^{t,e}_\tau)|\tau,J_\tau,\kappa,T^{sel}_\tau]=\\&\lim_{t\to\infty}\mathbb{E}^{\hat\psi_0}[f\circ F(e_{t-\varsigma_k},...,e_{t-\varsigma_1},\tau,J_\tau,\kappa,T^{sel}_\tau)|\tau,J_\tau,\kappa,T^{sel}_\tau]=\\&\mathbb{E}^{\hat\psi_0}[f\circ F(e^*_0,e^*_{\varsigma_k-\varsigma_{k-1}}...,e^*_{\varsigma_k-\varsigma_1},\tau,J_\tau,\kappa,T^{sel}_\tau)|\tau,J_\tau,\kappa,T^{sel}_\tau]\end{aligned} \tag{5.4.54}$$

Getting another expectation, knowing that τ is finite and $\|\psi^{t,e}_\tau\|\le\|\hat\psi_0\|$ a.s. yields that

$$\begin{aligned}&\lim_{t\to\infty}\mathbb{E}^{\hat\psi_0}[\psi^{t,e}_\tau]=\\&\mathbb{E}^{\hat\psi_0}[F(e^*_0,e^*_{\varsigma_k-\varsigma_{k-1}}...,e^*_{\varsigma_k-\varsigma_1},\tau,J_\tau,\kappa,T^{sel}_\tau)],\end{aligned} \tag{5.4.55}$$

and hence the limit exists and is bounded by $\|\hat\psi_0\|_\infty$. Similarly,

$$\begin{aligned}&\lim_{t\to\infty}\mathbb{E}^{\hat\psi_0}[f(\psi^{t,e}_\tau)]=\\&\mathbb{E}^{\hat\psi_0}[f\circ F(e^*_0,e^*_{\varsigma_k-\varsigma_{k-1}}...,e^*_{\varsigma_k-\varsigma_1},\tau,J_\tau,\kappa,T^{sel}_\tau)]\end{aligned} \tag{5.4.56}$$

for $f\in\bar{\mathcal{C}}(\mathcal{C}^*)$. ■

5.5 Convergence of generators

This section is devoted to the convergence of generators of MRE to FVRE. Before setting the convergence of generator processes, we need to extend the generators of the measure-valued Moran processes in a convenient sense. The Moran process takes values in $\mathcal{P}^N(I)$. Also all functions in $\tilde{\mathfrak{F}}_N$ have domain $\mathcal{P}^N(I)$. On the other hand, the FV process is a $\mathcal{P}(I)$-valued Markov process, and $\mathcal{P}(I)$ is the domain of polynomials in $\tilde{\mathfrak{F}}$. In order to measure the distance of the elements of $\tilde{\mathfrak{F}}$ and $\tilde{\mathfrak{F}}_N$, for $N \in \mathbb{N}$, we need to extend the functions in the second algebra to take all measures of $\mathcal{P}(I)$.

Let $S' \subset S$, and consider time-dependent linear operators $A : \mathbb{R}_+ \times \mathcal{D}(A) \to \mathcal{B}(S)$ and $B : \mathbb{R}_+ \times \mathcal{D}(B) \to \mathcal{B}(S')$ with $\mathcal{D}(A) \subset \mathcal{B}(S)$ and $\mathcal{D}(B) \subset \mathcal{B}(S')$. For $f \in \mathcal{D}(A)$ and $g \in \mathcal{D}(B)$, let $g^f \in \mathcal{B}(S)$ be

$$g^f(x) = \begin{cases} g(x) & \text{if } x \in S' \\ f(x) & \text{if } x \in S \setminus S'. \end{cases} \tag{5.5.1}$$

Set

$$D := \mathcal{D}(B, A) = \{g^f \in \mathcal{B}(S) : f \in \mathcal{D}(A), g \in \mathcal{D}(B)\}, \tag{5.5.2}$$

and define, for $t \geq 0$, the time-dependent linear operator $B^A : \mathbb{R}_+ \times D \to \mathcal{B}(S)$ by

$$B^A(t)g^f(x) = \begin{cases} B(t)g(x) & \text{if } x \in S' \\ A(t)f(x) & \text{if } x \in S \setminus S'. \end{cases} \tag{5.5.3}$$

It is clear that $(f, g) \mapsto g^f$ is bilinear with respect to the function addition. Moreover, if $\mathcal{D}(A)$ and $\mathcal{D}(B)$ are algebras, then so is D. The time-dependent linear operator (B^A, D) is called the extension of $(B, \mathcal{D}(B))$ with respect to $(A, \mathcal{D}(A))$.

For a moment, denote by $\|.\|_\infty = \|.\|_\infty^S$ the sup-norm on $\mathcal{B}(S)$. We extend the notion of supnorm to restrictions of functions to a subdomain of S. More precisely, for

$$h_1, h_2 \in \bigcup_{\emptyset \neq S' \subset \tilde{S} \subset S} \mathcal{B}(\tilde{S}), \tag{5.5.4}$$

define

$$\|h_1 - h_2\|_\infty^{S'} = \sup_{x \in S'} |h_1(x) - h_2(x)|. \tag{5.5.5}$$

Then, the following properties are trivial.

$$\begin{aligned} \|g^f - f\|_\infty &= \|g - f\|_\infty^{S'} \\ \|B^A g^f - Af\|_\infty &= \|Bg - Af\|_\infty^{S'} \end{aligned} \tag{5.5.6}$$

Let $\iota_{S',S}$ be the natural embedding from S' into S, and let $m_0 \in \mathcal{P}(S')$, and, as before, denote by $\iota_{S',S} * m_0$ its push-forward measure under $\iota_{S',S}$. If an S'-valued measureable stochastic process $\zeta = (\zeta_t)_{t\geq 0}$ is a solution to the martingale problem $(B, \mathcal{D}(B), m_0)$, then its image under the natural embedding, $(\iota_{S',S}(\zeta_t))_{t\geq 0}$, is a solution to the martingale problem $(B^A, \mathcal{D}(B, A), \iota_{S',S} * m_0)$.

The following proposition is a generalization of Lemma 4.5.1 [23].

Proposition 5.5.1. *Let S be a separable metric space, $S_n \subset S$, $\mathcal{D} \subset \bar{\mathcal{C}}(S)$, $\mathcal{D}_n \subset \mathcal{B}(S_n)$, for $n \in \mathbb{N}$. Consider time-dependent linear operators*

$$A : \mathbb{R}_+ \times \mathcal{D} \to \bar{\mathcal{C}}(S) \tag{5.5.7}$$

and

$$A_n : \mathbb{R}_+ \times \mathcal{D}_n \to \mathcal{B}(S_n). \tag{5.5.8}$$

Denote the sup-norm on S by $\|.\|_\infty$, and let

$$\|.\|_{\infty,n} := \|.\|_\infty^{S_n} \quad for\ n \in \mathbb{N}, \tag{5.5.9}$$

where the right side is defined as before. Let $m_0^n \in \mathcal{P}(S_n)$, for $n \in \mathbb{N}$, and $m_0 \in \mathcal{P}(S)$. Let Z_n be a solution of the martingale problem $(A_n, \mathcal{D}_n, m_0^n)$ (with sample paths in $D_{S_n}[0,\infty) \subset D_S[0,\infty)$) for every $n \in \mathbb{N}$. Assume that, for any $f \in \mathcal{D}$, there exists a sequence $(f_n)_{n\in\mathbb{N}}$, with $f_n \in \mathcal{D}_n$ for every $n \in \mathbb{N}$, such that

(i) $\lim_{n\to\infty} \|f_n - f\|_{\infty,n} = 0$

(ii) $\lim_{n\to\infty} \|A_n(s)f - A(s)f_n\|_{\infty,n} = 0$ *for a.e.* $s \geq 0$.

(iii) For any $t \geq 0$

$$\sup_{n\in\mathbb{N}, s\leq t} \|A_n(s)f_n\|_{\infty,n} < \infty. \tag{5.5.10}$$

If Z is an S-valued stochastic process with the initial distribution m_0 such that $Z_n \Rightarrow Z$ in $D_S[0,\infty)$, as $n \to \infty$, then Z is a solution of the martingale problem $(A, \mathcal{D}, m_0)$.

Proof: Let $\tilde{Z}_n(t) = \iota_{S_n,S}(Z_n(t))$, for $t \geq 0$, where $\iota_{S_n,S}$ is the natural embedding from S_n into S. As explained above, $\tilde{Z}_n$ is a solution to the martingale problem $(A_n^A, \mathcal{D}(A_n, A), \tilde{m}_0^n)$, for $n \in \mathbb{N}$, where A_n^A is the extension of A_n with respect to A with the domain $\mathcal{D}(A_n, A)$ (defined as before), and $\tilde{m}_0^n = \iota_{S_n,S} * m_0^n$ is the image (push-forward measure) of m_0^n under $\iota_{S_n,S}$. In order for Z to be a solution to $(A, \mathcal{D}, m_0)$, it is necessary and sufficient that for any $k \in \mathbb{N}$, $0 \leq t_1 \leq ... \leq t_k \leq t \leq s$, $h_i \in \bar{\mathcal{C}}(S)$, for $i = 1, ..., k$, and $f \in \mathcal{D}$

$$\mathbb{E}[\{f(Z(s)) - f(Z(t)) - \int_t^s A(r)f(Z(r))dr\} \prod_{i=1}^k h_i(Z(t_i))] = 0. \tag{5.5.11}$$

Under the assumptions, for any $f \in \mathcal{D}$, there exists a sequence of $f_n \in \mathcal{D}_n$ such that

$$\lim_{n\to\infty} \|f_n^f - f\|_\infty = \lim_{n\to\infty} \|f_n - f\|_{\infty,n} = 0 \tag{5.5.12}$$

and

$$\lim_{n\to\infty} \|A_n^A(s)f_n^f - A(s)f\|_\infty = \lim_{n\to\infty} \|A_n(s)f_n - A(s)f\|_{\infty,n} = 0 \tag{5.5.13}$$

for a.e. $s \geq 0$. Let $k \in \mathbb{N}$, $h_i \in \bar{\mathcal{C}}(S)$, for $i = 1, ..., k$. Let all $0 \leq t_1 \leq ... \leq t_k \leq t \leq s$ be in the times of continuity for Z, i.e. they belong to $\{r \geq 0 : Z(r) = Z(r^-)\ a.s.\}$ which contains all positive real numbers except possibly at most countable ones. Since $\tilde{Z}_n \Rightarrow Z$ in $D_S[0, \infty)$, as $n \to \infty$, and f is bounded continuous, we have

$$\lim_{t\to\infty} \mathbb{E}[A_n^A(r)f_n^f(\tilde{Z}_n(r))] = \mathbb{E}[A(r)f(Z(r))] \tag{5.5.14}$$

for a.e. $r \geq 0$ (for all $r \geq 0$ except possibly at most a countable number of points). Thus, by (iii),

$$\lim_{t\to\infty} \int_t^s \mathbb{E}[A_n^A(r)f_n^f(\tilde{Z}_n(r))]dr = \int_t^s \mathbb{E}[A(r)f(Z(r))]dr. \tag{5.5.15}$$

Therefore, as $\tilde{Z}_n$ is a solution to $(A_n^A, \mathcal{D}(A_n, A), \tilde{m}_0^n)$ for any $n \in \mathbb{N}$,

$$0 = \lim_{n\to\infty} \mathbb{E}[\{f_n^f(\tilde{Z}_n(s)) - f_n^f(\tilde{Z}_n(t)) - \int_t^s A_n^A(r)f_n^f(\tilde{Z}_n(r))dr\} \prod_{i=1}^k h_i(Z(t_i))] =$$
$$\mathbb{E}[\{f(Z(s)) - f(Z(t)) - \int_t^s A(r)f(Z(r))dr\} \prod_{i=1}^k h_i(Z(t_i))] = 0. \tag{5.5.16}$$

Since $\tilde{Z}$ is right continuous a.s., by (iii) and continuity of f, the last equality holds for any choice of $k \in \mathbb{N}$ and $t_1, ..., t_k, t, s \geq 0$. ■

Remark 5.5.2. *One can replace (ii) and (iii) in the assumptions of the last proposition by*

(ii)' There exists a measure-zero subset of $\mathbb{R}_+$, namely U, such that for any $t \geq 0$, $\lim_{n\to\infty} \|A_n(s)f_n - A(s)f\|_{\infty,n} = 0$ uniformly on $s \in [0,t] \cap U^c$.

(iii)' For any $t \geq 0$

$$\sup_{s\leq t} \|A(s)f\|_\infty < \infty. \tag{5.5.17}$$

In fact $(ii)'$ and $(iii)'$ conclude (ii) and (iii).

To apply the last proposition in our problem, we must verify the validity of the assumptions (i),(ii), and (iii), for the generators of MRE and FVRE. We can see that the generators are uniformly bounded in a very strong sense.

Recall that for $\tilde{e} \in D_E[0,\infty)$, $\tilde{\mathcal{G}}_{\tilde{e}} = \tilde{\mathcal{G}}^{res} + \tilde{\mathcal{G}}^{mut} + \tilde{\mathcal{G}}^{sel}_{\tilde{e}}$ and $\tilde{\mathcal{G}}^N_{\tilde{e}} = \tilde{\mathcal{G}}^{res,N} + \tilde{\mathcal{G}}^{mut,N} + \tilde{\mathcal{G}}^{sel,N}_{\tilde{e}}$ for $N \in \mathbb{N}$. Also, from now on in this chapter, we denote by $\|.\|$ the supnorm on $\mathcal{B}(\mathcal{P}(I))$, and by $\|.\|_N := \|.\|^{\mathcal{P}^N(I)}$ the supnorm on restrictions on $\mathcal{B}(\mathcal{P}^N(I))$. Also, similarly to the last section, we denote by $\|.\|_\infty$ the supnorm on $\mathcal{B}(I^{\mathbb{N}})$, specially we use this notation for the functions on $\mathcal{C}^*$, i.e. the state space of the dual process, and we denote by $\|.\|_{\infty,N} = \|.\|^{I^N}_\infty$ the supnorm on restrictions on $\mathcal{B}(I^N)$.

Proposition 5.5.3. *For any $\tilde{\Phi} := \tilde{\Phi}^f \in \tilde{\mathfrak{F}}$,*

$$\sup_{s\geq 0, \tilde{e}\in D_E[0,\infty)} \|\tilde{\mathcal{G}}_{\tilde{e}}(s)\tilde{\Phi}\| < \infty \tag{5.5.18}$$

Proof: Assume that $f \in \mathcal{C}_n(I^{\mathbb{N}})$. As we saw in the proof of Proposition 5.4.8, for any $i, j \leq n$ and $y \in I$, $f \circ \sigma_{ij}$ and $f \circ \sigma_i^y$ are defined by setting restrictions on the n first variables of f, that is they are restrictions of f on a subdomain, and therefore,

$$\begin{aligned} &\|f \circ \sigma_i^y\|_\infty \leq \|f\|_\infty \\ &\|f \circ \sigma_{ij}\|_\infty \leq \|f\|_\infty \\ &. \end{aligned} \tag{5.5.19}$$

Thus, for any $m \in \mathcal{P}(I)$,

$$\begin{aligned} &|\tilde{\mathcal{G}}^{res}\tilde{\Phi}^f(m)| = \\ &\tfrac{\gamma}{2}|\sum_{i,j=1}^{n} < m^{\otimes\mathbb{N}}, f \circ \sigma_{ij} - f > | \leq \\ &\tfrac{\gamma}{2}\tbinom{n}{2}(\|f \circ \sigma_{ij}\|_\infty + \|f\|_\infty) \leq \\ &\gamma\tbinom{n}{2}\|f\|_\infty. \end{aligned} \tag{5.5.20}$$

Also,

$$\begin{aligned} &|\tilde{\mathcal{G}}^{mut}\tilde{\Phi}^f(m)| = \\ &= |\beta' \sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, B_i' f - f > +\beta'' \sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, B_i'' f - f > | \leq \\ &\beta'(\sum_{i=1}^{n}(\|B_i' f\|_\infty + \|f\|_\infty) + \beta''(\sum_{i=1}^{n}(\|B_i'' f\|_\infty + \|f\|_\infty) \leq \\ &4\beta\|f\|_\infty, \end{aligned} \tag{5.5.21}$$

and for any $\tilde{e} \in D_E[0, \infty)$ and $s \geq 0$

$$\tilde{\mathcal{G}}^{sel}_{\tilde{e}}(s)\tilde{\Phi}^f(m) = \alpha \sum_{i=1}^{n} < m^{\otimes\mathbb{N}}, \tilde{e}_i(s)f - \tilde{e}_{n+1}(s)f > \leq 2n\alpha\|f\|_\infty. \tag{5.5.22}$$

Therefore, there exists a constant $C > 0$ such that for any $s \geq 0$ and $\tilde{e} \in D_E[0, \infty)$

$$|\tilde{\mathcal{G}}(s)\tilde{\Phi}^f(m)| < C\|f\|_\infty < \infty. \tag{5.5.23}$$

■

Proposition 5.5.4. *Let $\tilde{e}_N, \tilde{e} \in D_E[0,\infty)$, for $N \in \mathbb{N}$, such that $\tilde{e}_N \to \tilde{e}$ in $D_E[0,\infty)$, as $n \to \infty$. For any $\tilde{\Phi} := \tilde{\Phi}^f \in \tilde{\mathfrak{F}}$, there exists a sequence of functions $\tilde{\Phi}_N := \tilde{\Phi}_N^{f_N} \in \tilde{\mathfrak{F}}_N$, for $N \in \mathbb{N}$, such that*

(i) $\lim\limits_{N\to\infty} \|\tilde{\Phi}_N - \tilde{\Phi}\|_N = 0.$

(ii) For a.e. $t \geq 0$

$$\lim_{N\to\infty} \|\tilde{\mathcal{G}}^N_{\tilde{e}_N}(t)\tilde{\Phi}_N - \tilde{\mathcal{G}}_{\tilde{e}}(t)\tilde{\Phi}\|_N = 0. \tag{5.5.24}$$

(iii)

$$\sup_{N\in\mathbb{N}, s\geq 0} \|\tilde{\mathcal{G}}^N_{\tilde{e}_N}(s)\tilde{\Phi}_N\|_N < \infty \tag{5.5.25}$$

Proof: To simplify the notation, when it comes to applying it in the proof, we denote

$$\tilde{e}^i_N(s)(x) := \tilde{e}_N(s)(x_i), \tag{5.5.26}$$

for $s \geq 0$ and $x = (x_1, x_2, ...)$ (or $x = (x_1, ..., x_N)$). We assume $f \in \mathcal{C}_n(I^{\mathbb{N}})$, for fixed $n \in \mathbb{N}$. Consider an arbitrary sequence of injective maps $\eta^N : I^N \to I^{\mathbb{N}}$, for $N \in \mathbb{N}$, such that each η^N is identical on the first N coordinates (e.g. $\eta^N : (x_1, x_2, ..., x_N) \mapsto (y_1, y_2, ...)$ where $x_i = y_i$ for $i \leq N$ and, for $i > N$, $y_i = c$ for a fixed $c \in I$). Note that, as we deal with the limit and the supremum, for $N \leq n$, neither the value of $\tilde{\Phi}_N$, nor the value of $\tilde{\mathcal{G}}^N_{\tilde{e}_N}\tilde{\Phi}_N$ are important. So we assume these functions are 0 functions, for $N \leq n$. For $m \in \mathcal{P}^N(I)$, recall the definition of $m^{(N)}$ from (5.2.4) and set

$$\tilde{m}^{(N)} := \eta^N * m^{(N)} \quad N \in \mathbb{N}, \tag{5.5.27}$$

that is the push-forward measure of $m^{(N)}$ under η^N. It is clear that for any function $g \in C(I^{\mathbb{N}})$

$$\int_{I^{\mathbb{N}}} g d\tilde{m}^{(N)} = \int_{I^N} g \circ \eta^N dm^{(N)}. \tag{5.5.28}$$

As f depends only on the first n variables, we can define $\bar{f} \in \mathcal{C}(I^n)$ by

$$\bar{f}(x_1, x_2, ..., x_n) = f(x_1, x_2, ..., x_n, x_{n+1}, ...), \tag{5.5.29}$$

for an arbitrary choice of x_i for $i > n$. Let $f_N := f \circ \eta^N$.

$$\begin{aligned}
&\|\Phi_N^{f_N} - \Phi^f\|_N = \\
&\sup_{m\in\mathcal{P}^N(I)} | < m^{(N)}, f\circ\eta^N > - < m^{\otimes \mathbb{N}}, f > | = \\
&\sup_{m\in\mathcal{P}^N(I)} | < \tilde{m}^{(N)}, f > - < m^{\otimes \mathbb{N}}, f > | = \\
&\sup_{m\in\mathcal{P}^N(I)} | \int_{I^n} \bar{f}(x) m(dx_1) \prod_{i=2}^{n} \frac{Nm(dx_i) - \sum_{j=1}^{i-1}\delta_{x_j}}{N-i+1} - \int_{I^n} \bar{f}(x) \prod_{i=1}^{n} m(dx_i)| \leq \\
&\|f\|_\infty \sup_{m\in\mathcal{P}^N(I)} | \sum_{(x_1,\dots,x_n)\in supp(m)^n} (m(x_1) \prod_{i=2}^{n} \frac{Nm(x_i) - \sum_{j=1}^{i-1}\delta_{x_j}}{N-i+1} - \prod_{i=1}^{n} m(x_i))| \leq \\
&\|f\|_\infty (\tfrac{C}{N} + o(\tfrac{1}{N}))
\end{aligned} \tag{5.5.30}$$

for a constant C. This yields (i).

To prove (ii), first we observe that for $i, j \leq N$

$$f_N \circ \sigma_{ij} = f \circ \eta^N \circ \sigma_{ij} = f \circ \sigma_{ij} \circ \eta^N. \tag{5.5.31}$$

Let tr_{ij} be the transposition operator on i and j. Since f depends only on the first n variables, for $N \geq n$

$$\begin{aligned}
&\tilde{\mathcal{G}}^{res,N} \tilde{\Phi}_N^{f_N}(m) = \\
&\tfrac{\gamma}{2} \sum_{i,j=1}^{N} < m^{(N)}, f\circ\sigma_{ij}\circ\eta^N - f\circ\eta^N >= \\
&\tfrac{\gamma}{2} \sum_{i,j=1}^{N} < \tilde{m}^{(N)}, f\circ\sigma_{ij} - f >= \\
&\tfrac{\gamma}{2} \sum_{i,j=1}^{n} < \tilde{m}^{(N)}, f\circ\sigma_{ij} - f > + \\
&\tfrac{\gamma}{2} \sum_{i=n+1}^{N} \sum_{j=1}^{n} < \tilde{m}^{(N)}, f\circ\sigma_{ij} - f >= \\
&\tfrac{\gamma}{2} \sum_{i,j=1}^{n} < \tilde{m}^{(N)}, f\circ\sigma_{ij} - f > + \\
&\tfrac{\gamma}{2} \sum_{i=n+1}^{N} \sum_{j=1}^{n} < \tilde{m}^{(N)}, f\circ tr_{ij} - f > .
\end{aligned} \tag{5.5.32}$$

But the last term vanishes because of exchangeability of $m^{(N)}$. Therefore, similarly to the proof of (i),

$$\begin{aligned}&\|\tilde{\mathcal{G}}^{res,N}\tilde{\Phi}_N^{f_N}-\tilde{\mathcal{G}}^{res}\tilde{\Phi}^f\|_N\leq\\&\tfrac{\gamma}{2}\sup_{m\in\mathcal{P}^N(I)}\sum_{i,j=1}^{n}|<\tilde{m}^{(N)},f\circ\sigma_{ij}-f>-<m^{\otimes\mathbb{N}},f\circ\sigma_{ij}-f>|<\\&\tfrac{C_1}{N}+o(\tfrac{1}{N})\end{aligned}\tag{5.5.33}$$

for a constant $C_1>0$. Similarly, for mutation

$$\begin{aligned}&\|\tilde{\mathcal{G}}^{mut,N}\tilde{\Phi}_N^{f_N}-\tilde{\mathcal{G}}^{mut}\tilde{\Phi}^f\|_N\leq\\&\sup_{m\in\mathcal{P}^N(I)}\sum_{i=1}^{n}|<\tilde{m}^{(N)},\beta'(B_i'f-f)+\beta''(B_i''f-f)>\\&-<m^{\otimes\mathbb{N}},\beta'(B_i'f-f)+\beta''(B_i''f-f)>|<\\&\tfrac{C_2}{N}+o(\tfrac{1}{N}).\end{aligned}\tag{5.5.34}$$

To verify this for the selection, first note that as $\tilde{e}_N\to\tilde{e}$ in $D_E[0,\infty)$, as $N\to\infty$, for every positive real number $t\geq 0$, except possibly a countable number of them, we have $\tilde{e}_N(t)\to\tilde{e}(t)$. The selection operator is very similar to the resampling one, except, here, the constant rate $\frac{\gamma}{2}$ is replaced by a time-dependent càdlàg fitness, and hence, the terms corresponding to $i=n+1,...N, j=1,...,n$ do not necessarily vanish.

More explicitly, for a continuity time $s\geq 0$ of $\tilde{e}$

$$\begin{aligned}&\tfrac{\alpha}{N}\sum_{i=n+1}^{N}\sum_{j=1}^{n}<\tilde{m}^{(N)},\tilde{e}_N^i(s)f\circ\sigma_{ij}-\tilde{e}_N^i(s)f>=\\&\tfrac{\alpha}{N}\sum_{i=n+1}^{N}\sum_{j=1}^{n}<\tilde{m}^{(N)},(\tilde{e}_N^j(s)f)\circ\sigma_{ij}-\tilde{e}_N^i(s)f>=\\&\tfrac{\alpha}{N}\sum_{i=n+1}^{N}\sum_{j=1}^{n}<\tilde{m}^{(N)},(\tilde{e}_N^j(s)f)\circ tr_{ij}-\tilde{e}_N^i(s)f>=\\&\tfrac{\alpha}{N}\sum_{i=n+1}^{N}\sum_{j=1}^{n}<\tilde{m}^{(N)},\tilde{e}_N^j(s)f-\tilde{e}_N^i(s)f>=\\&\tfrac{\alpha(N-n)}{N}\sum_{j=1}^{n}<\tilde{m}^{(N)},\tilde{e}_N^j(s)f-\tilde{e}_N^{n+1}(s)f>.\end{aligned}\tag{5.5.35}$$

Therefore, for constants C_3 and C_4,

$$
\begin{aligned}
&\|\tilde{\mathcal{G}}^{sel,N}_{\tilde{e}_N}(s)\tilde{\Phi}^{f_N}_N - \tilde{\mathcal{G}}^{sel}_{\tilde{e}}(s)\tilde{\Phi}^f\|_N = \\
&\sup_{m\in\mathcal{P}^N(I)} |\tfrac{\alpha}{N}\sum_{i,j=1}^{N} < \tilde{m}^{(N)}, \tilde{e}^i_N(s)(f\circ\sigma_{ij} - f) > \\
&-\alpha\sum_{i=1}^{n} < m^{\otimes N}, \tilde{e}_i(s)f - \tilde{e}_{n+1}(s)f > | \le \\
&\sup_{m\in\mathcal{P}^N(I)} |\tfrac{\alpha}{N}\sum_{i,j=1}^{n} < \tilde{m}^{(N)}, \tilde{e}^i_N(s)(f\circ\sigma_{ij} - f) > |+ \\
&\sup_{m\in\mathcal{P}^N(I)} |\tfrac{\alpha}{N}\sum_{i=n+1}^{N}\sum_{j=1}^{n} < \tilde{m}^{(N)}, \tilde{e}^i_N(s)f\circ\sigma_{ij} - \tilde{e}^i_N(s)f > \\
&-\alpha\sum_{i=1}^{n} < m^{\otimes N}, \tilde{e}_i(s)f - \tilde{e}_{n+1}(s)f > | \le \\
&\tfrac{C_3}{N} + o(\tfrac{1}{N}) + \\
&\sup_{m\in\mathcal{P}^N(I)} |\alpha\tfrac{N-n}{N}\sum_{i=1}^{n} < \tilde{m}^{(N)}, \tilde{e}^i_N(s)f - \tilde{e}^{n+1}_N(s)f > \\
&-\alpha\sum_{i=1}^{n} < m^{\otimes N}, \tilde{e}_i(s)f - \tilde{e}_{n+1}(s)f > |,
\end{aligned}
\tag{5.5.36}
$$

where the right hand side is bounded by

$$
\begin{aligned}
&\tfrac{C_3}{N} + o(\tfrac{1}{N}) + \\
&\alpha\sum_{i=1}^{n} \|f\|_\infty\{\|\tilde{e}^i_N(s) - \tilde{e}_i(s)\|_\infty + \|\tilde{e}^{n+1}_N(s) - \tilde{e}_{n+1}(s)\|_\infty\}
\end{aligned}
\tag{5.5.37}
$$

Hence, there exists C_5 such that

$$
\|\tilde{\mathcal{G}}^{sel,N}_{\tilde{e}_N}(s)\tilde{\Phi}^{f_N}_N - \tilde{\mathcal{G}}^{sel}_{\tilde{e}}(s)\tilde{\Phi}^f\|_N < \tfrac{C_5}{N} + o(\tfrac{1}{N}) \tag{5.5.38}
$$

This finishes the proof of (ii).

For part (iii), similarly to the proof of Proposition 5.5.3, there exists a constant $C > 0$ such that for any $m \in \mathcal{P}^N(I)$, $s \ge 0$

$$
|\tilde{\mathcal{G}}^N_{\tilde{e}_N}(s)\Phi^{f_N}_N(m)| \le C\|f_N\|_{\infty,N} = C\|f\circ\eta^N\|_{\infty,N} = C\|f\|_\infty \tag{5.5.39}
$$

∎

Proposition 5.5.5. *Let $\tilde{e}_N, \tilde{e} \in D_E[0,\infty)$, and suppose that $\tilde{e}_N \to \tilde{e}$ in $D_E[0,\infty)$. For any $\tilde{\Phi} = \tilde{\Phi}^f \in \tilde{\mathfrak{F}}$*

$$\lim_{N\to\infty} \|\tilde{\mathcal{G}}_{\tilde{e}_N}(s)\tilde{\Phi} - \tilde{\mathcal{G}}_{\tilde{e}}(s)\tilde{\Phi}\| = 0 \tag{5.5.40}$$

for every $s \geq 0$ except possibly a countable number of real numbers. Moreover,

$$\sup_{N\in\mathbb{N}, s\geq 0} \|\tilde{\mathcal{G}}_{\tilde{e}_N}(s)\tilde{\Phi}^f\| < \infty \tag{5.5.41}$$

Proof: As $\tilde{e}_N \to \tilde{e}$ in $D_E[0,\infty)$, for every positive real numbers $s \geq 0$, except possibly countable ones $\tilde{e}_N(s) \to \tilde{e}(s)$ in E, as $N \to \infty$. The resampling and mutation rates, for both generators, are identical. Thus we need to verify that the limit is 0 for the selection terms. To this end, for any $m \in \mathcal{P}(I)$, and any continuity point $s \geq 0$ of $\tilde{e}$,

$$\begin{aligned}
&\|\tilde{\mathcal{G}}_{\tilde{e}_N}(s)\tilde{\Phi} - \tilde{\mathcal{G}}_{\tilde{e}}(s)\tilde{\Phi}\| \leq \\
&\sum_{i=1}^{n} \sup_{m\in\mathcal{P}(I)} |< m, \tilde{e}_N^i(s)f - \tilde{e}_N^{n+1}(s)f > - < m, \tilde{e}_i(S)f - \tilde{e}_{n+1}(s)f > | \\
&\leq \sum_{i=1}^{n} \|(\tilde{e}_N^i(s)f - \tilde{e}_i(s)f) - (\tilde{e}_N^{n+1}(s)f - \tilde{e}_{n+1}(s)f)\|_\infty \leq \\
&\sum_{i=1}^{n} \|f\|_\infty(\|\tilde{e}_N^i(s) - \tilde{e}_i(s)\|_\infty + \|\tilde{e}_N^{n+1}(s) - \tilde{e}_{n+1}(s)\|_\infty).
\end{aligned} \tag{5.5.42}$$

The last term converges to 0 and this yields the result.

For the second part, write

$$\sup_{N\in\mathbb{N}, s\geq 0} \|\tilde{\mathcal{G}}_{\tilde{e}_N}(s)\tilde{\Phi}^f\| \leq \sup_{\tilde{e}\in D_E[0,\infty), s\geq 0} \|\tilde{\mathcal{G}}_{\tilde{e}}(s)\tilde{\Phi}^f\| < \infty, \tag{5.5.43}$$

where the last inequality follows Proposition 5.5.3. ■

Proposition 5.5.6. *Let $M \in \mathbb{N}$, and let $\tilde{e}_N, \tilde{e} \in D_E[0,\infty)$, and suppose that $\tilde{e}_N \to \tilde{e}$ in $D_E[0,\infty)$. Then for any $\tilde{\Phi}_M = \tilde{\Phi}_M^f \in \tilde{\mathfrak{F}}_M$*

$$\lim_{N\to\infty} \|\tilde{\mathcal{G}}^M_{\tilde{e}_N}(s)\tilde{\Phi}_M - \tilde{\mathcal{G}}^M_{\tilde{e}}(s)\tilde{\Phi}_M\|_M = 0 \tag{5.5.44}$$

for every $s \geq 0$ *except possibly a countable number of real numbers. Moreover,*

$$\sup_{N\in\mathbb{N}, s\geq 0} \|\tilde{\mathcal{G}}^M_{\tilde{e}_N}(s)\tilde{\Phi}_M\|_M < \infty \tag{5.5.45}$$

Proof: The proof is similar to the proof of Proposition 5.5.5. Again, resampling and mutation terms of both generators are the same, and for a continuity point $s \geq 0$ of the function $\tilde{e}$

$$\begin{aligned}
&\|\tilde{\mathcal{G}}^M_{\tilde{e}_N}(s)\tilde{\Phi}_M - \tilde{\mathcal{G}}^M_{\tilde{e}}(s)\tilde{\Phi}_M\|_M \leq \\
&\sum_{i,j=1}^{N} \|\tilde{e}^i_N(s) - \tilde{e}_i(s)\|_\infty \|f\circ\sigma_{ij} - f\|_\infty \leq \\
&2(N-1)\sum_{i=1}^{N} \|\tilde{e}^i_N(s) - \tilde{e}_i(s)\|_\infty \|f\|_\infty,
\end{aligned} \tag{5.5.46}$$

where the last term is converging to 0 as $N \to \infty$. As before, there exists a constant $C > 0$ such that for any $m \in \mathcal{P}^M(I)$ and $s \geq 0$

$$|\tilde{\mathcal{G}}^M_{\tilde{e}_N}(s)\Phi_M(m)| \leq C\|f\|_{\infty,M} < \infty. \tag{5.5.47}$$

■

5.6 Convergence of MRE to FVRE

In this section we prove the wellposedness of the FVRE martingale problem (Theorem 5.2.8), and also prove convergence of MRE to FVRE (Theorem 5.2.10). In the previous sections, we prepared all necessary tools to construct FVRE from MRE. We proved the uniqueness of the FVRE martingale problem, convergence of generators, and other required properties. What remained is the proof of tightness that is relatively simple, due to compactness of the state spaces I and E and uniform boundedness of the fitness process. This section essentially is devoted to the problem of tightness, and proves it for $\{\mu_N^{\tilde{e}_N}\}_{N\in\mathbb{N}}$, $\{\mu^{\tilde{e}_N}\}_{N\in\mathbb{N}}$, and $\{\mu_M^{\tilde{e}_N}\}_{N\in\mathbb{N}}$, where $\tilde{e}$ and $\tilde{e}_N$, for $N \in \mathbb{N}$, are càdlàg functions in $D_E[0,\infty)$ and $M \in \mathbb{N}$. We apply a modification of Remark 4.5.2 [23] which best fits our problem.

Lemma 5.6.1. *Let S be a Polish space, and $S_n \subset S$. Consider $\mathcal{D} \subset \bar{\mathcal{C}}(S)$ that contains an algebra that separates points and vanishes nowhere. Let $\mathcal{D}_n \subset \mathcal{B}(S_n)$ and consider time-dependent linear operators*

$$A_n : \mathbb{R}_+ \times \mathcal{D}_n \to \mathcal{B}(S_n) \tag{5.6.1}$$

for $n \in \mathbb{N}$. For any $N \in \mathbb{N}$, suppose there exists an S_n-valued solution $Z_n = (Z_n(s))_{s\geq 0}$ (with sample paths in $D_{S_n}[0,\infty) \subset D_S[0,\infty)$) to the martingale problem $(A_n, \mathcal{D}_n, m_0^n)$ where $m_0^n \in \mathcal{P}(S_n)$. Assume:

(i) For any $f \in \mathcal{D}$, there exists a sequence $(f_n)_{n\in\mathbb{N}}$, $f_n \in \mathcal{D}_n$, such that

$$\|f_n - f\|_{\infty,n} \to 0, \tag{5.6.2}$$

as $n \to \infty$ (here $\|.\|_{\infty,n}$ is the general norm $\|.\|_\infty^{S_n}$ defined in the beginning of Section 5.5).

(ii) For any $t \geq 0$, there exists $r > 1$, such that

$$\limsup_n \mathbb{E}[(\int_0^t |A_n(s) f_n(Z_n(s))|^r ds)^{\frac{1}{r}}] < \infty. \tag{5.6.3}$$

(iii) (Compact containment condition)

For any $\varepsilon, t > 0$, there exists a compact set $K_{\varepsilon,t} \subset S$ such that

$$\inf_n \mathbb{P}[Z_n(s) \in K_{\varepsilon,t} \quad for\ 0 \leq s \leq t\} \geq 1 - \varepsilon \tag{5.6.4}$$

Then $\{Z_n\}_{n\in\mathbb{N}}$ is relatively compact (equivalently tight) in $D_S[0,\infty)$.

Proof: The set $\bar{\mathcal{D}}$ contains an algebra that separates points and vanishes nowhere, and hence it is dense in $\bar{\mathcal{C}}(S)$ in the topology of uniform convergence on compact sets. As the compact containment condition holds, and Z_n takes sample paths in $D_S[0,\infty)$ for any n, applying Theorem 3.9.1 [23], it suffices to show for any $f \in \mathcal{D}$,

$f \circ Z_n = (f \circ Z_n(s))_{s \geq 0}$ is tight in $D_{\mathbb{R}}[0,\infty)$. Theorem 3.9.4 [23] gives certain criteria under which $f \circ Z_n$ is tight, namely for any $t \geq 0$

$$\limsup_n \mathbb{E}[\sup_{s\in[0,t]\cap\mathbb{Q}} |f_n(Z_N(s)) - f(Z_n(s))] < \|f_n - f\|_{\infty,n}. \tag{5.6.5}$$

But the last term converges to 0 as $n \to \infty$. Thus (ii) and the fact that Z_n are solutions to martingale problems $(A_n, \mathcal{D}_n, m_0^n)$ yields the result. ■

Lemma 5.6.2. *For any $M \in \mathbb{N}$ and $(\tilde{e}_N)_{N\in\mathbb{N}} \subset D_E[0,\infty)$, the sequences $(\mu_N^{\tilde{e}_N})_{N\in\mathbb{N}}$ and $(\mu^{\tilde{e}_N})_{N\in\mathbb{N}}$ are tight in $D_{\mathcal{P}(I)}[0,\infty)$, and $(\mu_M^{\tilde{e}_N})_{N\in\mathbb{N}}$ is tight in $D_{\mathcal{P}^M(I)}[0,\infty)$.*

Proof: First note that the compact containment condition always holds due to the compactness of the state space $\mathcal{P}(I)$. Propositions 5.5.4, 5.5.5, and 5.5.6 guarantee the conditions (i) and (ii) of Lemma 5.6.1. Of course, condition (i) holds for $\{\mu^{\tilde{e}_N}\}_{N\in\mathbb{N}}$ by setting $\mathcal{D}_N = \tilde{\mathfrak{F}}$ and $\tilde{\Phi}^{f_N} = \tilde{\Phi}$ for any $N \in \mathbb{N}$. Similarly, in the case of $\{\mu_M^{\tilde{e}_N}\}_{N\in\mathbb{N}}$, we set $\mathcal{D}_N = \tilde{\mathfrak{F}}_M$, for any $N \in \mathbb{N}$, and the sequence of functions, $\Phi_M^{f_N}$, all identical to $\tilde{\Phi}_M$. Further, $\tilde{\mathfrak{F}}$ and $\tilde{\mathfrak{F}}_N$ separate points and vanishes nowhere (Propositions 5.2.6, 5.2.2) on $\mathcal{P}(I)$ and $\mathcal{P}^N(I)$, respectively. This finishes the proof. ■

Now we are ready to prove Theorems 5.2.8 and 5.2.10.

Proof of Theorems 5.2.8 and 5.2.10

For Theorem 5.2.8, it suffices to prove existence of a solution to the martingale problem $(\tilde{\mathcal{G}}_{\tilde{e}}, \tilde{\mathfrak{F}}, \tilde{P}_0)$ for every $\tilde{e}$ (uniqueness of such a martingale problem has been proved (cf. Proposition 5.1.8). Then wellposedness of the quenched martingale problem follows immediately. Also, for Theorem 5.2.10, integrating over Ω, part (ii) follows part(i), automatically. Therefore, we concentrate on the proof of existence of

$(\tilde{\mathcal{G}}_{\tilde{e}}, \tilde{\mathfrak{F}}, \tilde{P}_0)$, for any $\tilde{e} \in D_E[0,\infty)$, and proof of part (i) in Theorem 5.2.10.

Let M be an arbitrary natural number. By assumption and continuity of $\tilde{P}_0$ and $\tilde{P}_0^M$, we have the convergence of initial distributions, i.e.

$$\begin{aligned} &\tilde{P}_0^N(\tilde{e}_N) \to \tilde{P}_0(\tilde{e}) \quad in\ \mathcal{P}(\mathcal{P}(I)) \\ &\tilde{P}_0(\tilde{e}_N) \to \tilde{P}_0(\tilde{e}) \ \ in \quad \mathcal{P}(\mathcal{P}(I)) \\ &\tilde{P}_0^M(\tilde{e}_N) \to \tilde{P}_0^M(\tilde{e}) \ \ in \quad \mathcal{P}(\mathcal{P}^M(I)). \end{aligned} \tag{5.6.6}$$

Propositions 5.5.1, 5.5.4, 5.5.5, and 5.5.6 combined with the uniqueness of martingale problems $(\tilde{\mathcal{G}}_{\tilde{e}}, \tilde{\mathfrak{F}}, \tilde{P}_0(\tilde{e}))$, $(\tilde{\mathcal{G}}^M_{\tilde{e}_N}, \tilde{\mathfrak{F}}_M, \tilde{P}_0^M(\tilde{e}))$ (Propositions 5.4.6 and 5.2.4) ensure that any convergent subsequence of $(\mu_N^{\tilde{e}_N})_{N\in\mathbb{N}}$ $((\mu^{\tilde{e}_N})_{N\in\mathbb{N}}$ and $(\mu_M^{\tilde{e}_N})_{N\in\mathbb{N}}$, respectively) converges weakly in the corresponding Skorokhod topology to the unique solution of time-dependent martingale problem $(\tilde{\mathcal{G}}_{\tilde{e}}, \tilde{\mathfrak{F}}, \tilde{P}_0(\tilde{e}))$ $((\tilde{\mathcal{G}}_{\tilde{e}}, \tilde{\mathfrak{F}}, \tilde{P}_0(\tilde{e}))$, $(\tilde{\mathcal{G}}^M_{\tilde{e}_N}, \tilde{\mathfrak{F}}_M, \tilde{P}_0^M(\tilde{e}))$, respectively). By tightness of $(\mu_N^{\tilde{e}_N})_{N\in\mathbb{N}}$ $((\mu^{\tilde{e}_N})_{N\in\mathbb{N}}$ and $(\mu_M^{\tilde{e}_N})_{N\in\mathbb{N}}$, respectively) , Lemma 5.6.2, the weak limits exist. This yields part (i) of Theorem 5.2.10. Note that, for any $\tilde{e} \in D_E[0,\infty)$, letting $\tilde{e}_N = \tilde{e}$ (or $\tilde{e}_N = \tilde{e}_M$, respectively) for all $N \in \mathbb{N}$, we have also proved that

$$\begin{aligned} &\mu_N^{\tilde{e}} \Rightarrow \mu^{\tilde{e}} \\ &\mu_N^{\tilde{e}_M} \Rightarrow \mu^{\tilde{e}_M} \end{aligned} \tag{5.6.7}$$

in $D_{\mathcal{P}(I)}[0,\infty)$ as $N \to \infty$. Also, in particular for any $\tilde{e} \in D_E[0,\infty)$, setting $\tilde{e}_N = \tilde{e}$ for all $N \in \mathbb{N}$, implies existence of a solution to $(\tilde{\mathcal{G}}_{\tilde{e}}, \tilde{\mathfrak{F}}, \tilde{P}_0)$, and this, together with the uniqueness, Proposition 5.4.6, deduces wellposedness.

Note that we do not need continuity of $\tilde{\mathcal{P}}_0$, except to prove convergence results $\mu^{\tilde{e}_N} \Rightarrow \mu$ and $\mu_M^{\tilde{e}_N} \Rightarrow \mu_M$ in the corresponding Skorokhod topology, as $N \to \infty$.

5.7 Continuity of sample paths of FVRE

The purpose of this section is to prove the continuity of sample paths for the FVRE process. We make use of the criteria developed recently by Depperschmidt et al. in

[13] (see also [1]). To formulate the sufficient conditions under which FVRE takes continuous paths a.s., we shall introduce the concept of first and second order operators. We follow the definitions and the proof of Section 4 in [13].

Definition 18. *Let L be a Banach space and suppose $\mathcal{D} \subset L$ contains an algebra $\mathcal{A}$. A linear operator G on L with the domain $\mathcal{D}$ is said to be a first order operator with respect to $\mathcal{A}$ if for any $f \in \mathcal{A}$*

$$Gf^2 - 2fGf = 0. \tag{5.7.1}$$

It is said to be a second order operator, if it is not a first order operator, and for every $f \in \mathcal{A}$

$$Gf^3 + 3f^2Gf - 3fGf^2 = 0. \tag{5.7.2}$$

The following lemma is an extension of Proposition 4.5 in [13] to the case of time-dependent martingale problems.

Lemma 5.7.1. *Let S be a Polish space, and consider $\mathcal{D} \subset \bar{\mathcal{C}}(S)$ containing a countable algebra $\mathcal{A}$ that separates points of S and contains constant functions. Let $G : \mathbb{R}_+ \times \mathcal{D} \to \mathcal{B}(S)$ be a time-dependent linear operator such that, for any $t \geq 0$,*

$$G(t) = G^1(t) + G^2(t), \tag{5.7.3}$$

where $G^1(t)$ and $G^2(t)$ are first and second order linear operators for every $t \geq 0$, respectively. Assume that for any $t \geq 0$ and $f \in \mathcal{A}$, $G(.)f$ is uniformly bounded on $[0, t]$, i.e.

$$\sup_{s \leq t} |G(s)f| < \infty. \tag{5.7.4}$$

Let $m_0 \in \mathcal{P}(S)$. Then, any general solution $Z = (Z_t)_{t \geq 0}$ of the martingale problem $(G, \mathcal{D}, m_0)$, has sample paths in $\mathcal{C}([0, \infty), S)$ a.s..

Remark 5.7.2. *Note that under the assumption, any general solution of the martingale problem $(G, \mathcal{D}, m_0)$ is a solution, i.e. takes its sample paths in $D_S[0,\infty)$ a.s. (cf. Theorem 4.3.6 [23]).*

Proof: We follow the proof of Proposition 4.5 in [13]. Let $Z = (Z_t)_{t\geq 0}$ be a solution to the martingale problem $(G, \mathcal{D}, m_0)$. For any $f \in \mathcal{A}$, first we prove $(f(Z_t))_{t\geq 0}$ has continuous paths a.s.. For $f \in \mathcal{A}$ and $x, y \in S$, let $F_{f,y}(x) := f(x) - f(y)$. Since $\mathcal{A}$ is an algebra, for any $y \in S$ and $f \in \mathcal{D}$, $F_{f,y} \in \mathcal{A}$, and hence $F^2_{f,y}, F^4_{f,y} \in \mathcal{A}$. Thus

$$F^2_{f,y}(Z_t) - \int_0^t G(u) F^2_{f,y}(Z_u)du \tag{5.7.5}$$

is a martingale with respect to the canonical filtration. In particular, for $t \geq s$

$$\mathbb{E}[F^2_{f,Z_s}(Z_t)] = \int_s^t \mathbb{E}[G(u) F^2_{f,Z_s}(Z_u)]du \leq C_1(t-s) \tag{5.7.6}$$

for a constant $C_1 > 0$. Also,

$$\begin{aligned}
&\mathbb{E}[(f(Z_t) - f(Z_s))^4] = \\
&\mathbb{E}[F^4_{f,Z_s}(Z_t)] = \quad (\textit{by Lemma } 4.4.\ [13]) \\
&\int_s^t \mathbb{E}[F^2_{f,Z_s}(Z_u)(6GF^2_{f,Z_s}(Z_u) - 8F_{f,Z_s}(Z_u)GF_{f,Z_s}(Z_u))]du \leq \\
&C_2 \int_s^t \mathbb{E}[F^2_{f,Z_s}(Z_u)]du \leq \\
&C_1C_2 \int_s^t (u-s)du = \tfrac{C_1C_2(t-s)^2}{2}.
\end{aligned} \tag{5.7.7}$$

Continuity of $(f(Z_t))_{t\geq 0}$ follows from Proposition 3.10.3 [23]. The remainder of the proof is identical to that of Lemma 4.4 in (Depperschmidt et al. [13]). ■

Lemma 5.7.3. *Let $\tilde{e} \in D_E[0,\infty)$. Then*

(a) $\tilde{\mathcal{G}}^{mut}$ is first order.

(b) For any $t \geq 0$, $\tilde{\mathcal{G}}^{sel}_{\tilde{e}}(t)$ is first order.

(c) $\tilde{\mathcal{G}}^{res}$ is second order.

Proof: See Proposition 4.10 in [13]. ∎

Proof of Theorem 5.2.9

Continuity of the sample paths of μ^e a.s. follows the continuity of sample paths of $\mu^{\tilde{e}}$ for every $\tilde{e} \in D_E[0,\infty)$. The latter is a consequence of Propositions 5.2.6, 5.5.3, Theorem 5.2.8, and Lemmas 5.7.1 and 5.7.3.

5.8 An ergodic theorem for FVRE

This section proves the main ergodic theorem, Theorem 5.2.12, for the FV annealed-environment process. Before giving a complete proof, we show that the semigroup of FVRE has Feller property, that is for any $0 \leq s \leq t$, the semigroup is from $\mathcal{C}(\mathcal{P}(I))$ to $\mathcal{C}(\mathcal{P}(I))$.

For any $0 \leq s \leq t$ and $\tilde{e} \in D_E[0,\infty)$, let $p^{\tilde{e}}(s,x;t,dy)$ be the transition probability of $\mu^{\tilde{e}}$ and let $p^e(s,x;t,dy)$ be the transition probability of μ^e. Denote by $(T^e_{s,t})_{0\leq s\leq t}$ the semigroup of μ^e, i.e. for $f \in \mathcal{C}(\mathcal{P}(I))$ and $m \in \mathcal{P}(I)$

$$T^e_{s,t}f(m) = \int_{\mathcal{P}(I)} f(m')p^e(s,m;t,dm'). \tag{5.8.1}$$

Proposition 5.8.1. *Let e be a fitness process. Then, $(T^e_{s,t})_{0\leq s\leq t}$ is a Feller semigroup, i.e. for any $0 \leq s \leq t$ and for any $f \in \mathcal{C}(\mathcal{P}(I))$, $T^e_{s,t}f \in \mathcal{C}(\mathcal{P}(I))$. In other words, for any $0 \leq s \leq t$*

$$T^e_{s,t} : \mathcal{C}(\mathcal{P}(I)) \to \mathcal{C}(\mathcal{P}(I)). \tag{5.8.2}$$

Proof: Recall from Remark 5.4.5 the definition of $(\mu_r^{+s,e})_{r\geq 0}$:

$$\mu_r^{+s,e} = \mu_{r+s}^e \tag{5.8.3}$$

Let $m_0^n \to m_0$ in $\mathcal{P}(I)$, as $n \to \infty$, and $\hat{\psi}_0 \in \mathcal{C}^*$. For any $0 \leq s \leq t$, the duality relation (in average) follows Proposition 5.4.3 and Remark 5.4.5, and since $\hat{\psi}_0$ is bounded continuous depending on a finite number of variables, we have

$$\begin{aligned}&\lim_{n\to\infty} \mathbb{E}^{m_0^n}[< \mu_{t-s}^{+s,e}, \hat{\psi}_0 >] = \lim_{n\to\infty} \mathbb{E}^{\hat{\psi}_0}[< m_0^n, \psi_{t-s}^{t,e} >] = \\ &\mathbb{E}^{\hat{\psi}_0}[< m_0, \psi_{t-s}^{t,e} >] = \mathbb{E}^{m_0}[< \mu_{t-s}^{+s,e}, \hat{\psi}_0 >].\end{aligned} \tag{5.8.4}$$

As $\{< ., \hat{\psi}_0 >\}_{\hat{\psi}_0 \in \mathcal{C}^*}$ is measure-determining, for any $f \in \mathcal{C}(\mathcal{P}(I))$

$$\begin{aligned}&\lim_{n\to\infty} T_{s,t}^e f(m_0^n) = \\ &\lim_{n\to\infty} \int_{\mathcal{P}(I)} f(m')p^e(s, m_0^n; t, dm') = \\ &\lim_{n\to\infty} \mathbb{E}^{m_0^n}[f(\mu_{t-s}^{+s,e})] = \\ &\mathbb{E}^{m_0}[f(\mu_{t-s}^{+s,e})] = \\ &T_{s,t}^e f(m_0).\end{aligned} \tag{5.8.5}$$

■

Proof of Theorem 5.2.12

By duality relation (in average), Proposition 5.4.3, and Theorem 5.4.11, for any $m_0 \in \mathcal{P}(I)$ and $\hat{\psi}_0 \in \mathcal{C}^*$

$$\lim_{t\to\infty} \mathbb{E}^{m_0}[< \mu_t^e, \hat{\psi}_0 >] = \lim_{t\to\infty} \mathbb{E}^{\hat{\psi}_0}[< m_0, \psi_t^{t,e} >] = \lim_{t\to\infty} \mathbb{E}_0^{\hat{\psi}}[\psi_t^{t,e}], \tag{5.8.6}$$

where the limit on the right hand side of the last equality exists and does not depend on $m_0 \in \mathcal{P}(I)$. Since $\mathcal{P}(I)$ is compact, $\{\mu_t^e\}_{t\geq 0}$ is tight and therefore there exist some convergent subsequences. Let $(t_i)_{i\in\mathbb{N}}$ and $(s_i)_{i\in\mathbb{N}}$ be two strictly increasing sequences of positive real numbers, and let $\{\mu^e_{t_i}\}_{i\in\mathbb{N}}$ and $\{\mu^e_{s_i}\}_{i\in\mathbb{N}}$ (with $\mu^e_0 = m_0$) be such that

$$\mu^e_{t_i} \Rightarrow \mu_1^{e,m_0}(\infty) \tag{5.8.7}$$

$$\mu^e_{s_i} \Rightarrow \mu_2^{e,m_0}(\infty) \tag{5.8.8}$$

in $\mathcal{P}(I)$ as $N \to \infty$, where $\mu_i^{e,m_0}(\infty)$, for $i = 1, 2$, are random measures in $\mathcal{P}(I)$. For any $\hat{\psi}_0 \in \mathcal{C}^*$

$$\begin{aligned} &\mathbb{E}[< \mu_1^{e,m_0}(\infty), \hat{\psi}_0 >] = \lim_{i\to\infty} \mathbb{E}^{m_0}[< \mu^e_{t_i}, \hat{\psi}_0 >] = \lim_{t_i\to\infty} \mathbb{E}^{\hat{\psi}_0}[\psi^{t_i,e}_{t_i}] = \\ &\lim_{s_i\to\infty} \mathbb{E}^{\hat{\psi}_0}[\psi^{s_i,e}_{s_i}] = \lim_{i\to\infty} \mathbb{E}^{m_0}[< \mu^e_{s_i}, \hat{\psi}_0 >] = \mathbb{E}[< \mu_2^{e,m_0}(\infty), \hat{\psi}_0 >]. \end{aligned} \tag{5.8.9}$$

As $\lim_{t_i\to\infty} \mathbb{E}^{\hat{\psi}_0}[\psi^{t_i,e}_{t_i}] = \lim_{s_i\to\infty} \mathbb{E}^{\hat{\psi}_0}[\psi^{s_i,e}_{s_i}]$ does not depend on m_0, so do $\mu_i^{e,m_0}(\infty)$, for $i = 1, 2$. Hence, there exists a random probability measure $\mu^e(\infty) \in \mathcal{P}(I)$ such that for any $m_0 \in \mathcal{P}(I)$, conditioning on $\mu^e_0 = m_0$,

$$\mu^e_t \Rightarrow \mu^e(\infty). \tag{5.8.10}$$

For part (ii), it is sufficient to prove that conditioning on any initial distribution of (μ^e_t, e_t), namely $\tilde{m}_0 \in \mathcal{P}(\mathcal{P}(I) \times E)$,

$$\lim_{t\to\infty} \mathbb{E}[f_1(\mu_t) f_2(e_t)] \quad f_1 \in \mathcal{C}(\mathcal{P}(I)),\ f_2 \in \mathcal{C}(E) \tag{5.8.11}$$

exists and does not depend on $\tilde{m}_0$. In that case, since $\mathcal{P}(I) \times E$ is compact, any convergent subsequence of (μ^e_t, e_t) converges weakly to a unique limit, and from part (i) and assumption, then

$$(\mu^e_t, e_t) \Rightarrow (\mu^e(\infty), e(\infty)). \tag{5.8.12}$$

To prove the existence of the limit for any arbitrary $\hat{\psi}_0 \in \mathcal{C}^*$ and $m_0 \in \mathcal{P}(I)$, write

$$\begin{aligned} &\mathbb{E}[< \mu^e_t, \hat{\psi}_0 > f_2(e_t)] = \\ &\mathbb{E}[\mathbb{E}[< \mu^e_t, \hat{\psi}_0 > |e] f_2(e_t)] = \\ &\mathbb{E}[\mathbb{E}[< m_0, \psi^{t,e}_t > |e] f_2(e_t)] \end{aligned} \tag{5.8.13}$$

But, conditioning on $\tau, J_\tau, \kappa, T^{sel}_\tau$, knowing the fact that τ is finite a.s. and does not depend on e, and replacing the continuous function f in (5.4.54)by $\psi_t^{t,e}e(t)$, we can see the limit of the last term in the last equality exists and does not depend on the choice of m_0 and $\hat{\psi}_0$. (Recall that $\{< ., f >\}_{f\in\mathcal{C}^*}$ is measure and convergence-determining.) Now let ν^* be the distribution of $(\mu^e(\infty), e(\infty))$. Let T_t^* be the semigroup of the joint annealed-environment process, (μ_t^e, e_t). For $F \in \mathcal{C}(\mathcal{P}(I) \times E)$, $\tilde{m}_0 \in \mathcal{P}(\mathcal{P}(I) \times E)$, and $s \geq 0$,

$$\int_{\mathcal{P}(I)\times E} T_s^* F d\nu^* = \lim_{t\to\infty} \int_{\mathcal{P}(I)\times E} T_{t+s}^* F d\tilde{m}_0 = \int_{\mathcal{P}(I)\times E} F d\nu^*. \tag{5.8.14}$$

The last equation holds for any $\tilde{m}_0$, including all invariant measures. Hence the uniqueness holds.

5.9 Conclusion

Natural selection is one of the most important mechanisms in evolutionary biology. The interaction between a population and environment as well as the effect of other evolutionary mechanisms (e.g. mutation and genetic drift) on the fitness of types, must be addressed in studying the genetic variations of an organism. The environmental fluctuations change the evolutionary behaviour of a population. If the environment fluctuates very fast, then the population cannot adapt to it, and this increases the randomness (noise) in the system [45, 44]. In the case that there exists an equilibrium for the environment, rapid environmental changes cause the population to behave, on average, as it would behave in the equilibrium environment. Indeed, the adaptation in the population is not as fast as the changes in environment and, therefore, the environment can adopt itself only with the average information from the rapidly fluctuating environment. Several interesting questions about interaction of the environment and the genetic variations have been addressed in the literature. However, because of the mathematical complexity, many of them have not been studied in depth.

In this chapter we studied different aspects of the Fleming-Viot population model in random time-dependent environment in its generality with compact environmental and type spaces. We have formulated quenched martingale problems and characterized the FVRE process as a unique solution of a quenched martingale problem. We have formulated and developed the notion of duality for time-dependent and quenched martingale problems and applied the method of duality to prove uniqueness of the FVRE martingale problem. Simultaneously with the proof of wellposedness, we showed that the Moran processes in random environments converge weakly to FVRE in $D_{\mathcal{P}(I)}[0,\infty)$. Also, we studied the long-time behaviour of FVRE, with weakly ergodic or stationary fitness processes, via the behaviour of its dual family of processes and observed that the behaviour of equilibrium population does not depend on the initial states and, in fact, the "loss of memory" property holds for the FVRE process.

An interesting project is to compare the effect of different environmental changes by studying the Radon-Nikodym derivative of laws of FVRE (MRE, respectively) in different environments. This has been done partly, by a new method called branching representation of the Radon-Nikodym derivative[33]. We have studied change of measures for MRE in finite and infinite times. This can be applied to study the mutual differential entropy between the measure-valued process and the environment (the fitness process). A Shannon-McMillan-Breiman type theorem can be studied in random environment with the same method of branching representation. Another interesting problem is studying the fitness flux which provides a tool to measure the interaction of the environment and the population[45].

Once we construct the FVRE process, one can talk about the joint annealed-environment process. In fact the generator of the the joint process is written as follows.

Definition 19. *For $f \in \mathcal{B}_n(I^{\mathbb{N}} \times E)$, a joint polynomial is a function $\Phi^f = \Phi : \mathcal{P}(I) \times E \to \mathbb{R}$ is defined by*

$$\Phi^f(m, \hat{e}) := < m^{\otimes\mathbb{N}}, f^{\hat{e}} > = \int_{I^{\mathbb{N}}} f^{\hat{e}}(x) m^{\otimes\mathbb{N}}(dx) \quad for\ m \in \mathcal{P}(I), \hat{e} \in E, \tag{5.9.1}$$

where $m^{\otimes\mathbb{N}}$ is the $\mathbb{N}$-fold product measure of m, and $f^{\hat{e}}(.) = f(., \hat{e})$.

Let $\mathfrak{F}_b = \{\Phi^f : f \in \mathcal{B}(I^N \times E)\}$ and $\mathfrak{F}^k = \{\Phi^f : f \in \bar{\mathcal{C}}^k(I^N \times E)\}$, and let $\mathfrak{F} = \mathfrak{F}^0$. In fact we can show that $\mathfrak{F}_b$ and $\mathfrak{F}^k$ are algebras of functions.

The generator of Fleming-Viot-environment process (FVE) on $\mathfrak{F}_b$ with domain $\mathfrak{F}$ is given by $\mathcal{G} = \mathcal{G}^{res} + \mathcal{G}^{mut} + \mathcal{G}^{sel} + \mathcal{G}^{env}$, where the terms on the right hand side of the equality denote the resampling, mutation, selection, and environment generators, respectively. Let $f \in \bar{C}_n(I^{\mathbb{N}} \times E)$. The resampling generator is given by

$$\mathcal{G}^{res}\Phi^f(m, \hat{e}) = \frac{\gamma}{2} \sum_{i,j=1}^{n} < m^{\otimes\mathbb{N}}, f^{\hat{e}} \circ \sigma_{i,j} - f^{\hat{e}} > \tag{5.9.2}$$

For mutation, we have

$$\begin{aligned} \mathcal{G}^{mut}\Phi^f(m) &= \beta \sum_{i \geq 1} < m^{\otimes\mathbb{N}}, B_i f^{\hat{e}} - f^{\hat{e}} > \\ &= \beta' \sum_{i \geq 1} < m^{\otimes\mathbb{N}}, B_i^{'} f^{\hat{e}} - f^{\hat{e}} > + \beta'' \sum_{i \geq 1} < m^{\otimes\mathbb{N}}, B_i^{''} f^{\hat{e}} - f^{\hat{e}} >, \end{aligned} \tag{5.9.3}$$

where

$$B_i f(x, \hat{e}) = B_i f^{\hat{e}}(x) := \int_I f^{\hat{e}} \circ \sigma_i^y(x) q(x_i, dy) \tag{5.9.4}$$

$$B_i' f(x, \hat{e}) = B_i' f^{\hat{e}}(x) := \int_I f^{\hat{e}} \circ \sigma_i^y(x) q'(dy) \tag{5.9.5}$$

$$B_i'' f(x, \hat{e}) = B_i^{''} f^{\hat{e}}(x) := \int_I f^{\hat{e}} \circ \sigma_i^y(x) q''(x_i, dy). \tag{5.9.6}$$

To write the selection generator $\mathcal{G}^{sel,N}$, let $\chi_i : \bar{I}_i \times E \to [0,1]$ be defined by $\chi_i(\underline{a}, \hat{e}) = \hat{e}(\underline{a}_i)$ for $\underline{a} \in \bar{I}_i$ where $\underline{a}_i$ is the ith component of $\underline{a}$. Then $\mathcal{G}^{sel,N}$ is a linear operator from $\mathfrak{F}_N$ to $\mathfrak{F}_N$, and is given by

$$\mathcal{G}^{sel} : \mathfrak{F} \to \mathfrak{F} \tag{5.9.7}$$

$$\mathcal{G}^{sel}\Phi^f(m, \hat{e}) = \alpha \sum_i < m^{\otimes \mathbb{N}}, \chi_i^{\hat{e}} f^{\hat{e}} - \chi_{n+1}^{\hat{e}} f^{\hat{e}} > . \tag{5.9.8}$$

Let $e = (e_t)_{t \in \mathbb{R}_+}$ be a Markov fitness with semigroup T_t^{env} and generator A^{env}. The environment generator is

$$\mathcal{G}^{env}\Phi^f(m, \hat{e}) = \lim_{t \to 0} \frac{1}{t} \int_{I^{\mathbb{N}}} (T_t^{env} f(u, \hat{e}) - f(u, \hat{e})) m^{\otimes \mathbb{N}}(du) = < m^{\otimes \mathbb{N}}, (Af)^{\hat{e}} > . \tag{5.9.9}$$

Many interesting questions can be addressed in the relation of environment and the FV process. In an ongoing project, we are studying the joint dual process and other behaviours of the joint annealed-environment processes. This includes the change of measures for the joint process and the population systems with interactions with fitness functions [34].

Bibliography

[1] Bakry, D., and Émery, M.: Diffusions hypercontractives. In Sminaire de Probabilits, XIX, 1983/84. Lecture Notes in Mathematics, 1123 177206. Springer, Berlin, 1985.

[2] Berestycki, N., and Durrett, R.: A phase transition in the random transposition random walk. Probability theory and related fields, v. 136, n. 2, p. 203-233, 2006.

[3] Billingsley, P.: Convergence of probability measures. Second edition. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, New York, 1999.

[4] Blanchette, M., Bourque, G., and Sankoff, D.: Breakpoint phylogenies. Genome informatics, v. 8, p. 25-34, 1997.

[5] Bollobás, B.: Random Graphs, 2nd edition. Cambridge University Press, 2001.

[6] Bourque, G., and Pevzner, P.A.: Genome-scale evolution: Reconstructing gene orders in the ancestral species. Genome Research, **12**: 26–36 2002.

[7] Caprara, A.: The reversal median problem. INFORMS Journal on Computing, **15**: 93–113, 2003.

[8] Dawson, D.A.: Measure-valued Markov processes. In Hennequin, P.L. editor, 'Ecole d'été de Probabilités de Saint-Flour XXI - 1991, volume 1541 of Lecture Notes in Mathematics, pages 1-260, Springer Berlin Heidelberg, 1993.

[9] Dawson, D.A.: Introductory lectures on stochastic population systems. Lecture notes in 2009 PIMS summer school. No. 451, Technical Report, 2010.

[10] Dawson, D.A., and Greven, A.: Spatial Fleming-Viot models with selection and mutation. Lecture Notes in Mathematics, 2092. Springer, Cham, 2014.

[11] Dawson, D.A., and Hochberg, K.J.: Wandering random measures in the Fleming-Viot model. The Annals of Probability, p. 554-580, 1982.

[12] Dawson, D.A., and Kurtz, T.G.: Applications of duality to measure-valued diffusion processes. In: Advances in Filtering and Optimal Stochastic Control. Springer Berlin Heidelberg, p. 91-105, 1982.

[13] Depperschmidt, A., Greven, A., and Pfaffelhuber, P.: Tree-valued FlemingViot dynamics with mutation and selection. The Annals of Applied Probability, v. 22, n. 6, p. 2560-2615, 2012.

[14] Donnelly, P., and Kurtz, T.G.: A countable representation of the Fleming-Viot measure-valued diffusion. The Annals of Probability, v. 24, n. 2, p. 698-742, 1996.

[15] Donnelly, P., and Kurtz, T.G.: Particle representations for Measure-Valued population models. Ann. Probab. 27, n. 1, 166–205, 1999.

[16] Donnelly, P.,and Kurtz, T.G.: Genealogical processes for Fleming-Viot models with selection and recombination. Ann. Appl. Probab. 9, n. 4, 1091–1148, 1999.

[17] Dress, A., Huber, K.T., Koolen, J., Moulton, V., and Spillner, A.: Basic phylogenetic combinatorics. Cambridge University Press, 2012.

[18] Durrett, R.: Ten lectures on particle systems. Lectures on probability theory (Saint-Flour, 1993), 97201, Lecture Notes in Mathematics. 1608, Springer, Berlin, 1995.

[19] Durrett, R.: Probability models for DNA sequence evolution, Springer series on probability and its applications, 2nd edition, 2008.

[20] Durrett, R.: Probability: theory and examples. Fourth edition. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010.

[21] Dynkin, E.B.: Markov processes. Volume I and II. Springer Berlin Heidelberg, 1965.

[22] Etheridge, A.M.: An introduction to superprocesses. Providence, RI: American Mathematical Society, 2000.

[23] Ethier, S.N., and Kurtz, T.G.: Markov Processes: Characterization and Convergence. Wiley, New York, 1986.

[24] Ethier, S.N., and Kurtz, T.G.: Fleming-Viot processes in population genetics. SIAM Journal on Control and Optimization. 31(2):345-86, 1993.

[25] Ewens, W.J.: Mathematical Population Genetics, Springer-Verlag, 1979.

[26] Fertin, G., Labarre, A., Rusu, I., Tannier, É., and Vialette, S.: Combinatorics of genome rearrangements. MIT Press, Cambridge, MA, 2009.

[27] Fleming, W., and Viot, M.: Some Measure-Valued Markov Processes in Population Genetics Theory, Indiana Univ. Math. J. 28, n. 5, p. 817843, 1979.

[28] Greven, A., Pfaffelhuber, P., and Winter, A.: Tree-valued resampling dynamics Martingale problems and applications. Probability Theory and Related Fields, 155(3-4), pp.789-838, 2013.

[29] Haghighi, M., and Sankoff, D.: Medians seek the corners, and other conjectures. BMC Bioinformatics, **13**: S19:S5, 2012.

[30] Hannenhalli, S., and Pevzner, P.A.: Transforming cabbage into turnip: Polynomial algorithm for sorting signed permutations by reversals. Journal of the ACM, **46**: 1–27, 1999.

[31] Holley, R.A., and Liggett, T.M.: Ergodic theorems for weakly interacting infinite systems and the voter model. The annals of probability, p. 643-663, 1975.

[32] Huson, D., Nettles, S., and Warnow, T.: Disk-covering, a fast-converging method for phylogenetic tree reconstruction. Journal of Computational Biology, **6**: 369–386, 1999.

[33] Jamshidpey, A.: Criteria for absolute continuity of laws of Moran processes in random and deterministic environments: a branching representation of the Radon-Nikodym derivative, *in preparation*.

[34] Jamshidpey, A.: Fleming-Viot models in random environments, *in preparation*.

[35] Jamshidpey, A.: Asymptotic tree-indexed random walks, geodesic trees, and geometric median problem on symmetric groups, *in preparation*.

[36] Jamshidpey, A., Jamshidpey, A., and Sankoff, D.: Sets of medians in the non-geodesic pseudometric space of unsigned genomes with breakpoints. BMC Genomics, **1**5, S6:S3, 2014.

[37] Jamshidpey, A., and Sankoff, D.: Phase change for the accuracy of the median value in estimating divergence time. BMC Bioinformatics, **1**4, S15:S7, 2013.

[38] Jamshidpey, A., and Sankoff, D.: Asymptotic medians of random permutations sampled from reversal random walks, *submitted*, 2015.

[39] Jansen, S., and Kurt, N.: On the notion(s) of duality for Markov processes. Probab. Surv. 29;11:59-120, 2014.

[40] Kallenberg, O.: Foundations of modern probability. Second edition, Springer-Verlag, 2002.

[41] Krone, S.M., and Neuhauser, C.: Ancestral processes with selection. Theoretical population biology. 30;51(3):210-37, 1997.

[42] Kurtz, T.: Martingale problems for conditional distributions of Markov processes. Electronic Journal of Probability, v.3, n.9, p.1-29, 1998.

[43] Liggett, T.: Interacting particle systems. Springer-Verlag, 2005.

[44] Mustonen, V., and Lässig, M.: From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. Trends in Genetics, v. 25, n. 3, p. 111-119, 2009.

[45] Mustonen, V., and Lässig, M.: Fitness flux and ubiquity of adaptive evolution. Proceedings of the National Academy of Sciences, v.107, n.9, p.4248-4253, 2010.

[46] Neuhauser, C., and Krone, S.M.: The genealogy of samples in models with selection. Genetics. 1;145(2):519-34, 1997.

[47] Pazy, A.: Semigroups of linear operators and applications to partial differential equations. Applied Mathematical Sciences, 44. Springer-Verlag, New York, 1983.

[48] Sample, C., and Steel, M.: Phylogenetics. Oxford Lecture Series in Mathematics and its Applications 24, Oxford University Press, 2003.

[49] Sankoff, D., Abel, Y., and Hein, J.: A Tree – A Window – A Hill; Generalization of nearest-neighbour interchange in phylogenetic optimisation. Journal of Classification, **11**: 209–232, 1994.

[50] Sankoff, D., and Blanchette, M.: The median problem for breakpoints in comparative genomics. Computing and combinatorics, p. 251-263, 1997.

[51] Sankoff, D., Cedergren, R.J., and Lapalme, G.: Frequency of insertion/deletion, transversion and transition in the evolution of 5S ribosomal RNA. Journal of Molecular Evolution, **7**:133–149, 1976.

[52] Sankoff, D., Sundaram, G., and Kececioglu, J.: Steiner points in the space of genome rearrangements. International Journal of the Foundations of Computer Science, **7**: 1–9, 1996.

[53] Setubal, J.C., and Meidanis, J.: Introduction to Computational Molecular Biology, PWS Publishing Company, 1997.

[54] Shiga, T.: An interacting system in population genetics. Journal of Mathematics of Kyoto University. 20(2):213-42, 1980.

[55] Shiga, T.: Diffusion processes in population genetics. Journal of Mathematics of Kyoto University, 21(1):133-51, 1981.

[56] Stroock, D.W., and Varadhan, S.R.S.: Diffusion processes with continuous coefficients, I. Comm. Pure Appl. Math., 22: 345400, 1969.

[57] Stroock, D.W., and Varadhan, S.R.S.: Diffusion processes with continuous coefficients, II. Comm. Pure Appl. Math., 22: 479530, 1969.

[58] Stroock, D.W., and Varadhan, S.R.S.: Multidimensional Diffusion Processes. Springer-Verlag, 1979.

[59] Székely, L.A., and Yang, Y.: On the expectation and variance of the reversal distance. Acta Univ. Sapientiae, Mathematica, **1**: 5–20, 2009.

[60] Tang, J., and Moret, B.: Scaling up accurate phylogenetic reconstruction from gene-order data. Bioinformatics, **19**: i305–i312, 2003.

[61] Tannier, É., Zheng, C., and Sankoff, D.: Multichromosomal median and halving problems under different genomic distances. BMC Bioinformatics, **10**: 120, 2009.

[62] Tavaré, S.: Ancestral inference in population genetics. Lectures on probability theory and statistics, Lecture Notes in Mathematics, 1837, Springer, Berlin, p. 1188, 2004.

[63] Varadhan, S.R.S.: Stochastic processes. Courant Lecture Notes in Mathematics, 16. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2007.

[64] Xu, A.W., Alain, B., and Sankoff, D.: Poisson adjacency distributions in genome comparison: multichromosomal , circular, signed and unsigned cases. Bioinformatics, **2**4, p. i146-i152, 2008.

[65] Zhang, M., Arndt, W., and Tang, J.: An exact solver for the DCJ median problem. Pacific Symposium on Biocomputing, p. 138–149, 2009.

[66] Zheng, C., and Sankoff, D.: On the PATHGROUPS approach to rapid small phylogeny. BMC Bioinformatics, **12**: S4, 2011.