UNIVERSITÀ COMMERCIALE "LUIGI BOCCONI" – MILANO

Institute of Quantitative Methods

Ph.D. in Statistics – Cycle XIX

Urn-based Particle Processes for Fleming-Viot Models in Bayesian Nonparametrics

Dissertation by:

Matteo Ruggiero

Matr. 948000

UNIVERSITÀ COMMERCIALE "LUIGI BOCCONI"
INSTITUTE OF QUANTITATIVE METHODS

The thesis "**Urn-based Particle Processes for Fleming-Viot Models in Bayesian Nonparametrics**" by **Matteo Ruggiero** is recommended for acceptance by the members of the delegated committee, as stated by the enclosed reports, in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

Dated: January 2007

Research Supervisor:	**Prof. Stephen G. Walker,** **University of Kent, UK**
Internal Examiner:	**Prof. Sonia Petrone,** **Università "Luigi Bocconi", Milano**
External Examiner:	**Dr. Dario Spanò,** **University of Oxford, UK**

UNIVERSITÀ COMMERCIALE "LUIGI BOCCONI"

Date: **January 2007**

Author: **Matteo Ruggiero**

Title: **Urn-based Particle Processes for Fleming-Viot Models in Bayesian Nonparametrics**

Department: **Institute of Quantitative Methods**

Signature of Author

Alla mia ombra

- Would you tell me, please, which way I ought to go from here?

- That depends a good deal on where you want to get to.

- I don't much care where.

- Then it doesn't much matter which way you go.

- . . . so long as I get somewhere.

- Oh, you're sure to do that, if only you walk long enough.

Lewis Carrol, *Alice in wonderland*

It is the face of our own shadow that glowers at us across the Iron Curtain.

Carl Gustav Jung, *Man and His Symbols*

Table of Contents

Acknowledgements

The author has been supported from September 15th 2003 to September 14th 2006 by a postgraduate fellowship awarded by Fondazione "Fratelli Confalonieri", Milano, Italy.

Thanks

The present work, which completes the course of my doctoral studies, would simply not have been possible without Professor Stephen G. Walker of University of Kent, UK. His help, encouragements, comprehension during my year spent in Canterbury have been at the base of my ability of beginning and bringing to an end this thesis. To him are directed my most profound thanks.

I also would like to thank: Professor Pietro Muliere of Università "L. Bocconi", for the chances i have been given; and my parents, of course, for the endless support, which we often take for granted but never should.

My thoughts go then to Petra, Davide, Gabriella and Chiara, faithful adventure-mates all the way through the uncertain path we embarked on more than three years ago. I wish each of you will manage to find his way;

Andrew, who witnessed and shared my changes of mood during a whole year;

Elena and Tommaso, cornerstones of my way back and forth, always ready to provide shelter to a modern pilgrim;

Andrea, Claudio, Igor, Luca, Mattia, Paolo, which in a way or another have proved to be the lighthouse in my long journeys;

and finally, Lori, which appeared in my life only when the road had started to be downhill, but which is also when one can lose control and get lost. Her presence has been soothing from the very first moment, and her empathic understanding has been complete ever since. I hope ours could be an everlasting dance.

Abstract

The Fleming-Viot process is a probability-measure-valued diffusion which arises as the large population limit of a wide class of models in population genetics. In the selectively neutral case its stationary distribution is known to be the Dirichlet process, but its connections with Bayesian nonparametrics at a deeper level are still to be explored.

In this work, by means of known and newly defined generalised Pòlya-urn schemes, several types of pure jump particle processes are introduced, describing the evolution in time of an exchangeable population. By considering the empirical measure of the individuals at each time point, the jump processes provide an explicit construction of four different formulations of the Fleming-Viot process, corresponding to the cases of neutrality, viability selection, haploid and diploid fertility selection. Weak convergence in the Skorohod space of every constructed measure-valued process to the respective Fleming-Viot diffusion holds. The stationary distribution of each case is derived and shown to be the de Finetti measure of the infinite sequence of individuals; in presence of viability selection the stationary distribution turns out to be the two-parameter Poisson-Dirichlet process.

Introduction

For a stochastic model in population genetics, the state of a population at a single time point can be represented as a probability distribution. In the simplest cases this can be thought of as a list of the genetic types present, with their respective frequencies. It is then natural to model the evolution in time of such distribution of frequencies as a measure-valued process. A probability-measure-valued process is, in general, a stochastic process which takes values in the space of Borel probability measures $\mathscr{P}(E)$, that is the set of all probability measures defined on $(E, \mathscr{B}(E))$, where E is a metric space and $\mathscr{B}(E)$ is the Borel sigma algebra on E.

The Fleming-Viot process is a measure-valued diffusion which arises as the large population limit of a wide class of population genetics models. Population genetics diffusions are often motivated by first considering a class of prelimiting finite-population models, where the dynamics are specified in terms of the behaviour of the individuals. Each individual has a type, which in the infinitely-many-alleles case is represented as a point in a complete separable metric space E. The composition of the population is then naturally interpreted as a probability measure on E.

In the last few decades the Fleming-Viot process has been object of a considerable amount of research, and together with the Dawson-Watanabe process, which arises from branching models, is one of the more well studied measure-valued processes. In particular, the Fleming-Viot diffusion has been considered in different formulations, which depend on the amount and type of evolutionary mechanisms that the process describes.

The purpose of this work is to exploit already known and newly defined urn models in order to construct several countable particle processes, representing the evolution in time of a genetic population, in such a way that a close connection between Fleming-Viot models and population genetics on the one hand and Bayesian nonparametrics and urn schemes on the other is established. From a population genetics standpoint, the constructions cover the cases of neutrality (in two different ways), viability selection, haploid and diploid fertility selection. Given the particle process, the associated process whose value at every time point is that of the empirical measure of the particles, is shown to converge weakly, in each defined model, to a certain formulation of the Fleming-Viot measure-valued diffusion, and the stationary distribution is derived.

The construction via urn models presents two important aspects: it provides insight into the dynamics of the individuals, enabling to explain the behaviour of the process underlying the measure-valued diffusion; and secondly, it allows to exploit the features of a Bayesian nonparametric approach, yielding a simple derivation of the stationary distribution of the measure-valued processes so constructed. The newly introduced urn-based approach, which constitutes a different perspective toward Fleming-Viot models with respect to the existent literature, thus proves to be a useful and powerful tool in the study of population genetics, simplifying both the stage of construction and the *ex post* analysis.

The dissertation is divided into two parts. The first, which counts four chapters, introduces the theoretical instruments which will constitute the operating background of the remainder of the dissertation. Except some cases in which some auxiliary results were needed, we tried to include in this preliminary part only the results that will actually be used or referenced to in the second part. Needless to say, the first four chapters are by no means intended to be self contained or exhaustive in their respective topics, but rather a collection of the notions which are strictly necessary in dealing with the second part, to be consulted whenever needed. All topics in the first part are therefore taken from the existent literature and appropriately referenced.

In the first chapter, after briefly fixing the notation on Markov processes, we mention the basic notions on semi-Markov process and the connection between the Gibbs sampler algorithm and Markov chains. We then introduce infinitesimal generators and semigroup operators, which will be broadly used throughout the work, and define the martingale problem of Stroock and Varadhan.

Chapter two deals with exchangeable sequences generated by urn-schemes and some Bayesian nonparametric priors related to them, namely the Dirichlet process and the two-parameter Poisson-Dirichlet process. The mixture of Dirichlet process model, which will be an auxiliary tool in a couple of proofs, is also sketched.

Chapter three is about convergence of probability measures, both in the topology of weak convergence induced by the Prohorov metric, and in the Skorohod topology of the càdlàg space $D_E[0,\infty)$.

The last chapter of the introductory part, after introducing some terminology typical of a population genetics context, defines Fleming-Viot models, emphasising their characterisation in terms of generators, and presents a few discrete representations, developed in the last decade, which converge, in the infinite population limit, to Fleming-Viot models.

The second part is concerned, as already mentioned, with the construction of particle processes which lead to Fleming-Viot models and which are based on urn schemes. The topics covered in this part, unless otherwise stated, can be considered new, from the definitions and constructions of the particle processes to the results on convergence and stationarity. If, on the contrary, a quoted result is taken from the literature, this is clearly stated and the source (or where to find it in the introductory part) referenced. The above criterion does not apply only to the technique for deriving the associated process of empirical measures, which is not new and can be found in Chapter four, with the exception of Chapter six, where a different approach is introduced.

Chapter five, which is the first of the second part, is somewhat a model for the successive, which, although with some differences, more or less follow the same pattern. First a pure jump Markov particle process on the space E^n is defined, as a vector of n exchangeable variables which are generated by the Blackwell-MacQueen urn and evolve in time according to some kernel, also based on the Blackwell-MacQueen prediction rule. Then the associated process, with sample paths in the càdlàg space $D_{\mathscr{P}(E)}[0,\infty)$, whose value at each t is given by the empirical measure of the particle process at t, is considered, and the infinitesimal generators of both processes computed. By means of the generator, the infinite population limit shows that the measure-valued process for large n becomes a neutral diffusion model, that is a Fleming-Viot process in the case of parent-independent non-atomic mutation and with no other evolutionary mechanism different apart from genetic drift. Weak convergence in the Skorohod topology is proved and the stationary distribution, which in this case is known to be the Dirichlet process, is derived. The derivation, which in particular exploits the properties of the Gibbs sampler algorithm, is simplified by the use of urn models in a Bayesian context, thus justifying a new proof for a known result, whose scheme will be used for the following unknown cases.

Chapter six develops a different measure-valued process associated with the particle process defined in Chapter five. This new construction starts from the observation that a probability measure sampled from a posterior Dirichlet can be represented in terms of mixture of a weighted sum of the observations to which the Dirichlet process is conditioned and a sample from a Dirichlet prior. It is shown that a measure-valued process based on this representation can be constructed, such that it is asymptotically Markov and converges weakly to the neutral diffusion model. Its stationary distribution is shown to be the Dirichlet process for every size of the population, and not only at the infinite population limit.

In Chapter seven we consider the case of the Fleming-Viot process with viability selection, that is selection that affects the ability to survive longer than others, and attempt the construction of a Markov particle process such that the de Finetti measure of the particles at every time point is the two-parameter Poisson-Dirichlet process. The attempt fails due to unboundedness of the generators, and some conjectures on the reasons in terms of properties of the model are provided in the following

Chapter eight, where the attempt is repeated relaxing the assumption of Markovianity, and this time succeeds. Results similar to Chapter five on weak convergence and stationarity of the associated measure-valued process are obtained. In particular the two-parameter Poisson-Dirichlet process is shown to be the stationary distribution of the Fleming-Viot process with viability selection.

Chapter nine and ten deal with the Fleming-Viot process with fertility selection, haploid and diploid respectively, that is selection that affects the ability to generate offspring in the cases of single alleles or pairs. Before following a construction analogous to the previous chapters, this time we need to generalise the prediction rule associated with the Blackwell-MacQueen urn, in such a way that allows the inclusion of the selection mechanism. Convergence to the respective Fleming-Viot process is shown, and in particular it holds when the differences between the fitnesses of distinct individuals are of the order of the inverse of the population size. The stationary distribution, which is known in both cases, is derived by appealing to the Dirichlet process mixture model in hierarchical framework.

Part I

PRELIMINARIES AND THEORETICAL BACKGROUND

Chapter 1

Markov processes and generators

In this chapter we introduce the basic definitions concerning semigroup operators and generators, which are of primary importance in the study of Markov processes and will constitute the main operating tools in the second part of this work. This chapter is not intended to be exhaustive on the topic, but only a compact reference to be used later on.

We first briefly recall a few notions about Markov processes, followed by the definition of the Gibbs sampler algorithm and its connection with Markov chains. Generators and semigroups are then introduced, and finally we mention the martingale problem of Stroock and Varadhan.

For a comprehensive account on the topic see for example Ethier and Kurtz (1986), from which, except for the section on the Gibbs sampler, are taken the results reported in this chapter.

1.1 Markov processes

Let $\{X(t), t \geq 0\}$ be a stochastic process defined on a probability space $(\Omega, \mathscr{F}, P)$ with values in a metric space E, and let $\mathscr{F}_t^X = \sigma(X(s) : s \leq t)$. Then $\{X(t), t \geq 0\}$ is a *Markov process* if

$$P\{X(t+s) \in \Gamma |\, \mathscr{F}_t^X\} = P\{X(t+s) \in \Gamma |\, X(t)\} \tag{1.1.1}$$

for all $s, t \geq 0$ and $\Gamma \in \mathscr{B}(E)$, where $\mathscr{B}$ denotes Borel sigma algebra.

A function $P(t, x, \Gamma)$ defined on $[0, \infty) \times E \times \mathscr{B}(E)$ is a time homogeneous *transition function* if

$$P(t, x, \cdot) \in \mathscr{P}(E) \tag{1.1.2}$$

$$P(0, x, \cdot) = \delta_x \tag{1.1.3}$$

$$P(\cdot, \cdot, \Gamma) \in B([0, \infty) \times E) \tag{1.1.4}$$

where $\mathscr{P}(E)$ is the set of probability measures on E, δ_x is the point mass at x and $B(\cdot)$ denotes the space of bounded functions, and also

$$P(t+s, x, \Gamma) = \int P(s, y, \Gamma) P(t, x, \mathrm{d}y) \tag{1.1.5}$$

for $s, t \geq 0$, $x \in E$ and $\Gamma \in \mathscr{B}(E)$. Expression (1.1.5) is called *Chapman-Kolmogorov equation*.

A transition function $P(t, x, \Gamma)$ is a *transition function for a time homogeneous Markov process* $\{X(t), t \geq 0\}$ if

$$P\{X(t+s) \in \Gamma |\, \mathscr{F}_t^X\} = P(s, X(t), \Gamma) \tag{1.1.6}$$

for all $s, t \geq 0$ and $\Gamma \in \mathscr{B}(E)$, or, equivalently, if

$$E[f(X(t+s))|\,\mathscr{F}_t^X] = \int f(y)P(s, X(t), \mathrm{d}y) \tag{1.1.7}$$

for all $s, t \geq 0$ and $f \in B(E)$.

The probability measure $\nu_0 \in \mathscr{P}(E)$ given by $\nu_0(\Gamma) = P(X(0) \in \Gamma)$ is called *initial distribution* of $\{X(t), t \geq 0\}$.

A transition function and the initial distribution determine the *finite-dimensional distributions* of $\{X(t), t \geq 0\}$, by

$$P\{X(0) \in \Gamma_0, X(t_1) \in \Gamma_1, \ldots, X(t_n) \in \Gamma_n\} = \tag{1.1.8}$$

$$= \int_{\Gamma_0 \times \ldots \times \Gamma_{n-1}} P(t_n - t_{n-1}, y_{n-1}, \Gamma_n) P(t_{n-1} - t_{n-2}, y_{n-2}, \mathrm{d}y_{n-1}) \ldots P(t_1, y_0, \mathrm{d}y_1)\nu_0(\mathrm{d}y_0).$$

In particular we have the following theorem.

Theorem 1.1.1 (Existence of Markov process)

Let $P(t, x, \Gamma)$ satisfy (1.1.2)-(1.1.5) *and let $\nu_0 \in \mathscr{P}(E)$. If for each $t \geq 0$ the probability measure $\int P(t, x, \cdot)\nu_0(dx)$ is tight (which is the case if the metric space (E, r) is complete and separable), then there exists a Markov process $\{X(t), t \geq 0\}$ taking values in E whose finite-dimensional distributions are uniquely determined by* (1.1.8).

Let P_x denote the measure on $\mathscr{B}(E)^{[0,\infty)}$ given by Theorem 1.1.1 with $\nu_0 = \delta_x$, and let $\{X(t), t \geq 0\}$ be the corresponding coordinate process, that is $X(t, \omega) = \omega(t)$. It follows from (1.1.4) and (1.1.8) that $P_x(B)$ is a Borel measurable function of x for $B = \{X(0) \in \Gamma_0, \ldots, X(t_n) \in \Gamma_n\}$, $0 < t_1 < \cdots < t_n$, $\Gamma_0, \Gamma_1, \ldots, \Gamma_n \in \mathscr{B}(E)$. In fact this is true for all $B \in \mathscr{B}(E)^{[0,\infty)}$.

Proposition 1.1.2 *Let P_x be as above. Then $P_x(B)$ is a Borel measurable function of x for each $B \in \mathscr{B}(E)^{[0,\infty)}$.*

Let $\{Y_n, n \in \mathbb{N}\}$ be a discrete parameter process defined on $(\Omega, \mathscr{F}, P)$ with values in E, and let $\mathscr{F}_n^Y = \sigma(Y(k) : k \leq n)$. Then $\{Y_n, n \in \mathbb{N}\}$ is a *Markov chain* if

$$P\{Y(n+m) \in \Gamma |\, \mathscr{F}_n^Y\} = P\{Y(n+m) \in \Gamma |\, Y(n)\} \tag{1.1.9}$$

for all $m, n \geq 0$ and $\Gamma \in \mathscr{B}(E)$. A function $\mu(x, \Gamma)$ defined in $E \times \mathscr{B}(E)$ is a *transition function* if

$$\mu(x, \cdot) \in \mathscr{P}(E) \tag{1.1.10}$$

and

$$\mu(\cdot, \Gamma) \in B(E) \tag{1.1.11}$$

for $x \in E$ and $\Gamma \in \mathscr{B}(E)$. A transition function $\mu(x, \Gamma)$ is a *transition function for a time homogeneous Markov chain* $\{Y_n, n \in \mathbb{N}\}$ if

$$P\{Y(n+1) \in \Gamma |\, \mathscr{F}_n^Y\} = \mu(Y(n), \Gamma) \tag{1.1.12}$$

for $n \geq 0$ and $\Gamma \in \mathscr{B}(E)$. Note that

$$P\{Y(n+m) \in \Gamma |\, \mathscr{F}_n^Y\} = \int \ldots \int \mu(y_{m-1}, \Gamma)\mu(y_{m-2}, \mathrm{d}y_{m-1}) \ldots \mu(Y(n), \mathrm{d}y_1).$$

As before, the probability measure $\nu_0 \in \mathscr{P}(E)$ given by $\nu_0(\Gamma) = P\{Y(0) \in \Gamma\}$ is called the initial distribution for Y. The analogues of Theorem 1.1.1 and Proposition 1.1.2 can be easily derived.

Let $\{X(t), t \geq 0\}$, defined on $(\Omega, \mathscr{F}, P)$, be an E-valued Markov process with respect to a filtration $\{\mathscr{G}_t\}$. Suppose $P(t, x, \Gamma)$ is a transition function for $\{X(t), t \geq$

$0\}$, and let τ be a $\{\mathscr{G}_t\}$-stopping time (i.e. a random variable taking values in $[0, \infty]$ such that $\{\tau \leq t\} \in \mathscr{G}_t$ for all $t \geq 0$), with $\tau < \infty$ a.s. Then X is *strong Markov at* τ if

$$P\{X(\tau + t) \in \Gamma | \mathscr{G}_t\} = P(t, X(\tau), \Gamma) \tag{1.1.13}$$

for all $t \geq 0$ and $\Gamma \in \mathscr{B}(E)$, or equivalently, if

$$E[f(X(\tau + t)) | \mathscr{G}_t] = \int f(y) P(t, X(\tau), \mathrm{d}y) \tag{1.1.14}$$

for all $t \geq 0$ and $f \in B(E)$. $\{X(t), t \geq 0\}$ is a *strong Markov process* with respect to $\{\mathscr{G}_t\}$ if it is strong Markov at τ for all $\{\mathscr{G}_t\}$-stopping times τ with $\tau < \infty$ a.s. .

A process $\{X(t), t \geq 0\}$ is *adapted* to a filtration $\{\mathscr{F}_t\}$ (or $\{\mathscr{F}_t\}$-adapted) if $X(t)$ is $\{\mathscr{F}_t\}$-measurable for each $t \geq 0$. A process is $\{\mathscr{F}_t\}-$*progressive* if for each $t \geq 0$ the restriction of $\{X(t), t \geq 0\}$ to $[0, t] \times \Omega$ is $\mathscr{B}[0, t] \times \{\mathscr{F}_t\}$-measurable. Note that if X is $\{\mathscr{F}_t\}$-progressive, then it is $\{\mathscr{F}_t\}$-adapted and measurable, but the converse is not necessarily true.

1.2 Semi-Markov processes

In this section we define the semi-Markov processes and the strictly connected Markov renewal processes, used in Chapter 8. Semi-Markov processes were introduced by Lévy (1954) and Smith (1955) independently and simultaneously, while Markov renewal processes were studied in detail by Pyke (1961a,b). In this brief review we follow Cinlar (1975).

Suppose E is a finite set, and assume to have on a probability space $(\Omega, \mathscr{F}, P)$

random variables

$$X_n : \Omega \to E$$

$$T_n : \Omega \to \mathbb{R}_+$$

defined for each $n \in \mathbb{N}$ so that $0 = T_0 \leq T_1 \leq T_2 \leq \cdots$. These elements are said to form a *Markov renewal process* with state space E provided that

$$\begin{aligned} &P\{X_{n+1} = j, T_{n+1} - T_n \leq t |\, X_0, \ldots, X_n; T_0, \ldots, T_n\} \\ &\qquad = P\{X_{n+1} = j,\, T_{n+1} - T_n \leq t |\, X_n\} \end{aligned} \tag{1.2.1}$$

for all $n \in \mathbb{N}$, $j \in E$, $t \geq 0$. If the process is time homogeneous, then

$$P\{X_{n+1} = j,\, T_{n+1} - T_n \leq t |\, X_n = i\} = Q(i, j, t) \tag{1.2.2}$$

independent of n. The family of probabilities

$$Q = \{Q(i, j, t) :\ i, j \in E,\, t \geq 0\}$$

is called a *semi-Markov kernel*. Assume $Q(i, j, 0) = 0$ for all $i, j \in E$.

Defining now

$$P(i, j) = \lim_{t \to \infty} Q(i, j, t) \tag{1.2.3}$$

we see that

$$P(i, j) \geq 0 \qquad i, j \in E \tag{1.2.4}$$

$$\sum_{j \in E} P(i, j) = 1 \qquad i \in E \tag{1.2.5}$$

that is the $P(i, j)$ are the transition probabilities for some Markov chain with state space E. From (1.2.1) and (1.2.3) it follows that

$$P\{X_{n+1} = j |\, X_0, \ldots, X_n; T_0, \ldots, T_n\} = P(X_n, j) \tag{1.2.6}$$

for all $n \in \mathbb{N}$ and $j \in E$. This implies in particular the following proposition.

Proposition 1.2.1 $\{X_n,\ n \in \mathbb{N}\}$ *is a Markov chain with state space* E *and transition probability matrix* P.

If $P(i,j) = 0$ for some pair (i,j), then $Q(i,j,t) = 0$ for all $t \geq 0$ and we define the ratio $Q(i,j,t)/P(i,j)$ to be unity. With this convention we define

$$G(i,j,t) = \frac{Q(i,j,t)}{P(i,j)} \qquad i,j \in E, \quad t \geq 0.$$

Then, for each pair (i,j), the function $G(i,j,\cdot)$ is a distribution function. From (1.2.2) and (1.2.6) we see that

$$G(i,j,t) = P\{\, T_{n+1} - T_n \leq t |\, X_n = i, X_{n+1} = j\}$$

Using this interpretation together with (1.2.1) we can show by induction the following.

Proposition 1.2.2 *For any integer* $n \in \mathbb{N}$ *and reals* $u_i, \dots, u_n \in \mathbb{R}_+$

$$P\{T_1 - T_0 \leq u_1,\ T_2 - T_1 \leq u_2, \dots,\ T_n - T_{n-1} \leq U_n |\ X_0, X_1, \dots\}$$
$$= G(X_0, X_1; u_1)\ G(X_1, X_2; u_2) \dots G(X_{n-1}, X_n; u_n)$$

that is, the increments $T_1 - T_0,\ T_2 - T_1, \dots$ *are conditionally independent given the Markov chain* $X_0, X_1, \dots$.

In particular, if the state space E consists of a single point, then the increments are independent and identically distributed nonnegative random variables.

Corollary 1.2.3 *If E consists of a single point then $\{T_n,\, n \in \mathbb{N}\}$ is a renewal process.*

Proposition 1.2.1 and Corollary 1.2.3 justify the phrase *Markov renewal process*, indicating a generalisation of a Markov chain and of a renewal process. The full justification is however contained in Proposition 1.2.1 and the following.

Proposition 1.2.4 *Let $i \in E$ be fixed and define $S_0^i, S_1^i, \ldots$ to be the successive T_n for which $X_n = i$. Then $\{S_n^i,\, n \in \mathbb{N}\}$ is a (possibly delayed) renewal process.*

Thus to each state i there corresponds a renewal process $\{S_n^i,\, n \in \mathbb{N}\}$. The superposition of all these renewal processes gives the points T_n, $n \in \mathbb{N}$. The renewal process which contributed the point T_n is the ith one if and only if $X_n = i$. The types of the successive points, namely $X_0, X_1, \ldots,$ form a Markov chain.

Proposition 1.2.5 *If E is finite, then $\sup_n T_n = +\infty$ almost surely.*

By weeding out those $\omega \in \Omega$ for which $\sup_n T_n(\omega) < \infty$, we may assume that $\sup_n T_n = +\infty$ for all ω. Then, for any $\omega \in \Omega$ and $t \geq 0$, there is some integer n such that $T_n(\omega) \leq t < T_{n+1}(\omega)$.

Definition 1.2.6 (Semi-Markov process)

Let X_n and T_n be defined as above. A continuous time parameter process $\{Y(t), t \geq 0\}$ with state space E such that

$$Y_t = X_n \quad \text{on} \quad \{T_n \leq t < T_{n+1}\}$$

is called a *semi-Markov process* with state space E and semi-Markov transition kernel $Q = \{Q(i,j,t)\}$.

We may think of Y_t as the state at time t of some system or particle which moves from one state to another with random sojourn times in between. The length of a sojourn interval $[T_n, T_{n+1})$ is a random variable whose distribution depends both on the state X_n being visited and the state X_{n+1} to be entered next. The successive states visited form a Markov chain and, conditional on that sequence, the successive sojourn times are independent.

We may obtain a realization of the semi-Markov process Y in the following manner. Suppose we have the transition probabilities $P(i,j)$ satisfying (1.2.4)-(1.2.5) prescribed along with a distribution function $G(i,j,t)$ for each pair (i,j). Suppose the initial state is to be i. First the next state to be entered is sampled from the distribution $P(i,\cdot)$. If the outcome is j, then a sojourn time u is sampled from the distribution $G(i,j,\cdot)$. The function y_t is set to be i for all $t < u$ and y_u is set to be j. The second step starts by sampling the next state to be entered from the distribution $P(i,\cdot)$. If the outcome is k, then a sojourn time v is sampled from the distribution $G(i,k,\cdot)$. The function y_t is now set to be j for all $t \in [u, u-v)$ and y_{u+v} is set to be

k, and so on. The resulting function Y_t is a realization of the semi-Markov process Y with semi-Markovian kernel Q given by $Q(i,j,t) = P(i,j)\,G(i,j,t)$.

Finally, a word of justification for the name semi-Markov process. If the semi-Markov kernel Q is of the form

$$Q(i,j,t) = P(i,j)\left(1 - e^{-\lambda(i)t}\right) \tag{1.2.7}$$

then one can show that

$$P\{Y_{t+s} = j|\, Y_u,\, u \leq t\} = P\{Y_{t+s} = j|\, Y_t\} \tag{1.2.8}$$

for all $t, s \geq 0$ and $j \in E$. Furthermore

$$P\{Y_{t+s} = j|\, Y_t = i\} = P_s(i,j)$$

is independent of t. In other words, under (1.2.7) the semi-Markov process becomes a (temporally homogeneous) Markov process. Thus, with respect to Markov processes, the novel feature of the semi-Markov process is the freedom allowed in the choice of the distribution of the sojourn times. This freedom, however, is achieved at the expense of the Markov property (1.2.8) which, instead of holding for all t, holds now only for the jump times T_n.

1.3 The Gibbs sampler

Before dealing with generators, it is useful to briefly introduce the Gibbs sampler algorithm and stress its relationship with Markov chains.

Initially introduced by Geman and Geman (1984), who studied image-processing models, the Gibbs sampler is a computationally-intensive technique which has officially entered the statistical literature since Gelfand and Smith (1990) revealed its potential in a wide variety of problems of Bayesian inference. For a review see Smith and Roberts (1993) and Besag and Green (1993). See also Tanner and Wong (1987).

The Gibbs sampler is an algorithm that generates a sequence of samples from the joint probability distribution of two or more random variables, without having to calculate it. A special case of the Metropolis-Hastings algorithm (see Metropolis *et al.*, 1953; Hastings, 1970), and thus an example of a Markov Chain Monte Carlo algorithm, often referred to as MCMC, the Gibbs sampler is named after the american J.W. Gibbs (1839-1903), one of the very first theoretical physicists.

The Gibbs sampler is applicable when the joint distribution is not known explicitly, but the conditional distribution of each variable is known. The procedure generates a sample from the distribution of each variable in turn, conditional on the current values of the other variables; the sequence of samples determines a Markov chain, whose stationary distribution is the sought-after joint distribution.

Gibbs sampling is particularly well-adapted to sampling the posterior distribution of a Bayesian model, since Bayesian models are typically specified as a collection of conditional distributions, and hence can be used to approximate joint distributions (as with a histogram), or to compute integrals (such as an expected value).

Let $p(x_1, \ldots, x_n)$ be the joint distribution of the random vector $X_1, \ldots, X_n$, and assume it is difficult to directly generate the values of such a vector, but it is relatively easy to generate, for each i, a random variable having the conditional distribution of X_i given all the others, which is called full conditional distribution.

Algorithm 1.3.1 (Gibbs sampler)

1. Choose an arbitrary starting point $\mathbf{X}^0 = (x_1^0, \ldots, x_n^0)$, and set $i = 0$.

2. Generate $\mathbf{X}^{i+1} = (x_1^{i+1}, \ldots, x_n^{i+1})$ as follows

   - Sample $X_1^{i+1} \sim p\,(x_1|\, x_2^i, \ldots, x_n^i)$ and call it x_1^{i+1}
   - Sample $X_2^{i+1} \sim p\,(x_2|\, x_1^{i+1}, x_3^i, \ldots, x_n^i)$ and call it x_2^{i+1}

   $\vdots$

   - Sample $X_n^{i+1} \sim p\,(x_n|\, x_1^{i+1}, \ldots, x_{n-1}^{i+1})$ and call it x_n^{i+1}

3. set $i = i + 1$ and go to step 2.

Gelfand and Smith (1990) showed that the sequence $\{\mathbf{X}^i, i \geq 1\}$ is a Markov chain, with stationary distribution is given by $p\,(x_1, \ldots, x_n)$. To see it intuitively, suppose that $\mathbf{X}^0$ has distribution $p\,(x_1, \ldots, x_n)$. Then it is easy to see that at any point in the algorithm, $x_1^i, \ldots, x_{j-1}^i, x_j^{i-1}, \ldots, x_n^{i-1}$ will be the value of a random vector with distribution $p\,(x_1, \ldots, x_n)$. For instance, if X_j^i takes value x_j^i, then

$$
\begin{aligned}
P\{X_1^1 = x_1,& X_2^0 = x_2, \ldots, X_n^0 = x_n\} \\
&= P\{X_1^1 = x_1|\, X_2^0 = x_2, \ldots, X_n^0 = x_n\}P\{X_2^0 = x_2, \ldots, X_n^0 = x_n\} \\
&= P\{X_1 = x_1|\, X_2 = x_2, \ldots, X_n = x_n\}P\{X_2 = x_2, \ldots, X_n = x_n\} \\
&= p\,(x_1, \ldots, x_n). \qquad (1.3.1)
\end{aligned}
$$

Thus $p\,(x_1, \ldots, x_n)$ is a stationary distribution, and provided the Markov chain is irreducible a aperiodic, it is the limiting probability vector for the Gibbs sampler. In particular, it also follows that $p\,(x_1, \ldots, x_n)$ would be limiting distribution even if the

Gibbs sampler were not systematic in following the natural order of the variables, as described in Algorithm 1.3.1. Indeed, even if the component whose value was to be changed was always randomly determined, $p\,(x_1, \ldots, x_n)$ would be the stationary distribution of the chain, provided every component is visited infinitely often (which corresponds to the requirement of irreducibility and aperiodicity). The two different systems of choice are called respectively *deterministic scan* and *random scan*.

The Gibbs sampler will be often exploited in the second part, although in its properties rather than in a computational sense.

1.4 Semigroups and generators

In this section we present the essential definitions concerning semigroups and generators of random processes.

Consider a one-parameter family $\{T(t), t \geq 0\}$ of bounded linear operators on a Banach space L, i.e. on a normed linear space which is complete, that is where every Cauchy sequence converges with respect to the metric induced by the norm. Assume the norm to be $\|\cdot\|$.

Definition 1.4.1 (Semigroup)

The family $\{T(t), t \geq 0\}$ is called a *semigroup* if $T(0) = I$ and $T(s+t) = T(s)T(t)$ for all $s, t \geq 0$.

Definition 1.4.2 (Strongly continuous contraction semigroup)

A semigroup $\{T(t), t \geq 0\}$ on L is said to be:

- *strongly continuous* if $\lim_{t \to 0} T(t)f = f$ for every $f \in L$;
- a *contraction* semigroup if $\| T(t)\| \leq 1$ for all $t \geq 0$, where $\| \cdot \|$ is the norm associated with the space L.

Proposition 1.4.3 *Let $\{T(t), t \geq 0\}$ be a strongly continuous semigroup on L. Then, for each $f \in L$, $t \mapsto T(t)f$ is a continuous function from $[0, \infty)$ into L.*

Assume that $L = C_0$ is the Banach space of continuous functions vanishing at infinite on a locally compact (i.e. such that every point has a neighborhood which is itself contained in a compact set) separable metric space S, with the sup norm.

Definition 1.4.4 (Positive semigroup)
A semigroup $\{T(t), t \geq 0\}$ on C_0 is said to be *positive* if $T(t)$ is a positive operator for each $t \geq 0$, i.e. if it maps non negative functions into non negative functions.

Definition 1.4.5 (Conservative semigroup)
A semigroup $\{T(t), t \geq 0\}$ on C_0 is said to be *conservative* if $T(t)1 = 1$ for all $t \geq 0$.

Definition 1.4.6 (Feller semigroup)
A strongly continuous, positive, contraction semigroup, whose generator is conservative, is called a *Feller semigroup*.

A (possibly unbounded) *linear operator* A on L is a linear mapping whose domain $\mathscr{D}(A)$ is a subspace of L and whose range $\mathscr{R}(A)$ lies in L. The *graph* of A is given by

$$\mathscr{G}(A) = \{(f, Af) : f \in \mathscr{D}(A)\} \subset L \times L. \tag{1.4.1}$$

Note that $L \times L$ is itself a Banach space with componentwise addiction and scalar multiplication and norm $\|(f, g)\| = \|f\| + \|g\|$. A is said to be *closed* if $\mathscr{G}(A)$ is a closed subspace of $L \times L$.

Definition 1.4.7 (Infinitesimal generator)
The *infinitesimal generator* of a semigroup $\{T(t), t \geq 0\}$ on L is the linear operator A defined by

$$Af = \lim_{t \to 0} \frac{T(t)f - f}{t}. \tag{1.4.2}$$

The domain $\mathscr{D}(A)$ of A is the subspace of all $f \in L$ for which this limit exists.

Remark 1.4.8 The operator (1.4.2), in which the limit is intended to be bounded pointwise, is also known as the *weak* generator, as defined in Dynkin (1965). Cf. also (1.4.10) below.

We now state a few properties of generators.

Proposition 1.4.9 *Let $\{T(t), t \geq 0\}$ be a strongly continuous semigroup on L with generator A.*

(*a*) *If* $f \in L$ *and* $t \geq 0$, *then* $\int_0^t T(s) f \mathrm{d}s \in \mathscr{D}(A)$ *and*

$$T(t)f - f = A \int_0^t T(s) f \mathrm{d}s \tag{1.4.3}$$

(*b*) *If* $f \in \mathscr{D}(A)$ *and* $t \geq 0$, *then* $T(t)f \in \mathscr{D}(A)$ *and*

$$\frac{\mathrm{d}}{\mathrm{d}t} T(t) f = A T(t) f = T(t) A f \tag{1.4.4}$$

(*c*) *If* $f \in \mathscr{D}(A)$ *and* $t \geq 0$, *then*

$$T(t)f - f = \int_0^t A T(s) f \mathrm{d}s = \int_0^t T(s) A f \mathrm{d}s. \tag{1.4.5}$$

Corollary 1.4.10 *If* A *is the generator of a strongly continuous semigroup* $\{T(t), t \geq 0\}$ *on* L, *then* $\mathscr{D}(A)$ *is dense in* L *and* A *is closed.*

Proposition 1.4.11 *A Feller semigroup is uniquely determined by its generator.*

Proposition 1.4.12 *Let* $\{T(t), t \geq 0\}$ *and* $\{S(t), t \geq 0\}$ *be two strongly continuous contraction semigroups on* L *with generators* A *and* B *respectively. If* $A = B$, *then* $T(t) = S(t)$ *for all* $t \geq 0$.

The following theorem, due to Hille and Yosida, characterizes the class of linear operators A on a Banach space L which are the generator of a Feller semigroup. A linear operator A on L is said to be *dissipative* if $\| \lambda f - Af \| \geq \lambda \| f \|$ for every $f \in \mathscr{D}(A)$ and $\lambda > 0$.

Theorem 1.4.13 [Hille-Yosida]

A linear operator A on L is the generator of strongly continuous contraction semigroup on L if and only if:

(a) $\mathscr{D}(\mathscr{A})$ *is dense in L;*

(b) A *is dissipative;*

(c) $\mathscr{R}(\lambda - A) = L$ *for some* $\lambda > 0$.

Let A be a closed linear operator on L. If, for some real λ, $\lambda - A(\equiv \lambda I - A)$ is one-to-one, $\mathscr{R}(\lambda - A) = L$ and $(\lambda - A)^{-1}$ is a bounded linear operator on L, then λ is said to belong to the *resolvent set* $\rho(A)$ of A, and $R_\lambda = (\lambda - A)^{-1}$ is called the *resolvent* (at λ) of A.

We now possess the basic notions which enable us to understand how generators can be used to study Markov processes. The main reason why generators are considered instruments of such importance in the analysis of stochastic processes relies on the fact that transition functions, although somehow appealing in view of the possible insight they can yield in the problem, ordinarily cannot be obtained explicitly. Consequently, directly defining a transition function is usually not a useful method of specifying a Markov process.

Let $P(t, x, \Gamma)$ be a transition function. Then

$$T(t)f(x) \equiv \int f(y)P(t, x, \mathrm{d}y) \tag{1.4.6}$$

defines a measurable contraction semigroup on $B(E)$, the space of bounded functions on E, by the Chapman-Kolmogorov property (1.1.5). Namely, the fact that the transition operator $\int f(y)P(t,x,\mathrm{d}y)$ has the semigroup property is a necessary and sufficient condition for the transition kernel $P(t,x,\Gamma)$ to satisfy the Chapman-Kolmogorov equation (Kallenberg, 1997, Lemma 17.1). This gives a first explanation of the importance of semigroup operators for the study of Markov processes.

Let $\{T(t), t \geq 0\}$ be a semigroup on a closed subspace $L \subset B(E)$. With reference to (1.1.7) we say that an E-valued Markov process X *corresponds to* $\{T(t), t \geq 0\}$ if

$$E[f(X(t+s))|\,\mathscr{F}_t^X] = T(s)f(X(t)) \tag{1.4.7}$$

for all $s, t \geq 0$ and $f \in L$. Trivially, if $\{T(t), t \geq 0\}$ is given by a transition function as in (1.4.6), then (1.4.7) is just (1.1.7).

The next proposition explains the connection between the semigroup associated to a Markov process and the finite-dimensional distributions. Given the importance of the result, we report a brief sketch of the proof.

Proposition 1.4.14 *Let E be separable. Let $\{X(t), t \geq 0\}$ be an E-valued Markov process, with initial distribution ν_0, corresponding to a semigroup $\{T(t), t \geq 0\}$ on a closed subspace $L \subset B(E)$. If L is separating (i.e. if whenever $P, Q \in \mathscr{P}(E)$ and $\int f\mathrm{d}P = \int f\mathrm{d}Q$ for $f \in L$, we have $P = Q$), then $\{T(t), t \geq 0\}$ and ν_0 determine the finite-dimensional distributions of X.*

Proof. For $f \in L$ and $t \geq 0$, we have

$$\begin{aligned}\int f(y)P\{X(t) \in \mathrm{d}y\} &= E[f(X(t))] \\ &= E[E[f(X(t))|\mathscr{F}_0^X]] \\ &= E[T(t)f(X(0))] = \int T(t)f(x)\nu_0(\mathrm{d}x). \qquad (1.4.8)\end{aligned}$$

Since L is separating, $\nu_t(\Gamma) \equiv P\{X(t) \in \Gamma\}$ is determined. Similarly if $f \in L$ and $g \in B(E)$, then for $0 \leq t_1 \leq t_2$

$$\begin{aligned}E[f(X(t_1))g(X(t_2))] &= E[f(X(t_1))T(t_2 - t_1)g(X(t_1))] \\ &= \int f(x)T(t_2 - t_1)g(x)\nu_{t_1}(\mathrm{d}x) \qquad (1.4.9)\end{aligned}$$

and the joint distribution of $X(t_1)$ and $X(t_2)$ is determined. Proceeding in this manner, the proposition can be proved by induction. □

We conclude the section by defining the full generator of a process. A semigroup $\{T(t), t \geq 0\}$ on L is said to be *measurable* if $T(\cdot)f$ is measurable as a function on $([0, \infty), \mathscr{B}[0, \infty))$ for each $f \in L$.

Definition 1.4.15 (Full generator)

The *full generator* of a measurable contraction semigroup $\{T(t), t \geq 0\}$ on L is defined as the set

$$\hat{A} = \left\{(f, g) \in L \times L : \ T(t)f - f = \int_0^t T(s)g\mathrm{d}s, \quad t \geq 0\right\}. \qquad (1.4.10)$$

With reference to (1.4.10) cf. also Proposition 1.4.9-(c). Note that $\hat{A}$ is not, in general, single-valued. For example, if $L = B(\mathbb{R})$ with the sup norm and $T(t)f(x) \equiv f(x+t)$, then $(0,t) \in \hat{A}$ for each $g \in B(\mathbb{R})$ that is zero almost everywhere with respect to the Lebesgue measure.

1.5 The martingale problem

Since the finite-dimensional distributions of a Markov process are determined by a corresponding semigroup $\{T(t), t \geq 0\}$, they are in turn determined by its full generator $\hat{A}$ or by a sufficiently large set $A \subset \hat{A}$. One of the best approaches for determining when a set is "sufficiently large" is through the martingale problem of Stroock and Varadhan, which is based on the observation in the following proposition.

Proposition 1.5.1 *Let X be an E-valued, progressive Markov process with transition function $P(t,x,\Gamma)$ and let $\{T(t), t \geq 0\}$ and $\hat{A}$ be as above. If $(f,g) \in \hat{A}$ then*

$$M(t) \equiv f(X(t)) - \int_0^t g(X(s))\mathrm{d}s \tag{1.5.1}$$

is an $\{\mathscr{F}_t^X\}$-martingale.

The idea of Stroock and Varadhan (1979) is that of using this martingale property as a means of characterizing the Markov process associated with a given generator A. By a *solution of the martingale problem for A* we mean a measurable stochastic process $\{X(t), t \geq 0\}$ with values in the metric space E defined on some probability space $(\Omega, \mathscr{F}, P)$ such that for each $(f, g) \in A$, (1.5.1) is a martingale with respect to the filtration

$$*\mathscr{F}_t^X \equiv \mathscr{F}_t^X \vee \sigma\left(\int_0^s h(X(u))du : s \leq t, h \in B(E)\right). \tag{1.5.2}$$

Note that if $\{X(t), t \geq 0\}$ is progressive, in particular if it is right continuous, then $*\mathscr{F}_t^X = \mathscr{F}_t^X$. In general, every event in $*\mathscr{F}_t^X$ differs from an event in $\mathscr{F}_t^X$ by an event of probability zero. If $\{\mathscr{G}_t\}$ is a filtration with $\mathscr{G}_t \supset *\mathscr{F}_t^X$ for all $t \geq 0$, and (1.5.1) is a $\{\mathscr{G}_t\}$-martingale for all $(f, g) \in A$, we say that X is a *solution of the martingale problem for A with respect to* $\{\mathscr{G}_t\}$. When an initial distribution $\mu \in \mathscr{P}(E)$ is specified, we say that a solution X of the martingale problem for A is a *solution of the martingale problem for* (A, μ) if $PX(0)^{-1} = \mu$.

Usually X has sample paths in $D_E[0, \infty)$, the space of *càdlàg*[1] functions from $[0, \infty)$ to E. It is convenient to call a probability measure $P \in \mathscr{P}(D_E[0, \infty))$ a *solution of the martingale problem for A* (or for (A, μ)) if the coordinate process defined on $(D_E[0, \infty), \mathscr{S}_E{}^2, P)$ by

$$X(t, \omega) \equiv \omega(t) \qquad \omega \in D_E[0, \infty), \quad t \geq 0 \tag{1.5.3}$$

is a solution of the martingale problem for A (or for (A, μ)) as defined above.

[1]See Section 3.3 below.
[2] See Section 3.4 below.

Note that a measurable process $\{X(t), t \geq 0\}$ is a solution of the martingale problem for A if and only if

$$\begin{aligned}
0 &= E\left[\left(f(X(t_{n+1})) - f(X(t_n)) - \int_{t_n}^{t_{n+1}} g(X(s))\mathrm{d}s\right)\prod_{k=1}^{n} h_k(X(t_k))\right] \\
&= E\left[f(X(t_{n+1}))\prod_{k=1}^{n} h_k(X(t_k))\right] - E\left[f(X(t_n))\prod_{k=1}^{n} h_k(X(t_k))\right] \\
&\quad - \int_{t_n}^{t_{n+1}} E\left[g(X(s))\prod_{k=1}^{n} h_k(X(t_k))\right]\mathrm{d}s
\end{aligned}$$

whenever $0 \leq t_1 < \cdots < t_{n+1}$, $(f, g) \in A$ and $h_1, \ldots, h_n \in B(E)$. Consequently the statement that a measurable process is a solution of a martingale problem is a statement about its finite-dimensional distributions. Note also that if $A^1 \subset A^2$, then any solution of the martingale problem for A^2 is also a solution for A^1, but the converse is not necessarily true.

We say that *uniqueness* holds for solutions of the martingale problem for (A, μ) if any two solutions have the same finite-dimensional distributions.

Definition 1.5.2 (Well posed martingale problem)
We say that the martingale problem for (A, μ) is *well-posed* if there exists a solution of the martingale problem for (A, μ) and uniqueness holds.

If this is true for all $\mu \in \mathscr{P}(E)$, then the martingale problem for A is said to be *well-posed*. We say that the martingale problem for (A, μ) is well-posed in $D_E[0, \infty)$ (respectively in $C_E[0, \infty)$, the space of continuous functions from $[0, \infty)$ to E) if there is a unique solution $P \in \mathscr{P}(D_E[0, \infty))$ ($P \in \mathscr{P}(C_E[0, \infty))$). Note that a martigale problem may be well-posed in $D_E[0,\infty)$ without being well-posed, that is uniqueness may hold under the restriction that the solution have sample paths in $D_E[0,\infty)$, but not in general. The terminology $D_E[0,\infty)$ martingale problem and $C_E[0,\infty)$ martingale problem indicate when a specific sample path behavior is required.

The following theorem says essentially that a Markov process is the unique solution of the martingale problem for its generator.

Theorem 1.5.3 (Markov solution of a martingale problem)
Let E be separable, and let $A \in B(E) \times B(E)$ be linear and dissipative. Suppose there exists $A' \in A$, A' linear and such that $\mathscr{R}(\lambda - A') = \mathscr{D}(A') \equiv L$ for some $\lambda > 0$, and L is separating. Let $\mu \in \mathscr{P}(E)$ and suppose X is a solution of the martingale problem for (A,μ). Then X is a Markov process corresponding to the semigroup on L generated by the closure of A', and uniqueness holds for the martingale problem for (A,μ).

Under the conditions of Theorem 1.5.3 every solution of the martingale problem for (A,μ) is Markovian. The following theorem shows that uniqueness of the solution of the martingale problem always implies the Markov property.

Theorem 1.5.4 (Uniqueness implies Markov solution)
Let E be separable, and let $A \in B(E) \times B(E)$. Suppose that for each $\mu \in \mathscr{P}(E)$ any two solutions X, Y of the martingale problem for (A,μ) have the same one-dimensional distributions, that is for each $t > 0$

$$P\{X(t) \in \Gamma\} = P\{Y(t) \in \Gamma\} \qquad \Gamma \in \mathscr{B}(E). \tag{1.5.4}$$

Then any solution of the martingale problem for A with respect to a filtration $\{\mathscr{G}_t\}$ is a Markov process with respect to $\{\mathscr{G}_t\}$, and any two solutions of the martingale problem for (A, μ) have the same finite-dimensional distributions (i.e. (1.5.4) *implies uniqueness).*

Although we do not want enter into details on the problem of the existence of solutions of a martingale problem, we mention a result that will be useful later. The following lemma, that concludes this section, indicates that one of the simplest ways of obtaining solutions is as weak limits of solutions of approximating martingale problems.

Lemma 1.5.5 *Let $A \subset \bar{C}(E) \times \bar{C}(E)$ and $A_n \subset B(E) \times B(E)$, $n = 1, 2, \ldots$. Suppose that for each $(f, g) \in A$ there exist $(f_n, g_n) \in A_n$ such that*

$$\lim_{n \to \infty} ||\, f_n - f|| = \lim_{n \to \infty} ||\, g_n - g|| = 0 \tag{1.5.5}$$

where $||\cdot||$ is the sup norm. If, for each n, X_n is a solution of the martingale problem for A_n with sample paths in $D_E[0, \infty)$, and if $X_n \Rightarrow X$[3], then X is a solution of the martingale problem for A.

[3] See Chapter 3 below.

Chapter 2

Urn-based exchangeable sequences and related priors

Urn schemes have a long tradition in the literature, the most famous probably being the Pólya urn, introduced by Eggenberger and Pólya (1923) for describing contagious diseases, and in the decades have proved to be powerful tools in handling with probabilistic and statistical models. In Bayesian nonparametrics, urn models have a role of primary importance, given that they allow alternative representations of widely used priors: from the celebrated result of Blackwell and MacQueen (1973) regarding the Dirichlet process to more recent works, like for example those of Pitman (1995, 1996b) on the two-parameter Poisson-Dirichlet process. These are indeed the priors that will be treated here, which will be often referenced to in the second part.

After introducing a general urn scheme for exchangeable sequences, we give the definition of the Dirichlet process, and describe the urn scheme that generates a sequence of random variables whose de Finetti measure is a Dirichlet process itself. The same is done for the two-parameter Poisson-Dirichlet process. Eventually we briefly introduce the Dirichlet process mixture model model, which will be used in the last chapters.

2.1 Exchangeable sequences

Consider the partition of $\{1,\ldots,n\}$ denoted by $\mathbf{p}_n = \{C_{j,n} : j = 1,\ldots,k\}$, where $C_{j,n}$ is the j-th set of the partition. Then $\mathbf{p}_n$ is made of k sets, where k is the number of unique values $x_1^*,\ldots,x_k^*$ in the observed sequence $x_1,\ldots,x_n$. Thus $x_l = x_j^*$ for each $l \in C_{j,n}$, with $j = 1,\ldots,k$. Define also $c_{j,n}$ to be the cardinality of the set $C_{j,n}$, such that $\sum_{j=1}^k c_{j,n} = n$.

Let now ν_0 be a finite probability measure on a complete separable metric space E, and $X_1, X_2, \ldots$ a sequence generated by the generalised Pólya-urn scheme defined by

$$X_1 \sim \nu_0 \tag{2.1.1}$$

and for $n \geq 1$

$$X_{n+1} | \, X_1,\ldots,X_n \sim \frac{q_{0,n}}{\sum_{j=0}^k q_{j,n}} \, \nu_0 + \sum_{j=1}^k \frac{q_{j,n}}{\sum_{j=0}^k q_{j,n}} \, \delta_{x_j^*} \tag{2.1.2}$$

where $q_{0,n}$ and $q_{j,n}$ are weights to be defined and δ_x is the point mass at x.

A sufficient condition for the sequence $X_1, X_2, \ldots$ to be exchangeable is given by Ishwaran and Zarepour (2003).

Condition 2.1.1 (Exchangeability)

A sufficient condition for the sequence defined by (2.1.1) and (2.1.2) *to be exchangeable* *is that jointly:*

(1) ν_0 *is a non atomic non null probability measure;*

(2) $q_{0,n} = \psi_0(k)$ *for some fixed non negative real valued function* ψ_0 *of the number of clusters;*

(3) $q_{j,n} = \psi(c_{j,n})$ *for some fixed non negative real valued functions* ψ *of the cardinality of each cluster;*

(4) $\sum_{j=0}^{k} q_{j,n} = \xi(n)$ *for some fixed positive real valued function* ξ *of the number of already observed values.*

The one-step-backward version of (2.1.2) in the exchangeable case therefore is

$$X_n|\, X_1, \ldots, X_{n-1} \sim \frac{\psi_0(k)}{\xi(n-1)}\, \nu_0 + \sum_{j=1}^{k} \frac{\psi(c_{j,n-1})}{\xi(n-1)}\, \delta_{x_j^*} \tag{2.1.3}$$

where it is clear that all the parameters, included k, are computed on $n-1$ elements. From (2.1.3), by exchangeability, we can determine the full conditional distribution for any X_i, i.e.

$$X_i|\, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n \sim \frac{\psi_0(k_i)}{\xi(n-1)}\, \nu_0 + \sum_{j=1}^{k_i} \frac{\psi(c_{j,n-1})}{\xi(n-1)}\, \delta_{x_j^*} \tag{2.1.4}$$

where k_i denotes the number of unique values in $\mathbf{x}_{-i} = \{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\}$, and $c_{j,n-1}$ is now obviously computed on $\mathbf{x}_{-i}$ instead of $\{x_1, \ldots, x_{n-1}\}$.

For notational simplicity we will often use c_j, instead of $c_{j,n-1}$, say, when this creates no ambiguity.

A general result for infinite exchangeable sequences is the celebrated de Finetti's representation theorem, which states that the law of an infinite exchangeable sequence of random variables is a mixture of laws of i.i.d. sequences.

Theorem 2.1.2 (de Finetti's representation)
Let $\{X_n,\, n \geq 1\}$ be a sequence of random variables taking values in a complete separable metric space E, endowed with its Borel σ-field $\mathscr{B}(E)$. The sequence $\{X_n,\, n \geq 1\}$ is exchangeable if and only if there exists a probability measure ν on $\mathscr{P}(E)$, the set of all probability measures on E, endowed with the topology of weak convergence, such that

$$P\{(X_1, X_2, \dots) \in A\} = \int_{\mathscr{P}(E)} \mu^\infty(A)\nu(\mathrm{d}\mu)$$

for all $A \in \mathscr{B}(E^\infty)$. Moreover ν is the unique probability distribution of the weak limit of

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$$

and is known as the de Finetti measure of the sequence $X_1, X_2, \dots$.

In the next sections we will provide some examples of functions $\psi_0(k)$ and $\psi(c_{j,n})$ which lead to well known prior distribution on the set $\mathscr{P}(E)$, namely the Dirichlet process and the two-parameter Poisson-Dirichlet process. For more details on exchangeability, see for example Aldous (1985).

2.2 The Dirichlet process

The Dirichlet process is a probability measure on the space $\mathscr{P}(E)$, i.e. the set of probability distributions on E, introduced by Ferguson (1973) as a prior for use in a Bayesian nonparametric framework. Before stressing the derivation of the Dirichlet process from the construction given in the previous section, we proceed with some definitions.

Definition 2.2.1 (Dirichlet distribution)

Let $\alpha_i > 0$ for $i = 1, \dots, n$, $n \geq 1$. The vector $(p_1, \dots, p_n)$ is said to have the Dirichlet distribution with parameter $(\alpha_1, \dots, \alpha_n)$, denoted by $\mathcal{D}(\alpha_1, \dots, \alpha_n)$, if its density is

$$\Pi(p_1, \dots, p_n) = \frac{\Gamma(\sum_{i=1}^n \alpha_i)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_n)} p_1^{\alpha_1 - 1} \dots p_{n-1}^{\alpha_{n-1}-1} \big(1 - \sum_{i=1}^{n-1} p_i\big)^{\alpha_n - 1} \tag{2.2.1}$$

for $(p_1, \dots, p_{n-1})$ in

$$\Delta_n = \big\{(p_1, \dots, p_{n-1}) : p_i \geq 0 \text{ for } i = 1, \dots, n-1, \sum p_i \leq 1\big\}$$

and where

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \mathrm{d}t.$$

If any $\alpha_i = 0$, we still have a Dirichlet distribution by setting the corresponding $p_i = 0$ and interpreting (2.2.1) on a lower-dimensional set.

The Dirichlet distribution admits representation in terms of gamma variables. Let $Z_1, \dots, Z_n$ be independent gamma random variables with parameter $\alpha_i \geq 0$, for $i = 1, \dots, n$. Then

$$\left(\frac{Z_1}{\sum_{i=1}^n Z_i}, \dots, \frac{Z_n}{\sum_{i=1}^n Z_i}\right) \tag{2.2.2}$$

has distribution $\mathcal{D}(\alpha_1, \ldots, \alpha_n)$.

The Dirichlet process is the natural generalisation of the Dirichlet distribution defined above.

Definition 2.2.2 (Dirichlet process)

Let α be a finite measure on $(E, \mathscr{B}(E))$. The unique probability measure on $\mathscr{P}(E)$ such that for every partition $A_1, \ldots, A_n$, with $A_i \in \mathscr{B}(E)$ for each i,

$$(\mu(A_1), \ldots, \mu(A_n)) \sim \mathcal{D}(\alpha_1, \ldots, \alpha_n) \tag{2.2.3}$$

is called Dirichlet process with parameter α, denoted by $\mathscr{D}(\alpha)$ or equivalently Π_α.

If $\mu \sim \mathscr{D}(\alpha)$, then $E(\mu(A)) = \bar{\alpha} = \alpha(A)/\alpha(E)$, the Dirichlet distribution being a generalisation of the beta distribution, and $E(\mu) = \bar{\alpha}$ is called the prior expectation of P.

The following theorem states the property of conjugacy for the Dirichlet process.

Theorem 2.2.3 (Posterior distribution)

For each μ in $\mathscr{P}(E)$, let $X_1, \ldots, X_n$ be i.i.d. μ and let μ itself be distributed as $\mathscr{D}(\alpha)$, where α is a finite measure. The posterior distribution of μ given $X_1, \ldots, X_n$ is

$$\Pi_\alpha(\cdot \mid X_1 = x_1, \ldots, X_n = x_n) = \Pi_{\alpha + \sum_{i=1}^n \delta_{x_i}}(\cdot) \tag{2.2.4}$$

where δ_x denotes point mass at x.

The following property of the Dirichlet distribution yields a useful representation of the posterior distribution, used in Chapter 6. See Ghosh and Ramamoorthi (2003).

Proposition 2.2.4 *Let α_1 and α_2 be two measures on E and μ_1 and μ_2 be two independent k-dimensional Dirichlet random vectors with parameters α_1 and α_2 respectively. If Y independent of μ_1, μ_2 is distributed as* beta($\alpha_1(E)$,$\alpha_2(E)$)*, then*

$$Y\,\mu_1 + (1-Y)\,\mu_2 \;\sim\; \mathcal{D}(\alpha_1+\alpha_2).$$

This can be easily seen using the representation (2.2.2). An interesting consequence is the following.

Proposition 2.2.5 *If μ is $\mathcal{D}(\alpha)$ as in Proposition* 2.2.4 *and Y is independent of μ and distributed as a* beta(n,$\alpha(E)$)*, then*

$$Y\,\delta_{(1,0,\ldots,0)} + (1-Y)\,\mu \;\sim\; \mathcal{D}(\alpha\{1\}+n, \alpha\{2\}, \ldots, \alpha\{k\}).$$

This follows if we think of $\delta_{(1,0,\ldots,0)}$ as a Dirichlet with parameter $(n, 0, \ldots, 0)$. A corresponding statement holds if $(1, 0, \ldots, 0)$ is replaced by any vector with a 1 at one coordinate and 0 at the other coordinates.

The next result states that if the random probability μ has probability distribution Π_α, then μ is discrete almost surely.

Theorem 2.2.6 (Discreteness)

Let Π_α be a Dirichlet process of parameter α. Then

$$\Pi_\alpha\left(\{\mu \in \mathscr{P}(E): \ \mu(\{x \in E: \ \mu(\{x\}) > 0\}) = 1\}\right) = 1.$$

Ferguson (1973) also introduced an alternative construction of the Dirichlet process. First we need a definition.

Definition 2.2.7 (Poisson-Dirichlet distribution)

Let $\{Z_t, \, t \geq 0\}$ be a stochastic process with independent increments such that, for each t, $Z_t \sim \text{Gamma}(t,1)$[1]. Let $V_1 > V_2 > \ldots$ the ordered sizes of the jumps of the process on $[0, \theta]$. Then, for $n \geq 1$, the random variables

$$p_n = \frac{V_n}{Z_\theta} \tag{2.2.5}$$

are said to have the Poisson-Dirichlet distribution with parameter θ, denoted by $(p_n)_{n \geq 1} \sim \mathcal{PD}(0, \theta)$.

Then we have the following.

Definition 2.2.8 (Dirichlet process)

Let α be a finite measure on E, let $p_1, p_2, \ldots$ have the Poisson-Dirichlet distribution

[1]Such a process is called gamma subordinator.

with parameter $\alpha(E)$, and $Y_1, Y_2, \ldots$ be i.i.d. $\bar{\alpha}$, where $\bar{\alpha} = \alpha/\alpha(E)$. Then the random probability measure

$$\sum_{i=1}^{\infty} p_i \delta_{Y_i} \tag{2.2.6}$$

is a Dirichlet process of parameter α.

The following is a constructive definition of the Dirichlet process, introduced by Sethuraman (1994), which will be useful as a starting point for the definition of the two-parameter Poisson-Dirichlet process in Section 2.3 below.

Definition 2.2.9 (Sethuraman construction)

Let $W_1, W_2, \ldots$ be i.i.d. beta(1,$\alpha(E)$) random variables, and define the sequence of weights $p_1, p_2, \ldots$ such that $p_1 = W_1$ and, for $n \geq 2$

$$p_n = W_n \prod_{i=1}^{n-1} (1 - W_i). \tag{2.2.7}$$

Also let $Y_1, Y_2, \ldots$ be i.i.d. $\bar{\alpha}$, where $\bar{\alpha} = \alpha/\alpha(E)$. Then the random probability measure

$$\sum_{i=1}^{\infty} p_i \delta_{Y_i} \tag{2.2.8}$$

is a Dirichlet process of parameter α.

Remark 2.2.10 The construction (2.2.7) is called *residual allocation model*, also known as *stick-breaking scheme*.

The similarity of the representations of the Dirichlet process in Definitions 2.2.8 and 2.2.9 is striking, and is no coincidence. The explanation of the connection between them will be given in Section 2.3 below.

Recall now the urn-based sequence constructed in Section 2.1, which is exchangeable if Condition 2.1.1 holds, and consider the special case in which the parameters are

$$\psi_0(k_i) = \theta \tag{2.2.9}$$

$$\psi(c_j) = c_j \tag{2.2.10}$$

recalling that k_i denotes the number of unique values in the subvector $\mathbf{x}_{-i}$ and c_j denotes the cardinality of the j-th cluster. Hence we have

$$\begin{aligned}\xi(n-1) &= \sum_{j=0}^{k_i} q_{j,n} \\ &= \psi_0(k_i) + \sum_{j=1}^{k_i} \psi(c_j) \\ &= \theta + \sum_{j=1}^{k_i} c_j = \theta + n - 1.\end{aligned}$$

Substituting (2.2.9) and (2.2.10) in (2.1.3) we obtain

$$X_n|\, X_1, \ldots, X_{n-1} \sim \frac{\theta}{\theta + n - 1}\, \nu_0 + \sum_{j=1}^{k_i} \frac{c_j}{\theta + n - 1}\, \delta_{x_j^*} \tag{2.2.11}$$

which, expanding the cluster notation, can be written

$$X_n|\, X_1, \ldots, X_{n-1} \sim \frac{\theta}{\theta + n - 1}\, \nu_0 + \sum_{i=1}^{n-1} \frac{1}{\theta + n - 1}\, \delta_{x_i}. \tag{2.2.12}$$

The scheme (2.2.12) is a generalised Pólya-urn, usually called Blackwell-MacQueen urn scheme. This is due to the fact that Blackwell and MacQueen (1973) showed that the urn scheme (2.2.12) generates a sequence of random variables whose de Finetti measure is a Dirichlet process, shedding new light on the newly introduced Dirichlet process prior.

Definition 2.2.11 (Pólya sequence) A sequence $\{X_n,\, n \geq 1\}$ of random variables with values in the complete separable metric space E is called a Pólya sequence with parameter α if for every $A \subset E$

$$P(X_1 \in B) = \frac{\alpha(A)}{\alpha(E)}$$

and

$$P(X_{n+1} \in B|\, X_1, \ldots, X_n) = \frac{\alpha(A) + \sum_{i=1}^{n} \delta_{X_i}(A)}{\alpha(E) + n}.$$

The result of Blackwell and MacQueen (1973) is the following. Denote

$$m_n = \frac{\alpha + \sum_{i=1}^{n} \delta_{X_i}}{\alpha(E) + n}.$$

Theorem 2.2.12 (Pólya sequences and Dirichlet process)
Let $\{X_n,\, n \geq 1\}$ be a Pólya sequence with parameter α. Then

(a) *as $n \to \infty$, with probability one m_n weakly converges to a discrete random probability measure μ;*

(b) *μ has distribution $\mathscr{D}(\alpha)$, i.e. a Dirichlet process with parameter α;*

(c) *given μ, the variables $X_1, X_2, \dots$ are independent with distribution μ, that is the Pólya sequence $\{X_n,\, n \geq 1\}$ is exchangeable and its de Finetti measure is $\mathscr{D}(\alpha)$.*

2.3 The two-parameter Poisson-Dirichlet process

The two-parameter Poisson-Dirichlet process, extensively described in Pitman and Yor (1997) and thus also known as Pitman-Yor process, is a distribution on the space of probability measures which generalises the Dirichlet process, in that the latter can be obtained as a special case of the former. The definition will be given in several steps.

Definition 2.3.1 (GEM distribution)
Let $\sigma \in [0, 1)$ and $\theta > -\sigma$. Let $W_1, W_2, \dots$ be independent random variables with

$$W_n \sim \text{beta}(1 - \sigma, \theta + n\sigma) \tag{2.3.1}$$

and define $Q_1, Q_2, \dots$ by the residual allocation model (2.2.7). Then $Q_1, Q_2, \dots$ is said to have the GEM distribution with parameters σ and θ, denoted by $GEM(\sigma, \theta)$.

Ewens (1988) named the above distribution GEM after Griffths, Engen, and McCloskey, who contributed to the study of the special case with $\sigma = 0$. The Sethuraman construction (2.2.9) is based upon the $GEM(0, \theta)$ distribution.

In view of the next definition, consider a sequence of probabilities $p_1, p_2, \ldots,$ which represent the frequencies in a population of infinitely-many distinct type, labeled by $1, 2, \ldots$. Sampling from this population induces a permutation of $p_1, p_2, \ldots$ according to the order in which the different types are observed, denoted by $\tilde{p}_1, \tilde{p}_2, \ldots$. That is, $\tilde{p}_1 = p_n$ if the first observation is of type n, $\tilde{p}_2 = p_m$ if the next observation not of type n is of type m, and so on.

Definition 2.3.2 (Size-biased permutation)

Let $p_1, p_2, \ldots$ be a sequence of probabilities, and define the sequence $\tilde{p}_1, \tilde{p}_2, \ldots$ by

$$P\{\tilde{p}_1 = p_n | \, p_1, p_2, \ldots\} = p_n \qquad (2.3.2)$$

and, for $i \geq 1$

$$P\{\tilde{p}_{i+1} = p_n | \, \tilde{p}_1, \ldots, \tilde{p}_i, p_1, p_2, \ldots\} = \frac{p_n}{1 - \tilde{p}_1 - \cdots - \tilde{p}_i} \, \mathrm{I}(p_n \neq \tilde{p}_1, \ldots, \tilde{p}_i). \qquad (2.3.3)$$

Then $(\tilde{p}_n)_{n\geq 1}$ is called a size-biased permutation of $(p_n)_{n\geq 1}$.

Definition 2.3.3 (ISBP)

If

$$(\tilde{p}_n)_{n\geq 1} \stackrel{d}{=} (p_n)_{n\geq 1}$$

then we say that $(p_n)_{n\geq 1}$ is invariant under size-biased permutations, or ISBP.

The following theorem is due to McCloskey (1965).

Theorem 2.3.4 (Poisson-Dirichlet and GEM)
Suppose $(p_n)_{n\geq 1} \sim \mathcal{PD}(0,\theta)$ *and let* $(\tilde{p}_n)_{n\geq 1}$ *be a size-biased permutation of* $(p_n)_{n\geq 1}$. *Then* $(\tilde{p}_n)_{n\geq 1} \sim GEM(0,\theta)$.

Given the previous result, since the order of the terms in the series (2.2.6) and (2.2.8) is clearly irrelevant, we can now understand why Ferguson's and Sethuraman's representations yield the same process.

The next theorem is from Pitman (1996a).

Theorem 2.3.5 (Maximal ISBP family)
Let $(p_n)_{n\geq 1}$ *be such that* $p_n > 0$ *a.s. for all* n *and* $\sum_n p_n = 1$, *and suppose* $(p_n)_{n\geq 1}$ *follows the residual allocation model* (2.2.7) *for independent* W_i. *Then* $(p_n)_{n\geq 1}$ *is ISBP if and only if* $(p_n)_{n\geq 1} \sim GEM(\sigma,\theta)$ *for some* $\sigma \in [0,1)$ *and* $\theta > -\sigma$.

That is, the $GEM(\sigma,\theta)$ distribution is the maximal non-degenerate ISBP family that we can realize under the residual allocation model.

Definition 2.3.6 (Two-parameter Poisson-Dirichlet distribution)
Suppose $(p_n)_{n\geq 1} \sim GEM(\sigma,\theta)$. Then the ranked sequence $(p_{(n)}) = (p_{(1)}, p_{(2)}, \dots)$ has a (two-parameter) Poisson-Dirichlet distribution with parameters (σ,θ), denoted by $\mathcal{PD}(\sigma,\theta)$.

So the size-biased permutation of a $\mathcal{PD}(\sigma,\theta)$ distribution is $GEM(\sigma,\theta)$, and the ranked sequence of a $GEM(\sigma,\theta)$ distribution is $\mathcal{PD}(\sigma,\theta)$.

Definition 2.3.7 (Two-parameter Poisson-Dirichlet process)
Let ν_0 be a probability measure on E, and let $Y_1, Y_2, \ldots$ be i.i.d. ν_0. Let $\sigma \in [0,1)$ and $\theta > -\sigma$, and let $(p_n)_{n\geq 1} \sim \mathcal{PD}(\sigma,\theta)$. The random probability measure

$$\sum_{i=1}^{\infty} p_i \delta_{Y_i} \tag{2.3.4}$$

is called Poisson-Dirichlet process with parameters (σ,θ), denoted by $\mathscr{PD}(\nu_0;\sigma,\theta)$ or simply by $\mathscr{PD}(\sigma,\theta)$.

Observe that for $\sigma = 0$, (2.3.4) reduces to (2.2.6), that is

$$\mathscr{PD}(\nu_0; 0, \theta) \equiv \mathscr{D}(\theta\nu_0). \tag{2.3.5}$$

Remark 2.3.8 Note that, as for the Dirichlet case, we may use $(p_n)_{n\geq 1} \sim GEM(\sigma,\theta)$ in (2.3.4) and get an equivalent definition of the two-parameter Poisson-Dirichlet process.

Return now to the framework of Condition 2.1.1, where the general urn scheme was

$$X_n | \, X_1, \ldots, X_{n-1} \sim \frac{\psi_0(k)}{\xi(n-1)}\, \nu_0 + \sum_{j=1}^{k} \frac{\psi(c_{j,n-1})}{\xi(n-1)}\, \delta_{x_j^*} \tag{2.3.6}$$

and take the special case of the parameters

$$\psi_0(k) = \theta + \sigma k \tag{2.3.7}$$

$$\psi(c_{j,n-1}) = c_{j,n-1} - \sigma \tag{2.3.8}$$

where $0 \leq \sigma < 1$ and $\theta > -\sigma$. Then $\xi(n-1)$ equals again $\theta + n - 1$, since

$$\begin{aligned}\xi(n-1) &= \psi_0(k) + \sum_{j=1}^{k} \psi(c_{j,n-1}) \\ &= \theta + \sigma k + \sum_{j=1}^{k}(c_{j,n-1} - \sigma) = \theta + n - 1.\end{aligned}$$

and we obtain the generalised Pólya urn scheme

$$X_n | \, X_1, \ldots, X_{n-1} \sim \frac{\theta + \sigma k}{\theta + n - 1}\, \nu_0 + \frac{1}{\theta + n - 1} \sum_{j=1}^{k} (c_{j,n-1} - \sigma)\delta_{x_j^*}. \tag{2.3.9}$$

Pitman (1995, 1996b) showed that (2.3.9) is the prediction rule for the random discrete probability measure μ defined by

$$\sum_{i=1}^{\infty} p_i \delta_{Y_i} \tag{2.3.10}$$

where $p_1 = V_1$, $p_n = V_n \prod_{i=1}^{n-1}(1 - V_i)$ for $n \geq 2$ and the V_i's are i.i.d. beta$(1 - \sigma, \theta + n\sigma)$ random variables independent of the Y_i's, which are i.i.d. ν_0. That is the two-parameter Poisson-Dirichlet process (cf. Remark 2.3.8). In other words, (2.3.9) generates a sequence of random variables which are i.i.d. μ, conditionally on μ which is sampled from $\mathscr{PD}(\sigma, \theta)$ (see also Ishwaran and James, 2001; Ishwaran and Zarepour, 2003).

Note that for $\sigma = 0$ we recover the Blackwell-MacQueen urn scheme, hence the Dirichlet case.

2.4 The Dirichlet process mixture model

The Dirichlet process mixture model, to be distinguished from the mixture of Dirichlet process model of Antoniak (1974), was first considered by Lo (1984). Developed in the framework of Ferguson (1973), its purpose is to achieve a probability measure supported by densities. A random density function is chosen as

$$f_G(y) = \int K(y|\,\theta)\mathrm{d}G(\theta)$$

where G is distributed as a Dirichlet process. Provided $K(y|\,\theta)$ is a density for each θ, then f_G will be a density. In the model of Antoniak (1974) the Dirichlet process was the integrand instead of the mixing measure. The problem of obtaining posterior summaries can be solved via computer simulation from the posterior distribution. Escobar (1988) was the first to introduce a simulation method for this model, based on Markov Chain Monte Carlo algorithms (see Section 1.3).

The basic normal mixture model, considered in Escobar and West (1995), is described as follows. Suppose that

$$\begin{aligned} Y_1, \ldots, Y_n|\,\pi_1, \ldots, \pi_n &\sim \prod_{i=1}^{n} N(\mu_i, V_i) \\ \pi_i = (\mu_i, V_i) &\overset{iid}{\sim} G \qquad i = 1, \ldots, n \end{aligned} \tag{2.4.1}$$

that is the data $Y_1, \ldots, Y_n$ are conditionally independent given $\pi_1, \ldots, \pi_n$ and normally distributed with parameters $\pi_i = (\mu_i, V_i)$, for $i = 1, \ldots, n$, where the parameters are distributed according to some prior distribution G on $\mathbb{R} \times \mathbb{R}^+$. Then, given the first n observation $y_1, \ldots, y_n$, the distribution of Y_{n+1} is a mixture of normals, with mixing measure given by the posterior predictive distribution of $\pi_{n+1}|\,y_1, \ldots, y_n$. If the common prior distribution for the π_i is modeled as a Dirichlet process, then the data come from a Dirichlet mixture of normals (see Escobar, 1988, 1994; Ferguson, 1983; West, 1990).

A key feature of the model is the discreteness of G under the Dirichlet process assumption (see Theorem 2.2.6). Suppose for example that $G \sim \mathscr{D}(\theta, \nu_0)$, where $\mathscr{D}(\theta, \nu_0)$ is the Dirichlet process with prior expectation ν_0 and precision parameter θ. Denote $\pi = (\pi_1, \ldots, \pi_n)$ and $\pi_{-i} = (\pi_1, \ldots, \pi_{i-1}, \pi_{i+1}, \ldots, \pi_n)$. Then the conditional prior for $\pi_i | \, \pi_{-i}$ is

$$\pi_i | \, \pi_{-i} \sim \frac{\theta \nu_0(\pi_i) + \sum_{j=1, j\neq i}^{n} \delta_{\pi_j}(\pi_i)}{\theta + n - 1} \tag{2.4.2}$$

where $\delta_x(y)$ denotes a point mass at $y = x$. Similarly, the distribution of $\pi_{n+1} | \, \pi$ is given by

$$\pi_{n+1} | \, \pi_1 \ldots, \pi_n \sim \frac{\theta \nu_0(\pi_{n+1}) + \sum_{i=1}^{n} \delta_{\pi_i}(\pi_{n+1})}{\theta + n}. \tag{2.4.3}$$

Thus, given the sample π of size n from G, the next observation π_{n+1} represent a new value with probability $\theta/(\theta + n)$ and is otherwise drawn uniformly among the first n values.

Denote now with $\pi_j^* = (\mu_j^*, V_j^*)$ the k distinct values in π_{-i}, for $j = 1, \ldots, k$, and with c_j the number of occurrences of π_j^* in π_{-i}. Immediately (2.4.2) reduces to the mixture of fewer components, that is

$$\pi_i | \, \pi_{-i} \sim \frac{\theta \nu_0(\pi_i) + \sum_{j=1}^{k} c_j \delta_{\pi_j^*}(\pi_i)}{\theta + n - 1}. \tag{2.4.4}$$

Consider now the general formulation of the model. Dropping the assumption of normality generalises the hierarchical structure given in (2.4.1), so that we have

$$\begin{aligned}
&Y_1,\dots,Y_n|\,\pi_1,\dots,\pi_n \sim \prod_{i=1}^{n} K(\cdot|\,\pi_i)\\
&\pi_i|\,G \overset{iid}{\sim} G \qquad i=1,\dots,n\\
&G \sim \mathscr{D}(\theta,\nu_0).
\end{aligned}$$

Note that it is possible to include in the model a further level in the hierarchy, to account for prior distributions for the parameters of the Dirichlet process. For simplicity we omit this case.

It is well known that G can be integrated out of the nonparametric model, leading to the revised, but equivalent, hierarchical model

$$\begin{aligned}
&Y_i|\,\pi_i \sim K(\cdot|\,\pi_i) \qquad i=1,\dots,n\\
&(\pi_1,\dots,\pi_n) \sim \nu_0(\pi_1)\prod_{i=2}^{n}\frac{\theta\nu_0(\pi_i)+\sum_{j=1}^{i-1}\delta_{\pi_j}(\pi_i)}{\theta+i-1}.
\end{aligned}$$

The posterior distribution is mathematically intractable and only sampling strategies are going to be able to obtain posterior summaries of interest. The idea of Escobar (1988) was to sample the posterior distribution of the parameters $(\pi_1,\dots,\pi_n)$ using a Markov chain, which he constructed by successively sampling from the full conditional densities $p\,(\pi_i|\,\pi_{-i},y_1,\dots,y_n)$, for $i=1,\dots,n$, where

$$p\,(\pi_i|\,\pi_{-i},y_1,\dots,y_n) \propto K(y_i|\,\pi_i)\left(\theta\nu_0(\pi_i)+\sum_{j=1,j\neq i}^{n}\delta_{\pi_j}(\pi_i)\right). \qquad (2.4.5)$$

This algorithm is what we know to be a Gibbs sampler (cf. Section 1.3).

We conclude the section with a few other references on the Dirichlet process mixture model. A review on the topic is given in MacEachern (1998). MacEachern (1994) develop an alternative sampling strategy still based on the Gibbs sampler. For more recent work on simulation of the Dirichlet process mixture model see MacEachern and Müller (1998), MacEachern *et al.* (1999), Neal (2000), Kottas and Gelfand (2001), Ishwaran and Zarepour (2000), Ishwaran and James (2002).

Chapter 3

Convergence of probability measures

In this chapter we report a few notions on weak convergence of probability measures, with particular attention to the space which will be the natural setting of the processes defined in the second part, that is the space of càdlàg functions, i.e. functions with discontinuities of the first kind. Details about the topics of this chapter can be found in Billingsley (1968) or Ethier and Kurtz (1986).

3.1 Weak convergence

Let (E, d) be a metric space. Denote with $\mathscr{B}(E)$ the σ-algebra of Borel subsets of E, and with $\mathscr{P}(E)$ the family of Borel probability measures on E. Let also $\bar{C}(E)$ be the space of real-valued bounded continuous functions on (E, d), with norm $\| f\| = \sup_{x\in E} | f(x)|$.

Definition 3.1.1 (Weak convergence)

A sequence $\{P_n\} \subset \mathscr{P}(E)$ is said to *converge weakly* to $P \in \mathscr{P}(E)$ if

$$\lim_{n\to\infty} \int f \mathrm{d}P_n = \int f \mathrm{d}P \qquad f \in \bar{C}(E).$$

The distribution of an E-valued random variable X, denoted by PX^{-1}, is the element of $\mathscr{P}(E)$ given by $PX^{-1}(B) = P\{X \in B\}$. A sequence $\{X_n\}$ of E-valued random variables is said to *converge in distribution* to the E-valued random variable X if $\{PX_n^{-1}\}$ converges weakly to PX^{-1}, or equivalently, if

$$\lim_{n\to\infty} E[f(X_n)] = E[f(X)] \qquad f \in \bar{C}(E). \tag{3.1.1}$$

Weak convergence is denoted by $P_n \Rightarrow P$ and convergence in distribution by $X_n \xrightarrow{d} X$. If $E = \mathbb{R}$, then (3.1.1) is equivalent to

$$\lim_{n\to\infty} P\{X_n \leq x\} = P\{X \leq x\} \tag{3.1.2}$$

for all x at which the right side of (3.1.2) is continuous.

Call A a P-continuity set if $A \in \mathscr{B}(E)$ and $P(\partial A) = 0$.

Theorem 3.1.2 (Portemanteau)

Let $\{P_n\} \subset \mathscr{P}(E)$ and $P \in \mathscr{P}(E)$. Then the following are equivalent:

(a) $P_n \Rightarrow P$.

(b) $\lim_{n\to\infty} \int f \mathrm{d}P_n = \int f \mathrm{d}P$ *for all uniformly continuous* $f \in \bar{C}(E)$.

(c) $\limsup_n P_n(F) \leq P(F)$ *for all closed sets* $F \subset E$.

(d) $\liminf_n P_n(G) \geq P(G)$ *for all open sets* $G \subset E$.

(e) $\lim_{n\to\infty} P_n(A) = P(A)$ *for all* P*-continuity sets* $A \subset E$.

3.2 The topology of weak convergence

A topology on the set $\mathscr{P}(E)$ of all probability measures on E can be induced by the Prohorov metric.

Definition 3.2.1 (Prohorov metric)

Let (E, d) be a metric space, and define

$$\rho(P, Q) = \inf\{\varepsilon > 0 : \ P(F) \leq Q(F^\varepsilon) + \varepsilon \text{ for all } F \in \mathscr{C}\}$$

where $\mathscr{C}$ is the collection of closed subsets of E and

$$F^\varepsilon = \left\{ x \in E : \ \inf_{y \in F} d(x, y) < \varepsilon \right\}.$$

ρ is a metric, called Prohorov metric.

Theorem 3.2.2 (Prohorov metric VS Portmanteau)

Let $\{P_n\} \subset \mathscr{P}(E)$ *and* $P \in \mathscr{P}(E)$, *and denote with* ρ *the Prohorov metric and with*

$$\lim_{n\to\infty} \rho(P_n, P) = 0 \tag{3.2.1}$$

convergence in the Prohorov metric. Then the following hold.

- *Let* (E, d) *be arbitrary. Then* (3.2.1) *implies* (a) *through* (e) *of Theorem* 3.1.2.
- *Let* (E, d) *be separable. Then* (3.2.1) *and* (a) *through* (e) *of Theorem* 3.1.2 *are equivalent.*

The metric space $(\mathscr{P}(E), \rho)$ is complete and separable whenever (E, d) is. Note that while separability is a topological property, completeness is a property of the metric.

Theorem 3.2.3 **($(\mathscr{P}(E), \rho)$ separable and complete)**

- *If* E *is separable, then* $\mathscr{P}(E)$ *is separable.*
- *If in addition* (E, d) *is complete, then* $(\mathscr{P}(E), \rho)$ *is complete.*

A common approach for verifying the convergence of a sequence of of Borel probability measures on (E, d) is to first show that the sequence is contained in some compact set and then to show that every convergent subsequence must converge to the same element. It is then comprehensible that a characterisation of the compact subsets of $\mathscr{P}(E)$ is crucial. This characterisation is given by the theorem of Prohorov that relates compactness to the notion of tightness. Let us first state these two concepts.

Definition 3.2.4 (Tightness)

A probability measure $P \in \mathscr{P}(E)$ is said to be *tight* if for each $\varepsilon > 0$ there exists a compact set $K \subset E$ such that

$$P(K) \geq 1 - \varepsilon.$$

A family of probability measures $\mathscr{M} \subset \mathscr{P}(E)$ is tight if for each $\varepsilon > 0$ there exists a compact set $K \subset E$ such that

$$\inf_{P \in \mathscr{M}} P(K) \geq 1 - \varepsilon.$$

Lemma 3.2.5 (Tightness in Polish spaces)

If (E, d) is complete and separable, then each $P \in \mathscr{P}(E)$ is tight.

Definition 3.2.6 (Relative compactness)

A family of probability measures $\mathscr{M} \subset \mathscr{P}(E)$ is said to be *relatively compact* if every sequence of elements of $\mathscr{M}$ contains a weakly convergent subsequence. That is, for every sequence $\{P_n\}$ in $\mathscr{M}$ there exists a subsequence $\{P_{n_i}\}$ and a probability measure Q defined on E (but not necessarily an element of $\mathscr{M}$) such that $P_{n_i} \Rightarrow Q$.

We can now state Prohorov's result.

Theorem 3.2.7 (Prohorov's theorem)

- *If $\mathscr{M}$ is tight, then it is relatively compact.*
- *Suppose E is separable and complete. If $\mathscr{M}$ is relatively compact, then it is tight.*

The next results are concerned with countably infinite product spaces.

Proposition 3.2.8 (Completeness and separability of product spaces)

Let (E_k, d_k), for $k = 1, 2, \ldots$ be metric spaces, and define the metric space (E, d) by letting $E = \prod_{k=1}^{\infty} E_k$ and

$$d(x, y) = \sum_{k=1}^{\infty} \frac{1}{2^k}(d_k(x_k, y_k) \wedge 1)$$

for all $x, y \in E$.

- *If all E_k's are separable, then E is separable.*
- *If all (E_k, d_k)'s are complete, then (E, d) is complete.*
- *If all E_k's are separable, then $\mathscr{B}(E) = \prod_{k=1}^{\infty} \mathscr{B}(E_k)$.*

Proposition 3.2.9 (Tightness in product spaces)

Let (E_k, d_k), for $k = 1, 2, \ldots$ be metric spaces, and define the metric space (E, d) by letting $E = \prod_{k=1}^{\infty} E_k$ and

$$d(x, y) = \sum_{k=1}^{\infty} \frac{1}{2^k}(d_k(x_k, y_k) \wedge 1)$$

for all $x, y \in E$. Let $\{P_\alpha\} \subset \mathscr{P}(E)$ (where α range over some index set) and for $k = 1, 2, \ldots$ and each α, define $P_\alpha^k \in \mathscr{P}(E_k)$ to be the k*-th marginal distribution of P_α (i.e. $P_\alpha^k = P_\alpha \pi_k^{-1}$, where the projection $\pi_k : E \to E_k$ is given by $\pi_k(x) = x_k$). Then $\{P_\alpha\}$ is tight if and only if $\{P_\alpha^k\}$ is tight for $k = 1, 2, \ldots$.*

3.3 The space $D_E[0,\infty)$ and the Skorohod topology

For the remainder of the chapter, let (E,r) denote a metric space, and let q denote the metric $r \wedge 1$.

Most part of stochastic processes present the property of having right and left limits at each time point for almost every sample path. It has become conventional to assume that sample paths are actually right continuous (as it usually can be done without altering the finite-dimensional distributions). Thus $D_E[0,\infty)$ conventionally denotes the space of *càdlàg* (an acronym for "Continu à droite, limité à gauche") functions $x:\ [0,\infty) \to E$, that is such that for each t

$$x(t) = \lim_{s \to t^+} x(s)$$

and

$$x(t^-) \triangleq \lim_{s \to t^-} x(s).$$

By convention, $x(0) = x(0^-) = \lim_{s \to 0^-} x(s)$.

As it turns out, the functions in $D_E[0,\infty)$ are better behaved than one can expect.

Lemma 3.3.1 *If $x \in D_E[0,\infty)$, then it has at most countably many points of discontinuity.*

Two functions x and y are near one another in the uniform topology used for the space of continuous functions if the graph of $x(t)$ can be carried onto the graph of $y(t)$ by a uniformly small perturbation of the ordinates, with the abscissas kept fixed. In the space D this is not enough, and we have to allow also a uniformly small deformation of the time scale. The Skorohod topology embodies this idea, and is described as follows.

We define a metric on $D_E[0,\infty)$ under which it is a separable metric space if E is, and is complete if (E,r) is complete. Let λ' be the collection of (strictly) increasing functions λ mapping $[0,\infty)$ into $[0,\infty)$ (in particular $\lambda(0)=0$, $\lim_{t\to\infty}\lambda(t)=\infty$ and λ is continuous). Let Λ be the set of Lipschitz continuous functions $\lambda\in\lambda'$ such that

$$\begin{aligned}\gamma(\lambda) &\equiv \operatorname*{ess\,sup}_{t\geq 0} |\log\lambda'(t)| \\ &= \sup_{s>t\geq 0}\left|\log\frac{\lambda(s)-\lambda(t)}{s-t}\right| < \infty.\end{aligned}$$

For $x,y\in D_E[0,\infty)$, define

$$d(x,y)=\inf_{\lambda\in\Lambda}\left[\gamma(\lambda)\vee\int_0^\infty e^{-u}d(x,y,\lambda,u)du\right]$$

where

$$d(x,y,\lambda,u)=\sup_{t\geq 0}\; q(x(t\wedge u),y(\lambda(t)\wedge u)).$$

It follows that, given $\{x_n\},\{y_n\}\subset D_E[0,\infty)$,

$$\lim_{n\to\infty} d(x_n,y_n)=0$$

if and only if there exists $\{\lambda_n\}\subset\Lambda$ such that $\lim_{n\to\infty}\gamma(\lambda_n)=0$ and

$$\lim_{n\to\infty} m\{u\in[0,u_0]:\; d(x_n,y_n,\lambda_n,u)\geq\varepsilon\}=0$$

for every $\varepsilon>0$ and $u_0>0$, where m is Lebesgue measure. Moreover, since

$$\operatorname*{ess\,sup}_{t\geq 0} |\,\lambda'(t)-1|\leq 1-e^{-\gamma(\lambda)}\leq\gamma(\lambda)$$

for every $\lambda \in \Lambda$, $\lim_{n\to\infty} \gamma(\lambda_n) = 0$ implies that

$$\lim_{n\to\infty} \sup_{0\le t\le T} |\,\lambda_n(t) - t| = 0$$

for all $T > 0$.

Then d defines a metric, and the topology induced on $D_E[0,\infty)$ by the metric d is called *Skorohod topology*.

Theorem 3.3.2 **($(D_E[0,\infty), d)$ separable and complete)**

- *If E is separable, then $D_E[0,\infty)$ is separable.*
- *If (E, r) is complete, then $(D_E[0,\infty), d)$ is complete.*

3.4 Convergence in distribution in $D_E[0,\infty)$

We are interested in weak convergence of elements of $\mathscr{P}(D_E[0,\infty))$. Let $\mathscr{S}_E$ denote the Borel σ-algebra of $D_E[0,\infty)$. The next result states that $\mathscr{S}_E$ is just the σ-algebra generated by the coordinate random variables.

Proposition 3.4.1 (Borel σ-algebra of $D_E[0,\infty)$)
For each $t \ge 0$, define $\pi_t : D_E[0,\infty) \to E$ by $\pi_t(x) = x(t)$. Then

$$\mathscr{S}_E \supset \mathscr{S}'_E \equiv \sigma(\pi_t : 0 \le t < \infty) = \sigma(\pi_t : t \in D)$$

where D is any dense subset of $[0,\infty)$. If E is separable, then $\mathscr{S}_E = \mathscr{S}'_E$.

A $D_E[0,\infty)$-valued random variable is a stochastic process with sample paths in $D_E[0,\infty)$, although the converse need not be true if E is not separable. Assuming E is separable, let $\{X_\alpha\}$, where α ranges over some index set, be a family of stochastic processes with sample paths in $D_E[0,\infty)$. Let $\{P_\alpha\} \subset \mathscr{P}(D_E[0,\infty))$ be the family of associated probability distributions, i.e. $P_\alpha(B) = P\{X_\alpha \in B\}$ for all $B \in \mathscr{S}_E$.

Definition 3.4.2 (Relative compactness in $D_E[0,\infty)$)

We say that $\{X_\alpha\}$ is relatively compact if $\{P_\alpha\}$ is, i.e. if the closure of $\{P_\alpha\}$ in $\mathscr{P}(D_E[0,\infty))$ is compact.

The conditions for compactness are usually stated in terms of the following modulus of continuity. For $x \in D_E[0,\infty)$, $\delta > 0$ and $T > 0$, define

$$w'(x,\delta,T) = \inf_{\{t_i\}} \max_i \sup_{s,t \in [t_{i-1},t_i)} r(x(s),x(t))$$

where $\{t_i\}$ ranges over all partitions of the form $0 = t_0 < t_1 < \cdots < t_{n-1} < T \le t_n$, with $\min_{1\le i\le n}(t_i - t_{i-1}) > \delta$ and $n \ge 1$.

Lemma 3.4.3 (Modulus of continuity of $D_E[0,\infty)$)

For each $x \in D_E[0,\infty)$ and $T > 0$, $w'(x,\delta,T)$ is right continuous in δ, and

$$\lim_{\delta\to 0} w'(x,\delta,T) = 0.$$

The next theorem gives a criterion for $\{X_\alpha\}$ to be relatively compact.

Theorem 3.4.4 (Criterion for relative compactness)

Let (E, r) be complete and separable, and let $\{X_\alpha\}$ be a family of process with sample paths in $D_E[0, \infty)$. Then $\{X_\alpha\}$ is relatively compact if and only if the following two conditions hold:

(a) *For every $\eta > 0$ and rational $t \geq 0$, there exists a compact set $\Gamma_{\eta,t} \subset E$ such that*

$$\inf_\alpha P\{X_\alpha(t) \in \Gamma_{\eta,t}\} \geq 1 - \eta.$$

(b) *For every $\eta > 0$ and $T > 0$, there exists $\delta > 0$ such that*

$$\sup_\alpha P\{w'(X_\alpha, \delta, T) \geq \eta\} \leq \eta. \tag{3.4.1}$$

The following theorem states weak convergence in $D_E[0, \infty)$.

Theorem 3.4.5 (Weak convergence in $D_E[0, \infty)$)

Let E be separable, and let X_n, $n = 1, 2, \ldots,$ be processes with sample paths in $D_E[0, \infty)$. If $\{X_n\}$ is relatively compact and there exists a dense set $D \subset [0, \infty)$ such that

$$\big(X_n(t_1), \ldots, X_n(t_k)\big) \Rightarrow \big(X(t_1), \ldots, X(t_k)\big)$$

for every finite set $\{t_1, \ldots, t_k\} \subset D$, then $X_n \Rightarrow X$.

The next theorem gives different conditions for sequences of processes to converge to a Markov process, in particular allowing the limiting process to be determined as a solution of a martingale problem (see Section 1.5).

Let $\{\mathscr{G}_t^n\}$ be a filtration and let $\mathscr{L}_n$ be the space of real-valued $\{\mathscr{G}_t^n\}$-progressive processes ξ satisfying

$$\sup_{t\leq T} E[|\,\xi(t)|] < \infty$$

for each $T > 0$. Let $\hat{\mathscr{A}_n}$ be the collection of pairs $(\xi, \varphi) \in \mathscr{L}_n \times \mathscr{L}_n$ such that

$$\xi(t) - \int_o^t \varphi(s)\mathrm{d}s$$

is a $\{\mathscr{G}_t^n\}$-martingale. Note that if X_n is a $\{\mathscr{G}_t^n\}$-progressive solution of the martingale problem for $A_n \subset B(E) \times B(E)$ with respect to $\{\mathscr{G}_t^n\}$, then $(f \circ X_n, g \circ X_n) \in \hat{\mathscr{A}_n}$ for each $(f, g) \in A_n$.

Theorem 3.4.6 (Conditions for weak convergence)
Let (E, r) be complete and separable. Let $A \subset \bar{C}(E) \times \bar{C}(E)$ and $\nu \in \mathscr{P}(E)$, and suppose that the $D_E[0,\infty)$ martingale problem for (A, ν) has at most one solution. Suppose X_n, $n = 1, 2, \ldots,$ is a $\{\mathscr{G}_t^n\}$-adapted process with sample paths in $D_E[0,\infty)$, $\{X_n\}$ is relatively compact, $PX_n(0)^{-1} \Rightarrow \nu$, and $M \subset \bar{C}(E)$ is separating (cf. Proposition 1.4.14). *Then the following are equivalent:*

(a) *There exists a solution X of the $D_E[0,\infty)$ martingale problem for (A, ν), and $X_n \Rightarrow X$.*

(b) *There exists a countable set $\Gamma \subset [0,\infty)$ such that for each $(f, g) \in A$ and $T > 0$, there exist $(\xi_n, \varphi_n) \in \hat{\mathscr{A}_n}$ such that*

$$\sup_n \sup_{s\leq T} E[|\,\xi_n(s)|] < \infty, \tag{3.4.2}$$

$$\sup_n \sup_{s\leq T} E[|\,\varphi_n(s)|] < \infty, \tag{3.4.3}$$

$$\lim_{n\to\infty} E\left[\big(\xi_n(t) - f(X_n(t))\big)\prod_{i=1}^{k} h_i(X_n(t_i))\right] = 0 \tag{3.4.4}$$

and

$$\lim_{n\to\infty} E\left[\big(\varphi_n(t) - g(X_n(t))\big)\prod_{i=1}^{k} h_i(X_n(t_i))\right] = 0 \tag{3.4.5}$$

for all $k \geq 0$, $0 \leq t_1 < t_2 < \cdots < t_k \leq t \leq T$ *with* $t_i, t \notin \Gamma$, *and* $h_1, \ldots, h_k \in M$.

If a sequence of processes approximates a Markov process, it is reasonable to expect the processes to be approximately Markovian in some sense. One way of expressing it is included in the following remark.

Remark 3.4.7 (Approximate Markovianity)

Note that (3.4.4) and (3.4.5) are implied respectively by

$$\lim_{n\to\infty} E[|\, \xi_n(t) - f(X_n(t))|] = 0 \tag{3.4.6}$$

and

$$\lim_{n\to\infty} E[|\, \varphi_n(t) - g(X_n(t))|] = 0. \tag{3.4.7}$$

If this conditions hold, then

$$\lim_{n\to\infty} E\Big[\big|E\big[f(\mathbf{X}^n(t+s))|\, \mathscr{F}_t^{\mathbf{X}^n}\big] - T(s)f(\mathbf{X}^n(t))\big|\Big] = 0 \tag{3.4.8}$$

where $\{T(s)\}$ is the semigroup corresponding to the limiting process.

Corollary 3.4.8 (Conditions for weak convergence II)

Suppose in Theorem 3.4.6 *that* $X_n = \eta_n \circ Y_n$ *and* $\{\mathscr{G}_t^n\} = \{\mathscr{F}_t^{Y_n}\}$, *where* Y_n *is a progressive Markov process in a metric space* E_n *corresponding to a measurable contraction semigroup* $\{T_n(t)\}$ *with full generator* $\hat{A}_n$, *and* $\eta_n : E_n \to E$ *is Borel measurable. Then condition* (b) *of Theorem* 3.4.6 *is equivalent to the following:*

(c) *There exists a countable set* $\Gamma \subset [0, \infty)$ *such that for every* $(f, g) \in A$ *and* $T > 0$, *there exist* $(f_n, g_n) \in \hat{A}_n$ *such that* $\{(\xi_n, \varphi_n)\} = \{(f_n \circ Y_n, g_n \circ Y_n)\}$ *satisfies* (3.4.2)-(3.4.5) *for all* $k \geq 0$, $0 \leq t_1 < \cdots < t_k \leq t \leq T$, *with* $t_i, t \notin \Gamma$ *and* $h_1, \ldots, h_k \in M$.

Chapter 4

Fleming-Viot models

The Fleming-Viot process, together with the Dawson-Watanabe process, which arises from branching models (see for example Dynkin, 1994; Etheridge, 2000), is among the most studied measure-valued processes in modern pure and applied probability, and concerned, like the latter, with stochastic population models.

The Fleming-Viot process was first introduced in population genetics by Fleming and Viot (1979), as a generalisation of the Wright-Fisher diffusion process with a finite number of types (see for example Ewens, 1979), which Otha and Kimura (1973) had previously extended to the countably infinite case, in the so-called stepwise-mutation model. Successively, Ethier and Kurtz (1981) showed that the infinitely-many-neutral-alleles diffusion model has a unique stationary distribution and is reversible and ergodic; see also Kimura and Crow (1964), Watterson (1976), Ethier and Kurtz (1986, 1987). A variation of the infinitely-many-neutral-alleles model is given by the infitely-many-sites model; see Kimura (1969, 1971), Watterson (1975), Ethier and Griffiths (1987). Ethier and Kurtz (1992) described a Markov chain on E^n

whose empirical measure converges almost surely to a random variable distributed as a Dirichlet process, which Ethier and Kurtz (1994) showed to be the stationary distribution of the neutral Fleming-Viot process. Ethier and Griffiths (1993) derived an explicit form for the transition function of the Fleming-Viot process, providing a result on the rate of convergence to the stationary distribution; see also Walker *et al.* (2007) for a Bayesian nonparametric approach to finding the transition function. Donnelly and Kurtz (1996) introduced a countable representation of the neutral Fleming-Viot process, in terms of the empirical measure of a process of particles, which opens new possibilities in terms of approach to the subject using ancestry analysis and coalescent's theory (Kingman, 1982). Donnelly and Kurtz (1999b) generalised such discrete construction to more general measure-valued population models, including the Dawson-Watanabe processes. Donnelly and Kurtz (1999a) extended the particle construction for Fleming-Viot models to the case where mechanisms of selection and recombination are included.

In this chapter we present a brief summary on Fleming-Viot models and a certain type of countable representation, either of which will constitute a reference point in the second part, where some new processes which lead to Fleming-Viot models will be introduced.

After introducing the basic terminology of population genetics, we will give an essential account on Fleming-Viot processes, emphasizing the characterization of the process in terms of its infinitesimal generator. We will then state the main results in the literature with particular attention to the stationary distributions, in the cases they are known. Finally we will trace the main features of the countable representations introduced by Donnelly and Kurtz (1996, 1999a), in the case of neutrality and in the presence of selection.

4.1 Population genetics

Population genetics is the study of the distribution and change in allele frequencies, or more generally in populations of individuals. Its primary founders were Sewall Wright, J. B. S. Haldane and Ronald Fisher, who also laid the foundations for the related discipline of quantitative genetics.

The forces that drive evolution, which are generally considered to be natural selection, genetic drift, mutation, and migration, are represented in probabilistic models by stochastic transition mechanisms. A brief overview of their meanings will thus be useful to identify the features of different components of the probabilistic models.

Natural selection occurs when individuals differ in reproductive output for functional reasons, that is when differences in reproduction follow from the fact that individuals differ from each other in their ability to tackle the challenges posed by their internal biology and by the biological and physical environment. More practically, natural selection acts either by making some genotypes more likely than others to have offspring, which is known as *fecundity selection*, or by making some genotypes more likely to survive longer than others, which is known as *viability selection*.

Genetic drift is the term used in population genetics to refer to the statistical drift over time of allele frequencies in a finite population due to random sampling effects in the formation of successive generations. In a narrower sense, genetic drift refers to the expected population dynamics of neutral alleles (those defined as having no positive or negative impact on fitness). Whereas natural selection describes the tendency of beneficial alleles to become more common over time (and detrimental ones less common), genetic drift refers to the fundamental tendency of any allele to vary randomly in frequency over time due to statistical variation alone, so long as it does not comprise all or none of the distribution.

Genetic drift is opposed in this regard by genetic *mutation*, which introduces novel variants into the population according to its own random processes. Mutations are considered the driving force of evolution. *Neutral* mutations are defined as mutations whose effects do not influence the fitness of either the species or the individuals who make up the species. That is, the differences within individuals do not influence the fitness.

In the literature there are also models with *migration* and *recombination*, the former occurring when living things move from one biome to another, biome denoting a major regional group, and the latter occurring when two randomly selected individuals from the population generate offspring.

In Fleming-Viot models, introduced in the next section, necessary simplification of reality brings to consider genetic drift and mutation as the two basic components of a population genetics model. Fertility and viability selection, among others, are instead considered additional features which can help describing more complex mechanisms, but which do not substitute the intrinsic forces of evolution given by the formers.

4.2 The Fleming-Viot process

4.2.1 Generalities

The neutral K-type diffusion model, also known as the Wright-Fisher process (see for instance Ewens, 1979), is one of the most widely used discrete stochastic models in population genetics, described as follows. The process describes a population of individuals which can be of K different types, and takes values in the $(K-1)$-dimensional simplex

$$\Delta_K = \{p = (p_1, \ldots, p_K) : p_i \geq 0, \, i = 1, \ldots, K, \sum_{i=1}^{K} p_i = 1\} \tag{4.2.1}$$

where p_i denotes the proportion of the population that is of type i. The process is characterized in terms of the generator

$$\mathscr{A} = \frac{1}{2} \sum_{i,j=1}^{K} p_i(\delta_{ij} - p_j) \frac{\partial^2}{\partial p_i \partial p_j} + \sum_{j=1}^{K} \left(\sum_{i=1}^{K} q_{ij} p_i \right) \frac{\partial}{\partial p_j} \tag{4.2.2}$$

where $(q_{ij})_{1 \leq i,j \leq k}$, the infinitesimal matrix for a Markov process in the set of types, describes the mutation structure. Here q_{ij} is the intensity of a mutation from type i to type j and $q_{ii} = -\sum_{j:j \neq i} q_{ij}$. The domain of $\mathscr{A}$ is

$$\mathscr{D}(\mathscr{A}) = \{F|_{\Delta_K} : F \in C^2(\mathbb{R}^K)\}$$

where C^2 is the set of twice-differentiable continuous functions.

Except for some technical requirements on (q_{ij}), the same description is valid when the set of possible types is countably infinite. One such example is given in Otha and Kimura (1973), namely the stepwise-mutation model, in which the set of types is $\mathbb{Z}$

and

$$q_{ij} = \begin{cases} \frac{1}{2}\theta & \text{if } j = i \pm 1 \\ -\theta & \text{if } j = i \\ 0 & \text{otherwise} \end{cases} \tag{4.2.3}$$

for some $\theta > 0$.

Fleming and Viot (1979) generalised this approach, covering the case of an uncountably infinite set of types. More precisely, they replaced $\{1, \dots, K\}$ by a compact metric space E, called type space, Δ_K by $\mathscr{P}(E)$, the set of Borel probability measures on E with the topology of weak convergence, and $\mathscr{A}$ by

$$\begin{aligned} \mathbb{A}\varphi(\mu) =& \frac{1}{2} \sum_{1 \leq j \neq i \leq m} \Big(\langle f_i f_j, \mu \rangle - \langle f_i, \mu \rangle \langle f_j, \mu \rangle \Big) F_{z_i z_j} \Big(\langle f_1, \mu \rangle, \dots, \langle f_m, \mu \rangle \Big) \\ &+ \frac{1}{2}\theta \sum_{i=1}^{m} \Big(\langle P f_i, \mu \rangle - \langle f_i, \mu \rangle \Big) F_{z_i} \Big(\langle f_1, \mu \rangle, \dots, \langle f_m, \mu \rangle \Big). \end{aligned} \tag{4.2.4}$$

with domain

$$\begin{aligned} \mathscr{D}(\mathbb{A}) = \Big\{ & \varphi \in C(\mathscr{P}(E)) : \varphi(\mu) \equiv F(\langle f_1, \mu \rangle, \dots, \langle f_m, \mu \rangle), \\ & F \in C^2(\mathbb{R}^m), f_1 \dots, f_m \in C(E), m \in \mathbb{N} \Big\} \end{aligned} \tag{4.2.5}$$

where $\langle f, \mu \rangle = \int_E f \mathrm{d}\mu$ and F_{z_i} denotes the partial derivative of F with respect to z_i, with $z_i = \langle f_i, \mu \rangle$. Here $\theta > 0$ is twice the mutation intensity and $P : C(E) \mapsto C(E)$ is given by

$$Pf(x) = \int_E f(y) P(x, \mathrm{d}y) \tag{4.2.6}$$

where $P(x, \mathrm{d}y)$ is the distribution of the type of a mutant offspring of a parent of type x.

More generally we can write the generator $\mathbb{A}$ as

$$\begin{aligned}\mathbb{A}\varphi(\mu) =&\frac{1}{2}\int_E\int_E \mu(\mathrm{d}x)\big(\delta_x(\mathrm{d}y)-\mu(\mathrm{d}y)\big)\frac{\partial^2\varphi(\mu)}{\partial\mu(x)\partial\mu(y)}\\ &+\int_E \mu(\mathrm{d}x)B\left(\frac{\partial\varphi(\mu)}{\partial\mu(\cdot)}\right)(x)\end{aligned} \tag{4.2.7}$$

where

$$\frac{\partial\varphi(\mu)}{\partial\mu(x)} = \lim_{\epsilon\to 0^+}\frac{\varphi(\mu+\epsilon\delta_x)-\varphi(\mu)}{\epsilon} \tag{4.2.8}$$

and we take $\mathscr{D}(\mathbb{A})$ to be

$$\begin{aligned}\mathscr{D}(\mathbb{A}) = \Big\{&\varphi\in B(\mathscr{P}(E)) : \varphi(\mu)\equiv F(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle),\\ &F\in C^2(\mathbb{R}^m), f_1\ldots,f_m\in\mathscr{D}(B), m\in\mathbb{N}\Big\}.\end{aligned} \tag{4.2.9}$$

Here B is the generator of a Feller semigroup on $C(E)$ (cf. Section 1.4) and $\delta_x\in\mathscr{P}(E)$ denotes the unit mass at $x\in E$. The formulation (4.2.4) is from Fleming and Viot (1979), whereas (4.2.7) is due to Dawson and Hochberg (1982).

The resulting probability-measure-valued diffusion process is referred to as a *Fleming-Viot process*, or alternatively *neutral diffusion model* or also *infinitely-many-neutral-alleles model*. In particular, when we refer to (4.2.4), (4.2.5) and (4.2.6), we say that the Fleming-Viot process has parameters E, θ and P.

B is known as the *mutation operator*, so that the second term in (4.2.7) is responsible for mutation, while the first describes genetic drift. If the chances of an individual mutating, replicating or dying do not depend on its type, then the process is said to be *neutral*.

A more general class of Fleming-Viot processes can be described in terms of generators of the type

$$\begin{aligned}\mathbb{A}\varphi(\mu) =&\frac{1}{2}\int_E\int_E \mu(\mathrm{d}x)(\delta_x(\mathrm{d}y)-\mu(\mathrm{d}y))\frac{\partial^2\varphi(\mu)}{\partial\mu(x)\partial\mu(y)}\\ &+\int_E \mu(\mathrm{d}x)B\left(\frac{\partial\varphi(\mu)}{\partial\mu(\cdot)}\right)(x)\\ &+\int_E\int_E \mu(\mathrm{d}x)\mu(\mathrm{d}y)R\left(\frac{\partial\varphi(\mu)}{\partial\mu(\cdot)}\right)(x,y)\\ &+\int_E\int_E \mu(\mathrm{d}x)\mu(\mathrm{d}y)(\sigma(x,y)-\langle\sigma,\mu^2\rangle)\frac{\partial\varphi(\mu)}{\partial\mu(\cdot)}\end{aligned} \tag{4.2.10}$$

where

$$Rg(x,y) = \varrho\int_E \big[g(z)-g(x)\big]H\big((x,y),\mathrm{d}z\big) \tag{4.2.11}$$

is the *recombination operator*, with recombination intensity $\varrho \geq 0$, $H\big((x,y),\mathrm{d}z\big)$ is a one step transition function on $E^2\times\mathscr{B}(E)$ and $\sigma\in B_{\mathrm{sym}}(E^2)$ is the bivariate symmetric *selection intensity function*. The domain of $\mathbb{A}$ is defined as above.

When

$$\varphi(\mu) = F(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle) = F(\langle\mathbf{f},\mu\rangle) \tag{4.2.12}$$

then $\mathbb{A}$ reduces to

$$\begin{aligned}\mathbb{A}\varphi(\mu) =&\frac{1}{2}\sum_{1\leq k\neq i\leq m}(\langle f_if_j,\mu\rangle-\langle f_i,\mu\rangle\langle f_j,\mu\rangle)\,F_{z_iz_j}(\langle\mathbf{f},\mu\rangle)\\ &+\sum_{i=1}^m\big(\langle Bf_i,\mu\rangle+\langle Rf_i,\mu^2\rangle\big)\,F_{z_i}(\langle\mathbf{f},\mu\rangle)\\ &+\sum_{i=1}^m\big(\langle(f_i\circ\pi)\,\sigma,\mu^2\rangle-\langle f_i,\mu\rangle\langle\sigma,\mu^2\rangle\big)\,F_{z_i}(\langle\mathbf{f},\mu\rangle)\end{aligned} \tag{4.2.13}$$

where π is the projection of E^2 onto its first coordinate.

Another choice for the domain of $\mathbb{A}$ that is often useful, and that we will actually use in the second part, is the set of all $\varphi \in B(\mathscr{P}(E))$ of the form

$$\varphi(\mu) = \langle f, \mu^m \rangle \tag{4.2.14}$$

where $m \in \mathbb{N}$, μ^m denotes the m-fold product measure and $f \in B(E^k)$. For a locally compact space E consider the Feller semigroup generated by the closure of B on $\hat{C}(E)$, the space of real continuous functions vanishing at infinity (if E is compact[1], then $\hat{C}(E) = C(E)$), that is

$$T(t)f(x) = \int_E f(\xi)P(t, x, \mathrm{d}\xi).$$

For each $m \geq 1$, define the semigroup on $B(E^m)$ by

$$T_m(t)f(x_1, \dots, x_m) = \int_E \dots \int_E f(\xi_1, \dots, \xi_m)P(t, x_1, \mathrm{d}\xi_1)\dots P(t, x_m, \mathrm{d}\xi_m)$$

and denote with $B^{(m)}$ its generator. We will write

$$B^{(m)}f(x_1, \dots, x_m) = \sum_{i=1}^{m} B_i f(x_1, \dots, x_m) \tag{4.2.15}$$

where B_i is $B^{(m)}$ applied to f as a function of the ith variable alone, when the right side is defined. In addition, for each $m \geq 2$ and $1 \leq i < j \leq m$ define $\Phi_{ij} : B(E^m) \to B(E^{m-1})$ by

$$\Phi_{ij} f(x_1, \dots, x_m) = f(x_1, \dots, x_{j-1}, x_i, x_{j+1}, \dots, x_m).$$

For each $m \geq 1$ and $1 \leq i \leq m$, define $M_i : B(E^m) \to B(E^{m+1})$ by

$$M_i f(x_1, \dots, x_{m+1}) = \int_E f(x_1, \dots, x_{i-1}, \xi, x_{i+1}, \dots, x_m) H\big((x_i, x_{m+1}), \mathrm{d}\xi\big)$$

[1]Note that the locally compact case can be reduced to the compact case by a one-point compactification.

and $K_i : B(E^m) \to B(E^{m+2})$ by

$$K_i f(x_1, \ldots, x_{m+2}) = \frac{\bar{\sigma} + \sigma(x_i, x_{m+1}) - \sigma(x_{m+1}, x_{m+2})}{2\bar{\sigma}} f(x_1, \ldots, x_m) \qquad (4.2.16)$$

where $\bar{\sigma} = sup_{x,y,z \in E} |\, \sigma(x,y) - \sigma(y,z)|$ and we assume $0/0 = 0$. Then when (4.2.14) holds, (4.2.10) reduces to

$$\begin{aligned} \mathbb{A}\varphi(\mu) = & \sum_{1 \le i < j \le m} \big(\langle \Phi_{ij} f, \mu^{m-1} \rangle - \langle f, \mu^m \rangle \big) \\ & + \langle B^{(m)} f, \mu^m \rangle + \varrho \sum_{i=1}^{m} \big(\langle M_i f, \mu^{m+1} \rangle - \langle f, \mu^m \rangle \big) \\ & + 2\bar{\sigma} \sum_{i=1}^{m} \big(\langle K_i f, \mu^{m+2} \rangle - \langle f, \mu^m \rangle \big) + \bar{\sigma} k \langle f, \mu^m \rangle \end{aligned} \qquad (4.2.17)$$

Note that since $\langle \Phi_{ki} f, \mu^{m-1} \rangle = \langle \Phi_{ik} f, \mu^{m-1} \rangle$ implies

$$\frac{1}{2} \sum_{1 \le k \ne i \le n} (\langle \Phi_{ki} f, \mu^{m-1} \rangle - \langle f, \mu^m \rangle) = \sum_{1 \le k < i \le n} (\langle \Phi_{ki} f, \mu^{m-1} \rangle - \langle f, \mu^m \rangle), \qquad (4.2.18)$$

using also (4.2.15) and (4.2.16) the generator (4.2.17) can be written (cf. Ethier and Kurtz, 1994)

$$\begin{aligned} \mathbb{A}\varphi(\mu) = & \frac{1}{2} \sum_{1 \le i \ne j \le m} \big(\langle \Phi_{ij} f, \mu^{m-1} \rangle - \langle f, \mu^m \rangle \big) \\ & + \sum_{i=1}^{m} \langle B f_i, \mu^m \rangle + \varrho \sum_{i=1}^{m} \big(\langle M_i f, \mu^{m+1} \rangle - \langle f, \mu^m \rangle \big) \\ & + \sum_{i=1}^{m} \big(\langle \sigma_{i,m+1}(\cdot,\cdot) f, \mu^{m+1} \rangle - \langle \sigma_{m+1,m+2}(\cdot,\cdot) f, \mu^{m+2} \rangle \big) . \end{aligned} \qquad (4.2.19)$$

where $\sigma_{i,j}(\cdot,\cdot)$ denotes $\sigma(x_i, x_j)$. Moreover, if we have no recombination, i.e. $\varrho = 0$, and the selection function is haploid instead of diploid, i.e. we have (cf. Donnelly and Kurtz, 1999a)

$$\sigma(x, \mu) = \int_E \sigma(x, y) \mu(\mathrm{d}y), \qquad (4.2.20)$$

then in place of (4.2.19) we have

$$\mathbb{A}\varphi(\mu) = \sum_{i=1}^{m} \langle Bf_i, \mu^m \rangle + \frac{1}{2} \sum_{1 \le i \ne j \le m} \big(\langle \Phi_{ij} f, \mu^{m-1} \rangle - \langle f, \mu^m \rangle \big)$$
$$+ \sum_{i=1}^{m} \big(\langle \sigma_i(\cdot, \mu) f, \mu^m \rangle - \langle \sigma(\cdot, \mu) \otimes f, \mu^{m+1} \rangle \big) \qquad (4.2.21)$$

where $\sigma(\cdot) \otimes f(x_1, \ldots, x_m)$ denotes $\sigma(x_{m+1}) f(x_1, \ldots, x_m)$. Observe that σ in (4.2.21) describes fertility selection (cf. Section 4.3.2 below for fertility vs. viability selection).

From (4.2.21) the neutral diffusion model is recovered when $\sigma \equiv 0$ and

$$Bg(x) = \frac{1}{2} \theta \int_E \big[g(y) - g(x) \big] P(x, \mathrm{d}y) \qquad (4.2.22)$$

(cf. Ethier and Kurtz, 1993; Donnelly and Kurtz, 1996).

The next result is essentially from Ethier and Kurtz (1987), except part (a) which is due to Fleming and Viot (1979).

Theorem 4.2.1 (Martingale problem for Fleming-Viot process) *Let E be locally compact and define B, R and σ as above. Suppose that the closure of B generates a Feller semigroup on $\hat{C}(E)$. Then*

(a) *uniqueness of solutions of the martingale problem holds for $\mathbb{A}$ defined by* (4.2.4) *and* (4.2.5), *with specified initial distribution;*

(b) *uniqueness of solutions of the martingale problem holds for $\mathbb{A}$ defined by* (4.2.12) *and* (4.2.13), *with specified initial distribution;*

(c) *the martingale problems for $\mathbb{A}$ defined by* (4.2.12) *and* (4.2.13) *and by* (4.2.14) *and* (4.2.17) *are equivalent;*

(d) *suppose that* $R \equiv 0$. *Then the* $C_{\mathscr{P}(E)}[0,\infty)$ *martingale problem for* $\mathbb{A}$ *is well posed.*

4.2.2 Further properties

In this section we state the basic properties of the Fleming-Viot process.

The following result, which states that the Fleming-Viot process lives in $\mathscr{P}_a(E)$, the set of purely atomic Borel probability measures on E, is due to Ethier and Kurtz (1986, 1987).

Theorem 4.2.2 (Atomicity of Fleming-Viot process)
Let $\{\mu_t, t \geq 0\}$ *be a Fleming-Viot process with type space* E, *generator given by* (4.2.14) *and* (4.2.17), *mutation operator*

$$Bg(x) = \frac{1}{2}\,\theta(x) \int_E \big[g(y) - g(x)\big]\,P(x, dy) \tag{4.2.23}$$

no recombination and selection intensity function given by $\sigma \in B_{\mathrm{sym}}(E^2)$. *Then*

$$P\{\mu_t \in \mathscr{P}_a(E) \text{ for all } t > 0\} = 1.$$

If Π *is a stationary distribution for such a Fleming-Viot process, then* $\Pi(\mathscr{P}_a(E)) = 1$.

If we assume that the space E is $\{1, \ldots, K\}$ and the support of ν_0 is all of E, and if we identify $\mathscr{P}(E)$ with Δ_K as in (4.2.1), then the resulting Fleming-Viot process, as already seen, has generator given by (4.2.2). In the special case of parent independent mutation, that is

$$q_{ij} = \frac{1}{2}\theta_j \equiv \frac{1}{2}\theta\nu_0(\{j\}) > 0 \tag{4.2.24}$$

for $i, j \in \{1, \ldots, K\}$, $i \neq j$, Wright (1949) discovered that there is a unique stationary distribution $\pi \in \mathscr{P}(\Delta_K)$ which is absolutely continuous with respect to the $(K-1)$-Lebesgue measure on Δ_K, with density

$$\pi(\mathrm{d}p) = \frac{\Gamma(\theta_1 + \cdots + \theta_K)}{\Gamma(\theta_1)\ldots\Gamma(\theta_K)} p_1^{\theta_1 - 1} \ldots p_K^{\theta_K - 1} \mathrm{d}p_1 \ldots \mathrm{d}p_{K-1}, \tag{4.2.25}$$

namely the Dirichlet distribution with parameters $\theta_1, \ldots, \theta_K$. See Ethier and Kurtz (1981) for proof.

An analogous result was established by Shiga (1990) for the Fleming-Viot process with generator (4.2.7), in the case

$$Bf(x) = \frac{1}{2}\theta \int_E [f(y) - f(x)]\nu_0(\mathrm{d}y) \tag{4.2.26}$$

where $\theta > 0$ and $\nu_0 \in \mathscr{P}(E)$. Namely, there is a unique stationary distribution $\Pi_{\theta,\nu_0} \in \mathscr{P}(\mathscr{P}(E))$, given by the distribution of the $\mathscr{P}(E)$-valued random variable μ characterized by the property that whenever $\Lambda_1, \ldots, \Lambda_K$, for $K \geq 2$, is a partition of E into Borel sets, then $(\mu(\Lambda_1), \ldots, \mu(\Lambda_K))$ has Dirichlet distribution with parameters $\theta\nu_0(\Lambda_1), \ldots, \theta\nu_0(\Lambda_K)$. Ethier and Kurtz (1994) showed that

$$\Pi_{\theta,\nu_0}(\mathrm{d}\mu) = P\left\{\sum_{i=1}^{\infty} \rho_i \delta_{\xi_i} \in \mathrm{d}\mu\right\} \tag{4.2.27}$$

where $(\rho_1, \rho_2, \ldots)$ have the Poisson-Dirichlet distribution with parameter θ (Kingman, 1975; cf. Definition 2.2.7) and $\xi_1, \xi_2, \ldots$ are i.i.d. from ν_0, and independent of $(\rho_1, \rho_2, \ldots)$. That is, when (4.2.26) holds the stationary distribution of the neutral diffusion model is the Dirichlet process with parameters (θ, ν_0) (cf. Definition 2.2.8). Of course $(\rho_1, \rho_2, \dots)$ assumes values in

$$\Delta_\infty = \left\{ p = (p_1, p_2, \dots) : p_1 \geq p_2 \geq \dots \geq 0, \sum_{i=1}^{\infty} p_i = 1 \right\}, \tag{4.2.28}$$

and in particular, from Theorem 2.2.6, Π_{θ,ν_0} is concentrated on $\mathscr{P}_a(E)$, the set of purely atomic Borel probability measures on E. Note that this is consistent with Theorem 4.2.2.

Shimakura (1977, 1981) and Griffiths (1979) derived an explicit formula for the transition density of the Wright-Fisher diffusion in the finite case assuming (4.2.24). This had previously been done by Malécot (1948), Goldberg (1950) and Crow and Kimura (1956) in the one dimensional case ($K = 2$). Ethier and Griffiths (1993) generalised this result, deriving the transition function (since the transition density does not exist in general) of the Fleming-Viot process with mutation (4.2.26). Define $\mu_n(\mathbf{x}) : E^n \to \mathscr{P}(E)$ to be the empirical measure determined by the (not necessarily distinct) points of the vector $\mathbf{x} = (x_1, \dots, x_n)$, i.e.

$$\mu_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}. \tag{4.2.29}$$

Given $\theta \geq 0$, let $\{D_t, t \geq 0\}$ be a pure death process in $\mathbb{Z}_+ \bigcup \{\infty\}$ starting at infinity, with death rates

$$\lambda_n = \frac{n(\theta + n - 1)}{2} \tag{4.2.30}$$

for $n \geq 0$ (∞ is an entrance boundary), and define

$$d_n^\theta(t) = P\{D_t = n\} \tag{4.2.31}$$

for $n \geq 0$, $t > 0$. It is known (see for example Tavaré, 1984) that

$$d_n^\theta(t) = \begin{cases} 1 - \sum_{m=1}^{\infty} (\theta + 2m - 1)(m!)(-1)^{m-1}\theta_{(m-1)} e^{-\lambda_m t} & \text{if } n = 0 \\ \sum_{m=n}^{\infty} (\theta + 2m - 1)(m!)(-1)^{m-n} \binom{m}{n} (\theta + n)_{(m-1)} e^{-\lambda_m t} & \text{if } n \geq 1 \end{cases}$$

where $a_{(n)} = a(a+1)\ldots(a+n-1)$. Then we have the following.

Theorem 4.2.3 (Transition function of Fleming-Viot process)

The Fleming-Viot process with type space E and mutation operator (4.2.26) *has transition function $P(t, \mu, \mathrm{d}\nu)$ given, for each $t > 0$ and $\mu \in \mathscr{P}(E)$, by*

$$P(t, \mu, \cdot) = d_0^\theta(t)\, \Pi_{\theta,\nu_0}(\cdot) + \sum_{n=1}^{\infty} d_n^\theta(t) \int_{E^n} \mu^n(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_n)\, \Pi_{\theta+n,(\theta+n)^{-1}(\theta\nu_0 + n\mu_n(x_1,\ldots,x_n))}(\cdot) \tag{4.2.32}$$

where $\mu^n \in \mathscr{P}(E^n)$ denotes the n*-fold product measure $\mu \times \cdots \times \mu$.*

In particular, for each $t > 0$ and $\mu \in \mathscr{P}(E)$, $P(t, \mu, \cdot)$ is a mixture of probability distributions of the form (4.2.27). This is again consistent with Theorem 4.2.2.

When $E = \{1, \ldots, K\}$, Theorem 4.2.3 includes the case (4.2.24) and, more generally, the case in which

$$q_{ij} = \frac{1}{2}\theta_j \geq 0 \qquad i, j \in \{1, \ldots, K\}, i \neq j \qquad \theta_1 + \cdots + \theta_K > 0. \tag{4.2.33}$$

In this case Shimakura (1981) derived the transition function and Griffiths (1979) its absolute continuous part.

Shiga (1990) proved a strong ergodic theorem in the setting of Theorem 4.2.3.

Theorem 4.2.4 (Ergodic theorem)

In the setting of Theorem 4.2.3*, for each* $\mu \in \mathscr{P}(E)$

$$\lim_{t\to 0} \|P(t,\mu,\cdot) - \Pi_{\theta,\nu_0}(\cdot)\|_{var} = 0 \tag{4.2.34}$$

where Π_{θ,ν_0} *is as in* (4.2.27) *and* $\|\cdot\|_{var}$ *denotes the total variation norm*

$$\|\nu_1 - \nu_2\|_{var} = \sup_{\Gamma\in\mathscr{B}(E)} |\nu_1(\Gamma) - \nu_2(\Gamma)|\,.$$

So far we have seen the properties of the neutral diffusion model, whose features can thus be considered fairly well explored. The same cannot be said for the more complex formulations, i.e. the non neutral cases. It is actually more correct to say that at the state of the art not much is known. For what concerns the aim of this work, the most relevant results in the literature are the following.

Ethier and Kurtz (1994) showed that the Fleming-Viot process with generator (4.2.19) with no recombination, that is $\varrho = 0$, mutation operator (4.2.26) and bounded symmetric selection intensity function $\sigma \in B_{\text{sym}}(E^2)$, has stationary distribution $\Pi_\sigma \in \mathscr{P}(\mathscr{P}(E))$ defined by

$$\Pi_\sigma(\mathrm{d}\mu) = C\, e^{\langle\sigma,\mu^2\rangle}\, \Pi_{\theta,\nu_0}(\mathrm{d}\mu) \tag{4.2.35}$$

where

$$\langle\sigma,\mu^2\rangle = \iint \sigma(x,y)\mu(\mathrm{d}x)\mu(\mathrm{d}y)$$

and C is a constant of proportionality. Observe that this is the case of fertility selection; cf. (4.2.21). In the haploid case, that is when $\sigma \in B(E)$, the stationary distribution of the corresponding Fleming-Viot process is

$$\Pi_\sigma(\mathrm{d}\mu) = C\, e^{2\langle\sigma,\mu\rangle}\, \Pi_{\theta,\nu_0}(\mathrm{d}\mu) \tag{4.2.36}$$

where

$$\langle \sigma, \mu \rangle = \int \sigma(x)\mu(\mathrm{d}x)$$

See Ethier and Shiga (2000).

Further, Donnelly and Kurtz (1999a) showed, as we will see in the next section, that when the mutation process B is uniformly ergodic in the Fleming-Viot process with selection and recombination, whose generator is (4.2.19), then the stationary distribution is unique and is the de Finetti measure of the individuals of the population.

4.3 Countable representations for Fleming-Viot models

In section 4.2 we have seen that the Fleming-Viot process is a Markov process with sample paths in $\mathscr{P}(E)$, the space of probability measures on E. It turns out that the Fleming-Viot process arises most naturally as the limit in distribution of certain sequences of Markov chains in population genetics, describing the dynamics of the individuals. In this sense, measure-valued diffusions are motivated by first considering a class of prelimiting finite-population models. The dynamics in such discrete contexts are easily specified in terms of the behavior of the individuals in the population, and the composition of the population is naturally represented as a measure on the set of possible types.

In this section we present the fundamental traits of such countable representations, following the works of Donnelly and Kurtz (1996, 1999a), with particular emphasis on the features that will be most useful for comparison with the results introduced in second part.

4.3.1 Neutral models

Donnelly and Kurtz (1996) introduced an E^n-valued process $\{\mathbf{X}^n(t), t \geq 0\}$, where $\mathbf{X}^n = (X_1, \ldots, X_n)$, representing the evolution in time of a population of size n (therefore also called particle process), such that the empirical measure

$$\mu_n(t) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(t)} \tag{4.3.1}$$

in the infinite population limit is a Fleming-Viot process. A process of this type appears implicitly in the work of Dawson and Hochberg (1982).

Let E be compact (if it is locally compact, take its one-point compactification). The generator of the Fleming-Viot process in the n-sized population case corresponds to (4.2.17) when there is no recombination nor selection, that is

$$\mathbb{A}^n \varphi(\mu) = \sum_{1 \leq i < j \leq n} \big(\langle \Phi_{ij} f, \mu^{n-1} \rangle - \langle f, \mu^n \rangle \big) + \langle B^{(n)} f, \mu^n \rangle = \langle A^n f, \mu^n \rangle \tag{4.3.2}$$

for all $f \in \mathscr{D}(B^{(n)})$, where $B^{(n)}$ is as in (4.2.15). The neutral model is recovered when (4.2.26) holds. The generator of the particle process is instead given by

$$A^n f(x_1, \ldots, x_n) = \sum_{1 \leq i < j \leq n} \Big(f\big(\theta_{ij}(x_1, \ldots, x_n)\big) - f(x_1, \ldots, x_n) \Big) + B^{(n)} f(x_1, \ldots, x_n) \tag{4.3.3}$$

where $\theta_{ij}(x_1, \ldots, x_n)$ is the element of E^n obtained from $(x_1, \ldots, x_n)$ by replacing the jth component with the ith. In particular, the jth component of the vector, referred to also as the particle at level j, evolves according to the mutation process until it "looks down" to level i, for some $i < j$, and changes its value to the value of the ith coordinate.

Given that, for $f \in B(E^m)$, $m < n$,

$$\sup_{\mathbf{x} \in E^n} |\Gamma f(\mathbf{x}) - \varphi(\mu_n)| \to 0 \tag{4.3.4}$$

where

$$\Gamma f(\mathbf{x}) = \frac{1}{n^m} \sum_{1 \le i_1 \ne \ldots \ne i_m \le n} f(x_{i_1}, \ldots, x_{i_m}) \tag{4.3.5}$$

and $\varphi(\mu) = \langle f, \mu^m \rangle$, it turns out that, as $n \to \infty$,

$$\sup_{\mathbf{x} \in E^n} |\Gamma A^n f(\mathbf{x}) - \mathbb{A}\varphi(\mu_n)| \to 0 \tag{4.3.6}$$

where $\mathbb{A}$ is (4.2.21) without selection. Under appropriate conditions on the convergence of the mutation operator and the assumption that $\mu_n(0)$ converges in distribution, (4.3.6) implies that $\{\mu_n(t), t \ge 0\}$ converges in distribution in the Skorohod space $D_{\mathscr{P}(E)}[0, \infty)$ to the Fleming-Viot process with generator $\mathbb{A}$ (see Donnelly and Kurtz, 1996). Heuristically, this is due to the fact that a statement about the generator of a process is ultimately concerned with the finite-dimensional distributions (see Section 1.4), and since a càdlàg function has at most a countable number of discontinuities, (4.3.6) holds for almost all t.

Denote now with $\{T^n(t)\}$ the Feller semigroup defined on $C(E^\infty)$ corresponding to A^n. Then, viewing $C(E^n)$ as a closed subspace of $C(E^\infty)$, the following identity holds

$$E_\mu\Big[\langle f, \mu_t^n \rangle\Big] = \langle T^n(t) f, \mu^n \rangle = E_{\mu^\infty}^{\mathbf{X}}\Big[f\big(X_1(t), \ldots, X_n(t)\big)\Big] \tag{4.3.7}$$

for all $f \in C(E^n)$ and $t \ge 0$, where μ_t denotes the Fleming-Viot process and

$$E_\mu\big[g(\mu_t^n)\big] = E\big[g(\mu_t^n) |\, \mu_0^n = \mu\big]$$

that is E_μ denotes the expectation for the Fleming-Viot process under the assumption that the initial state is μ, and

$$E^X_{\mu^\infty}\big[f(\mathbf{X}^n(t))\big] = E\big[f(\mathbf{X}^n(t))|\ \mathbf{X}(0) \sim \mu^\infty\big].$$

that is $E^{\mathbf{X}}_{\mu^\infty}$ denotes the expectation for the particle system under the assumption that $X_1(0), X_2(0), \ldots$ are i.i.d. with common distribution μ. Let $\nu \in \mathscr{P}(\mathscr{P}(E))$. By (4.3.7)

$$\int_{\mathscr{P}(E)} E_\mu[\langle f, \mu^n_t\rangle]\nu(\mathrm{d}\mu) = \int_{\mathscr{P}(E)} E^{\mathbf{X}}_{\mu^\infty}[f(X_1(t), \ldots, X_n(t))]\nu(\mathrm{d}\mu) \tag{4.3.8}$$

for all $f \in B(E^n)$ and $t \geq 0$. The left-hand side is just the expectation for a Fleming-Viot process with initial distribution ν, and the right-hand side is the expectation for the particle system under the assumption that $(X_1(0), X_2(0), \ldots)$ is an exchangeable sequence with

$$P\{X_1(0) \in \Gamma_1, \ldots, X_n(0) \in \Gamma_n\} = \int_{\mathscr{P}(E)} \prod_{i=1}^n \mu(\Gamma_i)\nu(\mathrm{d}\mu). \tag{4.3.9}$$

The identity (4.3.8) implies that

$$P\{X_1(t) \in \Gamma_1, \ldots, X_n(t) \in \Gamma_n\} = \int_{\mathscr{P}(E)} \prod_{i=1}^n \mu(\Gamma_i)\nu_t(\mathrm{d}\mu) \tag{4.3.10}$$

for all $t \geq 0$, where ν_t is the distribution at time t of the Fleming-Viot process with initial distribution ν. Consequently, we see that if $(X_1(0), X_2(0), \ldots)$ is an exchangeable sequence, then $(X_1(t), X_2(t), \ldots)$ is an exchangeable sequence, and the limit of the empirical measure

$$\mu_\infty(t) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \delta_{X_i(t)} \tag{4.3.11}$$

has the same distribution as μ_t. In fact, the process $\{\mu_n(t), t \geq 0\}$ is a version of the Fleming-Viot process, with type space E and mutation operator B. The proof, given in Donnelly and Kurtz (1996), is based on a coupling argument. In particular it is shown that the E^n-valued process $\mathbf{X}^n = (X_1, \dots, X_n)$ can be coupled to an E^n-valued process $\{\mathbf{Y}^n(t), t \geq 0\}$ with generator

$$\tilde{A}^n f(x_1, \dots, x_n) = \frac{1}{2} \sum_{1 \leq i \neq j \leq n} \Big(f\big(\theta_{ij}(x_1, \dots, x_n)\big) - f(x_1, \dots, x_n) \Big) + B^{(n)} f(x_1, \dots, x_n)$$

in such a way that

$$\eta_n(t) \equiv \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i(t)} = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(t)} \tag{4.3.12}$$

for all $t \geq 0$. The process $\{\mathbf{Y}^n(t), t \geq 0\}$ can be thought of as a variant of a Moran model[2] (Moran, 1958a,b,c) in which mutation takes place continuously rather than only at birth/death events. The symmetry of $\{\mathbf{Y}^n(t), t \geq 0\}$ implies the exchangeability of $\{\mathbf{X}^n(t), t \geq 0\}$. The sequence of processes of empirical measures $\{\eta_n(t), t \geq 0\}$ converges to the neutral diffusion model.

An explicit construction of the particle system can be given in terms of a collection of unit Poisson processes $\{N_{ij}, 1 \leq i < j < \infty\}$ where N_{ij} determines the times at which the jth level looks down at the ith level, and a collection of independent random mappings $U_{jk} : E \times [0, \infty) \to E$ that determines the values at the jth level between the kth and $(k+1)$th look-downs from the jth level. That is, for $x \in E$ and $j, k \geq 0$, $U_{jk}(x, \cdot)$ is a version of the mutation process starting from x, $X_1(t) = U_{10}(X_1(0), t)$ for all $t \geq 0$, $X_j(t) = U_{j0}(X_j(0), t)$ until the first look-down from level j, and if the kth look-down from the jth level occurs at time τ and the look-down is to level i, then up until the time of the $(k+1)$th look-down form level j, $X_j(t) = U_{jk}(X_i(\tau), t - \tau)$.

[2] Moran introduced overlapping generations.

It is worth observing that this particle construction not only carries the Fleming-Viot process, but contains within it the familiar genealogical processes, and in particular Kingman's coalescent (Kingman, 1982), thus providing a unification of two previously separate approaches to infinite population genetics models.

4.3.2 Models with selection

Donnelly and Kurtz (1999a) extend this representation to Fleming-Viot models which include recombination and selection. The neutral construction given in Donnelly and Kurtz (1996), i.e. (4.3.2) and (4.3.3), is augmented with two types of additional events on each level. The first type results in potential discontinuities on a level due to additional births (*fertility selection*) or deaths (*viability selection*), which depend on the type on the level and the current value of the empirical measure. The second type of additional event (which will not be treated here) is induced by *recombinations*, and results in discontinuities in which the type on a level is replaced by a random choice according to a Markov kernel, which depends on the current type, and a random choice from the empirical measure. In both cases of selection, the Fleming-Viot process arises when all differences between the fitnesses of distinct individuals are of the order of the inverse of the population size. The key to the discrete representation of the measure-valued diffusion is a particular labelling of the individuals that will constitute a variation of a Moran model. With this labelling, the dynamics of the individuals still make sense as the population size goes to infinity, thus enabling to study the genealogical structure of the population in the limit as well.

Such an E^{∞}-valued particle process is then characterized as the unique solution of an infinite system of ordinary stochastic differential equations and as the unique solution of an associated martingale problem. Further, provided $\mathbf{X}^n(0)$ is exchangeable, solutions to either the system of stochastic differential equations or to the martingale problem are exchangeable for every t. Exchangeability yields a one-to-one correspondence between stationary distributions for X and those for the measure-valued process, the latter being the de Finetti measure for the former.

Consider a collection of n particles, each of which has a *type* represented as a point in a complete, separable, metric space E, that is $\mathbf{X}^n(t) = (X_1^n(t), \dots, X_n^n(t)) \in E^n$. Between birth and death events, the type of the particles evolve independently according to a Markov process with generator B, which is the *mutation process*. Given the observed values $\mathbf{x}^n = (x_1, \dots, x_n) \in E^n$ (when there is no ambiguity on the population size, $\mathbf{x}^n$ will be denoted by $\mathbf{x}$), let $\mu_n = n^{-1} \sum_{i=1}^n \delta_{x_i}$ and let β be a nonnegative, bounded, measurable function defined on $E \times \mathscr{P}(E)$. Then

$$A_0^n f(\mathbf{x}) = \sum_{i=1}^n B_i f(\mathbf{x}) + \frac{1}{2} \sum_{1 \le i \ne j \le n} \left(1 + \frac{2}{n}\beta(x_j, \mu_n)\right) (f(\eta_j(\mathbf{x}|\, x_i)) - f(\mathbf{x})) \quad (4.3.13)$$

is the generator for a Moran particle model[3] with *viability selection*. For $\mathbf{x} \in E^n$ and $z \in E$, $\eta_j(\mathbf{x}|\, z)$ is the element of E^n obtained from $\mathbf{x}$ by replacing x_j by z. The function β governs selection, in the sense that a larger β corresponds to a higher probability for its argument of being selected to be replaced with another particle. Further, B is assumed to be a bounded operator, that is for $g \in B(E)$

$$Bg(z) = \lambda(z) \int_E (g(y) - g(z)) q(z, \mathrm{d}y) \quad (4.3.14)$$

where λ is a nonnegative, bounded, measurable function on E, q is a transition function on E and B_i is just B operating on f as a function of x_i. If B is bounded,

[3]Cf. footnote at page 87.

then A_0^n is also a bounded operator, and existence and uniqueness of solutions of the martingale problems for A_0^n are immediate (see Section 1.5). Consider now a second generator, given by

$$\begin{aligned} A^n f(\mathbf{x}) = & \sum_{i=1}^{n} B_i f(\mathbf{x}) + \sum_{1\le i<j\le n} (f(\theta_j(\mathbf{x}|\, x_i)) - f(\mathbf{x})) \\ & + \sum_{1\le i\ne j\le n} \left(\frac{1}{n}\beta(x_j,\mu_n)\right)(f(\eta_j(\mathbf{x}|\, x_i)) - f(\mathbf{x})) \end{aligned} \tag{4.3.15}$$

where $\theta_j(\mathbf{x}|\, x_i) = (x_1,\dots,x_{j-1},x_i,x_j,\dots,x_{n-1})$, that is a copy of the particle at level i is inserted at level j and the top level particle is killed.

For $m \le n$, let $\mu^{(m)}$ be the probability measure on E^m defined by

$$\mu^{(m)} = \frac{1}{n(n-1)\dots(n-m+1)} \sum \delta_{(x_{i_1},\dots,x_{i_m})} \tag{4.3.16}$$

where the sum is taken over all choices of $1 \le i_1,\dots,i_m \le n$ with $i_k \ne i_l$. Denoting with $\mathscr{P}^n(E) \subset \mathscr{P}(E)$ the set of purely atomic probability measures on E with atoms multiples of $1/n$, for $f \in B(E^n)$ define $\phi \in B(\mathscr{P}^n(E))$ by

$$\phi(\mu) = \langle f, \mu^{(n)}\rangle \tag{4.3.17}$$

and $\mathbb{A}^n\phi$ by

$$\mathbb{A}^n\phi(\mu) = \langle A^n f, \mu^{(n)}\rangle. \tag{4.3.18}$$

Then the following relationship between A_0^n and A^n holds:

$$\mathbb{A}^n\phi(\mu) = \langle A_0^n f, \mu^{(n)}\rangle = \langle A^n f, \mu^{(n)}\rangle. \tag{4.3.19}$$

In particular, denoting with $\tilde{\mathbf{X}}$ and $\mathbf{X}$ the solutions of the martingale problems respectively for A_0^n and A^n, if $\tilde{\mathbf{X}}(0)$ and $\mathbf{X}(0)$ have the same exchangeable distribution then, for each $t \geq 0$, $\tilde{\mathbf{X}}(t)$ and $\mathbf{X}(t)$ have the same exchangeable distribution and the respective processes of empirical measures have the same distribution on $D_{\mathscr{P}(E)}[0,\infty)$.

Viewing A^n as an operator on functions in $B(E^\infty)$, where components with indices greater than n do not vary, if $f \in B(E^m)$ then as n grows to infinity the generator (4.3.15) converges to

$$\begin{aligned} Af(\mathbf{x}) = & \sum_{i=1}^m B_i f(\mathbf{x}) + \sum_{1\leq i<j\leq m} (f(\theta_j(\mathbf{x}|\, x_i)) - f(\mathbf{x})) \\ & + \sum_{j=1}^m \int_E \beta(x_j,\mu)\,(f(\eta_j(\mathbf{x}|\, y)) - f(\mathbf{x}))\,\mu(\mathrm{d}y). \end{aligned} \tag{4.3.20}$$

Observing that for large n and $\mu \in \mathscr{P}^n(E)$, $\mu^{(n)}$ is essentially product measure (cf. (7.2.10) below), for $\varphi(\mu) = \langle f, \mu^m\rangle$ the limit of $\mathbb{A}^n$ in (4.3.19) is

$$\begin{aligned} \mathbb{A}\varphi(\mu) = & \sum_{i=1}^m \langle B_i f, \mu^m\rangle + \frac{1}{2}\sum_{1\leq i\neq j\leq m} (\langle \Phi_{ij} f - f, \mu^m\rangle) \\ & + \sum_{j=1}^m \left(\langle \beta(\cdot,\mu)\otimes f, \mu^{m+1}\rangle - \langle \beta_j(\cdot,\mu) f, \mu^m\rangle\right). \end{aligned} \tag{4.3.21}$$

The meaning of this convergence is that given $X^n(0)$ exchangeable and $X^n(0) \Rightarrow X(0)$, the sequence of processes (X^n, Z_n), with $Z_n(t) = n^{-1}\sum_{i=1}^n \delta_{X_i^n(t)}$, is relatively compact in $D_{E^\infty\times\mathscr{P}(E)}[0,\infty)$ in the sense of convergence in distribution in the Skorohod topology, and assuming f, $B_i f$ and β are all continuous, any limit point (X, Z) will have the property that

$$f(X(t)) - \int_0^t Af(X(s), Z(s))\mathrm{d}s \tag{4.3.22}$$

is an $\mathscr{F}_t^{X,Z}$-martingale, and

$$\phi(Z(t)) - \int_0^t \mathbb{A}\phi(Z(s))\mathrm{d}s \tag{4.3.23}$$

is an $\mathscr{F}_t^Z$-martingale.

By the exchangeability of $X^n(t)$ and the definition of Z_n,

$$E[f(X_1^n(t),\ldots,X_m^n(t))] = E[\langle f, Z_n^{(m)}(t)\rangle] \tag{4.3.24}$$

and passing to the limit we have

$$E[f(X_1(t),\ldots,X_m(t))] = E[\langle f, Z^m(t)\rangle]. \tag{4.3.25}$$

It follows that conditionally on $Z(t)$, $X_1(t), X_2(t), \ldots$ are i.i.d. with distribution $Z(t)$; hence $X(t)$ is exchangeable with de Finetti measure given by the distribution of $Z(t)$, and

$$Z(t) = \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^m \delta_{X_i(t)} \qquad \text{a.s.} \tag{4.3.26}$$

from which it follows that Z is $\mathscr{F}_t^Z$-adapted and $\mathscr{F}_t^{X,Z} = \mathscr{F}_t^Z$.

Analogous construction and results can be given for the case with *fertility selection*, which affect the birth rate rather than the death rate. The Moran model (4.3.13) now becomes

$$A_0^n f(\mathbf{x}) = \sum_{i=1}^n B_i f(\mathbf{x}) + \frac{1}{2}\sum_{1\le i\ne j\le n}\left(1+\frac{2}{n}\sigma(x_i,\mu_n)\right)(f(\eta_j(\mathbf{x}|\,x_i)) - f(\mathbf{x})) \tag{4.3.27}$$

where a larger σ means a larger rate of reproduction, and again σ is a nonnegative, bounded, measurable function on $E\times\mathscr{P}(E)$. Note that the fertility selection function depends on the type of the parent rather than on the type of particle to be replaced. The generator of the ordered model (4.3.15) becomes

$$\begin{aligned} A^n f(\mathbf{x}) =& \sum_{i=1}^n B_i f(\mathbf{x}) + \sum_{1\le i<j\le n} (f(\theta_j(\mathbf{x}|\,x_i)) - f(\mathbf{x})) \\ &+ \sum_{i=1}^{n-1}\sum_{j=1}^n \left(\frac{1}{n}\sigma(x_i,\mu_n)\right)(f(\eta_j(\mathbf{x}|\,x_i)) - f(\mathbf{x})) \end{aligned} \tag{4.3.28}$$

while the limiting operators are now

$$Af(\mathbf{x},\mu) = \sum_{i=1}^{m} B_i f(\mathbf{x}) + \sum_{1\le i<j\le m} \left(f(\theta_j(\mathbf{x}|\, x_i)) - f(\mathbf{x})\right) + \sum_{j=1}^{m} \int_E \sigma(z,\mu)\left(f(\eta_j(\mathbf{x}|\, z)) - f(\mathbf{x})\right)\mu(\mathrm{d}z) \qquad (4.3.29)$$

in place of (4.3.20) and

$$\mathbb{A}\varphi(\mu) = \sum_{i=1}^{m} \langle B_i f, \mu^m\rangle + \frac{1}{2}\sum_{1\le i\le j\le m} \left(\langle \Phi_{ij} f - f, \mu^m\rangle\right) + \sum_{j=1}^{m} \left(\langle \sigma_j(\cdot,\mu) f, \mu^m\rangle - \langle \sigma(\cdot,\mu)\otimes f, \mu^{m+1}\rangle\right) \qquad (4.3.30)$$

in place of (4.3.21).

The point of introducing generators with a different labelling like (4.3.15) and (4.3.28) is that in the infinite population limit the Moran-type generators are not well behaved. The ordered model generators instead guarantee a proper convergence that enables to exploit the genealogical environment which is embedded in the particle system, in particular in terms of tracking ancestry from a certain node of the system backwards.

Part II

URN-BASED PARTICLE PROCESSES FOR FLEMING-VIOT MODELS

Chapter 5

Neutral diffusion model

In this chapter we construct a particle process which models the evolution in time of an exchangeable population and leads to the neutral[1] Fleming-Viot process, with the purpose of stressing the relationship between the latter and Bayesian nonparametrics. In particular we define a pure jump Markov process, based on the Blackwell-MacQueen urn scheme (see Section 2.2), with sample paths in $D_{E^n}[0,\infty)$, for some locally compact Polish space E, whose associated process of empirical measures weakly converges to the neutral Fleming-Viot process in the space $D_{\mathscr{P}(E)}[0,\infty)$, endowed with the Skorohod topology. We also show a new derivation of the stationary distribution of the neutral Fleming-Viot diffusion, which we know is the Dirichlet process.

[1] Recall from Section 4.2.1 that a process in population genetics is said to be neutral when the probability of a particle mutating, replicating or dying does not depend on its type.

5.1 Definition of the particle process

Let E be a locally compact complete separable metric space. For fixed n, consider a collection of n particles, each of which has a type represented by a point in the type space E, so that $\{\mathbf{X}^n(t), t \geq 0\}$ is a process defined on a probability space $(\Omega, \mathcal{F}, P)$ taking values in E^n, and $\mathbf{X}^n(t) = \big(X_1(t), \ldots, X_n(t)\big) \in E^n$ is the vector of coordinates of the process at time $t \geq 0$ (we will drop the superscript n when it causes no ambiguity). Throughout the second part a variable X_i in the vector $\mathbf{X}^n$ will be equivalently referred to as the ith component of the vector, the ith individual in the population, or the particle at level i.

Let now $\mathbf{X}(0) = (X_1(0), \ldots, X_n(0))$ be the initial state[2] of the process, where $X_1(0), \ldots, X_n(0)$ are generated by a Blackwell-MacQueen urn scheme of parameters (θ, ν_0), where $\theta > 0$ and ν_0 is a finite, non-atomic, positive measure on E. That is

$$\big(X_1(0), \ldots, X_n(0)\big) \sim \prod_{i=1}^{n} \frac{\theta\nu_0(\mathrm{d}x_i) + \sum_{k=1}^{i-1} \delta_{x_k}(\mathrm{d}x_i)}{\theta + i - 1}; \tag{5.1.1}$$

cf. (2.2.12).

Let

$$K^n(i, \mathrm{d}x_{i,t}, \mathrm{d}t; \mathbf{x}_{-i}) \tag{5.1.2}$$

be a transition kernel which maps $\{1, \ldots, n\} \times \mathscr{B}(E) \times [0, \infty) \times E^{n-1}$ into $[0, 1]$, where the value of $K^n(i, \mathrm{d}x_{i,t}, \mathrm{d}t; \mathbf{x}_{-i})$ denotes the probability of updating X_i with $\mathrm{d}x_{i,t}$ in time $\mathrm{d}t$, given the subvector $\mathbf{x}_{-i}$, obtained from $\mathbf{x}$ by removing the i-th coordinate (the time at which $\mathbf{x}_{-i}$ is referred to in (5.1.2) will be clear shortly). Consider the

[2]Note that specifying the initial distribution is not an essential requirement; cf. Remark 5.5.1 below.

marginal distributions

$$K_1^n(i) = K^n(i, E, [0, \infty), E^{n-1})$$

$$K_3^n(\mathrm{d}t) = K^n(\{1, \ldots, n\}, E, \mathrm{d}t, E^{n-1}).$$

Denote now

$$K_1^n(i) = \pi_i^n \tag{5.1.3}$$

where $\pi_i^n > 0$, for $i = 1, \ldots, n$, and $\sum_i \pi_i^n = 1$, that is π_i^n is the probability of choosing the i-th coordinate to be updated, and with

$$K^n(\mathrm{d}x_{i,t}, t; i, \mathbf{x}_{-i}) = \frac{K^n(i, \mathrm{d}x_{i,t}, [0, t]; \mathbf{x}_{-i})}{\pi_i^n}, \tag{5.1.4}$$

where $K^n(\cdot, \cdot, t; \cdot) = K^n(\cdot, \cdot, [0, t]; \cdot)$, the conditional distribution of updating the i-th coordinate not later than t, given that i is the drawn index. Finally define $\{S_m, m \geq 1\}$ to be a process of holding times such that $S_m : \Omega' \mapsto [0, \infty)$, denoting with $T_j = \sum_{m=1}^j S_m$ the renewal times. Let

$$K_3^n([0, t]) = P^n\{S_m \leq t\} = Q^n(t) \qquad t \geq 0, \ \forall m \geq 1 \tag{5.1.5}$$

be the distribution of the holding times, which is assumed to be independent of both departure and arrival state, but does depend on the population size.

Given the initial state $(X_1(0), \ldots, X_n(0))$ and $S_0 = 0$, the process can be described as follows. For $j \geq 1$, after the $(j-1)$-th renewal time $T_{j-1} = \sum_{m=1}^{j-1} S_m$, the vector remains constant in all its components until it reaches time $T_j = \sum_{m=1}^j S_m$, at which one component has a jump, so described.

Algorithm 5.1.1 (Jump of the particle process)

1. An index i is sampled from $\{1,\dots,n\}$ according to (5.1.3).

2. The subvector $\mathbf{x}_{-i}$ is set such that

$$X_k(t_j) = x_{k,t_j^-} \qquad k \neq i \tag{5.1.6}$$

   where x_{k,t_j^-} is the observed value $X_k(t_j^-)$.

3. $X_i(t_j)$ is sampled from the $(n-1)$-th predictive density associated with (5.1.1), that is

$$Pr\{X_i(t_j) \in \mathrm{d}x_{i,t_j} | i, \mathbf{x}_{-i,t_j}\} = \frac{\theta\nu_0(\mathrm{d}x_{i,t_j}) + \sum_{1\leq k\leq n,\, k\neq i} \delta_{x_{k,t_j}}(\mathrm{d}x_{i,t_j})}{\theta + n - 1} \tag{5.1.7}$$

4. The process is constant until T_{i+1}, when the procedure is repeated.

Observe that if we define

$$N(t) = \sup\{k \geq 0 : \sum_{m=1}^{k} S_m \leq t\}$$

to be the process which counts how many renewal times occur up to time t, then the process $\{\mathbf{Y}_{N(t)}, t \geq 0, N(t) \in \mathbb{N}\}$, such that

$$\mathbf{Y}_{N(t)} = \left(Y_{1,N(t)}, \dots, Y_{n,N(t)}\right) \tag{5.1.8}$$

where

$$Y_{i,N(t)} = X_i(t), \tag{5.1.9}$$

is a Markov chain, with state space E^n, embedded at jump times in the continuous time process $\{\mathbf{X}(t), t \geq 0\}$, the former describing the successive states of the latter. Note that $\{\mathbf{Y}_{N(t)}, t \geq 0, N(t) \in \mathbb{N}\}$ is the Markov chain obtained by applying the Gibbs sampler algorithm (see Section 1.3) to the initial vector $\mathbf{X}(0) = (X_1(0), \ldots, X_n(0))$, where the updatings follow a random scan according to (5.1.3).

Thus $\{\mathbf{X}(t), t \geq 0\}$ is an E^n-valued pure jump process, whose components are piecewise constant. Therefore, when in (5.1.2) we write $K(i, \mathrm{d}x_{i,t}, \mathrm{d}t; \mathbf{x}_{-i})$, it is now clear that the values of the subvector $\mathbf{x}_{-i}$ are referred indifferently to any t lying in the interval $[\max\{T_j : T_j < t\}, t]$, since $\mathbf{x}_{-i}$ is constant in that interval.

As a consequence, (5.1.7) is independent of the interval length between two successive renewal times, so that, using (5.1.5) and remembering that the holding times are independent of the initial state, we can write

$$\begin{aligned} K^n(\mathrm{d}x_{i,t}, t; i, \mathbf{x}_{-i}) &= K_3^n([0,t]) Pr\{X_i \in \mathrm{d}x_i |\, i, t, \mathbf{x}_{-i}\} \\ &= Q^n(t) Pr\{X_i \in \mathrm{d}x_i |\, i, \mathbf{x}_{-i}\}. \end{aligned} \tag{5.1.10}$$

If we further assume that

$$Q^n(t) = 1 - e^{-\lambda_n t} \tag{5.1.11}$$

then the Markov property for $\{\mathbf{X}(t), t \geq 0\}$ follows directly. Indeed the process $\mathbf{X}(t)$ can be considered as a superposition of a Poisson point process on the real line with intensity rate λ_n and the Markov chain $\mathbf{Y}_{N(t)}$, given by (5.1.9), where the Poisson point process determines the time points at which the transitions of $\mathbf{Y}_{N(t)}$ occur.

5.2 Derivation of the generators

Let $T_i^n(t)f(\mathbf{x})$ be the semigroup transition operator associated to the process $\mathbf{X}^n(t)$, conditionally on the transition regarding coordinate i, namely

$$\begin{aligned} T_i^n(t)f(\mathbf{x}_0) &= E\big(f(\mathbf{X}^n(t)|\mathbf{X}^n(0) = \mathbf{x}_0^n\big) \\ &= \int f(\mathbf{x}_t^n)K^n(\mathrm{d}x_{i,t}, t; i, \mathbf{x}_{-i,0}^n) \end{aligned} \tag{5.2.1}$$

where $K^n(\mathrm{d}x_{i,t}, t; i, \mathbf{x}_{-i,0}^n)$ is (5.1.4), and f is a real-valued, bounded, Borel measurable function on E. Then the infinitesimal generator, conditional on the i-th component being updated, is

$$\begin{aligned} A_i^n f(\mathbf{x}_0^n) &= \lim_{t\downarrow 0} \frac{1}{t}\big[T_i^n(t)f(\mathbf{x}_0^n) - f(\mathbf{x}_0^n)\big] \\ &= \lim_{t\downarrow 0} \frac{1}{t}\int \big[f(\mathbf{x}_t^n) - f(\mathbf{x_0})\big]K^n(\mathrm{d}x_{i,t}, t; i, \mathbf{x}_{-i,0}^n) \end{aligned}$$

which, using (5.1.10) and (5.1.11), becomes

$$\begin{aligned} &= \lim_{t\downarrow 0} \frac{1}{t}\int \big[f(\mathbf{x}_t^n) - f(\mathbf{x_0})\big]Q^n(t)Pr\{X_i \in \mathrm{d}x_i | i, \mathbf{x}_{-i}\} \\ &= \lim_{t\downarrow 0} \frac{1 - e^{-\lambda_n t}}{t}\int \big[f(\mathbf{x}_t^n) - f(\mathbf{x_0})\big]Pr\{X_i \in \mathrm{d}x_i | i, \mathbf{x}_{-i}\} \\ &= \lambda_n \int \big[f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})\big]\frac{\theta\nu_0(\mathrm{d}y) + \sum_{k\neq i}^n \delta_{x_k}(\mathrm{d}y)}{\theta + n - 1} \end{aligned} \tag{5.2.2}$$

where $\eta_i(\mathbf{x}|\, y)$ denotes the vector obtained from $\mathbf{x}$ by replacing x_i with y. Now, assuming that (5.1.3) is uniform, i.e.

$$\pi_i^n = \frac{1}{n} \qquad i = 1, \ldots, n \tag{5.2.3}$$

the generator of the process $\mathbf{X}^n$ is given by

$$A^n f(\mathbf{x}) = \sum_{i=1}^{n} \frac{\lambda_n}{n} \int \big[f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})\big] \frac{\theta\nu_0(\mathrm{d}y) + \sum_{k\neq i}^{n} \delta_{x_k}(\mathrm{d}y)}{\theta + n - 1}$$
$$= \sum_{i=1}^{n} \frac{\lambda\theta}{n(\theta+n-1)} \int \big[f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})\big]\nu_0(\mathrm{d}y)$$
$$+ \sum_{i=1}^{n} \frac{\lambda}{n(\theta+n-1)} \int \big[f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})\big] \sum_{k\neq i} \delta_{x_k}(\mathrm{d}y) \tag{5.2.4}$$

Now, since $\int f(x)\delta_y(\mathrm{d}x) = f(y)$, if we also let λ_n be

$$\lambda_n = \frac{n(\theta+n-1)}{2} \tag{5.2.5}$$

we obtain

$$\sum_{i=1}^{n} \frac{1}{2}\theta \int \big[f(\eta_i(\mathbf{x}|\, x_k)) - f(\mathbf{x})\big]\nu_0(\mathrm{d}x_k) + \frac{1}{2}\sum_{i=1}^{n}\sum_{k\neq i}^{n} \big[f(\eta_i(\mathbf{x}|\, x_k)) - f(\mathbf{x})\big] \tag{5.2.6}$$

which can can be written

$$A^n f(\mathbf{x}) = \sum_{i=1}^{n} B_i f(\mathbf{x}) + \frac{1}{2} \sum_{1\leq k\neq i\leq n} \big[f(\eta_i(\mathbf{x}|\, x_k)) - f(\mathbf{x})\big] \tag{5.2.7}$$

where B_i is the bounded operator

$$Bg(x) = \frac{1}{2}\theta \int \big[g(y) - g(x)\big]\nu_0(\mathrm{d}y) \tag{5.2.8}$$

acting on f as a function of x_i, for $g \in B(E)$. Note that B is the generator of a pure jump Markov process on E.

Let us proceed now by defining $\mu_n \in \mathscr{P}^n(E)$ by

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \tag{5.2.9}$$

where $\mathscr{P}^n(E) \subset \mathscr{P}(E)$ is the set of atomic probability measures on E with masses proportional to $1/n$, and for $f \in B(E^m)$, $m < n$, defining $\Gamma^n f \in B(E^m)$ by

$$\Gamma^n f(\mathbf{x}) = \frac{1}{n^m} \sum_{1\leq i_1\neq\ldots\neq i_m\leq n} f_{i_1\ldots i_m}(\mathbf{x}) \tag{5.2.10}$$

where $f_{i_1\dots i_m}(\mathbf{x}) = f(x_{i_1},\dots,x_{i_m})$ (we drop the superscript of Γ^n if it does not cause ambiguity). Note that for $f \in B(E^m)$, by simply applying the operator B to $g = \Gamma f$, we have

$$\begin{aligned}\sum_{i=1}^{n} B_i\Gamma f(\mathbf{x}) =& \sum_{i=1}^{n}\frac{1}{2}\theta\int\big[\Gamma f(\eta_i(\mathbf{x}|\,y)) - \Gamma f(\mathbf{x})\big]\nu_0(\mathrm{d}y)\\ =& \sum_{i=1}^{m}\frac{1}{2}\theta\int\big[\Gamma f(\eta_i(\mathbf{x}|\,y)) - \Gamma f(\mathbf{x})\big]\nu_0(\mathrm{d}y)\\ &+\sum_{i=m+1}^{n}\frac{1}{2}\theta\int\big[\Gamma f(\eta_i(\mathbf{x}|\,y)) - \Gamma f(\mathbf{x})\big]\nu_0(\mathrm{d}y)\\ =& \sum_{i=1}^{m}\frac{1}{2}\theta\int\big[\Gamma f(\eta_i(\mathbf{x}|\,y)) - \Gamma f(\mathbf{x})\big]\nu_0(\mathrm{d}y)\\ =& \sum_{i=1}^{m} B_i\Gamma f(\mathbf{x})\end{aligned}$$

since, for $m+1 \le i \le n$, x_i is not an argument of f and $f(\eta_i(\mathbf{x}|\,y)) = f(\mathbf{x})$. For the same reason

$$\begin{aligned}&\sum_{1\le k\ne i\le n}\big[\Gamma f(\eta_i(\mathbf{x}|\,x_k)) - \Gamma f(\mathbf{x})\big] =\\ &\quad=\sum_{i=1}^{m}\sum_{1\le k\le n:\,k\ne i}\big[\Gamma f(\eta_i(\mathbf{x}|\,x_k)) - \Gamma f(\mathbf{x})\big]\\ &\quad\quad+\sum_{i=m+1}^{n}\sum_{1\le k\le n:\,k\ne i}\big[\Gamma f(\eta_i(\mathbf{x}|\,x_k)) - \Gamma f(\mathbf{x})\big]\\ &\quad=\sum_{i=1}^{m}\sum_{1\le k\le n:\,k\ne i}\big[\Gamma f(\eta_i(\mathbf{x}|\,x_k)) - \Gamma f(\mathbf{x})\big].\end{aligned} \tag{5.2.11}$$

Further

$$\begin{aligned}
\sum_{j=1}^{m} B_j \Gamma f(\mathbf{x}) &= \sum_{j=1}^{m} \frac{1}{2}\theta \int \left[\Gamma f(\eta_j(\mathbf{x}|\, y)) - \Gamma f(\mathbf{x})\right] \nu_0(\mathrm{d}y) \\
&= \sum_{j=1}^{m} \frac{1}{2}\theta \int \Bigg[\frac{1}{n^m} \sum_{1\le i_1 \neq \ldots \neq i_m \le n} f_{i_1\ldots i_m}(\eta_{i_j}(\mathbf{x}|\, y)) \\
&\quad - \frac{1}{n^m} \sum_{1\le i_1 \neq \ldots \neq i_m \le n} f_{i_1\ldots i_m}(\mathbf{x})\Bigg] \nu_0(\mathrm{d}y) \\
&= \sum_{j=1}^{m} \frac{1}{2}\theta \int \frac{1}{n^m} \sum_{1\le i_1 \neq \ldots \neq i_m \le n} \left[f_{i_1\ldots i_m}(\eta_{i_j}(\mathbf{x}|\, y)) - f_{i_1\ldots i_m}(\mathbf{x})\right] \nu_0(\mathrm{d}y) \\
&= \sum_{j=1}^{m} \frac{1}{n^m} \sum_{1\le i_1 \neq \ldots \neq i_m \le n} \frac{1}{2}\theta \int \left[f_{i_1\ldots i_m}(\eta_{i_j}(\mathbf{x}|\, y)) - f_{i_1\ldots i_m}(\mathbf{x})\right] \nu_0(\mathrm{d}y) \\
&= \sum_{j=1}^{m} \frac{1}{n^m} \sum_{1\le i_1 \neq \ldots \neq i_m \le n} B_{i_j} f_{i_1\ldots i_m}(\mathbf{x}) \\
&= \sum_{j=1}^{m} \Gamma B_j f(\mathbf{x})
\end{aligned}$$

and

$$
\begin{aligned}
&\sum_{j=1}^{m} \sum_{1\le k\le n:\, k\neq j} \Big(\Gamma f(\eta_j(\mathbf{x}|\, x_k)) - \Gamma f(\mathbf{x})\Big) \\
&=\sum_{j=1}^{m} \Bigg[\sum_{1\le k\le m:\, k\neq j} \Bigg(\frac{1}{n^m} \sum_{1\le i_1\neq\ldots\neq i_m\le n} f_{i_1\ldots i_m}(\eta_{i_j}(\mathbf{x}|\, x_{i_k})) - \frac{1}{n^m} \sum_{1\le i_1\neq\ldots\neq i_m\le n} f_{i_1\ldots i_m}(\mathbf{x}) \Bigg) \\
&\quad + \sum_{l\notin\{i_1,\ldots,i_m\}} \Bigg(\frac{1}{n^m} \sum_{1\le i_1\neq\ldots\neq i_m\le n} f_{i_1\ldots i_m}(\eta_{i_j}(\mathbf{x}|\, x_l)) - \frac{1}{n^m} \sum_{1\le i_1\neq\ldots\neq i_m\le n} f_{i_1\ldots i_m}(\mathbf{x}) \Bigg) \Bigg] \\
&=\sum_{j=1}^{m} \Bigg[\sum_{1\le k\le m:\, k\neq j} \frac{1}{n^m} \sum_{1\le i_1\neq\ldots\neq i_m\le n} \Big(f_{i_1\ldots i_m}(\eta_{i_j}(\mathbf{x}|\, x_{i_k})) - f_{i_1\ldots i_m}(\mathbf{x}) \Big) \\
&\quad + \sum_{l\notin\{i_1,\ldots,i_m\}} \frac{1}{n^m} \sum_{1\le i_1\neq\ldots\neq i_m\le n} \Big(f_{i_1\ldots i_m}(\eta_{i_j}(\mathbf{x}|\, x_l)) - f_{i_1\ldots i_m}(\mathbf{x}) \Big) \Bigg] \\
&=\sum_{j=1}^{m} \sum_{1\le k\le m:\, k\neq j} \frac{1}{n^m} \sum_{1\le i_1\neq\ldots\neq i_m\le n} \Big(f_{i_1\ldots i_m}(\eta_{i_j}(\mathbf{x}|\, x_{i_k})) - f_{i_1\ldots i_m}(\mathbf{x}) \Big) \\
&\quad + \sum_{j=1}^{m} \sum_{l\notin\{i_1,\ldots,i_m\}} \frac{1}{n^m} \sum_{1\le i_1\neq\ldots\neq i_m\le n} \Big(f_{i_1\ldots i_m}(\eta_{i_j}(\mathbf{x}|\, x_l)) - f_{i_1\ldots i_m}(\mathbf{x}) \Big) \\
&=\sum_{j=1}^{m} \sum_{1\le k\le m:\, k\neq j} \Gamma\Big(f(\eta_j(\mathbf{x}|\, x_k)) - f(\mathbf{x}) \Big) \\
&\quad + \sum_{j=1}^{m} (n-m)\Gamma\Big(f(\mathbf{x}) - f(\mathbf{x}) \Big) \\
&= \sum_{1\le k\neq j\le m} \Gamma\Big(f(\eta_j(\mathbf{x}|\, x_k)) - f(\mathbf{x}) \Big)
\end{aligned}
\tag{5.2.12}
$$

where in the fourth equality we have used the fact that

$$
\begin{aligned}
\sum_{l\notin\{i_1,\ldots,i_m\}} \Gamma\Phi_{l\cdot} f &= \sum_{l\notin\{i_1,\ldots,i_m\}} \frac{1}{n^m} \sum_{1\le i_1\neq\ldots\neq i_m\le n} \Phi_{l i_j} f_{i_1\ldots i_m} \\
&= \sum_{l\notin\{i_1,\ldots,i_m\}} \frac{1}{n^m} \sum_{1\le i_1\neq\ldots\neq i_m\le n} f_{i_1\ldots i_m} \\
&=(n-m)\Gamma f
\end{aligned}
\tag{5.2.13}
$$

where Φ_{ki} denotes the function of $m-1$ variables obtained by setting the i-th and the k-th variable in f equal. Loosely speaking, the second identity in (5.2.13) is due to the fact that when x_l is not an argument of f, then the operator $\Phi_{l\cdot}$ does not cause an effective change in f; this is because the two sums are taken over all possible choices of indices which are different from one another and belong to the set $\{1,\ldots,n\}$, thus already contemplating the case

$$\Phi_{li_j} f_{i_1\ldots i_m} = f_{i_1\ldots i_{j-1} l i_{i+1} i_m}$$

since that l is indeed an index different from all the others picked from the remainder of the set. If for example we have $\{1,2,3\}$, then

$$\sum_{1\le i_1\ne i_2\le 3} f_{i_1 i_2} = f_{12} + f_{13} + f_{23}$$

and, for example, $\Phi_{23} f_{13} = f_{12}$ is already in the sum.

Therefore we can write, for $f \in B(E^m)$,

$$\begin{aligned} A^n \Gamma f(\mathbf{x}) &= \sum_{j=1}^{m} B_j \Gamma f(\mathbf{x}) + \frac{1}{2}\sum_{j=1}^{m} \sum_{1\le k\le n:\, k\ne j} \big[\Gamma f(\eta_j(\mathbf{x}|\, x_k)) - \Gamma f(\mathbf{x})\big] \\ &= \sum_{j=1}^{m} \Gamma B_{i_j} f(\mathbf{x}) + \frac{1}{2} \sum_{1\le k\ne j\le m} \Gamma\left(\Phi_{kj} f(\mathbf{x}) - f(\mathbf{x})\right). \end{aligned} \tag{5.2.14}$$

This version of the operator A^n will be used in the next Section to derive the infinite population limit for the generator of the process of empirical measures associated with the E^n-valued process $\mathbf{X}^n$.

5.3 Infinite population limit

Define

$$\varphi(\mu) = \langle f, \mu^m \rangle = \int_{E^m} f d\mu^m \tag{5.3.1}$$

where μ^m denotes the m-fold product measure. For $n \to \infty$

$$\sup_{\mathbf{x} \in E^n} |\Gamma f(\mathbf{x}) - \langle f, \mu_{\mathbf{x}}^m \rangle| \longrightarrow 0 \tag{5.3.2}$$

where Γ is defined in (5.2.10) (cf. Donnelly and Kurtz, 1996, p. 701). This is due to the fact that

$$\begin{aligned}
\langle f, \mu_n^m \rangle =& \int_{E^m} f(x_1, \ldots, x_m) \mu_n(\mathrm{d}x_1) \ldots \mu_n(\mathrm{d}x_m) \\
=& \frac{1}{n^m} \int_{E^m} f(x_1, \ldots, x_m) \sum_{1 \le i_1 \le n} \delta_{x_{i_1}}(\mathrm{d}x_1) \cdots \sum_{1 \le i_m \le n} \delta_{x_{i_m}}(\mathrm{d}x_m) \\
=& \frac{1}{n^m} \sum_{1 \le i_1 \le n} \cdots \sum_{1 \le i_m \le n} f(x_{i_1}, \ldots, x_{i_m}) \\
=& \frac{1}{n^m} \sum_{1 \le i_1 \ne \cdots \ne i_m \le n} f(x_{i_1}, \ldots, x_{i_m}) + \frac{1}{n^m} \sum_{\text{at least 2 } i_j \text{ equal}} f(x_{i_1}, \ldots, x_{i_m})
\end{aligned} \tag{5.3.3}$$

where, defining $n_{(m)} = n(n-1)\ldots(n-m+1)$, the left-hand sum has $n_{(m)}$ terms, while the right-hand sum has $n^m - n_{(m)}$, and $n_{(m)}$ approaches n^m as $n \to \infty$.

Defining

$$\mathbb{A}\varphi(\mu) = \sum_{i=1}^{m} \langle B_i f, \mu^m \rangle + \frac{1}{2} \sum_{1 \le k \ne i \le m} \langle \Phi_{ki} f - f, \mu^m \rangle \tag{5.3.4}$$

we have

$$
\begin{aligned}
&\sup_{\mathbf{x}\in E^n} |A^n \Gamma f(\mathbf{x}) - \mathbb{A}\varphi(\mu_n)| \\
&\quad = \sup_{\mathbf{x}\in E^n} \left| \sum_{i=1}^m \Gamma B_i f(\mathbf{x}) + \frac{1}{2} \sum_{1\le k\neq i\le m} \Gamma\left(\Phi_{ki} f(\mathbf{x}) - f(\mathbf{x})\right) \right. \\
&\qquad\qquad \left. - \sum_{i=1}^m \langle B_i f, \mu_n^m\rangle - \frac{1}{2} \sum_{1\le k\neq i\le m} \langle \Phi_{ki} f - f, \mu_n^m\rangle \right| \\
&\quad = \sup_{\mathbf{x}\in E^n} \left| \sum_{i=1}^m \left[\Gamma B_i f(\mathbf{x}) - \langle B_i f, \mu_n^m\rangle\right] \right. \\
&\qquad\qquad \left. + \frac{1}{2} \sum_{1\le k\neq i\le m} \left[\Gamma\left(\Phi_{ki} f(\mathbf{x}) - f(\mathbf{x})\right) - \langle \Phi_{ki} f - f, \mu_n^m\rangle\right] \right| \\
&\quad \le \sum_{i=1}^m \sup_{\mathbf{x}\in E^n} |\Gamma B_i f(\mathbf{x}) - \langle B_i f, \mu_n^m\rangle| \\
&\qquad\qquad + \frac{1}{2} \sum_{1\le k\neq i\le m} \sup_{\mathbf{x}\in E^n} \left|\Gamma\left(\Phi_{ki} f(\mathbf{x}) - f(\mathbf{x})\right) - \langle \Phi_{ki} f - f, \mu_n^m\rangle\right|. \qquad (5.3.5)
\end{aligned}
$$

By (5.3.2), last expression converges to zero as $n \to \infty$, so that

$$\sup_{\mathbf{x}\in E^n} |A^n \Gamma f(\mathbf{x}) - \mathbb{A}\varphi(\mu_n)| \longrightarrow 0 \qquad (5.3.6)$$

where $\mathbb{A}\varphi(\mu)$, defined in (5.3.4), is the generator of the neutral Fleming-Viot process (cf. generator (4.2.21) with $\sigma \equiv 0$) which, as seen in Section 4.2.2, lives in $\mathscr{P}_a(E)$, the set of purely atomic Borel probability measures on E. Since also

$$\lim_{n\to\infty} A^n \Gamma f(\mathbf{x}) = \lim_{n\to\infty} \Gamma A^n f(\mathbf{x}) = \langle Af, \mu_n^m\rangle \qquad (5.3.7)$$

for some limiting operator A, assuming the existence of such an A, in virtue of (5.3.6) we can identify

$$\mathbb{A}\varphi(\mu_n) = \langle Af, \mu_n^m\rangle. \qquad (5.3.8)$$

Loosely speaking, (5.3.8) says that taking the E^∞-valued process $\mathbf{X}$ and applying the generator for the Fleming-Viot process to the empirical measure of the X_i's is the same as applying the limit, if it exists, of the generator (5.2.7) to the E^∞-valued process and then taking the empirical measure. More formally, given the process $\mathbf{X}$, the conditional expectation of a function φ of the limiting empirical measure μ_n taken on the space E^∞ or on the space $\mathscr{P}^n(E)$ (of atomic measures with masses multiples to $1/n$) yields the same result. Note that this is the same concept expressed in (4.3.7).

We conclude the section by deriving a simplified version of generator (5.3.4) which is often used, under the assumption of multiplicative f. Note that if $f(x_1, \dots, x_m) = f(x_1) \dots f(x_m)$, then we have

$$
\begin{aligned}
\langle f, \mu^m \rangle &= \int_{E^m} f(x_1, \dots, x_m) \mu^m(\mathrm{d}x_1 \times \dots \times \mathrm{d}x_m) \\
&= \int_{E^m} f(x_1) \dots f(x_m) \mu(\mathrm{d}x_1) \dots \mu(\mathrm{d}x_m) \\
&= \langle f_1, \mu \rangle \dots \langle f_m, \mu \rangle
\end{aligned}
$$

so that $\varphi(\mu) = \langle f_1, \mu\rangle \dots \langle f_m, \mu\rangle$. Thus the first term of (5.3.4) becomes

$$
\begin{aligned}
&\sum_{i=1}^{m}\langle B_i f, \mu^m\rangle = \\
&=\sum_{i=1}^{m}\int_{E^m} B_i f(x_1,\dots,x_m)\mu(\mathrm{d}x_1)\dots\mu(\mathrm{d}x_m) \\
&=\frac{1}{2}\theta\sum_{i=1}^{m}\int_{E^m}\int_E f_i(x_i)\nu_0(\mathrm{d}x_i)\prod_{l\neq i} f_l(x_l)\mu(\mathrm{d}x_1)\dots\mu(\mathrm{d}x_m) \\
&\quad-\frac{1}{2}\theta\sum_{i=1}^{m}\int_{E^m}\prod_{i=1}^{m} f_i(x_i)\mu(\mathrm{d}x_1)\dots\mu(\mathrm{d}x_m) \\
&=\frac{1}{2}\theta\sum_{i=1}^{m}\int_E Pf_i(x_i)\mu(\mathrm{d}x_i)\prod_{l\neq i}\int_E f_l(x_l)\mu(\mathrm{d}x_l) \\
&\quad-\frac{1}{2}\theta m\prod_{i=1}^{m}\int_E f_i(x_i)\mu(\mathrm{d}x_i) \\
&=\frac{1}{2}\theta\sum_{i=1}^{m}\langle Pf_i,\mu\rangle\prod_{l\neq i}\langle f_l,\mu\rangle-\frac{1}{2}\theta m\prod_{i=1}^{m}\langle f_i,\mu\rangle \\
&=\frac{1}{2}\theta\sum_{i=1}^{m}\Big[\langle Pf_i,\mu\rangle-\langle f_i,\mu\rangle\Big]\prod_{l\neq i}\langle f_l,\mu\rangle \\
&=\frac{1}{2}\theta\sum_{i=1}^{m}\Big[\langle Pf_i,\mu\rangle-\langle f_i,\mu\rangle\Big]F_{z_i}\big(\langle f_1,\mu\rangle,\dots,\langle f_m,\mu\rangle\big) \qquad (5.3.9)
\end{aligned}
$$

where

$$Pf(x) = \int_E f(x)\nu_0(\mathrm{d}x) \qquad (5.3.10)$$

and F_{z_i} is the partial derivative of F with respect to $z_i = \langle f_i, \mu\rangle$. The second term of

(5.3.4) becomes

$$
\begin{aligned}
\sum_{1\le k\neq i\le m}&\Big[\langle \Phi_{ki} f, \mu^m\rangle - \langle f, \mu^m\rangle\Big] = \\
&= \sum_{1\le k\neq i\le m}\Big[\langle f_i,\mu\rangle\langle f_k,\mu\rangle \prod_{l\neq i,k}\langle f_l,\mu\rangle - \prod_{i=1}^{m}\langle f_i,\mu\rangle\Big] \\
&= \sum_{1\le k\neq i\le m}\Big[\langle f_i f_k,\mu\rangle \prod_{l\neq i,k}\langle f_l,\mu\rangle - \prod_{i=1}^{m}\langle f_i,\mu\rangle\Big] \\
&= \sum_{1\le k\neq i\le m}\Big[\langle f_i f_k,\mu\rangle - \langle f_i,\mu\rangle\langle f_k,\mu\rangle\Big] \prod_{l\neq i,k}\langle f_l,\mu\rangle \\
&= \sum_{1\le k\neq i\le m}\Big[\langle f_i f_k,\mu\rangle - \langle f_i,\mu\rangle\langle f_k,\mu\rangle\Big] F_{z_i z_k}\big(\langle f_1,\mu\rangle,\dots,\langle f_m,\mu\rangle\big)
\end{aligned}
$$

where in the second equality we used the fact that

$$
\int_E f_i(x)\mu(\mathrm{d}x)\int_E f_k(y)\mu(\mathrm{d}y)\Big|_{y=x} = \int_E \mu(\mathrm{d}x)\int_E f_i(x) f_k(x)\mu(\mathrm{d}x).
$$

The generator of the neutral Fleming-Viot process (5.3.4) reduces to

$$
\begin{aligned}
\mathbb{A}F(\mu) =&\frac{1}{2}\theta\sum_{i=1}^{m}\Big[\langle Pf_i,\mu\rangle - \langle f_i,\mu\rangle\Big] F_{z_i}\big(\langle f_1,\mu\rangle,\dots,\langle f_m,\mu\rangle\big) \\
&+\frac{1}{2}\sum_{1\le k\neq i\le m}\Big[\langle f_i f_k,\mu\rangle - \langle f_i,\mu\rangle\langle f_k,\mu\rangle\Big] F_{z_i z_k}\big(\langle f_1,\mu\rangle,\dots,\langle f_m,\mu\rangle\big)
\end{aligned}
\tag{5.3.11}
$$

which is (4.2.4). Note that this can also be written

$$
\begin{aligned}
\mathbb{A}F(\mu) =&\frac{1}{2}\theta\sum_{i=1}^{m}\Big[\langle Pf_i,\mu\rangle-\langle f_i,\mu\rangle\Big]F_{z_i}\big(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle\big)\\
&+\frac{1}{2}\sum_{1\le k\neq i\le m}\Big[\langle f_if_k,\mu\rangle-\langle f_i,\mu\rangle\langle f_k,\mu\rangle\Big]F_{z_iz_k}\big(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle\big)\\
=&\frac{1}{2}\theta\sum_{i=1}^{m}\langle P_if,\mu\rangle F_i\big(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle\big)-\frac{1}{2}\theta mF\big(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle\big)\\
&+\frac{1}{2}\sum_{1\le k\neq i\le m}\langle f_if_k,\mu\rangle F_{ik}\big(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle\big)\\
&-\frac{1}{2}m(m-1)F\big(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle\big)\\
=&\frac{1}{2}\theta\sum_{i=1}^{m}\langle P_if,\mu\rangle F_{z_i}\big(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle\big)\\
&+\frac{1}{2}\sum_{1\le k\neq i\le m}\langle f_if_k,\mu\rangle F_{z_iz_k}\big(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle\big)\\
&-\frac{m(\theta+m-1)}{2}F\big(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle\big)\\
=&\frac{1}{2}\theta\sum_{i=1}^{m}\langle f_i,\nu_0\rangle F_{z_i}\big(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle\big)\\
&+\sum_{1\le k<i\le m}\langle f_if_k,\mu\rangle F_{z_iz_k}\big(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle\big)\\
&-\lambda_mF\big(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle\big)
\end{aligned}
\tag{5.3.12}
$$

(which appears in Ethier and Griffiths, 1993, proof of Theorem 1.1), where in the last equality we used the fact that

$$
\int_E\int_E f(x)\nu_0(\mathrm{d}x)\mu(\mathrm{d}x)=\int_E\mu(\mathrm{d}x)\int_E f(x)\nu_0(\mathrm{d}x)=\int_E f(x)\nu_0(\mathrm{d}x)
$$

and that

$$
\frac{1}{2}\sum_{1\le k\neq i\le n}\big(\langle\Phi_{ki}f,\mu^m\rangle-\langle f,\mu^m\rangle\big)=\sum_{1\le k<i\le n}\big(\langle\Phi_{ki}f,\mu^m\rangle-\langle f,\mu^m\rangle\big)
\tag{5.3.13}
$$

the latter of which follows from the identity $\langle\Phi_{ki}f,\mu^m\rangle=\langle\Phi_{ik}f,\mu^m\rangle$.

5.4 Weak convergence

Define $\mu_{\mathbf{X}^n} : E^n \times [0,\infty) \to \mathscr{P}(E)$ by

$$\mu_{\mathbf{X}^n}(t) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i^n(t)}.$$

that is $\mu_{\mathbf{X}^n}$ is the process of empirical measures of the X_i's, and observe that it has sample paths in $D_{\mathscr{P}(E)}[0,\infty)$. The above construction implies the weak convergence of $\{\mu_{\mathbf{X}^n}(t), t \geq 0\}$ to the Fleming-Viot process with generator (5.3.4), denoted by $\{\mu_t, t \geq 0\}$.

Theorem 5.4.1 *Let $\{\mu_{\mathbf{X}^n}(t), t \geq 0\}$ be a process with values in $\mathscr{P}(E)$ as in the above construction, and let $\{\mu_t, t \geq 0\}$ be a Fleming-Viot process with generator* (5.3.4)*. Assume that $P\mu_{\mathbf{X}^n}^{-1}(0) \Rightarrow \mu_0$, for $\mu_0 \in \mathscr{P}(E)$. Then*

$$\{\mu_{\mathbf{X}^n}(t), t \geq 0\} \Rightarrow \{\mu_t, t \geq 0\} \tag{5.4.1}$$

in $D_{\mathscr{P}(E)}[0,\infty)$, where $\Rightarrow$ means convergence in distribution in the Skorohod topology.

PROOF 1. Since the mutation operator does not depend on the population size, the thesis follows directly from (5.3.2) and (5.3.6), analogously to Section 4.3.1. Cf. (4.3.4) and (4.3.6). □

PROOF 2. Assume the process $\{\mu_{\mathbf{X}^n}(t), t \geq 0\}$ is relatively compact[3] in $D_{\mathscr{P}(E)}[0,\infty)$, which follows from Donnelly and Kurtz (1996). Define

$$\{\, (\xi_n, \psi_n) \,\} = \{ \big(g_n(\mathbf{X}^n), h_n(\mathbf{X}^n)\big) \} \tag{5.4.2}$$

[3]Note that, for a single time point, if E is compact, then $(\mathscr{P}(S), \rho)$ is compact, where ρ is Prohorov's metric (which induces the topology of weak convergence), and any collection of distributions on $(\mathscr{P}(S), \rho)$ is relatively compact in the weak topology.

where, for $f \in B(E^m)$,

$$g_n = \Gamma^n f \tag{5.4.3}$$

as in (5.2.10), and

$$h_n = A^n g_n = A^n \Gamma^n f \tag{5.4.4}$$

with A^n as in (5.2.7). Then for each $T \geq 0$

$$\begin{aligned}\sup_{n\to\infty}\sup_{s\leq T} E[|\xi_n(s)|] &= \\ &= \sup_{n\to\infty}\sup_{s\leq T} E[|g_n(\mathbf{X}^n(s))|] \\ &= \sup_{n\to\infty}\sup_{s\leq T} E[|\Gamma^n f(\mathbf{X}^n(s))|] \\ &= \sup_{n\to\infty}\sup_{s\leq T} E\left[\left|\frac{1}{n^m}\sum_{1\leq i_1\neq\ldots\neq i_m\leq n} f\left(X^n_{i_1}(s),\ldots,X^n_{i_m}(s)\right)\right|\right] < \infty \end{aligned} \tag{5.4.5}$$

since $f \in B(E^m)$, and

$$\begin{aligned}\sup_{n\to\infty}\sup_{s\leq T} E[|\psi_n(s)|] &= \sup_{n\to\infty}\sup_{s\leq T} E[|h_n(\mathbf{X}^n(s))|] \\ &= \sup_{n\to\infty}\sup_{s\leq T} E[|A^n\Gamma^n f(\mathbf{X}^n(s))|] < \infty\end{aligned} \tag{5.4.6}$$

since A^n is a bounded operator. Further, let

$$\mathbb{A}' = \{(\varphi', \mathbb{A}\varphi') : \varphi' \in \mathscr{D}(\mathbb{A})\} \subset B(\mathscr{P}(E)) \times B(\mathscr{P}(E)) \tag{5.4.7}$$

where

$$\mathscr{D}(\mathbb{A}) = \{\varphi' = \langle f, \mu^m\rangle :\ f \in B(E^m),\ \mu \in \mathscr{P}^n(E),\ n \geq m \geq 1\}\,. \tag{5.4.8}$$

and observe that for each $(g,h) = (\varphi, \mathbb{A}\varphi) \in \mathbb{A}'$ where $\varphi = \langle f, \mu^m\rangle$ and μ is purely atomic, there exist $(g_n, h_n) \in \widehat{\mathbb{A}}^n$, where

$$\widehat{\mathbb{A}}^n = \{(\phi', \mathbb{A}^n\phi') : \phi' \in \mathscr{D}(\mathbb{A}^n)\} \subset B(\mathscr{P}(E)) \times B(\mathscr{P}(E)) \tag{5.4.9}$$

and, since $\Gamma^n f = n^{-m} n_{(m)} \langle f, \mu^{(m)} \rangle$, with $n_{(m)} = n(n-1)\dots(n-m+1)$ (cf. (5.3.3) and (4.3.16)),

$$\mathscr{D}(\mathbb{A}^n) = \{\phi' = \Gamma^n f : \; f \in B(E^m), \; n \geq m \geq 1\}, \tag{5.4.10}$$

such that, for $t \leq T$ and $t \neq k$, $k = 0, 1, \dots$, (5.3.2) implies

$$\begin{aligned} \lim_{n\to\infty} E\left[\xi_n(t) - g(\mu_{\mathbf{x}^n}(t))\right] &= \lim_{n\to\infty} E\left[g_n(\mathbf{X}^n(t)) - \varphi(\mu_{\mathbf{x}^n}(t))\right] \\ &= \lim_{n\to\infty} E\left[\Gamma^n f(\mathbf{X}^n(t)) - \langle f, \mu^m_{\mathbf{x}^n}(t)\rangle\right] = 0 \end{aligned} \tag{5.4.11}$$

and (5.3.6) implies

$$\begin{aligned} \lim_{n\to\infty} E\left[\psi_n(t) - h(\mu_{\mathbf{x}^n}(t)\right] &= \lim_{n\to\infty} E\left[h_n(\mathbf{X}^n(t)) - \mathbb{A}\varphi(\mu_{\mathbf{x}^n}(t))\right] \\ &= \lim_{n\to\infty} E\left[A^n \Gamma^n f(\mathbf{X}^n(t)) - \mathbb{A}\varphi(\mu^m_{\mathbf{x}^n}(t)\right] = 0. \end{aligned} \tag{5.4.12}$$

Noting that $(\mathscr{P}(E), \rho)$, where ρ is Prohorov's metric, is separable and complete since E is (see Lemma 3.2.3), the thesis follows Corollary 3.4.8, given (5.4.5), (5.4.6), (5.4.11), (5.4.12) together with Remark 3.4.7, the relative compactness of $\{\mu_{\mathbf{x}^n}(t), t \geq 0\}$ and the well-posedness of the martingale problem for the generator (5.3.4) (cf. Fleming and Viot, 1979, Section 7 and 8). □

5.5 Stationary distribution

In this section we provide a new proof for an already known result. Yet, the Bayesian construction of the process given in the previous sections not only remarkably simplifies the derivation of the stationary distribution, but allows to stress the connection, also for what concerns the population dynamics, between the neutral diffusion model and Bayesian nonparametrics.

We know from Section 4.2.2 that if the neutral Fleming-Viot process with generator (5.3.4) has mutation

$$Bg(x) = \frac{1}{2}\theta \int \big[g(y) - g(x)\big]\nu_0(\mathrm{d}y) \tag{5.5.1}$$

then the stationary distribution is given by the Dirichlet process of parameters (θ, ν_0), denoted by Π_{θ,ν_0}. We now show that the above construction is coherent with this.

Consider the embedded E^n-valued chain (5.1.9) given by

$$\mathbf{Y}^n_{N(t)} = \mathbf{X}^n(t). \tag{5.5.2}$$

As already observed in Section 5.1, choosing uniformly a component to be replaced with a random sample from (5.1.7), that is from

$$\frac{\theta\nu_0 + \sum_{k\neq i}^n \delta_{x_k}}{\theta + n - 1} \tag{5.5.3}$$

amounts to implement a Gibbs sampler algorithm (see Section 1.3), to the vector $\mathbf{Y}^n = (Y_1, \dots, Y_n)$, where (5.5.3) is the full conditional distribution. By the properties of the Gibbs sampler, the procedure generates a Markov chain on E^n with stationary distribution $\mathcal{P}_n$ given by

$$\mathcal{P}_n = \prod_{i=1}^n \frac{\theta\nu_0 + \sum_{k=1}^{i-1} \delta_{y_k}}{\theta + i - 1}. \tag{5.5.4}$$

Remark 5.5.1 Even before stationarity, if $\mathcal{P}_n$ is the distribution of the vector $\mathbf{Y}^n$, after one transition $\mathbf{Y}^n$ is still distributed according to $\mathcal{P}_n$. This is due to the fact that if $(Y_1, \dots, Y_n) \sim \mathcal{P}_n$, no matter which Y_i is removed, the marginal distribution of $(Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n)$ is $\mathcal{P}_{n-1}$, and since Y_i' is sampled from (5.5.3), then

$$(Y_1, \dots, Y_{i-1}, Y_i', Y_{i+1}, \dots, Y_n) \sim \mathcal{P}_n.$$

Hence, if $\mathcal{P}_n$ is the initial distribution, at every step the sequence $(Y_1, \dots, Y_n)$ is exchangeable with distribution $\mathcal{P}_n$.

It is well known that

$$(Y_1, \dots, Y_n) \sim \int_{\mathscr{P}(E)} \mu^n(\mathrm{d}\cdot)\Pi_{\theta,\nu_0}(\mathrm{d}\mu) \tag{5.5.5}$$

where Π_{θ,ν_0} is the Dirichlet process, μ^n is the m-fold product measure and μ is a random probability measure sampled from Π_{θ,ν_0}. Recall now that the continuous time process, in which the chain is embedded, is constant between consecutive jumps. This implies that, for given n, (5.5.5) holds not only for every step of the chain, that is for every jump time of the continuous time process, but for each $t \geq 0$, and hence that (5.5.5) is the stationary distribution of the E^n-valued continuous time process $\mathbf{X}^n$.

From Theorem 2.1.2 (see also Aldous (1985)), it follows that Π_{θ,ν_0} is the unique distribution of the weak limit of the sequence

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \tag{5.5.6}$$

for $n \to \infty$. Thus Π_{θ,ν_0} is the de Finetti measure of the sequence $(X_1, X_2, \dots)$ at stationarity, but this holds for each $t \geq 0$ if the initial state has distribution $\mathcal{P}_n$ (cf. Remark 5.5.1). Moreover, since the $C_{\mathscr{P}(E)}[0, \infty)$ martingale problem for $\mathbb{A}$ given by

(5.3.4) is well posed (see Fleming and Viot, 1979, Section 7 and 8), from Lemma A.2 (see Appendix) it follows that Π_{θ,ν_0} is the stationary distribution of the process with generator (5.3.4). That is, the Dirichlet process of parameters (θ, ν_0) is the stationary distribution of the neutral diffusion model.

5.6 Discussion

It is known that the Dirichlet process is the unique stationary distribution of the neutral Fleming-Viot process with a certain mutation operator. Donnelly and Kurtz (1996) introduced a countable representation for this type of measure-valued diffusion, which permits to study the genealogical structure of the individuals.

In this chapter we introduced a particle process, based on the Blackwell-MacQueen Pólya-urn scheme, which leads to a slightly different representation than that of Donnelly and Kurtz (1996), and whose associated process of empirical measures weakly converges to the neutral diffusion model, that is the neutral Fleming-Viot. The process provides insight into the dynamics of the individuals and establish a solid connection between urn processes and population genetics. In particular, urn processes succeed in describing coherently the basic mechanisms upon which the evolution of a population of alleles lies, and prove to be powerful but easy-to-handle tools in the hands of the probabilist willing to analyze complex systems as those embraced by population genetics.

The model also shad new light on population genetics processes from a Bayesian nonparametric perspective, providing a construction of the sequence which coherently juxtaposes, via the Blackwell-MacQueen scheme, the role of the individuals with respect to the stationary distribution of the measure-valued diffusion, the latter being the de Finetti measure of the formers. In the next chapter the Bayesian approach toward the neutral diffusion model will be even deepened, showing that one can yield the Fleming-Viot process with a "purely" Bayesian construction.

Chapter 6

Neutral diffusion model via a process of posterior distributions

This chapter introduces a new type of measure-valued process which converges to the neutral Fleming-Viot diffusion, thus embedding the neutral diffusion model in a purely Bayesian framework and generalising the approaches of Donnelly and Kurtz (1996) and Donnelly and Kurtz (1999a), which are recovered as a special case.

So far the literature on Fleming-Viot models has moved from the observation that the state of a population can be represented as a probability distribution, with values on a simplex or on the set of probability measures according to whether there is a countable or uncountable number of types. It has then been natural to think at the empirical measure of the individuals as a sensible and relatively easy-to-handle distribution representing a population. This is in particular the approach taken by the particle processes of Donnelly and Kurtz (1996, 1999a) (see Section 4.3), but also here in the previous chapter (and in the remainder of the work).

The difference between Donnelly and Kurtz (1996, 1999a) and the construction developed in Chapter 5 lies in the fact that in the formers the emphasis of the model was mostly on the empirical measure, whilst in the latter it was on the random distribution function underlying the predictive scheme. It is nonetheless true that also the measure-valued process of Chapter 5 described the empirical measure of the population, and only at the infinite population limit the model exploited its relationship with the de Finetti measure of the sequence. As a consequence, in both cases the Dirichlet process is the stationary distribution of the measure-valued process only in the infinite population limit. The construction of the previous chapter, though, suggests that it could be possible to provide a model whose stationary distribution is the Dirichlet process for every size of the population, deepening the connection between de Finetti measure and individuals, with respect to Fleming-Viot models. This chapter develps this idea.

Starting from the same particle process of Chapter 5, we define a measure-valued process whose value, at every time point, is a representation of a random distribution function sampled from a posterior Dirichlet process, conditional on the particles at the same time point. In particular this can be seen as a mixture of a weighed empirical measure and a random distribution function from a prior Dirichlet process. We thus have again two processes with sample paths in the space of càdlàg functions respectively from $[0, \infty)$ to E^n and from $[0, \infty)$ to $\mathscr{P}(E)$. The measure-valued process is shown to converge to the neutral Fleming-Viot diffusion, and its stationary distribution is the prior Dirichlet process for every size of the population.

6.1 The particle process

For ease of reference, this section summarizes the particle process construction of Sections 5.1 and 5.2.

Define an E^n-valued Markov jump process, with sample paths in the càdlàg space $D_{E^n}[0,\infty)$, as follows. Let $x_1,\ldots,x_n$ be a collection of exchangeable particles, representing a population of size n. Given the initial value of $x_1,\ldots,x_n$, at every transition a particle is randomly selected, with uniform probability, and replaced with a sample of size 1 from the predictive density associated with the Blackwell-MacQueen urn-scheme (cf. (2.2.12)) conditionally on the other $n-1$ particles, whose value is set equal to their previous one until the next transition. That is, if x_i is selected to be replaced, say, then its new value is sampled from

$$p_n(\mathrm{d}x_i|\, x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n) = \frac{\theta\,\nu_0(\mathrm{d}x_i) + \sum_{k\leq n, k\neq i}\delta_{x_k}(\mathrm{d}x_i)}{\theta+n-1} \tag{6.1.1}$$

where θ is a positive real and ν_0 is a non atomic measure on E. The holding times between two consecutive jumps are exponential of parameter

$$\lambda_n = \frac{n(\theta+n-1)}{2}.$$

The infinitesimal generator of this process is given by

$$\begin{aligned} A^n f(\mathbf{x}) &= \sum_{i=1}^n \frac{\lambda_n}{n}\int \big[f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})\big]\frac{\theta\nu_0(\mathrm{d}y) + \sum_{k\neq i}^n \delta_{x_k}(\mathrm{d}y)}{\theta+n-1} \\ &= \sum_{i=1}^n B_i f(\mathbf{x}) + \frac{1}{2}\sum_{1\leq k\neq i\leq n}\big[f(\eta_i(\mathbf{x}|\, x_k)) - f(\mathbf{x})\big] \end{aligned} \tag{6.1.2}$$

where $\eta_i(\mathbf{x}|\, y)$ denotes the vector obtained from $x_1,\ldots,x_n$ by replacing x_i with y, and B_i is the bounded operator B acting on f as a function of x_i, with

$$Bg(x) = \frac{1}{2}\theta\int\big[g(y)-g(x)\big]\nu_0(\mathrm{d}y) \tag{6.1.3}$$

for $g \in B(E)$.

Remark 6.1.1 (a) The Markov property for the particle process follows immediately from the following observation. The transitions of the process describe a Gibbs sampler (see Section 1.3) on the joint density $p_n(x_1, \ldots, x_n)$ of the sequence of x's, (6.1.1) being the full conditional of x_i. Hence the sampler generates a Markov chain which is embedded at the jump times in the E^n-valued process, and $p_n(x_1, \ldots, x_n)$ is its stationary distribution. Since between consecutive jumps the value of the vector is constant, p_n is also the stationary distribution of the continuous time process.

(b) Note also that if the initial value of $x_1, \ldots, x_n$ is generated by the Blackwell-MacQueen urn, then $p_n(x_1, \ldots, x_n)$ is the distribution of the n-size vector for every $t \geq 0$, and not only at stationarity.

6.2 Posterior representation and main result

Consider the E^n-valued particle process defined in the previous section. For fixed $t \geq 0$, we are interested in the random measure sampled from a posterior Dirichlet process conditional on the value of the particles at t, i.e.

$$F_n(t) | \, x_1(t), \ldots, x_n(t) \sim \Pi_{\theta,\nu_0}(\, \cdot \, | \, x_1(t), \ldots, x_n(t)). \tag{6.2.1}$$

Recalling Proposition 2.2.4 and 2.2.5, consider $y_1, y_2, y_3 \stackrel{iid}{\sim} \nu_0$ and, given the partition of E in A_1, A_2, A_3,

$$\delta_{y_1} \sim \mathcal{D}\big(\delta_{y_1}(A_1), \delta_{y_1}(A_2), \delta_{y_1}(A_3)\big)$$
$$\delta_{y_2} \sim \mathcal{D}\big(\delta_{y_2}(A_1), \delta_{y_2}(A_2), \delta_{y_2}(A_3)\big)$$
$$\delta_{y_3} \sim \mathcal{D}\big(\delta_{y_3}(A_1), \delta_{y_3}(A_2), \delta_{y_3}(A_3)\big)$$

and

$$(\tau_1, \tau_2, \tau_3) \sim \mathcal{D}\big(\delta_{y_1}(E), \delta_{y_2}(E), \delta_{y_3}(E)\big) = \mathcal{D}(1,1,1).$$

Then from Proposition 2.2.4 it follows that

$$\tau_1\delta_{y_1} + \tau_2\delta_{y_2} + \tau_3\delta_{y_3} \sim \mathcal{D}\left(\sum_1^3 \delta_{y_i}(A_1), \sum_1^3 \delta_{y_i}(A_2), \sum_1^3 \delta_{y_i}(A_3)\right),$$

i.e. $\sum_i \tau_i \delta_{y_i}$ has a Dirichlet distribution of parameter $\sum_i \delta_{y_i}$. This can obviously be done for every partition of E, so that we can write $F_n(t)$ in (6.2.1) conditional on the x's as

$$F_n(t) = \alpha_n Z_n(t) + (1-\alpha_n)H_n \tag{6.2.2}$$

where

$$Z_n(t) = \sum_{i=1}^n \omega_{i,n}(t)\delta_{x_i(t)} \tag{6.2.3}$$

and the $\omega_{i,n}$'s are weights such that

$$(\omega_{1,n}(t), \dots, \omega_{n,n}(t)) \sim \mathcal{D}(1, \dots, 1) \tag{6.2.4}$$

and also

$$\alpha_n \sim \text{beta}(n, \theta)$$

and

$$H_n \sim \Pi_{\theta,\nu_0}(\,\cdot\,)$$

$\Pi_{\theta,\nu_0}(\,\cdot\,)$ denoting the prior Dirichlet process. In (6.2.2) the vector $(\omega_{1,n}(t),\ldots,\omega_{n,n}(t))$ is independently resampled at every t, while α_n and H_n change only with n. Note that

$$E(\alpha_n) = \frac{n}{n+\theta} \to 1$$

and

$$V(\alpha_n) = \frac{n\,\theta}{(n+\theta)^2(n+1+\theta)} \to 0$$

so that $\alpha_n \to 1$ in probability for $n \to \infty$. Then, recalling from (5.3.4) that the neutral diffusion model has generator

$$\mathbb{A}\varphi(\mu) = \sum_{i=1}^{m}\langle B_i f, \mu^m\rangle + \frac{1}{2}\sum_{1\le k\ne i\le m}\langle \Phi_{ki}f - f, \mu^m\rangle \tag{6.2.5}$$

we claim the following.

Theorem 6.2.1 *Let $\{\mu(t), t \ge 0\}$ denote the neutral Fleming-Viot process with generator* (6.2.5)*, and let $F_n(t)$ be given by* (6.2.2)*. Then $\{F_n(t), t \ge 0\}$ is asymptotically Markov and*

$$\{F_n(t), t \ge 0\} \Rightarrow \{\mu(t), t \ge 0\} \tag{6.2.6}$$

in the space $D_{\mathscr{P}(E)}[0,\infty)$, where $\Rightarrow$ denotes convergence in distribution in the Skorohod topology.

In order to prove Theorem 6.2.1 we need a few intermediate results.

Lemma 6.2.2 *If* (6.2.1) *holds for* $t \geq 0$*, then it holds for* $t + \tau > t$*.*

PROOF. If in the time interval τ no transition occur, it is trivial. Given $x_1(t), \ldots, x_n(t)$, assume the successive transition occurs at $t + \tau$, when x_i is selected to be replaced with z. As seen in section 6.1, z is either $z \sim \nu_0$ or $z \sim \delta_{x_k}$, with $k \neq i$. Then we can write $F_n(t + \tau)$ as

$$\begin{aligned} F_n(t+\tau) &= \alpha_n Z_n(t+\tau) + (1-\alpha_n)H_n \qquad (6.2.7) \\ &= \alpha_n \left(\sum_{j \neq i}^{n} \omega_{j,n}(t+\tau)\delta_{x_j(t+\tau)} + \omega_{i,n}(t+\tau)\delta_z \right) + (1-\alpha_n)H_n \\ &= \alpha_n \left(\sum_{j \neq i}^{n} \omega_{j,n}(t+\tau)\delta_{x_j(t)} + \omega_{i,n}(t+\tau)\delta_z \right) + (1-\alpha_n)H_n \end{aligned}$$

given that when x_i is replaced all other coordinates do not change. Thus $F_n(t+\tau)$ is a random sample from

$$\begin{aligned} &\Pi_{\theta,\nu_0}(\,\cdot \mid x_1(t+\tau), \ldots, x_n(t+\tau)) = \\ &\qquad = \Pi_{\theta,\nu_0}(\,\cdot \mid \eta_i(\mathbf{x}^n(t)|z)) \\ &\qquad = \Pi_{\theta,\nu_0}(\,\cdot \mid x_1(t), \ldots, x_{i-1}(t), z, x_{i+1}(t), \ldots, x_n(t)) \end{aligned}$$

where $\eta_i(\mathbf{x}|z)$ denotes $(x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n)$. □

6.3 Generator of the measure-valued process

Consider the weights defined in (6.2.4). For $m < n$ let $\omega_{j_1,\ldots,j_m,n}(t)$ be a system of weights obtained by defining, for $1 \leq j_1 \neq \cdots \neq j_m \leq n$,

$$\omega_{j_1,\ldots,j_m,n}(t) = \omega_{j_1,n}(t) \frac{\omega_{j_2,n}(t)}{1-\omega_{j_1,n}(t)} \cdots \frac{\omega_{j_m,n}(t)}{1-\sum_{l=1}^{m-1}\omega_{j_l,n}(t)}. \qquad (6.3.1)$$

The weights in (6.3.1) correspond to the probability of picking m elements from an n-dimensional vector without replacement, once at each x_i is assigned a weight ω_i. That is, the probability of picking the first is $\omega_{j_1,n}$, then the weights are normalised, so that the probability of picking the second is $\omega_{j_2,n}/(1-\omega_{j_1,n})$, and so on. If we pick all n elements, the last weight is obviously one. Define also, for $m < n$, the probability measure

$$F_n^{(m)}(t) = \alpha_n Z_n^{(m)}(t) + (1-\alpha_n)H_n^m \tag{6.3.2}$$

where

$$Z_n^{(m)}(t) = \sum_{1\le j_1\neq\cdots\neq j_m\le n} \omega_{j_1,\dots,j_m,n}(t)\delta_{(x_{j_1}(t),\dots,x_{j_m}(t))}$$

and H_n^m denotes the m-fold product measure of H_n. Finally, let $\phi(F_n(t))$ be, for $f \in B(E^m)$,

$$\phi(F_n(t)) = \langle f, F_n^{(m)}(t)\rangle \tag{6.3.3}$$

where $\langle f, \mu\rangle = \int f\mathrm{d}\mu$.

Remark 6.3.1 Note that if the weights are uniform, $\omega_{j_1,\dots,j_m,n}(t)$ simplifies to $1/n_{(m)}$, where

$$n_{(m)} = n(n-1)\dots(n-m+1).$$

Thus (6.3.2) can be seen as a generalisation of the probability measure

$$\mu^{(m)} = \frac{1}{n_{(m)}} \sum_{1\le i_1\neq\cdots\neq i_m\le n} \delta_{(x_{i_1},\dots,x_{i_m})} \tag{6.3.4}$$

used in Donnelly and Kurtz (1999a) (cf. (4.3.16)) to derive the generator of the Fleming-Viot process, which is recovered with uniform weights and α degenerate in

1. Under the same conditions, (6.3.3) can be seen as a generalisation, for $f \in B(E^m)$, of the function

$$\Gamma_m f(\mathbf{x}^n) = \frac{1}{n_{(m)}} \sum_{1 \le i_1 \ne \ldots \ne i_m \le n} f(x_{i_1}, \ldots, x_{i_m}) \tag{6.3.5}$$

used in Donnelly and Kurtz (1996), since

$$\begin{aligned} \langle f, F_n^{(m)}(t) \rangle = \int_{E^m} & f(y_1, \ldots, y_m) \\ & \times \Bigg[\alpha_n \sum_{1 \le i_1 \ne \cdots \ne i_m \le n} \omega_{i_1, \ldots, i_m, n}(t) \delta_{x_{i_1}(t), \ldots, x_{i_m}(t)}(\mathrm{d}y_1, \ldots, \mathrm{d}y_m) \\ & \qquad + (1 - \alpha_n) H_n(\mathrm{d}y_1, \ldots, \mathrm{d}y_m) \Bigg] \end{aligned}$$

which, for $\alpha_n \sim \delta_1$, reduces to

$$\sum_{1 \le i_1 \ne \cdots \ne i_m \le n} \omega_{i_1, \ldots, i_m, n}(t) f(x_{i_1}(t), \ldots, x_{i_m}(t))$$

and with uniform weights to (6.3.5). Note that including replacement in (6.3.1) yields the function $\Gamma f(\mathbf{x}^n)$, obtained by replacing $n_{(m)}$ with n^m in (6.3.5), also used in Donnelly and Kurtz (1996) (cf. (4.3.5)), which is asymptotically equivalent to (6.3.5). □

Lemma 6.3.2 *For $f \in B(E^m)$, the infinitesimal generator of the process $\{F_n(t),\ t \ge 0\}$ is given by*

$$\mathbb{A}^n \phi(F_n) = \sum_{i=1}^{m} \alpha_n \langle B_i f, Z_n^{(m)} \rangle + \frac{1}{2} \sum_{1 \le k \ne i \le m} \alpha_n \left[\langle \Phi_{ki} f, Z_n^{(m)} \rangle - \langle f, Z_n^{(m)} \rangle \right]$$

where B_i *is*

$$Bg(z) = \frac{1}{2}\theta \int [g(y) - g(z)]\, \nu_0(dy)$$

applied to the i*-th coordinate, and*

$$\Phi_{ki} f(x_1, \ldots, x_n) = f(x_1, \ldots, x_{i-1}, x_k, x_{i+1}, \ldots, x_n).$$

PROOF. Denote $\xi \sim \nu_0$. The generator of the measure-valued process can be written

$$\begin{aligned} \mathbb{A}^n \phi(F_n(t)) = \lim_{\tau \downarrow 0} \frac{1}{\tau} \int_{\mathscr{P}(E)} \Big[\phi(F_n(t+\tau)) - \phi(F_n(t))\Big] \Big(1 - e^{-\lambda_n \tau}\Big) \\ \times \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\theta}{\theta+n-1} \Pi_n\Big(\mathrm{d}F_n^{(n)}(t+\tau) | \mathbf{x}_{-i}(t), \xi\Big) \right.\right. \\ \left.\left. + \frac{1}{\theta+n-1} \sum_{k \neq i}^{n} \Pi_n\Big(\mathrm{d}F_n^{(n)}(t+\tau) | \mathbf{x}_{-i}(t), x_k(t)\Big) \right] \right\}. \end{aligned} \tag{6.3.6}$$

Since the probability of having two jumps in $[t, t+\tau]$ is $o(\tau)$, denoting with a prime a variable computed in $t+\tau$ when $\tau \downarrow 0$, we get

$$\begin{aligned} \mathbb{A}^n \phi(F_n) = & \sum_{i=1}^{n} \frac{\theta\, \lambda_n}{n(\theta+n-1)} \int_{\mathscr{P}(E)} \Big[\langle f, F_n^{\prime(n)} \rangle - \langle f, F_n^{(n)} \rangle\Big]\, \Pi_n(\mathrm{d}F_n^{\prime(n)} | \mathbf{x}_{-i}, \xi) \\ & + \sum_{1 \le k \neq i \le n} \frac{\lambda_n}{n(\theta+n-1)} \\ & \quad \times \int_{\mathscr{P}(E)} \Big[\langle f, F_n^{\prime(n)} \rangle - \langle f, F_n^{(n)} \rangle\Big]\, \Pi_n(\mathrm{d}F_n^{\prime(n)} | \mathbf{x}_{-i}, x_k) \\ = & \sum_{i=1}^{n} \frac{\theta\, \lambda_n}{n(\theta+n-1)} \\ & \quad \times \int_E \int_{\mathscr{P}(E)} \Big[\langle f, F_n^{\prime(n)} \rangle - \langle f, F_n^{(n)} \rangle\Big]\, \Pi_n(\mathrm{d}F_n^{\prime(n)} | \mathbf{x}_{-i}, z) \nu_0(\mathrm{d}z) \\ & + \sum_{1 \le k \neq i \le n} \frac{\lambda_n}{n(\theta+n-1)} \\ & \quad \times \int_E \int_{\mathscr{P}(E)} \Big[\langle f, F_n^{\prime(n)} \rangle - \langle f, F_n^{(n)} \rangle\Big]\, \Pi_n(\mathrm{d}F_n^{\prime(n)} | \mathbf{x}_{-i}, z) \delta_{x_k}(\mathrm{d}z). \end{aligned}$$

Recalling that $\lambda_n = 2^{-1}n(\theta + n - 1)$ and using (6.2.2) and (6.2.7), we can write $\mathbb{A}^n\phi(F_n)$ as

$$\begin{aligned}
&\sum_{i=1}^{n} \frac{1}{2}\theta \int_E \left[\langle f, \alpha_n Z_n^{\prime(n)} + (1-\alpha_n)H_n^n\rangle - \langle f, \alpha_n Z_n^{(n)} + (1-\alpha_n)H_n^n\rangle\right] \nu_0(\mathrm{d}z) \qquad (6.3.7)\\
&+ \frac{1}{2} \sum_{1\le k\ne i\le n}\\
&\quad\times \int_E \left[\langle f, \alpha_n Z_n^{\prime(n)} + (1-\alpha_n)H_n^n\rangle - \langle f, \alpha_n Z_n^{(n)} + (1-\alpha_n)H_n^n\rangle\right] \delta_{x_k}(\mathrm{d}z)\\
&= \sum_{i=1}^{n} \left[\frac{1}{2}\theta \int_E \alpha_n \langle f, Z_n^{\prime(n)}\rangle \nu_0(\mathrm{d}z) - \alpha_n\langle f, Z_n^{(n)}\rangle\right]\\
&+ \frac{1}{2} \sum_{1\le k\ne i\le n} \left[\int_E \alpha_n\langle f, Z_n^{\prime(n)}\rangle \delta_{x_k}(\mathrm{d}z) - \alpha_n\langle f, Z_n^{(n)}\rangle\right]
\end{aligned}$$

Now, if P_i is $Pg(y) = \frac{1}{2}\theta \int g(y)\nu_0(\mathrm{d}y)$ applied to the i-th coordinate,

$$\begin{aligned}
\frac{1}{2}\theta \int_E \langle f, Z_n^{\prime(n)}\rangle \nu_0(\mathrm{d}z) &= \frac{1}{2}\theta \int_{E^{n+1}} f(y_1,\dots,y_n)\left[Z_n^{\prime(n)}(\mathrm{d}y_1,\dots,\mathrm{d}y_n)\right]\nu_0(\mathrm{d}z) \qquad (6.3.8)\\
&= \frac{1}{2}\theta \int_{E^{n+1}} f(y_1,\dots,y_n)\\
&\quad\times \left[\sum_{1\le j_1\ne\cdots\ne j_n\le n} \omega'_{j_1,\dots,j_n,n}\delta_{(x_{j_1},\dots,x_{j_{i-1}},z,x_{j_{i+1}},\dots,x_{j_n})}(\mathrm{d}y_1,\dots,\mathrm{d}y_n)\right]\nu_0(\mathrm{d}z)\\
&= \frac{1}{2}\theta \int_E \sum_{1\le j_1\ne\cdots\ne j_n\le n} \omega'_{j_1,\dots,j_n,n} f(x_{j_1},\dots,x_{j_{i-1}},z,x_{j_{i+1}},\dots,x_{j_n})\nu_0(\mathrm{d}z)\\
&= \sum_{1\le j_1\ne\cdots\ne j_n\le n} \omega'_{j_1,\dots,j_n,n} P_{j_i} f(x_{j_1},\dots,x_{j_n})\\
&= \int_{E^n} P_i f(y_1,\dots,y_n)\left[\sum_{1\le j_1\ne\cdots\ne j_n\le n} \omega'_{j_1,\dots,j_n,n}\delta_{(x_{j_1},\dots,x_{j_n})}(\mathrm{d}y_1,\dots,\mathrm{d}y_n)\right]\\
&= \langle P_i f, Z_n^{(n)}\rangle
\end{aligned}$$

Analogously,

$$\begin{aligned}
\int_E \langle f, Z_n^{\prime(n)} \rangle \delta_{x_k}(\mathrm{d}z) &= \int_{E^{n+1}} f(y_1,\dots,y_n) \left[Z_n^{\prime(n)}(\mathrm{d}y_1,\dots,\mathrm{d}y_n) \right] \delta_{x_k}(\mathrm{d}z) \qquad (6.3.9)\\
&= \int_{E^{n+1}} f(y_1,\dots,y_n) \\
&\quad \times \left[\sum_{1\le j_1\neq\cdots\neq j_n\le n} \omega'_{j_1,\dots,j_n,n} \delta_{(x_{j_1},\dots,x_{j_{i-1}},z,x_{j_{i+1}},\dots,x_{j_n})}(\mathrm{d}y_1,\dots,\mathrm{d}y_n) \right] \delta_{x_{j_k}}(\mathrm{d}z)\\
&= \int_E \sum_{1\le j_1\neq\cdots\neq j_n\le n} \omega'_{j_1,\dots,j_n,n} f(x_{j_1},\dots,x_{j_{i-1}},z,x_{j_{i+1}},\dots,x_{j_n}) \delta_{x_{j_k}}(\mathrm{d}z)\\
&= \sum_{1\le j_1\neq\cdots\neq j_n\le n} \omega'_{j_1,\dots,j_n,n} \Phi_{j_k j_i} f(x_{j_1},\dots,x_{j_n})\\
&= \int_{E^n} \Phi_{ki} f(y_1,\dots,y_n) \left[\sum_{1\le j_1\neq\cdots\neq j_n\le n} \omega'_{j_1,\dots,j_n,n} \delta_{(x_{j_1},\dots,x_{j_n})}(\mathrm{d}y_1,\dots,\mathrm{d}y_n) \right]\\
&= \langle \Phi_{ki} f, Z_n^{(n)} \rangle
\end{aligned}$$

where Φ_{ki} denotes the function obtained by replacing the i-th with the k-th variable in f. Using (6.3.8) and (6.3.9), (6.3.7) becomes

$$\sum_{i=1}^n \alpha_n \left[\langle P_i f, Z_n^{(n)} \rangle - \langle f, Z_n^{(n)} \rangle \right] + \frac{1}{2} \sum_{1\le k\neq i\le n} \alpha_n \left[\langle \Phi_{ki} f, Z_n^{(n)} \rangle - \langle f, Z_n^{(n)} \rangle \right].$$

Note that for $f \in B(E^m)$, $m < n$, when x_i is not an argument of f we have $P_i f = f$ and $\Phi_{ki} f = f$, so that

$$\sum_{i=m+1}^n \alpha_n \left[\langle P_i f, Z_n^{(n)} \rangle - \langle f, Z_n^{(n)} \rangle \right] = 0$$

and

$$\sum_{i=m+1}^n \sum_{k=1,k\neq i}^n \alpha_n \left[\langle \Phi_{ki} f, Z_n^{(n)} \rangle - \langle f, Z_n^{(n)} \rangle \right] = 0.$$

Further, when x_i is an argument of f but x_k is not, we have $\langle \Phi_{ki} f, \mu \rangle = \langle f, \mu \rangle$, so that

$$\sum_{i=1}^m \sum_{k=m+1}^n \alpha_n \left[\langle \Phi_{ki} f, Z_n^{(n)} \rangle - \langle f, Z_n^{(n)} \rangle \right] = 0.$$

Hence, for $f \in B(E^m)$, we have

$$\mathbb{A}^n \phi(F_n) = \sum_{i=1}^{m} \alpha_n \langle B_i f, Z_n^{(m)} \rangle + \frac{1}{2} \sum_{1 \le k \ne i \le m} \alpha_n \left[\langle \Phi_{ki} f, Z_n^{(m)} \rangle - \langle f, Z_n^{(m)} \rangle \right] \tag{6.3.10}$$

where B_i is

$$Bg(z) = \frac{1}{2}\theta \int [g(y) - g(z)]\, \nu_0(\mathrm{d}y)$$

applied to the i-th coordinate. □

Lemma 6.3.3 *The process $\{F_n(t),\, t \ge 0\}$ has stationary distribution $\Pi_{\theta,\nu_0}(\,\cdot\,)$.*

PROOF. From Lemma 6.2.2 it follows that (6.2.1) holds for every $t \ge 0$. We can find the stationary distribution of $F_n(t)$ from (6.2.1) by integrating out the x's, that is

$$F_n(t) \sim \int_{E^n} \Pi_{\theta,\nu_0}(\,\cdot \mid x_1(t), \ldots, x_n(t)) \nu_0(\mathrm{d}x_1) \ldots \nu_0(\mathrm{d}x_n).$$

From Corollary 1.1 of Antoniak (1974) it follows that $F_n(t) \sim \Pi_{\theta,\nu_0}(\,\cdot\,)$. □

We are now able to prove the result given in Theorem 6.2.1.

PROOF OF THEOREM 6.2.1. Denote $\mu_n = n^{-1} \sum_i \delta_{x_i}$ and call μ its limit for $n \to \infty$. We know that (cf. also Donnelly and Kurtz, 1996)

$$\begin{aligned}
&\sup_{\mathbf{x}^n \in E^n} \left| \frac{n_{(m)}}{n^m} \langle f, \mu_n^{(m)} \rangle - \langle f, \mu_n^m \rangle \right| \qquad (6.3.11)\\
&= \sup_{\mathbf{x}^n \in E^n} \left| \int_{E^m} f(y_1, \ldots, y_m) \frac{1}{n^m} \sum_{1 \le i_1 \ne \cdots \ne i_m \le n} \delta_{x_{i_1}, \ldots, x_{i_m}}(\mathrm{d}y_1, \ldots, \mathrm{d}y_m) \right. \\
&\qquad \left. - \int_{E^m} f(y_1, \ldots, y_m) \frac{1}{n^m} \sum_{1 \le i_1, \ldots, i_m \le n} \delta_{x_{i_1}, \ldots, x_{i_m}}(\mathrm{d}y_1, \ldots, \mathrm{d}y_m) \right| \\
&= \sup_{\mathbf{x}^n \in E^n} \frac{1}{n^m} \left| \sum_{1 \le i_1 \ne \cdots \ne i_m \le n} f(x_{i_1}, \ldots, x_{i_m}) - \sum_{1 \le i_1, \ldots, i_m \le n} f(x_{i_1}, \ldots, x_{i_m}) \right| \to 0
\end{aligned}$$

for $n \to \infty$, where $\mu_n^{(m)}$ is (6.3.4). Recalling (6.3.1), observe now that $\omega_{i_1,\ldots,i_m,n}$ converges to $1/n_{(m)}$ in probability, since $E(\omega_i) = 1/n$ and

$$Var(\omega_i) = \frac{n-1}{n^2(n+1)}$$

from which it follows that (6.3.11) holds also replacing $\mu_n^{(m)}$ with $Z_n^{(m)}$. Hence, for large n,

$$\sup_{\mathbf{x}^n \in E^n} \left| \langle f, Z_n^{(m)} \rangle - \langle f, \mu_n^m \rangle \right| \to 0 \tag{6.3.12}$$

and recall that α_n converges to 1 (see Section 6.2). From (6.3.12) and Lemma 6.3.2 it follows that

$$\sup_{\mathbf{x}^n \in E^n} |\mathbb{A}^n \phi(F_n) - \mathbb{A}\varphi(\mu)| \to 0$$

where $\mathbb{A}\varphi(\mu)$ is

$$\mathbb{A}\varphi(\mu) = \sum_{i=1}^{m} \langle B_i f, \mu^m \rangle + \frac{1}{2} \sum_{1 \le k \ne i \le m} \left[\langle \Phi_{ki} f, \mu^{m-1} \rangle - \langle f, \mu^m \rangle \right] \tag{6.3.13}$$

which is the generator of the neutral Fleming-Viot process as in (6.2.5). The above computation implies (6.2.6) (cf. Section 4.3.1).

The asymptotical Markovianity of $\{F_n(t),\, t \ge 0\}$ follows from the convergence of the generator, together with the well-posedness of the martingale problem for (6.3.13) (see Theorem 4.2.1), which in turn implies the Markovianity of the unique solution (see Theorem 1.5.3). □

Corollary 6.3.4 *The neutral diffusion model with generator* (6.3.13) (*or* (6.2.5)) *has stationary distribution given by the Dirichlet process* $\Pi_{\theta,\nu_0}(\,\cdot\,)$.

PROOF. It follows immediately from Theorem 6.2.1 and Lemma 6.3.3. □

Remark 6.3.5 (a) An alternative proof of Corollary 6.3.4 follows from the fact that the weak limit of $n^{-1}\sum_i \delta_{x_i}$ is a purely atomic measure having distribution given by the de Finetti measure of the sequence of x's, which is the Dirichlet process $\Pi_{\theta,\nu_0}(\,\cdot\,)$ (see Aldous (1985)). Observing that the martingale problem for (6.3.13) is well-posed, then the thesis follows from Lemma A.2 (see Appendix).

(b) In the special case of uniform weights in (6.2.4) and α_n degenerate in 1, $\{F_n(t),\, t \geq 0\}$ reduces to the process of empirical measures of the particles, in the sense of the construction in Donnelly and Kurtz (1996) and Donnelly and Kurtz (1999a) (cf. also Remark 6.3.1 above). In this case, as we know from Ethier and Kurtz (1994), the argument in Corollary 6.3.4 still works, following from part (a) of this remark.

It seems however more natural to have a process whose stationary distribution in the n-sized population case is the same as its limit for an infinite population, which is the case of the present construction (cf. Lemma 6.3.3) but does not hold for the n-sized empirical measure. This could be useful when an approximation of the limiting process for a finite population is needed, in that working with the same distribution has self-evident advantages.

6.4 Discussion

This chapter introduces a measure-valued process, based on an E^n-valued Markov particle process, which is different from the approach on the matter in the existent literature. Instead of considering the process of empirical measures of the n particles, we exploits a representation for a random distribution function sampled from a posterior Dirichlet process conditional on the particles, described as a mixture of a weighted empirical measure and a random distribution function sampled from a prior Dirichlet process. The process based on such mixture is shown to converge to the neutral diffusion model. Also, its stationary distribution is the Dirichlet process for every size of the population, the n-sized population case having therefore the same stationary distribution as its limit for large n.

The construction, which embeds the neutral diffusion model in a purely Bayesian framework, constitutes a generalisation of the previous approaches (cf. Remark 6.3.1 and 6.3.5), which are recovered as a special case.

Chapter 7

Viability selection I: Markov case

In Section 4.3.2 we have seen that Donnelly and Kurtz (1999a) give a countable representation of the Fleming-Viot process with viability selection. They show that the stationary distribution of such Fleming-Viot process is the de Finetti measure of the individuals, and also that it is absolutely continuous with respect to the stationary distribution of the neutral diffusion model, that is with respect to the Dirichlet process (see Section 5.5 or Section 4.2.2).

In this chapter we attempt to provide an explicit construction, similar to that of Chapter 5, of the Fleming-Viot process with viability selection. We proceed by defining an E^n-valued Markov jump process, based on the urn-scheme which characterises the two-parameter Poisson-Dirichlet process, and compute the generator of the associated process of empirical measures. The generators of both processes turn out to be potentially unbounded, preventing from proving any asymptotic result concerning existence or convergence.

7.1 The particle process

Consider, in the setting of Section 5.1, a pure jump process $\mathbf{X}^n = (X_1^n, \ldots, X_n^n)$ on E^n, for $n \geq 1$, where now the holding times are no longer driven by a single Poisson process but by a collection of Poisson processes of intensity $\lambda_{n,l,i} = \lambda_n(x_l, x_i, \mu_n)$, $1 \leq l \neq i \leq n$. The transitions are as follows. Instantaneously after the $(m-1)$th renewal, $m \geq 1$, with probability $\pi_i^{\sigma,n} = \pi^{\sigma,n}(x_i, \mu_n)$ a component i is chosen to be replaced (a particle dies in the population); given i, an l different from i is sampled with probability $p_{l,i}^n = p^n(x_l, x_i, \mu_n)$, and a holding time S_m is sampled from an exponential distribution of parameter $\lambda_{n,l,i}$. That is, once i is chosen, the level $l \neq i$ determines the parameter of the exponential from which the successive holding time is drawn. Note that in the above definitions x_l, x_i and the empirical measure μ_n are all values referred to the starting state, which ensures that the process be Markov. Observe also that the level l which determines the parameter $\lambda_{n,l,i}$ has nothing to do with the level that will be duplicated to replace the dying particle x_i.

At the following renewal time the process is set to be

$$X_k(t_m) = x_{k,t_m^-} \qquad k \neq i \tag{7.1.1}$$

where $x_{k,t_m^-} = x_k(t_m^-)$, and

$$X_i(t_m)|i, \mathbf{x}_{-i,t_m} \sim \frac{\theta + \sigma k_i}{\theta + n - 1}\,\nu_0 + \frac{1}{\theta + n - 1}\sum_{j=1}^{k_i}(c_{j^*} - \sigma)\delta_{x_j^*} \tag{7.1.2}$$

where, as in Section 2.1, k_i denotes the number of unique values x_j^* in $\mathbf{x}_{-i} = \{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\}$, and c_{j^*} is the cardinality of the set of x_h such that $x_h = x_j^*$, where $h = 1, \ldots, i-1, i+1, \ldots, n$. The density in (7.1.2) is the predictive rule that generates a sequence of random variables which are i.i.d. from a random distribution function sampled from a two-parameter Poisson-Dirichlet process, denoted by $\mathscr{P}\mathscr{D}(\sigma,\theta)$. See Section 2.3.

Proceeding as in 5.2, the generator for this process is

$$\begin{aligned}
A^n f(\mathbf{x}) &= \sum_{1\le l\ne i\le n} \pi_i^{\sigma,n}\, p_{l,i}^n\, \lambda_{n,l,i} \int \big[f(\eta_i(\mathbf{x}|\,y)) - f(\mathbf{x})\big] \\
&\quad \times \left(\frac{\theta+\sigma k_i}{\theta+n-1}\,\nu_0(\mathrm{d}y) + \frac{1}{\theta+n-1}\sum_{j=1}^{k_i}(c_{j^*}-\sigma)\delta_{x_j^*}(\mathrm{d}y)\right) \\
&= \sum_{1\le l\ne i\le n} \pi_i^{\sigma,n}\, p_{l,i}^n\, \lambda_{n,l,i}\frac{\theta+\sigma k_i}{\theta+n-1} \int \big[f(\eta_i(\mathbf{x}|\,y)) - f(\mathbf{x})\big]\nu_0(\mathrm{d}y) \\
&\quad + \sum_{1\le l\ne i\le n}\sum_{j=1}^{k_i} \pi_i^{\sigma,n}\, p_{l,i}^n\, \lambda_{n,l,i}\frac{c_{j^*}-\sigma}{\theta+n-1}\big[f(\eta_i(\mathbf{x}|\,x_j^*)) - f(\mathbf{x})\big] \\
&= \sum_{i=1}^n\sum_{l\ne i}^n \pi_i^{\sigma,n}\, p_{l,i}^n\, \lambda_{n,l,i}\frac{\theta+\sigma k_i}{\theta+n-1} \int \big[f(\eta_i(\mathbf{x}|\,y)) - f(\mathbf{x})\big]\nu_0(\mathrm{d}y) \\
&\quad + \sum_{i=1}^n\sum_{l\ne i}^n\sum_{h\ne i}^n \pi_i^{\sigma,n}\, p_{l,i}^n\, \lambda_{n,l,i}(\theta+n-1)^{-1}\left(1-\frac{\sigma}{c_h}\right)\big[f(\eta_i(\mathbf{x}|\,x_h)) - f(\mathbf{x})\big]
\end{aligned} \tag{7.1.3}$$

where in the third equality we used the fact that

$$\sum_{j=1}^{k_i} 1 = \sum_{\substack{1\le h\le n\\ h\ne i}} \frac{1}{c_h}\,. \tag{7.1.4}$$

where c_h is the cardinality defined above. Note that we have c_h instead of c_{h^*} because the sum is taken over all values in the sequence, and not only over the unique values. Then (7.1.4) is implied by the trivial fact that $\sum_l 1/c_h = 1$, where the sum is taken over all l such that $x_l = x_h$.

Let now the probability of choosing a particle at level i be

$$\pi_i^{\sigma,n} = \frac{1}{n}\left(1+\frac{2\sigma\tilde\beta(x_i,\mu_n)}{n}\right) \tag{7.1.5}$$

where $\sigma > 0$ is the parameter in (7.1.2), $\tilde{\beta}_i = \tilde{\beta}(x_i, \mu_n)$ is such that[1]

$$\begin{cases} \tilde{\beta}_i \in \left(-\frac{n}{2\sigma}, \frac{n^2}{2\sigma} - \frac{n}{2\sigma}\right) \\ \sum_i \tilde{\beta}_i = 0 \end{cases} \tag{7.1.6}$$

and

$$\begin{aligned} \sum_{i=1}^{n} \pi_i^{\sigma,n} &= \sum_{i=1}^{n} \frac{1}{n}\left(1 + \frac{2\sigma\tilde{\beta}(x_i, \mu_n)}{n}\right) \\ &= \sum_{i=1}^{n} \frac{1}{n} + \frac{2\sigma}{n} \sum_{i=1}^{n} \tilde{\beta}(x_i, \mu_n) = 1. \end{aligned} \tag{7.1.7}$$

Given that i is the level at which the particle dies, let then the probability of sampling a holding time from an exponential distribution of parameter $\lambda_{n,l,i}$ be

$$p_{l,i}^n = (n - 1 - \sigma k_i)^{-1} \left(1 - \frac{\sigma}{c_l}\right) \tag{7.1.8}$$

such that

$$\begin{aligned} \sum_{l \neq i}^{n} p_{l,i}^n &= \sum_{l \neq i}^{n} (n - 1 - \sigma k_i)^{-1} \left(1 - \frac{\sigma}{c_l}\right) \\ &= (n - 1 - \sigma k_i)^{-1} \left(\sum_{l \neq i}^{n} - \sum_{l \neq i}^{n} \frac{\sigma}{c_l}\right) \\ &= (n - 1 - \sigma k_i)^{-1} \left(n - 1 - \sum_{j=1}^{k_i} \sigma\right) \\ &= (n - 1 - \sigma k_i)^{-1} (n - 1 - \sigma k_i) = 1 \end{aligned}$$

where in the third equality we used again (7.1.4). Finally, for $1 \leq l \neq i \leq n$, let the Poisson rates be

$$\lambda_{n,l,i} = \left(1 - \frac{\sigma}{c_l}\right)^{-1} \frac{n(\theta + n - 1)}{2}. \tag{7.1.9}$$

[1] For a more detailed discussion of the properties of $\tilde{\beta}(x_i, \mu_n)$ see Section 7.3 below.

From (7.1.3) we obtain

$$
\begin{aligned}
A^n f(\mathbf{x}) = {} & \sum_{i=1}^{n}\sum_{l\neq i}^{n}\frac{1}{n}\Bigg\{\left(1+\frac{2\sigma\tilde{\beta}(x_i,\mu_n)}{n}\right)(n-1-\sigma k_i)^{-1}\left(1-\frac{\sigma}{c_l}\right)\\
& \times\left(1-\frac{\sigma}{c_l}\right)^{-1}\frac{n(\theta+n-1)}{2}\,\frac{\theta+\sigma k_i}{\theta+n-1}\int\big[f(\eta_i(\mathbf{x}|\,y))-f(\mathbf{x})\big]\nu_0(\mathrm{d}y)\Bigg\}\\
& +\sum_{i=1}^{n}\sum_{l\neq i}^{n}\sum_{h\neq i}^{n}\frac{1}{n}\Bigg\{\left(1+\frac{2\sigma\tilde{\beta}(x_i,\mu_n)}{n}\right)(n-1-\sigma k_i)^{-1}\\
& \times\left(1-\frac{\sigma}{c_l}\right)\left(1-\frac{\sigma}{c_l}\right)^{-1}\frac{n(\theta+n-1)}{2}(\theta+n-1)^{-1}\\
& \times\left(1-\frac{\sigma}{c_h}\right)\big[f(\eta_i(\mathbf{x}|\,x_h))-f(\mathbf{x})\big]\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
= {} & \sum_{i=1}^{n}\sum_{l\neq i}^{n}\frac{1}{2}\Bigg\{\left(1+\frac{2\sigma\tilde{\beta}(x_i,\mu_n)}{n}\right)(n-1-\sigma k_i)^{-1}\\
& \times(\theta+\sigma k_i)\int\big[f(\eta_i(\mathbf{x}|\,y))-f(\mathbf{x})\big]\nu_0(\mathrm{d}y)\Bigg\}\\
& +\sum_{i=1}^{n}\sum_{l\neq i}^{n}\sum_{h\neq i}^{n}\frac{1}{2}\Bigg\{\left(1+\frac{2\sigma\tilde{\beta}(x_i,\mu_n)}{n}\right)(n-1-\sigma k_i)^{-1}\\
& \times\left(1-\frac{\sigma}{c_h}\right)\big[f(\eta_i(\mathbf{x}|\,x_h))-f(\mathbf{x})\big]\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
= {} & \sum_{i=1}^{n}\sum_{l\neq i}^{n}\frac{1}{2}(\theta+\sigma k_i)\left(1+\frac{2\sigma\tilde{\beta}(x_i,\mu_n)}{n}\right)\frac{1}{n-1-\sigma k_i}\\
& \times\int\big[f(\eta_i(\mathbf{x}|\,y))-f(\mathbf{x})\big]\nu_0(\mathrm{d}y)\\
& +\sum_{i=1}^{n}\sum_{h\neq i}^{n}\sum_{h\neq i}^{n}\frac{1}{2}\Bigg\{\left(1+\frac{2\sigma\tilde{\beta}(x_i,\mu_n)}{n}\right)(n-1-\sigma k_i)^{-1}\\
& \times\left(1-\frac{\sigma}{c_h}\right)\big[f(\eta_i(\mathbf{x}|\,x_h))-f(\mathbf{x})\big]\Bigg\}
\end{aligned}
$$

$$= \sum_{i=1}^{n} \frac{1}{2}(\theta + \sigma k_i) \left(1 + \frac{2\sigma\tilde{\beta}(x_i, \mu_n)}{n}\right) \frac{n-1}{n-1-\sigma k_i}$$
$$\times \int \big[f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})\big]\nu_0(\mathrm{d}y)$$
$$+ \sum_{i=1}^{n}\sum_{h\neq i}^{n} \frac{1}{2}\left(1 + \frac{2\sigma\tilde{\beta}(x_i, \mu_n)}{n}\right)\big[f(\eta_i(\mathbf{x}|\, x_h)) - f(\mathbf{x})\big] \tag{7.1.10}$$

where in the third equality we used the fact that, nothing depending on l, we can switch the sum on l with a sum on h, and in the last equality the fact that $\sum_{l\neq i} 1 = n-1$ and

$$\sum_{h\neq i}^{n} (n-1-\sigma k_i)^{-1}\left(1 - \frac{\sigma}{c_h}\right) = 1. \tag{7.1.11}$$

Hence the generator (7.1.10) of the E^n-valued particle process can be written

$$A^n f(\mathbf{x}) = \sum_{i=1}^{n} B_i^{\gamma} f(\mathbf{x}) + \frac{1}{2}\sum_{1\leq l\neq i\leq n}\left(1 + \frac{2\beta(x_i, \mu_n)}{n}\right)\big[f(\eta_i(\mathbf{x}|\, x_h)) - f(\mathbf{x})\big] \tag{7.1.12}$$

where the mutation operator is

$$B^{\gamma} g(x) = \gamma(x, \mu_n) \int \big[f(y)) - f(x)\big]\nu_0(\mathrm{d}y). \tag{7.1.13}$$

The coefficient of the mutation operator is

$$\gamma(x, \mu_n) = \frac{1}{2}(\theta + \sigma k_x)\left(1 + \frac{2\beta(x, \mu_n)}{n}\right)\frac{n-1}{n-1-\sigma k_x} \tag{7.1.14}$$

where

$$\beta(x, \mu_n) = \sigma\,\tilde{\beta}(x, \mu_n) \tag{7.1.15}$$

and k_x denotes the generic for k_i. Note that (7.1.13) is not necessarily bounded.

Note in (7.1.3), or in the first expressions of (7.1.10), that the particle at level l which determines the exponential parameter $\lambda_{n,l,i}$ is, as already observed, sampled from the subvector $\mathbf{x}_{-i}$ immediately after the previous transition, and has nothing to do with the particle at level h that figures in c_h and in $\eta_i(\mathbf{x}|\, x_h)$, and that is duplicated to replace the particle at level i. The particle at level h is indeed part of the predictive density (7.1.2), and thus of the arrival state. Hence $p_{l,i}^n$ and $\lambda_{n,l,i}$, which are functions of x_i and x_l, but not of x_h, depend only on the starting state and the process is Markov. Thus the main difference between this setting and that of Section 5.1, besides obviously the transition density, is that now the process of holding times is a vector of non homogeneous Poisson processes, whereas for the neutral particle process we had a single homogeneous Poisson process.

Observe also that for $\sigma = 0$, $\pi_i^{\sigma,n}$ in (7.1.5) reduces to $\pi_i^n = n^{-1}$ in (5.2.3), $\lambda_{n,l,i}$ in (7.1.9) reduces to $\lambda_n = 2^{-1}n(\theta + n - 1)$ in (5.2.5), $p_{l,i}^n$ in (7.1.8) reduces to $(n-1)^{-1}$, $\beta \equiv 0$ and γ becomes $\theta/2$, simplifying B^γ in (7.1.13) to B in (5.2.8), so that we recover the parameters of the neutral case; that is, the level i of the particle to be replaced is selected with uniform probability, the holding times are sampled from an exponential distribution of parameter $2^{-1}n(\theta + n - 1)$ with probability one, and (7.1.12) simplifies to (5.2.7). This is coherent with the fact that for $\sigma = 0$ the two-parameter Poisson-Dirichlet process $\mathscr{PD}(\sigma, \theta)$ reduces to the Dirichlet process Π_{θ,ν_0}, which is the de Finetti measure of the infinite exchangeable sequence (at stationarity) that form the particle process with generator (5.2.7).

The main difference between (7.1.12) and the generator of the neutral model (5.2.7) lies in two aspects of the same nature: mutation and viability (the capability of living) are no longer independent of the type of the individual. This can be seen from the term $\gamma(\cdot, \mu_n)$, which alters the uniformity in the chances of the individuals of mutating, i.e. of being selected to be replaced with a sample from ν_0; and from $\beta(\cdot, \mu_n)$, which alters the uniformity in the chances of the individuals of dying, i.e. of being selected to be replaced with the copy of a particle from a different level. It is now clear why the model is no longer neutral, that is no longer independent of the type of the individuals, and presents features of viability selection, i.e. mechanisms, like the selection function β, which alter the viability of the individuals.

In particular, the larger $\beta(x_i, \mu_n)$ is, the more likely the i-th particle is to die and be replaced by a copy of a randomly selected particle, so large β reduces the viability of an individual. In population genetics terms, this corresponds to low fitness. Note also that $\beta(x_i, \mu_n)$ is governed by two parameters: $\tilde{\beta}(x_i, \mu_n)$, which determines the weight associated with the particle at level i, and σ, the constant of the two-parameter Poisson-Dirichlet process, which determines the weight of the selection function with respect to the uniform sampling from the empirical distribution. This is somewhat coherent with the role of σ in the predictive distribution of the urn associated with the two-parameter Poisson-Dirichlet process, given by (7.1.2); in the prediction rule, indeed, σ determines the balance between uniform and non-uniform sampling from the empirical measure, acting as a counterweight on the whole cluster formed by individuals of the same type.

7.2 The associated measure-valued process

In this section we derive the generator of the $\mathscr{P}(E)$-valued process of the empirical measures of $\mathbf{X}^n$, and show that it has the form of the generator of the Fleming-Viot process with selection.

For $m < n$, let $\mu^{(m)}$ be the probability measure on E^m defined by

$$\mu^{(m)} = \frac{1}{n(n-1)\dots(n-m+1)} \sum \delta_{(x_{i_1},\dots,x_{i_m})} \tag{7.2.1}$$

where the sum is taken over all choices of $1 \le i_1, \dots, i_m \le n$ with $i_k \ne i_l$. Note that for $\mu \in \mathscr{P}^n(E)$, where $\mathscr{P}^n(E) \subset \mathscr{P}(E)$ is the set of atomic probability measures on E with masses proportional to $1/n$, $\mu^{(m)}$ depends only on μ, not on the particular choice of $\mathbf{x}$. For $f \in B(E^n)$, define $\phi \in B(\mathscr{P}^n(E))$ by

$$\phi(\mu) = \langle f, \mu^{(n)} \rangle \tag{7.2.2}$$

and $\mathbb{A}^n\phi$ by

$$\mathbb{A}^n\phi(\mu) = \langle A^n f, \mu^{(n)} \rangle; \tag{7.2.3}$$

see Section 4.3.2. Denote now with Φ_{ki} the function of $n-1$ variables obtained by setting the i-th variable in f equal to the k-th, that is

$$\Phi_{ki} f(x_1, \dots, x_n) = f(x_1, \dots, x_{i-1}, x_k, x_{i+1}, \dots, x_n) \tag{7.2.4}$$

(note that a function g of $n-1$ variables can be viewed as a function of n variables and $\langle g, \mu^n \rangle = \langle g, \mu^{n-1} \rangle$). From (7.1.12) and (7.2.3) we obtain, for $\mu \in \mathscr{P}^n(E)$,

$$\begin{aligned}
\mathbb{A}^n \phi(\mu) =& \int_{E^n} \sum_{i=1}^n B_i^\gamma f d\mu^{(n)} + \int_{E^n} \frac{1}{2} \sum_{1\le k\ne i\le n} (\Phi_{ki} f - f) d\mu^{(n)} \\
&+ \int_{E^n} \sum_{1\le k\ne i\le n} \frac{1}{n} \beta(x_i, \mu)(\Phi_{ki} f - f) d\mu^{(n)} \\
=& \sum_{i=1}^n \int_{E^n} B_i^\gamma f d\mu^{(n)} + \frac{1}{2} \sum_{1\le k\ne i\le n} \int_{E^n} (\Phi_{ki} f - f) d\mu^{(n)} \\
&+ \sum_{1\le k\ne i\le n} \int_{E^n} \frac{1}{n} \beta(x_i, \mu)(\Phi_{ki} f - f) d\mu^{(n)} \\
=& \sum_{i=1}^n \langle B_i^\gamma f, \mu^{(n)} \rangle + \frac{1}{2} \sum_{1\le k\ne i\le n} \big(\langle \Phi_{ki} f, \mu^{(n)} \rangle - \langle f, \mu^{(n)} \rangle \big) \\
&+ \frac{1}{n} \sum_{1\le k\ne i\le n} \big(\langle \beta_i(\cdot, \mu) \Phi_{ki} f, \mu^{(n)} \rangle - \langle \beta_i(\cdot, \mu) f, \mu^{(n)} \rangle \big) \qquad (7.2.5)
\end{aligned}$$

where $\beta_i(\cdot, \mu) = \beta(x_i, \mu)$. Note that if $f \in B(E^m)$, $m \le n$, then for $m+1 \le i \le n$ we have $\Phi_{ki} f = f$ (cf. (7.2.4)) and $B_i^\gamma f = f$ (cf. 7.1.13), that implies

$$\begin{aligned}
&\sum_{i=m+1}^n \langle B_i^\gamma f, \mu^{(m)} \rangle + \frac{1}{2} \sum_{i=m+1}^n \sum_{k\ne i} \big(\langle \Phi_{ki} f, \mu^{(m)} \rangle - \langle f, \mu^{(m)} \rangle \big) \\
&+ \frac{1}{n} \sum_{i=m+1}^n \sum_{k\ne i} \big(\langle \beta_i(\cdot, \mu) \Phi_{ki} f, \mu^{(m)} \rangle - \langle \beta_i(\cdot, \mu) f, \mu^{(m)} \rangle \big) = 0 \qquad (7.2.6)
\end{aligned}$$

and for $1 \le i \le m$ and $m+1 \le k \le m$ we have $\langle \Phi_{ki} f, \mu^{(m)} \rangle = \langle f, \mu^{(m)} \rangle$, that implies

$$\frac{1}{2} \sum_{i=1}^m \sum_{k=m+1}^n \big(\langle \Phi_{ki} f, \mu^{(m)} \rangle - \langle f, \mu^{(m)} \rangle \big) = 0. \qquad (7.2.7)$$

Hence, for $f \in B(E^m)$, $m \le n$, (7.2.5) becomes

$$\begin{aligned}
\mathbb{A}^n \phi(\mu) =& \sum_{i=1}^m \langle B_i^\gamma f, \mu^{(m)} \rangle + \frac{1}{2} \sum_{1\le k\ne i\le m} \big(\langle \Phi_{ki} f, \mu^{(m)} \rangle - \langle f, \mu^{(m)} \rangle \big) \\
&+ \frac{1}{n} \sum_{1\le k\ne i\le m} \big(\langle \beta_i(\cdot, \mu) \Phi_{ki} f, \mu^{(m)} \rangle - \langle \beta_i(\cdot, \mu) f, \mu^{(m)} \rangle \big) \\
&+ \frac{n-m}{n} \sum_{i=1}^m \big(\langle \beta(\cdot, \mu) \otimes f, \mu^{(m+1)} \rangle - \langle \beta_i(\cdot, \mu) f, \mu^{(m)} \rangle \big) \qquad (7.2.8)
\end{aligned}$$

where $h = \beta(\cdot, \mu) \otimes f$ is defined by

$$h(x_1, \dots, x_m, \mu) = \beta(x_{m+1}, \mu) f(x_1, \dots, x_m). \tag{7.2.9}$$

In (7.2.8) $\beta_i(\cdot, \mu) f$ has been substituted by $\beta(\cdot, \mu) \otimes f$ because the i-th variable in f has been replaced by the k-th, which was not an argument of f, so that f is still a function of m variables and not of $m - 1$. But still $\beta(\cdot, \mu)$ has to express the weight associated to the outgoing particle, which, since we are integrating on the domain E, we can treat as a generic particle x_{m+1}.

Further observe that

$$\begin{aligned}
\mu^m =& \frac{1}{n} \sum_{i_1} \delta_{x_{i_1}} \dots \frac{1}{n} \sum_{i_m} \delta_{x_{i_m}} \\
=& \frac{1}{n^m} \sum_{1 \le i_1, \dots, i_m \le n} \delta_{x_{i_1}, \dots, x_{i_m}} \\
=& \frac{1}{n^m} \sum_{1 \le i_1 \ne \dots \ne i_m \le n} \delta_{x_{i_1}, \dots, x_{i_m}} + \frac{1}{n^m} \sum_{\text{at least two } i_j \text{ equal}} \delta_{x_{i_1}, \dots, x_{i_m}} \\
=& \frac{n_{(m)}}{n^m} \mu^{(m)} + \frac{1}{n^m} \sum_{\text{at least two } i_j \text{ equal}} \delta_{x_{i_1}, \dots, x_{i_m}}
\end{aligned} \tag{7.2.10}$$

where $n_{(m)} = n(n-1) \dots (n-m+1)$ and the second sum has $n^m - n_{(m)}$ terms. For $n \to \infty$ we have $n_{(m)}/n^m \to 1$, and hence $\mu^{(m)} \approx \mu^m$. Therefore, for large n, the limit of $\mathbb{A}^n$ is

$$\begin{aligned}
\mathbb{A}\varphi(\mu) =& \sum_{i=1}^{m} \langle B_i^\gamma f, \mu^m \rangle + \frac{1}{2} \sum_{1 \le k \ne i \le m} \langle \Phi_{ki} f - f, \mu^m \rangle \\
&+ \sum_{i=1}^{m} \left(\langle \beta(\cdot, \mu) \otimes f, \mu^{m+1} \rangle - \langle \beta_i(\cdot, \mu) f, \mu^m \rangle \right)
\end{aligned} \tag{7.2.11}$$

where $\varphi = \langle f, \mu^m \rangle$ as in (5.3.1). If (7.2.11) exists, this is the generator of the Fleming-Viot process with viability selection, with a particular mutation process (cf. (4.3.21)).

Note that the existence of the limiting operator (7.2.11) is subordinated to the boundedness of the function $\gamma(\cdot, \mu_n)$, given in (7.1.14), which in turn determines the boundedness of B^γ. It is therefore impossible to state any result on the relative compactness of this measure-valued process, let alone the weak convergence and the stationary distribution. In the next chapter we will adopt a different approach, which will allow to construct a different process and to claim such results.

7.3 The selection intensity function

Since the same assumptions will be done in the next chapter, in this section we discuss the properties of the selection intensity function $\beta(x_i, \mu_n) = \sigma\tilde{\beta}(x_i, \mu_n)$. In (7.1.6) we have defined $\tilde{\beta}_i$ to be such that

$$\tilde{\beta}_i \in \left(-\frac{n}{2\sigma}, \frac{n^2}{2\sigma} - \frac{n}{2\sigma}\right) \tag{7.3.1}$$

for all $i = 1, \ldots, n$, and

$$\sum_{i=1}^{n} \tilde{\beta}_i = 0. \tag{7.3.2}$$

for all $n \geq 1$ (note that the dependence on n of $\tilde{\beta}$ is implicit in the fact that it is function of the empirical measure μ_n). Therefore it is enough to choose, for example, $\tilde{\beta}_i \in (-1/2\sigma, 0)$, which means $\beta \in (-1/2, 0)$, to satisfy (7.3.1) for each $n \geq 1$, so that β is bounded for every dimension of the population. In particular we want to show that a negative β does not represent a problem in the generator of the measure-valued process (7.2.11).

Consider in general $\beta_{\min}$ and $\beta_{\max}$ such that $\beta_i \in [\beta_{\min}, \beta_{\max}]$ for all i, with $-1/2 \leq \beta_{\min} < \beta_{\max} \leq 0$. Take now $\beta_i' = \beta_i - \beta_{\min}$, so that $\beta_i' \in [0, \beta_{\max} + |\beta_{\min}|]$, i.e. β' is a bounded non negative function. Then from the third term of (7.2.11) we have

$$\begin{aligned}
&\sum_{i=1}^{m} \left(\langle \beta'(\cdot,\mu)\otimes f, \mu^{m+1}\rangle - \langle \beta_i'(\cdot,\mu) f, \mu^{m}\rangle\right) \\
&= \sum_{i=1}^{m} \left(\langle \beta(\cdot,\mu)\otimes f, \mu^{m+1}\rangle - \langle \beta_{\min} f, \mu^{m+1}\rangle - \langle \beta_i(\cdot,\mu) f, \mu^{m}\rangle + \langle \beta_{\min} f, \mu^{m}\rangle\right) \\
&= \sum_{i=1}^{m} \left(\langle \beta(\cdot,\mu)\otimes f, \mu^{m+1}\rangle - \langle \beta_{\min} f, \mu^{m}\rangle - \langle \beta_i(\cdot,\mu) f, \mu^{m}\rangle + \langle \beta_{\min} f, \mu^{m}\rangle\right) \\
&= \sum_{i=1}^{m} \left(\langle \beta(\cdot,\mu)\otimes f, \mu^{m+1}\rangle - \langle \beta_i(\cdot,\mu) f, \mu^{m}\rangle\right).
\end{aligned}$$

Hence the sign of β does not matter, and we can assume (7.3.1) and (7.3.2).

7.4 Discussion

In this chapter we introduced a particle process given by an n-dimensional exchangeable sequence, whose transition density is based on the predictive rule of the urn-scheme which characterizes the two-parameter Poisson-Dirichlet process. We then computed the generator of the associated process of empirical measures, and showed that it has the form of the generator of a Fleming-Viot process with selection, except for the mutation rate, which is a potential source of unboundedness. The lack of boundedness of the generators does not allow to prove any asymptotic result.

The construction given in Section 7.1 is the only way we found to obtain a generator that resembles (4.3.21) keeping at the same time the two features given by: the two-parameter Poisson-Dirichlet process, as the distribution of the random measure governing the sequence; and the Markov property for the particle process. If on the one hand it is true that Theorem 4.2.1 above says that the martingale problem for the generator of the Fleming-Viot process with selection is well posed, hence a solution to the martingale problem which is a Markov process does exists, on the other hand it is also true that the first step in generalising the Dirichlet process is toward a sampling method which takes into account exchangeable partitions, that is the two-parameter Poisson-Dirichlet process. In the trade-off between the two, the latter is the direction we will take.

In the next chapter we will show that relaxing the assumption of markovianity we can define a particle process on $D_{E^n}[0,\infty)$, whose de Finetti measure is given by the two-parameter Poisson-Dirichlet process, which leads to a measure-valued process which converges to the Fleming-Viot model with viability selection.

Chapter 8

Viability selection II: semi-Markov case

In the setting of the previous chapter, relaxing the assumption of Markovianity enables us to provide a construction of the Fleming-Viot process with viability selection. We define an E^n-valued semi-Markov jump process whose associated process of empirical measures weakly converges, in the Skorohod space, to the Fleming-Viot process with viability selection. The stationary distribution of such Fleming-Viot process is shown to be the two-parameter Poisson-Dirichlet process, also known as Pitman-Yor process.

8.1 The particle process

Consider a variation of the model constructed in Section 7.1 in which we relax the assumption of Markovianity. Namely, consider a pure jump process $\mathbf{X}^n = (X_1^n, \ldots, X_n^n)$ on E^n, for $n \geq 1$, where now, conditionally on the fact that the particle to be replaced is at level i, for $i = 1, \ldots, n$, the transitions depend on the incoming particle j and are governed by the semi-Markov kernel

$$G^n(\mathbf{x}_0, \mathrm{d}\mathbf{x}_t, \mathrm{d}t) = \sum_{j=0}^{k_i} P_j(x_i, \mathrm{d}y) Q^n_{j,i,\mathbf{x}}(t) \tag{8.1.1}$$

where k_i is the number of unique values in the subvector $\mathbf{x}_{-i}$, $Q^n_{j,i,\mathbf{x}}(t)$ is an exponential distribution of parameter $\tilde{\lambda}_{n,j,i}$, for $j = 0, 1, \ldots, k_i$, and

$$P_0(x_i, \mathrm{d}y) = \frac{\theta + \sigma k_i}{\theta + n - 1}\, \nu_0(\mathrm{d}y) \tag{8.1.2}$$

$$P_j(x_i, \mathrm{d}y) = \frac{c_{j^*} - \sigma}{\theta + n - 1}\, \delta_{x^*_j}(\mathrm{d}y) \qquad j = 1, \ldots, k_i \tag{8.1.3}$$

where $j = 0$ denotes that the particle comes from the non atomic measure ν_0 and c_{j^*} is the cardinality of the set of x_h such that $x_h = x^*_j$, not counting x_i. Note that it is not necessary to specify if k_i and $Q^n_{j,i,\mathbf{x}}(t)$ in (8.1.1) are computed on $\mathbf{x}_0$ or $\mathbf{x}_t$, since once removed the component at level i of both, they are equal.

The transitions are as follows. After the $(m-1)$th renewal, for $m \geq 1$, with probability $\tilde{\pi}^{\sigma,n}_i = \tilde{\pi}^{\sigma,n}(x_i)$ a component i is chosen to be replaced. The components different from the ith are set equal to their previous values as usual. The incoming particle that replaces the ith is then sampled from the density

$$\frac{\theta + \sigma k_i}{\theta + n - 1}\, \nu_0 + \frac{1}{\theta + n - 1} \sum_{j=1}^{k_i} (c_{j^*} - \sigma) \delta_{x^*_j} \tag{8.1.4}$$

which is the prediction rule that generates an i.i.d. sequence from a random distribution function sampled from a two-parameter Poisson-Dirichlet process of parameters σ and θ, denoted by $\mathscr{PD}(\sigma, \theta)$ (see Section 2.3). The chosen incoming particle determines the parameter of the exponential distribution from which a holding time is sampled, which will be $\tilde{\lambda}_{n,j,i} = \tilde{\lambda}_n(x^*_j, x_i, \mu_n)$ if the new particle is set equal to x^*_j, for

$j = 0, 1, \ldots, k_i$, or $\tilde{\lambda}_{n,0,i}$ if the new particle is from ν_0. Thus the distribution of the holding times $Q^n_{j,i,\mathbf{x}}(t)$ depends on both the outgoing particle x_i, which belongs to the starting state, and the incoming particle x^*_j, which belongs to the arrival state, from which the semi-Markovianity. See Section 1.2.

Proceeding as in Section 5.2, the generator for this process is

$$\begin{aligned}
A^n f(\mathbf{x}) &= \lim_{t\downarrow 0} \frac{1}{t} \sum_{i=1}^{n} \tilde{\pi}_i^{\sigma,n} \int \big[f(\mathbf{x}^n_t) - f(\mathbf{x}^n_0)\big] G^n(\mathbf{x}_0, \mathrm{d}\mathbf{x}_t, \mathrm{d}t) \\
&= \lim_{t\downarrow 0} \frac{1}{t} \sum_{i=1}^{n} \tilde{\pi}_i^{\sigma,n} \int \big[f(\mathbf{x}^n_t) - f(\mathbf{x}^n_0)\big] \sum_{j=0}^{k_i} P_j(x_i, \mathrm{d}y) Q^n_{j,i,\mathbf{x}}(t) \\
&= \lim_{t\downarrow 0} \frac{1}{t} \sum_{i=1}^{n} \tilde{\pi}_i^{\sigma,n} \int \big[f(\mathbf{x}^n_t) - f(\mathbf{x}^n_0)\big] \\
&\quad \times \Bigg\{ \left(1 - e^{-\tilde{\lambda}_{n,0,i} t}\right) \frac{\theta + \sigma k_i}{\theta + n - 1}\, \nu_0(\mathrm{d}y) \\
&\qquad + \sum_{j=1}^{k_i} \left(1 - e^{-\tilde{\lambda}_{n,j,i} t}\right) \frac{c_{j^*} - \sigma}{\theta + n - 1}\, \delta_{x^*_j}(\mathrm{d}y) \Bigg\} \\
&= \sum_{i=1}^{n} \tilde{\pi}_i^{\sigma,n} \int \big[f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})\big] \\
&\quad \times \left\{ \tilde{\lambda}_{n,0,i} \frac{\theta + \sigma k_i}{\theta + n - 1}\, \nu_0(\mathrm{d}y) + \sum_{j=1}^{k_i} \tilde{\lambda}_{n,j,i} \frac{c_{j^*} - \sigma}{\theta + n - 1}\, \delta_{x^*_j}(\mathrm{d}y) \right\} \\
&= \sum_{i=1}^{n} \tilde{\pi}_i^{\sigma,n}\, \tilde{\lambda}_{n,0,i}\, \frac{\theta + \sigma k_i}{\theta + n - 1} \int \big[f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})\big] \nu_0(\mathrm{d}y) \\
&\quad + \sum_{i=1}^{n} \sum_{j=1}^{k_i} \tilde{\pi}_i^{\sigma,n}\, \tilde{\lambda}_{n,j,i}\, \frac{c_{j^*} - \sigma}{\theta + n - 1} \big[f(\eta_i(\mathbf{x}|\, x^*_j)) - f(\mathbf{x})\big] \\
&= \sum_{i=1}^{n} \tilde{\pi}_i^{\sigma,n}\, \tilde{\lambda}_{n,0,i}\, \frac{\theta + \sigma k_i}{\theta + n - 1} \int \big[f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})\big] \nu_0(\mathrm{d}y) \\
&\quad + \sum_{i=1}^{n} \sum_{l\neq i}^{n} \tilde{\pi}_i^{\sigma,n}\, \tilde{\lambda}_{n,l,i}\, (\theta + n - 1)^{-1} \left(1 - \frac{\sigma}{c_l}\right) \big[f(\eta_i(\mathbf{x}|\, x_l)) - f(\mathbf{x})\big] \qquad (8.1.5)
\end{aligned}$$

where in the last equality we used the fact that

$$\sum_{j=1}^{k_i} 1 = \sum_{\substack{1\le h\le n\\ h\ne i}} \frac{1}{c_h} . \tag{8.1.6}$$

Let now the probability of choosing a particle at level i be

$$\tilde{\pi}_i^{\sigma,n} = \frac{1}{n}\left(1 + \frac{2\sigma\tilde{\beta}(x_i,\mu_n)}{n}\right) \tag{8.1.7}$$

(cf. (7.1.7) above) where σ is the parameter in (8.1.4) and $\tilde{\beta}_i = \tilde{\beta}(x_i,\mu_n)$ is such that

$$\begin{cases} \tilde{\beta}_i \in \left(-\frac{n}{2\sigma}, \frac{n^2}{2\sigma} - \frac{n}{2\sigma}\right) \\ \sum_i \tilde{\beta}_i = 0 \end{cases} \tag{8.1.8}$$

(see Section 7.3). Let now the Poisson rates be

$$\tilde{\lambda}_{n,l,i} = \left(1 - \frac{\sigma}{c_l}\right)^{-1} \frac{n(\theta+n-1)}{2} \tag{8.1.9}$$

for $l = 1,\dots,n$, $l \ne i$, and

$$\tilde{\lambda}_{n,0,i} = \left(\frac{\theta}{\theta+\sigma k_i}\right) \frac{n(\theta+n-1)}{2}. \tag{8.1.10}$$

From (8.1.5) we obtain

$$\begin{aligned} A^n f(\mathbf{x}) =& \sum_{i=1}^{n} \frac{1}{n}\left(1 + \frac{2\sigma\tilde{\beta}(x_i,\mu_n)}{n}\right) \left(\frac{\theta}{\theta+\sigma k_i}\right) \frac{n(\theta+n-1)}{2} \frac{\theta+\sigma k_i}{\theta+n-1} \\ & \times \int \big[f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})\big]\nu_0(\mathrm{d}y) \\ & + \sum_{i=1}^{n}\sum_{l\ne i}^{n} \frac{1}{n}\left(1 + \frac{2\sigma\tilde{\beta}(x_i,\mu_n)}{n}\right)\left(1 - \frac{\sigma}{c_l}\right)^{-1} \frac{n(\theta+n-1)}{2} \\ & \times (\theta+n-1)^{-1}\left(1 - \frac{\sigma}{c_l}\right)\big[f(\eta_i(\mathbf{x}|\, x_l)) - f(\mathbf{x})\big] \\ =& \sum_{i=1}^{n} \frac{1}{2}\theta\left(1 + \frac{2\sigma\tilde{\beta}(x_i,\mu_n)}{n}\right) \int \big[f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})\big]\nu_0(\mathrm{d}y) \\ & + \frac{1}{2}\sum_{1\le l\ne i\le n}\left(1 + \frac{2\sigma\tilde{\beta}(x_i,\mu_n)}{n}\right)\big[f(\eta_i(\mathbf{x}|\, x_l)) - f(\mathbf{x})\big] \end{aligned} \tag{8.1.11}$$

Note in (8.1.5), or in (8.1.11), that the particle at level l which determines the exponential parameter $\tilde{\lambda}_{n,l,i}$, for $l = 1, \dots, n$, $l \neq i$, is the same particle that is duplicated to replace the particle at level i, where we have conventionally set $l = 0$ to mean that the new particle comes from ν_0 instead of coming from the empirical measure (actually from the subvector $\mathbf{x}_{-i}$). We can write (8.1.11) as (cf. (7.1.12) above)

$$A^n f(\mathbf{x}) = \sum_{i=1}^{n} B_i^\beta f(\mathbf{x}) + \frac{1}{2} \sum_{1 \leq l \neq i \leq n} \left(1 + \frac{2\beta(x_i, \mu_n)}{n}\right) \left[f(\eta_i(\mathbf{x}|\, x_l)) - f(\mathbf{x})\right] \quad (8.1.12)$$

where the mutation operator now is (cf. (7.1.13) above)

$$B^\beta g(x) = \frac{1}{2}\theta \left(1 + \frac{2\beta(x_i, \mu_n)}{n}\right) \int \left[f(y)) - f(x)\right] \nu_0(\mathrm{d}y) \quad (8.1.13)$$

and

$$\beta(x, \mu_n) = \sigma\, \tilde{\beta}(x, \mu_n). \quad (8.1.14)$$

Note in particular that, since β is a bounded function (cf. Section 7.3), the mutation operator B^β is bounded, and so is the whole generator (8.1.12).

The boundedness of the generator (8.1.12) has thus been obtained at the expense of the Markov property. The reason why we have chosen to stick with a certain type of parametrisation for the model, namely that described by (8.1.1), (8.1.7), (8.1.9) and (8.1.10), will be clear in the next sections, when we will be able to state substantial results which connect the Fleming-Viot process with selection and the two-parameter Poisson-Dirichlet process.

As it held for the model of Chapter 7 (cf. in particular the end of Section 7.1), observe that for $\sigma = 0$, $\tilde{\pi}_i^{\sigma,n}$ in (8.1.7) reduces to $\pi_i^n = n^{-1}$ in (5.2.3), both $\tilde{\lambda}_{n,l,i}$ in (8.1.9) and $\tilde{\lambda}_{n,0,i}$ in (8.1.10) reduce to $\lambda_n = 2^{-1}n(\theta + n - 1)$ in (5.2.5), and also $\beta \equiv 0$, so that B^β in (8.1.13) simplifies to B in (5.2.8). Hence, the parameters of the neutral case are once again recovered, the collection of Poisson processes simplifies to a single Poisson process and the generator (8.1.12) simplifies to (5.2.7), the generator of the particle process for the neutral Fleming-Viot. Again, this is the least we expected, since for $\sigma = 0$ the two-parameter Poisson-Dirichlet process $\mathscr{PD}(\sigma, \theta)$ reduces to the Dirichlet process Π_{θ,ν_0}, which is the stationary distribution of the measure-valued process with generator (5.3.4).

The difference between the generator (7.1.12) for the Markov jump process and the generator (8.1.12) for the semi-Markov one lies only in the coefficient of the mutation operator, which in the Markov case was

$$\gamma(x, \mu_n) = \frac{1}{2}(\theta + \sigma k_x)\left(1 + \frac{2\beta(x, \mu_n)}{n}\right)\frac{n-1}{n-1-\sigma k_x} \tag{8.1.15}$$

and now has become

$$\frac{1}{2}\theta\left(1 + \frac{2\beta(x_i, \mu_n)}{n}\right). \tag{8.1.16}$$

The interpretation of the reason why a semi-Markov model has been able to counterbalance the unboundedness of the generator (7.1.12), inherited by the mutation rate (8.1.15), and therefore succeed where the Markov model had failed, could be given by the following argument. In the Markov case, the mutation rate (8.1.15) is a function of the number of clusters k_i, i.e. of the number of different types in the population once excluded x_i. Indeed in the predictive distribution (8.1.4), associated to the two-parameter Poisson-Dirichlet process, which is at the base of the process' construction, the probability of sampling from the nonatomic measure ν_0 does not vanish at infinity like in the predictive of the Blackwell-MacQueen urn-scheme (5.1.7), which in Chapter 5 was shown to be the process underlying the neutral Fleming-Viot, and this because the weight of ν_0 itself is a function of k_i. Thus a sort of "'vicious circle"' between ν_0 and its weight (as a function of k_i) takes gradually control of the sampling process, a self enforcing mechanism according to which the higher the number of different values in the vector, the higher the probability of sampling the next one from ν_0, which will provide a yet-non-observed value with probability one. With the introduction of a semi-Markov kernel, we have been able to discriminate among the different arrival states, and in particular to set a specific parameter for the holding time that is drawn when the next particle is sampled from ν_0. In this way one can counterbalance the mechanism between k_i and ν_0 by slowing down the mutation rate, that is progressively decreasing the speed of the mutation operator, or, from a different standpoint, progressively decreasing the probability of sampling from ν_0, as the number of different types in the population increases. This is clearly seen from the structure of $\tilde{\lambda}_{n,0,i}$ in (8.1.10), which is a decreasing function of k_i.

8.2 The associated measure-valued process

In this section we proceed similarly to Section 7.2, in order to derive the generator of the process of the empirical measures of the exchangeable sequence given by the particle process of Section 8.1, and show that for large n it converges to the generator of the Fleming-Viot process with viability selection.

For $m < n$, let $\mu^{(m)}$ be as above the probability measure on E^m defined by

$$\mu^{(m)} = \frac{1}{n(n-1)\dots(n-m+1)} \sum \delta_{(x_{i_1},\dots,x_{i_m})} \tag{8.2.1}$$

where the sum is taken over all choices of $1 \le i_1, \dots, i_m \le n$ with $i_k \neq i_l$, and define, for $f \in B(E^n)$, $\phi \in B(\mathscr{P}^n(E))$ by

$$\phi(\mu) = \langle f, \mu^{(n)} \rangle \tag{8.2.2}$$

and $\mathbb{A}^n \phi$ by

$$\mathbb{A}^n \phi(\mu) = \langle A^n f, \mu^{(n)} \rangle; \tag{8.2.3}$$

see Section 4.3.2. Also let $\Phi_{ki} : E^n \to E^{n-1}$ be

$$\Phi_{ki} f(x_1, \ldots, x_n) = f(x_1, \ldots, x_{i-1}, x_k, x_{i+1}, \ldots, x_n) \tag{8.2.4}$$

for $1 \le k \ne i \le n$ as before. From (8.1.12) and (8.2.3) we obtain, for $\mu \in \mathscr{P}^n(E)$,

$$\begin{aligned}
\mathbb{A}^n \phi(\mu) =& \int_{E^n} \sum_{i=1}^n B_i^\beta f d\mu^{(n)} + \int_{E^n} \frac{1}{2} \sum_{1 \le k \ne i \le n} (\Phi_{ki} f - f) d\mu^{(n)} \\
&+ \int_{E^n} \sum_{1 \le k \ne i \le n} \frac{1}{n} \beta_i(\cdot, \mu)(\Phi_{ki} f - f) d\mu^{(n)} \\
=& \sum_{i=1}^n \int_{E^n} B_i^\beta f d\mu^{(n)} + \frac{1}{2} \sum_{1 \le k \ne i \le n} \int_{E^n} (\Phi_{ki} f - f) d\mu^{(n)} \\
&+ \sum_{1 \le k \ne i \le n} \int_{E^n} \frac{1}{n} \beta_i(\cdot, \mu)(\Phi_{ki} f - f) d\mu^{(n)} \\
=& \sum_{i=1}^n \langle B_i^\beta f, \mu^{(n)} \rangle + \frac{1}{2} \sum_{1 \le k \ne i \le n} \big(\langle \Phi_{ki} f, \mu^{(n)} \rangle - \langle f, \mu^{(n)} \rangle \big) \\
&+ \frac{1}{n} \sum_{1 \le k \ne i \le n} \big(\langle \beta_i(\cdot, \mu) \Phi_{ki} f, \mu^{(n)} \rangle - \langle \beta_i(\cdot, \mu) f, \mu^{(n)} \rangle \big)
\end{aligned} \tag{8.2.5}$$

where $\beta_i(\cdot, \mu) = \beta(x_i, \mu)$. If $f \in B(E^m)$, $m \le n$, using (7.2.6) and (7.2.7), (8.2.5) becomes

$$\mathbb{A}^n\phi(\mu) = \sum_{i=1}^{m}\langle B_i^{\beta} f, \mu^{(m)}\rangle + \frac{1}{2}\sum_{1\le k\ne i\le m}\left(\langle \Phi_{ki} f, \mu^{(m)}\rangle - \langle f, \mu^{(m)}\rangle\right)$$
$$+ \frac{1}{n}\sum_{1\le k\ne i\le m}\left(\langle \beta_i(\cdot,\mu)\Phi_{ki} f, \mu^{(m)}\rangle - \langle \beta_i(\cdot,\mu) f, \mu^{(m)}\rangle\right)$$
$$+ \frac{n-m}{n}\sum_{i=1}^{m}\left(\langle \beta(\cdot,\mu)\otimes f, \mu^{(m+1)}\rangle - \langle \beta_i(\cdot,\mu) f, \mu^{(m)}\rangle\right) \tag{8.2.6}$$

where $h = \beta(\cdot,\mu)\otimes f$ is defined by

$$h(x_1,\ldots,x_m,\mu) = \beta(x_{m+1},\mu) f(x_1,\ldots,x_m) \tag{8.2.7}$$

(see 7.2.9 for comments on the role of $\beta\otimes f$).

Recalling from (7.2.10) that, for large n, $\mu^{(m)}$ is essentially product measure, and noting that for n growing to infinity (8.1.16) converges to $2^{-1}\theta$, (8.2.6) converges to

$$\mathbb{A}\varphi(\mu) = \sum_{i=1}^{m}\langle B_i f, \mu^m\rangle + \frac{1}{2}\sum_{1\le k\ne i\le m}\langle \Phi_{ki} f - f, \mu^m\rangle$$
$$+ \sum_{i=1}^{m}\left(\langle \beta(\cdot,\mu)\otimes f, \mu^{m+1}\rangle - \langle \beta_i(\cdot,\mu) f, \mu^m\rangle\right), \tag{8.2.8}$$

with B as in (5.2.8) and φ as in (5.3.1), which is the generator of the Fleming-Viot process with viability selection (cf. 4.3.21).

In the next section it will be shown that the process of empirical measures with generator (8.2.5) weakly converges to the Fleming-Viot process with selection in the space $D_{\mathscr{P}(E)}[0,\infty)$.

8.3 Weak convergence

To show that the process of empirical measure $\{\mu_{\mathbf{x}^n}(t), t\ge 0\}$ weakly converges to the Fleming-Viot process with generator (8.2.8), denoted by $\{\mu_t, t\ge 0\}$, we proceed similarly to Section 5.4.

Define $\mu_{\mathbf{x}^n} : E^n \times [0,\infty) \to \mathscr{P}(E)$ by

$$\mu_{\mathbf{x}^n}(t) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i^n(t)}.$$

that is $\mu_{\mathbf{x}^n}$ is the process of empirical measures of the X_i's, and observe that it has sample paths in $D_{\mathscr{P}(E)}[0,\infty)$. Then the above construction implies the weak convergence of $\{\mu_{\mathbf{x}^n}(t), t \geq 0\}$ to the Fleming-Viot process with generator (8.2.8), denoted by $\{\mu_t, t \geq 0\}$.

Theorem 8.3.1 *Let $\{\mu_{\mathbf{x}^n}(t), t \geq 0\}$ be a process with values in $\mathscr{P}(E)$ as in the above construction, and let $\{\mu_t, t \geq 0\}$ be a Fleming-Viot process with generator* (8.2.8)*. Assume that $P\mu_{\mathbf{x}^n}^{-1}(0) \Rightarrow \mu_0$, for $\mu_0 \in \mathscr{P}(E)$. Then $\{\mu_{\mathbf{x}^n}(t), t \geq 0\}$ is asymptotically Markov and*

$$\{\mu_{\mathbf{x}^n}(t), t \geq 0\} \Rightarrow \{\mu_t, t \geq 0\} \tag{8.3.1}$$

in $D_{\mathscr{P}(E)}[0,\infty)$, where $\Rightarrow$ means convergence in distribution in the Skorohod topology.

Proof 1. Since the mutation operator (8.1.13) converges to (5.2.8), the above computation implies that for $f \in B(E^m)$

$$\sup_{\mathbf{x}\in E^n} |\langle f, \mu_{\mathbf{x}^n}^{(m)}\rangle - \varphi(\mu_{\mathbf{x}^n})| \longrightarrow 0 \tag{8.3.2}$$

and

$$\sup_{\mathbf{x}\in E^n} |\langle A^n f, \mu_{\mathbf{x}^n}^{(m)}\rangle - \mathbb{A}\varphi(\mu_{\mathbf{x}^n})| \longrightarrow 0 \tag{8.3.3}$$

and (8.3.1) follows analogously to Section 4.3.1.

The asymptotical Markovianity of $\{F_n(t),\, t \geq 0\}$ follows from the convergence of the generator, together with the well-posedness of the martingale problem for (8.2.8) (see Theorem 4.2.1), which in turn implies the Markovianity of the unique solution (see Theorem 1.5.3). See also Remark 8.3.2 below. □

PROOF 2. Assume the process $\{\mu_{\mathbf{x}^n}(t), t \geq 0\}$ is relatively compact in $D_{\mathscr{P}(E)}[0,\infty)$, which follows from Donnelly and Kurtz (1999a).

Define

$$\{\,(\xi_n, \psi_n)\,\} = \{\big(g_n(\mathbf{X}^n), h_n(\mathbf{X}^n)\big)\} \tag{8.3.4}$$

where

$$g_n(\mathbf{X}^n) = \langle f, \mu_{\mathbf{x}^n}^{(m)} \rangle = \phi(\mu_{\mathbf{x}^n}) \tag{8.3.5}$$

as in (8.2.2), and

$$h_n(\mathbf{X}^n) = \langle A^n f, \mu_{\mathbf{x}^n}^{(m)} \rangle = \mathbb{A}^n \phi(\mu_{\mathbf{x}^n}) \tag{8.3.6}$$

as in (8.2.3). Note that

For each $T \geq 0$

$$\sup_{n\to\infty} \sup_{s \leq T} E[|\xi_n(s)|] = \sup_{n\to\infty} \sup_{s \leq T} E[|\langle f, \mu_{\mathbf{x}^n}^{(m)}(s) \rangle|] < \infty \tag{8.3.7}$$

since $f \in B(E^m)$, and

$$\sup_{n\to\infty} \sup_{s \leq T} E[|\psi_n(s)|] = \sup_{n\to\infty} \sup_{s \leq T} E[|\langle A^n f, \mu_{\mathbf{x}^n}^{(m)}(s) \rangle|] < \infty \tag{8.3.8}$$

since A^n in (8.1.12) is a bounded operator. Further, let

$$\widehat{\mathbb{A}} = \{(\varphi', \mathbb{A}\varphi') : \varphi' \in \mathscr{D}(\mathbb{A})\} \subset B(\mathscr{P}(E)) \times B(\mathscr{P}(E)) \tag{8.3.9}$$

where

$$\mathscr{D}(\mathbb{A}) = \{\varphi' = \langle f, \mu^m \rangle :\ f \in B(E^m),\ \mu \in \mathscr{P}^n(E),\ n \geq m \geq 1\} \tag{8.3.10}$$

(cf. (1.4.10), satisfied by means of Proposition 1.4.9-(c)). For each $(g,h)=(\varphi,\mathbb{A}\varphi)\in\widehat{\mathbb{A}}$, where $\varphi=\langle f,\mu^m\rangle$ and μ is purely atomic, there exist $(g_n,h_n)\in\widehat{\mathbb{A}}^n$, where

$$\widehat{\mathbb{A}}^n=\{(\phi',\mathbb{A}^n\phi'):\phi'\in\mathscr{D}(\mathbb{A}^n)\}\subset B(\mathscr{P}(E))\times B(\mathscr{P}(E)) \tag{8.3.11}$$

with

$$\mathscr{D}(\mathbb{A}^n)=\left\{\phi'=\langle f,\mu^{(m)}\rangle:\ f\in B(E^m),\ \mu\in\mathscr{P}^n(E),\ n\geq m\geq 1\right\} \tag{8.3.12}$$

such that, letting $\Gamma\subset[0,\infty)$ be the set of jump points of the process $\mathbf{X}^n$, for $t\notin\Gamma$ (7.2.10) implies that

$$\begin{aligned}\lim_{n\to\infty}E\left[\xi_n(t)-g(\mu_{\mathbf{x}^n}(t))\right]&=\lim_{n\to\infty}E\left[g_n(\mathbf{X}^n(t))-g(\mu_{\mathbf{x}^n}(t))\right]\\&=\lim_{n\to\infty}E\left[\phi(\mu_{\mathbf{x}^n}(t))-\varphi(\mu_{\mathbf{x}^n}(t))\right]\\&=\lim_{n\to\infty}E\left[\langle f,\mu_{\mathbf{x}^n}^{(m)}(t)\rangle-\langle f,\mu_{\mathbf{x}^n}^{m}(t)\rangle\right]=0\end{aligned} \tag{8.3.13}$$

and the convergence of (8.2.6) to (8.2.8) implies that

$$\begin{aligned}\lim_{n\to\infty}E\left[\psi_n(t)-h(\mu_{\mathbf{x}^n}(t))\right]&=\lim_{n\to\infty}E\left[h_n(\mathbf{X}^n(t))-h(\mu_{\mathbf{x}^n}(t))\right]\\&=\lim_{n\to\infty}E\left[\mathbb{A}^n\phi(\mu_{\mathbf{x}^n}(t))-\mathbb{A}\varphi(\mu_{\mathbf{x}^n}(t))\right]=0.\end{aligned} \tag{8.3.14}$$

Noting that $(\mathscr{P}(E),\rho)$, where ρ is Prohorov's metric, is separable and complete since E is (see Lemma 3.2.3), the thesis follows Corollary 3.4.8, given (8.3.7), (8.3.8), (8.3.13), (8.3.14) together with Remark 3.4.7, the relative compactness of $\{\mu_{\mathbf{x}^n}(t),t\geq 0\}$ and the well-posedness of the martingale problem for the generator (8.2.8) (cf. Theorem 4.2.1-(d) and note that uniqueness of the solution, not depending on the symmetry of σ, holds when the latter is univariate; cf. also Ethier and Kurtz, 1987, pag. 77). □

Remark 8.3.2 Note that (8.3.13) and (8.3.14) imply that

$$\lim_{n\to\infty} E\Big[\big|E\big[f(\mathbf{X}^n(t+s))|\mathscr{F}_t^{\mathbf{X}^n}\big] - T(s)f(\mathbf{X}^n(t))\big|\Big] = 0 \tag{8.3.15}$$

(cf. Remark 3.4.7 above, or Ethier and Kurtz, 1986, Remark 4.8.3(a)) where $\{T(s)\}$ is the semigroup corresponding to the limiting process. In this sense we can say that the sequence of processes $\{\mathbf{X}^n\}_{n\geq 1}$ approximates a Markov process.

8.4 Stationary distribution

In Sections 4.2.2 and 5.5 we saw that the stationary distribution of the neutral Fleming-Viot process with generator (5.3.4) and mutation

$$Bg(x) = \frac{1}{2}\theta \int \big[g(y) - g(x)\big]\nu_0(\mathrm{d}y) \tag{8.4.1}$$

is the Dirichlet process of parameters (θ, ν_0). In this section we show that the two-parameter Poisson-Dirichlet process is the stationary distribution of the Fleming-Viot process with viability selection, whose generator is (8.2.8).

Consider the embedded E^n-valued chain (5.1.9), given by

$$\mathbf{Y}^n_{N(t)} = \mathbf{X}^n(t). \tag{8.4.2}$$

The transitions described in Section 8.1, restricted to the embedded chain, consist of randomly choosing a variable from the vector $\mathbf{Y}^n_{N(t)}$, which is replaced with a sample from

$$\frac{\theta + \sigma k_i}{\theta + n - 1}\,\nu_0 + \frac{1}{\theta + n - 1}\sum_{j=1}^{k_i}(c_{j^*} - \sigma)\,\delta_{y_j^*} \tag{8.4.3}$$

where i is the level of the component replaced, and k_i and c_{j^*} are as defined at the beginning of the chapter. As previously observed, this is like implementing a Gibbs sampler (see Section 1.3) to $\mathbf{Y}^n = (Y_1, \ldots, Y_n)$, where (8.4.3) is the full conditional distribution. By the properties of the Gibbs sampler, $\mathbf{Y}^n_{N(t)}$ is a Markov chain on E^n with stationary distribution $\mathcal{P}'_n$ given by

$$\mathcal{P}'_n = \prod_{i=1}^{n} \frac{(\theta + \sigma k_i)\, \nu_0 + \sum_{j=1}^{k_i} (c_{j^*} - \sigma)\, \delta_{y_j^*}}{\theta + i - 1}. \tag{8.4.4}$$

Also, if $\mathcal{P}'_n$ is the initial distribution (which we assume from now on), at every step the sequence $(Y_1, \ldots, Y_n)$ is exchangeable with distribution $\mathcal{P}'_n$.

From Pitman (1996b) (see also Ishwaran and Zarepour, 2003) we know that when the sequence $(Y_1, \ldots, Y_n)$ is sampled according to the prediction rule (8.4.3), and thus has distribution (8.4.4), then

$$Y_i \mid \mu \overset{iid}{\sim} \mu \qquad i = 1, \ldots, n$$

$$\mu \sim \mathscr{PD}(\sigma, \theta)$$

and this means

$$(Y_1, \ldots, Y_n) \sim \int_{\mathscr{P}(E)} \mu^n(\mathrm{d}\,\cdot)\widetilde{\Pi}_{\sigma,\theta}(\mathrm{d}\mu) \tag{8.4.5}$$

where $\mathscr{PD}(\sigma, \theta)$ and $\widetilde{\Pi}_{\sigma,\theta}$ denote the two-parameter Poisson-Dirichlet process with parameters σ and θ, μ is a random probability measure sampled from $\widetilde{\Pi}_{\sigma,\theta}$ and μ^n is the n-fold product measure. Since now the continuous time process is constant between consecutive jumps, for given n (8.4.5) holds not only for the jump times of the continuous time process, in which the chain $\mathbf{Y}^n_{N(t)}$ is embedded, but for each $t \geq 0$, and hence is the stationary distribution of the E^n-valued continuous time process $\mathbf{X}^n(t)$.

From Theorem 2.1.2 (see also Aldous, 1985),

$$\mu_{\mathbf{x}^\infty} = \operatorname*{weak-lim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \sim \widetilde{\Pi}_{\sigma,\theta} \tag{8.4.6}$$

that is for each $t \geq 0$ the two-parameter Poisson-Dirichlet process is the distribution of the weak limit of the empirical measure of the particle process.

Since, in addition, the $C_{\mathscr{P}(E)}[0,\infty)$ martingale problem for $\mathbb{A}$ given by (8.2.8) is well posed (cf. Theorem 4.2.1-(d) and note that uniqueness of the solution, not depending on the symmetry of σ, holds when the latter is univariate; cf. also Ethier and Kurtz, 1987, pag. 77), from Lemma A.2 (see Appendix) it follows that $\widetilde{\Pi}_{\sigma,\theta}$ is the stationary distribution for the Flemig-Viot process with generator (8.2.8).

8.5 Discussion

In the previous chapter we attempted to construct the Fleming-Viot process with viability selection via a Markov particle process based on the prediction rule which characterizes the two-parameter Poisson-Dirichlet process, but the resulting generators turned out to be unbounded.

In the present chapter we relaxed the assumption of Markovianity and introduced a particle process, whose transitions are governed by a semi-Markov kernel, which is Markov in the limit for large n, and whose associated process of empirical measures weakly converges to the Fleming-Viot process with viability selection. It is shown that the stationary distribution of such Fleming-Viot process is the two-parameter Poisson-Dirichlet process.

Chapter 9

Haploid fertility selection

In Section 4.2.2 we have seen that the Fleming-Viot process with mutation operator $Bf = \frac{1}{2}\theta(\langle f, \nu_0 \rangle - f)$ and haploid fertility selection function $\sigma \in B(E)$, then its stationary distribution is given by $\Pi_\sigma(\mathrm{d}\mu) = Ce^{2\langle \sigma, \mu \rangle}\Pi_{\theta,\nu_0}(\mathrm{d}\mu)$, where C is a constant of proportionality.

The purpose of this chapter is to provide an explicit construction, similar to that of the previous chapters, of such Fleming-Viot process. We will proceed by considering a generalised version of the predictive density associated with the Blackwell-MacQueen urn-scheme, thus implicitly generalising the Dirichlet process prior. In particular, we introduce the novel feature of allowing a non-uniform sampling from the empirical distribution given the first n observations, and use the new density to describe the Markov jump process in E^n. By considering the associated process of empirical measures, we then show that in a special case for σ one obtains in the limit the Fleming-Viot process with haploid fertility selection. It is then shown that appealing to a Dirichlet process mixture model yields a simple derivation of the stationary distribution of the measure-valued diffusion.

9.1 Generalisation of the predictive density

In this section the key predictive density is introduced, in such a way that allows a general choice for the function which alters the weights in the empirical distribution, without affecting exchangeability, and it is shown that with a simple rule we recover the Dirichlet case.

Let $X_1, \dots, X_n$ be exchangeable in the space E. Then it is well known that the joint density $p_n(x_1, \dots, x_n)$ is the same under any permutation of the vector $(x_1, \dots, x_n)$. If this holds for all n then there exists a de Finetti measure, or in the Bayesian parlance, a prior on the space of distribution functions $\mathscr{P}(E)$. Consider now a generalisation of $p_n(x_1, \dots, x_n)$, for each fixed n, by introducing a sequence of fixed functions $\tilde{\sigma}_n(x)$ and constructing

$$q_n(x_1, \dots, x_n) \propto p_n(x_1, \dots, x_n) \prod_{i=1}^{n} \tilde{\sigma}_n(x_i). \tag{9.1.1}$$

We will assume throughout the chapter that $\tilde{\sigma}_n$ is chosen, to ensure that q_n exists for each n. The density q_n is clearly exchangeable. Of course this does not imply the existence of a de Finetti measure.

The $(n-1)$ predictive density is available; that is we remove one of the elements, say x_i, and predict a replacement for it. This predictive is given by

$$q_n(\mathrm{d}x |\, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \frac{\tilde{\sigma}_n(x)\, p_n(\mathrm{d}x |\, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)}{\int \tilde{\sigma}_n(x)\, p_n(\mathrm{d}x |\, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)}. \tag{9.1.2}$$

We will concentrate on the case when p_n is derived from the Dirichlet process prior (see Section 2.2). If now

$$x_1, \ldots, x_n |\, \mu \overset{iid}{\sim} \mu$$

then marginally

$$p_n(\mathrm{d}x_1, \ldots, \mathrm{d}x_n) = \nu_0(\mathrm{d}x_1) \prod_{i=2}^{n} \frac{\theta\, \nu_0(\mathrm{d}x_i) + \sum_{l<i} \delta_{x_l}(\mathrm{d}x_i)}{\theta + i - 1}$$

and the $(n-1)$ predictive density is

$$p_n(\mathrm{d}x |\, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = \frac{\theta\, \nu_0(\mathrm{d}x) + \sum_{l \neq i}^{n} \delta_{x_l}(\mathrm{d}x)}{\theta + n - 1} \qquad (9.1.3)$$

(cf. (5.1.7)). Now let us consider our more general model with the $\tilde{\sigma}_n$ function. From (9.1.2) we have

$$\begin{aligned} q_n(\mathrm{d}x |\, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) &= \frac{\tilde{\sigma}_n(x)\, p_n(\mathrm{d}x |\, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)}{\int \tilde{\sigma}_n(x)\, p_n(\mathrm{d}x |\, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)} \\ &= \frac{\tilde{\sigma}_n(x) \left(\theta\, \nu_0(\mathrm{d}x) + \sum_{l \neq i}^{n} \delta_{x_l}(\mathrm{d}x)\right)}{\int \tilde{\sigma}_n(x) \left(\theta\, \nu_0(\mathrm{d}x) + \sum_{l \neq i}^{n} \delta_{x_l}(\mathrm{d}x)\right)} \frac{\theta + n - 1}{\theta + n - 1} \\ &= \frac{\theta \tilde{\sigma}_n(x) \nu_0(\mathrm{d}x) + \sum_{l \neq i}^{n} \tilde{\sigma}_n(x)\, \delta_{x_l}(\mathrm{d}x)}{\int \theta \tilde{\sigma}_n(x) \nu_0(\mathrm{d}x) + \sum_{l \neq i}^{n} \tilde{\sigma}_n(x)\, \delta_{x_l}(\mathrm{d}x)} \end{aligned}$$

so that the predictive density can be written

$$q_n(\mathrm{d}x |\, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = \frac{\theta_n\, \nu_n(\mathrm{d}x) + \sum_{l \neq i}^{n} \tilde{\sigma}_n(x_l)\, \delta_{x_l}(\mathrm{d}x)}{\theta_n + \sum_{l \neq i}^{n} \tilde{\sigma}_n(x_l)} \qquad (9.1.4)$$

where

$$\theta_n = \theta \int \tilde{\sigma}_n(x)\, \nu_0(\mathrm{d}x) \qquad (9.1.5)$$

and

$$\nu_n(\mathrm{d}x) = \frac{\tilde{\sigma}_n(x)\, \nu_0(\mathrm{d}x)}{\int \tilde{\sigma}_n(x)\, \nu_0(\mathrm{d}x)}. \qquad (9.1.6)$$

Observe that in (9.1.4) a larger $\tilde{\sigma}_n$ implies a larger probability for the particle x_l of being selected to replace x_i, i.e. the larger the $\tilde{\sigma}_n$ the higher the fitness (the fertility) of the individual who is going to have an offspring.

Note also that we recover the predictive from the Dirichlet process, that is (9.1.3), when $\tilde{\sigma}_n(x) \equiv 1$ for all n, which means that the selection is uniform and does not favour any type of individual (i.e. the model is neutral).

In the next section we construct the E^n-valued and the $\mathscr{P}(E)$-valued process based on (9.1.4), derive their generators in a special case for the function $\tilde{\sigma}$, and show that in the limit for n we have convergence to the Fleming-Viot process with fertility selection.

9.2 The particle process and the associated measure-valued process

In this section we construct an E^n-valued Markov jump process, whose transitions are based on the new predictive density, and show that, with a particular choice for the introduced function $\tilde{\sigma}$, the generator of the associated process of empirical measures, in the infinite population limit, converges to that of a Fleming-Viot process with haploid fertility selection, i.e. when $\sigma \in B(E)$ as in (4.2.21). The neutral Fleming-Viot process, whose stationary distribution is Π_{θ,ν_0}, is recovered with the same rule that simplifies the new predictive density to the Dirichlet case.

Consider, as in the previous chapters, a vector of n particles, each of which has a type represented by a point in the space E. The transitions are as follows. A particle

x_i is selected with probability $\pi_i^n = 1/n$ for $1 \leq i \leq n$, and a random holding time is sampled from an exponential distribution of parameter $\lambda_{n,i} = \lambda_n(x_i)$. At the next transition, the ith particle is replaced with a random sample from the new predictive, that is

$$q_n(\mathrm{d}x|\, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = \frac{\theta_n\, \nu_n(\mathrm{d}x) + \sum_{l\neq i}^n \tilde{\sigma}_n(x_l)\, \delta_{x_l}(\mathrm{d}x)}{\theta_n + \sum_{l\neq i}^n \tilde{\sigma}_n(x_l)} \tag{9.2.1}$$

so that the next state is set to be $(x_1, \ldots, x_{i-1}, x_i^{'}, x_{i+1}, \ldots, x_n)$, where $x_i^{'}$ has density (9.2.1). Proceeding as in 5.2, the generator for this process is

$$\begin{aligned} A^n f(\mathbf{x}) =& \sum_{i=1}^n \pi_i^n\ \lambda_{n,i} \int \left[f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})\right] \left(\frac{\theta_n\, \nu_n(\mathrm{d}y) + \sum_{l\neq i}^n \tilde{\sigma}_n(x_l)\, \delta_{x_l}(\mathrm{d}y)}{\theta_n + \sum_{l\neq i}^n \tilde{\sigma}_n(x_l)}\right) \\ =& \sum_{i=1}^n \frac{\pi_i^n \lambda_{n,i}\ \theta_n}{\left(\theta_n + \sum_{l\neq i} \tilde{\sigma}_n(x_l)\right)} \int [f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})]\nu_n(\mathrm{d}y) \\ &+ \sum_{i=1}^n \sum_{l\neq i}^n \frac{\pi_i^n \lambda_{n,i}\ \tilde{\sigma}_n(x_l)}{\left(\theta_n + \sum_{l\neq i} \tilde{\sigma}_n(x_l)\right)} [f(\eta_i(\mathbf{x}|\, x_l)) - f(\mathbf{x})] \end{aligned}$$

where $\eta_i(\mathbf{x}|\, y) = (x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n)$. Letting now

$$\lambda_{n,i} = \frac{n\left(\theta_n + \sum_{l\neq i} \tilde{\sigma}_n(x_l)\right)}{2} \tag{9.2.2}$$

we obtain

$$\sum_{i=1}^n \frac{1}{2}\theta_n \int [f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})]\nu_n(\mathrm{d}y) + \sum_{1\leq l\neq i\leq n} \frac{1}{2}\tilde{\sigma}_n(x_l)[f(\eta_i(\mathbf{x}|\, x_l)) - f(\mathbf{x})].$$

Consider now a particular choice for $\tilde{\sigma}_n$, which is the special case we will concentrate exclusively on, that is

$$\tilde{\sigma}_n(x) = 1 + \frac{2}{n}\sigma(x)$$

where σ is a bounded nonnegative measurable function on E. This yields

$$\begin{aligned} A^n f(\mathbf{x}) = & \sum_{i=1}^{n} \frac{1}{2}\theta \int [f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})] \left(1 + \frac{2}{n}\sigma(y)\right) \nu_0(\mathrm{d}y) \\ & + \frac{1}{2} \sum_{1\le l\neq i\le n} [f(\eta_i(\mathbf{x}|\, x_l)) - f(\mathbf{x})] \\ & + \sum_{1\le l\neq i\le n} \frac{\sigma(x_l)}{n}[f(\eta_i(\mathbf{x}|\, x_l)) - f(\mathbf{x})]. \end{aligned} \tag{9.2.3}$$

Let now $\mu^{(m)}$, $m \le n$, be the probability measure on E^m defined by

$$\mu^{(m)} = \frac{1}{n(n-1)\dots(n-m+1)} \sum \delta_{(x_{i_1},\dots,x_{i_m})}$$

where the sum is taken over all choices of $1 \le i_1, \dots, i_m \le n$ with $i_k \neq i_l$. For $f \in B(E^n)$, define $\phi \in B(\mathscr{P}^n(E))$, where $\mathscr{P}^n(E) \subset \mathscr{P}(E)$ is the set of purely atomic probability measures on E whose atoms are multiples of $1/n$, by

$$\phi(\mu) = \langle f, \mu^{(n)} \rangle \tag{9.2.4}$$

and $\mathbb{A}^n\phi$ by

$$\mathbb{A}^n\phi(\mu) = \langle A^n f, \mu^{(n)} \rangle; \tag{9.2.5}$$

see Section 4.3.2. Then the generator for the process of empirical measures in the

n-dimensional case is

$$\begin{aligned}
\mathbb{A}^n\phi(\mu) =& \int_{E^n} \sum_{i=1}^n B_i^n f \mathrm{d}\mu^{(n)} + \int_{E^n} \frac{1}{2} \sum_{1\le l\ne i\le n} (\Phi_{li} f - f)\mathrm{d}\mu^{(n)} \\
&+ \int_{E^n} \sum_{1\le l\ne i\le n} \frac{1}{n}\sigma_l(\cdot)(\Phi_{li} f - f)\mathrm{d}\mu^{(n)} \\
=& \sum_{i=1}^n \int_{E^n} B_i^n f \mathrm{d}\mu^{(n)} + \frac{1}{2} \sum_{1\le l\ne i\le n} \int_{E^n} (\Phi_{li} f - f)\mathrm{d}\mu^{(n)} \\
&+ \sum_{1\le l\ne i\le n} \int_{E^n} \frac{1}{n}\sigma_l(\cdot)(\Phi_{li} f - f)\mathrm{d}\mu^{(n)} \\
=& \sum_{i=1}^n \langle B_i^n f, \mu^{(n)}\rangle + \frac{1}{2} \sum_{1\le l\ne i\le n} \big(\langle \Phi_{li} f, \mu^{(n)}\rangle - \langle f, \mu^{(n)}\rangle\big) \\
&+ \frac{1}{n} \sum_{1\le l\ne i\le n} \big(\langle \sigma_l(\cdot)\Phi_{li} f, \mu^{(n)}\rangle - \langle \sigma_l(\cdot) f, \mu^{(n)}\rangle\big) \qquad (9.2.6)
\end{aligned}$$

where $\sigma_l(\cdot)$ denotes $\sigma(x_l)$ and

$$B^n f(x) = \frac{1}{2}\theta \int [f(z) - f(x)]\{1 + 2\sigma(z)/n\}\nu_0(\mathrm{d}z). \qquad (9.2.7)$$

We can now appreciate in (9.2.6) the role of σ in affecting the weight of the particle x_l to be selected as a replacement for the outgoing particle x_i. In population genetics such a σ is called fertility selection function.

Note now that if $f \in B(E^m)$, $m < n$, then

$$\begin{aligned}
\mathbb{A}^n\phi(\mu) =& \sum_{i=1}^m \langle B_i^n f, \mu^{(m)}\rangle + \frac{1}{2} \sum_{1\le l\ne i\le m} \langle \Phi_{li} f - f, \mu^{(m)}\rangle \\
&+ \frac{1}{n} \sum_{1\le l\ne i\le m} \big(\langle \sigma_l(\cdot)\Phi_{li} f, \mu^{(m)}\rangle - \langle \sigma_l(\cdot) f, \mu^{(m)}\rangle\big) \\
&+ \frac{n-m}{n} \sum_{i=1}^m \big(\langle \sigma_i(\cdot) f, \mu^{(m)}\rangle - \langle \sigma(\cdot)\otimes f, \mu^{(m+1)}\rangle\big)
\end{aligned}$$

where $\sigma(\cdot)\otimes f$ denotes $\sigma(x_{m+1})f(x_1,\ldots,x_m)$. Since, for large n, $\mu^{(m)}$ is essentially the product measure (cf. (7.2.10)), the limiting operator is

$$\begin{aligned}\mathbb{A}\phi(\mu) =& \sum_{i=1}^{m}\langle B_i f, \mu^m\rangle + \frac{1}{2}\sum_{1\le l\neq i\le m}\langle \Phi_{li} f - f, \mu^m\rangle \\ &+ \sum_{i=1}^{m}\left(\langle \sigma_i(\cdot) f, \mu^m\rangle - \langle \sigma(\cdot)\otimes f, \mu^{m+1}\rangle\right).\end{aligned} \tag{9.2.8}$$

where B^n has converged to B given by (5.2.8). $\mathbb{A}$ is the generator of the Fleming-Viot process with haploid fertility selection (cf. (4.2.21) and (4.3.30)). Observe that for $\sigma \equiv 0$, i.e. when there is no selection, we have $\tilde{\sigma}_n \equiv 1$ for every n, whence (9.2.7) reduces to (5.2.8), $\lambda_{n,i}$ in (9.2.2) reduces to λ_n in (5.2.5), and (9.2.8) reduces to

$$\mathbb{A}\phi(\mu) = \sum_{i=1}^{m}\langle A_i f, \mu^m\rangle + \frac{1}{2}\sum_{1\le l\neq i\le m}\langle \Phi_{li} f - f, \mu^m\rangle \tag{9.2.9}$$

recovering the generator of the neutral Fleming-Viot (cf. 5.3.4). This is coherent with the fact that for $\tilde{\sigma}_n \equiv 1$ we recover from (9.1.4) the predictive distribution for the Blakwell-MacQueen urn (5.1.7) and (9.1.3), which is associated with the Dirichlet process Π_{θ,ν_0}. Indeed Π_{θ,ν_0} is the stationary distribution of the process with transition function given by (5.1.7) and generator given by (5.3.4) (see Section 5.5).

9.3 Weak convergence

In this section we show that the process of empirical measure $\{\mu_{\mathbf{x}^n}(t), t \ge 0\}$ defined in Section 9.2 weakly converges to the Fleming-Viot process with generator (9.2.8), denoted by $\{\mu_t, t \ge 0\}$. Since the proof provided in Section 8.3 does not depend on the specific parametrisation of the model, there is no need to provide new proofs for the present one. Indeed, the key features on which all the results rely upon are the completeness and separability of the metric space E, the fact that the process is constant between jumps together with the Poisson rates, no matter their specific form, the convergence of the generators, showed in the previous section, and the well-posedness of the associated martingale problem.

Define $\mu_{\mathbf{x}^n} : E^n \times [0,\infty) \to \mathscr{P}(E)$ by

$$\mu_{\mathbf{x}^n}(t) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i^n(t)}.$$

that is $\mu_{\mathbf{x}^n}$ is the process of empirical measures of the X_i's.

Theorem 9.3.1 *Let $\{\mu_{\mathbf{x}^n}(t), t \geq 0\}$ be a process with values in $\mathscr{P}(E)$ as in the above construction, and let $\{\mu_t, t \geq 0\}$ be a Fleming-Viot process with generator* (9.2.8). *Assume that $P\mu_{\mathbf{x}^n}^{-1}(0) \Rightarrow \mu_0$, for $\mu_0 \in \mathscr{P}(E)$. Then*

$$\{\mu_{\mathbf{x}^n}(t), t \geq 0\} \Rightarrow \{\mu_t, t \geq 0\} \tag{9.3.1}$$

in $D_{\mathscr{P}(E)}[0,\infty)$, where $\Rightarrow$ means convergence in distribution in the Skorohod topology.

PROOF. The proof is the same as that of Theorem 8.3.1, except for the asymptotic Markovianity that is not needed here since the particle process is Markov. □

9.4 Stationary distribution

In this section we derive the stationary distribution of the Fleming-Viot process with haploid selection, with generator given by (9.2.8), by appealing to the Dirichlet process mixture model in hierarchical framework.

Theorem 9.4.1 *Let E be a locally compact complete separable metric space, and let $\{\mu_t, t \geq 0\}$ be the Fleming-Viot process on $\mathscr{P}(E)$ with generator given by (9.2.8). Then*

$$\Pi_\infty(\mathrm{d}\mu) = C \exp\left\{2\int \sigma(x)\,\mu(\mathrm{d}x)\right\}\Pi_{\theta,\nu_0}(\mathrm{d}\mu) \tag{9.4.1}$$

is the stationary distribution of $\{\mu_t, t \geq 0\}$, where Π_{θ,ν_0} denotes the Dirichlet process with parameters (θ, ν_0) and C is a constant.

Proof. Consider the Dirichlet process mixture model in hierarchical framework (see Section 2.4), by introducing a vector of variables $(y_1, \ldots, y_n)$ such that $y_i|\,x_i \sim K_n(\cdot|\,x_i)$, for some density K_n such that $K_n(1|\,x) = \tilde{\sigma}_n(x)$. Note that $K_n(1|\,x)$ is not to be regarded as a probability but rather as a density evaluated at 1. Also the value 1 is irrelevant, it could be any fixed number. Then further assume that

$$x_1, \ldots, x_n|\,\mu \overset{iid}{\sim} \mu$$

with $\mu \sim \Pi_{\theta,\nu_0}$. Now

$$p_n(y_1 = 1, \ldots, y_n = 1|\,x_1, \ldots, x_n) = \prod_{i=1}^n \tilde{\sigma}_n(x_i)$$

from which we obtain

$$p_n(x_1, \ldots, x_n|\,y_1 = 1, \ldots, y_n = 1) \propto p_n(x_1, \ldots, x_n) \prod_{i=1}^n \tilde{\sigma}_n(x_i)$$

Analogously to the previous chapters, since at each jump we are sampling from the full conditional density (9.1.4), we are effectively describing a Gibbs sampler, which ensures that the E^n-valued Markov chain embedded in the process at jump times is stationary, with stationary distribution q_n given by (9.1.1) (see Section 1.3). Note that the chain is not reversible, since, from (9.2.1), reversibility would imply that $\tilde{\sigma}$ be constant, i.e. there is no selection. Therefore we can consider the embedded chain as being a Gibbs sampler on

$$p_n(x_1, \ldots, x_n |\, y_1 = 1, \ldots, y_n = 1) = q_n(x_1, \ldots, x_n)$$

which is the stationary distribution of the process, since between consecutive jumps the process is constant.

Extending this Gibbs sampler to include the random distribution function, we would have a sampler over

$$(x_1, \ldots, x_n, \mu |\, y_1 = 1, \ldots, y_n = 1),$$

where at each point μ is sampled from $(\mu |\, x_1, \ldots, x_n, y_1 = 1, \ldots, y_n = 1)$. Hence, the marginal stationary distribution for the measure valued process is $(\mu |\, y_1 = 1, \ldots, y_n = 1)$. We can work out what this is. Now

$$p_n(y_1 = 1, \ldots, y_n = 1 | \mu) = \left\{ \int \tilde{\sigma}_n(x)\, \mu(\mathrm{d}x) \right\}^n$$

and so the stationary distribution for μ is

$$\Pi_n(\mathrm{d}\mu) \propto \left\{ \int \tilde{\sigma}_n(x)\, \mu(\mathrm{d}x) \right\}^n \Pi_{\theta,\nu_0}(\mathrm{d}\mu).$$

We now want to see what happens to Π_n as $n \to \infty$, since this will be the stationary distribution of the measure valued process corresponding to (9.2.8). This follows since $x_1, \ldots, x_n | \mu, y_1 = 1, \ldots, y_n = 1$ are i.i.d. from μ and $\mu \sim \Pi_n$. Note that μ is not the empirical measure of the sequence $x_1, \ldots, x_n$, but rather the random distribution function conditionally on which $x_1, \ldots, x_n$ are i.i.d.. It follows that the weak limit

$$\underset{n\to\infty}{\text{weak-lim}}\, \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \tag{9.4.2}$$

has distribution given by the limit of Π_n. See Aldous (1985).

Observe now that from Theorem 3.2.3, since E is separable and complete, so $(\mathscr{P}(E), \rho)$ is, where ρ is Prohorov metric (which induces the topology of weak convergence), and from and Lemma 3.2.5 every probability measure on a complete separable metric space is tight. Hence the sequence $\{\Pi_n, n \geq 1\} \subset \mathscr{P}(\mathscr{P}(E))$ is tight.

Letting now $n \to \infty$, and recalling that

$$\tilde{\sigma}_n(x) = 1 + \frac{2}{n}\sigma(x) \tag{9.4.3}$$

the limit of Π_n is

$$\Pi_\infty(\mathrm{d}\mu) = C \,\exp\left\{2\int \sigma(x)\,\mu(\mathrm{d}x)\right\} \Pi_{\theta,\nu_0}(\mathrm{d}\mu),$$

where C is a constant (cf. with (4.2.35)). That is, Π_∞ is the de Finetti measure of the infinite exchangeable sequence $x_1, x_2, \ldots$, conditional on $y_1 = 1, y_2 = 1, \ldots$, for each $t \geq 0$.

In addition, the $C_{\mathscr{P}(E)}[0,\infty)$ martingale problem for $\mathbb{G}$, given by (9.2.8), is well posed (see Theorem 4.2.1-(d); note that uniqueness of solution, not depending on the symmetry of σ, holds when the latter is univariate; cf. also Ethier and Kurtz (1987), pag. 77). From Lemma A.2 (see Appendix) it now follows that Π_∞ is the stationary distribution for the Fleming-Viot process with generator $\mathbb{G}$. □

Clearly, when $\sigma \equiv 0$ we recover the Dirichlet process.

Remark 9.4.2 Note that (9.4.1) cannot be deduced directly from (4.2.35) since here the function σ is asymmetric.

9.5 Discussion

In this chapter we introduced a class of predictive densities which generalise those of the Blackwell-MacQueen urn-scheme, in such a way that allows a non-uniform sampling from the empirical distribution via the function $\tilde{\sigma}_n(x)$. Then we considered a jump Markov process on E^n with transitions based on the predictive density; such a process, constructed following the same pattern of the previous chapters, has holding times which are exponential random variables with parameter that depends on the particle that will be replaced in the successive jump, and leads to show that the associated process of empirical measures in the limit for $n \to \infty$ converges to the Fleming-Viot process with haploid selection.

By considering an extension of the E^n-valued process to include a random distribution function and then looking at its distribution as n grows to infinity, we obtain the stationary distribution of this class of Fleming-Viot processes, as an application of the Dirichlet process mixture model in a hierarchical framework. In the case of identical to zero selection the results simplify coherently to obtain the Dirichlet process as the stationary distribution of the neutral Fleming-Viot diffusion.

An obviously interesting development is to investigate a bivariate selection function $\sigma \in B(E^2)$. The conjecture is that it can be treated by generalising the starting point of the proof of Theorem 1 in an appropriate way, possibly restricting to the case of symmetric functions, so that one can yield the stationary distribution (4.2.35) of Ethier and Kurtz (1994) for the associated Fleming-Viot process with diploid fertility selection, via some $\mathscr{P}(E)$-valued process. This will be the topic of the next chapter.

Chapter 10

Diploid fertility selection

In this chapter we reformulate the construction given in Chapter 9 for the case of bivariate symmetric selection intensity function, that is $\sigma \in B_{\text{sym}}(E^2)$. After defining a different generalisation of the Blackwell-MacQueen urn-scheme, we construct the appropriate E^n-valued particle process, and the associated measure-valued process is shown to converge to the Fleming-Viot process with diploid fertility selection, whose generator is (cf. (4.2.19))

$$\begin{aligned} &\frac{1}{2} \sum_{1 \le i \ne j \le m} \left(\langle \Phi_{ij} f, \mu^{m-1} \rangle - \langle f, \mu^m \rangle \right) + \sum_{j=1}^{m} \langle B_j f, \mu^m \rangle \\ &+ \sum_{j=1}^{m} \left(\langle \sigma_{j\cdot}(\cdot,\cdot) f, \mu^{m+1} \rangle - \langle \sigma(\cdot,\cdot) \otimes f, \mu^{m+2} \rangle \right). \end{aligned} \tag{10.0.1}$$

Again, appealing to the Dirichlet process mixture model leads to a simple derivation of the stationary distribution of the diffusion, which Ethier and Kurtz (1994) showed to be (see Section 4.2.2)

$$\Pi_\sigma(\mathrm{d}\mu) = C e^{\langle \sigma, \mu^2 \rangle} \Pi_{\theta,\nu_0}(\mathrm{d}\mu) \tag{10.0.2}$$

where $\langle \sigma, \mu^2 \rangle = \iint \sigma(x,y)\mu(\mathrm{d}x)\mu(\mathrm{d}y)$. Note that the topic of the present chapter does not constitute a proper generalisation of the univariate case of Chapter 9, since from $\sigma(x,y) \in B_{\text{sym}}(E^2)$ one cannot recover $\sigma(x) \in B(E)$ as a special case.

10.1 The underlying model

For each n even, let P_n denote a pairing of $\{1,\ldots,n\}$, such that given P_n, k is paired with j_k. Consider the following Dirichlet process mixture model in hierarchical framework:

$$\tilde{p}_n(y_1 = 1, \ldots, y_n = 1 \mid x_1, \ldots, x_n, P_n) = \prod_k n\, \tilde{\sigma}_n(x_k, x_{j_k}) \tag{10.1.1}$$

$$x_1, \ldots, x_n \mid \mu \overset{i.i.d.}{\sim} \mu$$

$$\mu \sim \Pi_{\theta,\nu_0}$$

$$P_n \sim \pi(P_n) \propto 1.$$

The product in (10.1.1), which is taken over $n/2$ terms that cover all pairs, is to be regarded as the density of the vector $y_1, \ldots, y_n$ computed at $(1, \ldots, 1)$, conditionally on $x_1, \ldots, x_n$ and on the pairing. The bounded symmetric function $\tilde{\sigma}_n(x,y) \in B_{\text{sym}}(E^2)$ is assumed to be chosen for each n. The vector $x_1, \ldots, x_n$ is exchangeable, that is $x_1, \ldots, x_n$ are i.i.d. μ conditionally on μ; μ is a random distribution function distributed as a Dirichlet process of parameters (θ, ν_0), denoted by Π_{θ,ν_0}. Last, $\pi(P_n)$ is the distribution of the pairing, assumed to be uniform.

From Section 2.2 we know that given a sample of size $n-1$ from a random distribution function which is a Dirichlet process prior, the predictive density for the next observation is

$$p_n(\mathrm{d}x_n|\, x_1, \ldots, x_{n-1}) = \frac{\theta\, \nu_0(\mathrm{d}x_n) + \sum_{k=1}^{n-1} \delta_{x_k}(\mathrm{d}x_n)}{\theta + n - 1}.$$

For this reason the notation $p_n(x_1, \ldots, x_n)$ will be used to denote the unconditional joint density of $x_1, \ldots, x_n$.

We are interested in constructing a Markov process which converges to the Fleming-Viot process with diploid selection, and derive its stationary distribution. Given the hierarchical model, in the next section a predictive distribution for the x's conditionally on the y's will be derived, which will be the transition density of the E^n-valued particle process we will define in Section 10.4. The hierarchical model will also play a central role in the derivation of the de Finetti measure of the infinite exchangeable sequence $x_1, x_2, \ldots$ conditionally on $y_1, y_2, \ldots$, in that the key step will be Gibb sampling the joint law of the x's and μ, conditionally on the y's, which will be done in Section 10.3. Finally it will remain to show that the so found de Finetti measure is the distribution, when n tends to infinity and for fixed $t \geq 0$, of the limiting empirical measure of the particle forming the E^n-valued Markov process, and is also the stationary distribution of the measure-valued process associated with the constructed particle process, that is of the Fleming-Viot process with diploid selection.

10.2 Conditional predictive density

Let $p_n(x_1, \ldots, x_n)$ be the exchangeable density associated with the Blackwell-MacQueen urn-scheme as above. The hierarchical model induces a generalisation of p_n via the functions $\tilde{\sigma}_n(x, y)$, by writing

$$q_n(x_1, \ldots, x_n, P_n | \mathbf{y} = \mathbf{1}) \propto \; p_n(x_1, \ldots, x_n) \prod_k \tilde{\sigma}_n(x_k, x_{j_k}) \tag{10.2.1}$$

where $\mathbf{y} = \mathbf{1}$ denotes $y_1 = 1, \ldots, y_n = 1$, the conditioning on which will be from now on implicit. Removing one element of the vector $x_1, \ldots, x_n$, say x_i, then the predictive, jointly with P_n, is

$$q_n(\mathrm{d}x_i, P_n | \mathbf{x}_{-i}) \propto \; p_n(\mathrm{d}x_i | \mathbf{x}_{-i}) \; \tilde{\sigma}_n(x_i, x_{j_i})$$

where $\mathbf{x}_{-i}$ denotes $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$. Now

$$q_n(P_n | \mathbf{x}_{-i}) \propto \int p_n(\mathrm{d}x_i | \mathbf{x}_{-i}) \; \tilde{\sigma}_n(x_i, x_j);$$

and since from (10.2.1) we can write

$$q_n(\mathrm{d}x_i | \mathbf{x}_{-i}, P_n) = \frac{p_n(\mathrm{d}x_i | \mathbf{x}_{-i}) \; \tilde{\sigma}_n(x_i, x_{j_i})}{\int p_n(\mathrm{d}x_i | \mathbf{x}_{-i}) \; \tilde{\sigma}_n(x_i, x_{j_i})}$$

we obtain

$$\begin{aligned} q_n(\mathrm{d}x_i | \mathbf{x}_{-i}) &\propto \sum_{j \neq i} q_n(P_n | \mathbf{x}_{-i}) \frac{p_n(\mathrm{d}x_i | \mathbf{x}_{-i}) \; \tilde{\sigma}_n(x_i, x_j)}{\int p_n(\mathrm{d}x_i | \mathbf{x}_{-i}) \; \tilde{\sigma}_n(x_i, x_j)} \\ &\propto p_n(\mathrm{d}x_i | \mathbf{x}_{-i}) \sum_{j \neq i}^{n} \tilde{\sigma}_n(x_i, x_j). \end{aligned}$$

Thus we have

$$q_n(\mathrm{d}x_i | \mathbf{x}_{-i}) = \frac{p_n(\mathrm{d}x_i | \mathbf{x}_{-i}) \sum_{j \neq i}^{n} \tilde{\sigma}_n(x_i, x_j)}{\int p_n(\mathrm{d}x_i | \mathbf{x}_{-i}) \sum_{j \neq i}^{n} \tilde{\sigma}_n(x_i, x_j)}. \tag{10.2.2}$$

When p_n is derived from the Dirichlet process prior, the predictive for x_i is

$$p_n(\mathrm{d}x | \mathbf{x}_{-i}) = \frac{\theta \, \nu_0(\mathrm{d}x) + \sum_{k \neq i}^{n} \delta_{x_k}(\mathrm{d}x)}{\theta + n - 1} \tag{10.2.3}$$

and (10.2.2) can be written,

$$q_n(\mathrm{d}x_i|\,\mathbf{x}_{-i}) = \frac{\theta\sum_{j\neq i}^n \tilde{\sigma}_n(x_i,x_j)\;\nu_0(\mathrm{d}x_i) + \sum_{k\neq i}^n\sum_{j\neq i}^n\;\tilde{\sigma}_n(x_i,x_j)\;\delta_{x_k}(\mathrm{d}x_i)}{\int\Big(\theta\sum_{j\neq i}^n \tilde{\sigma}_n(x_i,x_j)\;\nu_0(\mathrm{d}x_i) + \sum_{k\neq i}^n\sum_{j\neq i}^n\;\tilde{\sigma}_n(x_i,x_j)\;\delta_{x_k}(\mathrm{d}x_i)\Big)}$$
$$= \frac{\theta_n\,\nu_n(\mathrm{d}x_i) + \sum_{k\neq i}^n\sum_{j\neq i}^n \tilde{\sigma}_n(x_i,x_j)\,\delta_{x_k}(\mathrm{d}x_i)}{\theta_n + \sum_{k\neq i}^n\sum_{j\neq i}^n \tilde{\sigma}_n(x_k,x_j)} \tag{10.2.4}$$

where θ_n and ν_n denote

$$\theta_n = \theta\int\sum_{j\neq i}^n \tilde{\sigma}_n(x_i,x_j)\,\nu_0(\mathrm{d}x_i) \tag{10.2.5}$$

and

$$\nu_n(\mathrm{d}x_i) = \frac{\sum_{j\neq i}^n \tilde{\sigma}_n(x_i,x_j)\,\nu_0(\mathrm{d}x_i)}{\int\sum_{j\neq i}^n \tilde{\sigma}_n(x_i,x_j)\,\nu_0(\mathrm{d}x_i)}. \tag{10.2.6}$$

The predictive (10.2.4) will be the transition density of the Markov particle process defined in Section 10.4, and the full conditional distribution at the base of the Markov chain constructed via a Gibbs sampler in the next section.

Observe in (10.2.4) that a larger $\tilde{\sigma}_n$ implies a larger probability for the the first coordinate of being selected to update x_i, that means the larger the $\tilde{\sigma}_n$ the higher the fitness of the individual who is going to have an offspring. In population genetics such function describes the intensity of fertility selection. When $\tilde{\sigma}_n(x,y)\equiv 1$ for all n, we recover the Dirichlet case, that is (10.2.3).

Note that from (10.2.1) it is also possible to derive the distribution of the pairing. Indeed

$$q_n(P_n) \propto \int p_n(x_1,\ldots,x_n)\prod_k \tilde{\sigma}_n(x_k,x_{j_k})\,\mathrm{d}x_1\ldots\mathrm{d}x_n$$

from which also emerge the key role of the selection function. A pair with higher fitness will give a higher value of $\tilde{\sigma}_n$, so that those individuals which are fitter when paired will increase the probability of that specific pair occurring.

10.3 Gibbs sampling the model

Consider now a Gibbs sampler algorithm (see Section 1.3) implemented on

$$(x_1, \ldots, x_n, \mu)$$

where at each iteration $x_1, \ldots, x_n$ are sampled from the full conditionals (10.2.4), that is $q_n(\mathrm{d}x_i | \mathbf{x}_{-i})$, and μ is sampled from the Dirichlet posterior $\Pi_{\theta,\nu_0}(\cdot | x_1, \ldots, x_n)$.

The stationary distribution of the E^n-valued Markov chain generated by $x_1, \ldots, x_n$ is given by $q_n(x_1, \ldots, x_n)$. Further, since

$$\tilde{p}_n(y_1 = 1, \ldots, y_n = 1 | \mu, P_n) = \left\{ n \iint \tilde{\sigma}_n(x, y)\mu(\mathrm{d}x)\mu(\mathrm{d}y) \right\}^{n/2}$$

which does not depend on P_n, the stationary distribution of the chain of random distribution functions is

$$\Pi_n(\mathrm{d}\mu) \propto \left\{ n \iint \tilde{\sigma}_n(x, y)\mu(\mathrm{d}x)\mu(\mathrm{d}y) \right\}^{n/2} \Pi_{\theta,\nu_0}(\mathrm{d}\mu). \tag{10.3.1}$$

Note that $\mathscr{P}(E)$ with Prohorov's metric is separable and complete, so that $\{\Pi_n, n \geq 1\}$ is tight. If we now put

$$\tilde{\sigma}_n(x, y) = \frac{1}{n} + \frac{2}{n^2}\sigma(x, y)$$

and take the limit as $n \to \infty$, we obtain

$$\Pi_\infty(\mathrm{d}\mu) \propto \exp\left\{ \iint \sigma(x, y)\mu(\mathrm{d}x)\mu(\mathrm{d}y) \right\} \Pi_{\theta,\nu_0}(\mathrm{d}\mu)$$

which is the stationary distribution of the chain of random distribution functions when the sample size grows to infinity, and is also the de Finetti measure of the infinite exchangeable sequence $x_1, x_2, \ldots$.

10.4 The particle process and the associated measure-valued process

In this section we construct an E^n-valued Markov particle process based on (10.2.4) and an associated $\mathscr{P}(E)$-valued process, derive their respective generators in a special case for the function $\tilde{\sigma}_n$, and show that in the limit for large n the latter converges to the generator of the Fleming-Viot process with diploid fertility selection.

Consider a vector of n particles. Instantaneously after each transition, a particle x_i is uniformly selected, for $1 \leq i \leq n$, and a holding time is sampled from an exponential distribution of parameter $\lambda_{n,i} = \lambda_n(x_i)$. At the next transition, the ith particle is replaced with a random sample from (10.2.4). The process is clearly Markov. Note that there is a Markov chain embedded at jump times, and since the transition densities are given by $q_n(\mathrm{d}x_i|\,\mathbf{x}_{-i})$, the chain is otherwise obtained by implementing a Gibbs sampler on $q_n(x_1,\ldots,x_n)$, of which (10.2.4) is the full conditional. This ensures that q_n is the stationary distribution of either the E^n-valued chain and, given that between jumps the vector is constant, the process.

The generator of the E^n-valued process is

$$
\begin{aligned}
A^n f(\mathbf{x}) =& \sum_{i=1}^{n} \pi_i^n \,\lambda_{n,i} \int \big[f(\eta_i(\mathbf{x}|\,y)) - f(\mathbf{x})\big] \\
&\times \left(\frac{\theta_n\,\nu_n(\mathrm{d}y) + \sum_{k\neq i}^{n}\sum_{j\neq i}^{n} \tilde{\sigma}_n(y,x_j)\,\delta_{x_k}(\mathrm{d}y)}{\theta_n + \sum_{k\neq i}^{n}\sum_{j\neq i}^{n}\tilde{\sigma}_n(x_k,x_j)} \right) \\
=& \sum_{i=1}^{n} \frac{\lambda_{n,i}\,\theta_n}{n\left(\theta_n + \sum_{k\neq i}^{n}\sum_{j\neq i}^{n}\tilde{\sigma}_n(x_k,x_j)\right)} \int [f\eta_i(\mathbf{x}|\,y) - f(\mathbf{x})]\nu_n(\mathrm{d}y) \\
&+ \sum_{i=1}^{n}\sum_{k\neq i}^{n}\sum_{j\neq i}^{n} \frac{\lambda_{n,i}\,\tilde{\sigma}_n(x_k,x_j)}{n\left(\theta_n + \sum_{k\neq i}^{n}\sum_{j\neq i}^{n}\tilde{\sigma}_n(x_k,x_j)\right)} [f\eta_i(\mathbf{x}|\,x_k) - f(\mathbf{x})]
\end{aligned}
$$

where $\eta_i(\mathbf{x}|z) = (x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n)$. If we let the Poisson rates be

$$\lambda_{n,i} = \frac{n\left(\theta_n + \sum_{k\neq i}^n \sum_{j\neq i}^n \tilde{\sigma}_n(x_k, x_j)\right)}{2} \tag{10.4.1}$$

we obtain

$$\sum_{i=1}^n \frac{1}{2}\,\theta_n \int [f(\eta_i(\mathbf{x}|\, y)) - f(\mathbf{x})]\nu_n(\mathrm{d}y) + \sum_{1\leq k\neq i\neq j\leq n} \frac{1}{2}\,\tilde{\sigma}_n(x_k, x_j)[f(\eta_i(\mathbf{x}|\, x_k)) - f(\mathbf{x})]. \tag{10.4.2}$$

Consider now a particular choice for $\tilde{\sigma}_n$ given earlier, i.e.

$$\tilde{\sigma}_n(x, y) = \frac{1}{n} + \frac{2}{n^2}\,\sigma(x, y) \tag{10.4.3}$$

where σ is a bounded symmetric function on E^2. Note that if $\sigma(x, y) \equiv 0$, then (10.2.5) reduces to θ, (10.2.6) to ν_0 (cf. (10.2.3)) and $\lambda_{n,i}$ to $n(\theta + n - 1)/2$, as in the neutral case (cf. Chapter 5). Using also (10.2.5), (10.2.6) and (10.4.3) in (10.4.2) yields

$$\begin{aligned}
A^n f(\mathbf{x}) =& \sum_{i=1}^n \sum_{j\neq i}^n \frac{1}{2}\theta \int [f\eta_i(\mathbf{x}|\, y) - f(\mathbf{x})] \left(\frac{1}{n} + \frac{2\sigma(y, x_j)}{n^2}\right) \nu_0(\mathrm{d}y) \\
&+ \frac{1}{2n} \sum_{1\leq k\neq i\neq j\leq n} [f\eta_i(\mathbf{x}|\, x_k) - f(\mathbf{x})] \\
&+ \frac{1}{n^2} \sum_{1\leq k\neq i\neq j\leq n} \sigma(x_k, x_j)[f\eta_i(\mathbf{x}|\, x_k) - f(\mathbf{x})] \\
=& \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n B_i^{n,\sigma_j} f(\mathbf{x}) \\
&+ \frac{1}{2} \sum_{1\leq k\neq i\leq n} [f\eta_i(\mathbf{x}|\, x_k) - f(\mathbf{x})] \\
&+ \frac{1}{n^2} \sum_{1\leq k\neq i\neq j\leq n} \sigma(x_k, x_j)[f\eta_i(\mathbf{x}|\, x_k) - f(\mathbf{x})]
\end{aligned}$$

where

$$B^{n,\sigma_j} f(x) = \frac{1}{2}\theta \int [f(y) - f(x)] \left(1 + \frac{2\sigma(y, x_j)}{n}\right) \nu_0(\mathrm{d}y)$$

and B_i^{n,σ_j} is the operator B^{n,σ_j} applied to the i-th coordinate.

Define now for $m \leq n$ the probability measure on E^m

$$\mu^{(m)} = \frac{1}{n(n-1)\dots(n-m+1)} \sum_{1\leq i_1 \neq \dots \neq i_m \leq n} \delta_{(x_{i_1},\dots,x_{i_m})}$$

and for $f \in B(E^n)$

$$\phi(\mu) = \langle f, \mu^{(n)} \rangle$$

and

$$\mathbb{A}^n \phi(\mu) = \langle A^n f, \mu^{(n)} \rangle$$

where $\langle f, \mu \rangle = \int f \mathrm{d}\mu$. Then the generator for the process of empirical measures in the n-dimensional case is

$$\begin{aligned}\mathbb{A}^n \phi(\mu) = &\frac{1}{n} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \langle B_i^{n,\sigma_j} f, \mu^{(n)} \rangle \\ &+ \frac{1}{2} \sum_{1\leq k \neq i \leq n} \langle \Phi_{ki} f - f, \mu^{(n)} \rangle \\ &+ \frac{1}{n^2} \sum_{1\leq k\neq i \neq j \leq n} \langle \sigma_{k,j}(\cdot,\cdot)(\Phi_{ki} f - f), \mu^{(n)} \rangle\end{aligned}$$

where $\sigma_{k,j}(\cdot,\cdot)$ denotes $\sigma(x_k, x_j)$, and $\Phi_{ki} f$ is the function of f where the coordinate at level k has replaced the coordinate at level i.

Observe now that for $f \in B(E^m)$, $m < n$,

$$\sum_{i=m+1}^{n} \sum_{j\neq i}^{n} \langle B_i^{n,\sigma_j} f, \mu^{(m)} \rangle = 0,$$

$$\sum_{i=m+1}^{n} \sum_{k\neq i}^{n} \langle \Phi_{ki} f - f, \mu^{(n)} \rangle = 0$$

and

$$\sum_{i=m+1}^{n} \sum_{k\neq i}^{n} \sum_{j\neq i}^{n} \langle \sigma_{k,j}(\cdot,\cdot)(\Phi_{ki} f - f), \mu^{(n)} \rangle$$

given that in all cases x_i is not an argument of f and thus f does not change, that is $\Phi_{ki} f = f$, and also

$$\sum_{i=1}^{m} \sum_{k=m+1}^{n} \langle \Phi_{ki} f - f, \mu^{(n)} \rangle = 0$$

given that

$$\langle \Phi_{ki} f, Z_n^{(n)} \rangle = \langle f, Z_n^{(n)} \rangle$$

when x_k of f. Hence, when $f \in B(E^m)$, $m < n$,

$$\begin{aligned}
\mathbb{A}^n \phi(\mu) = & \frac{1}{n} \sum_{i=1}^{m} \sum_{j\neq i}^{m} \langle B_i^{n,\sigma_j} f, \mu^{(m)} \rangle \\
& + \frac{n-m}{n} \sum_{i=1}^{m} \langle B_i^{n,\sigma_{m+1}} f, \mu^{(m+1)} \rangle \\
& + \frac{1}{2} \sum_{1\leq k\neq i\leq m} \langle \Phi_{ki} f - f, \mu^{(m)} \rangle \\
& + \frac{1}{n^2} \sum_{1\leq k\neq i\neq j\leq m} \big(\langle \sigma_{kj}(\cdot,\cdot) \Phi_{ki} f, \mu^{(m)} \rangle - \langle \sigma_{kj}(\cdot,\cdot) f, \mu^{(m)} \rangle \big) \\
& + \frac{n-m}{n^2} \sum_{i=1}^{m} \sum_{j\neq i}^{m} \big(\langle \sigma_{ij}(\cdot,\cdot) f, \mu^{(m)} \rangle - \langle \sigma_{\cdot j}(\cdot,\cdot) f, \mu^{(m+1)} \rangle \big) \\
& + \frac{n-m}{n^2} \sum_{1\leq k\neq i\leq m} \big(\langle \sigma_{k\cdot}(\cdot,\cdot) \Phi_{ki} f, \mu^{(m+1)} \rangle - \langle \sigma_{k\cdot}(\cdot,\cdot) f, \mu^{(m+1)} \rangle \big) \\
& + \frac{(n-m)^2}{n^2} \sum_{i=1}^{m} \big(\langle \sigma_{i\cdot}(\cdot,\cdot) f, \mu^{(m+1)} \rangle - \langle \sigma(\cdot,\cdot) \otimes f, \mu^{(m+2)} \rangle \big)
\end{aligned} \tag{10.4.4}$$

where

$$\sigma_{h\cdot}(\cdot,\cdot) f = \sigma(x_h, x_{m+1}) f(x_1, \ldots, x_m)$$

and

$$\sigma(\cdot,\cdot) \otimes f = \sigma(x_{m+1}, x_{m+2}) f(x_1, \ldots, x_m).$$

Note that in the fifth term we have σ_{ij} since with the operator Φ_{ki} the ith argument of f is now x_k, which is also the difference between the two f's in the last term, which justifies the different dimension of integration.

Given now that

$$B^{n,\sigma_j} f(x) = \frac{1}{2}\theta \int [f(y) - f(x)] \left(1 + \frac{2\sigma(y, x_j)}{n}\right) \nu_0(\mathrm{d}y)$$

converges to

$$Bf(x) = \frac{1}{2}\theta \int [f(y) - f(x)] \nu_0(\mathrm{d}y)$$

we have that

$$\langle B_i^{n,\sigma_{m+1}} f, \mu^{(m+1)} \rangle$$

converges to

$$\langle B_i f, \mu^{(m+1)} \rangle = \langle B_i f, \mu^{(m)} \rangle$$

due to the fact that the $(m+1)$-th dimension in (10.4.5) vanishes in the limit.

Since in addition, for large n, $\mu^{(m)}$ is essentially the product measure, the limiting operator is

$$\begin{aligned} \mathbb{A}\phi(\mu) = & \sum_{i=1}^{m} \langle B_i f, \mu^m \rangle + \frac{1}{2} \sum_{1 \leq k \neq i \leq m} \langle \Phi_{ki} f - f, \mu^m \rangle \\ & + \sum_{i=1}^{m} \left(\langle \sigma_{i\cdot}(\cdot, \cdot) f, \mu^{(m+1)} \rangle - \langle \sigma(\cdot, \cdot) \otimes f, \mu^{(m+2)} \rangle \right) \end{aligned} \tag{10.4.5}$$

which is the generator of the Fleming-Viot process with diploid fertility selection, that is (10.0.1). See Section 9.3 for relative compactness conditions and weak convergence.

Observe that for $\sigma \equiv 0$, i.e. when there is no selection, (10.4.5) reduces to the generator of the neutral Fleming-Viot, whose stationary distribution is the Dirichlet process. This is coherent with the generalisation of the predictive distribution described in Section 10.2, since $\sigma \equiv 0$ reduces the new predictive to the the Blackwell-MacQueen case.

10.5 Stationary distribution

As stated in the introduction, we expect the measure-valued process with generator (10.4.5) to have stationary distribution given by (10.0.2), as showed by Ethier and Kurtz (1994). In this section we provide a different proof of this result, based on the construction of the previous sections. In Section 10.3 the use of the Gibbs sampler enabled us to elicit the stationary distribution of the chain of random distribution functions. What remains to do is to connect the de Finetti measure of the sequence with the empirical measure of the particles, when the population size goes to infinity.

Theorem 10.5.1 *Let E be a locally compact complete separable metric space, and let $\{\mu_t, t \geq 0\}$ be the Fleming-Viot process on $\mathscr{P}(E)$ with generator given by* (10.4.5)*. Then*

$$\Pi_\infty(\mathrm{d}\mu) = C \exp\left\{\int_{E^2} \sigma(x, y)\, \mu(\mathrm{d}x)\mu(\mathrm{d}y)\right\} \Pi_{\theta,\nu_0}(\mathrm{d}\mu) \tag{10.5.1}$$

is the stationary distribution of $\{\mu_t, t \geq 0\}$, where Π_{θ,ν_0} denotes the Dirichlet process with parameters (θ, ν_0) and C is a constant.

Proof. Since the transition density of the E^n-valued particle process is given by (10.2.4), it follows that $x_1, \ldots, x_n |\, \mu, y_1 = 1, \ldots, y_n = 1$ are i.i.d. μ and $\mu \sim \Pi_n$, where Π_n is (10.3.1), from which

$$\underset{n\to\infty}{\text{weak-lim}}\,\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$$

has distribution Π_∞ (see for example Aldous (1985)). That is, for fixed $t \geq 0$ the limiting distribution for large n of the empirical measure of the particles $x_1, \ldots, x_n$ is given by the de Finetti measure of the infinite exchangeable sequence $x_1, x_2, \ldots$ conditional on $y_1 = 1, y_2 = 1, \ldots$. Also, as noted in Section 10.4, the measure valued process is constant between two consecutive jumps of the particle process, and this implies that Π_n is the de Finetti measure of the particles for each $t \geq 0$.

Further, since the $C_{\mathscr{P}(E)}[0,\infty)$ martingale problem for (10.4.5) is well posed (cf. Theorem 4.2.1-(d)), the result follows from from Lemma A.2 (see Appendix). □

Clearly, when $\sigma(x, y) \equiv 0$ we recover the Dirichlet process.

10.6 Discussion

In this chapter we extended the construction of Chapter 9 to the case of bivariate symmetric selection intensity function, from which, as observed in Remark 9.4.2, one cannot directly reduce to the former. The pattern has been essentially the same, namely introducing a class of predictives which generalise those of the Blackwell-MacQueen urn-scheme, defining a jump Markov process on E^n with transitions based on the predictive density, and constructing the associated process of empirical measures, which is shown to converge to the Fleming-Viot process with diploid fertility selection.

The Dirichlet process mixture model with hierarchy proves again to be a useful tool in the derivation of the stationary distribution of the Fleming-Viot process with selection, which in this case had been shown by Ethier and Kurtz (1994).

Appendix

Stationary distributions

Let $A \subset B(E) \times B(E)$ and suppose the martingale problem for A is well posed.

Definition A.1 (Stationary distribution)

$\mu \in \mathscr{P}(E)$ is a stationary distribution for A if every solution X of the martingale problem for (A, μ) is a stationary process. That is, if

$$P\{X(t+s_1) \in \Gamma_1, \ldots, X(t+s_k) \in \Gamma_k\}$$

is independent of $t \geq 0$ for all $k \geq 1$, $0 \leq s_1 < \cdots < s_k$ and $\Gamma_1, \ldots \Gamma_k \in \mathscr{B}(E)$.

The following lemma, from Ethier and Kurtz (1986), shows that to check that μ is a stationary distribution it is sufficient to consider only the one-dimensional distributions.

Lemma A.2 (Conditions for stationary distribution)

Let $A \subset B(E) \times B(E)$ and suppose the martingale problem for A is well posed. Let

$\mu \in \mathscr{P}(E)$ and let X be a solution of the martingale problem for (A, μ). Then μ is a stationary distribution for A if and only if $X(t)$ has distribution μ for all $t \geq 0$.

Proof. The necessity is immediate. For sufficiency observe that $X_t \equiv X(t + \cdot)$ is a solution of the martingale problem for (A, μ), and hence, by uniqueness, has the same finite-dimensional distributions as X. □

Conclusions

This work introduces countable representations of several types of Fleming-Viot process; such representations are defined as E^n-valued jump processes whose transitions are based on different urn schemes. Via the associated processes of empirical measures, and in one case with a construction that exploits a posterior Dirichlet representation, results on weak convergence and stationary distributions are derived in each case covered.

If, on the one hand, a limit of the models presented is that they are not suited for ancestry analysis, as are for example the models of Donnelly and Kurtz (1999a), on the other hand the arguments used in the constructions and proofs, in our opinion, present some interesting and promising aspects that differ from the existent literature on the matter.

The first is the use of urn models to construct the particle process. The only attempt in this sense to our knowledge is found in Ethier and Kurtz (1992), where the authors use a Pólya-type urn to derive, via bounded convergence of martingales, the stationary distribution of the neutral diffusion model. In our case urns are used extensively, and their explanatory power cover, from a population genetics viewpoint, neutrality, viability selection, haploid and diploid fertility selection, and can be considered as pillars of the whole building. The resulting framework is overall simple and the theoretical tools needed to yield the results are not extremely sophisticated. This points out the versatility of urn schemes in modeling and analyzing stochastic models, and particularly their usefulness for stylizing complex systems in such a way that enables to derive their properties in an easy and intuitive way. This work thus suggests that the "field of attraction" of urn models reaches population genetics processes and measure-valued diffusions.

The second is the constructive approach. Instead of starting from a generator which is considered suitable for the model one wants to describe, and then solving the martingale problem for it, in a sort of "backward" approach which is spread in the literature, we have adopted a "forward" or constructive approach, such that we first construct the particle process, then we compute its generator and pass to the process of empirical measures (or posterior representations) and finally we derive the stationary distribution. If one can argue that this is seemingly an *ad hoc* procedure, because in the definition of the particle process one has obviously in mind the desired goal, the type of constructions and results provided testify in the opposite direction, since at least half of them use well known urn processes and yield well known prior distributions, and, perhaps more importantly, all of them perfectly match n-dimensional measure-valued processes which already exist in the literature and converge to Fleming-Viot processes. That is, between the urn scheme and the stationary distribution that the urn scheme suggests there is a compulsory check point consisting in whether the defined process matches or not a known Fleming-Viot model. The whole work is obviously based on this.

The third is the Bayesian context. By means of the urn models, the arguments provided show that in a Bayesian nonparametric framework some known and new results in population genetics can be obtained, once the model is defined, by simply appealing to the de Finetti's representation theorem, as seen in all sections about stationarity. This means that the emphasis is always on the random probability measure underlying the exchangeable sequence, rather than on the empirical measure, as it happens in the literature. It is enough to think for instance at the proof of Theorem 9.4.1 or the whole construction of Chapter 6, where instead of considering the empirical measure of the particles and trying to elicit its distribution, the attention is pointed toward the distribution of the random measure which governs the urn process, thus deviating from the orthodox approach and joining it again with a weak limit argument. The intrinsic difficulty of finding the distribution of an empirical measure could be a reason why some of the representations of Fleming-Viot models in the literature still have unknown stationary distributions. By contrast, the methodology adopted in the present work enables to find the stationary distribution of each different introduced process. A Bayesian nonparametric approach helps simplifying the techniques needed for the proofs, and hence constitutes a beneficial and profitable setting in which to embed a stochastic model such as those discusses.

Bibliography

Aldous, D. J. (1985). Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour XIII-1983*, volume 1117 of *Lecture Notes in Mathematics*, pages 1–198. Springer-Verlag, Berlin.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.*, **2**, 1152–1174.

Besag, J. and Green, P. J. (1993). Spatial statistics and Bayesian computation. *J. Royal Stat. Soc. Ser. B*, **55**, 25–37.

Billingsley, P. (1968). *Convergence of probability measures*. Wiley, New York.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Stat* ., **1**, 353–355.

Cinlar, E. (1975). Markov renewal theory: a survey. *Management Science*, **21**, 727–752.

Crow, J. F. and Kimura, M. (1956). Some genetic problems in natural populations. In *Proc. Third Berkeley Symp. Math. Statist. Probab.*, volume 4, pages 1–22. Univ. California Press, Berkeley.

Dawson, D. A. and Hochberg, K. J. (1982). Wandering random measures in the Fleming-Viot model. *Ann. Probab.*, **10**, 554–580.

Donnelly, P. and Kurtz, T. G. (1996). A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Probab.*, **24**, 698–742.

Donnelly, P. and Kurtz, T. G. (1999a). Genealogical processes for Fleming-Viot models with selection and recombination. *Ann. Appl. Probab.*, **9**, 1091–1148.

Donnelly, P. and Kurtz, T. G. (1999b). Particle representation for measure-valued population models. *Ann. Probab.*, **27**, 166–205.

Dynkin, E. B. (1965). *Markov Processes, I*. Springer, Berlin.

Dynkin, E. B. (1994). *An Introduction to Branching Measure-valued Processes*. American Mathematical Society, Providence, RI.

Eggenberger, F. and Pólya, G. (1923). Über die statistik verketteter vorgänge. *Zeitschrift für angewandte mathemathik und mechanik*, **3**, 279–289.

Escobar, M. D. (1988). *Estimating the means of several normal populations by nonparametric estimation of the distribution of the means*. Ph.D. thesis, Department of Statistics, Yale University.

Escobar, M. D. (1994). Estimating normal means with a Dirichlet process prior. *J. Amer. Statist. Assoc.*, **89**, 268–277.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.*, **90**, 57–588.

Etheridge, A. M. (2000). *An Introduction to Superprocesses*. American Mathematical Society, Providence, RI.

Ethier, S. N. and Griffiths, R. C. (1987). The infinitely-many-sites model as a measure-valued diffusion. *Ann. Probab.*, **15**, 515–545.

Ethier, S. N. and Griffiths, R. C. (1993). The transition function of a Fleming-Viot process. *Ann. Probab.*, **21**, 1571–1590.

Ethier, S. N. and Kurtz, T. G. (1981). The infinitely-many-neutral-alleles diffusion model. *Adv. in Appl. Probab.*, **13**, 429–452.

Ethier, S. N. and Kurtz, T. G. (1986). *Markov processes: characterization and convergence*. Wiley, New York.

Ethier, S. N. and Kurtz, T. G. (1987). The infinitely-many-alleles model with selection as a measure-valued diffusion. In M. Kimura, G. Kallianpur, and T. Hida, editors, *Stochastic Models in Biology*, volume 70 of *Lecture Notes in Biomathematics*, pages 72–86. Springer-Verlag, Berlin.

Ethier, S. N. and Kurtz, T. G. (1992). On the stationary distribution of the neutral diffusion model in population genetics. *Ann. Appl. Probab.*, **2**, 24–35.

Ethier, S. N. and Kurtz, T. G. (1993). Fleming-Viot processes in population genetics. *Siam J. Control and Optimisation*, **31**, 345–386.

Ethier, S. N. and Kurtz, T. G. (1994). Convergence to Fleming-Viot processes in the weak atomic topology. *Stoch. Proc. and Their Applications*, **54**, 1–27.

Ethier, S. N. and Shiga, T. (2000). A Fleming-Viot process with unbounded selection. *J. Math. Kyoto Univ.*, **40**, 337–361.

Ewens, J. W. (1979). *Mathematical Population Genetics*. Springer-Verlag, New York.

Ewens, J. W. (1988). Population genetics thoery - the past and the future. In *Mathematical and statistical problems in evolution*. University of Montréal press, Montréal.

Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *Ann. Statist.*, **1**, 209–230.

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In H. Rizvi and J. Rustagi, editors, *Recent advances in statistics*. New York Academic Press.

Fleming, W. H. and Viot, M. (1979). Some measure-valued Markov processes in population genetics theory. *Indiana University Mathematics J.*, **28**, 817–843.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, **85**, 398–409.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, **6**, 721–741.

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian nonparametrics*. Springer, New York.

Goldberg, S. (1950). *On a singular diffusion equation*. Ph.D. thesis, Cornell Univ.

Griffiths, R. C. (1979). A transition density expansion for a multi-allele diffusion model. *Adv. in Appl. Probab.*, **11**, 310–325.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.*, **96**, 161–173.

Ishwaran, H. and James, L. F. (2002). Approximate Dirichlet process computing in finite normal mixtures: smoothing and prior information. *J. Comput. Graph. Statist.*, **11**, 508–533.

Ishwaran, H. and Zarepour, M. (2000). Markov Chain Monte Carlo in approximate Dirichlet and beta two parameter process hierarchical models. *Biometrika*, **87**, 371–390.

Ishwaran, H. and Zarepour, M. (2003). Random probability measures via Pólya sequences: revisiting the Blackwell-MacQueen urn scheme. *ArXiv:math.PR/030904, http://uk.arxiv.org/PS_cache/math/pdf/0309/0309041.pdf*.

Kallenberg, O. (1997). *Foundations of Modern Probability*. Springer-Verlag, New York.

Kimura, M. (1969). The number of heterozygous nucleotide sites mantained in a finite population due to steady flux of mutation. *Genetics*, **61**, 893–903.

Kimura, M. (1971). Theoretical foundations of population genetics at the molecular level. *Theoret. Popn. Biol.*, **2**, 174–208.

Kimura, M. and Crow, J. F. (1964). The number of alleles that can be mantained in a finite population. *Genetics*, **49**, 725–738.

Kingman, J. F. C. (1975). Random discrete distributions. *J. Royal Stat. Soc. Ser. B*, **37**, 1–22.

Kingman, J. F. C. (1982). The coalescent. *Stoch. Proc. and Their Applications*, **13**, 235–248.

Kottas, A. and Gelfand, A. E. (2001). Bayesian semiparametric median regression modelling. *J. Amer. Statist. Assoc.*, **11**, 289–305.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates I. Density estimates. *Ann. Statist.*, **12**, 351–357.

Lévy, P. (1954). Processus semi-markoviens. *Proc. Int. Congr. Math. (Amsterdam)*, **3**, 416–426.

MacEachern, S. N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Comm. Statist. Simulation Comput.*, **23**, 727–741.

MacEachern, S. N. (1998). Computational methods for mixtures of Dirichlet process models. In D. Dey, P. Müller, and D. Sinha, editors, *Practical nonparametric and semiparametric Bayesian statistics*, pages 23–43. Springer, New York.

MacEachern, S. N. and Müller, P. (1998). Estimating mixtures of Dirichlet process models. *J. Comput. Graph. Statist.*, **7**, 223–238.

MacEachern, S. N., Clyde, M., and Liu, J. S. (1999). Sequential importance sampling for nonparametrics Bayes models: the next generation. *Canad. J. Statist.*, **27**, 251–267.

Malécot, G. M. (1948). *Les Mathématiques de l'Hérédité*. Masson et Cie, Paris.

McCloskey, J. W. (1965). *A model for the distribution of individuals by species in an environment*. Ph.D. thesis, Michigan State University.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chemical Physics*, **21**, 1087–1091.

Moran, P. A. P. (1958a). A general theory of the distribution of gene frequencies i. overlapping generations. *Proc. Roy. Soc. London B*, **149**, 102–111.

Moran, P. A. P. (1958b). A general theory of the distribution of gene frequencies ii. non-overlapping generations. *Proc. Roy. Soc. London B*, **149**, 113–116.

Moran, P. A. P. (1958c). Random processes in genetics. *Proc. Camb. Phil. Soc.*, **54**, 60–71.

Neal, R. M. (2000). Markov chain samping methods for Dirichlet process mixture models. *J. Comput. Graph. Statist.*, **9**, 249–265.

Otha, T. and Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res. Camb.*, **22**, 201–204.

Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. theory and related fields*, **102**, 145–158.

Pitman, J. (1996a). Random discrete distributions invariant under size-biased permutation. *Adv. App. Probab.*, **28**, 525–539.

Pitman, J. (1996b). Some developments of the Blackwell-MacQueen urn scheme. In T. Ferguson, L. Shapley, and J. MacQueen, editors, *Statistics, probability and* *game theory*, volume 30 of *IMS Lecture Notes - Monograph Series*, pages 245–267. Hayward, CA: Institute of Mathematical Statistics.

Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, **25**, 855–900.

Pyke, R. (1961a). Markov renewal processes: definition and preliminary properties. *Ann. Math. Stat.*, **32**, 1231–1242.

Pyke, R. (1961b). Markov renewal processes with finitely many states. *Ann. Math. Stat.*, **32**, 1243–1259.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica*, **4**, 639–650.

Shiga, T. (1990). A stochastic equation based on a Poisson system for a class of measure-valued diffusion processes. *J.Math. Kyoto Univ.*, **30**, 245–279.

Shimakura, N. (1977). Equations différentielles provenant de la génétique des populations. *Tôhoku Math. J.*, **29**, 287–318.

Shimakura, N. (1981). Formulas for diffusion approximations of some gene frequency models. *J. Math. Kyoto Univ.*, **21**, 19–45.

Smith, A. F. M. and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *J. Royal Stat. Soc. Ser. B*, **1**, 3–23.

Smith, W. L. (1955). Regenerative stochastic processes. *Proc. Roy. Soc. Ser. A*, **232**, 6–31.

Stroock, D. W. and Varadhan, S. R. S. (1979). *Multidimensional diffusion processes*. Springer-Verlag, Berlin.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distribution by data augmentation (with discussion). *J. Amer. Statist. Assoc.*, **82**, 528–550.

Tavaré, S. (1984). Line-of-descent and genealogical processes, and their applications in population genetics models. *Theoret. Popn. Biol.*, **26**, 119–164.

Walker, S. G., Hatjispyros, S. J., and Nicoleris, T. (2007). The Fleming-Fiot process and Bayesian nonparametrics. *Ann. Appl. Probab.* to appear.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoret. Popn. Biol.*, **13**, 639–651.

Watterson, G. A. (1976). The stationary distribution of the infinitely-many neutral alleles diffusion model. *J. Appl. Probab.*, **13**, 639–651.

West, M. (1990). Bayesian kernel density estimation. Discussion paper 90-A02. Duke University, Institute of Statistics and Decision Sciences.

Wright, S. (1949). Adaptation and selection. In G. Jepson, E. Mayr, and G. Simpson, editors, *Genetics, paleontology and evolution*, pages 365–389. Princeton Univ. Press, Princeton, NJ.

Index