

Introductory Lectures on Stochastic Population Systems

Donald A. Dawson

*School of Mathematics and Statistics
Carleton University
Ottawa, Canada*

Contents

1	Introduction	7
2	Stochastic models in biology: a historical overview	9
2.1	Classical deterministic population dynamics	9
2.1.1	The Logistic Equation	9
2.1.2	The Lotka-Volterra Equations for Competing Species	10
2.1.3	The SIR Epidemic Model	10
2.1.4	Population models and dynamical systems	10
2.2	Small population effects	11
2.2.1	The Bienamyé-Galton-Watson Branching Process (BGW)	11
2.2.2	Reed-Frost epidemic model	12
2.2.3	Multitype populations and the Wright-Fisher Model	12
2.3	The Role of Stochastic Analysis	13
2.3.1	Diffusion approximations of branching and Wright-Fisher processes	14
2.4	Darwinian selection	15
2.4.1	Fisher's large population approximation	15
2.4.2	Selection and Genetic Drift	16
2.5	Spatially structured population systems	16
2.5.1	Super Random Walk	16
2.5.2	Stepping Stone Models	17
2.5.3	Spatial spread of advantageous genes and epidemics	18
2.6	Complex population dynamics	18
2.6.1	Basic questions on modelling complex and evolving populations	18
2.6.2	The role and analysis of stochasticity	19
3	Branching Processes I	21
3.1	Basic Concepts and Results on Branching Processes	21
3.1.1	Bienamyé-Galton-Watson processes	21
3.1.2	Bienamyé-Galton-Watson trees	24
3.1.3	Supercritical branching	28
3.1.4	The general branching model of Crump-Mode-Jagers	29
3.1.5	Multitype branching	34
3.2	Multilevel branching	35
4	Branching Processes II	37
4.1	Birth and Death Processes	37
4.1.1	Linear birth and death processes	37
4.1.2	Semigroups and generating functions	38

4.2	Critical branching	40
4.3	Feller's continuous state branching	41
4.4	Diffusion process limits	44
4.4.1	Convergence to Feller's continuous state branching process	44
4.5	The critical BGW tree	47
4.5.1	The rooted BGW tree as a metric space	47
4.5.2	The contour functions	48
4.5.3	Real trees	48
4.5.4	Excursions from zero and real trees	49
4.5.5	The Aldous Continuum Random Tree	50
4.5.6	Conditioned limit theorem for the critical BGW tree	50
4.5.7	Aldous Invariance Principle for BGW trees	51
4.6	Remark on general continuous state branching	53
5	Wright-Fisher Processes	55
5.1	Introductory remarks	55
5.2	Wright-Fisher Markov Chain Model	55
5.2.1	Types in population genetics	57
5.2.2	Finite population resampling in a diploid population	57
5.2.3	Diploid population with mutation and selection	58
5.3	Diffusion Approximation of Wright-Fisher	59
5.3.1	Neutral 2-allele Wright-Fisher model	59
5.3.2	K-allele Wright-Fisher Diffusion	64
5.4	Stationary measures	66
5.4.1	The Invariant Measure for the neutral K-alleles WF Diffusion	66
5.4.2	Convergence of stationary measures of $\{p^N\}_{N \in \mathbb{N}}$	68
6	Infinitely many types models	69
6.1	Introduction	69
6.1.1	Motivation	69
6.1.2	Plan	69
6.2	Measure-valued Wright-Fisher Markov chain	70
6.3	The neutral Fleming-Viot process with mutation generator A	71
6.4	The Infinitely Many Alleles Model	74
6.4.1	Projective Limit Construction of the Infinitely Many Alleles Model	74
6.5	The Jirina Branching Process	76
6.6	Invariant Measures of the IMA and Jirina Processes	77
6.6.1	The Moran (Gamma) Subordinator	77
6.6.2	Representation of the Infinitely Many Alleles Equilibrium	78
6.6.3	The Poisson-Dirichlet Distribution	80
6.6.4	The GEM Representation	81
6.6.5	Application of the Poisson-Dirichlet distribution: The Ewens Sampling Formula	83
7	Martingale Problems and Dual Representations	87
7.1	Introduction	87
7.2	The Jirina martingale problem	88
7.2.1	Stochastic Integrals wrt Martingale Measures	89
7.2.2	Uniqueness and stationary measures for the Jirina Martingale Problem	91
7.3	The infinitely many alleles martingale problem	94

7.4	Dual martingale problems	96
7.5	Dual representation of the neutral Fleming-Viot process	99
7.5.1	The General Neutral F.V. Process	99
7.5.2	Equivalent Formulation of the martingale problem	99
7.5.3	The dual representation of the Fleming-Viot process	100
7.5.4	The Kingman coalescent	101
7.6	Interactions via change of measure- the Girsanov formula	101
8	Genealogy and History	107
8.1	Introduction	107
8.2	Family Structure of the neutral Fleming-Viot Process	107
8.2.1	Fleming-Viot with Feller Mutation Semigroup	107
8.2.2	Two Finite Particle Systems and Moment Measures	109
8.2.3	Extension to Infinite Particle Systems: Some Preliminaries	113
8.2.4	The Countable Donnelly-Kurtz Look-Down Process	114
8.2.5	Genealogy and the Kingman Coalescent	119
8.2.6	Polyá Urn Scheme	121
8.3	Dynamics of population structure and history	122
8.3.1	The tree-valued Fleming-Viot process	122
8.3.2	The Historical Process	124
8.4	Some Applications of the Coalescent and Look-Down Process	125
8.4.1	Ergodicity for the Fleming-Viot Process	125
8.4.2	Atomic Structure of the Infinitely Many Alleles Model	125
8.4.3	The Infinitely Many Sites Model	127
8.4.4	Wandering Distributions	128
8.4.5	Support Properties.	129
8.5	Generalizations of the Kingman coalescent	130
8.5.1	Coalescent with time varying population size	130
8.5.2	Generalized coalescent	131
9	Spatially Structured Models	133
9.1	Introduction	133
9.1.1	Spatial dynamics	133
9.1.2	Random Walks	133
9.2	Branching random walk and branching Brownian motions	135
9.3	Interacting particle systems	136
9.3.1	Coalescing random walks	136
9.3.2	Voter Model	137
9.3.3	Biased voter model	139
9.3.4	Oriented percolation	140
9.3.5	Contact process	140
9.3.6	Interacting birth and death processes	141
9.4	Interacting diffusions	141
9.4.1	Interacting Feller diffusions.	141
9.4.2	The Wright-Fisher stepping stone model.	142
9.4.3	Mean-field limit of exchangeable interacting diffusions	142
9.5	Measure-valued branching processes	143
9.5.1	Super-Brownian motion	143
9.5.2	Super-Brownian Motion as the Limit of Branching Brownian Motion	144
9.5.3	The Poisson Cluster Representation	149

9.5.4	The Palm measure	150
9.5.5	Excursions	150
9.5.6	The Historical Branching Process	151
9.5.7	A skew product construction	153
9.5.8	The Donnelly-Kurtz countable particle representation	154
9.5.9	Brownian Snake Representation of SBM X_t	154
9.5.10	Catalytic SBM	154
9.6	Local properties of critical branching systems	155
9.6.1	Support Properties of SBM	155
9.6.2	Brief review of sample path properties	155
9.7	Spatial dynamical invariance principles	156
9.7.1	Methods of Proof	160
9.7.2	Proof of the invariance principle for interacting Wright-Fisher diffusions .	163
9.7.3	Applications of the SBM Invariance Principle	165
9.7.4	Spatial Lotka-Volterra models	166
9.7.5	Application of SBM to a spatial epidemic model	166
10	Spatial systems in large space and time scales	169
10.1	Migration processes on Abelian groups	169
10.1.1	Transience-Recurrence Properties	170
10.1.2	Random walks and Lévy processes in \mathbb{R}^d	171
10.2	The Persistence-Extinction Dichotomy for Critical Branching Systems	172
10.2.1	Clumping in Low Dimensions	173
10.2.2	Ergodic Behaviour	174
10.3	The Equilibrium Clan Decomposition	175
10.3.1	An extension of Historical Brownian motion	175
10.3.2	The historical process conditioned to live forever	175
10.3.3	Convergence to equilibrium for the historical process	177
10.3.4	Clan dynamics	178
10.4	Neutral Stepping Stone Models	179
10.4.1	The two type stepping stone model	179
10.4.2	The infinitely many types stepping stone model	179
10.4.3	The Dual Process Representation	180
10.4.4	Spatial homogeneity and the local-fixation coexistence dichotomy	181
10.4.5	Historical and genealogical structure of the neutral stepping stone model	186
11	Mutation-Selection Systems	191
11.1	The infinite population dynamics of mutation, selection and recombination.	191
11.1.1	Selection	191
11.1.2	Riemannian structure on Δ_{K-1}	192
11.1.3	Mutation-Selection	196
11.1.4	Multiple loci and recombination	196
11.2	Infinitely many types Fleming-Viot	197
11.2.1	Dual representation with mutation, selection and recombination	198
11.2.2	Girsanov formula for Fleming-Viot with Mutation and Selection	200
11.3	Long-time behaviour	201
11.3.1	Fixation in finite population systems without mutation	201
11.3.2	The Equilibrium Infinitely Many Alleles Model with Selection	202
11.3.3	Remarks on further developments	204
11.4	Remarks on the Bibliography	205

12 Appendix I: Martingales and large deviations	207
12.1 Martingales and uniform integrability	207
12.1.1 Burkholder-Davis-Gundy inequalities	208
12.2 Large Deviations	208
13 Appendix II: Measures and Topologies	211
13.1 Measures and weak convergence	211
13.1.1 Borel measures on Polish space	211
13.1.2 Weak convergence of measures	212
13.2 Projective Limits of Measures and Processes	212
13.3 Random measures	213
13.4 Topologies on path spaces	214
13.4.1 Sufficient conditions for tightness	214
13.4.2 The Joffe-Métivier criteria for tightness of D-semimartingales	215
13.4.3 Tightness of measure-valued processes	217
13.5 The Gromov-Hausdorff metric on the space of compact metric spaces	219
13.5.1 Metric measure spaces	219
13.6 Riemannian metrics and gradient	220
14 Appendix III: Markov Processes	223
14.1 Operator semigroups	223
14.2 Some basic result for one dimensional diffusions	223
14.2.1 Boundary behaviour classification	224
14.2.2 Excursions	224

Chapter 1

Introduction

These notes are based on lectures and courses given at the University of Erlangen, the Britton Lectures at McMaster University, the 2009 PIMS Summer School in Probability and Carleton University.

Historically, the modelling of biological populations has been an important stimulus for the development of stochastic processes. The revolutionary changes in the biological sciences over the past 50 years have created many new challenges and open problems. At the same time probabilists have developed new classes of stochastic processes such as interacting particle systems and measure-valued processes and made advances in stochastic analysis that make possible the modelling and analysis of populations having complex structures and dynamics. This course will focus on these developments. In particular stochastic processes that model populations distributed in space as well as their genealogies and interactions will be considered. This will include branching particle systems, interacting Wright-Fisher diffusions, Fleming-Viot processes and superprocesses. Basic methodologies including martingale problems, diffusion approximations, dual representations, coupling methods, random measures and particle representations will be introduced.

Chapter 2

Stochastic models in biology: a historical overview

2.1 Classical deterministic population dynamics

We begin our historical review with some basic models from demography, ecology and epidemiology.

The mathematical formulation of the growth of an age-structured population was developed by Euler (1760) ([221]). The (female) birth rate at time t $\{B(t)\}_{t \geq 0}$ satisfies the renewal equation

$$(2.1) \quad B(t) = \int_0^t B(t-s)(1 - L(s))m(s)ds.$$

This leads to exponential growth $B(t) \sim e^{\alpha t}$ where the *Malthusian parameter* α is given by the *characteristic equation of demography* (Euler-Lotka equation)

$$(2.2) \quad 1 = \int_0^\infty e^{-\alpha s}(1 - L(s))m(s)ds,$$

where $m(s)$ is the average birth rate for an individual of age s and $L(s)$ is the cumulative distribution function of the lifetime of an individual.

The implications of exponential growth of the human population was the subject of the famous writings of Thomas Malthus - *Essay on the Principle of Population* - (1798) which had a major impact and which was one of the influences on Darwin.

2.1.1 The Logistic Equation

Verhulst (1838) ([551]) introduced the *logistic equation* which describes the more realistic situation in which resources are limited and the death rate increases as the resources are exhausted.

$$(2.3) \quad \frac{dx}{dt} = \alpha x(1 - \frac{x}{N}), \quad x(0) \geq 0, \quad \alpha \geq 0$$

$$(2.4) \quad x(t) = \frac{Nx(0)e^{\alpha t}}{N + x(0)(e^{\alpha t} - 1)} \rightarrow N \text{ as } t \rightarrow \infty$$

Here $N < \infty$ is interpreted as the *carrying capacity* of the environment in which the population lived.

2.1.2 The Lotka-Volterra Equations for Competing Species

Equations to model the competition between species in ecology were proposed by (Lotka (1925) [412] and Volterra (1926) [553]):

$$(2.5) \quad \frac{dx_1}{dt} = r_1 x_1 \left(1 - \frac{x_1}{K_1} - a_{12} \frac{x_2}{K_1} \right)$$

$$(2.6) \quad \frac{dx_2}{dt} = r_2 x_2 \left(1 - \frac{x_2}{K_2} - a_{21} \frac{x_1}{K_2} \right)$$

Coexistence, that is, a *stable equilibrium* with both species present occurs if $\frac{1}{a_{21}} > \frac{K_1}{K_2} > a_{12}$.

Remark 2.1 Gause (1934) proposed the competitive-exclusion principle that states two species cannot stably coexist if they occupy the same niche, for example, if $a_{12} = a_{21}$.

2.1.3 The SIR Epidemic Model

A classical model for the progress of an epidemic due to Kermack and McKendrick (1927) [351] is given by the system of ode:

$$(2.7) \quad \frac{dS}{dt} = -\beta SI, \quad \frac{dI}{dt} = \beta SI - \gamma I, \quad \frac{dR}{dt} = \gamma I,$$

$$(2.8) \quad S(0) > 0, \quad I(0) > 0, \quad R(0) = 0.$$

Here S denotes the population of susceptible individuals, I the population of infectious individuals and R the population of removed individuals.

The *epidemiological threshold* quantity is defined by

$$(2.9) \quad R_0 = \frac{\beta S(0)}{\gamma} (\text{reproductive ratio})$$

If $R_0 < 1$, then the infected population never increases whereas if $R_0 > 1$ the epidemic “will spread”.

2.1.4 Population models and dynamical systems

As suggested by these elementary examples, the modeling of interacting multitype populations leads to a rich area of dynamical systems and there exists an immense literature in this field.

For example the extension of the Lotka-Volterra equations to N interacting species is given by the system:

$$(2.10) \quad \frac{x_i(t)}{dt} = r_i x_i \left(1 - \sum_{j=1}^N \alpha_{ij} x_j \right), \quad i = 1, \dots, N.$$

These multispecies dynamical systems can have very complex behavior including limit cycles or chaotic behaviour. In fact S. Smale (1976) [525] proved that for $N \geq 5$ these systems can exhibit any asymptotic behavior.

2.2 Small population effects

Deterministic models provide good approximations to the growth of large (noncritical) populations but for small populations and “nearly critical populations” it is essential to take account of their inherent discrete nature and randomness.

2.2.1 The Bienamyé-Galton-Watson Branching Process (BGW)

The importance of the fact that individuals produce a random number of offspring and the possibility exists that the population can become extinct led Bienaym   (1845) [36], and Galton-Watson (1874) [564] to introduce this probabilistic model.

The population size at generation n is denoted X_n . Starting with $X_1 = 1$, at each generation each individual gives rise to a random number of children as follows:

$$X_{n+1} = \sum_{k=1}^{X_n} \xi_k, \text{ where the } \{\xi_k\}_{k=1,\dots,X_n} \text{ are independent.}$$

Generating functions provide a basic tool for developing these processes. The generating function for the offspring distribution is given by:

$$f(s) = E[s^{\xi}] = \sum_{k=0}^{\infty} p_k s^k, \quad 0 \leq s \leq 1$$

$$f'(1) = m = \text{mean offspring size.}$$

The key relation is

$$E[s^{X_n}] = f_n(s), \text{ where } f_{n+1} = f[f_n(s)].$$

The **extinction probability**, is defined by $q := P(X_n = 0 \text{ for some } n < \infty)$.

Theorem 2.2 (Steffensen (1930, 1932)) If $m \leq 1$ (critical, subcritical branching), then $q = 1$. If $m > 1$ (supercritical branching), the q is the unique nonnegative solution in $[0, 1]$

$$(2.11) \quad s = f(s).$$

If $m < \infty$, then

$$\frac{X_n}{m^n} \text{ is a martingale.}$$

Propagation of initial randomness

Theorem 2.3 (Hawkins and Ulam (1944), Yaglom (1947), Harris (1948), [278]) If $m > 1$, $\sigma^2 = f''(1) + f'(1) - (f'(1))^2 < \infty$, then

$$\frac{X_n}{m^n} \rightarrow W, \text{ in } L^2 \text{ and a.s. as } n \rightarrow \infty$$

and

$$EW = 1, \quad Var(W) = \frac{\sigma^2}{(m^2 - m)}, \quad P(W = 0) = q.$$

2.2.2 Reed-Frost epidemic model

A probabilistic analogue of the SIR epidemic model known as the Reed-Frost model is given as follows. We consider an initial population of susceptible individuals $S_0 = N$ and one infected individual $I_0 = 1$.

$$S_{t+1} \sim \text{Bin}(S_t, (1-p)^{I_t}), \quad t \in \mathbb{N},$$

that is, in each time unit a susceptible individual has probability p of meeting each infected individual and one such contact results in infection. Individuals are infected during one time period so that $I_{t+1} = S_t - S_{t+1}$. If $1-p = e^{-\lambda/N}$, then Von Bahr and Martin-Löf (1980) [554] showed that as $N \rightarrow \infty$ the critical threshold is $\lambda = 1$.

2.2.3 Multitype populations and the Wright-Fisher Model

The celebrated work of Mendel (1865) [436] on the inheritance of traits and its rediscovery around 1900 led to the development of the field of genetics. The modern theory of mathematical genetics was initiated in the work of Wright (1931), (1932) [581],[582] and Fisher (1930) [236]. They introduced a probabilistic model of finite population sampling that serves as a starting point for modern population genetics. This model deals with a population of individuals of different types. As a mathematical idealization they assume that the total population is constant in time and they focus on the changes in the relative proportions of the different types of individual. The key ingredients are:

- Fixed finite population size N
- Typespace (*alleles*)

$$E_K = \{1, \dots, K\}$$

- $X_n(i)$ is the number of individuals of type i , at generation n .

Let $\mathcal{N}(E_K)$ denote the counting measures on E . Then dynamics are defined by a Markov chain $X_n = (X_n(1), \dots, X_n(K))$ with state space

$$\{(x_1, \dots, x_K) \in \mathcal{N} : \sum_{i=1}^K x_i = N\}.$$

The intuitive idea leading to the transition mechanism for the *neutral model* is that first each individual in the n th generation produces a large number of potential offspring. Then in a second stage the population is pruned back (culling) so that the total population remains N (this can be thought of as an analogue of carrying capacity). Based on the *neutral assumption*, that is each of the individuals in produced in the first stage has equal probability of being selected, the $(n+1)$ st generation consists of N individuals of types $\{1, \dots, K\}$ obtained by

- *multinomial sampling* from the empirical distribution

$$\begin{aligned} P(X_{n+1} = (y_1, \dots, y_K) | X_n = (x_1, \dots, x_K)) \\ = \frac{N!}{y_1! y_2! \dots y_K!} \left(\frac{x_1}{N}\right)^{y_1} \dots \left(\frac{x_K}{N}\right)^{y_K} \end{aligned}$$

An important feature of this process is the loss of information (diversity) leading to **fixation**, that is, the long time survival of exactly one type. To see this note that

$$p_n(i) \text{ is a martingale where } p_n(i) = \frac{X_n(i)}{N}$$

and

$$p_n(i) \rightarrow 0 \text{ or } 1 \text{ as } n \rightarrow \infty \text{ for each } i \text{ w.p.1.}$$

The dual perspective

If we choose k individuals at random from generation $n + 1$ and look backwards in time to identify the parents in the n th generation, by an elementary conditional probability calculation, we see that *each individual in generation $n + 1$ picks its parents “at random”*. This naturally leads to the notion of *identity by descent* introduced by Malécot (1941) [420], that is, two individuals are identical by descent if they have a common ancestor (and no mutations have occurred).

2.3 The Role of Stochastic Analysis

Basic developments in stochastic analysis:

The development of stochastic population modelling was made possible by the remarkable developments in stochastic analysis.

- Markov chains and processes (1906) [427], Kolmogorov (1931), [383]
- Brownian motion Wiener (1923) [579], Lévy (1948) [401]
- Ito stochastic calculus (1942), (1946), (1951) [313]
- Markov processes and their semigroup characterization
Feller (1951) [232], Itô-McKean (1965).

Given a Markov process $\{X(t)\}_{t \geq 0}$ with state space E (for example, compact metric space) and $f \in C(E)$ (bounded continuous functions on E) and $x \in E$, let

$$T_t f(x) = E_x(f(X(t))$$

$\{T_t\}$ is said to be strongly continuous if $\|T_t f - f\| \rightarrow 0$ as $t \downarrow 0$ for $f \in C(E)$.

Then for some class $f \in D(G) \subset C(E)$ the generator G acting on f is defined by

$$Gf(x) = \lim_{t \downarrow 0} \frac{T_t f(x) - f(x)}{t} \text{ exists and } \in C(E).$$

Conditions under $(D(G), G)$ defines a strongly continuous semigroup were obtained in the celebrated Hille-Yosida Theorem (1948) [282], [586].

2.3.1 Diffusion approximations of branching and Wright-Fisher processes

In two seminal papers Feller (1939) [231], Feller (1951) [232] developed diffusion process approximations to the branching process and the Wright-Fisher model. These are now referred to as the Feller continuous state branching process (CSBP) and Wright-Fisher diffusion process, respectively. These serve as the “bridge” between the discrete world of individuals and generations and the world of differential equations and dynamical systems. It simultaneously maintains the power of analysis of the latter world and the random finite population effects of the former world.

We now state the two basic results.

Theorem 2.4 *Nearly critical BGW processes to Feller CSBP Branching*

Consider a sequence of BGW processes with mean offspring sizes $m_N = 1 + \frac{m}{N}$ and constant finite variance. Assume that $N^{-1}X_0^N \rightarrow X_0$ as $N \rightarrow \infty$. Then

$$\left\{ \frac{1}{N}X_{\lfloor Nt \rfloor}^N, t \geq 0 \right\} \Longrightarrow \{X_t : t \geq 0\}$$

where the convergence is in the sense of weak convergence of càdlàg processes. The limiting process is a continuous process with state space $[0, \infty)$ and the generator of the associated semigroup is given by

$$Gf(x) = mx \frac{\partial f}{\partial x} + \frac{\gamma}{2}x \frac{\partial^2 f}{\partial x^2} \quad \text{for some } \gamma > 0.$$

See Section 4.3 for details.

Theorem 2.5 *Wright-Fisher Diffusion*

Consider a sequence of K -type Wright-Fisher Markov chains X^N with total population N and assume that $N^{-1}X_0^N \rightarrow \mathbf{p}_0 \in \mathcal{P}(E_K)$ (probability measures on E_K) as $N \rightarrow \infty$. Then

$$\{\mathbf{p}_N(t) : t \geq 0\} \equiv \left\{ \frac{1}{N}X_{\lfloor Nt \rfloor}^N, t \geq 0 \right\} \Longrightarrow \{\mathbf{p}(t) : t \geq 0\}$$

where $\{\mathbf{p}(t) : t \geq 0\}$ is a Markov diffusion process with values in the simplex

$$\Delta_{K-1} = \{(p_1, \dots, p_K) : p_i \geq 0, \sum_{i=1}^K p_i = 1\}$$

and generator acting on functions $f(\mathbf{p}) = f(p_1, \dots, p_K)$:

$$G^{(K)}f(\mathbf{p}) = \frac{1}{2} \sum_{i,j=1}^K p_i (\delta_{ij} - p_j) \frac{\partial^2 f(\mathbf{p})}{\partial p_i \partial p_j}.$$

Remark 2.6 In the case $K = 2$ it suffices to keep track of $p_1 \in [0, 1]$ and then the generator is

$$Gf(p_1) = \frac{1}{2}p_1(1-p_1) \frac{\partial^2 f(p_1)}{\partial p_1^2}.$$

In terms of Itô’s SDE, this satisfies

$$dp_1(t) = \sqrt{p_1(t)(1-p_1(t))} dw(t), \quad p_1(0) \in [0, 1],$$

where $\{w(t)\}_{t \geq 0}$ is a standard Brownian motion.

See Section 5.2 for details.

2.4 Darwinian selection

Darwin's theory of evolution (*On the Origin of Species*, (1859 [106]) was based on the concept on natural selection based on the differential reproductive success of the different types of individual. In a seminal paper in 1924 J.B.S. Haldane [274], [285] initiated the modern synthesis of Darwinism evolution and Mendelian genetics (Mendel (1865)) and formulated the notion of *fitness*.

2.4.1 Fisher's large population approximation

A deterministic mathematical model incorporating the notion of fitness was developed by Fisher [236] as follows. Consider an infinite diploid population (organisms have a type $(i, j) \in E \times E$) which reproduce sexually with random mating, that is, the offspring type is obtained by choosing two individuals at random (parents) and choosing one of the homologous pairs from each parent. The type space is $E \times E$ (genotype determined by the gametes i and j).

Let $x_i(t)$ be the amount of gamete i in the population at time t and p_i denote the frequency $p_i = \frac{x_i}{\sum x_i}$.

Let $V(i, j) = V(j, i)$ = “diploid fitness” of the genotype (i, j) . The instantaneous fitness, $V(i, p)$ of the i th gamete is defined by

$$V(i, p) = \sum_j p_j V(i, j)$$

and the mean fitness is defined by

$$\bar{V} = \sum_i V(i, p)p_i = \sum_{ij} p_i p_j V(i, j).$$

(The *haploid case* is similar to the additive case $V(i, j) = V(i) + V(j)$.)

In Fisher's formulation the population sizes x_i satisfy the differential equations

$$\frac{dx_i}{dt} = x_i V(i, p), \quad i = 1, \dots, K$$

and therefore the proportions $\{p_i\}$ satisfy the equations:

$$(2.12) \quad \frac{dp_i}{dt} = p_i(V(i, p) - \bar{V}), \quad i = 1, \dots, K$$

Theorem 2.7 (*Fisher's Fundamental Theorem*) ([236])

- (a) Mean fitness $\bar{V}(t)$ increases on the trajectories of $\mathbf{p}(t)$.
- (b) The rate of change of the mean $\bar{V}(t)$ along orbits is proportional to the variance.

See Section 12.1 for details.

2.4.2 Selection and Genetic Drift

In contrast to Fisher, Wright [581], [582] considered genetic drift (due to finite population effects) to play an important role in evolution and developed his *shifting balance theory of evolution*. The relative importance to evolution of different mechanisms remains a subject of investigation and debate (see for example Barton-Turelli (1997) [24], Ohta-Gillespie (1996) [464]).

To introduce the interplay between selection and *genetic drift* we consider the diffusion approximation to the finite population model with two types (1, 2) and *haploid selection*. Let $p_1(t)$ denote the proportion of type 1. Let type 1 have fitness $s > 0$ and type 2 have fitness 0. Then the diffusion approximation limit $p_1(t)$ satisfies the SDE

$$(2.13) \quad dp_1(t) = sp_1(t)(1 - p_1(t))dt + \sqrt{\gamma p_1(t)(1 - p_1(t))}dw(t)$$

Here γ is proportional to the inverse of the *effective population size*.

The relation between the probability laws of $p_1(t)$ with and without selection (i.e. setting $s = 0$) follows from the Cameron-Martin-Girsanov representation ([493], Chapt. VIII) as follows.

Theorem 2.8 *Let $P_{[0,t]}^s$ be the probability law on $C_{[0,1]}([0,t])$ of the solution of (2.13). The Radon-Nikodym derivative on $\mathcal{P}(C_{[0,1]}([0,t]))$ is given by:*

$$\frac{dP_{[0,t]}^s}{dP_{[0,t]}^0} = \exp\left(\frac{s}{\gamma}(p_1(t) - p_1(0)) - \frac{s^2}{\gamma} \int_0^t p_1(s)(1 - p_1(s))ds\right).$$

The deviation of very large but finite populations from the deterministic (infinite population) limit can be analysed by considering the asymptotics as $\gamma \rightarrow 0$ and using Freidlin-Wentzell (see [244]) large deviation methods.

2.5 Spatially structured population systems

The above models assume that any new individual can be chosen from any member of the population and the members of the population interact in an exchangeable manner at any time. However real populations are distributed in space and reproduce and compete locally. Spatial models play an essential role in the study of population systems. We begin with a basic formulation with discrete “geographic space”.

To begin we consider a population with subpopulations located at sites on the a finite or countable set S , for example the lattice $S = \mathbb{Z}^d$, in which individual migrate between sites with migration rates given by a symmetric random walk kernel $\{p_{\xi-\xi'}\}_{\xi,\xi' \in S}$.

2.5.1 Super Random Walk

The branching random walk (see Athreya-Ney [8]) extends the basic Bienamyé-Galton-Watson process to the situation in which individuals are located in a countable space S .

The diffusion limit of branching random walks on S leads to a system of stochastic differential equations, now called *super random walks* (SRW), as follows:

$$\begin{aligned} dx_\xi(t) &= \sum_{\xi' \in S} p_{\xi-\xi'}(x_{\xi'}(t) - x_\xi(t))dt + \sqrt{\gamma x_\xi(t)}dw_\xi(t) \\ x_\xi(0) &\geq 0 \end{aligned}$$

where $\{w_\xi(\cdot)\}_{\xi \in S}$ is a system of independent Brownian motions. We note that if S is infinite, then the space of configurations $[0, \infty)^S$ is infinite dimensional. We can also identify this with the set of locally finite measures on S , $\mathcal{M}(S)$. The state at time t is then given by a random measure on S (see Moyal (1962) [444] for an early formulation of spatial population processes). A key tool in the study of an important class of random measures is the Laplace functional.

Theorem 2.9 *The transition Laplace Functional of the SRW is given by*

$$E_{\mathbf{x}(0)} \exp \left(- \sum_{\xi \in S} \varphi(\xi) x_\xi(t) \right) = \exp \left(- \sum_{\xi \in S} v_t(\xi) x_\xi(0) \right)$$

for $\varphi \in C_+(S)$, where v_t is the unique solution of

$$v_t(\xi) = S_t \varphi(\xi) - \int_0^t S_{t-s} (v_s^2(\xi)) ds$$

where $\{S_t\}$ is semigroup on $C_b(S)$ with generator

$$(2.14) \quad Gf(\xi) = \sum_{\xi'} p_{\xi-\xi'} (f(\xi') - f(\xi)).$$

The measure-valued analogue of this system on \mathbb{R}^d , now called super-Brownian motion, was introduced by S. Watanabe (1968) [563] in the context of branching processes and by Dawson (1975)[107] in the context of stochastic evolution equations.

2.5.2 Stepping Stone Models

The introduction of spatial models in population genetics goes back to Wright's island model [581] and the work of Malécot (1941) [420], (1948) [421], (1949) [422], Kimura (1953) [360], and Sawyer (1976) [501].

We will consider below the Wright-Fisher two-type diffusion stepping stone model with selection

$$\begin{aligned} dx_\xi(t) &= c \sum_{\xi' \in S} p_{\xi-\xi'} (x_{\xi'}(t) - x_\xi(t)) dt \\ &\quad + sx_\xi(t)(1-x_\xi(t))dt + \sqrt{2x_\xi(t)(1-x_\xi(t))} dw_\xi(t) \\ x_\xi(0) &= \theta \in [0, 1] \quad \forall \xi \end{aligned}$$

This arises from a collection of Wright-Fisher populations at demes $\xi \in S$ in which there is probability $\frac{cp_{\xi-\xi'}}{N}$ that an individual in generation $n+1$ is the offspring of an individual at deme $\xi' \neq \xi$ in generation n .

The stepping stone model is closely related to the voter model which was introduced by Clifford, Sudbury (1973) [74], Holley and Liggett (1975) [283]. The voter model is a $\{0, 1\}^S$ -valued Markov jump process and is one of the principal examples of the class of interacting particle systems introduced by Spitzer (1970) [528] and Dobrushin [150] (1971) and extensively developed over the past 35 years.

2.5.3 Spatial spread of advantageous genes and epidemics

An important feature of spatial models is spatial spread, for example, of a mutant type or epidemic. A classic example is the wave of advance of an advantageous gene modelled by the celebrated Fisher-KPP equation (1937) (Fisher ([237], Kolmogorov, Petrovsky and Piscounov [379]):

$u(t, x)$ = proportion of advantageous type 1 at x at time t

$$\begin{aligned} u_t &= \frac{1}{2}u_{xx} + u(1-u) \\ u(0, x) &= 1_{(-\infty, 0]}(x) \end{aligned}$$

The fundamental results of Kolmogorov, Petrovsky and Piscounov establish the existence of travelling wave solutions at speed $\sqrt{2}$. Relations between this travelling wave and the maximal displacement of branching Brownian motion were used by Bramson (1983) [52] to obtain fine properties of this phenomenon.

2.6 Complex population dynamics

A generalization of the N species Lotka-Volterra equations is give by the nonlinear dynamical system:

$$(2.15) \quad \frac{dx_i(t)}{dt} = F_i(x_1(t)), \quad i = 1, \dots, N.$$

As mentioned above these multispecies dynamical systems can have very complex behavior including limit cycles or chaotic behaviour. Questions of the complexity, robustness, diversity and spatial structure have been the subject of much research and debate much of it stimulated by the work of Hutchinson (1957) [305] and MacArthur (1955) [417], [418], [419]. Some information can be gained by analyzing the behaviour near stationary points, that is (x_1, \dots, x_N) where $F_i(x_1, \dots, x_N) = 0 \forall i$ by considering the linearized system around these points and in particular the spectra of the resulting matrices. One approach to these questions was developed in the 1973 paper of May [429] in which he assumed that the matrices are random and used results on random matrix theory to look at the relation between species number N and complexity (measured by the proportion of non-zero matrix elements). Limitations of this analysis have been pointed out [75] and the assumption of randomness does not reflect the dynamical mechanisms involved. A reasonably robust ecological systems is constantly being tested by the emergence of new mutant types. For a robust system most such mutations and deleterious and are eliminated. However rare mutant types can cause the collapse of the system or move the system to a new attractor and a higher or lower order of organization. *This means that the ecosystem dynamics is itself subject to evolutionary forces.* A theoretical framework to classify the types of dynamical system to develop and their stability remains elusive. However stochastic effects are important here in that the generation and survival at low population sizes of mutant types clearly plays an important role.

2.6.1 Basic questions on modelling complex and evolving populations

We next take a quick look at some of the basic questions on the modelling of complex evolving populations.

- How do we model a complex multilevel population?
- Individuals - type, internal state and dynamics, geographical location, fitness, interaction matrix.
- reproduction and modification mechanisms: birth, death and mutation rates,
- Family structure and genealogy
- migration dynamics
- Spatial distribution of types: role of spatial and hierarchical network structures
- Networks of interactions between individuals and groups of individuals, for example, competition and or cooperation between types, for example, in ecology and economics. Stability and collapse.
- How do we relate the different levels of description: microscopic, mesoscopic and macroscopic?
- How do we describe the development of population composition and structure in different space and time scales going from the microscopic to evolutionary scales.
- emergence of new levels of organization

2.6.2 The role and analysis of stochasticity

We have seen some of the classical stochastic population models in our historical survey. Today, in view of the revolutionary developments in biology over the past 50 years there is an endless richness of biological phenomena and models. There is a huge literature on both deterministic and stochastic models. Once can ask: what is the role of stochasticity in the development of complex populations? Where do deterministic models fail and the role of randomness is essential? There is no simple answer but we make the following four observations:

- Evolution is an interplay of nonlinear dynamics, finite population fluctuations and rare random events at which fundamental transitions occur in the population composition and dynamics.
- Extinction and the loss of information due to finite population resampling.
- The creation of diversity and information is a result of selective forces acting on randomly produced mutant types. The latter can be viewed as a random search through a potentially infinite search space.
- Demographic and environmental stochasticity is ubiquitous and plays a role analogous to molecular motion in statistical physics. The levels of these sources of randomness influences the nature of quasiequilibria and non-equilibrium phase transitions.

Fortunately over the past fifty year, stochastic analysis has undergone major developments in many directions, partly in response to these challenges. This has produced a range of tools that have proved effective in addressing some significance issues in the sciences including the biological sciences and hold potential to address even more of these questions in the future. The objective of these notes is to introduce some of the basic ideas and tools and to provide some pointers to the growing literature in this field.

Chapter 3

Branching Processes I: Supercritical growth and population structure

The fundamental characteristic of biological populations is that individuals undergo birth and death and that individuals carry information passed on from their parents at birth. Furthermore there is a randomness in this process in that the number of births that an individual gives rise to is in general not deterministic but random. Branching processes model this process under simplifying assumptions but nevertheless provide the starting point for the modelling and analysis of such populations. In this chapter we present some of the central ideas and key results in the theory of branching processes.

3.1 Basic Concepts and Results on Branching Processes

3.1.1 Bienamyé-Galton-Watson processes

The Bienamyé-Galton-Watson branching process (BGW process) is a Markov chain on $\mathbb{N}_0 := \{0, 1, 2, \dots\}$. The discrete time parameter is interpreted as the generation number and X_n denotes the number of individuals alive in the n 'th generation. Generation $(n + 1)$ consists of the offspring of the n th generation as follows:

- each individual i in the n th generation produces a random number ξ_i with distribution

$$p_k = P[\xi_i = k], \quad k \in \mathbb{N}_0$$

- $\xi_1, \xi_2, \dots, \xi_{X_n}$ are independent.

Let $X_0 = 1$. Then for $n \geq 0$

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_i, \quad \{\xi_i\} \text{ independent}$$

We assume that the mean number of offspring

$$m = \sum_{i=1}^{\infty} ip_i < \infty.$$

The BGW process is said to be *subcritical* if $m < 1$, *critical* if $m = 1$ and *supercritical* if $m > 1$.

A basic tool in the study of branching processes is the *generating function*

$$(3.1) \quad f(s) = E[s^\xi] = \sum_{k=0}^{\infty} p_k s^k, \quad 0 \leq s \leq 1.$$

Then

$$(3.2) \quad f'(1) = m, \quad f''(1) = E[\xi(\xi - 1)] \geq 0.$$

Let

$$f_n(s) = E[s^{X_n}], \quad n \in \mathbb{N}.$$

Then conditioned on X_n , and using the independence of the $\{\xi_i\}$,

$$f_{n+1}(s) = E[s^{\sum_{i=1}^{X_n} \xi_i}] = E[f(s)^{X_n}] = f_n(f(s)) = f(f_n(s)).$$

Note that $f(0) = P[\xi = 0] = p_0$ and

$$P[X_{n+1} = 0] = f(f_n(0)) = f(P[X_n = 0])$$

Then if $m > 1$, $p_0 > 0$, $P[X_n = 0] = f_n(0) \uparrow q$ where q is the smallest nonnegative root of

$$f(s) = s,$$

and if $m \leq 1$, $P[X_n = 0] \uparrow 1$. Note that 1 and q are the only roots of $f(s) = s$.

Since $E[X_{n+1}|X_n] = mX_n$,

$$(3.3) \quad W_n := \frac{X_n}{m^n} \text{ is a martingale and } \lim_{n \rightarrow \infty} W_n = W \text{ exists a.s.}$$

Proposition 3.1 *We have $P[W = 0] = q$ or 1, that is, conditioned on nonextinction either $W = 0$ a.s. or $W > 0$ a.s.*

Proof. It suffices to show that $P[W = 0]$ is a root of $f(s) = s$. The i th individual of the first generation has a descendant family with a martingale limit which we denote by $W^{(i)}$. Then $\{W^{(i)}\}_{i=1,\dots,X_1}$ are independent and have the same distribution as W . Therefore

$$(3.4) \quad W = \frac{1}{m} \sum_{i=1}^{X_1} W^{(i)}$$

and therefore $W = 0$ if and only if for all $i \leq X_1$, $W^{(i)} = 0$. Conditioning on X_1 implies that

$$(3.5) \quad P[W = 0] = E(P(W^{(i)} = 0)^{X_1}) = f(P[W = 0]).$$

Therefore $P[W = 0]$ is a root of $f(s) = s$. ■

Remark 3.2 *In the case $\text{Var}(X_1) = \sigma^2 < \infty$ we can show by induction that*

$$(3.6) \quad \begin{aligned} \text{Var}(X_n) &= \frac{\sigma^2 m^n (m^n - 1)}{m^2 - m}, & m \neq 1, \\ &= n\sigma^2, & m = 1 \end{aligned}$$

Then if $m > 1$ the martingale $\frac{X_n}{m^n}$ is uniformly integrable and $E(W) = 1$. Moreover $\frac{X_n}{m^n} \rightarrow W$ in L^2 and

$$(3.7) \quad \text{Var}(W) = \frac{\sigma^2}{m^2 - m} > 0 \quad (\text{see Harris [278] Theorem 8.1}).$$

If $m > 1$, $\sigma^2 = \infty$, a basic question concerns the nature of the random variable W and the question whether or not $\frac{X_n}{m^n} \rightarrow W$ in L^1 . The question was settled by a celebrated result of Kesten and Stigum which we present in Theorem 3.6 below. We first introduce some further basic notions.

Bienamyé-Galton-Watson process with immigration (BGWI)

The Bienamyé-Galton-Watson process with offspring distribution $\{p_k\}$ and immigration process $\{Y_n\}_{n \in \mathbb{N}_0}$ satisfies

$$(3.8) \quad X_{n+1} = \sum_{i=1}^{X_n} \xi_i + Y_{n+1},$$

where the ξ_i are iid with distribution $\{p_k\}$.

Let \mathcal{F}^Y be the σ -field generated by $\{Y_k : k \geq 1\}$ and $X_{n,k}$ be the number of descendants at generation n of the individuals who immigrated in generation k . Then the total number of individuals in generation n is $X_n = \sum_{k=1}^n X_{n,k}$.

For $k < n$ the random variable $W_{n,k} = X_{n,k}/m^{n-k}$ has the same law as \tilde{X}_{n-k}/m^{n-k} where \tilde{X}_n is the BGW process with Y_k initial particles. Therefore

$$(3.9) \quad E\left[\frac{X_{n,k}}{m^{n-k}}\right] = Y_k.$$

Now consider the subcritical case $m < 1$. If $\{Y_i\}$ are i.i.d. with $E[Y_i] < \infty$, then the Markov chain X_n has a stationary measure with mean $\frac{E[Y]}{1-m}$.

Next consider the supercritical case $m > 1$. Then

$$(3.10) \quad E\left[\frac{X_n}{m^n} \mid \mathcal{F}^Y\right] = E\left[\frac{1}{m^n} \sum_{k=1}^n X_{n,k} \mid \mathcal{F}^Y\right] = \sum_{k=1}^n \frac{1}{m^k} E\left[\frac{X_{n,k}}{m^{n-k}} \mid \mathcal{F}^Y\right] = \sum_{k=1}^n \frac{Y_k}{m^k}.$$

If $\sup_k E[Y_k] < \infty$, then

$$(3.11) \quad \lim_{n \rightarrow \infty} \frac{E[X_n]}{m^n} = \sum_{k=1}^{\infty} \frac{E[Y_k]}{m^k} < \infty.$$

A dichotomy in the more subtle case $E[Y_k] = \infty$ is provided by the following theorem of Seneta.

Theorem 3.3 (Seneta (1970) [506]) *Let X_n denote the BGW process with mean offspring $m > 1$, $X_0 = 0$ and with i.i.d. immigration process Y_n .*

(a) *If $E[\log^+ Y_1] < \infty$, then $\lim \frac{X_n}{m^n}$ exists and is finite a.s.*

(b) *If $E[\log^+ Y_1] = \infty$, then $\limsup \frac{X_n}{c^n} = \infty$ for every constant $c > 0$.*

Proof. The theorem is a consequence of the following elementary result.

Lemma 3.4 Let Y, Y_1, Y_2, \dots be nonnegative iid rv. Then a.s.

$$(3.12) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} Y_n = \begin{cases} 0, & \text{if } E[Y] < \infty \\ \infty, & \text{if } E[Y] = \infty \end{cases}$$

Proof. Recall that $E[Y] = \int_0^\infty P(Y > x)dx$. This gives $\sum_n P(\frac{Y}{n} > c) < \infty$ for any $c > 0$ if $E[Y] < \infty$ and the result follows by Borel-Cantelli. If $E[Y] = \infty$, then $\sum P(\frac{Y}{n} > c) = \infty$ for any $c > 0$ and the result follows by the second Borel-Cantelli Lemma since the Y_n are independent. ■

Proof of (a). By (3.10)

$$(3.13) \quad E\left[\frac{X_n}{m^n} | \mathcal{F}^Y\right] = \sum_{k=1}^n \frac{Y_k}{m^k}.$$

Since here we assume $E[\log^+ Y_1] < \infty$, Lemma 3.4 gives $\limsup_{k \rightarrow \infty} \frac{Y_k}{c^k} < \infty$ for any $c > 0$. Therefore the series given by the last expression in (3.13) converges a.s. and therefore $\lim_{n \rightarrow \infty} E\left[\frac{X_n}{m^n} | \mathcal{F}^Y\right]$ exists and is finite a.s. This implies (a)

Proof of (b). If $E[\log^+ Y_1] = \infty$, then by Lemma 3.4 $\limsup_{n \rightarrow \infty} \frac{\log^+ Y_n}{n} = \infty$ a.s. Therefore for any $c > 0$

$$(3.14) \quad \limsup_{n \rightarrow \infty} \frac{Y_n}{c^n} = \infty$$

a.s. Since $X_n \geq Y_n$, (b) follows.
■

3.1.2 Bienamyé-Galton-Watson trees

In addition to the keeping track of the total population of generation $n+1$ in a BGW process it is useful to incorporate genealogical information, for example, which individuals in generation $n+1$ have the same parent in generation n . This leads to a natural family tree structure which was introduced in the papers of Joffe and Waugh (1982), (1985), [332], [333] in their determination of the distribution of kin numbers and developed in the papers of Chauvin (1986) [70] and Neveu (1986) [458].

A convenient representation of the BGW random tree is as follows. Let $u = (i_1, \dots, i_n)$ denote an individual in generation n who is the i_n th child of the i_{n-1} -th child of \dots of the i_1 -th child of the ancestor, denoted by \emptyset . The *space of individuals* (vertices) is given by

$$(3.15) \quad \mathcal{I} = \{\emptyset\} \cup \cup_{n=1}^{\infty} \mathbb{N}^n.$$

Given $u = (u_1, \dots, u_m)$, $v = (v_1, \dots, v_n) \in \mathcal{I}$, we denote the composition by $uv := (u_1, \dots, u_m, v_1, \dots, v_n)$

A *plane rooted tree* \mathcal{T} with root \emptyset is a subset of \mathcal{I} such that

1. $\emptyset \in \mathcal{T}$,
2. If $v \in \mathcal{T}$ and $v = uj$ for some $u \in \mathcal{I}$ and $j \in \mathbb{N}$, then $u \in \mathcal{T}$.
3. For every $u \in \mathcal{T}$, there exists a number $k_u(\mathcal{T}) \geq 0$, such that $uj \in \mathcal{T}$ if and only if $1 \leq j \leq k_u(\mathcal{T})$.

A plane tree can be given the structure of a graph in which a parent is connected by an edge to each of its offspring.

Let \mathbf{T} be the set of all plane trees. If $t \in \mathbf{T}$ let $[t]_n$ be the set of rooted trees whose first n levels agree with those of t . Let \mathbf{V} denote the set of connected sequences in \mathcal{I} , $\emptyset, v_1, v_2, \dots$, which do not backtrack. Given $t \in \mathbf{T}$, let $V(t)$ denote the set of paths in t . If v_n is a vertex at the n th level, let $[t; v]_n$ denote the set of trees with distinguished paths such that the tree is in $[t]_n$, $v \in V(t)$ and the path goes through v_n .

Given a finite plane tree \mathcal{T} the *height* $h(\mathcal{T})$ is the maximal generation of a vertex in \mathcal{T} and $\#(\mathcal{T})$ denotes the number of vertices in \mathcal{T} . Let \mathbf{T}_n be the set of trees of height n .

A *random tree* is given by a probability measure on \mathbf{T} . Given an offspring distribution $\mathcal{L}(\xi) = \{p_k\}_{k \in \mathbb{N}}$, the corresponding BGW tree is constructed as follows:

Let the initial individual be labelled \emptyset . Give it a random number of children denoted $1, 2, \dots, \xi_\emptyset$.

Then each of these has a random number of children, for example i has children denoted $(i, 1), \dots, (i, \xi_i)$ etc. Each of these has children, for example (i, j) has $\xi_{i,j}$ children labelled $(i, j, 1), \dots, (i, j, \xi_{i,j})$, etc. Then considering the first n generations in this way we obtain a probability measure P_n^{BGW} on \mathbf{T}_n .

The probability measures, P_n^{BGW} form a consistent family and induce a probability measure P^{BGW} on \mathbf{T} , the law of the BGW random tree.

Let

$$(3.16) \quad Z_n = \text{number of vertices in the tree at level } n.$$

Then by the construction it follows that Z_n is a version of the BGW process and we can think of the BGW tree as an enriched version of the BGW process.

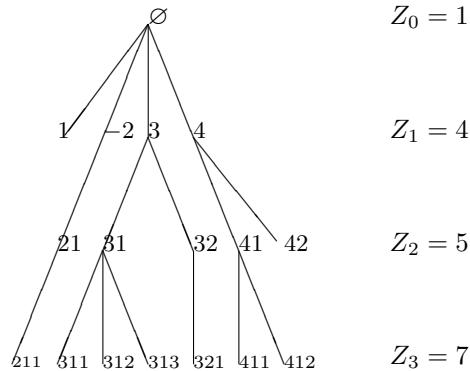


Figure 3.1: BGW Tree

The size-biased BGW tree

The fundamental notion of size-biasing has many applications. It will be used below in the proof of Lyons, Pemantle and Peres (1995) [416] of some basic results on Bienamyé-Galton-Watson processes (see Theorem 3.6 below).

To exploit this notion for branching processes we consider the *size-biased offspring distribution*

$$(3.17) \quad \hat{p}_k = \frac{kp_k}{m}, \quad k = 1, 2, \dots$$

We denote by $\hat{\xi}$ a random variable having the size biased offspring distribution. The size-biased BGW tree \hat{T} is constructed as follows:

- the initial individual is labelled \emptyset ; \emptyset has a random number $\hat{\xi}_{\emptyset}$ of children (with the size-biased offspring distribution) \hat{p} ,
- one of the children of \emptyset is selected at random and denoted v_1 and given an independent size-biased number $\hat{\xi}_{v_1}$ of children,
- the other children of \emptyset are independently assigned ordinary BGW descendant trees with offspring number ξ ,
- again one of the children of v_1 is selected at random and denoted v_2 and given an independent size-biased number $\hat{\xi}_{v_2}$ of children,
- this process is continued and produces the size-biased BGW tree \hat{T} which is *immortal* and infinite distinguished path v which we call the *backbone*.

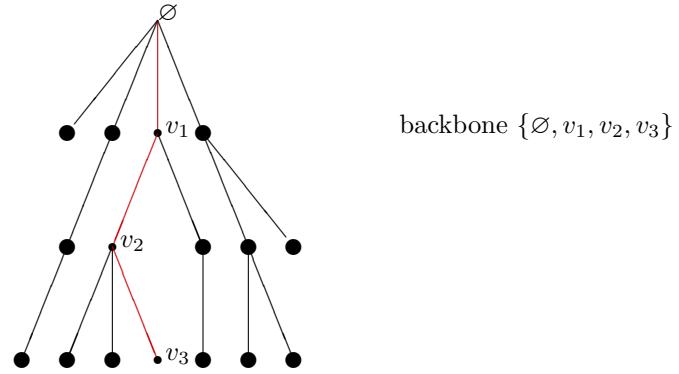


Figure 3.2: Size-biased BGW Tree

Define the measure $\bar{P}_*^{BGW} \in \mathcal{P}(\mathbf{T} \times \mathbf{V})$ to be the joint distribution of the random tree \hat{T} and backbone $\{v_0, v_1, v_2, \dots\}$. Let \bar{P}^{BGW} denote the marginal distribution of \hat{T} . We can view the vertices off the backbone (v_0, v_1, \dots) of the size-biased tree as a branching process with immigration in which the immigrants are the siblings of the individuals on the backbone. The distribution of the number of immigrants at generation n , Y_n , is given by the law $\hat{\xi} - 1$.

Given a tree t let $[t]_n$ denote the tree restricted to generations $1, \dots, n$. Let $Z_n(t)$ denote the number of vertices in the tree at the n th level (generation) and $\mathcal{F}_n = \sigma([t]_n)$. Let

$$(3.18) \quad W_n(t) := \frac{Z_n(t)}{m^n}$$

denote the martingale associated to a tree t with $Z_n(t)$ vertices at generation n .

Lemma 3.5 (a) *The Radon-Nikodym derivative of the marginal distribution $\bar{P}^{BGW}|_{\mathcal{F}_n}$ of \hat{T} with respect to $P^{PGW}|_{\mathcal{F}_n}$ is given by*

$$(3.19) \quad \frac{d\bar{P}_n^{BGW}}{dP_n^{BGW}}(t) = W_n(t).$$

(b) *Under the measure \bar{P}_*^{BGW} , the vertex v_n at the n th level of the tree \hat{T} in the random path (v_0, v_1, \dots) is uniformly distributed on the vertices at the n th level of \hat{T} .*

Proof.

We will verify that

$$(3.20) \quad \bar{P}_*^{BGW}[t, v]_n = \frac{1}{m^n} P^{BGW}[t]_n$$

and therefore

$$(3.21) \quad \bar{P}^{BGW}[t]_n = W_n(t) P^{BGW}[t]_n.$$

First observe that the

$$(3.22) \quad \begin{aligned} \bar{P}_*^{BGW}(Z_1 = k, v_1 = i) &= \frac{kp_k}{m} \cdot \frac{1}{k} \\ &= \frac{p_k}{m} = \frac{1}{m} P(\xi = k), \text{ for } i = 1, \dots, k. \end{aligned}$$

since v_1 is randomly chosen from the offspring $(1, \dots, \hat{\xi}_\emptyset)$.

Now consider $[\hat{T}, v]_{n+1}$. We can construct this by first selecting $\hat{\xi}_0$ and v_1 and then following the next n generations of the resulting descendant tree and backbone as well as the BGW descendant trees of the remaining $\hat{\xi}_0 - 1$ vertices in the first generation. If $\hat{\xi}_\emptyset(t) = k$ we denote the resulting descendant trees by $t^{(1)}, t^{(2)}, \dots, t^{(k)}$.

Let $v_{n+1}(t)$ be a vertex (determined by a position in the lexicographic order) at level $n+1$. It determines $v_1(t)$ and the descendant tree $t^{(v_1)}$ that it belongs to. If $\hat{\xi}_\emptyset(t) = k$, $v_1(t) = i$, then we obtain

$$(3.23) \quad \bar{P}_*^{BGW}[t; v]_{n+1} = \frac{p_k}{m} \cdot \bar{P}_*^{BGW}[t^{(i)}; v_{n+1}]_n \cdot \prod_{j=1, j \neq i}^k P^{BGW}[t^{(j)}]_n.$$

Then by induction for each n

$$(3.24) \quad \bar{P}_*^{BGW}[t; v]_n = \frac{1}{m^n} P^{BGW}[t]_n$$

for each of the $Z_n(t)$ positions v in the lexicographic order at level n and $[t]_n$. Consequently we have obtained the *martingale change of measure*

$$(3.25) \quad \bar{P}_*^{BGW}[t]_n = \frac{Z_n(t)}{m^n} P^{BGW}[t]_n$$

and

$$(3.26) \quad \bar{P}_*^{BGW}[v = i|t] = \frac{1}{Z_n(t)} \text{ for } i = 1, \dots, Z_n(t).$$

■

For an infinite tree t we define

$$(3.27) \quad W(t) := \limsup_{n \rightarrow \infty} W_n(t).$$

Note that in the critical and subcritical cases the measures P^{BGW} and \bar{P}^{BGW} are singular since the P^{BGW} - probability of nonextinction is zero. The question as to whether or not they are singular in the supercritical case will be the focus of the next subsection.

3.1.3 Supercritical branching

As mentioned above if $0 < m < \infty$, then under P^{BGW}

$$(3.28) \quad W_n = \frac{Z_n}{m^n}$$

is a martingale and converges to a random variable W a.s. as $n \rightarrow \infty$. The characterization of the limit W in the supercritical case, $m > 1$, under minimal conditions was obtained in the following theorem of Kesten and Stigum (1966) [352]. The proof given below follows the “conceptual proof” of Lyons, Pemantle and Peres (1995) [416].

Theorem 3.6 (Kesten-Stigum (1966) [352]) *Consider the BGW process with offspring ξ and mean offspring size m . If $1 < m < \infty$, the following are equivalent*

- (i) $P^{BGW}[W = 0] = q$
- (ii) $E^{BGW}[W] = 1$
- (iii) $E[\xi \log^+ \xi] < \infty$

Proof. By Lemma 3.5

$$(3.29) \quad \frac{d\bar{P}_n^{BGW}}{dP_n^{BGW}}(t) = W_n(t)$$

where the left side denotes the Radon-Nikodym derivative wrt $\mathcal{F}_n = \sigma([t]_n)$.

Note that $P^{BGW}(W = 0) \geq q$ where $q = P^{BGW}(E_0)$ where $E_0 := \{Z_n = 0 \text{ for some } n < \infty\}$ (extinction probability). Moreover, since $\mathcal{F}_n \uparrow \mathcal{F} = \sigma(t)$, we have the Radon-Nikodym dichotomy (see Theorem 12.4)

$$(3.30) \quad W = 0, \quad P^{BGW} - \text{a.s.} \quad \Leftrightarrow P^{BGW} \perp \bar{P}^{BGW} \quad \Leftrightarrow W = \infty \quad \bar{P}^{BGW} - \text{a.s.}$$

and

$$(3.31) \quad \int W dP^{BGW} = 1 \quad \Leftrightarrow \bar{P}^{BGW} \ll P^{BGW} \quad \Leftrightarrow W < \infty \quad \bar{P}^{BGW} - \text{a.s.}$$

Now recall (3.13) that the size-biased tree can be represented as a branching process with immigration in which the distribution of the number of immigrants at generation n , Y_n , is given by the law $\hat{\xi} - 1$, that is

$$(3.32) \quad E[Z_n | \mathcal{Y}] = \sum_{k=1}^n \frac{Y_k}{m^k}.$$

If $E[\log^+ \hat{\xi}] = \frac{1}{m} E[\xi \log^+ \xi] = \frac{1}{m} \sum_{k=1}^{\infty} kp_k \log k = \infty$, then

$$(3.33) \quad W = \lim_{n \rightarrow \infty} \frac{Z_n}{m^n} = \infty, \quad \bar{P}^{BGW} \text{ a.s.}$$

by Theorem 3.3 (b). Therefore $P^{BGW}(W = 0) = 1$ by (3.30).

If $E[\xi \log^+ \xi] < \infty$, then $E[\log^+ \hat{\xi}] = \sum_{k=1}^{\infty} kp_k \log k < \infty$. and by Theorem 3.3(a)

$$(3.34) \quad \lim_{n \rightarrow \infty} E\left(\frac{Z_n}{m^n} \mid \mathcal{F}^{\mathcal{Y}}\right) = \sum_{k=1}^{\infty} \frac{Y_k}{m^k} < \infty, \quad \bar{P}^{BGW} \text{ a.s.}$$

and therefore

$$(3.35) \quad W = \lim_{n \rightarrow \infty} \frac{Z_n}{m^n} < \infty, \quad \bar{P}^{BGW} - \text{a.s.}$$

Then $E^{BGW}[W] = \int W dP^{BGW} = 1$ by (3.31).

Finally, since by Proposition 3.1 $P^{BGW}(W = 0) = q$ or 0, we obtain (i). \blacksquare

Remark 3.7 The supercritical branching model is the basic model for a growing population with unlimited resources. A more realistic model is a spatial model in which resources are locally limited but the population can grow by spreading spatially. A simple deterministic model of this type is the Fisher-KPP equation. We will consider the analogous spatial stochastic models in a later chapter.

3.1.4 The general branching model of Crump-Mode-Jagers

We now consider a far-reaching generalization of the Bienamé-Galton-Watson process known as a Crump-Mode Jagers (CMJ) process ([105], [320]). This is a process with time parameter set $[0, \infty)$ consisting of finitely many individuals at each time.

With each individual x we denote its birth time τ_x , lifetime λ_x and reproduction process ξ_x . The latter is a point process which gives the sequence of birth times of individuals. $\xi_x(t)$ is the number of offspring produced (during its lifetime) by an individual x born at time 0 during $[0, t]$. The intensity of ξ_x , called the *reproduction function* is defined by

$$(3.36) \quad \mu(t) = E[\xi(t)].$$

The lifetime distribution is defined by

$$(3.37) \quad L(u) = P[\lambda \leq u].$$

We begin with one individual \emptyset which we assume is born at time $\tau_{\emptyset} = 0$. The reproduction processes ξ_x of different individuals are iid copies of ξ .

The basic probability space is

$$(3.38) \quad (\Omega_{\mathcal{I}}, \mathcal{B}_{\mathcal{I}}, P_{\mathcal{I}}) = \prod_{x \in \mathcal{I}} (\Omega_x, \mathcal{B}_x, P_x)$$

where \mathcal{I} is given as in (3.15) and (ξ_x, λ_x) are random variables defined on $(\Omega_x, \mathcal{B}_x, P_x)$ with distribution as above.

We then determine the birth times $\{\tau_x, x \in \mathcal{I}\}$ as follows:

$$(3.39) \quad \begin{aligned} \tau_{\emptyset} &= 0, \\ \tau_{(x', i)} &= \tau_{x'} + \inf \{u : \xi_{x'}(u) \geq i\}. \end{aligned}$$

Note that for individuals never born $\tau_x = \infty$.

Let

$$(3.40) \quad Z_t = \sum_{x \in \mathcal{I}} 1_{\tau_x \leq t < \lambda_x}, \quad T_t = \sum_{x \in \mathcal{I}} 1_{\tau_x \leq t}$$

that is, the *number of individuals alive at time t* and *total number of births before time t*, respectively.

For $\lambda > 0$ we define

$$(3.41) \quad \xi^\lambda(t) := \int_0^t e^{-\lambda u} \xi(du).$$

The *Malthusian parameter* α is defined by the equation

$$(3.42) \quad E[\xi^\alpha(\infty)] = 1$$

that is,

$$(3.43) \quad \int_0^\infty e^{-\alpha t} \mu(dt) = 1.$$

The *stable average age of child-bearing* is defined as

$$(3.44) \quad \beta = \int_0^\infty t \tilde{\mu}(dt) \text{ where } \tilde{\mu}(dt) = e^{-\alpha t} \mu(dt).$$

Example 3.8 Consider a population in which individuals have an internal state space, say \mathbb{N} . Assume that the individual starts in state 0 at its time of birth and its internal state changes according to a Markov transition mechanism. Finally assume that when it is in state i it produces offspring at rate λ_i .

Definition 3.9 Characteristics of an individual: A characteristic of an individual is given by a process $\phi : \mathbb{R} \times \Omega \rightarrow \mathbb{R}_+$ which is given by a $\mathcal{B}(\mathbb{R}) \times \sigma(\xi)$ -measurable non-negative function satisfying $\phi(t) = 0$ for $t < 0$, let

$$(3.45) \quad Z_t^\phi = \sum_{x \in \mathcal{I}} \phi_x(t - \tau_x)$$

denote the process counted with characteristic ϕ .

Example 3.10 If $\phi^a(t) = 1_{[0, \inf(a, \lambda))}(t)$, then $Z_t^{\phi^a}$ counts the number of individuals alive at time t whose ages are less than a .

The following fundamental generalization of the Kesten-Stigum theorem was developed in papers of Doney (1972), (1976) [166], [167], and Nerman (1981) [453].

Theorem 3.11 Consider a CMJ process with malthusian parameter α and assume that $\beta < \infty$.

(a) [166] Then as $t \rightarrow \infty$, $e^{-\alpha t} Z_t$ converges in distribution to mW_∞ where

$$(3.46) \quad m = \frac{\int_0^\infty e^{-\alpha s}(1 - L(s))ds}{\beta}$$

and W_∞ is a random variable (see Proposition 3.13) and

(b) The following are equivalent:

$$(3.47) \quad E[\xi^\alpha(\infty) \log^+ \xi^\alpha(\infty)] < \infty$$

$$(3.48) \quad E[W] > 0$$

$$(3.49) \quad E[e^{-\alpha t} Z_t] \rightarrow E[W] \quad \text{as } t \rightarrow \infty$$

$$(3.50) \quad W > 0 \text{ a.s. on } \{T_t \rightarrow \infty\}.$$

(c) [453] Under the condition that there exists a non-increasing integrable function g such that

$$(3.51) \quad E[\sup_t \frac{(\xi^\alpha(\infty) - \xi^\alpha(t))}{g(t)}] < \infty,$$

then $e^{-\alpha t} Z_t$ converges a.s. as $t \rightarrow \infty$.

Remark 3.12 A sufficient condition is the existence of non-increasing integrable function g such that

$$(3.52) \quad \int_0^\infty \frac{1}{g(t)} e^{-\alpha t} \mu(dt) < \infty.$$

(See Nerman [453] (5.4)).

Comments on Proofs

(b) The equivalence statements can be proved in this general case following the same lines as that of Lyons, Pemantle and Peres - see Olofsson (1996) [468].

(a) - convergence in distribution was proved by Doney (1972) [166]. However the almost sure convergence required some basic new ideas since we can no longer directly use the martingale convergence theorem since Z_t is not a martingale in the general case. The a.s. convergence was proved by Nerman [453]. We will not give Nerman's long detailed technical proof of this result but will now introduce the key tool used in its proof and which is of independent interest, namely, an underlying intrinsic martingale W_t discovered by Nerman [453] and then give an intuitive idea of the remainder of the proof.

Denote the mother of x by $m(x)$ and let

$$(3.53) \quad \mathcal{I}_t = \{x \in \mathcal{I} : \tau_{m(x)} \leq t < \tau_x < \infty\},$$

the set of individuals whose mothers are born before time t but who themselves are born after t

Consider the individuals ordered by their times of birth

$$(3.54) \quad 0 = \tau_{x_1} \leq \tau_{x_2} \leq \dots$$

Define $\mathcal{A}_n = \sigma$ -algebra generated by $\{(\tau_{x_i}, \xi_{x_i}, \lambda_{x_i}) : i = 1, \dots, n\}$ Recall (3.40) and let $\mathcal{F}_t = \mathcal{A}_{T_t}$.

Define

$$(3.55) \quad W_t := \sum_{x \in \mathcal{I}_t} e^{-\alpha \tau_x}.$$

Proposition 3.13 (*Nerman (1981) [453]*) (a) The process $\{W_t, \mathcal{F}_t\}$ is a non-negative martingale with $E[W_t] = 1$.

(b) There exists a random variable $W_\infty < \infty$ such that $W_t \rightarrow W_\infty$ a.s. as $t \rightarrow \infty$.

Proof. Define

$$(3.56) \quad R_0 = 1,$$

$$R_n = 1 + \sum_{i=1}^n e^{-\alpha \tau_{x_i}} (\xi_{x_i}^\alpha(\infty) - 1), \quad n = 1, 2, \dots$$

Equivalently, letting $\tau_{(x_i, k)}$ denote the time of birth of the k th offspring of x_i ,

$$(3.57) \quad R_n = 1 + \sum_{i=1}^n \sum_{k=1}^{\xi_{x_i}(\infty)} e^{-\alpha \tau_{(x_i, k)}} - \sum_{i=1}^n e^{-\alpha \tau_{x_i}}$$

so that R_n is a weighted (weights $e^{-\alpha \tau_x}$) sum of children of the first n individuals.

We next show that (R_n, \mathcal{A}_n) is a non-negative martingale. R_n and $\tau_{x_{n+1}}$ are \mathcal{A}_n -measurable and $\xi_{x_{n+1}}^\alpha$ is independent of \mathcal{A}_n and

$$(3.58) \quad E[\xi_{x_{n+1}}^\alpha(\infty)] = \mu_\alpha(\infty) = 1.$$

Therefore

$$(3.59) \quad E[R_{n+1} - R_n] = e^{-\alpha \tau_{x_{n+1}}} E[\xi_{x_{n+1}}^\alpha - 1] = 0.$$

Next we observe that since $\mathcal{I}(t)$ consists of exactly the children of the first T_t individuals to be born after t , it follows that $W_t = R_{T_t}$.

Note that for fixed t , $\{T_t > k\} = \{\tau_{x_n} \leq t\} \in \mathcal{A}_n$ and therefore T_t is an increasing family of integer-valued stopping times with respect to $\{\mathcal{A}_n\}$. Therefore $\{W_t\}$ is a supermartingale with respect to the filtration $\{\mathcal{A}_{T_t}\}$.

Since $E[T_t] < \infty$ and

$$(3.60) \quad E[|R_{n+1} - R_n| \mid \mathcal{A}_n] = e^{-\alpha \tau_{x_{n+1}}} E[|\xi_{x_{n+1}}^\alpha(\infty) - 1|] \leq 2.$$

a standard argument (e.g. Breiman [58] Prop. 5.33) implies that $E[W_t] = E[R_{T_t}] = 1$ and $\{W_t\}$ is actually a martingale.

(b) This follows from (a) and the martingale convergence theorem.

■

Remark 3.14 We now sketch an intuitive explanation for the proof of the a.s. convergence of $e^{-\alpha t} Z_t$ using Proposition 3.13. This is based on the relation between W_t and Z_t which is somewhat indirect. To give some idea of this, let

$$(3.61) \quad W_{t,c} = \sum_{x \in \mathcal{I}_{t,c}} e^{-\alpha \tau_x},$$

where

$$(3.62) \quad \mathcal{I}_{t,c} = \{x = (x', i) : \tau_{x'} \leq t, t + c < \tau_x < \infty\}.$$

Note that if we consider the characteristic χ^c defined by

$$(3.63) \quad \chi^c(s) = (\xi^\alpha(\infty) - \xi^\alpha(s + c))e^{\alpha s} \text{ for } s \geq 0,$$

then

$$(3.64) \quad W_{t,c} = e^{-\alpha t} Z_t^{\chi^c}$$

where

$$(3.65) \quad Z_t^{\chi^c} = \sum_{x \in \mathcal{I}} \chi_x^c(t - \tau_x), \quad \chi_x^c(s) = (\xi_x^\alpha(\infty) - \xi_x^\alpha(s + c))e^{\alpha s}.$$

Note that $\lim_{c \rightarrow 0} W_{t,c} = W_t$ and $\lim_{c \rightarrow 0} Z_t^{\chi^c} = Z_t^\chi$ where

$$(3.66) \quad \chi(s) = \int_s^\infty e^{-\alpha(u-s)} \xi(du).$$

Then $Z_t^\chi \rightarrow m_\chi W_\infty$, a.s. where

$$(3.67) \quad m_\chi = \frac{\int_0^\infty e^{-\alpha t} (1 - L(t)) dt}{\beta}.$$

In the special case where ξ is stationary then the distribution of $\chi(s)$ does not depend on s . Then Z_t^χ is a sum of Z_t i.i.d. random variables and therefore as $t \rightarrow \infty$, Z_t^χ should approach a constant times Z_t by the law of large numbers.

Stable age distribution

The notion of the stable age distribution of a population is a basic concept in demography going back to Euler. The stable age distribution in the deterministic setting of the Euler-Lotka equation (2.2) is

$$(3.68) \quad U(\infty, ds) = \frac{(1 - L(s))e^{-\alpha s} ds}{\int_0^\infty (1 - L(s))e^{-\alpha s} ds}.$$

It was introduced into the study of branching processes by Athreya and Kaplan (1976) [9]. Let Z_t^a denote the number of individuals of age $\leq a$. The normalized age distribution at time t is defined by

$$(3.69) \quad U(t, [0, a)) := \frac{Z_t^a}{Z_t}, \quad a \geq 0.$$

Theorem 3.15 (Nerman [453] Theorem 6.3 - Convergence to stable age distribution) Assume that ξ satisfies the conditions of Theorem 3.11. Then on the event $T_t \rightarrow \infty$,

$$(3.70) \quad U(t, [0, a)) \rightarrow \frac{\int_0^a (1 - L(u))e^{-\alpha u} du}{\int_0^\infty (1 - L(u))e^{-\alpha u} du} \text{ a.s. as } t \rightarrow \infty.$$

3.1.5 Multitype branching

A central idea in evolutionary biology is the differential growth rates of different types of individuals. Multitype branching processes provide a starting point for our discussion of this basic topic.

Consider a multitype BGW process with K types. Let $\xi^{(i,j)}$ be a random variable representing the number of particles of type j produced by one type i particle in one generation.

Let $Z^{(j)}$ be the number of particles of type j in generation n and $\mathbf{Z}_n := (Z_n^{(1)}, \dots, Z_n^K)$.

For $\mathbf{k} = (k_1, \dots, k_K)$, let $p_{\mathbf{k}}^{(i)} = P[\xi^{(i,j)} = k_j, j = 1, \dots, K]$. Assume that

$$(3.71) \quad \mathbf{M} = (m_{(i,j)})_{i,j=1,\dots,K}, \\ m_{(i,j)} = E[\xi^{(i,j)}] < \infty \quad \forall i, j.$$

Then

$$(3.72) \quad E(\mathbf{Z}_{m+n} | \mathbf{Z}_m) = \mathbf{Z}_m \mathbf{M}^n, \quad m, n \in \mathbb{N}.$$

The behaviour of $E[\mathbf{Z}_n]$ as $n \rightarrow \infty$ is then obtained from the classical Perron-Frobenius Theorem:

Theorem 3.16 (Perron-Frobenius) *Let \mathbf{M} be a nonnegative $K \times K$ matrix. Assume that \mathbf{M}^n is strictly positive for some $n \in \mathbb{N}$. Then \mathbf{M} has a largest positive eigenvalue ρ which is a simple eigenvalue with positive right and left normalized eigenvectors $\mathbf{u} = (u_i)$ ($\sum u_i = 1$) and $\mathbf{v} = (v_i)$ which are the only nonnegative eigenvectors. Moreover*

$$(3.73) \quad \mathbf{M}^n = \rho^n \mathbf{M}_1 + \mathbf{M}_2^n$$

where $\mathbf{M}_1 = (u_i v_j)_{i,j \in \{1, \dots, K\}}$ normalized by $\sum_i j u_i v_j = 1$. Moreover $\mathbf{M}_1 \mathbf{M}_2 = \mathbf{M}_2 \mathbf{M}_1 = 0$, $\mathbf{M}_1^n = \mathbf{M}_1$.

Finally,

$$(3.74) \quad |\mathbf{M}_2^n| = O(\alpha^n)$$

for some $0 < \alpha < \rho$.

The analogue of the Kesten-Stigum theorem stated above is given as follows.

Theorem 3.17 (Kesten-Stigum (1966) [352]), (Kurtz, Lyons, Pemantle and Peres (1997) [388])

(a) *There is a scalar random variable W such that*

$$(3.75) \quad \lim_{n \rightarrow \infty} \frac{\mathbf{Z}_n}{\rho^n} = W \mathbf{u} \text{ a.s.}$$

and $P[W > 0] > 0$ iff

$$(3.76) \quad E\left[\sum_{i,j=1}^J \xi^{(i,j)} \log^+ \xi^{(i,j)}\right] < \infty.$$

(b) *Almost surely, conditioned on nonextinction,*

$$(3.77) \quad \lim_{n \rightarrow \infty} \frac{\mathbf{Z}_n}{|\mathbf{Z}_n|} = \mathbf{u}.$$

3.2 Multilevel branching

Consider a *host-parasite population* in which the individuals in the host population reproduce by BGW branching and the population of parasite on a given host also develop by an independent BGW branching. This is an example of a *multilevel branching system*.

A *multilevel population system* is a hierarchically structured collection of objects at different levels as follows:

E_0 denotes the set of possible types of level 1 object,
for $n \geq 1$ each level $(n + 1)$ object is given by a collection of level n object including their their multiplicities.

Multilevel branching dynamics

Consider a continuous time branching process such that

- for $n \geq 1$, when a level n object branches, all its offspring are copies of it
- if $n \geq 2$, then the offspring contains a copy of the set of level- $n - 1$ objects contained in the parent level n object.
- let γ_n the level n branching rate and by $f_n(s)$ the level n offspring generating function.

Then the questions of extinction, classification into critical, subcritical and supercritical case and growth asymptotics in the supercritical case are more complex than the single level branching case. See for example, Dawson and Wu (1996) [128].

Chapter 4

Branching Processes II: Convergence of critical branching to Feller's CSB

4.1 Birth and Death Processes

4.1.1 Linear birth and death processes

Branching processes can be studied in discrete or continuous time. We now consider a classical continuous time version. This is a continuous time Markov chain, $\{X_t\}_{t \geq 0}$ with state space \mathbb{N}_0 and with linear birth and death rates, b and d and let $V = b + d \geq 0$. This corresponds to a branching system in which (independently) each particle can die or be replaced by two offspring in the interval $[t, t + \Delta t]$ with probability $V\Delta t + o(\Delta t)$. This means that the time until the first branch (birth-death event) is an exponential random variable with mean $\frac{1}{V}$. V is called the branching rate. When the particle “branches” it dies with probability $\frac{d}{b+d}$ and is replaced by two descendants with probability $\frac{b}{b+d}$. Note that this process can be built directly on a probability space containing a sequence of iid exponential (1) rv’s and a sequence of iid Bernoulli ($p = \frac{b}{b+d}$) rv’s (or a sequence of iid Uniform $[0, 1]$ rv’s) and this description can be used to generate a simulation of the model. The special case in which $d = 0$ is called the *Yule* process.

The birth and death process can also be realized on a probability space (Ω, \mathcal{F}, P) on which independent Poisson random measures N_1, N_2 on \mathbb{R}_+^2 are defined. Then the birth and death process is defined via a stochastic differential equation driven by the Poisson noises, namely,

$$(4.1) \quad X_t = x_0 + \int_0^t \int_0^{bX(s_-)} N_1(du, ds) - \int_0^t \int_0^{dX(s_-)} N_2(du, ds).$$

This equation has a pathwise unique càdlàg solution which is a continuous time Markov chain with the required transition rates. See Li-Ma [407].

Let P_{x_0} denote the resulting probability law on $D_{\mathbb{N}_0}([0, \infty))$, the space of càdlàg functions from $[0, \infty)$ to \mathbb{N}_0 .

4.1.2 Semigroups and generating functions

Given the Markov process X_t we can associate a Markov semigroup $\{T_t : t \geq 0\}$ of operators on the Banach space $C_0(\mathbb{N}_0)$ (the space of bounded functions on \mathbb{N}_0 , with limits at infinity) as follows:

$$T_t f(x_0) := E_{x_0}(f(X_t)) = \int f(x) P_{x_0}(X_t \in dx).$$

This semigroup determines the finite dimensional distributions of the Markov chain. This semigroup satisfies the conditions of the Hille-Yosida theorem with generator given by

$$\begin{aligned} Gf(n) &= \frac{dT_tf(n)}{dt}|_{t=0} \\ &= bn(f(n+1) - f(n)) + dn(f(n-1) - f(n)) \end{aligned}$$

Now consider the *Laplace function* of X_t starting with one particle at time 0:

$$L(t, \theta) := E_{x_0}(e^{-\theta X_t}), \text{ with } x_0 = 1, \theta \geq 0$$

Noting the outcome at the first branching time and using the independence of the particle and its offspring when a birth occurs, we obtain the nonlinear renewal-type equation

$$L(t, \theta) = e^{-Vt}e^{-\theta} + \frac{d}{b+d}(1 - e^{-Vt}) + V \frac{b}{b+d} \int_0^t e^{-Vu} L^2(t-u, \theta) du$$

Alternately, note that we can represent the jump in X_t at a branching time by the addition of an independent random variable ζ with Laplace transform $E(e^{-\theta\zeta}) = \frac{b}{b+d}e^{-\theta} + \frac{d}{b+d}e^\theta$. Since the branching occurs at linear rate VX_t at time t , we get

$$\begin{aligned} \frac{\partial L(t, \theta)}{\partial t} &= \lim_{\Delta \rightarrow 0} \frac{L(t + \Delta, \theta) - L(t, \theta)}{\Delta} \\ &= \lim_{\Delta \rightarrow 0} \left(\frac{E(E(e^{-\theta X_{t+\Delta}} | X_{t-})) - E(e^{-\theta X_{t-}})}{\Delta} \right) \\ &= \lim_{\Delta \rightarrow 0} \frac{V\Delta[E(X_{t-}E(e^{-\theta(X_{t-}+\zeta)} | X_{t-})) - E(X_{t-}e^{-\theta X_{t-}})] + o(\Delta)}{\Delta} \\ &= -\{V \frac{b}{b+d}(e^{-\theta} - 1) + V \frac{d}{b+d}(e^\theta - 1)\} \frac{\partial L(t, \theta)}{\partial \theta}. \end{aligned}$$

Here we have used $E(Xe^{-\theta X}) = -\frac{\partial L(\theta)}{\partial \theta}$. So we then have the first order PDE

$$(4.2) \quad \frac{\partial L(t, \theta)}{\partial t} + V[\frac{b}{b+d}(e^{-\theta} - 1) + \frac{d}{b+d}(e^\theta - 1)] \frac{\partial L(t, \theta)}{\partial \theta} = 0, \quad L(0, \theta) = e^{-\theta}.$$

We can solve this by finding the *characteristic curves* $(t(s), \theta(s))$ in the (t, θ) plane along which $L(t(s), \theta(s))$ is constant (refer to Garabedian (1964) [248], John (1982) [335] or Delgado (1997) [149]). We write this as

$$(4.3) \quad \frac{\partial}{\partial s} L(t(s), \theta(s)) = L_1 \frac{\partial t(s)}{\partial s} + L_2 \frac{\partial \theta(s)}{\partial s} = 0$$

where L_1, L_2 denote the first partial derivatives with respect to t, θ respectively. Comparing (4.3) with (4.2) leads to the *characteristic equations*

$$(4.4) \quad \frac{\partial L(s)}{\partial s} = 0, \quad \frac{\partial t(s)}{\partial s} = 1, \quad \frac{\partial \theta(s)}{\partial s} = h(\theta) = b(e^{-\theta} - 1) + d(e^\theta - 1)$$

For $b \neq d$ we obtain the characteristic curve

$$(4.5) \quad \frac{(e^{-\theta} - 1)e^{(b-d)t}}{be^{-\theta} - d} = \text{constant}.$$

and general solution

$$(4.6) \quad L(t, \theta) = \Psi\left(\frac{(e^{-\theta} - 1)e^{(b-d)t}}{be^{-\theta} - d}\right)$$

where Ψ is a differentiable function. From the initial condition we have

$$(4.7) \quad \Psi\left(\frac{e^{-\theta} - 1}{be^{-\theta} - d}\right) = e^{-\theta X_0}.$$

Solving for Ψ we obtain for $b \neq d$ the solution

$$(4.8) \quad L(t, \theta) = \left(\frac{d(e^{-\theta} - 1)e^{(b-d)t} - (be^{-\theta} - d)}{b(e^{-\theta} - 1)e^{(b-d)t} - (be^{-\theta} - d)} \right)^{X_0}$$

and for $b = d$

$$(4.9) \quad L(t, \theta) = \left(\frac{1 - (bt - 1)(e^{-\theta} - 1)}{1 - bt(e^{-\theta} - 1)} \right)^{X_0}.$$

Remark 4.1 Note that the form of the Laplace transforms (4.8), (4.9) implies the branching property, namely, if $X_0 = X_{0,1} + X_{0,2}$, then the probability law of X_t is identical to the distribution of the sum of independent random variables $X_{t,1} + X_{t,2}$ where $X_{t,i}$ are versions of the linear birth and death process with initial conditions $X_{0,1}, X_{0,2}$.

Distribution function, moments, extinction probability

Setting b, d as the birth and death rates. Then replacing θ by $-\ln z$ in $L_t(\theta)$ we obtain the probability generating function

$$(4.10) \quad G_t(z) = L(t, -\ln z) = \sum_{k=0}^{\infty} z^k p_k(t).$$

Then expanding in a power series in z we can obtain the standard formula

$$(4.11) \quad p_0(t) = f(t),$$

$$(4.12) \quad p_n(t) = (1 - f(t))(1 - g(t))g(t)^{n-1}, \quad n \geq 1$$

where

$$(4.13) \quad f(t) = \frac{d(e^{(b-d)t} - 1)}{be^{(b-d)t} - d}, \quad g(t) = \frac{b(e^{(b-d)t} - 1)}{be^{(b-d)t} - d}.$$

Similarly if $b = d = \frac{V}{2}$, then

$$(4.14) \quad p_n(t) = \frac{(bt)^{n-1}}{(1 + bt)^{n+1}}, \quad n \geq 1,$$

$$(4.15) \quad p_0(t) = \frac{bt}{1+bt}.$$

Then the extinction probability is

$$(4.16) \quad \lim_{t \rightarrow \infty} p_0(t) = \lim_{t \rightarrow \infty} \frac{d(e^{(b-d)t} - 1)}{be^{(b-d)t} - d} = \begin{cases} 1 & \text{if } b \leq d, \\ \frac{d}{b} & \text{if } b > d. \end{cases}$$

Recalling that

$$(4.17) \quad E(X_t) = -\left. \frac{\partial L_t(\theta)}{\partial \theta} \right|_{\theta=0},$$

$$(4.18) \quad E((X_t)^2) = \left. \frac{\partial^2 L_t(\theta)}{\partial \theta^2} \right|_{\theta=0}.$$

we can obtain

$$(4.19) \quad E(X_t) = X_0 e^{(b-d)t},$$

$$(4.20) \quad E((X_t)^2) = (X_0)^2 e^{2(b-d)t} + \frac{X_0(b+d)}{b-d} e^{(b-d)t} (e^{(b-d)t} - 1), \quad b \neq d$$

$$(4.21) \quad E((X_t)^2) = (X_0)^2 + 2bt, \quad \text{if } b = d.$$

4.2 Critical branching

Exponential growth of a population is unrealistic and therefore supercritical branching models describe only the growth of a population as long as the resources are unlimited. Otherwise logistic competition comes into play. We will return to this circle of questions throughout this course.

Only critical branching processes have the property that the mean population size is stable but as shown above the critical branching process actually suffers extinction with probability one. Nevertheless critical branching processes have played a key role in the development of stochastic population models. We will later see that a key feature of critical branching is the limiting behavior of the process conditioned on non-extinction up to time t and letting $t \rightarrow \infty$. We now give two formulations of the resulting behavior.

Theorem 4.2 Consider the BGW process Z_n with mean offspring size $m = 1$. Suppose that $\sigma^2 := \text{Var}(\xi) = E[\xi^2] - 1 \leq \infty$. Then

(i) *Kolmogorov*

$$(4.22) \quad \lim_{n \rightarrow \infty} nP[Z_n > 0] = \frac{2}{\sigma^2}$$

(ii) *Yaglom:* If $\sigma < \infty$, then the conditional distribution of $\frac{Z_n}{n}$ given $Z_n > 0$ converges as $n \rightarrow \infty$ to an exponential law with mean $\frac{\sigma^2}{2}$.

We refer the proof of this to the literature [8], [416].

Theorem 4.3 Consider the critical linear birth and death process, $\{X_t\}$ with $\alpha = \frac{1}{2}$, $b = d = \frac{V}{2}$. Then

- (i) Extinction probability: $\lim_{t \rightarrow \infty} p_0(t) = 1$.
- (ii) Expected extinction time: Let $\tau := \inf\{t : X_t = 0\}$ Then $E[\tau] = \infty$.
- (iii) Exponential limit law: conditioned on $X_t \neq 0$,

$$\frac{X_t}{t} \Rightarrow Y$$

where Y is exponential with mean b .

Proof. The proof is based on the explicit form of the generating function (4.9).

- (i) From (4.15), $p_0(t) = \frac{bt}{bt+1} \rightarrow 1$ as $t \rightarrow \infty$.
- (ii) The expected extinction time is infinite

$$E(\tau) = \int_0^\infty (1 - p_0(t))dt = \int_0^\infty \frac{1}{bt+1} dt = \infty.$$

- (iii) From (4.9),

(4.23)

$$\begin{aligned} E(e^{-\frac{X_t \theta}{t}} | X_t \neq 0) &= \frac{L(t, \frac{\theta}{t}) - P(X_t = 0)}{1 - P(X_t = 0)} \\ &= \frac{1 - (bt - 1)(e^{-\frac{\theta}{t}} - 1)1 - bt(e^{-\frac{\theta}{t}} - 1) - \frac{bt}{bt+1}}{\frac{1}{bt+1}} \\ \lim_{t \rightarrow \infty} E(e^{-\frac{X_t \theta}{t}} | X_t \neq 0) &= \frac{1}{1 + b\theta} \end{aligned}$$

and which is the Laplace transform of the exponential distribution with mean b . ■

4.3 Feller's continuous state branching process (CSBP)

Consider the Itô stochastic differential equation (SDE)

$$dX_t = mX_t dt + \sqrt{\gamma X_t} dW_t, \quad X_0 = x \geq 0$$

where $\{W_t\}$ is a standard Brownian motion. This equation has a non-Lipschitz coefficient but its pathwise uniqueness follows from the Yamada-Watanabe theorem [585].

Using Itô's lemma one can then check that the generator of the resulting diffusion process acting on $D(G) = \{f \in C_0^2(\mathbb{R}_+), xf_x, xf_{xx} \in C_0(\mathbb{R}_+^d)\}$ satisfies

$$(4.24) \quad Gf(x) = mx \frac{\partial f}{\partial x} + \frac{1}{2} \gamma x \frac{\partial^2 f}{\partial x^2}$$

and therefore X_t is a realization of the Feller CSBP process.

Proposition 4.4 (*Laplace transform and extinction probability*)

(a) *The Laplace transform is given by*

$$(4.25) \quad L(\theta, t) = \mathbb{E}_x \exp(-\theta X_t) = \exp(-u(t)x)$$

where $u(s)$ satisfies the equation:

$$(4.26) \quad \frac{\partial u}{\partial s} = mu - \frac{\gamma}{2}u^2 \quad u(0) = \theta.$$

(b) *In the critical case $m = 0$*

$$(4.27) \quad P_x(x_t = 0) = \exp\left(-\frac{x}{\gamma t}\right).$$

Proof. Assume that $\theta(s) \geq 0$ is differentiable. Then applying Itô's lemma ([493], Theorem 3.3, Remark 1) to $F(\theta, x) = e^{-\theta x}$ we have

$$\begin{aligned} (4.28) \quad & F(\theta(t), X_t) - F(\theta(0), X_0) \\ &= m \int_0^t X_s F_2(\theta(s), X_s) ds + \int_0^t F_2(\theta(s), X_s) \sqrt{\gamma X_s} dW_s \\ &+ \int_0^t F_1(\theta(s), X_s) d\theta(s) + \frac{\gamma}{2} \int_0^t X_s F_{22}(\theta(s), X_s) ds \end{aligned}$$

Noting that $E(X_s e^{-\theta X_s}) = -L_1(\theta, s)$ we obtain

$$\begin{aligned} (4.29) \quad & \frac{\partial L(\theta(s), s)}{\partial s} = L_1(\theta(s), s) \frac{d\theta(s)}{ds} - m\theta(s)L_1(\theta(s), s) + \frac{\gamma}{2}\theta(s)^2 L_1(\theta(s), s) \\ & \text{with } L(\theta, 0) = e^{-\theta x} \end{aligned}$$

If u is a solution of

$$(4.30) \quad \frac{\partial u(\theta, s)}{\partial s} = mu(\theta, s) - \frac{\gamma}{2}u^2(\theta, s), \quad u(\theta, 0) = \theta$$

then the derivative with respect to s

$$\frac{\partial}{\partial s} L(u(\theta, t-s), s) = 0, \quad 0 \leq s \leq t$$

and therefore

$$(4.31) \quad \mathbb{E}_x(e^{-\theta X_t}) = L(\theta, t) = L(u(\theta, t), 0) = e^{-u(\theta, t)x}.$$

(b) Solving (4.30) we get

$$(4.32) \quad u(\theta, t) = \frac{\theta}{(1 + t\gamma\theta)}, \quad \text{if } m = 0$$

$$(4.33) \quad u(\theta, t) = \frac{\theta m e^{mt}}{m + \gamma\theta(e^{mt} - 1)}, \quad \text{if } m \neq 0.$$

If $m = 0$

$$(4.34) \quad P_{X_0}(x_t = 0) = \lim_{\theta \rightarrow \infty} e^{-x_0 u(\theta, t)} = e^{-\frac{x_0}{\gamma t}}.$$

■

Remark 4.5 An immediate consequence of (4.31) is that for each t , X_t is an infinitely divisible random variable. In fact the law of X_t corresponds to the law of the sum of a Poisson distributed number of independent exponential random variables. These facts will provide an important tool for the study of these processes and their infinite dimensional generalizations.

Feller CSBP with immigration

Adding an immigration term ct to X_t , one obtains the continuous state branching with immigration process (CBI), and can verify (see e.g. Li (2006) [406]) the following:

Proposition 4.6 Consider the continuous subcritical branching process with immigration (CBI), given by the SDE:

$$(4.35) \quad dY_t = cdt - bY_t dt + \sqrt{\gamma Y_t} dW_t, \quad Y_0 = y_0, \quad b, c > 0.$$

(a) The Laplace transform of the distribution of Y_t is given by:

$$(4.36) \quad \mathbb{E}_{y_0} \exp(-\theta Y(t)) = e^{-y_0 u(t) - \int_0^t cu(s) ds}; \quad \frac{\partial u}{\partial t} = -bu - \frac{\gamma}{2}u^2 \quad u(0) = \theta > 0.$$

(b) In the subcritical case Y_t converges to equilibrium, $Y_t \Rightarrow Y_\infty$ as $t \rightarrow \infty$, where Y_∞ has the gamma distribution with Laplace transform

$$(4.37) \quad L(\theta) = \frac{c}{[(b + \gamma\theta) - \gamma\theta e^{-bt}]^{1/\gamma}}.$$

Proof. (a) This can be proved using the method of Theorem 4.4. Alternately, we can prove this by consider the process with immigrants coming according to $\frac{1}{K} \sum \delta_{y_i}$ where $\{y_i\}$ are the points of a Poisson process with rate K and letting $K \rightarrow \infty$.

(b) We obtain

$$(4.38) \quad L_t(\theta) = \frac{c}{[(b + \gamma\theta) - \gamma\theta e^{-bt}]^{1/\gamma}}.$$

from (a) by simple integration of (4.36). (b) then follows by taking $t \rightarrow \infty$. ■

Remark 4.7 The critical Feller CSBP with immigration

$$(4.39) \quad \begin{aligned} dY_t &= \beta dt + 2\sqrt{Y_t} dW_t \\ Y_0 &= y_0 \end{aligned}$$

is the square of a β -dimensional Bessel process. (See Revuz Yor [493] where this is called a $BESQ^\beta$ process). For $\beta \geq 2$, $\{0\}$ is polar. For $0 < \beta < 2$, $\{0\}$ is instantaneously reflecting. For $0 < \beta < 1$ the set $\{t : X_t = 0\}$ is a perfect set. (See Revuz Yor [493] Chap. XI.)

4.4 Diffusion limits of critical and nearly critical branching processes

4.4.1 Convergence to Feller's continuous state branching process

In a celebrated paper Feller (1951) [232] developed the diffusion approximation to branching processes using semigroup methods.

Theorem 4.8 (*Convergence of B+D and BGW processes to Feller CSBP*)

(a) Consider the sequence of birth and death process, $\{X_t^K\}$, $K \in \mathbb{N}$, with linear birth and death rates $b_K = 1 + \frac{m}{2K}$, $d_K = 1 - \frac{m}{2K}$ with $X_0^K = \lfloor Kz \rfloor$. Assume that $\frac{\lfloor Kz \rfloor}{K} \rightarrow x$ and let

$$(4.40) \quad Z_t^K := \frac{1}{K} X_{Kt}^K.$$

Then as $K \rightarrow \infty$

$$(4.41) \quad \{Z_t^K\}_{t \geq 0} \Longrightarrow \{Z_t\}_{t \geq 0},$$

where $\{Z_t\}_{t \geq 0}$ is a CSBP with generator G given by (4.24) with $\gamma = 1$ and $Z_0 = x$. The convergence is in the sense of weak convergence of probability measures on $D_{[0,\infty)}([0,\infty))$ and the limiting process is a.s. continuous.

(b) Consider a sequence of BGW processes $\{X_k^N\}$ with mean offspring sizes

$$(4.42) \quad E(\xi^N) = m_N = 1 + \frac{m}{N}$$

and offspring variances

$$(4.43) \quad \text{Var}(\xi^N) = \gamma > 0.$$

Let

$$(4.44) \quad Z_t^N := \frac{1}{N} X_{\lfloor Nt \rfloor}^N.$$

Assume that $Z_0^N \rightarrow Z_0$ as $N \rightarrow \infty$. Then

$$(4.45) \quad \{Z_t^N\}_{t \geq 0} \Longrightarrow \{Z_t\}_{t \geq 0},$$

that is, Z_t^N converges in distribution on $D_{[0,\infty)}([0,\infty))$ to a Markov diffusion process, $\{Z_t\}_{t \geq 0}$, called the Feller continuous state branching process (CSBP). The generator of the CSBP $\{\bar{Z}_t\}$ acting on functions $f \in C_0^2([0,\infty))$ is given by

$$(4.46) \quad Gf(x) = mx \frac{\partial f}{\partial x} + \frac{1}{2} \gamma x \frac{\partial^2 f}{\partial x^2}.$$

Proof. (a) The proof follows a standard program for weak convergence of processes, namely,

- the convergence of the finite dimensional distributions,
- proof that the laws of the processes $P^K \in \mathcal{P}(D_{[0,\infty)}([0,\infty)))$ are tight.

To show that the finite dimensional distributions converge, first substitute birth and death rates $b = 1 + \frac{m}{2K}$, $d = 1 - \frac{m}{2K}$, in (4.8) to obtain the Laplace transform of Z_t^K with $Z_0^K = \lfloor Kz \rfloor$ as follows:

(4.47)

$$\begin{aligned} E(e^{-\theta Z_t^K}) &= L^K(t, \theta) \\ &= \left(-\frac{K(e^{-\frac{\theta}{K}} - 1)(e^{mt} - 1) - \frac{m}{2}(e^{-\frac{\theta}{K}} - 1)e^{mt} - \frac{m}{2}(e^{-\frac{\theta}{K}} + 1)}{K(e^{-\frac{\theta}{K}} - 1)(e^{mt} - 1) + \frac{m}{2}(e^{-\frac{\theta}{K}} - 1)e^{mt} - \frac{m}{2}(e^{-\frac{\theta}{K}} + 1)} \right)^{\lfloor Kz \rfloor} \\ &= \left(-\frac{(-\theta)(e^{mt} - 1) + \frac{\theta^2}{2K} - \frac{m}{2}(e^{-\frac{\theta}{K}} - 1)e^{mt} - \frac{m}{2}(e^{-\frac{\theta}{K}} + 1) + O(K^{-2})}{-\theta(e^{mt} - 1) + \frac{\theta^2}{2K} + \frac{m}{2}(e^{-\frac{\theta}{K}} - 1)e^{mt} - \frac{m}{2}(e^{-\frac{\theta}{K}} + 1) + O(K^{-2})} \right)^{\lfloor Kz \rfloor} \\ &\longrightarrow \exp \left(-\frac{m\theta z e^{mt}}{m + \theta(e^{mt} - 1)} \right). \end{aligned}$$

This coincides (see Proposition 4.4) with the Laplace transform at time t of the diffusion process with $Z_0 = x$ and with generator

$$(4.48) \quad Gf(x) = mx \frac{\partial f}{\partial x} + \frac{1}{2}x \frac{\partial^2 f}{\partial x^2}.$$

Using the Markov property and the continuity of the transition probability in x we can then obtain convergence of the finite dimensional distributions.

To complete the proof we must verify that the probability laws of $\{Z_t^K\}_{t \geq 0}$ denoted by $P^K \in \mathcal{P}(D_{[0, \infty)}([0, \infty))$ are tight. We will use the Aldous condition. We first verify that given $\delta > 0$ there exists $0 < L < \infty$ such

$$(4.49) \quad \sup_K P^K(\sup_{0 \leq t \leq T} X^K(t) > L) \leq \delta.$$

Note that the generator of $Z_t^K = \frac{X_{Kt}^K}{K}$ is

$$(4.50) \quad G^K f\left(\frac{n}{K}\right) = \frac{n}{K} \cdot K^2 [f\left(\frac{n+1}{K}\right) + f\left(\frac{n-1}{K}\right) - 2f\left(\frac{n}{K}\right)] + \frac{mn}{2K} \cdot K [f\left(\frac{n+1}{K}\right) - f\left(\frac{n-1}{K}\right)].$$

Then

$$(4.51) \quad M_t^K := Z_t^K - m \int_0^t Z_s^K ds \quad \text{is a martingale.}$$

By Gronwall's inequality

$$(4.52) \quad \sup_{0 \leq t \leq T} Z_t^K \leq \sup_{0 \leq t \leq T} |M_t^K| e^{mt}.$$

Applying Doob's maximal inequality to M_t^K

$$(4.53) \quad P(\sup_{0 \leq t \leq T} |M_t^K| \geq R) \leq \frac{E((M_T^K)^2)}{R^2}.$$

It remains to compute $E((M_T^K)^2)$. We have

$$(4.54) \quad E((M_T^K)^2) \leq E(Z_T^K)^2 + 2|m| \int_0^T E(Z_s^K Z_T^K) ds + \int_0^T \int_0^T E(Z_s^K Z_t^K) ds dt.$$

Using (4.20) we can check that

$$(4.55) \quad E[(Z_t^K)^2] \leq Z_0^K e^{mt} \frac{(e^{mt} - 1)}{m} + (Z_0^K)^2 e^{2mt}.$$

A simple calculation then yields

$$(4.56) \quad E((M_T^K)^2) \leq C(T, z)$$

where $C(T, z)$ does not depend on K which proves (4.49).

We can then apply the Aldous sufficient condition for tightness, namely, given stopping times $\tau_K \leq T$ and $\delta_K \downarrow 0$ as $K \rightarrow \infty$

$$(4.57) \quad \lim_{K \rightarrow \infty} P^K(|Z_{\tau_K + \delta_K}^K - Z_{\tau_K}^K| > \varepsilon) = 0.$$

First note that $X_{\tau_K}^K$ is tight so we can take a convergent subsequence. Then by Skorohod's representation we can put these on a common probability space so that there is a.s. convergence. In this setting assume that $X_{\tau_{K_n}}^{K_n} \rightarrow x$. It now suffices to prove that $X_{\tau_{K_n} + \delta_{K_n}}^{K_n}$ converges in distribution to x . Then by the strong Markov property we have

$$(4.58) \quad \begin{aligned} & E(e^{-\theta(Z_{\tau_K + \delta_K}^K - Z_{\tau_K}^K)} - e^{-\theta Z_{\tau_K}^K} | Z_{\tau_K}^K) \\ &= \left(-\frac{K(e^{-\frac{\theta}{K}} - 1)(e^{m\delta_K} - 1) - \frac{m}{2}(e^{-\frac{\theta}{K}} - 1)e^{m\delta_K} - \frac{m}{2}(e^{-\frac{\theta}{K}} + 1)}{K(e^{-\frac{\theta}{K}} - 1)(e^{m\delta_K} - 1) + \frac{m}{2}(e^{-\frac{\theta}{K}} - 1)e^{m\delta_K} - \frac{m}{2}(e^{-\frac{\theta}{K}} + 1)} \right)^{\lfloor KZ_{\tau_K}^K \rfloor} - e^{-\theta Z_{\tau_K}^K} \\ &\longrightarrow 0 \text{ on } \{ \sup_{0 \leq t \leq T} Z^K(t) \leq L \} \text{ as } K \rightarrow \infty. \end{aligned}$$

Therefore $Z_{\tau_K + \delta_K}^K - Z_{\tau_K}^K \rightarrow 0$ in distribution and for $\varepsilon, \eta > 0$ we can find K_0 such that

$$(4.59) \quad P^K(|Z_{\tau_K + \delta_K}^K - Z_{\tau_K}^K| > \varepsilon) < 2\eta, \quad \forall K \geq K_0.$$

This completes the proof of tightness.

(b) See Ethier and Kurtz ([212] Chapter 9, Theorem 1.3) for a proof based on a semigroup convergence theorem (e.g. [212], Chap. 1, Theorem 6.5). This involves showing that

$$(4.60) \quad \lim_{N \rightarrow \infty} \sup_{x=\frac{\ell}{N}, \ell \in \mathbb{N}} |N(T_N f(x) - f(x)) - Gf(x)| = 0 \quad \forall f \in C_c^\infty([0, \infty)),$$

where

$$(4.61) \quad T_N f(x) = E[f(\frac{1}{N} \sum_{k=1}^{Nx} \xi_k^N)], \quad x \in \{\frac{\ell}{N}, \ell \in \mathbb{N}\}$$

and where $\{\xi_k^N\}$ are i.i.d. satisfy (4.42), (4.43).

■

Remark 4.9 These results can also be proved using the martingale problem formulation in the same way as is carried out below for the Wright-Fisher model.

4.5 The critical BGW tree

4.5.1 The rooted BGW tree as a metric space

We begin by recalling that a BGW tree $\mathcal{T} \in \mathbf{T}$ with root \emptyset is a graph in which the vertices are a subset of

$$(4.62) \quad \mathcal{I} = \emptyset \cup \bigcup_{n=1}^{\infty} \mathbb{N}_0$$

satisfying conditions (3.1.2). Recall that if $x = (i_1, \dots, i_n) \in \mathcal{T}$ is said to be in generation n , denoted by $H_N(x) = n$ where $N = \#(\mathcal{T})$. The edges are given by the set of pairs of the form $(i_1, \dots, i_n), (i_1, \dots, i_n, j)$.

The *lexicographic order* is an order relation on the vertices of \mathcal{T} defined as follows. We say that $x = (i_1, \dots, i_n)$ and $y = (j_1, \dots, j_m)$ have a *last common ancestor* at generation $\ell \geq 1$ if

$$(4.63) \quad (i_1, \dots, i_\ell) = (j_1, \dots, j_\ell) \text{ and } i_{\ell+1} \neq j_{\ell+1} \text{ (or is empty).}$$

Given \mathcal{T} with $\#(\mathcal{T}) = N$ we can order the vertices in lexicographic order $\emptyset, x_1, x_2, \dots, x_{N-1}$. We can then embed it in the plane so x_i appears to the left of x_j if $i < j$.

The corresponding *height function* $H_N(k)$ of a tree of size $\#(\mathcal{T}) = N$ is defined by

$$(4.64) \quad H_N(k) := |x_k|, \quad 0 \leq k \leq N-1$$

where $|x|$ denotes the generation of x .

Note that the number of visits of $H_N(k)$ to n gives the population size at generation n , that is,

$$(4.65) \quad Z_n = \sum_{k=0}^{N-1} 1_{\{n\}}(H_N(k))$$

where $1_{\{n\}}$ denotes the indicator function.

We now define a distance between the individuals in \mathcal{T} . If we assign length 1 to each edge then a metric $d_{\mathcal{T}}(x, y)$ can be defined on \mathcal{T} by

$$(4.66) \quad d_{\mathcal{T}}(x, y) := \text{the length of the shortest path in } \mathcal{T} \text{ from } x \text{ to } y.$$

Since the critical BGW tree is a.s. finite this produces a compact metric space and is an example of random compact rooted real tree which we define below.

Remark 4.10 Note that a reordering of the offspring (in the lexicographic order) defines a root preserving isometry. We can then associate to \mathcal{T} the corresponding equivalence class of plane trees (modulo the family of root preserving isometries). This equivalence class is characterized by $(\#(\mathcal{T}), \emptyset, d_{\mathcal{T}}(., .))$.

We now briefly introduce the reduced tree at generation n . We denote the set of n th generation individuals

$$(4.67) \quad X_n = \mathcal{T} \cap \mathbb{N}_0^n.$$

The *reduced tree*

$$(4.68) \quad \mathcal{T}_n^R := \{x \in \mathcal{T} : x = (i_1, \dots, i_r), r = 1, \dots, n, \text{ such that } \exists (i_1, \dots, i_n) \in X_n\}.$$

We also define a metric on X_n by $d_n(x, y) := n - \ell$ if the last common ancestor of x, y is in generation $\ell < n$. It is easy to verify that d_n is an *ultrametric*, that is,

$$(4.69) \quad d_n(x, y) \leq \max(d_n(x, z), d_n(z, y)) \quad \text{for any } z \in X_n.$$

4.5.2 The contour functions

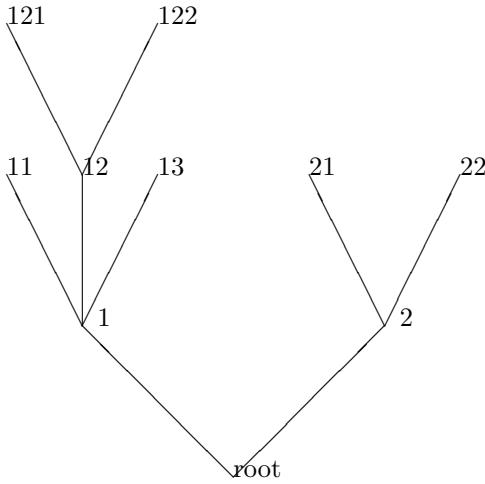
Given a tree \mathcal{T} with $\#(\mathcal{T}) < \infty$ we define the *contour function*

$$(4.70) \quad C^{\mathcal{T}} = C^{\mathcal{T}}(t) : 0 \leq t \leq 2(\#(\mathcal{T}) - 1)$$

which is obtained by taking a particle that starts from the root of \mathcal{T} and visits continuously all edges at speed one, moving away from the root if possible otherwise going backwards along the edge leading to the root and respecting the lexicographical order of vertices. The domain of $C^{\mathcal{T}}$ can be extended to $[0, \infty)$ by setting $C^{\mathcal{T}}(t) = 0$ for $t > 2(\#(\mathcal{T}) - 1)$. In other words, $C^{\mathcal{T}}$ is a piecewise linear process given by the distance from the root as we move through the tree.

We have considered above the Yaglom conditioned limit theorem (Theorem 4.1) for a critical BGW process. Similarly it is of interest to consider the conditioned BGW process conditioned on $\#(\mathcal{T})$. In order to formulate results for this we need to introduce two additional notions, real trees and the Gromov-Hausdorff metric.

Figure 4.1: BGW Tree and contour function, $N = 10$



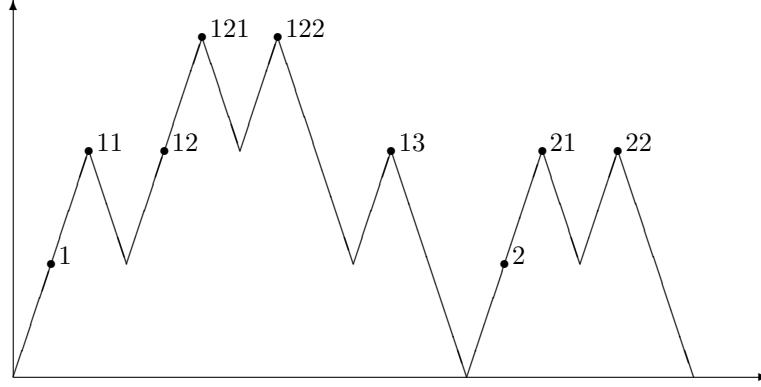
4.5.3 Real trees

Following Evans [229] and Le Gall [398] we now introduce the notion of real trees and their coding. See Dress and Terhalle [170], [171] for general background on “tree theory”.

Definition 4.11 A metric space (\mathcal{T}, d) is a *real tree* if the following two properties hold for every $(x, y) \in \mathcal{T}$.

- there is a unique isometric map $f_{x,y}$ from $[0, d(x,y)]$ into \mathcal{T} such that $f_{x,y}(0) = x$ and $f_{x,y}(d(x,y)) = y$

Contour function



- If q is a continuous injective map from $[0, 1]$ into \mathcal{T} , such that $q(0) = x$, $q(1) = y$, then

$$(4.71) \quad q([0, 1]) = f_{x,y}([0, d(x, y)]).$$

A rooted real tree is a real tree (\mathcal{T}, d) with a distinguished vertex \emptyset called the root.

As explained above it is natural to consider the equivalence class \mathbb{T} of real trees (\mathcal{T}, d) modulo the family of root preserving isometries. Since this results in a collection of compact metric spaces, it can be furnished with the Gromov-Hausdorff metric d_{GH} (see Appendix II, section 13.5).

(Recall that $d_{GH}((E_1, d_1), (E_2, d_2))$ is given by the infimum of the Hausdorff distances of the images of $(E_1, d_1), (E_2, d_2)$ under the set of isometric embeddings of $(E_1, d_1), (E_2, d_2)$, respectively, into a common compact metric space (E_0, d_0) .)

Proposition 4.12 (Evans, Pitman, Winter (2003) [228]). *The space of real trees furnished with the Gromov-Hausdorff topology, (\mathbb{T}, d_{GH}) , is Polish.*

Remark 4.13 A metric space (E, d) can be embedded isometrically into a real tree iff the four point condition

$$(4.72) \quad d(x, y) + d(u, v) \leq \max(d(x, u) + d(y, v), d(x, v) + d(y, u))$$

is satisfied for all 4-tuples u, v, x, y (Dress (1984) , [169])

4.5.4 Excursions from zero and real trees

Consider a continuous function $g : [0, \infty) \rightarrow [0, \infty)$ with non-empty compact support such that $g(0) = 0$ and $g(s) = 0 \forall s > \inf\{t >: g(t) = 0\}$ (we call this an positive excursion from 0). For $s, t \geq 0$, let

$$(4.73) \quad m_g(s, t) = \inf_{r \in [s \wedge t, s \vee t]} g(r),$$

$$(4.74) \quad d_g(s, t) = g(s) + g(t) - 2m_g(s, t).$$

It is easy to check that d_g is symmetric and satisfies the triangle inequality. Let \mathcal{T}_g denote the quotient space $[0, \infty)/\equiv$ where $s \equiv t$ if $d_g(s, t) = 0$. Then it can be verified that the metric space (\mathcal{T}_g, d_g) is a real tree (Le Gall (2006) [398], Theorem 2.1).

Given g the ancestral relationships can be reconstructed by noting that s is an ancestor of t , $s \prec t$ iff $g(s) = \inf_{[s,t]} g(r)$

Let $(\mathbb{C}, \|\cdot\|) := (\{(g, d_g) : g \text{ a positive excursion from } 0, d_g = \text{sup norm metric}\})$.

It can be verified that (e.g. Le Gall (2006) [398], Lemma 2.3)) the mapping from $(\mathbb{C}, \|\cdot\|)$ to (\mathbb{T}, d_{GH}) is continuous, that is, for two continuous functions g, g' such that $g(0) = g'(0) = 0$:

$$(4.75) \quad d_{GH}(\mathcal{T}_g, \mathcal{T}_{g'}) \leq 2 \|g - g'\|.$$

4.5.5 The Aldous Continuum Random Tree

Let $\{B_t\}_{t \geq 0}$ be a standard Brownian motion and

$$(4.76) \quad \tau_1 := \sup\{t \in [0, 1] : B_t = 0\}, \quad \tau_2 := \inf\{t \geq 1 : B_t = 0\}.$$

Then the *Brownian excursion* is a nonhomogeneous Markov process defined as follows:

$$(4.77) \quad B_t^e := \frac{1}{\sqrt{(\tau_2 - \tau_1)}} B(\tau_1 + t(\tau_2 - \tau_1)), \quad 0 \leq t \leq 1.$$

It can be shown (see Itô-McKean) [315] that the marginal PDF is given by

$$(4.78) \quad f(t, x) = \frac{2x^2}{\sqrt{2\pi t^3(1-t)^3}} e^{-\frac{x^2}{2t(1-t)}}.$$

Definition 4.14 (Aldous continuum random tree) *Let $(B_t^e)_{0 \leq t \leq 1}$ be a normalized Brownian excursion (extended to $[0, \infty)$ by setting $B_t^e = 0$ for $t > 1$). The corresponding random real tree (\mathcal{T}^e, d^e) is called the continuum random tree (CRT). We denote by $P^{CRT} \in \mathcal{P}((\mathbb{T}, d_{GH}))$ the probability law of \mathcal{T}^e .*

The CRT was introduced by Aldous (1991-1993) in a series of papers [4], [5] and [6].

4.5.6 Conditioned limit theorem for the critical BGW tree

Consider the special case of a BGW process with geometric offspring distribution, that is,

$$(4.79) \quad p_k = P(\xi = k) = \frac{1}{2^{k+1}}, \quad k = 0, 1, 2, \dots,$$

Lemma 4.15 *For the offspring distribution (4.79) the contour process $C^\mathcal{T}$ is given by a simple random walk $\{S_k\}$ with*

$$(4.80) \quad P(S_{k+1} - S_k = \pm 1) = \frac{1}{2}.$$

Proof. This can be verified by first noting that in this case the number of jumps from 0 to 1 corresponds to the number of offspring of the initial vertex. Now let $\tau_1^k, \tau_2^k, \dots$ denote the

times of visits to height k . Consider the m th such visit to height k , $k \geq 1$. The corresponding vertex is the offspring of a vertex at height $k - 1$, say, the ℓ th offspring. Then

$$(4.81) \quad C(\tau_m^k + 1) = \begin{cases} k + 1, & \text{with probability } P(\xi \geq \ell + 1 | \xi \geq \ell) = \frac{1}{2} \\ k - 1 & \text{with probability } P(\xi = \ell | \xi \geq \ell) = \frac{1}{2} \end{cases}$$

■

Proposition 4.16 *Let $P^{BGW}(\cdot | \#(\mathcal{T}) = n) \in \mathcal{P}((\mathbb{T}, d_{GH}))$ denote the probability law of the BGW tree with offspring distribution (4.79) conditioned to have n vertices. Then*

$$(4.82) \quad P^{BGW}\left(\frac{\mathcal{T}}{2\sqrt{2n}} | \#(\mathcal{T}) = n\right) \Rightarrow P^{CRT}$$

in the sense of weak convergence in $\mathcal{P}((\mathbb{T}, d_{GH}))$.

Proof. Letting $S_0 = 0$, and $N = \min\{k > 0 : S_k = 0\}$ and conditioning on $N = 2n$ we have the contour process for this BGW process to have total population n . Note that this is simply an excursion of the simple random walk conditioned to first return to the origin at time $N = 2n$, S_k^N . But it is known that which rescaled converges as $n \rightarrow \infty$ to a Brownian excursion from 0 (see Durrett, Iglehart and Miller (1977) [173]).

$$(4.83) \quad \left(\frac{S^N(\lfloor 2nu \rfloor)}{2\sqrt{2n}}\right)_{0 \leq u \leq 1} \Rightarrow (B_u^e)_{0 \leq u \leq 1}.$$

where B^e is the standard Brownian excursion. Using (4.75) and the continuous mapping theorem ([38], Theorem 2.7) this implies that the laws of the corresponding BGW trees converge to the CRT as $n \rightarrow \infty$. ■

4.5.7 Aldous Invariance Principle for BGW trees

A remarkable result of Aldous is the *invariance principle* for scaling limit of critical BGW tree, that is, the CRT arises as the limit for the entire class of critical BGW processes with aperiodic offspring distributions having finite second moments.

Theorem 4.17 (*Invariance principle for BGW trees - Aldous (1993)* [6], Theorem 23.)

Consider the critical BGW tree with offspring distribution μ . Assume that μ is aperiodic with variance $\sigma^2 < \infty$. Then the distribution of the rescaled tree

$$\frac{\sigma}{2\sqrt{n}} \mathcal{T}$$

under the probability measure $P^{BGW}(\cdot | \#(\mathcal{T}) = n)$ (i.e. conditioning that total population up to extinction is n) converges as $n \rightarrow \infty$ to the law of CRT.

Proof. The proof is given in [6]. It is too long and complex to include here.

However some of the ideas behind the proof are as follows. Using (4.75) we see that the result would follow if the rescaled contour process

$$(4.84) \quad \left(\frac{\sigma}{2\sqrt{n}} C^{\mathcal{T}}(2nt) : 0 \leq t \leq 1\right)$$

under the probability measure $P^{BGW}(\cdot | \#(\mathcal{T}) = n)$ converges in distribution to the normalized Brownian excursion. In the general case can no longer be represented by the excursion of a simple random walk. Aldous (1993) [6] proof of the invariance result is based on a characterization of the distribution of the CRT. Marckert and Mokkadem (2003) [423] gave an alternate proof (assuming the offspring distribution has exponential moments) involving only the contour and height functions. In particular they proved that for any critical offspring distribution with variance σ^2 the weak convergence of the rescaled contour function and height processes. Their key idea is to couple the height process to the random walk (“depth-first queue process”)

$$(4.85) \quad S_n(j) = \sum_{i=0}^{j-1} (\xi_i - 1), \quad 1 \leq j \leq n,$$

that is, with jump distribution is given by $q_i = p_{i+1}$, $i = -1, 0, 1, 2, \dots$, conditioned by $S_n(0) = 0$, $S_n(i) \geq 0$, $1 \leq i \leq n-1$, $S_n(n) = -1$. Then

$$(4.86) \quad H_n(\ell) = \text{Card}\{j : 0 \leq j \leq \ell-1, \min_{0 \leq k \leq \ell-j} S_n(j+k) = S_n(j)\}, \quad 0 \leq \ell < n-1.$$

They then obtain exponential bounds on deviations between the height process and the conditioned random walk S_n to prove that

$$(4.87) \quad \left(\frac{H_n(nt)}{\sqrt{n}} \right)_{0 \leq t \leq 1} \Rightarrow \left(\frac{2}{\sigma} B^e(t) \right)_{0 \leq t \leq 1}.$$

■

Remark 4.18 It has also been proved that starting the BGW process with n individuals then the rescaled height function

$$(4.88) \quad \left\{ \frac{1}{n} H_n(\lfloor n^2 t \rfloor) \right\}_{t \geq 0} \rightarrow (H_t)_{t \geq 0} \quad \text{with } H_0 = 1$$

where

$$(4.89) \quad H_t = (B_t - \inf_{0 \leq s \leq t} B_s)$$

where B_t is a Brownian motion, that is, H_t is reflecting Brownian motion. (See [397]).

Recall that the Ray-Knight Theorem ([494], 52.1) states that if B_t is a Brownian motion with local time $\{\ell_t^a\}$

$$(4.90) \quad T := \inf\{u : \ell_u^0 > 1\},$$

then the Brownian local time $\{\ell_T^a : a \geq 0\}$ has the same law as the Feller CSB satisfying

$$(4.91) \quad dZ_t = 2\sqrt{Z_t} dW_t, \quad Z_0 = 1.$$

In other words the local time of the height process is a version of the Feller CSB starting at 1. More precisely, the initial mass $Z_0 = 1$ corresponds to the local time at 0 of a reflecting Brownian motion on $[0, T]$ and for $t \geq 0$ $Z_t = \ell_T^a$, that is the occupation density of the reflecting Brownian motion.

4.6 Remark on general continuous state branching

By the basic result of Silverstein [518] the general continuous state branching process has log-Laplace equation

$$(4.92) \quad u_t(\lambda) + \int_0^t \psi(u_s(\lambda)) = \lambda,$$

with

$$(4.93) \quad \psi(u) = \alpha u + \beta u^2 + \int_0^\infty (e^{-ru} - 1 + ru) \nu(dr)$$

where $\alpha, \beta \geq 0$ and ν is a σ -finite measure on $(0, \infty)$ such that $\int(r \wedge r^2)\nu(dr) < \infty$. This include the class of $(1 + \beta)$ CSB which arise as limits of BGW processes in which the offspring distribution has infinite second moments and are related to stable processes and other Lévy processes. The genealogical structure, stable continuum trees and convergence of the contour process in this general setting have been developed by Duquesne and LeGall [189] but we do not consider this major topic here.

Chapter 5

Wright-Fisher Processes

5.1 Introductory remarks

The BGW processes and birth and death processes we have studied in the previous chapters have the property that

$$(5.1) \quad X_n \rightarrow 0 \text{ or } \infty, \quad a.s.$$

A more realistic model is one in which the population grows at low population densities and tends to a steady state near some constant value. The Wright-Fisher model that we consider in this chapter (and the corresponding Moran continuous time model) assume that the total population remains at a constant level N and focusses on the changes in the relative proportions of the different types. Fluctuations of the total population, provided that they do not become too small, result in time-varying resampling rates in the Wright-Fisher model but do not change the main qualitative features of the conclusions.

The branching model and the Wright-Fisher idealized models are complementary. The branching process model provides an important approximation in two cases:

- If the total population density becomes small then the critical and near critical branching process provides an useful approximation to compute extinction probabilities.
- If a new type emerges which has a competitive advantage, then the supercritical branching model provides a good approximation to the growth of this type as long as its contribution to the total population is small.

Models which incorporate multiple types, supercritical growth at low densities and have non-trivial steady states will be discussed in a later chapter. The advantage of the idealized models we discuss here is the possibility of explicit solutions.

5.2 Wright-Fisher Markov Chain Model

The classical neutral Wright-Fisher (1931) model is a discrete time model of a population with constant size N and types $E = \{1, 2\}$. Let X_n be the number of type 1 individuals at time n . Then X_n is a Markov chain with state space $\{0, \dots, N\}$ and transition probabilities:

$$P(X_{n+1} = j | X_n = i) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}, \quad j = 0, \dots, N.$$

In other words at generation $n+1$ this involves binomial sampling with probability $p = \frac{X_n}{N}$, that is, the current empirical probability of type 1. Looking backwards from the viewpoint of generation $n+1$ this can be interpreted as having each of the N individuals of the $(n+1)$ st generation “pick their parents at random” from the population at time n .

Similarly, the neutral K -allele Wright Fisher model with types $E_K = \{e_1, \dots, e_K\}$ is given by a Markov chain X_n with state space $\setminus(E_K)$ (counting measures) and

$$(5.2) \quad P(X_{n+1} = (\beta_1, \dots, \beta_K) | X_n = (\alpha_1, \dots, \alpha_K)) \\ = \frac{N!}{\beta_1! \beta_2! \dots \beta_K!} \left(\frac{\alpha_1}{N} \right)^{\beta_1} \dots \left(\frac{\alpha_K}{N} \right)^{\beta_K}$$

In this case the binomial sampling is simply replaced by multinomial sampling.

Consider the multinomial distribution with parameters (N, p_1, \dots, p_K) . Then the moment generating function is given by

$$(5.3) \quad M(\theta_1, \dots, \theta_K) = E(\exp(\sum_{i=1}^K \theta_i X_i)) = \left(\sum_{i=1}^K p_i e^{\theta_i} \right)^N$$

Then

$$(5.4) \quad E(X_i) = Np_i, \quad \text{Var}(X_i) = Np_i(1-p_i),$$

and

$$(5.5) \quad \text{Cov}(X_i, X_j) = -Np_i p_j, \quad i \neq j.$$

Remark 5.1 We can relax the assumptions of the Wright-Fisher model in two ways. First, if we relax the assumption of the total population constant, equal to N , we obtain a Wright-Fisher model with variable resampling rate (e.g. Donnelly and Kurtz [164] and Kaj and Krone [340]).

To introduce the second way to relax the assumptions note that we can obtain the Wright-Fisher model as follows. Consider a population of N individuals in generation n with possible types in E_K , Y_1^n, \dots, Y_N^n . Assume each individual has a Poisson number of offspring with mean m , (Z_1, \dots, Z_N) and the offspring is of the same type as the parent. Then

$$\text{conditioned on } \sum_{i=1}^N Z_i = N,$$

the resulting population $(Y_1^{(n+1)}, \dots, Y_N^{(n+1)})$ is multinomial $(N; \frac{1}{N}, \dots, \frac{1}{N})$, that is, we have a multitype (Poisson) branching process conditioned to have constant total population N . If we then define

$$(5.6) \quad p_{n+1}(i) = \frac{1}{N} \sum_{j=1}^N 1(Y_j^{(n+1)} = i), \quad i = 1, \dots, K,$$

then $(p_{n+1}(1), \dots, p_{n+1}(K))$ is multinomial $(N; p_n(1), \dots, p_n(K))$ where

$$(5.7) \quad p_n(i) = \frac{1}{N} \sum_{j=1}^N 1(Y_j^n = i), \quad i = 1, \dots, K.$$

We can generalize this by assuming that the offspring distribution of the individuals is given by a common distribution on \mathbb{N}_0 . Then again conditioned the total population to have constant size N the vector $(Y_1^{n+1}, \dots, Y_N^{n+1})$ is exchangeable but not necessarily multinomial. This exchangeability assumption is the basis of the Cannings Model (see e.g. Ewens [230]).

A basic phenomenon of neutral Wright-Fisher without mutation is *fixation*, that is, the elimination of all but one type at a finite random time. To see this note that for each $j = 1, \dots, K$, $\delta_j \in \mathcal{P}(E_K)$ are absorbing states and $X_n(j)$ is a martingale. Therefore $X_n \rightarrow X_\infty$, *a.s.* Since $\text{Var}(X_{n+1}) = NX_n(1 - X_n)$, this means that $X_\infty = 0$ or 1 , *a.s.* and X_n must be 0 or 1 after a finite number of generations (since only the values $\frac{k}{N}$ are possible).

5.2.1 Types in population genetics

The notion of type in population biology is based on the *genotype*. The genotype of an individual is specified by the *genome* and this codes *genetic information* that passes, possibly modified, from parent to offspring (parents in sexual reproduction). The genome consists of a set of *chromosomes* (23 in humans). A chromosome is a single molecule of DNA that contains many genes, regulatory elements and other nucleotide sequences. A given position on a chromosome is called a *locus* (*loci*) and may be occupied by one or more *genes*. Genes code for the production of a protein. The different variations of the gene at a particular locus are called *alleles*. The ordered list of loci for a particular genome is called a *genetic map*. The *phenotype* of an organism describes its structure and behaviour, that is, how it interacts with its environment. The relationship between genotype and phenotype is not necessarily 1-1. The field of *epigenetics* studies this relationship and in particular the mechanisms during cellular development that produce different outcomes from the same genetic information.

Diploid individuals have two homologous copies of each chromosome, usually one from the mother and one from the father in the case of sexual reproduction. Homologous chromosomes contain the same genes at the same loci but possibly different alleles at those genes.

5.2.2 Finite population resampling in a diploid population

For a diploid population with K -alleles e_1, \dots, e_K at a particular gene we can focus on the set of types denoted by $E_K^{2\circ}$ consisting of the set of $\frac{K(K+1)}{2}$ unordered pairs (e_i, e_j) . The genotype (e_i, e_j) is said to be homozygous (at the locus in question) if $e_i = e_j$, otherwise heterozygous.

Consider a finite population of N individuals. Let

$$P_{ij} = \text{proportion of type } (e_i, e_j)$$

Then, p_i , the proportion of allele e_i is

$$p_i = P_{ii} + \frac{1}{2} \sum_{j \neq i} P_{ij}.$$

The probability $\{P_{ij}\}$ on $E_K^{2\circ}$ is said to be a Hardy-Weinberg equilibrium if

$$(5.8) \quad P_{ij} = (2 - \delta_{ij}) p_i p_j.$$

This is what is obtained if one picks independently the parent types e_i and e_j from a population having proportions $\{p_i\}$ (in the case of sexual reproduction this corresponds to “random mating”).

Consider a diploid Wright-Fisher model with N individuals therefore $2N$ genes with random mating. This means that an individual at generation $(n + 1)$ has two genes randomly chosen from the $2N$ genes in generation n .

In order to introduce the notions of identity by descent and genealogy we assume that in generation 0 each of the $2N$ genes correspond to different alleles. Now consider generation n . What is the probability, F_n , that an individual is homozygous, that is, two genes selected at

random are of the same type (homozygous)? This will occur only if they are both descendants of the same gene in generation 0.

First note that in generation 1, this means that an individual is homozygous only if the same allele must be selected twice and this has probability $\frac{1}{2N}$. In generation $n + 1$ this happens if the same gene is selected twice or if different genes are selected from generation n but they are identical alleles. Therefore,

$$(5.9) \quad F_1 = \frac{1}{2N}, \quad F_n = \frac{1}{2N} + (1 - \frac{1}{2N})F_{n-1}.$$

Let $H_n := 1 - F_n$ (heterozygous). Then

$$(5.10) \quad H_1 = 1 - \frac{1}{2N}, \quad H_n = (1 - \frac{1}{2N})H_{n-1}, \quad H_n = (1 - \frac{1}{2N})^n$$

Two randomly selected genes are said to be *identical by descent* if they are the same allele. This will happen if they have a *common ancestor*. Therefore if $T_{2,1}$ denotes the time in generations back to the common ancestor we have

$$(5.11) \quad P(T_{2,1} > n) = H_n = (1 - \frac{1}{2N})^n, \quad n = 0, 1, 2, \dots,$$

$$(5.12) \quad P(T_{2,1} = n) = \frac{1}{2N}(1 - \frac{1}{2N})^{n-1}, \quad n = 1, 2, \dots$$

Similarly, for k randomly selected genes they are identical by descent if they all have a common ancestor. We can consider the time $T_{k,1}$ in generations back to the most recent common ancestor of k individuals randomly sampled from the population. We will return to discuss the distribution of $T_{k,1}$ in the limit as $N \rightarrow \infty$ in Chapter 9.

5.2.3 Diploid population with mutation and selection

In the previous section we considered only the mechanism of resampling (genetic drift). In addition to genetic drift the basic genetic mechanisms include mutation, selection and recombination. In this subsection we consider the Wright-Fisher model incorporating mutation and selection.

For a diploid population of size N with mutation, selection and resampling the *reproduction cycle* can be modelled as follows (cf [212], Chap. 10). We assume that in generation 0 individuals have genotypic proportions $\{P_{ij}\}$ and therefore the proportion of type i (in the population of $2N$ genes) is

$$p_i = P_{ii} + \frac{1}{2} \sum_{j \neq i} P_{ij}.$$

Stage I:

In the first stage diploid cells undergo meiotic division producing haploid gametes (single chromosomes), that is, meiosis reduces the number of sets of chromosomes from two to one. The resulting *gametes* are haploid cells; that is, they contain one half a complete set of chromosomes. When two gametes fuse (in animals typically involving a sperm and an egg), they form a *zygote* that has two complete sets of chromosomes and therefore is diploid. The zygote receives one set of chromosomes from each of the two gametes through the *fusion* of the two gametes. By the

assumption of random mating, then in generation 1 this produces zygotes in *Hardy-Weinberg proportions* $(2 - \delta_{ij})p_i p_j$.

Stage II: Selection and Mutation.

Selection. The resulting zygotes can have different viabilities for survival. The viability of (e_i, e_j) has viability V_{ij} . Then the proportions of surviving zygotes are proportional to the product of the viabilities and the Hardy-Weinberg proportions, that is,

$$(5.13) \quad P_{k,\ell}^{\text{sel}} = \frac{V_{k\ell} \cdot (2 - \delta_{k\ell})p_k p_\ell}{\sum_{k' \leq \ell'} (2 - \delta_{k'\ell'})V_{k'\ell'} p_{k'} p_{\ell'}}$$

Mutation. We assume that each of the 2 gametes forming zygote can (independently) mutate with probability p_m and that if a gamete of type e_i mutates then it produces a gamete of type e_j with probability m_{ij} .

(5.14)

$$\begin{aligned} P_{ij}^{\text{sel,mut}} &= (1 - \frac{1}{2}\delta_{ij}) \sum_{k \leq \ell} (m_{ki} m_{\ell j} + m_{kj} m_{\ell i}) P_{k\ell}^{\text{sel}} \\ &= (1 - \frac{1}{2}\delta_{ij}) \sum_{k \leq \ell} (m_{ki} m_{\ell j} + m_{kj} m_{\ell i}) \frac{V_{k\ell} \cdot (2 - \delta_{k\ell})p_k p_\ell}{\sum_{k' \leq \ell'} (2 - \delta_{k'\ell'})V_{k'\ell'} p_{k'} p_{\ell'}} \end{aligned}$$

Stage III: Resampling. Finally random sampling reduces the population to N adults with proportions P_{ij}^{next} where

$$(5.15) \quad (P_{ij}^{\text{next}})_{i \leq j} \sim \frac{1}{N} \text{multinomial}(N, (P_{ij}^{\text{sel,mut}})_{i \leq j}).$$

We then obtain a population of $2N$ gametes with proportions

$$(5.16) \quad p_i^{\text{next}} = P_{ii}^{\text{next}} + \frac{1}{2} \sum_{j \neq i} P_{ij}^{\text{next}}.$$

Therefore we have defined the process $\{X_n^N\}_{n \in \mathbb{N}}$ with state space $\mathcal{P}^N(E_K)$. If X_n^N is a Markov chain we defined the transition function

$$P(X_{n+1}^N = (p_1^{\text{next}}, \dots, p_K^{\text{next}}) | X_n^N = (p_1, \dots, p_K)) = \pi_{p_1, \dots, p_K}(p_1^{\text{next}}, \dots, p_K^{\text{next}})$$

where the function π is obtained from (5.14), (5.15), (5.16). See Remark 5.7.

5.3 Diffusion Approximation of Wright-Fisher

5.3.1 Neutral 2-allele Wright-Fisher model

As a warm-up to the use of diffusion approximations we consider the case of 2 alleles A_1, A_2 , ($k = 2$). Let X_n^N denote the number of individuals of type A_1 at the n th generation. Then as above $\{X_n^N\}_{n \in \mathbb{N}}$ is a Markov chain.

Theorem 5.2 (*Neutral case without mutation*) Assume that $N^{-1}X_0^N \rightarrow p_0$ as $N \rightarrow \infty$. Then

$$\{p_N(t) : t \geq 0\} \equiv \{N^{-1}X_{[Nt]}^N, t \geq 0\} \implies \{p(t) : t \geq 0\}$$

where $\{p(t) : t \geq 0\}$ is a Markov diffusion process with state space $[0, 1]$ and with generator

$$(5.17) \quad Gf(p) = \frac{1}{2}p(1-p)\frac{d^2}{dp^2}f(p)$$

if $f \in C^2([0, 1])$. This is equivalent to the pathwise unique solution of the SDE

$$\begin{aligned} dp(t) &= \sqrt{p(t)(1-p(t))}dB(t) \\ p(0) &= p_0. \end{aligned}$$

Proof. Note that in this case X_{n+1}^N is Binomial(N, p_n) where $p_n = \frac{X_n^N}{N}$. Then from the Binomial formula,

$$\begin{aligned} E_{X_n^N}\left(\frac{X_{n+1}^N}{N}\right) &= \frac{X_n^N}{N} \\ E_{X_n^N}\left[\left(\frac{X_{n+1}^N}{N} - \frac{X_n^N}{N}\right)^2 \mid \frac{X_n^N}{N}\right] &= \frac{1}{N} \left(\frac{X_n^N}{N} \left(1 - \frac{X_n^N}{N}\right)\right). \end{aligned}$$

We can then verify that

$$(5.18) \quad \{p_N(t) := N^{-1}X_{[Nt]}^N : t \geq 0\} \text{ is a martingale}$$

with

$$\begin{aligned} (5.19) \quad E(p_N(t_2) - p_N(t_1))^2 &= E \sum_{k=\lfloor Nt_1 \rfloor}^{\lfloor Nt_2 \rfloor} (p_N(\frac{k+1}{N}) - p_N(\frac{k}{N}))^2 \\ &= \frac{1}{N} E \sum_{k=\lfloor Nt_1 \rfloor}^{\lfloor Nt_2 \rfloor} p_N(\frac{k}{N})(1 - p_N(\frac{k}{N})) \end{aligned}$$

and then that

$$(5.20) \quad M_N(t) = p_N^2(t) - \frac{1}{N} \sum_{k=0}^{\lfloor Nt \rfloor} p_N(\frac{k}{N})(1 - p_N(\frac{k}{N}))$$

is a martingale.

Let $P_{p_N}^N \in \mathcal{P}(D_{[0,1]}([0, \infty)))$ denote the probability law of $\{p_N(t)\}_{t \geq 0}$ with $p_N(0) = p_N$. From this we can prove that the sequence $\{P_{p_N(0)}^N\}_{N \in \mathbb{N}}$ is tight on $\mathcal{P}(D_{[0,\infty)}([0, 1]))$. To verify this as in the previous chapter we use Aldous criterion $P_{p_N(0)}^N(p_N(\tau_N + \delta_N) - p_N(\tau_N) > \varepsilon) \rightarrow 0$ as $N \rightarrow \infty$ for any stopping times $\tau_N \leq T$ and $\delta_N \downarrow 0$. This follows easily from the strong Markov property, (5.19) and Chebyshev's inequality. Since the processes $p_N(\cdot)$ are bounded it then follows that for any limit point P_{p_0} of $P_{p_N(0)}^N$ we have

$$\begin{aligned} \{p(t)\}_{t \geq 0} &\text{ is a bounded martingale with } p(0) = p_0 \text{ and with increasing process} \\ (5.21) \quad \langle p \rangle_t &= \int_0^t p(s)(1 - p(s))ds. \end{aligned}$$

Since the largest jump of $p_N(\cdot)$ goes to 0 as $N \rightarrow \infty$ the limiting process is continuous (see Theorem 17.14 in the Appendix). Also, by the Burkholder-Davis-Gundy inequality we have

$$(5.22) \quad E((p(t_2) - p(t_1))^4) \leq \text{const} \cdot (t_2 - t_1)^2,$$

so that $p(t)$ satisfies Kolmogorov's criterion for a.s. continuous.

We can then prove that there is a unique solution to this *martingale problem*, that is, for each p there exists a unique probability measure on $C_{[0,\infty)}([0,\infty))$ satisfying (5.21) and therefore this defines a Markov diffusion process with generator (5.17).

The uniqueness can be proved by determining all joint moments of the form

$$(5.23) \quad E_p((p(t_1)^{k_1} \dots (p(t_\ell))^{k_\ell})), \quad 0 \leq t_1 < t_2 < \dots < t_\ell, \quad k_i \in \mathbb{N}$$

by solving a closed system of differential equation. It can also be proved using duality and this will be done in detail below (Chapter 7) in a more general case.) ■

We now give an illustrative application of the diffusion approximation, namely the calculation of expected fixation times.

Corollary 5.3 (*Expected fixation time.*) *Let $\tau := \inf\{t : p(t) \in \{0, 1\}\}$ denote the fixation time of the diffusion process. Then*

$$E_p[\tau] = g(p) = -[p \log p + (1-p) \log(1-p)].$$

Proof. Let $f \in C^2([0, 1])$, $f(0) = f(1) = 0$. Let $g_f(p) := \int_0^\infty T_s f(p) ds$, and note that as $f \uparrow 1_{(0,1)}$ this converges to the expected time spent in $(0, 1)$. Since $p(t) \rightarrow \{0, 1\}$ as $t \rightarrow \infty$, a.s., we can show that

$$G \left(\int_0^t T_s f(p) ds \right) = \int_0^t GT_s f(p) ds = T_t f(p) - f(p) \rightarrow 0 - f(p) \text{ as } t \rightarrow \infty,$$

that is,

$$Gg(p) = -f(p)$$

where G is given by (5.17).

Applying this to a sequence of C^2 functions increasing to $1_{(0,1)}$ we get

$$\begin{aligned} E_p(\tau) &= \int_0^\infty P_p(\tau > t) dt \\ &= \int_0^\infty T_t 1_{(0,1)}(p) dt \\ &= g(p) \end{aligned}$$

We then obtain $g(p)$ by solving the differential equation $Gg(p) = -1$ with boundary conditions $g(0) = g(1) = 0$ to obtain

$$E_p[\tau] = g(p) = -[p \log p + (1-p) \log(1-p)].$$

■ Let τ_N denote the fixation time for $N^{-1}X_{[Nt]}$. We want to show that

$$(5.24) \quad E_{\frac{x_0^N}{N}}[\tau_N] \rightarrow E_{p_0}[\tau] \text{ if } \frac{X_0^N}{N} \rightarrow p \text{ and } N \rightarrow \infty.$$

However note that τ is not a continuous function on $D([0, \infty), [0, 1])$. The weak convergence can be proved for $\tau^\varepsilon = \inf\{t : p(t) \notin (\varepsilon, 1 - \varepsilon)\}$ (because there is no “slowing down” here). To complete the proof it can be verified that for $\delta > 0$

$$(5.25) \quad \lim_{\varepsilon \rightarrow 0} \limsup_{N \rightarrow \infty} P(|\tau_N^\varepsilon - \tau_N| > \delta) = 0$$

(see Ethier and Kurtz, [212], Chapt. 10, Theorem 2.4).

2-allele Wright-Fisher with mutation

For each N consider a Wright-Fisher population of size M_N and with mutation rates $m_{12} = \frac{u}{N}$, $A_1 \rightarrow A_2$ and $m_{21} = \frac{v}{N}$, $A_2 \rightarrow A_1$.

In this case X_{n+1}^N is Binomial(M_N, p_n) with

$$(5.26) \quad p_n = \left(1 - \frac{u}{N}\right) \frac{X_n^N}{M_N} + \frac{v}{N} \left(1 - \frac{X_n^N}{M_N}\right).$$

We now consider

$$(5.27) \quad p_N(t) = \frac{1}{M_N} X_{\lfloor Nt \rfloor}.$$

If we assume that

$$(5.28) \quad \gamma = \lim_{N \rightarrow \infty} \frac{N}{M_N},$$

then the diffusion approximation is given by the diffusion process p_t with generator

$$(5.29) \quad Gf(p) = \frac{\gamma}{2} p(1-p) \frac{\partial^2}{\partial p^2} + [-up + v(1-p)] \frac{\partial}{\partial p}.$$

In this case the domain of the generator involves boundary conditions at 0 and 1 (see [212], Chap. 8, Theorem 1.1) but we will not need this.

Remark 5.4 Note that the diffusion coefficient is proportional to the inverse population size. Below for more complex models we frequently think of the diffusion coefficient in terms of inverse effective population size.

Error estimates

Consider a haploid Wright-Fisher population of size M with mutation rates $m_{12} = u$, $m_{21} = v$.

Let $p_t^{(M,u,v)}$ denote the diffusion process with generator (5.29) with $\gamma = \frac{1}{M}$. Then if $\alpha, \beta \geq 0$, the law of

$$(5.30) \quad \{Z_t^{(\alpha,\beta)}\}_{t \geq 0} := p_t^{(M, \frac{\alpha}{M}, \frac{\beta}{M})}$$

is independent of M and is a Wright-Fisher diffusion with generator

$$(5.31) \quad Gf(p) = \frac{\gamma}{2} p(1-p) \frac{\partial^2}{\partial p^2} + [-\alpha p + \beta(1-p)] \frac{\partial}{\partial p}.$$

The assumption of mutation rates of order $O(\frac{1}{N})$ corresponds to the case in which both mutation and genetic drift are of the same order and appear in the limit as population sizes goes to ∞ . Other only one of the two mechanisms appears in the limit as $N \rightarrow \infty$.

On the other hand one can consider the diffusion process as an approximation to the finite population model. Ethier and Norman ([210]) obtained an estimate of the error due to the diffusion approximation in the calculation of the expected value of a smooth function of the nth generation allelic frequency.

To formulate their result consider the Wright-Fisher Markov chain model $\{X_n^{(M,u,v)}\}$ with population size M and one-step mutation probabilities $m_{12} = u$, $m_{21} = v$ and $p_t^{(M,u,v)}$ the Wright-Fisher diffusion with generator (5.29) with $\gamma = \frac{1}{M}$.

Theorem 5.5 (Ethier and Norman [210]) *Assume that $f \in C^6([0, 1])$. Then for $n \in \mathbb{N}_0$,*

$$(5.32) \quad \begin{aligned} & |E_x(f(X_n^{(M,u,v)}) - E_x(f(p_n^{(M,u,v)}))| \\ & \leq \frac{\max(u, v)}{2} \cdot \|f^{(1)}\| + \frac{1}{M} \left(\frac{1}{8} \|f^{(2)}\| + \frac{1}{216\sqrt{3}} \|f^{(3)}\| \right) \\ & + \frac{9 \max(u^2, v^2)}{2} \left(\sum_{j=1}^6 \|f^{(j)}\| \right) + \frac{7}{16M^2} \sum_{j=2}^6 \|f^{(j)}\| \end{aligned}$$

where $\|f^{(j)}\|$ is the sup of the j th derivative of f .

We do not include a proof but sketch the main idea. Let

$$(5.33) \quad (S_n f)(x) := E_x[f(X_n^{(M,u,v)})],$$

$$(5.34) \quad (T_t f)(x) := E_x[f(p_t^{(M,u,v)})].$$

If $g \in C_b^6([0, \infty))$, then we have the Taylor expansions

$$(5.35) \quad (T_1 g)(x) = g(x) + (Gg)(x) + \frac{G^2 g(x)}{2} + \omega_2 \frac{\|G^3 g\|}{6}, \quad |\omega_2| \leq 1$$

and

$$(5.36) \quad \begin{aligned} (S_1 g)(x) &= g(x) + \sum_{j=1}^5 E_x[(X_1^{(M,u,v)} - x)^j] \frac{g^{(j)}(x)}{j!} + \omega_1 E_x[(X_1^{(M,u,v)} - x)^6] \frac{\|g^{(6)}\|}{6!}, \quad |\omega_1| \leq 1. \end{aligned}$$

We then obtain

$$(5.37) \quad \|S_1 g - T_1 g\|_M \leq \sum_{j=1}^6 \gamma_j \|g^{(j)}\|$$

for some constants γ_j .

The proof is then completed using the inequality

$$(5.38) \quad \|S_n f - T_n f\|_M \leq \sum_{k=0}^{n-1} \|(S_1 - T_1)T_k\|_M$$

where $\|\cdot\|_M$ is the sup norm on $\{\frac{j}{M} : j = 1, \dots, M\}$.

5.3.2 K-allele Wright-Fisher Diffusion

Now consider the K -allele Wright-Fisher Markov chain $\{X_k^{2N}\}_{k \in \mathbb{N}}$ with $2N$ gametes present in each generation and assume that the mutation rates and fitnesses satisfy

$$(5.39) \quad m_{ij} = \frac{q_{ij}}{2N}, \quad i \neq j, \quad m_{ii} = 1 - \frac{m}{N}, \quad m = \sum_j q_{ij}$$

$$(5.40) \quad V_{ij} = 1 + \frac{\sigma_{ij}}{2N} + O\left(\frac{1}{N^2}\right).$$

We now consider the Markov process with state space

$$(5.41) \quad \Delta_{K-1} := \{(p_1, \dots, p_K) : p_i \geq 0, \sum_{i=1}^K p_i = 1\}.$$

defined by

$$(5.42) \quad \{p^{2N}(t) : t \geq 0\} \equiv \left\{ \frac{1}{2N} X_{[2Nt]}^{2N}, \quad t \geq 0 \right\}.$$

Theorem 5.6 Assume that $2N^{-1} X_0^{2N} \rightarrow p$ as $N \rightarrow \infty$ in Δ_{K-1} .

Then the laws of the càdlàg processes $\{p_N(t) := \frac{1}{2N} X_t^{2N}\}_{t \geq 0}$ are tight and for any limit point and function $f(p) = f(p_1, \dots, p_{K-1}) \in C^2(\Delta_{K-1})$,

$$(5.43) \quad M_f(t) := f(p(t)) - \int_0^t G^K f(p(s)) ds \quad \text{is a martingale}$$

where

$$(5.44) \quad \begin{aligned} & G^K f(p) \\ &= \frac{1}{2} \sum_{i,j=1}^{K-1} p_i (\delta_{ij} - p_j) \frac{\partial^2 f(p)}{\partial p_i \partial p_j} \\ &+ \sum_{i=1}^{K-1} \left[m \left(\sum_{j=1, j \neq i}^K q_{ji} p_j - p_i \right) + p_i \left(\sum_{j=1}^K \sigma_{ij} p_j - \sum_{k,\ell}^K \sigma_{k\ell} p_k p_\ell \right) \right] \frac{\partial f(p)}{\partial p_i}. \end{aligned}$$

The martingale problem (5.43) has a unique solution which determines a Markov diffusion process $\{p(t) : t \geq 0\}$ called the K-allele Wright-Fisher diffusion.

Proof. Following the pattern of the 2-allele neutral case the proof involves three steps which we now sketch.

Step 1. The tightness of the probability laws P^N of $\{p^{2N}(\cdot)\}$ on $D_{\Delta_{K-1}}([0, \infty))$ can be proved using Aldous criterion.

Step 2. Proof that for any limit point of P^N and $i = 1, \dots, K$

$$(5.45) \quad \begin{aligned} M_i(t) &:= p_i(t) - p_i(0) - \int_0^t \left[m \left(\sum_{j=1}^K q_{ji} p_j(s) - p_i(s) \right) \right. \\ &\quad \left. + p_i(s) \left(\sum_{j=1}^K \sigma_{ij} p_j(s) - \sum_{k,\ell}^K \sigma_{k\ell} p_k(s) p_\ell(s) \right) \right] ds \end{aligned}$$

is a martingale with quadratic covariation process

$$(5.46) \quad \langle M_i, M_j \rangle_t = \frac{1}{2} \int_0^t p_j(s)(\delta_{ij} - p_i(s))ds$$

To verify this, let $\mathcal{F}_{\frac{k}{2N}} = \sigma\{p_i^{2N}(\frac{\ell}{2N}) : \ell \leq k, i = 1, \dots, K\}$. Then we have for $k \in \mathbb{N}$

$$\begin{aligned} & E[p_i^{2N}(\frac{k+1}{2N}) - p_i^{2N}(\frac{k}{2N}) | \mathcal{F}_{\frac{k}{2N}}] \\ &= \frac{1}{2N} \left[m \left(\sum_{j=1, j \neq i}^K \frac{q_{ji}}{m} p_j^{2N}(\frac{k}{2N}) - p_i^{2N}(\frac{k}{2N}) \right) \right. \\ (5.47) \quad & \left. + \left(\sum_{j=1}^K \sigma_{ij} p_j^{2N}(\frac{k}{2N}) - \sum_{k, \ell=1}^K \sigma_{k\ell} p_k^{2N}(\frac{k}{2N}) p_\ell^{2N}(\frac{k}{2N}) \right) \right] \\ &+ o(\frac{1}{2N}) \end{aligned}$$

$$(5.48) \quad \text{Cov}(p_i^{2N}(\frac{k+1}{2N}), p_j^{2N}(\frac{k+1}{2N}) | \mathcal{F}_{\frac{k}{2N}}) = \frac{p_i^{2N}}{2N}(\frac{k}{2N})(\delta_{ij} - p_j^{2N}(\frac{k}{2N})) + o(\frac{1}{N})$$

Remark 5.7 The Markov property for X_n^N follows if in the resampling step the $\{P_{ij}^{\text{sel,mut}}\}$ are in Hardy-Weinberg proportions which implies that the $\{p_i^{\text{next}}\}$ are

$$(5.49) \quad \text{multinomial}(2N, (p_1^{\text{sel,mut}}, \dots, p_K^{\text{sel,mut}})).$$

This is true without selection or with multiplicative selection $V_{ij} = V_i V_j$ (which leads to haploid selection in the diffusion limit) but not in general. In the diffusion limit this can sometimes be dealt with by the $O(\frac{1}{N^2})$ term in (5.40). In general the diffusion limit result remains true but the argument is more subtle. The idea is that the selection-mutation changes the allele frequencies more slowly than the mechanism of Stages I and III which rapidly bring the frequencies to Hardy-Weinberg equilibrium - see [212], Chap. 10, section 3.

Then for each N and i

$$\begin{aligned} (5.50) \quad M_i^{2N}(t) &:= p_i^{2N}(t) - p_i^{2N}(0) - \int_0^t \left[m \left(\sum_{j=1, j \neq i}^K q_{ji} p_j^{2N}(s) - p_i^{2N}(s) \right) \right. \\ &\quad \left. + p_i^{2N}(s) \left(\sum_{j=1}^K \sigma_{ij} p_j^{2N}(s) - \sum_{k, \ell} \sigma_{k\ell} p_k^{2N}(s) p_\ell^{2N}(s) \right) \right] ds^N + o(\frac{1}{N}) \end{aligned}$$

is a martingale and for $i, j = 1, \dots, K$

$$\begin{aligned} (5.51) \quad & E[(M_i^{2N}(t_2) - M_i^{2N}(t_1))(M_j^{2N}(t_2) - M_j^{2N}(t_1))] \\ &= \frac{1}{2N} E \sum_{k=\lfloor 2N t_1 \rfloor}^{\lfloor 2N t_2 \rfloor} p_i^{2N}(\frac{k}{2N})(\delta_{ij} - p_j^{2N}(\frac{k}{2N})) + o(\frac{1}{N}). \end{aligned}$$

Step 3. Proof that there exists a unique probability measure on $C_{\Delta_{K-1}}([0, \infty))$ such that (5.45) and (5.46) are satisfied.

Uniqueness can be proved in the neutral case, $\sigma \equiv 0$, by showing that moments are obtained as unique solutions of a closed system of differential equations.

■

Remark 5.8 *The uniqueness when σ is not zero follows from the dual representation developed in the next chapter.*

5.4 Stationary measures

A special case of a theorem in Section 8.3 implies that if the matrix (q_{ij}) is irreducible, then the Wright-Fisher diffusion is ergodic with unique stationary distribution.

5.4.1 The Invariant Measure for the neutral K-alleles WF Diffusion

Consider the neutral K -type Wright-Fisher diffusion with type-independent mutation (Kingman's "house-of-cards" mutation model) with generator

$$G^K f(p) = \frac{1}{2} \sum_{i,j=1}^{K-1} p_i(\delta_{ij} - p_j) \frac{\partial^2 f(p)}{\partial p_i \partial p_j} + \frac{\theta}{2} \sum_{i=1}^{K-1} (\nu_i - p_i) \frac{\partial f(p)}{\partial p_i}.$$

where the *type-independent mutation kernel* is given by $\nu \in \Delta_{K-1}$.

Theorem 5.9 (*Wright [?], Griffiths [267]*) *The Dirichlet distribution $D(p_1, \dots, p_n)$ on Δ_{K-1} with density*

$$\begin{aligned} \Pi_K(dp) &= \frac{\Gamma(\theta_1 + \dots + \theta_K)}{\Gamma(\theta_1) \dots \Gamma(\theta_K)} p_1^{\theta_1-1} \dots p_K^{\theta_K-1} dp_1 \dots dp_{K-1} \\ \theta_j &= \theta \nu_j, \quad \nu \in \mathcal{P}(1, \dots, K) \end{aligned}$$

is a reversible stationary measure for the neutral K -alleles WF diffusion with $\gamma = 1$.

In the case $K = 2$ this is the Beta distribution

$$(5.52) \quad \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)} x_1^{\theta_1-1} (1-x_1)^{\theta_2-1} dx_1.$$

Proof. (cf. [211]) Reversibility and stationarity means that when Π_K is the initial distribution, then $\{p(t) : 0 \leq t \leq t_0\}$ has the same distribution as $\{p(t_0 - t) : 0 \leq t \leq t_0\}$. In terms of the strongly continuous semigroup $\{T(t)\}$ on $C(\Delta_{K-1})$ generated by G a necessary and sufficient condition (see Fukushima and Stroock (1986) [246]) for reversibility with respect to Π_K is that

$$\int g T(t) f d\Pi_K = \int f T(t) g d\Pi_K \quad \forall f, g \in C(\Delta_{K-1}), \quad t \geq 0$$

or equivalently that

$$\int g G f d\Pi_K = \int f G g d\Pi_K \quad \forall f, g \in D(G)$$

or for f, g in a core for G (see Appendix I).

Since the space of polynomials in p_1, \dots, p_K is a core for G it suffices by linearity to show that

$$\int gGfd\Pi = \int fGgd\Pi \quad \forall f = f_\alpha, g = f_\beta$$

where $f_\alpha = p_1^{\alpha_1} \dots p_K^{\alpha_K}$. Let $|\alpha| = \sum \alpha_i$.

Then

$$\begin{aligned} & \int f_\beta Gf_\alpha d\Pi_K \\ &= \frac{1}{2} \int [\sum_{i=1}^K \alpha_i(\alpha_i + \theta_i - 1) f_{\alpha+\beta-e^i} - |\alpha|(|\alpha| + \sum_{i=1}^K \theta_i - 1) f_{\alpha+\beta}] d\Pi_K \\ &= \frac{1}{2} \left\{ \sum_{i=1}^K \frac{\alpha_i(\alpha_i + \theta_i - 1)}{\alpha_i + \beta_i + \theta_i - 1} - \frac{|\alpha|(|\alpha| + \sum \theta_i - 1)}{|\alpha| + |\beta| + \sum \theta_i - 1} \right\} \\ &\cdot \frac{\Gamma(\alpha_1 + \beta_1 + \theta_1) \dots \Gamma(\alpha_K + \beta_K + \theta_K)}{\Gamma(|\alpha| + |\beta| + \sum \theta_i - 1)} \frac{\Gamma(\sum \theta_i)}{\Gamma(\theta_1) \dots \Gamma(\theta_K)}. \end{aligned}$$

To show that this is symmetric in α, β , let $h(\alpha, \beta)$ denote the expression within $\{\dots\}$ above. Then

$$\begin{aligned} & h(\alpha, \beta) - h(\beta, \alpha) \\ &= \sum \frac{\alpha_i^2 - \beta_i^2 + (\alpha_i - \beta_i)(\theta_i - 1)}{\alpha_i + \beta_i + \theta_i - 1} - \frac{|\alpha|^2 - |\beta|^2 + (|\alpha| - |\beta|)(\sum \theta_i - 1)}{|\alpha| + |\beta| + \sum \theta_i - 1} \\ &= \sum (\alpha_i - \beta_i) - (|\alpha| - |\beta|) \\ &= 0 \end{aligned}$$

■

Corollary 5.10 Consider the mixed moments:

$$m_{k_1, \dots, k_K} = \int \dots \int_{\Delta_{K-1}} p_1^{k_1} \dots p_K^{k_K} \Pi_K(dp)$$

Then

$$m_{k_1, \dots, k_K} = \frac{\Gamma(\theta_1) \dots \Gamma(\theta_K)}{\Gamma(\theta_1 + \dots + \theta_K)} \frac{\Gamma(\theta_1 + \dots + \theta_K + k_1 + \dots + k_K)}{\Gamma(\theta_1 + k_1) \dots \Gamma(\theta_K + k_K)}.$$

Stationary measure with selection

If selection (as in (5.44)) is added then the stationary distribution is given by the “Gibbs-like” distribution

$$(5.53) \quad \Pi_\sigma(dp) = C \exp \left(\sum_{i,j=1}^K \sigma_{ij} p_i p_j \right) \Pi_K(dp_1 \dots dp_{K-1})$$

and this is reversible. (This is a special case of a result that will be proved in a later section.)

5.4.2 Convergence of stationary measures of $\{p^N\}_{N \in \mathbb{N}}$

It is of interest to consider the convergence of the stationary measures of the Wright-Fisher Markov chains to (5.53). A standard argument applied to the Wright-Fisher model is as follows.

Theorem 5.11 Convergence of Stationary Measures. *Assume that the diffusion limit, $p(t)$, has a unique invariant measure, ν and that ν_N is an invariant measure for $p^N(t)$. Then*

$$(5.54) \quad \nu_N \Rightarrow \nu \text{ as } N \rightarrow \infty.$$

Proof. Denote by $\{T_t\}_{t \geq 0}$ the semigroup of the Wright-Fisher diffusion. Since the state space is compact, the space of probability measure is compact. and therefore the sequence ν_N is tight $M_1(\Delta_{K-1})$. Given a limit point $\tilde{\nu}$ and a subsequence $\nu_{N'}$ that converges weakly to $\tilde{\nu} \in M_1(\Delta_{K-1})$ it follows that for $f \in C(\Delta_{K-1})$,

$$\begin{aligned} \int T(t) f d\tilde{\nu} &= \lim_{N' \rightarrow \infty} \int T(t) f d\nu_{N'} \quad (\text{by } \nu_{N'} \Rightarrow \nu) \\ &= \lim_{N' \rightarrow \infty} \int T_{N'}(2N't) f d\nu_{N'} \quad (\text{by } p_N \Rightarrow p) \\ &= \lim_{N' \rightarrow \infty} \int f d\nu_{N'} \quad (\text{by inv. of } \nu_{N'}) \\ &= \int f d\tilde{\nu} \quad (\text{by } \nu_{N'} \Rightarrow \tilde{\nu}). \end{aligned}$$

Therefore $\tilde{\nu}$ is invariant for $\{T(t)\}$ and hence $\tilde{\nu} = \nu$ by assumption of the uniqueness of the invariant measure for $p(t)$. That is, any limit points of $\{\nu_N\}$ coincides with ν and therefore $\nu_N \Rightarrow \nu$. ■

Properties of the Dirichlet Distribution

1. Consistency under merging of types.

Under $D(\theta_1, \dots, \theta_n)$, the distribution of $(X_1, \dots, X_k, 1 - \sum_{i=1}^k X_i)$ is

$$D(\theta_1, \dots, \theta_k, \theta_{k+1} + \dots + \theta_n)$$

and the distribution of $\frac{X_k}{1 - \sum_{i=1}^{k-1} X_i} = \frac{X_k}{\sum_{i=k}^K X_i}$ is $Beta(\theta_k, \sum_{i=k+1}^K \theta_i)$.

2. Bayes posterior under random sampling

Consider the n-dimensional Dirichlet distribution, $\mathcal{D}(\alpha)$ with parameters $(\alpha_1, \dots, \alpha_n)$. Assume that some phenomena is described by a random probability vector $p = (p_1, \dots, p_n)$. Let $\mathcal{D}(\alpha)$ be the “prior distribution of the vector p . Now let us assume that we take a sample and observe that N_i of the outcome are i . Now compute the posterior distribution of p given the observations $N = (N_1, \dots, N_n)$ as follows: Using properties of the Dirichlet distribution we can show that it is

$$\begin{aligned} P(p \in dx | N) &= \frac{1}{Z} \frac{x_1^{\alpha_1} \dots x_n^{\alpha_n} x_1^{N_1} \dots x_n^{N_n}}{\int x_1^{\alpha_1} \dots x_n^{\alpha_n} x_1^{N_1} \dots x_n^{N_n} dx_1 \dots dx_n} \\ &= \frac{1}{Z'} x_1^{(\alpha_1 + N_1)} \dots x_n^{(\alpha_n + N_n)}. \end{aligned}$$

That is,

$$(5.55) \quad P(p \in \cdot | N) \text{ is } \mathcal{D}(\alpha_1 + N_1, \dots, \alpha_n + N_n).$$

Chapter 6

Infinitely many types models

6.1 Introduction

6.1.1 Motivation

We have considered particle systems above with finitely many types. In the 1970's with the advent of electrophoresis and molecular biology, new models were needed in which the number of types were not fixed. In many cases the number of types can be random and new types can be introduced at random times. Several models began to appear at that time involving infinitely many types, for example the *ladder* or *stepwise mutation model* of Ohta and Kimura (1973) [465] (which could model for example continuous characteristics). Another model was one in which no attempt to model the structure of types was made but in which new types can be introduced (leading to the *infinitely many alleles* model) (Kimura and Crow (1964) [362]). In this model we take $[0, 1]$ as the type space. Then when a new type is needed we can choose a type in $[0, 1]$ by sampling from the uniform distribution on $[0, 1]$. The *infinitely many sites model* introduced by Kimura in 1969 provides an idealization of the genome viewed as a sequence of nucleotides (A,T,C,G). These processes now form the basis for molecular population genetics.

More generally, such infinitely many type models provide the possibility of coding information at a number of levels and provide a powerful tool for the study of complex systems. For example we can code historical information, genealogical information, and information about the random environment that has been visited. In addition it allows for individuals with internal structure described by an internal state space and state transition dynamics.

6.1.2 Plan

The objective of this chapter is to construct two basic infinitely-many-type processes, formulated as measure-valued processes, by taking the projective limit of the finite type Feller CSB and Wright-Fisher diffusions. These are the *Jirina measure-valued branching process* and *infinitely many alleles model* of Crow and Kimura. The latter has played a central role in population genetics. We will establish a relation between the invariant measures of these two processes that allows us to obtain the basic properties of the *Poisson-Dirichlet distribution* and the *Griffiths, Engen and McCloskey (GEM) representation*.

We begin by considering the diffusion limit of a measure-valued generalization of the Wright-Fisher Markov chain in the setting of semigroup theory. This process, the *Fleming-Viot process*, includes the infinitely many alleles model as a special case. In the next Chapter we reformulate these processes in terms of measure-valued martingale problems and develop techniques for

working with more general classes of measure-valued processes including the class of superprocesses and the class of Fleming-Viot processes with selection, mutation and recombination.

6.2 Measure-valued Wright-Fisher Markov chain

We now consider a Wright-Fisher model of a population of N individuals in which the space of types is a separable metric space E . The process is then a Markov chain $\{p_n^N\}_{n \in \mathbb{N}}$ with state space

$$\mathcal{P}^N(E) = \left\{ \mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}, \{x_1, \dots, x_N\} \in E \right\} \subset \mathcal{P}(E).$$

In this case the mutation process is a Markov chain on E with probability transition function $P(x, dy)$ giving the type distribution of the offspring of a type x parent if mutation occurs. Let $V > 0$ be a measurable function on E with $V(x)$ interpreted as the (haploid) fitness of a type x individual.

Then as in the finitely many type case X_n^N is a Markov chain in $\mathcal{P}^N(E)$ with one step transition function $P(\mu, d\mu^{\text{next}})$. This is obtained by noting that X_{n+1}^N is a random probability measure on $\mathcal{P}^N(E)$ given by:

$$(6.1) \quad X_{n+1}^N = \frac{1}{N} \sum_{i=1}^N \delta_{y_i}$$

where y_1, \dots, y_N are i.i.d. $\mu_n^*(dy)$ where

$$\mu_n^*(dy) = \int_E \left(\frac{V(x) X_n^N(dx)}{\int V(x) X_n^N(dx)} \right) P(x, dy),$$

that is, as before selection first and then mutation and sampling.

Example 6.1 *Infinitely many alleles model [362]. $E = [0, 1]$ and*

$$(6.2) \quad P(x, dz) = (1 - m)\delta_x(dz) + m \int_0^1 \delta_y(dz)\lambda(dy)$$

where λ is Lebesgue measure on $[0, 1]$.

Example 6.2 *The infinitely many sites model was introduced by Kimura [364], [365]. (See also Ethier and Griffiths (1987) [213])*

Infinitely many sites model. $E = [0, 1]^{\mathbb{Z}^+}$

$$P(\mathbf{x}, dy) = (1 - m)\delta_{\mathbf{x}}(dy) + m \int_0^1 \delta_{\{\xi, \mathbf{x}\}}\lambda(d\xi)$$

Here we interpret ξ as the locus on the genome where the last mutation occurred.

The number of segregating sites is the number of homologous DNA positions that differ in a sample of m sequences. They are used to investigate phylogenetic relationships. The location of polymorphisms within humans are also used to determine the potential differences in reactions of individuals to medical treatments.

See Section 8.3.3 for the analysis of segregating sites.

Example 6.3 Ladder model of Ohta and Kimura (1973) [465]. This stepwise-mutation model was introduced to describe the distribution of allelic types distinguishable as signed electrical charges in gel electrophoresis experiments.

Here $E = \mathbb{Z}$ and

$$(6.3) \quad P(x, dy) = (1 - m)\delta_x(\cdot) + \frac{m}{2}\delta_{x+1}(\cdot) + \frac{m}{2}\delta_{x-1}(\cdot).$$

6.3 The neutral Fleming-Viot process with mutation generator A

The infinitely many alleles diffusion of Crow and Kimura can be studied as an infinite dimensional diffusion (i.e. countably many types) (see Ethier and Kurtz (1981) [211]) in which a mutation always leads to a new type. However it is advantageous to reformulate it as a measure-valued process. This process was introduced by Fleming and Viot in 1979 [242]. We will show that it arises as the diffusion limit of the measure-valued Wright-Fisher model.

We will now derive the Fleming-Viot process under some simplifying assumptions using semigroup methods. The general case will be dealt with below in the martingale problem setting.

Assumptions

- Let E be a compact metric space.
- $V \equiv 1$, that is, we omit the selection effect.
- We consider a mutation process given by a Feller process on E with generator $(D(A), A)$ and semigroup $\{S_t : t \geq 0\}$ on $C(E)$. A will be called the mutation operator for the Fleming-Viot process.

We assume that $D(A)$ contains an algebra $D_0(A)$ that separates points and $S_t : D_0(A) \rightarrow D_0(A)$. Then linear combinations of functions in $D_0(A)$ form an algebra of functions separating points and therefore is dense in $C(E)$ and therefore measure-determining. We also assume that A arises as the limit of a sequence of mutation Markov chains on E with transition kernels $\{P_N(\cdot, \cdot)\}_{N \in \mathbb{N}}$, that is for $f \in D(A)$,

$$(6.4) \quad N(\langle f, P_N \rangle - f) \rightarrow Af \text{ as } N \rightarrow \infty$$

uniformly on E .

The state space for the Fleming-Viot (FV) process is $\mathcal{P}(E)$, the set of Borel probability measures on E with the topology of weak convergence. For $f \in C(E)$, $\mu \in \mathcal{P}$ we denote $\langle f, \mu \rangle = \int f d\mu$.

We will now obtain the neutral FV process as the limit of neutral (i.e. $V \equiv 1$) Wright-Fisher Markov chains in the diffusion time scale,

$$(6.5) \quad p^N(t) = X_{[Nt]}^N \in \mathcal{P}(E)$$

where X_n^N is defined by (6.1) with mutation transition functions $P^N(x, dy)$.

In order to identify the limiting generator for a $\mathcal{P}(E)$ -valued diffusion we need a measure-determining family of test functions. Consider the algebra \mathcal{D} of nice functions on $\mathcal{P}(E)$ containing the functions:

$$(6.6) \quad F(\mu) = \langle f_1, \mu \rangle \dots \langle f_n, \mu \rangle$$

with $n \geq 1$ and $f_1, \dots, f_n \in D(A)$. This algebra of functions is measure-determining in $\mathcal{P}(\mathcal{P}(E))$.

Notation 6.4 For $F \in \mathcal{D}$, $x \in E$ we define

$$\begin{aligned}\frac{\partial F(\mu)}{\partial \mu(x)} &= \lim_{\varepsilon \rightarrow 0} \frac{F(\mu + \varepsilon \delta_x) - F(\mu)}{\varepsilon} \Big|_{\varepsilon=0} = \frac{\partial F(\mu + \varepsilon \delta_x)}{\partial \varepsilon} \Big|_{\varepsilon=0} \\ \frac{\partial^2 F(\mu)}{\partial \mu(x) \partial \mu(y)} &= \frac{\partial^2 F(\mu + \varepsilon_1 \delta_x + \varepsilon_2 \delta_y)}{\partial \varepsilon_1 \partial \varepsilon_2} \Big|_{\varepsilon_1=\varepsilon_2=0}\end{aligned}$$

Proposition 6.5 Let p_n^N denote the measure-valued Wright-Fisher Markov chain (6.1) under the above assumptions. Then $p^N(t) = X_{[Nt]}^N \Rightarrow p_t$ where $\{p_t\}_{t \geq 0}$ is a $\mathcal{P}(E)$ -valued Markov process with generator

$$\begin{aligned}GF(\mu) &= \sum_{1 \leq i < j \leq n} (\langle f_i f_j, \mu \rangle - \langle f_i, \mu \rangle \langle f_j, \mu \rangle) \prod_{\ell: \ell \neq i, j} \langle f_\ell, \mu \rangle + \sum_i \langle Af_i, \mu \rangle \prod_{\ell: \ell \neq i} \langle f_\ell, \mu \rangle \\ &= \frac{1}{2} \left[\int \frac{\partial^2 F(\mu)}{\partial \mu(x) \partial \mu(y)} \delta_x(dy) \mu(dx) - \int \frac{\partial^2 F(\mu)}{\partial \mu(x) \partial \mu(y)} \mu(dx) \mu(dy) \right] \\ &\quad + \int A \frac{\partial F(\mu)}{\partial \mu(x)} \mu(dx).\end{aligned}$$

for all $F \in \mathcal{D}$.

Proof. Here we follow the Ethier-Kurtz semigroup approach. Using the Kurtz semigroup convergence theorem ([212], Chap. 1, Theorem 6.5, Proposition 3.7 and Chap. 4, Theorem 2.5 -see Appendix III Theorems 14.1, 14.2). Using these results it suffices to show that for $F \in \mathcal{D}$,

$$(6.7) \quad \lim_{N \rightarrow \infty} NE_\mu[F(p_{\frac{1}{N}}^N) - F(\mu)] = \lim_{N \rightarrow \infty} NE_\mu[F(X_1^N) - F(\mu)] = GF(\mu)$$

uniformly in $\mu \in \mathcal{P}(E)$.

First note that for $f_1, \dots, f_n \in C(E)$

$$E_\mu(F(X_1^N)) = E_\mu[\langle f_1, X_1^N \rangle \dots \langle f_n, X_1^N \rangle], \quad F(\mu) = \langle f_1, \mu \rangle \dots \langle f_n, \mu \rangle$$

where $X_1^N = \frac{1}{N} \sum_{i=1}^N \delta_{Y_i}$ and Y_1, \dots, Y_N are i.i.d. μP^N . Hence

$$\begin{aligned}E_\mu[\langle f_1, X_1^N \rangle \dots \langle f_n, X_1^N \rangle] &= \frac{1}{N^n} E \left[\sum_{i=1}^N f_1(Y_i) \dots \sum_{i=1}^N f_n(Y_i) \right] \\ &= \sum_{k=1}^n \frac{N^{[k]}}{N^n} \sum_{\beta \in \pi(n, k)} \prod_{j=1}^k \left\langle \left\langle \prod_{i \in \beta_j} f_i, \mu P^N \right\rangle, \mu \right\rangle \\ &= \sum_{k=1}^n \frac{N^{[k]}}{N^n} \sum_{\beta \in \pi(n, k)} \prod_{j=1}^k \left\langle \left\langle \prod_{i \in \beta_j} f_i, P^N \right\rangle, \mu \right\rangle\end{aligned}$$

where $N^{[k]} = \frac{N!}{(N-k)!}$, $\pi(n, k)$ is the set of partitions β of $\{1, \dots, n\}$ into k nonempty subsets β_1, \dots, β_k , labelled so that $\min \beta_1 < \dots < \min \beta_k$.

Only the terms involving $k = n, n-1$ contribute in the limit. To see this note that we can choose n different Y_i 's in $N(N-1) \dots (N-n+1) = N^n - \frac{n(n-1)}{2} N^{n-1} + O(N^{n-2})$ ways and

$n - 1$ different Y_i 's in $N(N - 1) \dots (N - n + 2) = N^{n-1} - O(N^{n-2})$ ways. For $k = n - 2$ we can choose k different Y_i 's in $O(N^{n-2})$ ways, etc.

$$\begin{aligned} & NE_\mu[F(X_1^N) - F(\mu)] \\ &= N \left\{ \frac{N^{[n]}}{N^n} \prod_{j=1}^n \langle f_j, \mu P_N \rangle + \frac{N^{[n-1]}}{N^n} \sum_{1 \leq i < j \leq n} \langle f_i f_j, \mu P_N \rangle \prod_{\ell: \ell \neq i, j} \langle f_\ell, \mu P_N \rangle \right. \\ &\quad \left. + O(N^{-2}) - \prod_{j=1}^n \langle f_j, \mu \rangle \right\} \\ &= N \left\{ \left(1 - \frac{n(n-1)}{2N}\right) \prod_{j=1}^n \langle f_j, \mu P_N \rangle \right. \\ &\quad \left. + \frac{1}{N} \sum_{1 \leq i < j \leq n} \langle f_i f_j, \mu P_N \rangle \prod_{\ell \neq i, j} \langle f_\ell, \mu P_N \rangle - \prod_{j=1}^n \langle f_j, \mu \rangle \right\} + O\left(\frac{1}{N}\right) \end{aligned}$$

Note that $\lim_{N \rightarrow \infty} \langle f, \mu P_N \rangle = \langle f, \mu \rangle$. Now let $b_j = \langle f_j, \mu \rangle$ and $a_j = \langle f_j, \mu P_N \rangle$ and recall that

$$(6.8) \quad \lim_{N \rightarrow \infty} N(\langle f, \mu P_N \rangle - \langle f, \mu \rangle) = \langle Af, \mu \rangle$$

Then using this together with the collapsing sum

$$\begin{aligned} & a_1 \dots a_n + (a_1 \dots a_{n-1} b_n - a_1 \dots a_n) + (a_1 \dots a_{n-2} b_{n-1} b_n - a_1 \dots a_{n-1} b_n) \\ &+ (b_1 \dots b_n - a_1 b_2 \dots b_n) - b_1 \dots b_n = 0 \\ a_1 \dots a_n - b_1 \dots b_n &= \sum_k (a_1 \dots a_k b_{k+1} \dots b_n - a_1 \dots a_{k+1} b_{k+2} \dots b_n). \end{aligned}$$

or rewriting

$$(6.9) \quad \prod_{j=1}^n \langle f_j, \mu P_N \rangle = \prod_{j=1}^n [(\langle f_j, \mu \rangle + (\langle f_j, \mu P_N \rangle - \langle f_j, \mu \rangle))]$$

we obtain

$$\begin{aligned} NE_\mu[F(X_1^N) - F(\mu)] &= \sum_{1 \leq i < j \leq n} (\langle f_i f_j, \mu P_N \rangle - \langle f_i, \mu P_N \rangle \langle f_j, \mu P_N \rangle) \prod_{\ell: \ell \neq i, j} \langle f_\ell, \mu P_N \rangle \\ &\quad + \sum_{i=1}^n \langle Af_i, \mu \rangle \prod_{j: j < i} \langle f_j, \mu \rangle \prod_{j: j > i} \langle f_j, \mu P_N \rangle + O(N^{-1}) \\ &= GF(\mu) + o(1) \end{aligned}$$

uniformly in μ .

The completes the verification of condition (6.7). ■

6.4 The Infinitely Many Alleles Model

This is a special case of the Fleming-Viot process which has played a crucial role in modern population biology. It has type space $E = [0, 1]$ and *type-independent* mutation operator with mutation source $\nu_0 \in \mathcal{P}([0, 1])$

$$\begin{aligned} Af(x) &= \theta \left(\int p(x, dy) f(y) - f(x) \right) \\ &= \theta \left(\int f(y) \nu_0(dy) - f(x) \right). \end{aligned}$$

Since A is a bounded operator we can take indicator functions of intervals in $D(A)$. If we have a partition $[0, 1] = \cup_{j=1}^K B_j$ where the B_j are intervals, consider the set $D(G)$ of functions

$$(6.10) \quad F(\mu) = \langle f_1, \mu \rangle \dots \langle f_n, \mu \rangle$$

with $n \geq 1$ and where the functions f_1, \dots, f_n are finite linear combinations of indicator functions of the intervals $\{A_j\}$. Then the function $GF(\mu)$ can be written in the same form and we can prove that the Δ_{K-1} -valued process $\{p_t(A_1), \dots, p_t(A_K)\}$ is a version of the K -allele process with generator

$$\begin{aligned} (6.11) \quad G^K f(p) &= \frac{1}{2} \sum_{i,j=1}^{K-1} p_i (\delta_{ij} - p_j) \frac{\partial^2 f(p)}{\partial p_i \partial p_j} + \theta \sum_{i=1}^{K-1} (\nu_0(A_i) - p_i) \frac{\partial f(p)}{\partial p_i}. \end{aligned}$$

We will next give an explicit construction of this process that allows us to derive a number of interesting properties of this important model.

6.4.1 Projective Limit Construction of the Infinitely Many Alleles Model

Let $\mu, \nu_0 \in \mathcal{P}(E)$, $\mathcal{C} = C_{[0, \infty)}([0, \infty))$. Let U denote the collection of finite partitions $u = (A_1^u, \dots, A_{|u|}^u)$ of E into measurable subsets in $\mathcal{B}(E)$ and $|u|$ denotes the number of sets in the partition u . We place a partial ordering on U as follows:

$$v \succ u$$

if v is a refinement of u . We can also identify partitions with the finite algebras of subsets of E they generate. Given a partition we define the probability measure, P_u on \mathcal{C}^u as the law of the Wright-Fisher diffusion with generator

$$\begin{aligned} G^{(K)} f(p) &= \frac{1}{2} \sum_{i,j=1}^{K-1} p_i (\delta_{ij} - p_j) \frac{\partial^2 f(p)}{\partial p_i \partial p_j} \\ &\quad + \frac{1}{2} \sum_{i=1}^{K-1} \theta(\nu_i - p_i) \frac{\partial f(p)}{\partial p_i} \\ \nu_i &:= \nu_0(A_j) \end{aligned}$$

and initial measure μ , that is, the law of $(p_t(A_1^u), \dots, p_t(A_{|u|}^u))$ (and the additive extension of this to the algebra generated by u).

Remark 6.6 Recall that the associated Markov transition function is determined by the joint moments as follows.

Since the family of functions $p_1^{k_1} \dots p_{K-1}^{k_{K-1}}$ belong to $D(G^{(K)})$ we can apply $G^{(K)}$ and obtain the following system of equations for the joint moments:

$$(6.12) \quad m_{k_1, \dots, k_{K-1}}(t) := E[p_1^{k_1}(t) \dots p_{K-1}^{k_{K-1}}(t)],$$

$$\begin{aligned} \frac{\partial}{\partial t} m_{k_1, \dots, k_{K-1}}(t) &= \frac{1}{2} \sum_i k_i(k_i - 1)m_{k_1, \dots, k_{i-1}, k_{i+1}, \dots, k_{K-1}}(t) \\ &\quad - \frac{1}{2} \sum_{i \neq j} k_i k_j m_{k_1, \dots, k_K}(t) \\ &\quad + \frac{\theta}{2} \sum_{i=1}^{K-1} \nu_i k_i m_{k_1, \dots, k_{j-1}, k_{j+1}, \dots, k_K}(t) \\ &\quad - \frac{\theta}{2} \sum_{i=1}^{K-1} k_i m_{k_1, \dots, k_{K-1}}(t) \end{aligned}$$

Since this system of linear equations is closed, there exists a unique solution which characterizes the K -allele Wright-Fisher diffusion.

In a similar way we can apply this to the function corresponding to the coalescence of two partition elements

$$\begin{aligned} f(p) &= \tilde{f}(\tilde{p}) \\ \tilde{p} &= (\tilde{p}_1, \dots, \tilde{p}_{K-1}) \\ &= (p_1, \dots, p_{\ell-1}, p_{\ell+1}, \dots, p_{k-1}, p_{k+1}, \dots, p_{K-1}, (p_\ell + p_k)) \end{aligned}$$

$$\begin{aligned} G^{(K)} f(p) &= \frac{1}{2} \sum_{i,j=1}^{K-2} \tilde{p}_i (\delta_{ij} - \tilde{p}_j) \frac{\partial^2 f(\tilde{p})}{\partial \tilde{p}_i \partial \tilde{p}_j} \\ &\quad + \frac{\theta}{2} \sum_{i=1}^{K-2} (\tilde{\nu}_i - \tilde{p}_i) \frac{\partial \tilde{f}(\tilde{p})}{\partial \tilde{p}_i} \\ &= G^{(K-1)} \tilde{f}(\tilde{p}) \end{aligned}$$

In other words we have consistency under coalescence of the partition elements. Because of uniqueness this implies that the process $\tilde{p}(t) = (\tilde{p}_1(t), \dots, \tilde{p}_{K-1}(t))$ coincides with the $(K-1)$ -allele Wright-Fisher diffusion.

We denote the canonical projections $\pi_u : \mathcal{C}^{\mathcal{B}(E)} \rightarrow \mathcal{C}^u$ and $\pi_{uv} : \mathcal{C}^v \rightarrow \mathcal{C}^u$ if $v \succ u$ such that $\pi_u = \pi_{uv} \pi_v$, $v \succ u$.

The family $\{P_u\}_{u \in U}$ forms a projective system of probability laws, that is for every pair, (u, v) , $v \succ u$, $\{P_u\}$ then satisfies

$$(6.13) \quad \pi_{uv}(P_v) = P_u, \quad P_u(B) = P_v(\pi_{uv}^{-1}(B)).$$

Therefore, by Theorem 13.6 (in Appendix I) there exists a projective limit measure, that is, a probability measure P_∞ on $\mathcal{C}^{\mathcal{B}([0,1])}$ such that for any $u \in U$, $\pi_u P_\infty = P_u$.

For fixed t (or any finite set of times) we can identify the projective limit,

$$(6.14) \quad \{\tilde{p}_t(A) : A \in \mathcal{B}([0, 1])\}$$

with an element of $\mathcal{X}([0, 1])$, the space of all finitely additive, non-negative, mass one measures on $[0, 1]$, equipped with the projective limit topology, i.e., the weakest topology such that for all Borel subset B of $[0, 1]$, $\mu(B)$ is continuous in μ . Under this topology, $\mathcal{X}([0, 1])$ is Hausdorff. The σ -algebra \mathcal{B} of the space $\mathcal{X}([0, 1])$ is the smallest σ -algebra such that for all Borel subset B of $[0, 1]$, $\mu(B)$ is a measurable function of μ .

For fixed $t \in [0, \infty)$, $\tilde{p}_t(\cdot)$ is a.s. a finitely additive measure, that is, a member of $\mathcal{X}[0, 1]$ and satisfies the conditions of Theorem 13.8 in the Appendices (conditions 1,2 follow immediately from the construction, 3 follows since for any $A \in \mathcal{B}([0, 1])$ $E(p_t(A)) \leq \max(\mu(A), \nu_0(A))$ and (4) is automatic since all measures are bounded by 1). Therefore for fixed t this determines a unique countably additive version $p_t(\cdot)$, that is, a random countable additive measure $p_t \in \mathcal{P}([0, 1])$ a.s. Similarly, taking two times t_1, t_2 we obtain a the joint distribution of a pair (p_{t_1}, p_{t_2}) of random probability measures. We can then verify that $t \rightarrow \int f(x)p_t(dx)$ is a.s. continuous for countable convergence determining class of functions so that there is an a.s. continuous version with respect to the topology of weak convergence.

Remark 6.7 *We can carry out the same construction assuming that for each $u \in U$ the Wright-Fisher diffusion starts with the stationary Dirichlet measure and obtain by the projective limit a probability measure on $\mathcal{P}(E)$ which for any partition has the associated Dirichlet distribution.*

6.5 The Jirina Branching Process

In 1964 Jirina [337] gave the first construction of a measure-valued branching process. The state space is the space of finite measures on $[0, 1]$, $M_f([0, 1])$, $\nu_0 \in M_1([0, 1])$. We will construct a version of this process with immigration by a projective limit construction.

Given a partition (A_1, \dots, A_K) of $[0, 1]$ let $\{X_t(A_i) : t \geq 0, i = 1, \dots, K\}$ satisfy the SDE (Feller CSB plus immigration):

$$(6.15) \quad \begin{aligned} dX_t(A_i) &= c(\nu_0(A_i) - X_t(A_i))dt + \sqrt{2\gamma X_t(A_i)}dW_t^{A_i} \\ X_0(A_i) &= \mu(A_i) \end{aligned}$$

where ν_0 is in $\mathcal{P}([0, 1])$ and for each i , $W_t^{A_i}$ is a standard Brownian motion and for $i \neq j$ $W_t^{A_i}$ and $W_t^{A_j}$ are independent.

We can then verify that the processes $X_t(A_i) : i = 1, \dots, K$ are independent and as $t \rightarrow \infty$, $X_t(A_i)$ converges in distribution to a stationary measure $X_\infty(A_i)$ with density which satisfies

$$f_i(x) = \frac{1}{Z}x^{\theta_i-1}e^{-\theta x}, \quad x > 0$$

where $\theta = \frac{c}{\gamma}$, $\theta_i = \theta\nu_0(A_i)$.

This can be represented by $X_\infty(A) = \theta^{-1}G(\theta\nu_0(A))$ where $\theta = \frac{c}{\gamma}$ and

$$\begin{aligned} \mathcal{L}\{(X_\infty(A_1), \dots, X_\infty(A_K))\} &= \\ \mathcal{L}\left\{\frac{1}{\theta}[G(\theta_1), G(\theta_1 + \theta_2) - G(\theta_1), \dots, G(\theta) - G(\theta - \theta_K)]\right\} \end{aligned}$$

where $G(s)$ is the *Moran subordinator* - see subsection 6.6.1 below.

For $u = (A_1^u, \dots, A_{|u|}^u) \in U$ (defined as in the last subsection) let $\{P_u = \mathcal{L}(\{(X_t(A_1), \dots, X_t(A_{|u|})) : t \geq 0, A \in u\})\}$. Then the collection $\{P_u\}_{u \in U}$ forms a projective system and as in the previous

section there exists a projective limit measure P_∞ on $(C_{[0,\infty)}([0,\infty)))^{\mathcal{B}([0,1])}$. Moreover for fixed $t \in [0,\infty)$, $X_t(\cdot)$ is a.s. a finitely additive measure that is regular (on a countable generating subset of $\mathcal{B}([0,1])$) we obtain a unique countably additive version (recall Theorem 13.8). Thus, $\{X_t(\cdot) : t \geq 0\}$ is a measure-valued process and again we can obtain an a.s. continuous $M_F([0,1])$ -valued version. This $M_F([0,1])$ -valued process is called the *Jirina process*.

Corollary 6.8 *The stationary measure for the Jirina process is given by the random measure*

$$(6.16) \quad X_\infty(A) = \frac{1}{\theta} \int_0^1 1_A(x) dG(\theta s), \quad A \in \mathcal{B}([0,1])$$

where $G(\cdot)$ is the Moran gamma subordinator.

6.6 Invariant Measures of the IMA and Jirina Processes

6.6.1 The Moran (Gamma) Subordinator

We begin by recalling the the *Gamma distribution* with parameter $\alpha > 0$ given by the density function

$$g_\alpha(u) = u^{\alpha-1} e^{-u} / \Gamma(\alpha)$$

and Laplace transform of g_α is

$$\int_0^\infty g_\alpha(y) e^{-\lambda y} dy = \frac{1}{(1+\lambda)^\alpha}, \quad \lambda > -1,$$

The Moran subordinator $\{G(\alpha) : \alpha \geq 0\}$ is an increasing process with stationary independent increments $G(\alpha_2) - G(\alpha_1)$, $\alpha_1 < \alpha_2$ given by $g_{\alpha_2 - \alpha_1}$.

Lévy representation

Lemma 6.9

$$(6.17) \quad E\left(e^{-\lambda G(\alpha)}\right) = \exp\left(-\alpha \int_0^\infty (1 - e^{-u\lambda}) \frac{e^{-u}}{u} du\right).$$

Proof. Note that

$$\begin{aligned} \frac{\partial}{\partial \lambda} \int_0^\infty (1 - e^{-\lambda z}) z^{-1} e^{-z} dz &= \int_0^\infty (e^{-\lambda z}) e^{-z} dz = \frac{1}{1+\lambda} \\ \int_0^\infty (1 - e^{-\lambda z}) z^{-1} e^{-z} dz &= \log(1 + \lambda) \end{aligned}$$

Hence we have the Lévy-Khintchin representation with Lévy measure $\frac{e^{-z}}{z}$, $z > 0$

$$(6.18) \quad \frac{1}{(1+\lambda)^\alpha} = \exp\left\{-\alpha \int_0^\infty (1 - e^{-\lambda z}) z^{-1} e^{-z} dz\right\}.$$

■

Poisson representation

The *Poisson random field* with intensity measure μ is a random counting measure Π on a space S . $\Pi(A_i), \Pi(A_j)$ are independent if $i \neq j$ and $\Pi(A)$ is Poisson with parameter $\mu(A)$.

Theorem 6.10 (*Campbell's Theorem.*) *Let Π be a Poisson random field with intensity $\mu \in M(S)$ and $f : S \rightarrow \mathbb{R}$, $\Sigma = \sum_{x \in \Pi} f(x) = \int f(x)\Pi(dx)$ converges a.s. if and only if*

$$\int_S \min(|f(x)|, 1)\mu(dx) < \infty$$

and then

$$E(e^{s \int f(x)\Pi(dx)}) = \exp(\int (e^{sf(x)} - 1)\mu(dx)), \quad s \in R$$

provided the integral on the right exists.

Now consider the Poisson random measure on $[0, 1] \times (0, \infty)$

$$(6.19) \quad \Xi_\theta = \sum \delta_{\{x, u\}}$$

with intensity measure

$$\theta \nu_0(dx) \frac{e^{-u}}{u} du.$$

Let $\tilde{X}_\infty(A) := \int_A \int_0^\infty u \Xi_\theta(dx, du)$. Then by Campbell's Theorem

$$(6.20) \quad \begin{aligned} E(e^{-\lambda \tilde{X}_\infty(A)}) &= E(e^{-\lambda \int_A \int_0^\infty u \Xi_\theta(dx, du)}) \\ &= e^{-\theta \nu_0(A) \int (1 - e^{-\lambda u}) \frac{e^{-u}}{u} du}. \end{aligned}$$

Hence we can represent equilibrium of the Jirina process by the random measure with Poisson representation $\{X_\infty(A) : A \in \mathcal{B}([0, 1])\}$ by

$$(6.21) \quad X_\infty(A) = \theta^{-1} \int_A \int_0^\infty u \Xi_\theta(dx, du)$$

and this can be obtained as the projective limit of the finite systems.

If ν_0 is Lebesgue measure on $[0, 1]$ then have that the $\{X_\infty([0, s])\}_{0 \leq s \leq 1} = \{G(s)\}_{0 \leq s \leq 1}$ where $G(s)$ is the Moran subordinator with increments $G(s_2) - G(s_1)$ having the Gamma $\theta(s_2 - s_1)$ distribution $\theta = \frac{c}{\gamma}$.

6.6.2 Representation of the Infinitely Many Alleles Equilibrium

Recall (Theorem 5.9) that the Dirichlet distribution $\text{Dirichlet}(\theta_1, \dots, \theta_n)$ has the joint density on relative to $(n-1)$ -dimensional Lebesgue measure on Δ_{n-1} given by

$$f(p_1, \dots, p_{n-1}) = \frac{\Gamma(\theta_1 + \dots + \theta_n)}{\Gamma(\theta_1) \dots \Gamma(\theta_n)} p_1^{\theta_1-1} p_2^{\theta_2-1} \dots p_n^{\theta_n-1}.$$

Recall that if the θ are large the measure concentrates away from the boundary whereas if the θ are small things concentrate near the boundary corresponding to highly disparate p with a few large p_j and the others small. For example if the θ_j are small but equal there is a high probability that at least one of the p_j is much greater than average; and which value or values of j have large p_j is a matter of chance.

Proposition 6.11 Let X_∞ denote the equilibrium random measure for the Jirina process and consider a partition $[0, 1] = \cup_{i=1}^n A_i$ and define

$$(6.22) \quad Y(A_i) := \frac{X_\infty(A_i)}{X_\infty([0, 1])} = \frac{G(\theta|A_i|)}{G(\theta)}.$$

Then the family $(Y(A_1), \dots, Y(A_K))$ is **independent** of $X_\infty([0, 1])$ and has as distribution the Dirichlet($\theta_1, \dots, \theta_K$) where $\theta_j = \theta \nu_0(A_j)$.

Proof. Let Y be Gamma(θ) and (P_1, \dots, P_K) Dirichlet($\theta_1, \dots, \theta_K$) with Y and (P_1, \dots, P_K) independent, and define (Y_1, \dots, Y_K) by

$$(6.23) \quad Y_i := Y P_i.$$

We will verify that (Y_1, \dots, Y_K) has the joint probability density function

$$(6.24) \quad g(y_1, \dots, y_K) = \prod_{i=1}^K u_i^{\theta_i-1} e^{-u_i} / \Gamma(\theta_i).$$

Consider the 1-1 transformation $(Y_1, Y_2, \dots, Y_K) \leftrightarrow (Y, P_2, \dots, P_K)$ with Jacobian

$$(6.25) \quad |J| = \left\{ \left| \frac{\partial x_1, \dots, x_K}{\partial y_1, \dots, y_K} \right|, \quad x_1 = y, x_2 = p_2, \dots, x_K = p_K \right\} = \frac{1}{y^{K-1}}.$$

By independence of Y and (P_1, \dots, P_K) , we obtain the joint density of (Y_1, \dots, Y_K) as

$$\begin{aligned} g(y_1, \dots, y_K) &= f(p_1, \dots, p_K|Y) f_Y(y) |J| \\ &= f(p_1, \dots, p_K) \frac{1}{\Gamma(\theta)} y^{\theta-1} e^{-y} \frac{1}{y^{K-1}} \\ &= \frac{\Gamma(\theta)}{\Gamma(\theta_1) \dots \Gamma(\theta_K)} \\ &\quad \cdot \left(\frac{y_1}{\sum y_i} \right)^{\theta_1-1} \dots \left(\frac{y_K}{\sum y_i} \right)^{\theta_K-1} \frac{1}{\Gamma(\theta)} (\sum y_i)^{(\theta-1)} e^{-\sum y_i} (\sum y_i)^{-(K-1)} \\ &= \prod_{i=1}^K \frac{1}{\Gamma(\theta_i)} y_i^{\theta_i-1} e^{-y_i} \end{aligned}$$

Note that this coincides with the Dirichlet($\theta_1, \dots, \theta_K$) distribution. ■

Corollary 6.12 The invariant measure of the infinitely many alleles model can be represented by the random probability measure

$$(6.26) \quad Y(A) = \frac{X_\infty(A)}{X_\infty([0, 1])}, \quad , A \in \mathcal{B}([0, 1]).$$

where $X_\infty(\cdot)$ is the equilibrium of the above Jirina process and $Y(\cdot)$ and $X_\infty([0, 1])$ are independent.

Reversibility

Recall that the Dirichlet distribution is a *reversible* stationary measure for the K -*type* Wright-Fisher model with house of cards mutation (Theorem 5.9). From this and the projective limit construction it can be verified that $\mathcal{L}(Y(\cdot))$ is a reversible stationary measure for the infinitely many alleles process. Note that reversibility actually characterizes the IMA model among neutral Fleming-Viot processes with mutation, that is, any mutation mechanism other than the “type-independent” or “house of cards” mutation leads to a stationary measure that is not reversible (see Li-Shiga-Yau (1999) [405]).

6.6.3 The Poisson-Dirichlet Distribution

Without loss of generality we can assume that ν_0 is Lebesgue measure on $[0, 1]$. This implies that the IMA equilibrium is given by a random probability measure which is pure atomic

$$(6.27) \quad p_\infty = \sum_{i=1}^{\infty} a_i \delta_{x_i}, \quad \sum_{i=1}^{\infty} a_i = 1, \quad x_i \in [0, 1]$$

in which the $\{x_i\}$ are i.i.d. $U([0, 1])$ and the atom sizes $\{a_i\}$ correspond to the normalized jumps of the Moran subordinator. Let (ξ_1, ξ_2, \dots) denote the reordering of the atom sizes $\{a_i\}$ in decreasing order.

The Poisson-Dirichlet $PD(\theta)$ distribution is defined to be the distribution of the infinite sequence $\xi = (\xi_1, \xi_2, \dots)$ which satisfies

$$\xi_1 \geq \xi_2 \geq \dots, \quad \sum_k \xi_k = 1.$$

This sequence is given by

$$(6.28) \quad \xi_k = \frac{\eta_k(\theta)}{G(\theta)} = \frac{\eta_k}{\sum \eta_\ell}, \quad k = 1, 2, \dots,$$

where $\eta_k = \eta_k(\theta)$ is the height of the k th largest jump in $[0, \theta]$ of the Moran process (subordinator), G and $G(\theta) = \sum_{\ell=1}^{\infty} \eta_\ell$.

Properties of the Poisson-Dirichlet Distribution

Recalling (6.19), (6.21) we note that the set of heights of the jumps of $G(\cdot)$ in $[0, \theta]$ form a Poisson random field Π_θ on $(0, \infty)$ with intensity measure

$$\theta \frac{e^{-u}}{u} du.$$

We can then give a direct description of $PD(\theta)$ in terms of such a Poisson random field. If $\eta_1 \geq \eta_2 \geq \eta_3 \geq \dots$ are the points of such a random field ordered by size then

$$\xi_k = \frac{\eta_k}{\sum_{\ell=1}^{\infty} \eta_\ell}$$

defines a sequence ξ having the distribution $PD(\theta)$.

By the law of large numbers (for the Poisson) we get

$$\lim_{t \rightarrow 0} \frac{\#\{k : \eta_k > t\}}{L(t)} = 1$$

with probability one where

$$L(t) = \int_t^\infty \theta \frac{e^{-u}}{u} du \sim -\theta \log t.$$

Thus

$$\#\{k : \eta_k > t\} \sim -\theta \log t$$

$$\eta_{\theta \log \frac{1}{t}} \approx t \quad \text{as } t \rightarrow 0.$$

$$\eta_k \approx e^{-k/\theta}$$

Thus ξ_k decays exponentially fast

$$-\log \xi_k \sim \frac{k}{\theta} \quad \text{as } k \rightarrow \infty.$$

The Distribution of Atom Sizes

We now introduce the random measure on $(0, \infty)$,

$$Z_\theta((a, b)) = \frac{\Xi_\theta([0, 1] \times (a, b))}{G(\theta)}$$

$$\int_0^\infty u Z_\theta(du) = 1.$$

This is the *distribution of normalized atom sizes* and this just depends on the normalized ordered atoms and hence is independent of $X_\infty([0, 1])$. Intuitively, as $\theta \rightarrow \infty$, $Z_\theta(\theta du)$ converges in some sense to

$$\frac{e^{-u}}{u} du.$$

To give a precise formulation of this we first note that

$$\int_0^\infty u^k \left(\frac{e^{-u}}{u} \right) du = \Gamma(k) = (k-1)!$$

Then one can show (see Griffiths (1979), [267]) that

$$(6.29) \quad \lim_{\theta \rightarrow \infty} \theta^{k-1} \int_0^\infty u^k Z_\theta(du) = (k-1)!$$

and there is an associated CLT

$$(6.30) \quad \sqrt{\theta} \frac{\theta^{k-1} \int u^k Z_\theta(du) - (k-1)!}{(k-1)!} \Rightarrow N(0, \sigma_k^2),$$

with $\sigma_k^2 = \frac{(2k-1)! - (k!)^2}{((k-1)!)^2}$ Joyce, Krone and Kurtz (2002) [338]. Also see Dawson and Feng (2006) [147] for the related large deviation behaviour.

6.6.4 The GEM Representation

Without loss of generality we can assume $\nu_0 = U[0, 1]$. Consider a partition of $[0, 1]$ into K intervals of equal length. Then the random probability

$$\vec{p}_K = (p_1, \dots, p_K)$$

has the symmetric Dirichlet distribution $D(\alpha, \dots, \alpha)$ with $\alpha = \frac{\theta}{K}$.

Randomized Ordering via Size-biased sampling

Let \mathcal{N} be a random variable having values in $\{1, 2, \dots, K\}$ in such a way that

$$P(\mathcal{N} = k | \vec{p}_K) = p_k, \quad (1 \leq k \leq K)$$

Then a standard calculation shows that the vector

$$\vec{p}' = (p_{\mathcal{N}}, p_1, \dots, p_{\mathcal{N}-1}, p_{\mathcal{N}+1}, \dots, p_K)$$

has distribution (cf. (5.55))

$$D(\alpha + 1, \alpha, \dots, \alpha)$$

It follows that $(p_{\mathcal{N}}, 1 - p_{\mathcal{N}})$ has the Dirichlet distribution (Beta distribution)

$$D(\alpha + 1, (K - 1)\alpha)$$

so that it has probability density function

$$\frac{\Gamma(K\alpha + 1)}{\Gamma(\alpha + 1)\Gamma(K\alpha - \alpha)} p^{\alpha} (1 - p)^{(K-1)\alpha - 1}.$$

Given $v_1 = p_{\mathcal{N}}$, the conditional distribution of the remaining components of \vec{p} is the same as that of $(1 - p_{\mathcal{N}})p^{(1)}$, where the $(K - 1)$ -vector $\vec{p}^{(1)}$ has the symmetric distribution $D(\alpha, \dots, \alpha)$.

We say that $p_{\mathcal{N}}$ is obtained from \vec{p} by size-biased sampling. This process may now be applied to $\vec{p}^{(1)}$ to produce a component, v_2 with distribution

$$\frac{\Gamma((K - 1)\alpha + 1)}{\Gamma(\alpha + 1)\Gamma((K - 1)\alpha - \alpha)} p^{\alpha} (1 - p)^{(K-2)\alpha - 1}$$

and a $(K - 2)$ vector $\vec{p}^{(2)}$ with distribution $D(\alpha, \dots, \alpha)$. This is an example of Kolmogorov's *stick breaking* process.

Theorem 6.13 (a) As $K \rightarrow \infty$, with $K\alpha = \theta$ constant, the distribution of the vector $\vec{q}_K = (q_1, q_2, \dots, q_K)$ converges weakly to the GEM distribution with parameter θ , that is the distribution of the random probability vector $\vec{q} = (q_1, q_2, \dots)$ where

$$q_1 = v_1, \quad q_2 = (1 - v_1)v_2, \quad q_3 = (1 - v_1)(1 - v_2)v_3, \quad \dots$$

with $\{v_k\}$ are i.i.d. with Beta density ($\text{Beta}(1, \theta)$)

$$\theta(1 - p)^{\theta - 1}, \quad 0 \leq p \leq 1$$

(b) If $\vec{q} = (q_1, q_2, \dots)$ is reordered (by size) as $\vec{p} = (p_1, p_2, \dots)$, that is i.e. p_k is the k th largest of the $\{q_j\}$, then \vec{p} has the Poisson-Dirichlet distribution, $\text{PD}(\theta)$.

Proof. (a) Let $(p_1^K, \dots, p_k^K) \in \Delta_{K-1}$ be a random probability vector obtained by decreasing size reordering of a probability vector sampled from the distribution $D(\alpha, \dots, \alpha)$ with $\alpha = \frac{\theta}{K}$. Then let (q_1^K, \dots, q_k^K) be the size-biased reordering of (p_1^K, \dots, p_k^K) . Then as shown above we can rewrite this as

$$(6.31) \quad q_1^K = v_1^K, \quad q_2^K = (1 - v_1^K)v_2^K, \quad q_3^K = (1 - v_1^K)(1 - v_2^K)v_3^K, \quad \dots$$

where v_1^K, \dots, v_{K-1}^K are independent and v_r^K has pdf

$$(6.32) \quad \frac{\Gamma((K-r)\alpha+1)}{\Gamma(\alpha+1)\Gamma((K-r)\alpha-\alpha)} u^\alpha (1-u)^{(K-r-1)\alpha-1}, \quad 0 \leq u \leq 1.$$

Now let $K \rightarrow \infty$ with $K\alpha = \theta$. Then

$$\frac{\Gamma(K\alpha+1)}{\Gamma(\alpha+1)\Gamma(K\alpha-\alpha)} p^\alpha (1-p)^{(K-1)\alpha-1} \rightarrow \theta(1-p)^{\theta-1}.$$

and

$$\begin{aligned} & \frac{\Gamma((K-r)\alpha+1)}{\Gamma(\alpha+1)\Gamma((K-r)\alpha-\alpha)} p^\alpha (1-p)^{(K-r-1)\alpha-1} \\ &= \frac{\Gamma((K-r)\theta/K+1)}{\Gamma(\theta/K+1)\Gamma((K-r)\theta/K-\theta/K)} p^{\theta/K} (1-p)^{(K-r-1)\theta/K-1} \\ &\rightarrow \theta(1-p)^\theta \end{aligned}$$

Thus the distributions of the first m components of the vector \vec{q}_K converge weakly to the distribution of the first m components of the random (infinite) probability vector \vec{q} defined by

$$q_1 = v_1, q_2 = (1-v_1)v_2, q_3 = (1-v_1)(1-v_2)v_3, \dots$$

where $\{v_k\}$ are i.i.d. with Beta density ($\text{Beta}(1, \theta)$)

$$\theta(1-p)^{\theta-1}, \quad 0 \leq p \leq 1.$$

(b) By the projective limit construction, the $\text{PD}(\theta)$ distribution arises as the limit in distribution of the ordered probability vectors (p_1^K, \dots, p_k^K) . Then the size-biased reorderings converge in distribution to the size-biased reordering q_1, q_2, \dots of the probability vector p_1, p_2, p_3, \dots . Clearly the decreasing-size reordering of q_1, q_2, \dots reproduces p_1, p_2, \dots . ■

Remark 6.14 *The distribution of sizes of the age ordered alleles in the infinitely many alleles model is given by the GEM distribution. The intuitive idea is as follows. By exchangeability at the individual level the probability that the k th allele at a given time survives the longest (time to extinction) among those present at that time is proportional to p_k the frequency of that allele in the population at that time. Observing that the ordered survival times correspond to ages under time reversal, the result follows from reversibility. See Ethier [215] for justification of this argument and a second proof of the result.*

Remark 6.15 *There is a two parameter analogue of the Poisson-Dirichlet introduced by Perman, Pitman and Yor (1992) [482] that shares some features with the PD distribution. See Feng (2009) [234] for a recent detailed exposition.*

6.6.5 Application of the Poisson-Dirichlet distribution: The Ewens Sampling Formula

In analyzing population genetics data under the neutral hypothesis it is important to know the probabilities of the distribution of types obtained in taking a random sample of size n . For example, this is used to test for neutral mutation in a population.

Consider a random sample of size n chosen from the random vector $\xi = (\xi_1, \xi_2, \dots)$ chosen from the distribution $\text{PD}(\theta)$.

We first compute the probability that they are all of the same type. Conditioned on ξ this is $\sum_{k=1}^{\infty} \xi_k^n$ and hence the unconditional probability is

$$h_n = E \left\{ \sum_{k=1}^{\infty} \xi_k^n \right\}$$

Using Campbell's formula we get

$$\begin{aligned} & E \left\{ \sum_{k=1}^{\infty} \eta_k^n \right\} \\ &= \frac{d}{ds} \left(E(e^{\int_0^1 \int_0^{\infty} (e^{sz^n} - 1) \frac{\theta e^{-z}}{z} dz} \right) |_{s=0} \\ &= \int z^n \frac{\theta e^{-z}}{z} dz = \theta(n-1)! \end{aligned}$$

Also

$$E \left(\left(\sum_{k=1}^{\infty} \eta_k \right)^n \right) = \frac{\Gamma(n+\theta)}{\Gamma(\theta)}.$$

By the Gamma representation (6.28) for the ordered jumps of the Gamma subordinator we get

$$\begin{aligned} E \left\{ \sum_{k=1}^{\infty} \eta_k^n \right\} &= E \left\{ \left(\sum_{k=1}^{\infty} \eta_k \right)^n \sum_{k=1}^{\infty} \xi_k^n \right\} \\ &= E \left(\left(\sum_{k=1}^{\infty} \eta_k \right)^n \right) E \left\{ \sum_{k=1}^{\infty} \xi_k^n \right\} \text{ by independence} \end{aligned}$$

Therefore

$$h_n = \frac{\theta \Gamma(\theta)(n-1)!}{\Gamma(n+\theta)} = \frac{(n-1)!}{(1+\theta)(2+\theta)\dots(n+\theta-1)}$$

In general in a sample of size n let

$$\begin{aligned} a_1 &= \text{number of types with 1 representative} \\ a_2 &= \text{number of types with 2 representatives} \\ &\dots \\ a_n &= \text{number of types with } n \text{ representatives} \end{aligned}$$

Of course,

$$a_i \geq 0, \quad a_1 + 2a_2 + \dots + na_n = n.$$

We can also think of this as a partition

$$\alpha = 1^{a_1} 2^{a_2} \dots n^{a_n}$$

Let $P_n(\alpha)$ denote the probability that the sample exhibits the partition α . Note that $P_n(n^1) = h_n$.

Proposition 6.16 (*Ewens sampling formula*)

$$(6.33) \quad P_n(\mathbf{a}) = P_n(a_1, \dots, a_n) = \frac{n! \Gamma(\theta)}{\Gamma(n + \theta)} \prod_{j=1}^n \left(\frac{\theta^{a_j}}{j^{a_j} a_j!} \right).$$

Proof. We can select the partition of $\{1, \dots, n\}$ into subsets of sizes (a_1, \dots, a_n) as follows. Consider a_1 boxes of size 1, … a_n boxes of size n . Then the number of ways we can distribute $\{1, \dots, n\}$ is $n!$ but we can reorder the a_i boxes in $a_i!$ ways and there are $(j!)$ permutations of the indices in each of the a_j partition elements with j elements. Hence the total number of ways we can do the partitioning to $\{1, \dots, n\}$ is $\frac{n!}{\prod(j!)^{a_j} a_j!}$.

Now condition on the vector $\xi = (\xi(1), \xi(2), \dots)$. The probability that we select the types (ordered by their frequencies is then given by)

$$P_n(\mathbf{a}|\xi) = \frac{n!}{\prod(j!)^{a_j} a_j!} \sum_{I_{a_1, \dots, a_n}} \xi(k_{11}) \xi(k_{12}) \dots \xi(k_{1a_1}) \xi(k_{21})^2 \dots \xi(k_{2a_2})^2 \xi(k_{31})^3 \dots$$

where the summation is over

$$I_{a_1, \dots, a_n} := \{k_{ij} : i = 1, 2, \dots; j = 1, 2, \dots, a_i\}$$

Hence using the Gamma representation we get

$$\begin{aligned} & \frac{\Gamma(n + \theta)}{\Gamma(\theta)} P_n(\mathbf{a}) \\ &= \frac{n!}{\prod(j!)^{a_j} a_j!} E \left\{ \sum \eta(k_{11}) \eta(k_{12}) \dots \eta(k_{1a_1}) \eta(k_{21})^2 \dots \eta(k_{2a_2})^2 \eta(k_{31})^3 \dots \right\} \end{aligned}$$

But

$$\begin{aligned} & E \left\{ \sum_{i,j} \eta(k_{11}) \eta(k_{12}) \dots \eta(k_{1a_1}) \eta(k_{21})^2 \dots \eta(k_{2a_2})^2 \eta(k_{31})^3 \dots \right\} \\ &= \prod_{j=1}^n E \left\{ \sum_{k=1}^{\infty} \eta(k)^j \right\}^{a_j} = \prod_{j=1}^n \left\{ \int_0^{\infty} z^j \theta \frac{e^{-z}}{z} dz \right\}^{a_j} \text{ by Campbell's thm} \\ &= \prod_{j=1}^n \{\theta(j-1)!\}^{a_j} \end{aligned}$$

Therefore substituting we get

$$P_n(\mathbf{a}) = \frac{n! \Gamma(\theta)}{\Gamma(n + \theta)} \prod_{j=1}^n \left(\frac{\theta^{a_j}}{j^{a_j} a_j!} \right).$$

■

Chapter 7

Martingale Problems and Dual Representations

7.1 Introduction

We have introduced above the basic mechanisms of branching, resampling and mutation mainly using generating functions and semigroup methods. However these methods have limitations and in order to work with a wider class of mechanisms we will introduce some additional tools of stochastic analysis in this chapter. The martingale method which we use has proved to be a natural framework for studying a wide range of problems including those of population systems. The general framework is as follows:

- the object is to specify a Markov process on a Polish space E in terms of its probability laws $\{P_x\}_{x \in E}$ where P_x is a probability measure on $C_E([0, \infty))$ or $D_E([0, \infty))$ satisfying $P_x(X(0) = x) = 1$.
- the probabilities $\{P_x\} \in \mathcal{P}(C_E([0, \infty)))$ satisfy a *martingale problem* (MP). One class of martingale problems is defined by the set of conditions of the form

$$(7.1) \quad F(X(t)) - \int_0^t GF(X(s))ds, \quad F \in \mathcal{D} \quad (\mathcal{D}, G) - \text{martingale problem}$$

is a P_x martingale where G is a linear map from \mathcal{D} to $C(E)$, and $\mathcal{D} \subset C(E)$ is measure-determining.

- the martingale problem MP has one and only one solution.

Two martingale problems MP_1, MP_2 are said to be *equivalent* if a solution to MP_1 problem is a solution to MP_2 and vice versa.

In our setting the existence of a solution is often obtained as the limit of a sequence of probability laws of approximating processes. The question of uniqueness is often the more challenging part. We introduce the method of dual representation which can be used to establish uniqueness for a number of basic population processes. However the method of duality is applicable only for special classes of models. We introduce a second method, the Cameron-Martin-Girsanov type change of measure which is applicable to some basic problems of stochastic population systems. Beyond the domain of applicability of these methods, things are much more challenging. Some

recent progress has been made in a series of papers of Athreya, Barlow, Bass, Perkins [10], Bass-Perkins [29], [31] but open problems remain.

We begin by reformulating the Jirina and neutral IMA Fleming-Viot in the martingale problem setting. We then develop the Girsanov and duality methods in the framework of measure-valued processes and apply them to the Fleming-Viot process with selection.

7.2 The Jirina martingale problem

By our projective limit construction of the Jirina process (with $\nu_0 = \text{Lebesgue}$), we have a probability space $(\Omega, \mathcal{F}, \{X^\infty : [0, \infty) \times \mathcal{B}([0, 1]) \rightarrow [0, \infty)\}, P)$ such that a.s. $t \rightarrow X_t^\infty(A)$ is continuous and $A \rightarrow X^\infty(A)$ is finitely additive. We can take a modification, X , of X^∞ such that a.s. $X : [0, \infty) \rightarrow M_F([0, 1])$ is continuous where $M_F([0, 1])$ is the space of (countably additive) finite measures on $[0, 1]$ with the weak topology. We then define the filtration

$$\mathcal{F}_t : \sigma\{X_s(A) : 0 \leq s \leq t, A \in \mathcal{B}([0, 1])\}$$

and \mathcal{P} , the σ -field of predictable sets in $\mathbb{R}_+ \times \Omega$ (ie the σ -algebra generated by the class of \mathcal{F}_t -adapted, left continuous processes).

Recall that for a fixed set A the Feller CSB with immigration satisfies

$$(7.2) \quad X_t(A) - X_0(A) - \int_0^t c(\nu_0(A) - X_s(A))ds = \int_0^t \sqrt{2\gamma X_t(A)} dw_t^A$$

which is an L^2 -martingale.

Moreover, by polarization

$$(7.3) \quad \langle M(A_1), M(A_2) \rangle_t = \gamma \int_0^t X_s(A_1 \cap A_2) ds$$

and if $A_1 \cap A_2 = \emptyset$, then the martingales $M(A_1)_t$ and $M(A_2)_t$ are orthogonal. This is an example of an *orthogonal martingale measure*.

Therefore for any Borel set A

$$(7.4) \quad M_t(A) := X_t(A) - X_0(A) - \int_0^t c(\nu_0(A) - X_s(A))ds$$

is a martingale with increasing process

$$(7.5) \quad \langle M(A) \rangle_t = \gamma \int_0^t X_s(A) ds.$$

We note that we can define integrals with respect to an orthogonal martingale measure (see next subsection) and show that (letting $X_t(f) = \int f(x)X_t(dx)$ for $f \in \mathcal{B}([0, 1])$)

$$(7.6) \quad M_t(f) := X_t(f) - X_0(f) - \int_0^t c(\nu_0(f) - X_s(f))ds = \int f(x)M_t(dx)$$

which is a martingale with increasing process

$$(7.7) \quad \langle M(f) \rangle_t = \gamma \int_0^t f^2(x)X_s(dx) ds.$$

This suggests the martingale problem for the Jirina process which we state in subsection 7.2.2.

7.2.1 Stochastic Integrals wrt Martingale Measures

A general approach to martingale measures and stochastic integrals with respect to martingale measures was developed by Walsh [562]. We briefly review some basic results.

Consider the collection of \mathcal{S} simple functions ψ of the form

$$(7.8) \quad \psi(t, \omega, x) = \sum_{i=1}^K \psi_{i-1}(\omega) \phi_i(x) 1_{(t_{i-1}, t_i]}(t)$$

for some $\phi_i \in b\mathcal{B}(E)$, $\psi \in b\mathcal{F}_{t_{i-1}}$, $0 = t_0 < t_1 \dots < t_K \leq \infty$. The predictable σ -field $\mathcal{P}r$ on $\mathbb{R}_+ \times \Omega \times E$ is the σ -field generated by the class of simple functions of the form (7.8).

For $\psi \in \mathcal{S}$, define

$$M_t(\psi) := \int_0^t \int \psi(s, x) dM(s, x) = \sum_{i=1}^K \psi_{i-1}(M_{t \wedge t_i}(\phi_i) - M_{t \wedge t_{i-1}}(\phi_i))$$

Then $M_t(\psi) \in \mathcal{M}_{loc}$ (the space of \mathcal{F}_t local martingales) and

$$\langle M(\psi)_t \rangle = \int_0^t X_s(\gamma \psi_s^2) ds.$$

Let

$$\mathcal{L}_{loc}^2 = \left\{ \psi : \mathbb{R}_+ \times \Omega \times E \rightarrow \mathbb{R} : \psi \text{ is } \mathcal{P}r\text{-measurable, } \int_0^t X_s(\psi_s^2) ds < \infty, \forall t > 0 \right\}$$

Lemma 7.1 *For any $\psi \in \mathcal{L}_{loc}^2$ there is a sequence $\{\psi_n\}$ in \mathcal{S} such that*

$$P \left(\int_0^n \int (\psi_n - \psi)^2(s, \omega, x) \gamma(x) X_s(dx) ds > 2^{-n} \right) < 2^{-n}.$$

Proof. Let $\bar{\mathcal{S}}$ denote the set of bounded $\mathcal{P}r$ -measurable functions which can be approximated as above. $\bar{\mathcal{S}}$ is closed under \rightarrow^{bp} . Using $\mathcal{H}_0 = \{f_{i-1}(\omega) \phi_i(x), \phi \in b\mathcal{E}, f_{i-1} \in b\mathcal{F}_{t_{i-1}}, \phi_i \in b\mathcal{E}\}$, we see that $\psi(t, \omega, x) = \sum_{i=1}^K \psi_{i-1}(\omega, x) 1_{(t_{i-1}, t_i]}(t)$ is in $\bar{\mathcal{S}}$ for any $\psi_{i-1} \in b(\mathcal{F}_{t_{i-1}} \times \mathcal{E})$. If $\psi \in b(\mathcal{P}r)$, then

$$\psi_n(s, \omega, x) = 2^n \int_{(i-1)2^{-n}}^{i2^{-n}} \psi(r, \omega, x) dr \text{ is } s \in (i2^{-n}, (i+1)2^{-n}], i = 1, 2, \dots$$

satisfies $\psi_n \in \bar{\mathcal{S}}$ by the above. For each (ω, x) , $\psi_n(s, \omega, x) \rightarrow \psi(s, \omega, x)$ for Lebesgue a.a. s by Lebesgue's differentiation theorem and it follows easily that $\psi \in \bar{\mathcal{S}}$. Finally if $\psi \in \mathcal{L}_{loc}^2$, the obvious truncation argument and dominated convergence (set $\psi_n = (\psi \wedge n) \vee (-n)$) completes the proof. ■

Proposition 7.2 *There is a unique linear extension of $M : \mathcal{S} \rightarrow \mathcal{M}_{loc}$ (the space of local martingales) to a map $M : \mathcal{L}_{loc}^2 \rightarrow \mathcal{M}_{loc}$ such that $M_t(\psi)$ is a local martingale with increasing process $\langle M(\psi) \rangle_t$ given by*

$$\langle M(\psi) \rangle_t := \int_0^t \gamma X_s(\psi_s^2) ds \quad \forall t \geq 0 \text{ a.s. } \forall \psi \in \mathcal{L}_{loc}^2.$$

Proof. We can choose $\psi_n \in \mathcal{S}$ as in the Lemma. Then

$$\begin{aligned}\langle M(\psi) - M(\psi_n) \rangle_n &= \langle M(\psi - \psi_n) \rangle_n = \gamma \int_0^n X_s (\gamma(\psi(s) - \psi_n(s))^2 ds \\ P(\langle M(\psi) - M(\psi_n) \rangle_n > 2^{-n}) &< 2^{-n}\end{aligned}$$

The $\{M_t(\psi_n)\}_{t \geq 0}$ is Cauchy and using Doob's inequality and the Borel-Cantelli Lemma we can define $\{M_t(\psi)\}_{t \geq 0}$ such that

$$\sup_{t \leq n} |M_t(\psi) - M_t(\psi_n)| \rightarrow 0 \text{ a.s. as } n \rightarrow \infty.$$

This yields the required extension and its uniqueness. ■

Note that it immediately follows by polarization that if $\psi, \phi \in \mathcal{L}_{\text{loc}}^2$,

$$\langle M(\phi), M(\psi) \rangle_t = \gamma \int_0^t X_s (\phi_s \psi_s) ds$$

Moreover, in this case $M_t(\psi)$ is a L^2 -martingale, that is,

$$E(\langle M(\psi) \rangle_t) = \gamma \int_0^t E(X_s(\psi_s^2)) ds < \infty$$

provided that

$$\psi \in \mathcal{L}^2 = \{\psi \in \mathcal{L}_{\text{loc}}^2 : E\left(\int_0^t X_s(\psi_s^2) ds\right) < \infty, \forall t > 0\}.$$

Remark 7.3 Walsh (1986) [562] defined a more general class of martingale measures on a measurable space (E, \mathcal{E}) for which the above construction of stochastic integrals can be extended. $\{M_t(A) : t \geq 0, A \in \mathcal{E}\}$ is an L^2 -martingale measure wrt \mathcal{F}_t iff

- (a) $M_0(A) = 0 \quad \forall A \in \mathcal{E}$,
- (b) $\{M_t(A), t \geq 0\}$ is an \mathcal{F}_t -martingale for every $A \in \mathcal{E}$,
- (c) for all $t > 0$, M_t is an L^2 -valued σ -finite measure.

The martingale measure is worthy if there exists a σ -finite “dominating measure” $K(\cdot, \cdot, \cdot, \omega)$, on $\mathcal{E} \times \mathcal{E} \times \mathcal{B}(\mathbb{R}_+)$, $\omega \in \Omega$ such that

- (a) K is symmetric and positive definite, i.e. for any $f \in b\mathcal{E} \times \mathcal{B}(\mathbb{R}_+)$,

$$\int \int \int f(x, s) f(y, s) K(dx, dy, ds) \geq 0$$

- (b) for fixed A, B , $\{K(A \times B \times (0, t]), t \geq 0\}$ is \mathcal{F}_t -predictable
- (c) $\exists E_n \uparrow E$ such that $E\{K(E_n \times E_n \times [0, T])\} < \infty \forall n$,
- (d) $|\langle M(A), M(A) \rangle_t| \leq K(A \times A \times [0, t])$.

7.2.2 Uniqueness and stationary measures for the Jirina Martingale Problem

A probability law, $\mathbb{P}_\mu \in \mathcal{P}(C_{M_F([0,1])}([0, \infty)))$, is a solution of the *Jirina martingale problem*, if under \mathbb{P}_μ , $X_0 = \mu$ and

$$(7.9) \quad \begin{aligned} M_t(\phi) &:= X_t(\phi) - X_0(\phi) - \int_0^t c(\nu_0(\phi) - X_s(\phi))ds, \\ &\text{is a } L^2, \mathcal{F}_t\text{-martingale } \forall \phi \in b\mathcal{B}([0, 1]) \text{ with increasing process} \\ \langle M(\phi) \rangle_t &= \gamma \int_0^t X_s(\phi^2)ds, \text{ that is,} \\ M_t^2(\phi) - \langle M(\phi) \rangle_t &\text{ is a martingale.} \end{aligned}$$

Remark 7.4 This is equivalent to the martingale problem

$$(7.10) \quad M_F(t) = F(X_t) - \int_0^t GF(X(s))ds \quad \text{is a martingale}$$

for all $F \in \mathcal{D} \subset C(M_F([0, 1]))$ where

$$\mathcal{D} = \{F : F(\mu) = \prod_{i=1}^n \mu(f_i), f_i \in C([0, 1]), i = 1, \dots, n, n \in \mathbb{N}\}$$

and

$$\begin{aligned} GF(\mu) &= c \int \left[\int \frac{\partial F(\mu)}{\partial \mu(x)} \nu_0(dx) - \frac{\partial F(\mu)}{\partial \mu(x)} \right] \mu(dx) \\ &\quad + \frac{\gamma}{2} \int \int \frac{\partial^2 F(\mu)}{\partial \mu(x) \partial \mu(y)} (\delta_x(dy) \mu(dx) - \mu(dx) \mu(dy)) \end{aligned}$$

Theorem 7.5 There exists one and only one solution $\mathbb{P}_\mu \in \mathcal{P}(C_{M_F([0,1])}([0, \infty)))$ to the martingale problem (7.9). This defines a continuous $M_F([0, 1])$ -valued continuous strong Markov process.

(b) (Ergodic Theorem) Given any initial condition, X_0 , the law of X_t converges weakly to a limiting distribution as $t \rightarrow \infty$ with Laplace functional

$$(7.11) \quad E(e^{-\int_0^1 f(x)X_\infty(dx)}) = \exp \left(-\frac{2c}{\gamma} \int_0^1 \log(1 + \frac{f(x)}{\theta}) \nu_0(dx) \right).$$

This can be represented by

$$(7.12) \quad X_\infty(A) = \frac{1}{\theta} \int_0^1 f(s) G(\theta ds)$$

where G is the Gamma (Moran) subordinator (recall (6.17)).

Proof. Outline of method. As discussed above the projective limit construction produced a solution to this martingale problem.

A fundamental result of Stroock and Varadhan ([531] Theorem 6.2.3) is that in order to prove that the martingale problem has at most one solution it suffices to show that the one-dimensional marginal distributions $\mathcal{L}(X_t), t \geq 0$, are uniquely determined. Moreover in order

to determine the law of a random measure, X , on $[0, 1]$ it suffices to determine the Laplace functional.

The main step of the proof is to verify that if P_μ is a solution to the Jirina martingale problem, $t > 0$, and $f \in C_+([0, 1])$, then

$$(7.13) \quad E_\mu(e^{-X_t(f)}) = e^{-\mu(\psi(t)) - c\nu_0(\int_0^t \psi(t-s)ds)}$$

where

$$\begin{aligned} \frac{d\psi(s, x)}{ds} &= -c\psi(s, x) - \frac{\gamma}{2}\psi^2(s, x), \\ \psi(0, x) &= f(x). \end{aligned}$$

STEP 1: -discretization

We can choose a sequence of partitions $\{A_1^n, \dots, A_{K_n}^n\}$ and $\lambda_1^n, \dots, \lambda_{K_n}^n$ such that

$$(7.14) \quad \sum_{i=1}^{K_n} \lambda_i^n 1_{A_i^n} \uparrow f(\cdot).$$

We next show that for a partition $\{A_1, \dots, A_K\}$ of $[0, 1]$ and $\lambda_i \geq 0$, $i = 1, \dots, K$,

$$\begin{aligned} \exp(-\sum_{i=1}^K \lambda_i X_t(A_i)) &= \exp\left(-\sum_{i=1}^K \psi_i(t) X_0(A_i) - \sum c\nu_0(A_i) \int_0^t \psi_i(t-s) ds\right) \\ \frac{d\psi_i}{ds} &= -c\psi_i - \frac{\gamma}{2}\psi_i^2 \\ \psi_i(0) &= \lambda_i. \end{aligned}$$

To verify this first note that by Itô's Lemma, for fixed t and $0 \leq s \leq t$,

$$\begin{aligned} d\psi_i(t-s)X_s(A_i) &= X_s(A_i)d\psi_i(t-s) + \psi_i(t-s)dX_s(A_i) \\ &= X_s(A_i)d\psi_i(t-s) + \psi_i(t-s)c\nu_0(A_i) \\ &\quad - c\psi_i(t-s)X_s(A_i) + \psi_i(t-s)dM_s(A_i) \end{aligned}$$

and

$$\begin{aligned} X_t(A_i)\psi_i(0) - X_0(A_i)\psi_i(t) &= - \int_0^t X_s(A_i)\dot{\psi}_i(t-s)ds + c\nu_0(A_i) \int_0^t \psi_i(t-s)ds \\ &\quad - c \int_0^t X_s(A_i)\psi_i(t-s)ds + N_t(A_i) \end{aligned}$$

where $\{N\}_{0 \leq s \leq t}$ is an orthogonal martingale measure with

$$\begin{aligned} N_s(A_i) &= \int_0^s \psi_i(t-u)M(A_i, du) \\ \langle N(A_i) \rangle_s &= \frac{\gamma}{2} \int_0^s \psi_i^2(t-u)X_u(A_i)du \\ \langle N(A_i), N(A_j) \rangle_s &= 0 \text{ if } i \neq j. \end{aligned}$$

Again using Itô's lemma, for $0 \leq s \leq t$

$$\begin{aligned} de^{-X_s(A_i)\psi_i(t-s)} &= \dot{\psi}_i(t-s)e^{-X_s(A_i)\psi_i(t-s)}ds - \psi_i(t-s)e^{-X_s(A_i)\psi_i(t-s)}dX_s(A_i) \\ &\quad + \frac{\gamma}{2}e^{-X_s(A_i)\psi_i(t-s)}\psi_i^2(t-s)X_s(A_i)ds \\ &= \dot{\psi}_i(t-s)e^{-X_s(A_i)\psi_i(t-s)}ds + c\psi_i(t-s)e^{-X_s(A_i)\psi_i(t-s)}X_sds \\ &\quad + c\nu_0(A)\psi_i(t-s)e^{-X_s(A_i)\psi_i(t-s)}ds \\ &\quad + \frac{\gamma}{2}e^{-X_s(A_i)\psi_i(t-s)}\psi_i^2(t-s)X_sds + dN_s(A_i) \\ &= c\nu_0(A_i)\psi_i(t-s)e^{-X_s(A_i)\psi_i(t-s)}ds + dN_s(A_i) \end{aligned}$$

Then by the method of integrating factors we can get

$$\tilde{N}_s(A_i) = e^{(-X_s(A_i)\psi_i(t-s)+c\nu_0(A_i)\int_s^t \psi_i(t-u)du)}, \quad 0 \leq s \leq t,$$

is a bounded non-negative martingale that can be represented as

$$(7.15) \quad \tilde{N}_t(A_i) - \tilde{N}_0(A_i) = \int_0^t e^{-\zeta_i(s)}dN_s(A_i).$$

where

$$(7.16) \quad \zeta_i(s) = \left(c\nu_0(A_i) \int_s^t \psi_i(t-u)du \right).$$

Noting that the martingales $\tilde{N}_t(A_i), \tilde{N}_t(A_j)$ are orthogonal if $i \neq j$ we can conclude that

$$(7.17) \quad e^{-\sum_i (X_s(A_i)\psi_i(t-s)-c\nu_0(A_i)\int_s^t \psi_i(t-u)du)}, \quad 0 \leq s \leq t,$$

is a bounded martingale. Therefore for each n

$$(7.18) \quad E \left[e^{-\sum_{i=1}^{K_n} (X_t(A_i^n)\psi_i^n(0))} \right] = e^{-\sum_{i=1}^{K_n} (X_0(A_i^n)\psi_i^n(t)-c\nu_0(A_i^n)\int_s^t \psi_i^n(t-u)du)}$$

STEP 2: Completion of the proof

Taking limits as $n \rightarrow \infty$ and dominated convergence we can then show that the Laplace functional

$$E(e^{-X_t(f)}) = e^{-X_0(\psi(t))-c\nu_0(\int_0^t \psi(t-s)ds)}$$

where

$$\begin{aligned} \frac{d\psi(s, x)}{ds} &= -c\psi(s, x) - \frac{\gamma}{2}\psi^2(s, x), \\ \psi(0, x) &= f(x). \end{aligned}$$

Therefore the distribution of $X_t(f)$ is determined for any non-negative continuous function, f , on $[0, 1]$. Since the Laplace functional characterizes the law of a random measure, this

proves that the distribution at time t is uniquely determined by the martingale problem. This completes the proof of uniqueness.

(b) Recall that $\psi(\cdot, \cdot)$ satisfies

$$\begin{aligned}\frac{d\psi(x, s)}{ds} &= -c\psi(x, s) - \frac{\gamma}{2}\psi^2(x, s), \\ \psi(x, 0) &= f(x)\end{aligned}$$

Solving, we get

$$\psi(x, t) = \frac{f(x)e^{-ct}}{1 + \frac{f(x)}{\theta} - \frac{f(x)}{\theta}e^{-ct}}, \quad \theta = \frac{2c}{\gamma}$$

Next, note that $\psi(x, t) \rightarrow 0$ as $t \rightarrow \infty$ and

$$\begin{aligned}\int_0^\infty \psi(x, s) ds &= \int_0^\infty \frac{f(x)e^{-ct}}{1 + \frac{f(x)}{\theta} - \frac{f(x)}{\theta}e^{-ct}} dt = \int \frac{(-\frac{f(x)}{c})d(e^{-ct})}{1 + \frac{f(x)}{\theta} - \frac{f(x)}{\theta}(e^{-ct})} \\ &= \frac{2}{\gamma} \int_0^1 \frac{\frac{f(x)}{\theta}du}{1 + \frac{f(x)}{\theta} - \frac{f(x)}{\theta}u} = \frac{2}{\gamma} \int_0^{\frac{f(x)}{\theta}} \frac{\frac{f(x)}{\theta}du}{1 + \frac{f(x)}{\theta} - \frac{f(x)}{\theta}u} \\ &= \frac{2}{\gamma} \log(1 + \frac{f(x)}{\theta})\end{aligned}$$

Therefore

$$E(e^{-X_t(f)}) \rightarrow e^{-\frac{2c}{\gamma} \int \log(1 + \frac{f(x)}{\theta}) \nu_0(dx)}$$

This coincides with the Laplace functional of

$$\frac{1}{\theta} \int f(s)G(\theta ds)$$

where $G(\cdot)$ is the Moran subordinator. Therefore $X_\infty(f)$ can be represented as

$$X_\infty(f) = \frac{1}{\theta} \int f(s)G(\theta ds).$$

■

Remark 7.6 A more general class of measure-valued branching processes, known as superprocesses or Dawson-Watanabe processes will be discussed in Section 9.4.

7.3 The infinitely many alleles martingale problem

From the construction above we can show that the probability law of the infinitely many alleles Fleming-Viot process $\{X_t : t \geq 0\}$ on $C([0, \infty), M_1([0, 1]))$ satisfies the *martingale problem*

$$(7.19) \quad M_t(\phi) := X_t(\phi) - X_0(\phi) - \int_0^t c(\nu_0(\phi) - X_s(\phi))ds,$$

is a $L^2 \mathcal{F}_t$ martingale $\forall \phi \in b\mathcal{E}$ with increasing process

$$(7.20) \quad \langle M(\phi) \rangle = \gamma \int_0^t (X_s(\phi^2) - X_s(\phi)^2)ds.$$

We now show that this martingale problem completely characterizes the process. First note that $M_t(\phi)$ extends to a martingale measure with covariation

$$\langle M(A), M(B) \rangle_t = \int_0^t Q(X_s; A, B) ds$$

where

$$Q(\mu; dx, dy) = \mu(dx)\delta_x(dy) - \mu(dx)\mu(dy).$$

We observe that M is a worthy martingale measure with dominating measure $K(dx, dy, ds) = Q(X_s; A, B)ds$, because

$$|\int_0^t Q(X_s; A, A) ds| \leq \int_0^t |Q(X_s; A, A)| ds \leq t.$$

In order to prove that the martingale problem is well-posed we will introduce moment measures.

Let X be a random probability measure on the Polish space (E, \mathcal{E}) . The n th moment measure is a probability measure on E^n defined as follows:

$$M_n(dx_1, \dots, dx_n) = E(X(dx_1), \dots, X(dx_n))$$

M_n is the probability law of n -exchangeable E -valued random variables (Z_1, \dots, Z_n) .

Lemma 7.7 (a) *A random probability measure X on E is uniquely determined by its moment measures of all orders.*

(b) *The sequence $\{X_n\}$ of random probability measures with moment measures $\{M_{n,m}, n, m \in \mathbb{N}\}$ converges weakly to a random probability measure X with moment measures $\{M_m\}$ as $n \rightarrow \infty$ if and only if $M_{n,m} \Rightarrow M_m$ for each $m \in \mathbb{N}$.*

Theorem 7.8 *There exists a unique solution, \mathbb{Q} , to the martingale problem (7.19), (7.20).*

Proof. The existence has been proved above.

By the result of Stroock and Varadhan it suffices to show that the one-dimensional marginal distributions are uniquely determined. But from the Lemma, to determine the law of the random measure, Z_t , on $[0, 1]$ it suffices to determine all the moment measures.

But by Ito's Lemma, the moment measures satisfy the following system of equations

$$\begin{aligned} & \frac{\partial M_n(t; dx_1, \dots, dx_n)}{\partial t} \\ &= \sum_{i=1}^n c[M_{n-1}(t; dx_1, \dots, \cancel{x}_i, \dots, dx_n)\nu_0(dx_i) - M_n(t; dx_1, \dots, dx_n)] \\ & - \frac{1}{2}\gamma n(n-1)M_n(t; dx_1, \dots, dx_n) \\ & + \frac{1}{2}\gamma \sum_i \sum_{j \neq i} M_{n-1}(t; dx_1, \dots, dx_{i-1}, \cancel{dx}_i, dx_{i+1}, \dots, dx_n)\delta_{x_j}(dx_i) \end{aligned}$$

$$M_n(0; dx_1, \dots, dx_n) = \mu(dx_1) \dots \mu(dx_n).$$

Then

$$M_1(t, dx) = e^{-ct}X_0(dx) + (1 - e^{-ct})\nu_0(dx)$$

and we can then solve the remaining equations recursively. This implies that all the moment measures of Z_t are uniquely determined by the martingale problem. Hence the martingale problem has a unique solution. ■

7.4 Dual martingale problems

Dual processes play an important role in the study of interacting particle systems (see Liggett [409]). A dual representation for the Fleming-Viot process was introduced in Dawson and Hochberg (1982) [111]. The following generalization with applications to measure-valued processes was established in (Dawson-Kurtz (1982) [112]). Here we give the main ideas and refer [120], Sect. 5.5 for the details.

To give the main idea we first present the theorem in a simplified case.

Theorem 7.9 (*Dual Representation*)

Let E_1, E_2 be Polish spaces and $F(\cdot, \cdot)$, $GF(\cdot, \cdot)$, $HF(\cdot, \cdot) \in \mathcal{B}_b(E_1 \times E_2)$, $\beta \in \mathcal{B}_b(E_2)$ and $P_x : E_1 \rightarrow \mathcal{P}(D_{E_1}(0, \infty))$ and $Q_y : E_2 \rightarrow \mathcal{P}(D_{E_2}(0, \infty))$

Assume that

$$(7.21) \quad \begin{aligned} F(X(t), y) - \int_0^t GF(X(s), y)ds &\text{ is a } P_{X(0)} \text{ martingale for each } y \in E_2 \\ F(x, Y(t)) - \int_0^t HF(x, Y(s))ds &\text{ is a } Q_{Y(0)} \text{ martingale for each } x \in E_1 \end{aligned}$$

and

$$(7.22) \quad GF(x, y) = HF(x, y) + \beta(y)F(x, y).$$

Then

$$(7.23) \quad E_x^X(F(X(t), y)) = E_y^Y \left(F(x, Y(t)) \exp\left(\int_0^t \beta(Y(s))ds\right) \right), \quad 0 < t < T$$

Proof. Let

$$(7.24) \quad \Phi(s, t) := E_x^X \otimes E_y^Y \left(F(X(s), Y(t)) \exp\left(\int_0^t \beta(Y(u))du\right) \right)$$

$$(7.25) \quad \Phi(t, 0) = E_x^X(F(X(t), y))$$

$$(7.26) \quad \Phi(0, t) = E_y^Y(F(x, Y(t)))$$

$$(7.27) \quad \Phi_1(s, t) = E_x^X \otimes E_y^Y \left(GF(X(s), Y(t)) \exp\left(\int_0^t \beta(Y(u))du\right) \right)$$

$$(7.28) \quad \Phi_2(t, s) = E_x^X \otimes E_y^Y \left([HF(X(t), Y(s)) + \beta Y(s)F(X(t), Y(s))] \exp\left(\int_0^s \beta(Y(u))du\right) \right)$$

where Φ_1, Φ_2 are the first partial derivatives with respect to the first and second variables. Under the assumptions, $\Phi_1(s, t-s), \Phi_2(s, t-s)$, $0 \leq s \leq t$ exist and are uniformly bounded .

Therefore

$$(7.29) \quad \Phi(0, t) - \Phi(t, 0) = \int_0^t \frac{\partial}{\partial s} \Phi(s, t-s) = \int_0^t (\Phi_1(s, t-s) - \Phi_2(s, t-s))ds = 0$$

In applications the assumption that $\beta(\cdot)$ and $GF(\cdot, \cdot)$ are bounded needs to be relaxed. The following extension (see [120], Cor. 5.5.3) provides the required conditions.

Proposition 7.10 *Assume that*

- (i) $F \in C_b(E_1 \times E_2)$, and $\{F(\cdot, y) : y \in E_2\}$ is measure-determining on E_1
- (ii) there exist stopping times $\tau_K \uparrow t$ such that

$$(7.30) \quad \left\{ (1 + \sup_x |GF(x, Y(\tau_K))|) \cdot \exp\left(\int_0^{\tau_K} |\beta(Y(u))| du\right) \right\}_K$$

are Q_{δ_y} – uniformly integrable for all $y \in E_2$

and (iii) $Q_{\delta_y}(Y(s-) \neq Y(s)) = 0$ for each $s \geq 0$, that is, no fixed discontinuities.

Then the G -martingale problem is well-posed and for all $y \in E_2$

$$(7.31) \quad P_\mu(F(X(t), y)) = \int_{E_1} \mu(dx) \left(Q_{\delta_y}[F(x, Y(t))] \exp\left(\int_0^t \beta(Y(u)) du\right) \right).$$

Example 7.11 (*The Wright-Fisher diffusion with polynomial drift*)

Let $\Delta_{d-1} = \{(x_1, \dots, x_d), x_i \geq 0, i = 1, \dots, d, \sum_{i=1}^d x_i \leq 1\}$

Then consider the Wright-Fisher diffusion $\{x(t)\}$ with generator

$$G = \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i(x) \frac{\partial}{\partial x_i}$$

where $\{a_{i,j}(x)\}$ is the real symmetric non-negative definite matrix, $\{a_{ij}(x)\} = \{x_i(\delta_{ij} - x_j)\}$ and the drift coefficient $b_i(x)$ is a polynomial satisfying certain natural boundary conditions on Δ_{d-1} to ensure that the process remains in Δ_{d-1} .

Shiga (1981) [511] obtained a dual in terms of a family of functions $\{\phi_\alpha\}_{\alpha \in \Gamma}$, $\phi_\alpha \in D(G)$ defined by

$$\phi_\alpha(x_1, \dots, x_d) = \prod_{i=1}^d x_i^{\alpha_i}, \quad \alpha = (\alpha_1, \dots, \alpha_d) \in \Gamma$$

and showed that

$$G\phi_\alpha = \sum_\beta Q_{\alpha,\beta}(\phi_\beta - \phi_\alpha) + h_\alpha \phi_\alpha$$

where $Q = \{Q_{\alpha,\beta}\}$ defines a conservative Markov chain α_t with state space Γ . Then the following identity follows from Proposition 7.10:

$$(7.32) \quad E_x[\phi_\alpha(x(t))] = E_\alpha[\phi_{\alpha_t}(x) \exp\left(\int_0^t h_{\alpha_u} du\right)], \quad 0 \leq t \leq t_0$$

provided that

$$E_\alpha\left[\exp\left(\int_0^{t_0} |h_{\alpha_u}| du\right)\right] < \infty \quad \forall \alpha \in \Gamma$$

Therefore the corresponding Wright-Fisher martingale problem is well-posed.

Example 7.12 (*Markov chains*) Consider a continuous time Markov chain with state space $E_K = \{1, \dots, K\}$ and transition rates

$$(7.33) \quad i \rightarrow j \text{ with rate } m_{ij}, \quad i \neq j, \quad m_{ii} = - \sum_{j \neq i} m_{ij}.$$

Let \mathcal{PK} denote the collection of subsets of E_K and define the function $F : \mathcal{PK} \times E_K$ by

$$(7.34) \quad F(A, j) = 1_A(j)$$

Now consider the Markov \mathcal{A}_t chain with state space \mathcal{PK} and transition rates

$$(7.35) \quad A \rightarrow A \cup \{j\} \text{ at rate } \sum_{\ell \in A} m_{j\ell}, \quad j \in A^c$$

$$(7.36) \quad A \rightarrow A \setminus \{j\} \text{ at rate } \sum_{\ell \in A^c} m_{j\ell} \quad \text{if } j \in A$$

$$(7.37) \quad GF(A, j) = \sum_{\ell} m_{j\ell} (1_A(\ell) - 1_A(j)) = \sum_{\ell \in A} m_{j\ell} (1 - F(A, j)) - \sum_{\ell \in A^c} m_{j\ell} F(A, j)$$

Then

$$(7.38) \quad HF(A) = \sum_{k \in A^c} \left[\left(\sum_{\ell \in A} m_{k\ell} \right) (F(A \cup \{k\}) - F(A)) + \left(\sum_{k \in A} \left(\sum_{\ell \in A^c} m_{k\ell} \right) (F(A \setminus \{k\}) - F(A)) \right) \right]$$

and therefore

$$(7.39) \quad \begin{aligned} HF(A, j) &= \sum_{k \in A^c} \left(\sum_{\ell \in A} m_{k\ell} \right) (1_{(A \cup \{k\})}(j) - 1_A(j)) + \sum_{k \in A} \left(\sum_{\ell \in A^c} m_{k\ell} \right) (1_{(A \setminus \{k\})}(j) - 1_A(j)) \\ &= \sum_{\ell \in A} m_{j\ell} (1 - F(A, j)) - \sum_{\ell \in A^c} m_{j\ell} F(A, j). \end{aligned}$$

By duality we have

$$(7.40) \quad E_j(1_{\ell}(X_t)) = E_{\{\ell\}}(1_{\mathcal{A}_t}(j)).$$

Remark 7.13 If $\{m_{ij}\}$ is irreducible, then the Markov chain \mathcal{A}_t has two traps \emptyset and E_K . It is easy to verify that \mathcal{A}_t is absorbed at a trap with probability one. This together with (7.40) implies (the elementary result) that $P_j(x(t) = \ell)$ converges as $t \rightarrow \infty$ to a stationary measure π_{ℓ} with

$$(7.41) \quad \pi_{\ell} = P_{\{\ell\}}(\mathcal{A}_t \rightarrow E_K), \quad \ell \in E_K$$

and that $\lim_{t \rightarrow \infty} P_j(X(t) = \ell)$ is independent of j .

7.5 Dual representation of the neutral Fleming-Viot process

The method of dual representation plays an important role in the study of Fleming-Viot processes and will be frequently used below. To introduce this we first consider the special case of a neutral Fleming-Viot process with a nice mutation process.

7.5.1 The General Neutral F.V. Process

Let E be a compact metric space, A be a linear operator defined on $D(A) \subset C(E)$ and assume that the closure of A generates a Feller semigroup, $\{S_t : t \geq 0\}$ on $C(E)$. A probability measure \mathbb{P}_μ on $C([0, \infty), M_f(E))$ is said to be a solution of the *neutral Fleming-Viot martingale problem* $\text{MIP}_{(A, Q, 0)}$ with initial condition μ if

$$\mathbb{P}_\mu(X_0 = \mu) = 1$$

and for each $\phi \in C_b^+(E) \cap D(A)$

$$M_t^0(\phi) := \langle \phi, X_t \rangle - \langle \phi, X_0 \rangle - \int_0^t \langle A\phi, X_s \rangle ds$$

where M_t^0 defines a martingale measure $M^0(ds, dx)$ with covariance

$$\begin{aligned} \langle M^0(dx), M^0(dy) \rangle_t &= \gamma \int_0^t Q(X_s; dx, dy) ds \\ Q(\mu; dx, dy) &= \delta_x(dy)\mu(dx) - \mu(dx)\mu(dy). \end{aligned}$$

Theorem 7.14 *There exists a unique solution to the $\text{MIP}_{(A, Q, 0)}$ martingale problem.*

Proof. This will be proved in the following section. ■

7.5.2 Equivalent Formulation of the martingale problem

We now turn to an equivalent formulation of the Fleming-Viot process that will be needed for the application of the dual representation in the next chapter.

Let $F \in D(G) \subset C^2(\mathcal{P}(E))$

$$(7.42) \quad GF(\mu) = \int_E \left(A \frac{\delta F(\mu)}{\delta \mu(x)} \right) \mu(dx) + \frac{\gamma}{2} \int_E \int_E \frac{\delta^2 F(\mu)}{\delta \mu(x) \delta \mu(y)} Q(\mu; dx, dy)$$

where $Q(\mu, dx, dy) := \mu(dx)\delta_x(dy) - \mu(dx)\mu(dy)$.

Now consider function $F(\mu, (f, n)) = \int \cdots \int f(x_1, \dots, x_n) \mu^n(dx)$ with $f \in C(E^n)$, $n \in \mathbb{N}$ and

$$(7.43) \quad \mu^n(dx) = \mu(dx_1) \dots \mu(dx_n).$$

Then

$$(7.44) \quad GF(\mu, (f, n)) = \langle \mu^n, A^{(n)} f \rangle + \frac{\gamma}{2} \sum_{i \neq j} \left(\langle \mu^{n-1}, \tilde{\Theta}_{ij} f \rangle - \langle \mu^n, f \rangle \right)$$

$$(7.45) \quad (\tilde{\Theta}_{ij}f)(y_1, \dots, y_{N-1}) := f(x_1, \dots, x_N)$$

On the right side of (7.45)

$$(7.46) \quad \begin{aligned} x_k &= y_k \text{ for } k < i \vee j, k \neq i \wedge j \\ x_{i \vee j} &= x_{i \wedge j} = y_{i \wedge j} \\ x_k &= y_{k-1} \text{ for } k > i \vee j. \end{aligned}$$

7.5.3 The dual representation of the Fleming-Viot process

The Fleming-Viot process has state space $\mathcal{P}(E)$. We assume that the mutation process has semigroup S_t with generator A and there exists an algebra of functions $D_0(E)$ dense in $C(E)$ and $S_t : D_0(E) \rightarrow D_0(E)$.

We can then consider the extension of the mutation process to E^n , $n \geq 1$ corresponding to n i.i.d. copies of the basic mutation process and with generator $A^{(n)} = \sum_{i=1}^n A_i$ where A_i denotes the action of A on the i th variable.

Let

$$(7.47) \quad E_2 := \{(f, n) : f \in (D_0(E))^n \cap, n \in \mathbb{N}\}.$$

Define $F : \mathcal{P}(E) \times E_2 \rightarrow \mathbb{R}$ by

$$(7.48) \quad F(\mu, (f, n)) = \int_{E^n} f_n(x_1, \dots, x_n) \mu(dx_1) \dots \mu(dx_n).$$

Now consider the Fleming-Viot process with generator:

$$(7.49) \quad GF(\mu, (f, n)) = \int_E \left(A \frac{\partial F(\mu, (f, n))}{\partial \mu(x)} \right) \mu(dx) + \frac{\gamma}{2} \int_E \int_E \frac{\partial^2 F(\mu, (f, n))}{\partial \mu(x) \partial \mu(y)} Q(\mu; dx, dy)$$

and note that for each $\mu \in \mathcal{P}(E)$ this coincides with

$$(7.50) \quad HF(\mu, (f, n)) = F(\mu, (A^{(n)} f, n)) + \frac{\gamma}{2} \sum_{j=1}^n \sum_{k \neq j} [F(\mu, (\tilde{\Theta}_{jk} f, n)) - F(\mu, (f, n))]$$

where $\tilde{\Theta}_{jk} : (D_0(E))^n \rightarrow (D_0(E))^{n-1}$ is defined by (7.45).

Then H is the generator of a càdlàg process with values in E_2 and law $\{Q_f : f \in E_2\}$ which evolves as follows:

- $Y(t)$ jumps from $(D_0(E)^n, n)$ to $(D_0(E)^{n-1}, n-1)$ at rate $\frac{1}{2}\gamma n(n-1)$
- at the time of a jump, f is replaced by $\tilde{\Theta}_{jk} f$
- between jumps, $Y(t)$ is deterministic on $D_0(E)^n$ and evolves according to the semigroup (S_t^n) with generator $A^{(n)}$.

Theorem 7.15 (a) Let $(\{X(t)\}_{t \geq 0}, \{P_\mu : \mu \in \mathcal{P}(E)\})$ be a solution to the Fleming-Viot martingale problem and the process $(\{Y(t)\}_{t \geq 0}, \{Q_{(f,n)} : (f, n) \in E_2\})$ be defined as above. Then
(a) these processes are dual, that is,

$$(7.51) \quad P_\mu(F(X(t), (f, n))) = Q_f(F(\mu, Y(t))), \quad (f, n) \in E_2.$$

(b) The martingale problem is well-posed and the Fleming-Viot process is a strong Markov process.

Proof. In this case for $(f, n) \in E_2$ $\mu \in \mathcal{P}(E)$,

$$(7.52) \quad GF(\mu, (f, n)) = HF(\mu, (f, n))$$

and the uniqueness follows from Theorem (7.9). (b) follows by the Stroock-Varadhan Theorem. \blacksquare

7.5.4 The Kingman coalescent

Consider the special case with no mutation, that is, $A \equiv 0$. Then we can represent the dual process $Y(t)$ with $Y(0) = (f, n)$ as follows.

$$(7.53) \quad Y(t) = (f_t, n_t)$$

where $n_t \leq n$ and there is a map

$$(7.54) \quad \pi_t : \{1, \dots, n\} \rightarrow \{1, \dots, n_t\}$$

and $f_t \in C(E^{n_t})$ given by

$$(7.55) \quad f_t(y_1, \dots, y_{n_t}) = f(x_1, \dots, x_n) \quad \text{with } x_i = y_{\pi_t(i)}, \quad i = 1, \dots, n.$$

In other words π_t is a process with values in the set of partitions of $\{1, \dots, n\}$ and n_t is a pure death process with deaths rate $\gamma \binom{k}{2}$ where $n_t = k$. This partition-valued process is the *Kingman coalescent* [368] and plays an important role in population genetics.

7.6 Interactions via change of measure- the Girsanov formula

We have established uniqueness for the Fleming-Viot process with diploid selection using duality in the previous section. However for more general state dependent fitness functions there is no natural dual process. Instead we can use a change of measure argument based on a Girsanov-type formula. Here we consider the Girsanov transformation for measure-valued processes introduced in [110]. We give here a version suitable for applications in later chapters on selection and logistic competition.

Preliminaries

We begin by reviewing with some general notions of stochastic analysis.

Definition. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ be a complete probability space such that \mathcal{F}_0 contains all P -null sets of \mathcal{F} and \mathcal{F}_t is right continuous.

An \mathcal{F}_t -adapted càdlàg process, Y , is a (*classical*) *semimartingale* if there exists processes N and B with $N_0 = B_0 = 0$ and

$$Y_t = Y_0 + N_t + B_t$$

where N_t is a local martingale and B_t is a finite variation process.

A generalization of the classical Girsanov Theorem to semimartingales due to Meyer is as follows.

Theorem 7.16 ([488], Chap. III, Theorem 20)

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ be as above. Let X be a classical semimartingale under \mathbb{P} , with decomposition $X = M + A$ where M is a local martingale. Let \mathbb{Q} be equivalent to \mathbb{P} with

$$Z_t = E^{\mathbb{P}} \left[\frac{d\mathbb{Q}}{d\mathbb{P}} \mid \mathcal{F}_t \right]$$

Then X is also a classical semimartingale under \mathbb{Q} and has a decomposition $X = M^{\mathbb{Q}} + C$ where

$$M_t^{\mathbb{Q}} = M_t - \int_0^t \frac{1}{Z_s} d[Z, M]_s$$

is a \mathbb{Q} -local martingale and $C = X - M^{\mathbb{Q}}$ is a \mathbb{Q} -finite variation process.

Proof. The idea is to use Itô's Lemma to show that $Z_t^{-1} M_t^{\mathbb{Q}}$ is a \mathbb{P} -martingale. ■

Let $\{M_t\}$ be a martingale and $f(\cdot) \in L_{\text{loc}}^2(M)$. Then by Itô's Lemma the stochastic exponential

$$(7.56) \quad Z_t := \exp \left(\int_0^t f_s dM_s - \frac{1}{2} \int_0^t f_s^2 d\langle M \rangle_s \right) \quad \text{which satisfies } dZ_t = Z_t f_t dM_t$$

is a local martingale.

Proposition 7.17 (Novikov's condition) A sufficient condition for the stochastic exponential Z_t to be a martingale is that

$$E \left[\frac{1}{2} \exp \left(\int_0^t f_s^2 d\langle M \rangle_s \right) \right] < \infty.$$

See Ikeda-Watanabe [309] Theorem 5.3.

Girsanov's Transformation for Measure-Valued Processss

Let E be a locally compact space,

- $\mathcal{D}(A)$ be a measure-determining linear subspace of $C_b(E)$ containing constants and A a linear mapping from $\mathcal{D}(A) \rightarrow C_b(E)$
- $Q : M_F(E) \rightarrow M_F(E \times E)$ is continuous, and

$$(7.57) \quad Q(\mu, B, B) \leq K\mu(B), \quad K < \infty, \quad B \in \mathcal{B}(E),$$

- V is be a measurable function $V : [0, \infty) \times M_F(E) \times E \rightarrow R$

Then a probability measure \mathbb{P}_μ on $C([0, \infty), M_F(E))$ is said to solve the *martingale problem* $\text{MMP}_{(A, Q, V)}$ with initial condition μ if

$$\mathbb{P}_\mu(X_0 = \mu) = 1$$

and for each $\phi \in \mathcal{D}(A)$

$$\begin{aligned} M_t^V(\phi) &:= \langle \phi, X_t \rangle - \langle \phi, X_0 \rangle - \int_0^t \langle A\phi, X_s \rangle ds \\ &\quad - \int_0^t \int \int \phi(x) V(s, X_s, y) (Q(X_s; dx, dy)) ds \end{aligned}$$

is a P_μ -martingale with increasing process

$$(7.58) \quad \langle M^V(\phi) \rangle_t = \int_0^t \int \int \phi(x) \phi(y) Q(X_s, dx, dy) ds.$$

Then M_t^V defines a martingale measure $M^V(ds, dx)$ with covariation

$$(7.59) \quad \langle M^V(dx), M^V(dy) \rangle_t = \int_0^t Q(X_s; dx, dy) ds.$$

Theorem 7.18 Assume that \mathbb{P}_μ is the unique solution of the martingale problem $\text{MIP}_{(A, Q, 0)}$ and that \mathbb{P}_μ -a.s.

$$(7.60) \quad \int_0^t \int \int V(s, X_s, x) V(s, X_s, y) Q(X_s; dx, dy) ds < \infty, \quad \forall t > 0.$$

Define the \mathbb{P}_μ continuous local martingales:

$$(7.61) \quad N_t^V = \frac{1}{\gamma} \int_0^t \int V(s, X_s, y) M^0(ds, dy)$$

$$(7.62) \quad \langle N^V \rangle_t = \int_0^t \int \int V(s, X_s, x) V(s, X_s, y) Q(X_s; dx, dy) ds$$

and the stochastic exponential

$$(7.63) \quad Z_t^V := \exp \left(N_t^V - \frac{1}{2} \langle N^V \rangle_t \right).$$

(a) (Existence) Assume that (7.60) holds \mathbb{P}_μ -a.s. Then $\mathbb{Q}_\mu := Z_t^V \mathbb{P}_\mu$ is a solution to the (A, Q, V) local martingale problem.

(b) (Uniqueness) If \mathbb{Q}_μ is any solution of the martingale problem $\text{MIP}_{(A, Q, V)}$ such that (7.60) holds \mathbb{Q}_μ a.s., then

$$\frac{d\mathbb{Q}_\mu}{d\mathbb{P}_\mu}|_{\mathcal{F}_t} = Z_t^V$$

and therefore there is only one such solution.

Proof. Note that

$$(7.64) \quad M_t^V(\phi) = M_t^0(\phi) - \langle N^V, M^0(\phi) \rangle_t.$$

(a) Assume that \mathbb{P}_μ is a solution to the $(A, Q, 0)$ martingale problem and that (7.60) holds \mathbb{P}_μ -a.s. By Itô's formula we have

$$\begin{aligned} dZ_t^V M_t^V(\phi) &= d(Z_t^V (M_t^0(\phi) - \langle N_t^V, M_t^V(\phi) \rangle)) \\ &= (M_t^0(\phi) - \langle N_t^V, M_t^V(\phi) \rangle) dZ_t^V + Z_t^V dM_t^0(\phi) - Z_t^V \langle N^V, M^0(\phi) \rangle_t + dZ_t \cdot dM_t^0 \\ &= (M_t^0(\phi) - \langle N_t^V, M_t^V(\phi) \rangle) dZ_t^V + Z_t^V dM_t^0(\phi) \end{aligned}$$

since by (7.56) $dZ_t^V \cdot dM_t^0 = Z_t^V dN_t^V dM_t^0(\phi) = Z_t^V d\langle M^0(\phi), N^V \rangle_t$

$$\begin{aligned}
 (7.65) \quad & d(Z_t M_t^V(\phi)) = d(Z_t(M_t^0(\phi) - \frac{1}{2} - \int_0^t \int \int \phi(x)V(s, X_s, y)(Q(X_s; dx, dy))ds)) \\
 &= M_t^0(\phi)dZ_t + Z_t(dM_t^0 - \int \int \phi(x)V(s, X_s, y)(Q(X_s; dx, dy))dt) \\
 &\quad + Z_t \langle M_t^0(\phi), \int_0^t \int V(s, X_s, y)M^0(ds, dy) \rangle \\
 &= M_t^0(\phi)dZ_t + Z_t dM_t^0(\phi)
 \end{aligned}$$

so that $M_t^V(\phi)$ is a local martingale under $\mathbb{Q}_\mu := Z_t^V \mathbb{P}_\mu$. Moreover the quadratic variation of M_t^V is

$$(7.66) \quad \langle M^V(\phi), M^V(\phi) \rangle_t = \int_0^t \int \int \phi(x)\phi(y)Q(X_s; dx, dy)ds, \quad \mathbb{Q}_\mu - a.s$$

In other words \mathbb{Q}_μ is a solution to the (A, Q, V) local martingale problem.

(b) Now assume that \mathbb{Q}_μ is a solution to the (A, Q, V) -martingale problem and (7.60) holds \mathbb{Q}_μ a.s. The same argument as (a) implies that $M_t^0(\phi)$ is a local martingale under $Z_t^{-V} \mathbb{Q}$ and $M_t^0(\phi)Z_t^{-V}$ is a \mathbb{Q}_μ -local martingale. Let

$$(7.67) \quad \tau_n = \inf\{t : \int_0^t [\int \int (V(s, X_s, x)V(s, X_s, y) + 1)Q(X_s; dx, dy) + 1]ds \geq n\} \leq n$$

Since $\langle N^V \rangle_{t \wedge \tau_n}$ is bounded, Novikov's criterion implies that $Z_{t \wedge \tau_n}^{-V}$ is a martingale and

$$(7.68) \quad d\mathbb{P}_{\mu, n} := Z_{t \wedge \tau_n}^{-V} d\mathbb{Q}_\mu$$

defines probability and $M_t^0(\phi)$ is a \mathbb{P}_n -local martingale and

$$(7.69) \quad \langle M_{t \wedge \tau_n}^0(\phi), M_{t \wedge \tau_n}^0(\phi) \rangle = \int_0^{t \wedge \tau_n} \int \int \phi(x)\phi(y)Q(X_s; dx, dy)ds \quad \forall t \geq 0, \quad \mathbb{P}_n - a.s.$$

Since this is bounded it is integrable and therefore $M_t^0(\phi)$ is a \mathbb{P}_n martingale. Let $\mathbb{P}_n^{ex} | \mathcal{F}_{\tau_n} = \mathbb{P}_n | \mathcal{F}_{\tau_n}$ and $\mathbb{P}_n^{ex}(X_{\tau_n+} | \mathcal{F}_{\tau_n}) = \mathbb{P}_{X_{\tau_n}}(\cdot)$. Then \mathbb{P}_n^{ex} solves the $(A, Q, 0)$ martingale problem and therefore since we assumed that this is well-posed, we have $\mathbb{P}_n^{ex} = \mathbb{P}_\mu$. Therefore (7.60) implies that

$$(7.70) \quad \mathbb{P}_n^{ex}(\tau_n < t) = \mathbb{P}_\mu(\tau_n < t) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Then since we assume that \mathbb{Q}_μ solves the (A, Q, V) martingale problem and (7.60) holds \mathbb{Q}_μ a.s., M_t^{-V} is a \mathbb{Q}_μ non-negative local martingale. Then

$$\begin{aligned}
 E^{\mathbb{Q}_\mu}(Z_t^{-V}) &\geq E^{\mathbb{Q}_\mu}(Z_{t \wedge \tau_n}^{-V} 1_{\tau_n \geq t}) \\
 &= E^{\mathbb{Q}_\mu}(Z_{t \wedge \tau_n}^{-V}) - E^{\mathbb{Q}_\mu}(Z_{t \wedge \tau_n}^{-V} 1_{\tau_n < t}) \\
 &= 1 - \mathbb{P}_n(\tau_n < t) \rightarrow 1 \text{ as } n \rightarrow \infty.
 \end{aligned}$$

Hence Z_t^{-V} is a \mathbb{Q}_μ -martingale and we define

$$(7.71) \quad \tilde{\mathbb{P}}_\mu | \mathcal{F}_t = Z_t^{-V} d\mathbb{Q}_\mu | \mathcal{F}_t.$$

But then we can verify that $M_t^0(\phi)$ is a $\tilde{\mathbb{P}}_\mu$ local martingale and therefore $\tilde{\mathbb{P}}_\mu = \mathbb{P}_\mu$. This implies that

$$(7.72) \quad \mathbb{Q}_\mu|_{\mathcal{F}_t} = \frac{1}{Z_t^{-V}} \mathbb{P}_\mu = Z_t^V \mathbb{P}_\mu$$

and therefore it is unique. ■

Proposition 7.19 *If $\sup |V(s, \mu, x)| \leq V_0$ (constant), then Z_t^V is a martingale under \mathbb{P}_μ and $\mathbb{Q}_\mu := Z_t^V \mathbb{P}_\mu$ is the unique law that satisfies $\text{MIP}_{(A, Q, V)}$.*

Proof. Define

$$(7.73) \quad \tau_n = \inf\{t : \int_0^t [\int \int (V(s, X_s, x) V(s, X_s, y) + 1) Q(X_s; dx, dy) + 1] ds \geq n\} \leq n$$

$$V^n(s, X, x) = 1(s \leq \tau_n) V(s, X, x).$$

Then as in the proof of the Theorem $\mathbb{Q}_{\mu,n} := Z_{\tau_n}^V d\mathbb{P}_\mu$ is a solution to the $\text{MIP}_{(A, Q, V^n)}$ -martingale problem. Taking $\phi(x) \equiv 1$ we have

$$\begin{aligned} E^{\mathbb{Q}_n}(X_t(1)) &= \mu(1) + E^{\mathbb{Q}_n}\left(\int_0^{t \wedge \tau_n} X_s(V(s, X_s, \cdot)) ds\right) \\ &\leq \mu(1) + V_0 E^{\mathbb{Q}_n}\left(\int_0^{t \wedge \tau_n} X_s(1) ds\right) \end{aligned}$$

Then by Gronwall's inequality $E^{\mathbb{Q}_n}(X_t(1)) \leq \mu(1)e^{V_0 t}$ and therefore

$$\begin{aligned} E^{\mathbb{Q}_n}\left(\int_0^t [\int \int (V(s, X_s, x) V(s, X_s, y) + 1) Q(X_s; dx, dy)] ds\right) \\ \leq (V_0^2 + 1)\mu(1)e^{V_0 t} t + 2t = K(t). \end{aligned}$$

and then by Chebyshev $\mathbb{Q}_n(\tau_n < t) \leq K(t)/n \rightarrow 0$ and $n \rightarrow \infty$. Then

$$\begin{aligned} E^{\mathbb{P}_\mu}(Z_t^V) &\geq E^{\mathbb{P}_\mu}(Z_{t \wedge \tau_n}^V 1_{\tau_n \geq t}) \\ &= E^{\mathbb{P}_\mu}(Z_{t \wedge \tau_n}^V) - E^{\mathbb{P}_\mu}(Z_{t \wedge \tau_n}^V 1_{\tau_n < t}) \\ &= 1 - \mathbb{Q}_n(\tau_n < t) \end{aligned}$$

and therefore $E^{\mathbb{P}_\mu}(Z_t^V) = 1$ and Z_t^V is a \mathbb{P}_μ -martingale.

Moreover,

$$(7.74) \quad E^{\mathbb{Q}_\mu} \langle M^V(\phi) \rangle_t = E^{\mathbb{Q}_\mu} \left(\int_0^t [\int \int (\phi(x) \phi(x) Q(X_s; dx, dy)] ds \right) < \infty$$

and therefore $M_t^V(\phi)$ is a \mathbb{Q}_μ -martingale. ■

Remark 7.20 Ethier and Shiga (2000, 2002) [219], [220] established the Girsanov formula for a Fleming-Viot process with unbounded selection which arises as the diffusion limit of a model of Tachida (1991) [539] with type space \mathbb{R} , house of cards mutation with mutation source $\nu_0 =$

$N(0, \sigma_0^2)$ and fitness function $V(x) = x$. (The model was proposed by Tachida in the context of an ongoing discussion of the roles of genetic drift and weak selection in protein evolution.)

Overbeck, Röckner and Schmuland (1995) studied Fleming-Viot processes with interactive selection using Dirichlet forms.

Evans and Perkins (1994) [223] extend the Girsanov formula to the case of interacting species modeled by two interacting super-Brownian motions with either competition or predation. In the case of predation in which collisions effect only the prey they establish existence and uniqueness in dimensions one, two and three.

We will return to a systematic discussion of mutation-selection systems in Chapter 12.

Chapter 8

Genealogy and History

8.1 Introduction

In this Chapter we will focus on a neutral Fleming-Viot process with (or without) mutation. If for the moment we ignore mutation then we can focus on the family relations among the members of the population, for example the ancestral relation between a finite random sample from the population at a fixed time. This was the purpose of the Kingman coalescent that has become a standard tool of population genetics. With mutation it is also of interest to trace the mutational history of an individual and its ancestors. This is the purpose of the historical process. In the case of the infinitely many sites model the mutational history of an individual is built into the state of the individual and in this context we can explore the genealogy and mutational history in a unified manner. More generally, the idea is that giving the individuals labels or coding for certain genealogical or historical information can be a useful mathematical tool.

We will introduce some important tools in studying this richer structure including the Kingman coalescent (Kingman (1982) [368], [356]), the look-down process (Donnelly-Kurtz [163]), the tree-valued Fleming-Viot process (Greven-Pfaffelhuber-Winter [265]) and the analogue of the historical process (Dawson-Perkins [118]).

We will return to the question of the genealogical structure of Fleming-Viot processes with both mutation and selection in Chapter 12.

8.2 Family Structure of the neutral Fleming-Viot Process

8.2.1 Fleming-Viot with Feller Mutation Semigroup

We assume that the space of types is a compact set, E . The mutation process is assumed to be a Feller process with Feller semigroup $\{S_t : t \geq 0\}$ on $C(E)$ and with transition function $p_t(x, dy)$. We assume that the generator, is the closure of $(A, D(A))$ where A is a linear operator defined on a linear subspace, $D(A)$, of $C(E)$. By the theory of Feller processes, without loss of generality, we can assume that $D(A)$ contains a countable subset that is convergence determining and that there exists a càdlàg version of the process, $(D([0, \infty), E), (\mathcal{D}_t)_{t \geq 0}, \{P_x : x \in E\})$.

A probability measure \mathbb{P}_μ on $C([0, \infty), \mathcal{P}(E))$ is said to be a solution of the *neutral Fleming-Viot martingale problem* $\text{MIP}_{(A, \gamma Q, 0)}$ (wrt $D(A)$) with initial condition μ and resampling rate function $\gamma \in C([0, \infty), \mathbb{R}^+)$ if

$$\mathbb{P}_\mu(X_0 = \mu) = 1$$

and for each $\phi \in C_b^+(E) \cap D(A)$,

$$M_t(\phi) := \langle \phi, X_t \rangle - \langle \phi, X_0 \rangle - \int_0^t \langle A\phi, X_s \rangle ds$$

where M_t defines a martingale measure $M(ds, dx)$ with covariance

$$\begin{aligned} \langle M(dx), M(dy) \rangle_t &= \int_0^t \gamma(s) Q(X_s; dx, dy) ds \\ Q(\mu; dx, dy) &= \delta_x(dy)\mu(dx) - \mu(dx)\mu(dy). \end{aligned}$$

Equivalent Martingale Problem

For each $n \geq 1$, define the Feller semigroup $\{S_t^{(n)} : t \geq 0\}$ on $C(E^n)$ by

$$\begin{aligned} S_t^{(n)} f(x_1, \dots, x_n) \\ := \int \cdots \int f(y_1, \dots, y_n) p_t(x_1, dy_1) \dots p_t(x_n, dy_n) \end{aligned}$$

and let $A^{(n)}$ denote its generator.

We can also consider integrals with respect to the martingale measure of the form

$$\begin{aligned} M_t(f) := \int \cdots \int \gamma(s) f(x_1, \dots, x_n) M(ds, dx_1) M(ds, dx_2) \dots M(ds, dx_n), \\ f \in C_{sym}(E^n) \end{aligned}$$

and using Itô's lemma verify that

$$\langle M(f) \rangle_t = \sum_{1 \leq i < j \leq n} \int_0^t \gamma(s) \left(\langle \Phi_{ij}^{(n)} f, X_s^{n-1} \rangle - \langle f, X_s^n \rangle \right) ds$$

where

$$\begin{aligned} (8.1) \quad \Phi_{ij}^{(n)} : C(E^n) &\rightarrow C(E^{n-1}) \\ \Phi_{ij}^{(n)} f(x_1, \dots, x_{n-1}) &:= f(x_1, \dots, x_i, \dots, x_{j-1}, x_i, x_j, \dots, x_{n-1}). \end{aligned}$$

(Hint: First do this for linear combinations of functions of the form $f(x_1, \dots, x_n) = \prod_{i=1}^n \varphi_i(x_i)$ and then take limits in L^2 .)

Now consider the collection, $D(G)$, of subset of $C(M_1(E))$ of all linear combinations of functions of the form

$$(8.2) \quad F_f(\mu) = F(f, \mu) := \langle f, \mu^n \rangle, \quad n \in \mathbb{N}, \quad f \in C_{sym}(E^n) \cap D(A^{(n)}).$$

A function of the form (8.2) is called a *polynomial of degree n*.

(Note that $D(G)$ is an algebra of functions in $C(M_1(E))$ that separates points and is therefore dense by the Stone-Weierstrass theorem.)

Note that since the resampling rate $\gamma(t)$ is not assumed to be constant we have a time-inhomogeneous Markov process. In this case we can verify that for each $f \in C_{sym}(E^n) \cap D(A^{(n)})$,

$$(8.3) \quad M_f(t) := F_f(X_t) - \int_0^t G_t F_f(X_s) ds \quad \text{is a martingale}$$

where

$$(8.4) \quad G_s F_f(\mu) = \left\langle A^{(n)} f, \mu^n \right\rangle + \sum_{1 \leq i < j \leq n} \gamma(s) (\left\langle \Phi_{ij}^{(n)} f, \mu^{n-1} \right\rangle - \langle f, \mu^n \rangle), \quad 0 \leq s \leq t.$$

We thus obtain a second martingale problem formulation for the Fleming-Viot process, $\mathbb{MP}_{(D(G), \gamma\Phi)}$. It turns out that this is equivalent to the $\mathbb{MP}_{(A, \gamma Q, 0)}$.

A Function-valued Dual

For each $\mu \in M_1(E)$, let

$$(8.5) \quad H_s F(f, \mu) = \left\langle A^{(n)} f, \mu^n \right\rangle + \sum_{1 \leq i < j \leq n} \gamma(t-s) (\left\langle \Phi_{ij}^{(n)} f, \mu^{n-1} \right\rangle - \langle f, \mu^n \rangle) \quad 0 \leq s \leq t,$$

where we can interpret H_s as the generator of a function valued dual process

$$(8.6) \quad \{Y_s : 0 \leq s \leq t\}$$

which has jumps $f \rightarrow \Phi_{ij}^{(n)} f$ at rate $\gamma(s)$ for each pair $0 \leq i < j \leq n$, and between jumps evolves according to the semigroup $f \rightarrow S_t^{(m)} f$ if $f \in C(E^m)$.

This yields the duality relation

$$E_\mu[\langle F(f, X_t) \rangle] = E_f[F(Y_t, \mu)]$$

where E_f denotes expectation with respect to the law of the function-valued process, Y_t , starting at $Y_0 = f$.

Proof. The proof is analogous to the proof of Theorem 7.9 but modified to take into account the that we are now working with a time inhomogeneous process. The key step requires that

$$(8.7) \quad G_s F(X(s), Y(t-s)) = H_{t-s}(f(X(s), Y(t-s)))$$

which is satisfied by the choice (8.5). ■

Remark 8.1 This is a consequence of the fact that the dual process looks backwards in time.

8.2.2 Two Finite Particle Systems and Moment Measures

Countable exchangeable particle representations of random measures and their genealogical structures have proved to be useful in study the properties of random measures and measure-valued processes (see [111]). In the study of Fleming-Viot processes a construction (now known as the *look-down process*) of Donnelly and Kurtz ([163]) has become a standard tool in this subject. This will be described below. For simplicity, in this section we consider the homogeneous case, $\gamma(t) \equiv \gamma$.

The n-Particle Look-Down Process

We begin with a graphical construction. For each $N \in \mathbb{N}$ let $\mathcal{I}_N = \{1, \dots, N\}$ and consider a collection of independent rate γ Poisson point processes:

$$(8.8) \quad \{(N_{j,i}^{LD}(t))_{t \geq 0}\}_{1 \leq i < j \leq N}.$$

We consider N levels indexed by $i = 1, \dots, N$ and at each jump of the process $N_{j,i}$ we draw a vertical arrow from level j down to level i . At time t a path from i back to j at time s is specified by $n \in \mathbb{N}$, a sequence $i = i_n \geq i_{n-1} > \dots > i_1 > i_0 = j$ and times $s \leq u_1 \leq u_2 \leq \dots \leq u_n \leq t$ where the last jump (back in time) from level i_k was to level i_{k-1} , and occurred at time u_k .

Given (i, t) there is a unique ancestor at time $0 \leq s < t$

$$(8.9) \quad A_s(i, t) \in \mathcal{I}_N, \quad 0 \leq s \leq t$$

such that there is a path from $(A_s(i, t), s)$ to (i, t) .

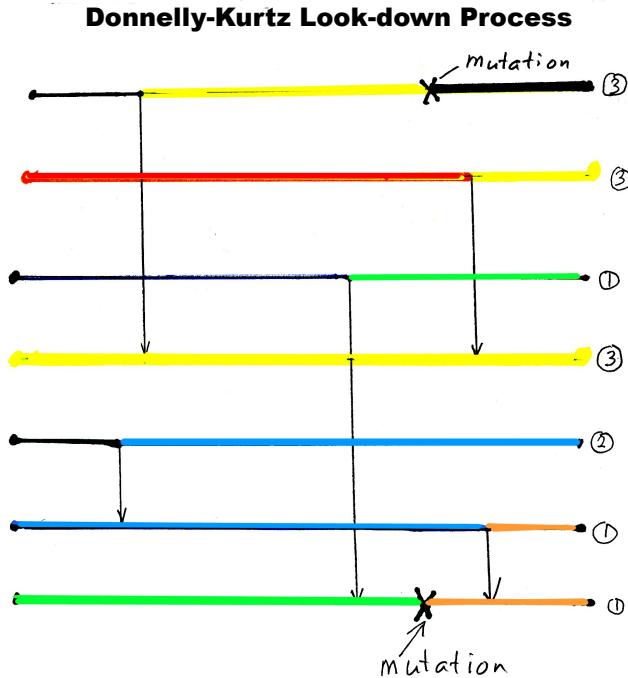


Figure 8.1: Graphical construction of the look-down process

We can then define a pseudometric on \mathcal{I}_N (as in [265])

$$(8.10) \quad d_t(i, j) := \begin{cases} 2(t - \sup\{s \in [0, t] : A_s(i, t) = A_s(j, t)\}) & \text{if } A_0(i, t) = A_0(j, t), \\ 2t + r_0(A_0(i, t), A_0(j, t)), & \text{if } A_0(i, t) \neq A_0(j, t). \end{cases}$$

This induces as usual a metric space by passing to equivalence classes.

At time t we can decompose $\{1, \dots, N\}$ into equivalence classes where two points i, j belong to the same class if $A_0(i, t) = A_0(j, t)$. Each equivalence class then defines a tree and the set of equivalence classes defines a forest. We then obtain a *forest-valued process* that becomes a *rooted tree-valued process* with a constant number, N , of leaves after a finite time.

Now consider a system of N particles $\zeta_1(t), \dots, \zeta_N(t)$ moving in the space E according to a Markov process with generator

$$(8.11) \quad \begin{aligned} \mathcal{C}^N f(x_1, \dots, x_N) \\ = A^{(N)} f(x_1, \dots, x_N) + \sum_{1 \leq i < j \leq N} \gamma \{f(\theta_{ij}(x_1, \dots, x_N)) - f(x_1, \dots, x_N)\} \end{aligned}$$

and where $\theta_{ij} : E^N \rightarrow E^N$ is defined by $\theta_{ij}(x_1, \dots, x_N)$, $i < j$, is the element of E^N obtained from (x_1, \dots, x_N) by replacing the j th component by the i th.

\mathcal{C}_t^N can be identified as the generator of an N -particle system (the finite Donnelly-Kurtz look-down process) in which the dynamics is as follows:

- at rate γ the particle with label j makes a jump to the location in E of particle i (with $i < j$)
- between jumps of the previous type the particles perform independent copies of the mutation process.

The Moran n-Particle Process

The second n-particle process in E , i.e. E^n -valued process, $Y^{(n)}$, has generator

$$\begin{aligned} \tilde{\mathcal{C}}^n f(x_1, \dots, x_n) \\ = \frac{1}{2} \sum_{i \neq j} \gamma \{f(\theta_{ij}(x_1, \dots, x_n)) - f(x_1, \dots, x_n)\} + A^{(n)} f(x_1, \dots, x_n). \end{aligned}$$

and define

$$\eta_t^{(n)} \equiv \frac{1}{n} \sum_{i=1}^n \delta_{Y_i^{(n)}(t)}.$$

This is the continuous time *Moran model* for a population of size n . (We will see below that the sequence $\{\eta_t^{(n)} : t \geq 0\}$ is tight and that every limit point satisfies the Fleming-Viot martingale problem.)

Lemma 8.2 *The empirical measure processes*

$$\eta_t^{(n)} \equiv \frac{1}{n} \sum_{i=1}^n \delta_{Y_i^{(n)}(t)} \stackrel{\mathcal{L}}{=} Z_t^{(n)} := \frac{1}{n} \sum_{i=1}^n \delta_{\zeta_i^{(n)}(t)}$$

and the resulting measure-valued process is Markov.

Proof. To verify this note that both satisfy the same martingale problem given by $G|\{F_f : f \in C_{sym}(E^n)\}$. To verify this it suffices to check that the generators C^n and \tilde{C}^n agree on symmetric functions. But if $f(x_1, \dots, x_n)$ is symmetric, then

$$\begin{aligned} C^n f(x_1, \dots, x_n) &= \sum_{1 \leq i < j \leq n} \gamma \{f(\theta_{ij}(x_1, \dots, x_n)) - f(x_1, \dots, x_n)\} + A^{(n)} f(x_1, \dots, x_n) \\ &= \frac{1}{2} \sum_{i \neq j} \gamma \{f(\theta_{ij}(x_1, \dots, x_n)) - f(x_1, \dots, x_n)\} + A^{(n)} f(x_1, \dots, x_n) \\ &= \tilde{C}^n f(x_1, \dots, x_n). \end{aligned}$$

It then that it suffices to show the solutions to these two martingale problems both have the same one dimensional marginals. But this follows since they have the same moment measures

$$E_\mu \left[\int \cdots \int f(x_1, \dots, x_k) \prod_{i=1}^k \eta_t^{(n)}(dx_i) \right] = E_\mu \left[\int \cdots \int f(x_1, \dots, x_k) \prod_{i=1}^k X_t^{(n)}(dx_i) \right]$$

for any $k, n \in \mathbb{N}$ and that these quantities are constant for $n > k$. But the terms inside the $[.]$ is a sum of symmetric functions of (x_1, \dots, x_n) and therefore the expectations agree. ■

In other words we have two particle encodings of the same measure-valued process.

Remark 8.3 *Donnelly and Kurtz [163] show that one can also construct a coupling of $(\zeta_1(t), \dots, \zeta_n(t))$ and (Y_1, \dots, Y_n) so that $\zeta^{(n)}$ is a random permutation of $Y^{(n)}$ that is independent of $\eta_t^{(n)}$.*

Remark 8.4 *The relationship*

$$GF_f(\mu) = GF(f, \mu) = \langle \mathcal{C}^N f, \mu^N \rangle, \forall f \in D(A^{(N)}) \cap C(E^N) \quad \forall \mu \in M_1(E)$$

implies that for any solution to the Fleming-Viot martingale problem, $\{X_t\}$, and solution $\{x_1(t), x_2(t), \dots, x_N(t)\}$ of the \mathcal{C} -martingale problem and $N \in \mathbb{N}$,

(8.12)

$$\begin{aligned} & E_{\mu^N} \left[\int \cdots \int f(x_1(t), \dots, x_N(t)) \right] \\ &= \int \cdots \int E_{x_1, \dots, x_N} f(x_1(t), \dots, x_N(t)) \mu(dx_1) \dots, \mu(dx_N) \\ &= E_\mu \left[\int \cdots \int f(x_1, \dots, x_N) \prod_{i=1}^N X_t(dx_i) \right] = E_\mu(F(f, X_t)). \end{aligned}$$

(The proof of this uses the fact that both can be represented by the same function-valued dual $\{Y(t)\}_{t \geq 0}$.)

Tree-valued Moran processes

The tree-valued Moran process is constructed as in Greven-Pfaffelhuber-Winter [265]. It is obtained in the as described above for the look-down tree process based on the corresponding graphical description.

As before $\mathcal{I}_N = \{1, \dots, N\}$. Let $\{N_{i,j}^M : 1 \leq i, j \leq N\}$ be a realization of a family of rate $\gamma/2$ Poisson point processes. We say that for $i, j \in \mathcal{I}_N$ and for $0 < s < t < \infty$ there is a path of descent from (i, s) to (j, t) if there exists n , $s \leq u_1 < u_2 \leq \dots < u_n \leq t$ and $i_1, \dots, i_n \in \mathcal{I}_N$ such that for all $k \in \{1, \dots, n-1\}$ (and putting $i_0 = i$ and $i_n = j$), $N_{i_{k-1}, i_k}^M[u_{k-1}; u_k] = N_{i_{k-1}, i_k}^M\{u_k\} = 1$ and $N_{m, i_{k-1}}^M[u_{k-1}, u_k] = 0$ for all $m \in \mathcal{I}_N$ as well as $N_{m, i}^M[s, u_1] = N_{m, j}^M(u_n, t] = 0$. We define $A_s(i, t)$ and the pseudometric $d_t(i, j)$ as in (8.10) and (8.9).

Let \mathbb{M} denote the set of equivalence classes of metric measure spaces (see Appendix I, Section 16.5.1). We call $\mathcal{U}^N = (\mathcal{U}_t^N)_{t \geq 0}$ the *tree-valued Moran dynamics* with population size N , where for $t \geq 0$ where $\mathcal{U}_t^N \in \mathbb{M}$ is the equivalence class of the metric measure space

$$(8.13) \quad \mathcal{U}_t^N := \overline{(\mathcal{I}, d_t^N, \frac{1}{N} \sum_{i \in \mathcal{I}} \delta_i)}.$$

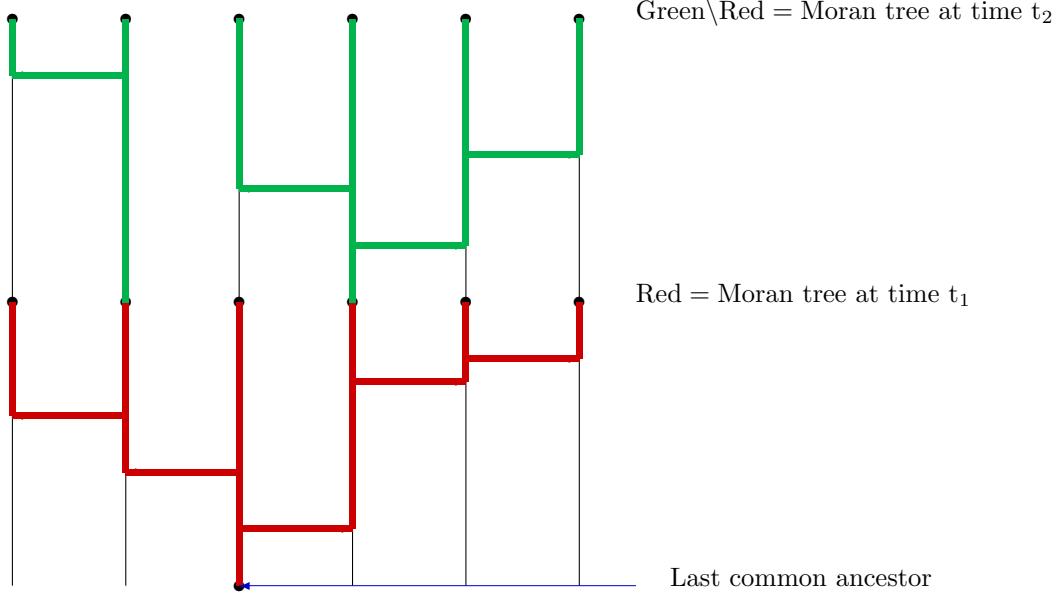


Figure 8.2: Graphical Representation of the Moran Tree at two times.

8.2.3 Extension to Infinite Particle Systems: Some Preliminaries

In this section we state for convenience some well-known results that are needed in the next section.

Theorem 8.5 (*de Finetti's Theorem*)

(a) Let $P \in M_1(M_1(E))$. Then there exists a sequence $\{Z_n\}$ of E -valued exchangeable random variables defined on a probability space $(\Omega, \mathcal{G}, P_{dF})$ with $\Omega = E^{\mathbb{N}}$, such that (Z_1, \dots, Z_n) has joint distribution

$$P^{(n)}(dx_1, \dots, dx_n) = \int_{M_1(E)} \mu(dx_1) \dots \mu(dx_n) P_{dF}(d\mu), \quad n \in \mathbb{N}$$

(b) Consider the sequence of $\{Z_n\}$ of E -valued exchangeable random variables. Let

$$X_n(\omega) = \frac{1}{n} \sum_{i=1}^n \delta_{Z_i}.$$

Then

$$X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega) \in M_1(E) \text{ exists for } P_{dF} \text{ a.e. } \omega$$

where the limit is taken in the weak topology on $M_1(E)$ and X has probability law P .

(c) Let $\mathcal{G}_n = \sigma(X_n, Z_{n+1}, Z_{n+2}, \dots)$. Then $\mathcal{G}_\infty := \cap \mathcal{G}_n$ is the σ -algebra of exchangeable events. Then X is \mathcal{G}_∞ -measurable and conditioned on \mathcal{G}_∞ , $\{Z_n\}$ is a sequence of i.i.d. random variables with marginal distribution X .

Proof. See [120], Theorem 11.2.1. ■

Corollary 8.6 *Given $f \in C(E)$, and $\varepsilon > 0$, there exists $C > 0$ and $\eta > 0$ (both depending only on ε and $\|f\|$) such that*

$$P\left(\left|\int f(x)X_n(dx) - \int f(x)X(dx)\right| \geq \varepsilon\right) \leq Ce^{-n\eta}.$$

Proof. This follows immediately from the de Finetti disintegration and Azuma's inequality (see Appendix Lemma 12.6). ■

8.2.4 The Countable Donnelly-Kurtz Look-Down Process

In this section we describe the look-down process of Donnelly and Kurtz ([163]). This process provides a representation of the Fleming-Viot process in terms of an exchangeable infinite particle system.

Note that the processes with generators $\{\mathcal{C}^n : n \in \mathbb{N}\}$ are consistent. Then taking the projective limit we obtain an infinite particle system described as follows. For any $n \in \mathbb{N}$ and $f \in D(A^{(n)})$, let

$$(8.14) \quad \begin{aligned} \mathcal{C}f(x_1, \dots, x_n) \\ = A^{(n)}f(x_1, \dots, x_n) + \sum_{1 \leq i < j \leq n} \gamma\{f(\theta_{ij}(x_1, \dots, x_n)) - f(x_1, \dots, x_n)\} \end{aligned}$$

and where $\theta_{ij} : E^n \rightarrow E^n$ is defined as above.

\mathcal{C} can be identified as the (time-dependent) generator of an ∞ -particle system, $(\zeta_1(t), \zeta_2(t), \dots)_{t \geq 0}$, (the Donnelly-Kurtz look-down process) in which the dynamics is as follows:

- at rate γ at time t the particle with label j makes a jump to the location of particle i (with $i < j$)
- between jumps of the previous type the particles perform independent copies of the mutation process with generator A .

Remark 8.7 *The relationship*

$$GF_f(\mu) = \langle \mathcal{C}^n f, \mu^n \rangle, \forall f \in D(A^{(n)}) \cap C(E^n) \quad \forall \mu \in M_1(E)$$

implies that for any solution to the Fleming-Viot martingale problem, $\{X_t\}$, and the solution $\{\zeta_1(t), \zeta_2(t), \dots\}$ of the \mathcal{C} -martingale problem and $n \in \mathbb{N}$,

$$E_{\mu^n} \left[\int \cdots \int g(\zeta_1(t), \dots, \zeta_n(t)) \right] = E_\mu \left[\int \cdots \int g(x_1, \dots, x_n) \prod_{i=1}^n X_t(dx_i) \right].$$

(The proof of this uses the fact that both can be represented by the same function-valued dual.)

Lemma 8.8 *If $(\zeta_1(0), \zeta_2(0), \dots)$ is an exchangeable sequence, then for any fixed t , $(\zeta_1(t), \zeta_2(t), \dots)$ is an exchangeable sequence,*

Proof. From the above we have

$$\int_{M_1(E)} E_\mu[\langle f, X_t^n \rangle] \nu(d\mu) = \int_{M_1(E)} E_{\mu^\infty}[f(\zeta_1(t), \dots, \zeta_n(t))] \nu(d\mu).$$

The left side is the expectation for a Fleming-Viot process with initial distribution ν and the right side is the expectation for the particle system under the assumption that $(\zeta_1(0), \zeta_2(0), \dots)$ is an exchangeable sequence with

$$P(\zeta_1(0) \in B_1, \dots, \zeta_n(0) \in B_n) = \int_{M_1(E)} \prod_{i=1}^n \mu(B_i) \nu(d\mu)$$

Then

$$P(\zeta_1(t) \in B_1, \dots, \zeta_n(t) \in B_n) = \int \int \prod_{i=1}^n X_t(B_i) P_\mu(dX) \nu(d\mu)$$

for all $t \geq 0$ where P_μ is the law of the Fleming-Viot process starting at μ and ν is the law of X_0 . Hence if $(\zeta_1(0), \zeta_2(0), \dots)$ is an exchangeable sequence, then for fixed t , $(\zeta_1(t), \zeta_2(t), \dots)$ is an exchangeable sequence. ■

Since $(\zeta_1(t), \zeta_2(t), \dots)$ is an exchangeable sequence, the corresponding de Finetti measure

$$Z_t := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \delta_{\zeta_i(t)}$$

exists a.s. in the weak topology and has the same distribution as X_t . Let $\mathcal{G}_t^n = \sigma\{Z^n(s), \zeta_{n+1}(s), \zeta_{n+2}(s), \dots\}$, $\mathcal{G}_t = \cap_n \mathcal{G}_t^n$.

It also follows from de Finetti's theorem that

$$(8.15) \quad E[f(\zeta_1(t), \dots, \zeta_k(t)) | \mathcal{G}_t] = \langle f, Z^{(k)}(t) \rangle.$$

In fact, we will show that the process $\{Z_t : t \geq 0\}$ is a version of the Fleming-Viot process. Given a convergence determining class of functions $\{f_n\}$ (with $\|f_n\| \leq 1$ for each n) on E consider the metric ρ defined by

$$\rho(\mu, \nu) = \sum_n \frac{1}{2^n} |\langle f_n, \mu \rangle - \langle f_n, \nu \rangle|.$$

Theorem 8.9 (Donnelly-Kurtz [163]) (a) Let $\zeta = (\zeta_1, \zeta_2, \dots)$ be a Markov process in E^∞ with generator C and suppose that $(\zeta_1(0), \zeta_2(0), \dots)$ is exchangeable. Then for each $t > 0$, $(\zeta_1(t), \zeta_2(t), \dots)$ is exchangeable, and the process given by the de Finetti measure

$$\begin{aligned} Z_t &= \lim_{n \rightarrow \infty} Z_t^{(n)} \\ Z_t^{(n)} &:= \frac{1}{n} \sum_{i=1}^n \delta_{\zeta_i(t)} \end{aligned}$$

is a continuous $M_1(E)$ -valued process that is a solution to the Fleming-Viot martingale problem with mutation operator A .

(b) With probability one, $Z_t^{(n)}$ converges uniformly (in t) in the weak topology on $M_1(E)$.

Proof. (a) We first note that the sequence $\{Z_t^{(n)}\}$ is tight in $D([0, \infty), M_1(E))$. Since there is a countable convergence determining class of functions in $D(G)$ it suffices to prove that for $\{F_f(Z_t^{(n)})\}$ is tight in $D([0, \infty), \mathbb{R})$, with $f \in D(A^{(n)})$ for some $n \in \mathbb{N}$. But since

$$M_f(t) := F_f(Z_t^{(m)}) - \int_0^t GF_f(Z_s^{(m)})ds$$

is a bounded martingale for $m > n$, we can verify this by applying Lemma (Appendix I, 13.15). Moreover since the jump sizes go uniformly to zero, this implies that $\{f(Z_t^{(\infty)})\}$ is continuous, a.s. (e.g. Theorem 10.2, Chapt. 3, Ethier and Kurtz [212]).

Moreover from the above we know that the moment equations are the moment equations of the Fleming-Viot process. (This can be extended to joint moments at a finite set of times $t_1 < t_2 < \dots < t_k$.) This gives the existence of a solution to the Fleming-Viot martingale problem. By the uniqueness proved above this means that that Z is a version of Fleming-Viot.

(b) By Corollary (8.6) we have

$$P \left\{ \left| \int f(x) Z_n(t, dx) - \int f(x) Z(t, dx) \right| \geq \varepsilon \right\} \leq C e^{-n\eta}$$

where C and η depend only on ε and $\|f\|$. We will show that in fact the processes $Z_t^{(n)}$ converges uniformly in t to a solution of the Fleming-Viot martingale problem. Since there is a countable convergence determining class of functions in $D(A)$ it suffices to prove this for $f(Z_t^{(n)})$, $f \in D(A)$.

Let $R_i(t, h) = 1$ if ζ_i “looks down” during the time interval $(t, t+h]$ and 0 otherwise. Note that $P(R_i(t, h) = 0) = e^{-(i-1)h}$. For $\varepsilon > 0$

$$\begin{aligned} (8.16) \quad & P \left(\sup_{t \leq s < t+h} \left| \int f(x) Z^{(n)}(s, dx) - \int f(x) Z^{(n)}(t, dx) \right| \geq \varepsilon \right) \\ &= P \left(\sup_{t \leq s < t+h} \left| \frac{1}{n} \sum_{i=1}^n f(\zeta_i(s)) - \frac{1}{n} \sum_{i=1}^n f(\zeta_i(t)) \right| \geq \varepsilon \right) \\ &\leq P \left(\sup_{t \leq s < t+h} \left| \frac{1}{n} \sum_{i=1}^n \left[(f(\zeta_i(s)) - f(\zeta_i(t)) - \int_t^s A f(\zeta_i(u)) du) \right] \right. \right. \\ &\quad \left. \left. 1_{\{R_i=0\}} \right| > \frac{\varepsilon}{4} \right) \\ &\quad + P \left((2\|f\| + h\|Af\|) \left| \frac{1}{n} \sum_{i=1}^n (R_i(t, h) - 1 + e^{-(i-1)h}) \right| > \frac{\varepsilon}{4} \right) \\ &\quad + P \left(\frac{1}{n} \sum_{i=1}^n \int_t^{t+h} |A f(X_i(u))| du > \frac{\varepsilon}{4} \right) \\ &\quad + P \left((2\|f\| + h\|Af\|) \frac{1}{n} \sum_{i=1}^n (1 - e^{-(i-1)h}) \geq \frac{\varepsilon}{4} \right) \end{aligned}$$

(The first and third terms come from $1_{\{R_i=0\}}$. The second and fourth terms comes from observing that on $1_{\{R_i=1\}}$ the increment is bounded by $(2\|f\| + h\|Af\|)$).

The independence of the R_i from the evolution of the ζ_i between look-downs implies that the process in the first term on the right is a martingale. By Doob's inequality this is less than or equal to

$$\inf_{\lambda > 0} \frac{1}{\Phi(\lambda)} E \left\{ \Phi \left(\lambda \times \frac{1}{n} \sum_{i=1}^n \left[(f(\zeta_i(t+h)) - f(\zeta_i(t)) - \int_t^{t+h} A f(\zeta_i(u)) du) \right] (1 - R_i(t, h)) \right) \right\}$$

for any convex function Φ . Then by part (b) of the large deviation Lemma 12.6 (for a sum of bounded independent zero mean r.v.'s) that there exists C and $\eta > 0$ (depending only on ε and $(\|f\| + h\|Af\|)$) such this term is bounded by $C e^{-n\eta}$.

The second term is bounded by a similar expression. The third and fourth terms are zero if $h\|Af\| < \frac{\varepsilon}{4}$ and $(2\|f\| + h\|Af\|)(1 - e^{-h(n-1)}) < \frac{\varepsilon}{4}$. Therefore C and η may be selected

depending only on ε , $\|f\|$ and $\|Af\|$ such that for h sufficiently small,

$$P \left(\sup_{t \leq s < t+h} \left| \int f(x) Z^{(n)}(s, dx) - \int f(x) Z^{(n)}(t, dx) \right| \geq \varepsilon \right) \leq C e^{-\eta n}.$$

Let $h_n \rightarrow 0$ slowly enough so that $\sum e^{-\eta n}/h_n < \infty$ for every $\eta > 0$ and fast enough so that $nh_n \rightarrow 0$ (e.g. $h_n = n^{-2}$). For $T > 0$, let $H_{T,n} = \{kh_n : k \leq T/h_n\}$. Let $f \in D(A)$, recall that $\int f(x) Z(\cdot, dx)$ is continuous and define

$$D_n = \left\{ \sup_{t \leq T} \sup_{s \leq t+h_n} \left| \int f(x) Z(s, dx) - \int f(x) Z(t, dx) \right| < \varepsilon \right\}.$$

Note that $D_n \subset D_{n+1}$, and by continuity of $\int f(x) Z(\cdot, dx)$, $P(D_n) \rightarrow 1$. For n sufficiently large (h_n sufficiently small),

$$\begin{aligned} & P \left(\left\{ \sup_{t \leq T} \left| f(x) Z_n(t, dx) - \int f(x) Z(t, dx) \right| \geq 3\varepsilon \right\} \cap D_n \right) \\ & \leq \sum_{t \in H_{T,n}} P \left(\left\{ \left| f(x) Z_n(t, dx) - \int f(x) Z(t, dx) \right| \geq \varepsilon \right\} \right) \\ & + \sum_{t \in H_{T,n}} P \left(\sup_{s \leq t+h_n} \left\{ \int f(x) Z_n(s, dx) - \int f(x) Z_n(t, dx) \geq \varepsilon \right\} \right) \\ & \leq \frac{2CT}{h_n} e^{-\eta n}. \end{aligned}$$

where we have used the Corollary to de Finetti's theorem for the first term. Summing over n , the right side converges, and Borel-Cantelli and the properties of D_n ensure that

$$\lim_{n \rightarrow \infty} \sup_{t \leq T} \left| \int f(x) Z_n(t, dx) - \int f(x) Z(t, dx) \right| = 0$$

and the proof is complete. ■

Corollary 8.10 *Let τ be a finite $\{\mathcal{G}_t\}$ stopping time. Then $\{\zeta_1(\tau), \zeta_2(\tau), \dots\}$ is exchangeable and*

$$Z(\tau) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum \delta_{\zeta_i(\tau)}.$$

Proof. If τ is discrete, then (8.15) implies the exchangeability and $Z(\tau) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum \delta_{\zeta_i(\tau)}$. For the general case, let τ_n be a decreasing sequence of $\{\mathcal{G}_t\}$ stopping times converging to τ . By the right continuity of the ζ , $(\zeta_1(\tau_n), \zeta_2(\tau_n), \dots) \rightarrow (\zeta_1(\tau), \zeta_2(\tau), \dots)$ exchangeability follows. This together with the uniform convergence gives $Z(\tau) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum \delta_{\zeta_i(\tau)}$. ■

Explicit Construction of the Look-Down Process

Theorem 8.11 *Let $\{S_t\}$ be a Feller semigroup on $C(E)$ with E compact. Then (a) for each $x \in E$ there exists a probability measure P_x on $\mathcal{B}(D_E)$ satisfying*

$$P_x(\omega(0) = x) = 1,$$

and for $s \leq t$

$$P_x(f(\omega(t)) | \sigma(\omega(u) : u \leq s) = (S_{t-s}f)(\omega(s)), P_x\text{-a.s. } \forall f \in C(E)$$

(b) There exists a standard probability space $(\Omega^A, \mathcal{F}^A, Q_A)$ and a measurable mapping $\zeta : (E \times \Omega^A, \mathcal{E} \otimes \mathcal{F}^A) \rightarrow (\mathcal{D}_E, \mathcal{B}(\mathcal{D}_E))$ such that for each $x \in E$

$$Q_A(\{\omega : \zeta(x, \omega) \in B\}) = P_x(B) \quad \forall B \in \mathcal{B}(\mathcal{D}_E)$$

Furthermore, $\zeta(\cdot, \omega)$ is continuous at x for Q_A -a.e. ω , for each $x \in E$.

Proof. (a) It is well-known that a Feller process has a càdlàg version. (This can be verified by obtaining it as the weak limit of the jump processes associated with the Yosida approximation, $A_n = A(I - \frac{1}{n}A)^{-1}$) of A .

(b) The mapping $x \rightarrow P_x$ from E to $M_1(\mathcal{D}_E)$ is continuous if the latter is given the weak topology. To see this consider first the finite dimensional distributions, $0 < t_1 < \dots < t_n$

$$E_x(f_n(\omega(t_n)) \dots f_1((t_1))) = (S_{t_1}f_1 \dots S_{t_{n-1}-t_{n-2}}(f_{n-1}S_{t_n-t_{n-1}}f_n))(x)$$

By the Feller property, $S_t f(x)$ is continuous in x for all choices of f_1, \dots, f_n . This implies that the finite dimensional distributions are continuous in x . It now suffices to show that the measures $\{P_x : x \in E\}$ are relatively compact in $D([0, \infty), E)$. To get tightness of paths in $D([0, \infty), E)$ it suffices to show tightness of $\{f(x(t)) : t \geq 0\}$ in $D([0, \infty), \mathbb{R})$ for each $f \in D(A)$. Since $f(x(t)) - \int_0^t Af(x(s))ds$ is a bounded martingale for each $f \in D(A)$, we can verify the latter condition by using the tightness lemma. Hence we get that if $x_n \rightarrow x$, then $P_{x_n} \Rightarrow P_x$. Since the mapping $x \rightarrow P_x$ is continuous, existence of a representation $(\Omega^A, \mathcal{F}, Q_A, \{\xi(x)\}_{x \in E})$, $\xi : \Omega^A \times E \rightarrow \mathcal{D}_E$ (for each $x \in E$, $\xi(\cdot, x)$ is measurable on Ω^A) follows from the extension of Skorohod's almost sure representation theorem due to Blackwell and Dubins (1983) [41]. In their representation for each $x \in E$, $\xi(\cdot, \cdot)$ is almost surely continuous at x . It remains to show that there exists a jointly measurable version. We will construct a jointly measurable function $\zeta(\cdot, \cdot)$ such that at each for each $x \in E$,

$$\zeta(\omega, x) = \xi(\omega, x) \text{ a.e. } \omega$$

that is, ζ is a version of ξ . To do this let $\{x_m : m = 1, 2, \dots\}$ be an enumeration of a countable dense set in E . Let ρ be a complete separable metric on $D([0, \infty), E)$. Consider the finite measurable partitions of $E^{(n)} = \cup_m E_m^{(n)}$ where

$$x \in E_m^{(n)} \text{ if } x_m \text{ is the closest among } \{x_m, m \leq n\} \text{ to } x$$

and in the case of ties x is assigned to the smallest such x_m . We then define the jointly measurable functions

$$\tilde{\zeta}_n(\omega, x) = \xi(\omega, x_m) \in D([0, \infty), E) \text{ if } x \in E_m^{(n)}.$$

In particular $\zeta_n(\omega, x_k) = \xi(\omega, x_k)$ for all sufficiently large n . Now define the jointly measurable function

$$\eta(\omega, x) = \lim_{n \rightarrow \infty} \max_{n', n'' \geq n} \{\rho(\tilde{\zeta}_{n'}(\omega, x), \tilde{\zeta}_{n''}(\omega, x))\}$$

and the jointly measurable function

$$\zeta(\omega, x) := 1_{\eta=0}(\omega, x) \lim_{n \rightarrow \infty} \tilde{\zeta}_n(\omega, x) + 1_{\eta>0}(\omega, x) \zeta_x^0$$

where ζ_x^0 is the constant function $\zeta_x^0 \equiv x$. Note that for each $x \in E$

$$\zeta(\omega, x) = \xi(\omega, x) \text{ } Q_A\text{-a.e. } \omega$$

since $\xi(\omega, x_m) \rightarrow \xi(\omega, x)$ if $x_m \rightarrow x$ for a.e. ω by the defining property of ξ . This means that ζ is a version of ξ . ■

Example 8.12 For Brownian motion we can simply take $\zeta(\omega, x) = x + W(\cdot)$ where $W(\cdot)$ is a standard Brownian motion starting at 0.

Theorem 8.13 Let $(\Omega^P, \mathcal{F}^P, Q_P)$ denote a probability space on which there is defined a rate 1 Poisson process, N and $(\Omega^A, \mathcal{F}^A, Q_A)$ the probability space on which we have defined the mutation process, ζ , as above. Given $\zeta(\mathbf{0}) = \{\zeta_1(0), \zeta_2(0), \dots\}$ there exists a measurable process, $\{\zeta_i(t) : i \in \mathbb{N}\}$ on the probability space

$$\Omega^{LD} := ((\Omega^A, \mathcal{F}^A, Q_A)^{\mathbb{N}} \times (\Omega^A, \mathcal{F}^A, Q_A)^{\mathbb{N}^3} \times (\Omega^P, \mathcal{F}^P, Q_P)^{\mathbb{N}^2})$$

with law given by that of the look-down process started at $\zeta(0)$.

Proof. $\omega \in \Omega^{LD}$ has the form $\omega = ((\omega^1)_{i \in \mathbb{N}}, (\omega^2)_{ijk \in \mathbb{N}^3}, (\omega^3)_{ji \in \mathbb{N}^2})$. For $i \in \mathbb{N}$, let $U_{i0} = \omega_i^1$, for $1 \leq j < i < \infty$ and $k \geq 1$ let $U_{jik} = \omega_{jik}^2$ and for $1 \leq j < i < \infty$, let $N_{ji} = \omega_{ji}^3$. Thus the $\{U_{jik}, U_{i0}\}$ are independent copies of the mutation process and $\{N_{ij}\}$ are independent Poisson processes. Put $N_i = \sum_{j:j < i} N_{ji}$. The dynamics of the system $\{\zeta_i(\cdot)\}_{i \in \mathbb{N}}$ is as follows.

- Until the first jump in N_i , ζ_i evolves according to $U_{i0}(\zeta_i(0), \cdot)$.
- If the k th jump of N_{ji} , occurs at time τ_{ji}^k , then ζ_i assumes the value of ζ_j at time τ_{ji}^k and then evolves according to $U_{jik}(\zeta_i(\tau_{ji}^k), \cdot)$ until the next jump of N_i .

It is then easy to verify that the system $\{\zeta_i(\cdot)\}_{i \in \mathbb{N}}$ is a version of the look-down process. ■

Remark 8.14 The process satisfies the following system of stochastic integral equations. For any $f \in C(E)$,

$$\begin{aligned} f(\zeta_i(t)) &= f(U_{i,0}(\zeta_i(0), t)) \mathbf{1}(\sum_{j < i} N_{ji}(t) = 0) \\ &+ \sum_{j=0}^{i-1} \int_0^t f(U_{j,i,N_{ji}(s)}((\zeta_j(s), t-s))) - f(\zeta_i(s-)) dN_{ji}(s). \end{aligned}$$

8.2.5 Genealogy and the Kingman Coalescent

In this section we describe the embedding of the genealogical tree in the countable particle system.

If the mutation process is stationary, then we can consider the look-down process on the time interval $(-\infty, \infty)$ and assume that $\{\zeta_1(t) : -\infty < t < \infty\}$ is stationary. In this stationary case we can trace the ancestry of a particle by following the process backward in time. For $s < t$ we define $a_j(s, t)$ to be the level of the ancestor at time s of the j th level particle at time t . To be precise, for $s < t$ let $N_j(s, t] = \sum_{i:i < j} N_{ij}(s, t]$. Define $\gamma_j(t) = \sup\{u < t : N_j(u, t] > 0\}$ (last time before t that the particle j passed on its type to a particle at a lower level). Let $\alpha_j(\gamma_j(t))$ be the index i such that $\gamma_j(t) \in N_{ij}$, that is, the index of . Define $a_j(s, t) = j$ for $\gamma_j(t) \leq s < t$ and $a_j(s, t) = \alpha_j(\gamma_j(t))$ for $\gamma_{\alpha_j(\gamma_j(t))}(\gamma_j(t)) \leq s < \gamma_j(t)$ and extend the definition of $a_j(s, t)$ to all $s < t$ in the obvious way.

For $0 < s < t < \infty$, let $\Gamma_n(s, t) := \{a_j(s, t) : j = 1, \dots, n\}$, that is, $\Gamma_n(s, t)$ is the set of indices of particles at time s that have descendants among the 1st n particles at time t . Then letting $N_{t,n}(s) := |\Gamma_n(s, t)|$ to denote the cardinality of $\Gamma_n(s, t)$ it follows that $D_n(u) =$

$|\Gamma_n((t-u)-, t)|$ is a pure death process with transition intensity $\binom{k}{2}$ from state k and $D_n(0) = n$. Let $\tau_{n,k} = \inf\{u \geq 0 : D_n = k\}$. Then

$$E[\tau_{n,k}] = \sum_{m=k+1}^n \frac{1}{\binom{m}{2}} = \frac{2}{k} - \frac{2}{n}$$

converges as $n \rightarrow \infty$. Then $D(u) = \lim_{n \rightarrow \infty} D_n(u) < \infty$ for all $u > 0$.

Let $N_{t,n}(s)$ denote the number of distinct individuals (ancestors) at time s that have descendants among $(\zeta_1(t), \dots, \zeta_n(t))$ and let $\Gamma_n(s, t)$ denote the collection of indices of these $N_{t,n}(s)$ particles. Since $N_{t,n}(s)$ is monotone increasing in s , we can associate a binary branching Feller process in a natural way. For $u > 0$ let $R^n(u)$ denote the equivalence relation of $\{1, \dots, n\}$ where i and j are in the same equivalence class iff they have the same ancestors at time $t-u$, that is, $i \sim j$ if $a_i((t-u)-, t) = a_j((t-u)-, t)$. Let \mathfrak{C}^n denote the set of equivalence classes on $\{1, \dots, n\}$. Let $D_{t,n}(u) := N_{t,n}(t-u)-$, $0 \leq u \leq t$, the number of equivalence classes in $R^n(u)$ (we take right continuous versions of all processes.)

Theorem 8.15 (Kingman) (a) The \mathbb{N} -valued process $\{D_{t,n}(u) : n \geq 0\}$ is a pure death process with death rates $d = \binom{k}{2}$ and $N_{t,n}(t-0) = n$.

(b) Let $N_t(s) := \lim_{n \rightarrow \infty} N_{t,n}(s)$ denote the number of distinct ancestors at time s of the infinite set of particles $\{\zeta_1(t), \zeta_2(t), \dots\}$. Then for $s < t$, $N_t(s) < \infty$, a.s.

(c) Let $D_t(u) := N_t((t-u)-)$. Then $\{D_t(u) : u > 0\}$ is a Markov pure death process started from an entrance boundary at ∞ with death rates $d_k = \binom{k}{2}$.

Proof. (a) The times between jumps in the n -particle look-down process are i.i.d. exponential r.v.'s with mean $\frac{2}{n(n-1)}$. Therefore the time since the last look-down is exponential with mean $\frac{2}{n(n-1)}$. To obtain the second-to-last look-down time we then consider the resulting $(n-1)$ -particle system and the distribution of its last look-down is exponential with mean $\frac{2}{(n-1)(n-2)}$. Continuing in this way we get that the time between the $(k-1)$ st last look-down and k th last look-down is an exponential r.v. with mean $\frac{2}{(n-k+1)(n-k)}$. Since the times between these look-downs are also independent we conclude that $\{D_{t,n}(s) : s \geq 0\}$ is a pure death process with death rates $d_k = \frac{k(k-1)}{2}$.

(b) Let $\tau_{n,k} := \inf\{s : D_{t,n}(s) = k\}$. Then $\tau_{n,k} = 2E_1/(n(n-1)) + \tau_{n-1,k}$ where E_1 and $\tau_{n-1,k}$ are independent r.v. and E_1 is exponential with mean 1. From this we obtain the representation $\tau_{n,k} = 2E_1/(n(n-1)) + \dots + 2E_{n-k}/(k+1)k$ where $\{E_m\}$ are iid $\text{Exp}(1)$ r.v.'s. Since $\sum_{j=k}^{\infty} \frac{1}{j(j+1)} < \infty$, we conclude that $\lim_{s \rightarrow \infty} \tau_{n,k} < \infty$ a.s. Consequently, $N_t(s) < \infty$ a.s. if $s < t$.

(c) This follows from (a), the consistency of the processes $\{N_{t,n}(\cdot) : n \in \mathbb{N}\}$ and the construction of N_t as the projective limit of the $\{N_{t,n}(\cdot) : n \in \mathbb{N}\}$. ■

Note that the processes $\{R^n(u)\}_{n \in \mathbb{N}}$ are consistent and we can take the projective limit, $\{R(u)\}$. This process can be described as follows. Let $\Gamma(s, t)$ be the collection of indices of particles at time $s < t$ that have a descendent at time t in the infinite look-down processes. By the last Theorem (b) $\Gamma(s, t)$ is a.s. finite and is therefore associated with an equivalence relation on \mathbb{N} having a finite number of equivalence classes which we denote by $R(t-s)$. In other words, $R(u)$ is the equivalence relation on \mathbb{N} in which i and j belong to the same equivalence class iff they have the same ancestors at time $t-u$. Let $\mathfrak{C} \subset 2^{\mathbb{N} \times \mathbb{N}}$ denote the set of equivalence

relations on \mathbb{N} with the subspace topology when $2^{\mathbb{N} \times \mathbb{N}}$ is given the product topology. Then \mathfrak{C} is a compact metrizable space. A probability measure on \mathfrak{C} is called exchangeable if it is invariant under permutations on \mathbb{N} .

From the limiting argument above we conclude that $R(s)$ is a \mathfrak{C} -valued continuous time Markov chain called *Kingman's coalescent* which is characterized by the property that its restriction to $\{1, \dots, n\}$ is the coalescent described above.

We will now consider the genealogical development in “forward time”. Note that $t - \tau_1 = \sup\{s : \text{all particles at time } t \text{ have a common ancestor at time } s\}$. Define $\bar{D}(s) = D((\tau_1 - s)-)$, $\bar{R}(s) = R((\tau_1 - s)-)$. Then \bar{D} is a pure birth process with $\bar{D}(0) = 2$ and birth rates $\frac{k(k-1)}{2}$ and a.s. finite explosion time $\hat{\tau}_\infty := \lim_{k \rightarrow \infty} \hat{\tau}_k$ where $\hat{\tau}_k := \inf\{s : \bar{D}(s) = k\}$. We denote by \bar{D}^t, \bar{R}^t the corresponding processes conditioned on $\{\hat{\tau}_\infty = t\}$.

Coalescent - Reduced Look-down

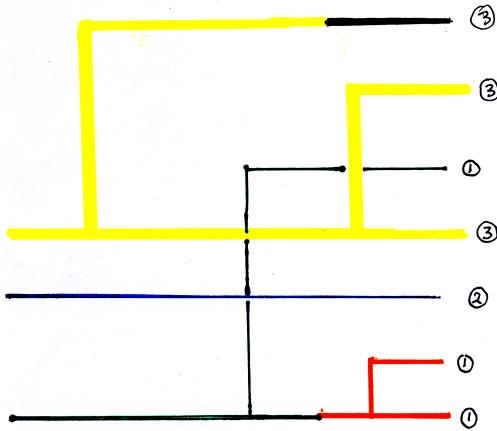


Figure 8.3: Coalescent

8.2.6 Polyá Urn Scheme

Theorem 8.16 (*Polyá Urn Representation*) (a) For each j , the limit

$$\tilde{M}_j(s) = \lim_{u \uparrow t} \frac{N_t(s, u, j)}{\sum_k N_t(s, u, k)}, \quad j = 1, \dots, N_t(s)$$

exists a.s. where $N_t(s, u, j)$ is the number of atoms at time u with common ancestor j at time s .

(b) At each time s , the random vector $(\tilde{M}_1(s), \dots, \tilde{M}_{N_t(s)}(s))$ is uniformly distributed over the simplex $\Delta_{D(s)-1}$, that is, it is distributed as the Dirichlet $D(1, \dots, 1)$.

Proof. (a) For fixed $s > 0$ we consider an urn model involving $N_t(s)$ types. At each jump time, u , $N_t(\cdot)$, $s < u < t$, a particle of type j is added with probability

$$\tilde{M}_j(s) = \lim_{u \uparrow t} \frac{N_t(s, u, j)}{\sum_k N_t(s, u, k)}, \quad j = 1, \dots, N_t(s)$$

that is, it is an $N_t(s)$ -type Polyá urn. It follows from the results of Blackwell and Kendall (1964) [40] on the Martin boundary of the Polyá urn process (and the fact that we have conditioned on $\{\tau = t\}$) that the vector $(\tilde{M}_j(s), \dots, \tilde{M}_{N_t(s)}(s))$ is uniformly distributed over the simplex $\Delta_{N_t(s)-1}$. Moreover the same result implies that at the time of a split, the mass $\tilde{M}_j(u)$ is divided into two equivalence classes of masses $U\tilde{M}_j(u)$ and $(1-U)\tilde{M}_j(u)$ where U is an independent $U[0, 1]$ r.v. Recall that if an m -type Polya urn is started with n_j initial particles of type j , then the joint distribution of the limiting proportions (x_1, \dots, x_m) in Δ_{m-1} is distributed via the Dirichlet $D(n_1, \dots, n_m)$. Using this we will show that if a split occurs at time s then

$$P(j\text{th class splits} | \tilde{M}_1(\tau_k-), \dots, \tilde{M}_{k-1}(\tau_k-)) = \tilde{M}_j(\tau_k-), \quad j = 1, \dots, k-1.$$

Taking advantage of the fact that merging types in a Polya urn yields a new Pólya urn, it suffices to prove this for a two type urn in which case we obtain

$$f_{m,n}(x, 1-x) = \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)} x^{m-1} (1-x)^{n-1}, \quad x \in [0, 1]$$

Then

$$\begin{aligned} P(1\text{st class splits} | \tilde{M}_1 = x, \tilde{M}_2 = 1-x) \\ = \frac{\frac{m}{n+m} f_{m+1,n}(x, 1-x)}{\frac{m}{n+m} f_{m+1,n}(x, 1-x) + \frac{n}{n+m} f_{m,n+1}(x, 1-x)} = x. \end{aligned}$$

Finally it is clear from the above construction that the $\{\hat{M}_j(\tau_k-) : j = 1, \dots, k-1; k \in \mathbb{N}\}$ is independent of $\{\tilde{D}^t(u) : u > 0\}$. This completes the proof that the Pólya urn scheme and coalescent yield the same probabilistic mechanism, that is,

$$Law(\tilde{M}_1(s), \dots, \tilde{M}_{N_t(s)} : s < t) = Law(M_1(s), \dots, M_{N_t(s)} : s < t).$$

(b) follows from the result of Blackwell and Kendall [40]. ■

Corollary 8.17 (a) At the time of a split, one equivalence class of mass M_C is split into two equivalence classes of masses, $M_{C'} = UM_C$ and $M_{C''} = (1-U)M_C$ where U is an independent uniform $(0, 1)$ random variable.

(b) For each C , the probability that C splits is given by M_C .

Proof. Kingman (1982) [368]. ■

8.3 Dynamics of population structure and history

8.3.1 The tree-valued Fleming-Viot process

Recently, Greven-Pfaffelhuber-Winter (2008) [265] have identified the analogue of the Fleming-Viot limit of Moran processes in the enriched framework of Moran tree processes (without mutation). This required some new concepts and techniques, in particular the notion of metric measure space and the Gromov-weak topology (see Appendix I, 13.5).

Let \mathbb{M} denote the set of equivalence classes of metric measure space with the Gromov-weak topology. Let \mathbb{M}_c denote the subset of compact metric measure spaces.

The resulting *Greven-Pfaffelhuber-Winter tree-valued Fleming-Viot process* has as state space space the set \mathbb{U} of ultrametric measure spaces furnished with the Gromov-weak topology. The process is characterized by a martingale problem analogous to the Fleming-Viot martingale

problem and again the proof of uniqueness is based on duality. We briefly outline some of the main ingredients.

We first define a class of test functions needed to define the generator.

A function $\Phi = \Phi^{n,\phi} : \mathbb{M} \rightarrow \mathbb{R}$ is called a polynomial of degree n if n is the minimal number such that there exists a function $\phi : \mathbb{R}_+^{\binom{\mathbb{N}}{2}} \rightarrow \mathbb{R}$ but and ϕ depends only on $(r_{i,j})_{1 \leq i < j \leq n}$ such that if $\chi = \overline{(X, r, \mu)}$,

$$(8.17) \quad \begin{aligned} \Phi(\chi) &= \langle \nu^\chi, \phi \rangle := \int_{\mathbb{R}_+^{\binom{\mathbb{N}}{2}}} \nu^\chi(d\bar{r}) \phi(\bar{r}) \quad \text{where } \bar{r} := (r_{i,j})_{1 \leq i < j} \\ &= \int_{X^n} \phi(\{r(x_i, x_j)\}_{1 \leq i < j \leq n}) \mu(dx_1) \dots \mu(dx_n). \end{aligned}$$

Let $\Pi^1 := \{\Phi^{n,\phi} : n \in \mathbb{N}, \phi \in C_b(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})\}$

For $\Phi \in \Pi^1$, The define

$$(8.18) \quad \Omega^\uparrow \Phi = \Omega^{\uparrow, \text{grow}} \Phi + \Omega^{\uparrow, \text{res}} \Phi$$

$$(8.19) \quad \Omega^{\uparrow, \text{grow}} \Phi(v) := \langle \nu^v, \text{div}(\phi) \rangle,$$

$$(8.20) \quad \text{div}(\phi) := 2 \sum_{1 \leq i < j \leq n} \partial \phi / \partial r_{i,j}.$$

$$(8.21) \quad \Omega^{\uparrow, \text{res}} \Phi(v) := \frac{\gamma}{2} \sum_{1 \leq k, \ell \leq n} (\langle \nu^v, \phi \circ \theta_{k,\ell} \rangle - \langle \nu^v \rangle),$$

$$(8.22) \quad \theta_{k,\ell}((r_{i,j})_{1 \leq i < j}) := \begin{cases} r_{i,j} & i, j \neq 1 \\ r_{i \wedge k, i \vee k} & j = \ell \\ r_{j \wedge k, j \vee k} & i = \ell. \end{cases}$$

Theorem 8.18 ([265], Theorem 1) Let $\mathbf{P}_0 \in \mathcal{P}(\mathbb{U})$. The $(\mathbf{P}_0, \Omega^\uparrow, \Pi^1)$ -martingale problem has a unique solution $\mathcal{U} = (\mathcal{U}_t)_{t \geq 0}$. Moreover a.s.

- (i) \mathcal{U} has continuous sample paths in \mathbb{U}
- (ii) For all $t > 0$ $\mathcal{U}_t \in \mathbb{U}_c = \mathbb{U} \cap \mathbb{M}_c$
- (iii) For all $t > 0$, $\nu^{\mathcal{U}_t}((0, \infty)^{\binom{\mathbb{N}}{2}}) = 1$ and

$$(8.23) \quad \{t \in [0, \infty) : \nu^{\mathcal{U}_t}((0, \infty)^{\binom{\mathbb{N}}{2}}) < 1\}$$

has Lebesgue measure 0.

Finally, they establish the weak convergence of the tree-valued Moran processes to the tree-valued Fleming-Viot process.

Theorem 8.19 (Greven-Pfaffelhuber-Winter [265], Theorem 2) (*Convergence of the tree-valued Moran to Fleming-Viot dynamics*). For $N \in \mathbb{N}$, let \mathcal{U}^N be the tree-valued Moran dynamics with population size N , and let $\mathcal{U} = (\mathcal{U}_t)_{t \geq 0}$ be the tree-valued Fleming-Viot dynamics. If $\mathcal{U}_0^N \Rightarrow \mathcal{U}_0$ as $N \rightarrow \infty$, weakly in the Gromov-weak topology (see Appendix ?), then

$$(8.24) \quad \mathcal{U}_{\substack{N \rightarrow \infty}}^N \rightharpoonup \mathcal{U}$$

weakly in the Skorohod topology on $\mathcal{D}_{\mathbb{U}}([0, \infty))$.

Uniqueness is again proved by duality where the dual is a tree-valued Kingman coalescent.

8.3.2 The Historical Process

We have incorporated mutation above in the explicit construction of the look-down process in the case of a Feller mutation process. We can also allow for more general mutation processes including time inhomogeneous processes.

An important special case is the *historical process* (Dawson-Perkins (1991) [118]). In this case the mutation process is a path-valued process $(t, Y^t) := (t, Y(\cdot \wedge t)) \in \mathbb{R}_+ \times D([0, \infty), E)$. The state space is $\hat{E} = \{(t, Y^t) : t \geq 0, y \in D([0, \infty), E)\}$ with the subspace topology from $\mathbb{R}_+ \times D([0, \infty), E)$. If $y, w \in D([0, \infty), E)$ and $s \geq 0$, let

$$(y/s/w)(t) = \begin{cases} y(t) & t < s \\ w(t-s) & t \geq s \end{cases}$$

Note that \hat{E} is Polish since it is a closed subset of the Polish space $\mathbb{R}_+ \times D([0, \infty), E)$.

Definition 8.20 Let $W_t : D([0, \infty), \hat{E}) \rightarrow \hat{E}$ denote the coordinate maps and for $(s, y) \in \hat{E}$, define $\hat{P}_{s,y}$ on $D([0, \infty), \hat{E})$ with its Borel σ -algebra $\hat{\mathcal{D}}$, by

$$\hat{P}_{s,y}(W_\cdot \in A) = P_{y(s)}((s + \cdot, y/s/Y^\cdot) \in A),$$

i.e. under $\hat{P}_{s,y}$ we run ξ up to some s and then tag a copy of Y starting at $y(s)$.

Lemma 8.21 $(W, (\hat{P}_{s,y})_{(s,y) \in \hat{E}})$ is a strong Markov process with semigroup

$$\hat{P}_t : C_b(\hat{E}) \rightarrow C_b(\hat{E}).$$

Proof. See Perkins. ■

Remark 8.22 This Lemma remains true if we simply assume that the mutation process satisfies: $x \rightarrow P_x$ from E to $M_1(D_E)$ is continuous. This was proved above for a Feller process.

If we assume condition

$$(8.25) \quad (s, y) \rightarrow \hat{P}_{s,y} \text{ from } \hat{E} \text{ to } M_1(D([0, \infty), \hat{E}) \text{ is continuous}$$

then we may carry out the explicit construction of the look-down process exactly as above. The resulting process is the *historical look-down process*.

Remark 8.23 The historical process can be defined even if the mutation process is non-Markovian. For example we can consider the case in which the mutation process is a fractional Brownian motion. In this case we modify the above definition as follows:

$$\hat{P}_{s,y}(W_\cdot \in A) = P((s + \cdot, y/s/Y^\cdot) \in A | Y(u) = y(u), u \leq s),$$

8.4 Some Applications of the Coalescent and Look-Down Process

8.4.1 Ergodicity for the Fleming-Viot Process

As a first application of the look-down process we establish ergodicity for the Fleming-Viot process under the assumption that the mutation process is ergodic.

Theorem 8.24 *Assume that $\theta > 0$ and the mutation process has a unique stationary distribution $\pi(dx)$ on E . Then the Fleming-Viot process, X_t has a unique stationary distribution on $M_1(E)$.*

Proof. Note that $\zeta_1(t) = U_{10}(\zeta_1(0), t)$ for all $t \geq 0$. Therefore if the mutation process has a stationary distribution, then we can assume that $\zeta_1(t)$ is defined for $-\infty < t < \infty$ and otherwise define the process as above on $(-\infty, \infty)$. It is then easy to verify iteratively that $\{\zeta_1(t), \dots, \zeta_n(t)\}$, $n = 1, 2, \dots$ are stationary. We will show below that w.p.1 all particles have a common ancestor at a finite time in the past. It is then easy to check that the resulting E^∞ -valued process will be stationary, as will the Fleming-Viot process. ■

Remark 8.25 *For the infinitely many alleles model it can be shown (cf. Ethier) that*

$$\|P_{\mu,t}(\cdot) - \Pi_\theta\| \leq 1 - P(D(t) = 0)$$

where $P_{\mu,t}$ is the law at time t of the infinitely many alleles model with parameter θ at time t and $\|\cdot\|$ is the total variation norm, and $D(t)$ is the pure death process starting at ∞ with death rates

$$d_k = \frac{1}{2}k(k-1+\theta), \quad k \geq 0.$$

8.4.2 Atomic Structure of the Infinitely Many Alleles Model

Recall that in this case the mutation operator is bounded. Note if there is no mutation, then at time t the infinite collection of particle have only finitely many ancestors at time zero and therefore only finitely many types (i.e. a finite set of atoms). In the case of a bounded mutation rate, at most countably many mutations occur (finitely many along each $\zeta_i(s)$, $0 \leq s \leq t$). Therefore there are at most countably many types (or countably many atoms). To show that there are actually countably many types note that the mutations occur according to independent Poisson processes for the countably many particles in the representation. Now consider the sequence of time of jumps in the Kingman coalescent process D_n . Note that

$$\int_0^t D(u)du = \infty \text{ w.p.1}$$

Therefore the Poisson processes running along the genealogical tree ending at time t has infinitely many jumps. Hence there are infinitely many types present at a fixed time t w.p.1.

Schmuland's Theorem

For an infinite-alleles model, with probability 1, there will be times at which the Fleming-Viot measure consists of a single atom iff $\theta < 1$. If $\theta \geq 1$, then there will always be an infinite number of atoms.

Proof. This was first proved by Schmuland [504] using Dirichlet forms. (Here we give a proof due to Donnelly and Kurtz.)

Let $S_1(t)$ denote the size of the largest atom in $Z(t)$ and define $\tau_1 = \inf\{t : S_1(t) = 1\}$. Define recursively, $\alpha_1 = \inf\{t : S_1(t) \geq \frac{3}{4}\}$, $\beta_k = \inf\{t > \alpha_k : S_1(t) \leq \frac{1}{2}\}$ and $\alpha_{k+1} = \inf\{t > \beta_k : S_1(t) \geq \frac{3}{4}\}$. Fix a time interval $[\alpha_k, \beta_k]$ and define $\tilde{S}(t) = S_1(\alpha_k + t)$. By the strong Markov property, we can let g be the indicator of the location of the largest atom at time α_k . Noting that this location does not change during the time interval $[\alpha_k, \beta_k]$, therefore (by the martingale problem) \tilde{S} satisfies

$$\tilde{S}(t) = S_1(\alpha_k) + \int_0^t \sqrt{\tilde{S}(s) - \tilde{S}^2(s)} dW_k(s) - \int_0^t \frac{\theta}{2} \tilde{S}(s) ds, \quad t < \beta_k - \alpha_k$$

for some standard Brownian motion W_k . This corresponds to a Wright-Fisher diffusion with generator

$$G_\theta f(x) = \frac{1}{2}x(1-x)f''(x) - \frac{\theta}{2}xf'(x)$$

This process has “speed measure” given by

$$\begin{aligned} m(x) &= \frac{2}{x(1-x)} e^{\int_x^1 \frac{\theta y}{y(1-y)} dy} \\ &= \frac{2}{x(1-x)} \frac{1}{(1-x)^\theta}. \end{aligned}$$

Recall that a boundary point 1 of a diffusion in natural scale is accessible (cf. Breiman) iff

$$\int^1 (1-x)m(x)dx < \infty.$$

Noting that for the Wright-Fisher diffusion the scale function, $s(x) \sim \text{const} \cdot x$ near 1, this becomes

$$\int_{\frac{1}{2}}^1 \frac{2}{x} \frac{1}{(1-x)^\theta} dx < \infty,$$

that is, $\theta < 1$ (note that the scale function, $s(x) \sim \text{const} \cdot x$ near 1. That is

$$P(\tilde{S}(t) = 1 \text{ for some } s \in [\alpha_k, \beta_k]) > 0$$

iff $\theta < 1$. S_1 can reach 1 iff $\theta < 1$. If $\theta < 1$, a renewal argument shows that $\tau_1 < \infty$, a.s.. If $\theta \geq 1$, $\tau_1 > \beta_k$ for every k . Since $(\beta_k - \alpha_k)$ is iid by the strong Markov property for S_1 , we must have $\lim_{k \rightarrow \infty} \beta_k = \infty$ and hence $\tau_1 = \infty$.

The proof that there are infinitely many atoms if $\theta \geq 1$ proceeds by considering the sum of the sizes of the two largest atoms and showing that there will always be at least three, etc. ■

Remark 8.26 Remark 8.27 This was originally proved by B. Schmuland in the Fleming-Viot case using Dirichlet forms and calculating capacities.

8.4.3 The Infinitely Many Sites Model

The infinitely-many-sites model has $E = [0, 1]^{\mathbb{Z}_+}$ and mutation kernel

$$P(\mathbf{x}, \cdot) = \int_0^1 \delta_{(\xi, \mathbf{x})}(\cdot) d\xi.$$

Here we interpret $[0, 1]$ as the sites in a DNA string and $\mathbf{x} = (x_1, x_2, \dots, 0, 0, 0, \dots)$ where x_1, x_2, \dots denotes the sites at which mutations have occurred with x_1 denotes the latest mutation, x_2 the site of the second most recent mutation, etc. Note that we can identify this with the historical infinitely many types model by reinterpreting the jump to $x \in [0, 1]$ as the site of the most recent mutation (rather than as a label for a new type). This model is used in the analysis of large sets of DNA sequence data. In particular, given a finite population the analysis begins by identifying the sites at which at least two members of the population differ this indicating a mutation has occurred in one of their ancestors (after the most recent common ancestor at which all members of the population had identical DNA sequences).

Theorem 8.28 (Ethier-Griffiths [213], Theorem 2.3) *The infinitely many sites process $X(t)$ has a unique stationary distribution $P_{st} \in \mathcal{P}(\mathcal{P}(E))$ and is ergodic $X(t) \Rightarrow X(\infty)$.*

We can now consider a random sample of size n , that is, a point in E^n chosen according to $(X(\infty))^{\otimes n}$ and the corresponding moment measures. A site $z \in [0, 1]$ is said to be *segregating* with respect to the sample if it appears in at least one but not all of the n sequences.

Given an ordered k -tuple $(\mathbf{x}_1, \dots, \mathbf{x}_k) \in E^k$ it forms a *tree* if

- a the coordinates of \mathbf{x}_1 are distinct
- b if $i, i' \in \{1, \dots, k\}$, and $j, j' \in \mathbb{Z}_+$, and $x_{ij} = x_{i'j'}$, then $x_{i,j+\ell} = x_{i'j'+\ell}$ for all $\ell \geq 1$,
- c there exists $j_1, \dots, j_d \in \mathbb{Z}_+$ such that $x_{1,j_1} = \dots = x_{k,j_k}$, that is, they have a common ancestor

Let

$$(8.26) \quad \mathcal{P}_a^0(E) := \{\mu \in \mathcal{P}_a(E) : \mu^n(\mathcal{T}_n) = 1 \forall n \in \mathbb{N}\}.$$

Then

$$(8.27) \quad P(X(t) \in \mathcal{P}_a^0(E) \forall t > 0) = 1, \quad \text{and } P_{st}(\mathcal{P}_a^0) = 1.$$

We can classify the tree structures into equivalence classes where two trees are equivalent if they are equal after a relabeling of $[0, 1]$.

For $i, j \in \mathbb{N}$ let $T_{i,j}$ be the equivalence class of trees of the form

$$(8.28) \quad ((\mathbf{x}_0, \dots, \mathbf{x}_{i-1}, \mathbf{z}_0, \mathbf{z}_1, \dots), (\mathbf{y}_0, \dots, \mathbf{y}_{j-1}, \mathbf{z}_0, \mathbf{z}_1, \dots))$$

where the $\mathbf{x}_0, \dots, \mathbf{x}_{i-1}, \mathbf{y}_0, \dots, \mathbf{y}_{j-1}, \mathbf{z}_0, \dots$ are distinct.

Watterson (1975) obtained the distribution of the number of segregating sites as follows. Let

$$(8.29) \quad p_{i,j} = \int \mu^2(T_{i,j}) P_{st}(d\mu).$$

Then it can be shown that

$$(8.30) \quad p_{i,j} = \binom{i+j}{i} \left(\frac{\theta}{2(1+\theta)} \right)^{i+j} \frac{1}{1+\theta}.$$

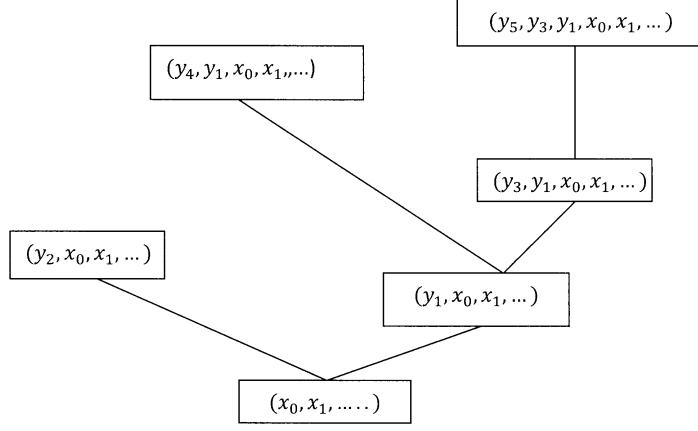


Figure 8.4: Infinitely Many Sites - Segregating sites

Let S_k denote the number of segregating sites in a random sample of size k . Then the distribution of S_2 is geometric,

$$(8.31) \quad P(S_2 = n) = \sum_{i=0}^n \binom{n}{i} \left(\frac{\theta}{2(1+\theta)}\right)^n \frac{1}{1+\theta} = \left(\frac{\theta}{1+\theta}\right)^n \frac{1}{1+\theta}.$$

Tavaré (1984) [542] proved that

$$(8.32) \quad P(S_n = s) = \frac{n-1}{\theta} \sum_{j=1}^{n-1} (-1)^{j-1} \binom{n-2}{j-1} \left(\frac{\theta}{j+\theta}\right)^{s+1}.$$

Given k sequences $\mathbf{x}_1, \dots, \mathbf{x}_k$ in $E = [0, 1]^{\mathbb{Z}^+}$ and a vector of multiplicities $\mathbf{n} = (\mathbf{n}_1, \dots, \mathbf{n}_k)$, consider the equivalence class of trees containing

$$(8.33) \quad (\mathbf{x}_1, \dots, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_k).$$

Then an analogue of the Ewens sampling formula is to determine the distribution of a finite sample taken from the stationary distribution of the infinitely many sites model. In particular given an ordered random sample of size n , what is the probability that the sample has tree structure T with multiplicities (n_1, \dots, n_k) . These questions have been studied by Griffiths (1982) [268], Wakeley (1998) [558], Ethier and Griffiths [213], and Griffiths and Tavaré [270].

8.4.4 Wandering Distributions

In 1973 Ohta and Kimura introduced the *stepwise mutation model* to describe electrophoretically detectable alleles in a population. In this case the different alleles are represented by

points on \mathbb{Z}^1 and the mutation semigroup is that of simple random walk on \mathbb{Z}^1 . In simulations they discovered the tendency for the entire population to be somewhat spread out but essentially to wander around as a loose clump. Moran investigated this in [445]. The explanation for this was given by Kingman using the coalescent [368], (see [377] for an interesting history of this and its role in the origins of the coalescent).

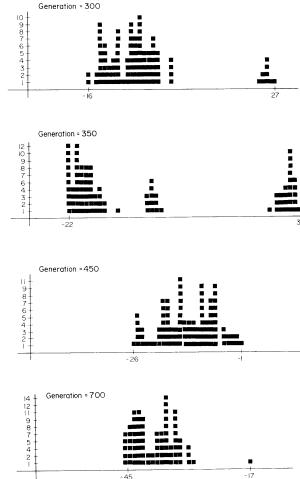


FIG. 1. Computer simulation of allelic frequency distribution for a population of size 100 evolving according to a discrete generation Ohta-Kimura ladder model with mutation probability .4.

Figure 8.5: Wandering distribution

The analogous phenomenon with continuous types was given by Dawson and Hochberg (1982) [111] using the infinite particle representation. In fact consider the particle ζ_1 in the lookdown process. Then it follows a simple random walk on \mathbb{Z}^1 . Moreover the entire population branches off from ζ_1 at a finite time in the past. It can be shown that the relative to the particle ζ_1 , the evolving population cloud approaches an equilibrium thus describing what was called the “wandering distribution”.

8.4.5 Support Properties.

Consider the Fleming-Viot process in which the mutation semigroup is the standard Brownian motion semigroup in \mathbb{R}^d . This model, in the case of \mathbb{R}^1 , arises as the limit of the stepwise mutation model on $\varepsilon\mathbb{Z}^1$ when $\varepsilon \rightarrow 0$. In the general case it serves as a model of a population described by d continuous characteristics.

Theorem 8.29 *Consider the Fleming-Viot process with Brownian motion in \mathbb{R}^d as the mutation process. Then at a fixed time (a) the closed support of X_t is compact with probability one (b) if $d > 2$, then the measure is supported on a set of Lebesgue measure zero.*

Proof. We begin by determining the distribution of the times $\tau_{\infty,n}$ in the look down process. Let $\tau_{\infty,n} := \inf\{s : |\Gamma(s,t)| \geq n\}$ and $B_n := n(t - \tau_{\infty,n})$. Then

$$B_n := n \sum_{k=n+1}^{\infty} \frac{2E_k}{k(k+1)}.$$

where $\{E_k\}$ are i.i.d. exponential one r.v.'s. Using standard exponential estimates (see Dawson and Vinogradov (1992)) it can be verified that for $0 < \nu < \frac{1}{2}$,

$$P(|B_n - 2| > n^{-\nu}) \leq 2e^{-3n^{1-2\nu}/4} + c/n^{1+\nu}$$

Therefore, by Borel-Cantelli there exists a random $n(\nu)$ such that

$$\begin{aligned} |B_n - 2| &\leq n^{-\nu} \quad \forall n \geq n(\nu), \quad P\text{-a.s. and} \\ B_n &\leq 3 \quad \forall n \geq n(\nu), \quad P\text{-a.s.} \end{aligned}$$

and

$$(t - \tau_{\infty,n}) \leq \frac{3}{n} \text{ if } n \geq n(\nu), \quad P\text{-a.s.}$$

We then observe that for the Brownian mutation process, in the time interval $(\tau_{\infty,n}, t]$ a particle has a displacement that is normally distributed with variance B_n/n (and different particles have independent displacements). Let A_{n+1} be the event that some particle with index in $\Gamma(\tau_{\infty,2^{n+1}}, t)$ is a distance greater than ε_n from its ancestor at time τ_{2^n} . Then $P(A_{n+1}) \leq c2^{n+1}e^{-\frac{2^n \varepsilon_n^2}{3}}$ for some constant c . Taking $\varepsilon_n = 2^{-n/(2+\eta)}$ with $0 < \eta < \frac{1}{2}$, we get

$$\begin{aligned} &\sum_{n=1}^{\infty} P(A_{n+1}) \\ &\leq \sum_{n=1}^{\infty} c2^{n+1}e^{-\frac{2^{n\eta}}{3}} \\ &< \infty. \end{aligned}$$

Then by the Borel-Cantelli only finitely many of the A_n occur.

$$\text{Let } \delta_n = \sum_{k \geq n} \varepsilon_k = \frac{2^{-n/(2+\eta)}}{1 - 2^{-1/(2+\eta)}}.$$

(a) Then for all n sufficiently large,

$$\text{supp}(X_t) \subset \bigcup_{j \in \Gamma(\tau_{\infty,2^n}, t)} B(\delta_n, \zeta_j(\tau_{\infty,2^n})).$$

where $B(\delta, \zeta)$ denotes a ball of radius δ centered at ζ . This implies that $\text{supp}(X_t)$ is contained in bounded subset of \mathbb{R}^d and is therefore compact.

(b) This implies that $\text{supp}(X_t)$ is contained in a set of Lebesgue measure less than or equal to

$$c2^{n+1} \left(\frac{2^{-n/(2+\eta)}}{1 - 2^{-1/(2+\eta)}} \right)^d \rightarrow 0 \text{ as } n \rightarrow \infty$$

if $d > 2$. ■

8.5 Generalizations of the Kingman coalescent

8.5.1 Coalescent with time varying population size

Since constant population size is unrealistic it is important to determine to what extent the results obtained from coalescent theory are robust, that is, what happens to the ancestral structure if the population size is randomly-varying in time. Under the assumption that the individuals in the population are exchangeable and the rescaled backward population size process converges to a continuous time Markov chain Kaj and Krone [340] show that the ancestral process is a stochastic time change of the Kingman coalescent.

8.5.2 Generalized coalescent

In the Kingman coalescent the jumps correspond to the coalescence of exactly two clusters. In (1999) Pitman [486] and Sagitov [499] introduced coalescents with multiple collisions called Λ -coalescent in which many clusters can merge simultaneously into a single cluster. The relation between these Lambda-coalescents and a class of discontinuous Fleming-Viot processes was established in the 2005 paper of Birkner, Blath, Capaldo, Etheridge, Möhle, Schweinsberg and Wakolbinger [39]. Here we just briefly describe these objects.

A generalized coalescent process is a Markov process $\{\Pi(t)\}_{0 \leq t \leq T}$ with state space given by the space of partitions of \mathbb{N} and such that the law is invariant under permutations of \mathbb{N} - see [?].

Consider a Markov process with state space $\mathcal{P}([0, 1])$ and with generator

$$(8.34) \quad GF(\mu) = \int_{(0,1]} y^{-2} \Lambda(dy) \int \mu(da)(F((1-y)\mu + y\delta_a) - F(\mu))$$

where Λ is a finite measure on $[0, 1]$ with $\Lambda(\{0\}) = 0$.

Consider functions of the form $F(\mu) = \int \dots \int f(a_1, \dots, a_p) \mu(da_1) \dots \mu(da_p)$. Then the generator has the form

$$GF(\mu) = \sum_{J \subset \{1, \dots, p\}, |J| \geq 2} \beta_{p,J}^\Lambda \int \mu(da_1) \dots \mu(da_p) (f(a_1^J, \dots, a_p^J) - f(a_1, \dots, a_p))$$

$$\beta_{p,J}^\Lambda = \int_{[0,1]} y^{j-2} (1-y)^{p-j} \Lambda(dy)$$

where a_1^J, \dots, a_p^J denotes the coalescence of the $a_j \in J$. This process is called a *generalized Fleming-Viot process*.

To set up the connection between the Fleming-Viot process and the coalescent process, fix a time T and pick a sequence of individuals labelled $1, 2, 3, \dots$ independently and uniformly on $[0, 1]$. Then for $t \leq T$ let $\Pi(t)$ denote collecting together individuals having the same ancestor at time $T-t$. This results in The Λ -coalescent process. Then by Kingman's theory of exchangeable partitions, for every $t \geq 0$ each block of $\Pi(t)$ has an asymptotic frequency and the ranked sequence of these frequencies yields a Markov process called the mass-coalescent. The has the same distribution as the ranked sequence of jump sizes of the Fleming-Viot process.

Pitman [486] proved that the Λ -coalescent has proper ordered frequencies, that is,

$$(8.35) \quad \sum_i f(\pi_i) = 1 \text{ if and only if } \int_0^1 \Lambda(dx)x^{-1} = \infty.$$

Consider the case where $\nu([\varepsilon, 1])$ is regularly varying with index $-\gamma$ as $\varepsilon \rightarrow 0$. If $1 < \gamma < 2$, then the coalescent *comes down from infinity*, that is, $\Pi(t)$ with $t > 0$ has finitely many blocks.

Example 8.30 *The classical Fleming-Viot corresponds to $\Lambda = \delta_0$.*

Example 8.31 *Another special case, the Bolthausen-Sznitman coalescent is given by $\Lambda = U([0, 1])$. This has interesting connections to the random energy model of Derrida which arises in statistical physics and Neveu's CSBP. In contrast to the Kingman coalescent, the Bolthausen-Sznitman coalescent does not come down from infinity but has infinitely many clusters at all times.*

Example 8.32 If $\Lambda = \text{Beta}(2 - \alpha, \alpha)$ for $0 < \alpha < 2$, the Fleming-Viot process Y_t is associated to a time-changed α -stable continuous state measure-valued branching process X_t as follows:

$$(8.36) \quad Y_t = \frac{X_{\tau(t)}}{X_{\tau(t)}([0, 1])}.$$

In the case (8.35), Bertoin and LeGall [33], [34] also obtain the generalized Fleming-Viot process as the solution for $x \in [0, 1]$ of the stochastic equation

$$(8.37) \quad Y_t(x) = x + \int_{[0,1] \times (0,1) \times (0,1]} N(ds, du, dr) r(1_{\{u \leq Y_{s-}(x)\}} - Y_{s-}(x))$$

where N is an \mathcal{F}_t Poisson point process with intensity $dt \otimes du \otimes \nu(dr)$ with $\nu(dx) = \frac{\Lambda(dx)}{x^2}$. This has been generalized by Dawson and Li [148] to include the case in which $\Lambda(\{0\}) > 0$.

Chapter 9

Spatially Structured and Measure-valued Models

9.1 Introduction

In this chapter we consider spatially structured population systems in the context of measure-valued processes on a discrete (countable) space S_1 or more generally on a Polish space S . In addition an objective is consider the scaling limit of particle systems so that we also consider the situation of parametric families $\{S_\varepsilon\}_{\varepsilon \in (0,1]}$ or $\{S_N\}_{N \in \mathbb{N}} \subset S$. We denote by $\mathcal{N}_f(S_1)$, $\mathcal{M}_f(S)$ the space of finite counting measures, respectively, finite Borel measures on S .

9.1.1 Spatial dynamics

The spatial systems we consider involve two basic types of dynamical mechanism:

- Migration between sites usually described by a random walk in the discrete case or Markov process in the general case,
- Local interactions at a site such as reproduction, competition, etc.

In the discrete case we consider a finite or countable set of sites, S_1 , together with an irreducible continuous time Markov chain on S_1 . The Markov chain describes the migration of individuals between sites.

9.1.2 Random Walks

We often consider the special case in which the migration Markov chain is a random walk on a countable (additive) abelian group S_1 with transition kernel $p(x, y) = p(x - y)$ where $p(\cdot)$ is a symmetric probability kernel on S_1 with $p(0) = 0$. The corresponding continuous time random walk with transition rate γ is usually denoted by $q_{x,y}^\gamma = \gamma p(x - y)$.

Example 9.1 $S_1 = \mathbb{Z}^d \subset \mathbb{R}^d = S$.

Let $p(\cdot)$ is a finite range kernel on \mathbb{Z}^d which satisfies

$$(9.1) \quad \sum_{x \in \mathbb{Z}^d} x^i p(x) = 0, \text{ for } i = 1, \dots, d.$$

$$(9.2) \quad \sum_{x \in \mathbb{Z}^d} x^i x^j p(x) = \delta_{i,j} \sigma^2, \quad i, j = 1, \dots, d,$$

where $\delta_{i,j}$ is Kronecker's delta.

Proposition 9.2 (a) The random walk is transient if and only if $d \geq 3$.

(b) Consider the rescaled lattice $S_\varepsilon = \sqrt{\varepsilon} \mathbb{Z}^d$ and random walk with kernel $p_\varepsilon(x) = p(x/\sqrt{\varepsilon})$ and jump rate $\frac{\gamma}{\varepsilon}$. Then the scaling limit of the random walk is $\{\sqrt{\gamma} \sigma B_t\}_{t \geq 0}$ where B_t is a standard Brownian motion on \mathbb{R}^d .

Remark 9.3 We can also consider infinite range random walk kernels for which the scaling limit is a α -symmetric stable process, $0 < \alpha < 2$.

Example 9.4 Island Model. For the finite group $S_N = \{0, \dots, N-1\}$ (with addition modulo N) we consider the random walk

$$(9.3) \quad p(x) = \begin{cases} \frac{1}{N-1} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Example 9.5 For $N \in \mathbb{N}$ we denote by S_N the Hierarchical Lattice Ω_N^0 and the corresponding rescaled lattices Ω_N^j defined by:

$$\begin{aligned} \Omega_N^j = \{(\xi_\ell)_{\ell \in \mathbb{Z}, \ell \geq -j} : \xi_\ell \in \{0, 1, \dots, N-1\}, \\ \exists \ell_0, \xi_\ell = 0 \forall \ell \geq \ell_0.\} \end{aligned}$$

The group operation is componentwise addition modulo N .

The hierarchical distance between two points ξ and η is

$$|\xi - \eta| := \min\{k \in \mathbb{Z} : \xi_m = \eta_m \forall m \geq k\}.$$

We also introduce the metric

$$d_j(\xi, \eta) = N^{|\xi - \eta|+1} \text{ if } -j < |\xi - \eta|.$$

Then $(\Omega_N^\infty, d_\infty)$ is a totally disconnected, locally compact abelian group. (See Evans [222].)

Let $c, d, \alpha \in (0, \infty)$. The rescaled (c, d, α) -random walk on Ω_N^j has jump rates

$$(9.4) \quad q_{\xi, \eta}^{(j)} = \sum_{k=-j+1}^{\infty} \frac{c^{k-1} N^{(k-1)(1-\frac{\alpha}{d})}}{N^{k-1}} \frac{1}{N^{j-|k|}} \mathbf{1}_{B_k^j(\xi)}(\eta),$$

that is, a jump to a point in ball $B_k^j = \{\zeta \in \Omega_N^j : |\zeta - \xi| \leq k\}$ is taken at rate $\frac{c^{k-1}}{N^{k-1}}$, $k \geq -j+1$ and the point to which it jumps is chosen at random in the ball $B_k^j(\xi)$.

Proposition 9.6 (a) The $(c, d, 2)$ random walk is transient if and only if $d > 2$ or $d = 2$, $c > 1$. (Sawyer-Felsenstein [503])

(b) The scaling limit of the rescaled (c, d) -random walks on Ω_N^j as $j \rightarrow \infty$ is the Evans-Lévy process on Ω_N^∞ with parameters (c, d) . (See Evans [222].)

(c) (Evans-Fleischmann [224], Prop. 13) The Lévy process has a jointly continuous local time if $c < 1$ (this corresponds to dimension $d = 2-$).

The (c, d, α) -random walks also mimic α -stable processes in terms of the potential operators (Dawson-Gorostiza-Waklobinger) [140], [145]. The potential operator of this hierarchical random walk has a kernel of the form

$$G_N(x) = \text{const.} N^{-|x|(1-1/(\gamma+1))},$$

where $\gamma = \frac{d}{\alpha} - 1$ (hence $d > \alpha$), this can be written as

$$(9.5) \quad G_{N,\gamma}(x) = \text{const.} \rho(x)^{-(d-\alpha)}$$

where

$$\rho(x) = N^{|x|/d}.$$

$\rho(x)$ is the “Euclidean radial distance” of x from 0, so that the volume of a ball of radius ρ grows like ρ^d . Therefore the potential operator of the (c, d, α) -random walk and that for the α -stable process in \mathbb{R}^d have the same asymptotic decay.

In the next three sections we describe some of the basic spatial systems including interacting particle systems, interacting diffusions and measure-valued processes.

9.2 Branching random walk and branching Brownian motions

Branching random walks and branching diffusions have a long history. A general theory of branching Markov processes was developed in a series of three papers by Ikeda, Nagasawa and Watanabe in 1968, 1969 [308]. The application of branching random fields to genetics was introduced by Sawyer (1975) [500].

We consider a branching random walk (BRW). The dynamics are given by:

- Birth and death at rate γ :

$$\delta_x \rightarrow (k \text{ particles}) \delta_x + \dots + \delta_x \text{ w.p. } p_k, \quad \delta_x \rightarrow \emptyset \text{ w.p. } p_0,$$

$$\mathcal{G}(z) = \sum_{k=0}^{\infty} z^k p_k \quad \text{offspring distribution generating function.}$$

- Spatial random walk in S_1 with kernel $p(\cdot)$

$$\delta_x \rightarrow \delta_y \text{ with rate } p(y-x)$$

The BRW is critical, subcritical, supercritical depending on $m = \sum_k kp_k = 1, < 1, > 1$, respectively.

We can write the generator of the branching rate walk as follows: $\mathcal{D} = \{F : F(\mu) = f(\mu(\phi)) = f(\langle \phi, \mu \rangle), \phi \in \mathcal{B}_b(S_1), f \in C(\mathbb{R})\}$ and for $F \in \mathcal{D}$,

$$\begin{aligned} GF(\mu) &= \sum_x \mu(x) \sum_y p(y) [F(\mu + \delta_{x+y} - \delta_x) - F(\mu)] \\ &\quad + \kappa \sum_x \mu(x) \sum_{k=0}^{\infty} p_k [F(\mu + (k-1)\delta_x) - F(\mu)] \\ &= \sum_x \mu(x) \sum_y p(y) [f(\mu(\phi) + \phi(x+y) - \phi(x)) - f(\mu(\phi))] \\ &\quad - \kappa \sum_x \mu(x) \sum_{k=0}^{\infty} p_k [f(\mu(\phi) + (k-1)\phi(x)) - f(\mu(\phi))] \end{aligned}$$

Let $\{S_t : t \geq 0\}$ denote the semigroup acting on $\mathcal{B}_b(S_1)$ associated to the random walk. Now define the Laplace functional

$$(9.6) \quad u(t, x) = P_{\delta_x}(e^{-X_t(\phi)}).$$

Then conditioning at the first birth-death event we obtain

$$(9.7) \quad u(t, x) = (S_t e^{-\phi})(x) e^{-\kappa t} + \kappa \int_0^t e^{-\kappa s} (S_s \mathcal{G}(u(t-s, \cdot)))(x) ds.$$

Note that this is also valid if we replace the random walk by a Lévy process on a locally compact abelian group (for example Brownian motion on \mathbb{R}^d with semigroup $\{S_t : t \geq 0\}$).

Proposition 9.7 *The martingale problem for G is well posed and the Laplace functional of the solution is the unique solution of equation (9.7).*

The system of branching Brownian motions (BBM) is defined in the same way with $S = \mathbb{R}^d$ with offspring produced at the location of the parent and between branching the particles perform independent Brownian motions. (For non-local branching see Z. Li [406].)

Remark 9.8 *We sometimes combine the reproduction and spatial jump by replacing the reproduction and migration by a single mechanism in which an offspring produced by a birth immediately moves to a new location obtained by taking a jump with kernel p_ε , that is, $\delta_x \rightarrow \delta_x + \delta_y$.*

We also consider the $\mathcal{N}(S)$ -measure-valued process $\{X_t\}$ in which each particle has mass η , that is,

$$X_t(A) = \eta \sum_{i=1}^{N(t)} \delta_{x_i(t)}, \quad A \subset S$$

where $x_i(t)$ denotes the location of the i th particles at time t .

Supercritical BRW and BBM

There is an important relation between supercritical branching Brownian motions and the Fisher-KPP equation. This relation was developed by McKean [434] and Bramson [52].

A basic question concerns the geometrical properties of the supercritical branching random walk. Biggins [37] has proved that the set $\mathcal{I}^{(n)}$ of positions occupied by n th generation individuals rescaled by a factor $\frac{1}{n}$ has asymptotic shape \mathcal{I} where \mathcal{I} is a convex set.

9.3 Interacting particle systems

9.3.1 Coalescing random walks

Let S_1 be a countable abelian group and $p(x)$ a symmetric finite range kernel on S_1 , $p(0) = 0$. The state space is $\{0, 1\}^{S_1}$.

Consider a collection of particles on S_1 undergoing random walks which are independent up to a collision time. If two particles collide (occupy the same site), then the two particles instantaneously coalesce and are replaced by a single particle.

Now consider the associated a set-valued process $A(t)$ corresponding to the set of points occupied by a coalescing random walk. The state space is the collection of finite subsets of S_1 , $\{\eta \in \{0, 1\}^{S_1}, \eta(x) = 0, a.a. x\}$. Then the transitions are

$$(9.8) \quad A \rightarrow A \cup \{y\} \setminus \{x\} \text{ at rate } q_{x,y} = p(x - y).$$

Now consider the family of functions on

$$(9.9) \quad F(A, \eta) = \prod_{x \in A} \eta(x),$$

$A = \text{finite subset of } S_1, \eta \in \{0, 1\}^{S_1}$

We denote the generator by H . We observe that the generator H applied to functions of the form F (defined in (9.9)) satisfies

$$(9.10) \quad HF(\{x_1, \dots, x_n\}, \eta) \\ = \sum_{i=1}^n \sum_y q_{x_i, y} (F(\{x_1, \dots, x_n\} \setminus x_i \cup y, \eta) - F(\{x_1, \dots, x_n\}, \eta))$$

Coalescing random walks with delay

A variation of the system of coalescing walks is the coalescing random walk with delay. In this case when a pair of particles occupy the same site, then they coalesce at a random time with exponential distribution, that is, the particles coalesce with finite rate κ . These exponential random variable at different sites are independent.

An important quantity is the probability that two particles starting at the same site eventually coalesce. To determine this let q_0 be the rate at which a jump from 0 occurs, $\tau_0 = \inf\{t > 0 : Z_t \neq 0\}$ and $\tau_1 := \inf\{t > \tau_0 : Z_t = 0\}$. Then $E[\tau_0] = \frac{1}{q_0}$. Let $q_e = 1 - P(\tau_1 < \infty)$ (escape probability).

Starting two particles at 0 we compute the probability that they do not eventually coalesce, γ_e , as

$$(9.11) \quad \gamma_e = \frac{2q_0}{2q_0 + \kappa} \cdot [q_e + (1 - q_e)\gamma_e]$$

so that

$$(9.12) \quad \gamma_e = \frac{2q_0 q_e}{\kappa + 2q_0 q_e}.$$

9.3.2 Voter Model

The voter model was independently introduced by Clifford, Sudbury (1973) [74] and Holley and Liggett (1975) [283].

Again, let S_1 be a countable abelian group and $p(x)$ a symmetric finite range kernel on S_1 . The state space is $\{0, 1\}^{S_1}$. The voter model ξ_t is defined by the transitions

$$\eta \rightarrow \eta_{x,y}, x \neq y, \quad \text{at rate } q_{x,y} \text{ where}$$

$$\eta_{x,y}(z) = \eta(z) \text{ if } z \neq x$$

$$\eta_{x,y}(x) = \eta(y).$$

Acting on functions of finite support, the generator can be written

$$(9.13) \quad Gf(\eta) = \sum_x c(x, \eta)[f(\eta_x) - f(\eta)]$$

where

$$(9.14) \quad \eta_x(y) = \begin{cases} \eta(y), & \text{if } x \neq y \\ 1 - \eta(y), & \text{if } x = y \end{cases}$$

and

$$(9.15) \quad c(x, \eta) = \begin{cases} \sum_y q_{x,y} \eta(y), & \text{if } \eta(x) = 0 \\ \sum_y q_{x,y} [1 - \eta(y)] & \text{if } \eta(x) = 1 \end{cases}$$

Let $f_x(\eta) := 1(\eta(x) = 1)$. Now consider the functions

$$(9.16) \quad F(A, \eta) = \prod_{x \in A} \eta(x),$$

A is a finite subset of S , $\eta \in \{0, 1\}^S$

Noting that $(f(\eta_x) - f(\eta)) = 1 - 2f_x(\eta)$ and a change occurs at site x with rate $\sum_y q_{y,x} [f_x(\eta)(1 - f_y(\eta)) + f_y(\eta)(1 - f_x(\eta))] = \sum_y p(y, x) [f_x(\eta) + f_y(\eta) - 2f_x(\eta)f_y(\eta)]$ Then

$$(9.17) \quad GF(\{x_1, \dots, x_n\}, \eta) = G \left(\prod_{i=1}^n f_{x_i}(\eta) \right)$$

$$= \sum_{i=1}^n \left(\sum_{y \neq x_1, \dots, x_n} q_{y,x_i} (f_y - f_{x_i}) \prod_{j \neq i} f_{x_j} + \sum_{k=1}^n p(x_k, x_i) (f_{x_k} - f_{x_i}) \prod_{j \neq i} f_{x_j} \right)$$

$$= \sum_{i=1}^n \sum_y q_{y,x_i} (F(\{x_1, \dots, x_n\} \setminus x_i \cup y, \eta) - F(\{x_1, \dots, x_n\}, \eta))$$

Now consider a set valued process $A(t)$ corresponding to the set of points of a coalescing random walk. Then the transitions are

$$(9.18) \quad A \rightarrow A \cup \{y\} \setminus \{x\} \text{ at rate } q_{x,y}.$$

We denote the generator by H . We observe that the generator H applied to functions of the form $F(\cdot, \eta)$, $\eta \in \{0, 1\}^S$ (defined in (9.9)) satisfies

$$(9.19) \quad HF(\{x_1, \dots, x_n\}, \eta) = GF(\{x_1, \dots, x_n\}, \eta).$$

We then apply the dual representation Theorem 7.9.

Remark 9.9 In the case $S = \mathbb{Z}$, a direct relation between the voter model and its dual can be demonstrated by a graphical construction.

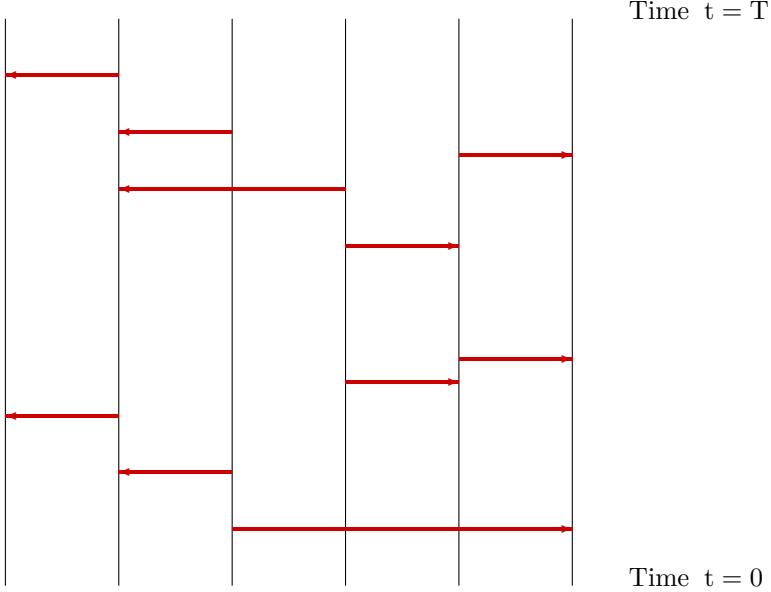


Figure 9.1: Graphical Representation of the Voter Model.

As in the case of birth and death processes we can also obtain a representation of the voter model by a stochastic integral equation (cf. Mueller-Tribe (1995) [448], Kurtz-Protter (1996) [387]) as follows.

Let $\{\Lambda_t(x, y) : x, y \in S\}$ be a family of independent Poisson processes with rates $p(y - x)$. Consider the system of stochastic integral equations

$$(9.20) \quad \xi_t(x) = \xi_0(x) + \sum_y \int_0^t [\xi_{s-}(y) - \xi_{s-}(x)] d\Lambda_s(x, y)$$

with $x \in S$, $t \geq 0$,

$$(9.21) \quad \xi_0(x) = 0, \text{ a.a.x.}$$

Proposition 9.10 *The system (9.20) has a unique solution and has pregenerator G.*

Proof. This is a special case of a result in Kurtz and Protter [387], Chap. 9. ■

A key property of the voter model is that the system started with a single non-zero site always dies out. This follows since $\sum_{x \in S} \xi_t(x)$ is a martingale.

9.3.3 Biased voter model

The biased voter model is a modification of the voter model that arose from the Williams-Bjerknes (1972) tumour growth model. Basic results on this model were obtained by Bramson and Griffeath (1980,1981) [49], [51] and Lanchier-Neuhauser (2007) [393].

It is a spin system with generator

$$(9.22) \quad Gf(\eta) = \sum_x c(x, \eta)[f(\eta_x) - f(\eta)]$$

where

$$(9.23) \quad \eta_x(y) = \begin{cases} \eta(y), & \text{if } x \neq y \\ 1 - \eta(y), & \text{if } x = y. \end{cases}$$

The biased voter model spin rates are given by:

$$(9.24) \quad c(x, \eta) = \begin{cases} \beta \sum_y q_{x,y} \eta(y), & \text{if } \eta(x) = 0 \\ \sum_y q_{x,y} [1 - \eta(y)] & \text{if } \eta(x) = 1 \end{cases}$$

where $\beta \geq 0$. If $\beta > 1$, the dual process is a coalescing branching random walk. If $\beta > 1$ the opinion 1 is favoured and growing clusters of 1's can form. In fact Bramson and Griffeath [51] prove conditioned on non-extinction that there is a growing region whose radius grows linearly and that the occupied regions has an asymptotic shape.

9.3.4 Oriented percolation

Consider the lattice $\mathbb{Z}_+ \times \mathbb{Z}^d$. Points (n_1, x_1) and (n_2, x_2) are neighbours iff $n_2 = n_1 + 1$ and x_1, x_2 are neighbours in \mathbb{Z}^d . We consider the bonds between such neighbours and designate them open with probability p and closed with probability $1 - p$. Percolation occurs by the flow through open bonds. We say that (n, x) can be reached from (m, y) if there is a sequence of open bonds joining them. We consider the cluster $C_{(0,0)}$ consisting of points that can be reached from $(0, 0)$. Also let

$$(9.25) \quad \xi_n^0 = \{x : (0, 0) \rightarrow (n, x)\}.$$

We define

$$(9.26) \quad p_c = \inf\{p : P(\xi_n^0 \neq \emptyset \forall n) > 0\}.$$

In the case $d = 1$, Liggett [410] proved that $p_c \leq \frac{2}{3}$ and it is known that $p_c \geq .6446$ (cf. Durrett (1985) [160]).

See Durrett (1984) [176] and Durrett-Tanaka (1989) [181] for the basic properties.

9.3.5 Contact process

The contact process was introduced by Mollison (1977) [443], and basic results were obtained by Griffeath (1981) [266], Durrett-Griffeath (1982) [174], and Durrett-Schonmann (1987) [177].

The contact process on $S_1 = \mathbb{Z}^d$ has transition rates:

$$(9.27) \quad c(x, \eta) = \begin{cases} \lambda \sum_{|y-x|=1} \eta(y), & \text{if } \eta(x) = 0 \\ 1 & \text{if } \eta(x) = 1 \end{cases}$$

where $\lambda > 0$.

There is a critical value λ_c of λ such that there is a positive probability that the contact process on \mathbb{Z}^d does not die out for $\lambda > \lambda_c$ and below which dies out with probability 1. See Durrett (1988) [178] for an introduction to contact process on \mathbb{Z}^d . Recent results on the contact process on hierarchical group are given in Athreya and Swart [13].

9.3.6 Interacting birth and death processes

We next consider a class of processes which have been used to model chemical reaction diffusion systems. The state space is $(\mathbb{Z}_+)^{S_1}$ and the generator has the form

$$(9.28) \quad \begin{aligned} Gf(x) = & \sum_{\xi \in S_1} \{ \lambda_1(x_\xi)[f(x + e_\xi) - f(x)] + \lambda_2(x_\xi)[f(x - e_\xi) - f(x)] \} \\ & + \sum_{\xi \neq \xi'} x_\xi q_{\xi, \xi'} [f(x - e_\xi + e_{\xi'}) - f(x)] \end{aligned}$$

where $\lambda_1(\cdot), \lambda_2(\cdot) \geq 0$. If $\lambda_1(k) = \beta_0 + \beta_1 k, \lambda_2(k) = \delta_1 k + \delta_2 k^2$, the construction and uniqueness for such systems has been established by M.-F. Chen (see for example [71]). In the hydrodynamic limit they give rise to reaction-diffusion equations.

If $\beta_0, \beta_1, \delta_1, \delta_2 > 0$, this is known as Schlögl's first model. If $\beta_0 = 0, \beta_1 > 0, -\delta_1 = \delta_2 > 0$, this corresponds to branching coalescing model (BC-model).

9.4 Interacting diffusions

We now consider processes X_t with local state space $[0, \infty)^M, M \in \mathbb{N}$ and configuration space $([0, \infty)^M)^{S_1}$ where S_1 is a countable abelian group.

- $g : [0, \infty) \rightarrow [0, \infty)$ is locally Lipschitz continuous
- $g^{-1}((0, \infty)) = (0, b)$ for some $b \in (0, \infty]$,
- $g(z) \leq C(1 + z^2)$ for some $C < \infty$.

Consider the system of stochastic differential equations

$$dX_t^{(i)}(x) = [\sum q_{x,y}(X_t^{(i)}(y) - X_t^{(i)}(x))]dt + \sqrt{g(X^{(i)}(x))} dW_t^{(i)}(x)$$

$$i = 1, \dots, M, \quad x \in S_1$$

where $\{(W_t^{(i)}(x))_{i=1, \dots, M}\}_{x \in S_1}$ are independent M -dimensional Wiener processes and $\{q_{x,y}\}_{x,y \in S_1}$ are the transition rates for a symmetric random walk on S_1 .

Assume that there exists $\{\gamma_i : i \in S_1\}$, a positive summable reference measure satisfying

$$(9.29) \quad \sum_i \gamma_i q_{ij} \leq \Gamma \gamma_j, \quad j \in S_1, \quad \text{for some constant } \Gamma.$$

The Spitzer-Liggett space is defined as $E := \{z \in [0, b]^{S_1} : \|z\| < \infty\}, \|z\| = \sum_i \gamma_i |z_i|$ with the topology of componentwise convergence.

Conditions for existence and uniqueness of solutions to these systems were obtained by Shiga and Shimizu (1980) [510].

9.4.1 Interacting Feller diffusions.

The case $M = 1, g(x) = \gamma x, \gamma > 0$, describes a system of interacting Feller CSBP processes.

9.4.2 The Wright-Fisher stepping stone model.

The special case $M = 2$, $X^1, X^2 \geq 0$, $X^1 + X^2 \equiv 1$, setting $X_t = X_t^1$, satisfying

$$\begin{aligned} dX_t(x) &= \sum_{y \in S_1} p_{y-x}(X_t(y) - X_t(x))dt \\ &\quad + sX_t(x)(1 - X_t(x))dt + \sqrt{2\gamma X_t(x)(1 - X_t(x))}dW_t(x) \\ x_0(x) &= \theta \in [0, 1] \quad \forall x \in S_1 \end{aligned}$$

with $s = 0$ describes the stepping stone diffusion approximation to the neutral stepping stone model introduced by Malécot (1948) [421], and studied by Kimura (1953) [360], Kimura-Weiss (1964) [363], Nagylaki (1974) [450], and Sawyer (1976) [501].

Remark 9.11 *The voter model corresponds to the limiting case $s = 0$, $\gamma \rightarrow \infty$. The additional term with coefficient s represents the non-neutral case in which type 1 has fitness $s \neq 0$.*

Refer to Section 10.4 for the development of this model.

9.4.3 Mean-field limit of exchangeable interacting diffusions

Consider the system of exchangeable diffusions on $S = \{0, \dots, N-1\}$

$$\begin{aligned} dx_\xi(t) &= c(x_{\xi[1]}(t) - x_\xi(t))dt + \sqrt{2g(x_\xi(t))}dw_\xi(t) \\ x_\xi(0) &= \theta_0 \quad \forall \xi \\ x_{\xi[1]}(t) &:= \frac{1}{N} \sum_{\xi=1}^N x_\xi(Nt) \quad \text{mean-field process} \end{aligned}$$

Renormalized system:

$$(9.30) \quad Z_1^N(t) = x_{\xi[1]}(Nt), \quad dZ_1^N(t) = -cZ_1^N(t)dt + \sqrt{\frac{2}{N} \sum_{\xi=1}^N g(x_\xi(Nt))} dw(t).$$

Stationary measures:

$$\Gamma_\theta^g(A) := \frac{1}{Z(g)} \int_A \frac{1}{g(x)} \exp \left[\int_\theta^x \frac{\theta - y}{g(y)} dy \right] dx$$

$$\mathcal{F}(g)(\theta) := \int g(x) \Gamma_\theta^g(dx)$$

Theorem 9.12 *As $N \rightarrow \infty$*

$$(9.31) \quad Z_1^N(\cdot) \Rightarrow Z_1(t)$$

where $Z_1(t)$ satisfies

$$dZ_1(t) = -cZ_1(t)dt + \sqrt{2g_1(Z_1(t))} dw(t).$$

where

$$g_1 = \mathcal{F}(g).$$

Proof. See Dawson-Greven (1993) [121]. ■

9.5 Measure-valued branching processes

9.5.1 Super-Brownian motion

Introduction

Super-Brownian motion (SBM) is a measure-valued branching process which generalizes the Jirina process. It was constructed by S. Watanabe (1968) [563] as a continuous state branching process and Dawson (1975) [107] in the context of SPDE. The lecture notes by Dawson (1993) [120] and Etheridge (2000) [205] provide introductions to measure-valued processes. The books of Dynkin [195], [196], Le Gall [397], Perkins [487] and Li [408] provide comprehensive developments of various aspects of measure-valued branching processes. In this section we begin with a brief introduction and then survey some aspects of superprocesses which are important for the study of stochastic population models. Section 9.5 gives a brief survey of the small scale properties of SBM and Chapter 10 deals with the large space-time scale properties.

Of special note is the discovery in recent years that super-Brownian motion arises as the scaling limit of a number of models from particle systems and statistical physics. An introduction to this class of *SBM invariance principles* is presented in Section 9.6 with emphasis on their application to the voter model and interacting Wright-Fisher diffusions. A discussion of the invariance properties of Feller CSB in the context of a renormalization group analysis is given in Chapter 11.

The SBM Martingale Problem

Let $(D(A), A)$ be the generator of a Feller process on a locally compact metric space (E, d) and $\gamma \geq 0$. The probability laws $\{P_\mu : \mu \in M_f(E)\}$ on $C([0, \infty), M_f(E))$ of the superprocess associated to (A, a, γ) can be characterized as the unique solution of the following martingale problem:

$$M_t(\varphi) := \langle \varphi, X_t \rangle - \int_0^t \langle A\varphi, X_s \rangle ds$$

is a P_μ -martingale with increasing process

$$\langle M(\varphi) \rangle_t = \int_0^t \gamma \langle \varphi, X_s \rangle ds$$

for each $\varphi \in D(A)$.

Equivalently, it solves the martingale problem

$$\begin{aligned} GF &= \int A \frac{\delta F}{\delta \mu(x)} \mu(dx) \\ &\quad + \frac{\gamma}{2} \iint \frac{\delta^2 F}{\delta \mu(x) \delta \mu(y)} \delta_x(dy) \mu(dx) \end{aligned}$$

$$D(G) := \{F(\mu) = e^{-\mu(\varphi)}, \varphi \in \mathcal{B}_+(\mathbb{R}^d)\}$$

The special case $E = [0, 1]$, $Af(x) = [\int f(y)\nu_0(dy) - f(x)]dy$ is the Jirina process. The special case $E = \mathbb{R}^d$ $A = \frac{1}{2}\Delta$ on $D(A) = C_b^2(\mathbb{R}^d)$ is called super-Brownian motion.

9.5.2 Super-Brownian Motion as the Limit of Branching Brownian Motion

Given a system of branching Brownian motions on $S = \mathbb{R}^d$ and $\varepsilon > 0$ we consider the measure-valued process, X^ε , with particle mass $m_\varepsilon = \varepsilon$ and branching rate $\gamma_\varepsilon = \frac{\gamma}{\varepsilon}$, that is,

$$(9.32) \quad X^\varepsilon(t) = m_\varepsilon \sum_{j=1}^{N(t)} \delta_{x_j(t)}$$

where $N(t)$ denotes the number of particles alive at time t and $x_1(t), \dots, x_{N(t)}$ denote the locations of the particles at time t . Given an initial set of particles, let $\mu_\varepsilon = m_\varepsilon \sum_{j=1}^{N(0)} \delta_{x_j(0)}$, let $P_{\mu_\varepsilon}^\varepsilon$ denote the probability law of X^ε on $D_{M_F(\mathbb{R}^d)}([0, \infty))$. Let $\{\mathcal{F}_t\}_{t \geq 0}$ be the canonical filtration on $D([0, \infty), M_F(\mathbb{R}^d))$.

Notation 9.13 $\mu(\phi) = \langle \phi, \mu \rangle = \int \phi d\mu$.

Let $C(M_F(\mathbb{R}^d)) \supset D(G_\varepsilon) := \{F(\mu) = f(\langle \phi, \mu \rangle) : f \in C_b^2(\mathbb{R}), \phi \in C_b^2(\mathbb{R}^d)\}$. Then $D(G_\varepsilon)$ is measure-determining on $M_F(\mathbb{R}^d)$ ([120], Lemma 3.2.5.).

Then using Itô's Lemma, it follows that $P_{\mu_\varepsilon}^\varepsilon \in \mathcal{P}(D([0, \infty), M_F(\mathbb{R}^d)))$ satisfies the G^ε -martingale problem where for $F \in D(G^\varepsilon)$,

$$\begin{aligned} G^\varepsilon F(\mu) &= f'(\mu(\phi))\mu\left(\frac{1}{2}\Delta\phi\right) + \frac{\varepsilon}{2}f''(\mu(\phi))\mu(\nabla\phi \cdot \nabla\phi) \\ &\quad + \frac{\gamma}{2\varepsilon^2} \int [f(\mu(\phi) + \varepsilon\phi(x)) + f(\mu(\phi) - \varepsilon\phi(x)) - 2f(\mu(\phi))] \mu(dx). \end{aligned}$$

We can also obtain the Laplace functional of $X^\varepsilon(t)$ using Proposition 9.7 with $\{S_t : t \geq 0\}$ the Brownian motion semigroup on $C(\mathbb{R}^d)$ and $\mathcal{G}(z) = \frac{1}{2} + \frac{1}{2}z^2$.

Theorem 9.14 Assume that $X^\varepsilon(0) = \mu_\varepsilon \Rightarrow \mu$ as $\varepsilon \rightarrow 0$.

Then
(a) $P_{\mu_\varepsilon}^\varepsilon \xrightarrow{\varepsilon \rightarrow 0} \mathbb{P}_\mu \in \mathcal{P}(C_{M_F(\mathbb{R}^d)}([0, \infty))$ and \mathbb{P}_μ is the unique solution to the martingale problem: for all $\phi \in C_b^2(\mathbb{R}^d)$,

$$(9.33) \quad M_t(\phi) := X_t(\phi) - \mu(\phi) - \int_0^t X_s\left(\frac{1}{2}\Delta\phi\right) ds$$

is an (\mathcal{F}_t^X) -martingale starting at zero with increasing process

$$\langle M(\phi) \rangle_t = \gamma \int_0^t X_s(\phi^2) ds.$$

(b) The Laplace functional of X_t is given by

$$(9.34) \quad \mathbb{P}_\mu \left(e^{-\int \phi(x) \mathbb{X}_t(dx)} \right) = e^{-\int v_t(x) \mu(dx)}.$$

where $v(t, x)$ is the unique solution of

$$(9.35) \quad \frac{\partial v(t, x)}{\partial t} = \frac{1}{2}\Delta v(t, x) - \frac{\gamma}{2}v^2(t, x), \quad v_0 = \phi \in C_{+, b}^2(\mathbb{R}^d).$$

(c) The total mass process $\{X_t(\mathbb{R}^d)\}_{t \geq 0}$ is a Feller CSBP.

Proof.

Step 1. Tightness of probability laws of X^ε on $D_{M_F(\mathbb{R}^d)}([0, \infty)$ and a.s. continuity of limit points. In order to prove tightness it suffices to prove that for $\delta > 0$ there exists a compact subset $K \subset \mathbb{R}^d$ and $0 < L < \infty$ such that

(9.36)

$$P_{\mu_\varepsilon}^\varepsilon(\sup_{0 \leq t \leq T} X_t(K^c) > \delta) < \delta, \quad P_{\mu_\varepsilon}^\varepsilon(\sup_{0 \leq t \leq T} X_t(1) > L) < \delta$$

and

(9.37) $P_{\mu_\varepsilon}^\varepsilon \circ (X_t(\phi))^{-1}$ is tight in $D_{\mathbb{R}}([0, \infty))$ for $\phi \in C_c^2(\mathbb{R}^d)$.

This can be checked by standard moment and martingale inequality arguments. For example for (9.36) it suffices to show that

(9.38) $\sup_{0 < \varepsilon \leq 1} \sup_{\delta > 0} E(\sup_{0 \leq t \leq T} \langle e^{-\delta \|x\|}(1 + \|x\|^2), \mathbf{X}^\varepsilon(t) \rangle) < \infty,$

and (9.37) can be verified using the Joffe-Métivier criterion (see Appendix, (13.4.2)). The a.s. continuity of any limit point then follows from Theorem 13.14 since the maximum jump size in X^ε is ε .

Moreover, if \mathbb{P}_μ is a limit point, it is also easy to check (cf. Lemma 12.2) that for ϕ in $C_b^2(\mathbb{R}^d)$, $M_t(\phi)$ is a \mathbb{P}_μ -martingale and $(F_1(\mu) = \mu(\phi), F_2(\mu) = \mu(\phi)^2)$

$$\begin{aligned} \langle M(\phi) \rangle_t &= \lim_{\varepsilon \rightarrow 0} \int_0^t (G_\varepsilon F_2(X_s) - 2F_1(X_s)G_\varepsilon F_1(X_s)) ds \\ &= \gamma \int_0^t X_s(\phi^2) ds. \end{aligned}$$

As pointed out above, (9.33) and Ito's formula yields an equivalent formulation of the martingale problem, namely: for $f \in C_b^2(\mathbb{R})$, $\phi \in C_b^2(\mathbb{R}^d)$, and $F(\mu) = f(\mu(\phi))$,

(9.39) $F(X_t) - \int_0^t GF(X_s) ds$ is a \mathbb{P}_μ -martingale

where

$$GF(\mu) = f'(\mu(\phi))\mu(\frac{1}{2}\Delta\phi) + \frac{\gamma}{2}f''(\mu(\phi))\mu(\phi^2).$$

Step 2. (Uniqueness) In order to prove (b) we first verify that \mathbb{P}_μ also solves the following time dependent martingale problem. Let $\psi : [0, \infty) \times E \rightarrow [0, \infty)$ such that ψ , $\frac{\partial}{\partial s}\psi$ and $\Delta\psi$ are bounded and strongly continuous in $C_b(\mathbb{R}^d)$. Assume that

(9.40) $\left\| \frac{\psi(s+h, \cdot) - \psi(s, \cdot)}{h} - \frac{\partial}{\partial s}\psi(s, \cdot) \right\|_\infty \rightarrow 0 \quad \text{as } h \rightarrow 0.$

Then

(9.41) $\exp(-X_t(\psi_t)) + \int_0^t \exp(-X_s(\psi_s))X_s((A + \frac{\partial}{\partial s})\psi_s) ds - \frac{\gamma}{2} \int_0^t \exp(-X_s(\psi_s))X_s(\psi_s^2) ds$

is a \mathbb{P}_μ -martingale. Let $\mathbb{P}_\mu^{\mathcal{F}_t}$ denote the conditional expectation with respect to \mathcal{F}_t under \mathbb{P}_μ .

To prove (9.41) first note that applying (9.39) to $\exp(-\mu(\phi))$ with $\phi \in C_b^2(\mathbb{R}^d)$, we obtain

$$(9.42) \quad \mathcal{E}_t(\phi) = \exp(-X_t(\phi)) + \int_0^t \exp(-X_s(\phi)) X_s(A\phi) ds - \frac{\gamma}{2} \int_0^t \exp(-X_s(\phi)) X_s(\phi^2) ds$$

is a \mathbb{P}_μ -martingale.

Next take

$$(9.43) \quad \begin{aligned} u(s, X_t) &= \exp(-X_t(\psi_s)), & v(s, X_t) &= \exp(-(X_t(\psi_s)) X_t(\frac{\partial}{\partial s} \psi_s)), \text{ and} \\ w(s, X_t) &= \exp(-X_t(\phi))(X_t(A\psi_s)) \end{aligned}$$

so that for $t_2 > t_1$

$$(9.44) \quad u(t_2, X_{t_2}) - u(t_1, X_{t_2}) = - \int_{t_1}^{t_2} v(s, X_{t_2}) ds.$$

Then using (9.42) we have

$$(9.45) \quad \begin{aligned} \mathbb{P}_\mu^{\mathcal{F}_{t_1}}[u(t_1, X_{t_2}) - u(t_1, X_{t_1})] &= -\mathbb{P}_\mu^{\mathcal{F}_{t_1}} \left[\int_{t_1}^{t_2} w(s, X_s) ds \right] \\ &\quad + \frac{\gamma}{2} \mathbb{P}_\mu^{\mathcal{F}_{t_1}} \left[\int_{t_1}^{t_2} u(s, X_s) X_s(\psi_s^2) ds \right]. \end{aligned}$$

Let Λ^n be a partition of $[t_1, t_2]$ with $\text{mesh}(\Lambda^n) \rightarrow 0$ and

$$\begin{aligned} \psi^n(s, x) &: = \sum_{i=1}^n \psi(t_i^n, x) 1_{[t_i^n, t_{i+1}^n)}(s) \\ X^n(s) &: = \sum_{i=1}^n X_{t_{i+1}^n} 1_{[t_i^n, t_{i+1}^n)}(s) \end{aligned}$$

Let $u^n(t, X_t) := \exp(-X_t(\psi_t^n))$.

Then by (9.45)

$$\begin{aligned} \mathbb{P}_\mu^{\mathcal{F}_{t_1}}[u^n(t_2, X_{t_2}) - u^n(t_1, X_{t_1})] &= -\mathbb{P}_m^{\mathcal{F}_{t_1}} \left[\int_{t_1}^{t_2} \exp(-X_s^n(\psi_s)) X_s^n(\frac{\partial}{\partial s} \psi_s) ds \right] \\ &\quad - \mathbb{P}_\mu^{\mathcal{F}_{t_1}} \left[\int_{t_1}^{t_2} \exp(-X_s(\psi_s^n)) X_s(A\psi_s^n) ds \right] \\ &\quad + \frac{\gamma}{2} \mathbb{P}_\mu^{\mathcal{F}_{t_1}} \left[\int_{t_1}^{t_2} \exp(-X_s(\psi_s^n)) X_s((\psi_s^n)^2) ds \right]. \end{aligned}$$

Standard arguments show that this converges to

$$\begin{aligned} \mathbb{P}_\mu^{\mathcal{F}_{t_1}}[u(t_2, X_{t_2}) - u(t_1, X_{t_1})] &= -\mathbb{P}_m^{\mathcal{F}_{t_1}} \left[\int_{t_1}^{t_2} \exp(-X_s(\psi_s)) X_s(\frac{\partial}{\partial s} \psi_s) ds \right] \\ &\quad - \mathbb{P}_\mu^{\mathcal{F}_{t_1}} \left[\int_{t_1}^{t_2} \exp(-X_s(\psi_s)) X_s(A\psi_s) ds \right] \\ &\quad + \frac{\gamma}{2} \mathbb{P}_\mu^{\mathcal{F}_{t_1}} \left[\int_{t_1}^{t_2} \exp(-X_s(\psi_s)) X_s((\psi_s)^2) ds \right] \end{aligned}$$

which completes the proof of (9.41).

Now let $v_t = V_t\phi$ be the unique solution (see [473]) of

$$(9.46) \quad \frac{\partial v_t}{\partial t} = Av_t - \frac{\gamma}{2}v_t^2, \quad v_0 = \phi \in C_{+,b}^2(\mathbb{R}^d).$$

Then v_t satisfies (9.40). Applying (9.41) we deduce that $\{\exp(-X_s(v_{t-s}))\}_{0 \leq s \leq t}$ is a martingale. Equating mean values at $s = 0$ and $s = t$ we get the fundamental equality

$$(9.47) \quad \mathbb{P}_\mu(\exp(-X_t(\phi)) = \exp(-\mu(V_t\phi))).$$

The extension from $\phi \geq 0$ in $D(A)$ to $\phi \geq 0$ in $b\mathcal{E}$ follows easily by considering the “weak form”

$$V_t\phi = P_t\phi - \frac{\gamma}{2} \int_0^t P_{t-s}(V_s\phi)^2 ds$$

and then taking bounded pointwise limits.

(9.47) proves the uniqueness of the law $\mathbb{P}_\mu(X_t \in \cdot)$ for any solution of (MP) and hence the uniqueness of \mathbb{P}_μ (see [531] or [212], Chapt. 4, Theorem 4.2).

(c) follows by taking $\phi \equiv 1$ and comparing the Laplace transforms of the transition measures.

■

Corollary 9.15 (*Infinite divisibility*) \mathbb{P}_μ is an infinitely divisible probability measure with canonical representation

$$(9.48) \quad \begin{aligned} & -\log(\mathbb{P}_\mu(\exp(-X_t(\phi)))) \\ &= -\int \log(\mathbb{P}_{\delta_x}(\exp(-X_t(\phi)))\mu(dx) = \int_{M_F(\mathbb{R}^d) \setminus \{0\}} (1 - e^{\nu(\phi)}) R_t(x, d\nu) \end{aligned}$$

where the canonical measure $R_t(x, d\nu) \in M_F(M_F(\mathbb{R}^d) \setminus \{0\})$ satisfies

$R_t(x, M_F(\mathbb{R}^d) \setminus \{0\}) = \frac{2}{\gamma t}$ and the normalized measure is an exponential probability law with mean $\frac{\gamma t}{2}$.

The infinite divisibility of SBM allow us to use the theory of infinitely divisible random measure (see e.g. Dawson (1992) (1993), [119], [120]) to obtain detailed properties of the process.

Weighted occupation time

If $\{X_t : t \geq 0\}$ is a super-Brownian motion, then we defined the associated *weighted occupation time* $Y_y : t \geq 0$ as

$$(9.49) \quad Y_t(A) = \int_0^t X_s(A) ds, \quad A \in \mathcal{B}(\mathbb{R}^d).$$

Theorem 9.16 (*Iscoe (1986) [310]*)

Let $\mu \in M_F(\mathbb{R}^d)$ and $\phi, \psi \in C_{c,+}^2(\mathbb{R}^d)$. Then the joint Laplace functional of X_t and Y_t is given by

$$(9.50) \quad E_\mu \left[e^{-\langle \psi, X_t \rangle - \langle \phi, Y_t \rangle} \right] = e^{-\int (U_t^\phi \psi)(x) \mu(dx)}$$

where $(U_t^\phi \psi)(x)$ is the unique solution of

$$(9.51) \quad \begin{aligned} \frac{\partial u(t, x)}{\partial t} &= \frac{1}{2} \Delta u(t, x) - \frac{\gamma}{2} u^2(t, x) + \phi(x), \\ u(0, x) &= \psi(x). \end{aligned}$$

Proof. Now consider two semigroups on $C_{0,+}(\mathbb{R}^d)$:

- $V_t\psi$ given by the solution of

$$(9.52) \quad v(t) = S_t\psi - \int_0^t S_{t-s}(v(s))^2 ds,$$

- W_t given by

$$(9.53) \quad W_t\psi = \psi + t\phi, \quad \dot{W}_t\psi = \phi.$$

Using the iterated conditioning and the Markov property at the time points $\{\frac{N-1}{N}t, \frac{N-2}{N}t, \frac{N-3}{N}t, \dots, \frac{1}{N}t\}$ we obtain

$$(9.54) \quad \begin{aligned} E_\mu & \left(\exp \left[-\langle \psi, X_y \rangle - \int_0^t \langle \phi, X_s \rangle ds \right] \right) \\ &= \lim_{N \rightarrow \infty} \exp \left[-\langle (V_{\frac{t}{N}} W_{\frac{t}{N}})^N \psi, \mu \rangle \right] \end{aligned}$$

Then by the Trotter-Lie product formula (cf. Chorin et al [72])

$$(9.55) \quad U_t = \lim_{N \rightarrow \infty} \left(V_{\frac{t}{N}} W_{\frac{t}{N}} \right)^N \quad \text{on } C_{0,+}(\mathbb{R}^d).$$

and therefore

$$(9.56) \quad E_\mu \left(\exp \left[-\langle \psi, X_t \rangle - \int_0^t \langle \phi, X_s \rangle ds \right] \right) = \exp(-\langle U_t\psi, \mu \rangle)$$

noting that the interchange of limit and integral is justified by dominated convergence since

$$(9.57) \quad (V_{\frac{t}{N}} W_{\frac{t}{N}})^N \psi \leq (S_{\frac{t}{N}} W_{\frac{t}{N}})^N \psi \leq S_t\psi + (1+t)\|\phi\|.$$

Finally, we can verify that the semigroup U_t^ϕ defined by (9.55) satisfies (9.51). ■

As an application of this Iscoe established the *compact support property* of super-Brownian motion:

Theorem 9.17 *Let $\{X_t : t \geq 0\}$ be super-Brownian motion with initial measure δ_0 . Then*

$$(9.58) \quad P_{\delta_0} \left(\sup_{0 \leq t < \infty} X_t(\mathbb{R}^d \setminus \overline{B(0, R)}) > 0 \right) = 1 - e^{-\frac{u(0)}{R^2}},$$

where $u(\cdot)$ is the solution of

$$(9.59) \quad \begin{aligned} \Delta u(x) &= u^2(x), \quad x \in B(0, 1), \\ u(x) &\rightarrow \infty \text{ as } x \rightarrow \partial B(0, 1). \end{aligned}$$

Proof. Iscoe (1988) [312]. ■

Convergence of renormalized BRW and interacting Feller CSBP

A number of different variations of rescaled branching systems on \mathbb{R}^d can be proved to converge to SBM. For example, the following two results can be proved following the same basic steps.

Theorem 9.18 *Let $\varepsilon = \frac{1}{N}$. Consider a sequence of branching random walks X_t^ε on $\varepsilon\mathbb{Z}^d$ with random walk kernel $p^\varepsilon(\cdot)$ which satisfies (9.2), particle mass $m_\varepsilon = \varepsilon$, branching rate $\gamma^\varepsilon = \frac{\gamma}{\varepsilon}$ and assume that $X_0^\varepsilon \Rightarrow X_0$ in $M_F(\mathbb{R}^d)$. Then $\{X_t^\varepsilon\}_{t \geq 0} \Rightarrow \{X_t\}_{t \geq 0}$ where X_t is a super-Brownian motion on \mathbb{R}^d with $A = \frac{\sigma^2}{2}\Delta$ and branching rate γ .*

Remark 9.19 *The analogue of these results for more general branching mechanisms with possible infinite second moments are established in Dawson (1993), [120] Theorem 4.6.2.*

Theorem 9.20 *Consider a sequence of interacting Feller CSBP in which the rescaled random walks converge to Brownian motion. Then the interacting Feller CSBP converge to SBM.*

9.5.3 The Poisson Cluster Representation

Let X be an infinitely divisible random measure on a Polish space E with finite expected total mass $E[X(E)] < \infty$. Then (cf. [120], Theorem 3.3.1) there exists a measure $X_D \in M_F(E)$ and a measure $R \in M(M_F(E) \setminus \{0\})$ satisfying

$$\int (1 - e^{-\mu(E)}) R(d\mu) < \infty$$

and such that

$$-\log(P(e^{-X(\phi)})) = X_D(\phi) + \int (1 - e^{-\nu(\phi)}) R(d\nu).$$

X_D is called the deterministic component and R is called the *canonical measure*. For example, for a Poisson random measure with intensity Λ , $R(d\nu) = \int \delta_{\delta_x}(d\nu)\Lambda(dx)$. If we replace each Poisson point, x , with a random measure (*cluster*) with probability law $R(x, d\nu)$ then we obtain

$$R(d\nu) = \int \Lambda(dx)R(x, d\nu).$$

In the case of super Brownian motion with $X_0 = \mu \in M_F(\mathbb{R}^d)$, X_t for $t > 0$ is infinitely divisible with $X_D = 0$ and canonical measure $R_t(d\nu) = \int \mu(dx)R_t(x, d\nu)$ where

$$V_t \phi(x) = \int (1 - e^{-\nu(\phi)}) R_t(x, d\nu), \quad \phi \geq 0$$

(see e.g. Dawson and Perkins (1991) [118]). Since $(V_t \theta)(x) = \frac{2\theta}{(2+\theta\gamma)t}$, $R_t(x, M_F(\mathbb{R}^d) \setminus \{0\}) = \lim_{\theta \rightarrow \infty} (V_t \theta)(x) = \frac{2}{\gamma t}$ and

$$\int e^{-\theta\nu(1)} R_t(x, d\nu) = \frac{(\frac{2}{\gamma t})^2}{(\frac{2}{\gamma t}) + \theta}.$$

Hence $R_t(x, \nu(1) \in \cdot)$ is $\frac{2}{\gamma t}$ times an exponential law with mean $\frac{\gamma t}{2}$. Then using the above and (4.27) one can check that

$$\int e^{-\theta\nu(\phi)} R_t(x, d\nu) = \lim_{\varepsilon \downarrow 0} \varepsilon^{-1} \mathbb{P}_{\varepsilon\delta_x}(e^{-\theta X_t(\phi)} 1(X_t > 0)).$$

Hence $R_t(x, \cdot)$ can be interpreted as the unnormalized distribution of X_t starting with infinitesimal mass at x (when it is nonzero).

9.5.4 The Palm measure

We now consider the locally size-biased law which is given by the Campbell measure

$$(9.60) \quad \bar{R}_t(x, A \times B) = \nu(A)1_B(\nu)R_t(x, d\nu), \quad A \in \mathcal{B}(\mathbb{R}^d), B \in \mathcal{B}(M_F(\mathbb{R}^d))$$

The *Palm measure* at $y \in \mathbb{R}^d$ is defined by

$$(9.61) \quad (R_t(x, d\mu))_y = \frac{\bar{R}_t(x, dy \times d\mu)}{I(dy)}, \quad I(A) = \int \mu(A)R_t(x, d\mu)$$

Remark 9.21 In the case of a single point $X_t = x_t \delta_e$ (instead of \mathbb{R}^d),

$$(9.62) \quad (R_t(x, B))_y = \int_B \frac{\mu(e)R_t(x, d\mu)}{I(e)}$$

we obtain the size-biased distribution of the mass at the point.

The Laplace functional of the Palm measure is given by (see [118])

$$(9.63) \quad \int e^{-\mu(\phi)}(R_t(x, d\mu))_y = P_y \left(e^{-\gamma \int_0^t (V_r \phi)(w_r) dr} | w_t = x \right)$$

where P_y is the law of Brownian motion started at y and $V_t \phi$ satisfies

$$(9.64) \quad V_t \phi(x) = S_t \phi(x) - \frac{\gamma}{2} \int_0^t S_u (V_{y-u} \phi)^2 du.$$

9.5.5 Excursions

$w \in C([0, \infty), M_F(E))$ is an $M_F(E)$ -valued with lifetime τ starting at x if

$$\begin{aligned} w_0 &= 0, \quad \tau(w) > 0, \quad w_t \neq 0, \quad 0 < t < \tau(w), \quad w_t = 0, \quad t > \tau(w) \\ w_t(1)^{-1} w_t &\Rightarrow \delta_x \text{ as } t \rightarrow 0+. \end{aligned}$$

Let $C_x^0 = C_x^0([0, \infty), M_F(E) \setminus \{0\})$ denote the set of all excursion paths.

Theorem 9.22 (*Super-excursions and canonical measure*) (El Karoui and Roelly [204], Li and Shiga [404]) There is a unique σ -finite kernel $R(x, \cdot)$ from E to C_x^0 such that

- (i) for each $x \in E$ $R(x, \cdot)$ is supported by C_x^0
- (ii) $\int w_t(1)R(x, dw) \rightarrow 1$ as $t \rightarrow 0+$, and
- (iii) The measure $R(x, \cdot)$ is Markov with the same transition laws as super-Brownian motion and has t -marginal distribution equal to $R_t(x, \cdot)$.
- (iv)

$$E_m(e^{-X_t(\phi)}) = e^{(-\int \int 1_{\tau(w) \geq t} (1 - e^{-w_t(\phi)}) R(x, dw) m(dx))}.$$

Remark 9.23 $R(x, \{\tau > t\}) = \frac{2}{\gamma t}$ and we can verify that

$$R(x, B \cap \{\tau > t\}) = \lim_{\varepsilon \rightarrow 0+} \frac{P_{\varepsilon \delta_x}(B \cap \{X_t > 0\})}{\varepsilon}$$

The *integrated super-excursion* (ISE) is defined to be the random measure on \mathbb{R}^d given by

$$(9.65) \quad \mathcal{I}(B) := \int_0^\infty w_s(B) ds \quad \text{conditioned on } \mathcal{I}(\mathbb{R}^d) = 1.$$

It has been proved that ISE also describes the scaling limit of random trees in high dimensions (see Derbez and Slade (1997) [155] and is conjectured to arise in the scaling limit of critical percolation clusters in high dimensions (see Slade (2002), (2006) [521], [522] for a complete exposition of these results.)

9.5.6 The Historical Branching Process

An enrichment of SBM called the the Historical Brownian motion (HBM) was introduced by Dawson and Perkins (1991) [118] and studied by Dynkin (1991) [193], Le Gall (1991) [403] and Donnelly and Kurtz (1996) [163].

Given the Brownian motion ξ , the *path-valued process* $\bar{\xi}(t) = \xi(\cdot \wedge t)$ is a time-inhomogeneous continuous strong Markov process taking values in the Polish space $C = C(\mathbb{R}_+, E)$. We let $P_{r,y}$ (here $y(\cdot \wedge r) = y$) denote the law of $\bar{\xi}$ starting at y at time r . Historical Brownian motion is obtained by replacing ξ with the path-valued Brownian motion $\bar{\xi}$ in the general superprocess construction outlined above.

We can consider the corresponding empirical process for the BBM

$$H_t^\varepsilon = \sum_{\alpha \sim t} m_\varepsilon \delta_{\bar{\xi}_\alpha(t)}.$$

By a limiting procedure analogous to the above (but involving additional technical considerations), we get the ξ -historical process $\{H_t\}$ and law $\mathbb{Q}_{\tau,m}$ on Ω_H . Let $y^t := y(\cdot \wedge t)$, $C^t := \{y : y = y^t\}$, $M_F(C)^t = \{m \in M_F(C) : y^t = y \text{ } m - \text{a.e. } y\}$, \mathcal{C} is the Borel σ -field of C and \mathcal{C}_t is the sub- σ -field generated by the paths up to time t .

$$\begin{aligned} \Omega_H &= \{H \in C(\mathbb{R}_+, M_F(C)) : H_t \in M_F(C)^t \ \forall t \geq 0\}, \\ \mathcal{H}[s, t] &= \sigma(H_u : s \leq u \leq t), \quad \mathcal{H}[s, \infty) = \sigma(H_u : u \geq s). \end{aligned}$$

Foe $E = \mathbb{R}^d$ let $C_{0,\infty}^d$ denote the continuous functions from $[0, \infty)$ to \mathbb{R}^d and $D(\bar{A})$ denote the subset of $f \in C_{0,\infty}^d$ of the form

$$f(y) = g(y(t_1), \dots, y(t_n)), \quad 0 \leq t_1 \leq t_2 \dots \leq t_n,$$

where g is infinitely differentiable and constant outside a compact set.

For $f \in D(\bar{A})$ and $t > 0$, let

$$\bar{A}_t f(y) := \frac{1}{2} \sum_{i=1}^d \sum_{k=0}^{n-1} \sum_{\ell=0}^{n-1} 1(t < t_{k+1} \wedge t_{\ell+1}) g_{x_{kd+i}, x_{\ell d+i}}(y^t(t_1), \dots, y^t(t_n)).$$

where $g_{x_{kd+i}, x_{\ell d+i}}$ denote the second partial derivatives of g with respect to $x_{kd+i}, x_{\ell d+i}$, $k, \ell = 1, \dots, n-1$, $i = 1, \dots, d$.

Theorem 9.24 (*Perkins (1995) [480]*) If $\tau \geq 0$ and $m \in M_F(C)^\tau$, $\mathbb{Q}_{r,m}$ is the unique law on $(\Omega_H, \mathcal{H}[\tau, \infty))$ such that

$$(9.66) \quad \begin{aligned} & \text{for each } f \in D(\bar{A}), \quad M_t(f) = H_t(f) - m(f) - \int_r^t H_s(\bar{A}_s f) ds, \quad t \geq r \\ & \text{is an } (\mathcal{H}[r, \infty))\text{-martingale, starting at zero when } t = r, \\ & \text{and with quadratic variation} \\ & \langle M(f) \rangle_t = \int_r^t \gamma H_s(f^2) ds. \end{aligned}$$

The Laplace functional characterizing H has the form

$$\begin{aligned} \mathbb{Q}_{r,m}(\exp(-H_t(\phi))) &= \exp(-m(V_{r,t}\phi)) \\ V_{r,t}\phi(y) &= S_{r,t}\phi(y) - \frac{\gamma}{2} \int_r^t S_{r,s}(V_{s,t}\phi)^2 ds, \quad y \in C^t \\ S_{r,t}\phi(y) &= P_{r,y(r)}\phi(y/r/\xi(\cdot \wedge t)) \end{aligned}$$

where if $y, \xi \in C_{0,\infty}^d$, $r \in [0, \infty)$, then $(y/r/\xi) \in C_{0,\infty}^d$ is defined by

$$(y/r/\xi)(u) := \begin{cases} y(u) & \text{if } u < r \\ \xi(u-r) & \text{if } u \geq r \end{cases}.$$

Corollary 9.25 Let $\pi_t : C \rightarrow E$, $\pi_t(y) := y(t)$. Then .

$\pi_t(H_t)$ is a version of the SBM X_t

Remark 9.26 If A_1, \dots, A_n are disjoint sets in \mathcal{C}_r and $f_i(t, y) = 1_{A_i}(y)$, then clearly $f_i \in D(A)$ and $A_{\tau, m}f_i = 0$. (HMP) shows that under $\mathbb{Q}_{r,m}$, $X_t^i = H_{r+t}(A_i)$, $i = 1, \dots, n$, are independent FB processes starting at $m(A_i)$, $i = 1, \dots, n$. Therefore

$$(9.67) \quad \mathbb{Q}_{r,m}(\exp(-\sum_i^n \lambda_i H_{\tau+t}(A_i))) = \exp(-\sum_{i=1}^n 2\lambda_i m(A_i)/(2 + \lambda_i \gamma t)), \quad \lambda_i \geq 0,$$

and

$$(9.68) \quad \mathbb{Q}_{r,m}(H_{\tau+s}(A_i) = 0 \ \forall s \geq t) = \exp(-2m(A_i)/(\gamma t)).$$

Proposition 9.27 (*Historical Cluster Representation*) Let $r \leq s < t$. Under $\mathbb{Q}_{r,m}$, the conditional distribution of $H_t(\{y : y^s \in \cdot\})$ given $\mathcal{H}[r, s]$ is the law of $\sum_{i=1}^M \delta_{y_i} m_i$, where $\{y_1, \dots, y_M\}$ are the points of a Poisson point process with intensity $H_s(\cdot)2\gamma^{-1}(t-s)^{-1}$ and $\{m_1, \dots, m_M\}$ are independent exponential masses with mean $\gamma(t-s)/2$.

Proof. As above we may take $s = r$ and argue unconditionally. An easy calculation shows the Laplace functional of the above Poisson cluster random measure, Ξ , is

$$(9.69) \quad \mathbb{P}(\exp(-\Xi(\phi))) = \exp\left(-\int 2\phi(y)(2 + \phi(y)\gamma(t-r))^{-1} m(dy)\right).$$

If $\phi(y) = \sum_1^n \lambda_i 1_{B_i}(y)$ for B_1, \dots, B_n disjoint Borel sets in C , (9.67) implies

$\mathbb{Q}_{r,m}\left(\exp(-\int \phi(\bar{y}(\tau)) H_t(d\bar{y}))\right)$ is also given by the right side of (9.69). By taking limits of simple functions we see that the Laplace functionals of the random measures in question are equal and hence so are their laws. \square

Remark 9.28 Perkins developed the historical stochastic calculus which is an analogue of Itô stochastic calculus for the historical superprocess (see Perkins (1992), [479], [480]) and in Evans-Perkins (1998) [225] used it to obtain a model of competing superprocesses with killing according to a collision local time. See Perkins (2002) [487] for a comprehensive exposition of these developments.

9.5.7 A skew product construction

Etheridge and March [206] proved that between normalized SBM conditioned to have constant mass is a Fleming-Viot process. Perkins then proved that (unconditioned) normalized SBM could be viewed as a Fleming-Viot process with time-dependent resampling rate inversely to the total mass process for the SBM (Perkins Disintegration Theorem [478]). We leave the precise statement and proof of these results to Chapter 12 but using these ideas now give an informal construction of SBM from the Feller CSB and the Fleming-Viot process.

Consider the following $\mathbb{R}^+ \times M_1(E)$ -valued martingale problem for $(x(t), Y(t))$

(9.70) $x(t)$ is a martingale with increasing process

$$\langle x \rangle_t = \int_0^t x(s) ds$$

Let

$$(9.71) \quad \tau := \inf\{t : x(t) = 0\}.$$

In other words $x(t)$ is a critical Feller CSBP process.

Perkins [478] introduced a Fleming-Viot process Y_t with time varying resampling rate $1/x(t)$ that satisfies the following martingale problem:

$$\begin{aligned} M_t(\varphi) &:= \{\langle \varphi, Y_t \rangle - \int_0^t \langle A\varphi, Y_s \rangle ds\}_{0 \leq t < \tau} \text{ is a martingale} \\ \langle M(\varphi) \rangle_t &= \int_0^t \frac{1}{x(s)} [\langle \varphi^2, Y_s \rangle - \langle \varphi, Y_s \rangle^2] ds \end{aligned}$$

Now consider the process

$$X(t) := x(t)Y_t$$

Let us verify that X is a solution to the super-A martingale problem. Applying Ito's Lemma to the semimartingale $\langle \varphi, X(t) \rangle$ we obtain

$$\begin{aligned} d\langle \varphi, X(t) \rangle &= x_t \langle A\varphi, Y(t) \rangle + x_t dM_t(\varphi) + \langle \varphi, Y_t \rangle dx_t \\ &= \langle A\varphi, X(t) \rangle + \tilde{M}_t(\varphi) \\ \tilde{M}_t(\varphi) &:= x_t dM_t(\varphi) + \langle \varphi, Y_t \rangle dx_t \\ \langle \tilde{M}(\varphi) \rangle_t &= \int_0^t \left\{ \frac{x_s^2}{x_s} [\langle \varphi^2, Y_s \rangle - \langle \varphi, Y_s \rangle^2] + \langle \varphi, Y_s \rangle^2 x_s \right\} ds \\ &= \int_0^t \{ \langle \varphi^2, X_s \rangle - x_s \langle \varphi, Y_s \rangle^2 + \langle \varphi, Y_s \rangle^2 x_s \} ds \\ &= \int_0^t \langle \varphi^2, X_s \rangle ds. \end{aligned}$$

9.5.8 The Donnelly-Kurtz countable particle representation

Donnelly-Kurtz (1999) [165] developed a countable particle representation that includes both the Fleming-Viot process and superprocess in the same spirit as the lookdown process construction of the Fleming-Viot process but we do not consider this here.

9.5.9 Brownian Snake Representation of SBM X_t

In (1991) Le Gall [403] developed the *Brownian snake representation* of super-Brownian motion. This is based on a deep connection between Brownian excursions and Feller CSB that was suggested by the discovery by Neveu and Pitman (1989) [456] of a natural branching process structure within a Brownian excursion. We have already met this in the discussion of the continuum random tree.

The Brownian snake is based on reflecting Brownian motion $\{\zeta_t\}_{t \geq 0} \in \mathbb{R}_+$ which serves as the *lifetime process*. Let $C_s([0, \infty), \mathbb{R}^d)$ denote the set of continuous paths in \mathbb{R}^d stopped at time s . Then given $\{\zeta_t\}_{t \geq 0}$ the Brownian snake $W_s \in C_s([0, \infty), \mathbb{R}^d)$ and is characterized as follows: for $s_1 < s_2$

$$W_{s_1}(u) = W_{s_2}(u), \quad u \leq \min_{s \in [s_1, s_2]} \zeta_s$$

the continuations are given by independent Brownian paths $u \geq \min_{s \in [s_1, s_2]} \zeta_s$

The excursion measure of the Brownian snake starting at x is defined by

$$(9.72) \quad N_x(df d\omega) = n(df) \Theta_x^f(d\omega)$$

where $n(df)$ is the Itô excursion measure (on the set of positive excursions) and Θ_x^f is the law of the Brownian snake started at x with lifetime process f .

Let

$$(9.73) \quad \mathcal{W} = \cup_{t \geq 0} C([0, t], \mathbb{R}^d), \quad \zeta_w = t \text{ if } w \in C([0, t], \mathbb{R}^d).$$

$$\int_{\mathbb{R}^d} \varphi(x) X_t(dx) = \int_{\mathcal{W} \times \mathbb{R}^d} \left[\int_0^{\sigma(W)} \varphi(W_s(\zeta_s)) L_{ds}^t \right] \Pi(dw dx)$$

$(L_s^a)_{s \geq 0}$ local time of ζ at a and $\sigma(W)$ is the lifetime of the excursion and where Π is a Poisson field with intensity

$$\pi(dw dx) = \int_{\mathbb{R}^d} N_y(dw) \otimes \delta_y(dx) X_0(dy)$$

and N_y is the Itô excursion measure for the Brownian snake from the trivial path at y . The corresponding measures on paths $\{W_u\}_{0 \leq u \leq \zeta_s}$ gives the *historical process*.

9.5.10 Catalytic SBM

The branching rate γ of SBM is assumed to be constant. A superprocess in which the branching rate γ is replaced by a non-negative function or measure is called catalytic SBM. See Dawson-Fleischmann [139], [142] for an exposition.

9.6 Local properties of critical branching systems

9.6.1 Support Properties of SBM

We first derive some properties using the skew product representation and the lookdown process for the Fleming-Viot process.

Theorem 9.29 (*Iscoe's Compact Support Property*) Consider a super Brownian motion in \mathbb{R}^d with compactly supported initial measure. Then at any fixed time, $t \geq 0$, the closed support of X_t is compact with probability one.

Proof. Using the above theorem, this follows immediately from the compact support property for the Fleming-Viot process, Theorem 8.29. ■

Theorem 9.30 (*R. Tribe*) Consider super-Brownian motion $\{X_t\}_{t \geq 0}$. Let $\tau := \inf\{t : X_t(1) = 0\}$. Then

$$\lim_{t \rightarrow \tau^-} Z(t) = \delta_{\zeta_1(\tau)} \text{ a.s.}$$

Proof. Note that τ is predictable and $(\zeta_1(\tau-), \zeta_2(\tau-), \dots)$ is exchangeable. It is known that for the Feller branching diffusion $\{x(t) : t \geq 0\}$ that

$$\int_0^{\tau^-} \frac{1}{x(s)} ds = \infty.$$

From this it follows that each ζ_i has lookdowns to ζ_0 at times arbitrarily close to τ^- . The result then follows from the Hölder continuity of the Brownian paths. ■

9.6.2 Brief review of sample path properties

There is an extensive literature on the sample path properties of SBM $\{X_t\}_{t \geq 0}$. We do not consider this in detail but briefly mention some basic properties.

- In dimensions $d = 1$, $X_t(dx) = \tilde{X}_t(x)dx$ has a jointly continuous density which is given by a weak solution of stochastic partial differential equation (SPDE)

$$d\tilde{X}_t(x) = \frac{1}{2}\Delta\tilde{X}_t(x)dt + \sqrt{\tilde{X}_t(x)}W(dt, dx)$$

$$W(dt, dx) = \text{white noise}$$

(Konno-Shiga (1988) [381], Reimers (1989) [492])

Strong uniqueness is an open problem.

- In dimensions $d \geq 2$ X_t is a.s. a singular measure (D-Hochberg (1979) [109]). The Hausdorff measure properties of the support of SBM were established by Perkins (1989) [476] for dimensions $d \geq 3$, and by Le Gall and Perkins (1995) [396] in dimension $d = 2$: There is a universal constant $c_0 \in (0, \infty)$ such that

$$X_t(A) = c \cdot \phi_d - m(A \cap S(X_t)), \text{ a.s.}$$

where

$$S(X_t) = \text{closed support of } X_t$$

and

$$\phi_{\geq 3}(r) = r^2 \log \log \frac{1}{r}, \quad \phi_2(r) = r^2 \log^+(1/r) \log^+ \log^+(1/r).$$

For a comprehensive development of these and other sample path properties of super-Brownian motion refer to Perkins (2002) [487].

9.7 Spatial dynamical invariance principles

SBM arose in a natural context as the limit of rescaled branching random walks, branching Brownian motions, interacting Feller CSBP. There is a natural invariance principle here, namely, for a large class of migration mechanism in the domain of attraction of Brownian motion and branching mechanisms having finite second moments the limit is SBM. However more surprising, it has been discovered that super-Brownian motion also arises in the scaling limit of systems that have no immediate branching interpretation. Examples related to finite excursions of SBM arose in lattice trees where David Aldous (1993) conjectured that if $d > 8$ the rescaled tree of size N , X_N should converge to (ISE) and Derbez and Slade (1997) [155] proved this in sufficiently high dimensions. Examples also arise in interacting particle systems. Durrett and Perkins (1999) [186] proved that a class of rescaled contact processes converge to SBM and Cox-Durrett-Perkins (2000) [96] proved that a class of rescaled voter models converge to SBM. Other examples are *voter model clusters* Bramson-Cox-Le Gall (2001) [57], and interacting diffusions Cox-Klenke (2003) [98]. The proofs of many of these results involved weak convergence of solutions of martingale problems in the same spirit as the proof of the convergence of branching Brownian motion to SBM. The proofs in Derbez-Slade [155] and the proof due to van der Hofstad and Slade (2003) [300] that critical oriented percolation above 4+1 dimensions converges to SBM uses lace expansion methods to prove the convergence of the moment measures of the finite dimensional distributions.

Scaling limits of long-range voter models in $d = 1$

We begin by describing a class of *long range voter models* on \mathbb{Z}^1 . Let

$$(9.74) \quad S_N := \left\{ \frac{x}{\widetilde{M}_N \cdot N} : x \in \mathbb{Z}^1 \right\}$$

and probability distribution on S_N defined by

$$(9.75) \quad p_N(x) := \begin{cases} \frac{1}{2[\widetilde{M}_N \sqrt{N}]} & \text{if } x \text{ is one of the } 2[\widetilde{M}_N \sqrt{N}] \text{ equally spaced points in } (-\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{N}}) \\ & \quad := 0, \text{ otherwise.} \end{cases}$$

Let $\xi_t^N(x) = \xi_{Nt}(x\sqrt{N})$ denote the biased voter model with state space $\{0, 1\}^{S_N}$ with voting rate for a site with opinion 1 is $\gamma_\theta(N) = \widetilde{M}_N(N + \theta\sqrt{N})$, $\theta \geq 0$ and is $\widetilde{M}_N N$ for a site with opinion 0 and voting kernel $p_N(x, y) = p_N(x - y)$.

Define

$$(9.76) \quad X_t^N = \frac{1}{N} \sum_{x \in S_N} \xi_t^N(x) \delta_x.$$

Mueller and Tribe introduce the space

$$(9.77) \quad \mathcal{C} = \{f : \mathbb{R} \rightarrow [0, \infty) \text{ continuous with } |f(x)|e^{\lambda|x|} \rightarrow 0 \text{ as } |x| \rightarrow \infty \forall \lambda < 0\}.$$

Theorem 9.31 (Mueller-Tribe [448] Theorem 2) Let $\widetilde{M}_N \equiv 2$ and let the approximate densities of X_t^N be defined by

$$(9.78) \quad u_N(t, x) = \frac{\sum_{x \sim y} \xi_t^N(y)}{\sum_{x \sim y} 1} \quad \text{where } x \sim y \text{ iff } |x - y| \leq N^{-1/2}.$$

Assuming that the initial approximate densities converge in \mathcal{C} to $u(0)$, then the approximate densities converge in distribution to a continuous \mathcal{C} -valued process $u(t)$ which is a solution of the stochastic partial differential equation (SPDE)

$$(9.79) \quad \frac{\partial u}{\partial t} = \frac{1}{6} \frac{\partial^2}{\partial x^2} + 2\theta_v u(1-u) + \sqrt{4u(1-u)} \dot{W}$$

with initial condition $u(0)$ and where \dot{W} is space-time white noise.

Remark 9.32 This SPDE is a Fisher-KPP equation driven by Fisher-Wright noise.

In contrast to this, (for the unbiased voter model) the special one-dimensional case of the Cox-Durrett-Perkins long-range voter model in which $\widetilde{M}_N \rightarrow \infty$ to be presented in the next subsection has as scaling limit SBM. The essential difference is the CDP scaling is such that the measure is more sparsely distributed on the voters and the local density of sites occupied by opinion 1 goes to 0.

Theorem 9.33 (CDP [96]) Assume that $\widetilde{M}_N \rightarrow \infty$ as $N \rightarrow \infty$. Let P_N denote the law of X_t^N on $D_{M_F(\mathbb{R}^1)}([0, \infty))$. Assume that $\sum \xi_0^N(x) < \infty$ and $X_0^N \rightarrow X_0$ in $M_F(\mathbb{R}^1)$ as $N \rightarrow \infty$. Then

$$(9.80) \quad P_N \Rightarrow P_{X_0}^{1,2,1/3}$$

as $N \rightarrow \infty$ where $P^{d,2,1/3}$ is the law of super-Brownian motion with branching rate 2 and $\sigma^2 = \frac{1}{3}$ where $P_{X_0}^{\gamma, \sigma^2}$ denotes the law of SBM in \mathbb{R}^d with initial measure $X_0 \in M_F(\mathbb{R}^d)$ and log-Laplace equation

$$(9.81) \quad \frac{\partial u}{\partial t} = \frac{\sigma^2}{2} \Delta u - \frac{\gamma}{2} u^2.$$

The rescaled voter model in \mathbb{Z}^d , $d \geq 3$

Assume that the kernel $p(x - y)$ is irreducible and symmetric, $p(0) = 0$ and $\sum_{x \in \mathbb{Z}^d} x^i x^j p(x) = \delta_{ij} \sigma^2$. Let $\xi_t(x)$ denote the voter model on \mathbb{Z}^d with kernel $p(\cdot)$.

Let $S_N = \frac{\mathbb{Z}^d}{\sqrt{N}}$, $\xi_t^N(x) := \xi_{Nt}(x\sqrt{N})$, $x \in S_N$, and consider the measure-valued process on \mathbb{R}^d defined by

$$(9.82) \quad X_t^N = \frac{1}{N} \sum_{x \in S_N} \xi_t^N(x) \delta_x.$$

Theorem 9.34 ([96], Theorem 1.2) Let P_N denote the law of X_t^N on $D_{M_F(\mathbb{R}^d)}([0, \infty))$ with $d \geq 3$. Assume that $\sum \xi_0^N(x) < \infty$ and $X_0^N \Rightarrow X_0$ in $M_F(\mathbb{R}^1)$ as $N \rightarrow \infty$. Then

$$(9.83) \quad P_N \Rightarrow P_{X_0}^{d, 2\gamma_e, \sigma^2} \text{ as } N \rightarrow \infty$$

where γ_e is the escape probability

$$(9.84) \quad \gamma_e = P_0(p(\cdot)\text{-RW never returns to 0}).$$

Remark 9.35 The case $d = 2$ is more subtle. In this case there is a logarithmic correction and one considers the sequence of measure-valued processes

$$(9.85) \quad X_t^N = \frac{\log N}{N} \sum_{x \in S_N} \xi_t^N(x) \delta_x.$$

With this scaling Cox-Durrett-Perkins [96] Theorem 1.2 prove the weak convergence to $P_{X_0}^{2\gamma_e, \sigma^2}$.

Rescaled long-range voter models

To describe the results for the long range voter model in higher dimensions we let M_N be a sequence of positive constants with $M_N \rightarrow \infty$ as $N \rightarrow \infty$ and consider the voter model on $S_N = \left\{ \frac{x}{M_N \sqrt{N}} : x \in \mathbb{Z}^d \right\}$ with probability kernels $p_N(\cdot)$ defined as follows:

For each N let W_N denote a random variable with values in $\frac{(\mathbb{Z}^d \setminus \{0\})}{M_N}$ and with uniform distributed on $(\frac{(\mathbb{Z}^d \setminus \{0\})}{M_N}) \cap I$ where $I = [-1, 1]^d$

Then

- W_N and $-W_N$ have the same distribution,
- $\lim_{N \rightarrow \infty} E(W_N^i W_N^j) = \delta_{ij} \sigma^2$ with $\sigma^2 = \frac{1}{3}$, and
- the family $\{|W_N|^2\}_{N \in \mathbb{N}}$ is uniformly integrable.

Then define the kernel

$$(9.86) \quad p_N(x) := P\left(\frac{W_N}{\sqrt{N}} = x\right)$$

Let $\xi_t^N(x) = \xi_{Nt}(x\sqrt{N})$ denote the voter model on $\{0, 1\}^{\mathbb{S}_N}$ with rate N and voting kernel $p_N(x, y) = p_N(x - y)$ and let P_N be the law of the measure-valued process

$$(9.87) \quad X_t^N = \frac{1}{N} \sum_{x \in \mathbb{S}_N} \xi_t^N(x) \delta_x$$

Theorem 9.36 (Cox, Durrett, Perkins (2000) [96])

Assume that

$$(9.88) \quad \sum_x \xi_0^N(x) < \infty,$$

$$(9.89) \quad X_0^N \Rightarrow X_0 \text{ in } M_F(\mathbb{R}^d) \text{ as } N \rightarrow \infty$$

and

$$(9.90) \quad \begin{cases} M_N/\sqrt{N} \rightarrow \infty, & \text{in } d = 1 \\ M_N^2/(\log N) \rightarrow \infty, & \text{in } d = 2 \\ M_N \rightarrow \infty, & \text{in } d \geq 3, \end{cases} .$$

Then $P_N \Rightarrow P_{X_0}^{d, 2, 1/3}$ as $N \rightarrow \infty$.

Interacting diffusions - convergence to SBM

Assume that the kernel $p(x - y)$ on $S = \mathbb{Z}^d$ is irreducible and symmetric, $p(0) = 0$ and $\sum_{x \in \mathbb{Z}^d} x^i x^j p(x) = \delta_{ij} \sigma^2$.

Now consider the system of interacting diffusions on \mathbb{Z}^d :

$$dX_t(x) = [\sum q_{x,y}(X_t(y)) - X_t(x)]dt + \sqrt{g(X_t(x))} dW_t(x), \quad x \in \mathbb{Z}^d$$

where $q_{x,y} = p(x - y)$ and $\{(W_t(x))_{t \geq 0}\}_{x \in \mathbb{Z}^d}$ are independent Wiener processes.

Let P_N denote the probability laws on $C_{M_F(\mathbb{R}^d)}([0, \infty))$ of the measure-valued processes

$$(9.91) \quad X_t^N := \frac{1}{N} \sum_{i \in \mathbb{Z}^d} X_{Nt}(i) \delta_{i/\sqrt{N}} \in M_F\left(\frac{\mathbb{Z}^d}{\sqrt{N}}\right).$$

Theorem 9.37 (Cox and Klenke (2003) [98], Theorem 1) Assume that $X_0^N \Rightarrow X_0$ in $M_F(\mathbb{R}^d)$ with $d \geq 3$ and that

$$(9.92) \quad g(x) = \kappa x(1-x)^+, \quad x \geq 0, \quad \kappa > 0,$$

Then

$$(9.93) \quad P^N \Rightarrow P^{d,\gamma,\sigma^2}$$

where

$$(9.94) \quad \gamma = \lim_{\theta \rightarrow 0} \frac{1}{\theta} \int g(x(0)) \nu_\theta(dx) = \gamma_e \cdot \kappa$$

ν_θ is the invariant measure for the interacting Wright-Fisher system on \mathbb{Z}^d with intensity θ and γ_e is given by (9.12).

Cox and Klenke [98] conjectured that the same result holds for the more general case with

- $g : [0, \infty) \rightarrow [0, \infty)$ is locally Lipschitz continuous
- $g^{-1}((0, \infty)) = (0, b)$ for some $b \in (0, \infty]$,
- $g(z) \leq C(1 + z^2)$ for some $C < \infty$.

In this case the conjectured limit is P^{d,γ,σ^2} with

$$(9.95) \quad \gamma = \lim_{\theta \downarrow 0} \frac{1}{\theta} \int g(x(0)) \nu_\theta(dx)$$

where ν_θ is the unique stationary measure on \mathbb{Z}^d with intensity θ .

Remark 9.38 Note that in Cox-Klenke the limiting branching rate is obtained from the derivative at 0 of the $\mathcal{F}g$. In this thinning out we get super-Brownian excursions. (cf. Bramson, Cox Le Gall).

Remark 9.39 Cox and Klenke also describe the long range case and note that the invariant measure involved then can be described by the mean-field equation.

9.7.1 Methods of Proof

The proofs of these results involve three main steps

- formulation of the sequence of measure-valued processes in terms of martingale problems
- establishing tightness of the probability laws on $D_{M_F}(\mathbb{R}^d)$
- verifying any any limit point satisfies the SBM martingale problem.

Step 1: Reformulation as a martingale problem

Starting from the set of stochastic equations (9.20), we can characterize the rescaled voter model as the solution of the stochastic integral equations

$$(9.96) \quad \xi_t^N(x) = \xi_0^N(x) + \sum_y \int_0^t [\xi_{s-}^N(y) - \xi_{s-}^N(x)] s \Lambda_s^N(x, y), \quad x \in S_N,$$

where $\{\Lambda_t^N(x, y) : x, y \in S_N\}$ is a system of independent Poisson processes with rates $N p_N(y - x)$. Then the measure-valued processes X_t^N can be characterized by a martingale problem as follows.

Theorem 9.40 (Cox, Durrett, Perkins (2000) Theorem 2.2)
Let $\phi \in C_b^{1,3}([0, \infty) \times \mathbb{R}^d)$. Then

$$(9.97) \quad X_t^N(\phi) = X_0^N(\phi) + \int_0^t X_s^N(\phi_1(s) + \mathcal{A}_N \phi(s)) ds + M_t^N(\phi),$$

where $\phi_1(s)$ denotes the partial derivative $\frac{\partial \phi(s, \cdot)}{\partial s}$ and

$$(9.98) \quad \mathcal{A}_N \phi(s, x) = N \sum_y p_N(x - y) (\phi(s, y) - \phi(s, x)),$$

and

$$(9.99) \quad M_t^N(\phi) = \frac{1}{N} \sum_x \sum_y \int_0^t \phi(s, x) (\xi_{s-}(y) - \xi_{s-}(x)) \hat{\Lambda}_s(x, y), \quad 0 \leq t \leq T,$$

$\hat{\Lambda}_t^N(x, y) = \Lambda_t^N(x, y) - N p_N(x - y) t$ is a cadlag, square integrable (\mathcal{F}_t) -martingale.
The predictable square function is given by

$$(9.100) \quad \langle M^N(\phi) \rangle_t = \int_0^t [2X_s^N(\phi^2(s) V_N(s)) + \varepsilon_s^N(\phi)] ds,$$

where

$$(9.101) \quad \begin{aligned} V_N(t, x) &= \sum_y p_N(y - x) 1\{\xi_t(y) = 0\} = \text{density of vacant sites near } x, \\ &= \lim_{\lambda \rightarrow \infty} \sum_y p_N(y - x) e^{-\lambda X_t^N(y)} \end{aligned}$$

and $\varepsilon_s^N(\phi)$ satisfies

$$(9.102) \quad E \left(\sup_{s \leq T} |\varepsilon_s^N(\phi)|^2 \right) \rightarrow 0, \text{ as } N \rightarrow \infty, \text{ for any } T > 0.$$

(iii) For any $T > 0$,

$$(9.103) \quad E \int_0^T X_s^N (|\mathcal{A}_N \phi(s) - \frac{\sigma^2}{2} \Delta \phi(s)|) ds \rightarrow 0 \text{ as } N \rightarrow \infty.$$

Steps 2 and 3: The Cox-Durrett-Perkins Criteria for weak convergence to SBM

Cox,Durrett and Perkins formulate a general set of additional conditions for a sequence of $M_F(\mathbb{R}^d)$ -valued martingale problems satisfying the conclusions of Theorem 9.40 which if satisfied imply that the solutions to these martingale problems converge weakly to SBM.

Assumptions:

(I1) There is a finite $\gamma > 0$ such that, for all $\phi \in C_0^\infty(\mathbb{R}^d)$ and $T > 0$, as $N \rightarrow \infty$,

$$(9.104) \quad E \left[\left(\int_0^T X_s^N (\{V_N(s) - \gamma\} \phi^2) ds \right)^2 \right] \rightarrow 0.$$

(I2) For all $T > 0$ there exists a finite C_T such that $\lim_{T \downarrow 0} C_T = 0$ and for all N ,

$$(9.105) \quad \int_0^T E^{\xi_0^N} [X_s^N (V_N(s))] ds \leq C_T X_0^N ((1)).$$

(I3) There is a $\theta \in (0, 1]$ and a finite $C(\varepsilon, T, K)$ such that for all $N \in \mathbb{N}$ all cutoffs $0 < \varepsilon, K < \infty$ and all pairs of times $\varepsilon \leq s \leq t \leq T$, we have

$$(9.106) \quad \sup \left\{ E \left[\left(\int_s^t X_r^N (V_N(r)) dr \right)^2 \right] : X_0^N ((1)) \leq K \right\} \leq C(\varepsilon, T, K) |t - s|^{1+\theta}.$$

Theorem 9.41 Assume (I1)-(I3). Then $P_N \Rightarrow P_{X_0}^{2\gamma, \sigma^2}$.

Proof. We refer to [96] for the details but outline the main ideas here. The proof involves two parts. The first is the proof of the tightness of the processes on $D([0, \infty), M_F(\mathbb{R}^d))$ and all limit points are supported by $C([0, \infty), M_F(\mathbb{R}^d))$. This follows the same general lines as in our examples - we omit the details.

The second part shows that every limit point P of P_N satisfies the SBM martingale problem. By Skorohod's theorem, given $P_{N_k} \Rightarrow P$ we may assume there is X, X^{N_k} defined on a probability space (Ω, \mathcal{F}, P) such that

$$(9.107) \quad X^{N_k} \rightarrow X, \text{P-a.s. in } D([0, \infty), M_F(\mathbb{R}^d)).$$

Then a standard argument shows that for $T > 0$,

$$(9.108) \quad \lim_{k \rightarrow \infty} \sup_{t \leq T} \left| \int_0^t X_s^{N_k} (\mathcal{A}_{N_k} \phi) ds - \int_0^t X_s (\sigma^2 \Delta \phi / 2) ds \right| = 0, \text{ P -a.s.}$$

Let

$$(9.109) \quad M_t(\phi) := X_t(\phi) - X_0(\phi) - \int_0^t X_s(\sigma^2 \Delta \phi / 2) ds.$$

Then using (9.103), some calculus and the fact that $M(\phi)$ is continuous to derive uniform convergence on compacts from convergence on D one can verify

$$(9.110) \quad \lim_{k \rightarrow \infty} \sup_{t \leq T} |M_t^{N_k}(\phi) - M_t(\phi)| = 0, \quad P - \text{a.s.}$$

A stochastic calculus computation starting with the system of stochastic integral equations (9.96) yields

$$(9.111) \quad E\langle M^N(\phi) \rangle_t^2 = E \left(\int_0^t [2X_s^N(\phi^2 V_{N,s}) + \varepsilon_s^N(\phi)] ds \right)^2$$

where $|\varepsilon_s^N(\phi)| \leq C_\phi X_s^N(1)/\sqrt{N}$. Then from (I1), Theorem 9.40(ii) and the elementary estimate

$$(9.112) \quad E(\sup_{s \leq T} (X_s^N(1))^2) \leq C(X_0^N(1) + (X_0^N(1))^2)$$

it follows that that for $T > 0$, $\sup_N E(\langle M^N(\phi) \rangle_T^2) < \infty$.

Then using Burkholder's inequality (see appendix Theorem 12.5) and noting that $|\Delta M^N(\phi)(t)| \leq \|\phi\|_\infty(N)^{-1}$, we have

$$(9.113) \quad \sup_N E \left(\sup_{t \leq T} |M_t^N(\phi)|^4 \right) < \infty.$$

Fix $0 \leq t_1 < t_2 < \dots < t_n \leq s < t$ and test functions $h_i : M_F(\mathbb{R}^d) \rightarrow \mathbb{R}$ that are bounded and continuous for $1 \leq i \leq n$. Now Theorem 9.40, (9.107),(9.110),(10.23) and dominated convergence imply that

$$E \left((M_t(\phi) - M_s(\phi)) \prod_1^n h_i(X_{t_i}^{N_k}) \right) = 0.$$

Therefore under P , $M_t(\phi)$ is a continuous \mathcal{F}_t^X -martingale where \mathcal{F}_t^X is a canonical right-continuous filtration generated by X . Also (9.107),(9.110),(10.23) imply that

$$\begin{aligned} & E \left(\left(M_t(\phi)^2 - M_s(\phi)^2 - \int_s^t X_r(2\gamma\phi^2) dr \right) \prod_1^n h_i(X_{t_i}^{N_k}) \right) \\ & \lim_{k \rightarrow \infty} E \left(\left(M_t^{N_k}(\phi)^2 - M_s^{N_k}(\phi)^2 - \int_s^t X_r^{N_k}(2\gamma\phi^2) dr \right) \prod_1^n h_i(X_{t_i}^{N_k}) \right) \end{aligned}$$

(9.101), (9.102) and (I1) show that the above equals

$$(9.114) \quad \lim_{k \rightarrow \infty} E \left(\left(M_t^{N_k}(\phi)^2 - M_s^{N_k}(\phi)^2 - (\langle M^{N_k}(\phi) \rangle_t - \langle M^{N_k}(\phi) \rangle_s) \right) \prod_1^n h_i(X_{t_i}^{N_k}) \right),$$

which is 0 by Theorem 9.40. This shows that $\langle M(\phi) \rangle_t = \int_0^t X_s(2\gamma\phi^2) ds$ for all $t \geq 0$, P-a.s. Therefore P satisfies the SBM martingale problem $(MP)_{X_0}^{2\gamma, \sigma^2}$. Therefore the law of X equals $P_{X_0}^{2\gamma, \sigma^2}$. ■

9.7.2 Proof of the invariance principle for interacting Wright-Fisher diffusions

In this case we have

$$(9.115) \quad M_t^N(\varphi) := X_t^N(\varphi) - X_0^N(\varphi) - \int_0^t X_s^N(\mathcal{A}_N \varphi) ds$$

is a continuous square integrable martingale with quadratic variation process

$$(9.116) \quad \langle M^N(\varphi) \rangle_t = \int_0^t \Gamma_s^N(\varphi^2) ds,$$

where

$$(9.117) \quad \begin{aligned} \Gamma_s^N &:= \frac{1}{N} \sum_{x \in \frac{\mathbb{Z}^d}{\sqrt{N}}} g(NX_s^N(\{x\})) \delta_x \\ &= \frac{1}{N} \sum_{i \in \mathbb{Z}^d} g(NX_{sN}(i)) \delta i / \sqrt{N}. \end{aligned}$$

Therefore the key step Cox-Klenke [98] in verifying the CDG conditions is to prove that

$$(9.118) \quad \varepsilon_{K,\phi}^{N,\gamma}(t) := \sup\{E[|(\Gamma_t^N - \gamma X_t^N)(\phi^2)| : X_0^N(1) \leq K]\} \rightarrow 0 \text{ as } N \rightarrow \infty.$$

Verification via Duality calculations

The verification is achieved using a duality calculation which involves the coalescing random walk with two particles. The difference between the voter and Wright-Fisher cases is that in the former colliding particles coalesce instantaneously whereas for Wright-Fisher they coalesce *with delay*. We now sketch the proof in the Wright-Fisher case.

In the case when g has Wright-Fisher form $g(x) = \kappa x(1-x)^+$, γ_e given by (9.12) this becomes

(9.119)

$$\varepsilon_{K,\varphi}^{N,\gamma}(t) = \kappa \sup \left\{ \left| E \left[\sum_{x \in \mathbb{Z}^d/\sqrt{N}} ((1 - \gamma_e) X_t^N(\{x\}) - NX_t^N(\{x\})^2) \varphi^2(x) \right] \right| : X_0^N(1) \leq K \right\}.$$

Let p_t be the transition function for the random walk, Z .

The dual involves the coalescing random walk when the particles coalesce at rate κ when they are at the same site, namely,

$$(9.120) \quad E[X_t(z^1) X_t(z^2)] = E^{(z^1, z^2)}[X_0(Z_t^1) X_0(Z_t^2)].$$

Consider two random walks Z^1 and Z^2 and let A_t be the event that they have coalesced by time t and $A := \cup_{t \geq 0} A_t$. Therefore for $\varphi \in C_c^2(\mathbb{R}^d)$,

(9.121)

$$\begin{aligned} NE[X_t^N(\{i/\sqrt{N}\})^2] &= \frac{1}{N} E[X_{Nt}(i)^2] \\ &= \frac{1}{N} E^i[X_0(Z_{Nt}^1); A_{Nt}] + \frac{1}{N} E^i[X_0(Z_{Nt}^1) X_0(Z_{Nt}^2) : A_{Nt}^c]. \end{aligned}$$

By the CLT, there exists a constant $C < \infty$ such that

$$(9.122) \quad p_t(0, j) \leq \left(\frac{1}{t^{d/2} \wedge 1}\right) \cdot C, \quad j \in \mathbb{Z}^d, t > 0.$$

Then since $\frac{1}{N} \sum_j X_0^N(j) \leq K$, $X_0^N(j) \leq NK$,

$$\begin{aligned} (9.123) \quad & \sum_i \frac{1}{N} E^i[X_0(Z_{Nt}^1)X_0(Z_{Nt}^2)]\varphi(i/\sqrt{N})^2 \\ & \leq \frac{1}{N} \sup_{j,k \in \mathbb{Z}^d} p_{Nt}(i,j)p_{Nt}(i,k)X_0^N(j)X_0^N(k) \\ & \leq \frac{C}{N} t^{-d} N^{-d} N^2 K^2 \leq .CC_\varphi K^2 t^{-d} N^{1-d/2}, \end{aligned}$$

where C_φ depends only on φ .

Hence by dominated convergence it suffices to show that

$$\begin{aligned} (9.124) \quad & \sup_i \tilde{\varepsilon}^{N,\gamma,i}(t) \\ & := N^{\frac{d}{2}-1} \sup_i \left\{ |E^i[X_0(Z_{Nt}^1); A_{Nt}^c] - \gamma_e E^i[X_0(Z_{Nt}^1)]| : X_0^N(1) \leq K \right\} \xrightarrow[N \rightarrow \infty]{} 0 \end{aligned}$$

The intuitive reason for this is that

$$\lim_{T \rightarrow \infty} P[A_T^c] = \gamma_e$$

and the distribution of Z_{Nt}^1 and the conditional distribution of Z_{Nt}^1 given A_{Nt}^c are close.

To make this precise, let $\delta > 0$ and fix $T_0 > 0$ be such that

$$(9.125) \quad |P(A_T^c) - \gamma_e| \leq \frac{(T/2)^{d/2}}{C} \delta, \text{ for all } T \geq T_0$$

with C as in (9.122).

We next obtain an upper bound on the probability that Z^1 and Z^2 coalesce between times T and tN and end at time tN at j .

Noting that

$$(9.126) \quad P^i[A_{Nt} \cap A_T^c \cap \{Z_{Nt}^1 = j\}] = 1 - E^i[e^{-\kappa \int_T^{Nt} 1(Z_r^1 = Z_r^2) dr} 1(Z_{Nt}^1 = j)]$$

and using Jensen's inequality we have

$$(9.127) \quad P^i[A_{Nt} \cap A_T^c \cap \{Z_{Nt}^1 = j\}] \leq \kappa E^i \left[\int_T^{Nt} 1(Z_r^1 = Z_r^2) dr \cdot 1(Z_{Nt}^1 = j) \right].$$

Therefore

$$\begin{aligned} (9.128) \quad & P^i[A_{Nt} \cap A_T^c \cap \{Z_{Nt}^1 = j\}] \\ & \leq \kappa \int_T^{Nt} dr \sum_{k \in \mathbb{Z}^d} p_r(i, k) p_r(i, k) p_{Nt-r}(k, j) \\ & \leq \kappa p_{Nt}(i, j) \int_T^{Nt} dr \sup_k p_r(0, k) \\ & \leq \kappa C^2 t^{-d/2} N^{-d/2} \int_T^{Nt} r^{-d/2} dr \\ (9.129) \quad & \leq \frac{2\kappa C^2 t^{-d/2}}{d-2} T^{1-d/2} N^{-d/2}. \end{aligned}$$

Choosing T_0 large enough we can also assume that

$$(9.130) \quad \sup_{j \in \mathbb{Z}^d} P^i[A_{Nt} \cap A_T^c \cap \{Z_{Nt}^1 = j\}] \leq \delta N^{-d/2}, \quad T \geq T_0.$$

Let $R > 0$ be such that

$$(9.131) \quad P^i[|Z_{T_0}^1| > R] < \frac{\delta}{(1 + \gamma_e)(1 + (2/t)^{d/2}C)}.$$

Using (9.122) and the Markov property at time T_0 , we get for $N \geq 2T_0/t$

$$(9.132) \quad (1 + \gamma_e)P^i[|Z_{T_0}^1| > R; Z_{Nt}^1 = j] \leq \delta N^{-d/2}.$$

Using the CLT again there exists $N_0 \geq 2T_0/t$ such that for all $N_0 \geq N$ and $|k| < R$

$$(9.133) \quad |p_{Nt-T_0}(k, j) - p_{Nt-T_0}(0, j)| < \frac{\delta}{1 + \gamma_e} N^{d/2}.$$

Combining (9.130), (9.132), (9.133), (9.122), (9.131) and (9.125) and using the Markov property we get, for $N \geq N_0$.

$$\begin{aligned} (9.134) \quad & |P^i[Z_{Nt}^1 = j; A_{Nt}^c] - \gamma_e P^i[Z_{Nt}^1 = j]| \\ & \leq |P^i[Z_{Nt}^1 = j; A_{T_0}^c] - \gamma_e P^i[Z_{Nt}^1 = j]| + \delta N^{-d/2} \\ & \leq \sum_{|k| < R} |P^i[Z_{Nt}^1 = j; Z_{T_0}^1 = k; A_{T_0}^c] - \gamma_e P^i[Z_{Nt}^1 = j; Z_{T_0}^1 = k]| + 2\delta N^{-d/2} \\ & \leq \sum_{|k| < R} p_{Nt-T_0}(k, j) |(P^i[Z_{T_0}^1 = k; A_{T_0}^c] - \gamma_e P^i[Z_{T_0}^1 = k])| + 2\delta N^{-d/2} \\ & \leq |P^i[|Z_{T_0}^1| < R; A_{T_0}^c] - \gamma_e P^i[|Z_{T_0}^1| < R]| \cdot p_{Nt-T_0}(0, j) + 3\delta N^{-d/2} \\ & \leq |P^i[A_{T_0}^c] - \gamma_e| \cdot C(2/t)^{d/2} N^{-d/2} + 4\delta N^{-d/2} \\ (9.135) \quad & \leq 5\delta N^{-d/2}. \end{aligned}$$

Recall that

$$(9.136) \quad \tilde{\varepsilon}_K^{N, \gamma_e, i}(t) = N^{d/2-1} \sup \sum_{j \in \mathbb{Z}^d} X_0(j) |[P^i(Z_{Nt}^1 = j; A_{Nt}^c)] - \gamma_e [P^i(Z_{Nt}^1 = j)]|$$

Since the estimate holds for all j , and $\sum X_0(j) \leq NK$,

$$(9.137) \quad \limsup_{N \rightarrow \infty} \tilde{\varepsilon}_K^{N, \gamma_e, i}(t) \leq 5K\delta.$$

Since $\delta > 0$ was arbitrary, (9.124) follows.

9.7.3 Applications of the SBM Invariance Principle

Critical parameter of the contact process

Durrett and Perkins [186] used this to obtain sharp asymptotics for the critical parameter of the long-range contact process which improve upon the results of Bramson, Durrett and Swindle (1989) [54].

Extinction time for a voter model cluster

Under the assumptions of Theorem 9.34 Cox-Perkins (2004) [100] prove that

$$(9.138) \quad \lim_{N \rightarrow \infty} P(\tau_0^N > t) = P(\tau_0 > t) = 1 - \exp\left(-\frac{X_0(1)}{t\gamma_d}\right), \quad t > 0$$

where

$$(9.139) \quad \tau_0^N = \inf\{t > 0 : X_t^N(1) = 0\}, \quad \tau_0 = \inf\{t > 0 : X_t(1) = 0\}.$$

Note that this does not immediately follow from the invariance principle since τ_0 is not a continuous function on $D_{M_F}(\mathbb{R}^d)([0, \infty))$.

Lotka-Volterra models

In a series of papers Cox, Durrett and Perkins obtain deep results on the region of survival and coexistence for the Lotka-Volterra model of competing species - see section ???. Recently,

Voter model clusters

Bramson, Cox and LeGall [57] showed that the rescaled voter model cluster on \mathbb{Z}^d in dimensions $d \geq 2$ conditioned on non-extinction at time t converges to the canonical measure of SBM and that the rescaled support converges (wrt Hausdorff metric) to the support of the super-Brownian excursion.

9.7.4 Spatial Lotka-Volterra models

Neuhauser and Pacala [454] introduced a stochastic spatial model for the competition between two species based on the Lotka-Volterra equations. Their model takes into account the local competition between species but allows for spatial segregation of the species. The model is an interacting particle model on \mathbb{Z}^d with state space $\{0, 1\}^{\mathbb{Z}^d}$ and dynamics given by a perturbed voter model with rates dependent at a site on the density of the two types in a neighbourhood of the site. This models the effects of short range spatial dispersion and demographic stochasticity as well as the role of *interspecific and intraspecific competition*.

Neuhauser-Pacala (1999) [454] prove that the local competitive interactions reduce the parameter region where coexistence occurs in the classical mean-field (deterministic) model and spatial segregation of the species in parts of the parameter region where the classical model predicts coexistence.

Cox and Perkins (2007), (2008) and Cox, Durrett and Perkins use a super-Brownian invariance principle for the perturbed voter model to obtain more precise information on the coexistence region for $d \geq 3$ and new results on the survival region for $d \geq 2$.

9.7.5 Application of SBM to a spatial epidemic model

The SIR epidemic model consider a population with S susceptible, I infected and R removed individuals. Recall that the deterministic model is given by the system of ODE

$$\frac{dS}{dt} = -\beta SI, \quad \frac{dI}{dt} = \beta SI - \gamma I, \quad \frac{dR}{dt} = \gamma I$$

The Reed-Frost stochastic model is defined as follows:

$$(9.140) \quad S_{t+1} \sim \text{Bin}(S_t, (1-p)^{I_t}), \quad 1-p = e^{-\lambda/n}, \quad S_0 = N$$

Martin-Löf (1998) proved that this has the *critical threshold* $\lambda = 1$. Moreover if $I(0) = N^\alpha$, then the scaling limit as $N \rightarrow \infty$ has a phase transition at $\alpha = 1/3$ (Martin-Löf (1998) [428], Aldous (1997) [7]).

We now consider a spatial analogue of the Reed-Frost model due to Lalley (2008) [390]. At each site on \mathbb{Z}^d there are N individuals (types SIR). Infected models remain infected one unit of time and then recover and become immune. At time $t = 0, 1, 2, \dots$, for each pair (i_x, y_y) (infected at x and susceptible at y) the susceptible individual at y becomes infected with probability $p_N(x, y)$. Assume that $p_N(\cdot, \cdot)$ is nearest neighbour simple random walk. Scale the village size N so that the expected number of infections by a contagious individual in a healthy population is 1 so that the epidemic is critical, that is

$$(9.141) \quad p_N(x, y) = \frac{1}{(2d+1)N} \quad \text{if } |x - y| = 1, \quad = 0 \text{ otherwise.}$$

Theorem 9.42 (Lalley (2008) [390]) Let $Y_t^N(x)$ be the number infected at time t at site x in a critical SIR epidemic with village size N and initial configuration $Y_0^N(x)$. Fix $\alpha > 0$ and let $X^N(t, x)$ be the renormalized particle density function process obtained by linear interpolation in x from the valued

$$(9.142) \quad X^N(t, x) = \frac{Y_{[N^\alpha t]}^N(N^{\alpha/2}x)}{N^{\alpha/2}} \quad \text{for } x \in \mathcal{Z}/N^{\alpha/2}.$$

Assume that there is a compact interval J such that the initial particle density functions $X^N(0, x)$ have support in J and assume that $X^N(0, x)$ converge in $C_b(\mathbb{R})$ to a function $X(0, x)$. Then as $N \rightarrow \infty$

$$(9.143) \quad X^N(t, x) \Rightarrow X(t, x)$$

where $X(t, x)$ is the density of a SBM with initial density $X(0, x)$ and killing rate θ where

$$(9.144) \quad \begin{aligned} \theta(x, t) &= 0 && \text{if } \alpha < \frac{2}{5} \\ \theta(x, t) &= \int_0^t X(x, s) ds && \text{if } \alpha = \frac{2}{5} \end{aligned}$$

The idea is that with this scaling up to the extinction time there should be about $O(N^{\alpha/2})$ particles per site and that since the extinction time is $O(N^\alpha)$ the total attrition rate (removed individuals) per generation is $O(N^{5\alpha/2})$. Therefore if $\alpha = 2/5$ then a non-trivial proportion of the population is removed.

Lalley and Zheng [391] have obtained similar results in dimensions 2 and 3 using the absolute continuity of super-Brownian motion local time in these dimensions. Lalley, Perkins and Zheng (2009) establish the existence of a phase transition in 2 and 3 dimensions.

Chapter 10

Spatial systems in large space and time scales

In this chapter we consider critical spatial branching systems and interacting neutral Fleming-Viot processes in large space and time scales. The behaviour of these systems is determined by potential theoretic properties of the migration process such as transience or recurrence. We begin with a brief review of some basic notions.

10.1 Migration processes on Abelian groups

In this section we give a brief review of the basic notions of random walks and Lévy processes on groups on abelian groups following [145].

Let S be a locally compact (additive) Abelian group with countable base and with Haar measure ρ . A discrete time random walk, $\{W_n\}_{n \in \mathbb{Z}_+}$, is prescribed by a transition function

$$P(x, dy) := P(W_{n+1} \in dy | W_n = x) = p(d(y - x))$$

where p is a probability measure on S . The corresponding k -step transition function is

$$P^k(x, dy) := P(W_{n+k} \in dy | W_n = x).$$

A continuous time random walk $\{W_t : t \geq 0\}$ with jump rate 1 is then defined by the transition function

$$P_t(x, dy) := P_x(W_t \in dy), \quad t \geq 0,$$

$$P_t(x, dy) = \sum_{k=0}^{\infty} \frac{e^{-t} t^k}{k!} P^k(x, dy).$$

A natural generalization of continuous time random walks is the notion of Lévy process.

Definition 10.1 A S -valued process $\{X_t : t \geq 0\}$ is a Lévy process if it is stochastically continuous and has stationary and independent increments.

We associate to a Lévy process a semigroup $\{T_t : t \geq 0\}$ on $\mathcal{B}_c(S)$, the space of bounded measurable functions on S with compact support, as follows:

$$T_t \varphi(x) = E_x(\varphi(X_t)),$$

The Green potential of X is the operator

$$G\varphi = \int_0^\infty T_t \varphi dt, \quad \varphi \in \mathcal{B}_c(S).$$

The *fractional operator powers* of G are given by

$$G^\zeta \varphi = \frac{1}{\Gamma(\zeta)} \int_0^\infty t^{\zeta-1} T_t \varphi dt, \quad \zeta > 0 \quad \varphi \in \mathcal{B}_c(S).$$

10.1.1 Transience-Recurrence Properties

In order to review the definitions of transience and recurrence, (following [146]) we consider the *last exit time*, L_A , of X from a non-empty set A defined by

$$L_A := \sup\{t > 0 : X_t \in A\} \quad (\text{if } \{t > 0 : X_t \in A\} \neq \emptyset)$$

Definition 10.2 *The Lévy process X_t on S is transient if for any compact set K*

$$P(L_K < \infty) = 1.$$

and recurrent if it is not transient.

The following result in the spirit of Sato and Watanabe [?], [?] is the basis for a finer classification of the transience properties of random walks in terms of the moments of last exit times.

Proposition 10.3 *Assume that X_t is transient, for any compact set $K \subset S$*

$$\sup_{x \in K} G1_K(x) < \infty$$

and for any compact set C contained in the interior of K

$$\inf_{x \in C} G1_K(x) > 0.$$

Then there exist positive constants c_1 and c_2 such that for all $\zeta > 0$ and $x \in S$

$$c_1 G^{\zeta+1} 1_C(x) \leq E_x L_C^\zeta \leq c_2 G^{\zeta+1} 1_K(x).$$

Proof. See [146], Proposition 2.2.1. ■

Definition 10.4 *The degree of transience, γ , of a transient Lévy process X is defined by*

$$\gamma := \sup\{\zeta > 0 : E_0 L_K^\zeta < \infty \quad \text{for all compact } K\},$$

or equivalently

$$\gamma := \sup\{\zeta > 0 : G^{\zeta+1} \varphi < \infty \text{ for } \varphi \in C_c^+(S)\}$$

where $C_c^+(S)$ denotes the space of nonnegative continuous functions on S with compact support.

Remark 10.5 Sato and Watanabe introduced the set

$$(10.1) \quad \mathcal{T} := \{\zeta > 0 : E_0 L_K^\zeta < \infty \text{ for all compact } K\}.$$

In [146] we consider the extended set

$$\mathcal{T} := \{\zeta > -1 : \int_1^\infty t^\zeta T_t \varphi dt < \infty \text{ for all } \varphi \in C_c^+(S)\},$$

and we call

$$\gamma := \sup\{\zeta > -1 : \zeta \in \mathcal{T}\}$$

the degree of the process. This coincides with the degree of transience if $\gamma > 0$, and if $-1 < \gamma < 0$, we call γ the degree of recurrence of the process.

Given $k \in \mathbb{Z}_+$, the process is said to be (cf. [140])

k - strongly transient if $k \in \mathcal{T}$, and

k - weakly-transient if $k-1 \in \mathcal{T}$ and $k \notin \mathcal{T}$.

Remark 10.6 The degree of transience can be viewed as a generalization of the notion of “critical dimension”. Note that $G^{\zeta+1}\varphi$ at $\zeta = \gamma$ can be either finite or infinite - we will give examples of both possibilities below.

10.1.2 Random walks and Lévy processes in \mathbb{R}^d .

In this section we briefly review the classical results on random walks and Lévy processes in \mathbb{Z}^d and \mathbb{R}^d .

First recall that symmetric nearest neighbour random walks in \mathbb{Z}^d are recurrent in dimensions $d = 1, 2$ and transient in dimensions $d \geq 3$. Moreover, since the rate of decay of the transition probabilities for simple symmetric d -dimensional random walk is $p_t(0, 0) \sim \text{const.} t^{-d/2}$, its degree is $\gamma = d/2 - 1$.

We next recall the classical characterization of Lévy processes in \mathbb{R}^d .

Theorem 10.7 (Lévy-Khintchine representation) A Lévy process in \mathbb{R}^d has the characteristic function (i.e. Fourier transform)

$$\begin{aligned} & E[e^{i(z, X_t)}] \\ &= \exp \left[t \left(-\frac{1}{2}(z, Az) + \int_{\mathbb{R}^d} (e^{i(z, x)} - 1 - i(z, x)1_{\{|x| \leq 1\}}(x)) \nu(dx) + i(m, z) \right) \right] \end{aligned}$$

where A is a symmetric nonnegative definite $d \times d$ matrix, ν is a measure on $\mathbb{R}^d \setminus \{0\}$ satisfying $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}^d} (|x|^2 \wedge 1) \nu(dx) < \infty$, and $m \in \mathbb{R}^d$.

For the proof see [499].

The case $A = Id$, $\nu = 0$, $m = 0$ is the standard Brownian motion and the case $A = 0$, $m = 0$ and $\nu(dx) = |x|^{-\alpha-d} dx$, is the symmetric α -stable process.

Proposition 10.8 For the α -stable process on \mathbb{R}^d the degree is

$$(10.2) \quad \gamma = \frac{d}{\alpha} - 1$$

and in this case

$$\int_0^t s^\gamma T_s \varphi ds \sim \text{const.} \cdot \log t \rightarrow \infty$$

as $t \rightarrow \infty$.

The distribution of jumps of the α -stable process has “long tails”.

10.2 The Persistence-Extinction Dichotomy for Critical Branching Systems

Consider the super-Brownian motion in \mathbb{R}^d with initial measure $X_0 = m\lambda$, $m > 0$ where λ is Lebesgue measure. If $\sup |\phi(x)| \cdot (1 + |x|^2)^{\frac{p}{2}} < \infty$, then the solution, $v_t(x) = V[\phi](t, x)$, to

$$(10.3) \quad \frac{\partial v_t}{\partial t} = Av_t - \frac{\gamma}{2}v_t^2,$$

with $A = \frac{\Delta}{2}$ is integrable and integrating both sides with respect to Lebesgue measure gives

$$\int v_t(x)dx = \frac{\gamma}{2} \int_0^t \int v_s^2(x)dxds.$$

Therefore the large time limit of the Laplace functional

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbb{P}_{m\lambda}(\exp(-X_t(\phi))) &= \lim_{t \rightarrow \infty} \exp\left(-m \int v_t(x)dx\right) \\ &= \lim_{t \rightarrow \infty} \exp\left(-\frac{m\gamma}{2} \int_0^t \int v_s^2(x)dxds\right) \end{aligned}$$

exists for every $\phi \in \mathcal{B}_+$ since the right side is monotone in t . Therefore X_t converges in distribution as $t \rightarrow \infty$ to a random measure on \mathbb{R}^d with probability law which we denote by \mathbb{P}_m^{eq} . Replacing ϕ by $\theta\phi$, $\theta > 0$, and evaluating the first and second derivatives with respect to θ at $\theta = 0$, we can verify that the first and second moments are given by

$$\mathbb{P}_{m\lambda}(X_t(\phi)) = m \int \phi(x)dx$$

and

$$\begin{aligned} \mathbb{P}_{m\lambda}(X_t(\phi)^2) &= m^2 \left(\int \phi(x)dx \right)^2 + \gamma m \int_0^t \int \left(\int p_s(y-z)\phi(z)dz \right)^2 dyds \\ &= m^2 \left(\int \phi(x)dx \right)^2 + \gamma m \int_0^t \left(\int p_{2s}(z_1-z_2)\phi(z_1)\phi(z_2)dz_1dz_2 \right) ds. \end{aligned}$$

Recalling that for the Brownian motion transition kernel $\int_0^\infty p_s(z)ds$ diverges if $d = 1, 2$ and is given by $\frac{2c_d}{|z|^{d-2}}$ if $d \geq 3$, we obtain

$$\begin{aligned} \mathbb{P}_{m\lambda}(X_t(\phi)^2) &\uparrow \infty \text{ if } d = 1, 2 \\ &\uparrow m^2 \left(\int \phi(x)dx \right)^2 + \gamma mc_d \int \int \frac{\phi(z_1)\phi(z_2)}{|z_1 - z_2|^{d-2}} dz_1 dz_2 \text{ if } d \geq 3. \end{aligned}$$

If $d \geq 3$, the above imply that $\{X_t(\phi)\}_{t \geq 0}$ are uniformly integrable and $\mathbb{P}_{m\lambda}(X_\infty(\phi)) = m\lambda(\phi)$, that is, the limiting equilibrium random measure \mathbb{P}_m^{eq} has the same intensity, m , as the initial intensity - this behaviour is called *persistence*. Bramson, Cox and Greven (1997) [56] proved that $\{\mathbb{P}_m^{eq} : m \in [0, \infty)\}$ is in fact the set of all extremal invariant measures.

Theorem 10.9 [108] *Let X_∞ denote the equilibrium random measure for super-Brownian motion in \mathbb{R}^d with mean measure $E(X_\infty(A)) = \lambda(A)$. Let*

$$(10.4) \quad \langle X_\infty^K, \phi \rangle = \int \phi\left(\frac{x}{K}\right) X_\infty(dx),$$

and

$$(10.5) \quad V(\phi) := \gamma \left(\int \int |z_1 - z_2|^{-(d-2)} \phi(z_1) \phi(z_2) dz_1 dz_2 \right).$$

Then the rescaled fluctuations

$$(10.6) \quad \frac{\langle X_\infty^K, \phi \rangle - \langle \lambda, \phi \rangle}{K^{\frac{d+2}{2}} V(\phi)} \Rightarrow Z_\infty$$

where Z_∞ is the Gaussian free field, that is, the Gaussian random field with covariance kernel

$$(10.7) \quad \frac{1}{|x - y|^{d-2}}.$$

The divergence of the second moment in the low dimensional case suggests that the behaviour is qualitatively different in these dimensions. It was proved in Dawson (1977) [108] that in this case the spatially homogeneous super Brownian motion with $X_0 = m\lambda$ suffers local extinction, that is, $X_t(A) \rightarrow 0$ in probability as $t \rightarrow \infty$ for any bounded set A . Iscoe (1986b) [311] has shown that $X_t(A) \xrightarrow{a.s.} 0$ for any bounded set if $d = 1$ and that this result is false if $d = 2$. In dimensions $d = 1, 2$ Bramson, Cox and Greven ([55]) have established that δ_0 is the only measure which is invariant for the process X_t and that for any locally finite initial measure the system undergoes local extinction or explodes thus ruling out the possibility of an invariant measure with infinite mean.

10.2.1 Clumping in Low Dimensions

In order to describe the low dimensional behavior of X_t with $X_0 = \lambda$ (Lebesgue) in more detail we introduce the space-time-mass rescaling

$$\begin{aligned} X_t^{K,\xi}(A) &:= K^{-\xi} X_{Kt}(K^{\frac{\xi}{d}} A) \\ X_0^{K,\xi}(A) &= |A|. \end{aligned}$$

Then

$$\begin{aligned} \mathbb{P}_\lambda(\exp(-X_t^{K,\xi}(\phi))) &= \exp(-\lambda(V_{Kt}\phi_K)) \quad \text{with} \\ \phi_K(x) &:= K^{-\xi} \phi(K^{-\frac{\xi}{d}} x). \end{aligned}$$

Note that

$$\tilde{v}(t, x) := K^\xi V_{Kt} \phi_K(K^{-\frac{\xi}{d}} x)$$

satisfies

$$\begin{aligned} \frac{\partial \tilde{v}(t, x)}{\partial t} &= K^{1-\frac{2\xi}{d}} \Delta \tilde{v}(t, x) - \frac{\gamma}{2} K^{1-\xi} \tilde{v}(t, x)^2 \\ \tilde{v}(0, x) &= \phi(x) \end{aligned}$$

and therefore $X^{K,\xi}$ is equivalent to a super Brownian motion with ‘‘diffusion coefficient’’ $K^{1-\frac{2\xi}{d}}$ and ‘‘branching coefficient’’ $\frac{\gamma}{2} K^{1-\xi}$. The branching term dominates in the $K \rightarrow \infty$ limit and the diffusion term dominates in the $K \rightarrow 0$ limit if $d < 2$ and the opposite occurs if $d > 2$.

Theorem 10.10 (Dawson and Fleischmann 1988) [115] (a) Let $d < 2$. Then $X^{K,\xi} \xrightarrow{K \rightarrow \infty} 0$ if $\xi < 1$ and $X^{K,\xi} \xrightarrow{K \rightarrow \infty} \lambda$ if $\xi > 1$

(b) If $d = 1$ and $\xi = 1$, then X_K converges in distribution as $K \rightarrow \infty$ to the pure atomic process $\{X_t^0\}_{t \geq 0}$ in which X_t^0 is Poisson with intensity $(\frac{\gamma}{2}t)X(0)$ and the mass of each atom evolves according to a Feller continuous state branching.

(c) If $d = 2$, then $X^{K,1} \xrightarrow{D} X$, that is X is self-similar.

Remark 10.11 (b) suggests that for $d = 1$ at time K there are clumps of size K with interclump distance K .

In the case $d = 2$, the phenomenon of *diffusive clustering* arises. This is made precise in the following result of Klenke.

Theorem 10.12 (Klenke (1997) [[375], Theorem 2]) Let $d = 2$, and $I = (-\infty, 1]$. For $\alpha \in I$, let

$$X_t^\alpha(B) := t^{-\alpha} X_t(t^{\alpha/2}B).$$

Then in the sense of finite dimensional distributions

$$\mathcal{L}^{\frac{(\log t)\lambda}{8\pi}}[\{X_t^\alpha(B)\}_{\alpha \in I}] \xrightarrow{t \rightarrow \infty} \mathcal{L}^1[\{Z_{1-\alpha}\}_{\alpha \in I} \cdot \lambda(B)]$$

where Z is a FB process with $Z_0 = 1$.

In the case $d = 2$ Theorem 10.10 (c) provides a link between the small scale and large scale behaviours. In particular it implies that

$$X_{Kt}(B(0, 1)) \xrightarrow{D} \frac{X_t(B(0, K^{-1/2}))}{K^{-1}}.$$

For $t > 0$ the left side goes to zero in probability as $K \rightarrow \infty$ because of the local extinction result which then shows that the local density at time t is 0 which implies that it does not have a non-trivial absolutely continuous component.

10.2.2 Ergodic Behaviour

The extinction-persistence result implies that if ϕ has compact support, then $X_t(\phi)$ converges to zero in probability if $d \leq 2$ and converges in distribution to a non-degenerate limit if $d > 2$. This can be extended to an ergodic theorem in the latter case.

Theorem 10.13 (Iscoe (1986b) ([311]), Fleischmann and Gärtner (1986) ([238])).

- (a) For $d > 2$ with probability one, $\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t X_s ds = \lambda$ (in the vague topology).
- (b) For $d = 2$, as $t \rightarrow \infty$ $\frac{1}{t} \int_0^t X_s ds$ converges a.s. in the vague topology to $\eta\lambda$ where η is a non-degenerate infinitely divisible random variable with mean one.

Remark 10.14 (b) implies that in critical dimension $d = 2$, $X_t(\phi)$ goes to zero in probability but there are arbitrarily large values of t for which $X_t(\phi)$ is large, therefore the large clumps revisit bounded sets at arbitrarily large times, that is, they have a recurrence property.

Remark 10.15 See the series of papers of Bojdecki, Gorostiza and Talarczyk [42], [43], [44], [45] for recent advances in the classification of the occupation time fluctuation structure of branching systems.

10.3 The Equilibrium Clan Decomposition

In this subsection we consider the structure of equilibrium states of SBM following the development in Dawson-Perkins [118], [186].

10.3.1 An extension of Historical Brownian motion

We will consider the historical process with time parameter in $(-\infty, \infty)$. Let $C_{s,t}^d = C([s, t), \mathbb{R}^d)$, $C^d = C_{-\infty, \infty}^d$ and $(C^d)^s = \{y \in C^d : y(\cdot) = y(\cdot \wedge s)\}$. Also let $M_p(C_{s,t}^d)$ denote the space of measures on $C_{s,t}^d$ with marginal distributions in $M_p(\mathbb{R}^d)$. Let μ_λ^t denote the law (on C^d) of $\bar{\xi}_t := \{\xi(s \wedge t)\}_{s \in (-\infty, \infty)}$ when $\{\xi(t) : t \in \mathbb{R}\}$ is a Brownian motion with 0-marginal given by Lebesgue measure. The transition Laplace functional is given by: for $\phi \in p\mathcal{B}(C^d)$

$$\mathbb{Q}_{s,m}(\exp(-H_t(\phi))) = \exp\left(-\int V_{s,t}\phi(y)m(dy)\right), \quad m \in M_p(C^d)^s$$

where

$$V_{s,t}\phi(y) = S_{s,t}\phi(y) - \frac{\gamma}{2} \int_s^t S_{s,r}((V_{r,t}\phi)^2)dr.$$

If $H_s = \mu_\lambda^s$, then H_t is an infinitely divisible random measure with canonical measure $\int R_{s,t}(y, A)\mu_\lambda^s(dy)$ where R is characterized by

$$(10.8) \quad V_{s,t}\phi(y) = \int (1 - e^{-\nu(\phi)})R_{s,t}(y, d\nu).$$

Let $s < u < t$. By the Historical Cluster Representation $H_t(\{y : y^u \in \cdot\})$ (under $\mathbb{Q}_{s,m}$) can be represented as a Poisson random field of clan measures in $M_p(C_{s,t}^d)$ with Poisson intensity $\frac{2}{\gamma(t-u)}H_u$. The typical clan measure $\Xi_t(y')$ can be interpreted as the descendent population from an individual alive at time u with history y' and $r_{s,u}\Xi_t = \Xi_t(C^d)\delta_{y'}$ where for $\mu \in M_F(C_{s,t}^d)$, $s \leq u \leq t$, $r_{s,u}\mu(A) := \mu(\{y : y(\cdot \wedge u) \in A\})$.

10.3.2 The historical process conditioned to live forever

For $t > s$ let H_t^I denote a realization of the historical process H_t conditioned to stay alive forever starting from a finite initial measure $H_s = \eta_s \in M_p(C_{-\infty,s}^d)$. The law $\mathbb{Q}^I(H_t^I \in \cdot)$ is defined rigorously as the weak limit as $T \rightarrow \infty$ of $\mathbb{Q}_{s,\eta_s}(H_t \in \cdot | H_T \neq 0)$. Using Remark 9.26 we have

$$\begin{aligned} \mathbb{Q}_{s,\eta_s}^I(H_t^I \in A) &= \lim_{T \rightarrow \infty} \mathbb{Q}_{s,\eta_s}(H_t \in A | H_T \neq 0) \\ &= \lim_{T \rightarrow \infty} \frac{\mathbb{Q}_{s,\eta_s}(1_A(H_t)\mathbb{Q}_{t,H_t}(H_T \neq 0))}{\mathbb{Q}_{s,\eta_s}(H_T \neq 0)} \\ &= \lim_{T \rightarrow \infty} \frac{\mathbb{Q}_{s,\eta_s}(1_A(H_t)(1 - e^{-\frac{2H_t(1)}{\gamma(T-t)}}))}{1 - e^{-\frac{2\eta_s(1)}{\gamma(T-s)}}} \\ &= \eta_s(\mathbf{1})^{-1} \mathbb{Q}_{s,\eta_s}(\mathbf{1}_A(H_t)H_t(\mathbf{1})) \end{aligned}$$

The corresponding *normalized Campbell measure* of H_t is

$$\tilde{\mathbb{Q}}_{s,\eta_s}(H_t^I \in A, \bar{\xi}_t \in B) := (\eta_s(\mathbf{1}))^{-1} \mathbb{Q}_{s,\eta_s}[\mathbf{1}_A(H_t)H_t(B)].$$

Let

$$U_{s,t}(\phi, \psi)(y) := \int \nu(\phi) e^{-\nu(\psi)} R_{s,t}(y, d\nu).$$

Lemma 10.16 *Let $\phi, \psi \in \mathcal{B}_+(C)$. Then*

$$\begin{aligned} U_{s,t}(\phi, \psi)(y) &= \frac{\partial}{\partial \theta} V_{s,t}(\theta \phi + \psi)|_{\theta=0} \\ &= P_{s,y} \left(\phi(\bar{\xi}_t) \exp \left(- \int_s^t \gamma V_{u,t} \psi(\bar{\xi}_u) du \right) \right) \end{aligned}$$

Proof.

Since (recall the connections between $R_{s,t}$ and $V_{s,t}$)

$$\int e^{-\nu(\psi)} \nu(\phi) R_{s,t}(y, d\nu) = \frac{\partial}{\partial \theta} V_{s,t}(\theta \phi + \psi)|_{\theta=0},$$

the first equality is clear. Recall

$$V_{s,t} \phi(y) = S_{s,t} \phi(y) - \frac{\gamma}{2} \int_s^t S_{s,r} ((V_{r,t} \phi)^2) dr,$$

and so, replacing ϕ with $\theta \phi + \psi$ and differentiating with respect to θ we get

$$U_{s,t}(\phi, \psi) = S_{s,t} \phi - \gamma \int_s^t S_{s,r} ((V_{r,t} \psi) U_{r,t}) dr.$$

Then by a Feynman-Kac argument for the Brownian path process we get

$$U_{s,t}(\phi, \psi)(y) = P_{s,y} \left(\phi(\bar{\xi}_t) \exp \left(- \int_s^t \gamma V_{u,t} \psi(\bar{\xi}_u) du \right) \right).$$

Remark 10.17 *Let $R_{s,t}^{\bar{\xi}}(y, d\nu)$ denote the Palm measure associated with $R_{s,t}(y, d\nu)$, that is, it is the regular conditional probability on $M_p(C_{s,t}^d)$ for $A \times B \rightarrow \int \int 1_A(\xi) d\nu 1_B(\nu) R_{s,t}(y, d\nu)$ given $\bar{\xi}$. The above Lemma implies that*

$$(10.9) \quad \int \exp(-\nu(\psi)) R_{s,t}^{\bar{\xi}}(y, d\nu) = \exp \left(- \int_s^t \gamma V_{u,t} \psi(\bar{\xi}_u) du \right).$$

Intuitively, $R_{s,t}^{\bar{\xi}}(y, d\nu)$ is the law of a random measure obtained if ξ throws off historical excursions at a constant rate γ on $[s, t]$. We let $\Xi_{s,t}^I(\xi)$ denote such an immortal clan, i.e. a random measure with law $R_{s,t}^{\bar{\xi}}(y, \cdot)$.

Proposition 10.18 *Under $\tilde{\mathbb{Q}}_{s,\eta_s}$,*

- (a) $\{\bar{\xi}_t\}_{t \geq s}$ is a path-valued Brownian motion with initial law $(\eta_s(1)^{-1} \eta_s(dy))$, and
- (b) the regular conditional law, $Q_{s,t}^{\bar{\xi}}$ of H_t^I , given $\bar{\xi}$ has Laplace functional

$$Q_{s,t}^{\bar{\xi}}[\exp^{-< H_t^I, \psi >}] = \int \left(\int \exp(-\nu(\psi)) R_{s,t}^{\bar{\xi}}(y, d\nu) \right) \eta_s(dy) \eta_s(1)^{-1} Q_{s,\eta_s}(\exp(-< H_t, \psi >)).$$

Proof.

$$\begin{aligned}
& \tilde{\mathbb{Q}}_{s,\eta_s} \left(\phi(\bar{\xi}) \exp^{-\langle H_t^I, \psi \rangle} \right) = (\eta_s(\mathbf{1}))^{-1} \mathbb{Q}_{s,\eta_s} [\exp^{-\langle H_t, \psi \rangle} H_t(\phi)] \\
&= -(\eta_s(\mathbf{1}))^{-1} \frac{\partial}{\partial \theta} \mathbb{Q}_{s,\eta_s} (\exp^{-\langle H_t, \theta\phi + \psi \rangle})|_{\theta=0} \\
&= -(\eta_s(\mathbf{1}))^{-1} \frac{\partial}{\partial \theta} \left(\exp^{-\eta_s(V_{s,t}(\theta\phi + \psi))} \right)|_{\theta=0} \\
&= \exp(-\eta_s(V_{s,t}(\psi))) P_{s,\eta_s/\eta_s(1)} \left(\phi(\bar{\xi}_t) \exp\left(-\int_s^t \gamma V_{u,t} \psi(\bar{\xi}_u) du\right) \right) \\
&\quad (\text{use Lemma 10.16}) \\
&= \int P_{(s,y)} \left(\phi(\bar{\xi}_t) \int \exp(-\nu(\psi)) R_{s,t}^{\bar{\xi}}(y, d\nu) \right) \eta_s(dy) / \eta_s(1) Q_{s,\eta_s}(\exp(-\langle H_t, \psi \rangle))
\end{aligned}$$

which proves the result.

Remark 10.19 From the perspective of a typical immortal particle chosen according to H_t^I , H_t^I is the sum of $\Xi_{s,t}^I(\xi)$ and an independent copy of H_t . The former is the contribution to H_t^I of cousins of ξ .

10.3.3 Convergence to equilibrium for the historical process

Recall that if $H_s = \mu_\lambda^s$, then H_t is an infinitely divisible random measure with canonical measure $R_{s,t} = \int R_{s,t}(y, A) \mu_\lambda^s(dy)$ where R is characterized by $V_{s,t}\phi(y) = \int (1 - e^{-\nu(\phi)}) R_{s,t}(y, d\nu)$.

Integrate both sides of (10.9) with respect to μ_λ^s to see that $R_{s,t}^{\bar{\xi}}$ is still the Palm measure of the canonical measure $\int R_{s,t}(y, d\nu) \mu_\lambda^s(dy)$ of H_t .

As $s \rightarrow -\infty$, the total mass of $\Xi_{s,t}^I(\xi)$ goes to infinity (set $\psi = 1$ in Proposition 10.18 and let $s \rightarrow -\infty$). However

Proposition 10.20 (a) As $s \rightarrow -\infty$, $\Xi_{s,t}^I(\bar{\xi})$ converges to a locally finite measure iff $d \geq 3$.
(b) As $s \rightarrow -\infty$, $R_{s,t}$ converges vaguely to a measure $R_{-\infty,t}$ such that $\int \mu R_{-\infty,t}(d\mu)$ is locally finite.

(c) $R_{-\infty,t}$ is “an entrance law” for H in the sense that if H_s has the infinite law $R_{-\infty,s}$ and $t > s$, then H_t has law $R_{-\infty,t}$.

Proof.

(a) The Laplace functional in (10.9) converges as $s \rightarrow -\infty$ by monotonicity. In the case $d \geq 3$, we can obtain the mean measure of $\Xi_{s,0}^I(\bar{\xi})$ for a trajectory ξ as follows. Differentiate the expression for the Laplace functional in (10.9) to see that for ϕ continuous with compact support

$$\lim_{s \rightarrow -\infty} E \left[\int_{\mathbb{R}^d} \phi(y_t) \Xi_{s,t}^I(\bar{\xi}, dy) \right] = \gamma \lim_{s \rightarrow -\infty} \int_{s-t}^0 \int p(|r|, y - \xi(t+r)) \phi(y) dy dr < \infty$$

and hence $\Xi_{-\infty,t}^I$ is a.s. locally finite.

(b) Without loss of generality we can take $t = 0$ and $\psi(y) = \tilde{\psi}(y_0), \phi(y) = \tilde{\phi}(y_0)$, when $\tilde{\phi}, \tilde{\psi}$ are continuous with compact support. Then for $d \geq 3$, the above gives

$$\begin{aligned} & \lim_{s \rightarrow -\infty} \int \int \mu(\psi)\mu(\phi)R_{s,0}(y,d\mu)\mu_\lambda^s(dy) \\ &= \lim_{s \rightarrow -\infty} \int \int \psi(\xi)\mu(\phi)R_{s,t}^{\bar{\xi}}(d\mu)P_{s,\mu_\lambda^s}(d\xi) \\ &= \gamma \lim_{s \rightarrow -\infty} \int \int \int_s^0 p(2|r|, z-y)\phi(z)\psi(y)drdzdy \\ &< \infty. \end{aligned}$$

(c) See [118] Theorem 6.4.

Theorem 10.21 (Dawson-Perkins (1991), [118], Theorem 6.3) *The stationary random measure X_∞ of the super-Brownian motion in \mathbb{R}^d is an infinitely divisible random measure with intensity λ and with canonical measure given by $\hat{R}(B) = R_{-\infty,0}(\{\pi_0\mu \in B\})$ where $\pi_0\mu(A) := \mu(\{\xi : \xi_0 \in A\})$, $A \in \mathcal{B}(\mathbb{R}^d)$, $\mu \in M_p(C^d)$.*

Remark 10.22 *This leads to a description of the super-Brownian motion in equilibrium as a countable collection of infinite clan measures.*

An infinite clan containing an individual located at 0 at time 0, denoted by $\Xi_{-\infty,0}^I$, has law $\int R_{-\infty,0}^{\bar{\xi}}\Pi_0(d\xi)$. It is constructed by running a Brownian trajectory, ξ , backwards to time $-\infty$ and then collecting the mass at time 0 corresponding to $R_{-\infty,0}^{\bar{\xi}}$.

10.3.4 Clan dynamics

The clan $\{\Xi_{-\infty,t}^I\}_{t \geq 0}$ then evolves as a historical Brownian motion with initial condition $\Xi_{-\infty,0}^I$. It is an easy consequence of this description that the infinite clan is self-similar in the sense that

$$(10.10) \quad K^{-2}\Xi_{-\infty,0}^I(\{y : y_0 \in KA\}) \stackrel{\mathcal{D}}{=} \Xi_{-\infty,0}^I(\{y : y_0 \in A\}) \quad \forall A \in \mathcal{B}(\mathbb{R}^d).$$

There has been considerable recent interest in the dynamics of these infinite clans and this has led to a second dimensional dichotomy - namely clan recurrence or transience. To describe this consider the total weighted occupation time that a given clan spends in the unit ball at the origin $B(0,1)$. In dimensions $d > 2$, (10.10) and spherical symmetry implies that the infinite clan of an individual at the origin and has mean density proportional to

$$(10.11) \quad \frac{1}{|x|^{d-2}}.$$

Therefore if we exclude the clan mass initially in a ball of radius 2 (which by the extinction property of Feller branching has a.s. finite lifetime), then

$$\begin{aligned} E[\int_0^\infty & \Xi_{-\infty,t}^I(\{y : y_t \in B(0,1), y_0 \notin B(0,2)\})dt] \\ &= \int_0^\infty \int_{B(0,2)^c} \frac{1}{|y|^{d-2}} \left[\int_{B(0,1)} p_t(x-y)dx \right] dy dt \\ &= \int_{B(0,2)^c} \left[\int_{B(0,1)} \frac{1}{|x-y|^{d-2}} dx \right] \frac{1}{|y|^{d-2}} dy \\ &< \infty \text{ iff } d \geq 5. \end{aligned}$$

Hence in dimensions $d \geq 5$,

$$(10.12) \quad \int_0^\infty \Xi_{-\infty,t}^I(\{y : y_t \in B(0,1)\}) dt < \infty, \text{ a.s.}$$

- this behaviour is called *clan transience*. Analogous results of Stöckl and Wakolbinger [530] show that the corresponding infinite clan for a branching particle system in dimensions $d \geq 5$ gives positive mass to the unit ball over only a finite time horizon.

10.4 Neutral Stepping Stone Models

10.4.1 The two type stepping stone model

The neutral two type stepping stone model on a countable abelian group S with migration kernel $p(\cdot)$ is given by the system

$$\begin{aligned} dX_t(x) &= \sum_{y \in S_1} p_{y-x}(X_t(y) - X_t(x)) dt \\ &\quad + \sqrt{2X_t(x)(1-X_t(x))} dW_t(x) \\ x_0(x) &\in [0, 1], \quad x \in S \end{aligned}$$

This process can be embedded in the infinitely many types stepping stone model which we now consider.

10.4.2 The infinitely many types stepping stone model

Consider a collection (finite or countable) of subpopulations (demes), indexed by S . The subpopulation at $\xi \in S$ at time t is described by a probability distribution $X_\xi(t)$ over a space $E = [0, 1]$ of possible types (alleles). In other words, $X_\xi(t) \in \mathcal{P}(E)$, the set of probability measures on E so that the state space is

$$(10.13) \quad (\mathcal{P}([0, 1]))^S.$$

Within each subpopulation there is mutation, selection and finite population sampling. Mutation is assumed to produce a new type chosen by sampling from a fixed source distribution $\theta \in \mathcal{P}(E)$. Selection is prescribed by a fitness function $V(x)$ in the haploid case or by $V(x, y) = V(y, x)$ in the diploid case. Migration from site ξ to site ξ' is assumed to occur via a symmetric random walk with rates $q_{\xi, \xi'} = p(\xi - \xi')$. Finally Fleming-Viot continuous sampling is assumed to take place within each subpopulation. It is a basic property of this model that for any $t > 0$, $X_\xi(t)$ is a purely atomic random measure (with countably many atoms) and therefore can be represented in the form

$$X_\xi(t) = \sum_{k \in I} m_{\xi, k}(t) \delta_{y_k}$$

where $m_{\xi, k}(t) \geq 0$ denotes the proportion of the population in subpopulation ξ of type $y_k \in E$ at time t . Note that in this model two individuals are related if and only if they are of the same type.

We denote the vector $\{\mu_\xi\}_{\xi \in S}$ by $\bar{\mu}$. The generator is then given by

$$(10.14) \quad \begin{aligned} GF(\bar{\mu}) &= c \cdot \sum_{\xi \in S} \int_{[0,1]} \frac{\partial F(\bar{\mu})}{\partial \mu_\xi(u)} (\theta(du) - \mu_\xi(du)) \\ &\quad + \sum q_{\xi,\xi'} \int_{[0,1]} \frac{\partial F(\bar{\mu})}{\partial \mu_\xi(u)} (\mu_{\xi'}(du) - \mu_\xi(du)) \\ &\quad + \frac{\gamma}{2} \sum \int_{[0,1]} \int_{[0,1]} \frac{\partial^2 F(\bar{\mu})}{\partial \mu_\xi(u) \partial \mu_\xi(v)} Q_{\mu_\xi}(du, dv) \\ X_{0,\xi} &= \nu \quad \forall \xi, \quad Q_\mu(du, dv) = \mu(du)\delta_u(dv) - \mu(du)\mu(dv). \end{aligned}$$

The first term corresponds to mutation with source distribution θ , the second to spatial migration and the last to continuous resampling. The resampling rate coefficient γ is inversely proportional to the effective population size of a deme.

This existence and uniqueness of this system of interacting Fleming-Viot processes was established by Vaillancourt [549] and Handa [286].

The questions which we wish to investigate are

- the distribution in a given subpopulation, that is what is the joint distribution of the $\{m_{\xi,k}\}$
- the spatial distribution of relatives
- how are these affected by the migration geometry.

10.4.3 The Dual Process Representation

Given $n \in \mathbb{N}$ consider the collection

$$(10.15) \quad \begin{aligned} \Pi_n &= \{\bar{\eta} := (\eta, \pi)\} : \text{ where} \\ &\pi \text{ is a partition of } \{1, \dots, n\}, \text{ that is,} \\ &\pi : \{1, \dots, n\} \rightarrow \{1, \dots, |\pi|\} \text{ with } |\pi| \leq n, \\ &\eta : \{1, \dots, |\pi|\} \rightarrow S. \end{aligned}$$

Now consider the family of functions in $C((\mathcal{P}([0,1]))^S \times \Pi)$ of the form

$$(10.16) \quad F_f(\bar{\mu}, \bar{\eta}) := \int_{[0,1]} \cdots \int_{[0,1]} f(u_{\pi(1)}, \dots, u_{\pi(n)}) \mu_{\eta_1}(du_1) \cdots \mu_{\eta_{|\pi|}}(du_{|\pi|})$$

with $f \in C([0,1]^n)$.

We now consider a continuous time Markov chain, $\bar{\eta}_t = (\eta_t, \pi_t)$, with state space Π_n and jump rates:

- the partition elements perform continuous time symmetric random walks on S with rates $q_{\xi,\xi'}$ and in addition a partition element can jump to $\{\infty\}$ with rate c (once a partition element reaches ∞ it remains there without change of further coalescence).

- each pair of partition elements during the period they reside at an element of S (but not $\{\infty\}$) coalesce at rate γ to the partition element equal to the union of the two partition elements.

Let H denote the generator of $\bar{\eta}$. Then for a function of the form (10.16)

$$(10.17) \quad HF_f(\bar{\mu}, \bar{\eta}) = GF_f(\bar{\mu}, \bar{\eta}).$$

We then obtain the dual relationship

$$(10.18) \quad E(F_f(X_t, (\eta, \pi))) = E(F_f(X_0, (\eta_t, \pi_t)))$$

and this proves that the infinitely many types stepping stone martingale problem is well-posed.

Remark 10.23 *Given the dual we can construct a spatially structured coalescent that describes the ancestral structure of a sample of a finite number of individuals located at the same or different sites.*

Note that this is essentially equivalent to the coalescent geographically structured populations introduced by developed by Notohara (1990) [460] and Takahata (1991) [540].

Remark 10.24 *Note that as $\gamma \rightarrow \infty$ the dual converges to the dual of the voter model and we can regard the voter model as the limit as $\gamma \rightarrow \infty$ of the interacting Fisher-Wright diffusions.*

10.4.4 Spatial homogeneity and the local-fixation coexistence dichotomy

In this subsection we consider the neutral stepping stone model without mutation.

Theorem 10.25 *(Dawson-Greven-Vaillancourt (1995) [124], Theorem 0.1)*

Let S be a countable abelian group and consider the infinitely many types stepping stone model with no mutation ($c = 0$). Assume that the initial random field $\{X_\xi(0)\}_{\xi \in S}$ is spatially stationary, ergodic, weakly mixing and has single site mean measure satisfying

$$(10.19) \quad E\left(\int g(u)X_\xi(0, du)\right) = \int g(u)\theta(du), \quad \theta \in \mathcal{P}[0, 1].$$

(a) If $q_{\xi, \xi'}$ is a symmetric transient random walk on S , then the stepping stone process $\{X_\xi(t)\}_{\xi \in S}$ converges in distribution to a nontrivial invariant $\mathcal{P}([0, 1])$ -valued random field $\{X_\xi(\infty)\}_{\xi \in S}$ which also has single site mean measure θ . $\{X_\xi(\infty)\}_{\xi \in S}$ is spatially homogeneous (that is, the law is invariant under translations on S), ergodic and weakly mixing, in particular

$$(10.20) \quad E(\langle x_\xi, f \rangle \langle x_\zeta, f \rangle) \rightarrow \langle \mu, f \rangle^2 \text{ as } d(\xi, \zeta) \rightarrow \infty, \quad \forall f \in L_\infty([0, 1]).$$

(b) In (a) the equilibrium state decomposes into countably many coexisting infinite families, namely,

$$(10.21) \quad X_\xi(\infty) = \sum_{k=1}^{\infty} a_{\xi, k} \delta_{y_k}$$

with $\sum_x i a_{\xi, k} = \infty$ for each k .

(c) If p_ξ is recurrent, then the set of invariant measures is a convex set with extremal invariant measures are δ_a , $a \in [0, 1]$, that is, there is local fixation, and

$$(10.22) \quad \mathcal{L}(\{X_\xi(\infty)\}_{\xi \in S}) = \int (\delta_y)^S \theta(dy).$$

Proof. We sketch the main steps of the proof.

The proof uses the dual representation (10.18), (??). Note that $|\pi_t|$ is monotone decreasing so that we can define

$$(10.23) \quad \pi_\infty = \lim_{t \rightarrow \infty} \pi_t, \quad \pi_\infty = \{\pi_\infty(1), \dots, \pi_\infty(n)\}.$$

Then we note that $\hat{\eta}$ is prescribed by a *coalescing random walk with delay*. We let $Z(t)$ be a random walk on S with transition kernel $\{q_{\xi, \xi'}\}$. Since we have assumed that the random walk is symmetric, then the difference process $Z_1(t) - Z_2(t)$, where Z_1, Z_2 are independent copies of the random walk, is a random walk with jump rates $2q_{\xi, \xi'}$. We can assume that the system of coalescing random walks with delay is constructed on a probability space on which the sequence $\{Z_i(t)\}_{i \in \mathbb{N}}$ of independent random walks and an independent collection of exponentially distributed random variables are defined.

Lemma 10.26 *If the q -random walk is recurrent, then*

(a)

$$(10.24) \quad \mathcal{L}(\hat{\eta}_t) - \mathcal{L}((Z(t); \{1, \dots, n\})) \Rightarrow 0 \text{ as } t \rightarrow \infty$$

Given two initial sites 0 and $\xi \neq 0$ and $(\eta, (\{1\}, \{2\})), \eta_1 = 0, \eta_2 = \xi$,

$$(10.25) \quad P(\pi_t = \{1, 2\}) \leq \text{const} \cdot \int_0^t P(Z(s) = 0) ds \leq \frac{\text{const}}{|\xi|^{d-2}}$$

where $Z(s)$ is a random walk starting at ξ and with jump rate 2.

(b) If the q -random walk is transient, then

$$(10.26) \quad \mathcal{L}(\eta_t | |\pi_\infty| = k) - \mathcal{L}(Z_1(t), \dots, Z_k(t)) \Rightarrow 0 \text{ as } t \rightarrow \infty$$

and $P(|\pi_\infty| = 1) < 1$ provided that $|\pi_0| \neq 1$.

Proof. If the random walk is recurrent, then $Z_1(t) - Z_2(t)$ visits 0 infinitely often and therefore they must coalesce with probability one.

If the random walk is transient, then there exists a random time $\sigma < \infty$ a.s. such that σ is the last coalescence time in the system $\hat{\eta}_t$. Denote by $\xi^1(t), \dots, \xi^{|\pi_\infty|}(t)$ the position of the partition elements at time $\sigma + t$. The system η_u for times $u = s + t + \sigma$ behaves like a system of $|\pi_\infty|$ random walks in s starting at $\xi^1(t) \dots \xi^{|\pi_\infty|}(t)$ and conditioned on never meeting. Since for every pair $i \neq j$ $\xi^i(t) - \xi^j(t) \rightarrow \infty$ as $t \rightarrow \infty$, the event that ξ^i and ξ^j never meet after time t tends to one as $t \rightarrow \infty$. It remains to show that the distance between the distributions of the system of $|\pi_\infty|$ independent random walks starting at $(\xi^1(t) \dots \xi^{|\pi_\infty|}(t))$ and starting at $(0, \dots, 0)$ tends to 0 as $s \rightarrow \infty$. This is verified using a coupling by randomized stopping times due to Greven (1987) [261] and a result of Choquet and Deny on transient random walks (see Spitzer [529], Ch 6. T1) - see ([124] for details).

■

We also note the following elementary result on random probability measures.

Lemma 10.27 *Let X_1, X_2 be a random probability measures on $[0, 1]$, having the same mean measures $E(X_i) = \theta \in \mathcal{P}([0, 1])$, that is, a measurable map from a probability space (Ω, \mathcal{F}, P) to $\mathcal{P}([0, 1])$.*

(a) If

$$(10.27) \quad E\left[\left(\int g(y)X_i(dy)\right)^2\right] = E\left[\int g^2(y)X(dy)\right] \quad \forall g \in C([0, 1]),$$

then

$$(10.28) \quad X_i(\omega) = \delta_{y(\omega)} \quad \text{for a.e. } \omega \in \Omega \text{ and } \omega \rightarrow y \text{ is measurable.}$$

(b) If in addition,

$$(10.29) \quad E\left[\left(\int g(y)X_1(dy) \int g(y)X_2(dy)\right)\right] = E\left[\int g^2(y)X_1(dy)\right] \quad \forall g \in C([0, 1]),$$

then

$$(10.30) \quad X_1(\omega) = X_2(\omega) = \delta_{y(\omega)} \quad \text{for a.e. } \omega \in \Omega.$$

We return to the proof of the theorem.

(a) Recurrent Case.

Step 1. Let $m = 2$ and take $f(u_1, u_2) = g(u_1)g(u_2), \eta = (\xi, \xi)$. Then by the dual representation and Lemma 10.26

$$\begin{aligned} & E\left(\int_0^1 g(u)X_\xi(t, du)\right)^2 \\ (10.31) \quad &= E\left(\int g^2(u)X_{\eta_t^1}(0, du)1(\pi_t = \{1, 2\}) \right. \\ & \quad \left. + \int g(u)X_{\eta^1}(t, du) \cdot \int g(u)X_{\eta^2}(t, du)1(\pi_t = \{\{1\}, \{2\}\})\right) \\ &= E\left(\int g^2(u)X_{\eta_t^1}(0, du)\right) + o(t) \end{aligned}$$

Therefore in the limit by Lemma 10.27(a) we have

$$(10.32) \quad X_\xi(\infty, du) = \delta_y, \text{ a.s.}$$

Since $\mathcal{L}(X_\xi(t)) \in \mathcal{P}(\mathcal{P}([0, 1]))$, the set $\{\mathcal{L}(X_\xi(t))\}_{t \geq 0}$ is weakly relatively compact. By (10.38) a weak limit point must be concentrated on

$$(10.33) \quad M = \{\delta_u : u \in [0, 1]\}$$

that is, $\mathcal{L}(X_\xi(\infty)) = \int_0^1 \delta_{\delta_u} H_\xi(du)$ with $H_\xi \in \mathcal{P}([0, 1])$. But we have

$$(10.34) \quad E\langle X_\xi(t), f \rangle = \langle \theta, f \rangle$$

so that for a limit point $\mathcal{L}(\{X_\xi(\infty)\}_{\xi \in S})$

$$(10.35) \quad E\langle X_\xi(\infty), f \rangle = \langle \theta, f \rangle \quad \forall f \in C([0, 1]).$$

Therefore $H_\xi = \theta$

$$(10.36) \quad \mathcal{L}(X_\xi(\infty)) = \int_0^1 \delta_{\delta_u} \theta(du).$$

Step 2. In order to show consensus of the components occurs for $t \rightarrow \infty$ take $m = 2$, $f(u_1, u_2) = g(u_1)g(u_2)$ but use $\eta = (\xi^1, \xi^2)$ with $\xi^1 \neq \xi^2$. Then again using Lemma 10.26

$$\begin{aligned}
& E \left(\int_0^1 g(u) X_{\xi^1}(t, du) \int_0^1 g(u) X_{\xi^2}(t, du) \right) \\
(10.37) \quad &= E \left(\int g^2(u) X_{\eta_t^1}(0, du) 1(\pi_t = \{1, 2\}) \right. \\
&\quad \left. + \int g(u) X_{\eta_t^1}(0, du) \cdot \int g(u) X_{\eta_t^2}(0, du) 1(\pi_t = \{\{1\}, \{2\}\}) \right) \\
&= E \left(\int g^2(u) X_{\eta_t^1}(0, du) \right) + o(t)
\end{aligned}$$

The result then follows from Lemma 10.27(b), that is

$$(10.38) \quad (X_{\xi^1}(\infty), X_{\xi^2}(\infty)) = (\delta_y, \delta_y) \text{ for some random } y, \text{ a.s.}$$

where

$$(10.39) \quad P(y \in (a, b)) = \theta((a, b)).$$

Step 3. We can obtain the analogue of (10.38) for any finite ξ^1, \dots, ξ^k . Therefore we obtain

$$(10.40) \quad \mathcal{L}((x_\xi(t))_{\xi \in S}) \Rightarrow \int \delta_{(\delta_u)^s} \theta(du)$$

and the proof of (a) is complete.

(b) Transient case. To prove convergence of $\mathcal{L}(t)$ as $t \rightarrow \infty$ we first recall that

$$(10.41) \quad \pi_t \rightarrow \pi_\infty, (cf.(10.23)).$$

Let $n \in \mathbb{N}$ and $f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$. Then by the dual representation

$$\begin{aligned}
E_{X(0)}(F(X(t), (\eta, \pi))) &= E_{(\eta, \pi)}(F(X(0), (\eta_t, \pi_t))) \\
&= \sum_{m=1}^n E_{(\eta, \pi)} \left(\langle X_{\eta_t^1}(0), \prod_{i \in \pi_t(1)} f_i \rangle, \dots, \langle X_{\eta_t^m}(0), \prod_{i \in \pi_t(m)} f_i \rangle 1(|\pi_t| = m) \right) \\
&\rightarrow E_{(\eta, \pi)} \left(\langle \theta, \prod_{i \in \pi_\infty(1)} f_i \rangle, \dots, \langle \theta, \prod_{i \in \pi_\infty(|\pi_\infty|)} f_i \rangle \right)
\end{aligned}$$

where we have used the fact that $|Z_i(t) - Z_j(t)| \rightarrow \infty$ in probability as $t \rightarrow \infty$ and the weak mixing property of the initial random field so that for $i \neq j$

$$\begin{aligned}
& \lim_{t \rightarrow \infty} E \left(\int f_1(x) X_{Z_i(t)}(0, dx) \int f_2(y) X_{Z_j(t)}(0, dy) \right) \\
&= \lim_{t \rightarrow \infty} E \left(\int f_1(x) X_{Z_i(t)}(0, dx) \right) E \left(\int f_2(y) X_{Z_j(t)}(0, dy) \right) \\
&= \int f_1(x) \theta(dx) \int f_2(y) \theta(dy).
\end{aligned}$$

This implies the convergence of the laws \mathcal{L}_t .

The proof of the weak mixing property is obtained by noting that if $|\eta_1 - \eta_2| \rightarrow \infty$, then

$$(10.42) \quad P(\pi_\infty = (\{1\}, \{2\})) \rightarrow 1 \quad \text{as } t \rightarrow \infty.$$

The proof that the limiting law \mathcal{L}_∞ is an invariant measure for the dynamics is standard. ■

Remark 10.28 (*Population structure in 2 dimensions*)

The phenomenon of diffusive clustering in dimension $d = 2$ was discovered by Cox and Griffeath (1986) [79].

More recently, coalescing random walks used to study the coalescence time and identity by descent between 2 randomly chosen individuals on a 2-d torus (Cox and Durrett (2002) [97], Cox, Durrett, Zähle (2005) [102])

Homozygosity in large time scales

Given a probability measure μ on $[0, 1]$ the *homozygosity* is defined by

$$(10.43) \quad \int_0^1 \int_0^1 1_{x=y} \mu(dx) \mu(dy) = \sum_{i=1}^{\infty} a_i^2$$

where $\{a_i\}$ are the masses of the atoms (if any) in μ , that is $\mu = \sum a_i \delta_{y_i} + \mu_{diff}$ and μ_{diff} is the non-atomic component of the measure.

It follows from Theorem 10.25 that in the recurrent case for any $L \in \mathbb{N}$

$$(10.44) \quad \lim_{t \rightarrow \infty} E \left[\frac{1}{N(L)} \sum_{|j| \leq L} \langle X_\xi(t) \otimes X_0(t), I_\Delta \rangle \right] = 1.$$

where $I_\Delta = \{(x, y) : x = y\}$ and $N(L)$ denotes the number of sites in a ball of radius L and for the transient case

$$(10.45) \quad \lim_{t \rightarrow \infty} E [\langle X_0(t) \otimes X_0(t), I_\Delta \rangle] < 1$$

Theorem 10.29 Consider the stepping stone model on \mathbb{Z}^d and random walk kernel given by a nearest neighbour random walk. Let $d \geq 3$ and $X_0 = \nu$, with ν nonatomic. Then
(a)

$$(10.46) \quad \lim_{L \rightarrow \infty} \frac{1}{L^d} \sum_{|j| \leq L} \langle X_\xi(\infty) \otimes X_0(\infty), I_\Delta \rangle = 0,$$

(b) Each allelic type present at equilibrium has infinite total mass in \mathbb{Z}^d but has zero spatial density.

In addition, if $X(0)$ is given the stationary measure, then

(c)

$$(10.47) \quad \int_0^\infty \langle X_0(t) \otimes X_0(0), I_\Delta \rangle dt < \infty$$

if and only if $d \geq 5$.

Proof. (a) We briefly sketch the argument. We note that if $\pi = (\{1\}, \{2\})$, $\eta_1 = 0, \eta_2 = \xi$ then

$$(10.48) \quad \lim_{t \rightarrow \infty} E[< X_\xi(t) \otimes X_0(t), I_\Delta >] \leq P_{(\eta, \pi)}(\pi_t = \{1, 2\})$$

since if coalescence does not occur, the expected homozygosity is 0. But the probability that two random walks Z_0 and Z_ξ starting at 0 and ξ coalesce by time t satisfies

$$\begin{aligned} (10.49) \quad & \lim_{t \rightarrow \infty} P(\text{coalesce by time } t) \\ &= \lim_{t \rightarrow \infty} E(1 - e^{-\gamma \int_0^t P(Z_0(s) = Z_\xi(s)) ds}) \\ &\leq (1 - e^{-\gamma \int_0^t P(Z_0(s) = Z_\xi(s)) ds}) \sim \frac{1}{|\xi|^{d-2}}. \end{aligned}$$

The result follows by summing and dividing by L^d .

(c) is the analogue of (10.12). ■

Family decomposition and renormalization of the fluctuation field

The decomposition of the infinitely many types stepping stone model and the related voter model provides a tool for the study of the renormalized fluctuation field. (Recall that the difference between the stepping stone model and the voter model is that coalescence of the random walks occurs with delay for the stepping stone model but is instantaneous for the voter model. Otherwise the structure of the infinite clusters is similar.) The following special case of a theorem of I. Zähle illustrates this.

Theorem 10.30 [587] Consider the equilibrium voter model $\{X_\xi(t)\}_{\xi \in \mathbb{Z}^d} \in \{0, 1\}^{\mathbb{Z}^d}$ with nearest neighbour simple random walk kernel. For a bounded function ϕ with bounded support let

$$(10.50) \quad Z_r(\phi) := \frac{\sum_{\xi \in \mathbb{Z}^d} [X_\xi(\infty) - E(X_\xi(\infty))] \phi(\frac{\xi}{r})}{r^{\frac{d+2}{2}}}$$

If $d \geq 3$, then as $r \rightarrow \infty$, Z_r converges weakly to the Gaussian free field on \mathbb{R}^d , that is, the Gaussian field on \mathbb{R}^d with covariance kernel $\frac{1}{r^{\frac{d-2}{2}}}$.

Remark 10.31 Recall that the dual of the voter model and the dual for the 2 type Wright-Fisher diffusion differ only in that for the voter model the coalescence is instantaneous and for the Wright-Fisher model coalescence occurs with delay. Using this observation the basic strategy of the proof of this theorem which involves the “infinite colour” decomposition can be applied to the case of the Wright-Fisher diffusion.

10.4.5 Historical and genealogical structure of the neutral stepping stone model

In order to describe the historical and genealogical structure of the neutral stepping stone model we need the analogues of the historical and ancestral processes described in earlier chapters. In this direction, Greven, Limic and Winter (2005) [262] have introduced the historical interacting Moran model and the historical interacting Fisher-Wright diffusions on a countable abelian group S . We now briefly describe their formulation.

Consider the type space $E = \{0, 1\}$. Then the Moran model is given by a locally finite population with the following dynamics

- each individual moves in S independently according to a random walk with transition function

$$(10.51) \quad p_t(x, y) = \sum_{n \geq 0} p^{(n)}(x, y) \frac{t^n e^{-t}}{n!}, \quad x, y \in S$$

- resampling: each pair of individuals dies at rate γ and is replaced by a new pair of individual where each new individual adopts a type by choosing the parent independently from the “dying pair”.

This system of particles defines a measure $\eta \in M(S \times E)$. As state space we take the Liggett-Spitzer space

$$(10.52) \quad \mathcal{E}_S := \{\eta \in M(E \times E) : \sum_S \eta(\{x\} \times K) \alpha(\{x\}) < \infty\}$$

where α is a finite measure on S such that

$$(10.53) \quad \sum_y p(x, y) \alpha(\{y\}) \leq \Gamma \alpha(\{x\}).$$

Consider the two type Fisher-Wright system given by

$$(10.54) \quad d\zeta_t(x) = \sum_y (p(x, y) - \delta(x, y)) \zeta_t(x) + \sqrt{g(\zeta_t(x))} dw_t^x, \quad x \in S$$

where $g(z) = \gamma z(1-z)$.

This can be obtained as a limit as follows: let $\theta \in [0, 1]$, $(\eta^\rho)_{\rho \in \mathbb{R}_+}$ have law concentrated on \mathcal{E}_S such that $\mathcal{L}[\eta(\cdot \times E)] \in \mathcal{E}_S$ and be translation invariant and ergodic with total mass per site intensity $\rho > 0$ and $\mathcal{L}[\eta(\cdot \times \{1\})] \in \mathcal{E}_S$ and translation invariant and ergodic with total mass per site intensity $\theta\rho$.

$$(10.55) \quad \hat{\eta}_t^\rho(x) := \frac{\eta_t^\rho(\{x\}) \times \{1\}}{\eta_t^\rho(\{x\} \times E)} \mathbf{1}_{\{\eta_t^\rho(\{x\} \times E) \neq 0\}}$$

that is the relative frequency of type 1 individuals.

Theorem 10.32

$$(10.56) \quad \mathcal{L}^\eta[\hat{\eta}^\rho] \Rightarrow \mathcal{L}^\theta[\zeta]$$

where \mathcal{L}^θ denotes that $\zeta_0(x) = \theta$ for all $x \in S$.

At each time we can also consider for each individual alive at time t its *line of descent* given by a path $y \in D_{S \times E}([0, \infty))$. This path follows the random walk in reversed time from the time t until the birth time of the individual. At that time the parent particle from whom the type has been inherited provides the continuation of the path back to its birth place. This is continued until we reach time 0. The path at times < 0 and $> t$ are set to be constant equal to their values at times 0 and t .

We then get a measure

$$(10.57) \quad \eta_t^* \in \mathcal{N}(D_{S \times E}([0, \infty))).$$

Letting t vary we get the historical interacting Moran model.

Consider

$$(10.58) \quad \hat{\eta}_t^\rho(x) := \frac{\eta_t^{*,\rho}(A_{\{x\},t})}{\eta_t^{*,\rho}(E_{\{x\},t})} 1_{\{\eta_t^{\rho,*}(E_{\{x\}}) \neq 0\}}$$

where

$$(10.59) \quad E_{A,t} := \{y \in D_{S \times E}([0, \infty)) : y_t \in A \times E\},$$

We obtain the historical interacting Fisher-Wright process, $\zeta^* = (\zeta_t^*)_{t \geq 0}$, as the limit of $\eta_t^{*,\rho}$ as $\rho \rightarrow \infty$. It is a $M(D_{S \times E}(\mathbb{R}))$ -valued process.

Generator: First consider the generator \tilde{A} of the path process. Its generator is defined on a class of functions

$$(10.60) \quad \Phi(s, y), \quad s \in \mathbb{R}, y \in D_{S \times E}(\mathbb{R})$$

as follows. For $j = 1, \dots, n$ let

$$(10.61) \quad g_j : \mathbb{R} \times S \times E \rightarrow \mathbb{R}$$

where g_j are bounded and C^1 in the time variable. For $0 < t_2 < t_3 < \dots < t_n$, $n \in \mathbb{N}$ define

$$(10.62) \quad \Phi(t, y) = \prod_{j=1}^n g_j(t, y_{t \wedge t_j}).$$

Let \mathcal{A} denote the algebra of functions generated by these functions. Denote by A the generator of the random walk on $S \times E$ $Af(x, k) = \sum_{z \in S} (a(x, z) - \delta(x, z))f(z, k)$. Then we define

$$(10.63) \quad \begin{aligned} \tilde{A}\Phi(t, y) = & \prod_{j=1}^k \left[\left(\frac{\partial}{\partial t} + A \right) \prod_{j=k+1}^n g_j(t, y_{t \wedge t_j}) \right] \\ & + \left[\frac{\partial}{\partial t} \prod_{j=1}^h g_j(t, y_{t \wedge t_j}) \right] \left[\prod_{j=k+1}^n g_j(t, y_{t \wedge t_j}) \right] \end{aligned}$$

We also use the notation $y^r = y_{\cdot \wedge r}$ for $y \in D_{S \times E}([0, \infty))$ and $\eta^{*,r}$ for a measure concentrated on paths stopped at time r . Also π_S^*, π_E^* denote the obvious projections on $D_S(\mathbb{R}), D_E(\mathbb{R})$.

The martingale problem for the historical interacting Fisher-Wright system is analogous to the historical branching martingale problem (Theorem 9.24). It is formulated as follows.

For $\Phi \in \mathcal{A}$ and (t, s) , $t \geq s$

$$(10.64) \quad \left\{ \langle \zeta_y^*, \Phi(t, \cdot) \rangle - \langle \zeta_y^*, \Phi(s, \cdot) \rangle - \int_s^t \langle \zeta_r^{*,r}, (\tilde{A}\Phi)(r, \cdot) \rangle dr \right\}_{t \geq s}$$

is a martingale with increasing process

$$(10.65) \quad \left(\int_s^t \int_{(D_{S \times E}(\mathbb{R}))^2} I^r(y, y') \Phi(r, y) \Phi(r, y') \zeta^{*,r}(dy) (\zeta^{*,r}(dy') - \delta_y(dy')) dr \right)_{t \geq s}.$$

where

$$(10.66) \quad I^t(y, y') = \begin{cases} 1 & \text{if } (\pi_S^* y)_t = (\pi_S^* y')_t \\ 0, & \text{otherwise} \end{cases}$$

Theorem 10.33 (*Greven-Winter-Limic [262], Theorem 2.*) *The martingale problem (10.64), (10.66) is well-posed and arises as the diffusion limit of η^* .*

Remark 10.34 *Greven-Limic-Winter develop a particle representation starting with the corresponding look-down process in which particles are assigned labels in a countable subset of $[0, \infty)$. They construct this on a probability space with an additional randomization of labels at lookdowns and in this way construct the interacting Moran models and also the interacting Fisher-Wright processes on a common probability space following the program of Donnelly and Kurtz ([163]).*

Chapter 11

Mutation-Selection Systems

The basic mechanisms of population biology are mutation, selection, recombination and genetic drift. In the previous chapter we concentrated on mutation and genetic drift. In this chapter we introduce mathematical models of recombination and selection. However it should be emphasized that these are idealizations of highly complex biological processes and there is an immense biological literature including empirical investigation, theoretical models of varying degrees of complexity and simulation studies. For example the concept of *fitness* is an abstract notion that in the biological context can involve fitness at the level of a single gene, genome or phenotype. At the level of the genome this can involve the interaction between genes (*epistasis*) and various models of such interactions have been proposed (see e.g. Gavrilets [254]). One of the continuing issues is the question of the *levels of selection* (see e.g. Brandon and Burian (1984) [47], Lloyd (2005) [411], Okasha (2006), [467]) which include notions of group selection, kin selection, inclusive fitness (see Hamilton (1964) [277]) and so on. For example, *inclusive fitness* represents the effective overall contribution of an individual including its own reproductive success as well as its contribution (due to its behavior) to the fitness of its genetic kin.

Our aim in this chapter is to introduce some mathematical aspects of the interplay of mutation, selection and genetic drift.

11.1 The infinite population dynamics of mutation, selection and recombination.

11.1.1 Selection

The investigation of infinite population models with mutation, recombination and selection leads to an interesting class of dynamical systems (see Hofbauer and Sigmund (1988) [292] and Bürger [60], [61]). These are obtained as special cases of the general FV process by setting $\gamma = 0$ and serve as approximations to systems in which the number of individuals N is very large.

One of the objectives of this chapter is to investigate in one setting the extent to which the behavior of the finite system differs from that of the infinite system.

Consider an infinite diploid population without mutation or recombination (i.e. $\gamma = 0$, $A = 0$, $\rho = 0$) with K types of gametes. The unordered pair $\{i, j\}$ represents the genotype determined by the gametes i and j . Let $x_i(t)$ be the amount of copies of gamete i in the population at time t and p_i denote the frequency $p_i = \frac{x_i}{\sum x_i}$.

Let $V(i, j) = V(j, i) = b_{i,j} - d_{i,j}$ where $b_{i,j}$ and $d_{i,j}$ are the birth and death rates of the genotype. The fitness, $V(i)$ of the i th gamete is defined by

$$V(i) = \sum_j p_j V(i, j)$$

and the mean fitness is defined by

$$\bar{V}(p) = \bar{V} = \sum_i V(i)p_i = \sum_{ij} p_i p_j V(i, j).$$

Then the population sizes x_i satisfy the equations

$$\dot{x}_i = x_i \sum_j V(i, j) \frac{x_j}{|x|}, \quad i = 1, \dots, K$$

Proposition 11.1 *The proportions $\{p_i\}$ satisfy the equations:*

$$\dot{p}_i = p_i(V(i) - \bar{V}), \quad i = 1, \dots, K$$

Proof. This can be derived from the \dot{x} equations by the substitution $x_i = |x|p_i$ giving

$$\dot{p}_i|x| + p_i(\sum_j \dot{x}_j) = |x|p_i V(i)$$

which yields

$$\dot{p}_i + p_i(\sum_j p_j V(j)) = p_i V(i)$$

and the result immediately follows. ■

11.1.2 Riemannian structure on Δ_{K-1}

The deterministic differential equations of selection have played an important role in the development of population genetics. A useful tool in their analysis was a geometrical approach developed by Shahshahani and Akin. We next give a brief introduction to this idea.

Let M be a smooth manifold. The tangent space at x , $T_x M$ can be identified with the space of tangents at x to all smooth curves through x . The tangent bundle $TM = \{(p, v) : p \in M, v \in T_p M\}$.

Definition 11.2 *A Riemannian metric on M is a smooth tensor field*

$$g : C^\infty(TM) \otimes C^\infty(TM) \rightarrow C_0^\infty(M)$$

such that for each $p \in M$,

$$g(p)|_{T_p M \otimes T_p M} : T_p M \otimes T_p M \rightarrow \mathbb{R}$$

with

$$g(p) : (X, Y) \rightarrow \langle X, Y \rangle_{g(p)}$$

where $\langle X, Y \rangle_{g(p)}$ is an inner product on $T_p M$.

Definition 11.3 The directional derivative in direction v is defined by

$$\begin{aligned}\partial_v f(x) &= \lim_{t \rightarrow 0} \frac{f(x + tv) - f(x)}{t} \\ &= \sum v_i \frac{\partial f(x)}{\partial x_i}\end{aligned}$$

The gradient $\nabla_g f(x)$ is defined by

$$\langle \nabla_g f(x), v \rangle_g = \partial_v f(x) \quad \forall v \in T_x M.$$

Example 11.4 Consider the d -dimensional manifold $M = \mathbb{R}^d$ and $\mathbf{a}(\cdot)$ be a smooth map from M to $\mathbb{R}^d \otimes \mathbb{R}^d$ (($d \times d$)-matrices). We will write

$$\begin{aligned}\mathbf{a}(x) &= (a_{ij}(x)) \\ \mathbf{a}^{-1}(x) &= (a^{ij}(x))\end{aligned}$$

Assume that

$$\sum a^{ij}(x) u_i u_j \geq \gamma \sum u_j^2, \quad \gamma > 0.$$

The tangent space $T_x M \cong \mathbb{R}^d$ and we define a Riemannian metric on M by

$$g_{\mathbf{a}(x)}(\mathbf{u}, \mathbf{v}) := \sum_{i,j=1}^d a_{ij}(x) u^i v^j.$$

The associated Riemannian gradient and norm are

$$\begin{aligned}(\nabla_{\mathbf{a}} f)^i &= \sum_j a^{ij} \frac{\partial f}{\partial x_j} \\ \|u\|_{\mathbf{a}(x)}^2 &= \sum_{ij} a_{ij}(x) u^i u^j.\end{aligned}$$

The Shahshahani metric and gradient on Δ_{K-1}

Let $M_K = \mathbb{R}_+^K := \{x \in \mathbb{R}^K, x = (x_1, \dots, x_K), x_i > 0 \text{ for all } i\}$ is a smooth K -dimensional manifold.

Shahshahani introduced the following Riemannian metric on M_K

$$\begin{aligned}\langle u, v \rangle_g &= g_x(u, v) := \sum_{i=1}^K |x| \frac{u_i v_i}{x_i} \\ |x| &= \sum x_i\end{aligned}$$

$\|\cdot\|_g$ and $\nabla_g F$ will denote the corresponding norm and gradient. We have

$$(\nabla_g F)^i = \sum_i \frac{x^i}{|x|} \frac{\partial F}{\partial x^i} \frac{\partial}{\partial x^i}$$

Recall that the simplex $\Delta_{K-1} := \{(p_1, \dots, p_K) : p_i \geq 0, \sum_{i=1}^K p_i = 1\}$. The interior of the simplex $\Delta_{K-1}^0 = \mathbb{R}_+^K \cap \Delta_{K-1}$ is a $(K-1)$ -dimensional submanifold of M_K . We denote by $T_p \Delta_{K-1}^0$ the tangent space to Δ_{K-1}^0 at p . Then g induces a Riemannian metric on $T_p \Delta_{K-1}^0$.

Basic Facts

We have the Shahshahani inner product on $\Delta K - 1$ at a point $p \in \Delta K - 1$:

$$(11.1) \quad \langle u, v \rangle_p = \sum_{i=1}^K \frac{u_i v_i}{p_i}.$$

1. $T_p \Delta_{K-1}^0$ can be viewed as the subspace of $T_p M_K$ of vectors, v , satisfying $\langle p, v \rangle_g = 0$ if we identify p with an element of $T_p M_K$.

Proof. Recall that $T_p \Delta_{K-1}^0$ is given by tangents to all smooth curves lying in Δ_{K-1}^0 . Therefore if $v \in T_p \Delta_{K-1}^0$, then $v = q - p$ where $p, q \in \Delta_{K-1}^0$ and therefore $\sum_{i=1}^K v_i = 0$. Therefore,

$$\sum_i p_i \frac{1}{p_i} v_i = 0.$$

2. If $F : \Delta_{K-1}^0 \rightarrow \mathbb{R}$ is smooth, then the Shahshahani gradient is

$$(\nabla_g F)_i = p_i \left(\frac{\partial F}{\partial p_i} - \sum_j p_j \frac{\partial F}{\partial p_j} \right).$$

Proof. From the definition, $\nabla_g F$ is the orthogonal projection on the subspace $T_p \Delta_{K-1}^0$ of

$$(\nabla_g F)_i = p_i \frac{\partial F}{\partial p_i}$$

and therefore we must have $\sum_i (\nabla_g F)_i = 0$. This then gives the result. \blacksquare

Remark 11.5 This (Shahshahani) gradient coincides with the gradient on Δ_{K-1} associated with the K -alleles Wright-Fisher model and appears in the description of the rate function for large deviations from the infinite population limit (see below).

Theorem 11.6 The dynamical system $\{\mathbf{p}(t) : t \geq 0\}$ is given by

$$\dot{\mathbf{p}}(t) = \frac{1}{2} (\nabla_g (\mathbf{p}(t)) \bar{V}) (\mathbf{p}(t)).$$

Proof. From the above, applying the Shahshahani gradient to \bar{V} , we get

$$\begin{aligned} (\nabla_g \bar{V})_i &= 2 \left(p_i V(i) - p_i \sum_j p_j V(j) \right) \\ &= 2p_i(V(i) - \bar{V}). \end{aligned}$$

Theorem 11.7 (Fisher's Fundamental Theorem)

- (a) Mean fitness increases on the trajectories of $p(t)$.
- (b) The rate of change of the mean $\bar{V}(t)$ along orbits is proportional to the variance.
- (c) At an equilibrium point the eigenvalues of the Hessian must be real.

Proof. (a) follows immediately from (b).

(b)

$$\begin{aligned} d\bar{V}(t) &= \langle \nabla_g \bar{V}(\mathbf{p}(t)), \dot{\mathbf{p}}(t) \rangle_{g(\mathbf{p}(t))} \\ &= 2 \langle \dot{p}(t), \dot{p}(t) \rangle_{g(\mathbf{p}(t))} = 2 \left(\sum_i p_i(t)(V(i) - \bar{V}(t))^2 \right) \\ &= 2 \left(\sum_i p_i(t)V(i)^2 - \bar{V}(t)^2 \right) \\ &= 2Var_{\mathbf{p}(t)}(\mathbf{V}) \geq 0. \end{aligned}$$

(b) It also follows from the gradient form that the Hessian is symmetric (matrix of mixed second partials of \bar{V}). ■

Theorem 11.8 (*Kimura's Maximum Principle*) “Natural selection acts so as to maximize the rate of increase in the average fitness of the population.”

Proof. This simply follows from the property that the directional derivative $\partial_v \bar{V}$ is maximal in the direction of the gradient. ■

Example 11.9 Consider a two type ($\{1, 2\}$) population with frequencies $(p_1, p_2) = (p, 1 - p)$.

$$V(i, j) = av(i) + bv(j) + c\delta_{ij}$$

(When $c = 0$ we have the additive (or haploid) model. When $a = 0$ and $c > 0$ we have the heterozygote advantage model.)

In this case

$$\begin{aligned} \bar{V}(p_1, p_2) &= ap_1v(1) + bp_2v(2) + cp_1p_2 \\ &= V(p, 1 - p) = ap(v(1) - v(2)) + bv(2) + cp(1 - p) \end{aligned}$$

Then depending on the choice of $a, b, v(1), v(2)$, the optimum value of p can range between 0 and 1.

Remark 11.10 For the multilocus situation there is the Fisher-Price-Ewens version (e.g. Frank (1997) [243], Ewens [230]). This is also related to the secondary theorem of natural selection of Robertson (1966) [?] which relates the rate of change of a quantitative character under selection in terms of the covariance of the character and fitness.

The above equations are special cases of the class of *replicator equations* of the form

$$(11.2) \quad \frac{dp_i(t)}{dt} = p_i(t)(f_i(\mathbf{p}(t)) - \sum_i p_i f_i(\mathbf{p}(t))), \quad i = 1, \dots, K$$

where $\{f_i(\mathbf{p})\}_{i=1, \dots, K}$ is a vector field on Δ_{K-1} . In the linear case $f_i(\mathbf{p}) = \sum_j a_{ij} p_j$ these are equivalent to the the Lotka-Volterra equations

$$(11.3) \quad \frac{dx_i(t)}{dt} = x_i(t) \left(r_i + \sum_{j=1}^n K_{ij} x_j(t) \right), \quad i = 1, \dots, K-1$$

by setting $p_i(t) = \frac{x_i(t)}{\sum_i x_i(t)}$.

11.1.3 Mutation-Selection

The replicator equations that include both mutation and selection are given by

$$(11.4) \quad \frac{dp_i(t)}{dt} = p_i(t)(V(i) - \bar{V}) + m(\sum_{j \neq i} q_{ji}p_j - p_i)$$

where m is the mutation rate and for each j , q_{ji} , $i \neq j$ is the probability that type j mutates to type i and $\sum_{i \neq j} q_{ji} = 1$.

Theorem 11.11 *The mutation-selection dynamical system is a Shahshahani gradient system if and only if*

$$(11.5) \quad q_{ji} = q_i \quad \forall j,$$

(that is type-independent mutation as in the infinitely many alleles model). In the latter case the potential is

$$(11.6) \quad W(p) = \bar{V}(p) - H(q|p), \quad H(q|p) = -\sum_{i=1}^n q_i \log p_i.$$

Proof. See Hofbauer and Sigmund [292], Chapt. VI, Theorem 1. ■

We will see below that there is a far-reaching analogue of this for the stochastic (finite population) generalizations.

Remark 11.12 *In general the deterministic mutation-selection equations are not a gradient system and can exhibit complex dynamics - for example, a stable limit cycle (Hofbauer and Sigmund [292], 25.4). An interesting special case is the diploid case with three types - two favourable and mutation. Baake [23] showed that these can exhibit stable limit cycles. Hofbauer (1985) [291] also showed this for selection mutation models with cyclic mutation.*

Smale [525] pointed out that for n types, $n \geq 5$, dynamical systems on the simplex can have complex behaviour. He gave an example that “may not be approximated by a structurally stable, dynamical system, or it may have strange attractors with an infinite number of periodic solutions”. Some further basic results on competitive systems are covered by Hirsch (1982), (1985), (1988) [290] and Liang and Jiang (2003) [?].

11.1.4 Multiple loci and recombination

Multiloci models give rise to dynamical systems that have been extensively studied. They give rise to a large class of dynamical systems that can have complex behaviour. Akin [3] analyzed the simplest two loci model with selection and recombination and proved that in general this is not a gradient system and that periodic orbits can exist. We briefly sketch the simplest example.

Consider a two-loci model with two alleles at each loci. We denote the types by $1 = AB, 2 = Ab, 3 = aB, 4 = ab$ and with gamete frequencies

$$(11.7) \quad p_{AB}, p_{Ab}, p_{aB}, p_{ab}, \quad p_A = p_{AB} + p_{Ab}, \quad p_B = p_{AB} + p_{aB}, \quad p_a = p_{aB} + p_{ab}, \quad p_b = p_{Ab} + p_{ab}.$$

Then the *measure of linkage disequilibrium* is defined as

$$(11.8) \quad d := p_{AB}p_{ab} - p_{Ab}p_{bA}$$

so that $d = 0$ if $p_{AB} = p_A p_B$, etc. The diploid fitness function is denoted by $V(i, j)$. Some natural assumptions are that

$$(11.9) \quad m_{ij} = m_{ji}, \quad m_{14} = m_{23} = 0.$$

There are 10 zygotic types $AB/AB, Ab/AB, \dots, ab/ab$ and the corresponding fitness table

	AB	Ab	aB	ab
AB	w_{11}	w_{12}	w_{13}	w_{14}
Ab	w_{21}	w_{22}	w_{23}	w_{24}
aB	w_{31}	w_{32}	w_{33}	w_{34}
ab	w_{41}	w_{42}	w_{43}	w_{44}

The recombination vectorfield

$$(11.11) \quad R = rbd\xi_i, \quad i = 1, 2, 3, 4$$

where r is the recombination rate, b is the birth rate for double heterozygotes, d is the linkage disequilibrium and

$$(11.12) \quad \xi = (1, -1, -1, 1)$$

so that

$$(11.13) \quad d\xi = p - \pi(p)$$

where $\pi(p)$ has the same marginals as p but in linkage equilibrium (independent loci).

The system of differential equations for the frequencies of types 1, 2, 3, 4 with selection and recombination are

$$(11.14) \quad \frac{dp_i}{dt} = p_i(V(i) - \bar{V}) - rbd\xi_i \quad i = 1, 2, 3, 4$$

where

$$(11.15) \quad V(i) = \sum_{j=1}^4 p_j V(i, j), \quad \bar{V} = \sum_{i=1}^4 p_i V(i), \quad d = p_1 p_4 - p_2 p_3.$$

In the case $V \equiv 0$ the system approaches linkage equilibrium. However Akin [3] showed that there exist fitness functions V and parameters b, r such that the system exhibits a Hopf bifurcation leading to cyclic behaviour. More generally, multilocus systems can exhibit many types of complex behaviour (see for example, Kirzhner, Korol and Nevo (1996) [373] and Lyubich and Kirzhner (2003) [413]).

11.2 Infinitely many types Fleming-Viot

We now consider the Fleming-Viot process with selection and recombination and establish uniqueness using a dual representation of Ethier and Kurtz.

In Chapter 6 we showed that the martingale problem for the Fleming-Viot process with mutation selection and recombination is well-posed and defines a $\mathcal{P}(E)$ -valued Markov diffusion process. In this chapter we focus on mutation and selection but also give a brief introduction to some aspects of recombination. In evolutionary theory mutation plays an important role in producing novelty and maintaining diversity while selection eliminates deleterious mutations and makes possible the emergence and fixation of rare advantageous mutations. From a more abstract viewpoint this can be viewed as a search process which generates new information.

11.2.1 Dual representation with mutation, selection and recombination

As above we consider the mutation generator A and the bounded diploid fitness function For $V \in \mathcal{B}_{\text{sym}}(E \times E)$, set $\bar{V} = \sup_{x,y,z} |V(x,y) - V(y,z)|$. Without loss of generality we can assume that $\bar{V} = 1$ and define the selection coefficient $s > 0$ and selection operators

$$(11.16) \quad V_{im}f(x_1, \dots, x_{m+2}) = (V(x_i, x_{m+1}) - V(x_{m+1}, x_{m+2}))f(x_1, \dots, x_m).$$

For $f \in \mathcal{D}(A^{(n)}) \cap \mathcal{B}(E^n)$, define $F(f, \mu) = \int f d\mu^n$ and

$$(11.17) \quad GF(f, \mu) = F(A^{(n)}f, \mu) + \gamma \sum_{1 \leq i < j \leq n} (F(\Theta_{ij}f, \mu) - F(f, \mu)) + s \sum_{i=1}^n F(V_{in}f, \mu).$$

For $f \in C_{\text{sim}}(E^{\mathbb{N}})$, with $\mathbf{n}(f) = n$, and $f \in \mathcal{D}(A^n) \cap \mathcal{B}(E^n)$, let

$$(11.18) \quad Kf := \sum_{i=1}^n A_i f + \gamma \sum_{j=1}^n \sum_{k \neq j} [\Theta_{jk}f - f] + s \sum_{i=1}^n [V_{in}f - f].$$

where $\tilde{\Theta}_{jk}$, $\mathbf{n}(f)$ are defined as in section 7.5.

If $\beta(f) := s\mathbf{n}(f)$, then

$$(11.19) \quad GF(f, \mu) = F(Kf, \mu) + \bar{V}(\mathbf{n}(f))F(f, \mu),$$

and $\sup_{\mu \in M_1(E)} |F(Kf, \mu)| \leq \text{const} \cdot \mathbf{n}(f)$.

Let $\rho \geq 0$ and $\eta(x_1, x_2, \Gamma)$ be a transition function from $E \times E \rightarrow E$. For $i = 1, \dots, m$ define $R_{im} : \mathcal{B}(E^m) \rightarrow \mathcal{B}(E^{m+1})$ by

$$(11.20) \quad R_{im}f(x_1, \dots, x_{m+1}) = \int f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_m) \eta(x_i, x_{m+1}, dz)$$

and assume that $R_{im} : C_b(E^m) \rightarrow C_b(E^{m+1})$. The R_{im} are called the recombination operators for the process and ρ is called the recombination rate.

Given $V \in \mathcal{B}_{\text{sym}}(E \times E)$, with $\bar{V} := \sup_{x,y,z} |V(x,y) - V(y,z)| < \infty$, define the selection operators

$$(11.21) \quad V_{im}f(x_1, \dots, x_{m+2}) = \frac{V(x_i, x_{m+1}) - V(x_{m+1}, x_{m+2})}{\bar{V}} f(x_1, \dots, x_m) \quad \text{for } i = 1, \dots, m.$$

For $f \in \mathcal{D}(A^{(n)}) \cap \mathcal{B}(E^n)$, define $F(f, \mu) = \int f d\mu^n$ and

$$(11.22) \quad \begin{aligned} GF(f, \mu) &= F(A^{(n)}f, \mu) + \gamma \sum_{1 \leq i < j \leq n} (F(\tilde{\Theta}_{ij}f, \mu) - F(f, \mu)) \\ &\quad + \rho \sum_{i=1}^n (F(R_{in}f, \mu) - F(f, \mu)) + \bar{V} \sum_{i=1}^n F(V_{in}f, \mu). \end{aligned}$$

For $f \in C_{\text{sim}}(E^{\mathbb{N}})$, with $\mathbf{n}(f) = n$, and $f \in \mathcal{D}(A^n) \cap \mathcal{B}(E^n)$, let

$$(11.23) \quad Hf := \sum_{i=1}^n A_i f + \gamma \sum_{j=1}^n \sum_{k \neq j} [\tilde{\Theta}_{jk}f - f] + \rho \sum_{i=1}^n [R_{in}f - f] + \bar{V} \sum_{i=1}^n [V_{in}f - f].$$

If $\beta(f) := \bar{V}\mathbf{n}(f)$, then

$$(11.24) \quad GF(f, \mu) = F(Hf, \mu) + \beta(f)F(f, \mu),$$

and $\sup_{\mu \in M_1(E)} |F(Hf, \mu)| \leq \text{const} \cdot \mathbf{n}(f)$.

Theorem 11.13 *Let G satisfy the above conditions and assume that the mutation process with generator A has a version with sample paths in $D_E[0, \infty)$. Then for each $\mu \in \mathcal{P}(E)$ there exists a unique solution P_μ of the martingale problem for G .*

Proof. (Ethier-Kurtz (1987) [214]) The uniqueness will be proved by constructing a dual representation.

Let N be a jump Markov process taking non-negative integer values with transition intensities

$$(11.25) \quad q_{m,m-1} = \gamma m(m-1), \quad q_{n,m+2} = \bar{V}m, \quad q_{m,m+1} = \rho m, \quad q_{i,j} = 0 \text{ otherwise.}$$

For $1 \leq i \leq m$, let $\{\tau_k\}$ be the jump times of N , $\tau_0 = 0$, and let $\{\Gamma_k\}$ be a sequence of random operators which are conditionally independent given M and satisfy

$$(11.26) \quad P(\Gamma_k = \Theta_{ij}|N) = \frac{2}{N(\tau_k-)N(\tau_k)} \mathbf{1}_{N(\tau_k-) - N(\tau_k) = 1}, \quad 1 \leq i < j \leq N(\tau_k-)$$

$$(11.27) \quad P(\Gamma_k = R_{im}|N) = \frac{1}{m} \mathbf{1}_{\{N(\tau_k-) = m, N(\tau_k) = m+1\}}$$

$$(11.28) \quad P(\Gamma_k = V_{im}|N) = \frac{1}{m} \mathbf{1}_{\{N(\tau_k-) = m, N(\tau_k) = m+2\}}.$$

For $f \in C_{\text{sim}}(E^{\mathbb{N}})$, define the $C_{\text{sim}}(E^{\mathbb{N}})$ -valued process Y with $Y(0) = f$ by

$$(11.29) \quad Y(t) = S_{t-\tau_k} \Gamma_k S_{\tau_k-\tau_{k-1}} \Gamma_{k-1} \dots \Gamma_1 S_{\tau_1} f, \quad \tau_k \leq \tau_{k+1}.$$

Then for any solution P_μ to the martingale problem for G and $f \in C_{\text{sim}}(E^{\mathbb{N}})$ we get the FK-dual representation

$$(11.30) \quad P_\mu[F(f, X(t))] = Q_f \left[F(Y(t), \mu) \exp \left(\bar{V} \int_0^t \mathbf{n}(Y(u)) du \right) \right]$$

which establishes that the martingale problem for G is well-posed. Since the function $\beta(f) = \bar{V}\mathbf{n}(f)$ is not bounded we must verify condition (7.30). This follows from the following lemma due to Ethier and Kurtz (1998) [218], Lemma 2.1.

■

Lemma 11.14 *Let $N(t) = \mathbf{n}(Y(t))$ be as above, $\tau_K := \inf\{t : N(t) \geq K\}$ and $\theta > 0$. Then there exists a function $R(n) \geq \text{const} \cdot n^2$ and a constant $L > 0$ such that*

$$(11.31) \quad E \left[R(N(t \wedge \tau_K)) \exp \left(\theta \int_0^{t \wedge \tau_K} N(s) ds \right) | N(0) = n \right] \leq F(n)e^{Lt}, \quad \forall K \geq 1,$$

and given $N(0) = n$, $\left\{ N(t \wedge \tau_K) \exp \left(\bar{V} \int_0^{t \wedge \tau_K} N(s) ds \right) : K \geq 1 \right\}$ are uniformly integrable.

Proof. The integer-valued process $N(t)$ is a birth and death process with jump rates

$$(11.32) \quad m \rightarrow (m+1) \text{ rate } \rho m, \quad m \rightarrow m+2 \text{ rate } \bar{V}m, \quad m \rightarrow m-1 \text{ rate } \gamma m(m-1).$$

Let Q denote the corresponding generator. Take $R(m) := (m!)^\beta$, with $\beta < \frac{1}{2}$. Then

$$\begin{aligned} & QR(m) + \theta m R(m) \\ &= \gamma m(m-1)(R(m-1) - R(m)) + \rho m(R(m+1) - R(m)) \\ &\quad + \theta m(R(m+2) - R(m)) + \theta m R(m) \\ &= -\gamma O(m^2)(m!)^\beta + \rho O(m^2)(m!)^\beta + \theta O(m^{1+2\beta})(m!)^\beta \end{aligned}$$

Since the negative term dominates for large m if $0 < \beta < \frac{1}{2}$ and $\gamma > 0$, we can choose $L > 0$ such that

$$(11.33) \quad QR(m) + \theta m R(m) \leq L.$$

The optional sampling theorem implies that for $\tau_K := \inf\{t : N(t) \geq K\}$ and $N(0) = m$

$$\begin{aligned} & E \left[\exp \left(\theta \int_0^{t \wedge \tau_k} N(s) ds \right) | N(0) = m \right] \\ & \leq E \left[R(N(t \wedge \tau_k)) \exp \left(\theta \int_0^{t \wedge \tau_k} N(s) ds \right) | N(0) = m \right] \\ & \leq R(m) + E \left[\int_0^{t \wedge \tau_k} \exp \left(\theta \int_0^u N(s) ds \right) (QR(N(u)) + \theta N(u) R(N(u))) du | N(0) = m \right] \\ & \leq R(m) + LE \left[\int_0^{t \wedge \tau_k} \exp \left(\theta \int_0^u N(s) ds \right) du | N(0) = m \right] \end{aligned}$$

and the lemma follows by Gronwall's inequality. ■

11.2.2 Girsanov formula for Fleming-Viot with Mutation and Selection

Recall that the Fleming-Viot martingale problem $\mathbb{MP}_{(A, \gamma Q, 0)}$ corresponds to the case

$$\langle M(A), M(A) \rangle_t = \gamma \int_0^t Q(X_s; A, A) ds$$

where

$$Q(\mu; dx, dy) = \mu(dx)\delta_x(dy) - \mu(dx)\mu(dy).$$

and that M is a worthy martingale measure.

Now consider a time-dependent diploid fitness function $V : [0, \infty) \times E \times E \rightarrow \mathbb{R}$ with

$\|V\|_\infty < \infty$. . Then the FV martingale problem $\mathbb{MP}_{(A,Q,V)}$ is

$$\begin{aligned} M^V(\phi)_t &:= \langle X_t, \phi \rangle - \int_0^t \langle X_s, A\phi \rangle ds \\ &\quad - \int_0^t \int \left[\int V(s, x, y) X_s(dy) - \int \int V(s, y, z) X_s(dy) X_s(dz) \right] \phi(x) X_s(dx) ds \\ &= \langle X_t, \phi \rangle - \int_0^t \langle X_s, A\phi \rangle ds \\ &\quad - \int_0^t \int \int \left[\left(\int \frac{V(s, y, z)}{\gamma} X_s(dz) \right) \gamma Q(X_s, dx, dy) \right] \phi(x) ds \\ \langle M^V(\phi) \rangle_t &= \gamma \int_0^t \int \int \phi(x) \phi(y) Q(X_s, dx, dy) ds. \end{aligned}$$

We then apply Theorem 7.18 to conclude that this martingale problem has a unique solution \mathbb{P}^V and that the Radon-Nikodym derivative

$$Z_t^V := \frac{d\mathbb{P}^V}{d\mathbb{P}^0}|_{\mathcal{F}_t}$$

where \mathbb{P}^0 is the unique solution to $\mathbb{MP}_{(A,\gamma Q,0)}$ is given by

$$\begin{aligned} Z_t^V &:= \exp \left(\frac{1}{\gamma} \int_0^t \int V(s, X_s, y) M^0(ds, dy) \right. \\ &\quad \left. - \frac{1}{2\gamma^2} \int_0^t \int \int V(s, X_s, x) V(s, X_s, y) \gamma Q(X_s; dx, dy) ds \right). \end{aligned}$$

where we write

$$V(s, X_s, x) = \int V(s, z, x) X_s(dz).$$

11.3 Long-time behaviour of systems with finite population resampling, mutation and selection

Systems with finite population resampling can have rather different long-time behaviour than the corresponding infinite population systems. One essential difference is that even high fitness types can be lost due to resampling and in the absence of mutation the system can eventually become unitype. On the other hand if the mutation process can regenerate all types, then the system can reach equilibrium in which all types are present. We now consider these two situations.

11.3.1 Fixation in finite population systems without mutation

In the previous section we have considered the infinite population system with selection but no mutation. In this case Fisher's fundamental theorem states that such a system evolves to one

of maximal population fitness. But what happens in the finite population case, $\gamma > 0$? We first observe that if $V \equiv 0$, then $\{X_t(A) : t \geq 0\}$ is a bounded martingale and

$$(11.34) \quad X_t(A) \xrightarrow[t \rightarrow \infty]{} \begin{cases} 1 & \text{with probability } X_0(A) \\ 0 & \text{with probability } (1 - X_0(A)). \end{cases}$$

Therefore

$$X_t \xrightarrow{t \rightarrow \infty} \delta_x \text{ with } x \in A \text{ with probability } X_0(A)$$

that is, the system experiences ultimate “fixation”. If we add selection to this, ultimate fixation still occurs. However if γ is small then the tendency is for the limiting types to be those of higher fitness.

11.3.2 The Equilibrium Infinitely Many Alleles Model with Selection

In order to have a non-degenerate equilibrium a source of new types through mutation is required. In this section we consider the type independent infinitely many alleles mutation together with selection. If ν_0 is a non-atomic measure, then mutation always leads to a new type and thus provides a mechanism to guarantee sufficient diversity on which selection can act.

We first obtain the ergodic theorem in this case.

Proposition 11.15 (*Ergodicity of IMA Fleming-Viot with selection*)

Consider the infinitely many allele type Fleming-Viot process on $[0, 1]$, mutation source $\nu_0 \in \mathcal{P}([0, 1])$ and $X_0 = \mu$. Then

$$(11.35) \quad X_t \Rightarrow X_\infty$$

where X_∞ is a random probability measure on $[0, 1]$. The process is ergodic and $\mathcal{L}(X_\infty)$ is the unique stationary measure.

Proof. This follows immediately from the dual representation. Note that for the type-independent mutation

$$(11.36) \quad Af(x) = \int_0^1 f(y)\nu_0(dy) - f(x)$$

and then

$$(11.37) \quad q_{1,0} > 0.$$

Due to the quadratic death rate the process $\mathbf{n}(Y(t))$ is recurrent and starting at $n \geq 1$ returns to 1 with probability 1. But then the probability that the dual reaches the trap $C^0([0, 1])$ is one and the limiting dual is a constant. This implies the result. ■

We now identify the resulting stationary measure.

Let \mathbb{P}_∞^0 denote the probability measure on $C_{\mathcal{P}([0, 1])}(-\infty, \infty)$ corresponding to the reversible stationary measure, with one dimensional marginal distribution $\Pi_\gamma^0(d\mu)$, for the neutral infinitely many alleles model (recall the representation in terms of the Moran subordinator). Assume that V is symmetric and $V(s, x, y) = V(x, y) = V(y, x)$.

The following results is the infinitely many types analogue of a result of Wright [?].

Theorem 11.16 *The infinitely many alleles model with selection has a reversible stationary measure given by*

$$\Pi_\gamma^V(d\mu) = \frac{1}{Z} e^{\frac{V(\mu)}{\gamma}} \Pi_\gamma^0(d\mu)$$

where Z is a normalizing constant.

Proof. Let X_0 have distribution

$$\frac{1}{Z} e^{\frac{V(X_0)}{\gamma}} \Pi_\gamma^0(dX_0)$$

Recall that to verify that this is a reversible equilibrium measure it suffices to show that for any two continuous functions, f and g , on $[0, 1]$

$$\mathbb{P}_\infty(f(X_0)g(X_t)) = \mathbb{P}_\infty(g(X_0)f(X_t)).$$

But

$$\begin{aligned} & \mathbb{P}_\infty(f(X_0)g(X_t)) \\ &= \frac{1}{Z} \int f(X_0) e^{\frac{V(X_0)}{\gamma}} g(X_t) \mathbb{P}_{X_0}^V(d\{X_s : 0 \leq s \leq t\}) \Pi_\gamma^0(dX_0) \\ &= \frac{1}{Z} \int f(X_0) e^{\frac{V(X_0)}{\gamma}} Z_t^V \mathbb{P}_\infty^0(dX_0) g(X_t). \end{aligned}$$

By Theorem 7.18

$$\begin{aligned} Z_t^V := & \exp \left(\frac{1}{\gamma} \int_0^t \int V(X_s, y) M^0(ds, dy) \right. \\ & \left. - \frac{1}{2\gamma^2} \int_0^t \int \int V(X_s, x) V(X_s, y) \gamma Q(X_s; dx, dy) ds \right) \end{aligned}$$

where

$$\begin{aligned} M_s^0(dy) &= X_s - \int_0^s A^* X_u du \\ &= X_s - \int_0^s c[\nu_0 - X_u] du \end{aligned}$$

As a preparation, note that by Ito's lemma,

$$\begin{aligned} & d_t \left(\int \int V(x, y) X_t(dx) X_t(dy) \right) \\ &= \int \int V(x, y) X_t(dx) d_t X_t(dy) + \int \int V(x, y) X_t(dy) d_t X_t(dx) \\ &+ \int \int V(X_s, x) V(X_s, y) \gamma Q(X_s; dx, dy) \end{aligned}$$

Hence by symmetry in x and y and Ito's Lemma,

$$\begin{aligned} \frac{1}{\gamma} \int_0^t \int \int V(x, y) X_s(dx) d_s X_s(dy) &= \frac{1}{2\gamma} \left[\int \int V(x, y) X_t(dx) X_t(dy) \right. \\ &\quad \left. - \int \int V(x, y) X_0(dx) X_0(dy) \right] \\ &\quad - \frac{1}{2\gamma^2} \int_0^t \int \int V(X_s, x) V(X_s, y) \gamma Q(X_s; dx, dy) \end{aligned}$$

Therefore

$$\begin{aligned}
& \log(e^{\frac{1}{\gamma}V(X_0)}Z_t^V) \\
&= \int_0^t \int \frac{1}{\gamma}V(X_s, y)M^0(ds, dy) + \frac{1}{\gamma} \int \int V(x, y)X_0(dx)X_0(dy) \\
&\quad - \frac{1}{2\gamma^2} \int_0^t \int \int V(X_s, x)V(X_s, y)\gamma Q(X_s; dx, dy)ds \\
&= \frac{1}{\gamma} \int_0^t \int \int V(x, y)X_s(dx)d_sX_s(dy) - \frac{c}{\gamma} \int_0^t \int \int V(x, y)X_s(dx)(\nu_0(dy) - X_s(dy))ds \\
&\quad - \frac{1}{2\gamma^2} \int_0^t \int \int V(X_s, x)V(X_s, y)\gamma Q(X_s; dx, dy)ds \Big) + \frac{1}{\gamma} \int \int V(x, y)X_0(dx)X_0(dy) \\
&= \frac{1}{2\gamma} [\int \int V(x, y)X_t(dx)X_t(dy) + \int \int V(x, y)X_0(dx)X_0(dy)] \\
&\quad - \frac{1}{\gamma^2} \int_0^t \int \int V(X_s, x)V(X_s, y)\gamma Q(X_s; dx, dy)ds \\
&\quad - \frac{c}{\gamma} \int_0^t \int \int V(x, y)X_s(dx)(\nu_0(dy) - X_s(dy))ds
\end{aligned}$$

This is symmetric with respect to the direction of time. Also under \mathbb{P}_∞^0 , $\{X_t : t \in \mathbb{R}\}$ is stationary and reversible. Therefore we conclude that

$$E(f(X_0)g(X_t)) = E(f(X_t)g(X_0))$$

Therefore $\frac{1}{Z}e^{\frac{V(\mu)}{2\gamma}}\Pi_\gamma^0(d\mu)$ is a reversible invariant measure. ■

Corollary 11.17 Consider the K -allele case with $c = \gamma$ and $\nu_0(dx) = dx$. Assume that $V(p)$ is continuous and has a unique global maximum $p_0 \in \Delta_{K-1}$. Then as $\gamma \rightarrow 0$, $\Pi_\gamma^V \Rightarrow \delta_{p_0}$.

Proof. In this case $\Pi_\gamma^0(dp)$ is the Dirichlet (1) distribution on Δ_{K-1} . Let

$$N_p^\varepsilon := \{p : V(p_0) - V(p) \leq \varepsilon\}$$

Then for any $\varepsilon > 0$, $\Pi_\gamma^0(N_{p_0}^\varepsilon) > 0$. It is then easy to check that

$$\Pi_\gamma^V((N_{p_0}^\varepsilon)^c) \rightarrow 0 \text{ as } \gamma \rightarrow 0.$$

■

11.3.3 Remarks on further developments

The Gibbs form of the invariant measure is suggestive. One can ask if there is a reversible equilibrium for other mutation processes. However the fact that the type-independent mutation is the only mutation process for which the equilibrium is reversible was proved by Li, Shiga and Ya (1999) [405]. This is the analogue of the result of Hofbauer and Sigmund mentioned above. The reason for the reversibility of the IMA mutation is that this mutation mechanism “erases” all historical information. For the related infinitely many sites mutation historical information is preserved and the stationary state is not reversible.

A natural setting for the study of the macroscopic development of population systems that incorporates finite local capacity, finite population resampling, spatial migration, mutation and

selection is the *stepping stone model with infinitely many sites or hierarchical mutation and general state-dependent fitness*, $V(x, \mu)$ of types. These systems do not have reversible stationary measures but other methods including mean-field methods can yield partial information on the large scale behaviour of these systems. This is currently an active field of research.

11.4 Remarks on the Bibliography

The topics covered in these notes have focussed on basic tools from stochastic analysis and the basic mechanisms of reproduction, mutation and selection as well as spatial migration. These methods and processes serve as building blocks for the study of more complex systems. The Bibliography below contains numerous references to the vast and growing literature on developments along these lines both in stochastic analysis related to population models and references from the biological literature which deal with questions to which stochastic population models have contributed or have the potential to contribute to.

Chapter 12

Appendix I: Martingales and large deviations

12.1 Martingales and uniform integrability

Definition 12.1 A family of integrable random variables $\{X_\alpha\}$ is uniformly integrable if

$$(12.1) \quad \sup_{\alpha} E[|X_\alpha 1_{|X_\alpha|>L}] \rightarrow 0 \text{ as } L \rightarrow \infty.$$

A sufficient condition for uniform integrability is that there exists a non-negative increasing function such that $\lim_{s \rightarrow \infty} \frac{g(s)}{s} = \infty$ and

$$(12.2) \quad \sup_{\alpha} E(g(|X_\alpha|)) < \infty.$$

Lemma 12.2 (a) If $X_n \Rightarrow X$, then $E(|X|) \leq \liminf_{n \rightarrow \infty} E(|X_n|)$. If $p_2 > p_1 > 0$ and $\sup_n E(|X_n|^{p_2}) < \infty$, then $E(|X|^{p_1}) = \lim_{n \rightarrow \infty} E(|X_n|^{p_1})$.

(b) If a sequence of martingales $X_n(\cdot)$ converge weakly to $X(\cdot)$ in $D_{\mathbb{R}}([0, \infty))$ and for some $\delta > 0$

$$(12.3) \quad \sup_n \sup_{0 \leq t \leq T} E(|X_n(t)|^{1+\delta}) < \infty$$

for each $T < \infty$, then $X(\cdot)$ is a martingale.

Theorem 12.3 Let $\{X_n\}$ be a martingale and τ a stopping time with $E(\tau) < \infty$. Assume that

$$(12.4) \quad E(|X_{n+1} - X_n| | X_1, \dots, X_n) \leq \alpha < \infty, \quad n \leq \tau.$$

Then

$$(12.5) \quad EX_\tau = E(X_1).$$

Proof. See Breiman [58] Prop. 5.33. ■

Application to the Radon-Nikodym Theorem

Let P and Q be two probability measures on (Ω, \mathcal{F}) and \mathcal{F}_n an increasing sequence of sub- σ -algebras which generate \mathcal{F} . Assume that $Q \ll P$ on \mathcal{F}_n for all n and

$$(12.6) \quad R_n = \left. \frac{dQ}{dP} \right|_{\mathcal{F}_n}.$$

Theorem 12.4 $((R_n)_{n \in \mathbb{N}}, \mathcal{F}_n)$ is a martingale. R_n is uniformly integrable if and only if $Q \ll P$ on \mathcal{F} and if $Q \ll P$, then $R_\infty = \lim_{n \rightarrow \infty} R_n$ is the Radon-Nikodym derivative. If $Q \perp P$, then $R_n \rightarrow 0$ a.s. with respect to P and $R_n \rightarrow \infty$ with respect to Q .

Reference: See, for example, Durrett [185].

12.1.1 Burkholder-Davis-Gundy inequalities

For every continuous local martingale M and $p > 0$ there exists $0 < c < C < \infty$ such that

$$(12.7) \quad cE([M]_\infty^{p/2}) \leq E((M_\infty^*)^p) \leq CE([M]_\infty^{p/2})$$

where $M_t^* := \sup_{s \leq t} |M_s|$.

Theorem 12.5 (Burkholder's inequality (Burkholder (1973)))

For $p > 1$ and martingale M with jumps ΔM (Burkholder's inequality (Burkholder (1973)) Theorem 21.1) is

$$(12.8) \quad E[(M_t^*)^p] \leq CE[\langle M \rangle^p] + E(\sup_{s \leq t} |\Delta M_s|^p).$$

12.2 Large Deviations

Lemma 12.6 (a) Let X_1, \dots, X_N be bounded iid r.v.'s with mean μ and $|X_i| \leq K$ a.s. and $S_N = \sum_{i=1}^N X_i$. Then there exists constants C, η depending on K and ε so that

$$P\left(\left|\frac{S_N}{N} - \mu\right| \geq \varepsilon\right) < Ce^{-N\eta}.$$

(b) Suppose that $\{X_n\}$ are random variables such that $E[X_{n+1}|S_n] = 0$ and $E[X_{n+1}^2|S_n] \leq v$. Then (Azuma's inequality): for $x > 0$

$$P\left(\frac{|S_n|}{n} \geq x\right) \leq \exp\left(-nH\left(\frac{x+v}{1+v} \mid \frac{v}{1+v}\right)\right)$$

where

$$(12.9) \quad H(p|p_0) = p \log \frac{p}{p_0} + (1-p) \log \frac{1-p}{1-p_0}.$$

Proof. (a)(cf. Dembo and Zeitouni [154], Theorem 2.2.3) Without loss of generality we can assume that $K = 1$ and $\mu = 0$. Then Cramér's Theorem states that

$$\limsup \frac{1}{N} \log P\left(\left|\frac{S_N}{N}\right| \geq \varepsilon\right) \leq \inf_{|x| > \varepsilon} \Lambda^*(x)$$

where $\Lambda^*(x)$ is the Legendre transform

$$(12.10) \quad \Lambda^*(x) := \sup_x \{\lambda x - \Lambda(\lambda)\}$$

$$(12.11) \quad \Lambda(\lambda) := \log E[e^{\lambda X_1}].$$

Then a simple calculation shows that

$$\begin{aligned} \Lambda^*(x) &\geq H\left(\frac{x+1}{2} \mid \frac{1}{2}\right) \\ H\left(p \mid \frac{1}{2}\right) &= -(p \log 2p + (1-p) \log 2(1-p)) \\ &> 0. \end{aligned}$$

(b) See Dembo and Zeitouni [154] exercise 2.2.29 for hints on the proof. ■

Chapter 13

Appendix II: Measures and Topologies

13.1 Measures on Polish spaces and weak convergence

Throughout these lectures measures on a number of Polish topological spaces will play an essential role. In addition, the construction of various limiting objects involves weak convergence of probability measures on these spaces. In this section we collect some basic facts on certain basic classes of Polish spaces and the topology of weak convergence of probability measures on Polish spaces.

13.1.1 Borel measures on Polish space

Definition 13.1 A Polish space is a separable completely metrizable topological space; that is, a space homeomorphic to a complete metric space (E, d) that has a countable dense subset.

A basic property of a Polish space is that a subset A is relatively compact iff it is totally bounded with respect to d .

Theorem 13.2 Every finite measure μ on the Borel σ -algebra $\mathcal{E} = \mathcal{B}(E)$ of a complete separable metric space (E, d) is Radon, that is, regular relative to the family of compact subsets.

Proof. We first prove that $\mu(E) = \sup_{K \subset E, K \text{ compact}} \mu(K)$. Let $\{x_k\}$ be a dense subset and $B(x, \varepsilon)$ the ball with center x and radius ε . For $\varepsilon > 0$, choose integers N_1, N_2, \dots such that

$$P(\bigcup_{k=1}^{N_n} B(x_k, \frac{1}{n})) \geq 1 - \frac{\varepsilon}{2^n}$$

Let K be the closure of $\bigcap_{n \geq 1} \bigcup_{k=1}^{N_n} B(x_k, \frac{1}{n})$. Then K is totally bounded and hence compact and

$$P(K^c) < \varepsilon.$$

For the completion, see D.L. Cohn, Measure Theory, Birkhäuser. ■

Sometimes we will consider a special class of metric spaces, the *ultrametric spaces* on which the strong triangle inequality is satisfied

$$(13.1) \quad d(x, y) \leq \max(d(x, z), d(y, z)).$$

Example 13.3 Consider a rooted real tree \mathcal{T} with root \emptyset . Then for any t the level set

$$(13.2) \quad X^t := \{x \in \mathcal{T} : d(\emptyset, x) = t\}$$

is an ultrametric space.

13.1.2 Weak convergence of measures

A basic reference for weak convergence of probability measures on a complete separable metric space (E, d) is Ethier-Kurtz [212], Chap. 3. Let $\mathcal{P}(E)$ denote the set of Borel probability measures on E . The following are basic properties

1. $(\mathcal{P}(E), \rho)$ is a complete separable metric space where ρ is the Prohorov metric.
2. A subset $\mathcal{M} \subset \mathcal{P}(E)$ is relatively compact if and only if it is (uniformly) *tight*, that is, for each $\varepsilon > 0$ there exists a compact set $K \subset E$ such that

$$(13.3) \quad \inf_{P \in \mathcal{M}} P(K) \geq 1 - \varepsilon.$$

13.2 Projective Limits of Measures and Processes

(Reference: P.A. Meyer (1966) [440])

Let (E, \mathcal{E}) be a measurable space, T an index set and U the collection of finite subsets of T . We denote by π_u the canonical projection of E^T on E^u , $u \in U$. The mappings π_u, π_{uv} are measurable over the natural product σ -fields and

$$\begin{aligned} \pi_{uv}\pi_v &= \pi_u, \quad u \subset v \\ \pi_{uv}\pi_{vw} &= \pi_{uw}, \quad u \subset v \subset w \end{aligned}$$

Suppose we are given a probability law P_u on every measurable space (E^u, \mathcal{E}^u) , $u \in U$. The family $(P_u)_{u \in U}$ constitutes a projective systems of probability laws if for every pair of elements, $u \subset v$,

$$\begin{aligned} \pi_{uv}(P_v) &= P_u \\ P_u(A) &= P_v(\pi_{uv}^{-1}(A)) \end{aligned}$$

The projective system admits a *projective limit* if there exists a probability law P on (E^T, \mathcal{E}^T) such that

$$\pi_u(P) = P_u \text{ for every } u \in U$$

The collection of subsets of E^T of the form $\pi_u^{-1}(A_u)$ ($u \in U, A_u \in \mathcal{E}^u$), denoted by \mathcal{E}_0^T , forms a *paving* (i.e. a collection of subsets of E^T containing \emptyset, E^T) is closed under finite unions and taking complements. If we put for $A = \pi_u^{-1}(A_u) \in \mathcal{E}_0^T$,

$$P(A) = P_u(A_u)$$

we obtain a function of A independent of the representation of A . Given that the paving \mathcal{E}_0^T generates the σ -field \mathcal{E}^T the uniqueness of the projective limit follows.

Definition 13.4 A measure P on (Ω, \mathcal{F}) is said to be regular with respect to a semicompact paving \mathcal{T} (every countable family having the finite intersection property has the countable intersection property) whose elements are measurable if

$$P(B) = \sup_{\substack{A \in \mathcal{T} \\ A \subset B}} P(A) \text{ for every } B \in \mathcal{F}$$

Theorem 13.5 *It suffices to show this for B in a generating algebra $\mathcal{F}_0 \subset \mathcal{F}$ and the extension to a regular probability law is unique.*

Proof. Use the monotone class theorem. See Meyer (1966). ■

Theorem 13.6 (Neveu) *Suppose that there exists, for each $t \in T$, a semicompact paving $\mathcal{K}_t \subset \mathcal{E}$ such that the law P_t is regular wrt \mathcal{K}_t . Then the projective system $(P_u)_{u \in U}$ admits a projective limit.*

Proof. We can suppose that E belongs to each of the pavings \mathcal{K}_t . Let $\mathcal{K}_u, \mathcal{K}_T$ denote the closure under $(\cup f, \cap c)$ of the product paving $\prod_{t \in u} \mathcal{K}_t, \prod_{t \in T} \mathcal{K}_t$. Each of these pavings is semicompact. Next denote by \mathcal{K}_T^0 the paving on E^T consisting of subsets of the form $\pi_u^{-1}(A_u) (u \in U, A_u \in \mathcal{K}_u)$, $\mathcal{K}_T^0 \subset \mathcal{K}_T$ and hence is semicompact. Let us first show that the law P_u is regular relative to \mathcal{K}_u . To do this we apply the previous theorem using \mathcal{F}_0 the collection of finite unions of sets of the form $\prod_{t \in u} A_t (A_t \in \mathcal{E})$ and for \mathcal{K}_0 the collection of finite unions of sets of the form $\prod_{t \in u} K_t (K_t \in \mathcal{K}_t)$. We must verify $P_u(\prod_{t \in u} A_t) = \sup_{\substack{K_t \in \mathcal{K}_t \\ K_t \subset A_t}} P(\prod_{t \in u} K_t)$. For $\varepsilon > 0$ take $\varepsilon_t > 0 (t \in u)$ such that $\sum \varepsilon_t = \varepsilon$. Choose for each $t \in u$ a set $K_t \in \mathcal{K}_t$ such that $K_t \subset A_t$ and $P(A_t \setminus K_t) \leq \varepsilon_t$. Denote by B_t the inverse image in E^u of $A_t \setminus K_t$ under the projection of E^u onto $E^{\{t\}}$

$$P_u \left[\left(\prod_{t \in u} A_t \right) \setminus \left(\prod_{t \in u} K_t \right) \right] \leq P_u(\cup_{t \in u} B_t) \leq \sum_{t \in u} P_t(A_t \setminus K_t) \leq \varepsilon$$

Hence P_u is regular wrt \mathcal{K}_u for each $u \in U$. Therefore P is defined on \mathcal{E}_0^T and can be extended uniquely to a regular probability law on \mathcal{E} . ■

We state the following theorem without proof. This implies, for example, that every probability law on the Baire σ -field of a compact space is regular.

Theorem 13.7 *Let \mathcal{K} be a semicompact paving on the set Ω , closed under $(\cup f, \cap c)$ such that the complement of every element of \mathcal{K} belongs to \mathcal{K}_σ . Every probability law P on the σ -fields generated by \mathcal{K} is then regular relative to \mathcal{K} .*

Proof. See Meyer (1966). ■

Remark: Note that we can generalise the above result of Neveu to let U be a *right-filtering partially ordered set*, i.e. given $u, v \in U$, $\exists w \succ u, w \succ v$, and a family of probabilities P_u on (E_u, \mathcal{E}_u) , and measurable maps $\pi_{uv} : (E_v, \mathcal{E}_v) \rightarrow (E_u, \mathcal{E}_u)$, $u \subset v$.

$$\pi_{uv}\pi_v = \pi_u, u \subset v$$

$$\pi_{uv}\pi_{vw} = \pi_{uw}, u \subset v \subset w$$

13.3 Random measures

Let E be Polish space with a countable base \mathcal{D} for the topology, and \mathcal{A} the algebra generated by \mathcal{D} . For $A_1, \dots, A_n \in \mathcal{A}$ let $P_{\{A_1, \dots, A_n\}}$ be a distribution $(R_+)^n$. Assume that

1. consistency $P_{\{A_1, \dots, A_n\}} = P_{\{A_1, \dots, A_n, E\}}$
2. if $A, B \in \mathcal{A}$ and $A \cap B = \emptyset$, then

$$(13.4) \quad P_{A, B, A \cup B}((x, y, z) \in \mathbb{R}^3 : x + y = z) = 1,$$

3. if $\{A_n\} \subset \mathcal{A}$ and $A_n \downarrow \emptyset$, then for all $\varepsilon > 0$

$$(13.5) \quad \lim_{n \rightarrow \infty} P_{A_n}([\varepsilon, \infty)) = 0.$$

4. for $\varepsilon > 0$ there exists a compact subset K_ε such that $P(K_\varepsilon^c([\varepsilon, \infty)) < \varepsilon$

Theorem 13.8 *Assume (1-4) above. Then there is a unique probability, P , on $(M_F(E), \mathcal{B}(M_F(E)))$ such that*

$$(13.6) \quad P(\{\mu \in M_F(E) : (\mu(A_1), \dots, \mu(A_n)) \in C\}) = P_{A_1, \dots, A_n}(C), \quad C \in \mathcal{B}(\mathbb{R}_+^n).$$

Proof. See for example, Jagers (1974) [318], Harris (1968) [279].

13.4 Topologies on path spaces

Definition 13.9 *Let $\mu_i, \mu \in \mathcal{M}_f$. Then $(\mu_n)_{n \in \mathbb{N}}$ converges weakly to μ as $n \rightarrow \infty$, denoted $\mu_n \Rightarrow \mu$ iff and only if*

$$(13.7) \quad \int f d\mu_n \xrightarrow[n \rightarrow \infty]{} \int f d\mu \quad \forall f \in C_b(E)$$

Given a Polish space (E, d) we consider the space $C_E([0, \infty))$ with the metric

$$(13.8) \quad \tilde{d}(f, g) = \sum_{n=1}^{\infty} 2^{-n} \sup_{0 \leq t \leq n} |f(t) - g(t)|.$$

Then $(C_E([0, \infty)), \tilde{d})$ is also a Polish space. To prove weak convergence in $\mathcal{P}(C_E([0, \infty)), \tilde{d})$ it suffices to prove tightness and the convergence of the finite dimensional distributions.

Similarly the space $D_E([0, \infty)$ of càdlàg functions from $[0, \infty)$ to E with the Skorohod metric \tilde{d} is a Polish space where

$$(13.9) \quad \tilde{d}(f, g) = \inf_{\lambda \in \Lambda} \left(\gamma(\lambda) + \int_0^\infty e^{-u} \left(1 \wedge \sup_t d(f(t \wedge u), g(t \wedge u)) \right) \right)$$

where Λ is the set of continuous, strictly increasing functions on $[0, \infty)$ and for $\lambda \in \Lambda$,

$$(13.10) \quad \gamma(\lambda) = 1 + \left(\sup_t |t - \lambda(t)| \vee \sup_{s \neq t} \left| \frac{\log(\lambda(s) - \lambda(t))}{s - t} \right| \right).$$

Theorem 13.10 (Ethier-Kurtz) (Ch. 3, Theorem 10.2) *Let X_n and X be processes with sample paths in $D_E([0, \infty)$ and $X_n \Rightarrow X$. Then X is a.s. continuous if and only if $J(X_n) \Rightarrow 0$ where*

$$(13.11) \quad J(x) = \int_0^\infty e^{-u} \left[\sup_{0 \leq t \leq u} d(X(t), x(t-)) \right].$$

13.4.1 Sufficient conditions for tightness

Theorem 13.11 (Aldous (1978)) *Let $\{P_n\}$ be a sequence of probability measures on $D([0, \infty), \mathbb{R})$ such that*

- for each fixed t , $P_n \circ X_t^{-1}$ is tight in \mathbb{R} ,
- given stopping times τ_n bounded by $T < \infty$ and $\delta_n \downarrow 0$ as $n \rightarrow \infty$

$$(13.12) \quad \lim_{n \rightarrow \infty} P_n(|X_{\tau_n + \delta_n} - X_{\tau_n}| > \varepsilon) = 0,$$

or

- $\forall \eta > 0 \exists \delta, n_0$ such that

$$(13.13) \quad \sup_{n \geq n_0} \sup_{\theta \in [0, \delta]} P_n(|X_{\tau_n + \theta} - X_{\tau_n}| > \varepsilon) \leq \eta.$$

Then $\{P_n\}$ are tight.

13.4.2 The Joffe-Métivier criteria for tightness of D-semimartingales

We recall the Joffe Métivier criterion ([334]) for tightness of locally square integrable processes.

A càdlàg adapted process X , defined on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ with values in \mathbb{R} is called a *D-semimartingale* if there exists a càdlàg function $A(t)$, a linear subspace $D(L) \subset C(\mathbb{R})$ and a mapping $L : (D(L) \times \mathbb{R} \times [0, \infty) \times \Omega) \rightarrow \mathbb{R}$ with the following properties:

1. for every $(x, t, \omega) \in \mathbb{R} \times [0, \infty) \times \Omega$ the mapping $\phi \rightarrow L(\phi, x, t, \omega)$ is a linear functional on $D(L)$ and $L(\phi, \cdot, t, \omega) \in D(L)$,
2. for every $\phi \in D(L)$, $(x, t, \omega) \rightarrow L(\phi, x, t, \omega)$ is $\mathcal{B}(\mathbb{R}) \times \mathcal{P}$ -measurable, where \mathcal{P} is the predictable σ -algebra on $[0, \infty) \times \Omega$, (\mathcal{P} is generated by sets of the form $(s, t] \times F$ where $F \in \mathcal{F}_s$ and s, t are arbitrary)
3. for every $\phi \in D(L)$ the process M^ϕ defined by

$$(13.14) \quad M^\phi(t, \omega) := \phi(X_t(\omega) - \phi(X_0(\omega)) - \int_0^t L(\phi, X_{s-}(\omega), s, \omega) dA_s,$$

is a locally square integrable martingale on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$,

4. the functions $\psi(x) := x$ and ψ^2 belong to $D(L)$.

The functions

$$(13.15) \quad \beta(x, t, \omega) := L(\psi, x, t, \omega)$$

$$(13.16) \quad \alpha(x, t, \omega) := L((\psi)^2, x, t, \omega) - 2x\beta(x, t, \omega)$$

are called the *local characteristics of the first and second order*.

Theorem 13.12 Let $X^m = (\Omega^m, \mathcal{F}^m, \mathcal{F}_t^M, P^m)$ be a sequence of D-semimartingales with common $D(L)$ and associated operators L^m , functions A^m, α^m, β^m . Then the sequence $\{X^m : m \in \mathbb{N}\}$ is tight in $D_{\mathbb{R}}([0, \infty)$ provided the following conditions are satisfied:

1. $\sup_m E|X_0^m|^2 < \infty$,

2. there is a $K > 0$ and a sequence of positive adapted processes $\{\{C_t^m : t \geq 0\} \text{ on } \Omega^m\}_{m \in \mathbb{N}}$ such that for every $m \in \mathbb{N}, x \in \mathbb{R}, \omega \in \Omega^m$,

$$(13.17) \quad |\beta_m(x, t, \omega)|^2 + \alpha_m(x, t, \omega) \leq K(C_t^m(\omega) + x^2)$$

and for every $T > 0$,

$$(13.18) \quad \sup_m \sup_{t \in [0, T]} E|C_t^m| < \infty, \text{ and } \lim_{k \rightarrow \infty} \sup_m P^m(\sup_{t \in [0, T]} C_t^m \geq k) = 0,$$

3. there exists a positive function γ on $[0, \infty)$ and a decreasing sequence of numbers (δ_m) such that $\lim_{t \rightarrow 0} \gamma(t) = 0$, $\lim_{m \rightarrow \infty} \delta_m = 0$ and for all $0 < s < t$ and all m ,

$$(13.19) \quad (A^m(t) - A^m(s)) \leq \gamma(t - s) + \delta_m.$$

4. if we set

$$(13.20) \quad M_t^m := X_t^m - X_0^m - \int_0^t \beta_m(X_s^m, s, \cdot) dA_s^m,$$

then for each $T > 0$ there is a constant K_T and m_0 such that for all $m \geq m_0$, then

$$(13.21) \quad E(\sup_{t \in [0, T]} |X_t^m|^2) \leq K_T(1 + E|X_0^m|^2),$$

and

$$(13.22) \quad E(\sup_{t \in [0, T]} |M_t^m|^2) \leq K_T(1 + E|X_0^m|^2),$$

Corollary 13.13

Assume that for $T > 0$ there is a constant K_T such that

$$(13.23) \quad \sup_m \sup_{t \leq T, x \in \mathbb{R}} (|\alpha_m(t, x)| + |\beta_m(t, x)|) \leq K_T, \text{ a.s.}$$

$$(13.24) \quad \sum_m (A^m(t) - A^m(s)) \leq K_T(t - s) \text{ if } 0 \leq s \leq t \leq T,$$

and

$$(13.25) \quad \sup_m E|X_0^m|^2 < \infty,$$

and M_t^m is a square integrable martingale with $\sup_m E(|M_T^m|^2) \leq K_T$. The the $\{X^m : m \in \mathbb{N}\}$ are tight in $D_{\mathbb{R}}([0, \infty)$.

Criteria for continuous processes

Now consider the special case of probability measures on $C([0, \infty), \mathbb{R}^d)$. This criterion is concerned with a collection $(X^{(n)}(t))_{t \geq 0}$ of semimartingales with values in \mathbb{R}^d with continuous paths. First observe that by forming

$$(13.26) \quad (< X^{(n)}(t), \lambda >)_{t \geq 0}, \quad \lambda \in \mathbb{R}^d$$

we obtain \mathbb{R} -valued semi-martingales. If for every $\lambda \in \mathbb{R}^d$ the laws of these projections are tight on $C([0, \infty), \mathbb{R})$ then this is true for $\{\mathcal{L}[(X^{(n)}(t))_{t \geq 0}], n \in \mathbb{N}\}$. The tightness criterion for \mathbb{R} -valued semimartingales is in terms of the so-called local characteristics of the semimartingales.

For Itô processes the local characteristics can be calculated directly from the coefficients. For example, if we have a sequence of semimartingales X^n that are also a Markov processes with generators:

$$(13.27) \quad L^{(n)} f = \left(\sum_{i=1}^d a_i^n(x) \frac{\partial}{\partial x_i} + \sum_{i=1}^d \sum_{j=1}^d b_{i,j}^n(x) \frac{\partial^2}{\partial x_i \partial x_j} \right) f$$

then the local characteristics are given by

$$(13.28) \quad a^n = (a_i^n)_{i=1, \dots, d}, \quad b^n = (b_{i,j}^n)_{i,j=1, \dots, d}.$$

The Joffe-Métivier criterion implies that if

$$(13.29) \quad \sup_n \sup_{0 \leq t \leq T} E[|a^n(X^{(n)}(t))| + |b^n(X^{(n)}(t))|^2] < \infty,$$

$$(13.30) \quad \lim_{k \rightarrow \infty} \sup_n P(\sup_{0 \leq t \leq T} (|a^n(X^{(n)}(t))| + |b^n(X^{(n)}(t))|) \geq k) = 0$$

then $\{\mathcal{L}[(X^{(n)}(t))_{t \geq 0}], n \in \mathbb{N}\}$ are tight in $C([0, \infty), \mathbb{R})$. See [334] for details.

Theorem 13.14 (Ethier-Kurtz [212] Chapt. 3, Theorem 10.2) Let

$$(13.31) \quad J(x) = \int_0^\infty e^{-u} [J(x, u) \wedge 1] du, \quad J(x, u) = \sup_{0 \leq t \leq u} d(x(t), x(t-)).$$

Assume that a sequence of processes $X_n \Rightarrow X$ in $D_E([0, \infty))$. Then X is a.s. continuous if and only if $J(X_n) \Rightarrow 0$.

13.4.3 Tightness of measure-valued processes

Lemma 13.15 (Tightness Lemma).

(a) Let E be a compact metric space and $\{P_n\}$ a sequence of probability measures on $D([0, \infty), M_1(E))$. Then $\{P_n\}$ is compact if and only if there exists a linear separating set $D \subset C(E)$ such that $t \rightarrow \int f(x) X_t(\omega, dx)$ is relatively compact in $D([0, \infty), [-\|f\|, \|f\|])$ for each $f \in D$.

(b) Assume that $\{P_n\}$ is a family of probability measures on $D([0, \infty), [-K, K])$ such that for $0 \leq t \leq T$, there are bounded predictable processes $\{v_i(\cdot) : i = 1, 2\}$ such that for each n

$$M_{i,n}(t) := x(\omega, t)^i - \int_0^t v_{i,n}(\omega, s) ds, \quad i = 1, 2$$

are P_n -square integrable martingales with

$$\sup_n E_n(\sup_s (|v_{2,n}(s)| + |v_{1,n}(s)|)) < \infty.$$

Then the family $\{P_n\}$ is tight.

(c) In (b) we can replace the $i = 2$ condition with: for any $\varepsilon > 0$ there exists f and $v_{f,n}$ such that

$$\sup_{[-K, K]} |f_\varepsilon(x) - x^2| < \varepsilon$$

and

$$M_{f,n}(t) := f_\varepsilon(x(\omega, t)) - \int_0^t v_{f_\varepsilon, n}(\omega, s) ds$$

$$\sup_n E_n(\sup_s |v_{f_\varepsilon, n}(s)|) < \infty.$$

Proof. (a) See e.g. Dawson, [120] Section 3.6.
(b) Given stopping times τ_n and $\delta_n \downarrow 0$ as $n \rightarrow \infty$.

$$\begin{aligned} & E_n [(x(\tau_n + \delta_n) - x(\tau_n))^2] \\ &= \{E_n[x^2(\tau_n + \delta_n) - x^2(\tau_n)] - 2E_n[x(\tau_n)(x(\tau_n + \delta_n) - x_n(\tau_n))]\} \\ &\leq E_n \left[\int_{\tau_n}^{\tau_n + \delta_n} |v_{2,n}(s)| ds + 2K \int_{\tau_n}^{\tau_n + \delta_n} |v_{1,n}(s)| ds \right] \\ &\leq \delta_n \sup_n E_n(\sup_s (|v_{2,n}(s)| + |v_{1,n}(s)|)) \\ &\rightarrow 0 \text{ as } \delta_n \rightarrow 0. \end{aligned}$$

The result then follows by Aldous' condition.

(c)

$$\begin{aligned} & E_n [(x(\tau_n + \delta_n) - x(\tau_n))^2] \\ &= \{E_n[x^2(\tau_n + \delta_n) - x^2(\tau_n)] - 2E_n[x(\tau_n)(x(\tau_n + \delta_n) - x_n(\tau_n))]\} \\ &\leq E_n(f(x(\tau_n + \delta_n)) - f(x(\tau_n))) + 2K \int_{\tau_n}^{\tau_n + \delta_n} |v_{1,n}(s)| ds + 2\varepsilon \\ &\leq E_n \left[\int_{\tau_n}^{\tau_n + \delta_n} |v_{f_\varepsilon, n}(s)| ds + 2K \int_{\tau_n}^{\tau_n + \delta_n} |v_{1,n}(s)| ds \right] + 2\varepsilon \\ &\leq \delta_n \sup_n E_n(\sup_s (|v_{f_\varepsilon, n}(s)| + |v_{1,n}(s)|)) + 2\varepsilon \end{aligned}$$

Hence for any $\varepsilon > 0$

$$\begin{aligned} & \lim_{\delta_n \rightarrow 0} \sup_n E_n [(x(\tau_n + \delta_n) - x(\tau_n))^2] \\ &\leq \lim_{n \rightarrow \infty} \delta_n \sup_n E_n(\sup_s (|v_{f_\varepsilon, n}(s)| + |v_{1,n}(s)|)) + 2\varepsilon \\ &= 2\varepsilon. \end{aligned}$$

and the result again follows from Aldous criterion. ■

Remark 13.16 These results can be also used to prove tightness in the case of non-compact E . However in this case an additional step is required, namely to show that for $\varepsilon > 0$ and $T > 0$ there exists a compact subset $K_{T,\varepsilon} \subset E$ such that

$$P_n[D([0, T], K_{T,\varepsilon})] > 1 - \varepsilon \quad \forall n.$$

Remark 13.17 Note that if P_n is a tight sequence of probability measures on $D([0, T], \mathbb{R})$ such that $P_n(\sup_{0 \leq s \leq T} |x(s) - x(s-)| \leq \delta_n) = 1$ and $\delta_n \rightarrow 0$ as $n \rightarrow \infty$, then for any limit point P_∞ , $P_\infty(C([0, T], \mathbb{R})) = 1$.

13.5 The Gromov-Hausdorff metric on the space of compact metric spaces

Let E be a metric space and B_1, B_2 two subsets. Then the Hausdorff distance is defined by

$$(13.32) \quad d_H(K_1, K_2) = \inf\{\varepsilon \geq 0 : K_1 \subset V_\varepsilon(K_2), K_2 \subset V_\varepsilon(K_1)\}$$

where $V_\varepsilon(K)$ denotes the ε -neighbourhood of K . This defines a pseudometric, $d_H(B_1, B_2) = 0$ iff they have the same closures.

If X and Y are two compact metric spaces. The *Gromov-Hausdorff metric* $d_{GH}(X, Y)$ is defined to be the infimum of all numbers $d_H(f(X), g(Y))$ for all metric spaces M and all isometric embeddings $f : X \rightarrow M$ and $g : Y \rightarrow M$ and where d_{Haus} denotes Hausdorff distance between subsets in M . d_{GH} is a pseudometric with $d_{GH}(K_1, K_2) = 0$ iff they are isometric.

Now let (\mathbb{K}, d_{GH}) denote the class of compact metric spaces (modulo isometry) with the Gromov-Hausdorff metric. Then (\mathbb{K}, d_{GH}) is complete.

See Gromov [272] and Evans [229] for detailed expositions on this topic.

13.5.1 Metric measure spaces

The notion of *metric measure space* was developed by Gromov [272] (called mm spaces there). It is given by a triple (X, r, μ) where (X, r) is a metric space such that $(\text{supp}(\mu), r)$ is complete and separable and $\mu \in \mathcal{P}(X)$ is a probability measure on (X, r) . Let \mathbb{M} be the space of equivalence classes of metric measure spaces (whose elements are not themselves metric spaces - see remark (2.2(ii)) in [265]) with equivalence in the sense of measure-preserving isometries. The distance matrix map is defined for $n \leq \infty$

$$(13.33) \quad X^n \rightarrow \mathbb{R}_+^{\binom{n}{2}}, \quad ((x_i)_{i=1,\dots,n}) \rightarrow (r(x_i, x_j))_{1 \leq i < j \leq n}$$

and we denote by $R^{(X, r)}$ the map that sends a sequence of points to its infinite distance matrix.

Then the *distance matrix distribution* of (X, r, μ) (representative of equivalence class) is defined by

$$(13.34) \quad \nu^{(X, r, \mu)} := R^{(X, r)} - \text{pushforward of } \mu^{\otimes \mathbb{N}} \in \mathcal{P}(\mathbb{R}_+^{\binom{\mathbb{N}}{2}}).$$

Since this depends only on the equivalence class it defined the mapping $\kappa \rightarrow \nu^\kappa$ for $\kappa \in \mathbb{M}$. Gromov [272] (Section 3 $\frac{1}{2}$.5) proved that a metric measure space is characterized by its distance matrix distribution.

Greven, Pfaffelhuber and Winter (2008) [265] introduced the *Gromov-weak topology*. In this topology a sequence $\{\chi_n\}$ converges Gromov-weakly to χ in \mathbb{M} if and only if $\Phi(\chi_n)$ converges to $\Phi(\chi)$ in \mathbb{R} for all polynomial in Π .

In [265], Theorem 1, they proved that \mathbb{M} equipped with the Gromov-weak topology is Polish.

An important subclass is the set of ultrametric measure spaces given by the closed subset of \mathbb{M}

$$(13.35) \quad \mathbb{U} := \{ u \in \mathbb{M} : u \text{ is ultra-metric}\}.$$

13.6 Riemannian metrics and gradient

Let M be a smooth manifold. The tangent space at x , $T_x M$ can be identified with the space of tangents at x to all smooth curves through x . The tangent bundle $TM = \{(p, v) : p \in M, v \in T_p M\}$.

Definition 13.18 A Riemannian metric on M is a smooth tensor field

$$g : C^\infty(TM) \otimes C^\infty(TM) \rightarrow C_0^\infty(M)$$

such that for each $p \in M$,

$$g(p)|_{T_p M \otimes T_p M} : T_p M \otimes T_p M \rightarrow \mathbb{R}$$

with

$$g(p) : (X, Y) \rightarrow \langle X, Y \rangle_{g(p)}$$

where $\langle X, Y \rangle_{g(p)}$ is an inner product on $T_p M$.

Definition 13.19 The directional derivative in direction v is defined by

$$\begin{aligned} \partial_v f(x) &= \lim_{t \rightarrow 0} \frac{f(x + tv) - f(x)}{t} \\ &= \sum v_i \frac{\partial f(x)}{\partial x_i} \end{aligned}$$

The gradient $\nabla_g f(x)$ is defined by

$$\langle \nabla_g f(x), v \rangle_g = \partial_v f(x) \quad \forall v \in T_x M.$$

Example 13.20 Consider the d -dimensional manifold $M = \mathbb{R}^d$ and $\mathbf{a}(\cdot)$ be a smooth map from M to $\mathbb{R}^d \otimes \mathbb{R}^d$ ($(d \times d)$ -matrices). We will write

$$\begin{aligned} \mathbf{a}(x) &= (a_{ij}(x)) \\ \mathbf{a}^{-1}(x) &= (a^{ij}(x)) \end{aligned}$$

Assume that

$$\sum a^{ij}(x) u_i u_j \geq \gamma \sum u_j^2, \quad \gamma > 0.$$

The tangent space $T_x M \cong \mathbb{R}^d$ and we define a Riemannian metric on M by

$$g_{\mathbf{a}(x)}(\mathbf{u}, \mathbf{v}) := \sum_{i,j=1}^d a_{ij}(x) u^i v^j.$$

The associated Riemannian gradient and norm are

$$\begin{aligned} (\nabla_{\mathbf{a}} f)^i &= \sum_j a^{ij} \frac{\partial f}{\partial x_j} \\ \|u\|_{\mathbf{a}(x)}^2 &= \sum_{ij} a_{ij}(x) u^i u^j. \end{aligned}$$

The Shahshahani metric and gradient on Δ_{K-1}

Let $M_K = \mathbb{R}_+^K := \{x \in \mathbb{R}^K, x = (x_1, \dots, x_K), x_i > 0 \text{ for all } i\}$ is a smooth K-dimensional manifold.

Shahshahani introduced the following Riemannian metric on M_K

$$\langle u, v \rangle_g = g_x(u, v) := \sum_{i=1}^K |x| \frac{u_i v_i}{x_i}$$

$$|x| = \sum x_i$$

$\|\cdot\|_g$ and $\nabla_g F$ will denote the corresponding norm and gradient. We have

$$(\nabla_g F)^i = \sum_i \frac{x^i}{|x|} \frac{\partial F}{\partial x^i} \frac{\partial}{\partial x^i}$$

Recall that the simplex $\Delta_{K-1} := \{(p_1, \dots, p_K) : p_i \geq 0, \sum_{i=1}^K p_i = 1\}$. The interior of the simplex $\Delta_{K-1}^0 = \mathbb{R}_+^K \cap \Delta_{K-1}$ is a $(K-1)$ -dimensional submanifold of M_K . We denote by $T_p \Delta_{K-1}^0$ the tangent space to Δ_{K-1}^0 at p . Then g induces a Riemannian metric on $T_p \Delta_{K-1}^0$.

Basic Facts

We have the Shahshahani inner product on Δ_{K-1} at a point $p \in \Delta_{K-1}$:

$$(13.36) \quad \langle u, v \rangle_p = \sum_{i=1}^K \frac{u_i v_i}{p_i}.$$

1. $T_p \Delta_{K-1}^0$ can be viewed as the subspace of $T_p M_K$ of vectors, v , satisfying $\langle p, v \rangle_g = 0$ if we identify p with an element of $T_p M_K$.

Proof. Recall that $T_p \Delta_{K-1}^0$ is given by tangents to all smooth curves lying in Δ_{K-1}^0 . Therefore if $v \in T_p \Delta_{K-1}^0$, then $v = q - p$ where $p, q \in \Delta_{K-1}^0$ and therefore $\sum_{i=1}^K v_i = 0$. Therefore,

$$\sum_i p_i \frac{1}{p_i} v_i = 0.$$

2. If $F : \Delta_{K-1}^0 \rightarrow \mathbb{R}$ is smooth, then the Shahshahani gradient is

$$(\nabla_g F)_i = p_i \left(\frac{\partial F}{\partial p_i} - \sum_j p_j \frac{\partial F}{\partial p_j} \right).$$

Proof. From the definition, $\nabla_g F$ is the orthogonal projection on the subspace $T_p \Delta_{K-1}^0$ of

$$(\nabla_g F)_i = p_i \frac{\partial F}{\partial p_i}$$

and therefore we must have $\sum_i (\nabla_g F)_i = 0$. This then gives the result.

Chapter 14

Appendix III: Markov Processes

14.1 Operator semigroups

See Ethier-Kurtz, [212] Chap.1.

Consider a strongly continuous semigroup $\{T_t\}$ with generator G and domain $D(G)$. A subset $D_0 \subset D(G)$ is a *core* if the closure of $G|_{D_0}$ equals G . If D_0 is dense and $T_t : D_0 \rightarrow D_0$ for all t , then it is a core.

Theorem 14.1 (*Kurtz semigroup convergence Theorem [212], Chap. 1, Theorem 6.5*) Let L, L_n be Banach spaces and $\pi_n : L \rightarrow L_n$ is a bounded linear mapping and $\sup_n \|\pi_n\| < \infty$. We say $f_n \in L_n \rightarrow f \in L$ if $\lim_{n \rightarrow \infty} \|f_n - \pi_n f\| = 0$.

For $n \in \mathbb{N}$ let T_n be a contraction on a Banach space L_n , let $\varepsilon_n > 0$, $\lim_{n \rightarrow \infty} \varepsilon_n = 0$. Let $\{T(t)\}$ be a strongly continuous contraction semigroup on L with generator A and let D be a core for A . Then the following are equivalent:

- (a) For each $f \in L$, $T_n^{\lfloor t/\varepsilon_n \rfloor} \pi_n f \rightarrow T(t)f$, for all $t \geq 0$, uniformly on bounded intervals.
- (b) For each $f \in D$ there exists $f_n \in L_n$ such that $f_n \rightarrow f$ and $A_n f_n \rightarrow Af$.

Theorem 14.2 [212] Chap. 4, Theorem 2.5.

Let E be locally compact and separable. For $n = 1, 2, \dots$ let $\{T_n(t)\}$ be a Feller semigroup on $C_0(E)$ and suppose that X_n is a Markov process with semigroup $\{T_n(t)\}$ and sample paths in $D_E([0, \infty))$. Suppose that $\{T(t)\}$ is a Feller semigroup on $C_0(E)$ such that for each $f \in C_0(E)$

$$(14.1) \quad \lim_{n \rightarrow \infty} T_n(t)f = T(t)f, \quad t \geq 0.$$

If $\{X_n(0)\}$ has limiting distribution $\nu \in \mathcal{P}(E)$, then there is a Markov process X corresponding to $\{T(t)\}$ with initial distribution ν and sample paths in $D_E([0, \infty))$ with initial distribution ν and sample paths in $D_E([0, \infty))$ and $X_n \Rightarrow X$.

14.2 Some basic result for one dimensional diffusions

Basic References: Itô-McKean [315], Karlin and Taylor [343], Revuz and Yor [493].

14.2.1 Boundary behaviour classification

Consider a diffusion process on $[0, L]$ or $[0, \infty)$ with drift and diffusion coefficients $b(x)$ and $\sigma^2(x)$. The scale function $S(x)$ is defined by

$$(14.2) \quad S(x) = \int_{x_0}^x s(u)du, \quad s(x) = \exp\left(-\int_{x_0}^x \frac{2b(u)}{\sigma^2(u)}du\right),$$

and the speed measure

$$(14.3) \quad M([x_1, x_2]) = \int_{x_1}^{x_2} m(x)dx, \quad m(x) = \frac{1}{\sigma^2(x)s(x)}.$$

Feller introduced a classification of boundary points as follows: Applied to the boundary point 0 this becomes (following Karlin and Taylor [343]):

$$(14.4) \quad \Sigma(0) = \int_0^x S(0, u)m(u)du, \quad N(0) = \int_0^x (S(x) - S(u))m(u)du.$$

Then

- 0 is an *entrance* boundary if $S(0, x] = \infty$ and $N(0) < \infty$.
- 0 is an *exit* boundary if $\Sigma(0) < \infty$ but $M(0, x] = \infty$
- 0 is a regular boundary if $S(0, x] < \infty$ and $M(0, x] < \infty$.

For the Feller CSBP process, 0 is an exit boundary. For the Feller CSBP process with immigration (4.39)

$$(14.5) \quad s(x) = \exp\left(-\int_{x_0}^x \frac{2\beta}{4x}\right) = \left(\frac{x}{x_0}\right)^{-\beta/2}, \quad m(x) = \frac{1}{4x\left(\frac{x}{x_0}\right)^{-\beta/2}} = \frac{4}{x_0}x^{1-\frac{\beta}{2}}$$

14.2.2 Excursions

The Brownian excursion

Let $\{B(t)\}_{t \geq 0}$ be a standard Brownian motion and

$$(14.6) \quad \tau_1 := \sup\{t \in [0, 1] : B(t) = 0\}, \quad \tau_2 := \inf\{t \geq 1 : B(t) = 0\}.$$

Then the Brownian excursion is a nonhomogeneous Markov process defined as follows:

$$(14.7) \quad B^e(t) := \frac{1}{\sqrt{(\tau_2 - \tau_1)}}B(\tau_1 + t(\tau_2 - \tau_1)), \quad 0 \leq t \leq 1.$$

It can be shown (see Itô-McKean) [315] that the marginal PDF is given by

$$(14.8) \quad f(t, x) = \frac{2x^2}{\sqrt{2\pi t^3(1-t)^3}} e^{-\frac{x^2}{2t(1-t)}}.$$

Itô's excursion measure $n(de)$ is the σ -finite measure on $C(\mathbb{R}_+, \mathbb{R}_+)$ obtained by

$$(14.9) \quad n(de) = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} P_\varepsilon(de)$$

where $P_\varepsilon(de)$ is the distribution of Brownian motion started at ε and stopped at the first time it hits $\zeta(e) = \inf\{s > 0 : e(s) = 0\}$. Then the *normalized Brownian excursion* is given by $n(de|\zeta(e) = 1)$.

Excursions of non-negative diffusions

Now consider a general diffusion on $[0, \infty)$ which is regular on $(0, \infty)$ and with 0 as an absorbing boundary and with laws $\{P_x : 0 \leq x < \infty\}$. Let T_y be the hitting time of y and that 0 is an exit point for the diffusion.

Revuz and Yor introduced the *excursion law of the diffusion* in terms of a σ -finite measure Λ on C . Under Λ the trajectories come in from zero according to an entrance law, then move according to the diffusion.

Let $s(x)$ be a scale function for the diffusion. Since we assume the absorbing point 0 can be reached with positive probability from $x > 0$ we can take

$$(14.10) \quad s(0) = 0, \quad s(x) > 0 \quad \text{for } x > 0$$

and s is defined uniquely up to a constant factor by

$$(14.11) \quad P_x(T_y < \infty) = \frac{s(x)}{s(y)}, \quad 0 < x < y < \infty.$$

Then there exists a σ -finite excursion law \mathbb{Q} on

$$(14.12) \quad W_0 := \{w \in C([0, \infty), \mathbb{R}^+), w(0) = 0, w(t) > 0 \text{ for } 0 < t < \zeta \\ \text{for some } \zeta \in (0, \infty)\}$$

obtained as follows. Denoting by P^ε the law of the process started with $w(0) = \varepsilon$ and $\varepsilon > 0$, \mathbb{Q} is given by:

$$(14.13) \quad \mathbb{Q}(\cdot) = \lim_{\varepsilon \rightarrow 0} \frac{P^\varepsilon(\cdot)}{S(\varepsilon)},$$

where $S(\cdot)$ is the scale function of the diffusion defined by the relation,

$$(14.14) \quad P_\varepsilon(T_\eta < \infty) = \frac{S(\varepsilon)}{S(\eta)}, \quad 0 < \varepsilon < \eta < \infty.$$

Bibliography

- [1] E. Akin (1979). The geometry of population genetics, Lect. Notes in Biomath. 31, Springer-Verlag.,
- [2] E. Akin (1982). Cycling in simple genetic systems, J. Math. Biology 13, 305-324.
- [3] E. Akin (1983). Hopf bifurcation in the two locus genetic model, Memoirs AMS 284.
- [4] D. Aldous (1991) The continuum random tree I. Ann. Probab. 19, 1-28.
- [5] D. Aldous (1991) The continuum random tree II: An overview. In Stochastic Analysis (M.T. Barlow, N.H. Bingham, eds), 23-70, Cambridge Univ. Press.
- [6] D. Aldous (1993) The continuum random tree III. Ann. Probab. 21, 248-289.
- [7] D. Aldous (1997). Brownian excursions, critical random graphs and the multiplicative coalescent, Ann. Probab. 25, 812-854.
- [8] K.B. Athreya and P.E. Ney (1972) Branching Processes, Springer-Verlag.
- [9] K.B. Athreya and N. Kaplan (1976). Convergence of the age distribution in the one-dimensional age-dependent branching process, Ann. Probab. 4, 38-50.
- [10] S.R. Athreya, M.T. Barlow, R.F. Bass and E.A. Perkins, Degenerate stochastic differential equations and super-Markov chains, Probab. Theory Related Fields, 123(2002), 484-520.
- [11] S.R. Athreya and J.M. Swart (2005). Branching-coalescing particle systems, PTRF 131, 376-414.
- [12] S.R. Athreya, R.F. Bass, M. Gordina and E.A. Perkins (2005). Infinite dimensional stochastic differential equations of Ornstein-Uhlenbeck type, arXiv:math.PR\ 0503165.
- [13] S.R. Athreya and J.M. Swart (2008). Survival of contact process on the hierarchical group arXiv:0808.3732v1
- [14] E. Baake (1995). Diploid models on sequence space,, J. Biol. Syst. 3,, 343, 349-.
- [15] E. Baake and W. Gabriel (1999). Biological evolution through mutation, selection and drift: an introductory review, arXiv:cond-mat/9907372v1.
- [16] M. Baake and E. Baake (2003). An exactly solved model for mutation, recombination and selection, Canad. J. Math. 55, 3-41.
- [17] J.-B. Baillon, Ph. Clment, A. Greven, F. den Hollander (1995). On the attracting orbit of a non-linear transformation arising from renormalization of hierarchically interacting diffusions. I. The compact case. Canad. J. Math. 47, no. 1, 3-27.
- [18] J. -B. Baillon, Ph. Clment, A. Greven, F. den Hollander (1997). On the attracting orbit of a nonlinear transformation arising from renormalization of hierarchically interacting diffusions. II. The non-compact case. J. Funct. Anal. 146, no. 1, 236-298.

- [19] N.T.J. Bailey (1975). The Mathematical Theory of Infectious Diseases and its Applications. Griffin, London
- [20] A.-L. Barabasi and Z.N. Oltvai (2004). Network biology: understanding the cell's functional organization, *Nature Reviews , Genetics* Vol. 5, 101-113.
- [21] A.D. Barbour, S.N. Ethier and R.C. Griffiths (2000). A transition function for a diffusion model with selection, *Ann. Appl. Probab.* 10, 123-162.
- [22] N. H. Barton, A. M. Etheridge and A.K. Sturm (2004). Coalescence in a random background, *Ann. Appl. Probab.* 14, 754-785.
- [23] N.H. Barton (1995). A general model for the evolution of recombination. *Genet Res* 65, 123-144.
- [24] N.H. Barton and M. Turelli (1997). A critique of Sewall Wright's shifting balance theory of evolution, *Evolution* 31, 643-671.
- [25] N.H. Barton and B. Charlesworth (1998). Why sex and recombination? *Science* 281: 1986-1990.
- [26] N. H. Barton, J. Polechov (2005). The limitations of adaptive dynamics as a model of evolution: *Journal of Evolutionary Biology* 19: 1186-1190.
- [27] N.H. Barton and S.P. Otto (2005). Evolution of recombination due to random drift, *Genetics* 169, 2353-2370.
- [28] N.H. Barton and H.P. de Vladar (2009). Statistical mechanics and the evolution of quantitative traits, *Genetics*, to appear.
- [29] R.F. Bass and E.A. Perkins. Degenerate stochastic differential equations with Hölder continuous coefficients and super-Markov chains, *Trans. Amer. Math. Soc.* 355(2003), 373-405.
- [30] R.F. Bass and E. Perkins (2004). Countable systems of degenerate stochastic differential equations with applications to super-Markov chains, *Electron. J. Probab.* 9, 634673.
- [31] R.F. Bass and E.A. Perkins (2008) Degenerate stochastic differential equations arising from catalytic branching networks, preprint.
- [32] J. Bertoin and J.F. Le Gall (2000). The Bolthausen-Sznitman coalescent and the genealogy of continuous-state branching processes, *PTRF*, 117, 249-266.
- [33] J. Bertoin and J.F. Le Gall (2003). Stochastic flows associated to coalescent processes, *Probab. Theory Related Fields* 126 (2003) 261-288.
- [34] J. Bertoin and J.F. Le Gall (2006). Stochastic flows associated to coalescent processes III, Limit Theorems, *Ill. J. Math.* 50, 147-181.
- [35] J. Bertoin (2006). Random Fragmentation and Coagulation Processes, Cambridge Univ. Press.
- [36] I.J. Bienaymé (1845). De la loi de multiplication et de la dure des familles. *Soc. Philomath. Paris Extraits*, Sr. 5, 3739.
- [37] J.D. Biggins (1978). The Asymptotic Shape of the Branching Random Walk, *Advances in Applied Probability*, Vol. 10, No. 1 (Mar., 1978), pp. 62-84
- [38] P. Billingsley (1999). Convergence of Probability Measures. John Wiley, 2nd edition.
- [39] M. Birkner, J. Blath, M. Capaldo, A. Etheridge, M. Möhle, J. Schweinsberg, A. Wakolbinger. Alpha-stable branching and beta-coalescents, *Electron. J. Probab.* 10(2005), 303-325.
- [40] D. Blackwell and D.G. Kendall (1964) The Martin boundary for Polya's urn scheme and an application to stochastic population growth, *J. Appl. Prob.* 1, 284-296.

- [41] D. Blackwell and L.E. Dubins (1983) An extension of Skorohod's almost sure representation theorem, Proc. A.M.S. 89, 691-692.
- [42] T. Bojdecki, L. G. Gorostiza and A. Talarczyk (2006). Limit theorems for occupation time fluctuations of branching systems I: Long-range dependence, Stochastic. Process. Appl. 116, pp. 1-18.
- [43] T. Bojdecki, L. G. Gorostiza and A. Talarczyk (2006). Limit theorems for occupation time fluctuations of branching systems II: Critical and large dimensions functional, Stochastic Process. Appl. 116, pp. 19-35.
- [44] T. Bojdecki, L. G. Gorostiza and A. Talarczyk (2007). A long range dependence stable process and an infinite variance branching system, Ann. Probab. 35, pp. 500-527.
- [45] T. Bojdecki, L. G. Gorostiza and A. Talarczyk (2007). Occupation time fluctuations of an infinite variance branching system in large dimensions, Bernoulli 13, pp. 20-39.
- [46] B.M. Bolker and S.W. Pacala (1997). Using moment equations to understand stochasticaly driven spatial pattern formation in ecological systems, Theor. Pop. Biol. 52, 179-197.
- [47] R.M. Brandon and R.M. Burian (1984). Genes, organisms and populations, MIT Press.
- [48] M.D. Bramson and D. Griffeath (1979) Renormalizing the 3-dimensional voter model, Ann. Probab. 7, 418-432.J.-B.
- [49] M. Bramson and D. Griffeath (1980) Clustering and dispersion rates for some interacting particle systems on \mathbb{Z} , Ann. Probab. 8, 183-213.
- [50] M. Bramson and D. Griffeath (1980). On the Williams-Bjerknes tumour model II, Camb. Phil Soc. 88, 339-357.
- [51] M. Bramson and D. Griffeath (1981). On the Williams-Bjerknes tumour model I, Ann. Probab. 9, 173-185.
- [52] M. Bramson (1983). Convergence of solutions of the Kolmogorov equation to travelling waves, Memoirs AMS 285.
- [53] M. Bramson, J.T. Cox and D. Griffeath (1988) Occupation time large deviation of the voter model, Prob. Th. Rel. Fields 77, 401-413.
- [54] M. Bramson, R. Durrett and G. Swindle (1989). Statistical mechanics of crabgrass, Ann. Probab. 17, 444-481.
- [55] M. Bramson, J.T. Cox and A. Greven (1993) Ergodicity of a critical spatial branching process in low dimensions, Ann. Probab. 21, 1946-1957.
- [56] M. Bramson, J.T. Cox and A. Greven (1997). Extremal invariant measures of critical spatial branching processes in high dimensions, Annals Probab. 25, 56-70.
- [57] M. Bramson, J.T. Cox and J-F Le Gall (2001) Super-Brownian limits of voter model clusters, Ann Probab. 29(2001), 1001-1032.
- [58] L. Breiman (1968) Probability, Addison-Wesley.
- [59] D.C. Brydges, S.E. Evans and J.Z. Imbrie. Self-avoiding walk on a hierarchical lattice in four dimensions, Ann. Prob. 20 (1992), 82-124.
- [60] R. Bürger (1989). Mutation-selection models in population genetics and evolutionary game theory, Acta Applicandae Mathematicae 14, 75-89.
- [61] R. Bürger: "The mathematical theory of selection, recombination, and mutation", Wiley 2000.

- [62] R.S. Cantrell and C. Cosner (2003). Spatial Ecology via Reaction-diffusion Equations, John Wiley and Sons.
- [63] R. Cerf (1996). The dynamics of mutation-selection algorithms with large population sizes, Ann. l'Institut H. Poincaré B, 32, 455-508.
- [64] R. Cerf (1998). Asymptotic convergence of genetic algorithms, Adv. Appl. Prob. 30, 521-550.
- [65] Champagnat N., R. Ferriere and G. Ben Arous (2001). The canonical equation of adaptive dynamics: a mathematical view. Selection 2:73-84
- [66] N. Champagnat (2006) A microscopic interpretation for adaptive dynamics trait substitution sequence models, SPA 116, 1127-1160.
- [67] N. Champagnat and A. Lambert (2007). Evolution of discrete populations and the canonical diffusion of adaptive dynamics, Ann. Appl. Probab. 17, 102-155.
- [68] N. Champagnat and S. Méléard (2008). Polymorphic evolution sequence and evolutionary branching, Ecole Polytechnique CMA preprint.
- [69] N. Champagnat (2009). Large deviations for singular and degenerate diffusio models in adaptive dynamics, arXiv:0903.2345v1.
- [70] B. Chauvin (1986). Arbres et processus de Bellman Harris, Annales de l'I.H.P. B 22, 209-232.
- [71] M.-F. Chen (1992, 2004). From Markov chains to non-equilibrium particle systems, 1st ed. 1992, 2nd ed. 2004, World Scientific Publ. Co.
- [72] A. J. Chorin, T. J. R. Hughes, M. F. McCracken, J. E. Marsden (1978). Product formulas and numerical algorithms, Comm. Pure and Appl. Math., 31, 205-256.
- [73] F.B. Christiansen (1999). Population genetics of multiple loci, John Wiley.
- [74] P. Clifford and A. Sudbury (1973). A model for spatial conflict, Biometrika 60, 581-588.
- [75] J.E. Cohen and C.M. Newman (1985). When will a large complex system be stable?, J. Theor. Biol. 113, 153-156.
- [76] J.T. Cox and D. Griffeath (1983) Occupation time limit theorems for the voter model, Ann. Probab. 11, 876-893.
- [77] J.T. Cox and D. Griffeath (1985) Occupation times for critical branching Brownian motions, Ann. Probab. 13, 1108-1132.
- [78] J.T. Cox and D. Griffeath (1985). Large deviations for some infinite particle system occupation times, Contemp. Math., 41, 43-53.
- [79] J.T. Cox and D. Griffeath (1986) Diffusive clustering in the two dimensional voter model, Ann. Probab. 14, 347-370.
- [80] J.T. Cox (1988) Some limit theorems for voter model occupation times, Ann. Probab. 16, 1559-1569.
- [81] J.T. Cox (1989). Coalescing random walks and voter model consensus times on the torus in \mathbb{Z}^d , Ann. Probab. 17, 1333-1366.
- [82] J.T. Cox and D. Griffeath (1988) Recent results on the stepping stone model, IMA vol. 8.
- [83] J.T. Cox and D. Griffeath (1990) Mean field asymptotics for the planar stepping stone model, Proc. London Math. Soc. 61, 189-208.

- [84] J.T. Cox and A. Greven (1990). On the long term behavior of some finite particle systems, Probab. Th. Rel Fields 85, 195-237.
- [85] J.T. Cox and A. Greven (1991) On the long term behavior of some finite particle systems: a critical dimension example, Random Walks, Brownian Motion and Interacting Particle Systems, R. Durrett and H. Kesten Birkhäuser, 203-213.
- [86] J.T. Cox (1992) On the ergodic theory of critical branching Markov chains, Stoch. Proc. Appl.
- [87] J.T. Cox and A. Greven (1993) The finite systems scheme: an abstract theorem and a new example. Proc. Workshop and Conference on Measure-valued Processes, Stochastic Partial Differential Equations and Interacting Systems, CRM.
- [88] J.T. Cox and A. Greven (1994) Ergodic theorems for infinite systems of locally interacting diffusions, Ann. Probab. 22(2), 833-853.
- [89] J.T. Cox and A. Greven (1994). The finite systems scheme: an abstract theorem and a new example, in CRM Proc. 5, 55-68, Amer. Math. Soc.
- [90] J.T. Cox, A. Greven and T. Shiga (1995) Finite and Infinite systems of interacting diffusions, Probab. Th. Rel. Fields 103(2), 165-197.
- [91] J.T. Cox, K. Fleischmann, A. Greven (1996). Comparison of interacting diffusions and an application to their ergodic theory, Prob. Th. Rel. Fields 105(4), 513-528.
- [92] T. Cox, R. Durrett and E. Perkins (1999). Rescaled contact processes converge to super-Brownian motion in two or more dimensions, Prob. Th. Rel. Fields 114, 309-399,
- [93] T. Cox, R. Durrett and E. Perkins (1999). Rescaled particle systems converging to super-Brownian motion, 269-284, in Perplexing Problems in Probability: Festschrift in Honor of Harry Kesten, ed. M. Bramson and R. Durrett, Birkhauser.
- [94] J.T. Cox, A. Klenke and E.A. Perkins (2000) Convergence to equilibrium and linear systems duality, in vol. 26, Conf Proc., Can. Math. Soc. 4-66, AMS.
- [95] J.T. Cox and A. Klenke (2000) Recurrence and ergodicity of interacting particle systems, Prob. Th. Rel. Fields 116(2), 239-255.
- [96] T. Cox, R. Durrett and E.A. Perkins (2000) Rescaled voter models converge to super-Brownian motion, Ann. Probab. 28, 185-234.
- [97] J.T. Cox and R. Durrett (2002) The stepping stone model: formulas expose old myths, Ann. Appl. Probab. 12, 1348-1377.
- [98] J.T. Cox and A. Klenke (2003). Rescaled interacting diffusions converge to super Brownian motion, Ann. Appl. Probab. 13, 501-513.
- [99] J.T. Cox, D.A. Dawson and A. Greven, *Mutually Catalytic Super Branching Random Walks: Large Finite Systems and Renormalization Analysis*, Memoirs of the American Mathematical Society 809, American Mathematical Society, Providence, Rhode Island, 2004.
- [100] J.T. Cox and E.A. Perkins (2004) An application of the voter-model-super-Brownian motion invariance principle, Ann. Inst. H. Poincaré Probab. Statist. 40(2004), 25-32.
- [101] J.T. Cox and E.A. Perkins (2005). Rescaled Lotka-Volterra models converge to super-Brownian motion, Ann Probab 33, 904-947.
- [102] J.T. Cox, R. Durrett and I. Zähle (2005) The stepping stone model, II. Genealogies and the infinite sites model. Ann. Appl. Probab. 15, 671-699.

- [103] T. Cox and E.A. Perkins (2007) Survival and coexistence in stochastic spatial Lotka-Volterra models, *Probab. Theory Rel. Fields* 139, 89-142.
- [104] J. F. Crow and M. Kimura (1970). An introduction to population genetics theory, Harper and Row, New York.
- [105] K. Crump and C.J. Mode (1968), (1969). A general age-dependent branching process, I,II, *J. Math. Anal. Appl.* 24, 494-508, 25, 8-17.
- [106] C. Darwin (1859). On the Origin of Species by Means of Natural Selection, John Murray, London
- [107] D.A. Dawson (1975). Stochastic evolution equations and related measure-valued processes, *J. Mult. Analysis* 5, 1-52.
- [108] D.A. Dawson (1977). The critical measure diffusion, *Z. Wahr. verw Geb.* 40, 125-145.
- [109] D.A. Dawson and K.J. Hochberg (1979). The carrying dimension of a stochastic measure diffusion, *Ann. Probab.* 7, 693-703.
- [110] D.A. Dawson (1978). Geostochastic calculus, *Canadian Journal of Statistics* 6, 143-168.
- [111] D.A. Dawson and K.J. Hochberg (1982). Wandering random measures in the Fleming-Viot process, *Ann. Prob.* 10, 554-580.
- [112] D.A. Dawson and T.G. Kurtz (1982). Applications of duality to measure-valued diffusions, *Springer Lecture Notes in Control and Inf. Sci.* 42, 177-191.
- [113] D.A. Dawson and K.J. Hochberg (1983). Qualitative behavior of a selectively neutral allelic model, *Theor. Pop. Biol.* 23, 1-18.
- [114] D.A. Dawson and K.J. Hochberg (1985). Function-valued duals for measure-valued processes with applications, *Contemporary Mathematics* 41, 55-69.
- [115] D.A. Dawson and K. Fleischmann (1988). Strong clumping of critical space-time branching models in subcritical dimensions, *Stoch. Proc. Appl.* 30, 193-208.
- [116] D.A. Dawson and J. Gärtner (1989). Large deviations, free energy functional and quasipotential for a mean-field model of interacting diffusions, *Memoirs of the American Mathematical Society*, **398**, 94+(iv) pages.
- [117] D.A. Dawson and K.J. Hochberg and Y. Wu (1990). Multilevel branching systems, in *Proc. Bielefeld Encounters in Mathematics and Physics* (1989), World Scientific, 93-107.
- [118] D.A. Dawson and E.A. Perkins (1991). Historical processes. *Mem. Amer. Math. Soc.* 454.
- [119] D.A. Dawson (1992). Infinitely divisible random measures and superprocesses, in *Stochastic Analysis and Related Topics*, H. Korezlioglu and A.S. Ustunel, eds., *Progress in Probability* vol. 31, Birkhauser, 1-130.
- [120] D.A. Dawson (1993). Measure-valued Markov processes. In P.L. Hennequin, editor, *École d'été de probabilités de Saint Flour XXI-1991*, volume 1541 of *Lecture Notes in Mathematics*, pages 1–260. Springer-Verlag, Berlin.
- [121] D.A. Dawson and A. Greven, Multiple scale analysis of interacting diffusions, *Probab. Theory Relat. Fields* 95 (1993) 467–508.
- [122] D.A. Dawson and A. Greven (1993). Hierarchical models of interacting diffusions: multiple time scale phenomena, phase transition and pattern of cluster-formation, *Probab. Theory Related Fields* **96**, 435-473.

- [123] D.A. Dawson and J. Gärtner (1994). Multilevel large deviations and interacting diffusions, *Probab. Th. Rel. Fields*, 98(1994), 423-487.
- [124] D.A. Dawson, A. Greven and J. Vaillancourt (1995). Equilibria and quasiequilibria for infinite collections of interacting Fleming-Viot processes, *Trans. Amer. Math. Soc.* 347, 2277-2360.
- [125] D.A. Dawson and P. March (1995). Resolvent estimates for Fleming-Viot operators and uniqueness of solutions to related martingale problems, *J. Functional Anal.* 132, 417-472.
- [126] D.A. Dawson, K.J. Hochberg and V. Vinogradov (1996). High-density limits of hierarchically structured branching-diffusing populations, *Stochastic Processes and their Applications*, 62, 191-222.
- [127] D.A. Dawson and A. Greven (1996). Multiple space-time scale analysis for interacting branching models, *Electronic Journal of Probability*, 1, paper 14, 1-84.
- [128] D.A. Dawson and Y. Wu (1996). Multilevel multitype models of an information system, *I.M.A. Volume 84, Classical and Modern Branching Processes*, eds. K.B. Athreya and P. Jagers, Springer-Verlag, pages 57-72.
- [129] D.A. Dawson Hierarchical and mean-field stepping stone models (1997) Hierarchical and mean-field stepping stone models, *I.M.A. Volume 87, Progress in Population Genetics and Human Evolution*, eds. P. Donnelly and S. Tavaré, 287-298.
- [130] D.A. Dawson and E.A. Perkins (1988). Long-time behavior and coexistence in a mutually catalytic branching model, 26(3):1088-1138.
- [131] D.A. Dawson and E.A. Perkins (1999). *Measure-valued processes and Renormalization of Branching Particle Systems*, in Stochastic Partial Differential Equations: Six Perspectives, eds. R. Carmona and B. Rozovskii, American Mathematical Society Mathematical Surveys and Monographs Vol. **64**, pp 45-106.
- [132] D.A. Dawson and A. Greven (1999) Hierarchically interacting Fleming-Viot processes with selection and mutation, *Electronic J. Probab.* Vol. 4, paper 4, pages 1-81.
- [133] D.A. Dawson, L.G. Gorostiza and A. Wakolbinger (2001). Occupation time fluctuations in branching systems, *J. Theor. Probab.* 14, 729-796.
- [134] D.A. Dawson, L.G. Gorostiza and A. Wakolbinger, Degrees of transience and recurrence and hierarchical random walks, *Potential Analysis* 22 (2005) 305-350.
- [135] D.A. Dawson and A. Greven (2003). State dependent spatial branching processes and their long-time behavior, *EJP* 8, paper 4.
- [136] D. A. Dawson and A. Greven and F. den Hollander and R. Sun and J. M. Swart (2008). The renormalization transformation for two-type branching models, *Ann. l'Institut H. Poincaré*.
- [137] D.A. Dawson and A. Greven (2009). On the effects of migration in spatial Fleming-Viot models with selection and mutation, in preparation.
- [138] D.A. Dawson and S. Feng (2001). Large deviations for the Fleming-Viot process with neutral mutation and selection II, *Stochastic Processes and their Applications* 92, 131-162.
- [139] D.A. Dawson and K. Fleischmann (2000). Catalytic and mutually catalytic branching, in *Infinite Dimensional Stochastic Analysis*, eds. Ph. Clément, F. den Hollander, J. van Neerven and B. de Pagter, Royal Netherlands Academy, Amsterdam, pp. 145-170.
- [140] D.A. Dawson, L.G. Gorostiza and A. Wakolbinger (2001). Occupation time fluctuations in branching systems, *J. Theor. Probab.* 14, 729-796.

- [141] D.A. Dawson and S. Feng (2001). Large deviations for the Fleming-Viot process with neutral mutation and selection II, Stochastic Processes and their Applications 92, 2001, 131-162.
- [142] D.A. Dawson and K. Fleischmann (2002). Catalytic and mutually catalytic super-Brownian motions, in Proceedings of the 3rd Ascona Conference on Stochastic Analysis, Progress in Probability 52 (2002), 89-110, Birkhauser.
- [143] D.A. Dawson, L.G. Gorostiza and Z. Li (2002). Nonlocal branching superprocesses and some related models, Acta Appl. Math. 74, 93-112.
- [144] D.A. Dawson, A.M. Etheridge, K. Fleischmann, L. Mytnik, E.A. Perkins and J. Xiong (2002). Mutually catalytic branching in the plane: finite measure states, Annals of Probability, 30, 1681-1762.
- [145] D.A. Dawson, L.G. Gorostiza and A. Wakolbinger (2004), Hierarchical random walks, in *Asymptotic Methods in Stochastics*, Fields Institute Communications and Monograph Series 44, 173-194, Amer. Math. Soc.
- [146] D.A. Dawson, L.G. Gorostiza and A. Wakolbinger (2005), Degrees of transience and recurrence and hierarchical random walk, Potential Analysis, vol. 22, no.4, 305-350.
- [147] D.A. Dawson and S. Feng (2006). Asymptotic behavior of Poisson Dirichlet Distribution for large mutation rate, Annals Appl. Probab. 16, 562-582.
- [148] D.A. Dawson and Z. Li (2010). Stochastic equations, stochastic flows and measure-valued processes, preprint.
- [149] M. Delgado (1997). The Lagrange-Charpit method, SIAM Rev. 39, 298-304.
- [150] R.L. Dobrushin (1971) Markov processes with a large number of locally interacting components: existence of a limit process and its ergodicity. Problems Inform. Transmission 7, 149-164.
- [151] C. Dellacherie and P.A. Meyer (1978,1982). Probabilities and Potential A,B, North Holland.
- [152] C.F. Delwiche (1999). Tracing the web of plastid diversity through the tapestry of life. Am Nat 154: S164S177.
- [153] C.F. Delwiche (2000). Gene transfer between organisms. In: McGraw-Hill 2001 yearbook of science and technology. New York: McGraw- Hill. pp. 193197.
- [154] A. Dembo and O. Zeitouni (1993). Large Deviations and Applications, Jones and Bartlett Publishers, Boston.
- [155] E. Derbez and G. Slade (1997) Lattice trees and super-Brownian motion, Canad. Math. Bull 40(1997), 19-38.
- [156] U. Dieckmann and M. Doebeli (1999). On the origin of species by sympatric speciation. Nature 400, 354-357.
- [157] U. Dieckmann and R. Law (1996). The dynamical theory of coevolution. J. Math. Biol. 34, 579-612.
- [158] E. Domingo, ed. (2006). Quasispecies: Concept and Implications for Virology (Current Topics in Microbiology and Immunology), Springer
- [159] P. Donnelly (1984). The transient behavior of the Moran model in population genetics, Math. Proc. Camb. Phil. Soc. 95, 345-358.
- [160] P. Donnelly (1985). Dual processes and an invariance result for exchangeable models in population genetics, J. Math. Biology 23, 103-118.

- [161] P. Donnelly (1986). Partition structures, Polya urns, the Ewens sampling formula and the age of alleles, *Theor. Pop. Biol.* 30, 271-288.
- [162] P. Donnelly and S. Tavaré (1986). The age of alleles and a coalescent, *Adv. Appl. Prob.* 18, 1-19.
- [163] P. Donnelly and T.G. Kurtz (1996). A countable representation of the Fleming-Viot measure-valued branching diffusion, *Ann. Probab.* 24, 1-16.
- [164] P. Donnelly and T.G. Kurtz (1991). Genealogical processes for Fleming-Viot models with selection and recombination. *Ann. Appl. Probab.* 9, no. 4, 1091–1148.
- [165] P. Donnelly and T.G. Kurtz (1999). Particle representations for measure-valued population models, *Annals of Prob.* 27, 166-205.
- [166] R.A. Doney (1972). A limit theorem for a class of supercritical branching processes, *J. Appl. Probab.* 9, 707-724.
- [167] R.A. Doney (1976). On single and multi-type general age-dependent branching processes A limit theorem for a class of supercritical branching processes, *Adv. Appl. Probab.* 13, 239-246
- [168] M. Draief (2006). Epidemic processes on complex networks. The effect of topology on the spread of epidemics, *Physica A* 363, 120-131.
- [169] A. Dress (1984). Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: A note on combinatorial properties of metric spaces, *Adv. in Math.* 53 321- 402.
- [170] A.W.M. Dress and W.F. Terhalle (1996). The real tree, *Adv. Math.* 120, 283-301.
- [171] A. W. M. Dress, V. Moulton, and W. Terhalle, T-Theory (1996) An overview, *European Journal of Combinatorics* 17, 161-175.
- [172] M. Drmota (2009). Random trees, Springer.
- [173] R.T. Durrett, R.D.L. Iglehart and D.R. Miller (1977). Weak convergence to Brownian meander and Brownian excursion, *Ann. Probab.* 5, 117-129.
- [174] R. Durrett and D. Griffeath (1982). Contact processes in several dimensions, *ZFW* 59, 539-552.
- [175] R. Durrett and D. Griffeath (1983). Supercritical contact processes on \mathbb{Z} , *Ann. Prob.* 11, 1-15.
- [176] R. Durrett (1984). Oriented percolation in two dimensions, *Ann. Probab.* 12, 999-1040.
- [177] R. Durrett and R.H. Schonmann (1987). Stochastic growth models, in *Percolation Theory and the Ergodic Theory of Interacting Particle Systems*, ed. H. Kesten, IMA vol, Springer-Verlag.
- [178] R. Durrett (1988). *Lecture Notes on Particle Systems and Percolation*, Wadsworth and Brooks/Cole.
- [179] R. Durrett (1988). Crabgrass, measles and gypsy moths: an introduction to interacting particle systems, *Math. Intel.* 10, 37-47.
- [180] R. Durrett (1989). A new method for proving the existence of phase transitions, preprint.
- [181] R. Durrett and N.I. Tanaka (1989). Scaling inequalities for oriented percolation, *J. Stat. Phys.* 55, 981-995.
- [182] R. Durrett. Ten Lectures on Particle Systems, Ecole d'été de Saint Flour XXIII, Lecture Notes in Math 1608, 97-201.
- [183] R. Durrett and S. Levin (1994). The importance of being discrete (and spatial), *Theor. Pop. Biol.* 46, 363-394.

- [184] R. Durrett and C. Neuhauser (1994). Particle systems and reaction diffusion equations, Ann. Probab. 22, 289-333.
- [185] R. Durrett (1996) Probability: theory and examples, 2nd edition, Duxbury Press, Belmont CA.
- [186] R. Durrett and E.A. Perkins (1999). Rescaled contact processes converge to super-Brownian motion in two or more dimensions, Probab. Th. Relat. Fields 114, 309-399.
- [187] R. Durrett, J. Schweinsberg (2005). Rick Random partitions approximating the coalescence of lineages during a selective sweep. Ann. Appl. Probab. 15, 1591–1651.
- [188] R. Durrett (2008). Probability Models for DNA Sequence Evolution, Springer Series on Probability and its Applications 2nd ed.
- [189] T. Duquesne and J.-F. Le Gall (2002). Random trees, Lévy processes and spatial branching processes, Astérisque 281, Soc. Math. de France.
- [190] M. Durinx and J.A.J. Metz (2004), *Multi-type branching processes and adaptive dynamics of structured populations*, in Branching Processes in Biology: Variation, Growth and Extinction, eds. P. Haccou, P. Jagers and V. Vatutin, Cambridge Studies in Adaptive Dynamics 5, Cambridge UP
- [191] B.B. Dybiec, A. Kleczkowski and C.A. Gilligan (2009). Modelling control of epidemics spreading by long-range interactions, J. Royal Soc., to appear.
- [192] E.B. Dynkin (1991). Branching particle systems and superprocesses, Ann. Probab. 19:1157-1194.
- [193] E.B. Dynkin (1991). Path processes and historical superprocesses. Probab. Theory Relat. Fields 90, 1-36 (1991).
- [194] E.B. Dynkin (1993). Superprocesses and partial differential equations. Ann. Probab. 21, 1185–1262.
- [195] E.B. Dynkin (1994) An introduction to Branching Measure-Valued Processes, CRM Monograph Series, Vol. 6, Amer. Math. Soc., Providence.
- [196] E.B. Dynkin (2002). Diffusions, Superdiffusions and Partial Differential Equations, Amer. Math. Soc. Colloquium Publ 50.
- [197] M. Eigen (1971). Selforganization of matter and the evolution of biological macromolecules Naturwissenschaften 58, 465-523 (1971).
- [198] M. Eigen and P. Schuster (1979). The hypercycle: A principle of natural self-organization, Springer-Verlag.
- [199] M. Eigen, W. Gardiner, P. Schuster, and R. Winkler-Oswatsch (1981). The Origin of the Genetic Information. Sci.Am. 244(4) 88-118.
- [200] M. Eigen (1992). Steps towards life, Oxford University Press.
- [201] M. Eigen (1993) Viral quasispecies, Sci. Amer., July 42-49.
- [202] N. Eldredge and S.J. Gould (1972). Punctuated equilibria: an alternative to phyletic gradualism, in T.J.M. Schopf, ed., Models in Paleobiology. San Francisco: Freeman Cooper. pp. 82-115.
- [203] N. Eldredge and S.J. Gould (1977). Paleobiology 3, 115-151.
- [204] N. El Karoui and S. Roelly (1991). Propriétés de martingales, explosion et représentation de Lévy-Khinchine d'une classes de processus de branchement à valeurs mesures, Stoch. Proc. Appl. 38, 239-266.

- [205] A.M. Etheridge (2000). An introduction to Superprocesses, American Math. Soc.
- [206] A. Etheridge and P. March (1991). A note on superprocesses, *Probab. Th. Rel. Fields* 89, 141-147.
- [207] A.M. Etheridge (2004). Survival and extinction in a locally regulated population, *Ann. Applied Prob.* 14, 188-214.
- [208] A. Etheridge, P. Pfaffelhuber and A. Wakolbinger (2007). How often does the ratchet click? Heacts, heuristics, asymptotics, arXiv:0709.2775v1 [math.PR].
- [209] S.N. Ethier (1976). A class of degenerate diffusion processesoccurring in population genetics, *Comm. Pure Appl. Math.* 29, 483-493.
- [210] S.N. Ethier and M.F. Norman (1977). Error estimate for the diffusion approximation of the Wright-Fisher model, *Proc. N.A.S. USA* 74, 5096-5098.
- [211] S.N. Ethier and T. Kurtz (1981). The infinitely many neutral alleles diffusion model, *Adv. Appl. Prob.* 13, 429-452.
- [212] S.N. Ethier and T.G. Kurtz (1986). *Markov Processes: Characterization and Convergence*, Wiley, New York.
- [213] S.N. Ethier and R.C. Griffiths (1987). The infinitely many sites model as a measure-valued diffusion, *Ann. Probab.* 15, 515-545.
- [214] S.N. Ethier and T.G. Kurtz (1987). The infinitely many alleles model with selection as a measure-valued diffusion, *Springer Lecture Notes in Biomathematics* 70, 72-86.
- [215] S.N. Ethier (1990) The distribution of the frequencies of age-ordered alleles in a diffusion model, *Adv. in Appl. Probab.* 22, 519-532.
- [216] S.N. Ethier and T.G. Kurtz (1992) On the stationary distribution of the neutral diffusion model in population genetics, *Ann. Appl. Probab.* 2, 24-35.
- [217] S.N. Ethier and T.G. Kurtz (1993). Some measure-valued processes in population genetics. *SIAM J. Control Opt.* **31**, 345–386.
- [218] S.N. Ethier and T.G. Kurtz (1998). Coupling and ergodic theorems for Fleming-Viot processes, *Ann. Probab.* 26, 533-561.
- [219] S.N. Ethier and T. Shiga (2000). A Fleming-Viot process with unbounded selection, *J. Math. Kyoto Univ.* 40, 337-361.
- [220] S.N. Ethier and T. Shiga (2002). A FlemingViot process with unbounded selection, II. In *Markov Processes and Controlled Markov Chains*. Z. Hou, J. A. Filar, A. Chen, eds. Kluwer, 2002, pp. 305322.
- [221] L. Euler (1760). Recherches générales sur la mortalité: la multiplication du genre humain. *Mem. Acad. Sci, Berlin* 16 :144-64
- [222] S.N. Evans (1989) Local properties of Lévy processes on a totally disconnected group, *J. Theor. Probab.* 2, 209-259.
- [223] S.N. Evans and E.A. Perkins (1994). Measure-valued branchng diffusions with singular interactions, *Canad. J. Math.* 46, 120-168.
- [224] S.N. Evans and K. Fleischmann (1996). Cluster formation in a stepping-stone modek with continuous hierachically structured sites, *Ann. Probab.* 24, 1926-1952.

- [225] S.N. Evans and E.A. Perkins (1998). Collision local times, historical stochastic calculus and competing superprocesses, *Elect. J. Probab.* 3, paper 5.
- [226] S. N. Evans, D. Steinsaltz and Kenneth Wachter: A generalized model of mutation-selection balance with applications to aging. *Advances in Applied Mathematics*, 35: 1 (2005), pp. 1633.
- [227] S.N. Evans and D. Steinsaltz (2006). Damage segregation at fissioning may increase growth rates: a superprocess model, arXiv:q-bio.PE/0608008
- [228] S.N. Evans, J.W. Pitman and A. Winter (2006). Rayleigh processes, real trees and root growth with re-grafting, *Probab. Th. Rel. Fields*, 134, 81-126.
- [229] S.N. Evans (2007) Probability and Real Trees: École D'Été de Probabilités de Saint-Flour XXXV-2005 Springer LNM 1920.
- [230] W. J. Ewens, Mathematical Population Genetics, Springer-Verlag, 1979.
- [231] W. Feller (1939). Die Grundlagen der Volterraischen Theorie des Kampfes uns Dasein in wahrscheinlichkeitstheorie Behandlung, *Acta Biotheoretica* A. 5, 11-40.
- [232] W. Feller (1951). Diffusion Processes in genetics, Proc. 2nd Berkeley Symp. Math. Statist. Prob. Univ. of California Press, 227-246.
- [233] S. Feng, B. Schmuland, J. Vaillancourt and X. Zhou (2008). Reversibility of Interacting Fleming-Viot processes with mutation, selection and recombination, preprint.
- [234] S. Feng (2009). Poisson-Dirichlet Distribution and Related Topics, book to appear.
- [235] R. Ferriere, S. Méléard, N. Fournier and N. Champagnat (2004). The mathematics of Darwinian evolution: from stochastic individual processes to adaptive dynamics, prpublications MODAL'X.
- [236] R. A. Fisher. The Genetical Theory of Natural Selection. Clarendon Press, Oxford, 1930.
- [237] R.A. Fisher (1937) The wave of advance of advantageous genes, *Annals of Eugenics* 7, 353.
- [238] K. Fleischmann and J. Gärtner (1986). Occupation time process at a critical point, *Math. Nachr.* 125, 275-290.
- [239] K. Fleischmann and A. Greven (1994). Diffusive clustering in an infinite system of hierarchically interacting diffusions, *Probab. Th. Rel. Fields* 98, 517-566.
- [240] K. Fleischmann and A. Greven (1996). Time-space analysis of the cluster formation in interacting diffusions, *Elect. J. Probab.* 1.
- [241] W.H. Fleming (1979). Equilibrium distributions of continuous polygenic traits. *SIAM J. Appl. Math.* 36, 148-168.
- [242] W. Fleming and M. Viot (1979). Some measure-valued Markov processes in population genetics theory, *Indiana Univ. Math. J* 28, 817-843.
- [243] S.A. Frank (1997). The Price equation, Fisher's fundamental theorem, kin selection and causal analysis, *Evolution*, 51(6), 1997, pp. 1712-1729
- [244] M.I. Freidlin and A.D. Wentzell (1984). Random Perturbations of Dynamical Systems, Springer-Verlag, New York.
- [245] N. Fournier and S. Méléard (2004). A microscopic probabilistic description of a locally regulated population and macroscopic approximations, *An. Appl. Probab.* 14, 1880-1919.
- [246] M. Fukushima and D. Stroock (1986). Reversibility of solutions to martingale problems, *Adv. Math. Suppl. Stud.* 9, 107-123, Academic Press.

- [247] F. Galton and H.W. Watson (1874). On the probability of extinction of families. *J. Anthropol. Inst.* 4, 138-144.
- [248] P.R. Garabedian (1964). *Partial Differential Equations*, Wiley.
- [249] J. Gärtner (1988). On the McKean-Vlasov limit for interacting diffusions, *Math. Nachr.* 137, 197-248.
- [250] G.F. Gause (1934). *The struggle for existence*. Baltimore, MD: Williams and Wilkins.
- [251] S. Gavrilets (2004) *Fitness Landscapes and the Origin of Species*. Monographs in Population Biology, Princeton Univ. Press.
- [252] J. Geiger (2000). Poisson point process limits in size-biased Galton-Watson trees, *Elect. J. Probab.* 5, paper 17.
- [253] John H. Gillespie (1991). *The Causes of Molecular Evolution*, Oxford Univ. Press.
- [254] John H. Gillespie (2004) *Population Genetics*, 2nd ed., Johns Hopkins University Press, Baltimore.
- [255] J.H. Gillespie Genetic drift in an infinite population: the pseudohitchhiking model, *Genetics* 155: 090-919.
- [256] J.H. Gillespie (1999) The role of population size in molecular evolution, *Theor. Pop. Biol.* 55, 145-156.
- [257] I.V. Girsanov (1960). On transforming a certain class of stochastic processes by absolutely continuous substitution of measures, *The. Probab. and its Applications* 5, 314-330.
- [258] J. Peter Gogarten and J.P. Townsend (2005) Horizontal gene transfer, genome innovation and evolution, *Nature Reviews/Microbiology*, 3(2005), 679-687.
- [259] O. Görnerup and J.P. Crutchfield (2006). Objects that make objects: the population dynamics of structural complexity, *J.R. Soc. Interface* 3, 345-349.
- [260] O. Görnerup and J.P. Crutchfield (2008). Hierarchical self-organization in the finitary process soup, *Artif. Life*, 14, 245-254.
- [261] A. Greven (1987). Couplings of Markov Chains by randomized stopping times, *PTFR* 75, 195-212.
- [262] A. Greven, V. Limic and A. Winter (2005) Representation theorems for interacting Moran models, interacting Fisher-Wright diffusions and applications, *Elect. J. Probab.* 10, 1286-1358.
- [263] A. Greven (2006). Multiscale analysis of population models, 2005 les Houches Summer School on Mathematical Statistical Physics eds. A. Bovier, F. Dunlop, A. Van Enter, F. den Hollander, pp. 547-606, Elsevier.
- [264] A. Greven, P. Pfaffelhuber and A. Winter (2008), Convergence in distribution of random metric spaces, (Λ coalescent measure trees), *PTRF*, to appear.
- [265] A. Greven, P. Pfaffelhuber and A. Winter (2008). Tree-valued resampling dynamics: martingale problems and applications, arXiv:0806.2224v1.
- [266] D. Griffeath (1981). The basic contact process, *SPA* 11, 151-168.
- [267] R.C. Griffiths (1979). On the distribution of allele frequencies in a diffusion model. *Theoretical Population Biology*, 15, 140-158.
- [268] R.C. Griffiths (1982). The number of alleles and segregating sites in a neutral infinite-sites model with no recombination, *J. Appl. Probab.* 18, 42-51.

- [269] R.C. Griffiths and P, Marjoram (1996). Ancestral inference from samples of DNA sequences with recombination, *J. Comp. Biol.* 3, 479-502.
- [270] R.C. Griffiths and S. Tavar (1995). Unrooted genealogical tree probabilities in the infinitely-many-sites model. *Mathematical Biosciences* 127, 77-98.
- [271] R.C. Griffiths and P, Marjoram (1997). An ancestral recombination graph, in *Progress in Population Genetics and Human Evolution*, IMA Vol. in Math. and its Applications 87, 257-270.
- [272] M. Gromov (1999). Metric spaces for Riemannian and non-Riemannian spaces, *Progress in Math.* 152, Birkhäuser, Boston.
- [273] J. Haigh (1978). The accumulation of deleterious genes in a population - Muller's ratchet. *Theor. Pop. Biol.* 33, 677-702.
- [274] J.B.S. Haldane (1924). A mathematical theory of natural and artificial selection. *Proc. Cambridge Phil. Soc.* 1, 158-163.
- [275] J. Haldane (1930). Theoretical genetics of auto-polyploids. *J. Genetics* 22, 359-372.
- [276] J.B. S. Haldane (1932). *The Causes of Evolution*, Longmans Green and Co, London.
- [277] W.D. Hamilton (1964). The Genetical Evolution of Social Behaviour I and II, *J. Theor. Biol.* v7, pp 1-16, and 17-52.
- [278] T. Harris (1963). *The Theory of Branching Processes*, Grundlehren der math. Wissenschaften 119, Springer-Verlag.
- [279] T. Harris (1968). Counting measures, monotone random set functions, *Z. Wahr. Verw. Geb.* 10, 102-119.
- [280] D.Hartl and A.Clark, *Principles of Population Genetics*, Sinauer Associates, Inc., 1989.
- [281] H. He (2008). Rescaled Lotka-Volterra Models Converge to Super Stable Processes, preprint.
- [282] E. Hille (1948). Functional Analysis and Semi-groups, *Am. Math. Soc. Colloq. Publ.* 31, New York.
- [283] R. Holley and T.M. Liggett (1975). Ergodic theorems for weakly interacting infinite systems and the voter model, *Ann. Probab.* 3, 643-663.
- [284] P. Haccou, P. Jagers and V.A. Vatutin (2005), *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge. U. Press, Cambridge.
- [285] J.B.S. Haldane (1932). *The Causes of Evolution*, Harper and Brothers.
- [286] K. Handa (1990). A measure-valued diffusion process describing the stepping stone model with infinitely many alleles, *Stoch. Proc. Appl.* 36. 269-296.
- [287] I. Hanski (1999). *Metapopulation Ecology*, Oxford Univ. Press.
- [288] T. Hara and G. Slade (2000) The scaling limit of the incipient infinite cluster in high-dimensional percolation II, Integrated super-Brownian excursion, *J. Math. Phys* 41(2000), 1244-1293.
- [289] H. He (2008) Rescaled Lotka-Volterra Models Converge to Super Stable Processes, preprint.
- [290] M. Hirsch (1982), (1985), (1988). Systems of Differential equations which are competitive or cooperative I,II,III, *SIAM J. Math. Anal.* 13, 167-179; 16, 423-439, *Nonlinearity* 1, 51-71.
- [291] J. Hofbauer (1985). The selection mutation equation, *SIAM J. Appl. Math* 41, 1-7.

- [292] J. Hofbauer and K. Sigmund (1988) The theory of evolution and dynamical systems, Cambridge University Press.
- [293] J. Hofbauer and K. Sigmund (1998). Evolutionary Games and Population Dynamics, Cambridge Univ. Press.
- [294] B. Hlldobler, E.O. Wilson (2008). The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies. W.W. Norton and Co.
- [295] J. Holland (1975). Adaptation in natural and artificial systems, Univ. of Michigan Press.,
- [296] J.H. Holland (1995). Hidden Order, Helix Books, Addison-Wesley.
- [297] F.M. Hoppe (1987). The sampling theory of neutral alleles and an urn model in population genetics, *J. Math. Biol.* 25(1987), 123-159.
- [298] F.M. Hoppe (1984). Plya-like urns and the Ewens' sampling formula. *J. Math. Biol.* 20 (1984), no. 1, 91-94.
- [299] R. van der Hofstad (2006). Infinite canonical super-Brownian motion and scaling limits, *Comm. Math. Phys.* 265, 547-583.
- [300] R. van der Hofstad and G. Slade (2003) Convergence of critical oriented percolation to super-Brownian motion above 4+1 dimensions, *Ann. Inst. H. Poincare Probab. Statist.*, 39, 413-485.
- [301] R. van der Hofstad and A. Sakai (2008). Convergence of the critical finite range contact process to super-Brownian motion above the upper critical dimension: I. The higher point functions, arXiv:0809.1712v.
- [302] U. Horst (1999) Fluctuations in a stock market model with interacting agents - the mean-field case, discussion paper 106, SFB 373, Humboldt Univ. Berlin
- [303] S.P. Hubbell (2001). The Unified Neutral Theory of Biodiversity and Biogeography, Princeton Univ. Press.
- [304] L.D. Hurst (2009) Genetics and the understanding of selection, *Nature Reviews. Genetics*, 10,83-93.
- [305] G.E. Hutchinson(1957). Concluding remarks. Cold Spring Harbour Symposium on Quantitative Biology 22, 415-427.
- [306] M. Hutzenhaller and A. Wakolbinger. Ergodic behaviour of locally regulated branching populations, *Annals of Applied Probability*. Volume 17, Number 2 (2007), 474-501
- [307] M. Iizuka, H. Tachida and H. Matsuda (2002). A neutral model with fluctuating population size and its effective size, *Genetics* 161, 381-388.
- [308] N. Ikeda, M. Nagasawa and S. Watanabe (1968), (1969). Branching Markov processes I,II,III, *J. Math. Kyoto Univ.* 8, 233-278; 9 95-160.
- [309] N. Ikeda and S. Watanabe (1981). Stochastic Differential Equations and Diffusion Processes, North Holland.
- [310] I. Iscoe (1986). A weighted occupation time for a class of measure-valued critical branching Brownian motion, *Probab. Theor. Rel. Fields* 71, 85-116.
- [311] I. Iscoe (1986b). Ergodic theory and a local occupation time for a class of measure-valued critical branching Brownian motion, *Stochastics* 18, 197-143.

- [312] I. Iscoe (1988) On the supports of measure-valued critical branching Brownian motion, Ann. Probab. 16(1), 200-221.
- [313] K. Itô (1961). On stochastic differential equations, Mem. Amer. Math. Soc. No. 4.
- [314] K. Ito (1961) Lectures on Stochastic Processes, Tata Institute.
- [315] K. Itô and H.P. McKean, Jr. Diffusion processes and their sample paths, 1965.
- [316] K. Itô (1970) Poisson point process attached to Markov processes, Proc. Sixth Berkeley Symp. Math. Stat. Prob. vol. 3, Univ. of Californian, Berkeley, 225-239.
- [317] P. Jagers (1969). A general stochastic model for population development. Skand. Aktuarietidskr. 52, 84-103.
- [318] P. Jagers (1974). Random measures and point processes, in Advances in probability and related topics, eds. P. Ney and S. Port, Marcel Dekker.
- [319] P. Jagers (1975). Branching processes with biological applications, Wiley.
- [320] P. Jagers (1975). Branching Processes with Biological Applications. John Wiley and Sons, London etc.
- [321] P. Jagers (1981). How probable is it to be firstborn? Math. Biosci. 59, 1-15.
- [322] P. Jagers and O. Nerman (1984). The growth and composition of populations, Adv Appl. Prob. 16, 221-259.
- [323] P. Jagers (1989). The Markov structure of population growth, Acta Appl. Math. 14, 103-114.,
- [324] P. Jagers and O. Nerman (1996), The asymptotic composition of supercritical, multi-type branching populations. Sminaire de probabilités de Strasbourg, 30, Springer Lecture Notes in Mathematics, 1626, 40-54.
- [325] P. Jagers, F. Klebaner and S. Sagitov (2007), On the path to extinction. Proc. Nat. Acad. Sci. 104, 6107-6111.
- [326] P. Jagers and S. Sagitov(2008). General branching processes in discrete time as random trees. Bernoulli 14:4, 949-962
- [327] R. Jain, M. Rivera and J. Lake (1999). Horizontal gene transfer among genomes: the complexity hypothesis, Proc. Nat. Acad. Sci. USA 96, 3801-3806.
- [328] S. Jain and S. Krishna (2002) Large extinctions in an evolutionary model: The Role of Innovation and Keystone Species, Proc. NAS USA 99, 2055-2060.
- [329] R. Jain, M. Rivera, J. Moore and J. Lake (2002). Horizontal gene transfer in microbial genome evolution, Theor. Pop. Biol. 61, 489-495.
- [330] H.K. Janssen, K. Oerding, F. van Wijland and H.J. Hilhorst (1999). Levy-flight spreading of epidemic processes leading to percolating clusters, Eur. Phys. J. B 7 137.
- [331] A. Joffe (1978) Remarks on the structure of trees with applications to supercritical Galton-Watson processes, in *Advances in Probability and related topics* 5 (Branching processes), ed. A. Joffe and P. Ney.
- [332] A. Joffe and W.A. O'N. Waugh (1982). Exact distributions of kin numbers in a Galton-Watson process, J. Appl. Prob. 19, 767-775.
- [333] A. Joffe and W.A. O'N. Waugh (1982). The kin number problem in a multitype Galton-Watson population, J. Appl. Prob. 22, 37-47.

- [334] A. Joffe and M. Métivier (1986). Weak convergence of sequences of semi-martingales with application to multitype branching processes. *Adv. Appl. Probab.* 18, 20-65.
- [335] F. John (1982). Partial Differential Equations, 4th edition. Fritz John. Springer.
- [336] M. Jirina (1957). Stochastic branching processes with continuous state space, *Czechosl. Math. J.* 8, 292-313.
- [337] M. Jirina (1964). Branching processes with measure-valued states, In *Trans. Third Prague Conf. on Inf. Th.* 333-357.
- [338] P. Joyce, S.M. Krone and T.G. Kurtz (2002). Gaussian limits associated with the Poisson-Dirichlet distribution and the Ewens sampling formula. *Ann. Appl. Probab.*, **12**, No. 1, 101-124.
- [339] P. Joyce, S.M. Krone and T.G. Kurtz (2003). When can one detect overdominant selection in the infinite-alleles model? *Ann. Appl. Probab.*, **13**, No. 1, 181-212.
- [340] I. Kaj and S.M. Krone (2003) The Coalescent Process in a Population with Stochastically Varying Size, *J. of Appl. Probab.* Vol. 40, 33-48
- [341] O. Kallenberg (1977) Stability of critical cluster fields, *Math. Nachr.* 77, 7-43.
- [342] O. Kallenberg (1983). Random Measures, 3rd ed., Akademie Verlag and Academic Press, Berlin.
- [343] S. Karlin and H.M. Taylor (1981). A second course in stochastic processes, Academic Press.
- [344] S.A. Kauffman and S.A. Levin (1987). Towards a general theory of adaptive walks on rugged landscapes. *J. Theoret. Biol.* 128 1145
- [345] S. Kauffman (1993) The Origins of Order: Self-Organization and Selection in Evolution, Oxford University Press.
- [346] S.A. Kauffman (1995). At home in the universe: The Search for the Laws of Self-Organization and Complexity, Oxford Univ. Press.
- [347] N. Keyfitz (1968). Introduction to the mathematics of population, Addison-Wesley, Reading, Mass.
- [348] S. Kauffman (2000). Investigations, Oxford University Press.
- [349] L. Keller (1999) Levels of selection in evolution, Princeton University Press.
- [350] D.G. Kendall (1966). Branching processes since 1873. *J. London Math. Soc* 41, 385-486.
- [351] W. Kermack and A. McKendrick (1927). A contribution to the mathematical theory of epidemics. *Proc. R. Soc. London A* 115, 700-721. *Proc. R. Soc. London A* 115, 700-721.
- [352] H. Kesten and B.P. Stigum (1966). A limit theorem for multidimensional Galton-Watson processes, *Ann. Math. Statist.* 37, 1211-1223.
- [353] J.F.C. Kingman (1977) The population structure associated with the Ewens sampling formula, *Theor. Pop. Biol.* 11, 274-283.
- [354] J.F.C. Kingman (1980) The mathematics of Genetic Diversity, CBMS Regional Conf Series in Appl. Math, 34, SIAM.
- [355] J.F.C. Kingman (1982) The coalescent, *Stoch. Proc. Appl.* 13, 235-248.
- [356] J.F.C. Kingman (1982) The genealogy of large populations, *J. Appl. Probab.* 19A, 27-43
- [357] J.F.C. Kingman (1993). Poisson Processes, Clarendon Press.

- [358] J.F.C. Kingman (2000). Origins of the Coalescent: 19741982, Genetics, Vol. 156, 1461-1463.
- [359] M. Kimura (1983) The Neutral Theory of Molecular Evolution, Cambridge Univ. Press.
- [360] M. Kimura (1953) Stepping stone model of population. Ann. Rep. Nat Inst Genetics, Japan 3: 62-63.
- [361] M. Kimura (1964) Diffusion models in population genetics, J. Appl. Prob. 1, 177-232.
- [362] M. Kimura and J.F. Crow (1964). The number of alleles that can be maintained in a finite population, Genetics 49, 725-738.
- [363] M. Kimura and G.H. Weiss (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance, Genetics 49, 561-576.
- [364] M. Kimura (1969). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations, Genetics 61, 893-903.
- [365] M. Kimura (1971). Theoretical foundations of population genetics at the molecular level, Theor. Pop. Biol. 2, 174-208.
- [366] M. Kimura (1983). Diffusion model of intergroup selection, with special reference to evolution of an altruistic gene, Proc. Nat. Acad. Sci. 80, 6317-6321.
- [367] J.F.C. Kingman (1976). Coherent random walks arising in some genetical models. Proc. R. Soc. Lond. Ser. A 351: 1931.
- [368] J.F.C. Kingman (1982) The coalescent, Stoch. Proc. Appl. 13, 235-248.
- [369] J.F.C. Kingman (1982) The genealogy of large populations, J. Appl. Probab. 19A, 27-43
- [370] J.F.C. Kingman (1993). Poisson Processes, Clarendon Press.
- [371] C. Kipnis and C.M. Newman (1985), *The metastable behavior of infrequently observed weakly random one dimensional diffusion processes*, SIAM J. Appl. Math. 45, 972-982.
- [372] M. Kirkpatrick, T. Johnson and N. Barton (2002). General models of multilocus evolution, Genetics 161, 1727-1750.
- [373] V.M. Kirzhner, A.B. Korol and E. Nevo (1996). Complex dynamics of multilocus systems subject to cyclical selection, Proc. N.A.S. USA 93, 6532-6535.
- [374] A. Klenke (1996) Different clustering regimes in systems of hierarchically interacting diffusions, Ann. Prob. 24(2), 660-697.
- [375] A. Klenke (1997) Multiple scale analysis of clusters in spatial branching models, Ann. Probab. 25(4), 1670-1711.
- [376] A. Klenke (1998) Clustering and invariant measures for spatial branching models with infinite variance, Ann. Probab. 26(3), 1057-1087.
- [377] A. Klenke (2000) Diffusive clustering of interacting Brownian motions on \mathbb{Z}^2 , Stoch. Proc. Appl.
- [378] A. Klenke (2000). Longtime Behavior of Stochastic Processes with Complex Interactions, Habilitation Thesis, University Erlangen.
- [379] A. Kolmogorov, I. Petrovsky and N. Piscounov (1937). Etude de l'équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique, Moscow Univ. Bulletin of Mathematics 1,1.

- [380] A.N. Kolmogorov and N.A. Dmitriev (1947). Branching stochastic processes, Doklady Akad. Nauk SSSR 56, 7-10.
- [381] N. Konno and T. Shiga (1988). Stochastic partial differential equations for some measure-valued diffusions, Probab. Theory Rel Fields 79, 201-225.
- [382] S.M. Krone and C. Neuhauser (1997) Ancestral processes with selection, Theoret. Pop. Biol. 51 (1997), 210-237.
- [383] A.N. Kolmogorov (1931). Über die analytischen methoden in der wahrscheinlichkeitsrechnung, Math. Ann. 104, 415-458.
- [384] S. Krone and C. Neuhauser (1997b). The genealogy of samples in models with selection, Genetics 145, 519-534.
- [385] S. Krone and C. Neuhauser (1997), *Ancestral processes with selection*, Theor. Pop. Biol. 51, 210-237.
- [386] S. Krone and M. Nordborg (2002), *Separation of Time Scales and Convergence to the Coalescent in Structured Populations*. In Modern Developments in Theoretical Population Genetics, pp. 194-232, M. Slatkin and M. Veuille, editors. Oxford University Press.
- [387] T.G. Kurtz and P. Protter (1996). Weak convergence of stochastic integrals and differential equations. Infinite dimensional case. In *Probabilistic Models for Nonlinear Partial Differential Equations*, Springer Lecture Notes in Math 1627, 197-285.
- [388] T.G. Kurtz, R. Lyons, R. Pemantle and Y. Peres (1997). A conceptual proof of the Kesten-Stigum theorem for multi-type branching processes, in *Classical and Modern Branaching Processes*, K. Athreya and P. Jagers, eds., Springer.
- [389] T.G. Kurtz and E. Rodrigues (2008). Poisson representations of branching Markov and measure-valued branching processes, preprint.
- [390] S.P. Lalley (2008). Spatial epidemics: critical behavior in one dimension, Probab. Theory Relat. Fields.
- [391] S.P. Lalley and X. Zheng (2009). Spatial epidemics and local times for critical branching random walks in dimensions 2 and 3, arXiv:0901.0246v1
- [392] S.P. Lalley, E.A. Perkins and X. Zheng (2009). A phase transition for spatial SIR epidemics, preprint.
- [393] N. Lanchier and C. Neuhauser (2007). Voter model and biased voter model in heterogeneous environments, J. Appl. Probab. 44, 770-787.
- [394] M. Lässig, F. Tria and L. Peliti (2003). Evolutionary games and quasispecies, Europhysics Letters preprint arXiv:cond-mat/0209086v2
- [395] J.F. Le Gall (1991) Brownian excursions, trees and measure-valued branching processes, Ann. Probab. 19, 1399-1439.
- [396] J.F. Le Gall and E.A. Perkins (1995) The Hausdorff measure of the support of two-dimensional super-Brownian motion, Ann. Probab. 23, 1719-1747.
- [397] J.-F. Le Gall (1999) Spatial Branching Processes, Random Snakes and Partial Differential Equations, Birkhauser.
- [398] J.-F. Le Gall (2006). Random real trees, Ann. Fac. Sci. Toulouse, 15, 35-62.
- [399] S.A. Levin (1999). Fragile Dominion, Princeton Univ. Press.

- [400] R. Levins (1968). Evolution in a changing environment, Princeton Univ. Press.
- [401] P. Lévy (1948). Processus Stochastiques et Mouvement Browien, Gauthier-Villars, Paris.
- [402] R.C. Lewontin (1974). The Genetic Basis of Evolutionary Change, Princeton Univ. Press.
- [403] W.H. Li and D. Graur (1991). Fundamentals of molecular evolution, Sinauer.
- [404] Z. Li and T. Shiga (1995) Measure-valued branching diffusions: immigrations, excursions and limit theorems, J. Math. Kyoto Univ.
- [405] Z. Li, T. Shiga, and L. Yao (1999). A reversibility problem for Fleming-Viot processes. Electronic Communications in Probability 4, 6576.
- [406] Z. Li (2006). *Branching processes with immigration and related topics*, Frontiers of Mathematics in China 1, 73-97.
- [407] Z. Li and C. Ma (2008). Catalytic discrete state branching models and related limit theorems, J. Theor. Probab. 21, 936-965.
- [408] Z. Li (2009). Measure-valued Markov branching processes, book to appear.
- [409] T.M. Liggett (1985). Interacting Particle Systems, Springer-Verlag.
- [410] T.M. Liggett (1995). Survival of Discrete Time Growth Models, with Applications to Oriented Percolation, Ann. Appl. Probab. Volume 5, Number 3 (1995), 613-636.
- [411] E. Lloyd (2005). Units and Levels of Selection, in *The Stanford Encyclopedia of Philosophy* (Fall 2005 Edition), Edward N. Zalta (ed.),
- [412] A.J. Lotka (1925). Elements of Physical Biology, Williams and Wilkins, Baltimore.
- [413] Y. Lyubich and V. Kirzhner (2003). Mathematical framework for phenotypical selection and epistasis, J. Theor. Biol. 221, 625-638
- [414] D. Ludwig (1974). Stochastic Population Theories. Lecture Notes in Biomathematics 3, Springer-Verlag, Berlin.
- [415] T. Lux (1995). Herd behaviour, bubbles and crashes, The Economic Journal 105, 881-896.
- [416] R. Lyons, R. Pemantle and Y. Peres (1995). Conceptual proofs of $L \log L$ criteria for mean behaviour of branching processes, Ann. Probab. 3, 1125-1138.
- [417] R. MacArthur (1955). Fluctuations of animal populations and a measure of community stability, Ecology 36, 533-536.
- [418] R.H. MacArthur and E.O. Wilson (1967). Theory of island biogeography, Princeton Univ. Press.
- [419] R.H. MacArthur (1972). Geographical Ecology: Patterns in the Distribution of Species. Harper and Row, N.Y.
- [420] G. Malécot (1941). Étude mathématique des populations mendéliennes, Ann. Univ. Lyon Sci. Ser. A 4, 45-60.
- [421] G. Malécot (1948). Les mathématiques de l'hérédité, Masson, Paris.
- [422] G. Malécot (1949). Les processus de la génétique, Coll. Inst. Cent. Nat. Rech. Sci 13, 121-126.
- [423] J.-F. Marckert and A. Mokkadem (2003). The depth first processes of Galton-Watson trees converge to the same Brownian excursion, Ann. Probab. 31, 1655-1678.

- [424] R. Margalef (1968). Perspectives in Ecological Theory, Univ. of Chicago Press.
- [425] L. Margulis (1981). Symbiosis in cell evolution - Freeman San Francisco.
- [426] L. Margulis and D. Sagan (2002). Acquiring Genomes. A theory of the origins of species, Basic Books, 2002.
- [427] A.A. Markov (1906). "Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga". Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete, 2-ya seriya, tom 15, pp 135-156.
- [428] A. Martin-Löf (1998). The final size of a nearly critical epidemic, and the first passage time of a Wiener process to a parabolic barrier, J. Appl. Prob. 35, 671-682.
- [429] R.M. May 1973. Stability and Complexity in Model Ecosystems. Princeton University Press
- [430] J. Maynard Smith (1982). Evolution and the theory of games, Cambridge University Press.
- [431] J. Maynard Smith and E. Szathmary (1995). The Major Transitions in Evolution, Oxford U. Press.
- [432] A.J.McKane and T.J. Newman (2004). Stochastic models in population biology and their deterministic analogs, Phys. Rev. E 70, 041902.
- [433] H.P. McKean (1969). Propagation of chaos for a class of nonlinear parabolic equations, Lecture Series in Differential Equations 7, Catholic University, 41-47, Van Nostrand.
- [434] H.P. McKean, Application of Brownian motion to the equation of KPP, Comm. Pure Appl. Math. 28 (1975), pp 323-331.
- [435] R. Mehrotra, V. Soni and S. Jain (2007), Diversity begets stability in an evolving network, arXiv.0705.1075v1.
- [436] J. G. Mendel (1965). Versuche ber Pflanzenhybriden Verhandlungen des naturforschenden Vereines in Brnn, Bd. IV fr das Jahr, 1865, 3-47.
- [437] J.A.J. Metz, S.A.H. Geritz, G. Meszna, F.J.A. Jacobs and J.S. van Heerwaarden (1996). Adaptive dynamics, a geometrical study of nearly faithful reproduction. In: Stochastic and Spatial Structures of Dynamical Systems. North-Holland, Amsterdam.
- [438] J.A.J. Metz, R.M. Nisbet, S.A.H. Geritz (1992). How should we define fitness for general ecological scenarios. Trends in Ecology and Evolution 7: 198-202.
- [439] L.A. Meyers (2007). Contact network epidemiology: bond percolation applied to infectious disease prediction and control, Bull. AMS 44, 63-86.
- [440] P. A. Meyer (1966). Probability and potentials, Blaisdell, New York.
- [441] C.J. Mode (1971). Multitype Branching Processes. Elsevier, New York. Nerman, O. and Jagers, P. (1984). The stable doubly infinite. pedigree process of supercritical branching populations. Z. Wahrsch. Verw. Gebiete 65, 445-460.
- [442] D. Mollison, The rate of spatial propagation of simple epidemics, In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, vol. 3, pp 579-614.
- [443] D. Mollison (1977). Spatial contact models for ecological and epidemic spread. J. Roy. Statist. Soc. B 39: 283-326
- [444] J.E. Moyal The general theory of stochastic population processes (1962), Acta Math., 108, 131.

- [445] P.A.P. Moran (1975). Wandering distributions and the electrophoretic profile. *Theor. Popul. Biol.* 8: 318330.
- [446] P.A.P. Moran (1958). Random processes in genetics, *Proc. Camb. Phil. Soc.* 54, 60-72.
- [447] C. Mueller and R. Tribe (1994). A phase transition for a stochastic PDE related to the contact process, *PTRF* 100, 131-156.
- [448] C. Mueller and R. Tribe (1995). Stochastic p.d.e.'s arising from long range contact and long range voter processes, *Probab. Th. Rel. Fields* 102, 519-545.
- [449] H. Mühlenbein and D. Schlierkamp-Voosen (1995). Analysis of selection, mutation and recombination in genetic algorithms, *Lecture Notes in Computer Science*, vol. 899.
- [450] T. Nagylaki (1974) The decay of genetic variability in geographically structured populations, *Proc. Nat. Acad. Sci USA* 71, 2932-2936.
- [451] T. Nagylaki (1976). The geographic structure of populations, *MAA Studies in Mathematical Biology*.
- [452] M. Nei (1987). *Molecular Evolutionary Genetics*, Columbia Univ. Press.
- [453] O. Nerman (1981). On the convergence of supercritical general (C-M-J) branching processes, *Z. Wahr. verw. Geb.* 57, 365-395.
- [454] C. Neuhauser and S. W. Pacala (1999). An explicitly spatial version of the Lotka-Volterra model with interspecific competition, *Ann. Appl. Probab.* 9, 12261259
- [455] C.M. Newman, J.E. Cohen and C. Kipnis (1985). Neo-darwinian evolution implies punctuated equilibria, *Letters to Nature*, vol 315, (30 May 1985), 400-401.
- [456] J. Neveu and J.W. Pitman (1989) The branching process in a Brownian excursion, *Séminaire de Probabilités XXIII*, LNM 1372, 248-257, Springer.
- [457] M. E. J. Newman (2002). Spread of epidemic disease on networks. *Physical Review E* 66, 016128
- [458] J. Neveu (1986). Arbres et processus de Galton-Watson, *Annales de l'Inst. H. Poincaré* 22.
- [459] E. van Nimwegen, J.P. Crutchfield and M. Mitchell (1997). Finite populations induce metastability in evolutionary search, *Physics Letters A*.
- [460] M. Notohara (1990). The coalescent and the genealogical process in geographically structured population, *J. Math. Biol.* 29, 59-75.
- [461] M. Notohara and T. Shiga (1980) Convergence to genetically uniform state in stepping stone models of populations genetics, *J. Math. Biology* 10, 281-294.
- [462] M.A. Nowak (2006). *Evolutionary Dynamics*, Belknap Press, Harvard Univ.
- [463] Y. Ogura and N. Shimakura (1987). Stationary solutions and their stability for Kimura's diffusion model with intergroup selection, *J. Math Kyoto Univ.* 27, 305-347, 635-655.
- [464] T. Ohta and J.H. Gillespie (1996). Development of neutral and nearly neutral theories, *Theor. Pop. Biol.* 49, 128-142.
- [465] T. Ohta and M. Kimura (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population, *Genet. Res. Cambridge* 22, 201-204.
- [466] S. Okasha (2003). Recent work on the levels of selection problem, *Human Nature Review* 3, 349-356.

- [467] S. Okasha (2006). Evolution and the Levels of Selection, Oxford University Press.
- [468] P. Olofsson (1996) The $x \log x$ condition for general branching processes, *J. Appl. Prob.* 33, 940-948.
- [469] P. Olofsson (2008) Size-biased branching population measures and the multi-type $x \log x$ condition, preprint.
- [470] S. Ott and T. Lenormand (2002). Resolving the paradox of sex and recombination, *Nature Reviews - Genetocs*, 3, 252-261.
- [471] L. Overbeck, M. Röckner and B. Schmuland (1995). An analytic approach to Fleming-Viot processes, *Ann. Probab* 23, 1-36.
- [472] J. Paulsson (2002). Multilevel selection on plasmid replication, *Genetics* 161, 1373-1384.
- [473] A. Pazy (1983). Semigroups of Linear Operators and Applications to Partial Differential Equations, Springer-Verlag, Berlin.
- [474] A.S. Perelson and C. A. Macken (1995). Protein evolution on partially correlated landscapes. *Proc. Natl. Acad. Sci. USA* 92:96579661
- [475] E.A. Perkins (1988). A space-time property of a class of measure-valued branching diffusions, *Trans. Amer. Math. Soc.* 305, 743–795.
- [476] E.A. Perkins (1989) The Hausdorff measure of the closed support of super-Brownian motion, *Ann. Inst. H. Poincaré Probab. Stat.* 25, 205-224.
- [477] E.A. Perkins (1990). Polar sets and multiple points for super-Brownian motion. *Ann. Probab.* 18, 453-491.
- [478] E.A. Perkins (1991). Conditional Dawson-Watanabe processes and Fleming-Viot processes, in Seminar on Stoch. Processes 142-155, Birkhäuser, Boston.
- [479] E.A. Perkins (1992). Measure-valued branching with spatial interactions, *Probab. Th. Related Fields* 94, 189-245.
- [480] E.A. Perkins (1995). On the martingale problem for interactive measure-valued branching diffusions, *Mem. Amer. Math. Soc.* 115, 1-89.
- [481] E.A. Perkins (2002) Dawson-Watanabe Superprocesses and Measure-valued Diffusions, in École d'été de Saint Flour XXIX-1999, LNM 1781, Springer.
- [482] M. Perman, J. Pitman and M. Yor (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Relat. Fields* 92, 21-39.
- [483] M. Pigliucci and J. Kaplan (2006). Making sense of evolution: The conceptual foundations of evolutionary biology, University of Chicago Press.
- [484] P. Pfaffelhuber and A. Wakolbinger (2006). The process of most recent common ancestor in an evolving coalescent, *Stoch. Proc. Appl.* 116, 1836-1859.
- [485] S.L. Pimm (1984). The complexity and stability of ecosystems, *Nature*, vol. 307 no. 321-326.
- [486] J. Pitman (1999). Coalescents with multiple collisions, *Ann. Probab.* 27, 1870-1902.
- [487] J. Pitman (2002). Combinatorial Stochastic Processes, Saint Flour 2002.
- [488] P. Protter (1990). Stochastic Integration and Differential Equations, Springer.

- [489] D.A. Rand, H. B. Wilson, and J. M. McGlade (1994). Dynamics and evolution: evolutionary stable attractors, invasion exponents and phenotype dynamics. *Phil. Trans. Roy. Soc. Lond. B* 343:261-283.
- [490] M.L. Reed (1997). Algebraic structure of genetic inheritance, *Bull. A.M.S.* 34, 107-130.
- [491] H.K. Reeve and B. Hölldobler (2007) The emergence of a superorganism through intergroup competition, *Proc. N.A.S.* 104, 9736-9740.
- [492] M. Reimers (1989). One-dimensional stochastic partial differential equations and the branching measure diffusion. *Prob. Th. Rel. Fields* 81, 319-340.
- [493] D. Revuz and M. Yor (1991) Continuous martingales and Brownian motion, Springer.
- [494] L.C.G. Rogers and D. Williams (1987) Diffusions, Markov processes and martingales, vol. 2, Wiley.
- [495] T. Röhl, A. Traulsen, J.C. Claussen and H.G. Schuster (2008). Stochastic gain in finite populations, *Phys. Rev. E* 78, 026108-1 - 02618-8.
- [496] I.M. Rouzine, A. Rodrigo and J.M. Coffin (2001) Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology, *Microbiology and Molecular Biology Reviews*, 65, 151-185.
- [497] D. Roze and N.H. Barton (2006). The Hill-Robertson effect and the evolution of recombination, *Genetics* 173, 1793-1811.
- [498] S. Sagitov (1999) The general coalescent with asynchronous mergers of ancestral lines, *J. Appl. Probab.* 26, 1116-1125.
- [499] K.-I. Sato (1999). Lévy Processes and Infinitely Divisible Distributions, Cambridge Univ. Press.
- [500] S. Sawyer (1975) An application of branching random fields to genetics, In *Probabilistic methods in Differential Equations*, Springer Lecture Notes in Math. 451, 100-112.
- [501] S. Sawyer (1976). Results for stepping stone models for migration in population genetics, *Adv. Appl. Prob.* 4, 538-562.
- [502] S. Sawyer (1976) Branching diffusion processes in population genetics, *Adv. Appl. Probab.* 8, 659-689.
- [503] S. Sawyer and J. Felsenstein (1983) Isolation by distance in a hierarchically clustered population, *J. Appl. Probab.* 20, 1-10..
- [504] B. Schmuland (1991). A result on the infinitely many neutral alleles diffusion model, *J. App. Prob.* 28, 253-267.
- [505] P. Schuster and P.F. Stadler (2003). Networks in molecular evolution, *Complexity* Vol. 8, 34-42.
- [506] E. Senata (1970). On the supercritical branching process with immigration, *Math. Biosci.* 7, 9-14.
- [507] T. Shiga (1981) Diffusion processes in population genetics, *J. Math. Kyoto Univ.* 21, 133-151.
- [508] T. Shiga (1988). Stepping stone models in population genetics and population dynamics, In *Stochastic Processes in Physics and Engineering* (S. Albeverio, P. Blanchard, M. Hazewinkel and L. Streit, eds.) 345-355, Reidel, Dordrecht.

- [509] S. Shashahani (1979). A new mathematical framework for the study of linkage and selection, Memoirs of the American Mathematical Society 211.
- bibitemS-80 T. Shiga (1980) An interacting system in population genetics I and II, J. Math. Kyoto Univ. 20, 213-242, 723-733.
- [510] T. Shiga and A. Shimizu (1980). Infinite-dimensional stochastic differential equations and their applications, J. Math. Kyoto Univ. 20(3), 395-416.
- [511] T. Shiga (1981). Diffusion processes in population genetics, J. Math. Kyoto Univ. **21**:133–151.
- [512] N. Shimakura (1985). Existence and uniqueness for a diffusion model of intergroup selection, J. Math. Kyoto Univ. 25, 775-788.
- [513] T. Shiga and K. Uchiyama (1986) Stationary states and their stability of the stepping stone model involving mutation and selection, Probab. Th. Rel. Fields 73, 87-117.
- [514] T. Shiga J. Math. Kyoto Univ. (1987). Existence and uniqueness of solutions for a class of non-linear diffusion equations, J. Math. Kyoto Univ. 27, 195-215.
- [515] T. Shiga (1992). Ergodic theorems and exponential decay of sample paths for certain interacting diffusion systems, Osaka J. Math. 29, 789-807.
- [516] T. Shiga (1994) Stationary distribution problem for interacting diffusion systems, CRM Proc.
- [517] N. Shimakura (1988). A nonlinear parabolic mixed problem arising in population genetics, Journées Équations aux dérivées partielles (1988), 1-20.
- [518] M.L. Silverstein (1969). Continuous state branching semigroups, Z. Wahr. verw. Geb. 14, 96-112.
- [519] A.B. Simonson, J.A. Servin, R.G. Skophamer, C.W. Herbold, M.C. Rivera and J.A. Lake (2007). Decoding the genomic tree of life, Proc. N.A.S. USA 102, 6608-6613.
- [520] Ya. Sinai, Theory of Phase Transitions: Rigorous Results, Pergamon Press (1982).
- [521] G. Slade. Scaling limits and super-Brownian motion, Notices AMS 49 (2002), 1056-1067.
- [522] G. Slade. The Lace Expansion and its Applications. Springer Lecture Notes in Mathematics, Ecole d'Été Probabilité Saint-Flour, Vol. 1879. Springer (2006).
- [523] P.F. Slade and J. Wakeley (2005). The structured ancestral selection graph and the many-demes limit, Genetics 169, 1117-1131
- [524] S. Smale (1967). Differential Dynamical Systems Bull. Amer. Math. Soc. 73, 747–817.
- [525] S. Smale (1976) On the Differential Equations of Species in Competition, J. Math. Biol. 3, 5.
- [526] R.V. Solé (1999). Statistical mechanics of network models of macroevolution and extinction, in Statistical Mechanics of Biocomplexity, Springer Lecture Notes in Physics 727.
- [527] R.V. Solé and J. Bascompte (2006). Self-organization in Complex Ecosystems, Princeton University Press.
- [528] F. Spitzer (1970) Interactions of Markov processes, Adv. in Math.
- [529] F. Spitzer (1964). Principles of Random Walk, Van Nostrand.
- [530] A. Stöckl and A. Wakolbinger (1994) On clan recurrence and transience in time stationary branching Brownian particle systems. In Measure-valued Processes, Stochastic Partial Differential Equations and Interacting Systems, ed. D. Dawson, CRM Lecture Notes and Monographs 5, 213-219.

- [531] D.W. Stroock and S.R.S. Varadhan (1979) Multidimensional diffusion processes, Springer-Verlag.
- [532] A. Sturm and J.M. Swart (2007). Voter models with heterozygosity selection, arXiv:math.PR/070155v1.
- [533] M.A. Suchard (2005). Stochastic models for horizontal gene transfer: taking a random walk through gene tree space, *Genetics*, 170, 419-431.
- [534] J.M. Swart (2000) Clustering of linearly interacting diffusions and universality of their long-time limit distribution, *Probab. Th. Rel. Fields* 118, 574-594.
- [535] E. Szathmry and L. Demeter (1987). Group selection of early replicators and the origin of life. *Journal of Theoretical Biology* 128, 463-486.
- [536] E. Szathmry (2006). The origin of replicators and reproducers. *Phil. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 1761-1776.
- [537] A.S. Sznitman (1984) Non linear reflecting diffusion processes and the propagation of chaos and fluctuations associated, *J. Funct. Anal.* 56(3), 311-336.
- [538] A.S. Sznitman (1989) Topics in the Propagation of Chaos. Ecole d'été de Probabilités de Saint-Flour XIX-1989, LNM 1464.
- [539] H. Tachida (1991). A study on a nearly neutral mutation model in finite populations, *Genetics* 128, 183-192.
- [540] N. Takahata (1991). Genealogy of neutral genes and spreading of selected mutations in a geographically structured population, *Genetics*, 129, 585-595
- [541] E. Tannenbaum and E.I. Shakhnovich (2005). Semiconservative replication, genetic repair, and many-gened genomes: extending the quasispecies paradigm to living systems, *Phys. of Life Reviews* 2, 290-317.
- [542] S. Tavaré (1984). Lines of descent and genealogical processes, and their applications in population genetics models, *Theor. Pop. Biol.* 26, 119-164.
- [543] S. Tavaré (2003). Ancestral inference in population genetics, Saint Flour 2001, Springer Lecture Notes 1837.
- [544] D. Tilman and P.M. Kareiva (1997). Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions, Princeton University Press.
- [545] A. Traulsen and M.A. Nowak (2006) Evolution of cooperation by multilevel selection, *Proc. N.A.S.* 103, 10952-10955.
- [546] A. Traulsen, J.C. Claussen and C. Hauert (2005). Coevolutionary dynamics: from finite to infinite populations, *Phys. Rev. Letters* 95, 238701-1 - 238701-4.
- [547] R. Tribe (1992). The behaviour of superprocesses near extinction, *Ann. Prob.* 20, 286-311.
- [548] R. Tribe (1996). A travelling wave solution to the Kolmogorov equation with noise, *Stochastic, Stochastics Reports* 56, 317-340.
- [549] J. Vaillancourt (1990). Interacting Fleming-Viot processes, *Stochastic Process. Appl.* 36, 45-57.
- [550] L. Van Valen (1973). A new evolutionary law. *Evolutionary Theory*, 1:1-30.
- [551] P.F. Verhulst (1838). Notice sur la loi que la population suit dans son accroissement, *Corr. Math. et Phys.* X, 268-281.

- [552] T.L. Vincent and J.S. Brown (2005). Evolutionary Game Theory, Natural Selection and Darwinian Dynamics, Cambridge Univ. Press.
- [553] V. Volterra (1926). Variazioni e fluttuazioni del numero d'individui in specie animali conviventi, Mem. Acad. Lineci Roma, 2, 31.
- [554] B. Von Bahr and A. Martin-Lf (1980). Threshold Limit Theorems for Some Epidemic Processes, Advances in Applied Probability, Vol. 12, No. 2, (Jun., 1980), pp. 319-349.
- [555] M. J. Wade (1992), *Sewall Wright, gene interaction and the shifting balance theory*, Oxford Surveys Evol. Biol. 8, 35-62.
- [556] M.J. Wade (2007) The co-evolutionary genetics of ecological communities, Nature reviews—Genetics Vol. 8, 185-195.
- [557] L. Wahl and D.C. Krakauer (2000). Models of experimental evolution: the role of genetic chance and selective necessity, Genetics 156, 1437-1448.
- [558] J. Wakeley (1998). Segragating sites in Wright's island model, Theor. Pop. Biol. 53, 166-174.
- [559] J. Wakeley (2005). The limits of theoretical population genetics, Genetics 169, 1-7.
- [560] J. Wakeley (2008). Coalescent Theory, Roberts and Co., Woodbury New York.
- [561] A. Wakolbinger (1995) Limits of spatial branching populations, Bernoulli 1(1-1), 171-189.
- [562] J.B. Walsh (1986). An introduction to stochastic partial differential equations, Ecole d 'été de Saint Flour XIV, Springer Lecture Notes in Math. 1180, 265-439.
- [563] S. Watanabe (1968) A limit theorem of branching processes and continuous state branching processes, J. Math. Kyoto Univ. 8, 141-167.
- [564] H.W. Watson and Francis Galton (1874). On the probability of extinction of families, J. Anthropol. Inst. Great Britain and Ireland 4, 138-144.
- [565] G.A. Watterson (1962). Some theoretical aspects of diffusion theory in population genetics, Ann. Math. Statist., 33, 939-957.
- [566] G.A. Watterson (1975) on the number of segregating sites in genetic models without recombination, Theor. Pop. Biol. 7, 256-276.
- [567] G.A. Watterson (1976). The stationary distribution of the infinitely many neutral alleles diffusion model, J. Appl. Prob. 13, 639-651.
- [568] H.W. Watson and Francis Galton (1874). On the probability of extinction of families, J. Anthropol. Inst. Great Britain and Ireland 4, 138-144.
- [569] J.D. Watson and J.H.C. Crick (1953). A structure for deoxyribose nucleic acid, Nature 171, 737-738.
- [570] D.J. Watts, R. Muhamad, D.C. Medina and P.S. Dodds (2005). Multiscale, resurgent epidemics in a hierarchical metapopulation model, Proc. N.A.S. USA 102, 11157-11162.
- [571] D. Waxman and S. Gavrillets (2005). "20 Questions on Adaptive Dynamics" J. of Evolutionary Biology 18: 1139-1154.
- [572] D. Waxman and S. Gavrillets (2005). "Issues of terminology, gradient dynamics, and the ease of sympatric speciation in Adaptive Dynamics" J. of Evolutionary Biology 18: 1214-1219
- [573] G.F. Webb, E. M.C. D'Agata, P. Magal and S. Ruan (2005). A model of antibiotic-resistant bacterial epidemics in hospitals, Proc. NAS, USA 102, 13343-13348.

- [574] J.W. Weibull (1995). Evolutionary Game Theory, MIT Press.
- [575] K.G. Wilson (1971). The Renormalization Group and Critical Phenomena I: Renormalization Group and the Kadanoff Scaling Picture. Phys. Rev. B4 3174 (1971).
- [576] U. Witt (2003). The Evolving Economy: Essays on the Evolutionary Approach to Economics, Edward Elgar, Cheltenham.
- [577] C.R. Woese (2002). On the evolution of cells, Proc. Nat. Acad. Sci. 99, no. 13, 8742-8747.
- [578] C.R. Woese (2004). A new biology for a new century, Microbiology and Molecular Biology Reviews 2004, 173-186.
- [579] N. Wiener (1923). Differential space, J. Math. and Phys. 2, 132-174.
- [580] H.M. Wilkinson-Herbots (1998). Genealogy and subpopulation differentiation under various models of population structure, J. Math. Biology 37, 535-585.
- [581] S. Wright (1931). Evolution in mendelian populations, Genetics 16: 97-159.
- [582] S. Wright (1932), *The roles of mutation, inbreeding, crossbreeding and selection in evolution*, Proc. 6th Int. Congress Genet., Vol. 1, 356-366.
- bibitemW-49 S. Wright (1949). Adaptation and selection. In *Genetics, Paleontology and Evolution* (G.L. Jepson, E. Mayr and G.G. Simpson,eds.) Princeton Univ. Press.
- [583] S. Wright (1977), *Evolution and the Genetics of Populations*, Vol. 3(1977), 443-473; vol. 4 (1978), 460-476, Univ. of Chicago Press.
- [584] Y. Wu (1994). Asymptotic behavior of the two level measure-valued branching process, Ann. Probab. 22.
- [585] T. Yamada and S. Watanabe (1971). On the uniqueness of solutions of stochastic differential equations, J. Math. Kyoto Univ. 11, 155-167.
- [586] K. Yosida (1948). On the differentiability and the representation of one-parameter semi-groups of linear operators, J. Math. Soc. Japan 1, 15-21.
- [587] I. Zähle (2001). Renormalization of the voter model in equilibrium, Annals Probab. 29, 1262-1302.