A mutation-selection model with recombination for general genotypes

Steven N. Evans

David Steinsaltz

Kenneth W. Wachter

Author address:

Steven N. Evans, Department of Statistics, 367 Evans Hall, University of California, Berkeley, CA 94720-3860, U.S.A.
E-mail address: `evans@stat.berkeley.edu`

David Steinsaltz, Department of Statistics, 1 South Parks Road, Oxford, OX1 3TG, UNITED KINGDOM
E-mail address: `steinsal@stats.ox.ac.uk`

Kenneth W. Wachter, Department of Demography, 2232 Piedmont Avenue, University of California, Berkeley, CA 94720-2120, U.S.A.
E-mail address: `wachter@demog.berkeley.edu`

Contents

Chapter 1. Introduction 1
- 1.1. Informal description of the limit model 3
- 1.2. Example I: Mutation counting 7
- 1.3. Example II: Polynomial selective costs 9
- 1.4. Example III: Demographic selective costs 10
- 1.5. Comments on the literature 11
- 1.6. Overview of the remainder of the work 12

Chapter 2. Definition, existence, and uniqueness of the dynamical system 13
- 2.1. Spaces of measures 13
- 2.2. Definition of the dynamical system 14
- 2.3. Existence and uniqueness of solutions 15
- 2.4. Lemmas used in the proof of existence and uniqueness 16

Chapter 3. Equilibria 21
- 3.1. Introductory example: One-dimensional systems 22
- 3.2. Introductory example: Multiplicative selective cost 23
- 3.3. Fréchet derivatives 26
- 3.4. Existence of equilibria via perturbation 27
- 3.5. Concave selective costs 28
- 3.6. Iterative computation of the minimal equilibrium 33
- 3.7. Stable equilibria in the concave setting via perturbation 33
- 3.8. Equilibria for demographic selective costs 35

Chapter 4. Mutation-selection-recombination in discrete time 39
- 4.1. Mutation and selection in discrete-time 39
- 4.2. Recombination in discrete-time 41
- 4.3. The discrete-time model 42

Chapter 5. Hypotheses and statement of the convergence result 43
- 5.1. Shattering of point processes 43
- 5.2. Statement of the main convergence result 44

Chapter 6. Convergence with complete Poissonization 47

Chapter 7. Preparatory lemmas for the main convergence result 51
- 7.1. Consequences of shattering 51
- 7.2. Estimates for Radon-Nikodym derivatives 53
- 7.3. Palm measures 58
- 7.4. Comparisons with complete Poissonization 59

Chapter 8. Proof of convergence of the discrete-generation systems 69

Chapter 9. Convergence to Poisson of iterated recombination 75

Appendix A. An expectation approximation 79

Bibliography 81

Abstract

We investigate a continuous time, probability measure valued dynamical system that describes the process of mutation-selection balance in a context where the population is infinite, there may be infinitely many loci, and there are weak assumptions on selective costs. Our model arises when we incorporate very general recombination mechanisms into a previous model of mutation and selection from Steinsaltz, Evans and Wachter (2005) and take the relative strength of mutation and selection to be sufficiently small. The resulting dynamical system is a flow of measures on the space of loci. Each such measure is the intensity measure of a Poisson random measure on the space of loci: the points of a realization of the random measure record the set of loci at which the genotype of a uniformly chosen individual differs from a reference wild type due to an accumulation of ancestral mutations. Our main motivation for working in such a general setting is to provide a basis for understanding mutation-driven changes in age-specific demographic schedules that arise from the complex interaction of many genes, and hence to develop a framework for understanding the evolution of aging.

Received by the editor October 16, 2009.

2000 *Mathematics Subject Classification.* Primary 60G57, 92D15; Secondary 37N25, 60G55, 92D10.

Key words and phrases. measure-valued, dynamical system, population genetics, quasi-linkage equilibrium, Poisson random measure, Wasserstein metric, Palm measure, shadowing, stability, attractivity.

SNE supported in part by grants DMS-04-05778 and DMS-09-07630 from the National Science Foundation (U.S.A.).

DR supported by grant K12-AG00981 from the National Institute on Aging (U.S.A.) and a Discovery Grant from the National Science and Engineering Research Council (Canada).

KWW supported by grant P01-008454 from the National Institute on Aging (U.S.A.) and by the Miller Institute for Basic Research in Science at U.C. Berkeley.

CHAPTER 1

Introduction

A principal goal of population genetics is to understand how the mechanisms of mutation, selection and recombination determine the manner in which the distribution of genotypes in a population changes over time. The distribution of genotypes dictates, to a large degree, the distribution of phenotypes, an object that is both more readily observable and of more immediate practical interest. Conversely, the differential survival of genotypes is mediated almost exclusively by their phenotypic expression.

More specifically, *mutation accumulation* models that incorporate the countervailing forces of recurring, slightly deleterious mutations and persistent selection have been staples of evolutionary theory [**Bür00**]. In order to use these models to explain phenomena such as aging that are presumed to result from the combination of many small mutational impacts, it is necessary to adopt a multilocus perspective. However, as Pletcher and Curtsinger [**PC98**] point out, early progress in this area has relied on simplifying assumptions that are severely limiting and possibly unfounded: equal impacts from all mutations, additive effects of mutations on age-specific survival, and the existence of mutations that impact a specific, narrow range of ages.

Several general multilocus formalisms have been proposed. Notable among these is that of [**KJB02**], which is designed to allow almost any conceivable regime of selection, mating, linkage, mutation, and phenotypic effects. Such frameworks have not been exploited for studies of aging, in part because so much detail is counterproductive when what concerns us is not the fate of any individual allele but rather the mass of overlapping, age-varying phenotypic effects that are central to standard theories of aging.

K. Dawson [**Daw99**] applied a variant of the Kimura-Maruyama rare-allele approximation [**KM66, Bür00**] (see also [**Kon82**]) to aging. While this less detailed view of the genome is more amenable to theoretical analysis, it is also not suited to describing the interacting phenotypic contribution of multiple loci.

In a previous paper [**SEW05**], we proposed a model, which we also describe briefly here in the development leading to (1.4), that overcame many of these limitations. That model leads to computable solutions for mutation-selection equilibria, the hazard functions which such equilibria imply, and the time evolution of the population distribution of genotypes.

In this work, we define and analyze a parallel model that, in essence, incorporates recombination into the model of [**SEW05**]. The key assumptions behind the model in this work are:

- the population is infinite,
- the genome may consist of infinitely many or even a continuum of loci,
- reproduction is sexual, in that each individual has two parents,

- the mechanism of genetic recombination randomly shuffles together the genomes of the parents in order to obtain the genome of the offspring,
- mating is random and individuals are identical except for their genotypes, so the population at any time can be completely described by a probability measure on the space of possible genotypes — heuristically, the empirical distribution of genotypes in the population or, equivalently, the probability distribution of the genotype of a randomly sampled individual,
- individuals are *haploid* — an individual has only one copy of each gene rather than copies from each of two parents,
- down a lineage, mutant alleles only accumulate — there is no back-mutation to cancel out the alleles introduced by earlier mutation events,
- fitness is calculated for individuals rather than for mating pairs,
- a genotype becomes less fit when it accumulates additional mutant alleles, but otherwise selective costs are arbitrary,
- recombination acts on a faster time scale than mutation or selection — the common quasi-linkage equilibrium (QLE) assumption.

Our general picture of the genome and the processes of mutation, selection and recombination is similar to that of [**BT91, KJB02**]. Whereas [**BT91, KJB02**] invoke the QLE assumption to justify treating the effects of alleles at different loci as nearly independent, we present a detailed asymptotic treatment in a standard, explicit scaling regime.

More specifically, we establish conditions under which a discrete-time process with the above features converges to a continuous-time, deterministic, dynamical system that has its state space the probability measures on the set of possible genotypes. At any point in time, the genotype distribution is in the complete linkage equilibrium represented by the distribution of a Poisson point process. The points of a realization of the Poisson process represent loci at which a randomly sampled individual's genotype has accumulated ancestral mutations away from an original "reference" genotype which we will call the *wild type*. Because a Poisson process is defined by its intensity measure, which in our case is a finite measure on the set of loci, the asymptotic model can be more simply described by a deterministic dynamical system that moves about in the space of such measures.

In order to establish such a convergence result it is, of course, not enough simply to show that the distribution of the genotype of a randomly chosen individual at a fixed time converges to the distribution of a Poisson random measure. Rather, we must keep track of the accumulated perturbations that arise over time from the effect of mutation, selection and recombination, and demonstrate the convergence of the entire time evolution of the genotype distribution to a dynamical system of Poisson distributions.

Moreover, we establish conditions under which the continuous-time dynamical system has equilibria, investigate when the system converges to an equilibrium from the pure wild type genotype, and obtain results about the stability and attractivity of that particular equilibrium when it is present.

Our models provide a basis for the rigorous study of a number of questions in the biodemography of longevity, as surveyed in [**Wac03**]:

- the adequacy of Charlesworth's [**Cha01**] proposed explanation of Gompertz hazards and mortality plateaus as a consequence of mutation accumulation,

- the causes of the hyperexponential hazards described by Horiuchi [**Hor03**],
- extensions to age-structured settings of Haldane's principle that, in its original form, equates the mutation rate to population decline of fitness,
- contrasts between proportional and additive mutation effects on hazards as discussed in [**Bau05, Bau08**],
- the possibility of mortality rates diverging to infinity at ages before the end of reproduction, which have been termed "Walls of Death".

We pursue these matters in the companion papers [**WES08, WSE08**].

1.1. Informal description of the limit model

In this section we describe the asymptotic model and motivate why it is reasonable that such a dynamical system should arise as the limit of a sequence of discrete-time systems with the features listed above.

Denote by $\mathcal{M}$ the collection of *loci* in the portion of the genome that is of interest to us. There is a distinguished reference wild type genotype, and each locus represents a "position" at which the genotype of an individual may differ from that of the wild genotype. We allow the set $\mathcal{M}$ to be quite general and do not necessarily think of it as a finite collection of physical DNA base positions or a finite collection of genes. For example, the proposed explanation for the Gompertz mortality curve and mortality plateaus at extreme ages in Charlesworth [**Cha01**] suggest taking $\mathcal{M}$ to be a class of functions from $\mathbb{R}_+$ to $\mathbb{R}_+$: the value of such a function at time $t \geq 0$ represents an additional increment to the mortality hazard rate at age t conferred by a mutation away from the wild type at this locus. Some structure on $\mathcal{M}$ is necessary to accommodate rigorous probability theory, so we take $\mathcal{M}$ to be a complete, separable metric space.

The genotype of an individual is specified by the set of loci at which there has been a mutation somewhere along the ancestral lineage leading to that individual. More precisely, a genotype is an element of the space $\mathcal{G}$ of integer–valued finite Borel measures on $\mathcal{M}$. An element of $\mathcal{G}$ is a finite sum $\sum_i \delta_{m_i}$, where δ_m is the unit point mass at the locus $m \in \mathcal{M}$. The measure $\sum_i \delta_{m_i}$ corresponds to a genotype that has ancestral mutations at loci $m_1, m_2, \ldots$. The wild type genotype is thus the null measure. We do not require that the loci $m_i \in \mathcal{M}$ be distinct. We thus allow several copies of a mutation. This is reasonable, since we are not identifying mutations with changes in nucleotide sequences in a one-to-one manner.

For example, if $\mathcal{M}$ is finite, so we might as well take $\mathcal{M} = \{1, 2, \ldots, N\}$ for some positive integer N, then $\mathcal{G}$ is essentially the Cartesian product $\mathbb{N}^N$ of N copies of the nonnegative integers: A genotype is of the form $\sum_{j=1}^N n_j \delta_j$, indicating that an ancestral mutation is present n_j times at locus j, and we identify such a genotype with the nonnegative integer vector $(n_1, n_2, \ldots, n_N)$.

Recall that the population is infinite and all that matters about an individual is the individual's genotype, so that the dynamics of the population are described by the proportions of individuals with genotypes that belong to the various subsets of $\mathcal{G}$. We are thus led to consider a family of probability measures P_t, $t \geq 0$, on $\mathcal{G}$, where $P_t(G)$ for some subset $G \subseteq \mathcal{G}$ represents the proportion of individuals in the population at time t that have genotypes belonging to G. Note that $P_t(G)$ also may be thought of as the probability that an individual chosen uniformly at random from the population will have genotype belonging to the set G. In other words, P_t is the distribution of a random finite integer-valued measure on $\mathcal{M}$. For example, if $\mathcal{M} = \{1, 2, \ldots, N\}$ and we identify $\mathcal{G}$ with the Cartesian product $\mathbb{N}^N$ as above, then $P_t(\{(n_1, n_2, \ldots, n_N)\})$ represents the probability that an individual chosen uniformly at random from the population will have n_j ancestral mutations at locus j for $j = 1, 2, \ldots, N$.

We next indicate how we model mutation, selection and recombination to obtain the evolution dynamics for P_t, $t \geq 0$.

Mutation alone. Suppose that there is only mutation and no selection or recombination. In this case all individuals present in the population at a give time die and reproduce at the same rate because differing genotypes do not confer differing selective costs and mutations accumulate down lineages because they cannot be replaced by recombination.

We describe the mutation process using a finite measure ν on the space of loci $\mathcal{M}$, where $\nu(B)$ for $B \subseteq \mathcal{M}$ gives the rate at which mutations from the ancestral wild type belonging to the set B accumulate along a given lineage.

Write $P_t\Phi = \int_{\mathcal{G}} \Phi(g)\, dP_t(g)$ for some test function $\Phi : \mathcal{G} \to \mathbb{R}$. That is, $P_t\Phi$ is the expected value of the real-valued random variable obtained by applying the function Φ to the genotype of an individual chosen uniformly at random from the population. The content of our assumptions is that when there is only mutation P_t, $t \geq 0$, evolves according to

$$\frac{d}{dt} P_t \Phi = P_t \left(\int_{\mathcal{M}} [\Phi(\cdot + \delta_m) - \Phi(\cdot)]\ \nu(dm) \right). \tag{1.1}$$

For example, when $\mathcal{M} = \{1, 2, \ldots, N\}$ we have the system of ordinary differential equations

$$\frac{d}{dt} P_t(\{\mathbf{n}\}) = \sum_{j=1}^{N} \nu(\{j\}) \left[P_t(\{\mathbf{n} - \mathbf{e_j}\}) - P_t(\{\mathbf{n}\}) \right],$$

where $\mathbf{e_j}$ is the j^{th} coordinate vector. This equation is, of course, just a special case of the usual equation (see, for example, Section III.1.2 of [**Bür00**]) describing evolution due to mutation of type frequencies in a population where the set of types is $\mathbb{N}^N$ and mutation from type $\mathbf{n}$ to type $\mathbf{n} + \mathbf{e_j}$ occurs at rate $\nu(\{j\})$.

The evolution equation (1.1) is of the form

$$\frac{d}{dt} P_t = A P_t$$

for a certain linear operator A. We recognize that A is the the infinitesimal generator of a $\mathcal{G}$-valued Lévy process, and hence (1.1) has the following explicit probabilistic solution. Let Π denote a Poisson random measure on $\mathcal{M} \times \mathbb{R}_+$ with intensity measure $\nu \otimes \lambda$, where λ is Lebesgue measure; that is, Π is a random integer-valued Borel measure such that:

1. The random variable $\Pi(A)$ is Poisson with expectation $(\nu \otimes \lambda)(A)$ for any Borel subset A of $\mathcal{M} \times \mathbb{R}_+$.
2. If $A_1, A_2, \ldots, A_n$ are disjoint Borel subsets of $\mathcal{M} \times \mathbb{R}_+$, then the random variables $\Pi(A_k)$ are independent.

Define a $\mathcal{G}$-valued random variable Z_t (that is, Z_t is a random finite integer-valued measure on $\mathcal{M}$) by

$$Z_t := \int_{\mathcal{M} \times [0,t]} \delta_m \, d\Pi((m, u)).$$

Then,

$$P_t\Phi = \mathbb{E}\left[\Phi(W + Z_t)\right],$$

where W is a random measure on $\mathcal{M}$ that has distribution P_0 and is independent of Π. In particular, if P_0 is itself the distribution of a Poisson random measure, then P_t will also be the distribution of a Poisson random measure. If we write ρ_t for the intensity measure associated with P_t (that is, ρ_t is the measure on $\mathcal{M}$ defined by $\rho_t(M) := \int_{\mathcal{G}} g(M)\, dP_t(g)$ for $M \subseteq \mathcal{M}$), then ρ_t evolves according to the simple dynamics

$$\rho_t(M) = \rho_0(M) + t\,\nu(M).$$

Selection alone. Now suppose there is only selection and no mutation or recombination. We specify the fitnesses of different genotypes by a *selective cost function* $S : \mathcal{G} \to \mathbb{R}_+$. The difference $S(g') - S(g'')$ for $g', g'' \in \mathcal{G}$ is the difference in the rate of sub-population growth between the sub-population of individuals with genotype g'' and the sub-population of individuals with genotype g'. We make the normalizing assumption $S(0) = 0$ and suppose that

$$S(g + h) \geq S(h), \quad g, h \in \mathcal{G}, \tag{1.2}$$

in line with our assumption above that genotypes with more accumulated mutations are less fit.

It follows that at time $t \geq 0$ the per individual rate of increase of the proportion of the population of individuals with genotype g' is $P_tS - S(g')$. More formally,

$$\begin{aligned} \frac{d}{dt} P_t\Phi &= -\,P_t(\Phi \cdot [S - P_tS]) \\ &= -\int_{\mathcal{G}} \Phi(g') \left[S(g') - \int_{\mathcal{G}} S(g'')\, dP_t(g'')\right] dP_t(g'). \end{aligned} \tag{1.3}$$

For example, when $\mathcal{M} = \{1, 2, \ldots, N\}$ we have

$$\frac{d}{dt} P_t(\{\mathbf{n}'\}) = -\left[S(\mathbf{n}') - \sum_{\mathbf{n}''} P_t(\{\mathbf{n}''\}) S(\mathbf{n}'')\right] P_t(\{\mathbf{n}'\}).$$

If S is *non-epistatic*, that is, S has the additive property

$$S(\sum_i \delta_{m_i}) = \sum_i S(\delta_{m_i}),$$

then the selective effects of different mutations don't interact. In particular, if P_0 is the distribution of a Poisson random measure, then P_t will also be the distribution of a Poisson random measure and, writing ρ_t for the intensity measure associated with P_t as before, we have

$$\rho_t(dm') = \rho_0(dm') - \int_0^t \left[S(\delta_{m'}) - \int_{\mathcal{M}} S(\delta_{m''}) \rho_s(dm'')\right] \rho_s(dm')\, ds.$$

However, if S is epistatic (that is, non-additive), then P_t will, in general, not be the distribution of a Poisson random measure — even when P_0 is.

Combining mutation and selection. If there is mutation and selection, but no recombination, then the appropriate evolution equation for P_t comes from simply combining equations (1.1) and (1.3):

$$\begin{aligned}\frac{d}{dt}P_t\Phi = P_t\left(\int_{\mathcal{M}}[\Phi(\cdot+\delta_m)-\Phi(\cdot)]\;\nu(dm)\right) \\ -\;P_t(\Phi\cdot[S-P_tS]).\end{aligned}\tag{1.4}$$

This is the model introduced and analyzed at length in [**SEW05**] using the Feynman-Kac formula. When $\mathcal{M}=\{1,2,\ldots,N\}$, (1.4) is a special case of the classical system of ordinary differential equations for mutation and selection in continuous-time – see Section III.1.2 of [**Bür00**] for the derivation of an analytic solution that agrees with the one that arises from a Feynman-Kac analysis.

Recombination alone. The effect of recombination is to choose an individual uniformly at random from the population at some rate and replace the individual's genotype g' with a genotype of the form $g'(\cdot\cap M)+g''(\cdot\cap M^c)$, where g'' is the genotype of another randomly chosen individual, M is a subset of $\mathcal{M}$ chosen according to a suitable random mechanism, and M^c is the complement of M. Thus, recombination randomly shuffles together two different genotypes drawn from the population. In order to specify the recombination mechanism fully, we would need to specify the recombination rate and the distribution of the *segregating set* M. Suppose, however, that we are in the following regime:

- recombination acts alone (that is, there is no mutation or selection),
- the initial population P_0 has the property that there does not exist an $m\in\mathcal{M}$ with $P_0(\{g\in\mathcal{G}:g(\{m\})>0\})>0$ (that is, no single mutation from the ancestral wild type is possessed by a positive proportion of the initial population),
- the mechanism for choosing the *segregating* set M is such that, loosely speaking, if m' and m'' are two loci then there is positive probability that $m'\in M$ and $m''\in M^c$ (that is, no region of the genome $\mathcal{M}$ is immune from the shuffling effect of recombination).

Then, under these condition the probability measure P_t will converge as $t\to\infty$ to the distribution of a Poisson random measure on $\mathcal{M}$ with the same intensity measure as P_0. Moreover, the speed of this convergence increases with the recombination rate and so if we take the the recombination rate to be effectively infinite, then the distribution P_t will be essentially Poisson for all $t>0$ with the same intensity measure as P_0, irrespective of the details of the recombination mechanism.

Combining mutation, selection and recombination. We have seen that if P_0 is the distribution of a Poisson random measure, then mutation preserves this property. On the other hand epistatic selection drives the population distribution away from Poisson, while increasing rates of recombination push it towards Poisson. Thus, when all three processes operate and we consider a limiting regime where recombination acts on a much faster time scale than selection and recombination, we expect asymptotically that if the initial condition P_0 is the distribution of a Poisson random measure on $\mathcal{M}$, then P_t will also be the distribution of a Poisson random measure for all $t>0$. Let ρ_t denote the intensity measure of P_t (so that ρ_t is a finite measure on the space $\mathcal{M}$ of loci). If we write X^π for a Poisson random measure on $\mathcal{M}$ with intensity measure π, then we expect from combining the observations above that ρ_t should satisfy the evolution equation

$$\rho_t(dm) = \rho_0(dm) + t\,\nu(dm) - \int_0^t \mathbb{E}\left[S(X^{\rho_s} + \delta_m) - S(X^{\rho_s})\right] \rho_s(dm)\,ds. \tag{1.5}$$

We define the rigorous counterpart of (1.5) in Chapter 2 and establish the existence and uniqueness of solutions. Furthermore, we show in Chapter 4 that our dynamical equation is indeed a limit of a sequence of standard discrete generation, mutation-selection-recombination models.

Given that it involves computing an expected value for a quite general Poisson process, equation (1.5) may look rather forbidding. However, for certain reasonable choices of selective costs the infinite sum can be evaluated explicitly, leading to a simpler and more intuitive looking system. We will consider three such cases in the following three sections. First, though, we make a useful observation about how loci may be "clumped together" in (1.5).

REMARK 1.1. Consider two instances $(\rho'_t)_{t\geq 0}$ and $(\rho''_t)_{t\geq 0}$ of the dynamical system (1.5) with respective locus spaces $\mathcal{M}'$ and $\mathcal{M}''$, associated genotype spaces $\mathcal{G}'$ and $\mathcal{G}''$, mutation intensity measures ν' and ν'', and selective costs S' and S''. Suppose that there is a Borel measurable map T from $\mathcal{M}'$ onto $\mathcal{M}''$ such that:

- the initial measure ρ''_0 on $\mathcal{M}''$ is the push-forward of the initial measure ρ' on $\mathcal{M}'$ by the map T,
- the mutation intensity measure ν'' on $\mathcal{M}''$ is the push-forward of the mutation intensity measure ν' on $\mathcal{M}'$ by the map T,
- the selective cost S' on $\mathcal{G}'$ has the property that $S'(g') = S'(h')$ whenever the push-forwards of $g', h' \in \mathcal{G}'$ by T are the same,
- the selective cost $S''(g'')$ on $\mathcal{G}''$ is given by the common value of $S'(g')$ for all $g' \in \mathcal{G}'$ that have push-forward by T equal to g''.

Then, for each $t > 0$, the measure ρ''_t on $\mathcal{M}''$ is the push-forward of the measure ρ'_t on $\mathcal{M}'$ by the map T. Intuitively, we have a situation in which for a given for $g'' \in \mathcal{G}''$ any two genotypes in the set $T^{-1}(g'') \subseteq \mathcal{G}'$ are indistinguishable in terms of their associated selective cost, and so we may identify any two such genotypes as being the same. Of course, we cannot recover the finer description ρ'_t from the coarser one ρ''_t in general, but if we let P'_t and P''_t be the population distributions of genotypes corresponding to ρ'_t and ρ''_t (that is, P'_t and P''_t are the distributions of Poisson random measures on $\mathcal{M}'$ and $\mathcal{M}''$ with intensity measures ρ'_t and ρ''_t), then the push-forward of P'_t by S' is the same as the push-forward of P''_t by S'': the population distributions of selective costs agree whether we use the fine or the coarse description of genotypes.

1.2. Example I: Mutation counting

The simplest special case of our framework occurs when there are many loci but the selective cost of a genotype g only depends on the total number of loci $g(\mathcal{M})$ at which there have been ancestral mutations away from the wild type.

For example, suppose that the space $\mathcal{M}$ of loci is the unit interval $[0, 1]$ and the selective cost S is of the form $S(g) = s(g(\mathcal{M}))$ for some non-decreasing function $s : \mathbb{N}_0 \to \mathbb{R}_+$ with $s(0) = 0$.

We may apply Remark 1.1 and "replace" the locus space $\mathcal{M}$ by a single point. Let $q := \nu(\mathcal{M})$ be the total rate at which mutations occur and write $r_t := \rho_t(\mathcal{M})$ for the expected number of ancestral mutations in the genotype of an individual chosen at random from the population at time $t \geq 0$. It follows from Remark 1.1 that the function r evolves autonomously according to the (non-linear) ordinary differential equation

$$\dot{r} = q - \psi(r),$$

where

$$\begin{aligned}\psi(x) &:= x \sum_{k=0}^{\infty} (s(k+1) - s(k)) e^{-x} \frac{x^k}{k!} \\ &= e^{-x} \sum_{k=1}^{\infty} s(k) \frac{x^k}{k!} (k - x).\end{aligned}$$

One example of this simplified model, which lines up with models familiar from earlier literature, assumes the cost per mutation to be a constant $\bar{s}$, so that $s(k) = \bar{s}k$ and $\psi(x) = \bar{s}x$, and the starting point to be the null genotype. In this case, we readily compute that the intensity at time t is $r_t = (q/\bar{s})\big(1 - \exp(-\bar{s}t)\big)$. The intensity converges monotonically to the equilibrium value $q/\bar{s}$, the elementary expression for mutation-selection equilibrium going back to Haldane. Thus, the limiting distribution for the number of mutations from the ancestral wild type is Poisson with mean $q/\bar{s}$. Our assumption about recombination leads to simpler answers than Kimura and Maruyama's treatment of mutation counting without recombination (see [**KM66**]).

When costs are multiplicative, in the sense that $s(k) := 1 - \exp(-k\sigma)$ for some constant $\sigma > 0$, we have $\psi(x) = ax\exp(-ax)$, where $a := 1 - \exp(-\sigma)$. An equilibrium exists only if the mutation rate q is below the maximum of $r\exp(-r)$, namely e^{-1}, and in that case r_t converges monotonically to the smallest positive solution of the equation $q - \psi(x) = 0$. The solution is $x = W(-q)/a$, where $W(z) := \sum_{n=1}^{\infty} (-n)^{n-1} \frac{z^n}{n!}$ is Lambert's W function – the power series expansion is a simple application of the technique known variously as reversion of series or the Lagrange inversion formula. These properties for multiplicative costs are generalized in Section 3.2. Note also that when σ is small the equilibrium is approximately q/σ, as one would expect from the observation that in this case $s(k)$ is approximately σk for small k and hence this model is approximately the additive one of the previous paragraph with $\bar{s} = \sigma$.

If we are interested only in how the population distribution of selective costs evolves, then we need consider only $(r_t)_{t\geq 0}$, rather than $(\rho_t)_{t\geq 0}$. However, we should be somewhat careful about how we interpret the biological import of this simplification. We justified the dynamical system (1.5) as describing the evolving population distribution of genotypes defined in terms of the locus space $\mathcal{M}$ in a population undergoing mutation, selection and recombination. Mathematically, we see that in this example we can replace the locus space $[0,1]$ by a single point for the purposes of studying the dynamics of the distribution of selective costs, but this does not mean that biologically the multilocus model is identical with a single locus model. As we show later, our instance of (1.5) with locus space $[0,1]$ arises as a limit of discrete generation models in which Poisson random measures appear because of the manner in which recombination breaks up and shuffles together genotypes from different individuals. Even though our instance of (1.5) with locus space a single point is mathematically well-defined, it can't arise as a limit of such discrete generation models because with a single locus there is no way that recombination can drive genotype distributions towards Poisson.

1.3. Example II: Polynomial selective costs

Recall that when $\mathcal{M} = \{1, 2, \ldots, N\}$ we can encode genotypes as ordered $N-$tuples of nonnegative integers, where the entry in the k^{th} coordinate is the number of ancestral mutations present at locus k. Then, (1.5) becomes the system of ordinary differential equations

$$\frac{d}{dt}\rho_t(\{j\}) = \nu(\{j\}) - \rho_t(\{j\}) \sum_{\mathbf{n}} [S(\mathbf{n} + \mathbf{e_j}) - S(\mathbf{n})] \prod_{k=1}^{N} e^{-\rho_t(\{k\})} \frac{\rho_t(\{k\})^{n_k}}{n_k!}. \tag{1.6}$$

As in the previous example, we should not think of $\{1, 2, \ldots, N\}$ as being the "real" locus space. Rather, we should imagine that there is something like a continuum of loci which may be partitioned into N sub-classes in such a way that the loci in each sub-class have indistinguishable selective effects, and (1.6) is a reduced description that comes from applying the observation in Remark 1.1.

A natural family of selective costs is given by those of the polynomial form

$$S(g) = \sum_{I} \alpha_I g^I,$$

where the sum is taken over all nonempty subsets $I \subseteq \{1, \ldots, N\}$ and we adopt the usual multi-index convention that for a vector v the notation v^I denotes the product $\prod_{i \in I} v_i$. The constant $\alpha_{\{i\}}$ for $1 \leq i \leq N$ measures the selective cost of mutation i alone, whereas the constant α_I for a subset $I \subseteq \{1, \ldots, N\}$ of cardinality greater than one measures the selective cost attributable to interactions between all of the mutations in I over and above that attributable to interactions between mutations in proper subsets of I.

The system of ordinary differential equations (1.6) becomes

$$\dot{\rho_k} = \nu_k - \sum_{I \in \mathcal{I}_k} \alpha_I \rho^I, \quad 1 \leq k \leq N, \tag{1.7}$$

where we write $\rho_k := \rho(\{k\})$ and $\nu_k := \nu(\{k\})$, and where $\mathcal{I}_k$ denotes the collection of subsets of $\{1, \ldots, N\}$ that contain k, see [**CE09**].

It is shown in [**CE09**] that if $\nu_k > 0$ for all k (that is, mutations may occur at all loci), $\alpha_{\{i\}} > 0$ for all i (that is, the individual effect of any mutation is deleterious, in keeping with our general assumption on the selective cost), and $\alpha_I \geq 0$ for all subsets I (that is, the synergistic effects of individually deleterious mutations are never beneficial), then the system of equations (1.7) has a unique equilibrium point in the positive orthant $\mathbb{R}_+^N$. Moreover, this equilibrium is globally attractive; that is, the system converges to the equilibrium from any initial conditions in $\mathbb{R}_+^N$. The condition $\alpha_I \geq 0$ for all I certainly implies our standing assumption that $S(g+h) \geq S(g)$ for all $g, h \in \mathcal{G}$, but it is strictly stronger. It is also shown in [**CE09**] that the analogue of this result for general $\mathcal{M}$ holds with a suitable definition of polynomial selective costs in terms of sums of integrals against products of the measure g with itself.

1.4. Example III: Demographic selective costs

The following model is discussed in [**WES08, WSE08**], where there is a more extensive discussion of the demographic assumptions.

For the moment, suppose that the space of loci $\mathcal{M}$ is general. Write $\ell_x(g)$ for the probability that an individual with genotype $g \in \mathcal{G}$ lives beyond age $x \in \mathbb{R}_+$. At age x, the corresponding *cumulative hazard* and *hazard function* are thus $-\log \ell_x(g)$ and $d/dx(-\log \ell_x(g))$, respectively. Suppose that the infinitesimal rate that an individual at age $x \in \mathbb{R}_+$ has offspring is f_x, independently of the individual's genotype. For individuals with genotype g, the size of the next generation relative to the current one is $\int_0^\infty f_x \ell_x(g)\, dx$.

Suppose further that there is a background hazard λ and that an ancestral mutation at locus $m \in \mathcal{M}$ contributes an increment $\theta(m,x)$ to the hazard function at age x. Thus, the probability that an individual with genotype $g \in \mathcal{G}$ lives beyond age $x \in \mathbb{R}_+$ is

$$\ell_x(g) = \exp\left(-\lambda x - \int_{\mathcal{M}} \theta(m,x)\, dg(m)\right).$$

The corresponding selective cost is

$$S(g) = \int_0^\infty f_x\, \ell_x(0)\, dx - \int_0^\infty f_x\, \ell_x(g)\, dx$$

(recall that selective costs represent relative rates of increase, and we have adopted the normalizing convention that $S(0) = 0$).

Note that

$$S(g + \delta_m) - S(g) = \int_0^\infty \left(1 - e^{-\theta(m,x)}\right) f_x\, \ell_x(g)\, dx$$

and

$$\mathbb{E}\left[\ell_x(X^\pi)\right] = \exp\left(-\lambda x - \int_{\mathcal{M}} \left(1 - e^{-\theta(m,x)}\right)\, \pi(dm)\right)$$

for any finite measure π on $\mathcal{M}$ (recall that X^π is a Poisson random measure on $\mathcal{M}$ with intensity measure π).

An important issue for demographic applications is whether the solution ρ_t, $t \geq 0$, of (1.5) converges to an equilibrium ρ_∞ as $t \to \infty$ and, if so, what are the features of that equilibrium – in particular, what we can say about $\mathbb{E}[\ell_x(X^{\rho_\infty})]$, the probability that a randomly chosen individual from the equilibrium population lives beyond age x. It is not hard to show that if the limit ρ_∞ exists, then it must be absolutely continuous with respect to the mutation rate measure ν and have a Radon-Nikodym derivative r_∞ that satisfies

$$1 = r_\infty(m) \int_0^\infty (1 - e^{-\theta(m,x)})\, f_x\, \mathbb{E}[\ell_x(X^{\rho_\infty})]\, dx.$$

We study such equilibria further in Section 3.8 in and in [**WES08, WSE08**]. In particular, we show in [**WES08**] that if $\mathcal{M} = \mathbb{R}_+$, the mutation rate measure ν is absolutely continuous with respect to Lebesgue measure, and θ is of the form $\theta(m,x) = \eta(m)\mathbf{1}_{x \geq m}$, so that we consider mutations which each provide a point mass increment to the hazard at a specific age, then the equilibrium equation turns out to be equivalent to a second-order, non-linear, ordinary differential equation in one variable that can be solved explicitly.

1.5. Comments on the literature

We make some brief remarks about the substantial literature on multilocus deterministic models in population genetics and its relation to our work.

A very comprehensive recent reference is Reinhard Bürger's book [**Bür00**] (see also Bürger's review paper [**Bür98**]). As well as giving an overview of the classical models for finitely many alleles at each of a finite number of loci, these works consider at length deterministic haploid *continuum-of-alleles* models in which individuals have a *type* that is thought of as the contribution of a gene to a given quantitative trait. The type belongs to a general state space that represents something like the trait value (in which case the state space is a subset of $\mathbb{R}$) and is often thought of as the combined effect of a multilocus genotype. Each type has an associated fitness, which is some fairly arbitrary function from the type space to $(0, 1]$. However, the models do not explicitly incorporate a family of loci, the configuration of alleles present at those loci, or a function describing the fitness of a configuration. Rather, everything is cast in terms of how fit each type is and how likely one type is to mutate into another.

Certain classes of mutation-selection models without recombination are solved explicitly in [**WBG98, BW01**] using ideas from statistical mechanics. Such models may be thought of as either multilocus systems with complete linkage or structured single locus systems. We also mention the constellation of papers [**BB03, Baa05, Baa07, Baa01**] presenting a deterministic model of population change due to recombination alone.

Finally, we make two comments about our model in order to distinguish it from others in the literature that superficially might seem to have similar features.

- We do **not** present a Markovian stochastic model such as the one in [**SH92**], where the population is described in terms of an evolving Poisson random measure that keeps track of the proportion of alleles at each site that are mutant (with a typical site being purely wild type and only exceptional sites having a positive proportion of mutants present). For us, the evolution of the proportions of different genotypes in the population is described by a deterministic dynamical system living on a space of probability measures. If we sample from the evolving probability measure at some fixed time, then the resulting individual's genotype is a Poisson random measure.
- Linkage arises in our discrete-time approximation models as the dependence between loci, a natural consequence of non-additive selection costs, is only partially broken up in any finite number of rounds of recombination. However, linkage does not appear in the limit model. That is, if we sample from the limit population at some time, then the fact that the resulting individual's genotype is described by a Poisson random measure means that the presence of ancestral mutations from wild type in one part of the genotype is independent of the presence of mutations in another part. Our limit theorem may therefore be thought of as delineating the relative strengths of mutation, selection and recombination that lead asymptotically to a situation in which the Poissonizing effect of recombination wins out over the interactions introduced by non-additive selection. Understanding how these two forces counteract each other is far from trivial and is the most demanding technical task of the present

work. We cannot stress too strongly that we have not *a priori* assumed that linkage is not present.

1.6. Overview of the remainder of the work

We define the measure-valued dynamical system (1.5) rigorously in Chapter 2 and establish the existence and uniqueness of solutions.

We investigate in Chapter 3 whether the dynamical system has equilibria and whether these equilibria are stable and attractive.

We devote Chapters 4 to 8 to showing that the dynamical system is a limit of discrete-generation, infinite-population models. We define the discrete-generation models in Chapter 4 and we state the convergence theorem and its hypotheses in Chapter 5. As a first step towards the proof of the convergence result, we show in Chapter 6 that an analogous convergence result holds when the recombination mechanism is replaced by a complete Poissonization operation that destroys all linkage between loci in a single step. The proof of the actual convergence result is quite involved and requires a number of technical estimates that, loosely speaking, bound the extent to which selection reintroduces linkage that has been partially removed by recombination. We present these preliminary results in Chapter 7. We complete the proof of the convergence theorem in Chapter 8.

The passage from the discrete-generation mutation-selection-recombination model to its continuous-time limit involves the "linearization" of operators that describe mutation and selection. This linearization is essentially a consequence of the fact that $\log \mathbb{E}\left[e^{-T}\right]$ is approximately $-\mathbb{E}[T]$ when the nonnegative random variable T is close to the being constant. A simple quantitative version of this observation is used several times in this work. We state the result and prove it in Appendix A.

CHAPTER 2

Definition, existence, and uniqueness of the dynamical system

2.1. Spaces of measures

Recall from the Introduction that among the fundamental ingredients in our model are:

- the complete, separable metric space $\mathcal{M}$ of loci;
- a finite Borel measure ν on $\mathcal{M}$ called the "mutation measure", because it describes the rate at which mutations occur in regions of the genome;
- the space $\mathcal{G}$ of integer–valued finite Borel measures on $\mathcal{M}$ – an element of $\mathcal{G}$ represents a genotype thought of as the set of loci at which there have been ancestral mutations away from the reference wild type, so that, in particular, the null measure represents the wild genotype.

Recall also that the state of the population at time $t \geq 0$ in our model is a probability measure P_t on $\mathcal{G}$ that is the distribution of a Poisson random finite measure on $\mathcal{M}$. The distribution of such a Poisson random measure is determined by its intensity measure, which is a finite Borel measure on $\mathcal{M}$. The following notation will be useful:

Notation 2.1. Denote by $\mathcal{H}$ the space of finite signed Borel measures on $\mathcal{M}$. Let $\mathcal{H}^+$ be the subset of $\mathcal{H}$ consisting of nonnegative measures.

For $\pi \in \mathcal{H}$, write $\pi^+, \pi^- \in \mathcal{H}^+$ for the positive and negative parts of π appearing in the Hahn-Jordan decomposition. Thus, $\pi = \pi^+ - \pi^-$.

We next recall the definition of the Wasserstein metric (also transliterated as Vasershtein, or otherwise). This metric provides a unified way of topologizing the various spaces of measures that we use.

Notation 2.2. Given a metric space (E, d), let Lip be the space of functions $f : E \to \mathbb{R}$ such that

$$\|f\|_{\mathrm{Lip}} := \sup_x |f(x)| + \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)} < \infty. \tag{2.1}$$

Define a norm $\|\cdot\|_{\mathrm{Was}}$ on the space of finite signed Borel measures on (E, d) by

$$\|\pi\|_{\mathrm{Was}} := \sup \left\{ |\pi[f]| : \|f\|_{\mathrm{Lip}} \leq 1 \right\}, \tag{2.2}$$

where

$$\pi[f] := \int f d\pi. \tag{2.3}$$

An extensive account may be found in [**Rac91, RR98**], while the properties used here are described in Problem 3.11.2 of [**EK86**].

We may build Wasserstein metrics on top of Wasserstein metrics. We start by taking $\mathcal{M}$ for the metric space X. Then $\mathcal{G}$, as a space of measures on $\mathcal{M}$, has a Wasserstein metric. Next we take $\mathcal{G}$ with its Wasserstein metric for the metric space X, and obtain a Wasserstein metric on the finite signed measures on $\mathcal{G}$, including the measures P_t. Also, with $\mathcal{M}$ again playing the role of X, we obtain a Wasserstein metric on $\mathcal{H}$.

We note that the designation "Wasserstein metric" is often applied to the analogous definition where the constraining Lipschitz norm $\|\cdot\|_{\text{Lip}}$ does not include the supremum norm term. This latter distance is always greater than or equal to the Wasserstein metric as we are defining it. It is equivalent to the Kantorovich-Rubinstein distance, which is a member of the class of Monge-Kantorovich distances, a class defined by a single parameter p; the Kantorovich-Rubinstein distance is the element of this class corresponding to $p = 1$. Details may be found in [**Vil03**, Section 7.1], [**Vil09**, Chapter 6] or [**AGS05**, Section 7.1].

2.2. Definition of the dynamical system

Recall that the state of the population at time $t \geq 0$ is given by a probability measure P_t on $\mathcal{G}$ that is the distribution of a Poisson random measure with intensity measure ρ_t. Our informal description of the evolution of ρ_t, and hence of P_t, is (see equation (1.5))

$$\rho_t(dm) = \rho_0(dm) + t\,\nu(dm) - \int_0^t \mathbb{E}\left[S(X^{\rho_s} + \delta_m) - S(X^{\rho_s})\right] \rho_s(dm)\,ds, \tag{2.4}$$

where we write X^π for a Poisson random measure on $\mathcal{M}$ with intensity measure $\pi \in H^+$.

In order to formalize this definition, it will be convenient to introduce the following two objects.

Definition 2.3.

- Define $F : \mathcal{M} \times \mathcal{H}^+ \to \mathbb{R}_+$ by

$$F_\pi(x) := \mathbb{E}\big[S(X^\pi + \delta_x) - S(X^\pi)\big] \text{ for } x \in \mathcal{M} \text{ and } \pi \in \mathcal{H}^+. \tag{2.5}$$

- Define the operator $D : \mathcal{H}^+ \to \mathcal{H}^+$ by

$$\frac{d(D\pi)}{d\pi}(x) := F_\pi(x). \tag{2.6}$$

That is, for any bounded $f : \mathcal{M} \to \mathbb{R}$,

$$\int_{\mathcal{M}} f(x)\,(D\pi)(dx) = \int_{\mathcal{M}} f(x) F_\pi(x)\,\pi(dx). \tag{2.7}$$

Formally, a solution to (1.5) is an $\mathcal{H}^+$-valued function ρ that is continuous (with respect to the topology of weak converge of measures) and satisfies

$$\rho_t = \rho_0 + t\nu - \int_0^t D\rho_s\,ds \tag{2.8}$$

for all $t \geq 0$.

Equation (2.8) involves the integration of a measure-valued function, and such an integral can have a number of different meanings. We require only the following notion: If $\eta : \mathbb{R}_+ \to \mathcal{H}$ is a a Borel function, then for $t \geq 0$ the integral $\mathcal{I}_t = \int_0^t \eta_s \, ds$ is the element of $\mathcal{H}$ satisfying

$$\mathcal{I}_t(A) = \int_0^t \eta_s(A) \, ds \text{ for every Borel } A \subseteq \mathcal{M}. \tag{2.9}$$

This integral certainly exists (and is unique) if the function η is continuous with respect to the topology of weak convergence of measures. For more information about integration on infinite dimensional spaces, see chapter 2 of [**DU77**].

2.3. Existence and uniqueness of solutions

Theorem 2.4. *Fix a mutation measure $\nu \in \mathcal{H}^+$ and a selective cost $S : \mathcal{G} \to \mathbb{R}_+$, that satisfies the conditions*

- $S(0) = 0$,
- $S(g) \leq S(g+h)$ *for all* $g, h \in \mathcal{G}$,
- *for some constant* K, $\big|S(g) - S(h)\big| \leq K \big\|g - h\big\|_{\mathrm{Was}}$, *for all* $g, h \in \mathcal{G}$.

Then, equation (2.8) *has a unique solution for any* $\rho_0 \in \mathcal{H}^+$.

Proof. Fix a time horizon $T > 0$ and let $c > 0$ be a constant that will be chosen later. Write $C([0,T], \mathcal{H})$ for the Banach space of continuous $\mathcal{H}$-valued functions on $[0,T]$, equipped with the norm

$$\|\alpha\|_c = \sup_{0 \leq t \leq T} e^{-ct} \|\alpha_t\|_{\mathrm{Was}}.$$

Denote by Γ the closed subset of $C([0,T], \mathcal{H})$ consisting of $\mathcal{H}$-valued functions α with $\alpha_0 = \rho_0$ and

$$\alpha_t^+(\mathcal{M}) \leq \rho_0(\mathcal{M}) + t\nu(\mathcal{M})$$

for $0 \leq t \leq T$ (recall from Notation 2.1 that the measure α_t^+ is the positive part in the Hahn-Jordan decomposition of the signed measure α_t).

Define a map $\Delta : \mathcal{H} \to \mathcal{H}$ by

$$(\Delta\alpha)_t = \rho_0 - \int_0^t D\alpha_s^+ ds + t\nu.$$

Note that Δ maps Γ into itself. Moreover, for $\alpha, \beta \in \Gamma$,

$$\begin{aligned}
\|\Delta\alpha - \Delta\beta\|_c &\leq \sup_{0 \leq t \leq T} e^{-ct} \int_0^t \big\|\Delta\alpha_s^+ - \Delta\beta_s^+\big\|_{\mathrm{Was}} \, ds \\
&\leq \sup_{0 \leq t \leq T} e^{-ct} \int_0^t K\big(2 + 6\big\{\alpha_s^+(\mathcal{M}) \wedge \beta_s^+(\mathcal{M})\big\}\big) \big\|\alpha_s - \beta_s\big\|_{\mathrm{Was}} \, ds \\
&\qquad \text{by Lemma 2.9 below} \\
&\leq \sup_{0 \leq t \leq T} e^{-ct} \int_0^t K\big(2 + 6\big\{\rho_0(\mathcal{M}) + s\nu(\mathcal{M})\big\}\big) e^{cs} \big\|\alpha - \beta\big\|_c \, ds \\
&\leq \sup_{0 \leq t \leq T} e^{-ct} \Big[K(2 + 6\rho_0(\mathcal{M})) \frac{e^{ct} - 1}{c} + 6K\nu(\mathcal{M}) \frac{(ct-1)e^{ct} + 1}{c^2}\Big] \|\alpha - \beta\|_c.
\end{aligned}$$

Thus, $\Delta : \Gamma \to \Gamma$ is a contraction, provided c is chosen sufficiently large.

It follows from the Contraction Mapping Theorem that the equation

$$\rho_t = \Delta\rho_t = \rho_0 - \int_0^t D\rho_s^+ \, ds + t\nu \tag{2.10}$$

has a unique solution in Γ. Furthermore, any function in $\mathcal{H}$ that is a solution to (2.10) must automatically be in Γ. Therefore, the solution is unique.

This unique solution is a function taking values in the set of signed measures $\mathcal{H}$, and it remains to show that it actually takes values in the subset $\mathcal{H}^+$ of nonnegative measures. For any Borel set $A \subseteq \mathcal{M}$,

$$\rho_t(A) = \rho_0(A) - \int_0^t \int_A F_{\rho_s^+}(x)\, d\rho_s^+(x)\, ds + t\nu(A).$$

In particular, $t \mapsto \rho_t(A)$ is continuous. For any Borel set $A \subset \mathcal{M}$ we have

$$\rho_t^+(A) \le \rho_0(A) + t\nu(A).$$

Since $F_\pi(x) \le K$ for all x,

$$\begin{aligned}\rho_t(A) &\ge \rho_0(A) - K\int_0^t \big(\rho_0(A) + s\nu(A)\big)\, ds + t\nu(A)\\ &= (1-Kt)\rho_0(A) + t\left(1 - \frac{Kt}{2}\right)\nu(A).\end{aligned}$$

Hence, $\rho_t(A) \ge 0$ for $0 \le t \le 1/K$. Because this is true for all A, we have $\rho_t \in \mathcal{H}^+$ for $0 \le t \le 1/K$. Iterating this argument, with the time 0 replaced successively by the times $1/K, 2/K, \ldots$ gives the result. □

2.4. Lemmas used in the proof of existence and uniqueness

We assume in this section that the hypotheses of Theorem 2.4 hold.

LEMMA 2.5. *The function F_π is in* Lip *for each finite measure $\pi \in \mathcal{H}^+$, and*

$$\sup_{\pi \in \mathcal{H}^+} \|F_\pi(\cdot)\|_{\mathrm{Lip}} \le 2K.$$

PROOF. By definition,

$$\begin{aligned}\|F_\pi(\cdot)\|_{\mathrm{Lip}} &= \sup_x |\mathbb{E}[S(X^\pi + \delta_x) - S(X^\pi)]|\\ &\quad + \sup_{x\ne y} |\mathbb{E}[S(X^\pi + \delta_x) - S(X^\pi)] - \mathbb{E}[S(X^\pi + \delta_y) - S(X^\pi)]| / d(x,y)\\ &\le \sup_x K\|\delta_x\|_{\mathrm{Was}} + \sup_{x\ne y} K\|\delta_x - \delta_y\|_{\mathrm{Was}}/d(x,y)\\ &\le 2K.\end{aligned}$$

□

NOTATION 2.6. Given a measure $q \in \mathcal{H}^+$, define Π_q to be the distribution of the Poisson random measure with intensity q. Thus, if P is the distribution of a Poisson random finite measure on $\mathcal{M}$ (that is, a Poisson distributed random element of $\mathcal{G}$), then $P = \Pi_{\mu P}$, where $\mu P \in \mathcal{H}^+$ is the intensity measure associated with P (that is, $\mu P[F] = \int_{\mathcal{G}} g[F]\, dP(g)$).

The next result probably already exists in some form in the literature, but we have been unable to find a reference. It says that if the intensities of two Poisson random measures are close in the Wasserstein sense, then the same is true of their distributions. In the statement of the result, the Wasserstein distance on the left-hand side of the inequality is between probability measures on the space $\mathcal{G}$, while the Wasserstein distance on the right-hand side is between finite measures on $\mathcal{M}$.

Lemma 2.7. *For two finite measures $\pi', \pi'' \in \mathcal{H}^+$*

$$\big\|\Pi_{\pi'} - \Pi_{\pi''}\big\|_{\mathrm{Was}} \leq 4\big\|\pi' - \pi''\big\|_{\mathrm{Was}}.$$

Proof. Fix $f : \mathcal{G} \to \mathrm{Lip}$ with $\|f\|_{\mathrm{Lip}} \leq 1$. Fix $\pi', \pi'' \in \mathcal{H}^+$. If $\pi' = \pi'' = 0$ there is nothing to prove. Therefore, suppose without loss of generality that $\pi' \neq 0$ and $\pi'(\mathcal{M}) \geq \pi''(\mathcal{M})$. Set $\pi^* = (\pi''(\mathcal{M})/\pi'(\mathcal{M}))\pi' \in \mathcal{H}^+$. We have

$$\big|\Pi_{\pi'}[f] - \Pi_{\pi''}[f]\big| \leq \big|\Pi_{\pi'}[f] - \Pi_{\pi^*}[f]\big| + \big|\Pi_{\pi^*}[f] - \Pi_{\pi''}[f]\big|.$$

Note that if $X^{\pi'}$ and X^{π^*} are any two Poisson random measures on the same probability space with distributions $\Pi_{\pi'}$ and Π_{π^*}, respectively, then

$$\begin{aligned}\big|\Pi_{\pi'}[f] - \Pi_{\pi^*}[f]\big| &= \big|\mathbb{E}\left[f(X^{\pi'}) - f(X^{\pi^*})\right]\big| \\ &\leq 2\mathbb{P}\big\{X^{\pi'} \neq X^{\pi^*}\big\}\end{aligned}$$

because $|f(g)| \leq 1$ for all $g \in \mathcal{G}$. In particular, if we first build $X^{\pi'}$ and then construct X^{π^*} by the usual "thinning" procedure of independently keeping each point of $X^{\pi'}$ with probability $\pi''(\mathcal{M})/\pi'(\mathcal{M})$ and discarding it with the complementary probability, we have

$$\begin{aligned}\big|\Pi_{\pi'}[f] - \Pi_{\pi^*}[f]\big| &\leq 2\mathbb{P}\big\{X^{\pi'}(\mathcal{M}) \neq X^{\pi^*}(\mathcal{M})\big\} \\ &\leq 2\sum_{k=0}^{\infty} e^{-\pi'(\mathcal{M})}\frac{\pi'(\mathcal{M})^k}{k!}\left[1 - \left(\frac{\pi''(\mathcal{M})}{\pi'(\mathcal{M})}\right)^k\right] \\ &\leq 2\sum_{k=0}^{\infty} e^{-\pi'(\mathcal{M})}\frac{\pi'(\mathcal{M})^k}{k!}k\left[1 - \frac{\pi''(M)}{\pi'(\mathcal{M})}\right] \\ &= 2\pi'(\mathcal{M})\left[1 - \frac{\pi''(M)}{\pi'(\mathcal{M})}\right] \\ &= 2\big|\pi'(\mathcal{M}) - \pi''(\mathcal{M})\big| \\ &\leq 2\big\|\pi'(\mathcal{M}) - \pi''(\mathcal{M})\big\|_{\mathrm{Was}}.\end{aligned}$$

Setting $r = \pi^*(\mathcal{M}) = \pi''(\mathcal{M})$,

$$
\begin{aligned}
&\bigl|\Pi_{\pi^*}[f] - \Pi_{\pi''}[f]\bigr| \\
&\leq \sum_{k=0}^{\infty} \frac{e^{-r}}{k!} \Biggl| \int \cdots \int f\left(\sum_{\ell=1}^{k} \delta_{y_\ell}\right) \pi^*(dy_1) \cdots \pi^*(dy_k) \\
&\qquad\qquad - \int \cdots \int f\left(\sum_{\ell=1}^{k} \delta_{y_\ell}\right) \pi''(dy_1) \cdots \pi''(dy_k) \Biggr| \\
&\leq \sum_{k=0}^{\infty} \frac{e^{-r}}{k!} \sum_{m=0}^{k-1} \Biggl| \int \cdots \int f\left(\sum_{\ell=1}^{k} \delta_{y_\ell}\right) \pi^*(dy_1) \cdots \pi^*(dy_m)\pi''(dy_{m+1}) \cdots \pi''(dy_k) \\
&\qquad - \int \cdots \int f\left(\sum_{\ell=1}^{k} \delta_{y_\ell}\right) \pi^*(dy_1) \cdots \pi^*(dy_{m+1})\pi''(dy_{m+2}) \cdots \pi''(dy_k) \Biggr| \\
&\leq \sum_{k=0}^{\infty} \frac{e^{-r}}{k!} r^{k-1} k \sup_{g \in \mathcal{G}} \left| \int f(g + \delta_z)\pi^*(dz) - \int f(g + \delta_z)\pi''(dz) \right| \\
&\leq \|\pi^* - \pi''\|_{\mathrm{Was}} \\
&\leq \|\pi' - \pi''\|_{\mathrm{Was}} + \bigl|\pi'(\mathcal{M}) - \pi''(\mathcal{M})\bigr| \\
&\leq 2\|\pi^* - \pi''\|_{\mathrm{Was}}.
\end{aligned}
$$

Putting these bounds together yields

$$
\bigl|\Pi_{\pi'}[f] - \Pi_{\pi''}[f]\bigr| \leq 4\|\pi' - \pi''\|_{\mathrm{Was}},
$$

as required. □

Lemma 2.8. *For two finite measures $\pi', \pi'' \in \mathcal{H}^+$,*

$$
\sup_{x \in \mathcal{M}} |F_{\pi'}(x) - F_{\pi''}(x)| \leq 8K\|\pi' - \pi''\|_{\mathrm{Was}}.
$$

Proof. Fix $x \in \mathcal{M}$. Define $G : \mathcal{G} \to \mathbb{R}_+$ by $G(g) := (S(g + \delta_x) - S(g))/2K$. Then, $\|G\|_{\mathrm{Lip}} \leq 1$ by the Lipschitz assumption on S. Note that $F_\pi(x) = 2K\Pi_\pi[G]$. By definition of the Wasserstein distance,

$$
|F_{\pi'}(x) - F_{\pi''}(x)| = 2K\bigl|\Pi_{\pi'}[G] - \Pi_{\pi''}[G]\bigr| \leq 2K\bigl\|\Pi_{\pi'}[G] - \Pi_{\pi''}[G]\bigr\|_{\mathrm{Was}}.
$$

The lemma now follows from Lemma 2.7. □

Lemma 2.9. *For two finite signed measures $\sigma, \tau \in \mathcal{H}$,*

$$
\|D\sigma^+ - D\tau^+\|_{\mathrm{Was}} \leq K(2 + 8\{\sigma^+(\mathcal{M}) \wedge \tau^+(\mathcal{M})\})\,\|\sigma - \tau\|_{\mathrm{Was}}.
$$

Proof. Suppose without loss of generality that $\sigma^+(\mathcal{M}) \le \tau^+(\mathcal{M})$. By Lemmas 2.5 and 2.8,

$$\begin{aligned}
&\Big|\int f(x)\, D\sigma^+(dx) - \int f(x)\, D\tau^+(dx)\Big| \\
&\qquad\qquad\le \left|\int F(x,\sigma^+)f(x)\,\sigma^+(dx) - \int F(x,\tau^+)f(x)\,\sigma^+(dx)\right| \\
&\qquad\qquad\qquad + \left|\int F(x,\tau^+)f(x)\,\sigma^+(dx) - \int F(x,\tau^+)f(x)\,\tau^+(dx)\right| \\
&\qquad\qquad\le 8K\|\sigma^+ - \tau^+\|_{\mathrm{Was}}\|f\|_\infty \sigma^+(\mathcal{M}) + 2K\|f\|_{\mathrm{Lip}}\|\sigma^+ - \tau^+\|_{\mathrm{Was}} \\
&\qquad\qquad\le (2K + 8K\sigma^+(\mathcal{M}))\|f\|_{\mathrm{Lip}}\|\sigma - \tau\|_{\mathrm{Was}},
\end{aligned}$$

where we have used the fact that $\|f'f''\|_{\mathrm{Lip}} \le \|f'\|_{\mathrm{Lip}}\|f''\|_{\mathrm{Lip}}$ for $f', f'' \in \mathrm{Lip}$. $\square$

CHAPTER 3

Equilibria

One of the primary problems concerning our dynamical systems is to understand their asymptotic behavior. We begin the analysis of these asymptotics by identifying when fixed points of the motion exist, and then examining whether there is convergence to these fixed points from suitable initial conditions.

We assume throughout this chapter that the assumptions of Theorem 2.4 always hold.

Definition 3.1. A finite *fixed point* or *equilibrium* for the motion (the terms will be used interchangeably) is a measure $\rho_* \in \mathcal{H}_+$ at which the driving vector field vanishes. That is, ρ_* is absolutely continuous with respect to ν, with Radon-Nikodym derivative satisfying

$$F_{\rho_*} \frac{d\rho_*}{d\nu} = 1. \tag{3.1}$$

The equilibrium ρ_* is called *stable* if for every neighborhood V of ρ_* there is a neighborhood $U \subset V$, such that $\rho_t \in V$ for all times t if $\rho_0 \in U$. It is called *attractive* if it is stable and there is a neighborhood U_0 of ρ_* such that $\lim_{t\to\infty} \rho_t = \rho_*$ whenever $\rho_0 \in U_0$.

We introduce the terms *box-stable* and *box-attractive* when the above definitions hold if "neighborhoods" in the above definitions are replaced by sets of the form

$$B(\tilde\rho, \tilde\rho') := \{\rho : \tilde\rho < \rho < \tilde\rho'\}, \tag{3.2}$$

where $\tilde\rho \le \rho_* \le \tilde\rho'$, and both measures $\tilde\rho - \rho_*$ and $\tilde\rho' - \rho_*$ are mutually absolutely continuous with respect to ν.

Remark 3.2. Note that box-stability (respectively, box-attractivity) is a weaker condition than stability (respectively, attractivity), because boxes do not contain open neighborhoods in the topology induced by the Wasserstein metric (that is, the topology of weak convergence), but open neighborhoods do contain boxes.

Of course, it is not obvious that the motion needs to have fixed points, since ν could dominate all fitness costs. This possibility is easy to see in the one-dimensional case (when $\mathcal{M}$ is a single point), which we described in Section 1.2 and which we revisit in Section 3.1. However, we show in Section 3.4 that at least one fixed point exists when the mutation measure ν is small enough. In order to go further, we need to impose additional assumptions. For the case of multiplicative selection costs, Section 3.2 gives a complete description of the fixed points, of which there can be 0, 1 or 2. In the remainder of this chapter we then impose a weaker assumption, namely that the selective cost is concave, in the sense that the marginal cost of an additional mutation decreases as more mutations are added to the genotype. (The formal definition is given as the last condition in Theorem 3.11.) Under this assumption, we show in Section 3.5 that trajectories starting from 0 increase monotonically, and we give a sufficient condition for them to converge to a finite equilibrium. This equilibrium is dominated by any other equilibrium. If the driving vector field at nearby points above this minimal equilibrium are in the negative "orthant", then trajectories starting above the equilibrium will converge down to it, so the minimal equilibrium will be box attractive. We give an iterative procedure for computing the minimal equilibrium that avoids following the dynamical system to large times in Section 3.6. Moreover, the iteration is a useful theoretical tool in Section 3.7, where we establish that the equilibria shown to exist for suitably small ν in Section 3.4 are in fact the same as the minimal equilibria in the concave setting and that these equilibria are then box stable. Finally, in Section 3.8 we apply the above results to the demographic selection cost introduced in Section 1.4

These results cover many cases of substantive interest, although they do not settle all relevant questions. Cases with very small $\nu(\mathcal{M})$ do not wholly put on display the rich array of differences between the full non-linear model and non-epistatic additive models. However, the conditions of Corollary 3.14 and Theorem 3.15 can often be verified in specific cases, and we have found that they can hold when $\nu(\mathcal{M})$ is only moderately small.

The example discussed in [**WSE08**] provides numerical evidence that there can be convergence to an equilibrium even without uniformly monotone growth.

3.1. Introductory example: One-dimensional systems

Suppose as in Section 1.2 that the space of loci $\mathcal{M}$ consists of a single point. Then, the space of genotypes $\mathcal{G}$ may be identified with the natural numbers $\mathbb{N}_0$, the selective cost S is simply an increasing function from $\mathbb{N}_0$ to $\mathbb{R}_+$, and the mutation measure ν is a positive constant. The space $\mathcal{H}_+$ of finite measures measures on $\mathcal{G}$ may also be identified with $\mathbb{R}_+$. The dynamical system $(\rho_t)_{t\geq 0}$ is $\mathbb{R}_+$-valued and satisfies

$$\dot{\rho}_t = \nu - F_{\rho_t}\rho_t, \tag{3.3}$$

where

$$F_\rho \rho = e^{-\rho}\sum_{k=1}^{\infty}\frac{\rho^k(k-\rho)}{k!}S(k). \tag{3.4}$$

The system has equilibria at solutions to the equation $\rho F_\rho = \nu$ and we discussed some special cases in Section 1.2. In general, it is possible to construct selective costs for which the number of equilibria is arbitrarily large for a given mutation rate. For example, suppose the selective cost has magnitude 1 for 1 to 5 mutations and magnitude 2 for 6 or more. The resulting function $-\rho F_\rho$ is shown in Figure 1. The number of equilibria may be 0, 1, 2, 3, or 4, depending on the value of ν.

However, there are a few things that we can say quite generally about one-dimensional systems:

- As long as S is not identically 0, the same will be true for F_ρ, so there will be at least one equilibrium for ν sufficiently small.
- The smallest equilibrium will be attractive, unless it corresponds to a local minimum of $-F_\rho\rho$, in which case it will attract only trajectories coming from below; trajectories starting above the equilibrium will then be repelled.

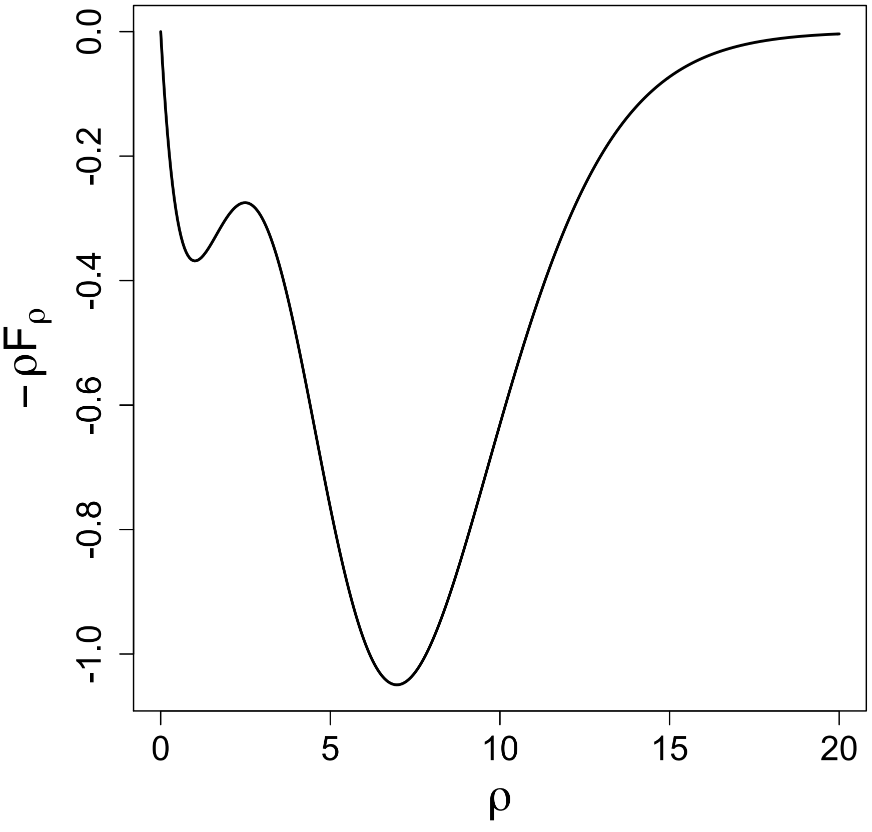

FIGURE 1. Plot of the function $-\rho F_\rho$ when $S(k) = \mathbf{1}_{\{k \geq 1\}} + \mathbf{1}_{\{k \geq 7\}}$.

- If S is bounded above, then $-F_\rho \rho$ is bounded below, and so there is no equilibrium for ν sufficiently large.

3.2. Introductory example: Multiplicative selective cost

It is probably apparent to the reader by now that one cannot hope to get anything like a closed-form solution for the dynamical system (2.8) for general selective costs. However, we first show in this section that it is possible to solve (2.8) explicitly when the selective cost has the form

$$S(g) = 1 - \exp\left(-\int_{\mathcal{M}} \sigma(m)\, dg(m)\right) \tag{3.5}$$

for some $\sigma : \mathcal{M} \to \mathbb{R}_+$. Moreover, it is possible to analyze in this case whether $\lim_{t\to\infty} \rho_t$ exists, and we will couple this analysis with a comparison argument in later sections to give sufficient conditions for the existence of a limit for more general selective costs.

Suppose that ρ_0 is absolutely continuous with respect to ν. It follows from (2.8) that ρ_t will have a density ϕ_t against ν, for all $t \geq 0$. These densities will solve the equation

$$\frac{d\phi_t(m)}{dt} = 1 - \mathbb{E}[S(X^{\rho_t} + \delta_m) - S(X^{\rho_t})]\,\phi_t(m).$$

Put $b_k(t) := \int \exp(-k\sigma(m))\phi_t(m)\,d\nu(m)$ for $k \in \mathbb{N}$, then

$$\frac{d\phi_t(m)}{dt} = 1 - (1 - \exp(-\sigma(m)))\exp(b_1(t) - b_0(t))\,\phi_t(m). \tag{3.6}$$

This is an ordinary differential equation for each $m \in \mathcal{M}$ that has the solution

$$\begin{aligned}\phi_t(m) = \exp&\left(-\int_0^t (1 - \exp(-\sigma(m)))\exp(b_1(s) - b_0(s))\,ds\right) \\ &\times \left[\phi_0(m) + \int_0^t \exp\left(-\int_0^s (1 - \exp(-\sigma(m)))\exp(b_1(r) - b_0(r))\,dr\right)ds\right].\end{aligned}$$

Thus, we have reduced the (in general) infinite collection of coupled ordinary differential equations to the problem of finding two functions, b_0 and b_1.

Of course, this will benefit us only if we have an autonomous system of equations in just the functions b_0 and b_1. To this end, set $a_k := \int \exp(-k\sigma(m))\,d\nu(m)$ for $k \in \mathbb{N}$. Then, from (3.6) we have

$$\frac{db_k(t)}{dt} = a_k + \exp(b_1(t) - b_0(t))(b_{k+1}(t) - b_k(t)). \tag{3.7}$$

Introduce the generating functions

$$A(z) := \sum_{k=0}^{\infty} a_k \frac{z^k}{k!}$$

and

$$B(z,t) := \sum_{k=0}^{\infty} b_k(t) \frac{z^k}{k!}.$$

The system of ordinary differential equations (3.7) then becomes the first order, linear partial differential equation

$$\frac{\partial B(z,t)}{\partial t} = A(z) + \exp(b_1(t) - b_0(t))\left[\frac{\partial B(z,t)}{\partial z} - B(z,t)\right].$$

We may find the general solution of this PDE via the method of characteristic curves. Once we get the general solution (which will involve the unknown functions b_0 and b_1), we have to impose the conditions

$$B(0,t) = b_0(t)$$

and

$$\frac{\partial B(0,t)}{\partial z} = b_1(t)$$

to get the solution we are after.

Using *Mathematica*, the PDE has the solution

$$\begin{aligned}B(z,t) = \int_0^t &\exp\left(-\int_u^t \exp(b_1(s) - b_0(s))\,ds\right) A\left(z + \int_u^t \exp(b_1(s) - b_0(s))\,ds\right) du \\ &+ \exp\left(-\int_0^t \exp(b_1(s) - b_0(s))\,ds\right) C\left(z + \int_0^t \exp(b_1(s) - b_0(s))\,ds\right)\end{aligned}$$

where $C(z) := \sum_{k=0}^{\infty} b_k(0)\frac{z^k}{k!}$.

Therefore, the functions b_0 and b_1 solve the system of equations

$$\begin{aligned} b_0(t) = &\int_0^t \exp\left(-\int_u^t \exp(b_1(s) - b_0(s))\, ds\right) A\left(\int_u^t \exp(b_1(s) - b_0(s))\, ds\right) du \\ &+ \exp\left(-\int_0^t \exp(b_1(s) - b_0(s))\, ds\right) C\left(\int_0^t \exp(b_1(s) - b_0(s))\, ds\right) \end{aligned}$$

and

$$\begin{aligned} b_1(t) = &\int_0^t \exp\left(-\int_u^t \exp(b_1(s) - b_0(s))\, ds\right) A'\left(\int_u^t \exp(b_1(s) - b_0(s))\, ds\right) du \\ &+ \exp\left(-\int_0^t \exp(b_1(s) - b_0(s))\, ds\right) C'\left(\int_0^t \exp(b_1(s) - b_0(s))\, ds\right). \end{aligned}$$

Note that this system of ordinary differential equations it autonomous: it depends only on the unknown functions b_0 and b_1 themselves, together with the known functions A (determined by the given mutation and selection) and C (determined by the initial conditions). Hence, these equations could at least be solved numerically. (Of course, we started with a system of as many equations as there are points in $\mathcal{M}$. This reduction to a system of two coupled ordinary differential equations is only advantageous if the space $\mathcal{M}$ has more than two points.)

We can now obtain a necessary and sufficient condition for the existence of an equilibrium for (2.8) when the selection cost is the multiplicative one of (3.5). Note that if ρ_* is an equilibrium, then we have

$$\nu(dm) - \mathbb{E}[S(X^{\rho_*} + \delta_m) - S(X^{\rho_*})]\, \rho_*(dm) = 0, \tag{3.8}$$

and so ρ_* has a density ϕ_* against ν that satisfies

$$\phi_*(m) = \frac{\exp(b_0^* - b_1^*)}{1 - \exp(-\sigma(m))},$$

where $b_k^* := \int \exp(-k\sigma(m))\phi_*(m)\, d\nu(m)$ for $k \in \{0, 1\}$ (cf. (3.6)). Therefore, an equilibrium exists if and only if

$$\int \frac{1}{1 - \exp(-\sigma(m))}\, d\nu(m) < \infty$$

and there is a constant $c > 0$ such that

$$\begin{aligned} c &= \int \frac{\exp(c)}{1 - \exp(-\sigma(m))}\, d\nu(m) - \int \exp(-\sigma(m)) \frac{\exp(c)}{1 - \exp(-\sigma(m))}\, d\nu(m) \\ &= \exp(c)\, \nu(\mathcal{M}), \end{aligned}$$

in which case $c = b_0^* - b_1^*$. Such a constant exists if and only if

$$\nu(\mathcal{M}) \le \sup_{x \ge 0} x \exp(-x) = e^{-1}.$$

Note that there are three possible cases:

- If $\nu(\mathcal{M}) < e^{-1}$, then there are two equilibria, corresponding to the two distinct solutions of $ce^{-c} = \nu(\mathcal{M})$. The equilibrium corresponding to the smaller c is attractive, while that corresponding to the larger c is unstable.
- If $\nu(\mathcal{M}) > e^{-1}$, then there is no equilibrium.
- If $\nu(\mathcal{M}) = e^{-1}$, then there is a single equilibrium, which is unstable.

3.3. Fréchet derivatives

We need to introduce some machinery on derivatives of curves and vector fields in order to analyze equilibria in more generality. Consider a Banach space $(X, \|\cdot\|_X)$ and a closed convex cone $X_+ \subseteq X$. For $x \in X_+$, let U_x be the closed convex cone $\{t(x'-x) : t \geq 0,\, x' \in X_+\}$. Consider another Banach spaces.$(Y, \|\cdot\|_Y)$. Extending the usual definition slightly, we say that a map $\Phi : X_+ \to Y$, is Fréchet differentiable at $x \in X_+$, if there is map $D_x\Phi : U_x \to Y$ with the properties

$$\lim_{x' \to x,\, x' \in X_+} \|x'-x\|_X^{-1} \Big(\Phi(x') - \Phi(x) - D_x\Phi[x'-x] \Big) = 0,$$

$D_x\Phi[tz] = tD_x\Phi[z]$ for $t \geq 0$ and $z \in U_x$, $D_x\Phi[z' + z''] = D_x\Phi[z'] + D_x\Phi[z'']$ for $z', z'' \in U_x$, and, for some constant C, $\|D_x[z]\|_Y \leq C\|z\|_X$ for $z \in U_x$. It is not difficult to show that if Φ is differentiable at x, then $D_x\Phi$ is uniquely defined.

As usual, we say that a curve $\psi : I \to X$, where $I \subseteq \mathbb{R}$ is an interval, is differentiable at $t \in I$ if the limit

$$\dot{\psi}_t = \lim_{t' \to t} (t'-t)^{-1}(\psi_{t'} - \psi_t)$$

exists.

For the sake of completeness, we record the following standard fundamental theorem of calculus and chain rule.

Lemma 3.3. *Consider an interval $I \subseteq \mathbb{R}$ and a Banach spaces $(X, \|\cdot\|_X)$. Suppose that a curve $\psi : I \to X$ is differentiable at every $t \in I$ and the curve $t \mapsto \dot{\psi}_t$ is continuous. Then,*

$$\psi_b - \psi_a = \int_a^b \dot{\psi}_t \, dt$$

for all $a, b \in I$ with $a < b$.

Lemma 3.4. *Consider an interval $I \subseteq \mathbb{R}$, two Banach spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, and a closed convex cone $X_+ \subseteq X$. Suppose for some $t \in I$ that the curve $\psi : I \to X_+$ is differentiable at $t \in I$ and the map $\Phi : X_+ \to Y$ is differentiable at ψ_t. Then, the curve $\Phi \circ \psi : I \to Y$ is differentiable at t with derivative $D_{\psi_t}\Phi[\dot{\psi}_t]$.*

We also have a particular analogue of the product rule. Write $C_b(\mathcal{M}, \mathbb{R})$ for the Banach space of bounded continuous functions from $\mathcal{M}$ to $\mathbb{R}$ equipped with the supremum norm.

Lemma 3.5. *Consider an interval $I \subseteq \mathbb{R}$ and two curves $\gamma : I \to \mathcal{H}_+$ and $f : I \to C_b(\mathcal{M}, \mathbb{R})$. Suppose that γ and f are differentiable at $t \in I$. Define a curve $\beta : I \to \mathcal{H}_+$ by $\beta_u := f_u \cdot \gamma_u$, $u \in I$; that is, β_u is the element of $\mathcal{H}$ that has Radon-Nikodym derivative f_u with respect to γ_u. Then, β is differentiable at t with*

$$\dot{\beta}_t = \dot{f}_t \cdot \gamma_t + f_t \cdot \dot{\gamma}_t.$$

Proof. The follows from Lemma 3.4 with $X = C_b(\mathcal{M}, \mathbb{R}) \times \mathcal{H}_+$, $Y = \mathcal{H}$, $\Phi(e, \eta) = e \cdot \eta$, and $\psi = (f, \gamma)$ upon showing that the map Φ is differentiable at any $(e, \eta) \in C_b(\mathcal{M}, \mathbb{R}) \times \mathcal{H}$ with

$$D_{e,\eta}\Phi[(e', \eta')] = e' \cdot \eta + e \cdot \eta'$$

and the curve ψ is differentiable at t with $\dot{\psi}_t = (\dot{f}_t, \dot{\gamma}_t)$. Both proofs are straightforward and we leave them to the reader. □

For $\eta \in \mathcal{H}_+$ and $m', m'' \in \mathcal{M}$, set

$$K_\eta(m', m'') := \mathbb{E}\Big[S(X^\eta + \delta_{m'} + \delta_{m''}) - S(X^\eta + \delta_{m''}) - S(X^\eta + \delta_{m'}) + S(X^\eta)\Big].$$

By our standing assumption the conditions of Theorem 2.4 are in place, the map $(\eta, m', m'') \mapsto K_\eta(m', m'')$ is bounded. Ideas similar to those behind Lemma 2.7 and Lemma 2.8 establish that this kernel gives the Fréchet derivative of the map $\eta \mapsto F_\eta(\cdot)$.

Lemma 3.6. *The mapping $\eta \mapsto F_\eta(\cdot)$ from $\mathcal{H}_+$ to $C_b(\mathcal{M}, \mathbb{R})$ is Fréchet differentiable at every point $\eta' \in \mathcal{H}_+$ with derivative $D_{\eta'}F$ given by*

$$D_{\eta'}F[\eta''](m') = \int_{\mathcal{M}} K_{\eta'}(m', m'')\, d\eta''(m'').$$

Moreover, straightforward coupling arguments establish the following bounds, where we recall that the constant K is such that $|S(g) - S(h)| \le K\|g - h\|_{\mathrm{Was}}$ for all $g, h \in \mathcal{G}$.

Lemma 3.7. *For any $\rho, \rho', \eta \in \mathcal{H}_+$ with $\rho \le \rho'$,*

$$\big\|D_\rho F[\eta]\big\|_\infty \le 2K\eta(\mathcal{M}),$$

and

$$\big\|D_\rho F[\eta] - D_{\rho'}F[\eta]\big\|_\infty \le 16K\big(\rho'(\mathcal{M}) - \rho(\mathcal{M})\big)\eta(\mathcal{M}).$$

3.4. Existence of equilibria via perturbation

We define a family of dynamical systems for $u \in \mathbb{R}_+$ by

$$\rho_t^{(u)} = \rho_0^{(u)} + ut\nu - \int_0^t D\rho_s^{(u)} ds. \tag{3.9}$$

That is, we apply the equation (2.8) with the mutation measure ν replaced by the multiple $u\nu$.

Theorem 3.8. *Suppose the selective cost of a nonzero genotype is bounded below. That is,*

$$\inf\{S(\delta_m) : m \in \mathcal{M}\} = \inf\{S(g) : g \in \mathcal{G},\, g \ne 0\} > 0.$$

Then, there exists $U > 0$ such that there is a finite equilibrium for the equation (3.9) for all $u \in [0, U]$. That is, there exists a measure $\rho_*^{(u)} \in \mathcal{H}^+$ such that

$$u\nu = F_{\rho^{(u)}} \cdot \rho^{(u)}.$$

Proof. If it exists, $\rho_*^{(u)}$ has a density with respect to ν, which we write as $p^{(u)} \in C_b(\mathcal{M}, \mathbb{R}_+)$. The equilibrium equation then becomes

$$F_{p^{(u)}} \cdot p^{(u)} = u, \tag{3.10}$$

where we adopt the convention that F_r for a function r is F_ρ where $\rho(dm) = r(m)\nu(dm)$. The theorem is proved if we can show that there is a differentiable family of functions $p^{(u)}$ that satisfy $p^{(0)} = 0$ and (3.10) in a neighborhood of $u = 0$.

Differentiating with respect to u and applying Lemma 3.6, we get

$$\left[\int_{\mathcal{M}} K_{p^{(u)}}(m', m'')\, \frac{dp^{(u)}}{du}(m'')\, d\nu(m'')\right] p^{(u)}(m') + F_{p^{(u)}}(m') \frac{dp^{(u)}}{du}(m') = 1. \tag{3.11}$$

For each $p \in C_b(\mathcal{M}, \mathbb{R})$ we may define the bounded linear operator $T_p : C_b(\mathcal{M}, \mathbb{R}) \to C_b(\mathcal{M}, \mathbb{R})$ by

$$T_p(q) := \left[\int_{\mathcal{M}} K_{p_+}(m', m'')\, q(m'')\, d\nu(m'') \right] p_+(m') + F_{p_+}(m') q(m'), \tag{3.12}$$

where $p_+ \in C_b(\mathcal{M}, \mathbb{R}_+)$ is defined by $p_+(m) = (p(m))_+$, $m \in \mathcal{M}$. Let $\mathcal{D}$ be the subset of $p \in C_b(\mathcal{M}, \mathbb{R})$ such that T_p is invertible. It follows that $0 \in \mathcal{D}$, because $T_0(q)(m') = S(\delta_{m'})q(m')$ and $\inf\{S(\delta_m) : m \in \mathcal{M}\} > 0$ by assumption. A standard result in operator theory (see Lemma VII.6.1 of [**DS88**]) tells us that the invertible operators form an open set in the operator norm topology, so that $\mathcal{D}$ includes all p such that $\|T_p - T_0\|$ is sufficiently small, where $\|\cdot\|$ denotes the operator norm. By Lemma 2.8 we can bound $\|T_p - T_0\|$ by $10K\|p\|_\infty$, from which we conclude that $\mathcal{D}$ includes an open ball around 0.

Define a map $L : \mathcal{D} \to C_b(\mathcal{M}, \mathbb{R}_+)$ by

$$L(p) := T_p^{-1}(\mathbf{1}),$$

where $\mathbf{1} \in C_b(\mathcal{M}, \mathbb{R}_+)$ is the function with constant value 1. The map L is continuous, and

$$\begin{aligned} 0 &= T_p\big(L(p)\big) - T_0\big(L(0)\big) \\ &= \big(T_0 + (T_p - T_0)\big)\big(L(0) + (L(p) - L(0))\big) - T_0\big(L(0)\big) \\ &= (T_p - T_0)L(0) + T_0(L(p) - L(0)) + (T_p - T_0)(L(p) - L(0)). \end{aligned}$$

Thus,

$$L(p) = L(0) - T_0^{-1}\Big((T_p - T_0)L(0) - (T_p - T_0)(L(p) - L(0))\Big).$$

Since $L(0)(m)$ is bounded away from 0, it follows that there is a neighborhood V of 0 such that $0 < \inf_{p \in V} \inf_{m \in \mathcal{M}} L(p)(m)$. Hence, by standard results on existence and uniqueness of solutions to ordinary differential equations in a Banach space, the ordinary differential equation $\frac{dp^{(u)}}{du} = L(p^{(u)})$ with initial condition $p^{(0)} = 0$ has a solution on an interval $[0, U]$ and this solution takes values in $C_b(\mathcal{M}, \mathbb{R}_+)$. Thus, $(p^{(u)})_{0 \le u \le U}$ satisfies (3.11) and hence also (3.10). □

3.5. Concave selective costs

For an important class of examples — including the demographic example of Section 1.4 — the selective costs are concave, in the sense that the marginal cost of adding a given mutation becomes smaller, the more other mutations are already present. Formally, this is stated in Definition 3.9. Under a few mild constraints, we show in Theorem 3.11 that concave selective costs yield monotonic solutions $(\rho_t)_{t \ge 0}$ when started from the pure wild type population 0, and hence such systems must either diverge to a measure with infinite total mass or converge to an element of $\mathcal{H}_+$. Corollary 3.14 gives a further condition that is sufficient to ensure the limit is an element of $\mathcal{H}_+$. Our conditions for monotone increase in ρ over time turn out to be satisfied quite generally for the applications we have investigated. The conditions for the existence of a limit in $\mathcal{H}_+$, on the other hand, are not always satisfied, and there are important cases (discussed in [**WES08**]) for which ρ_t increases to a measure with infinite total mass.

Definition 3.9. A selective cost function S is *concave* if

$$S(g + h + k) - S(g + h) \le S(g + k) - S(g) \text{ for all } g, h, k \in \mathcal{G}. \tag{3.13}$$

LEMMA 3.10. *A selective cost S is concave if and only if*

$$S(g+\delta_m+\delta_{m'}) - S(g+\delta_m) \le S(g+\delta_{m'}) - S(g)$$

for all $g \in \mathcal{G}$ and $m, m' \in \mathcal{M}$.

PROOF. Since elements of $\mathcal{G}$ have finite integer mass, we can prove (3.13) by induction on $n := \max\{h(\mathcal{M}), k(\mathcal{M})\}$. Our assumption is equivalent to (3.13) in the case $n = 1$. Assume now that (3.13) holds whenever $\max\{h(\mathcal{M}), k(\mathcal{M})\} \le n - 1$. Suppose $h(\mathcal{M}) = n$ and $k(\mathcal{M}) \le n - 1$. Let m be in the support of h, and let $h = \tilde{h} + \delta_m$. Then

$$\begin{aligned} S(g{+}h+k) &- S(g+h) - S(g+k) + S(g) \\ &= S(g+k+\tilde{h}+\delta_m) - S(g+\tilde{h}+\delta_m) - S(g+k) + S(g) \\ &= \big[S(g+\delta_m+k+\tilde{h}) - S(g+\delta_m+\tilde{h}) - S(g+\delta_m+k) + S(g+\delta_m)\big] \\ &\qquad + \big[S(g+\delta_m+k) - S(g+\delta_m) - S(g+k) + S(g)\big]. \end{aligned}$$

Since $\tilde{h}$ and k both have mass smaller than n, each of the terms in brackets is ≤ 0 by the induction hypothesis. To complete the induction, we need only address the case when $h(\mathcal{M}) = k(\mathcal{M}) = n$; this proceeds exactly as above. □

THEOREM 3.11. *Fix a mutation measure $\nu \in \mathcal{H}^+$ and a selective cost $S : \mathcal{G} \to \mathbb{R}_+$, that satisfies the conditions*

- $S(0) = 0$,
- $S(g) \le S(g + h)$ *for all* $g, h \in \mathcal{G}$,
- *for some constant K,* $\big|S(g) - S(h)\big| \le K\big\|g - h\big\|_{\mathrm{Was}}$, *for all* $g, h \in \mathcal{G}$,
- $\sup_{g\in\mathcal{G}} S(g) < \infty$,
- $S(g + h + k) - S(g + h) \le S(h + k) - S(h)$ *for all* $g, h, k \in \mathcal{G}$.

If $\dot{\rho}_0 \ge 0$ (≤ 0), then the solution of equation (2.8) guaranteed by Theorem 2.4 satisfies $\rho_s \le \rho_t$ ($\rho_s \ge \rho_t$) for all $0 \le s \le t < \infty$.

PROOF. Recall that $F_\eta(m) = \mathbb{E}[S(X^\eta + \delta_m) - S(X^\eta)]$ for $\eta \in \mathcal{H}_+$ and $m \in \mathcal{M}$. It is clear from the assumptions that $\eta \mapsto F_\eta(\cdot)$ is a map from $\mathcal{H}_+$ to $C_b(\mathcal{M}, \mathbb{R})$.

The curve ρ is differentiable at each $t \ge 0$ and satisfies

$$\dot{\rho}_t = \nu - F_{\rho_t} \cdot \rho_t. \tag{3.14}$$

The right-hand side is continuous in t. By Lemma 3.3, it then suffices to show that $\dot{\rho}_t \ge 0$ for all $t \ge 0$.

By Lemma 3.4 and Lemma 3.6, the curve $t \mapsto F_{\rho_t}$, $t \in \mathbb{R}_+$, is differentiable, with

$$\frac{d}{dt}F_{\rho_t} = D_{\rho_t}F[\dot{\rho}_t],$$

and, by Lemma 3.5,

$$\ddot{\rho}_t := \frac{d}{dt}\dot{\rho}_t = -\Big(D_{\rho_t}F[\dot{\rho}_t]\Big) \cdot \rho_t - F_{\rho_t} \cdot \dot{\rho}_t.$$

Define

$$\gamma_t := -\exp\left\{\int_0^t F_{\rho_s}\,ds\right\}\dot{\rho}_t.$$

Then,

$$
\begin{aligned}
\frac{d\gamma_t}{dt} &= -\exp\left\{\int_0^t F_{\rho_s} ds\right\}\left(F_{\rho_t}\dot{\rho}_t + \ddot{\rho}_t\right)\\
&= -\exp\left\{\int_0^t F_{\rho_s} ds\right\}\Big(F_{\rho_t}\dot{\rho}_t - \Big(D_{\rho_t}F[\dot{\rho}_t]\Big)\cdot\rho_t - F_{\rho_t}\cdot\dot{\rho}_t\Big) \qquad (3.15)\\
&= \exp\left\{\int_0^t F_{\rho_s} ds\right\}\Big(-D_{\rho_t}F\left[\exp\left\{-\int_0^t F_{\rho_s} ds\right\}\cdot\gamma_t\right]\Big)\cdot\rho_t.
\end{aligned}
$$

Suppose now that $\gamma_0 = -\dot{\rho}_0 \leq 0$. For any Borel set $B \subseteq \mathcal{M}$

$$
\begin{aligned}
\gamma_t(B) &\leq \gamma_t(B) - \gamma_0(B)\\
&= \int_0^t \Big(\int_B \Big[\int_{\mathcal{M}} -K_{\rho_s}(m', m'')\exp\left\{-\int_0^s F_{\rho_r}(m'')\, dr\right\} d\gamma_s(m'')\Big]\\
&\qquad\qquad \times \exp\left\{\int_0^s F_{\rho_r}(m')\, dr\right\} d\rho_s(m')\Big)\, ds.
\end{aligned}
$$

Put $\beta_t := \sup_{B\subset\mathcal{M}} \gamma_t(B)$, where the supremum is taken over Borel sets. Because $-K$ and F are both nonnegative and bounded and since $\rho_s(\mathcal{M}) \leq s\nu(\mathcal{M})$, for any positive T there is a positive C_T such that

$$
\beta_t \leq C_T \int_0^t \beta_s ds
$$

for $0 \leq t \leq T$. By Grönwall's Inequality [**SY02**, D.1], this implies that $\beta_t \leq 0$ for all t. It follows that the measure γ_t is nonpositive. Thus, $\dot{\rho}_t$ (which differs from γ_t by a strictly negative Radon-Nikodym factor) is nonnegative.

If $\dot{\rho}_0 \leq 0$, then we define γ_t to be $+\exp\left\{\int_0^t F_{\rho_s} ds\right\}\dot{\rho}_t$, and then the rest of the proof carries through as above. ☐

Corollary 3.12. *Suppose the conditions of Theorem 3.11 hold, and there exists $\rho_{**} \in \mathcal{H}_+$ satisfying the equilibrium condition*

$$
\nu(dm) = \mathbb{E}\left[S(X^{\rho_{**}} + \delta_m) - S(X^{\rho_{**}})\right]\rho_{**}(dm)
$$

for equation (2.8). *For the process started at $\rho_0 = 0$, $\rho_t \uparrow \rho_* \in \mathcal{H}_+$, where $\rho_* \leq \rho_{**}$ and ρ_* is also an equilibrium for* (2.8).

The proof of the following comparison result is straightforward and is left to the reader.

Lemma 3.13. *Consider two selective cost functions S' and S'' that satisfy the conditions of Theorem 3.11. Let ρ' and ρ'' be the corresponding solutions of* (2.8). *Suppose that $S'(g + \delta_m) - S'(g) \leq S''(g + \delta_m) - S''(g)$ for all $g \in \mathcal{G}$ and $m \in \mathcal{M}$ and that $\rho'_0 \geq \rho''_0$. Then, $\rho'_t \geq \rho''_t$ for all $t \geq 0$.*

If the conditions of Theorem 3.11 hold, trajectories starting from 0 will either converge as time goes to infinity to an equilibrium state in $\mathcal{H}_+$ or diverge to a measure with infinite total mass. We therefore wish to consider conditions that will ensure the existence of an equilibrium with finite total mass. One approach is to compare the concave selective cost to a multiplicative selective cost. This produces the small benefit over the general existence result of Theorem 3.8 of providing an explicit value of $\nu(\mathcal{M})$ that is "small enough" to guarantee the existence of finite equilibria.

COROLLARY 3.14. *Suppose that the selective cost* $S : \mathcal{G} \to \mathbb{R}$ *satisfies the conditions of Theorem 3.11 and also satisfies the bound*

$$S(g + \delta_{m'}) - S(g) \geq \xi\,[1 - \exp(-\tau(m'))] \exp\left(-\int_{\mathcal{M}} \tau(m'')\,dg(m'')\right) \tag{3.16}$$

for all $m' \in \mathcal{M}$ *for some constant* $\xi > 0$ *and function* $\tau : \mathcal{M} \to \mathbb{R}_+$ *such that*

$$\int_{\mathcal{M}} \frac{1}{1 - \exp(-\tau(m))}\,d\nu(m) < \infty.$$

Suppose also that ρ_0 *is the null measure* 0 *and* $\nu(\mathcal{M}) \leq e^{-1}\xi$*. Then, there exists* $\rho_* \in \mathcal{H}_+$ *such that* $\rho_t \uparrow \rho_*$ *as* $t \to \infty$*.*

On the other hand, if the reverse inequality

$$S(g + \delta_{m'}) - S(g) \leq \xi\,[1 - \exp(-\tau(m'))] \exp\left(-\int_{\mathcal{M}} \tau(m'')\,dg(m'')\right) \tag{3.17}$$

holds and $\nu(\mathcal{M}) > e^{-1}\xi$*, then* $\lim_{t\to\infty} \rho_t(\mathcal{M}) = \infty$*.*

PROOF. Suppose first that (3.16) holds. Let ρ'' be the solution of 2.8 with selective cost $S''(g) = \xi \int_{\mathcal{M}} (1 - \exp(-\tau(m))\,dg(m)$ and initial condition $\rho''_0 = \rho_{**}$, where

$$\rho_{**}(dm) = \frac{\exp(c)}{1 - \exp(-\tau(m))}\,\nu(dm)$$

with $c\xi = \exp(c)\nu(\mathcal{M})$ (such a c exists by the assumption that $\nu(\mathcal{M}) \leq e^{-1}\xi$). It follows from the results of Section 3.2 that $\rho''_t = \rho_{**}$ for all $t \geq 0$.

Apply Lemma 3.13 with $S' = S$ and and $\rho'_0 = 0$ to conclude that $\rho_t \leq \rho_{**}$ for all $t \geq 0$. It follows from Theorem 3.11 that $\rho_t \uparrow \rho_*$ as $t \to \infty$ for some $\rho_* \in \mathcal{H}_+$ with $\rho_* \leq \rho_{**}$.

Now suppose instead that the upper bound (3.17) holds. We define ρ'' as before, with selective cost S'', but with initial condition $\rho''_0 = 0$. Lemma 3.13 implies then that $\rho_t \geq \rho''_t$ for all $t \geq 0$. We know that ρ''_t is increasing in t, and there is no finite equilibrium. Suppose $R := \lim_{t\to\infty} \rho_t(\mathcal{M}) < \infty$. Then for any Borel set A, the quantity $\rho''_t(A)$ is increasing in t and bounded by R, so it converges to a limit $\rho''_*(A)$. It is easy to check that $A \mapsto \rho''_*(A)$ is a measure $\mathcal{H}_+$ and that ρ''_t converges to ρ''_* in the Wasserstein metric (that is, in the topology of weak convergence). From equation (2.8) we know that for any Borel set A, and any $s < t$,

$$0 \leq \rho''_t(A) - \rho''_s(A) = \int_s^t \big(\nu(A) - D''\rho''_u(A)\big)\,du,$$

where the operator D'' is the analogue of the operator D when the selective cost S is replaced by the selective cost S''. We conclude that $\nu(A) - D''\rho''_u(A) \geq 0$ for all $u \geq 0$, and so

$$\big|\rho''_*(A) - \rho''_t(A)\big| = \left|\int_t^\infty \big(\nu(A) - D''\rho''_u(A)\big)\,du\right| \downarrow 0 \text{ as } t \to \infty.$$

Since $D''\rho$ is continuous in ρ by Lemma 2.5, it follows that the integrand on the right-hand side converges to $\nu(A) - D''\rho''_*(A)$, which must then be 0. Since this is true for all Borel sets A, it would follow that ρ''_* would be an equilibrium for the dynamical system with selective cost S'', contradicting the fact that no such equilibrium exists. □

The monotone growth of systems with concave fitness cost functions allows us to derive a simple sufficient condition for stability of the "minimum equilibrium" ρ_* — the fixed point to which the process converges when started from 0. We know that all trajectories that start strictly below ρ_* converge asymptotically to ρ_*. Since the vector field vanishes at ρ_*, this implies that the derivative of the vector field $-D_{\rho_*}F[\eta]\cdot\rho_* - F_{\rho_*}\cdot\eta$ is nonpositive for all positive directions η. The equilibrium is box stable if this nonpositivity extends to a neighborhood of ρ_*, which can be guaranteed if the derivative is actually bounded away from 0, measured by its Radon-Nikodym derivative with respect to ν.

Theorem 3.15. *(i) Suppose the selective cost function satisfies the conditions of Theorem 3.11, and the curve $(\rho_t)_{t\geq 0}$, started at $\rho_0 = 0$, converges to a finite fixed point ρ_*. Suppose further that*

$$k := \inf_{m\in\mathcal{M}} D_{\rho_*}F[\nu](m)\frac{d\rho_*}{d\nu}(m) + F_{\rho_*}(m) > 0.$$

Then the fixed point is box stable.

(ii) Moreover, if $\mathcal{M}$ is compact and the equation

$$D_{\rho_*}F[\eta]\cdot\rho_* + F_{\rho_*}\cdot\eta = 0$$

has no solution $\eta \in \mathcal{H}_+$ that is absolutely continuous with respect to ν, then ρ_* is box attractive.

Proof. Consider part (i). Let $\psi_s := \rho_* + s\nu$, and let $\phi_s = \nu - F_{\psi_s}\cdot\psi_s$, be the driving vector field evaluated at the point ψ_s. Applying Lemma 3.7, we see that

$$\begin{aligned}\inf_{m\in\mathcal{M}}&\Big(D_\rho F[\nu](m)\frac{d\rho}{d\nu}(m) + F_\rho(m)\Big)\\ &> k - 10Ks\nu(\mathcal{M}) - 16Ks\nu(\mathcal{M})^2\left(\sup_{m'\in\mathcal{M}}\frac{d\rho_*}{d\nu}(m) + s\right).\end{aligned}$$

For positive s small enough this bound is positive. Then, $\phi_0 = 0$ and

$$\begin{aligned}\phi_s &= \int_0^s \dot\phi_r dr\\ &= \int_0^s \big(-D_{\phi_r}F[\nu] - F_{\phi_r}\cdot\nu\big)dr\\ &< 0 \text{ for } s \text{ sufficiently small.}\end{aligned}$$

Applying Theorem 3.11, we see that when $\rho_0 = \psi_s$, the trajectory ρ_t is monotonically decreasing. Thus, for all times $t \geq 0$ we have $\rho_t \in B(\rho_*, \psi_s)$.

Now consider part (ii). Suppose that there is no $\eta \in \mathcal{H}_+$ that is absolutely continuous with respect to ν and satisfies $D_{\rho_*}F[\eta]\cdot\rho_* + F_{\rho_*}\cdot\eta = 0$. For all s sufficiently small, the trajectory starting from ψ_s is monotonically decreasing, and converges to an equilibrium $\rho_*^{(s)}$. If ρ_* were not box attractive, then these equilibria would be distinct from $\rho_* = \rho_*^{(0)}$. Since the $\rho_*^{(s)}$ are all between ρ_* and $\rho_* + s\nu$, we know that $\rho_*^{(s)} - \rho_*$ is absolutely continuous with respect to ν. Consider the measures $\pi^{(s)}$, defined by

$$\pi^{(s)} = \frac{\rho_*^{(s)} - \rho_*}{\rho_*^{(s)}(\mathcal{M}) - \rho_*(\mathcal{M})}.$$

These are probability measures on $\mathcal{M}$. Since $\mathcal{M}$ is compact, the space of probability measures on $\mathcal{M}$ is also compact (recall that the Wasserstein norm induces the topology of weak convergence), so there is an accumulation point. Let $s_1, s_2, \dots$ be a sequence converging to u, with $\pi^{(s_i)} \to \eta$. Since the vector field vanishes at all of these points, the derivative in direction η vanishes as well, contradicting our assumption, and hence proving that ρ_* is, in fact, box attractive. $\square$

3.6. Iterative computation of the minimal equilibrium

The measure $\pi \in \mathcal{H}_+$ is an equilibrium for the dynamic equation (2.8) if $F_\pi \cdot \pi = \nu$, where we recall that $F_\pi \cdot \pi$ is the measure that has Radon-Nikodym derivative F_π with respect to π. This implies that π is absolutely continuous with respect to ν with a Radon-Nikodym derivative p that satisfies

$$F_p(m)p(m) = 1, \quad m \in M, \tag{3.18}$$

where, with a slight abuse of notation, we write F_q for F_κ when $\kappa \in \mathcal{H}_+$ is a measure that has Radon-Nikodym derivative q with respect to ν. Under the conditions of Theorem 3.11, if (3.18) has a solution, then it has a minimal solution that arises as the Radon-Nikodym derivative with respect to ν of $\lim_{t\to\infty} \rho_t$ when ρ is the solution of (2.8) with $\rho_0 = 0$.

If one is only interested in finding the minimal equilibrium numerically, then it would be desirable to be able to do so without having to solve (2.8). An obvious approach to that problem is to define a sequence of functions $p_n : \mathcal{M} \to \mathbb{R}_+$ inductively by $p_0 = 0$ and

$$p_{n+1} = \frac{1}{F_{p_n}}, \quad n \geq 0.$$

It is clear that if S is concave, then $F_{p''} \leq F_{p'}$ for $p' \leq p''$. Because $p_0 = 0 \leq p_1$, it follows that $p_0 \leq p_1 \leq \dots$. Moreover, if p_{**} is a solution of (3.18) such that $\int_{\mathcal{M}} p_{**}(m)\, \nu(dm) < \infty$, then $p_n \uparrow p_* \leq p_{**}$ as $n \to \infty$ for some function $p_* : \mathcal{M} \to \mathbb{R}_+$ and the measure $p_* \cdot \nu$ is the minimal equilibrium of the dynamical equation (2.8).

Remark 3.16. If $\pi \in \mathcal{H}_+$ is an equilibrium for (2.8) with Radon-Nikodym derivative p against ν, then, from (3.18),

$$\begin{aligned}\pi(\mathcal{M}) &= \int_{\mathcal{M}} p(m)\, d\nu(m) = \int_{\mathcal{M}} \frac{1}{F_p(m)}\, d\nu(m) \\ &\geq \int_{\mathcal{M}} \frac{1}{\sup_{g\in\mathcal{G}} S(g+\delta_m) - S(g)}\, d\nu(m) = \int_{\mathcal{M}} \frac{1}{S(\delta_m)}\, d\nu(m)\end{aligned}$$

under the concavity assumption. Thus, a necessary condition for the existence of an equilibrium in $\mathcal{H}_+$ is that the last integral is finite.

3.7. Stable equilibria in the concave setting via perturbation

Suppose that the selective cost is concave. In Section 3.4 we constructed equilibria for sufficiently small mutation measures. We know from Section 3.5 that all trajectories starting below an equilibrium $\tilde{\rho}$ converge asymptotically to an equilibrium that is also dominated by $\tilde{\rho}$. This leaves open the question of whether the equilibrium constructed by perturbing the dynamical system away from $\nu \equiv 0$ is the same as the minimal equilibrium ρ_* to which the system converges when started from 0.

THEOREM 3.17. *(i) Suppose the selective cost function satisfies the conditions of Theorem 3.11. For* $U > 0$ *sufficiently small, there is a unique* $p : [0, U] \to C_b(\mathcal{M}, \mathbb{R}_+)$ *solving the equation*

$$\left[\int_{\mathcal{M}} K_{p^{(u)}}(m', m'')\, \dot{p}^{(u)}(m'')\, d\nu(m'')\right] p^{(u)}(m') + F_{p^{(u)}}(m')\dot{p}^{(u)}(m') = 1,$$

with $p^{(0)} \equiv 0$. *The measure* $p^{(u)}\nu \in \mathcal{H}_+$ *is the minimal equilibrium for the system with mutation measure* $u\nu$ *for all* $u \in [0, U]$. *Furthermore, the minimal equilibria so realized for* $u < U$ *are box stable.*

(ii) Moreover, if $\mathcal{M}$ *is compact and the equation*

$$D_{\rho_*}F[\eta] \cdot \rho_* + F_{\rho_*} \cdot \eta = 0$$

has no solution $\eta \in \mathcal{H}_+$ *with* η *absolutely continuous with respect to* ν, *then* ρ_* *is box attractive.*

PROOF. Following the method of Section 3.6, we may compute the minimal equilibrium as the limit of the iteration

$$p_{n+1} = \frac{u}{F_{p_n}},$$

with $p_0 \equiv 0$. Let q_n be the corresponding iterates for $v\nu$, where $0 \le u < v$.

The concavity of S implies that $p_n \le q_n$ for all n and

$$\begin{aligned} q_{n+1} - p_{n+1} &= \frac{v}{F_{q_n}} - \frac{u}{F_{p_n}} \\ &= \frac{vF_{p_n} - uF_{q_n}}{F_{p_n}F_{q_n}} \\ &= \frac{(v-u)F_{p_n}}{F_{p_n}F_{q_n}} + \frac{u(F_{p_n} - F_{q_n})}{F_{p_n}F_{q_n}} \\ &\le a|v-u| + b\|q_n - p_n\|_\infty \end{aligned}$$

for appropriate constants a and b, if p_n and q_n both converge to finite equilibria, by Lemma 2.8.

Thus,

$$\|q_{n+1} - p_{n+1}\|_\infty \le a|v-u| + b\|q_n - p_n\|_\infty.$$

Moreover, we can make the constant b less than 1 by taking u and v sufficiently small. Iterating, we see that

$$\|q_{n+1} - p_{n+1}\|_\infty \le c|v-u| + b^n\|q_0 - p_0\|_\infty,$$

where $c = a/(1-b)$. Hence, the corresponding minimal equilibria, say p_* and q_*, satisfy

$$\|q_* - p_*\|_\infty \le c|v-u|.$$

Therefore, if we write $u \mapsto p^{(u)}$ for the curve of densities of minimal equilibria, then

$$0 \le p^{(v)}(m) - p^{(u)}(m) \le c(v-u)$$

for all $m \in \mathcal{M}$ for all $0 \le u < v$ sufficiently small. It follows that $u \mapsto p^{(u)}(m)$ is Lebesgue almost everywhere differentiable, and is the integral of its derivative. Since $p^{(u)}$ satisfies the relation $p^{(u)}F_{p^{(u)}} = u$ for every u, we see by differentiating with respect to u that $p^{(u)}$ satisfies the differential equation in the statement. By standard uniqueness results for ordinary differential equations, it is the unique solution.

Fix any $u \in [0, U)$. The dynamical system with mutation measure $U\nu$ converges to a finite equilibrium $p^{(U)}\nu \in \mathcal{H}_+$, where $p^{(U)} \geq p^{(u)}$. Let $(\rho_t)_{t\geq 0}$ be the dynamical system with mutation measure $U\nu$ started from $\rho_0 = p^{(u)} \cdot \nu$ and let $(\rho'_t)_{t\geq 0}$ be the reverse – the dynamical system with mutation measure $u\nu$ started from $\rho'_0 = p^{(U)} \cdot \nu$. Thus, $(\rho_t)_{t\geq 0}$ starts below its equilibrium, and $(\rho'_t)_{t\geq 0}$ starts above its equilibrium. We have

$$\begin{aligned} \dot{\rho}'_0 &= u\nu - F_{p^{(U)}} p^{(U)} \cdot \nu = (u - U)\nu \leq 0, \\ \dot{\rho}_0 &= u\nu - F_{p^{(u)}} p^{(u)} \cdot \nu = (U - u)\nu \geq 0. \end{aligned}$$

Therefore, $p^{(U)} > p^{(u)}$, and we can conclude from Theorem 3.11 that the measure ρ_t stays bounded between the measures $p^{(u)} \cdot \nu$ and $p^{(U)} \cdot \nu$ for all times t, and hence the minimal equilibrium $p^{(u)} \cdot \nu$ is box stable.

Now consider part (ii). We know that the system with mutation measure $u\nu$ converges monotonically downward to an equilibrium when started from $p^{(v)}\nu$. The final part of the proof, showing that this equilibrium is in fact $p^{(u)} \cdot \nu$, proceeds exactly as in the previous proof of Theorem 3.15. ☐

3.8. Equilibria for demographic selective costs

As an example of Corollary 3.14, suppose that S is the demographic selective cost of Section 1.4, so that

$$S(g + \delta_{m'}) - S(g) = \int_0^\infty \left(1 - e^{-\theta(m',x)}\right) f_x \exp\left(-\lambda x - \int_{\mathcal{M}} \theta(m'', x)\, dg(m'')\right) dx.$$

This selection cost is concave, allowing us to apply the results of Sections 3.5 and 3.7. If $\sup_{m,x} \theta(m, x) < \infty$ and $\inf_m \inf_{x\in B} \theta(m, x) > 0$ for some set B such that $\int_B f_x\, dx > 0$, then the conditions of Corollary 3.14 hold, with the function τ a suitable constant and the constant ξ sufficiently small. In other words, if there is a range of fertile ages over which all deleterious mutations reduce survival by at least some minimal amount, then selection keeps the total intensity from going to infinity provided $\nu(\mathcal{M}) \leq e^{-1}\xi$.

On the other hand, when $\sup_{m,x} \theta(m, x) < \infty$ holds (the lower bound is then irrelevant) there is also a (larger) constant τ and ξ sufficiently large such that the reverse inequality (3.17) holds. We then conclude that there is no convergence when $\nu(\mathcal{M})$ is too large.

For example, suppose that S is the demographic selective cost of Section 1.4, with $\mathcal{M} = [\alpha, \beta]$ for $0 < \alpha < \beta < \infty$, ν is a constant multiple of Lebesgue measure on $\mathcal{M}$, f_x is constant, and $\theta(m, x) = \eta \mathbf{1}_{[m,\beta]}(x)$ for some constant η. Such a simplified model has featureless fertility between two ages that represent the onset and end of reproduction, mutations associated with effects at specific ages, constant mutation rate during the reproductive span, and equal increments to the hazard from all mutations. Then, for a suitable constant ζ,

$$S(g + \delta_m) - S(g) \leq \zeta \left(1 - e^{-\eta}\right) \left[\exp(-\lambda m) - \exp(-\lambda\beta)\right] \leq \zeta(m - \beta).$$

Because

$$\int_\alpha^\beta \frac{1}{m - \beta}\, dm = \infty,$$

an equilibrium with finite total mass does not exist. Of course, (3.18) may have a solution p such that $p \cdot \nu$ has infinite total mass and so does not belong to $\mathcal{H}_+$. In the setting of Theorem 3.11, the measure $\pi = p \cdot \nu$ could still arise as the increasing limit of the solution ρ_t to the dynamic equation (2.8) and be such that population average selective cost $\mathbb{E}[S(X^\pi)]$ would be finite. These more delicate possibilities are treated in [**WES08**].

Observe, now, that Theorem 3.15 immediately tells us that the minimal equilibrium is box stable when ν is sufficiently small. In order to show that it is box attractive, though, we need to apply Theorem 3.15, which requires that we show the derivative to be nonsingular in a neighborhood of the equilibrium. An equilibrium π for the demographic selective cost is a measure $\pi(dm) = p(m)\nu(dm)$ where p satisfies

$$\left[\int_0^\infty (1-e^{-\theta(m',x)}) f_x \exp\left(-\lambda x - \int_{\mathcal{M}} (1-e^{-\theta(m'',x)})\, p(m'')\, d\nu(m'')\right) dx\right] p(m) = 1.$$

If we differentiate the left-hand side at p in the direction of the function q, (giving us the derivative of the vector field) we obtain

(3.19)
$$\begin{aligned}
&-\Bigg[\int_0^\infty (1-e^{-\theta(m',x)}) f_x \exp\left(-\lambda x - \int_{\mathcal{M}} (1-e^{-\theta(m'',x)})\, p(m'')\, d\nu(m'')\right) \\
&\qquad\qquad \times \left(-\int_{\mathcal{M}} (1-e^{-\theta(m''',x)})\, q(m''')\, d\nu(m''')\right) dx\Bigg]\, p(m') \\
&\quad - \left[\int_0^\infty (1-e^{-\theta(m',x)}) f_x \exp\left(-\lambda x - \int_{\mathcal{M}} (1-e^{-\theta(m'',x)})\, p(m'')\, d\nu(m'')\right) dx\right]\; q(m') \\
&= \frac{1}{p(m')}\Bigg[p(m')^2 \int_0^\infty (1-e^{-\theta(m',x)}) f_x \exp\left(-\lambda x - \int_{\mathcal{M}} (1-e^{-\theta(m'',x)})\, p(m'')\, d\nu(m'')\right) \\
&\qquad\qquad \times \left(\int_{\mathcal{M}} (1-e^{-\theta(m''',x)})\, q(m''')\, d\nu(m''')\right) dx \quad - q(m')\Bigg].
\end{aligned}$$

As we have already observed, if $\sup_{m,x} \theta(m,x) < \infty$ and $\inf_m \inf_{x \in B} \theta(m,x) > 0$ for some set B such that $\int_B f_x \, dx > 0$, then

$$S(g + \delta_{m'}) - S(g) \geq \xi\, [1 - \exp(-\tau)] \exp(-\tau\, g(\mathcal{M}))$$

for all $m' \in \mathcal{M}$ for suitable constants $\xi > 0$ and $\tau > 0$. If $\nu(\mathcal{M}) \leq e^{-1}\xi$, then there is a minimal equilibrium π with

$$\pi \leq \frac{\gamma(\nu(\mathcal{M}))\xi}{\nu(\mathcal{M})(1-\exp(-\tau))}\nu,$$

where $\gamma(z)$ is the solution of $c\xi = \exp(c)z$ for $z \leq e^{-1}\xi$. Hence $\pi = p \cdot \nu$ with

$$\sup_{m \in \mathcal{M}} p(m) \leq \frac{\gamma(\nu(\mathcal{M}))\xi}{\nu(\mathcal{M})(1-\exp(-\tau))}.$$

Observe that

$$\lim_{z \downarrow 0} \frac{\gamma(z)}{z} = \frac{1}{\xi},$$

so we can vary ν (but keep the selective cost fixed) and conclude that

$$\sup\{p(m) : m \in \mathcal{M},\ \nu(\mathcal{M}) \leq e^{-1}\xi\} =: \hat{p} < \infty.$$

The supremum over $m' \in \mathcal{M}$ of the quantity in brackets on the right-hand side of (3.19) is at most

$$\sup_{m \in \mathcal{M}} q(m) \left[-1 + \nu(\mathcal{M}) \, \hat{p}^2 \int_0^\infty f_x e^{-\lambda x} \, dx \right],$$

which is negative for $\nu(\mathcal{M})$ sufficiently small. Thus, Theorem 3.15 implies that the minimal equilibrium is box attractive when $\nu(\mathcal{M})$ is sufficiently small.

CHAPTER 4

Mutation-selection-recombination in discrete time

We devote the remainder of this work to establishing rigorously the claim that we argued heuristically in the Introduction: that the dynamical system (1.5), which we defined more formally in (2.8), should be the limit of a sequence of infinite population mutation-selection-recombination models in the standard asymptotic regime where selection and mutation are weak relative to recombination and both scale at the same infinitesimal rate in the limit.

More specifically we show that our continuous time model is indeed a limit of the sort of discrete generation, infinite population mutation-selection-recombination models considered in [**BT91, KJB02**], once such models have been extended to incorporate our more general definition of genotypes. Such a result not only justifies our model as a tractable approximation to more familiar models in the literature, but, as we remarked in the Introduction, it also marks out the range of relative strengths for mutation, selection, and recombination where this approximation can be expected to be satisfactory, and thus where the resulting conclusions that underly applications to the demographic study of longevity can be trusted.

4.1. Mutation and selection in discrete-time

Consider first the situation without recombination.

Recall that the distribution of genotypes in a population is described by a probability measure on $\mathcal{G}$, the space of finite integer-valued measures on the set $\mathcal{M}$ of loci.

We consider a sequence of models indexed by the positive integers. In the n^{th} model each new birth gets a random set of extra mutations away from the ancestral wild type added to those it inherited from its parents. The loci at which these mutations occur are distributed as the Poisson random measure $X^{\nu/n}$ with intensity ν/n. The fitness of a genotype $g \in \mathcal{G}$ is $e^{-S(g)/n}$. We speed up time by a factor of n so that n generations pass in 1 unit of time. The net result is that in a finite, non-zero amount of time the population is asymptotically subject to finite, non-zero "amounts" of mutation and selection.

We now define operators $\mathfrak{M}_n$ and $\mathfrak{S}_n$ acting on the space of probabilities on $\mathcal{G}$ that describe the transformation in genotype distribution by, respectively, one round of mutation and one round of selection. The intuition behind these definitions is described in the Introduction.

Notation 4.1. Given a probability measure P on $\mathcal{G}$, define new probability measures $\mathfrak{M}_n P$ and $\mathfrak{S}_n P$ by

$$\mathfrak{M}_n P[F] := \int_{\mathcal{G}} \mathbb{E}[F(g + X^{\nu/n})]\, dP(g) \tag{4.1}$$

and

$$\mathfrak{S}_n P[F] := \frac{\int_{\mathcal{G}} e^{-S(g)/n} F(g)\, dP(g)}{\int_{\mathcal{G}} e^{-S(g)/n}\, dP(g)} \tag{4.2}$$

for a Borel function $F : \mathcal{G} \to \mathbb{R}_+$.

Note that when n is large,

$$\mathbb{P}\{X^{\nu/n} = 0\} = \exp(-\nu(\mathcal{M})/n) \approx 1 - \nu(\mathcal{M})/n,$$

and conditional on the event $\{X^{\nu/n} \neq 0\}$ the random measure $X^{\nu/n}$ is approximately a unit point mass with location in $\mathcal{M}$ chosen according to the probability measure $\nu(\cdot)/\nu(\mathcal{M})$. Hence, for any probability measure P on $\mathcal{G}$

$$\lim_{n\to\infty} n\big(\mathfrak{M}_n P[F] - P[F]\big) = \int_{\mathcal{G}} \left(\int_{\mathcal{M}} F(g + \delta_m) - F(g)\, \nu(dm) \right) dP(g). \tag{4.3}$$

In particular, if F is of the form $F(g) := e^{-g[f]}$ for some Borel function $f : \mathcal{M} \to \mathbb{R}_+$, then

$$\mathfrak{M}_n P[F] = P\left[F \cdot e^{\nu[e^{-f}-1]/n}\right] \tag{4.4}$$

and so

$$\lim_{n\to\infty} n\big(\mathfrak{M}_n P[F] - P[F]\big) = \nu\big[e^{-f} - 1\big] P[F]. \tag{4.5}$$

Note also that $e^{-S(g)/n} \approx 1 - S(g)/n$ when n is large and so, when $P[S]$ is finite,

$$\lim_{n\to\infty} n\big(\mathfrak{S}_n P[F] - P[F]\big) = P[S \cdot F] - P[S]P[F]. \tag{4.6}$$

When we start with a population genotype distribution P, the population genotype distribution after one generation of mutation and selection is $\mathfrak{M}_n\mathfrak{S}_n P$ (assuming that selection precedes mutation). A trajectory of the resulting discrete-time model is defined by iteration: Given an initial population genotype distribution P, the population genotype distribution after m generations is $\big(\mathfrak{M}_n\mathfrak{S}_n\big)^m P$.

If, after speeding up time in the n^{th} model, the resulting sequence of trajectories has a continuously differentiable limit P_t, this limit should satisfy the equation

$$\begin{aligned}
\lim_{\epsilon\downarrow 0} \epsilon^{-1}\big(P_{t+\epsilon}[F] - P_t[F]\big) &= \lim_{n\to\infty} n\big(\mathfrak{M}_n\mathfrak{S}_n P[F] - P[F]\big) \\
&= \lim_{n\to\infty} n\big(\mathfrak{S}_n P[F] - P[F]\big) + \lim_{n\to\infty} n\big(\mathfrak{M}_n\mathfrak{S}_n P[F] - \mathfrak{S}_n P[F]\big) \\
&= P[SF] - P[S]P[F] + \nu\big[e^{-f} - 1\big] P[F]
\end{aligned}$$

for test functions $F : \mathcal{G} \to \mathbb{R}_+$ of the form $F(g) = e^{-g[f]}$ for some Borel function $f : \mathcal{M} \to \mathbb{R}_+$. For this special choice of test function, this is precisely the dynamical equation defining the recombination-free process that we introduced in [**SEW05**] and derived heuristically in equation (1.4). In fact, these test functions are enough to consider, since they determine probability distributions on $\mathcal{G}$. Formally, a proof that the discrete-time process converges to this continuous-time process would require that we prove the existence of the continuously differentiable limit, a fact that we assumed above.

4.2. Recombination in discrete-time

We now introduce recombination. Imitating [**BT91**], we think of a recombination event as taking two genotypes $g', g'' \in \mathcal{G}$ from the population and replacing the genotype g' in the population by the genotype g defined by $g(A) := g'(A \cap R) + g''(A \cap R^c)$, where $R \subseteq \mathcal{M}$ is the particular "segregating set" for the recombination event. That is, the new individual with genotype g has the same accumulated mutations as the individual with genotype g' (respectively, g'') for "loci" in the set R (respectively, R^c). (Recall that we work in an abstract framework in which loci are just places at which mutations from wild type can occur rather than concrete physical loci *per se*). From a biological point of view, the recombination process would more legitimately be defined on a linear sequence of loci, as in [**BT91**], so that one can picture recombination caused by crossovers during meiosis. Working with a somewhat abstract space of loci is perhaps a weakness, but also a useful simplification, of our approach.)

We will, of course, think of g' and g'' as being chosen independently at random according to the particular probability measure describing the distribution of genotypes in the population. We will also imagine that the segregating set is chosen at random via some suitable mechanism. In order to discuss random sets rigorously, we follow formalism described in [**Ken74**] and define a σ-algebra on sets of Borel subsets of $\mathcal{M}$ by the requirement that all incidence functions with Borel subsets are measurable. A consequence of this definition is that if Ξ is a random Borel set and κ is a finite measure, then $\kappa(\Xi)$ is a real-valued random variable. We suppose there is a probability distribution $\mathcal{R}$ that describes the distribution of the random set of loci that segregate together. We will always assume, without loss of generality, that $\mathcal{R}$ is *symmetric* in the sense that

$$\mathcal{R}(A) = \mathcal{R}\big(\{R^c \,:\, R \in A\}\big), \tag{4.7}$$

where R^c denotes the complement of the set R.

Notation 4.2. For any Borel measure g on $\mathcal{M}$ and Borel subset R of $\mathcal{M}$, define the Borel measure g_R on $\mathcal{M}$ by

$$g_R(A) := g(A \cap R)$$

for Borel subsets $A \subset \mathcal{M}$. Given the distribution $\mathcal{R}$ of a random subset of $\mathcal{M}$, define the corresponding recombination operator that maps the space of Borel probability measures on $\mathcal{G}$ into itself by

$$\mathfrak{R}P[f] := \int_{\mathcal{B}(\mathcal{M})} \int_{\mathcal{G}} \int_{\mathcal{G}} f(g'_R + g''_{R^c})\, dP(g')\, dP(g'')\, d\mathcal{R}(R), \tag{4.8}$$

where P is a Borel probability measure on $\mathcal{G}$, $f : \mathcal{G} \to \mathbb{R}$ is a bounded Borel function, and $\mathcal{B}(\mathcal{M})$ is the collection of Borel subsets of $\mathcal{M}$.

Thus, if P describes the distribution of genotypes in the population, then $\mathfrak{R}P$ describes the distribution of a genotype that is obtained by picking two genotypes g' and g'' independently according to P and then forming a composite genotype that agrees with g' on the set R and g'' on the set R^c, where the segregating set R is picked according to the distribution $\mathcal{R}$. Recall that μP denotes the intensity measure of P; that is,

$$\mu P(A) := \int_{\mathcal{G}} g(A)\, dP(g). \tag{4.9}$$

REMARK 4.3. Note that $\mu\mathfrak{R}P = \mu P$; that is, the intensity measure of the point process describing the genotype of a randomly chosen individual is left unchanged by recombination. Also, note that if P is the distribution of a Poisson random measure, then $\mathfrak{R}P = P$.

4.3. The discrete-time model

NOTATION 4.4. Given a probability measure P_0 on $\mathcal{G}$, set $Q_k := (\mathfrak{R}\mathfrak{M}_n\mathfrak{S}_n)^k P_0$. That is, if P_0 describes the population at generation 0, then Q_k describes the population at generation k, where we suppose that each generation is produced from the previous one by the successive action of selection, mutation, and recombination.

CHAPTER 5

Hypotheses and statement of the convergence result

5.1. Shattering of point processes

We expect recombination to break up dependencies between different parts of the genome, so that $\mathfrak{R}^k P$ should be approximately $\Pi_{\mu P}$ when k is large, where we recall that $\Pi_{\mu P}$ is the distribution of the Poisson random measure with intensity μP. In order that this happens for a given distribution P, it must generically be the case that there is a positive probability that the segregating set and its complement will both intersect any set with positive μP mass in two sets that each have positive μP mass. The following condition (with $\lambda = \mu P$) will be useful for establishing quantitative bounds on the rate with which $\mathfrak{R}^k P$ converges to $\Pi_{\mu P}$.

Definition 5.1. Given a recombination measure $\mathcal{R}$ and a finite measure λ on $\mathcal{M}$, we say that the pair $(\mathcal{R}, \lambda)$ is *shattering* if there is a positive constant α such that for any Borel set A,

$$\begin{aligned} \lambda(A)^3 &\le \alpha \left[\lambda(A)^2 - 2\int \lambda(A\cap R)^2 \, d\mathcal{R}(R)\right] \\ &= 2\alpha \int \lambda(A\cap R)\lambda(A\cap R^c)\, d\mathcal{R}(R) \end{aligned} \tag{5.1}$$

For example, suppose that N is a finite simple point process on $\mathcal{M} = (0,1]$. We think of $\mathcal{M}$ in this case as a physical chromosome and N as the set of crossover points formed during meiosis. Write $0 < T_1 < \ldots < T_K < 1$ for the successive points of $\mathcal{M}$. Set $T_0 = 0$ and $T_{K+1} = 1$. Let Z be a $\{0,1\}$-valued random variable that is independent of N with $\mathbb{P}\{Z = 0\} = \frac{1}{2}$. Define $\mathcal{R}$ to be the distribution of the random set given by

$$\begin{aligned} &(T_0, T_1] \cup (T_2, T_3] \cup \ldots, \quad \text{if } Z = 0, \\ &(T_1, T_2] \cup (T_3, T_4] \cup \ldots, \quad \text{if } Z = 1. \end{aligned}$$

Take λ to be any diffuse probability measure (that is, λ has no atoms). Suppose that there is a constant c such that $\mathbb{P}\{N(u,w] = 1\} \ge c\lambda((u,w])$ for $0 < u < w \le 1$. This will be the case, for example, if N is Poisson process with intensity measure bounded below by a positive multiple of λ or N consists of a single point with distribution bounded below by a positive multiple of λ. Also, for most "reasonable" simple point processes with intensity $C\lambda$ for some constant C, it will be the case that $\mathbb{P}\{N(u,w] = 1\} \approx \mathbb{E}[N(u,w]] = C\lambda((u,w])$ when $|u - w|$ is small, where the notation $\approx$ indicates that the ratio of the two sides is close to 1.

Then,

$$\begin{aligned}
&\lambda(A)^2 - 2\int \lambda(A \cap R)^2 \, d\mathcal{R}(R) \\
&\quad = \lambda(A)^2 - 2\frac{1}{2}\int_0^1 \int_0^1 \mathbf{1}_A(u)\mathbf{1}_A(w)\mathbb{P}\{N((u,w]) = 0 \mod 2\} \, d\lambda(u)\, d\lambda(w) \\
&\quad = \int_0^1 \int_0^1 \mathbf{1}_A(u)\mathbf{1}_A(w)\mathbb{P}\{N((u,w]) = 1 \mod 2\} \, d\lambda(u)\, d\lambda(w) \\
&\quad \geq \int_0^1 \int_0^1 \mathbf{1}_A(u)\mathbf{1}_A(w)\mathbb{P}\{N((u,w]) = 1\} \, d\lambda(u)\, d\lambda(w) \\
&\quad \geq c\int_0^1 \int_0^1 \mathbf{1}_A(u)\mathbf{1}_A(w)\lambda((u,w]) \, d\lambda(u)\, d\lambda(w).
\end{aligned}$$

Observe that

$$\begin{aligned}
\lambda(A)^3 &= 3! \iiint_{\{0<u<v<w\leq 1\}} \mathbf{1}_A(u)\mathbf{1}_A(v)\mathbf{1}_A(w)\, d\lambda(u)\, d\lambda(v)\, d\lambda(w) \\
&\leq 3! \iiint_{\{0<u<v<w\leq 1\}} \mathbf{1}_A(u)\mathbf{1}_A(w)\, d\lambda(u)\, d\lambda(v)\, d\lambda(w) \\
&= 3! \iint_{\{0<u<w\leq 1\}} \mathbf{1}_A(u)\mathbf{1}_A(w)\lambda((u,w])\, d\lambda(u)\, d\lambda(w) \\
&= 3\int_0^1 \int_0^1 \mathbf{1}_A(u)\mathbf{1}_A(w)\lambda((u,w])\, d\lambda(u)\, d\lambda(w).
\end{aligned}$$

Thus, the pair $(\mathcal{R}, \lambda)$ is shattering with constant $\alpha = 3/c$.

REMARK 5.2. Note that if the pair $(\mathcal{R}, \lambda)$ is shattering with some constant α, then it follows from (5.1) that $\lambda(\{x\}) = 0$ for all $x \in \mathcal{M}$, and so the measure λ is diffuse.

REMARK 5.3. It can be shown that if the pair $(\mathcal{R}, \mu P)$ is shattering and there is a constant β such that

$$\int g(A)\mathbf{1}_{\{g(A)\geq 2\}}\, dP(g) \leq \beta \mu P(A)^2$$

for any Borel set $A \subseteq \mathcal{M}$, then $\mathfrak{R}^k P$ converges to $\Pi_{\mu P}$ as $k \to \infty$.

5.2. Statement of the main convergence result

We can now state our main result that justifies the continuous-time model (2.8) as a limit of discrete-time models in which recombination acts on a faster time scale than mutation and selection. We use the notation $\lfloor x \rfloor$ to denote the greatest integer less than or equal to the real number x.

THEOREM 5.4. *Let $(\rho_t)_{t\geq 0}$ be the measure-valued dynamical system of* (2.8) *whose existence is guaranteed by Theorem 2.4. Suppose in addition to the hypotheses of Theorem 2.4 that the pair $(\mathcal{R}, \nu)$ (respectively, $(\mathcal{R}, \rho_0)$) consisting of the recombination measure and the mutation intensity measure (respectively, the recombination measure and the initial intensity) is shattering, the initial measure P_0 is Poisson* *(with intensity ρ_0), and the selective cost S is bounded with $\sigma = \sup_{g \in \mathcal{G}} S(g)$. Then, for any $T > 0$,*

$$\lim_{n\to\infty} \sup_{0\le t\le T} \left\|\Pi_{\rho_t} - Q_{\lfloor tn\rfloor}\right\|_{\mathrm{Was}} = 0.$$

This result is proved in Chapter 8 as a consequence of a similar result in Chapter 6 in which the partially Poissonizing action of recombination is replaced by complete Poissonization in each generation. Although it is intuitively reasonable that the effect of recombination should be asymptotically indistinguishable from that of complete Poissonization when the per-generation effect of selection is sufficiently small, the technical difficulties are substantial, and we collect some necessary preliminary results in Chapter 7.

REMARK 5.5. We have assumed throughout, for notational convenience, a particular order of operations: In each generation there is first selection, then mutation, then recombination. This order has **no special significance**, and it is easy to see that the proofs would hold equally well for another order, or if the same total amounts of mutation and selection were split up into multiple bouts within a generation, whether before or after recombination.

REMARK 5.6. Because of Remark 5.2, the hypotheses of Theorem 5.4 imply that the mutation intensity measure ν and the initial intensity ρ_0 are both diffuse. It follows easily from this that each probability measure Q_k assigns all of its mass to the set of elements of $\mathcal{G}$ that have atoms of mass one; that is, every Q_k is the distribution of a *simple* point process.

From now on, we assume without further comment that all the probability measures on $\mathcal{G}$ we consider are distributions of simple point processes.

CHAPTER 6

Convergence with complete Poissonization

Recall from Notation 4.4 that $Q_k := (\mathfrak{R}\mathfrak{M}_n\mathfrak{S}_n)^k P_0$. Theorem 5.4 says that, under suitable conditions, $Q_{\lfloor tn \rfloor}$ is close to Π_{ρ_t} when n is large. The intuition behind this result is that $\mathfrak{R}$ drives distributions on $\mathcal{G}$ towards Poisson faster than $\mathfrak{S}_n$ drive ones away. As a first step in the proof of this theorem, we prove a simpler result in which the operator $\mathfrak{R}$ is replaced by the operator that immediately replaces a distribution on $\mathcal{G}$ by the distribution of a Poisson random measure with the same intensity measure. This latter proposition is essentially a *shadowing* result, but none of the standard shadowing theorems, such as those in [**CKP95**], can be applied as stated to this setting. In any case, the direct proof is not difficult.

Notation 6.1. The complete Poissonization operator $\mathfrak{P}$ acts on a probability measure P on $\mathcal{G}$ by $\mathfrak{P}P := \Pi_{\mu P}$. That is, $\mathfrak{P}P$ is the distribution of the Poisson random measure with the same intensity measure μP as P.

Notation 6.2. Set

$$Q'_k := (\mathfrak{P}\mathfrak{M}_n\mathfrak{S}_n)^k P_0$$

and

$$\pi_k := \mu Q'_k.$$

That is, Q'_k the distribution of the Poisson random measure resulting from k iterations of mutation and selection of intensity $1/n$, with complete Poissonization after every round, and π_k is the corresponding intensity measure.

Proposition 6.3. *There are constants A, B, C, depending on K, $\nu(\mathcal{M})$, and $\rho_0(\mathcal{M})$, such that for every $T > 0$*

$$\sup_{0 \le t \le T} \|\pi_{\lfloor tn \rfloor} - \rho_t\|_{\mathrm{Was}} \le e^{AT}(BT + C)n^{-1}.$$

In particular,

$$\lim_{n\to\infty} \sup_{0 \le t \le T} \left\| \Pi_{\rho_t} - Q'_{\lfloor tn \rfloor} \right\|_{\mathrm{Was}} = 0.$$

Proof. For any $0 \le s \le t$,

$$\|\rho_t - \rho_s\|_{\mathrm{Was}} \le \|(t-s)\nu\|_{\mathrm{Was}} + \left\| \int_s^t F_{\rho_u}\rho_u \, du \right\|_{\mathrm{Was}}.$$

The first term on the right is simply $(t-s)\|\nu\|_{\mathrm{Was}} = (t-s)\nu(\mathcal{M})$. For the second term we use Lemma 2.5 and the triangle inequality, together with the universal bound $\|\rho_u\|_{\mathrm{Was}} \le \|\rho_0\|_{\mathrm{Was}} + u\|\nu\|_{\mathrm{Was}}$, to obtain the bound

$$2K \int_s^t \|\rho_u\|_{\mathrm{Was}} \, du \le 2K(t-s)\big(\rho_0(\mathcal{M}) + t\nu(\mathcal{M})\big).$$

Putting these bounds together,

$$\left\|\rho_t-\rho_s\right\|_{\mathrm{Was}} \le (t-s)\nu(\mathcal{M})+2K(t-s)\big(\rho_0(\mathcal{M})+t\nu(\mathcal{M})\big). \tag{6.1}$$

Note that

$$\frac{d(\pi_{m+1}-\nu/n)}{d\pi_m}(x)=\frac{\mathbb{E}[e^{-S(X^{\pi_m}+\delta_x)/n}]}{\mathbb{E}[e^{-S(X^{\pi_m})/n}]}, \tag{6.2}$$

so that

$$\begin{aligned}\left\|\pi_{m+1}-\frac{\nu}{n}-\pi_m\left(1-\frac{1}{n}F_{\pi_m}(\cdot)\right)\right\|_{\mathrm{Was}}&\\ \le n^{-2}\cdot 2e^{2K/n}(\mathbb{E}[S(X^{\pi_m})^2]+K^2)\exp\left\{\mathbb{E}[S(X^{\pi_m})/n]\right\}.&\end{aligned} \tag{6.3}$$

Next,

$$\begin{aligned}\left\|\pi_m\left(1-\frac{1}{n}F_{\pi_m}\right)-\pi_m\left(1-\frac{1}{n}F_{\rho_{m/n}}\right)\right\|_{\mathrm{Was}}&\\ \le \frac{1}{n}\sup_{x\in\mathfrak{M}}\left|F_{\pi_m}(x)-F_{\rho_{m/n}}(x)\right|&\\ \le \frac{8K}{n}\left\|\pi_m-\rho_{m/n}\right\|_{\mathrm{Was}}&\end{aligned} \tag{6.4}$$

by Lemma 2.8, and

$$\left\|\pi_m\left(1-\frac{1}{n}F_{\rho_{m/n}}\right)-\rho_{m/n}\left(1-\frac{1}{n}F_{\rho_{m/n}}\right)\right\|_{\mathrm{Was}}\le\left\|\pi_m-\rho_{m/n}\right\|_{\mathrm{Was}} \tag{6.5}$$

since $F_\rho(x)$ is always nonnegative. Finally,

$$\begin{aligned}&\left\|\rho_{(m+1)/n}-\frac{\nu}{n}-\rho_{m/n}\left(1-\frac{1}{n}F_{\rho_{m/n}}\right)\right\|_{\mathrm{Was}}\\ &\quad\le\left\|\int_0^{1/n}F_{\rho_{m/n}}\rho_{m/n}-F_{\rho_{s+m/n}}\rho_{s+m/n}\,ds\right\|_{\mathrm{Was}}\\ &\quad\le\int_0^{1/n}\sup_{x\in\mathcal{M}}\left|F_{\rho_{m/n}}(x)-F_{\rho_{s+m/n}}(x)\right|\left\|\rho_{m/n+s}\right\|_{\mathrm{Was}}ds\\ &\quad\quad+\sup_{x\in\mathcal{M}}F_{\rho_{m/n}}(x)\int_0^{1/n}\left\|\rho_{m/n}-\rho_{s+m/n}\right\|_{\mathrm{Was}}ds\\ &\quad\le\int_0^{1/n}8Ks\left(\frac{m+1}{n}\nu(\mathcal{M})+\rho_0(\mathcal{M})\right)ds\\ &\quad\quad+K\int_0^{1/n}\left\|s\nu-\int_0^s D\rho_{u+m/n}\,du\right\|_{\mathrm{Was}}ds\\ &\quad\le Kn^{-2}\left(4+\frac{1}{2}\right)\frac{m+1}{n}\big(\nu(\mathcal{M})+\rho_0(\mathcal{M})\big)+K\nu(\mathcal{M})n^{-2}.\end{aligned} \tag{6.6}$$

Thus, there are constants a and b such that

$$\|\pi_{m+1}-\rho_{(m+1)/n}\|_{\mathrm{Was}}\le\frac{a}{n}\|\pi_m-\rho_{m/n}\|_{\mathrm{Was}}+\frac{bT+c}{n^2}.$$

Hence,

$$\|\pi_m-\rho_{m/n}\|_{\mathrm{Was}}\le n^{-1}e^{aT}\frac{bT+c}{a}.$$

Combining this inequality with the inequality (6.1), the proposition follows immediately. □

CHAPTER 7

Preparatory lemmas for the main convergence result

7.1. Consequences of shattering

We introduced the recombination probability measure $\mathcal{R}$ as the distribution of a random subset of $\mathcal{M}$ with the property that the subset and its complement have the same distribution, and we think of the subset and its complement as segregating set of loci. These two sets form a *partition* of $\mathcal{M}$ and, with a slight abuse of notation, we will use also use the notation $\mathcal{R}$ for the distribution of this random partition of $\mathcal{M}$ into two sets. In the same vein, we define the probability measure $\mathcal{R}_k$ on partitions of $\mathcal{M}$ to be the distribution of the coarsest random partition that is finer than each of the independent partitions $\{R_j, R_j^c\}$, $1 \le j \le k$, where each partition $\{R_j, R_j^c\}$ has distribution $\mathcal{R}$. That is, $\mathcal{R}_k$ is the distribution of the partition of $\mathcal{M}$ consisting of the non-empty sets of the form $\tilde{R}_1 \cap \cdots \tilde{R}_k$, where $\tilde{R}_j$ is either R_j or R_j^c. Using standard notation, we write this partition as $\{R_1, R_1^c\} \wedge \cdots \wedge \{R_k, R_k^c\}$. Note that the number of sets in this partition is at most 2^k, but may be smaller.

With this notation, the product of k copies of the recombination operator $\mathfrak{R}$ can be written as

$$\mathfrak{R}^k = \int \mathfrak{R}_{\mathcal{A}} d\mathcal{R}_k(\mathcal{A}), \tag{7.1}$$

where, for a finite partition $\mathcal{A} = \{A_1, \ldots, A_K\}$ of $\mathcal{M}$ into Borel sets, the *annealed recombination operator* $\mathfrak{R}_{\mathcal{A}}$ is the operator on probability measure on $\mathcal{G}$ defined by

$$\mathfrak{R}_{\mathcal{A}} Q[F] := \int \cdots \int F\left(g_{A_1}^{(1)} + \cdots + g_{A_K}^{(K)}\right) dQ(g^{(1)}) \cdots dQ(g^{(K)}) \tag{7.2}$$

for a probability measure Q on $\mathcal{G}$ and a function $F : \mathcal{G} \to \mathbb{R}$. The annealed recombination operator should be thought of as representing the combined effect of some number k of successive recombination events involving specified segregating sets $R_1, \ldots, R_k$ such that $\mathcal{A} = \{R_1, R_1^c\} \wedge \cdots \wedge \{R_k, R_k^c\}$. Note that if $\mathcal{A}$ and $\mathcal{A}'$ are partitions, then $\mathfrak{R}_{\mathcal{A}} \mathfrak{R}_{\mathcal{A}'} = \mathfrak{R}_{\mathcal{A} \wedge \mathcal{A}'}$.

Given some probability measure P on $\mathcal{G}$, the relation (7.1) indicates that in order for $\mathfrak{R}^k P$ to be approximately Poisson with intensity μP for large k it will, in general, be necessary that a partition of $\mathcal{M}$ distributed according to $\mathcal{R}_k$ typically consists of sets that all have small μP mass, so that a sample from $\mathfrak{R}^k P$ consists of a mosaic of many small pieces each taken from genomes sampled independently from the population described by P. The following is a convenient way of measuring the extent to which a partition of $\mathcal{M}$ is made up of sets that each have small mass with respect to some reference measure.

Notation 7.1. Suppose that $\mathcal{A} = (A_1, \dots, A_k)$ is a partition of $\mathcal{M}$, and λ a measure on $\mathcal{M}$. For $r > 0$ set

$$|\mathcal{A}|_r^{(\lambda)}(t) := \sum_{i=1}^{k} \lambda(A_i)^r.$$

Recall the definition of a shattering pair from Definition 5.1.

Lemma 7.2. *If the pair $(\mathcal{R}, \lambda)$ is shattering with constant α, then for all $k \geq 1$,*

$$\int |\mathcal{A}|_2^{(\lambda)} \, d\mathcal{R}_k(\mathcal{A}) \leq \frac{\alpha^*}{k+1},$$

where

$$\alpha^* = \lambda(\mathcal{M})^2 \vee 2\alpha\lambda(\mathcal{M}).$$

Proof. Define a sequence of random partitions $\mathcal{A}_0, \mathcal{A}_1, \dots$ of $\mathcal{M}$ by setting $\mathcal{A}_0 = \{\mathcal{M}\}$ and $\mathcal{A}_k = \{R_1, R_1^c\} \wedge \cdots \wedge \{R_k, R_k^c\}$ for $k \geq 1$, where the partitions $\{R_j, R_j^c\}$, $j \geq 1$, are independent and identically distributed with common distribution $\mathcal{R}$. Set $X_k = |\mathcal{A}_k|_2^{(\lambda)}$. Then,

$$\int |\mathcal{A}|_2^{(\lambda)} \, d\mathcal{R}_k(\mathcal{A}) = \mathbb{E}[X_k].$$

By the symmetry of $\mathcal{R}$,

$$\begin{aligned}
\mathbb{E}\left[X_{k+1} \,\middle|\, \mathcal{A}_1, \dots, \mathcal{A}_k\right] &= \int \left(\sum_{A \in \mathcal{A}_k} \lambda(A \cap R)^2 + \lambda(A \cap R^c)^2 \right) d\mathcal{R}(R) \\
&= \sum_{A \in \mathcal{A}_k} 2 \int \lambda(A \cap R)^2 \, d\mathcal{R}(R) \\
&\leq \sum_{A \in \mathcal{A}_k} \lambda(A)^2 (1 - \alpha\lambda(A)) \\
&= |\mathcal{A}_k|_2^{(\lambda)} - \alpha |\mathcal{A}_k|_3^{(\lambda)}.
\end{aligned}$$

It is always the case for any partition $\mathcal{A}$ of $\mathcal{M}$ that

$$|\mathcal{A}|_3^{(\lambda)} \geq \big(|\mathcal{A}|_1^{(\lambda)}\big)^{-1} \big(|\mathcal{A}|_2^{(\lambda)}\big)^2 = \lambda(\mathcal{M})^{-1} \big(|\mathcal{A}|_2^{(\lambda)}\big)^2.$$

Thus, setting $c = 2/\alpha\lambda(\mathcal{M})$,

$$\mathbb{E}\left[X_{k+1} \,\middle|\, \mathcal{A}_1, \dots, \mathcal{A}_k\right] \leq X_k (1 - cX_k).$$

Applying Jensen's inequality to the concave function $x(1 - cx)$, we see that

$$\mathbb{E}[X_{k+1}] \leq \mathbb{E}\big[X_k(1 - cX_k)\big] \leq (\mathbb{E}[X_k])\big(1 - c\mathbb{E}[X_k]\big). \tag{7.3}$$

We complete the proof by induction. We have $X_0 = \lambda(\mathcal{M})^2$. Suppose that $\mathbb{E}[X_k] \leq \alpha^*/(k+1)$. Since $\alpha^* \geq 1/c$,

$$\begin{aligned}
\mathbb{E}[X_{k+1}] &\leq \frac{\alpha^*}{k+1} \left(1 - \frac{c\alpha^*}{k+1}\right) \\
&\leq \frac{\alpha^*}{k+1} \cdot \frac{k}{k+1} \\
&\leq \frac{\alpha^*}{k+2}.
\end{aligned}$$

□

The following corollary is immediate.

Corollary 7.3. *Suppose that the pairs* $(\mathcal{R}, \rho_0)$ *and* $(\mathcal{R}, \nu)$ *are both shattering with the same constant* α. *Let* $\nu_t = t\nu + \rho_0$. *Then, for all* $k \geq 1$,

$$\int |\mathcal{A}|_2^{(\nu_t)} \, d\mathcal{R}_k(\mathcal{A}) \leq \frac{\alpha^*}{k+1},$$

where

$$\alpha^* := (2\nu_t(\mathcal{M})^2) \vee (4\alpha\nu_t(\mathcal{M})).$$

7.2. Estimates for Radon-Nikodym derivatives

Let X be the canonical random variable on $\mathcal{G}$; that is, X is defined by $X(g) = g$. Given a probability measure Q on $\mathcal{G}$, a Borel subset A of $\mathcal{M}$, and a measure $g \in \mathcal{G}$, we may define a regular conditional distribution $Q[\cdot \mid X_A = g_A]$.

Recall the definition of the annealed recombination operator $\mathfrak{R}_{\mathcal{A}}$ from (7.2).

Lemma 7.4. *For a finite partition* $\mathcal{A}$ *of* $\mathcal{M}$ *into Borel subsets and a probability measure* Q *on* $\mathcal{G}$, *the Radon–Nikodym derivative of the probability measure* $\mathfrak{R}_{\mathcal{A}}\mathfrak{S}_n Q$ *with respect to the probability measure* $\mathfrak{R}_{\mathcal{A}} Q$ *is given by*

$$\frac{d\mathfrak{R}_{\mathcal{A}}\mathfrak{S}_n Q}{d\mathfrak{R}_{\mathcal{A}} Q} = e^{-\hat{S}_{Q,\mathcal{A}}/n},$$

where

$$\begin{aligned} &\hat{S}_{Q,\mathcal{A}}(g) := \\ &\qquad - n \sum_{A \in \mathcal{A}} \left(\log Q \left[e^{-S(X)/n} \,\middle|\, X_A = g_A \right] - \log Q \left[e^{-S(X)/n} \right] \right). \end{aligned}$$

Proof. Write $\mathcal{A} = \{A_1, \ldots, A_K\}$. By definition, for any measurable $F : \mathcal{G} \to [0, 1]$,

$$\begin{aligned} (\mathfrak{R}_{\mathcal{A}}\mathfrak{S}_n Q)[F] &= \int \cdots \int F\left(g^{(1)}_{A_1} + \cdots + g^{(K)}_{A_K}\right) \, d\mathfrak{S}_n Q(g^{(1)}) \cdots d\mathfrak{S}_n Q(g^{(K)}) \\ &= Q\left[e^{-S(X)/n}\right]^{-K} \int \cdots \int F\left(\sum_{i=1}^K g^{(i)}_{A_i}\right) \\ &\quad \times \exp\left\{-\sum_{i=1}^K S(g^{(i)})/n\right\} \, dQ(g^{(1)}) \cdots dQ(g^{(K)}) \\ &= Q\left[e^{-S(X)/n}\right]^{-K} \int \cdots \int F(g) \\ &\quad \times \left(\prod_{i=1}^K \int_{\mathcal{G}} e^{-S(X)/n} dQ(X \,\middle|\, X_{A_i} = g_{A_i})\right) \, d\mathfrak{R}_{\mathcal{A}} Q(g) \\ &= (\mathfrak{R}_{\mathcal{A}} Q)[e^{-\hat{S}_{Q,\mathcal{A}}/n} F]. \end{aligned}$$

□

Lemma 7.5. *For a probability measure* Q *on* $\mathcal{G}$, *a partition* $\mathcal{A}$ *of* $\mathcal{M}$ *into Borel sets, and measures* $g, g' \in \mathcal{G}$,

$$\left|\hat{S}_{Q,\mathcal{A}}(g) - \hat{S}_{Q,\mathcal{A}}(g')\right| \leq \sigma(g(\mathcal{M}) + g'(\mathcal{M})).$$

Proof. We may write

$$\begin{aligned}\hat{S}_{Q,\mathcal{A}}(g) = -n\sum_{A\in\mathcal{A}}&\left(\log Q\left[e^{-S(X)/n} \,\middle|\, X_A = g_A\right] - \log Q\left[e^{-S(X)/n} \,\middle|\, X_A = 0\right]\right)\\ &- n\sum_{A\in\mathcal{A}}\left(\log Q\left[e^{-S(X)/n} \,\middle|\, X_A = 0\right] - \log Q\left[e^{-S(X)/n}\right]\right).\end{aligned}$$

Consider the summand in the first sum corresponding to the set A in the partition $\mathcal{A}$. If $g_A = 0$, then the summand is 0. If $g_A > 0$, then the summand is bounded above by

$$\begin{aligned}&\log Q[e^{-S(X)/n}1\{X_A = g_A\}] + \log Q[1\{X_A = 0\}]\\ &\quad - \log Q[e^{-S(X)/n}1\{X_A = 0\}] - \log Q[1\{X_A = g_A\}]\\ &\le \log Q[e^{0}1\{X_A = g_A\}] + \log Q[1\{X_A = 0\}]\\ &\quad - \log Q[e^{-\sigma/n}1\{X_A = 0\}] - \log Q[1\{X_A = g_A\}]\\ &\le \sigma/n,\end{aligned}$$

and, by a similar argument, it is bounded below by $-\sigma/n$. Thus, the first sum is bounded in absolute value by $\sigma\#\{A \in \mathcal{A} : g(A) > 0\} \le \sigma g(\mathcal{M})$.

Note that $\hat{S}_{Q,\mathcal{A}}(g) - \hat{S}_{Q,\mathcal{A}}(g')$ is the difference of the first sum and the corresponding quantity with g replaced by g'; that is, the second sum and its counterpart for g' coincide and hence cancel each other in the difference. □

For some purposes, it will be useful to keep track of the temporal order in which mutations appeared in genotype. We denote by $\mathcal{G}^*$ the space of finite sequences of genotypes. That is,

$$\mathcal{G}^* := \mathcal{G} \sqcup \mathcal{G}^2 \sqcup \cdots, \tag{7.4}$$

where $\sqcup$ denotes disjoint union. We think of an element of $\mathcal{G}^i$ as recording the mutations from the ancestral wild type that appear in each of i consecutive generations. There are natural "projection" operations $\Sigma : \mathcal{G}^* \to \mathcal{G}$ defined for $g = (g_i, \ldots, g_2, g_0) \in \mathcal{G}^{i+1}$ (it will be convenient for us later to index sequences in reverse order) by

$$\Sigma(g) := g_0 + g_2 + \cdots + g_i$$

and

$$\Psi_j(g) = \begin{cases} g_j, & 0 \le j \le i,\\ 0, & j > i.\end{cases}$$

In essence, Σ removes the "labels" that record in which generation the various mutations from the ancestral wild type occurred, while Ψ_j isolates mutations that occurred in generation j. Note that $\Sigma = \sum_{j=0}^{\infty} \Psi_j$.

A probability measure Q on $\mathcal{G}^*$ may be thought of as a sequence of subprobability measures $(Q^{(i)})_{i=0}^{\infty}$, where $Q^{(i)}$ is the portion of the Q concentrated on $i+1$-tuples. Thus, $\sum_{i=0}^{\infty} Q^{(i)}(\mathcal{G}^{i+1}) = 1$. We interpret Q as the distribution of a random sequence $(X_I, \ldots, X_2, X_0)$, with $Q^{(i)}(\mathcal{G}^{i+1})$ being the probability of the event $\{I = i\}$ that the sequence has length $i+1$ and $Q^{(i)}(\cdot)/Q^{(i)}(\mathcal{G}^{i+1})$ being the joint distribution of $(X_i, \ldots, X_2, X_0)$ conditional on this event.

Given a probability measure Q on $\mathcal{G}^*$, we define, with a slight abuse of notation, probability measures ΣQ and $\Psi_j Q$ on $\mathcal{G}$ by

$$(\Sigma Q)[F] := Q[F \circ \Sigma] \quad \text{and} \quad (\Psi_j Q)[F] := Q[F \circ \Psi_j]$$

for a Borel function $F : \mathcal{G} \to \mathbb{R}_+$.

We now define analogues $\mathfrak{M}_n^*$, $\mathfrak{S}_n^*$, $\mathfrak{R}^*$, and $\mathfrak{P}^*$ of the mutation, selection, recombination and complete Poissonization operators $\mathfrak{M}_n$, $\mathfrak{S}_n$, $\mathfrak{R}$, and $\mathfrak{P}$ with $\mathcal{G}$ replaced by $\mathcal{G}^*$.

NOTATION 7.6. Consider a probability measure Q on $\mathcal{G}^*$ thought of as the distribution of a finite sequence $(X_I, \ldots, X_0)$ of random measures.

Define the probability measure $\mathfrak{M}_n^* Q$ on $\mathcal{G}^*$ to be the distribution of the finite sequence $(X_{I+1}, X_I, \ldots, X_0)$, where X_{I+1} is a Poisson random measure with intensity measure ν/n that is independent of $(X_I, \ldots, X_0)$.

Similarly, define a new probability measure $\mathfrak{S}_n^* Q$ on $\mathcal{G}^*$ by setting

$$\begin{aligned}(\mathfrak{S}_n^* Q)^{(i)}[F] &:= \frac{\int_{\mathcal{G}^i} \exp\{-S(\Sigma(g))/n\} F(g)\, dQ^{(i)}(g)}{\int_{\mathcal{G}^*} \exp\{-S(\Sigma(g))/n\}\, dQ(g)} \\ &= \frac{Q[\exp\{-S(X_I + \cdots + X_0)/n\} F(X_I, \ldots, X_0) \mathbf{1}\{I = i\}]}{Q[\exp\{-S(X_I + \cdots + X_0)/n\}]}\end{aligned}$$

for a Borel function $F : \mathcal{G}^* \to \mathbb{R}_+$.

The probability measure $\mathfrak{R}^* Q$ is the distribution of

$$((X'_J)_R + (X''_J)_{R^c}, \ldots, (X'_0)_R + (X''_0)_{R^c}),$$

where the segregating set R is a pick from $\mathcal{R}$, J is independent of R and distributed as I, and, conditional on $J = i$, $(X'_J, \ldots, X'_0)$ and $(X''_J, \ldots, X''_0)$ are independent picks from the distribution of $(X_I, \ldots, X_0)$ under Q conditional on $I = i$.

The probability measure $\mathfrak{P}^* Q$ is the distribution of $(Y_J, \ldots, Y_0)$, where J is distributed as I and conditional on $J = i$ the random measures $Y_J, \ldots, Y_0$ are independent and Poisson, with the intensity measure associated with the conditional distribution of Y_k being the same as the intensity measure associated with the conditional distribution of X_k given $I = i$ for $0 \le k \le i$.

Observe that we have the four intertwining relations

$$\Sigma \mathfrak{M}_n^* = \mathfrak{M}_n \Sigma, \ \Sigma \mathfrak{S}_n^* = \mathfrak{S}_n \Sigma, \ \Sigma \mathfrak{R}^* = \mathfrak{R} \Sigma, \Sigma \mathfrak{P}^* = \mathfrak{P} \Sigma.$$

These equalities confirm that "starred" operators agree with the "unstarred" ones once the labeling that identified in which generation a mutation occurred is removed.

We will think of our initial condition P_0 as a probability measure on $\mathcal{G}^*$ that is concentrated on sequences of length one.

NOTATION 7.7. Put

$$P_k := (\mathfrak{M}_n)^k P_0. \tag{7.5}$$

Set

$$\begin{aligned} Q_k^* &:= (\mathfrak{R}^* \mathfrak{M}_n^* \mathfrak{S}_n^*)^k P_0, \\ P_k^* &:= (\mathfrak{M}_n^*)^k P_0. \end{aligned} \tag{7.6}$$

Note that both probability measures Q_k^* and P_k^* are concentrated on sequences of length $(k+1)$. Note also that if $(X_k,\dots,X_0)$ is distributed according to P_k^*, then $X_1,\dots,X_k$ are independent Poisson random measures, each with intensity measure ν/n, and X_0 is independent with distribution P_0. As we expect, $Q_k = \Sigma Q_k^*$ and $P_k = \Sigma P_k^*$, where we recall that $Q_k = (\mathfrak{R}\mathfrak{M}_n\mathfrak{S}_n)^k P_0$ and $P_k := (\mathfrak{M}_n)^k P_0$. If $P_0 = \Pi_{\rho_0}$ (that is, P_0 is the distribution of a Poisson random measure with intensity measure $\mu P_0 = \rho_0$), then P_k^* is the distribution of a $(k+1)$-tuple $(X_k,\dots,X_0)$ of independent Poisson random measures on $\mathcal{M}$, where X_j has intensity measure $\mu\Psi_j P_k^*$ given by ρ_0 when $j=0$, and ν/n otherwise. In this case, P_k is the distribution of a Poisson random measure with intensity $\frac{k}{n}\nu + \rho_0$.

LEMMA 7.8. *For a positive integer k, and sequence of measures $(g_k,\dots,g_0) \in \mathcal{G}^*$,*

$$\begin{aligned}&\frac{dQ_k^*}{dP_k^*}(g_k,\dots,g_0)\\ &\qquad = \int \exp\left\{-\frac{1}{n}\sum_{i=0}^{k-1} \hat{S}_{Q_i,\mathcal{A}_i}(g_i+\cdots+g_0)\right\} d\mathcal{R}(A_1)\cdots d\mathcal{R}(A_k),\end{aligned}$$

where the partition $\mathcal{A}_i$ in the integral is $\{A_{i+1},A_{i+1}^c\}\wedge\cdots\wedge\{A_k,A_k^c\}$.

PROOF. This follows by repeated application of Lemma 7.4, and the fact that $\mathfrak{R}^*$ commutes with $\mathfrak{M}_n^*$. □

LEMMA 7.9. *Suppose that $P_0 = \Pi_{\rho_0}$. For P_k^*-almost every $g \in \mathcal{G}^*$,*

$$\begin{aligned}&\exp\left\{-\frac{\sigma k}{n}\Sigma g(\mathcal{M}) - \big(e^{\sigma k/n}-1\big)\mu P_k(\mathcal{M})\right\}\\ &\quad\le \frac{dQ_k^*}{dP_k^*}(g)\\ &\quad\le \exp\left\{\frac{\sigma k}{n}\Sigma g(\mathcal{M}) + \big(1-e^{-\sigma k/n}\big)\mu P_k(\mathcal{M})\right\}.\end{aligned}$$

PROOF. Recall that $\mu P_k = \rho_0 + \frac{k}{n}\nu$. By Lemmas 7.5 and 7.8, for any $g \in \mathcal{G}^*$,

$$\begin{aligned}\left|\log\frac{dQ_k^*/dP_k^*(g)}{dQ_k^*/dP_k^*(0)}\right| &\le \frac{k}{n}\sup_{\mathcal{A}}\sup_{0\le i\le k}\left|\hat{S}_{Q_i,\mathcal{A}}(g_0+\cdots+g_i)\right|\\ &\le \frac{\sigma k\Sigma g(\mathcal{M})}{n}.\end{aligned}$$

Thus,

$$\begin{aligned}\frac{dQ_k^*}{dP_k^*}(g) &= \frac{\frac{dQ_k^*/dP_k^*(g)}{dQ_k^*/dP_k^*(0)}}{\int_{\mathcal{G}}\frac{dQ_k^*/dP_k^*(h)}{dQ_k^*/dP_k^*(0)}\,dP_k^*(h)}\\ &\le \frac{e^{\sigma k\Sigma g(\mathcal{M})/n}}{\int_{\mathcal{G}^*} e^{-\sigma k\Sigma h(\mathcal{M})/n}\,dP_k^*(h)}\\ &= \exp\left\{\sigma\frac{k}{n}\Sigma g(\mathcal{M})\right\}\exp\left\{\left(1-e^{-\sigma k/n}\right)\mu P_k(\mathcal{M})\right\}.\end{aligned}$$

The lower bound is similar. □

For the next result, we recall Campbell's Theorem [**DVJ88**, (6.4.11)], which says that

$$\int_{\mathcal{G}} F(g)g[f]\,d\Pi_\pi(g) = \int_{\mathcal{M}} \Pi_\pi[F(\cdot + \delta_x)]f(x)\,d\pi(x)$$

for a finite measure π on $\mathcal{M}$. In particular,

$$\int_{\mathcal{G}} g[h]\exp\{-cg(\mathcal{M})\}\,d\Pi_\pi(g) = \exp\{-(1-e^{-c})\pi(\mathcal{M})\}\int_{\mathcal{M}} \exp\{-c\}h(x)\,d\pi(x)$$

for any constant $c \geq 0$.

COROLLARY 7.10. *Suppose that $P_0 = \Pi_{\rho_0}$. For $0 \leq j \leq k < \infty$ and μP_k-almost every $x \in \mathcal{M}$,*

$$\begin{aligned}
&\exp\Big\{-\big(e^{\sigma k/n} - e^{-\sigma k/n}\big)\mu P_k(\mathcal{M}) - \sigma k/n\Big\} \\
&\quad\leq \frac{d\mu\Psi_j Q_k^*}{d\mu\Psi_j P_k^*}(x) \\
&\quad\leq \exp\Big\{\big(e^{\sigma k/n} - e^{-\sigma k/n}\big)\mu P_k(\mathcal{M}) + \sigma k/n\Big\}.
\end{aligned}$$

PROOF. By the first inequality of Lemma 7.9, we have for a Borel function $h : \mathcal{M} \to \mathbb{R}_+$ that

$$\begin{aligned}
\mu\Psi_j Q_k^*[h] &= \int_{\mathcal{G}^*} \Psi_j g[h]\,dQ_k^*(g) \\
&= \int_{\mathcal{G}^*} \Psi_j g[h]\frac{dQ_k^*}{dP_k^*}(g)\,dP_k^*(g) \\
&\geq \int_{\mathcal{G}^*} \Psi_j g[h]\exp\Big\{-\frac{\sigma k}{n}\Sigma g(\mathcal{M}) - \big(e^{\sigma k/n} - 1\big)\mu P_k(\mathcal{M})\Big\}\,dP_k^*(g) \\
&= \exp\Big\{-\big(e^{\sigma k/n} - 1\big)\mu P_k(\mathcal{M})\Big\} \\
&\quad\times \int_{\mathcal{G}^*} \Psi_j g[h]\exp\Big\{-\frac{\sigma k}{n}\Psi_j g(\mathcal{M})\Big\}\exp\Big\{-\frac{\sigma k}{n}\sum_{i\neq j}\Psi_i g(\mathcal{M})\Big\}\,dP_k^*(g) \\
&= \exp\Big\{-\big(e^{\sigma k/n} - 1\big)\mu P_k(\mathcal{M})\Big\} \\
&\quad\times \exp\Big\{-\big(1 - e^{-\sigma k/n}\big)\mu\Psi_j P_k^*(\mathcal{M})\Big\}\int_{\mathcal{M}} \exp\Big\{-\frac{\sigma k}{n}\Big\}h(x)\,d\mu\Psi_j P_k^*(x) \\
&\quad\times \exp\Big\{-\big(1 - e^{-\sigma k/n}\big)\sum_{i\neq j}\mu\Psi_j P_k^*(\mathcal{M})\Big\} \\
&= \exp\Big\{-\big(e^{\sigma k/n} - e^{-\sigma k/n}\big)\mu P_k(\mathcal{M}) - \sigma k/n\Big\}\mu\Psi_j P_k^*[h].
\end{aligned}$$

The first inequality of the result follows. The proof of the second inequality is similar. □

The next result is an elementary consequence of Corollary 7.10.

Corollary 7.11. *Suppose that* $P_0 = \Pi_{\rho_0}$. *For* $0 \le k < \infty$ *and* μP_k*-almost every* $x \in \mathcal{M}$,

$$\begin{aligned}&\exp\Big\{-\big(e^{\sigma k/n} - e^{-\sigma k/n}\big)\mu P_k(\mathcal{M}) - \sigma k/n\Big\}\\ &\le \frac{d\mu Q_k}{d\mu P_k}(x)\\ &\le \exp\Big\{\big(e^{\sigma k/n} - e^{-\sigma k/n}\big)\mu P_k(\mathcal{M}) + \sigma k/n\Big\}.\end{aligned}$$

7.3. Palm measures

For various probability measures Q on $\mathcal{G}$, we will need to understand something about the probability measure that corresponds to conditioning a pick from Q to place a point at the locus $x \in \mathcal{M}$. When the intensity measure μQ is finite, the object that makes rigorous sense of such conditioning is the corresponding *Palm measure* $\mathcal{P}_Q^{(x)}$. For a Borel set $A \subseteq \mathcal{G}$, $\mathcal{P}_Q^{(x)}(A)$ is the Radon–Nikodym derivative

$$\frac{d\mu_A Q}{d\mu Q}(x),$$

where $\mu_A Q$ is the finite measure on $\mathcal{M}$ defined by

$$\mu_A Q[f] := \int_A g[f]\, dQ(g)$$

(see, for example, [**Kal84**, (1.7)]). The Palm measure is defined for μQ-almost every $x \in \mathcal{M}$. If Q is the distribution Π_π of a Poisson random random measure with finite intensity measure π, then $\mathcal{P}_Q^{(x)}[F] = Q[F(\cdot + \delta_x)]$ for π-almost every $x \in \mathcal{M}$ (see, for example, [**DVJ88**, Example 12.1(b)]). It follows from the definition of the Palm measure that if $F : \mathcal{G} \to \mathbb{R}_+$ and $f : \mathcal{M} \to \mathbb{R}_+$ are Borel functions, then

$$\int_{\mathcal{G}} F(g) g[f]\, dQ(g) = \int_{\mathcal{M}} \mathcal{P}_Q^{(x)}[F] f(x)\, d\mu Q(x).$$

Note that Campbell's theorem follows from applying this observation when Q is the distribution of a Poisson random measure.

Corollary 7.12. *Suppose that* $P_0 = \Pi_{\rho_0}$. *For every positive integer* k, μQ_k*-almost every* $x \in \mathcal{M}$ *(equivalently,* μP_k*-almost every* $x \in \mathcal{M}$ *or* $(\rho_0 + \nu)$*-almost every* $x \in \mathcal{M}$*), and every Borel function* $F : \mathcal{G} \to \mathbb{R}_+$,

$$\begin{aligned}\mathcal{P}_{Q_k}^{(x)}[F] \le \exp&\left\{\frac{2\sigma k}{n} + \big(e^{\sigma k/n} + 1 - 2e^{-\sigma k/n}\big)\mu P_k(\mathcal{M})\right\}\\ &\times \int_{\mathcal{G}} F(g + \delta_x) \exp\left\{\frac{\sigma k}{n} g(\mathcal{M})\right\} dP_k(g).\end{aligned}$$

Proof. By Corollary 7.11,

$$\frac{d\mu P_k}{d\mu Q_k}(x) \le \exp\Big\{\big(e^{\sigma k/n} - e^{-\sigma k/n}\big)\mu P_k(\mathcal{M}) + \sigma k/n\Big\} \tag{7.7}$$

for $(\rho_0 + \nu)$-almost every x.

Consider Borel functions $F : \mathcal{G} \to \mathbb{R}_+$ and $f : \mathcal{M} \to \mathbb{R}_+$. From the second inequality of Lemma 7.9, the inequality (7.7), and the characterization of the Palm measure of a Poisson random measure we have

$$\begin{aligned}
&\int_{\mathcal{M}} \mathcal{P}_{Q_k}^{(x)}[F] f(x)\, d\mu Q_k(x) \\
&\quad = \int_{\mathcal{G}} F(g) g[f]\, dQ_k(g) \\
&\quad = \int_{\mathcal{G}} F(g) g[f] \frac{dQ_k}{dP_k}(g)\, dP_k(g) \\
&\quad = \int_{\mathcal{M}} \left(\int_{\mathcal{G}} F(g) \frac{dQ_k}{dP_k}(g)\, d\mathcal{P}_{P_k}^{(x)}(g) \right) f(x)\, d\mu P_k(x) \\
&\quad \le \int_{\mathcal{M}} \left(\int_{\mathcal{G}} F(g+\delta_x) \exp\left\{ \frac{\sigma k}{n}(g+\delta_x)(\mathcal{M}) + \big(1 - e^{-\sigma k/n}\big)\mu P_k(\mathcal{M}) \right\} dP_k(g) \right) \\
&\qquad \times f(x)\, d\mu P_k(x) \\
&\quad = \exp\left\{ \frac{\sigma k}{n} + \big(1 - e^{-\sigma k/n}\big)\mu P_k(\mathcal{M}) \right\} \\
&\qquad \times \int_{\mathcal{M}} \left(\int_{\mathcal{G}} F(g+\delta_x) \exp\left\{ \frac{\sigma k}{n}(g)(\mathcal{M}) \right\} dP_k(g) \right) f(x) \frac{d\mu P_k}{d\mu Q_k}(x)\, d\mu Q_k(x) \\
&\quad \le \exp\left\{ \frac{\sigma k}{n} + \big(1 - e^{-\sigma k/n}\big)\mu P_k(\mathcal{M}) \right\} \exp\Big\{ \big(e^{\sigma k/n} - e^{-\sigma k/n}\big)\mu P_k(\mathcal{M}) + \frac{\sigma k}{n} \Big\} \\
&\qquad \times \int_{\mathcal{M}} \left(\int_{\mathcal{G}} F(g+\delta_x) \exp\left\{ \frac{\sigma k}{n}(g)(\mathcal{M}) \right\} dP_k(g) \right) f(x)\, d\mu Q_k(x) \\
&\quad = \exp\left\{ \frac{2\sigma k}{n} + \big(e^{\sigma k/n} + 1 - 2e^{-\sigma k/n}\big)\mu P_k(\mathcal{M}) \right\} \\
&\qquad \times \int_{\mathcal{M}} \left(\int_{\mathcal{G}} F(g+\delta_x) \exp\left\{ \frac{\sigma k}{n}(g)(\mathcal{M}) \right\} dP_k(g) \right) f(x)\, d\mu Q_k(x),
\end{aligned}$$

and the result follows. □

7.4. Comparisons with complete Poissonization

Recall from Notation 6.1 that the complete Poissonization operator $\mathfrak{P}$ acts on a probability measure P on $\mathcal{G}$ by $\mathfrak{P}P := \Pi_{\mu P}$. Also, recall from Notation 6.2 that $Q'_k := (\mathfrak{P}\mathfrak{S}_n\mathfrak{M}_n)^k P_0$ is the analogue of the sequence of probability measures $Q_k := (\mathfrak{R}\mathfrak{S}_n\mathfrak{M}_n)^k P_0$ from Notation 4.4 that is of primary interest to us in Theorem 5.4, with the recombination operator $\mathfrak{R}$ replaced by the complete Poissonization operator $\mathfrak{P}$. Finally, recall from Notation 7.7 that the counterpart of Q_k for $\mathcal{G}^*$ (that is, for the setting in which we keep track of the generation in which mutations from the ancestral wild type occurred) is $Q_k^* = (\mathfrak{R}^*\mathfrak{S}_n^*\mathfrak{M}^*)^k P_0$.

The next result is a consequence of Corollary 7.10 and the observation that if π' and π'' are two finite measures on $\mathcal{M}$ and π' is absolutely continuous with respect to π'', then

$$\frac{d\Pi_{\pi'}}{d\Pi_{\pi''}}(g) = \exp\left\{ g\left[\log\left(\frac{d\pi'}{d\pi''} \right) \right] - \pi'(\mathcal{M}) + \pi''(\mathcal{M}) \right\}.$$

COROLLARY 7.13. *Suppose that* $P_0 = \Pi_{\rho_0}$. *For* P_k^**-almost every* $g \in \mathcal{G}^*$,

$$\begin{aligned}
&\exp\Big\{\Big(-\big(e^{\sigma k/n} - e^{-\sigma k/n}\big)\mu P_k(\mathcal{M}) - \sigma k/n\Big)\,\Sigma g(\mathcal{M}) - \mu Q_k(\mathcal{M}) + \mu P_k(\mathcal{M})\Big\} \\
&\quad \le \frac{d\mathfrak{P}^* Q_k^*}{dP_k^*}(g) \\
&\quad \le \exp\Big\{\Big(\big(e^{\sigma k/n} - e^{-\sigma k/n}\big)\mu P_k(\mathcal{M}) + \sigma k/n\Big)\,\Sigma g(\mathcal{M}) - \mu Q_k(\mathcal{M}) + \mu P_k(\mathcal{M})\Big\}.
\end{aligned}$$

LEMMA 7.14. *Suppose that* $Q = \Pi_\pi$ *for some finite measure* π *on* $\mathcal{M}$. *The probability measure* $\mathfrak{P}\mathfrak{S}_n Q$ *on* $\mathcal{G}$ *is absolutely continuous with respect to* Q *with Radon-Nikodym derivative*

$$\frac{d\mathfrak{P}\mathfrak{S}_n Q}{dQ}(g) = e^{-\hat{S}'_Q(g)/n}, \tag{7.8}$$

where

$$\hat{S}'_Q(g) := g[s_Q] - n\frac{\int_{\mathcal{G}} h(\mathcal{M}) e^{-S(h)/n}\, dQ(h)}{\int_{\mathcal{G}} e^{-S(h)/n}\, dQ(h)} + n\int_{\mathcal{G}} h(\mathcal{M})\, dQ(h), \tag{7.9}$$

with

$$\begin{aligned}
s_Q(x) :&= -n\log\int_{\mathcal{G}} e^{-S(h)/n}\, d\mathcal{P}_Q^{(x)}(h) + n\log\int_{\mathcal{G}} e^{-S(h)/n}\, dQ(h) \\
&= -n\log\int_{\mathcal{G}} e^{-S(h+\delta_x)/n}\, dQ(h) \;+\; n\log\int_{\mathcal{G}} e^{-S(h)/n}\, dQ(h) \qquad (7.10)\\
&= -n\log\left(\frac{d\mu(\mathfrak{P}\mathfrak{S}_n Q)}{d\pi}(x)\right).
\end{aligned}$$

PROOF. Consider (7.10). The first equality is simply the characterization of the Palm distribution for Poisson random measures that we have already noted. For the second equality, take a Borel function $\phi : \mathcal{M} \to \mathbb{R}_+$, and define $\Phi : \mathcal{G} \to \mathbb{R}_+$ by $\Phi(g) = g[f]$. Then,

$$\begin{aligned}
\mu\mathfrak{P}\mathfrak{S}_n[\phi] &= \mu\mathfrak{S}_n Q[\phi] \\
&= \mathfrak{S}_n Q[\Phi] \\
&= \frac{Q[e^{-S/n}\Phi]}{Q[e^{-S/n}]} \\
&= \int \frac{\mathcal{P}_Q^{(x)}[e^{-S/n}]}{Q[e^{-S/n}]}\phi(x)\, d\pi(x) \\
&= \int e^{-s_Q(x)/n}\phi(x)\, d\pi(x)
\end{aligned}$$

by Campbell's Theorem, and the equality follows.

Next consider (7.8). By (7.10),

$$\frac{d\mathfrak{P}\mathfrak{S}Q}{dQ}(g) = \frac{e^{-g[s_Q/n]}}{\int e^{-h[s_Q/n]}\, dQ(h)}.$$

Now,

$$\int e^{-h[s_Q/n]}\, dQ(h) = \exp\left\{-\int \left(e^{-s_Q(x)/n} - 1\right)\, d\pi(x)\right\},$$

and, by another application of Campbell's Theorem,

$$\int \left(e^{-s_Q(x)/n} - 1\right) d\pi(x) = \left(\int \frac{\mathcal{P}_Q^{(x)}[e^{-S/n}]}{Q[e^{-S/n}]}\, d\pi(x)\right) e^{-\pi(\mathcal{M})}$$

$$= \frac{\int e^{-S(h)/n} h(\mathcal{M}) dQ(h)}{\int e^{-S(h)/n}\, dQ(h)} e^{-\int h(\mathcal{M})\, dQ(h)}.$$

Putting these equations together proves (7.8). □

The analogue of Q_k^* with the complete Poissonization operator $\mathfrak{P}^*$ replacing the recombination operator $\mathfrak{R}^*$ (equivalently, the counterpart of Q'_k for $\mathcal{G}^*$) is

$$Q_k'^* := (\mathfrak{P}^* \mathfrak{S}_n^* \mathfrak{M}_n^*)^k P_0.$$

As expected, we have the intertwining relation $\Sigma Q_k'^* = Q'_k$.

Recall from Definition 2.3 that if π is a finite measure on $\mathcal{M}$ and $x \in \mathcal{M}$, then $F_\pi(x) := \mathbb{E}[S(X^\pi + \delta_x) - S(X^\pi)]$, where, as usual, X^π is Poisson random measure on $\mathcal{M}$ with intensity measure π (that is, X^π has distribution Π_π).

PROOF. Observe that

$$\left|s_Q(x) - F_\pi(x)\right| \le n \left|\log \int_{\mathcal{G}} e^{-S(h+\delta_x)/n}\, dQ(h) - \int_{\mathcal{G}} S(h+\delta_x)/n\, dQ(h)\right|$$

$$+ n \left|\log \int_{\mathcal{G}} e^{-S(h)/n}\, dQ(h) - \int_{\mathcal{G}} S(h)/n\, dQ(h)\right|,$$

and the result follows from Lemma A.1. □

Recall from Lemma 7.4 that $d\mathfrak{R}_{\mathcal{A}}\mathfrak{S}_n Q / d\mathfrak{R}_{\mathcal{A}} Q = \exp\{-\hat{S}_{Q,\mathcal{A}}/n\}$, where Q is any probability measure on $\mathcal{G}$, $\mathcal{A}$ is a partition of $\mathcal{M}$ into a finite number of Borel sets, $\mathfrak{R}_{\mathcal{A}}$ is the annealed recombination operator of (7.2), $\hat{S}_{Q,\mathcal{A}}$ is defined in Lemma 7.4. In particular, $\hat{S}_{Q,\mathcal{A}}$ depends implicitly on n.

COROLLARY 7.15. *Consider a finite Borel partition $\mathcal{A}$ of $\mathfrak{M}$. Suppose that Q a probability measure on $\mathcal{G}$ under which the masses of sets in the partition $\mathcal{A}$ are independent. Let $\widetilde{Q}$ be defined to have $d\widetilde{Q}/dQ = \exp\{-\hat{S}_{Q,\mathcal{A}}/n\}$. Then,*

$$\frac{d\mathfrak{P}\tilde{Q}}{d\mathfrak{P}Q} = e^{-\hat{S}'_{\tilde{Q}}/n}.$$

PROOF. By assumption, $Q = \mathfrak{R}_{\mathcal{A}} Q$. By Lemma 7.4 we may write

$$\widetilde{Q} = \mathfrak{R}_{\mathcal{A}} \mathfrak{S}_n Q$$

Thus,

$$\frac{d\mathfrak{P}\widetilde{Q}}{d\mathfrak{P}Q} = \frac{d\mathfrak{P}\mathfrak{S}_n Q}{d\mathfrak{P}Q} = e^{-\hat{S}'_{\tilde{Q}}/n}$$

by Lemma 7.14. □

LEMMA 7.16. *There is a constant κ depending on $\sigma := \sup_{g \in \mathcal{G}} S(g)$ but not on n such that for a probability measure Q on $\mathcal{G}$, a finite Borel partition $\mathcal{A}$ of $\mathcal{M}$, and* *a point* $g \in \mathcal{G}$,

$$\begin{aligned} &\left|\hat{S}_{Q,\mathcal{A}}(g) - \hat{S}'_Q(g)\right| \\ &\quad \le \kappa\biggl(\frac{1}{n}g(\mathcal{M}) + \frac{1}{n}Q[X(\mathcal{M})] \\ &\qquad + \sum_{A\in\mathcal{A}} Q\bigl\{X(A)>0\bigr\}^2 + \sum_{A\in\mathcal{A}} Q[X(A)\mathbf{1}_{\{X(A)\ge 2\}}] \\ &\qquad + \int_{\mathcal{M}} Q\{X(A(x))>0\}\,dg(x) + \int_{\mathcal{M}} \mathcal{P}_Q^{(x)}\bigl\{X(A(x))\ge 2\bigr\}\,dg(x) \\ &\qquad + g\left\{x\in\mathcal{M} : Q\{X(A(x))>0\} > \frac{1}{2}\right\} + \#\{A\in\mathcal{A} : g(A)\ge 2\}\biggr), \end{aligned}$$

where $A(x) \subseteq \mathcal{M}$ *is the unique set in the partition* $\mathcal{A}$ *that contains the point* $x \in \mathcal{M}$.

Proof. We begin by establishing a bound on the difference between $\hat{S}_{Q,\mathcal{A}}(g)$ and the quantity

$$\sum_{A\in\mathcal{A}} \bigl(Q\bigl[S \,\big|\, X_A = g_A\bigr] - Q\bigl[S \,\big|\, X_A = 0\bigr]\bigr) - \Gamma_Q,$$

where

$$\Gamma_Q := Q[S(X)X(\mathcal{M})] - Q[S(X)] - Q[X(\mathcal{M})].$$

By definition (see Lemma 7.4),

$$\begin{aligned} &\hat{S}_{Q,\mathcal{A}}(g) \\ &\quad = -n\sum_{A\in\mathcal{A}}\Bigl(\log Q\Bigl[e^{-S(X)/n} \,\big|\, X_A = g_A\Bigr] - \log Q\Bigl[e^{-S(X)/n} \,\big|\, X(A)=0\Bigr]\Bigr) \\ &\qquad - n\sum_{A\in\mathcal{A}}\Bigl(\log Q\Bigl[e^{-S(X)/n} \,\big|\, X(A)=0\Bigr] - \log Q\Bigl[e^{-S(X)/n}\Bigr]\Bigr). \end{aligned}$$

In the first sum, a summand can only be nonzero when $g(A) > 0$. By Lemma A.1, the first term is is thus approximately

$$\sum_{A\in\mathcal{A} \,:\, g(A)>0} \bigl(Q\bigl[S(X) \,\big|\, X_A = g_A\bigr] - Q\bigl[S(X) \,\big|\, X(A)=0\bigr]\bigr),$$

with an error bounded by

$$2n\frac{1}{2}\frac{\sigma^2}{n^2}e^{\sigma/n}g(\mathcal{M}) = \frac{\sigma^2}{n}e^{\sigma/n}g(\mathcal{M}).$$

The second term requires more care, because the sum extends over all (the potentially very large number of) sets of the partition $\mathcal{A}$. Fix one of these sets A. Then,

$$\begin{aligned} &\log Q\Bigl[e^{-S(X)/n} \,\big|\, X(A)=0\Bigr] - \log Q\Bigl[e^{-S(X)/n}\Bigr] \\ &\quad = \log\frac{Q\bigl[e^{-S(X)/n}\mathbf{1}_{\{X(A)=0\}}\bigr]}{Q\bigl[e^{-S(X)/n}\bigr]\,Q\{X(A)=0\}}. \end{aligned} \tag{7.11}$$

Note for $y \ge 1$ that

$$0 \le (y-1) - \log y = \int_1^y 1 - \frac{1}{z}\,dz \le \left(1-\frac{1}{y}\right)(y-1) = \frac{(y-1)^2}{y}, \tag{7.12}$$

and a similar argument shows that the same bound holds for $0 < y \leq 1$. That is, we have the approximation $\log(y) \approx y-1$, $y > 0$, with error bounds $y-1-(y-1)^2/y \leq \log(y) \leq y-1$.

Applying this approximation to the right-hand side of (7.11), the first-order term is

$$\begin{aligned}\frac{Q\left[e^{-S(X)/n}\mathbf{1}_{\{X(A)=0\}}\right]}{Q\left[e^{-S(X)/n}\right]Q\left\{X(A)=0\right\}} - 1 &= \frac{\operatorname{Cov}_Q[e^{-S(X)/n}, \mathbf{1}_{\{X(A)=0\}}]}{Q\left[e^{-S(X)/n}\right]Q\left\{X(A)=0\right\}} \\ &= \frac{\operatorname{Cov}_Q[1-e^{-S(X)/n}, \mathbf{1}_{\{X(A)>0\}}]}{Q\left[e^{-S(X)/n}\right]Q\left\{X(A)=0\right\}} \\ &\approx \frac{1}{n}\operatorname{Cov}_Q[S(X), \mathbf{1}_{\{X(A)>0\}}].\end{aligned} \tag{7.13}$$

The difference between the second and third quantities in (7.13) is

$$\begin{aligned}&\frac{Q\left[\left(1-e^{-S(X)/n}-S(X)/n-Q\left[1-e^{-S(X)/n}-S(X)/n\right]\right)\mathbf{1}_{\{X(A)>0\}}\right]}{Q\left[e^{-S(X)/n}\right]Q\left\{X(A)=0\right\}} \\ &\quad + Q\left[\left(S(X)/n - Q\left[S(X)/n\right]\right)\mathbf{1}_{\{X(A)>0\}}\right] \\ &\quad \times \left(\frac{1}{Q[e^{-S(X)/n}]Q\{X_A=0\}} - 1\right),\end{aligned} \tag{7.14}$$

From the inequalities $-y^2/2 \leq 1-e^{-y}-y \leq 0$, $y \geq 0$, the absolute value of (7.14) is bounded by

$$e^{\sigma/n}\frac{\sigma^2}{2n^2}\frac{Q\{X(A)>0\}}{Q\{X(A)=0\}} + e^{\sigma/n}\frac{\sigma}{n}Q\{X(A)>0\}\left(\frac{\sigma}{n}+\frac{Q\{X(A)>0\}}{Q\{X(A)=0\}}\right). \tag{7.15}$$

This bound is only useful when $Q\{X(A)>0\}$ is small, so we further note that the absolute value of (7.14) is also bounded by

$$\begin{aligned}&\left|\frac{\operatorname{Cov}_Q[1-e^{-S(X)/n}, \mathbf{1}_{\{X(A)=0\}}]}{Q\left[e^{-S(X)/n}\right]Q\left\{X(A)=0\right\}}\right| + \frac{1}{n}\left|\operatorname{Cov}_Q[S(X), \mathbf{1}_{\{X(A)>0\}}]\right| \\ &\quad \leq e^{\sigma/n}\frac{\sigma}{n} + \frac{\sigma}{n}.\end{aligned} \tag{7.16}$$

The contribution coming from the bound $(y-1)^2/y$ on the error in the first order approximation $\log(y) \approx y-1$ is bounded by

$$\begin{aligned}&\left(\frac{\operatorname{Cov}_Q[e^{-S(X)/n}, \mathbf{1}_{\{X(A)=0\}}]}{Q[e^{-S(X)/n}]Q\{X(A)=0\}}\right)^2 \frac{Q\left[e^{-S(X)/n}\right]Q\left\{X(A)=0\right\}}{Q\left[e^{-S(X)/n}\mathbf{1}_{\{X(A)=0\}}\right]} \\ &\quad = \frac{\operatorname{Cov}_Q[1-e^{-S(X)/n}, \mathbf{1}_{\{X(A)>0\}}]^2}{Q[e^{-S(X)/n}]Q\{X(A)=0\}Q\left[e^{-S(X)/n}\mathbf{1}_{\{X(A)=0\}}\right]} \\ &\quad \leq e^{2\sigma/n}\frac{\sigma^2}{n^2}\frac{Q\{X(A)>0\}^2}{Q\{X(A)=0\}^2}.\end{aligned} \tag{7.17}$$

Once again, this bound is only useful when $Q\{X(A)>0\}$ is small, so we further note that the left-hand side of (7.17) is also bounded by

$$e^{2\sigma/n}\frac{\sigma^2}{n^2}. \tag{7.18}$$

Putting these bounds together, we see that

$$\begin{aligned}&-n\left(\log Q\left[e^{-S(X)/n}\,\middle|\,X(A)=0\right]-\log Q\left[e^{-S(X)/n}\right]\right)\\&\qquad\approx \operatorname{Cov}_Q[S,\mathbf{1}_{\{X(A)>0\}}],\end{aligned}\tag{7.19}$$

with error bounded by

$$\begin{aligned}&e^{\sigma/n}\frac{\sigma^2}{n}Q\{X(A)>0\}\\&\quad+e^{\sigma/n}\sigma Q\{X(A)>0\}\left(\frac{\sigma}{n}+2Q\big\{X(A)>0\big\}\right)\\&\quad+4e^{2\sigma/n}\frac{\sigma^2}{n}Q\{X(A)>0\}^2,\end{aligned}\tag{7.20}$$

when $Q\{X(A)=0\}\geq 1/2$, and by

$$e^{\sigma/n}\sigma+\sigma+e^{2\sigma/n}\frac{\sigma^2}{n},\tag{7.21}$$

otherwise.

Note that

$$\begin{aligned}\left|\Gamma_Q-\sum_{A\in\mathcal{A}}\operatorname{Cov}_Q[S(X),\mathbf{1}_{\{X(A)>0\}}]\right|&=\left|\sum_{A\in\mathcal{A}}\operatorname{Cov}_Q[S(X),X(A)-\mathbf{1}_{\{X(A)>0\}}]\right|\\&\leq\sigma\sum_{A\in\mathcal{A}}Q[X(A)\mathbf{1}_{\{X(A)\geq 2\}}]\end{aligned}\tag{7.22}$$

Summing (7.19), (7.20), and (7.21) over $A\in\mathcal{A}$ and combining the result with (7.22), we see that

$$\hat{S}_{Q,\mathcal{A}}(g)\approx\sum_{A\in\mathcal{A}}\big(Q\big[S\,\big|\,X_A=g_A\big]-Q\big[S\,\big|\,X_A=0\big]\big)-\Gamma_Q\tag{7.23}$$

with an error bounded by

$$\begin{aligned}&e^{\sigma/n}\frac{\sigma^2}{n}g(\mathcal{M})+2e^{\sigma/n}\frac{\sigma^2}{n}Q[X(\mathcal{M})]\\&\quad+\left(2e^{\sigma/n}\sigma+4e^{2\sigma/n}\frac{\sigma^2}{n}+4\left(e^{\sigma/n}\sigma+\sigma+e^{2\sigma/n}\frac{\sigma^2}{n}\right)\right)\sum_{A\in\mathcal{A}}Q\big\{X(A)>0\big\}^2\\&\quad+\sigma\sum_{A\in\mathcal{A}}Q[X(A)\mathbf{1}_{\{X(A)\geq 2\}}],\end{aligned}\tag{7.24}$$

where we used the inequality

$$\mathbf{1}\Big\{Q\{X(A)=0\}\leq\frac{1}{2}\Big\}\leq 4Q\{X(A)>0\}^2.$$

We now bound the difference between $\hat{S}'_Q(g)$ and the approximation to $\hat{S}_{Q,\mathcal{A}}(g)$ in (7.23). We stress that here we are not assuming that Q is the distribution of a Poisson random measure, and so, in the definition of $\hat{S}'_Q(g)$ from (7.9), namely

$$\hat{S}'_Q(g)=g[s_Q]-n\frac{Q[X(\mathcal{M})e^{-S(X)/n}]}{Q[e^{-S(X)/n}]}+nQ[X(\mathcal{M})],$$

the definition of the function s_Q is the one appearing in the first line of (7.10), namely

$$s_Q(x) = -n\log \mathcal{P}_Q^{(x)}[e^{-S(X)/n}] + n\log Q[e^{-S(X)/n}],$$

rather than the definitions on the other lines of (7.10) that are only equivalent in the Poisson case. The absolute value of the difference is at most

$$\begin{aligned}&\sum_{A\in\mathcal{A}:g(A)=1} \left|g_A[s_Q] - \big(Q\big[S\,\big|\,X_A = g_A\big] - Q\big[S\,\big|\,X_A=0\big]\big)\right|\\ &\quad+\sum_{A\in\mathcal{A}:g(A)>1} \left|g_A(s_Q) - \big(Q\big[S\,\big|\,X_A = g_A\big] - Q\big[S\,\big|\,X_A=0\big]\big)\right|\\ &\quad+\left|-n\frac{Q[X(\mathcal{M})e^{-S(X)/n}]}{Q[e^{-S(X)/n}]} + nQ[X(\mathcal{M})] + \Gamma_Q\right|.\end{aligned}\tag{7.25}$$

We will consider each of the terms in (7.25) in turn.

Note first of all that, by Lemma A.1,

$$\begin{aligned}&\left|n\log Q[e^{-S(X)/n}] + Q\big[S(X)\,\big|\,X_A=0\big]\right|\\ &\quad\le \left|Q[S(X)] - \frac{Q[S(X)\mathbf{1}_{\{X(A)=0\}}]}{Q\{X(A)=0\}}\right| + e^{\sigma/n}\frac{\sigma^2}{2n}\\ &\quad\le \sigma\frac{Q\{X(A)>0\}}{Q\{X(A)=0\}} + e^{\sigma/n}\frac{\sigma^2}{2n}\\ &\quad\le 2\sigma Q\{X(A)>0\} + e^{\sigma/n}\frac{\sigma^2}{2n}\end{aligned}\tag{7.26}$$

if $Q\{X(A)>0\}\le\frac{1}{2}$, and otherwise the left-hand side is bounded by σ.

Note next, that, by definition of the Palm measure, for any Borel subset $B\subseteq\mathcal{G}$ we have

$$\left|\mathcal{P}_Q^{(x)}(B) - Q\left(B\,\middle|\,X_A=\delta_x\right)\right| \le \mathcal{P}_Q^{(x)}\{X(A)\ge 2\}$$

for μQ-almost every $x\in\mathcal{M}$. Combining this observation with Lemma A.1 gives

$$\begin{aligned}&\left|n\log\mathcal{P}_Q^{(x)}[e^{-S(X)/n}] + Q\big[S(X)\,\big|\,X_A=\delta_x\big]\right|\\ &\quad\le e^{\sigma/n}\frac{\sigma^2}{2n} + \sigma\mathcal{P}_Q^{(x)}\big\{X(A)\ge 2\big\}.\end{aligned}\tag{7.27}$$

Integrating the inequalities (7.26) and (7.27) against the measure g_A and summing over A yields that the first sum in (7.25) is at most

$$
\begin{aligned}
&\sum_{A\in\mathcal{A}:g(A)=1}\int_{\mathcal{M}} \bigl|s_Q(x) - \bigl(Q\bigl[S \bigm| X_A=\delta_x\bigr] - Q\bigl[S \bigm| X_A=0\bigr]\bigr)\bigr|\; dg_A(x)\\
&\le \sum_{A\in\mathcal{A}:g(A)=1}\int_A 2\sigma Q\{X(A)>0\} + e^{\sigma/n}\frac{\sigma^2}{2n} + \sigma\mathbf{1}\left\{Q\{X(A)>0\}>\frac{1}{2}\right\}\\
&\qquad + e^{\sigma/n}\frac{\sigma^2}{2n} + \sigma\mathcal{P}_Q^{(x)}\bigl\{X(A)\ge 2\bigr\}\,dg_A(x)\\
&\le e^{\sigma/n}\frac{\sigma^2}{n}g(\mathcal{M})\\
&\qquad + \int_{\mathcal{M}} 2\sigma Q\{X(A(x))>0\} + \sigma\mathcal{P}_Q^{(x)}\bigl\{X(A(x))\ge 2\bigr\}\,dg(x)\\
&\qquad + \sigma g\left\{x\in\mathcal{M} : Q\{X(A(x))>0\}>\frac{1}{2}\right\}.
\end{aligned}
\tag{7.28}
$$

Turning to the second sum in (7.25), we have

$$
\Bigl|n\log\mathcal{P}_Q^{(x)}[e^{-S(X)/n}] + Q\bigl[S(X)\bigm| X_A=g_A\bigr]\Bigr| \le \sigma.
\tag{7.29}
$$

Combining this with (7.26), integrating with respect to g_A, and summing over A yields that the second sum in (7.25) is at most

$$
\begin{aligned}
&e^{\sigma/n}\frac{\sigma^2}{2n}g(\mathcal{M}) + \int_{\mathcal{M}} 2\sigma Q\{X(A(x))>0\}\,dg(x)\\
&\quad + \sigma g\left\{x\in\mathcal{M} : Q\{X(A(x))>0\}>\frac{1}{2}\right\}\\
&\quad + \sigma\#\{A\in\mathcal{A} : g(A)\ge 2\}.
\end{aligned}
\tag{7.30}
$$

The third term in (7.25) is

$$
\begin{aligned}
&\left|-n\frac{Q[X(\mathcal{M})e^{-S(X)/n}]}{Q[e^{-S(X)/n}]} + nQ[X(\mathcal{M})] + Q[X(\mathcal{M})S(X)] - Q[X(\mathcal{M})]Q[S(X)]\right|\\
&\quad\le \left|-n\left(Q[X(\mathcal{M})e^{-S(X)/n}] - Q[X(\mathcal{M})]Q[e^{-S(X)/n}]\right)\left(\frac{1}{Q[e^{-S(X)/n}]}-1\right)\right|\\
&\qquad + \Bigl|-n\left(Q[X(\mathcal{M})e^{-S(X)/n}] - Q[X(\mathcal{M})]Q[e^{-S(X)/n}]\right)\\
&\qquad\quad - (Q[X(\mathcal{M})S(X)] - Q[X(\mathcal{M})]Q[S(X)])\Bigr|\\
&\quad= \left|nQ\left[(X(\mathcal{M}) - Q[X(\mathcal{M})])(1-e^{-S(X)/n})\right]\frac{1-Q[e^{-S(X)/n}]}{Q[e^{-S(X)/n}]}\right|\\
&\qquad + \left|nQ\left[(X(\mathcal{M}) - Q[X(\mathcal{M})])(1-e^{-S(X)/n} - S(x)/n)\right]\right|\\
&\quad\le Q[X(\mathcal{M})]\left(e^{\sigma/n}\frac{\sigma^2}{n} + \frac{1}{2}\frac{\sigma^2}{n}\right).
\end{aligned}
\tag{7.31}
$$

It follows from substituting the bounds (7.28), (7.30), and (7.31) into (7.25) that the difference between $\hat{S}'_Q(g)$ and the approximation to $\hat{S}_{Q,\mathcal{A}}(g)$ in (7.23) is bounded in absolute value by

$$
\begin{aligned}
& e^{\sigma/n}\frac{3\sigma^2}{2n}g(\mathcal{M}) \\
& + \int_{\mathcal{M}} 4\sigma Q\{X(A(x))>0\} + \sigma\mathcal{P}_Q^{(x)}\big\{X(A(x))\geq 2\big\}\,dg(x) \\
& + 2\sigma g\left\{x\in\mathcal{M} : Q\{X(A(x))>0\} > \frac{1}{2}\right\} \\
& + \sigma\#\{A\in\mathcal{A} : g(A)\geq 2\} \\
& + \left(e^{\sigma/n}\frac{\sigma^2}{n} + \frac{1}{2}\frac{\sigma^2}{n}\right)Q[X(\mathcal{M})].
\end{aligned}
\tag{7.32}
$$

The result follows upon combining this last bound with the bound (7.24) on the absolute value of the difference between $\hat{S}_{Q,\mathcal{A}}(g)$ and the approximation to $\hat{S}_{Q,\mathcal{A}}(g)$ in (7.23). □

CHAPTER 8

Proof of convergence of the discrete-generation systems

With Proposition 6.3 in hand, it will suffice to show for a number of generations k with $0 \le k \le Tn$ that $Q_k = (\mathfrak{R}\mathfrak{M}_n\mathfrak{S}_n)^k P_0$, the probability measure that describes a population after k rounds of selection, mutation, and recombination, is close to $Q'_k = (\mathfrak{P}\mathfrak{M}_n\mathfrak{S}_n)^k P_0$, the probability measure that describes a population after k rounds of selection, mutation, and complete Poissonization.

By Lemma 2.7,

$$\begin{aligned} \|Q_k - Q'_k\|_{\mathrm{Was}} &\le \|Q_k - \mathfrak{P}Q_k\|_{\mathrm{Was}} + \|\mathfrak{P}Q_k - Q'_k\|_{\mathrm{Was}} \\ &\le \|Q_k - \mathfrak{P}Q_k\|_{\mathrm{Was}} + 4\|\mu Q_k - \mu Q'_k\|_{\mathrm{Was}}, \end{aligned} \tag{8.1}$$

where we recall that μQ_k and $\mu Q'_k$ are the intensity measures of Q_k and Q'_k.

Set

$$a_k := \|Q_k - \mathfrak{P}Q_k\|_{\mathrm{Was}}$$

and

$$b_k := \|\mu Q_k - \mu Q'_k\|_{\mathrm{Was}}.$$

Given a Borel function $f : \mathfrak{M} \to [-1, 1]$, define the function $F : \mathcal{G} \to \mathbb{R}$ by $F(g) := g[f]$. Note that

$$Q_{k+1}[F] = \mu Q_{k+1}[f] = \mu(\mathfrak{R}\mathfrak{M}_n\mathfrak{S}_n)Q_k[f] = \frac{\nu[f]}{n} + \mathfrak{S}_n Q_k[f],$$

where we use the facts that, for any probability measure Q on $\mathcal{G}$, $\mu\mathfrak{R}Q = \mu Q$ and $\mu\mathfrak{M}_n Q = \mu Q + \nu/n$. Thus, from the inequalities $-y^2/2 \le 1 - e^{-y} - y \le 0$, $y \ge 0$,

$$\begin{aligned} b_{k+1} &\le \sup\left\{\left|\mathfrak{S}_n Q_k[f] - \mathfrak{S}_n Q'_k[f]\right| : \|f\|_{\mathrm{Lip}} \le 1\right\} \\ &\le e^{\sigma/n}\left|Q_k\left[fe^{-S/n}\right] - Q'_k\left[fe^{-S/n}\right]\right| \\ &\le e^{\sigma/n}\left(|Q_k[f] - Q'_k[f]| + \frac{1}{n}\left|Q_k[S\cdot f] - Q'_k[S\cdot f]\right| + \frac{2\sigma^2}{n^2}\right) \\ &\le e^{\sigma/n}\left(b_k + \frac{2\sigma}{n}a_k + \frac{2\sigma^2}{n^2}\right). \end{aligned} \tag{8.2}$$

We now bound a_k. It follows from the relations $\Sigma Q_k^* = Q_k$, $\Sigma \mathfrak{P}^* = \mathfrak{P}\Sigma$, and $\Sigma P_k^* = P_k$ that

$$\begin{aligned}\bigl\|Q_k - \mathfrak{P}Q_k\bigr\|_{\mathrm{Was}} &= \sup\{|Q_k[\phi] - \mathfrak{P}Q_k[\phi]| : \|\phi\|_{\mathrm{Lip}} \le 1\}\\ &\le \sup\{|Q_k[\phi] - \mathfrak{P}Q_k[\phi]| : \|\phi\|_\infty \le 1\}\\ &= \sup\{|Q_k^*[\phi\circ\Sigma] - \mathfrak{P}^*Q_k^*[\phi\circ\Sigma]| : \|\phi\|_\infty \le 1\}\\ &\le \sup\{|Q_k^*[\phi^*] - \mathfrak{P}^*Q_k^*[\phi^*]| : \|\phi^*\|_\infty \le 1\}\\ &= \int \left|\frac{dQ_k^*}{dP_k^*}(g) - \frac{d\mathfrak{P}^*Q_k^*}{dP_k^*}(g)\right| \, dP_k^*(g)\end{aligned}$$

So, from the inequality $\log y \le y - 1$,

$$\begin{aligned}\bigl\|Q_k - \mathfrak{P}Q_k\bigr\|_{\mathrm{Was}} \le \int &\left|\log\frac{dQ_k^*}{dP_k^*}(g) - \log\frac{d\mathfrak{P}^*Q_k^*}{dP_k^*}(g)\right|\\ &\times \frac{dQ_k^*}{dP_k^*}(g) \vee \frac{d\mathfrak{P}^*Q_k^*}{dP_k^*}(g)\, dP_k^*(g)\end{aligned}$$

By Lemma 7.8,

$$\begin{aligned}&\frac{dQ_k^*}{dP_k^*}(g_k,\dots,g_0)\\ &\quad = \int \exp\left\{-\frac{1}{n}\sum_{i=0}^{k-1} \hat{S}_{Q_i,\mathcal{A}_i}(g_i + \dots + g_0)\right\} d\mathcal{R}(A_1)\cdots d\mathcal{R}(A_k),\end{aligned}$$

where the partition $\mathcal{A}_i$ in the integral is $\{A_{i+1}, A_{i+1}^c\} \wedge \dots \wedge \{A_k, A_k^c\}$, and, by Lemma 7.9,

$$\frac{dQ_k^*}{dP_k^*}(g) \le \exp\left\{\frac{\sigma k}{n}\Sigma g(\mathcal{M}) + \bigl(1 - e^{-\sigma k/n}\bigr)\mu P_k(\mathcal{M})\right\}.$$

By repeated application of Corollary 7.15,

$$\frac{d\mathfrak{P}^*Q_k^*}{dP_k^*}(g_k,\dots,g_0) = \exp\left\{-\frac{1}{n}\sum_{i=0}^{k-1} \hat{S}'_{Q_i}(g_0 + \dots + g_i)\right\},$$

and, by Corollary 7.13,

$$\begin{aligned}&\frac{d\mathfrak{P}^*Q_k^*}{dP_k^*}(g)\\ &\quad\le \exp\left\{\Bigl(\bigl(e^{\sigma k/n} - e^{-\sigma k/n}\bigr)\mu P_k(\mathcal{M}) + \sigma k/n\Bigr)\Sigma g(\mathcal{M}) - \mu Q_k(\mathcal{M}) + \mu P_k(\mathcal{M})\right\}.\end{aligned}$$

Observe that

$$\begin{aligned}&\left\{\frac{\sigma k}{n}\Sigma g(\mathcal{M}) + \bigl(1 - e^{-\sigma k/n}\bigr)\mu P_k(\mathcal{M})\right\}\\ &\quad \vee \left\{\Bigl(\bigl(e^{\sigma k/n} - e^{-\sigma k/n}\bigr)\mu P_k(\mathcal{M}) + \sigma k/n\Bigr)\Sigma g(\mathcal{M}) - \mu Q_k(\mathcal{M}) + \mu P_k(\mathcal{M})\right\}\\ &\qquad \le \Bigl(\bigl(e^{\sigma k/n} - e^{-\sigma k/n}\bigr)\mu P_k(\mathcal{M}) + \sigma k/n\Bigr)\Sigma g(\mathcal{M}) + \mu P_k(\mathcal{M})\\ &\qquad \le \Bigl(\bigl(e^{\sigma k/n} - e^{-\sigma k/n}\bigr)\bigl(\rho_0(\mathcal{M}) + \nu(\mathcal{M})k/n\bigr) + \sigma k/n\Bigr)\Sigma g(\mathcal{M})\\ &\qquad\quad + \bigl(\rho_0(\mathcal{M}) + \nu(\mathcal{M})k/n\bigr)\\ &\qquad \le \tilde{c}\Sigma g(\mathcal{M}) + \check{c},\end{aligned}$$

where $\tilde{c}$ and $\breve{c}$ are constants that depend on σ, T, $\rho_0(\mathcal{M})$, and $\nu(\mathcal{M})$. Therefore,

$$\begin{aligned} &\left\|Q_k - \mathfrak{P}Q_k\right\|_{\mathrm{Was}} \\ &\quad \leq \int \left| \frac{1}{n} \sum_{i=0}^{k-1} \hat{S}'_{Q_i}(g_0 + \cdots + g_i) - \frac{1}{n} \sum_{i=0}^{k-1} \hat{S}_{Q_i,\mathcal{A}_i}(g_0 + \cdots + g_i) \right| \\ &\qquad \times \exp\{\tilde{c}\Sigma g(\mathcal{M}) + \breve{c}\}\, dP_k^*(g)\, d\mathcal{R}(A_1) \cdots d\mathcal{R}(A_k) \\ &\quad \leq \frac{1}{n} \sum_{i=0}^{k-1} \int \left| \hat{S}'_{Q_i}(g) - \hat{S}_{Q_i,\mathcal{A}}(g) \right| \exp\{\tilde{c} g(\mathcal{M}) + \check{c}\}\, dP_i(g)\, d\mathcal{R}_{k-i}(\mathcal{A}), \end{aligned} \tag{8.3}$$

where $\check{c}$ is a constant that depends on σ, T, $\rho_0(\mathcal{M})$, and $\nu(\mathcal{M})$ such that

$$\int \exp\{\tilde{c}(g_{k-i+1} + \cdots + g_k)(\mathcal{M}) + \breve{c}\}\, d\Pi_{\nu/n}^{\otimes k-i}(g_k, \ldots, g_{k-i+1}) \leq \exp\{\check{c}\},$$

and we recall from Section 7.1 that $\mathcal{R}_{k-i}$ is the probability measure on partitions of $\mathcal{M}$ that is the push-forward of the product measure $\mathcal{R}^{\otimes(k-i)}$ by the map $(A_{i+1}, \ldots, A_k) \mapsto \{A_{i+1}, A_{i+1}^c\} \wedge \cdots \wedge \{A_k, A_k^c\}$.

Recall from Chapter 7.3 that $\mathcal{P}_Q^{(x)}$ denotes the Palm measure at $x \in \mathcal{M}$ of a probability measure Q on $\mathcal{G}$. Recall also for a finite measure π on $\mathcal{M}$ and bounded functions $f : \mathcal{M} \to \mathbb{R}_+$ and $h : \mathcal{M} \to \mathbb{R}_+$ that

$$\begin{gathered} \int \exp\{g[f]\}\, d\Pi_\pi(g) = \exp\{\pi[e^f - 1]\} \\ \int g[h] \exp\{g[f]\}\, d\Pi_\pi(g) = \pi[h e^f] \exp\{\pi[e^f - 1]\} \end{gathered} \tag{8.4}$$

From Lemma 7.16,

$$\begin{aligned} &\int \left| \hat{S}'_{Q_i}(g) - \hat{S}_{Q_i,\mathcal{A}}(g) \right| \exp\left\{\tilde{c} g(\mathcal{M}) + \check{c}\right\}\, dP_i(g) \\ &\quad \leq \int \kappa \Bigg(\frac{1}{n} g(\mathcal{M}) + \frac{1}{n} Q_i[X(\mathcal{M})] \\ &\qquad + \sum_{A \in \mathcal{A}} Q_i\big\{X(A) > 0\big\}^2 + \sum_{A \in \mathcal{A}} Q_i[X(A) \mathbf{1}_{\{X(A) \geq 2\}}] \\ &\qquad + \int_{\mathcal{M}} Q_i\{X(A(x)) > 0\}\, dg(x) + \int_{\mathcal{M}} \mathcal{P}_{Q_i}^{(x)}\big\{X(A(x)) \geq 2\big\}\, dg(x) \\ &\qquad + g\left\{x \in \mathcal{M} : Q_i\{X(A(x)) > 0\} > \frac{1}{2}\right\} + \#\{A \in \mathcal{A} : g(A) \geq 2\} \Bigg), \\ &\quad \times \exp\left\{\tilde{c} g(\mathcal{M}) + \check{c}\right\}\, dP_i(g), \end{aligned} \tag{8.5}$$

for any Borel partition $\mathcal{A}$ of $\mathcal{M}$, where $A(x)$ is the set in $\mathcal{A}$ that contains x.

We next break (8.5) up into a sum of integrals and, ignoring multiplicative constants that depend on σ, T, $\rho_0(\mathcal{M})$ and $\nu(\mathcal{M})$, estimate each integral individually. We will apply (8.4) repeatedly without explicit mention. We will write $c_1, c_2, \ldots$ for constants that depend on σ, T, $\rho_0(\mathcal{M})$ and $\nu(\mathcal{M})$, and sometimes use expectation rather than integral notation.

To begin,

$$\frac{1}{n} P_i\left[X(\mathcal{M}) \exp\left\{\tilde{c} X(\mathcal{M})\right\}\right] \leq \frac{c_1}{n}.$$

From Lemma 7.9,

$$\begin{aligned}
&\frac{1}{n}Q_i[X(\mathcal{M})]P_i\left[\exp\left\{\tilde{c}X(\mathcal{M})\right\}\right] \\
&\quad\le \frac{1}{n}P_i\left[\exp\left\{\frac{\sigma i}{n}X(\mathcal{M}) + \left(1-e^{-\sigma i/n}\right)\mu P_i(\mathcal{M})\right\}\right] P_i\left[\exp\left\{\tilde{c}X(\mathcal{M})\right\}\right] \\
&\quad\le \frac{c_2}{n}.
\end{aligned}$$

Again from Lemma 7.9,

$$\begin{aligned}
&Q_i\big\{X(A) > 0\big\} \\
&\quad\le Q_i[X(A)] \\
&\quad\le P_i\left[X(A)\exp\left\{\frac{\sigma i}{n}X(\mathcal{M}) + \left(1-e^{-\sigma i/n}\right)\mu P_i(\mathcal{M})\right\}\right] \\
&\quad\le c_3\mu P_i(A),
\end{aligned}$$

and so

$$\sum_{A\in\mathcal{A}} Q_i\big\{X(A)>0\big\}^2 P_i\left[\exp\left\{\tilde{c}X(\mathcal{M})\right\}\right] \le c_4 \sum_{A\in\mathcal{A}} (\mu P_i(A))^2.$$

Similarly,

$$\begin{aligned}
&Q_i[X(A)\mathbf{1}_{\{X(A)\ge 2\}}] \\
&\quad\le P_i\left[X(A)\mathbf{1}_{\{X(A)\ge 2\}}\exp\left\{\frac{\sigma i}{n}X(\mathcal{M}) + \left(1-e^{-\sigma i/n}\right)\mu P_i(\mathcal{M})\right\}\right] \\
&\quad= P_i\left[X(A)(1-\mathbf{1}_{\{X(A)\le 1\}}\exp\left\{\frac{\sigma i}{n}X(\mathcal{M}) + \left(1-e^{-\sigma i/n}\right)\mu P_i(\mathcal{M})\right\}\right] \\
&\quad= \exp\left\{\left(1-e^{-\sigma i/n}\right)\mu P_i(\mathcal{M})\right\}\left(P_i\left[X(A)\exp\left\{\frac{\sigma i}{n}X(\mathcal{M})\right\}\right]\right. \\
&\qquad\left. - P_i\left[\exp\left\{\frac{\sigma i}{n}X(\mathcal{M}\setminus A)\right\}\right]P_i\{X(A)=1\}\right) \\
&\quad\le c_5\left(\mu P_i(A)\exp\left\{\frac{\sigma i}{n}\right\}\exp\left\{\mu P_i(\mathcal{M})\left(\frac{\sigma i}{n}-1\right)\right\}\right. \\
&\qquad\left. - \exp\left\{\mu P_i(\mathcal{M}\setminus A)\left(\frac{\sigma i}{n}-1\right)\right\}\mu P_i(A)\exp\{-\mu P_i(A)\}\right) \\
&\quad\le c_6(\mu P_i(A))^2,
\end{aligned}$$

and so

$$\sum_{A\in\mathcal{A}} Q_i[X(A)\mathbf{1}_{\{X(A)\ge 2\}}]P_i\left[\exp\left\{\tilde{c}X(\mathcal{M})\right\}\right] \le c_7 \sum_{A\in\mathcal{A}} (\mu P_i(A))^2.$$

Also,

$$\begin{aligned}
&P_i\left[\int_{\mathcal{M}} Q_i\{X(A(x))>0\}\,dX(x)\exp\{\tilde{c}X(\mathcal{M})\}\right]\\
&\quad= P_i\left[\sum_{A\in\mathcal{A}}\int_{\mathcal{M}} Q_i\{X(A)>0\}\mathbf{1}_{\{x\in A\}}\,dX(x)\exp\{\tilde{c}X(\mathcal{M})\}\right]\\
&\quad= P_i\left[\sum_{A\in\mathcal{A}} Q_i\{X(A)>0\}X(A)\exp\{\tilde{c}X(\mathcal{M})\}\right]\\
&\quad\le c_8\sum_{A\in\mathcal{A}}(\mu P_i(A))^2.
\end{aligned}$$

by arguments along the lines of those above.

By Corollary 7.12,

$$\begin{aligned}
&P_i\left[\int_{\mathcal{M}} \mathcal{P}^{(x)}_{Q_i}\bigl\{X(A(x))\ge 2\bigr\}\,dX(x)\ \exp\{\tilde{c}X(\mathcal{M})\}\right]\\
&\le \exp\left\{\frac{2\sigma k}{n}+\bigl(e^{\sigma k/n}+1-2e^{-\sigma k/n}\bigr)\mu P_i(\mathcal{M})\right\}\\
&\quad\times P_i\Biggl[\int_{\mathcal{M}}\left(\int \mathbf{1}_{\{(g+\delta_x)(A(x))\ge 2\}}\exp\left\{\frac{\sigma k}{n}g(\mathcal{M})\right\}\,dP_i(g)\right)\,dX(x)\\
&\qquad\times\exp\{\tilde{c}X(\mathcal{M})\}\Biggr]\\
&\le c_9\sum_{A\in\mathcal{A}}P_i\Biggl[\int_{\mathcal{M}}\left(\int \mathbf{1}_{\{g(A)\ge 1\}}\exp\left\{\frac{\sigma k}{n}g(\mathcal{M})\right\}\,dP_i(g)\right)\mathbf{1}_{\{x\in A\}}\,dX(x)\\
&\qquad\times\exp\{\tilde{c}X(\mathcal{M})\}\Biggr]\\
&= c_9\sum_{A\in\mathcal{A}}P_i\left[\mathbf{1}_{\{X(A)\ge 1\}}\exp\left\{\frac{\sigma k}{n}X(\mathcal{M})\right\}\right]P_i\left[X(A)\exp\{\tilde{c}X(\mathcal{M})\}\right]\\
&\le c_{10}\sum_{A\in\mathcal{A}}(\mu P_i(A))^2,
\end{aligned}$$

via arguments we have used before.

Observe that

$$g\left\{x\in\mathcal{M}:Q_i\{X(A(x))>0\}>\frac{1}{2}\right\}\le 2\int_{\mathcal{M}}Q_i\{X(A(x))>0\}\,dg(x),$$

and so

$$\begin{aligned}
&P_i\left[X\left\{x\in\mathcal{M}:Q_i\{X(A(x))>0\}>\frac{1}{2}\right\}\ \exp\{\tilde{c}X(\mathcal{M})\}\right]\\
&\quad\le c_{11}\sum_{A\in\mathcal{A}}(\mu P_i(A))^2,
\end{aligned}$$

from above.

Lastly,

$$\begin{aligned}&\sum_{A\in\mathcal{A}} P_i\left[\mathbf{1}_{\{X(A)\geq 2\}}\ \exp\{\tilde{c}X(\mathcal{M})\}\right]\\&\quad=\sum_{A\in\mathcal{A}} P_i\left[\mathbf{1}_{\{X(A)\geq 2\}}\ \exp\{\tilde{c}X(A)\}\right] P_i\left[\exp\{\tilde{c}X(\mathcal{M}\setminus A)\}\right]\\&\quad\leq c_{12}\sum_{A\in\mathcal{A}}(\mu P_i(A))^2.\end{aligned}$$

Substituting these estimates into (8.5) and recalling (8.3) gives

$$\begin{aligned}&\left\|Q_k-\mathfrak{P}Q_k\right\|_{\mathrm{Was}}\\&\quad\leq\frac{1}{n}\sum_{i=0}^{k-1}\left(\frac{\hat{c}}{n}+\bar{c}\int\sum_{A\in\mathcal{A}}(\mu P_i(A))^2\,d\mathcal{R}_{k-i}(\mathcal{A})\right),\end{aligned}$$

where $\hat{c}$ and $\bar{c}$ are constants that depend on σ, T, $\rho_0(\mathcal{M})$ and $\nu(\mathcal{M})$.

It follows from Corollary 7.3 that

$$a_k=\left\|Q_k-\mathfrak{P}Q_k\right\|_{\mathrm{Was}}\leq\frac{\beta'}{n}\sum_{i=0}^{k-1}\frac{1}{k-i+1}\leq\beta''\frac{\log n}{n},$$

where β' and β'' are constants that depend on σ, T, $\rho_0(\mathcal{M})$, $\nu(\mathcal{M})$, and the constants that appear in the shattering condition for the measures ρ_0 and ν with respect to the recombination operation $\mathbb{R}$.

It is clear that a_k tends to 0 as $n\to\infty$. Moreover, from (8.2),

$$b_{k+1}\leq e^{\sigma/n}\left(b_k+\frac{2\sigma}{n}\beta''\frac{\log n}{n}+\frac{2\sigma^2}{n^2}\right)\leq e^{\sigma/n}b_k+\gamma\frac{\log n}{n^2}$$

for a suitable constant γ that does not depend on n, so that b_k also tends to 0 as $n\to\infty$, completing the proof.

CHAPTER 9

Convergence to Poisson of iterated recombination

Our main result, Theorem 5.4, states that when n is large the discrete generation mutation-selection process, when started in a probability measure on the genotype space $\mathcal{G}$ that is the distribution of Poisson random measure on the locus space $\mathcal{M}$, stays close to the set of such probability measures out to numbers of generations that are of order n. In particular, this conclusion holds when the discrete generation starts with the entire population being completely wild type at every locus, so that the initial condition is the distribution of the Poisson random measure with zero intensity.

In this section we provide a complement to this observation by showing that the recombination process acting alone rapidly shuffles the distribution of a non-Poisson random measure on $\mathcal{M}$ to produce a probability measure on $\mathcal{G}$ that is close to the distribution of a Poisson random measure on $\mathcal{M}$.

Definition 9.1. A distribution P on $\mathcal{G}$ is *dispersive* if there is a constant β such that for any Borel $A \subset \mathcal{M}$,

$$\int g(A)\mathbf{1}_{\{g(A)\geq 2\}} dP(g) \leq \beta \mu P(A)^2.$$

Of course, the distribution of a Poisson random measure is always dispersive.

Proposition 9.2. *For all n and k, $\mu\mathfrak{R}^k P = \mu P$. If the recombination measure is shattering with respect to μP, and P is dispersive, then $\mathfrak{R}^k P$ converges to $\mathfrak{P}P$, as $k \to \infty$, with*

$$\|\mathfrak{R}^k P - \mathfrak{P}P\|_{\mathrm{Was}} \leq (3\beta + 2)\left(\nu(\mathcal{M})^2 \vee 2\alpha\nu(\mathcal{M})\right)(k+1)^{-1}. \tag{9.1}$$

Proof. We note first that

$$\mu\mathfrak{R}^{k+1}P(B) = \int\int_{\mathcal{G}}\int_{\mathcal{G}} (g_1\big|_R + g_2\big|_{R^c})(B)\, d\mu\mathfrak{R}^k P(g_1)\, d\mu\mathfrak{R}^k P(g_2)\, d\mathcal{R}(R) = \mu_n(B).$$

Let $\mathcal{A} = (A_1, \ldots, A_N)$ be a partition of $\mathcal{M}$. Then, $\mathfrak{R}_{\mathcal{A}}P$ defines a random genotype that has independent components on each A_i, with corresponding distributions $P\big|_{A_i}$. The distance between two probabilities on genotypes is the sum of the distances between the restrictions to a partition of $\mathcal{M}$. Thus,

$$\begin{aligned}\|\mathfrak{R}^k_{\mathcal{A}}P - \Pi_{\mu P}\|_{\mathrm{Was}} &= \sum_{i=1}^N \|\mathfrak{R}^k_{\mathcal{A}}P\big|_{A_i} - \Pi_{\mu P}\big|_{A_i}\|_{\mathrm{Was}} \\ &\leq \sum_{i=1}^N \|P\big|_{A_i} - \Pi_{\mu P}\big|_{A_i}\|_{\mathrm{Was}}.\end{aligned} \tag{9.2}$$

Pick any Borel subset A of $\mathcal{X}$. Define μ_A to be the measure on A defined for Borel subsets $B \subseteq A$ by

$$\mu_A(B) = \int_{\mathcal{G}} g(B)\mathbf{1}_{\{g(A)=1\}} dP(g).$$

Clearly $\mu_A \leq \mu P$. Also define the sub-probability measure P_A on genotypes restricted to A by

$$P_A[f] = \int_{\mathcal{G}} f(g\big|_A)\mathbf{1}_{\{g(A)\leq 1\}} dP(g).$$

Then, μ_A is the first-moment measure of P_A. Note also that

$$\mu P(A) = \mu_A(A) + \int_{\mathcal{G}} g(A)\mathbf{1}_{\{g(A)\geq 2\}}. \tag{9.3}$$

We have

$$\begin{aligned} &\big\|P\big|_A - \Pi_{\mu P}\big|_A\big\|_{\text{Was}} \\ &\quad \leq \big\|P\big|_A - P_A\big\|_{\text{Was}} + \big\|P_A - \Pi_{\mu_A}\big\|_{\text{Was}} + \big\|\Pi_{\mu_A} - \Pi_{\mu P}\big|_A\big\|_{\text{Was}}. \end{aligned} \tag{9.4}$$

We bound first

$$\big\|P\big|_A - P_A\big\|_{\text{Was}} \leq P\big\{g : g(A) \geq 2\big\} \leq \beta \mu P(A)^2. \tag{9.5}$$

Next, for any $f : \mathcal{G} \to \mathbb{R}$ with $\|f\|_{\text{Lip}} \leq 1$,

$$\begin{aligned} \big|P_A f - \Pi_{\mu_A} f\big| &\leq \big|P\big\{g : g(A) = 0\big\} - \Pi_{\mu_A}\big\{g : g(A) = 0\big\}\big| \\ &\qquad + \big|P_A f \mathbf{1}_{\{g:g(A)=1\}} - \Pi_{\mu_A} f \mathbf{1}_{\{g:g(A)=1\}}\big| + \Pi_{\mu_A}\big\{g : g(A) \geq 2\big\} \\ &\leq \Big|1 - \mu_A(A) + P\big\{g : g(A) \geq 2\big\} - e^{-\mu_A(A)}\Big| \\ &\qquad + \Big|\mu_A(f)\left(1 - e^{-\mu_A(A)}\right)\Big| + \frac{\mu_A(A)^2}{2} \\ &\leq P\big\{g : g(A) \geq 2\big\} + \frac{\mu P(A)^2}{2} + \mu P(A)^2 + \frac{\mu P(A)^2}{2}. \end{aligned}$$

Thus,

$$\big\|P_A - \Pi_{\mu_A}\big\|_{\text{Was}} \leq \big(\beta + 2\big)\mu P(A)^2. \tag{9.6}$$

Finally, by Lemma 2.7,

$$\big\|\Pi_{\mu_A} - \Pi_{\mu P}\big|_A\big\|_{\text{Was}} \leq \big\|\mu_A - \mu P\big|_A\big\|_{\text{Was}} + \big|\mu_A(A) - \mu P(A)\big| + \frac{1}{2}\big(\mu_A(A)^2 + \mu P(A)^2\big)$$

For any $f : A \to \mathbb{R}$ with $\|f\|_{\text{Lip}} \leq 1$, since μ_A is the first-moment measure of P_A, we have by Campbell's Theorem [**DVJ88**, (6.4.11)],

$$\begin{aligned} \big|\mu_A f - \mu P f\big| &= \left|\int_{\mathcal{G}} g[f] dP_A(g) - \int_{\mathcal{G}} g[f] dP(g)\right| \\ &= \int_{\mathcal{G}} g[f]\mathbf{1}_{\{g(A)\geq 2\}} dP(g). \end{aligned}$$

Thus,

$$\big\|\Pi_{\mu_A} - \Pi_{\mu P}\big|_A\big\|_{\text{Was}} \leq \beta \mu P(A)^2. \tag{9.7}$$

Putting (9.5), (9.6) and (9.7) into (9.4), we get

$$\big\|P\big|_A - \Pi_{\mu P}\big|_A\big\|_{\text{Was}} \leq (3\beta + 2)\mu P(A)^2 \tag{9.8}$$

By (9.2), then,

$$\left\|\mathfrak{R}^k P - \Pi_{\mu P}\right\|_{\mathrm{Was}} \le (3\beta + 2)\mathbb{E}|\pi|_2, \tag{9.9}$$

where the expectation is taken with respect to partitions π chosen from the distribution $\mathfrak{R}^k$. Applying Lemma 7.2 completes the proof of (9.1). □

Note that Proposition 9.2 is essentially a point-process version of Le Cam's Poisson convergence result of [**LC60**].

APPENDIX A

An expectation approximation

The following lemma, which gives error bounds for this approximation, was used several times in this work.

LEMMA A.1. *Let T be a nonnegative random variable with finite second moment. Then,*

$$-\mathbb{E}[T] + \frac{1}{2}\operatorname{Var}(T)e^{\mathbb{E}[T]} \geq \log \mathbb{E}\left[e^{-T}\right] \geq -\mathbb{E}[T].$$

and so

$$0 \leq \log \mathbb{E}\left[e^{-T}\right] \geq +\mathbb{E}[T] \leq \frac{1}{2}\operatorname{Var}(T)e^{\mathbb{E}[T]}.$$

In particular, if T is bounded by a constant τ, then

$$0 \leq \log \mathbb{E}\left[e^{-T}\right] \geq +\mathbb{E}[T] \leq \frac{1}{2}\tau^2 e^{\tau}.$$

PROOF. The second inequality holds directly from Jensen's inequality, by the convexity of $x \mapsto e^{-x}$.

The function $x \mapsto e^{-x} - \frac{x^2}{2}$ is concave for $x \geq 0$, so that by Jensen's inequality,

$$\mathbb{E}\left[e^{-T} - \frac{T^2}{2}\right] \leq e^{-\mathbb{E}[T]} - \frac{\mathbb{E}[T]^2}{2}.$$

Consequently,

$$\mathbb{E}\left[e^{-T}\right] \leq e^{-\mathbb{E}[T]} + \frac{1}{2}\operatorname{Var}(T) = e^{-\mathbb{E}[T]}\left(1 + \frac{1}{2}\operatorname{Var}[T]e^{\mathbb{E}[T]}\right).$$

Taking logarithms of both sides, and using the bound $\log(1+x) \leq x$ completes the proof. $\square$

Bibliography

[AGS05] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2005.

[Baa01] Ellen Baake, *Mutation and recombination with tight linkage*, J. Math. Biol. **42** (2001), no. 5, 455–488.

[Baa05] Michael Baake, *Recombination semigroups on measure spaces*, Monatsh. Math. **146** (2005), no. 4, 267–278.

[Baa07] ———, *Addendum to: "Recombination semigroups on measure spaces" [Monatsh. Math.* ***146*** *(2005), no. 4, 267–278; mr2191729]*, Monatsh. Math. **150** (2007), no. 1, 83–84.

[Bau05] Annette Baudisch, *Hamilton's indicators of the force of selection*, Proceedings of the National Academy of Sciences **102** (2005), no. 23, 8263–8.

[Bau08] ———, *Inevitable aging? : contributions to evolutionary-demographic theory*, Springer, Berlin, 2008.

[BB03] Michael Baake and Ellen Baake, *An exactly solved model for mutation, recombination and selection*, Canad. J. Math. **55** (2003), no. 1, 3–41.

[BT91] N. H. Barton and M. Turelli, *Natural and sexual selection on many loci*, Genetics **127** (1991), 229–55.

[Bür98] Reinhard Bürger, *Mathematical properties of mutation selection models*, Genetica **102/103** (1998), 279—298.

[Bür00] Reinhard Bürger, *The mathematical theory of selection, recombination, and mutation*, John Wiley & Sons, Chichester, New York, 2000.

[BW01] Ellen Baake and Holger Wagner, *Mutation—selection models solved exactly with methods of statistical mechanics*, Genet. Res., Camb. **78** (2001), 93—117.

[CE09] Aubrey Clayton and Steven N. Evans, *Mutation-selection balance with recombination: convergence to equilibrium for polynomial selection costs*, SIAM J. Appl. Math. **69** (2009), no. 6, 1772–1792.

[Cha01] Brian Charlesworth, *Patterns of age-specific means and genetic variances of mortality rates predicted by the mutation-accumulation theory of ageing*, J. Theor. Bio. **210** (2001), no. 1, 47–65.

[CKP95] Brian A. Coomes, Hüseyin Koçak, and Kenneth J. Palmer, *A shadowing theorem for ordinary differential equations*, Z. Angew. Math. Phys. **46** (1995), no. 1, 85–106.

[Daw99] Kevin J. Dawson, *The dynamics of infinitesimally rare alleles, applied to the evolution of mutation rates and the expression of deleterious mutations*, Theor. Popul. Biol. **55** (1999), 1–22.

[DS88] Nelson Dunford and Jacob T. Schwartz, *Linear operators. Part I*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1988.

[DU77] J. Diestel and J. J. Uhl, Jr., *Vector measures*, American Mathematical Society, Providence, R.I., 1977, With a foreword by B. J. Pettis, Mathematical Surveys, No. 15.

[DVJ88] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes*, Springer Series in Statistics, Springer-Verlag, New York, 1988.

[EK86] Stewart Ethier and Thomas Kurtz, *Markov processes: Characterization and convergence*, John Wiley & Sons, 1986.

[Hor03] Shiro Horiuchi, *Interspecies differences in the lifespan distribution: Humans versus invertebrates*, Life Span: Evolutionary, Ecological, and demographic perspectives (James R. Carey and Shripad Tuljapurkar, eds.), Population and Development Review, vol. 29 (Supplement), The Population Council, New York, 2003, pp. 127–151.

[Kal84] Olav Kallenberg, *An informal guide to the theory of conditioning in point processes*, International Statistical Review **52** (1984), no. 2, 151–64.

[Ken74] D. G. Kendall, *Foundations of a theory of random sets*, Stochastic geometry (a tribute to the memory of Rollo Davidson), Wiley, London, 1974, pp. 322–376.

[KJB02] Mark Kirkpatrick, Toby Johnson, and Nick Barton, *General models of multilocus evolution*, Genetics **161** (2002), 1727–50.

[KM66] Motoo Kimura and Takeo Maruyama, *The mutational load with epistatic gene interactions in fitness*, Genetics **54** (1966), 1337–1351.

[Kon82] A.S. Kondrashov, *Selection against harmful mutations in large sexual and asexual populations*, Genet. Res. **40** (1982), 325–332.

[LC60] Lucien Le Cam, *An approximation theorem for the Poisson binomial distribution*, Pacific Journal of Mathematics **10** (1960), 1181–1197.

[PC98] Scott D. Pletcher and James W. Curtsinger, *Mortality plateaus and the evolution of senescence: Why are old-age mortality rates so low?*, Evolution **52** (1998), no. 2, 454–64.

[Rac91] Svetlozar T. Rachev, *Probability metrics and the stability of stochastic models*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Ltd., Chichester, 1991. MR MR1105086 (93b:60012)

[RR98] Svetlozar T. Rachev and Ludger Rüschendorf, *Mass transportation problems. Vol. I*, Probability and its Applications (New York), Springer-Verlag, New York, 1998, Theory.

[SEW05] David Steinsaltz, Steven N. Evans, and Kenneth W. Wachter, *A generalized model of mutation-selection balance with applications to aging*, Adv. Appl. Math. **35** (2005), no. 1, 16–33.

[SH92] Stanley A. Sawyer and Daniel L. Hartl, *Population genetics of polymorphism and divergence*, Genetics **132** (1992), 1161–1176.

[SY02] George R. Sell and Yuncheng You, *Dynamics of evolutionary equations*, Applied Mathematical Sciences, vol. 143, Springer Verlag, 2002.

[Vil03] Cédric Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.

[Vil09] ———, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new.

[Wac03] Kenneth W. Wachter, *Hazard curves and life span prospects*, Life Span: Evolutionary, Ecological, and demographic perspectives (James R. Carey and Shripad Tuljapurkar, eds.), Population and Development Review, vol. 29 (Supplement), The Population Council, New York, 2003, pp. 270–291.

[WBG98] Holger Wagner, Ellen Baake, and Thomas Gerisch, *Ising quantum chain and sequence evolution*, J. Statist. Phys. **92** (1998), no. 5-6, 1017–1052.

[WES08] Kenneth W. Wachter, Steven N. Evans, and David R. Steinsaltz, *The age-specific forces of natural selection and walls of death*, U.C. Berkeley Department of Statistics Technical Report No. 757, `http://arxiv.org/abs/0807.0483`, 2008.

[WSE08] Kenneth W. Wachter, David R. Steinsaltz, and Steven N. Evans, *Vital rates from the action of mutation accumulation*, U.C. Berkeley Department of Statistics Technical Report No. 762, `http://arxiv.org/abs/0808.3622`, 2008.