

**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	4
1.2. Example I: Mutation counting	9
1.3. Example II: Polynomial selective costs	10
1.4. Example III: Demographic selective costs	11
1.5. Comments on the literature	12
1.6. Overview of the remainder of the work	14
Chapter 2. Definition, existence, and uniqueness of the dynamical system	15
2.1. Spaces of measures	15
2.2. Definition of the dynamical system	17
2.3. Existence and uniqueness of solutions	18
2.4. Lemmas used in the proof of existence and uniqueness	19
2.5. Density form of the dynamical system	21
2.6. Mutation measures with infinite total mass	25
Chapter 3. Equilibria	27
3.1. Introductory example: One-dimensional systems	28
3.2. Introductory example: Multiplicative selective costs	29
3.3. Fréchet derivatives	33
3.4. Existence of equilibria via perturbation	34
3.5. Concave selective costs	36
3.6. Concave selective costs: Existence and stability of equilibria	42
3.7. Iterative computation of the minimal equilibrium	45
3.8. Stable equilibria in the concave setting via perturbation	45
3.9. Equilibria for demographic selective costs	47
Chapter 4. Mutation-selection-recombination in discrete time	51
4.1. Mutation and selection in discrete-time	51
4.2. Recombination in discrete-time	53
4.3. Recombination trees and annealed recombination	55
4.4. Vintages	59
Chapter 5. Shattering and the formulation of the convergence result	63
5.1. Shattering of random measures	63
5.2. Consequences of shattering	66
5.3. Convergence to Poisson of iterated recombination	68
5.4. Atoms in the initial intensity	71
5.5. Preview of the main convergence result	72
Chapter 6. Convergence with complete Poissonization	75

Chapter 7. Supporting lemmas for the main convergence result	81
7.1. Estimates for Radon-Nikodym derivatives	81
7.2. Comparisons with complete Poissonization	91
Chapter 8. Convergence of the discrete-generation system	103
8.1. Outline of the proof	103
8.2. The convergence theorem	104
Appendix A. Results cited in the text	111
A.1. Gronwall's Inequality	111
A.2. Two expectation approximations	111
A.3. Identities for Poisson random measures	112
A.4. Bounds for Poisson random measures	114
A.5. Bounds for Radon-Nikodym derivatives	116
Bibliography	123
Index of notation	125
Index	127
Index	127

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 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.

We establish the existence and uniqueness of the dynamical system, provide conditions for the existence and stability of equilibrium states, and prove that our continuous-time dynamical system is the limit of a sequence of discrete-time 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.

Received by the editor June 27, 2014.

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.).

DS supported by grant K12-AG00981 from the National Institute on Aging (U.S.A.); a Discovery Grant from the National Science and Engineering Research Council (Canada); and a grant from the New Dynamics of Ageing program coordinated by the Economics and Social Research Council (U.K.).

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 governs, 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.

Mathematical models have particular salience for the study of senescence. Progressive physical deterioration and mortality rates rising across the lifespan are not, on the face of it, characteristics that natural selection would be expected to favor. Peter Medawar, in his inaugural lecture at University College London delivered in 1951 [Med52] called the origin of senescence “an unsolved problem in biology.” He went on to outline a possible solution. His idea, called “mutation accumulation” presents senescence as a side-effect of the interplay between natural selection and mutation. Shaping evolutionary thinking for the last half-century, Medawar’s idea is the centerpiece of the present work.

Medawar’s proposal was complemented in 1957 by ideas of George C. Williams [Wil57] directing attention to alleles with multiple, that is, pleiotropic, effects on fitness, some enhancing and some impairing the production and survival of progeny. This concept of “antagonistic pleiotropy” has taken its place beside the concept of “mutation accumulation”; together they are the twin guiding themes for evolutionary theories of aging. The relevance of both concepts to senescence turns on the truism that deleterious effects that only take hold later in life sacrifice less of an organism’s remaining potential for reproduction and nurturing than effects that act early. Medawar leads us to think of an equilibrium between mutation and selection in which a population is maintained at less than optimal fitness. Williams depicts alleles that could in constant environments go on to fixation.

It is important to emphasize that the mutations under consideration here are germ-line mutations passed on from generation to generation over stretches of evolutionary time. Mutations may also occur within cells of a single individual during the individual’s lifespan and be passed on and spread through cell division. These “somatic mutations” may also be important for aging. Although the models in [SEW05] might turn out to have some relevance to somatic mutation, the focus in this work is on germ-line mutations.

There is a rich literature on models for mutation, selection, and recombination, described further in Section 1.5. Our models build on that tradition but differ from much of it in being specially designed to accommodate general forms of detailed demographic structure. In leading applications, mutant alleles are associated with representations of age-specific effects on demographic schedules. One goal is to

provide structure from which predictions of the shapes of curves of probabilities of death as functions of age can be derived from assumptions about rates and kinds of mutations. This enterprise has come to the fore with the discovery of common features in the age-specific demographic schedules of populations from a variety of species which seem to call out for evolutionary explanations and which have stimulated the field of biodemography ([WF97] [CT03]).

An early precedent for the kinds of models we develop here is found in the well-known work of Motoo Kimura and Takeo Maruyama [KM66]. Our demographic formulations grow out of the work of Brian Charlesworth [Cha94], [Cha01].

Recapitulating, mutation accumulation models that incorporate the counterbalancing forces of recurring, slightly deleterious mutations and persistent selection have been staples of evolutionary theory [Bö0]. 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 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.

Kevin Dawson [Daw99] applied a variant of the Kimura-Maruyama rare-allele approximation [KM66, Bö0] (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 (1.5), 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 following.

- 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 may be freely specified.
- Recombination acts on a faster time scale than mutation or selection — in other words, the common quasi-linkage equilibrium (QLE) assumption holds.

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 dynamical system with the above features converges to a continuous-time, deterministic dynamical system that has as 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 random measure. The points of a realization of the Poisson random measure represent loci at which a randomly sampled individual’s genotype has accumulated ancestral mutations away from an original “reference” genotype which we call the *wild type*. Because a Poisson random measure 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 probability measures.

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 companion papers [WES08, WSE10].

1.1. Informal description of the limit model

In this section we describe the asymptotic model and explain by way of motivation 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. We present a formal development of the model in Sections 2.1 and 2.2 of Chapter 2. The exposition in [CE09] was based on an early version of the present text.

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 age $x \geq 0$ represents an additional increment to mortality at age x 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, \dots . 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 do not identify 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 $\mathcal{M} = \{1, 2, \dots, N\}$ for some positive integer N , then \mathcal{G} is essentially the Cartesian product \mathbb{N}_0^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, \dots, 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 we may also think of $P_t(G)$ as the probability that an individual chosen uniformly at random from the population has a 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 $\mathcal{M} = \{1, 2, \dots, N\}$ and we identify \mathcal{G} with the Cartesian product \mathbb{N}_0^N as above,

then $P_t(\{(n_1, n_2, \dots, n_N)\})$ represents the probability that an individual chosen uniformly at random from the population has n_j ancestral mutations at locus j for $j = 1, 2, \dots, N$.

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 given time die and reproduce at the same rate, because differing genotypes do not confer differing selective costs. 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 bounded Borel test function $\Phi : \mathcal{G} \rightarrow \mathbb{R}$. 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 when there is only mutation is that P_t , $t \geq 0$, evolves according to

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

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

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

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 [BÖ0]) 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 infinitesimal generator of a \mathcal{G} -valued Lévy process, and hence (1.1) has the following explicit probabilistic solution:

Let $\tilde{\Pi}$ denote a Poisson random measure on $\mathcal{M} \times \mathbb{R}_+$ with intensity measure $\nu \otimes \tilde{\lambda}$, where \otimes denotes the product of measures and $\tilde{\lambda}$ is Lebesgue measure. That is, $\tilde{\Pi}$ is a random integer-valued Borel measure such that:

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

$$(1.2) \quad Z_t := \int_{\mathcal{M} \times [0, t]} \delta_m d\tilde{\Pi}((m, u)).$$

Then,

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

where W is a random measure on \mathcal{M} that has probability measure P_0 and is independent of $\tilde{\Pi}$. In particular, if P_0 is itself the distribution of a Poisson random measure, then P_t is also 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 $\rho_t(A) := \int_{\mathcal{G}} g(A) dP_t(g)$ for $A \subseteq \mathcal{M}$), then ρ_t evolves according to the simple dynamics

$$\rho_t(A) = \rho_0(A) + t\nu(A).$$

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* $S : \mathcal{G} \rightarrow \mathbb{R}_+$. 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

$$(1.3) \quad S(g + h) \geq S(h), \quad g, h \in \mathcal{G},$$

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 $P_t[S] - S(g')$. More formally,

$$(1.4) \quad \begin{aligned} \frac{d}{dt} P_t[\Phi] &= -P_t[\Phi \cdot (S - P_t S)] \\ &= - \int_{\mathcal{G}} \Phi(g') \left(S(g') - \int_{\mathcal{G}} S(g'') dP_t(g'') \right) dP_t(g'). \end{aligned}$$

For example, when $\mathcal{M} = \{1, 2, \dots, N\}$ 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, if S has the additive property

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

then the selective effects of different mutations do not interact. In particular, if P_0 is the distribution of a Poisson random measure, then P_t is also 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, in the more general case in which S is epistatic (that is, non-additive), then P_t is, in general, not 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 (1.4):

$$(1.5) \quad \frac{d}{dt}P_t[\Phi] = P_t \left[\int_{\mathcal{M}} (\Phi(\cdot + \delta_m) - \Phi(\cdot)) d\nu(m) \right] - P_t[\Phi \cdot (S - P_t S)].$$

This is the model introduced and analyzed at length in [SEW05] using the Feynman-Kac formula. When $\mathcal{M} = \{1, 2, \dots, N\}$, (1.5) 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 [Bö0] for the derivation of an analytic solution that agrees with the one that arises from a Feynman-Kac analysis.

The principal result from [SEW05] is the following:

PROPOSITION 1.1. *Let Z_t be the random measure distributed according to the Poisson probability measure $\bar{\Pi}$ in (1.2). Suppose that there is a positive T such that*

$$(1.6) \quad \mathbb{E} \left[\exp \left(- \int_0^t S(Z_u) du \right) S(Z_t) \right] < \infty$$

for all $t \in [0, T)$. Then, (1.5) has a solution on $[0, T)$, given by

$$(1.7) \quad P_t[\Phi] = \frac{\mathbb{E} \left[\exp \left(- \int_0^t S(Z_u) du \right) \Phi(Z_t) \right]}{\mathbb{E} \left[\exp \left(- \int_0^t S(Z_u) du \right) \right]}.$$

Notice that the effect of selection may be understood as reweighting a Poisson measure, rather than explicitly removing mutations from the population. This will be an important theme in the convergence proof that starts in Chapter 4.

The model without recombination in [SEW05], like the model presented here, is an infinite-population model. Alleles are not subject to genetic drift. The fittest genotypes present in the initial population are not lost, no matter how rare they are. A finite-population model without recombination defined along the same lines would be vulnerable to Mueller's ratchet, the process in which the fittest classes of genotypes in an asexual population can be successively lost through drift, carrying the population to extinction. A thorough discussion of the ratchet and the associated advantages of sex and recombination with references is found in [Bö0], pages 303–308. In the face of Mueller's ratchet, a finite-population version of the model in [SEW05] would not be viable. This observation underscores the importance of incorporating recombination, which avoids the ratchet by allowing fittest classes to be reconstituted in every generation.

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 (but see Remark 1.2 and Section 5.4 below).
- 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 conditions, the probability measure P_t converges as $t \rightarrow \infty$ 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 recombination rate to be effectively infinite, then the probability measure P_t is 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 P_t is also the distribution of a Poisson random measure for all $t > 0$. As before, ρ_t denotes the intensity measure of P_t (so that ρ_t is a finite measure on the space \mathcal{M} of loci). In anticipation of Notation 2.1, we write X^π for a Poisson random measure on \mathcal{M} with intensity measure π . Combining our previous observations, we expect that ρ_t should satisfy the evolution equation

$$(1.8) \quad \rho_t(dm) = \rho_0(dm) + t\nu(dm) - \int_0^t \mathbb{E}[S(X^{\rho_s} + \delta_m) - S(X^{\rho_s})] \rho_s(dm) ds.$$

We define the rigorous counterpart of (1.8) in Chapter 2 and establish the existence and uniqueness of solutions. Furthermore, we show in Chapters 4 ff. 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 random measure, (1.8) may look rather forbidding. However, for certain reasonable choices of selective costs we can evaluate the integral explicitly, leading to a simpler and more intuitive system. We consider three such cases in the following three sections. First, though, we make a useful observation about how we may clump loci together in (1.8).

REMARK 1.2. Consider two instances $(\rho'_t)_{t \geq 0}$ and $(\rho''_t)_{t \geq 0}$ of the dynamical system (1.8) 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 following hold.

- The initial measure ρ_0'' on \mathcal{M}'' is the push-forward of the initial measure ρ_0' 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 $g'' \in \mathcal{G}''$ 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 write P_t' and P_t'' for 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'' . That is, 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 $s : \mathbb{N}_0 \rightarrow \mathbb{R}_+$ with $s(0) = 0$.

We may apply Remark 1.2 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 1.2 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 assumes the starting point to be the null genotype. In this case, we readily compute that the intensity at time t is $r_t = (q/\bar{s})(1 - \exp\{-\bar{s}t\})$. The intensity converges monotonically to the equilibrium value q/\bar{s} , the elementary expression for mutation-selection equilibrium going back to J.B.S. 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 different, simpler answers than Kimura and Maruyama's treatment of mutation counting without recombination (see [KM66]).

For another example, consider the case when costs are multiplicative, in the sense that $s(k) := 1 - \exp\{-k\theta\}$ for some constant $\theta > 0$. We have $\psi(x) = ax \exp\{-ax\}$, where $a := 1 - \exp\{-\theta\}$. An equilibrium exists only if the mutation rate q is below the maximum of $r \exp\{-r\}$, namely $1/e$, and in that case r_t converges monotonically to the smallest positive solution of the equation $q - \psi(x) = 0$. The solution is $x = (1/a) \sum_{n=1}^{\infty} (-n)^{n-1} (-q)^n / n!$ where the summed expression is Lambert's W function evaluated at the negative argument $-q$. 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 θ is small the equilibrium is approximately q/θ , as one would expect from the observation that in this case $s(k)$ is approximately θk for small k and hence this model is approximately the additive one of the previous paragraph with $\bar{s} = \theta$.

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 (1.8) 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 (1.8) 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 (1.8) with locus space a single point is mathematically well-defined, it cannot 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. Single-point mutation spaces require the special treatment described in Remark 1.2.

1.3. Example II: Polynomial selective costs

Recall that when $\mathcal{M} = \{1, 2, \dots, N\}$ 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, (1.8) becomes the system of ordinary differential equations

$$(1.9) \quad \frac{d}{dt} \rho_t(\{j\}) = \nu(\{j\}) - \rho_t(\{j\}) \sum_{\mathbf{n}} [S(\mathbf{n} + \mathbf{e}_j) - S(\mathbf{n})] \prod_{k=1}^N e^{-\rho_t(\{k\})} \frac{\rho_t(\{k\})^{n_k}}{n_k!}.$$

As in the previous example, we should not think of $\{1, 2, \dots, N\}$ as being the "real" locus space. Rather, we should imagine that there is something like a continuum of loci which are partitioned into N sub-classes in such a way that the loci in each sub-class have indistinguishable selective effects, and (1.9) is a reduced description that comes from applying the observation in Remark 1.2.

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 $I \subseteq \{1, \dots, N\}$ 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 $I \subseteq \{1, \dots, N\}$ 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 (1.9) becomes

$$(1.10) \quad \dot{\rho}_k = \nu_k - \sum_{I \in \mathcal{I}_k} \alpha_I \rho^I, \quad 1 \leq k \leq N,$$

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

It is shown in [CE09] that if $\nu_k > 0$ for all k (that is, if mutations may occur at all loci), if $\alpha_{\{i\}} > 0$ for all i (that is, if the individual effect of any mutation is deleterious, in keeping with our general assumption on the selective cost), and if $\alpha_I \geq 0$ for all subsets I (that is, the synergistic effects of individually deleterious mutations are never beneficial), then the system of equations (1.10) 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, WSE10], 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* is $-\log \ell_x(g)$ and the *hazard function* is its derivative with respect to x when it exists. . 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 $\int_0^\infty f_x \ell_x(g) dx$, the “Net Reproduction Ratio” or “NRR”.

Suppose further that there is a background hazard λ and that an ancestral mutation at locus $m \in \mathcal{M}$ contributes a bounded increment $\theta(m, x)$ to the cumulative hazard function at age x . This function θ with two arguments is a generalization of the constant cost θ in Section 1.2 and of a single-argument cost $\theta(m)$ in play in examples with multiplicative costs in Section 3.2. The probability that an individual with genotype $g \in \mathcal{G}$ lives beyond age $x \in \mathbb{R}_+$ is

$$(1.11) \quad \ell_x(g) = \exp \left(-\lambda x - \int_{\mathcal{M}} \theta(m, x) dg(m) \right).$$

The corresponding selective cost is

$$(1.12) \quad S(g) = \int_0^\infty f_x \ell_x(0) dx - \int_0^\infty f_x \ell_x(g) dx$$

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

Marginal selective costs are given by the expression

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

Expected marginal costs have the expected value of $\ell_x(g)$ under the integral on the right-hand side. When the genotype g is a realization of a Poisson random measure X^π whose intensity is the finite measure π on \mathcal{M} , this expected survivorship function can be found from the Poisson identity

$$(1.13) \quad \mathbb{E} \left[\exp \left\{ - \int_{\mathcal{M}} \phi(m) dX^\pi(m) \right\} \right] = \exp \left\{ - \int_{\mathcal{M}} (1 - e^{-\phi(m)}) d\pi(m) \right\}.$$

The identity applies to bounded Borel functions $\phi : \mathcal{M} \rightarrow \mathbb{R}$. It is proved, for instance, in [Kin93, 3.15]. It also follows from Campbell's Theorem (see Proposition A.8).

The identity gives the expression, in terms of the expectation operator \mathbb{E} ,

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

An important issue for demographic applications is whether the solution ρ_t , $t \geq 0$, of (1.8) converges to an equilibrium ρ_* as $t \rightarrow \infty$ and, if so, what are the features of that equilibrium. In particular, what we can say about $\mathbb{E}[\ell_x(X^{\rho_*})]$, 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

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

We study such equilibria further in Section 3.9 and in [WES08, WSE10]. In particular, in [WES08], we consider mutations which each provide a point mass increment to the hazard at a specific age. When $\mathcal{M} = \mathbb{R}_+$ and the mutation rate measure ν is absolutely continuous with respect to Lebesgue measure, and θ is of the form $\theta(m, x) = \eta(m) \mathbf{1}_{\{x \geq m\}}$, (where, as before, $\mathbf{1}_A$ is the indicator function of the set A) 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 for r_* .

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 reference is Reinhard Bürger's book [B00] along with 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 envisioned 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 regarded 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 treated either as multilocus systems with complete linkage or as 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. In our setting, the Poisson probability measures constitute, very roughly speaking, a hypersurface within the space of all probability measures and recombination can be viewed as an operation that pushes an arbitrary probability measure “towards” that hypersurface. General results about the convergence of discrete-time dynamical systems to a continuous-time one that is forced onto a submanifold by a suitable vector field may be relevant here, but we are not aware of a particular framework of this kind into which we can fit our results.

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 convergence theorem in Chapter 8 therefore delineates 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 absent.

1.6. Overview of the remainder of the work

We define the measure-valued dynamical system (1.8) 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 preview 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 state the convergence theorem and complete its proof in Chapter 8.

The appendix contains relevant versions of results from the literature as well as technical results about Poisson random measures and Radon-Nikodym derivatives that are used throughout the main text.

Definition, existence, and uniqueness of the dynamical system

2.1. Spaces of measures

Our model, as presented in the Introduction, has four fundamental ingredients:

- a 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} ;
- a selective cost function $S : \mathcal{G} \rightarrow \mathbb{R}_+$ with $S(0) = 0$ and $S(g + h) \geq S(h)$ for $g, h \in \mathcal{G}$.

An element of \mathcal{G} represents a “genotype” regarded as the set of loci at which there have been ancestral mutations away from the reference wild type. 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 measure on \mathcal{M} . The distribution of such a Poisson random measure is determined by its intensity measure, which is in general a locally-finite Borel measure on \mathcal{M} and which is a finite measure when the mutation measure ν has finite total mass. For most results, we do assume that ν has finite total mass, but a brief discussion of infinite-mass cases is given in Section 2.6.

NOTATION 2.1.

- Denote by \mathcal{H} the Banach space of finite signed Borel measures on \mathcal{M} equipped with the norm $\|\cdot\|_{\text{Was}}$ defined below.
- 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, $\pi = \pi^+ - \pi^-$.
- For any measure P on \mathcal{G} , let μP be the intensity measure associated with P ; that is, for a nonnegative Borel function f on \mathcal{M} ,

$$\int_{\mathcal{M}} f(m) d(\mu P)(m) = \int_{\mathcal{G}} \int_{\mathcal{M}} f(m) dg(m) dP(g).$$

- For $\pi \in \mathcal{H}^+$, write X^π for a Poisson random measure on \mathcal{M} with intensity measure π and Π_π for the distribution of this Poisson random measure.

The norms on our spaces of measures are based on a metric from the class of metrics named for Leonid Wasserstein (also transliterated as Vasershtein, or otherwise). In our investigations we have to deal with spaces of measures at a hierarchy of different levels, including \mathcal{G} , \mathcal{H} , and finite signed measures on \mathcal{G} . The Wasserstein metrics provide a unified way of topologizing all these various spaces.

NOTATION 2.2.

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

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

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

$$(2.2) \quad \|\pi\|_{\text{Was}} := \sup \left\{ \left| \int f d\pi \right| : \|f\|_{\text{Lip}} \leq 1 \right\}.$$

The Wasserstein metric metrizes the topology of weak convergence of measures on the finite signed Borel measures on (E, d) .

- For any measure π in such a space of signed measures, write

$$(2.3) \quad \pi[f] := \int f d\pi.$$

- Write

$$\sigma := \sup_{g, h \in \mathcal{G}, g \neq h} \frac{|S(g) - S(h)|}{\|g - h\|_{\text{Was}}}$$

for the (possibly infinite) Lipschitz constant of the selective cost function S .

We take advantage of the versatility of these definitions to build Wasserstein metrics on top of Wasserstein metrics. We start by taking \mathcal{M} for the metric space E . 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 E , 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 E , we obtain a Wasserstein metric on \mathcal{H} .

An extensive account of Wasserstein metrics may be found in [Rac91, RR98]. The properties used here are described in Problem 3.11.2 of [EK86]. In particular, it is shown that our version is indeed a metric and that \mathcal{H} as well as \mathcal{G} and the finite signed measures on \mathcal{G} are complete in this norm. The Wasserstein distance between two probability measures is bounded above by their total variation distance.

We note that the designation ‘‘Wasserstein metric’’ is also 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].

In some places we need to use the total variation norm, defined for a finite signed Borel measure π as

$$(2.4) \quad \|\pi\|_{\text{TV}} := \sup \left\{ \left| \int f d\pi \right| : \sup_x |f(x)| \leq 1 \right\}.$$

There is a corresponding Lipschitz constant for a function $F : \mathcal{G} \rightarrow \mathbb{R}$, defined as

$$(2.5) \quad \text{Lip}_{\text{TV}} F := \sup_{g \neq g'} \frac{|F(g) - F(g')|}{\|g - g'\|_{\text{TV}}}.$$

This yields the identity

$$(2.6) \quad \text{Lip}_{\text{TV}} F = \sup_{g \in \mathcal{G}, x \in \mathcal{M}} |F(g + \delta_x) - F(g)|.$$

Lastly, we sometimes compare two equivalent probability measures P and Q on \mathcal{G} by considering the quantity

$$(2.7) \quad \Theta(P, Q) := \text{Lip}_{\text{TV}} \log \frac{dQ}{dP} = \sup_{g \neq g'} \frac{\left| \log \frac{dQ}{dP}(g) - \log \frac{dQ}{dP}(g') \right|}{\|g - g'\|_{\text{TV}}}.$$

Note that $\Theta(P, P) = 0$. Conversely, if $\Theta(P, Q) = 0$, then $\log dQ/dP$ is a constant and so dQ/dP is also a constant; however, this constant must be 1 since P and Q are probability measures, and hence $P = Q$. Note also that $\Theta(P, Q) = \Theta(Q, P)$ because $\log dP/dQ = \log(dQ/dP)^{-1} = -\log dQ/dP$. Finally, if R is another probability measure equivalent to P and Q , then

$$\log \frac{dR}{dP} = \log \frac{dR}{dQ} \cdot \frac{dQ}{dP} = \log \frac{dR}{dQ} + \log \frac{dQ}{dP}$$

and so $\Theta(P, R) \leq \Theta(P, Q) + \Theta(Q, R)$. Therefore, Θ restricted to a suitable class of probability measures is a metric.

2.2. Definition of the dynamical system

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 motivated by (1.8), restated here.

$$(2.8) \quad \rho_t(dm) = \rho_0(dm) + t \nu(dm) - \int_0^t \mathbb{E}[S(X^{\rho_s} + \delta_m) - S(X^{\rho_s})] \rho_s(dm) ds,$$

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

DEFINITION 2.3.

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

$$(2.9) \quad F_\pi(x) := \mathbb{E}[S(X^\pi + \delta_x) - S(X^\pi)]$$

for $x \in \mathcal{M}$ and $\pi \in \mathcal{H}^+$.

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

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

That is, for any bounded Borel function $f : \mathcal{M} \rightarrow \mathbb{R}$ and π in \mathcal{H}^+ ,

$$\int_{\mathcal{M}} f(x) d(D\pi)(x) = \int_{\mathcal{M}} f(x) F_\pi(x) d\pi(x).$$

With this notation, our evolution equation (2.8) becomes

$$(2.10) \quad \rho_t = \rho_0 + t\nu - \int_0^t D\rho_s ds$$

A solution is an \mathcal{H}^+ -valued function ρ that is continuous with respect to the Wasserstein metric and therefore with respect to the topology of weak convergence of measures and satisfies (2.10) for all $t \geq 0$. Equation (2.10) involves the integration of a measure-valued function, and such an integral can have a number of different meanings. We require only that if $\eta : \mathbb{R}_+ \rightarrow \mathcal{H}$ is a Borel function, then for $t \geq 0$ the integral $\int_0^t \eta_s ds$ is the element of \mathcal{H} satisfying, for every Borel $A \subseteq \mathcal{M}$,

$$(2.11) \quad \left(\int_0^t \eta_s ds \right) (A) = \int_0^t \eta_s(A) ds$$

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

We now prove the existence and uniqueness of solutions to (2.10) with the assistance of three lemmas that are proved below in Section 2.4. The proof is an application of the standard iterative method of Charles Émile Picard.

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

- $S(0) = 0$,
- $S(g) \leq S(g + h)$ for all $g, h \in \mathcal{G}$,
- the Lipschitz constant σ of the selective cost S (in the Wasserstein metric) is finite.

Then, (2.10) 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. Temporarily, write $C([0, T], \mathcal{H})$ for the Banach space of continuous \mathcal{H} -valued functions on $[0, T]$, equipped with the norm

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

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 $\Delta : C([0, T], \mathcal{H}) \rightarrow C([0, T], \mathcal{H})$ by

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

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

$$\|\Delta\alpha - \Delta\beta\|_c \leq \sup_{0 \leq t \leq T} e^{-ct} \int_0^t \|D\alpha_s^+ - D\beta_s^+\|_{\text{Was}} ds.$$

By Lemma 2.7 below, the norm inside the integral on the right has the following bound

$$\|D\alpha_s^+ - D\beta_s^+\|_{\text{Was}} \leq \sigma(2 + 8\{\alpha_s^+(\mathcal{M}) \wedge \beta_s^+(\mathcal{M})\})\|\alpha_s - \beta_s\|_{\text{Was}}.$$

We conclude that

$$\begin{aligned} \|\Delta\alpha - \Delta\beta\|_c &\leq \sup_{0 \leq t \leq T} e^{-ct} \int_0^t \sigma(2 + 8\{\rho_0(\mathcal{M}) + s\nu(\mathcal{M})\})e^{cs}\|\alpha - \beta\|_c ds \\ &\leq \sup_{0 \leq t \leq T} e^{-ct} \left[\sigma(2 + 8\rho_0(\mathcal{M}))\frac{e^{ct} - 1}{c} + 8\sigma\nu(\mathcal{M})\frac{(ct - 1)e^{ct} + 1}{c^2} \right] \|\alpha - \beta\|_c. \end{aligned}$$

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

It follows from the Contraction Mapping Theorem that the equation

$$(2.12) \quad \rho_t = \Delta\rho_t = \rho_0 + t\nu - \int_0^t D\rho_s^+ ds$$

has a unique solution in Γ . Furthermore, any function in \mathcal{H} that is a solution to (2.12) must automatically be in Γ . Therefore, the solution is unique.

It remains to show that it actually takes values in the subset \mathcal{H}^+ of nonnegative measures. For any Borel set $A \subseteq \mathcal{M}$,

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

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

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

The Lipschitz condition on S in terms of the Wasserstein metric implies, via Lemma 2.7 below, that $F_\pi(m) \leq \sigma$ for all $m \in \mathcal{M}$ and $\pi \in \mathcal{H}^+$. Then,

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

Hence, $\rho_t(A) \geq 0$ for $0 \leq t \leq 1/\sigma$. Because this holds for all Borel sets $A \subseteq \mathcal{M}$, we have $\rho_t \in \mathcal{H}^+$ for $0 \leq t \leq 1/\sigma$. Iterating this argument, with the time 0 replaced successively by the times $1/\sigma, 2/\sigma, \dots$ gives the result. \square

2.4. Lemmas used in the proof of existence and uniqueness

We assume in this section that the hypotheses of Theorem 2.4 hold and prove the lemmas on which that theorem relies.

LEMMA 2.5. *The function F_π is Lipschitz in the Wasserstein metric for each finite measure $\pi \in \mathcal{H}^+$, and, in terms of the Lipschitz constant σ of the selective cost function S ,*

$$\sup_{\pi \in \mathcal{H}^+} \|F_\pi(\cdot)\|_{\text{Lip}} \leq 2\sigma$$

and

$$\sup_{\pi \in \mathcal{H}^+, m \in \mathcal{M}} F_\pi(m) \leq \sigma.$$

Furthermore, for all $m \in \mathcal{M}$ and all $\pi \in \mathcal{H}^+$,

$$(2.13) \quad \frac{1}{F_\pi(m)} \leq \frac{\exp\{\pi(\mathcal{M})\}}{S(\delta_m)}.$$

PROOF. By definition,

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

Furthermore,

$$\begin{aligned} F_\pi(m) &= \mathbb{E}[(S(X^\pi + \delta_x) - S(X^\pi)) (\mathbf{1}_{\{X^\pi=0\}} + \mathbf{1}_{\{X^\pi>0\}})] \\ &\geq S(\delta_m) \mathbb{E}[\mathbf{1}_{\{X^\pi=0\}}] = S(\delta_m) \exp\{-\pi(\mathcal{M})\}, \end{aligned}$$

implying the bound on $1/F_\pi$. □

LEMMA 2.6. For two finite measures $\pi', \pi'' \in \mathcal{H}^+$,

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

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

$$|F_{\pi'}(x) - F_{\pi''}(x)| = 2\sigma |\Pi_{\pi'}[\Phi] - \Pi_{\pi''}[\Phi]| \leq 2\sigma \|\Pi_{\pi'} - \Pi_{\pi''}\|_{\text{Was}}.$$

The lemma now follows from Lemma A.12. □

LEMMA 2.7. For two finite signed measures $\alpha, \beta \in \mathcal{H}$,

$$\|D\alpha^+ - D\beta^+\|_{\text{Was}} \leq \sigma(2 + 8\{\alpha^+(\mathcal{M}) \wedge \beta^+(\mathcal{M})\}) \|\alpha - \beta\|_{\text{Was}}.$$

PROOF. Suppose without loss of generality that $\alpha^+(\mathcal{M}) \leq \beta^+(\mathcal{M})$. By Lemmas 2.5 and 2.6, for any Lipschitz test function f ,

$$\begin{aligned} &\left| \int f(x) dD\alpha^+(x) - \int f(x) dD\beta^+(x) \right| \\ &\leq \left| \int F_{\alpha^+}(x) f(x) d\alpha^+(x) - \int F_{\beta^+}(x) f(x) d\alpha^+(x) \right| \\ &\quad + \left| \int F_{\beta^+}(x) f(x) d\alpha^+(x) - \int F_{\beta^+}(x) f(x) d\beta^+(x) \right| \\ &\leq 8\sigma \|\alpha^+ - \beta^+\|_{\text{Was}} \|f\|_{\infty} \alpha^+(\mathcal{M}) + 2\sigma \|f\|_{\text{Lip}} \|\alpha^+ - \beta^+\|_{\text{Was}} \\ &\leq (2\sigma + 8\sigma \alpha^+(\mathcal{M})) \|f\|_{\text{Lip}} \|\alpha - \beta\|_{\text{Was}}. \end{aligned}$$

We have used the fact that $\|f'f''\|_{\text{Lip}} \leq \|f'\|_{\text{Lip}} \|f''\|_{\text{Lip}}$ for $f', f'' \in \text{Lip}$. □

2.5. Density form of the dynamical system

The solutions provided by Theorem 2.4 are absolutely continuous with respect to $\rho_0 + \nu$ and so can be written out in terms of Radon-Nikodym derivatives (that is, densities) with respect to that reference measure. We show below that, more generally, the Radon-Nikodym derivatives with respect to suitable reference measures ζ belong to the space $L_+^\infty(\mathcal{M}, \zeta)$ of nonnegative functions that are essentially-bounded for ζ . This parallel approach of viewing our dynamical system as taking values in a space of functions rather than a space of measures was first developed in [CE09]. It provides another route to defining the informal dynamical system rigorously and establishing the existence and uniqueness of solutions.

Of course our analysis as a whole cannot be carried through in $L^\infty(\mathcal{M}, \zeta)$. The singleton mutations δ_m in \mathcal{G} with which selective costs are measured can not be viewed as elements of $L^\infty(\mathcal{M}, \zeta)$ for some measure ζ equivalent to $\rho_0 + \nu$ unless ρ_0 or ν has atoms, and the simultaneous consideration of a continuum of non-equivalent initial measures ρ_0 can not be accommodated within a single $L^\infty(\mathcal{M}, \zeta)$. For the full theory, spaces of measures equipped with Wasserstein metrics or tools of similar generality are required. However, proofs of existence and uniqueness for fixed initial conditions and mutation measures can be carried out in the $L^\infty(\mathcal{M}, \zeta)$ setting. Moreover, this approach facilitates an extension of the theory to some cases where the mutation measure ν has infinite total mass and it also justifies a step in the proof of Lemma 3.14 in our chapter on equilibria.

NOTATION 2.8.

- The measure ζ on \mathcal{M} is any measure with respect to which the initial state ρ_0 and the mutation measure ν are both absolutely continuous, with bounded Radon-Nikodym derivatives $r_0(m)$ and $q_\nu(m)$. (In practice, it is convenient to take ζ to be the sum of ν and the part of ρ_0 , if any, orthogonal to ν).
- Denote by $L^\infty(\mathcal{M}, \zeta)$ the Banach space of equivalence classes of ζ -essentially bounded functions on \mathcal{M} and by $L_+^\infty(\mathcal{M}, \zeta)$ its subset consisting of equivalence classes of nonnegative functions.
- Write \mathcal{K} for the space of finite signed measures on \mathcal{M} that are absolutely continuous with respect to ζ with ζ -essentially bounded Radon-Nikodym derivatives, equipped with the $L^\infty(\mathcal{M}, \zeta)$ norm via the usual bijection between $L^\infty(\mathcal{M}, \zeta)$ and \mathcal{K} .
- Write \mathcal{K}_+ for the subset of \mathcal{K} consisting of nonnegative measures.

Instead of working forward from conditions on the selective cost S via the lemmas of Section 2.4, we impose a bound and Lipschitz condition directly on the whole contribution from $D\rho$. Anticipating the remarks in Section 2.6, we offer proofs that do not rely on the finiteness of $\nu(\mathcal{M})$.

THEOREM 2.9. *Fix a mutation measure ν and starting state ρ_0 in \mathcal{H}^+ and a selective cost $S : \mathcal{G} \rightarrow \mathbb{R}_+$ such that $S(0) = 0$, and $S(g) \leq S(g + h)$ for all $g, h \in \mathcal{G}$. Let ν and ρ_0 be absolutely continuous with respect to a measure ζ in \mathcal{H}^+ with bounded Radon-Nikodym derivatives $q_\nu(m)$ and $r_0(m)$.*

Suppose the function $(\pi, m) \mapsto F_\pi(m)$ defined from S in Definition 2.3 the following conditions.

- *There is a constant σ such that for all $\pi \in \mathcal{H}^+$ and $m \in \mathcal{M}$, $F_\pi(m) \leq \sigma$.*

- For all pairs of measures $\alpha = a(m)\zeta(dm)$ and $\beta = b(m)\zeta(dm)$ in \mathcal{K}_+ ,

$$\|F_\alpha(m)a(m) - F_\beta(m)b(m)\|_\infty \leq \sigma(\|a\|_\infty \wedge \|b\|_\infty) \|\alpha - \beta\|_\infty.$$

Then, the dynamic equation (2.10) has a unique solution in $C(\mathbb{R}_+, \mathcal{K}_+)$.

PROOF. The proof mimics the proof of Theorem 2.4. The space \mathcal{H} is replaced by \mathcal{K} , \mathcal{H}^+ is replaced by \mathcal{K}_+ , and the norm $\|\cdot\|_{\text{Was}}$ is replaced by $\|\cdot\|_\infty$. The analogue of the subset Γ consists of \mathcal{K} -valued functions α with $\alpha_0 = \rho_0$ and

$$\|\alpha_t^+\|_\infty \leq \|r_0(m) + tq_\nu(m)\|_\infty.$$

The assumptions imposed on F take the place of the bounds from Lemmas 2.7 and 2.5. \square

Theorem 2.9 gives us for each $t \geq 0$ a measure which, as an element of \mathcal{K}_+ , has a Radon-Nikodym derivative with respect to ζ . *A priori*, the Radon-Nikodym derivative is defined separately for each $t \geq 0$ as a function of $m \in \mathcal{M}$ and is only unique up to ζ -null sets. The behavior of a particular choice for the ensemble of Radon-Nikodym derivatives is, for fixed $m \in \mathcal{M}$, not guaranteed to be even a measurable function of $t \geq 0$. We now show that we can define versions of the Radon-Nikodym derivatives that are in fact continuously differentiable functions of $t \geq 0$ for every $m \in \mathcal{M}$.

THEOREM 2.10. *Suppose that ρ_t is a solution to the dynamic equation (2.10) in $C(\mathbb{R}_+, \mathcal{K}_+)$ for selective costs satisfying the conditions of Theorem 2.9. Then, there exists a Borel function $(t, m) \mapsto r_t(m)$ on $\mathbb{R}_+ \times \mathcal{M}$ such that for all $t \geq 0$ the function $m \mapsto r_t(m)$ is a Radon-Nikodym derivative of ρ_t with respect to ζ and such that for all $m \in \mathcal{M}$ the function $t \mapsto r_t(m)$ is continuously differentiable with a derivative which satisfies*

$$\dot{r}_t(m) = q_\nu(m) - F_{\rho_t}(m) r_t(m).$$

PROOF. We will construct a curve of *a priori* new measures ξ_t with Radon-Nikodym derivatives $m \mapsto x_t(m)$ that have nice behavior in $t \geq 0$ for fixed $m \in \mathcal{M}$. We then prove that, as measures, $\xi = \rho$. Finally, we replace the arbitrary choice of Radon-Nikodym derivative r_t for ρ_t whose existence is guaranteed by Theorem 2.9 with x_t .

Because $t \mapsto \rho_t$ is continuous, it follows from a monotone class argument that $t \mapsto \rho_t(A)$ is Borel measurable for all $A \in \mathcal{B}(\mathcal{M})$ (recall that $\mathcal{B}(\mathcal{M})$ is the Borel σ -field on \mathcal{M}). We have $\mathcal{B}(\mathcal{M}) = \bigvee_{k \in \mathbb{N}} \mathcal{F}_k$, where $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots$ is a suitable increasing sequence of finitely generated sub- σ -fields. Suppose for $k \in \mathbb{N}$ that \mathcal{F}_k consists of the empty set and unions of collections of sets drawn from the Borel partition $\{A_{k,1}, \dots, A_{k,n(k)}\}$ of \mathcal{M} . Then, we may suppose that

$$r_t(m) := \begin{cases} \lim_{k \rightarrow \infty} \sum_{j=1}^{n(k)} \frac{\rho_t(A_{k,j})}{\zeta(A_{k,j})} \mathbf{1}_{A_{k,j}}(m), & \text{if the limit exists,} \\ 0, & \text{otherwise} \end{cases}$$

(see, for example, Section III-1 of [Nev75]). In particular, we may suppose that the map $(t, m) \rightarrow r_t(m)$ is Borel measurable.

When ρ_t is already known, for each fixed $m \in \mathcal{M}$ the function $F_{\rho_t}(m)$ is just some known Borel function of t and the Radon-Nikodym derivative $m \mapsto r_0(m)$ is just some known bounded function. Consider, then, for each fixed $m \in \mathcal{M}$ the

following ordinary integral equation for an unknown function $x_t(m)$, analogous to our dynamic equation

$$(2.14) \quad x_t(m) = r_0(m) + q_\nu(m)t - \int_0^t F_{\rho_s}(m) x_s(m) ds.$$

For every $m \in \mathcal{M}$, (2.14) has the continuously differentiable solution

$$(2.15) \quad x_t(m) = \exp \left\{ - \int_0^t F_{\rho_u}(m) du \right\} \\ \times \left(x_0(m) + q_\nu(m) \int_0^t \exp \left\{ \int_0^u F_{\rho_v}(m) dv \right\} du \right).$$

For $t \geq 0$, define measures ξ_t by $\xi_t(dm) = x_t(m)\zeta(dm)$. For ξ_t , for any Borel set A ,

$$(2.16) \quad \xi_t(A) = \int_A x_t(m) d\zeta(m) \\ = \int_A (r_0(m) + q_\nu(m)t) d\zeta(m) - \int_A \left(\int_0^t F_{\rho_s}(m) x_s(m) ds \right) d\zeta(m).$$

For ρ_t , with Radon-Nikodym derivatives r_t , we have a similar equation but with the integral over $m \in \mathcal{M}$ inside the integral over $s \in \mathbb{R}_+$, namely

$$(2.17) \quad \rho_t(A) = \int_A (r_0(m) + q_\nu(m)t) d\zeta(m) - \int_0^t \left(\int_A F_{\rho_s}(m) r_s(m) d\zeta(m) \right) ds.$$

We want to reverse the order of integration in the first equation for ξ . We can apply Fubini's Theorem so long as the function $(s, m) \mapsto x_s(m) \in \mathbb{R}$ is Borel measurable, which is the case (because of the equation defining $x_t(m)$) when $(s, m) \mapsto F_{\rho_s}(m)$ is Borel measurable. As we have already remarked, the map $s \mapsto \rho_s$ is continuous by construction, and the map $(\pi, m) \mapsto F_\pi(m)$ is always Borel measurable. (Under the conditions of Theorem 2.4 it is jointly continuous.) Thus, we can apply Fubini's Theorem.

After reversing the order of integration in (2.16), we subtract (2.17) from (2.16): The terms in the initial state and the mutation measure cancel

$$\xi_t(A) - \rho_t(A) = \int_0^t \int_A F_{\rho_s}(m) (x_s(m) - r_s(m)) d\zeta(m) ds.$$

Since $F_\pi(m) \leq \sigma$ for all π in \mathcal{H}^+ ,

$$|\xi_t(A) - \rho_t(A)| \leq \int_0^t \sigma \|x_s - r_s\|_\infty \left| \int_A d\zeta(m) \right| ds.$$

Thus,

$$(2.18) \quad |\xi_t(A) - \rho_t(A)| \leq \sigma \left(\int_0^t \|x_s - r_s\|_\infty ds \right) \zeta(A).$$

Define $\beta_t = \|x_t - r_t\|_\infty$. For Borel sets A with strictly positive measure, and particularly for sets A_n , $n \in \mathbb{N}$, defined by

$$A_n := \{m \in \mathcal{M} : |x_t(m) - r_t(m)| \geq \beta_t - 1/n\}.$$

we can divide both sides of (2.18) by $\zeta(A)$. The resulting left-hand side is no less than $\beta_t - 1/n$ for all $n \in \mathbb{N}$. Taking the supremum over $n \in \mathbb{N}$, we conclude for all

$t \geq 0$ that

$$\beta_t \leq \sigma \int_0^t \beta_s ds.$$

Gronwall's Inequality (see Appendix A.1) then forces $\beta_t \equiv 0$. It follows that $x_t(m) = r_t(m)$ outside of a set of $m \in \mathcal{M}$ with ζ measure zero so that, for every t , $x_t(m)$ is itself a Radon-Nikodym derivative of ρ_t with respect to ζ . In other words, $\xi_t = \rho_t$ and ξ_t satisfies the dynamic equation. We replace our original function $(t, m) \mapsto r_t(m)$ with the function $(t, m) \mapsto x_t(m)$ with the property that $t \mapsto x_t(m)$ is continuously differentiable for every $m \in \mathcal{M}$. To spare notation, we henceforth write $r_t(m)$ for $x_t(m)$. When we differentiate this new family of Radon-Nikodym derivatives with respect to t , we find, pointwise for all $t \geq 0$ and $m \in \mathcal{M}$ that

$$\dot{r}_t(m) = q_\nu(m) - F_{\rho_t}(m) r_t(m).$$

□

A continuously differentiable function $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ may have a set of zeros that is quite messy, but its positive part can still be expressed as an integral in a straightforward way. For applications of Theorem 2.10, the following elementary fact from real analysis is helpful.

LEMMA 2.11. *Suppose that the function $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ is an absolutely continuous function with $h(0) = 0$ and derivative \dot{h} defined Lebesgue-a.e. Write J for the indicator function of the set $\{t \in \mathbb{R}_+ : h(t) > 0\}$. Then, for all $t \in \mathbb{R}_+$,*

$$h^+(t) = h(t)J(t) = \int_0^t \dot{h}(s) J(s) ds.$$

PROOF. The fundamental theorem of calculus holds for absolutely continuous functions and so $h(t) = \int_0^t \dot{h}(s) ds$ for all $t \in \mathbb{R}_+$. Note that the function $t \mapsto g(t) := h(t)^+$ is also absolutely continuous. Therefore, the nonnegative function g is differentiable almost everywhere with respect to Lebesgue measure and $g(t) = \int_0^t \dot{g}(s) ds$ for all $t \in \mathbb{R}_+$.

Choose a point $t \in \mathbb{R}_+$ in the set of full Lebesgue measure where h is differentiable. Consider the following four alternatives.

- (1) If $h(t) > 0$, then $g(s) = h(s)$ for s in a neighborhood of t , and so g is differentiable at t and $\dot{g}(t) = \dot{h}(t)$.
- (2) If $h(t) < 0$, then $g(s) = 0$ in a neighborhood of t , and so g is differentiable at t and $\dot{g}(t) = 0$.
- (3) If $h(t) = 0$ and t is not an isolated point of the set $\{s \in \mathbb{R}_+ : h(s) = 0\}$, then there exists a sequence $(t_n)_{n \in \mathbb{N}}$ such that $t = \lim_{n \rightarrow \infty} t_n$ with $t_n \neq t$ and $h(t_n) = 0$. Then, $\dot{h}(t) = \lim_{n \rightarrow \infty} h(t_n)/(t_n - t) = \lim_{n \rightarrow \infty} g(t_n)/(t_n - t) = 0$. Since h is differentiable at t , $h(t_j)/(t_j - t)$ converges to zero for any sequence t_j converging to t . For any j , $g(t_j)$ either equals $h(t_j)$ or equals zero, so $g(t_j)/(t_j - t)$ also converges to zero, meaning that g is differentiable at t and $\dot{g}(t) = 0$.
- (4) If $h(t) = 0$ and t is an isolated point of the set $\{s \in \mathbb{R}_+ : h(s) = 0\}$, then g need not be differentiable at t , but the set of such t is countable and hence Lebesgue null.

Thus, $\dot{g}(t) = \dot{h}(t)J(t)$ for Lebesgue-a.e. $t \in \mathbb{R}_+$ and we conclude that

$$h^+(t) = g(t) = \int_0^t \dot{g}(s) ds = \int_0^t \dot{h}(s) J(s) ds$$

for all $t \in \mathbb{R}_+$. □

2.6. Mutation measures with infinite total mass

The assumption that the mutation measure ν has a finite total mass underlies our development of the model, our use of Wasserstein metrics, and our main theorems in Chapters 5 through 8. However, the dynamic system itself, taken in isolation, can be defined for some cases when $\nu(M)$ is infinite and the selective cost S is chosen suitably. For example, the case where ν is a multiple of Lebesgue measure on the positive reals is important for demographic applications. This section sketches how the dynamical system for such measures can be handled by viewing the state space of the dynamical system as a subset of $L^\infty(\mathcal{M}, \zeta)$ for a suitable reference measure ζ .

NOTATION 2.12.

- Write $\bar{\mathcal{H}}$ for the space of signed measures on \mathcal{M} such that the trace on any bounded set Borel set $A \subset \mathcal{M}$ belongs to \mathcal{H} . The space $\bar{\mathcal{H}}$ is equipped with the σ -field generated by the maps $\pi \mapsto \pi(A)$ as A ranges over the bounded Borel subsets of \mathcal{M} . (The trace of a measure π on a Borel set A is the measure $\pi(\cdot \cap A)$.)
- Denote the subset of nonnegative measures in $\bar{\mathcal{H}}$ by $\bar{\mathcal{H}}_+$.
- Denote the set of integer-valued elements of $\bar{\mathcal{H}}_+$ by $\bar{\mathcal{G}}$.

Fix $\zeta \in \bar{\mathcal{H}}_+$, and identify $L_+^\infty(\mathcal{M}, \zeta)$ with $\bar{\mathcal{K}}_+$, the set of elements of $\bar{\mathcal{H}}_+$ that are absolutely continuous with respect to ζ with essentially bounded Radon-Nikodym derivatives.

If $\pi \in \bar{\mathcal{H}}_+$, then the Poisson random measure X^π with intensity measure π takes values in the set $\bar{\mathcal{G}}$ almost-surely. Suppose that the selective cost function $S : \bar{\mathcal{G}} \rightarrow \mathbb{R} \cup \{+\infty\}$ is Borel measurable with $S(0) = 0$ and $S(g) \leq S(g+h)$. Then, $\pi \mapsto \mathbb{E}[S(X^\pi)]$ is a Borel measurable map from $\bar{\mathcal{K}}_+$ to $\mathbb{R} \cup \{+\infty\}$ and the map $(\pi, m) \mapsto \mathbb{E}[S(X^\pi + \delta_m)]$ from $\bar{\mathcal{K}}_+ \times \mathcal{M}$ to $\mathbb{R} \cup \{+\infty\}$ is also Borel measurable.

Suppose that the selective cost is such that

$$\mathbb{E}[S(X^\pi)] \leq \mathbb{E}[S(X^\pi + \delta_m)] < \infty$$

for all $\pi \in \bar{\mathcal{K}}_+$ and $m \in \mathcal{M}$. Then,

$$(\pi, m) \mapsto \mathbb{E}[S(X^\pi + \delta_m) - S(X^\pi)] =: F_\pi(m)$$

is a well-defined and Borel measurable map from $\bar{\mathcal{K}}_+ \times \mathcal{M}$ to \mathbb{R}_+ .

REMARK 2.13. An examination of the proofs in Section 2.5 of Theorems 2.9 and 2.10 shows that they apply in this broader setting and do not depend on the finiteness of $\nu(\mathcal{M})$.

CHAPTER 3

Equilibria

One of the primary problems concerning our dynamical systems is to understand their asymptotic behavior over time. We begin the analysis of these asymptotics by identifying when fixed points of the dynamical system 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 dynamical system (we use the terms 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

$$(3.1) \quad F_{\rho_*} \frac{d\rho_*}{d\nu} = 1.$$

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 $t \geq 0$ if $\rho_0 \in U$. It is called *attractive* if it is stable and there is a neighborhood U_0 of ρ_* such that $\lim_{t \rightarrow \infty} \rho_t = \rho_*$ 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

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

where $\tilde{\rho} \leq \rho_* \leq \tilde{\rho}'$, 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 dynamical system 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 3.12.) 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 converge down to it, so the minimal equilibrium is box attractive. We give an iterative procedure for computing the minimal equilibrium that avoids following the dynamical system to large times in Section 3.7. Moreover, the iteration is a useful theoretical tool in Section 3.8, 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 3.9 we apply the above results to the demographic selective 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 3.17 and Theorem 3.18 can often be verified in specific cases, and we have found that they can hold when $\nu(\mathcal{M})$ is only moderately small.

3.1. Introductory example: One-dimensional systems

Suppose as in Section 1.2 that selective costs depend only on the number of mutations and the original space of loci has been pushed forward into a space \mathcal{M} consisting of a single point. The space of genotypes \mathcal{G} may be identified with nonnegative integers \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 on \mathcal{G} may also be identified with \mathbb{R}_+ . The dynamical system $(\rho_t)_{t \geq 0}$ is \mathbb{R}_+ -valued and satisfies

$$(3.3) \quad \dot{\rho}_t = \nu - F_{\rho_t} \rho_t,$$

where

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

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 3.1. The number of equilibria may be 0, 1, 2, 3, or 4, depending on the value of ν .

However, we can say quite generally the following things about one-dimensional systems.

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

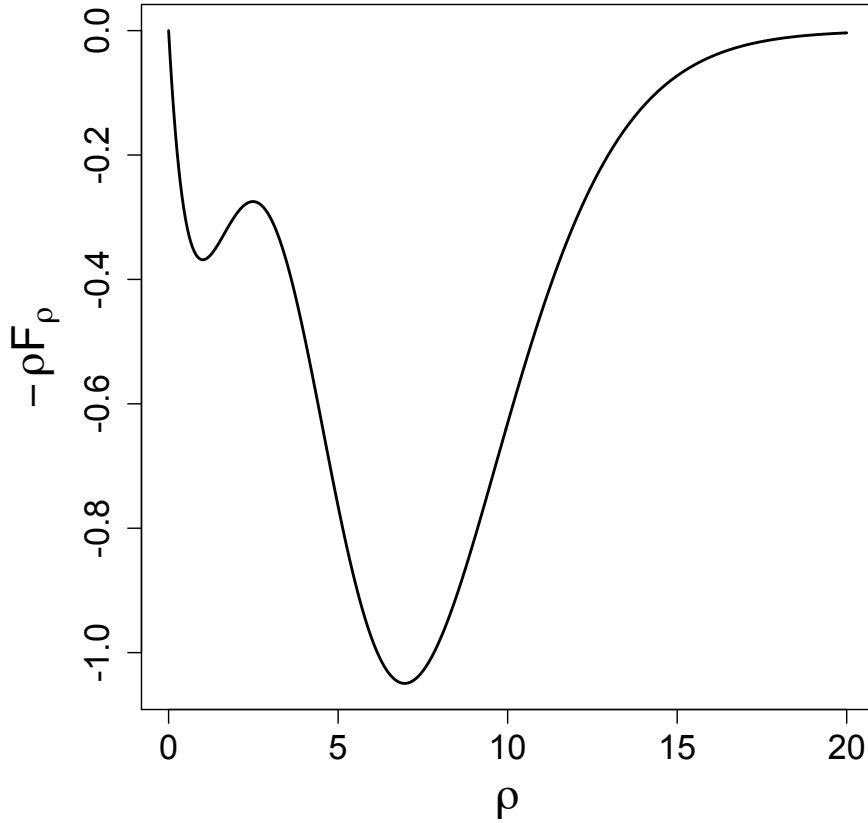


FIGURE 3.1. Plot of the function $-\rho F_\rho$ when $S(k) = \mathbf{1}_{\{k \geq 1\}} + \mathbf{1}_{\{k \geq 6\}}$.

- 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 costs

It is probably apparent that one cannot hope to get anything like a closed-form solution for the dynamical system (2.10) for general selective costs. However, we show in this section that it is possible to solve (2.10) explicitly when the selective cost has the multiplicative form

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

for some $\theta : \mathcal{M} \rightarrow \mathbb{R}_+$. Moreover, in this case it is possible to analyze whether $\lim_{t \rightarrow \infty} \rho_t$ exists, and we 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 Theorem 2.10 that the measure ρ_t has a Radon-Nikodym derivative $m \mapsto r_t(m)$ against ν for all $t \geq 0$ such that for each $m \in \mathcal{M}$ the function $t \mapsto r_t(m)$ is continuously differentiable and solves the equation

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

Substituting from (3.5), we see that the expected marginal cost is given by the product of $1 - e^{-\theta(m)}$ and $\mathbb{E}[\exp\{-\sum_{m' \in g} \theta(m')\}]$. The first factor $1 - e^{-\theta(m)}$ is $S(\delta_m)$, the selective cost of a singleton mutation, which is the same as $F_0(m)$. The second factor, the expectation, can be evaluated via the Poisson identity (1.13) and it equals $\exp\{-\int (1 - e^{-\theta(m')}) d\rho_t(m)\}$.

We now define analogues of moments for $k \in \mathbb{N}$ by the equations

$$\begin{aligned} a_k &= \int e^{-k\theta(m')} d\nu(m') \\ (3.6) \quad b_k(t) &= \int e^{-k\theta(m')} r_t(m') d\nu(m') \end{aligned}$$

Our differential equation becomes

$$(3.7) \quad \frac{dr_t(m)}{dt} = 1 - (1 - e^{-\theta(m)}) e^{b_1(t) - b_0(t)} r_t(m).$$

This is an ordinary differential equation for each $m \in \mathcal{M}$ whose solution, a special case of (2.15), is given by

$$\begin{aligned} r_t(m) &= \exp \left\{ - \int_0^t (1 - \exp\{-\theta(m)\}) \exp\{b_1(s) - b_0(s)\} ds \right\} \\ &\quad \times \left[r_0(m) + \int_0^t \exp \left\{ \int_0^s (1 - \exp\{-\theta(m)\}) \exp\{b_1(r) - b_0(r)\} dr \right\} ds \right]. \end{aligned}$$

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

Of course, this benefits us only if we have an autonomous system of equations in just the functions b_0 and b_1 . Combining (3.7) with (3.6) we have

$$(3.8) \quad \frac{db_k(t)}{dt} = a_k + \exp\{b_1(t) - b_0(t)\} (b_{k+1}(t) - b_k(t)).$$

We introduce three generating functions

$$\begin{aligned} A(z) &:= \sum_{k=0}^{\infty} a_k z^k / k! \\ B(z, t) &:= \sum_{k=0}^{\infty} b_k(t) z^k / k! \\ C(z) &:= B(z, 0) = \sum_{k=0}^{\infty} b_k(0) z^k / k!. \end{aligned}$$

The system of ordinary differential equations (3.8) 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 obtain the general solution (which involves the unknown functions b_0 and b_1), we have to impose the boundary conditions

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

and

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

to select the solution to our original differential equation.

Using *Mathematica*, the PDE is seen to have the solution

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

Therefore, the functions b_0 and b_1 solve the following system of equations expressed in terms of the generating functions $A(z)$ and $C(z)$ and their derivatives $A'(z)$ and $C'(z)$

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

and

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

Note that this system of ordinary differential equations is 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 with 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 (2.10) in \mathcal{H}^+ when the selection cost is the multiplicative one of (3.5). Note that if ρ_* is an equilibrium, then we have

$$(3.9) \quad \nu(dm) - \mathbb{E}[S(X^{\rho_*} + \delta_m) - S(X^{\rho_*})] \rho_*(dm) = 0,$$

and so ρ_* has a Radon-Nikodym derivative r_* with respect to ν that satisfies

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

where $b_k^* := \int \exp\{-k\theta(m)\} r_*(m) d\nu(m)$ for $k \in \{0, 1\}$ (cf. (3.7)). Although the equation for r_* sometimes makes sense even when ν and ρ have infinite total mass, as discussed in Section 2.6, here we are only considering ν and ρ in \mathcal{H}^+ , the space of finite positive measures. An equilibrium exists in this space if and only if, first,

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

and, second, the conditions defining b_0^* and b_1^* are consistent with the formula for r_* , so that there is a constant $c > 0$ such that

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

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

$$\nu(\mathcal{M}) \leq \sup_{x \geq 0} x e^{-x} = e^{-1}.$$

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.

We are now in a position to give an example which previews the general approaches in the following sections. Take $\mathcal{M} = [1, \infty) \subset \mathbb{R}_+$. We generate multiplicative selective costs from the function $\theta(m) = \log(m/(m-1))$, so that

$$S(\delta_m) = F_0(m) = 1 - e^{-\theta(m)} = 1/m.$$

We take mutation measure to be the measure $\nu(dm) = u(2/m^3) dm$ with total mass $\nu(\mathcal{M}) = u$. We have just seen that when $u < e^{-1}$ there exists an equilibrium whose Radon-Nikodym derivative with respect to ν is given by

$$r_*(m) = e^{c(u)} m.$$

Here $c(u)$ is the smaller root of the equation $ce^{-c} = u$.

The equilibrium Radon-Nikodym derivative is not bounded. It increases linearly with $m \in \mathcal{M}$, even though the equilibrium measure $\rho_*(dm) = r_*(m)\nu(dm)$ has finite total mass and does belong to \mathcal{H}^+ . If we want a bounded continuous function of $m \in \mathcal{M}$, we need to turn to the product $S(\delta_m)r_*(m)$. In this example the product equals $\exp\{c(u)\}\mathbf{1}$, where $\mathbf{1}$ is the function taking the value 1 for every $m \in \mathcal{M}$.

For every $m \in \mathcal{M}$, the product is a continuously differentiable function of the total mass u of the mutation measure and is in fact for small u the solution of the differential equation

$$\frac{d}{du} S(\delta_m)r_*(m) = \mathbf{1} e^{c(u)} \frac{dc(u)}{du}.$$

If we did not know r_* , we could arrive at this equation by differentiating (3.9). We exploit this approach to give a general construction of equilibria for small $\nu(\mathcal{M})$ in Section 3.4.

The scale factor in this example with multiplicative costs can be obtained by iteration of the map $c \mapsto ue^c$ starting from $c = 0$. This scalar-valued iteration corresponds to an iterative scheme for generating r_* itself, namely $r \mapsto 1/F_{r\nu}$. For selective costs which are concave, in a sense defined in Section 3.5, an iterative scheme along these lines enables us to prove stability results for equilibria in Section 3.8.

3.3. Fréchet derivatives

We need 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 space $(Y, \|\cdot\|_Y)$. Extending the usual definition slightly, we say that a map $\Phi : X_+ \rightarrow Y$, is Fréchet differentiable at $x \in X_+$, if there is map $D_x\Phi : U_x \rightarrow Y$ with the properties

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

$D_x\Phi[tz] = tD_x\Phi[z]$ 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 , $\|D_x\Phi[z]\|_Y \leq C\|z\|_X$ 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. The definitions depend on the choice of cone. We shall be applying them to the cone \mathcal{H}^+ and taking care to restrict directions to such positive measures.

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

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

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 space $(X, \|\cdot\|_X)$. Suppose that a curve $\psi : I \rightarrow X$ 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 $\psi : I \rightarrow X_+$ is differentiable at $t \in I$ and the map $\Phi : X_+ \rightarrow Y$ is differentiable at ψ_t . Then, the curve $\Phi \circ \psi : I \rightarrow Y$ is differentiable at t with derivative $D_{\psi_t}\Phi[\dot{\psi}_t]$.*

We also have a particular analogue of the product rule. Recall that \mathcal{H} is the Banach space of finite signed measures on \mathcal{M} equipped with the Wasserstein metric, and \mathcal{H}^+ is the cone of positive measures within \mathcal{H} .

NOTATION 3.5.

- 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.
- For $\eta \in \mathcal{H}^+$ and $m', m'' \in \mathcal{M}$, define the kernel

$$\tilde{K}_\eta(m', m'') := \mathbb{E} \left[S(X^\eta + \delta_{m'} + \delta_{m''}) - S(X^\eta + \delta_{m''}) - S(X^\eta + \delta_{m'}) + S(X^\eta) \right].$$

LEMMA 3.6. *Consider an interval $I \subseteq \mathbb{R}$ and two curves $\gamma : I \rightarrow \mathcal{H}^+$ and $f : I \rightarrow C_b(\mathcal{M}, \mathbb{R})$. Suppose that γ and f are differentiable at $t \in I$. Define a curve $\beta : I \rightarrow \mathcal{H}^+$ 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. This 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. \square

By our standing assumption, the conditions of Theorem 2.4 are in place. By the Lipschitz condition on S , with constant σ , via Lemma 2.5, the absolute value of the sum of the first two terms in \tilde{K} is bounded by $\sigma \|\delta_{m''}\|_{\text{Was}} = \sigma$ and the absolute value of the sum of the second two terms is bounded by $\sigma \|\delta_{m'}\|_{\text{Was}} = \sigma$. No separate uniform bound on S is required. Thus, the map $(\eta, m', m'') \mapsto \tilde{K}_\eta(m', m'')$ is bounded by 2σ . Ideas similar to those behind Lemma A.12 and Lemma 2.6 establish that this kernel gives the Fréchet derivative of the map $\eta \mapsto F_\eta(\cdot)$.

LEMMA 3.7. *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*

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

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

LEMMA 3.8. *For any $\rho, \rho', \eta \in \mathcal{H}^+$ with $\rho \leq \rho'$,*

$$\|D_\rho F[\eta]\|_\infty \leq 2\sigma \eta(\mathcal{M}),$$

and

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

3.4. Existence of equilibria via perturbation

We now proceed to prove the existence of equilibria when the total mutation rate is sufficiently small. We fix a selective cost function S and define a family of mutation measures $\nu^{(u)}$ for $u \in \mathbb{R}_+$ and some $\zeta \in \mathcal{H}^+$ by

$$\nu^{(u)} = u \zeta.$$

In other words, the mutation measures are scalar multiples of each other differing only in total mass. A family of corresponding dynamical systems starting from the null state is given by

$$(3.10) \quad \rho_t^{(u)} = ut\zeta - \int_0^t D\rho_s^{(u)} ds.$$

An equilibrium for any one of these systems has a Radon-Nikodym derivative with respect to ζ .

THEOREM 3.9. *Consider a selective cost function S with $S(0) = 0$ that satisfies the monotonicity and Lipschitz conditions of Theorem 2.4. Suppose, moreover, that*

$$\inf_{m \in \mathcal{M}} S(\delta_m) > 0.$$

Then, there exists $U > 0$ and a curve $(u, m) \mapsto p^{(u)}(m)$ for $(u, m) \in [0, U] \times \mathcal{M}$ such that the following hold.

- *For each $u \in [0, U]$ the function $m \mapsto p^{(u)}(m)$ is a bounded continuous function of $m \in \mathcal{M}$.*
- *For each $m \in \mathcal{M}$, the function $u \mapsto p^{(u)}(m)$ is a continuously differentiable function of $u \in [0, U]$.*
- *The measure*

$$\rho^{(u)}(dm) := p^{(u)}(m) \zeta(dm)$$

is a finite equilibrium for (3.10) in \mathcal{H}^+ for all $u \in [0, U]$. That is,

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

PROOF. If measures $\rho^{(u)}$ do exist satisfying the conditions of the theorem, then we expect their Radon-Nikodym derivatives to satisfy an equation obtained by differentiating both sides of

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

Taking advantage of our expression for the Fréchet derivative of F_ρ and applying Lemmas 3.6 and 3.7 with $\eta' = p^{(u)}\zeta$ and $\eta'' = \frac{dp^{(u)}}{du}\zeta$, the desired relationship comes out to be

$$(3.11) \quad \mathbf{1} = F_{p^{(u)}\zeta}(m) \frac{dp^{(u)}}{du}(m) - p^{(u)}(m) \int_{\mathcal{M}} (-\tilde{K}_{p^{(u)}\zeta}(m, m')) \frac{dp^{(u)}}{du}(m') d\zeta(m').$$

Here $\mathbf{1} \in C_b(\mathcal{M}, \mathbb{R}_+)$ is the function with constant value 1.

Our strategy is to start with an operator equation of similar form, show that it does have solutions in $C_b(\mathcal{M}, \mathbb{R}_+)$, derive a differential equation from them, solve it, and identify $p^{(u)}$ with a suitable function of the solution.

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

$$(3.12) \quad T_p(q) := p_+(m') \left[\int_{\mathcal{M}} \tilde{K}_{p_+}(m', m'') q(m'') d\zeta(m'') \right] + F_{p_+}(m') q(m').$$

Let \mathcal{D} be the set of functions $p \in C_b(\mathcal{M}, \mathbb{R})$ 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 (here only) $\|\cdot\|$ denotes the operator norm.

By Lemmas 3.8 and 2.6 we can bound $\|T_p - T_0\|$ in terms of the Lipschitz constant σ on S by

$$\begin{aligned} \|T_p - T_0\| &\leq 2\sigma\zeta(\mathcal{M})\|p\|_\infty + 8\sigma\|p_+\|_{\text{was}} \\ &\leq 10\sigma\zeta(\mathcal{M})\|p\|_\infty. \end{aligned}$$

We conclude that \mathcal{D} includes an open ball around 0.

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

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

$$\begin{aligned} \mathbf{0} &= \mathbf{1} - \mathbf{1} = T_p[L(p)] - T_0[L(0)] \\ &= T_0[L(p) - L(0)] + (T_p - T_0)[L(0)] + (T_p - T_0)[L(p) - L(0)]. \end{aligned}$$

Thus,

$$L(p) = L(0) - T_0^{-1} \left((T_p - T_0)[L(0)] + (T_p - T_0)[L(p) - L(0)] \right).$$

Now T_0^{-1} is the diagonal operator that multiplies a function of $m \in \mathcal{M}$ by the bounded function $m \mapsto 1/S(\delta_m) = L(0)$, and we can combine our bound on the operator norm of $T_p - T_0$ to show that for a suitable constant c

$$\begin{aligned} \|L(p) - L(0)\|_\infty &\leq \|L(0)\|_\infty \left(\|T_p - T_0\| \|L(0)\|_\infty + \|T_p - T_0\| \|L(p) - L(0)\|_\infty \right) \\ &\leq c\|p\|_\infty \|L(0)\|_\infty^2 + c\|p\|_\infty \|L(0)\|_\infty \|L(p) - L(0)\|_\infty. \end{aligned}$$

By requiring $\|p\|_\infty$ to be small enough, we can make the first term on the right arbitrarily small and make $1 - c\|p\|_\infty \|L(0)\|_\infty$ arbitrarily close to 1, so we can make $\|L(p) - L(0)\|_\infty$ arbitrarily small. Since $L(0)(m)$ is bounded away from 0, it follows that there is a neighborhood $\mathcal{D}' \subset \mathcal{D}$ of 0 such that $0 < \inf_{p \in \mathcal{D}'} \inf_{m \in \mathcal{M}} L(p)(m)$.

Furthermore, again by Lemma VII.6.1 of [DS88], within this neighborhood, the map from an operator to its inverse is a homeomorphism and

$$\|L(p') - L(p'')\|_\infty \leq \frac{\|L(p')\|_\infty^2 \|T_{p'} - T_{p''}\|}{1 - \|T_{p'} - T_{p''}\| \|L(p')\|_\infty}.$$

It follows that $L(p)$ satisfies a Lipschitz condition.

Then, 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, for $0 \leq u \leq U$, $p^{(u)} = \int_0^u L(p^{(v)}) dv$ satisfies the requirements in the conclusion of the theorem. \square

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 3.10. Under a few mild constraints, we show in Theorem 3.12 that concave selective costs yield monotonic solutions $(\rho_t)_{t \geq 0}$ when started from the pure wild type population $\rho_0 = 0$, and hence such

systems must either diverge to a measure with infinite total mass or converge to an element of \mathcal{H}^+ . Corollary 3.17 gives a further condition that is sufficient to ensure that 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 3.10. A selective cost function S is *concave* if

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

LEMMA 3.11. A selective cost S is concave if and only if

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

for all $g \in \mathcal{G}$ and $m, m' \in \mathcal{M}$. Moreover, if S is concave, then

$$F_\pi(m) \geq F_{\pi+\eta}(m)$$

for $\pi, \eta \in \mathcal{H}^+$ and $m \in \mathcal{M}$.

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

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

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 case proceeds exactly as above.

Finally, the Poisson random measures $X^{\pi+\eta}$ is distributed like the sum of independent copies of the Poisson random measures X^π and X^η , and so

$$\begin{aligned} F_{\pi+\eta}(m) &= \mathbb{E}[S(X^\pi + X^\eta + \delta_x) - S(X^\pi + X^\eta)] \\ &\leq \mathbb{E}[S(X^\pi + \delta_x) - S(X^\pi)] \\ &= F_\pi(m). \end{aligned}$$

□

THEOREM 3.12. Fix a mutation measure $\nu \in \mathcal{H}^+$ and a selective cost $S : \mathcal{G} \rightarrow \mathbb{R}_+$, that satisfies the conditions

- $S(0) = 0$,
- $S(g) \leq S(g+h)$ for all $g, h \in \mathcal{G}$,
- for some constant σ , $|S(g) - S(h)| \leq \sigma \|g - h\|_{\text{Was}}$, for all $g, h \in \mathcal{G}$,
- $S(g+h+k) - S(g+h) \leq S(g+k) - S(g)$ for all $g, h, k \in \mathcal{G}$.

If $\dot{\rho}_0 \geq 0$ then the solution of (2.10) guaranteed by Theorem 2.4 satisfies $\rho_s \leq \rho_t$ for all $0 \leq s \leq t < \infty$. If $\dot{\rho}_0 \leq 0$, then the solution satisfies $\rho_s \geq \rho_t$ for all $0 \leq s \leq t < \infty$

PROOF. By Definition 2.3, $F_\eta(m) = \mathbb{E}[S(X^\eta + \delta_m) - S(X^\eta)]$ for $\eta \in \mathcal{H}^+$ and $m \in \mathcal{M}$. By Lemmas 2.5 and 2.6, thanks to the Lipschitz bound on S , it is clear that $\eta \mapsto F_\eta(\cdot)$ is a continuous map from \mathcal{H}^+ to $C_b(\mathcal{M}, \mathbb{R})$.

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

$$(3.14) \quad \dot{\rho}_t = \nu - F_{\rho_t} \cdot \rho_t.$$

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

By Lemma 3.4 and Lemma 3.7, the curve $t \mapsto F_{\rho_t}$, $t \in \mathbb{R}_+$, is differentiable, with

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

The value of this derivative at m' equals the integral $\int_{\mathcal{M}} \tilde{K}_{\rho_t}(m', m'') \dot{\rho}_t(dm'')$, where the kernel \tilde{K} is defined in Notation 3.5. By the concavity condition on S , for all m' and m'' in \mathcal{M} ,

$$0 \leq -\tilde{K}_{\rho_t}(m', m'') \leq F_{\rho_t}(m').$$

Furthermore, by Lemma 3.6,

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

Suppose now that $\dot{\rho}_0 \geq 0$. Define a negatively-rescaled version of $t \mapsto \dot{\rho}_t$ by

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

Then,

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

By assumption, $\gamma_0 = -\dot{\rho}_0 \leq 0$. For any Borel set $B \subseteq \mathcal{M}$

$$\gamma_t(B) \leq \gamma_t(B) - \gamma_0(B) = \int_0^t \frac{d}{ds} \gamma_s(B) ds.$$

Writing out the derivative in terms of the kernel \tilde{K} and expressing $\dot{\rho}_t$ in terms of γ_t , we obtain

$$\begin{aligned} \gamma_t(B) &\leq \int_0^t \left(\int_B \left[\int_{\mathcal{M}} -\tilde{K}_{\rho_s}(m', m'') \exp\left\{-\int_0^s F_{\rho_u}(m'') du\right\} d\gamma_s(m'') \right] \right. \\ &\quad \left. \times \exp\left\{\int_0^s F_{\rho_u}(m') du\right\} d\rho_s(m') \right) ds. \end{aligned}$$

Let $\gamma_s = \gamma_s^+ - \gamma_s^-$ be the Hahn-Jordan decomposition of γ_s into its positive and negative parts. Since $-\tilde{K}$ and the exponential factors are nonnegative, γ_s can be

replaced by its positive part in the inequality for $\gamma_t(B)$ thanks to the upper bound on $-\tilde{K}$. The inner integral over m'' is bounded above by $\gamma_s^+(\mathcal{M})F_{\rho_s}(m')$. Hence,

$$(3.15) \quad \gamma_t(B) \leq \int_0^t \gamma_s^+(\mathcal{M}) \left(\int_B F_{\rho_s}(m') \exp \left\{ \int_0^s F_{\rho_u}(m') du \right\} \rho_s(dm') \right) ds.$$

For any positive T , set C_T equal to the product of two bounds, namely, first, the bound on the exponential multiplier

$$\exp \left\{ \int_0^s F_{\rho_u}(m') du \right\} \leq \sigma T$$

and, second, the bound of the integral

$$\int_B F_{\rho_s}(m') \rho_s(dm') \leq \sigma \rho_s(B) \leq \sigma(\rho_0(\mathcal{M}) + T\nu(\mathcal{M})).$$

Both bounds hold for $s \leq t \leq T$ by Lemma 2.5, and $\rho_s(\mathcal{M}) \leq \rho_0(\mathcal{M}) + s\nu(\mathcal{M})$ because F and ρ are nonnegative in (3.14).

Since C_T is finite, we have shown that

$$(3.16) \quad \gamma_t(B) \leq C_T \int_0^t \gamma_s^+(\mathcal{M}) ds.$$

Put $\beta_t := \sup_{B \subseteq \mathcal{M}} \gamma_t(B)$, where the supremum is taken over Borel sets including the null set, so that β_t is nonnegative. Now, $\gamma_s^+(\mathcal{M}) = \beta_s$. Hence, we have shown that

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

for $0 \leq t \leq T$. By Gronwall's Inequality (see Appendix A.1), this equation implies that $\beta_t \equiv 0$ 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. This finishes the proof of the claim for the case $\dot{\rho}_0 \geq 0$.

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 the rest of the proof carries through as before. \square

COROLLARY 3.13. *Suppose the conditions of Theorem 3.12 hold and there exists $\rho_{**} \in \mathcal{H}^+$ satisfying the equilibrium condition*

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

for (2.10). For the dynamic system $(\rho_t)_{t \geq 0}$ started at $\rho_0 = 0$, $\rho_t \uparrow \rho_* \in \mathcal{H}^+$, where $\rho_* \leq \rho_{**}$ and ρ_* is also an equilibrium for (2.10).

The following Comparison Lemma, also proved via Gronwall's Inequality, is a powerful tool for applications. It treats pairs of solutions ρ'_t and ρ''_t in which, informally speaking, ρ' starts ahead of ρ'' and the marginal selective costs slowing the progress of ρ' are always less than the marginal costs slowing ρ'' . In such a race ρ' always keeps the lead.

LEMMA 3.14. *Consider two selective cost functions S' and S'' that satisfy the conditions of Theorem 3.12. Let ρ' and ρ'' be the corresponding solutions of (2.10). 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, for all $t \geq 0$ we have $\rho'_t \geq \rho''_t$.*

PROOF. Define a signed measure $\xi_t = \rho_t'' - \rho_t'$ with Hahn-Jordan decomposition $\xi_t = \xi_t^+ - \xi_t^-$. We seek to prove that the positive part ξ_t^+ is zero for all $t \geq 0$.

Set

$$\eta_t := \rho_t' \wedge \rho_t'' = \rho_t'' - \xi_t^+ = \rho_t' - \xi_t^-$$

and

$$\beta_t := \sup_{A \subseteq \mathcal{M}} \{\xi_t(A)\} = \|\rho_t'' - \eta_t\|_{TV},$$

where the supremum is over the Borel subsets of \mathcal{M} . Here $\|\cdot\|_{TV}$ is the total variation norm, which, as remarked in Section 2.1, dominates the Wasserstein metric on \mathcal{H} . The function β is nonnegative, and we proceed to show that it is actually zero, so that there is no Borel set on which ρ'' is bigger than ρ' .

Let $(m, \rho) \mapsto F'_\rho(m)$ be the expected cost function corresponding to S' , and $(m, \rho) \mapsto F''_\rho(m)$ be the expected cost function corresponding to S'' . As before, via Lemmas 2.5 and 2.6, for fixed ρ , $F'_\rho(\cdot)$ and $F''_\rho(\cdot)$ can be regarded as elements of the space $C_b(\mathcal{M}, \mathbb{R})$ of bounded continuous functions on \mathcal{M} equipped with the supremum norm. Because $\eta \leq \rho'$, our assumptions imply the inequalities

$$F'_{\rho'}(m) \leq F''_{\rho'}(m) \leq F''_\eta(m)$$

for all $m \in \mathcal{M}$.

By Lemma 2.6 we have the Lipschitz bound

$$\sup_{m \in \mathcal{M}} \left(F''_{\rho_t'}(m) - F''_{\eta_t}(m) \right) = \left\| F''_{\rho_t'} - F''_{\eta_t} \right\|_\infty \leq 8\sigma \|\rho_t'' - \eta_t\|_{\text{Was}} \leq 8\sigma\beta_t.$$

By Lemma 2.5 we also have $\|F''\| \leq 2\sigma$. Bearing in mind that F'' is nonnegative, we see that

$$\begin{aligned} (3.17) \quad \dot{\xi}_t &= F'_{\rho_t'} \cdot \rho_t' - F''_{\rho_t'} \cdot \rho_t'' \\ &= -(F''_{\rho_t^*} - F'_{\rho_t'}) \cdot \rho_t' + (F''_{\rho_t^*} - F''_{\rho_t'}) \cdot \rho_t'' + (F''_{\rho_t^*}) \cdot (-1) \cdot (\rho_t'' - \rho_t') \\ &\leq 0 + 8\sigma\beta_t\rho_t'' + 2\sigma\xi_t^-. \end{aligned}$$

Our assumption that $\rho_0' \geq \rho_0''$ makes $\xi_0^+ = 0$. Let $(s, m) \mapsto x_s(m)$ be the function supplied by Theorem 2.10 with the properties that the function $m \mapsto x_s(m)$ is a Radon-Nikodym derivative of ξ_s with respect to the measure $\zeta := \rho_0 + \nu$ for every $s \geq 0$ and for every $m \in \mathcal{M}$ the function $s \mapsto x_s(m)$ is a continuously differentiable function. Write J for the indicator function of the set of $\{(s, m) \in \mathbb{R}_+ \times \mathcal{M} : x_s(m) > 0\}$. For a Borel set A we have, by Lemma 2.11,

$$\xi_t^+(A) = \int_0^t \left(\int_A J(s, m) d\dot{x}_s(m) d\zeta(m) \right) ds.$$

Since $J(s, m)x_s^-(m)$ vanishes, we also have

$$\int_A J(s, m)\dot{x}_s(m) d\zeta(m) \leq 8K\beta_s \int_A J(s, m) d\rho_s''(m).$$

Thus, the inequality (3.17) on $\dot{\xi}_s$ implies

$$\xi_t^+(A) \leq 8K \int_0^t \beta_s \rho_s''(\mathcal{M}) ds.$$

Note that $\sup_{s \in [0, T]} \rho'_s(\mathcal{M}) \leq \rho''_0(\mathcal{M}) + T\nu(\mathcal{M})$. Taking the supremum over Borel sets A on the left-hand side, we have a new constant K' such that

$$\beta_t \leq K' \int_0^t \beta_s ds.$$

Gronwall's Inequality (Appendix A.1) then gives $\beta_t = 0$ for all t in the interval $[0, T]$, and hence for all $t \geq 0$. Thus, $\rho''_t \leq \rho'_t$ for all $t \geq 0$. \square

An alternative proof along the same lines takes advantage of the L^∞ norm. It is crafted so that it does not depend on the assumption of finite total mass for ν .

LEMMA 3.15. *Consider two selective cost functions S' and S'' and corresponding solutions ρ' and ρ'' under the conditions of Theorem 2.9. Suppose for all $g \in \mathcal{G}$ and $m \in \mathcal{M}$ that*

$$S'(g + \delta_m) - S'(g) \leq S''(g + \delta_m) - S''(g).$$

Suppose also that

$$\rho'_0 \geq \rho''_0.$$

Then, for all $t \geq 0$ we have

$$\rho'_t \geq \rho''_t.$$

PROOF. Set $\eta_t = \rho'_t \wedge \rho''_t$. Theorem 2.10 supplies functions $(t, m) \mapsto r'_t(m)$ and $(t, m) \mapsto r''_t(m)$ such that $m \mapsto r'_t(m)$ and $m \mapsto r''_t(m)$ are the Radon-Nikodym derivatives of ρ'_t and ρ''_t with respect to $\zeta = \rho'_0 + \nu$ for each $t \geq 0$, and $t \mapsto r'_t(m)$ and $t \mapsto r''_t(m)$ are continuously differentiable for all $m \in \mathcal{M}$. Write q_ν for the Radon-Nikodym derivative of ν with respect to ζ . For $\pi \in \mathcal{K}^+$, let F'_π and F''_π be the expected cost functions corresponding respectively to S' and S'' .

Set $x_t(m) = r''_t(m) - r'_t(m)$. Let J be the indicator function of the subset of $\{(t, m) \in \mathbb{R}_+ \times \mathcal{M} : x_t(m) > 0\}$, which, by Theorem 2.10 is Borel measurable. We now appeal to Lemma 2.11. Separately for every $m \in \mathcal{M}$ for all $t \geq 0$

$$\begin{aligned} x_t(m)J(t, m) &= x_0(m) + \int_0^t \dot{x}_s(m)J(s, m)ds \\ &= x_0(m) + \int_0^t \left[F'_{\rho'_s}(m) r'_s(m) - F''_{\rho'_s}(m) r''_s(m) \right] J(s, m) ds. \end{aligned}$$

We write the integrand as the sum of three terms as follows

$$\begin{aligned} \dot{x}_s(m)J(s, m) &= + \left[F''_{\eta_s}(m) - F''_{\rho'_s}(m) \right] r''_s(m)J(s, m) \\ &\quad - \left[F''_{\eta_s}(m) - F'_{\rho'_s}(m) \right] r'_s(m)J(s, m) \\ &\quad + \left[F''_{\eta_s}(m) \right] (-1) (r''_s(m) - r'_s(m)) J(s, m). \end{aligned}$$

The third term is never positive, since $J(s, m)$ vanishes whenever $r''_s(m) - r'_s(m)$ is negative. The second term is never positive, since the assumed inequality on the marginal costs makes $F'_{\rho'_s}(m) \leq F''_{\rho'_s}(m)$ for all $m \in \mathcal{M}$ and the concavity condition arranges for $\eta_s \leq \rho'_s$ to imply $F''_{\eta_s}(m) - F''_{\rho'_s}(m) \geq 0$ for all $s \geq 0$ and $m \in \mathcal{M}$. In contrast, the first term is never negative and the factor $J(s, m)$ is redundant, by the same concavity argument applied to ρ''_s .

The Lipschitz condition on F'' bounds the first term by the quantity

$$\sigma \|\rho''_s - \eta_s\|_\infty r''_s(m) = \sigma \|x_s(m)J(s, m)\|_\infty r''_s(m).$$

By assumption, the contribution of the starting state x_0 is negative, so we conclude for all $m \in \mathcal{M}$ that

$$x_t(m)J(t, m) \leq \int_0^t \|x_s(m)J(s, m)\|_\infty r_s''(m) ds.$$

For $t \in [0, T]$, the Radon-Nikodym derivative $r_s''(m)$ is bounded by $r_0''(m) + q_\nu(m)T$, so there is a new constant K' such that the essential supremum of the left hand side, namely $\beta(t) := \|x_s(m)J(s, m)\|_\infty$, satisfies

$$\beta_t \leq K' \int_0^t \beta_s ds.$$

Gronwall's Inequality (Appendix A.1) then forces $\beta_t \equiv 0$ for $t \in [0, T]$ and so for all $t \geq 0$. \square

REMARK 3.16. The proof of Lemma 3.15 does not depend on the finiteness of $\nu(\mathcal{M})$ and the lemma remains valid in the broader setting developed in Section 2.6.

3.6. Concave selective costs: Existence and stability of equilibria

If the conditions of Theorem 3.12 hold, trajectories starting from 0 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 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 3.9 of providing an explicit value of $\nu(\mathcal{M})$ that is small enough to guarantee the existence of finite equilibria.

COROLLARY 3.17. (a) *Suppose that the selective cost $S : \mathcal{G} \rightarrow \mathbb{R}$ satisfies the conditions of Theorem 3.12 and also satisfies the bound*

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

for all $m' \in \mathcal{M}$ for some constant $\xi > 0$ and function $\tau : \mathcal{M} \rightarrow \mathbb{R}_+$ 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 $t \rightarrow \infty$.

(b) *Conversely, if there exist some constant ξ and function τ such that the reverse inequality*

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

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

PROOF. Consider part (a). Let ρ'' be the solution of (2.10) 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 3.14 with $S' = S$ and $\rho'_0 = 0$ to conclude that $\rho_t \leq \rho_{**}$ for all $t \geq 0$. It follows from Theorem 3.12 that $\rho_t \uparrow \rho_*$ as $t \rightarrow \infty$ for some $\rho_* \in \mathcal{H}^+$ with $\rho_* \leq \rho_{**}$.

Now consider part (b). We define ρ'' as before, with selective cost S'' , but with initial condition $\rho''_0 = 0$. Lemma 3.14 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 $R := \lim_{t \rightarrow \infty} \rho_t(\mathcal{M}) < \infty$. 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 in \mathcal{H}^+ and that ρ''_t converges to ρ''_* in the Wasserstein metric (that is, in the topology of weak convergence). From (2.10) we know that for any Borel set A , and any $s < t$,

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

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

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

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. \square

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_* , that is, the fixed point to which the dynamical system 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 3.18. (a) *Suppose that the selective cost function satisfies the conditions of Theorem 3.12, and the curve $(\rho_t)_{t \geq 0}$, started at $\rho_0 = 0$, converges to a finite fixed point ρ_* . Suppose further that*

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

Then, the fixed point is box stable.

(b) *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 (a). Define an intensity measure ψ_s which is offset above the equilibrium ρ_* by a constant multiple of the mutation measure ν . That is, $\psi_s := \rho_* + s\nu$. Let $\phi_s = \nu - F_{\psi_s} \cdot \psi_s$ be the driving vector field evaluated at the point ψ_s . We express the derivative of ϕ_s with respect to s in terms of the quantity whose infimum is bounded below in the definition of C_{ρ_*} . This quantity would be denoted by $T_p[\mathbf{1}] \cdot \nu$ in the notation of (3.12) with $\rho_* = p \cdot \nu$, but p does not necessarily belong to $C_b(\mathcal{M}, \mathbb{R}_+)$.

The derivative of ϕ_s with respect to s is given by

$$\begin{aligned} -\frac{d\phi_s}{ds} &= (D_{\rho_*+s\nu}F)[\nu](\rho_* + s\nu) + F_{\rho_*+s\nu} \cdot \nu \\ &= [(D_{\rho_*+s\nu}F)[\nu] - (D_{\rho_*}F)[\nu]](p + ms) \cdot \nu \\ &\quad + (D_{\rho_*}F)[\nu]\rho_* + (D_{\rho_*}F)[\nu]s\nu \\ &\quad + (F_{\rho_*+s\nu} - F_{\rho_*}) \cdot \nu \\ &\quad + (F_{\rho_*}) \cdot \nu. \end{aligned}$$

Lemma 3.8 guarantees that

$$|[(D_{\rho_*+s\nu}F)[\nu] - (D_{\rho_*}F)[\nu]]| \leq 16\sigma s\nu(\mathcal{M})^2,$$

where σ is the Lipschitz constant for the selective cost S . By the same lemma,

$$|D_{\rho_*}F[\nu]s| \leq 2\sigma s\nu(\mathcal{M}).$$

By Lemma 2.6,

$$|F_{\rho_*+s\nu} - F_{\rho_*}| \leq 8\sigma s\nu(\mathcal{M}).$$

The remaining two terms are the terms bounded below by the assumption that $C_{\rho_*} > 0$. We have shown that the derivative of ϕ_s with respect to s has a Radon-Nikodym derivative with respect to ν bounded below for all $m \in \mathcal{M}$ by

$$C_{\rho_*} - s(10\sigma\nu(\mathcal{M}) + 16\sigma\nu(\mathcal{M})^2(\|p\|_\infty + s)).$$

For small enough positive s this lower bound is positive. For such s , starting from $\phi_0 = 0$,

$$\phi_s = \int_0^s \frac{d\phi_u}{du} du = - \int_0^s (D_{\psi_u}F[\nu] + F_{\psi_u} \cdot \nu) du < 0.$$

Applying Theorem 3.12, we see that when $\rho_0 = \psi_s$, the trajectory ρ_t is monotonically decreasing. Thus, for all times $t \geq 0$ we have $\rho_t \in B(\rho_*, \psi_s)$, where $B(\rho_*, \psi_s)$ is the box defined in (3.2).

Now consider part (b). 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 and belongs to \mathcal{H}^+ . 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 metric 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 $\pi^{(s_i)} \rightarrow \eta$. 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

3.7. Iterative computation of the minimal equilibrium

The measure $\pi \in \mathcal{H}^+$ is an equilibrium for the dynamic equation (2.10) if it is true that $F_\pi \cdot \pi = \nu$. Recall that $F_\pi \cdot \pi$ is the measure that has Radon-Nikodym derivative F_π with respect to π . The equation implies that π is absolutely continuous with respect to ν with a Radon-Nikodym derivative p that satisfies

$$(3.18) \quad F_{p\nu}(m)p(m) = 1, \quad m \in \mathcal{M}.$$

Under the conditions of Theorem 3.12, if (3.18) has a solution, then it has a minimal solution that arises as the Radon-Nikodym derivative with respect to ν of $\lim_{t \rightarrow \infty} \rho_t$ when ρ is the solution of (2.10) 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 (2.10). An obvious approach to that problem is to define a sequence of functions $p_n : \mathcal{M} \rightarrow \mathbb{R}_+$ inductively by $p_0 = 0$ and

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

By Lemma 3.11, it is clear that if S is concave, then $F_{p'\nu} \geq F_{p''\nu}$ for $p' \leq p''$. Because $p_0 = 0 \leq p_1$, it follows that $p_0 \leq p_1 \leq p_2 \leq \dots$. Moreover, if p_{**} is any solution of (3.18) then p_{**} has to be a fixed point of the map $p \mapsto 1/F_{p\nu}$, so $p_n \leq p_{**}$ for all n . If there is a solution p_{**} such that $\int_{\mathcal{M}} p_{**}(m) \nu(dm) < \infty$, then $p_n \uparrow p_* \leq p_{**}$ as $n \rightarrow \infty$ for some function $p_* : \mathcal{M} \rightarrow \mathbb{R}_+$ such that $p_*\nu$ assigns finite mass to \mathcal{M} . Thus, if a limit exists in \mathcal{H}^+ for ρ_t started from $\rho_0 = 0$ as time goes to infinity, it is greater than or equal to the iterative limit $p_*\nu$. But such a limit over time is also less than or equal to any other equilibrium by Lemma 3.14, the Comparison Lemma. So, starting from zero, when the iterative limit and the limit over time exist, they are the same limit, the so-called “minimal equilibrium”.

REMARK 3.19. If $\pi \in \mathcal{H}^+$ is an equilibrium for (2.10) 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.

3.8. 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 ρ_{**} converge asymptotically to an equilibrium that is also dominated by ρ_{**} . 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 3.20. *Suppose the selective cost function satisfies the Lipschitz condition, the concavity condition, and the other conditions of Theorem 3.12. Suppose also that $\inf_{m \in \mathcal{M}} S(\delta_m) > 0$. Then the following hold.*

- (a) *For $U > 0$ sufficiently small, there is a unique $p : [0, U] \rightarrow C_b(\mathcal{M}, \mathbb{R}_+)$ solving the equation*

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

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.

- (b) *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 3.7, 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 $0 \leq u < v$.

The concavity of S implies via Lemma 3.11 that $p_n \leq q_n$ for all n and

$$\begin{aligned} q_{n+1} - p_{n+1} &= \frac{v}{F_{q_n}} - \frac{u}{F_{p_n}} \\ &= \frac{v-u}{F_{q_n}} + \frac{u(F_{p_n} - F_{q_n})}{F_{p_n}F_{q_n}}. \end{aligned}$$

By Lemma 2.6 the factor $F_{p_n} - F_{q_n}$ in the second term on the right is no greater than $8\sigma \|q_n - p_n\|_{\infty} \nu(\mathcal{M})$.

We now need upper bounds on the reciprocals of F_p and F_q . Thanks to the lower bound on $S(\delta_m)$, the conditions of Theorem 3.9 are satisfied. That theorem provides a curve of equilibria for u in $[0, U]$ for some $U > 0$. To avoid confusion with the minimal equilibria that we are constructing here, denote these equilibria from Theorem 3.9 as $x^{(u)}(m)\nu(dm)$. By Lemma 3.14, $x^{(u)} \leq x^{(U)}$, and because $x^{(u)}$ is a fixed point of the map $p \rightarrow u/F_{p\nu}$ and $0 \leq x^{(u)}$, we have $p_n \leq x^{(u)} \leq x^{(U)}$ for all n . Similarly $q_n \leq x^{(v)} \leq x^{(U)}$ for $0 < u \leq v \leq U$. It follows from (2.13) that

$$(3.19) \quad \frac{1}{F_{p_n}(m)} \leq \frac{1}{F_{q_n}(m)} \leq \frac{\exp\left\{\int_{\mathcal{M}} x^{(U)}(m') \nu(m')\right\}}{\inf_{m' \in \mathcal{M}} S(\delta_{m'})}.$$

For fixed U , write a for the right-hand side of the inequality (3.19), which is constant over values of $m \in \mathcal{M}$, and write $b = 8\sigma\nu(\mathcal{M})a^2u$. We now have

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

By choosing U' in $[0, U]$ small enough and requiring $0 < u \leq v < U'$, we can make b less than 1. Iterating while bearing in mind that $q_0 - p_0 = 0$, we find

$\|q_{n+1} - p_{n+1}\|_\infty \leq (a/(1-b))|v-u|$ for all n . Hence, with $c = a/(1-b)$, the corresponding minimal equilibria, say p_* and q_* , satisfy

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

We now drop the stars subscripting equilibria and write $u \mapsto p^{(u)}$ for the curve of Radon-Nikodym derivatives of minimal equilibria whose dependence on u we have been investigating. We have just verified the Lipschitz condition

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

for all $m \in \mathcal{M}$. The condition guarantees 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 $u < U'$, we see by differentiating with respect to u that $p^{(u)}$ satisfies the differential equation in the statement of the theorem. By standard uniqueness results for ordinary differential equations, it is the unique solution with $p^{(0)} = 0$. It therefore agrees with the solution provided by Theorem 3.9, here called $x^{(u)}$ but there (and henceforth) called $p^{(u)}$, defined for all u in $[0, U]$.

It remains to verify box stability and box attractivity. Fix $u \in [0, U]$. 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 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

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

Therefore, $p^{(U)} > p^{(u)}$, and we can conclude from Theorem 3.12 that the measure ρ_t stays bounded between the measures $p^{(u)} \cdot \nu$ and $p^{(U)} \cdot \nu$ for all times $t \geq 0$, and hence the minimal equilibrium $p^{(u)} \cdot \nu$ is box stable.

Now consider part (b). 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 3.18. \square

3.9. Equilibria for demographic selective costs

As an example of Corollary 3.17, suppose that S is the demographic selective cost of Section 1.4, so that

$$(3.20) \quad \begin{aligned} &S(g + \delta_m) - S(g) \\ &= \int_0^\infty \left(1 - e^{-\theta(m,x)}\right) f_x \exp\left(-\lambda x - \int_{\mathcal{M}} \theta(m',x) dg(m')\right) dx. \end{aligned}$$

Here, as before, θ specifies an increment to the cumulative hazard function. This selective cost is concave, allowing us to capitalize on the results of Sections 3.5 and 3.8 so long as the θ are such that the required Lipschitz condition on the selective cost is satisfied, a requirement that is messy to formulate in general but easy to verify for most specific applications.

LEMMA 3.21. *Suppose that the selective cost function S is the demographic cost of (3.20). Suppose also the age-specific profiles θ satisfies the following two conditions.*

- The supremum $\tau := \sup_{m \in \mathcal{M}, x \in \mathbb{R}_+} \theta(m, x)$ is finite.
- The infimum $\inf_{m \in \mathcal{M}, x \in B} \theta(m, x) > 0$ is strictly positive for some Borel set $B \subseteq \mathbb{R}_+$ such that the integral $\int_B f_x dx$ is strictly positive.

Then,

$$\inf_{m \in \mathcal{M}} S(\delta_m) > 0$$

and

$$S(g + \delta_m) - S(g) \geq e^{-\tau g(\mathcal{M})} \inf_{m \in \mathcal{M}} S(\delta_m)$$

for all $g \in \mathcal{G}$.

PROOF. For all $m \in \mathcal{M}$,

$$\begin{aligned} S(\delta_m) &= F_0(m) = \int (1 - e^{-\theta(m,x)}) f_x e^{-\lambda x} dx \\ &\geq \int_B \left(1 - \exp\left\{-\inf_{m \in \mathcal{M}, y \in B} \theta(m, y)\right\}\right) f_x e^{-\lambda x} dx > 0. \end{aligned}$$

We also have

$$\begin{aligned} S(g + \delta_m) - S(g) &\geq \int_0^\infty \left(1 - e^{-\theta(m,x)} \right) f_x e^{-\lambda x} \exp\left\{-\sup_{m'' \in \mathcal{M}, x'' \in \mathbb{R}_+} \theta(m'', x'') g(\mathcal{M})\right\} dx \\ &= e^{-\tau g(\mathcal{M})} \int_0^\infty \left(1 - e^{-\theta(m,x)} \right) f_x e^{-\lambda x} dx \\ &\geq e^{-\tau g(\mathcal{M})} \inf_{m \in \mathcal{M}} S(\delta_m). \end{aligned}$$

□

It follows from the first implication in Corollary 3.17 that the dynamical system started from the null state with these costs converges to a finite equilibrium ρ_* in \mathcal{H}^+ . 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 $\nu(\mathcal{M}) \leq e^{-1}\xi$.

The converse implication in Corollary 3.17 cannot be readily applied to this setting, at least not with a constant function τ . However, in some cases we can give a direct proof that no equilibrium exists with finite total mass.

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. Similar models were introduced and studied by Brian Charlesworth in [Cha01]. For a suitable constant K' ,

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

The bound on the right-hand side is independent of g , so the expected value of the marginal cost has the same bound. Any equilibrium ρ_* must satisfy $\rho_* = \nu/F_{\rho_*}$. Since

$$\int_\alpha^\beta \frac{1}{F_{\rho_*}(m)} dm \geq \int_\alpha^\beta \frac{1}{\beta - m} dm = \infty,$$

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 3.12, the measure $\pi = p \cdot \nu$ could still arise as the increasing limit of the solution ρ_t to the dynamic equation (2.10) and be such that the population average selective cost $\mathbb{E}[S(X^\pi)]$ would be finite. These more delicate possibilities are treated in [WES08].

We now investigate the box stability and box attractivity of the minimal equilibrium.

THEOREM 3.22. *Suppose the selective cost function S has the demographic form specified in (3.20), (or in (1.11) and (1.12)) with $\rho_0 = 0$ and $\int f_x dx < \infty$. Suppose the age-specific effect profiles θ are such that the selective cost satisfies the Lipschitz condition of Theorem 2.4 and also satisfies the following two conditions.*

- *The supremum $\tau := \sup_{m \in \mathcal{M}, x \in \mathbb{R}_+} \theta(m, x)$ is finite.*
- *The infimum $\inf_{m \in \mathcal{M}, x \in B} \theta(m, x)$ is strictly positive for some Borel set $B \subseteq \mathbb{R}_+$ such that the integral $\int_B f_x dx$ is strictly positive.*

Set $\xi := \inf_{m \in \mathcal{M}} S(\delta_m)/(1 - e^{-\tau})$. Then, when $\nu(\mathcal{M}) < (\xi/e) \wedge (1/e)$ the measures ρ_t converge monotonically as $t \rightarrow \infty$ to an equilibrium ρ_ which is box stable. If \mathcal{M} is also compact, this equilibrium is box attractive.*

PROOF. For a measure $\rho \in \mathcal{H}^+$, we write $A_\rho(x)$ for the aggregate population net maternity function, that is,

$$\begin{aligned} A_\rho(x) &= \mathbb{E}[f_x \ell_x(X^\rho)] \\ &= f_x e^{-\lambda x} \exp\left(-\int_{\mathcal{M}} (1 - e^{-\theta(m,x)}) d\rho(m)\right). \end{aligned}$$

The expected marginal cost F_ρ is given by

$$F_\rho(m') = \int (1 - e^{-\theta(m',x)}) A_\rho(x) dx,$$

and the kernel \tilde{K} for computing its Fréchet derivative defined in Notation 3.5 turns out to be

$$\tilde{K}(m', m'') = - \int (1 - e^{-\theta(m',x)})(1 - e^{-\theta(m'',x)}) A_\rho(x) dx.$$

The existence of a finite minimal equilibrium $\rho_* \in \mathcal{H}^+$ when $\nu(\mathcal{M}) < \xi/e$ is guaranteed, as we have seen, by Corollary 3.17. For sufficiently small values of $\nu(\mathcal{M})$, ρ_* is the same as the equilibrium guaranteed by Theorem 3.9 because we can bound $S(\delta_m)$ from below thanks to Lemma 3.21.

We derive an upper bound on ρ_* . By hypothesis, $\nu(\mathcal{M})$ is strictly less than the smaller of ξ/e and $1/e$. From the Comparison Lemma 3.14 and the discussion in Section 3.2 (substituting ν/ξ for ν), we know that ρ_* is bounded above by an equilibrium solution with multiplicative costs. Specifically,

$$(1 - e^{-\tau})\rho_* \leq e^c \cdot \nu$$

where c is the unique solution in $[0, 1]$ of the equation $ce^{-c} = z$ for $z = \nu(\mathcal{M})/\xi$. As we have seen in Section 1.2, this solution exists when $\nu(\mathcal{M}) \leq \xi/e$ and can be expressed in terms of Lambert's W-function, dating back to 1758. Moreover, the map $z \mapsto c$ is convex, lying below the line $z \mapsto ez$, so our hypothesis $\nu(\mathcal{M}) < 1/e$

yields $c\xi < 1$, implying $e^c \nu(\mathcal{M}) \leq 1 - \epsilon$ for some $\epsilon > 0$. Since ρ_* has a bounded Radon-Nikodym derivative, call it r_* , with respect to ν , the function $F_{\rho_*}(m)$ is bounded below, away from zero.

To establish the claims about stability and attractivity, we appeal to Theorem 3.18. Consider any measure η in \mathcal{H}^+ which is absolutely continuous with respect to ν with Radon-Nikodym derivative $q_\eta(m)$, supplying a direction for a perturbation of ρ_* . As before, we put $\psi_s = \rho_* + s\eta$ and $\phi_s = \nu - F_{\psi_s} \cdot \psi_s$. The derivative at $s = 0$ of $-\phi_s$ with respect to s is the measure $T_{r_*}(q_\eta) \cdot \nu$ in the notation of (3.12). Expressing the Fréchet derivative of F_{ρ_*} in terms of the kernel \tilde{K} , this measure is given by

$$\begin{aligned} & T_{r_*}(q_\eta) \cdot \nu(dm') \\ &= -\rho_*(dm') \int_{\mathcal{M}} \left[\int (1 - e^{-\theta(m',x)})(1 - e^{-\theta(m'',x)}) A_{\rho_*}(x) dx \right] \eta(dm'') \\ &+ \eta(dm') \int (1 - e^{-\theta(m',x)}) A_{\rho_*}(x) dx. \end{aligned}$$

Examine the nonnegative function of m' and x multiplying $-\rho_*(dm')$ in the first term on the right-hand side. Because $1 - \exp\{-\theta(m'',x)\} \leq 1 - \exp\{-\tau\}$ for all m'' and x , this multiplier is bounded above by $(1 - \exp\{-\tau\})\eta(\mathcal{M})F_{\rho_*}(m')$. The nonnegative function multiplying $\eta(dm')$ is $F_{\rho_*}(m')$. Hence,

$$(3.21) \quad T_{r_*}(q_\eta) \cdot \nu \geq (F_{\rho_*}) (\eta - \eta(\mathcal{M})(1 - e^{-\tau})\rho_*).$$

Taking advantage of the bound on ρ_* , in terms of Radon-Nikodym derivatives,

$$(3.22) \quad T_{r_*}(q_\eta) \geq (F_{\rho_*}) (q_\eta - (1 - \epsilon)\eta(\mathcal{M})/\nu(\mathcal{M})).$$

We now use Theorem 3.18. For box stability we put $\eta = \nu$, so that $q_\eta \equiv 1$, and, as required, $\inf_{m \in \mathcal{M}} T_{r_*}(\mathbf{1})(m)$ is bounded below by $\inf_{m \in \mathcal{M}} F_{\rho_*}(m)\epsilon > 0$.

For box attractivity, when \mathcal{M} is compact, we let η be any probability measure in \mathcal{H}^* . As required, the left-hand side of (3.22) cannot be almost-surely zero, since the integral over \mathcal{M} of the right-hand side with respect to ν is bounded below by $\inf_{m \in \mathcal{M}} F_{\rho_*}(m)\epsilon$. \square

With demographic costs, assuming \mathcal{M} to be compact is not an onerous restriction. Taking advantage of the boundedness of demographic costs, \mathcal{M} can often be embedded in a compact space without disrupting the Lipschitz condition on S , and Theorem 3.22 then comes into force.

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: The continuous-time dynamical system (1.8), which we defined formally in (2.10), is the limit of a sequence of discrete-time 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 a limit of the sort of discrete generation, infinite population mutation-selection-recombination models considered by Barton and Turelli [BT91] and Kirkpatrick, Johnson, and Barton [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

This section describes the operations of mutation and selection in the discrete-time setting, leaving the description of recombination to Section 4.2.

We consider a sequence of models indexed by the positive integers. In the n^{th} model, time is sped up by a factor of n , so that n generations pass in 1 unit of time. Unlike familiar diffusion approximations for genetic processes, all models in the sequence are infinite-population models; n does not have any reference to population size. The distribution of genotypes in the population is described, as before, by a probability measure on \mathcal{G} , the space of finite integer-valued measures on the set \mathcal{M} of loci.

At each generation each new birth gets a random set of extra mutations away from the ancestral wild type added to those mutant alleles it inherited from its parents. In the n^{th} model in our sequence, 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 $e^{-S(g)/n}$. The n 's in the denominators are what, in effect, speed up time. Mutation and selective cost per unit time are maintained. 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 the four principal operators that act on the space of probabilities on \mathcal{G} . The operators \mathfrak{M}_n , \mathfrak{S}_n , \mathfrak{R} , and \mathfrak{P} describe the transformation in genotype distribution by, respectively, one round of mutation, one round of selection, one

round of recombination, and a process of complete Poissonization in which a genotype with possibly linked loci is replaced by one in which there is no linkage. The \mathfrak{M}_n , \mathfrak{S}_n , and \mathfrak{P} operators are defined in this section. The \mathfrak{R} operator, requiring further structure, is defined in the following section. The intuition behind these definitions is described in the Introduction. In this chapter we use the notation F without a subscript to stand for a bounded Borel test function. There is no connection with the subscripted function F_π defined in Definition 2.3. We also use the letter Q to stand for a generic probability measure, so that P remains available for Poisson probability measures like the measures that describe the continuous-time dynamical system.

NOTATION 4.1. Given a probability measure Q on \mathcal{G} , define new probability measures $\mathfrak{M}_n Q$, $\mathfrak{S}_n Q$, and $\mathfrak{P}Q$ by their action on a bounded Borel test function $F : \mathcal{G} \rightarrow \mathbb{R}_+$ as follows

$$(4.1) \quad \mathfrak{M}_n Q[F] := \int_{\mathcal{G}} \mathbb{E}[F(g + X^{\nu/n})] dQ(g)$$

$$(4.2) \quad \mathfrak{S}_n Q[F] := \frac{\int_{\mathcal{G}} \exp\{-S(g)/n\} F(g) dQ(g)}{\int_{\mathcal{G}} \exp\{-S(g)/n\} dQ(g)}$$

$\mathfrak{R}Q[F]$ is defined in Notation 4.2

$$(4.3) \quad \mathfrak{P}Q[F] := \Pi_{\mu Q}[F].$$

We now show that as n becomes large, our discrete-time operators, considered separately, come to agree with their continuous-time counterparts. 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 $\nu(\cdot)/\nu(\mathcal{M})$. In other words, $X^{\nu/n}$ takes the value δ_m with probability approximately $\nu(dm)$ and takes the value zero otherwise. For any probability measure Q on \mathcal{G}

$$(4.4) \quad \lim_{n \rightarrow \infty} n(\mathfrak{M}_n Q[F] - Q[F]) = \int_{\mathcal{G}} \left(\int_{\mathcal{M}} (F(g + \delta_m) - F(g)) d\nu(m) \right) dQ(g).$$

In particular, if F is of the form $F(g) := \exp\{-g[f]\}$ for some Borel function $f : \mathcal{M} \rightarrow \mathbb{R}_+$, then, by the Poisson identity (1.13),

$$(4.5) \quad \mathfrak{M}_n Q[F] = Q[F] \cdot \exp\{\nu[e^{-f} - 1]/n\}$$

and so

$$(4.6) \quad \lim_{n \rightarrow \infty} n(\mathfrak{M}_n Q[F] - Q[F]) = \nu[e^{-f} - 1]Q[F].$$

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

$$(4.7) \quad \lim_{n \rightarrow \infty} n(\mathfrak{S}_n Q[F] - Q[F]) = -Q[S \cdot F] + Q[S]Q[F].$$

When we start with a population genotype distribution Q and assume that selection precedes mutation, the population genotype distribution after one generation of mutation and selection is $\mathfrak{M}_n \mathfrak{S}_n Q$. A trajectory of the resulting discrete-time model is defined by iteration. That is, given an initial population genotype distribution Q_0 , the population genotype distribution after k generations is $(\mathfrak{M}_n \mathfrak{S}_n)^k Q_0$.

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

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

for test functions $F : \mathcal{G} \rightarrow \mathbb{R}_+$ of the form $F(g) = \exp\{-g[f]\}$ for some Borel function $f : \mathcal{M} \rightarrow \mathbb{R}_+$. For this special choice of test function, this is precisely the equation defining dynamical system without recombination that we introduced in [SEW05] and derived heuristically in (1.5). In fact, these test functions are enough to consider, since they determine probability measures on \mathcal{G} . Formally, a proof that the discrete-time dynamical system converges to this continuous-time one in the absence of recombination would require that we prove the existence of the continuously differentiable limit, a fact that we assumed above. We forgo such a proof, since we wish to incorporate recombination.

4.2. Recombination in discrete-time

We now introduce recombination. Imitating [BT91], we think of a recombination event as taking two genotypes $g^{(1)}, g^{(2)} \in \mathcal{G}$ from the population and replacing the genotype $g^{(1)}$ in the population by the genotype g defined by $g(A) := g^{(1)}(A \cap R) + g^{(2)}(A \cap R^c)$, 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 $g^{(1)}$ (respectively, $g^{(2)}$) for “loci” in the set R (respectively, R^c).

As described in the Introduction, we are working 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 might reasonably be defined on a linear sequence of loci, as in [BT91], with ends of chromosomes at pre-determined positions, and with recombination caused by crossovers during meiosis. Our abstract space of loci accommodates such a concrete representation but also leaves us freedom to specify loci using structures related to their demographic effects rather than their chromosomal positions.

Of course, we think of $g^{(1)}$ and $g^{(2)}$ as being chosen independently at random according to the particular probability measure describing the distribution of genotypes in the population. We 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 A is a random Borel set and κ is a finite measure, then $\kappa(A)$ is a real-valued random variable. We suppose that there is a probability measure \mathcal{R} that describes the distribution of the random set

of loci that segregate together. We always assume, without loss of generality, that \mathcal{R} is *symmetric* in the sense that

$$(4.8) \quad \mathcal{R}(A) = \mathcal{R}(\{R^{\complement} : R \in A\}),$$

where R^{\complement} 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 $A \subseteq \mathcal{M}$. Given the (symmetric) probability measure \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

$$(4.9) \quad \mathfrak{R}Q[F] := \int_{\mathcal{B}(\mathcal{M})} \int_{\mathcal{G}} \int_{\mathcal{G}} F(g_R^{(1)} + g_{R^{\complement}}^{(2)}) dQ(g^{(1)}) dQ(g^{(2)}) d\mathcal{R}(R),$$

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

Thus, if Q describes the distribution of genotypes in the population, then $\mathfrak{R}Q$ describes the distribution of a genotype that is obtained by picking two genotypes $g^{(1)}$ and $g^{(2)}$ independently according to Q , picking a segregating set R according to the probability measure \mathcal{R} , and forming a composite genotype that agrees with $g^{(1)}$ on the set R and $g^{(2)}$ on the set R^{\complement} . Where confusion with subscripts might result, the customary notation $g|_R$ is also used for the restriction g_R of g to the set R .

Recall from Notation 2.1 that μQ denotes the intensity measure of Q ; that is,

$$(4.10) \quad \mu Q(A) := \int_{\mathcal{G}} g(A) dQ(g).$$

REMARK 4.3. Note that $\mu \mathfrak{R}Q = \mu Q$; that is, the intensity measure of the random measure describing the genotype of a randomly chosen individual is left unchanged by recombination. Specifically, for Borel sets B ,

$$\begin{aligned} \mu \mathfrak{R}Q(B) &= \int_{\mathcal{G}} \int_{\mathcal{G}} (g^{(1)}(R \cap B) + g^{(2)}(R^{\complement} \cap B)) dQ(g^{(1)}) dQ(g^{(2)}) d\mathcal{R}(R) \\ &= \int (\mu Q(R \cap B) + \mu Q(R^{\complement} \cap B)) d\mathcal{R}(R) = \mu Q(B). \end{aligned}$$

Moreover, if P is the distribution of a Poisson random measure, then $\mathfrak{R}P = P$.

We are now in a position to put the operators for recombination, mutation, and selection together to describe the whole discrete-time dynamical system.

NOTATION 4.4.

- Q_0 is the distribution of any simple, \mathcal{G} -valued random measure and describes the distribution of genotypes in the initial population;
- $P_0 = \mathfrak{P}Q_0 = \Pi_{\mu Q_0} = \Pi_{\rho_0}$ is the Poisson probability measure with the same intensity as Q_0 ;
- $Q_k := (\mathfrak{M}_n \mathfrak{S}_n)^k Q_0$ and describes the distribution of genotypes in the population after k generations of selection, mutation, and recombination;
- $P_k := (\mathfrak{M}_n)^k P_0$ is the Poisson probability measure on \mathcal{G} that results from k generations of mutation in the absence of selection starting from P_0 ;

- $O_k := (\mathfrak{P}\mathfrak{M}_n\mathfrak{S}_n)^k P_0$ is the Poisson probability measure resulting from k iterations of mutation and selection with intensity rescaled by $1/n$ and complete Poissonization (rather than recombination) after every round.

4.3. Recombination trees and annealed recombination

In order to establish the convergence of our discrete-generation dynamical system to our continuous-time dynamical system, we need to understand the effects of mutation, selection and recombination not only over a single generation, as we have done in Sections 4.1 and 4.2, but also over a number of successive generations. Over k generations, each of the operators \mathfrak{M}_n , \mathfrak{S}_n , and \mathfrak{R} acts k times. However, these operators are interleaved: there is first an \mathfrak{S}_n , then an \mathfrak{M}_n , then an \mathfrak{R} , then another \mathfrak{S}_n , then another \mathfrak{M}_n , and so on, and the operators in this product do not all commute with each other. For our convergence proof, we want to be able to decompose the impact of each kind of operator over k generations into its impact on the mutant alleles that first appear in a particular generation. Separating out the k -fold effects of each operator requires further structure to keep track of the generation in which a given mutant allele first appears. In this section we define structure for k -fold iterates of the recombination operator in terms of a version of a genealogical tree. In the next sections we define structure for separating out effects of \mathfrak{M}_n and of \mathfrak{S}_n over k generations by introducing a concept of “vintage”, tracking the vintages of mutant alleles, and defining associated “starred” versions of our basic operators that are defined on genotypes that have been decomposed according to the allele vintages.

The iterates of our recombination operator implicitly create a genealogical tree. The genotype of an individual sampled from the population at generation k is formed from the genotypes of two parents sampled from the population, which are formed from the genotypes of four grandparents, and so on back to 2^k individuals sampled from the initial population at generation zero. Taken together, these $2^{k+1} - 1$ individuals comprise the nodes of a complete rooted binary tree. The root stands for the individual sampled at generation k and the leaves stand for the earliest ancestors back in the initial population at generation 0. Each node except the leaves has two edges (branches) directed in from parent nodes and each node except the root has one branch directed out to the offspring node.

For each choice of the time-scaling factor n , we have a finite number of generations $k_{\max} = \lfloor nT \rfloor$ up to time T . The tree of ancestors for an individual in generation k just introduced may be envisioned as a subtree of the tree built from a root node at generation k_{\max} . Call this latter tree \mathcal{L} .

The tree structure defines a partial ordering among nodes in the usual fashion. We label nodes in any convenient way, and write $v \prec \ell$ if node v is an ancestor of node ℓ , that is, if there exists a path from v to the root without duplicate edges containing ℓ . Nodes prior in the ordering are prior in time. We count generations starting with the leaves at generation zero. Since the letter “ g ” is already in use for genotypes, we write $V(\ell)$ for the generation (“vintage”) of node ℓ .

The recombination structure is implemented by assigning complementary segregating sets B and $B^{\mathfrak{G}}$ to the two branches connecting a node to its parents: the genotype of the corresponding individual in the set B (respectively, $B^{\mathfrak{G}}$) coincides with that of the first (respectively, second) parent in the set B (respectively, $B^{\mathfrak{G}}$).

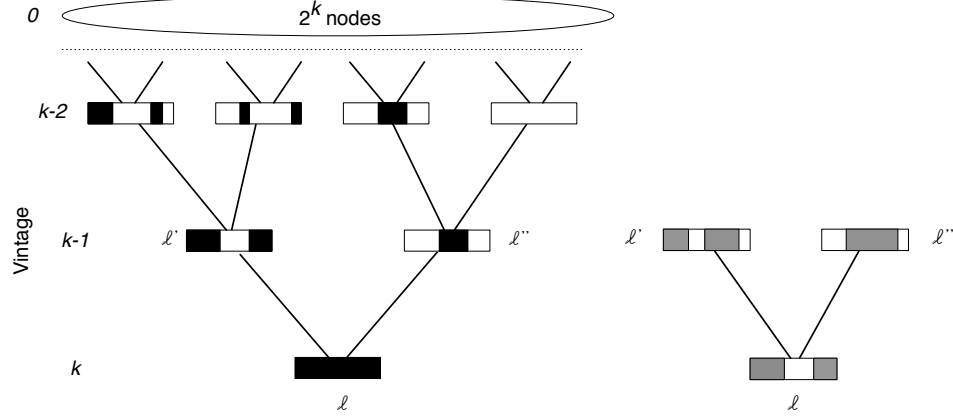


FIGURE 4.1. The ancestor tree for an individual of vintage k . For purposes of illustration, the mutation space \mathcal{M} is the unit interval. The first picture shows in black the sets $W(v, \ell)$ at the node v , for v in the first two ancestor generations preceding ℓ , so vintages k , $k-1$, and $k-2$. The second picture shows in gray the recombination set R_ℓ (vintage k) and $R_{\ell'}$ and $R_{\ell''}$ (vintage $k-1$) that generated these partitions. Note that $W(\ell, \ell) = \mathcal{M}$. Note too that some sets in the partition will be empty, meaning that all genetic information from that ancestor has been lost through recombination.

More specifically, we first assign each node ℓ an independent set R_ℓ with distribution \mathcal{R} . (Sets for the leaves go unused.) Suppose that the parents of node ℓ are ℓ' and ℓ'' . We label the branch connecting ℓ' (respectively, ℓ'') to ℓ with the label ℓ' (respectively, ℓ''). The complementary segregating sets assigned to the branches labeled ℓ' and ℓ'' are, respectively, $B_{\ell'} = R_\ell$ and $B_{\ell''} = R_\ell^c$. There is some ambiguity in this construction due to the choice of which branch receives R_ℓ or its complement, and this can be resolved by embedding the complete rooted binary tree in the plane so that we can distinguish between the “left” and “right” ancestors of a node. Probabilistically, however, this ambiguity is immaterial thanks to the symmetry of the recombination measure – any way of resolving it will lead to an assignment of segregating sets to branches that has the same distribution.

NOTATION 4.5. The genetic legacy W for any pair of nodes satisfying $v \preceq \ell$ is the subset of \mathcal{M} given by

$$W(v, \ell) := \bigcap \{B_\alpha : v \preceq \alpha \prec \ell\}.$$

The intersection of the empty set is taken to be \mathcal{M} . For other pairs v, ℓ such that $v \not\preceq \ell$, $W(v, \ell) := \emptyset$. See Figure 4.1.

When ℓ' and ℓ'' are ancestors (not necessarily parents) of a node ℓ and belong to the same generation, the sets $W(\ell', \ell)$ and $W(\ell'', \ell)$ are disjoint subsets of \mathcal{M} , since the branches leading to the first common descendant pass through sets R and R^c that are complements of each other. The sets $W(v, \ell)$ for fixed ℓ over all v satisfying $v \prec \ell$ and $V(v) = j$ for some generation j cover \mathcal{M} and so form a partition of \mathcal{M} (We adopt the convention that a partition may contain empty sets.)

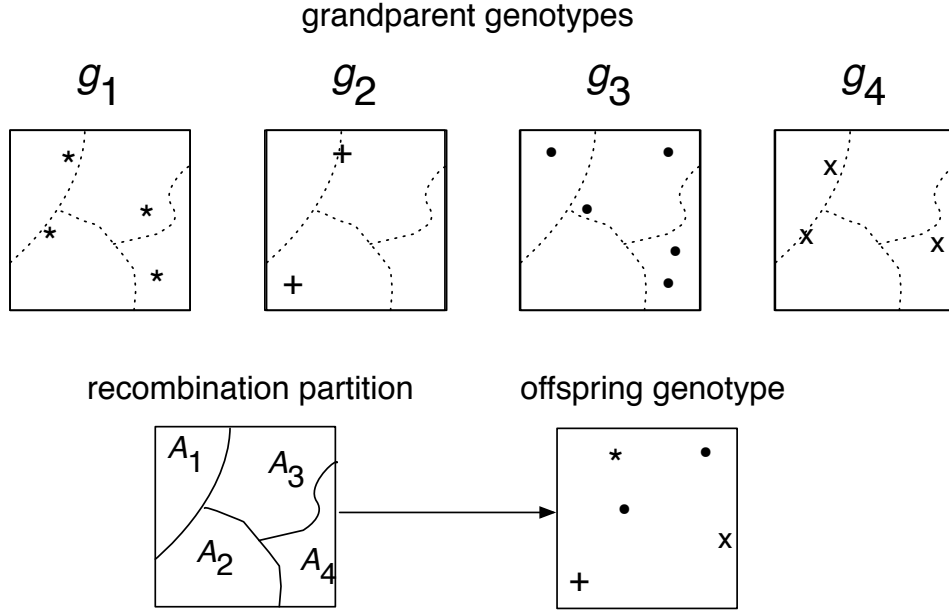


FIGURE 4.2. Illustration of the operation on genotypes induced by a partition \mathcal{A} that underlies the action of the annealed recombination $\mathfrak{R}_{\mathcal{A}}$ operator on a collection of probability measures on \mathcal{G} . Here the mutation space \mathcal{M} is the unit square, and the partition \mathcal{A} is the collection of four sets $\{A_1, A_2, A_3, A_4\}$ shown in the lower left. The mutations of four grandparent genotypes are shown at the top, and these genotypes (shown with a different symbol for each grandparent to clarify the origins of each one) are filtered through the partition to produce the offspring genotype at lower right.

We are now in a position to specify the combined action of k applications of the recombination operator.

NOTATION 4.6. For any finite partition $\mathcal{A} = \{A_1, \dots, A_L\}$ of \mathcal{M} into Borel sets, some possibly empty, the *annealed recombination operator* $\mathfrak{R}_{\mathcal{A}}$ is the operator acting on probability measures Q in \mathcal{G} defined by the condition

$$(4.11) \quad \mathfrak{R}_{\mathcal{A}}Q[F] := \int \cdots \int F \left(g^{(1)}|_{A_1} + \cdots + g^{(L)}|_{A_L} \right) dQ(g^{(1)}) \cdots dQ(g^{(L)})$$

for bounded Borel functions $F : \mathcal{G} \rightarrow \mathbb{R}$. Here, $g|_A$ is the restriction of the integer-valued measure g to the subset A , also written g_A as in Notation 4.2 in contexts where no confusion with indexing subscripts would result. See Figure 4.2.

In other words, if Y_1, \dots, Y_L are i.i.d. random measures with common distribution Q , then $\mathfrak{R}_{\mathcal{A}}Q$ is the distribution of the random measure $Y_1|_{A_1} + \cdots + Y_L|_{A_L}$.

Via annealing, the genetic legacies from ancestors at generation zero generate the k -th iterate of the recombination operator as follows:

LEMMA 4.7. *Let $\ell \in \mathcal{L}$ be a node with $V(\ell) = k$. Let $\mathcal{A} = (A_1 \dots A_{2^k})$ be the random partition of \mathcal{M} consisting of the 2^k sets $\{W(\alpha, \ell) : \alpha \prec \ell \text{ and } V(\alpha) = 0\}$, constructed from the random segregating sets $\{R_v : v \preceq \ell\}$. Then, for any probability measure Q on \mathcal{G} and bounded Borel function $F : \mathcal{G} \rightarrow \mathbb{R}$,*

$$\mathfrak{R}^k Q[F] = \mathbb{E} [\mathfrak{R}_{\mathcal{A}} Q[F]].$$

PROOF. The claim holds trivially for $k = 0$. We proceed by induction. Suppose the claim holds up to generation $k - 1$ and thus for the parent nodes ℓ' and ℓ'' of node ℓ . For all $\omega \prec \ell$, either $\omega \preceq \ell'$ or $\omega \preceq \ell''$, and $W(\omega, \ell) = W(\omega, \ell') \cap R_{\ell}$ or $W(\omega, \ell) = W(\omega, \ell'') \cap R_{\ell}^c$, respectively. For the genotypes of the parent nodes, write

$$\Gamma'((g_v)_{v \preceq \ell', V(v)=0}) = \sum_{\substack{\omega \preceq \ell', \\ V(\omega)=0}} (g_{\omega})|_{W(\omega, \ell')}$$

and

$$\Gamma''((g_v)_{v \preceq \ell'', V(v)=0}) = \sum_{\substack{\omega \preceq \ell'', \\ V(\omega)=0}} (g_{\omega})|_{W(\omega, \ell'')}.$$

Then, by the induction hypothesis,

$$\mathfrak{R}^{k-1} Q[F] = \int \mathbb{E} [F(\Gamma'(\mathbf{g}))] d \bigotimes_{\substack{v \preceq \ell', \\ V(v)=0}} Q_v(\mathbf{g}),$$

where each Q_v is a copy of Q , and the analogous equation holds for Γ'' and ℓ'' . Applying the operator \mathfrak{R} to the expressions for \mathfrak{R}^{k-1} gives

$$\mathfrak{R}^k Q[F] = \iint \mathbb{E} \left[F(\Gamma'(\mathbf{g}'|R_{\ell}) + \Gamma''(\mathbf{g}''|R_{\ell}^c)) \right] d \bigotimes_{\substack{v \preceq \ell', \\ V(v)=0}} Q_v(\mathbf{g}') d \bigotimes_{\substack{v \preceq \ell'', \\ V(v)=0}} Q_v(\mathbf{g}'').$$

Substituting for Γ' and Γ'' shows this last expression is equal to the claimed formula. \square

We use the tree structure of \mathcal{L} to construct a nested sequence of random partitions $\mathcal{A}_0(\ell) \supseteq \mathcal{A}_1(\ell) \supseteq \dots \supseteq \mathcal{A}_{V(\ell)}(\ell)$ of \mathcal{M} for an arbitrary node $\ell \in \mathcal{L}$. There is one set in the partition $\mathcal{A}_j(\ell)$ for each ancestor of the individual ℓ in generation j . This set represents the portion of that ancestor's genotype inherited by the individual ℓ through the shuffling together of genotypes due to recombination.

NOTATION 4.8. For $\ell \in \mathcal{L}$ and $k \leq V(\ell)$, define a random partition $\mathcal{A}_k(\ell)$ of \mathcal{M} by

$$\begin{aligned} \mathcal{A}_k(\ell) &:= \{W(v, \ell) : V(v) = k \text{ with } v \prec \ell\} \\ &= \left\{ \bigcap \{B_{\alpha} : v \preceq \alpha \prec \ell\} : V(v) = k \text{ with } v \prec \ell \right\}. \end{aligned}$$

REMARK 4.9. Note that $\mathcal{A}_k(\ell)$ is a refinement of $\mathcal{A}_{k+1}(\ell)$, and $\mathcal{A}_{V(\ell)}(\ell)$ is the trivial partition $\{\mathcal{M}\}$. Note also that for $j \geq k$ the distribution of the sequence of partitions $(\mathcal{A}_0(\ell), \mathcal{A}_1(\ell), \dots, \mathcal{A}_k(\ell))$ is the same for all nodes ℓ with $V(\ell) = j$. Recall that R_{ℓ} is the recombination set at node ℓ . Suppose that ℓ has parents ℓ'

and ℓ'' and that the sets R_ℓ and R_ℓ^c are associated with the branches labeled ℓ' and ℓ'' . Then, for $k \leq V(\ell) - 1$,

$$\mathcal{A}_k(\ell) = \{A' \cap R_\ell : A' \in \mathcal{A}_k(\ell')\} \cup \{A'' \cap R_\ell^c : A'' \in \mathcal{A}_k(\ell'')\}.$$

NOTATION 4.10. Fix a node $\ell \in \mathcal{L}$ with $V(\ell) = k$. Write $(\mathcal{A}_{k,0}, \mathcal{A}_{k,1}, \dots, \mathcal{A}_{k,k})$ for the nested sequence of random partitions $(\mathcal{A}_0(\ell), \mathcal{A}_1(\ell), \dots, \mathcal{A}_k(\ell))$.

REMARK 4.11. The distribution of $(\mathcal{A}_{k,0}, \mathcal{A}_{k,1}, \dots, \mathcal{A}_{k,k})$ does not depend on the choice of the node ℓ with vintage k . Note also that for $0 \leq j \leq k$ the sequence $(\mathcal{A}_{k-j,0}, \mathcal{A}_{k-j,1}, \dots, \mathcal{A}_{k-j,k-j})$ has the same distribution as the sequence $(\mathcal{A}_{k,j}, \mathcal{A}_{k,j+1}, \dots, \mathcal{A}_{k,k})$. We can re-state the conclusion of Lemma 4.7 as

$$\mathfrak{R}^k Q[F] = \mathbb{E} [\mathfrak{R}_{\mathcal{A}_{k,0}} Q[F]].$$

4.4. Vintages

The annealed recombination operator allows us to condition on choices of segregating sets while we isolate k -generational effects of mutation and selection. We next consider how the process of mutation relates to the recombination tree. Any element of \mathcal{M} assigned mass by the genotype for a node at generation k has to have entered at some earlier generation, either in one of the nodes for ancestors in the initial population (distributed according to Q_0) or as a new mutation at one of the intermediate nodes in the tree of ancestors (distributed as a Poisson random measure with intensity measure ν/n). In our infinite-population setting, selection reweights the probabilities of outcomes. It does not remove elements of \mathcal{M} or alter their pattern of descent from nodes of the tree. Although Q_0 may not be a Poisson probability measure, one of our assumptions will require it to be absolutely continuous with respect to the Poisson probability measure P_0 with the same intensity measure. It turns out that the difference between Q_0 and P_0 can be treated as a process of reweighting, analogous to the reweighting resulting from selection. With this strategy in mind, we use Poisson random measures to build a joint distribution of random genotypes across the tree, a template to which reweighting is applied.

Specifically, we assign to each node v in \mathcal{L} an independent Poisson random measure Γ_v with distribution $P_0 = \Pi_{\rho_0}$ when $V(v) = 0$ and with distribution $\Pi_{\nu/n}$ otherwise. Given a choice of the segregating sets R_ℓ , which define the genetic legacies W at each node, consider restrictions of the Poisson measures to the legacy sets and put

$$(4.12) \quad \Xi_\ell = \sum_{v \preceq \ell} \Gamma_v |_{W(v,\ell)}.$$

These random elements of \mathcal{G} are independent across nodes within any one generation but are dependent across nodes down any line of descent. They correspond to the genotypes of the ancestors of a randomly chosen individual in generation k_{\max} in the absence of selection starting from a population with genotype distribution P_0 . With this representation, which underlies the crucial Lemma 7.5 at the start of Chapter 7, we may separate out the effects of mutation over k generations from the effects of the other operators.

Any mutant allele found in the genotype of a reference individual sampled from the population at generation k originates in the Γ_v term for some ancestral node. However, knowing the genotype at generation k and knowing the segregating sets for the tree of ancestral nodes is not enough to identify the origin of each

mutation. A mutant allele found in the subset of \mathcal{M} inherited from a particular great-grandparent might have originated from that great-grandparent or it might come from a new mutation in the same portion of the parental genome. What is missing is information on the generation, or vintage, from which the mutation originates. Because we need to be able to condition on the identities of the mutations that are still extant in the genome at generation k , while we average over all the other contributions to the Γ_v , we introduce additional structure to keep track of vintages.

We can rewrite (4.12) as

$$(4.13) \quad \Xi_\ell = \sum_{j \leq V(\ell)} \sum_{\substack{v \leq \ell, \\ V(v)=j}} \Gamma_v |_{W(v,\ell)}.$$

The term

$$\sum_{\substack{v \leq \ell, \\ V(v)=j}} \Gamma_v |_{W(v,\ell)}$$

separates out the contribution to Ξ_ℓ due to mutant alleles that first appeared in generation j in the lineage of the individual associated with node ℓ . This suggests that we imagine that the genotype g of an individual in generation k is decomposed as a sum $g = g_0 + \dots + g_k$, where g_j , $j < k-1$, is the contribution to g that first appeared in an ancestor in the past generation j and g_k is the contribution of new mutations that first appeared in the individual herself. We thus have a richer structure where a genotype $g \in \mathcal{G}$ is replaced by the $(k+1)$ -tuple $\mathbf{g} = (g_0, g_1, \dots, g_{k-1}, g_k)$ in the product space \mathcal{G}^{k+1} that decomposes g according to vintage. In most contexts there is no danger of confusing the subscripts for vintage components of \mathbf{g} with subscripts for Borel sets to which genotypes g may be restricted.

The $k+1$ -tuple giving the decomposition according to vintage of an individual drawn at random from the population has a joint distribution which is a probability measure on \mathcal{G}^{k+1} . This object describes the vintages of the current stock of mutations retained in the genomes of individuals in generation k . The force of selection does not depend on the decomposition by vintages – it only matters what mutations are present in an individual’s genotype, not when they entered her lineage. But because selection alters the weights of whole genotypes in the population it necessarily alters the overall distribution of the decomposition by vintages. Recombination shuffles contributions within vintages. Mutation adds a new component to the decomposition by vintages at each generation.

In order to have a single space on which all our vintage-specific distributions are defined, we form the disjoint union \mathcal{G}^* of the spaces of tuples for successive generations starting with the initial population for $k = 0$ on up. That is,

$$(4.14) \quad \mathcal{G}^* := \mathcal{G}^1 \sqcup \mathcal{G}^2 \sqcup \mathcal{G}^3 \sqcup \dots,$$

where \sqcup denotes disjoint union. In principle, a probability measure on this space can assign positive probability to tuples in several different generations and can thus be thought of as a sum of sub-probability measures on the different disjoint pieces. In practice, the probability measures that we consider are always concentrated on the tuples for some particular generation. We can always return from our big tuple space \mathcal{G}^* to \mathcal{G} itself by adding up the components associated with each vintage.

The operators for mutation, selection, and recombination \mathfrak{M}_n , \mathfrak{S}_n , and \mathfrak{R} , which we have defined to act on probability measures on \mathcal{G} , all have counterparts acting on the probability measures on our big tuple space \mathcal{G}^* . Just as \mathfrak{M}_n , \mathfrak{S}_n , and \mathfrak{R} combine to take the distribution of genotypes in one generation to the distribution in the next, these new operators combine to take the distribution of genotypes decomposed by vintage in each generation into the distribution in the next. We denote these new operators by adding a star to the notation for the corresponding operators acting on probability measures on \mathcal{G} .

In order to define these new operators, we first need to introduce a little more notation. We usually represent a generic element of \mathcal{G}^* by the boldface \mathbf{g} . There are natural “projection” operations $\Sigma : \mathcal{G}^* \rightarrow \mathcal{G}$ defined for $\mathbf{g} = (g_0, g_1, \dots, g_k) \in \mathcal{G}^{k+1}$ by

$$\Sigma(\mathbf{g}) := g_0 + g_1 + \dots + g_k$$

and

$$\Psi_j(\mathbf{g}) := \begin{cases} g_j, & 0 \leq j \leq k, \\ 0, & j > k. \end{cases}$$

In essence, Σ removes the labels that record the vintages of the various mutations from the ancestral wild type, while Ψ_j isolates mutations with vintage j . Note that, as operators, $\Sigma = \sum_{j=0}^{\infty} \Psi_j$.

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 $F : \mathcal{G} \rightarrow \mathbb{R}_+$.

Although we shall only be employing probability measures on \mathcal{G}^* that are in fact concentrated on some \mathcal{G}^k , we give a brief account of the structure of a general probability measure on the disjoint union \mathcal{G}^* . As we remarked above, a probability measure Q on \mathcal{G}^* may be thought of as a sequence of sub-probability measures $(Q^{(k)})_{k=0}^{\infty}$, where $Q^{(k)}$ is the portion of Q concentrated on $(k+1)$ -tuples from \mathcal{G} . Thus, $\sum_{i=0}^{\infty} Q^{(i)}(\mathcal{G}^{i+1}) = 1$. We interpret such a Q as the distribution of a finite, random length sequence of random measures (X_0, X_1, \dots, X_I) , with $Q^{(k)}(\cdot; \cdot)/Q^{(k)}(\mathcal{G}^{k+1})$ being the joint distribution of (X_0, X_1, \dots, X_k) conditional on the event $\{I = k\}$, and $Q^{(k)}(\mathcal{G}^{k+1})$ being the probability of the event $\{I = k\}$.

NOTATION 4.12. Consider a probability measure Q on \mathcal{G}^* that is concentrated on \mathcal{G}^{k+1} for some $k \in \mathbb{N}$. Write $(X_0 \dots X_k)$ for a $k+1$ -tuple of random measures with conditional distribution $Q^{(k)}$.

- Define a new probability measure $\mathfrak{M}_n^* Q$ on \mathcal{G}^* to be the distribution of $(X_0 \dots X_k, Y)$, where Y is an independent Poisson random measure with intensity measure ν/n .
- Define a new probability measure $\mathfrak{S}_n^* Q$ on \mathcal{G}^* that is concentrated on \mathcal{G}^{k+1} by

$$\begin{aligned} (\mathfrak{S}_n^* Q)^{(k)}[F] &:= \frac{\int_{\mathcal{G}^k} \exp\{-S(\Sigma(g))/n\} F(g) dQ^{(k)}(g)}{\int_{\mathcal{G}^k} \exp\{-S(\Sigma(g))/n\} dQ^{(k)}(g)} \\ &= \frac{\mathbb{E}[\exp\{-S(X_0 + \dots + X_k)/n\} F(X_0, \dots, X_k)]}{\mathbb{E}[\exp\{-S(X_0 + \dots + X_k)/n\}]} \end{aligned}$$

for any bounded Borel function $F : \mathcal{G}^* \rightarrow \mathbb{R}_+$.

- Define a new probability measure \mathfrak{R}^*Q on \mathcal{G}^* that is concentrated on \mathcal{G}^{k+1} by

$$\int F(\mathbf{g}) d\mathfrak{R}^*Q(\mathbf{g}) := \int \int F(\mathbf{g}^{(1)}|A + \mathbf{g}^{(2)}|A^{\mathfrak{L}}) dQ(\mathbf{g}^{(1)}) dQ(\mathbf{g}^{(2)}) d\mathcal{R}(A)$$

for any bounded Borel function $F : \mathcal{G}^* \rightarrow \mathbb{R}_+$.

- Define a new probability measure \mathfrak{P}^*Q on \mathcal{G}^* that is concentrated on \mathcal{G}^{k+1} by letting $\mathfrak{P}^*Q^{(k)}$ be the distribution of (Y_0, \dots, Y_k) , where the random measure Y_j is Poisson with the same intensity measure as X_j and Y_0, \dots, Y_k are independent.

Observe that we have the four intertwining relations

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

These equalities confirm that starred operators agree with the unstarred ones once the labeling by vintage is removed.

Our initial conditions P_0 and Q_0 can also be regarded as probability measures on \mathcal{G}^* that are concentrated on sequences of length 1.

NOTATION 4.13. Set

$$(4.15) \quad \begin{aligned} Q_k^* &:= (\mathfrak{R}^* \mathfrak{M}_n^* \mathfrak{S}_n^*)^k Q_0 \\ P_k^* &:= (\mathfrak{M}_n^*)^k P_0 \\ O_k^* &:= (\mathfrak{P}^* \mathfrak{S}_n^* \mathfrak{M}_n^*)^k P_0. \end{aligned}$$

The operator O_k^* is the analogue of Q_k^* with the complete Poissonization operator \mathfrak{P}^* replacing the recombination operator \mathfrak{R}^* .

Note that all three probability measures O_k^* , Q_k^* and P_k^* are concentrated on genotype sequences of length $(k+1)$. Note also that if (X_0, \dots, X_k) 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, $Q_k = \Sigma Q_k^*$, $O_k = \Sigma O_k^*$, and $P_k = \Sigma P_k^*$, where we recall that $Q_k = (\mathfrak{R} \mathfrak{M}_n \mathfrak{S}_n)^k Q_0$, $O_k = (\mathfrak{P} \mathfrak{M}_n \mathfrak{S}_n)^k Q_0$, and $P_k := (\mathfrak{M}_n)^k P_0$.

Shattering and the formulation of the convergence result

5.1. Shattering of random measures

We expect recombination to break up dependencies between different parts of the genome, so that when k is large $\mathfrak{R}^k P$ should be approximately $\Pi_{\mu P}$, the distribution of the Poisson random measure with intensity μP in the notation established in Notation 2.1.

In order that this approximation should hold for a given probability measure P , it must generically be the case that there is a positive probability that the segregating set and its complement both intersect any set with positive μP mass in two sets that each have positive μP mass. The following condition (with $\lambda = \mu P$) is key to establishing quantitative bounds on the rate with which $\mathfrak{R}^k P$ converges to $\Pi_{\mu P}$.

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

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

We show in Section 5.3 that if the pair $(\mathcal{R}, \mu Q)$ is shattering and P satisfies a further simple condition, then $\mathfrak{R}^k Q$ converges to $\Pi_{\mu Q}$ as $k \rightarrow \infty$.

The concept of shattering is illustrated by the following elementary example.

REMARK 5.2. Suppose that N is a finite simple random measure 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 $0 < T_1 < \dots < T_L < 1$ for the successive points of \mathcal{M} . Set $T_0 = 0$ and $T_{L+1} = 1$. 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

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

Take λ to be any diffuse probability measure (that is, λ has no atoms). Suppose that there is a constant c such that $\mathbb{P}\{N(u, w] = 1\} \geq c\lambda((u, w])$ for $0 < u < w \leq 1$. This is the case, for example, if N is Poisson random measure 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 random measures with intensity $C\lambda$ for some constant C , it is 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.

For the recombination measure \mathcal{R} in this example we have for the right-hand side of (5.1)

$$\begin{aligned}
& \lambda(A)^2 - 2 \int \lambda(A \cap R)^2 \mathcal{R}(dR) \\
&= 2 \left[\int_0^1 \int_0^w \mathbf{1}_A(u) \mathbf{1}_A(w) \lambda(du) \lambda(dw) \right. \\
&\quad \left. - 2 \int_0^1 \int_0^w \mathbf{1}_A(u) \mathbf{1}_A(w) \mathbb{P}\{N((u, w]) = 0 \pmod{2}\} \lambda(du) \lambda(dw) \right] \\
&= 2 \int_0^1 \int_0^w \mathbf{1}_A(u) \mathbf{1}_A(w) \mathbb{P}\{N((u, w]) = 1 \pmod{2}\} \lambda(du) \lambda(dw) \\
&\geq 2 \int_0^1 \int_0^w \mathbf{1}_A(u) \mathbf{1}_A(w) \mathbb{P}\{N((u, w]) = 1\} \lambda(du) \lambda(dw) \\
&\geq 2c \int_0^1 \int_0^w \mathbf{1}_A(u) \mathbf{1}_A(w) \lambda((u, w]) \lambda(du) \lambda(dw).
\end{aligned}$$

As for the left-hand side of (5.1) observe that

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

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

REMARK 5.3. The meaning of the shattering condition may also be clarified by the following observation. Suppose that \mathcal{M} is equipped with a metric δ , that λ is a probability measure, and for some constant $c > 0$

$$\begin{aligned}
p(r) &:= \inf\{\mathcal{R}\{R : m' \in R, m'' \in R^c\} : m', m'' \in \mathcal{M}, \delta(m', m'') \geq r\} \\
&\geq c \sup\{\lambda\{m'' \in \mathcal{M} : \delta(m', m'') \leq r\} : m' \in \mathcal{M}\} =: c\varphi(r)
\end{aligned}$$

for all $r \in \mathbb{R}_+$. Loosely speaking, this condition says that the probability that two loci inherit their contents from different parents dominates a multiple of the λ mass

of a ball with radius the distance between the two loci. By a change of variables,

$$\begin{aligned}
\int \lambda(A \cap R) \lambda(A \cap R^{\complement}) \mathcal{R}(dR) &= \int_A \int_A \mathcal{R}\{R : m' \in R, m'' \in R^{\complement}\} \lambda(dm'') \lambda(dm') \\
&\geq \int_A \int_A p(\delta(m', m'')) \lambda(dm'') \lambda(dm') \\
&\geq c \int_A \int_{\mathcal{M}} \mathbf{1}_{\{\delta(m', m'') \leq \varphi^{-1}(\nu(A))\}} \varphi(\delta(m', m'')) \lambda(dm'') \lambda(dm') \\
&= c \lambda(A) \frac{1}{2} \lambda(A)^2,
\end{aligned}$$

Thus, (\mathcal{R}, λ) is shattering with constant $\alpha = c^{-1}$.

We collect some elementary observations about the shattering condition in the following lemma.

LEMMA 5.4. *Suppose that the pair (\mathcal{R}, λ) is shattering with some constant α .*

- (a) *If λ is non-zero, then $\alpha \geq 2\lambda(\mathcal{M})$.*
- (b) *The measure λ is necessarily diffuse.*
- (c) *If the pair (\mathcal{R}, η) is also shattering for the same recombination measure and same shattering constant α , then the pair $(\mathcal{R}, \lambda + \eta)$ is shattering with constant 4α .*
- (d) *For any $t \geq 0$ the pair $(\mathcal{R}, t\lambda)$ is shattering with constant $t\alpha$.*
- (e) *If η is another measure on \mathcal{M} with $\bar{H} := \sup |\log d\eta/d\lambda| < \infty$, then (\mathcal{R}, η) is shattering with constant $\alpha e^{5\bar{H}}$.*

PROOF. (a) Suppose that $\lambda(\mathcal{M}) > 0$. Then, from (5.1),

$$\lambda(\mathcal{M}) \leq 2\alpha \int \frac{\lambda(R)}{\lambda(\mathcal{M})} \frac{\lambda(R^{\complement})}{\lambda(\mathcal{M})} d\mathcal{R}(R) \leq 2\alpha \frac{1}{4}.$$

(b) It is clear from (5.1) that $\lambda(\{m\}) = 0$ for all $m \in \mathcal{M}$.

(c) By Jensen's Inequality, $(\lambda(A)/2 + \eta(A)/2)^3 \leq \lambda(A)^3/2 + \eta(A)^3/2$, and the sum of two terms $\lambda(A \cap R)\lambda(A \cap R^{\complement}) + \eta(A \cap R)\eta(A \cap R^{\complement})$ is less than or equal to the sum of four terms resulting from the product $(\lambda + \eta)(A \cap R)(\lambda + \eta)(A \cap R^{\complement})$.

(d) For any Borel set A ,

$$(t\lambda)(A)^3 = t^3 \lambda(A)^3 \leq 2t\alpha \int (t\lambda)(A \cap R)(t\lambda)(A \cap R^{\complement}) d\mathcal{R}(R).$$

(e) For any Borel set A ,

$$\begin{aligned}
\eta(A)^3 &\leq e^{3\bar{H}} \lambda(A)^3 \\
&\leq e^{3\bar{H}} \cdot 2\alpha \int \lambda(A \cap R) \lambda(A \cap R^{\complement}) d\mathcal{R}(R) \\
&\leq 2\alpha e^{5\bar{H}} \cdot \int \eta(A \cap R) \eta(A \cap R^{\complement}) d\mathcal{R}(R).
\end{aligned}$$

□

REMARK 5.5. Most pairs (\mathcal{R}, λ) that arise naturally in applications are shattering, but there do exist pairs for which the recombination measure \mathcal{R} splits every set with positive λ -measure into sets with positive λ -measure and the pair nonetheless fails to be shattering. For example, let \mathcal{M} be the positive real axis. Let λ have

a density with respect to Lebesgue measure equal to 2^{-k} on $[k, k+1)$. Consider random partitions of the axis into two parts $[0, x)$ and $[x, \infty)$ (taken for the sake of symmetry in random order), where the recombination measure chooses x according to a probability density equal to $2^{-k}(1/k)(x-k)^{-1+1/k}$ on $[k, k+1)$. This density assigns the same measure to each whole interval $[k, k+1)$ as λ , namely 2^{-k} .

For each k , consider the set $A_k = [k, k+1)$. Put $y = x - k$. If the pair was shattering, there would be some constant α such that for all k

$$(5.2) \quad \frac{1}{2\alpha} \leq \frac{1}{\lambda(A_k)} \int \frac{\lambda(A \cap R)}{\lambda(A_k)} \frac{\lambda(A_k \cap R^c)}{\lambda(A_k)} \mathcal{R}(dR).$$

With probability $1 - \lambda(A_k)$, the partition does not cut the interval A_k , that is, x falls outside the interval, and the integrand on the right-hand side vanishes. The rescaled measure $\mathcal{R}(dR)/\lambda(A_k)$ restricted to A_k is a probability measure, and (5.2) can be rewritten

$$\frac{1}{2\alpha} \leq \int_0^1 y(1-y)(1/k)y^{-1+1/k} dy = \frac{1}{k+1} - \frac{1}{2k+1} \leq \frac{1}{k+1}.$$

It follows that there is no finite shattering constant α which satisfies the required condition for all the sets A_k , establishing that this pair (\mathcal{R}, λ) is not shattering.

5.2. Consequences of shattering

We are now in a position to state and prove a pair of results that quantify the extent to which $\mathfrak{R}^k Q$ for some probability measure Q on \mathcal{G} is a mosaic of many small pieces, each taken from genomes sampled independently from the population described by Q , where we measure the size of a subset of \mathcal{M} using its μQ mass.

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. We apply it to the random nested partitions defined in Notation 4.8.

NOTATION 5.6. Suppose that $\mathcal{A} = \{A_1, \dots, A_L\}$ is a partition of \mathcal{M} , λ is a measure on \mathcal{M} , and $r > 0$. Set

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

LEMMA 5.7. *Suppose that the pair (\mathcal{R}, λ) is shattering with constant α . Then, for all $k \in \mathbb{N}_0$,*

$$\mathbb{E} \left[|\mathcal{A}_{k,0}|_2^{(\lambda)} \right] < \frac{\alpha \lambda(\mathcal{M})}{k+1}.$$

PROOF. Recall that $\mathcal{A}_{k+1,0}$ is the random partition

$$\mathcal{A}_0(\ell) := \{W(v, \ell) : V(v) = 0 \text{ with } v \prec \ell\}$$

for some $\ell \in \mathcal{L}$ with $V(\ell) = k+1$ (the choice of ℓ is arbitrary insofar as the distribution of $\mathcal{A}_{k+1,0}$ is concerned). By construction, a set of the form $W(v, \ell)$ with $V(v) = 0$ and $v \prec \ell$ can be written as either $W(\gamma, \ell) \cap R_\gamma$ or $W(\gamma, \ell) \cap R_\gamma^c$, where $V(\gamma) = 1$ and $\gamma \prec \ell$, and so

$$|\mathcal{A}_{k+1,0}|_2^{(\lambda)} = \sum_{V(\gamma)=1, \gamma \prec \ell} \left[\lambda(W(\gamma, \ell) \cap R_\gamma)^2 + \lambda(W(\gamma, \ell) \cap R_\gamma^c)^2 \right].$$

Set $X_j := |\mathcal{A}_{j,0}|_2^{(\lambda)}$ for $j \in \mathbb{N}_0$ and recall the notation

$$\mathcal{A}_1(\ell) := \{W(\gamma, \ell) : V(\gamma) = 1 \text{ with } \gamma \prec \ell\}.$$

From the above and the symmetry of \mathcal{R} ,

$$\begin{aligned} \mathbb{E}[X_{k+1} \mid \mathcal{A}_1(\ell)] &= \sum_{A \in \mathcal{A}_1(\ell)} 2 \int \lambda(A \cap R)^2 d\mathcal{R}(R) \\ &\leq \sum_{A \in \mathcal{A}_1(\ell)} \lambda(A)^2 (1 - (1/\alpha)\lambda(A)) \\ &= |\mathcal{A}_1(\ell)|_2^{(\lambda)} - (1/\alpha)|\mathcal{A}_1(\ell)|_3^{(\lambda)}. \end{aligned}$$

For any partition $\mathcal{A} = \{A_1, \dots, A_L\}$ of \mathcal{M} , it follows from the Cauchy-Schwarz inequality that

$$\begin{aligned} |\mathcal{A}|_2^{(\lambda)} &= \sum_{i=1}^L \lambda(A_i)^2 \\ &= \sum_{i=1}^L \lambda(A_i)^{\frac{1}{2}} \lambda(A_i)^{\frac{3}{2}} \\ &\leq \left(\sum_{i=1}^L \lambda(A_i)^{\frac{2}{3}} \right)^{\frac{1}{2}} \left(\sum_{i=1}^L \lambda(A_i)^{\frac{6}{3}} \right)^{\frac{1}{2}} \\ &= \left(|\mathcal{A}|_1^{(\lambda)} \right)^{\frac{1}{2}} \left(|\mathcal{A}|_3^{(\lambda)} \right)^{\frac{1}{2}} \\ &= (\lambda(\mathcal{M}))^{\frac{1}{2}} \left(|\mathcal{A}|_3^{(\lambda)} \right)^{\frac{1}{2}}. \end{aligned}$$

Thus,

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

and so

$$\mathbb{E}[X_{k+1} \mid \mathcal{A}_1(\ell)] \leq \tilde{X}_k (1 - \tilde{X}_k/c),$$

where $c = \alpha\lambda(\mathcal{M})$ and $\tilde{X}_k = |\mathcal{A}_1(\ell)|_2^{(\lambda)}$. By construction, the random partition $\mathcal{A}_1(\ell)$ has the same distribution as $\mathcal{A}_{k,0}$. Applying Jensen's inequality to the concave function $x(1 - x/c)$, we see that

$$\begin{aligned} \mathbb{E}[X_{k+1}] &\leq \mathbb{E}[\tilde{X}_k (1 - \tilde{X}_k/c)] \\ &\leq \mathbb{E}[\tilde{X}_k] (1 - \mathbb{E}[\tilde{X}_k]/c) \\ &= \mathbb{E}[X_k] (1 - \mathbb{E}[X_k]/c), \end{aligned}$$

and so

$$\frac{c}{\mathbb{E}[X_{k+1}]} \geq \frac{c}{\mathbb{E}[X_k]} + \frac{c}{c - \mathbb{E}[X_k]} > \frac{c}{\mathbb{E}[X_k]} + 1$$

provided $\mathbb{E}[X_k] < 1$. We have $X_0 = \lambda(\mathcal{M})^2$, which is strictly smaller than $c = \alpha\lambda(\mathcal{M})$ since the shattering constant α is necessarily greater than $\lambda(\mathcal{M})$. It follows that $\mathbb{E}[X_k] < c/(1+k)$ for all $k \geq 0$. \square

5.3. Convergence to Poisson of iterated recombination

The purpose of the shattering condition is to guarantee that recombination bring the population distribution of genotypes back to that of a Poisson random measure, even in the face of the countervailing force of selection. However, before we proceed to our main result in which recombination, mutation, and selection are intermingled, it is instructive to see how recombination does its job in the absence of selection. This section shows that, under the shattering condition, 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} .

Along with the concept of annealed recombination, our result on convergence of iterated recombination in the absence of mutation and selection also requires the following condition on the initial distribution of genotypes in the population.

We cannot expect iterated recombination to send an arbitrary probability measure Q on \mathcal{G} to Poisson. In the extreme situation where the intensity measure μQ is diffuse and yet a realization of Q is not simple, recombination is unable to break up atoms of size greater than one, and so no amount of recombination will produce a result that is close to a Poisson random measure with intensity μQ . We have, of course, adopted the standing assumption that all of our probability measures on \mathcal{G} are the distributions of simple random measures, but it is clear that if Q is, in some sense, close to the distribution of a random measure that is not simple, then the convergence of $\mathfrak{R}^k Q$ to the Poisson measure $\Pi_{\mu Q}$ can at best be arbitrarily slow. The following condition quantifies how far a probability measure Q is from one with a realization that has atoms and it will enable us to provide a bound on the rate that $\mathfrak{R}^k Q$ converges to $\Pi_{\mu Q}$.

DEFINITION 5.8. A probability measure Q on \mathcal{G} is *dispersive* if there is a constant β such that for any Borel set $A \subseteq \mathcal{M}$,

$$\int_{\mathcal{G}} g(A) \mathbf{1}_{\{g(A) \geq 2\}} dQ(g) \leq \beta \mu Q(A)^2.$$

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

THEOREM 5.9. *Suppose that the pair $(\mathcal{R}, \mu Q)$ is shattering and Q is dispersive with constant β . Then, $\mathfrak{R}^k Q$ converges to the Poisson measure $\Pi_{\mu Q}$ as $k \rightarrow \infty$, with*

$$(5.3) \quad \|\mathfrak{R}^k Q - \Pi_{\mu Q}\|_{\text{Was}} \leq (6\beta + 2)(\alpha \mu Q(\mathcal{M}))(k + 1)^{-1}.$$

PROOF. Recall that for a measure $g \in \mathcal{G}$ and a Borel set $A \in \mathcal{B}(\mathcal{M})$ that $g_A \in \mathcal{G}$ is the measure $g(\cdot \cap A)$. Write $Q|_A$ for the push-forward of the probability measure Q by the map $g \mapsto g_A$. (This notation is used only within the context of this proof, and is not to be confused with the use of $|$ elsewhere to denote the restriction of a measure to a set.)

Let $\mathcal{A} = \{A_1, \dots, A_L\}$ be a partition of \mathcal{M} . Then, by definition, the push-forward of $\mathfrak{R}_{\mathcal{A}} Q$ by the map $g \mapsto (g_{A_1}, \dots, g_{A_L})$ is $Q|_{A_1} \otimes \dots \otimes Q|_{A_L}$. Similarly, the push-forward of $\Pi_{\mu Q}$ by the map $g \mapsto (g_{A_1}, \dots, g_{A_L})$ is $\Pi_{\mu Q}|_{A_1} \otimes \dots \otimes \Pi_{\mu Q}|_{A_L}$. Suppose that $F : \mathcal{G} \rightarrow \mathbb{R}$ is Lipschitz with Lipschitz norm $\|F\|_{\text{Lip}}$ at most 1. Define $\Phi : \mathcal{G}^L \rightarrow \mathbb{R}$ by $\Phi(g_1, \dots, g_L) = F(g_1 + \dots + g_L)$. Note for fixed $g_1, \dots, g_{i-1}, g_{i+1}, \dots, g_L$ that the function $g \mapsto \Phi(g_1, \dots, g_{i-1}, g, g_{i+1}, \dots, g_L)$ is Lipschitz with Lipschitz

norm at most 1 and so, writing \otimes for the product of measures,

$$\begin{aligned}
& \left| \left(\bigotimes_{i=1}^L Q|_{A_i} - \bigotimes_{i=1}^L \Pi_{\mu Q}|_{A_i} \right) [\Phi] \right| \\
& \leq \sum_{i=1}^L \left| \left(\bigotimes_{j=1}^{i-1} Q|_{A_j} \otimes \bigotimes_{k=i}^L \Pi_{\mu Q}|_{A_k} - \bigotimes_{j=1}^i Q|_{A_j} \otimes \bigotimes_{k=i+1}^L \Pi_{\mu Q}|_{A_k} \right) [\Phi] \right| \\
& = \sum_{i=1}^L \left| \left(\bigotimes_{j=1}^{i-1} Q|_{A_j} \otimes (Q|_{A_i} - \Pi_{\mu Q}|_{A_i}) \otimes \bigotimes_{k=i+1}^L \Pi_{\mu Q}|_{A_k} \right) [\Phi] \right| \\
& \leq \sum_{i=1}^L \| \mathfrak{R}_{\mathcal{A}} Q|_{A_i} - \Pi_{\mu Q}|_{A_i} \|_{\text{Was}}.
\end{aligned}$$

Therefore,

$$(5.4) \quad \| \mathfrak{R}_{\mathcal{A}} Q - \Pi_{\mu Q} \|_{\text{Was}} \leq \sum_{i=1}^L \| Q|_{A_i} - \Pi_{\mu Q}|_{A_i} \|_{\text{Was}}.$$

Given a Borel subset $A \subseteq \mathcal{M}$, define $\hat{\pi}$ to be the measure on \mathcal{M} given by

$$\hat{\pi}(B) = \int_{\mathcal{G}} g(B \cap A) \mathbf{1}_{\{g(A)=1\}} dQ(g).$$

Clearly $\hat{\pi} \leq \mu Q$. Also, define the sub-probability measure \hat{Q} on \mathcal{G} by

$$\hat{Q}[F] = \int_{\mathcal{G}} F(g|_A) \mathbf{1}_{\{g(A) \leq 1\}} dQ(g).$$

Observe that $\hat{\pi}$ is the intensity measure of \hat{Q} . Note also that

$$(5.5) \quad \mu Q(A) = \hat{\pi}(A) + \int_{\mathcal{G}} g(A) \mathbf{1}_{\{g(A) \geq 2\}} dQ(g).$$

We have

$$(5.6) \quad \begin{aligned} & \| Q|_A - \Pi_{\mu Q}|_A \|_{\text{Was}} \\ & \leq \| Q|_A - \hat{Q} \|_{\text{Was}} + \| \hat{Q} - \Pi_{\hat{\pi}} \|_{\text{Was}} + \| \Pi_{\hat{\pi}} - \Pi_{\mu Q}|_A \|_{\text{Was}}. \end{aligned}$$

We bound the first term on the right by

$$(5.7) \quad \| Q|_A - \hat{Q} \|_{\text{Was}} \leq Q\{g \in \mathcal{G} : g(A) \geq 2\} \leq \beta \mu Q(A)^2,$$

since Q is dispersive with constant β .

For a bound on the second term on the right of (5.6), for any $F : \mathcal{G} \rightarrow \mathbb{R}$ with $\|F\|_{\text{Lip}} \leq 1$, we can write $\hat{\pi}[F(\delta.)]$ for $\int F(\delta_m) d\hat{\pi}(m)$, and we have

$$\begin{aligned} |\hat{Q}[F] - \Pi_{\hat{\pi}}[F]| &\leq |Q\{g \in \mathcal{G} : g(A) = 0\} - \Pi_{\hat{\pi}}\{g \in \mathcal{G} : g(A) = 0\}| \\ &\quad + |\hat{Q}[F\mathbf{1}_{\{g \in \mathcal{G} : g(A)=1\}}] - \Pi_{\hat{\pi}}[F\mathbf{1}_{\{g \in \mathcal{G} : g(A)=1\}}]| \\ &\quad + \Pi_{\hat{\pi}}\{g \in \mathcal{G} : g(A) \geq 2\} \\ &\leq \left|1 - \hat{\pi}(A) + Q\{g \in \mathcal{G} : g(A) \geq 2\} - e^{-\hat{\pi}(A)}\right| \\ &\quad + \left|\hat{\pi}[F(\delta.)] \left(1 - e^{-\hat{\pi}(A)}\right)\right| + \frac{\hat{\pi}(A)^2}{2} \\ &\leq Q\{g \in \mathcal{G} : g(A) \geq 2\} + \frac{\mu Q(A)^2}{2} + \mu Q(A)^2 + \frac{\mu Q(A)^2}{2}. \end{aligned}$$

Thus,

$$(5.8) \quad \|\hat{Q} - \Pi_{\hat{\pi}}\|_{\text{Was}} \leq (\beta + 2)\mu Q(A)^2.$$

Finally, for our bound on the third term of (5.6), we have, by Lemma A.12,

$$\|\Pi_{\hat{\pi}} - \Pi_{\mu Q}|_A\|_{\text{Was}} \leq 4\|\hat{\pi} - \mu Q(\cdot \cap A)\|_{\text{Was}},$$

because the intensity measure of $\Pi_{\mu Q}|_A$ is $\mu Q(\cdot \cap A)$. For any $f : A \rightarrow \mathbb{R}$ with $\|f\|_{\text{Lip}} \leq 1$,

$$\begin{aligned} |\hat{\pi}[f] - \mu Q[f]| &= \left| \int_{\mathcal{G}} g[f] d\hat{Q}(g) - \int_{\mathcal{G}} g[f\mathbf{1}_A] dQ(g) \right| \\ &= \left| \int_{\mathcal{G}} g[f\mathbf{1}_A]\mathbf{1}_{\{g(A)=1\}} dQ(g) - \int_{\mathcal{G}} g[f\mathbf{1}_A] dQ(g) \right| \\ &= \left| \int_{\mathcal{G}} g[f\mathbf{1}_A]\mathbf{1}_{\{g(A) \geq 2\}} dQ(g) \right| \\ &\leq Q\{g \in \mathcal{G} : g(A) \geq 2\} \\ &\leq \beta \mu Q(A)^2. \end{aligned}$$

Thus,

$$(5.9) \quad \|\Pi_{\hat{\pi}} - \Pi_{\mu Q}|_A\|_{\text{Was}} \leq 4\beta \mu Q(A)^2.$$

Putting (5.7), (5.8) and (5.9) into (5.6), we get

$$(5.10) \quad \|Q|_A - \Pi_{\mu Q}|_A\|_{\text{Was}} \leq (6\beta + 2)\mu Q(A)^2.$$

By (5.4), then,

$$\|\mathfrak{R}_{\mathcal{A}}Q - \Pi_{\mu Q}\|_{\text{Was}} \leq (6\beta + 2) \sum_{i=1}^L \mu Q(A_i)^2.$$

It follows, via Jensen's Inequality, that

$$(5.11) \quad \|\mathfrak{R}^k Q - \Pi_{\mu Q}\|_{\text{Was}} = \|\mathbb{E}[\mathfrak{R}_{\mathcal{A}_{k,0}}Q] - \Pi_{\mu Q}\|_{\text{Was}} \leq (6\beta + 2) \mathbb{E} \left[|\mathcal{A}_{k,0}|_2^{(\mu Q)} \right].$$

Applying Lemma 5.7 completes the proof of (5.3). \square

Note that Theorem 5.9 is essentially a random measure version of Le Cam's Poisson convergence result of [LC60].

Theorem 5.9 is about the situation where the initial distribution of genotypes in the population satisfies both a shattering condition and a dispersive condition

and recombination acts alone. Our main convergence theorem in Chapter 8 posits an initial distribution of genotypes that satisfies a shattering condition along with an hypothesis on its Radon-Nikodym derivative with respect to the distribution of the Poisson measure with the same intensity measure, but no dispersive condition. When the selective cost S and the mutation rate ν are set to zero, the main convergence theorem also covers the situation in which recombination acts alone. The relationship between these two results on iterated recombination is clarified by the following observation, where we recall the definition of the quantity $\Theta(P, Q)$ from (2.7).

PROPOSITION 5.10. *Suppose that $P = \Pi_\pi$ and Q is a probability measure on \mathcal{G} such that $\mu Q = \pi$ and $\Theta(P, Q) < \infty$. Then, there exists a positive constant β for which the dispersive condition*

$$\int_{\mathcal{G}} g(A) \mathbf{1}_{\{g(A) \geq 2\}} dQ(g) \leq \beta \mu Q(A)^2$$

holds.

PROOF. Set $\hat{H} := \Theta(P, Q)$. Lemma A.13, to be proved in the appendix, shows that the Lipschitz condition implies that

$$\int_{\mathcal{G}} g(A) \mathbf{1}_{\{g(A) \geq 2\}} dQ(g) \leq \int_{\mathcal{G}} g(A) \mathbf{1}_{\{g(A) \geq 2\}} \exp\{\hat{H}(g(\mathcal{M}) + \pi(\mathcal{M}))\} dP(g).$$

Splitting \mathcal{M} up into two sets A and A^c and using the independence properties of the Poisson probability measure P , we can bound this quantity by a constant multiple of

$$\begin{aligned} & \int_{\mathcal{G}} g(A) \mathbf{1}_{\{g(A) \geq 2\}} \exp\{\hat{H}g(A)\} dP(g) \\ &= \sum_{k \geq 2} k e^{\hat{H}k} e^{-\pi(A)} \frac{\pi(A)^k}{k!} \\ &= e^{-\pi(A)} e^{\hat{H}\pi(A)} [\exp\{e^{\hat{H}}\pi(A)\} - 1] \\ &\leq e^{-\pi(A)} e^{\hat{H}\pi(A)} e^{\hat{H}\pi(A)} \exp\{e^{\hat{H}}\pi(A)\}. \end{aligned}$$

The last expression is bounded by a constant multiple of $\pi(A)^2$, so the dispersive condition

$$\int_{\mathcal{G}} g(A) \mathbf{1}_{\{g(A) \geq 2\}} dQ(g) \leq \beta \mu Q(A)^2$$

holds. □

5.4. Atoms in the initial intensity

We have discussed the shattering condition in the context of the mutation measure ν . We shall also be assuming that the initial intensity measure ρ_0 for the starting state P_0 is shattering. This hypothesis excludes initial intensity measures with atoms. The question arises as to whether atoms could be allowed. The answer is no.

The reason for excluding atoms is illustrated by an artificial case in which the mutation measure ν is zero and ρ_0 has a single atom at a point $m \in \mathcal{M}$. As time

passes m remains the only mutation present in the population, since recombination does not introduce mutations not already present. Each occurring element of \mathcal{G} consists of an integer mass at m . We identify \mathcal{G} with the nonnegative integers \mathbb{N}_0 and simply write g for $g(m)$. We can then think of S as a function from \mathbb{N}_0 to \mathbb{R}_+ . Take the initial probability measure Q_0 to be the distribution of a Poisson random variable with mean $\lambda = \rho_0(m)$. As usual, the notation $\lfloor t \rfloor$ denotes the greatest integer less than or equal to the real number t .

After $\lfloor tn \rfloor$ generations of selection, in the notation of Notation 4.4, the population is described by a measure Q which, by (4.2), equals

$$\begin{aligned} Q_{\lfloor tn \rfloor}(dg) &= \mathfrak{S}_n^{\lfloor tn \rfloor} dQ_0(g) \\ &= \frac{\exp\{-\lfloor tn \rfloor S(g)/n\} dQ_0(g)}{\int_{\mathcal{G}} \exp\{-\lfloor tn \rfloor S(h)/n\} dQ_0(h)}. \end{aligned}$$

This mass converges as $n \rightarrow \infty$ to

$$\exp\{-tS(g)\}Q_0(dg) / \int_{\mathcal{G}} \exp\{-tS(h)\}Q_0(dh),$$

which, in our case, is a probability measure on the nonnegative integers that assigns to each integer g the mass

$$\left(\exp\{-tS(g)\}e^{-\lambda}(\lambda^g / g!) \right) / \left(\sum_{k=0}^{\infty} \exp\{-tS(k)\}e^{-\lambda}(\lambda^k / k!) \right).$$

This probability measure is not Poisson except in the non-epistatic case where $S(g) = cg$ for some constant $c > 0$. In short, in the presence of an atom in the initial intensity, the operation of selection drives the distribution of genotypes in the population away from that of a Poisson random measure, and recombination is powerless to restore its Poisson character. For this reason, something akin to the shattering condition for the recombination measure and the initial intensity is essential.

5.5. Preview of the main convergence result

We can now state the hypotheses and conclusions of our main result, the result that justifies the continuous-time model (2.10) as a limit of discrete-time models in which recombination acts on a faster time scale than mutation and selection. The formal statement of the theorem and its proof are given in Chapter 8, Theorem 8.1. The proof is based on supporting results established in Chapters 6 and 7. The statement of the theorem is given here by way of preview, so as to make the purposes served by the supporting theorems and lemmas in those chapters easier to discern.

We continue to use the notation $\lfloor x \rfloor$ to denote the greatest integer less than or equal to the real number x .

Hypotheses:

Let $(\rho_t)_{t \geq 0}$ be the measure-valued dynamical system of (2.10) whose existence is guaranteed by Theorem 2.4. Suppose that the selective cost function S satisfies the hypotheses of Theorem 2.4, namely

- $S(0) = 0$,
- $S(g) \leq S(g + h)$ for all $g, h \in \mathcal{G}$,
- for some constant σ , $|S(g) - S(h)| \leq \sigma \|g - h\|_{\text{was}}$, for all $g, h \in \mathcal{G}$.

In addition, suppose that the following assumptions are in force.

- The pair (\mathcal{R}, ν) consisting of the recombination measure and the mutation measure is shattering.
- The pair (\mathcal{R}, ρ_0) consisting of the recombination measure and the initial intensity is shattering.
- The initial measure Q_0 is equivalent to its Poissonization $P_0 := \mathfrak{P}Q_0 = \Pi_{\mu Q_0} = \Pi_{\rho_0}$, and $\Theta(P_0, Q_0) < \infty$.

Conclusions:

Then, for any $T > \epsilon > 0$,

$$\lim_{n \rightarrow \infty} \sup_{\epsilon \leq t \leq T} \|\Pi_{\rho_t} - Q_{\lfloor tn \rfloor}\|_{\text{Was}} = 0.$$

If, in addition, the initial measure $Q_0 = P_0$ is Poisson, then this equation holds for $\epsilon = 0$.

REMARK 5.11. 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 the proofs 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 5.12. Because of Lemma 5.4, the hypotheses of Theorem 8.1 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* integer-valued random measure.

From now on, we assume without further comment that all the probability measures on \mathcal{G} we consider are distributions of simple integer-valued random measures.

Before beginning the formal presentation in the following chapters, we outline the strategy of the proof. The distance in the Wasserstein metric between our continuous-time dynamical system and our discrete-generation system at any point in time for any given value of the generation-scaling parameter n can be broken into three pieces:

- (1) the distance between the dynamical system and (with a suitable re-scaling of time) the discrete-generation system O_k with complete Poissonization after every generation;
- (2) the distance between O_k and the discrete-generation system with Poissonization only at the end of k steps, $\mathfrak{P}Q_k$, which we may call *end-state Poissonization*;
- (3) the distance between $\mathfrak{P}Q_k$ and the discrete-generation system Q_k itself.

Chapter 6 is devoted to showing that the first distance, the one between Π_{ρ_t} and $O_{\lfloor kt \rfloor}$, is of order $1/n$. This part of the proof is the only one in which the dynamical system comes into play, and it is also the only one in which recombination does not have a role because both systems are built (implicitly for the first and explicitly for the second) with complete Poissonization rather than a finite amount

of recombination. Here we must focus primarily on understanding the operation of selection when the selective cost is scaled by $1/n$.

Chapter 7 lays the groundwork for bounds on the second distance and the third distance. The second distance is a distance between the distributions of two Poisson random measures, which, as we know from Lemma A.12, can be bounded in terms of the distance between their intensity measures. The probability measure $\mathfrak{P}Q_k$ has the same intensity measure as Q_k , so the quantity to be bounded is the distance between μO_k and μQ_k . We show in Chapter 8 that such a bound can be derived from a bound on the third distance, that between the probability measures $\mathfrak{P}Q_k$ and Q_k .

The required bound on the third distance has to hold for every generation between zero and $\lfloor Tn \rfloor$. The approach in the proof is to pick any intermediate generation as a provisional starting point and look at the dynamical systems after k steps beyond such a starting point. The dynamical systems differ by some terms that involve the initial states and some terms which primarily involve the selection costs accumulated over k steps. Chapter 7 is largely devoted to showing that recombination drives down the difference terms involving the starting states by a factor of $k+1$, while terms involving selection costs remain bounded by a constant multiple of k/n . It follows that taking $k = \sqrt{n}$ steps beyond any starting point makes the former kind of terms drop by $1/\sqrt{n}$ before the latter kind of terms grow by more than $1/\sqrt{n}$. The dynamical systems are getting closer to each other over a time interval of length $k/n = 1/\sqrt{n}$ which becomes infinitesimal as n goes to infinity.

In our infinite-population setting, as we have said, the process of selection is a process which reweights the probabilities of finding various combinations of mutant alleles in the genotype of a randomly selected individual. The tree structure of recombination events can be specified prior to the introduction of selection. Because the transformation of probability measures by reweighting is at issue, it makes sense to represent the probability measures Q_k and O_k by their Radon-Nikodym derivatives with respect to a suitable underlying probability measure. The Poisson probability measure P_k serves this purpose. These probability measures, and their starred counterparts that keep track of the vintages of mutations from wild type, are separate objects for each separate generation. Joint distributions of variables across generations and sample paths of time-serial processes are not in the foreground. However, in the background lie the tree-structure of recombination and the accumulating effects of selection which do reach across generations and do lead to such joint distributions. The decomposition of genotypes in terms of vintages allows the relevant features of these background joint distributions to be inferred from current states, and so facilitates the proof.

Convergence with complete Poissonization

We embark on the proof of our main convergence result, Theorem 8.1, by establishing in this chapter a result, important in its own right, which provides the first step in that proof. For this result, Theorem 6.3, the recombination operator \mathfrak{R} is replaced by the operator \mathfrak{P} defined in (4.3) that immediately replaces a probability measure on \mathcal{G} by the distribution of a Poisson random measure with the same intensity measure. In place of $Q_k := (\mathfrak{M}_n \mathfrak{S}_n)^k Q_0$, we study the sequence of probability measures $O_k := (\mathfrak{P} \mathfrak{M}_n \mathfrak{S}_n)^k P_0$, defined in Notation 4.4. Both the measures O_k and the measures Π_{ρ_t} that define our continuous-time dynamical system are Poisson measures, determined by their intensity measures. It is simpler to work with the intensity measures, defined on \mathcal{M} , than with probability measures defined on \mathcal{G} . The relationship between discrete generations and continuous time, mediated by the scaling parameter n , has to figure prominently in the proof. Theorem 6.3 is essentially a *shadowing* result, but none of the standard shadowing theorems, such as those in [CKP95], seem to cover this case.

REMARK 6.1. In this and the following chapters we frequently make use of multiple “constants”, numbered successively as c_1, c_2, \dots or C_1, C_2, \dots . These constants may depend on the time horizon T , the Lipschitz constant σ for the selective cost S , the total mass $\nu(\mathcal{M})$ of the mutation measure, and the total mass $\rho_0(\mathcal{M})$ of the intensity measure for the initial population.

NOTATION 6.2. Write $\pi_k := \mu O_k$ for the intensity measure of the Poisson probability measure O_k .

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

$$\sup_{0 \leq t \leq T} \|\pi_{\lfloor tn \rfloor} - \rho_t\|_{\text{Was}} \leq \frac{A}{n} (1+T)T \exp\{B(1+T)T\},$$

and hence

$$\sup_{0 \leq t \leq T} \|O_{\lfloor tn \rfloor} - \Pi_{\rho_t}\|_{\text{Was}} \leq 4 \frac{A}{n} (1+T)T \exp\{B(1+T)T\}.$$

PROOF. It suffices by Lemma A.12 to establish the first inequality. Note that

$$(6.1) \quad \sup_{0 \leq t \leq T} \|\pi_{\lfloor tn \rfloor} - \rho_t\|_{\text{Was}} \leq \sup_{0 \leq k \leq \lfloor Tn \rfloor} \left[\|\pi_k - \rho_{k/n}\|_{\text{Was}} + \sup_{k/n \leq t \leq (k+1)/n \wedge T} \|\rho_t - \rho_{k/n}\|_{\text{Was}} \right].$$

We obtain a bound on the first term on the right-hand side of (6.1), namely,

$$(6.2) \quad \sup_{0 \leq k \leq \lfloor Tn \rfloor} \|\pi_k - \rho_{k/n}\|_{\text{Was}}$$

by establishing a bound for $\|\pi_{k+1} - \rho_{(k+1)/n}\|_{\text{Was}}$ in terms of $\|\pi_k - \rho_{k/n}\|_{\text{Was}}$ and then iterating that bound.

Given $\pi \in \mathcal{H}^+$ and the associated Poisson random measure X^π with intensity π , define $H_\pi : \mathcal{M} \rightarrow \mathbb{R}_+$ by

$$H_\pi(x) := \frac{\mathbb{E}[\exp\{-S(X^\pi + \delta_x)/n\}]}{\mathbb{E}[\exp\{-S(X^\pi)/n\}]}.$$

By the assumptions on the selective cost S and the inequality

$$e^{-b} - e^{-c} = \int_b^c e^{-z} dz \leq (c - b)e^{-a}$$

for $a \leq b \leq c$, we have

$$(6.3) \quad 0 \leq H_\pi(x) \leq 1, \quad x \in \mathcal{M},$$

and

$$(6.4) \quad \begin{aligned} |H_\pi(x) - H_\pi(y)| &\leq \frac{\mathbb{E}[\exp\{-S(X^\pi)/n\}] |S(X^\pi + \delta_x)/n - S(X^\pi + \delta_y)/n|}{\mathbb{E}[\exp\{-S(X^\pi)/n\}]} \\ &\leq \frac{\sigma}{n} \|\delta_x - \delta_y\|_{\text{Was}} \\ &= \frac{\sigma}{n} d(x, y), \end{aligned}$$

where d here is the metric on \mathcal{M} . We take advantage of the Lipschitz constant σ for the selective cost S to write $0 \leq S(g) \leq \sigma g(\mathcal{M})$, which implies for any $\pi \in \mathcal{H}^+$ that

$$\mathbb{E}[\exp\{-S(X^\pi)/n\}] \geq \mathbb{E}\left[\exp\left\{-\frac{\sigma}{n} X^\pi(\mathcal{M})\right\}\right] = \exp\left\{e^{-\sigma\pi(\mathcal{M})/n}\right\}.$$

Thus, for any $\pi', \pi'' \in \mathcal{H}^+$,

$$(6.5) \quad \begin{aligned} &|H_{\pi'}(x) - H_{\pi''}(x)| \\ &= \left| \mathbb{E}\left[\exp\{-S(X^{\pi''})/n\}\right] \mathbb{E}\left[\exp\{-S(X^{\pi'} + \delta_x)/n\}\right] \right. \\ &\quad \left. - \mathbb{E}\left[\exp\{-S(X^{\pi'})/n\}\right] \mathbb{E}\left[\exp\{-S(X^{\pi''} + \delta_x)/n\}\right] \right| \\ &\quad \left/ \left(\mathbb{E}\left[\exp\{-S(X^{\pi'})/n\}\right] \mathbb{E}\left[\exp\{-S(X^{\pi''})/n\}\right] \right) \right| \\ &\leq e^{\sigma(\pi'(\mathcal{M}) + \pi''(\mathcal{M}))/n} \left(\left| \mathbb{E}\left[\exp\{-S(X^{\pi''})/n\}\right] \mathbb{E}\left[\exp\{-S(X^{\pi'} + \delta_x)/n\}\right] \right. \right. \\ &\quad \left. \left. - \mathbb{E}\left[\exp\{-S(X^{\pi'})/n\}\right] \mathbb{E}\left[\exp\{-S(X^{\pi''} + \delta_x)/n\}\right] \right| \right. \\ &\quad \left. + \left| \mathbb{E}\left[\exp\{-S(X^{\pi'})/n\}\right] \mathbb{E}\left[\exp\{-S(X^{\pi'} + \delta_x)/n\}\right] \right. \right. \\ &\quad \left. \left. - \mathbb{E}\left[\exp\{-S(X^{\pi'})/n\}\right] \mathbb{E}\left[\exp\{-S(X^{\pi''} + \delta_x)/n\}\right] \right| \right) \end{aligned}$$

and

$$\begin{aligned}
& |H_{\pi'}(x) - H_{\pi''}(x)| \\
& \leq e^{\sigma(\pi'(\mathcal{M}) + \pi''(\mathcal{M}))/n} \left(\left| \mathbb{E} \left[\exp\{-S(X^{\pi''})/n\} \right] - \mathbb{E} \left[\exp\{-S(X^{\pi'})/n\} \right] \right| \right. \\
& \quad \left. + \left| \mathbb{E} \left[\exp\{-S(X^{\pi'} + \delta_x)/n\} \right] - \mathbb{E} \left[\exp\{-S(X^{\pi''} + \delta_x)/n\} \right] \right| \right) \\
(6.6) \quad & = e^{\sigma(\pi'(\mathcal{M}) + \pi''(\mathcal{M}))/n} \\
& \quad \times \left(\left| \mathbb{E} \left[1 - \exp\{-S(X^{\pi''})/n\} \right] - \mathbb{E} \left[1 - \exp\{-S(X^{\pi'})/n\} \right] \right| \right. \\
& \quad \left. + \left| \mathbb{E} \left[1 - \exp\{-S(X^{\pi'} + \delta_x)/n\} \right] - \mathbb{E} \left[1 - \exp\{-S(X^{\pi''} + \delta_x)/n\} \right] \right| \right).
\end{aligned}$$

Now we may apply Lemma A.11 with $\beta = 0$ and $C = \sigma/n$, followed by Lemma A.12, to obtain

$$|H_{\pi'}(x) - H_{\pi''}(x)| \leq e^{\sigma(\pi'(\mathcal{M}) + \pi''(\mathcal{M}))/n} \cdot \frac{8\sigma}{n} \|\pi' - \pi''\|_{\text{Was}}.$$

Now, returning to the task of bounding the quantity in (6.2),

$$\begin{aligned}
(6.7) \quad \|\pi_{k+1} - \rho_{(k+1)/n}\|_{\text{Was}} & \leq \left\| \pi_{k+1} - \frac{\nu}{n} - H_{\pi_k} \cdot \pi_k \right\|_{\text{Was}} \\
& \quad + \left\| H_{\pi_k} \cdot \pi_k - H_{\rho_{k/n}} \cdot \pi_k \right\|_{\text{Was}} \\
& \quad + \left\| H_{\rho_{k/n}} \cdot \pi_k - H_{\rho_{k/n}} \cdot \rho_{k/n} \right\|_{\text{Was}} \\
& \quad + \left\| \rho_{(k+1)/n} - \frac{\nu}{n} - H_{\rho_{k/n}} \cdot \rho_{k/n} \right\|_{\text{Was}}.
\end{aligned}$$

Consider the first term on the right-hand side of (6.7). By definition,

$$\pi_{k+1} = \mu O_{k+1} = \mu \mathfrak{P} \mathfrak{M}_n \mathfrak{S}_n O_k.$$

Thus,

$$\pi_{k+1} - \frac{\nu}{n} = \mu \mathfrak{P} \mathfrak{M}_n \mathfrak{S}_n O_k - \frac{\nu}{n} = \mu \mathfrak{M}_n \mathfrak{S}_n O_k - \frac{\nu}{n} = \mu \mathfrak{S}_n O_k = \mu \mathfrak{S}_n \Pi_{\pi_k}.$$

Hence, for a bounded Borel function $f : \mathcal{M} \rightarrow \mathbb{R}$,

$$\begin{aligned}
\int_{\mathcal{M}} f(x) d\left(\pi_{k+1} - \frac{\nu}{n}\right)(x) & = \frac{\int_{\mathcal{G}} g[f] \exp\{-S(g)/n\} d\Pi_{\pi_k}(g)}{\int_{\mathcal{G}} \exp\{-S(g)/n\} d\Pi_{\pi_k}(g)} \\
& = \frac{\int_{\mathcal{M}} f(x) \int_{\mathcal{G}} \exp\{-S(g + \delta_x)/n\} d\Pi_{\pi_k}(g) d\pi_k(x)}{\int_{\mathcal{G}} \exp\{-S(g)/n\} d\Pi_{\pi_k}(g)},
\end{aligned}$$

where we have again made use of Campbell's Theorem (see Proposition A.8).

Equivalently,

$$(6.8) \quad \pi_{k+1} - \frac{\nu}{n} = H_{\pi_k} \cdot \pi_k,$$

and so the first term on the right-hand side of (6.7) is zero.

It follows from (6.8) and (6.3) that

$$(6.9) \quad \pi_k(\mathcal{M}) \leq \rho_0(\mathcal{M}) + \frac{k}{n} \nu(\mathcal{M}), \quad k \geq 0.$$

Combining (6.9) with (6.5) establishes that the second term on the right-hand side of (6.7) is bounded by

$$(6.10) \quad \begin{aligned} & \frac{8\sigma}{n} e^{\sigma(\pi'(\mathcal{M})+\pi''(\mathcal{M}))/n} \|\pi_k - \rho_{k/n}\|_{\text{Was}} \pi_k(\mathcal{M}) \\ & \leq \frac{8\sigma}{n} e^{\sigma(\pi'(\mathcal{M})+\pi''(\mathcal{M}))/n} \left(\rho_0(\mathcal{M}) + \frac{k}{n} \nu(\mathcal{M}) \right) \|\pi_k - \rho_{k/n}\|_{\text{Was}}. \end{aligned}$$

From (6.3) and (6.4), the third term on the right-hand side of (6.7) is bounded by

$$(6.11) \quad \|H_{\rho_{k/n}}\|_{\text{Lip}} \|\pi_k - \rho_{k/n}\|_{\text{Was}} \leq \left(1 + \frac{\sigma}{n}\right) \|\pi_k - \rho_{k/n}\|_{\text{Was}}.$$

For the last term on the right-hand side of (6.7) we have

$$(6.12) \quad \begin{aligned} & \left\| \rho_{(k+1)/n} - \frac{\nu}{n} - H_{\rho_{k/n}} \cdot \rho_{k/n} \right\|_{\text{Was}} \\ & = \left\| \rho_{k/n} + \frac{1}{n} \nu + \int_0^{\frac{1}{n}} F_{\rho_{s+k/n}} \cdot \rho_{s+k/n} ds - \frac{\nu}{n} - H_{\rho_{k/n}} \cdot \rho_{k/n} \right\|_{\text{Was}} \\ & \leq \left\| \left(1 - H_{\rho_{k/n}} + \frac{1}{n} F_{\rho_{k/n}}\right) \cdot \rho_{k/n} \right\|_{\text{Was}} \\ & \quad + \int_0^{\frac{1}{n}} \|F_{\rho_{s+k/n}} \cdot \rho_{s+k/n} - F_{\rho_{k/n}} \cdot \rho_{k/n}\|_{\text{Was}} ds. \end{aligned}$$

By Lemma A.2 we know for any $\rho \in \mathcal{H}^+$ that

$$\begin{aligned} -\frac{1}{n} \mathbb{E}[S(X^\rho)] & \leq \log \mathbb{E} \left[\exp \left\{ -\frac{1}{n} S(X^\rho) \right\} \right] \\ & \leq -\frac{1}{n} \mathbb{E}[S(X^\rho)] + \frac{1}{2n^2} \mathbb{E}[S(X^\rho)^2] \exp \left\{ \frac{1}{n} \mathbb{E}[S(X^\rho)] \right\}, \end{aligned}$$

and a similar bound holds with X^ρ replaced by $X^\rho + \delta_x$. Note also that

$$\mathbb{E}[S(X^\rho)] \leq \mathbb{E}[S(X^\rho + \delta_x)] \leq \sigma \mathbb{E}[(X^\rho(\mathcal{M}) + 1)] = \sigma(\rho(\mathcal{M}) + 1)$$

and

$$\mathbb{E}[S(X^\rho)^2] \leq \mathbb{E}[S(X^\rho + \delta_x)^2] \leq \sigma^2 \mathbb{E}[(X^\rho(\mathcal{M}) + 1)^2] = \sigma^2(\rho(\mathcal{M})^2 + 3\rho(\mathcal{M}) + 1).$$

Applying the obvious bound

$$(6.13) \quad \rho_t(\mathcal{M}) \leq \rho_0(\mathcal{M}) + t\nu(\mathcal{M}), \quad t \geq 0$$

and the inequality $1 - y \leq e^{-y} \leq 1 - y + \frac{y^2}{2}$ (for $y \geq 0$), it follows that

$$(6.14) \quad \begin{aligned} & \left| 1 - H_{\rho_{k/n}}(x) + \frac{1}{n} F_{\rho_{k/n}}(x) \right| \\ & = \left| 1 - \exp \{ \mathbb{E}[S(X^{\rho_{k/n}}) - S(X^{\rho_{k/n}} + \delta_x)] \} \right| \\ & \quad + \frac{1}{n} \mathbb{E} [S(X^{\rho_{k/n}} + \delta_x) - S(X^{\rho_{k/n}})] \left| + \frac{C_1}{n^2} \right. \\ & \leq \frac{C_2}{n^2}. \end{aligned}$$

Thus,

$$(6.15) \quad \left\| \left(1 - H_{\rho_{k/n}} + \frac{1}{n} F_{\rho_{k/n}} \right) \cdot \rho_{k/n} \right\|_{\text{Was}} \leq \frac{C_2}{n^2} \rho_{k/n}(\mathcal{M})$$

$$\leq \frac{C_3}{n^2}$$

for a suitable constant C_3 .

From Lemma 2.5 and (6.13), for any $0 \leq u \leq v$,

$$(6.16) \quad \|\rho_v - \rho_u\|_{\text{Was}} \leq (v - u)\nu(\mathcal{M}) + 2\sigma \int_u^v (\rho_0(\mathcal{M}) + w\nu(\mathcal{M})) dw$$

$$\leq (v - u)((1 + 2\sigma v)\nu(\mathcal{M}) + 2\sigma\rho_0(\mathcal{M})).$$

In particular, for $0 \leq s \leq \frac{1}{n}$,

$$(6.17) \quad \left\| F_{\rho_{s+k/n}} \cdot \rho_{s+k/n} - F_{\rho_{k/n}} \cdot \rho_{k/n} \right\|_{\text{Was}}$$

$$\leq \left\| F_{\rho_{s+k/n}} \cdot \rho_{s+k/n} - F_{\rho_{s+k/n}} \cdot \rho_{k/n} \right\|_{\text{Was}}$$

$$+ \left\| F_{\rho_{s+k/n}} \cdot \rho_{k/n} - F_{\rho_{k/n}} \cdot \rho_{k/n} \right\|_{\text{Was}}$$

$$\leq \left\| F_{\rho_{s+k/n}} \right\|_{\text{Lip}} \left\| \rho_{s+k/n} - \rho_{k/n} \right\|_{\text{Was}}$$

$$+ \sup_{x \in \mathcal{M}} |F_{\rho_{s+k/n}} - F_{\rho_{k/n}}| \rho_{k/n}(\mathcal{M})$$

$$\leq \frac{2\sigma}{n} ((1 + 2\sigma(k+1)/n)\nu(\mathcal{M}) + 2\sigma\rho_0(\mathcal{M}))$$

$$+ \frac{8\sigma}{n} ((1 + 2\sigma(k+1)/n)\nu(\mathcal{M}) + 2\sigma\rho_0(\mathcal{M}))$$

$$\times \left(\rho_0(\mathcal{M}) + \frac{k}{n}\nu(\mathcal{M}) \right),$$

where we combined (6.16) and (6.13) with Lemma 2.5 and Lemma 2.6 in the last inequality.

Hence,

$$(6.18) \quad \int_0^{\frac{1}{n}} \left\| F_{\rho_{s+k/n}} \cdot \rho_{s+k/n} - F_{\rho_{k/n}} \cdot \rho_{k/n} \right\|_{\text{Was}} ds \leq \frac{C_4}{n^2},$$

for a suitable constant C_4 .

Combining (6.15) and (6.18), we see from (6.12) that

$$(6.19) \quad \left\| \rho_{(k+1)/n} - \frac{\nu}{n} - H_{\rho_{k/n}} \cdot \rho_{k/n} \right\|_{\text{Was}} \leq \frac{C_5}{n^2}$$

for a suitable constant C_5 .

Recall from (6.8) that the first term on the right-hand side of (6.7) is zero and bound the remaining terms using (6.10), (6.11) and (6.19) to obtain

$$(6.20) \quad \|\pi_{k+1} - \rho_{(k+1)/n}\|_{\text{Was}} \leq \left(1 + \frac{C_6}{n} \right) \|\pi_k - \rho_{k/n}\|_{\text{Was}} + \frac{C_7}{n^2}$$

for $(k+1) \leq nT$, where C_6 and C_7 are constants. Iterating (6.20) leads to

$$\begin{aligned}
 \|\pi_{k+1} - \rho_{(k+1)/n}\|_{\text{Was}} &\leq \sum_{j=0}^k \left(1 + \frac{C_6}{n}\right)^j \frac{C_7}{n^2} \\
 (6.21) \qquad \qquad \qquad &\leq nT \exp\left\{\frac{C_6}{n}\right\}^{nT} \frac{C_7}{n^2} \\
 &= \exp\{C_8\} \frac{C_9}{n}
 \end{aligned}$$

for $(k+1) \leq nT$.

The result now follows by using (6.21) and (6.16) to bound the first and second terms, respectively, on the right-hand side of (6.1). \square

Supporting lemmas for the main convergence result

7.1. Estimates for Radon-Nikodym derivatives

In this chapter we establish a set of lemmas which provide the central technical tools used in the proof of the main convergence result in Chapter 8. As previewed in Chapter 5, we work with weights in the form of Radon-Nikodym derivatives, because selection, in our infinite population, is a process of reweighting of population frequencies of genotypes. Furthermore, we suppose that our initial population Q_0 can be treated as a reweighted version of the Poisson probability measure $P_0 = \mathfrak{P}Q_0$. Thus, it is natural to express our full discrete-generation system in terms of Radon-Nikodym derivatives with respect to a family of Poisson probability measures.

The substantive effects of selection have been treated in Chapter 6. That chapter shows how, in the presence of selection, the completely Poissonized version of the discrete-generation system aligns with the continuous-time dynamical system. From this point on in the story, attention focuses on short stretches of time which become infinitesimal as the scaling parameter n goes to infinity. Over these short stretches, we show that the effect of selection can be treated as negligible whereas the total effect of recombination tracks closely that of complete Poissonization.

We thus require bounds on the overall effects of selection, acting along with mutation and recombination, over k generations. We are interested in the outcome at the k -th generation, but the effects of selection accumulate across generations and depend on the genotypes of the ancestors, portions of whose genomes are recombined into the genotype of any individual sampled from the population at generation k . Reconstruction of the history of accumulating selection is made possible by the decomposition of elements of \mathcal{G} according to vintage and thus by the starred operators which update in each generation the population distribution of mutations from ancestral wild type labeled according to vintage.

The capstone of this chapter, Lemma 7.15 is a bound on the difference between the starred operator Q_k^* and its own Poissonization $\mathfrak{P}Q_k^*$. This bound is obtained by combining an approximation for dQ_k^*/dP_k^* derived in this section with an approximation for $d\mathfrak{P}Q_k^*/dP_k^*$ derived in the following section. In both approximations, terms that are retained arise from contributions of the initial probability measures. Contributions from selection are absorbed into an error term bounded by a multiple of k/n .

We begin with an elementary observation concerning how the action of the starred recombination operator \mathfrak{A}^* may be represented in terms of Radon-Nikodym derivatives.

LEMMA 7.1. *Suppose that P^* is a Poisson random measure. Then, for $\mathbf{g} \in \mathcal{G}^{k+1}$,*

$$\frac{d\mathfrak{R}^*Q^*}{dP^*}(\mathbf{g}) = \int_{\mathcal{B}(\mathcal{M})} \int_{\mathcal{G}^{k+1}} \frac{dQ^*}{dP^*}(\mathbf{g}|_A + \mathbf{g}'|_{A^c}) \frac{dQ^*}{dP^*}(\mathbf{g}'|_A + \mathbf{g}|_{A^c}) dP^*(\mathbf{g}') d\mathcal{R}(A).$$

PROOF. By Corollary A.7, for any bounded Borel function $F : \mathcal{G}^{k+1} \rightarrow \mathbb{R}$,

$$\begin{aligned} \mathfrak{R}^*Q^*[F] &= \int_{\mathcal{B}(\mathcal{M})} \int_{\mathcal{G}^{k+1}} \int_{\mathcal{G}^{k+1}} F(\mathbf{g}|_A + \mathbf{g}'|_{A^c}) dQ^*(\mathbf{g}) dQ^*(\mathbf{g}') d\mathcal{R}(A) \\ &= \int_{\mathcal{B}(\mathcal{M})} \int_{\mathcal{G}^{k+1}} \int_{\mathcal{G}^{k+1}} F(\mathbf{g}|_A + \mathbf{g}'|_{A^c}) \frac{dQ^*}{dP^*}(\mathbf{g}) \frac{dQ^*}{dP^*}(\mathbf{g}') dP^*(\mathbf{g}) dP^*(\mathbf{g}') d\mathcal{R}(A) \\ &= \int_{\mathcal{B}(\mathcal{M})} \int_{\mathcal{G}^{k+1}} \int_{\mathcal{G}^{k+1}} F(\mathbf{g}) \frac{dQ^*}{dP^*}(\mathbf{g}|_A + \mathbf{g}'|_{A^c}) \frac{dQ^*}{dP^*}(\mathbf{g}'|_A + \mathbf{g}|_{A^c}) dP^*(\mathbf{g}') dP^*(\mathbf{g}) d\mathcal{R}(A). \end{aligned}$$

This is equivalent to the claim. \square

We next develop apparatus for reconstructing the ancestral genotypes that are consistent with a particular genotype (decomposed by vintages) that is observed at the end of k generations. At each ancestral generation, there is a partition of \mathcal{M} (generally including empty sets), as presented in Section 4.3. Each set in this partition corresponds to an ancestor in the given generation. Mutant alleles within the set are retained from the genotype of the ancestor while mutant alleles outside the set are lost through recombination. We work with functions that assign to each set a surrogate for the missing piece of the ancestral genome. This additional bookkeeping facilitates our development of bounds on the “de-Poissonizing” effects of selection.

DEFINITION 7.2. Let \mathcal{A} be a partition of \mathcal{M} into Borel sets and let $\xi : \mathcal{A} \rightarrow \mathcal{G}$ be a family of genotypes indexed by the sets of \mathcal{A} . Given $F : \mathcal{G} \rightarrow \mathbb{R}$ and $g \in \mathcal{G}$, set

$$\tilde{F}_{\mathcal{A}}(g, \xi) = \sum_{A \in \mathcal{A}} F(g|_A + \xi(A)|_{A^c}).$$

Note that if \mathcal{A} is the trivial partition $\{\mathcal{M}\}$, then $\tilde{F}_{\mathcal{A}}(g, \xi) = F(g)$ for any $\xi : \mathcal{A} \rightarrow \mathcal{G}$.

LEMMA 7.3. *Fix a Borel partition \mathcal{A} of \mathcal{M} and a family of genotypes $\xi : \mathcal{A} \rightarrow \mathcal{G}$. Then, for any $F : \mathcal{G} \rightarrow \mathbb{R}$,*

$$\left| \tilde{F}_{\mathcal{A}}(g, \xi) - \tilde{F}_{\mathcal{A}}(g', \xi) \right| \leq \text{Lip}_{\text{TV}} F \cdot \|g - g'\|_{\text{TV}}.$$

PROOF. We have

$$\begin{aligned} \left| \tilde{F}_{\mathcal{A}}(g, \xi) - \tilde{F}_{\mathcal{A}}(g', \xi) \right| &\leq \sum_{A \in \mathcal{A}} |F(g|_A + \xi(A)|_{A^c}) - F(g'|_A + \xi(A)|_{A^c})| \\ &\leq \sum_{A \in \mathcal{A}} \text{Lip}_{\text{TV}} F \cdot \|g|_A - g'|_A\|_{\text{TV}} \\ &= \text{Lip}_{\text{TV}} F \cdot \|g - g'\|_{\text{TV}}. \end{aligned}$$

\square

NOTATION 7.4. Recall the ancestral tree \mathcal{L} and the random recombination sets R_{β} and random genotypes Γ_{β} indexed by the nodes $\beta \in \mathcal{L}$ introduced in Section 4.4. Fix a node $\ell \in \mathcal{L}$ with vintage $V(\ell) = k$. Any set A in the random

partition $\mathcal{A}_i(\ell) = \mathcal{A}_{k,i}$ corresponds to a node $\omega \preceq \ell$ with vintage $V(\omega) = i$. Define $\xi_{k,i} : \mathcal{A}_{k,i} \rightarrow \mathcal{G}$ by $\xi_{k,i}(A) := \Xi_\omega = \sum_{v \preceq \omega} \Gamma_v |_{W(v,\omega)}$.

We now state and prove a lemma which gives an exact representation of the Radon-Nikodym derivative of the starred measure Q_k^* with respect to P_k^* at generation k in terms of ancestral episodes of selection. The factors that make the total probability integrate to 1 are absorbed into a normalization constant C .

LEMMA 7.5. *Set $H := \log dQ_0/dP_0$. Suppose for $k \in \mathbb{N}_0$ that the partitions $\mathcal{A}_{k,0}, \dots, \mathcal{A}_{k,k}$ are defined using the node $\ell \in \mathcal{L}$ with vintage $V(\ell) = k$. Then,*

$$\begin{aligned} \frac{dQ_k^*}{dP_k^*}(g_0, \dots, g_k) &= C \mathbb{E} \left[\exp \left\{ \tilde{H}_{\mathcal{A}_{k,0}}(g_0, \xi_{k,0}) - \frac{1}{n} \sum_{i=0}^{k-1} \tilde{S}_{\mathcal{A}_{k,i}}(g_0 + \dots + g_i, \xi_{k,i}) \right\} \right] \\ &= C \mathbb{E} \left[\exp \left\{ \sum_{\omega \preceq \ell} J_{V(\omega)} \left((g_0 + \dots + g_{V(\omega)}) |_{W(\omega,\ell)} + (\Xi_\omega) |_{W(\omega,\ell)^c} \right) \right\} \right], \end{aligned}$$

where C does not depend on (g_0, \dots, g_k) and the functions $J_i : \mathcal{G} \rightarrow \mathbb{R}$ are given by $J_0(g) := H(g) - S(g)/n$ and $J_i(g) := -S(g)/n$ for $1 \leq i \leq k-1$. The expectation is taken with respect to the random collection of genotypes $\xi_{k,i}$ (or Ξ_ℓ) and the random partitions $\mathcal{A}_{k,i}$ (or $W(\omega, \ell)$).

PROOF. By definition,

$$\begin{aligned} \tilde{H}_{\mathcal{A}_{k,0}}(g, \xi_{k,0}) &= \sum_{A \in \mathcal{A}_{k,0}} H(g|_A + \xi_{k,0}(A)|_{A^c}) \\ &= \sum_{\omega \preceq \ell, V(\omega)=0} H(g|_{W(\omega,\ell)} + (\Xi_\omega)|_{W(\omega,\ell)^c}) \end{aligned}$$

and

$$\begin{aligned} \tilde{S}_{\mathcal{A}_{k,i}}(g, \xi_{k,i}) &= \sum_{A \in \mathcal{A}_{k,i}} S(g|_A + \xi_{k,i}(A)|_{A^c}) \\ &= \sum_{\omega \preceq \ell, V(\omega)=i} S(g|_{W(\omega,\ell)} + (\Xi_\omega)|_{W(\omega,\ell)^c}), \end{aligned}$$

so

$$\begin{aligned} \tilde{H}_{\mathcal{A}_{k,0}}(g_0, \xi_{k,0}) - \frac{1}{n} \sum_{i=0}^{k-1} \tilde{S}_{\mathcal{A}_{k,i}}(g_0 + \dots + g_i, \xi_{k,i}) \\ = \sum_{\omega \preceq \ell} J_{V(\omega)} \left((g_0 + \dots + g_{V(\omega)}) |_{W(\omega,\ell)} + (\Xi_\omega) |_{W(\omega,\ell)^c} \right). \end{aligned}$$

The equality of the two expectations follows.

The remainder of the proof is by induction on k . For $k = 0$, the partition $\mathcal{A}_{0,0}$ is the trivial partition consisting of the single set \mathcal{M} , and $\tilde{H}_{\mathcal{A}_{0,0}}(g, \xi) = H(g)$, and so the result is obvious. Now suppose the result is true for $k-1 \geq 0$.

By definition, $Q_k^* = \mathfrak{R}^* \mathfrak{M}_n^* \mathfrak{S}_n^* Q_{k-1}^*$ and $P_k^* = \mathfrak{M}_n^* P_{k-1}^*$. Thus,

$$\begin{aligned}
& \frac{dQ_k^*}{dP_k^*}(g_0, \dots, g_k) \\
&= \frac{d\mathfrak{R}^* \mathfrak{M}_n^* \mathfrak{S}_n^* Q_{k-1}^*}{d\mathfrak{M}_n^* P_{k-1}^*}(g_0, \dots, g_k) \\
(7.1) \quad &= \int_{\mathcal{B}(\mathcal{M})} \int_{\mathcal{G}^{k+1}} \frac{d\mathfrak{M}_n^* \mathfrak{S}_n^* Q_{k-1}^*}{d\mathfrak{M}_n^* P_{k-1}^*}(\mathbf{g}|_A + \mathbf{g}'|_{A^c}) \\
&\quad \times \frac{d\mathfrak{M}_n^* \mathfrak{S}_n^* Q_{k-1}^*}{d\mathfrak{M}_n^* P_{k-1}^*}(\mathbf{g}'|_A + \mathbf{g}|_{A^c}) d\mathfrak{M}_n^* P_{k-1}^*(\mathbf{g}') d\mathcal{R}(A) \\
&= \int_{\mathcal{B}(\mathcal{M})} \int_{\mathcal{G}^{k+1}} \frac{d\mathfrak{S}_n^* Q_{k-1}^*}{dP_{k-1}^*}((g_0, \dots, g_{k-1})|_A + (g'_0, \dots, g'_{k-1})|_{A^c}) \\
&\quad \times \frac{d\mathfrak{S}_n^* Q_{k-1}^*}{dP_{k-1}^*}((g'_0, \dots, g'_{k-1})|_A + (g_0, \dots, g_{k-1})|_{A^c}) dP_{k-1}^*(\mathbf{g}') d\mathcal{R}(A),
\end{aligned}$$

by Lemma 7.1 and the observation that for any two probability measures Q^* and P^* on \mathcal{G}^{k-1} we have

$$\frac{d\mathfrak{M}_n^* Q^*}{d\mathfrak{M}_n^* P^*}(g_0, \dots, g_k) = \frac{d(\Pi_{\nu/n} \otimes Q^*)}{d(\Pi_{\nu/n} \otimes P^*)}(g_0, \dots, g_k) = \frac{dQ^*}{dP^*}(g_0, \dots, g_{k-1}).$$

Let the parents of the node ℓ be the two nodes $\ell', \ell'' \in \mathcal{L}$ with common vintage $k-1$. Suppose first of all that the random partitions $\mathcal{A}_{k-1,0}, \dots, \mathcal{A}_{k-1,k-1}$ are defined using the node ℓ' . By the induction hypothesis,

$$\begin{aligned}
& \frac{d\mathfrak{S}_n^* Q_{k-1}^*}{dP_{k-1}^*}(g_0, \dots, g_{k-1}) \\
(7.2) \quad &= C_1 \mathbb{E} \left[\exp \left\{ \sum_{\omega \preceq \ell'} J_{V(\omega)} \left((g_0 + \dots + g_{V(\omega)})|_{W(\omega, \ell')} + (\Xi_\omega)|_{W(\omega, \ell')^c} \right) \right\} \right] \\
&\quad \times \exp \left\{ -\frac{1}{n} S(g_0 + \dots + g_{k-1}) \right\} \\
&= C_1 \mathbb{E} \left[\exp \left\{ \sum_{\omega \preceq \ell'} J_{V(\omega)} \left((g_0 + \dots + g_{V(\omega)})|_{W(\omega, \ell')} + (\Xi_\omega)|_{W(\omega, \ell')^c} \right) \right\} \right]
\end{aligned}$$

for some constant C_1 . Of course, the same equation holds if ℓ' is replaced by ℓ'' .

Recall that R_ℓ is a $\mathcal{B}(\mathcal{M})$ -valued random variable with distribution \mathcal{R} and the collection of random sets $\{R_\omega : \omega \preceq \ell\} = \{R_\ell\} \cup \{R_\omega : \omega \preceq \ell'\} \cup \{R_\omega : \omega \preceq \ell''\}$ is independent. By convention, the genotype of the individual ℓ coincides with that of the parent ℓ' on the set R_ℓ and with that of the parent ℓ'' on the complementary set R_ℓ^c . For $0 \leq i \leq k-1$, define a random genotype by

$$Y_i := \sum_{\substack{\omega \preceq \ell', \\ V(\omega)=i}} (\Gamma_\omega)|_{W(\omega, \ell') \cap R_\ell^c} + \sum_{\substack{\omega \preceq \ell'', \\ V(\omega)=i}} (\Gamma_\omega)|_{W(\omega, \ell'') \cap R_\ell}.$$

Note that the conditional distribution of $\mathbf{Y} := (Y_0, \dots, Y_{k-1})$ given R_ℓ is P_{k-1}^* , so P_{k-1}^* is also the unconditional distribution of \mathbf{Y} and \mathbf{Y} is independent of R_ℓ . Thus,

the distribution of

$$\left(R_\ell, (Y_0, \dots, Y_{k-1})|_{R_\ell^{\mathfrak{C}}}, R_\ell^{\mathfrak{C}}, (Y_0, \dots, Y_{k-1})|_{R_\ell} \right)$$

is the push-forward of the probability measure $\mathcal{R} \otimes P_{k-1}^*$ by the map

$$(A, (g_0, \dots, g_{k-1})) \mapsto \left(A, (g_0, \dots, g_{k-1})|_{A^{\mathfrak{C}}}, A^{\mathfrak{C}}, (g_0, \dots, g_{k-1})|_A \right),$$

which is in turn that of the push-forward of the probability measure $\mathcal{R} \otimes P_{k-1}^* \otimes P_{k-1}^*$ by the map

$$(A, (g'_0, \dots, g'_{k-1}), (g''_0, \dots, g''_{k-1})) \mapsto \left(A, (g'_0, \dots, g'_{k-1})|_{A^{\mathfrak{C}}}, A^{\mathfrak{C}}, (g''_0, \dots, g''_{k-1})|_A \right).$$

Therefore, substituting (7.2) and the analogous equation with ℓ' replaced by ℓ'' into (7.1) gives

$$\begin{aligned} & \frac{dQ_k^*}{dP_k^*}(g_0, \dots, g_k) \\ &= \mathbb{E} \left[\frac{d\mathfrak{S}_n^* Q_{k-1}^*}{dP_{k-1}^*}((g_0, \dots, g_{k-1})|_{R_\ell} + (Y_0, \dots, Y_{k-1})|_{R_\ell^{\mathfrak{C}}}) \right. \\ & \quad \left. \times \frac{d\mathfrak{S}_n^* Q_{k-1}^*}{dP_{k-1}^*}((Y_0, \dots, Y_{k-1})|_{R_\ell} + (g_0, \dots, g_{k-1})|_{R_\ell^{\mathfrak{C}}}) \right] \\ &= C_1^2 \mathbb{E} \left[\mathbb{E} \left[\exp \left\{ \sum_{\omega \preceq \ell'} J_{V(\omega)} \left((g_0 + \dots + g_{V(\omega)})|_{W(\omega, \ell') \cap R_\ell} \right. \right. \right. \right. \\ & \quad \left. \left. \left. + (Y_0 + \dots + Y_{V(\omega)})|_{W(\omega, \ell') \cap R_\ell^{\mathfrak{C}}} + (\Xi_\omega)|_{W(\omega, \ell')^{\mathfrak{C}}} \right) \right\} \middle| R_\ell \right] \right. \\ & \quad \left. \times \mathbb{E} \left[\exp \left\{ \sum_{\omega \preceq \ell''} J_{V(\omega)} \left((g_0 + \dots + g_{V(\omega)})|_{W(\omega, \ell'') \cap R_\ell^{\mathfrak{C}}} \right. \right. \right. \right. \right. \\ & \quad \left. \left. \left. + (Y_0 + \dots + Y_{V(\omega)})|_{W(\omega, \ell'') \cap R_\ell} + (\Xi_\omega)|_{W(\omega, \ell'')^{\mathfrak{C}}} \right) \right\} \middle| R_\ell \right] \right] \\ &= C_1^2 \mathbb{E} \left[\exp \left\{ \sum_{\omega \preceq \ell'} J_{V(\omega)} \left((g_0 + \dots + g_{V(\omega)})|_{W(\omega, \ell') \cap R_\ell} \right. \right. \right. \right. \\ & \quad \left. \left. \left. + (Y_0 + \dots + Y_{V(\omega)})|_{W(\omega, \ell') \cap R_\ell^{\mathfrak{C}}} + (\Xi_\omega)|_{W(\omega, \ell')^{\mathfrak{C}}} \right) \right\} \right. \\ & \quad \left. + \sum_{\omega \preceq \ell''} J_{V(\omega)} \left((g_0 + \dots + g_{V(\omega)})|_{W(\omega, \ell'') \cap R_\ell^{\mathfrak{C}}} \right. \right. \\ & \quad \left. \left. \left. + (Y_0 + \dots + Y_{V(\omega)})|_{W(\omega, \ell'') \cap R_\ell} + (\Xi_\omega)|_{W(\omega, \ell'')^{\mathfrak{C}}} \right) \right\} \right]. \end{aligned}$$

Now, $\omega \prec \ell$ if and only if either $\omega \preceq \ell'$ or $\omega \preceq \ell''$. If $\omega \preceq \ell'$, then, by definition, $W(\omega, \ell) = W(\omega, \ell') \cap R_\ell$. Also by definition,

$$\begin{aligned} & (Y_0 + \cdots + Y_{V(\omega)})|_{W(\omega, \ell') \cap R_\ell^{\mathfrak{C}}} \\ &= \sum_{i=0}^{V(\omega)} \left[\sum_{\substack{v \preceq \ell', \\ V(v)=i}} (\Gamma_v)|_{W(v, \ell') \cap R_\ell^{\mathfrak{C}}} + \sum_{\substack{v \preceq \ell'', \\ V(v)=i}} (\Gamma_v)|_{W(v, \ell'') \cap R_\ell} \right] \Big|_{W(\omega, \ell') \cap R_\ell^{\mathfrak{C}}} \\ &= \sum_{v \preceq \omega} (\Gamma_v)|_{W(v, \ell') \cap R_\ell^{\mathfrak{C}}}, \end{aligned}$$

since for $\omega \preceq \ell'$ and v with $V(v) \leq V(\omega)$,

$$(W(v, \ell'') \cap R_\ell) \cap (W(\omega, \ell') \cap R_\ell^{\mathfrak{C}}) = \emptyset$$

and

$$(W(v, \ell') \cap R_\ell^{\mathfrak{C}}) \cap (W(\omega, \ell') \cap R_\ell^{\mathfrak{C}}) = \begin{cases} W(v, \ell') \cap R_\ell^{\mathfrak{C}}, & \text{if } v \preceq \omega, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Lastly,

$$(\Xi_\omega)|_{W(\omega, \ell')^{\mathfrak{C}}} = \left(\sum_{v \preceq \omega} (\Gamma_v)|_{W(v, \omega)} \right) \Big|_{W(\omega, \ell')^{\mathfrak{C}}} = \sum_{v \preceq \omega} (\Gamma_v)|_{W(v, \omega) \cap W(\omega, \ell')^{\mathfrak{C}}}$$

and

$$(\Xi_\omega)|_{W(\omega, \ell)^{\mathfrak{C}}} = \sum_{v \preceq \omega} (\Gamma_v)|_{W(v, \omega) \cap W(\omega, \ell)^{\mathfrak{C}}}.$$

Thus,

$$\begin{aligned} (Y_0 + \cdots + Y_{V(\omega)})|_{W(\omega, \ell') \cap R_\ell^{\mathfrak{C}}} + (\Xi_\omega)|_{W(\omega, \ell')^{\mathfrak{C}}} &= \left(\sum_{v \preceq \omega} \Gamma_v \right) \Big|_{W(\omega, \ell)^{\mathfrak{C}}} \\ &= (\Xi_\omega)|_{W(\omega, \ell)^{\mathfrak{C}}} \end{aligned}$$

because

$$W(\omega, \ell)^{\mathfrak{C}} = (W(\omega, \ell') \cap R_\ell)^{\mathfrak{C}} = W(\omega, \ell')^{\mathfrak{C}} \cup R_\ell^{\mathfrak{C}}.$$

Similarly, if $\omega \preceq \ell''$, then $W(\omega, \ell) = W(\omega, \ell'') \cap R_\ell^{\mathfrak{C}}$ and

$$(Y_0 + \cdots + Y_{V(\omega)})|_{W(\omega, \ell'') \cap R_\ell} + (\Xi_\omega)|_{W(\omega, \ell'')^{\mathfrak{C}}} = (\Xi_\omega)|_{W(\omega, \ell)^{\mathfrak{C}}},$$

completing the induction. \square

Recall from (2.7) that for two probability measures P and Q we denote the total-variation Lipschitz constant of $\log dQ/dP$ by $\Theta(P, Q)$.

LEMMA 7.6. *Suppose that $\hat{H} := \Theta(P_0, Q_0) < \infty$. Then, $\Theta(P_k^*, Q_k^*)$ and $\Theta(P_k, Q_k)$ are both bounded by $\hat{H} + \frac{k}{n}\sigma$ for all $k \in \mathbb{N}_0$.*

PROOF. Consider first the claim for $\log dQ_k^*/dP_k^*$. It suffices to check for any $\mathbf{g} \in \mathcal{G}^{k+1}$, $x \in \mathcal{M}$, and $0 \leq j \leq k$ that

$$(7.3) \quad \left| \log \frac{dQ_k^*}{dP_k^*}(\mathbf{g} + \delta_x^{(j)}) - \log \frac{dQ_k^*}{dP_k^*}(\mathbf{g}) \right| \leq \hat{H} + \frac{k}{n}\sigma,$$

where $\delta_x^{(j)} := (0, \dots, 0, \delta_x, 0, \dots, 0) \in \mathcal{G}^{k+1}$ with δ_x in the j^{th} coordinate.

Suppose first that $j = 0$. By Lemma 7.5,

$$\frac{dQ_k^*}{dP_k^*}(\mathbf{g}) = C\mathbb{E} \left[\exp \left\{ \tilde{H}_{\mathcal{A}_{k,0}}(g_0 + \delta_x, \xi_{k,0}) - \frac{1}{n} \sum_{i=0}^{k-1} \tilde{S}_{\mathcal{A}_{k,i}}(g_0 + \dots + g_i + \delta_x, \xi_{k,i}) \right\} \right].$$

It follows from Lemma 7.3 that

$$\begin{aligned} & \tilde{H}_{\mathcal{A}_{k,0}}(g_0, \xi_{k,0}) - \hat{H} - \frac{1}{n} \sum_{i=0}^{k-1} \tilde{S}_{\mathcal{A}_{k,i}}(g_0 + \dots + g_i, \xi_{k,i}) - \frac{k}{n}\sigma \\ & \leq \tilde{H}_{\mathcal{A}_{k,0}}(g_0 + \delta_x, \xi_{k,0}) - \frac{1}{n} \sum_{i=0}^{k-1} \tilde{S}_{\mathcal{A}_{k,i}}(g_0 + \dots + g_i + \delta_x, \xi_{k,i}) \\ & \leq \tilde{H}_{\mathcal{A}_{k,0}}(g_0, \xi_{k,0}) + \hat{H} - \frac{1}{n} \sum_{i=0}^{k-1} \tilde{S}_{\mathcal{A}_{k,i}}(g_0 + \dots + g_i, \xi_{k,i}) + \frac{k}{n}\sigma, \end{aligned}$$

and (7.3) follows immediately. The proof for $1 \leq j \leq k$ is similar.

In order to establish the claim for $\log dQ_k/dP_k$, it suffices to check for any $g \in \mathcal{G}$ and $x \in \mathcal{M}$ that

$$(7.4) \quad \left| \log \frac{dQ_k}{dP_k}(g + \delta_x) - \log \frac{dQ_k}{dP_k}(g) \right| \leq \hat{H} + \frac{k}{n}\sigma.$$

Define \mathbf{X} to be the canonical random variable on \mathcal{G}^* . Observe that

$$\frac{dQ_k}{dP_k}(g) = P_k^* \left[\frac{dQ_k^*}{dP_k^*}(\mathbf{X}) \mid \sum \mathbf{X} = g \right]$$

Under P_k^* , the conditional distribution of \mathbf{X} given the event $\{\sum \mathbf{X} = g = \sum_{j=1}^{g(\mathcal{M})} \delta_{x_j}\}$, is the distribution of the \mathcal{G}^{k+1} -valued random variable

$$\sum_{j=1}^{g(\mathcal{M})} \delta_{x_j^{(\mathbf{i}_j)}},$$

where the $\{0, 1, \dots, k\}$ -valued random variables \mathbf{i}_j , $1 \leq j \leq g(\mathcal{M})$, are independent with distribution given by

$$\mathbb{P}\{\mathbf{i}_j = 0\} = 1 - k\pi(x_j), \quad \mathbb{P}\{\mathbf{i}_j = i\} = \pi(x_j) \text{ for } 1 \leq i \leq k,$$

with

$$\pi(x) := \frac{d\nu/n}{d\mu P_k}(x) = \frac{d\nu/n}{d[\rho_0 + k\nu/n]}(x).$$

Consequently, for a bounded measurable function $F : \mathcal{G}^{k+1} \rightarrow \mathbb{R}$, we have

$$P_k^*[F(\mathbf{X}) \mid \sum \mathbf{X} = g + \delta_x] = \mathbb{E} \left[P_k^*[F(\mathbf{X} + \delta_x^{(\mathbf{i})}) \mid \sum \mathbf{X} = g] \right],$$

where \mathbf{i} is a $\{0, 1, \dots, k\}$ -valued random variable with distribution

$$\mathbb{P}\{\mathbf{i} = 0\} = 1 - k\pi(x), \quad \mathbb{P}\{\mathbf{i} = i\} = \pi(x) \text{ for } 1 \leq i \leq k.$$

Thus,

$$\frac{dQ_k}{dP_k}(g + \delta_x) = \mathbb{E} \left[P_k^* \left[\frac{dQ_k^*}{dP_k^*}(\mathbf{X}) \cdot \frac{dQ_k^*/dP_k^*(\mathbf{X} + \delta_x^{(i)})}{dQ_k^*/dP_k^*(\mathbf{X})} \mid \Sigma \mathbf{X} = g \right] \right].$$

It is immediate from (7.3) that

$$\exp \left\{ -\hat{H} - \frac{\sigma k}{n} \right\} \leq \frac{dQ_k^*/dP_k^*(\mathbf{g} + \delta_x^{(i)})}{dQ_k^*/dP_k^*(\mathbf{g})} \leq \exp \left\{ \hat{H} + \frac{\sigma k}{n} \right\},$$

for $0 \leq i \leq k$, and so (7.4) holds. \square

COROLLARY 7.7. *Suppose that $\hat{H} := \Theta(P_0, Q_0) < \infty$. Fix $k \in \mathbb{N}_0$ and set $c = \hat{H} + \frac{k}{n}\sigma$. Then,*

$$\exp\{-c(g(\mathcal{M}) + e^c \mu P_k(\mathcal{M}))\} \leq \frac{dQ_k}{dP_k}(g) \leq \exp\{c(g(\mathcal{M}) + \mu P_k(\mathcal{M}))\}$$

for all $g \in \mathcal{G}$, and

$$\exp\{-c(\Sigma \mathbf{g}(\mathcal{M}) + e^c \mu P_k(\mathcal{M}))\} \leq \frac{dQ_k^*}{dP_k^*}(\mathbf{g}) \leq \exp\{c(\Sigma \mathbf{g}(\mathcal{M}) + \mu P_k(\mathcal{M}))\}$$

for all $\mathbf{g} \in \mathcal{G}^{k+1}$.

PROOF. By Lemma 7.6 we have that $\Theta(P, Q) \leq c$, and the inequalities for dQ_k/dP_k follow from Corollary A.14.

The proof of the inequalities for dQ_k^*/dP_k^* is similar. \square

COROLLARY 7.8. *Suppose that $\hat{H} := \Theta(P_0, Q_0) < \infty$. Then, there are constants c and $\epsilon > 0$ (depending on \hat{H}) such that for any $k \leq \epsilon n$ and any $\mathbf{g} \in \mathcal{G}^{k+1}$,*

$$\left| \log \frac{dQ_k^*}{dP_k^*}(\mathbf{g}) - \log \mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} P_0 [\exp\{H(g_0|_A + X|_{A^c})\}] \right] \right| \leq c \frac{k}{n} (1 + \Sigma \mathbf{g}(\mathcal{M})).$$

PROOF. Set $H = dQ_0/dP_0$. From Lemma 7.5 and Lemma 7.3, we know that there is a constant C (independent of \mathbf{g}) such that, for the random map $\xi_{k,0} : \mathcal{A}_{k,0} \rightarrow \mathcal{G}$ introduced in Lemma 7.5,

$$\begin{aligned} & C \mathbb{E} \left[\exp \left\{ \tilde{H}_{\mathcal{A}_{k,0}}(g_0, \xi_{k,0}) - \frac{\sigma k}{n} \Sigma \mathbf{g}(\mathcal{M}) \right\} \right] \\ & \leq \frac{dQ_k^*}{dP_k^*}(\mathbf{g}) \\ & \leq C \mathbb{E} \left[\exp \left\{ \tilde{H}_{\mathcal{A}_{k,0}}(g_0, \xi_{k,0}) + \frac{\sigma k}{n} \Sigma \mathbf{g}(\mathcal{M}) \right\} \right]. \end{aligned}$$

(Note: We include the expectation with respect to the random partition $\mathcal{A}_{k,0}$ in the $\mathbb{E}[\cdot]$.) Substituting in the definition of \tilde{H} , we find that there is a constant C'

such that

$$\begin{aligned}
& C' \exp\left\{-\frac{\sigma k}{n} \Sigma \mathbf{g}(\mathcal{M})\right\} \mathbb{E}\left[\exp\left\{\sum_{A \in \mathcal{A}_{k,0}} H(g_0|_A + \xi_{k,0}(A)|_{A^c})\right\}\right] \\
& \leq \frac{dQ_k^*(\mathbf{g})}{dP_k^*(\mathbf{g})} \\
& \leq C' \exp\left\{\frac{\sigma k}{n} \Sigma \mathbf{g}(\mathcal{M})\right\} \mathbb{E}\left[\exp\left\{\sum_{A \in \mathcal{A}_{k,0}} H(g_0|_A + \xi_{k,0}(A)|_{A^c})\right\}\right].
\end{aligned}$$

Since the random measures $\{\xi_{k,0}(A) : A \in \mathcal{A}_{k,0}\}$ are i.i.d. with common distribution P_0 , this becomes

$$\begin{aligned}
& C' \mathbb{E}\left[\prod_{A \in \mathcal{A}_{k,0}} P_0\left[\exp\left\{-\frac{\sigma k}{n} \Sigma \mathbf{g}(A) + H(g_0|_A + X|_{A^c})\right\}\right]\right] \\
& \leq \frac{dQ_k^*(\mathbf{g})}{dP_k^*(\mathbf{g})} \\
& \leq C' \mathbb{E}\left[\prod_{A \in \mathcal{A}_{k,0}} P_0\left[\exp\left\{\frac{\sigma k}{n} \Sigma \mathbf{g}(A) + H(g_0|_A + X|_{A^c})\right\}\right]\right].
\end{aligned}$$

By Lemma A.13,

$$\begin{aligned}
& \left(\int \mathbb{E}\left[\prod_{A \in \mathcal{A}_{k,0}} P_0\left[\exp\left\{\frac{\sigma k}{n} \Sigma \mathbf{g}'(A) + H(g'_0|_A + X|_{A^c})\right\}\right] dP_k^*(\mathbf{g}')\right]\right)^{-1} \\
& \times \mathbb{E}\left[\prod_{A \in \mathcal{A}_{k,0}} P_0\left[\exp\left\{-\frac{\sigma k}{n} \Sigma \mathbf{g}(A) + H(g_0|_A + X|_{A^c})\right\}\right]\right] \\
& \leq \frac{dQ_k^*(\mathbf{g})}{dP_k^*(\mathbf{g})} \\
& \leq \left(\int \mathbb{E}\left[\prod_{A \in \mathcal{A}_{k,0}} P_0\left[\exp\left\{-\frac{\sigma k}{n} \Sigma \mathbf{g}'(A) + H(g'_0|_A + X|_{A^c})\right\}\right] d(P_k^*)(\mathbf{g}')\right]\right)^{-1} \\
& \times \mathbb{E}\left[\prod_{A \in \mathcal{A}_{k,0}} P_0\left[\exp\left\{\frac{\sigma k}{n} \Sigma \mathbf{g}(A) + H(g_0|_A + X|_{A^c})\right\}\right]\right].
\end{aligned}$$

We have for any set $A \in \mathcal{B}(\mathcal{M})$ that

$$\begin{aligned}
& \int P_0\left[\exp\left\{\frac{\sigma k}{n} \Sigma \mathbf{g}(A) + H(g_0|_A + X|_{A^c})\right\}\right] dP_k^*(\mathbf{g}) \\
& = \int P_0\left[\exp\left\{\frac{\sigma k}{n} (g_1 + \dots + g_k)(A) + \frac{\sigma k}{n} g_0(A) + H(g_0|_A + X|_{A^c})\right\}\right] dP_k^*(\mathbf{g}) \\
& = \int \exp\left\{\frac{\sigma k}{n} (g_1 + \dots + g_k)(A)\right\} dP_k^*(\mathbf{g}) \cdot P_0\left[\exp\left\{H(X) + \frac{\sigma k}{n} X(A)\right\}\right] \\
& = \exp\left\{\frac{\nu(A)k}{n} (e^{\sigma k/n} - 1)\right\} \cdot P_0\left[\exp\left\{H(X) + \frac{\sigma k}{n} X(A)\right\}\right]
\end{aligned}$$

because $\mathbf{g} \mapsto (g_1 + \cdots + g_k)(A)$ is a Poisson random variable with mean $\frac{k}{n}\nu(A)$ under P_k^* and $\mathbf{g} \mapsto g_0$ is a Poisson random measure with the distribution P_0 under P_k^* .

Using the inequality $e^x - 1 \leq xe^x$ for $x \geq 0$, we have

$$\begin{aligned} \exp\left\{\frac{\nu(A)k}{n}(e^{\sigma k/n} - 1)\right\} &\leq \exp\left\{\frac{\sigma\nu(A)k^2}{n^2}e^{\sigma k/n}\right\} \\ &\leq \exp\left\{\frac{c_1\nu(A)k^2}{n^2}\right\}, \end{aligned}$$

where c_1 is a constant depending on σ and T and $k/n \leq \epsilon$. By the same inequality, the fact that $P_0[e^{H(X)}] = Q_0[1] = 1$, and Corollary A.14,

$$\begin{aligned} &P_0\left[\exp\left\{H(X) + \frac{\sigma k}{n}X(A)\right\}\right] \\ &\leq P_0\left[\exp\{H(X)\}\left(1 + \frac{\sigma k}{n}X(A)\exp\left\{\frac{\sigma k}{n}X(A)\right\}\right)\right] \\ &\leq 1 + \frac{\sigma k}{n}\mu P_0(A)\exp\left\{2\mu P_0(A^c)\hat{H} + \hat{H} + \frac{\sigma k}{n} + \mu P_0(A)(e^{\hat{H} + \sigma k/n} - 1)\right\} \\ &\leq \exp\left\{c_2\frac{\mu P_0(A)k}{n}\right\}, \end{aligned}$$

where c_2 is a constant depending on σ , \hat{H} , T , $\nu(\mathcal{M})$, and $\mu P_0(\mathcal{M}) = \rho_0(\mathcal{M})$.

Thus, for any partition \mathcal{A} ,

$$\begin{aligned} &\int \prod_{A \in \mathcal{A}} P_0\left[\exp\left\{\frac{\sigma k}{n}\Sigma \mathbf{g}(A) + H(g_0|_A + X|_{A^c})\right\}\right] dP_k^*(\mathbf{g}) \\ &\leq \exp\left\{c_1\frac{\nu(\mathcal{M})k^2}{n^2} + c_2\frac{\rho_0(\mathcal{M})k}{n}\right\} \\ &\leq \exp\left\{c_3\frac{k}{n}\right\}, \end{aligned}$$

for a constant c_3 depending on σ , \hat{H} , T , $\nu(\mathcal{M})$, and $\rho_0(\mathcal{M})$.

On the other hand, using similar arguments and the inequality $e^{-x} \geq 1 - x$, $x \geq 0$, we get for any $A \in \mathcal{B}(\mathcal{M})$ that

$$\begin{aligned} &\int P_0\left[\exp\left\{-\frac{\sigma k}{n}\Sigma \mathbf{g}(A) + H(g_0|_A + X|_{A^c})\right\}\right] dP_k^*(\mathbf{g}) \\ &= \exp\left\{\frac{\nu(A)k}{n}(e^{-\sigma k/n} - 1)\right\} \cdot P_0\left[\exp\left\{H(X) - \frac{\sigma k}{n}X(A)\right\}\right] \\ &\geq \exp\left\{-c_4\frac{\nu(A)k^2}{n^2}\right\} - c_5\frac{\rho_0(A)k}{n}, \end{aligned}$$

where c_4 and c_5 are constants depending on \hat{H} , T , $\nu(\mathcal{M})$, and $\rho_0(\mathcal{M})$.

We may then find constants c_6 and $\epsilon > 0$ such that

$$\exp\left\{-c_4\frac{\nu(A)k^2}{n^2}\right\} - c_5\frac{\rho_0(A)k}{n} \geq \exp\left\{-c_6\frac{k}{n}(\nu(A) + \rho_0(A))\right\}$$

for all $k \leq \epsilon n$, and so

$$\begin{aligned} & \int \prod_{A \in \mathcal{A}} P_0 \left[\exp \left\{ -\frac{\sigma k}{n} \Sigma \mathbf{g}(A) + H(g_0|_A + X|_{A^c}) \right\} \right] dP_k^*(\mathbf{g}) \\ & \geq \exp \left\{ -c_6 \frac{k}{n} (\nu(\mathcal{M}) + \rho_0(\mathcal{M})) \right\}. \end{aligned}$$

Combining the upper and lower bounds from above, we see that there is a constant c_7 such that

$$\begin{aligned} & -c_7 \frac{k}{n} (1 + \Sigma \mathbf{g}(\mathcal{M})) + \log \mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} P_0 [\exp \{H(g_0|_A + X|_{A^c})\}] \right] \\ & \leq \log \frac{dQ_k^*}{dP_k^*}(\mathbf{g}) \\ & \leq c_7 \frac{k}{n} (1 + \Sigma \mathbf{g}(\mathcal{M})) + \log \mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} P_0 [\exp \{H(g_0|_A + X|_{A^c})\}] \right] \end{aligned}$$

for $k \leq \epsilon n$. \square

REMARK 7.9. The approximation to $\log dQ_k^*/dP_k^*$ in Corollary 7.8 involves a product over the (potentially very large) partition $\mathcal{A}_{k,0}$ and looks rather unwieldy. We may rewrite the approximation as follows to make it apparent where the non-trivial contributions to the product arise:

$$\begin{aligned} & \prod_{A \in \mathcal{A}_{k,0}} P_0 [\exp \{H(g_0|_A + X|_{A^c})\}] \\ (7.5) \quad & = \prod_{A \in \mathcal{A}_{k,0}} \frac{P_0 [\exp \{H(g_0|_A + X|_{A^c})\}]}{P_0 [\exp \{H(X|_{A^c})\}]} \times \prod_{A \in \mathcal{A}_{k,0}} \frac{P_0 [\exp \{H(X|_{A^c})\}]}{P_0 [\exp \{H(X)\}]} \\ & = \prod_{\substack{A \in \mathcal{A}_{k,0} \\ g(A) \geq 1}} \frac{P_0 [\exp \{H(g_0|_A + X|_{A^c})\}]}{P_0 [\exp \{H(X|_{A^c})\}]} \times \prod_{A \in \mathcal{A}_{k,0}} \frac{P_0 [\exp \{H(X|_{A^c})\}]}{P_0 [\exp \{H(X)\}]} \end{aligned}$$

This is a trivial consequence of the fact that $P_0[e^{H(X)}] = 1$.

7.2. Comparisons with complete Poissonization

Recall from Notation 4.1 that the Poissonization operator \mathfrak{P} acts on a probability measure P on \mathcal{G} by $\mathfrak{P}P := \Pi_{\mu P}$. Also, recall from Notation 4.4 that $O_k := (\mathfrak{P}\mathfrak{S}_n\mathfrak{M}_n)^k P_0$ 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 8.1, with the recombination operator \mathfrak{R} replaced by the complete Poissonization operator \mathfrak{P} . Finally, recall from Notation 4.13 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 $Q_k^* = (\mathfrak{R}^*\mathfrak{S}_n^*\mathfrak{M}^*)^k P_0$.

LEMMA 7.10. *Suppose that $P = \Pi_{\pi}$ for some finite measure $\pi \in \mathcal{H}^+$ and Q is an equivalent measure. Put $H := \log dQ/dP$. Set*

$$\tau(m) := \log \int_{\mathcal{G}} \exp\{H(g + \delta_m)\} dP(g).$$

- (a) *The probability measure $\mathfrak{P}Q$ has intensity measure μQ that is absolutely continuous with respect to π with Radon-Nikodym derivative*

$$\frac{d\mu\mathfrak{P}Q}{d\pi}(x) = \frac{d\mu Q}{d\pi}(x) = \exp\{\tau(x)\}.$$

- (b) *The probability measure $\mathfrak{P}Q$ is absolutely continuous with respect to P , with Radon-Nikodym derivative satisfying*

$$\begin{aligned} \log \frac{d\mathfrak{P}Q}{dP}(g) &= g[\tau] - \int_{\mathcal{M}} \left[\int_{\mathcal{G}} \exp\{H(g + \delta_x)\} dP(g) \right] d\pi(x) + \pi(\mathcal{M}) \\ &= g[\tau] - \int_{\mathcal{G}} g(\mathcal{M}) \exp\{H(g)\} dP(g) + \pi(\mathcal{M}). \end{aligned}$$

PROOF. For a Borel function $f : \mathcal{M} \rightarrow \mathbb{R}_+$,

$$\begin{aligned} \int_{\mathcal{M}} f(x) d(\mu\mathfrak{P}Q)(x) &= \int_{\mathcal{M}} f(x) d(\mu Q)(x) \\ &= \int_{\mathcal{G}} g[f] dQ(g) \\ &= \int_{\mathcal{G}} g[f] \exp\{H(g)\} dP(g) \\ &= \int_{\mathcal{M}} f(x) \exp\{\tau(x)\} d\pi(x), \end{aligned}$$

where we used Campbell's Theorem (see Proposition A.8) for the penultimate equality. Part (a) follows. It then follows immediately from Lemma A.4 that $\mathfrak{P}Q$ is absolutely continuous with respect to P , with

$$\log \frac{d\mathfrak{P}Q}{dP}(g) = g[\tau] - \int_{\mathcal{M}} \left[\int_{\mathcal{G}} \exp\{H(g + \delta_x)\} dP(g) \right] d\pi(x) + \pi(\mathcal{M}),$$

giving the first claim in part (b). The second claim in part (b) follows by another application of Campbell's Theorem. \square

LEMMA 7.11. *Suppose that $Q_0 = O_0 = P_0$.*

- (a) *For all $x \in \mathcal{M}$ and $k \in \mathbb{N}_0$,*

$$\left| \log \frac{d\mu O_k}{d\mu P_k}(x) \right| \leq \frac{\sigma k}{n}.$$

- (b) *For all $g \in \mathcal{G}$ and $k \in \mathbb{N}_0$,*

$$\left| \log \frac{dO_k}{dP_k}(g) - \log \frac{dO_k}{dP_k}(0) \right| \leq \frac{k\sigma}{n} g(\mathcal{M}).$$

- (c) *For all $g \in \mathcal{G}$ and $k \in \mathbb{N}_0$,*

$$\left| \log \frac{dO_k}{dP_k}(g) \right| \leq \frac{k\sigma}{n} \left(g(\mathcal{M}) + e^{k\sigma/n} \mu P_k(\mathcal{M}) \right).$$

PROOF. The proof is by induction. We start with $O_0 = P_0$, so $dO_0/dP_0 = 1$. Let $\pi_k := \mu P_k = \rho_0 + \frac{k}{n}\nu$, and $\pi'_k := \mu O_k$. Put $\alpha_k := \sup_{x \in \mathcal{M}} |\log d\pi'_k/d\pi_k(x)|$.

By Lemma A.4 and Lemma 7.10 we know that $d\mu\mathfrak{P}\mathfrak{S}_n O_k/d\mu P_k(x) = e^{\tau_k(x)}$, where $\tau_k(x) := \log P_k[\exp\{H(X + \delta_x)\}]$ and

$$\begin{aligned} H(g) &:= \log \frac{d\mathfrak{S}_n O_k}{dO_k}(g) + \log \frac{dO_k}{dP_k}(g) \\ &= \log \frac{\exp\{-S(g)/n\}}{O_k[\exp\{-S/n\}]} + g \left[\log \frac{d\mu O_k}{d\mu P_k} \right] - \mu P_k \left[\frac{d\mu O_k}{d\mu P_k} - 1 \right] \\ &= -\frac{S(g)}{n} - \log O_k[\exp\{-S(X)/n\}] + g \left[\log \frac{d\pi'_k}{d\pi_k} \right] - \pi_k \left[\frac{d\pi'_k}{d\pi_k} - 1 \right]. \end{aligned}$$

For any $g \in \mathcal{G}$ and $x \in \mathcal{M}$ we have

$$|H(g + \delta_x) - H(g)| \leq \frac{\sigma}{n} + \alpha_k.$$

Thus,

$$\begin{aligned} \tau_k(x) &\leq \log P_k[\exp\{H(X) + \alpha_k + \sigma/n\}] \\ &\leq \log P_k[\exp\{H(X)\}] + \alpha_k + \frac{\sigma}{n} = \alpha_k + \frac{\sigma}{n} \end{aligned}$$

and

$$\begin{aligned} \tau_k(x) &\geq \log P_k[\exp\{H(X) - \alpha_k - \sigma/n\}] \\ &\geq \log P_k[\exp\{H(X)\}] - \alpha_k - \frac{\sigma}{n} = -\alpha_k - \frac{\sigma}{n}, \end{aligned}$$

implying that $\sup |t_k(x)| \leq \alpha_k + \sigma/n$. By Lemma A.16, it follows that $\alpha_{k+1} \leq \alpha_k + \sigma/n$, and so $\alpha_k \leq k\sigma/n$ for all $k \in \mathbb{N}_0$, establishing part (a).

Applying Lemma A.4, we see that

$$\left| \log \frac{dO_k}{dP_k}(g) - \log \frac{dO_k}{dP_k}(0) \right| \leq \alpha_k g(\mathcal{M}) \leq \frac{k\sigma}{n} g(\mathcal{M})$$

and

$$\left| \log \frac{dO_k}{dP_k}(g) \right| \leq \alpha_k g(\mathcal{M}) + (e^{\alpha_k} - 1) \pi_k(\mathcal{M}) \leq \frac{k\sigma}{n} \left(g(\mathcal{M}) + e^{k\sigma/n} \mu P_k(\mathcal{M}) \right),$$

establishing parts (b) and (c). \square

We now extend Lemma 7.10 to probability measures on \mathcal{G}^* . Recall that a $(k+1)$ -tuple of independent Poisson random measures on \mathcal{M} with intensities π_0, \dots, π_k , may, as discussed in Section 4.4, be thought of as a single Poisson random measure on the set $\mathcal{M} \times \{0, \dots, k\}$ with intensity $\sum \pi_i \otimes \delta_{\{i\}}$.

COROLLARY 7.12. *Suppose that $P^* := \Pi_{\pi_0} \otimes \dots \otimes \Pi_{\pi_k}$ for $\pi_0, \dots, \pi_k \in \mathcal{H}^+$ and Q^* is any probability measure on \mathcal{G}^{k+1} such that $\mu\Psi_j Q^*$ is absolutely continuous with respect to $\mu\Psi_j P^* = \pi_j$ for $0 \leq j \leq k$. Then, $\mathfrak{P}^* Q^*$ is absolutely continuous with respect to P^* , with Radon-Nikodym derivative satisfying*

$$\log \frac{d\mathfrak{P}^* Q^*}{dP^*}(\mathbf{g}) = \sum_{i=0}^k g_i[\tau^{(i)}] - \sum_{i=0}^k \pi_i \left[\exp\{\tau^{(i)}\} - 1 \right]$$

where

$$\tau^{(i)}(x) := \log \int_{\mathcal{G}^{k+1}} \frac{dQ^*}{dP^*}(h_0, \dots, h_{i-1}, h_i + \delta_x, \dots, h_k) dP^*(h_0, \dots, h_k).$$

PROOF. By definition $\mathfrak{P}^*Q^* = \mathfrak{P}\Psi_0Q^* \otimes \cdots \otimes \mathfrak{P}\Psi_kQ^*$ with $\mathfrak{P}\Psi_jQ^* = \Pi_{\mu\Psi_jQ^*}$ for $0 \leq j \leq k$. Thus,

$$\log \frac{d\mathfrak{P}^*Q^*}{dP^*}(g_0, \dots, g_k) = \sum_{j=0}^k \log \frac{d\mathfrak{P}\Psi_jQ^*}{d\Psi_jP^*}(g_j)$$

and the result follows then directly from Lemma 7.10. \square

LEMMA 7.13. *Suppose that $\hat{H} := \Theta(P_0, Q_0) < \infty$. Set $H := \log dQ_0/dP_0$. Let $\tau^{(i)}$, $0 \leq i \leq k$, be as in Corollary 7.12 with $Q^* = Q_k^*$ and $P^* = P_k^*$. Then, there are positive constants ϵ and c , depending on \hat{H} but not depending on k or n , such that*

$$\left| \tau^{(0)}(x) - \log P_0[\exp\{H(X + \delta_x)\}] \right| \leq c \frac{k}{n}$$

and

$$\left| \tau^{(i)}(x) \right| \leq c \frac{k}{n}$$

for $1 \leq i \leq k \leq n\epsilon$.

PROOF. We apply Corollary 7.8. Since the Poisson random measures $\xi_{k,0}(A)$ are independent as A ranges over the random partition $\mathcal{A}_{k,0}$, there are constants c_1 and ϵ such that for $k \leq \epsilon n$,

$$\begin{aligned} & C e^{-c_1 k(1+\Sigma \mathbf{g}(\mathcal{M}))/n} \mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} P_0[\exp\{H(g_0|_A + X|_{A^c})\}] \right] \\ & \leq \frac{dQ_k^*}{dP_k^*}(\mathbf{g}) \\ & \leq C e^{c_1 k(1+\Sigma \mathbf{g}(\mathcal{M}))/n} \mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} P_0[\exp\{H(g_0|_A + X|_{A^c})\}] \right], \end{aligned}$$

where C does not depend on \mathbf{g} , but may depend on k and n . We recombine terms to obtain

$$\begin{aligned} & C \exp\left\{-c_1 \frac{k}{n} \left(1 + \sum_{i=1}^k g_i(\mathcal{M})\right)\right\} \mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} P_0 \left[\exp \left\{ H(g_0|_A + X|_{A^c}) - \frac{c_1 k}{n} g_0(A) \right\} \right] \right] \\ & \leq \frac{dQ_k^*}{dP_k^*}(\mathbf{g}) \\ & \leq C \exp\left\{c_1 \frac{k}{n} \left(1 + \sum_{i=1}^k g_i(\mathcal{M})\right)\right\} \mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} P_0 \left[\exp \left\{ H(g_0|_A + X|_{A^c}) + \frac{c_1 k}{n} g_0(A) \right\} \right] \right]. \end{aligned}$$

Note that $\mathbf{g} \mapsto \sum_{i=1}^k g_i(\mathcal{M})$ and $\mathbf{g} \mapsto g_0$ are independent under P_k^* , with the former having a Poisson distribution with parameter $k\nu(\mathcal{M})/n$. A bound on the

constant C can be imported from Lemma A.13, leading to the following inequality

$$\begin{aligned}
\tau^{(0)}(x) &= \log \int \frac{dQ_k^*}{dP_k^*}(g_0 + \delta_x, \dots, g_k) dP_k^*(\mathbf{g}) \\
&\leq \log \mathbb{E} \left[\int \exp \left\{ c_1 \frac{k}{n} \left(1 + \sum_{i=1}^k g_i(\mathcal{M}) \right) \right\} \right. \\
&\quad \times \prod_{A \in \mathcal{A}_{k,0}} P_0 \left[\exp \left\{ H((g_0 + \delta_x)|_A + X|_{A^c}) + \frac{c_1 k}{n} (g_0 + \delta_x)(A) \right\} \right] dP_k^*(\mathbf{g}) \\
&\quad - \log \mathbb{E} \left[\int \exp \left\{ -c_1 \frac{k}{n} \left(1 + \sum_{i=1}^k g_i(\mathcal{M}) \right) \right\} \right. \\
&\quad \times \prod_{A \in \mathcal{A}_{k,0}} P_0 \left[\exp \left\{ H(g_0|_A + X|_{A^c}) - \frac{c_1 k}{n} g_0(A) \right\} \right] dP_k^*(\mathbf{g}) \\
&= \frac{3c_1 k}{n} + 2 \frac{k\nu(\mathcal{M})}{n} (e^{c_1 k/n} - 1) \\
&\quad + \log \mathbb{E} \left[\int \prod_{A \in \mathcal{A}_{k,0}} P_0 \left[\exp \left\{ H((g' + \delta_x)|_A + X|_{A^c}) + \frac{c_1 k}{n} g'(A) \right\} \right] dP_0(g') \right] \\
&\quad - \log \mathbb{E} \left[\int \prod_{A \in \mathcal{A}_{k,0}} P_0 \left[\exp \left\{ H(g'|_A + X|_{A^c}) - \frac{c_1 k}{n} g'(A) \right\} \right] dP_0(g') \right].
\end{aligned}$$

Applying Corollary A.7, we see that there is a constant c_2 such that

$$\begin{aligned}
\tau^{(0)}(x) &\leq \frac{c_2 k}{n} + \log \mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} P_0 \left[\exp \left\{ H(X + (\delta_x)_A) + \frac{c_1 k}{n} X(A) \right\} \right] \right. \\
&\quad \left. \times P_0 \left[\exp \left\{ H(X) - \frac{c_1 k}{n} X(A) \right\} \right]^{-1} \right] \\
&\leq \frac{c_2 k}{n} + \log \mathbb{E} \left[P_0 \left[\exp \left\{ H(X + \delta_x) + \frac{c_1 k}{n} X(A(x)) \right\} \right] \right. \\
&\quad \left. \times \prod_{A \in \mathcal{A}_{k,0}} \frac{P_0 [\exp\{H(X) + c_1 k X(A)/n\}]}{P_0 [\exp\{H(X) - c_1 k X(A)/n\}]} \right],
\end{aligned}$$

where $A(x)$ denotes the element of $\mathcal{A}_{k,0}$ containing $x \in \mathcal{M}$.

By Lemma A.15 there are constants ϵ_2 and c_3 such that for $k/n \leq \epsilon_2$ the product is bounded by

$$\exp \left\{ \sum_{A \in \mathcal{A}_{k,0}} c_3 \frac{2c_1 k}{n} \rho_0(A) \right\} \leq \exp \left\{ 2c_1 c_3 \frac{k}{n} \rho_0(\mathcal{M}) \right\}.$$

Thus, there is a constant c_4 such that for $k \leq \epsilon' n$, where $\epsilon' = \epsilon \wedge \epsilon_2$,

$$(7.6) \quad \tau^{(0)}(x) \leq \frac{c_4 k}{n} + \log \mathbb{E} \left[P_0 \left[\exp \left\{ H(X + \delta_x) + \frac{c_1 k}{n} X(A(x)) \right\} \right] \right].$$

We proceed by writing

$$\begin{aligned}
(7.7) \quad & P_0 \left[\exp \left\{ H(X + \delta_x) + \frac{c_1 k}{n} X(A(x)) \right\} \right] \\
&= P_0 [\exp\{H(X + \delta_x)\}] \\
&\quad \times \left(1 + \frac{P_0 [\exp\{H(X + \delta_x)\}] (\exp\{c_1 k X(A(x))/n\} - 1)}{P_0 [\exp\{H(X + \delta_x)\}]} \right).
\end{aligned}$$

Using again the relation $e^x - 1 \leq |x|e^{|x|}$, and the fact that P_0 is Poisson, hence the restrictions of X to disjoint sets are independent under P_0 , the expectation in the numerator above may be rewritten as

$$\begin{aligned}
& P_0 \left[\exp \left\{ H(X|_{A(x)^c}) \right\} \cdot \exp \left\{ H(X + \delta_x) - H(X|_{A(x)^c}) \right\} (\exp\{c_1 k X(A(x))/n\} - 1) \right] \\
&\leq P_0 \left[\exp \left\{ H(X|_{A(x)^c}) \right\} \right] \cdot P_0 \left[\frac{c_1 k X(A(x))}{n} \exp \left\{ \left(\hat{H} + c_1 \frac{k}{n} \right) \cdot X(A(x)) + \hat{H} \right\} \right]. \\
&\leq \frac{c_1 k}{n} P_0 \left[\exp \left\{ H(X|_{A(x)^c}) \right\} \right] \cdot \rho_0(A(x)) c_5 e^{(c_5 - 1)\rho_0(A(x))}.
\end{aligned}$$

where $c_5 := \exp \left\{ \hat{H} + c_1 T \right\}$.

For the denominator, we have

$$P_0 [\exp\{H(X + \delta_x)\}] \geq P_0 \left[\exp \left\{ H(X|_{A(x)^c}) \right\} \right] \cdot \exp \left\{ -\hat{H}(\rho_0(A(x)) + 1) \right\}.$$

Combining these bounds and using the inequality $\log(1 + x) \leq x$ for $x \geq 0$, we may find a constant c_6 (depending on $\rho_0(\mathcal{M})$ and \hat{H}) such that

$$\log \left(1 + \frac{P_0 [\exp\{H(X + \delta_x)\}] (\exp\{c_1 k X(A(x))/n\} - 1)}{P_0 [\exp\{H(X + \delta_x)\}]} \right) \leq c_6 \rho_0(A(x)) \frac{k}{n}.$$

Combining this with (7.6) and (7.7) gives us a constant c_7 such that

$$(7.8) \quad \tau^{(0)}(x) \leq \frac{c_7 k}{n} + \log P_0 [\exp\{H(X + \delta_x)\}]$$

for $k/n \leq \epsilon'$. Nearly identical calculation give us a lower bound, completing the proof of the first claim.

The second claim follows from similar, but simpler, arguments. \square

COROLLARY 7.14. *Under the hypotheses of Lemma 7.13,*

$$\left| \log \frac{d\mathfrak{P}^* Q_k^*}{dP_k^*}(\mathbf{g}) - \int \log P_0 [\exp\{H(X + \delta_x)\}] dg_0(x) \right| \leq \frac{ck}{n} (1 + \Sigma \mathbf{g}).$$

PROOF. By Corollary 7.12, the quantity on the left-hand side of the claimed inequality is at most

$$\begin{aligned}
& \left| \int \tau^{(0)}(m) dg_0(m) - \int \log P_0 [\exp\{H(X + \delta_m)\}] dg_0(m) \right| \\
& + \left| \sum_{j=1}^k \int \tau^{(j)}(m) dg_j(m) \right| \\
(7.9) \quad & + \left| (-1) \int (\exp\{\tau^{(0)}(y)\} - 1) d\pi_0(y) \right| \\
& + \left| (-1) \sum_{j=1}^k \int \exp\{\tau^{(j)}(y)\} d\pi_j(y) \right| \\
& + k\nu(\mathcal{M})/n.
\end{aligned}$$

We bound the quantities on each of the five lines of (7.9) separately. For the first line, the first part of Lemma 7.13 provides a bound of $(ck/n)g_0(\mathcal{M})$. For the second line, the second part of Lemma 7.13 provides a bound of $(ck/n) \sum_{j=1}^k g_j(\mathcal{M})$. For the third line, the same lemma lets us write

$$P_0 [\exp\{H(X + \delta_y)\}] e^{-ck/n} \leq \exp\{\tau^{(0)}(y)\} \leq P_0 [\exp\{H(X + \delta_y)\}] e^{ck/n}.$$

By Campbell's Theorem A.8, the integral with respect to $\pi_0(dy)$ of the left-hand side of this last expression is $\pi_0(\mathcal{M}) \exp\{-ck/n\}$ and the corresponding integral on the right-hand side is $\pi_0(\mathcal{M}) \exp\{+ck/n\}$. Subtracting $\pi_0(\mathcal{M})$ and making use once more of the inequality $|e^x - 1| \leq |x|e^{|x|}$, we conclude that the quantity on the third line is bounded by $\pi_0(\mathcal{M})(ck/n)e^{ce}$ when $k/n \leq \epsilon$.

For the fourth line, again by Campbell's Theorem, the integral is the sum of the total intensities of g_1 to g_k with respect to Q_k^* , that is to say,

$$\int (g_1(\mathcal{M}) + \dots + g_k(\mathcal{M})) \frac{dQ_k^*}{dP_k^*} dP_k^*(\mathbf{g}).$$

By Corollary 7.7, this nonnegative quantity is bounded above by

$$\int (g_1(\mathcal{M}) + \dots + g_k(\mathcal{M})) \exp(cg_0(\mathcal{M}) + \dots + cg_k(\mathcal{M}) + c\mu P_k(\mathcal{M})) dP_k^*(\mathbf{g}).$$

Under P_k^* , the random measures $\mathbf{g} \mapsto g_0$ and $\mathbf{g} \mapsto g_1 + \dots + g_k$ are independent Poisson random measures with respective intensity measures ρ_0 and $k\nu/n$, and so

$$\exp\{\pi_0(\mathcal{M})(e^c - 1 + c)\} \times \exp\{c + (k/n)\nu(\mathcal{M})(e^c - 1 + c)\} \times \nu(\mathcal{M})k/n,$$

or, in effect, a constant multiple of k/n . The fifth line is already so bounded. Gathering terms and setting a new constant equal to the maximum of the constants from all the lines, the bound in the corollary is established. \square

LEMMA 7.15. *Suppose that $\hat{H} := \Theta(P_0, Q_0) < \infty$ and and the pair $(\mathcal{R}, \mu P_0)$ is shattering with constant α . Then, there exists a constant c (depending on T , α , $\rho_0(\mathcal{M}) = \mu Q_0(\mathcal{M}) = \mu P_0(\mathcal{M})$, and \hat{H}), such that*

$$\int \int \left| \frac{d\mathfrak{P}^* Q_k^*}{dQ_k^*}(\mathbf{g}) - 1 \right| dQ_k^*(\mathbf{g}) \leq c \left(\frac{1}{k+1} \vee \frac{k}{n} \right).$$

PROOF. Combining Corollary 7.8 (in the form of (7.5)), with Corollary 7.14, we see that there are positive constants ϵ and c_1 , depending on \hat{H} , such that for $k/n \leq \epsilon$,

$$\begin{aligned}
(7.10) \quad & -\log \frac{d\mathfrak{P}^* Q_k^*}{dQ_k^*}(\mathbf{g}) = \log \frac{dQ_k^*}{dP_k^*}(\mathbf{g}) - \log \frac{d\mathfrak{P}^* Q_k^*}{dP_k^*}(\mathbf{g}) \\
& = \left[\int \log P_0 [\exp\{H(X + \delta_x)\}] dg_0(x) \right. \\
& \quad \left. - \log \mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} \frac{P_0 [\exp\{H(g_0|_A + X|_{A^c})\}]}{P_0 [\exp\{H(X|_{A^c})\}]} \right] + C + \gamma(\mathbf{g}) \frac{k}{n} \Sigma \mathbf{g}(\mathcal{M}), \right]
\end{aligned}$$

where $|\gamma(\mathbf{g})| \leq c_1$ and

$$C = -\log \mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} P_0 [\exp(H(X|_{A^c}))] \right]$$

is a quantity not depending on \mathbf{g} .

Observe that the product in the second term of (7.10) is universally bounded above and below as follows

$$(7.11) \quad \exp\{-\hat{H}g_0(\mathcal{M})\} \leq \prod_{A \in \mathcal{A}_{k,0}} \frac{P_0 [\exp\{H((g_0)_A + X_{A^c})\}]}{P_0 [\exp\{H(X_{A^c})\}]} \leq \exp\{\hat{H}g_0(\mathcal{M})\}.$$

We decompose the product as

$$\begin{aligned}
(7.12) \quad & \mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} \frac{P_0 [\exp\{H(g_0|_A + X|_{A^c})\}]}{P_0 [\exp\{H(X|_{A^c})\}]} \right] \\
& = \mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} \frac{P_0 [\exp\{H(g_0|_A + X|_{A^c})\}]}{P_0 [\exp\{H(X|_{A^c})\}]} \mathbf{1} \left\{ \bigvee_{A \in \mathcal{A}_{k,0}} g_0(A) \leq 1 \right\} \right] \\
& \quad + \mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} \frac{P_0 [\exp\{H(g_0|_A + X|_{A^c})\}]}{P_0 [\exp\{H(X|_{A^c})\}]} \mathbf{1} \left\{ \bigvee_{A \in \mathcal{A}_{k,0}} g_0(A) \geq 2 \right\} \right].
\end{aligned}$$

The logarithm of the product in the first expectation on the right is

$$\begin{aligned}
(7.13) \quad & \left(\log \prod_{A \in \mathcal{A}_{k,0}} \frac{P_0 [\exp\{H(g_0|_A + X|_{A^c})\}]}{P_0 [\exp\{H(X|_{A^c})\}]} \right) \mathbf{1} \left\{ \bigvee_{A \in \mathcal{A}_{k,0}} g_0(A) \leq 1 \right\} \\
& = \int \log \frac{P_0 [\exp\{H(\delta_x + X|_{A(x)^c})\}]}{P_0 [\exp\{H(X|_{A^c})\}]} dg_0(x) \\
& = \int \log P_0 [\exp\{H(X + \delta_x)\}] dg_0(x) \\
& \quad + \int \log \frac{P_0 [\exp\{H(\delta_x + X|_{A(x)^c})\}]}{P_0 [\exp\{H(\delta_x + X)\}]} dg_0(x) \\
& \quad - \int \log \frac{P_0 [\exp\{H(X|_{A(x)^c})\}]}{P_0 [\exp\{H(X)\}]} dg_0(x).
\end{aligned}$$

We next find bounds for the second and third terms on the right of (7.13). For any set $A \in \mathcal{B}(\mathcal{M})$,

$$\begin{aligned}
P_0 [\exp\{H(\delta_x + X)\}] &= P_0 [\exp\{H(X|_{A^c} + \delta_x)\} \exp\{H(X + \delta_x) - H(X|_{A^c} + \delta_x)\}] \\
&\leq P_0 [\exp\{H(X|_{A^c} + \delta_x)\} e^{\hat{H}X(A)}] \\
&= P_0 [\exp\{H(X|_{A^c} + \delta_x)\}] \cdot P_0 [\exp\{\hat{H}X(A)\}] \\
&= P_0 [\exp\{H(X|_{A^c} + \delta_x)\}] \cdot \exp\left\{\rho_0(A) \left(e^{\hat{H}} - 1\right)\right\} \\
&\leq P_0 [\exp\{H(X|_{A^c} + \delta_x)\}] \cdot \exp\{\rho_0(A)\check{H}\},
\end{aligned}$$

where $\check{H} := \hat{H}e^{\hat{H}}$. A similar calculation gives an analogous lower bound for the second term on the right of (7.13), yielding

$$\left| \int \log \frac{P_0 [\exp\{H(\delta_x + X|_{A(x)^c})\}]}{P_0 [\exp\{H(\delta_x + X)\}]} dg_0(x) \right| \leq \check{H} \int \rho_0(A(x)) dg_0(x).$$

The third term on the right of (7.13) may be bounded in the same way. Thus,

$$\begin{aligned}
&\left| \log \prod_{A \in \mathcal{A}_{k,0}} \frac{P_0 [\exp\{H((g_0)_A + X_{A^c})\}]}{P_0 [\exp\{H(X_{A^c})\}]} \mathbf{1} \left\{ \bigvee_{A \in \mathcal{A}_{k,0}} g_0(A) \leq 1 \right\} \right. \\
&\quad \left. - \int \log P_0 [\exp\{H(X + \delta_x)\}] dg_0(x) \right| \\
&\leq 2\check{H} \int \rho_0(A(x)) dg_0(x).
\end{aligned}$$

Equivalently,

$$\begin{aligned}
&\mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} \frac{P_0 [\exp\{H((g_0)_A + X_{A^c})\}]}{P_0 [\exp\{H(X_{A^c})\}]} \mathbf{1} \left\{ \bigvee_{A \in \mathcal{A}_{k,0}} g_0(A) \leq 1 \right\} \right] \\
&= \exp \left\{ \int \log P_0 [\exp\{H(X + \delta_x)\}] dg_0(x) \right\} \\
&\quad \times \mathbb{E} \left[\exp\{R(g_0, \mathcal{A}_{k,0})\} \mathbf{1} \left\{ \bigvee_{A \in \mathcal{A}_{k,0}} g_0(A) \leq 1 \right\} \right],
\end{aligned}$$

where $R(g, \mathcal{A}_{k,0})$ is a random variable satisfying

$$|R(\mathbf{g}, \mathcal{A}_{k,0})| \leq 2\check{H} \sum_{\substack{A \in \mathcal{A}_{k,0} \\ g_0(A)=1}} \rho_0(A).$$

We see from Lemma A.3 that

$$\begin{aligned} -\mathbb{E} \left[2\check{H} \sum_{\substack{A \in \mathcal{A}_{k,0}: \\ g_0(A)=1}} \rho_0(A) \right] &\leq \log \mathbb{E} \left[\exp \left\{ 2\check{H} \sum_{\substack{A \in \mathcal{A}_{k,0}: \\ g_0(A)=1}} \rho_0(A) \right\} \right] \\ &\leq 2\check{H} e^{2\check{H}\rho_0(\mathcal{M})} \mathbb{E} \left[\sum_{\substack{A \in \mathcal{A}_{k,0}: \\ g_0(A)=1}} \rho_0(A) \right]. \end{aligned}$$

Combining this with (7.11) and (7.12) we may find constants c_2, c_3 (depending on \hat{H} and $\rho_0(\mathcal{M})$) such that

$$\begin{aligned} \left| \int \log P_0 \left[\exp \{ H(X + \delta_x) \} \right] dg_0(x) - \log \mathbb{E} \left[\prod_{A \in \mathcal{A}_{k,0}} \frac{P_0 \left[\exp \{ H(g_0|_A + X|_{A^c}) \}} \right]}{P_0 \left[\exp \{ H(X|_{A^c}) \}} \right]} \right] \right| \\ \leq c_2 \mathbb{E} \left[\sum_{\substack{A \in \mathcal{A}_{k,0}: \\ g_0(A)=1}} \rho_0(A) \right] + c_3 \mathbb{P} \left\{ \bigvee_{A \in \mathcal{A}_{k,0}} g_0(A) \geq 2 \right\}. \end{aligned}$$

We may now apply part (b) of Lemma A.13 with

$$f_2(\mathbf{g}) := \gamma(\mathbf{g}) \frac{k}{n} \Sigma \mathbf{g}(\mathcal{M}) + c_2 \mathbb{E} \left[\sum_{\substack{A \in \mathcal{A}_{k,0}: \\ g_0(A)=1}} \rho_0(A) \right] + c_3 \mathbb{P} \left\{ \bigvee_{A \in \mathcal{A}_{k,0}} g_0(A) \geq 2 \right\},$$

and $f_1(\mathbf{g}) = 0$, using the fact that $f_2(\mathbf{g})$ is uniformly bounded by a constant c_4 , to obtain

$$\int \left| \frac{d\mathfrak{P}^* Q_k^*}{dQ_k^*}(\mathbf{g}) - 1 \right| dQ_k^*(\mathbf{g}) \leq 2e^{2c_4} Q_k^* [f_2(X)].$$

By Corollary 7.7 there are positive constants c_5 and c_6 such that

$$\frac{dQ_k^*}{dP_k^*}(\mathbf{g}) \leq c_6 e^{c_5 \Sigma \mathbf{g}(\mathcal{M})}.$$

Thus,

(7.14)

$$\begin{aligned} Q_k^* [f_2(X)] &\leq c_7 \left(\frac{k}{n} P_k [X(\mathcal{M})] + \int e^{c_5 g'(\mathcal{M})} \mathbb{E} \left[\sum_{\substack{A \in \mathcal{A}_{k,0}: \\ g'(A)=1}} \rho_0(A) \right] dP_0(g') \right. \\ &\quad \left. + \int e^{c_5 g'(\mathcal{M})} \mathbb{P} \left\{ \bigvee_{A \in \mathcal{A}_{k,0}} g'(A) \geq 2 \right\} dP_0(g') \right) \end{aligned}$$

for a constant c_7 . Note that $\Sigma \mathbf{g} = g_0 + (g_1 + \cdots + g_k)$ and $\mathbf{g} \mapsto g_1 + \cdots + g_k$ is independent of $\mathbf{g} \mapsto g_0$ under P_k^* , and so the resulting integral is simply included in c_7 .

We may bound the indicator of the event $\{\bigvee_A g'(A) \geq 2\}$ from above by the sum $\sum_{A \in \mathcal{A}_{k,0}} \mathbf{1}\{g'(A) \geq 2\}$. Exchanging the order of the integrals in the second

and third terms, we then obtain

$$\begin{aligned}
& \int e^{c_5 g'(\mathcal{M})} \mathbb{E} \left[\sum_{\substack{A \in \mathcal{A}_{k,0}: \\ g'(A)=1}} \rho_0(A) \right] dP_0(g') + \int e^{c_5 g'(\mathcal{M})} \mathbb{P} \left\{ \bigvee_{A \in \mathcal{A}_{k,0}} g'(A) \geq 2 \right\} dP_0(g') \\
& \leq \mathbb{E} \left[\sum_{A \in \mathcal{A}_{k,0}} \int e^{c_5 g'(\mathcal{M})} \left(\mathbf{1}_{\{g'(A) \geq 1\}} \rho_0(A) + \mathbf{1}_{\{g'(A) \geq 2\}} \right) dP_0(g') \right] \\
& = \mathbb{E} \left[\sum_{A \in \mathcal{A}_{k,0}} e^{\rho_0(A^{\mathbb{G}})(e^{c_5}-1)} \int e^{c_5 g'(A)} \left(\mathbf{1}_{\{g'(A) \geq 1\}} \rho_0(A) + \mathbf{1}_{\{g'(A) \geq 2\}} \right) dP_0(g') \right] \\
& \leq \mathbb{E} \left[\sum_{A \in \mathcal{A}_{k,0}} e^{\rho_0(A^{\mathbb{G}})(e^{c_5}-1)} (\rho_0(A) e^{c_5}) \exp\{(e^{c_5}-1)\rho_0(A)\} \rho_0(A) \right. \\
& \quad \left. + \frac{1}{2} (\rho_0(A) e^{c_5})^2 \exp\{(e^{c_5}-1)\rho_0(A)\} \right] \\
& = \mathbb{E} \left[e^{\rho_0(\mathcal{M})(e^{c_5}-1)} \sum_{A \in \mathcal{A}_{k,0}} \rho_0(A)^2 (e^{c_5} + e^{2c_5}/2) \right].
\end{aligned}$$

This allows us to conclude for some constant c_8 that

$$Q_k^*[f_2(X)] \leq c_7 \frac{k}{n} P_k[X(\mathcal{M})] + c_8 \mathbb{E} \left[\sum_{A \in \mathcal{A}_{k,0}} \rho_0(A)^2 \right].$$

We apply Lemma 5.7 to the second term on the right-hand side and find a constant c_9 , which now depends on the shattering constant α , such that

$$Q_k^*[f_2(X)] \leq c_7 \frac{k}{n} P_k[X(\mathcal{M})] + \frac{c_9}{k+1}.$$

We also have $P_k[X(\mathcal{M})]$ equal to $(\rho_0(\mathcal{M}) + k\nu(\mathcal{M})/n)$, which is bounded by a constant for $k/n \leq T$. It follows that

$$\int \left| \frac{d\mathfrak{P}^* Q_k^*}{dQ_k^*}(\mathbf{g}) - 1 \right| dQ_k^*(\mathbf{g}) \leq \frac{c_{10}k}{n} + \frac{c_{11}}{k+1},$$

completing the proof. \square

Convergence of the discrete-generation system

8.1. Outline of the proof

We have now assembled all the ingredients for our proof of our main convergence result previewed in Chapter 5. This result establishes the convergence of our discrete-generation dynamical system to our continuous-time dynamical system as the scaling parameter n goes to infinity. This parameter n governs the rate of recombination relative to rates of selection and mutation. The overall strategy behind the proof has been described at the end of Chapter 5. Here we outline some more detailed considerations that enter into the proof.

Thanks to the result on complete Poissonization proved in Chapter 6, the task that remains involves only two Wasserstein distances which must be bounded uniformly with bounds that go to zero as n goes to infinity. One is the distance within the Poisson family between the probability measures O_k and $\mathfrak{P}Q_k$. The other is the distance between the Poisson family and the discrete-generation system Q_k itself, gauged by the distance between Q_k and its own Poissonization $\mathfrak{P}Q_k$.

The distance within the Poisson family can be bounded by the Wasserstein distance between the corresponding intensity measures. A critical observation, encapsulated in the inequality (8.3), is that the intensity measures deviate from each other appreciably only when Q_k deviates from Poisson. This reduces the problem to bounding the distance from the Poisson family, the distance between Q_k and $\mathfrak{P}Q_k$.

This problem is addressed in two parts, by parts requiring results established in Chapter 7. First we adduce the formula for the Radon-Nikodym derivative of Q_k with respect to P_k in Lemma 7.5. This formula is exact, in principle, but is probably impossible to compute exactly in most cases. Next we rely on two approximations derived in subsequent lemmas.

- Uniformly on $k \leq Tn$, there is a bound on $\Theta(P_0, Q_0)$.
- For $k \ll n$, there is a bound on $\|\mathfrak{P}Q_k - Q_k\|_{\text{Was}}$ of the form $c(1/k) \vee (k/n)$, where the constant c depends on $\Theta(P_0, Q_0)$.

We bring these consequences to bear by noting that there is no reason why we need to start at the original P_0 and Q_0 . Since $\Theta(P_L, Q_L)$ is bounded for any $L \leq Tn$, we have a bound $\|\mathfrak{P}Q_{L+\sqrt{n}} - Q_{L+\sqrt{n}}\|_{\text{Was}} \leq c/\sqrt{n}$ for any L .

This approach covers all k in the desired range, except in a small neighborhood of 0. If the initial probability measure Q_0 is Poisson (that is, $Q_0 = P_0$), then we have a bound $\|\mathfrak{P}Q_k - Q_k\|_{\text{Was}} \leq ck/n$ for small k , and this covers the gap at 0. If Q_0 is not Poisson, then there is a failure of uniform convergence in a neighborhood of 0: as $n \rightarrow \infty$, the non-Poisson probability measure at time 0 jumps immediately to its Poissonization at time $0+$.

8.2. The convergence theorem

THEOREM 8.1. *Let $(\rho_t)_{t \geq 0}$ be the measure-valued dynamical system of (2.10) whose existence is guaranteed by Theorem 2.4. Suppose that the selective cost function S satisfies the hypotheses of Theorem 2.4, namely*

- $S(0) = 0$,
- $S(g) \leq S(g + h)$ for all $g, h \in \mathcal{G}$,
- for some constant σ , $|S(g) - S(h)| \leq \sigma \|g - h\|_{\text{Was}}$, for all $g, h \in \mathcal{G}$.

In addition, suppose that the following assumptions are in force.

- *The pair (\mathcal{R}, ν) consisting of the recombination measure and the mutation measure is shattering.*
- *The pair (\mathcal{R}, ρ_0) consisting of the recombination measure and the initial intensity is shattering.*
- *The initial measure Q_0 is equivalent to its Poissonization $P_0 := \mathfrak{P}Q_0 = \Pi_{\rho_0}$, and $\Theta(P_0, Q_0) < \infty$.*

Then, for any $T > \epsilon > 0$,

$$\lim_{n \rightarrow \infty} \sup_{\epsilon \leq t \leq T} \|\Pi_{\rho_t} - Q_{\lfloor tn \rfloor}\|_{\text{Was}} = 0.$$

If, in addition, the initial measure $Q_0 = P_0$ is Poisson, then this equation holds for $\epsilon = 0$.

PROOF. The proof proceeds by establishing a bound for the total distance $\|\Pi_{\rho_t} - Q_k\|_{\text{Was}}$, a bound uniform across generations k in the set $\{k \in \mathbb{N}_0 : \lfloor \epsilon n \rfloor \leq k \leq \lfloor Tn \rfloor\}$ which goes to zero as the scaling parameter n goes to infinity. By the triangle inequality we have

$$(8.1) \quad \|\Pi_{\rho_t} - Q_k\|_{\text{Was}} \leq \|\Pi_{\rho_t} - O_k\|_{\text{Was}} + \|O_k - \mathfrak{P}Q_k\|_{\text{Was}} + \|\mathfrak{P}Q_k - Q_k\|_{\text{Was}}$$

The first of the three distances on the right-hand side has already been shown in Theorem 6.3 to be bounded by $1/n$ times a constant depending on T . In other words, the measure describing the dynamical system is close to the measure $O_k = (\mathfrak{M}_n \mathfrak{S}_n)^k Q_0$, the Poisson probability measure that describes a population after k rounds of selection, mutation, and complete Poissonization.

What remains is to bound the second and third of the distance terms. We need to show that the Poisson measure O_k is close to the Poisson measure $\mathfrak{P}Q_k$, and that $\mathfrak{P}Q_k$ is close to Q_k itself, where $Q_k = (\mathfrak{M}_n \mathfrak{S}_n)^k Q_0$ is the probability measure that describes a population after k rounds of selection, mutation, and recombination.

By Lemma A.12 the second term on the right in (8.1) is bounded by a multiple of the Wasserstein distance between the intensities μO_k and μQ_k (the latter of which is the same as $\mu \mathfrak{P}Q_k$):

$$(8.2) \quad \|O_k - \mathfrak{P}Q_k\|_{\text{Was}} \leq 4 \|\mu O_k - \mu Q_k\|_{\text{Was}}$$

We introduce symbols for the two terms for which we are now seeking bounds.

Set

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

and

$$b_k := \|\mu Q_k - \mu O_k\|_{\text{Was}}.$$

We treat b_k first, eventually obtaining a recursion equation for it that turns out to involve a_k .

Given a Borel function $f : \mathfrak{M} \rightarrow [-1, 1]$ with $\|f\|_{\text{Lip}} \leq 1$, define the function $F : \mathcal{G} \rightarrow \mathbb{R}$ by $F(g) := g[f] = \int f(m) dg(m)$. The Lipschitz condition on f implies

$$|F(g') - F(g'')| = \left| \int_{\mathcal{M}} f(m) dg'(m) - \int_{\mathcal{M}} f(m) dg''(m) \right| \leq \|g' - g''\|_{\text{Was}}.$$

By definition, the integral $\mu Q[f]$ is the expected value $Q[F]$ for any probability measure Q .

By definition of the family $(Q_k)_{k \geq 0}$,

$$\mu Q_{k+1}[f] = \mu(\mathfrak{R}\mathfrak{M}_n\mathfrak{S}_n Q_k)[f].$$

The recombination operator \mathfrak{R} leaves intensities unchanged, and the mutation operator \mathfrak{M}_n adds the measure ν/n to the intensity of any probability measure on \mathcal{G} . Thus,

$$\mu Q_{k+1}[f] = \frac{\nu[f]}{n} + \mathfrak{S}_n Q_k[F].$$

The analogous expression holds for μO_{k+1} and O_k . By definition of the Wasserstein metric,

$$(8.3) \quad b_{k+1} \leq \sup \{ |\mathfrak{S}_n Q_k[F] - \mathfrak{S}_n O_k[F]| : \|f\|_{\text{Lip}} \leq 1 \}.$$

Use of this equation requires an expression for the difference $|\mathfrak{S}_n Q_k[F] - \mathfrak{S}_n O_k[F]|$. For any Borel function $F : \mathcal{G} \rightarrow \mathfrak{R}$,

$$(8.4) \quad \begin{aligned} & |\mathfrak{S}_n Q_k[F] - \mathfrak{S}_n O_k[F]| \\ &= \left| \frac{Q_k [F \exp\{-S/n\}]}{Q_k [\exp\{-S/n\}]} - \frac{O_k [F \exp\{-S/n\}]}{O_k [\exp\{-S/n\}]} \right| \\ &= Q_k [\exp\{-S/n\}]^{-1} O_k [\exp\{-S/n\}]^{-1} \\ &\quad \times \left| Q_k [\exp\{-S/n\}] O_k [F \exp\{-S/n\}] \right. \\ &\quad \left. - Q_k [F \exp\{-S/n\}] O_k [\exp\{-S/n\}] \right|. \end{aligned}$$

The product of the two reciprocal factors on the right-hand side of (8.4) is bounded, via Jensen's Inequality, by $\exp((Q_k[S] + O_k[S])/n)$. The remaining factor, with all its terms, is easier to handle if we rewrite it in terms of expected values of $1 - \exp\{-S(g)/n\}$ and $F(g)(1 - \exp\{-S(g)/n\})$, quantities that are small when n is large.

Set $H(g) := F(g)(1 - \exp\{-S(g)/n\})$. Then, the final factor in (8.4) equals

$$\begin{aligned}
& (1 - Q_k[1 - \exp\{-S/n\}])O_k[H - F] - (1 - O_k[1 - \exp\{-S/n\}])Q_k[H - F] \\
&= Q_k[F] - O_k[F] - Q_k[H] + O_k[H] \\
&\quad + Q_k[H - F]O_k[1 - \exp\{-S/n\}] - O_k[H - F]Q_k[1 - \exp\{-S/n\}] \\
&= O_k[H - F]\left(\mathfrak{P}Q_k[1 - \exp\{-S/n\}] - Q_k[1 - \exp\{-S/n\}]\right) \\
&\quad + Q_k[H - F]\left(O_k[1 - \exp\{-S/n\}] - \mathfrak{P}Q_k[1 - \exp\{-S/n\}]\right) \\
&\quad + \left(Q_k[F] - O_k[F]\right)\mathfrak{P}Q_k[\exp\{-S/n\}] \\
&\quad + \left(\mathfrak{P}Q_k[H] - Q_k[H]\right)\mathfrak{P}Q_k[\exp\{-S/n\}] \\
&\quad + \left(O_k[H] - \mathfrak{P}Q_k[H]\right)\mathfrak{P}Q_k[\exp\{-S/n\}].
\end{aligned}$$

We have $|Q_k[H - F]| \leq Q_k[|F|]$ and $|O_k[H - F]| \leq O_k[|F|]$, and $\mathfrak{P}Q_k[\exp\{-S/n\}] \leq 1$. We may then bound (8.4) by

$$\begin{aligned}
& |\mathfrak{S}_n Q_k[F] - \mathfrak{S}_n O_k[F]| \\
&\leq \exp\{(Q_k[S] + O_k[S])/n\} \left(\left| Q_k[F] - O_k[F] \right| \right. \\
&\quad + \left| \mathfrak{P}Q_k[H] - O_k[H] \right| \\
(8.5) \quad &\quad + Q_k[|F|] \cdot \left| \mathfrak{P}Q_k[1 - \exp\{-S/n\}] - O_k[1 - \exp\{-S/n\}] \right| \\
&\quad + \left| \mathfrak{P}Q_k[H] - Q_k[H] \right| \\
&\quad \left. + O_k[|F|] \cdot \left| \mathfrak{P}Q_k[1 - \exp\{-S/n\}] - Q_k[1 - \exp\{-S/n\}] \right| \right).
\end{aligned}$$

Using Lemma 7.6 and Lemma 7.11, we may find positive constants c_1, c_2 such that

$$(8.6) \quad c_1 e^{-c_2 g(\mathcal{M})} \leq \frac{dQ_k}{dP_k}(g) \leq c_1 e^{c_2 g(\mathcal{M})}, \quad c_1 e^{-c_2 g(\mathcal{M})} \leq \frac{dO_k}{dP_k}(g) \leq c_1 e^{c_2 g(\mathcal{M})}.$$

If we write $c_3 := c_1(\sigma \vee 1)\mu P_{\lceil Tn \rceil}(\mathcal{M})e^{(e^{c_2}-1)\mu P_{\lceil Tn \rceil}(\mathcal{M})}$, then

$$\begin{aligned}
(8.7) \quad Q_k[S] &\leq \int c_1 e^{c_2 g(\mathcal{M})} \sigma g(\mathcal{M}) dP_k(g) \leq c_3, \\
&\quad \mathfrak{P}Q_k[S] \leq c_3, \quad Q_k[|F|] \leq c_3, \\
O_k[S] &\leq c_3, \quad \mathfrak{P}O_k[S] \leq c_3, \quad O_k[|F|] \leq c_3.
\end{aligned}$$

As $F(g) = g[f]$,

$$(8.8) \quad \left| Q_k[F] - O_k[F] \right| = \left| \mu Q_k[f] - \mu O_k[f] \right| \leq \|\mu Q_k - \mu O_k\|_{\text{Was}} \leq b_k.$$

Observe that $|H(g)| \leq \sigma g(\mathcal{M})^2/n$ and for $g, g' \in \mathcal{G}$,

$$\begin{aligned} |H(g') - H(g)| &\leq |g[f]| |\exp\{-S(g)/n\} - \exp\{-S(g')/n\}| \\ &\quad + |g[f] - g'[f]| (1 - \exp\{-S(g')/n\}) \\ &\leq (g(\mathcal{M}) \vee g'(\mathcal{M})) \cdot \frac{\sigma}{n} \|g - g'\|_{\text{Was}} + \|g - g'\|_{\text{Was}} \cdot \frac{(g(\mathcal{M}) \vee g'(\mathcal{M}))\sigma}{n} \\ &\leq \frac{2\sigma(g(\mathcal{M}) \vee g'(\mathcal{M}))}{n} \|g - g'\|_{\text{Was}}. \end{aligned}$$

We may then apply Lemma A.11 with $\beta = 1$ and $C = 4\sigma/n$ to obtain

$$\left| \mathfrak{P}Q_k[H] - O_k[H] \right| \leq 64\sigma n^{-1} (\mu Q_k(\mathcal{M}) \vee \mu O_k(\mathcal{M})) b_k.$$

A simpler version of this computation completes the bound of the third line in (8.5) by

$$\begin{aligned} &\left| \mathfrak{P}Q_k[H] - O_k[H] \right| \\ (8.9) \quad &+ Q_k[|F|] \left| \mathfrak{P}Q_k[1 - \exp\{-S/n\}] - O_k[1 - \exp\{-S/n\}] \right| \\ &\leq c_4 \frac{b_k}{n}. \end{aligned}$$

for an appropriate constant c_4 .

We would like to bound the remaining terms with respect to the Wasserstein distances among the measures involved, but this is not possible directly, because the function H is neither bounded nor Lipschitz. To get around this problem we truncate F and S , and then make use of the Poisson bounds that we have on the probability of large values of these functions. Fix some positive M , and define

$$\begin{aligned} F^{(M)}(g) &:= \text{sgn}(F(g)) (|F(g)| \wedge M), \\ S^{(M)}(g) &:= S(g) \wedge (\sigma M), \\ H^{(M)}(g) &:= F^{(M)}(g) \cdot (1 - \exp\{-S^{(M)}(g)/n\}). \end{aligned}$$

The function $n(3M^2\sigma)^{-1}H^{(M)}$ is bounded by $1/3$ and has Lipschitz constant bounded by $2/3$, so $\|n(3M^2\sigma)^{-1}H^{(M)}\|_{\text{Lip}} \leq 1$.

Then

$$\begin{aligned} \left| Q_k[H] - \mathfrak{P}Q_k[H] \right| &= \left| Q_k[F(1 - \exp\{-S/n\})] - \mathfrak{P}Q_k[F(1 - \exp\{-S/n\})] \right| \\ &\leq \left| Q_k[F^{(M)}(1 - \exp\{-S^{(M)}/n\})] - \mathfrak{P}Q_k[F^{(M)}(1 - \exp\{-S^{(M)}/n\})] \right| \\ &\quad + \mathfrak{P}Q_k[(|F| - M) \mathbf{1}_{\{|F| > M\}}] + Q_k[(|F| - M) \mathbf{1}_{\{|F| > M\}}] \\ &\quad + \mathfrak{P}Q_k \left[|F| \frac{|S - \sigma M|}{n} e^{-\sigma M} \mathbf{1}_{\{|S| > \sigma M\}} \right] \\ &\quad + Q_k \left[|F| \frac{|S - \sigma M|}{n} e^{-\sigma M} \mathbf{1}_{\{|S| > \sigma M\}} \right] \\ &\leq \frac{3M^2\sigma}{n} \|Q_k - \mathfrak{P}Q_k\|_{\text{Was}} + 4P_k \left[(g(\mathcal{M}) - M)^2 \mathbf{1}_{\{g(\mathcal{M}) \geq M\}} \cdot c_1 e^{c_2 g(\mathcal{M})} \right] \\ &\leq \frac{3M^2\sigma}{n} \|Q_k - \mathfrak{P}Q_k\|_{\text{Was}} + c_1 e^{c_2 e^{c_2} \mu P_k(\mathcal{M})} (\mu P_k(\mathcal{M}) e^{c_2})^M \cdot \frac{1}{(M-2)!} \end{aligned}$$

by Lemma A.10, where we used the fact that

$$|F(g)|e^{-\sigma M} \leq g(\mathcal{M})e^{-\sigma M} \leq (e\sigma)^{-1} + (g(\mathcal{M}) - M) \text{ for } g(\mathcal{M}) \geq M$$

in the penultimate inequality. A similar (but simpler) calculation together with Stirling's formula shows that we may find constants c_5 and c_6 such that the fourth line of (8.5) is bounded by

$$(8.10) \quad \begin{aligned} & \left| Q_k[H] - \mathfrak{P}Q_k[H] \right| \\ & + O_k[|F|] \left| Q_k[1 - \exp\{-S/n\}] - \mathfrak{P}Q_k[1 - \exp\{-S/n\}] \right| \\ & \leq \frac{3M^2\sigma}{n} a_k + c_5 \left(\frac{c_6}{M} \right)^{M-2}. \end{aligned}$$

Taking $M = \lceil \log n \rceil$ and combining this with (8.3), (8.8), (8.9) and (8.10), we may find constants c_7, c_8, c_9 such that

$$(8.11) \quad b_{k+1} \leq e^{c_9/n} b_k + \frac{c_7 \log^2 n}{n} a_k + \frac{c_8}{n^{\log \log n}}.$$

We now take the important step mentioned in Section 8.1 of considering intermediate states of the systems $(Q_k)_{k \geq 0}$ and $(O_k)_{k \geq 0}$ as starting states for the purposes of applying the bounds from Chapter 7 over time intervals that are short in the scaling limit. The starting state Q_0 from which $(Q_k)_{k \geq 0}$ and $(O_k)_{k \geq 0}$ are derived can be quite general, although there are a few conditions that must be satisfied for the bounds in Chapter 7 to apply. If an intermediate state Q_L can be shown to satisfy these conditions, then we can treat Q_L as a new starting state, count generations forward from L , and apply the results of Chapter 7. Recall that some of the bounds in these lemmas go down like $1/(k+1)$ and some of them go up like k/n . The best bounds are achieved by going forward on the order of $k = \lfloor \sqrt{n} \rfloor$ steps from an intermediate state at generation L with $0 \leq L \leq Tn - k$. The result will be a bound on a_{L+k} .

Define

$$\begin{aligned} \tilde{Q}_k &:= Q_{L+k}, \\ \tilde{P}_k &:= P_{L+k}, \\ \tilde{Q}_k^* &:= (\mathfrak{R}^* \mathfrak{M}_n^* \mathfrak{S}_n^*)^k Q_L, \\ \tilde{P}_k^* &:= (\mathfrak{M}_n^*)^k P_L. \end{aligned}$$

In order to apply the lemmas of Chapter 7, it is necessary to check a Lipschitz condition and a shattering condition. With regard to the Lipschitz condition, Lemma 7.6 guarantees the constant upper bound

$$\Theta(\tilde{P}_0, \tilde{Q}_0) \leq \Theta(P_0, Q_0) + T\sigma.$$

With regard to the shattering condition, it is necessary for the recombination measure \mathfrak{R} to be shattering with respect to the initial intensity, which for \tilde{Q} is the measure μQ_L .

We verify the shattering condition by an argument analogous to the proof of Lemma 5.4. Set $\pi_L = \mu P_L$ and choose any Borel set A of \mathcal{M} . According to the upper bound on Radon-Nikodym derivatives in Corollary 7.7, there is a positive

constant c such that

$$\mu Q_L(A) = \int g(A) \frac{dQ_L}{dP_L}(g) dP_L(g) \leq \int g(A) \exp\{cg(\mathcal{M}) + c\pi_L(\mathcal{M})\} dP_L(g).$$

The right-hand side can be evaluated via Campbell's Theorem, Proposition A.8, and equals

$$\exp(c + \pi_L(\mathcal{M})(e^c - 1 + c)) \pi_L(A).$$

Thus, $\mu Q_L(A)$ is bounded above by $\pi_L(A)$ times a constant depending on $\pi_L(\mathcal{M})$ but not on A . Similarly the lower bound on Radon-Nikodym derivatives in Corollary 7.7 provides lower bounds on $\mu Q_L(A \cap R)$ and $\mu Q_L(A \cap R^c)$ for any segregating set R , bounds which are multiples of $\pi_L(A \cap R)$ and $\pi_L(A \cap R^c)$. Because π_0 is assumed to be shattering with respect to \mathcal{R} , so is π_L by Lemma 5.4. Hence, we can find a new shattering constant α' such that

$$(\mu Q_L(A))^3 \leq 2\alpha' \int \mu Q_L(A \cap R) \mu Q_L(A \cap R^c) d\mathcal{R}(R),$$

and μQ_L is shattering.

A bound can now be established on α_{L+k} . Recall that P_k^* is the Poisson measure on \mathcal{G}^* obtained by starting with the Poisson measure P_0 and solely applying the mutation operator \mathcal{M}_n^* (without selection or recombination) to the distribution of genotypes broken down by vintages.

It follows from the relations $\Sigma \tilde{Q}_k^* = \tilde{Q}_k$, $\Sigma \mathfrak{P}^* = \mathfrak{P} \Sigma$, and $\Sigma \tilde{P}_k^* = \tilde{P}_k$ that

$$\begin{aligned} a_{L+k} &= \|\tilde{Q}_k - \mathfrak{P} \tilde{Q}_k\|_{\text{Was}} = \sup\{|\tilde{Q}_k[\phi] - \mathfrak{P} \tilde{Q}_k[\phi]| : \|\phi\|_{\text{Lip}} \leq 1\} \\ &\leq \sup\{|\tilde{Q}_k[\phi] - \mathfrak{P} \tilde{Q}_k[\phi]| : \|\phi\|_{\infty} \leq 1\} \\ &= \sup\{|\tilde{Q}_k^*[\phi \circ \Sigma] - \mathfrak{P}^* \tilde{Q}_k^*[\phi \circ \Sigma]| : \|\phi\|_{\infty} \leq 1\} \\ &\leq \sup\{|\tilde{Q}_k^*[\phi^*] - \mathfrak{P}^* \tilde{Q}_k^*[\phi^*]| : \|\phi^*\|_{\infty} \leq 1\} \\ &= \int \left| 1 - \frac{d\mathfrak{P}^* \tilde{Q}_k^*}{d\tilde{Q}_k^*}(\mathbf{g}) \right| d\tilde{Q}_k^*(\mathbf{g}) \\ &\leq c_9 \left(\frac{1}{k+1} \vee \frac{k}{n} \right) \end{aligned}$$

for some constant c_9 , by Lemma 7.15.

Setting $k = \lfloor \sqrt{n} \rfloor$ and combining this with the bound (8.11), we obtain for all $nT \geq j \geq \sqrt{n}$,

$$(8.12) \quad \begin{aligned} b_{j+1} &\leq e^{c_6/n} b_j + c_7 \frac{\log^2 n}{n} a_j + c_8 n^{-\log \log n} \\ a_j &\leq c_9 n^{-1/2}, \end{aligned}$$

Thus, there are constants c_6 and c_{10} such that

$$(8.13) \quad b_{j+1} \leq e^{c_6/n} b_j + c_{10} \frac{\log^2 n}{n^{3/2}}.$$

For $0 \leq j \leq \sqrt{n}$,

$$(8.14) \quad \begin{aligned} b_0 &= 0, \\ b_{j+1} &\leq e^{c_6/n} b_j + c_7 \frac{\log^2 n}{n} a_j + c_8 n^{-\log \log n}, \\ a_j &\leq c_9 (j+1)^{-1}, \end{aligned}$$

implying that $b_j \leq c_{11} \log^3 n/n$ for some constant c_{11} . Taking this bound on $b_{\lfloor \sqrt{n} \rfloor}$ as a starting point for iterating (8.13) we conclude that there is a constant c_{12} such that for $nT \geq j \geq \sqrt{n}$,

$$b_j \leq c_{12} \frac{\log^2 n}{n^{1/2}}.$$

Returning to (8.2), this implies that

$$\|Q_k - O_k\|_{\text{Was}} \leq c \frac{\log^2 n}{n^{1/2}}.$$

Combined with Theorem 6.3, this immediately implies

$$\lim_{n \rightarrow \infty} \sup_{\epsilon \leq t \leq T} \|\Pi_{\rho_t} - Q_{\lfloor tn \rfloor}\|_{\text{Was}} = 0$$

for $\epsilon > 0$.

When Q_0 is already Poisson, we begin with $a_0 = 0$, allowing us to extend (8.12) to k in the range $0 \leq k < \sqrt{n}$, completing the proof. \square

APPENDIX A

Results cited in the text

A.1. Gronwall's Inequality

We make frequent use of the following version of the inequality discovered by Thomas Hakon Gronwall in 1919 and extended by Richard Bellman in 1943, so we include a statement for easy reference. A proof may be found in Appendix 5.1 of [EK86].

PROPOSITION A.1. *Let f be a nonnegative Borel measurable function on \mathbb{R}_+ that is bounded on bounded intervals. Suppose there exists a constant $L > 0$ such that for all $t \geq 0$*

$$f(t) \leq L \int_0^t f(s) ds.$$

Then, $f(t) = 0$ for all t .

A.2. Two expectation approximations

The following two lemmas give error bounds for approximations used several times in this work.

LEMMA A.2. *Let Y be a nonnegative random variable with finite second moment. Then,*

$$-\mathbb{E}[Y] \leq \log \mathbb{E}[\exp\{-Y\}] \leq -\mathbb{E}[Y] + (1/2) \text{Var}(Y) \exp\{\mathbb{E}[Y]\}$$

and so

$$0 \leq \log \mathbb{E}[\exp\{-Y\}] + \mathbb{E}[Y] \leq (1/2) \text{Var}(Y) \exp\{\mathbb{E}[Y]\}.$$

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

$$0 \leq \log \mathbb{E}[\exp\{-Y\}] + \mathbb{E}[Y] \leq (1/2)\tau^2 e^\tau.$$

PROOF. Jensen's Inequality applied to the convex function $y \mapsto \exp\{-y\}$ implies $-\mathbb{E}[Y] \leq \log \mathbb{E}[\exp\{-Y\}]$.

For nonnegative y , the function $y \mapsto (1/2)y^2 - \exp\{-y\}$ is also convex. Jensen's Inequality implies

$$(1/2)(\mathbb{E}[Y])^2 - \exp\{-\mathbb{E}[Y]\} \leq \mathbb{E}[(1/2)Y^2 - \exp\{-Y\}]$$

Consequently,

$$\begin{aligned} \mathbb{E}[\exp\{-Y\}] &\leq \exp\{-\mathbb{E}[Y]\} + (1/2) \text{Var}(Y) \\ &= \exp\{-\mathbb{E}[Y]\} (1 + (1/2) \text{Var}[Y] \exp\{\mathbb{E}[Y]\}). \end{aligned}$$

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

LEMMA A.3. *Let Y be a random variable with finite first moment. Then,*

$$\mathbb{E}[Y] \leq \log \mathbb{E}[\exp\{Y\}] \leq \mathbb{E}[|Y| \exp\{|Y|\}].$$

PROOF. The first inequality is immediate from Jensen's inequality and the convexity of $y \mapsto \exp\{y\}$. For the second inequality, use the inequalities $\log(1+y) \leq y$, $y \geq -1$, and $\exp\{x\} - 1 \leq x \exp\{x\}$, $x \geq 0$, to get

$$\begin{aligned} \log \mathbb{E}[\exp\{Y\}] &\leq \mathbb{E}[\exp\{Y\}] - 1 \\ &\leq \mathbb{E}[\exp\{|Y|\}] - 1 \\ &\leq \mathbb{E}[|Y| \exp\{|Y|\}]. \end{aligned}$$

□

A.3. Identities for Poisson random measures

The following elementary lemma is well-known and follows readily from the fact that the conditional distribution of the Poisson random measure X^π given the event $\{X^\pi(\mathcal{M}) = n\}$ is the distribution of the random measure $\sum_{i=1}^n \delta_{Z_i}$, where Z_1, \dots, Z_n are i.i.d. \mathcal{M} -valued random variables with common distribution $\pi/\pi(\mathcal{M})$.

LEMMA A.4. *Suppose that $\pi', \pi'' \in \mathcal{H}^+$ and π' is absolutely continuous with respect to π'' . Then the Poisson probability measure $\Pi_{\pi'}$ is absolutely continuous with respect to $\Pi_{\pi''}$ with Radon-Nikodym derivative*

$$\begin{aligned} \frac{d\Pi_{\pi'}}{d\Pi_{\pi''}}(g) &= \exp \left\{ g \left[\log \left(\frac{d\pi'}{d\pi''} \right) \right] - \pi'(\mathcal{M}) + \pi''(\mathcal{M}) \right\} \\ &= \exp \left\{ g \left[\log \left(\frac{d\pi'}{d\pi''} \right) \right] - \int_{\mathcal{M}} \left(\frac{d\pi'}{d\pi''}(x) - 1 \right) d\pi''(x) \right\}. \end{aligned}$$

LEMMA A.5. *Let X', X'' and Y', Y'' be two i.i.d. pairs of Poisson random measures on \mathcal{M} with common intensity measure $\pi \in \mathcal{H}^+$. Fix a partition $\{A, A^{\mathfrak{C}}\}$ of \mathcal{M} . Then, the vector of random measures $(X'_A + X''_{A^{\mathfrak{C}}}, X', X'')$ has the same distribution as $(Y', Y'_A + Y''_{A^{\mathfrak{C}}}, Y''_A + Y'_{A^{\mathfrak{C}}})$. Consequently, for bounded Borel functions $F, H', H'' : \mathcal{G} \rightarrow \mathbb{R}$*

$$\begin{aligned} &\int_{\mathcal{G}} \int_{\mathcal{G}} F(g'_A + g''_{A^{\mathfrak{C}}}) H'(g') H''(g'') d\Pi_{\pi}(g') d\Pi_{\pi}(g'') \\ &= \int_{\mathcal{G}} \int_{\mathcal{G}} F(h) H'(h'_A + h''_{A^{\mathfrak{C}}}) H''(h''_A + h'_{A^{\mathfrak{C}}}) d\Pi_{\pi}(h') d\Pi_{\pi}(h''). \end{aligned}$$

PROOF. Note that the vector of random measures $(X'_A, X'_{A^{\mathfrak{C}}}, X''_A, X''_{A^{\mathfrak{C}}})$ has the same distribution as $(Y'_A, Y''_{A^{\mathfrak{C}}}, Y''_A, Y'_{A^{\mathfrak{C}}})$. Then observe that

- $Y' = Y'_A + Y'_{A^{\mathfrak{C}}}$,
- $X' = X'_A + X'_{A^{\mathfrak{C}}}$, and $X'' = X''_A + X''_{A^{\mathfrak{C}}}$
- $(X'_A + X''_{A^{\mathfrak{C}}})_A = X'_A$ and $(X'_A + X''_{A^{\mathfrak{C}}})_{A^{\mathfrak{C}}} = X''_{A^{\mathfrak{C}}}$.

□

An important property of Poisson random measures, which we use in our discussion of the starred probability measures, is that we can construct a Poisson random measure on a product space by starting with a Poisson random measure that lays down point on one component, and then conditional on that scatter of

points choosing the corresponding second coordinates independently according to some Markov kernel. One clear statement of this property is Theorem 6.3.2 of [Cin11]. A version tailored to our particular application is the following.

PROPOSITION A.6. *Suppose that $\pi_0, \dots, \pi_k \in \mathcal{H}^+$. Then, the push-forward of the probability measure $\Pi_{\pi_0} \otimes \Pi_{\pi_1} \cdots \otimes \Pi_{\pi_k}$ on \mathcal{G}^{k+1} by the map*

$$(g_0, g_1, \dots, g_k) \mapsto g_0 \otimes \delta_0 + g_1 \otimes \delta_1 + \cdots + g_k \otimes \delta_k$$

is the distribution of a Poisson random measure on $\mathcal{G} \times \{0, 1, \dots, k\}$ with intensity measure $\sum_{j=0}^k \pi_j \otimes \delta_j$.

For the sake of completeness, we record the following result that follows from combining Lemma A.5 and Proposition A.6. Recall that for $\mathbf{g} = (g_0, \dots, g_k) \in \mathcal{G}^{k+1}$ and $A \in \mathcal{B}(\mathcal{M})$ we define $\mathbf{g}_A := ((g_0)_A, \dots, (g_k)_A)$. We can identify \mathbf{g} with the measure on $\mathcal{M} \times \{0, \dots, k\}$ that assigns mass $g_i(B)$ to the set $B \times \{i\}$, and \mathbf{g}_A can be identified in the same way with the restriction of that measure to the set $A \times \{0, \dots, k\}$.

COROLLARY A.7. *Let P be the distribution of a Poisson random measure on $\mathcal{M} \times \{0, 1, \dots, k\}$. Fix a Borel set $A \subseteq \mathcal{M}$. Identifying P with a probability measure on \mathcal{G}^{k+1} , we have for bounded Borel functions $F, H', H'' : \mathcal{G}^{k+1} \rightarrow \mathbb{R}$ that*

$$\begin{aligned} & \int_{\mathcal{G}^{k+1}} \int_{\mathcal{G}^{k+1}} F(\mathbf{g}'_A + \mathbf{g}''_{A^c}) H'(\mathbf{g}') H''(\mathbf{g}'') dP(\mathbf{g}') dP(\mathbf{g}'') \\ &= \int_{\mathcal{G}^{k+1}} \int_{\mathcal{G}^{k+1}} F(\mathbf{h}) H'(\mathbf{h}'_A + \mathbf{h}''_{A^c}) H''(\mathbf{h}'_A + \mathbf{h}''_{A^c}) dP(\mathbf{h}') dP(\mathbf{h}''). \end{aligned}$$

The following version of Campbell's Theorem is one of a cluster of results relating to Palm probabilities inspired by 1909 work of N. R. Campbell. It follows immediately from two results proved in [DVJ07], Proposition 13.1.IV (an equality involving general Palm kernels) and Example 13.1c (an expression for the Palm kernel of a Poisson Process).

PROPOSITION A.8. *Suppose that $\pi \in \mathcal{H}^+$. For bounded Borel functions $f : \mathcal{M} \rightarrow \mathbb{R}$ and $F : \mathcal{G} \rightarrow \mathbb{R}$*

$$\begin{aligned} \int_{\mathcal{G}} F(g) g[f] d\Pi_{\pi}(g) &= \int_{\mathcal{M}} \int_{\mathcal{G}} F(g + \delta_m) d\Pi_{\pi}(g) f(m) \pi(dm) \\ &= \int_{\mathcal{M}} \Pi_{\pi}[F(\cdot + \delta_m)] f(m) \pi(dm). \end{aligned}$$

LEMMA A.9. *For a bounded Borel function $h : \mathcal{M} \rightarrow \mathbb{R}$ and $c \in \mathbb{R}$,*

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

PROOF. If $c \leq 0$, then the function $F(g) = e^{cg(\mathcal{M})}$ is bounded, so this is a fairly direct consequence of Campbell's Theorem, Proposition A.8. The formula cannot be immediately applied to positive c because the function $e^{cg(\mathcal{M})}$ would not be bounded. But if we take some positive integer B and define $F_B(g) := e^{c(g(\mathcal{M}) \wedge B)}$,

then

$$\begin{aligned}
\int_{\mathcal{G}} F_B(g)g[h] d\Pi_{\pi}(g) &= \int_{\mathcal{M}} \Pi_{\pi}[F_B(\cdot + \delta_x)]h(x) d\pi(x) \\
&= \pi[h] \cdot \sum_{k=0}^{\infty} e^{-\pi(\mathcal{M})} \frac{\pi(\mathcal{M})^k}{k!} e^{c((k+1)\wedge B)} \\
&= \pi[h] \cdot e^c e^{\pi(\mathcal{M})(e^c-1)} + \sum_{k=B}^{\infty} e^{-\pi(\mathcal{M})} \frac{\pi(\mathcal{M})^k}{k!} (e^{cB} - e^{ck}) \\
&\rightarrow \pi[h] \cdot e^c e^{\pi(\mathcal{M})(e^c-1)} \text{ as } B \rightarrow \infty.
\end{aligned}$$

The result then follows by Dominated Convergence. \square

A.4. Bounds for Poisson random measures

LEMMA A.10. *For any nonnegative constant c , and positive integers $s > r$,*

$$\begin{aligned}
\int_{\mathcal{G}} (g(\mathcal{M}) - s)^r \exp\{cg(\mathcal{M})\} \mathbf{1}_{\{g(\mathcal{M}) \geq s\}} d\Pi_{\pi}(g) \\
\leq \frac{1}{(s-r)!} (\pi(\mathcal{M})e^c)^s \exp\{(e^c-1)\pi(\mathcal{M})\}.
\end{aligned}$$

PROOF. We write

$$\begin{aligned}
&\int_{\mathcal{G}} (g(\mathcal{M}) - s)^r \exp\{cg(\mathcal{M})\} \mathbf{1}_{\{g(\mathcal{M}) \geq s\}} d\Pi_{\pi}(g) \\
&= e^{-\pi(\mathcal{M})} \sum_{j=0}^{\infty} \frac{(\pi(\mathcal{M})e^c)^{j+s}}{(j+s)!} j^r \\
&= e^{-\pi(\mathcal{M})} (\pi(\mathcal{M})e^c)^s \cdot \sum_{j=0}^{\infty} \frac{(\pi(\mathcal{M})e^c)^j}{j!} \frac{j!}{(j+s)!} j^r.
\end{aligned}$$

Now,

$$\begin{aligned}
\frac{j!}{(j+s)!} j^r &= \frac{j^r}{(j+1)(j+2)\cdots(j+s)} \\
&= \frac{1}{(j+1)(j+2)\cdots(j+s-r)} \cdot \frac{j^r}{(j+s-r+1)\cdots(j+s)} \leq \frac{1}{(s-r)!},
\end{aligned}$$

and the result follows. \square

The next result 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 A.11. *Consider a function $H : \mathcal{G} \rightarrow \mathbb{R}$ for which there are constants $C \in \mathbb{R}_+$ and $\beta \in \mathbb{N}_0$ such that for all $k \in \mathbb{N}$*

$$\sup_{\substack{g \neq g' \in \mathcal{G} \\ g(\mathcal{M}) \vee g'(\mathcal{M}) \leq k}} \frac{|H(g) - H(g')|}{\|g - g'\|_{\text{Was}}} + \sup_{\substack{g'' \in \mathcal{G} : g''(\mathcal{M}) \leq k \\ z \in \mathcal{M}}} |H(g'' + \delta_z) - H(g'')| \leq Ck^{\beta}.$$

Then, for $\pi', \pi'' \in \mathcal{H}^+$,

$$|\Pi_{\pi'}[H] - \Pi_{\pi''}[H]| \leq 4C(2\beta(\pi'(\mathcal{M}) \vee \pi''(\mathcal{M}) \vee 1))^\beta \|\pi' - \pi''\|_{\text{Was}}.$$

PROOF. 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

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

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

$$\begin{aligned} |\Pi_{\pi'}[H] - \Pi_{\pi^*}[H]| &= |\mathbb{E}[H(X^{\pi'}) - H(X^{\pi^*})]| \\ &\leq 2C\mathbb{E}[(X^{\pi'}(\mathcal{M}) \vee X^{\pi^*}(\mathcal{M}))^\beta \mathbf{1}\{X^{\pi'} \neq X^{\pi^*}\}]. \end{aligned}$$

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

$$\begin{aligned} |\Pi_{\pi'}[H] - \Pi_{\pi^*}[H]| &\leq 2C \sum_{k=0}^{\infty} e^{-\pi'(\mathcal{M})} \frac{\pi'(\mathcal{M})^k}{k!} k^\beta \left[1 - \left(\frac{\pi''(\mathcal{M})}{\pi'(\mathcal{M})} \right)^k \right] \\ &\leq 2C \sum_{k=1}^{\infty} e^{-\pi'(\mathcal{M})} \frac{\pi'(\mathcal{M})^{k-1}}{(k-1)!} k^\beta [\pi'(\mathcal{M}) - \pi''(\mathcal{M})] \\ &= 2C\mathbb{E}[(X^{\pi'}(\mathcal{M}) + 1)^\beta] [\pi'(\mathcal{M}) - \pi''(\mathcal{M})], \end{aligned}$$

where we used the inequality $a^k - b^k \leq ka^{k-1}(a - b)$ for $a \geq b \geq 0$. Note that if Y is an integer-valued Poisson random variable with mean λ , then writing $S(\beta, k)$ for the number of partitions of a set of size β into k blocks (that is, the Stirling number of the second kind) and $B_\beta = \sum_{k=0}^{\beta} S(\beta, k)$ for the number of partitions of a set with β members (that is, the Bell number)

$$\mathbb{E}[Y^\beta] = \sum_{k=0}^{\beta} S(\beta, k) \lambda^k \leq B_\beta (\lambda^\beta \vee 1) \leq \beta^\beta (\lambda^\beta \vee 1),$$

because $B_\beta \leq \beta^\beta$ (any partition of a set with β elements that has k blocks defines $\beta(\beta - 1) \cdots (\beta - k + 1)$ different functions from the set into itself by assigning a distinct element of the set to be the image of all the elements in each block, and there are β^β functions from the set into itself). Therefore,

$$\begin{aligned} \mathbb{E}[(Y + 1)^\beta] &\leq 2^\beta (\mathbb{E}[Y^\beta] \vee 1^\beta) \\ &\leq 2^\beta (\beta^\beta (\lambda^\beta \vee 1) \vee 1) \\ &= (2\beta)^\beta (\lambda \vee 1)^\beta. \end{aligned}$$

Thus,

$$\begin{aligned} |\Pi_{\pi'}[H] - \Pi_{\pi^*}[H]| &\leq 2C(2\beta)^\beta (\pi'(\mathcal{M}) \vee 1)^\beta |\pi'(\mathcal{M}) - \pi''(\mathcal{M})| \\ &\leq 2C((2\beta)(\pi'(\mathcal{M}) \vee 1))^\beta \|\pi' - \pi''\|_{\text{Was}}. \end{aligned}$$

On the other hand, setting $r = \pi^*(\mathcal{M}) = \pi''(\mathcal{M})$,

$$\begin{aligned}
& |\Pi_{\pi^*}[H] - \Pi_{\pi''}[H]| \\
& \leq \sum_{k=0}^{\infty} \frac{e^{-r}}{k!} \left| \int \dots \int H \left(\sum_{\ell=1}^k \delta_{y_\ell} \right) \pi^*(dy_1) \dots \pi^*(dy_k) \right. \\
& \quad \left. - \int \dots \int H \left(\sum_{\ell=1}^k \delta_{y_\ell} \right) \pi''(dy_1) \dots \pi''(dy_k) \right| \\
& \leq \sum_{k=0}^{\infty} \frac{e^{-r}}{k!} \sum_{m=0}^{k-1} \left| \int \dots \int H \left(\sum_{\ell=1}^k \delta_{y_\ell} \right) \pi^*(dy_1) \dots \pi^*(dy_m) \pi''(dy_{m+1}) \dots \pi''(dy_k) \right. \\
& \quad \left. - \int \dots \int H \left(\sum_{\ell=1}^k \delta_{y_\ell} \right) \pi^*(dy_1) \dots \pi^*(dy_{m+1}) \pi''(dy_{m+2}) \dots \pi''(dy_k) \right|.
\end{aligned}$$

Observe that the function $z \mapsto H(g + \delta_z)/C(1 + g(\mathcal{M}))^\beta$ is Lipschitz with Lipschitz constant 1. Hence, by arguments similar to those above,

$$\begin{aligned}
& |\Pi_{\pi^*}[H] - \Pi_{\pi''}[H]| \\
& \leq \sum_{k=0}^{\infty} \frac{e^{-r}}{k!} r^{k-1} k \sup_{g: g(\mathcal{M})=k} \left| \int H(g + \delta_z) \pi^*(dz) - \int H(g + \delta_z) \pi''(dz) \right| \\
& \leq \sum_{k=0}^{\infty} \frac{e^{-r}}{k!} r^{k-1} k C (k+1)^\beta \|\pi^* - \pi''\|_{\text{Was}} \\
& \leq C((2\beta)(r \vee 1))^\beta \|\pi^* - \pi''\|_{\text{Was}} \\
& \leq C((2\beta)(r \vee 1))^\beta (\|\pi' - \pi''\|_{\text{Was}} + |\pi'(\mathcal{M}) - \pi''(\mathcal{M})|) \\
& \leq 2C((2\beta)(\pi'(\mathcal{M}) \vee 1))^\beta \|\pi' - \pi''\|_{\text{Was}}.
\end{aligned}$$

Putting these bounds together and recalling we had assumed provisionally that $\pi'(\mathcal{M}) \geq \pi''(\mathcal{M})$ yields

$$|\Pi_{\pi'}[H] - \Pi_{\pi''}[H]| \leq 4C(2\beta(\pi'(\mathcal{M}) \vee \pi''(\mathcal{M}) \vee 1))^\beta \|\pi' - \pi''\|_{\text{Was}},$$

as required. \square

The next result, is a special case of Lemma A.11. It probably already exists in some form in the literature, but we have been unable to find a reference.

LEMMA A.12. *For two finite measures $\pi', \pi'' \in \mathcal{H}^+$*

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

A.5. Bounds for Radon-Nikodym derivatives

LEMMA A.13. *Let Q and Q' be equivalent probability measures on some measurable space \mathcal{X} . Suppose there are functions and $f_1, f_2 : \mathcal{X} \rightarrow \mathbb{R}$ such that*

$$C \exp\{f_1(x) - f_2(x)\} \leq \frac{dQ'}{dQ}(x) \leq C \exp\{f_1(x) + f_2(x)\}.$$

(a) Then,

$$\frac{\exp\{f_1(x) - f_2(x)\}}{\int \exp\{f_1(y) + f_2(y)\} dQ(y)} \leq \frac{dQ'}{dQ}(x) \leq \frac{\exp\{f_1(x) + f_2(x)\}}{\int \exp\{f_1(y) - f_2(y)\} dQ(y)}.$$

(b) Moreover, if $f_1, f_2 : \mathcal{X} \rightarrow \mathbb{R}_+$, then

$$\int_{\mathcal{X}} \left| \frac{dQ'}{dQ}(x) - 1 \right| dQ(x) \leq 2 \exp\{Q[f_2]\} Q[(f_1 + f_2) \exp\{f_1 + f_2\}].$$

PROOF. We have

$$C \int_{\mathcal{X}} \exp\{f_1(x) - f_2(x)\} dQ(x) \leq 1 \leq C \int_{\mathcal{X}} \exp\{f_1(x) + f_2(x)\} dQ(x).$$

So,

$$Q[\exp\{f_1 + f_2\}]^{-1} \leq C \leq Q[\exp\{f_1 - f_2\}]^{-1}$$

and (a) follows.

Now,

$$\begin{aligned} \int_{\mathcal{X}} \left| \frac{dQ'}{dQ}(x) - 1 \right| dQ(x) &\leq \int_{\mathcal{X}} \left| \frac{dQ'}{dQ}(x) - C \right| dQ(x) + |C - 1| \\ &= \int_{\mathcal{X}} \left| \frac{dQ'}{dQ}(x) - C \right| dQ(x) + \left| C - \int_{\mathcal{X}} \frac{dQ'}{dQ}(x) dQ(x) \right| \\ &\leq 2 \int_{\mathcal{X}} \left| \frac{dQ'}{dQ}(x) - C \right| dQ(x). \end{aligned}$$

By Jensen's Inequality,

$$C \leq Q[\exp\{f_1 - f_2\}]^{-1} \leq Q[\exp\{-f_2\}]^{-1} \leq \exp\{Q[f_2]\}.$$

Note also that if $a, b \geq 1$, then $|\frac{a}{b} - 1| \leq ab - 1$; this is obvious when $a \geq b$, and when $a < b$ the inequality is equivalent to $ab^2 - 2b + a = a(b-1)^2 + 2(a-1)b > 0$. Thus,

$$\begin{aligned} \left| \frac{dQ'}{dQ}(x) - C \right| &\leq C (\exp\{f_1(x) + f_2(x)\} - 1) \\ &\leq \exp\{Q[f_2]\} (f_1(x) + f_2(x)) \exp\{f_1(x) + f_2(x)\}, \end{aligned}$$

where we have used the inequality $e^y - 1 \leq ye^y$ for $y \geq 0$, and the result follows. \square

COROLLARY A.14. Set $Q = \Pi_{\pi}$ for some $\pi \in \mathcal{H}^+$ and suppose that the probability measure Q' is equivalent to Q with $\hat{H} := \Theta(Q, Q') < \infty$.

(a) Then, writing $H := dQ'/dQ$,

$$-\hat{H}g(\mathcal{M}) - \pi(\mathcal{M})(e^{\hat{H}} - 1) \leq H(g) \leq \hat{H}g(\mathcal{M}) - \pi(\mathcal{M})(e^{-\hat{H}} - 1).$$

(b) Hence,

$$-\hat{H}(g(\mathcal{M}) + \pi(\mathcal{M})e^{\hat{H}}) \leq H(g) \leq \hat{H}(g(\mathcal{M}) + \pi(\mathcal{M})).$$

(c) For any $\beta \in \mathbb{R}$,

$$\left| \log \int \exp\{H(g) + \beta g(\mathcal{M})\} dQ(g) \right| \leq \pi(\mathcal{M}) \left(\hat{H} + \exp\{\hat{H} + \beta\} - 1 \right).$$

(d) For any Borel set $A \subseteq \mathcal{M}$,

$$\begin{aligned} \int \exp\{H(g) + \beta g(A)\} g(A) dQ(g) \\ \leq \pi(A) \exp\left\{2\pi(A^{\mathbb{G}})\hat{H} + \hat{H} + \beta + \pi(A) \left(\exp\{\hat{H} + \beta\} - 1\right)\right\}. \end{aligned}$$

PROOF. Apply Lemma A.13 with $f_1(g) = 0$ and $f_2(g) = \hat{H}g(\mathcal{M})$ to get

$$\frac{\exp\{-\hat{H}g(\mathcal{M})\}}{\int \exp\{\hat{H}g(\mathcal{M})\} Q(dg)} \leq H(g) \leq \frac{\exp\{\hat{H}g(\mathcal{M})\}}{\int \exp\{-\hat{H}g(\mathcal{M})\} Q(dg)}.$$

Since

$$\int \exp\{\hat{H}g(\mathcal{M})\} Q(dg) = \exp\{\pi(\mathcal{M})(e^{\hat{H}} - 1)\}$$

and

$$\int \exp\{-\hat{H}g(\mathcal{M})\} Q(dg) = \exp\{\pi(\mathcal{M})(e^{-\hat{H}} - 1)\},$$

(a) follows immediately. Part (b) comes from combining part (a) with the inequality $|e^y - 1| \leq |y|(e^y \vee 1)$ for all real y .

Hence,

$$\begin{aligned} \int_{\mathcal{G}} \exp\{H(g) + \beta g(\mathcal{M})\} dQ(g) &\leq \int_{\mathcal{G}} e^{\hat{H}\pi(\mathcal{M})} e^{(\hat{H}+\beta)g(\mathcal{M})} dQ(g) \\ &= \exp\left\{\pi(\mathcal{M}) \left(\hat{H} + e^{\hat{H}+\beta} - 1\right)\right\}. \end{aligned}$$

The lower bound in part (c) is obtained similarly.

Finally, since $g \mapsto g(A^{\mathbb{G}})$ and $g \mapsto g(A)$ are independent under Q ,

$$\begin{aligned} \int_{\mathcal{G}} \exp\{H(g) + \beta g(A)\} g(A) dQ(g) \\ = \int_{\mathcal{G}} \exp\{H(g_{A^{\mathbb{G}}})\} dQ(g) \cdot \int_{\mathcal{G}} \exp\{H(g) - H(g_{A^{\mathbb{G}}}) + \beta g(A)\} g(A) dQ(g) \\ \leq \exp\left\{\pi(A^{\mathbb{G}}) \left(\hat{H} + e^{\hat{H}} - 1\right)\right\} \cdot \int_{\mathcal{G}} \exp\{(\hat{H} + \beta)g(A)\} g(A) dQ(g). \end{aligned}$$

By Campbell's theorem, Proposition A.8,

$$\begin{aligned} \int_{\mathcal{G}} \exp\{(\hat{H} + \beta)g(A)\} g(A) dQ(g) \\ = \int_{\mathcal{M}} \int_{\mathcal{G}} \exp\{(\hat{H} + \beta)(g + \delta_m)(A)\} \mathbf{1}_A(m) dQ(g) d\pi(dm) \\ = \int_{\mathcal{M}} \exp\{(\hat{H} + \beta)\delta_m(A)\} \mathbf{1}_A(m) d\pi(dm) \cdot \int_{\mathcal{G}} \exp\{(\hat{H} + \beta)g(A)\} dQ(g) \\ \leq \pi(A) e^{(\hat{H}+\beta)} \exp\left\{\pi(A) \left(e^{\hat{H}+\beta} - 1\right)\right\}, \end{aligned}$$

and part (d) follows. \square

LEMMA A.15. Suppose that $Q = \Pi_{\pi}$ for some $\pi \in \mathcal{H}^+$ and the probability measure Q' is equivalent to Q with $\hat{H} := \Theta(Q, Q') < \infty$.

(a) For any real β ,

$$\begin{aligned} -\pi(A) &\leq \log \int \exp\{\mathbf{1}_{\{g(A)\geq 1\}}(H(g)+\beta g(A))\} dQ(g) \\ &\leq \pi(A) \left(\hat{H} + e^{\beta+\hat{H}} - 1 \right). \end{aligned}$$

(b) There are positive constant c and ϵ , depending only on \hat{H} and $\pi(\mathcal{M})$, such that for any β, β' with $-\epsilon \leq \beta' \leq 0 \leq \beta \leq \epsilon$,

$$\log \frac{\int \exp\{H(g) + \beta g(A)\} dQ(g)}{\int \exp\{H(g) - \beta' g(A)\} dQ(g)} \leq c(\beta + \beta')\pi(A).$$

PROOF. For $i \geq 0$ let $\gamma_i := Q[\exp\{H(X)\} \mid X(A)=i]$. Then

$$Q[\exp\{H(X_{A\mathfrak{C}})\}] e^{-i\hat{H}} \leq \gamma_i \leq Q[\exp\{H(X_{A\mathfrak{C}})\}] e^{i\hat{H}}.$$

We also have

$$\begin{aligned} 1 &= Q[\exp\{H(X)\}] \\ &\geq Q[\exp\{H(X_{A\mathfrak{C}})\}] Q[\exp\{-\hat{H}X(A)\}] \\ &= Q[\exp\{H(X_{A\mathfrak{C}})\}] e^{-\hat{H}\pi(A)} \end{aligned}$$

and

$$\begin{aligned} 1 &= Q[\exp\{H(X)\}] \leq Q[\exp\{H(X_{A\mathfrak{C}})\}] Q[\exp\{\hat{H}X(A)\}] \\ &\leq Q[\exp\{H(X_{A\mathfrak{C}})\}] e^{\hat{H}\pi(A)}, \end{aligned}$$

where $\check{H} := \hat{H}e^{\hat{H}}$. Moreover,

$$\begin{aligned} \int \exp\{\mathbf{1}_{\{g(A)\geq 1\}}(H(g)+\beta g(A))\} dQ(g) &= e^{-\pi(A)} \sum_{i=0}^{\infty} \gamma_i \frac{(\pi(A)e^{\beta})^i}{i!} \\ &\leq e^{\pi(A)(\hat{H}-1)} \sum_{i=0}^{\infty} \frac{(\pi(A)e^{\hat{H}+\beta})^i}{i!} \\ &\leq \exp\left\{\pi(A) \left(\hat{H} + e^{\beta+\hat{H}} - 1 \right)\right\}. \end{aligned}$$

We now proceed to estimate

$$\begin{aligned} \sum_{i=0}^{\infty} \gamma_i \pi(A)^i / i! &= \sum_{i=0}^{\infty} Q[\exp\{H(X)\} \mid X(A)=i] Q\{X(A)=i\} = 1 \\ \sum_{i=1}^{\infty} \gamma_i \pi(A)^i (e^{i\beta} - 1) / i! &\leq e^{\pi(A)\check{H}} \beta \sum_{i=1}^{\infty} \left(\pi(A)e^{\beta+\hat{H}} \right)^i / (i-1)! \\ &= \beta \pi(A) \exp\left\{ \beta + \hat{H} + \pi(A) \left(\check{H} + e^{\beta+\hat{H}} \right) \right\}, \\ \sum_{i=1}^{\infty} \gamma_i \pi(A)^i (e^{-i\beta'} - 1) / i! &\geq -e^{\pi(A)\check{H}} \beta' \sum_{i=1}^{\infty} \left(\pi(A)e^{\hat{H}} \right)^i / (i-1)! \\ &= -\beta' \pi(A) \exp\left\{ \hat{H} + \pi(A) \left(\check{H} + e^{\hat{H}} \right) \right\}. \end{aligned}$$

We can then estimate the ratio

$$\begin{aligned} \frac{\int \exp\{H(g) + \beta g(A)\} dQ(g)}{\int \exp\{H(g) - \beta' g(A)\} dQ(g)} &= \frac{\sum_{i=0}^{\infty} \gamma_i (\pi(A) e^{\beta})^i / i!}{\sum_{i=0}^{\infty} \gamma_i (\pi(A) e^{-\beta'})^i / i!} \\ &= \frac{1 + \sum_{i=1}^{\infty} \gamma_i \pi(A)^i (e^{i\beta} - 1) / i!}{1 + \sum_{i=1}^{\infty} \gamma_i \pi(A)^i (e^{-i\beta'} - 1) / i!} \\ &\leq \frac{1 + C\beta\pi(A)}{1 - C\beta'\pi(A)} \end{aligned}$$

for a positive constant C (depending on \hat{H} and $\pi(\mathcal{M})$), part (a) follows.

For $i \geq 1$ let $\gamma_i := Q[\exp\{H(X)\} \mid X(A)=i]$, and let $\gamma_0 := 1$. Then, for $i \geq 1$,

$$Q[\exp\{H(X_{A^c})\}] e^{-i\hat{H}} \leq \gamma_i \leq Q[\exp\{H(X_{A^c})\}] e^{i\hat{H}}.$$

We also have

$$\begin{aligned} 1 &= Q[\exp\{H(X)\}] \\ &\geq Q[\exp\{H(X_{A^c})\}] Q[\exp\{-\hat{H}X(A)\}] \\ &= Q[\exp\{H(X_{A^c})\}] e^{-\hat{H}\pi(A)} \end{aligned}$$

and

$$\begin{aligned} 1 &= Q[\exp\{H(X)\}] \\ &\leq Q[\exp\{H(X_{A^c})\}] Q[\exp\{\hat{H}X(A)\}] \\ &= Q[\exp\{H(X_{A^c})\}] e^{\hat{H}\pi(A)}. \end{aligned}$$

Thus,

$$(A.1) \quad |\log Q[\exp\{H(X_{A^c})\}]| \leq \check{H}\pi(A),$$

where $\check{H} := \hat{H}e^{\hat{H}}$. We have

$$\begin{aligned} \int \exp\{\mathbf{1}_{\{g(A) \geq 1\}}(H(g) + \beta g(A))\} dQ(g) &= e^{-\pi(A)} \sum_{i=0}^{\infty} \gamma_i \frac{(\pi(A) e^{\beta})^i}{i!} \\ &\leq e^{\pi(A)(\hat{H}-1)} \sum_{i=0}^{\infty} \frac{(\pi(A) e^{\hat{H}+\beta})^i}{i!} \\ &\leq \exp\left\{\pi(A) \left(\hat{H} + e^{\beta+\hat{H}} - 1\right)\right\}. \end{aligned}$$

We now proceed to estimate

$$\begin{aligned} \sum_{i=1}^{\infty} \gamma_i \pi(A)^i (e^{i\beta} - 1) / i! &\leq e^{\pi(A)\check{H}} \beta \sum_{i=1}^{\infty} \left(\pi(A) e^{\beta+\hat{H}}\right)^i / (i-1)! \\ &= e^{\pi(A)\check{H}} \beta \pi(A) \exp\left\{\beta + \hat{H} + \pi(A) \left(\check{H} + e^{\beta+\hat{H}}\right)\right\}, \\ \sum_{i=1}^{\infty} \gamma_i \pi(A)^i (e^{-i\beta'} - 1) / i! &\geq -e^{\pi(A)\check{H}} \beta' \sum_{i=1}^{\infty} \left(\pi(A) e^{\hat{H}}\right)^i / (i-1)! \\ &= -\beta' \pi(A) \exp\left\{\hat{H} + \pi(A) \left(\check{H} + e^{\hat{H}}\right)\right\}. \end{aligned}$$

We can then estimate the ratio

$$\begin{aligned} \frac{\int \exp\{\mathbf{1}_{\{g(A) \geq 1\}}\} (H(g) + \beta g(A)) dQ(g)}{\int \exp\{\mathbf{1}_{\{g(A) \geq 1\}}\} (H(g) - \beta' g(A)) dQ(g)} &= \frac{\sum_{i=0}^{\infty} \gamma_i (\pi(A) e^{\beta})^i / i!}{\sum_{i=0}^{\infty} \gamma_i (\pi(A) e^{-\beta'})^i / i!} \\ &= \frac{\sum_{i=0}^{\infty} \gamma_i \pi(A)^i / i! + \sum_{i=1}^{\infty} \gamma_i \pi(A)^i (e^{i\beta} - 1) / i!}{\sum_{i=0}^{\infty} \gamma_i \pi(A)^i / i! + \sum_{i=1}^{\infty} \gamma_i \pi(A)^i (e^{-i\beta'} - 1) / i!}. \end{aligned}$$

Thus, there is a positive constant C (depending on \hat{H} and $\pi(\mathcal{M})$) such that

$$R \leq \frac{1 + C\beta\pi(A)}{1 - C\beta'\pi(A)}.$$

Part (b) follows immediately. \square

LEMMA A.16. *Let $\mu, \mu' \in \mathcal{H}^+$ be equivalent measures on \mathcal{M} with*

$$\alpha := \sup_{m \in \mathcal{M}} \left| \log \frac{d\mu'}{d\mu}(m) \right| < \infty.$$

Then, for any measure $\xi \in \mathcal{H}^+$,

$$\sup_{m \in \mathcal{M}} \left| \log \frac{d[\mu' + \xi]}{d[\mu + \xi]}(m) \right| \leq \alpha.$$

PROOF. Since $e^\alpha \geq 1$, for any nonnegative function $F : \mathcal{M} \rightarrow \mathbb{R}_+$,

$$\begin{aligned} \int_{\mathcal{M}} F(x) d[\mu' + \xi](x) &= \int_{\mathcal{M}} F(x) \frac{d\mu'}{d\mu}(x) d\mu(x) + \int_{\mathcal{M}} F(x) d\xi(x) \\ &\leq \int_{\mathcal{M}} F(x) e^\alpha d\mu(x) + e^\alpha \int_{\mathcal{M}} F(x) d\xi(x) \\ &\leq e^\alpha \int_{\mathcal{M}} F(x) d[\mu + \xi](x). \end{aligned}$$

Thus,

$$\frac{d[\mu' + \xi]}{d[\mu + \xi]}(x) \leq e^\alpha \quad \text{for } [\mu + \xi]\text{-a.e. } x \in \mathcal{M}.$$

An essentially identical calculation gives a corresponding lower bound $e^{-\alpha}$ and 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.
- [Bö0] Reinhard Bürger, *The mathematical theory of selection, recombination, and mutation*, John Wiley, Chichester, New York, 2000.
- [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*, Demographic Research Monographs, Springer Verlag, 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.
- [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.
- [Cha94] Brian Charlesworth, *Evolution in age-structured populations*, Cambridge University Press, Cambridge, 1994.
- [Cha01] ———, *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.
- [Çın11] Erhan Çınlar, *Probability and stochastics*, Graduate Texts in Mathematics, vol. 261, Springer, New York, 2011. MR 2767184
- [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.
- [CT03] James R. Carey and Shripad Tuljapurkar (eds.), *Life span: Evolutionary, ecological, and demographic perspectives*, Population Council, 2003, Pop. Dev. Rev. vol. 29 suppl.
- [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.
- [DVJ07] Daryl J. Daley and David Vere-Jones, *An introduction to the theory of point processes*, 2 ed., Springer Series in Statistics, vol. 2, Springer-Verlag, New York, 2007.

- [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.
- [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.
- [Kin93] John F. C. Kingman, *Poisson processes*, Oxford University Press, Oxford, 1993.
- [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.
- [Med52] Peter Medawar, *An unsolved problem in biology: An inaugural lecture delivered at University College, London, 6 December, 1951*, H. K. Lewis and Co., London, 1952.
- [Nev75] J. Neveu, *Discrete-parameter martingales*, revised ed., North-Holland Publishing Co., Amsterdam, 1975, Translated from the French by T. P. Speed, North-Holland Mathematical Library, Vol. 10. MR MR0402915 (53 #6729)
- [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.
- [RR98] Svetlozar T. Rachev and Ludger Rüschendorf, *Mass transportation problems. Vol. I*, Probability and its Applications (New York), Springer-Verlag, New York, 1998.
- [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.
- [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.
- [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*, Tech. Report 757, Department of Statistics, University of California at Berkeley, 2008, Available at <http://arxiv.org/abs/0807.0483>.
- [WF97] Kenneth W. Wachter and Caleb E. Finch (eds.), *Between zeus and the salmon: The biodemography of longevity*, National Academy Press, Washington, D.C., 1997.
- [Wil57] George C. Williams, *Pleiotropy, natural selection, and the evolution of senescence*, Evolution **11** (1957), 398–411.
- [WSE10] Kenneth W. Wachter, David R. Steinsaltz, and Steven N. Evans, *Vital rates from the action of mutation accumulation*, Journal of Population Ageing (2010), Currently only published online.

Index of notation

\mathcal{M} , 4 \mathcal{G} , 4 δ_m , 4 \mathbb{N}_0 , 4 P_t , 4 ν , 5 Φ , 5 $P_t[\Phi]$, 5 \mathbb{R} , 5 $\tilde{\Pi}$, 5 $\tilde{\lambda}$, 5 \otimes , 5 Z_t , 6 ρ_t , 6 $S(g)$, 6 \mathbb{N}_0 , 9 v^I , 11 $\ell_x(g)$, 11 f_x , 11 $\theta(m, x)$, 11 λ , 11 \mathbb{E} , 12 ρ_* , 12 r_* , 12 \mathcal{H} , 15 \mathcal{H}^+ , 15 π^+, π^- , 15 μP , 15 X^π , 15 Π_π , 15 Lip , 16 $\ \cdot\ _{\text{Lip}}$, 16 $\ \cdot\ _{\text{Was}}$, 16 $\pi[f]$, 16 σ , 16 $\ \cdot\ _{\text{TV}}$, 17 Lip_{TV} , 17 Θ , 17 F_π , 17 $D\pi$, 17 $C([0, T], \mathcal{H})$, 18 $\ \alpha\ _c$, 18 Γ , 18	Δ , 18 ζ , 21 $r_t(m)$, 21 $q_\nu(m)$, 21 $L^\infty(\mathcal{M}, \zeta)$, 21 L_+^∞ , 21 \mathcal{K} , 21 \mathcal{K}^+ , 21 $J(t)$, 24 $\tilde{\mathcal{H}}$, 25 $\tilde{\mathcal{G}}$, 25 ρ_* , 27 $B(\tilde{\rho}, \tilde{\rho}')$, 27 $\theta(m)$, 29 a_k , 30 b_k , 30 $A(z)$, 30 $B(z, t)$, 30 $C(z)$, 30 D_x , 33 $C_b(\mathcal{M}, \mathbb{R})$, 34 \tilde{K} , 34 $p^{(u)}$, 35 $\mathbf{1}$, 35 T_p , 35 \mathcal{D} , 35 $L(p)$, 35 γ_t , 38 β_t , 39 C_T , 39 ρ_{**} , 39 $\ \cdot\ _{\text{TV}}$, 40 β_t , 40 $J(s, m)$, 40 p_n , 45 $A_\rho(x)$, 49 \mathfrak{M}_n , 52 \mathfrak{S}_n , 52 \mathfrak{P} , 52 $g^{(j)}$, 53 g_A , 54 \mathcal{R} , 54 \mathfrak{R} , 54
--	--

\mathcal{B} , 54
 $R^{\mathbb{C}}$, 54
 Q_k , 55
 P_k , 55
 O_k , 55
 \mathcal{L} , 55
 $v \prec \ell$, 55
 $V(\ell)$, 55
 $W(v, \ell)$, 56
 $\mathfrak{R}_{\mathcal{A}}$, 57
 $\mathcal{A}_k(\ell)$, 58
 Ξ_{ℓ} , 59
 Γ_v , 59
 \mathcal{G}^* , 60
 \sqcup , 60
 \mathbf{g} , 61
 $\Sigma(\mathbf{g})$, 61
 $\Psi_j(\mathbf{g})$, 61
 ΣQ , 61
 $\Psi_j Q$, 61
 $\mathfrak{M}_n^* Q$, 61
 $\mathfrak{S}_n^* Q$, 61
 $\mathfrak{R}^* Q$, 62
 $\mathfrak{P}^* Q$, 62
 Q_k^* , 62
 P_k^* , 62
 O_k^* , 62
 $|\mathcal{A}|_r^{(\lambda)}$, 66
 $[\cdot]$, 72
 π_k , 75
 H_{π} , 76
 $\tilde{F}_{\mathcal{A}}$, 82
 $\xi_{k,i}$, 83
 \hat{H} , 86
 \mathbf{i}_j , 87
 $\tau(m)$, 92
 $S(\beta, k)$, 115
 $B_{\beta, k}$, 115

Index

- antagonistic pleiotropy, 1
- attractive equilibrium, 11, 27, 29, 32, 43, 46, 49
- Bürger, Reinhard, 2, 5, 7, 12
- Barton, Nicholas, 51, 53
- Barton, Nicolas, 3, 53
- Bell number, 115
- biodemography, 2, 3
- box attractive, 27, 43, 46, 49
- box stable, 27, 43, 46, 49
- Campbell's Theorem, 12, 77, 92, 97, 109, 113
- Campbell, N.R. , 113
- Charlesworth, Brian, 2–4, 48
- comparisons lemma, 21, 39, 43
- complete Poissonization, 52
- concave costs, 36
- contraction mapping theorem, 19
- cumulative hazard, 11, 47
- Curtsinger, James, 2
- Dawson, Kevin, 2
- demographic costs, 11, 36, 47, 49
- dispersive probability measure, 68, 69
- dynamical system, vii, 3, 8, 10, 17, 28, 35
- epistatic effects, 7, 8
- equilibrium, 1, 10–12, 27, 29, 31, 32, 35, 39, 42, 43, 45, 46, 49
- Fréchet derivatives, 33–35, 38, 49
- genetic legacy, 56
- genotype, 1, 3, 4, 15
- Gompertz hazards, 4
- Gronwall's Inequality, 24, 39, 41, 42, 111
- Gronwall, Thomas H., 111
- Hahn-Jordan decomposition, 15, 18, 38
- Haldane's Principle, 3
- Haldane, John B. S., 9
- hazard function, 2, 11, 47, 48
- Horiuchi, Shiro, 3
- intensity measure, vii, 3, 6, 8, 54
- Kantorovich-Rubinstein distance, 16
- Kimura, Motoo, 2
- Lévy process, 5
- Lagrange inversion formula, 10
- Lambert's W function, 10
- LeCam, Lucien, 70
- longevity, 3
- Maruyama, Takeo, 2
- Medawar, Peter, 1
- minimal equilibrium, 45
- Monge-Kantorovich distance, 16
- Mueller's ratchet, 7
- multiplicative costs, 10, 29, 32, 42
- mutation accumulation, 1
- mutation counting, 9
- net maternity function, 49
- net reproduction ratio, 11
- one-dimensional systems, 28
- Picard, Charles E., 18
- Pletcher, Scott, 2
- Poisson identity, 12
- polynomial costs, 10, 11
- quasi-linkage equilibrium, 3
- recombination, 2, 7, 8, 57, 63, 68, 73, 104
- recombination measure, 54
- recombination operator, 54
- recombination tree, 55
- segregating set, 8, 53–55
- selective cost, 3, 5, 6, 9–11, 15, 18, 25, 30, 73, 104
- shattering, 63, 66, 68
- somatic mutations, 1
- stable equilibrium, 27, 32, 43, 46, 49
- Stirling number, 115
- thinning, 115
- total variation norm, 17, 40, 82

Turelli, Michael, 3, 51, 53

vintage, 55, 60

walls of death, 4

Wasserstein metric, 15, 16, 18, 27, 40, 73,
114

Wasserstein, Leonid, 15

Williams, George C., 1