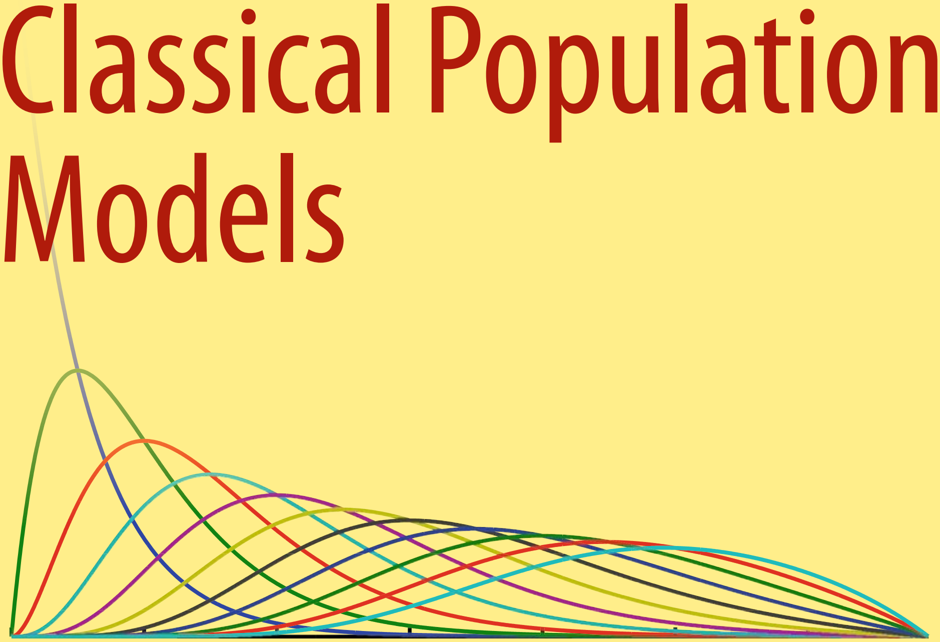
Probability Theory and Stochastic Modelling 102

Raphaël Cerf · Joseba Dalmau

The Quasispecies Equation and Classical Population Models



Springer

Probability Theory and Stochastic Modelling

Volume 102

Probability Theory and Stochastic Modelling publishes cutting-edge research monographs in probability and its applications, as well as postgraduate-level textbooks that either introduce the reader to new developments in the field, or present a fresh perspective on fundamental topics.

Books in this series are expected to follow rigorous mathematical standards, and all titles will be thoroughly peer-reviewed before being considered for publication.

Probability Theory and Stochastic Modelling covers all aspects of modern probability theory including:

- Gaussian processes
- Markov processes
- Random fields, point processes, and random sets
- Random matrices
- Statistical mechanics, and random media
- Stochastic analysis
- High-dimensional probability

as well as applications that include (but are not restricted to):

- Branching processes, and other models of population growth
- Communications, and processing networks
- Computational methods in probability theory and stochastic processes, including simulation
- Genetics and other stochastic models in biology and the life sciences
- Information theory, signal processing, and image synthesis
- Mathematical economics and finance
- Statistical methods (e.g. empirical processes, MCMC)
- Statistics for stochastic processes
- Stochastic control, and stochastic differential games
- Stochastic models in operations research and stochastic optimization
- Stochastic models in the physical sciences

Probability Theory and Stochastic Modelling is a merger and continuation of Springer's Stochastic Modelling and Applied Probability and Probability and Its Applications series.

More information about this series at https://link.springer.com/bookseries/13205

Raphaël Cerf • Joseba Dalmau

The Quasispecies Equation and Classical Population Models

Springer

Raphaël Cerf
LMO, Université Paris-Sud
CNRS, Université Paris-Saclay
Orsay, France

DMA, Ecole Normale Supérieure
CNRS, PSL Research University
Paris, France

Joseba Dalmau
NYU-ECNU Institute
of Mathematical Sciences
NYU Shanghai
Shanghai, China

ISSN 2199-3130 ISSN 2199-3149 (electronic)
Probability Theory and Stochastic Modelling
ISBN 978-3-031-08662-5 ISBN 978-3-031-08663-2 (eBook)
https://doi.org/10.1007/978-3-031-08663-2

Mathematics Subject Classification (2020): 60-02, 60Jxx, 92D25, 92Bxx

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

In a seminal paper published in the early 1970s, Manfred Eigen introduced a simple system of differential equations describing the evolution of an infinite population of macromolecules undergoing chemical reactions. The trajectories of this system converge to a unique equilibrium, called the “quasispecies” equilibrium.

The book by Raphaël Cerf and Joseba Dalmau revisits in depth the notion of quasispecies and demonstrates its remarkable universality in population dynamics, far beyond the original problem considered by Eigen. It explains how and why the quasispecies equilibrium can describe the long-term behavior of most classical models in population dynamics, whether they are deterministic or stochastic, including Moran–Kingman, multitype Galton–Watson, Wright–Fisher, continuous branching and Moran models. The common unifying thread is the fact that the quasispecies equilibrium is the normalized Perron–Frobenius eigenvector of the natural matrix encoding the fitness and mutation probabilities of the macromolecules.

The authors have recently made significant contributions to our understanding of finite population models, in the regime where genotypes are large (compared to the size of the population) and mutations are small. The book offers a beautiful synthesis of these works, the most salient features of which are: several explicit formulas describing the quasispecies distribution and their links with classical combinatorial identities, the phase transition separating a quasispecies regime from a disordered regime and a full proof for the Wright–Fisher model.

With the exception of a few classical results, all the results of the first four parts are rigorously demonstrated. The proofs are elegant, powerful and always accessible. The subsequent parts present some conjectures and more technical results, guiding the reader to further open questions. The text is a pleasure to read and can easily be used in several courses both in probability and Markov chains, population dynamics or mathematical ecology. In addition, the specialist, whether she/he is a mathematician or a theoretical ecologist or biologist, will find powerful ideas here for investigating finite population models.

University of Neuchâtel, March 2022 *Michel Benaïm*

Contents

Part III Error Threshold in Finite Populations

Part IV Proof for Wright–Fisher

Part V Class-Dependent Fitness Landscapes

Part VI A Glimpse at the Dynamics

Chapter 1
Introduction

We are surrounded by a huge heterogeneity of living beings: insects, plants, animals and humans. Even the creatures belonging to the same species present an extraordinary variability. And yet, what we can see with the naked eye, is but a tiny fraction of the realm of the living. Indeed, bacteria, viruses, prions and dozens of other microbes interact with us everyday without us even noticing. No matter how different, the one feature that we all share, is that we have all been shaped by the means of evolution: a careful equilibrium between mutation and selection. Mutations are responsible for having introduced all the changes in our genetic information, from a remote past until today, making us look as we do, while selection, caused by a combination of many internal as well as external forces, has preserved our lineage through history, by making it successful where many others have perished.

The present text focuses on equilibrium, on the subtle balance between selection and mutation to which we owe the vast genetic heterogeneity in many of the living populations. Imagine a population evolving in an environment that selects certain genotypes over some others (selection meaning that the fittest genotypes produce, on average, more children than their less fit companions), and that mutations allow for the fit genotypes to become unfit and vice versa. On one hand, if selection is mild, and the mutation rate is very high, the most fit individuals will have no advantage, since their genotype will immediately mutate and become unfit. In fact, if we observe such a population, we will see how the different genotypes come and go, and all of them will eventually vanish to let new genotypes appear, thus never reaching an equilibrium. On the other hand, if selection is very strong, and the mutation rate is very low, the fittest genotype will take over the whole population, thus leaving no place for any variability in the population.

We focus here on the intermediate situation, where selection and mutation compensate each other, and the population reaches an equilibrium. The main questions that we try to answer are: can this equilibrium be described as a function of the mutation and selection parameters? And if yes, to what extent is this description sensitive to the choice of the model? There exist several models describing the evolution of a population under mutation and selection, which encompass different features: the population size may be finite or infinite, constant or variable, the dynamics might

R. Cerf, J. Dalmau, *The Quasispecies Equation and Classical Population Models*,
Probability Theory and Stochastic Modelling 102,
https://doi.org/10.1007/978-3-031-08663-2_1

be deterministic or random, the different generations may or may not overlap, the mutations may happen during reproduction, or at any time of the life cycle. We consider throughout this text a series of models with these different features, and we study the equilibrium of the resulting mathematical processes.

Our goal is to present a unified picture of the mutation selection equilibrium, which is valid for several classical models in an adequate regime. Our starting point is the quasispecies equation, a general non-linear equation that describes the mutation-selection equilibrium of all the different models considered throughout the text. This equation arises naturally in Eigen's model [31], and it was Manfred Eigen and Peter Schuster who coined the term quasispecies to refer to the mutation-selection equilibrium of this model [33]. Notice that the quasispecies model of Eigen is dynamical in nature; what we call the quasispecies equation is the mere equilibrium equation which arises from Eigen's model.

There exists a huge literature on Eigen's model and quasispecies theory. Our goal here is not to present a synthesis of all the works on the quasispecies subject. For readers who wish to learn more on the various aspects of quasispecies theory and its application to the study of viruses, we refer to the books [27, 28, 30] and to the review papers [7, 13, 29]. Let us try instead to explain the objectives of our text and its relationship with other works on quasispecies.

Eigen's model is defined through a set of differential equations which describe the evolution of an infinite population of macromolecules [31]. Within this framework, Eigen and Schuster studied a specific stylized landscape, called the single peak landscape or sharp peak landscape, in which there is only one favored sequence, called the Master sequence [32]. Despite the simplicity of the model, they were able to derive very interesting results, namely the existence of an error threshold, and the formation of a specific population structure which they called a quasispecies. These concepts became widely used in biology to discuss the evolution of a population driven by mutation and selection, especially in the context of virus populations. Viruses being simpler organisms, and their mutation rates being very high, the concepts of quasispecies and error threshold are particularly relevant in understanding populations of viruses, as shows the extensive use of the term even in the more recent literature aimed at understanding the recent outbreak of the SARS-CoV-2 virus [1, 40, 48, 49, 87]. However, there is a major theoretical obstacle to apply Eigen's model to viruses, which was raised in [50]. Indeed, Eigen's model was initially formulated for an infinite population, whose genotype has finite length, whereas biological populations are finite with a typical population size much smaller than the number of possible genotypes. In addition to that, several finite population models of evolution have also been studied over the last decades, which incorporate stochastic effects, in the field of population genetics. The most classical ones are the Moran model, the Wright–Fisher model and the Galton–Watson process (see for instance the classical book [35]). This leads to a natural debate: is quasispecies theory suitable to describe the evolution of viruses, and what is its link with population genetics?

This question was fully addressed by Wilke [86]. Wilke argued that there is no disagreement between the population genetics of haploid, asexually replicating organisms and quasispecies theory, and he demonstrated it for the model of evolution of a single locus with two alleles, and for the mutational load studied by Kimura and Maruyama [56]. Furthermore, Wilke discussed whether quasispecies theory applies to finite populations. Several works aim at building a model for a finite size population which derives from Eigen's model, by introducing an approximation scheme to the deterministic differential equations, which incorporates stochastic effects: Alves and Fontanari [2], Demetrius, Schuster and Sigmund [25], McCaskill [62], Gillespie [41], Weinberger [85], and more recently Musso [67] and Dixit, Srivastava, Vishnoi [26]. In a very influential paper, Nowak and Schuster [69] constructed a birth and death model to approximate Eigen's model. Finite size corrections to Eigen's model have also been computed with the help of complex methods from statistical physics [3, 58, 73, 75, 76]. Moreover, various computer simulations have demonstrated that the predictions of quasispecies theory can be observed for finite populations. Comas, Moya and González-Candelas [19] studied how the population size affects the survival of the flattest. Ochoa [70] performed extensive simulations in the context of genetic algorithms. However these theoretical studies and these simulations deal with very stylized fitness landscapes and simple reproduction mechanisms, which are still very far from the complexity of real viruses. In the end, Wilke [86] concludes that there is nothing that could contradict the existence of the quasispecies effect in finite populations, and at the same time there is no true experimental evidence in its favor. The debate on the relevance of quasispecies concepts to the study of viruses is still ongoing [47, 72], and it seems to be quite open up to now.

This text is essentially a mathematical development of the questions discussed in Wilke's paper [86]. Namely, we wish to study further the mathematical links between Eigen's model and classical finite population models. Over the past few years, we have been investigating this question by examining successively several models. In each of these models, we found out that a quasispecies can be formed in a suitable asymptotic regime of the parameters [15, 16, 17, 20, 22]. However each model required a different treatment and lengthy proofs. Naturally, we tried to unify these results and to understand the common features which lead to the formation of a quasispecies. Currently, it seems to us that the central object which can be recovered in each case is the quasispecies equation, namely the equilibrium equation associated to Eigen's model. This is why this review text is centered on this quasispecies equation.

This text has several goals. A first goal is to show how the quasispecies equation is naturally linked with classical population models. This is essentially a mathematical formalization of one of the key points exposed in Wilke's paper [86]. A second goal is to study the quasispecies equation itself. Obviously, this equation is extremely complex and a rigorous mathematical analysis can be conducted only in specific cases. We discuss first the case of a finite genotype space, then we move on to the sharp peak landscape and finally to the case of class-dependent fitness landscapes. While studying the Moran model, we obtained an explicit formula for the quasispecies distribution on the sharp peak landscape [17], which appeared again for the Wright–Fisher model [20] and the Galton–Watson process [22]. We finally understood that this formula was in fact a solution of the limiting quasispecies equation, as we show in this text. A third goal is to show how the quasispecies and the error threshold phenomenon emerge in finite population models (which was in fact the original motivation of the works [15, 16, 17, 20, 22]). We tried to streamline the different proofs of our previous works in order to present a more general robust approach to these apparently similar results. To this end, we relied on ideas coming from the theory of random perturbations of dynamical systems of Freidlin and Wentzell [38].

At the end of the day, one may wonder whether it is worth developing all these mathematical techniques to prove facts which were certainly well known by theoretical biologists. On one hand, the mathematics associated to these apparently simple models is rich and beautiful, and certainly deserves to be investigated, in order to give precise answers to the previous questions. For instance, a delicate point, which is also raised by Wilke [86], is to understand the dependence of the error threshold on the population size. The approach presented here shows that, for a quasispecies to be formed, the population size has to be at least of the same order as the genotype length. On the other hand, these investigations lead to the development of new simple formulas for the quasispecies distributions. Because of the simplicity of the models, these formulas will not help directly to check the validity of the models, yet we believe they constitute a small modest step in this direction. Indeed, deep sequencing techniques yield a huge amount of data on the genotype of the viruses, and we need theoretical models to explain the structure of these data. For instance, in the case of class-dependent fitness landscapes, we have obtained a formula which allows us to recover the fitness landscape if we are given the concentration of each Hamming class at equilibrium. Unfortunately, because of all the simplifying features that lead to it, this formula is unlikely to be realistic. Nevertheless, we hope that it is a good starting point for theoretical discussions, and that it will be extended in due time when operational models of real fitness landscapes will be available.

Of course, the results presented here are not valid for any model of mutation selection. The formula for the quasispecies distribution seems to depend on some specific assumptions, for instance it is crucial that mutations occur only during a reproduction event. In the well-studied Crow–Kimura model, mutations and reproduction events are decoupled, and the equilibrium quasispecies equation is different from the one we consider here, so we choose not to speak about this class of models. However, several very interesting mathematical works have investigated the quasispecies theory within the framework of the Crow–Kimura model with the aim of finding formulas for the quasispecies. An important mathematical contribution is the papers [8, 45], where general criteria for the existence of an error threshold are discussed. In [8], the quasispecies equilibrium is characterized with the help of an approximate variational principle, under adequate approximation hypothesis on the mutation and reproduction rates, both for the Crow–Kimura and the Eigen model. These results are more general and robust than the results we present here. However, as noted in [12, 13], one has to check carefully the limiting procedures required to apply these results. In a series of recent works, Bratus, Novozhilov and Semenov [12, 13, 78] and Novozhilov and Semenov [79, 80, 81] study the quasispecies equation. They derive many interesting properties of the solutions under some symmetry assumptions on the fitness landscape, using a spectral approach. Their framework is also more general than ours. In the case of the sharp peak landscape for the Crow–Kimura model, they manage in [12, 79] to obtain exact expressions of the quasispecies distribution, with a wealth of additional information concerning the speed of convergence and the error threshold. We believe that the work we do here for Eigen's model can also be done for the Crow–Kimura model. More precisely, each finite population model considered here has its counterpart, in the form of a model where the mutations and the reproduction events are decoupled, and the equilibrium equation of the Crow–Kimura model should be recovered in some adequate asymptotic regime. In this scheme, the counterpart to the combinatorial formulas for the asymptotic quasispecies distribution are to be found in [12, 79]. Finally, the papers [78, 80] implement an approach similar to [12, 79] in order to solve the quasispecies equation of Eigen's model with various fitness landscapes. In the case of the sharp peak landscape, they also obtain a beautiful exact formula for the solution to the quasispecies equation, valid for any genotype length.

Let us finally describe the structure of the text. The quasispecies equation constitutes the backbone of the exposition, it is introduced in part I and all the subsequent parts are closely related to it. The central results are presented in parts II and III. In part II, the sharp peak landscape is introduced. We explain the error threshold phenomenon and we give an explicit formula for the quasispecies distribution. While part II deals with Eigen's model, in part III we present the counterpart of the error threshold in classical finite population models, namely the Wright–Fisher model and the Moran model. A full detailed proof of the main result for the Wright–Fisher model is given in part IV. In part V, we consider class-dependent fitness landscapes, which give rise to generalized quasispecies distributions. Part VI deals with the dynamical aspects of the models.

Part I
Finite Genotype Space

Overview of Part I

Instead of starting by introducing the different models, we begin by giving the equilibrium equation or quasispecies equation right away, which can be derived in a very simple manner, just by thinking about what mutation-selection equilibrium must mean. Our first chapter focuses on solving the equilibrium equation, that is, on characterizing it as a function of the selection and mutation parameters. Chapters 3 and 4 introduce the different models we will deal with in the rest of the text. In chapter 3, we introduce three models, the common feature of all these models being that the different generations do not overlap, in particular the time is discrete. The first of them is the Moran–Kingman model, where the population is taken to be infinite. The second model is the Galton–Watson model, where the population is finite, but varies over time, while the third is the Wright–Fisher model, where the population is also finite, but constant over time. In chapter 4, we introduce three continuous time models with overlapping generations, namely: Eigen's model for an infinite population, the continuous branching process for a finite population with variable size, and the Moran model for a finite and constant-size population. The common features shared by all the models we consider are:

- The population is well mixed, that is, there is no geographical structure, and the proportions of the different genotypes suffice to give a full description of a population.
- Individuals die at reproduction, either their own, or some other individuals'.
- Mutations happen during reproduction.

In addition to introducing the models, we also show in chapters 3 and 4 how the quasispecies equation arises in all of these models.

Chapter 2
The Quasispecies Equation

In this chapter, we first introduce the general quasispecies equation. We then present the classical Perron–Frobenius theorem and apply it to solve the quasispecies equation in the case where the set of genotypes is finite, under some additional assumptions.

2.1 The Equilibrium Equation

We consider a population of individuals evolving under the conjugate effects of mutation and selection. Individuals reproduce, yet the reproduction mechanism is error-prone, and mutations occur constantly. These mutations drive the genotypes away from the current equilibrium.

Let us introduce some notation in order to describe the model precisely. We denote by E the set of the possible genotypes (the set E might be finite of infinite). Generic elements of E are denoted by the letters u, v. The Darwinian fitness of an individual having genotype u is denoted by $\mathsf{A}(u)$, and can be thought of as its mean number of offspring. Let us denote by $c(u)$, $u \in \mathsf{E}$, the fraction of individuals of type u in the population at equilibrium. Without mutations, the quantity $c(u)$ would be proportional to $\mathsf{A}(u)$. When mutations occur in each reproduction cycle, an individual of type u might appear as the result of mutations from offspring of other types. Let us denote by $\mathsf{M}(v, u)$ the probability that the offspring of an individual of type v is of type u. We call $(\mathsf{M}(u, v), u, v \in \mathsf{E})$ the mutation matrix. Of course, we have

$$\forall\, v \in \mathsf{E} \qquad \sum_{u \in \mathsf{E}} \mathsf{M}(v, u) = 1\,.$$

At equilibrium, the fraction $c(u)$ of individuals of type u in the population has to be proportional to the mean production of individuals of type u, that is, there exists an $\alpha > 0$ such that

R. Cerf, J. Dalmau, *The Quasispecies Equation and Classical Population Models*,
Probability Theory and Stochastic Modelling 102,
https://doi.org/10.1007/978-3-031-08663-2_2

$$\forall\, u \in \mathsf{E} \qquad c(u) \;=\; \alpha \sum_{v\in \mathsf{E}} c(v)\mathsf{A}(v)\mathsf{M}(v,u)\,.$$

Summing these equations over u, we get

$$1 \;=\; \alpha \sum_{v\in \mathsf{E}} c(v)\mathsf{A}(v)\,.$$

The sum on the right-hand side represents the mean fitness of the population at equilibrium. Therefore α has to be equal to the inverse of the mean fitness of the population at equilibrium, and we conclude that the fractions $c(u)$ satisfy the following set of equations:

$$\forall\, u \in \mathsf{E} \qquad c(u) \sum_{v\in \mathsf{E}} c(v)\mathsf{A}(v) \;=\; \sum_{v\in \mathsf{E}} c(v)\mathsf{A}(v)\mathsf{M}(v,u) \tag{2.1}$$

subject to the constraint

$$\forall\, u \in \mathsf{E} \qquad c(u) \geq 0\,, \qquad \sum_{u\in \mathsf{E}} c(u) \;=\; 1\,. \tag{2.2}$$

In chapters 3 and 4, we will show how these equations characterize the equilibrium in several classical models in population genetics and mathematical biology. One of these models is Eigen's model [31], who studied the equilibrium equations in detail, and found that for certain choices of the selection and mutation parameters A and M, the above equilibrium has the following feature: the fittest genotype has a positive but possibly low concentration, and the mutants that are close to the fittest genotypes have positive concentrations too. Eigen and Schuster [33] coined the term quasispecies in order to describe this kind of equilibrium, as opposed to a species, where the fittest genotype would have a proportion close to 1. Due to the relevance of the concept of quasispecies in several areas of biology, we shall refer to the system of equations (2.1) as the quasispecies equation or the equilibrium equation. From part II onwards, we focus on the particular choices of A and M that are more pertinent from the quasispecies perspective, but before doing so we make an attempt at solving the equilibrium equation for arbitrary A and M. Unfortunately, the quasispecies equation cannot be solved analytically in general. We shall therefore focus on a more specific framework. Throughout this chapter, we consider the case where the space of genotypes E is a finite set. The case where E is infinite is much more delicate and mathematically challenging, and it cannot be analyzed in full generality.

2.2 The Perron–Frobenius Theorem

When the space of genotypes is finite, the key tool to solve the quasispecies equation (2.1) is the famous Perron–Frobenius theorem [82]. We state here a simplified version for symmetric matrices, which will be enough for our purposes, and for which the proof is considerably simpler than for the general case.

Proposition 2.1 *Let B be a square matrix, which is symmetric, and whose entries are all positive. Then its eigenvalues are real, the largest eigenvalue λ is positive, and the corresponding eigenspace has dimension one. Moreover there exists an eigenvector associated to λ whose coordinates are all positive. Finally any eigenvector of B whose coordinates are all non-negative is associated to λ.*

Proof Since B is symmetric and real, all its eigenvalues are real. The sum of its eigenvalues is equal to its trace, which is positive, thus the largest eigenvalue of B is positive, let us call it λ. Let $y = (y(u))_{u \in \mathsf{E}}$ be an eigenvector associated to λ:

$$\forall\, u \in \mathsf{E} \qquad \lambda y(u) \;=\; \sum_{v \in \mathsf{E}} B(u, v) y(v)\,.$$

We can assume that the Euclidean norm of y is 1, i.e., $\langle y, y \rangle = 1$, where $\langle \cdot, \cdot \rangle$ denotes the standard scalar product in $\mathbb{R}^{\mathsf{E}}$. Multiplying the previous equation by $y(u)$ and summing over $u \in \mathsf{E}$, we get

$$\lambda \;=\; \sum_{u,v \in \mathsf{E}} y(u) B(u, v) y(v) \;=\; \langle y, By \rangle\,.$$

Let us denote by $|y|$ the vector $(|y(u)|)_{u \in \mathsf{E}}$. Since the entries of B are positive, we deduce from the previous identity that

$$\lambda \;=\; \langle y, By \rangle \;\leq\; \langle |y|, B|y| \rangle \;\leq\; \sup_{z : \langle z, z \rangle = 1} \langle z, Bz \rangle\,.$$

However, since B is symmetric and real, the last supremum is precisely equal to λ. Therefore all the previous inequalities were in fact equalities. Since all the entries of B are positive, we conclude that all the entries of y have the same sign. The eigenvector identity implies furthermore that no entry of y is null. So far, we have proved that an eigenvector associated to λ has all its entries positive, or all negative, and none of them is zero. Let y, z be two eigenvectors associated to λ. We choose a real number α so that the first coordinate of $y - \alpha z$ vanishes. Since we have $B(y - \alpha z) = \lambda(y - \alpha z)$, necessarily $y - \alpha z = 0$. Thus the eigenspace associated to λ has dimension one. Finally, let y be an eigenvector associated to λ with positive coordinates and let z be another eigenvector of B with non-negative coordinates, associated to an eigenvalue μ. The eigenvalue identity implies that μ is positive and that all the coordinates of z are positive as well. We can then find $\alpha > 0$ sufficiently small so that $z(u) \geq \alpha y(u)$ for $u \in \mathsf{E}$. We have then, for any $n \geq 1$,

$$\langle z, B^n z \rangle \;=\; \mu^n \langle z, z \rangle \;\geq\; \langle \alpha y, B^n \alpha y \rangle \;=\; \alpha^2 \lambda^n \langle y, y \rangle\,.$$

Sending n to infinity, we conclude that $\mu \geq \lambda$, therefore $\mu = \lambda$. □

2.3 Solutions

We shall now use the Perron–Frobenius theorem to solve the quasispecies equation in the case where the set of genotypes is finite, and under some additional hypothesis. Suppose that $(c(u))_{u\in\mathsf{E}}$ is a solution to the system (2.1) which satisfies the constraint (2.2). Let λ be the mean fitness, given by

$$\lambda = \sum_{v\in\mathsf{E}} c(v)\mathsf{A}(v)$$

and let us set $d(v) = \sqrt{\mathsf{A}(v)}c(v)$ for $v \in \mathsf{E}$. These new variables satisfy

$$\forall\, u \in \mathsf{E} \qquad \lambda d(u) = \sum_{v\in\mathsf{E}} d(v)\sqrt{\mathsf{A}(v)}\mathsf{M}(v,u)\sqrt{\mathsf{A}(u)}\,. \tag{2.3}$$

Therefore $(d(u))_{u\in\mathsf{E}}$ is an eigenvector of the matrix $\sqrt{\mathsf{A}(v)}\mathsf{M}(v,u)\sqrt{\mathsf{A}(u)}$. The question of the existence and uniqueness of the solutions will be settled with the help of a result from linear algebra and the following hypothesis.

Hypothesis 2.2 We suppose that the genotype space E is finite, that the fitness function A is positive, that the mutation matrix M is symmetric and that all its entries are positive.

Suppose that hypothesis 2.2 holds. We can apply proposition 2.1 to the matrix

$$B(u,v) = \sqrt{\mathsf{A}(v)}\mathsf{M}(v,u)\sqrt{\mathsf{A}(u)}\,.$$

If $(d(u))_{u\in\mathsf{E}}$ is a solution to (2.3) with non-negative entries, then λ has to be the largest eigenvalue of B and $(d(u))_{u\in\mathsf{E}}$ is an eigenvector associated to λ. Since the corresponding eigenspace has dimension one, there is a unique choice satisfying the constraint (2.2). Therefore, under hypothesis 2.2, the system (2.1) admits a unique solution satisfying the constraint (2.2). In fact, this result still holds if we relax the hypothesis that the mutation matrix M is symmetric. We would then make appeal to the general Perron–Frobenius theorem [82] to get the conclusion.

Notation. Throughout part I, we assume that hypothesis 2.2 holds. We define the matrix W by setting

$$\forall\, u,v \in \mathsf{E} \qquad W(u,v) = \mathsf{A}(u)\mathsf{M}(u,v)\,. \tag{2.4}$$

We call the matrix W the mean reproduction matrix. For $u,v \in \mathsf{E}$, the quantity $W(u,v)$ represents the mean number of offspring of type v produced by an individual of type u. We denote by λ the Perron–Frobenius eigenvalue of the matrix W and by c^* the associated positive left eigenvector, normalized so that the sum of its components is equal to 1. The vector c^* is the unique solution of the quasispecies equation (2.1) which satisfies the constraint (2.2). The link between the quasispecies equation and the Perron–Frobenius eigenvector has been known for a long time and it is used in many works, for instance [77, 78, 86].

Chapter 3
Non-Overlapping Generations

In this chapter, we present three models of population genetics, namely the Moran–Kingman model, the Galton–Watson model, and the Wright–Fisher model. We show how to relate them with the quasispecies equation. A fundamental feature shared by these three models is that their successive generations are non-overlapping, meaning that the whole population is fully resampled from one generation to the next.

3.1 The Moran–Kingman Model

We begin by introducing the linear model, one of the simplest models for the evolution of a population with selection and mutation. Let us denote by $\mathsf{N}_n(u)$ the number of individuals of type u in the generation n. The linear model assumes that an individual of type v produces offspring at a rate proportional to its fitness $\mathsf{A}(v)$, and that a proportion $\mathsf{M}(v,u)$ of the offspring mutates and becomes of type u, thus $\mathsf{N}_{n+1}(u)$ is given by the formula

$$\forall\, u \in \mathsf{E} \qquad \mathsf{N}_{n+1}(u) = \sum_{v \in \mathsf{E}} \mathsf{N}_n(v)\mathsf{A}(v)\mathsf{M}(v,u)\,.$$

The trouble with this formula is that the sum is not necessarily an integer. To get around this problem, a natural approach is to develop stochastic population models, in such a way that the above formula describes the evolution of the mean number of individuals. The archetype of this kind of model is the Galton–Watson branching process, which is the object of the next section 3.2. If we introduce in addition a constraint on the total size of the population, then we would get the classical Wright–Fisher model, which is introduced in section 3.3. Yet the randomness adds an extra layer of complexity and stochastic models are considerably harder to study. Another simpler possibility is to consider the proportions of each type of individual in the population, instead of their numbers, as Moran and Kingman did in the late seventies [57, 65, 66]. Let us denote by $\mathsf{c}_n(u)$ the proportion of individuals of type u

R. Cerf, J. Dalmau, *The Quasispecies Equation and Classical Population Models*,
Probability Theory and Stochastic Modelling 102,
https://doi.org/10.1007/978-3-031-08663-2_3

in the generation n. The model proposed by Moran is given by

$$\forall\, u \in \mathsf{E} \qquad \mathsf{c}_{n+1}(u) \;=\; \frac{\sum_{v\in\mathsf{E}} \mathsf{c}_n(v)\mathsf{A}(v)\mathsf{M}(v,u)}{\sum_{v\in\mathsf{E}} \mathsf{c}_n(v)\mathsf{A}(v)}\,. \tag{3.1}$$

We call this model the Moran–Kingman model; it is not to be confused with the well-known Moran model, which is a stochastic model for the evolution of a finite population. Let us introduce an adequate framework to study this model. We consider the finite-dimensional simplex $\mathcal{S}$,

$$\mathcal{S} \;=\; \Big\{ c \in [0,1]^{\mathsf{E}} : \sum_{u\in\mathsf{E}} c(u) = 1 \Big\}\,,$$

and we define a map Φ from $\mathcal{S}$ to $\mathcal{S}$ by

$$\forall\, u \in \mathsf{E} \qquad \Phi(c)(u) \;=\; \frac{\sum_{v\in\mathsf{E}} c(v)\mathsf{A}(v)\mathsf{M}(v,u)}{\sum_{v\in\mathsf{E}} c(v)\mathsf{A}(v)}\,. \tag{3.2}$$

The Moran–Kingman model is the dynamical system on $\mathcal{S}$ defined by the iteration of the map Φ:

$$\forall\, n \geq 0 \qquad \mathsf{c}_{n+1} \;=\; \Phi(\mathsf{c}_n)\,. \tag{3.3}$$

The main result concerning the Moran–Kingman model is given in the following theorem.

Theorem 3.1 *The dynamical system (3.1) has a unique fixed point, which coincides with c^*. Moreover, for every $c \in \mathcal{S}$, the dynamical system (3.1) with initial condition $\mathsf{c}_0 = c$ converges to c^*, i.e.,*

$$\lim_{n\to\infty} \mathsf{c}_n \;=\; c^*\,.$$

Proof The equilibrium equation for the dynamical system (3.1) is the fundamental equation (2.1), and, as we have seen, it has a unique solution on the simplex $\mathcal{S}$, given by the vector c^*. Using the mean reproduction matrix W defined in (2.4), we have

$$\forall\, n \geq 1 \quad \forall\, u \in \mathsf{E} \qquad \mathsf{c}_n(u) \;=\; \frac{(\mathsf{c}_0 W^n)(u)}{|\mathsf{c}_0 W^n|_1}\,,$$

where $|c|_1$ is the sum of the absolute values of the components of the vector c, i.e.,

$$\forall c \in \mathbb{R}^{\mathsf{E}} \qquad |c|_1 \;=\; \sum_{u\in\mathsf{E}} |\, c(u)\,|\,.$$

The asymptotic behavior of the powers of the matrix W is given by

$$\forall\, u, v \in \mathsf{E} \qquad \lim_{n\to\infty} \frac{1}{\lambda^n} W^n(u,v) \;=\; d^*(u)\, c^*(v)\,,$$

where λ is the Perron–Frobenius eigenvalue of the matrix W, and d^* the right eigenvector associated to it, normalized so that the scalar product of d^* and c^* is equal to 1. For a proof of this result, see for instance theorem 1.2 in [82]. We deduce