UPPSALA DISSERTATIONS IN MATHEMATICS

115

Multi-trait Branching Models with Applications to Species Evolution

Daniah Tahir

Department of Mathematics
Uppsala University
UPPSALA 2019

Dissertation presented at Uppsala University to be publicly examined in Room 80101, Ångströmlaboratoriet, Lägerhyddsvägen 1, Uppsala, Friday, 14 June 2019 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Asger Hobolth (Aarhus University, Denmark).

Abstract
Tahir, D. 2019. Multi-trait Branching Models with Applications to Species Evolution. *Uppsala Dissertations in Mathematics* 115. 55 pp. Uppsala: Department of Mathematics. ISBN 978-91-506-2765-7.

This thesis provides an analysis of the evolution of discrete traits and their effect on the birth and survival of species using the theory of supercritical, continuous time Markov branching processes. We present a branching modeling framework that incorporates multi-trait diversification processes associated with the emergence of new species, death of existing species, and transition of species carrying one type of a trait to another. The trait-dependent speciation, extinction, and transition help in interpreting the relationships between traits on one hand, and linking together the diversification process with molecular evolution on the other. Various multitype species branching models are applied in order to examine the evolutionary patterns in known data sets, such as the impact of outcrossing and selfing mating systems on the diversification rates of species, and the analysis of virulent behavior of pathogenic bacterial strains in different environments. Stochastic equations and limit theorems for branching processes help scrutinize the long time asymptotics of the models under an asymmetry in change of types, and under various schemes of rescaling. In addition, we explore diversity-dependent processes in which, instead of allowing supercritical growth of population sizes, the increase in species numbers is regulated by modifying the species branching rates. The use of probabilistic methods in a setting of population genetics leads to an analogy between biallelic frequency models and binary trait species tree models. To obtain an approximation for a Markov branching process of species evolution over a long geological time scale, we methodically utilize the theory of diffusion processes. Overall, our results show that branching models can be effectively used to seek to comprehend the diversification patterns in species during the process of evolution.

Keywords: Markov models, branching processes, density dependence, discrete traits, species trees, diversification rates, diffusion approximation

Daniah Tahir, Department of Mathematics, Box 480, Uppsala University, SE-75106 Uppsala, Sweden.

ISSN 1401-2049
ISBN 978-91-506-2765-7
urn:nbn:se:uu:diva-380975 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-380975)

Dedicated to my beloved parents

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Tahir, D., Glémin, S., Lascoux, M., Kaj, I. (2019) Modeling a trait-dependent diversification process coupled with molecular evolution on a random species tree. *J. Theor. Biol. 461:189–203*.

II Tahir, D., Kaj, I., Bartoszek, K., Majchrzak, M., Parniewski, P., Sakowski, S. Using multitype branching models to analyze bacterial pathogenicity. *Submitted* (Feb. 2019).

III Kaj, I., Tahir, D. (2019) Stochastic equations and limit results for some two-type branching models. *Stat. Probab. Lett. 150:35–46*.

IV Kaj, I., Glémin, S., Tahir, D., Lascoux, M. Analysis of diversity-dependent species evolution using concepts in population genetics. *Manuscript*.

Contents

Introduction

Stochastic models based on naturally occurring random phenomena, such as molecular evolution and the growth of families of species, can be used to analyze the variability that characterizes such processes. This work utilizes various stochastic models, in particular, continuous time Markov branching models, to describe the evolution of discrete traits on species trees. The first part of the thesis not only provides a brief introduction to the mathematical theory behind some basic stochastic processes along with applications to biological models, it also serves as background reading for the forthcoming research work presented in the second part of the thesis, which consists of four appended articles.

In Part I of this thesis, we review basic nucleotide substitution models as well as fundamental models in the field of population genetics, such as the acclaimed Wright-Fisher model and the Wright-Fisher diffusion, and emphasize their relevance as Markov and diffusion models. Moreover, we outline the general theory of discrete time, continuous time, and continuous state Markov branching processes and discuss their applications to various population models. Specifically, we utilize the mathematical framework of continuous time branching processes to explore multitype trait-dependent and density-dependent models based on the differences in diversification rates during the evolution of species. Brief summaries of our research results are also provided in Part I.

In Part II of the thesis, we present our findings in the form of four appended papers. In Paper I, we model the impact of traits on species diversification rates and on molecular evolution. For this purpose, we first develop a probabilistic model for a random binary trait species tree, in which the number of species are represented by a two-type, continuous time Markov branching process. Then, we describe a trait-dependent substitution process which runs as a Poisson process along the branches of the species tree. From the analysis, we conclude that trait-dependent diversification processes can have a strong affect on molecular evolutionary rates. In Paper II, we extend the two-type model studied earlier, to four-type branching models in order to investigate pathogenic characteristics in *E. coli* bacteria. We survey existing fundamental theorems regarding the behavior of branching processes in the long time limit and apply our findings to a clinical data set of virulent and nonvirulent bacterial strains. We show that estimates of various parameter values of multitype branching models can be effectively used to provide information on the limits of the proportions of bacterial strains in different states of the models.

In the remaining papers, we examine continuous time branching models that incorporate population size dependence. In such models, species are not allowed to grow in size without a bound, instead their numbers are regulated by the population size itself. In Paper III, we analyze stochastic equations and limit results for both the basic branching process and its population size dependent version. We also explore scaled branching models and describe the fluctuations around the limit of these processes using the central limit theorem. In Paper IV, we again study a binary trait species branching process, this time using a population genetics setting. For this purpose, we construct a species modeling framework on an evolutionary time scale, and obtain a diffusion approximation of the process in order to provide an analogy with the Wright-Fisher diffusion model. The analytical structure developed is finally applied on models which integrate different forms of diversity-dependence.

Part I:
Background and summary

1. Stochastic processes with applications in evolutionary biology

A stochastic (or random) process is a mechanism by which one can assess the relationship between sequences of random events. It is defined as a family of random variables $\{X_t : t \in T\}$, where t is a parameter running over a suitable index set T. The state space S of a stochastic process gives the range of all possible values for the random variables. The index set T is usually taken as a set of times, and the parameter t may represent discrete units of time with $T = \{0,1,2,\ldots\}$, or continuous time in which $T = [0,\infty)$. Depending on the nature of the time parameter, X_t is either a *discrete time* or a *continuous time* stochastic process, whereas depending on the state space, X_t is either a *discrete state* or a *continuous state* stochastic process.

1.1 Markov models

Markov processes: A Markov process is a random process which satisfies the *Markov property*; given its current state, the probability of any future behavior of the process is independent of knowledge concerning its past behavior [46].

Discrete time Markov chain: A stochastic process $\{X_n : n = 0,1,2,\ldots\}$ is a discrete time Markov chain if

$$\mathbb{P}(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i) = \mathbb{P}(X_{n+1} = j | X_n = i),$$

for all time points n and all states $i_0, i_1, \ldots, i_{n-1}, i, j$ in the state space S. The homogeneous state *transition probability* for X_n is given by $p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i)$, where $0 \leq p_{ij} \leq 1$ and $\sum_j p_{ij} = 1$, $i = 1,\ldots,n$. The transition probabilities can be displayed as entries of the transition probability matrix, $P = \{p_{ij} : i,j = 1,\ldots,n\}$, such that the rows of the matrix sum up to 1. Furthermore, $p_{ij}(n) = \sum_k p_{ik}(m) p_{kj}(n-m)$, for all $0 < m < n$, i.e., for a process currently in state i, the probability that it will be in state j after n transitions is equal to the product of the probability that it starts in state i and reaches an

intermediate state k after m transitions, and the probability that it goes from state k to state j after $n-m$ transitions [18].

Continuous time Markov chain: A stochastic process $\{X_t : t \geq 0\}$ is a continuous time Markov chain if, given the value of X_s, the values of X_u, $u \leq s$, do not influence the value of X_{t+s}, or

$$\mathbb{P}(X_{t+s} = j | X_s = i, X_u = k) = \mathbb{P}(X_{t+s} = j | X_s = i),$$

for all s, t, $u \geq 0$ and i, j, $k \in \{1,2,\ldots\}$. The homogeneous transition probability function for X_t is given as $p_{ij}(t) = P(X_t = j | X_0 = i)$, i.e., the probability that the Markov chain currently in state i will be in state j after an additional time $t > 0$, with $0 \leq p_{ij}(t) \leq 1$ and $\sum_j p_{ij}(t) = 1$, for $i = 1,2,\ldots$. Moreover, $p_{ij}(t_1 + t_2) = \sum_k p_{ik}(t_1) p_{kj}(t_2)$ [18].

Whenever a continuous time Markov chain enters a state i, it spends there a certain amount of holding time, which is exponentially distributed with rate, say, a_i. After the holding time has elapsed, the process exits state i with rate a_i and transitions to another state j with probability p_{ij}. The rate at which the transition occurs from state i to j is given by $q_{ij} = a_i p_{ij}$. The transition rates q_{ij} are represented by the entries of the *transition rate matrix* $Q = \{q_{ij}\}$, each row of which should sum up to zero, hence the diagonal entries of Q are $q_{ii} = -\sum_{i \neq j} q_{ij}$, with $-q_{ii}$ being the rate at which the Markov chain leaves state i. The matrix Q determines the *transition probability matrix* $P(t) = \{p_{ij}(t)\}$, $t > 0$, which is given by the solution $P(t) = e^{Qt}$ of the ODE

$$\frac{dP(t)}{dt} = P(t)Q, \quad P(0) = I,$$

where I is the identity matrix [49].

Population genetics

A number of biological processes comprise of units that reproduce, for instance, individuals in demographic models and genes in population genetics models. Population genetics is the study of genetic variation within populations. It predicts the changes in the frequencies of alleles in populations over time and under various conditions. Biological applications of discrete and continuous time Markov chains can be found in the field of population genetics, by way of the classical Wright-Fisher and Moran models, respectively, which we will now elaborate in detail. Contrary to other models presented in later sections, the total size is assumed fixed in the Wright-Fisher and Moran models. Moreover, we observe that different forms of a gene (alleles) in a population genetics framework can be considered analogous to different traits in species branching processes. We implement this concept in Paper IV to analyze various properties of a branching model of binary trait species evolution under a geological time scale.

(i) The Wright-Fisher model

The Wright-Fisher (WF) model, introduced by S. Wright and R. A. Fisher in the early 20th century, is a model that describes *genetic drift*, a phenomenon defined as the random fluctuations in a finite populations' allele frequencies that occur during the transfer of alleles from one generation to the next. Mathematically, the WF model provides a simple application of the discrete time Markov chain. Consider a population with a constant size N and discrete generations, i.e., in each generation, the whole population reproduces simultaneously and is replaced by its offspring. To achieve genetic drift in the model, it is assumed that in each generation, alleles are randomly sampled with replacement (independently and with equal probability) and then transmitted to the next generation. The population consists of individuals carrying alleles of two types, say, type 0 and type 1. Let $\{X_n : n = 0,1,2,\ldots\}$ be the number of type 0 alleles in the nth generation and let $f_n = X_n/N$ be the type 0 allele frequency. Then, X_n is a discrete time Markov chain with state space $S = \{0,1,\ldots,N\}$ and transition probability

$$p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j},$$

which is a binomial distribution with parameters N and i/N. It follows that $\mathbb{E}(X_{n+1}|X_n = i) = i$, and therefore, $\mathbb{E}(f_{n+1}|f_n) = f_n$, i.e., the frequency of type 0 (and hence type 1) alleles is expected to remain the same from one generation to the next. Moreover, the states 0 and N are absorbing states of the Markov chain, thus, eventually when $X_n = 0$ (or N), $X_{n+k} = 0$ (or N) for all $k \geq 0$. In this case, the allele is said to be lost ($f_n = 0$) or fixed ($f_n = 1$) in the population [32, 33]. Figure 1.1 gives a diagrammatic representation of the WF model.

The WF model can be extended to incorporate the effect of mutations on allele frequencies [32]. Assume that a type 0 allele mutates to a type 1 allele with mutation rate δ_{01}, and type 1 to type 0 mutation occurs with rate δ_{10}. Then the expected allele frequency of type 0, given that type 0 allele frequency in generation n is f_n, is

$$\mathbb{E}(f_{n+1}) = (1 - \delta_{01})f_n + \delta_{10}(1 - f_n).$$

In this case, the transition probabilities for X_t are given by

$$p_{ij} = \binom{N}{j} \left(\frac{i}{N}\big(1 - \delta_{01}\big) + \big(1 - \frac{i}{N}\big)\delta_{10}\right)^j \left(\frac{i}{N}\delta_{01} + \big(1 - \frac{i}{N}\big)\big(1 - \delta_{10}\big)\right)^{N-j},$$

for $i, j = 0,1,\ldots,N$. If $\delta_{10}\delta_{01} > 0$, allele fixation does not occur in any state, instead, an equilibrium value is eventually reached as $n \to \infty$. At the equilibrium, there is no change in the allele frequencies, thus $\mathbb{E}(f_{n+1}) = f_n = \mathfrak{f}$ (say). This gives $(1 - \delta_{01})\mathfrak{f} + \delta_{10}(1 - \mathfrak{f}) = \mathfrak{f}$ and rearranging,

$$\mathfrak{f} = \frac{\delta_{10}}{\delta_{10} + \delta_{01}}.$$

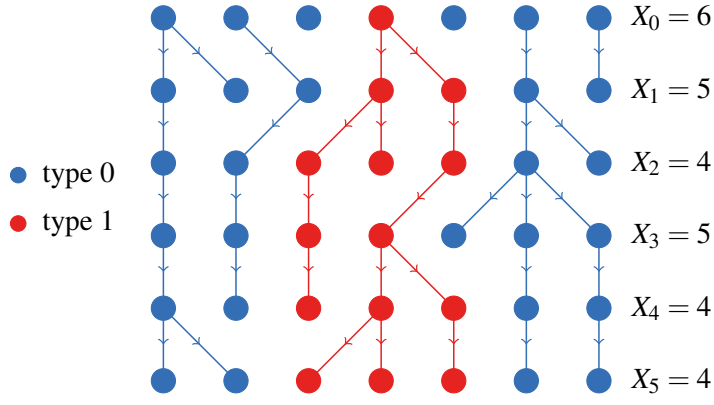

Figure 1.1. The Wright-Fisher model.

To investigate the fluctuations in the WF allele frequencies under the influence of selection, assume that type 0 alleles have a selective advantage over type 1 alleles, so that the relative fitness of type 0 and type 1 alleles is $1+s$ and 1, respectively, where $s > 0$ is the selection coefficient. In other words, all individuals with type 0 alleles and $1/(1+s)$ individuals with type 1 alleles survive. The Markov chain X_n has transition probabilities

$$p_{ij} = \binom{N}{j} \left(\frac{(1+s)i}{N+si}\right)^j \left(\frac{N-i}{N+si}\right)^{N-j},$$

for $i, j = 0, 1, \ldots, N$. The expected allele frequency of type 0 is given as

$$\mathbb{E}(f_{n+1}) = \frac{(1+s)f_n}{1+sf_n},$$

where f_n is the frequency of type 0 allele in generation n [20].

(ii) The Moran model

The Moran model, presented by P. Moran in 1958, is a continuous time analogue of the WF model. Mathematically, the Moran model is a birth and death process; a continuous time Markov chain $\{X_t : t \geq 0\}$ for which $p_{ij} = 0$ whenever $j \neq i+1$ or $j \neq i-1$, for $i, j = 1, 2, \ldots$, i.e., only those transitions (from a state i) can occur which either cause an increase in state by one $(i+1)$, known as a birth, or a decrease in state by one $(i-1)$, known as death. Transition rates q_{ij} for birth and death from state i are given by $q_{i(i+1)} = a_i p_{i(i+1)}$ and $q_{i(i-1)} = a_i p_{i(i-1)}$, respectively, where $q_{i(i+1)} + q_{i(i-1)} = a_i$ is the rate at which the process transitions out of state i [18].

Biologically, there are two essential differences between the WF and Moran models. Firstly, unlike the WF model which evolves in discrete generations,

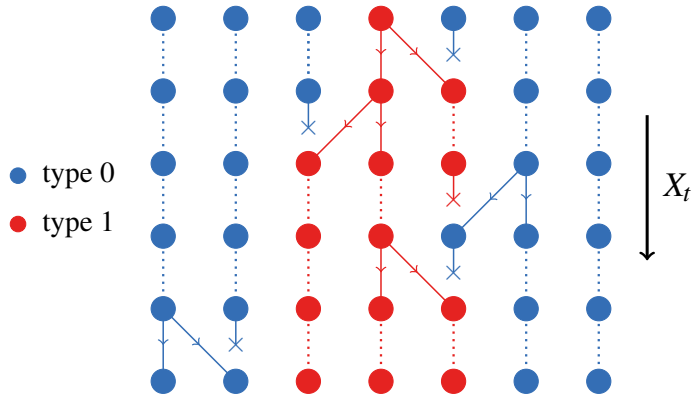

Figure 1.2. Diagrammatic representation of the Moran model. Here, the population is tracked at only those points in time where a birth-death event occurs.

the Moran model assumes overlapping generations. Secondly, while in the WF model, individuals are allowed to have any number of offspring from 0 to N, in the Moran model, genetic drift is caused by a sampling process which involves randomly choosing an individual to reproduce, immediately followed by the death of another randomly chosen individual, so that the population size remains constant at N. As before, let the population consist of individuals carrying two types of alleles, 0 and 1, and let $\{X_t : t \geq 0\}$ denote the number of type 0 alleles at time t. Then, X_t is a birth and death process with state space $S = \{0, 1, \ldots, N\}$, and transition rates given by

$$\lambda_i = \frac{i}{N}\left(1 - \frac{i}{N}\right), \quad \mu_i = \left(1 - \frac{i}{N}\right)\frac{i}{N}, \quad i \in \{0, 1, \ldots, N\}.$$

The above relations are motivated as follows: if the number of type 0 alleles is i, then to go from i to $i+1$ in one birth-death event, the individual randomly selected to reproduce (after an exponential holding time with rate one) should carry a type 0 allele with probability i/N, and the individual randomly selected to die should carry a type 1 allele with probability $1 - i/N$, hence the birth rate λ_i. Similarly, the number of type 0 alleles decrease from i to $i-1$ with death rate μ_i [33]. This process is represented in Figure 1.2.

Molecular evolution

Over the past few decades, there has been a massive increase in the availability of genetic sequence data which has led to a breakthrough in comprehending the mechanisms of molecular evolutionary processes. In this section, we study the changes that occur in gene sequences over time by surveying some basic models of molecular evolution, such as nucleotide and codon substitution models, which exploit the theory of Markov processes.

(i) Nucleotide substitution models

The hereditary material of living organisms (DNA) is composed of molecules known as nucleotides, which consist of a sugar and phosphate molecule, along with one of the four nitrogen bases, adenine (A), thymine (T), guanine (G) and cytosine (C). Continuous time Markov chains can be used to model the changes that occur randomly in the order of these bases, or the nucleotide sequence. Furthermore, the distance between a pair of nucleotide sequences, measured in terms of the average number of substitutions per site, forms the basis for phylogenetic tree reconstruction.

JC69 model: Consider a continuous time Markov chain $\{X_t : t \geq 0\}$, where X_t denotes the state of the chain at time t, which could be A, T, G, or C. Let the substitution rate from a nucleotide i to a nucleotide j be q_{ij}, $i, j \in S = \{\text{A, T, G, C}\}$. If q_{ij} has the same value, say α, for all $i \neq j$, then the resulting model is known as the JC69 (Jukes and Cantor 1969) model, and the rate matrix Q is given by

$$Q = \begin{pmatrix} -q_{\text{A}} & q_{\text{AT}} & q_{\text{AG}} & q_{\text{AC}} \\ q_{\text{TA}} & -q_{\text{T}} & q_{\text{TG}} & q_{\text{TC}} \\ q_{\text{GA}} & q_{\text{GT}} & -q_{\text{G}} & q_{\text{GC}} \\ q_{\text{CA}} & q_{\text{CT}} & q_{\text{CG}} & -q_{\text{C}} \end{pmatrix} = \begin{pmatrix} -3\alpha & \alpha & \alpha & \alpha \\ \alpha & -3\alpha & \alpha & \alpha \\ \alpha & \alpha & -3\alpha & \alpha \\ \alpha & \alpha & \alpha & -3\alpha \end{pmatrix}.$$

The transition probabilities $p_{ij}(t)$ that a nucleotide i will be a different nucleotide j after a time t, are given by the transition probability matrix as

$$P(t) = \{p_{ij}(t)\} = e^{Qt} = \sum_{k=0}^{\infty} \frac{Q^k t^k}{k!} = \begin{pmatrix} p_1(t) & p_2(t) & p_2(t) & p_2(t) \\ p_2(t) & p_1(t) & p_2(t) & p_2(t) \\ p_2(t) & p_2(t) & p_1(t) & p_2(t) \\ p_2(t) & p_2(t) & p_2(t) & p_1(t) \end{pmatrix},$$

where $p_1(t) = (1 + 3e^{-4\alpha t})/4$ and $p_2(t) = (1 - e^{-4\alpha t})/4$. As $t \to \infty$, $p_{ij}(t) = 1/4$ for all i, j. Thus, the stationary distribution, i.e., the probability that the chain is in a state k, $k \in S$, as $t \to \infty$, is given by $\pi = (\pi_{\text{A}}, \pi_{\text{T}}, \pi_{\text{G}}, \pi_{\text{C}}) = (1/4, 1/4, 1/4, 1/4)$, where $\pi P(t) = \pi$ and $\pi Q = 0$ [49].

Consider two sequences that are separated by a time t. Then, the distance between them under the JC69 model is $d = 3\alpha t$, which is the expected number of substitutions per site for each nucleotide. The probability that a site is occupied by a nucleotide, different from the ancestral sequence, is given by $p = 3p_2(t) = 3(1 - e^{-4d/3})/4$ [49]. Letting $\hat{p}$ be the proportion of sites that are different between the two sequences and equating p with $\hat{p}$, gives an estimate $\hat{d}$ of the distance between two sequences as

$$\hat{d} = -\frac{3}{4} \ln(1 - \frac{4}{3}\hat{p}).$$

K80 model: Substitutions between two pyrimidines (T and C) or two purines (A and G), known as transitions, are known to occur at different rates than

transversions, substitutions between a pyrimidine and a purine. To account for this, a model known as the K80 model was proposed by Kimura in 1980. Letting the substitution rates be μ for transitions and ν for transversions, the rate matrix Q becomes

$$Q = \begin{pmatrix} -(2\nu+\mu) & \nu & \mu & \nu \\ \nu & -(2\nu+\mu) & \nu & \mu \\ \mu & \nu & -(2\nu+\mu) & \nu \\ \nu & \mu & \nu & -(2\nu+\mu) \end{pmatrix},$$

while the transition probability matrix $P(t)$ is

$$P(t) = \begin{pmatrix} p_1(t) & p_3(t) & p_2(t) & p_3(t) \\ p_3(t) & p_1(t) & p_3(t) & p_2(t) \\ p_2(t) & p_3(t) & p_1(t) & p_3(t) \\ p_3(t) & p_2(t) & p_3(t) & p_1(t) \end{pmatrix},$$

where $p_1(t) = \big(1+e^{-4d/(\kappa+2)}+2e^{-2d(\kappa+1)/(\kappa+2)}\big)/4$, $p_2(t) = \big(1+e^{-4d/(\kappa+2)} - 2e^{-2d(\kappa+1)/(\kappa+2)}\big)/4$, and $p_3(t) = \big(1-e^{-4d/(\kappa+2)}\big)/4$, with $d = (2\nu+\mu)t$ and $\kappa = \mu/\nu$. As before, the stationary distribution is given by $\pi = (1/4, 1/4, 1/4, 1/4)$ [49].

The value $d = (2\nu+\mu)t$ gives the distance between two sequences that are separated by a time t, where the expected number of transitions and transversions per site are μt and $2\nu t$, respectively. The probability that a site has nucleotides with transitional difference is given as $A = p_2(t)$, and the probability that a site has nucleotides with transversional difference is $B = 2p_3(t)$. Moreover, let $\hat{A}$ and $\hat{B}$ be the proportion of sites that are different between the two sequences due to transitions and transversions, respectively. Equating A with $\hat{A}$ and B with $\hat{B}$ gives an estimate of the transition distance $\widehat{\mu t}$ and transversion distance $\widehat{2\nu t}$ as

$$\widehat{\mu t} = -\frac{1}{2}\ln(1-2\hat{A}-\hat{B}) + \frac{1}{4}\ln(1-2\hat{B}), \quad \widehat{2\nu t} = -\frac{1}{2}\ln(1-2\hat{B}).$$

Depending on the rates of nucleotide substitutions, various models with more realistic assumptions have also been studied, in which $q_{ij} \neq q_{ji}$ for all i, j, and the stationary distribution $\pi = (\pi_A, \pi_T, \pi_G, \pi_C)$ has unequal proportions of the four nucleotide bases, e.g., TN93 (Tamura and Nei 1993), HYK84 (Hasegawa et al. 1985), F84 (Felsenstein 1984), etc. [49].

(ii) Codon substitution models

The genetic code is a sequence of nucleotides that determines the formation of amino acids, the building blocks of proteins. The genetic code consists of 64 possible combinations of three adjacent nucleotides, which form a unit known as the codon. 61 out of 64 codons, called sense codons, correspond to one

of the 20 amino acids during protein synthesis, while the remaining are stop signals.

Just like nucleotide substitution models discussed above, Markov chains can be used to describe the substitutions that occur in codons which constitute protein-coding genes. Two types of substitutions may occur: a nucleotide substitution in a codon which does not change the amino acid being coded for, is called a *synonymous* substitution, while a substitution which causes a change in the encoded amino acid is a *nonsynonymous* substitution. The state space S of the continuous time Markov chain consists of the 61 sense codons. The entries q_{ij} of the rate matrix Q give the rate of change of codons from i to j, $i, j \in S$, $i \neq j$, as

$$q_{ij} = \begin{cases} 0, & \text{if } i \text{ and } j \text{ differ at two or more codon positions,} \\ \pi_j, & \text{if } i \text{ and } j \text{ differ by a synonymous transversion,} \\ \omega\pi_j, & \text{if } i \text{ and } j \text{ differ by a nonsynonymous transversion,} \\ \kappa\pi_j, & \text{if } i \text{ and } j \text{ differ by a synonymous transition,} \\ \kappa\omega\pi_j, & \text{if } i \text{ and } j \text{ differ by a nonsynonymous transition,} \end{cases}$$

where π_j is the equilibrium proportion of j, κ is the ratio of transitions to transversions, and ω is the ratio of nonsynonymous to synonymous substitutions [49]. Further, it is assumed that mutations occur independently and simultaneous changes at more than one positions do not occur. Different assumptions can be made about π_j, e.g., each codon could have the same frequency, $1/61$, or they could be estimated from the three sets of nucleotide frequencies for the three codon positions (F3 $\times$ 4 model), etc. The transition probability matrix $P(t)$ can be calculated using numerical methods [49].

Distances between protein-coding sequences are calculated separately for synonymous and nonsynonymous substitutions. One way to do this is by *counting methods* which are similar to distance calculation in nucleotide substitution models given above, in which synonymous and nonsynonymous sites as well as synonymous and nonsynonymous differences are first counted, and the proportions of differences is then calculated [49].

1.2 Diffusion models

So far, we have considered jump processes with constant sample paths, in which the given system enters a discrete state and after spending a certain amount of holding time in that state, it transitions to another state, and so on. We now explore diffusion processes — continuous state Markov processes for which the sample paths are continuous functions of time.

Stochastic differential equations (or SDEs) provide a representation of the dynamic behavior of continuous stochastic processes. An SDE for the process

$\{X_t : t \geq 0\}$ is given by

$$dX_t = a(X_t,t)\,dt + b(X_t,t)\,dB_t, \quad t \geq 0, \quad X_0 = x_0,$$

where B_t is the standard Brownian motion, and the functions $a(X_t,t)$ and $b(X_t,t)$ describe short-term growth and short-term variability, respectively. The solution of the SDE is obtained using the integral equation

$$X_t = X_0 + \int_0^t a(X_u,u)\,du + \int_0^t b(X_u,u)\,dB_u.$$

X_t is said to be a *strong solution* of the SDE if both integrals in the above stochastic equation exist for all $t > 0$ [18].

Diffusion processes: A diffusion process $\{X_t : t \geq 0\}$, whose state space S is an interval I over the real line, has continuous sample paths, and the mean and variance of the infinitesimal displacements in the process are given by the limits

$$\mu(x,t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{E}\left(X_{t+\Delta t} - X_t \mid X_t = x\right),$$

and

$$\sigma^2(x,t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{E}\left(\{X_{t+\Delta t} - X_t\}^2 \mid X_t = x\right),$$

respectively, for all $x \in I$. The functions $\mu(x,t)$ and $\sigma^2(x,t)$ are known as the infinitesimal mean or *drift* parameter and the infinitesimal variance or *diffusion* parameter of X_t, respectively. To visualize the diffusion process, consider a physical system with state X_t, $t \geq 0$, in which is introduced a random input B_t, the standard Brownian motion. Then, a small increment, dX_t, in X_t over a small time interval $(t, t+dt)$ can be expressed in terms of an SDE as

$$dX_t = \mu(x,t)\,dt + \sigma(x,t)\,dB_t,$$

where dB_t is an increment in B_t over the time interval $(t, t+dt)$ [21].

An example of a diffusion process over the interval $(-\infty, +\infty)$ is the Brownian motion, for which the drift parameter is 0 and the diffusion parameter is a positive constant σ^2. Another example is the *Brownian motion with drift*, in which both drift and diffusion parameters are constant functions, i.e., $\mu(x,t) = \mu$ and $\sigma^2(x,t) = \sigma^2$. The *mean reverting Ornstein-Uhlenbeck* process X_t is another diffusion process, which is the solution of the SDE

$$dX_t = \beta(\mu - X_t)\,dt + \sigma\,dB_t,$$

where β is the rate of mean reversion and μ is the constant long term mean. Furthermore, discrete state stochastic processes can also be approximated using continuous state diffusion models, a classical example of which is the Wright-Fisher diffusion model as described below.

Wright-Fisher diffusion model

Due to the incorporation of complex and more realistic biological features, such as mutation and selection, exact calculations of probabilistic functions under the discrete state WF model often become impossible. One solution is to approximate the model using a diffusion process, by which various complex properties of the WF model can be characterized by only two quantities: the mean change in the process and the variation around the mean value.

Consider a finite population of size N, with individuals carrying two types of alleles, type 0 and type 1. As before, assume that mutations between the two types occur with rates δ_{01} and δ_{10}, and the relative fitnesses of the two types are $1+s:1$. Recall that under the effect of mutation, the expected proportion of offspring carrying type 0 alleles, given that there are i type 0 alleles in the parent generation, is

$$p_i^{mut} = \frac{i}{N}(1-\delta_{01}) + \big(1-\frac{i}{N}\big)\delta_{10}.$$

Also, recall that due to selection, given i type 0 alleles in the parent generation, the expected fraction of offspring with type 0 alleles is

$$p_i^{sel} = \frac{(1+s)i}{N+si}.$$

For the construction of the WF diffusion process, the mutation rates and selection coefficient are scaled with N as

$$\delta_{01} = \frac{\rho_{01}}{N}, \quad \delta_{10} = \frac{\rho_{10}}{N}, \quad s = \frac{\gamma}{N}, \quad \rho_{01}, \rho_{10}, \gamma > 0.$$

Let X_n be the number of type 0 alleles in the nth generation, and let $X_{[Nt]}$ be the scaled version of the process at *generation time* $[Nt]$. Thus, the time is scaled in such a way, that in one unit $t=1$ of the process $X_{[Nt]}$, N generations have lapsed in the original process X_n. Moreover, let $\xi_t^N = N^{-1}X_{[Nt]}$ be the fraction of type 0 alleles in $[Nt]$ generations. In the limit $N\to\infty$, the process ξ_t^N converges to the WF diffusion process ξ_t [21].

Letting $\xi_{t+\Delta t}^N - \xi_t^N$ be the change in the proportion of type 0 alleles over a time interval of length $\Delta t = 1/N$, the infinitesimal mean and variance functions for the WF diffusion under mutation are obtained as

$$\begin{aligned}\mu(\xi,t) &= \lim_{N\to\infty} N\Big(p_i^{mut} - \frac{i}{N}\Big) = \lim_{N\to\infty}\Big(-\rho_{01}\frac{i}{N} + \rho_{10}\big(1-\frac{i}{N}\big)\Big)\\ &= -\rho_{01}\xi + \rho_{10}(1-\xi),\end{aligned}$$

and

$$\begin{aligned}\sigma^2(\xi,t) &= \lim_{N\to\infty} N\Big(\frac{1}{N}p_i^{mut}(1-p_i^{mut})\Big) = \lim_{N\to\infty}\Big(\frac{i}{N}\big(1-\frac{i}{N}\big) + O\big(\frac{i}{N}\big)\Big)\\ &= \xi(1-\xi),\end{aligned}$$

respectively. Similarly, the drift and diffusion parameters for the WF diffusion process under the effect of selection are obtained as

$$\begin{aligned}\mu(\xi,t) &= \lim_{N\to\infty} N\Big(p_i^{sel} - \frac{i}{N}\Big) = \lim_{N\to\infty}\Big(\gamma\frac{i}{N}(1-\frac{i}{N}) + O(\frac{i}{N})\Big)\\ &= \gamma\xi(1-\xi),\end{aligned}$$

and

$$\begin{aligned}\sigma^2(\xi,t) &= \lim_{N\to\infty} N\Big(\frac{1}{N}p_i^{sel}(1-p_i^{sel})\Big) = \lim_{N\to\infty}\Big(\frac{i}{N}(1-\frac{i}{N}) + O(\frac{i}{N})\Big)\\ &= \xi(1-\xi),\end{aligned}$$

respectively. Combining the above functions, the WF diffusion ξ_t is thus a solution of the SDE

$$d\xi_t = -\rho_{01}\xi_t\,dt + \rho_{10}(1-\xi_t)\,dt + \gamma\xi_t(1-\xi_t)\,dt + \sqrt{\xi_t(1-\xi_t)}\,dB_t,$$

where B is a Brownian motion [21].

2. Branching processes with applications to species diversification

In this section, we outline general properties of discrete time, continuous time, and continuous state Markov branching processes. We also provide applications of branching processes in the form of various species diversification models and density-dependent population growth models.

2.1 Discrete time branching processes

Consider the evolution of a population that starts with X_0 number of particles at time 0. After one unit of time, each particle independently produces a random number ξ of offspring according to the probability distribution

$$\mathbb{P}(\xi = k) = p_k, \quad k = 0,1,2,\ldots, \quad p_k \geq 0, \quad \sum_0^\infty p_k = 1.$$

Hence, in the first generation, the total number X_1 of particles is the sum of independent observations of ξ. Similarly, a second generation of X_2 particles is produced, independently of each other and independent of particles already existing, and so on. Let X_n denote the total number of particles in the nth generation, and note that if $X_n = 0$, then $X_{n+m} = 0$ for all $m \geq 0$. The process $\{X_n : n = 0,1,2,\ldots\}$ is called a *discrete time Markov branching process* or the *Galton-Watson process* [2, 20], with transition probabilities

$$p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i) = \begin{cases} \mathbb{P}(\xi_1 + \ldots + \xi_i = j) & \text{if} \quad i \geq 1,\ j \geq 0, \\ 0 & \text{if} \quad i = 0,\ j > 0, \\ 1 & \text{if} \quad i = 0,\ j = 0, \end{cases}$$

where $\{\xi_k\}_{k=1}^i$ are the numbers of particles produced independently by the i individuals in generation n. A key feature of this process is the additive property; a branching process with j initial particles, $j \geq 1$, is equal to summing up j independent branching processes for which $X_0 = 1$ [2]. Figure 2.1 is a diagram representing a discrete time Markov branching process.

Generating function: Let

$$f(s) = \sum_{j=0}^\infty p_j s^j, \quad |s| \leq 1,$$

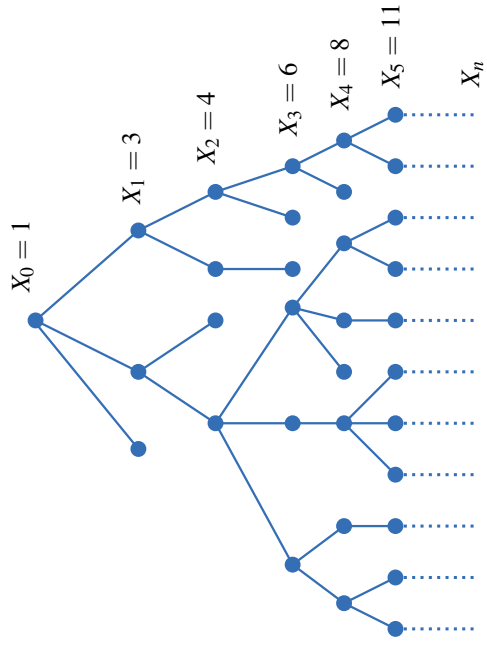

Figure 2.1. Pictorial representation of the Galton-Watson process.

be the generating function associated with ξ, and observe that $\sum_{j=0}^{\infty} p_{1j} s^j = \sum_{j=0}^{\infty} p_j s^j = f(s)$. Since ξ_k, $k = 1, \ldots, i$, are independent and identically distributed random variables with a common probability generating function $f(s)$, then using the property that the generating function of a sum of independent integer valued random variables is equal to the product of their generating functions, $[f(s)]^i$ is the generating function corresponding to $\xi_1 + \ldots + \xi_i$, that is,

$$\sum_{j=0}^{\infty} p_{ij} s^j = [f(s)]^i, \quad i \geq 1, \quad |s| \leq 1.$$

Hence, p_{ij} is the coefficient of s^j in the power series expansion of $[f(s)]^i$ [20].

Mean values and extinction probability: The expected value of the branching process $\{X_n : n = 0, 1, 2, \ldots\}$ is given by the derivative of $f(s)$ at $s = 1$, i.e.,

$$\mathbb{E}(X_n) = [f'(1)]^n = m^n,$$

where $m = f'(1) = \sum_{j=0}^{\infty} p_{1j} j = \mathbb{E}(X_1)$ [2].

The extinction probability of $\{X_n\}$ is the probability that the population will eventually go extinct ($\mathbb{P}\{X_n = 0\}$ for some n), and is given by the smallest nonnegative root, q, of the equation

$$f(s) = s.$$

q is 1 if $m \leq 1$ and less than 1 if $m > 1$. Hence, the population definitely dies out if the mean number of offspring produced per individual does not exceed 1. Moreover, $\mathbb{P}\{X_n = k\} = 0$ as $n \to \infty$ for $k \geq 1$, which implies that $X_n \xrightarrow{a.s.} 0$ or $X_n \xrightarrow{a.s.} \infty$. The branching process X_n is called *supercritical* if $m > 1$, *critical* if $m = 1$, and *subcritical* if $m < 1$ [2].

Multitype branching process: We can extend the above discrete time Markov branching process from one dimension to r dimensions as follows. Let $\mathscr{E}_r$ be the r dimensional Euclidean space and let

$$\mathbb{Z}_r^+ = \{(x_1, \ldots, x_r) : x_i \in \mathbb{Z},\ x_i \geq 0,\ i = 1, \ldots, r\},$$

i.e., the set of all points in $\mathscr{E}_r$ with nonnegative integer coordinates. Also, let $\mathscr{C}_r = \{(x_1, \ldots, x_r) : 0 \leq x_i \leq 1,\ i = 1, \ldots, r\}$. The probability of production of particles in an r-type process is given by

$$\boldsymbol{p_j} = \left(p_{\boldsymbol{j}}^{(1)}, \ldots, p_{\boldsymbol{j}}^{(r)}\right), \quad \boldsymbol{j} \in \mathbb{Z}_r^+, \quad \sum_{\boldsymbol{j}} p_{\boldsymbol{j}}^{(i)} = 1,\ \forall\, i = 1, \ldots, r.$$

Here, each $p_{\boldsymbol{j}}^{(i)} = p_{j_1,\ldots,j_r}^{(i)}$ gives the probability of a type i particle creating j_1 type 1 offspring, j_2 type 2 offspring, $\ldots$, j_r type r offspring. The associated generating function is given as

$$\boldsymbol{f}(\boldsymbol{s}) = \left(f^{(1)}(\boldsymbol{s}), \ldots, f^{(r)}(\boldsymbol{s})\right) = \sum_{\boldsymbol{j} \in \mathbb{Z}_r^+} \boldsymbol{p_j} \boldsymbol{s}^{\boldsymbol{j}}, \quad \boldsymbol{s} = (s_1, \ldots, s_r) \in \mathscr{C}_r,$$

where each

$$f^{(i)}(\boldsymbol{s}) = \sum_{j_1,\ldots,j_r \geq 0} p_{j_1,\ldots,j_r}^{(i)} s_1^{j_1} \cdot \ldots \cdot s_r^{j_r}, \quad i = 1, \ldots, r,$$

determines the distribution of the number of various types of offspring produced by a type i particle [2].

Now, let $\boldsymbol{X}_n = \left(X_n^{(1)}, \ldots, X_n^{(r)}\right)$ be a Markov chain on $\mathbb{Z}_r^+$, where each $X_n^{(i)}$, $i = 1, \ldots, r$, denotes the number of type i particles in the nth generation. Then, $\{\boldsymbol{X}_n : n = 0, 1, 2, \ldots\}$ is a *multitype (r-type), discrete time Markov branching process*, or the *multitype Galton-Watson process*, with transition probabilities given by

$$p_{\boldsymbol{ij}} = \mathbb{P}(\boldsymbol{X}_{n+1} = \boldsymbol{j} | \boldsymbol{X}_n = \boldsymbol{i}), \quad \boldsymbol{i}, \boldsymbol{j} \in \mathbb{Z}_r^+.$$

In terms of generating functions, $p_{\boldsymbol{ij}}$ is the coefficient of $\boldsymbol{s}^{\boldsymbol{j}}$ in $[\boldsymbol{f}(\boldsymbol{s})]^{\boldsymbol{i}}$ [2]. Figure 2.2 provides a diagram of a two-type, discrete time Markov branching process. We will discuss further properties of the multidimensional Markov branching processes for the continuous time case in the coming section.

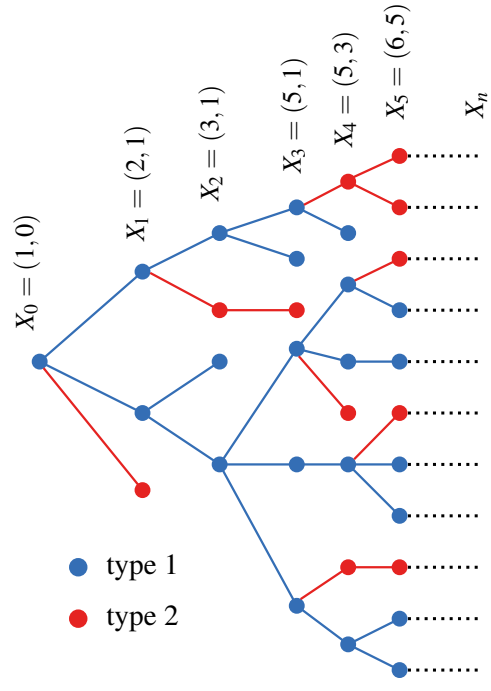

Figure 2.2. Pictorial representation of a two-type Galton-Watson process.

2.2 Continuous time branching processes

Consider a stochastic process $\{X_t : t \geq 0\}$ where X_t denotes the number of particles at time t. Let the life times of each particle be independent of one another and exponentially distributed random variables. Then X_t is said to be a *one dimensional, continuous time Markov branching process* [2] if

a) X_t is a continuous time Markov chain on the set of nonnegative integers.

b) The transition probabilities $p_{ij}(t)$ satisfy the property

$$\sum_{j=0}^{\infty} p_{ij}(t)s^j = \Big[\sum_{j=0}^{\infty} p_{1j}(t)s^j\Big]^i, \quad i \geq 0, \quad |s| \leq 1.$$

Without going into the properties of the above branching process, we skip directly to its generalization to r dimensions. A stochastic process $\boldsymbol{X}_t = \big(X_1(t),\ldots,X_r(t)\big)$, where each $X_i(t)$, $i = 1,\ldots,r$, represents the number of type i particles at time t, $t \geq 0$, is a *multitype, continuous time Markov branching process* if

a) $\boldsymbol{X}_t$ is a continuous time Markov chain on $\mathbb{Z}_r^+$, and

b) the transition probabilities $p_{\boldsymbol{ij}}(t)$ satisfy

$$\sum_{\boldsymbol{j}\in\mathbb{Z}_r^+} p_{\boldsymbol{ij}}(t)\boldsymbol{s}^{\boldsymbol{j}} = \prod_{k=1}^{r}\Big[\sum_{\boldsymbol{j}\in\mathbb{Z}_r^+} p_{\boldsymbol{e}_k\boldsymbol{j}}(t)\boldsymbol{s}^{\boldsymbol{j}}\Big]^{i_k},$$

where $\boldsymbol{i} = (i_1, \ldots, i_r) \in \mathbb{Z}_r^+$, $\boldsymbol{s} = (s_1, \ldots, s_r) \in \mathscr{C}_r$, and $\boldsymbol{e}_k = (0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 being the kth component [2].

Generating function: We first define an infinitesimal parameter $\boldsymbol{a}$ as

$$\boldsymbol{a} = (a_1, \ldots, a_r), \quad a_i \geq 0, \quad i = 1, \ldots, r,$$

where a_i represents the exponentially distributed life length of type i. Each type i produces offspring of r types according to a distribution given by coordinates of the vector

$$\boldsymbol{p_j} = \left(p_{\boldsymbol{j}}^{(1)}, \ldots, p_{\boldsymbol{j}}^{(r)}\right), \quad \sum_{\boldsymbol{j} \in \mathbb{Z}_r^+} p_{\boldsymbol{j}}^{(i)} = 1, \ i = 1, \ldots, r,$$

where $p_{\boldsymbol{j}}^{(i)}$ gives the probability of a type i particle creating j_1 type 1 offspring, j_2 type 2 offspring, and so on. All particles produce independently of each other and of past events. Similar to the discrete time multitype branching process, the generating function is given as

$$\boldsymbol{f}(\boldsymbol{s}) = \left(f^{(1)}(\boldsymbol{s}), \ldots, f^{(r)}(\boldsymbol{s})\right), \quad \boldsymbol{s} \in \mathscr{C}_r,$$

where $f^{(i)}(\boldsymbol{s}) = \sum_{\boldsymbol{j}} p_{\boldsymbol{j}}^{(i)} \boldsymbol{s}^{\boldsymbol{j}}$, $i = 1, \ldots, r$, determines the distribution of the number of offspring of various types to be produced by a type i particle. Further, the infinitesimal generating function $\boldsymbol{u}(\boldsymbol{s})$ is given by

$$\boldsymbol{u}(\boldsymbol{s}) = \left(u^{(1)}(\boldsymbol{s}), \ldots, u^{(r)}(\boldsymbol{s})\right), \quad \boldsymbol{s} = (s_1, \ldots, s_r) \in \mathscr{C}_r,$$

with $u^{(i)}(\boldsymbol{s}) = a_i\left(f^{(i)}(\boldsymbol{s}) - s_i\right)$, $i = 1, \ldots, r$ [2].

Expected values: The mean matrix of the branching process is given as $M(t) = \{m_{ij}(t) : i, j = 1, \ldots, r\}$, where

$$m_{ij}(t) = \mathbb{E}[X_j(t) | X_i(t) = 1] < \infty.$$

There exists a matrix $A = \{a_{ij} : i, j = 1, \ldots, r\}$, which we call the *mean offspring matrix*, such that

$$a_{ij} = a_i\Big(\frac{\partial f^{(i)}(\boldsymbol{s})}{\partial s_j}\Big|_{\boldsymbol{s}=(1,\ldots,1)} - \delta_{ij}\Big), \quad \delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{otherwise,} \end{cases}$$

and

$$M(t) = e^{At} = \sum_{q=0}^{\infty} \frac{A^q t^q}{q!}.$$

The process $\boldsymbol{X}_t$ is said to be *positively regular* if for some t_0 $(0 < t_0 < \infty)$, $m_{ij}(t_0) > 0 \ \forall \ i, j$. Then, a positive and simple eigenvalue $\rho(t_0)$ of $M(t_0)$ exists, whose magnitude is larger than all the other eigenvalues. The eigenvalues of $M(t)$ have the form $e^{\lambda_i t}$, where λ_i, $i = 1, \ldots, r$, are eigenvalues of A and classify the branching process as

$$\begin{cases} \text{supercritical} & \text{if} \quad \lambda_1 > 1, \\ \text{critical} & \text{if} \quad \lambda_1 = 1, \\ \text{subcritical} & \text{if} \quad \lambda_1 < 1, \end{cases}$$

with λ_1 being the largest eigenvalue of A [2].

Extinction probability: The extinction probability is given by $\boldsymbol{q} = (q^{(1)}, \ldots, q^{(r)})$, where each $q^{(i)}$ is the minimal solution of the equation

$$\boldsymbol{f}(\boldsymbol{s}) = \boldsymbol{s} \quad \text{or} \quad \boldsymbol{u}(\boldsymbol{s}) = \boldsymbol{0} = (0, \ldots, 0).$$

As in the discrete case, the branching process $\boldsymbol{X}_t \to \infty$ or $\boldsymbol{X}_t \to \boldsymbol{0}$, with probability 1, as $t \to \infty$ [2]. Contrary to the predicted extinction behavior of branching processes, in reality it is often seen that finite biological populations tend to reach a state of balance over long periods of time. In Papers III and IV, we show that such populations can be modeled by using branching processes which have been modified to include population size dependence.

2.3 Continuous state branching processes

In the previous sections, we described discrete time as well as continuous time branching processes, whose state space was always the discrete set of nonnegative integers. Here, we focus on stochastic branching processes with continuous state space, i.e., the whole nonnegative part of the Euclidean space.

Let $(E, \mathscr{E})$ be a measurable space, i.e., E is a set and $\mathscr{E}$ is a σ-algebra on E. The function $P_{s,t}(x,A)$, $A \in \mathscr{E}$, $x \in E$, $0 \leq s < t < \infty$, is called a *Markov transition function* on $(E, \mathscr{E})$ if for a fixed s, t and x, $A \mapsto P_{s,t}(x,A)$ is a probability measure on $(E, \mathscr{E})$, for a fixed A, s and t, $x \mapsto P_{s,t}(x,A)$ is $\mathscr{E}$-measurable, and for $0 \leq s < u < t$, $P_{s,t}(x,A) = \int_E P_{s,u}(x,dy) P_{u,t}(y,A)$ [8].

Definition 2.3.1. Let $(E, \mathscr{B})$ be a measurable space, where $E = [0, \infty)$ and $\mathscr{B}$ is the Borel σ-algebra on E. A *continuous branching function* is a family $\{P_t(x,B)\}$ of functions, with $B \in \mathscr{B}$, $x \in E$, $t \geq 0$, which satisfies the following conditions.

a) It is a Markov transition function on the Borel sets of $[0, \infty)$, such that $P_t\big(x, [0, \infty)\big) = 1$.
b) There exist some $t > 0$ and $x > 0$ such that $P_t(x, \{0\}) < 1$, i.e., the trivial case that the process starting at some x instantly dies, is ruled out.
c) The Markov transition function satisfies the following property: for any $x, y \geq 0$ and $t \geq 0$,

$$P_t(x+y, B) = \int P_t(x, B-u) P_t(y, du)$$

for each $B \in \mathscr{B}$. Hence, starting the process at $x+y$ is the same as the sum of two independent processes starting at x and at y [2, 24].

A *continuous state branching process* $\{X_t : t \geq 0\}$ is a Markov process on $[0,\infty)$ with transition probabilities given by a continuous branching function. ■

Some classes of continuous state branching processes can be constructed as limits of sequences of discrete time Markov branching processes as follows. Let $\{X_n : n = 0,1,2,\ldots\}$ be a discrete time Markov branching process and consider a sequence X_n^N, $N = 1,2,\ldots$, of such processes, with $X_0^N = N$, and offspring distribution given by $\mathbb{P}(\xi_N = k) = p_k^N$, $k = 0,1,2,\ldots$. Let the generating function $f_N(s)$ of X_n^N depend upon N in such a way, so that

$$\mathbb{E}(X_1^N) = f_N'(1) = 1 + \frac{\alpha_N}{N}, \quad \mathrm{Var}(X_1^N) = \beta_N,$$

where $\alpha_N \to \alpha \in \mathbb{R}$ and $\beta_N \to \beta > 0$ as $N \to \infty$. Next, consider the continuous time process $X_{[Nt]}^N$ scaled over Nt generations and let

$$Y_t^N = \frac{1}{N} X_{[Nt]}^N, \quad 0 \leq t < \infty.$$

As $N \to \infty$, Y_t^N converges to the continuous state branching process Y_t, which is the solution of the SDE

$$dY_t = \alpha Y_t\, dt + \sqrt{\beta Y_t}\, dB_t, \quad t \geq 0,$$

where B_t is a Brownian motion. While the original discrete process X_n gave the number of individuals in each generation, we can intuitively think of the continuous limit Y_t representing a process in which infinitely many births and deaths occur constantly [34]. The above SDE, representing the continuous approximation of the discrete Galton-Watson branching process, is known as the *Feller* diffusion equation [15]. These types of equations are discussed further in Paper IV.

2.4 Trait-dependent models

Discrete traits, such as selfing and outcrossing mating systems in flowering plants, and continuous traits, such as body size in mammals, have been suggested to affect the rates of diversification — the difference between birth and death rates — in species [12]. On the other hand, phylogenetic trees contain patterns of diversification and can help understand the changes in speciation and extinction rates, as well as the rates of transition between character states of a species. In recent years, a number of continuous time branching models have been developed that use information from phylogenetic trees and maximum likelihood analysis to study the effect of traits on species diversification. In this section, we review some of these models, such as the BiSSE [29], MuSSE [11], and ClaSSE [16] models, which in short we will call the SE — speciation and extinction — models. The SE models provide a parameter estimation method to test for correlations between phenotypic traits and

diversification rates in lineages and to determine if rates of character change in species depend on the character state or not. In contrast to earlier studies where the shape of a phylogenetic tree was considered fixed and characters were assumed to evolve independently along branches of the fixed tree, the SE models can simultaneously predict a phylogeny and the evolution of character states which differ in speciation and extinction rates [27].

(i) Binary state speciation and extinction model

A two-type Markov branching model, named the 'binary state speciation and extinction' (BiSSE) model, was proposed by Maddison et al. [29] to simultaneously study character change between binary characters, 0 and 1, and assess their impact on diversification rates. This model examines the effect of two characters on birth and death rates, while at the same time, it accounts for possible transitions between the two states. The model has six parameters: λ_0, λ_1 are the birth rates and μ_0, μ_1 are the death rates for type 0 and type 1 species, respectively, while q_{01} and q_{10} represent the rates of character change from type 0 to type 1 and vice versa, respectively. If we let K_t to be the number of type 0 species and L_t the number of type 1 species at $t \geq 0$, the BiSSE model is in fact a two-type, continuous time Markov branching process with branching rates

$$(k,\ell) \mapsto \begin{cases} (k+1,\ell) & \lambda_0 k \\ (k-1,\ell) & \mu_0 k \\ (k,\ell+1) & \lambda_1 \ell \\ (k,\ell-1) & \mu_1 \ell \\ (k-1,\ell+1) & q_{01} k \\ (k+1,\ell-1) & q_{10} \ell. \end{cases}$$

A diagrammatic representation of the BiSSE model is given in Figure 2.3.

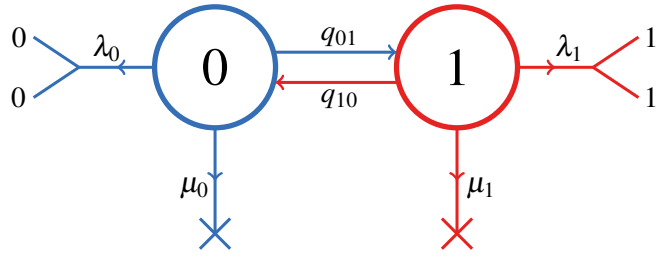

Figure 2.3. The BiSSE model.

For the parameter estimation analysis in the BiSSE model, it is assumed in [29] that a phylogenetic tree with known branch lengths and character states of all extant species is provided. A likelihood analysis is then carried out, i.e., given the model with six parameters, the probability D that a lineage would evolve into a tree, which is identical to the observed tree, is calculated. In order

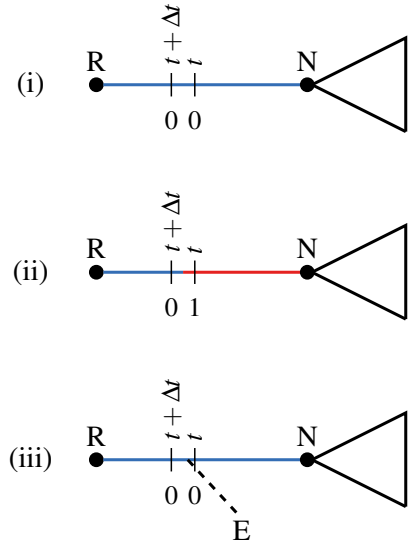

Figure 2.4. Events that can happen along a branch in the BiSSE model, with state 0 at time $t+\Delta t$: (i) no event takes place, (ii) character change occurs, (iii) extinction of the right lineage (or the left one) occurs, denoted by E. Here, R represents the root of the tree and N denotes the node from which the observed clade originates.

to achieve this, the probability of an event — such as state change, speciation, extinction — that may have happened in a small time interval, $(t, t+\Delta t)$, is computed. A diagrammatic representation of all possible events is shown in Figure 2.4. The calculations move backward in time, that is, from the tips towards the root of the tree. The probabilities are obtained along the branches, at the nodes, and at the root of the tree, and then added up. Letting $\Delta t \to 0$, a system of ODEs is formulated that can be solved numerically, and parameter estimates, which maximize the likelihood, are obtained. In practice, the BiSSE model can be employed using the function 'make.bisse()' in the *diversitree* package [13] of `R` [38].

The BiSSE model has been used extensively in recent years to analyze the process of diversification in various biological scenarios. For example, the authors of [23] used the BiSSE framework to study diversification of bellflowers in the mountains of South America and showed that species occurring at high elevations had higher speciation rates as compared to species at low elevation. This model was also employed in [37] to study the diversification patterns of the *Nymphalidae* butterfly family feeding on the *Solanaceae* plant family, and in [40], to show that speciation rates of the Caribbean lizard of genus *Anolis* have declined independently on various islands of the West Indies. The BiSSE model is utilized in Paper I of this thesis to obtain diversification rates for a collection of selfing and outcrossing plant species in the Geraniaceae family, and to illustrate the effect of diversification process on the estimates of molec-

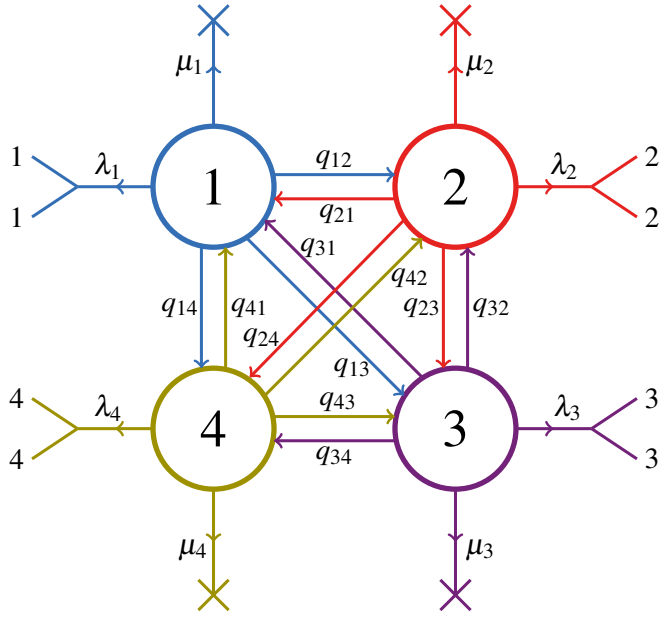

Figure 2.5. The MuSSE model with 4 states.

ular evolutionary rates. Apart from making inferences from phylogenetic data, we also study general mathematical properties of binary branching and asymmetric transition models in Paper I, that have not been explored earlier.

(ii) Multistate speciation and extinction model

The multistate speciation and extinction (MuSSE) model was introduced by FitzJohn [11] as an extension of the BiSSE branching model to discrete traits with multiple character states. Figure 2.5 shows a diagrammatic representation of a MuSSE model with four states, in which λ_i are the speciation rates and μ_i are the extinction rates of type i individuals, $i = 1 \ldots 4$, while q_{ij} are the rates of character change from type i to type j, $i, j = 1 \ldots 4$. The parameter estimates of this model are obtained in the same way as in the BiSSE model, and it can be accessed through the *diversitree* package [13] of `R` using the function 'make.musse()'. The MuSSE model has also been used considerably to test the impact of multiple traits on diversification rates. For instance, it was used to study the correlation between dioecious and monoecious species of conifers in [25], to disprove the hypothesis that evolution of a high number of vertebrae in reptiles allows for greater diversification regarding taxonomy and body shape in [7], and in [47], a 4-type MuSSE model was used to study patterns of diversification in speciation rates of *Exocelina* beetles in four different regions of the world. In Paper II of this thesis, various versions of a MuSSE model with 4 states and 10 parameters were employed to examine pathogenic behavior in a clinical data set [4] of *E. coli* bacterial strains.

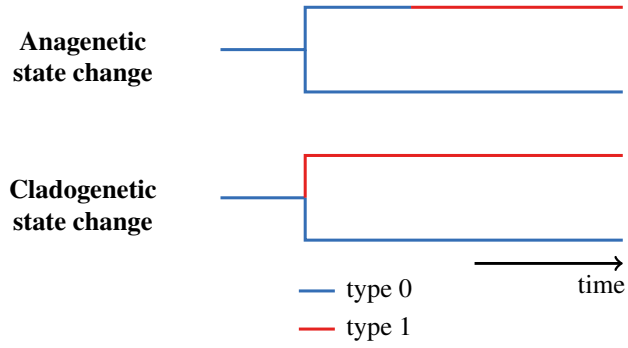

Figure 2.6. Pictorial representation of anagenetic and cladogenetic character change from type 0 to type 1 on a branch of a phylogenetic tree running in continuous time.

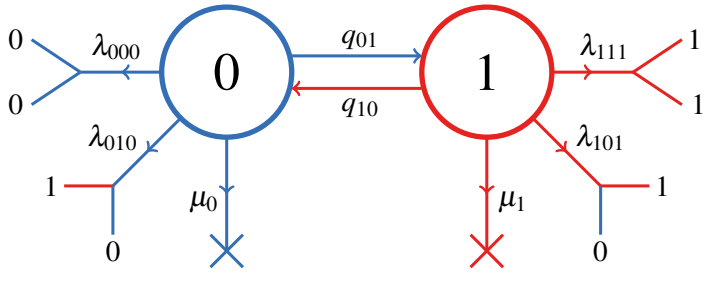

Figure 2.7. The ClaSSE model.

(iii) Cladogenetic state speciation and extinction model

The cladogenetic state speciation and extinction, or ClaSSE, model was introduced by Goldberg and Igic [16] as a modified version of the BiSSE model with different modes of character change. While the authors of the BiSSE model considered transitions between the two types to be anagenetic, that is, character change in a phylogenetic tree (from type 0 to type 1 and vice versa) occurs instantaneously along a branch, the ClaSSE model takes into account cladogenetic transitions as well — when a type 0 or type 1 species is replaced by a combination of both types, 0 and 1, as illustrated in Figure 2.6. The ClaSSE model is basically a two-type, continuous time Markov branching process with branching rates depicted in Figure 2.7, in which λ_{000} and λ_{111} are the speciation rates of the two types, μ_0 and μ_1 are the corresponding extinction rates, q_{01} and q_{10} are the rates of anagenetic state change, and λ_{010} and λ_{101} are the rates of cladogenetic state change. Similar to previous SE models, the ClaSSE model can be accessed through the package *diversitree* in `R` [38], using the functions 'make.classe()' and 'find.mle()'. This model has been widely used in recent research, see for example [3, 22, 44]. Paper I of this thesis utilizes the ClaSSE model to analyze diversification patterns in outcrossing and selfing plant species.

(iv) Limitations of the SE models

Apart from the models described above, many more derivatives of the BiSSE model have been developed, such as the GeoSSE model [17] for estimating region-dependent diversification rates associated with geographic character states, and the QuaSSE model [12] to accommodate character evolution for continuous traits, among others. Numerous studies have been conducted on the application of the SE models to various data sets in order to study the effects of traits on species diversification. However, doubts have been raised regarding the accuracy of the SE methods, and much research has been carried out to explore the statistical power of these parameter estimation methods, especially the original BiSSE model [26]. Some studies show that the power of BiSSE analysis is affected by low sample size and a high tip ratio bias (when one character state is more frequent among the tips of the phylogeny), thus, to get unbiased results, trees should be of reasonably large size and the number of tips should be more than 300 [9]. Other studies predict that the power of BiSSE and related models does not depend on the sample size, instead it is the assumption of the root state of the tree that effects the final results [14]. It is shown that decreasing the number of parameters in the analyses and letting parameters equal to one another, increases the power of estimation. In fact, the author of the MuSSE model and the *diversitree* package of R warns about the power of estimating transition rates in the MuSSE model: 'With more than 9 states, q_{ij} can be ambiguous' [13].

Depending on the shape of the phylogenetic tree, the SE models may falsely predict a neutral trait having an effect on the dynamics of diversification [42, 43]. It is also possible that only one clade has a high diversification rate associated with the trait under consideration, but the diversification is so strong, that it is enough to return high rates for the whole tree [5]. To prevent false positives and to account for any underlying unmeasured factors that could have an effect on diversification, the hidden state speciation and extinction (HiSSE) model was formulated [5]. This model assumes that related to each observed state, there is a hidden trait in the model, which is assigned an unknown state in all the tips of the tree. In this way, any effects on diversification could be attributed to the hidden trait and would divert focus from the neutral trait [42].

Another issue with the SE models is that of *pseudoreplication*, i.e., to infer major results when in reality only few evolutionary changes have occurred in a character involving the trait under consideration [28, 43]. A solution is to perform pairwise comparisons; selecting many pairs of species and then checking if the difference in one character consistently predicts a difference in a second character [28]. A model named as 'fast, intuitive state-dependent speciation extinction' (FiSSE) [41] was introduced recently, to provide a statistical test to analyze binary trait-dependent diversification rates in phylogenies. FiSSE does not require the use of parameters or an underlying model for species diversification, instead, it provides estimates of the so called quasi-

parameters (which can be related to speciation rates in the previous SE models) that provide distributions for branch lengths associated with each character state. Power of the estimated quasi-parameters is judged by comparing the observed values to the values obtained from a simulated null distribution [41].

The limitations of the SE models listed above are mainly statistical in nature and related to problem of inference. There is however a limitation of the models themselves; the species are allowed to grow supercritically, without any bound on their total sizes. To overcome this issue, we now describe diversity-dependent processes, in which population numbers increase only up to a certain limit and eventually reach a stationary state.

2.5 Diversity-dependent models

All over the natural world, there exists a vast variation in the diversity of species. Species diversification is not an easy concept to study; since speciation and extinction processes occur on the pylogenetic time scale of millions of years, it is difficult to estimate diversification rates from fossil data [31]. Hence the nature and causes of the increase in species diversity are debated, and has led to various conflicting view points. According to one belief, variance in species richness is due to variation in speciation and extinction rates [39]. Another reason given is that evolution causes new species to occupy uninhabited niches and adapt in new environments [6]. There exists a hypothesis which proposes that expansionist growth models can effectively describe the diversification of families of species. On the other hand, it is also widely believed that there exists the concept of a *carrying capacity*; species diversity can increase only up to a point where it can be supported within a given niche space [6, 30].

In accordance with the latter theory, we discuss here the concept of diversity dependent diversification that provides a method of modulating a populations' increasing numbers. It describes the effect of competition on speciation and extinction rates, which may be thought of as 'macroevolutionary' rates over a long geological timescale [39]. For this type of a model, we are interested in a situation in which net growth is regulated by the total size of the system. As an example, we define a supercritical, continuous time Markov branching model $\{K_t : t \geq 0\}$, where K_t denotes the total number of species at time t, and $K_0 = 1$. Let λ and μ be the rates of speciation and extinction of species, respectively, such that the net diversification rate $\lambda - \mu > 0$. To introduce diversity-dependence in the model, λ is replaced by $\lambda(1 - k/c)$, under the condition that $(1 - k/c) = 0$ if $k \geq c$, where c represents the carrying capacity. Let $\bar{k}$ be the equilibrium diversity, that is, the number of species when the birth rate equals the death rate [39]. Thus, letting $\lambda(1 - k/c) = \mu$, $\bar{k} = c(1 - \mu/\lambda)$ defines the equilibrium diversity for the model. Figure 2.8 shows the effect of density dependence on speciation and extinction rates of the species.

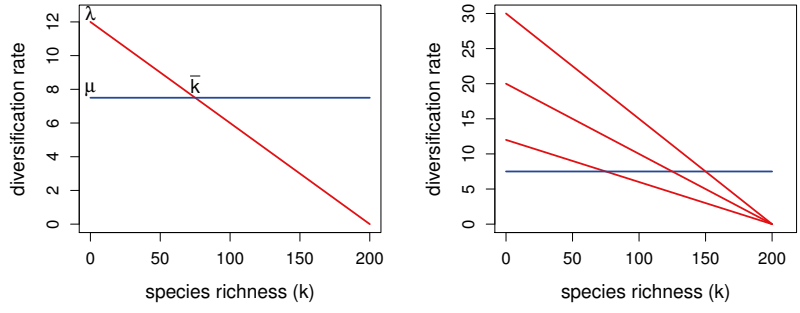

Figure 2.8. Left: plot of λ and μ versus species richness k, for $\lambda = 12$, $\mu = 7.5$ and $c = 200$, with $\bar{k}$ being the equilibrium diversity. Right: plot of diversification rates versus k for $\lambda = 12, 20, 30$ (red lines), $\mu = 7.5$ (blue line), and $c = 200$, showing that an increase in λ, increases the rate of approach towards the equilibrium diversity.

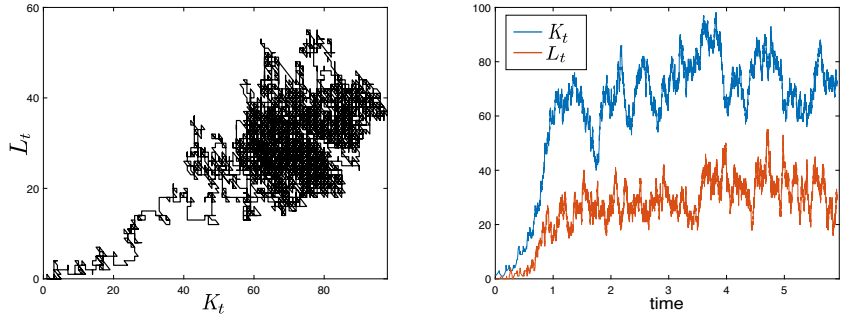

Figure 2.9. A simulation of the branching process X_t with $\lambda_0 = 12$, $\lambda_1 = 8$, $\mu_0 = 2.5$, $\mu_1 = 20$, $\delta = 5$, and $c = 200$. The left plot shows a trace of K_t and L_t, while the right plot gives the corresponding path of K_t and L_t versus time. It can be seen that the species numbers increase initially, but later reach a 'quasi stationary' state.

We now extend the above model to a two-type, continuous time Markov branching model, $X_t = (K_t, L_t)$, $t \geq 0$, where K_t is number of type 0 species, L_t is the number of type 1 species at time t, and $X_0 = (1,0)$. This is similar to the model studied in Papers I and III. The speciation rates for type 0 and type 1 species are λ_0 and λ_1, while the death rates are μ_0 and μ_1, respectively. Moreover, type 0 species are allowed to transition to type 1 species with rate δ. The net growth rates of type 0 and type 1 species are $\gamma_0 = \lambda_0 - \mu_0 - \delta > 0$ and $\gamma_1 = \lambda_1 - \mu_1 < \gamma_0$, respectively. We again consider a logistic type model in which density-dependence is assumed only on the speciation parameter of type 0 species, hence λ_0 is replaced by $\lambda_0(1 - k/c)$, while the remaining parameters remain unchanged. Figure 2.9 shows a simulation of the branching process for arbitrary parameter values.

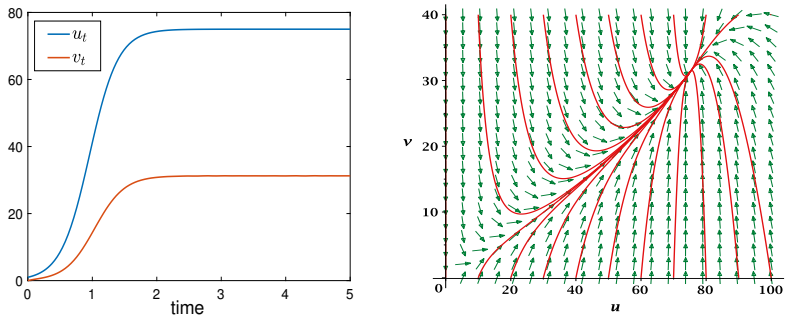

Figure 2.10. Plots showing solutions (left) and a phase portrait (right) of the ODE system when $\lambda_0 = 12$, $\delta = 5$, $\gamma_0 = 4.5$, and $\gamma_1 = -12$.

We can obtain a deterministic approximation of the branching process as follows. Consider the process $X_t^m = (K_t^m, L_t^m)$, which starts at time $t = 0$ with m type 0 species. This is equivalent to summing up m i.i.d copies of the original model X_t, which starts with one type 0 species. Define $Y_t^m = (U_t^m, V_t^m)$, $t \geq 0$, where $U_t^m = K_t^m/m$ and $V_t^m = L_t^m/m$. As $m \to \infty$, $Y_t^m = (U_t^m, V_t^m) \xrightarrow{a.s.} \mathbb{E}(X_t) = (u_t, v_t)$, which is given by the solution of the ODE system

$$\frac{d}{dt}\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} \gamma_0 - \lambda_0 u_t/c & 0 \\ \delta & \gamma_1 \end{pmatrix}\begin{pmatrix} u_t \\ v_t \end{pmatrix}, \quad u_0 = 1, \quad v_0 = 0.$$

For arbitrary parameter values, solutions of the above system are represented in Figure 2.10 (left panel). By letting $du_t/dt = 0$ and $dv_t/dt = 0$, the equilibrium solutions of the ODE system are obtained as $(\hat{u}, \hat{v}) = (0, 0)$ and $(\bar{u}, \bar{v}) = (c\gamma_0/\lambda_0, -c\delta\gamma_0/\gamma_1\lambda_0)$, where the latter is a positive equilibrium since $\gamma_1 < 0$. The Jacobian matrix J of the system is given as

$$J = \begin{pmatrix} \gamma_0 - 2\lambda_0 u_t/c & 0 \\ \delta & \gamma_1 \end{pmatrix}.$$

At the equilibrium $(\hat{u}, \hat{v})$, J has eigenvalues $\gamma_0 > 0$ and $\gamma_1 < 0$, hence $(\hat{u}, \hat{v})$ is an unstable saddle point, while at $(\bar{u}, \bar{v})$, the matrix J has eigenvalues $-\gamma_0 < 0$ and $\gamma_1 < 0$, hence $(\bar{u}, \bar{v})$ is asymptotically stable. A phase diagram showing the behavior of solutions near the equilibrium points is given in Figure 2.10 (right panel). Further properties of the model, such as the description of fluctuations around the scaled deterministic limit, are provided in Paper III. In Paper IV, we further explore various density dependent models and compare them with population genetics models of allele frequencies.

3. Summary of papers

3.1 Paper I

In Paper I, we study trait evolution and trait-dependent diversification in families of species, as explored before in e.g., [16, 29], and we also link together these processes with trait-dependent molecular evolution. In order to achieve this, we propose a mathematical model that describes the evolution of binary traits on a random species tree, which runs from the time of origin, 0, to the time of observation, t. The two traits, 0 and 1, develop on the tree, with the emergence of new species, extinction of existing species, and asymmetric transition of species from type 0 to type 1, according to a supercritical, continuous time Markov branching process. Simultaneously, depending on their trait, the species also accumulate mutations over time according to a Poisson process. In the Paper, we also examine various characteristics of binary-trait species trees, such as the expected sizes and expected branch lengths of *reduced trees* — trees from which extinct lineages have been removed, and thus consist of only extant species at the time of observation.

The continuous time Markov branching model describing the evolution of traits is given by $X_t = (K_t, L_t)$, $t \geq 0$, where K_t and L_t represent the number of type 0 and type 1 species, respectively, at time t and $X_0 = (1,0)$. The branching rates of the asymmetric model are

$$(k,\ell) \mapsto \begin{cases} (k+1,\ell) & \lambda_0 k \\ (k-1,\ell+1) & (1-p)\delta k \\ (k-1,\ell) & \mu_0 k \\ (k,\ell+1) & p\delta k + \lambda_1 \ell \\ (k,\ell-1) & \mu_1 \ell, \end{cases}$$

where λ_0, λ_1 are the speciation rates, and μ_0, μ_1 are the extinction rates of type 0 and type 1 species, respectively, while $p\delta$ and $(1-p)\delta$ are the rates of cladogenetic and anagenetic transition, respectively, from type 0 to type 1 species. Since the model is analyzed under supercritical conditions, we assume $\gamma_0 = \lambda_0 - (1-p)\delta - \mu_0 > 0$, while $\gamma_1 = \lambda_1 - \mu_1 < \gamma_0$. A simulation of the two-type branching process X_t, under arbitrary parameter values, is given in Figure 3.1. A diagrammatic representation of a two-trait reduced species tree is provided in Figure 3.2 (i). The tree is composed of two parts, as depicted in Figure 3.2 (ii); a single reduced type 0 tree, and a type 1 tree which consists of only disjoint branches. Separate analyses of the reduced type 0 and type 1 species trees yields the following tree properties.

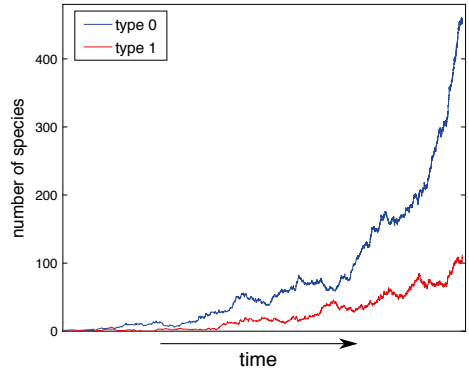

Figure 3.1. Simulation of the branching model X_t under supercritical conditions, with parameter values: $\lambda_0 = 9$, $\lambda_1 = 5$, $\mu_0 = 6$, $\mu_1 = 10$, $p = 0.5$ and $\delta = 2$. The growth of type 0 and type 1 species is represented in blue and red colors, respectively.

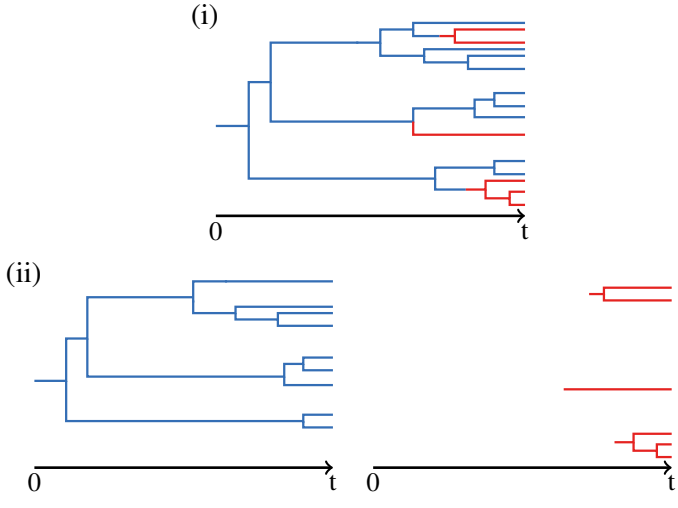

Figure 3.2. (i) A two-trait reduced species tree, with type 0 species represented in blue, and type 1 species in red. (ii) A reduced type 0 species tree on the left, and disjoint type 1 species clusters on the right.

a) Let A_t be the total branch length of the reduced type 0 species tree, corresponding to the blue colored tree in Figure 3.2 (ii). The expected value of A_t, conditional on nonextinction, is obtained as

$$\mathbb{E}(A_t|K_t > 0) = \int_0^t \frac{\lambda_0 e^{\gamma_0 t} - \mu_0 - (1-p)\delta}{\lambda_0 e^{\gamma_0 (t-s)} - \mu_0 - (1-p)\delta} ds.$$

b) Let B_t be the total branch length of the reduced type 1 species tree, corresponding to the red sub-trees in Figure 3.2 (ii). The expected value of B_t, conditional on nonextinction, is derived as

$$\mathbb{E}(B_t|K_t>0)=\delta\int_0^t\int_0^s\Big(\frac{\gamma_1 e^{\gamma_1(t-u)}}{\lambda_1 e^{\gamma_1(t-s)}-\mu_1}\Big)\times$$
$$\Big(\frac{\lambda_0 e^{\gamma_0 u}-\mu_0-(1-p)\delta}{\gamma_0}+\frac{\lambda_0(\mu_0+(1-p)\delta)(e^{\gamma_0 u}-1)(e^{\gamma_0(t-u)}-1)}{\gamma_0(\lambda_0 e^{\gamma_0 t}-\mu_0-(1-p)\delta)}\Big)duds.$$

Molecular evolution is introduced on the species tree by assuming that mutations accumulate through a Poisson process running along all branches of the tree. We let ω_0 and ω_1 be the rates of fixation of mutations in type 0 and type 1 species, respectively, with $\omega_0 \leq \omega_1 < 1$. Thus, for a species which carries trait 0 (trait 1) over a fixed time duration t, the value of dN/dS — defined as the normalized ratio of nonsynonymous to synonymous substitutions — is ω_0 (ω_1). An estimate of dN/dS over the whole species tree is given by

$$dN/dS \approx \frac{\omega_0\mathbb{E}(A_t|K_t>0)+\omega_1\mathbb{E}(B_t|K_t>0)}{\mathbb{E}(A_t|K_t>0)+\mathbb{E}(B_t|K_t>0)}.$$

We applied our binary branching and transition model on a real phylogenetic tree consisting of plant species from the Geraniaceae family, with two types of mating systems: outcrossing and selfing. From the analysis, we inferred the following results.

a) The value of dN/dS on the species tree increases initially, and after a sufficient amount of time has passed, reaches a limiting steady state.

b) Using estimates of the total branch length, T_{tot}, of the the phylogenetic tree, and $T^1_{\text{tot}} = \mathbb{E}(B_t|K_t>0)$, which can be calculated using estimates of the model parameters, an estimate for ω_1 is derived as

$$\omega_1 \approx \omega_0 + (dN/dS - \omega_0)\frac{T_{\text{tot}}}{T^1_{\text{tot}}},$$

where it is assumed that the estimates of ω_0 and dN/dS for the species tree have been found using existing methods (see e.g., [48]).

c) Trait-dependent diversification processes can have a strong impact on the estimates of molecular evolutionary rates, such as ω_1.

Contribution: I worked on developing the methodology, participated in writing the paper, performed the analysis of outcrossing/selfing plant species of the Geraniaceae family, and provided the simulations and illustrations.

3.2 Paper II

The motivation behind Paper II was to explore the possibility of utilizing the theory of branching processes into analyzing virulence in bacterial strains. For that purpose, we first reviewed fundamental properties of multitype, continuous time Markov branching processes as well as their behavior in the long time

limit. Then, we applied multitype branching models to examine pathogenicity in *E. coli* strains, and performed an in depth analysis on the limits of proportions of bacteria in different states of the models. The strains used in this study were isolated from human hosts, and obtained from a previously published [4] data set of pathogenic and nonpathogenic *E. coli* bacteria.

In this work, we survey n-type branching processes given by $X_t = \big(X_1(t), \ldots, X_n(t)\big)$, where each $X_i(t)$, $i = 1, \ldots, n$, denotes the number of type i particles at time t, $t \geq 0$. The limit theorems reviewed in the article, obtained from earlier works of Athreya and Ney [2] and Janson [19], describe how X_t behaves as $t \to \infty$. From the theorems, we infer that the limit behavior of the process can be completely characterized by the eigenvalues and corresponding eigenvectors of the mean offspring matrix of X_t. For the subsequent application to *E. coli* strains data, we formulate a 4-type branching model, since the data set we use comprises of bacterial strains divided into 4 categories: pathogenic and nonpathogenic bacteria in the intestine, and, pathogenic and nonpathogenic bacteria in the urinary tract. An *E. coli* bacterial strain is considered pathogenic only if it carries an agent known as a *virulence factor*. The branching rates of the model, $X_t = \big(X_1(t), \ldots, X_4(t)\big)$, are given as

$$(x_1,x_2,x_3,x_4) \mapsto \begin{cases} (x_1+1,x_2,x_3,x_4) & \lambda_1 x_1 \\ (x_1,x_2+1,x_3,x_4) & \lambda_2 x_2 \\ (x_1,x_2,x_3+1,x_4) & \lambda_3 x_3 \\ (x_1,x_2,x_3,x_4+1) & \lambda_4 x_4 \\ (x_1-1,x_2+1,x_3,x_4) & q_{12} x_1 \\ (x_1+1,x_2-1,x_3,x_4) & q_{21} x_2 \\ (x_1,x_2,x_3-1,x_4+1) & q_{34} x_3 \\ (x_1,x_2,x_3+1,x_4-1) & q_{43} x_4 \\ (x_1,x_2+1,x_3-1,x_4) & q_{32} x_3 \\ (x_1,x_2+1,x_3,x_4-1) & q_{42} x_4, \end{cases}$$

where λ_i, $i = 1, \ldots, 4$, are the speciation rates of type i strains, q_{ij}, $i, j = 1, \ldots, 4$, are the transition rates from type i to type j, and the initial state X_0 is assumed to be either $(0,0,1,0)$ or $(0,0,0,1)$. A diagrammatic representation of various parameters used in the model is given in Figure 3.3. We test out different versions of the model by either utilizing both q_{32} and q_{42}, or letting one of them to be equal to zero. This is because *E. coli* strains are known to travel from the gastrointestinal tract, their natural habitat, to the bladder and cause urinary tract infections [45]. To estimate various parameter values of X_t, for a total of 9 virulence factors, we make use of the MuSSE model [11] incorporated in the package '*diversitree*' of the software `R` [38]. We also obtain limiting values for the proportions of pathogenic and nonpathogenic bacterial strains in different states. From the analysis, we inferred some interesting results that are listed below.

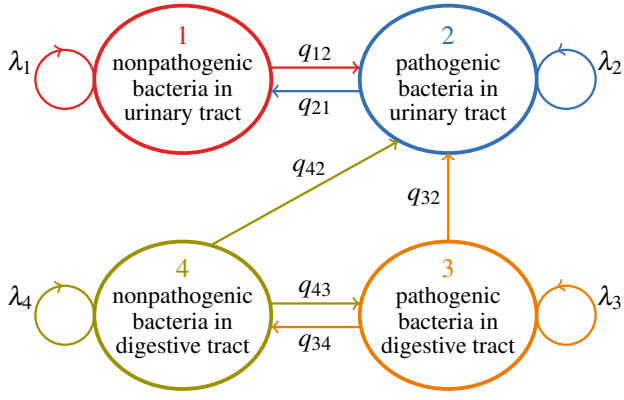

Figure 3.3. Diagram showing various states and parameters of the 4-type model. Here, λ_i represent the speciation rates of type i strains, q_{ij} represent the rates of transition from type i to type j, and extinction rates are assumed to be zero.

a) We can successfully use multitype branching processes to answer biological questions regarding virulent behavior in bacterial strains. For instance, we deduced that *E. coli* bacteria lose their pathogenic capability at higher rates as compared to gaining it, in both urinary and digestive tracts of human hosts. This confirms the fact that since it is costly to maintain virulence factors, *E. coli* do not remain pathogenic unless the conditions are agreeable for host invasion. We also concluded that virulent *E. coli* strains speciate at faster rates as compared to nonvirulent strains, confirming the result that was previously obtained in [4].

b) The MuSSE model, which incorporates maximum likelihood techniques, can be used to effectively estimate parameter values in multitype models, provided that the number of unknown parameters is 'reasonable'. During the analysis, we saw that in most of the parameter estimation scenarios, the models were a better fit to the given data set, if the parameters were constrained in some manner, for example, when some parameters were set to be equal to zero or pairs of parameters were set to be equal to one another.

c) The estimated parameter values can be used to provide information on the almost sure limits of the proportion of bacterial strains in various states. The probability of maintaining virulence corresponds to the sum of limiting proportions of bacteria in the two pathogenic states. From the analysis, we concluded that this probability varied significantly in bacterial strains and it depended on the virulence factor under consideration.

d) The application of limit theorems can be used to obtain plausible confidence regions, in the form of confidence ellipsoids, for the long term proportions of *E. coli* strains in different states of the models.

Contribution: I wrote the paper and carried out the analysis on *E. coli* strains data set.

3.3 Paper III

In Paper III, we study different forms of a supercritical, binary state, continuous time Markov branching process. These include 1) a classical two-type branching process with one sided transitions, and 2) a modified version of the basic model, altered to include the effect of population size dependence. We review existing functional and central limit theorems regarding the long time behavior of two-type branching models [2, 19], and also provide new results concerning the existence and uniqueness of solutions to stochastic equations. Moreover, as an illustration for the theoretical part of the Paper, we apply our results to a model of binary branching and one sided transition (similar to the one discussed in Paper I).

We consider a two-type branching process $\{X_t : t \geq 0\}$, which satisfies a stochastic equation of the form

$$X_t = X_0 + \int_0^t AX_s\,ds + M_t, \quad t \geq 0, \quad A = \begin{bmatrix} \gamma_0 & 0 \\ \delta & \gamma_1 \end{bmatrix},$$

where $X_t = (X_t^0, X_t^1)'$, with X_t^0 and X_t^1 being the number of type 0 and type 1 units at time t, respectively, $X_0 = (1,0)'$, $\gamma_0 > 0$ and $\gamma_1 < \gamma_0$ are the net growth rates of type 0 and type 1 particles, respectively, $\delta \geq 0$ is the the average rate of transition from type 0 to type 1, and $(M_t)_{t\geq 0}$ is a martingale term. We also examine population size dependent versions of the branching process given by

$$X_t = X_0 + \int_0^t A(X_s)X_s\,ds + M_t, \quad t \geq 0,$$

in which the entries of A have been replaced with suitable state-dependent functions. We let $M_z^e(ds,du)$ be Poisson random measures on $\mathbb{R}_+^2$ with intensity $q_z^e\,dsdu$, where $e \in \{0,1\}$ denotes the two types, and $z = (z^0, z^1) \in \mathcal{Z}_e$ are the set of jumps that can occur due to type e branching. Then, the basic branching process is proved to be a strong solution of the stochastic equation

$$X_t = X_0 + \sum_e \sum_{z\in\mathcal{Z}_e} z \int_0^t \int_0^{X_{s-}^e} M_z^e(ds,du),$$

while the population size dependent branching process, under additional assumptions, is shown to be a unique strong solution X_t of

$$X_t = X_0 + \sum_e \sum_{z\in\mathcal{Z}_e} z \int_0^t \int_0^{\beta_z^e(X_{s-})} M_z^e(ds,du), \quad \beta_z^e(x) = q_z^e(x)x^e/q_z^e.$$

Limit results: The eigenvalues γ_0 and γ_1 of the matrix A, and the left and right eigenvectors, u and v, corresponding to the largest eigenvalue γ_0, characterize the limit behavior of the basic branching model X_t. The fundamental limit result for supercritical branching processes is given as

$$e^{-\gamma_0 t}X_t \xrightarrow{a.s.} Wv,$$

where W is a nonnegative random variable [1]. Additional limit results, obtained from [19] and adapted to our basic branching model, are given below.
a) If $\gamma_1 < \gamma_0/2$, then as $t \to \infty$,

$$e^{-\gamma_0 t/2}\left(X^1_{t+x} - \delta X^0_{t+x}/(\gamma_0 - \gamma_1)\right) \xrightarrow{d} W^{1/2} e^{\gamma_0 x/2} U(x),$$

where $U(x)$ is an Ornstein-Uhlenbeck process and $x \in \mathbb{R}$.
b) If $\gamma_1 = \gamma_0/2$, then as $t \to \infty$,

$$t^{-1/2} e^{-\gamma_0 xt/2}\left(X^1_{xt} - 2\delta X^0_{xt}/\gamma_0\right) \xrightarrow{d} W^{1/2}\,\sigma B(x),$$

where σ^2 represents the variance and $\{B(x), x \geq 0\}$ is a Brownian motion.
c) If $\gamma_0/2 < \gamma_1 < \gamma_0$, then as $t \to \infty$,

$$e^{-\gamma_1 t}\left(X^1_t - \delta X^0_t/(\gamma_0 - \gamma_1)\right) \xrightarrow{a.s.} W_1,$$

where W_1 is a real-valued random variable.

We also study scaled branching processes and use the central limit theorem to understand the fluctuations around the limit of such processes. Let $(X^{(j)}_t)$, $j \geq 1$, be i.i.d. copies of the branching process with $X^{(j)}_0 = (1,0)'$, and let $\widehat{X}^n_t = n^{-1}\sum_{j=1}^n X^{(j)}_t$, with $\widehat{X}^n_0 = (1,0)'$. Letting $\widehat{X}^n_0 \xrightarrow{a.s.} x_0$ as $n \to \infty$, we show that $\widehat{X}^n_t$ converges to $x_t = \mathbb{E}(X_t | X_0 = x_0) = e^{At}x_0$, which is a solution of $x_t = x_0 + \int_0^t A x_s\,ds$. The fluctuation process given by $Y^n_t = \sqrt{n}(\widehat{X}^n_t - x_t)$, is proved to converge to Y_t, which is a solution of the SDE

$$Y_t = Y_0 + \int_0^t A Y_s\,ds + \sum_e \sum_{z \in \mathscr{Z}_e} z \int_0^t \sqrt{q^e_z x^e_s}\,dB^{z,e}_s,$$

where $\{B^{z,e}_t\}$ is a family of independent standard Brownian motions.

We obtain a similar result for the population size dependent process, using [10]. Thus, the associated scaled process $\widehat{X}^n_t$ for this case converges to x_t, which is a solution of $x_t = x_0 + \int_0^t F(x_s)\,ds$, where $F(x) = A(x)x$. Letting the normalized deviation process $V^n_t = \sqrt{n}(\widehat{X}^n_t - x_t)$, and assuming that V^n_0 converges to a constant v_0, we see that V^n_t converges to V_t, which solves

$$V_t = v_0 + \int_0^t \partial F(x_s) V_s\,ds + \sum_e \sum_{z \in \mathscr{Z}_e} z \int_0^t \sqrt{q^e_z(x_s) x^e_s}\,dB^{z,e}_s, \quad t \geq 0,$$

where $\partial F(x) = \begin{bmatrix} \frac{\partial}{\partial x^0}F^0(x) & \frac{\partial}{\partial x^1}F^0(x) \\ \frac{\partial}{\partial x^0}F^1(x) & \frac{\partial}{\partial x^1}F^1(x) \end{bmatrix}$, with $F(x) = \left(F^0(x), F^1(x)\right)'$.

Contribution: I participated in the writing process of the paper, and contributed towards applying the limit results to a binary branching and transition model.

3.4 Paper IV

The motivation behind Paper IV was to extend the diversification model studied in Paper I, in which species are allowed to grow in size without bound, to a more realistic model that prevented unlimited increase in the number of species. For that purpose, we first constructed a framework in an evolutionary time scale, which provided a correlation between processes involved in species diversification models and processes in population genetics models. Specifically, we analyzed a two-trait species branching model over a scale of generations, and compared the proportion of species carrying one of the traits, to allele frequencies in a bi-allelic Wright-Fisher (WF) diffusion process. Then, we applied the population genetics approach and the long time scaling regime to not only discuss different cases of density-dependent processes based on various parameter assumptions, but also to compare our work with similar types of models that have been studied previously, for example, in [35, 36].

Scaling of parameters in the bi-allelic (0 and 1) WF model is attained in a manner similar to the one described in Section 1.2. Also, recall that as the population size $N \to \infty$, the scaled WF frequency process ξ_t^N converges to the WF diffusion ξ_t, which is given by the solution of the SDE

$$d\xi_t = \gamma\xi_t(1-\xi_t)\,dt - \rho_{01}\xi_t\,dt + \rho_{10}(1-\xi_t)\,dt + \sqrt{\xi_t(1-\xi_t)}\,dB_t,$$

where ρ_{01} and ρ_{10} are the mutation rates scaled with N, γ is the scaled selection coefficient, and B_t is a Brownian motion. For the construction of the species model, we consider a two-type, continuous time branching process, $X_u = (K_u, L_u)$, $u \geq 0$, with K_u being the number of type 0 species and L_u the number of type 1 species. We also consider an alternative representation (P_u, R_u), where

$$P_u = \frac{K_u}{K_u + L_u} \quad \text{and} \quad R_u = K_u + L_u$$

is the fraction of type 0 species and the total number of species, respectively. We let λ_0 and λ_1 be the speciation rates, μ_0 and μ_1 the extinction rates of type 0 and type 1 species, respectively. The species are also allowed to transition both ways, hence, type 0 (type 1) species transition to type 1 (type 0) with rate $\delta_{01}(\delta_{10})$. Let $X_u^{(n)} = (K_u^{(n)}, L_u^{(n)})$ denote the branching process scaled by n, with $\lambda_i^{(n)}$, $\mu_i^{(n)}$, $i = 0,1$, and $\delta_{01}^{(n)}$, $\delta_{10}^{(n)}$ the corresponding parameters of speciation, extinction and transition for the two types. As $n \to \infty$,

$$\mu_i^{(n)} \to \mu_i, \quad \lambda_i^{(n)} \to \mu_i, \quad n(\lambda_i^{(n)} - \mu_i^{(n)}) \to \beta_i, \quad n\delta_{01}^{(n)} \to \rho_{01}, \quad n\delta_{10}^{(n)} \to \rho_{10},$$

where μ_i, β_i, $i = 0,1$, and ρ_{01}, ρ_{10} are the scaled parameters which control speciation/extinction, net diversification and transition of species on an evolutionary time scale. Note that we take μ_i as the reference parameter for both speciation and extinction events for trait i, and let β_i be the evolutionary net

diversification rate of trait i, $i = 0,1$. The process $X_t^n = n^{-1}X_{nt}^{(n)}$ gives the frequency of the two types of species on an evolutionary time scale of nt generations. Correspondingly, the alternative representation is

$$P_t^n = \frac{K_{nt}^{(n)}}{K_{nt}^{(n)} + L_{nt}^{(n)}} \quad \text{and} \quad R_t^n = \frac{K_{nt}^{(n)} + L_{nt}^{(n)}}{n},$$

where P_t^n is comparable to the frequency process ξ_t^N in the WF model. As $n \to \infty$, X_t^n converges to a continuous state branching process $\mathscr{X}_t$, and (P_t^n, R_t^n) converges to $(\mathscr{P}_t, \mathscr{R}_t)$, which is the solution of a system of SDEs given by

$$\begin{aligned} d\mathscr{P}_t &= \mathscr{P}_t(1-\mathscr{P}_t)\Big(\beta_0 - \beta_1 - \frac{2(\mu_0-\mu_1)}{\mathscr{R}_t}\Big)\,dt - \rho_{01}\mathscr{P}_t\,dt + \rho_{10}(1-\mathscr{P}_t)\,dt \\ &\quad + \sqrt{2\mathscr{P}_t\big(1-\mathscr{P}_t\big)\big(\mu_0(1-\mathscr{P}_t) + \mu_1\mathscr{P}_t\big)\frac{1}{\mathscr{R}_t}}\,dB_t^-, \\ d\mathscr{R}_t &= \mathscr{R}_t\big(\beta_0\mathscr{P}_t + \beta_1(1-\mathscr{P}_t)\big)\,dt + \sqrt{2\mathscr{R}_t\big(\mu_0\mathscr{P}_t + \mu_1(1-\mathscr{P}_t)\big)}\,dB_t^+, \end{aligned}$$

where B_t^-, B_t^+ are standard Brownian motions. From the above system, we deduce the following results.

a) The equation in $\mathscr{P}_t$ is analogous to the WF diffusion process with population size dependent selection, i.e., comparing with the WF diffusion equation, the selection coefficient γ equals $\beta_0 - \beta_1 - 2(\mu_0 - \mu_1)/\mathscr{R}_t$. On the other hand, the equation in $\mathscr{R}_t$ is a measure of the total species richness, with trait dependent drift and diffusion functions.

b) If we let $\beta_0 = \beta_1$, the equation in $\mathscr{P}_t$ is a WF diffusion process with population size dependent drift and diffusion parameters, whereas the equation in $\mathscr{R}_t$ is a Feller-type diffusion process in which the diffusion parameter is regulated by the trait proportions.

c) If we let $\mu_0 = \mu_1$, the equation in $\mathscr{P}_t$ is again a WF diffusion, but only the genetic drift term is population size dependent, while the equation in $\mathscr{R}_t$ is a Feller-type diffusion process in which the drift function is modulated by the trait frequencies. Moreover, the Brownian motions B^- and B^+ are independent when $\mu_0 = \mu_1$.

d) If we let $\beta_0 = \beta_1$ and $\mu_0 = \mu_1$, the equation in $\mathscr{P}_t$ is a WF diffusion with mutation and population size dependent genetic drift but no selection, while the equation in $\mathscr{R}_t$ is a standard Feller diffusion equation.

We also study the effect of diversity-dependence in species families, and examine three cases as follows.

1) We consider a logistic type model in which the scaled diversification parameter, β_i, $i = 0,1$, is replaced by $\beta_i(1 - \mathscr{R}_t/c)_+$, where c represents the carrying capacity. Putting $\mathscr{R}_t/c = \widetilde{\mathscr{R}}_t$, and then letting $c \to \infty$, we deduce

from the resulting ODE in $\mathscr{P}_t$, that selection acts on the net diversification rates β_i, and is regulated by the population size process.

2) We consider another logistic model, such that the diversity-dependence acts directly on the speciation parameter, $\lambda_i^{(n)}$, $i = 0,1$, in the population time scale. We derive again the SDE system, put $\mathscr{R}_t/c = \widetilde{\mathscr{R}}_t$, and then letting $c \to \infty$, we infer from the resulting ODE system that under the conditions $\rho_{01} = \rho_{10} = 0$, $\beta_0 < \mu_0$, $\beta_1 < \mu_1$, as $t \to \infty$ if $\beta_0/\mu_0 > \beta_1/\mu_1$ then type 0 gets fixed in the population while $\beta_0/\mu_0 < \beta_1/\mu_1$ leads to type 1 getting fixed.

3) We replace μ_i by $\mu_i\mathscr{R}_t$, $i = 0,1$, that is, population size dependence is achieved by letting the extinction rates increase with increasing diversity. As a result, the equation in $\mathscr{P}_t$ becomes independent of the total size of the system, and is thus analogous to the WF diffusion process, but with a frequency-dependent variance function. Moreover, the equation in $\mathscr{R}_t$ is found to have a solution which is a form of a geometric Brownian motion.

Contribution: I performed the background research on density dependent species diversification, and helped writing the introduction and various diversity dependent cases in the paper.

4. Sammanfattning på svenska

Matematiska modeller ger kvantitativ och kvalitativ information om naturligt förekommande fenomen. Stokastiska modeller baserade på slumpmässiga biologiska processer, som evolutionära genetiska mekanismer och befolkningstillväxt, kan användas för att analysera variabiliteten som kännetecknar dessa processer. I denna avhandling använder vi olika stokastiska modeller, i synnerhet Markov-förgreningsmodeller, för att försöka förstå egenskapernas utveckling och deras effekt på artens födelse och överlevnad. För detta ändamål undersöker vi först den allmänna teorin om diskreta och kontinuerliga tidsförgreningsprocesser, och sedan tillämpar vi det på olika biologiska modeller. Vi presenterar våra resultat i fyra artiklar, av vilka korta sammanfattningar presenteras nedan.

I Artikel I modellerar vi samtidigt effekten av egenskaper på artdiversifieringshastigheter och på molekylär evolution. För detta ändamål utvecklar vi först en probabilistisk modelleringsram för ett slumpmässigt, binärt träd, där antalet arter och deras egenskaper är representerade genom en asymmetrisk, två-typ, kontinuerlig tid Markov-förgreningsprocess. Därefter utforskar vi olika egenskaper hos de reducerade träden, såsom de förväntade storlekarna och grenarnas längder. Slutligen beskriver vi en egenskapsberoende substitutionsprocess längs grenarna på artens träd, genom en Poissonprocess. Vi demonstrerar också våra metoder genom att tillämpa dem på en fylogeni av växtarter och drar slutsatsen att egenskapsberoende diversifieringsprocesser kan ha en stark inverkan på molekylär evolution.

Vi utökar de två typerna av förgreningsprocesser som studerats ovan, till fyra typer av Markov-förgreningsmodeller i Artikel II, för att analysera patogena egenskaper hos *E. coli* bakteriestammar. Vi studerar först flera generella egenskaper för förgreningsprocesser med flera typer. Vi undersöker också några grundläggande befintliga teorem som förklarar beteendet hos förgreningsprocesser i lång tidshorisont. Sedan tillämpar vi våra resultat på en publicerad klinisk data-mängd av virulenta och icke-virulenta *E. coli* bakteriestammar. Vi kan dra slutsatsen från analysen att kända maximala sannolikhetstekniker kan utnyttjas för att effektivt estimera parametervärdena i flera typer av förgreningsmodeller och de uppskattade parametrarna kan sedan användas för att ge information om gränserna för proportionerna av bakteriestammar i olika tillstånd av modeller.

Artikel III ger ytterligare utökningar av de två-typ, asymmetriska förgreningsmodellerna som studerades tidigare i Artikel I, till förgreningsmodeller

som innefattar beroendet av populationsstorlek. Således får arterna i detta fall inte växa obegränsat. Istället styrs deras tillväxt av själva populationsstorleken. Vi analyserar stokastiska ekvationer och begränsar resultaten för både grundprocessen och dess populationsstorleksberoende version. Vi undersöker även befintliga teorem beträffande beteendet hos dessa processer under en lång tidsperiod och under olika skalningar. Det visas att processens gränsbeteende kan kännetecknas av egenvärdena och egenvektorerna hos motsvarande medelvärdes-matris. Dessutom analyserar vi skalade versioner av förgreningsmodellerna med stora talens lag och beskriver fluktuationsprocessen runt gränserna för dessa processer med hjälp av centrala gränsvärdessatsen.

Motiverad av den matematiska analysen av densitetsberoende modeller i Artikel III analyserar vi liknande typer av processer i Artikel IV, den här gången med hjälp av begrepp från populationsgenetik. För detta ändamål konstruerar vi först ett ramverk för modellering på en evolutionär tidsskala. Därför skalas rymd, tid och modellparametrar av speciering, utdöende och övergång över en lång geologisk tidsskala. Dessutom får vi en diffusionsapproximation av processen för att tillhandahålla en analogi med Wright Fisher-diffusionsprocesser. Slutligen tillämpas det analytiska ramverk som utvecklas på olika populationsberoende modeller.

5. Acknowledgements

First of all, I express my utmost gratitude to my PhD supervisor, Professor Ingemar Kaj, for his support, dedication and guidance. I have been extremely lucky to have him as my supervisor, whose hard work and brilliance has been a source of constant inspiration. For the last five years, his office door has always been open for me, without the need of setting any prior time for a meeting. Never before have I had the opportunity to work with someone who is not only exceptional as a researcher, but also an outstanding man in other walks of life. He has been my teacher, my advisor, my mentor and my friend. Thank you Ingemar, for being kind and understanding, for selflessly giving me infinitely many hours of your valuable time, for taking responsibility and doing all the worrying so that I didn't have to, and for granting me the golden opportunity of coming to Sweden, for which I will be forever in your debt.

I would also like to thank my second supervisor, Professor Martin Lascoux; without his biological expertise and knowledge, this thesis would not have been possible. His anecdotes from all over the world and his cheerful disposition always made me feel welcome at the biology department. I also extend my gratitude to Sylvain Glémin, an excellent researcher at EBC, for putting up patiently with my incessant questions, replying promptly to all my emails, and for being so nice and kindhearted. The fruitful discussions at EBC with Martin and Sylvain were a source of valuable help and motivation.

Krzysztof Bartoszek, I thank you for including me in the project with your Polish colleagues. Without your help and support, Paper II would not be a part of this thesis. Your code writing skills in `R` proved to be of great benefit. Thank you for helping me out and patiently answering my questions not only about Mathematics, but also about life in Sweden. I was fortunate to attend the academic teaching course alongside you.

I am much obliged to Uppsala University, the Department of Mathematics and the Centre for Interdisciplinary Mathematics for giving me the opportunity of getting my PhD in Sweden. My special gratitude to the heads of the Mathematics department: Johan Tysk, Warwick Tucker, and Volodymyr Mazorchuk and various directors of CIM: Elisabeth Larsson, Erik Ekström and Jordi-Lluis Figueras, for not only funding my PhD, but also for sending me to great conferences and giving me a chance to see other countries of Europe.

I express my gratitude to everyone in the Mathematics administration: Lisbeth Juso, Alma Kirlic, Olga Kaj, Fredrik Lannergård, Lina Flygerfeldt, Mattias Landelius, Shumin Pan, thank you for your constant effort and support.

Inga-Lena Assarsson, thank you for being so kind and helpful, and always providing me with assistance. Elisabeth Bill, mere words cannot show you how grateful I feel towards you; you have been my friend, companion and confidant. Thank you for always having time for me, for helping me survive in Sweden, for giving me great advice, and for answering my ridiculous queries about PhD and government and taxes and insurance and everything else.

I thank all my teachers at different departments of Uppsala University with whom I took various courses over the past five years — Martin Lascoux, Sylvain Glémin, Mattias Jakobsson, Thijessen Naidoo, Jesper Svedberg, Per Sjödin, Claus Rüffler, Erik Broman, Stefan Pålsson, Silvelyn Zwanzig, Jannika Chronholm, Thomas Lennerfors, Maja Elmgren, Ulf Holmbäck, Cecilia Gustavsson, Felix Ho, Behrang Mahjani, Carl Nettelblad, Svante Janson — you have helped mould me into the person I am today.

My deep appreciation to Inger Sigstam for always respecting my teaching wishes. To the teachers I had the pleasure of working with — Yevgen, Andreas, Rebekka, Ulf, Natalia, Fredrik, Tobias, Magnus, Oleksandra, Julian, Fabian, Laertis, Yu Liu, Orjän, Thomas, Maksim, Jens — thank you all, I had a great time working along side you. A special thanks to Sergio Duchens who helped me with many tight scheduling conflicts.

I give my heartfelt thanks to all my friends and colleagues; Giane, Dmytro, Noemi, Viktoria, Sebastian, Dan, David, Björn, Elin, Yuqiong, Emilia, Eva, Ylva, Zahedeh, Juozas, Konstantinos, Raazesh, Veronica, Brendan, Gustav, Carina, Lili, Linnéa, and my neighbors in house 7; Colin, Fabian, Johanna, Rolf, Cecilia, Svante, Xing, Fiona and Marcus, for being a part of my every day life. A special thanks to my office mate Azza: for understanding all my troubles, Yevgen: for helping with Stats problems and for giving me great advice, Jakob and Andrea: for assisting me with thesis related issues, Erik: for being my loyal friend, Helena, Victor and Andreas: for helping me translate the English $\rightarrow$ Swedish summary for this thesis, Debleena: for her motivating chats in Urdu, Filipe and Tilo: for being compassionate neighbors, Anna: for being so sweet and helpful, Marta and Natalia: for guiding me in my early years at UU. Thanks also to my friends, Rabia, Rabab and Maria, for extending their help in difficult times. Finally, I thank the cleaning crew at the Math department for always keeping my office neat and tidy.

Last but not least, my dearest beloved family — my mother and father, my grandma, my younger brothers Abdullah and AbdurRahman — thank you all for your constant support and prayers. Without your love and encouragement, this PhD and my life, so far away from home, would not have been possible.

6. References

[1] ATHREYA, K. B. Some results on multitype continuous time Markov branching processes. *Ann. Math. Stat. 39* (1968), 347–357.

[2] ATHREYA, K. B., AND NEY, P. E. *Branching Processes*. Springer-Verlag, Berlin, 1972.

[3] BACON, C. D., VELASQUEZ-PUENTES, F. J., HOORN, C., AND ANTONELLI, A. Iriarteeae palms tracked the uplift of Andean Cordilleras. *J. Biogeogr. 45* (2018), 1653–1663.

[4] BARTOSZEK, K., MAJCHRZAK, M., SAKOWSKI, S., SZELIGOWSKA, A. B. K., KAJ, I., AND PARNIEWSKI, P. Predicting pathogenicity behavior in *Escherichia coli* population through a state dependent model and TRS profiling. *PLOS Comput. Biol. 14* (2018), e1005931.

[5] BEAULIEU, J. M., AND O' MEARA, B. C. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Syst. Biol. 65* (2016), 583–601.

[6] BENTON, M. J., AND EMERSON, B. C. How did life become so diverse? The dynamics of diversification according to the fossil record and molecular phylogenetics. *Paleontology 50* (2007), 23–40.

[7] BERGMANN, P. J., AND IRSCHICK, D. J. Vertebral evolution and the diversification of squamate reptiles. *Evolution 66* (2011), 1044–1058.

[8] CINLAR, E. *Probability and Stochastics*. Springer, New York, 2011.

[9] DAVIS, M. P., MIDFORD, P. E., AND MADDISON, W. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC Evol. Biol. 13* (2013), 38.

[10] ETHIER, S. N., AND KURTZ, T. G. *Markov Processes, Characterization and Convergence*. John Wiley & Sons, New York, 1986.

[11] FITZJOHN, R. G. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol. 3* (2010), 1084–1092.

[12] FITZJOHN, R. G. Quantitative traits and diversification. *Syst. Biol. 59* (2012), 619–633.

[13] FITZJOHN, R. G., GOLDBERG, E. E., AND FORD, K. M. *Diversitree: comparative phylogenetic analyses of diversification*, 2012. R package version 0.9-2.

[14] GAMISCH, A. Notes on the statistical power of the binary state speciation and extinction (BiSSE) model. *Evol. Bioinform. 12* (2016), 165–174.

[15] GAN, X., AND WAXMAN, D. Singular solution of the Feller diffusion equation via a spectral decomposition. *Phys. Rev. E 91* (2015), 012123.

[16] GOLDBERG, E. E., AND IGIĆ, B. Tempo and mode in plant breeding system evolution. *Evolution 66* (2012), 3701–3709.

[17] GOLDBERG, E. E., LANCASTER, L. T., AND REE, R. H. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Syst. Biol. 60* (2011), 451–465.

[18] IBE, O. C. *Markov Processes for Stochastic Modeling*. Academic Press, New York, 2009.
[19] JANSON, S. Functional limit theorems for multitype branching processes and generalized Polya urns. *Stoch. Proc. Appl. 110* (2004), 177–245.
[20] KARLIN, S., AND TAYLOR, H. M. *A First Course in Stochastic Processes*. Academic Press, New York, 1975.
[21] KARLIN, S., AND TAYLOR, H. M. *A Second Course in Stochastic Processes*. Academic Press, New York, 1981.
[22] KIRCHHOFF, K. N., HAUFFE, T., STELBRINK, B., ALBRECHT, C., AND WILKE, T. Evolutionary bottlenecks in brackish water habitats drive the colonization of fresh water by stingrays. *J. Evol. Biol. 30* (2017), 1576–1591.
[23] LAGOMARSINO, L. P., CONDAMINE, F. L., ANTONELLI, A., MULCH, A., AND DAVIS, C. C. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytol. 210* (2016), 1430–1442.
[24] LAMPERTI, J. Continuous state branching processes. *Bull. Amer. Math. Soc. 73* (1967), 382–386.
[25] LESLIE, A. B., BEAULIEU, J. M., CRANE, P. R., AND DONOGHUE, M. J. Explaining the distribution of breeding and dispersal syndromes in conifers. *Proc. R Soc. B 280* (2013), 20131812.
[26] LI, Y. Exploring the power of BiSSE. Master's thesis, Linköping University, 2018.
[27] MADDISON, W. P. Confounding asymmetries in evolutionary diversification and character change. *Evolution 60* (2006), 1743–1746.
[28] MADDISON, W. P., AND FITZJOHN, R. G. The unsolved challenge to phylogenetic correlation tests for categorical characters. *Syst. Biol. 64* (2015), 127–136.
[29] MADDISON, W. P., MIDFORD, P. E., AND OTTO, S. P. Estimating a binary character's effect on speciation and extinction. *Syst. Biol. 56* (2007), 701–710.
[30] MARSHALL, C. R., AND QUENTAL, T. B. The uncertain role of diversity dependence in species diversification and the need to incorporate time-varying carrying capacities. *Phil. Trans. R. Soc. B 371* (2016), 20150217.
[31] MORLON, H. Phylogenetic approaches for studying diversification. *Ecol. Lett. 17* (2014), 508–525.
[32] NIELSEN, R., AND SLATKIN, M. *An introduction to population genetics*. Sinauer Associates, Inc., Sunderland, Msssachusetts, 2013.
[33] OTTO, S. P., AND DAY, T. *A biologist's guide to mathematical modeling in ecology and evolution*. Princeton University Press, New Jersey, 2007.
[34] PARDOUX, E. Continuous branching processes: the discrete hidden in the continuous: Dedicated to Claude Lobry. *ARIMA Rev. Afr. Rech. Inform. Math. Appl. 9* (2008), 211–229.
[35] PARSONS, T. L., AND QUINCE, C. Fixation in haploid populations exhibiting density dependence I: The non-neutral case. *Theor. Popul. Biol. 72* (2007), 121–135.
[36] PARSONS, T. L., QUINCE, C., AND PLOTKIN, J. B. Absorption and fixation times for neutral and quasi-neutral populations with density dependence. *Theor. Popul. Biol. 74* (2008), 302–310.
[37] PEÑA, C., AND ESPELAND, M. Diversity dynamics in Nymphalidae

butterflies: Effect of phylogenetic uncertainty on diversification rate shift estimates. *PLoS ONE 10* (2015), e01209282.

[38] R CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[39] RABOSKY, D. L. Diversity-dependence, ecological speciation, and the role of competition in macroevolution. *Annu. Rev. Ecol. Evol. Syst. 44* (2013), 481–502.

[40] RABOSKY, D. L., AND GLOR, R. E. Equilibrium speciation dynamics in a model adaptive radiation of island lizards. *PNAS 107* (2010), 22178–22183.

[41] RABOSKY, D. L., AND GOLDBERG, E. E. FiSSE: A simple nonparametric test for the effects of a binary character on lineage diversification rates. *Evolution 71* (2017), 1432–1442.

[42] RABOSKY, D. L., AND GOLGBERG, E. E. Model inadequacy and mistaken inferences of trait-dependent speciation. *Syst. Biol. 64* (2015), 340–355.

[43] ROLLAND, J., AND SALAMIN, N. Niche width impacts vertebrate diversification. *Global Ecol. Biogeogr. 25* (2016), 1252–1263.

[44] SERRANO, M. L., ROLLAND, J., CLARK, J. L., SALAMIN, N., AND PERRET, M. Hummingbird pollination and the diversification of angiosperms: an old and successful association in Gesneriaceae. *Proc. R. Soc. B 284* (2017), 20162816.

[45] SIVICK, K. E., AND MOBLEY, H. L. T. Waging war against uropathogenic *Escherichia coli*: winning back the urinary tract. *Infect. Immun. 78* (2010), 568–585.

[46] TAYLOR, H. M., AND KARLIN, S. *An Introduction to Stochastic Modeling*. Academic Press, New York, 1984.

[47] TOUSSAINT, E. F. A., HENDRICH, L., SHAVERDO, H., AND BALKE, M. Mosaic patterns of diversification dynamics following the colonization of Melanesian islands. *Sci. Rep. 5* (2015), 16016.

[48] YANG, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci. 13* (1997), 555–556.

[49] YANG, Z. *Computational Molecular Evolution*. Oxford University Press, Oxford, 2006.

Part II:
Research papers

This part of the thesis consists of four appended research papers. Paper I deals with the application of a supercritical, two-type, continuous time Markov branching process to the evolution of binary traits on a random species tree, as well as a molecular evolutionary process which runs as a Poisson process along all branches of the tree. The study of two-type branching models on the naturally occurring process of mating systems evolving in plant species, then leads to further applications in the form of: (i) analysis of pathogenicity in clinical bacterial strains data sets using four-type Markov branching models in Paper II, and (ii) the extension of the application of branching processes in populations with unbounded growth, to modified population models in which the net growth is restricted. This motivates Paper III in which continuous time branching models with population size dependence are analyzed in a framework of stochastic equations, and fundamental theorems for the behavior of these models under a long time limit are surveyed. The investigation of such systems demands even further applications, and this finally leads to Paper IV, in which, under a long geological time scale, various cases of diversity-dependent species models are explored and also compared with frequency processes in population genetics models.